The Annals of Statistics
1995, Vol. 23, No. 3, 1015–1028

TESTING A TIME SERIES FOR DIFFERENCE STATIONARITY

BY B. P. M. MCCABE AND A. R. TREMAYNE

University of British Columbia and University of Sydney

This paper addresses the problem of testing the hypothesis that an observed series is difference stationary. The alternative hypothesis is that the series is another nonstationary process; in particular, an autoregressive model with a random parameter is used. A locally best invariant test is developed assuming Gaussianity, and a representation of its asymptotic distribution as a mixture of Brownian motions is found. The performance of the test in finite samples is investigated by simulation. An example is given where the difference stationary assumption for a well-known data series is rejected.

1. Introduction. The work of Box and Jenkins (1976) has been instrumental in the popularity of the autorgressive-moving average (ARMA) class of statistical models for stationary time series. The ARMA class of models can also deal with data exhibiting homogeneous nonstationarity, since such data can be reduced to stationarity by differencing. Hence, a cornerstone of practical time series modeling is the acceptability of the difference stationary assumption, that is, that a series can be modeled using the ARIMA (stationary ARMA after differencing) class.

In standard significance tests of the null hypothesis of difference stationarity [see, e.g., Dickey (1976), Fuller (1976) and Phillips (1987)], the null of a unit root is typically tested against the alternative of stationarity. However, there is a strong need to be able to distinguish the difference stationary class from other types of nonstationary model. A failure of the Dickey–Fuller test to reject the null hypothesis of difference stationarity could be due to the low power of such a test against the actual (nonstationary) data-generating process rather than the acceptability of the presence of a unit root. Since modeling strategies such as the cointegration methods popularized by Engle and Granger (1987) rely crucially on the presence of unit roots, the sole use of standard unit root tests against stationary alternatives may result in poor approximations to the underlying data-generating processes.

In recent years, in parallel with the development of methods of nonlinear time series analysis [Tong (1990) and Granger and Teräsvirta (1993)], it has become apparent that many series are not difference stationary and may be nonstationary in a nonhomogeneous way. In this paper we provide a locally best invariant test of the null of difference stationarity (a unit root) against the alternative of a randomized unit root. If the null hypothesis is accepted,

Received June 1992; revised August 1994.
AMS 1991 *subject classifications*. Primary 62M10; secondary 62F03, 62F05.
Key words and phrases. Autoregression, Brownian motion, difference stationarity, locally best invariant, random coefficient, weak convergence.

then the analyst may proceed with an ARIMA analysis, while if it is rejected a nonstationary, and perhaps nonlinear, modeling methodology should be entertained.

In this paper we propose a test for coefficient constancy which is locally best invariant on Gaussian assumptions. A representation of the asymptotic distribution of the test statistic as a mixture of Brownian motions, with the normality assumption previously used being replaced by a suitable mixing condition is found. The test is now no longer locally best invariant, but the statistic does provide a valid significance test of wide applicability. Following this, we report the results of simulation experiments which provide critical values for use with the test. Some limited evidence on its empirical power properties is also given. The last main section applies the test to the IBM data set provided as Series B by Box, Jenkins and Reinsel (1994).

2. The model and some simple properties. In the stationary case, a suitable justification for use of the ARMA class of models is provided by the famous Wold decomposition. Cramér's (1961) extension to Wold's theorem [see, e.g., Priestley (1981), page 858] implies that any nonstationary stochastic process may be written as an ARMA process with coefficients that are allowed to vary with time. The special case of random coefficient autoregressive (RCA) models has been well examined in the literature. Nicholls and Quinn (1982) study the stationary case, and Tyssedal and Tjøstheim (1988) provide practical application of this class of model; these papers contain many other references. Also, Tong (1990) devotes a number of passages to RCA models and he points out that other important classes of nonlinear models like autoregressive conditional heteroscedastic models and bilinear formulations can be regarded as special cases.

Consider an autoregressive representation with time varying coefficients which are modeled as realizations of a sequence of random variables. In the first-order autoregressive case, this leads to a model of the form

$$y_t = \alpha_t y_{t-1} + \varepsilon_t,$$

where ε_t is a suitable innovation and α_t a sequence of random variables (which may be correlated with ε_t). This is, of course, not the only way to model time varying coefficients, but it is mathematically convenient. Further, should the ARIMA specification be rejected, the search for an alternative model may be conducted within the RCA class.

Suppose that $y_0, y_1, \dots, y_T$ are observable random variables generated, condition on y_0, according to

$$w_t = \alpha_t w_{t-1} + \varepsilon_t, \qquad t = 1, \dots, T, \tag{2.1}$$

where $w_t = y_t - y_0$ and $w_0 = 0$. The random coefficients α_t are assumed independent of one another with a constant mean of unity and variance ω^2, and the innovations ε_t are zero-mean independent random variables with variance σ^2. The processes ε_t and α_t are assumed mutually dependent with covariance $\psi\omega^2$ at time t and independent otherwise; it is further assumed that the two processes are not contemporaneously linearly dependent, so that $|\psi| < \sigma^2$.

The mean of w_t, conditional on $w_{t-1}, w_{t-2}, \dots$, is w_{t-1} and the conditional variance of w_t is $\omega^2 w_{t-1}^2 + 2\psi\omega^2 w_{t-1} + \sigma^2$. When ε_t and α_t are Gaussian, the joint density of the w_t is

$$k \cdot \prod_{t=2}^{T} \left(\omega^2 w_{t-1}^2 + 2\psi\omega^2 w_{t-1} + \sigma^2\right)^{-1/2}$$
$$\times \exp\left[-\tfrac{1}{2}\sum_{t=2}^{T} (w_t - w_{t-1})^2 \left(\omega^2 w_{t-1}^2 + 2\psi\omega^2 w_{t-1} + \sigma^2\right)^{-1}\right]$$
$$\times \sigma^{-1} \exp\left(-\tfrac{1}{2} w_1^2/\sigma^2\right),$$

by the usual conditioning argument. In what follows, most products and summations are from 2 to T and so the limits on such symbols are hereafter omitted unless they differ from this.

3. The derivation of the locally best invariant test. The testing problem of interest, that is, testing $\omega^2 = 0$, is invariant to scale transformations. Accordingly a maximal invariant, distributed independently of σ^2, is given by

$$z_t = w_t / w_1, \qquad t = 2, \dots, T.$$

For convenience z_1 and σ^2 are both set to unity. To find the joint distribution of $z_2, \dots, z_T$, define a dummy variable $D = w_1$. The Jacobian of the transformation is $|D|^{T-1}$, and the joint density of the z_t and D is

$$\begin{aligned} &G(z_2, \dots, z_T, D) \\ &\quad = k \cdot |D|^{T-1} \prod \left(\omega^2 z_{t-1}^2 D^2 + 2\psi\omega^2 z_{t-1} D + 1\right)^{-1/2} \\ &\qquad \times \exp\left\{-\tfrac{1}{2}\sum (z_t - z_{t-1})^2 D^2 \left(\omega^2 z_{t-1}^2 D^2 + 2\psi\omega^2 z_{t-1} D + 1\right)^{-1}\right\} \\ &\qquad \times \exp\left\{-\tfrac{1}{2} D^2\right\} = G(\omega^2, D), \end{aligned} \tag{3.1}$$

introducing a compact notation with the arguments which are now of concern.

Ideally, one would now wish to find the distribution of the maximal invariant by integrating out D and then differentiating the log of the resultant density with respect to ω^2 to find the locally best invariant (LBI) test [see Ferguson (1967), page 235]. Unfortunately, it does not seem possible to integrate (3.1) with respect to D directly.

An alternative approach is to interchange the order of the integration and differentiation operations. Of course this requires that both the original and resulting integrals be uniformly convergent. The following lemma shows that this is legitimate.

LEMMA 1.

$$\partial \int G(\omega^2, D)\, dD = \int \partial G(\omega^2, D)\, dD,$$

where $\partial = d/d\omega^2$ *and* $\omega^2 \le M$, *M being some finite number.*

The proof of this lemma is given in the Appendix. To derive the LBI test, it remains to differentiate the joint density with respect to ω^2, set it to zero and integrate out the dummy variable D. Finally, the resultant expression should be divided by the null density to provide the appropriate test. This leads to the following theorem.

THEOREM 1. *In model* (2.1), *when* $\{\alpha_t\}$ *and* $\{\varepsilon_t\}$ *are (possibly contemporaneously correlated) sequences of iid normal random variables and* $T\ (= 2n)$ *is even, the LBI test of* $\omega^2 = 0$ *against* $\omega^2 > 0$ *is given by* $Z^* = Z + Z_1$, *where*

$$Z = (T+2)T\varphi^{-4} \sum (w_t - w_{t-1})^2 w_{t-1}^2 - T\varphi^{-2} \sum w_{t-1}^2,$$

$$Z_1 = \psi f(T)\sqrt{2\pi}\left[\sum (w_t - w_{t-1})^2 w_{t-1}\varphi^{-3} g(T+2) - \sum w_{t-1}\varphi^{-1} g(T)\right].$$

In these expressions $f(2n) = [2^{n-1}(n-1)!]^{-1}$, $g(2n) = (2r)!/(2^r r!)$ *and* $\varphi^2 = \Sigma(w_t - w_{t-1})^2 + w_1^2$. *The test rejects for large values of* Z^*.

PROOF. Differentiating (3.1) with respect to ω^2 and equating to zero gives

$$-\frac{k}{2}\exp\left\{-\tfrac{1}{2}D^2\left[\sum (z_t - z_{t-1})^2 + 1\right]\right\}|D|^{T-1}\left(D^2 \sum z_{t-1}^2 + 2\psi D \sum z_{t-1}\right)$$
$$+\frac{k}{2}\exp\left\{-\tfrac{1}{2}D^2\left[\sum (z_t - z_{t-1})^2 + 1\right]\right\}|D|^{T-1}\left\{\sum (z_t - z_{t-1})^2 D^2 \right.$$
$$\left.\times\left(z_{t-1}^2 D^2 + 2\psi z_{t-1} D\right)\right\}.$$

Define $\phi^2 = [\Sigma(z_t - z_{t-1})^2 + 1]^{-1}$ and set $T = 2n$. To integrate with respect to D, we require certain moments of a normal random variable X having mean zero and variance σ^2. The odd-order moments are zero and the even ones are well known. The absolute odd-order moments are

$$E\left[|X|^{2n+1}\right] = \sigma^{2n+1} 2^n n! (2/\pi)^{1/2}.$$

On performing the integrations we get

$$\frac{k}{2}\left[\sum (z_t - z_{t-1})^2 z_{t-1}^2 \phi^{2n+4} 2^{n+2}(n+1)! + 2\psi \sum (z_t - z_{t-1})^2 \right.$$
$$\left. \times z_{t-1}\phi^{2n+3} g(2n+2)\sqrt{2\pi}\right]$$
$$-\frac{k}{2}\left[\sum z_{t-1}^2 \phi^{2n+2} 2^{n+1} n! + 2\psi \sum z_{t-1}\phi^{2n+1} g(2n)\sqrt{2\pi}\right].$$

Dividing by the null density $k\phi^{2n}2^n(n-1)!$ yields the LBI test as

$$\tfrac{1}{2}\Big[(T+2)T\phi^4\sum(z_t-z_{t-1})^2z_{t-1}^2+\psi\phi^3 f(T)g(T+2)\times\sqrt{2\pi}\sum(z_t-z_{t-1})z_{t-1}^2\Big]-\tfrac{1}{2}\Big[T\phi^2\sum z_{t-1}^2+\psi\phi f(T)g(T)\sqrt{2\pi}\sum z_{t-1}\Big].$$

Substituting $z_t = w_t/w_1$ and rearranging completes the proof. □

When T is odd a different statistic arises. The proof is similar to Theorem 1 and is thus omitted, but it is presented in McCabe and Tremayne (1989) as their Theorem 2. The LBI test in this case is just Z, the term Z_1 no longer being required. The statistic Z^* differs from Z in the important respect that, through Z_1, it does depend on ψ, which reflects the strength of the covariance between α_t and ε_t, although we shall see later that this does not affect the asymptotic behavior of the statistic. Thus Z may be used for all T, at any rate asymptotically.

The factorial expressions inherent in Z^* given in Theorem 1 can be simplified by employing Stirling's approximation and replacing $m!$ by $\sqrt{2\pi m}\, m^m e^{-m}$. By writing $f(2n)g(2n+2)$ as

$$n(n+1)(2n+2)!/\big\{2^{2n}[(n+1)!]^2\big\}$$

and so on, we can reexpress Z^* as

$$Z^* = Z + 2\psi \times\Big[T(T+2)^{1/2}\sum(w_t-w_{t-1})^2 w_{t-1}\varphi^{-3} - T^{1/2}\sum w_{t-1}\varphi^{-1}\Big]. \tag{3.2}$$

A neater expression for the Z^* statistic can be obtained by ignoring asymptotically negligible terms in (3.2) and reformulating it as

$$\tilde{\sigma}^{-4}\sum w_{t-1}^2\big\{(w_t-w_{t-1})^2-\tilde{\sigma}^2\big\} + 2\psi\tilde{\sigma}^{-3}\sum w_{t-1}\big[(w_t-w_{t-1})^2-\tilde{\sigma}^2\big], \tag{3.3}$$

where $\tilde{\sigma}^2 = T^{-1}\Sigma(w_t - w_{t-1})^2$. Since $\varepsilon_t = w_t - w_{t-1}$, the statistic is essentially a weighted sum of the difference between ε_t^2 and its overall average, $\tilde{\sigma}^2$. Corollary 1 below shows that the second term in (3.3) is asymptotically negligible, relative to the first, so the limiting behavior of Z^* can be established by finding that of Z.

4. The asymptotic distribution of Z. The exact distribution of Z, even under Gaussianity, does not seem to be tractable and so recourse to asymptotic theory must be made. However, the assumption of Gaussian white noise innovations in (2.1) can be relaxed considerably without affecting the asymptotic behavior of the statistic. In the current context a stationary and mixing assumption will be satisfactory.

ASSUMPTION 1. The sequence η_t, say, is stationary and satisfies the following:

(i) $E(\eta_t) = 0$;
(ii) $E|\eta_t|^p]^{1/p} < \infty$ for some $p > \beta > 2$; and
(iii) $\{\eta_t\}$ is α-mixing with coefficients α_m of size $-p\beta/(p-\beta)$.

As a consequence of stationarity $T^{-1}\operatorname{var}(\Sigma_1^T \eta_t) \to \sigma_\eta^2$ as $T \to \infty$, $0 < \sigma_\eta^2 < \infty$. Define a partial sum process (PSP) for any $\{\eta_t\}$ as a function on $D[0,1]$ by

$$\mathbb{W}_T(r) = T^{-1/2}\sigma_\eta^{-1} \sum_{1}^{\langle Tr \rangle} \eta_t$$

where $\langle\ \rangle$ means integer part of. Assumption 1 implies that $\mathbb{W}_T(r)$ converges weakly (written $\Rightarrow$) to $\mathbb{W}(r)$, a Brownian motion on $C[0,1]$. This is also true when η_t is a vector when σ_η^2 is interpreted as a covariance matrix with σ_η^{-1} satisfying $\sigma_\eta^{-1}\sigma_\eta^{-1\prime} = \sigma_\eta^{-2}$ and the limiting process is a vector of independent Brownian motions. Theorem 2, which is proved in the Appendix, gives the limiting distribution of Z under Assumption 1. It makes extensive use of the results of Hansen (1992).

THEOREM 2. *Define* $\sigma_\varepsilon^2 = E(\varepsilon_t^2)$, $\tilde{\sigma}^2 = \Sigma_1^T \varepsilon^2/T$ *and* $w_t = w_{t-1} + \varepsilon_t$. *Let* $\eta_t = \{\varepsilon_t, \varepsilon_t^2 - \sigma_\varepsilon^2\}$, *and let* σ^2, κ^2 *and* $\rho\sigma\kappa$ *be the distinct elements of the covariance matrix* σ_η^2. *Then, under Assumption* 1 *for* η_t,

$$Z_T = T^{-3/2}\sigma^{-2}\kappa^{-1} \sum w_{t-1}^2(\varepsilon_t^2 - \tilde{\sigma}^2) \Rightarrow \int_0^1 \left[\mathbb{W}_1(t)^2 - \int_0^1 \mathbb{W}_1(s)^2\,ds\right] d\mathbb{W}^*(t),$$

where $\mathbb{W}_1(r)$ *is a Brownian motion*, $\mathbb{W}^*(r) = \rho\mathbb{W}_1 + \sqrt{(1-\rho^2)}\,\mathbb{W}_2$ *and* $\mathbb{W}_2$ *is a Brownian motion independent of* $\mathbb{W}_1$.

The parameter ρ measures the limiting correlation between the two components of the stationary η_t and the form of the limiting distribution, involving ρ as it does, indicates that the statistic is operational when this correlation is known. If ε_t is iid and the distribution of ε_t is symmetric, $\rho = 0$, for then $\rho = E(\varepsilon_t^3)/\sigma\kappa$. For more general stationary ε_t, the symmetry of the distribution of the innovations in its Wold decomposition is sufficient to guarantee that ε_t has a symmetric distribution and that $\rho = 0$ because $E(\varepsilon_t\varepsilon_s^2) = 0$, $t \neq s$. This we shall assume in practical applications.

The statistic Z_T of Theorem 2 is a scaled version of the statistic Z arising when T is odd; see Theorem 1, (3.2) and (3.3). For T even we have the following corollary, which is also proved in the Appendix.

COROLLARY 1. *The quantity* $2\psi T^{-3/2}\sigma^{-1}\kappa^{-1}[\Sigma w_{t-1}(\varepsilon_t^2 - \tilde{\sigma}^2)]$ *converges to zero in probability.*

The effect of this corollary is that the statistics for T odd and even have the same asymptotic distribution, so that Z_T may be used for both.

Often it is not reasonable to assume that the innovation sequence $\{\varepsilon_t\}$ is white noise, and allowance must be made for its stationary dependence properties. If, in the absence of further information, $\{\varepsilon_t\}$ is simply assumed to be stationary mixing, we may use the result that σ_η^2, defined in conjunction with Assumption 1, may be written in terms of the spectral density of the process, that is, $\sigma_\eta^2 = 2\pi f(0)$, where $f(\cdot)$ is the spectral density of $\{\eta_t\}$. However, $\hat{f}$, some consistent estimator of f, may converge only slowly to f [see, e.g., Box, Jenkins and Reinsel (1994), page 40], so that this may not be a feasible strategy in small samples.

Another approach which could be used, and which is easy to implement, is to assume that the correlation structure in $\{\varepsilon_t\}$ can be modeled as an ARMA process. In the spirit of the augmented Dickey–Fuller test [Dickey and Fuller (1981)], we may assume that this ARMA process may be satisfactorily approximated by an autoregressive process of sufficiently high order. Then the test statistic is constructed from the residuals $\hat{\varepsilon}_t$ in the regression of Δy_t on $\Delta y_{t-1}, \dots, \Delta y_{t-p}$ for suitably chosen p. More formally we have the following corollary.

COROLLARY 2. *If ε_t is driven by an autoregressive process of order p^*, and Z_T of Theorem 2 is computed from the residuals $\hat{\varepsilon}_t$ in the regression of Δy_t on $\Delta y_{t-1}, \dots, \Delta y_{t-p}$, $p \geq p^*$, the limiting distribution of the test statistic is unaffected. This is because the PSP's of $\{\hat{\varepsilon}_t\}$ and $\{\hat{\varepsilon}_t^2\}$ converge weakly to the same limits that one obtains when the true ε's are used.*

The proof of this is straightforward and follows along the lines of Leybourne and McCabe [(1989), Section 4]. Of course, if p^* of Corollary 2 is equal to zero, so that the model innovations have no correlation structure, the statistic Z_T of Theorem 2 may be constructed using the following corollary.

COROLLARY 3. *Under the assumption that ε_t is a zero-mean, independently and identically distributed sequence with finite fourth moments, Theorem 2 continues to hold when σ^2 is replaced by $\tilde{\sigma}^2$ and κ^2 by $\tilde{\kappa}^2 = \sum(\varepsilon_t^2 - \tilde{\sigma}^2)^2/T$.*

This follows from the weak law of large numbers and the continuous mapping theorem. In Section 6 both Corollary 2 and Corollary 3 are used.

5. Tables and power of the Z-test. The distribution of Z_T is untabulated, so we conducted some simulation experiments at various sample sizes (with $\rho = 0$) to provide critical values. We used the model

$$y_t = \alpha_t y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{NID}(0,1),\ \alpha_t \sim \mathrm{NID}(1, \omega^2). \tag{5.1}$$

In addition, ε_t and α_t were independent. Results under the null $\omega^2 = 0$, are given in Table 1. It provides empirical percentiles from 90 to 99 in steps of 1% for the statistic. This corresponds to a range of significance levels from 10% to 1% and is tabulated because the statistic rejects for values in the

TABLE 1
Significance points for Z_T when $\rho = 0$

Significance level (%)	Sample size (T)				
	50	100	200	500	1000
	Z_T	Z_T	Z_T	Z_T	Z_T
10	0.47	0.47	0.48	0.51	0.49
9	0.51	0.51	0.52	0.55	0.54
8	0.56	0.56	0.57	0.61	0.59
7	0.61	0.62	0.63	0.67	0.65
6	0.69	0.70	0.71	0.75	0.72
5	0.77	0.79	0.80	0.84	0.81
4	0.87	0.91	0.91	0.96	0.92
3	1.04	1.06	1.07	1.11	1.07
2	1.27	1.28	1.32	1.37	1.29
1	1.66	1.72	1.74	1.90	1.70

Notes: (i) This table is based upon using pseudorandom Gaussian random variables. Entries are based on 100,000 replications. (ii) The null model is $y_t = y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \text{NID}(0, 1)$.

right-hand tail. Entries are based on 100,000 replications for each T. Use of the (weak convergence) invariance principle implies that the critical values in Table 1 apply for all symmetric innovation sequences that satisfy Assumption 1 with $\rho = 0$.

It is also of interest to investigate whether the test is able to discern correctly that random coefficient variation is present in small data sets that are often encountered in practice. Table 2 presents the results of limited simulations carried out to assess the empirical power of the test using (5.1). The rejection frequencies given demonstrate that the test has quite good power properties, even for fairly small values of ω^2 and moderate sample sizes like 50 or 100. As is to be expected, rejection frequencies increase with sample size.

TABLE 2
Empirical power of Z

ω^2	Sample size (T)				
	50	100	200	500	1000
	Z_T	Z_T	Z_T	Z_T	Z_T
0.001	0.07	0.11	0.21	0.42	0.62
0.01	0.20	0.35	0.51	0.71	0.85
0.1	0.36	0.43	0.49	0.50	0.56

Notes: (*i*) *This table is based on pseudorandom Gaussian random variables. Entries given are the proportion of rejections in* 10,000 *replications at the* 5% *level.* (*ii*) *The model is* $y_t = \alpha_t y_{t-1} +_\tau$, $\varepsilon_t \sim \text{NID}(0, 1)$, $\alpha_t \sim \text{NID}(1, \omega^2)$. (*iii*) *The null hypothesis tested is* $\omega^2 = 0$.

6. An example using the IBM stock price data set. We now apply the test developed earlier in the paper to a famous data series, to indicate its potential efficacy. The data is the set of 369 daily observations on IBM common stock closing prices given by Box, Jenkins and Reinsel (1994) as their Series B. Some controversy has surrounded the analysis of this series. For example, the authors themselves (Section 8.2.3), after preliminary analysis, divide the data into two, arguing that the data-generating mechanism may have changed about midway through. They then found a significant first-order moving average component for the first half of the data, but a much smaller, insignificant, coefficient for the second half. These findings could be construed as indicating that the data in the first half are ARIMA(0, 1, 1), while those in the second half are ARIMA(0, 1, 0). The Z procedure is now applied to test the adequacy of these two models.

Testing the hypothesis of difference stationarity for the first half of the data (184 observations) requires allowance for the correlation in the first differences because of the significant MA component found. Since the data set is not large, we employ the suggestion of Corollary 2 and approximate the correlation structure by an AR(p). In this example, setting $p = 4$ results in an insignificant Q-statistic based on the residuals, and we thus deemed the correlation due to the MA component to have been effectively eliminated. The value of Z based on the residuals from the ARIMA(4, 1, 0) fitted to the data was -0.88. This is not significant at any conventional significance level and thus the difference stationarity may be maintained.

For the second half of the data the proposed model is ARIMA(0, 1, 0). Thus there is no requirement to eliminate correlation via the augmenting procedure used above. The statistic Z, calculated in conjunction with Corollary 3, yields a value of $Z = 2.50$. This is significant at all conventional significance levels. Thus there is ample indication that the ARIMA(0, 1, 0) hypothesis for the second half of the data can be decisively rejected and that the data is not difference stationary.

Further applications of the theory developed in this paper may be found in Leybourne, McCabe and Tremayne (1994), where allowance is made for the presence of trends which are a feature of many macroeconomic data sets. The authors provide critical values with allowance for trends, although these differ from those given in the present paper because of the extra nuisance parameters.

7. Conclusion. The general conclusion to be reached from the theoretical developments and the empirical example of this paper is that it may be useful for applied workers to entertain alternatives to unit root hypotheses other than that of a stationary autoregression. This is particularly the case in view of the widespread use of ARIMA modeling.

The paper develops an optimal test of a difference stationary null hypothesis against another nonstationary alternative process characterized by random coefficient variation. The asymptotic distribution of the proposed statistic is derived under mixing and is tabulated. Sections 5 and 6 provide useful illustrations of the fact that not only does the test possess power in simulations, but also that it may be of value in practical applications.

APPENDIX

PROOF OF LEMMA 1. Consider the integral on the left-hand side in he lemma. Now

$$\int_B^\infty G(\omega^2, D)\, dD \le k^* \int_B^\infty |D|^{T-1} \exp\{-D^2/2)\, dD, \tag{A.1}$$

for a number k^* independent of ω^2. This follows since the product and exponent terms in $G(\cdot)$ may be bounded using (A.2a) and (A.2b):

$$\begin{aligned}(A.2a)\quad &\left(\omega^2 z_{t-1}^2 D^2 + 2\psi\omega^2 z_{t-1} D + 1\right)\\ &\qquad = \left(\omega z_{t-1} D + \omega\psi\right)^2 + \left(1 - \omega^2\psi^2\right)\\ &\qquad \ge 1 - \omega^2\psi^2 \ge 1 - M\psi^2 > 0;\end{aligned}$$

$$(A.2b)\quad \exp\left\{-\tfrac{1}{2}\sum (z_t - z_{t-1})^2 D^2 \left(\omega^2 z_{t-1}^2 D^2 + 2\psi\omega^2 z_{t-1} D + 1\right)^{-1}\right\} \le 1.$$

In (A.2a), noting the assumption that α_t and ε_t are not perfectly correlated and remembering σ^2 is unity ensures that $\omega^2\psi^2 < 1$. The integral (over the whole real line) on the right-hand side of (A.1) is an absolute moment of a Gaussian distribution. Hence it follows that the tail areas may be made arbitrarily small and so $\int G(\omega^2, D)\, dD$ is uniformly convergent.

Turning to the integral on the right-hand side of the lemma and differentiating $G(\cdot)$ with respect to ω^2 gives

$$\begin{aligned}&-\frac{k^*}{2}|D|^{T-1} \prod_2^T \left(\omega^2 z_{t-1}^2 D^2 + 2\psi\omega^2 z_{t-1} D + 1\right)^{-1/2}\\ &\quad\times \sum \left(z_{t-1}^2 D^2 + 2\psi z_{t-1} D\right)\left(\omega^2 z_{t-1}^2 D^2 + 2\psi z_{t-1} D + 1\right)^{-1}\\ &\quad\times \exp\left\{-\tfrac{1}{2}\sum (z_t - z_{t-1})^2 D^2 \left(\omega^2 z_{t-1}^2 D^2 + 2\psi\omega^2 z_{t-1} D + 1\right)^{-1}\right\}\\ &\quad\times \exp\left\{-\tfrac{1}{2} D^2\right\}\end{aligned}$$

plus

$$\begin{aligned}&\frac{k^*}{2} \cdot |D|^{T-1} \prod_2^T \left(\omega^2 z_{t-1}^2 D^2 + 2\psi\omega^2 z_{t-1} D + 1\right)^{-1/2}\\ &\quad\times \exp\left\{-\tfrac{1}{2}\sum (z_t - z_{t-1})^2 D^2 \left(\omega^2 z_{t-1}^2 D^2 + 2\psi\omega^2 z_{t-1} D + 1\right)^{-1}\right\}\\ &\quad\times \exp\left\{-\tfrac{1}{2} D^2\right\}\left\{\sum (z_t - z_{t-1})^2 D^2 \left(\omega^2 z_{t-1}^2 D^2 + 2\psi\omega^2 z_{t-1} D + 1\right)^{-2}\right.\\ &\qquad\qquad\qquad \left.\times \left(z_{t-1}^2 D^2 + 2\psi z_{t-1} D\right)\right\}\end{aligned}$$

Using (A2), the first term of this integral is bounded, independent of ω^2. The integral of the second term in this derivative expression with respect to D may be shown to be less than or equal to the integral of

$$\frac{k^*}{2}|D|^{T-1}\exp\left\{-\frac{D^2}{2}\right\}\left\{\sum(z_t - z_{t-1})^2\left(\omega^2 z_{t-1}^2 D^2 + 2\psi\omega^2 z_{t-1}D + 1\right)^{-2} \times\left(z_{t-1}^2 D^4 + 2\psi z_{t-1}D^3\right)\right\}$$

$$\leq \frac{k^*}{2}|D|^{T-1}\exp\left\{-\frac{D^2}{2}\right\}\left\{\sum(z_t - z_{t-1})^2 \times(1 - M\psi^2)^{-2}\left(z_{t-1}^2 D^4 + 2\psi z_{t-1}D^3\right)\right\},$$

again using (A.2a). However, collecting powers of D, the integral of the last expression is just the sum of scaled, absolute moments of a normal distribution, independent of ω^2. Thus the tail areas can be made arbitrarily small independent of ω^2, and the lemma is established. □

LEMMA A1 (Joint convergence). *Under Assumption* 1, *let* $\eta_t = \{\varepsilon_t, \varepsilon_t^2 - \sigma_\varepsilon^2\}$, *and let* σ^2, κ^2 *and* $\rho\sigma\kappa$ *be elements of the covariance matrix* σ_η^2 *defined in conjunction with the assumption. Then the PSP's of* $\{\varepsilon_t\}$ *and* $\{\varepsilon_t^2\}$ *jointly converge*; *in particular*,

$$\left[\mathbb{W}_T^2\ \mathbb{W}_T^*\right]' \Rightarrow \left[\mathbb{W}_1^2\ \mathbb{W}^*\right]',$$

where $\mathbb{W}_T$ *is the PSP of* $\{\varepsilon_t\}$ *and* $\mathbb{W}_T^*$ *that of* $\{\varepsilon_t^2\}$. *In the limit process*, $\mathbb{W}_1$ *is a Brownian motion*, $\mathbb{W}^* = \rho\mathbb{W}_1 + \sqrt{(1-\rho^2)}\,\mathbb{W}_2$ *and* $\mathbb{W}_2$ *is a Brownian motion independent of* $\mathbb{W}_1$.

PROOF. Set $\sigma_\varepsilon^2 = E(\varepsilon_t^2)$. Define $\mathbb{W}_T(r)$ to be the partial sum process (PSP) of ε_t, and $\mathbb{W}_T^*(r)$ to be that of ε_t^2, that is

$$\mathbb{W}_T(r) = T^{-1/2}\sigma^{-1}\sum_1^{\langle Tr\rangle}\varepsilon_t, \qquad \mathbb{W}_T^*(r) = T^{-1/2}\kappa^{-1}\sum_1^{\langle Tr\rangle}\left(\varepsilon_t^2 - \sigma_\varepsilon^2\right).$$

It then follows from the mixing assumption that

$$\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}^{-1/2}\left[\mathbb{W}_T \quad \mathbb{W}_T^*\right]' \Rightarrow \left[\mathbb{W}_1 \quad \mathbb{W}_2\right]'$$

where $\mathbb{W}_1$ and $\mathbb{W}_2$ are independent Brownian motions. Inverting the correlation matrix gives

$$\left[\mathbb{W}_T \quad \mathbb{W}_t^*\right]' \Rightarrow \left[\mathbb{W}_1, \rho\mathbb{W}_1 + \sqrt{(1-\rho^2)}\,\mathbb{W}_2\right]' = \left[\mathbb{W}_1 \quad \mathbb{W}^*\right]'.$$

The lemma now follows from the continuous mapping theorem (CMT). □

The proofs of Lemma A2 and Theorem 2 rely heavily on Hansen (1992) for technique and notation.

LEMMA A2 (Remainder term). *When $U_T = \mathbb{W}_T^2$ and $V_T = \mathbb{W}_T^*$, the martingale approximation term Λ_T [see Hansen (1992), equation (3)] converges to zero in probability.*

PROOF. Now,

$$\Lambda_T = T^{-1/2} \sum_{t=1}^{T} \{U_{Tt} - U_{T(t-1)}\} z_t - T^{-1/2} U_{TT} z_{T+1}, \tag{A.3}$$

where $U_{Ti} = (T^{-1/2}\sigma^{-1}\Sigma_1^{i-1}\varepsilon_t)^2$ and $z_t = \Sigma_1^{\infty} E[\varepsilon_{t+k}^2 - \sigma_\varepsilon^2 | \mathfrak{F}_t]$. Note that z_t is based on the mixing process $\{\varepsilon_t^2 - \sigma_\varepsilon^2\}$ and thus, as in Hansen [(1992), (A3)],

$$T^{-1/2} \sup_{t \le T} |z_t| \to_p 0. \tag{A.4}$$

Looking at the last term in (A.3) first,

$$|T^{-1/2} U_{TT} z_{T+1}| \le |U_{TT}| \cdot |T^{-1/2} z_{T+1}|.$$

Under Assumption 1, the PSP of $\{\varepsilon_t\}$ converges and the first term is $O_p(1)$, while the second converges to zero in probability by (A.4). Thus the product vanishes asymptotically. As to the summation term in (A.3), redefine $U_{Ti} = W_{Ti}^2$ and note that

$$\begin{aligned} U_{Ti} - U_{T(i-1)} &= \left(W_{Ti} - W_{T(i-1)}\right)^2 + 2 W_{Ti} W_{T(i-1)} \\ &= T^{-1}\sigma^{-2}\varepsilon_i^2 + 2T^{-1}\sigma^{-2} \sum_1^{i-1} \varepsilon_t \sum_1^{i-2} \varepsilon_t. \end{aligned}$$

Hence

$$\begin{aligned} T^{-1/2} \sum_{t=1}^{T} \{U_{Tt} - U_{T(t-1)}\} z_t = {}& T^{-3/2}\sigma^{-2} \sum_{t=1}^{T} \varepsilon_t^2 z_t \\ &+ 2T^{-3/2}\sigma^{-2} \sum_{t=1}^{T} \left(\sum_1^{i-1} \varepsilon_j \sum_1^{i-2} \varepsilon_j \right) z_t. \end{aligned}$$

Start with the first term:

$$T^{-3/2}\sigma^{-2} \sum_{t=1}^{T} \varepsilon_i^2 z_t = T^{-3/2}\sigma^{-2} \sum_{t=1}^{T} \left(\varepsilon_t^2 - \sigma_\varepsilon^2\right) z_t + T^{-3/2}\sigma^{-2}\sigma_\varepsilon^2 \sum_{t=1}^{T} z_t.$$

By Hansen [(1992), Theorem 4.1], both terms converge in probability to zero because of the presence of the extra $T^{-1/2}$ here. As for the other term, Hölder's inequality followed by by the weak convergence of the PSP of $\{\varepsilon_t\}$ and (A4) show that it, too, converges in probability to zero. Thus the lemma is proved. □

Under Assumption 1, and in view of Corollary 1, we are led to consider the asymptotic behavior only of Z_T as given in Theorem 2.

PROOF OF THEOREM 2. Using the properties of deviations from means, the statistic of interest may be rewritten as

$$T^{-3/2}\sigma^{-2}\kappa^{-1}\sum\left(W_t-\overline{W}\right)\left(\varepsilon_t^2-\sigma_\varepsilon^2\right),$$

where $W_t=(\Sigma_1^{t-1}\varepsilon_j)^2$ and $\overline{W}=T^{-1}\Sigma W_t$. Employing the partial sum process notation of Lemma A1,

$$\begin{aligned}Z_T&=\int_0^1\left\{\mathbb{W}_T(t)^2-\int_0^1\mathbb{W}_T(s)^2\,ds\right\}d\mathbb{W}_T^*(t)\\&=\int_0^1\mathbb{W}_T(t)^2\,d\mathbb{W}_T^*(t)-\int_0^1\mathbb{W}_T(s)^2\,ds\,\mathbb{W}_T^*(1).\end{aligned}$$

Next apply the martingale approximation of Hansen (1992) to the stochastic integral in the first term to get

$$Z_T=\int_0^1\mathbb{W}_T(t)^2\,d\mathbb{Y}_T(t)+\Lambda_T-\int_0^1\mathbb{W}_T(s)^2\,ds\,\mathbb{W}_T^*(1),$$

where $\mathbb{Y}_T(t)$ follows Hansen's notation and represents a martingale process. Using Lemma A1, in conjunction with Hansen [(1992), Theorem 2.1] and the CMT, it is seen that

$$\begin{aligned}Z_T&\Rightarrow\int_0^1\mathbb{W}_1(t)^2\,d\mathbb{W}^*(t)-\int_0^1\mathbb{W}_1(s)^2\,ds\,\mathbb{W}^*(1)\\&=\int_0^1\left[\mathbb{W}_1(t)^2-\int_0^1\mathbb{W}_1(s)^2\,ds\right]d\mathbb{W}^*(t)\end{aligned}$$

once it is established that $\Lambda_T\to_p 0$ by Lemma A2. □

PROOF OF COROLLARY 1. The proof of this corollary is very similar to that of Theorem 2, so we shall use the same notation and merely sketch the details. Write $W_t=\Sigma_1^t\varepsilon_j$ so that

$$\begin{aligned}&T^{-3/2}\sigma^{-1}\kappa^{-1}\left[\sum w_{t-1}\left(\varepsilon_t^2-\tilde{\sigma}^2\right)\right]\\&\quad=T^{-1/2}T^{-1}\sigma^{-1}\kappa^{-1}\left[\sum\left(W_{t-1}-\overline{W}\right)\left(\varepsilon_t^2-\sigma_\varepsilon^2\right)\right]\\&\quad=T^{-1/2}\int_0^1\left\{\mathbb{W}_T(t)-\int_0^1\mathbb{W}_T(s)\,ds\right\}d\mathbb{W}_T^*(t)\\&\quad=T^{-1/2}\left[\int_0^1\left\{\mathbb{W}_T(t)-\int_0^1\mathbb{W}_T(s)\,ds\right\}d\mathbb{Y}_T(t)+\Lambda_T\right].\end{aligned}$$

Now the integral expression is $O_p(1)$ by arguments similar to Theorem 2, and so multiplication by $T^{-1/2}$ ensures the product is $o_p(1)$. Reasoning identical to that in Lemma A2 also shows that $T^{-1/2}\Lambda_T\to_p 0$. □

REFERENCES

BOX, G. E. P. and JENKINS, G. M. (1976). *Time Series Analysis: Forecasting and Control*, 2nd ed. Holden-Day, San Francisco.

BOX, G. E. P., JENKINS, G. M. and REINSEL, G. C. (1994). *Time Series Analysis: Forecasting and Control*, 3rd ed. Prentice-Hall, Englewood Cliffs, NJ.

CRAMÉR, H. (1961). On some classes of nonstationary stochastic processes. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **2** 55–77. Univ. California Press, Berkeley.

DICKEY, D. A. (1976). Hypothesis testing for nonstationary time series. Ph.D. dissertation, Iowa State Univ.

DICKEY, D. A. and FULLER, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* **49** 1057–1072.

ENGLE, R. F. and GRANGER, C. W. J. (1987). Co-integration and error correction: representation, estimation and testing. *Econometrica* **55** 251–276.

FERGUSON, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach.* Academic Press, New York.

FULLER, W. A. (1976). *Introduction to Statistical Time Series.* Wiley, New York.

GRANGER, C. W. J. and TERÄSVIRTA, T. (1993). *Modeling Nonlinear Economic Relationships.* Oxford Univ. Press.

HANSEN, B. E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* **8** 489–500.

LEYBOURNE, S. J. and McCABE, B. P. M. (1989). On the distribution of some tests for coefficient constancy. *Biometrika* **76** 167–177.

LEYBOURNE, S. J., McCABE, B. P. M. and TREMAYNE, A. R. (1994). Can economic time series be differenced to stationarity? Unpublished manuscript.

McCABE, B. P. M. and TREMAYNE, A. R. (1989). Testing if a time varying autoregressive model is a random walk. Unpublished paper presented at the European Meeting of the Econometric Society, Munich.

NICHOLLS, D. F. and QUINN, B. G. (1982). *Random Coefficient Autoregressive Models: An Introduction.* Springer, New York.

PHILLIPS, P. C. B. (1987). Time series regression with a unit root. *Econometrica* **55** 277–302.

PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series.* Academic Press, London.

TONG, H. (1990). *Non-linear Time Series: A Dynamical Systems Approach.* Clarendon, Oxford.

TYSSEDAL, J. S. and TJØSTHEIM, D. (1988). An autoregressive model with suddenly changing parameters and an application to stock market prices. *J. Roy. Statist. Soc. Ser. C* **37** 353–369.

FACULTY OF COMMERCE
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, BC
CANADA V6T 1Z2

DEPARTMENT OF ECONOMETRICS
UNIVERSITY OF SYDNEY
NSW 2006
AUSTRALIA