Statistics and Probability Letters 99 (2015) 135–142

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Optimal jackknife for unit root models

Ye Chen[a,b,*], Jun Yu[b]

[a] International School of Economics and Management, Capital University of Economics and Business, Beijing 100070, PR China
[b] School of Economics, Singapore Management University, Singapore 178903, Singapore

ARTICLE INFO

Article history:
Received 21 July 2014
Received in revised form 9 December 2014
Accepted 11 December 2014
Available online 26 December 2014

Keywords:
Bias reduction
Variance reduction
Jackknife
Autoregression

ABSTRACT

A new jackknife method is introduced to remove the first order bias in unit root models. It is optimal in the sense that it minimizes the variance among all the jackknife estimators of the form considered in Phillips and Yu (2005) and Chambers and Kyriacou (2013) after the number of subsamples is selected. Simulations show that the new jackknife reduces the variance of that of Chambers and Kyriacou by about 10% for any selected number of subsamples without compromising bias reduction. The results continue to hold true in near unit root models.

1. Introduction

Many estimators suffer from finite sample bias in dynamic models. Subsampling methods have been found useful to reduce the bias. The jackknifing method of Quenouille (1949) is a widely used approach to achieving bias reduction. The basic idea of this method is to use a subsampling technique to estimate the bias, and then to subtract the bias estimate from the initial (biased) estimator. The bias estimate is formed through linear combinations of full sample estimate and subsample estimates. Under mild conditions, the jackknife estimator can remove the first order bias.

The bootstrap method of Efron (1979) generalizes the jackknife for bias reduction. It was subsequently found that the bootstrap was more effective in reducing the bias than the jackknife; see for example, Hall (1992) and Shao and Tu (1995) for more detailed discussions. Nevertheless, the jackknife remains appealing for its ease of implementation. In addition, it is computationally not much more expensive than the initial estimator. Moreover, it is often found that the jackknife continues to reduce the bias when the error distribution is misspecified; see for example, Phillips and Yu (2005, PY hereafter).

In the context of a discrete time unit root model, Chambers and Kyriacou (2013, CK hereafter) pointed out that the jackknife of PY cannot completely remove the first order bias. A revised jackknife was proposed in CK and was shown to perform better than the PY estimator for bias reduction. While the jackknife of CK reduces the bias of the original estimator, it tends to increase its variance, as other jackknife estimators. The variance increases over the original estimator because it is constructed from the subsample estimates that have a larger variance.

In this paper, we propose an improved jackknife estimator for unit root models. Our estimator is optimal in the sense that it not only removes the first order bias, but also leads to a smaller variance than that of CK. Unlike the estimators of CK, the weights are not the same across different subsamples. Optimal weights are derived and the finite sample performance of the new estimator is examined. It is found that the optimal jackknife estimator offers about 10% reduction in variance over the CK estimator without compromising bias reduction. When the model is known to have a unit root, there is no need to

* Correspondence to: School of Economics, Singapore Management University, 90 Stamford Rd, Singapore 178903, Singapore.
E-mail addresses: chen.ye.2009@phdecons.smu.edu.sg (Y. Chen), yujun@smu.edu.sg (J. Yu).

http://dx.doi.org/10.1016/j.spl.2014.12.014
0167-7152/

estimate the persistent parameter. However, when the root is close to one, we provide evidence that our optimal jackknife continues to work well in finite sample.

Let the parameter of interest be β. Let $\widetilde{\beta}$ be the full sample estimator of β, $\widetilde{\beta}_j$ be the estimator from the jth ($j = 1, \ldots, m$) subsample of length l (i.e., $m \times l = n$ and n is the sample size of the full sample), $\widetilde{\beta}^{PY}$, $\widetilde{\beta}^{CK}$, $\widetilde{\beta}^{CY}$ be the jackknife estimators of PY, CK, and the present paper, respectively. Following CK, we define $Z = \int_0^1 WdW / \int_0^1 W^2$, $Z_j = \int_{(j-1)/m}^{j/m} WdW / \int_{(j-1)/m}^{j/m} W^2$, where W is a standard Brownian motion, $\mu = E(Z)$ and $\mu_j = E(Z_j)$.

2. Optimal jackknife for unit root models

2.1. Jackknife methods of PY and CK

Considering a simple unit root model with initial value $y_0 = O_p(1)$:

$$y_t = \beta y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma_\varepsilon^2),\ t = 1, \ldots, n,\ \text{with } \beta = 1. \tag{1}$$

With the available data $\{y_t\}_{t=0}^n$, the least squares (LS) estimator of β is

$$\widetilde{\beta} = \sum_{t=1}^n y_{t-1} y_t \Big/ \sum_{t=1}^n y_{t-1}^2.$$

When ε_t is normally distributed, $\widetilde{\beta}$ is also the ML estimator of β, conditional on y_0.

Following the original work of Quenouille (1949), PY (2005) suggests splitting the full sample into m subsamples and utilizes the subsample estimators of β to achieve bias reduction with the following formula:

$$\widetilde{\beta}_m^{PY} = \frac{m}{m-1}\widetilde{\beta} - \frac{1}{m-1}\left(\frac{1}{m}\sum_{j=1}^m \widetilde{\beta}_j\right) = \widetilde{\beta} - \frac{1}{m-1}\left(\frac{1}{m}\sum_{j=1}^m \widetilde{\beta}_j - \widetilde{\beta}\right). \tag{2}$$

To check the validity of this jackknife method, consider the following Nagar approximation:

$$E\left(\widetilde{\beta}\right) = \beta + \frac{b_1}{n} + o\left(n^{-1}\right), \qquad E\left(\widetilde{\beta}_j\right) = \beta + \frac{b_1}{l} + o\left(l^{-1}\right), \tag{3}$$

which can be derived from a set of mild conditions, as $n, l \to \infty$. Taking the expectation in (2) and using (3), we have $E\left(\widetilde{\beta}_m^{PY}\right) = \beta + o\left(n^{-1}\right)$, confirming the validity of this jackknife method. In essence, the jackknife method of PY estimates the bias in the initial estimator $\widetilde{\beta}$ by $\frac{1}{m-1}\left(\frac{1}{m}\sum_{j=1}^m \widetilde{\beta}_j - \widetilde{\beta}\right)$. Particularly effective bias reduction can be achieved by splitting the full sample into two subsamples (i.e. $m = 2$) and the estimator becomes:

$$\widetilde{\beta}^{PY} = 2\widetilde{\beta} - \frac{1}{2}\left(\widetilde{\beta}_1 + \widetilde{\beta}_2\right). \tag{4}$$

Both PY and Chambers (2013) have reported evidence to support this method for the purpose of bias reduction in different contexts. $\widetilde{\beta}_m^{PY}$ tends to lead to a larger variance than $\widetilde{\beta}$ because the variance of $\widetilde{\beta}_j$ is larger than that of $\widetilde{\beta}$. So in general there is a bias–variance trade-off.

The Nagar approximation is a general result and can be verified by Sargan's (1976) theorem. Given the mild conditions under which Sargan's theorem holds, it is rather surprising that the standard jackknife fails to remove the first order bias in the unit root model. This failure was first documented in CK (2013). The basic argument of CK is that in (3), b_1 is not constant any more in the unit root model. Instead, it depends on the initial condition. As the initial condition varies across subsamples, the jackknife cannot eliminate the first order bias. Specifically, the limit distribution of $l(\widetilde{\beta}_j - 1)$ is $\int_{(j-1)/m}^{j/m} WdW / \left(m \int_{(j-1)/m}^{j/m} W^2\right)$ whose expectation depends on j. To eliminate the first order asymptotic bias, based on m subsamples, CK proposes the following modified jackknife estimator:

$$\widetilde{\beta}_m^{CK} = b_m^{CK}\widetilde{\beta} - \delta_m^{CK}\sum_{j=1}^m \widetilde{\beta}_j, \tag{5}$$

where

$$b_m^{CK} = \frac{\sum_{j=1}^m \mu_j}{\sum_{j=1}^m \mu_j - \mu}, \qquad \delta_m^{CK} = \frac{\mu}{m\left(\sum_{j=1}^m \mu_j - \mu\right)}. \tag{6}$$

When $\mu_1 = \cdots = \mu_m = \mu$, $b_m^{CK} = m/(m-1) = b_m^{PY}$, and $\delta_m^{CK} = 1/\left(m^2 - m\right)$. Under model (1), CK showed that $\mu = \mu_1 = -1.7814$, $\mu_2 = -1.1382$, $\mu_3 = -0.9319$, $\mu_4 = -0.8143$, etc. That is, the bias becomes smaller and smaller as we go deeper and deeper into subsampling. Substituting these expected values into the formula (6), we can calculate the weights.

Table 1
Weights assigned to the full- and sub-sample estimators for alternative jackknife methods.

Methodology	$m=2$			$m=3$			
	Full sample	First subsample	Second subsample	Full sample	First subsample	Second subsample	Third subsample
Standard jackknife	2.0000	0.5000	0.5000	1.5000	1/6	1/6	1/6
Jackknife of CK	2.5651	0.7825	0.7825	1.8605	0.2868	0.2868	0.2868
Jackknife of CY	2.8390	0.6771	1.1619	2.0260	0.2087	0.3376	0.4797

Table 1 reports the weights when $m = 2$ and $m = 3$. We also report the weights of PY for comparison. An important feature in the CK jackknife is that the weight assigned to all subsample estimates is the same. Among all possible values of m, CK proposed to choose m to minimize the root mean squared errors (RMSE).

2.2. Optimal jackknife

In this paper, we introduce a new jackknife estimator, which not only removes the first order bias but also minimizes its variance for any given m, including the optimal m proposed by CK. To do so, we select the weights, b_m^{CY} and $\{a_{j,m}^{CY}\}_{j=1}^m$, to minimize the variance of the new jackknife estimator defined by $\widetilde{\beta}_m^{CY} = b_m^{CY}\widetilde{\beta} - \sum_{j=1}^m a_{j,m}^{CY}\widetilde{\beta}_j$, i.e.,

$$\min_{b_m^{CY}, \{a_{j,m}^{CY}\}_{j=1}^m} Var\left(\widetilde{\beta}_m^{CY}\right), \tag{7}$$

subject to two constraints:

$$b_m^{CY} = \sum_{j=1}^m a_{j,m}^{CY} + 1, \tag{8}$$

$$b_m^{CY}\mu = m\sum_{j=1}^m a_{j,m}^{CY}\mu_j, \tag{9}$$

where $\mu = \mu_1$. These two constraints are used to ensure the first order bias is fully removed. The first order conditions with respect to $a_{j,m}^{CY}$ are:

$$\begin{aligned} 0 = {} & b_m^{CY}\left[2\frac{m(\mu-\mu_j)}{(m-1)\mu}Var(\widetilde{\beta}) - 2\frac{\mu - m\mu_j}{(m-1)\mu}Cov(\widetilde{\beta},\widetilde{\beta}_1) - 2Cov(\widetilde{\beta},\widetilde{\beta}_j)\right] + a_{1,m}^{CY} \\ & \times\left[2\frac{\mu - m\mu_j}{(m-1)\mu}Var(\widetilde{\beta}_1) - 2\frac{m(\mu-\mu_j)}{(m-1)\mu}Cov(\widetilde{\beta},\widetilde{\beta}_1) + 2Cov(\widetilde{\beta},\widetilde{\beta}_j)\right] + \cdots + \sum_{i=2}^m a_{i,m}^{CY} \\ & \times\left[-2\frac{m(\mu-\mu_i)}{(m-1)\mu}Cov(\widetilde{\beta},\widetilde{\beta}_i) + 2\frac{\mu - m\mu_i}{(m-1)\mu}Cov(\widetilde{\beta}_1,\widetilde{\beta}_i) + 2Cov(\widetilde{\beta}_i,\widetilde{\beta}_j)\right], \end{aligned} \tag{10}$$

for $j = 2, \ldots, m$. In addition, we have:

$$b_m^{CY} = a_{2,m}^{CY}\frac{m(\mu-\mu_2)}{(m-1)\mu} + \cdots + a_{m,m}^{CY}\frac{m(\mu-\mu_m)}{(m-1)\mu} + \frac{m}{m-1},$$

$$a_{1,m}^{CY} = a_{2,m}^{CY}\frac{\mu - m\mu_2}{(m-1)\mu} + \cdots + a_{m,m}^{CY}\frac{\mu - m\mu_m}{(m-1)\mu} + \frac{1}{m-1}.$$

To eliminate the first order bias, one must first obtain $\mu, \mu_2, \ldots, \mu_m$, as CK did. To minimize the variance of the new estimator, one must calculate the exact variances and covariances of the finite sample distributions. However, it is known in the literature that the exact moments are difficult to obtain analytically in dynamic models. To simplify the derivations, we propose to approximate the moments of the finite sample distributions by those of the limit distributions, but will check the quality of these approximations in simulations.

The variances can be computed by combining the techniques of White (1961) and CK. We refer to Chen and Yu (2013, page 8) for more details. Note that:

$$n^2 Var(\widetilde{\beta}) = E\left(\frac{\int_0^1 WdW}{\int_0^1 W^2}\right)^2 - \mu^2 + o(1).$$

Similarly, the variance of the subsample estimators is:

$$l^2 Var(\widetilde{\beta}_j) = E\left(\frac{\int_{(j-1)/m}^{j/m} WdW}{m\int_{(j-1)/m}^{j/m} W^2}\right)^2 - \mu_j^2 + o(1), \quad j = 1, 2, \ldots, m.$$

Table 2
Variances of subsample estimators.

jth subsample	$l^2Var(\widetilde{\beta}_j)$	jth subsample	$l^2Var(\widetilde{\beta}_j)$
1	10.1123	7	2.8375
2	5.3612	8	2.6660
3	4.2839	9	2.5238
4	3.7065	10	2.4034
5	3.3268	11	2.2995
6	3.0507	12	2.2087

Let $N(a,b) = \int_a^b WdW$, $D(a,b) = \int_a^b W^2$ $(0 \leqslant a < b \leqslant 1)$, and $M_{a,b}(\theta_1,\theta_2)$ denote the joint moment generating function (MGF) of $N(a,b)$ and $D(a,b)$. Following Magnus (1986), we use the following expression in numerical integrations:

$$E\left(\frac{N(a,b)}{D(a,b)}\right)^2 = \int_0^\infty \theta_2 \frac{\partial^2 M_{a,b}(\theta_1,-\theta_2)}{\partial\theta_1^2}\Big|_{\theta_1=0} d\theta_2.$$

The above integration is computed numerically using the MATLAB functions *quadgk* and *quad2dggen*.

Using the expression for $M_{a,b}(\theta_1,\theta_2)$ from CK, we obtain the approximate variance for the full sample estimator and subsample estimators in the discrete time unit root model:

$$n^2Var(\widetilde{\beta}) = l^2Var(\widetilde{\beta}_1) = 10.1123 + O(n^{-1}).$$
$$l^2Var(\widetilde{\beta}_2) = 5.3612 + O(n^{-1}).$$

Table 2 lists the variances of all the subsample estimators for $m = 1,\ldots,12$. It can be seen that the variance of the subsample estimator decreases as j increases. The largest difference occurs between $j = 1$ and $j = 2$. If m is allowed to go to infinity, the limit distribution of the jackknife converges to that of the LS, as pointed out by CK.

To calculate the covariances, we note that:

$$n^2Cov(\widetilde{\beta},\widetilde{\beta}_j) = E\left(\frac{\int_0^1 WdW}{\int_0^1 W^2}\frac{\int_{(j-1)/m}^{j/m} WdW}{\int_{(j-1)/m}^{j/m} W^2}\right) - m\mu\mu_j + O(n^{-1}), \quad 1 \leqslant j \leqslant m.$$

$$n^2Cov(\widetilde{\beta}_i,\widetilde{\beta}_j) = E\left(\frac{\int_{(i-1)/m}^{i/m} WdW}{\int_{(i-1)/m}^{i/m} W^2}\frac{\int_{(j-1)/m}^{j/m} WdW}{\int_{(j-1)/m}^{j/m} W^2}\right) - m^2\mu_i\mu_j + O(n^{-1}), \quad 1 \leqslant i < j \leqslant m.$$

Hence, we need to compute the covariance between the limit distribution of the full sample estimator and that of any subsample estimator, and the covariance between any two subsample limit distributions. The following lemma and proposition obtain the expression for the MGF of the covariances.

Lemma 2.1. *Let $M_{a,b,c,d}(\theta_1,\theta_2,\varphi_1,\varphi_2)$ denote the MGF of $N(a,b)$, $N(c,d)$, $D(a,b)$ and $D(c,d)$ with $(0 \leqslant a < b \leqslant 1)$ and $(0 \leqslant c < d \leqslant 1)$. Then the expectation of $\frac{N(a,b)}{D(a,b)}\frac{N(c,d)}{D(c,d)}$ is given by:*

$$E\left(\frac{N(a,b)}{D(a,b)}\frac{N(c,d)}{D(c,d)}\right) = \int_0^\infty\int_0^\infty \frac{\partial^2 M_{a,b,c,d}(\theta_1,-\theta_2,\varphi_1,-\varphi_2)}{\partial\theta_1\partial\varphi_1}\Big|_{\theta_1=0,\varphi_1=0} d\theta_2 d\varphi_2. \tag{11}$$

The following proposition obtains the expression for the MGF of $N(a,b)$, $N(c,d)$, $D(a,b)$ and $D(c,d)$, and the covariances.

Proposition 2.1. *The MGF $M_{0,a,b,1}(\theta_1,\theta_2,\varphi_1,\varphi_2)$ is given by*

$$\begin{aligned} M_{0,a,b,1}(\theta_1,\theta_2,\varphi_1,\varphi_2) &= \exp\left[\frac{a\lambda-\theta_1-s\varphi_1}{2}\right][1-(2p+\eta-\lambda)\varpi^2]^{-1/2} \\ &\times\left[\cosh(e\lambda)-\frac{\theta_1}{\lambda}\sinh(e\lambda)\right]^{-1/2}\left[\cosh(s\eta)-\frac{(\theta_1-\lambda)\kappa_b+\varphi_1+\lambda}{\eta}\sinh(s\eta)\right]^{-1/2}, \end{aligned} \tag{12}$$

with $e = 1-b$, $s = b-a$, $\xi = \lambda = \sqrt{-2\theta_2}$, $\eta = \sqrt{-2\theta_2-2\varphi_2}$, $\varpi_b^2 = \frac{\exp(2\lambda e)-1}{2\lambda}$, $\kappa_b = \left[1-(\theta_1-\lambda)\varpi_b^2\right]^{-1}\exp(2\lambda e)$, $\kappa_a = \frac{\exp(2\eta s)}{1-[\varphi_1+(\theta_1-\lambda)\kappa_b+(\lambda-\eta)]\varpi_a^2}$, $p = \frac{[\varphi_1+(\theta_1-\lambda)\kappa_b+(\lambda-\eta)]\kappa_a-\varphi_1}{2}$, $\varpi_a^2 = \frac{\exp(2\eta s)-1}{2\eta}$, and $\varpi^2 = \frac{\exp(2a\lambda)-1}{2\lambda}$.

The MGF $M_{a,b,c,d}(\theta_1,\theta_2,\varphi_1,\varphi_2)$ is given by

$$\begin{aligned} M_{a,b,c,d}(\theta_1,\theta_2,\varphi_1,\varphi_2) &= \exp\left(\frac{-e\varphi_1-s\theta_1}{2}\right)\{1-a[(2p+\theta_1-\eta)\kappa_a-\theta_1+\eta]\}^{-1/2} \\ &\times[1-(c-b)(\varphi_1-\lambda)(\kappa_c-1)]^{-1/2}\left[\cosh(e\lambda)-\frac{\varphi_1}{\lambda}\sinh(e\lambda)\right]^{-1/2}\left[\cosh(s\eta)-\frac{2p+\theta_1}{\eta}\sinh(s\eta)\right]^{-1/2}, \end{aligned}$$

Table 3
Approximate variances and covariances for the full sample and subsamples when $m = 2, 3$.

Subsamples	$m=2$			$m=3$			
Covariance	$n\widetilde{\beta}$	$l\widetilde{\beta}_1$	$l\widetilde{\beta}_2$	$n\widetilde{\beta}$	$l\widetilde{\beta}_1$	$l\widetilde{\beta}_2$	$l\widetilde{\beta}_3$
$n\widetilde{\beta}$	10.1123	5.0188	5.7932	10.1123	3.3443	4.0769	4.3796
$l\widetilde{\beta}_1$	5.0188	10.1123	1.1053	3.3443	10.1123	1.1053	0.4287
$l\widetilde{\beta}_2$	5.7932	1.1053	5.3612	4.0769	1.1053	5.3612	0.8978
$l\widetilde{\beta}_3$				4.3796	0.4287	0.8978	4.2839

with $e = d - c$, $s = b - a$, $\varpi_c^2 = \frac{\exp(2e\lambda)-1}{2\lambda}$, $\varpi_a^2 = \frac{\exp(2\eta s)-1}{2\eta}$, $\kappa_a = \left[1-(2p+\theta_1-\eta)\varpi_a^2\right]^{-1}\exp(2\eta s)$, $\kappa_c = \left[1-(\varphi_1-\lambda)\varpi_c^2\right]^{-1}\exp(2e\lambda)$ *and* $p = \frac{(\varphi_1-\lambda)(\kappa_c-1)}{2[1-(c-b)(\varphi_1-\lambda)(\kappa_c-1)]}$.

Chen and Yu (2013, page 22–23) gives the expression of the second derivative of the MGF $M_{0,a,b,1}(\theta_1, \theta_2, \varphi_1, \varphi_2)$ and $M_{a,b,c,d}(\theta_1, \theta_2, \varphi_1, \varphi_2)$, which is used to compute the numerical value of covariance. When $m = 2$, we have following approximate covariances between the full sample estimator and the two subsample estimators:

$$n^2 Cov(\widetilde{\beta}, \widetilde{\beta}_1) = 10.0376 + O(n^{-1}); \tag{13}$$

$$n^2 Cov(\widetilde{\beta}, \widetilde{\beta}_2) = 11.5863 + O(n^{-1}); \tag{14}$$

$$n^2 Cov(\widetilde{\beta}_1, \widetilde{\beta}_2) = 4.4212 + O(n^{-1}). \tag{15}$$

There are several interesting findings here. First, the covariances between the full sample estimator and the second subsample estimator are similar to, but slightly larger than that between the full sample estimator and the first subsample estimator, although the variance of the second subsample estimator is smaller. This is because the correlation between the full sample estimator and the second subsample estimator is larger due to the increased order of magnitude of the initial condition. Second, these two covariances are much larger than the covariance between the two subsample estimators. This is not surprising as the data used in the two subsamples estimators do not overlap.

Table 3 summarizes the approximate values of the variances and covariances when $m = 2, 3$. Given the values of variances and covariances, we further compute the optimal jackknife estimator when $m = 2$:

$$\widetilde{\beta}_{JK}^{CY} = 2.8390\widetilde{\beta} - (0.6771\widetilde{\beta}_1 + 1.1619\widetilde{\beta}_2). \tag{16}$$

and the optimal jackknife estimator when $m = 3$:

$$\widetilde{\beta}_{JK}^{CY} = 2.0260\widetilde{\beta} - (0.2087\widetilde{\beta}_1 + 0.3376\widetilde{\beta}_2 + 0.4797\widetilde{\beta}_3). \tag{17}$$

Clearly the weights assigned to the subsample estimates are not the same in our method.

It can be easily shown that the results derived for the discrete time unit root model can be adapted to the following continuous time unit root model:

$$dy_t = -\kappa y dt + \sigma dW_t, \quad \text{with } \kappa = 0. \tag{18}$$

The bias in $\widetilde{\kappa}$ is substantial as shown in Yu (2012). Although it was recently shown in Bao et al. (2013) that the exact moment of LS estimator of κ does not exist, the moment we try to approximate can be thus understood as the pseudo moment. Chen and Yu (2013) derives the optimal jackknife weights for the LS estimator of κ using long span asymptotics for $\widetilde{\kappa}$.

3. Simulation studies

We design three experiments to evaluate the performance of alternative jackknife methods under different sample sizes with the number of replications being 10,000. It is reasonable to consider small sample sizes since we focus on the finite sample property. As a benchmark, we always report the original full sample estimator which is the LS estimator.

In Table 4, we simulate data from Model (1) with a Gaussian error ε_t. In the first experiment, we set $m = 2$ which is chosen by CK to minimize the bias. We compare the CK jackknife method based on weights from Table 1, the CY jackknife method based on (16) where the weights are derived from the approximate variances and covariances, the CY jackknife method where the weights are calculated from the exact variances and covariances obtained from the finite sample distributions. We measure the efficiency gain of the proposed method over the CK method based on the ratio of variances as well as the ratio of RMSEs. Since the weights are obtained based on the variances and the covariances of the limit distributions but not on the finite sample distributions, it is useful to examine the approximation error from using the limit distributions. Although it is difficult to obtain the analytical expressions for the variances and the covariances of the finite sample distribution, they can be computed using simulated data in a Monte Carlo study, provided the number of replications is large enough. The bias, variance, RMSE, and relative efficiency for the estimates of β are reported in the second block of Table 4. The bias is very similar for the CK method and the two CY methods in all cases. The variance of the two CY methods is significantly smaller than that of CK in all cases with the reduction in variance being about 10%. Consequently, the RMSE is smaller for the CY

Table 4
Finite sample performance of alternative jackknife estimators for discrete time unit root models.

n	Statistics	Bias-minimizing values of m						RMSE-minimizing values of m			
		m	LS	CK	CY	RE	Exact CY	m	CK	Exact CY	RE
24	2	Bias	−0.068	−0.012	−0.013		−0.014	4	−0.025	−0.027	
		100 ∗ Var	1.538	3.635	3.186	0.876	3.159		2.230	1.978	0.887
		10 ∗ RMSE	1.415	1.910	1.790	0.937	1.783		1.513	1.431	0.946
48	2	Bias	−0.036	−0.003	−0.004		−0.004	6	−0.009	−0.009	
		100 ∗ Var	0.405	0.903	0.791	0.876	0.787		0.449	0.400	0.891
		10 ∗ RMSE	0.729	0.951	0.890	0.936	0.888		0.676	0.639	0.946
96	2	Bias	−0.018	−0.001	−0.001		−0.001	8	−0.003	−0.003	
		100 ∗ Var	0.106	0.235	0.208	0.888	0.208		0.104	0.092	0.880
		10 ∗ RMSE	0.374	0.485	0.457	0.942	0.456		0.324	0.304	0.939
192	2	Bias	−0.009	0.000	0.000		0.000	8	−0.001	−0.001	
		100 ∗ Var	0.028	0.061	0.054	0.887	0.054		0.027	0.024	0.882
		10 ∗ RMSE	0.191	0.247	0.233	0.942	0.233		0.165	0.155	0.939

methods. Although the exact CY method provides a smaller variance, the difference between the two CY methods is very small, suggesting that the CY based on the limit distributions works well. Table 4 of Chen and Yu (2013) gives the simulation results when $m = 3$.

In the second experiment, the same model is estimated but we follow CK by choosing m to minimize the RMSE. We then compare the CK jackknife and the CY jackknife based on the same m. The CY jackknife is computed using the exact variances and covariances for the purpose of easy implementation. The bias, variance, RMSE, and relative efficiency for the estimates of β are reported in the third block of Table 4. It can be seen that the optimal value for m changes with the sample size. With the RMSE-minimizing m being used, however, the variance of the new jackknife continues to be smaller than that of CK with the reduction in variance being about 10%.

The simulation results obtained above are based on the assumption that the true model has a unit root. In practice, however, the persistence parameter is often unknown and has to be estimated although the unit root model is very popular empirically. Based on the estimator of slope coefficient, it is not easy to tell whether the true value is one or not. To check the robustness of our results in the persistent case, we now compare the finite sample performance of three jackknife estimators in the context of an AR model with a root local to unity. Following Phillips (1987) and Chan and Wei (1988), the data generating process considered is:

$$y_t = \beta y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid\, N(0, 1),\ t = 1, 2, \ldots, n,$$

with $\beta = 1 + c/n$, $(-\infty < c < \infty)$. We investigate the finite sample performance of alternative estimates of β using a sample size ranging from 12 to 108. The local to unity parameter c is set at 0.1, 0.5 and 1 for the local to unity from the explosive side, and −0.1, −0.5 and −1 for the local to unity from the stationary side. Table 5 reports the bias, variance and RMSE of the CK method and the two CY methods. In all cases, the bias is always close to zero, suggesting the jackknife methods continue to work well for reducing the bias in the local to unit root case. The variance for the two CY methods continues to be smaller than that for the CK method in all cases. Moreover, although the exact CY method provides a smaller variance, the difference between the two CY methods is quite small, suggesting that the weights obtained from the unit root model approximate well the exact weights in the local to unit root models.[1]

4. Conclusion

This paper has introduced a new jackknife procedure for unit root models that offers an improvement over the jackknife methodology of CK (2013). The proposed estimator is optimal in the sense that it minimizes the variance of the jackknife estimator while removing the first order bias. The new method works well both in the unit root model and in the local to unit root model. Simulations have shown that the new method reduces the variance by about 10% relative to the estimator of CK without compromising the bias. The results hold true when an optimal number of subsamples is used in both jackknife methods. There exist some other models for which the asymptotic theory depends on the initial condition, i.e. (Phillips and Magdalinos, 2009). Examples include explosive processes. It may be interesting to extend the results in the present paper to cover these models, although it is not pursued in the present paper. It is useful to point out that for a unit root model with an unknown intercept case, although fitting an intercept increases the bias of LS estimator, the asymptotic theory does not depend on the initial value.

[1] However, as pointed out by a referee, the limit distribution depends on the localization parameter. Hence, it is expected that the jackknife weight should also depend on this parameter.

Table 5
Finite sample performance of alternative jackknife methods for the discrete time local to unit root model for $m = 2$, where RE means the efficiency of CK relative to CY and the Exact CY (ExCY).

n	c	−0.1					−0.5					−1				
	Statistics	LS	CK	CY	RE	ExCY	LS	CK	CY	RE	ExCY	LS	CK	CY	RE	ExCY
12	Bias	−0.128	−0.057	−0.060		−0.061	−0.128	−0.055	−0.056		−0.056	−0.125	−0.050	−0.048		−0.048
	100 ∗ Var	5.616	17.493	15.539	0.888	15.425	5.738	17.354	15.571	0.897	15.521	5.903	17.413	16.093	0.924	16.092
	10 ∗ RMSE	2.695	4.221	3.988	0.945	3.975	2.714	4.201	3.985	0.949	3.979	2.734	4.202	4.040	0.961	4.040
36	Bias	−0.047	−0.005	−0.006		−0.006	−0.048	−0.005	−0.004		−0.004	−0.048	−0.003	−0.001		−0.001
	100 ∗ Var	0.723	1.611	1.401	0.870	1.390	0.758	1.629	1.454	0.892	1.452	0.804	1.677	1.572	0.938	1.566
	10 ∗ RMSE	0.971	1.270	1.185	0.933	1.180	0.993	1.277	1.206	0.945	1.206	1.018	1.295	1.254	0.968	1.251
48	Bias	−0.036	−0.003	−0.003		−0.003	−0.036	−0.002	−0.001		−0.001	−0.037	−0.001	0.001		0.001
	100 ∗ Var	0.410	0.908	0.803	0.884	0.801	0.430	0.925	0.844	0.913	0.844	0.456	0.950	0.903	0.951	0.895
	10 ∗ RMSE	0.734	0.953	0.896	0.940	0.895	0.750	0.962	0.919	0.955	0.919	0.770	0.974	0.950	0.975	0.946
96	Bias	−0.019	−0.001	−0.001		−0.001	−0.019	−0.001	0.000		0.000	−0.019	0.000	0.001		0.000
	100 ∗ Var	0.107	0.236	0.211	0.894	0.211	0.114	0.243	0.224	0.923	0.224	0.122	0.252	0.241	0.956	0.238
	10 ∗ RMSE	0.377	0.486	0.459	0.945	0.459	0.387	0.493	0.473	0.960	0.473	0.399	0.502	0.491	0.978	0.488
108	Bias	−0.017	−0.001	−0.001		−0.001	−0.017	−0.001	−0.001		−0.001	−0.017	0.000	0.000		0.000
	100 ∗ Var	0.087	0.182	0.163	0.898	0.163	0.092	0.188	0.175	0.932	0.175	0.099	0.195	0.189	0.970	0.187
	10 ∗ RMSE	0.338	0.427	0.404	0.948	0.404	0.348	0.434	0.418	0.965	0.418	0.359	0.442	0.435	0.985	0.432
n	c	0.1					0.5					1				
	Statistics	LS	CK	CY	RE	ExCY	LS	CK	CY	RE	ExCY	LS	CK	CY	RE	ExCY
12	Bias	−0.128	−0.058	−0.062		−0.063	−0.127	−0.058	−0.065		−0.067	−0.124	−0.056	−0.065		−0.072
	100 ∗ Var	5.558	17.636	15.694	0.890	15.577	5.446	17.888	15.947	0.891	15.814	5.306	17.729	15.231	0.859	14.648
	10 ∗ RMSE	2.684	4.239	4.010	0.946	3.997	2.659	4.269	4.045	0.948	4.032	2.617	4.247	3.957	0.932	3.895
36	Bias	−0.046	−0.006	−0.007		−0.007	−0.045	−0.006	−0.008		−0.009	−0.042	−0.005	−0.008		−0.012
	100 ∗ Var	0.706	1.604	1.380	0.861	1.361	0.672	1.595	1.353	0.848	1.315	0.632	1.570	1.293	0.824	1.203
	10 ∗ RMSE	0.959	1.268	1.177	0.928	1.169	0.934	1.264	1.166	0.922	1.150	0.900	1.254	1.140	0.909	1.103
48	Bias	−0.035	−0.004	−0.004		−0.004	−0.034	−0.004	−0.005		−0.006	−0.032	−0.003	−0.006		−0.008
	100 ∗ Var	0.400	0.898	0.779	0.867	0.772	0.382	0.884	0.746	0.844	0.726	0.359	0.868	0.718	0.827	0.677
	10 ∗ RMSE	0.725	0.948	0.883	0.932	0.880	0.707	0.941	0.866	0.920	0.854	0.681	0.932	0.849	0.911	0.826
96	Bias	−0.018	−0.001	−0.001		−0.001	−0.018	−0.001	−0.002		−0.002	−0.016	−0.001	−0.002		−0.003
	100 ∗ Var	0.104	0.234	0.206	0.883	0.206	0.098	0.229	0.197	0.862	0.195	0.090	0.222	0.184	0.830	0.174
	10 ∗ RMSE	0.371	0.484	0.454	0.940	0.454	0.359	0.479	0.445	0.929	0.442	0.342	0.471	0.429	0.912	0.419
108	Bias	−0.016	−0.001	−0.001		−0.001	−0.016	−0.001	−0.002		−0.002	−0.015	−0.001	−0.002		−0.003
	100 ∗ Var	0.085	0.180	0.160	0.889	0.160	0.080	0.178	0.156	0.880	0.155	0.073	0.172	0.146	0.848	0.142
	10 ∗ RMSE	0.334	0.425	0.400	0.943	0.400	0.323	0.422	0.396	0.939	0.395	0.308	0.415	0.382	0.922	0.377

Acknowledgment

Jun Yu would like to acknowledge the financial support from Singapore Ministry of Education Academic Research Fund Tier 2 under the grant number MOE2011-T2-2-096.

Appendix

Proof of Lemma 2.1. Taking the derivative of MGF with respect to θ_1, we get:

$$\frac{\partial M_{a,b,c,d}(\theta_1, -\theta_2, \varphi_1, -\varphi_2)}{\partial \theta_1} = E\left[N(a,b)\exp(\theta_1 N(a,b) - \theta_2 D(a,b) + \varphi_1 N(c,d) - \varphi_2 D(c,d))\right].$$

Setting $\theta_1 = 0$, taking the derivative with respect to φ_1, and then evaluating it at $\varphi_1 = 0$, we have,

$$\left\{\partial\left[\left.\frac{\partial M_{a,b,c,d}(\theta_1, -\theta_2, \varphi_1, -\varphi_2)}{\partial \theta_1}\right|_{\theta_1=0}\right] \Big/ \partial\varphi_1\right\}\Big|_{\varphi_1=0} = E\left\{N(a,b)N(c,d)\exp\left[-\theta_2 D(a,b) - \varphi_2 D(c,d)\right]\right\}.$$

Consequently,

$$\int_0^\infty \int_0^\infty \left\{\partial\left[\left.\frac{\partial M_{a,b,c,d}(\theta_1, -\theta_2, \varphi_1, -\varphi_2)}{\partial \theta_1}\right|_{\theta_1=0}\right] \Big/ \partial\varphi_1\Big|_{\varphi_1=0}\right\} d\theta_2 d\varphi_2 = E\left[\frac{N(a,b)}{D(a,b)}\frac{N(c,d)}{D(c,d)}\right].$$

Proof of Proposition 2.1. It can be found in Chen and Yu (2013, page 21–28).

References

Bao, Y., Ullah, A., Zinde-Walsh, V., 2013. On existence of moment of mean reversion estimator in linear diffusion models. Econom. Lett. 120, 146–148.
Chambers, M.J., 2013. Jackknife estimation and inference in stationary autoregressive models. J. Econometrics 172, 142–157.

Chambers, M.J., Kyriacou, M., 2013. Jackknife estimation with a unit root. Statist. Probab. Lett. 83, 1677–1682.
Chan, N.H., Wei, C.Z., 1988. Limiting distributions of least squares estimates of unstable autoregressive processes. Ann. Statist. 16, 367–401.
Chen, Y., Yu, J., Optimal jackknife for discrete time and continuous time unit root models, School of Economics, Singapore Management University, Working Paper, 2013.
Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Statist. 7, 1–26.
Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer-Verlag, New York.
Magnus, J.R., 1986. The exact moments of a ratio of quadratic forms in Normal variables. Ann. d'Économ. Stat. 4, 95–109.
Phillips, P.C.B., 1987. Toward a unified asymptotic theory for autoregression. Biometrika 74, 533–547.
Phillips, P.C., Magdalinos, T., 2009. Unit root and cointegrating limit theory when initialization is in the infinite past. Econometric Theory 25 (6), 1682–1715.
Phillips, P.C., Yu, J., 2005. Jackknifing bond option prices. Rev. Financ. Stud. 18 (2), 707–742.
Quenouille, M.H., 1949. Approximate tests of correlation in time-series. J. R. Stat. Soc. Ser. B 11, 68–84.
Sargan, J.D., 1976. Econometric estimators and the edgeworth approximation. Econometrica 44, 421–448.
Shao, J., Tu, D., 1995. The Jackknife and Bootstrap. Springer, New York.
White, J.S., 1961. Asymptotic expansions for the mean and variance of the serial correlation coefficient. Biometrika 48, 85–94.
Yu, J., 2012. Bias in the estimation of the mean reversion parameter in continuous time models. J. Econometrics 169, 114–122.