Econometric Theory, **25**, 2009, 637–657. Printed in the United States of America.

COMMENTARIES ON "Unit Root Testing in Practice: Dealing with Uncertainty over the Trend and Initial Condition," by David I. Harvey, Stephen J. Leybourne, and A.M. Robert Taylor

PATRICK MARSH*
University of York
doi:10.1017/S0266466608090191

The paper devises operational procedures for testing for a unit root in a time series. It does so under the conditions of uncertainty about whether a deterministic linear trend is present and also whether the initial condition for the observed series is asymptotically negligible. The objective is to provide practitioners with clear and simple procedures that are also well justified in terms of their statistical properties. That is, their size should be, approximately, known under very general assumptions and their power at least as large as any competitor. Judged according to the twin criteria of statistical optimality and operational simplicity the authors provide procedures that are, in my view, unmatched.

Any commentary on this paper should pay serious attention to those twin requirements of the practitioner, optimality and simplicity. This is particularly so because it is precisely the issues of whether or not to include a linear trend or how to incorporate unknown initial conditions that make satisfying them both simultaneously especially difficult. The recommended procedures involve either a pre-test determining which of two unit root tests to use or, even simpler, a union of their rejection regions. Given that the underlying statistics are well-known standard variants of established tests, there can be no doubt that the paper has provided the literature with tests that are extremely simple to implement and interpret. With luck they will soon be incorporated into our standard econometric packages.

This commentary will focus upon the optimality of the procedures. Uncertainty over the inclusion of a linear trend is seen to be related to a classic information loss problem, and the suggested procedures an optimal response to it. Similarly, although the issue of the initial condition is not explicitly addressed via a formal invariance argument, they nonetheless obey one. These concepts of optimality

*Address correspondence to Patrick Marsh, Department of Economics, University of York, Heslington, York YO10 5DD, United Kingdom; e-mail: pwnm1@york.ac.uk.

© 2009 Cambridge University Press 0266-4666/09 $15.00

https://doi.org/10.1017/S0266466608090191 Published online by Cambridge University Press

matter greatly for the practitioner; procedures are unlikely to be adopted if there is an expectation that a "better" one might soon be developed.

Optimal Unit Root Tests

All of the proposed procedures are based on established procedures, with comparisons among them appearing in Section 2. There are no uniformly most powerful unit root tests. However, numerical evidence (here and elsewhere) suggests that Dickey–Fuller tests applied via quasi-differencing (*DF-QD* tests), developed in Elliott, Rothenberg, and Stock (1996), have a slender but consistent advantage over, for example, ordinary least squares (OLS) residual based tests or those based on the von Neumann ratio, as in Schmidt and Phillips (1992). It should be noted, however, that no formal optimality theory justifies the *DF-QD* tests. Such theory applies instead to the likelihood ratio, L_T^*, tests.

It should equally be noted that no test obeying a formal optimality criterion, whether point or locally optimal, is ever used in practice, though such tests provide necessary benchmarks for our numerical experiments. There is a seemingly bewildering array of criteria by which we may declare a test "optimal." What this paper brings into extremely sharp focus is that the issue of whether or not a linear trend should be included has a potentially far greater impact upon power than the actual choice of test.

The recent literature does contain examples where the chosen underlying procedure may be beaten. However, either the conditions required are very restrictive, such as in Francke and de Vos (2007), and/or they apply only to the case where the hypothesized unit root is assumed to be in the levels of y, rather than in the error process, u, as in Marsh (2007a). An open question pertaining to an optimality/simplicity trade-off remains. The quasi-differencing parameter $\bar{c}$ is chosen so that at $\bar{c}$ the power envelope is 50%. More general frameworks will allow for additional regressors, such as additional polynomial trends, dummy variables, breaks in the trends, and so on. Does each such model require the calculation of a unique $\bar{c}$, as demanded by optimality, or is the simple solution of using $\bar{c} = 13$ in all cases sufficient? Although this is, of course, beyond the scope of this paper, it would become a serious issue for the obvious extensions to uncertainty over polynomial trends or trend breaks.

Tests and (Uncertainty over) Linear Trends

The formulation of the model in the first two equations of the paper, along with Assumption 1, has become standard in the literature. To recapitulate, the data y_t are determined by the three equations

$$y_t = \mu + \beta t + u_t, \tag{1}$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \tag{2}$$

https://doi.org/10.1017/S0266466608090191 Published online by Cambridge University Press

$$\varepsilon_t = C(L)e_t. \tag{3}$$

Crucially in this literature, only ρ in (2) is an interest parameter, whereas μ and β in (1) and the conditional variance of e_t, σ^2, and the coefficients of $C(L)$ in (3) are nuisance parameters. Of immediate concern to the practitioner, therefore, is that the size of any procedure should be either known or predictably accurate over all values of these unknowns. As highlighted in the paper the MAIC (modified Akaike information criterion) procedure of Ng and Perron (2001) has been demonstrated to deliver such for the parameters in (3). Traditionally, the approach to dealing with the parameters in (1) has been to consider tests that are invariant with respect to a trend and/or a constant.

The impact on power of requiring invariance with respect to the trend compared to just the constant is clear both in the paper (see Figure 1) and in the literature before it. This loss of power is borne regardless of the value of β, including if $\beta = 0$, but we do not impose that restriction. The crucial element in the paper, that is, in Section 3, is that there is uncertainty as to whether (possibly asymptotically at an appropriate rate) $\beta = 0$ or not.

As the discussion in Section 3.3.3 concludes, the choice really comes down to one between the pre-test (PT) and the union of rejections (UR) based procedures. The latter is favored in the paper, in part because the conservative strategy for it slightly outperforms the former in terms of its size. However, it is worth highlighting here an underlying statistical justification for them, particularly the former. The issue of whether or not to include a linear trend is essentially one of information loss. Specifically, if $\beta = 0$ is imposed in (1), then at the unit root Fisher information in any test invariant only to a constant will generally be of order $O(T^2)$, as given by the asymptotic Hessian, for example, derived via a Gaussian likelihood. On the other hand, when we do not impose $\beta = 0$ information falls to zero; see Marsh (2007b, Thm. 1, part (i)). Consequently, if $\beta = 0$ but this is not imposed then a practitioner is unnecessarily losing all of the information in the sample about ρ at the very point that the practitioner is testing.

The PT procedures, and also those based upon the weighted average, are directly underpinned by statistical optimality theory, not explored in the paper. Although uncertainty over the trend in the paper is characterized by considering local alternatives of the form $\beta = \kappa w_\varepsilon T^{-1/2}$, consider instead letting $\beta = \gamma\beta_0$, where γ is a Bernoulli random variable, say, $\gamma \sim \mathcal{B}(\pi)$, with

$$\gamma = \begin{cases} 0 & \text{with probability } 1-\pi \\ 1 & \text{with probability } \pi \end{cases}. \tag{4}$$

That is, any data set may have been taken from a data generating process (DGP) with a trend with probability π and from a DGP without one with probability $(1-\pi)$. Thus here the uncertainty is modeled as a mixture of the no trend and fixed trend cases.

https://doi.org/10.1017/S0266466608090191 Published online by Cambridge University Press

The pre-test based tests can all be written

$$PT\left(|t_\beta|\right) = \left(DF\text{-}QD^\mu\right)\left(1 - \mathbb{I}\left(|t_\beta| > cv_\alpha\right)\right) + \left(DF\text{-}QD^\tau\right)\mathbb{I}\left(|t_\beta| > cv_\alpha\right),$$

where t_β is a consistent test of size $\alpha \times 100\%$, of $H_0^\beta : \beta = 0$. Asymptotically,

$$\lim_{T\to\infty} \Pr\left[|t_\beta| > cv_\alpha \,|\, \gamma = 1\right] = 1,$$

$$\lim_{T\to\infty} \Pr\left[|t_\beta| > cv_\alpha \,|\, \gamma = 0\right] = \alpha,$$

so that each pre-test can be written

$$PT\left(|t_\beta|\right) = \left(DF\text{-}QD^\mu\right)\left(1 - \hat{\gamma}\right) + \left(DF\text{-}QD^\tau\right)\hat{\gamma}, \tag{5}$$

where $\hat{\gamma}$ is also a Bernoulli random variable with

$$\hat{\gamma} \sim \mathcal{B}(\pi + \alpha(1-\pi)).$$

Because $\hat{\gamma}$ has a distribution not depending on the interest parameter, ρ, it is *ancillary*, and the decision rule given for the pre-test procedures is in fact one based upon the *conditional* asymptotic distribution of $PT(|t_\beta|)$, *given* $\hat{\gamma}$. In fact all pre-test procedures may be cast in a similar light. What is critical about the problem at hand is the extreme asymptotic difference in information, conditional on the outcome of $\hat{\gamma}$, being zero if $\hat{\gamma} = 1$ and $O(T^2)$ otherwise.

By analyzing the uncertainty on the inclusion of a linear trend in the context of the mixing implied by (4), the problem is, asymptotically, analogous to Cox's original justification for conditional inference, as detailed in Fraser (2004). There, samples may be drawn randomly from one of two sources having the same mean but one with high precision and the other low. For arguments explored at length in Fraser's review, conditioning upon the ancillary statistic is more informative about the precision with which the mean may be identified from the sample at hand. This is almost exactly the situation described earlier—almost in that $\hat{\gamma}$ does not exactly identify whether or not the model implying high precision was chosen.

Statistical examination of the operationally simpler UR procedure is less straightforward. By definition, the UR test statistic is

$$\begin{aligned} UR &= DF\text{-}QD^\mu\left(\mathbb{I}(DF\text{-}QD^\mu < -1.94)\right) + DF\text{-}QD^\tau\left(\mathbb{I}(DF\text{-}QD^\mu \geq -1.94)\right) \\ &= DF\text{-}QD^\mu\left(1 - \gamma^*\right) + DF\text{-}QD^\tau\left(\gamma^*\right), \end{aligned} \tag{6}$$

with

$$\gamma^* = \begin{cases} 1 & \text{if } \mathbb{I}(DF\text{-}QD^\mu \geq -1.94) \\ 0 & \text{otherwise} \end{cases},$$

which is of the same form as the $PT(t_1)$ test, except that γ^* is not ancillary. However, as is explained in Remark 6 in the paper, if t_1 is the t-statistic for testing

https://doi.org/10.1017/S0266466608090191 Published online by Cambridge University Press

$\beta = 0$, upon first-differencing, then the events, $|t_1| > 1.96$ and $DF\text{-}QD^{\mu} < -1.94$ are asymptotically mutually exclusive. With $\kappa = 1$, we have

$$DF\text{-}QD^{\mu} \geq -1.94 \Longleftrightarrow t_1^2 \geq 1 - 1.94\sqrt{\int_0^1 (r + W_c(r))^2\, dr}, \tag{7}$$

and because $\int_0^1 (r + W_c(r))^2\, dr$ is $O_p(1)$ and t_1 is a consistent test, when $\gamma = 1$, we find

$$\lim_{T\to\infty} \Pr\left[DF\text{-}QD^{\mu} \geq -1.94\right] = 1;$$

however, when $\gamma = 0$ this probability depends, because this is a unit root test, upon whether or not $c = 0$. As a consequence, the mixing over the two statistics implied by (6) will differ from that in (5), with

$$\gamma^{*} \sim \mathcal{B}(\pi + \bar{\alpha}_c(1 - \pi)),$$

where $\bar{\alpha}_c$ is the probability that $DF\text{-}QD^{\mu} \geq -1.94$, given that $\gamma = 0$.

The UR procedure differs from the PT procedures in terms of the statistical properties of the conditioning variable. However, they are both very simple combinations of the two tests having proven power properties for their given model. Combining a binary view of the uncertainty with the results of Marsh (2007b) it is clear that, asymptotically, inference based upon these tests can be viewed as being conditional upon an information recovering statistic, which is strictly ancillary for the PT statistics. Given that any future competitor must share this basic structure, or else not recover information, and that it is difficult to think of simpler combinations of the underlying tests, the prospect of these tests being bettered seems very remote. That, ultimately, the UR procedure is preferred on the grounds of simplicity and slightly better finite–sample size properties should not provoke controversy.

The Initial Condition

The second objective of the paper is to suggest procedures that are successful in the light of uncertainty over the asymptotic negligibility of the initial condition. Again there are two issues at play, the practical and the theoretical. The practitioner needs to consider whether the series begins with the initial observation or whether be or she is merely observing some fraction of a process that may have been going on, unobserved, for an unspecified, possibly infinite, length of time. Given the success of the UR procedure for dealing with uncertainty over the trend, the same is also applied in this case.

The possibility of a nonnegligible initial condition is captured in Assumption 3, with two possibilities, either

https://doi.org/10.1017/S0266466608090191 Published online by Cambridge University Press

$$\text{(a)} \quad u_1 = \mu_\alpha \quad \text{or}$$

$$\text{(b)} \quad u_1 \sim N\left(0, \frac{\sigma_\alpha^2 \omega_\varepsilon^2}{1-\rho_T^2}\right), \qquad \rho_T = 1 - c/T. \tag{8}$$

Because the platform upon which the procedures are built is that of the construction of invariant tests it is important to notice that the unit root hypothesis testing problem remains invariant (with respect to the constant and trend) for (b) but not for (a). Invariance follows from the unconditional mean of y_t not depending upon the autoregressive parameter. If $u_1 = \mu_\alpha$, then we easily find

$$\mathrm{E}[y_t] = \mu + t\beta + \rho_T^{t-1}\mu_\alpha.$$

Because the problem is no longer invariant it is hardly surprising that tests built on the methods developed in Dufour and King (1991), such as the Elliott, Rothenberg, and Stock (1996) tests, do not necessarily work. Numerical evidence of such is readily found in Figures 6 and 7 of the paper.

It is perfectly possible to incorporate a nonzero initial condition and retain the desirable properties implied by invariance. All that is necessary is that $\mathrm{E}[y_t]$ is not a function of ρ_T for all t. Rewriting equations (1)–(3) as

$$(y_t - \mu - \beta t) - \rho_T\,(y_{t-1} - \mu - \beta(t-1)) = C(L)e_t,$$

then for $t = 1$, we require that

$$\mathrm{E}[y_1] = \mathrm{E}[\rho_T\,(y_0 - \mu)] + \beta + \mu = \tilde{\mu}_1, \tag{9}$$

where $\tilde{\mu}_1$ does not depend on ρ_T. Because $(y_0 - \mu) = u_1$, then assumption (b) in (8) is seen to satisfy this condition.

More generally we need not assume $\mathrm{E}[u_1] = 0$, provided (9) holds. By treating $\tilde{\mu}_1$ as a nuisance parameter entering only the equation for y_1, tests invariant with respect to this follow by simply including a dummy variable on the initial observation. It is, however, important to stress that such tests are invariant only in the sense that they do not depend on the nuisance parameter $\tilde{\mu}_1$. They are not invariant with respect to the initial condition itself. Moreover, as follows immediately from the Frisch–Waugh theorem, regressing out the dummy variable implies nothing more than starting the regression at time $t = 2$. That is, with an initial condition generated under (b) in (8), optimal tests then follow by detrending via an OLS regression, beginning at $t = 2$. The procedures described throughout the paper do precisely that. The *DF-QD* procedure fails even in the latter case, because quasi-differencing (or any kind of differencing) implies test statistics explicitly depending on y_1.

There are technical reasons for preferring an assumption on the initial condition that preserves invariance of the testing problem. However, economic justifications may be more appropriate; for instance, many recorded series do begin following some structural break. Whether assumption (a) in (8) appropriately captures this

https://doi.org/10.1017/S0266466608090191 Published online by Cambridge University Press

is a mater for debate, however. For example, would a break in the middle of the observed series be similarly modeled? Notice also that the condition required for invariance to be maintained does not constrain the variance of u_1 in any way. For example, u_1 could be generated by a (local-to) unit root process begun in the infinite past; its mean will remain zero.

Despite this, ample numerical evidence is given in the paper to give sufficient confidence in the suggested procedure, regardless of the assumption on the initial condition. Ultimately, to the practitioner, that is probably all that matters.

REFERENCES

Dufour, J.-M. & M.L. King (1991) Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or nonstationary AR(1) errors. *Journal of Econometrics* 47, 115–143.

Elliott, C., T.J. Rothenberg, & J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.

Francke, M.K. & A.F. de Vos (2007) Marginal likelihood and unit roots. *Journal of Econometrics* 137, 708–728.

Fraser, D.A.S. (2004) Ancillaries and conditional inference (with commentaries). *Statistical Science* 19, 333–369.

Marsh, P. (2007a) Constructing optimal tests on a lagged dependent variable. *Journal of Time Series Analysis* 28, 723–743.

Marsh, P. (2007b) The available information for invariant tests of a unit root. *Econometric Theory* 23, 686–710.

Ng, S. & P. Perron (2001) Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519–1554.

Schmidt, P. & P.C.B. Phillips (1992) LM tests for a unit root in the presence of deterministic trends. Oxford Bulletin of Economics and Statistics 54, 257–287.

ULRICH K. MÜLLER*
Princeton University
doi:10.1017/S0266466608090208

When testing for an autoregressive unit root in practice, uncertainty over the initial condition and the potential presence of a linear trend are important issues that need to be addressed. Also, it is well known that for standard data generating processes, t-statistics of augmented Dickey–Fuller-type regressions yield unit root tests with good small-sample properties relative to other approaches. David Harvey, Stephen Leybourne, and Robert Taylor (abbreviated HLT in what follows) ingeniously combine the information of different Dickey–Fuller-type t-statistics to construct unit root tests for both forms of uncertainty. The two resulting union of rejections procedures are computationally straightforward, and, as the authors demonstrate, they inherit the attractive small-sample properties of the

*The author thanks Graham Elliott and Mark Watson for helpful discussions and the NSF for financial support via grant SES-0518036. Address correspondence to Ulrich Müller, Department of Economics, Princeton University, Princeton, NJ 08544, USA; e-mail: umueller@princeton.edu.

https://doi.org/10.1017/S0266466608090191 Published online by Cambridge University Press