Econometric Theory, **19**, 2003, 885–922. Printed in the United States of America.
DOI: 10.1017/S0266466603196016

AR(1) MODELS, UNIT ROOTS, AND ADJUSTED PROFILE LIKELIHOOD

PEKKA PERE
University of Helsinki

An unobserved components and a conventional first-order autoregressive (AR) model with constant are analyzed with the adjusted profile likelihood of Cox and Reid (1987, *Journal of the Royal Statistical Society, Series B* 49, 1–39; 1993, *Journal of the Royal Statistical Society, Series B* 55, 467–71). Both the unobserved components model and the Cox–Reid adjustment can provide more accurate estimates of an AR coefficient of unity. The unobserved components model yields more powerful unit-root tests. In general the most powerful test utilizes the adjustment. Under the unobserved components model the adjusted statistics follow the asymptotic distributions better than the unadjusted.

1. INTRODUCTION

The adjusted profile (concentrated) likelihoods (APLs) of Cox and Reid (1987, 1993) aim to decrease distortions caused by estimation of nuisance parameters when interest focuses on a single parameter.[1] The description exactly matches the typical setup in unit-root testing, so APL should be of interest in the analysis of unit roots. The suitability of the theory is not apparent, however, because it has been set out with a stationary world in mind in the first place and the theoretical background for the nonstationary case is unexplored. Nevertheless, the Cox–Reid adjusted statistics, for a conventional potentially nonstationary first-order autoregressive (AR(1)) model with constant, behave sensibly, implying that the theory may be of use in the nonstationary case, too (Pere, 1997, 2000).

The emphasis is on applying the APL to an unobserved components, nonstationary AR(1) process with constant analyzed by Bhargava (1986). The model (or close relatives of it) has become popular recently (cf. Ahn, 1993; Andrews, 1993; Elliot, Rothenberg, and Stock, 1996; Lee and Schmidt, 1994; Pantula, Gonzalez-Farias, and Fuller, 1994; Robinson, 1994; Saikkonen and Luukkonen, 1997; Schmidt and Lee, 1991; Schmidt and Phillips, 1992; Zivot, 1994). An attractive property of the model is that the parameters are orthogonal in a

The basis of this research is my D. Phil. thesis (1997) for the University of Oxford. I am grateful to David Hendry for supervision and to Pentti Saikkonen for most helpful advice. I appreciate also the CBRT Young Economist Award with which I was honored after presenting a previous version at the 1998 erc/METU International Conference on Economics. I remain responsible for any errors. Financial support by the Yrjö Jahnsson Foundation and Academy of Finland is gratefully acknowledged. Address correspondence to: Pekka Pere, Department of Economics, P.O. Box 54, 00014 University of Helsinki, Finland; email: Pekka.Pere@Helsinki.FI.

© 2003 Cambridge University Press 0266-4666/03 $12.00

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

sense to be defined in Section 4, which is beneficial when applying the Cox–Reid theory. The model is compared to a conventional AR(1) with or without a constant in terms of interpretation; accuracy of estimation; and, in the case of the former model, power of unit-root tests. The latter model is excluded from the power comparison because the associated tests are not invariant with respect to the initial value of the time series under this model and the null hypothesis of a unit root.

The AR(1) models are introduced in Section 2. Unit-root asymptotics for the maximum likelihood estimate (MLE), Wald, and likelihood ratio (LR) statistics, with some other likelihood results, mainly for the unobserved components model, are given in Section 3. The profile likelihood (PL) is convenient in these derivations. The APL is presented in Section 4. The adjusted statistics and their unit-root asymptotics are then derived for the unobserved components model in Section 5. Corresponding results, originally derived in Pere (1997, 2000), are presented in Section 6 for the conventional model. Small-sample distributions and the accuracy of the estimators are compared by simulation in Sections 7, 8, and 9. Sections 10 and 11 concern near unit-root asymptotics. These results could be used to characterize the power of the new tests in relation to the asymptotic power envelope developed in Elliot et al. (1996). Sections 12–14 pertain to the power of the new and established unit-root tests. The new tests include those based on the Cox–Reid adjusted statistics for the conventional model. Finally, estimation and testing are discussed in Section 15. Proofs and details of the simulations are in the Appendixes.

Symbols for models, tests, and (nonprofile) estimates of the AR coefficient are tagged by superscripts (B, μ, or 0) according to the model in question. A subscript (ml or a) indicates the type of statistic involved. Symbols of general meaning or unique to a model and intermediate statistics (concerning the unobserved components model) are not tagged. Whether a statistic is calculated for observations 0 to T or 1 to T is not indicated by a tag either, but the case is made clear in the text.

2. ON AR(1) MODELS

The model favored by Bhargava (1986) exploits an elegant manner of adding a constant to an AR(1) model:

$$y_t = \gamma + x_t,$$

$$x_t = \psi x_{t-1} + \epsilon_t,$$

or

$$y_t = \begin{cases} \gamma + \psi x_0 + \epsilon_1 & \text{for } t = 1 \\ \gamma(1-\psi) + \psi y_{t-1} + \epsilon_t & \text{for } t = 2,\ldots,T. \end{cases} \quad \textbf{(1)}$$

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

Here $\epsilon_t \sim \text{NID}(0, \sigma^2)$ (normally independently distributed), $\sigma^2 > 0$, and $x_0 = 0$. The Cox–Reid (1993) notation is used, so the nuisance parameters or the constant and the innovation variance are denoted by $\boldsymbol{\phi} = [\gamma\ \sigma^2]'$ and the parameter of interest or the autoregressive (AR) coefficient by ψ. The focus is on the unit-root value $\psi = 1$, but values $|\psi| < 1$ are considered, too. This unobserved components model is labeled $\text{AR}^B(1)$. An intriguing feature of the model is that the unconditional mean of the process is γ, so the interpretation of the constant does not depend on whether the popular null hypothesis of a unit root is valid or not. This is a major difference from the conventional model described next (cf. (3)).

Handling of the first observation differentiates the $\text{AR}^B(1)$ model from that of the more usual formulation:

$$y_t = \alpha + \psi y_{t-1} + \epsilon_t \quad \text{for } t = 1, \ldots, T, \tag{2}$$

where y_0 is a fixed initial value. The complication here is that the constant (α) becomes a drift under $\psi = 1$, making the customary Dickey–Fuller (1979) unit-root tests, based on (2), invalid. To induce a nondrifted random walk under $\psi = 1$, the extra assumption of $\alpha = 0$ under $\psi = 1$ should be made. Model (2) will carry the tag $\text{AR}^{\mu}(1)$.

The first model corresponds to the process

$$y_t = \begin{cases} \psi^t x_0 + \gamma + \sum_{i=1}^{t} \psi^{t-i} \epsilon_i & \text{if } |\psi| < 1 \\ x_0 + \gamma + \sum_{i=1}^{t} \epsilon_i & \text{if } \psi = 1 \end{cases} \tag{3}$$

instead of

$$y_t = \begin{cases} \psi^t y_0 + \alpha \dfrac{1 - \psi^t}{1 - \psi} + \sum_{i=1}^{t} \psi^{t-i} \epsilon_i & \text{if } |\psi| < 1 \\ y_0 + \alpha t + \sum_{i=1}^{t} \epsilon_i & \text{if } \psi = 1, \end{cases} \tag{4}$$

which results from the latter model. The initial value x_0 occurs in (1) and (3) to clarify the potential role of x_0 and for the invariance and similarity results in Appendix C. The MLE for ψ should in general take x_0 into account if it were allowed to be nonzero. Empirically x_0 would be likely to be unobserved and difficult to estimate (Schmidt and Phillips, 1992). Hence $x_0 = 0$ is assumed, except that the invariance and similarity results entail that the unit-root fractiles in Section 7 apply also when $x_0 \neq 0$.

The first process is assumed to start from the neighborhood of its long-run mean (γ, assuming that σ^2 is not large relative to γ). The assumption is evident if the process is nonstationary ($\psi = 1$, in which case the process shrinks to

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

a random walk that has started from γ). A similar assumption is often also made when a time series has been in operation for some time and the process is known to be stable ($|\psi| < 1$). The model gains credibility when the process truly originates at $t = 1$ as opposed to this being the moment to begin recording of the time series.

Models (1) and (2) yield the same process in two circumstances:

- under a unit root and zero constant (α) for the $AR^{\mu}(1)$ process, γ then being understood to account for y_0 of the $AR^{\mu}(1)$ process;
- under stationarity when the processes share a common AR coefficient and long-run mean and the $AR^{\mu}(1)$ process is initialized from it so that $\gamma = \alpha/(1 - \psi) = y_0$.

The $AR^{B}(1)$ process lies in a sense between the $AR^{\mu}(1)$ and the stationary AR(1) process with constant where the first observation already follows the stationary distribution. A feature shared with the former is the nonconstancy of the variance of y_t and the constancy of the mean with the latter. If $\gamma = \alpha/(1 - \psi)$ and $|\psi| < 1$, then the differences between the $AR^{B}(1)$ and $AR^{\mu}(1)$ processes, if any, diminish for large T, and the processes unify with the stationary process. The $AR^{B}(1)$ process shares perhaps more characteristics with the $AR^{\mu}(1)$ than with the stationary process: the latter cannot accommodate $\psi = 1$, but the $AR^{\mu}(1)$ and $AR^{B}(1)$ processes can. The variance of y_t ($\sigma^2(1 - \psi^{2t})/(1 - \psi^2)$ or $\sigma^2 t$ if $\psi = 1$) is the same under the $AR^{\mu}(1)$ and $AR^{B}(1)$ processes. If $x_0 \neq 0$ were allowed for in (1), the mean of y_t would exhibit start-up effects and the common feature of the $AR^{B}(1)$ and the stationary process would disappear. As a whole the $AR^{B}(1)$ process can be comprehended as an AR(1) process with constant $\gamma(1 - \psi)$ and start-up effects altering the variance. Furthermore, if $\gamma = 0$ then the process shrinks to the AR(1) without a constant

$$y_t = \psi y_{t-1} + \epsilon_t \quad \text{for } t = 1, \ldots, T \tag{5}$$

or (2) with $\alpha = y_0 = 0$. This model, but allowing for $y_0 \neq 0$, will be denoted by $AR^{0}(1)$. However, under this model the MLE of ψ will be calculated conditional on y_1 for reasons explained next.

Even though the conditions for the processes (1), (2), and (5) to agree are met, there are aspects to take account of in choosing the model to use. If the time series is due to the $AR^{B}(1)$ process (with a potentially nonzero constant) then the $AR^{0}(1)$ or $AR^{\mu}(1)$ model could be applied, too. With them the analysis would have to be conducted conditional on the first observation, so one would have one observation less for estimation. If the data follow an $AR^{0}(1)$ or $AR^{\mu}(1)$ process, the $AR^{B}(1)$ model would be valid (for $y_1, \ldots, y_T$ leaving y_0 aside) with an equal number of observations available for estimation. However, if y_0 is known to be the long-run mean, so that the processes can match in general, then one can simply deduct it from the time series and apply the parsimonious $AR^{0}(1)$ model. Therefore later, when the models are compared by simulation, the data are generated by the $AR^{B}(1)$ model and ψ is estimated conditional on y_1 when the $AR^{\mu}(1)$ or $AR^{0}(1)$ model is used.

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

The potential benefit from using the $AR^B(1)$ model is not confined to avoiding losing an observation for conditioning. The $AR^0(1)$ and $AR^B(1)$ models have much in common in terms of asymptotics as shown next.

3. PROFILE LIKELIHOOD ANALYSIS AND UNIT-ROOT ASYMPTOTICS FOR THE $AR^B(1)$ MODEL

Pantula et al. (1994) refer to unpublished proofs for the asymptotic unit-root distributions of this section. Proofs introducing formulas needed for the APL and demonstrating the convenience of the PL are given subsequently.

The first observation is treated separately in the log-likelihood under the $AR^B(1)$ model:

$$l(\psi,\gamma,\sigma^2;\mathbf{y}) = -\frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left\{(y_1-\gamma)^2 + \sum_{t=2}^{T}[y_t-\gamma(1-\psi)-\psi y_{t-1}]^2\right\}.$$

The PL has the standard appearance at the outset:

$$l_p(\psi) = -\frac{T}{2}\log\hat{\sigma}_\psi^2, \tag{6}$$

but

$$\begin{aligned}\hat{\sigma}_\psi^2 &= T^{-1}\left\{(y_1-\hat{\gamma}_\psi)^2 + \sum_{t=2}^{T}[y_t-\hat{\gamma}_\psi(1-\psi)-\psi y_{t-1}]^2\right\}\\ &= T^{-1}\left\{\sum_{t=1}^{T}y_t^2 - 2\psi\sum_{t=2}^{T}y_t y_{t-1} + \psi^2\sum_{t=2}^{T}y_{t-1}^2 - \frac{\left[y_1+(1-\psi)\left(\sum_{t=2}^{T}y_t - \psi\sum_{t=2}^{T}y_{t-1}\right)\right]^2}{1+(T-1)(1-\psi)^2}\right\}\end{aligned} \tag{7}$$

and

$$\hat{\gamma}_\psi = \frac{y_1+(1-\psi)\sum_{t=2}^{T}(y_t-\psi y_{t-1})}{1+(T-1)(1-\psi)^2}. \tag{8}$$

The formula for $\hat{\sigma}_\psi^2$ reveals the particular handling of the first observation. In the unit-root case, it alone carries information on γ. Indeed, $\hat{\gamma}_\psi$ equals y_1 when evaluated at $\psi = 1$, so the MLE of γ ($\hat{\gamma}_{ml}$) is a kind of dummy that is in general inconsistent for γ under $\psi = 1$ (consistency of the MLE of ψ—denoted $\hat{\psi}_{ml}^B$ for

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

this model—is shown subsequently). The model has some resemblance to the Neyman–Scott (1948) incidental parameters model, to which APL has been successfully applied (Cox and Reid, 1993): in both models a sole observation carries all of the information on a nuisance parameter, yielding an inconsistent MLE. By contrast, the analogous estimate for α equals $(y_T - y_0)/T$, or a drift is sought under model (2) and $\psi = 1$.

The likelihood equation for ψ,

$$\frac{\partial l_p(\psi)}{\partial \psi} = -\frac{T(\partial \hat{\sigma}_\psi^2/\partial \psi)}{2\hat{\sigma}_\psi^2} = 0, \tag{9}$$

is equivalent to a fifth-order polynomial in ψ, so analytical solutions do not necessarily exist. The formula for $\partial \hat{\sigma}_\psi^2/\partial \psi$ is given in Appendix A. The equation could still be solved or the PL maximized numerically to find $\hat{\psi}_{ml}^B$. The equation has at least one real root because the order of the polynomial is odd.

The MLE for a known γ is

$$\hat{\psi}_\gamma = \frac{\sum_{t=2}^{T} (y_t - \gamma)(y_{t-1} - \gamma)}{\sum_{t=2}^{T} (y_{t-1} - \gamma)^2}. \tag{10}$$

The PL is helpful in proving that $\hat{\gamma}_{ml}$ tends in probability to y_1 or is $O_p(1)$ when $\psi = 1$ (Appendix A). The $O_p(1)$ness of $\hat{\gamma}_{ml}$ and the well-known orders of magnitude and convergencies of expressions such as $T^{-2}\sum_{t=2}^{T} y_{t-1}^2$ suffice to prove that

$$\begin{aligned} T(\hat{\psi}_{ml}^B - 1) &= T\left[\frac{\sum_{t=2}^{T} (y_t - \hat{\gamma}_{ml})(y_{t-1} - \hat{\gamma}_{ml})}{\sum_{t=2}^{T} (y_{t-1} - \hat{\gamma}_{ml})^2} - 1\right] \\ &= T\left[\frac{\sum_{t=2}^{T} y_t y_{t-1} - \sum_{t=2}^{T} y_{t-1}^2 - \hat{\gamma}_{ml}(y_T - y_1)}{\sum_{t=2}^{T} y_{t-1}^2 - 2\hat{\gamma}_{ml}\sum_{t=2}^{T} y_{t-1} + (T-1)\hat{\gamma}_{ml}}\right] \\ &= \frac{T^{-1}\sum_{t=2}^{T} y_{t-1}\epsilon_t}{T^{-2}\sum_{t=2}^{T} y_{t-1}^2 - O_p(T^{-1/2}) + O_p(T^{-1})} - O_p(T^{-1/2}) \\ &\Rightarrow \frac{\frac{1}{2}\{[B(1)]^2 - 1\}}{\int [B(r)]^2\, dr} \end{aligned} \tag{11}$$

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

under $\psi = 1$ and regardless of γ. Here $\Rightarrow$ signifies weak convergence, $B(r)$ is standard Wiener process with variance r, and $\int$ denotes integral over (0,1).

Remarkably, the $AR^B(1)$ model and the assumption $\psi = 1$ alone—with no restrictions on the constant—yield a (standardized) MLE with the same asymptotic distribution as $\hat{\psi}^0_{ml}$ or the MLE of ψ for the $AR^0(1)$ model (e.g., Dickey and Fuller, 1979). Heuristically, the MLE picks $\psi = 1$ and $\gamma(1 - \psi) = 0$ or the random-walk process as T increases. This is in sharp contrast with the $AR^\mu(1)$ model where α remains a nuisance even asymptotically. The corresponding asymptotic distribution, under $\psi = 1$ and $\alpha = 0$, is then (11) with the integrands replaced by $B(r) - \int B(s)\, ds$ (see (24)) (Phillips and Perron, 1988). This MLE is far more biased and scattered than the MLE(s) associated with the distribution in (11) (e.g., Pere, 2000). Moreover, if α is not zero then the asymptotics changes (Dickey, 1976, pp. 58–61; West, 1988). A conjecture is that the bias and variance of $\hat{\psi}^B_{ml}$ remain smaller for a finite but reasonable T. Despite the common unit-root asymptotics of $\hat{\psi}^B_{ml}$ and $\hat{\psi}^0_{ml}$, it does not follow that the two MLEs could not take very different values (Section 7).

Parallel results can be derived for the (signed square root of the) Wald statistic for the null hypothesis $\psi = \psi_0$:

$$\sqrt{W^B_{ml}} = (\hat{\psi}^B_{ml} - \psi_0)\sqrt{-\frac{\partial^2 l_p(\psi)}{\partial \psi^2}\bigg|_{\psi=\hat{\psi}^B_{ml}}} \tag{12}$$

(Seber and Wild, 1989, pp. 39–42). An analytic expression of $\sqrt{W^B_{ml}}$ cannot be given, the analytic form of the MLE being unknown. However, the numerical value, for a specific sample, can be exposed straightforwardly after substituting the numerical value of the MLE into the formulas defining $\sqrt{W^B_{ml}}$. The details are in Appendix A.

Under the null of a unit root, $\sqrt{W^B_{ml}}$ tends to a Dickey–Fuller distribution:

$$\sqrt{W^B_{ml}} \Rightarrow \frac{\frac{1}{2}\{[B(1)]^2 - 1\}}{\left\{\int [B(r)]^2\, dr\right\}^{1/2}}$$

(Appendix A). This is the asymptotic distribution that arises also when the analogous Wald statistic is calculated from the $AR^0(1)$ model; again the $AR^\mu(1)$ model would yield a different distribution or (24) (Dickey and Fuller, 1979).

A comparable LR statistic is

$$LR^B_{ml} = \left\{\frac{\exp[l_p(\psi)]|_{\psi=\hat{\psi}^B_{ml}}}{\exp[l_p(\psi)]|_{\psi=1}}\right\}^{2/T} = \frac{\hat{\sigma}^2_\psi|_{\psi=1}}{\hat{\sigma}^2_\psi|_{\psi=\hat{\psi}^B_{ml}}}. \tag{13}$$

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

The asymptotic unit-root distribution of this statistic, after standardization, is

$$T(LR_{ml}^{B}-1) \Rightarrow \frac{\frac{1}{4}\{[B(1)]^2-1\}^2}{\int [B(r)]^2\,dr}$$

(Appendix A). It can be shown by standard unit-root asymptotic results that this is the distribution of the corresponding LR statistic, $T(LR_{ml}^{0}-1)$, under the $AR^0(1)$ model also (proof omitted). The $T(LR_{ml}^{0}-1)$ statistic essentially equals the usual $F=t^2$ statistic, with obvious notation, for the null of a unit root. It gives intuition for the fact that the preceding asymptotic distribution of $T(LR_{ml}^{B}-1)$ is the square of that of $\sqrt{W_{ml}^{B}}$.

The agreement of the asymptotic unit-root distributions of the statistics extends to the case of a near unit root (Section 10). By simple arguments, the agreement holds also in the stable case (for the MLE and Wald statistics considered by the author).

4. ADJUSTED PROFILE LIKELIHOOD

In this section, ψ and $\boldsymbol{\phi}$ are considered as a general parameter of interest and a vector of nuisance parameters, respectively. The APL of Cox and Reid (1993) is

$$l_a(\psi) = l_p(\psi) - \frac{1}{2}\log|\mathbf{J}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\psi,\hat{\boldsymbol{\phi}}_{\psi})| + \hat{c}(\psi-\hat{\psi}_{ml}). \tag{14}$$

Here $\mathbf{J}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\psi,\hat{\boldsymbol{\phi}}_{\psi})$ is the $k\times k$ observed information matrix for $\boldsymbol{\phi}$ or $-\partial^2 l(\psi,\boldsymbol{\phi})/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'$ (with (i,j)th element $j_{\phi_i\phi_j}$) evaluated at $\hat{\boldsymbol{\phi}}_{\psi}$, the coefficient c is defined by

$$c \equiv \mathrm{tr}\left(\frac{\partial}{\partial\boldsymbol{\phi}}(\mathbf{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1}\mathbf{i}_{\boldsymbol{\phi}\psi})'\right), \tag{15}$$

its MLE being $\hat{c}=c|_{(\hat{\psi}_{ml},\hat{\boldsymbol{\phi}}_{ml})}$ or c evaluated at the MLEs of ψ and $\boldsymbol{\phi}$, $\mathbf{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}$ is the Fisher information matrix with (i,j)th component $i_{\phi_i\phi_j}$, and $\mathbf{i}_{\boldsymbol{\phi}\psi}$ is the Fisher information vector with $(i,1)$th element $i_{\phi_i\psi}$ or $\mathbf{i}_{\boldsymbol{\phi}\psi}=E[-\partial^2 l(\psi,\boldsymbol{\phi})/\partial\boldsymbol{\phi}\partial\psi]$.

If ψ and $\boldsymbol{\phi}$ are *orthogonal* or $\mathbf{i}_{\boldsymbol{\phi}\psi}=\mathbf{0}$ then the coefficient c, and its MLE $\hat{c}$, collapse to zero and, helpfully, only the two first terms of the APL need to be calculated. This will turn out to be the case under the $AR^B(1)$ model. Oddly enough, the coefficient can equal zero even if the parameters are not orthogonal, as in the $AR^B(1)$ model with an unrestricted x_0 in Appendix C. In the $AR^{\mu}(1)$ model the parameters are nonorthogonal and the coefficient has to be calculated (Pere, 1997, 2000). Some intuition for the APL under orthogonality is given in the next section.

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

The value of ψ that maximizes (14) is the adjusted estimator (AE) of ψ or $\hat{\psi}_a$. An adjusted Wald statistic for the null hypothesis $\psi = 1$ is defined as

$$W_a \equiv (\hat{\psi}_a - 1)^2 \left[-\frac{\partial^2 l_a(\psi)}{\partial \psi^2} \bigg|_{\psi=\hat{\psi}_a} \right] \tag{16}$$

(Pere, 1997, 2000). An analogous adjusted LR statistic is

$$LR_a = \left\{ \frac{\exp[l_a(\psi)]|_{\psi=\hat{\psi}_a}}{\exp[l_a(\psi)]|_{\psi=1}} \right\}^{2/(T-3)}, \tag{17}$$

where the power is set for convenience only for the application to follow. The statistics and the corresponding asymptotic distributions, under the $\mathrm{AR}^B(1)$ model with a unit root, are calculated in the next section.

5. ADJUSTED PROFILE LIKELIHOOD ANALYSIS AND UNIT-ROOT ASYMPTOTICS FOR THE $\mathrm{AR}^B(1)$ MODEL

Under the $\mathrm{AR}^B(1)$ model ψ is advantageously orthogonal with respect to γ and σ^2 and the latter to each other:

$$\begin{bmatrix} i_{\psi\gamma} \\ i_{\psi\sigma^2} \\ i_{\gamma\sigma^2} \end{bmatrix} = E \begin{bmatrix} \sigma^{-2}\left[\sum_{t=2}^{T}(y_t-\gamma)+\sum_{t=2}^{T}(y_{t-1}-\gamma)-2\psi\sum_{t=2}^{T}(y_{t-1}-\gamma)\right] \\ \sigma^{-4}\sum_{t=2}^{T}[y_t-\gamma(1-\psi)-\psi y_{t-1}](y_{t-1}-\gamma) \\ \sigma^{-4}\left\{y_1-\gamma+(1-\psi)\sum_{t=2}^{T}[y_t-\gamma(1-\psi)-\psi y_{t-1}]\right\} \end{bmatrix} = \mathbf{0}. \tag{18}$$

The following equality was already touched on when (8) was contemplated:

$$i_{\gamma\gamma} = \sigma^{-2}[1 + (T-1)(1-\psi)^2]. \tag{19}$$

The number 1 in (19) arises from the first observation, whereas the term $(T-1)(1-\psi)^2$ ensues from the rest of the observations. Hence the first observation remains the most informative on γ as long as $\psi \neq 0$. If $\psi = 0$ then all observations are equally informative.

The final information measure to note is an observed measure:

$$j_{\sigma^2\sigma^2} = -T/2\sigma^4 + \sigma^{-6}\left\{(y_1-\gamma)^2 + \sum_{t=2}^{T}[y_t-\gamma(1-\psi)-\psi y_{t-1}]^2\right\}.$$

These calculations yield $-\frac{1}{2}\log|\mathbf{J}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\psi, \hat{\boldsymbol{\phi}}_\psi)| \propto \frac{3}{2}\log\hat{\sigma}_\psi^2 - \frac{1}{2}\log[1 + (T-1)\times(1-\psi)^2]$ and the APL:

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

$$l_a(\psi) \propto -\frac{(T-3)}{2}\log\hat{\sigma}_\psi^2 - \frac{1}{2}\log[1+(T-1)(1-\psi)^2]. \quad \textbf{(20)}$$

The term $-\frac{1}{2}\log[1+(T-1)(1-\psi)^2]$ pushes $\hat{\psi}_a^B$ to the right of $\hat{\psi}_{ml}^B$ on the real axis if $\hat{\psi}_{ml}^B < 1$, but it shrinks $\hat{\psi}_a^B$ downward if $\hat{\psi}_{ml}^B > 1$. The term is nonpositive and penalizes large values of $(1-\psi)^2$ or small values of ψ when the parameter space is restricted to $(-1,1]$. Absorbingly, the term diminishes as ψ approaches 1 and vanishes at 1. A divergent maximum of the APL can be excluded because there is a bounded solution for the minimum of $\hat{\sigma}_\psi^2$ by Section 3 and the potential divergence of the second term of (20) to minus infinity would be boosted if ψ diverged.

A way to interpret the Cox–Reid adjustment in general is that it penalizes values of ψ that provide relatively more information on the nuisance parameters. The phenomenon is quite explicit here: $i_{\gamma\gamma}$ is minimized at $\psi = 1$, and other more informative values are penalized as a simple function of $i_{\gamma\gamma}$.

A kind of degrees-of-freedom correction emerges to the multiplier of $\log\hat{\sigma}_\psi^2$. By numerical experiments, this adjustment may increase or decrease the estimate compared to the case where $T-3$ is replaced by the usual T (cf. (6)). An enlargement arises if $\hat{\psi}_{ml}^B < 1$, whereas a subtraction results if $\hat{\psi}_{ml}^B > 1$. Therefore the correction strengthens the adjustment due to the term $-\frac{1}{2}\log[1+(T-1)(1-\psi)^2]$.

It can be inferred from (20) that the MLE and AE share the same asymptotic distribution under a unit root: the terms $\hat{\sigma}_\psi^2$ and $[1+(T-1)(1-\psi)^2]$ are of the same magnitude for $\psi \in (-1,1]$ ($O_p(T)$ or $O_p(1)$ when evaluated at $\psi \in (-1,1)$ or $\psi = 1$, respectively). Logs are taken of both terms in (20), but the second term becomes negligible for large T because the first term is multiplied by $-(T-3)/2$. Hence the first term determines both the MLE and AE for large T.

The adjusted likelihood equation

$$\frac{\partial l_a(\psi)}{\partial\psi} = -\frac{(T-3)}{2\hat{\sigma}_\psi^2}\frac{\partial\hat{\sigma}_\psi^2}{\partial\psi} + \frac{(T-1)(1-\psi)}{1+(T-1)(1-\psi)^2} = 0 \quad \textbf{(21)}$$

can be shown to be equivalent to a fifth-order polynomial in ψ, so numerical methods are needed to find $\hat{\psi}_a^B$. The equation is guaranteed to possess a real root because the order of the polynomial is odd.

The second derivative of the APL is needed to construct the W_a^B statistic of (16):

$$\frac{\partial^2 l_a(\psi)}{\partial\psi^2} = -\frac{(T-3)}{2}\left[\frac{\partial^2\hat{\sigma}_\psi^2/\partial\psi^2}{\hat{\sigma}_\psi^2} - \frac{(\partial\hat{\sigma}_\psi^2/\partial\psi)^2}{\hat{\sigma}_\psi^4}\right] - \frac{(T-1)[1-(T-1)(1-\psi)^2]}{[1+(T-1)(1-\psi)^2]^2}. \quad \textbf{(22)}$$

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

The formula for $\partial^2\hat{\sigma}^2_\psi/\partial\psi^2$ is supplied in Appendix A. Despite absence of an analytical formula, the numerical value of W_a^B is easily found by substituting $\hat{\psi}_a^B$ in (16) and (22).

The adjustment does not alter the asymptotic unit-root distribution of the (signed square root of the) Wald statistic:

$$\sqrt{W_a^B} = T(\hat{\psi}_a^B - 1)$$

$$\cdot\sqrt{\left\{\frac{(T-3)}{2T}\left[\frac{T^{-1}\partial^2\hat{\sigma}^2_\psi/\partial\psi^2}{\hat{\sigma}^2_\psi} - \frac{(T^{-1/2}\partial\hat{\sigma}^2_\psi/\partial\psi)^2}{\hat{\sigma}^4_\psi}\right] + \frac{(T-1)[1-(T-1)(1-\psi)^2]}{T^2[1+(T-1)(1-\psi)^2]^2}\right\}\Bigg|_{\psi=\hat{\psi}_a^B}}$$

$$\Rightarrow \frac{\frac{1}{2}\{[B(1)]^2 - 1\}}{\int[B(r)]^2\,dr}\sqrt{\frac{1}{2}\left\{\frac{2\sigma^2\int[B(r)]^2\,dr}{\sigma^2} - \frac{0}{\sigma^4}\right\} + 0}$$

$$= \frac{\frac{1}{2}\{[B(1)]^2 - 1\}}{\sqrt{\int[B(r)]^2\,dr}}.$$

The convergence to zero of the additive adjustment $(T-1)[1-(T-1)\times(1-\psi)^2]/\{T^2[1+(T-1)(1-\psi)^2]^2\}$, evaluated at $\hat{\psi}_a^B$, is easily reasoned: $(T-1)(1-\hat{\psi}_a^B)^2$ is $O_p(T^{-1})$ under $\psi = 1$, so the effect is $O_p(T)$ over T^2. The other convergencies are explained in Appendix A.

The adjusted LR statistic is from (17)

$$LR_a^B = \frac{\hat{\sigma}^2_\psi|_{\psi=1}}{\{\hat{\sigma}^2_\psi[1+(T-1)(1-\psi)^2]^{1/(T-3)}\}|_{\psi=\hat{\psi}_a^B}}.$$

The term $\{[1+(T-1)(1-\psi)^2]^{1/(T-3)}\}|_{\psi=\hat{\psi}_a^B}$ does not play a part in the asymptotic distribution of $T(LR_a^B - 1)$ under a unit root: $(T-3)^{-1}\log\{1+(T-1)(1-\hat{\psi}_a^B)^2\}$ converges in probability to zero, so $[1+(T-1)\times(1-\hat{\psi}_a^B)^2]^{1/(T-3)}$ converges in probability to unity. It follows that $T(LR_a^B - 1)$ shares the same asymptotic unit-root distribution with $T(LR_{ml}^B - 1)$ or

$$T(LR_a^B - 1) \Rightarrow \frac{\frac{1}{4}\{[B(1)]^2 - 1\}^2}{\int[B(r)]^2\,dr}.$$

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

6. STATISTICS AND UNIT-ROOT ASYMPTOTICS FOR THE AR$^{\mu}$(1) MODEL

Under the AR$^{\mu}$(1) model the MLE of ψ and the Wald statistic for the unit-root null hypothesis are

$$\hat{\psi}_{ml}^{\mu} = \frac{\sum_{t=1}^{T} (y_t - \bar{y})(y_{t-1} - \bar{y}_{-1})}{\sum_{t=1}^{T} (y_{t-1} - \bar{y}_{-1})^2},$$

$$\sqrt{W_{ml}^{\mu}} = (\hat{\psi}_{ml}^{\mu} - 1)\sqrt{\frac{-T}{(\hat{\psi}_{ml}^{\mu})^2 - \hat{1}}},$$

where $\bar{y} = T^{-1}\sum_{t=1}^{T} y_t$, $\bar{y}_{-1} = T^{-1}\sum_{t=1}^{T} y_{t-1}$, $\hat{1} = \sum_{t=1}^{T}(y_t - \bar{y})^2 / \sum_{t=1}^{T}(y_{t-1} - \bar{y}_{-1})^2 \geq 0$, and the formulation of the Wald statistic is akin to (12) (Pere, 2000). The estimator $\hat{\psi}_{ml}^{\mu}$ is also the ordinary least-squares (OLS) estimator. The AE for ψ under the AR$^{\mu}$(1) model is essentially

$$\hat{\psi}_{a}^{\mu} = \hat{\psi}_{ml}^{\mu} + \frac{T-3}{2\hat{c}} - \sqrt{\frac{(T-3)^2}{4\hat{c}^2} + (\hat{\psi}_{ml}^{\mu})^2 - \hat{1}}, \tag{23}$$

where $\hat{c}$ is $(1-\psi)^{-1} - (1-\psi^T)/[T(1-\psi)^2]$ evaluated at $\hat{\psi}_{ml}^{\mu}$. The formula for $\hat{\psi}_{a}^{\mu}$ applies for a large T, but in the simulations that follow the exact definition given in Pere (2000) is employed. The adjusted Wald statistic of (16) is

$$\sqrt{W_{a}^{\mu}} = (\hat{\psi}_{a}^{\mu} - 1)\sqrt{-\frac{(T-3)[\psi^2 - \hat{1} - 2\psi\hat{\psi}_{ml}^{\mu} + 2(\hat{\psi}_{ml}^{\mu})^2]}{(\hat{1} - 2\psi\hat{\psi}_{ml}^{\mu} + \psi^2)^2}\Bigg|_{\psi=\hat{\psi}_{a}^{\mu}}}$$

(Pere, 2000).

The conventional statistics tend to the following well-known unit-root distributions as T increases:

$$T(\hat{\psi}_{ml}^{\mu} - 1) \Rightarrow \frac{\frac{1}{2}\{[B(1)]^2 - 1\} - B(1)\int B(r)\,dr}{\int [B^{\mu}(r)]^2\,dr} = DF^{\mu},$$

$$\sqrt{W_{ml}^{\mu}} \Rightarrow \frac{\frac{1}{2}\{[B(1)]^2 - 1\} - B(1)\int B(r)\,dr}{\left\{\int [B^{\mu}(r)]^2\,dr\right\}^{1/2}}, \tag{24}$$

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

where $B^{\mu}(r) = B(r) - \int_0^1 B(s)\,ds$ (e.g., Pere, 2000).[2] The corresponding adjusted statistics tend instead to

$$T(\hat{\psi}_a^{\mu} - 1) \Rightarrow DF^{\mu} + \frac{\exp(DF^{\mu}) - (1 + DF^{\mu})}{(DF^{\mu})^2} \left\{ \int [B^{\mu}(r)]^2\,dr \right\}^{-1},$$

$$\sqrt{W_a^{\mu}} \Rightarrow \left[DF^{\mu} + \frac{\exp(DF^{\mu}) - (1 + DF^{\mu})}{(DF^{\mu})^2 \int [B^{\mu}(r)]^2\,dr} \right] \left\{ \int [B^{\mu}(r)]^2\,dr \right\}^{1/2}$$

(Pere, 2000). The LR statistics are not inspected under the $AR^{\mu}(1)$ model.[3]

7. FRACTILES FOR THE UNIT-ROOT $AR^{B}(1)$ MODEL

The small-sample unit-root fractiles of the standardized coefficients (SCs), $T(\hat{\psi}_{ml}^{B} - 1)$ and $T(\hat{\psi}_a^{B} - 1)$, are reported in Tables 1 and 4 (based on simulations reported in Appendix D). These fractiles (and those in Tables 6, 7, 8, and 9 for the Wald and LR statistics) apply irrespective of the nuisance parameters γ and σ^2, or the initial value x_0, now relaxed to take any finite value, by the invariance and similarity results in Appendix C.

The median of the distribution of $T(\hat{\psi}_{ml}^{B} - 1)$ is negative and the left tail is the longer, which suggest a biased estimator for all sample sizes. For observations in excess of 50 the fractiles essentially monotonically tend toward zero, or the distribution becomes more compact.

The fractiles are compared with the (unreported) corresponding distribution under the $AR^{0}(1)$ model in Table 2. Two sample sizes are printed on most lines of the table (say, 24–25). The larger is the last observed time period and the

TABLE 1. Fractiles of $T(\hat{\psi}_{ml}^{B} - 1)$ for $\psi = 1$

T/Fr.	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
25	−17.2	−14.4	−12.1	−9.5	−1.05	1.08	1.50	1.89	2.42
50	−18.2	−14.7	−11.7	−8.4	−0.92	1.00	1.38	1.74	2.21
75	−18.2	−14.3	−11.0	−7.6	−0.90	0.97	1.34	1.69	2.13
100	−18.0	−13.8	−10.4	−7.1	−0.89	0.96	1.33	1.68	2.13
250	−16.0	−11.8	−9.0	−6.2	−0.86	0.94	1.31	1.64	2.07
500	−14.9	−11.2	−8.6	−6.0	−0.86	0.94	1.30	1.63	2.04
1,000	−14.3	−10.8	−8.2	−5.8	−0.85	0.94	1.29	1.62	2.04
∞	−13.6	−10.4	−8.0	−5.7	−0.85	0.93	1.28	1.61	2.03
ASE	0.18	0.14	0.10	0.07	0.02	0.01	0.01	0.02	0.03

Note: ASE is the average of the estimated standard errors of the fractile across the finite-sample sizes.

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

TABLE 2. Differences between the fractiles of $T(\hat{\psi}^B_{ml} - 1)$ and $T(\hat{\psi}^0_{ml} - 1)$ for $\psi = 1$

T/Fr.	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
24–25	−5.3	−5.0	−4.7	−4.1	−0.2	0.1	0.1	0.1	0.1
49–50	−5.4	−4.8	−4.0	−2.9	−0.1				0.1
74–75	−5.1	−4.2	−3.2	−2.0	−0.1				
99–100	−4.7	−3.6	−2.5	−1.4					
249–250	−2.5	−1.5	−1.0	−0.5					
499–500	−1.3	−0.8	−0.5	−0.3					
999–1,000	−0.7	−0.4	−0.2	−0.1					
∞									

Note: The larger sample sizes apply to the $AR^B(1)$ model and the smaller to the $AR^0(1)$ model. An empty cell marks no difference to one decimal place.

smaller is the number of observations available for estimation of the $AR^0(1)$ model (Section 2). (Parallel notation applies to Tables 3, 12, and 13.)

Engrossingly, when the sample size is small the left-tail fractiles under the $AR^B(1)$ model lie far to the left of the corresponding left-tail fractiles under the $AR^0(1)$ model. It increases the bias and spread of the MLE in comparison to the latter model. Moreover, the distribution under the former converges to the asymptotic fairly slowly. The $AR^B(1)$ model establishes invariance with respect to the initial value at the cost of inaccuracy in estimation in comparison with the $AR^0(1)$ model and does not yield small-sample distributions that are as dense as the asymptotic distribution suggests. The small-sample left-tail fractiles under the $AR^B(1)$ model fit better with (and lie in general to the right of) those for the $AR^\mu(1)$ model (Table 3). In large samples these fractiles naturally differ.

TABLE 3. Differences between the fractiles of $T(\hat{\psi}^B_{ml} - 1)$ and $T(\hat{\psi}^\mu_{ml} - 1)$ for $\psi = 1$

T/Fr.	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
24–25	−0.6	−0.4	−0.1	0.4	3.0	1.8	1.5	1.3	1.1
49–50	0.3	0.7	1.3	2.1	3.3	1.8	1.5	1.2	1.0
74–75	1.0	1.6	2.3	3.2	3.3	1.8	1.4	1.2	1.0
99–100	1.6	2.4	3.1	3.8	3.4	1.8	1.5	1.2	1.0
249–250	4.1	4.7	4.9	4.9	3.5	1.8	1.4	1.2	1.0
499–500	5.5	5.5	5.4	5.2	3.5	1.8	1.4	1.2	1.0
999–1,000	6.1	6.0	5.8	5.4	3.5	1.8	1.4	1.2	1.0
∞	7.0	6.5	6.1	5.5	3.5	1.8	1.4	1.2	1.0

Note: The larger sample sizes apply to the $AR^B(1)$ model and the smaller to the $AR^\mu(1)$ model. An empty cell marks no difference to one decimal place.

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

TABLE 4. Fractiles of $T(\hat{\psi}_a^B - 1)$ for $\psi = 1$

T/Fr.	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
25	−15.5	−12.5	−10.0	−7.1	−0.87	1.02	1.40	1.75	2.21
50	−16.2	−12.6	−9.7	−6.5	−0.85	0.97	1.34	1.68	2.12
75	−16.2	−12.3	−9.2	−6.3	−0.86	0.95	1.32	1.65	2.08
100	−16.0	−12.0	−9.0	−6.1	−0.86	0.95	1.31	1.65	2.08
250	−14.7	−10.9	−8.4	−5.9	−0.85	0.94	1.30	1.63	2.05
500	−14.2	−10.7	−8.3	−5.8	−0.85	0.93	1.29	1.62	2.03
1,000	−14.0	−10.6	−8.1	−5.7	−0.85	0.94	1.29	1.62	2.03
∞	−13.6	−10.4	−8.0	−5.7	−0.85	0.93	1.28	1.61	2.03
ASE	0.18	0.13	0.11	0.06	0.01	0.01	0.01	0.02	0.02

Note: ASE is the average of the estimated standard errors of the fractile across the finite-sample sizes.

The 0.05th fractiles of $T(\hat{\psi}_{ml}^B - 1)$ are reported in Pantula et al. (1994) also. These fractiles are smaller than those in Table 1 for samples of 25 and 50 apparently because, by unreported simulations with the OPTMUM module of GAUSS 3.2.17 (1992), the simple algorithm employed in Pantula et al. (1994) yields left-tail fractiles that are too small.

Zivot (1994, p. 564), focusing on the unit-root case, reasons that $\hat{\psi}_{ml}^B$ takes similar values to $\hat{\psi}_{ml}^0 = \sum_{t=2}^T y_t y_{t-1} / \sum_{t=2}^T y_{t-1}^2$ or the MLE of ψ calculated from the AR^0(1) model treating y_1 as the initial value.[4] Indeed, it is shown in (A.5) (evaluated at $\delta = 0$) in Appendix A that $\hat{\psi}_{ml}^B = \hat{\psi}_{ml}^0 + O_p(T^{-3/2})$ for $\psi = 1$. However, unreported simulation experiments reveal that the similarity holds only under a (near) unit root and a fairly large sample size (cf. Table 2). Otherwise the estimators can take very different values.

The adjustment cuts both tails of the distribution of the MLE but especially the left tail (Tables 4 and 5). Moreover, in essence all the fractiles of the AE—

TABLE 5. Differences between the fractiles of $T(\hat{\psi}_a^B - 1)$ and $T(\hat{\psi}_{ml}^B - 1)$ for $\psi = 1$

T/Fr.	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
25	1.7	1.9	2.1	2.3	0.2	−0.1	−0.1	−0.1	−0.2
50	2.0	2.0	2.1	1.8	0.1			−0.1	−0.1
75	2.0	2.0	1.8	1.3					−0.1
100	2.0	1.8	1.5	0.9					
250	1.3	0.9	0.6	0.3					
500	0.7	0.5	0.3	0.2					
1,000	0.4	0.2	0.1	0.1					
∞									

Note: An empty cell marks no difference to one decimal place.

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

TABLE 6. Fractiles of $\sqrt{W_{ml}^B}$ for $\psi = 1$

T/Fr.	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
25	−3.47	−3.03	−2.65	−2.20	−0.52	1.01	1.44	1.81	2.25
50	−3.17	−2.74	−2.35	−1.88	−0.50	0.95	1.36	1.72	2.14
75	−3.03	−2.59	−2.19	−1.73	−0.50	0.92	1.33	1.68	2.08
100	−2.94	−2.49	−2.10	−1.66	−0.50	0.91	1.32	1.67	2.07
250	−2.66	−2.25	−1.92	−1.58	−0.50	0.90	1.30	1.64	2.04
500	−2.57	−2.20	−1.91	−1.60	−0.50	0.89	1.29	1.63	2.04
1,000	−2.54	−2.20	−1.92	−1.60	−0.50	0.90	1.29	1.63	2.02
∞	−2.56	−2.23	−1.94	−1.62	−0.50	0.89	1.28	1.62	2.00
ASE	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.02

Note: ASE is the average of the estimated standard errors of the fractile across the finite-sample sizes.

except the asymptotic fractiles copied from Table 1—lie closer both to zero and to the asymptotic fractiles than those of the MLE. The AE seems less biased and variable, or more accurate, and follows the common asymptotic distribution better than the MLE.

The fractiles of the Wald statistics $\sqrt{W_{ml}^B}$ and $\sqrt{W_a^B}$ are in Tables 6 and 7. The location and shape of the distributions vary less with the sample size than those of the SCs. Yet the distribution of $\sqrt{W_a^B}$ is more concentrated in general and matches much better with the common asymptotic distribution in small samples. Both statistics take a real value for all draws. However, unreported simulation experiments uncover that $\sqrt{W_a^B}$ can take a complex value if T is extremely small, such as 5.

Tables 8 and 9 contain the fractiles for the LR statistics. The adjustment cuts the tails of the small-sample distributions so that LR_a^B follows the common asymptotic distribution far better than LR_{ml}^B.

TABLE 7. Fractiles of $\sqrt{W_a^B}$ for $\psi = 1$

T/Fr.	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
25	−2.79	−2.31	−1.91	−1.48	−0.43	0.92	1.31	1.66	2.08
50	−2.63	−2.20	−1.85	−1.48	−0.46	0.90	1.30	1.65	2.05
75	−2.57	−2.16	−1.83	−1.48	−0.48	0.90	1.29	1.63	2.03
100	−2.54	−2.14	−1.82	−1.49	−0.49	0.89	1.29	1.64	2.03
250	−2.46	−2.11	−1.83	−1.53	−0.49	0.89	1.29	1.63	2.02
500	−2.48	−2.14	−1.87	−1.57	−0.50	0.88	1.28	1.63	2.03
1,000	−2.50	−2.17	−1.90	−1.59	−0.50	0.89	1.28	1.63	2.02
∞	−2.56	−2.23	−1.94	−1.62	−0.50	0.89	1.28	1.62	2.00
ASE	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02

Note: ASE is the average of the estimated standard errors of the fractile across the finite-sample sizes.

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

TABLE 8. Fractiles of $T(LR_{ml}^{B} - 1)$ for $\psi = 1$

T/Fr.	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
25	0.00029	0.0018	0.0071	0.0286	0.795	4.60	6.48	8.44	11.19
50	0.00025	0.0016	0.0063	0.0256	0.665	3.86	5.49	7.16	9.40
75	0.00024	0.0015	0.0062	0.0253	0.644	3.57	5.10	6.68	8.88
100	0.00024	0.0016	0.0062	0.0246	0.628	3.43	4.88	6.42	8.53
250	0.00025	0.0015	0.0061	0.0241	0.614	3.12	4.39	5.74	7.66
500	0.00023	0.0014	0.0058	0.0235	0.607	3.06	4.29	5.55	7.32
1,000	0.00023	0.0014	0.0058	0.0234	0.604	3.01	4.19	5.44	7.12
∞	0.00024	0.0015	0.0059	0.0236	0.602	2.97	4.12	5.32	6.92
ASE	0.00003	0.0001	0.0003	0.0010	0.008	0.03	0.05	0.07	0.11

Note: ASE is the average of the estimated standard errors of the fractile across the finite-sample sizes.

8. FRACTILES FOR THE UNIT-ROOT AR$^{\mu}$(1) MODEL

Fractiles of the finite-sample and the asymptotic unit-root distributions of $T(\hat{\psi}_a^{\mu} - 1)$ and $\sqrt{W_a^{\mu}}$ are reported in Tables 10 and 11. We found that $\sqrt{W_a^{\mu}}$ was real in all of the simulations. The fractiles are invariant with respect to the initial value (Pere, 2000). The constant α is fixed to zero, but the magnitude of σ^2 is irrelevant for the fractiles (proof omitted). The distributions are discussed and the details of the simulations are in Pere (1997, 2000).

9. COMPARING ESTIMATION ACCURACY

The bias, standard deviation (SD), and mean-square error (MSE) of the MLE and AE for a random walk under the different AR(1) models are shown in

TABLE 9. Fractiles of $T(LR_{a}^{B} - 1)$ for $\psi = 1$

T/Fr.	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
25	0.00025	0.0016	0.0062	0.0249	0.630	3.16	4.46	5.89	7.95
50	0.00023	0.0015	0.0060	0.0241	0.609	3.09	4.33	5.61	7.41
75	0.00024	0.0015	0.0059	0.0244	0.610	3.06	4.27	5.54	7.31
100	0.00023	0.0015	0.0060	0.0240	0.604	3.04	4.24	5.51	7.26
250	0.00024	0.0015	0.0060	0.0239	0.605	2.99	4.15	5.35	7.06
500	0.00023	0.0014	0.0058	0.0234	0.603	2.99	4.16	5.35	7.00
1,000	0.00023	0.0014	0.0058	0.0234	0.602	2.98	4.13	5.34	6.95
∞	0.00024	0.0015	0.0059	0.0236	0.602	2.97	4.12	5.32	6.92
ASE	0.00002	0.0001	0.0003	0.0009	0.008	0.03	0.04	0.06	0.09

Note: ASE is the average of the estimated standard errors of the fractile across the finite-sample sizes.

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

TABLE 10. Fractiles of $T(\hat{\psi}_a^{\mu} - 1)$ for $\psi = 1$

T/Fr.	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
24	−15.2	−12.6	−10.5	−8.3	−2.62	1.00	2.20	3.45	5.32
49	−17.0	−13.8	−11.4	−8.9	−2.78	0.83	1.94	3.10	4.81
74	−17.5	−14.2	−11.7	−9.1	−2.82	0.78	1.86	2.96	4.61
99	−17.9	−14.5	−11.9	−9.2	−2.88	0.74	1.79	2.87	4.50
249	−18.4	−14.8	−12.1	−9.4	−2.90	0.70	1.74	2.81	4.35
499	−18.6	−15.0	−12.3	−9.5	−2.91	0.69	1.68	2.72	4.24
999	−18.7	−15.1	−12.3	−9.5	−2.92	0.67	1.71	2.76	4.28
∞	−18.8	−15.1	−12.4	−9.6	−2.95	0.65	1.68	2.67	4.21
ASE	0.19	0.13	0.09	0.05	0.02	0.03	0.04	0.05	0.11

Note: ASE is the average of the estimated standard errors of the fractile across the sample sizes.

Table 12. The number of observations is 100 for estimation of the $AR^B(1)$ model and one less for the other models (Section 2). The AE and MLE are the same under the $AR^0(1)$ model (Pere, 1997).

The bias, SD, and MSE of the estimators decrease uniformly in absolute value from the top to the bottom row. The $AR^{\mu}(1)$ model is by far the worst with respect to each of the quantities. The statistics for the $AR^B(1)$ model are clearly smaller yet are inferior to the statistics for the $AR^0(1)$ model. The adjustment improves estimation accuracy in terms of bias and MSE under the $AR^{\mu}(1)$ and $AR^B(1)$ models and also in terms of SD under the latter. Under a unit root this holds even asymptotically under the former model (Pere, 2000). This is possible because the standard regularity conditions behind the usual optimality results for the MLE are not valid under nonstationarity.

TABLE 11. Fractiles of $\sqrt{W_a^{\mu}}$ for $\psi = 1$

T/Fr.	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
24	−3.40	−2.96	−2.60	−2.22	−0.97	0.44	0.94	1.35	1.76
49	−3.27	−2.89	−2.57	−2.22	−1.03	0.39	0.91	1.39	1.99
74	−3.23	−2.87	−2.56	−2.22	−1.04	0.38	0.89	1.37	2.01
99	−3.23	−2.86	−2.56	−2.22	−1.05	0.36	0.88	1.36	1.99
249	−3.17	−2.83	−2.54	−2.22	−1.07	0.35	0.86	1.36	2.03
499	−3.17	−2.83	−2.55	−2.22	−1.07	0.34	0.85	1.33	2.00
999	−3.16	−2.83	−2.55	−2.22	−1.07	0.34	0.86	1.36	2.02
∞	−3.15	−2.82	−2.54	−2.22	−1.08	0.33	0.84	1.33	2.00
ASE	0.02	0.02	0.01	0.01	0.01	0.01	0.02	0.02	0.04

Note: ASE is the average of the estimated standard errors of the fractile across the sample sizes.

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

TABLE 12. Bias (×100), SD (×100), and MSE (×1,000) of the MLE and AE for a random walk and $T = 99$–100

	MLE			AE		
Model	Bias	SD	MSE	Bias	SD	MSE
$AR^{\mu}(1)$	−5.29	4.35	4.7	−3.70	4.43	3.3
$AR^{B}(1)$	−2.18	4.00	2.1	−1.93	3.56	1.6
$AR^{0}(1)$	−1.74	3.11	1.3	−1.74	3.11	1.3

Note: The sample size is 100 for the $AR^{B}(1)$ model and 99 for the other models.

Unreported simulation experiments indicate that if the sample size is exceptionally small, say, 10 or less, then the AE is an overwhelmingly better estimator than the MLE under the unit-root $AR^{B}(1)$ model. This is due to the dominance of the unit-root-favoring term of the APL, $-\frac{1}{2}\log[1 + (T-1)(1-\psi)^2]$, when the sample is uninformative.

10. NEAR UNIT-ROOT ASYMPTOTICS UNDER THE $AR^{B}(1)$ MODEL

In the case of a near unit-root or $\psi = \exp(\delta/T)$, $\delta < 0$ (Phillips, 1987) corresponding asymptotic distributions apply:

$$T(\hat{\psi}_i^B - 1) \Rightarrow \frac{\frac{1}{2}\{[J_\delta(1)]^2 - 1\}}{\int [J_\delta(r)]^2\, dr},$$

$$\sqrt{W_i^B} \Rightarrow \frac{\frac{1}{2}\{[J_\delta(1)]^2 - 1\}}{\left\{\int [J_\delta(r)]^2\, dr\right\}^{1/2}},$$

$$T(LR_i^B - 1) \Rightarrow \frac{\frac{1}{4}\{[J_\delta(1)]^2 - 1\}^2}{\int [J_\delta(r)]^2\, dr}. \tag{25}$$

In the preceding expressions i is an index for ml or a, and $J_\delta(r) = B(r) + \delta\int_0^r \exp[(r-s)\delta]B(s)\, ds$ is an Ornstein–Uhlenbeck process. The proofs for $i = ml$ are in Appendix A. The adjusted statistics follow the same distributions as the unadjusted by similar logic as in the unit-root case (Section 5 and Appendix A). Phillips (1987) has discovered that the first two asymptotic distri-

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

butions apply (and it can be shown that the third distribution also applies) for the corresponding statistics under the $AR^0(1)$ model with a near unit root. Thus the unit-root and near unit-root distributions of the respective statistics are identical under the $AR^B(1)$ and $AR^0(1)$ models. It follows that the respective unit-root tests are asymptotically equally powerful against near unit-root alternatives whether or not the adjustment is employed and regardless of which of the two models is used.

11. NEAR UNIT-ROOT ASYMPTOTICS UNDER THE $AR^{\mu}(1)$ MODEL

The asymptotic distributions of $T(\hat{\psi}_{ml}^{\mu} - 1)$ and $\sqrt{W_{ml}^{\mu}}$, under $\psi = \exp(\delta/T)$ and $\alpha = 0$, are known to be

$$T(\hat{\psi}_{ml}^{\mu} - 1) \Rightarrow \frac{\frac{1}{2}\{[J_{\delta}(1)]^2 - 1\} - J_{\delta}(1)\int J_{\delta}(r)\,dr}{\int [J_{\delta}^{\mu}(r)]^2\,dr} = DF_{\delta}^{\mu},$$

$$\sqrt{W_{ml}^{\mu}} \Rightarrow \frac{\frac{1}{2}\{[J_{\delta}(1)]^2 - 1\} - J_{\delta}(1)\int J_{\delta}(r)\,dr}{\left\{\int [J_{\delta}^{\mu}(r)]^2\,dr\right\}^{1/2}},$$

where $J_{\delta}^{\mu}(r) = J_{\delta}(r) - \int_0^1 J_{\delta}(s)\,ds$ (e.g., Stock, 1994, p. 2772). It is proved in Appendix B that the adjusted statistics follow different distributions under the same assumptions:

$$T(\hat{\psi}_{a}^{\mu} - 1) \Rightarrow DF_{\delta}^{\mu} + \frac{\exp(DF_{\delta}^{\mu}) - (1 + DF_{\delta}^{\mu})}{(DF_{\delta}^{\mu})^2}\left\{\int [J_{\delta}^{\mu}(r)]^2\,dr\right\}^{-1},$$

$$\sqrt{W_{a}^{\mu}} \Rightarrow \left[DF_{\delta}^{\mu} + \frac{\exp(DF_{\delta}^{\mu}) - (1 + DF_{\delta}^{\mu})}{(DF_{\delta}^{\mu})^2\int [J_{\delta}^{\mu}(r)]^2\,dr}\right]\left\{\int [J_{\delta}^{\mu}(r)]^2\,dr\right\}^{1/2}.$$

The power of the respective unit-root tests against near unit-root alternatives is likely to differ, even asymptotically, under the $AR^{\mu}(1)$ model.

12. UNIT-ROOT TESTS AND POWER COMPARISONS FOR THE $AR^B(1)$ MODEL

The unit-root fractiles of $T(\hat{\psi}_{ml}^{B} - 1)$, $T(\hat{\psi}_{a}^{B} - 1)$, $\sqrt{W_{ml}^{B}}$, $\sqrt{W_{a}^{B}}$, LR_{ml}^{B}, and LR_{a}^{B} permit testing of $\psi = 1$ under the $AR^B(1)$ model for various sample sizes. Large

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

values of the LR statistics and small values of the other statistics suggest rejection of the null.

The tests based on the SCs will be denoted by ψ_{ml}^B and ψ_a^B with obvious notation. The corresponding Wald and LR tests will carry the labels τ_{ml}^B, τ_a^B, Φ_{ml}^B, and Φ_a^B. The simulated powers of the tests are in Table 13. The test size is 5%, and the ranges of ψ and T are 0.4–0.99 and 24–25 to 249–250. The larger sample sizes apply in this section.

The ψ_a^B test is as powerful as or slightly more powerful than the ψ_{ml}^B test.[5] The tests are about equally powerful at 250 observations, reflecting the asymptotic convergence of the estimates. There is little difference in power between the τ_{ml}^B and τ_a^B tests, except that τ_a^B is more powerful when the sample size is 25.

The Wald tests are as powerful as or more powerful than the corresponding SC tests when there are 25 observations, contrary to the well-known ranking under the AR$^\mu$(1) model (next section).[6] The τ_a^B test is, then, among the in-

TABLE 13. Power (×100) of the unit root tests when the size is 5%

T	ψ	ψ_{ml}^B	ψ_a^B	τ_{ml}^B	τ_a^B	Φ_{ml}^B	Φ_a^B	ψ_{ml}^μ	ψ_a^μ	τ_{ml}^μ	τ_a^μ
24–25	0.99	5.6	5.6	5.5	5.6	5.2	4.8	5.5	5.5	5.1	5.1
	0.95	7.9	8.0	7.9	8.0	7.3	6.3	7.7	7.6	5.7	5.6
	0.90	11.2	11.5	11.5	11.9	11.0	9.5	11.0	10.7	6.8	6.7
	0.85	14.9	15.7	15.8	16.4	15.5	13.4	14.9	14.3	8.4	8.2
	0.80	20.1	21.1	21.3	22.2	21.2	18.7	19.7	18.9	10.8	10.6
	0.70	34.2	35.9	36.3	37.8	36.4	32.7	33.2	31.8	19.0	18.7
	0.60	52.4	54.3	54.7	56.5	54.7	50.1	50.8	49.1	31.7	31.2
	0.50	71.1	72.7	72.9	74.1	72.2	67.8	68.9	67.3	48.3	47.7
	0.40	85.1	86.2	86.2	86.8	85.1	81.6	83.4	82.2	65.8	65.2
99–100	0.99	7.6	7.6	7.7	7.8	6.7	6.2	7.2	7.1	5.7	5.7
	0.95	27.1	27.8	28.3	28.9	25.3	24.1	19.3	18.7	11.6	11.4
	0.90	64.1	65.3	66.0	66.1	61.8	59.9	47.1	45.8	30.8	30.3
	0.85	90.5	90.9	90.2	89.1	88.2	86.5	78.5	77.3	60.9	60.4
	0.80	98.7	98.6	98.0	97.2	97.4	96.5	95.3	94.8	86.0	85.6
	0.70	100.0	100.0	99.9	99.9	99.9	99.8	100.0	100.0	99.6	99.6
	0.60			100.0	100.0	100.0	100.0			100.0	100.0
249–250	0.99	14.3	14.4	14.5	14.5	12.3	11.8	10.8	10.6	7.1	7.0
	0.95	85.7	85.8	84.3	83.8	81.6	80.6	62.7	61.4	44.6	44.1
	0.90	99.9	99.9	99.5	99.2	99.5	99.2	99.4	99.3	96.7	96.6
	0.85	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: The larger sample sizes apply to the AR^B(1) model and the smaller to the AR$^\mu$(1) model. Blank entries indicate that power equal to one to three decimal places was achieved already under a larger ψ.

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

spected tests and values of ψ, the most powerful. For larger sample sizes the most powerful test can be any of the SC or Wald tests.

The LR tests are never the most powerful and compare unfavorably to the rest of the tests. The Φ_{ml}^{B} test can possess more power than a SC or Wald test under $T = 25$ only, and the weakest test is Φ_{a}^{B}. Some intuition is that the LR tests are two-sided, whereas the other tests are one-sided.

13. UNIT-ROOT TESTS AND POWER COMPARISONS FOR THE AR$^{\mu}$(1) MODEL

The tests associated with the $\hat{\psi}_{ml}^{\mu}$, $\hat{\psi}_{al}^{\mu}$, $\sqrt{W_{ml}^{B}}$, and W_{a}^{B} statistics will be referred to as ψ_{ml}^{μ}, ψ_{a}^{μ}, τ_{ml}^{μ}, and τ_{a}^{μ}. Their powers are in Table 13 also.[7] The sample sizes correspond to those in the previous section by being smaller by one. The difference in sample sizes arises from generating the data from the AR^{B}(1) model but fitting the AR$^{\mu}$(1) model, as explained in Section 2.

The adjusted versions of the tests tend to be a bit weaker in contrast to the AR^{B}(1) case. The differences in power are in general larger between the SC than between the Wald tests. As a by-product, Table 13 affirms the Monte Carlo results of Dickey, Bell, and Miller (1986, Table 2), Dickey and Fuller (1979, p. 430), and Schmidt and Phillips (1992): the ψ_{ml}^{μ} test is often substantially more powerful than the τ_{ml}^{μ} test.

14. POWER COMPARISON BETWEEN THE TWO AR(1) MODELS

Even though the AR^{B}(1) and AR$^{\mu}$(1) models agree under the null, the time series, or the underlying processes against which the powers are evaluated, can differ despite a common AR coefficient. The time series match, up to the long-run mean, when $x_0 = y_0 - \alpha/(1 - \psi) = 0$ (Section 2) or when both time series have started from the respective long-run mean. The power figures for the models can be compared under this assumption meaningfully.

Underlining indicates the most powerful test in Table 13. In case of a tie a number of tests are indicated except on the lines where all tests achieve power of unity.

The ψ_{ml}^{B} test is in general more powerful than the ψ_{ml}^{μ} test. The difference in power can be very large but is small for 24–25 observations. The latter outcome is intuitively due to the similarity of the small-sample distributions of the associated MLEs (Table 3). The ψ_{a}^{B} test is the most powerful of the SC tests for all sample sizes (but see note 5). The τ_{ml}^{B} and τ_{a}^{B} tests tend to be far more powerful than the corresponding τ_{ml}^{μ} and τ_{a}^{μ} tests, so that at its best the τ_{a}^{B} test can possess more than twice the power of the τ_{ml}^{μ} or τ_{a}^{μ} test. This is largely because the powers of the Wald tests do not fall short of the powers of the SC tests under the AR^{B}(1) model as happens under the AR$^{\mu}$(1) model.

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

Tests gain power much faster under the $AR^B(1)$ than the $AR^\mu(1)$ model as the number of observations increases. For example, the powers ($\times 100$) for the ψ_{ml}^μ test under $\psi = 0.95$ at sample sizes of 24 and 249 are 7.7 and 62.7. The corresponding figures for the ψ_{ml}^B test are 7.9 and 85.7. The relative gain is even larger for the Wald tests: the analogous figures for them are 5.7 and 44.6 vs. 7.9 and 84.3.

The $AR^B(1)$ model always yields more powerful tests. The benefit in power ($\times 100$) from using the $AR^B(1)$ model is 5.7% units at its best for 24–25 observations (τ_a^B vs. ψ_{ml}^μ at $\psi = 0.6$) and swells to 19% units for 99–100 observations (τ_a^B vs. ψ_{ml}^μ at $\psi = 0.90$) and to 23.1% units for 249–250 observations (ψ_a^B vs. ψ_{ml}^μ at $\psi = 0.95$).

The adjustment does not necessarily increase power. The adjustment decreases power under the $AR^\mu(1)$ model and power of LR tests under the $AR^B(1)$ model. Nevertheless, in general the most powerful test employs the adjustment (but see note 5).

The τ_a^B test is, among the inspected tests and values of ψ, the most powerful at sample size of 25 and the most powerful for ψ close to unity at sample size of 99–100. Among the same sets, no test is always the most powerful for sample sizes of 99–100 and 249–250.

15. DISCUSSION

The $AR^B(1)$ model is an innovative way to present nuisance parameters orthogonally to the AR coefficient. Estimation of the constant does not affect the asymptotic distribution of the standardized MLE even in the unit-root case. The model can yield more accurate estimates and powerful tests of a unit root than the conventional model. A disadvantage is that the time series is restrictively assumed to initialize from the long-run mean.

The Cox–Reid adjustment is lucid and intuitive. Under the unit-root $AR^B(1)$ model a simultaneous cutback in bias, variance, and MSE over the MLE can be achieved with the adjustment in small samples.

The adjusted statistics follow the asymptotic unit-root distributions better than their unadjusted counterparts when the model is $AR^B(1)$. The same outcome has been found previously when the model is $AR^\mu(1)$ and the process is a random walk with drift or stable and $0 \leq \psi < 1$ (Pere, 2000) or when the process is a stationary AR(1) with the initial value obeying the stationary distribution (Cruddas, Reid, and Cox, 1989). Thus the Cox–Reid adjustment is able to bring the small-sample distribution closer to the asymptotic distribution whether the process is stable or nonstationary or whether the asymptotic distribution is standard (such as normal) or a unit-root distribution.

A point of statistical interest is that if the initial value of the unobserved AR(1) underlying the $AR^B(1)$ process is allowed to differ from zero then the parameters become nonorthogonal whereas the adjustment remains unmodified.

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

The derived tests are invariant with respect to the initial value and similar. The τ_a^B test is the most powerful for the smallest inspected sample size, but for larger samples no single test dominates when the model is $AR^B(1)$. Under the same model, an LR test is never the most powerful, and the performance of the LR tests is deficient. For the $AR^\mu(1)$ model the ψ_{ml}^μ test is the most powerful.

Powers under the two models can be meaningfully compared under the assumptions that the time series follow an $AR^B(1)$ process, that the $AR^\mu(1)$ model is estimated conditional on y_1, and that the time series have started from their respective long-run means. The $AR^B(1)$ model yields more powerful tests so that the difference in power increases with the sample size (within the inspected range) and can become very large.

The adjustment does not inevitably increase power, as exemplified by the LR tests for the $AR^B(1)$ model and the tests for the $AR^\mu(1)$ model. Nevertheless, the τ_a^B test appears the most powerful for very small sample sizes, and in general the most powerful tests utilize the adjustment. The benefit from the adjustment is clearer for small sample sizes, because the adjustment fades asymptotically.

Many empiricists use the τ_{ml}^μ test even though the ψ_{ml}^μ test is more powerful. Now with very small samples, at least, there is reason to employ a Wald test but one derived from the $AR^B(1)$ model. The finding matches with the result of Pantula et al. (1994, p. 452) that in general, apart from the ψ_{ml}^μ and τ_{ml}^μ tests, Wald-type tests tend to be more powerful if the time series starts from the long-run mean. No single test possesses the highest power for all sample size and parameter combinations.

The results on power would not change essentially (by unreported simulations) if the power comparisons were based on an equal number of observations or had the $AR^0(1)$ or $AR^\mu(1)$ model generated the data to which the $AR^B(1)$ model would have been fitted. The $AR^B(1)$ model would still furnish the more powerful tests, and the most powerful tests would still tend to employ the adjustment.

Robinson (1994) and Saikkonen and Luukkonen (1997) also utilize a model in which the interpretation of deterministic terms is not dependent on the stationarity status of the underlying time series. Robinson's unit-root tests enjoy asymptotic optimality properties against fractionally integrated alternatives, whereas the cointegration tests of the latter authors possess superior power performance over conventional tests. Because the asymptotic distribution theory in these papers is not affected by the deterministic terms, similarly to this paper, this contemporary way of introducing deterministicity seems beneficial in more general settings, too.

NOTES

1. Recently Hansen and Rahbek (2002) have dealt with multiple parameters of interest in a related fashion.

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

2. There is a misprint in the $B_*(1)$ formula in Pere (2000, p. 339). The term $B_*(1)$ should be denoted by $B_*(r)$ and be as $B^{\mu}(r)$ in the current notation.

3. Under the AR$^{\mu}(1)$ model with $(\psi, \alpha) = (1,0)$, one might presume that an LR unit-root test for the joint hypothesis $(\psi, \alpha) = (1,0)$ renders a more powerful test than one focusing on ψ only. However, in the simulations of Dickey and Fuller (1981) the test did not yield a gain in power. An explanation is that $\hat{\alpha}_{ml}$, with obvious notation, absorbs changes in the level of the time series due to an altered initial value, say, and is comparable to the in general inessential constant of an ordinary linear regression.

4. Zivot's formula (37) for $S^2(\rho, x_0)$ agrees with (7) here for $\hat{\sigma}^2_{\psi}$ after substituting ψ for ρ. Zivot wrote that "$S^2(\mu, x_0)$ . . . is minimized near $\tilde{\rho}$, the OLS estimate of ρ in the simple AR(1) model without an intercept." (Here $S^2(\mu, x_0)$ looks like a misprint for $S^2(\rho, x_0)$ because the former was not defined in the paper nor does it have ρ as an argument.)

5. The combination $\psi = 0.8$ and $T = 99$–100 is an exception.

6. The ψ^B_{ml} and τ^B_{ml} tests are an exception at $\psi = 0.99$.

7. The statistics $\sqrt{W^{\mu}_{ml}}$ and $\sqrt{W^{\mu}_{a}}$ employ the MLE of σ instead of the degrees-of-freedom corrected estimator as in Dickey and Fuller (1979). The estimator of σ affects the scale of the null distributions, but power is unaffected. The simulation algorithm accepts the null of a unit root if it comes across a noncalculable $\sqrt{W^{\mu}_{a}}$ (for details, see Pere, 1997, 2000). Another alternative would be to exclude such instances, which would, if anything, increase the power estimates for the τ^{μ}_{a} test.

REFERENCES

Abadir, K.M. (1995) A new test for nonstationarity against the stable alternative. *Econometric Theory* 11, 81–104.

Ahn, S.K. (1993) Some tests for unit roots in autoregressive-integrated-moving average models with deterministic trends. *Biometrika* 80, 855–868.

Andrews, D.W.K. (1993) Exactly median-unbiased estimation of first order autoregressive/unit root models. *Econometrica* 61, 139–165.

Bhargava, A. (1986) On the theory of testing for unit roots in observed time series. *Review of Economic Studies* 52, 369–384.

Cox, D.R. & N. Reid (1987) Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society, Series B* 49, 1–39.

Cox, D.R. & N. Reid (1993) A note on the calculation of adjusted profile likelihood. *Journal of the Royal Statistical Society, Series B* 55, 467–471.

Cruddas, A.M., N. Reid, & D.R. Cox (1989) A time series illustration of approximate conditional likelihood. *Biometrika* 76, 231–237.

Dickey, D.A. (1976) *Estimation and Hypothesis Testing in Nonstationary Time Series*. Ph.D. thesis. Iowa State University.

Dickey, D.A., W.R. Bell, & R.B. Miller (1986) Unit roots in time series models: Tests and implications. *American Statistician* 40, 12–26.

Dickey, D.A. & 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.

Dickey, D.A. & W.A. Fuller (1981) The likelihood ratio statistics for autoregressive time series with a unit root. *Econometrics* 49, 1057–1072.

Elliot, G., T.J. Rothenberg, & J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 4, 813–836.

The GAUSS System Version 3.0 (1992) Washington, D.C.: Aptech Systems.

Hansen, H. & A. Rahbek (2002) Approximate conditional unit root inference. *Journal of Time Series Analysis* 23, 1–28.

Lee, J. & P. Schmidt (1994) Unit root tests based on instrumental variables estimation. *International Economic Review* 35, 449–462.

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

Neyman, J. & E.L. Scott (1948) Consistent estimates based on partially consistent observations. *Econometrica* 16, 1–32.
Pantula, S.G., G. Gonzalez–Farias, & W.A. Fuller (1994) A comparison of unit-root test criteria. *Journal of Business and Economic Statistics* 12, 449–459.
Pere, P. (1997) *Adjusted Profile Likelihood Applied to Estimation and Testing of Unit Roots*. D.Phil. thesis, University of Oxford.
Pere, P. (2000) Adjusted estimates and Wald statistics for the AR(1) model with constant. *Journal of Econometrics* 98, 335–363.
Phillips, P.C.B. (1987) Towards a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.
Phillips, P.C.B. & P. Perron (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–346.
Robinson, P.M. (1994) Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association* 89, 1420–1437.
Saikkonen, P. & R. Luukkonen (1997) Testing cointegration in infinite order vector autoregressive processes. *Journal of Econometrics* 81, 93–126.
Schmidt, P. & J. Lee (1991) A modification of the Schmidt–Phillips unit root test. *Economics Letters* 36, 285–289.
Schmidt, P. & P.C.B. Phillips (1992) LM tests for a unit root in the presence of deterministic time trends. *Oxford Bulletin of Economics and Statistics* 54, 257–287.
Seber, G.A.F. & C.J. Wild (1989) *Nonlinear Regression*. New York: Wiley.
Stock, J.H. (1994) Unit Roots, Structural Breaks and Trends. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4. Amsterdam: North-Holland.
West, K.D. (1988) Asymptotic normality, when regressors have a unit root. *Econometrica* 56, 1397–1417.
Zivot, E. (1994) A Bayesian analysis of the unit root hypothesis within an unobserved components model. *Econometric Theory* 10, 552–578.

APPENDIX A: NEAR UNIT-ROOT ASYMPTOTICS FOR THE $AR^B(1)$ MODEL

Calculations in this Appendix are valid under

$$\psi = \exp(\delta/T) = 1 + \frac{\delta/T}{1!} + \frac{(\delta/T)^2}{2!} + \cdots = 1 + O(T^{-1}). \quad \textbf{(A.1)}$$

Straightforward calculations would show that the relevant sums entailing innovations and observations under the near unit-root $AR^B(1)$ process yield the standard asymptotic convergencies employed subsequently. The unit-root results follow by setting $\delta = 0$.

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

Proving That $\hat{\gamma}_{ml}$ Is $O_p(1)$ and Equals y_1 Asymptotically under a Near Unit Root. The PL for γ is

$$l_p(\gamma) \propto -\frac{T}{2}\log\hat{\sigma}_\gamma^2,$$

where

$$\begin{aligned}\hat{\sigma}_\gamma^2 &= T^{-1}\left\{(y_1-\gamma)^2+\sum_{t=2}^{T}[y_t-\gamma-\hat{\psi}_\gamma(y_{t-1}-\gamma)]^2\right\}\\ &= T^{-1}\left\{\sum_{t=1}^{T}(y_t-\gamma)^2-\frac{\left[\sum_{t=2}^{T}(y_t-\gamma)(y_{t-1}-\gamma)\right]^2}{\sum_{t=2}^{T}(y_{t-1}-\gamma)^2}\right\}\end{aligned} \tag{A.2}$$

and $\hat{\psi}_\gamma$ is defined in (10). Some algebra, the relation $y_t = \gamma(1-\psi)+\psi y_{t-1}+\varepsilon_t$ for $t \geq 2$, and well-known orders of magnitude of the sums imply that

$$\begin{aligned}\hat{\sigma}_\gamma^2 &= T^{-1}\left\{(y_1-\gamma)^2+\sum_{t=2}^{T}\epsilon_t^2\right.\\ &\qquad\left.-\frac{\left(\sum_{t=2}^{T}y_{t-1}\epsilon_t\right)^2-2\gamma\sum_{t=2}^{T}y_{t-1}\epsilon_t\sum_{t=2}^{T}\epsilon_t+\gamma^2\left(\sum_{t=2}^{T}\epsilon_t\right)^2}{\sum_{t=2}^{T}y_{t-1}^2-2\gamma\sum_{t=2}^{T}y_{t-1}+(T-1)\gamma^2}\right\}\\ &= T^{-1}\left\{(y_1-\gamma)^2+\sum_{t=2}^{T}\epsilon_t^2\right.\\ &\qquad\left.-\frac{[O_p(T)]^2-2\gamma O_p(T)O_p(T^{1/2})+\gamma^2[O_p(T^{1/2})]^2}{O_p(T^2)-2\gamma O_p(T^{3/2})+\gamma^2 O(T)}\right\}\\ &\propto (y_1-\gamma)^2-\frac{O_p(T^2)-2\gamma O_p(T^{3/2})+\gamma^2 O_p(T)}{O_p(T^2)-2\gamma O_p(T^{3/2})+\gamma^2 O(T)}.\end{aligned} \tag{A.3}$$

By simple arguments it can be confirmed from the preceding expression that $\hat{\gamma}_{ml}$—the solution to the minimization problem $\inf_\gamma(\hat{\sigma}_\gamma^2)$—cannot diverge. The minimization problem is thus $\inf_\gamma[(y_1-\gamma)^2-\gamma O_p(T^{-1/2})]$, which yields $\hat{\gamma}_{ml}=y_1$ for large T.

Approximations for $\hat{\psi}_{ml}^B$ under a Near Unit Root. Substitution of $y_t = \gamma+\psi(y_{t-1}-\gamma)+\varepsilon_t$ for $t \geq 2$ to the second line of (11) and the approximations $T(\psi-1)=\delta+O(T^{-1})=O(1)$ from (A.1) yield

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

$$T(\hat{\psi}^B_{ml}-1)=\frac{T(\psi-1)T^{-2}\sum_{t=2}^{T}y^2_{t-1}-\gamma T(\psi-1)T^{-2}\sum_{t=2}^{T}y_{t-1}+T^{-1}\sum_{t=2}^{T}y_{t-1}\varepsilon_t-\hat{\gamma}_{ml}T^{-1}(y_T-y_1)}{T^{-2}\sum_{t=2}^{T}y^2_{t-1}-2\hat{\gamma}_{ml}T^{-2}\sum_{t=2}^{T}y_{t-1}+T^{-2}(T-1)\hat{\gamma}^2_{ml}}$$

$$=\frac{[\delta+O(T^{-1})]T^{-2}\sum_{t=2}^{T}y^2_{t-1}-\gamma O(1)O_p(T^{-1/2})+T^{-1}\sum_{t=2}^{T}y_{t-1}\varepsilon_t}{T^{-2}\sum_{t=2}^{T}y^2_{t-1}-2O_p(1)O_p(T^{-1/2})+O(T^{-1})O_p(1)}$$

$$-\frac{O_p(1)O_p(T^{-1/2})}{T^{-2}\sum_{t=2}^{T}y^2_{t-1}-2O_p(1)O_p(T^{-1/2})+O(T^{-1})O_p(1)}.$$

One could find the asymptotic distribution directly from the preceding expression. Instead, an approximation is developed—to be used later—from which the distribution can be reasoned. From the preceding expression

$$T(\hat{\psi}^B_{ml}-1)=\frac{\delta T^{-2}\sum_{t=2}^{T}y^2_{t-1}+T^{-1}\sum_{t=2}^{T}y_{t-1}\varepsilon_t}{T^{-2}\sum_{t=2}^{T}y^2_{t-1}[1+O_p(T^{-1/2})]}+O_p(T^{-1/2})$$

$$=\delta+T(\hat{\psi}^0_{ml}-\psi)+O_p(T^{-1/2})$$

$$=T(\hat{\psi}^0_{ml}-1)+O_p(T^{-1/2}), \qquad \textbf{(A.4)}$$

where the last equality is a consequence of (A.1). It follows that

$$\hat{\psi}^B_{ml}=\hat{\psi}^0_{ml}+r, \qquad r=O_p(T^{-3/2}), \qquad \textbf{(A.5)}$$

under $\psi=\exp(\delta/T)$. The asymptotic near unit-root distribution of $T(\hat{\psi}^B_{ml}-1)$ is in (25). The distribution can be inferred from (A.4) or (A.5), because statistics of which $\hat{\psi}^0_{ml}$ is composed share the same asymptotic distributions whether the generating model is $\mathrm{AR}^B(1)$ or $\mathrm{AR}^0(1)$.

By $O_p(1)$ness of $T(\hat{\psi}^0_{ml}-1)$, (A.1), and (A.5):

$$\hat{\psi}^B_{ml}=\psi+s, \qquad s=O_p(T^{-1}), \qquad \textbf{(A.6)}$$

under $\psi=\exp(\delta/T)$.

Consistency of $\hat{\sigma}^2_\psi$ Evaluated at the MLE or AE under a Near Unit Root. Consistency of $\hat{\sigma}^2_\psi$, evaluated at $\hat{\psi}^B_{ml}$ of ψ under a near unit root, is easily proved by substituting (A.6) into (7). Consistency of $\hat{\sigma}^2_\psi|_{\psi=\hat{\psi}^B_a}$ can be proved similarly.

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

Asymptotic Distribution of the Wald Statistic under a Near Unit Root. An important ingredient of the Wald statistic is $\partial^2 l_p(\psi)/\partial\psi^2$:

$$\frac{\partial^2 l_p(\psi)}{\partial\psi^2} = -\frac{T}{2}\left[\frac{\partial^2\hat{\sigma}_\psi^2/\partial\psi^2}{\hat{\sigma}_\psi^2} - \frac{(\partial\hat{\sigma}_\psi^2/\partial\psi)^2}{\hat{\sigma}_\psi^4}\right]$$

(cf. (9)), where

$$\begin{aligned}\frac{\partial\hat{\sigma}_\psi^2}{\partial\psi} &= 2T^{-1}\sum_{t=2}^{T}[y_t - \hat{\gamma}_\psi(1-\psi) - \psi y_{t-1}](\hat{\gamma}_\psi - y_{t-1}) \\ &= 2T^{-1}\left\{-\sum_{t=2}^{T} y_t y_{t-1} + \psi\sum_{t=2}^{T} y_{t-1}^2 \right. \\ &\quad + \frac{(1-\psi)\left[y_1 + (1-\psi)\left(\sum_{t=2}^{T} y_t - \psi\sum_{t=2}^{T} y_{t-1}\right)\right]\sum_{t=2}^{T} y_{t-1}}{1+(T-1)(1-\psi)^2} \\ &\quad \left. + \frac{\left[y_1 + (1-\psi)\left(\sum_{t=2}^{T} y_t - \psi\sum_{t=2}^{T} y_{t-1}\right)\right]\left[\left(\sum_{t=2}^{T} y_t - \psi\sum_{t=2}^{T} y_{t-1}\right) - (T-1)(1-\psi)y_1\right]}{[1+(T-1)(1-\psi)^2]^2}\right\},\end{aligned} \tag{A.7}$$

$$\frac{\partial^2\hat{\sigma}_\psi^2}{\partial\psi^2} = 2T^{-1}\left\{\sum_{t=2}^{T}(y_{t-1} - \hat{\gamma}_\psi)^2 - \frac{\left\{\sum_{t=2}^{T}[y_t - \hat{\gamma}_\psi(1-\psi) - \psi y_{t-1}] + (1-\psi)\sum_{t=2}^{T}(y_{t-1} - \hat{\gamma}_\psi)\right\}^2}{1+(T-1)(1-\psi)^2}\right\},$$

and

$$\frac{\partial\hat{\gamma}_\psi}{\partial\psi} = -\frac{\sum_{t=2}^{T}[y_t - \hat{\gamma}_\psi(1-\psi) - \psi y_{t-1}] + (1-\psi)\sum_{t=2}^{T}(y_{t-1} - \hat{\gamma}_\psi)}{1+(T-1)(1-\psi)^2}.$$

The Wald statistic $\sqrt{W_{ml}^B}$, the ingredients appropriately standardized, is

$$\sqrt{W_{ml}^B} = T(\hat{\psi}_{ml}^B - 1)\left[-T^{-2}\frac{\partial^2 l_p(\psi)}{\partial\psi^2}\bigg|_{\psi=\hat{\psi}_{ml}^B}\right]^{1/2}.$$

The derivatives of $\hat{\sigma}_\psi^2$—of which $\partial^2 l_p(\psi)/\partial\psi^2$ is composed—become multiplied by powers of T. By definition, $\partial\hat{\sigma}_\psi^2/\partial\psi$, evaluated at $\hat{\psi}_{ml}^B$, equals zero. The asymptotic distribution of the standardized second derivative, evaluated at $\hat{\psi}_{ml}^B$, is nondegenerate:

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

$$T^{-1}\left(\frac{\partial^2\hat{\sigma}_\psi^2}{\partial\psi^2}\right)\Bigg|_{\psi=\hat{\psi}_{ml}^B}$$

$$= 2T^{-2}\left\{\sum_{t=2}^{T}(y_{t-1}-\hat{\gamma}_{ml})^2 - \frac{\left[\sum_{t=2}^{T}\hat{\epsilon}_t + (1-\hat{\psi}_{ml}^B)\sum_{t=2}^{T}(y_{t-1}-\hat{\gamma}_{ml})\right]^2}{1+(T-1)(1-\hat{\psi}_{ml}^B)^2}\right\}$$

$$= 2T^{-2}\sum_{t=2}^{T}y_{t-1}^2 + O_p(T^{-1/2})$$

$$\Rightarrow 2\sigma^2\int[J_\delta(r)]^2\,dr, \tag{A.8}$$

where $\hat{\epsilon}_t = y_t - \hat{\gamma}_{ml}(1-\hat{\psi}_{ml}^B) - \hat{\psi}_{ml}^B y_{t-1}$, $t = 2,\ldots,T$. By these results and the consistency of $\hat{\sigma}_\psi^2|_{\psi=\hat{\psi}_{ml}^B}$, the asymptotic distribution of $\sqrt{W_{ml}^B}$ is

$$\sqrt{W_{ml}^B} = T(\hat{\psi}_{ml}^B - 1)\left\{\frac{1}{2}\left[\frac{T^{-1}\partial^2\hat{\sigma}_\psi^2/\partial\psi^2}{\hat{\sigma}_\psi^2} - \frac{(T^{-1/2}\partial\hat{\sigma}_\psi^2/\partial\psi)^2}{\hat{\sigma}_\psi^4}\right]\Bigg|_{\psi=\hat{\psi}_{ml}^B}\right\}^{1/2}$$

$$\Rightarrow \frac{\frac{1}{2}\{[J_\delta(1)]^2-1\}}{\int[J_\delta(r)]^2\,dr}\left(\frac{1}{2}\left\{\frac{2\sigma^2\int[J_\delta(r)]^2\,dr}{\sigma^2} - \frac{0}{\sigma^4}\right\}\right)^{1/2}$$

$$= \frac{\frac{1}{2}\{[J_\delta(1)]^2-1\}}{\left\{\int[J_\delta(r)]^2\,dr\right\}^{1/2}},$$

where $\partial\hat{\sigma}_\psi^2/\partial\psi \equiv 0$ is substituted on the second line and convergencies (25) and (A.8) and consistency of $\hat{\sigma}_\psi^2|_{\psi=\hat{\psi}_{ml}^B}$ are employed.

Construction and Near Unit-Root Asymptotics of $T(LR_{ml}^B - 1)$. From (7) and (13):

$$LR_{ml}^B = \frac{\sum_{t=1}^{T}y_t^2 - 2\sum_{t=2}^{T}y_t y_{t-1} + \sum_{t=2}^{T}y_{t-1}^2 - y_1^2}{\sum_{t=1}^{T}y_t^2 - 2\hat{\psi}_{ml}^B\sum_{t=2}^{T}y_t y_{t-1} + (\hat{\psi}_{ml}^B)^2\sum_{t=2}^{T}y_{t-1}^2 - b_\psi}\Bigg|_{\psi=\hat{\psi}_{ml}^B}$$

$$= \frac{-2(\hat{\psi}_{ml}^0 - 1) + \dfrac{y_T^2 - y_1^2}{\sum_{t=2}^{T}y_{t-1}^2}}{-2\hat{\psi}_{ml}^B\hat{\psi}_{ml}^0 + (\hat{\psi}_{ml}^B)^2 + 1 + \dfrac{y_T^2}{\sum_{t=2}^{T}y_{t-1}^2} - \dfrac{b_\psi}{\sum_{t=2}^{T}y_{t-1}^2}}\Bigg|_{\psi=\hat{\psi}_{ml}^B},$$

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

where

$$b_\psi = \frac{\left[y_1 + (1-\psi)\left(\sum_{t=2}^{T} y_t - \psi \sum_{t=2}^{T} y_{t-1}\right)\right]^2}{1 + (T-1)(1-\psi)^2}.$$

The asymptotic distribution of $T(LR^B_{ml} - 1)$ under the null of a near unit root will be derived. For this purpose, a few intermediate results on orders of magnitude are needed. First, $\sum_{t=2}^{T} y_t - \hat{\psi}^B_{ml} \sum_{t=2}^{T} y_{t-1} = \sum_{t=2}^{T} (y_t - \hat{\psi}^B_{ml} y_{t-1})$ is $O_p(T^{1/2})$ by superconsistency of $\hat{\psi}^B_{ml}$. The numerator of $b_\psi|_{\psi=\hat{\psi}^B_{ml}}$ is hence

$$\left[y_1 + (1-\hat{\psi}^B_{ml})\left(\sum_{t=2}^{T} y_t - \hat{\psi}^B_{ml} \sum_{t=2}^{T} y_{t-1}\right)\right]^2 = [y_1 - (\hat{\psi}^B_{ml} - 1)O_p(T^{1/2})]^2$$

$$= y_1^2 + O_p(T^{-1/2}).$$

An asymptotic expression for the denominator of $b_\psi|_{\psi=\hat{\psi}^B_{ml}}$ is likewise found to be

$$1 + (T-1)(1-\hat{\psi}^B_{ml})^2 = 1 + (T-1)\{1 - [\hat{\psi}^0_{ml} + O_p(T^{-3/2})]\}^2$$

$$= 1 + (T-1)(\hat{\psi}^0_{ml} - 1)^2 + O_p(T^{-3/2})$$

by (A.5). Thus, $(b_\psi / \sum_{t=2}^{T} y_{t-1}^2)|_{\psi=\hat{\psi}^B_{ml}}$ is

$$\left.\frac{b_\psi}{\sum_{t=2}^{T} y_{t-1}^2}\right|_{\psi=\hat{\psi}^B_{ml}} = \frac{y_1^2 \Big/ \sum_{t=2}^{T} y_{t-1}^2 + O_p(T^{-5/2})}{1 + (T-1)(\hat{\psi}^0_{ml} - 1)^2 [1 + O_p(T^{-3/2})]}$$

$$= \frac{y_1^2 \Big/ \sum_{t=2}^{T} y_{t-1}^2}{1 + (T-1)(\hat{\psi}^0_{ml} - 1)^2} + O_p(T^{-5/2}).$$

These results and (A.5) yield the following expression for the denominator of LR^B_{ml}:

$$-2(\hat{\psi}^0_{ml} + r)\hat{\psi}^0_{ml} + (\hat{\psi}^0_{ml} + r)^2 + 1 + \frac{y_T^2}{\sum_{t=2}^{T} y_{t-1}^2} - \frac{y_1^2 \Big/ \sum_{t=2}^{T} y_{t-1}^2}{1 + (T-1)(\hat{\psi}^0_{ml} - 1)^2} + O_p(T^{-5/2})$$

$$= -(\hat{\psi}^0_{ml} + 1)(\hat{\psi}^0_{ml} - 1) + \frac{y_T^2}{\sum_{t=2}^{T} y_{t-1}^2} - \frac{y_1^2 \Big/ \sum_{t=2}^{T} y_{t-1}^2}{1 + (T-1)(\hat{\psi}^0_{ml} - 1)^2} + O_p(T^{-5/2}).$$

Finally, the asymptotic distribution can be derived:

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

$$T(LR^B_{ml}-1) = T\left[\frac{-2(\hat{\psi}^0_{ml}-1)+\dfrac{y_T^2-y_1^2}{\sum_{t=2}^{T}y_{t-1}^2}}{-(\hat{\psi}^0_{ml}+1)(\hat{\psi}^0_{ml}-1)+\dfrac{y_T^2}{\sum_{t=2}^{T}y_{t-1}^2}-\dfrac{y_1^2\Big/\sum_{t=2}^{T}y_{t-1}^2}{1+(T-1)(\hat{\psi}^0_{ml}-1)^2}+O_p(T^{-5/2})}-1\right]$$

$$= T\left[\frac{-2(\hat{\psi}^0_{ml}-1)+\dfrac{y_T^2-y_1^2}{\sum_{t=2}^{T}y_{t-1}^2}+(\hat{\psi}^0_{ml}+1)(\hat{\psi}^0_{ml}-1)-\dfrac{y_T^2}{\sum_{t=2}^{T}y_{t-1}^2}}{-(\hat{\psi}^0_{ml}+1)(\hat{\psi}^0_{ml}-1)+\dfrac{y_T^2}{\sum_{t=2}^{T}y_{t-1}^2}-\dfrac{y_1^2\Big/\sum_{t=2}^{T}y_{t-1}^2}{1+(T-1)(\hat{\psi}^0_{ml}-1)^2}+O_p(T^{-5/2})}\right.$$

$$\left.+\frac{\dfrac{y_1^2\Big/\sum_{t=2}^{T}y_{t-1}^2}{1+(T-1)(\hat{\psi}^0_{ml}-1)^2}-O_p(T^{-5/2})}{-(\hat{\psi}^0_{ml}+1)(\hat{\psi}^0_{ml}-1)+\dfrac{y_T^2}{\sum_{t=2}^{T}y_{t-1}^2}-\dfrac{y_1^2\Big/\sum_{t=2}^{T}y_{t-1}^2}{1+(T-1)(\hat{\psi}^0_{ml}-1)^2}+O_p(T^{-5/2})}\right]$$

(accuracy in the denominator can be relaxed)

$$= T\left[\frac{(\hat{\psi}^0_{ml}-1)^2-\dfrac{y_1^2}{\sum_{t=2}^{T}y_{t-1}^2}+\dfrac{y_1^2\Big/\sum_{t=2}^{T}y_{t-1}^2}{1+(T-1)(\hat{\psi}^0_{ml}-1)^2}+O_p(T^{-5/2})}{-(\hat{\psi}^0_{ml}+1)(\hat{\psi}^0_{ml}-1)+\dfrac{y_T^2}{\sum_{t=2}^{T}y_{t-1}^2}+O_p(T^{-2})}\right]$$

$$\Rightarrow \frac{\left\{\dfrac{\frac{1}{2}\{[J_\delta(1)]^2-1\}}{\int[J_\delta(r)]^2\,dr}\right\}^2}{-2\left\{\dfrac{\frac{1}{2}\{[J_\delta(1)]^2-1\}}{\int[J_\delta(r)]^2\,dr}\right\}+\dfrac{[\sigma J_\delta(1)]^2}{\sigma^2\int[J_\delta(r)]^2\,dr}}$$

$$= \left(\frac{\frac{1}{2}\{[J_\delta(1)]^2-1\}}{\left\{\sqrt{\int[J_\delta(r)]^2\,dr}\right\}^{1/2}}\right)^2. \qquad \textbf{(A.9)}$$

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

The convergence follows from (25), consistency of $\hat{\psi}^0_{ml}$ under the $\mathrm{AR}^B(1)$ model by (A.5) and (A.6), and Phillips (1987, p. 539).

APPENDIX B: NEAR UNIT-ROOT ASYMPTOTICS FOR THE $\mathrm{AR}^\mu(1)$ MODEL

The asymptotic distributions for the adjusted statistics under the $\mathrm{AR}^\mu(1)$ model with $\psi = \exp(\delta/T)$ and $\alpha = 0$ are derived. It is assumed that the statistics are calculated conditional on y_0.

Near Unit-Root Asymptotics for the MLE of ψ. The well-known asymptotic near unit-root distribution of $T(\hat{\psi}^\mu_{ml} - 1)$ is

$$T(\hat{\psi}^\mu_{ml} - 1) \Rightarrow \frac{\frac{1}{2}\{[J_\delta(1)]^2 - 1\} - J_\delta(1)\int J_\delta(r)\,dr}{\int [J^\mu_\delta(r)]^2\,dr}$$

$$= DF^\mu_\delta. \tag{B.1}$$

Recalling (A.1), it follows that $T(\hat{\psi}^\mu_{ml} - \psi) \Rightarrow DF^\mu_\delta - \delta$ or that

$$\hat{\psi}^\mu_{ml} = \psi + O_p(T^{-1}) \xrightarrow{p} \psi. \tag{B.2}$$

The final standard result to point out is that the denominator of $T(\hat{\psi}^\mu_{ml} - \psi)$ or $T^{-2}\sum_{t=1}^T (y_{t-1} - \bar{y}_{-1})^2 \Rightarrow \sigma^2 \int_0^1 [J^\mu_\delta(r)]^2\,dr$.

Near Unit-Root Asymptotics for the AE of ψ. The AE of (23) equals

$$\hat{\psi}^\mu_a = \hat{\psi}^\mu_{ml} + \frac{T-3}{2\hat{c}} - \frac{T-3}{2\hat{c}}\left\{1 + \frac{4\hat{c}^2}{(T-3)^2}[(\hat{\psi}^\mu_{ml})^2 - \hat{1}]\right\}^{1/2}$$

$$= \hat{\psi}^\mu_{ml} + \frac{T-3}{2\hat{c}} - \frac{T-3}{2\hat{c}}\left\{1 + \frac{2\hat{c}^2}{(T-3)^2}[(\hat{\psi}^\mu_{ml})^2 - \hat{1}] + r_{1,T}\right\}, \qquad r_{1,T} = O_p(T^{-2}),$$

$$= \hat{\psi}^\mu_{ml} - \frac{\hat{c}[(\hat{\psi}^\mu_{ml})^2 - \hat{1}]}{T-3} + r_T, \qquad r_T = O_p(T^{-2}). \tag{B.3}$$

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

The preceding approximation derives from the first-order Taylor expansion around unity of $\sqrt{1+x} = 1 + \frac{1}{2}x + r_1$. Here $x > -1$ is assumed and the remainder term $r_1 = -\frac{1}{8}(1+ax)^{-3/2}x^2$ where $a \in (0,1)$ by Taylor's theorem. The orders of magnitude of $r_{1,T}$ and r_T are due to $\hat{c}$ being $O_p(T)$ and $(\hat{\psi}_{ml}^{\mu})^2 - \hat{1}$ being $O_p(T^{-1})$ under a near unit root, as is shown subsequently.

To find the asymptotic near unit-root distribution of $\hat{c}$, one needs the corresponding distributions for $\hat{\psi}_{ml}^{\mu}$, $(\hat{\psi}_{ml}^{\mu})^T$, and $(\hat{\psi}_{ml}^{\mu})^2 - \hat{1}$. The derivation in Pere (2000, p. 348) applies here by substituting DF_{δ}^{μ} in place of DF_{μ} there. This gives

$$\hat{c}/T \Rightarrow \frac{\exp(DF_{\delta}^{\mu}) - (1 + DF_{\delta}^{\mu})}{(DF_{\delta}^{\mu})^2}. \tag{B.4}$$

The next task is to derive the asymptotic near unit-root distribution for $(\hat{\psi}_{ml}^{\mu})^2 - \hat{1} = (\hat{\psi}_{ml}^{\mu} + 1)(\hat{\psi}_{ml}^{\mu} - 1) - (\hat{1} - 1)$. The asymptotic distribution and probability limit of the terms in the multiplication have been derived already in (B.1) and (B.2), respectively. The last term can be written as

$$\hat{1} - 1 = \frac{\sum_{t=1}^{T} y_t^2 - T(\bar{y})^2 - \left[\sum_{t=1}^{T} y_{t-1}^2 - T(\bar{y}_{-1})^2\right]}{\sum_{t=1}^{T}(y_{t-1} - \bar{y}_{-1})^2}.$$

The asymptotics of the denominator is the same as that of $\hat{\psi}_{ml}^{\mu}$, so, after multiplication by T^{-2}, it converges weakly to $\sigma^2 \int_0^1 [J_{\delta}^{\mu}(r)]^2\, dr$. An approximation for the numerator, $y_T(y_T - 2\bar{y}_{-1}) + O_p(T^{1/2})$, can be found similarly to the case under a unit root (Pere, 2000, p. 356). The numerator multiplied by T^{-1} is

$$T^{-1}[y_T(y_T - 2\bar{y}_{-1}) + O_p(T^{1/2})] = T^{-1/2}y_T\left(T^{-1/2}y_T - 2T^{-3/2}\sum_{t=1}^{T} y_{t-1}\right) + O_p(T^{-1/2})$$

$$\Rightarrow \sigma^2 J_{\delta}(1)\left[J_{\delta}(1) - 2\int J_{\delta}(r)\, dr\right].$$

Thus

$$T(\hat{1} - 1) \Rightarrow \frac{J_{\delta}(1)\left[J_{\delta}(1) - 2\int J_{\delta}(r)\, dr\right]}{\int [J_{\delta}^{\mu}(r)]^2\, dr}.$$

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

It follows from (B.1), (B.2), and the preceding result that

$$
\begin{aligned}
T[(\hat{\psi}_{ml}^{\mu})^2 - \hat{1}] &= (\hat{\psi}_{ml}^{\mu} + 1)T(\hat{\psi}_{ml}^{\mu} - 1) - T(\hat{1} - 1) \\
&\Rightarrow 2\left[\frac{\frac{1}{2}\{[J_\delta(1)]^2 - 1\} - J_\delta(1)\int J_\delta(r)\,dr}{\int [J_\delta^{\mu}(r)]^2\,dr}\right] \\
&\quad - \frac{J_\delta(1)\left[J_\delta(1) - 2\int J_\delta(r)\,dr\right]}{\int [J_\delta^{\mu}(r)]^2\,dr} \\
&= -\left\{\int [J_\delta^{\mu}(r)]^2\,dr\right\}^{-1}. \qquad \textbf{(B.5)}
\end{aligned}
$$

Combining (B.1), (B.3), (B.4), and (B.5) yields

$$
\begin{aligned}
T(\hat{\psi}_a^{\mu} - 1) &= T\left\{\hat{\psi}_{ml}^{\mu} - \frac{\hat{c}[(\hat{\psi}_{ml}^{\mu})^2 - \hat{1}]}{T-3} + O_p(T^{-2}) - 1\right\} \\
&= T(\hat{\psi}_{ml}^{\mu} - 1) - [\hat{c}/(T-3)]T[(\hat{\psi}_{ml}^{\mu})^2 - \hat{1}]\} + O_p(T^{-1}) \\
&\Rightarrow DF_\delta^{\mu} + \frac{\exp(DF_\delta^{\mu}) - (1 + DF_\delta^{\mu})}{(DF_\delta^{\mu})^2}\left\{\int [J_\delta^{\mu}(r)]^2\,dr\right\}^{-1}. \qquad \textbf{(B.6)}
\end{aligned}
$$

Near Unit-Root Asymptotics for the Adjusted Wald Statistic. A formulation of $\sqrt{W_a^{\mu}}$ is $(\hat{\psi}_a^{\mu} - 1)[-\partial^2 l_a(\psi)/\partial\psi^2|_{\psi=\hat{\psi}_a^{\mu}}]^{1/2}$ (Pere, 2000, (5), (18), and (23)). The form for the ratio derived on p. 358 of Pere (2000) extends straightforwardly by the previous results to the near unit-root case. Hence

$$
\begin{aligned}
\sqrt{W_a^{\mu}} &= T(\hat{\psi}_a^{\mu} - 1)\left\{-\frac{(T-3)[(\hat{\psi}_{ml}^{\mu})^2 - \hat{1} + r_T^*]}{T\{-[(\hat{\psi}_{ml}^{\mu})^2 - \hat{1}] + r_T^*\}^2}\right\}^{1/2} \\
&\Rightarrow \left(DF_\delta^{\mu} + \frac{\exp(DF_\delta^{\mu}) - (1 + DF_\delta^{\mu})}{(DF_\delta^{\mu})^2}\left\{\int [J_\delta^{\mu}(r)]^2\,dr\right\}^{-1}\right) \\
&\quad \times \left(-\frac{-\left\{\int [J_\delta^{\mu}(r)]^2\,dr\right\}^{-1}}{\left\{\int [J_\delta^{\mu}(r)]^2\,dr\right\}^{-2}}\right)^{1/2} \\
&= \left[DF_\delta^{\mu} + \frac{\exp(DF_\delta^{\mu}) - (1 + DF_\delta^{\mu})}{(DF_\delta^{\mu})^2\int [J_\delta^{\mu}(r)]^2\,dr}\right]\left\{\int [J_\delta^{\mu}(r)]^2\,dr\right\}^{1/2}.
\end{aligned}
$$

Here r_T^* is $O_p(T^{-2})$. The convergence follows from (B.5) and (B.6).

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

APPENDIX C: INVARIANCE AND SIMILARITY RESULTS FOR THE $AR^B(1)$ MODEL

If the process is a random walk then a meaningful estimator and test statistic of ψ should not depend on the value from which the process was initialized (x_0). Such an invariance property of the MLE and AE under a unit root is proved subsequently for the $AR^B(1)$ model. Invariance of the associated Wald and LR statistics, under a unit root, follows straightforwardly. Next, invariance of the statistics with respect to multiplication of the data is verified. Similarity of the associated unit-root tests is a direct consequence of these results. The PL is again a useful tool for proving the claims. The scrutinized statistics are based on $x_0 = 0$ even though $x_0 \neq 0$ is allowed for. Related results for the $AR^{\mu}(1)$ statistics have been proved in Pere (2000).

Addition of a constant a to the observations converts $\hat{\gamma}_{\psi}$ to $\hat{\gamma}_{\psi} + a$ (see (8)). The residual variance of (7),

$$\hat{\sigma}_{\psi}^2 = T^{-1}\left\{[y_1 + a - (\hat{\gamma}_{\psi} + a)]^2 + \sum_{t=2}^{T}[y_t + a - (\hat{\gamma}_{\psi} + a)(1-\psi) - \psi(y_{t-1} + a)]^2\right\}$$

$$= T^{-1}\left\{(y_1 - \hat{\gamma}_{\psi})^2 + \sum_{t=2}^{T}[y_t - \hat{\gamma}_{\psi}(1-\psi) - \psi y_{t-1}]^2\right\},$$

remains unchanged. Thus $\hat{\psi}_{ml}^B$ is tied to its original value regardless of an added constant. The adjustment term of (21) also remains unmodified, as does $\hat{\psi}_a^B$.

If $\psi = 1$, then $x_0 \neq 0$ has an analogous effect on y_t as adding a constant to y_t after the generation of it. It follows that $\hat{\psi}_{ml}^B$ is undisturbed by $x_0 \neq 0$ if $\psi = 1$, and one can allow for $x_0 \neq 0$ when testing for a unit root via $\hat{\psi}_{ml}^B$.

Invariance of $\hat{\psi}_a^B$ with respect to x_0 under $\psi = 1$ is not obvious because some of the influence measures, on which the adjustments rest, are disturbed by x_0. The formula

$$\hat{\sigma}_{\psi,x_0}^2 \propto \sum_{t=1}^{T} y_t^2 - 2\psi\left(y_1 x_0 + \sum_{t=2}^{T} y_t y_{t-1}\right) + \psi^2\left(x_0^2 + \sum_{t=2}^{T} y_{t-1}^2\right)$$

$$- \frac{\left[y_1 - \psi x_0 + (1-\psi)\left(\sum_{t=2}^{T} y_t - \psi\sum_{t=2}^{T} y_{t-1}\right)\right]^2}{1 + (T-1)(1-\psi)^2}$$

permits a nonzero x_0 and is the basis for the following calculations (cf. (7) for $x_0 = 0$). A sole information measure of relevance needs revision:

$$i_{\psi\gamma} = \begin{cases} \sigma^{-2}x_0(1 + \psi - \psi^T) & \text{if } |\psi| < 1 \\ \sigma^{-2}x_0 & \text{if } \psi = 1. \end{cases}$$

The parameters are nonorthogonal unless $x_0 = 0$, so the term $\hat{c}(\psi - \hat{\psi}_{ml})$ should be included in APL (14). However, the term is superfluous because c equals zero:

$$c = \frac{\partial}{\partial\gamma}(i_{\gamma\gamma}^{-1} i_{\psi\gamma}) = 0,$$

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

because neither of the information measures in the parentheses is a function of γ ((18) and (19)). The conclusion is that $\hat{\psi}_a^B$ is not affected by x_0 under $\psi = 1$, so it can be used as a basis for a unit-root test allowing for $x_0 \neq 0$.

The proposed unit-root tests are invariant with respect to multiplication of the data by a constant: the tests are derivatives of $\hat{\sigma}_\psi^2$ and $\log \hat{\sigma}_\psi^2$. Multiplication of the y_t's by a constant h transfers the estimators to $h^2\hat{\sigma}_\psi^2$and $\log h^2 + \log \hat{\sigma}_\psi^2$ by (7). The factor h^2 does not affect maximization of the PL or APL, and hence the SC statistics, and cancels when the LR statistics are constructed. In the calculation of the Wald statistics $\log h^2$ differentiates out.

Similarity of the tests follows straightforwardly. A change in γ corresponds to addition of a constant to the observations and is irrelevant as explained before. Multiplying the innovations by h or altering σ^2 to $h\sigma^2$ has the same effect as multiplication of the y_ts by h if $\gamma = x_0 = 0$ (cf. (3)), which is inconsequential as explained previously. The same alteration is inconsequential even if $\gamma \neq 0$ or $x_0 \neq 0$ because any constant including γ or x_0 can be deducted from the data first, without affecting the analysis as reasoned earlier, and the previous argument can be applied next. The tests are hence similar whether $x_0 = 0$ or $x_0 \neq 0$.

APPENDIX D: ON THE SIMULATIONS

The GAUSS 3.2.17 (1992) simulation programs in essence (excluding the code for the LR tests) are given in Pere (1997). The advancement here is that the simulations entail 300,000 draws composed of 10 subsamples with different seeds and the power results for the $\mathrm{AR}^B(1)$ model involve 100,000 draws. Such splitting is recommended by Abadir (1995) for strongly autocorrelated processes to lessen dependence of the results on a seed. The set of seeds is common for each value of ψ when simulating power. The GAUSS code for performing the simulations is available from the author upon request.

In the estimation of the asymptotic fractiles Brownian motions are approximated with random walks (z_t) of length 10,000. For example, $\int [B(r)]^2\,dr$ is approximated by $\sum_{t=1}^{10,000} z_t^2/10{,}000^2$.

The statistics for the $\mathrm{AR}^B(1)$ model are found numerically with the default optimization algorithm in the OPTMUM 3.1.4 module of GAUSS using initial value of unity for ψ. Under a unit root, the magnitude of γ does not theoretically alter the MLE. With OPTMUM, however, a large γ may decelerate convergence apparently indefinitely (Pere, 1997), so it was fixed at zero for the present results. Relatedly, the program may not reproduce exactly the same results if two different values of σ^2 are experimented with. The explanation is that if OPTMUM enters a random search in an iteration, different sequences of time series arise despite a common set of seeds for two Monte Carlo experiments.

Appropriate convergence of the algorithm was confirmed by recording the proportion of cases in which the residual variance is not smaller at the MLE, according to the algorithm. The proportion was very small when simulating the fractiles (Table D.1). In the few inspected incidents, numerical inaccuracy was the explanation because the variances departed in the eighth decimal place, say.

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press

TABLE D.1. Percentage of cases where the residual variance took a smaller value under the AE than the MLE when simulating the unit-root distributions ($AR^B(1)$)

T	25	50	75	100	250	500	1,000
%	0.14	0.07	0.03	0.02	0.01	0.04	0.19

Table 2 is constructed so that for each draw of a time series the MLEs under the $AR^B(1)$ and $AR^0(1)$ model are computed. For the latter model the MLE is conditional on y_1, which is random instead of fixed as in (5), but it should not be consequential. The fractiles for the MLE under the $AR^\mu(1)$ model, needed for Table 3, are calculated conditional on y_1, too, but the randomness does not affect the fractiles (Appendix C). Table 12 is constructed analogously to Tables 2 and 3. The fractiles (for each sample size) and the results on power (for each value of ψ), respectively, are based on the same seeds but not necessarily the same time series as explained earlier.

https://doi.org/10.1017/S0266466603196016 Published online by Cambridge University Press