

## Article

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

Marcus J. Chambers <sup>1,\*</sup> and Maria Kyriacou <sup>2</sup><sup>1</sup> Department of Economics, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK<sup>2</sup> Department of Economics, University of Southampton, Southampton SO17 1BJ, UK; m.kyriacou@soton.ac.uk

\* Correspondence: mchamb@essex.ac.uk; Tel.: +44-120-687-2756; Fax: +44-120-687-2724

Received: 29 September 2017; Accepted: 22 February 2018; Published: 5 March 2018

**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 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 and a quantity that is related to the long-run variance of the disturbance process, which are typically unknown in practice, and so, this dependence is characterised fully and a discussion provided of the issues that arise in practice in the most general settings.

**Keywords:** Jackknife; bias reduction; near-unit root; moment generating function

**JEL Classification:** C22

## 1. Introduction

Throughout his career, Peter Phillips has made important contributions to knowledge across the broad spectrum of econometrics and statistics, providing inspiration to many other researchers along the way. This paper builds on two of the strands of Peter's research, namely jackknife bias reduction and the analysis of nonstationary time series. Indeed, our own work on the jackknife (Chambers 2013, 2015; Chambers and Kyriacou 2013) was inspired by Peter's work on this topic with Jun Yu, published as Phillips and Yu (2005), and the current contribution also extends the results on moment generating functions (MGFs) contained in Phillips (1987a).

The jackknife has been proven 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 to accommodate more general time series settings. Within the class of stationary autoregressive time series models, Phillips and Yu (2005) show that the jackknife can effectively reduce bias in the pricing of bond options in finance, while Chambers (2013) analyses the performance of jackknife methods based on a variety of sub-sampling procedures. In subsequent work, Chambers and Kyriacou (2013) demonstrate that the usual jackknife construction in the time series case has to be amended when a unit root is present, while Chen and Yu (2015) show that a variance-minimising jackknife can be constructed in a unit root setting that also retains its bias reduction properties. In addition, Kruse and Kaufmann (2015) 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 original full-sample 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 fully the first-order term in the resulting jackknife estimator's bias expansion. 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 eliminate fully the first-order bias in the near-unit root setting. Near-unit root models have attracted a great deal of interest in time series owing, amongst other things, to their ability to capture better the effects of sample size in the vicinity of a unit root, to explore analytically the power properties of unit root tests and to allow the development of an integrated asymptotic theory for both stationary and non-stationary autoregressions; see [Phillips \(1987a\)](#) and [Chan and Wei \(1987\)](#) for details. We find that jackknife estimators can be constructed in the presence of a near-unit root that achieve this aim of bias reduction. Jackknife estimators have the advantage of incurring only a very slight additional computational burden, unlike alternative resampling and simulation-based methods such as the bootstrap and indirect inference. Furthermore, they are applicable in a wide variety of estimation frameworks and work well in finite sample situations in which the prime objective is bias reduction. Although the bootstrap is often a viable candidate for bias reduction, it was shown by [Park \(2006\)](#) that the bootstrap is inconsistent in the presence of a near-unit root, and hence, jackknife methods offer a useful alternative in these circumstances.

The development of a jackknife estimator that achieves bias reduction in the near-unit root case is not simply a straightforward application of previous results. While [Chambers and Kyriacou \(2013\)](#) first pointed out that under unit root non-stationarity, the effect of the sub-sample initial conditions does not vanish asymptotically, thereby affecting asymptotic expansions of sub-sample estimator bias and the resulting jackknife weights as compared to the stationary case, the extension of these results to a local-to-unity setting is not obvious. With a near-unit root, the autoregressive parameter plays an important role, and it is therefore necessary to derive the appropriate asymptotic expansion of sub-sample estimator bias for this more general case, as well as the MGFs of the relevant limiting distributions that can be used to construct the appropriate jackknife weights. The derivation of such results is challenging in itself and is a major reason why we focus on the bias-minimising jackknife, rather than attempting to derive results for the variance-minimising jackknife of [Chen and Yu \(2015\)](#).

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 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 then goes on to explore, via simulations, the performance of the proposed 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. Section 5 contains some concluding comments, and all proofs are contained in the Appendix A.

The following notation will be used throughout the paper. The symbol  $\stackrel{d}{=}$  denotes equality in distribution;  $\stackrel{d}{\rightarrow}$  denotes convergence in distribution;  $\stackrel{p}{\rightarrow}$  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  $\int_0^1 J_c(r)^2 dr$ , are 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 is to 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 model with a near-unit root is defined as follows.

**Assumption 1.** The sequence  $y_1, \dots, y_n$  satisfies:

$$y_t = \rho y_{t-1} + u_t, \quad t = 1, \dots, n, \quad (1)$$

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:

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

where  $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$ ,  $E(\epsilon_t^4) < \infty$ ,  $\delta(z) = \sum_{j=0}^{\infty} \delta_j z^j$ ,  $\delta_0 = 1$  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  $c > 0$ . Strictly speaking, the autoregressive parameter should be denoted  $\rho_n$  to emphasise its dependence on the sample size,  $n$ , but we use  $\rho$  for notational convenience. 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:

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} u_t \Rightarrow \sigma W(r) \quad \text{as } n \rightarrow \infty \quad (3)$$

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). In economic and financial time series, they offer a flexible mechanism of modelling highly persistent series whose autoregressive roots are generally close, but not exactly equal, to unity. Ordinary least squares (OLS) regression on (1) yields:

$$y_t = \hat{\rho} y_{t-1} + \hat{u}_t, \quad t = 1, \dots, n, \quad (4)$$

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

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

where  $\eta = \sigma_u^2 / \sigma^2$ ,  $\sigma_u^2 = E(u_t^2) = \sigma_\varepsilon^2 \sum_{j=0}^\infty \delta_j^2$  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)  $\eta = 1$ .

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,  $\hat{\rho}_j$  ( $j = 1, \dots, m$ ), 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. In this approach, the full sample of  $n$  observations is divided into  $m$  sub-samples, each of length  $\ell$ , so that  $n = m \times \ell$ . The generic form of jackknife estimator is given by:

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

where the weights are determined so as to eliminate the first-order finite sample bias. 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 the appropriate weights are given by  $w_1 = m/(m-1)$  and  $w_2 = -1/(m-1)$ , in which case:

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

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  $\hat{\rho}$ . However, in the pure unit root setting ( $c = 0$ ), [Chambers and Kyriacou \(2013\)](#) demonstrated that the sub-sample estimators do not share the same limit distribution as the full-sample estimator, which means that the expansions for the bias of the sub-sample estimators are incorrect, and hence, the weights defined above do not eliminate fully the first-order bias. It is therefore important to investigate this issue in the more general setting of a near-unit root with a view toward deriving the appropriate weights for eliminating the first-order bias term.

## 2.2. Sub-Sample Properties

In order to explore the sub-sample properties, let:

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

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:

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

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  $y_1, \dots, y_n$  satisfy Assumption 1. Then, if  $m$  is fixed as  $n \rightarrow \infty$  (and hence,  $\ell \rightarrow \infty$ ):

$$\begin{aligned} (a) \quad & \frac{1}{\ell^2} \sum_{t \in \tau_j} y_t^2 \Rightarrow \sigma^2 m^2 \int_{(j-1)/m}^{j/m} J_c^2; \\ (b) \quad & \frac{1}{\ell} \sum_{t \in \tau_j} y_{t-1} u_t \Rightarrow \sigma^2 m \int_{(j-1)/m}^{j/m} J_c dW + \frac{1}{2}(\sigma^2 - \sigma_u^2); \\ (c) \quad & \ell(\hat{\rho}_j - \rho) \Rightarrow Z_{c,j}(\eta) = \frac{\int_{(j-1)/m}^{j/m} J_c dW + \frac{1}{2m}(1 - \eta)}{m \int_{(j-1)/m}^{j/m} J_c^2}, \quad j = 1, \dots, m, \\ & \eta = \sigma_u^2 / \sigma^2. \end{aligned}$$

where the functional  $Z_{c,j}(\eta)$  is implicitly defined.

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:

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

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 defined following (6) 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:

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

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:

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

where  $\xi$  is a standard normal random variable distributed independently of  $W$ . Taking expectations in (9), using the independence of  $\xi$  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:

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

see, also, Phillips (2012, 2014). In the more general setting of the model in Assumption 1, and assuming that  $u_t$  is Gaussian, Theorem 1 of Perron (1996) established that:

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

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. The following result extends the type of expansion in (11) to the sub-sample estimators.

**Theorem 2.** Let  $y_1, \dots, y_n$  satisfy Assumption 1 and, in addition, assume that  $u_t$  is Gaussian. Then, for  $j = 1, \dots, m$ ,

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

where  $\xi_{1j} \sim N(0, s_j^2)$ ,  $\xi_{2j} \sim N(0, 1)$  and:

$$\begin{aligned} s_j^2 &= \frac{(\sqrt{m} - 1)^2}{2cm} \left( e^{cj/m} - e^{c(j-1)/m} \right)^2 + \frac{(m - 2\sqrt{m} - 2)}{2cm} \left( e^{cj/m} - e^{c(j-1)/m} \right) \\ &\quad + \frac{2(1 + \sqrt{m})}{m^2} e^{cj/m}. \end{aligned}$$

The form of the expansion for  $\ell(\hat{\rho}_j - \rho)$  in Theorem 2 is similar to that for the full-sample estimator, but also 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  $y_0 = 0^1$ , 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 ( $\xi_{1j}$  and  $\xi_{2j}$ ) 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, thereby enabling the construction of a jackknife estimator that eliminates the first-order bias.

<sup>1</sup> If  $y_0 \neq 0$ , then additional random variables appear in the numerator and denominator of the bias, thereby complicating the derivation of the required expectation. In the case of  $c = 0$ , the simulation results reported in Table 2.3 of Kyriacou (2011) indicate that the bias of  $\hat{\rho}$  increases with  $y_0/\sigma$ , but that the jackknife continues to be an effective method of bias reduction.

### 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  $v^2 = (e^{2ac} - 1)/(2c)$ ,

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

(b) The individual MGFs for  $N_c$  and  $D_c$  are given by, respectively,

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

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

where  $\lambda_1 = (c^2 + 2c\theta_1)^{1/2}$  and  $\lambda_2 = (c^2 - 2\theta_2)^{1/2}$ .

(c) Let:

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

Then, the expectation of  $N_c/D_c$  is given by:

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

where:

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

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

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

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



The MGFs for the two functionals in Theorem 3 have potential applications in a wide range of sub-sampling problems with near-unit root processes. A potential application of the joint MGF in Part (a) of Theorem 3 is in the computation of the cumulative and probability density functions of the distributions  $Z_{c,j}(\eta)$  when setting  $a = (j-1)/m$  and  $b = j/m$ . For example, the probability density function of  $mZ_{c,j}(1)$  is given by (with  $i^2 = -1$ ):

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

see, for example, Perron (1991, p. 221), who performs this type of 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.

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:

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

these expressions can be found; for example, Gonzalo and Pitarakis (1998, Lemma 3.1). Some further special cases of interest that follow from Theorem 3 are presented below.

**Corollary to Theorem 3.** (a) Let  $[a, b] = [0, 1]$  so that  $N_c = \int_0^1 J_c(r) dW(r)$  and  $D_c = \int_0^1 J_c(r)^2 dr$ . Then:

$$\begin{aligned} M_c(\theta_1, \theta_2) &= \exp\left(-\frac{(\theta_1 + c)}{2}\right) \left[ \cosh(\lambda) - \frac{1}{\lambda} (\theta_1 + c) \sinh(\lambda) \right]^{-1/2}, \\ M_{N_c}(\theta_1) &= \exp\left(-\frac{(\theta_1 + c)}{2}\right) \left[ \cosh(\lambda_1) - \frac{1}{\lambda_1} (\theta_1 + c) \sinh(\lambda_1) \right]^{-1/2}, \\ M_{D_c}(\theta_2) &= \exp\left(-\frac{c}{2}\right) \left[ \cosh(\lambda_2) - \frac{c}{\lambda_2} \sinh(\lambda_2) \right]^{-1/2}, \end{aligned}$$

while taking the limit as  $c \rightarrow 0$  yields:

$$\begin{aligned} M_0(\theta_1, \theta_2) &= \exp\left(-\frac{\theta_1}{2}\right) \left[ \cosh(\lambda_0) - \frac{\theta_1}{\lambda_0} \sinh(\lambda_0) \right]^{-1/2}, \\ M_{N_0}(\theta_1) &= e^{-\theta_1/2} (1 - \theta_1)^{-1/2}, \\ M_{D_0}(\theta_2) &= [\cosh(\lambda_0)]^{-1/2}, \end{aligned}$$

where  $\lambda_0 = \sqrt{-2\theta_2}$ .

(b) Let  $[a, b] = [(j-1)/m, j/m]$  so that  $N_c = \int_{(j-1)/m}^{j/m} J_c(r) dW(r)$  and  $D_c = \int_{(j-1)/m}^{j/m} J_c(r)^2 dr$  ( $j = 1, \dots, m$ ). Then:

$$\begin{aligned} M_c(\theta_1, \theta_2) &= \exp\left(-\frac{(\theta_1 + c)}{2m}\right) \left[ \cosh\left(\frac{\lambda}{m}\right) - \frac{1}{\lambda} \left[ \theta_1 + c + v_{j-1}^2 (\theta_1^2 + 2\theta_2) \right] \sinh\left(\frac{\lambda}{m}\right) \right]^{-1/2}, \\ M_{N_c}(\theta_1) &= \exp\left(-\frac{(\theta_1 + c)}{2m}\right) \left[ \cosh\left(\frac{\lambda_1}{m}\right) - \frac{1}{\lambda_1} (\theta_1 + c + v_{j-1}^2 \theta_1^2) \sinh\left(\frac{\lambda_1}{m}\right) \right]^{-1/2}, \end{aligned}$$



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

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

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

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

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

The results in Part (a) of the corollary are relevant in the full-sample case, and the result for  $M_0(\theta_1, \theta_2)$  goes back to [White \(1958\)](#). The results in Part (b) of the corollary are pertinent to the sub-sampling issues being investigated here in the case of a near-unit root, with the unit root ( $c = 0$ ) result for  $M_0(\theta_1, \theta_2)$  having been first derived by [Chambers and Kyriacou \(2013\)](#).

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:

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

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 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:

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

where  $\lambda$  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.,

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

this follows straightforwardly from (14). It is also of interest to examine what happens when the local-to-unity parameter  $c \rightarrow -\infty$ , 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:

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

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

- (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  $\ell$  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  $c \rightarrow -\infty$ ) 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

The discussion following Theorem 2 indicates that the weights defining an optimal jackknife estimator, which removes first-order bias in the local-to-unity setting, depend on the quantities:

$$\begin{aligned} \mu_c(\eta) &= E(Z_c(\eta)) = \mu_c(1) + \lambda_1(\eta) E \left( \frac{1}{D_c} \right), \\ \mu_{c,j}(\eta) &= E(Z_{c,j}(\eta)) = \mu_{c,j}(1) + \lambda_2(\eta, m) E \left( \frac{1}{D_{c,j}} \right), \quad j = 1, \dots, m, \end{aligned}$$

where  $\lambda_1(\eta) = (1 - \eta)/2$ ,  $\lambda_2(\eta, m) = \lambda_1(\eta)/m^2$  and:

$$D_c = \int_0^1 J_c^2, \quad D_{c,j} = \int_{(j-1)/m}^{j/m} J_c^2, \quad j = 1, \dots, m.$$

In particular, Part (c) of Theorem 3 can be used to evaluate the quantities:

$$\mu_c(1) = E \left( \frac{N_c}{D_c} \right), \quad \mu_{c,j}(1) = E \left( \frac{N_{c,j}}{m D_{c,j}} \right), \quad j = 1, \dots, m,$$

where we have defined:

$$N_c = \int_0^1 J_c dW, \quad N_{c,j} = \int_{(j-1)/m}^{j/m} J_c dW, \quad j = 1, \dots, m.$$

The relevant MGFs for evaluating  $\mu_c(1)$  and  $\mu_{c,j}(1)$  are given in the corollary to Theorem 3.

Table 1 contains the values of  $\mu_{c,j}(1)$  for values of  $m$  and  $c$  as follows:  $m = \{1, 2, 3, 4, 6, 8, 12\}$  and  $c = \{-50, -20, -10, -5, -1, 0, 1\}$ .<sup>2</sup> The entries for  $m = 1$  correspond to  $\mu_c(1) = E(Z_c(1))$  in view of the distributional equivalence of  $Z_c(1)$  and  $Z_{c,1}(1)$  discussed following the corollary. 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. The values of the sub-sample expectations when  $c = 0$  are seen from Table 1 to be independent of  $m$  and to increase with  $j$ . Note that  $\mu_{0,1}(1) = -1.7814$  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  $\mu_{0,j}(\eta)$  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  $\mu_{c,j}(\eta)$  can be used to define optimal weights for the jackknife estimator that achieve the aim of first-order bias removal in the presence of a near-unit root. The result is presented in Theorem 5.

**Theorem 5.** Let  $\bar{\mu}_c(\eta) = \mu_c(\eta) - \sum_{j=1}^m \mu_{c,j}(\eta)$ . Then, under Assumption 1, an optimal jackknife estimator is given by:

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

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

Theorem 5 shows the form of the optimal weights for the jackknife estimator when the process (1) has a near-unit root. It can be seen that the weights depend not only on the value of  $c$ , but also on the value of  $\eta$ , both of which are unknown in practice. The authors in Chambers and Kyriacou (2013) and Chen and Yu (2015) have emphasised the case  $c = 0$  and  $\eta = 1$  and have reported simulation results highlighting the good bias-reduction properties of appropriate jackknife estimators in that case. When  $c \neq 0$  and  $\eta = 1$ , the optimal weights in Theorem 5 simplify to:

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

The values of  $\mu_{c,j}(1)$  in Table 1 can be utilised to derive these optimal weights for the jackknife estimator in this case; these are reported in Table 2 for the values of  $m$  and  $c$  used in Table 1, along with the values of the standard weights that are applicable in stationary autoregressions. The entries in Table 1 show 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 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  $\mu_{c,j}(1)$  by examining the appropriate limits as  $c \rightarrow -\infty$ , although we do not pursue such an investigation here.

The relationship between the optimal weights when  $\eta = 1$  and when  $\eta \neq 1$  is not straightforward. Noting that:

$$\bar{\mu}_c(\eta) = \bar{\mu}_c(1) + \lambda_1(\eta)E\left(\frac{1}{D_c}\right) - \lambda_2(\eta, m)\sum_{j=1}^m E\left(\frac{1}{D_{c,j}}\right)$$

and that:

$$\sum_{j=1}^m \mu_{c,j}(\eta) = \sum_{j=1}^m \mu_{c,j}(1) + \lambda_2(\eta, m)\sum_{j=1}^m E\left(\frac{1}{D_{c,j}}\right)$$

<sup>2</sup> The integrals were computed numerically using an adaptive quadrature method in the `integrate1d` routine in Gauss 17.

we find that:

$$w_{1,c}^*(\eta) = - \frac{\sum_{j=1}^m \mu_{c,j}(1) + \lambda_2(\eta, m) \sum_{j=1}^m E\left(\frac{1}{D_{c,j}}\right)}{\bar{\mu}_c(1) + \lambda_1(\eta) E\left(\frac{1}{D_c}\right) - \lambda_2(\eta, m) \sum_{j=1}^m E\left(\frac{1}{D_{c,j}}\right)}.$$

This expression can be manipulated to write  $w_{1,c}^*(\eta)$  explicitly in terms of  $w_{1,c}^*(1)$  as follows:

$$w_{1,c}^*(\eta) = \frac{w_{1,c}^*(1) - \frac{\lambda_2(\eta, m)}{\bar{\mu}_c(1)} \sum_{j=1}^m E\left(\frac{1}{D_{c,j}}\right)}{1 + \frac{\lambda_1(\eta)}{\bar{\mu}_c(1)} E\left(\frac{1}{D_c}\right) - \frac{\lambda_2(\eta, m)}{\bar{\mu}_c(1)} \sum_{j=1}^m E\left(\frac{1}{D_{c,j}}\right)}. \quad (15)$$

The second weight is obtained simply as  $w_{2,c}^*(\eta) = 1 - w_{1,c}^*(\eta)$ . In situations where  $\eta \neq 1$ , which essentially reflects cases where  $u_t$  does not have a white noise structure, the optimal weights can be obtained from the entries in Table 2 (at least for the relevant values of  $c$ ) using (15), but knowledge is still required not only of  $c$  and  $\eta$ , but also the expectations of the inverses of  $D_c$  and  $D_{c,j}$ . The latter can be computed numerically; Equation (2.3) of Meng (2005) shows that:

$$E\left(\frac{1}{D_c}\right) = \int_0^\infty M_{D_c}(-\theta_2) d\theta_2, \quad E\left(\frac{1}{D_{c,j}}\right) = \int_0^\infty M_{D_{c,j}}(-\theta_2) d\theta_2, \quad j = 1, \dots, m,$$

where  $M_{D_c}(\theta_2)$  and  $M_{D_{c,j}}(\theta_2)$  are the MGFs of  $D_c$  and  $D_{c,j}$ , respectively, which can be obtained from the corollary to Theorem 3.

**Table 1.** Values of  $\mu_{c,j}(1) = E(Z_{c,j}(1))$ .

$j \setminus c$	−50	−20	−10	−5	−1	0	1
$m = 1$							
1	−1.9995	−1.9972	−1.9912	−1.9758	−1.8818	−1.7814	−1.5811
$m = 2$							
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	−0.8920
$m = 3$							
1	−1.9962	−1.9838	−1.9595	−1.9175	−1.8234	−1.7814	−1.7283
2	−1.9412	−1.8613	−1.7502	−1.5921	−1.2722	−1.1382	−0.9791
3	−1.9412	−1.8613	−1.7500	−1.5845	−1.1515	−0.9319	−0.6759
$m = 4$							
1	−1.9939	−1.9758	−1.9439	−1.8973	−1.8138	−1.7814	−1.7427
2	−1.9225	−1.8214	−1.6891	−1.5210	−1.2411	−1.1382	−1.0210
3	−1.9225	−1.8214	−1.6879	−1.5021	−1.1016	−0.9319	−0.7410
4	−1.9225	−1.8214	−1.6879	−1.5006	−1.0396	−0.8143	−0.5643
$m = 6$							
1	−1.9884	−1.9594	−1.9175	−1.8698	−1.8037	−1.7814	−1.7564
2	−1.8867	−1.7502	−1.5921	−1.4268	−1.2085	−1.1382	−1.0616
3	−1.8867	−1.7500	−1.5845	−1.3812	−1.0482	−0.9319	−0.8059
4	−1.8867	−1.7500	−1.5843	−1.3732	−0.9697	−0.8143	−0.6472
5	−1.8867	−1.7500	−1.5842	−1.3717	−0.9243	−0.7348	−0.5331
6	−1.8867	−1.7500	−1.5842	−1.3715	−0.8958	−0.6761	−0.4450
$m = 8$							
1	−1.9823	−1.9439	−1.8973	−1.8526	−1.7984	−1.7814	−1.7629
2	−1.8530	−1.6891	−1.5210	−1.3686	−1.1915	−1.1382	−1.0813
3	−1.8530	−1.6879	−1.5021	−1.2991	−1.0203	−0.9319	−0.8381
4	−1.8530	−1.6879	−1.5006	−1.2815	−0.9326	−0.8143	−0.6893
5	−1.8530	−1.6879	−1.5005	−1.2766	−0.8795	−0.7348	−0.5829
6	−1.8530	−1.6879	−1.5005	−1.2752	−0.8444	−0.6761	−0.5008
7	−1.8530	−1.6879	−1.5005	−1.2748	−0.8201	−0.6302	−0.4345
8	−1.8530	−1.6879	−1.5005	−1.2747	−0.8027	−0.5931	−0.3793
$m = 12$							
1	−1.9693	−1.9175	−1.8698	−1.8324	−1.7929	−1.7814	−1.7693
2	−1.7916	−1.5921	−1.4268	−1.3016	−1.1742	−1.1382	−1.1007
3	−1.7916	−1.5845	−1.3812	−1.1979	−0.9916	−0.9319	−0.8698
4	−1.7916	−1.5842	−1.3732	−1.1612	−0.8943	−0.8143	−0.7313
5	−1.7916	−1.5842	−1.3717	−1.1464	−0.8328	−0.7348	−0.6335
6	−1.7916	−1.5842	−1.3715	−1.1403	−0.7904	−0.6761	−0.5585
7	−1.7916	−1.5842	−1.3714	−1.1376	−0.7595	−0.6302	−0.4981
8	−1.7916	−1.5842	−1.3714	−1.1365	−0.7362	−0.5931	−0.4477
9	−1.7916	−1.5842	−1.3714	−1.1360	−0.7183	−0.5622	−0.4047
10	−1.7916	−1.5842	−1.3714	−1.1358	−0.7041	−0.5358	−0.3674
11	−1.7916	−1.5842	−1.3714	−1.1357	−0.6928	−0.5131	−0.3346
12	−1.7916	−1.5842	−1.3714	−1.1356	−0.6837	−0.4931	−0.3055

**Table 2.** Values of standard and optimal jackknife weights ( $\eta = 1$ ).

$m :$	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: $c = -50$						
$w_{1,c}^*(1)$	2.0206	1.5156	1.3470	1.2122	1.1544	1.1016
$w_{2,c}^*(1)$	−1.0206	−0.5156	−0.3470	−0.2122	−0.1544	−0.1016
Optimal weights: $c = -20$						
$w_{1,c}^*(1)$	2.0521	1.5385	1.3670	1.2292	1.1698	1.1151
$w_{2,c}^*(1)$	−1.0521	−0.5385	−0.3670	−0.2292	−0.1698	−0.1151
Optimal weights: $c = -10$						
$w_{1,c}^*(1)$	2.1026	1.5741	1.3969	1.2535	1.1909	1.1325
$w_{2,c}^*(1)$	−1.1026	−0.5741	−0.3969	−0.2535	−0.1909	−0.1325
Optimal weights: $c = -5$						
$w_{1,c}^*(1)$	2.1923	1.6336	1.4445	1.2898	1.2213	1.1565
$w_{2,c}^*(1)$	−1.1923	−0.6336	−0.4445	−0.2898	−0.2213	−0.1565
Optimal weights: $c = -1$						
$w_{1,c}^*(1)$	2.4605	1.7956	1.5678	1.3788	1.2937	1.2117
$w_{2,c}^*(1)$	−1.4605	−0.7956	−0.5678	−0.3788	−0.2937	−0.2117
Optimal weights: $c = 0$						
$w_{1,c}^*(1)$	2.5651	1.8605	1.6176	1.4147	1.3228	1.2337
$w_{2,c}^*(1)$	−1.5651	−0.8605	−0.6176	−0.4147	−0.3228	−0.2337
Optimal weights: $c = 1$						
$w_{1,c}^*(1)$	2.5689	1.8773	1.6355	1.4311	1.3373	1.2455
$w_{2,c}^*(1)$	−1.5689	−0.8773	−0.6355	−0.4311	−0.3373	−0.2455

In practice, however, the values of  $c$  and  $\eta$  are still required in order to construct the optimal estimator. Although the localization parameter  $c$  from the model defined under Assumption 1 is identifiable, it is not possible to estimate it consistently, and attempts to do so require a completely different formulation of the model; see, for example, Phillips et al. (2001), who propose a block local-to-unity framework to consistently estimate  $c$ , although this approach does not appear to have been pursued subsequently. Furthermore  $\eta$  depends on an estimator of the long-run variance  $\sigma^2$ , which is a notoriously difficult quantity to estimate in finite samples. In view of these unresolved challenges and following earlier work on jackknife estimation of autoregressive models with a (near-)unit root, we focus on the case  $\eta = 1$ , but allow  $c \neq 0$  with particular attention paid to unit root and locally-stationary processes, i.e.,  $c \leq 0$ .

Our simulations examine the performance of five estimators<sup>3</sup> of the parameter  $\rho = e^{c/n}$ . The baseline estimator is the OLS estimator in (4), the bias of which the jackknife estimators aim to reduce. Three jackknife estimators with the generic form:

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

<sup>3</sup> The Gauss codes used for the jackknife estimators are available from the authors on request.

are also considered, each differing in the choice of weights  $w_1$ ; in all cases,  $w_2 = 1 - w_1$ . The standard jackknife sets  $w_1 = m / (m - 1)$ ; the optimal jackknife sets  $w_1 = w_{1,c}^*$ ; and the unit root jackknife sets  $w_1 = w_{1,0}^*$ . The standard jackknife removes fully the first-order bias in stationary autoregressions, but does not do so in the near-unit root framework, in which the optimal estimator achieves this goal. However, the optimal estimator is infeasible because it relies on the unknown parameter  $c$ .<sup>4</sup> We therefore also consider the feasible unit root jackknife obtained by setting  $c = 0$ . In addition, we consider the jackknife estimator of [Chen and Yu \(2015\)](#), which is of the form:

$$\hat{\rho}_f^{CY} = w_1 \hat{\rho} + \sum_{j=1}^m w_{2,j} \hat{\rho}_j.$$

The weights are chosen so as to minimise the variance of the estimator in addition to providing bias reduction in the case  $c = 0$ . Because the choice of weights is a more complex problem for this type of jackknife estimator, [Chen and Yu](#) only provide results for the cases  $m = 2$  and  $m = 3$ , in which case the weights are  $w_1 = 2.8390$ ,  $w_{2,1} = 0.6771$ ,  $w_{2,2} = 1.1619$  and  $w_1 = 2.0260$ ,  $w_{2,1} = 0.2087$ ,  $w_{2,2} = 0.3376$ ,  $w_{2,3} = 0.4797$ , respectively; see Table 1 of [Chen and Yu \(2015\)](#).

Table 3 reports the bias of the five estimators obtained from 100,000 replications of the model in Assumption 1 with  $u_t \sim \text{iid } N(0, 1)$  and  $y_0 = 0$  using  $m = 2$  for each of the jackknife estimators; this value has been found to provide particularly good bias reduction in a number of studies, including [Phillips and Yu \(2005\)](#), [Chambers \(2013\)](#), [Chambers and Kyriacou \(2013\)](#) and [Chen and Yu \(2015\)](#). The particular values of  $c$  are  $c = \{-10, -5, -1, 0\}$ , which focus on the pure unit root case, as well as locally stationary processes, and four sample sizes are considered, these being  $n = 24, 48, 96$  and  $192$ . The corresponding values of  $\rho$  are:  $\{0.5833, 0.7917, 0.8958, 0.9479\}$  ( $c = -10$ );  $\{0.7917, 0.8958, 0.9479, 0.9740\}$  ( $c = -5$ );  $\{0.9583, 0.9792, 0.9896, 0.9948\}$  ( $c = -1$ ); and  $\rho = 1$  for all values of  $n$  when  $c = 0$ . The values of  $\rho$  when  $c = -10$  are some way from unity for the smaller sample sizes, which suggests that the standard jackknife might perform well in these cases.

The value of the bias of the estimator producing the minimum (absolute) bias for each  $c$  and  $n$  is highlighted in bold in Table 3. The results show the substantial reduction in bias that can be achieved with jackknife estimators, the superiority of the optimal estimator being apparent as  $c$  becomes more negative, although the unit root jackknife also performs well in terms of bias reduction.

Table 4 contains the corresponding RMSE values for the jackknife estimators using  $m = 2$ , as well as the RMSE corresponding to the RMSE-minimising values of  $m$ , which are typically larger than  $m = 2$  and are also reported in the table. The RMSE value of the estimator producing the minimum RMSE for each  $c$  and  $n$  is highlighted in bold. In fact, the optimal jackknife estimator, although constructed to eliminate first order bias, manages to reduce the OLS estimator's RMSE and outperforms the [Chen and Yu \(2015\)](#) jackknife estimator in both bias and RMSE reduction, although the latter occurs at a larger number of sub-samples. The results show that use of larger values of  $m$  tends to produce smaller RMSE than when  $m = 2$ , and again, the optimal jackknife performs particularly well when  $c$  becomes more negative. The performance of the unit root jackknife is also impressive, suggesting that it is a feasible alternative to the optimal estimator when the value of  $c$  is unknown.

Although in itself important, bias is not the only feature of a distribution that is of interest, and hence, the RMSE values in Table 4 should also be taken into account when assessing the performance of the estimators. The substantial bias reductions obtained with the bias-minimising value of  $m = 2$  come at the cost of a larger variance that ultimately feeds through into a larger RMSE compared with the OLS estimator  $\hat{\rho}$ . This can be offset, however, by using the larger RMSE-minimising values of  $m$  that, despite having a larger bias than when  $m = 2$ , are nevertheless able to reduce the variance sufficiently to result in a smaller RMSE than  $\hat{\rho}$ .

<sup>4</sup> In the simulations, we are taking  $\eta = 1$  as known.



**Table 3.** Bias of OLS and jackknife estimators ( $m = 2$ ) when  $\eta = 1$ .

Estimator	$n = 24$	$n = 48$	$n = 96$	$n = 192$
$c = 0$				
OLS	−0.0663	−0.0351	−0.0180	−0.0091
Standard jackknife	−0.0343	−0.0155	−0.0072	−0.0035
Optimal/unit root jackknife	<b>−0.0163</b>	<b>−0.0045</b>	<b>−0.0011</b>	<b>−0.0003</b>
Chen-Yu jackknife	−0.0264	−0.0144	−0.0110	−0.0102
$c = -1$				
OLS	−0.0675	−0.0365	−0.0188	−0.0096
Standard jackknife	−0.0323	−0.0145	−0.0067	−0.0032
Optimal jackknife	−0.0161	−0.0044	−0.0011	<b>−0.0002</b>
Unit root jackknife	<b>−0.0124</b>	<b>−0.0021</b>	<b>0.0002</b>	0.0004
Chen-Yu jackknife	−0.0190	−0.0099	−0.0087	−0.0089
$c = -5$				
OLS	−0.0589	−0.0350	−0.0188	−0.0098
Standard jackknife	−0.0193	−0.0088	−0.0038	−0.0017
Optimal jackknife	−0.0116	−0.0037	<b>−0.0009</b>	<b>−0.0001</b>
Unit root jackknife	<b>0.0031</b>	0.0061	0.0048	0.0029
Chen-Yu jackknife	0.0037	<b>0.0027</b>	−0.0017	−0.0051
$c = -10$				
OLS	−0.0437	−0.0310	−0.0178	−0.0095
Standard jackknife	−0.0114	−0.0055	−0.0022	−0.0009
Optimal jackknife	−0.0080	<b>−0.0029</b>	<b>−0.0006</b>	<b>−0.0001</b>
Unit root jackknife	<b>0.0069</b>	0.0089	0.0066	0.0039
Chen-Yu jackknife	0.0090	0.0071	0.0013	−0.0035

In order to assess the robustness of the jackknife estimators, some additional bias results are presented in Table 5 that correspond to values of  $\eta < 1$ , while the estimators are based on the assumption that  $\eta = 1$ , as in the preceding simulations.<sup>5</sup> The results correspond to two different specifications for  $u_t$  that enable data to be generated that are consistent with different values of  $\eta$ . The first specifies  $u_t$  to be a first-order moving average (MA(1)) process, so that  $u_t = \epsilon_t + \theta\epsilon_{t-1}$  where  $\epsilon_t \sim \text{iid } N(0, 1)$ ; in this case  $\eta = (1 + \theta^2)/(1 + \theta)^2$ . The second specification is a first-order autoregressive (AR(1)) process of the form  $u_t = \phi u_{t-1} + \epsilon_t$ , in which case  $\eta = (1 - \phi)^2/(1 - \phi^2)$ . In the MA(1) case, we have chosen  $\theta = 0.5$  in order to give an intermediate value of  $\eta = 0.5556$ , while in the AR(1) case, we have chosen  $\phi = 0.9$  to give a small value of  $\eta = 0.0526$ . As in Table 3, the value of the bias of the estimator producing the minimum (absolute) bias for each  $c$  and  $n$  is highlighted in bold. Table 5 shows, in the MA case, that the jackknife estimators are able to reduce bias when  $c = 0$ , but none of them is able to do so when  $c = -1$  or  $c = -5$ . In the AR case, with a smaller value of  $\eta$ , the jackknife estimators are still able to deliver bias reduction, albeit to a lesser extent than when  $\eta = 1$ , and it is the unit root jackknife of Chambers and Kyriacou (2013) that achieves the greatest bias reduction in this case. These results are indicative of the importance of knowing  $\eta$  and suggest that developing methods to allow for  $\eta \neq 1$  is important from an empirical viewpoint.

<sup>5</sup> We thank a referee for suggesting that we investigate the performance of the estimators when  $\eta < 1$ .

**Table 4.** RMSE of OLS and jackknife estimators when  $\eta = 1$ .

Estimator	$n = 24$	$n = 48$	$n = 96$	$n = 192$
$c = 0$				
OLS	0.1366	0.0719	0.0371	0.0187
Standard jackknife ( $m = 2$ )	0.1482	0.0766	0.0396	0.0199
Standard jackknife ( $m = 4, 6, 6, 8$ )	<b>0.1310</b>	0.0659	0.0333	0.0165
Optimal/unit root jackknife ( $m = 2$ )	0.1753	0.0915	0.0479	0.0242
Optimal/unit root jackknife ( $m = 4, 8, 12, 12$ )	0.1383	<b>0.0642</b>	<b>0.0312</b>	<b>0.0154</b>
Chen-Yu jackknife ( $m = 2$ )	0.1640	0.0864	0.0462	0.0249
Chen-Yu jackknife ( $m = 3$ )	0.1392	0.0719	0.0374	0.0188
$c = -1$				
OLS	0.1428	0.0762	0.0396	0.0200
Standard jackknife ( $m = 2$ )	0.1524	0.0797	0.0415	0.0209
Standard jackknife ( $m = 4, 6, 6, 8$ )	<b>0.1368</b>	0.0698	0.0355	0.0176
Optimal jackknife ( $m = 2$ )	0.1724	0.0908	0.0477	0.0242
Optimal jackknife ( $m = 4, 8, 12, 12$ )	0.1416	<b>0.0679</b>	0.0334	<b>0.0165</b>
Unit root jackknife ( $m = 2$ )	0.1778	0.0939	0.0495	0.0251
Unit root jackknife ( $m = 4, 8, 12, 12$ )	0.1435	0.0680	<b>0.0333</b>	<b>0.0165</b>
Chen-Yu jackknife ( $m = 2$ )	0.1710	0.0916	0.0489	0.0262
Chen-Yu jackknife ( $m = 3$ )	0.1477	0.0778	0.0408	0.0207
$c = -5$				
OLS	0.1626	0.0901	0.0476	0.0243
Standard jackknife ( $m = 2$ )	0.1745	0.0944	0.0498	0.0253
Standard jackknife ( $m = 6, 6, 8, 8$ )	<b>0.1615</b>	0.0855	0.0442	0.0223
Optimal jackknife ( $m = 2$ )	0.1813	0.0982	0.0520	0.0265
Optimal jackknife ( $m = 6, 8, 12, 12$ )	0.1641	<b>0.0847</b>	<b>0.0432</b>	<b>0.0216</b>
Unit root jackknife ( $m = 2$ )	0.1975	0.1078	0.0576	0.0295
Unit root jackknife ( $m = 6, 12, 12, 12$ )	0.1710	0.0852	0.0433	0.0219
Chen-Yu jackknife ( $m = 2$ )	0.2066	0.1138	0.0610	0.0318
Chen-Yu jackknife ( $m = 3$ )	0.1857	0.1014	0.0540	0.0279
$c = -10$				
OLS	<b>0.1809</b>	0.1037	0.0558	0.0288
Standard jackknife ( $m = 2$ )	0.1971	0.1096	0.0584	0.0300
Standard jackknife ( $m = 8, 8, 12, 12$ )	0.1853	0.1016	0.0534	0.0272
Optimal jackknife ( $m = 2$ )	0.2003	0.1114	0.0595	0.0306
Optimal jackknife ( $m = 4, 12, 12, 12$ )	0.1877	<b>0.1015</b>	<b>0.0530</b>	<b>0.0270</b>
Unit root jackknife ( $m = 2$ )	0.2175	0.1217	0.0656	0.0340
Unit root jackknife ( $m = 6, 12, 12, 12$ )	0.1985	0.1032	0.0539	0.0277
Chen-Yu jackknife ( $m = 2$ )	0.2328	0.1315	0.0711	0.0372
Chen-Yu jackknife ( $m = 3$ )	0.2151	0.1202	0.0649	0.0338

**Table 5.** Bias of OLS and jackknife estimators ( $m = 2$ ) when  $\eta \neq 1$ .

Estimator	$n = 24$	$n = 48$	$n = 96$	$n = 192$
$\eta = 0.5556$ (MA case)				
$c = 0$				
OLS	−0.0191	−0.0105	−0.0054	−0.0028
Standard jackknife	−0.0059	<b>−0.0018</b>	<b>−0.0006</b>	<b>−0.0002</b>
Optimal/unit root jackknife ( $\eta = 1$ )	<b>0.0016</b>	0.0031	0.0022	0.0012
Chen-Yu jackknife	0.0062	0.0054	0.0034	0.0018
$c = -1$				
OLS	<b>−0.0060</b>	<b>−0.0040</b>	<b>−0.0022</b>	<b>−0.0012</b>
Standard jackknife	0.0110	0.0068	0.0038	0.0020
Optimal jackknife ( $\eta = 1$ )	0.0188	0.0118	0.0065	0.0034
Unit root jackknife	0.0206	0.0129	0.0072	0.0037
Chen-Yu jackknife	0.0275	0.0167	0.0091	0.0048
$c = -5$				
OLS	<b>0.0631</b>	<b>0.0304</b>	<b>0.0150</b>	<b>0.0074</b>
Standard jackknife	0.0885	0.0457	0.0234	0.0118
Optimal jackknife ( $\eta = 1$ )	0.0934	0.0486	0.0250	0.0127
Unit root jackknife	0.1028	0.0543	0.0281	0.0143
Chen-Yu jackknife	0.1129	0.0601	0.0312	0.0159
$\eta = 0.0526$ (AR case)				
$c = 0$				
OLS	0.0524	0.0237	0.0107	0.0049
Standard jackknife	0.0306	0.0137	0.0066	0.0033
Optimal/unit root jackknife ( $\eta = 1$ )	<b>0.0183</b>	<b>0.0081</b>	<b>0.0043</b>	<b>0.0024</b>
Chen-Yu jackknife	0.0293	0.0143	0.0076	0.0039
$c = -1$				
OLS	0.0801	0.0386	0.0184	0.0089
Standard jackknife	0.0624	0.0305	0.0153	0.0077
Optimal jackknife ( $\eta = 1$ )	0.0542	0.0268	0.0139	0.0071
Unit root jackknife	<b>0.0524</b>	<b>0.0260</b>	<b>0.0135</b>	<b>0.0070</b>
Chen-Yu jackknife	0.0645	0.0331	0.0172	0.0088
$c = -5$				
OLS	0.2116	0.1086	0.0547	0.0273
Standard jackknife	0.2096	0.1070	0.0541	0.0271
Optimal jackknife ( $\eta = 1$ )	0.2092	0.1067	0.0539	0.0271
Unit root jackknife	<b>0.2084</b>	<b>0.1061</b>	<b>0.0537</b>	<b>0.0270</b>
Chen-Yu jackknife	0.2212	0.1138	0.0575	0.0289

## 5. 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 then 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 at finite sample bias reduction in a simulation study and, with a suitable choice of the number of sub-samples, is shown to be able to reduce the overall finite sample RMSE, as well.

The theoretical findings in Sections 3 and 4 show how first-order approximations on sub-sample estimators can be used along with the well-known full-sample results of Phillips (1987a) for finite-sample refinements. The jackknife uses analytical (rather than simulation-based) results to

achieve bias reduction at minimal computational cost along the same lines as indirect inference methods based on analytical approximations in Phillips (2012) and Kyriacou et al. (2017). Apart from computational simplicity, an evident advantage of analytical-based methods over simulation-based alternatives such as bootstrap or (traditional, simulation-based) indirect inference methods is that they require no distributional assumptions on the error term.

Despite its success in achieving substantial bias reduction in finite samples, as shown in the simulations, a shortcoming 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, as well as on a quantity related to the long-run variance of the disturbances.<sup>6</sup> However, our theoretical results in Sections 3 and 4 reveal precisely how these quantities affect the optimal weights and therefore can, in principle, be used to guide further research into the development of a feasible data-driven version of the jackknife within this framework. Such further work is potentially useful in view of the simulations in Tables 3 and 4 highlighting that (feasible) jackknife estimators are an effective bias and RMSE reduction tool in a local unit root setting, even if they do not fully remove first order bias. Moreover, the results obtained in Theorems 1–4 can be utilised in a wide range of sub-sampling situations outside that of jackknife estimation itself.

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  $3$ , although extending their approach to arbitrary  $c$  and  $m$  represents a challenging task. However, considerable progress has been made in this direction by Stoykov (2017), who builds upon our results and also proposes a two-step jackknife estimator that incorporates an estimate of  $c$  to determine the jackknife weights. The estimation model could also be extended to include an intercept and/or a time trend. The presence of an intercept will affect the limit distributions by replacing the Ornstein–Uhlenbeck processes by demeaned versions thereof, which will also have an effect on the finite sample biases. Such effects have been investigated by Stoykov (2017), who shows that substantial reductions in bias can still be achieved by jackknife methods. 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 multivariate possibilities could be envisaged.

**Acknowledgments:** We thank the Editors, two anonymous referees and Marian Stoykov for helpful comments on this piece of work. 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.

**Author Contributions:** Both authors contributed equally to the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A.

**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:

$$y_t = \sum_{j=1}^t e^{c(t-j)/n} u_j + e^{ct/n} y_0. \quad (\text{A1})$$

<sup>6</sup> It should also be noted that we have assumed  $y_0 = 0$  in deriving the jackknife weights, whereas the expansion in Theorem 2 suggests that the bias functions (and, hence, the jackknife weights) will depend in a non-trivial and more complicated way on  $y_0$  when  $y_0 \neq 0$ . Although we have not investigated the issue further here, the results in Kyriacou (2011) suggest that jackknife methods can still provide bias reduction even when  $y_0 \neq 0$ .

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

$$X_n(r) = \frac{1}{\sqrt{n}} S_{[nr]} = \frac{1}{\sqrt{n}} S_{j-1}, \quad \frac{j-1}{n} \leq r < \frac{j}{n}. \quad (\text{A2})$$

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

(a) In view of (A1) and (A2), 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:

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

Now, as  $n \rightarrow \infty$ ,

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

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

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

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:

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

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 (7).  $\square$

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

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

Next, as before, we have:

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

Now:

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

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

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

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

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

where  $\xi_{1j}$  ( $j = 1, \dots, m$ ) is defined by:

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

The stated distribution of  $\xi_{1j}$  then follows using the property that:

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

to calculate the variances and covariances; see Perron (1991, p. 234). In particular:

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

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

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



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

**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:

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

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

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

Let  $\mu_{J_c}$  and  $\mu_Y$  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  $f(J_c)$ ,

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

Using the change of measure, we obtain:

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

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

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

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

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

As the parameter  $\lambda$  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  $Y(b)|\mathcal{F}_0^a \sim N(\mu, \omega^2)$ , where:

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

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

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

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

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

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

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:

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

With the change of measure, the expectation of interest becomes:

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

However,  $\eta$  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:

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

Now,  $V(a) = \int_0^a e^{c(a-r)} dW(r)$ , and so,  $V(a) \sim N(0, v^2)$  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:

$$H_c(\theta_1, \theta_2) = \exp(-(b - a)\lambda)(1 - \delta\omega^2)(1 - \delta(k-1)v^2). \quad (\text{A3})$$

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) The individual MGFs follow straightforwardly from Part (a), noting that  $M_{N_c}(\theta_1) = M_c(\theta_1, 0)$  and  $M_{D_c}(\theta_2) = M_c(0, \theta_2)$ .

(c) 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  $\lambda$ ; this results in:

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

It is also convenient to define:

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

Combining the results above yields:

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

Integrating with respect to  $\theta_2$  yields the result in the theorem.

**Proof of Corollary to Theorem 3.** The results follow from Theorem 3 noting that:

- (a)  $b - a = 1$  and  $v^2 = 0$ ;
- (b)  $b - a = 1/m$  and  $\lim_{c \rightarrow 0} v_{j-1}^2 = (j-1)/m$ .

**Derivation of (14).** From (A3), 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).  $\square$

**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  $\lambda$ ,  $\delta$  and  $v^2$  as  $c \rightarrow -\infty$ . The following asymptotic expansions facilitate this:

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

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

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

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

Combining these results, we find that:

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

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).  $\square$

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

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

where  $\mu_c(\eta)$  is defined in the theorem. From the definition of  $\hat{\rho}_j^*$ , taking expectations yields:

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

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

- (i)  $w_{1,c}^*(\eta) + w_{2,c}^*(\eta) = 1$ , and
- (ii)  $w_{1,c}^*(\eta) \mu_c(\eta) + w_{2,c}^*(\eta) \sum_{j=1}^m \mu_{c,j}(\eta) = 0$ .

Solving these two conditions simultaneously yields the stated weights.  $\square$

## References

- Abadir, Karim M. 1993. The limiting distribution of the autocorrelation coefficient under a unit root. *Annals of Statistics* 21: 1058–70.
- Chambers, Marcus J. 2013. Jackknife estimation and inference in stationary autoregressive models. *Journal of Econometrics* 172: 142–57.
- Chambers, Marcus J. 2015. A jackknife correction to a test for cointegration rank. *Econometrics* 3: 355–75.
- Chambers, Marcus J., and Maria Kyriacou. 2013. Jackknife estimation with a unit root. *Statistics and Probability Letters* 83: 1677–82.
- Chan, Ngai H., and Ching-Zong Wei. 1987. Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics* 15: 1050–63.
- Chen, Ye, and Jun Yu. 2015. Optimal jackknife for unit root models. *Statistics and Probability Letters* 99: 135–42.
- Gonzalo, Jesus, and Jean-Yves Pitarakis. 1998. On the exact moments of nonstandard asymptotic distributions in an unstable AR(1) with dependent errors. *International Economic Review* 39: 71–88.
- Kruse, Robinson, and Hendrik Kaufmann, 2015. Bias-corrected estimation in mildly explosive autoregressions. Paper presented at Annual Conference 2015: Economic Development—Theory and Policy, Verein für Socialpolitik/German Economic Association, Muenster, Germany, September 6–9.
- Kyriacou, Maria. 2011. *Jackknife Estimation and Inference in Non-stationary Autoregression*. Ph.D. thesis, University of Essex, Colchester, UK.
- Kyriacou, Maria, Peter C. B. Phillips, and Francesca Rossi. 2017. Indirect inference in spatial autoregression. *The Econometrics Journal* 20: 168–89.
- Magnus, Jan R. 1986. The exact moments of a ratio of quadratic forms in normal variables. *Annales d'Économie et de Statistique* 4: 95–109.
- Meng, Xiao-Li. 2005. From unit root to Stein's estimator to Fisher's  $k$  statistics: If you have a moment, I can tell you more. *Statistical Science* 20: 141–62.
- Park, Joon. 2006. A bootstrap theory for weakly integrated processes. *Journal of Econometrics* 133: 639–72.
- Perron, Pierre. 1989. The calculation of the limiting distribution of the least-squares estimator in a near-integrated model. *Econometric Theory* 5: 241–55.
- Perron, Pierre. 1991. A continuous time approximation to the unstable first-order autoregressive process: The case without an intercept. *Econometrica* 59: 211–36.
- Perron, Pierre. 1996. The adequacy of asymptotic approximations in the near-integrated autoregressive model with dependent errors. *Journal of Econometrics* 70: 317–50.
- Phillips, Peter C. B. 1987a. Towards a unified asymptotic theory for autoregression. *Biometrika* 74: 535–47.
- Phillips, Peter C. B. 1987b. Time series regression with a unit root. *Econometrica* 55: 277–301.
- Phillips, Peter C. B. 2012. Folklore theorems, implicit maps, and indirect inference. *Econometrica* 80: 425–54.
- Phillips, Peter C. B. 2014. On confidence intervals for autoregressive roots and predictive regression. *Econometrica* 82: 1177–95.
- Phillips, Peter C. B., Hyungsik Roger Moon, and Zhijie Xiao. 2001. How to estimate autoregressive roots near unity. *Econometric Theory* 17: 26–69.
- Phillips, Peter C. B., and Jun Yu. 2005. Jackknifing bond option prices. *Review of Financial Studies* 18: 707–42.
- Quenouille, Maurice H. 1956. Notes on bias in estimation. *Biometrika* 43: 353–60.
- Stoykov, Marian Z. 2017. *Optimal Jackknife Estimation of Local to Unit Root Models*. Colchester: Essex Business School, preprint.
- Tanaka, Katsuto. 1996. *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*. New York: Wiley.
- Tukey, John W. 1958. Bias and confidence in not-quite large samples. *Annals of Mathematical Statistics* 29: 614.
- White, John S. 1958. The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics* 29: 1188–97.

