Unit Root Testing with Unstable Volatility

Brendan K. Beare*†

Nuffield College, University of Oxford

May 5, 2008

Abstract: It is known that unit root test statistics may not have the usual asymptotic properties when the variance of innovations is unstable. In particular, persistent changes in volatility can cause the size of unit root tests to differ from the nominal level. In this paper we propose a class of modified unit root test statistics that are robust to the presence of unstable volatility. The modification is achieved by purging heteroskedasticity from the data using a kernel estimate of volatility prior to the application of standard tests. In the absence of deterministic trend components, this approach delivers test statistics that achieve standard asymptotics under the null hypothesis of a unit root. When the data are homoskedastic, the local power of unit root tests is unchanged by our modification. We use Monte Carlo simulations to compare the finite sample performance of our modified tests with that of existing methods of correcting for unstable volatility.

JEL classifications: C14, C22.

Keywords and phrases: unit root, heteroskedasticity, nonstationary volatility.

*Address: Nuffield College, Oxford OX1-1NF, United Kingdom. Email: brendan.beare@nuffield.ox.ac.uk.

†This paper began life in 2004 as part of the solution to a take-home exam given by Peter Phillips in his course on time series econometrics at Yale University. Earlier versions of the paper were circulated in 2004 and 2005 under the title "Robustifying unit root tests to permanent changes in innovation variance." I thank Don Andrews, Juergen Gaul, Peter Phillips, Rob Taylor, and two anonymous referees for helpful comments, and the Cowles Foundation for financial support.

1 Introduction

In recent years, a number of researchers have devoted considerable attention to the problem of testing for a unit root in a time series whose innovations display unstable volatility. It is now known that the asymptotic properties of unit root test statistics under unstable volatility may differ from the properties that obtain under homoskedasticity. Thus, if unstable volatility is not properly taken into account, spurious inference about the presence or absence of a unit root may result.

Hamori and Tokihisa (1997) appear to be the first to observe that a simple level shift in innovation variance partway through a sample invalidates the usual unit root asymptotics. Subsequently, Kim *et. al.* (2002) showed that ignoring the effect of a volatility shift could result in tests with asymptotic rejection rates of as high as 40% under the null hypothesis of a unit root, using a nominal size of 5%. Cavaliere (2004b) studied the properties of unit root tests under a more flexible model of volatility that could incorporate smooth changes in variance as well as multiple level shifts, while Boswijk (2001) studied the case of near-integrated stochastic volatility. Related work by Dahl and Levine (2006), Phillips and Xu (2006) and Xu and Phillips (2008) has considered the impact of unstable volatility in stationary autoregressions.

The main contribution of this paper is the development of a new approach to unit root testing that is robust to the presence of heteroskedasticity of quite general form. The basic idea is to scale the increments of a time series using a nonparametric estimate of their variance, rendering them approximately homoskedastic. Standard unit root tests may then be applied to the rescaled data. We show that, under suitable conditions including the absence of a deterministic trend component, the null asymptotic properties of tests constructed in this fashion are the same as one would expect under homoskedasticity. Moreover, applying this approach in a homoskedastic context leads to no loss of local power.

Several other methods of testing for a unit root in the presence of unstable volatility have been proposed in recent years. Cavaliere and Taylor (2007ab, 2008ab) and Boswijk (2005) have proposed a total of four testing procedures. We review those approaches here, and compare them to each other and to our own approach using small sample simulations.

The paper is organized as follows. Section 2 introduces the heteroskedastic unit root model, and in Section 3 we review the existing literature on unit root testing under unstable volatility. A new approach to unit root testing under unstable volatility is developed in Section 4. We report our small sample simulation results in Section 5, and conclude in Section 6. Proofs of lemmas and theorems are collected in Appendix A, and tables in Appendix B.

2 The Heteroskedastic Model

The econometric model we consider is very similar to that of Cavaliere (2004b), Boswijk (2005), and Cavaliere and Taylor (2007ab, 2008ab), although some of our technical conditions differ. Let $\{\varepsilon_t : t \in \mathbb{N}\}$ be a stationary sequence of random variables with zero mean and unit variance, and define the triangular array of random variables $\{y_{t,n} : 1 \leq t \leq n, n \in \mathbb{N}\}$ by

$$\begin{aligned} y_{t,n} &= z_t + x_{t,n} & t &= 0, \ldots, n \\ x_{t,n} &= \phi_n x_{t-1,n} + \omega\left(\tfrac{t}{n}\right)\varepsilon_t & t &= 1, \ldots, n \\ x_{0,n} &= O_p(1). \end{aligned}$$

z_t is a deterministic process of some form; cases of leading interest include $z_t = 0$, $z_t = \alpha$, and $z_t = \alpha + \beta t$. $y_{t,n}$ is thus the sum of a deterministic process z_t and an autoregressive process $x_{t,n}$. The variance of the innovations of $x_{t,n}$ is governed by the function $\omega : [0,1] \to \mathbb{R}_+$.

Our focus in this paper is on the sequence of autoregressive coefficients ϕ_n. In general we will set $\phi_n = \exp(-c/n)$, where $c \geq 0$. When $c = 0$, $y_{t,n}$ is a unit root process, while when $c > 0$, $y_{t,n}$ is a near-unit root process. We seek to construct a test of the unit root hypothesis that is well behaved when ω is nonconstant.

The volatility function ω is assumed to satisfy the following condition.

Assumption 2.1 *ω is twice continuously differentiable, and strictly positive.*

Assumption 2.1 is stronger than the comparable assumptions imposed by Cavaliere and Taylor (2007ab, 2008ab), who allow ω to be discontinuous, and by Boswijk (2005), who assumes continuity but not differentiability. Our need for stronger smoothness conditions on ω stems from our use of a kernel estimator for ω in the construction of our heteroskedasticity-robust unit root tests in Section 4 below. Assumption 2.1 is such that uniformly consistent estimation of ω is possible, and such that the derivative of our estimator is also well behaved. Nevertheless, the assumption of continuity is not critical for our tests to be well behaved, as will be clear from the small sample simulations reported in Section 5.

An implicit assumption on ω is that it is nonstochastic. Cavaliere and Taylor (2007a, 2008ab) share this assumption, while Boswijk (2005) and Cavaliere and Taylor (2007b) allow ω to be stochastic. Allowing ω to be stochastic in this paper would seem to be of rather limited benefit, as our assumption that ω is twice continuously differentiable rules out the majority of stochastic processes used to model volatility in practice. Nevertheless, in the small sample simulations reported in Section 5, we will see that the performance of the heteroskedasticity-robust tests introduced in this paper seems comparable to the performance of existing heteroskedasticity-robust tests when ω is the exponential of a Brownian motion or Ornstein-Uhlenbeck process.

The sequence of innovations $\{\varepsilon_t\}$ is assumed to satisfy the following condition.

Assumption 2.2 *There exists $p > 4$ such that $E|\varepsilon_0|^p < \infty$, and such that the α-mixing coefficients of $\{\varepsilon_t\}$ satisfy $\sum_{j=1}^{\infty} \alpha_j^{2(1/r-1/p)} < \infty$ for some $r \in [2p/(p-2), 4]$. The long-run variance $\lambda^2 = \sum_{j=-\infty}^{\infty} E(\varepsilon_t \varepsilon_{t+j})$ satisfies $\lambda^2 > 0$.*

Assumption 2.2 allows for very general forms of serial dependence in ε_t, but rules out processes that are strongly dependent or very fat tailed. The role of the quantity r will become clear in Assumption 4.4 below. Cavaliere and Taylor (2007ab, 2008ab) make comparable assumptions in the framework of linear processes, while Boswijk (2005) assumes the innovations ε_t to be independent of one another. Note that the mixing condition in Assumption 2.2 implies that $\lambda^2 < \infty$.

3 Unit Root Testing under Heteroskedasticity

In this section we survey the existing literature on unit root testing in the context of the heteroskedastic model described in the previous section. We begin with the fundamental results of

Cavaliere (2004b) concerning the effect of heteroskedasticity on the behavior of standard unit root tests. Next, we discuss a class of tests proposed by Cavaliere and Taylor (2008a), which achieve standard unit root asymptotics by applying a nonparametrically estimated time-transformation to the data prior to testing. Finally, we discuss several unit root tests that use simulated critical values to adjust for the effect of heteroskedasticity - these are the tests of Boswijk (2005) and Cavaliere and Taylor (2007ab, 2008b). The relative merits of all these tests, and of the new tests we shall propose in Section 4, will be studied in the small sample simulations in Section 5.

3.1 Standard Tests

The properties of unit root tests when applied to the model described in Section 2 were studied in detail by Cavaliere (2004b). Suppose we impose the unit root hypothesis by setting $\phi_n = 1$. Cavaliere showed, under conditions somewhat more general than Assumptions 2.1 and 2.2 above, that

$$S_n(\cdot) := \frac{1}{\sqrt{n}} \sum_{t=1}^{[n\cdot]} \omega\left(\frac{t}{n}\right) \varepsilon_t \Rightarrow \lambda B_\omega(\cdot) \tag{3.1}$$

as $n \to \infty$, where

$$B_\omega(r) := \int_0^r \omega(s) \mathrm{d}B(s),$$

and B is a standard Brownian motion. The symbol "$\Rightarrow$" denotes weak convergence in the usual sense. This is Lemma 1 of Cavaliere (2004b). Lemma 3 of Cavaliere is relevant to the more general case where $\phi_n = \exp(-c/n)$ with $c \geq 0$; in this case, we have

$$S_n^c(\cdot) := \frac{1}{\sqrt{n}} \sum_{t=1}^{[n\cdot]} \exp\left(-c\frac{[n\cdot] - t}{n}\right) \omega\left(\frac{t}{n}\right) \varepsilon_t \Rightarrow \lambda J_\omega^c(\cdot) \tag{3.2}$$

as $n \to \infty$, where

$$J_\omega^c(r) := \int_0^r \exp(-c(r-s))\omega(s)\mathrm{d}B(s) = \int_0^r \exp(-c(r-s))\mathrm{d}B_\omega(s).$$

(3.1) and (3.2) generalize the well-known results in Phillips (1987ab), which obtain when ω is constant.

Using (3.1) and (3.2), it is relatively straightforward to see how the asymptotic behavior of standard unit root test statistics is affected when ω is nonconstant. Suppose we set $z_t = 0$, and let Z_n denote the coefficient-based test statistic proposed by Phillips (1987a) to test the unit root hypothesis in the absence of deterministic components. In this case, it can be shown under suitable technical conditions that

$$\begin{aligned} Z_n \quad &\to_d \quad \frac{\frac{1}{2}\left(J_\omega^c(1)^2 - J_\omega^c(0)^2 - 1\right)}{\int_0^1 J_\omega^c(r)^2 \mathrm{d}r} \\ &= \quad \frac{\frac{1}{2}\left(J_\omega^c(1)^2 - 1\right)}{\int_0^1 J_\omega^c(r)^2 \mathrm{d}r} \end{aligned} \tag{3.3}$$

as $n \to \infty$. This is part of Theorems 1 and 3 in Cavaliere (2004b). When $\phi_n = 1$, the limiting distribution in (3.3) in general differs from the Dickey-Fuller distribution that obtains when ω is constant, and so the actual size of tests based on this statistic may differ from the nominal size if the usual Dickey-Fuller critical values are used. Analogues to (3.3) can be demonstrated for various other unit root tests; see Cavaliere (2004b) and Cavaliere and Taylor (2007a). The general principle is that the Ornstein-Uhlenbeck process J^c appearing in the limiting distribution of standard test statistics is replaced by the scaled process J_ω^c. When our model for y_t includes deterministic components, the limiting distribution of test statistics that allow for deterministic components is typically obtained by replacing J_ω^c with the residual from a Hilbert projection of J_ω^c onto the space spanned by those components. The projection may be orthogonal, as in Phillips and Perron (1988), or non-orthogonal, as in Schmidt and Phillips (1992), Phillips and Lee (1996) or Elliott *et al.* (1996); see also the discussion in Phillips and Xiao (1998), Theorem 2 in Cavaliere (2004b), and Theorem 1 in Cavaliere and Taylor (2008b). As an example, the limiting distribution of a Phillips-Perron coefficient-based test statistic using demeaned data is given by replacing $J_\omega^c(\cdot)$ with $J_\omega^c(\cdot) - \int_0^1 J_\omega^c(r)\mathrm{d}r$ in (3.3).

Our discussion in this subsection has focused mostly on results due to Cavaliere (2004b). Earlier work on this subject includes Hamori and Tokihisa (1997), Kim *et al.* (2002), and Cavaliere (2004a). These authors considered the more restrictive case where ω is constant except at a single point of discontinuity. The results in Cavaliere (2004b) allow ω to be a strictly positive cadlag process exhibiting a finite number of points of discontinuity, and satisfying a uniform first-order Lipschitz condition between the discontinuities. This condition is evidently more general than our Assumption 2.1. Other related work includes Boswijk (2001), who studied the behavior of unit root tests when ω is a near-integrated stochastic process.

3.2 Time-Transformed Tests

In view of the results described in the previous subsection, it would clearly be of interest to develop new unit root tests whose null asymptotic behavior is unaffected when ω is allowed to be nonconstant. A class of tests with this property was proposed recently by Cavaliere and Taylor (2008a). The tests operate by applying a nonparametrically estimated time-transformation to the data, prior to the application of standard unit root tests. The motivation for this approach derives from Lemma 2 in Cavaliere (2004b), which states that $B_\omega(\cdot)$ is distributionally equivalent to $B(\eta(\cdot))$, where

$$\eta(r) := \frac{\int_0^r \omega(s)^2 \mathrm{d}s}{\int_0^1 \omega(s)^2 \mathrm{d}s}.$$

Note that η is proportional to the quadratic variation of B_ω.

The construction of the time-transformed test statistics proceeds in the following manner. Let us assume for simplicity that $z_t = 0$, so that there are no deterministic components in the model. One begins by forming the following nonparametric estimate of η:

$$\hat{\eta}_n(r) := \frac{\sum_{t=1}^{[nr]} \left(\Delta x_{t,n}\right)^2 - (nr - [nr]) \left(\Delta x_{[nr]+1,n}\right)^2}{\sum_{t=1}^{n} \left(\Delta x_{t,n}\right)^2},$$

where $\Delta x_{t,n} := x_{t,n} - x_{t-1,n}$. Minor modifications to this estimator are possible; see Cavaliere and Taylor (2008a) for details. Assuming that $\hat{\eta}_n$ is strictly increasing, it has a unique inverse, which we denote by $\hat{g}_n$, and we let g denote the inverse of η. One next forms the time-transformed series $x^\dagger_{t,n} := x_{[\hat{g}_n(t/n)n],n}$, $t = 0, \ldots, n$. Finally, a standard unit root test is applied directly to

$x^{\dagger}_{t,n}$. Cavaliere and Taylor (2008a) consider the application of the so called "M unit root tests" originally proposed by Perron and Ng (1996), Stock (1999) and Ng and Perron (2001), but it is clear that a much broader range of tests could be applied in this fashion. In general, the null limiting distribution of the test statistics will be the same as one would expect when ω is constant.

A heuristic explanation for the effectiveness of the time-transformation method runs as follows. Define $S^{\dagger}_n(\cdot) := n^{-1/2} x^{\dagger}_{[n\cdot],n}$. When $\phi_n = 1$, we have $S^{\dagger}_n(\cdot) = S_n\left(\hat{g}_n([n\cdot]/n)\right)$. Now, $S_n(\cdot) \Rightarrow \lambda B_\omega(\cdot)$ by (3.1), and it is shown by Cavaliere and Taylor (2008a) that, under suitable technical conditions, $\hat{g}_n$ is uniformly consistent for g. Thus, $S^{\dagger}_n(\cdot) \Rightarrow \lambda B_\omega(g(\cdot))$. Since $B_\omega(\cdot)$ is distributionally equivalent to $B(\eta(\cdot))$, we now have $S^{\dagger}_n(\cdot) \Rightarrow \lambda B(\eta(g(\cdot))) = \lambda B(\cdot)$, and so the time-transformed unit root process $x^{\dagger}_{t,n}$ satisfies a functional central limit theorem of the usual form.

It is possible to adapt the time-transformation method so that it can be applied when our model contains deterministic components; refer to Cavaliere and Taylor (2008a) for details. When $z_t = \alpha$, so that only a constant term is included, the time-transformation method continues to yield test statistics with null limiting distributions that are invariant to ω. Unfortunately, if z_t includes a time trend, the null limiting distributions are no longer pivotal and depend on ω. This property is shared by the new class of unit root tests we shall introduce in Section 4. Suitable critical values for the new tests and for the time transformed tests when z_t includes a time trend may still be achieved by simulation conditional on $\hat{g}_n$ (in the case of the time-transformed tests) or on a nonparametric estimate of ω (in the case of the tests proposed here).

3.3 Tests Using Simulated Critical Values

The motivation for the time-transformation method discussed in the previous subsection was to construct test statistics that achieve standard unit root asymptotics under the null hypothesis, even when ω is nonconstant. An alternative approach is to construct critical values that may be used validly with unit root test statistics whose null limiting distributions are not invariant to ω. This approach is used by Boswijk (2005) and by Cavaliere and Taylor (2007ab, 2008b).

Cavaliere and Taylor (2007a) propose to use the estimated time deformation $\hat{\eta}_n$ to simulate valid critical values for standard unit root tests. In particular, they consider the M tests mentioned in the previous subsection, but it is clear that their approach can be applied to a broader range of test statistics. For instance, in view of (3.3), critical values for the Phillips (1987a) coefficient-based test statistic in the absence of deterministic components can be extracted from the quantiles of the random variable

$$\frac{\frac{1}{2}\left(B(1)^2-1\right)}{\int_0^1 B(\hat{\eta}_n(r))^2\mathrm{d}r},$$

which can be simulated conditional on $\hat{\eta}_n = \eta$ by drawing realizations of B. In view of the uniform consistency of $\hat{\eta}_n$ and the distributional equivalence of $B_\omega(\cdot)$ and $B(\eta(\cdot))$ (note also that $\hat{\eta}_n(1) = 1$), critical values obtained in this fashion lead to tests based on standard statistics that are of the correct asymptotic size in the presence of nonconstant ω.

Cavaliere and Taylor (2007b, 2008b) consider an alternative approach to simulation involving the wild bootstrap. Again, their focus is on obtaining critical values for the M tests, but the approach is more broadly applicable. In the case of the Phillips (1987a) coefficient-based test absent deterministic components, one begins by forming a bootstrap sample of the form

$$x_{t,n}^b = \sum_{s=1}^{t} \hat{u}_{t,n} v_t^b$$

for $t = 1, \ldots, n$, where $\hat{u}_{t,n}$ are the residuals from a regression of $x_{t,n}$ on $x_{t-1,n}$, and v_t^b are an independent sequence of standard normal random variables. The initial bootstrap observation $x_{0,n}^b$ may be set equal to zero. One then computes the bootstrap test statistic Z_n^b according to

$$Z_n^b = n\left(\frac{\sum_{t=1}^n x_{t-1,n}^b x_{t,n}^b}{\sum_{t=1}^n \left(x_{t-1,n}^b\right)^2} - 1\right).$$

Note that there is no need to include a correction term for serial dependence in the bootstrap statistic, as the increments in the bootstrap sample are uncorrelated by construction. Asymptotically valid critical values for the test statistic Z_n may be extracted from the quantiles of Z_n^b, computed by drawing a suitably large number of bootstrap samples.

Boswijk (2005) proposes to test the unit root hypothesis using a new statistic based on the principle of maximum likelihood estimation. When the innovations are assumed to be Gaussian, the statistic is formed from a feasible generalized least squares (GLS) estimator of ϕ_n. In the absence of deterministic components, and assuming independence of the innovations ε_t, Boswijk's estimator for ϕ_n is given by

$$\hat{\phi}_n^{\mathrm{B}} = \frac{\sum_{t=1}^{n} \tilde{\omega}_n \left(\frac{t}{n}\right)^{-2} x_{t-1,n} x_{t,n}}{\sum_{t=1}^{n} \tilde{\omega}_n \left(\frac{t}{n}\right)^{-2} x_{t-1,n}^2}.$$

Here, $\tilde{\omega}_n(t/n)$ is a nonparametric estimate of $\omega(t/n)$ constructed from a one-sided kernel-weighted average of the squared increments $\Delta x_{s,n}^2$, $s = 0, \ldots, t-1$. Alternatively, estimation may be based on the squared residuals from a least squares regression of $x_{t,n}$ on $x_{t-1,n}$. Drawing on results due to Hansen (1995), Boswijk shows that, under suitable technical conditions, $\tilde{\omega}_n$ is uniformly consistent, and the feasible GLS estimator is adaptive in the sense that it achieves the same asymptotic distribution as the true GLS estimator. That is, we have

$$n\left(\hat{\phi}_n^{\mathrm{B}} - 1\right) \Rightarrow \frac{\int_0^1 V_\omega^c(r) \mathrm{d}B(r)}{\int_0^1 V_\omega^c(r)^2 \mathrm{d}r} - c, \tag{3.4}$$

where

$$V_\omega^c(r) = \int_0^r \exp(-c(r-s)) \frac{\omega(s)}{\omega(r)} \mathrm{d}B(s).$$

When $c = 0$, so that $\phi_n = 1$, the limiting distribution in (3.4) depends on ω. Asymptotically valid critical values for testing $\phi_n = 1$ can be obtained by simulating the limiting distribution in (3.4) conditional on $\tilde{\omega}_n = \omega$ and $c = 0$.

Finally, we note that Kim *et. al.* (2002) also proposed a method of testing for a unit root in the presence of heteroskedasticity, which involved simulating critical values conditional on an estimate of ω. Their approach required ω to be constant aside from a single point of discontinuity. We confine our discussion here to tests that are applicable when ω is allowed to vary in a more flexible manner.

4 A New Class of Heteroskedasticity-Robust Tests

In this section we propose a new approach to constructing unit root tests that are robust to the presence of heteroskedasticity. Our approach is similar in spirit to the time-transformation tests of Cavaliere and Taylor (2008a), in that we seek to transform the data in such a way that the null limiting distributions of standard unit root test statistics are the same as would be obtained under homoskedasticity. Rather than apply a time-transformation to the data, we will seek to rescale the increments of $x_{t,n}$ in such a way that they become approximately homoskedastic.

4.1 Construction of the Tests

We shall assume in this section that $z_t = \alpha$, so that our model potentially contains an intercept, but no trend term. The extension to deterministic time trends will be considered in Subsection 4.3. Let $u_{t,n} := \omega(t/n)\varepsilon_t$, and suppose we have a series of estimated residuals $\hat{u}_{t,n}$ satisfying

$$\sup_{1 \leq t \leq n} |\hat{u}_{t,n} - u_{t,n}| = O_p(n^{-1/2}) \tag{4.1}$$

when $\phi_n = \exp(-c/n)$, $c \geq 0$. Estimated residuals satisfying (4.1) are rather easy to obtain. One convenient possibility is to set $\hat{u}_{t,n} = y_{t,n} - y_{t-1,n}$. Another is to let $\hat{u}_{t,n}$ be the residuals from a regression of $y_{t,n}$ on $y_{t-1,n}$ and a constant. Both choices can easily be shown to satisfy (4.1) using (3.2).

We can obtain a nonparametric estimate of ω by applying the Nadaraya-Watson estimator to the squares of our estimated residuals $\hat{u}_{t,n}$. For $r \in [0,1]$, define

$$\hat{\omega}_n(r) = \left(\frac{\sum_{t=1}^{n} k\left(\frac{t/n - r}{h_n}\right) \hat{u}_{t,n}^2}{\sum_{t=1}^{n} k\left(\frac{t/n - r}{h_n}\right)} \right)^{1/2}. \tag{4.2}$$

Here, $k : \mathbb{R} \to \mathbb{R}$ is a kernel function and h_n is a suitably chosen bandwidth. We impose the following condition on k.

Assumption 4.1 *k is continuously differentiable and satisfies $\int k(x)dx > 0$, $\int |xk(x)|dx < \infty$,* *and* $\int |xk'(x)|dx < \infty$. *The Fourier transform of* k, *denoted* ψ, *satisfies* $\int |x\psi(x)|dx < \infty$.

Assumption 4.1 is satisfied by, for instance, the Gaussian kernel. Note that we do not require k to be one-sided, as does Boswijk (2005). The bandwidth h_n is assumed to satisfy the following condition.

Assumption 4.2 $h_n \to 0$ *and* $nh_n^4 \to \infty$ *as* $n \to \infty$.

Assumption 4.2 is quite strong, and forces the bandwidth h_n to shrink at a rate slower than $n^{-1/4}$. We use this condition to ensure that the derivative of $\hat{\omega}_n$ is well behaved. Results pertaining to the asymptotic behavior of the derivatives of nonparametric function estimates typically require fairly stringent bandwidth conditions; see e.g. Andrews (1995). In applications, h_n could be chosen by cross-validation (Wong, 1983), or by modified cross-validation (Chu and Marron, 1991) in the case of serially dependent innovations. Visual inspection of $\hat{\omega}_n$ for various choices of h_n may also be a pragmatic way to proceed in practice.

The estimated function $\hat{\omega}_n$ can be used to construct a modified version of $y_{t,n}$ that has approximately homoskedastic differences. Define

$$\begin{aligned} y_{t,n}^* &= \sum_{s=1}^{t} \frac{y_{s,n} - y_{s-1,n}}{\hat{\omega}_n\left(\frac{s}{n}\right)}, \quad t = 1, \ldots, n \\ y_{0,n}^* &= 0. \end{aligned}$$

We propose to test for a unit root by applying a standard unit root test to the rescaled series $y_{t,n}^*$. For instance, we could apply a Phillips-Perron coefficient-based test to $y_{t,n}^*$. Our test statistic would then be

$$Z_n^* = n\left(\hat{\phi}_n^* - 1\right) - \frac{\frac{1}{2}\left(\hat{\lambda}_n^{*2} - \hat{\sigma}_n^{*2}\right)}{\frac{1}{n^2}\sum_{t=1}^{n} y_{t-1,n}^{*2}}, \tag{4.3}$$

where $\hat{\phi}_n^*$ is the estimated slope coefficient in a least squares regression of $y_{t,n}^*$ on $y_{t-1,n}^*$ and a constant, $\hat{\varepsilon}_{t,n}$ are the residuals from that regression, and

$$\hat{\sigma}_n^{*2} = \frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_{t,n}^2,$$
$$\hat{\lambda}_n^{*2} = \sum_{j=-n+1}^{n-1}\kappa\left(\frac{j}{m_n}\right)\left(\frac{1}{n}\sum_{t=|j|+1}^{n}\hat{\varepsilon}_{t,n}\hat{\varepsilon}_{t-|j|,n}\right).$$

Here, $\kappa : \mathbb{R} \to \mathbb{R}$ is another kernel function, and m_n is another bandwidth parameter. We introduce two further technical conditions.

Assumption 4.3 *κ is symmetric and satisfies $|\kappa| \leq 1$ and $\kappa(0) = 1$, and is continuous at zero and almost everywhere else. Further, there exists a nonincreasing function $l : \mathbb{R} \to \mathbb{R}$ such that $|\kappa| \leq |l|$ and $\int_{-\infty}^{\infty} |xl(x)|dx < \infty$.*

Assumption 4.4 *$m_n \to \infty$ as $n \to \infty$, and $m_n = o(n^{1/2-1/r})$.*

Assumptions 4.3 and 4.4 are essentially equivalent to Assumption $\mathcal{K}$ of Cavaliere (2004b), taken from de Jong (2000), who uses the conditions to demonstrate the consistency of long run variance estimators. Note that the quantity r must be chosen to satisfy Assumption 2.2 above. In practical applications, the choices of κ and m_n could be guided by the discussion in Andrews (1991).

4.2 Asymptotic Properties

In this subsection we discuss the asymptotic properties of unit root test statistics constructed from the rescaled series $y_{t,n}^*$. We will show that tests based on $y_{t,n}^*$ will in general achieve standard null asymptotics; that is, the asymptotics that would obtain if the statistics were computed using a directly observed, homoskedastic unit root process. Assumptions 2.1, 2.2, 4.1, 4.2, 4.3 and 4.4 are maintained throughout.

Let us first consider the behavior of our nonparametric volatility estimator $\hat{\omega}_n$. Our first preliminary result states that $\hat{\omega}_n$ is a uniformly consistent estimator of ω.

Lemma 4.1 *$\hat{\omega}_n$ satisfies*

$$\sup_{r\in[0,1]} |\hat{\omega}_n(r) - \omega(r)| = o_p(1).$$

The proof of Lemma 4.1, and of all other results in this subsection and the sequel, may be found in Appendix A. The proof of Lemma 4.1 uses mostly standard techniques, as in e.g. Pagan and Ullah (1999). Continuity of ω is critical to obtaining uniform consistency.

We next consider the asymptotic behavior of $\hat{\omega}_n'$, the derivative of $\hat{\omega}_n$. Things here are complicated by the fact that $\hat{\omega}_n'(r)$ is not consistent for $\omega'(r)$ at the endpoints $r = 0$ and $r = 1$. (We define $\omega'(0)$ and $\omega'(1)$ as the one-sided limits of $\omega'(r)$ at $r = 0$ and $r = 1$ respectively.) Our next result concerns the behavior of $\hat{\omega}_n'$ over a subinterval that is trimmed away from the endpoints $r = 0$ and $r = 1$, but with the amount of trimming vanishing asymptotically.

Lemma 4.2 *Let τ_n be a sequence of small positive numbers satisfying $\tau_n \to 0$ and $h_n^{-1}\tau_n \to \infty$. Then $\hat{\omega}_n'$ satisfies*

$$\sup_{r \in [\tau_n, 1-\tau_n]} \left|\hat{\omega}_n'(r) - \omega'(r)\right| = o_p(1).$$

In terms of the uniform behavior of $\hat{\omega}_n'$ over $[0, 1]$, the best we can do is the following.

Lemma 4.3 *$\hat{\omega}_n'$ satisfies*

$$\sup_{r \in [0,1]} \left|\hat{\omega}_n'(r) - \omega'(r)\right| = O_p(1).$$

With Lemmas 4.1-4.3 in hand, we are in a position to develop our asymptotic theory for statistics based on $y_{t,n}^*$. For $c \geq 0$, define the stochastic process $\xi_\omega^c(r)$, $r \in [0, 1]$, by

$$\xi_\omega^c(r) := B(r) - c\int_0^r \omega(s)^{-1} J_\omega^c(s)\mathrm{d}s. \tag{4.4}$$

When $c = 0$, we obviously have $\xi_\omega^c = B$, a standard Brownian motion. When ω is constant, we have the following somewhat less obvious result.

Theorem 4.1 *When ω is constant, $\xi_\omega^c(r) = J^c(r) := \int_0^r \exp(-c(r-s))dB(s)$.*

Theorem 4.1 states that, under homoskedasticity, our process ξ_ω^c reduces to the Ornstein-Uhlenbeck process J^c. It is proved using the Fubini-Tonelli theorem and the well-known representation $J^c(r) = B(r) - c\int_0^r \exp(-c(r-s))B(s)\mathrm{d}s$, a consequence of Ito's lemma.

Our main result is as follows.

Theorem 4.2 *As $n \to \infty$, we have $\hat{\sigma}_n^* \to_p 1$, $\hat{\lambda}_n^* \to_p \lambda$, and $n^{-1/2}y_{[nr],n}^* \Rightarrow \lambda\xi_\omega^c(r)$.*

Theorem 4.2 establishes consistency of the estimators $\hat{\sigma}_n^*$ and $\hat{\lambda}_n^*$, and provides a version of the functional central limit theorem for $y_{t,n}^*$. It is sufficient to establish the asymptotic behavior of a variety of unit root test statistics when applied to $y_{t,n}^*$. For instance, it follows easily from Theorem 4.2 and the continuous mapping theorem that the rescaled Phillips-Perron test statistic Z_n^* defined in (4.3) satisfies

$$Z_n^* \to_d \frac{\frac{1}{2}\left(\tilde{\xi}_\omega^c(1)^2 - \tilde{\xi}_\omega^c(0)^2 - 1\right)}{\int_0^1 \tilde{\xi}_\omega^c(r)^2\mathrm{d}r}, \tag{4.5}$$

where $\tilde{\xi}_\omega^c$ is the demeaned process given by

$$\tilde{\xi}_\omega^c(r) := \xi_\omega^c(r) - \int_0^1 \xi_\omega^c(s)\mathrm{d}s.$$

Since $\xi_\omega^c = B$ when $c = 0$, (4.5) implies that Z_n^* achieves the usual Dickey-Fuller asymptotics when the null hypothesis of a unit root is true. Furthermore, Theorem 4.1 implies that, when ω is constant, the process $\tilde{\xi}_\omega^c(r)$ in (4.5) may be replaced with $\tilde{J}^c(r) := J^c(r) - \int_0^1 J^c(s)\mathrm{d}s$, a demeaned Ornstein-Uhlenbeck process. This yields an important property of Z_n^*: no local power is lost by applying the Phillips-Perron test to $y_{t,n}^*$ instead of to $y_{t,n}$ when ω is in fact constant. The other heteroskedasticity-robust tests described in Section 3 also share this property.

When ω is not constant, and $c > 0$, the limiting distribution of Z_n^* will in general be different to that of the unmodified Phillips-Perron test statistic. The relative local powers of the two tests will depend on the form of ω. Some small sample evidence on local power under various choices of ω will be reported in the following section.

The proof of Theorem 4.2 draws on Lemmas 4.1-4.3, which require strict smoothness conditions on ω to be true. In fact, under some additional technical conditions, it is possible to show that

Theorem 4.2 continues to hold when ω and ω' are subject to a finite number of discontinuities. The trick here is to modify Lemmas 4.1 and 4.2 so that the supremum over r is taken over a smaller set $A_n \subset [0,1]$ that is "trimmed" away from the discontinuities in ω and ω'. By allowing the trimming to vanish at a suitable rate, one can show that Theorem 4.2 continues to hold. A proof along these lines is substantially more complicated than the proof under our more strict smoothness conditions, in rather uninteresting ways, so we do not pursue it here. The small sample simulations reported in the following section provide ample evidence that the testing procedure proposed here is effective when ω is discontinuous.

4.3 Extension to Deterministic Time Trends

In this subsection we discuss the construction of unit root test statistics based on rescaled data when $z_t = \alpha + \beta t$, so that our model for $y_{t,n}$ includes both a constant term and a linear time trend. Insofar as the estimation of ω and ω' is concerned, the presence of a trend raises no new challenges. Estimated residuals $\hat{u}_{t,n}$ satisfying (4.1) are easy to obtain. For instance, adopting the approach to detrending taken by Schmidt and Phillips (1992), which is shown by Phillips and Lee (1996) to be asymptotically efficient when $c = 0$, we may set $\hat{u}_{t,n} = y_{t,n} - y_{t-1,n} - n^{-1}(y_{n,n} - y_{0,n})$. $\hat{\omega}_n$ can then be constructed using $\hat{u}_{t,n}$ as before, and Lemmas 4.1-4.3 follow in the same way. The rescaled data $y^*_{t,n}$ are constructed according to

$$
\begin{aligned}
y^*_{t,n} &= \sum_{s=1}^{t} \frac{y_{s,n} - y_{s-1,n} - \frac{1}{n}(y_{n,n} - y_{0,n})}{\hat{\omega}_n\left(\frac{s}{n}\right)}, \quad t = 1, \ldots, n \qquad (4.6) \\
y^*_{0,n} &= 0.
\end{aligned}
$$

The following result provides an analogue to Theorem 4.2 for the case of trending data.

Theorem 4.3 *As $n \to \infty$, $\hat{\sigma}^*_n \to_p 1$, $\hat{\lambda}^*_n \to_p \lambda$, and $n^{-1/2} y^*_{[nr],n} \Rightarrow \lambda \xi^c_\omega(r) - \lambda J^c_\omega(1) \int_0^r \omega(s)^{-1} ds$.*

Unfortunately, Theorem 4.3 makes it clear that we cannot expect unit root test statistics computed using $y^*_{t,n}$ to achieve standard Dickey-Fuller null asymptotics when we include a trend term in our model. Specifically, when $c = 0$, we have

$$\frac{1}{\sqrt{n}} y^*_{[nr],n} \Rightarrow \lambda B(r) - \lambda B_\omega(1) \int_0^r \omega(s)^{-1} \mathrm{d}s,$$

so that the partial sums of $y^*_{t,n}$ do not behave like those of a homoskedastic unit root process in general. The exception is when ω is constant, in which case $n^{-1/2} y^*_{[nr],n}$ converges weakly to a Brownian bridge scaled by λ, and standard null asymptotics may be applied. In this respect, our testing procedure is similar to the time-transformation procedure of Cavaliere and Taylor (2008a), which yields null asymptotics that are invariant to ω when the model includes only a constant, but dependent upon ω when the model includes a trend. In principle, we could still construct valid tests for a unit root using $y^*_{t,n}$ by simulating critical values conditional on our estimated volatility function $\hat{\omega}_n$, but the motivation for our procedure seems somewhat weak in this case.

Theorem 4.3 is specific to the method of detrending employed in our construction of $y^*_{t,n}$; that is, the method used by Schmidt and Phillips (1992). Analogous results that apply under different methods of detrending are easy to establish. In slightly loose terms, if we replace the term $n^{-1}(y_{n,n} - y_{0,n})$ in (4.6) with some other slope estimator $\hat{\beta}_n$, then Theorem 4.3 continues to hold, with the limiting distribution of $n^{-1/2}(\hat{\beta}_n - \beta)$ replacing $J^c_\omega(1)$. We thus continue to fail to achieve standard null asymptotics under heteroskedasticity.

5 Small Sample Simulations

In this section, we report the results of a Monte-Carlo simulation used to evaluate the finite sample properties of our proposed testing procedure, and of the procedures discussed in Section 3. All computations were carried out using *Ox* version 5.00, using 10,000 experimental replications. We make a number of special assumptions: (a) the sample size n is set equal to 100; (b) the innovations ε_t are taken to be i.i.d. standard normal random variables; (c) there are no deterministic terms in the data generating process; (d) all test statistics are based directly on an estimator of ϕ_n, allowing for a constant deterministic term, with no correction for serial dependence; (e) all simulated critical values are based on 1000 replications; (f) size calculations are based on a nominal size of 5%, and power calculations correspond to $\phi_n = 1 - 7/n = 0.93$. We confine ourselves to the i.i.d. framework so as to isolate the effect of heteroskedasticity on the tests, independent of the effect of serial dependence. Moreover, Boswijk (2005) does not discuss the extension of his test to data exhibiting serial dependence of general form.

The tests employed in the study are as follows.

$\boldsymbol{DF}$**:** The Dickey-Fuller coefficient statistic based on OLS regression of $y_{t,n}$ on $y_{t-1,n}$ and a constant, compared to the standard asymptotic critical value (i.e., -14.1).

$\boldsymbol{CT}$**:** The Dickey-Fuller coefficient statistic based on OLS regression of $y_{t,n}$ on $y_{t-1,n}$ and a constant, compared to a critical value simulated conditional on the estimated time-transformation $\hat{\eta}_n$, as in Cavaliere and Taylor (2007a).

$\boldsymbol{BS}$**:** The Dickey-Fuller coefficient statistic based on OLS regression of $y_{t,n}$ on $y_{t-1,n}$ and a constant, compared to a wild bootstrap critical value, as in Cavaliere and Taylor (2007b, 2008b).

$\boldsymbol{TT}$**:** The Dickey-Fuller coefficient statistic based on OLS regression of $y^{\dagger}_{t,n}$ on $y^{\dagger}_{t-1,n}$ and a constant, compared to the standard asymptotic critical value, as in Cavaliere and Taylor (2008a).

$\boldsymbol{RS}_h$**:** The Dickey-Fuller coefficient statistic based on OLS regression of $y^*_{t,n}$ on $y^*_{t-1,n}$ and a constant, compared to the standard asymptotic critical value, as described in Section 4.

$\boldsymbol{ML}_h$**:** The Boswijk coefficient statistic $n(\hat{\phi}^{\mathrm{B}}_n - 1)$ based on FGLS regression of $y_{t,n}$ on $y_{t-1,n}$ and a constant, compared to critical values simulated conditional on $\hat{\omega}_n$, as in Boswijk (2005).

The subscript h in the $\boldsymbol{RS}_h$ and $\boldsymbol{ML}_h$ tests refers to the bandwidth used to estimate ω. We have made one important change to the Boswijk (2005) test: the estimator of ω used to implement Boswijk's test is the same as that used to implement the test proposed here - that is, $\hat{\omega}_n$ defined above in (4.2). In Boswijk (2005), the suggested estimator is based on a one-sided kernel, whereas $\hat{\omega}_n$ as defined in Section 4 is in general based on a two-sided kernel. We do this to facilitate comparison between the two tests $\boldsymbol{RS}_h$ and $\boldsymbol{ML}_h$, and to avoid giving the Boswijk test an unfair disadvantage in the form of a less accurate estimate of ω. We use a Gaussian kernel to estimate ω, and allow the bandwidth h to take the values 0.05, 0.1, 0.2 and 0.4. Unfortunately, automatic bandwidth selection by cross-validation is computationally infeasible in a simulation of this size. In (4.2), we set $\hat{u}_{t,n} = y_{t,n} - y_{t-1,n}$.

We consider twenty different choices of ω in our simulations. The first nineteen of these are taken from the simulation reported in Cavaliere and Taylor (2007a), while the last is taken from Boswijk (2005). Table 1 displays the various choices of ω. (All tables are collected in Appendix B.) Model 1 is homoskedastic; models 2-5 feature a single variance break; models 6-9 feature two variance breaks; models 10-11 feature a trend in variance; models 12-17 are exponential Brownian motions or Ornstein-Uhlenbeck processes; and models 18-19 are GARCH(1,1) models. Finally, Model 20 is an exponential Ornstein-Uhlenbeck process considered by Boswijk (2005), denoted σ_3 in his study. This process is of particular interest, as Boswijk argues that the rescaled test proposed here has extremely low power when volatility is of this form. Note that, out of all twenty models, only the homoskedastic model and the two models of trending variance satisfy Assumption 2.1.

Size calculations under a nominal size of 5% are reported in Table 2. There are a few things to observe here. First, the rejection rates of the $\boldsymbol{DF}$ test are at times well in excess of the nominal 5% level, exceeding 10% in models 4, 8, 12 and 20, and reaching 25% in model 15. This finding is in line with numerous earlier studies in the literature. The $\boldsymbol{TT}$ test is substantially undersized, with a rejection rate below 4% in all twenty models, and below 3% in nine out of twenty models. Rejection rates for the $\boldsymbol{CT}$ and $\boldsymbol{BS}$ tests are generally close to 5%. The $\boldsymbol{CT}$ test performs particularly well, achieving a rejection rate between 5% and 6% in all twenty models, while the $\boldsymbol{BS}$ test is mildly oversized in model 15, with a rejection rate of 8.9%. Rejection rates for the $\boldsymbol{RS}_h$ and $\boldsymbol{ML}_h$ tests are, as one would expect, dependent on the bandwidth parameter h. When h is equal to 0.05, the tests tend to be mildly undersized, with the rejection rate for both tests falling below 4% in five of the twenty models. When h is equal to 0.1, the rejection rates for both tests are close to the 5% mark, although there is again some mild overrejection in model 15. When h is equal to 0.2, the size control of both tests deteriorates somewhat, with rejection rates exceeding 15% in model 15. When $h = 0.4$ we observe the most serious size distortion: the $\boldsymbol{RS}_{.4}$ and $\boldsymbol{ML}_{.4}$ tests achieve rejection rates of 10% or greater in three and five models respectively, and the rejection rate for both tests exceeds 20% in model 15. This is precisely what we would expect, since substantial oversmoothing in the estimation of ω causes $\hat{\omega}_n$ to become approximately flat, so that the behavior of the $\boldsymbol{RS}_h$ and $\boldsymbol{ML}_h$ tests approximates that of the $\boldsymbol{DF}$ test.

Size-adjusted power calculations are reported in Table 3. Since the $\boldsymbol{DF}$, $\boldsymbol{CT}$ and $\boldsymbol{BS}$ tests are all based on the same test statistic, the three columns are identical. The first thing to note is that all tests achieve approximately the same size-adjusted power in model 1, the homoskedastic model. This reflects the fact that the heteroskedasticity robust tests achieve the same local power as the $\boldsymbol{DF}$ test under homoskedasticity, as discussed in Section 4. The $\boldsymbol{TT}$ test fares rather poorly in general, achieving the lowest size-adjusted power in eight of the twenty models, or in twelve of the twenty models if we exclude the $\boldsymbol{RS}_h$ and $\boldsymbol{ML}_h$ tests when $h \neq 0.1$. Its performance is particularly bad in models 2, 3, 6, 7 and 11, and is even worse in view of the substantial undersizing exhibited in Table 2. The low power of the $\boldsymbol{TT}$ test should not be surprising given the nature of its construction: the time-transformation operates by sampling observations more frequently in times of high volatility, and less frequently in times of low volatility. This leads to a kind of "reverse GLS" effect, and consequently low power.

The size-adjusted powers of the $\boldsymbol{DF}/\boldsymbol{CT}/\boldsymbol{BS}$ test and the $\boldsymbol{RS}_{.1}$ and $\boldsymbol{ML}_{.1}$ tests tend to be fairly close. The $\boldsymbol{ML}_{.1}$ test achieves the highest or equal highest size-adjusted power in twelve of the twenty models, while the $\boldsymbol{RS}_{.1}$ test achieves the highest or equal highest size adjusted power in eight of the models. The $\boldsymbol{DF}/\boldsymbol{CT}/\boldsymbol{BS}$ test achieves the highest size-adjusted power in only one model. However, the $\boldsymbol{CT}$ and $\boldsymbol{BS}$ tests have the advantage of not requiring the selection of a bandwidth parameter. Table 3 shows that the size-adjusted power of the $\boldsymbol{RS}_h$ test can be reduced somewhat if h is chosen too small.

Some broad conclusions to be drawn from these results are as follows: (a) the $\boldsymbol{DF}$ test should be avoided in situations of potentially unstable volatility, due to its tendency to overreject when the null is true; (b) the $\boldsymbol{TT}$ test does relatively poorly compared to the other tests in terms of finite sample power, and is undersized in finite samples; (c) the empirical size of the $\boldsymbol{CT}$ and $\boldsymbol{BS}$ tests is close to the nominal size; (d) the empirical size of the $\boldsymbol{RS}_h$ and $\boldsymbol{BS}_h$ tests is close to the nominal size if h is chosen suitably, but the tests may overreject if h is too large; (e) the $\boldsymbol{RS}_h$ and $\boldsymbol{ML}_h$ tests may enjoy a mild size-adjusted power advantage over the $\boldsymbol{CT}$ and $\boldsymbol{BS}$ tests, but this is balanced by the fact that no bandwidth selection is required to implement the $\boldsymbol{CT}$ and $\boldsymbol{BS}$ tests.

Although only 3 of the 20 volatility models satisfy Assumption 2.1, the results in Tables 2-4 indicate that the rescaled unit root test we have proposed in this paper tends to perform well under much broader conditions. In particular, discontinuities in ω do not seem to present a problem. Perhaps most surprising is the performance of our test in model 20, and in the other models of integrated or near-integrated stochastic volatility. Boswijk (2005) presents simulation evidence indicating that, if ω were known and used in place of $\hat{\omega}_n$ to construct our rescaled test statistic, then the local power of our test under model 20 stays approximately constant at 5% as c increases. Yet in the simulation reported here, our test achieves power in excess of 20% with only 100 observations, provided that h is 0.1 or greater. A possible explanation for this apparent contradiction is that our test is not adaptive when ω is nondifferentiable. It would seem that this failure of adaptivity actually leads to an increase in power in certain circumstances - perhaps by smoothing away the nondifferentiability of ω. This conjecture is supported by the fact that the size-adjusted power of our test in model 20 drops markedly when h is equal to 0.05.

6 Conclusion

In this paper we have proposed an approach to unit root testing that works effectively in the presence of unstable volatility. Finite sample simulations indicate that our test performs quite well relative to existing tests of this sort. Moreover, when working with a model without a deterministic trend, our tests are convenient in that they require the usual critical values that apply under homoskedasticity.

A Mathematical Appendix

Proof of Lemma 4.1 Begin by writing

$$\hat{\omega}_n(r)^2 = \nu_n(r) + \eta_n(r),$$

where

$$\begin{aligned}
\nu_n(r) &= \frac{\sum_{t=1}^{n} k\left(\frac{t-nr}{nh_n}\right)\omega\left(\frac{t}{n}\right)^2 \varepsilon_t^2}{\sum_{t=1}^{n} k\left(\frac{t-nr}{nh_n}\right)}, \\
\eta_n(r) &= \frac{\sum_{t=1}^{n} k\left(\frac{t-nr}{nh_n}\right)\left(\hat{u}_{t,n}^2 - u_{t,n}^2\right)}{\sum_{t=1}^{n} k\left(\frac{t-nr}{nh_n}\right)}.
\end{aligned}$$

It suffices for us to verify the following three statements:

$$E\sup_{r\in[0,1]}\left|\nu_n(r) - E\nu_n(r)\right| = O\left(\frac{1}{\sqrt{n}h_n}\right) \tag{A.1}$$

$$\sup_{r\in[0,1]}\left|E\nu_n(r) - \omega(r)^2\right| = O(h_n) \tag{A.2}$$

$$\sup_{r\in[0,1]}\left|\eta_n(r)\right| = O_p\left(\frac{1}{\sqrt{n}h_n}\right). \tag{A.3}$$

(A.1) can be demonstrated using Fourier methods, as in the proof of uniform consistency of Nadaraya-Watson estimators under mixing conditions (see e.g. Pagan and Ullah, 1999, pp. 36-39, 44-47). (A.2) can be proved using simple arguments relating to the approximation of integrals by discrete sums, and a mean value expansion. To see that (A.3) holds, observe that

$$\eta_n(r) = -\frac{\frac{1}{nh_n}\sum_{t=1}^{n} k\left(\frac{t-nr}{nh_n}\right)\left(\hat{u}_{t,n} + u_{t,n}\right)\left(\hat{u}_{t,n} - u_{t,n}\right)}{\frac{1}{nh_n}\sum_{t=1}^{n} k\left(\frac{t-nr}{nh_n}\right)}. \tag{A.4}$$

We need only show that the numerator in (A.4) is uniformly $O_p\left(\frac{1}{\sqrt{nh_n}}\right)$. Boundedness of k and (4.1) yield

$$\begin{aligned}
\sup_{r\in[0,1]}\left|\frac{1}{nh_n}\sum_{t=1}^{n}k\left(\frac{t-nr}{nh_n}\right)(\hat{u}_{t,n}+u_{t,n})(\hat{u}_{t,n}-u_{t,n})\right| &= O_p\left(\frac{1}{\sqrt{nh_n}}\right)\cdot\frac{1}{n}\sum_{t=1}^{n}|\hat{u}_{t,n}+u_{t,n}| \\
&= O_p\left(\frac{1}{\sqrt{nh_n}}\right)\cdot\frac{1}{n}\sum_{t=1}^{n}|\varepsilon_t|+O_p\left(\frac{1}{nh_n}\right).
\end{aligned}$$

A law of large numbers for mixing sequences ensures that $\frac{1}{n}\sum_{t=1}^{n}|\varepsilon_t|=O_p(1)$. It follows that the numerator in (A.1) is uniformly $O_p\left(\frac{1}{\sqrt{nh_n}}\right)$. ■

Proof of Lemma 4.2 Observe that

$$\begin{aligned}
\hat{\omega}_n'(r)-\omega'(r) &= \frac{\frac{\mathrm{d}}{\mathrm{d}r}\hat{\omega}_n(r)^2}{2\hat{\omega}_n}-\frac{\frac{\mathrm{d}}{\mathrm{d}r}\omega(r)^2}{2\omega} \\
&= \frac{\omega(r)\left(\frac{\mathrm{d}}{\mathrm{d}r}\hat{\omega}_n(r)^2-\frac{\mathrm{d}}{\mathrm{d}r}\omega(r)^2\right)-\frac{\mathrm{d}}{\mathrm{d}r}\omega(r)^2(\hat{\omega}_n(r)-\omega(r))}{2\hat{\omega}_n(r)\,\omega(r)}.
\end{aligned}$$

In view of Lemma 4.1, it suffices for us to show that $\sup_{r\in[\tau_n,1-\tau_n]}\left|\frac{\mathrm{d}}{\mathrm{d}r}\hat{\omega}_n(r)^2-\frac{\mathrm{d}}{\mathrm{d}r}\omega(r)^2\right|=o_p(1)$. This will be true under the following three conditions:

$$E\sup_{r\in[0,1]}|\nu_n'(r)-E\nu_n'(r)| = O\left(\frac{1}{\sqrt{n}h_n^2}\right) \tag{A.5}$$

$$\sup_{r\in[\tau,1-\tau]}\left|E\nu_n'(r)-\frac{\mathrm{d}}{\mathrm{d}r}\omega(r)^2\right| = o(1) \tag{A.6}$$

$$\sup_{r\in[0,1]}|\eta_n'(r)| = O_p\left(\frac{1}{\sqrt{n}h_n^2}\right). \tag{A.7}$$

(A.5) and (A.7) can be proved using minor variations on the arguments used to prove (A.1) and (A.3) respectively. It remains to prove (A.6). Begin by writing

$$E\nu_n'(r)=\frac{E\nu_n(r)\frac{1}{nh_n^2}\sum_{t=1}^{n}k'\left(\frac{t-nr}{nh_n}\right)-\frac{1}{nh_n^2}\sum_{t=1}^{n}k'\left(\frac{t-nr}{nh_n}\right)\omega\left(\frac{t}{n}\right)^2}{\frac{1}{nh_n}\sum_{t=1}^{n}k\left(\frac{t-nr}{nh_n}\right)}. \tag{A.8}$$

Using (A.2) and simple arguments relating to the approximation of integrals by discrete sums, it can be shown that the first term in the numerator of (A.8) satisfies

$$\begin{aligned}
&\sup_{r\in[\tau_n,1-\tau_n]}\left|E\nu_n\left(r\right)\frac{1}{nh_n^2}\sum_{t=1}^{n}k'\left(\frac{t-nr}{nh_n}\right)-\omega\left(r\right)^2\frac{1}{h_n}\left(k\left(\frac{1-r}{h_n}\right)-k\left(\frac{-r}{h_n}\right)\right)\right| \\
&= O\left(\sup_{r\in[\tau_n,1-\tau_n]}\left|k\left(\frac{1-r}{h_n}\right)\right|+\sup_{r\in[\tau_n,1-\tau_n]}\left|k\left(\frac{-r}{h_n}\right)\right|\right)+o\left(1\right). \qquad \text{(A.9)}
\end{aligned}$$

Integration by parts and the mean value theorem can be used to show that the second term in the numerator of (A.8) satisfies

$$\begin{aligned}
&\sup_{r\in[\tau_n,1-\tau_n]}\left|\frac{1}{nh_n^2}\sum_{t=1}^{n}k'\left(\frac{t-nr}{nh_n}\right)\omega\left(\frac{t}{n}\right)^2-\left(\frac{\mathrm{d}}{\mathrm{d}r}\omega_n\left(r\right)^2\right)\frac{1}{nh_n}\sum_{t=1}^{n}k\left(\frac{t-nr}{nh_n}\right)\right. \\
&\qquad\left.-\frac{1}{h_n}\left(k\left(\frac{1-r}{h_n}\right)\omega(1)^2-k\left(\frac{-r}{h_n}\right)\omega(0)^2\right)\right| = o(1). \qquad \text{(A.10)}
\end{aligned}$$

Combining (A.8), (A.9) and (A.10), and applying the mean value theorem, we obtain

$$\begin{aligned}
&\sup_{r\in[\tau_n,1-\tau_n]}\left|E\nu_n'\left(r\right)-\frac{\mathrm{d}}{\mathrm{d}r}\omega\left(r\right)^2\right| \\
&= O\left(\sup_{r\in[\tau_n,1-\tau_n]}\left|k\left(\frac{1-r}{h_n}\right)\frac{\omega(r)^2-\omega(1)^2}{h_n}-k\left(\frac{-r}{h_n}\right)\frac{\omega(r)^2-\omega(0)^2}{h_n}\right|\right) \\
&\quad +O\left(\sup_{r\in[\tau_n,1-\tau_n]}\left|k\left(\frac{1-r}{h_n}\right)\right|\right)+O\left(\sup_{r\in[\tau_n,1-\tau_n]}\left|k\left(\frac{-r}{h_n}\right)\right|\right)+o(1) \\
&= O\left(\sup_{r\in[\tau_n,1-\tau_n]}\left|\frac{1-r}{h_n}k\left(\frac{1-r}{h_n}\right)\right|\right)+O\left(\sup_{r\in[\tau_n,1-\tau_n]}\left|\frac{-r}{h_n}k\left(\frac{-r}{h_n}\right)\right|\right) \\
&\quad +O\left(\sup_{r\in[\tau_n,1-\tau_n]}\left|k\left(\frac{1-r}{h_n}\right)\right|\right)+O\left(\sup_{r\in[\tau_n,1-\tau_n]}\left|k\left(\frac{-r}{h_n}\right)\right|\right)+o(1).
\end{aligned}$$

Since $xk(x)\to 0$ as $|x|\to\infty$, this proves (A.6). ■

Proof of Lemma 4.3 In view of (A.5) and (A.7), we need only show that $\sup_{r\in[0,1]}\left|E\nu_n'\left(r\right)-\frac{\mathrm{d}}{\mathrm{d}r}\omega\left(r\right)^2\right| = O(1)$. This may be proved in the same way as (A.6), except we observe that $\sup_{r\in[0,1]} \frac{r}{h_n} k\left(\frac{r}{h_n}\right) = O(1)$ and $\sup_{r\in[0,1]} k\left(\frac{r}{h_n}\right) = O(1)$. ■

Proof of Theorem 4.1 It suffices for us to show that

$$-c\int_0^r J^c(s)\mathrm{d}s = J^c(r) - B(r).$$

Begin by using the representation $J^c(r) = B(r) - c\int_0^r \exp(-c(r-s))B(s)\mathrm{d}s$ to write

$$-c\int_0^r J^c(s)\mathrm{d}s = -c\int_0^r B(s)\mathrm{d}s + c^2\int_0^r \left(\int_0^s \exp(-c(s-u))B(u)\mathrm{d}u\right)\mathrm{d}s.$$

Applying the Fubini-Tonelli theorem, we have

$$\begin{aligned} -c\int_0^r J^c(s)\mathrm{d}s &= -c\int_0^r B(s)\mathrm{d}s + c^2\int_0^r \left(\int_u^r \exp(-c(s-u))\mathrm{d}s\right) B(u)\mathrm{d}u \\ &= -c\int_0^r \exp(-c(r-u))B(u)\mathrm{d}u. \end{aligned}$$

Another application of $J^c(r) = B(r) - c\int_0^r \exp(-c(r-s))B(s)\mathrm{d}s$ completes the proof. ■

Proof of Theorem 4.2 We first prove $n^{-1/2}y^*_{[nr],n} \Rightarrow \lambda\xi^c_\omega(r)$. Using summation by parts, it can be shown that

$$\begin{aligned} n^{-1/2}y^*_{[nr],n} &= \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\hat{\omega}_n\left(\frac{t}{n}\right)^{-1}(x_{t,n} - x_{t-1,n}) \\ &= \frac{\omega\left(\frac{[nr]}{n}\right)}{\hat{\omega}_n\left(\frac{[nr]}{n}\right)}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\varepsilon_t\right) + \sum_{t=1}^{[nr]}\left(\frac{\omega\left(\frac{t}{n}\right)}{\hat{\omega}_n\left(\frac{t}{n}\right)} - \frac{\omega\left(\frac{t-1}{n}\right)}{\hat{\omega}_n\left(\frac{t-1}{n}\right)}\right)\left(\frac{1}{\sqrt{n}}\sum_{s=1}^{t-1}\varepsilon_s\right) \\ &\quad + \left(\exp\left(-\frac{c}{n}\right) - 1\right)\sum_{t=1}^{[nr]}\hat{\omega}_n\left(\frac{t}{n}\right)^{-1}\left(n^{-1/2}x_{t-1,n}\right). \end{aligned} \tag{A.11}$$

Using (3.2), Lemma 4.1, the continuous mapping theorem, and the approximation $\exp\left(-\frac{c}{n}\right) = 1 - \frac{c}{n} + O\left(\frac{1}{n^2}\right)$, we have

$$\left(\exp\left(-\frac{c}{n}\right) - 1\right)\sum_{t=1}^{[nr]} \hat{\omega}_n\left(\frac{t}{n}\right)^{-1}\left(n^{-1/2}x_{t-1,n}\right) \Rightarrow -c\lambda\int_0^r \omega\left(s\right)^{-1} J_\omega^c(s)\mathrm{d}s.$$

Lemma 4.1 and an FCLT for mixing sequences imply that

$$\frac{\omega\left(\frac{[nr]}{n}\right)}{\hat{\omega}_n\left(\frac{[nr]}{n}\right)}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\varepsilon_t\right) \Rightarrow \lambda B\left(r\right).$$

It remains to show that the second term on the right-hand side of (A.11) is uniformly $o_p(1)$. Using the mean value theorem, we have

$$\begin{aligned}
&\sup_{r\in[0,1]}\left|\sum_{t=1}^{[nr]}\left(\frac{\omega\left(\frac{t}{n}\right)}{\hat{\omega}_n\left(\frac{t}{n}\right)} - \frac{\omega\left(\frac{t-1}{n}\right)}{\hat{\omega}_n\left(\frac{t-1}{n}\right)}\right)\left(\frac{1}{\sqrt{n}}\sum_{s=1}^{t-1}\varepsilon_s\right)\right| \\
\leq\ & \sum_{t=1}^{n}\left|\frac{\omega\left(\frac{t}{n}\right)}{\hat{\omega}_n\left(\frac{t}{n}\right)} - \frac{\omega\left(\frac{t-1}{n}\right)}{\hat{\omega}_n\left(\frac{t-1}{n}\right)}\right|\left|\frac{1}{\sqrt{n}}\sum_{s=1}^{t-1}\varepsilon_s\right| \\
\leq\ & B_n\left(\tau\right)C_n\left(1,n\right) + B_n\left(0\right)C_n\left(1,[n\tau_n]+1\right) + B_n\left(0\right)C_n\left([n\left(1-\tau_n\right)]+1,n\right), \qquad \text{(A.12)}
\end{aligned}$$

where

$$\begin{aligned}
B_n\left(a\right) &= \sup_{r\in[a,1-a]}\left|\frac{\omega'\left(r\right)\hat{\omega}_n\left(r\right) - \hat{\omega}_n'\left(r\right)\omega\left(r\right)}{\hat{\omega}_n\left(r\right)^2}\right|, \\
C_n\left(a,b\right) &= \frac{1}{n}\sum_{t=a}^{b}\left|\frac{1}{\sqrt{n}}\sum_{s=1}^{t-1}\varepsilon_s\right|,
\end{aligned}$$

and where τ_n is chosen as in the statement of Lemma 4.2. An FCLT for mixing sequences and the continuous mapping theorem imply that

$$\begin{aligned} C_n(1,n) &= O_p(1) \\ C_n(1,[n\tau_n]+1) &= o_p(1) \\ C_n([n(1-\tau_n)]+1,n) &= o_p(1). \end{aligned}$$

From Lemma 4.1, Lemma 4.2 and Lemma 4.3, we have

$$\begin{aligned} B_n(0) &= O_p(1) \\ B_n(\tau_n) &= o_p(1). \end{aligned}$$

It follows from (A.12) that

$$\sup_{r\in[0,1]}\left|\sum_{t=1}^{[nr]}\left(\frac{\omega\left(\frac{t}{n}\right)}{\hat{\omega}_n\left(\frac{t}{n}\right)}-\frac{\omega\left(\frac{t-1}{n}\right)}{\hat{\omega}_n\left(\frac{t-1}{n}\right)}\right)\left(\frac{1}{\sqrt{n}}\sum_{s=1}^{t-1}\varepsilon_s\right)\right|=o_p(1).$$

This proves $n^{-1/2}y^*_{[nr],n}\Rightarrow\lambda\xi^c_\omega(r)$. $\hat{\sigma}^*_n\to_p 1$ can be proved easily using (3.2), Lemma 4.1 above, and a LLN for mixing sequences. $\hat{\lambda}^*_n\to_p\lambda$ can be proved in a relatively straightforward fashion by using Lemma 4.1 and the fact that $n^{-1/2}y^*_{[nr],n}\Rightarrow\lambda\xi^c_\omega(r)$ to show that

$$\hat{\lambda}^{*2}_n=\sum_{j=-n+1}^{n-1}\kappa\left(\frac{j}{m_n}\right)\left(\frac{1}{n}\sum_{t=|j|+1}^{n}\varepsilon_{t,n}\varepsilon_{t-|j|,n}\right)+o_p(1),$$

and then appealing to Theorem 2 of de Jong (2000). ■

Proof of Theorem 4.3 In view of Theorem 4.2, it suffices for us to show that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\frac{\frac{1}{n}(y_{n,n}-y_{0,n})-\beta}{\hat{\omega}_n\left(\frac{t}{n}\right)}\Rightarrow\lambda J^c_\omega(1)\int_0^r\omega(s)^{-1}\mathrm{d}s.$$

It is easy to see that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\frac{\frac{1}{n}\left(y_{n,n}-y_{0,n}\right)-\beta}{\hat{\omega}_n\left(\frac{t}{n}\right)} = \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\frac{\frac{1}{n}\left(x_{n,n}-x_{0,n}\right)}{\hat{\omega}_n\left(\frac{t}{n}\right)}$$

$$= \left(\frac{1}{\sqrt{n}}x_{n,n}+O_p\left(\frac{1}{\sqrt{n}}\right)\right)\frac{1}{n}\sum_{t=1}^{[nr]}\hat{\omega}_n\left(\frac{t}{n}\right)^{-1}.$$

Our desired result now follows from (3.2), the continuous mapping theorem, and Lemma 4.1. ■

B Tables

Table 1: *Models of heteroskedasticity considered in Tables 2-3*

Model	Volatility Function
1	$\omega(r) = 1$
2	$\omega(r) = 1\{r < 0.2\} + 3\{r \geq 0.2\}$
3	$\omega(r) = 1\{r < 0.8\} + 3\{r \geq 0.8\}$
4	$\omega(r) = 1\{r < 0.2\} + \frac{1}{3}\{r \geq 0.2\}$
5	$\omega(r) = 1\{r < 0.8\} + \frac{1}{3}\{r \geq 0.8\}$
6	$\omega(r) = 1\{r < 0.1\} + 3(0.1 \leq r < 0.9) + 1\{r \geq 0.9\}$
7	$\omega(r) = 1\{r < 0.4\} + 3\{0.4 \leq r < 0.6\} + 1\{r \geq 0.6\}$
8	$\omega(r) = 1\{r < 0.1\} + \frac{1}{3}\{0.1 \leq r < 0.9\} + 1\{r \geq 0.9\}$
9	$\omega(r) = 1\{r < 0.4\} + \frac{1}{3}\{0.4 \leq r < 0.6\} + 1\{r \geq 0.6\}$
10	$\omega(r) = 1 - \frac{2}{3}r$
11	$\omega(r) = 1 + 2r$
12	$\omega(r) = \exp(2B(r))$
13	$\omega(r) = \exp\left(2J^{10}(r)\right)$
14	$\omega(r) = \exp\left(2J^{20}(r)\right)$
15	$\omega(r) = \exp(4.5B(r))$
16	$\omega(r) = \exp\left(4.5J^{10}(r)\right)$
17	$\omega(r) = \exp\left(4.5J^{20}(r)\right)$
18	$\omega(t/n)^2 = 0.5 + 0.1\omega((t-1)/n)^2 \varepsilon_{t-1}^2 + 0.4\omega((t-1)/n)^2$, $\omega(1/n) = 1$
19	$\omega(t/n)^2 = 0.1 + 0.1\omega((t-1)/n)^2 \varepsilon_{t-1}^2 + 0.8\omega((t-1)/n)^2$, $\omega(1/n) = 1$
20	$\omega(r) = \exp\left(5J^{10}(r)\right)$

Table 2: *Size*

Model	$\boldsymbol{DF}$	$\boldsymbol{CT}$	$\boldsymbol{BS}$	$\boldsymbol{TT}$	$\boldsymbol{RS}_{.05}$	$\boldsymbol{RS}_{.1}$	$\boldsymbol{RS}_{.2}$	$\boldsymbol{RS}_{.4}$	$\boldsymbol{ML}_{.05}$	$\boldsymbol{ML}_{.1}$	$\boldsymbol{ML}_{.2}$	$\boldsymbol{ML}_{.4}$
1	4.6	5.4	5.3	3.2	4.0	4.1	4.3	4.5	4.2	4.7	5.0	5.1
2	4.1	5.8	5.5	3.2	4.0	4.1	4.0	4.0	4.6	4.8	4.5	4.5
3	8.3	5.7	5.8	2.7	3.8	4.4	5.3	6.5	4.4	5.1	6.1	7.3
4	14.7	5.2	6.4	2.7	3.9	4.6	5.7	9.0	3.9	4.6	6.4	10.3
5	5.4	5.4	5.6	3.1	3.9	4.2	4.8	5.1	4.5	5.0	5.5	5.7
6	3.5	5.6	5.4	3.1	3.4	3.3	3.4	3.5	4.0	3.8	3.9	4.0
7	6.9	5.9	5.8	2.4	4.1	4.8	5.9	6.7	4.5	5.2	6.6	7.5
8	15.1	5.2	6.6	3.1	4.2	5.5	8.1	12.5	4.5	6.3	9.5	14.2
9	5.6	5.4	5.4	3.2	4.1	4.5	5.0	5.3	4.1	4.9	5.7	6.2
10	7.0	5.2	5.5	3.2	4.0	4.3	4.8	5.5	4.3	4.9	5.5	6.4
11	4.5	5.7	5.5	3.5	3.9	3.8	3.9	4.1	4.3	4.6	4.7	4.7
12	13.1	5.5	6.2	2.6	4.5	5.2	7.5	10.3	4.4	5.6	8.2	11.5
13	6.0	5.7	5.8	2.9	4.1	4.6	5.1	5.6	4.3	5.1	5.7	6.3
14	5.2	5.7	5.5	3.0	4.2	4.3	4.6	4.9	4.3	5.0	5.4	5.8
15	25.0	5.7	8.9	2.8	4.9	8.6	16.8	22.3	3.5	7.6	16.6	23.1
16	9.7	5.8	6.1	2.0	4.6	5.2	7.0	8.4	3.5	5.2	7.5	9.4
17	7.0	5.6	5.6	2.1	4.3	4.7	5.5	6.5	3.6	4.8	6.0	7.2
18	4.8	5.4	5.4	3.3	4.1	4.2	4.4	4.7	4.3	4.9	5.2	5.4
19	5.3	5.7	5.7	3.5	4.2	4.5	4.8	5.2	4.6	5.2	5.6	5.8
20	10.5	5.6	6.2	1.7	4.5	5.6	7.2	9.0	3.4	5.3	8.1	10.0

Table 3: *Size-adjusted power*

Model	***DF***	***CT***	***BS***	***TT***	$\boldsymbol{RS}_{.05}$	$\boldsymbol{RS}_{.1}$	$\boldsymbol{RS}_{.2}$	$\boldsymbol{RS}_{.4}$	$\boldsymbol{ML}_{.05}$	$\boldsymbol{ML}_{.1}$	$\boldsymbol{ML}_{.2}$	$\boldsymbol{ML}_{.4}$
1	28.3	28.3	28.3	28.5	27.9	28.8	28.9	28.5	28.8	28.5	28.6	28.4
2	32.0	32.0	32.0	23.1	32.8	32.9	32.7	32.7	34.5	35.0	33.4	32.8
3	25.2	25.2	25.2	14.5	27.1	27.3	27.0	26.7	27.9	27.4	26.9	26.7
4	15.3	15.3	15.3	15.6	14.6	15.9	17.2	16.5	18.0	18.0	17.9	16.5
5	26.2	26.2	26.2	26.4	25.2	27.6	27.3	26.8	26.8	26.8	26.8	26.5
6	35.2	35.2	35.2	25.4	35.6	37.2	36.4	35.4	41.1	38.7	37.2	35.8
7	27.7	27.7	27.7	16.9	23.0	27.8	30.3	28.2	33.5	33.4	30.9	28.2
8	14.8	14.8	14.8	16.1	14.3	15.2	14.8	14.8	15.4	15.5	15.5	14.7
9	25.4	25.4	25.4	25.0	20.7	23.8	24.4	25.5	25.8	25.3	24.7	25.4
10	22.9	22.9	22.9	24.5	22.9	23.0	22.7	23.2	23.2	22.9	22.6	23.0
11	30.0	30.0	30.0	21.3	31.5	33.1	32.8	31.2	32.9	32.8	32.6	31.5
12	14.1	14.1	14.1	16.1	18.5	20.1	18.6	16.5	19.4	18.1	17.1	16.1
13	25.5	25.5	25.5	22.9	23.3	26.0	26.3	25.8	27.0	26.3	26.3	25.6
14	26.9	26.9	26.9	25.9	25.4	27.0	27.4	27.4	27.4	27.0	27.3	27.4
15	6.5	6.5	6.5	11.0	9.1	10.5	8.2	7.3	9.2	8.8	7.1	7.3
16	21.3	21.3	21.3	19.8	14.6	21.1	23.1	23.0	29.3	24.7	23.4	23.0
17	25.0	25.0	25.0	24.3	19.3	24.6	25.6	25.4	30.1	26.9	26.4	25.5
18	28.0	28.0	28.0	26.7	27.2	28.3	28.2	27.9	28.8	27.8	28.2	28.0
19	26.5	26.5	26.5	24.5	26.6	26.9	26.9	26.3	27.0	27.0	26.5	26.4
20	20.3	20.3	20.3	20.1	12.9	20.3	21.6	21.8	29.3	24.3	22.6	21.7

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59** 817–858.

Andrews, D. W. K. (1995). Nonparametric kernel estimation for semiparametric models. *Econometric Theory* **11** 560–596.

Boswijk, H. P. (2001). Testing for a unit root with near integrated volatility. Tinbergen Institute Discussion Paper 2001-077/4.

Boswijk, H. P. (2005). Adaptive testing for a unit root with nonstationary volatility. UvA Econometrics Discussion Paper 2005/07.

Cavaliere, G. (2004a). Testing stationarity under a permanent variance shift. *Econ. Letters* **82** 403–408.

Cavaliere, G. (2004b). Unit root tests under time-varying variances. *Econometric Rev.* **23** 259–292.

Cavaliere, G. and Taylor, A. M. R. (2007a). Testing for unit roots in time series models with non-stationary volatility. *J. Econometrics* **140** 919–947.

Cavaliere, G. and Taylor, A. M. R. (2007b). Heteroskedastic time series with a unit root. Mimeo, University of Nottingham.

Cavaliere, G. and Taylor, A. M. R. (2008a). Time-transformed unit root tests for models with non-stationary volatility. *J. Time Ser. Anal.* **29** 300–330.

Cavaliere, G. and Taylor, A. M. R. (2008b). Bootstrap unit root tests for time series with non-stationary volatility. *Econometric Theory* **24** 43–71.

Chu, C. -K. and Marron, J. S. (1991). Comparison of two bandwidth selectors with dependent errors. *Ann. Stat.* **19** 1906–1918.

Dahl, C. M. and Levine, M. (2006). Nonparametric estimation of volatility models with serially dependent innovations. *Statist. Probab. Lett.* **76** 2007–2016.

de Jong, R. M. (2000). A strong consistency proof for heteroskedasticity and autocorrelation consistent covariance matrix estimators. *Econometric Theory* **16** 262–268.

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

Hamori, S. and Tokihisa, A. (1997). Testing for a unit root in the presence of a variance shift. *Econ. Letters* **57** 245–253.

Hansen, B. E. (1995). Regression with nonstationary volatility. *Econometrica* **63** 1113–1132.

Kim, T. H., Leybourne, S. and Newbold, P. (2002). Unit root tests with a break in innovation variance. *J. Econometrics* **109** 365–387.

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

Pagan, A. R. and Ullah, A. (1999). *Nonparametric Econometrics*. Cambridge University Press.

Perron, P. and Ng, S. (1996). Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. *Rev. Econ. Stud.* **63** 435–463.

Phillips, P. C. B. (1987a). Time series regression with a unit root. *Econometrica* **55** 277–301.

Phillips, P. C. B. (1987b). Towards a unified asymptotic theory for autoregression. *Biometrika* **74** 535–547.

Phillips, P. C. B. and Lee, C. C. (1996). Efficiency gains from quasi-differencing under nonstationarity. In Robinson, P. M. and Rosenblatt, M. (eds.) *Essays in Honour of E. J. Hannan* 300–314. Springer-Verlag.

Phillips, P. C. B., and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika* **75** 335–346.

Phillips, P. C. B. and Xiao, Z. (1998). A primer on unit root testing. *J. Econ. Surveys* **12** 423–470.

Phillips, P. C. B., and Xu, K. -L. (2006). Inference in autoregression under heteroskedasticity. *J. Time Ser. Anal.* **27** 289–308.

Schmidt, P. and Phillips, P. C. B. (1992). LM tests for a unit root in the presence of deterministic trends. *Oxford Bull. Econ. Statist.* **54** 257–287.

Stock, J. H. (1999). A class of tests for integration and cointegration. In Engle, R. F. and White, H. (eds.) *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W. J. Granger* 137–167. Oxford University Press.

Wong, W. H. (1983). On the consistency of cross validation in kernel nonparametric regression. *Ann. Stat.* **11** 1136–1141.

Xu, K. -L. and Phillips, P. C. B. (2008). Adaptive estimation of autoregressive models with time-varying variance. *J. Econometrics* **142** 265–280.