CENTRE FOR ECONOMETRIC ANALYSIS
CEA@Cass

http://www.cass.city.ac.uk/cea/index.html

Cass Business School
Faculty of Finance
106 Bunhill Row
London EC1Y 8TZ

Jackknife Bias Reduction in Autoregressive Models with a Unit Root

Marcus J. Chambers and Maria Kyriacou

CEA@Cass Working Paper Series

WP–CEA–02-2012

Jackknife Bias Reduction in Autoregressive Models with a Unit Root

Marcus J. Chambers
University of Essex

and

Maria Kyriacou
University of Southampton

May 2011

ABSTRACT: This paper is concerned with the application of jackknife methods as a means of bias reduction in the estimation of autoregressive models with a unit root. It is shown that the usual jackknife estimator based on non-overlapping sub-samples does not remove fully the first-order bias as intended, but that an 'optimal' jackknife estimator can be defined that is capable of removing this bias. The results are based on a demonstration that the sub-sample estimators converge to different limiting distributions, and the joint moment generating function of the numerator and denominator of these distributions (which are functionals of a Wiener process over a sub-interval of [0,1]) is derived and utilised to extract the optimal weights. Simulations demonstrate the ability of the jackknife estimator to produce substantial bias reductions in the parameter of interest. It is also shown that incorporating an intercept in the regressions allows the standard jackknife estimator to be used and it is able also to produce substantial bias reduction despite the fact that the distributions of the full-sample and sub-sample estimators have greater bias in this case. Of interest, too, is the fact that the jackknife estimators can also reduce the overall root mean squared error compared to the ordinary least squares estimator, this requiring a larger (though still small) number of sub-samples compared to the value that produces maximum bias reduction (which is typically equal to two).

KEY WORDS: Jackknife; bias reduction; unit root; moment generating function.

ACKNOWLEDGEMENTS: The first author's research was funded by the Economic and Social Research Council under grant number RES-000-22-3082.

ADDRESS FOR CORRESPONDENCE: Professor Marcus J. Chambers, Department of Economics, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, England.
Tel: +44 1206 872756; fax: +44 1206 872724; e-mail: mchamb@essex.ac.uk.

1. INTRODUCTION

The nature of the bias of ordinary least squares (OLS) estimators of the parameters in stationary autoregressive (AR) processes has been extensively studied and its properties are well understood. Early contributions to this topic can be found in Marriott and Pope (1954), Kendall (1954) and Shenton and Johnson (1965), for example, who provide results for the first-order model, while Shaman and Stine (1988) have provided expansions of the bias in terms of the inverse of the sample size for more general AR processes of finite order p. Fewer theoretical results are available for such AR(p) processes that have a root on the unit circle, although Phillips (1987) provides an expansion in the case $p = 1$ with Gaussian innovations and a fixed initial value of zero. Simulation studies of the AR(1) model with a unit root, however, have found the OLS estimator to be significantly negatively biased in finite samples, and a number of methods have been proposed for eliminating or reducing the finite sample bias. Such methods include (but are not limited to): first-order bias correction, as explored by Orcutt and Winokur (1969); bootstrapping, an early motivation for its use in connection with bias reduction in AR models being Stine (1987); recursive mean adjustment, suggested for the unit root AR case by Shin and So (2001); and exact median unbiased estimation of the AR(1) model with a unit root proposed by Andrews (1993).

A longstanding method of bias reduction whose properties have been less extensively explored in AR models is the jackknife of Quenouille (1956), which was subsequently shown by Tukey (1958) to be a useful method for constructing a nonparameteric estimator of variance. In a recent application to bond option pricing in finance Phillips and Yu (2005) found a jackknife estimator to provide substantial reductions in bias compared to OLS and maximum likelihood estimators of the parameters in an AR(1) model arising from an underlying continuous time model of the interest rate. The jackknife estimator studied by Phillips and Yu (2005) is extremely straightforward to compute as it involves the use of only a small number of non-overlapping sub-samples, with considerable reductions in bias being obtained with just two, three or four such sub-samples. In view of the properties of many time series in economics and finance being characterised by finite-order AR models with a unit root, as well as the prevalence of such models in theoretical work in time series statistics and econometrics, it is therefore of interest to investigate whether the jackknife techniques described above can be applied in these models and, if so, to examine their properties.

In the context of an AR(p) model with a unit root we are able to show that the usual formulation of the jackknife estimator, as used by Phillips and Yu (2005), is no longer applicable in the sense that it is unable to remove fully the first-order bias of the OLS estimator as intended. The source of this failure is that the sub-sample estimators do not possess the same limiting distributions as the full-sample estimator and, because it is these distributions which motivate the bias expansions, the usual construction of the jackknife fails to work as intended. We demonstrate this feature by deriving the joint moment generating function (MGF) of the numerator and denominator of the limiting distribution of the sub-sample

estimators, both of which are functions of a Wiener process on sub-intervals of $[0,1]$. Furthermore this MGF can be used to compute the expectations of the ratios of the numerator to the denominator of the relevant limiting distributions, thereby enabling us to quantify the nature of the differences in expected values and also to compute 'optimal' jackknife weights that remove fully the first-order bias under a unit root as intended. Simulations are used to examine the extent of bias reduction possible in the $p = 1$ and $p = 2$ cases and is shown to be considerable. We also demonstrate an alternative method under which the usual jackknife weights are optimal, which simply requires the incorporation of an intercept in the regressions. Although the effect of this is to produce even greater bias in both the full- and sub-sample OLS estimators we demonstrate that the jackknife estimator can still eliminate a substantial amount of this bias. The results we obtain concerning the jackknife estimator provide a basis for examining (in future work) its use in testing for a unit root in an AR(p) model, while the results concerning the joint MGF of the numerator and denominator of the limiting distributions of sub-samples may have applications in other sub-sampling procedures under a unit root.

The paper is organised as follows. Section 2 defines the model and the jackknife estimator and derives the limiting distribution of the estimator in Theorem 1. The result concerning the joint MGF of the numerator and denominator of the sub-sample estimators is presented in Theorem 2 along with an expression for the expectation of the ratio of these two components. The optimal weights are presented in Theorem 3 and the performance of the optimal estimator is examined via a simulation exercise. Section 3 shows that the usual jackknife estimator works as intended if an intercept is included in the regressions although, as stated above, the distributions themselves have greater bias (the distributions being presented in Theorem 4). However, simulations reveal that the jackknife estimator is capable of large bias reductions here too. Although the jackknife is intended as a means of bias reduction it is also shown in the simulations that a reduction in the overall root mean squared error (RMSE) can be obtained by an appropriate choice of the number of non-overlapping sub-samples employed; this number tends to be greater than that which produces maximum bias reduction, although bias reduction remains a part of the overall reduction in RMSE. Some concluding comments, along with some directions for future research, are contained in Section 5, and all proofs are presented in an Appendix.

The following notation will be used throughout the paper. The symbol $\stackrel{d}{=}$ denotes equality in distribution; $\xrightarrow{d}$ denotes convergence in distribution; $\xrightarrow{p}$ denotes convergence in probability; $\Rightarrow$ denotes weak convergence of the relevant probability measures; and $W(r)$ denotes a Wiener process on $C[0,1]$, the space of continuous real-valued functions on the unit interval. Functionals of $W(r)$, such as $\int_0^1 W(r)^2 dr$, shall be denoted $\int_0^1 W^2$ for notational convenience.

2. JACKKNIFE BIAS REDUCTION WITH A UNIT ROOT

The focus is on a sequence of observations generated as follows.

Assumption 1. The sequence $y_1, \ldots, y_n$ satisfies

$$y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \epsilon_t, \quad t = 1, \ldots, n, \tag{1}$$

where ϵ_t is an iid$(0, \sigma_\epsilon^2)$ process with finite fourth moment and $y_{-p+1}, \ldots, y_0$ can be any $O_p(1)$ random variables (including constants). In addition, the equation

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p = 0$$

has a single root of $z = 1$ and all other roots lie outside the unit circle.

It is convenient to re-parameterise (1) as

$$y_t = \rho y_{t-1} + \sum_{j=1}^{p-1} \zeta_j \Delta y_{t-j} + \epsilon_t, \quad t = 1, \ldots, n, \tag{2}$$

where $\rho = \sum_{j=1}^{p} \phi_j$ and $\zeta_j = -\sum_{k=j+1}^{p} \phi_k$ $(j = 1, \ldots, p-1)$. When $\phi(1) = 0$, as under Assumption 1, it follows that $\rho = 1$ and $\Delta y_t = y_t - y_{t-1}$ has the representation $\zeta(L)\Delta y_t = \epsilon_t$, where $\zeta(z) = 1 - \sum_{j=1}^{p-1} \zeta_j z^j$ and the equation $\zeta(z) = 0$ has all roots outside the unit circle. OLS regression on (2) yields

$$y_t = \hat{\rho} y_{t-1} + \sum_{j=1}^{p-1} \hat{\zeta}_j \Delta y_{t-j} + e_t, \quad t = 1, \ldots, n, \tag{3}$$

where e_t denotes the regression residual, and it can be shown that $\hat{\rho}$ satisfies

$$n(\hat{\rho} - 1) = \frac{n^{-1} \sum_{t=1}^{n} y_{t-1} \epsilon_t}{n^{-2} \sum_{t=1}^{n} y_{t-1}^2} + o_p(1) \Rightarrow \frac{\zeta(1) \int_0^1 W dW}{\int_0^1 W^2} \quad \text{as} \quad n \to \infty; \tag{4}$$

a justification for this result can be found in the Appendix. The limiting distribution in (4) is skewed and the estimator suffers from significant negative bias in finite samples.

The finite sample bias has been demonstrated in a number of studies, both theoretically and using simulations. For example, Phillips (1987, Theorem 7.1) considered the Gaussian random walk

$$y_t = \rho y_{t-1} + \epsilon_t, \quad \rho = 1, \quad \epsilon_t \sim N(0, \sigma^2), \quad y_0 = 0,$$

demonstrating the validity of an asymptotic expansion for the normalised coefficient estima-

tor; it is given by

$$n(\hat{\rho}-1) \stackrel{d}{=} \frac{\int_0^1 W dW}{\int_0^1 W^2} - \frac{\eta}{\sqrt{2n}\int_0^1 W^2} + O_p(n^{-1}), \tag{5}$$

where η is a standard normal random variable distributed independently of W. Taking expectations in (5), using the independence of η and W, and noting that the expected value of the leading term is -1.781 (see, for example, Table 7.1 of Tanaka, 1996), the bias satisfies

$$E(\hat{\rho}-1) = -\frac{1.781}{n} + o(n^{-1}), \tag{6}$$

an expansion that can be used to motivate the use of the jackknife as a method of bias reduction; a similar, but not identical, expansion can be expected to hold in the more general setting of (1) in view of (4).

The jackknife offers a simple method of eliminating the leading bias term from expansions of the form (6). The jackknife estimator combines the full-sample estimator, $\hat{\rho}$, with a set of m sub-sample estimators, $\hat{\rho}_j$ $(j = 1, \ldots, m)$, the weights assigned to these components depending on the type of sub-sampling method employed. Phillips and Yu (2005) find the use of non-overlapping sub-samples to perform well in reducing bias in the estimation of stationary diffusions, and so it is this approach that shall be followed here. The jackknife estimator is

$$\hat{\rho}_J = w_{1m}\hat{\rho} + w_{2m}\frac{1}{m}\sum_{j=1}^{m}\hat{\rho}_j, \tag{7}$$

where the weights are given by $w_{1m} = m/(m-1)$ and $w_{2m} = -1/(m-1)$ and the length of each sub-sample is ℓ with $n = m \times \ell$. The weights are determined on the assumption that each sub-sample estimator also satisfies (6), so that

$$E(\hat{\rho}_j - 1) = -\frac{1.781}{\ell} + o(\ell^{-1}), \quad j = 1, \ldots, m.$$

In this case it can be shown that

$$\begin{aligned} E(\hat{\rho}_J - 1) &= \frac{m}{m-1}E(\hat{\rho}-1) - \frac{1}{m-1}\frac{1}{m}\sum_{j=1}^{m}E(\hat{\rho}_j - 1) \\ &= \frac{m}{m-1}\left(-\frac{1.781}{n} + o(n^{-1})\right) - \frac{1}{m-1}\left(-\frac{1.781}{\ell} + o(\ell^{-1})\right) \\ &= -\frac{1.781}{m-1}\left(mn^{-1} - \ell^{-1}\right) + o(n^{-1}) = o(n^{-1}), \end{aligned}$$

using the fact that $m/n = 1/\ell$. Under such circumstances the jackknife estimator is capable

of completely eliminating the $O(n^{-1})$ bias term in the estimator as compared to $\hat{\rho}$.

The problem with the argument above is that the sub-sample estimators do not share the same limiting distribution as the full-sample estimator, which means that the expansions for the bias of the sub-sample estimators are incorrect. To demonstrate this feature, let

$$\tau_j = \{(j-1)\ell + 1, \ldots, j\ell\}, \quad j = 1, \ldots, m,$$

denote the set of integers determining the observations in sub-sample j. The sub-sample estimator can be written, in view of (4), as

$$\ell\left(\hat{\rho}_j - 1\right) = \frac{\ell^{-1}\sum_{t\in\tau_j} y_{t-1}\epsilon_t}{\ell^{-2}\sum_{t\in\tau_j} y_{t-1}^2} + o_p(1), \quad j = 1, \ldots, m. \tag{8}$$

Theorem 1 states the limiting distributions of $\ell(\hat{\rho}_j - 1)$ $(j = 1, \ldots, m)$ and also of the jackknife estimator $n(\hat{\rho}_J - 1)$. In presenting the results it is convenient to define the functionals

$$Z(W, \delta) = \frac{\int_0^1 W dW}{\int_0^1 W^2}, \quad Z(W, \delta_j) = \frac{\int_{(j-1)/m}^{j/m} W dW}{\int_{(j-1)/m}^{j/m} W^2}, \quad j = 1, \ldots, m,$$

where the intervals $\delta = [0, 1]$ and $\delta_j = [(j-1)/m, j/m]$ denote the ranges of integration.

Theorem 1. Under Assumption 1, if $\ell \to \infty$ as $n \to \infty$:

(a) If m is fixed, $\ell(\hat{\rho}_j - 1) \Rightarrow m^{-1}\zeta(1)Z(W, \delta_j)$ $(j = 1, \ldots, m)$ and

$$n(\hat{\rho}_J - 1) \Rightarrow w_{1m}\zeta(1)Z(W, \delta) + w_{2m}\zeta(1)\sum_{j=1}^{m} m^{-1}Z(W, \delta_j);$$

(b) If $m^{-1} + mn^{-1} \to 0$, $n(\hat{\rho}_J - 1) \Rightarrow \zeta(1)Z(W, \delta)$.

Although it is natural to normalise $\hat{\rho}_j$ in part (a) of Theorem 1 by the sub-sample size ℓ, the stated result is valid only when m is fixed, otherwise the limiting distribution is degenerate. This is because both components of $Z(W, \delta_j)$, namely $\int_{(j-1)/m}^{j/m} W dW$ and $\int_{(j-1)/m}^{j/m} W^2$, are $O_p(1/m)$, which means that the stated distribution $m^{-1}Z(W, \delta_j)$ is also $O_p(1/m)$ due to the presence of m in the denominator. Multiplying by m, of course, provides the limit for $n(\hat{\rho}_j - 1)$ in terms of an $O_p(1)$ random variable and is valid even when m is not

held fixed. Note, too, that the numerator of $Z(W, \delta_j)$ also has the representation

$$\int_{(j-1)/m}^{j/m} W dW \stackrel{d}{=} \frac{1}{2}\left[W\left(\frac{j}{m}\right)^2 - W\left(\frac{j-1}{m}\right)^2 - \frac{1}{m}\right] \tag{9}$$

which follows from the Ito calculus; see, for example, equation (2.58) of Tanaka (1996, p.59). The familiar result, $\int_0^1 W dW = [W(1)^2 - 1]/2$, is a special case.

The limiting distribution of the jackknife estimator in Theorem 1 takes one of two forms, depending on whether m is fixed or is allowed to increase with n in conjunction with ℓ. When m is fixed the limiting distribution is a weighted average of the limiting distribution of $n(\hat{\rho} - 1)$ and of the sub-samples $\ell(\hat{\rho}_j - 1)$. Allowing m to increase with n results in the jackknife estimator inheriting the same limiting distribution as the full-sample estimator $\hat{\rho}$. Note, too, that the condition $m^{-1} + mn^{-1} \to 0$ also implies that $\ell \to \infty$ because $mn^{-1} = \ell^{-1}$.

The fact that the distributions $Z(W, \delta_j)$ in part (a) of Theorem 1 depend on j implies that the expansions for $E(\hat{\rho}_j - 1)$ that are used to derive the jackknife weights may not be correct under a unit root. The following result provides the joint moment generating function (MGF) of the numerator and denominator of the limiting distributions defined over a subinterval $\delta_{a,b} = [a, b]$ of $[0, b]$ where $0 \leq a < b$, which then enables the expectations of the limiting distributions in Theorem 1 to be calculated. The results are presented in greater generality than is required for the specific application under consideration because they may have more widespread use beyond this particular application.

Theorem 2. Let $N = \int_a^b W(r) dW(r)$ and $D = \int_a^b W(r)^2 dr$, where $W(r)$ is a Wiener process on $r \in [0, b]$ and $0 \leq a < b$. Then:

(a) The joint MGF of N and D is given by

$$M(\theta_1, \theta_2) = E \exp(\theta_1 N + \theta_2 D) = \exp\left(-\frac{\theta_1}{2}(b-a)\right) H(\theta_1, \theta_2)^{-1/2},$$

where, defining $\lambda = \sqrt{-2\theta_2}$,

$$H(\theta_1, \theta_2) = \cosh\left((b-a)\lambda\right) - \frac{1}{\lambda}\left[\theta_1 + a\left(\theta_1^2 - \lambda^2\right)\right] \sinh\left((b-a)\lambda\right).$$

(b) The expectation of N/D is given by

$$E\left(\frac{N}{D}\right) = \int_0^\infty \left.\frac{\partial M(\theta_1, -\theta_2)}{\partial \theta_1}\right|_{\theta_1 = 0} d\theta_2 = I_1(\delta_{a,b}) - I_2(\delta_{a,b}),$$

where $\delta_{a,b} = [a, b]$ and

$$I_1(\delta_{a,b}) = \frac{1}{2(b-a)} \int_0^\infty \frac{\sinh(v)}{[\cosh(v) + (b-a)^{-1} av \sinh(v)]^{3/2}} dv,$$

$$I_2(\delta_{a,b}) = \frac{1}{2(b-a)} \int_0^\infty \frac{v}{[\cosh(v) + (b-a)^{-1} av \sinh(v)]^{1/2}} dv.$$

Part (a) of Theorem 2 derives the joint MGF for the two functionals $N = \int W dW$ and $D = \int W^2$ on the interval $[a, b]$ and has potential applications in a wide range of sub-sampling problems with unit root processes. The individual MGFs for N and D, denoted $M_N(\theta_1) = M(\theta_1, 0)$ and $M_D(\theta_2) = M(0, \theta_2)$ respectively, follow straightforwardly and are given by

$$M_N(\theta_1) = \exp\left(-\frac{\theta_1}{2}(b-a)\right) \left[1 - (b-a)(\theta_1 + a\theta_1^2)\right]^{-1/2}, \tag{10}$$

$$M_D(\theta_2) = \left[\cosh\left((b-a)\lambda\right) + a\lambda \sinh\left((b-a)\lambda\right)\right]^{-1/2}, \tag{11}$$

respectively. Some special cases then result:

Example 1. When $[a, b] = [0, 1]$ we obtain

$$M_N(\theta_1) = e^{-\theta_1/2}(1-\theta_1)^{-1/2}, \quad M_D(\theta_2) = (\cosh(\lambda))^{-1/2},$$

while the joint MGF is

$$M(\theta_1, \theta_2) = \exp\left(-\frac{\theta_1}{2}\right) \left[\cosh(\lambda) - \frac{\theta_1}{\lambda} \sinh(\lambda)\right]^{-1/2},$$

a result that goes back to White (1958).

Example 2. The case of relevance for the non-overlapping jackknife sub-sampling is when $[a, b] = [(j-1)/m, j/m]$ and it follows that

$$M(\theta_1, \theta_2) = \exp\left(-\frac{\theta_1}{2m}\right) \left[\cosh\left(\frac{\lambda}{m}\right) - \frac{1}{\lambda}\left(\theta_1 + \frac{(j-1)}{m}(\theta_1^2 + 2\theta_2)\right) \sinh\left(\frac{\lambda}{m}\right)\right]^{-1/2},$$

$$M_N(\theta_1) = \exp\left(-\frac{\theta_1}{2m}\right) \left[1 - \frac{1}{m}\left(\theta_1 + \frac{(j-1)}{m}\theta_1^2\right)\right]^{-1/2},$$

$$M_D(\theta_2) = \left[\cosh\left(\frac{\lambda}{m}\right) + \frac{(j-1)\lambda}{m} \sinh\left(\frac{\lambda}{m}\right)\right]^{-1/2}.$$

Another potential use of the joint MGF in part (a) of Theorem 2 is in the computation of the cumulative and probability density functions of the distributions $m^{-1}Z(W, \delta_j)$. The

latter function is given by (with $i^2 = -1$)

$$pdf(z) = \frac{1}{2\pi i} \lim_{\epsilon_1 \to 0, \epsilon_2 \to \infty} \int_{\epsilon_1 < |\theta_1| < \epsilon_2} \left(\frac{\partial M(i\theta_1, i\theta_2)}{\partial \theta_2} \right)_{\theta_2 = -\theta_1 z} d\theta_1;$$

see, for example, Perron (1991, p.221) who performs this calculation for the distribution $Z(W, \delta)$, while Abadir (1993) derives a representation for the density function of $Z(W, \delta)$ in terms of a parabolic cylinder function. Of relevance later is the observation that, when $j = 1$, the MGF for N is the same as the MGF on $[0, 1]$ evaluated at θ_1/m, while that for D is the same as the full sample MGF evaluated at θ_2/m^2, implying that

$$\int_0^{1/m} W dW \overset{d}{=} \frac{1}{m} \int_0^1 W dW, \quad \int_0^{1/m} W^2 \overset{d}{=} \frac{1}{m^2} \int_0^1 W^2.$$

Furthermore, this implies that the limiting distribution of the first sub-sample estimator, $\ell(\hat{\rho}_1 - 1)$, is the same as that of the full-sample estimator, $n(\hat{\rho} - 1)$.

The result in part (b) of Theorem 2 is obtained by differentiating the MGF and constructing the appropriate integrals. Note that the usual (full-sample) result, where $a = 0$ and $b = 1$, is obtained as a special case:

$$I_1(\delta_{0,1}) = \frac{1}{2} \int_0^\infty \frac{\sinh(v)}{\cosh(v)^{3/2}} dv, \quad I_2(\delta_{0,1}) = \frac{1}{2} \int_0^\infty \frac{v}{\cosh(v)^{1/2}} dv;$$

see, for example, Gonzalo and Pitarakis (1998, Lemma 3.1). In the present situation of non-overlapping sub-samples, $a = (j-1)/m$ and $b = j/m$, resulting in

$$I_1(\delta_j) = \frac{m}{2} \int_0^\infty \frac{\sinh(v)}{[\cosh(v) + (j-1)v \sinh(v)]^{3/2}} dv,$$

$$I_2(\delta_j) = \frac{m}{2} \int_0^\infty \frac{v}{[\cosh(v) + (j-1)v \sinh(v)]^{1/2}} dv,$$

both of which depend on m. However, the limiting distribution of $\ell(\hat{\rho}_j - 1)$ is $N/(mD)$, and hence the expectation of this distribution does not depend on m. Table 1 contains the values of the normalised integrals $m^{-1} I_1(\delta_j)$ and $m^{-1} I_2(\delta_j)$ for values of $j = 1, \ldots, 12$, as well as the resulting expectations

$$\mu_j = E\left(m^{-1} Z(W, \delta_j)\right). \tag{12}$$

Numerical integration routines in Maple, Stata and Gauss were used to evaluate the integrals and produced the same results, at least to the degree of accuracy reported. For the reasons outlined above the expectation over $[0, 1/m]$ is the same as over $[0, 1]$, while the expectation increases monotonically in j. A simple explanation for the different properties of the sub-samples beyond $j = 1$ is that the initial values are of the same order of magnitude as the partial sums of the innovations, a topic to which we shall return later.

Table 1. Values of integrals and expectations for sub-samples

j	$m^{-1}I_1(\delta_j)$	$m^{-1}I_2(\delta_j)$	μ_j
1	1.000000	2.781430	−1.781430
2	0.267423	1.405632	−1.138209
3	0.163216	1.095145	−0.931929
4	0.118673	0.933003	−0.814330
5	0.093636	0.828454	−0.734818
6	0.077502	0.753586	−0.676084
7	0.066204	0.696450	−0.630246
8	0.057835	0.650934	−0.593099
9	0.051378	0.613532	−0.562154
10	0.046240	0.582067	−0.535827
11	0.042052	0.555105	−0.513053
12	0.038571	0.531656	−0.493085

When the process (1) has a unit root we can expect the expansions for $E(\hat{\rho}_j - 1)$ to be of the form

$$E\left(\hat{\rho}_j - 1\right) = \frac{\mu_j}{l} + o(\ell^{-1}), \quad j = 1, \ldots, m.$$

This information can be used to define the correct weights for the jackknife estimator that eliminate fully the first-order bias in $\hat{\rho}$. The optimal jackknife estimator is defined below.

Theorem 3. Let $\mu = E\left(Z(W, \delta)\right)$ and $\bar{\mu} = \mu - \sum_{j=1}^{m} \mu_j$, where the μ_j are defined in (12). Then, under Assumption 1, the optimal jackknife estimator is given by

$$\hat{\rho}_J^* = w_{1m}^* \hat{\rho} + w_{2m}^* \frac{1}{m} \sum_{j=1}^{m} \hat{\rho}_j,$$

where $w_{1m}^* = -\sum_{j=1}^{m} \mu_j / \bar{\mu}$ and $w_{2m}^* = \mu / \bar{\mu}$.

Theorem 3 shows the optimal weights for the jackknife estimator when the process (1) has a unit root. The values of μ_j in Table 1 can be utilised in Theorem 3 to derive the optimal weights for the jackknife estimator; these are reported in Table 2 for a range of values of m. It can be seen from Table 2 that the optimal weights are larger in (absolute) value than the standard weights that would apply if all the sub-sample distributions were the same.

Table 2. Values of standard and optimal jackknife weights

m	w_{1m}	w_{2m}	w^*_{1m}	w^*_{2m}
2	2.0000	−1.0000	2.5651	−1.5651
3	1.5000	−0.5000	1.8605	−0.8605
4	1.3333	−0.3333	1.6176	−0.6176
6	1.2000	−0.2000	1.4147	−0.4147
8	1.1429	−0.1429	1.3228	−0.3228
12	1.0909	−0.0909	1.2337	−0.2337

The effect of the variations in weights reported in Table 2 on the finite sample properties of the jackknife estimator has been explored in simulations, and the results are presented in Tables 3 and 4. The entries in Table 3 report the bias (and RMSE in parentheses) of $\hat{\rho}$, $\hat{\rho}_J$ and $\hat{\rho}^*_J$ obtained from 100,000 replications of the model with $p = 1$ (which corresponds to the Gaussian random walk process). Results are presented for the values of m that minimise the jackknife bias, denoted $\hat{\rho}_{J,B}$ and $\hat{\rho}^*_{J,B}$, as well as for the values of m that minimise the RMSE, denoted $\hat{\rho}_{J,R}$ and $\hat{\rho}^*_{J,R}$. These bias- and RMSE-minimising values of m are reported as superscripts in the table. Results are given in Table 4 for three examples when $p = 2$, these corresponding to values of the non-unit root of 1.25 ($\phi_1 = 1.8$, $\phi_2 = -0.8$), 2.5 ($\phi_1 = 1.4$, $\phi_2 = -0.4$), and 5 ($\phi_1 = 1.2$, $\phi_2 = -0.2$).

Table 3. Bias (RMSE) of OLS and jackknife estimators: $p = 1$

n :	24	48	96	192
$\hat{\rho}$	-0.0664	-0.0350	-0.0180	-0.0091
	(0.1368)	(0.0717)	(0.0370)	(0.0188)
$\hat{\rho}_{J,B}$	-0.0340^2	-0.0155^2	-0.0073^2	-0.0035^2
	(0.1486)	(0.0766)	(0.0394)	(0.0201)
$\hat{\rho}^*_{J,B}$	-0.0157^2	-0.0044^2	-0.0012^2	-0.0003^2
	(0.1760)	(0.0917)	(0.0475)	(0.0244)
$\hat{\rho}_{J,R}$	-0.0447^4	-0.0231^6	-0.0116^8	-0.0055^8
	(0.1313)	(0.0657)	(0.0333)	(0.0168)
$\hat{\rho}^*_{J,R}$	-0.0353^6	-0.0126^8	-0.0049^{12}	-0.0013^{12}
	(0.1352)	(0.0638)	(0.0312)	(0.0155)

NOTE: Superscripts denote the value of m.

Table 4. Bias (RMSE) of OLS and jackknife estimators: $p = 2$

n :	24	48	96	192
$\phi_1 = 1.8$, $\phi_2 = -0.8$				
$\hat{\rho}$	-0.0140	-0.0073	-0.0037	-0.0019
	(0.0460)	(0.0193)	(0.0087)	(0.0041)
$\hat{\rho}_{J,B}$	-0.0067^2	-0.0032^2	-0.0015^2	-0.0007^2
	(0.0684)	(0.0259)	(0.0107)	(0.0047)
$\hat{\rho}^*_{J,B}$	-0.0025^2	-0.0008^2	-0.0002^2	-0.0000^2
	(0.0927)	(0.0345)	(0.0138)	(0.0060)
$\hat{\rho}_{J,R}$	-0.0108^6	-0.0054^8	-0.0027^{12}	-0.0012^{12}
	(0.0511)	(0.0189)	(0.0081)	(0.0037)
$\hat{\rho}^*_{J,R}$	-0.0056^4	-0.0030^8	-0.0012^{12}	-0.0002^{12}
	(0.0701)	(0.0234)	(0.0090)	(0.0038)
$\phi_1 = 1.4$, $\phi_2 = -0.4$				
$\hat{\rho}$	-0.0393	-0.0211	-0.0108	-0.0055
	(0.0906)	(0.0457)	(0.0229)	(0.0114)
$\hat{\rho}_{J,B}$	-0.0191^2	-0.0094^2	-0.0044^2	-0.0021^2
	(0.1080)	(0.0513)	(0.0251)	(0.0124)
$\hat{\rho}^*_{J,B}$	-0.0077^2	-0.0027^2	-0.0007^2	-0.0001^2
	(0.1357)	(0.0634)	(0.0309)	(0.0153)
$\hat{\rho}_{J,R}$	-0.0287^6	-0.0149^6	-0.0069^8	-0.0033^8
	(0.0890)	(0.0420)	(0.0206)	(0.0101)
$\hat{\rho}^*_{J,R}$	-0.0172^6	-0.0070^8	-0.0027^{12}	-0.0007^{12}
	(0.1032)	(0.0431)	(0.0198)	(0.0096)
$\phi_1 = 1.2$, $\phi_2 = -0.2$				
$\hat{\rho}$	-0.0519	-0.0280	-0.0144	-0.0073
	(0.1138)	(0.0590)	(0.0301)	(0.0151)
$\hat{\rho}_{J,B}$	-0.0256^2	-0.0125^2	-0.0058^2	-0.0028^2
	(0.1301)	(0.0646)	(0.0325)	(0.0163)
$\hat{\rho}^*_{J,B}$	-0.0107^2	-0.0037^2	-0.0010^2	-0.0002^2
	(0.1599)	(0.0786)	(0.0395)	(0.0199)
$\hat{\rho}_{J,R}$	-0.0382^6	-0.0198^8	-0.0092^8	-0.0044^8
	(0.1100)	(0.0540)	(0.0270)	(0.0134)
$\hat{\rho}^*_{J,R}$	-0.0236^6	-0.0094^8	-0.0037^{12}	-0.0010^{12}
	(0.1224)	(0.0538)	(0.0256)	(0.0126)

NOTE: Superscripts denote the value of m.

In terms of bias it can be seen from Table 3 that the jackknife estimator $\hat{\rho}_{J,B}$ is capable of producing substantial bias reduction over $\hat{\rho}$, ranging from 49% at $n = 24$ through to 62% at $n = 192$; the bias-minimising values of m are equal to 2 for all four sample sizes. The bias reduction is still significant when the RMSE-minimising values of m are used, ranging from 33% at $n = 24$ to 40% at $n = 192$. However, the standard formulation does not take into account the differing means of the limiting sub-sample distributions, and it can be seen that the jackknife estimator with the optimal weights, $\hat{\rho}^*_J$, produces even more spectacular bias reductions, ranging from 76% at $n = 24$ to 97% at $n = 192$ for $\hat{\rho}^*_{J,B}$, and from 47% to 86% for $\hat{\rho}^*_{J,R}$. The effects of jackknifing on the RMSE are also interesting. When the pursuit of bias reduction is the objective it can be seen that $\hat{\rho}_{J,B}$ and $\hat{\rho}^*_{J,B}$ bear the cost of bias reduction in terms of larger variance and hence higher RMSE as compared to $\hat{\rho}$, the RMSE being almost 30% higher for the optimal estimator. But the results also show that $\hat{\rho}_{J,R}$ and $\hat{\rho}^*_{J,R}$ not only reduce bias but also reduce the overall RMSE compared to the full-sample estimator $\hat{\rho}$. These RMSE-minimising values of m are larger than the bias-minimising values and are seen to increase with n.

The results in Table 4, for the AR(2) model, are broadly in line with those in Table 3. In particular the jackknife estimators $\hat{\rho}_{J,B}$ and $\hat{\rho}^*_{J,B}$ are capable of substantial bias reduction at all sample sizes, although this comes at the expense of an increase in variance. The bias-minimising value of m is always equal to 2. However, choosing a larger value of m enables bias reduction to be combined with an overall reduction in RMSE, as can be seen by comparing $\hat{\rho}_{J,R}$ anf $\hat{\rho}^*_{J,R}$ with $\hat{\rho}$.

3. REGRESSION WITH AN INTERCEPT

The analysis of the previous section demonstrated that the distributions of the sub-sample estimators (used to construct the jackknife estimator) differ across sub-samples but can be used to define an optimal form of jackknife estimator under a unit root. Such an approach, however, requires knowledge of the unit root but can be useful when such information is available for the bias-reduced estimation of the remaining parameters in the AR(p) model. An alternative approach that does not require *a priori* knowledge of the unit root is examined below.

The source of the failure of the jackknife in the unit root setting is that the initial (or pre-sample) value in the sub-samples is the accumulated sum of all previous innovations and, being integrated, is therefore not eliminated in the asymptotics. To see this note that, under a unit root, the process Δy_t from (2) satisfies $\Delta y_t = u_t$, where $u_t = \zeta(L)^{-1}\epsilon_t$ is a stationary linear process, and hence the observations in sub-sample j satisfy

$$y_t = y_{t-1} + u_t = y_{(j-1)\ell} + \sum_{i=(j-1)\ell+1}^{t} u_i, \quad t = (j-1)\ell + 1, \ldots, j\ell; \tag{13}$$

the pre-sub-sample value, $y_{(j-1)\ell}$, is $O_p(\ell^{1/2})$ rather than $O_p(1)$ or a constant. The effect of the pre-sub-sample value on the asymptotics can be eliminated by incorporating an intercept in the regression, leading to

$$y_t = \tilde{\alpha} + \tilde{\rho} y_{t-1} + \sum_{j=1}^{p-1} \tilde{\zeta}_j \Delta y_{t-j} + \tilde{e}_t, \quad t = 1, \ldots, n, \tag{14}$$

where $\tilde{e}_t$ denotes the regression residual.

In the above framework the OLS estimator $\tilde{\rho}$ satisfies

$$n(\tilde{\rho} - 1) \Rightarrow \zeta(1) Z(W_0, \delta) \ \text{ as } \ n \to \infty, \tag{15}$$

where $W_0(r) = W(r) - \int_0^1 W(s)ds$ is a demeaned Wiener process. The standard jackknife estimator, based on (7), is given by

$$\tilde{\rho}_J = w_{1m} \tilde{\rho} + w_{2m} \frac{1}{m} \sum_{j=1}^{m} \tilde{\rho}_j, \tag{16}$$

where w_{1m} and w_{2m} are defined following (7) and the $\tilde{\rho}_j$ $(j = 1, \ldots, m)$ are the sub-sample estimators. Theorem 4 provides the limiting properties of $\ell(\tilde{\rho}_j - 1)$ and, hence, of $\tilde{\rho}_J$, which rely on the sub-sample demeaned Wiener processes

$$W_{j,m}(r) = W(r) - m \int_{(j-1)/m}^{j/m} W(s)ds, \quad j = 1, \ldots, m.$$

Theorem 4. Under Assumption 1, if $\ell \to \infty$ as $n \to \infty$:

(a) If m is fixed, $\ell(\tilde{\rho}_j - 1) \Rightarrow m^{-1} \zeta(1) Z(W_{j,m}, \delta_j)$ $(j = 1, \ldots, m)$ and

$$n(\tilde{\rho}_J - 1) \Rightarrow w_{1m} \zeta(1) Z(W_0, \delta) + w_{2m} \zeta(1) \sum_{j=1}^{m} m^{-1} Z(W_{j,m}, \delta_j);$$

(b) If $m^{-1} + mn^{-1} \to 0$, $n(\tilde{\rho}_J - 1) \Rightarrow \zeta(1) Z(W_0, \delta)$.

The limiting distributions of the sub-sample estimators in part (a) of Theorem 4 are expressed in terms of the demeaned Wiener processes $W_{j,m}$. Note that the usual demeaned process on $[0, 1]$, denoted W_0 following (15), is given by $W_{1,1}$ in this notation. The fact that

regression with an intercept eliminates the effects of the pre-sample value implies that

$$m^{-1}Z(W_{j,m},\delta_j)=\frac{\displaystyle\int_{(j-1)/m}^{j/m}W_{j,m}dW_{j,m}}{\displaystyle m\int_{(j-1)/m}^{j/m}W_{j,m}^2}\overset{d}{=}\frac{\displaystyle\int_0^1 W_0dW_0}{\displaystyle\int_0^1 W_0^2}=Z(W_0,\delta) \tag{17}$$

for all $j=1,\ldots,m$. Although regression with an intercept eliminates the effects of pre-sub-sample values, the effect on the limiting distributions is to actually increase the negative bias. In fact, $E[Z(W_0,\delta)]=-5.379$; see, for example, Table 7.2 of Tanaka (1996).

Tables 5 and 6 report the bias and RMSE of the estimators $\tilde{\rho}$ and $\tilde{\rho}_J$ obtained from 100,000 replications of the AR model with $p=1$ and $p=2$ respectively. Also included is the estimator $\tilde{\rho}_J^*$ which is based on regression with an intercept but uses the optimal weights employed by the estimator $\hat{\rho}_J^*$ in the regression without an intercept; it is defined by $\tilde{\rho}_J^*=w_{1m}^*\tilde{\rho}+(w_{2m}^*/m)\sum_{j=1}^m\tilde{\rho}_j$, where w_{1m}^* and w_{2m}^* are defined in Theorem 3. This enables the assessment of the effects of using the optimal weights in an inappropriate setting i.e. when the standard weights are, in fact, optimal.

Table 5. Bias (RMSE) of OLS and jackknife estimators in regression with intercept: $p=1$

n :	24	48	96	192
$\tilde{\rho}$	−0.1985	−0.1052	−0.0545	−0.0276
	(0.2524)	(0.1350)	(0.0706)	(0.0360)
$\tilde{\rho}_{J,B}$	-0.0399^2	-0.0116^2	-0.0035^2	-0.0008^2
	(0.2444)	(0.1316)	(0.0695)	(0.0360)
$\tilde{\rho}_{J,B}^*$	0.0497^2	0.0413^2	0.0253^2	0.0143^2
	(0.3166)	(0.1766)	(0.0949)	(0.0498)
$\tilde{\rho}_{J,R}$	-0.0673^4	-0.0356^8	-0.0152^{12}	-0.0044^{12}
	(0.2013)	(0.0992)	(0.0499)	(0.0248)
$\tilde{\rho}_{J,R}^*$	0.0317^6	0.0465^{12}	0.0466^{12}	0.0247^6
	(0.2288)	(0.1171)	(0.0720)	(0.0406)

NOTE: Superscripts denote the value of m.

Table 6. Bias (RMSE) of OLS and jackknife estimators in regression with intercept: $p = 2$

n :	24	48	96	192
$\phi_1 = 1.8$, $\phi_2 = -0.8$				
$\tilde{\rho}$	-0.0587	-0.0269	-0.0124	-0.0059
	(0.0893)	(0.0379)	(0.0169)	(0.0079)
$\tilde{\rho}_{J,B}$	0.0028^{6}	0.0041^{12}	0.0019^{2}	0.0006^{2}
	(0.1499)	(0.0550)	(0.0202)	(0.0088)
$\tilde{\rho}^{*}_{J,B}$	0.0514^{2}	0.0238^{2}	0.0100^{2}	0.0043^{2}
	(0.1898)	(0.0735)	(0.0296)	(0.0127)
$\tilde{\rho}_{J,R}$	0.0085^{4}	0.0063^{8}	0.0043^{12}	-0.0021^{12}
	(0.1031)	(0.0358)	(0.0144)	(0.0063)
$\tilde{\rho}^{*}_{J,R}$	0.0600^{3}	0.0308^{3}	0.0134^{3}	0.0056^{3}
	(0.1716)	(0.0676)	(0.0275)	(0.0116)
$\phi_1 = 1.4$, $\phi_2 = -0.4$				
$\tilde{\rho}$	-0.1362	-0.0684	-0.0337	-0.0169
	(0.1815)	(0.0900)	(0.0443)	(0.0221)
$\tilde{\rho}_{J,B}$	-0.0019^{2}	0.0001^{2}	0.0000^{8}	0.0001^{3}
	(0.2055)	(0.0958)	(0.0322)	(0.0190)
$\tilde{\rho}^{*}_{J,B}$	0.0739^{2}	0.0388^{2}	0.0197^{2}	0.0097^{2}
	(0.2893)	(0.1349)	(0.0650)	(0.0319)
$\tilde{\rho}_{J,R}$	-0.0101^{4}	-0.0047^{8}	-0.0007^{12}	0.0002^{12}
	(0.1620)	(0.0675)	(0.0312)	(0.0153)
$\tilde{\rho}^{*}_{J,R}$	0.0973^{4}	0.0565^{4}	0.0296^{4}	0.0147^{4}
	(0.2467)	(0.1139)	(0.0558)	(0.0273)
$\phi_1 = 1.2$, $\phi_2 = -0.2$				
$\tilde{\rho}$	-0.1740	-0.0890	-0.0444	-0.0224
	(0.2278)	(0.1161)	(0.0581)	(0.0292)
$\tilde{\rho}_{J,B}$	-0.0119^{2}	-0.0033^{2}	-0.0005^{2}	-0.0002^{2}
	(0.2448)	(0.1195)	(0.0594)	(0.0297)
$\tilde{\rho}^{*}_{J,B}$	0.0797^{2}	0.0451^{2}	0.0243^{2}	0.0124^{2}
	(0.3379)	(0.1657)	(0.0823)	(0.0415)
$\tilde{\rho}_{J,R}$	-0.0245^{4}	-0.0128^{8}	-0.0044^{12}	-0.0010^{12}
	(0.1938)	(0.0851)	(0.0405)	(0.0201)
$\tilde{\rho}^{*}_{J,R}$	0.1029^{4}	0.0753^{6}	0.0356^{4}	0.0185^{4}
	(0.2800)	(0.1345)	(0.0696)	(0.0351)

NOTE: Superscripts denote the value of m.

From Table 5 it can be seen that, not surprisingly, the estimator $\tilde{\rho}$ is more biased than $\hat{\rho}$, its theoretical first-order bias being $-5.379/n$ as opposed to $-1.781/n$. Compared to $\tilde{\rho}$ the estimator $\tilde{\rho}_{J,B}$ manages to reduce the bias by 80% at $n = 24$ rising to 97% at $n = 192$. The estimator $\tilde{\rho}^*_{J,B}$ also reduces bias (in absolute terms) but by not as much as $\tilde{\rho}_{J,B}$. It is also interesting to note that the RMSE of $\tilde{\rho}_{J,B}$ is less than that of $\tilde{\rho}$ at all sample sizes. The estimators $\tilde{\rho}_{J,R}$ and $\tilde{\rho}^*_{J,R}$ also show substantial bias reduction as compared to $\tilde{\rho}$ while $\tilde{\rho}_{J,R}$ shows a considerable overall reduction in RMSE. Overall the performance of the estimators $\tilde{\rho}^*_{J,B}$ and $\tilde{\rho}^*_{J,R}$ is, not surprisingly, inferior to that of $\tilde{\rho}_{J,B}$ and $\tilde{\rho}_{J,R}$ in the regression with an intercept, further supporting the fact that including the intercept removes the effects of pre-sub-sample values in the sub-sample regressions and, hence, the standard jackknife weights are the optimal weights in this case.

The results in Table 6, for the AR(2) model, are broadly in line with those in Table 5. In particular the standard jackknife weights are confirmed as being optimal in the regressions with an intercept, the estimators $\tilde{\rho}_{J,B}$ and $\tilde{\rho}_{J,R}$ being superior to $\tilde{\rho}^*_{J,B}$ and $\tilde{\rho}^*_{J,R}$, obtaining substantial bias reduction as well as smaller RMSEs compared to $\tilde{\rho}$.

4. CONCLUDING COMMENTS

This paper has been concerned with the application of jackknife methods as a means of bias reduction in the estimation of AR(p) models with a unit root. It has been shown that the usual jackknife estimator based on non-overlapping sub-samples, of the type used to great effect by Phillips and Yu (2005), does not remove fully the first-order bias as intended, but that an 'optimal' jackknife estimator can be defined that is capable of removing the first-order bias. The results are based on a demonstration that the sub-sample estimators converge to different limiting distributions, and the joint MGF of the numerator and denominator of these distributions (which are functionals of a Wiener process over a sub-interval of [0,1]) is derived and utilised to extract the optimal weights. Simulations for $p = 1$ and $p = 2$ demonstrate the ability of the jackknife estimator to produce substantial bias reductions in the parameter of interest. It is also shown that incorporating an intercept in the regressions allows the standard jackknife estimator to be used and it is able also to produce substantial bias reduction despite the fact that the distributions of the full-sample and sub-sample estimators have greater bias in this case. Of interest, too, is the fact that the jackknife estimators can also reduce the overall RMSE compared to the OLS estimator, this requiring a larger (though still small) number of sub-samples compared to the value that produces maximum bias reduction (which is typically equal to two).

The results in this paper can be useful in further research. The joint MGF of the numerator and denominator of the functionals of a Wiener process over sub-intervals of [0,1] is presented in sufficient generality that it may have applications in other problems using sub-sampling in a unit root setting. Using jackknife estimators as a basis for actually testing

for a unit root is another interesting avenue for future research and one that is being pursued by the authors.

APPENDIX: PROOFS

Equation (4). In order to verify (4) it is convenient to first stack the observations in the form

$$\mathbf{y} = \rho \mathbf{y}_L + \mathbf{X}\boldsymbol{\zeta} + \boldsymbol{\epsilon},$$

where $\mathbf{y} = (y_1, \ldots, y_n)'$, $\mathbf{y}_L = (y_0, \ldots, y_{n-1})'$, $\mathbf{X}$ is the $n \times (p-1)$ matrix with typical row $[\Delta y_{t-1}, \ldots, \Delta y_{t-(p-1)}]$ and $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_{p-1})'$. Partitioned regression formulae result in

$$\hat{\rho} - \rho = \frac{\mathbf{y}_L' \mathbf{M}_{\mathbf{X}} \boldsymbol{\epsilon}}{\mathbf{y}_L' \mathbf{M}_{\mathbf{X}} \mathbf{y}_L}$$

where $\mathbf{M}_{\mathbf{X}} = \mathbf{I}_{\mathbf{n}} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'$. The numerator can be written

$$\begin{aligned} n^{-1}\mathbf{y}_L' \mathbf{M}_{\mathbf{X}} \boldsymbol{\epsilon} &= n^{-1}\mathbf{y}_L'\boldsymbol{\epsilon} - n^{-3/2}\mathbf{y}_L'\mathbf{X}\left(n^{-1}\mathbf{X}'\mathbf{X}\right)^{-1} n^{-1/2}\mathbf{X}'\boldsymbol{\epsilon} \\ &= n^{-1}\mathbf{y}_L'\boldsymbol{\epsilon} + o_p(1) \Rightarrow \frac{\sigma_\epsilon^2}{\zeta(1)} \int_0^1 W dW, \end{aligned}$$

while the denominator is

$$\begin{aligned} n^{-2}\mathbf{y}_L' \mathbf{M}_{\mathbf{X}} \mathbf{y}_L &= n^{-2}\mathbf{y}_L'\mathbf{y}_L - n^{-3/2}\mathbf{y}_L'\mathbf{X}\left(n^{-1}\mathbf{X}'\mathbf{X}\right)^{-1} n^{-3/2}\mathbf{X}'\mathbf{y}_L \\ &= n^{-2}\mathbf{y}_L'\mathbf{y}_L + o_p(1) \Rightarrow \frac{\sigma_\epsilon^2}{\zeta(1)^2} \int_0^1 W^2, \end{aligned}$$

thereby justifying (4).

The following Lemma is used in the proof of Theorem 1.

Lemma A1. Under Assumption 1, if $\ell \to \infty$ as $n \to \infty$:

(a) $\ell^{-3/2} \sum_{t \in \tau_j} y_{t-1} \Rightarrow \sigma_\epsilon \psi(1) m^{3/2} \int_{(j-1)/m}^{j/m} W;$

(b) $\ell^{-2} \sum_{t \in \tau_j} y_{t-1}^2 \Rightarrow \sigma_\epsilon^2 \psi(1)^2 m^2 \int_{(j-1)/m}^{j/m} W^2;$

(c) $\ell^{-1} \sum_{t \in \tau_j} y_{t-1}\epsilon_t \Rightarrow \sigma_\epsilon^2 \psi(1) m \int_{(j-1)/m}^{j/m} W dW,$

where $\psi(1) = \zeta(1)^{-1}$.

Proof. Under a unit root y_t has the representation $\zeta(L)\Delta y_t = \epsilon_t$ which, given that the roots of the autoregressive polynomial lie outside the unit circle, can be inverted to yield $\Delta y_t = u_t$ where $u_t = \psi(L)\epsilon_t$ and $\psi(z) = \zeta(z)^{-1} = \sum_{j=0}^{\infty}\psi_j$. Applying the Beveridge-Nelson (BN) decomposition yields

$$y_t = \psi(1)S_t + \eta_t - \eta_0 + y_0, \quad t = 1, \ldots, n,$$

where $S_t = \sum_{j=1}^{t}\epsilon_j$, $\eta_t = \sum_{j=0}^{\infty}\alpha_i\epsilon_{t-j}$ and $\alpha_j = -\sum_{k=j+1}^{\infty}\psi_k$. In what follows we shall use the property that $n^{-1/2}S_{[nr]} \Rightarrow \sigma_\epsilon W(r)$ as $n \to \infty$. It is possible to write

$$S_{t-1} = S_{t-1}n\int_{(t-1)/n}^{t/n} dr = n\int_{(t-1)/n}^{t/n} S_{[nr]}dr$$

so that

$$\sum_{t\in\tau_j} S_{t-1} = n\sum_{t=(j-1)\ell+1}^{j\ell}\int_{(t-1)/n}^{t/n} S_{[nr]}dr = n\int_{(j-1)\ell/n}^{j\ell/n} S_{[nr]}dr = n\int_{(j-1)/m}^{j/m} S_{[nr]}dr,$$

in view of the fact that $(j-1)\ell/n = (j-1)/m$ and $j\ell/n = j/m$ in the limits of the integral. Similarly,

$$\sum_{t\in\tau_j} S_{t-1}^2 = n\int_{(j-1)/m}^{j/m} S_{[nr]}^2 dr.$$

It follows that, as $n \to \infty$,

$$\begin{aligned}
\ell^{-3/2}\sum_{t\in\tau_j} S_{t-1} &= m^{3/2}\int_{(j-1)/m}^{j/m} n^{-1/2}S_{[nr]}dr \Rightarrow \sigma_\epsilon m^{3/2}\int_{(j-1)/m}^{j/m} W, \\
\ell^{-2}\sum_{t\in\tau_j} S_{t-1}^2 &= m^2\int_{(j-1)/m}^{j/m}\left(n^{-1/2}S_{[nr]}\right)^2 dr \Rightarrow \sigma_\epsilon^2 m^2\int_{(j-1)/m}^{j/m} W^2, \\
\ell^{-1}\sum_{t\in\tau_j} S_{t-1}\epsilon_t &= \frac{1}{2}\left[\left(\ell^{-1/2}S_{j\ell}\right)^2 - \left(\ell^{-1/2}S_{(j-1)\ell}\right)^2 - \ell^{-1}\sum_{t\in\tau_j}\epsilon_t^2\right] \\
&\Rightarrow \frac{\sigma_\epsilon^2 m}{2}\left[W\left(\frac{j}{m}\right) - W\left(\frac{j-1}{m}\right) - \frac{1}{m}\right] \\
&\overset{d}{=} \sigma_\epsilon^2 m\int_{(j-1)/m}^{j/m} W dW,
\end{aligned}$$

the latter equality in distribution holding in view of (9). These expressions are used repeatedly in what follows.

(a) Using the BN decomposition of y_t we obtain

$$\begin{aligned}\ell^{-3/2}\sum_{t\in\tau_j} y_{t-1} &= \psi(1)\ell^{-3/2}\sum_{t\in\tau_j} S_{t-1} + \ell^{-3/2}\sum_{t\in\tau_j}\eta_{t-1} + \ell^{-1/2}(y_0-\eta_0)\\ &= \psi(1)\ell^{-3/2}\sum_{t\in\tau_j} S_{t-1} + o_p(1) \Rightarrow \sigma_\epsilon\psi(1)m^{3/2}\int_{(j-1)/m}^{j/m} W.\end{aligned}$$

(b) Expanding y_{t-1}^2 using the BN decomposition it follows that

$$\begin{aligned}\ell^{-2}\sum_{t\in\tau_j} y_{t-1}^2 &= \psi(1)^2\ell^{-2}\sum_{t\in\tau_j} S_{t-1}^2 + 2\psi(1)\ell^{-1}\sum_{t\in\tau_j} S_{t-1}\eta_{t-1}\\ &\quad +2\psi(1)(y_0-\eta_0)\ell^{-2}\sum_{t\in\tau_j} S_{t-1} + \ell^{-2}\sum_{t\in\tau_j}\eta_{t-1}^2\\ &\quad +2(y_0-\eta_0)\ell^{-2}\sum_{t\in\tau_j}\eta_{t-1} + \ell^{-1}(y_0-\eta_0)^2\\ &= \psi(1)^2\ell^{-2}\sum_{t\in\tau_j} S_{t-1}^2 + o_p(1) \Rightarrow \sigma_\epsilon^2\psi(1)^2m^2\int_{(j-1)/m}^{j/m} W^2.\end{aligned}$$

(c) Again, using the BN decomposition, it is possible to write

$$\ell^{-1}\sum_{t\in\tau_j} y_{t-1}\epsilon_t = \psi(1)\ell^{-1}\sum_{t\in\tau_j} S_{t-1}\epsilon_t + \ell^{-1}\sum_{t\in\tau_j}(\eta_{t-1}-\eta_0)\epsilon_t + y_0\ell^{-1}\sum_{t\in\tau_j}\epsilon_t.$$

Now $(\eta_{t-1}-\eta_0)\epsilon_t$ is a martingale difference sequence and so the second term converges in probability to zero, as does the third. We therefore have

$$\begin{aligned}\ell^{-1}\sum_{t\in\tau_j} y_{t-1}\epsilon_t &= \psi(1)\ell^{-1}\sum_{t\in\tau_j} S_{t-1}\epsilon_t + o_p(1)\\ &\Rightarrow \sigma_\epsilon^2\psi(1)m\int_{(j-1)/m}^{j/m} W dW.\end{aligned}$$

Proof of Theorem 1. (a) The result follows from parts (b) and (c) of Lemma A1 by noting that

$$\ell(\hat\rho_j - 1) = \frac{\ell^{-1}\sum_{t\in\tau_j} y_{t-1}\epsilon_t}{\ell^{-2}\sum_{t\in\tau_j} y_{t-1}^2} + o_p(1). \tag{18}$$

The result for $\hat\rho_J$ follows from the appropriate linear combination of the limiting distributions of $n(\hat\rho-1)$ in (4) and of $\ell(\hat\rho_j-1)$, using the continuous mapping theorem.
(b) Let $Z = O_p(1)$ denote the limit of $n(\hat\rho-1)$ and let $Z_j = O_p(1/m)$ denote the limit of $\ell(\hat\rho_j-1)$; see the comments in Remark 1 following the Theorem. When $m\to\infty$ it follows

that $w_{1m} \to 1$ and $w_{2m} \sum_{j=1}^{m} Z_j = -\sum_{j=1}^{m} Z_j/(m-1) = O_p(1/m) = o_p(1)$, thereby yielding the stated result.

Proof of Theorem 2. (a) Consider the two Ornstein-Uhlenbeck (O-U) processes, $X(t)$ and $Y(t)$, on $t \in [0, b]$, given by

$$\begin{aligned} dX(t) &= \gamma X(t)dt + dW(t), \quad X(0) = 0, \\ dY(t) &= \lambda Y(t)dt + dW(t), \quad Y(0) = 0, \end{aligned}$$

and let μ_X and μ_Y be the measures induced by X and Y respectively. These measures are equivalent and, by Girsanov's Theorem (see, for example, Theorem 4.1 of Tanaka, 1996),

$$\frac{d\mu_X}{d\mu_Y}(s) = \exp\left((\gamma - \lambda)\int_0^b s(t)ds(t) - \frac{(\gamma^2 - \lambda^2)}{2}\int_0^b s(t)^2 dt\right)$$

is the Radon-Nikodym derivative evaluated at $s(t)$, a random process on $[0, b]$ with $s(0) = 0$. We are interested in the case where $\gamma = 0$, so that $X(t) = W(t)$, and the change of measure will be used because

$$E\left(f(X)\right) = E\left(f(Y)\frac{d\mu_X}{d\mu_Y}(Y)\right).$$

Under $\gamma = 0$ we obtain

$$\begin{aligned} M(\theta_1, \theta_2) &= E\exp\left(\theta_1\int_a^b W dW + \theta_2\int_a^b W^2\right) \\ &= E\exp\left(\theta_1\int_a^b Y dY + \theta_2\int_a^b Y^2 - \lambda\int_0^b Y dY + \frac{\lambda^2}{2}\int_0^b Y^2\right). \end{aligned}$$

Now, using the Ito calculus, $\int_a^b Y dY = (1/2)[Y(b)^2 - Y(a)^2 - (b-a)]$, and so

$$\theta_1\int_a^b Y dY - \lambda\int_0^b Y dY = \frac{(\theta_1 - \lambda)}{2}Y(b)^2 - \frac{\theta_1}{2}Y(a)^2 - \frac{(\theta_1 - \lambda)}{2}b + \frac{\theta_1}{2}a,$$

while splitting the second integral yields

$$\theta_2\int_a^b Y^2 + \frac{\lambda^2}{2}\int_0^b Y^2 = \left(\theta_2 + \frac{\lambda^2}{2}\right)\int_a^b Y^2 + \frac{\lambda^2}{2}\int_0^a Y^2.$$

Hence

$$\begin{aligned} M(\theta_1, \theta_2) &= \exp\left(\frac{\theta_1}{2}a - \frac{(\theta_1 - \lambda)}{2}b\right) E\exp\left\{\frac{(\theta_1 - \lambda)}{2}Y(b)^2 - \frac{\theta_1}{2}Y(a)^2\right. \\ &\quad \left. + \left(\theta_2 + \frac{\lambda^2}{2}\right)\int_a^b Y^2 + \frac{\lambda^2}{2}\int_0^a Y^2\right\}. \end{aligned}$$

As the parameter λ is arbitrary, it is convenient to set $\lambda = \sqrt{-2\theta_2}$ so as to eliminate the term $\int_a^b Y^2$. We shall then proceed in two steps:

(i) Take the expectation of $M(\theta_1, \theta_2)$ conditional on $\mathcal{F}_0^a$, the sigma field generated by W on $[0, a]$;

(ii) Introduce another O-U process V and apply Girsanov's Theorem again to take the expectation with respect to $\mathcal{F}_0^a$.

Step (i). Conditional on $\mathcal{F}_0^a$, let $M(\theta_1, \theta_2; \mathcal{F}_0^a) = EM(\theta_1, \theta_2)|\mathcal{F}_0^a$, so that

$$\begin{aligned} M(\theta_1, \theta_2; \mathcal{F}_0^a) &= \exp\left(\frac{\theta_1}{2}a - \frac{(\theta_1 - \lambda)}{2}b\right) \exp\left(-\frac{\theta_1}{2}Y(a)^2 + \frac{\lambda^2}{2}\int_0^a Y^2\right) \\ &\quad \times E \exp\left(\frac{(\theta_1 - \lambda)}{2}Y(b)^2\right). \end{aligned}$$

Define $\mu = \exp((b-a)\lambda)Y(a)$ and $\omega^2 = (\exp(2(b-a)\lambda) - 1)/2\lambda$ so that, conditional on $\mathcal{F}_0^a$, $Y(b) \sim \mathrm{N}(\mu, \omega^2)$. Hence

$$E \exp\left(\frac{(\theta_1 - \lambda)}{2}Y(b)^2\right) = \exp\left(\frac{(\theta_1 - \lambda)}{2}kY(a)^2\right)\left[1 - (\theta_1 - \lambda)\omega^2\right]^{-1/2},$$

where $k = \exp(2(b-a)\lambda)/[1 - (\theta_1 - \lambda)\omega^2]$, and so

$$\begin{aligned} M(\theta_1, \theta_2; \mathcal{F}_0^a) &= \exp\left(\frac{\theta_1}{2}a - \frac{(\theta_1 - \lambda)}{2}b\right)\left[1 - (\theta_1 - \lambda)\omega^2\right]^{-1/2} \\ &\quad \times \exp\left\{\left(\frac{(\theta_1 - \lambda)}{2}k - \frac{\theta_1}{2}\right)Y(a)^2 + \frac{\lambda^2}{2}\int_0^a Y^2\right\}. \end{aligned}$$

Step (ii). We now introduce a new auxiliary process, $V(t)$, on $[0, a]$, given by

$$dV(t) = \eta V(t)dt + dw(t), \quad V(0) = 0,$$

and will make use of the change of measure

$$\frac{d\mu_Y}{d\mu_V}(s) = \exp\left((\lambda - \eta)\int_0^a s(t)ds(t) - \frac{(\lambda^2 - \eta^2)}{2}\int_0^a s(t)^2 dt\right)$$

in order to eliminate $\int_0^a Y^2$. We have $M(\theta_1, \theta_2) = EM(\theta_1, \theta_2; \mathcal{F}_0^a)$ and so

$$\begin{aligned} M(\theta_1, \theta_2) &= \exp\left(\frac{\theta_1}{2}a - \frac{(\theta_1 - \lambda)}{2}b\right)\left[1 - (\theta_1 - \lambda)\omega^2\right]^{-1/2} \\ &\quad \times E \exp\left\{\left(\frac{(\theta_1 - \lambda)}{2}k - \frac{\theta_1}{2}\right)Y(a)^2 + \frac{\lambda^2}{2}\int_0^a Y^2\right\}. \end{aligned}$$

With the change of measure the expectation of interest becomes

$$E\exp\left\{\left(\frac{(\theta_1-\lambda)}{2}k-\frac{\theta_1}{2}\right)V(a)^2+(\lambda-\eta)\int_0^a VdV+\frac{\eta^2}{2}\int_0^a V^2\right\}.$$

But η is arbitrary and so we set $\eta=0$ to eliminate $\int_0^a V^2$ and obtain

$$\begin{aligned}
&E\exp\left\{\left(\frac{(\theta_1-\lambda)}{2}k-\frac{\theta_1}{2}\right)V(a)^2+\lambda\int_0^a VdV\right\}\\
&= E\exp\left\{\left(\frac{(\theta_1-\lambda)}{2}k-\frac{\theta_1}{2}\right)V(a)^2+\frac{\lambda}{2}\left(V(a)^2-a\right)\right\}\\
&= \exp\left(-\frac{\lambda}{2}a\right)E\exp\left(\frac{(\theta_1-\lambda)(k-1)}{2}V(a)^2\right).
\end{aligned}$$

Under $\eta=0$ it follows that $V(a)\sim \mathrm{N}(0,a)$ and so

$$E\exp\left(\frac{(\theta_1-\lambda)(k-1)}{2}V(a)^2\right)=[1-(\theta_1-\lambda)(k-1)a]^{-1/2}.$$

Hence $M(\theta_1,\theta_2)=\exp(-\theta_1(b-a)/2)H(\theta_1,\theta_2)^{-1/2}$ where

$$H(\theta_1,\theta_2)=\exp(-(b-a)\lambda)(1-\delta\omega^2)-\exp(-(b-a)\lambda)a\delta(k-1)(1-\delta\omega^2)$$

and $\delta=\theta_1-\lambda$ for notational convenience. Let $z=(b-a)\lambda$. The first term is

$$\begin{aligned}
e^{-z}-\delta e^{-z}\frac{(e^{2z}-1)}{2\lambda} &= e^{-z}-\left(\frac{\theta_1}{\lambda}-1\right)\frac{(e^z-e^{-z})}{2}\\
&= \frac{(e^z+e^{-z})}{2}-\frac{\theta_1}{\lambda}\frac{(e^z-e^{-z})}{2}=\cosh z-\frac{\theta_1}{\lambda}\sinh z.
\end{aligned}$$

The second term involves the expression $(k-1)(1-\delta\omega^2)=e^{2z}-1+\delta\omega^2$ and so we obtain

$$\begin{aligned}
e^{-z}(k-1)(1-\delta\omega^2) &= e^z-e^{-z}+\delta e^{-z}\frac{(e^{2z}-1)}{2\lambda}\\
&= e^z-e^{-z}+\left(\frac{\theta_1}{\lambda}-1\right)\frac{(e^z-e^{-z})}{2}\\
&= \left(1+\frac{\theta_1}{\lambda}\right)\frac{(e^z-e^{-z})}{2}=\left(1+\frac{\theta_1}{\lambda}\right)\sinh z.
\end{aligned}$$

Combining these components yields the required expression for $H(\theta_1,\theta_2)$.

(b) From the definition of $M(\theta_1, \theta_2)$ we obtain

$$\begin{aligned}\frac{\partial M(\theta_1,\theta_2)}{\partial \theta_1} &= -\frac{(b-a)}{2}\exp\left(-\frac{\theta_1}{2}(b-a)\right)H(\theta_1,\theta_2)^{-1/2} \\ &\quad -\frac{1}{2}\exp\left(-\frac{\theta_1}{2}(b-a)\right)H(\theta_1,\theta_2)^{-3/2}\frac{\partial H(\theta_1,\theta_2)}{\partial \theta_1}.\end{aligned}$$

Partial differentiation of $H(\theta_1, \theta_2)$ yields

$$\frac{\partial H(\theta_1,\theta_2)}{\partial \theta_1} = -\frac{1}{\lambda}(1+2a\theta_1)\sinh((b-a)\lambda)$$

from which it follows that

$$\left.\frac{\partial H(\theta_1,\theta_2)}{\partial \theta_1}\right|_{\theta_1=0} = -\frac{1}{\lambda}\sinh((b-a)\lambda).$$

Also $H(0, \theta_2) = \cosh((b-a)\lambda) + a\lambda\sinh((b-a)\lambda)$.

Let $x = \sqrt{2\theta_2}$. Then, combining the results above,

$$\begin{aligned}\left.\frac{\partial M(\theta_1,-\theta_2)}{\partial \theta_1}\right|_{\theta_1=0} &= -\frac{(b-a)}{2}\left[\cosh((b-a)x) + ax\sinh((b-a)x)\right]^{-1/2} \\ &\quad +\frac{1}{2x}\frac{\sinh((b-a)x)}{\left[\cosh((b-a)x) + ax\sinh((b-a)x)\right]^{3/2}}.\end{aligned}$$

Integrating with respect to θ_2, and making the substitution $v = (b-a)\sqrt{2\theta_2}$, yields the result in the Theorem.

Proof of Theorem 3. To determine the weights for $\hat{\rho}_J^*$, note that

$$E(\hat{\rho}) = 1 + \frac{\mu}{n} + o(n^{-1}), \quad E(\hat{\rho}_j) = 1 + \frac{\mu_j}{n} + o(n^{-1}), \quad j = 1, \ldots, m,$$

where μ is defined in the Theorem. From the definition of $\hat{\rho}_J^*$, taking expectations yields

$$E(\hat{\rho}_J^*) = (w_{1m}^* + w_{2m}^*) + \frac{1}{n}\left(w_{1m}^*\mu + w_{2m}^*\sum_{j=1}^{m}\mu_j\right) + o(n^{-1}).$$

In order that $E(\hat{\rho}_J^*) = 1 + o(n^{-1})$ the requirement is that: (i) $w_{1m}^* + w_{2m}^* = 1$, and (ii) $w_{1m}^*\mu + w_{2m}^*\sum_{j=1}^{m}\mu_j = 0$. Solving these two conditions simultaneously yields the stated weights.

Proof of Theorem 4. (a) It is convenient to stack the observation for sub-samples in the form

$$\mathbf{y}_j = \mathbf{W}_j\boldsymbol{\beta} + \mathbf{X}_j\boldsymbol{\zeta} + \boldsymbol{\epsilon}_j, \quad j = 1, \ldots, m,$$

where $\mathbf{y}_j = [y_{(j-1)\ell+1}, \ldots, y_{j\ell}]'$, $\mathbf{X}_j$ is $\ell \times (p-1)$ with typical row $[\Delta y_{t-1}, \ldots, \Delta y_{t-(p-1)}]$ $(t \in \tau_j)$, $\boldsymbol{\beta} = [0, 1]'$, $\boldsymbol{\zeta} = [\zeta_1, \ldots, \zeta_{p-1}]'$ and

$$\mathbf{W}_j = \begin{bmatrix} 1 & y_{(j-1)\ell} \\ \vdots & \vdots \\ 1 & y_{j\ell-1} \end{bmatrix}.$$

Application of partitioned regression formulae results in

$$\tilde{\boldsymbol{\beta}}_j - \boldsymbol{\beta} = (\mathbf{W}_j'\mathbf{M}_j\mathbf{W}_j)^{-1}\mathbf{W}_j'\mathbf{M}_j\boldsymbol{\epsilon}_j, \quad j = 1, \ldots, m,$$

where $\mathbf{W}_j = \mathbf{I}_\ell - \mathbf{X}_j(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'$ and $\tilde{\boldsymbol{\beta}}_j$ denotes the OLS estimator of $\boldsymbol{\beta}$ in sub-sample j. Let

$$\mathbf{D}_\ell = \begin{bmatrix} \ell^{1/2} & 0 \\ 0 & \ell \end{bmatrix}.$$

Then

$$\mathbf{D}_\ell(\tilde{\boldsymbol{\beta}}_j - \boldsymbol{\beta}) = (\mathbf{D}_\ell^{-1}\mathbf{W}_j'\mathbf{M}_j\mathbf{W}_j\mathbf{D}_\ell^{-1})^{-1}\mathbf{D}_\ell^{-1}\mathbf{W}_j'\mathbf{M}_j\boldsymbol{\epsilon}_j.$$

The two components of this expression can be written

$$\mathbf{D}_\ell^{-1}\mathbf{W}_j'\mathbf{M}_j\mathbf{W}_j\mathbf{D}_\ell^{-1} = \mathbf{D}_\ell^{-1}\mathbf{W}_j'\mathbf{W}_j\mathbf{D}_\ell^{-1} - \ell^{-1/2}\mathbf{D}_\ell^{-1}\mathbf{W}_j'\mathbf{X}_j(\ell^{-1}\mathbf{X}_j'\mathbf{X}_j)^{-1}\ell^{-1/2}\mathbf{X}_j'\mathbf{W}_j\mathbf{D}_\ell^{-1},$$

$$\mathbf{D}_\ell^{-1}\mathbf{W}_j'\mathbf{M}_j\boldsymbol{\epsilon}_j = \mathbf{D}_\ell^{-1}\mathbf{W}_j'\boldsymbol{\epsilon}_j - \ell^{-1/2}\mathbf{D}_\ell^{-1}\mathbf{W}_j'\mathbf{X}_j(\ell^{-1}\mathbf{X}_j'\mathbf{X}_j)^{-1}\ell^{-1/2}\mathbf{X}_j'\boldsymbol{\epsilon}_j.$$

For notational convenience let $u_t = \Delta y_t$ and note that $\ell^{-1}\sum_{t\in\tau_j} u_{t-j}u_{t-k} \xrightarrow{p} \gamma_{|j-k|}$ where $\gamma_j = E(u_t u_{t-j})$. The components of the above expressions satisfy

$$\mathbf{D}_\ell^{-1}\mathbf{W}_j'\mathbf{W}_j\mathbf{D}_\ell^{-1} = \begin{bmatrix} 1 & \ell^{-3/2}\displaystyle\sum_{t\in\tau_j} y_{t-1} \\ \ell^{-3/2}\displaystyle\sum_{t\in\tau_j} y_{t-1} & \ell^{-2}\displaystyle\sum_{t\in\tau_j} y_{t-1}^2 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 1 & \sigma_\epsilon\psi(1)m^{3/2}\displaystyle\int_{(j-1)/m}^{j/m} W \\ \sigma_\epsilon\psi(1)m^{3/2}\displaystyle\int_{(j-1)/m}^{j/m} W & \sigma_\epsilon^2\psi(1)^2m^2\displaystyle\int_{(j-1)/m}^{j/m} W^2 \end{bmatrix} = \mathbf{Q}_j,$$

$$\ell^{-1/2}\mathbf{D}_\ell^{-1}\mathbf{W}_j'\mathbf{X}_j = \begin{bmatrix} \ell^{-1}\displaystyle\sum_{t\in\tau_j} u_{t-1} & \cdots & \ell^{-1}\displaystyle\sum_{t\in\tau_j} u_{t-(p-1)} \\ \ell^{-3/2}\displaystyle\sum_{t\in\tau_j} y_{t-1}u_{t-1} & \cdots & \ell^{-3/2}\displaystyle\sum_{t\in\tau_j} y_{t-1}u_{t-(p-1)} \end{bmatrix} \xrightarrow{p} \mathbf{0},$$

$$\ell^{-1}\mathbf{X}_j'\mathbf{X}_j = \begin{bmatrix} \ell^{-1}\sum_{t\in\tau_j} u_{t-1}^2 & \cdots & \ell^{-1}\sum_{t\in\tau_j} u_{t-1}u_{t-(p-1)} \\ \vdots & & \vdots \\ \ell^{-1}\sum_{t\in\tau_j} u_{t-(p-1)}u_{t-1} & \cdots & \ell^{-1}\sum_{t\in\tau_j} u_{t-(p-1)}^2 \end{bmatrix}$$

$$\xrightarrow{p} \begin{bmatrix} \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & & \vdots \\ \gamma_{p-2} & \cdots & \gamma_0 \end{bmatrix} = \mathbf{\Gamma},$$

$$\mathbf{D}_\ell^{-1}\mathbf{W}_j'\boldsymbol{\epsilon}_j = \begin{bmatrix} \ell^{-1/2}\sum_{t\in\tau_j} \epsilon_t \\ \ell^{-1}\sum_{t\in\tau_j} y_{t-1}\epsilon_t \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \sigma_\epsilon m^{1/2}[W(j/m) - W((j-1)/m)] \\ \sigma_\epsilon^2\psi(1)m\int_{(j-1)/m}^{j/m} W dW \end{bmatrix} = \mathbf{q}_j,$$

$$\ell^{-1/2}\mathbf{X}_j'\boldsymbol{\epsilon}_j = \begin{bmatrix} \ell^{-1/2}\sum_{t\in\tau_j} u_{t-1}\epsilon_t \\ \vdots \\ \ell^{-1/2}\sum_{t\in\tau_j} u_{t-(p-1)}\epsilon_t \end{bmatrix} \xrightarrow{d} \mathbf{h}_j \sim N(0, \sigma_\epsilon^2\mathbf{\Gamma}),$$

where the first element in $\mathbf{q}_j$ arises because

$$\ell^{-1/2}\sum_{t\in\tau_j}\epsilon_t = \ell^{-1/2}(S_{j\ell} - S_{(j-1)\ell}) = \ell^{-1/2}(S_{jn/m} - S_{(j-1)n/m})$$
$$\Rightarrow \sigma_\epsilon m^{1/2}[W(j/m) - W((j-1)/m)].$$

In view of the above it follows that, as $\ell \to \infty$,

$$\mathbf{D}_\ell^{-1}\mathbf{W}_j'\mathbf{M}_j\mathbf{W}_j\mathbf{D}_\ell^{-1} = \mathbf{D}_\ell^{-1}\mathbf{W}_j'\mathbf{W}_j\mathbf{D}_\ell^{-1} + o_p(1) \Rightarrow \mathbf{Q}_j,$$

$$\mathbf{D}_\ell^{-1}\mathbf{W}_j'\mathbf{M}_j\boldsymbol{\epsilon}_j = \mathbf{D}_\ell^{-1}\mathbf{W}_j'\boldsymbol{\epsilon}_j + o_p(1) \Rightarrow \mathbf{q}_j,$$

and hence $\mathbf{D}_\ell(\tilde{\boldsymbol{\beta}}_j - \beta) \Rightarrow \mathbf{Q}_j^{-1}\mathbf{q}_j$. Picking out the second element from this expression yields

$$\ell(\tilde{\rho}_j - 1) \Rightarrow \frac{\int_{(j-1)/m}^{j/m} W dW - m\left[W\left(\frac{j}{m}\right) - W\left(\frac{(j-1)}{m}\right)\right]\int_{(j-1)/m}^{j/m} W}{\psi(1)m\left[\int_{(j-1)/m}^{j/m} W^2 - m\left(\int_{(j-1)/m}^{j/m} W\right)^2\right]},$$

where the limiting distribution can be written in terms of the process $W_{j,m}(r)$ as in the statement of the Theorem.

(b) The proof follows as in part (b) of Theorem 1.

REFERENCES

Abadir, K.M. (1993), "The Limiting Distribution of the Autocorrelation Coefficient Under a Unit Root," *Annals of Statistics*, 21, 1058–1070.

Andrews, D.W.K. (1993), "Exactly Median Unbiased Estimation of First Order Autoregressive/Unit Root Models," *Econometrica*, 61, 139–165.

Gonzalo, J., and Pitarakis, J.-Y. (1998), "On the Exact Moments of Nonstandard Asymptotic Distributions in an Unstable AR(1) with Dependent Errors," *International Economic Review*, 39, 71–88.

Kendall, M.G. (1954), "Note on Bias in the Estimation of Autocorrelation," *Biometrika*, 41, 403–404.

Marriott, F.H.C., and Pope, J.A. (1954), "Bias in the Estimation of Autocorrelations," *Biometrika*, 41, 390–402.

Orcutt, G.H., and Winokur, H.S. (1969), "First Order Autoregression: Inference, Estimation, and Prediction," *Econometrica*, 37, 1–14.

Perron, P. (1991), "A Continuous Time Approximation to the Unstable First-Order Autoregressive Process: The Case without an Intercept," *Econometrica*, 59, 211-236.

Phillips, P. C. B. (1987), "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.

Phillips, P. C. B., and Yu, J. (2005), "Jackknifing Bond Option Prices," *Review of Financial Studies*, 18, 707–742.

Quenouille, M.H. (1956), "Notes on Bias in Estimation," *Biometrika*, 43, 353–360.

Shaman, P., and Stine, R. (1988), "The Bias of Autoregressive Coefficient Estimators," *Journal of the American Statistical Association*, 83, 842–848.

Shenton, L.R., and Johnson, W.I. (1965), "Moments of a Serial Correlation Coefficient," *Journal of the Royal Statistical Society Series B*, 27, 308–320.

Shin, D.W., and So, B.S. (2001), "Recursive Mean Adjustment for Unit Root Tests," *Journal of Time Series Analysis*, 22, 595–612.

Stine, R. (1987), "Estimating Properties of Autoregressive Forecasts," *Journal of the American Statistical Association*, 82, 1072–1078.

Tanaka, K. (1996), *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*, New York: Wiley.

Tukey, J.W. (1958), "Bias and Confidence in not-quite Large Samples," *Annals of Mathematical Statistics*, 29, 614.

White, J.S. (1958), "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case," *Annals of Mathematical Statistics*, 29, 1188–1197.