

SPECIAL SECTION: A WINNING ENTRY OF THE 1998 STUDENT PAPER COMPETITION

The Statistical Computing Section of the American Statistical Association held its annual Student Paper Competition in 1998. Winners in the competition received a prize and the opportunity to publish their winning entry in this special section of *JCGS*. The papers were subjected to an additional review by the Editorial Board of *JCGS*. We publish here the third of the four winning entries; two were published in the September 1999 issue. We congratulate the young authors for their excellent work.—*The Editor*

Adaptive Tests for Changing Unit Roots in Nonstationary Time Series

Carlo GRILLENZONI

This article considers tests for unit roots in time series models with varying parameters. The null hypothesis is that roots are unity against an alternative where they change over time. Tests statistics are based on recursive least squares (RLS) estimates having exponentially weighted (EW) observations. This method belongs to the class of nonparametric estimators and allows interesting computational and graphical aspects. Asymptotic properties are investigated as in kernel estimation, by allowing smoothing coefficients tending to zero. Under the null, we find that test statistics approach the distributions tabulated by Dickey and Fuller. Applications to real and simulated data show the validity of the method.

Key Words: Dickey–Fuller tests; Exponential window; Nonstationarity and instability; Random walks; Recursive estimators; Wiener process.

1. INTRODUCTION

Time series are usually represented by polynomial models whose stability properties depend on the location of their roots (see Box and Jenkins 1976). So-called unit roots lie on the unit circle of the complex plane—that is, have modulus one. Their presence has important consequences on the behavior and the inference of the models. In particular, unit roots imply instability and can generate persistent components of the real world, such as stochastic trends. On the other hand, they affect the asymptotic properties of the estimates of both parameters and predictors.

Carlo Grillenzoni is Associate Professor of Statistics, University Institute of Architecture of Venice, S. Croce 1957, 30135 Venezia, ITALY (E-mail: carlog@iuav.unive.it).

©1999 American Statistical Association, Institute of Mathematical Statistics,
and Interface Foundation of North America

Journal of Computational and Graphical Statistics, Volume 8, Number 4, Pages 763–778

Inference on unit-root processes is an important topic of mathematical statistics and has a long history. Seminal contributions are those of White (1959), Rao (1978), Dickey and Fuller (1979), and Phillips and Perron (1988) (see Hamilton 1994 for a review). Main findings are that least squares estimates have a faster convergence rate, but have a nonstandard distribution when models contain unit roots. Such distribution was derived in analytical form by Rao (1978), but can also be expressed as the ratio of functionals on Brownian motion. This feature is useful for carrying out simulations.

Testing for the presence of unit roots in a time series is useful both for identifying the nature of the process and for applying the model building methodology of Box and Jenkins (1976). Nowadays, Dickey–Fuller (DF) tests are the most popular tools used in this field. They are based on nonstandard distributions which have been simulated and tabulated by the two authors (see Fuller 1996). Apart from nonstandard aspects, DF tests resemble classical procedures and can be easily extended.

Except for the recent work by Granger and Swanson (1996), small attention has been devoted to the possible time-variability of unit roots. This aspect is important because regression parameters may change over time (see Grillenzoni 1994); moreover, it may contrast the tendency of unstable models to diverge. To be specific, real time series usually contain trends, but we cannot expect them to grow indefinitely over time. Now, roots wandering on the unit circle can stabilize the level of a process because they create changing points where the slope of the trends inverts.

Granger and Swanson (1996) assumed that roots change stochastically around unity and define a nonlinear model for their fluctuations. They showed that DF tests can distinguish between constant and stochastic unit roots, but the power decreases. This article assumes that roots change smoothly but in an unknown manner. In this case, a proper nonparametric estimator is recursive least squares (RLS) with discounted observations. I define recursive statistics for testing for unit roots and study their distribution as in kernel estimation. The method actually provides sequential DF tests which can also be used for checking the constancy over time of the roots.

The plan of the work is as follows: Section 2 provides background material on recursive estimation and tests for unit roots. Section 3 defines recursive DF tests and studies their distribution analytically and by simulations. Section 4 checks the validity of the method with applications to real and simulated data.

2. MODELS AND METHODS

Using the Box–Jenkins symbols, the basic representation we consider is an autoregressive (AR) process whose parameter (root) wanders on the unit circle:

$$\begin{aligned} Z_t &= \phi_t Z_{t-1} + a_t \\ a_t &\sim \text{iid}(0, \sigma^2), \quad Z_0 = a_0 \\ \bar{\phi} &= \lim_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n |\mathbf{E}(\phi_t)| = 1 \end{aligned} \quad (2.1)$$

Specifically, the sequence $\{\phi_t\}$ has time-average $\bar{\phi}$ with unit modulus and the input $\{a_t\}$ is a sequence of independent identically distributed (iid) variates having finite variance. In this way, (2.1) can be classified as an *evolving unstable* AR(1) process.

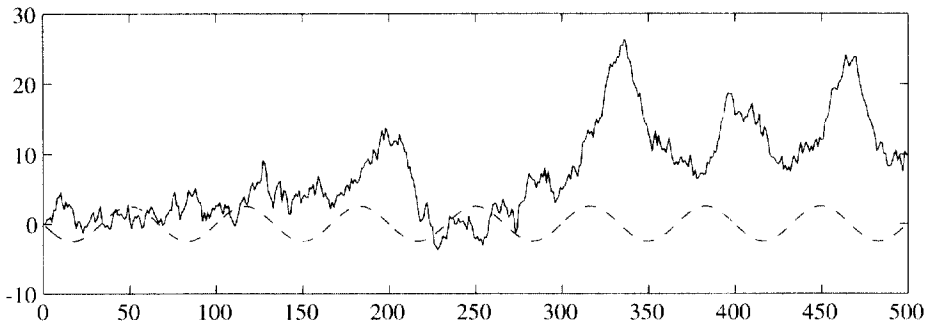


Figure 1. A realization of the process (2.1) with sinusoidal parameters.

On average, the process (2.1) is a random walk; however, fluctuations of the root inside and outside the unit circle determine more complex behavior, such as local stability as well as local trends. These features crucially depend on the path of the root, on which we do not make specific assumptions. In particular, ϕ_t may be stochastic or deterministic, linear or nonlinear, white noise or random walk, smoothly or suddenly changing. The sole restriction we establish concerns the average trajectory $\bar{\phi}$.

An Example. Figure 1 shows a realization of the process (2.1) with the designs $a_t \sim \text{IN}(0, 1)$, $Z_0 = 0$ and $\phi_t = 1 - .5 \sin(t/10.6)$, where $t = 1 \dots 500$. This specification is such that $\min(\phi_t) = .95$ and $\max(\phi_t) = 1.05$, with a period of length 66.6 . We may see that fluctuations of the parameter on the unit circle considerably influence the level of the series. The transient in Z_t depends on the fact that unstable processes are sensitive to the initial condition. This behavior may be reduced by choosing $Z_0 > 0$.

Inference on the model (2.1) in its general form is difficult, especially when parameters are stochastic. If $\{\phi_t\}$ is an AR process itself, then the Kalman filter (KF) can estimate its conditional mean $E(\phi_t | Z_{t-1}, Z_{t-2} \dots)$. When $\{\phi_t\}$ has an unknown dynamic, however, a nonparametric estimator must be defined. Consistently with the approach of local regression (see Hastie and Loader 1993 or Fan and Gijbels 1996), we may consider the least squares algorithm with exponentially weighted (EW) observations

$$\hat{\phi}_t(\lambda) = \frac{\sum_{i=2}^t \lambda^{t-i} Z_{i-1} Z_i}{\sum_{i=2}^t \lambda^{t-i} Z_{i-1}^2}, \quad 0 < \lambda < 1, \tag{2.2}$$

where λ is the discounting rate. When $\phi_t = \phi$ constant, it is easy to show that estimator (2.2) minimizes the weighted loss function $Q_t(\phi) = \sum_{i=2}^t \lambda^{t-i} (Z_i - \phi Z_{i-1})^2$.

The recursive implementation of (2.2) allows fast calculation and is more transparent in showing its ability to estimate the sequence $\{\phi_t\}$. As in Grillenzoni (1994) we have

$$\begin{aligned} R_t &= \lambda R_{t-1} + Z_{t-1}^2 \\ \hat{a}_t &= Z_t - \hat{\phi}_{t-1} Z_{t-1} \\ \hat{\phi}_t &= \hat{\phi}_{t-1} + R_t^{-1} Z_{t-1} \hat{a}_t \\ \tilde{a}_t &= Z_t - \hat{\phi}_t Z_{t-1} \\ \hat{\sigma}_t^2 &= \lambda \hat{\sigma}_{t-1}^2 + (1 - \lambda)(\hat{a}_t \tilde{a}_t). \end{aligned} \tag{2.3}$$

where R_t is the weighted sum of squared regressors, \hat{a}_t is the prediction error, \bar{a}_t is the recursive residual, and $\hat{\sigma}_t^2$ is an estimator of the residual variance σ_a^2 .

Algorithm (2.3) is algebraically equivalent to (2.2); however, since $t \geq 2$ it needs the starting values $(R_1, \hat{\phi}_1, \hat{\sigma}_1^2) > 0$. The adaptive ability of (2.3) depends on these values, and especially on the factor λ . Its action can be appreciated either by the fact that it gives more weight to recent observations or by preventing $R_t^{-1} \rightarrow 0$ in (2.3). Such a coefficient tends, therefore, to have the same role as the bandwidth in kernel type estimators. In the Appendix, Section A.1, we shed light on this point by defining the correspondence between (2.2)–(2.3) and classical nonparametric estimators.

The Problem. On the basis of the algorithm (2.3), this article aims to derive recursive tests for unit roots. We are also interested in using these methods for testing the constancy over time of the parameters of model (2.1). In both cases the system of hypotheses is given by

$$H_0 : \phi_t = 1 \quad \text{versus} \quad H_1 : \phi_t \neq 1. \tag{2.4}$$

Under the null, the model (2.1) becomes a true random walk and the Dickey–Fuller tests could be used. However, the DF statistics must be adapted to the new alternative; in particular, they must be made sensitive to parameter variations.

Typically, DF tests are based on the statistic $S_n = n(\hat{\phi}_n - 1)$, where $\hat{\phi}_n$ is the OLS estimator, n is the sample size, and 1 is the null. The hypothesis H_0 is rejected in favor of $H_1 : \phi \neq 1$ if S_n lies outside critical values tabulated by Fuller (1996, p. 640). Now, S_n can be adapted to the system (2.4), by replacing $\hat{\phi}_n$ with $\hat{\phi}_t(\lambda)$ and proceeding for each $t \geq 2$. However, two technical questions arise from this approach: (1) What is the quantity to be used in place of n ?; and (2) What is the distribution of the resulting statistic $S_t(\lambda)$? The answers will be provided in the next section. In the meantime, we note that the implicit sample size of (2.2)–(2.3) is given by $(\sum_{i=1}^t \lambda^{t-i}) \rightarrow 1/(1 - \lambda)$.

3. ADAPTIVE DICKEY–FULLER TESTS

As in the classical analysis (e.g., White 1958 or Hamilton 1994, p. 475) we preliminarily define bounding functions for the numerator and the denominator of (2.2). Under the null (2.4), it is well known that $Z_t = O_p(\sqrt{t})$, because $Z_t = \sum_{i=1}^t a_i$ and $\text{var}(Z_t) = t\sigma_a^2$. Therefore, using the asymptotic sample size $1/(1 - \lambda)$, it is easy to prove (see Appendix A.2)

$$\begin{aligned} \sum_{t=2}^n \lambda^{n-t} Z_{t-1} &= O_p\left(\frac{\sqrt{n}}{1-\lambda}\right) \\ \sum_{t=2}^n \lambda^{n-t} Z_{t-1}^2 &= O_p\left(\frac{n}{1-\lambda}\right) \\ \sum_{t=2}^n \lambda^{n-t} Z_{t-1} a_t &= O_p\left(\frac{\sqrt{n}}{\sqrt{1-\lambda}}\right). \end{aligned} \tag{3.1}$$

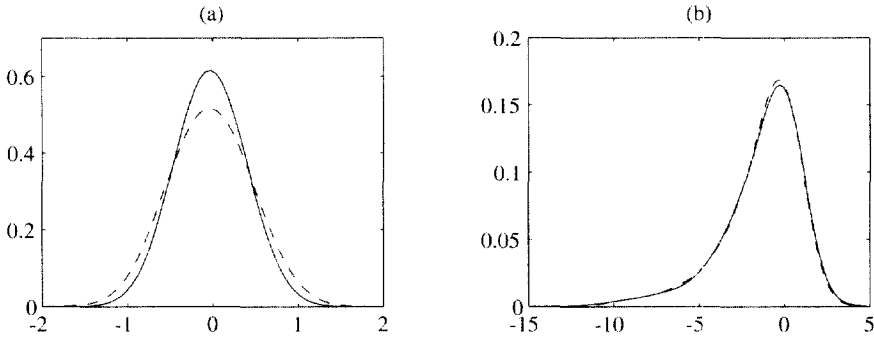


Figure 2. Kernel density estimates of OLS (---) and WLS (—) statistics of random walks: (a) $n^{-3/2} \sum_{t=1}^n Z_t$ (---), $n^{-1/2}(1-\lambda) \sum_{t=1}^n \lambda^{n-t} Z_t$ (—); (b) S_n (···), $S_n(\lambda)$ (—).

Statistic 1. On the basis of results (3.1), the adaptive DF test statistic becomes

$$\begin{aligned}
 S_t(\lambda) &= \frac{\sqrt{t}}{\sqrt{1-\lambda}} \left[\hat{\phi}_t(\lambda) - 1 \right] \\
 &= \left[\frac{(1-\lambda)}{t} \sum_{i=2}^t \lambda^{t-i} Z_{i-1}^2 \right]^{-1} \frac{\sqrt{1-\lambda}}{\sqrt{t}} \sum_{i=2}^t \lambda^{t-i} Z_{i-1} a_i. \quad (3.2)
 \end{aligned}$$

In fact, the fixed sequence $\sqrt{t/(1-\lambda)}$ that replaces n , arises from the product of the bounding terms of numerator and denominator of $[\hat{\phi}_t(\lambda) - 1]$ under the null.

The asymptotic analysis of (3.2) with $\lambda < 1$ is difficult because it does not converge to a well defined random variable; therefore, we need to let $\lambda \rightarrow 1$. On the other hand, the behavior of (3.2) by running $\lambda \rightarrow 1$ before $t \rightarrow \infty$ is trivial because the WLS estimator converges in probability to the OLS one. As in Grillenzoni (1996) this may be avoided by using a double limit operator $\lim_{\lambda \rightarrow 1} \lim_{t \rightarrow \infty} f(t, \lambda)$ in which the order is not invertible and by running t, λ at the same rate, namely $\lim_{\lambda \rightarrow 1} \lim_{t \rightarrow \infty} [t(1-\lambda)] = 1$. This approach resembles the analysis of kernel estimators (see Härdle 1991), and can be accomplished by using $\lambda_t = (1 - 1/t)$.

Now, under the null and the stated conditions, we can prove that (3.2) converges to the same limiting distribution as $S_n = n(\hat{\phi}_n - 1)$. This does not hold for its numerator and denominator taken separately, however. In particular, the asymptotic distribution is the ratio of two random variables (which may be expressed as functionals on Wiener process $W(\cdot)$) whose variances are half of those of the classical analysis. From Appendix A.4, as $t \rightarrow \infty, \lambda \rightarrow 1$ and $t(1-\lambda) \rightarrow 1$, we have

$$\frac{\sqrt{t}}{\sqrt{1-\lambda}} \left[\hat{\phi}_t(\lambda) - 1 \right] \xrightarrow{L} \frac{(\sigma_a^2/\sqrt{2}) \int_0^1 W(s) dW(s)}{(\sigma_a^2/\sqrt{2}) \int_0^1 W^2(s) ds}. \quad (3.3)$$

By deleting $(\sigma_a^2/\sqrt{2})$, one obtains the distribution introduced by White (1959) and tabulated by Dickey and Fuller (see Fuller 1996, p. 641), hence denoted as DF(1).

Simulation experiments were conducted to check the validity of the results in (3.3). Figure 2(a) compares the kernel density estimates of $n^{-3/2} \sum_{t=1}^n Z_t$ and $n^{-1/2}(1 -$

$\lambda) \sum_{t=1}^n \lambda^{n-t} Z_t$, with $\lambda = (1 - 1/n)$. Figure 2(b) repeats the exercise for the statistics S_n and (3.2). In both cases the designs in (2.1) were $a_t \sim \text{IN}(0, 1)$, $\phi_t = 1$, $Z_0 = a_0$; the number of replications was $N = 200$ and the sample size was $n = 100$.

One may note that only the variances in Figure 2(a) are different; in particular, that of OLS (.325) is nearly twice that of WLS (.164). Further experiments have also shown that this conclusion may be extended to other kinds of statistics, such as $n^{-2} \sum_{t=1}^n Z_t^2$ and $n^{-1}(1 - \lambda) \sum_{t=1}^n \lambda^{n-t} Z_t^2$. Finally, Figure 2(a) shows that bounded sums of random walks are nearly normally distributed across the samples, whereas Figure 2(b) exhibits the well known asymmetric profile of the DF(1) distribution.

Statistic 2. To avoid the problem of defining the “sample size” in (3.2), one may consider the t -type statistic used in tests for significance, namely $T_n = (\hat{\phi}_n - 1)/\widehat{\text{SE}}(\hat{\phi}_n)$. From (2.3), the corresponding WLS statistic and its distribution are

$$\sqrt{R_t \hat{\sigma}_t^{-2}} \left[\hat{\phi}_t(\lambda) - 1 \right] \xrightarrow{L} \frac{\int_0^1 W(s) dW(s)}{\left[\sqrt{2} \int_0^1 W^2(s) ds \right]^{1/2}}, \tag{3.4}$$

as $t \rightarrow \infty$, $\lambda \rightarrow 1$ and $t(1 - \lambda) \rightarrow 1$ (see the Appendix, Section A.4, for the proof). We note that (3.4) coincides with the asymptotic distribution of T_n [namely DF(2), because it also was tabulated by Dickey and Fuller (see Fuller 1996, p. 642)] except for the presence of the constant $(1/2)^{1/4}$.

To implement sequential tests for unit roots which directly refer to the critical values of the DF(2) distribution, one must modify the statistic in (3.4) as follows

$$\begin{aligned} T_t(\lambda) &= \sqrt{R_t \hat{\sigma}_t^{-2} \sqrt{1 + \lambda}} \left[\hat{\phi}_t(\lambda) - 1 \right] \\ &= \left[\frac{(1 + \lambda)^{1/2} \sum_{i=2}^t \lambda^{t-i} Z_i^2}{(1 - \lambda) \sum_{i=2}^t \lambda^{t-i} \hat{u}_i^2(\lambda)} \right]^{1/2} \left[\hat{\phi}_t(\lambda) - 1 \right], \end{aligned} \tag{3.5}$$

where $\hat{u}_t^2 = (\hat{a}_t \tilde{u}_t)$ are defined in (2.3). Multiplying by $(1 + \lambda)^{1/4}$ in (3.5) actually inflates the sum of squared regressors R_t , which is underestimated as a consequence of the exponential weighting. The result is loosely related to the sampling variance of (2.3) in the case of stationary series: $\widehat{\text{var}}(\hat{\phi}_t) = \hat{\sigma}_t^2 R_t^{-1}/(1 + \lambda)$, see Grillenzoni (1996) and (3.6). However, a fundamental difference is in the power of $(1 + \lambda)$.

Simulation experiments were conducted to check the validity of results (3.4) and (3.5). Under the same design conditions as Figure 2(a), Figure 3 compares the kernel density estimates of the statistics $T_n(\text{OLS})$ and $T_n(\lambda)$, obtained with $\lambda = (1 - 1/n)$. One may see that distribution of (3.5) is very close to that of T_n ; hence, the DF(2) tabulation can be used for it.

Testing. From the results (3.2)–(3.5) one can develop sequential tests for unit roots which have the same structure as those based on OLS estimates. In particular:

1. For each $t = 2, 3, \dots$ one rejects $H_0: \phi_t = 1$ when $S_t, T_t(\lambda)$ lie outside the critical values of the DF distributions; and
2. if H_0 is rejected for certain values of t and accepted for others, then parameter variability is also detected.

Some words of caution are necessary when using the above decision rules:

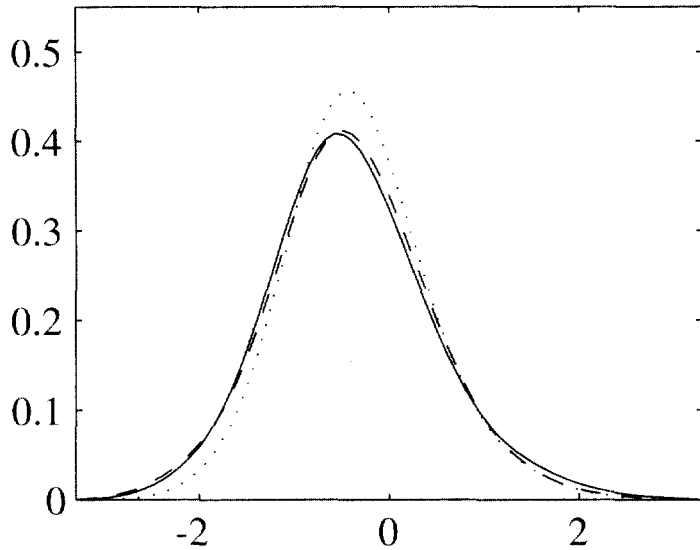


Figure 3. Kernel density estimates of T_n -statistics: OLS (---), (3.5) (—), (3.4) (···).

1. Under the null, the quantity $[\hat{\phi}_t(\lambda) - 1]$ converges in probability to zero as $t \rightarrow \infty$ and $\lambda \rightarrow 1$, but nothing can be said under the alternative (2.4), because the sequence $\{\phi_t\}$ has an unknown dynamic.
2. Results (3.3) and (3.4) are asymptotic ones, therefore the testing procedure holds for values of t, λ sufficiently large. The problem may be reduced for $T_t(\lambda)$ by using algorithm (2.3) and suitable design for the initial estimates $R_1, \hat{\phi}_1, \hat{\sigma}_1^2$.
3. From Figures 2(b) and 3, one may note that DF critical values are not symmetrical around 0, therefore the power of the tests under $H_1: \phi > 1$ is greater than that under $H_1: \phi < 1$. In other words, it is easier to identify a local trend rather than a local stability. This situation is dramatic for $S_t(\lambda)$.
4. As stated before, the method also serves as a test for the constancy of the root of model (2.1), which is unity in time average. However, an F test on the reduction of the sum $Q_n = \sum_{t=2}^n \hat{a}_t^2(\lambda)$, compared with that of the OLS estimator, is more powerful for this purpose (see Grillenzoni 1994).

With respect to classical DF tests, the effect of the coefficient $\lambda < 1$ is to reduce the “sample size” of the LS estimator. Hence, under H_0 we expect that statistics $S_t, T_t(\lambda)$ approach a steady-state value. However, owing to the sequence $\sqrt{t/(1-\lambda)}$, the value of $S_t(\lambda)$ is increasing and leads to the rejection of the null. In order to stabilize its behavior one may replace t large with $1/(1-\lambda)$, the *effective* number of observations of the estimator (2.3). This choice may, however, underestimate the test statistic.

An intermediate solution can be obtained from the dispersion of (2.2)–(2.3) in the case of stationary/stable time series $\{z_t\}$. Following Grillenzoni (1996), one has

$$\lim_{t \rightarrow \infty} E \left[\hat{\phi}_t(\lambda) - \phi \right]^2 = \left(\frac{1-\lambda}{1+\lambda} \right) E \left(z_{t-1}^2 \right)^{-1} \sigma_a^2 \tag{3.6}$$

(see also the Appendix, Section A.3). Compared with the variance of the OLS estimator, one may note that the ratio $(1-\lambda)/(1+\lambda)$ has the same role as $1/n$; therefore, it represents the *equivalent* number of observations. This term can be used for tuning $[\hat{\phi}_t(\lambda) - 1]$, whenever the series Z_t does not contain marked trends (i.e., roots greater than unity).

Summing up, the adaptive DF statistics developed in this article are given by

$$\tilde{S}_t(\lambda) = (1-\lambda)^{-1}(1+\lambda) \left[\hat{\phi}_t(\lambda) - 1 \right] \approx \text{DF}(1) \quad (3.7)$$

$$T_t(\lambda) = \sqrt{R_t \hat{\sigma}_t^{-2} \sqrt{1+\lambda}} \left[\hat{\phi}_t(\lambda) - 1 \right] \sim \text{DF}(2), \quad (3.8)$$

where the distributions hold for values of t, λ sufficiently large. Test procedures based on (3.7)–(3.8) enjoy differences and similarities as we now discuss.

Similarities. In both cases, decision rules for testing H_0 are those of the constant parameter case. Moreover, the statistics are relatively invariant to choice of the factor λ , because their components tend to balance each other. In practice, when λ decreases, the algorithm sensitivity increases and hence the distance $\|\hat{\phi}_t(\lambda) - 1\|$; on the other hand, the remaining components of (3.7)–(3.8) decrease.

Differences. Solution (3.8) has a great advantage over (3.7) in that it depends only indirectly on the sample size. Moreover, a similar feature holds for the DF(2) distribution, whose critical values do not vary substantially for $n > 25$ (see Fuller 1996, p. 642). Finally, for reasons discussed below (3.6) the statistic, \tilde{S}_t , can be used in practice only if the series Z_t does not contain marked trends.

4. APPLICATIONS AND SIMULATIONS

This section presents applications to real and simulated data that aim to check the validity of the framework developed in Section 3. Case studies focus on two financial datasets published in statistical books. This choice is motivated by the fact that for many financial processes the hypothesis of random walk (or unit root) models is theoretically and empirically verified. Largely untreated is the problem of parameter variability.

Application 1. First we consider the daily IBM stock price series published in Box and Jenkins (1976) and displayed in Figure 4(a). Figures 4(b) and (c) show recursive estimates of model (2.1) obtained with the algorithm (2.3) implemented with $\lambda = .97, .85$. Initial values $R_1, \hat{\phi}_1, \hat{\sigma}_1^2$ were obtained from OLS estimates over the first $n = 25, 15$ observations. Figures 4(d), (e), and (f) display the statistics (3.7)–(3.8) together with their 95% critical values.

One may note that estimates $\hat{\phi}_t$ vary around unity and statistics $T_t(\lambda)$ show that such variability is significant. In fact, the test rejects the null in certain periods and accepts it in others. Furthermore, rejection occurs both in favor of $\phi > 1$ (at the beginning) and $\phi < 1$ (at the end of the sample). Since Figures 4(e) and (f) lead to the same decision, one can also conclude that the recursive T test is relatively insensitive to the design of λ .

Application 2. The second application focuses on the weekly average of the HSI index series published by Tong (1990) and plotted in Figure 5(a). Unlike the IBM one, this series has a marked stochastic trend, that should lead to rejecting H_0 in favor of

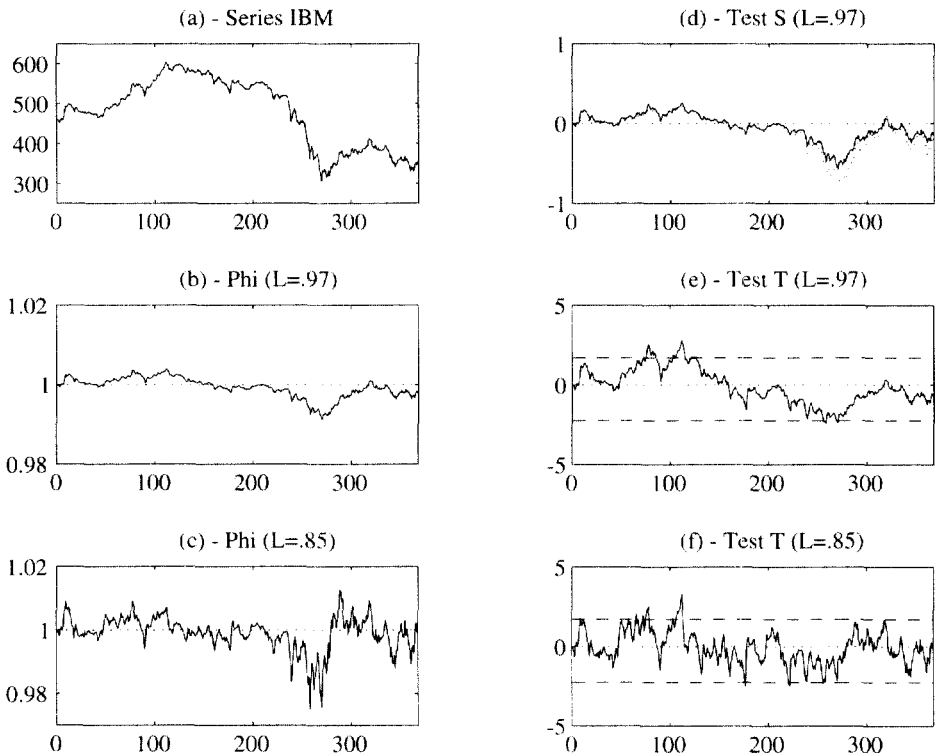


Figure 4. Results of Application 1: (a) IBM time series; (b)–(c) recursive estimates of ϕ_t with $\lambda = .97, .85$; (d) statistics (3.7) (—) and (3.2) (· · ·); (e)–(f) statistics (3.8) (—) and critical values (- - -).

$H_1 : \phi_t > 1$. Figures 5 (b) and (c) show the recursive estimates obtained with the same designs as in Application 1. In particular, $\lambda = .97$ belongs to the range $.95 \div .99$ which is usually suggested in the literature (e.g., Ljung and Soderstrom 1983), whereas the other is more extreme.

As regards as initialization of algorithm (2.3) with small sample OLS estimates, one can note that $n = 25, 15$ provide values of R_1 which render the initial variability of $\hat{\phi}_t$ consistent with the subsequent one. The Bayesian approach treats R_1^{-1} as the variance of the a priori distribution of ϕ_1 , and therefore recommends choosing it large—that is, R_1 small. However, in our experience this was not a good design, just because the scale of the series must be considered.

Figures 5(d), (e), and (f) show the path of statistics (3.7)–(3.8) together with their 95% critical values; the same conclusions as before can be drawn. In particular, test based on $T_t(\lambda)$ rejects H_0 in certain periods, detects significant variability of parameters, and is insensitive to the design of λ . Moreover, none of these findings were confirmed by the statistics $S_t, \tilde{S}_t(\lambda)$, whose values are uniformly insignificant on $2 \leq t \leq n$.

Simulation. We conclude with a simulation experiment that aims to check the performance of previous tests in situations of changing and constant parameters. For the process in Figure 1 with initial condition $Z_0 > 0$ (to reduce the problem of transient behavior), $N = 100$ replications of length $n = 200$ were fitted with algorithm (2.3). Best

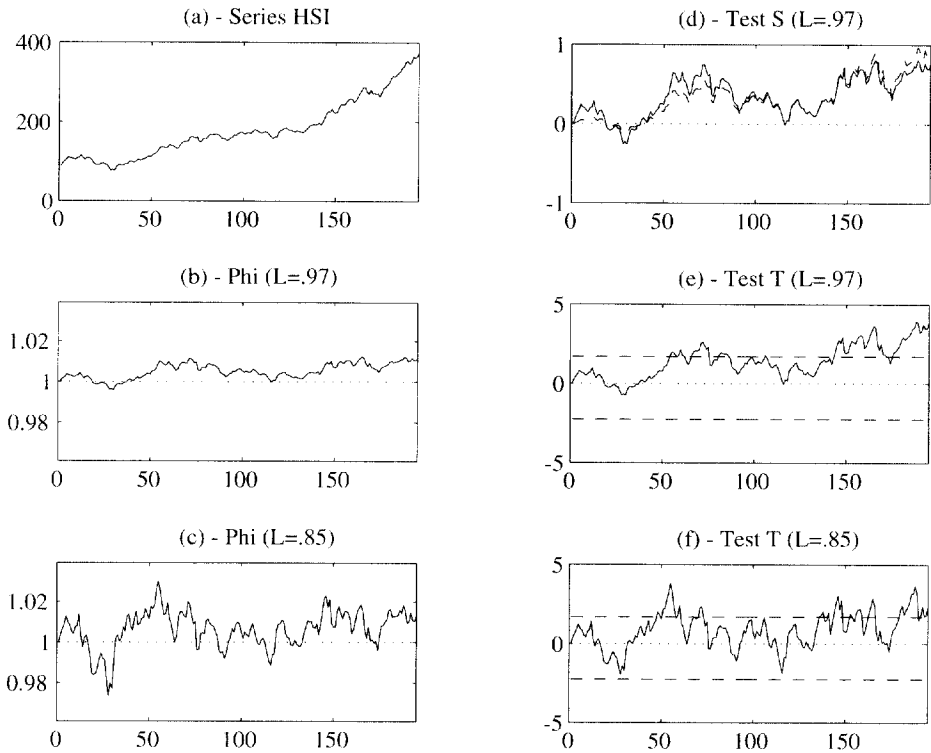


Figure 5. Results of Application 2: (a) HSI time series; (b) (c) recursive estimates of ϕ_t with $\lambda = .97, .85$; (d) statistics (3.7) (—) and (3.2) (- - -); (e)–(f) statistics (3.8) (—) and critical values (- - -).

MSE design was found to be $\lambda = .8$, whereas $R_1 > 0$ did not affect the average values $\bar{\phi}_t = 100^{-1} \sum_{i=1}^{100} \hat{\phi}_{it}, t = 1 \dots 200$ of the estimates, so it was selected as $R_1 = 1$, like $\hat{\phi}_1, \hat{\sigma}_1^2$.

Numerical results are displayed in Figure 6. Specifically: Figure 6(a) compares the paths of ϕ_t and $\bar{\phi}_t(\lambda)$. Figures 6(b) and (c) provide the mean values \bar{T}_t, \bar{S}_t of statistics (3.7)–(3.8), together with their 95% critical values. Figure 6(d) shows the values of \bar{T}_t generated under H_0 , together with their twice standard errors.

We can conclude that the estimator (2.3) has a good estimation ability because the paths of $\phi_t, \bar{\phi}_t(\lambda)$ are quite close. As in applications to real data, only $T_t(\lambda)$ detects departure from unity and time-variability of the root. Figure 6(b) also shows that such a decision is clearer when the root is greater than one. Finally, the experiment in Figure 6(d) serves to evaluate the power of the T test. As we may see, the performance is uniformly good.

Final Remarks. From previous applications, we have seen that recursive DF tests can detect departures from unity of a root on certain periods. A practical meaning when test statistics cross critical bands is that roots change significantly. Because the slopes of the series also change significantly at these points, the procedure may be used for identifying turning points in nonstationary processes. This result can have important applications both in signaling and forecasting problems.

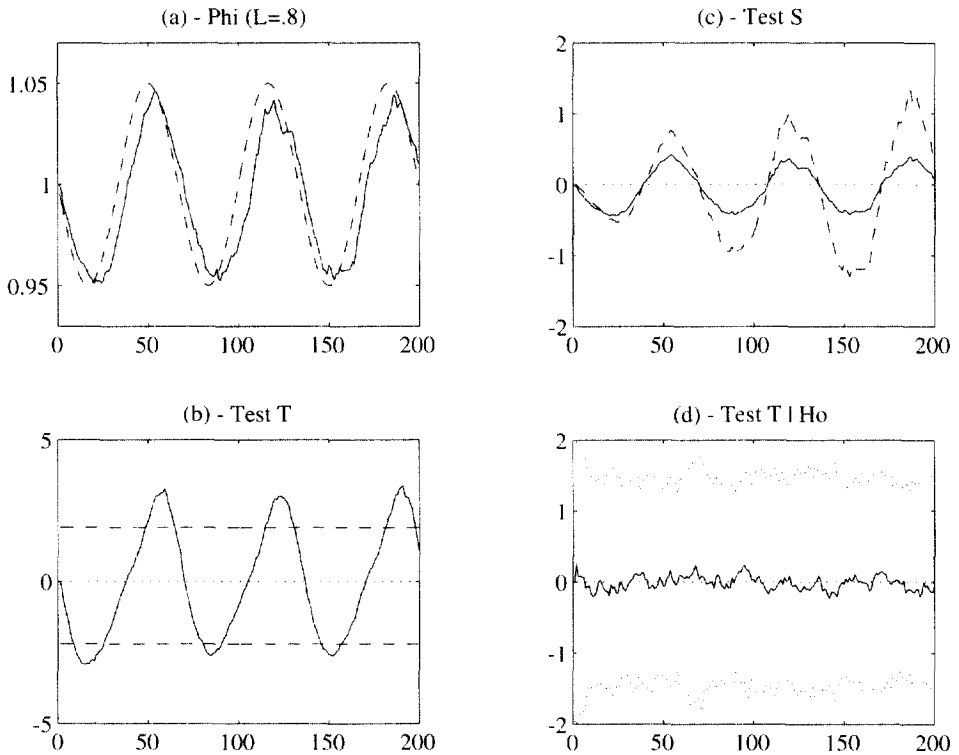


Figure 6. Results of Simulation: (a) parameter (---) and estimates (—); (b) statistics (3.8) (—) and critical values (---); (c) statistics (3.7) (—) and (3.2) (---); (d) statistics (3.8) (—) and 2SE (···) for a pure random walk.

It should be noted that our method is tuned on the trade-off between estimation sensitivity and estimation accuracy involved by the weighting rate. In particular, when λ decreases and the variability of test statistics increases, the equivalent sample size $(1 - \lambda)/(1 + \lambda)$ also decreases, so that the width of critical bands increases. As a consequence, the probability of rejecting H_0 when this is false (i.e., $\hat{\phi}_t \neq 1$ for the action of the sampling error) increases. However, unlike stationary processes (see Grillenzoni 1996), critical values do not vary significantly for $n > 25$. This invariance is due to the superconsistency property of the LS estimator when it is applied to unit-root processes.

Finally, our inferential method is adaptive, both in the sense of time-varying (as in engineering) and nonparametric (as in statistics). Here, the limit conditions developed in Section 3 resemble those in the analysis of kernel type estimators, where the bandwidth $h \rightarrow 0$ must satisfy $th^5 = O(1)$ (see Härdle 1991)—that is, $h \propto 1/\sqrt[5]{t}$. In our context the smoothing coefficient is $\lambda \rightarrow 1$, and must satisfy $t(1 - \lambda) = O(1)$, hence $(1 - \lambda) \propto 1/t$, as for the bandwidth.

APPENDIX: TECHNICAL DETAILS

A.1 WLS AND KERNEL ESTIMATION

The nonparametric nature of the framework (2.2)–(2.3) may be defined more specifically. It is well known that a kernel-type estimator for the regression function of the nonlinear model $Z_t = g(Z_{t-1}) + a_t$, minimizes the weighted criterion (see Härdle 1991)

$$P_t(z) = \sum_{i=2}^t w_i(z) [Z_i - g(z)]^2, \quad w_i(z) = \frac{K[(z - Z_{i-1})/h]}{t h \hat{f}_t(z)},$$

where $z \in \mathfrak{R}$ real, $K(\cdot)$ is the kernel, and $\hat{f}_t(z)$ is an estimator for the density. This approach can be used for deriving the kernel estimator for (2.1) when ϕ_t is a deterministic function. In this case, the regression function is $g(t, z) = \phi(t)z$ and conditionally on $z = Z_{t-1}$ the parametric part is $\phi(t)$; moreover, the weights of $P_t(\cdot)$ become $w_i(t) = K[(t - i)/h]/(th)$, because t is deterministic and $f(t) = 1$. Thus, from the relationship between the functionals $P(\cdot)$ and $Q(\cdot)$, the resulting algorithm is given by (2.2) with the weights λ^{t-i} just replaced by $K[(t - i)/h]$. Fundamental problems of this solution are the difficulty of obtaining the recursive implementation; moreover, it complicates the statistical analysis.

A.2 PROOF OF FORMULA (3.1)

We recall that any real random variable is as big as its standard deviation: $x = O_p(\sigma_x)$, thus the fundamental step is computing the variance of the sums in (3.1)

$$\begin{aligned} \text{var} \left(\sum_{t=1}^n \lambda^{n-t} Z_t \right) &= \text{E} \left(\sum_{t=1}^n \sum_{s=1}^n \lambda^{2n-t-s} Z_t Z_s \right) \\ &= \sum_{t=1}^n \sum_{s=1}^n \lambda^{n-t} \lambda^{n-s} t \sigma_a^2 \leq \frac{n \sigma^2}{(1 - \lambda^2)} \\ \text{var} \left(\sum_{t=1}^n \lambda^{n-t} Z_{t-1} a_t \right) &= \text{E} \left(\sum_{t=2}^n \sum_{s=2}^n \lambda^{2n-t-s} Z_{t-1} a_t a_s Z_{s-1} \right) \\ &= \sum_{t=2}^n \lambda^{2(n-t)} (t-1) \sigma_a^4 \leq \frac{n \sigma^4}{(1 - \lambda^2)} \leq \frac{n \sigma^4}{(1 - \lambda)}. \end{aligned}$$

The expression of $\text{var}(\sum_t \lambda^{n-t} Z_t^2)$ is given in (A.9), but is complicated. To prove (3.1) we recall $Z_t = O_p(\sqrt{t})$ so that $Z_t^2 = O_p(n)$ for $t \leq n$. Thus, $\sum_t \lambda^{n-t} Z_t^2 \leq n \sigma^4 / (1 - \lambda)$ in probability.

A.3 THE DISTRIBUTION OF EXPONENTIALLY WEIGHTED FUNCTIONALS

The first result we need to prove (3.3)–(3.4) is the extension of the so-called *functional* central limit theorem (CLT) to exponentially weighted sums of random variables.

Let

$$\tilde{a}_{[nr]} = \Lambda_{[nr]}^{-1} \sum_{t=1}^{[nr]} \lambda^{[nr]-t} a_t, \quad a_t \sim \text{iid}(0, \sigma^2), \tag{A.1}$$

where $0 \leq (r, \lambda) \leq 1$, $[\cdot] = \text{int}(\cdot)$ denotes the integer-value function and $\Lambda_n = (\sum_{t=1}^n \lambda^{n-t})$. Obviously, as $(r, \lambda) \rightarrow 1$ the above converges to the classical sample mean.

It should be noted that (A.1) uses only the first r th fraction of observations; in any event, even keeping $r < 1$ the central limit theorem still holds when $\lambda \rightarrow 1$:

$$\tilde{a}_{[nr]} = \sqrt{\Lambda_{[nr]}} \cdot \tilde{a}_{[nr]} \xrightarrow{L} N\left(0, \frac{1}{2}\sigma_a^2\right) \tag{A.2}$$

as $n \rightarrow \infty, \lambda \rightarrow 1$. The factor 1/2 in the variance of (A.2) can be explained by the fact that

$$\text{var}(\tilde{a}_{[nr]}) = \frac{\sum_{t=1}^{[nr]} \lambda^{2([nr]-t)} \sigma_a^2}{\sum_{t=1}^{[nr]} \lambda^{[nr]-t}} \xrightarrow{n} \frac{1/(1-\lambda^2)}{1/(1-\lambda)} \sigma_a^2 \xrightarrow{\lambda} \frac{1}{2} \sigma_a^2 \tag{A.3}$$

where, as in Section 3, we assume that the order of the limits cannot be inverted.

We now investigate if a similar property holds for stochastic functionals of r , such as $A_n(r) = \Lambda_n^{-1} \sum_{t=1}^{[nr]} \lambda^{[nr]-t} a_t$. For any given realization, this is a bounded step function defined on $j/n \leq r < (j+1)/n$ with $j = 0, 1 \dots n$; therefore

$$\sqrt{\Lambda_n} \cdot A_n(r) = \sqrt{\Lambda_{[nr]} \Lambda_n^{-1}} \cdot \tilde{a}_{[nr]} \xrightarrow{L} N\left(0, \frac{r}{2}\sigma_a^2\right) \tag{A.4}$$

as $n \rightarrow \infty, \lambda \rightarrow 1$ with $n(1-\lambda) \rightarrow 1$. This can be proved by using (A.2) and assuming $\lambda = (1 - 1/n)$; indeed

$$\Lambda_{[nr]} \Lambda_n^{-1} = \frac{\sum_{t=1}^{[nr]} \lambda^{[nr]-t}}{\sum_{t=1}^n \lambda^{n-t}} = \left[\sum_{t=1}^n \left(1 - \frac{1}{n}\right)^{n-t} \right]^{-1} \left[\sum_{t=1}^{[nr]} \left(1 - \frac{1}{[nr]}\right)^{[nr]-t} \right] \xrightarrow{n} r$$

which also holds by replacing the finite sum Λ_n by its limit $1/(1-\lambda)$.

The functional CLT for exponentially weighted sums can be obtained by noting that $B_n(r) = A_n(r)/[(1-\lambda)\sigma_a^2/2]^{1/2}$ tends to behave like a standard Brownian motion. Indeed for any $r_2 > r_1$ the statistics (13) are asymptotically independent, and (A.4) implies that

$$\left[B_n(r_2) - B_n(r_1) \right] = \frac{A_n(r_2) - A_n(r_1)}{\sqrt{(1-\lambda)\sigma_a^2/2}} \xrightarrow{L} N\left[0, (r_2 - r_1)\right] \tag{A.5}$$

as $n \rightarrow \infty, \lambda \rightarrow 1$ with $n(1-\lambda) \rightarrow 1$. By definition, a standard Brownian motion $W(\cdot)$ is a continuous-time process on $[0, 1]$ having independent Gaussian increments, therefore $B_n(\cdot) \rightarrow W(\cdot)$ in law.

As a conclusion, we may note that fundamental difference of the above with classical results is the factor 1/2 in the expressions (A.2)–(A.5). This does not mean that

exponential weighting allows for more efficiency; on the contrary, (A.1) has the same efficiency as the common n -sample mean whenever $(1 - \lambda)/(1 + \lambda) = 1/n$ (see (A.3) and (3.6)). Given $\lambda \ll 1$, this equivalence can be achieved only by processing a larger number of observations; for example, $\lambda = .95$ implies $n = 39$, but $.95^{39} \neq 0$ or $\sum_{t=1}^{39} .95^{(39-t)} \neq 1/(1 - .95)$.

A.4 APPLICATIONS OF THE CONTINUOUS MAPPING THEOREM

By viewing the weighted LS estimator (2.2) as a transformation of the statistic (A.1) the proofs of (3.3)–(3.4) can now be achieved by applying the so-called *continuous mapping theorem* (CMT) (e.g., Hall and Heyde 1980, p. 276) to (A.4)–(A.5). This theorem states that if $B_n(\cdot) \xrightarrow{L} W(\cdot)$ and $g(\cdot)$ is a continuous functional, then $g[B_n(\cdot)] \xrightarrow{L} g[W(\cdot)]$. Now, the key element of the application consists of recalling that the bounded process $z_{t,n} = Z_t/\sqrt{n}\sigma_a^2$, which corresponds to $B_n(t/n)$ in (A.5), converges in law to $W(r)$ as $t/n \rightarrow r$.

Let us start with

$$C_n(\lambda) = \frac{(1 - \lambda)}{\sqrt{n}} \sum_{t=1}^n \lambda^{n-t} Z_t = \frac{(1 - \lambda)}{\sqrt{n}} \sum_{t=1}^n \left(\sum_{j=1}^n \lambda^{n-j} \right) a_t \tag{A.6}$$

we are interested in analyzing the distribution of statistic (A.6) as $n \rightarrow \infty, \lambda \rightarrow 1$ at the same rate. This may be achieved by letting $\lambda = (1 - 1/n)$, which yields

$$C_n = \frac{1}{n} \sum_{t=1}^n \left(1 - \frac{1}{n} \right)^{n-t} \frac{Z_t}{\sqrt{n}} \xrightarrow{L} \frac{\sigma_a}{\sqrt{2}} \int_0^1 W(r) dr \tag{A.7}$$

as $n \rightarrow \infty$. Result (A.7) follows by noting that $Z_t/\sqrt{n} \rightarrow \sigma_a W(\cdot)$ and applying the CMT to the transformation $g(\cdot) = n^{-1} \sum_{t=1}^n (\cdot)$; moreover, it is known that the random variable $\int_0^1 W(r) dr \sim N(0, 1/3)$ (see Hamilton 1994 p. 485). Therefore, from (A.6) with $\lambda = (1 - 1/n)$ one may obtain

$$\text{var}(C_n) = \frac{1}{n^3} \sum_{t=1}^n \left[\sum_{j=1}^n \left(1 - \frac{1}{n} \right)^{n-j} \right]^2 \sigma_a^2 \xrightarrow{n} \frac{\sigma_a^2}{6} \tag{A.8}$$

which can easily be assessed by numerical evaluation.

Similarly, it can be shown that as $n \rightarrow \infty, \lambda \rightarrow 1$ with $n(1 - \lambda) \rightarrow 1$:

$$D_n(\lambda) = \frac{(1 - \lambda)}{n} \sum_{t=2}^n \lambda^{n-t} Z_{t-1}^2 \xrightarrow{L} \frac{\sigma_a^2}{\sqrt{2}} \int_0^1 W^2(r) dr \tag{A.9}$$

$$E_n(\lambda) = \frac{\sqrt{1 - \lambda}}{\sqrt{n}} \sum_{t=2}^n \lambda^{n-t} Z_{t-1} a_t \xrightarrow{L} \frac{\sigma_a^2}{\sqrt{2}} \int_0^1 W(r) dW(r). \tag{A.10}$$

In fact, application of the CMT yields $u_{t,n} = Z_{t-1}^2/n \rightarrow \sigma_a^2 W^2(\cdot)$ and $v_{t,n} = (Z_{t-1}/\sqrt{n})a_t \rightarrow \sigma_a^2 W(\cdot) dW(\cdot)$, because $a_t = (Z_t - Z_{t-1})$. Subsequently one can analyze the statistics $(1 - \lambda) \sum_{t=1}^n \lambda^{n-t} u_{t,n}$ and $\sqrt{1 - \lambda} \sum_{t=1}^n \lambda^{n-t} v_{t,n}$, with the same approach as in (A.6)–(A.7).

In particular, in the classical analysis it is well known that the sums $n^{-2} \sum_{t=2}^n Z_{t-1}^2$ and $n^{-1} \sum_{t=2}^n Z_{t-1} a_t$ converge to the same expressions as (A.9)–(A.10) but without the factor $1/\sqrt{2}$; the corresponding limit variances being $\sigma_a^4/3$ and $\sigma_a^4/2$, respectively, (see Fuller 1996, p. 367). Now, using the formula $\text{var}(\sum_t Z_t^2) = \sum_t \text{var}(Z_t^2) + 2 \sum_{i < j} \text{cov}(Z_i, Z_j)$ one can get

$$\text{var}\left(\sum_{t=2}^n \lambda^{n-t} Z_{t-1}^2\right) = \sum_{t=1}^{n-1} \lambda^{2(n-t)} (2t^2 \sigma_a^4 + t \mu_4) + 2 \sum_{j=1}^{n-2} \lambda^{2(n-1-j)} (2j^2 \sigma_a^4) (n-1-j). \quad (\text{A.11})$$

with $\mu_4 = E(a_t^4)$. Thus, proceeding as in (A.8) one can obtain $\text{var}(D_n) \rightarrow \sigma_a^4/6$, and similarly $\text{var}(E_n) \rightarrow \sigma_a^4/4$. These variances are half of the classical ones and motivate the factor $1/\sqrt{2}$ in expressions (A.9)–(A.10).

By expressing the above statistics in recursive form (i.e., letting $n = t$), from the ratio of (A.9) and (A.10) we prove the result (3.3) of Section 3. Moreover, using the same invariance principle one has

$$F_n(\lambda) = \left[\frac{(1-\lambda)}{n} \sum_{t=2}^n \lambda^{n-t} Z_{t-1}^2 \right]^{1/2} \xrightarrow{L} \left[\frac{\sigma_a^2}{\sqrt{2}} \int_0^1 W^2(r) dr \right]^{1/2} \quad (\text{A.12})$$

$n \rightarrow \infty, \lambda \rightarrow 1$ with $n(1-\lambda) \rightarrow 1$. Note that multiplying (3.2) by $F_t(\lambda)$, the quantity in (A.12) expressed in recursive form, one can obtain the statistic $\sqrt{R_t} \hat{\sigma}_a^{-1} [\hat{\phi}_t(\lambda) - 1]$. Hence, result (3.4) follows from the ratio of (A.10) and (A.12).

[Received June 1998. Revised February 1999.]

REFERENCES

- Box, G. E. P., and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control* (2nd ed.), San Francisco: Holden-Day.
- Dickey, D. A., and Fuller, W. A. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.
- Fan, J., and Gijbels, I. (1996), *Local Polynomial Modeling and its Applications*, New York: Chapman and Hall.
- Fuller, W. A. (1996), *Introduction to Statistical Time Series* (2nd ed.), New York: Wiley.
- Granger, C. W. J., and Swanson, N. R. (1996), "An Introduction to Stochastic Unit Root Processes," *Journal of Econometrics*, 74, 35–62.
- Grillenzoni, C. (1994), "Optimal Recursive Estimators of Dynamic Models," *Journal of the American Statistical Association*, 89, 777–787.
- (1996), "Testing for Causality in Real-Time," *Journal of Econometrics*, 73, 355–377.
- Hamilton, J. (1994), *Time Series Analysis*, Princeton, NJ: Princeton University Press.
- Hall, P., and Heyde, C. C. (1980), *Martingale Limit Theory and its Application*, New York: Academic Press.
- Härdle, W. (1991), *Smoothing Techniques with Implementation in S*, New York: Springer-Verlag.
- Hastie, T., and Loader, C. (1993), "Local Regression: Automatic Kernel Carpentry," *Statistical Science*, 8, 120–143.

- Ljung, L., and Soderstrom, T. (1983), *Theory and Practice of Recursive Identification*, Cambridge, MA: The MIT Press.
- Phillips, P. C. B., and Perron, P. (1988), "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.
- Rao, M. M. (1978), "Asymptotic Distribution of an Estimator of a Boundary Parameter of an Unstable Process." *The Annals of Statistics*, 16, 185–190.
- Tong, H. (1990), *Nonlinear Time Series Analysis: A Dynamical System Approach*, Oxford: Clarendon Press.
- White, J. S. (1958), "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case I," *Annals of Mathematical Statistics*, 29, 1188–1197.
- (1959), "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case II." *Annals of Mathematical Statistics*, 30, 831–834.