SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

Optimal Jackknife for Discrete Time and Continuous Time Unit Root Models

Ye Chen & Jun Yu

October 2011

Paper No. 12-2011

Optimal Jackknife for Discrete Time and Continuous Time Unit Root Models*

Ye Chen and Jun Yu
Singapore Management University

September 27, 2011

Abstract

Maximum likelihood estimation of the persistence parameter in the discrete time unit root model is known for suffering from a downward bias. The bias is more pronounced in the continuous time unit root model. Recently Chambers and Kyriacou (2010) introduced a new jackknife method to remove the first order bias in the estimator of the persistence parameter in a discrete time unit root model. This paper proposes an improved jackknife estimator of the persistence parameter that works for both the discrete time unit root model and the continuous time unit root model. The proposed jackknife estimator is optimal in the sense that it minimizes the variance. Simulations highlight the performance of the proposed method in both contexts. They show that our optimal jackknife reduces the variance of the jackknife method of Chambers and Kyriacou by at least 10% in both cases.

Keywords: Bias reduction, Variance reduction, Vasicek model, Long-span Asymptotics, Autoregression

JEL classification: C11, C15

1 Introduction

As a subsampling method, the jackknife estimator was first proposed by Quenouille (1949) for reducing the bias in the estimation of the serial correlation coefficient. Tukey (1958) used the jackknife method to estimate the variance. Miller (1974) reviewed the literature of the jackknife, emphasizing its properties in bias reduction and interval estimation. More recently, the jackknife method has been found useful

*Ye Chen, School of Economics and Sim Kee Boon Institute for Financial Economics, Singapore Management University, 90 Stamford Road, Singapore, 178903; Email: chenye.2009@phdecons.smu.edu.sg. Jun Yu, Sim Kee Boon Institute for Financial Economics, School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Road, Singapore 178903. Email: yujun@smu.edu.sg. We acknowledge comments from Peter Phillips and Denis Leung.

in econometrics. Hahn and Newey (2003) demonstrated the use of the jackknife to reduce the bias arising from the incidental parameters problem in the dynamic panel model with fixed effects. Phillips and Yu (2005, PY hereafter) introduced a jackknife method based on non-overlapping subsamples, and showed that it can be applied to the coefficients in continuous time models as well as the asset prices directly. Chiquoine and Hjalmarsson (2009) provided the evidence that the jackknife can be successfully applied to stock return predictability regressions.

In the time series context, the jackknife method proposed by PY is very easy to implement and also quite effective in reducing the bias. In effect, this jackknife estimator removes the first order bias in the original estimator by taking a linear combination of the full sample estimator and a set of subsample estimators. The validity for removing the first order bias can be explained by the Nagar approximation of the moments of the original estimator, in terms of the moments of the estimator's Taylor expansion as a polynomial of the sample moments of the data. In particular, under regular conditions, the first order bias is a reciprocal function of the sample size with a common coefficient. Consequently, the weights used for the subsamples are the same and are proportional to the weight used for the full sample. For stationary autoregressive models, Chambers (2010) obtained the limit distribution of the jackknife estimator of PY and examined its performance relative to alternative jackknifing procedures.

However, in the context of a discrete time unit root model, Chambers and Kyriacou (2010, CK hereafter) pointed out that the jackknife estimator of PY cannot completely remove the first order bias. This is because the first order bias of the least squares (LS) or maximum likelihood (ML) estimator of the autoregressive coefficient has distinctive functional forms in different subsamples. This distinctiveness is manifest in the fact that the limit distribution of the LS/ML estimator depends on the initial condition. A revised jackknife estimator with new weights was proposed in CK. Unlike the PY estimator, the weight used in CK for the full sample is no longer proportional to those for the subsamples. However, the weights used for the subsamples are required to be identical to each other in CK. CK showed that the modified jackknife estimate performs better than the PY estimator for bias reduction.

While the jackknife method of CK reduces the bias of the original estimator, it always increases the variance. This is not surprising because subsample estimators are used in the construction of the jackknife. As a result, the trade-off between the bias reduction and the increase in variance is important. Simulation results in CK suggest that their jackknife method sometimes gains enough in bias reduction without a significant increase in variance, so that there may be an overall gain in root mean square error (RMSE) for the discrete time unit root model.

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 minimizes the variance, and hence, has better finite sample properties than the CK estimator. Like the estimators of PY and CK, our optimal jackknife estimator is a linear combination of the full sample estimator and the subsample estimators. However, we do not require the weight for the full sample to be proportional to those for the subsample, nor do we require the weights for the subsample be the same. Indeed, all the weights are obtained by minimizing the variance under the condition that the first order bias is removed.

The calculation of the weights involves obtaining the covariances between the full sample estimator and the subsample estimators, and the covariances between the subsample estimators. To do so, we employ the joint moment generating function (MGF) of the limit distributions of these estimators. The idea is applied to both the discrete time and the continuous time unit root models. Optimal weights are derived and the finite sample performance of the new estimator is examined for both models. It is found that the optimal jackknife estimator offers over 10% reduction in variance over the CK estimator without compromising bias reduction in both cases.

The paper is organized as follows. Section 2 derives the general form of the optimal jackknife estimator in the discrete time unit root model. Section 3 extends the results to the continuous time unit root model where the bias of the persistent parameter is known to be more serious. Section 4 presents the Monte Carlo simulation results. Section 5 concludes. Appendix collects all the proofs.

Throughout the paper, we adopt the following notations. The full sample size is denoted by n, the time span of data by T, the parameter of interest by θ (or β or κ), and the true value of it by θ_0 (or β_0 or κ_0). The signal "$\Rightarrow$" is used to indicate the convergence of the associated probability measures as n or T goes to infinity while "$\equiv$" denotes the equality in distribution, and m the number of subsamples. $\widetilde{\theta}_j$ means the LS/ML estimator of θ from the jth subsample of sample length l (i.e., $m \times l = n$), $\widetilde{\theta}^{PY}$ is the jackknife estimator of θ proposed by PY, $\widetilde{\theta}^{CK}$ is the jackknife estimator proposed by CK, and $\widetilde{\theta}^{CY}$ is the jackknife estimator proposed in the present paper. Following CK, we define $Z = \int_0^1 WdW / \int_0^1 W^2$, $Z_j = \int_{(j-1)/m}^{j/m} WdW / \int_{(j-1)/m}^{j/m} W^2$, $\mu = E(Z)$ and $\mu_j = E(Z_j)$, where W is a standard Brownian motion.

2 Optimal Jackknife for the Discrete Time Unit Root Model

In this section, we first briefly review the literature, focusing especially on the approach of CK (2010). We then introduce our optimal jackknife estimator in the context of the discrete time unit root model.

2.1 A Literature Review

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

$$y_t = y_{t-1} + \varepsilon_t, \ \varepsilon_t \sim iid(0, \sigma_\varepsilon^2), \ t = 1, 2, \ldots, n. \tag{2.1}$$

The LS estimator of the autoregressive (AR) parameter of, say, β, is $\widetilde{\beta} = \sum_{t=1}^{n} y_{t-1}y_t / \sum_{t=1}^{n} y_{t-1}^2$. When ε_t is normally distributed, $\widetilde{\beta}$ is also the ML estimator of β, conditional on y_0. The limit distribution of $\widetilde{\beta}$ was obtained by Phillips (1987a) using the functional central limit theory and the continuous mapping theorem:

$$n\left(\widetilde{\beta}-1\right) \Rightarrow \frac{\int_0^1 W dW}{\int_0^1 W^2}. \tag{2.2}$$

The limit distribution is skewed for two reasons. First, $\int_0^1 W dW$ is not centered around zero. Second, $\int_0^1 W dW$ and $\int_0^1 W^2$ are dependent. This skewness gives rise to the first order bias in the LS/ML estimator, even asymptotically.

Under the assumption that the initial value y_0 is 0, Phillips (1987a) obtained the asymptotic expansion of the limit distribution of $n\left(\widetilde{\beta}-1\right)$:

$$n\left(\widetilde{\beta}-1\right) = \frac{\int_0^1 W dW}{\int_0^1 W^2} - \frac{1/\sqrt{2n}\xi}{\int_0^1 W^2} + O_p(n^{-1}), \tag{2.3}$$

where $\xi \equiv N(0,1)$, and is independent of W. Taking the expectation of (2.3) and noting that $\mu = E(Z) = -1.781$, we have

$$E\left(\widetilde{\beta}\right) - 1 = -\frac{1.781}{n} + O(n^{-2}). \tag{2.4}$$

This result is in contrast to the Kendall bias formula for the stationary AR(1) model (i.e. when $\beta_0 < 1$):

$$E\left(\widetilde{\beta}\right) - \beta_0 = -\frac{2\beta_0}{n} + O(n^{-1}). \tag{2.5}$$

Following the original work of Quenouille, PY (2005) utilized the subsample estimators of β to achieve bias reduction with the following linear combination:

$$\widetilde{\beta}_m^{PY} = \frac{m}{m-1}\widetilde{\beta} - \left(\frac{1}{m-1}\right)\left(\frac{1}{m}\sum_{j=1}^{m}\widetilde{\beta}_j\right), \tag{2.6}$$

where $\widetilde{\beta}$ is the LS/ML estimator of β based on the full sample, i.e., $y_1, \ldots, y_n$; $\widetilde{\beta}_j$ is the LS/ML estimator of β based on the jth subsample, i.e., $y_{(j-1)l+1}, \ldots, y_{jl}$. There are two important features in this estimator. First, the weight given to the full sample estimator is proportional to (i.e. m times as large as) that assigned to the average of the subsample estimators. Second, the weights are the same for all the subsample estimators. The validity of this jackknife method can be explained by a general result based on the Nagar expansion:

$$E\left(\widetilde{\beta}\right) = \beta_0 + \frac{b_1}{n} + O\left(n^{-2}\right), \tag{2.7}$$

which is assumed to hold for some constant b_1. Two observations can be made from this general result. First, to the first order, the bias is a reciprocal function of the sample size. This explains why in the PY method, the weight assigned to the full sample estimator is m times as big as that assigned to the subsample estimators. Second, to the first order, the bias is the same across different subsamples. This explains why in the PY method, the weights are the same for all the subsample estimators. It is easy to show that the jackknife estimator removes the first order bias, i.e., $E\left(\widetilde{\beta}_m^{PY}\right) = \beta_0 + O\left(n^{-2}\right)$, when (2.7) holds true.

Particularly effective bias reduction can be achieved by choosing $m = 2$. In this case, this estimator becomes:

$$\widetilde{\beta}^{PY} = 2\widetilde{\beta} - \frac{1}{2}\left(\widetilde{\beta}_1 + \widetilde{\beta}_2\right). \tag{2.8}$$

Both PY and Chambers (2010) have reported evidence to support this method for the purpose of bias reduction in different contexts. It should be pointed out that many jackknife procedures have been proposed, including the delete-1 jackknife of Tukey (1958), the delete-d jackknife of Wu (1986), and the moving-block jackknife of Chambers (2010).

The Nagar approximation is a general result and may be verified by Sargan's (1976) theorem. Given the mild conditions under which Sargan's (1976) theorem hold, it is perhaps rather surprising that the jackknife method fails to remove the first order bias in a unit root model. This failure was first documented in CK (2010). The basic argument of CK is that in (2.7), b_1 is not a constant any more in the unit root model. Instead, it depends on the initial condition. As the initial condition varies across different subsamples, the jackknife cannot eliminate the first order bias term. To be more specific, the limit distribution of the normalized subsamples estimator, $l(\widetilde{\beta}_j - 1)$, is $\int_{(j-1)/m}^{j/m} W dW / \left(m \int_{(j-1)/m}^{j/m} W^2\right)$. CK showed that $\mu_j = E\left(\int_{(j-1)/m}^{j/m} W dW / \int_{(j-1)/m}^{j/m} W^2\right)$ depends on j. To eliminate the first order asymptotic bias, CK proposed the following modified jackknife estimator:

$$\widetilde{\beta}_m^{CK} = b_m^{CK}\widetilde{\beta} - \sum_{j=1}^{m} \delta_m^{CK}\widetilde{\beta}_j, \tag{2.9}$$

where

$$b_m^{CK} = \frac{\sum_{j=1}^m \mu_j}{\sum_{j=1}^m \mu_j - \mu}, \delta_m^{CK} = \frac{\mu}{m\left(\sum_{j=1}^m \mu_j - \mu\right)}. \tag{2.10}$$

When $\mu_1 = \cdots = \mu_m = \mu$, $b_m^{CK} = m/(m-1) = b_m^{PY}$, and $\delta_m^{CK} = 1/\left(m^2 - m\right)$, the CK estimator is the same as the PY estimator. Under model (2.1), CK showed that $\mu = \mu_1 = -1.781430$, $\mu_2 = -1.138209$, $\mu_3 = -0.931929$, $\mu_4 = -0.814330$, etc. That is, the bias becomes smaller and smaller as we go deeper and deeper into subsampling. Substituting these expected values into the formulae (2.10), we can calculate the weights. Table 1 reports the weights when $m = 2$. We also report the weights of PY for comparison. As μ_2 is closer to zero than μ_1 and μ, a larger weight is assigned to the full sample

Table 1: Weights assigned to the full sample and subsample estimators for alternative jackknife methods

m	b_m^{PY}	δ_m^{PY}	b_m^{CK}	δ_m^{CK}	b_m^{CY}	$a_{1,m}^{CY}$	$a_{2,m}^{CY}$
2	2	0.5	2.5651	0.7825	2.8390	0.6771	1.1619

estimator, compared to the PY estimator.

The dependence of the bias on the initial condition is a very interesting finding which is explained by a recent result obtained in Phillips and Magdalinos (2009, PM hereafter). In PM, the initial condition, y_0, is assumed to be:

$$y_0(n) = \sum_{j=0}^{k_n} \varepsilon_{-j}, \tag{2.11}$$

with

$$k_n/n \to \tau \in [0, +\infty]. \tag{2.12}$$

When $\tau = 0$, $y_0(n)$ is said to be a recent past initialization, which obviously includes $y_0 = 0$ as a special case. When $\tau = (0, +\infty)$, $y_0(n)$ is said to be a distant past initialization. When $\tau = \infty$, $y_0(n)$ is said to be an infinite past initialization. PM showed that when $\tau \in [0, +\infty)$, the limit distribution of the LS/ML estimator of β under the unit root model is given by:

$$n\left(\widetilde{\beta} - 1\right) \Rightarrow \frac{\int_0^1 W_\tau^+ dW}{\int_0^1 \left(W_\tau^+\right)^2},$$

where $W_\tau^+(t) = W(t) + \sqrt{\tau} W_0(t)$, and W_0 is another standard Brownian motion but independent of W. If $\tau = 0$, the limit distribution is the same as in (2.2). When $\tau = +\infty$, the limit distribution of the LS/ML estimator of β under the unit root model is given by:

$$\sqrt{k_n n}\left(\widetilde{\beta} - 1\right) \Rightarrow C,$$

where C stands for a Cauchy variate. The dependence of the limit distribution and hence its first moment on the initial condition are clearly seen from the above results. The larger the value of τ, the greater is the importance of the initial condition, and the bigger is the share of $\sqrt{\tau} W_0(t)$ in $W_\tau^+(t)$.

For Model (2.1), $\tau = 0$ applied for the first sample and $\tau = jl/n$, for the $(j+1)th$ subsample. That is, for $j = 1, 2, \ldots, m-1$, we have $\tau = 0, 1/m, \ldots, (m-1)/m$. As we go deeper and deeper into subsampling, the bigger and bigger is the influence of the initial condition on the limit distribution. Not surprisingly, all the moments, including the mean and the variance, are different for different subsamples.

2.2 Optimal Jackknife

It is known that the jackknife estimator of PY increases the variance, when compared to the LS/ML estimator. The same feature is seen in the estimator of CK. In this paper, we introduce a new jackknife estimator, which can remove the first order bias as well as minimize the variance for a given m. The reason why it is possible to reduce the variance of the estimator of CK is that it is not necessary to assign an identical weight to different subsamples, given that the biases of these estimators are different.

To fix the idea, the new jackknife estimator is defined as:

$$\widetilde{\beta}_m^{CY} = b_m^{CY}\widetilde{\beta} - \sum_{j=1}^{m} a_{j,m}^{CY}\widetilde{\beta}_j,$$

where b_m^{CY} and $\{a_{j,m}^{CY}\}_{j=1}^m$ are the weights assigned to the full sample estimator and the subsample estimators, respectively. Unlike the jackknife estimators of PY and CK, we do not require the weights for different subsamples to be the same. Our objective is to select the weights $b_m^{CY}, \{a_{j,m}^{CY}\}_{j=1}^m$, to minimize the variance of the new estimator, i.e.:

$$\min_{b_m^{CY},\{a_{j,m}^{CY}\}_{j=1}^m} Var\left(b_m^{CY}\widetilde{\beta} - \sum_{j=1}^{m} a_{j,m}^{CY}\widetilde{\beta}_j\right), \tag{2.13}$$

subject to two constraints:

$$b_m^{CY} = \sum_{j=1}^{m} a_{j,m}^{CY} + 1, \tag{2.14}$$

$$b_m^{CY}\mu = m\sum_{j=1}^{m} a_{j,m}^{CY}\mu_j, \tag{2.15}$$

where $\mu = \mu_1$. These two constraints are used to ensure the first order bias of the LS/ML estimator is fully removed.

From the two constraints, we can express b_m^{CY} and $a_{1,m}^{CY}$ as functions of $a_{2,m}^{CY}, \ldots, a_{m,m}^{CY}$:

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

Substituting b_m^{CY} and $a_{1,m}^{CY}$ into the objective function (2.13), and then taking the derivative with respect to $a_{j,m}^{CY}$, we get:

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

for $j = 2, \cdots, m$. These first order conditions can be used to obtain analytical expressions for $a_{2,m}^{CY}, \ldots, a_{m,m}^{CY}$ and hence b_m^{CY} and $a_{1,m}^{CY}$, all as functions of $\mu, \mu_2, \ldots, \mu_m$, the variances and the covariances of the full sample and subsample estimators.

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 also calculate the exact variances and covariances of the finite sample distributions. However, it is known in the literature that the exact moments are analytically very difficult to obtain in dynamic models. To simplify the derivations, we approximate the moments of the finite sample distributions by those of the limit distributions. We shall check the quality of these approximations in the simulation studies. While the techniques proposed in White (1961) and in CK can be combined to compute the variances, additional effort is needed to compute the covariances. A technical contribution of the present paper is to show how to compute these covariances.

We now illustrate the optimal jackknife method in a special case where $m = 2$, that is, we split the full sample into two non-overlapping subsamples. The first subsample is made with the observations from y_1 to y_l, with $l = n/2$ and the initial condition 0, and the remainder belongs to the second subsample with the initial condition y_l.

Under the two non-overlapping subsample scheme, we have:

$$
\widetilde{\beta}_m^{CY} = b^{CY}\widetilde{\beta} - (\, a_1^{CY}\widetilde{\beta}_1 \;+\; a_2^{CY}\widetilde{\beta}_2 \,).
$$

The objective function and the constraints are:

$$
\begin{aligned}
\min_{b^{CY}, a_1^{CY}, a_2^{CY}} \quad & \left(b^{CY}\right)^2 Var(\widetilde{\beta}) + \sum_{j=1}^{2}\left(a_j^{CY}\right)^2 Var(\widetilde{\beta}_j) - 2b^{CY}\sum_{j=1}^{2} a_j^{CY} Cov(\widetilde{\beta},\widetilde{\beta}_j) + 2a_1^{CY}a_2^{CY}Cov(\widetilde{\beta}_1,\widetilde{\beta}_2), \\
s.t. \quad & b^{CY} = a_1^{CY} + a_2^{CY} + 1, \\
& b^{CY}\mu = 2a_1^{CY}\mu_1 + 2a_2^{CY}\mu_2.
\end{aligned}
$$

From the two constraints, we express b^{CY} and a_1^{CY} in terms of a_2^{CY}, that is:

$$b^{CY} = \frac{a_2^{CY}(2\mu - 2\mu_2)}{\mu} + 2, \tag{2.16}$$

$$a_1^{CY} = \frac{a_2^{CY}(\mu - 2\mu_2)}{\mu} + 1. \tag{2.17}$$

From CK, we know that $\mu = \mu_1 = -1.781430$, and $\mu_2 = -1.138209$, and hence, we have $b^{CY} = 0.7221a_2^{CY} + 2$ and $a_1^{CY} = -0.2779a_2^{CY} + 1$. The first order condition with respect to a_2^{CY} gives:

$$a_2^{CY} = -\frac{2.8884Var(\widetilde{\beta}) - 0.5558Var(\widetilde{\beta}_1) - 0.3326Cov(\widetilde{\beta},\widetilde{\beta}_1) - 4Cov(\widetilde{\beta},\widetilde{\beta}_2) + 2Cov(\widetilde{\beta}_1,\widetilde{\beta}_2)}{1.0429Var(\widetilde{\beta}) + 0.1545Var(\widetilde{\beta}_1) + 2Var(\widetilde{\beta}_2) + 0.8026Cov(\widetilde{\beta},\widetilde{\beta}_1) - 2.8884Cov(\widetilde{\beta},\widetilde{\beta}_2) - 1.1116Cov(\widetilde{\beta}_1,\widetilde{\beta}_2)}. \tag{2.18}$$

It is straightforward to check that this is the global minimizer as the objective function is quadratic and convex.

To calculate the weights, it is imperative to obtain $Var(\widetilde{\beta})$, $Var(\widetilde{\beta}_1)$, $Var(\widetilde{\beta}_2)$, $Cov(\widetilde{\beta},\widetilde{\beta}_1)$, $Cov(\widetilde{\beta},\widetilde{\beta}_2)$, and $Cov(\widetilde{\beta}_1,\widetilde{\beta}_2)$. Instead of obtaining the variances and the covariances from the finite sample distributions, we will calculate them from the corresponding asymptotic distributions.

First, the variances can be computed by combining the techniques of White (1961) and CK. Note that:

$$\begin{aligned}
n^2 Var(\widetilde{\beta}) &= n^2\left[E\left(\widetilde{\beta}^2\right) - E^2\left(\widetilde{\beta}\right)\right] \\
&= n^2\left[1 + \frac{2E\left(\int_0^1 WdW/\int_0^1 W^2\right)}{n} + \frac{E\left(\int_0^1 WdW/\int_0^1 W^2\right)^2}{n^2} + o(n^{-2})\right] \\
&\quad -n^2\left[1 + \frac{2E\left(\int_0^1 WdW/\int_0^1 W^2\right)}{n} + \frac{\left[E\left(\int_0^1 WdW/\int_0^1 W^2\right)\right]^2}{n^2} + o(n^{-2})\right] \\
&= E\left(\frac{\int_0^1 WdW}{\int_0^1 W^2}\right)^2 - \mu^2 + o(1)\ ,
\end{aligned}$$

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

We need to compute the first term on the right hand side of these two equations.

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 $D(a,b)$. White (1961) gave the formula for calculating the second moment of $N(a,b)/D(a,b)$ as:

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

Following Phillips (1987b), CK obtained the expression for $M_{a,b}(\theta_1,\theta_2)$:

$$M_{a,b}(\theta_1,-\theta_2) = \exp\left[-\frac{\theta_1}{2}(b-a)\right]\left\{\cosh\left[\lambda(b-a)\right] - \frac{1}{\lambda}\left[\theta_1 + a(\theta_1^2-\lambda^2)\right]\sinh\left[\lambda(b-a)\right]\right\}^{-1/2}, \tag{2.19}$$

where $\lambda = \sqrt{-2\theta_2}$. The following proposition calculates the variances.

Proposition 2.1 *The second derivative of $M_{a,b}$ with respect to θ_1, evaluated at $\theta_1 = 0$, is given by*

$$\frac{1}{4}(b-a)^2 H_0^{-1/2} + \frac{1}{2}H_0^{-3/2}\left[(b-a)H'(0) - H''(0)\right] + \frac{3}{4}H_0^{-5/2}\left(H'(0)\right)^2 := v(a,b), \tag{2.20}$$

where H, H_0, $H'(0)$, $H''(0)$ are given in the Appendix.

When $[a,b] = [0,1]$,

$$v(0,1) = \frac{1}{4}\cosh^{-1/2}(\lambda) - \frac{1}{2\lambda}\cosh^{-3/2}(\lambda)\sinh(\lambda) + \frac{3}{4\lambda^2}\cosh^{-5/2}(\lambda)\sinh^2(\lambda).$$

This leads to the approximate variance for the full sample estimator and the first subsample estimator in the discrete time unit root model:

$$n^2 Var(\widetilde{\beta}) = l^2 Var(\widetilde{\beta}_1) = 10.1123 + O(n^{-1}). \tag{2.21}$$

When $[a,b] = [1/2,1]$,

$$\begin{aligned} v(1/2,1) &= \frac{1}{16}\left[\cosh\left(\frac{\lambda}{2}\right) + \frac{\lambda}{2}\sinh\left(\frac{\lambda}{2}\right)\right]^{-1/2} - \frac{1}{4\lambda}\left[\cosh\left(\frac{\lambda}{2}\right) + \frac{\lambda}{2}\sinh\left(\frac{\lambda}{2}\right)\right]^{-3/2}\sinh\left(\frac{\lambda}{2}\right) \\ &\quad + \frac{3}{4\lambda^2}\left[\cosh\left(\frac{\lambda}{2}\right) + \frac{\lambda}{2}\sinh\left(\frac{\lambda}{2}\right)\right]^{-5/2}\sinh^2\left(\frac{\lambda}{2}\right). \end{aligned}$$

This leads to the approximate variance for the second subsample estimator in the discrete time unit root model:

$$l^2 Var(\widetilde{\beta}_2) = 5.3612 + O(n^{-1}). \tag{2.22}$$

Remark 2.2 *The variance of the full sample estimator, $\widetilde{\beta}$, normalized by n^2, is 10.1123. Interestingly, it is the same as that of the first subsample estimator, normalized by l^2. This equality arises because the full sample has the same initial value as the first subsample.*

Remark 2.3 *The variance of the first subsample estimator is about twice as large as that of the second subsample estimator. This difference is due to the distinctive initial conditions and may be made clearer*

Table 2: Variances of subsample estimators

j	Variance normalized by l^2
1	10.1122
2	5.3612
3	4.2839
4	3.7065
5	3.3268
6	3.0507
7	2.8375
8	2.6660
9	2.5238
10	2.4034
11	2.2995
12	2.2087

by examining the limit distribution of $l(\widetilde{\beta}_j - 1)$*,*

$$l(\widetilde{\beta}_j - 1) \Rightarrow \frac{\int_{(j-1)/m}^{j/m} W dW}{m \int_{(j-1)/m}^{j/m} W^2}. \tag{2.23}$$

In (2.23), the numerator plays an insignificant role in the variance calculation since it can be rewritten as a difference between two chi-square variates. The important part is in the denominator, which is a partial sum of chi-square variates. For the first subsample estimator, the denominator starts with the square of zero. However, the denominator of the second subsample begins with the square of a random variable whose magnitude is $O_p(\sqrt{n})$*. This random initialization of the second subsample increases the magnitude of the denominator and, hence, tends to reduce the absolute value of the ratio. To confirm our explanation, we carried out a simple Monte Carlo study, in which data were simulated from the Brownian motion with the sample length being set at 5000, and the number of replications set at 10,000. The mean of* $\int_0^{1/2} W^2$ *is* 0.1243*, whereas the mean of* $\int_{1/2}^1 W^2$ *is* 0.3694*. Not only is the mean affected, but also the variance and the entire distribution. In this case, the two variances are* 0.0202 *and* 0.2169*, respectively.*

Remark 2.4 *Table 2 lists the variances of all the subsample estimators when* $m = 12$*. 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 estimator will be the same as that of the LS estimator, as pointed out by CK.*

Second, to calculate the covariances, we note that:

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

$$n^2 Cov(\widetilde{\beta}_1, \widetilde{\beta}_2) = E\left(\frac{\int_0^{1/2} WdW}{\int_0^{1/2} W^2} \frac{\int_{1/2}^{1} WdW}{\int_{1/2}^{1} W^2}\right) - m^2\mu_1\mu_2 + O(n^{-1}) .$$

It is required to obtain the expected value of the cross product of random variables. As for the variances, we also calculate these expected values from the joint MGF given in the following lemma.

Lemma 2.5 *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:*

$$E\left(\frac{N(a,b)}{D(a,b)} \frac{N(c,d)}{D(c,d)}\right) = \int_0^\infty \int_0^\infty \frac{\partial M_{a,b,c,d}(\theta_1, -\theta_2, \varphi_1, -\varphi_2)}{\partial\theta_1 \partial\varphi_1} |_{\theta_1=0,\varphi_1=0}\, d\theta_2 d\varphi_2. \tag{2.24}$$

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 when $m = 2$.

Proposition 2.6 *The MGF $M_{0,1,0,1/2}(\theta_1, \theta_2, \varphi_1, \varphi_2)$ is given by*

$$\left\{\left[\cosh\left(\frac{\lambda_{0,1,0,1/2}}{2}\right) - \frac{\theta_1}{\lambda_{0,1,0,1/2}} \sinh\left(\frac{\lambda_{0,1,0,1/2}}{2}\right)\right]\left[\cosh\left(\frac{\eta_{0,1,0,1/2}}{2}\right) - \frac{\pi_{0,1,0,1/2}}{\eta_{0,1,0,1/2}} \sinh\left(\frac{\eta_{0,1,0,1/2}}{2}\right)\right]\right\}^{-1/2}$$
$$\exp\left(-\frac{\theta_1}{2} - \frac{\varphi_1}{4}\right);$$

the MGF $M_{0,1,1/2,1}(\theta_1, \theta_2, \varphi_1, \varphi_2)$ is given by:

$$\left\{\left[\cosh\left(\frac{\lambda_{0,1,1/2,1}}{2}\right) - \frac{\theta_1 + \varphi_1}{\lambda_{0,1,1/2,1}} \sinh\left(\frac{\lambda_{0,1,1/2,1}}{2}\right)\right]\left[\cosh\left(\frac{\eta_{0,1,1/2,1}}{2}\right) - \frac{\pi_{0,1,1/2,1}}{\eta_{0,1,1/2,1}} \sinh\left(\frac{\eta_{0,1,1/2,1}}{2}\right)\right]\right\}^{-1/2}$$
$$\exp\left(-\frac{\theta_1}{2} - \frac{\varphi_1}{4}\right);$$

and the MGF $M_{0,1/2,1/2,1}(\theta_1, \theta_2, \varphi_1, \varphi_2)$ is given by:

$$\left\{\left[\cosh\left(\frac{\lambda_{0,1/2,1/2,1}}{2}\right) - \frac{\varphi_1}{\lambda_{0,1/2,1/2,1}} \sinh\left(\frac{\lambda_{0,1/2,1/2,1}}{2}\right)\right]\left[\cosh\left(\frac{\eta_{0,1/2,1/2,1}}{2}\right) - \frac{\pi_{0,1/2,1/2,1}}{\eta_{0,1/2,1/2,1}} \sinh\left(\frac{\eta_{0,1/2,1/2,1}}{2}\right)\right]\right\}^{-1/2}$$
$$\exp\left(-\frac{\theta_1}{4} - \frac{\varphi_1}{4}\right),$$

where $\lambda_{a,b,cd}$, $\eta_{a,b,c,d}$ and $\pi_{a,b,c,d}$ are given in the Appendix. These MGFs, together with formula (2.24),

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

Variance	$n^2Var(\widetilde{\beta})$	$l^2Var(\widetilde{\beta}_1)$	$l^2Var(\widetilde{\beta}_2)$
	10.1123	10.1123	5.3612
Covariance	$n^2Cov(\widetilde{\beta},\beta_1)$	$n^2Cov(\widetilde{\beta},\widetilde{\beta}_2)$	$n^2Cov(\beta_1,\widetilde{\beta}_2)$
	10.0376	11.5863	4.4212

lead to the approximate covariances between the full sample estimator and the two subsample estimators:

$$n^2Cov(\widetilde{\beta},\widetilde{\beta}_1) = 10.0376 + O(n^{-1}); \tag{2.25}$$

$$n^2Cov(\widetilde{\beta},\widetilde{\beta}_2) = 11.5863 + O(n^{-1}); \tag{2.26}$$

$$n^2Cov(\widetilde{\beta}_1,\widetilde{\beta}_2) = 4.4212 + O(n^{-1}). \tag{2.27}$$

Remark 2.7 *There are several interesting findings in Proposition 2.6. 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.*

Remark 2.8 *Table 3 summarizes the approximate values of the variances and covariances when $m = 2$.*

Theorem 2.9 *The optimal weights for the jackknife estimator of β in the discrete time unit root model when $m = 2$ are $b^{CY} = 2.8390$, $a_1^{CY} = 0.6771$, and $a_2^{CY} = 1.1619$. Hence, the optimal jackknife estimator is*

$$\widetilde{\beta}_{JK}^{CY} = 2.8390\widetilde{\beta} - (0.6771\widetilde{\beta}_1 + 1.1619\widetilde{\beta}_2). \tag{2.28}$$

Remark 2.10 *The optimal jackknife estimator compares interestingly to the estimator of CK:*

$$\widetilde{\beta}_{JK}^{CK} = 2.5651\widetilde{\beta} \;- (0.7825\widetilde{\beta}_1 + 0.7825\widetilde{\beta}_2). \tag{2.29}$$

Relative to the CK estimator, our estimator gives a larger weight to the full sample estimator, and more so, to the second subsample estimator. The weight for the second subsample is nearly twice as much as that for the first subsample. Since the variance of the second subsample estimator is nearly half as that of the first sample estimator, not surprisingly, the variance of our estimator is smaller than that of CK's.

3 Optimal Jackknife for the Continuous Time Unit Root Model

In the section, we extend the result to a continuous time unit root model. In the continuous time stationary models, it is well known that the estimation bias of the mean reversion parameter depends on the span of the data, but not the number of observations. In empirically realistic situations, the time span (measured in number of years) is often small. Consequently, the bias can be much more substantial than that in the discrete time models; see for example, PY (2005) and Yu (2011).

Considering the following Vasicek model with $y_0 = 0$:

$$dy_t = -\kappa y dt + \sigma dW_t. \tag{3.1}$$

The parameter of interest here is κ that captures the persistence of the process. When $\kappa > 0$, the process is stationary and κ determines the speed of mean reversion. The observed data are assumed to be recorded discretely at $(h, 2h, \cdots, nh(=T))$ in the time interval $(0, T]$. So h is the sample interval, n is the total number of observations and T is the time span. In this case, Yu (2010) showed that the bias of the ML estimator of κ is:

$$E(\widetilde{\kappa}) - \kappa_0 = \frac{1}{2T}\left(3 + e^{2\kappa h}\right) + o(T^{-1}). \tag{3.2}$$

It is clear from (3.2) that the bias depends on T, κ, and h. When κ is close to 0, or h is close to 0, which is empirically realistic, the bias is about $2/T$. When T is not too big, this bias is very big, relative to the true value of κ. The result can be extended to the Vasicek model with an unknown mean and to the square root model; see Tang and Chen (2009). The large bias motivated PY (2005) to use the jackknife method.

When $\kappa = 0$, the model has a unit root, and the exact discrete time representation is a random walk. In this case, it can be shown that the bias of the ML estimator of κ is:

$$\frac{1.7814}{T} + o(T^{-1}). \tag{3.3a}$$

The bias formula for $\widetilde{\kappa}$ is similar to that for $\widetilde{\beta}$ in (2.4). However, the direction of the bias is opposite and the bias depends on T, not n. When T is not big, this bias is very big relative to the true value of κ. If $T \to \infty$, the so-called long span limit distribution of $\widetilde{\kappa}$ is given by:

$$T\left(\widetilde{\kappa}-1\right) \Rightarrow -\frac{\int_0^1 WdW}{\int_0^1 W^2}.$$

See, for example, Phillips (1987b) and Zhou and Yu (2010). Similarly, the long span limit distributions of the sub-sample estimators are:

$$T(\widetilde{\kappa}_j - 1) \Rightarrow -\frac{\int_{(j-1)/m}^{j/m} WdW}{T\int_{(j-1)/m}^{j/m} W^2}, j = 1, \ldots, m.$$

Obviously, the only difference between these limit distributions and those in the discrete time unit root model is the minus sign. Hence, the variances and covariances of the two sets of limit distribution are the same but the expected values of them change the sign. Of course, the rate of convergence changes from n to T. Consequently, we have the following theorem.

Theorem 3.1 *When $m = 2$, the approximate variance for the full sample estimator and the first subsample estimator in the continuous time unit root model is:*

$$T^2 Var(\widetilde{\kappa}) = \frac{1}{4}T^2 Var(\widetilde{\kappa}_1) = 13.2867 + O(T^{-1}). \tag{3.4}$$

Similarly, the approximate variance for the second subsample estimator in the continuous time unit root model is:

$$\frac{1}{4}T^2 Var(\widetilde{\kappa}_2) = 5.3612 + O(T^{-1}). \tag{3.5}$$

The approximate covariances between the full sample estimator and the two subsample estimators are:

$$T^2 Cov(\widetilde{\kappa}, \widetilde{\kappa}_1) = 10.0376 + O(T^{-1}); \tag{3.6}$$

$$T^2 Cov(\widetilde{\kappa}, \widetilde{\kappa}_2) = 11.5863 + O(T^{-1}); \tag{3.7}$$

$$T^2 Cov(\widetilde{\kappa}_1, \widetilde{\kappa}_2) = 4.4212 + O(T^{-1}). \tag{3.8}$$

The optimal jackknife estimator of κ in the continuous time unit root model when $m = 2$ is:

$$\widetilde{\kappa}_m^{CY} = 2.8390\widetilde{\kappa} \ - (\ 0.6771\widetilde{\kappa}_1 \ + \ 1.1619\widetilde{\kappa}_2\ ). \tag{3.9}$$

Remark 3.2 *The optimal jackknife estimator of κ in the continuous time unit root model has the same weights and the expression as that of β in the discrete time unit root model. This is not surprising because there are only two differences in the limit theory, the sign of the bias and the rate of convergence. These differences do not have any impact on the weights as they are canceled out.*

Remark 3.3 *The effect of the initial condition on the bias in the continuous time model was recently pointed out in Yu (2011). When the initial value is 0 in Model (3.1) with $\kappa > 0$, Yu showed that the bias of the ML estimator of κ is:*

$$E(\widetilde{\kappa}) - \kappa_0 = \frac{1}{2T}\left(3 + e^{2\kappa h}\right) + o(T^{-1}). \tag{3.10}$$

However, when the initial value is $N(0, \sigma^2/2\kappa)$ in Model (3.1) with $\kappa \searrow 0$, the bias of the ML estimator of κ is:

$$E(\widetilde{\kappa}) - \kappa_0 = \frac{1}{2T}\left(3 + e^{2\kappa h}\right) - \frac{2(1 - e^{-2n\kappa h})}{Tn(1 - e^{-2\kappa h})} + o(T^{-1}). \tag{3.11}$$

4 Monte Carlo Studies

In this section, we check the finite sample properties of the proposed jackknife against the CK jackknife using simulated data. This is important for two reasons. First, it is important to measure the efficiency gain of the proposed method. Second, the weights are obtained based on the variances and the covariances of the limit distributions but not based on the finite sample distributions. These approximations may or may not have impacts on the optimal weights. Hence, it is informative to examine the importance of the approximation error. Although it is difficult to obtain the analytical expressions for the variances and the covariances of the finite sample distribution, we can compute them from simulated data in a Monte Carlo study, provided the number of replications is large enough. In this paper, we always set the number of replications at 5,000.

4.1 Discrete Time Unit Root Model

First, we simulate data from the following discrete time unit root model with initial value $y_0 = 0$:

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

We evaluate the performance of alternative jackknife methods by applying them to five different sample sizes, i.e. $n = 12, 24, 48, 96, 108$. In particular, we compare three methods when $m = 2$, the CK jackknife method based on (2.29), the CY jackknife method based on (2.28) 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.

The results are reported in Table 4, where for each case we calculate the mean, the variance and the RMSE for the estimates of β, all across 5,000 sample paths. In addition, we calculate the ratio of the variances and the RMSEs for the CK estimates and the proposed CY estimates to measure the efficiency loss in using the CK estimate.

Table 4: Finite sample performance of alternative jackknife estimators for the discrete time unit root model when $m = 2$

n	Statistics	CK	CY	Efficiency of CK relative to CY	Exact CY
12	Bias	-0.0637	-0.0665	–	-0.0671
	Var*100	17.9035	16.1215	0.9005	16.0522
	RMSE*10	4.2789	4.0698	0.9511	4.0624
36	Bias	-0.0073	-0.0078	–	-0.0079
	Var*100	1.5185	1.3181	0.8680	1.3060
	RMSE*10	1.2344	1.1507	0.9322	1.1456
48	Bias	-0.0064	-0.0062	–	-0.0061
	Var*100	0.8648	0.7570	0.8753	0.7534
	RMSE*10	0.9322	0.8723	0.9357	0.8701
96	Bias	-0.0012	-0.0014	–	-0.0014
	Var*100	0.2266	0.1989	0.8777	0.1982
	RMSE*10	0.4762	0.4462	0.9370	0.4454
108	Bias	-0.0015	-0.0015	–	-0.0016
	Var*100	0.1761	0.1566	0.8892	0.1564
	RMSE*10	0.4199	0.3960	0.9432	0.3958

Several interesting results emerge from the table. First, the bias in the estimate of β for the CK method and the CY method is very similar in every case. Second, the variance in the estimate of β for the CY method is significantly smaller than that for the CK method in each case. The reduction in the variance is over 10% in all cases. Consequently, the RMSE is smaller for the CY method. Third, although the exact CY method provides a smaller variance, the difference between the two CY methods is so small, suggesting that the proposed CY works well, and it is not necessary to bear the additional computational cost associated with calculating the variances and covariances from the finite sample distributions. It is worth pointing out that even when $n = 12$, a very small sample size that could have bigger implications for the approximation error, the difference between the two CY methods is still negligible in terms of the bias, variance and RMSE.

To understand why the two CY estimates are so similar, Table 5 reports the weights obtained from the finite sample distributions. For the purpose of comparison, we also report the weights obtained from the limit distribution. It can be clearly seen that when $n \to \infty$, all the weights converge. When n is as small as 12, the weights are not very different from those obtained when n is infinite.

4.2 Continuous Time Unit Root Model

Second, we simulate data from the following continuous time unit root model with initial value $y_0 = 0$:

$$dy_t = -\kappa y dt + dW_t,$$

Table 5: Jackknife weights based on the finite sample distributions and the limit distribution for the discrete time unit root model.

n	b^{CY}	a_1^{CY}	a_2^{CY}
12	2.9044	0.6494	1.2525
36	2.9244	0.6442	1.2802
48	2.8993	0.6539	1.2453
96	2.8915	0.6569	1.2346
108	2.8699	0.6652	1.2047
∞	2.8390	0.6771	1.1619

Table 6: Finite sample performance of alternative jackknife estimators for the continuous time unit root model when $m = 2$.

n	h	T	Statistics	CK	CY	Efficiency of CK relative to CY	Exact CY
104	1/52	2	Bias	0.0026	0.0109	–	0.0116
			Var	6.0625	5.3898	0.8890	5.3832
			RMSE	2.4622	2.3216	0.9429	2.3202
260	1/52	5	Bias	-0.0080	-0.0069	–	-0.0069
			Var	0.9737	0.8779	0.9016	0.8778
			RMSE	0.9868	0.9370	0.9495	0.9369
504	1/252	2	Bias	-0.0071	-0.0101	–	-0.0107
			Var	6.0505	5.3099	0.8776	5.2893
			RMSE	2.4598	2.3043	0.9368	2.2999
1260	1/252	5	Bias	0.0055	0.0015	–	0.0008
			Var	0.9292	0.8199	0.8823	0.8178
			RMSE	0.9640	0.9055	0.9393	0.9043

with $\kappa = 0$. Here, we generate data based on its exact discrete form as $y_{t+h} = y_t + \varepsilon_t$, with $\varepsilon_t \sim N(0, h)$. As for the discrete time model, we employ the three jackknife method to estimate κ. Table 6 shows the results based on two sampling intervals $h = 1/52, 1/252$, corresponding to the weekly and daily frequency. The time span, T, is set at 2 years and 5 years. These settings are empirically realistic for modeling interest rates and volatility.

Several interesting results also emerge from the table. First. the bias in the estimate of κ for the CK method and the CY method is very similar in each case. Second, the variance in the estimate of κ for the CY method is always significantly smaller than those for the CK method. The reduction in the variance is at least 10%. Consequently, the RMSE is smaller for the CY method. Third, the difference between the two CY methods is very small, suggesting that the proposed CY works well and it is not necessary to bear the additional computational cost associated with calculating the variances and covariances from the finite sample distributions.

Table 7: Weights based on the finite sample distributions and the limit distribution

n	h	T	b^{CY}	a_1^{CY}	a_2^{CY}
104	1/52	2	2.8684	0.6658	1.2027
260	1/52	5	2.8483	0.6735	1.1748
504	1/252	2	2.8931	0.6563	1.2368
1260	1/252	5	2.8823	0.6605	1.2218
∞	fixed	∞	2.8390	0.6771	1.1619

To understand why the two CY estimates are so similar in the continuous time model, Table 7 reports the weights obtained from the finite sample distributions. For the purpose of comparison, we also report the weights obtained from the limit distribution. It can be clearly seen, as T (but not n) increases, all the weights get close to those obtained from the limit distribution. When T is as small as 2, the weights are not different from those obtained when T is infinite.

5 Conclusion

This paper has introduced a new jackknife procedure for unit root models that offers improvement over the jackknife methodology of CK (2010). The proposed estimator is optimal in the sense that it minimizes the variance of the jackknife estimator while maintaining the desirable property of being able to remove the first order bias. The new method is applicable to both discrete time and continuous time unit root models. Simulation studies have shown that this new method reduces the variance by more than 10% relative to the estimator of CK without compromising the bias.

Although it is not pursued in the present paper, it may be useful to further improve our method by using alternative values for m, and hence, further reduce the variance and RMSE. Furthermore, there are other models where the asymptotic theory is critically dependent on the initial condition. Examples would include near unit root models and explosive processes. It may be interesting to extend the results in the present paper to these models. We plan to report these results in our later work.

APPENDIX

A Proof of Proposition 2.1

The notations in the proposition are defined as follows.

$$H = H(\theta_1) = \cosh[\lambda(b-a)] - \frac{1}{\lambda}[\theta_1 + a(\theta_1^2 - \lambda^2)]\sinh[\lambda(b-a)];$$

$H_0 = H(0)$, i.e.:

$$H_0 = \cosh[\lambda(b-a)] + a\lambda \sinh[\lambda(b-a)];$$

$H^{'}(0)$ denotes the first derivative of $H(\theta_1)$ with respect to θ_1, evaluated at $\theta_1 = 0$, i.e.:

$$H^{'}(0) = \frac{\partial H}{\partial \theta_1} \mid_{\theta_1=0} = -\frac{1}{\lambda} \sinh[\lambda(b-a)].$$

Similarly, $H^{''}(0)$ denotes the second derivative of $H(\theta_1)$ with respect to θ_1, evaluated at $\theta_1 = 0$, i.e.:

$$H^{''}(0) = \frac{\partial^2 H}{\partial \theta_1^2} \mid_{\theta_1=0} = -\frac{2a}{\lambda} \sinh[\lambda(b-a)].$$

The first derivative of $M_{a,b}$ with respect to θ_1 can be written as:

$$\frac{\partial M_{a,b}}{\partial \theta_1} = \left(-\frac{b-a}{2}\right) \exp\left[-\frac{\theta_1(b-a)}{2}\right] H^{-1/2} - \frac{1}{2} \exp\left[-\frac{\theta_1(b-a)}{2}\right] H^{-3/2} \frac{\partial H}{\partial \theta_1}.$$

The second derivative of $M_{a,b}$ with respect to θ_1 is:

$$\begin{aligned} \frac{\partial^2 M_{a,b}}{\partial \theta_1^2} \quad = \quad & \frac{1}{4}(b-a)^2 \exp\left[-\frac{\theta_1(b-a)}{2}\right] H^{-1/2} + \frac{1}{2} \exp\left[-\frac{\theta_1(b-a)}{2}\right] H^{-3/2} \left[(b-a)\frac{\partial H}{\theta_1} - \frac{\partial^2 H}{\partial \theta_1^2}\right] \\ & + \frac{3}{4} \exp\left[-\frac{\theta_1(b-a)}{2}\right] H^{-5/2} \left(\frac{\partial H}{\partial \theta_1}\right)^2 \end{aligned}$$

Setting $\theta_1 = 0$ and by standard numerical integration techniques, we can obtain the results in Proposition 2.1.

B Proof of Lemma 2.5

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,

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

Consequently,

$$\begin{aligned}
& \int_0^\infty \int_0^\infty \left\{ \partial \left[\frac{\partial M_{a,b,c,d}(\theta_1, -\theta_2, \varphi_1, -\varphi_2)}{\partial \theta_1} \mid_{\theta_1=0} \right] / \partial \varphi_1 \mid_{\varphi_1=0} \right\} d\theta_2 d\varphi_2 \\
= & \int_0^\infty \int_0^\infty E\left\{ N(a,b)N(c,d) \exp\left[-\theta_2 D(a,b) - \varphi_2 D(c,d)\right]\right\} d\theta_2 d\varphi_2 \\
= & E \int_0^\infty N(c,d) \exp[-\varphi_2 D(c,d)] d\varphi_2 \int_0^\infty N(a,b) \exp[-\theta_2 D(a,b)] d\theta_2 \\
= & E\left[\frac{N(a,b)}{D(a,b)} \frac{N(c,d)}{D(c,d)} \right].
\end{aligned}$$

C Proof of Proposition 2.6

To prove this proposition, we follow CK.

(1) STEP 1: Deriving the MGF for $E\left(\frac{\int_0^1 WdW}{\int_0^1 W^2}\frac{\int_0^{1/2} WdW}{\int_0^{1/2} W^2}\right)$.

Let $X(t)$ and $Y(t)$ $(t \in [0,1])$ be the OU process defined by

$$dX(t) = \gamma X(t)dt + dW(t), X(0) = 0,$$

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

The measures induced by X and Y are denoted as μ_X and μ_Y, respectively. By Girsanov's Theorem,

$$\frac{d\mu_X}{d\mu_Y}(s) = \exp\left[(\gamma - \lambda)\int_0^1 s(t)ds(t) - \frac{\gamma^2 - \lambda^2}{2}\int_0^1 s(t)^2 dt\right],$$

where the left side is the Radon-Nikodym derivative evaluated at $s(t)$.

Let the MGF of $E\left(\frac{\int_0^1 WdW}{\int_0^1 W^2}\frac{\int_0^{1/2} WdW}{\int_0^{1/2} W^2}\right)$ be $M_{0,1,0,1/2}(\theta_1, \theta_2, \varphi_1, \varphi_2)$ so that

$$M_{0,1,0,1/2}(\theta_1, \theta_2, \varphi_1, \varphi_2) = E\left[\exp\left(\theta_1 \int_0^1 WdW + \theta_2 \int_0^1 W^2 + \varphi_1 \int_0^{1/2} WdW + \varphi_2 \int_0^{1/2} W^2\right)\right].$$

With the change of measure , we have the following formula by setting $\gamma = 0$:

$$\begin{aligned}
& M_{0,1,0,1/2}(\theta_1, \theta_2, \varphi_1, \varphi_2) \\
= & E\left[\exp\left(\theta_1 \int_0^1 YdY + \theta_2 \int_0^1 Y^2 + \varphi_1 \int_0^{1/2} YdY + \varphi_2 \int_0^{1/2} Y^2 - \lambda \int_0^1 YdY + \frac{\lambda^2}{2}\int_0^1 Y^2\right)\right].
\end{aligned}$$

By Ito's calculus, we get:

$$
\begin{aligned}
& M_{0,1,0,1/2}\left(\theta_1,\theta_2,\varphi_1,\varphi_2\right) \\
= \;& \exp\left(-\frac{\theta_1}{2}-\frac{\varphi_1}{4}+\frac{\lambda}{2}\right) \\
& E\left\{\exp\left[\left(\frac{\theta_1-\lambda}{2}\right)Y(1)^2+\frac{\varphi_1}{2}Y\left(\frac{1}{2}\right)^2+\left(\varphi_2+\theta_2+\frac{\lambda^2}{2}\right)\int_0^{1/2}Y^2+\left(\theta_2+\frac{\lambda^2}{2}\right)\int_{1/2}^{1}Y^2\right]\right\} \\
= \;& \exp\left(-\frac{\theta_1}{2}-\frac{\varphi_1}{4}+\frac{\lambda}{2}\right)E\left\{\exp\left[\left(\frac{\theta_1-\lambda}{2}\right)Y(1)^2+\frac{\varphi_1}{2}Y\left(\frac{1}{2}\right)^2+\varphi_2\int_0^{1/2}Y^2\right]\right\}
\end{aligned}
$$

where $\lambda=\sqrt{-2\theta_2}$. To calculate the expectation, we first take the expectation of $M_{0,1,0,1/2}(\theta_1,\theta_2,\varphi_1,\varphi_2)$ conditional on $\mathcal{F}_0^{1/2}$, which is the sigma field generated by W on $[0,\frac{1}{2}]$, and then change the measure by Girsanov's Theorem with introducing another OU process.

Note that,

$$
\begin{aligned}
& M_{0,1,0,1/2}(\theta_1,\theta_2,\varphi_1,\varphi_2,\mathcal{F}_0^{1/2}) = E\left[M_{0,1,0,1/2}(\theta_1,\theta_2,\varphi_1,\varphi_2)\mid \mathcal{F}_0^{1/2}\right] \\
= \;& \exp\left(-\frac{\theta_1}{2}-\frac{\varphi_1}{4}+\frac{\lambda}{2}\right)\exp\left[\frac{\varphi_1}{2}Y\left(\frac{1}{2}\right)^2+\varphi_2\int_0^{1/2}Y^2\right]E\left\{\exp\left[\left(\frac{\theta_1-\lambda}{2}\right)Y(1)^2\mid Y\left(\frac{1}{2}\right)\right]\right\}.
\end{aligned}
$$

The MGF of the noncentral χ^2 distribution, $(\frac{\theta_1-\lambda}{2})Y(1)^2$, is $[1-(\theta_1-\lambda)\varpi^2]^{-1/2}\exp\left[\frac{(\theta_1-\lambda)}{2}\zeta Y\left(\frac{1}{2}\right)^2\right]$ with $\varpi^2=\frac{\exp(\lambda)-1}{2\lambda}$ and $\zeta=\left[1-(\theta_1-\lambda)\varpi^2\right]^{-1}\exp(\lambda)$. Therefore,

$$
\begin{aligned}
& M_{0,1,0,1/2}\left(\theta_1,\theta_2,\varphi_1,\varphi_2,\mathcal{F}_0^{1/2}\right) \\
= \;& [1-(\theta_1-\lambda)\varpi^2]^{-1/2}\exp\left(-\frac{\theta_1}{2}-\frac{\varphi_1}{4}+\frac{\lambda}{2}\right)E\left(\exp\left\{\left[\frac{(\theta_1-\lambda)}{2}\zeta+\frac{\varphi_1}{2}\right]Y\left(\frac{1}{2}\right)^2+\varphi_2\int_0^{1/2}Y^2\right\}\right).
\end{aligned}
$$

Now we introduce another process $Z(t)$ on $[0,\frac{1}{2}]$, given by $dZ(t)=\eta Z(t)dt+dW(t)$, $Z(0)=0$. By changing the measure and setting $\varphi_2=\left(\lambda^2-\eta^2\right)/2$ to cancel out $\int_0^{1/2}Z^2$, we have

$$
\begin{aligned}
& E\left(\exp\left\{\left[\frac{(\theta_1-\lambda)}{2}\zeta+\frac{\varphi_1}{2}\right]Y\left(\frac{1}{2}\right)^2+\varphi_2\int_0^{1/2}Y^2\right\}\right) \\
= \;& E\left(\exp\left\{\left[\frac{(\theta_1-\lambda)}{2}\zeta+\frac{\varphi_1}{2}\right]Z\left(\frac{1}{2}\right)^2+(\lambda-\eta)\int_0^{1/2}ZdZ\right\}\right) \\
= \;& \exp\left(\frac{\eta-\lambda}{4}\right)E\left(\exp\left[\tau Z\left(\frac{1}{2}\right)^2\right]\right),
\end{aligned}
$$

with $\tau = \frac{(\theta_1-\lambda)}{2}\zeta + \frac{\varphi_1}{2} + \frac{\lambda-\eta}{2}$. Considering $Z(\frac{1}{2}) \sim N(0, \varpi_z^2)$ where $\varpi_z^2 = \frac{\exp(\eta)-1}{2\eta}$, we have

$$E\left(\exp\left\{\left[\frac{(\theta_1-\lambda)}{2}\zeta + \frac{\varphi_1}{2}\right] Y\left(\frac{1}{2}\right)^2 + \varphi_2\int_0^{1/2} Y^2\right\}\right) = [1-2\tau\varpi_z^2]^{-1/2}\exp\left(\frac{\eta-\lambda}{4}\right).$$

Finally by combining the equations above, we obtain

$$M_{0,1,0,1/2}(\theta_1,\theta_2,\varphi_1,\varphi_2) = \left\{[1-(\theta_1-\lambda)\varpi^2][1-2\tau\varpi_z^2]\exp\left(\frac{\eta+\lambda}{-2}\right)\right\}^{-1/2}\exp\left(-\frac{\theta_1}{2}-\frac{\varphi_1}{4}\right).$$

Note that,

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

and that

$$[1-2\tau\varpi_z^2]\exp\left(\frac{\eta}{-2}\right) = \frac{\exp\left(\frac{\eta}{2}\right)+\exp\left(-\frac{\eta}{2}\right)}{2} - \frac{\pi}{\eta}\frac{\exp\left(\frac{\eta}{2}\right)-\exp\left(-\frac{\eta}{2}\right)}{2} = \cosh\left(\frac{\eta}{2}\right) - \frac{\pi}{\eta}\sinh(\frac{\eta}{2}),$$

where $\pi = (\theta_1-\lambda)\zeta + \varphi_1 + \lambda$. Thus,

$$M_{0,1,0,1/2}(\theta_1,\theta_2,\varphi_1,\varphi_2) = H_{0,1,0,1/2}(\theta_1,\theta_2,\varphi_1,\varphi_2)^{-1/2}\exp\left(-\frac{\theta_1}{2}-\frac{\varphi_1}{4}\right),$$

with $H_{0,1,0,1/2}(\theta_1,\theta_2,\varphi_1,\varphi_2) = \left[\cosh\left(\frac{\lambda}{2}\right) - \frac{\theta_1}{\lambda}\sinh\left(\frac{\lambda}{2}\right)\right]\left[\cosh\left(\frac{\eta}{2}\right) - \frac{\pi}{\eta}\sinh\left(\frac{\eta}{2}\right)\right]$.

(2) STEP 2: Compute $\partial\left[\frac{\partial M_{a,b,c,d}(\theta_1,-\theta_2,\varphi_1,-\varphi_2)}{\partial\theta_1}\mid_{\theta_1=0}\right]/\partial\varphi_1\mid_{\varphi_1=0}$.

Denote $H_{0,1,0,1/2}(\theta_1,-\theta_2,\varphi_1,-\varphi_2)$ and $M_{0,1,0,1/2}(\theta_1,-\theta_2,\varphi_1,-\varphi_2)$ by H and M. Note that $M = H^{-1/2}\exp\left(-\frac{\theta_1}{2}-\frac{\varphi_1}{4}\right)$ where $\lambda = \sqrt{2\theta_2}$ and $\eta = \sqrt{2\theta_2+2\varphi_2}$. Taking the partial derivative with respect to θ_1 and evaluating it at $\theta_1 = 0$, we have,

$$\frac{\partial M}{\partial\theta_1}\mid_{\theta_1=0} = -\frac{1}{2}\exp\left(-\frac{\varphi_1}{4}\right)H_0^{-1/2} - \frac{1}{2}\exp\left(-\frac{\varphi_1}{4}\right)H_0^{-3/2}\left(\frac{\partial H}{\partial\theta_1}\mid_{\theta_1=0}\right).$$

Then taking the second derivative with respect to φ_1 and evaluating it at $\varphi_1 = 0$, we obtain

$$\begin{aligned}
&\left[\frac{\partial(\frac{\partial M}{\partial\theta_1}\mid_{\theta_1=0})}{\partial\varphi_1}\mid_{\varphi_1=0}\right] \\
=\ &\frac{1}{8}H_{00}^{-1/2} + H_{00}^{-3/2}\left\{\frac{1}{4}\left(\frac{\partial H_0}{\partial\varphi_1}\mid_{\varphi_1=0}\right) + \frac{1}{8}\left(\frac{\partial H}{\partial\theta_1}\mid_{\theta_1=0,\varphi_1=0}\right) - \frac{1}{2}\left[\frac{\partial\left(\frac{\partial H}{\partial\theta_1}\mid_{\theta_1=0}\right)}{\partial\varphi_1}\mid_{\varphi_1=0}\right]\right\} \\
&+\frac{3}{4}H_{00}^{-5/2}\left(\frac{\partial H_0}{\partial\varphi_1}\mid_{\varphi_1=0}\right)\left(\frac{\partial H}{\partial\theta_1}\mid_{\theta_1=0,\varphi_1=0}\right),
\end{aligned}$$

from which we require the following four expressions

$$H_{00} = H \mid_{\theta_1=0,\varphi_1=0} = \cosh\left(\frac{\lambda}{2}\right)\cosh\left(\frac{\eta}{2}\right) + \frac{\lambda}{\eta}\sinh\left(\frac{\eta}{2}\right)\sinh\left(\frac{\lambda}{2}\right);$$

$$\frac{\partial H_0}{\partial\varphi_1} \mid_{\varphi_1=0} = \frac{-1}{\eta}\cosh\left(\frac{\lambda}{2}\right)\sinh\left(\frac{\eta}{2}\right);$$

$$\frac{\partial H}{\partial\theta_1} \mid_{\theta_1=0,\varphi_1=0} = -\frac{1}{\lambda}\sinh\left(\frac{\lambda}{2}\right)\cosh\left(\frac{\eta}{2}\right) - \frac{1}{\eta}\sinh\left(\frac{\eta}{2}\right)\cosh\left(\frac{\lambda}{2}\right);$$

$$\frac{\partial(\frac{\partial H}{\partial\theta_1} \mid_{\theta_1=0})}{\partial\varphi_1} \mid_{\varphi_1=0} = \frac{1}{\lambda\eta}\sinh\left(\frac{\lambda}{2}\right)\sinh\left(\frac{\eta}{2}\right).$$

Finally, we have the following expression that is amenable for numerical integrations.

$$\begin{aligned}&\left[\frac{\partial(\frac{\partial M}{\partial\theta_1} \mid_{\theta_1=0})}{\partial\varphi_1} \mid_{\varphi_1=0}\right]\\ = \; & \frac{1}{8}H_{00}^{-1/2} + H_{00}^{-3/2}\left\{-\frac{1}{8\lambda}\sinh\left(\frac{\lambda}{2}\right)\cosh\left(\frac{\eta}{2}\right) - \frac{3}{8\eta}\sinh\left(\frac{\eta}{2}\right)\cosh\left(\frac{\lambda}{2}\right) + \frac{1}{2\lambda\eta}\sinh\left(\frac{\lambda}{2}\right)\sinh\left(\frac{\eta}{2}\right)\right\}\\ & + \frac{3}{4}H_{00}^{-5/2}\left[\frac{-1}{\eta}\sinh\left(\frac{\eta}{2}\right)\cosh\left(\frac{\lambda}{2}\right)\right]\left[-\frac{1}{\lambda}\sinh\left(\frac{\lambda}{2}\right)\cosh\left(\frac{\eta}{2}\right) - \frac{1}{\eta}\sinh\left(\frac{\eta}{2}\right)\cosh\left(\frac{\lambda}{2}\right)\right].\end{aligned}$$

By the same argument, we have the MGFs of $M_{0,1,1/2,1}(\theta_1,\theta_2,\varphi_1,\varphi_2)$ and $M_{0,1/2,1/2,1}(\theta_1,\theta_2,\varphi_1,\varphi_2)$. To distinguish λ, η and π in the three cases, we use subscripts.

To sum up:

The MGF $M_{0,1,0,1/2}(\theta_1,\theta_2,\varphi_1,\varphi_2)$ **is given by**

$$\left\{\left[\cosh\left(\frac{\lambda_{0,1,0,1/2}}{2}\right) - \frac{\theta_1}{\lambda_{0,1,0,1/2}}\sinh\left(\frac{\lambda_{0,1,0,1/2}}{2}\right)\right]\left[\cosh\left(\frac{\eta_{0,1,0,1/2}}{2}\right) - \frac{\pi_{0,1,0,1/2}}{\eta_{0,1,0,1/2}}\sinh\left(\frac{\eta}{2}\right)\right]\right\}^{-1/2}$$
$$\exp\left(-\frac{\theta_1}{2} - \frac{\varphi_1}{4}\right),$$

where $\lambda_{0,1,0,1/2} = \sqrt{-2\theta_2}$, $\eta_{0,1,0,1/2} = \sqrt{-2\theta_2 - 2\varphi_2}$, and $\pi_{0,1,0,1/2} = \left(\theta_1 - \lambda_{0,1,0,1/2}\right)\zeta_{0,1,0,1/2} + \varphi_1 + \lambda_{0,1,0,1/2}$ with $\zeta_{0,1,0,1/2} = \exp\left(\lambda_{0,1,0,1/2}\right)[1 - \left(\theta_1 - \lambda_{0,1,0,1/2}\right)\varpi^2_{0,1,0,1/2}]^{-1}$, $\varpi^2_{0,1,0,1/2} = \frac{\exp\left(\lambda_{0,1,0,1/2}\right)-1}{2\lambda_{0,1,0,1/2}}$, $\tau_{0,1,0,1/2} = \frac{\theta_1-\lambda_{0,1,0,1/2}}{2}\zeta_{0,1,0,1/2} + \frac{\varphi_1}{2} + \frac{\lambda_{0,1,0,1/2}-\eta_{0,1,0,1/2}}{2}$, and $\varpi^2_{z,0,1,0,1/2} = \frac{\exp\left(\eta_{0,1,0,1/2}\right)-1}{2\eta_{0,1,0,1/2}}$.

The MGF $M_{0,1,1/2,1}(\theta_1,\theta_2,\varphi_1,\varphi_2)$ **is given by**

$$\left\{\left[\cosh\left(\frac{\lambda_{0,1,1/2,1}}{2}\right) - \frac{\theta_1+\varphi_1}{\lambda_{0,1,1/2,1}}\sinh\left(\frac{\lambda_{0,1,1/2,1}}{2}\right)\right]\left[\cosh\left(\frac{\eta_{0,1,1/2,1}}{2}\right) - \frac{\pi_{0,1,1/2,1}}{\eta_{0,1,1/2,1}}\sinh\left(\frac{\eta_{0,1,1/2,1}}{2}\right)\right]\right\}^{-1/2}$$
$$\exp\left(-\frac{\theta_1}{2} - \frac{\varphi_1}{4}\right),$$

where $\lambda_{0,1,1/2,1} = \sqrt{-2\theta_2 - 2\varphi_2}$, $\eta_{0,1,1/2,1} = \sqrt{-2\theta_2}$ and $\pi_{0,1,1/2,1} = \left(\theta_1 + \varphi_1 - \lambda_{0,1,1/2,1}\right)\zeta_{0,1,1/2,1} - \varphi_1 + \lambda_{0,1,1/2,1}$ with $\zeta_{0,1,1/2,1} = \left[1 - \left(\theta_1 + \varphi_1 - \lambda_{0,1,1/2,1}\right)\varpi^2_{0,1,1/2,1}\right]^{-1} \exp\left(\lambda_{0,1,1/2,1}\right)$, $\varpi^2_{0,1,1/2,1} = \frac{\exp\left(\lambda_{0,1,1/2,1}\right)-1}{2\lambda_{0,1,1/2,1}}$, $\tau_{0,1,1/2,1} = \frac{\theta_1+\varphi_1-\lambda_{0,1,1/2,1}}{2}\zeta_{0,1,1/2,1} - \frac{\varphi_1}{2} + \frac{\lambda_{0,1,1/2,1} - \eta_{0,1,1/2,1}}{2}$, and $\varpi^2_{z,0,1,1/2,1} = \frac{\exp\left(\eta_{0,1,1/2,1}\right)-1}{2\eta_{0,1,1/2,1}}$.

The MGF $M_{0,1/2,1/2,1}(\theta_1,\theta_2,\varphi_1,\varphi_2)$ is given by

$$\left\{\left[\cosh\left(\frac{\lambda_{0,1/2,1/2,1}}{2}\right) - \frac{\varphi_1}{\lambda_{0,1/2,1/2,1}}\sinh\left(\frac{\lambda_{0,1/2,1/2,1}}{2}\right)\right]\left[\cosh\left(\frac{\eta_{0,1/2,1/2,1}}{2}\right) - \frac{\pi_{0,1/2,1/2,1}}{\eta_{0,1/2,1/2,1}}\sinh\left(\frac{\eta_{0,1/2,1/2,1}}{2}\right)\right]\right\}^{-1/2}$$
$$\exp\left(-\frac{\theta_1}{4} - \frac{\varphi_1}{4}\right),$$

where $\varpi^2 = \frac{\exp\left(\lambda_{0,1/2,1/2,1}\right)-1}{2\lambda_{0,1/2,1/2,1}}$, $\lambda_{0,1/2,1/2,1} = \sqrt{-2\varphi_2}$, $\eta_{0,1/2,1/2,1} = \sqrt{-2\theta_2}$, and $\pi_{0,1/2,1/2,1} = (\varphi_1 - \lambda_{0,1/2,1/2,1})\zeta_{0,1/2,1/2,1} + \theta_1 - \varphi_1 + \lambda_{0,1/2,1/2,1}$ with $\zeta_{0,1/2,1/2,1} = \left[1 - \left(\theta_1 - \lambda_{0,1/2,1/2,1}\right)\varpi^2_{0,1/2,1/2,1}\right]^{-1}\exp(\lambda_{0,1/2,1/2,1})$, $\tau_{0,1/2,1/2,1} = \frac{\varphi_1 - \lambda_{0,1/2,1/2,1}}{2}\zeta_{0,1/2,1/2,1} + \frac{\theta_1-\varphi_1}{2} + \frac{\lambda_{0,1/2,1/2,1}-\eta_{0,1/2,1/2,1}}{2}$, and $\varpi^2_z = \frac{\exp(\eta_{0,1/2,1/2,1})-1}{2\eta_{0,1/2,1/2,1}}$.

The computation of the second derivative of the MGF simply follows step 2.

For example, **the second derivative of MGF $M_{0,1,0,1/2}(\theta_1,-\theta_2,\varphi_1,-\varphi_2)$ is**

$$\left[\frac{\partial\left(\frac{\partial M_{0,1,0,1/2}(\theta_1,-\theta_2,\varphi_1,-\varphi_2)}{\partial\theta_1}\ |_{\theta_1=0}\right)}{\partial\varphi_1}\ |_{\varphi_1=0}\right]$$
$$= \frac{1}{8}H^{-1/2}_{00,0,1,0,1/2} + H^{-3/2}_{00,0,1,0,1/2}\{-\frac{1}{8\lambda_{0,1,0,1/2}}\sinh\left(\frac{\lambda_{0,1,0,1/2}}{2}\right)\cosh\left(\frac{\eta_{0,1,0,1/2}}{2}\right) -$$
$$\frac{3}{8\eta_{0,1,0,1/2}}\sinh\left(\frac{\eta_{0,1,0,1/2}}{2}\right)\cosh\left(\frac{\lambda_{0,1,0,1/2}}{2}\right) - \frac{1}{2\lambda_{0,1,0,1/2}\eta_{0,1,0,1/2}}\sinh\left(\frac{\lambda_{0,1,0,1/2}}{2}\right)\sinh\left(\frac{\eta_{0,1,0,1/2}}{2}\right)\}$$
$$+\frac{3}{4}H^{-5/2}_{00,0,1,0,1/2}\left[\frac{-1}{\eta_{0,1,0,1/2}}\sinh\left(\frac{\eta_{0,1,0,1/2}}{2}\right)\cosh\left(\frac{\lambda_{0,1,0,1/2}}{2}\right)\right]\times$$
$$\left[-\frac{1}{\lambda_{0,1,0,1/2}}\sinh\left(\frac{\lambda_{0,1,0,1/2}}{2}\right)\cosh\left(\frac{\eta_{0,1,0,1/2}}{2}\right) - \frac{1}{\eta_{0,1,0,1/2}}\sinh\left(\frac{\eta_{0,1,0,1/2}}{2}\right)\cosh\left(\frac{\lambda_{0,1,0,1/2}}{2}\right)\right].$$

and $H_{00,0,1,0,1/2} = \cosh\left(\frac{\lambda_{0,1,0,1/2}}{2}\right)\cosh\left(\frac{\eta_{0,1,0,1/2}}{2}\right) + \frac{\lambda_{0,1,0,1/2}}{\eta_{0,1,0,1/2}}\sinh\left(\frac{\eta_{0,1,0,1/2}}{2}\right)\sinh\left(\frac{\lambda_{0,1,0,1/2}}{2}\right)$ with $\lambda_{0,1,0,1/2} = \sqrt{2\theta_2}$ and $\eta_{0,1,0,1/2} = \sqrt{2\theta_2 + 2\varphi_2}$.

The second derivative of MGF $M_{0,1,1/2,1}(\theta_1, -\theta_2, \varphi_1, -\varphi_2)$ is

$$\left[\frac{\partial(\frac{\partial M_{0,1,1/2,1}(\theta_1,-\theta_2,\varphi_1,-\varphi_2)}{\partial\theta_1}|_{\theta_1=0})}{\partial\varphi_1}|_{\varphi_1=0}\right]$$

$$= \frac{1}{8}H^{-1/2}_{00,0,1,1/2,1} + H^{-3/2}_{00,0,1,1/2,1}\{-\frac{3}{8\lambda_{0,1,1/2,1}}\sinh\left(\frac{\lambda_{0,1,1/2,1}}{2}\right)\cosh\left(\frac{\eta_{0,1,1/2,1}}{2}\right)$$
$$-\frac{1}{8\eta_{0,1,1/2,1}}\sinh\left(\frac{\eta_{0,1,1/2,1}}{2}\right)\cosh\left(\frac{\lambda_{0,1,1/2,1}}{2}\right) + \frac{1}{2\lambda_{0,1,1/2,1}\eta_{0,1,1/2,1}}\sinh\left(\frac{\lambda_{0,1,1/2,1}}{2}\right)\sinh\left(\frac{\eta_{0,1,1/2,1}}{2}\right)\}$$
$$+\frac{3}{4}H^{-5/2}_{00,0,1,1/2,1}\left[\frac{-1}{\lambda_{0,1,1/2,1}}\sinh\left(\frac{\lambda_{0,1,1/2,1}}{2}\right)\cosh\left(\frac{\eta_{0,1,1/2,1}}{2}\right)\right]\times$$
$$\left[-\frac{1}{\lambda_{0,1,1/2,1}}\sinh\left(\frac{\lambda_{0,1,1/2,1}}{2}\right)\cosh\left(\frac{\eta_{0,1,1/2,1}}{2}\right) - \frac{1}{\eta_{0,1,1/2,1}}\sinh\left(\frac{\eta_{0,1,1/2,1}}{2}\right)\cosh\left(\frac{\lambda_{0,1,1/2,1}}{2}\right)\right],$$

and $H_{00,0,1,0,1/2} = \cosh\left(\frac{\lambda_{0,1,0,1/2}}{2}\right)\cosh\left(\frac{\eta_{0,1,0,1/2}}{2}\right) + \frac{\lambda_{0,1,0,1/2}}{\eta_{0,1,0,1/2}}\sinh\left(\frac{\eta_{0,1,0,1/2}}{2}\right)\sinh\left(\frac{\lambda_{0,1,0,1/2}}{2}\right)$ with $\lambda_{0,1,0,1/2} = \sqrt{2\theta_2 + 2\varphi_2}$ and $\eta_{0,1,0,1/2} = \sqrt{2\theta_2}$.

The second derivative of MGF $M_{0,1/2,1/2,1}(\theta_1, -\theta_2, \varphi_1, -\varphi_2)$ is

$$\left[\frac{\partial(M_{0,1/2,1/2,1}(\theta_1,-\theta_2,\varphi_1,-\varphi_2))}{\partial\varphi_1}|_{\varphi_1=0}\right]$$

$$= \frac{1}{16}H^{-1/2}_{00,0,1/2,1/2,1} + H^{-3/2}_{00,0,1/2,1/2,1}\{-\frac{1}{8\lambda_{0,1/2,1/2,1}}\sinh\left(\frac{\lambda_{0,1/2,1/2,1}}{2}\right)\cosh\left(\frac{\eta_{0,1/2,1/2,1}}{2}\right)$$
$$-\frac{1}{8\eta_{0,1/2,1/2,1}}\sinh\left(\frac{\eta_{0,1/2,1/2,1}}{2}\right)\cosh\left(\frac{\lambda_{0,1/2,1/2,1}}{2}\right) - \frac{1}{2\lambda_{0,1/2,1/2,1}\eta_{0,1/2,1/2,1}}\sinh\left(\frac{\lambda_{0,1/2,1/2,1}}{2}\right)\sinh\left(\frac{\eta_{0,1/2,1/2,1}}{2}\right)\}$$
$$+\frac{3}{4}H^{-5/2}_{00,0,1/2,1/2,1}\left[\frac{-1}{\lambda_{0,1/2,1/2,1}}\sinh\left(\frac{\lambda_{0,1/2,1/2,1}}{2}\right)\cosh\left(\frac{\eta_{0,1/2,1/2,1}}{2}\right)\right]\times$$
$$\left[-\frac{1}{\eta_{0,1/2,1/2,1}}\sinh\left(\frac{\eta_{0,1/2,1/2,1}}{2}\right)\cosh\left(\frac{\lambda_{0,1/2,1/2,1}}{2}\right)\right],$$

and $H_{00,0,1/2,1/2,1} = \cosh\left(\frac{\lambda_{0,1/2,1/2,1}}{2}\right)\cosh\left(\frac{\eta_{0,1/2,1/2,1}}{2}\right) + \frac{\lambda_{0,1/2,1/2,1}}{\eta_{0,1/2,1/2,1}}\sinh\left(\frac{\eta_{0,1/2,1/2,1}}{2}\right)\sinh\left(\frac{\lambda_{0,1/2,1/2,1}}{2}\right)$ with $\lambda_{0,1/2,1/2,1} = \sqrt{2\varphi_2}$ and $\eta_{0,1/2,1/2,1} = \sqrt{2\theta_2}$.

D Proof of Theorem 2.9

Using the results from Propositions 2.1 and 2.6, and substituting the values in Equation (2.18), we have the stated result.

References

Chambers, M. J. 2010, Jackknife estimation and inference in stationary autoregressive models, Department of Economics, University of Essex, Working paper.

Chambers, M.J., and M. Kyriacou, 2010, Jackknife bias reduction in the presence of a unit root, Department of Economics, University of Essex, Working paper.

Chiquoine, B., and E. Hjalmarsson, 2009, Jackknifing stock return predictions, *Journal of Empirical Finance*, 16, 793-803,

Miller, R.G, 1974, The jackknife - A review, *Biometrika*, 61, 1-15.

Hahn, J. and W. K. Newey, 2004, Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72, 1295-1319.

Phillips, P.C.B., 1987a, Time Series Regression with a Unit Root, *Econometrica*, 55, 277-301.

Phillips, P.C.B., 1987b, Toward a unified asymptotic theory for autoregression, *Biometrika*, 74, 533-547.

Phillips, P.C.B., and T. Magdalinos, 2009, Unit root and cointegrating limit theory when initialization is in the infinite past, *Econometric Theory*, 25, 1682-1715.

Phillips, P.C.B., and J. Yu, 2005, Jackknifing bond option prices, *Review of Financial Studies*, 18, 707-742.

Quenouille, M. H., 1949, Approximate Tests of Correlation in Time-Series, *Journal of the Royal Statistical Society. Series B,* 11, 68-84.

Quenouille, M. H., 1956, Notes on bias in estimation, *Biometrika* 43, 353-360.

Sargan, J. D., 1976, Econometric Estimators and the Edgeworth Approximation, *Econometrica*, 44, 421–448.

Schucany, W. R., Gray, H. L., Owen, D. B., 1971, On bias reduction in estimation, *Journal of the American Statistical Association*, 66, 524-533.

Shenton L.R., and W. L. Johnson, 1965, Moments of a serial correlation coefficient, *Journal of the Royal Statistical Society*, 27, 308–320.

Tang, C. Y., and S. X. Chen., 2009. Parameter estimation and bias correction for diffusion processes. *Journal of Econometrics*. 149, 65-81

Tukey, J. W., 1958, Bias and confidence in not-quite large samples, *Annals of Mathematical Statistics*, 29, 614.

White, J. S., 1961, Asymptotic Expansions for the Mean and Variance of the Serial Correlation Coefficient, *Biometrika*, 48, 85-94.

Wu, C. F. J., 1986, Jackknife, bootstrap, and other resampling methods in regression analysis, *The Annals of Statistics*, 14, 1261-1295.

Yu, J., 2011, Bias in the Estimation of the Mean Reversion Parameter in Continuous Time Models, *Journal of Econometrics*, forthcoming.

Zhou, Q. and J. Yu, 2010, Asymptotic distributions of the least squares estimator for diffusion process, Singapore Management University, Working Paper;.