Journal of Economic Dynamics and Control 15 (1991) 275–284. North-Holland

A critique of the application of unit root tests

John H. Cochrane*
University of Chicago, Chicago, IL 60637, USA

Received December 1988, final version received April 1990

This paper exploits the fact that any time series with a unit root can de decomposed into a stationary series and a random walk. Since the random walk component can have arbitrarily small variance, tests for unit roots or trend stationarity have arbitrarily low power in finite samples. Furthermore, there are unit root processes whose likelihood functions and autocorrelation functions are arbitrarily close to those of any given stationary processes and vice versa, so there are stationary and unit root processes for which the result of any inference is arbitrarily close in finite samples.

1. Introduction

Time series in macroeconomics and finance are commonly detrended or first differenced before they are analyzed. For many statistical purposes it seems important to distinguish which procedure is appropriate for a given series or group of series. For example, Nelson and Kang (1981) observed that if one detrends data that are actually generated by a random walk, one will infer a time series structure that is not in fact present. Conversely, if one takes first differences of data that are actually trend-stationary, one introduces a unit root into the moving average representation of the series.

Many tests have been devised to distinguish whether a series is difference-stationary (or contains a unit root in its autoregressive representation) and should be first differenced or whether it is trend-stationary and should be detrended. Dickey and Fuller (1979), Dickey, Bell, and Miller

*I thank John Huizinga and Jim Stock for detailed and valuable comments, and Robert Townsend for the discussion that provoked this paper. This research was partially supported by National Science Foundation grant SES-8809912.

(1986), Phillips (1987), Phillips and Perron (1986), and Lo and MacKinlay (1988) are a small sample.

Unit root tests are coming into widespread use in macroeconomics. It is now commonplace to pretest series to classify them as either trend-stationary or difference-stationary, and then impose that form in subsequent analysis. That subsequent analysis can include tests whose validity or form depends on the trend or difference stationarity of the series, estimation which imposes one or the other form, or direct reading of economic importance into the trend or difference stationarity of the series.

However, Schwert (1987), Lo and MacKinlay (1989), Blough (1988), and others have documented that tests for unit roots or trend stationarity can have low power against some specific alternatives. Essentially, they show that tests for a unit root have low power in finite samples against the local alternative of a root close to but below unity. Section 2 of this paper shows that any trend-stationary process has similar local unit root alternatives. These are formed by adding random walk with an arbitrarily small innovation variance to a given stationary series. Therefore, any test for a unit root or trend stationarity must have arbitrarily low power against some alternatives of the other class.

By itself, this is not a particularly serious problem. Any test whether a continuous parameter θ is equal to some value θ_0 has arbitrarily low power against alternatives $\theta_0 - \varepsilon$ in finite samples. However, in most such cases, the difference between θ_0 and $\theta_0 - \varepsilon$ is not particularly important, from either a statistical or an economic perspective. What makes the unit root question special is the impression that important statistical and economic issues hang on the difference between a root of precisely 1 and a root of $1 - \varepsilon$, or between a random walk component with innovation variance precisely 0 and a random walk component with innovation variance ε, in a way that (say) an elasticity of demand of -1.0 is not importantly different from an elasticity of -0.99.

The apparent statistical importance comes from the fact that the asymptotic distribution theory of many estimators or test statistics is quite sensitive to the presence or absence of a unit root, and in fact is discontinuous as the largest root approaches 1 or the random walk component approaches 0. Therefore, many authors have thought it important to pretest a series for inclusion into one or the other class, even if that test has low power, so that the 'correct' asymptotic distribution theory can be applied at a later stage.

However, section 3 of this paper shows that there are unit root alternatives whose autocorrelation function and likelihood function are arbitrarily close to those of stationary series. These local alternatives are again formed by adding random walk components, with sample size x innovation variance small compared to the variance of the stationary component. For these cases, the asymptotic distribution theory derived under the (false) assumption of trend stationarity may be a better guide than the (correct) asymptotic distribution theory derived under the assumption of a unit root. Similarly, the distribution of test statistics involving a stationary process whose largest autoregressive root is close to 1 might be better approximated by the unit root asymptotic theory than by the (correct) stationary theory.

Thus, even if a test for unit roots *could* successfully distinguish between a stationary process and a process with a unit root but small induced random walk component, or between a unit root process and a process with a large but less than unit root, such a test would not necessarily answer the question: which model provides the best approximation to the small sample distribution of estimates and test statistics? This observation calls into question the common methodology of pretesting for a unit root and then imposing the results of that test in subsequent analysis.

One example of the economic importance of unit roots is that with stationary dividend growth and discount rates, the price dividend ratio has a unit root if and only if there is a bubble. [See Hamilton and Whiteman (1985) and Cochrane (1989).] No matter how 'close' to a unit root, there is no bubble if the price dividend ratio is stationary and vice versa. Here the finding that unit roots cannot be distinguished from stationary series in finite samples tells us that the conceptually distinct economic models – bubble or no bubble – also cannot be distinguished in finite samples. Other examples of economic importance are hard to find. Christiano and Eichenbaum (1990) give several counterexamples.

The analysis in this paper is limited to univariate time series. However, the points carry over directly to the application of multivariate tests for the number of random walk components or, equivalently, tests for cointegration. [Stock and Watson (1988) and Phillips and Ouliaris (1987) are some examples of such tests. Cochrane and Sbordone (1988) present a multivariate extension of the Beveridge and Nelson Decomposition that shows how one can form cointegrated series by adding random walk components with variance increasing from zero.]

2. Tests for unit roots

Throughout, I assume that the time series under examination has either a *difference-stationary* Wold representation,

$$(1-L)y_t = \mu + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} = \mu + a(L)\varepsilon_t, \tag{1}$$

or a *trend-stationary* Wold representation,

$$y_t = \mu t + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} = \mu t + a(L)\varepsilon_t, \tag{2}$$

where $\mathrm{E}(\varepsilon_t | y_{t-1}, y_{t-2}, \dots) = 0$, $\mathrm{var}(\varepsilon_t) = \sigma_\varepsilon^2$, $a_0 = a(0) = 1$, $a(L)$ is invertible (its roots are all outside the unit circle), and $\sum_{j=0}^{\infty} a_j^2 < \infty$. In addition, I will assume that $\sum_{j=0}^{\infty} |a_j| < \infty$. Eq. (2) is a limiting case of (1), in which the polynomial $a(L)$ can be expressed as $(1-L)b(L)$, and $b(L)$ satisfies the above conditions. The point of tests for unit roots is to distinguish between series of the class (1) and those of the class (2).

Beveridge and Nelson (1981) give a constructive decomposition of series with a unit root (1) into a pure random walk with drift and a stationary component:

$$y_t = z_t + c_t,$$

$$z_t = \mu + z_{t-1} + a(1)\varepsilon_t, \qquad a(1) = \sum_{j=0}^{\infty} a_j, \tag{3}$$

$$c_t = a^*(L)\varepsilon_t, \qquad a_k^* = -\sum_{j=k+1}^{\infty} a_j.$$

An alternate statement [used by Granger and Engle (1987)] of this decomposition is that we can always rewrite the lag polynomial $a(L)$ in (1) as

$$(1-L)y_t = \mu + a(L)\varepsilon_t = \mu + (a(1) + (1-L)a^*(L))\varepsilon_t. \tag{4}$$

If $a(L)$ satisfies the conditions given under (1), $a^*(L)$ also satisfies those conditions, so the constructed $c_t = a^*(L)\varepsilon_t$ is a genuine 'stationary component'.

The representations (1) and (3)–(4) are equivalent, so 'difference-stationary', 'contains a unit root', and 'contains a random walk component' are equivalent. Since the decomposition (3)–(4) is constructed from the representation (1), any series with a representation (1) has such a decomposition. Trend-stationary series (2) also trivially have such a decomposition with the variance of changes to the random walk component equal to 0. Conversely, any combination of stationary and random walk components (with arbitrary correlation between their innovations) is stationary in first differences, and so has a representation (1) by the Wold decomposition theorem. Also, we can construct series with a representation (1) by independently picking any real number for $a(1)$ and any stationary stochastic process [satisfying the conditions under (2)] for $c_t = a^*(L)\varepsilon_t$.[1]

Various measures based on $a(1)$ can capture the importance of the unit root or random walk component. First, $z_t = \lim_{k\to\infty}(\mathrm{E}_t y_{t+k} - k\mu)$, i.e., z_t is the long-term forecast of y_t and $a(1)$ is the response of that long-term forecast to a unit innovation at time t. Second, $a(1)^2\sigma_\varepsilon^2$ is the innovation variance of the random walk component z_t and the spectral density of $(1-L)y_t$ at frequency 0. $a(1)\sigma_\varepsilon^2$ is also the innovation variance of the random walk in any decomposition of the series y_t into stationary and random walk components. Third, $(a(1)^2)\sigma_\varepsilon^2 = (1 + 2\sum_{j=1}^{\infty}\rho_j)\sigma_{\Delta y}^2$, where ρ_j is the jth autocorrelation coefficient of $(1-L)y_t$.

Furthermore, measures based on the quantity $a(1)$ are the *only* measures of the presence of a unit root in a finite sample, in that we can construct a trend-stationary series x_t that is 'just like' a given difference-stationary series y_t in every respect save $a(1)$. There are three ways to see this point. First, start with any difference-stationary process (1). Under normality or quadratic loss (or weak stationarity), we can completely characterize a sample of length T of the process by its $T-1$ autocovariances or $T-1$ periodogram ordinates. Then, we can always construct a trend-stationary process x_t such that its first difference $(1-L)x_t$ matches the periodogram ordinates of $(1-L)y_t$ at the ordinates $\{1\cdot 2\pi/T, 2\cdot 2\pi/T, \ldots, (T-1)\cdot 2\pi/T\}$. However, to make x_t trend-stationary, choose the periodogram ordinate of $(1-L)x_t$ at frequency 0 to be 0, while that of $(1-L)y_t$ is $a(1)^2\sigma_\varepsilon^2$. Hence this first ordinate *alone* distinguishes the trend-stationary series x_t from the difference stationary series y_t.[2]

Second, we can construct the trend-stationary series x_t from the difference-stationary y_t by choosing $x_t = \mu t + a^*(L)\varepsilon_t$. This is a trend-stationary series with the same stationary component as the difference stationary y_t. Since we can recover all of $a(L)$ in the representation of y_t except $a(1)$ from the a_j^* (see footnote 1), only $a(1)$ distinguishes this trend-stationary x_t from the corresponding difference-stationary y_t.

[1]It might seem that the conditions $a_0 = 1$ and $a_k^* = -\sum_{j=k+1}^{\infty} a_j$ uniquely determine $a(L)$ and hence $a(1)$, so the stationary and random walk components could not be picked independently: we know $a_0 = 1$ and the second condition implies $a_j = -(a_{j-1}^* - a_j^*)$ for $j \geq 1$. The reason this argument fails is that $a^*(L)$ is not necessarily normalized to $a_0^* = 1$. To construct y_t from a choice of $a^*(L)$ and $a(1)$, you can either pick $a_0^* = 1$ which is the conventional normalization, add an arbitrary $a(1)$, and then rescale the variance of ε_t so that $a_0 = 1$; or you can pick an arbitrary a_0^*. Either technique amounts to an arbitrary choice of the size of the random walk component added to the given stationary process.

[2]These propositions hold only for finite samples. Since the slope of the spectral density at 0 is also 0, the spectral density in an ε neighborhood of 0 does provide information about the spectral density at 0 asymptotically. See Phillips and Ouliaris (1987).

Third, recall that $(a(1)^2)\sigma_\varepsilon^2 = (1 + 2\sum_{j=1}^{\infty}\rho_j)\sigma_{\Delta y}^2$. For any finite set of N autocorrelations $\{\rho_1, \ldots, \rho_N\}$ of $(1-L)y_t$, we can construct a trend-stationary series x_t such that the first N autocorrelations of $(1-L)x_t$ match those of $(1-L)y_t$, but the infinite sum of the autocorrelations of $(1-L)x_t$ is 0.

In summary, periodogram ordinates other than that at frequency 0, aspects of the moving average $a(L)$ other than $a(1)$, or aspects of the autocorrelation of $(1-L)y_t$ other than their infinite sum carry *no* information about whether the series is trend- or difference-stationary.[3]

Thus any test for trend stationarity is a test of the hypothesis $a(1) = 0$ against the alternative $|a(1)| > 0$, and any test for unit roots is a test of the hypothesis $|a(1)| > 0$ against the alternative $a(1) = 0$. The likelihood function is continuous in the parameters of $a(L)$, so in any finite sample a test of the null hypothesis of trend stationarity has arbitrarily low power against the alternative of a unit root process with a small enough $a(1)$ or random walk component. Similarly, a test of the null hypothesis that the series has a unit root has no power against the adjacent alternative that the series is trend-stationary.

The only way out of this dilemma is add some restriction to the class of processes $a(L)$ that will be allowed. For example, we can test the hypothesis $a(1)^2 > \nu$ against $a(1) = 0$ or we can test the hypothesis $a(1) = 0$ against the alternative $a(1)^2 > \nu$, where ν is bounded away from 0. In particular, we can test the null hypothesis of a pure random walk $[a(1) = 1]$ against the alternative of trend stationarity $[a(1) = 0]$, and vice versa, by assuming away all the possibilities in between.

Alternately, we can assume that the slope of the spectral density is small in a region near 0, so that evidence from ordinates other than 0 can provide evidence about its value at the point $\omega = 0$. This is essentially the assumption behind the estimates Fama and French (1988), Huizinga (1987), Lo and MacKinlay (1988), and I myself [Cochrane (1988), Cochrane and Sbordone (1988)] used.

Every test in the literature includes some restriction of this sort. The restrictions are usually made to vanish as sample size increases, to derive the asymptotic distribution of the test under seemingly quite general assumptions. The test of Dickey and Fuller (1979) requires that $a(L)$ have an autoregressive representation of known finite order. Said and Dickey (1984) allow $a(L)$ to have an ARMA(p,q) representation, and allow the lag length used in autoregression to increase as $T^{1/3}$. Phillips and Perron (1986) and Stock and Watson (1988) require a maximum lag length for computed

[3]Note also that in all the interpretations given above, $a(1)$ is a property of the extremely long-run behavior of a series alone. The unit root question started as the question of the *first* autoregression coefficient: in $y_t = \mu + bt + \rho y_{t-1} + \varepsilon_t$, is $\rho = 1$ or is $\rho < 1$? What makes it a purely long-run question is the possibility of an arbitrary stationary time series process for ε_t.

autocorrelations, which increases as $T \to \infty$. The point of this paper is not to criticize those extra restrictions, but to emphasize that they must be included in a finite sample.

3. Other tests

Tests for unit roots are popular in part because the asymptotic distribution theory of many time series estimators or test statistics are sensitive to the difference between trend-stationary and difference-stationary series. Thus, it has seemed important to pretest a series to see in which class it falls, even if the tests have low power.

However, the point of the last section generalizes: there is always a difference-stationary series for which the results of any estimate or test (not just tests for unit roots) is arbitrarily close to the results of that estimate or test applied to a given stationary series. This point is demonstrated by showing that there are difference-stationary series whose likelihood functions and autocorrelation functions are arbitrarily close to those of any trend-stationary series.

Start from the representation (4). Then, we can write

$$y_k = y_0 + k\mu + a(1)[\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_k] + a^*(L)\varepsilon_t - a^*(L)\varepsilon_0. \quad (5)$$

We can construct a trend-stationary series by using the stationary component of (5). It has a corresponding representation:

$$x_k = x_0 + k\mu + a^*(L)\varepsilon_t - a^*(L)\varepsilon_0. \quad (6)$$

The conditional likelihood function (conditional on $\{y_0, \varepsilon_0, \varepsilon_{-1}, \ldots\}$ or $\{x_0, \varepsilon_0, \varepsilon_{-1}, \ldots\}$) of either process follows by inverting (5) or (6) for ε_t and knowledge of the distribution of ε_t.

Recall from section 2 that we can independently pick $a^*(L)$ as the lag polynomial of the moving average of any stationary process and $a(1)$ as any real number. From (5) and the fact that $\lim_{j\to\infty} a_j^* = 0$, the likelihood function for a set of difference-stationary observations $\{y_1, \ldots, y_T\}$ conditional on y_0 or on $\{y_0, \varepsilon_0, \varepsilon_{-1}, \ldots\}$ will be asymptotically dominated by the term following $a(1)$ no matter how small $a(1)$, so long as $a(1) > 0$. It is also clear that for any *finite* T, we may choose $a(1)$ small enough that the likelihood function for the difference-stationary process (5) and the trend-stationary process (6) are arbitrarily close.[4]

[4]Precisely, for any T, any set of parameters $a^*(L)$, and any $\delta > 0$, there is an $a(1) > 0$ such that the likelihood function generated from (6) and that generated from (5) differ by less than δ.

Alternately, consider the conditional autocorrelation function of y_k and y_{k+s}:[5]

$$\frac{\operatorname{cov}(y_k, y_{k+s} | y_0, \varepsilon_0, \varepsilon_{-1}, \dots)}{\operatorname{var}(y_k | y_0, \varepsilon_0, \varepsilon_{-1}, \dots)} = \frac{ka(1)^2 + a(1) \sum_{j=0}^{k-1} (a_j^* + a_{j+s}^*) + \sum_{j=0}^{k-1} a_j^* a_{j+s}^*}{ka(1)^2 + 2a(1) \sum_{j=0}^{k-1} a_j^* + \sum_{j=0}^{k-1} a_j^{*2}}. \tag{7}$$

Since $\lim_{j \to \infty} a_j^* = 0$, the second and third terms in both the numerator and denominator of (7) eventually must grow slower than linearly. Hence, so long as $a(1) > 0$, the first term in both the numerator and denominator eventually dominates the other terms as k grows, so that the conditional autocorrelation of y_t approaches 1 at all lags s for high enough k. On the other hand, if $a(1) = 0$, y_t is stationary, so its autocorrelations die out at large s for any k. This seems like a sharp difference in behavior between the cases $a(1) = 0$ and $a(1) > 0$.

However, in any finite sample we can only compute $T - 1$ autocorrelations. Hence, for any trend-stationary process (6) and for any desired degree of

[5]To derive eq. (7), start with the representation (5),

$$y_k = y_0 + k\mu + a(1)[\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_k] + a^*(L)\varepsilon_t - a^*(L)\varepsilon_0,$$

then

$$\mathrm{E}(y_k | y_0, \varepsilon_0, \varepsilon_{-1}, \dots) = y_0 + k\mu + \sum_{j=k}^{\infty} a_j^* \varepsilon_{k-j} - \sum_{j=0}^{\infty} a_j^* \varepsilon_{-j}$$

and

$$y_k - \mathrm{E}(y_k | y_0, \varepsilon_0, \varepsilon_{-1}, \dots) = \sum_{j=0}^{k-1} (a_j^* + a(1)) \varepsilon_{k-j},$$

and similarly for y_{k+s}. Then, the conditional covariance is

$$\operatorname{cov}(y_k y_{k+s} | y_0, \varepsilon_0, \varepsilon_{-1}, \dots) = \mathrm{E}\left(\left(\sum_{j=0}^{k-1} (a_j^* + a(1)) \varepsilon_{k-j}\right)\left(\sum_{j=0}^{k+s-1} (a_j^* + a(1)) \varepsilon_{k+s-j}\right)\right).$$

Expanding the product, keeping terms in ε_j^2,

$$\operatorname{cov}(y_k y_{k+s} | y_0, \varepsilon_0, \varepsilon_{-1}, \dots) = \sigma_\varepsilon^2 \sum_{j=0}^{k-1} (a_j^* + a(1))(a_{j+s}^* + a(1))$$

$$= \left(ka(1)^2 + a(1) \sum_{j=0}^{k-1} (a_j^* + a_{j+s}^*) + \sum_{j=0}^{k-1} a_j^* a_{j+s}^*\right) \sigma_\varepsilon^2.$$

The variance is the same expression evaluated at $s = 0$.

accuracy, we can pick a small enough $a(1)$ to construct a corresponding difference-stationary process (5) whose first $T-1$ autocorrelations are within the desired accuracy of the first $T-1$ autocorrelations of the trend-stationary process.

4. Concluding remarks

These results of this paper are not a criticism of the unit root tests per se. Rather, they are a warning that *application* of unit root tests without consideration for their low power and for the restrictions that they inevitably impose in a finite sample can be misleading. In particular, the results of unit root tests do not necessarily answer one important question, namely: which distribution theory provides a better small sample approximation?

The borderline cases discussed here are not just a technical curiosity. Fama and French (1988), Huizinga (1987), Campbell and Mankiw (1987), and I [Cochrane (1988)] found that stock prices, exchange rates, and GNP, respectively, are all potentially of the 'borderline' type. First differences of these series have large positive autocorrelations for the first few lags, and then a series of small negative autocorrelations at very high lags which may add up to a small $a(1)$. In each case the time scale required before $a(1)$ can be plausibly estimated is at least five or ten years.

References

Beveridge, Stephen and Charles R. Nelson, 1981, A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle', Journal of Monetary Economics 7, 151–174.

Blough, Stephen R., 1988, On the impossibility of testing for unit roots and cointegration in finite samples, Johns Hopkins University working paper in economics no. 211.

Campbell, John and N. Gregory Mankiw, 1987, Are output fluctuations transitory?, Quarterly Journal of Economics 102, 857–880.

Christiano, Lawrence J. and Martin Eichenbaum, 1990, Unit roots: Do we know and do we care?, Carnegie–Rochester Conference Series on Public Policy 32, forthcoming.

Cochrane, John H., 1988, How big is the random walk in GNP?, Journal of Political Economy 96, 893–920.

Cochrane, John H., 1989, Explaining the variance of price–dividend ratios, NBER working paper no. 3157.

Cochrane, John H. and Argia M. Sbordone, 1988, Multivariate estimates of the permanent components of GNP and stock prices, Journal of Economic Dynamics and Control 12, 255–296.

Dickey, D.A., W.R. Bell, and R.B. Miller, 1986, Unit roots series models: Tests and implications, American Statistician 40, 12–26.

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, Robert F. and C.W.J. Granger, 1987, Co-integration and error correction: Representation, estimation and testing, Econometrica 55, 251–276.

Fama, Eugene F. and Kenneth French, 1988, Permanent and temporary components of stock prices, Journal of Political Economy 96, 246–273.

Hamilton, James D. and Charles H. Whiteman, 1985, The observable implications of self-fulfilling expectations, Journal of Monetary Economics 16, 353–373.
Huizinga, John, 1987, An empirical investigation of the long run behavior of real exchange rates, Carnegie–Rochester Conference Series on Public Policy 27, 149–214.
Lo, Andrew and Craig MacKinlay, 1988, Stock market prices do not follow random walks: Evidence from a simple specification test, Review of Financial Studies 1, 41–66.
Lo, Andrew and Craig MacKinlay, 1989, The size and power of the variance ratio test in finite samples: A Monte Carlo investigation, Journal of Econometrics 40, 203–238.
Nelson, Charles R. and Heejoon Kang, 1981, Spurious periodicity in inappropriately detrended time series, Econometrica 49, 741–752.
Nelson, Charles R. and Charles I. Plosser, 1981, Trends and random walks in macroeconomic time series: Some evidence and implications, Journal of Monetary Economics 10, 139–162.
Phillips, P.C.B., 1987, Time series regression with a unit root, Econometrica 55, 277–301.
Phillips, P.C.B. and Ouliaris, 1987, Testing for cointegration, Cowles Foundation discussion paper no. 809.
Phillips, P.C.B. and Pierre Perron, 1986, Testing for a unit root in a time series regression, Manuscript.
Sargent, Thomas J., 1987, Macroeconomic theory (Academic Press, Orlando, FL).
Said, S.E. and D.A. Dickey, 1984, Testing for unit roots in autoregressive-moving average of unknown order, Biometrika 71, 599–607.
Schwert, G. William, 1987, Effects of model specification on tests for unit roots in macroeconomic data, Journal of Monetary Economics 20, 73.
Stock, James and Mark Watson, 1988, Testing for common trends, Journal of the American Statistical Association 83, 1097–1107.