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.

underlying $t$-statistics, at least for independent and identically distributed (i.i.d.) Gaussian disturbances.

What is more, under uncertainty over the magnitude of the initial condition, the suggested union of rejections test is "close to admissible" asymptotically for Gaussian driving disturbances: Müller and Elliott (2003) and Elliott and Müller (2006) derive unit root tests that maximize weighted average power over different values of the initial condition under the alternative, and HLT demonstrate that the local asymptotic power of the combination test is never much below (and sometimes above) the power of such a weighted average power maximizing test for a particular weighting function. Although not obviously optimal by construction, the union of rejections test by HLT is therefore "close to admissible" asymptotically in the sense that no test can exist with substantially higher asymptotic power for all values of the initial condition.

The issue of uncertainty over the presence and magnitude of a time trend has received much less attention in recent years (with the exception of Ayat and Burridge, 2000), and no corresponding weighted average power maximizing test over values of the (local) trend coefficient under the alternative has been derived. The remainder of this commentary provides such a derivation and identifies a particular weighting function so that HLT's union of rejections test under uncertainty over the trend has power that is never much below the power of the weighted average power maximizing test. In analogy to the case of uncertainty over the magnitude of the initial condition, this shows that HLT's union of rejections test is "close to admissible" asymptotically.

Both union of rejections tests thus come close to exploiting the available information efficiently, at least in large samples. To the extent that the implicit weighting functions correspond to the priorities of the applied researcher, these results strengthen the appeal of HLT's union of rejections tests for empirical work.

## Admissible Tests under Uncertainty about a Local Trend

From a statistical point of view, the value of the initial condition and the value of the local trend parameter are nuisance parameters for testing the unit root null hypothesis. There is an important asymmetry in their role, however: For translation invariant unit root tests, the initial condition is only present under the alternative, as changes of the initial condition under the null hypothesis are identical to translations. The value of the local trend, in contrast, potentially affects the distribution of translation invariant tests under both the null and alternative hypotheses.

To fix ideas, consider the same model as HLT with a local trend and standard normal errors, that is,

$$y_t = \mu + \beta t + u_t, \qquad t = 1, \cdots, T,$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \qquad t = 1, \cdots, T,$$

where $u_0 = 0$, $\rho = \rho_T = 1 - c/T$, $\beta = \beta_T = \kappa T^{-1/2}$, and $\varepsilon_t \sim$ i.i.d.$\mathcal{N}(0, 1)$. As Elliott, Rothenberg, and Stock (1996) (abbreviated ERS in what follows), suppose we are interested in deriving point optimal tests; that is, we seek to efficiently discriminate $H_0 : c = 0$ against $H_1 : c = c_1$ for some known positive value of $c_1$, which, following ERS, we choose to equal $c_1 = 13.5$. These hypotheses are composite, with nuisance parameters $\mu$ and $\kappa$ that appear under both the null and alternative hypotheses. ERS eliminate these nuisance parameters by invoking invariance to the group of transformations $\{y_t\}_{t=1}^T \mapsto \{y_t + m + bt\}_{t=1}^T$ for all $m, b \in \mathbb{R}$: Theorem 1 of Lehmann (1986, p. 285) shows that all invariant tests can be written as a function of a maximal invariant, and one such maximal invariant is the ordinary least squares (OLS) residuals of a regression of $y_t$ on $(1, t)$. It is quite apparent that the distribution of these residuals does not depend on the value of $\mu$ and $\kappa$, so that the nuisance parameters effectively drop out of the problem when imposing invariance. Furthermore, translation invariance $\{y_t\} \mapsto \{y_t + m\}$ for all $m \in \mathbb{R}$ seems entirely natural for many applications, because the mean of a series measured in logs depends on the units of measurement, and invariance ensures that the outcome of the unit root test remains consistent across choices of, say, a measurement in millions or billions. At the same time, the case for invariance to adding a linear trend seems less obvious. One might thus be interested, as HLT are, in constructing powerful tests that are only translation invariant, so that the local trend parameter $\kappa$ remains as a nuisance parameter under both the null and alternative hypotheses.

If it was known that $\kappa = \kappa_1$ under the alternative and $\kappa = 0$ under the null hypothesis, then following the same derivations as in ERS, one obtains that rejecting for large values of

$$\mathrm{LR}_T(\kappa_1) = \exp\bigg[ -\tfrac{1}{2}c_1((J_T(1) - \kappa_1)^2 - 1) + \kappa_1 J_T(1) - \tfrac{1}{2}\kappa_1^2 - \tfrac{1}{2}c_1^2 \int_0^1 (J_T(s) - \kappa_1 s)^2 ds \bigg]$$

with $J_T(s) = T^{-1/2}(y_{[sT]} - y_1) \Rightarrow \int_0^s e^{-c(s-r)} dW(r) + \kappa s$, where $W$ is a standard Wiener process, is the asymptotically translation invariant point-optimal test. More generally, if one wants to maximize weighted average power over alternative values of $\kappa = \kappa_1$, with an integrable weighting function $w$, against the null hypothesis with $\kappa = 0$ known, then one needs to reject for large values of

$$\int \mathrm{LR}_T(\kappa_1)\, dw(\kappa_1).$$

Of course, we do not actually know that $\kappa = 0$ under the null hypothesis, and so these considerations seem of no obvious value. But consider first a special case where $w$ is degenerate and puts all mass on $\kappa_1 = 0$, so that $\mathrm{LR}_T(0)$ is the efficient test statistic. Note that a test based on $\mathrm{LR}_T(0)$ is asymptotically equivalent to ERS's point-optimal test in the mean case for $c_1 = 13.5$. Now HLT's numerical analysis shows that the 5% level *DF-QD*$^\mu$ test, with critical value chosen to ensure asymptotically correct nominal level for $\kappa = 0$, has asymptotically rejection probability of less than 5% for any value of $\kappa \neq 0$. This suggests that the 5% level test that rejects for large values of $\mathrm{LR}_T(0)$, with critical value chosen for $\kappa = 0$, also controls asymptotic size over $\kappa$, and a numerical computation confirms that conjecture. Thus, the 5% level $\mathrm{LR}_T(0)$ test maximizes asymptotic weighted average power over alternative values of $\kappa = \kappa_1$ (with a weighting function $w$ that concentrates all mass at zero), while assuming a particular value of the nuisance parameter $\kappa$ under the null hypothesis (which happens to be zero, too), and has a rejection probability below 5% for any other value of the nuisance parameter $\kappa$ under the null hypothesis. These observations imply via Theorem 7 of Lehmann (1986, p. 104) that in large samples, putting all mass at $\kappa = 0$ is the least favorable distribution for this particular weighting function $w$, and that the asymptotic 5% level $\mathrm{LR}_T(0)$ test maximizes asymptotic power at $c = c_1$ and $\kappa = \kappa_1 = 0$ among all tests that have asymptotic rejection probability of no more than 5% under the composite null hypothesis that leaves $\kappa$ unspecified.

Now consider weighting functions $w$ that correspond to the distribution of a mixture between a mean zero normal with variance $V > 0$ (whose probability density function we denote by $\phi_V$) and a point mass at zero, with mixing probabilities $p$ and $1 - p$, respectively. By completing the square, one obtains

$$
\begin{aligned}
Q_T(p, V) &= p \int \mathrm{LR}_T(\kappa_1)\phi_V(\kappa_1)\, d\kappa_1 + (1 - p)\mathrm{LR}_T(0) \\
&= \exp\left[-\tfrac{1}{2}c_1(J_T(1)^2 - 1) - \tfrac{1}{2}c_1^2 \int_0^1 J_T(s)^2\, ds\right] \\
&\quad \times \Bigg( \frac{p}{\sqrt{V(1 + c_1 + \frac{1}{3}c_1^2 + V^{-1})}} \\
&\qquad \exp\left[\frac{1}{2}\frac{((1 + c_1)J_T(1) + c_1^2 \int_0^1 s J_T(s) ds)^2}{1 + c_1 + \frac{1}{3}c_1^2 + V^{-1}}\right] + 1 - p \Bigg),
\end{aligned}
$$

and *if* it continues to be the case that choosing the asymptotic critical value of $Q_T(p, V)$ for $\kappa = 0$ yields an asymptotic 5% level test that has smaller null rejection probability for $\kappa \neq 0$, then this test is also the asymptotic weighted average power maximizing test for the composite null hypothesis with $\kappa$ unconstrained. Numerical computations suggest this to be the case as long as $V \leq 0.9$.[1] By construction, the family of asymptotic 5% level tests $Q_T(p, V)$ indexed by $p \in [0, 1]$ and $V \in (0, 0.9]$ thus maximize asymptotic weighted average power over different values of $\kappa = \kappa_1$ for the local alternative with $c = c_1 = 13.5$ when $\varepsilon_t \sim$ i.i.d.$\mathcal{N}(0, 1)$ among all translation invariant tests. To show that HLT's union of rejections test is "close to admissible," we must thus identify a test from this family such that the asymptotic power of HLT's test when $c = c_1 = 13.5$ does not fall much below this test for any value of $\kappa$. A reasonably good match is obtained for $p = 0.45$ and $V = 0.9$. Figure 1 depicts the asymptotic power of HLT's

5% level union of rejections test of their Section 3.3.2 (denoted UR) and the 5% level test based on $Q_T(0.45, 0.9)$ for $c = c_1 = 13.5$ as a function of $|\kappa| \in [0, 1.5]$. By construction, no test can exist with uniformly higher asymptotic power than $Q_T(0.45, 0.9)$ across all values of $\kappa$. As can be seen from Figure 1, HLT's UR test has asymptotic power that is never more than 6.5 percentage points below the power of $Q_T(0.45, 0.9)$ (with the largest shortfall of power for approximately $|\kappa| = 0.25$), and it is more powerful than $Q_T(0.45, 0.9)$ when $|\kappa| > 0.6$ (and unreported results show this to be the case also for $|\kappa| > 1.5$). Thus, no test can exist with substantially higher asymptotic power than HLT's union of rejections test uniformly over $\kappa$, and so HLT's test is "close to admissible" asymptotically for the alternative with $c = c_1 = 13.5$.

These asymptotic optimality properties were derived under the assumption that $\varepsilon_t \sim$ i.i.d.$\mathcal{N}(0, 1)$. As Jansson (2008) demonstrates, it is in general possible to derive more powerful unit root tests for non-Gaussian i.i.d. driving disturbances. At the same time, the optimality results derived here in the i.i.d. Gaussian setting extend to the non-Gaussian case in some sense: if the unit root null hypothesis with a local trend is defined in terms of the weak convergence

$$\hat{\omega}_T^{-1} J_T(s) = T^{-1/2} \hat{\omega}_T^{-1} (y_{[sT]} - y_1) \Rightarrow W(s) + \kappa s \qquad \textbf{(1)}$$

for some "reasonable" long-run variance estimator $\hat{\omega}_T^2$, then it makes sense to impose on 5% level unit root tests to have asymptotic rejection probability of at most 5% for *all* (double-array) processes $\{y_t\}$ that satisfy (1). But under this constraint, the results of Müller (2007) imply that an asymptotically 5% level test based on $Q_T(0.45, 0.9)$ with $J_T$ replaced by $\hat{\omega}_T^{-1} J_T$ maximizes asymptotic weighted average power for *all* alternatives that satisfy $\hat{\omega}_T^{-1} J_T(s) \Rightarrow \int_0^s e^{-c_1(s-r)} dW(r) + \kappa s$ (with Gaussian or non-Gaussian $\{y_t\}$). So if HLT's union of rejections test is implemented with *DF-QD*$^{\mu}$ and *DF-QD*$^{\tau}$ computed via replacement of

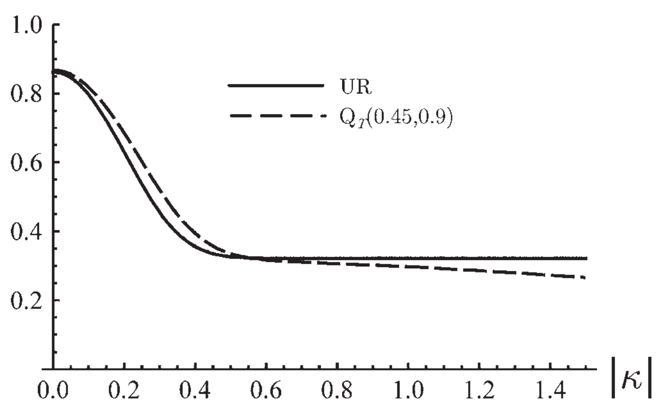


**FIGURE 1.** Asymptotic power for $c = 13.5$ as function of the local trend $\kappa$.

$W_c(s) + \kappa s$ by $\hat{\omega}_T^{-1} J_T(s)$ on the right-hand side of HLT's (7), then the test is "close to admissible" asymptotically in this more general sense.

## Conclusion

The union of rejection tests suggested by HLT are intuitive, are easy to implement, and, at least up to a good approximation, do not leave any information unexploited asymptotically. In addition, by virtue of the tests having been derived from Dickey–Fuller-type $t$-statistics, it is reasonable to expect them to have relatively good finite–sample properties also in the presence of autoregressive moving average–type short-run dynamics.

The exact mechanism for the relatively superior small-sample performance of Dickey–Fuller-type $t$-statistics is not well understood, which makes it difficult to judge what kind of regularity conditions the data need to possess to generate reliable results in practice. One alternative is to base inference exclusively on the difference in low-frequency variability implied by unit root and local-to-unity processes. Müller and Watson (2008) argue that in a macroeconomic context, a reasonable *definition* of an I(1) process might be that the below-business-cycle-frequency variability matches that of a Gaussian random walk. The low-frequency unit root test suggested in Müller and Watson (2008) tests this property directly, without taking a stand on higher frequency variability, facilitating the interpretation of a rejection in practice.

### *NOTE*

1. A calculation shows that a test based on $Q_T(p, V)$ has asymptotic null rejection probability converging to unity as $|\kappa| \to \infty$ when $p > 0$ and $V > (3c_1 + c_1^2)/(3 + 3c_1 + c_1^2) \approx 0.987$, for any finite critical value.

### *REFERENCES*


Ayat, L. & P. Burridge (2000) Unit root tests in the presence of uncertainty about the stochastic trend. *Journal of Econometrics* 95, 71–96.

Elliott, G. & U.K. Müller (2006) Minimizing the impact of the initial condition on testing for unit roots. *Journal of Econometrics* 135, 285–310.

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

Jansson, M. (2008) Semiparametric power envelopes for tests of the unit root hypothesis. *Econometrica* 76, 1103–1142.

Lehmann, E.L. (1986) *Testing Statistical Hypotheses*, 2nd ed. Wiley.

Müller, U.K. (2007) An Alternative Sense of Asymptotic Efficiency. Working paper, Princeton University.

Müller, U.K. & G. Elliott (2003) Tests for unit roots and the initial condition. *Econometrica* 71, 1269–1286.

Müller, U.K. & M.W. Watson (2008) Testing models of low-frequency variability. *Econometrica* 76, 979–1016.