Granger Centre
for time series econometrics

Granger Centre Discussion Paper Series

On the inconsistency of the unrestricted estimator of the information matrix near a unit root

by

Tassos Magdalinos

Granger Centre Discussion Paper No. 06/05

The University of Nottingham

On the inconsistency of the unrestricted estimator of the information matrix near a unit root[1]

Tassos Magdalinos
Department of Mathematics
University of York
York YO10 5DD, UK
am175@york.ac.uk

October 3, 2005

[1]The author thanks Karim Abadir and Peter Phillips for useful comments as well as the EPSRC and the Onassis Foundation for scholarship support.

Abstract

The unrestricted estimator of the information matrix is shown to be inconsistent for an autoregressive process with a root lying in a neighbourhood of unity with radial length proportional or smaller than n^{-1}, i.e. a root that takes the form $\rho = 1 + c/n^{\alpha}$, $\alpha \geq 1$. In this case the information evaluated at $\hat{\rho}_n$ converges to a non-degenerate random variable and contributes to the asymptotic distribution of a Wald test for the null hypothesis of a random walk versus a stable AR(1) alternative. With this newly derived asymptotic distribution the above Wald test is found to improve its performance. A non local criterion of asymptotic relative efficiency based on Bahadur slopes has been employed for the first time to the problem of unit root testing. The Wald test derived in the paper is found to be as efficient as the Dickey Fuller t ratio test and to outperform the non studentised Dickey Fuller test and a Lagrange Multiplier test.

Some keywords: Unit root distribution; Neighbourhoods of unity; Information matrix; Inconsistency; Wald test; Bahadur slopes.

1. Introduction

The derivation of statistical procedures for detecting the presence of a unit root in autoregressive processes has attracted a lot of attention among econometricians over the last two decades. Early contributions such as Dickey and Fuller (1979) and Evans and Savin (1981) derived tests for random walks against dynamically stable autoregressions. A Gaussian structure was imposed on the innovation errors in order to consider the problem in a likelihood inference framework. In this context, we review a Wald-type test based on the Fisher information $\mathcal{I}_n(\rho)$ (see (4) below) suggested by Evans and Savin (1981) and discussed further in Abadir (1993b). A basic assumption in Evans and Savin (1981), followed later by Abadir (1993b), is that the unrestricted estimator of the information matrix, $\mathcal{I}_n(\hat{\rho}_n)$ defined in (5), is consistent for $\mathcal{I}_n(\rho)$ even when the autoregressive parameter ρ lies in a viccinity of unity. It is one of the main purposes of this paper to show that this is not the case and discuss the consequences for the resulting Wald test.

The inconsistency of $\mathcal{I}_n(\hat{\rho}_n)$ is established in Theorem 2.3. It is shown that, after appropriate normalization, $\mathcal{I}_n(\hat{\rho}_n)$ has a non-degenerate asymptotic distribution when the autoregressive root lies in a neighbourhood of unity with radial length proportional or smaller than n^{-1}, i.e. a root that takes the form $\rho = 1 + c/n^{\alpha}$, $\alpha \geq 1$, $c \in \mathbb{R}$. This limiting distribution can be expressed as a deterministic function of the familiar random variables

$$Z_1 := \frac{W^2(1) - 1}{\int_0^1 W(s)^2 ds} \tag{1}$$

in the unit root case ($c = 0$ or $\alpha > 1$) and

$$Z_2 := \frac{\int_0^1 J_c(s) dW(s)}{\int_0^1 J_c(s)^2 ds} \tag{2}$$

in the local to unity case ($\alpha = 1$), where $W(\cdot)$ is standard Brownian motion on $D[0,1]$ and J_c is an Ornstein-Uhlenbeck process defined as $J_c(t) = \int_0^t e^{c(t-s)} dW(s)$.

In Section 3, the inconsistency result of Theorem 2.3 is applied to unit root testing. In the fashion of Evans and Savin (1981) and Abadir (1993b), we consider a Wald type test for the null hypothesis of a random walk versus the alternative of a stable root AR(1) process. A new expression for the asymptotic distribution under the null hypothesis is derived leading to a re-evaluation of the asymptotic properties of the test. To this end, we find the order of magnitude of the tails of the limiting distribution function by applying the transformation theorem and Abadir's (1993a) formula for the p.d.f. of Z_1. Use of the transformation theorem requires a non-trivial inversion carried out with the help of a special function known as the Lambert W function, a short account of which is provided in Appendix A.

Sections 4 and 5 deal with the issue of asymptotic efficiency of the Wald type test derived in Section 3 compared to other well known unit root tests. In Section

4, two direct methods have been used for power comparisons: the asymptotic power function and the rate of decay of the asymptotic size when power is kept fixed. In Section 5 a formal criterion of asymptotic relative efficiency due to Bahadur is introduced. To our knowledge, this is the first application of a non local measure of asymptotic relative efficiency to unit root testing. The conclusions of these relative efficiency considerations give some practical significance to the theory developed in the previous sections. With the null asymptotic distribution derived in Section 3, the Wald test based on the information matrix performs better than what is currently believed: it has the same approximate Bahadur slope as the Dickey Fuller t ratio, and outperforms commonly used tests, such as the Dickey Fuller test and the Lagrange Multiplier test defined in Solo (1984). Interestingly, these results are in conflict with some simulation studies (e.g. Dickey and Fuller, 1979), where the non studentised Dickey Fuller test is found to have better power properties than the t ratio test. This may be explained by the fact that, to a certain extent, Bahadur's approach to asymptotic relative efficiency places greater importance to minimal size rather than maximal power. All proofs are collected in Appendix B.

2. Inconsistency and asymptotic distribution of $\mathcal{I}_n(\hat{\rho}_n)$

Consider the family of processes given by

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad t = 2, ..., n; \quad \rho = 1 + \frac{c}{n^\alpha}, \ \alpha \geq 0, \ c \in \mathbb{R} \tag{3}$$

where y_1 is a constant and ε_t are i.i.d. $N(0, \sigma^2)$ random variables. The parametrisation of the autoregressive root in (3) defines a family of processes that includes various types of first order autoregression with different asymptotic behaviour. When $c = 0$, y_t is a unit root process. When $\alpha = 0$, y_t is a stable root process for $c \in (-2, 0)$ and an explosive process for $c \in (-\infty, -2) \cup (0, \infty)$. When $c \neq 0$ and $\alpha > 0$ the autoregressive root lies in a neighbourhood of unity. In this case, $\alpha = 1$ gives rise to the local to unity processes of Phillips (1987b) and Chan and Wei (1987), whereas $\alpha > 1$ implies that y_t behaves asymptotically as a unit root process even for $c \neq 0$. Finally, when $\alpha \in (0, 1)$ the autoregressive root lies in neighbourhoods of unity with radial length larger that n^{-1}. Processes with such "moderate deviations from unity" roots were discussed in recent work by Phillips and Magdalinos (2006) and Giraitis and Phillips (2006).

The log likelihood for a sample $(y_t)_{2 \leq t \leq n}$ from (3) can be written as

$$\ell(\rho, \sigma^2) = -\frac{n-1}{2} \log 2\pi - \frac{n-1}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=2}^{n} (y_t - \rho y_{t-1})^2,$$

giving rise to the maximum likelihood estimators

$$\hat{\rho}_n = \frac{\sum_{t=2}^n y_t y_{t-1}}{\sum_{t=2}^n y_{t-1}^2} \quad \text{and} \quad \hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{t=2}^{n} (y_t - \hat{\rho}_n y_{t-1})^2.$$

Assuming correct specification of the model, the top left element of the information matrix is given by

$$\mathcal{I}_n(\rho) = -E(\frac{\partial^2 \ell}{\partial \rho^2}) = \begin{cases} \frac{\rho^{2n-2}-1}{\rho^2-1}\frac{y_1^2}{\sigma^2} + \frac{\rho^{2n-2}-1}{(\rho^2-1)^2} - \frac{n-1}{\rho^2-1}, & \rho \neq 1 \\ (n-1)\frac{y_1^2}{\sigma^2} + \frac{(n-1)(n-2)}{2}, & \rho = 1. \end{cases} \tag{4}$$

Thus, the unrestricted estimator of the information is given by

$$\mathcal{I}_n(\hat{\rho}_n) = \frac{\hat{\rho}_n^{2n-2}-1}{\hat{\rho}_n^2-1}\frac{y_1^2}{\hat{\sigma}_n^2} + \frac{\hat{\rho}_n^{2n-2}-1}{(\hat{\rho}_n^2-1)^2} - \frac{n-1}{\hat{\rho}_n^2-1}. \tag{5}$$

In order to obtain the asymptotic distributions of $\hat{\rho}_n^n$ and $\mathcal{I}_n(\hat{\rho}_n)$ we need a modification of the continuous mapping theorem which deals with weak convergence of sequences of functions of a stochastic process. Let X_n, X be random variables on possibly different probability spaces for $n \in \mathbb{N}$. Denote by $\mathcal{B}$ the Borel σ-algebra over $\mathbb{R}$ and by $\mathbb{P}_X := PX^{-1}$ the distribution of X on $\mathcal{B}$. The following result is usually attributed to H. Rubin; see Kallenberg (2002, Theorem 4.27) for a proof.

2.1 Lemma. *Let h_n,$h : \mathbb{R} \to \mathbb{R}$ be Borel functions and define the Borel set*

$$E = \{x \in \mathbb{R} \mid \exists (x_n)_{n\in\mathbb{N}} \text{ such that } x_n \to x \text{ and } h_n(x_n) \nrightarrow h(x)\}.$$

If $X_n \Rightarrow X$ and $\mathbb{P}_X(E) = 0$, then $h_n(X_n) \Rightarrow h(X)$.

Lemma 2.1 is very useful in determining the asymptotic behaviour of $\hat{\rho}_n^n$ and hence that of $\mathcal{I}_n(\hat{\rho}_n)$. When the root of (3) lies in a neighbourhood of unity with $\alpha \geq 1$, $\hat{\rho}_n^n$ can be written as $\hat{\rho}_n^n = \left(1 + \frac{X_n}{n}\right)^n$, where $(X_n)_{n\in\mathbb{N}}$ is a process that converges in distribution to the random variables $\frac{1}{2}Z_1$ in the unit root case and $Z_2 + c$ in the local to unity case. Hence, in each of these cases Lemma 2.1 implies that $\hat{\rho}_n^n$ converges in distribution to the exponential of the weak limit of X_n. This gives rise to a non degenerate limit distribution for $n^{-2}\mathcal{I}_n(\hat{\rho}_n)$ when the autoregressive root is one, or lies in a viccinity of unity with radial parameter $\alpha \geq 1$. On the other hand, when $\alpha \in [0, 1)$ and $c \neq 0$ both $\hat{\rho}_n^n$ and $\mathcal{I}_n(\hat{\rho}_n)$ have, after suitable normalization, a constant probability limit as $n \to \infty$.

2.2 Lemma. *For the family of autoregressions defined by (3) with $\rho = 1 + c/n^\alpha$ we have, as $n \to \infty$*

(a) $\hat{\rho}_n^n \Rightarrow \exp(\frac{1}{2}Z_1)$, *when $c = 0$ or $\alpha > 1$.*

For $c \in \mathbb{R} \setminus \{0\}$, we have

(b) $\hat{\rho}_n^n \Rightarrow \exp(Z_2 + c)$, *when $\alpha = 1$*

(c) $\hat{\rho}_n^n \to_P 0$, *when $\alpha \in (0, 1)$ and $c < 0$, or $\alpha = 0$ and $c \in (-2, 0)$*

(d) $\rho^{-n}\hat{\rho}_n^n \to_P 1$, *when $\alpha \in (0, 1)$ and $c > 0$, or $\alpha = 0$ and $c \in (-\infty, -2) \cup (0, \infty)$*

where Z_1 and Z_2 are the random variables defined in (1) and (2).

2.3 Theorem. *For the family of autoregressions defined by (3) with $\rho = 1 + c/n^{\alpha}$, the asymptotic behaviour of $\mathcal{I}_n(\hat{\rho}_n)$ as $n \to \infty$ is given by the following.*

(a) *When $c = 0$ or $\alpha > 1$,*

$$n^{-2}\mathcal{I}_n(\hat{\rho}_n) \Rightarrow \frac{\exp(Z_1) - 1}{Z_1^2} - \frac{1}{Z_1}.$$

For $c \in \mathbb{R} \setminus \{0\}$, we have:

(b) *When $\alpha = 1$,*

$$n^{-2}\mathcal{I}_n(\hat{\rho}_n) \Rightarrow \frac{\exp\{2(Z_2 + c)\} - 1}{4(Z_2 + c)^2} - \frac{1}{2(Z_2 + c)}.$$

(c) *When $\alpha \in (0, 1)$ and $c < 0$, $n^{-(1+\alpha)}\mathcal{I}_n(\hat{\rho}_n) \to_P (-2c)^{-1}$.*

(d) *When $\alpha \in (0, 1)$ and $c > 0$, $n^{-2\alpha}\rho^{-2n}\mathcal{I}_n(\hat{\rho}_n) \to_P (4c^2)^{-1}$.*

(e) *When $\alpha = 0$ and $c \in (-2, 0)$, $n^{-1}\mathcal{I}_n(\hat{\rho}_n) \to_P (1 - \rho^2)^{-1}$.*

(f) *When $\alpha = 0$ and $c \in (-\infty, -2) \cup (0, \infty)$,*

$$\rho^{-2n}\mathcal{I}_n(\hat{\rho}_n) \to_P \frac{1}{\rho^2(\rho^2 - 1)}\frac{y_1^2}{\sigma^2} + \frac{1}{\rho^2(\rho^2 - 1)^2}.$$

2.4 Remarks.

(i) Parts (a) and (b) imply that $\mathcal{I}_n(\hat{\rho}_n)$ is inconsistent for the true information $\mathcal{I}_n(\rho)$ since, as $n \to \infty$, $n^{-2}[\mathcal{I}_n(\hat{\rho}_n) - \mathcal{I}_n(\rho)] \nrightarrow 0$ in probability. Taking limits in (4) yields $\lim_{n\to\infty} n^{-2}\mathcal{I}_n(\rho) = 1/2$ in the unit root case and

$$\lim_{n\to\infty} \frac{1}{n^2}\mathcal{I}_n(\rho) = \frac{e^{2c} - 1}{4c^2} - \frac{1}{2c}$$

in the local to unity case, which clearly do not agree with the weak limits for $n^{-2}\mathcal{I}_n(\hat{\rho}_n)$ given by parts (a) and (b) of Theorem 2.3. On the other hand, for the stationary and moderately stationary cases, $\lim_{n\to\infty} n^{-1}\mathcal{I}_n(\rho) = (1 - \rho^2)^{-1}$ and $\lim_{n\to\infty} n^{-(1-\alpha)}\mathcal{I}_n(\rho) = (-2c)^{-1}$, which coincide with parts (e) and (c) respectively. Similarly, for the explosive and moderately explosive cases, $\rho^{-2n}\mathcal{I}_n(\rho)$ and $\rho^{-2n}n^{-2\alpha}\mathcal{I}_n(\rho)$ converge to the constant probability limits of parts (f) and (d) respectively. Consequently, for the family of autoregressive processes considered in (3), a necessary and sufficient condition for consistent estimation of the information $\mathcal{I}_n(\rho)$ by $\mathcal{I}_n(\hat{\rho}_n)$ is $c \neq 0$ and $\alpha \in [0, 1)$.

(ii) The normalization of $\mathcal{I}_n(\hat{\rho}_n)$ varies continuously with the radial parameter α and covers the interval $\left[(1 + c)^{-2n}, n^{-1}\right]$, with $c \in (-\infty, -2) \cup (0, \infty)$, providing a smooth transition from explosive to unit root and stationary asymptotics.

3. Wald test

We now concentrate on autoregressive processes with a fixed root $\rho = 1 + c$ that does not depend on the sample size, i.e. processes defined by (3) with $\alpha = 0$. Given this statistical model, with ρ and σ^2 being unknown parameters, unit root tests are concerned with testing the hypothesis

$$H_0 : \rho = 1 \quad \text{versus} \quad H_1 : |\rho| < 1. \tag{6}$$

There is of course a wide range of tests for the unit root hypothesis (6) and we review some of the most commonly used in Sections 4 and 5. In this section, we discuss the following (signed square root) Wald test statistic proposed by Evans and Savin (1981) and Abadir (1993b):

$$T_n(\hat{\rho}_n) = (\hat{\rho}_n - 1)\, \mathcal{I}_n(\hat{\rho}_n)^{1/2}, \tag{7}$$

where $\mathcal{I}_n(\hat{\rho}_n)$ is given by (5). The asymptotic distribution of $T_n(\hat{\rho}_n)$ is an easy corollary of part (a) of Theorem 2.3 and the fact that $n\,(\hat{\rho}_n - 1) \Rightarrow \frac{1}{2} Z_1$.

3.1 Theorem. *As $n \to \infty$, $T_n(\hat{\rho}_n) \Rightarrow U := \frac{1}{2}\left[\exp(Z_1) - Z_1 - 1\right]^{\frac{1}{2}}$.*

By deriving the asymptotic distribution of the Wald-type statistic (7), Theorem 3.1 yields in effect a new test for the unit root hypothesis (6). It is important, therefore, to investigate its performance in terms of asymptotic power and size compared to other tests available in the literature. A direct way of conducting asymptotic power and size comparisons is to find the order of magnitude of the tails of the limiting distribution function. This approach has been followed by Abadir (1993b) and is based on his earlier result on the density and distribution functions of the Dickey Fuller distribution (cf. Abadir, 1993a). We derive similar results for U.

By equation (3.2) in Abadir (1993a) we know that the asymptotic behaviour of the density function of Z_1 is given by

$$f_{Z_1}(z) \sim \frac{1}{\sqrt{2}} \frac{e^{\frac{z}{8}}}{\sqrt{-3\pi z}} \quad \text{as } z \to -\infty, \tag{8}$$

where the asymptotic equivalence $f(x) \sim g(x)$ means $f(x)/g(x) \to 1$. In order to find the rate of decay of the density function of U, we need to consider the transformation $u = \frac{1}{2}\left(e^x - x - 1\right)^{\frac{1}{2}}$. Inverting this transformation gives

$$e^x - x - 1 = 4u^2 \quad \text{or} \quad x = -(1 + 4u^2) - \mathbb{W}\left(-e^{-(1+4u^2)}\right), \tag{9}$$

where $\mathbb{W}(\cdot)$ is the Lambert W function (cf. Appendix A) and the last equality is obtained as a consequence of equation (20). Using (19) of Appendix A, we obtain

$$\frac{dx}{du} = -8u + \frac{8u\mathbb{W}\left(-e^{-(1+4u^2)}\right)}{1 + \mathbb{W}\left(-e^{-(1+4u^2)}\right)} = -\frac{8u}{1 + \mathbb{W}\left(-e^{-(1+4u^2)}\right)}.$$

We can therefore use the transformation theorem to write the p.d.f. of U in terms of the p.d.f. of Z_1. For each $u \in \mathbb{R}$ we obtain

$$f_U(u) = \frac{8\,|u|}{|1 + \mathbb{W}\left(-e^{-(1+4u^2)}\right)|} f_{Z_1}\left[-(1+4u^2) - \mathbb{W}\left(-e^{-(1+4u^2)}\right)\right]. \tag{10}$$

We are interested in the behaviour of $f_U(u)$ as $|u| \to \infty$. By (18) of Appendix A, $\mathbb{W}(\cdot)$ is analytic on a neighbourhood around 0, so $\lim_{|u|\to\infty} \mathbb{W}\left(-e^{-(1+4u^2)}\right) = \mathbb{W}(0) = 0$. Therefore, since $-(1+4u^2) \to -\infty$ as $|u| \to \infty$, (10) and (8) imply that

$$f_U(u) \sim 8\,|u|\, f_{Z_1}\left[-(1+4u^2)\right] \sim k\phi(u) \quad \text{as } |u| \to \infty, \tag{11}$$

where $k := 4e^{-\frac{1}{8}}/\sqrt{3} = 2.038$ and $\phi(\cdot)$ is the standard normal density. Note that the tails of the p.d.f. of U are roughly twice the size of those of the standard normal density. Having established (11), it is straightforward to derive an analogous result for the distribution function of U.

3.2 Theorem. *The tail asymptotic behaviour of the distribution function of the random variable U is given by*

$$F_U(x) \sim (1-k)\mathbf{1}_{\mathbb{R}_+}(x) + k\Phi(x) \quad \text{as } |x| \to \infty,$$

where $\Phi(\cdot)$ denotes the standard normal distribution function.

4. Asymptotic properties of the Wald test

Application of the Wald statistic discussed in Section 3 requires first specifying appropriate critical regions. Since T_n diverges to $-\infty$ *a.s.* under H_1 (see (24)), a consistent test is obtained by considering critical regions of the form $CR_n(\rho) = \{T_n \leq c_n\}$, where $(c_n)_{n\in\mathbb{N}}$ is a sequence of constants. Denoting by a_n and Π_n the size and the power of the test respectively, we can write

$$a_n(\rho) = P_{H_0}\{T_n \leq c_n\} = F_U(c_n) + o(1) \quad \text{as } n \to \infty \tag{12}$$

$$\Pi_n(\rho) = P_{H_1}\{T_n \leq c_n\} = \Phi\left(c_n + \sqrt{n}\sqrt{\frac{1-\rho}{1+\rho}}\right) + o(1) \quad \text{as } n \to \infty. \tag{13}$$

In this section we derive expressions for the size and power of the Wald test of Section 3 and compare the results with other tests in the literature, using Abadir's (1993b) survey as the main reference. Three tests have been chosen to be compete against the test considered here. The first is based on the same statistic T_n but has the null asymptotic distribution given in Evans and Savin (1981) and Abadir (1993b), since $\mathcal{I}_n(\hat{\rho}_n)$ has been assumed consistent. This test is called in Abadir (1993b) "Exact

Wald" and denoted here by T_{0n}. The second test used for comparison is the Dickey Fuller t ratio:

$$T_{1n} := (\hat{\rho}_n - 1)\left(\frac{\sum_{t=2}^{n} y_{t-1}^2}{\hat{\sigma}_n^2}\right)^{\frac{1}{2}}.$$

The third test is a Lagrange Multiplier test defined in Solo (1984)

$$T_{2n} := (\hat{\rho}_n - 1)\left(\frac{\sum_{t=2}^{n} y_{t-1}^2}{\tilde{\sigma}_n^2}\right)^{\frac{1}{2}},$$

where $\tilde{\sigma}_n^2$ is the restricted maximum likelihood estimator $\tilde{\sigma}_n^2 = \frac{1}{n-1}\sum_{t=2}^{n}(y_t - y_{t-1})^2$. T_{1n} and T_{2n} have the same asymptotic distribution under the null, but since $\tilde{\sigma}_n^2$ is not consistent under H_1 (see Proposition B1 (c)), they diverge at different rates under the alternative hypothesis. In what follows, we will see that this has an impact in terms of asymptotic relative efficiency.

We start by the classical approach of fixing the size of the test to a satisfactory level and then comparing powers. For fixed size a, (12) gives rise to the critical region $CR_n(\rho, a) = \left\{T_n \leq F_U^{-1}(a) + o(1)\right\}$. Therefore, using (13) the power function becomes

$$\Pi_n(\rho, a) = \Phi\left(F_U^{-1}(a) + \sqrt{n}\sqrt{\frac{1-\rho}{1+\rho}}\right) + o(1) \quad \text{as } n \to \infty,$$

by continuity of $\Phi(\cdot)$. Therefore, (cf. Theorem 4.1 in Abadir (1993b)) the asymptotic power function of the test derived here differs from that of T_{0n}, T_{1n} and T_{2n} only through the quantile $F_U^{-1}(a)$. Since a is constant, it is clear that $F_U^{-1}(a)$ does not appear in the leading term of $\Pi_n(\rho, a)$, and all four tests are asymptotically equivalent.

Cochran (1952) has suggested an alternative approach for deciding between competing test statistics by fixing the power and comparing the rate at which the size converges to zero as $n \to \infty$. It turns out that this comparison can distinguish between the leading terms of the competing test statistics considered above. The leading terms of the asymptotic size functions of the test statistics T_{0n}, T_{1n} and T_{2n} have been obtained in Abadir (1993b, Theorem 4.3 (i)). We do the same for T_n.

For fixed power $\Pi \in (0,1)$, (13) implies that $c_n = \Phi^{-1}(\Pi) - \sqrt{n}\sqrt{\frac{1-\rho}{1+\rho}} + o(1)$ as $n \to \infty$. Hence, (12) and Theorem 3.2 give

$$\begin{aligned} a_n(\rho, \Pi) &= F_U\left(\Phi^{-1}(\Pi) - \sqrt{n}\sqrt{\frac{1-\rho}{1+\rho}}\right) + o(1) \\ &\sim k\Phi\left(\Phi^{-1}(\Pi) - \sqrt{n}\sqrt{\frac{1-\rho}{1+\rho}}\right) \quad \text{as } n \to \infty. \end{aligned}$$

Hence, the well-known asymptotic equivalence $\Phi(z) \sim -\phi(z)/z$, as $z \to -\infty$, implies that the leading term of the asymptotic size of T_n takes the form

$$a_n(\rho, \Pi) \sim \sqrt{\frac{k^2(1+\rho)}{2\pi n(1-\rho)}} \exp\left\{-n\frac{1-\rho}{2(1+\rho)}\right\} \quad \text{as } n \to \infty. \tag{14}$$

Comparing (14) with Theorem 4.3 (i) of Abadir (1993b) leads to different conclusions on the issue of asymptotic efficiency of the Wald test considered in this paper. The asymptotic size function of T_{0n}, the Wald test based on T_n when $\mathcal{I}_n(\hat{\rho}_n)$ is assumed consistent, converges to zero only at rate $n^{-\frac{1}{4}}e^{-\sqrt{n}}$. Consequently, with the null asymptotic distribution given in Theorem 3.1, the Wald test based on the information matrix performs better than what is currently believed. In fact, in the size ranking of Abadir (1993b) it moves from the last place to the first, together with the Dickey Fuller t ratio T_{1n}. Taking $k \simeq 2$, the leading term of the asymptotic size of T_{1n} is identical to the right hand side of (14). Finally, since under H_1 the coefficient of the exponential term of T_{1n} and T_n is greater in absolute value than that of T_{2n} $\left(\frac{1-\rho}{2(1+\rho)} > \frac{1-\rho}{4}\right)$, we expect that T_{1n} and T_n will outperform T_{2n}. This point will be made formal in the next section.

5. Bahadur slopes

As demonstrated by Abadir (1993b) and Section 4 above, with the notable exception of comparing asymptotic sizes, crude asymptotic power comparisons lead to asymptotically equivalent tests. Phillips (1987b) suggested a formal approach to power comparisons of unit root tests by developing an asymptotic theory for the sequence of local alternatives $\rho = 1 + c/n$, which is the basis for comparisons using Pitman's approach to asymptotic relative efficiency. This method was followed up by Nabeya and Tanaka (1990) and was successful in describing the limiting power properties of a variety of unit root tests. As in any comparison based on Pitman drifts, the relative performance of different tests is examined locally, i.e. only for alternatives that lie on an appropriate boundary of the null parameter space. No information is available on the performance of tests for the rest of H_1.

In this section we apply another formal notion of asymptotic relative efficiency, first introduced by Bahadur (1960), to the problem of unit root testing. It turns out that this approach yields simple analytical formulae that allow relative efficiency comparisons between different tests over the whole alternative parameter space. It has the additional advantage that, unlike the method of Nabeya and Tanaka (1990), it can deal with test statistics that are not the ratio of quadratic forms, such as the Dickey Fuller t ratio. We begin by giving a summary of Bahadur's approach, using Bahadur (1960, 1967), Serfling (1980) and Nikitin (1995) as the main references.

Given a sample space Ω and a parameter space Θ, consider the problem of testing a simple null hypothesis $\theta = \theta_0$ against $\theta \in \Theta \setminus \{\theta_0\}$ by using a statistic τ_n with distribution function $F_{\tau_n}(\cdot;\theta)$ when θ is the true value of the parameter. Suppose that the null hypothesis is rejected for large values of τ_n. An indicator of the significance of the observed data against the null hypothesis is given by the level attained by the test statistic. The (exact) level attained by τ_n is defined to be the random variable

$$L_n(\omega) = 1 - F_{\tau_n}(\tau_n(\omega);\theta_0) = P_{\theta_0}\{\omega' : \tau_n(\omega') > \tau_n(\omega)\}, \tag{15}$$

for $\omega, \omega' \in \Omega$. In other words, if after conducting a random experiment we observe $\tau_n(\omega)$, $L_n(\omega)$ represents the probability that a new random experiment will yield a value of τ_n greater than $\tau_n(\omega)$ when the null hypothesis is true. Thus, the level attained indicates the degree to which τ_n tends to reject the null hypothesis and between competing tests we prefer the one with the smallest level.

It is clear from the definition in (15) that calculation of the exact level attained by a test statistic τ_n requires determining the large deviation asymptotics of τ_n under H_0, i.e. the asymptotic behaviour as $n \to \infty$ of $P_{\theta_0}\{\tau_n > x\}$. In many statistical problems, including unit root testing, such large deviation results are not available. As a substitute Bahadur (1960) suggested considering instead $P\{\tau > x\}$, where τ is the weak limit of τ_n under H_0. This compromise gives rise to the concept of approximate level and Bahadur slopes. As before, between competing tests we prefer the one with the smallest approximate level and hence (see Definition 5.1) the largest approximate Bahadur slope.

5.1 Definition. *Suppose that $\tau_n \Rightarrow \tau$ under H_0. The approximate level attained by τ_n is defined to be the random variable $L_n^*(\omega) = 1 - F_\tau(\tau_n(\omega))$, where $F_\tau(\cdot)$ is the distribution function of τ. A function $c^*(\cdot)$ is called the approximate Bahadur slope of τ_n if, as $n \to \infty$,*

$$\frac{1}{n}\log L_n^* \to_{P_\theta} -\frac{1}{2}c^*(\theta) \quad \textit{for each } \theta \in \Theta \setminus \{\theta_0\}.$$

After discussing Bahadur's approach to asymptotic relative efficiency, we proceed to apply this method to the problem of testing for a unit root. The main result of this section, Theorem 5.2 below, derives the approximate Bahadur slopes of the unit root tests T_n, T_{1n} and T_{2n} defined in Section 4, as well as that of the Dickey Fuller test statistic $T_{3n} := n(\hat{\rho}_n - 1)$. The null asymptotic distributions of these test statistics can be expressed in terms of the random variables Z_1, defined in (1), and

$$Z_3 := \frac{1}{2}\frac{W^2(1) - 1}{\left(\int_0^1 W(s)^2 ds\right)^{1/2}}.$$

From Phillips (1987a) we know that $T_{3n} \Rightarrow \frac{1}{2}Z_1$ and $T_{1n}, T_{2n} \Rightarrow Z_3$ under H_0. Also, by Theorem 3.1 $T_n \Rightarrow U$ under H_0. In view of Definition 5.1, determining the approximate slopes of the test statistics considered above involves calculating the order of magnitude of the tails of the distribution functions of the limiting random variables Z_1, Z_3 and U. For U, this is done in Theorem 3.2. For $\frac{1}{2}Z_1$ and Z_3, Abadir (1993a) and Abadir (1995) have established

$$F_{\frac{1}{2}Z_1}(x) \sim \frac{2^{7/4}e^{\frac{x}{4}}}{\sqrt{-\frac{3\pi}{\sqrt{2}}x}}, \quad F_{Z_3}(x) \sim \frac{2}{\sqrt{2\pi}}\frac{e^{-\frac{x^2}{2}}}{-x} \quad \text{as } x \to -\infty. \tag{16}$$

Given Theorem 3.2 and (16), we can calculate the approximate Bahadur slopes of the test statistics T_n, T_{1n}, T_{2n} and T_{3n} by examining their asymptotic behaviour under the alternative hypothesis.

5.2 Theorem. *The test statistics T_n and T_{1n} have the same approximate Bahadur slope $c_1^*(\rho) = \frac{1-\rho}{1+\rho}$. The approximate Bahadur slope of the test statistics T_{2n} and T_{3n} is given by $c_2^*(\rho) = \frac{1-\rho}{2}$.*

5.3 Remarks.

(i) Since $c_1^*(\rho) > c_2^*(\rho)$ for all $\rho \in (-1, 1)$ the Dickey Fuller t ratio test T_{1n} and the Wald type test T_n derived in Section 3 are more efficient, in the Bahadur sense, than the non studentised Dickey Fuller test T_{3n} and the Solo LM test T_{2n}. Note that Theorem 5.2 provides a comparison which is not limited to a specific sequence of local alternatives but takes place over the whole alternative parameter space $(-1, 1)$. As expected, both $c_1^*(\rho)$ and $c_2^*(\rho)$ assume larger numerical values and thus lead to more efficient tests for alternatives ρ that lie far away from the null hypothesis $\rho = 1$.

(ii) As pointed out in Section 4, the Wald type test T_n is asymptotically more efficient than previously reported in Evans and Savin (1981) and Abadir (1993b). This is due to the null asymptotic distribution for T_n established in Section 3.

(iii) An interesting feature of Theorem 5.2 is that the Dickey Fuller t ratio test appears to be asymptotically more efficient than the non studentised Dickey Fuller test. This finding is related to the Bahadur approach of comparing slopes: since the slope of a test statistic is the probability limit of a monotonic function of the level attained, Bahadur's relative efficiency can be interpreted as a stochastic comparison between the levels attained by competing test statistics (cf. Serfling, 1980). Therefore, Bahadur's criterion for asymptotic relative efficiency examines minimal size rather than maximal power.

(iv) Since only approximate slopes have been considered, the conclusions of Theorem 5.2 are valid irrespectively of the distribution of the innovation errors in (3).

6. Conclusion

One of the standard results in classical statistical inference states that the information matrix $\mathcal{I}_n(\rho)$ can be estimated consistently by $\mathcal{I}_n(\hat{\rho}_n)$, where $\hat{\rho}_n$ is the maximum likelihood estimator of ρ. We have shown that this result does not apply to autoregressions that approach nonstationarity with rate n^{-1} or faster. In the unit root and local to unity cases $n^{-2}\mathcal{I}_n(\widehat{\rho}_n)$ is found to have a non degenerate weak limit, which contributes to the null asymptotic distribution of a Wald type test for the unit root hypothesis. This new limit result improves the performance of the above Wald test in terms of asymptotic efficiency relative to other unit root tests.

It is worth noting that the assumption of i.i.d. Gaussian innovations ε_t in (3) is not essential. Gaussianity has been assumed only in order to operate in a likelihood inference framework. The limit theory derived in this paper is invariant to the distribution of ε_t. Moreover, the main results of the paper are only slightly modified if we allow for correlated innovations. We can consider, for example, linear process errors $\varepsilon_t = \sum_{j=0}^{\infty} c_j \zeta_{t-j}$, where ζ_t is a sequence of i.i.d. $(0, \sigma^2)$ random variables and c_j is a sequence of constants satisfying $\sum_{j=1}^{\infty} j\,|c_j| < \infty$. In this case, the functional limit theory of Phillips and Solo (1992) implies that Lemma 2.2 and Theorem 2.3 continue to hold with the limiting random variables Z_1 and Z_2 replaced (respectively) by

$$Z_1' = \frac{W^2(1) - 1 + \frac{2\lambda}{\omega^2}}{\int_0^1 W(s)^2 ds} \quad \text{and} \quad Z_2' = \frac{\int_0^1 J_c(s)dW(s) + \frac{\lambda}{\omega^2}}{\int_0^1 J_c(s)^2 ds},$$

where $\omega^2 = \sigma^2 \left(\sum_{j=0}^{\infty} c_j\right)^2$ and $\lambda = \sum_{j=1}^{\infty} E\left(\varepsilon_j \varepsilon_0\right) = \sigma^2 \sum_{j=1}^{\infty} c_j \sum_{k=j+1}^{\infty} c_k$. Inference for the Wald test statistic can then be carried out after consistently estimating the nuisance parameters ω^2 and λ.

Appendix A. The Lambert W function

The Lambert W function (Corless *et al.*, 1996) is defined to be the multivalued inverse of the complex function $w \longmapsto we^w$, i.e. it is the function $\mathbb{W}$ on $\mathbb{C}$ satisfying

$$\mathbb{W}(z) \exp\{\mathbb{W}(z)\} = z. \tag{17}$$

By using the Lagrange inversion formula (see e.g. De Bruijn, 1961), we can obtain the following power series expansion for the principal branch of the Lambert W function:

$$\mathbb{W}(z) = \sum_{n=1}^{\infty} \frac{(-n)^n}{n!} z^n \quad |z| \leq \frac{1}{e}, \tag{18}$$

where the radius of convergence may be obtained using the ratio test. This implies that $\mathbb{W}$ is analytic around 0 and that $\mathbb{W}(0) = 0$.

Taking logarithms in (17), we get $\mathbb{W}(z) + \log \mathbb{W}(z) = \log z$. Differentiating the last expression with respect to z gives

$$\mathbb{W}'(z) = \frac{1}{z} \frac{\mathbb{W}(z)}{1 + \mathbb{W}(z)}. \tag{19}$$

The Lambert W function has been used in Section 3 because it makes possible the inversion of certain classes of functions. In particular, it provides an analytical solution to the equation $a^x - x = b$. From Corless *et al.* (1996),

$$a^x - x = b \Longleftrightarrow x = -b - \frac{\mathbb{W}(-a^{-b} \log a)}{\log a}. \tag{20}$$

Letting $\alpha = e$ and $b = 1 + 4u^2$ in (20) yields (9).

It is of independent interest to note that the Lambert W function has a very simple asymptotic expansion for large values of its argument. De Bruijn (1961) pp. 25-28 shows that for x real and positive,

$$\mathbb{W}(x) = \log x - \log\log x + O\left(\frac{\log\log x}{\log x}\right) \quad \text{as } x \to \infty.$$

Appendix B. Proofs

We begin by including a proposition on strong consistency of various estimators in the stable root case which facilitates the proof of Theorem 5.2.

Proposition B1. *For (3) with fixed $\rho \in (-1,1)$ the following hold as $n \to \infty$:*

(a) $\frac{1}{n}\sum_{t=2}^{n} y_{t-1}^2 \to_{a.s.} \sigma^2 \left(1-\rho^2\right)^{-1}$

(b) $\hat{\rho}_n \to_{a.s.} \rho$ *and* $\hat{\sigma}_n^2 \to_{a.s.} \sigma^2$

(c) $\tilde{\sigma}_n^2 \to_{a.s.} 2\left(1+\rho\right)^{-1}\sigma^2$

(d) $\hat{\rho}_n^n \to_{a.s.} 0$ *and* $n^{-1}\mathcal{I}_n(\hat{\rho}_n) \to_{a.s.} \left(1-\rho^2\right)^{-1}$.

Proof. Parts (a) and (b) are standard (see e.g. Brockwell and Davis, 1991). For part (c), we first show that

$$\frac{1}{n}\sum_{t=2}^{n} y_{t-1}\varepsilon_t \to_{a.s.} 0. \tag{21}$$

To see this, note that for any $t \geq 2$, $Ey_t^2 < K$, where $K := y_1^2 + \sigma^2\rho^{-2}\left(1-\rho^2\right)^{-1} < \infty$. Thus, monotone convergence yields

$$E\sum_{t=2}^{\infty} t^{-2} y_{t-1}^2 < K\sum_{t=2}^{\infty} t^{-2} < \infty,$$

implying that $\sum_{t=2}^{\infty} t^{-2} y_{t-1}^2 < \infty$ *a.s.*. Letting $\mathcal{F}_t = \sigma\left(\varepsilon_2, ..., \varepsilon_t\right)$, we obtain

$$\sum_{t=2}^{\infty} \frac{E\left(y_{t-1}^2\varepsilon_t^2 \middle| \mathcal{F}_{t-1}\right)}{t^2} = \sigma^2\sum_{t=2}^{\infty}\frac{y_{t-1}^2}{t^2} < \infty \quad a.s.$$

so (21) follows by the SLLN for martingales (Hall and Heyde, 1980, Theorem 2.18). Now $\tilde{\sigma}_n^2$ can be written as

$$\begin{aligned} \tilde{\sigma}_n^2 &= \frac{1}{n-1}\sum_{t=2}^{n}\left(y_t - y_{t-1}\right)^2 = \frac{1}{n-1}\sum_{t=2}^{n}\left[\varepsilon_t + \left(\rho-1\right)y_{t-1}\right]^2 \\ &= \frac{1}{n-1}\sum_{t=2}^{n}\varepsilon_t^2 + \frac{\left(\rho-1\right)^2}{n-1}\sum_{t=2}^{n} y_{t-1}^2 + o_{a.s.}\left(1\right) = \frac{2}{1+\rho}\sigma^2 + o_{a.s.}\left(1\right), \end{aligned}$$

by (21), part (a) and the SLLN. For part (d), strong consistency of $\hat{\rho}_n$ implies that $\log|\hat{\rho}_n| \to_{a.s.} \log|\rho| < 0$, since $|\rho| < 1$. Thus,

$$|\hat{\rho}_n^n| = \exp\{n\log|\hat{\rho}_n|\} = \exp\{n\log|\rho|\,[1+o_{a.s.}(1)]\} = o_{a.s.}(1).$$

The $a.s.$ limit of $n^{-1}\mathcal{I}_n(\hat{\rho}_n)$ follows from part (b) and the fact that $\hat{\rho}_n^n = o_{a.s.}(1)$. ■

Proof of Lemma 2.2. For the unit root case $c = 0$, we can write

$$\hat{\rho}_n^n = \left[1 + \frac{n(\hat{\rho}_n - 1)}{n}\right]^n = \left(1 + \frac{X_n}{n}\right)^n,$$

where, as $n \to \infty$, $X_n = n(\hat{\rho}_n - 1) \Rightarrow \frac{1}{2}Z_1$. For a sequence $(x_n)_{n\in\mathbb{N}}$ of real numbers, we know that, for all $x \in \mathbb{R}$,

$$x_n \to x \ \text{ implies } \ \left(1 + \frac{x_n}{n}\right)^n \to e^x. \tag{22}$$

Hence, part (a) for $c = 0$ follows by Lemma 2.1 with $h_n(x_n) = \left(1 + \frac{x_n}{n}\right)^n$, $h(x) = e^x$ and $E = \emptyset$. When $c \neq 0$ and $\alpha \geq 1$, we can write

$$\hat{\rho}_n^n = (\rho + \hat{\rho}_n - \rho)^n = \left[1 + \frac{cn^{1-\alpha} + n(\hat{\rho}_n - \rho)}{n}\right]^n = \left(1 + \frac{X_n'}{n}\right)^n$$

where, by Phillips (1987b),

$$X_n' = \begin{cases} n(\hat{\rho}_n - \rho) + o_P(1) \Rightarrow \frac{1}{2}Z_1, & \alpha > 1 \\ c + n(\hat{\rho}_n - \rho) \Rightarrow Z_2 + c, & \alpha = 1. \end{cases}$$

Thus, (22) and Lemma 2.1 yield part (a) for $\alpha > 1$ and part (b). For part (c), we know by Theorem 3.2 of Phillips and Magdalinos (2006) that $n^{\frac{1+\alpha}{2}}(\hat{\rho}_n - \rho) = O_P(1)$. Therefore, for $\alpha \in (0, 1)$ we obtain, as $n \to \infty$,

$$\begin{aligned}
\hat{\rho}_n^n &= \left[1 + \frac{c}{n^\alpha} + \frac{n^{\frac{1+\alpha}{2}}(\hat{\rho}_n - \rho)}{n^{\frac{1+\alpha}{2}}}\right]^n = \exp\left\{n\log\left[1 + \frac{c}{n^\alpha} + \frac{n^{\frac{1+\alpha}{2}}(\hat{\rho}_n - \rho)}{n^{\frac{1+\alpha}{2}}}\right]\right\} \\
&= \exp\left\{n\left[\frac{c}{n^\alpha} + \frac{n^{\frac{1+\alpha}{2}}(\hat{\rho}_n - \rho)}{n^{\frac{1+\alpha}{2}}} + O_P\left(\frac{1}{n^{2\alpha}}\right)\right]\right\} \\
&= \exp\left\{cn^{1-\alpha}\left[1 + O_P\left(\frac{1}{n^{\frac{1-\alpha}{2}\wedge\alpha}}\right)\right]\right\} = o_P(1).
\end{aligned}$$

When $\alpha = 0$, the asymptotic expansion of the logarithm of the second line above is not valid, however the required result is given by part (d) of Proposition B1.

Finally, for part (d), Theorem 4.3 of Phillips and Magdalinos (2006) implies that $n^\alpha \rho^n(\hat{\rho}_n - \rho) = O_P(1)$. Hence, proceeding as for part (c) we obtain

$$\begin{aligned} \rho^{-n}\hat{\rho}_n^n &= \left[1 + \frac{n^\alpha \rho^n(\hat{\rho}_n - \rho)}{n^\alpha \rho^{n+1}}\right]^n = \exp\left\{n \log\left[1 + \frac{n^\alpha \rho^n(\hat{\rho}_n - \rho)}{n^\alpha \rho^{n+1}}\right]\right\} \\ &= \exp\left\{O_P\left(n^{1-\alpha}\rho^{-n}\right)\right\} = 1 + o_P(1). \end{aligned}$$

Note that the asymptotic expansion of the logarithm of the second line above is valid for $\alpha = 0$, so the above argument includes the pure explosive case. ■

Proof of Theorem 2.3. Writing $\rho = 1 + \frac{c}{n^\alpha}$ we obtain

$$\begin{aligned} \hat{\rho}_n^2 - 1 &= (\hat{\rho}_n - \rho)^2 + 2(\hat{\rho}_n - \rho) + \rho^2 - 1 \\ &= 2(\hat{\rho}_n - \rho) + (\hat{\rho}_n - \rho)^2 + \frac{2c}{n^\alpha} + \frac{c^2}{n^{2\alpha}}. \end{aligned} \tag{23}$$

Thus, when $c = 0$ or $\alpha > 1$ (23) yields $n\left(\hat{\rho}_n^2 - 1\right) \Rightarrow Z_1$ and

$$\begin{aligned} \frac{1}{n^2}\mathcal{I}_n(\hat{\rho}_n) &= \frac{\hat{\rho}_n^{2n-2} - 1}{n^2(\hat{\rho}_n^2 - 1)^2} - \frac{n-1}{n^2\left(\hat{\rho}_n^2 - 1\right)} + \frac{\hat{\rho}_n^{2n-2} - 1}{n^2\left(\hat{\rho}_n^2 - 1\right)}\frac{y_1^2}{\hat{\sigma}_n^2} \\ &= \frac{\hat{\rho}_n^{2n-2} - 1}{\left[n(\hat{\rho}_n^2 - 1)\right]^2} - \frac{n-1}{n}\frac{1}{n\left(\hat{\rho}_n^2 - 1\right)} + O_P\left(\frac{1}{n}\right) \\ &\Rightarrow \frac{\exp(Z_1) - 1}{Z_1^2} - \frac{1}{Z_1}, \end{aligned}$$

by Lemma 2.2. For part (b), taking $\alpha = 1$ in (23) we obtain $n\left(\hat{\rho}_n^2 - 1\right) \Rightarrow 2(Z_2 + c)$, which together with Lemma 2.2 gives

$$\begin{aligned} \frac{1}{n^2}\mathcal{I}_n(\hat{\rho}_n) &= \frac{\hat{\rho}_n^{2n-2} - 1}{\left[n(\hat{\rho}_n^2 - 1)\right]^2} - \frac{n-1}{n}\frac{1}{n\left(\hat{\rho}_n^2 - 1\right)} + O_P\left(\frac{1}{n}\right) \\ &\Rightarrow \frac{\exp(\{2(Z_2 + c)\} - 1}{4(Z_2 + c)^2} - \frac{1}{2(Z_2 + c)}, \end{aligned}$$

thus showing part (b). For part (c), $\hat{\rho}_n - \rho = O_P\left(n^{-\frac{1+\alpha}{2}}\right)$ by Phillips and Magdalinos (2006), so (23) yields $n^\alpha\left(\hat{\rho}_n^2 - 1\right) = 2c + o_P(1)$ and

$$\frac{1}{n^{1+\alpha}}\mathcal{I}_n(\hat{\rho}_n) = -\frac{n-1}{n}\frac{1}{n^\alpha\left(\hat{\rho}_n^2 - 1\right)} + O_P\left(\frac{1}{n^{1-\alpha}}\right) = \frac{1}{-2c} + o_P(1).$$

For part (d), (23) yields $n^\alpha\left(\hat{\rho}_n^2 - 1\right) = 2c + O_P(n^{-\alpha})$, which together with Lemma 2.2 imply

$$\begin{aligned} \frac{\rho^{-2n}}{n^{2\alpha}}\mathcal{I}_n(\hat{\rho}_n) &= \frac{\rho^{-2n}\left(\hat{\rho}_n^{2n-2} - 1\right)}{n^{2\alpha}(\hat{\rho}_n^2 - 1)^2} + O_P\left(\frac{1}{n^\alpha}\right) \\ &= [1 + o_P(1)]\frac{\left(\rho^{-n}\hat{\rho}_n^n\right)^2}{\left[n^\alpha(\hat{\rho}_n^2 - 1)\right]^2} \to_P \frac{1}{4c^2}. \end{aligned}$$

Part (e) follows from Proposition B1 (d). For part (f), Lemma 2.2 and consistency of $\hat{\rho}_n$ and $\hat{\sigma}_n^2$ yield

$$\begin{aligned}\rho^{-2n}\mathcal{I}_n(\hat{\rho}_n) &= \frac{\left(\rho^{-n}\hat{\rho}_n^n\right)^2\hat{\rho}_n^{-2}}{\hat{\rho}_n^2-1}\frac{y_1^2}{\hat{\sigma}_n^2}+\frac{\left(\rho^{-n}\hat{\rho}_n^n\right)^2\hat{\rho}_n^{-2}}{(\hat{\rho}_n^2-1)^2}+O_P\left(n\rho^{-2n}\right)\\ &= \frac{1}{\rho^2\left(\rho^2-1\right)}\frac{y_1^2}{\sigma^2}+\frac{1}{\rho^2(\rho^2-1)^2}+o_P(1).\end{aligned}$$

This completes the proof of the theorem. ■

Proof of Theorem 3.1. Theorem 2.3 (a) and the fact that $n\left(\hat{\rho}_n-1\right)\Rightarrow\frac{1}{2}Z_1$ give

$$T_n(\hat{\rho}_n)=n\left(\hat{\rho}_n-1\right)\left[\frac{1}{n^2}\mathcal{I}_n(\hat{\rho}_n)\right]^{1/2}\Rightarrow\frac{1}{2}Z_1\left[\frac{\exp(Z_1)-1}{Z_1^2}-\frac{1}{Z_1}\right]^{1/2}=U. \quad ■$$

Proof of Theorem 3.2. First, consider the case $x\to\infty$. Letting $k=4e^{-\frac{1}{8}}/\sqrt{3}$,

$$\lim_{x\to\infty}\frac{1-F_U(x)}{1-\Phi(x)}=\lim_{x\to\infty}\frac{\int_x^\infty f_U(u)du}{\int_x^\infty\phi(u)du}=\lim_{x\to\infty}\frac{f_U(x)}{\phi(x)}=k,$$

by (11). Thus, $1-F_U(x)\sim k\left[1-\Phi(x)\right]$ or $F_U(x)\sim 1-k+k\Phi(x)\ \ (x\to\infty)$. An identical argument yields $F_U(x)\sim k\Phi(x)$ as $x\to-\infty$. ■

Proof of Theorem 5.2. We begin by examining the asymptotic behaviour of the various test statistics under $H_1:\rho\in(-1,1)$. T_n can be written as

$$\begin{aligned}\frac{1}{\sqrt{n}}T_n &= -(1-\rho)\sqrt{\frac{1}{n}\mathcal{I}_n(\hat{\rho}_n)}+(\hat{\rho}_n-\rho)\sqrt{\frac{1}{n}\mathcal{I}_n(\hat{\rho}_n)}\\ &= -\sqrt{\frac{1-\rho}{1+\rho}}+o_{a.s.}(1)\quad\text{as } n\to\infty,\end{aligned}\tag{24}$$

by parts (b) and (d) of Proposition B1. A similar calculation for the Dickey Fuller t ratio T_{1n} and Solo's LM test T_{2n} yields, for each fixed $\rho\in(-1,1)$,

$$\begin{aligned}\frac{1}{\sqrt{n}}T_{1n} &= (\rho-1)\left(\frac{\frac{1}{n}\sum_{t=2}^n y_{t-1}^2}{\hat{\sigma}_n^2}\right)^{\frac{1}{2}}+(\hat{\rho}_n-\rho)\left(\frac{\frac{1}{n}\sum_{t=2}^n y_{t-1}^2}{\hat{\sigma}_n^2}\right)^{\frac{1}{2}}\\ &= -(1-\rho)\left(\frac{1}{1-\rho^2}\right)^{\frac{1}{2}}+o_{a.s.}(1)=-\sqrt{\frac{1-\rho}{1+\rho}}+o_{a.s.}(1),\end{aligned}\tag{25}$$

by parts (a) and (b) of Proposition B1, and

$$\frac{1}{\sqrt{n}}T_{2n}=(\rho-1)\left(\frac{\frac{1}{n}\sum_{t=2}^n y_{t-1}^2}{\tilde{\sigma}_n^2}\right)^{\frac{1}{2}}+o_{a.s.}(1)=-\sqrt{\frac{1-\rho}{2}}+o_{a.s.}(1),\tag{26}$$

by Proposition B1 (b) and (c). (24) (25) and (26) imply that $T_n, T_{1n}, T_{2n} \to -\infty$ $a.s.$ under H_1. This means that H_0 in (6) is rejected for small values of the test statistics, so we need to replace τ_n and τ in Definition 5.1 with $-\tau_n$ and $-\tau$ respectively. With this modification, the approximate level attained by T_{1n} is given by

$$L_{1n}^* = 1 - F_{-Z_3}(-T_{1n}) = F_{Z_3}(T_{1n}) \sim \frac{2}{\sqrt{2\pi}} \frac{\exp\left\{-\frac{1}{2}T_{1n}^2\right\}}{-T_{1n}} \quad a.s.$$

by (16), since $T_{1n} \to -\infty$ $a.s.$ as $n \to \infty$. Thus, under H_1, we obtain $a.s.$ as $n \to \infty$

$$\begin{aligned} \frac{1}{n}\log L_{1n}^* &\sim -\frac{1}{2}\frac{1}{n}T_{1n}^2 - \frac{1}{n}\log\left(-T_{1n}\right) + O\left(\frac{1}{n}\right) \\ &= -\frac{1}{2}\left(\frac{1}{\sqrt{n}}T_{1n}\right)^2 + O\left(\frac{\log n}{n}\right), \end{aligned}$$

so (25) gives for each $\rho \in (-1, 1)$

$$\frac{1}{n}\log L_{1n}^* \to_{a.s.} -\frac{1}{2}\frac{1-\rho}{1+\rho} \quad \text{as } n \to \infty.$$

Thus, the approximate Bahadur slope of T_{1n} is given by $c_1^*(\rho) = \frac{1-\rho}{1+\rho}$. Since T_{2n} has the same null asymptotic distribution and rate of divergence under H_1 as T_{1n}, an identical argument yields $\frac{1}{n}\log L_{2n}^* \sim -\frac{1}{2}\left(\frac{1}{\sqrt{n}}T_{2n}\right)^2$ as $n \to \infty$ $a.s.$, so (26) implies

$$\frac{1}{n}\log L_{2n}^* \to_{a.s.} -\frac{1}{2}\frac{1-\rho}{2}, \quad \rho \in (-1, 1).$$

This shows that the approximate Bahadur slope of T_{2n} is given by $c_2^*(\rho) = \frac{1-\rho}{2}$.

For the Dickey Fuller test $T_{3n} = n(\hat{\rho}_n - 1)$ we obtain, for each $\rho \in (-1, 1)$, $\frac{1}{n}T_{3n} \to_{a.s.} -(1-\rho)$. The approximate level attained by T_{3n} is given by

$$L_{3n}^* = 1 - F_{-\frac{1}{2}Z_1}(-T_{3n}) = F_{\frac{1}{2}Z_1}(T_{3n}) \sim \frac{2^{7/4}\exp\left\{\frac{1}{4}T_{3n}\right\}}{\sqrt{-\frac{3\pi}{\sqrt{2}}T_{3n}}} \quad \text{as } n \to \infty \ a.s.$$

by (16), since $T_{3n} \to -\infty$ $a.s.$. Thus, we obtain, for any $\rho \in (-1, 1)$,

$$\frac{1}{n}\log L_{3n}^* \sim \frac{1}{4}\frac{1}{n}T_{3n} + O\left(\frac{\log n}{n}\right) \to -\frac{1}{2}\frac{1-\rho}{2} \quad \text{as } n \to \infty \ a.s.$$

Finally, for the Wald type test statistic T_n, the approximate level attained is given by

$$L_n^* = 1 - F_{-U}(-T_n) = F_U(T_n) \sim k\Phi\left(T_n\right) \quad \text{as } n \to \infty \ a.s.$$

by Theorem 3.2. Thus, as in the case of T_{1n}, the level attained is proportional to the standard normal distribution function evaluated at the test statistic. Since T_{1n} and T_n have the same asymptotic behaviour under H_1 (compare (25) and (24)) they have the same approximate Bahadur slope. ■

References

Abadir, K.M. (1993a). The limiting distribution of the autocorrelation coefficient under a unit root. *Annals of Statistics* 21, 1058-1070.

Abadir, K.M. (1993b). On the asymptotic power of unit root tests. *Econometric Theory* 9, 189-221.

Abadir, K.M. (1995). The limiting distribution of the t ratio under a unit root, *Econometric Theory* 11, 775-793.

Bahadur, R.R. (1960). Stochastic comparison of tests, *Annals of Mathematical Statistics* 31, 276-295.

Bahadur, R.R. (1967). Rates of convergence of estimates and test statistics. *Annals of Mathematical Statistics* 38, 303-324.

Brockwell P.J. & R.A. Davis (1991). *Time Series: Theory and Methods*, 2nd ed., Springer Series in Statistics.

Chan, N.H. and C.Z. Wei (1987). Asymptotic inference for nearly nonstationary AR(1) processes, *Annals of Statistics* 15, 1050-1063.

Cochran, W.G. (1952). The χ^2-goodness of fit. *Annals of Mathematical Statistics* 23, 315-345.

Corless R.M., G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey, D.E. Knuth (1996). On the Lambert W function. *Advances in Computational Mathematics* vol. 5, 329-359.

De Bruijn, N.G. (1961). *Asymptotic methods in analysis*. New York, Dover.

Dickey, D.A. and W.A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association* 74, 427–431.

Evans, G.B.A. & N.E. Savin (1981). Testing for unit roots: 1. *Econometrica* 49, 753-779.

Giraitis L. and P. C. B. Phillips (2006). "Uniform limit theory for stationary autoregression". *Journal of Time Series Analysis* (forthcoming).

Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and its Application*. Academic Press.

Kallenberg, O. (2002). *Foundations of Modern Probability*. Springer.

Nabeya, S. and K. Tanaka (1990). Limiting powers of unit root tests in time series regression, *Journal of Econometrics* 46, 247–271.

Nikitin, Y. (1995). *Asymptotic Efficiency of Non Parametric Tests.* Cambridge University Press.

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

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

Phillips, P. C. B. and T. Magdalinos (2006), Limit theory for moderate deviations from a unit root. *Journal of Econometrics* (forthcoming).

Phillips, P. C. B. and V. Solo (1992), "Asymptotics for Linear Processes," *Annals of Statistics* 20, 971–1001.

Serfling R. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley.

Solo, V. (1984). The order of differencing in ARIMA models. *Journal of the American Statistical Association.* 79, 916-921.

White, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case, *Annals of Mathematical Statistics* 29, 1188–1197.