# Jackknife Bias Reduction in the Presence of a Unit Root

Marcus J. Chambers
*University of Essex*

and

Maria Kyriacou
*University of Essex*

January 2010

**Abstract**

This paper analyses the properties of jackknife estimators of the first-order autoregressive coefficient when the time series of interest contains a unit root. It is shown that, when the sub-samples do not overlap, the sub-sample estimators have different limiting distributions from the full-sample estimator and, hence, the jackknife estimator in its usual form does not eliminate fully the first-order bias as intended. The joint moment generating function of the numerator and denominator of these limiting distributions is derived and used to calculate the expectations that determine the optimal jackknife weights. Two methods of avoiding this procedure are proposed and investigated, one based on inclusion of an intercept in the regressions, the other based on adjusting the observations in the sub-samples. Extensions to more general augmented Dickey-Fuller (ADF) regressions are also considered. In addition to the theoretical results extensive simulations reveal the impressive bias reductions that can be obtained with these computationally simple jackknife estimators and they also highlight the importance of correct lag-length selection in ADF regressions.

**Keywords**. Jackknife; bias reduction; unit root.

**J.E.L. classification numbers**. C12; C13; C22.

**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

Jackknife methods of bias reduction have recently been the subject of interest in a number of applications in econometrics. Hahn and Newey (2004) propose the use of the jackknife in a nonlinear panel data model to overcome the bias in fixed effects estimators caused by the incidental parameters problem. Their model is appropriate for *iid* (independent and identically distributed) data so that the delete-one jackknife of Quenouille (1956) and Tukey (1958) can be utilised. An alternative form of the jackknife estimator that can handle non-*iid* data has been employed by Dhaene and Jochmans (2009) also in a nonlinear panel data model but one in which the data are stationary in the time dimension, thereby allowing for dynamic effects. In a pure time series setting Chambers (2010) has investigated the use of jackknife methods in the estimation of stationary autoregressive models and Phillips and Yu (2005) have used jackknife techniques not only to estimate the parameters of a continuous time model but also to estimate bond option prices directly. In all of the above contributions the theoretical bias reduction properties of the jackknife estimators are supported by significant bias reductions obtained in simulation experiments. In view of the typically large negative biases that characterise estimators of autoregressive parameters in models containing unit roots it is therefore of interest to ascertain the extent of bias reduction that can be achieved by jackknife methods in such settings.

The focus of this paper is on jackknife estimation of autoregressive models which contain a unit root. The particular jackknife estimator considered is the one proposed by Phillips and Yu (2005) and also found to perform well in bias reduction by Chambers (2010), namely one based on non-overlapping sub-samples. It is also computationally very easy to implement with the greatest bias reduction typically found to be obtained using just two sub-samples.

The paper is organised as follows. Section 2 considers the Gaussian random walk model and derives the properties of the sub-sample estimators and of the jackknife estimator itself. This reveals that the estimator as usually defined does not work as intended because the sub-sample estimators have different limiting distributions, although these distributions can be used to derive an asymptotic expansion that motivates a set of optimal weights that ensure that the first-order bias is removed correctly. These weights depend on the means of these distributions, and so the joint moment generating function (MGF) of the numerator and denominator of the relevant distributions is derived and used to compute the required expectations and, hence, the optimal weights. Simulations reveal that, although the usual jackknife estimator is capable of reducing bias, the estimator based on the optimal weights produces further bias reductions as well as reductions in the root mean squared error (RMSE).

Section 3 analyses two alternative methods of ensuring that the usual jackknife weights remain optimal in the presence of a unit root. The source of the non-optimality is the fact that the pre-sub-sample values are of the same order of magnitude as the partial sum of disturbances in each sub-sample, and so their effect is not eliminated in the asymptotic analysis. One method of overcoming this is to simply incorporate an intercept in the regressions, while another is to subtract the pre-sub-sample value from the observations in each sub-sample. Both methods are shown to eliminate the first-order bias as intended, although their limiting distributions are different, one depending on demeaned Wiener processes, the other on standard Wiener processes. In simulations the method based on sub-sample adjustment appears to yield smaller bias and RMSE, at least for the random walk model.

The more challenging situation of augmented Dickey-Fuller (ADF) regressions is the topic of section 4, in which the two jackknife estimators of section 3 are used to estimate the ADF parameter of principal interest (i.e. the coefficient on the lagged level of the variable). The properties of the estimators are analysed in a simulation exercise in which the data are first detrended using the generalised least squares (GLS) approach of Elliott, Rothenberg and Stock (1996). While both methods are capable of reducing the bias in this key parameter it is the method based on including an intercept in the regressions that is capable of the most spectacular bias reductions, although correct choice of the number of lags to incorporate in the regression is a vital ingredient to the good performance of the estimator. Data-based methods are available for this purpose, such as the modified information criterion-based method of Ng and Perron (2001), and will be important if the bias-reduced jackknife estimators are to be used for actually testing the unit root hypothesis. This remains an interesting avenue for future research and is under investigation by the authors.

The following notation will be used throughout. The symbol $\overset{d}{=}$ denotes equality in distribution; $\overset{d}{\to}$ denotes convergence in distribution; $\overset{p}{\to}$ 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 ESTIMATION WITH A RANDOM WALK

In order to motivate the use of jackknife methods in models of use in empirical research we begin with a simple example in which the data are generated by the random walk

$$y_t = y_{t-1} + \epsilon_t, \quad \epsilon_t \sim iid\ N(0, \sigma_\epsilon^2), \quad t = 1, \ldots, n, \tag{1}$$

where $y_0$ can be any $O_p(1)$ random variable and is assumed to be observed by the econometrician. The ordinary least squares (OLS) regression of $y_t$ on $y_{t-1}$ will be denoted

$$y_t = \tilde{\beta} y_{t-1} + \tilde{\epsilon}_t, \quad t = 1, \ldots, n, \tag{2}$$

where $\tilde{\epsilon}_t$ denotes the regression residual. In this framework the OLS estimator satisfies

$$n(\tilde{\beta} - 1) = \frac{n^{-1} \sum_{t=1}^{n} y_{t-1} \epsilon_t}{n^{-2} \sum_{t=1}^{n} y_{t-1}^2} \Rightarrow \frac{\int_0^1 W dW}{\int_0^1 W^2} \quad \text{as} \quad n \to \infty. \tag{3}$$

The limiting distribution in (3) is skewed and the estimator suffers from significant negative bias in finite samples. Phillips (1987, Theorem 7.1) demonstrated the validity of an asymptotic expansion for the normalised coefficient estimator, assuming $y_0 = 0$, given by

$$n(\tilde{\beta} - 1) \overset{d}{=} \frac{\int_0^1 W dW}{\int_0^1 W^2} - \frac{\eta}{\sqrt{2n} \int_0^1 W^2} + O_p(n^{-1}), \tag{4}$$

where $\eta$ is a standard normal random variable distributed independently of $W$. Taking expectations in (4), using the independence of $\eta$ 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

$$E(\tilde{\beta} - 1) = -\frac{1.781}{n} + O(n^{-2}), \tag{5}$$

an expansion that motivates the use of the jackknife as a method of bias reduction.

The jackknife offers a simple method of reducing bias in estimators and test statistics by eliminating the leading bias term from expansions of the form of (5), which will serve as a useful reference point. The jackknife estimator combines the full-sample estimator, $\tilde{\beta}$, with a set of $m$ sub-sample estimators, $\tilde{\beta}_j$ ($j = 1, \ldots, m$), the weights assigned to these components depending on the type of sub-sampling employed. In the context of stationary autoregressive time series Chambers (2010) compares alternative methods of sub-sampling and finds the use of non-overlapping sub-samples to perform well in reducing bias, and so it is this approach upon which we shall concentrate here; see also Phillips and Yu (2005) for applications of this jackknife estimator to bond option pricing. The jackknife estimator is

$$\tilde{\beta}_J = \kappa_m \tilde{\beta} + \delta_m \frac{1}{m} \sum_{j=1}^{m} \tilde{\beta}_j, \tag{6}$$

where the weights are given by $\kappa_m = m/(m-1)$ and $\delta_m = -1/(m-1)$ and the length of each sub-sample is $\ell$ with $n = m \times \ell$. The weights are determined on the assumption that each sub-sample estimator also satisfies (5), so that

$$E(\tilde{\beta}_j - 1) = -\frac{1.781}{\ell} + O(\ell^{-2}), \quad j = 1, \ldots, m.$$

In this case it can be shown that

$$\begin{aligned} E(\tilde{\beta}_J - 1) &= \frac{m}{m-1} E(\tilde{\beta} - 1) - \frac{1}{m-1} \frac{1}{m} \sum_{j=1}^{m} E(\tilde{\beta}_j - 1) \\ &= \frac{m}{m-1} \left( -\frac{1.781}{n} + O(n^{-2}) \right) - \frac{1}{m-1} \left( -\frac{1.781}{\ell} + O(\ell^{-2}) \right) \\ &= -\frac{1.781}{m-1} \left( mn^{-1} - \ell^{-1} \right) + O(n^{-2}) = O(n^{-2}), \end{aligned}$$

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 $\tilde{\beta}$.

The problem with the argument above is that the sub-sample estimators do not share the same limiting distribution as the full-sample estimators, which means that the expansions for the bias of the sub-sample 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

$$\tilde{\beta}_j - 1 = \frac{\sum_{t \in \tau_j} y_{t-1} \epsilon_t}{\sum_{t \in \tau_j} y_{t-1}^2}, \quad j = 1, \ldots, m. \tag{7}$$

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

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

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

THEOREM 1. *Let $y_1, \ldots, y_n$ be generated by (1) with $y_0$ being any $O_p(1)$ random variable. Then, if $\ell \to \infty$ as $n \to \infty$:*

*(a) If $m$ is fixed, $\ell(\tilde{\beta}_j - 1) \Rightarrow m^{-1}Z(W, r_j)$ $(j = 1, \ldots, m)$ and*

$$n(\tilde{\beta}_J - 1) \Rightarrow \kappa_m Z(W,r) + \delta_m \sum_{j=1}^{m} m^{-1} Z(W, r_j);$$

*(b) If $m^{-1} + mn^{-1} \to 0$, $n(\tilde{\beta}_J - 1) \Rightarrow Z(W, r)$.*

**Remark 1.** Although it is natural to normalise $\tilde{\beta}_j$ in part (a) by the sub-sample size $\ell$, the stated result is valid only when $m$ is fixed, otherwise the limiting distribution is degenerate. This is because both components of $Z(W, r_j)$, 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 $m^{-1}Z(W, r_j)$ is also $O_p(1/m)$ due to the presence of $m$ in the denominator. Multiplying by $m$, of course, provides the limit for $n(\tilde{\beta} - 1)$ in terms of an $O_p(1)$ random variable and is valid even when $m$ is not held fixed.

**Remark 2.** Note that the numerator of $Z(W, r_j)$ also has the representation

$$\int_{(j-1)/m}^{j/m} W dW \overset{d}{=} \frac{1}{2}\left[W\left(\frac{j}{m}\right)^2 - W\left(\frac{j-1}{m}\right)^2 - \frac{1}{m}\right] \tag{8}$$

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

**Remark 3.** The limiting distribution of the jackknife estimator takes one of two forms, depending on whether $m$ is fixed or is allowed to increase with $n$ in conjunction with $\ell$. When $m$ is fixed the limiting distribution is a weighted average of the limiting distribution of $n(\tilde{\beta} - 1)$ and of the sub-samples $\ell(\tilde{\beta}_j - 1)$. Allowing $m$ to increase with $n$ results in the jackknife estimator inheriting the same limiting distribution as the full-sample estimator $\tilde{\beta}$. Hence unit-root inference could be based on $n(\tilde{\beta}_J - 1)$ using the same critical values as for $n(\tilde{\beta} - 1)$ which are widely tabulated; see, for example, Table B.5 in Hamilton (1994). Note, too, that the condition $m^{-1} + mn^{-1} \to 0$ also implies that $\ell \to \infty$ because $mn^{-1} = \ell^{-1}$.

**Remark 4.** The fact that the distributions $Z(W, r_j)$ in part (a) depend on $j$ implies that the usual weights used to construct the jackknife estimator may not be correct under a unit root, as alluded to above. The following result sheds light on this conjecture, in which it is convenient to define

$$\mu = E\left(Z(W,r)\right), \quad \mu_j = E\left(m^{-1}Z(W,r_j)\right), \quad j = 1,\ldots,m.$$

THEOREM 2. *Let $y_1,\ldots,y_n$ be generated by (1) and assume that $y_0 = 0$. Then:*

*(a) The sub-sample estimators satisfy*

$$\ell(\tilde{\beta}_j - 1) \overset{d}{=} \frac{\displaystyle\int_{(j-1)/m}^{j/m} W dW}{\displaystyle m\int_{(j-1)/m}^{j/m} W^2} - \frac{\eta_j}{\displaystyle m^3\sqrt{2\ell}\int_{(j-1)/m}^{j/m} W^2} + O_p(\ell^{-1}), \quad j = 1,\ldots,m,$$

*where $\eta_j$ $(j = 1,\ldots,m)$ are standard normal random variates distributed independently of each other and of $W$. It follows that*

$$E(\tilde{\beta}_j - 1) = \frac{\mu_j}{\ell} + O(\ell^{-2}), \quad j = 1,\ldots,m.$$

*(b) The optimal jackknife estimator is given by*

$$\tilde{\beta}_J^{opt} = \kappa_m^{opt}\tilde{\beta} + \delta_m^{opt}\frac{1}{m}\sum_{j=1}^{m}\tilde{\beta}_j,$$

*where $\kappa_m^{opt} = -\sum_{j=1}^m \mu_j/\bar{\mu}$, $\delta_m^{opt} = \mu/\bar{\mu}$, and $\bar{\mu} = \mu - \sum_{j=1}^m \mu_j$.*

**Remark 5.** Part (a) of Theorem 2 generalises the result of Phillips (1987) in (4) to the distribution of the sub-sample estimators. The resulting bias expansion clarifies the different properties of these sub-sample estimators, the effect of which is shown in part (b).

**Remark 6.** Part (b) of the Theorem shows the optimal weights for the jackknife estimator when the properties of the estimator differ in the sub-samples. Of course, knowledge of the means of the relevant sub-sample (limit) distributions is required in order to make this estimator operational. In the case of the full sample it is known that $\mu = -1.781$ but for the sub-samples the means are, as yet, unkown. Theorem 3 provides the required result.

THEOREM 3. *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 $0 \le a < b$. 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*

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

*where*

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

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

**Remark 7.** Part (a) of Theorem 3 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 $D$ follow straightforwardly and are given by

$$M_N(\theta_1) = M(\theta_1, 0) = \exp\left(-\frac{\theta_1}{2}(b-a)\right)\left[1 - (b-a)(\theta_1 + a\theta_1^2)\right]^{-1/2}, \quad (9)$$

$$M_D(\theta_2) = M(0, \theta_2) = \left[\cosh\left((b-a)\sqrt{-2\theta_2}\right) + a\sqrt{-2\theta_2}\sinh\left((b-a)\sqrt{-2\theta_2}\right)\right]^{-1/2}, \quad (10)$$

respectively. Some special cases then result.

(i) $[a,b] = [0,1]$
We obtain $M_N(\theta_1) = e^{-\theta_1/2}(1-\theta_1)^{-1/2}$ and $M_D(\theta_2) = (\cosh(\sqrt{-2\theta_2}))^{-1/2}$ while the joint MGF is

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

a result that goes back to White (1958).

(ii) $[a,b] = [(j-1)/m, j/m]$
This is the case of relevance for the non-overlapping jackknife sub-sampling, and it follows that (with $\lambda = \sqrt{-2\theta_2}$)

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

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

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

**Remark 8.** Another potential use of the joint MGF in part (a) is in the computation of the cumulative and probability density functions of the distributions $m^{-1}Z(W, r_j)$. 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 $Z(W, r)$ along with other densities.

**Remark 9.** 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 $\theta_1/m$, while that for $D$ is the same as the full sample MGF evaluated at $\theta_2/m^2$, implying that

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

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

**Remark 10.** The result in part (b) of Theorem 3 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:

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

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

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

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

both of which depend on $m$. However, the limiting distribution of $\ell(\tilde{\beta}_j - 1)$ is $N/(mD)$, and hence the expectation of this distribution does not depend on $m$. Table 1 contains the values of the integrals $I_1((j-1)/m, j/m)/m$ and $I_2((j-1)/m, j/m)/m$ for values of $j = 1, \ldots, 12$, as well as the resulting expectations $\mu_j$ defined in Theorem 2(a). 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.

**Remark 11.** The values of $\mu_j$ in Table 1 can be utilised in Theorem 2(b) to derive the optimal weights for the jackknife estimator; these are reported in Table 2. 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 1.** Values of integrals and expectations for sub-samples

| $j$ | $\frac{1}{m}I_1\left(\frac{(j-1)}{m},\frac{j}{m}\right)$ | $\frac{1}{m}I_2\left(\frac{(j-1)}{m},\frac{j}{m}\right)$ | $\mu_j$ |
|---|---|---|---|
| 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 |

**Table 2.** Values of standard and optimal jackknife weights

| $m$ | $\kappa_m$ | $\delta_m$ | $\kappa_m^{opt}$ | $\delta_m^{opt}$ |
|---|---|---|---|---|
| 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 Table 3. The entries in Table 3 report the bias and RMSE of $\tilde{\beta}$, $\tilde{\beta}_J$ and $\tilde{\beta}_J^{opt}$ as well as the ratios of these quantities for $\tilde{\beta}_J$ and $\tilde{\beta}_J^{opt}$ with respect to $\tilde{\beta}$ obtained from 100,000 replications of the random walk process (1). Results are presented for the values of $m$ that minimise the jackknife bias, as well as for the values of $m$ that minimise the RMSE. In terms of bias it can be seen that the jackknife estimator $\tilde{\beta}_J$ is capable of producing substantial bias reduction over $\tilde{\beta}$ ranging from 49% at $n = 24$ through to 62% at $n = 192$ based on the bias-minimising values of $m$ (which 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, $\tilde{\beta}_J^{opt}$, produces even more spectacular bias reductions, ranging from 76% at $n = 24$ to 97% at $n = 192$ for the bias minimising values of $m$, and from 47% to 86% for the RMSE-minimising values of $m$. The effects of jackknifing on the RMSE are also interesting. When the pursuit of bias reduction is the objective it can be seen that the estimators constructed using the bias-minimising values of $m$ bear the cost of bias reduction in terms of larger variance and hence higher RMSE as compared to $\tilde{\beta}$, the RMSE being almost 30% higher for the optimal estimator. But the results also show that the RMSE-minimising values of $m$ yield jackknife estimators that not only reduce bias but also reduce the overall RMSE compared to the full-sample estimator $\tilde{\beta}$. These RMSE-minimising values of $m$ are seen to increase with $n$.

**Table 3.** Bias and RMSE of OLS and jackknife estimators of $\beta$ in regression without an intercept

| $n$ | $\tilde{\beta}$ | $\tilde{\beta}_J$ | $m$ | $\tilde{\beta}_J/\tilde{\beta}$ | $\tilde{\beta}_J^{opt}$ | $m$ | $\tilde{\beta}_J^{opt}/\tilde{\beta}$ |
|---|---|---|---|---|---|---|---|
| Bias using bias-minimising values of $m$ | | | | | | | |
| 24 | −0.0664 | −0.0340 | 2 | 0.51 | −0.0157 | 2 | 0.24 |
| 48 | −0.0350 | −0.0155 | 2 | 0.44 | −0.0044 | 2 | 0.13 |
| 96 | −0.0180 | −0.0073 | 2 | 0.40 | −0.0012 | 2 | 0.07 |
| 192 | −0.0091 | −0.0035 | 2 | 0.38 | −0.0003 | 2 | 0.03 |
| Bias using RMSE-minimising values of $m$ | | | | | | | |
| 24 | −0.0664 | −0.0447 | 4 | 0.67 | −0.0353 | 6 | 0.53 |
| 48 | −0.0350 | −0.0231 | 6 | 0.66 | −0.0126 | 8 | 0.36 |
| 96 | −0.0180 | −0.0116 | 8 | 0.65 | −0.0049 | 12 | 0.27 |
| 192 | −0.0091 | −0.0055 | 8 | 0.60 | −0.0013 | 12 | 0.14 |
| RMSE using bias-minimising values of $m$ | | | | | | | |
| 24 | 0.1368 | 0.1486 | 2 | 1.09 | 0.1760 | 2 | 1.29 |
| 48 | 0.0717 | 0.0766 | 2 | 1.07 | 0.0917 | 2 | 1.28 |
| 96 | 0.0370 | 0.0394 | 2 | 1.06 | 0.0475 | 2 | 1.28 |
| 192 | 0.0188 | 0.0201 | 2 | 1.07 | 0.0244 | 2 | 1.30 |
| RMSE using RMSE-minimising values of $m$ | | | | | | | |
| 24 | 0.1368 | 0.1313 | 4 | 0.96 | 0.1352 | 6 | 0.99 |
| 48 | 0.0717 | 0.0657 | 6 | 0.92 | 0.0638 | 8 | 0.89 |
| 96 | 0.0370 | 0.0333 | 8 | 0.90 | 0.0312 | 12 | 0.84 |
| 192 | 0.0188 | 0.0168 | 8 | 0.89 | 0.0155 | 12 | 0.83 |

## 3. MODIFIED SUB-SAMPLING IN THE RANDOM WALK MODEL

The analysis of the previous section is extended by considering two distinct methods of ensuring that the sub-sample estimators have the same limiting distributions as the full-sample estimator in order for the usual jackknife estimator to remove the first-order bias in finite samples in the intended manner. The root of the failure of the jackknife in a 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 (1), the observations in sub-sample $j$ satisfy

$$y_t = y_{t-1} + \epsilon_t = y_{(j-1)\ell} + \sum_{i=(j-1)\ell+1}^{t} \epsilon_i, \quad t = (j-1)\ell + 1, \ldots, j\ell, \tag{11}$$

and so the pre-sample value, $y_{(j-1)\ell}$, is $O_p(\sqrt{(j-1)\ell})$ rather than $O_p(1)$ or a constant. The first method of dealing with this simply incorporates an intercept in the regression, while the second adjusts the sub-sample observations by simply subtracting the pre-sample value.

### 3.1. Regression with an Intercept

The data are assumed to satisfy (1), as before, but in this case the OLS regression is now

$$y_t = \hat{\alpha} + \hat{\beta} y_{t-1} + \hat{\epsilon}_t, \quad t = 1, \ldots, n, \tag{12}$$

where $\hat{\epsilon}_t$ denotes the regression residual and the presence of the intercept also ensures that $\hat{\beta}$ is invariant to $y_0$. In this framework the OLS estimator $\hat{\beta}$ satisfies

$$n(\hat{\beta} - 1) = \frac{n^{-1} \sum_{t=1}^{n} y_{t-1}\epsilon_t - \left(n^{-3/2} \sum_{t=1}^{n} y_{t-1}\right)\left(n^{-1/2} \sum_{t=1}^{n} \epsilon_t\right)}{n^{-2} \sum_{t=1}^{n} y_{t-1}^2 - \left(n^{-3/2} \sum_{t=1}^{n} y_{t-1}\right)^2} \Rightarrow Z(W_0, r) \quad \text{as} \quad n \to \infty, \tag{13}$$

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

$$\hat{\beta}_J = \kappa_m \hat{\beta} + \delta_m \frac{1}{m} \sum_{j=1}^{m} \hat{\beta}_j, \tag{14}$$

where $\kappa_m$ and $\delta_m$ are defined following (6) and the sub-sample estimators are

$$\hat{\beta}_j - 1 = \frac{\ell \sum_{t \in \tau_j} y_{t-1}\epsilon_t - \sum_{t \in \tau_j} y_{t-1} \sum_{t \in \tau_j} \epsilon_t}{\ell \sum_{t \in \tau_j} y_{t-1}^2 - \left(\sum_{t \in \tau_j} y_{t-1}\right)^2}, \quad j = 1, \ldots, m. \tag{15}$$

Theorem 4 provides the limiting properties of $\ell(\hat{\beta}_j - 1)$ and, hence, of $\hat{\beta}_J$, 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. *Let $y_1, \ldots, y_n$ be generated by (1) with $y_0$ being any $O_p(1)$ random variable. Then, if $\ell \to \infty$ as $n \to \infty$:*

*(a) If $m$ is fixed, $\ell(\hat{\beta}_j - 1) \Rightarrow m^{-1} Z(W_{j,m}, r_j)$ $(j = 1, \ldots, m)$ and*

$$n(\hat{\beta}_J - 1) \Rightarrow \kappa_m Z(W_0, r) + \delta_m \sum_{j=1}^{m} m^{-1} Z(W_{j,m}, r_j);$$

*(b) If $m^{-1} + mn^{-1} \to 0$, $n(\hat{\beta}_J - 1) \Rightarrow Z(W_0, r)$.*

**Remark 12.** 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 (13), 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

$$m^{-1}Z(W_{j,m},r_j)=\frac{\displaystyle\int_{(j-1)/m}^{j/m}W_{j,m}dW_{j,m}}{\displaystyle m\int_{(j-1)/m}^{j/m}W_{j,m}^2}\overset{d}{=}\frac{\displaystyle\int_0^1 W_0dW_0}{\displaystyle\int_0^1 W_0^2}=Z(W_0,r),\quad j=1,\ldots,m. \tag{16}$$

In view of this equivalence, a result analagous to Theorem 2 motivates the validity of the jackknife estimator (14), and is given below.

THEOREM 5. *Let $y_1,\ldots,y_n$ be generated by (1) with $y_0=0$. Then*

$$n(\hat{\beta}-1)\overset{d}{=}\frac{\displaystyle\int_0^1 W_0dW_0}{\displaystyle\int_0^1 W_0^2}-\frac{\eta_0}{\displaystyle\sqrt{2n}\int_0^1 W_0^2}+O_p(n^{-1}), \tag{17}$$

*where $\eta_0$ denotes a standard normal random variable distributed independently of $W_0$. It follows that*

$$E(\hat{\beta})=1-\frac{5.379}{n}+O(n^{-2}). \tag{18}$$

The form of the expansion in (17) mirrors the one in (4) except that the demeaned Wiener process $W_0$ replaces the standard Wiener process $W$. In addition to the limiting distribution of $\hat{\beta}$ being invariant to $y_0$ Theorem 5 shows that the same property also holds to $O_p(n^{-1})$. The expression for the bias in (18) follows by taking expectations in (17), using the fact that $\eta_0$ has mean zero and is independent of $W_0$, and noting that the mean of $\int_0^1 W_0dW_0/\int_0^1 W_0^2$ is equal to $-5.379$; see, for example, Table 7.2 of Tanaka (1996). The negative bias of $\hat{\beta}$ is clearly evident from the expression in (18). In view of (16) it follows that

$$E(\hat{\beta}_i)=1-\frac{5.379}{\ell}+O(\ell^{-2}),\quad i=1,\ldots,m,$$

and the argument used following (6) can be used here, correctly, to show that $\hat{\beta}_J$ satisfies $E(\hat{\beta}_J)=1+O(n^{-2})$, thereby eliminating the first-order bias term.

Table 4 reports the results of 100,000 replications of the random walk (1) and the resulting bias and RMSE properties of $\hat{\beta}$ and $\hat{\beta}_J$. Not surprisingly the estimator $\hat{\beta}$ is more biased than $\tilde{\beta}$, its theoretical first-order bias being $-5.379/n$ as opposed to $-1.781/n$. Compared to $\hat{\beta}$ the estimator $\hat{\beta}_J$ manages to reduce the bias by 80% at $n=24$ rising to 97% at $n=192$ using the bias-minimising values of $m=2$, the same as for $\tilde{\beta}_J$. However, compared to $\tilde{\beta}_J$, the estimator $\hat{\beta}_J$ achieves even greater bias reductions for $n>24$, reducing the bias of $\tilde{\beta}_J$ by a further 77% when $n=192$. This bias reduction does come at a cost, however. Although the RMSE of $\hat{\beta}_J$ is less than or equal to that of $\hat{\beta}$ for all sample sizes considered, it is greater than that of $\tilde{\beta}_J$ in all of the eight relevant cells in Table 4.

**Table 4.** Bias and RMSE of OLS and jackknife estimators of $\beta$ in regression with an intercept

| | | Bias-minimising $m$ | | | | RMSE-minimising $m$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | $\hat{\beta}$ | $\hat{\beta}_J$ | $m$ | $\hat{\beta}_J/\hat{\beta}$ | $\hat{\beta}_J/\tilde{\beta}_J$ | $\hat{\beta}_J$ | $m$ | $\hat{\beta}_J/\hat{\beta}$ | $\hat{\beta}_J/\tilde{\beta}$ |
| Bias | | | | | | | | | |
| 24 | −0.1985 | −0.0399 | 2 | 0.20 | 1.17 | −0.0673 | 4 | 0.34 | 1.51 |
| 48 | −0.1052 | −0.0116 | 2 | 0.11 | 0.74 | −0.0356 | 8 | 0.34 | 1.54 |
| 96 | −0.0545 | −0.0035 | 2 | 0.06 | 0.48 | −0.0152 | 12 | 0.28 | 1.31 |
| 192 | −0.0276 | −0.0008 | 2 | 0.03 | 0.23 | −0.0044 | 12 | 0.16 | 0.80 |
| RMSE | | | | | | | | | |
| 24 | 0.2524 | 0.2444 | 2 | 0.97 | 1.64 | 0.2013 | 4 | 0.80 | 1.53 |
| 48 | 0.1350 | 0.1316 | 2 | 0.97 | 1.72 | 0.0992 | 8 | 0.73 | 1.51 |
| 96 | 0.0706 | 0.0695 | 2 | 0.99 | 1.76 | 0.0499 | 12 | 0.71 | 1.50 |
| 192 | 0.0360 | 0.0360 | 2 | 1.00 | 0.80 | 0.0248 | 12 | 0.69 | 1.48 |

### 3.2. Sub-Sample Adjustment

The second method considered is to adjust the sub-sample observations to eliminate the effects of the pre-sub-sample value. The adjusted variables in each sub-sample are then

$$y_t^a = y_t - y_{(j-1)\ell}, \quad t = (j-1)\ell + 1, \ldots, j\ell. \tag{19}$$

From (11) it is apparent that $y_t^a = y_{t-1}^a + \epsilon_t$ or $y_t^a = S_t^a = \sum_{i=(j-1)\ell+1}^{t} \epsilon_i$ for $t \in \tau_j$, so that the adjustment effectively makes the pre-sample value equal to zero and re-initialises the innovations to begin at $t = (j-1)\ell + 1$. The sub-sample estimators based on $y_t^a$ satisfy

$$\ell(\beta_j^a - 1) = \frac{\ell^{-1} \sum_{t \in \tau_j} y_{t-1}^a \epsilon_t}{\ell^{-2} \sum_{t \in \tau_j} (y_{t-1}^a)^2},$$

and the corresponding jackknife estimator, $\beta_J^a$, combines $\tilde{\beta}$ with the $\beta_j^a$ in the usual way:

$$\beta_J^a = \kappa_m \tilde{\beta} + \delta_m \frac{1}{m} \sum_{j=1}^{m} \beta_j^a.$$

The effects of the sub-sample adjustments are characterised as follows, in which we define

$$W_{j,m}^a(r) = W(r) - W((j-1)/m), \quad r \in [(j-1)/m, j/m].$$

THEOREM 6. *Let $y_1, \ldots, y_n$ be generated by (1) with $y_0$ being any $O_p(1)$ random variable. Then, if $\ell \to \infty$ as $n \to \infty$:*

*(a) If $m$ is fixed, $\ell(\beta_j^a - 1) \Rightarrow m^{-1} Z(W_{j,m}^a, r_j)$ $(j = 1, \ldots, m)$ and*

$$n(\beta_J^a - 1) \Rightarrow \kappa_m Z(W, r) + \delta_m \sum_{j=1}^{m} m^{-1} Z(W_{j,m}^a, r_j);$$

*(b) If $m^{-1} + mn^{-1} \to 0$, $n(\beta_J^a - 1) \Rightarrow Z(W, r)$.*

**Remark 13.** The Wiener processes $W^a_{j,m}(r)$ have the same properties on the interval $[(j-1)/m, j/m)$ as does $W(r)$ on the interval $[0, 1/m)$, in particular that $W^a_{j,m}(r) \sim N(0, r-((j-1)/m))$ with $W^a_{j,m}((j-1)/m) = 0$. It follows, recalling the properties at the end of Remark 7, that

$$\int_{(j-1)/m}^{j/m} W^a_{j,m} dW^a_{j,m} \overset{d}{=} \int_0^{1/m} W dW \overset{d}{=} \frac{1}{m}\int_0^1 W dW,$$

$$\int_{(j-1)/m}^{j/m} (W^a_{j,m})^2 \overset{d}{=} \int_0^{1/m} W^2 \overset{d}{=} \frac{1}{m^2}\int_0^1 W^2.$$

Hence the limiting distributions of each sub-sample estimator, $\ell(\beta^a_j - 1)$, are the same as those of the full-sample estimator, $n(\tilde{\beta}-1)$, which means that the standard jackknife weights are appropriate in this case and $E(\beta^a_J) = 1 + O(n^{-2})$.

The finite sample performance of the estimator $\beta^a_J$ was assessed using 100,000 replications of the random walk (1); the results are summarised in Table 5. The bias reduction achieved by $\beta^a_J$ is impressive. Its bias is only 20% that of $\tilde{\beta}$ when $n = 24$, and only 2% when $n = 192$, even outperforming the estimator $\tilde{\beta}^{opt}_J$ (see Table 3). The bias of $\beta^a_J$ is also only 34% of the bias of $\hat{\beta}_J$ when $n = 24$ and just 25% when $n = 192$. Although the RMSE of $\beta^a_J$ is greater than the RMSE of $\tilde{\beta}$ at the bias-minimising value of $m = 2$, the results in Table 5 show that the estimator has both lower bias and RMSE than $\tilde{\beta}$ at the RMSE-minimising values of $m$. Furthermore, the RMSE of $\beta^a_J$ is less than the RMSE of $\hat{\beta}_J$. The evidence in Tables 4 and 5 suggests that estimation using the adjusted sub-samples is preferable to including a constant in the regression in terms of both bias and RMSE considerations, at least in the random walk model. We now investigate the performance of these two estimators in the more general setting of ADF regressions.

**Table 5.** Bias and RMSE of OLS and jackknife estimators of $\beta$ in regression with adjusted sub-sample data

| | | Bias-minimising $m$ | | | | RMSE-minimising $m$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| $n$ | $\tilde{\beta}$ | $\beta^a_J$ | $m$ | $\beta^a_J/\tilde{\beta}$ | $\beta^a_J/\hat{\beta}_J$ | $\beta^a_J$ | $m$ | $\beta^a_J/\tilde{\beta}$ | $\beta^a_J/\hat{\beta}_J$ |
| Bias | | | | | | | | | |
| 24 | −0.0664 | −0.0135 | 2 | 0.20 | 0.34 | −0.0364 | 6 | 0.55 | 0.54 |
| 48 | −0.0350 | −0.0036 | 2 | 0.10 | 0.31 | −0.0185 | 12 | 0.53 | 0.52 |
| 96 | −0.0180 | −0.0010 | 2 | 0.05 | 0.29 | −0.0051 | 12 | 0.28 | 0.34 |
| 192 | −0.0091 | −0.0002 | 2 | 0.02 | 0.25 | −0.0014 | 12 | 0.15 | 0.32 |
| RMSE | | | | | | | | | |
| 24 | 0.1368 | 0.1951 | 2 | 1.43 | 0.80 | 0.1469 | 6 | 1.07 | 0.73 |
| 48 | 0.0717 | 0.1021 | 2 | 1.42 | 0.78 | 0.0704 | 12 | 0.98 | 0.71 |
| 96 | 0.0370 | 0.0530 | 2 | 1.43 | 0.76 | 0.0350 | 12 | 0.94 | 0.70 |
| 192 | 0.0188 | 0.0270 | 2 | 1.43 | 0.75 | 0.0176 | 12 | 0.94 | 0.71 |

## 4. JACKKNIFE BIAS REDUCTION IN ADF REGRESSIONS

The ability of the jackknife to considerably reduce bias in the random walk model has been demonstrated clearly in the previous two sections. It is therefore of interest to investigate how well such methods perform in the more challenging setting of ADF regressions which underpin many unit root testing procedures, although we shall focus here on the issue of bias reduction rather than testing. Haldrup and Jansson (2006, p.259) note that biases in the estimation of a key parameter ($\beta_0$ in what follows) "are an important source of size distortion" in unit root tests, and so being able to reduce the bias may lead to tests with better size properties. A key feature for the jackknife to work as intended in eliminating the first-order bias is that the distributions of the sub-sample estimators must be the same as that of the full-sample estimator. The previous section showed that two methods, one including an intercept in the regression, the other subtracting the pre-sub-sample value from the observations in the sub-samples, both satisfy this property, and each will be considered in the ADF setting.

The variable of interest will be assumed to satisfy

$$y_t = d_t + u_t, \quad t = 1, \ldots, n, \tag{20}$$

where $d_t$ denotes a deterministic component and $u_t$ satisfies

$$u_t = \alpha u_{t-1} + v_t, \quad v_t = \delta(L)\epsilon_t, \quad \epsilon_t \sim iid(0, \sigma_\epsilon^2), \quad t = 1, \ldots, n, \tag{21}$$

$\delta(z) = \sum_{j=0}^{\infty} \delta_j z^j$ with $\sum_{j=0}^{\infty} j|\delta_j| < \infty$ and $L$ denotes the lag operator. Such a specification is consistent with $v_t$ being a stationary ARMA$(p, q)$ process of the form $\rho(L)v_t = \theta(L)\epsilon_t$ where $\rho(z) = \sum_{j=0}^{\infty} \rho_j z^j$ and $\theta(z) = \sum_{j=0}^{\infty} \theta_j z^j$, in which case $\delta(z) = \theta(z)/\rho(z)$, but it also allows for more general forms of linear processes. Under these assumptions $v_t$ satisfies the functional central limit theorem $n^{-1/2} \sum_{t=1}^{[nr]} v_t \Rightarrow \sigma W(r)$ on $C[0, 1]$, where $\sigma^2 = \sigma_\epsilon^2 \delta(1)^2$ denotes the long-run variance. The deterministic component, $d_t$, in (20) is assumed to be of the form $d_t = \psi' z_t$ where $z_t = [1, t, t^2, \ldots, t^p]'$, most interest focusing on the cases $p = 0$ and $p = 1$. We shall use the GLS detrending method of Elliott, Rothenberg and Stock (1996). Let $\bar{\alpha} = 1 + \bar{c}/n$ denote the detrending parameter, $\bar{c}$ being a suitably chosen constant, and, for any series $x_0, x_1, \ldots, x_n$, define the quasi-differenced variables $x_0^{\bar{\alpha}} = x_0$ and $x_t^{\bar{\alpha}} = x_t - \bar{\alpha} x_{t-1}$ ($t = 1, \ldots, n$). Then the detrended series is $y_t^d = y_t - \tilde{\psi}' z_t$, where $\tilde{\psi}$ is obtained from the OLS regression of $y_t^{\bar{\alpha}}$ on $z_t^{\bar{\alpha}}$. Elliott, Rothenberg and Stock (1996) recommend that when $p = 0$, $\bar{c} = -7$ and when $p = 1$, $\bar{c} = -13.5$, these values having been chosen so as to make the asymptotic local power function of tests tangent to the asymptotic Gaussian power envelope at the point where power equals one half. Under (20) and (21) the detrended series satisfies

$$y_t^d = \alpha y_{t-1}^d + v_t, \quad t = 1, \ldots, n, \tag{22}$$

the unit root manifesting itself when $\alpha = 1$. Ng and Perron (2001, p.1524) note that "the attractiveness of GLS detrending is that it estimates the deterministic function with more precision and leads to reduced bias" in estimates of the parameters of the ADF regression. It is possible that additional bias reduction can be obtained using jackknife methods.

The ADF testing approach is based on the regression

$$\Delta y_t^d = \beta_0 y_{t-1}^d + \sum_{j=1}^{k} \beta_j \Delta y_{t-j}^d + e_{tk}, \quad t = k+1, \ldots, n, \tag{23}$$

where $\beta_0 = \alpha - 1$. Hence in testing for a unit root the hypothesis to be tested is whether $\beta_0 = 0$. Said and Dickey (1984) showed that the lag length, $k$, needs to satisfy $k = o(n^{1/3})$ in order that $||\tilde{\beta} - \beta|| = o_p(1)$ when $d_t = 0$, where $\beta = (\beta_0, \beta_1, \ldots, \beta_k)'$, $\tilde{\beta}$ denotes the OLS estimator, and $|| \cdot ||$ denotes Euclidean norm; furthermore they showed that $\tilde{\beta}_0$ has the limiting distribution in (3). When GLS detrending is used the limiting distribution of $\tilde{\beta}_0$ changes. Elliott, Rothenberg and Stock (1996) show that, under the null of $\beta_0 = 0$, $n^{-1/2} y_{[nr]}^d \Rightarrow \sigma W(r)$ when $p = 0$, while when $p = 1$, $n^{-1/2} y_{[nr]}^d \Rightarrow \sigma V(r, \bar{c})$, where $V(r, \bar{c}) = W(r) - r[\lambda W(1) + 3(1-\lambda) \int_0^1 sW(s)ds]$ and $\lambda = (1-\bar{c})/(1-\bar{c}+\bar{c}^2/3)$. Under these conditions it follows that, as $n \to \infty$,

$$n\tilde{\beta}_0 \Rightarrow \frac{\frac{1}{2}[W(1)^2 - 1]}{\displaystyle\int_0^1 W^2} \quad (p = 0), \quad n\tilde{\beta}_0 \Rightarrow \frac{\frac{1}{2}[V(1)^2 - 1]}{\displaystyle\int_0^1 V^2} \quad (p = 1), \tag{24}$$

where, in the former representation, $\int_0^1 W dW$ in (3) has been replaced by the equivalent representation $[W(1)^2 - 1]/2$.

The two jackknife estimators considered work in the same way as in the previous section. When an intercept is included in the regressions the estimator of $\beta_0$ will be denoted $\hat{\beta}_{J,0}$ while when the sub-samples themselves are adjusted the estimator will be denoted $\beta_{J,0}^a$. Note that, in the latter case, the adjustments for sub-sample $j$ will be of the form $y_t^a = y_t^d - y_{(j-1)\ell}^d$ $(t = (j-1)\ell + 1, \ldots, j\ell)$ which implies that $\Delta y_t^a = \Delta y_t^d$ so, in effect, it is only the lagged level term in the ADF regression that relies on the adjustment.

Tables 6–9 contain the results of a simulation exercise to examine the extent to which the jackknife estimators are capable of reducing the bias of $\tilde{\beta}_0$ in ADF regressions. A total of 10,000 replications of each experiment were conducted, with $v_t$ in (22) satisfying either an AR, MA or white noise process, each being a special case of the ARMA process $v_t = \rho v_{t-1} + \epsilon_t + \theta \epsilon_{t-1}$ where $\epsilon_t$ is Gaussian white noise with unit variance. The AR parameter was set at either 0, 0.5 or 0.8 while the MA parameter was set equal to 0, $-0.5$ or $-0.8$. These values were chosen because AR(1) processes with large positive parameters and MA(1) processes with negative parameters have been found to cause greatest difficulties (particularly size distortions) for unit root tests. The simulations here are therefore designed to assess the extent of bias in the underlying estimator of $\beta_0$ that is used to construct the test statistics and the ability of the jackknife estimators to reduce this bias. Effective sample sizes of $n^* = 48, 96, 192$ were employed, where $n^* = n - k - 1$, and lags ranging from zero to $[12(n/100)^{1/4}]$ were considered, these equating to 10, 12 and 14 respectively (the corresponding values of $n$ were $n = 59, 109, 207$). Such large numbers of lags in the sub-samples obviously limit the number of sub-samples that can be used to construct the jackknife estimators. Entries in the Tables where insufficient degrees of freedom were available due to the combination of $m$ and $k$ are denoted 'na'.

**Table 6.** Bias of $\tilde{\beta}_0$ in ADF regression ($p = 0$)

| | $\rho$ | 0.0 | 0.0 | 0.0 | 0.5 | 0.8 |
|---|---|---|---|---|---|---|
| $k$ | $\theta$ | −0.8 | −0.5 | 0.0 | 0.0 | 0.0 |
| $n^* = 48$ | | | | | | |
| 0 | | −0.5992 | −0.2201 | −0.0503 | −0.0081 | 0.0094 |
| 1 | | −0.4736 | −0.1499 | −0.0515 | −0.0263 | −0.0114 |
| 2 | | −0.3953 | −0.1221 | −0.0513 | −0.0263 | −0.0115 |
| 3 | | −0.3521 | −0.1126 | −0.0526 | −0.0272 | −0.0120 |
| 4 | | −0.3156 | −0.1066 | −0.0528 | −0.0274 | −0.0122 |
| 5 | | −0.2981 | −0.1075 | −0.0548 | −0.0284 | −0.0128 |
| 6 | | −0.2776 | −0.1059 | −0.0548 | −0.0285 | −0.0129 |
| 7 | | −0.2692 | −0.1087 | −0.0569 | −0.0297 | −0.0137 |
| 8 | | −0.2569 | −0.1085 | −0.0572 | −0.0301 | −0.0139 |
| 9 | | −0.2542 | −0.1124 | −0.0598 | −0.0316 | −0.0147 |
| 10 | | −0.2467 | −0.1133 | −0.0608 | −0.0321 | −0.0150 |
| $n^* = 96$ | | | | | | |
| 0 | | −0.4367 | −0.1203 | −0.0231 | −0.0025 | 0.0054 |
| 1 | | −0.3103 | −0.0751 | −0.0234 | −0.0118 | −0.0049 |
| 2 | | −0.2446 | −0.0589 | −0.0232 | −0.0117 | −0.0049 |
| 3 | | −0.2069 | −0.0525 | −0.0235 | −0.0119 | −0.0050 |
| 4 | | −0.1803 | −0.0491 | −0.0234 | −0.0119 | −0.0050 |
| 5 | | −0.1634 | −0.0480 | −0.0237 | −0.0121 | −0.0051 |
| 6 | | −0.1490 | −0.0470 | −0.0237 | −0.0121 | −0.0051 |
| 7 | | −0.1402 | −0.0473 | −0.0241 | −0.0123 | −0.0052 |
| 8 | | −0.1319 | −0.0469 | −0.0240 | −0.0122 | −0.0052 |
| 9 | | −0.1268 | −0.0475 | −0.0244 | −0.0125 | −0.0053 |
| 10 | | −0.1218 | −0.0476 | −0.0245 | −0.0124 | −0.0053 |
| 11 | | −0.1192 | −0.0482 | −0.0248 | −0.0126 | −0.0054 |
| 12 | | −0.1155 | −0.0480 | −0.0247 | −0.0126 | −0.0054 |
| $n^* = 192$ | | | | | | |
| 0 | | −0.2992 | −0.0633 | −0.0107 | −0.0007 | 0.0030 |
| 1 | | −0.1939 | −0.0372 | −0.0107 | −0.0054 | −0.0022 |
| 2 | | −0.1456 | −0.0285 | −0.0107 | −0.0053 | −0.0022 |
| 3 | | −0.1186 | −0.0248 | −0.0107 | −0.0054 | −0.0022 |
| 4 | | −0.1006 | −0.0230 | −0.0107 | −0.0054 | −0.0022 |
| 5 | | −0.0888 | −0.0222 | −0.0107 | −0.0054 | −0.0022 |
| 6 | | −0.0797 | −0.0217 | −0.0107 | −0.0054 | −0.0022 |
| 7 | | −0.0736 | −0.0216 | −0.0108 | −0.0054 | −0.0022 |
| 8 | | −0.0683 | −0.0214 | −0.0107 | −0.0054 | −0.0022 |
| 9 | | −0.0647 | −0.0215 | −0.0108 | −0.0054 | −0.0022 |
| 10 | | −0.0615 | −0.0214 | −0.0108 | −0.0054 | −0.0022 |
| 11 | | −0.0596 | −0.0216 | −0.0109 | −0.0054 | −0.0022 |
| 12 | | −0.0574 | −0.0215 | −0.0108 | −0.0054 | −0.0022 |
| 13 | | −0.0563 | −0.0216 | −0.0109 | −0.0055 | −0.0022 |
| 14 | | −0.0548 | −0.0215 | −0.0108 | −0.0054 | −0.0022 |

**Table 7.** Ratios of bias of $\beta^a_{J,0}$ and $\hat{\beta}_{J,0}$ to bias of $\tilde{\beta}_0$ in ADF regression ($p = 0$)

| | | $\beta^a_{J,0}/\tilde{\beta}_0$ | | | | | $\hat{\beta}_{J,0}/\tilde{\beta}_0$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\rho$ | 0.0 | 0.0 | 0.0 | 0.5 | 0.8 | 0.0 | 0.0 | 0.0 | 0.5 | 0.8 |
| $k$ | $\theta$ | −0.8 | −0.5 | 0.0 | 0.0 | 0.0 | −0.8 | −0.5 | 0.0 | 0.0 | 0.0 |
| $n^* = 48$ | | | | $m = 2$ | | | | | $m = 4$ | | |
| 0 | | 0.97 | 0.66 | 0.43 | 0.61 | 0.06 | 1.11 | 0.77 | 0.23 | −0.29 | 0.37 |
| 1 | | 0.92 | 0.56 | 0.41 | 0.37 | 0.34 | 1.05 | 0.51 | 0.11 | −0.09 | −0.50 |
| 2 | | 0.89 | 0.54 | 0.44 | 0.41 | 0.38 | 0.99 | 0.38 | 0.06 | −0.14 | −0.55 |
| 3 | | 0.86 | 0.50 | 0.42 | 0.38 | 0.32 | 0.88 | 0.21 | −0.06 | −0.27 | −0.70 |
| 4 | | 0.85 | 0.52 | 0.46 | 0.41 | 0.33 | 0.80 | 0.13 | −0.10 | −0.32 | −0.80 |
| 5 | | 0.82 | 0.50 | 0.43 | 0.37 | 0.26 | 0.67 | −0.00 | −0.24 | −0.49 | −1.03 |
| 6 | | 0.81 | 0.51 | 0.45 | 0.39 | 0.27 | 0.58 | −0.08 | −0.32 | −0.59 | −1.20 |
| 7 | | 0.78 | 0.49 | 0.41 | 0.35 | 0.20 | 0.43 | −0.21 | −0.46 | −0.78 | −1.47 |
| 8 | | 0.78 | 0.50 | 0.43 | 0.36 | 0.19 | 0.34 | −0.28 | −0.57 | −0.95 | −1.72 |
| 9 | | 0.76 | 0.47 | 0.38 | 0.31 | 0.11 | 0.17 | −0.46 | −0.80 | −1.23 | −2.06 |
| 10 | | 0.75 | 0.48 | 0.40 | 0.31 | 0.07 | 0.05 | na | na | na | na |
| $n^* = 96$ | | | | $m = 2$ | | | | | $m = 6$ | | |
| 0 | | 0.85 | 0.48 | 0.29 | 0.61 | 0.00 | 1.13 | 0.63 | 0.14 | −0.38 | 0.27 |
| 1 | | 0.77 | 0.39 | 0.28 | 0.27 | 0.25 | 1.04 | 0.39 | 0.07 | −0.07 | −0.40 |
| 2 | | 0.73 | 0.38 | 0.29 | 0.28 | 0.27 | 0.95 | 0.27 | 0.04 | −0.10 | −0.43 |
| 3 | | 0.69 | 0.34 | 0.28 | 0.27 | 0.25 | 0.84 | 0.15 | −0.04 | −0.18 | −0.51 |
| 4 | | 0.67 | 0.35 | 0.29 | 0.28 | 0.26 | 0.76 | 0.10 | −0.07 | −0.20 | −0.55 |
| 5 | | 0.64 | 0.33 | 0.28 | 0.26 | 0.23 | 0.65 | 0.01 | −0.14 | −0.28 | −0.64 |
| 6 | | 0.63 | 0.35 | 0.29 | 0.28 | 0.25 | 0.58 | −0.02 | −0.17 | −0.32 | −0.69 |
| 7 | | 0.60 | 0.33 | 0.28 | 0.27 | 0.21 | 0.47 | −0.11 | −0.25 | −0.40 | −0.79 |
| 8 | | 0.60 | 0.35 | 0.29 | 0.28 | 0.22 | 0.41 | −0.14 | −0.28 | −0.45 | −0.85 |
| 9 | | 0.57 | 0.33 | 0.28 | 0.26 | 0.18 | 0.31 | −0.21 | −0.37 | −0.54 | −0.95 |
| 10 | | 0.57 | 0.35 | 0.29 | 0.26 | 0.18 | 0.25 | −0.25 | −0.41 | −0.58 | −1.02 |
| 11 | | 0.55 | 0.33 | 0.27 | 0.23 | 0.14 | 0.14 | −0.34 | −0.50 | −0.68 | −1.15 |
| 12 | | 0.56 | 0.34 | 0.27 | 0.24 | 0.14 | 0.09 | −0.37 | −0.54 | −0.73 | −1.22 |
| $n^* = 192$ | | | | $m = 2$ | | | | | $m = 12$ | | |
| 0 | | 0.70 | 0.32 | 0.20 | 0.88 | 0.06 | 1.08 | 0.45 | 0.07 | −0.64 | 0.16 |
| 1 | | 0.61 | 0.25 | 0.19 | 0.17 | 0.16 | 0.92 | 0.26 | 0.03 | −0.04 | −0.26 |
| 2 | | 0.56 | 0.24 | 0.19 | 0.18 | 0.17 | 0.81 | 0.18 | 0.02 | −0.06 | −0.28 |
| 3 | | 0.52 | 0.22 | 0.18 | 0.17 | 0.16 | 0.69 | 0.10 | −0.02 | −0.10 | −0.33 |
| 4 | | 0.49 | 0.22 | 0.19 | 0.18 | 0.16 | 0.61 | 0.06 | −0.04 | −0.13 | −0.35 |
| 5 | | 0.46 | 0.21 | 0.18 | 0.17 | 0.15 | 0.52 | 0.01 | −0.08 | −0.17 | −0.39 |
| 6 | | 0.45 | 0.22 | 0.19 | 0.18 | 0.16 | 0.45 | −0.01 | −0.10 | −0.18 | −0.42 |
| 7 | | 0.43 | 0.21 | 0.18 | 0.17 | 0.15 | 0.38 | −0.06 | −0.14 | −0.23 | −0.47 |
| 8 | | 0.42 | 0.22 | 0.19 | 0.18 | 0.15 | 0.33 | −0.08 | −0.16 | −0.25 | −0.50 |
| 9 | | 0.40 | 0.21 | 0.18 | 0.17 | 0.13 | 0.25 | −0.12 | −0.21 | −0.29 | −0.54 |
| 10 | | 0.40 | 0.22 | 0.19 | 0.17 | 0.14 | 0.21 | −0.14 | −0.22 | −0.31 | −0.57 |
| 11 | | 0.39 | 0.21 | 0.18 | 0.16 | 0.12 | 0.15 | −0.18 | −0.27 | −0.36 | −0.63 |
| 12 | | 0.39 | 0.22 | 0.18 | 0.16 | 0.12 | 0.12 | −0.20 | −0.29 | −0.39 | −0.66 |
| 13 | | 0.38 | 0.21 | 0.17 | 0.15 | 0.10 | 0.06 | −0.24 | −0.34 | −0.44 | −0.71 |
| 14 | | 0.38 | 0.21 | 0.17 | 0.15 | 0.11 | 0.03 | na | na | na | na |

**Table 8.** Bias of $\tilde{\beta}_0$ in ADF regression ($p = 1$)

| $k$ | $\rho$ / $\theta$ | 0.0 / −0.8 | 0.0 / −0.5 | 0.0 / 0.0 | 0.5 / 0.0 | 0.8 / 0.0 |
|---|---|---|---|---|---|---|
| $n^* = 48$ | | | | | | |
| 0 | | −0.8330 | −0.4346 | −0.1417 | −0.0526 | −0.0178 |
| 1 | | −0.7593 | −0.3449 | −0.1486 | −0.0809 | −0.0411 |
| 2 | | −0.7019 | −0.3117 | −0.1527 | −0.0833 | −0.0425 |
| 3 | | −0.6762 | −0.3064 | −0.1598 | −0.0875 | −0.0449 |
| 4 | | −0.6502 | −0.3055 | −0.1649 | −0.0905 | −0.0467 |
| 5 | | −0.6481 | −0.3161 | −0.1733 | −0.0951 | −0.0493 |
| 6 | | −0.6380 | −0.3222 | −0.1782 | −0.0979 | −0.0510 |
| 7 | | −0.6455 | −0.3360 | −0.1870 | −0.1031 | −0.0540 |
| 8 | | −0.6430 | −0.3441 | −0.1926 | −0.1065 | −0.0559 |
| 9 | | −0.6568 | −0.3606 | −0.2029 | −0.1123 | −0.0591 |
| 10 | | −0.6624 | −0.3715 | −0.2095 | −0.1158 | −0.0611 |
| $n^* = 96$ | | | | | | |
| 0 | | −0.6959 | −0.2710 | −0.0725 | −0.0246 | −0.0068 |
| 1 | | −0.5702 | −0.1933 | −0.0744 | −0.0389 | −0.0178 |
| 2 | | −0.4924 | −0.1664 | −0.0754 | −0.0394 | −0.0181 |
| 3 | | −0.4470 | −0.1576 | −0.0773 | −0.0405 | −0.0186 |
| 4 | | −0.4131 | −0.1539 | −0.0785 | −0.0411 | −0.0189 |
| 5 | | −0.3933 | −0.1546 | −0.0804 | −0.0422 | −0.0194 |
| 6 | | −0.3763 | −0.1554 | −0.0817 | −0.0429 | −0.0198 |
| 7 | | −0.3682 | −0.1585 | −0.0838 | −0.0440 | −0.0203 |
| 8 | | −0.3595 | −0.1603 | −0.0850 | −0.0446 | −0.0206 |
| 9 | | −0.3571 | −0.1640 | −0.0872 | −0.0458 | −0.0212 |
| 10 | | −0.3536 | −0.1664 | −0.0887 | −0.0466 | −0.0216 |
| 11 | | −0.3547 | −0.1704 | −0.0908 | −0.0477 | −0.0221 |
| 12 | | −0.3536 | −0.1726 | −0.0921 | −0.0485 | −0.0225 |
| $n^* = 192$ | | | | | | |
| 0 | | −0.5417 | −0.1561 | −0.0362 | −0.0115 | −0.0026 |
| 1 | | −0.3981 | −0.1031 | −0.0367 | −0.0188 | −0.0081 |
| 2 | | −0.3224 | −0.0859 | −0.0369 | −0.0189 | −0.0081 |
| 3 | | −0.2783 | −0.0794 | −0.0374 | −0.0191 | −0.0082 |
| 4 | | −0.2482 | −0.0764 | −0.0376 | −0.0193 | −0.0083 |
| 5 | | −0.2287 | −0.0756 | −0.0381 | −0.0195 | −0.0084 |
| 6 | | −0.2137 | −0.0753 | −0.0384 | −0.0197 | −0.0085 |
| 7 | | −0.2043 | −0.0759 | −0.0390 | −0.0200 | −0.0086 |
| 8 | | −0.1963 | −0.0763 | −0.0392 | −0.0201 | −0.0087 |
| 9 | | −0.1914 | −0.0771 | −0.0397 | −0.0204 | −0.0088 |
| 10 | | −0.1871 | −0.0777 | −0.0401 | −0.0206 | −0.0089 |
| 11 | | −0.1852 | −0.0788 | −0.0407 | −0.0208 | −0.0090 |
| 12 | | −0.1828 | −0.0793 | −0.0409 | −0.0210 | −0.0091 |
| 13 | | −0.1823 | −0.0803 | −0.0415 | −0.0213 | −0.0092 |
| 14 | | −0.1813 | −0.0809 | −0.0418 | −0.0214 | −0.0093 |

**Table 9.** Ratios of bias of $\beta^a_{J,0}$ and $\hat{\beta}_{J,0}$ to bias of $\tilde{\beta}_0$ in ADF regression ($p = 1$)

| | | $\beta^a_{J,0}/\tilde{\beta}_0$ | | | | | $\hat{\beta}_{J,0}/\tilde{\beta}_0$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\rho$ | 0.0 | 0.0 | 0.0 | 0.5 | 0.8 | 0.0 | 0.0 | 0.0 | 0.5 | 0.8 |
| $k$ | $\theta$ | −0.8 | −0.5 | 0.0 | 0.0 | 0.0 | −0.8 | −0.5 | 0.0 | 0.0 | 0.0 |
| $n^* = 48$ | | | | $m = 2$ | | | | | $m = 4$ | | |
| 0 | | 1.07 | 1.01 | 1.00 | 1.21 | 2.11 | 0.96 | 0.81 | 0.46 | 0.08 | −1.06 |
| 1 | | 1.07 | 1.00 | 0.99 | 0.99 | 1.00 | 0.90 | 0.65 | 0.41 | 0.30 | 0.13 |
| 2 | | 1.09 | 1.03 | 1.02 | 1.02 | 1.02 | 0.84 | 0.57 | 0.39 | 0.27 | 0.11 |
| 3 | | 1.09 | 1.02 | 1.00 | 0.98 | 0.96 | 0.77 | 0.47 | 0.32 | 0.19 | 0.01 |
| 4 | | 1.10 | 1.05 | 1.02 | 1.00 | 0.96 | 0.71 | 0.43 | 0.29 | 0.13 | −0.07 |
| 5 | | 1.10 | 1.04 | 1.00 | 0.95 | 0.89 | 0.62 | 0.35 | 0.19 | −0.02 | −0.27 |
| 6 | | 1.11 | 1.05 | 1.01 | 0.96 | 0.88 | 0.55 | 0.28 | 0.10 | −0.15 | −0.45 |
| 7 | | 1.10 | 1.03 | 0.98 | 0.91 | 0.81 | 0.43 | 0.15 | −0.08 | −0.40 | −0.78 |
| 8 | | 1.11 | 1.04 | 0.98 | 0.90 | 0.79 | 0.31 | 0.02 | −0.27 | −0.68 | −1.10 |
| 9 | | 1.10 | 1.00 | 0.93 | 0.82 | 0.66 | 0.05 | −0.28 | −0.62 | −1.25 | −1.78 |
| 10 | | 1.12 | 1.02 | 0.91 | 0.75 | 0.54 | na | na | na | na | na |
| $n^* = 96$ | | | | $m = 2$ | | | | | $m = 6$ | | |
| 0 | | 1.02 | 0.94 | 0.92 | 1.15 | 2.36 | 0.97 | 0.78 | 0.42 | 0.03 | −1.32 |
| 1 | | 1.02 | 0.93 | 0.92 | 0.91 | 0.92 | 0.91 | 0.62 | 0.37 | 0.27 | 0.09 |
| 2 | | 1.02 | 0.94 | 0.93 | 0.93 | 0.94 | 0.86 | 0.53 | 0.35 | 0.25 | 0.06 |
| 3 | | 1.02 | 0.94 | 0.92 | 0.91 | 0.89 | 0.78 | 0.44 | 0.29 | 0.18 | −0.02 |
| 4 | | 1.02 | 0.95 | 0.94 | 0.92 | 0.90 | 0.73 | 0.39 | 0.27 | 0.15 | −0.08 |
| 5 | | 1.01 | 0.95 | 0.92 | 0.90 | 0.85 | 0.65 | 0.32 | 0.20 | 0.06 | −0.21 |
| 6 | | 1.02 | 0.96 | 0.93 | 0.90 | 0.85 | 0.59 | 0.29 | 0.16 | −0.00 | −0.31 |
| 7 | | 1.02 | 0.95 | 0.91 | 0.87 | 0.80 | 0.51 | 0.21 | 0.06 | −0.14 | −0.50 |
| 8 | | 1.02 | 0.95 | 0.91 | 0.87 | 0.79 | 0.45 | 0.16 | −0.01 | −0.25 | −0.66 |
| 9 | | 1.02 | 0.94 | 0.89 | 0.84 | 0.73 | 0.34 | 0.05 | −0.17 | −0.46 | −0.96 |
| 10 | | 1.02 | 0.94 | 0.89 | 0.83 | 0.71 | 0.26 | −0.04 | −0.31 | −0.68 | −1.27 |
| 11 | | 1.01 | 0.92 | 0.85 | 0.78 | 0.63 | 0.11 | −0.23 | −0.58 | −1.06 | −1.81 |
| 12 | | 1.02 | 0.92 | 0.84 | 0.75 | 0.57 | −0.03 | −0.41 | −0.88 | −1.47 | −2.40 |
| $n^* = 192$ | | | | $m = 2$ | | | | | $m = 12$ | | |
| 0 | | 0.99 | 0.90 | 0.85 | 1.09 | 2.67 | 0.99 | 0.79 | 0.40 | −0.03 | −1.78 |
| 1 | | 0.98 | 0.87 | 0.84 | 0.84 | 0.85 | 0.94 | 0.61 | 0.34 | 0.24 | 0.03 |
| 2 | | 0.97 | 0.87 | 0.86 | 0.86 | 0.86 | 0.89 | 0.51 | 0.32 | 0.22 | −0.00 |
| 3 | | 0.96 | 0.87 | 0.85 | 0.84 | 0.81 | 0.81 | 0.42 | 0.26 | 0.14 | −0.11 |
| 4 | | 0.96 | 0.88 | 0.86 | 0.85 | 0.82 | 0.76 | 0.37 | 0.23 | 0.10 | −0.17 |
| 5 | | 0.96 | 0.87 | 0.84 | 0.82 | 0.76 | 0.68 | 0.29 | 0.16 | 0.00 | −0.32 |
| 6 | | 0.96 | 0.88 | 0.85 | 0.82 | 0.76 | 0.62 | 0.25 | 0.11 | −0.07 | −0.45 |
| 7 | | 0.95 | 0.87 | 0.83 | 0.79 | 0.70 | 0.53 | 0.17 | −0.00 | −0.22 | −0.67 |
| 8 | | 0.95 | 0.87 | 0.83 | 0.79 | 0.68 | 0.46 | 0.11 | −0.09 | −0.35 | −0.87 |
| 9 | | 0.95 | 0.86 | 0.80 | 0.75 | 0.62 | 0.36 | −0.01 | −0.25 | −0.57 | −1.21 |
| 10 | | 0.95 | 0.86 | 0.80 | 0.74 | 0.60 | 0.28 | −0.11 | −0.39 | −0.80 | −1.54 |
| 11 | | 0.94 | 0.84 | 0.77 | 0.69 | 0.51 | 0.14 | −0.29 | −0.66 | −1.19 | −2.14 |
| 12 | | 0.95 | 0.84 | 0.75 | 0.66 | 0.45 | −0.00 | −0.46 | −0.94 | −1.62 | −2.79 |
| 13 | | 0.94 | 0.81 | 0.70 | 0.57 | 0.31 | −0.28 | −0.82 | −1.50 | −2.55 | −3.93 |
| 14 | | 0.94 | 0.79 | 0.67 | 0.47 | 0.15 | na | na | na | na | na |

Table 6 contains the bias of $\tilde{\beta}_0$ when $p = 0$. In the case of MA(1) disturbances the bias is a decreasing function of lag length $k$ while in the AR(1) case it tends to increase with $k$ for the smallest sample size or is otherwise static. In the MA case a large number of lags is required because the regression is misspecified while in the AR case the regression is over-parameterised for $k > 1$. The bias is also smallest for the AR cases. The actual biases in Table 6 (and, later, in Table 8) serve as a useful reference point against which the ratios of jackknife bias to the bias of $\tilde{\beta}_0$ reported in later tables can be compared.

Table 7 reports the ratios of the bias of $\beta^a_{J,0}$ and $\hat{\beta}_{J,0}$ to the bias of $\tilde{\beta}_0$ when $p = 0$. A number of features are evident. First, the estimator $\beta^a_{J,0}$ is able to reduce bias uniformly across all parameterisations considered, with $m = 2$ producing the greatest bias reduction for each sample size. The proportion of bias of $\tilde{\beta}_0$ that is eliminated is, broadly, an increasing function of sample size and also increases as $v_t$ passes from an MA to an AR process. The bias of this jackknife estimator is also always of the same sign as $\tilde{\beta}_0$ (i.e. negative in all but three entries in Table 6). The performance of $\hat{\beta}_0$ is rather different to that of $\beta^a_{J,0}$. In some cases it is more biased than $\tilde{\beta}_0$ but is also capable of some spectacular reductions in bias over and above those achieved by $\beta^a_{J,0}$, particularly for the MA cases when using a large number of lags. The bias-minimising values of $m$ for $\hat{\beta}_0$ are also greater than the corresponding values for $\beta^a_{J,0}$ The bias performance of $\hat{\beta}_0$ is affected more by lag length $k$ than is the performance of $\beta^a_{J,0}$. The bias of $\hat{\beta}_0$ is also of a different sign to that of $\tilde{\beta}_0$ in the majority of entries in Table 7, particularly in the autoregressive cases. Correct choice of $k$ is therefore extremely important for $\hat{\beta}_0$.

Tables 8 and 9 report the same information as Tables 6 and 7 but this time for $p = 1$. Note that the presence of an intercept in the regressions renders the largest lag lengths infeasible when $n^a st = 48, 192$. The biases of $\tilde{\beta}_0$ in Table 8 are larger in absolute value than those in Table 6 but otherwise display the same broad patterns. The pattern of bias of the jackknife estimators in Table 9 is, however, different to that reported in Table 7. The estimator $\beta^a_{J,0}$ is unable to reduce bias at all under MA disturbances for $n^a st = 48$ but moderate reductions are obtained for $n = 192$. The performance of $\beta^a_{J,0}$ is somewhat better in the AR case and for the largest sample size. The estimator $\hat{\beta}_0$ appears to perform rather better than $\beta^a_{J,0}$ in reducing bias and once again the importance of correct lag length selection is evident. When the disturbances are MA(1) it is possible to virtually eliminate the bias provided $k$ is sufficiently large while the same is possible for smaller $k$ when the disturbances are an AR process. It therefore seems that the jackknife estimator $\hat{\beta}_0$, allied to a suitable method of lag length selection, provides a promising approach to bias reduction in ADF regressions.

## 5. CONCLUDING COMMENTS

This paper has analysed the properties of jackknife estimators when the time series of interest contains a unit root. Following previous work on the application of jackknife methods in stationary settings, such as Chambers (2010) and Phillips and Yu (2005), the focus has been on the use of non-overlapping sub-intervals, but it has been shown here that, in the setting of a random walk model, the usual jackknife weights are not applicable owing to the differing nature of the distributions of the sub-sample estimators. An expansion of the bias of the sub-sample estimators demonstrates that the leading term depends on the expectation of the limiting sub-sample distributions, which are not equal to one another. The values of these expectations can, however, be obtained via use of the joint moment generating function of the random variables that characterise the limiting distributions, and this MGF is derived in Theorem 3, which also reports the expectations of interest. A simulation experiment shows that, although the usual jackknife estimator is capable of reducing the bias of the OLS estimator in a random walk setting, the jackknife estimator based on the optimal weights does produce further gains in bias reduction and in RMSE.

Two methods of ensuring equivalence of the limiting sub-sample distributions are also considered, which means that the usual jackknife weights can be utilised. One method involves adjusting the sub-samples by subtracting the value of the pre-sub-sample observation, thereby effectively re-setting the pre-sample values of the sub-samples to zero. The other method simply incorporates an intercept into the full- and sub-sample regressions. Both methods are shown to produce sub-sample estimators sharing the same form of limiting distribution as the full-sample estimator, although these distributions are different between the two methods. Simulations show that both methods achieve bias reduction as intended, with the first capable of also achieving reductions in RMSE. These two methods are also applied in the more general and challenging setting of an ADF regression, where this time the second method is capable of producing the greatest bias reduction for the parameter of interest, provided that the number of lags in the regression is chosen appropriately. Data-based methods of doing this in ADF regressions have been proposed by Ng and Perron (2001), and it will be important to investigate how well the jackknife performs in conjunction with such lag selection techniques, not to mention the equally important extension of the techniques to provide a basis for actually conducting tests for unit roots. Such topics are under active investigation by the authors and the results will be reported in subsequent work.

*REFERENCES*


Chambers, M.J. (2010) Jackknife estimation and inference in stationary autoregressive models. Preprint.

Dhaene, G. & K. Jochmans (2009) Split-panel jackknife estimation of fixed effect models. Preprint.

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

Gonzalo, J. & J. Pitarakis (1998) On the exact moments of nonstandard asymptotic distributions in an unstable AR(1) with dependent errors. *International Economic Review* 39, 71–88.

Haldrup, N. & M. Jansson (2006) Improving size and power in unit root testing. In T.C. Mills and K. Patterson (eds.), *Handbook of Econometrics, Volume 1: Econometric Theory*, pp.252–277. Palgrave Macmillan.

Hahn, J. & W. Newey (2004) Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72, 1295–1319.

Hamilton, J.D. (1994) *Time Series Analysis.* Princeton University Press, Princeton.

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

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. & J. Yu (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.

Said, S.E. & D.A. Dickey (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71, 599-607.

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

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.

APPENDIX

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

LEMMA A1. *Let $y_1, \ldots, y_n$ be generated by (1) with $y_0$ being any $O_p(1)$ random variable. Then, if $\ell \to \infty$ as $n \to \infty$:*

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

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

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

**Proof.** Let $S_{[nr]} = \sum_{j=1}^{[nr]} \epsilon_j$, which satisfies $n^{-1/2} S_{[nr]} \Rightarrow \sigma_\epsilon W(r)$ as $n \to \infty$, and note that $y_t = y_0 + S_t$. 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 $n \to \infty$,

$$\ell^{-3/2} \sum_{t \in \tau_j} S_{t-1} = m^{3/2} \int_{(j-1)/m}^{j/m} n^{-1/2} S_{[nr]} dr \Rightarrow \sigma_\epsilon m^{3/2} \int_{(j-1)/m}^{j/m} W,$$

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

These expressions are used in summations of $y_{t-1}$ and $y_{t-1}^2$ in what follows.
(a) First note that $\sum_{t \in \tau_j} y_{t-1} = \ell y_0 + \sum_{t \in \tau_j} S_{t-1}$, so that

$$\ell^{-3/2} \sum_{t \in \tau_j} y_{t-1} = \ell^{-1/2} y_0 + \ell^{-3/2} \sum_{t \in \tau_j} S_{t-1} \Rightarrow \sigma_\epsilon m^{3/2} \int_{(j-1)/m}^{j/m} W$$

as required.
(b) As $y_{t-1}^2 = y_0^2 + 2y_0 S_{t-1} + S_{t-1}^2$ it follows that

$$\ell^{-2} \sum_{t \in \tau_j} y_{t-1}^2 = \ell^{-1} y_0^2 + 2\ell^{-1/2} y_0 \ell^{-3/2} \sum_{t \in \tau_j} S_{t-1} + \ell^{-2} \sum_{t \in \tau_j} S_{t-1}^2 \Rightarrow \sigma_\epsilon^2 m^2 \int_{(j-1)/m}^{j/m} W^2.$$

(c) It is possible to write

$$\sum_{t\in\tau_j} y_{t-1}\epsilon_t = \frac{1}{2}\left(y_{j\ell}^2 - y_{(j-1)\ell}^2 - \sum_{t\in\tau_j}\epsilon_t^2\right).$$

Now $\ell^{-1/2}y_{j\ell} = \ell^{-1/2}y_0 + \ell^{-1/2}S_{j\ell} = m^{1/2}n^{-1/2}S_{[nj/m]} + o_p(1) \Rightarrow \sigma_\epsilon m^{1/2}W(j/m)$ and, similarly, $\ell^{-1/2}y_{(j-1)\ell} \Rightarrow \sigma_\epsilon m^{1/2}W((j-1)/m)$. Furthermore, $\ell^{-1}\sum_{t\in\tau_j}\epsilon_t^2 \xrightarrow{p} \sigma_\epsilon^2$, so that

$$\begin{aligned}\ell^{-1}\sum_{t\in\tau_j} y_{t-1}\epsilon_t &\Rightarrow \frac{\sigma_\epsilon^2}{2}\left[mW\left(\frac{j}{m}\right)^2 - mW\left(\frac{j-1}{m}\right)^2 - 1\right]\\ &= \frac{\sigma_\epsilon^2 m}{2}\left[W\left(\frac{j}{m}\right)^2 - W\left(\frac{j-1}{m}\right)^2 - \frac{1}{m}\right].\end{aligned}$$

The stated result then holds because of (8). □

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

$$\ell(\tilde\beta_j - 1) = \frac{\ell^{-1}\sum_{t\in\tau_j} y_{t-1}\epsilon_t}{\ell^{-2}\sum_{t\in\tau_j} y_{t-1}^2}. \tag{25}$$

The result for $\tilde\beta_J$ follows from the appropriate linear combination of the limiting distributions of $n(\tilde\beta - 1)$ in (3) and of $\ell(\tilde\beta_j - 1)$ using the continuous mapping theorem.
(b) Let $Z = O_p(1)$ denote the limit of $n(\tilde\beta - 1)$ and let $Z_j = O_p(1/m)$ denote the limit of $\ell(\tilde\beta_j - 1)$; see the comments in Remark 1 following the Theorem. When $m\to\infty$ it follows that $\kappa_m \to 1$ and $\delta_m\sum_{j=1}^m Z_j = -\sum_{j=1}^m Z_j/(m-1) = O_p(1/m) = o_p(1)$, thereby yielding the stated result. □

**Proof of Theorem 2.** (a) Let $X_n(r) = S_{[nr]}/(\sigma_\epsilon\sqrt{n})$. The expansions are based on the representation

$$X_n(r) \overset{d}{=} W(r)\left(1 - \frac{1}{2}\frac{nr-[nr]}{nr}\right) + O_p(n^{-2});$$

see Phillips (1987, p.293). The numerator in (25) can be written

$$\ell^{-1}\sum_{t\in\tau_j} y_{t-1}\epsilon_t = \frac{1}{2}\left(\ell^{-1}y_{j\ell}^2 - \ell^{-1}y_{(j-1)\ell}^2 - \ell^{-1}\sum_{t\in\tau_j}(\epsilon_t^2 - \sigma_\epsilon^2) - \sigma_\epsilon^2\right).$$

It is easy to show that $\ell^{-1/2}y_{j\ell} = \sigma_\epsilon\sqrt{m}X_n(j/m)$ and $\ell^{-1/2}y_{(j-1)\ell} = \sigma_\epsilon\sqrt{m}X_n((j-1)/m)$. Now $n^{-1/2}\sum_{t=1}^n(\epsilon_t^2 - \sigma_\epsilon^2) \xrightarrow{d} \sqrt{2}\sigma_\epsilon^2\eta$ as $n\to\infty$ which can be written as

$$\frac{\sqrt{m}}{\sqrt{\ell}}\sum_{j=1}^m\sum_{t\in\tau_j}(\epsilon_t^2 - \sigma_\epsilon^2) \xrightarrow{d} \sqrt{2}\sigma_\epsilon^2\eta,$$

implying that, due to independence, in each sub-sample

$$\frac{\sqrt{m}}{\sqrt{2\ell}}\sum_{t\in\tau_j}(\epsilon_t^2 - \sigma_\epsilon^2) \xrightarrow{d} \frac{\sigma_\epsilon^2}{\sqrt{m}}\eta_j, \quad j = 1,\ldots,m,$$

where $\eta_j$ is a standard normal variate independent of $\eta_k$ $(k \neq j)$ and of $W$. Hence

$$\frac{1}{\sqrt{2\ell}} \sum_{t \in \tau_j} (\epsilon_t^2 - \sigma_\epsilon^2) \xrightarrow{d} \frac{\sigma_\epsilon^2}{m} \eta_j, \quad j = 1, \ldots, m,$$

so that

$$\begin{aligned}
\ell^{-1} \sum_{t \in \tau_j} y_{t-1}\epsilon_t &\stackrel{d}{=} \frac{\sigma_\epsilon^2 m}{2} \left( X_n \left( \frac{j}{m} \right)^2 - X_n \left( \frac{j-1}{m} \right)^2 - \frac{1}{m} \right) - \frac{\sigma_\epsilon^2 \eta_j}{m\sqrt{2\ell}} \\
&\stackrel{d}{=} \frac{\sigma_\epsilon^2 m}{2} \left( W \left( \frac{j}{m} \right)^2 - W \left( \frac{j-1}{m} \right)^2 - \frac{1}{m} \right) - \frac{\sigma_\epsilon^2 \eta_j}{m\sqrt{2\ell}} + O_p(\ell^{-1}) \\
&\stackrel{d}{=} \sigma_\epsilon^2 m \int_{(j-1)/m}^{j/m} W dW - \frac{\sigma_\epsilon^2 \eta_j}{m\sqrt{2\ell}} + O_p(\ell^{-1}).
\end{aligned}$$

Turning to the denominator of (25):

$$\begin{aligned}
\ell^{-2} \sum_{t \in \tau_j} y_{t-1}^2 = \ell^{-2} \sum_{t \in \tau_j} S_{t-1}^2 &= m^2 \int_{(j-1)/m}^{j/m} \left( \frac{1}{\sqrt{n}} S_{[nr]} \right)^2 dr \\
&= \sigma_\epsilon^2 m^2 \int_{(j-1)/m}^{j/m} X_n(r)^2 dr \\
&\stackrel{d}{=} \sigma_\epsilon^2 m^2 \int_{(j-1)/m}^{j/m} W^2 dr + O_p(\ell^{-1}).
\end{aligned}$$

Combining the numerator and denominator yields the stated result for $\ell(\tilde{\beta}_j - 1)$, and the expansion for the bias follows upon taking expectations and exploiting the fact that $\eta_j$ is independent of $W$.
(b) To determine the weights for $\tilde{\beta}_J^{opt}$, note that

$$E(\tilde{\beta}) = 1 + \frac{\mu}{n} + O(n^{-2}), \quad E(\tilde{\beta}_j) = 1 + \frac{\mu_j}{n} + O(n^{-2}), \quad j = 1, \ldots, m,$$

where $\mu$ is defined in the Theorem. From the definition of $\tilde{\beta}_J^{opt}$, taking expectations yields

$$E(\tilde{\beta}_J^{opt}) = (\kappa_m^{opt} + \delta_m^{opt}) + \frac{1}{n} \left( \kappa_m^{opt}\mu + \delta_m^{opt} \sum_{j=1}^m \mu_j \right) + O(n^{-2}).$$

In order that $E(\tilde{\beta}_J^{opt}) = 1 + O(n^{-2})$ the requirement is that: (i) $\kappa_m^{opt} + \delta_m^{opt} = 1$, and (ii) $\kappa_m^{opt}\mu + \delta_m^{opt} \sum_{j=1}^m \mu_j = 0$. Solving these two conditions simultaneously yields the stated weights. □

**Proof of Theorem 3.** (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 $\mu_X$ and $\mu_Y$ 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 evaluate 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

$$\theta_2 \int_a^b Y^2 + \frac{\lambda^2}{2} \int_0^b Y^2 = \left( \theta_1 + \frac{\lambda^2}{2} \right) \int_a^b Y^2 + \frac{\lambda^2}{2} \int_0^a Y^2.$$

Hence

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

As the parameter $\lambda$ 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 $Z$ 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

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

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

$$dZ(t) = \eta Z(t)dt + dw(t), \quad Z(0) = 0,$$

and will make use of the change of measure

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

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

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

But $\eta$ is arbitrary and so we set $\eta = 0$ to obtain

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

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

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

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

**Proof of Theorem 4.** (a) The normalised estimators are given by

$$\ell(\hat{\beta}_j - 1) = \frac{\ell^{-1}\sum_{t\in\tau_j} y_{t-1}\epsilon_t - \ell^{-3/2}\sum_{t\in\tau_j} y_{t-1}\ell^{-1/2}\sum_{t\in\tau_j}\epsilon_t}{\ell^{-2}\sum_{t\in\tau_j} y_{t-1}^2 - \left(\ell^{-3/2}\sum_{t\in\tau_j} y_{t-1}\right)^2}, \quad j = 1, \ldots, m.$$

Lemma A1 provides the asymptotics for all components except $\ell^{-1/2}\sum_{t\in\tau_j}\epsilon_t$, which satisfies

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

as $n \to \infty$. Combining this with the results in Lemma A1 yields

$$\ell(\hat{\beta}_j - 1) \Rightarrow \frac{\displaystyle\int_{(j-1)/m}^{j/m} W dW - m \int_{(j-1)/m}^{j/m} W \left[ W\left(\frac{j}{m}\right) - W\left(\frac{j-1}{m}\right) \right]}{m \left[ \displaystyle\int_{(j-1)/m}^{j/m} W^2 - m \left( \int_{(j-1)/m}^{j/m} W \right)^2 \right]}.$$

The result stated in the Theorem follows upon noting that the numerator and denominator can be written in terms of the demeaned process $W_{j,m}$. The result for $\hat{\beta}_J$ follows from the continuous mapping theorem.

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

**Proof of Theorem 5.** The proof applies the same arguments as in Theorem 2, albeit to the demeaned Wiener process $W_0$ rather than $W$, and so is omitted. □

**Proof of Theorem 6.** The proof follows along the same lines as the proofs of Theorems 1 and 4 and is therefore omitted. □