Jackknife Bias Reduction in the Presence of a Near-Unit Root

Marcus J. Chambers
University of Essex

and

Maria Kyriacou
University of Southampton

August 2016

Abstract: This paper considers the specification and performance of jackknife estimators of the autoregressive coefficient in a model with a near-unit root. The limit distributions of sub-sample estimators that are used in the construction of the jackknife estimator are derived and the joint moment generating function (MGF) of two components of these distributions is obtained and its properties are explored. The MGF can be used to derive the weights for an optimal jackknife estimator that removes fully the first-order finite sample bias from the estimator. The resulting jackknife estimator is shown to perform well in finite samples and, with a suitable choice of the number of sub-samples, is shown to reduce the overall finite sample root mean squared error as well as bias. However, the optimal jackknife weights rely on knowledge of the near-unit root parameter, which is typically unknown in practice, and so an alternative, feasible, jackknife estimator is proposed which achieves the intended bias reduction but does not rely on knowledge of this parameter. This feasible jackknife estimator is also capable of substantial bias and root mean squared error reductions in finite samples across a range of values of the near-unit root parameter and across different sample sizes.

Key words: Jackknife; bias reduction; near-unit root; moment generating function.

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

Address for correspondence: 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 jackknife has proved to be an easy-to-implement method of eliminating first-order estimation bias in a wide variety of applications in statistics and econometrics. Its genesis can be traced to Quenouille (1956) and Tukey (1958) in the case of independently and identically distributed (iid) samples while it has been adapted more recently for application in a variety of time series settings. These applications in time series include: Phillips and Yu (2005), who show how the jackknife can be used to reduce bias in the pricing of bond options in finance; Chambers (2013), who analyses the jackknife based on a variety of sub-sampling procedures in the setting of stationary autoregressive models; Chambers and Kyriacou (2013), who demonstrate that the usual jackknife construction in the time series case has to be amended when a unit root is present; Chen and Yu (2015), who show, also in the context of a unit root, that a variance-minimising jackknife can be constructed that also retains its bias reduction properties; and Kruse and Kaufmann (2015), who compare bootstrap, jackknife and indirect inference estimators in mildly explosive autoregressions, finding that the indirect inference estimator dominates in terms of root mean squared error but that the jackknife excels for bias reduction in stationary and unit root situations.

The usual motivation for a jackknife estimator relies on the existence of a Nagar-type expansion of the estimator's bias. Its construction proceeds by finding a set of weights that, when applied to a full-sample estimator and a set of sub-sample estimators, is able to eliminate the first-order term in the resulting bias expansion of the jackknife estimator. In stationary time series settings the bias expansions are common to both the full-sample and sub-sample estimators, but Chambers and Kyriacou (2013) pointed out that this property no longer holds in the case of a unit root. This is because the initial values in the sub-samples are no longer negligible in the asymptotics and have a resulting effect on the bias expansions, thereby affecting the optimal weights. Construction of a fully effective jackknife estimator relies, therefore, on knowledge of the presence (or otherwise) of a unit root.

In this paper we explore the construction of jackknife estimators that are effective in eliminating fully the first-order bias in the setting of a near-unit root. Such models have become of widespread interest in time series econometrics owing, amongst other things, to their ability to capture better the effects of sample size in the vicinity of a unit root and to explore analytically the power properties of unit root tests. We find that jackknife estimators can be constructed in the presence of a near-unit root that achieve this aim of bias reduction and, moreover, *a priori* knowledge of the near-unit root parameter is not even required for such optimal bias reduction. Jackknife estimators have the advantage of incurring only a very slight additional computational burden, unlike alternative resampling amd simulation methods such as the bootstrap and indirect inference. Furthermore they are applicable in a wide variety of estimation frameworks and work well in finite samples where the objective is bias reduction.

The paper is organised as follows. Section 2 defines the near-unit root model of interest and focuses on the limit distributions of sub-sample estimators, demonstrating that these limit distributions are sub-sample dependent. An asymptotic expansion of these limit distributions demonstrates the source of the failure of the standard jackknife weights in a near-unit root setting by showing that the bias expansion is also sub-sample dependent. In order to define a successful jackknife estimator it is necessary to compute the mean of these limit distributions and so section

3 derives the moment generating function (MGF) of two random variables that determine the limit distributions over an arbitrary sub-interval of the unit interval. Expressions for the computation of the mean of the ratio of the two random variables are derived using the MGF. Various properties of the MGF are established and it is shown that results obtained in Phillips (1987a) arise as a special case, including those that emerge as the near-unit root parameter tends to minus infinity.

Based on the results in sections 2 and 3 the optimal weights for the jackknife estimator are defined in section 4 which goes on to explore, via simulations, the performance of the estimator in finite samples. Consideration is given to the choice of the appropriate number of sub-samples to use when either bias reduction or root mean squared error (RMSE) minimisation is the objective. It is found that greatest bias reduction can be achieved using just two sub-samples while minimisation of RMSE – which, it should be stressed, is *not* the objective of the jackknife estimator – requires a larger number of sub-samples which increases with sample size.

Despite its success in achieving substantial bias reduction in finite samples, a drawback of the jackknife estimator, and an impediment to its use in practice, is the dependence of the optimal weights on the unknown near-unit root parameter. A feasible jackknife estimator is therefore proposed in section 5 and its performance is assessed in simulations which also include the median unbiased estimator (MUE) of Andrews (1993) and the indirect inference estimator (IIE) analysed by Phillips (2012) for comparison. Bootstrap alternatives were not considered owing to the results of Park (2006) which established that the bootstrap is inconsistent in the presence of a near-unit root. The simulations indicate the superior bias reduction properties of the feasible jackknife estimator although the MUE and IIE have smaller RMSEs. Section 6 contains some concluding comments, and all proofs are contained in the Appendix.

The following notation will be used throughout the paper. 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; $W(r)$ denotes a Wiener process on $C[0,1]$, the space of continuous real-valued functions on the unit interval; and $J_c(r) = \int_0^r e^{(r-s)c}dW(s)$ denotes the Ornstein-Uhlenbeck process which satisfies $dJ_c(r) = cJ_c(r)dr + dW(r)$ for some constant parameter c. Functionals of $W(r)$ and $J_c(r)$, such as $\int_0^1 J_c(r)^2 dr$, shall be denoted $\int_0^1 J_c^2$ for notational convenience where appropriate, and in stochastic integrals of the form $\int e^{cr} J_c$ it will be understood that integration is carried out with respect to r. Finally, L denotes the lag operator such that $L^j y_t = y_{t-j}$ for a random variable y_t.

2. Jackknife estimation with a near-unit root

2.1 The model and the standard jackknife estimator

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

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

$$y_t = \rho y_{t-1} + u_t, \quad t = 1, \ldots, n, \tag{1}$$

where $\rho = e^{c/n} = 1 + c/n + O(n^{-2})$ for some constant c, y_0 is an observable $O_p(1)$ random variable, and u_t is the stationary linear process

$$u_t = \delta(L)\epsilon_t = \sum_{j=0}^{\infty} \delta_j \epsilon_{t-j}, \quad t = 1, \ldots, n, \tag{2}$$

where $\epsilon_t \sim \text{iid}(0, \sigma_\epsilon^2)$, $E(\epsilon_t^4) < \infty$, $\delta(z) = \sum_{j=0}^{\infty} \delta_j z^j$, $\delta_0 = 1$ and $\sum_{j=0}^{\infty} j|\delta_j| < \infty$.

The parameter c controls the extent to which the near-unit root deviates from unity; when $c < 0$ the process is (locally) stationary whereas it is (locally) explosive when $c > 0$. The linear process specification for the innovations is consistent with u_t being a stationary ARMA(p, q) process of the form $\phi(L)u_t = \theta(L)\epsilon_t$, where $\phi(z) = \sum_{j=0}^{p} \phi_j z^j$, $\theta(z) = \sum_{j=0}^{q} \theta_j z^j$ and all roots of the equation $\phi(z) = 0$ lie outside the unit circle. In this case $\delta(z) = \theta(z)/\phi(z)$, but Assumption 1 also allows for more general forms of linear processes and is not restricted solely to the ARMA class. Under Assumption 1 u_t satisfies the functional central limit theorem

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t \Rightarrow \sigma W(r) \quad \text{as} \quad n \to \infty \tag{3}$$

on $C[0, 1]$, where $\sigma^2 = \sigma_\epsilon^2 \delta(1)^2$ denotes the long-run variance.

Equations of the form (1) have been used extensively in the literature on testing for an autoregressive unit root (corresponding to $c = 0$) and for examining the power properties of the resulting tests (by allowing c to deviate from zero). Ordinary least squares (OLS) regression on (1) yields

$$y_t = \hat{\rho} y_{t-1} + \hat{u}_t, \quad t = 1, \ldots, n, \tag{4}$$

where $\hat{u}_t$ denotes the regression residual, and it can be shown (see Phillips, 1987a) that $\hat{\rho}$ satisfies

$$n(\hat{\rho} - \rho) = \frac{\frac{1}{n}\sum_{t=1}^{n} y_{t-1}u_t}{\frac{1}{n^2}\sum_{t=1}^{n} y_{t-1}^2} \Rightarrow Z_c(\eta) = \frac{\int_0^1 J_c dW + \frac{1}{2}(1 - \eta)}{\int_0^1 J_c^2} \quad \text{as} \quad n \to \infty, \tag{5}$$

where $\eta = \sigma_u^2/\sigma^2$, $\sigma_u^2 = E(u_t^2) = \sigma_\epsilon^2 \sum_{j=0}^{\infty} \delta_j^2$ and the functional $Z_c(\eta)$ is implicitly defined. The limit distribution in (5) is skewed and the estimator suffers from significant negative bias in finite samples; see Perron (1989) for properties of the limit distribution for the case where $\sigma^2 = \sigma_u^2$ and (hence) $\eta = 1$.

To gain some idea as to the size of the bias in finite samples, Table 1 reports the results of a simulation exercise involving 100,000 replications of the process defined in (1) for sample sizes $n = \{24, 48, 96, 192\}$, local-to-unity parameter $c = \{-10, -5, -1, 0, 1\}$, $y_0 = 0$ and $u_t \sim \text{iid } N(0, 1)$. Table 1 reports the simulated bias, $\hat{\rho} - \rho$, and the mean of $n(\hat{\rho} - \rho)$, as well as the limit properties of these statistics. In the case of $\hat{\rho} - \rho$ the limit bias is zero for all values of c while for $n(\hat{\rho} - \rho)$ the reported values are the mean of the limit distribution Z_c, where

$$Z_c = Z_c(1) = \frac{\int_0^1 J_c dW}{\int_0^1 J_c^2};$$

these limit properties are obtained from calculations carried out in the next section.[1] It can be seen from Table 1 that although the bias of $\hat{\rho}$ decreases monotonically with n for all values of c, the means of the distribution of $n(\hat{\rho} - \rho)$ converge to negative values in the range -1.5812 to -1.9912 for the range of values of c considered. A detailed analysis of the bias of $\hat{\rho}$ in the vicinity of unity can be found in Phillips (2012).

The jackknife estimator offers a computationally simple method of bias reduction by combining 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, while the analysis of Chambers (2013) supports this result in the setting of stationary autoregressions, and so it is this approach that shall be followed here. The full sample of n observations is divided into m sub-samples, each of length ℓ, so that $n = m \times \ell$. The jackknife estimator is then defined by

$$\hat{\rho}_J = w_1 \hat{\rho} + w_2 \frac{1}{m} \sum_{j=1}^{m} \hat{\rho}_j, \tag{6}$$

where the weights are given by $w_1 = m/(m-1)$ and $w_2 = -1/(m-1)$. Assuming that the full-sample estimator and each sub-sample estimator satisfy a (Nagar-type) bias expansion of the form

$$E(\hat{\rho} - \rho) = \frac{a}{n} + O\left(\frac{1}{n^2}\right), \quad E(\hat{\rho}_j - \rho) = \frac{a}{\ell} + O\left(\frac{1}{\ell^2}\right),$$

it can be shown that

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

using the fact that $m/n = 1/\ell$. Under such circumstances the jackknife estimator is capable of completely eliminating the $O(1/n)$ bias term in the estimator as compared to $\hat{\rho}$. The problem with the argument above in the near-unit root setting, however, is that the sub-sample estimators do not share the same limit distribution as the full-sample estimator, which means that the expansions

[1] In fact, these values can be found in the row corresponding to $m = 1$ in Table 2.

for the bias of the sub-sample estimators are incorrect. These properties are demonstrated below, our results ultimately enabling an 'optimal' jackknife estimator to be defined.

2.2 Sub-sample properties

In order to explore the sub-sample properties let

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

denote the set of integers indexing the observations in each sub-sample. The sub-sample estimators can then be written, in view of (5), as

$$\ell\left(\hat{\rho}_j - \rho\right) = \frac{\dfrac{1}{\ell}\displaystyle\sum_{t\in\tau_j} y_{t-1}u_t}{\dfrac{1}{\ell^2}\displaystyle\sum_{t\in\tau_j} y_{t-1}^2}, \quad j = 1, \ldots, m. \tag{7}$$

Theorem 1 (below) determines the limiting properties of the quantities appearing in (7) as well as the limit distribution of $\ell(\hat{\rho}_j - \rho)$ itself.

Theorem 1. *Let $y_1, \ldots, y_n$ satisfy Assumption 1. Then, if m is fixed as $n \to \infty$ (and hence $\ell \to \infty$):*

(a) $\dfrac{1}{\ell^2}\displaystyle\sum_{t\in\tau_j} y_t^2 \Rightarrow \sigma^2 m^2 \int_{(j-1)/m}^{j/m} J_c^2;$

(b) $\dfrac{1}{\ell}\displaystyle\sum_{t\in\tau_j} y_{t-1}u_t \Rightarrow \sigma^2 m \int_{(j-1)/m}^{j/m} J_c dW + \frac{1}{2}(\sigma^2 - \sigma_u^2);$

(c) $\ell(\hat{\rho}_j - \rho) \Rightarrow Z_{c,j}(\eta) = \dfrac{\displaystyle\int_{(j-1)/m}^{j/m} J_c dW + \frac{1}{2m}(1-\eta)}{m\displaystyle\int_{(j-1)/m}^{j/m} J_c^2}, \quad (j = 1, \ldots, m),$

where the functional $Z_{c,j}(\eta)$ is implicitly defined and $\eta = \sigma_u^2/\sigma^2$.

The limit distribution in part (c) of Theorem 1 is of the same form as that of the full-sample estimator in (5) except that the integrals are over the subset $[(j-1)/m, j/m]$ of $[0,1]$ rather than the unit interval itself. Note, too, that the first component of the numerator of $Z_{c,j}(\eta)$ also has the representation

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

which follows from the Itô calculus and is demonstrated in the proof of part (b) of Theorem 1 in the Appendix. The familiar result, $\int_0^1 W dW = [W(1)^2 - 1]/2$, follows as a special case by setting $j = m = 1$ and $c = 0$.

The fact that the distributions $Z_{c,j}(\eta)$ in Theorem 1 depend on j implies that the expansions for $E(\hat{\rho}_j - \rho)$ that are used to derive the jackknife weights may not be correct under a near-unit root. When the process (1) has a near-unit root we can expect the expansions for $E(\hat{\rho}_j - \rho)$ to be of the form

$$E\left(\hat{\rho}_j - \rho\right) = \frac{\mu_{c,j}}{\ell} + O\left(\frac{1}{\ell^2}\right), \quad j = 1, \ldots, m;$$

indeed, we later justify this expansion and characterise $\mu_{c,j}$ precisely. Such expansions have been shown to hold in the unit root ($c = 0$) case as well as more generally when $c \neq 0$. For example, Phillips (1987b, Theorem 7.1) considered the Gaussian random walk (corresponding to (1) with $c = 0$, $\delta(z) = 1$, $y_0 = 0$ and u_t Gaussian) and demonstrated the validity of an asymptotic expansion for the normalised coefficient estimator; it is given by

$$n(\hat{\rho} - 1) \overset{d}{=} \frac{\int_0^1 W dW}{\int_0^1 W^2} - \frac{\xi}{\sqrt{2n} \int_0^1 W^2} + O_p\left(\frac{1}{n}\right), \tag{9}$$

where ξ is a standard normal random variable distributed independently of W. Taking expectations in (9), using the independence of ξ and W, and noting that the expected value of the leading term is -1.7814 (see, for example, Table 7.1 of Tanaka, 1996), the bias satisfies

$$E(\hat{\rho} - 1) = -\frac{1.7814}{n} + o\left(\frac{1}{n}\right); \tag{10}$$

see, also, Phillips (2012, 2014). In the more general setting of the model in Assumption 1 Perron (1996) established that

$$n(\hat{\rho} - \rho) \overset{d}{=} \frac{\int_0^1 J_c dW + \frac{1}{2}(1 - \eta) + \frac{y_0}{\sigma\sqrt{n}} \int_0^1 e^{cr} dW - \frac{v_f}{2\sigma^2\sqrt{n}}\xi}{\int_0^1 J_c^2 + 2\frac{y_0}{\sigma\sqrt{n}} \int_0^1 e^{cr} J_c} + O_p\left(\frac{1}{n}\right), \tag{11}$$

where $v_f^2 = 2\pi f_{u^2}(0)$ and $f_{u^2}(0)$ denotes the spectral density of $u_t^2 - \sigma_u^2$ at the origin, while more recently Mikusheva (2015) has derived a second-order expansion of the t-statistic in this model. The following result extends the type of expansion in (11) to the sub-sample estimators.

Theorem 2. *Let $y_1, \ldots, y_n$ satisfy Assumption 1 with $u_t \sim$ iid $N(0, \sigma_u^2)$. Then:*

$$\ell(\hat{\rho}_j - \rho) \overset{d}{=} \frac{\int_{(j-1)/m}^{j/m} J_c dW + \frac{1}{2m}(1 - \eta) + \frac{2y_0}{\sigma\sqrt{\ell}}\xi_{1j} - \frac{\xi_{2j}}{\sigma^2 m\sqrt{\ell}}}{m \int_{(j-1)/m}^{j/m} J_c^2 + \frac{2\sqrt{m}y_0}{\sigma\sqrt{\ell}}\sigma \int_{(j-1)/m}^{j/m} e^{cr} J_c} + O_p\left(\frac{1}{\ell}\right),$$

where

$$\xi_{1j} = e^{cj/m} J_c\left(\frac{j}{m}\right) - e^{c(j-1)/m} J_c\left(\frac{j-1}{m}\right) - \frac{2c}{\sqrt{m}} \int_{(j-1)/m}^{j/m} e^{cr} J_c \sim N(0, s^2),$$

$$s^2 = \frac{(\sqrt{m}-1)^2}{2cm}\left(e^{cj/m} - e^{c(j-1)/m}\right)^2 + \frac{(m-2\sqrt{m}-2)}{2cm}\left(e^{cj/m} - e^{c(j-1)/m}\right) + \frac{2(1+\sqrt{m})}{m^2} e^{cj/m},$$

and $\xi_{2j} \sim N(0, v_f^2)$.

The form of the expansion for $\ell(\hat{\rho}_j - \rho)$ in Theorem 2 is similar to that for the full-sample estimator but depends on m and j. Use of these expansions to derive expressions for the biases of $\hat{\rho}$ and $\hat{\rho}_j$ would be complicated due to the dependence on y_0. We therefore take $y_0 = 0$ which results in the following expectations:

$$E(\hat{\rho} - \rho) = \frac{E(Z_c(\eta))}{n} + O\left(\frac{1}{n^2}\right), \quad E(\hat{\rho}_j - \rho) = \frac{E(Z_{c,j}(\eta))}{\ell} + O\left(\frac{1}{\ell^2}\right),$$

these results utilising the independence of the normally distributed random variables (ξ_{1j} and ξ_{2j}) and the Wiener process W. The next section provides the form of the moment generating function that enables expectations of the functionals $Z_c(\eta)$ and $Z_{c,j}(\eta)$ to be computed.

3. A moment generating function and its properties

The following result provides the joint moment generating function (MGF) of two relevant functionals of J_c defined over a subinterval $[a, b]$ of $[0, b]$ where $0 \leq a < b$. Although our focus is on sub-intervals of $[0, 1]$ we leave b unconstrained for greater generality than is required for our specific purposes because the results may have more widespread use beyond our particular application.

Theorem 3. *Let* $N_c = \int_a^b J_c(r) dW(r)$ *and* $D_c = \int_a^b J_c(r)^2 dr$*, where* $J_c(r)$ *is an Ornstein-Uhlenbeck process on* $r \in [0, b]$ *with parameter* c*, and* $0 \leq a < b$*. Then:*

(a) The joint MGF of N_c *and* D_c *is given by*

$$M_c(\theta_1, \theta_2) = E\exp(\theta_1 N_c + \theta_2 D_c) = \exp\left(-\frac{(\theta_1 + c)}{2}(b-a)\right) H_c(\theta_1, \theta_2)^{-1/2},$$

where, defining $\lambda = (c^2 + 2c\theta_1 - 2\theta_2)^{1/2}$ *and* $v^2 = (e^{2ac} - 1)/(2c)$,

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

(b) Let

$$g(\theta_2) = \cosh\left((b-a)(c^2 + 2\theta_2)^{1/2}\right) - \left(c - 2v^2\theta_2\right) \frac{\sinh\left((b-a)(c^2 + 2\theta_2)^{1/2}\right)}{(c^2 + 2\theta_2)^{1/2}}.$$

Then the expectation of N_c/D_c is given by

$$E\left(\frac{N_c}{D_c}\right) = \int_0^\infty \left.\frac{\partial M_c(\theta_1, -\theta_2)}{\partial \theta_1}\right|_{\theta_1=0} d\theta_2 = I_1(a,b) + I_2(a,b) + I_3(a,b) + I_4(a,b),$$

where

$$I_1(a,b) = -\frac{(b-a)}{2}\exp\left(-\frac{c(b-a)}{2}\right)\int_0^\infty \frac{1}{g(\theta_2)^{1/2}} d\theta_2,$$

$$I_2(a,b) = -\frac{(c(b-a)-1)}{2}\exp\left(-\frac{c(b-a)}{2}\right)\int_0^\infty \frac{\sinh\left((b-a)(c^2+2\theta_2)^{1/2}\right)}{(c^2+2\theta_2)^{1/2} g(\theta_2)^{3/2}} d\theta_2,$$

$$I_3(a,b) = -\frac{c}{2}\exp\left(-\frac{c(b-a)}{2}\right)\int_0^\infty \left(c - 2v^2\theta_2\right)\frac{\sinh\left((b-a)(c^2+2\theta_2)^{1/2}\right)}{(c^2+2\theta_2)^{3/2} g(\theta_2)^{3/2}} d\theta_2,$$

$$I_4(a,b) = \frac{c(b-a)}{2}\exp\left(-\frac{c(b-a)}{2}\right)\int_0^\infty \left(c - 2v^2\theta_2\right)\frac{\cosh\left((b-a)(c^2+2\theta_2)^{1/2}\right)}{(c^2+2\theta_2) g(\theta_2)^{3/2}} d\theta_2.$$

The MGF for the two functionals in part (a) of Theorem 3 has potential applications in a wide range of sub-sampling problems with near-unit root processes. The individual MGFs for N_c and D_c, denoted $M_{N_c}(\theta_1) = M_c(\theta_1, 0)$ and $M_{D_c}(\theta_2) = M_c(0, \theta_2)$ respectively, follow straightforwardly and are given by

$$M_{N_c}(\theta_1) = \exp\left(-\frac{(\theta_1+c)}{2}(b-a)\right)\left[\cosh((b-a)\lambda_1) - \frac{1}{\lambda_1}\left(\theta_1 + c + v^2\theta_1^2\right)\sinh((b-a)\lambda_1)\right]^{-1/2}, \tag{12}$$

$$M_{D_c}(\theta_2) = \exp\left(-\frac{c}{2}(b-a)\right)\left[\cosh\left((b-a)\lambda_2\right) - \frac{1}{\lambda_2}\left(c + 2v^2\theta_2\right)\sinh\left((b-a)\lambda_2\right)\right]^{-1/2}, \tag{13}$$

where $\lambda_1 = (c^2 + 2c\theta_1)^{1/2}$ and $\lambda_2 = (c^2 - 2\theta_2)^{1/2}$. Some special cases then result:

Example 1. When $[a,b] = [0,1]$ it follows that $b - a = 1$ and $v^2 = 0$. In this case we find that

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

$$M_{N_c}(\theta_1) = \exp\left(-\frac{(\theta_1+c)}{2}\right)\left[\cosh(\lambda_1) - \frac{1}{\lambda_1}(\theta_1 + c)\sinh(\lambda_1)\right]^{-1/2},$$

$$M_{D_c}(\theta_2) = \exp\left(-\frac{c}{2}\right)\left[\cosh(\lambda_2) - \frac{c}{\lambda_2}\sinh(\lambda_2)\right]^{-1/2},$$

while taking the limit as $c \to 0$ yields

$$M_{N_0}(\theta_1) = e^{-\theta_1/2}(1-\theta_1)^{-1/2}, \quad M_{D_0}(\theta_2) = \left[\cosh(\lambda_0)\right]^{-1/2},$$

where $\lambda_0 = \sqrt{-2\theta_2}$; in this case the joint MGF is

$$M_0(\theta_1, \theta_2) = \exp\left(-\frac{\theta_1}{2}\right)\left[\cosh(\lambda_0) - \frac{\theta_1}{\lambda_0}\sinh(\lambda_0)\right]^{-1/2},$$

a result that goes back to White (1958).

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

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

where $v_{j-1}^2 = (\exp(2(j-1)c/m) - 1)/(2c)$. Taking the limit as $c \to 0$ results in

$$M_0(\theta_1, \theta_2) = \exp\left(-\frac{\theta_1}{2m}\right)\left[\cosh\left(\frac{\lambda_0}{m}\right) - \frac{1}{\lambda_0}\left[\theta_1 + \frac{(j-1)}{m}\left(\theta_1^2 + 2\theta_2\right)\right]\sinh\left(\frac{\lambda_0}{m}\right)\right]^{-1/2},$$

a result that has been used in Chambers and Kyriacou (2013). The individual MGFs for N_c and D_c in this case are given by

$$M_{N_c}(\theta_1) = \exp\left(-\frac{(\theta_1 + c)}{2m}\right)\left[\cosh\left(\frac{\lambda_1}{m}\right) - \frac{1}{\lambda_1}\left(\theta_1 + c + v_{j-1}^2\theta_1^2\right)\sinh\left(\frac{\lambda_1}{m}\right)\right]^{-1/2},$$

$$M_{D_c}(\theta_2) = \exp\left(-\frac{c}{2m}\right)\left[\cosh\left(\frac{\lambda_2}{m}\right) - \frac{1}{\lambda_2}\left(c + 2v_{j-1}^2\theta_2\right)\sinh\left(\frac{\lambda_2}{m}\right)\right]^{-1/2},$$

respectively, while the limits as $c \to 0$ are

$$M_{N_0}(\theta_1) = \exp\left(-\frac{c}{2m}\right)\left(1 - \frac{\theta_1}{m} - (j-1)\frac{\theta_1^2}{m^2}\right)^{-1/2},$$

$$M_{D_0}(\theta_2) = \exp\left(-\frac{c}{2m}\right)\left[\cosh\left(\frac{\lambda_0}{m}\right) - \frac{2(j-1)\theta_2}{m\lambda_0}\sinh\left(\frac{\lambda_0}{m}\right)\right]^{-1/2}.$$

Another potential use of the joint MGF in part (a) of Theorem 3 is in the computation of the cumulative and probability density functions of the distributions $m^{-1}Z_{c,j}(\eta)$. For example, the probability density function of $m^{-1}Z_{c,j}(1)$ is given by (with $i^2 = -1$)

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

see, for example, Perron (1991, p.221) who performs this calculation for the distribution Z_c, while Abadir (1993) derives a representation for the density function of Z_c in terms of a parabolic cylinder function.

It is also possible to use the above results to explore the relationship between the sub-sample distributions and the full-sample distribution. For example, it is possible to show that $M_{N_{c/m}}(\theta_1/m)$

on $[0,1]$ is equal to $M_{N_c}(\theta_1)$ for $j=1$ in the sub-samples, while $M_{D_{c/m}}(\theta_2/m^2)$ on $[0,1]$ is equal to $M_{D_c}(\theta_2)$ for $j=1$ in the sub-samples; an implication of this is that

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

Furthermore, this implies that the limit distribution of the first sub-sample estimator, $\ell(\hat{\rho}_1 - \rho)$, when $\rho = e^{c/n} = e^{c/m\ell}$, is the same as that of the full-sample estimator, $n(\hat{\rho}-\rho)$, when $\rho = e^{c/mn}$.

The result in part (b) of Theorem 3 is obtained by differentiating the MGF and constructing the appropriate integrals. When $c=0$ the usual (full-sample) result, where $a=0$ and $b=1$, can be obtained as a special case. Noting that $v^2=0$ in this case, and making the substitution $w=(c^2+2\theta_2)^{1/2}$, results in

$$I_1(0,1) = -\frac{1}{2}\int_0^\infty \frac{w}{\cosh(w)^{1/2}}dw, \quad I_2(0,1) = \frac{1}{2}\int_0^\infty \frac{\sinh(w)}{\cosh(w)^{3/2}}dw, \quad I_3(0,1)=0, \quad I_4(0,1)=0;$$

see, for example, Gonzalo and Pitarakis (1998, Lemma 3.1). In the case of non-overlapping sub-samples, where $a=(j-1)/m$ and $b=j/m$, making the substitution $w=(c^2+2\theta_2)^{1/2}/m$ and allowing $c\neq 0$ results in

$$I_{1,j} = -\frac{m}{2}\exp\left(-\frac{c}{2m}\right)\int_{c/m}^\infty \frac{w}{g(w)^{1/2}}dw,$$

$$I_{2,j} = \frac{(m-c)}{2}\exp\left(-\frac{c}{2m}\right)\int_{c/m}^\infty \frac{\sinh(w)}{g(w)^{3/2}}dw,$$

$$I_{3,j} = -\frac{c}{2m}\exp\left(-\frac{c}{2m}\right)\int_{c/m}^\infty \left[c\left(1+cv_{j-1}^2\right) - v_{j-1}^2 m^2 w^2\right]\frac{\sinh(w)}{w^2 g(w)^{3/2}}dw,$$

$$I_{4,j} = \frac{c}{2m}\exp\left(-\frac{c}{2m}\right)\int_{c/m}^\infty \left[c\left(1+cv_{j-1}^2\right) - v_{j-1}^2 m^2 w^2\right]\frac{\cosh(w)}{w g(w)^{3/2}}dw,$$

where $g(w) = \cosh(w) + mwv_{j-1}^2\sinh(w) - c(1+cv_{j-1}^2)\sinh(w)/(mw)$ and $I_{k,j} = I_k((j-1)/m, j/m)$ $(k=1,\ldots,4)$.

The sub-sample results with a near-unit root can be related to the full-sample results of Phillips (1987a). For example, the MGF in Theorem 3 has the equivalent representation

$$\begin{aligned} M_c(\theta_1,\theta_2) \quad = \quad & \left\{\frac{1}{2}\exp\left((\theta_1+c)(b-a)\right)\lambda^{-1}\left[(2\lambda+\delta)(1+\delta v^2)\exp(-z)\right.\right. \\ & \left.\left. - \left(2\lambda\delta v^2 + \delta(1+\delta v^2)\right)\exp(z)\right]\right\} \end{aligned} \qquad (14)$$

where λ and v^2 are defined in the Theorem, $z=(b-a)\lambda$ and $\delta = \theta_1 + c - \lambda$. When $a=0$, $b=1$ it follows that $v^2 = 0$ and the above expression nests the MGF in Phillips (1987a) i.e.

$$\begin{aligned} M_c(\theta_1,\theta_2) &= \left\{ \frac{1}{2}\exp(\theta_1+c)(c^2+2\theta_1-2\theta_2)^{-1/2} \right. \\ &\quad \times \left[\left(\theta_1+c+(c^2+2c\theta_1-2\theta_2)^{1/2}\right)\ \exp\left(-(c^2+2c\theta_1-\theta_2)^{1/2}\right) \right. \\ &\quad \left.\left. - \left(\theta_1+c-(c^2+2c\theta_1-2\theta_2)^{1/2}\right)\exp\left((c^2+2c\theta_1-2\theta_2)^{1/2}\right)\right] \right\}; \end{aligned}$$

this follows straightforwardly from (14). It is also of interest to examine what happens when the local-to-unity parameter $c \to -\infty$, as in Phillips (1987a) and other recent work on autoregression e.g. Phillips (2012). We present the results in Theorem 4 below.

Theorem 4. *Let $J_c(r)$ denote an Ornstein-Uhlenbeck process on $r \in [0, b]$ with parameter c, and let $0 \leq a < b$. Furthermore, define the functional*

$$K(c) = g(c)^{1/2}\left(\int_a^b J_c^2\right)^{-1}\left(\int_a^b J_c dW + \frac{1}{2}(1-\eta)\right),$$

where $\eta = \sigma_u^2/\sigma^2$ and

$$g(c) = E\left(\int_a^b J_c^2\right) = \frac{1}{4c^2}\left(\exp(2bc)-\exp(2ac)\right) + \left(\frac{1}{-2c}\right)(b-a).$$

Then, as $c \to -\infty$:

(a) $(-2c)^{1/2}\int_a^b J_c dW \Rightarrow N(0,(b-a))$;

(b) $(-2c)\int_a^b J_c^2 \xrightarrow{p} (b-a)$;

(c) $K(c) \Rightarrow N(0,1)$ *if* $\sigma_u^2 = \sigma^2$ *(and hence* $\eta = 1$*) and diverges otherwise.*

The functional $K(c)$ in Theorem 4 represents the limit distribution of the normalised estimator $g(c)^{1/2}\ell(\hat{\rho}_{a,b}-\rho)$, where ℓ denotes the number of observations in the sub-sample $[bn-\ell+1, bn]$ (so that $a = b-(1/m)$ in this case) and $\hat{\rho}_{a,b}$ is the corresponding estimator. However, as pointed out by Phillips (1987a), the sequential limits (large sample for fixed c, followed by $c \to -\infty$) are only indicative of the results one might expect in the stationary case and do not constitute a rigorous demonstration. The results in Theorem 4 also encompass the related results in Phillips (1987a) obtained when $a = 0$ and $b = 1$.

4. An 'optimal' jackknife estimator

An 'optimal' jackknife estimator can be defined based on an application of the results in Theorem 3 by noting that the limit distribution of $\ell(\hat{\rho}_j-\rho)$ in Theorem 1 can be written in terms of the quantities N_c and D_c in Theorem 3. Recalling that $\ell(\hat{\rho}_j-\rho) \Rightarrow Z_{c,j}(\eta)$ we find that $Z_{c,j}(\eta)$

has the representation

$$Z_{c,j}(\eta) \stackrel{d}{=} \frac{N_c + \frac{1}{2m}(1-\eta)}{mD_c} = Z_{c,j} + \frac{(1-\eta)}{2m^2 D_c},$$

where $Z_{c,j} = Z_{c,j}(1)$ for notational convenience. It follows that the expectations of the limit distribution take the form

$$E\left(Z_{c,j}(\eta)\right) = \mu_{c,j} + \frac{1}{2m^2}(1-\eta)E\left(\frac{1}{D_c}\right)$$

where

$$\mu_{c,j} = E\left(\frac{N_c}{mD_c}\right) = E(Z_{c,j}). \tag{15}$$

Table 2 contains the values of the expectations $\mu_{c,j}$ for a range of values of $m = \{1, 2, 3, 4, 6, 8, 12\}$ and $c = \{-50, -20, -10, -5, -1, 0, 1\}$. The entries for $m = 1$ correspond to $\mu_c = E(Z_c)$ and numerical integration routines were used to evaluate the integrals.[2] For a given combination of j and m it can be seen that the expectations increase as c increases, while for given c and j the expectations increase with m. 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. The values of the sub-sample expectations when $c = 0$ are seen from Table 2 to be independent of m and to increase with j. Note that $\mu_{0,1} = -1.7814$ corresponds to the expected value of the limit distribution of the full-sample estimator $\hat{\rho}$ under a unit root; see, for example, (10) and the associated commentary. The values of $\mu_{0,j}$ can be used to define jackknife weights under a unit root for different values of m; see, for example, Chambers and Kyriacou (2013). More generally the values of $\mu_{c,j}$ can be used to define optimal weights for the jackknife estimator that achieve the aim of first-order bias removal under a near-unit root. The result is presented in Theorem 5.

Theorem 5. *Let $\mu_c = E\left(Z_c\right)$ and $\bar{\mu}_c = \mu_c - \sum_{j=1}^m \mu_{c,j}$, where the $\mu_{c,j}$ are defined in (15). Then, under Assumption 1 and assuming that $\sigma^2 = \sigma_u^2$ (so that $\eta = 1$), an 'optimal' jackknife estimator is given by*

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

where $w_{1,c}^* = -\sum_{j=1}^m \mu_{c,j}/\bar{\mu}_c$ and $w_{2,c}^* = \mu_c/\bar{\mu}_c$.

Theorem 5 shows the 'optimal' weights for the jackknife estimator when the process (1) has a near-unit root. The values of $\mu_{c,j}$ in Table 2 can be utilised in Theorem 5 to derive these 'optimal' weights for the jackknife estimator; these are reported in Table 3 for a range of values of m and c. It can be seen from Table 3 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, and that they

[2] Romberg's method was used to compute the integrals and the computations were carried out using Gauss 14.

increase with c for given m. The optimal weights also converge towards the standard weights as c becomes more negative – this could presumably be demonstrated analytically using the properties of the MGF in constructing the $\mu_{c,j}$ by examining the appropriate limits as $c \to -\infty$, although we do not pursue such an investigation here.

The effect of the variations in weights reported in Table 3 on the finite sample properties of the jackknife estimator has been explored in simulations, and the results are presented in Tables 4 and 5. The entries in Table 4 report the bias of $\hat{\rho}$, $\hat{\rho}_J$ and $\hat{\rho}_J^*$ obtained from 100,000 replications of the model in Assumption 1 with $u_t \sim$ iid $N(0,1)$ and $y_0 = 0$, while Table 5 contains the corresponding RMSE values. Four sample sizes are considered, these being n =24, 48, 96 and 192, which allow for a range of values of m to be compared. 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 subscripts to the bias and RMSE values in the tables.

In terms of bias it can be seen from Table 4 that the jackknife estimator $\hat{\rho}_{J,B}$ is capable of producing substantial bias reduction over $\hat{\rho}$, for all values of c and for all four sample sizes. The bias-minimising value is seen to be $m = 2$ in all cases except when $c = 1$ and $n = 96$ in which case $m = 3$. However, using the 'optimal' weights in the jackknife estimator results in even further bias reduction over $\hat{\rho}_{J,B}$ across all values of c and all sample sizes; in all cases the bias-minimising value is $m = 2$. It can also be seen from Table 4 that the jackknife estimators based on the RMSE-minimising values of m also provide substantial bias reductions over $\hat{\rho}$ although not as great as $\hat{\rho}_{J,B}$ and $\hat{\rho}_{J,B}^*$. The RMSE-minimising values of m are also larger than the bias-minimising values, ranging from $m = 4$ through to $m = 12$. A similar finding is reported in Chambers (2013) in the case of stationary autoregressions.

Although in itself important, bias is not the only feature of a distribution that is of interest, and hence the RMSE values in Table 5 should also be taken into account when assessing the performance of the estimators. The substantial bias reductions obtained with the bias-minimising values of m are seen to come at the cost of a larger variance that ultimately feeds through into a larger RMSE compared with $\hat{\rho}$. This can be offset, however, by using the RMSE-minimising values of m that, despite having a larger bias than $\hat{\rho}_{J,B}$ and $\hat{\rho}_{J,B}^*$, are nevertheless able to reduce the variance sufficiently to result in a smaller RMSE than $\hat{\rho}$; this is true for both $\hat{\rho}_{J,R}$ and $\hat{\rho}_{J,R}^*$.

5. A feasible jackknife estimator

The analysis of previous sections has 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 near-unit root. A drawback of this approach, however, is that it requires knowledge of the near-unit root (in particular the parameter c). An alternative, feasible, approach that does not require *a priori* knowledge of the near-unit root parameter is examined below.

The source of the failure of the usual jackknife in the near-unit root setting is that the initial (or pre-sample) value in the sub-samples is no longer $O_p(1)$ and is therefore not eliminated in the asymptotics. To see this note that the observations in sub-sample j satisfy

$$y_{(j-1)\ell+k} = \rho^k y_{(j-1)\ell} + \sum_{i=1}^{k} \rho^{k-i} u_{(j-1)\ell+i}, \quad k = 1, \ldots, \ell. \tag{16}$$

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

$$y_t = \tilde{\alpha} + \tilde{\rho} y_{t-1} + \tilde{u}_t, \quad t = 1, \ldots, n, \tag{17}$$

where $\tilde{\alpha}$ and $\tilde{\rho}$ are the least squares estimators of the intercept and slope, respectively, and $\tilde{u}_t$ denotes the regression residual; see, for example, Davidson (2000, p.351) for a clear demonstration as to how the presence of an intercept eliminates the dependence of the estimators on the pre-sample value.

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

$$n(\tilde{\rho} - \rho) \Rightarrow Z_c^\mu(\eta) = \frac{\int_0^1 J_c^\mu dW + \frac{1}{2}(1-\eta)}{\int_0^1 (J_c^\mu)^2} \quad \text{as} \quad n \to \infty, \tag{18}$$

where $J_c^\mu(r) = J_c(r) - \int_0^1 J_c(s)ds$ is a demeaned Ornstein-Uhlenbeck process and $Z_c^\mu(\eta)$ is implicitly defined. The standard jackknife estimator, based on (6), is given by

$$\tilde{\rho}_J = w_1 \tilde{\rho} + w_2 \frac{1}{m} \sum_{j=1}^{m} \tilde{\rho}_j, \tag{19}$$

where w_1 and w_2 are defined following (6) and the $\tilde{\rho}_j$ $(j = 1, \ldots, m)$ are the sub-sample estimators obtained in regressions that include an intercept. Theorem 6 provides the limit properties of the normalised sums of y_t as well as $\ell(\tilde{\rho}_j - \rho)$.

Theorem 6. *Let $y_1, \ldots, y_n$ satisfy Assumption 1. Then, if m is fixed as $n \to \infty$ (and hence $\ell \to \infty$):*

(a) $\displaystyle \frac{1}{\ell^{3/2}} \sum_{t \in \tau_j} y_t \Rightarrow \sigma m^{3/2} \int_{(j-1)/m}^{j/m} J_c;$

(b) $\displaystyle \ell(\tilde{\rho}_j - \rho) \Rightarrow Z_{c,j}^\mu(\eta) = \frac{\int_{(j-1)/m}^{j/m} J_{c,j}^\mu dW + \frac{1}{2m}(1-\eta)}{m \int_{(j-1)/m}^{j/m} (J_{c,j}^\mu)^2}, \quad (j = 1, \ldots, m),$

where the functional $Z^{\mu}_{c,j}(\eta)$ *is implicitly defined and*

$$J^{\mu}_{c,j}(r) = J_c(r) - m\int_{(j-1)/m}^{j/m} J_c(s)ds, \quad (j = 1, \ldots, m).$$

Part (a) of Theorem 6 provides an additional convergence result that enables the limit properties of all terms in the expression for the normalised estimator to be determined. The limit distributions of the sub-sample estimators in part (b) of Theorem 6 are expressed in terms of the demeaned Ornstein-Uhlenbeck processes $J^{\mu}_{c,j}$. Although regression with an intercept eliminates the effects of pre-sub-sample values, the effect on the limit distributions is to actually increase the negative bias. In fact, the mean of the distribution $Z^{\mu}_0(1)$ in (18) is equal to -5.379; see, for example, Table 7.2 of Tanaka (1996). This compares with a mean value of -1.7814 for $Z_0(1)$.

Tables 6 and 7 report the bias and RMSE of the estimators $\tilde{\rho}$ and $\tilde{\rho}_J$ obtained from 100,000 replications of the model with uncorrelated $N(0, 1)$ disturbances. 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 $\tilde{\rho}^*_J = w^*_{1,c}\tilde{\rho} + (w^*_{2,c}/m)\sum_{j=1}^{m}\tilde{\rho}_j$, where $w^*_{1,c}$ and $w^*_{2,c}$ are defined in Theorem 5. 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'. As before the subscripts 'B' and 'R' refer to the jackknife estimators using the bias-minimising and RMSE-minimising values of m, respectively. In addition, to provide a further source of comparison, results obtained using the MUE of Andrews (1993), denoted $\tilde{\rho}_{MU}$, and the IIE considered by Phillips (2012), denoted $\tilde{\rho}_{II}$ (based on 10,000 simulated samples), are also included in the tables.

From Table 6 it can be seen that, not surprisingly, the estimator $\tilde{\rho}$ is more biased than $\hat{\rho}$ by a magnitude of approximately three. Both estimators $\tilde{\rho}_{J,B}$ and $\tilde{\rho}_{J,R}$ reduce the bias dramatically compared to $\tilde{\rho}$ itself, the former achieving particularly large bias reductions. The bias of $\tilde{\rho}_{J,B}$ is typically less than that of $\tilde{\rho}^*_{J,B}$ although not uniformly so, any exceptions occurring only at the smallest sample size ($n = 24$). The bias reduction achieved by the jackknife estimators also exceeds that achieved by the MUE and IIE in the vast majority of cases although it has to be acknowledged that these estimators are not designed with the explicit aim of bias reduction. In terms of RMSE Table 7 shows that, as in the case with no intercept, the bias-minimising jackknife estimators achieve their bias reduction at the cost of a higher variance that more than outweighs the reduction in bias and leads to a larger RMSE than $\tilde{\rho}$. However, reductions in RMSE are possible via an appropriate choice of m, with $\tilde{\rho}_{J,R}$ achieving the smallest RMSE of all the jackknife estimators. The RMSEs of the MUE and IIE do, however, tend to dominate, this being achieved by a much smaller variance of these estimators compared to the jackknife which outweigh their larger biases, a finding that is in accordance with Kruse and Kaufmann (2015).

6. Conclusions

This paper has analysed the specification and performance of jackknife estimators of the autoregressive coefficient in a model with a near-unit root. The limit distributions of sub-sample estimators that are used in the construction of the jackknife estimator are derived and the joint

MGF of two components of that distribution is obtained and its properties explored. The MGF can be used to derive the weights for an optimal jackknife estimator that removes fully the first-order finite sample bias from the OLS estimator. The resulting jackknife estimator is shown to perform well in finite samples and, with a suitable choice of the number of sub-samples, is shown to reduce the overall finite sample RMSE as well as bias. However, the optimal jackknife weights rely on knowledge of the near-unit root parameter, which is typically unknown in practice, and so an alternative jackknife is proposed, which is based on regression including an intercept and uses the standard weights appropriate in stationary settings. This feasible jackknife estimator is also capable of substantial bias and RMSE reductions in finite samples across a range of values of the near-unit root parameter and for different sample sizes.

The results in this paper could be utilised and extended in a number of directions. An obvious application would be in the use of jackknife estimators as the basis for developing unit root test statistics, the local-to-unity framework being particularly well suited to the analysis of the power functions of such tests. It would also be possible to develop, fully, a variance-minimising jackknife estimator along the lines of Chen and Yu (2015) who derived analytic results for $c = 0$ and $m = 2$ or 3. Extending their approach to arbitrary c and m appears to be feasible but would require care in the extensive and detailed derivations required. Applications of jackknife methods in multivariate time series settings are also possible, a recent example being Chambers (2015) in the case of a cointegrated system, but other possibilities can be foreseen.

Appendix

Proof of Theorem 1. The proofs of parts (a) and (b) rely on the solution to the stochastic difference equation generating y_t, which is given by

$$y_t = \sum_{j=1}^{t} e^{c(t-j)/n} u_j + e^{ct/n} y_0. \tag{20}$$

The normalised partial sums of u_t, $S_t = \sum_{j=1}^{t} u_j$, are also important, as is the functional

$$X_n(r) = \frac{1}{\sqrt{n}} S_{[nr]} = \frac{1}{\sqrt{n}} S_{j-1}, \quad \frac{j-1}{n} \leq r < \frac{j}{n}. \tag{21}$$

Under the conditions on u_t it follows that $X_n(r) \Rightarrow \sigma W(r)$ as $n \to \infty$. Taking each part in turn:

(a) In view of (20) and (21) the object of interest can be written

$$
\begin{aligned}
\frac{1}{\ell^2}\sum_{t\in\tau_j} y_t^2 &= \frac{1}{\ell^2}\sum_{t\in\tau_j}\left\{\sum_{j=1}^{t} e^{c(t-j)/n}u_j + e^{ct/n}y_0\right\}^2 \\
&= \frac{1}{\ell^2}\sum_{t\in\tau_j}\left\{\left(\sum_{j=1}^{t} e^{c(t-j)/n}u_j\right)^2 + 2e^{ct/n}y_0\sum_{j=1}^{t} e^{c(t-j)/n}u_j + e^{2ct/n}y_0^2\right\} \\
&= \frac{n^2}{\ell^2}\sum_{t\in\tau_j}\int_{(t-1)/n}^{t/n}\left(\sum_{j=1}^{t} e^{c(t-j)/n}\int_{(j-1)/n}^{j/n} dX_n(s)\right)^2 dr + o_p(1) \\
&= m^2\int_{(j-1)/m}^{j/m}\left(\int_0^r e^{c(r-s)}dX_n(s)\right)^2 dr + o_p(1) \\
&\Rightarrow \sigma^2 m^2\int_{(j-1)/m}^{j/m} J_c^2
\end{aligned}
$$

where, in the penultimate line, we note that $j\ell/n = j/m$ and $(j-1)\ell/n = (j-1)/m$ to give the limits of the outer integral.
(b) Squaring the difference equation for y_t, summing over $t \in \tau_j$ and noting that $e^{2c/n} = 1 + (2c/n) + O(n^{-2})$ we obtain

$$
\begin{aligned}
\frac{1}{\ell}\sum_{t\in\tau_j} y_t^2 &= \frac{e^{2c/n}}{\ell}\sum_{t\in\tau_j} y_{t-1}^2 + \frac{2e^{c/n}}{\ell}\sum_{t\in\tau_j} y_{t-1}u_t + \frac{1}{\ell}\sum_{t\in\tau_j} u_t^2 \\
&= \frac{1}{\ell}\sum_{t\in\tau_j} y_{t-1}^2 + \frac{2c}{n\ell}\sum_{t\in\tau_j} y_{t-1}^2 + \frac{2}{\ell}\sum_{t\in\tau_j} y_{t-1}u_t + \frac{1}{\ell}\sum_{t\in\tau_j} u_t^2 + o_p(1).
\end{aligned}
$$

Solving for the quantity of interest yields

$$
\begin{aligned}
\frac{1}{\ell}\sum_{t\in\tau_j} y_{t-1}u_t &= \frac{1}{2}\left\{\frac{1}{\ell}\sum_{t\in\tau_j} y_t^2 - \frac{1}{\ell}\sum_{t\in\tau_j} y_{t-1}^2 - \frac{2c}{n\ell}\sum_{t\in\tau_j} y_{t-1}^2 - \frac{1}{\ell}\sum_{t\in\tau_j} u_t^2\right\} + o_p(1) \\
&= \frac{1}{2}\left\{\left(\frac{1}{\sqrt{\ell}}y_{jl}\right)^2 - \left(\frac{1}{\sqrt{\ell}}y_{(j-1)l}\right)^2 - \frac{2c}{m\ell^2}\sum_{t\in\tau_j} y_{t-1}^2 - \frac{1}{\ell}\sum_{t\in\tau_j} u_t^2\right\} + o_p(1).
\end{aligned}
$$

Now, as $n \to \infty$,

$$
\frac{1}{\sqrt{\ell}}y_{jl} = \frac{\sqrt{m}}{\sqrt{n}}y_{jl} \Rightarrow \sigma\sqrt{m}J_c\left(\frac{j}{m}\right), \quad \frac{1}{\sqrt{\ell}}y_{(j-1)l} = \frac{\sqrt{m}}{\sqrt{n}}y_{(j-1)l} \Rightarrow \sigma\sqrt{m}J_c\left(\frac{j-1}{m}\right),
$$

noting that $j\ell = (j/m)n$ and $(j-1)\ell = ((j-1)/m)n$. It follows that

$$
\begin{aligned}
\frac{1}{\ell}\sum_{t\in\tau_j} y_{t-1}u_t &\Rightarrow \frac{1}{2}\left\{\sigma^2 mJ_c\left(\frac{j}{m}\right)^2 - \sigma^2 mJ_c\left(\frac{j-1}{m}\right)^2 - \frac{2c}{m}\sigma^2 m^2\int_{(j-1)/m}^{j/m} J_c^2\right\} - \frac{\sigma_u^2}{2} \\
&= \frac{\sigma^2 m}{2}\left\{J_c\left(\frac{j}{m}\right)^2 - J_c\left(\frac{j-1}{m}\right)^2 - 2c\int_{(j-1)/m}^{j/m} J_c^2\right\} - \frac{\sigma_u^2}{2}.
\end{aligned}
$$

Using the Itô calculus (see, for example, Tanaka, 1996, p.58) we obtain the following stochastic differential equation for $J_c(t)^2$:

$d[J_c(t)^2] = 2J_c(t)dJ_c(t) + dt;$

substituting $dJ_c(t) = cJ_c(t)dt + dW(t)$ then yields

$d[J_c(t)^2] = 2J_c(t)dW(t) + \left(1 + 2cJ_c(t)^2\right) dt.$

Integrating the above over $[(j-1)/m, j/m]$ we find that

$$J_c\left(\frac{j}{m}\right)^2 - J_c\left(\frac{j-1}{m}\right)^2 = 2\int_{(j-1)/m}^{j/m} J_c dW + 2c\int_{(j-1)/m}^{j/m} J_c^2 + \frac{1}{m},$$

and hence we obtain

$$\begin{aligned}
\frac{1}{\ell}\sum_{t\in\tau_j} y_{t-1}u_t &\Rightarrow \frac{\sigma^2 m}{2}\left(2\int_{(j-1)/m}^{j/m} J_c dW + \frac{1}{m}\right) - \frac{\sigma_u^2}{2} \\
&= \sigma^2 m\int_{(j-1)/m}^{j/m} J_c dW + \frac{1}{2}(\sigma^2 - \sigma_u^2)
\end{aligned}$$

as required.

(c) The result follows immediately from parts (a) and (b) in view of the representation of the normalised estimator in (7). □

Proof of Theorem 2. Proceeding as in the proof of Theorem 1 but retaining higher-order terms we find that

$$\begin{aligned}
\frac{1}{\ell^2}\sum_{t\in\tau_j} y_t^2 &= \frac{n^2}{\ell^2}\sum_{t\in\tau_j}\int_{(t-1)/n}^{t/n}\left(\sum_{j=1}^{t} e^{c(t-j)/n}\int_{(j-1)/n}^{j/n} dX_n(s)\right)^2 dr \\
&\quad + \frac{2n^{3/2}y_0}{\ell^2}\sum_{t\in\tau_j}\int_{(t-1)/n}^{t/n} e^{ct/n}\sum_{j=1}^{t} e^{c(t-j)/n}\int_{(j-1)/n}^{j/n} dX_n(s)dr \\
&\quad + \frac{n}{\ell^2}\sum_{t\in\tau_j}\int_{(t-1)/n}^{t/n} e^{2ct/n}dry_0^2 \\
&= m^2\int_{(j-1)/m}^{j/m}\left(\int_0^r e^{c(r-s)}dX_n(s)\right)^2 dr + \frac{2m^{3/2}y_0}{\sqrt{l}}\int_{(j-1)/m}^{j/m} e^{cr}\int_0^r e^{c(r-s)}dX_n(s)dr \\
&\quad + \frac{m}{\ell^2}\int_{(j-1)/m}^{j/m} e^{2cr}dry_0^2 \\
&\overset{d}{=} \sigma^2 m^2\int_{(j-1)/m}^{j/m} J_c^2 + \frac{2\sigma m^{3/2}y_0}{\sqrt{\ell}}\int_{(j-1)/m}^{j/m} e^{cr}J_c + O_p\left(\frac{1}{\ell}\right).
\end{aligned}$$

Next, as before, we have

$$\frac{1}{\ell}\sum_{t\in\tau_j} y_{t-1}u_t = \frac{1}{2}\left\{\left(\frac{1}{\sqrt{\ell}}y_{jl}\right)^2 - \left(\frac{1}{\sqrt{\ell}}y_{(j-1)l}\right)^2 - \frac{2c}{m\ell^2}\sum_{t\in\tau_j} y_{t-1}^2 - \frac{1}{\ell}\sum_{t\in\tau_j} u_t^2\right\} + O_p\left(\frac{1}{\ell}\right).$$

Now

$$\begin{aligned}\left(\frac{1}{\sqrt{\ell}}y_{jl}\right)^2 &\stackrel{d}{=} \left(\sigma\sqrt{m}J_c\left(\frac{j}{m}\right) + \frac{\sqrt{m}}{\sqrt{\ell}}e^{cj/m}y_0\right)^2 + O_p\left(\frac{1}{\ell}\right)\\ &\stackrel{d}{=} \sigma^2 m J_c\left(\frac{j}{m}\right)^2 + \frac{2\sigma m e^{cj/m}y_0}{\sqrt{\ell}}J_c\left(\frac{j}{m}\right) + O_p\left(\frac{1}{\ell}\right);\end{aligned}$$

a similar result holds for $(y_{(j-1)\ell}/\sqrt{\ell})^2$. Furthermore,

$$\frac{1}{\ell}\sum_{t\in\tau_j} u_t^2 = \frac{1}{\sqrt{\ell}}\left[\frac{1}{\sqrt{\ell}}\sum_{t\in\tau_j}(u_t^2 - \sigma_u^2)\right] + \sigma_u^2 \stackrel{d}{=} \frac{1}{\sqrt{\ell}}\xi_{2j} + \sigma_u^2 + O_p\left(\frac{1}{\ell}\right)$$

where $\xi_{2j} \sim N(0, v_f^2)$ $(j = 1, \ldots, m)$. Combining with the result for $(1/\ell^2)\sum_{t\in\tau_j} y_t^2$ we find that

$$\begin{aligned}\frac{1}{\ell}\sum_{t\in\tau_j} y_{t-1}u_t &\stackrel{d}{=} \frac{1}{2}\left\{\sigma^2 m\left(J_c\left(\frac{j}{m}\right)^2 - J_c\left(\frac{j-1}{m}\right)^2 - 2c\int_{(j-1)/m}^{j/m} J_c^2 - \frac{1}{m}\right) + \sigma^2\right.\\ &\quad + \frac{2\sigma m y_0}{\sqrt{\ell}}\left(e^{cj/m}J_c\left(\frac{j}{m}\right) - e^{c(j-1)/m}J_c\left(\frac{j-1}{m}\right)\right)\\ &\quad \left. - \frac{4c\sigma\sqrt{m}y_0}{\sqrt{\ell}}\int_{(j-1)/m}^{j/m} e^{cr}J_c - \frac{1}{\sqrt{\ell}}\xi_{2j} - \sigma_u^2\right\} + O_p\left(\frac{1}{\ell}\right)\\ &\stackrel{d}{=} \sigma^2 m\int_{(j-1)/m}^{j/m} J_c dW + \frac{1}{2}(\sigma^2 - \sigma_u^2) + \frac{2\sigma m y_0}{\sqrt{\ell}}\xi_{1j} - \frac{1}{\sqrt{\ell}}\xi_{2j} + O_p\left(\frac{1}{\ell}\right)\end{aligned}$$

where ξ_{1j} $(j = 1, \ldots, m)$ is defined in the Theorem. The stated distribution of ξ_{1j} then follows using the property that

$$EJ_c(r)J_c(s) = \frac{e^{c(r+s)} - e^{c(\max(r,s)-\min(r,s))}}{2c}$$

to calculate the variances and covariances; see Perron (1991, p.234). In particular it can be shown that

$$E\left(e^{cj/m}J_c\left(\frac{j}{m}\right) - e^{c(j-1)/m}J_c\left(\frac{j-1}{m}\right)\right)^2 = \frac{1}{2c}\left[(e^{2cj/m} - e^{2c(j-1)/m})^2 + e^{2cj/m} - e^{2c(j-1)/m}\right],$$

$$E\left(\int_{(j-1)/m}^{j/m} e^{cr}J_c\right)^2 = \frac{(e^{2cj/m} - e^{2c(j-1)/m})^2}{8c^3} - \frac{(e^{2cj/m} - e^{2c(j-1)/m})}{4c^3} + \frac{e^{2cj/m}}{2mc^2},$$

$$E\left(\left(e^{cj/m}J_c\left(\frac{j}{m}\right)-e^{c(j-1)/m}J_c\left(\frac{j-1}{m}\right)\right)\int_{(j-1)/m}^{j/m}e^{cr}J_c\right)$$
$$=\frac{(e^{2cj/m}-e^{2c(j-1)/m})^2}{4c^2}+\frac{(e^{2cj/m}-e^{2c(j-1)/m})}{4c^2}-\frac{e^{2cj/m}}{2mc},$$

which combine to determine s^2. The result for $\ell(\hat{\rho}_j-\rho)$ follows from the above results. □

Proof of Theorem 3. (a) The aim is to derive the joint MGF

$$M_c(\theta_1,\theta_2)=E\exp\left(\theta_1\int_a^b J_c dW+\theta_2\int_a^b J_c^2\right).$$

We begin by noting that

$$\int_a^b J_c dW=\frac{1}{2}\left[J_c(b)^2-J_c(a)^2-2c\int_a^b J_c^2-(b-a)\right]$$

so that the function of interest becomes

$$M_c(\theta_1,\theta_2)=\exp\left(-\frac{\theta_1(b-a)}{2}\right)E\exp\left(\frac{\theta_1}{2}\left[J_c(b)^2-J_c(a)^2\right]+(\theta_2-c\theta_1)\int_a^b J_c^2\right).$$

Evaluation of this expectation is aided by introducing the auxiliary O-U process $Y(t)$ on $t\in[0,b]$ with parameter λ, defined by

$$dY(t)=\lambda Y(t)dt+dW(t),\quad Y(0)=0.$$

Let μ_{J_c} and μ_Y denote the probability measures induced by J_c and Y respectively. These measures are equivalent and, by Girsanov's Theorem (see, for example, Theorem 4.1 of Tanaka, 1996),

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

is the Radon-Nikodym derivative evaluated at $s(t)$, a random process on $[0,b]$ with $s(0)=0$. The above change of measure will be used because, for a function $f(J_c)$,

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

Using the change of measure we obtain

$$\begin{aligned}M_c(\theta_1,\theta_2)\;=\;&\exp\left(-\frac{\theta_1(b-a)}{2}\right)E\exp\left(\frac{\theta_1}{2}\left[Y(b)^2-Y(a)^2\right]+(\theta_2-c\theta_1)\int_a^b Y^2\right.\\&\left.+(c-\lambda)\int_0^b YdY-\frac{(c^2-\lambda^2)}{2}\int_0^b Y^2\right).\end{aligned}$$

Now, using the Itô calculus, $\int_0^b YdY=(1/2)[Y(b)^2-b]$, and so

$$\frac{\theta_1}{2}\left[Y(b)^2-Y(a)^2\right]+(c-\lambda)\int_0^b YdY=\frac{(\theta_1+c-\lambda)}{2}Y(b)^2-\frac{\theta_1}{2}Y(a)^2-\frac{(c-\lambda)}{2}b,$$

while splitting the second integral involving Y^2 yields

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

Hence, defining $\delta = \theta_1 + c - \lambda$,

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

As the parameter λ is arbitrary, it is convenient to set $\lambda = (c^2 + 2c\theta_1 - 2\theta_2)^{1/2}$ so as to eliminate the term $\int_a^b Y^2$. We shall then proceed in two steps:

(i) Take the expectation in $M_c(\theta_1,\theta_2)$ 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 $\mathcal{F}_0^a$ we obtain

$$M_c(\theta_1,\theta_2;\mathcal{F}_0^a) = \exp\left(\frac{\theta_1 a - \delta b}{2}\right)\exp\left(-\frac{\theta_1}{2}Y(a)^2 - \frac{(c^2-\lambda^2)}{2}\int_0^a Y^2\right) E\left[\exp\left(\frac{\delta}{2}Y(b)^2\right)\middle|\mathcal{F}_0^a\right].$$

Now, from the representation $Y(b) = \exp((b-a)\lambda)Y(a) + \int_a^b \exp((b-r)\lambda)dW(r)$, it follows that $Y(b)|\mathcal{F}_0^a \sim \mathrm{N}(\mu,\omega^2)$, where

$$\mu = E\left(Y(b)|\mathcal{F}_0^a\right) = \exp((b-a)\lambda)Y(a),$$

$$\omega^2 = E\left[\left(Y(b) - E\left(Y(b)|\mathcal{F}_0^a\right)\right)^2|\mathcal{F}_0^a\right] = \frac{\exp(2(b-a)\lambda) - 1}{2\lambda}.$$

Hence, using Lemma 5 of Magnus (1986), for example,

$$E\left[\exp\left(\frac{\delta}{2}Y(b)^2\right)\middle|\mathcal{F}_0^a\right] = \exp\left(\frac{\delta}{2}kY(a)^2\right)\left(1-\delta\omega^2\right)^{-1/2},$$

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

$$M_c(\theta_1,\theta_2;\mathcal{F}_0^a) = \exp\left(\frac{\theta_1 a - \delta b}{2}\right)\left(1-\delta\omega^2\right)^{-1/2}\exp\left\{\left(\frac{\delta k - \theta_1}{2}\right)Y(a)^2 - \frac{(c^2-\lambda^2)}{2}\int_0^a Y^2\right\}.$$

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 $M_c(\theta_1, \theta_2) = EM_c(\theta_1, \theta_2; \mathcal{F}_0^a)$ and so

$$M_c(\theta_1, \theta_2) = \exp\left(\frac{\theta_1 a - \delta b}{2}\right) \left(1 - \delta\omega^2\right)^{-1/2} E \exp\left\{ \left(\frac{\delta k - \theta_1}{2}\right) Y(a)^2 - \frac{(c^2 - \lambda^2)}{2} \int_0^a Y^2 \right\}.$$

With the change of measure the expectation of interest becomes

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

But η is arbitrary and so we set $\eta = c$ in order to eliminate $\int_0^a V^2$. Furthermore, noting that $\int_0^a V dV = (1/2)[V(a)^2 - a]$, we obtain

$$E \exp\left\{ \left(\frac{\delta k - \theta_1}{2}\right) V(a)^2 + (\lambda - c) \int_0^a V dV \right\} = \exp\left(-\frac{(\lambda - c)}{2} a \right) E \exp\left(\frac{\delta(k-1)}{2} V(a)^2 \right).$$

Now $V(a) = \int_0^a e^{c(a-r)} dW(r)$ and so $V(a) \sim \mathrm{N}(0, v^2)$ where $v^2 = (e^{2ac} - 1)/(2c)$, hence

$$E \exp\left(\frac{\delta(k-1)}{2} V(a)^2 \right) = \left(1 - \delta(k-1)v^2\right)^{-1/2}.$$

It follows that $M_c(\theta_1, \theta_2) = \exp(-(\theta_1 + c)(b - a)/2) H_c(\theta_1, \theta_2)^{-1/2}$ where

$$H_c(\theta_1, \theta_2) = \exp(-(b-a)\lambda)(1 - \delta\omega^2)(1 - \delta(k-1)v^2). \tag{22}$$

Let $z = (b - a)\lambda$. Then

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

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

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

Noting that $\delta(\theta_1 + c + \lambda) = (\theta_1 + c - \lambda)(\theta_1 + c + \lambda) = (\theta_1 + c)^2 - \lambda^2 = \theta_1^2 + 2\theta_2$ and combining these components yields the required expression for $H_c(\theta_1, \theta_2)$.
(b) From the definition of $M_c(\theta_1, \theta_2)$ we obtain

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

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

$$\begin{aligned}\frac{\partial H_c(\theta_1,\theta_2)}{\partial\theta_1} &= c(b-a)\frac{\sinh(b-a)\lambda}{\lambda} + c\left[\theta_1 + c + v^2(\theta_1^2+2\theta_2)\right]\frac{\sinh(b-a)\lambda}{\lambda^3}\\ &\quad -\left(1+2\theta_1 v^2\right)\frac{\sinh(b-a)\lambda}{\lambda} - c(b-a)\left[\theta_1+c+v^2(\theta_1^2+2\theta_2)\right]\frac{\cosh(b-a)\lambda}{\lambda^2},\end{aligned}$$

which makes use of the results

$$\frac{\partial\cosh(b-a)\lambda}{\partial\theta_1} = c(b-a)\frac{\sinh(b-a)\lambda}{\lambda},\quad \frac{\partial\sinh(b-a)\lambda}{\partial\theta_1} = c(b-a)\frac{\cosh(b-a)\lambda}{\lambda}.$$

We need to evaluate $\partial H_c(\theta_1, \theta_2)/\partial\theta_1$ at $\theta_1 = 0$ and at $-\theta_2$, and this is facilitated by defining $x = (c^2 + 2\theta_2)^{1/2}$ to replace λ; this results in

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

It is also convenient to define

$$g(x) = H_c(0,-\theta_2) = \cosh(b-a)x - \left(c-2v^2\theta_2\right)\frac{\sinh(b-a)x}{x}.$$

Combining the results above yields

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

Integrating with respect to θ_2 yields the result in the Theorem. □

Derivation of (14). From (22) we can write

$$M_c(\theta_1,\theta_2) = \left\{\exp((\theta_1+c)(b-a))\exp(-z)(1-\delta\omega^2)(1-\delta(k-1)v^2)\right\}^{-1/2},$$

where $z = (b-a)\lambda$. It can be shown that

$$1-\delta(k-1)v^2 = \frac{1-\delta\omega^2-\delta v^2[\exp(2z)-(1-\delta\omega^2)]}{1-\delta\omega^2}$$

so that

$$(1-\delta\omega^2)(1-\delta(k-1)v^2) = (1+\delta v^2)(1-\delta\omega^2) - \delta v^2\exp(2z).$$

Multiplying by $\exp(-z)$ and noting that

$$\exp(-z)(1-\delta\omega^2) = \frac{2\lambda+\delta}{2\lambda}\exp(-z) - \frac{\delta}{2\lambda}\exp(z)$$

results in the expression for $M_c(\theta_1,\theta_2)$ in (14). □

Proof of Theorem 4. We can examine what happens to the quantities in parts (a) and (b) by considering the joint MGF of $(-2c)^{1/2}\int_a^b J_c dW$ and $(-2c)\int_a^b J_c^2$, which is given by

$$L_c(p,q) = M_c\left((-2c)^{1/2}p, -2cq\right).$$

Using (14) we need to examine the asymptotic properties of λ, δ and v^2 as $c\to-\infty$. The following asymptotic expansions facilitate this:

$$\begin{aligned}\lambda = (c^2+2c(-2c)^{1/2}p+4cq)^{1/2} &= (c^2-2^{3/2}(-c)^{3/2}p+4cq)^{1/2}\\ &= -c-2^{1/2}(-c)^{1/2}p-p^2-2q+O(|c|^{-1/2});\end{aligned}$$

$$\begin{aligned}\delta = (-2c)^{1/2}p+c-\lambda &= 2^{1/2}(-c)^{1/2}p+c+c+2^{1/2}(-c)^{1/2}p+p^2+2q+O(|c|^{-1/2})\\ &= 2^{3/2}(-c)^{1/2}p+2c+p^2+2q+O(|c|^{-1/2});\end{aligned}$$

$$\begin{aligned}2\lambda+\delta = (-2c)^{1/2}p+c+\lambda &= (-2c)^{1/2}p+c-c-2^{1/2}(-c)^{1/2}p-p^2-2q+O(|c|^{-1/2})\\ &= -p^2-2q+O(|c|^{-1/2});\end{aligned}$$

$$\begin{aligned}\delta v^2 &= (\exp(2ac)-1)\left(\frac{2^{3/2}(-c)^{1/2}p+2c+p^2+2q+O(|c|^{-1/2})}{-2(-c)}\right)\\ &= (\exp(2ac)-1)\left(-2^{1/2}(-c)^{-1/2}p+1-2^{-1/2}(-c)^{-1}p^2-(-c)^{-1}q+O(|c|^{-3/2})\right)\\ &\to -1 \quad\text{as}\quad c\to-\infty.\end{aligned}$$

Combining these results we find that

$$L_c(p,q)\to\exp\left\{\left(\frac{1}{2}p^2+q\right)(b-a)\right\}\quad\text{as}\quad c\to-\infty,$$

from which the results in (a) and (b) follow immediately. To establish (c), note that

$$K(c) = (b-a)^{1/2}\left((-2c)\int_a^b J_c^2\right)^{-1}\left((-2c)^{1/2}\int_a^b J_c dW + \frac{1}{2}(1-\eta)\right)+o_p(1).$$

The result then follows using (a) and (b). □

Proof of Theorem 5. To determine the weights for $\hat{\rho}_J^*$, note that

$$E(\hat{\rho}) = \rho + \frac{\mu_c}{n} + O\left(\frac{1}{n^2}\right), \quad E(\hat{\rho}_j) = \rho + \frac{\mu_{c,j}}{\ell} + O\left(\frac{1}{\ell^2}\right), \quad j = 1, \ldots, m,$$

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

$$E(\hat{\rho}_J^*) = (w_1^c + w_2^c)\rho + \frac{1}{n}\left(w_1^c \mu_c + w_2^c \sum_{j=1}^m \mu_{c,j}\right) + O\left(\frac{1}{n^2}\right).$$

In order that $E(\hat{\rho}_J^*) = \rho + O(1/n^2)$ the requirements are that:

(i) $w_1^c + w_2^c = 1$, and

(ii) $w_1^c \mu_c + w_2^c \sum_{j=1}^m \mu_{c,j} = 0$.

Solving these two conditions simultaneously yields the stated weights. □

Proof of Theorem 6. (a) The object of interest is, using (20) and (21),

$$\begin{aligned}
\frac{1}{\ell^{3/2}} \sum_{t \in \tau_j} y_t &= \frac{1}{\ell^{3/2}} \sum_{t \in \tau_j} \left\{ \sum_{j=1}^t e^{c(t-j)/n} u_j + e^{ct/n} y_0 \right\} \\
&= \frac{\sqrt{n}}{\ell^{3/2}} \sum_{t \in \tau_j} \sum_{j=1}^t e^{c(t-j)/n} \int_{(j-1)/n}^{j/n} dX_n(s) + o_p(1) \\
&= \frac{n^{3/2}}{\ell^{3/2}} \sum_{t \in \tau_j} \int_{(t-1)/n}^{t/n} \left\{ \sum_{j=1}^t \int_{(j-1)/n}^{j/n} e^{c(t-j)/n} dX_n(s) \right\} dr + o_p(1) \\
&= m^{3/2} \int_{(j-1)/m}^{j/m} \left\{ \int_0^r e^{c(r-s)} dX_n(s) \right\} dr + o_p(1) \\
&\Rightarrow \sigma m^{3/2} \int_{(j-1)/m}^{j/m} J_c
\end{aligned}$$

as required.

(b) From the usual least squares regression formulae we may write the normalised estimator as

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

We can make use of the convergence results in Theorem 1 and part (a) above allied to the following:

$$\frac{1}{\sqrt{\ell}} \sum_{t \in \tau_j} u_t = \frac{1}{\sqrt{\ell}} (S_{j\ell} - S_{(j-1)\ell}) = \frac{1}{\sqrt{\ell}} (S_{jn/m} - S_{(j-1)n/m}) \Rightarrow \sigma \sqrt{m} \left[W\left(\frac{j}{m}\right) - W\left(\frac{(j-1)}{m}\right) \right].$$

Combining these results yields

$$
\begin{aligned}
\ell(\tilde{\rho}_j - \rho) \quad &\Rightarrow \quad \frac{\sigma^2 m \int_{(j-1)/m}^{j/m} J_c dW + \frac{1}{2}(\sigma^2 - \sigma_u^2) - \sigma^2 m^2 \left[W\left(\frac{j}{m}\right) - W\left(\frac{j-1}{m}\right) \right] \int_{(j-1)/m}^{j/m} J_c}{\sigma^2 m^2 \int_{(j-1)/m}^{j/m} J_c^2 - \left(\sigma m^{3/2} \int_{(j-1)/m}^{j/m} J_c \right)^2} \\
&\overset{d}{=} \quad \frac{\int_{(j-1)/m}^{j/m} J_{c,j}^{\mu} dW + \frac{1}{2m}(1-\eta)}{m \int_{(j-1)/m}^{j/m} (J_{c,j}^{\mu})^2}
\end{aligned}
$$

where the second line follows by dividing the numerator and denominator by $\sigma^2 m$ and using the properties of $J_{c,j}^{\mu}$ to represent the relevant terms. $\square$

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.

Chambers, M.J., 2013. Jackknife estimation and inference in stationary autoregressive models. *Journal of Econometrics* 172, 142–157.

Chambers, M.J., 2015. A jackknife correction to a test for cointegration rank. *Econometrics* 3, 355–375.

Chambers, M.J., Kyriacou, M., 2013. Jackknife estimation with a unit root. *Statistics and Probability Letters* 83, 1677–1682.

Chen, Y., Yu, J., 2015. Optimal jackknife for unit root models. *Statistics and Probability Letters* 99, 135–142.

Davidson, J., 2000. *Econometric Theory.* Blackwell, Oxford.

Gonzalo, J., 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.

Kruse, R., Kaufmann, H., 2015. Bias-corrected estimation in mildly explosive autoregressions. Annual Conference 2015 (Muenster): Economic Development – Theory and Policy 112897, Verein fr Socialpolitik/German Economic Association.

Magnus, J.R., 1986. The exact moments of a ratio of quadratic forms in normal variables. *Annales d'Économie et de Statistique* 4, 95–109.

Mikusheva, A., 2015. Second order expansions of the T-statistic in AR(1) models. *Econometric Theory* 31, 426–448.

Park, J., 2006. A bootstrap theory for weakly integrated processes. *Journal of Econometrics* 133, 639–672.

Perron, P., 1989. The calculation of the limiting distribution of the least-squares estimator in a near-integrated model. *Econometric Theory* 5, 241–255.

Perron, P., 1991. A continuous time approximation to the unstable first-order autoregressive process: the case without an intercept. *Econometrica* 59, 211-236.

Perron, P., 1996. The adequacy of asymptotic approximations in the near-integrated autoregressive model with dependent errors. *Journal of Econometrics* 70, 317–350.

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

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

Phillips, P.C.B., 2012. Folklore theorems, implicit maps, and indirect inference. *Econometrica* 80, 425–454.

Phillips, P.C.B., 2014. On confidence intervals for autoregressive roots and predictive regression. *Econometrica* 82, 1177–1195.

Phillips, P. C. B., 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.

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.

Table 1
Bias of OLS estimator

n	c = -10	-5	-1	0	1
$\hat{\rho} - \rho$					
24	-0.0443	-0.0598	-0.0681	-0.0667	-0.0612
48	-0.0319	-0.0355	-0.0365	-0.0351	-0.0317
96	-0.0185	-0.0193	-0.0190	-0.0181	-0.0162
192	-0.0098	-0.0099	-0.0096	-0.0091	-0.0081
∞	0.0000	0.0000	0.0000	0.0000	0.0000
$n(\hat{\rho} - \rho)$					
24	-1.0632	-1.4352	-1.6344	-1.6008	-1.4448
48	-1.5312	-1.7040	-1.7520	-1.6848	-1.5216
96	-1.7760	-1.8528	-1.8240	-1.7376	-1.5552
192	-1.8816	-1.9008	-1.8432	-1.7472	-1.5552
∞	-1.9912	-1.9758	-1.8818	-1.7814	-1.5812

Table 2
Values of $\mu_{c,j} = E(Z_{c,j})$

$j \setminus c$	-50	-20	-10	-5	-1	0	1
$m = 1$							
1	−1.9995	−1.9972	−1.9912	−1.9758	−1.8818	−1.7814	−1.5812
$m = 2$							
1	−1.9981	−1.9912	−1.9758	−1.9439	−1.8408	−1.7814	−1.6969
2	−1.9604	−1.9043	−1.8214	−1.6891	−1.3295	−1.1382	−0.8916
$m = 3$							
1	−1.9962	−1.9838	−1.9595	−1.9175	−1.8233	−1.7814	−1.7282
2	−1.9412	−1.8613	−1.7502	−1.5921	−1.2720	−1.1382	−0.9785
3	−1.9412	−1.8613	−1.7500	−1.5845	−1.1514	−0.9319	−0.6746
$m = 4$							
1	−1.9939	−1.9758	−1.9439	−1.8973	−1.8137	−1.7814	−1.7426
2	−1.9225	−1.8214	−1.6891	−1.5210	−1.2407	−1.1382	−1.0200
3	−1.9225	−1.8214	−1.6879	−1.5021	−1.1013	−0.9319	−0.7396
4	−1.9225	−1.8214	−1.6879	−1.5006	−1.0393	−0.8143	−0.5623
$m = 6$							
1	−1.9884	−1.9594	−1.9175	−1.8698	−1.8034	−1.7814	−1.7561
2	−1.8867	−1.7502	−1.5921	−1.4268	−1.2076	−1.1382	−1.0601
3	−1.8867	−1.7500	−1.5845	−1.3812	−1.0475	−0.9319	−0.8042
4	−1.8867	−1.7500	−1.5843	−1.3732	−0.9691	−0.8143	−0.6452
5	−1.8867	−1.7500	−1.5842	−1.3717	−0.9237	−0.7348	−0.5306
6	−1.8867	−1.7500	−1.5842	−1.3715	−0.8953	−0.6761	−0.4421
$m = 8$							
1	−1.9823	−1.9439	−1.8973	−1.8526	−1.7980	−1.7814	−1.7625
2	−1.8530	−1.6891	−1.5210	−1.3686	−1.1904	−1.1382	−1.0795
3	−1.8530	−1.6879	−1.5021	−1.2991	−1.0193	−0.9319	−0.8361
4	−1.8530	−1.6879	−1.5006	−1.2815	−0.9318	−0.8143	−0.6872
5	−1.8530	−1.6879	−1.5005	−1.2766	−0.8787	−0.7348	−0.5806
6	−1.8530	−1.6879	−1.5005	−1.2752	−0.8437	−0.6761	−0.4982
7	−1.8530	−1.6879	−1.5005	−1.2748	−0.8194	−0.6302	−0.4316
8	−1.8530	−1.6879	−1.5005	−1.2747	−0.8021	−0.5931	−1.3762
$m = 12$							
1	−1.9693	−1.9175	−1.8698	−1.8324	−1.7924	−1.7814	−1.7688
2	−1.7916	−1.5921	−1.4268	−1.3016	−1.1727	−1.1382	−1.0987
3	−1.7916	−1.5845	−1.3812	−1.1979	−0.9903	−0.9319	−0.8678
4	−1.7916	−1.5842	−1.3732	−1.1612	−0.8931	−0.8143	−0.7292
5	−1.7916	−1.5842	−1.3717	−1.1464	−0.8317	−0.7348	−0.6313
6	−1.7916	−1.5842	−1.3715	−1.1403	−0.7894	−0.6761	−0.5562
7	−1.7916	−1.5842	−1.3714	−1.1376	−0.7586	−0.6302	−0.4957
8	−1.7916	−1.5842	−1.3714	−1.1365	−0.7354	−0.5931	−0.4452
9	−1.7916	−1.5842	−1.3714	−1.1360	−0.7174	−0.5622	−0.4021
10	−1.7916	−1.5842	−1.3714	−1.1358	−0.7033	−0.5358	−0.3647
11	−1.7916	−1.5842	−1.3714	−1.1357	−0.6921	−0.5131	−0.3319
12	−1.7916	−1.5842	−1.3714	−1.1356	−0.6830	−0.4931	−0.3027

Table 3

Values of standard and 'optimal' jackknife weights

$m:$	2	3	4	6	8	12
Standard weights						
w_1	2.0000	1.5000	1.3333	1.2000	1.1429	1.0909
w_2	−1.0000	−0.5000	−0.3333	−0.2000	−0.1429	−0.0909
Optimal weights: $c = -50$						
$w^*_{1,c}$	2.0206	1.5156	1.3470	1.2122	1.1544	1.1016
$w^*_{2,c}$	−1.0206	−0.5156	−0.3470	−0.2122	−0.1544	−0.1016
Optimal weights: $c = -20$						
$w^*_{1,c}$	2.0521	1.5385	1.3670	1.2292	1.1698	1.1151
$w^*_{2,c}$	−1.0521	−0.5385	−0.3670	−0.2292	−0.1698	−0.1151
Optimal weights: $c = -10$						
$w^*_{1,c}$	2.1026	1.5741	1.3969	1.2535	1.1909	1.1325
$w^*_{2,c}$	−1.1026	−0.5741	−0.3969	−0.2535	−0.1909	−0.1325
Optimal weights: $c = -5$						
$w^*_{1,c}$	2.1923	1.6336	1.4445	1.2898	1.2213	1.1565
$w^*_{2,c}$	−1.1923	−0.6336	−0.4445	−0.2898	−0.2213	−0.1565
Optimal weights: $c = -1$						
$w^*_{1,c}$	2.4605	1.7957	1.5680	1.3790	1.2940	1.2120
$w^*_{2,c}$	−1.4605	−0.7957	−0.5680	−0.3790	−0.2940	−0.2120
Optimal weights: $c = 0$						
$w^*_{1,c}$	2.5651	1.8605	1.6176	1.4147	1.3228	1.2337
$w^*_{2,c}$	−1.5651	−0.8605	−0.6176	−0.4147	−0.3228	−0.2337
Optimal weights: $c = 1$						
$w^*_{1,c}$	2.5696	1.8783	1.6367	1.4324	1.3385	1.2465
$w^*_{2,c}$	−1.5696	−0.8783	−0.6367	−0.4324	−0.3385	−0.2465

Table 4
Bias of OLS and jackknife estimators

n	$\hat{\rho}$	$\hat{\rho}_{J,B}$	$\hat{\rho}^*_{J,B}$	$\hat{\rho}_{J,R}$	$\hat{\rho}^*_{J,R}$
$c = -10$					
24	-0.0443	-0.0115_2	-0.0081_2	-0.0259_6	-0.0210_6
48	-0.0319	-0.0061_2	-0.0035_2	-0.0161_8	-0.0147_{12}
96	-0.0185	-0.0028_2	-0.0012_2	-0.0092_{12}	-0.0050_{12}
192	-0.0098	-0.0012_2	-0.0003_2	-0.0040_{12}	-0.0014_{12}
$c = -5$					
24	-0.0598	-0.0198_2	-0.0121_2	-0.0388_6	-0.0294_6
48	-0.0355	-0.0091_2	-0.0040_2	-0.0178_6	-0.0123_8
96	-0.0193	-0.0041_2	-0.0012_2	-0.0093_8	-0.0052_{12}
192	-0.0099	-0.0018_2	-0.0003_2	-0.0041_8	-0.0014_{12}
$c = -1$					
24	-0.0681	-0.0325_2	-0.0162_2	-0.0438_4	-0.0357_6
48	-0.0365	-0.0145_2	-0.0044_2	-0.0228_6	-0.0131_8
96	-0.0190	-0.0069_2	-0.0013_2	-0.0105_6	-0.0051_{12}
192	-0.0096	-0.0032_2	-0.0005_2	-0.0054_8	-0.0013_{12}
$c = 0$					
24	-0.0667	-0.0344_2	-0.0162_2	-0.0517_6	-0.0356_6
48	-0.0351	-0.0154_2	-0.0042_2	-0.0231_6	-0.0128_8
96	-0.0181	-0.0073_2	-0.0012_2	-0.0107_6	-0.0049_{12}
192	-0.0091	-0.0034_2	-0.0002_2	-0.0055_8	-0.0013_{12}
$c = 1$					
24	-0.0612	-0.0314_2	-0.0145_2	-0.0480_6	-0.0327_6
48	-0.0317	-0.0138_2	-0.0037_2	-0.0213_6	-0.0155_{12}
96	-0.0162	-0.0078_3	-0.0010_2	-0.0107_8	-0.0045_{12}
192	-0.0081	-0.0031_2	-0.0002_2	-0.0050_8	-0.0011_{12}

Subscripts denote the value of m.

Table 5
RMSE of OLS and jackknife estimators

n	$\hat{\rho}$	$\hat{\rho}_{J,B}$	$\hat{\rho}^*_{J,B}$	$\hat{\rho}_{J,R}$	$\hat{\rho}^*_{J,R}$
$c = -10$					
24	0.1820	0.1985_2	0.2018_2	0.1856_6	0.1877_6
48	0.1044	0.1101_2	0.1119_2	0.1022_8	0.1022_{12}
96	0.0560	0.0584_2	0.0594_2	0.0535_{12}	0.0530_{12}
192	0.0288	0.0300_2	0.0305_2	0.0272_{12}	0.0269_{12}
$c = -5$					
24	0.1640	0.1761_2	0.1829_2	0.1619_6	0.1637_6
48	0.0907	0.0950_2	0.0989_2	0.0859_6	0.0852_8
96	0.0477	0.0497_2	0.0519_2	0.0441_8	0.0431_{12}
192	0.0243	0.0253_2	0.0265_2	0.0222_8	0.0216_{12}
$c = -1$					
24	0.1436	0.1532_2	0.1731_2	0.1376_4	0.1396_6
48	0.0765	0.0802_2	0.0915_2	0.0700_6	0.0681_8
96	0.0395	0.0413_2	0.0475_2	0.0353_6	0.0332_{12}
192	0.0199	0.0209_2	0.0251_2	0.0176_8	0.0165_{12}
$c = 0$					
24	0.1368	0.1483_2	0.1755_2	0.1311_6	0.1347_6
48	0.0721	0.0770_2	0.0921_2	0.0660_6	0.0643_8
96	0.0370	0.0394_2	0.0477_2	0.0331_6	0.0310_{12}
192	0.0186	0.0199_2	0.0242_2	0.0164_8	0.0153_{12}
$c = 1$					
24	0.1280	0.1431_2	0.1732_2	0.1234_6	0.1292_6
48	0.0667	0.0736_2	0.0900_2	0.0616_6	0.0606_{12}
96	0.0340	0.0329_3	0.0465_2	0.0308_8	0.0291_{12}
192	0.0170	0.0189_2	0.0235_2	0.0152_8	0.0144_{12}

Subscripts denote the value of m.

Table 6

Bias of OLS, jackknife, MU and II estimators in regression with intercept

n	$\tilde{\rho}$	$\tilde{\rho}_{J,B}$	$\tilde{\rho}^*_{J,B}$	$\tilde{\rho}_{J,R}$	$\tilde{\rho}^*_{J,R}$	$\tilde{\rho}_{MU}$	$\tilde{\rho}_{II}$
$c = -10$							
24	-0.1204	-0.0094_2	-0.0011_3	-0.0347_6	-0.0118_6	-0.0248	0.0078
48	-0.0776	-0.0018_2	-0.0021_{12}	-0.0175_8	-0.0021_{12}	-0.0176	0.0026
96	-0.0432	-0.0001_4	0.0058_2	-0.0070_{12}	0.0097_{12}	-0.0103	0.0014
192	-0.0228	-0.0003_8	0.0037_2	-0.0007_{12}	0.0093_{12}	-0.0056	0.0003
$c = -5$							
24	-0.1584	-0.0205_2	0.0009_3	-0.0592_6	-0.0146_6	-0.0414	-0.0039
48	-0.0910	-0.0050_2	0.0046_{12}	-0.0260_8	0.0046_{12}	-0.0245	-0.0016
96	-0.0484	-0.0000_2	0.0093_2	-0.0105_{12}	0.0168_{12}	-0.0133	-0.0012
192	-0.0250	0.0001_4	0.0057_2	-0.0023_{12}	0.0141_{12}	-0.0070	-0.0011
$c = -1$							
24	-0.1992	-0.0464_2	0.0051_6	-0.0913_6	0.0051_6	-0.0837	-0.0556
48	-0.1070	-0.0153_2	-0.0269_2	-0.0385_8	0.0284_{12}	-0.0445	-0.0293
96	-0.0555	-0.0054_2	0.0177_2	-0.0166_{12}	0.0352_{12}	-0.0229	-0.0144
192	-0.0281	-0.0016_2	0.0133_2	-0.0050_{12}	0.0193_6	-0.0116	-0.0070
$c = 0$							
24	-0.1990	-0.0402_2	0.0161_8	-0.0679_4	0.0306_6	-0.0939	-0.0703
48	-0.1055	-0.0115_2	0.0416_2	-0.0359_8	0.0460_{12}	-0.0492	-0.0374
96	-0.0545	-0.0034_2	0.0255_2	-0.0152_{12}	0.0465_{12}	-0.0251	-0.0192
192	-0.0275	-0.0007_2	0.0145_2	-0.0043_{12}	0.0247_6	-0.0126	-0.0094
$c = 1$							
24	-0.1764	-0.0232_2	0.0485_8	-0.0659_6	0.0617_6	-0.1010	-0.0723
48	-0.0920	-0.0026_2	0.0483_2	-0.0229_8	0.0668_{12}	-0.0519	-0.0373
96	-0.0472	-0.0003_4	0.0288_2	-0.0082_{12}	0.0495_6	-0.0263	-0.0184
192	-0.0238	0.0003_8	0.0157_2	-0.0008_{12}	0.0240_4	-0.0131	-0.0094

Subscripts denote the value of m.

Table 7
RMSE of OLS, jackknife and MU estimators in regression with intercept

n	$\tilde{\rho}$	$\tilde{\rho}_{J,B}$	$\tilde{\rho}^*_{J,B}$	$\tilde{\rho}_{J,R}$	$\tilde{\rho}^*_{J,R}$	$\tilde{\rho}_{MU}$	$\tilde{\rho}_{II}$
$c = -10$							
24	0.2249	0.2509_2	0.2373_3	0.2164_6	0.2224_6	0.2139	0.2286
48	0.1353	0.1441_2	0.1212_{12}	0.1207_8	0.1212_{12}	0.1197	0.1236
96	0.0742	0.0681_4	0.0814_2	0.0635_{12}	0.0655_{12}	0.0637	0.0648
192	0.0388	0.0334_8	0.0426_2	0.0327_{12}	0.0349_{12}	0.0329	0.0333
$c = -5$							
24	0.2380	0.2531_2	0.2377_3	0.2100_6	0.2173_6	0.2017	0.2105
48	0.1354	0.1408_2	0.1129_{12}	0.1115_8	0.1129_{12}	0.1095	0.1097
96	0.0722	0.0752_2	0.0823_2	0.0573_{12}	0.0615_{12}	0.0575	0.0572
192	0.0372	0.0321_4	0.0429_2	0.0291_{12}	0.0337_{12}	0.0294	0.0295
$c = -1$							
24	0.2563	0.2516_2	0.2196_6	0.2069_6	0.2196_6	0.1807	0.1791
48	0.1391	0.1367_2	0.1710_2	0.1044_8	0.1117_{12}	0.0966	0.0938
96	0.0726	0.0721_2	0.0917_2	0.0522_{12}	0.0657_{12}	0.0503	0.0475
192	0.0369	0.0369_2	0.0505_2	0.0258_{12}	0.0373_6	0.0255	0.0239
$c = 0$							
24	0.2535	0.2451_2	0.2974_8	0.2021_4	0.2245_6	0.1722	0.1695
48	0.1360	0.1326_2	0.1777_2	0.1005_8	0.1169_{12}	0.0916	0.0902
96	0.0705	0.0697_2	0.0953_2	0.0498_{12}	0.0718_{12}	0.0475	0.0463
192	0.0358	0.0357_2	0.0495_2	0.0246_{12}	0.0405_6	0.0239	0.0234
$c = 1$							
24	0.2340	0.2259_2	0.3109_8	0.1849_6	0.2234_6	0.1587	0.1558
48	0.1242	0.1219_2	0.1671_2	0.0922_8	0.1235_{12}	0.0835	0.0809
96	0.0639	0.0501_4	0.0896_2	0.0456_{12}	0.0762_6	0.0429	0.0402
192	0.0323	0.0234_8	0.0464_2	0.0228_{12}	0.0410_4	0.0216	0.0214

Subscripts denote the value of m.