Sims, C. A., J. H. Stock, and M. W. Watson. (1990). Inference in linear time series models with some unit roots. *Econometrica* 58:113–144.
Stock, J. H. (1987). Asymptotic properties of least-squares estimators of cointegrating vectors. *Econometrica* 55:1035–1056.
———. (1990). A class of tests for integration and cointegration. Mimeo, Harvard University.
Stock, J. H., and M. W. Watson. (1988a). Testing for common trends. *Journal of the American Statistical Association* 83:1097–1107.
———. (1988b). Variable trends in economic time series. *Journal of Economic Perspectives* 2(3):147–174.
———. (1989a). Interpreting the evidence on money-income causality. *Journal of Econometrics* 40:161–182.
———. (1989b). A simple MLE of cointegrating vectors in higher order integrated systems. National Bureau of Economic Research Technical Working Paper No. 83, Cambridge.
Stock, J. H., and K. D. West. (1988). Integrated regressors and tests of the permanent income hypothesis. *Journal of Monetary Economics* 21:85–96.
Watson, M. W. (1986). Univariate detrending methods with stochastic trends. *Journal of Monetary Economics* 18:1–27.
West, K. D. (1988). Asymptotic normality when regressors have a unit root. *Econometrica* 56:1397–1418.
Zivot, E., and D. W. K. Andrews. (1990). Further evidence on the great crash, the oil price shock and the unit root hypothesis. Cowles Foundation Discussion Paper No. 944, Yale University.

Comment

JOHN H. COCHRANE[1]
University of Chicago and NBER

This paper is an outstanding survey of unit root econometrics. It is an enormous and confusing literature, and Campbell and Perron's 24 rules are a tremendous and very practical condensation. If you decide to run unit root tests, this is a good place to start.

Rather than pick on rule 22, or survey some fields that the authors left out of this already massive paper (such as the Bayesian view or fractional unit roots), I will devote my comments to some reservations on practical usefulness. The bottom line is that, as much as I admire this paper as a survey of what *econometricians* know about unit roots, I am not yet convinced that this is what *macroeconomists* should know about unit roots.

For the moment, there are two broad uses of unit root econometrics, and I think it is best to organize my thoughts about what macroeconomists need to know about unit root econometrics by how they use it.

1. University of Chicago and NBER. I thank Jim Stock and Mark Watson for helpful discussions in preparing these comments.

1. Pretests for Unit Roots and Cointegration

Many macroeconomists now start papers whose substantive interest is elsewhere with tables of unit root and cointegration tests. These tests are used to determine the specification (order of differencing, which ratios are stationary, nature of deterministic trends, etc.) and relevant asymptotic distribution theory for subsequent estimates and tests.

The problem with this procedure is that, in finite samples, unit roots and stationary processes cannot be distinguished. For any unit root process, there are "arbitrarily close" stationary processes, and vice versa.[2] Therefore, the search for tests will sharply distinguish the two classes in finite samples is hopeless.

Campbell and Perron discuss this point under the title "near-observational equivalence," and I will respond in a second. However, their paper implies a much more severe version of the same problem, namely the possibility of deterministic trends.

Here's the problem. Low-frequency movement can be generated by unit roots (random walk components) or it can be generated by deterministic trends, including linear trends, "breaking trends," shifts in means, sine waves, polynomials, etc. Unit root *tests* are based on measurements of low-frequency movement in a time series, so they are easily fooled by nonlinear trends.[3] Therefore, Campbell and Perron's repeated theme that "the proper handling of deterministic trends is a vital prerequisite for dealing with unit roots" is correct and sensible advice.

But, of course, one never knows the deterministic trends with great precision before analysis begins. Economic theory does not give any guidance. And there is no hope that we can use purely statistical techniques to isolate arbitrarily specified deterministic trends.[4]

Thus the *theme* of the paper strikes me as the stake through the vampire's heart. The proper handling of deterministic trends *is* a vital prerequisite for dealing with unit roots. But "proper handling" of deterministic trends is an impossible task. To a humble macro*economist* it would seem that an edifice of asymptotic distribution theory that depends crucially on *unknowable* quantities must be pretty useless in practice.

However, there is an argument that the "observational equivalence"

2. Take a unit root process and change the root to 0.999. That's a "close" stationary process. Conversely, take a stationary process and add to it a random walk with tiny innovation variance. That's a "close" unit root process.
3. For example, in an earlier paper, Perron (1989) showed that U.S. GNP seems to have a unit root when compared to a stationary process around a linear trend. But if one allows for a break in the trend during the great depression, then U.S. GNP seems to be stationary around this "breaking trend." Therefore, to determine if there is a unit root in U.S. GNP, it is vital to know whether or not there is a "breaking trend."
4. This observation is due to Sims (1989) and Christiano (1988).

problem may not matter that much. The finite sample statistical properties of "borderline" time series lie between the polar extremes predicted by the unit root and stationary asymptotics. Therefore, unit root tests may provide a guide to which asymptotic distribution gives a better approximation to the true finite sample distribution, even if it is "wrong." The unit root distribution may better describe a stationary AR(1) with a coefficient of 0.9999 in a finite sample than the "true" stationary distribution. Similarly, maybe a "breaking trend" model is a useful metaphor for a series with moderately persistent and transitory "business cycle" shocks, as well as rare and extremely persistent (but, obviously, not literally deterministic) "world war" shocks.

This is a dangerous argument, since it implicitly acknowledges that unit root tests cannot accomplish the mission for which they were designed, and that mission is not interesting. But it is useful to think about anyway.

The approximation argument has been made informally,[5] but the paper includes a neat Monte Carlo that starts to address it quantitatively. Campbell and Perron simulate data from an ARMA(1,1), and apply unit root tests. Then, they compare the out of sample forecasting performance of AR models in levels and AR models in differences. Here is the interesting finding: in cases in which the unit root test was fooled, it nonetheless correctly indicated which estimated AR model would provide better out of sample one-step ahead forecasts.

But this Monte Carlo is an example, and not a theorem. Whether unit root tests are a good guide depends on *for what purpose,* and it is likely that one can easily think up purposes for which they are not a good guide. In particular, one lesson I learned from the unit root wars is that model selection criteria designed to produce good one-step-ahead forecasts can be very misleading for inferring long-run properties of a time series.

That lesson suggests a counterexample, which I evaluated with a small Monte Carlo. The results are presented in Table 1. The most interesting row of the table is the ARMA(1,1)s (θ=0.5). Here, the AR(1) in differences provides the better *one*-step-ahead forecasts but the AR(1) in levels provides the better *20*- and *50*-step-ahead forecasts.[6] The reason is obvi-

5. Cochrane (1991a), and others, I am sure.
6. Campbell and Perron use longer ARs in forecasting. In my example, unit root tests may not pick the correct AR(1), where in Campbell and Perron's example they pick the correct AR(p), with p selected by a specific lag length selection procedure. The point of my example is that there *are* purposes for which unit root tests can be misleading. As explained below, I wanted to separate the lag length selection question from the unit root question. There is no reversal when the data generating process is an AR(1) (θ=0). Since the AR(1) in levels is the true model, the most one can hope for is that the AR(1) in

Table 1 AVERAGE MEAN SQUARED ERROR OF FORECASTS

		1-step-ahead			*20-steps-ahead*			*50-steps-ahead*		
ϕ	θ	True	Level	Diff.	True	Level	Diff.	True	Level	Diff.
0.95	0.0	1.00	1.03	1.04	8.9	11.1	16.3	10.2	14.5	31.4
0.95	0.5	1.00	1.27	1.10	19.5	27.5	36.5	22.3	47.1	71.5
0.98	0.0	1.00	1.03	1.03	14.0	17.9	21.1	21.9	36.9	55.3
0.98	0.5	1.00	1.28	1.08	30.7	43.9	46.8	48.7	144	125

Note: The Monte Carlo follows Campbell and Perron's procedure. (1) A 100 period sample is drawn from the process
$X_t = \phi X_{t-1} + u_t + \theta u_{t-1}$, u_t iid $N(0,1)$.
(2) An AR(1) in levels and an AR(1) in differences are fit to the 100 period sample by OLS. These are used to forecast X_{101}, X_{120}, and X_{150}. (Note: Campbell and Perron may use longer order ARs.) Also, a forecast is computed using the true ARMA(1,1) model. (3) A sample of $\{X_{101} \ldots X_{150}\}$ is drawn 25 times, and the mean squared error of each forecast is evaluated. (4) The whole procedure is repeated 5000 times to produce the average mean squared error.

ous: 20 and 50 steps ahead, the series has pretty much reverted to its mean. The levels model may completely miss the short-term dynamics, but it recognizes this crucial fact.

Thus, unit root tests do not *necessarily* provide a good guide to the right approximate model. This point is obvious, and not a criticism of anyone: in no field of statistics has anyone ever claimed that there is an estimator that is optimal for *every* loss function, and so here.

A second lesson I learned from the unit root wars is that the pure *unit root* question is much less important than other aspects of the modeling process. I think that lesson describes the Monte Carlo as well.

The Monte Carlo has five ingredients: (1) The choice of data-generating mechanism, an ARMA(1,1), (2) the choice of levels vs. differences, (3) the choice of family of approximate models, ARs, (4) the estimation procedure, OLS, and (5) lag length selection procedure, here driven by *t*-statistics on extra lags.

Of the five ingredients, it seems to me that the Monte Carlos say the *least* about the choice of levels vs. differences. (1) Estimates of the *true* model [ARMA(1,1)] ought to form better forecasts than *any* AR approximation.[7] Thus the choice of families of approximate models is important. (2) AR models in differences and AR models in levels can arbitrarily well approximate each other as well as the true ARMA(1,1), if one allows

differences will perform equally for one-step-ahead forecasts. That is why I had to construct stationary ARMA(1,1)s, not considered by Campbell and Perron, to make this example work.

7. Most estimation techniques amount to minimizing the one-step-ahead forecast error variance, so this statement is almost a theorem.

arbitrary lag lengths. Thus the lag length selection procedure is crucially important. (3) Similarly, if the true data-generating mechanism features unstructured mean reversion rather than a tight ARMA(1,1), it is likely that more loosely parameterized models will do better for long-run forecasts. (4) OLS selects parameters by matching the spectral density of the true model and the data over the whole frequency range; for long-run purposes, it may be better to use an estimation technique that emphasizes low-frequency aspects.[8]

As an example of all these points, one can estimate an AR(p) in levels by starting with an OLS estimate of an AR($p-1$) in differences and calculating the implied model in levels.[9] Thus, absent lag length restrictions and a stand on estimation, *nothing* is determined by the choice of levels and differences.

These points are meant as praise rather than criticism. Given the death blow Campbell and Perron dealt to unit roots tests by noticing that we must prespecify deterministic trends, the tests will be interesting *only* if we learn something about approximation issues. I just want to point out how subtle the issues can be, and to argue that Campbell and Perron's Monte Carlo is the *beginning* of a literature, rather than an epitaph.

And there is a long way to go. The forecasting question Campbell and Perron address is the *least* frequent use of unit root tests. Suppose one tests for cointegration, and then imposes the results of the test in subsequent analysis, such as VAR estimation or Granger causality tests. Do times when the unit root test indicate the wrong model also correspond to times when the asymptotic distribution theory based on the wrong model is a better approximation? Maybe yes, but maybe no. And how sensitive is this guide to the other aspects of the modeling process, especially hidden deterministic trends and lag length selection procedures? Nobody knows.

2. *Direct Estimates of Unit Roots*

The second use of unit root tests has been simply testing for unit roots and cointegration for its own sake. It is natural that each new time series technique gets tried out on every series in CITIBASE, and one has to write an introduction about the economic relevance of the test to get it

8. See the appendix to Cochrane (1988).

9. $(X_t - X_{t-1}) = a_0 + a_1(X_{t-1} - X_{t-2}) + \ldots a_{p-1}(X_{t-p+1} - X_{t-p}) + \epsilon_t$

implies

$$X_t = a_0 + (1+a_1)X_{t-1} + (a_2-a_1)\, X_{t-2} + \ldots - a_{p-1}X_{t-p} + \epsilon_t.$$

past the referees. This happened with Box–Jenkins techniques, Granger causality, and VARs, and is now going on with nonlinear time series models (both chaotic and ARCH variants) and fractional integration, as well as unit roots and cointegration. The question is whether such unit root and cointegration tests are worth pursuing much further.

My two reservations about this kind of work are (1) that not much of economic importance hinges on unit root or cointegration structure *per se,* and (2) that the unit root methodology described by Campbell and Perron can be quite misleading.

2.1. ECONOMIC INTERPRETATION OF UNIT ROOT TESTS

The impossibility of distinguishing unit roots and deterministic trends argues that, in Christiano and Eichenbaum's (1989) terminology, "we don't know," or, better, "we *can't* know," so it *must* be the case that "we don't care." Nothing of economic significance can hinge on an *unknowable* quantity. Though this is clear in the abstract, I think it is worth making the point directly.

Consider the still-studied question whether GNP contains a unit root or not. Why do we care? Initially, Nelson and Plosser (1982) argued that the presence of unit roots meant that shocks were persistent, and hence that most shocks to GNP were technology shocks. But with the advantage of hindsight, I think this interpretation has evaporated.

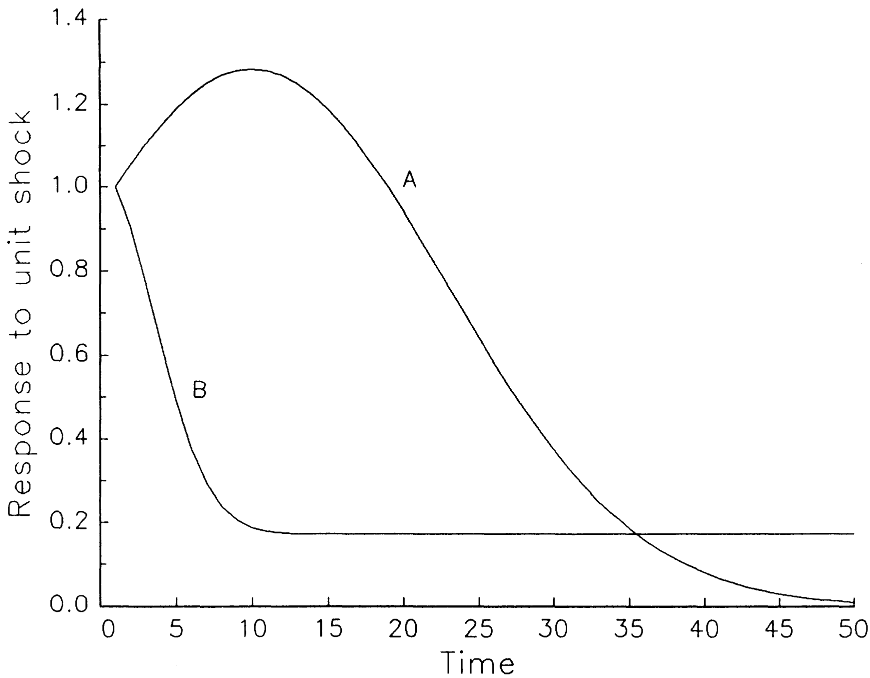
First, it is now clear that unit roots need have nothing to do with persistence. Consider the impulse–response functions plotted in Figure 1.

A series has a unit root if and only if the limit of its impulse–response function is nonzero. Thus, series B has a unit root, and series A is stationary. But the shock to series A is obviously much more "persistent," by any interesting measure. (I think the common confusion that unit root means persistence comes from thinking of unit roots as generalizations of random walks, rather than just difference stationary and arbitrarily autocorrelated time series.)

Again, one might argue for approximation. Series that are more likely to reject unit root tests may also be those with "less persistent" shocks. But, again, we do not know anything about the accuracy of such an approximation.

Second, it is also now clear that the persistence of univariate shocks tells us nothing about the source and nature of true shocks to the economy. At best, unit root tests and persistence measures uncover aspects of the univariate Wold representation, in which the shocks are errors from forecasts of GNP based on past GNP. These shocks are different objects from multivariate prediction shocks recovered from VARs, and

Figure 1 IMPULSE–RESPONSE FUNCTIONS



different objects again from the "true" shocks that impinge in the economy.[10] The persistence of univariate prediction error shocks can be a very misleading guide to the persistence of multivariate prediction error shocks, and both can be very misleading guides to the persistence of the true or underlying shocks.

2.2 UNIT ROOT TEST METHODOLOGY CAN BE MISLEADING

Even if one is just interested in examining the univariate time series properties of a given variable, unit root test methodology can be misleading. The unit roots question amounts to the specification of units: should we use levels or first differences (etc.). For most series we *know* the answer. GNP, consumption, investment, etc. belong in growth rates. Variables that are already rates, such as interest rates, inflation, and unemployment belong in levels. Ratios such as the dividend/price ratio, the consumption/GNP ratio, etc. belong in levels.

Unit root tests often suggest the opposite. I think the fact that they

10. See Hansen and Sargent (1991) for this point, and Cochrane (1991b) and Lippi and Richelin (1990) for examples and discussion in the unit root in GNP context.

suggest the opposite, and that they are wrong, is one of the most interesting things to come out of this literature.

For example, so long as you do not get too creative with breaking trends and structural shifts, any test tells you that interest rates have unit roots, and lag selection procedures indicate a near random walk structure. That model does quite well for one-step-ahead forecasting. Yet, interest rates are almost certainly stationary in levels. Interest rates were about 6% in ancient Babylon; they are about 6% now. The chances of a process with a random walk component displaying this behavior are infinitesimal.[11] Furthermore, the mean reversion of interest rates is economically important: it explains expected return premia in the term structure.[12]

Most unit root tests (again without overly creative deterministic trends) point to a unit root in postwar GNP and most lag selection procedures deliver a near random walk structure. But a short order ARMA in GNP growth misses its substantial and economically important transitory movement over business cycles.[13]

The dividend/price ratio fails most unit root tests, yet theory and common sense suggest that it must be stationary. It too, features very long swings. A researcher who blindly follows the advice of unit root tests and lag length selection procedures would miss the long-run mean reversion in returns forecast by dividend price ratios, and the useful fact that prices and dividends are cointegrated.

3. *Summary*

The central problem driving all the doubts I have expressed is that the pure statement that a series has a unit root (or that two series are cointegrated) is vacuous in a finite sample. Campbell and Perron (implicitly) and Sims (1989) emphasize the fact that unit roots are indistinguishable from nonlinear trends. Here and elsewhere I have emphasized the

11. Stan Fischer pointed out that I may have gotten the story wrong, and cited a figure near 25% for interest rates in ancient Babylon. Interest rates *were* around 6% in the middle ages, and the substance of the story goes through even starting at 25%. One way to make the argument a little more formal is to calculate

$$\Pr(|r_{1991}| < 100\% \mid r_{4000\text{B.C.}} = 6\%).$$

This probability is infitesimal if interest rates are or contain a random walk; it is near one if interest rates are an AR(1) with a coefficient of 0.99.

12. Fama and Bliss (1987).

13. Transitory movement is hard to document with any univariate method, but is clear in multivariate estimates. See Blanchard and Quah (1989), Cochrane and Sbordone (1988), Cochrane (1991b) among others.

fact that all unit root tests and estimated models come with lag length selection procedures, and the action is in the lags, not in the roots.

There is still some hope that unit root tests will provide useful approximations for some purposes. That hope must rest on implicit assumptions that one can, in fact, prespecify a lot about deterministic trends, and that modeling the low-frequency behavior of time series does not, in fact, require richer specifications than typical lag length selection procedures allow. Furthermore, whether unit root tests provide a useful approximation guide has to depend on for what purpose. *This* is what macroeconomists need to know about unit roots, and I hope Campbell and Perron's paper inspires them and others to find out.

I do not want to seem negative. I think we have learned a lot from the unit root journey. Among other positive results, (1) our handling of trends is much improved. Ten years ago, the Y variable in most models was stationary about a mean, and data were blithely detrended or Hodrick–Prescott filtered to match them with the model. Now most theoretical models are constructed to predict the appropriate stationarity inducing transformation. (For example, see the Rotemberg and Woodford paper in this volume.) (2) We are much more sensitive to the information in levels, and relations between levels. For example, Lucas (1988) and Stock and Watson (1991) use relations between levels to measure the income elasticity of money demand, and Ogaki and Park (1989) use relations between levels to measure preference parameters. (3) Cointegrated *representations* and error correction models are proving very useful. (4) As I mentioned before, we are aware of long-horizon mean reversion, and interesting long-horizon behavior of time series that is missed by the old AR(2) around a deterministic trend [or the new AR(1) in first differences].

However, in all these cases it is the *representation* machinery that is paying off. In most cases, one *knows* the unit root/cointegration structure. Relations between levels (#2 above) are equally informative whether they are relations between stochastic or deterministic trends. Thus, the *testing* machinery is not very useful and often misleading.

It is very hard to argue against the proposition "macroeconomists should know x" since more knowledge is never bad. But the statement that "macroeconomists should know" Campbell and Perron's 24 rules imply that empirical papers should start with a battery of tests for unit roots and cointegration with a variety of nonlinear and breaking trends, that empirical researchers should change the specification of their subsequent work in response to that battery of tests (otherwise, why bother?), and that editors and referees should complain loudly when such tables are not included or do not contain up-to-the minute methodology. The message of my comments is that one can appreciate the paper and the literature it summarizes and still disagree with that conclusion.

REFERENCES

Blanchard, Olivier Jean, and Danny Quah. (1989). The dynamic effects of aggregate supply and demand disturbances. *American Economic Review* 79(September):655–673.

Christiano, Lawrence J. (1988). Searching for a break in GNP. Manuscript, Federal Reserve Bank of Minneapolis.

Christiano, Lawrence J., and Martin Eichenbaum. (1989). Unit roots in GNP: Do we know and do we care? *Carnegie-Rochester Conference Series on Public Policy.*

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

———. (1991a). A critique of the application of unit roots tests. *Journal of Economic Dynamics and Control* 15(April):275–284.

———. (1991b). Univariate vs. multivariate forecasts of GNP growth, and stock returns: Evidence and implications for the persistence of shocks, detrending methods, and tests of the permanent income hypothesis. (February). (Revision of NBER working paper #3427, September 1990.)

Cochrane, John H., and Argia M. Sbordone (1988). Multivariate estimates of the permanent components in GNP and stock prices. *Journal of Economic Dynamics and Control* 12(June/July):255–296.

Fama, Eugene F., and Robert R. Bliss. (1987). The information in long-maturity forward rates. *American Economic Review* 77:680–683.

Hansen, Lars Peter, and Thomas J. Sargent. (1991). Two difficulties in interpreting vector autoregressions. In *Rational Expectations Econometrics.* Boulder and London: Westview Press.

Lippi, Marco, and Lucrezia Richelin. (1990). A note on measuring the dynamic effects of aggregate demand and supply distrubances. Working paper No. 90-06, Observatoire Français des Conjonctures Economiques, Paris.

Lucas, Robert E. (1988). Money demand in the United States: A quantitative review. *Carnegie-Rochester Conference Series on Public Policy* 29:137–168.

Nelson, Charles, and Charles Plosser (1982). Trends and random walks in macroeconomic time series. *Journal of Monetary Economics* 10: 139–162.

Ogaki, Masao, and Park, Joon Y. (1989). A cointegration approach to estimating preference parameters. Rochester Center for Economic Research Working paper 209, University of Rochester (December).

Perron, Pierre. (1989). The great crash, the oil price shock and the unit root hypothesis. *Econometrica* 57:1361–1401.

Sims, Christopher A. (1989). Modeling trends. Discussion paper 22, Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis (December).

Sims, Christopher A., and Harold Uhlig. (1991). Understanding unit rooters: A helicopter tour. *Econometrica* (in press).

Stock, James H., and Mark W. Watson. (1991). A simple estimator of cointegrating vectors in higher order integrated systems. Manuscript, Northwestern University.