

LM TESTS FOR A UNIT ROOT IN THE PRESENCE OF DETERMINISTIC TRENDS*

Peter Schmidt and Peter C. B. Phillips

I. INTRODUCTION

The most commonly used tests of the null hypothesis of a unit root in an observed time series are derivatives of the Dickey-Fuller tests (Dickey (1976), Fuller (1976), Dickey and Fuller (1979)). The Dickey-Fuller tests were developed for simple Gaussian random walks and the derivative procedures (notably Said and Dickey (1984), Phillips (1987) and Phillips and Perron (1988)) are intended to detect the presence of a unit root in a general integrated process of order one (I(1) process). The Dickey-Fuller tests are based on the regression of the observed variable (say, y) on its one-period lagged value, with the regression sometimes including an intercept and time trend; that is, they are based on regressions of the form:

$$y_t = \beta y_{t-1} + \varepsilon_t \quad (1)$$

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t \quad (2)$$

$$y_t = \alpha + \beta y_{t-1} + \delta t + \varepsilon_t \quad (3)$$

for $t = 1, 2, \dots, T$. The $\hat{\rho}$, $\hat{\rho}_\mu$, and $\hat{\rho}_\tau$ tests are based on the statistic $T(\hat{\beta} - 1)$, where $\hat{\beta}$ is the OLS estimator of β in (1), (2) and (3) respectively, while the $\hat{\tau}$, $\hat{\tau}_\mu$ and $\hat{\tau}_\tau$ tests are based on the t -statistics for the hypothesis $\beta = 1$ in the same three regressions. The former are coefficient tests, and the latter are t -ratio tests. Both types of test have time series extensions by the semiparametric correction method of Phillips (1987) and Phillips and Perron (1988). Only the t -ratio test is extended in the long autoregression method of Said and Dickey (1984).

Following the empirical work of Nelson and Plosser (1982), a common motivation for testing for a unit root is to test the hypothesis that a series is difference stationary against the alternative that it is trend stationary. That is, one wishes to test for a unit root in the presence of deterministic trend. Economists are especially interested in such tests because under the alterna-

*The authors thank the Editorial Board for helpful comments. Financial support of the National Science Foundation and wordprocessing by Glenda Ames are also gratefully acknowledged.

tive hypothesis of stationarity time series exhibit trend reversion characteristics, whereas under the null they do not. Unfortunately, the parameterizations in (1)–(3) above are not convenient for this purpose, because they handle level and trend in a clumsy and potentially confusing way. Equation (1) does not allow for non-zero level or trend either under the null hypothesis nor under the alternative, of course. Equations (2) and (3) do, but they suffer from the serious problem that the meanings of the parameters α and δ differ under the null and the alternative. For example, in equation (2) the parameter α represents trend when $\beta = 1$ (since the solution for y_t then includes the deterministic trend term αt), but it determines level when $\beta < 1$ (since y_t is then stationary around the level $\alpha/(1 - \beta)$). Similarly, in equation (3), when $\beta = 1$ the parameter α represents trend and δ represents quadratic trend, while under the alternative α determines level and δ determines trend. This confusion over the meanings of the parameters shows up in the properties of the Dickey-Fuller tests based on (2) and (3). The distributions of the Dickey-Fuller tests based on equation (2) depend on the nuisance parameter α , even under the null hypothesis (Evans and Savin (1984), Nankervis and Savin (1985), Schmidt (1990), Guilkey and Schmidt (1992)), and they are inconsistent against trend stationary alternatives (West (1987)) because equation (2) does not allow for trend under the alternative. The distributions of the Dickey-Fuller tests based on equation (3) are independent of α , but they do depend on δ under both the null and the alternative. The tests based on equation (3) therefore constitute similar tests with regard to the nuisance parameter α and they are consistent against trend stationary alternatives. But they allow for trend under the alternative by introducing a variable (t) that is irrelevant under the null and this is the price that is paid for the property of similarity with regard to α . The role of irrelevant regressors in achieving test similarity is considered more fully in the present context in Kiviet and Phillips (1992).

These difficulties can be avoided by regarding the equations (1)–(3) not as data generating processes (DGP's), but simply as regression equations used to generate test statistics. In this paper we consider the DGP to be as follows:

$$y_t = \psi + \xi t + X_t, X_t = \beta X_{t-1} + \varepsilon_t. \quad (4)$$

This DGP (or parameterization) has previously been considered by Dickey (1984), Bhargava (1986), and others in the context of unit root tests and it corresponds to the conventional components representation of a time series. Once again the unit root corresponds to $\beta = 1$. This paper will present tests that are extracted from the score or LM principle applied to (4) under the assumption that the ε_t are iid $N(0, \sigma_\varepsilon^2)$. However, it should be stressed that the advantages of the parameterization in (4) are deeper than just that it leads to our new tests. This parameterization allows for trend under both the null and the alternative, without introducing any parameters that are irrelevant under either. Indeed, the important attraction of this parameterization is that the meaning of the nuisance parameters ψ and ξ does not depend on whether the

unit root hypothesis is true: ψ represents level and ξ represents deterministic trend, whether $\beta = 1$ or not. As a result, the parameterization in (4) is useful in studying the properties of our new tests and also those of the Dickey-Fuller tests. In particular, the distributions of our tests and of the Dickey-Fuller tests under both the null and alternative hypothesis are independent of the nuisance parameters ψ , ξ and σ_ε .

The plan of the paper is as follows. Section II defines the new test statistics and compares them to the Dickey-Fuller $\hat{\rho}_\tau$ and $\hat{\tau}_\tau$ statistics. Section III gives results on the finite sample distributions of the statistics under the assumption of iid errors. Section IV provides the asymptotic distribution of the statistics under more general error assumptions, and gives extensions along the lines of Phillips (1987) and Phillips and Perron (1988) that are asymptotically robust to error autocorrelation and heteroskedasticity. Section V extends the tests to the case of deterministic trend that follows a higher order polynomial in time. Section VI provides some Monte Carlo evidence on the power of the tests. Finally, Section VII contains our conclusions.

II. UNIT ROOT TESTS BASED ON THE SCORE PRINCIPLE

We begin with the model as given in (4) above, where the errors ε_t are assumed to be iid $N(0, \sigma_\varepsilon^2)$ and where the initial condition X_0 is taken as fixed. We wish to derive the LM test of the hypothesis $\beta = 1$ in this model. The derivation is given in Appendix 1, and here we will give only a brief summary. The restricted MLE's (that is, the MLE's when we impose $\beta = 1$) of ξ and $\psi_X = \psi + X_0$ are as follows:

$$\bar{\xi} = \text{mean } \Delta y = (y_T - y_1)/(T-1) \quad (5)$$

$$\tilde{\psi}_X = y_1 - \bar{\xi} \quad (6)$$

(The parameters ψ and X_0 are identified separately under the alternative hypothesis, but not under the null hypothesis that β equals one.) Note that, as expected, the estimate of ξ comes from estimation of (4) in differences. Now define the 'residuals'

$$\hat{S}_t = y_t - \tilde{\psi}_X - \bar{\xi}t, \quad t = 1, \dots, T. \quad (7)$$

These are the residuals from the model (4) in levels, but where the parameters have been estimated from the model in differences. We note that $\hat{S}_1 = \hat{S}_T = 0$. A little algebra reveals the following alternative expressions for \hat{S}_t :

$$\hat{S}_t = \sum_{j=2}^t (\Delta y_j - \bar{\Delta y}) \quad (t \geq 2) \quad (8A)$$

$$= y_t - y_1 - (t-1)\bar{\Delta y} \quad (8B)$$

$$= [(T-1)y_t - (t-1)y_T - (T-t)y_1] \quad (8C)$$

$$\begin{aligned}
&= \sum_{j=2}^t (w_j - \bar{w}), \quad w_j = \beta^{j-1}(\beta - 1)X_0 + \varepsilon_j \\
&\quad + (\beta - 1) \sum_{i=0}^{j-1} \beta^i \varepsilon_{j-i} \quad (t \geq 2). \quad (8D)
\end{aligned}$$

In Appendix 1 we show that the score vector evaluated at the restricted MLE's is proportional to

$$\sum_{t=2}^T (\Delta y_t - \bar{\xi}) \hat{S}_{t-1}. \quad (9)$$

This is the numerator of the estimated regression coefficient of \hat{S}_{t-1} in the regression

$$\Delta y_t = \text{intercept} + \phi \hat{S}_{t-1} + \text{error} \quad (t = 2, \dots, T). \quad (10)$$

Denote the least squares estimate of ϕ by $\tilde{\phi}$. We then define the test statistics

$$\tilde{\rho} = T \tilde{\phi} \quad (11)$$

$$\tilde{\tau} = \text{usual } t\text{-statistic for } \phi = 0 \text{ in (10)}. \quad (12)$$

It is instructive to compare these statistics to the Dickey-Fuller statistics $\hat{\rho}_\tau$ and $\hat{\tau}_\tau$ based on regression (3). This is a regression of y on intercept, time trend and lagged y ; equivalently, it is a regression of Δy on the same variables. By the standard algebra of least-squares regression, it follows that this is in turn equivalent to the regression

$$\Delta y_t = \text{intercept} + \rho \hat{S}_{t-1} + \text{error} \quad (t = 2, \dots, T), \quad (13)$$

where \hat{S}_{t-1} is the residual from an *ordinary least squares* regression of y_{t-1} on an intercept and time trend. We then have $\hat{\rho}_\tau$ as the estimated coefficient of \hat{S}_{t-1} in (13), and $\hat{\tau}_\tau$ as the t -statistic for the hypothesis $\rho = 0$. Comparing (13) to (10), the only difference the new tests $\tilde{\rho}$ and $\tilde{\tau}$ and the Dickey-Fuller tests $\hat{\rho}_\tau$ and $\hat{\tau}_\tau$ is the nature of the residual upon which Δy_t is regressed. Both \hat{S}_{t-1} in (10) and \hat{S}_{t-1} in (13) are residuals in the levels equation for y_{t-1} , but the parameters used to calculate the residuals are estimated differently: the parameters used to calculate \hat{S}_{t-1} are estimated from the model in differences, while the parameters used to calculate \hat{S}_{t-1} are estimated from the model in levels. Given that y_t is $I(1)$ under the null hypothesis, the regression of y_{t-1} on intercept and time *in levels* is spurious in the sense of Granger and Newbold (1974) and Phillips (1986), so that the regression coefficients for the intercept and time trend do not converge to constants, but remain random even asymptotically. This 'extra' randomness makes \hat{S}_{t-1} complicated in a way that \hat{S}_{t-1} is not. Tests based on \hat{S}_{t-1} will have simpler properties than tests based on \hat{S}_{t-1} , and (depending on one's intuition) may also be expected to be more powerful.

Since $\Delta y_t - \bar{\xi} = \Delta \hat{S}_t$, the score vector given above in (9) can be rewritten as

$$\text{Score} = \sum_{t=2}^T (\Delta y_t - \bar{\xi}) \hat{S}_{t-1} = \sum_{t=2}^T \Delta \hat{S}_t \hat{S}_{t-1} = \sum_{t=2}^T (\Delta \hat{S}_t - \overline{\Delta \hat{S}}) \hat{S}_{t-1}, \quad (14)$$

where the last equality follows from mean $\overline{\Delta \hat{S}} = (\hat{S}_T - \hat{S}_1)/(T-1) = 0$. Thus the score vector is also the numerator of the estimated regression coefficient in the regressions

$$\Delta \hat{S}_t = \text{intercept} + \phi \hat{S}_{t-1} + \text{error} \quad (15A)$$

$$\Delta \hat{S}_t = \phi \hat{S}_{t-1} + \text{error}. \quad (15B)$$

Equation (15A) is identical to (10) above and yields the same test statistics, $\hat{\rho}$ and $\hat{\tau}$. However, if $\tilde{\phi}$ is the least squares estimate of ϕ in (15B), $\tilde{\phi} \neq \hat{\phi}$; the numerators are the same, but the denominators are different, because the sample mean of \hat{S}_{t-1} is non-zero. Let $\tilde{\rho}$ and $\tilde{\tau}$ be the test statistics from (15B), defined analogously to (11) and (12) above. These statistics are analysed by Schmidt and Lee (1991), who show that their power properties are generally inferior to those of $\hat{\rho}$ and $\hat{\tau}$, as will be explained more precisely below.

The tests $\tilde{\rho}$ and $\tilde{\tau}$ are closely related to earlier tests of Bhargava (1986). We show in Appendix 2 that the score vector in (9) or (14) can also be rewritten as follows:

$$\text{Score} = \sum_{t=2}^T \Delta \hat{S}_t \hat{S}_{t-1} = -\frac{1}{2} \sum_{t=2}^T \Delta \hat{S}_t^2 = -\frac{1}{2} \sum_{t=2}^T (\Delta y_t - \overline{\Delta y})^2. \quad (16)$$

Using this result, it is not hard to show that $\tilde{\rho}$ is almost equal to $-T/2$ times Bhargava's statistic R_2 defined in Bhargava (1986, p. 376, equation (40)), and that $\tilde{\rho}$ equals $-T/2$ times Bhargava's statistic N_2 (p. 377, equation (47)). See Appendix 2. Note also that since the model (4) has been estimated in differences \hat{S}_t is a partial sum process (see (18) below). Hence, the score (16) can be interpreted as a functional of this partial sum process \hat{S}_t and its component elements $\Delta \hat{S}_t$. As such, tests like $\tilde{\rho}$ and $\tilde{\tau}$ that are based on \hat{S}_t are related to the generic class of tests for a unit root developed recently by Stock (1990).

An interesting fact that also follows from (16) is that $\tilde{\tau}$ is a monotonic transformation of $\tilde{\rho}$. Specifically, we show in Appendix 2 that

$$\tilde{\tau} = \begin{cases} -\left[\frac{-2T}{(T-3)\tilde{\rho}} - \frac{1}{(T-3)} \right]^{-1/2} \tilde{\rho} < 0 \\ \left[\frac{2T}{(T-3)\tilde{\rho}} - \frac{1}{(T-3)} \right]^{-1/2} \tilde{\rho} > 0 \end{cases} \quad (17)$$

Thus the tests are equivalent. However, we will continue to discuss both tests, because the tests are not equivalent after various types of corrections for error autocorrelation (to be discussed later).

III FINITE SAMPLE DISTRIBUTION THEORY

The finite sample distributions of the test statistics $\tilde{\rho}$ and $\tilde{\tau}$ are complicated and will be tabulated by simulation. However, we first note the simple but important fact that the distributions of the test statistics under the null hypothesis are independent of the nuisance parameters ψ , ξ , X_0 and σ_ε . Thus the distributions of the test statistics under the null hypothesis depend only on the sample size (T). Bhargava derived his statistics R_2 and N_2 as best invariant tests, and this guarantees invariance with respect to ψ , ξ and σ_ε . However, a direct approach is simple and perhaps more revealing. Using equation (8D) above, we have, under H_0 ,

$$\tilde{S}_t = \sum_{j=2}^t (\varepsilon_j - \bar{\varepsilon}). \quad (18)$$

This does not depend on ξ , ψ or X_0 . Since $\tilde{\rho}$ and $\tilde{\tau}$ can be written as functions of \tilde{S}_t , $t=1, \dots, T$, their distributions are invariant to ξ , ψ and X_0 . Finally, the scale factor σ_ε also cancels out of all expressions for $\tilde{\rho}$ and $\tilde{\tau}$, so that their distributions are independent of σ_ε as well as the other nuisance parameters.

King (1981) and Dufour and King (1991) have similarly used the theory of invariance to yield test statistics independent of nuisance parameters. Their tests are designed to be point optimal and hence they are of rather different form than this paper's tests or Bhargava's.

Critical values for the test statistics $\tilde{\rho}$ and $\tilde{\tau}$ are given in Table 1A. These are calculated by a direct simulation using 50,000 replications. Random deviates were generated by the routines GASDEV and RAN3 of Press, Flannery, Teukolsky and Vetterling (1986); more detail on this random number generation scheme can be found in Guilkey and Schmidt (1992). We note in passing that the lower tail critical values are smaller in absolute value than the corresponding lower tail critical values for the Dickey-Fuller $\hat{\rho}_\tau$ and $\hat{\tau}_\tau$ tests.

Under the alternative hypothesis that β is not equal to one, the distributions of $\tilde{\rho}$ and $\tilde{\tau}$ are independent of ψ , ξ and σ_ε , but they depend on $X_0^* = X_0/\sigma_\varepsilon$. (They also depend on β and T , of course.) To show this, we make use of the representation of \tilde{S}_t in (8D) above. From (8D) it is clear that \tilde{S}_{t-1} is independent of ψ and ξ . It depends on X_0 unless $\beta=1$; and, given T , β and X_0^* , it has the same scale as σ_ε . Since both $\tilde{\rho}$ and its associated t -statistic depend only on the \tilde{S}_t , they are independent of ψ and ξ , but their distribution depends on X_0 when $\beta \neq 1$. Also, since the scale factor σ_ε enters the numerator and denominator of $\tilde{\rho}$ and the t -statistic in exactly the same way,

the distributions of $\tilde{\rho}$ and the t -statistic (for given T , β and X_0^*) are also independent of σ_ε . Furthermore, from (8D) it is also clear that, if the errors ε_t are symmetrically distributed, the distributions of $\tilde{\rho}$ and $\tilde{\tau}$ depend only on $|X_0^*|$, T and β .

De Jong *et al.* (1989) have proved the same invariance results for the Dickey-Fuller $\hat{\rho}_\tau$ and $\hat{\tau}_\tau$ tests. (Dickey (1984) had earlier noted the invariance of $\hat{\rho}_\tau$ and $\hat{\tau}_\tau$ to ψ and ξ , based on numerical results.) Under the null hypothesis, their distributions depend only on T , while under the alternative, they depend on T , β and X_0^* . If the ε_t are symmetrically distributed, dependence of the distributions on X_0^* is replaced by dependence on $|X_0^*|$.

Simulation evidence on the powers of the tests will be given in Section V.

IV ASYMPTOTICS

Following Phillips (1987) and Phillips and Perron (1988), we can relax the assumption that the ε_t are iid by considering the asymptotic distribution of the test statistics and correcting for serial dependence. We assume the same regularity conditions as Phillips and Perron (1988, p. 336); these put some limits on the degree of heterogeneity and autocorrelation allowed in the ε sequence but are otherwise fairly general. We define the two nuisance parameters

$$\sigma_\varepsilon^2 = \lim_{T \rightarrow \infty} T^{-1} E \left(\sum_{t=1}^T \varepsilon_t^2 \right) \quad (19)$$

$$\sigma^2 = \lim_{T \rightarrow \infty} T^{-1} E \left(\sum_{t=1}^T \varepsilon_t \right)^2 \quad (20)$$

and assume that $\sigma_\varepsilon^2, \sigma^2 > 0$. We also define the ratio $\omega^2 = \sigma_\varepsilon^2 / \sigma^2$.

The basic insight behind the asymptotic distribution theory is simple. From (18) we see that under the null hypothesis the \tilde{S} process is the cumulative sum process of the deviations from means of the ε process. Therefore \tilde{S} (appropriately normalized) converges to a Brownian bridge, and our statistics converge to simple functions of a Brownian bridge. The asymptotic distribution theory for our tests is accordingly simpler than the asymptotic distribution theory for the Dickey-Fuller $\hat{\rho}_\tau$ and $\hat{\tau}_\tau$ tests.

In Appendix 3 we derive the following asymptotic results for our tests:

$$\tilde{\rho} \rightarrow - \left(2 \int_0^1 \underline{V}^2 \right)^{-1} \omega^2 \quad (21)$$

$$\tilde{\tau} \rightarrow -(1/2) \left(\int_0^1 \underline{V}^2 \right)^{-1/2} \omega. \quad (22)$$

TABLE 1A
Critical Values for $\hat{\tau}$

T	0.01	0.025	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.975	0.99
25	-3.90	-3.50	-3.18	-2.85	-2.48	-2.24	-2.05	-1.89	-1.75	-1.60	-1.45	-1.28	-1.17	-1.08	-1.00
50	-3.73	-3.39	-3.11	-2.80	-2.46	-2.24	-2.06	-1.90	-1.76	-1.62	-1.47	-1.29	-1.16	-1.08	-0.99
100	-3.63	-3.32	-3.06	-2.77	-2.45	-2.24	-2.06	-1.90	-1.76	-1.62	-1.47	-1.29	-1.17	-1.07	-0.97
200	-3.61	-3.30	-3.04	-2.76	-2.45	-2.24	-2.06	-1.91	-1.77	-1.62	-1.47	-1.29	-1.16	-1.07	-0.97
500	-3.59	-3.29	-3.04	-2.76	-2.44	-2.22	-2.05	-1.90	-1.75	-1.61	-1.47	-1.29	-1.16	-1.07	-0.98
1000	-3.58	-3.28	-3.02	-2.75	-2.43	-2.22	-2.05	-1.90	-1.76	-1.62	-1.47	-1.29	-1.16	-1.07	-0.98
2000	-3.56	-3.27	-3.02	-2.75	-2.44	-2.22	-2.05	-1.90	-1.76	-1.62	-1.47	-1.29	-1.16	-1.07	-0.97

Critical Values for \hat{p}

T	0.01	0.025	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.975	0.99
25	-20.4	-17.9	-15.7	-13.4	-10.9	-9.27	-8.02	-6.98	-6.05	-5.21	-4.35	-3.43	-2.89	-2.51	-2.16
50	-22.8	-19.6	-17.0	-14.3	-11.4	-9.62	-8.26	-7.12	-6.14	-5.24	-4.35	-3.37	-2.78	-2.39	-2.03
100	-23.8	-20.4	-17.5	-14.6	-11.6	-9.76	-8.34	-7.17	-6.17	-5.23	-4.32	-3.35	-2.75	-2.32	-1.92
200	-24.8	-20.9	-17.9	-14.9	-11.8	-9.86	-8.42	-7.23	-6.20	-5.22	-4.30	-3.31	-2.72	-2.30	-1.89
500	-25.3	-21.3	-18.1	-15.0	-11.8	-9.78	-8.32	-7.15	-6.14	-5.19	-4.29	-3.31	-2.70	-2.28	-1.90
1000	-25.3	-21.3	-18.1	-15.0	-11.8	-9.78	-8.32	-7.15	-6.14	-5.19	-4.29	-3.31	-2.70	-2.28	-1.90
2000	-25.2	-21.2	-18.1	-15.0	-11.8	-9.85	-8.39	-7.19	-6.15	-5.21	-4.30	-3.32	-2.68	-2.27	-1.87

TABLE IB
Critical Values for $\hat{\tau}$

T	p=2				p=3				p=4			
	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10
25	-4.52	-4.09	-3.77	-3.41	-5.07	-4.62	-4.26	-3.89	-5.57	-5.09	-4.70	-4.30
50	-4.28	-3.93	-3.65	-3.34	-4.73	-4.37	-4.08	-3.77	-5.13	-4.78	-4.47	-4.15
100	-4.16	-3.84	-3.60	-3.31	-4.59	-4.29	-4.03	-3.72	-4.99	-4.66	-4.39	-4.10
200	-4.12	-3.81	-3.55	-3.28	-4.53	-4.24	-3.99	-3.69	-4.90	-4.60	-4.33	-4.06
500	-4.08	-3.80	-3.55	-3.26	-4.50	-4.20	-3.96	-3.68	-4.85	-4.56	-4.31	-4.03
1000	-4.06	-3.78	-3.53	-3.26	-4.49	-4.19	-3.95	-3.68	-4.83	-4.54	-4.31	-4.03
2000	-4.06	-3.78	-3.52	-3.26	-4.44	-4.16	-3.93	-3.67	-4.81	-4.52	-4.29	-4.01

Critical Values for $\hat{\rho}$

T	p=2				p=3				p=4			
	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10
25	-24.6	-22.2	-20.1	-17.8	-28.1	-25.8	-23.8	-21.5	-31.0	-28.8	-26.9	-24.7
50	-28.4	-25.1	-22.4	-19.5	-33.1	-29.8	-27.0	-24.0	-37.4	-34.1	-31.2	-28.1
100	-30.4	-26.6	-23.7	-20.4	-36.3	-32.3	-29.1	-25.4	-41.8	-37.4	-34.0	-30.2
200	-31.8	-27.5	-24.2	-20.7	-38.0	-33.6	-30.1	-26.1	-44.0	-39.2	-35.2	-31.2
500	-32.4	-28.2	-24.8	-21.0	-39.1	-34.4	-30.6	-26.6	-45.3	-40.3	-36.2	-31.8
1000	-32.5	-28.2	-24.6	-21.1	-39.5	-34.6	-30.8	-26.7	-45.7	-40.6	-36.6	-32.0
2000	-32.6	-28.3	-24.7	-21.1	-39.7	-34.4	-30.6	-26.7	-45.8	-40.5	-36.4	-31.9

Here $V(r)$ is a standard Brownian bridge on the interval $[0, 1]$ and $\underline{V}(r)$ is the demeaned Brownian bridge

$$\underline{V}(r) = V(r) - \int_0^1 V(r) dr. \quad (23)$$

The symbol ' \rightarrow ' in (21) and (22) signifies weak convergence of the associated probability measures.

Schmidt and Lee (1991, Section 2) demonstrates analogous results for the statistics $\tilde{\rho}$ and $\tilde{\tau}$. The only difference is that the Brownian bridge $V(r)$ replaces the demeaned Brownian bridge $\underline{V}(r)$ in (21) and (22). The result in (21) is also equivalent to the result of Nabeya and Tanaka (1990) for their statistic R_3 , which equals $1/(T \cdot \text{Bhargava's } R_2)$ and is therefore almost equal to $-1/(2\tilde{\rho})$. However, their algebraic expression is rather different, and so are their regularity conditions in the presence of error autocorrelation.

The asymptotic formulae (21) and (22) require only simple corrections to remove the effects of dependent and heterogeneous errors. Multiplying $\tilde{\rho}$ by a consistent estimate of $1/\omega^2 = (\sigma^2/\sigma_\varepsilon^2)$ yields a corrected test statistic whose asymptotic distribution is identical to the asymptotic distribution that $\tilde{\rho}$ would have under iid errors, so that the critical values given in Table 1 are asymptotically correct. Similarly, multiplying $\tilde{\tau}$ by a consistent estimate of $1/\omega = (\sigma/\sigma_\varepsilon)$ yields a corrected test statistic for which the critical values in Table 1 are asymptotically correct. These corrections are very simple in comparison with the corrections given in Phillips (1987) and Phillips and Perron (1988) for the Dickey-Fuller tests.

Estimation of σ^2 and σ_ε^2 can be performed along the lines suggested in Phillips (1987), Phillips and Ouliaris (1990), and Phillips and Perron (1988). In particular, the arguments given in Phillips and Ouliaris (1990) apply and the consistency of both tests requires that σ^2 and σ_ε^2 be estimated from regression residuals rather than first differences. Thus, let $\hat{\varepsilon}_t$ be the residuals from a least squares regression on (3). Then by arguments analogous to those of Theorem 4.2 in Phillips (1987) we find that the following estimates are consistent for the variance parameters under the null:

$$s^2 = T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t^2 \rightarrow_p \sigma_\varepsilon^2 \quad (24)$$

and

$$s^2(l) = T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t^2 + 2T^{-1} \sum_{s=1}^l \sum_{t=s+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-s} \rightarrow_p \sigma^2. \quad (25)$$

In the case of $s^2(l)$ we require that the lag truncation parameter $l \rightarrow \infty$ as $T \rightarrow \infty$. The rate $l = O(T^{1/2})$ will usually be satisfactory, as for the case of stationary sequences ε_t . Of course, many other consistent estimates of σ^2 are

available using a variety of lag windows (see Andrews (1991) for a discussion and analysis of alternatives, including data based choices of l).

With these estimates we construct $\hat{\omega}^2 = s^2/s^2(l)$ and the test statistics

$$Z(\rho) = \hat{\rho}/\hat{\omega}^2, \quad Z(\tau) = \hat{\tau}/\hat{\omega}. \quad (26)$$

Under the null hypothesis we have

$$Z(\rho) \rightarrow -\left(2 \int_0^1 Y^2\right)^{-1}, \quad Z(\tau) \rightarrow -(1/2) \left(\int_0^1 Y^2\right)^{-1/2}. \quad (27)$$

These limit distributions are free of nuisance parameters and they are negative almost surely. Under the alternative hypothesis that $|\beta| < 1$ we find that $Z(\rho) = O_p(T)$, $Z(\tau) = O_p(T^{1/2})$ as in Theorem 5.1 of Phillips and Ouliaris (1990). Thus, the statistics diverge under the alternative and the two tests are consistent, but at different rates as $T \rightarrow \infty$.

REMARK. As observed above, the construction of a consistent test requires the use of regression residuals rather than first differences. This means that a regression such as (3) is needed, at least at this stage, to remove nuisance parameters. Although this does not cause any loss in asymptotic local power, it seems likely that it will have finite sample effects in terms of some size distortion and power loss.

V EXTENSIONS TO HIGHER ORDER POLYNOMIAL TRENDS

We now wish to replace the linear deterministic trend in (4) with a higher order polynomial trend. To do so, we first consider the more general model

$$y_t = \alpha + Z_t \delta + X_t, \quad X_t = \beta X_{t-1} + \varepsilon_t, \quad (28)$$

where Z_t is at this point a general row vector of explanatory variables. The null hypothesis is $\beta = 1$, as before, and to construct the LM statistic we need to consider the differenced version of (28), namely

$$\Delta y_t = \Delta Z_t \delta + u_t \quad (29)$$

(where $u_t = \varepsilon_t$ under the null hypothesis). Define the restricted MLE's: $\hat{\delta}$ = OLS estimate of δ from (29), $\tilde{\psi}_X = y_1 - Z_1 \hat{\delta}$; and define

$$\tilde{S}_t = y_t - \tilde{\psi}_X - Z_t \hat{\delta}. \quad (30)$$

Finally, run the regression

$$\Delta y_t = \Delta Z_t \gamma + \phi \tilde{S}_{t-1} + \text{error}. \quad (31)$$

We again define $\tilde{\rho} = T\tilde{\phi}$, where $\tilde{\phi}$ is the least squares estimate of ϕ in (31), and $\tilde{\tau}$ = usual t statistic for the hypothesis $\phi = 0$.

We are specifically interested in the case $Z_t = [t, t^2, \dots, t^p]$, so that the model allows a p^{th} order polynomial trend. In this case the differenced model is

equivalent to a $(p-1)^{\text{th}}$ order time trend, and it is convenient to rewrite the above general expressions as follows. The model is

$$y_t = \sum_{j=0}^p a_j t^j + X_t, X_t = \beta X_{t-1} + \varepsilon_t. \quad (32)$$

The differenced model can be written as

$$\Delta y_t = \sum_{j=0}^{p-1} b_j t^j + u_t. \quad (33)$$

Define $\tilde{u}_1 = 0$ and $\tilde{u}_t = \text{OLS residual from (33), } t=2, \dots, T$. Then it is easy to show that \tilde{S} as defined in (30) can be calculated as the partial sum of the \tilde{u} :

$$\tilde{S}_t = \sum_{k=1}^t \tilde{u}_k, \quad (34)$$

and the regression (31) that defines the test statistics is simply

$$\Delta y_t = \sum_{j=0}^{p-1} c_j t^j + \phi \tilde{S}_{t-1} + \text{error}. \quad (35)$$

From the limit theory given in Appendix 3(ii) we have

$$\tilde{\rho} \rightarrow - \left(2 \int_0^1 Y_p^2 \right)^{-1} \omega^2, \quad \tilde{\tau} \rightarrow -(1/2) \left(\int_0^1 Y_p^2 \right)^{-1/2} \omega. \quad (36)$$

In the above formulae $Y_p(r)$ is a detrended p -level Brownian bridge; i.e.

$$Y_p(r) = V_p(r) - \sum_{j=0}^{p-1} \hat{\alpha}_j r^j \quad (37)$$

and

$$\hat{\alpha} = \operatorname{argmin}_{\alpha} \int_0^1 \left(V_p(r) - \sum_{j=0}^{p-1} \alpha_j r^j \right)^2 dr. \quad (38)$$

Here $V_p(r)$ is a Gaussian process which we call a p -level Brownian bridge. It can be defined in terms of standard Brownian motion $W(r)$ as:

$$V_p(r) = W(r) - \left(\int_0^1 dW(s) g(s)' \right) Q^{-1} q(r), \quad (39)$$

where $g(s)' = (1, s, \dots, s^{p-1})$, Q is $p \times p$ with $(i, j)^{\text{th}}$ element $q_{ij} = 1/(i+j-1)$ and $q(r)$ is $p \times 1$ with i^{th} element r^i/i . An alternative but more complicated

expression is given by MacNeill (1978, p. 426); our $V_p(r)$ is MacNeill's $B_{p-1}(t)$, which he terms a generalized Brownian bridge. As noted in Appendix 3(ii), $V_p(r)$ is tied down in the $[0, 1]$ interval with $V_p(0) = V_p(1) = 0$ just like a Brownian bridge. In fact when $p = 1$ we have $V_1(r) = V(r)$, a simple Brownian bridge; and when $p = 0$ we have $V_0(r) = W(r)$, a standard Brownian motion. Also as shown in (A3.8) of Appendix 3(ii), $V_p(r)$ is the weak limit of a standardized partial sum of detrended innovations. Thus, writing ε_t as the residual in the regression of ε_t on a time trend of order $p - 1$, viz.

$$\varepsilon_t = \sum_{j=0}^{p-1} \hat{\delta}_j t^j + \varepsilon_t, \quad (40)$$

we have

$$T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_t \rightarrow V_p(r). \quad (41)$$

The detrended process $V_p(r)$ is most easily interpreted as a Hilbert projection in $L_2[0, 1]$ of the process $V_p(r)$ on the orthogonal complement of the space spanned by the trend functions $\{r^j; j = 0, 1, \dots, p - 1\}$.

The nuisance parameter ω , or variance ratio ω^2 , that appears in the limit formulae (39) may be eliminated by transformation as discussed in the preceding section, leading to test statistics $Z(\rho)$ and $Z(\tau)$.

Table 1B gives the critical values for the test statistics $\tilde{\rho}$ and $\tilde{\tau}$, for $p = 2, 3$ and 4 (where p is the order of the deterministic polynomial trend in the model (32)). These are calculated as described in Section III.

VI POWER OF THE TESTS

In this section we perform some Monte Carlo experiments to compare the power of the tests proposed in this paper with the power of the Dickey-Fuller tests. In the Monte Carlo experiment we consider the performance of five tests: the Dickey-Fuller tests $\hat{\rho}_\mu$, $\hat{\tau}_\mu$, $\hat{\rho}_\tau$ and $\hat{\tau}_\tau$, and the test $\tilde{\rho}$ proposed in this paper. We do not display results for the test $\tilde{\tau}$ because, in the absence of corrections for error autocorrelation, $\tilde{\rho}$ and $\tilde{\tau}$ have exactly the same power. The focus of the experiments will be on power against trend stationary alternatives, and we do not expect the $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$ tests to have much power against such alternatives. These tests would be expected to be more powerful than the other four tests against level stationary alternatives, and it is interesting to see how much trend or how large a sample size it takes before the $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$ tests are dominated by the other tests. However, the main focus of the experiments is to compare the power of the new test $\tilde{\rho}$ with the power of the Dickey-Fuller $\hat{\rho}_\tau$ and $\hat{\tau}_\tau$ tests.

The parameters that are relevant are sample size (T), β , standardized trend (ξ^*) and standardized initial condition $X_0^* = X_0/\sigma_\varepsilon$. Under the null

hypothesis that $\beta=1$, the tests $\hat{\rho}$, $\hat{\rho}_r$ and $\hat{\tau}_r$ have distributions that are independent of the nuisance parameters ψ , ξ^* , X_0^* and σ_ε , while the distributions of $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$ depend on ξ^* but are independent of ψ , X_0^* and σ_ε . When β is not equal to one, the distributions of the $\hat{\rho}$, $\hat{\rho}_r$ and $\hat{\tau}_r$ tests depend only on T , β and X_0^* ; and when the ε_i are symmetrically distributed, they depend only on T , β and $|X_0^*|$.

The results of our experiments are given in Tables 2-3. The results were generated by a simulation using 20,000 replications; the random number generator was described in Section III. The tables give percentages of rejections for 5 percent lower tail tests. Other significance levels would tell essentially the same story.

Our first experiment, called Experiment 1 in Table 2, studies the size of the various tests under the null hypothesis. We set $T=100$, $\beta=1$, $X_0^*=0$ (its value is irrelevant anyway), and varied ξ^* . Specifically, we considered values

TABLE 2
Size and Power, 5% Lower Tail Tests, $T=100$

Exp. no.	T	β	ξ^*	X_0^*	$\hat{\tau}_\mu$	$\hat{\rho}_\mu$	$\hat{\tau}_r$	$\hat{\rho}_r$	$\hat{\rho}$
1	100	1	0	0	0.049	0.048	0.048	0.050	0.052
1	100	1	0.02	0	0.050	0.046	0.048	0.050	0.052
1	100	1	0.05	0	0.046	0.045	0.048	0.050	0.052
1	100	1	0.10	0	0.041	0.034	0.048	0.050	0.052
1	100	1	0.20	0	0.022	0.008	0.048	0.050	0.052
1	100	1	0.50	0	0.006	0.000	0.048	0.050	0.052
2	100	0.90	0	-10	0.686	0.317	0.304	0.120	0.033
2	100	0.90	0	-5	0.413	0.421	0.211	0.198	0.165
2	100	0.90	0	-2	0.334	0.459	0.191	0.234	0.252
2	100	0.90	0	-1	0.322	0.465	0.188	0.239	0.267
2	100	0.90	0	0	0.321	0.467	0.186	0.239	0.270
3	100	0.90	0	0	0.321	0.467	0.186	0.239	0.270
3	100	0.90	0.02	0	0.281	0.409	0.186	0.239	0.270
3	100	0.90	0.05	0	0.140	0.198	0.186	0.239	0.270
3	100	0.90	0.10	0	0.016	0.014	0.186	0.239	0.270
3	100	0.90	0.20	0	0.000	0.000	0.186	0.239	0.270
4A	100	1	0	0	0.049	0.048	0.048	0.050	0.052
4A	100	0.95	0	0	0.120	0.190	0.082	0.098	0.108
4A	100	0.90	0	0	0.321	0.467	0.186	0.239	0.270
4A	100	0.80	0	0	0.871	0.950	0.644	0.734	0.765
4B	100	1	0	-2	0.049	0.046	0.048	0.050	0.052
4B	100	0.95	0	-2	0.121	0.181	0.082	0.095	0.104
4B	100	0.90	0	-2	0.334	0.459	0.191	0.234	0.252
4B	100	0.80	0	-2	0.886	0.952	0.658	0.732	0.707

TABLE 3
Size and Power, 5% Lower Tail Tests, $T = 200, 500$

Exp. no.	T	β	ξ^*	X_0^*	$\hat{\tau}_\mu$	$\hat{\rho}_\mu$	$\hat{\tau}_\tau$	$\hat{\rho}_\tau$	$\bar{\rho}$
5	200	0.90	0	0	0.858	0.946	0.617	0.724	0.763
5	200	0.90	0	-1	0.860	0.947	0.620	0.720	0.752
5	200	0.90	0	-2	0.868	0.947	0.627	0.719	0.720
5	200	0.90	0	-5	0.908	0.949	0.677	0.710	0.526
6A	200	1	0	0	0.048	0.048	0.048	0.048	0.050
6A	200	0.95	0	0	0.310	0.460	0.178	0.233	0.266
6A	200	0.90	0	0	0.858	0.946	0.617	0.724	0.763
6A	200	0.80	0	0	1.00	1.00	0.999	1.00	0.997
6B	500	1	0	0	0.053	0.051	0.051	0.051	0.051
6B	500	0.95	0	0	0.967	0.994	0.819	0.897	0.914
6B	500	0.90	0	0	1.00	1.00	1.00	1.00	1.00

0, 0.02, 0.05, 0.10, 0.20 and 0.50 for ξ^* . All tests except $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$ should have size equal to the nominal critical value (0.05), and this is so apart from randomness. (With 20,000 replications, a 95% confidence interval around 0.05 is approximately [0.047, 0.053].) When $\xi^* = 0$, the $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$ tests should also have size equal to the nominal critical value, and they do apart from randomness. However, the size of the $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$ tests should decrease to zero as ξ^* increases, for fixed T , or as T increases for fixed ξ^* . For $T = 100$, we can see in Table 2 that the size of these tests does go to zero as ξ^* increases; it is nearly zero for ξ^* as large as 0.50. Results for $T = 200$ and $T = 500$ (presented in earlier versions of the paper, and available on request) confirm that the size distortion of the $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$ tests is larger for larger sample sizes; the larger T , the smaller the value of ξ^* required to produce substantial size distortions, and conversely.

The values of ξ^* considered here are empirically relevant. Recall that ξ^* is standardized trend, equal to ξ/σ_ϵ . We can estimate ξ by $\bar{\xi} = \text{mean } \Delta y = (y_T - y_1)/(T - 1)$; this is the MLE subject to $\beta = 1$, but it is a consistent estimate of ξ even if β is not equal to one. Similarly, imposing the unit root, the MLE of σ_ϵ^2 is the empirical variance of the Δy 's, and this is a consistent estimate even if y is stationary. Thus a consistent estimate of ξ^* is just the mean of the Δy 's divided by the standard deviation of the Δy 's. Schmidt (1990) provides values of this measure (which he calls 'standardized drift') for the Nelson-Plosser data, and values in the range [0.2, 0.5] are the norm. Smaller values of ξ^* might be expected in higher-frequency data, or in financial data, but it is nevertheless clear that the size distortions for the $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$ tests are potentially very serious for values of sample size and trend encountered in economic data.

Experiment 2, reported in Table 2, explores the effect of the initial condition X_0^* on the power of the various tests. We set $T=100$, $\beta=0.90$, and $\xi^*=0$. (The value of ξ^* would matter only for $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$.) We consider values of X_0^* ranging from -10 to zero; positive values of X_0 are unnecessary because only $|X_0^*|$ matters when the errors are symmetrically distributed. The $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$ tests are more powerful than the other four tests, because experiment 2 has $\xi^*=0$ (no trend). A more interesting comparison is between the powers of the Dickey-Fuller $\hat{\rho}_\tau$ and $\hat{\tau}_\tau$ tests and $\tilde{\rho}$ test. All three tests have power that is monotonic in $|X_0^*|$, but not all in the same direction: as $|X_0^*|$ increases, the power of $\hat{\tau}_\tau$ increases, while the powers of $\hat{\rho}_\tau$ and $\tilde{\rho}$ decrease. The power of the $\hat{\rho}_\tau$ test is higher than the power of the $\hat{\tau}_\tau$ test (the conventional wisdom), except when X_0^* is large, in which case this ranking reverses.

Comparing the new test $\tilde{\rho}$ to the Dickey-Fuller $\hat{\rho}_\tau$ and $\hat{\tau}_\tau$ tests, we see that the new test is more powerful than the Dickey-Fuller tests for $|X_0^*| \leq 2$ and less powerful for $|X_0^*| \geq 5$. A more detailed picture of the power curves for this case is given in Figure 1, which displays power (for $T=100$ and $\beta=0.9$) as a function of $|X_0^*|$, for $|X_0^*|$ between zero and five, for the $\tilde{\rho}$, $\hat{\rho}_\tau$ and $\hat{\tau}_\tau$ tests. As in Table 2, the $\tilde{\rho}$ test is most powerful for small X_0^* . The $\hat{\rho}_\tau$ test is more powerful than $\tilde{\rho}$ for $|X_0^*| > 2.6$, and the $\hat{\tau}_\tau$ test is more powerful than $\tilde{\rho}$ for $|X_0^*| > 3.6$. To put this into perspective, note that under stationarity the standard deviation of X_0^* is $(1-\beta^2)^{-1/2}$, which equals 2.29 for $\beta=0.9$. Thus $\tilde{\rho}$ is more powerful than $\hat{\rho}_\tau$ except when $|X_0^*|$ is about 1.1 standard deviations (of X_0^*) away from zero, so that under normality $\tilde{\rho}$ is more powerful than $\hat{\rho}_\tau$ with a probability of about 0.73. Similarly, $\tilde{\rho}$ is more powerful than $\hat{\tau}_\tau$ with a probability of about 0.88. Our summary of these results is that the new $\tilde{\rho}$ test dominates the Dickey-Fuller $\hat{\rho}_\tau$ and $\hat{\tau}_\tau$ tests except for values of X_0^* that are unreasonably large, but the reader is of course free to draw his or her own conclusions.

Experiment 3 systematically varies the trend parameter ξ^* , for $X_0^*=0$ and $\beta=0.90$. The results are consistent with the conclusion that $\xi^*>0.05$ is sufficient to make the power of the $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$ tests less than the power of the tests that explicitly allow for trend. It should be stressed that, empirically, this is a very small value of ξ^* , and these results argue strongly that it is a mistake to apply the $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$ tests to data with noticeable trend. Accordingly, we will not discuss the results for the $\hat{\rho}_\mu$ and $\hat{\tau}_\mu$ tests further.

Experiment 4, made up of parts 4A and 4B, is also reported in Table 2. This experiment varies β over the range from one to 0.80, for $X_0^*=0$ and -2 . There are no surprises in the results. Over this range of X_0^* , we expect the $\tilde{\rho}$ test to be more powerful than the $\hat{\rho}_\tau$ and $\hat{\tau}_\tau$ tests, and it is for all three values of $\beta \neq 1$.

Table 3 gives the results of some experiments done for $T=200$ and 500. The experiments consider power under the alternative: Experiment 5 varies X_0^* for $T=200$, $\xi^*=0$ and $\beta=0.90$; and Experiment 6 varies β for $T=200$ and 500 and $X_0^*=\xi^*=0$. The results are in line with the previous discussion and so their detailed analysis is left to the reader.

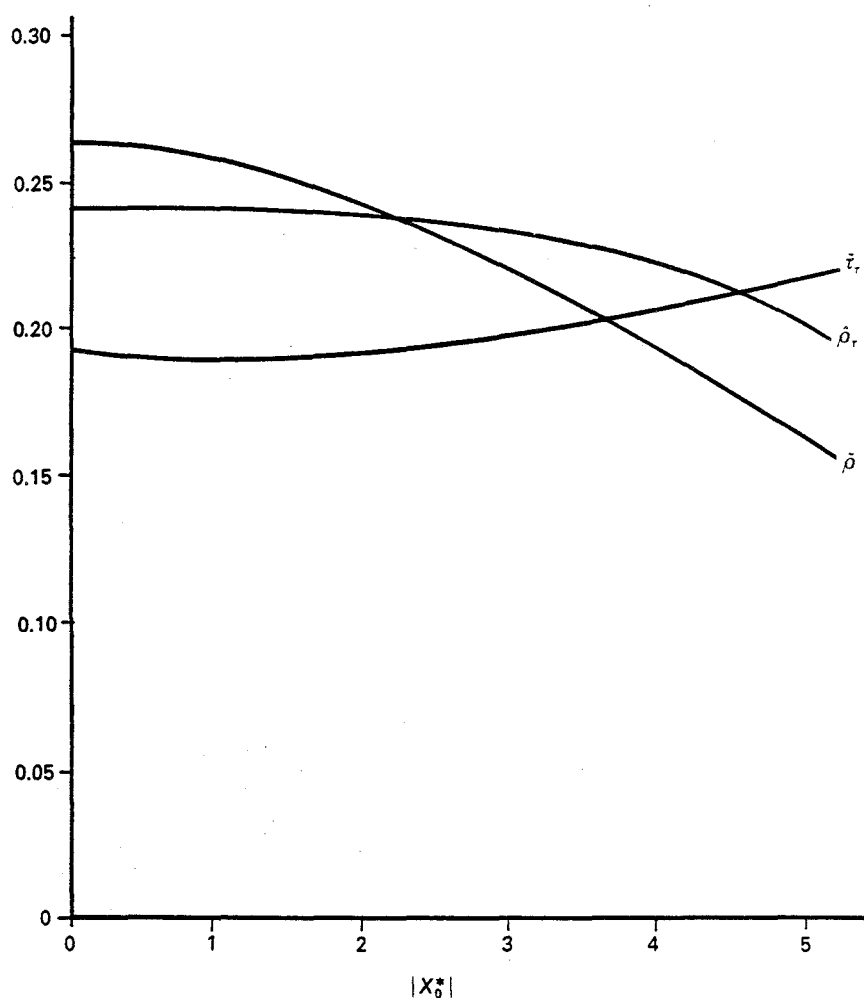


Fig. 1. Power of various tests, $T=100$, $\beta=0.9$

The results of Experiments 1-6 can be well summarized by a numerical response surface. We ran separate response surface regressions for each test ($\bar{\rho}$, $\tilde{\rho}_r$ and $\tilde{\rho}$) using the 17 distinct observations (combinations of T , β and X_0^*) in Experiments 1-6. After some experimentation we settled on the functional form:

$$\ln[(\text{Power} - 0.05)/(1 - \text{Power})] = \alpha_0 + \alpha_1 \ln T + \alpha_2 \ln(1 - \beta) + \alpha_3 |X_0^*| + \alpha_4 X_0^{*2}.$$

This functional form makes $\text{Power} \rightarrow 0.05$ as $\beta \rightarrow 1$ and $\text{Power} \rightarrow 1$ as $T \rightarrow \infty$ ($\beta \neq 1$), as it should. Also, in our experiments power appears similar for

T, β pairs such that $T(1 - \beta)$ is constant (e.g. the tabulated results for $T = 100, \beta = 0.9$ and $T = 200, \beta = 0.95$), which would support the restriction $\alpha_1 = \alpha_2$. Imposing this restriction did not change the results much. Finally, given our functional form, we need to drop one observation ($T = 100, \beta = 0.9, X_0^* = -10$) for $\hat{\rho}$ because $\text{Power} < 0.05$. We also eliminated two observations ($T = 200, \beta = 0.8, X_0^* = 0$ and $T = 500, \beta = 0.9, X_0^* = 0$) for which power was equal or nearly equal to unity. Our fitted response surfaces were as follows (standard errors in parentheses):

	α_0	α_1	α_2	α_3	α_4	R^2	Number of experiments
$\hat{\rho}_t$	-9.255 (0.720)	3.182 (0.147)	2.882 (0.159)	0.019 (0.069)	-0.015 (0.007)	0.985	15
$\hat{\tau}_t$	-8.924 (0.618)	3.015 (0.126)	2.862 (0.137)	0.001 (0.059)	0.006 (0.006)	0.987	15
$\hat{\rho}$	-8.873 (0.749)	3.078 (0.151)	2.747 (0.159)	0.008 (0.127)	-0.043 (0.025)	0.984	14

The main conclusion that emerges from our experiments is that the $\hat{\rho}$ test proposed in this paper is more powerful than the Dickey-Fuller $\hat{\rho}_t$ and $\hat{\tau}_t$ tests except when the initial condition term (X_0^*) is large in absolute value. The empirical relevance of this finding can be argued. From the point of view of estimation, X_0^* is a parameter. However, it is not identified when $\beta = 1$ (only the sum $\psi_X = \psi + X_0$ is then identified), and we should not expect to estimate it well when β is close to one. Therefore the possible strategy of choosing a test, for a given sample, on the basis of estimated X_0^* is probably not to be recommended. An alternative is to consider X_0^* as random and drawn from the stationary distribution of X_t/σ_ϵ , which under normality is $N[0, 1/(1 - \beta^2)]$, and ask how likely it is that X_0^* takes a value that favours the new $\hat{\rho}$ test over the Dickey-Fuller $\hat{\rho}_t$ and $\hat{\tau}_t$ tests. As reported above, the new test dominates the Dickey-Fuller tests over the most likely values of X_0^* ; for example, we reported (for $T = 100, \beta = 0.9$) that $\hat{\rho}$ is more powerful than $\hat{\tau}_t$ with a probability of about 0.88, where this probability is taken over the distribution of X_0^* . This is a reasonable calculation, but one may object that it fails to address the question of how large the power differences are over various regions of the distribution of X_0^* . In particular, $\hat{\rho}$ is modestly more powerful than $\hat{\tau}_t$ over most (in terms of probability) of the values of X_0^* , but it is considerably less powerful for some values of X_0^* .

An obvious way to see which tests are most powerful 'on average' is simply to treat X_0^* as drawn from the stationary distribution of X_t/σ_ϵ , and then to calculate the powers of the test *not* conditional on X_0^* . Table 4 gives the results of two experiments in which X_0^* is drawn from $N[0, 1/(1 - \beta^2)]$, with a different X_0^* drawn for each of the 50,000 replications of an experiment. This

TABLE 4
Size and Power, 5% Lower Tail Tests X_0^* Drawn from $N[0, (1 - \beta^2)^{-1}]$

Exp. no.	T	β	ξ^*	$\hat{\tau}_\mu$	$\hat{\rho}_\mu$	$\hat{\tau}_\tau$	$\hat{\rho}_\tau$	$\hat{\rho}$
7	100	0.95	0	0.126	0.167	0.081	0.090	0.098
7	100	0.90	0	0.333	0.447	0.187	0.225	0.242
7	100	0.85	0	0.640	0.776	0.390	0.464	0.483
7	100	0.80	0	0.880	0.951	0.647	0.732	0.728
7	100	0.70	0	0.997	1.000	0.961	0.982	0.954
8	500	0.99	0	0.119	0.170	0.081	0.094	0.098
8	500	0.98	0	0.312	0.445	0.184	0.225	0.237
8	500	0.97	0	0.601	0.763	0.374	0.452	0.464
8	500	0.95	0	0.996	0.994	0.828	0.892	0.869
8	500	0.90	0	1.000	1.000	1.000	1.000	0.999

corresponds to a calculation of the unconditional (on X_0^*) power of the various tests. Experiment 7 gives power for $T=100$ and for β between 0.70 and 0.95, while experiment 8 has $T=500$ and β between 0.90 and 0.99. Incidentally, the power of $\hat{\rho}$ for $T=100$ corresponds closely to the results of Bhargava (1986, Table II, p. 378) for his test R_2 , as would be expected.

The results of these experiments are easy to summarize. First, the $\hat{\rho}_\tau$ test dominates the $\hat{\tau}_\tau$ test. Second, the new $\hat{\rho}$ test is more powerful than the Dickey-Fuller $\hat{\rho}_\tau$ and $\hat{\tau}_\tau$ tests for T and β such that power is low, and less powerful than the Dickey-Fuller tests for T and β such that power is high. This last conclusion is intuitively reasonable. The differences between the $\hat{\rho}$ test and the Dickey-Fuller tests is the way in which level and trend are removed from the data; that is, the way in which parameters representing level and trend are estimated. The new test estimates these parameters from a regression in first differences, while the Dickey-Fuller tests estimate these parameters from a regression in levels. Estimation in differences is superior when the null is true, or presumably when it is close to being true; that is, when power is low. On the other hand, estimation in levels is superior when the null is far from being true; that is, when power is high.

Schmidt and Lee (1991) show that the comparison between the $\hat{\rho}$ and $\hat{\rho}$ tests is qualitatively similar to the comparison just given between the $\hat{\rho}$ test and the Dickey-Fuller tests. The $\hat{\rho}$ test is more powerful than the $\hat{\rho}$ test when power is low, and less powerful when power is high. However, the gain in power when power is low is quite small, while the loss in power when power is higher is more substantial. For this reason they recommend $\hat{\rho}$ rather than $\hat{\rho}$.

VII CONCLUDING REMARKS

This paper has proposed tests of the unit root hypothesis based on the LM (score) principle. They are based on a different parameterization than the Dickey-Fuller tests. The choice of a parameterization is to some extent a

matter of taste. However, the parameterization we use has two important advantages. First, the meaning of the parameters governing level and trend is independent of whether or not the unit root hypothesis is true. Second, the analysis of the distributional properties of both the new tests and the Dickey-Fuller $\hat{\rho}_\tau$ and $\hat{\tau}_\tau$ tests is simplified.

Although the new tests were not derived on the basis of considerations of invariance, they do have the property that their distributions under the null hypothesis are independent of the nuisance parameters reflecting level, trend and variance. Because they were derived as LM tests, they should be expected to have good local power properties. Our simulation results indicate that the comparison of their power to the power of the Dickey-Fuller $\hat{\rho}_\tau$ and $\hat{\tau}_\tau$ tests hinges on an initial conditions parameter (X_0^*), and that they should be more powerful than the Dickey-Fuller tests except for values of this parameter that are large enough to be unlikely. If X_0^* is treated as random and drawn from the stationary distribution of X_t/σ_ϵ , the new tests are more powerful than the Dickey-Fuller tests for T and β such that power is low, and less powerful for T and β such that power is high.

The LM test procedure used in this paper can be used in other settings, with appropriate modifications to the distributional theory. Essentially, we have tested the hypothesis that the error in a regression has a unit root, where the regressors form a deterministic trend. This could be extended to accommodate stochastic trends. Thus, if the regressor is an $I(1)$ variable rather than a deterministic trend, the unit root test becomes a cointegration test. In fact, the cointegration test of Hansen (1990) can be derived in this way. The change in the nature of the regressors will change the asymptotic distribution of the test statistics, but in ways that are presumably straightforward.

A reasonable extension of this paper is to find alternative ways to correct for error autocorrelation. It is intuitively reasonable, following Said and Dickey (1984), that we can (asymptotically) correct the τ statistic for the effects of error autocorrelation by including lagged values of Δy_t (or $\Delta \hat{S}_t$) in the regression (10) that generates the test statistics. Similarly, Lee and Schmidt (1991) consider IV tests, analogous to those of Hall (1989) but based on the regressions (15A) and (15B), and show that they have good size and power properties in the presence of MA(1) errors.

Michigan State University, MI 48824
Yale University, CN 06520-2125

REFERENCES

- Andrews, D. W. K. (1991). 'Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation'. *Econometrica*, Vol. 59, pp. 817-58.
Bhargava, A. (1986). 'On the Theory of Testing for Unit Roots in Observed Times Series'. *Review of Economic Studies*, Vol. 52, pp. 369-84.

- DeJong, D. N., Nankervis, J. C., Savin, N. E. and Whiteman, C. H. (1989). 'Integration Versus Trend Stationarity in Macroeconomic Time Series', Working Paper 89-99. Department of Economics, University of Iowa.
- Dickey, D. A. (1976). 'Estimation and Hypothesis Testing in Nonstationary Times Series', unpublished Ph.D. dissertation, Iowa State University.
- Dickey, D. A. (1984). 'Powers of Unit Root Tests'. *Proceedings of the American Statistical Association*, Business and Economic Statistics Section, pp. 489-93.
- Dickey, D. A. and Fuller, W. A. (1979). 'Distribution of the Estimators of Autoregressive Time Series with a Unit Root'. *Journal of the American Statistical Association*, Vol. 74, pp. 427-31.
- Dufour, J.-M. and King, M. L. (1989). 'Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary and Nonstationary AR(1) Errors'. *Journal of Econometrics*, Vol. 47, pp. 115-43.
- Evans, G. B. A. and Savin, N. E. (1984). 'Testing for Unit Roots: 2'. *Econometrica*, 52, 1241-69.
- Fuller, W. A. (1976). *Introduction to Statistical Time Series*, John Wiley and Sons, New York.
- Guilkey, D. K. and Schmidt, P. (1992). 'An Interpolation Test for a Unit Root in the Presence of Drift'. *Biometrika*, forthcoming.
- Granger, C. W. J. and Newbold, P. (1974). 'Spurious Regressions in Econometrics'. *Journal of Econometrics*, Vol. 2, pp. 121-30.
- Hall, A. (1989). 'Testing for a Unit Root in the Presence of Moving Average Errors'. *Biometrika*, Vol. 76, pp. 49-56.
- Hansen, B. E. (1990). 'A Powerful, Simple Test of Cointegration Using Cochrane-Orcutt', Working Paper 230, University of Rochester.
- King, M. L. (1981). 'Testing for Random Walk Disturbances in the Linear Regression Model', unpublished manuscript.
- Kiviet, J. F. and Phillips, G. D. A. (1992). 'An Exact Similar t -type Test for Unit Roots'. University of Amsterdam, mimeographed.
- Lee, J. and Schmidt, P. (1991). 'Unit Root Tests Based on Instrumental Variables Estimation', unpublished manuscript, Michigan State University.
- MacNeill, I. B. (1978). 'Properties of Sequences of Partial Sums of Polynomial Regression Residuals with Applications to Tests for Change of Regression at Unknown Times'. *Annals of Statistics*, Vol. 6, pp. 422-33.
- Nabeya, S. and Tanaka, K. (1990). 'Limiting Power of Unit-Root Tests in Time-Series Regression', *Journal of Econometrics*, Vol. 46, pp. 247-71.
- Nankervis, J. C. and Savin, N. E. (1985). 'Testing the Autoregressive Parameter with the t Statistic'. *Journal of Econometrics*, Vol. 27, pp. 143-61.
- Nankervis, J. C. and Savin, N. E. (1987). 'Finite Sample Distributions of t and F Statistics in an AR(1) Model with an Exogenous Variable'. *Econometric Theory*, Vol. 3, pp. 387-408.
- Nelson, C. R. and Plosser, C. I. (1982). 'Trends Versus Random Walks in Macroeconomic Time Series: Some Evidence and Implications'. *Journal of Monetary Economics*, Vol. 10, pp. 139-62.
- Ouliaris, S., Park, J. Y. and Phillips, P. C. B. (1989). 'Testing for a Unit Root in the Presence of a Maintained Trend', Chapter 1, in Raj, B. and Norwell, M. A. (eds.) *Advances in Econometrics and Modelling*, Kluwer Academic Publishers.
- Phillips, P. C. B. (1986). 'Understanding Spurious Regressions in Econometrics'. *Journal of Econometrics*, Vol. 33, pp. 311-40.

- Phillips, P. C. B. (1987). 'Time Series Regression with Unit Roots', *Econometrica*, Vol. 55, pp. 277-301.
- Phillips, P. C. B. and Ouliaris, S. (1990). 'Asymptotic Properties of Residual Based Tests for Cointegration', *Econometrica*, Vol. 58, pp. 165-94.
- Phillips, P. C. B. and Perron, P. (1988). 'Testing for a Unit Root in Time Series Regression', *Biometrika*, Vol. 75, pp. 335-46.
- Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1986). *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge.
- Said, S. E. and Dickey, D. A. (1984). Testing for Unit Roots in Autoregressive — Moving Average Models of Unknown Order'. *Biometrika*, Vol. 71, pp. 599-608.
- Schmidt, P. (1990). 'Dickey-Fuller Tests with Drift', in Fomby, T. and Rhodes, G. (eds.) *Advances in Econometrics: Cointegration, Spurious Regressions, and Unit Roots*, JAI Press, Greenwich, CT.
- Schmidt, P. and Lee, J. (1991). 'A Modification of the Schmidt-Phillips Unit Root Test', *Economics Letters*, Vol. 36, pp. 285-93.
- Stock, J. H. (1990). 'A Class of Tests for Integration and Cointegration', manuscript, Harvard University.
- West, K. D. (1987). 'A Note on the Power of Least Squares Tests for a Unit Root', *Economics Letters*, Vol. 24, pp. 249-52.

APPENDIX I

DERIVATION OF THE LM TEST

We begin with the model as given in equation (4) of the main text. It implies

$$y_1 = \psi + \beta X_0 + \xi + \varepsilon_1$$

$$y_t = \beta y_{t-1} + \psi(1 - \beta) + \xi(t + \beta - t\beta) + \varepsilon_t, t = 2, \dots, T. \quad (\text{A1.1})$$

We assume that the $\varepsilon_t (t = 1, \dots, T)$ are iid $N(0, \sigma^2)$ and we treat the initial condition X_0 as fixed. Since the Jacobian from $(\varepsilon_1, \dots, \varepsilon_T)$ to (y_1, \dots, y_T) is unity, we obtain the log likelihood

$$\ln L = \text{constant} - \frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \text{SSE}, \quad (\text{A1.2})$$

where

$$\text{SSE} = (y_1 - \psi - \beta X_0 - \xi)^2 + \sum_{t=2}^T [(y_t - \beta y_{t-1}) - \psi(1 - \beta) - \xi(t + \beta - t\beta)]^2. \quad (\text{A1.3})$$

At the maximum $\bar{\sigma}^2 = \text{SSE}/T$ and so the concentrated log likelihood is

$$\ln L^* = \text{const.} - \frac{T}{2} \ln(\text{SSE}/T). \quad (\text{A1.4})$$

To derive the MLE's subject to the restriction $\beta = 1$, we note that, when $\beta = 1$, SSE simplifies to

$$\text{SSE}_R = [y_1 - \psi_X - \xi]^2 + \sum_{t=2}^T (\Delta y_t - \xi)^2, \quad \psi_X = \psi + X_0 \quad (\text{A1.5})$$

and this is minimized by the restricted MLE's

$$\begin{aligned} \hat{\xi} &= \overline{\Delta y} = (y_T - y_1)/(T-1) \\ \tilde{\psi}_X &= y_1 - \hat{\xi} = (Ty_1 - y_T)/(T-1), \end{aligned}$$

as given in equations (5) and (6) of the main text.

To derive the LM test we need to calculate the efficient score, evaluated at $\hat{\beta} = 1$:

$$\frac{\partial \ln L^*(\sim)}{\partial \beta} = \frac{-1}{2\sigma^2} \frac{\partial \text{SSE}}{\partial \beta}. \quad (\text{A1.7})$$

It is easy to calculate that

$$\begin{aligned} \frac{\partial \text{SSE}}{\partial \beta} &= -2X_0(y_1 - \psi - \beta X_0 - \xi) - 2 \sum_{t=2}^T [y_{t-1} - \psi - \xi(t-1)] \cdot [(y_t - \beta y_{t-1}) \\ &\quad - \psi(1 - \beta) - \xi(t + \beta - t\beta)]. \end{aligned} \quad (\text{A1.8})$$

Define

$$\tilde{S}_{t-1} = y_{t-1} - \tilde{\psi}_X - \hat{\xi}(t-1), \quad (\text{A1.9})$$

as in equation (7) of the main text. Then (A1.8), evaluated at the restricted MLE's, becomes

$$\frac{\partial \text{SSE}}{\partial \beta} - 2 \sum_{t=2}^T (\Delta y_t - \hat{\xi}) \tilde{S}_{t-1}, \quad (\text{A1.10})$$

and the score becomes

$$\frac{\partial \ln L^*(\sim)}{\partial \beta} = \frac{1}{\hat{\sigma}^2} \sum_{t=2}^T (\Delta y_t - \hat{\xi}) \tilde{S}_{t-1} = \frac{1}{\hat{\sigma}^2} \sum_{t=2}^T (\Delta y_t - \hat{\xi})(\tilde{S}_{t-1} - \bar{\tilde{S}}). \quad (\text{A1.11})$$

(The last equality holds since $(\Delta y_t - \hat{\xi})$ sums to zero.)

The term $\sum (\Delta y_t - \hat{\xi}) \tilde{S}_{t-1}$ in (A1.11) is the numerator of the estimated coefficient (say $\hat{\phi}$) in the regression

$$\Delta y_t = \text{intercept} + \phi \tilde{S}_{t-1} + \text{error}, \quad (\text{A1.12})$$

as given also by equation (10) of the main text. It is also the numerator of the estimated coefficient in regressions (15A) and (15B), as explained in the main text.

To construct the LM test, we also need the information matrix. We calculate

$$\frac{\partial^2 \ln L^*}{\partial \beta^2} = \frac{-1}{\sigma^2} \left\{ X_0^2 + \sum_{t=2}^T [y_{t-1} - \psi - \xi(t-1)]^2 \right\}. \quad (\text{A1.13})$$

Evaluating (A1.13) at the restricted MLE's, and ignoring X_0 (which will be negligible asymptotically), we have

$$\frac{\partial^2 \ln L^*(\sim)}{\partial \beta^2} = \frac{1}{\hat{\sigma}^2} \sum_{t=2}^T \hat{S}_{t-1}^2. \quad (\text{A1.14})$$

We show below that the information matrix is asymptotically block diagonal between β and $[\psi_X, \xi]$. Therefore the LM statistic becomes

$$\text{LM} = \left(\frac{\partial \ln L^*(\sim)}{\partial \beta} \right)^2 / \frac{\partial^2 \ln L^*(\sim)}{\partial \beta^2} \quad (\text{A1.15})$$

and using (A1.11) and (A1.14) we have

$$\text{LM} = \frac{\left[\sum_{t=2}^T (\Delta y_t - \hat{\xi}) \hat{S}_{t-1} \right]^2}{\hat{\sigma}^2 \sum_{t=2}^T \hat{S}_{t-1}^2}. \quad (\text{A1.16})$$

This is the t -statistic for the hypothesis $\phi = 0$ in the regression (15B) of the main text.

Finally, it remains to show that the information matrix is block diagonal. A straightforward calculation yields

$$\frac{\partial^2 \ln L^*(\sim)}{\partial \beta \partial \xi} = \frac{-1}{\hat{\sigma}^2} \left\{ X_0 + \sum_{t=2}^T [(y_t - \hat{\xi})(t-1) + \hat{S}_{t-1}] \right\} \quad (\text{A1.17})$$

$$\frac{\partial^2 \ln L^*(\sim)}{\partial \beta \partial \psi_X} = -X_0 / \hat{\sigma}^2. \quad (\text{A1.18})$$

The appropriate normalization for the information matrix is T^{-2} , since T^{-2} times (A1.14) approaches a limiting distribution, in light of the convergence

$$T^{-2} \sum_i S_{i-1}^2 \rightarrow \sigma^2 \int_0^1 W(r)^2 dr, \quad S_i = \sum_i \varepsilon_i. \quad (\text{A1.19})$$

Here $W(r)$ is the standard Wiener process (Brownian motion) on $[0, 1]$. It is obvious that T^{-2} times (A1.18) approaches zero. The same is true of (A1.17). We have

$$T^{-3/2} \sum_i t \varepsilon_i \rightarrow \sigma \left[W(1) - \int_0^1 W(r) dr \right] \quad (\text{A1.20})$$

$$T^{-3/2} \sum_i S_{i-1} \rightarrow \sigma \int_0^1 W(r) dr. \quad (\text{A1.21})$$

Therefore $T^{-3/2}$ times (A1.17) has a limiting distribution, and T^{-2} times (A1.17) approaches zero.

APPENDIX 2 RELATIONSHIPS BETWEEN TEST STATISTICS

We first establish equation (16) of the main text.

LEMMA 1:

$$\sum_{i=2}^T \Delta \tilde{S}_i \tilde{S}_{i-1} = -\frac{1}{2} \sum_{i=2}^T \Delta \tilde{S}_i^2.$$

PROOF:

$$\begin{aligned} & \sum_{i=2}^T \Delta \tilde{S}_i^2 + 2 \sum_{i=2}^T \Delta \tilde{S}_i \tilde{S}_{i-1} \\ &= \sum_{i=2}^T \tilde{S}_i^2 + \sum_{i=2}^T \tilde{S}_{i-1}^2 - 2 \sum_{i=2}^T \tilde{S}_i \tilde{S}_{i-1} + 2 \sum_{i=2}^T \tilde{S}_i \tilde{S}_{i-1} - 2 \sum_{i=2}^T \tilde{S}_{i-1}^2 \\ &= \sum_{i=2}^T \tilde{S}_i^2 - \sum_{i=2}^T \tilde{S}_{i-1}^2 = \tilde{S}_T^2 - \tilde{S}_1^2 = 0. \end{aligned}$$

LEMMA 2:

$$\tilde{S}_i = [(T-1)y_i - (t-1)y_T - (T-t)y_1].$$

LEMMA 3: Define

$$\bar{S} = T^{-1} \sum_{i=1}^T S_i = T^{-1} \sum_{i=2}^{T-1} S_i.$$

Then

$$\bar{S} = \bar{y} - (y_1 + y_T)/2.$$

THEOREM 1: Bhargava's R_2 statistic can be written

$$R_2 = \frac{\sum_{i=2}^T \Delta S_i^2}{\sum_{i=1}^T (S_i - \bar{S})^2} = -2 \frac{\sum_{i=2}^T \Delta \bar{S}_i \bar{S}_{i-1}}{\sum_{i=1}^T (S_i - \bar{S})^2}.$$

PROOF: The numerator of R_2 is (equation (40), p. 376)

$$\begin{aligned} \sum_{i=2}^T \Delta y_i^2 - (y_T - y_1)^2 / (T-1) &= \sum_{i=2}^T (\Delta y_i - \overline{\Delta y})^2 \\ &= \sum_{i=2}^T \Delta S_i^2 = -2 \sum_{i=2}^T \Delta \bar{S}_i \bar{S}_{i-1} \end{aligned}$$

using Lemma 1. For the denominator, note that

$$S_i - \bar{S} = \{(T-1)y_i - (t-1)y_T - (T-t)y_1 - (T-1)[\bar{y} - (y_1 + y_T)/2]\} / (T-1)$$

using Lemmas 2 and 3 and a little algebra. Then the denominator of R_2 equals

$$\sum_{i=1}^T (S_i - \bar{S})^2$$

by inspection.

From Theorem 1 we see that R_2 is almost equal to $-2\tilde{\phi}$, or equivalently $-2\tilde{\rho}/T$. The only difference is that the denominator of R_2 is

$$\sum_{i=1}^T (S_i - \bar{S})^2 = \sum_{i=1}^T S_i^2 - T^{-1} \left(\sum_{i=1}^T S_i \right)^2$$

whereas the denominator of $\tilde{\phi}$ is

$$\sum_{i=2}^T (S_{i-1} - \bar{S}_{-1})^2 = \sum_{i=1}^T S_i^2 - (T-1)^{-1} \left(\sum_{i=1}^T S_i \right)^2.$$

Finally, we wish to establish equation (17) of the text.

THEOREM 2:

$$\tilde{\tau} = \begin{cases} - \left[\frac{-2T}{(T-3)\tilde{\rho}} - \frac{1}{(T-3)} \right]^{-1/2} & \tilde{\rho} < 0 \\ \left[\frac{2T}{(T-3)\tilde{\rho}} - \frac{1}{(T-3)} \right]^{-1/2} & \tilde{\rho} > 0 \end{cases}$$

PROOF: We will give the proof for $\tilde{\phi} < 0$; when $\tilde{\phi} > 0$, only a few signs need changing. We have

$$\begin{aligned} \tilde{\tau} &= \frac{(T-3)^{1/2} \tilde{\phi}}{\left[\frac{\Sigma \Delta \tilde{S}_t^2}{\Sigma (\tilde{S}_{t-1} - \bar{\tilde{S}}_{-1})^2} - \tilde{\phi}^2 \right]^{1/2}} = \frac{(T-3)^{1/2} \tilde{\phi}}{\left[\frac{-2 \Sigma \Delta \tilde{S}_t \tilde{S}_{t-1}}{\Sigma (\tilde{S}_{t-1} - \bar{\tilde{S}}_{-1})^2} - \tilde{\phi}^2 \right]^{1/2}} \quad \text{using Lemma 1} \\ &= \frac{(T-3)^{1/2} \tilde{\phi}}{(-2\tilde{\phi} - \tilde{\phi}^2)^{1/2}} = - \left[\frac{(T-3)\tilde{\phi}^2}{-2\tilde{\phi} - \tilde{\phi}^2} \right]^{1/2} \quad \text{assuming } \tilde{\phi} < 0 \\ &= - \left[\frac{-2}{(T-3)\tilde{\phi}} - \frac{1}{(T-3)} \right]^{-1/2} = - \left[\frac{-2T}{(T-3)\tilde{\rho}} - \frac{1}{(T-3)} \right]^{-1/2} \end{aligned}$$

APPENDIX 3 ASYMPTOTIC THEORY

(i) Linear Trends

We employ the functional limit theory used in Phillips and Perron (1988) and some of the subsidiary limit results on partial sums given there. Let $W(r)$ be a standard Brownian motion on the $[0, 1]$ interval, $V(r) = W(r) - rW(1)$ be a standard Brownian bridge and define

$$\underline{V}(r) = V(r) - \int_0^1 V(r) dr = W(r) + \left(\frac{1}{2} - r \right) W(1) - \int_0^1 W(r) dr.$$

Observe that in $L_2[0, 1]$, $\underline{V}(r)$ is the projection of $V(r)$ on the orthogonal complement of the constant function. Thus $\underline{V}(r)$ is simply a demeaned Brownian bridge.

From (18) we have

$$\tilde{S}_t = \sum_{s=2}^t (\varepsilon_s - \bar{\varepsilon})$$

so that

$$T^{-1/2} \hat{S}_{[Tr]} = T^{-1/2} S_{[Tr]} - ([Tr]/T) T^{-1/2} S_T \rightarrow \sigma(W(r) - rW(1)) = \sigma V(r)$$

With this in hand it is easy to see that

$$T^{-2} \sum_1^T (\hat{S}_{t-1} - \bar{\hat{S}})^2 \rightarrow \sigma^2 \int_0^1 V(r)^2 dr. \quad (\text{A3.1})$$

Further

$$\begin{aligned} T^{-1} \sum_1^T (\hat{S}_{t-1} - \bar{\hat{S}}) \varepsilon_t &= T^{-1} \sum_1^T (\hat{S}_{t-1} - \bar{\hat{S}}) (\varepsilon_t - \bar{\varepsilon}) \\ &= (1/2) \left\{ T^{-1} \left(\sum_1^T (\varepsilon_t - \bar{\varepsilon}) \right)^2 - T^{-1} \sum_1^T (\varepsilon_t - \bar{\varepsilon})^2 \right\} \rightarrow_p (-1/2) \sigma_\varepsilon^2. \end{aligned} \quad (\text{A3.2})$$

Now

$$\hat{\rho} = \left\{ T^{-2} \sum_1^T (\hat{S}_{t-1} - \bar{\hat{S}})^2 \right\}^{-1} \left\{ T^{-1} \sum_1^T (\hat{S}_{t-1} - \bar{\hat{S}}) \varepsilon_t \right\} \rightarrow_p (-1/2) \sigma_\varepsilon^2 / \sigma^2 \int_0^1 V(r)^2 dr \quad (\text{A3.3})$$

by joint convergence of the numerator and denominator.

Next observe that

$$\tilde{\tau} = \left\{ T^{-2} \sum_1^T (\hat{S}_{t-1} - \bar{\hat{S}})^2 \right\}^{1/2} \hat{\rho} / s \quad (\text{A3.4})$$

where s is the estimated standard error of the regression (10). Since $s \rightarrow_p \sigma_\varepsilon$ we obtain from (A3.1), (A3.3) and (A3.4) the following limit for $\tilde{\tau}$:

$$\tilde{\tau} \rightarrow (-1/2)(\sigma_\varepsilon/\sigma) \left(\int_0^1 V(r)^2 dr \right)^{-1/2}$$

(ii) Higher Order Trends

As in the linear trend case, the asymptotics are determined by the behaviour of the partial sum process \hat{S}_t in (37). Note that from the regression

$$\Delta y_t = \sum_{j=0}^{p-1} \tilde{b}_j t^j + \tilde{u}_t \quad (\text{A3.5})$$

We have, under the null hypothesis,

$$\tilde{u}_t = \varepsilon_t - \sum_{j=0}^{p-1} (\tilde{b}_j - b_j) t^j. \quad (\text{A3.6})$$

Let X be the trend regressor matrix in (A3.5) and define

$$D_T = \text{diag}(T, T^3, \dots, T^{2p-3}).$$

Then, in conventional regression notation, we have.

$$D_T^{1/2}(\tilde{b} - b) = (D_T^{-1/2} X' X D_T^{-1/2})^{-1} (D_T^{1/2} X' \varepsilon) \rightarrow \sigma Q^{-1} \int_0^1 g(r) dW(r) \quad (\text{A3.7})$$

where $Q = (q_{ij})$ with $q_{ij} = 1/(i+j-1)$ and $g(r)' = (1, r, \dots, r^{p-1})$. The weak convergence to (A3.7) follows in a straightforward way using the methods in Phillips (1987). Next, we observe that

$$T^{-j-1} \sum_{i=1}^{\lfloor Tr \rfloor} t^j \sim [Tr]^{j+1} / T^{j+1} (j+1) \rightarrow r^{j+1} / (j+1).$$

Hence

$$\begin{aligned} T^{-1/2} \sum_{i=1}^{\lfloor Tr \rfloor} \tilde{u}_t &= T^{-1/2} S_{\lfloor Tr \rfloor} - \sum_{j=0}^{p-1} (\tilde{b}_j - b_j) \left(T^{-1/2} \sum_{i=1}^{\lfloor Tr \rfloor} t^j \right) \\ &= T^{-1/2} S_{\lfloor Tr \rfloor} - (D_T^{1/2}(\tilde{b} - b))' D_T^{-1/2} \left(T^{-1/2} \sum_{i=1}^{\lfloor Tr \rfloor} t^j \right) \rightarrow \sigma W(r) \\ &\quad - \sigma \left(\int_0^1 dW(s) g(s)' \right) Q^{-1} q(r) \\ &= \sigma \left\{ W(r) - \left(\int_0^1 dW(s) g(s)' \right) Q^{-1} q(r) \right\} = \sigma V_p(r), \end{aligned} \quad (\text{A3.8})$$

where $q(r)$ is $p \times 1$ with j th element r^j/j . We shall call the Gaussian process $V_p(r)$ a p -level Brownian bridge. Like a conventional Brownian bridge, this process is tied down on the $[0, 1]$ interval since $V_p(0) = 0$ and

$$\begin{aligned} V_p(1) &= W(1) - \int_0^1 dW(s) g(s)' Q^{-1} q(1) \\ &= W(1) - \int_0^1 dW(s) g(s)' e_1 \end{aligned}$$

$$\begin{aligned}
&= W(1) - \int_0^1 dW(s) \\
&= 0
\end{aligned}$$

where we use the fact that $Q^{-1}q(1) = e_1$, the first unit vector.

From (38) in conventional regression notation we have

$$\tilde{\phi} = (\tilde{S}' Q_X \tilde{S})^{-1} (\tilde{S}' Q_X \Delta y), \quad (\text{A3.9})$$

where Q_X denotes the usual projection matrix on the orthogonal complement of the range of X . Now

$$T^{-2} \tilde{S}' Q_X \tilde{S} \rightarrow \sigma^2 \int_0^1 \underline{V}_p(r)^2 dr \quad (\text{A3.10})$$

where

$$\underline{V}_p(r) = V_p(r) - \sum_{j=0}^{p-1} \hat{\alpha}_j r^j$$

and

$$\hat{\alpha} = \operatorname{argmin}_{\alpha} \int_0^1 \left(V_p(r) - \sum_{j=0}^{p-1} \alpha_j r^j \right)^2 dr.$$

The process $\underline{V}_p(r)$ is a detrended p -level Brownian bridge and is the asymptotic equivalent of the regression projection $Q_X \tilde{S}$.

We also find that

$$T^{-1} \tilde{S}' Q_X \Delta y = T^{-1} \tilde{S}' Q_X \varepsilon = T^{-1} \tilde{S}' \underline{\varepsilon}$$

where $\underline{\varepsilon} = Q_X \varepsilon$. But from (A3.6) we have

$$\tilde{S}_t = \sum_1^t \left\{ \varepsilon_s - \sum_0^{p-1} (\tilde{b}_j - b_j) s^j \right\} = \sum_1^t \underline{\varepsilon}_s.$$

Thus

$$\begin{aligned}
T^{-1} \tilde{S}' \underline{\varepsilon} &= T^{-1} \sum_1^t \tilde{S}_{t-1} \underline{\varepsilon}_t \\
&= (2T)^{-1} \left\{ \left(\sum_1^t \underline{\varepsilon}_t \right)^2 - \sum_1^t \underline{\varepsilon}_t^2 \right\} \rightarrow -(1/2) \sigma_{\varepsilon}^2. \quad (\text{A3.11})
\end{aligned}$$

From (A3.9) to (A3.11) we deduce that

$$\tilde{\rho} = T\tilde{\phi} \rightarrow -(1/2)\sigma_\varepsilon^2 / \left(\sigma^2 \int_0^1 \underline{V}_\rho^2 \right) = - \left(2 \int_0^1 \underline{V}_\rho^2 \right)^{-1} \omega^2,$$

and in a similar fashion the limit of the ι ratio $\tilde{\tau}$ in (38) is found to be

$$\tilde{\tau} = -(1/2)\sigma_\varepsilon^2 / \sigma_\varepsilon \left(\sigma^2 \int_0^1 \underline{V}_\rho^2 \right)^{1/2} = -(1/2) \left(\int_0^1 \underline{V}_\rho^2 \right)^{-1/2} \omega.$$