Optimal Jackknife Estimation of Local to Unit Root Models

Marian Stoykov

University of Essex

January 2017

Abstract

This paper considers the application of the variance minimising jackknife estimator developed by Chen and Yu (2015) to local to unit root models. The weights used to construct the estimator depend on the variances of the full and each of the sub-sample estimators and the covariances between them. Thus, the joint moment generating functions (MGF) between each of the estimators have been derived to compute the asymptotic moments. Simulation studies demonstrate the excellent bias reducing performance of the estimator in addition to a reduced variance. A drawback of this construction is the dependence of the weights on the true local to unit root parameter, which is typically unknown. As such, this paper proposes a two-step "optimal" jackknife estimator to overcome the issue. Simulations outcome is encouraging for the usage of the two-step estimator.

Keywords: Local to unit root, asymptotic moments, jackknife, bias reduction, variance reduction, moment generating function.

1 Introduction

The least squares estimator of autoregressive processes is consistent but suffers from a bias in finite samples, a result supported by both analytical and simulation studies. For small samples the bias can be substantial and thus it would be beneficial to apply a bias reducing procedure or utilise an estimator different from OLS. The problem is challenging due to the fact that stationary, non-stationary and explosive processes have different properties. As a result, finding a unifying method that works for all cases is not easy.

One method that could prove useful is the jackknife estimator which is due to Quenouille (1956), and Tukey (1958) shows how one can use it to construct non-parametric estimates of variance. The jackknife estimator utilises re-sampling techniques and there are many ways in which it can be constructed. Miller (1974a) provides an excellent overview on the main ideas and Chambers (2013) discusses the main ways in which one could re-sample the data.

The original work of Quenouille and Tukey has led to a subsequent surge of literature on the jackknife. In terms of regression analysis, Miller (1974b) showed that the jackknife estimator is asymptotically normally distributed for i.i.d[1] errors when applied to the linear regression model. In addition, Duncan (1978) studies the finite sample capabilities of the jackknife estimator to construct confidence intervals when applied to non-linear regression models. Further contributions to the literature were made by Shao and Wu (1987) who proved the asymptotic unbiasdeness and consistency of three variance jackknife estimators when applied to linear regression models in the presence of error heteroskedasticity. In terms of instrumental variable estimation, Angrist et al. (1999) studied the finite sample properties of a jackknife instrumental variable estimator (JIVE) and found evidence that its performance is superior compared to the two-stage least squares (TSLS) and limited information maximum likelihood (LIML) estimators when applied to data with more instruments than endogenous variables. On the other hand, Blomquist and Dahlberg (1999) find no evidence that JIVE performs better than TSLS in small samples when the instruments are weak. Furthermore, the estimator of Angrist et al. has been criticised by Davidson and McKinnon (2006) who find further evidence on the bad performance of the JIVE in the case of weak instruments. Applications of the jackknife estimator to panel data fixed effects has been considered by Hahn and Newey (2004), and Dhaene et al. (2006) extend further to dynamic panel data models. In terms of pure time series, Kunsch (1989) used both the bootstrap and the jackknife to estimate standard errors in stationary time series and Phillips and Yu (2005) (henceforth PY) demonstrated the gains that can be obtained by applying the jackknife in reducing bias in option pricing in countinuous time models. Chambers (2013) explores the properties of the jackknife estimator when applied to stationary autoregressive processes, Chambers and Kyriacou (2013) (henceforth CK13) study the properties of the estimator in non-stationary processes and Chambers and Kyriacou (2016) (henceforth CK16) consider the local to unit root case. Lastly, Chen and Yu (2015) (henceforth CY) propose a more efficient modified version of the CK13 estimator which applies to unit root processes, which we shall label as an “optimal jackknife estimator”. The focus in this paper will be to extend the literature by constructing an ”optimal jackknife” for local to unit root processes, which is a natural

[1]The abbreviation i.i.d stands for independently and identically distributed.

extension of the CY estimator.

Section 2 introduces the non-overlapping jackknife estimator which PY, Chambers, CK and CY consider and the main advantages and drawbacks of the estimator such as the fact that it produces an outstanding performance in terms of bias reduction but does that at the expense of higher variance. Appropriate procedures for constructing the weights are discussed as well as the problems associated with them. The main issue with the weights is that they depend on the analytical solution of the bias which unfortunately has different forms in the stationary, non-stationary and explosive cases (Le Bretton and Pham (1988), Phillips (2012) and Stoykov (2017)). A further problem applicable to local to unit root cases is that the weights depend on the true parameter generating the data, something one is trying to estimate in the first place.

To set the stage, section 3 introduces the model. To construct the CY weights in the local to unit root case one needs to derive the finite sample variances of the full-sample and sub-sample estimators, and the covariances between them. CY state that the finite sample moments are difficult to obtain and thus they propose to use their asymptotic counterparts. As such, the section explores the way in which one can obtain the asymptotic moments by use of moment generating functions (MGF). Since the solutions of the MGFs involve integrals which cannot be solved analytically, one needs to apply numerical methods for evaluation. A discussion is also provided on some interesting features on the numerical results of the moments as the properties of the estimator change in the stationary, non-stationary and explosive cases. More specifically, in the explosive side the full-sample estimator follows a Cauchy distribution asymptotically (White (1958) and Phillips and Magdalinos (2007)) which also feeds through to the sub-sample estimators. Thus, the moments in that case should be interpreted as pseudo moments as it is a known fact that the Cauchy distribution is absent of moments.

Subsequent to all of the preliminaries, section 4 explores the "optimal" local to unit root jackknife estimator and discusses its properties and performance. Simulations (10,000 replications with normally distributed innovations) are carried out to try and explore the features of the estimator in comparison with the OLS, CK and CY estimators. The outcome of the exercise provides evidence in favour of the newly constructed estimator.

The estimator performs well in simulations, however, it is impossible to apply in practice as one needs to know the true parameter *a priori* to construct the weights. To overcome the problem, section 5 discusses a two-step "optimal" jackknife estimator. The idea behind it is to get an estimate of the true parameter that is assumed to have generated the data and use it to construct the weights and then use those weights in the second stage. Another simulation exercise (10,000 replications with normally distributed innovations) is performed to show that there is still some gain in applying that procedure. Lastly, section 6 concludes and the appendix depicts all proofs of theorems.

For notational simplicity the following notation shall be used throughout the paper. The symbol $\Rightarrow$ denotes convergence in distribution. $W(r)$ denotes a Wiener process on $C[0,1]$, the space of continuous real-valued functions on the unit interval. $J_\gamma(r) = \int_0^r e^{(r-s)\gamma} dW(s)$ denotes the Ornstein-Uhlenbeck process which satisfies the following stochastic differential equation $dJ_\gamma(r) = \gamma J_\gamma(r)dr + dW(r)$ for some constant parameter γ. The functionals $J_\gamma(r)$ and $W(r)$ will be denoted by J and W, respectively, such that functionals of the form $\int_a^b J_\gamma(r)dJ_\gamma(r)$ will be denoted as $\int_a^b JdJ$. We also make use of the lag operator L which translates into $L^j y_t = y_{t-j}$.

2 Jackknife estimation of autoregressive time series

The OLS estimator in autoregressive time series has been found to be biased but consistent and this has been supported by simulation and theoretical studies. Figure 3.1 shows a smoothed graph of the simulated bias for different sample sizes. The value of the autoregressive coefficient has been plotted on the x-axis and the value of the bias on the y-axis. It can also be observed that as the sample size increases the bias diminishes. Furthermore, the bias increases as the absolute value of the autoregressive parameter approaches unity.

The original jackknife (also sometimes referred to as "delete 1" jackknife) works only with i.i.d data which makes it inapplicable in time series settings as removing observations heuristically would affect the correlation structure. One way to try to maintain the structure of the process is to use non-overlapping blocks of sub-samples, as demonstrated by PY. Consider the researcher is interested in the population parameter θ. Further, assume the researcher can utilise a full sample estimator which satisfies

$$E(\hat{\beta}_n) = \beta + \frac{a_1}{n} + \frac{a_2}{n^2} + O\left(n^{-3}\right), \tag{1}$$

which is the case for most maximum likelihood estimations. Consider splitting the entire sample of n observations into m sub-samples, each of length l, such that $n = m \times l$. Assume the sub-sample estimators $\hat{\theta}_j$ $(j = 1, ..., m)$ within each block satisfy

$$E(\hat{\beta}_j) = \beta + \frac{a_1}{l} + \frac{a_2}{l^2} + O\left(l^{-3}\right). \tag{2}$$

Then, the jackknife is constructed as

$$\hat{\theta}_J = w_1\hat{\theta} + w_2\frac{1}{m}\sum_{i=1}^{m}\hat{\theta}_j.$$

Theorem 1 from Chambers ensures that setting up the weights as $w_1 = m/(m-1)$ and $w_2 = -1/(m-1)$ completely removes the first order bias of the estimator in stationary autoregressive settings as long as the expansions in (1) and (2) exist. Chambers analyses the asymptotic properties of the estimator in the stationary AR(p) process[2]

$$y_t = \alpha + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \epsilon_t, \quad t = 1, ..., n,$$

where $\epsilon_t \sim i.i.d(0, \sigma^2)$ with finite fourth moment and the roots of the equation $\phi(z) = 0$, where $\phi(z) = 1 - \phi_1 z - ... - \phi_p z^p$, lie outside the unit root circle. Rewrite the model as $y_t = x_t'\beta + \epsilon_t$, where $x_t = (1, y_{t-1}, ..., y_{t-p})'$ and $\beta = (\alpha, \phi_1, ..., \phi_p)$, such that the OLS estimator of the vector β is given by

$$\hat{\beta}_n = \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} \sum_{t=1}^{n} x_t y_t = \beta + \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} \sum_{t=1}^{n} x_t \epsilon_t.$$

The estimator also has the following asymptotic distribution

$$\sqrt{n}\left(\hat{\beta}_n - \beta\right) \Rightarrow N\left(0, \sigma^2 Q^{-1}\right),$$

where

[2] Following Chambers' notation.

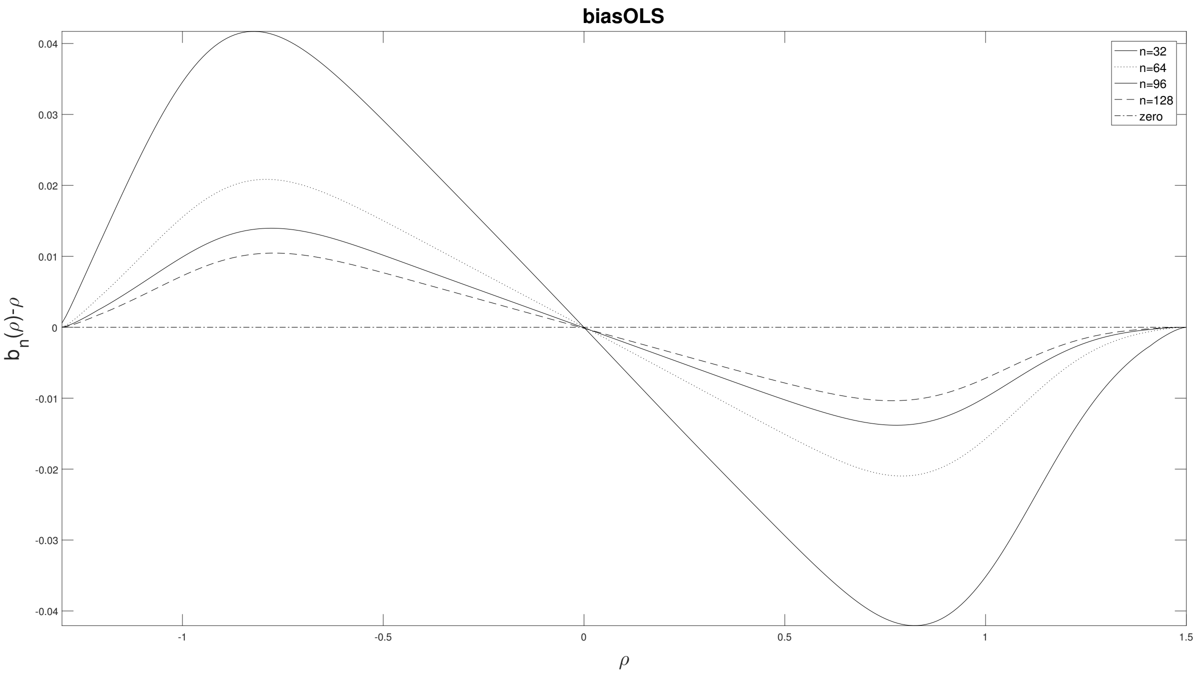

Figure 1: Bias of the OLS estimator for different sample sizes.

$$Q = E(x_t x_t') = \begin{bmatrix} 1 & \mu & \mu & \cdots & \mu \\ \mu & \gamma_0 + \mu^2 & \gamma_1 + \mu^2 & \cdots & \gamma_{p-1} + \mu^2 \\ \mu & \gamma_1 + \mu^2 & \gamma_0 + \mu^2 & \cdots & \gamma_{p-2} + \mu^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu & \gamma_{p-1} + \mu^2 & \gamma_{p-2} + \mu^2 & \cdots & \gamma_0 + \mu^2 \end{bmatrix},$$

with $\mu = E(y_t) = \alpha/\phi(1)$ and $\gamma_j = E(y_t y_{t-|j|})$. Chambers showed that the non-overlapping block jackknife has the same asymptotic distribution as the OLS, namely

$$\sqrt{n}\left(\hat{\beta}_J - \beta\right) \Rightarrow N\left(0, \sigma^2 Q^{-1}\right).$$

As such, the jackknife is consistent and asymptotically normally distributed. What is more, there is no asymptotic efficiency loss in comparison to OLS. Chambers also shows, by means of simulations, that the jackknife is robust to different specifications and also performs extremely well in comparison to the median-unbiased estimator and the bootstrap. However, there is the issue that the estimator has a higher variance in finite samples compared to its counterparts, as documented by Chambers.

The above mentioned weights work in stationary cases but not in unit root settings. CK13 argue that the constants in (2) are no longer identical across each of the sub-samples in the limit and the weights require modifications to take this into account. Consider

$$y_t = \beta y_{t-1} + \epsilon_t, \quad \epsilon_t \sim i.i.d(0, \sigma^2), \quad \beta = 1, \quad t = 1, ..., n,$$

where $y_0 = 0$. The constants from the $O(n^{-1})$ terms in (2) were derived by CK13 and are given by

$$\mu = E\left(\frac{\int_0^1 W dW}{\int_0^1 W^2}\right), \quad \mu_j = E\left(\frac{\int_{(j-1)/m}^{j/m} W dW}{\int_{(j-1)/m}^{j/m} W^2}\right), \quad j = 1, ..., m.$$

Then, Theorem 2 from CK13 defines the first-order asymptotic bias reducing jackknife as

$$\hat{\beta}_J^{opt} = \kappa^{opt}\hat{\beta}_n + \delta^{opt}\frac{1}{m}\sum_{i=1}^{m}\hat{\beta}_i$$

where $\kappa^{opt} = -\sum_{j=1}^{m}\mu_j/\bar{\mu}$, $\delta^{opt} = \mu/\bar{\mu}$ and $\bar{\mu} = \mu - \sum_{j=1}^{m}\mu_j$. Again, simulation results highlight the outstanding bias reduction performance with the expense of inflating the variance.

To try to mitigate the adverse effects CY propose to construct an optimisation problem, which aims to minimise the variance of the CK estimator. The new estimator is defined as[3]

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

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

subject to the constraints

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

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

The two constraints ensure that the first order term of the asymptotic bias is eliminated. To obtain the optimal weights one needs to calculate the means of the full and each of the sub-samples $(\mu, \mu_1, ..., \mu_m)$, their variances and the covariances between them, a topic which we shall return to in a subsequent section. By means of simulations, CY showed that the newly constructed estimator performs as well in terms of bias reduction and also has approximately 10% reduced variance in comparison to the CK13 estimator.

Theoretically, both estimators perform excellently, however applying them in practice becomes cumbersome. The weights, as they were derived, would be optimal only in a unit root case scenario, or having a value of the autoregressive parameter exactly equal to one. This would hardly be the case in any given empirical application. As such, CY analyse the performance of their estimator in local to unit root settings. However, their simulation study finds further evidence that the weights are no longer optimal the further away from unity the true parameter is. In this case, the CY estimator exhibits more distortions as it involves more terms that need to be calculated to construct the weights

[3]Following CY's notation.

(variances and covariances in addition to the means), which are theoretically incorrect once the parameter starts shifting away from unity.

Local to unit root processes also have the advantage that the asymptotic expansion satisfies

$$E(\hat{\beta}_n) = \beta + O\left(n^{-1}\right),$$

as $n \to \infty$ (Phillips (2012)), in contrast to just trying to think about stationary, nonstationary and explosive cases where the autocorrelation coefficients are considered fixed. In the latter case the OLS estimator has the following asymptotic expansion (due to Le Breton and Pham): as $n \to \infty$, for $|\beta| < 1$, $n(E(\hat{\beta}) - \beta)$ converges to -2β, for $|\beta| = 1$, $n(E(\hat{\beta}) - \beta)$ converges to -1.7814, for $|\beta| > 1$, $n^{-1/2}|\beta|^n(E(\hat{\beta}) - \beta)$ converges to $-2^{-1/2}\pi^{1/2}\beta^{-1}(\beta^2 - 1)^{3/2}$. As a result, constructing the weights in an ordinary fashion would be impossible, the reason being twofold: one does not know the true parameter *a priori* and, secondly, the function is discontinuous and unusable for autoregressive parameter values close to unity.

3 Optimal jackknife estimation in local to unit root models

Assume the variable of interest satisfies the following stochastic difference equation

$$y_t = \rho y_{t-1} + u_t, \qquad t = 1, \ldots, n, \tag{3}$$

where the process is initiated at any observable random variable bounded in probability $y_0 = O_p(1)$, $\rho = e^{\gamma/n} = 1 + \gamma/n + O(n^{-2})$ with γ being a constant and u_t is given by the following stationary process

$$u_t = \omega(L)\epsilon_t = \sum_{j=0}^{\infty} \omega_j \epsilon_{t-j},$$

where $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$, $E\epsilon_t^4 < \infty$, $\omega(z) = \sum_{j=0}^{\infty} \omega_j z^j$, $\omega_0 = 1$ and $\sum_{j=0}^{\infty} j|\omega_j| < \infty$.

No assumption will be made regarding the autoregressive coefficient in terms of whether it is smaller, equal or bigger to unity: if $\gamma < 0$ then the process is (locally) stationary, when $\gamma = 0$ the process is nonstationary and for $\gamma > 0$ the process is (locally) explosive. By construction, $\lim_{n\to\infty} \rho = 1$ and the model considered bares the name "local to unit root". The linear assumption on the error term allows for a great generality on the specification. One such specification could be that u_t is a stationary ARMA(p, q) process of the form $\phi(L)u_t = \theta(L)\epsilon_t$, where $\phi(z) = \sum_{j=0}^{p} \phi_j z^j$, $\theta(z) = \sum_{j=0}^{p} \theta_j z^j$ and all of the roots of the equation $\phi(z) = 0$ lie outside the unit circle. In this setting u_t satisfies the functional central limit theorem

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

on $C[0, 1]$, where $\sigma^2 = \sigma_\epsilon^2 \omega(1)^2$.

Consider splitting the entire sample of n observations into m sub-samples, each of length l such that $n = m \times l$. The OLS estimators for the full sample and each of the sub-samples are given by

$$\hat{\rho} = \frac{\sum_{t=1}^{n} y_t y_{t-1}}{\sum_{t=1}^{n} y_{t-1}^2} \text{ and } \hat{\rho}_j = \frac{\sum_{t \in \tau_j} y_t y_{t-1}}{\sum_{t \in \tau_j} y_{t-1}^2},$$

where $\tau_j = \{(j-1)l + 1, \ldots, jl\}$, for $j = 1, \ldots, m$. The limiting distributions of the full-sample and sub-sample estimators have been derived by Phillips (1987) and CK16, respectively, and are given by:

$$n(\hat{\rho} - \rho) \Rightarrow \frac{\int_0^1 JdW + \frac{1}{2}(1-\eta)}{\int_0^1 J^2},$$

$$l(\hat{\rho}_j - \rho) \Rightarrow \frac{\int_{(j-1)/m}^{j/m} JdW + \frac{1}{2m}(1-\eta)}{m \int_{(j-1)/m}^{j/m} J^2}.$$

where $\eta = \sigma_u^2/\sigma^2$ and $\sigma_u^2 = E(u_t^2) = \sigma_\epsilon^2 \sum_{j=0}^{\infty} \omega_j^2$. In this scenario, formulating the usual jackknife with weights given by $w_1 = m/(m-1)$ and $w_2 = -1/(m-1)$ fails to completely remove the first order bias as the different sub-samples have different limiting distributions, as observed by CK16. This issue is tackled by adjusting the second constraint. Construct the jackknife estimator as $\hat{\rho}_J = w\hat{\rho} - \sum_{j=1}^{m} w_j \hat{\rho}_j$ and minimise its variance subject to the constraints that ensure the first order bias term of the estimator is eliminated

$$\min_{w, \{w_j\}_{j=1}^m} Var(\hat{\rho}_J)$$

s.t.

$$w = \sum_{j=1}^{m} w_j + 1,$$

$$w\mu_\gamma = m \sum_{j=1}^{m} w_j \mu_{\gamma,j}.$$

The subscript 'γ' is used to distinguish from the case considered in *section* 2 where the constant γ was equal to zero. It is convenient to write the optimisation problem in a matrix form, which would also facilitate the simulations which are carried in a consequent section

$$\min_{\boldsymbol{w}} \boldsymbol{w}'\Omega\boldsymbol{w} \qquad \text{s.t.} \qquad \boldsymbol{\nu w} - 1 = 0 \text{ and } \boldsymbol{\mu w} = 0,$$

where $\boldsymbol{w} = [w, -w_1, -w_2, \ldots, -w_m]'$, $\boldsymbol{\nu} = [1, 1, \ldots, 1]$, a row vector of ones with $(m+1)$ elements, $\boldsymbol{\mu} = [\mu_\gamma, m\mu_{\gamma,1}, \ldots, m\mu_{\gamma,m}]$ and

$$\boldsymbol{\Omega} = \begin{bmatrix} VAR(\hat{\rho}) & COV(\hat{\rho}, \hat{\rho}_1) & \cdots & COV(\hat{\rho}, \hat{\rho}_m) \\ COV(\hat{\rho}, \hat{\rho}_1) & VAR(\hat{\rho}_1) & \cdots & COV(\hat{\rho}_1, \hat{\rho}_m) \\ \vdots & \vdots & \ddots & \vdots \\ COV(\hat{\rho}, \hat{\rho}_m) & COV(\hat{\rho}_1, \hat{\rho}_m) & \cdots & VAR(\hat{\rho}_m) \end{bmatrix}.$$

an $(m+1) \times (m+1)$ symmetric matrix due to $COV(\hat{\rho}, \hat{\rho}_1) = COV(\hat{\rho}_1, \hat{\rho})$. The corresponding Lagrangian is

$$L(\boldsymbol{w}; \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{w}'\Omega\boldsymbol{w} + \boldsymbol{\lambda}'(C\boldsymbol{w} - p),$$

where $\boldsymbol{\lambda} = [\lambda_1, \lambda_2]'$, $C = [\boldsymbol{\nu}'\boldsymbol{\mu}']'$ and $\boldsymbol{p} = [1, 0]'$. The 1/2 is introduced to facilitate the optimisation and does not change the optimising values $\boldsymbol{w}^*$. Taking the partial derivatives and setting them equal to zero yields

$$\frac{\partial L}{\partial \boldsymbol{w}} = \Omega\boldsymbol{w} + C'\boldsymbol{\lambda} = 0,$$
$$\frac{\partial L}{\partial \boldsymbol{\lambda}} = C\boldsymbol{w} - \boldsymbol{p} = 0.$$

The solution is given by

$$\begin{bmatrix} \boldsymbol{w}^* \\ \boldsymbol{\lambda}^* \end{bmatrix} = \begin{bmatrix} \Omega & C' \\ C & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \boldsymbol{p} \end{bmatrix}$$

Since Ω is positive definite and the constraints are linear, w^* produces a global minimum.

As a result of the construction of the minimisation problem one would need to calculate the variances of each of the estimators and the covariances between them. We will employ the procedure of CY and substitute the finite sample expressions with their asymptotic counterparts:

$$n^2 Var(\hat{\rho}) = E\left(\frac{\int_0^1 JdW}{\int_0^1 J^2}\right)^2 - \mu_\gamma^2 + O\left(n^{-1}\right),$$
$$l^2 Var(\hat{\rho}_j) = E\left(\frac{\int_{(j-1)/m}^{j/m} JdW}{m\int_{(j-1)/m}^{j/m} J^2}\right)^2 - \mu_{\gamma,j}^2 + O\left(n^{-1}\right), \quad j = 1, ..., m,$$
$$n^2 Cov(\hat{\rho}, \hat{\rho}_j) = E\left(\frac{\int_0^1 JdW}{\int_0^1 J^2}\frac{\int_{(j-1)/m}^{j/m} JdW}{\int_{(j-1)/m}^{j/m} J^2}\right) - m\mu_\gamma\mu_{\gamma,j} + O\left(n^{-1}\right), \quad 1 \leq j \leq m$$
$$n^2 Cov(\hat{\rho}_i, \hat{\rho}_j) = E\left(\frac{\int_{(i-1)/m}^{i/m} JdW}{\int_{(i-1)/m}^{i/m} J^2}\frac{\int_{(j-1)/m}^{j/m} JdW}{\int_{(j-1)/m}^{j/m} J^2}\right) - m^2\mu_{\gamma,i}\mu_{\gamma,j} + O\left(n^{-1}\right), \quad 1 \leq i < j \leq m$$

To get the required expectations we will follow the procedure of CK and CY and make use of the MGF. Let $N(a,b) = \int_a^b JdJ$ and $D(a,b) = \int_a^b J^2$ and let $M_{a,b}(\theta_1, \theta_2) = E\exp\left(\theta_1 \int_a^b JdJ + \theta_2 \int_a^b J^2\right)$ denote their joint MGF. Magnus (1986) showed that

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

and CK16 showed that

$$M_{a,b}(\theta_1, \theta_2) = \exp\left(-\frac{\theta_1 + \gamma}{2}(b-a)\right) H_{a,b}(\theta_1, \theta_2)^{-1/2}$$

where $H_{a,b}(\theta_1,\theta_2) = \cosh((b-a)\lambda) - \frac{1}{\lambda}[\theta_1 + \gamma + ((\theta_1+\gamma)^2 - \lambda^2)\nu^2]\sinh((b-a)\lambda)$, with $\lambda = \sqrt{\gamma^2 - 2\theta_2}$ and $\nu^2 = (e^{2a\gamma} - 1)/2\gamma$.

Proposition 1. *The second derivative of the MGF of $M_{a,b}(\theta_1,\theta_2)$ is given by*

$$\frac{\partial^2 M_{a,b}}{\partial\theta_1^2}\Big|_{\theta_1=0} = \exp\left(\frac{-\gamma(b-a)}{2}\right)\Bigg\{\frac{1}{4}(b-a)^2 H_0^{-1/2} + \frac{3(1+2\gamma\nu^2)^2}{4\lambda^2}H_0^{-5/2}\sinh^2((b-a)\lambda) + \\ + \frac{1}{2\lambda}H_0^{-3/2}[2\nu^2(1-(b-a)\gamma) - (b-a)]\sinh((b-a)\lambda)\Bigg\},$$

where $H_0 = \cosh((b-a)\lambda) - \frac{1}{\lambda}[\gamma + (\gamma^2 - \lambda^2)\nu^2]\sinh((b-a)\lambda)$, $\lambda = \sqrt{\gamma^2 - 2\theta_2}$, and $\nu^2 = (e^{2a\gamma} - 1)/2\gamma$.

The expression is amenable for numerical integration and can be used to derive the asymptotic variances. For the full-sample case $a = 0$ and $b = 1$, and for the sub-sample cases $a = (j-1)/m$ and $b = j/m$, for $j = 1, ..., m$. Note that the integrals from the asymptotic variances and covariances involve terms which are integrated with respect to a Brownian motion whereas the result from Proposition 3.1 involves terms which are integrated with respect to an O-U process. To overcome the problem we utilise the procedure of CK16. They showed that

$$N(a,b) = \int_{(j-1)/m}^{j/m} JdJ = \int_{(j-1)/m}^{j/m} JdW + \gamma\int_{(j-1/m)}^{j/m} J^2,$$

from which follows that

$$\int_{(j-1)/m}^{j/m} JdW = \int_{(j-1)/m}^{j/m} JdJ - \gamma\int_{(j-1)/m}^{j/m} J^2.$$

This also holds with $(j-1)/m = 0$ and $j/m = 1$. Plugging this into the asymptotic expression yields

$$\begin{aligned} n^2 Var(\hat{\rho}) &= E\left(\frac{\int_0^1 JdW}{\int_0^1 J^2}\right)^2 - \mu_\gamma^2 + O(n^{-1}) \\ &= E\left(\frac{\int_0^1 JdJ - \gamma\int_0^1 J^2}{\int_0^1 J^2}\right)^2 - \mu_\gamma^2 + O(n^{-1}) \\ &= E\left(\frac{\int_0^1 JdJ}{\int_0^1 J^2} - \gamma\right)^2 - \mu_\gamma^2 + O(n^{-1}) \\ &= E\left(\frac{\int_0^1 JdJ}{\int_0^1 J^2}\right)^2 - 2\gamma E\left(\frac{\int_0^1 JdJ}{\int_0^1 J^2}\right) + \gamma^2 - \mu_\gamma^2 + O(n^{-1}). \end{aligned}$$

The same procedure can be applied for each of the sub-sample terms

$$
\begin{aligned}
l^2 Var(\hat{\rho}_j) &= E\left(\frac{\int_{(j-1)/m}^{j/m} JdW}{m\int_{(j-1)/m}^{j/m} J^2}\right)^2 - \mu_{\gamma,j}^2 + O(n^{-1}) \\
&= E\left(\frac{\int_{(j-1)/m}^{j/m} JdJ - \gamma\int_{(j-1)/m}^{j/m} J^2}{m\int_{(j-1)/m}^{j/m} J^2}\right)^2 - \mu_{\gamma,j}^2 + O(n^{-1}) \\
&= E\left(\frac{\int_{(j-1)/m}^{j/m} JdJ}{m\int_{(j-1)/m}^{j/m} J^2} - \frac{\gamma}{m}\right)^2 - \mu_{\gamma,j}^2 + O(n^{-1}) \\
&= E\left(\frac{\int_{(j-1)/m}^{j/m} JdJ}{m\int_{(j-1)/m}^{j/m} J^2}\right)^2 - \frac{2\gamma}{m}E\left(\frac{\int_{(j-1)/m}^{j/m} JdJ}{m\int_{(j-1)/m}^{j/m} J^2}\right) + \frac{\gamma^2}{m^2} - \mu_{\gamma,j}^2 + O(n^{-1}),
\end{aligned}
$$

for $j = 1, ..., m$. The first terms on the right-hand side can be obtained from Proposition 1 and the second and fourth from CK16. The covariances can be obtain by utilising the same procedure.

Let $M_{a,b,c,d}(\theta_1, \theta_2, \varphi_1, \varphi_2) = \exp\left(\theta_1 \int_a^b JdJ + \theta_2 \int_a^b J^2 + \varphi_1 \int_c^d JdJ + \varphi_2 \int_c^d J^2\right)$ denote the MGF of $N(a,b)$, $N(c,d)$, $D(a,b)$ and $D(c,d)$ with $0 \le a < b \le 1$ and $0 \le c < d \le 1$. CY showed that

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

The following theorem derives the MGF of the four functionals for the two cases: for $a = 0$, $b = 1$ and $0 \le c < d \le 1$, which corresponds to the full sample - each sub-sample cases and for $0 \le a < b \le c < d \le 1$, which corresponds to the each sub-sample with each sub-sample case.

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

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

where $\zeta = \lambda = \sqrt{\gamma^2 - 2\theta_2}$, $\eta = \sqrt{\gamma^2 - 2\theta_2 - 2\varphi_2}$, $e = 1 - b$, $s = b - a$, $\varpi_b^2 = \frac{\exp(2e\lambda)-1}{2\lambda}$, $k_b = [1 - (\theta_1 + \gamma - \lambda)\varpi_b^2]^{-1}\exp(2e\lambda)$, $\varpi_a^2 = \frac{\exp(2s\eta)-1}{2\eta}$, $k_a = \left[1 - [(\theta_1 + \gamma - \lambda)k_b + \varphi_1 + (\lambda - \eta)]\varpi_a^2\right]^{-1}\exp(2s\eta)$, $\varpi^2 = \frac{\exp(2a\lambda)-1}{2\lambda}$ *and* $p = [(\theta_1 + \gamma - \lambda)k_b + \varphi_1 + (\lambda - \eta)]k_a - \varphi_1$.

For $0 \leq a < b \leq c < d \leq 1$*, the MGF* $M_{a,b,c,d}(\theta_1,\theta_2,\varphi_1,\varphi_2)$ *is given by*

$$M_{a,b,c,d}(\theta_1,\theta_2,\varphi_1,\varphi_2) =$$
$$\exp\left(-\frac{\theta_1}{2}s - \frac{\varphi_1}{2}e - \frac{\gamma}{2}d\right)\left[\cosh(a\gamma) - \frac{(p+\theta_1+\gamma-\eta)k_a - \theta_1 + \eta}{\gamma}\sinh(a\gamma)\right]^{-1/2}$$
$$\times\left[\cosh(q\gamma) - \frac{(\varphi_1+\gamma-\lambda)k_c - \varphi_1 + \lambda}{\gamma}\sinh(q\gamma)\right]^{-1/2}\left[\cosh(e\lambda) - \frac{\varphi_1+\gamma}{\lambda}\sinh(e\lambda)\right]^{-1/2}$$
$$\times\left[\cosh(s\eta) - \frac{(\varphi_1+\gamma-\lambda)(k_c-1)k_b + \theta_1 + \gamma}{\eta}\sinh(s\eta)\right]^{-1/2},$$

where $\delta = \zeta = \gamma$, $\lambda = \sqrt{\gamma^2 - 2\varphi_2}$, $\eta = \sqrt{\gamma^2 - 2\theta_2}$, $e = d - c$, $s = b - a$, $q = c - b$, $\varpi_c^2 = \frac{\exp(2e\lambda)-1}{2\lambda}$, $k_c = [1 - (\varphi_1+\gamma-\lambda)\varpi_c^2]^{-1}\exp(2e\lambda)$, $\varpi_b^2 = \frac{\exp(2(c-b)\gamma)-1}{2\gamma}$, $k_b = [1 - [(\varphi_1 + \gamma - \lambda)k_c - \varphi_1 + \lambda - \gamma]\varpi_b^2]^{-1}\exp(2q\gamma)$, $\varpi_a^2 = \frac{\exp(2s\eta)-1}{2\eta}$, $k_a = \left[1 - [(\varphi_1+\gamma-\lambda)(k_c-1)k_b + \theta_1 + (\gamma-\eta)]\varpi_a^2\right]^{-1}\exp(2s\eta)$, $\varpi^2 = \frac{\exp(2a\gamma)-1}{2\gamma}$ *and* $p = (\varphi_1+\gamma-\lambda)(k_c-1)k_b$.

The appendix contains the second derivative of both of the MGFs with respect to θ_1 and has not been included here as they are not essential for the discussion.

Again, as in the case with the variances, one needs to adjust the terms of the covariances. Starting with the full-sample with each sub-sample cases this can be achieved in the following way

$$\begin{aligned} n^2Cov(\hat{\rho},\hat{\rho}_j) &= E\left(\frac{\int_0^1 JdW}{\int_0^1 J^2}\frac{\int_{(j-1)/m}^{j/m} JdW}{\int_{(j-1)/m}^{j/m} J^2}\right) - m\mu_\gamma\mu_{\gamma,j} + O(n^{-1}) \\ &= E\left[\left(\frac{\int_0^1 JdJ}{\int_0^1 J^2} - \gamma\right)\left(\frac{\int_{(j-1)/m}^{j/m} JdJ}{\int_{(j-1)/m}^{j/m} J^2} - \gamma\right)\right] - m\mu_\gamma\mu_{\gamma,j} + O(n^{-1}) \\ &= E\left(\frac{\int_0^1 JdJ}{\int_0^1 J^2}\frac{\int_{(j-1)/m}^{j/m} JdJ}{\int_{(j-1)/m}^{j/m} J^2}\right) - \gamma E\left(\frac{\int_0^1 JdJ}{\int_0^1 J^2} + \frac{\int_{(j-1)/m}^{j/m} JdJ}{\int_{(j-1)/m}^{j/m} J^2}\right) + \\ &\quad + \gamma^2 - m\mu_\gamma\mu_{\gamma,j} + O(n^{-1}), \quad j = 1,...,m. \end{aligned}$$

The cases of each sub-sample with each sub-sample are straightforward:

$$\begin{aligned} n^2Cov(\hat{\rho}_i,\hat{\rho}_j) &= E\left(\frac{\int_{(i-1)/m}^{i/m} JdW}{\int_{(i-1)m}^{i/m} J^2}\frac{\int_{(j-1)/m}^{j/m} JdW}{\int_{(j-1)/m}^{j/m} J^2}\right) - m^2\mu_{\gamma,i}\mu_{\gamma,j} + O(n^{-1}) \\ &= E\left[\left(\frac{\int_{(i-1)m}^{i/m} JdJ}{\int_{(i-1)m}^{i/m} J^2} - \gamma\right)\left(\frac{\int_{(j-1)/m}^{j/m} JdJ}{\int_{(j-1)/m}^{j/m} J^2} - \gamma\right)\right] - m^2\mu_{\gamma,i}\mu_{\gamma,j} + O(n^{-1}) \\ &= E\left(\frac{\int_{(i-1)m}^{i/m} JdJ}{\int_{(i-1)m}^{i/m} J^2}\frac{\int_{(j-1)/m}^{j/m} JdJ}{\int_{(j-1)/m}^{j/m} J^2}\right) - \gamma E\left(\frac{\int_{(i-1)m}^{i/m} JdJ}{\int_{(i-1)m}^{i/m} J^2} + \frac{\int_{(j-1)/m}^{j/m} JdJ}{\int_{(j-1)/m}^{j/m} J^2}\right) + \\ &\quad + \gamma^2 - m^2\mu_{\gamma,i}\mu_{\gamma,j} + O(n^{-1}), \quad 1 \leq i < j \leq m. \end{aligned}$$

As in the asymptotic variances cases the first term can be obtained from Theorem 3 and the second and fourth terms from CK16.

Table 1 depicts the variance of the full-sample and each of the first six sub-sample estimators (the diagonal entries) and the covariances between them for different values of γ (the off diagonal entries). Although not included in the present paper, the patterns that are observed in the table follow for any number of sub-samples. It can be seen that the normalised variances increase as γ decreases. The variance of the full sample is bigger than the first's for negative γ, equal to it for $\gamma = 0$ and smaller than it for positive γ. This is a consequence of the solution of the MGF and the details can be found in CK16. Furthermore, the variances of each sub-sample are bigger than any of the consecutive sub-samples. In terms of the covariances, the covariance between the full-sample and each of the sub-samples increase as the number of the sub-sample used increase, i.e. $COV(\hat{\rho}, \hat{\rho}_1) < COV(\hat{\rho}, \hat{\rho}_2)$. This holds for all sub-samples with the exception of the last one. On the other hand, the pattern for sub-samples is the opposite, i.e $COV(\hat{\rho}_1, \hat{\rho}_2) > COV(\hat{\rho}_1, \hat{\rho}_3)$, etc. It should be noted that some of the entries in the table coincide with results from previous studies but most of them are new to the literature, namely Chen and Yu (2015) derived the variances of the full sample and sub-sample estimators for $\gamma = 0$ and Hansen (2014) derived the variances for the full-sample for negative values of γ.

Now that we have the asymptotic moments, we can construct the weights. Table 2 depicts those for different values of γ. Two things are worth noticing here. Firstly, as m increases the weight that is applied to the full sample estimator decreases for any γ. For example, when $\gamma = -10$ and $m = 100$, $w^* = 1.0371$. It seems that as $m \to \infty$, $w^* \to 1$. In other words the jackknife estimator converges to OLS as the number of sub-samples goes to infinity since the weights sum up to one. Thus the local to unit root optimal jackknife estimator shares the same feature as the standard jackknife analysed by Chambers (2013). Secondly, the estimator applies bigger weights to estimators which have lower variances. From Table 2 when $m = 6$, for $\gamma = -10$, $w^* = 1.2568$ and for $\gamma = 1$, $w^* = 1.5593$.

4 Simulation studies

This section has the aim to investigate the performance of the jackknife estimator in different settings via simulations. A comparison will be made between the "optimal" local to unit root jackknife and the following estimators: OLS, C_m, CY and CK16. The C_m estimator stands for Chambers' standard jackknife which utilises weights given by $w_1 = m/(m-1)$ and $w_2 = -1/(m-1)$, where each of the sub-samples has the same weight, namely w_2. The estimator is included for a couple of reasons. Firstly, it utilises optimal weights for the stationary case so it would be interesting to compare our estimator to C_m on the stationary side. Secondly, Kruse and Kaufmann (2013) find evidence that those weights perform best in terms of bias reduction for mildly explosive processes in the samples of smallest sizes when compared to different estimators: the bootstrap-aided estimator of Kim (2003), indirect inference and the approximately median-unbiased estimator of Roy and Fuller (2001). Thus, it would also be interesting to compare our estimator to C_m on the explosive side as well since, regardless of its excellent performance, it's theoretically suboptimal. Inclusion of the OLS estimator is self-explanatory. The CY jackknife is considered as it would be interesting to see how their weights perform for the full range of values for γ as the weighs are, by construction, suboptimal for values of γ different from 0. Comparison with the CK16 estimator is self-explanatory since the argument for the "optimal" jackknife estimator is that it performs as well as that of CK16 in terms of bias reduction but, in addition, also has a reduced variance. The comparison will be made in terms of bias and root mean-squared error (RMSE) since RMSE has the aim to capture the trade-off between bias reduction and an increase in the variance.

For the purpose of the simulation study, the replications are set at 1000 with the error term being drawn from a normal distribution with mean 0 and unit variance: $u_t \sim N(0,1)$. The number of observations considered start with $n = 24$ and are then doubled for each further scenario considered: $n \in \{24, 48, 96, 192\}$. The number of observations is chosen such that we can utilise a number of sub-samples. Furthermore, small samples are considered as the negative effects of the bias are most well-pronounced in those situations. The value of γ is taken such that it covers stationary, non-stationary and explosive processes.

4.1 Normal errors and zero initial condition

The model considered is the one given by (3) with normally distributed error terms and a zero initial condition. Table 3 depicts the results. When considering the bias it can be observed that the bias reduces as the sample size increases. Furthermore, it can also be seen that the bias of CK16 and the "optimal" jackknife considered in this paper is significantly lower in magnitude in comparison with OLS. For the range of values of γ considered the former two produce practically the same level of bias reduction. Secondly, the CY and C_m estimators are not always optimal. The former underperforms the further γ moves away from 0 whereas the latter underperforms on the explosive side (as discussed in previous sections). It is also interesting to note that CK16 and the "optimal" jackknife outperform C_m on the stationary side. This could be explained by the fact that they remove different first order terms and leave different second order etc. terms from the asymptotic bias. Note that the asymptotic bias on the stationary side is given by $E(\hat{\rho} - \rho) = -2\rho/n + O(n^{-2})$, which is different from the local to unit root's. Also, the "optimal" jackknife and CK16 also outperform C_m on the explosive side, which should not come as a surprise.

What is very encouraging in terms of the "optimal" estimator is Table 4 which contains a comparison between the RMSEs of the estimators. The subscripts denote the value of m. It can be observed that for more than half of the cases the estimator has the smallest RMSE. In particular, it always has a smaller RMSE in comparison with CK16, which is what the estimator is constructed to do. It should be noted that the entries in the table are for the values of the smallest RMSE across different values for m, which are not the same as the one which minimise the bias, typically for $m = 2$.

5 Two-step "optimal" jackknife estimation

The "optimal" jackknife estimator performs excellently in a controlled environment as the values of γ that are used to construct the weights are known. This, however, is not the case in applied situations as the idea of the estimator is to reduce the bias in the process of estimation of the autoregressive parameter. To try and overcome this problem we propose a two-step estimator. The idea of the estimator is to get an initial value of γ

in a first step, use it to construct the weights and then use those weights in the second step.

To try and back γ out we apply the following procedure. Firstly, run a regression and estimate $\hat{\rho}$ as an estimate of the true ρ. Since it is assumed that $\rho = e^{\gamma/n}$, we reverse and solve for γ, which is given by $\gamma = n \log \rho$. The estimator of γ would be given by $\hat{\gamma} = n \log \hat{\rho}$. The estimator chosen for the first step is OLS. In the second step we construct the weights with $\hat{\gamma}$ instead of the true γ. The last columns of Table 3 and 4 contain the performance of the estimator in comparison with the rest. We find two findings of importance. Firstly, for the bias minimising values of m the estimator performs extremely well and on some occasions it produces the smaller bias. Secondly, for the RMSE minimising values of m the estimator has a significantly lower bias in comparison to OLS and at the same time a smaller RMSE, which is extremely encouraging for the use of the estimator.

6 Conclusion

This paper has had the aim to construct an "optimal" jackknife estimator which attempts to overcome some of the problems with previous versions of the jackknife documented in the literature when autoregressive time series are considered. The "optimal" local to unit root jackknife estimator is constructed as a variance minimisation problem of the estimator considered by Chambers and Kyriacou (2016). The unifying method is applicable to stationary, non-stationary and explosive series. Simulation studies provide evidence that the newly constructed estimator performs outstandingly in terms of bias reduction and produces smaller variance than rival jackknife estimators for the bigger part of the autoregressive coefficient considered. As part of the construction of the estimator, one needs to calculate the variances of the full-sample and each of the sub-samples, and the covariances between all of them. As such, the paper also derives their asymptotic equivalents by means of moment generating functions and provides a discussion on some of their features. The paper also proposes a two-step estimator which tries to overcome the theoretical nature of constructing the weighs as they depend on the true parameter generating the data. Simulation studies show that the there is not much loss in the performance of the two-step estimator in comparison with the theoretical one. The two-step procedure is encouraging and could easily be utilised in applied frameworks. Future areas for exploration could include using the estimator for unit root testing.

Appendix

Proof of Proposition 1. Magnus (1986) showed that

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

CK showed that for $\gamma \neq 0$ the MGF of $\frac{\int_a^b J dJ}{\int_a^b J^2}$, where $0 \leq a < b \leq 1$, is given by

$$M_{\gamma;a,b}(\theta_1, \theta_2) = \exp\left(-\frac{\theta_1 + \gamma}{2}(b-a)\right) H_{\gamma;a,b}(\theta_1, \theta_2)^{-1/2}$$

where $\lambda = \sqrt{\gamma^2 - 2\theta_2}$, $\nu^2 = (e^{2a\gamma} - 1)/(2\gamma)$ and $H_{\gamma;a,b} = \cosh((b-a)\lambda) - (1/\lambda)[\theta_1 + \gamma + ((\theta_1+\gamma)^2 - \lambda^2)\nu^2]\sinh((b-a)\lambda)$. The first derivative is

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

Taking the second derivative of the MGF with respect to θ_1 and setting $\theta_1 = 0$ gives

$$\frac{\partial^2 M_{\gamma;a,b}}{\partial \theta_1^2}\Big|_{\theta_1=0} = \exp\left(\frac{-\gamma(b-a)}{2}\right)\left\{\frac{1}{4}(b-a)^2 H_0^{-1/2} + \frac{3}{4}H_0^{-5/2}\left[\frac{\partial H}{\partial \theta_1}\Big|_{\theta_1=0}\right]^2 + \right.$$
$$\left. + \frac{1}{2}H_0^{-3/2}\left[(b-a)\frac{\partial H}{\partial \theta}\Big|_{\theta_1=0} - \frac{\partial^2 H}{\partial \theta_1^2}\Big|_{\theta_1=0}\right]\right\}.$$

This requires the following three expressions

$$\begin{aligned}
H_0 &= \cosh((b-a)\lambda) - \frac{1}{\lambda}[\gamma + (\gamma^2 - \lambda^2)\nu^2]\sinh((b-a)\lambda) \\
\frac{\partial H}{\partial \theta_1}\Big|_{\theta_1=0} &= -\frac{1}{\lambda}[1 + 2\gamma\nu^2]\sinh((b-a)\lambda) \\
\frac{\partial^2 H}{\partial \theta_1^2}\Big|_{\theta_1=0} &= -\frac{2\nu^2}{\lambda}\sinh((b-a)\lambda).
\end{aligned}$$

Combining the results gives

$$\frac{\partial^2 M_{\gamma;a,b}}{\partial \theta_1}\Big|_{\theta_1=0} = \exp\left(\frac{-\gamma(b-a)}{2}\right)\left\{\frac{1}{4}(b-a)^2 H_0^{-1/2} + \frac{3(1+2\gamma\nu^2)^2}{4\lambda^2}H_0^{-5/2}\sinh^2((b-a)\lambda) + \right.$$
$$\left. + \frac{1}{2\lambda}H_0^{-3/2}[2\nu^2(1-(b-a)\gamma) - (b-a)]\sinh((b-a)\lambda)\right\},$$

an expression which is amenable for numerical integration. □

Proof of Theorem 1. Using the techniques of Chambers and Kyriacou (2012) and Chen and Yu (2011) one can derive the covariances by utilising the MGFs. Let $J(t)$ and $Y(t)$ ($t \in [0,1]$) be the O-U processes defined by

$$\begin{aligned}
dJ(t) &= \gamma J(t)dt + dW(t), \quad J(0) = 0 \\
dY(t) &= \lambda Y(t)dt + dW(t), \quad Y(0) = 0
\end{aligned}$$

Then by Girsanov's theorem $E(f(X)) = E(f(Y)\frac{d\mu_x}{d\mu_y}(s))$ where

$$\frac{d\mu_x}{d\mu_y}(s) = \exp\left\{(\gamma - \lambda)\int_0^1 s(t)dt - \frac{\gamma^2 - \lambda^2}{2}\int_0^1 s(t)^2 dt\right\}.$$

For $0 \le a < b \le 1$, let the MGF of $\frac{\int_0^1 JdJ}{\int_0^1 J^2}\frac{\int_a^b JdJ}{\int_a^b J^2}$ be $M_{0,1,a,b}(\theta_1, \theta_2, \varphi_1, \varphi_2)$ which is given by

$$M_{0,1,a,b}(\theta_1,\theta_2,\varphi_1,\varphi_2) = E\left[\exp\left\{\theta_1\int_0^1 JdJ + \theta_2\int_0^1 J^2 + \varphi_1\int_a^b JdJ + \varphi_2\int_a^b J^2\right\}\right].$$

Then by Girsanov's theorem

$$M_{0,1,a,b}(\theta_1,\theta_2,\varphi_1,\varphi_2)$$
$$= E\Bigg[\exp\Bigg\{\theta_1\int_0^1 YdY + \theta_2\int_0^1 Y^2 + \varphi_1\int_a^b YdY + \varphi_2\int_a^b Y^2 + (\gamma-\lambda)\int_0^1 YdY$$
$$- \frac{\gamma^2-\lambda^2}{2}\int_0^1 Y^2\Bigg\}\Bigg].$$

By the Ito Calculus $\int_a^b YdY = \frac{1}{2}\left[Y(b)^2 - Y(a)^2 - (b-a)\right]$, setting $\lambda = \sqrt{\gamma^2 - 2\theta_2}$ and denoting $s = b-a$ we have

$$M_{0,1,a,b}(\theta_1,\theta_2,\varphi_1,\varphi_2)$$
$$= \exp\left(-\frac{\theta_1}{2} - \frac{\varphi_1}{2}s - \frac{\gamma-\lambda}{2}\right) E\Bigg[\exp\Bigg\{\frac{(\theta_1+\gamma-\lambda)}{2}Y(1)^2 + \frac{\varphi_1}{2}[Y(b)^2 - Y(a)^2]$$
$$+ \varphi_2\int_a^b Y^2\Bigg\}\Bigg].$$

Now take the conditional expectation with respect to F_0^b, the sigma field generated by W on $[0,b]$

$$E[M_{0,1,a,b}(\theta_1,\theta_2,\varphi_1,\varphi_2)|F_0^b]$$
$$= \exp\left(-\frac{\theta_1}{2} - \frac{\varphi_1}{2}s - \frac{\gamma-\lambda}{2}\right)\exp\left(\frac{\varphi_1}{2}\left[Y(b)^2 - Y(a)^2\right] + \varphi_2\int_a^b Y^2\right)$$
$$\times E\exp\left(\frac{(\theta_1+\gamma-\lambda)}{2}Y(1)^2\Big| Y(b)\right).$$

Define $\mu_b = \exp(e\lambda)Y(b)$ and $\varpi_b^2 = (\exp(2e\lambda)-1)/(2\lambda)$, where $e = 1-b$, such that conditional on F_0^b, $Y(b) \sim N(\mu_b, \varpi_b^2)$. Then, by Lemma 5 of Magnus (1986)

$$E\exp\left(\frac{(\theta_1+\gamma-\lambda)}{2}Y(1)^2\Big| Y(b)\right) = [1-(\theta_1+\gamma-\lambda)\varpi_b^2]^{-1/2}\exp\left\{\left(\frac{\theta_1+\gamma-\lambda}{2}k_bY(b)^2\right)\right\},$$

where $k_b = [1-(\theta_1+\gamma-\lambda)\varpi_b^2]^{-1}\exp(2e\lambda)$. Thus,

$$E\left[M_{0,1,a,b}(\theta_1,\theta_2,\varphi_1,\varphi_2)|F_0^b\right]$$
$$= \left[1-(\theta_1+\gamma-\lambda)\varpi_b^2\right]^{-1/2}\exp\left(-\frac{\theta_1}{2} - \frac{\varphi_1}{2}s - \frac{\gamma-\lambda}{2}\right)$$
$$E\exp\left\{\left(\frac{\theta_1+\gamma-\lambda}{2}k_b + \frac{\varphi_1}{2}\right)Y(b)^2 - \frac{\varphi_1}{2}Y(a)^2 + \varphi_2\int_a^b Y^2\right\}.$$

Now, introduce another process on $[0,b]$ given by $dZ(t) = \eta Z(t)dt + dW(t),\ Z(0) = 0$ and apply Girsanov's theorem. The expectation of interest becomes

$$E\exp\Bigg\{\left(\frac{\theta_1+\gamma-\lambda}{2}k_b + \frac{\varphi_1}{2}\right)Z(b)^2 - \frac{\varphi_1}{2}Z(a)^2 + \varphi_2\int_a^b Z^2 + (\lambda-\eta)\int_0^b ZdZ$$
$$- \frac{\lambda^2-\eta^2}{2}\int_0^b Z^2\Bigg\}$$
$$= \exp\left(-\frac{\lambda-\eta}{2}b\right)E\exp\left\{\left(\frac{\theta_1+\gamma-\lambda}{2}k_b + \frac{\varphi_1}{2} + \frac{\lambda-\eta}{2}\right)Z(b)^2 - \frac{\varphi_1}{2}Z(a)^2 - \varphi_2\int_0^a Z^2\right\},$$

where $\eta = \sqrt{\gamma^2 - 2\theta_2 - 2\varphi_2}$. Now taking expectations w.r.t F_0^a yields

$$E\left[E\left(M_{0,1,a,b}(\theta_1,\theta_2,\varphi_1,\varphi_2)|F_0^b\right)|F_0^a\right]$$
$$= \left[1 - (\theta_1 + \gamma - \lambda)\varpi_b^2\right]^{-1/2} \exp\left(-\frac{\theta_1}{2} - \frac{\varphi_1}{2}s - \frac{\gamma - \lambda}{2} - \frac{\lambda - \eta}{2}b\right)$$
$$\times \exp\left(-\frac{\varphi_1}{2}Z(a)^2 - \varphi_2\int_0^a Z^2\right) E\exp\left\{\left(\frac{\theta_1 + \gamma - \lambda}{2}k_b + \frac{\varphi_1}{2} + \frac{\lambda - \eta}{2}\right)Z(b)^2\middle| Z(a)\right\}.$$

Define $\mu_a = \exp(s\eta)Z(a)$ and $\varpi_a^2 = (\exp(2s\eta) - 1)/(2\eta)$. Then, by Lemma 5 of Magnus

$$E\exp\left\{\left(\frac{\theta_1 + \gamma - \lambda}{2}k_b + \frac{\varphi_1}{2} + \frac{\lambda - \eta}{2}\right)Z(b)^2\middle| Z(a)\right\}$$
$$= \left[1 - [(\theta_1 + \gamma - \lambda)k_b + \varphi_1 + (\lambda - \eta)]\varpi_a^2\right]^{-1/2} \exp\left\{\left[\left(\frac{\theta_1 + \gamma - \lambda}{2}\right)k_b + \frac{\varphi_1}{2} + \frac{\lambda - \eta}{2}\right]k_a Z(a)^2\right\},$$

where $k_a = [1 - [(\theta_1 + \gamma - \lambda)k_b + \varphi_1 + (\lambda - \eta)]\varpi_a^2]^{-1}\exp(2s\eta)$. The MGF thus far is:

$$E\left[E\left(M_{0,1,a,b}(\theta_1,\theta_2,\varphi_1,\varphi_2)|F_0^b\right)|F_0^a\right]$$
$$= \left[1 - (\theta_1 + \gamma - \lambda)\varpi_b^2\right]^{-1/2}\left[1 - [(\theta_1 + \gamma - \lambda)k_b + \varphi_1 + (\lambda - \eta)]\varpi_a^2\right]^{-1/2}$$
$$\times \exp\left(-\frac{\theta_1}{2} - \frac{\varphi_1}{2}s - \frac{\gamma - \lambda}{2} - \frac{\lambda - \eta}{2}b\right)$$
$$\times E\exp\left\{\left[\left(\frac{\theta_1 + \gamma - \lambda}{2}\right)k_b + \frac{\varphi_1}{2} + \frac{\lambda - \eta}{2}\right]k_a Z(a)^2 - \varphi_2\int_0^a Z^2\right\}.$$

Now, introduce another process on $t \in [0, a]$ given by $dX(t) = \zeta X(t)dt + dW(t),\ X(0) = 0$. Applying Girsanov's theorem again to the expectation of interest yields

$$E\exp\left\{\left[\left(\frac{\theta_1 + \gamma - \lambda}{2}\right)k_b + \frac{\varphi_1}{2} + \frac{\lambda - \eta}{2}\right]k_a X(a)^2 - \varphi_2\int_0^a X^2 + (\eta - \zeta)\int_0^a XdX\right.$$
$$\left. - \frac{\eta^2 - \zeta^2}{2}\int_0^a X^2\right\}$$
$$= \exp\left(-\frac{\eta - \lambda}{2}a\right)E\exp\left\{\left[\left(\left(\frac{\theta_1 + \gamma - \lambda}{2}\right)k_b + \frac{\varphi_1}{2} + \frac{\lambda - \eta}{2}\right)k_a - \frac{\varphi_1}{2} + \frac{\eta - \lambda}{2}\right]X(a)^2\right\},$$

where $\zeta = \lambda$. Now, $X(a) \sim N(0, \varpi^2)$, where $\varpi^2 = (\exp(2a\lambda) - 1)/(2\lambda)$. Thus, the unconditional expectation is given by

$$E\exp\left\{\left[\left[\left(\frac{\theta_1 + \gamma - \lambda}{2}\right)k_b + \frac{\varphi_1}{2} + \frac{\lambda - \eta}{2}\right]k_a - \frac{\varphi_1}{2} + \frac{\eta - \lambda}{2}\right]X(a)^2\right\}$$
$$= \left[1 - [(\theta_1 + \gamma - \lambda)k_b + \varphi_1 + (\lambda - \eta)]k_a - \varphi_1 + (\eta - \lambda)]\varpi^2\right]^{-1/2}.$$

The MGF thus far is:

$$E\left[E\left(M_{0,1,a,b}(\theta_1,\theta_2,\varphi_1,\varphi_2)|F_0^b\right)|F_0^a\right] =$$
$$\left[1 - (\theta_1 + \gamma - \lambda)\varpi_b^2\right]^{-1/2}\left[1 - [(\theta_1 + \gamma - \lambda)k_b + \varphi_1 + (\lambda - \eta)]\varpi_a^2\right]^{-1/2}$$
$$\times \left[1 - [(\theta_1 + \gamma - \lambda)k_b + \varphi_1 + (\lambda - \eta)]\,k_a - \varphi_1 + (\eta - \lambda)]\,\varpi^2\right]^{-1/2}$$
$$\times \exp\left(-\frac{\theta_1}{2} - \frac{\varphi_1}{2}s - \frac{\gamma - \lambda}{2} - \frac{\lambda - \eta}{2}b - \frac{\eta - \lambda}{2}a\right).$$

Define $p = [(\theta_1 + \gamma - \lambda)k_b + \varphi_1 + (\lambda - \eta)]\, k_a - \varphi_1$, such that the MGF becomes

$$E\left[E\left(M_{0,1,a,b}(\theta_1, \theta_2, \varphi_1, \varphi_2)|F_0^b\right)|F_0^a\right] =$$
$$\left[\exp(-e\lambda)(1 - (\theta_1 + \gamma - \lambda)\varpi_b^2)\right]^{-1/2} \left[\exp(-s\eta)(1 - ((\theta_1 + \gamma - \lambda)k_b + \varphi_1 + (\lambda - \eta))\varpi_a^2)\right]^{-1/2}$$
$$\left[\exp(-a\lambda)(1 - (p + \eta - \lambda)\varpi^2)\right]^{-1/2} \exp\left(-\frac{\theta_1}{2} - \frac{\varphi_1}{2}s - \frac{\gamma}{2}\right).$$

Note that

$$\exp(-e\lambda)\left[1 - (\theta_1 + \gamma - \lambda)\varpi_b^2\right] = \exp(-e\lambda) - \left(\frac{\theta_1 + \gamma}{\lambda} - 1\right)\frac{\exp(e\lambda) - \exp(-e\lambda)}{2}$$
$$\frac{\exp(e\lambda) + \exp(-e\lambda)}{2} - \frac{(\theta_1 + \gamma)}{\lambda}\frac{(\exp(e\lambda) - \exp(-e\lambda))}{2} = \cosh(e\lambda) - \frac{\theta_1 + \gamma}{\lambda}\sinh(e\lambda).$$

In the same fashion

$$\begin{aligned} &\exp(-s\eta)\left(1 - ((\theta_1 + \gamma - \lambda)k_b + \varphi_1 + (\lambda - \eta))\varpi_a^2\right) \\ &= \cosh(s\eta) - \frac{(\theta_1 + \gamma - \lambda)k_b + \varphi_1 + \lambda}{\eta}\sinh(s\eta). \end{aligned}$$

Lastly,

$$\exp(-a\lambda)\left(1 - (p + \eta - \lambda)\varpi^2\right) = \cosh(a\lambda) - \frac{(p + \eta)}{\lambda}\sinh(a\lambda).$$

Thus

$$\begin{aligned} &M_{0,1,a,b}(\theta_1, \theta_2, \varphi_1, \varphi_2) = E\left[E\left(M_{0,1,a,b}(\theta_1, \theta_2, \varphi_1, \varphi_2)|F_0^b\right)|F_0^a\right] = \\ &\exp\left(-\frac{\theta_1}{2} - \frac{\varphi_1}{2}s - \frac{\gamma}{2}\right)\left[\cosh(a\lambda) - \frac{p + \eta}{\lambda}\cosh(a\lambda)\right]^{-1/2} \\ &\times \left[\cosh(e\lambda) - \frac{\theta_1 + \gamma}{\lambda}\sinh(e\lambda)\right]^{-1/2}\left[\cosh(s\eta) - \frac{(\theta_1 + \gamma - \lambda)k_b + \varphi_1 + \lambda}{\eta}\sinh(s\eta)\right]^{-1/2} \\ &= \exp\left(-\frac{\theta_1}{2} - \frac{\varphi_1}{2}s - \frac{\gamma}{2}\right)H_{0,1,a,b}(\theta_1, \theta_2, \varphi_1, \varphi_2)^{-1/2}, \end{aligned}$$

where

$$H_{0,1,a,b}(\theta_1, \theta_2, \varphi_1, \varphi_2) \equiv H = \left[\cosh(a\lambda) - \frac{p + \eta}{\lambda}\cosh(a\lambda)\right]\left[\cosh(e\lambda) - \frac{\theta_1 + \gamma}{\lambda}\sinh(e\lambda)\right]$$
$$\times \left[\cosh(s\eta) - \frac{(\theta_1 + \gamma - \lambda)k_b + \varphi_1 + \lambda}{\eta}\sinh(s\eta)\right],$$

$\zeta = \lambda = \sqrt{\gamma^2 - 2\theta_2}, \eta = \sqrt{\gamma^2 - 2\theta_2 - 2\varphi_2}, e = 1 - b, s = b - a, \varpi_b^2 = (\exp(2e\lambda) - 1)/(2\lambda),$
$k_b = \left[1 - (\theta_1 + \gamma - \lambda)\varpi_b^2\right]^{-1}\exp(2e\lambda), \varpi_a^2 = (\exp(2s\eta) - 1)/(2\eta),$
$k_a = \left[1 - [(\theta_1 + \gamma - \lambda)k_b + \varphi_1 + (\lambda - \eta)]\varpi_a^2\right]^{-1}\exp(2s\eta), \varpi^2 = (\exp(2a\lambda) - 1)/(2\lambda)$
$p = [(\theta_1 + \gamma - \lambda)k_b + \varphi_1 + (\lambda - \eta)]\, k_a - \varphi_1.$

By Lemma 2.1 of Chen and Yu (2015)

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

Taking the first partial derivative gives

$$\frac{\partial M}{\partial \theta_1}\Big|_{\theta_1=0} = \exp\left(-\frac{\gamma}{2}\right)\left\{-\frac{1}{2}\exp\left(-\frac{\varphi_1}{2}s\right)H_0^{-1/2} - \frac{1}{2}\exp\left(-\frac{\varphi_1}{2}s\right)H_0^{-3/2}\frac{\partial H}{\partial \theta_1}\Big|_{\theta_1=0}\right\}.$$

The second partial derivative is given by

$$\frac{\partial\left\{\frac{\partial M}{\partial \theta_1}\big|_{\theta_1=0}\right\}}{\partial \varphi_1}\Bigg|_{\varphi_1=0} = \exp\left(-\frac{\gamma}{2}\right)\left\{\frac{1}{4}sH_{00}^{-1/2} + \frac{3}{4}H_{00}^{-5/2}\left(\frac{\partial H_0}{\partial \varphi_1}\Big|_{\varphi_1=0}\right)\left(\frac{\partial H}{\partial \theta_1}\Big|_{\theta_1=0,\varphi_1=0}\right) + \right.$$

$$\left. + H_{00}^{-3/2}\left[\frac{1}{4}s\left(\frac{\partial H}{\partial \theta_1}\Big|_{\theta_1=0,\varphi_1=0}\right) + \frac{1}{4}\left(\frac{\partial H_0}{\partial \varphi_1}\Big|_{\varphi_1=0}\right) - \frac{1}{2}\left(\frac{\partial\left(\frac{\partial H}{\partial \theta_1}\big|_{\theta_1=0}\right)}{\partial \varphi_1}\Big|_{\varphi_1=0}\right)\right]\right\},$$

from which we need

$$\begin{aligned}
H_{00} &= \frac{1}{\lambda^2\eta}[(\pi\eta d - \Delta\lambda c)\lambda g - (\Delta\lambda d - \pi\eta c)\eta f],\\
\frac{\partial H_0}{\partial \varphi_1}\Big|_{\varphi_1=0} &= -\frac{1}{\lambda\eta}[\pi g + \Delta f]c,\\
\frac{\partial H}{\partial \theta_1}\Big|_{\theta_1=0,\varphi_1=0} &= -\frac{1}{\lambda^2\eta}[(\eta df + \lambda cg)\lambda b + (\lambda dg + \eta cf)\eta a], \text{and}\\
\frac{\partial\left(\frac{\partial H}{\partial \theta_1}\big|_{\theta_1=0}\right)}{\partial \varphi_1}\Big|_{\varphi_1=0} &= \frac{1}{\lambda\eta}[ag - bf]c,
\end{aligned}$$

where $\pi = \lambda b - \gamma a$, $\Delta = \gamma b - \lambda a$, $a = \sinh(e\lambda)$, $b = \cosh(e\lambda)$, $c = \sinh(s\eta)$, $d = \cosh(s\eta)$, $f = \sinh(a\lambda)$ and $g = \cosh(a\lambda)$.

For $0 \le a < b \le c < d \le 1$, $M_{a,b,c,d}(\theta_1, \theta_2, \varphi_1, \varphi_2)$ can be derived in the same fashion. Let $J(t)$ and $Y(t)$ ($t \in [0,1]$) be the O-U processes defined by

$$\begin{aligned}
dJ(t) &= \gamma J(t)dt + dW(t), \quad J(0) = 0\\
dY(t) &= \lambda Y(t)dt + dW(t), \;\; Y(0) = 0
\end{aligned}$$

For $0 \le a < b \le c < d \le 1$, let the MGF of $\frac{\int_a^b JdJ}{\int_a^b J^2}\frac{\int_c^d JdJ}{\int_c^d J^2}$ be $M_{a,b,c,d}(\theta_1, \theta_2, \varphi_1, \varphi_2)$ which is given by

$$\begin{aligned}
&M_{a,b,c,d}(\theta_1, \theta_2, \varphi_1, \varphi_2)\\
&= E\left[\exp\left\{\theta_1\int_a^b JdJ + \theta_2\int_a^b J^2 + \varphi_1\int_c^d JdJ + \varphi_2\int_c^d J^2\right\}\right]\\
&= E\left[\exp\left\{\theta_1\int_a^b YdY + \theta_2\int_a^b Y^2 + \varphi_1\int_c^d YdY + \varphi_2\int_c^d Y^2 + (\gamma - \lambda)\int_0^d YdY\right.\right.\\
&\left.\left. - \frac{\gamma^2 - \lambda^2}{2}\int_0^d Y^2\right\}\right],
\end{aligned}$$

in line of Girsanov's theorem. By using the Ito Calculus, setting $\lambda = \sqrt{\gamma^2 - 2\varphi_2}$ and denoting $s = b - a$, $e = d - c$ we have

$$M_{a,b,c,d}(\theta_1, \theta_2, \varphi_1, \varphi_2) = \exp\left(-\frac{\theta_1}{2}s - \frac{\varphi_1}{2}e - \frac{\gamma - \lambda}{2}d\right)$$

$$\times E\left[\exp\left\{\frac{\theta_1}{2}\left[Y(b)^2 - Y(a)^2\right] + \theta_2\int_a^b Y^2 + \frac{(\varphi_1 + \gamma - \lambda)}{2}Y(d)^2 - \frac{\varphi_1}{2}Y(c)^2 - \varphi_2\int_0^c Y^2\right\}\right].$$

Now take the conditional expectation with respect to F_0^c,

$$E\left[M_{0,1,a,b}(\theta_1,\theta_2,\varphi_1,\varphi_2)|F_0^c\right] = \exp\left(-\frac{\theta_1}{2}s - \frac{\varphi_1}{2}e - \frac{\gamma-\lambda}{2}d\right)$$
$$\times \exp\left(\frac{\theta_1}{2}\Big[Y(b)^2 - Y(a)^2\Big] + \theta_2\int_a^b Y^2 - \frac{\varphi_1}{2}Y(c)^2 - \varphi_2\int_0^c Y^2\right) E\exp\left(\frac{(\varphi_1+\gamma-\lambda)}{2}Y(d)^2\bigg|Y(c)\right).$$

Define $\mu_c = \exp(e\lambda)Y(c)$ and $\varpi_c^2 = (\exp(2e\lambda)-1)/(2\lambda)$, such that conditional on F_0^c, $Y(d) \sim N(\mu_c, \varpi_c^2)$. Then, by Lemma 5 of Magnus

$$E\exp\left(\frac{(\theta_1+\gamma-\lambda)}{2}Y(d)^2\bigg|Y(c)\right) = \left[1-(\varphi_1+\gamma-\lambda)\varpi_c^2\right]^{-1/2}\exp\left(\frac{\varphi_1+\gamma-\lambda}{2}k_cY(c)^2\right),$$

where $k_c = \left[1-(\varphi_1+\gamma-\lambda)\varpi_c^2\right]^{-1}\exp(2e\lambda)$. Thus,

$$\begin{aligned}
&E\left[M_{0,1,a,b}(\theta_1,\theta_2,\varphi_1,\varphi_2)|F_0^b\right]\\
&= \left[1-(\varphi_1+\gamma-\lambda)\varpi_c^2\right]^{-1/2}\exp\left(-\frac{\theta_1}{2}s - \frac{\varphi_1}{2}e - \frac{\gamma-\lambda}{2}d\right)\exp\left(\frac{\varphi_1+\gamma-\lambda}{2}k_cY(c)^2\right)\\
&\times \exp\left\{\frac{\theta_1}{2}\left[Y(b)^2-Y(a)^2\right] + \theta_2\int_a^b Y^2 + \left(\frac{\varphi_1+\gamma-\lambda}{2}k_c - \frac{\varphi_1}{2}\right)Y(c)^2 - \varphi_2\int_0^c Y^2\right\}.
\end{aligned}$$

Now, introduce another process on $[0,c]$ given by $dZ(t) = \delta Z(t)dt + dW(t)$, $Z(0)=0$ and apply Girsanov's theorem. The expectation of interest becomes

$$\begin{aligned}
&E\exp\left\{\frac{\theta_1}{2}\left[Z(b)^2-Z(a)^2\right] + \theta_2\int_a^b Z^2 + \left(\frac{\varphi_1+\gamma-\lambda}{2}k_c - \frac{\varphi_1}{2}\right)Z(c)^2 - \varphi_2\int_0^c Z^2\right.\\
&\left. + (\lambda-\delta)\int_0^c ZdZ - \frac{\lambda^2-\delta^2}{2}\int_0^c Z^2\right\}\\
&= \exp\left(-\frac{\lambda-\gamma}{2}c\right)E\exp\left\{\frac{\theta_1}{2}\left[Z(b)^2-Z(a)^2\right] + \theta_2\int_a^b Z^2 + \left(\frac{(\varphi_1+\gamma-\lambda)(k_c-1)}{2}\right)Z(c)^2\right\},
\end{aligned}$$

where $\delta = \gamma$. Now taking expectations w.r.t F_0^b yields

$$\begin{aligned}
&E\left[E\left(M_{a,b,c,d}(\theta_1,\theta_2,\varphi_1,\varphi_2)|F_0^c\right)|F_0^b\right]\\
&= \left[1-(\varphi_1+\gamma-\lambda)\varpi_c^2\right]^{-1/2}\exp\left(-\frac{\theta_1}{2}s - \frac{\varphi_1}{2}e - \frac{\gamma-\lambda}{2}d - \frac{\lambda-\gamma}{2}c\right)\\
&\exp\left(\frac{\theta_1}{2}\left[Z(b)^2-Z(a)^2\right] + \theta_2\int_a^b Z^2\right)E\exp\left\{\left(\frac{(\varphi_1+\gamma-\lambda)(k_c-1)}{2}\right)Z(c)^2\bigg|Z(b)\right\}.
\end{aligned}$$

Define $\mu_b = \exp(q\gamma)Z(b)$ and $\varpi_b^2 = (\exp(2q\gamma)-1)/(2\gamma)$, where $q = c-b$. Then, by Lemma 5 of Magnus

$$\begin{aligned}
&E\exp\left\{\left(\frac{(\varphi_1+\gamma-\lambda)(k_c-1)}{2}\right)Z(c)^2\bigg|Z(b)\right\}\\
&= \left[1-(\varphi_1+\gamma-\lambda)(k_c-1)\varpi_b^2\right]^{-1/2}\exp\left(\frac{(\varphi_1+\gamma-\lambda)(k_c-1)}{2}k_bZ(b)^2\right),
\end{aligned}$$

where $k_b = \left[1-(\varphi_1+\gamma-\lambda)(k_c-1)\varpi_b^2\right]^{-1}\exp(2q\gamma)$. The MGF thus far is

$$
\begin{aligned}
& E\left[E\left(M_{a,b,c,d}(\theta_1,\theta_2,\varphi_1,\varphi_2)|F_0^c\right)|F_0^b\right] \\
& = \left[1-(\varphi_1+\gamma-\lambda)\varpi_c^2\right]^{-1/2}\left[1-(\varphi_1+\gamma-\lambda)(k_c-1)\varpi_b^2\right]^{-1/2} \\
& \times \exp\left(-\frac{\theta_1}{2}s-\frac{\varphi_1}{2}e-\frac{\gamma-\lambda}{2}d-\frac{\lambda-\gamma}{2}c\right) \\
& \times \exp\left\{\left[\frac{(\varphi_1+\gamma-\lambda)(k_c-1)}{2}k_b+\frac{\theta_1}{2}\right]Z(b)^2-\frac{\theta_1}{2}Z(a)^2+\theta_2\int_a^b Z^2\right\}.
\end{aligned}
$$

Now, introduce another process on $t\in[0,b]$ given by $dX(t)=\eta X(t)dt+dW(t),\ X(0)=0$. Applying Girsanov's theorem again to the expectation of interest yields

$$
\begin{aligned}
& E\exp\left\{\left[\frac{(\varphi_1+\gamma-\lambda)(k_c-1)}{2}k_b+\frac{\theta_1}{2}\right]Z(b)^2-\frac{\theta_1}{2}Z(a)^2+\theta_2\int_a^b Z^2\right\} \\
& = E\exp\left\{\left[\left(\frac{(\varphi_1+\gamma-\lambda)(k_c-1)}{2}\right)k_b+\frac{\theta_1}{2}\right]X(b)^2-\frac{\theta}{2}X(a)^2+\theta_2\int_a^b X^2\right. \\
& \left.+(\gamma-\eta)\int_0^b XdX-\frac{\gamma^2-\eta^2}{2}\int_0^b X^2\right\} \\
& = \exp\left(-\frac{\gamma-\eta}{2}b\right)E\exp\left\{\left[\frac{(\varphi_1+\gamma-\lambda)(k_c-1)}{2}k_b+\frac{\theta_1}{2}+\frac{\gamma-\eta}{2}\right]X(b)^2-\frac{\theta_1}{2}X(a)^2\right. \\
& \left.-\theta_2\int_0^a X^2\right\},
\end{aligned}
$$

where $\eta=\sqrt{\gamma^2-2\theta_2}$. Now, take expectations with respect to F_0^a

$$
\begin{aligned}
& E\left\{E\left[E\left(M_{a,b,c,d}(\theta_1,\theta_2,\varphi_1,\varphi_2)|F_0^c\right)|F_0^b\right]|F_0^a\right\} \\
& = \left[1-(\varphi_1+\gamma-\lambda)\varpi_c^2\right]^{-1/2}\left[1-(\varphi_1+\gamma-\lambda)(k_c-1)\varpi_b^2\right]^{-1/2} \\
& \exp\left(-\frac{\theta_1}{2}s-\frac{\varphi_1}{2}e-\frac{\gamma-\lambda}{2}d-\frac{\lambda-\gamma}{2}c-\frac{\gamma-\eta}{2}b\right)\exp\left(-\frac{\theta_1}{2}X(a)^2-\theta_2\int_0^a X^2\right) \\
& E\exp\left\{\left(\frac{(\varphi_1+\gamma-\lambda)(k_c-1)}{2}k_b+\frac{\theta_1}{2}+\frac{\gamma-\eta}{2}\right)X(b)^2\middle|X(a)\right\}.
\end{aligned}
$$

Define $\mu_a=\exp(s\eta)X(a)$ and $\varpi_a=(\exp(2s\eta)-1)/(2\eta)$. Then by Lemma 5 of Magnus (1986)

$$
\begin{aligned}
& E\exp\left\{\left(\frac{(\varphi_1+\gamma-\lambda)(k_c-1)}{2}k_b+\frac{\theta_1}{2}+\frac{\gamma-\eta}{2}\right)X(b)^2\middle|X(a)\right\} \\
& = \left[1-[(\varphi_1+\gamma-\lambda)(k_c-1)k_b+\theta_1+(\gamma-\eta)]\varpi_a^2\right]^{-1/2} \\
& \exp\left(\frac{(\varphi_1+\gamma-\lambda)(k_c-1)k_b+\theta_1+(\gamma-\eta)}{2}k_aX(a)^2\right)
\end{aligned}
$$

where $k_a=[1-[(\varphi_1+\gamma-\lambda)(k_c-1)k_b+\theta_1+(\gamma-\eta)]\varpi_a^2]^{-1}\exp(2s\eta)$. Thus, the entire

MGF becomes

$$
\begin{aligned}
&E\left\{E\left[E\left(M_{a,b,c,d}(\theta_1,\theta_2,\varphi_1,\varphi_2)|F_0^c\right)|F_0^b\right]|F_0^a\right\}\\
&=\left[1-(\varphi_1+\gamma-\lambda)\varpi_c^2\right]^{-1/2}\left[1-(\varphi_1+\gamma-\lambda)(k_c-1)\varpi_b^2\right]^{-1/2}\\
&\left[1-[(\varphi_1+\gamma-\lambda)(k_c-1)k_b+\theta_1+(\gamma-\eta)]\varpi_a^2\right]^{-1/2}\\
&\exp\left(-\frac{\theta_1}{2}s-\frac{\varphi_1}{2}e-\frac{\gamma-\lambda}{2}d-\frac{\lambda-\gamma}{2}c-\frac{\gamma-\eta}{2}b\right)\\
&\exp\left\{\left[\frac{(\varphi_1+\gamma-\lambda)(k_c-1)k_b+\theta_1+(\gamma-\eta)}{2}k_a-\frac{\theta_1}{2}\right]X(a)^2-\theta_2\int_0^a X^2\right\}.
\end{aligned}
$$

Now introduce another process $dG(t)=\zeta G(t)dt+dW(t)$, $G(0)=0$ on $t\in[0,a]$. Applying Girsanov's theorem again to the expectation of interest yields

$$
\begin{aligned}
&E\exp\left\{\left[\frac{(\varphi_1+\gamma-\lambda)(k_c-1)k_b+\theta_1+(\gamma-\eta)}{2}k_a-\frac{\theta_1}{2}\right]X(a)^2-\theta_2\int_0^a X^2\right\}\\
&=E\exp\left\{\left[\frac{(\varphi_1+\gamma-\lambda)(k_c-1)k_b+\theta_1+(\gamma-\eta)}{2}k_a-\frac{\theta_1}{2}\right]G(a)^2-\theta_2\int_0^a G^2\right.\\
&\left.+(\eta-\zeta)\int_0^a GdG-\frac{\eta^2-\zeta^2}{2}\int_0^a G^2\right\}\\
&=\exp\left(-\frac{\eta-\gamma}{2}a\right)E\exp\left\{\left[\frac{(\varphi_1+\gamma-\lambda)(k_c-1)k_b+\theta_1+(\gamma-\eta)}{2}k_a-\frac{\theta_1}{2}+\frac{\eta-\gamma}{2}\right]G(a)^2\right\}.
\end{aligned}
$$

where $\zeta=\gamma$. Now $G(a)\sim N(0,\varpi^2)$, where $\varpi^2=(\exp(2a\gamma)-1)/(2\gamma)$. Thus,

$$
\begin{aligned}
&E\exp\left\{\left[\frac{(\varphi_1+\gamma-\lambda)(k_c-1)k_b+\theta_1+(\gamma-\eta)}{2}k_a-\frac{\theta_1}{2}+\frac{\eta-\gamma}{2}\right]G(a)^2\right\}\\
&=\left[1-\{[(\varphi_1+\gamma-\lambda)(k_c-1)k_b+\theta_1+(\gamma-\eta)]k_a-\theta_1+(\eta-\gamma)\}\varpi^2\right]^{-1/2}.
\end{aligned}
$$

The MGF becomes

$$
\begin{aligned}
&E\left\{E\left[E\left(M_{a,b,c,d}(\theta_1,\theta_2,\varphi_1,\varphi_2)|F_0^c\right)|F_0^b\right]|F_0^a\right\}\\
&=\left[1-(\varphi_1+\gamma-\lambda)\varpi_c^2\right]^{-1/2}\left[1-(\varphi_1+\gamma-\lambda)(k_c-1)\varpi_b^2\right]^{-1/2}\\
&\left[1-[(\varphi_1+\gamma-\lambda)(k_c-1)k_b+\theta_1+(\gamma-\eta)]\varpi_a^2\right]^{-1/2}\\
&\left[1-\{[(\varphi_1+\gamma-\lambda)(k_c-1)k_b+\theta_1+(\gamma-\eta)]k_a-\theta_1+(\eta-\gamma)\}\varpi^2\right]^{-1/2}\\
&\exp\left(-\frac{\theta_1}{2}s-\frac{\varphi_1}{2}e-\frac{\gamma}{2}(e+s)+\frac{\lambda}{2}e+\frac{\eta}{2}b\right).
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\exp(-e\lambda)\left[1-(\varphi_1+\gamma-\lambda)\varpi_c^2\right]=\exp(-e\lambda)-\left(\frac{\varphi_1+\gamma}{\lambda}-1\right)\frac{\exp(e\lambda)-\exp(-e\lambda)}{2}\\
&\frac{\exp(e\lambda)+\exp(-e\lambda)}{2}-\frac{(\theta_1+\gamma)}{\lambda}\frac{(\exp(e\lambda)-\exp(-e\lambda))}{2}=\cosh(e\lambda)-\frac{\theta_1+\gamma}{\lambda}\sinh(e\lambda).
\end{aligned}
$$

In the same fashion

$$\exp(-q\gamma)\left\{[(\varphi_1+\gamma-\lambda)k_c-\varphi_1+\lambda-\gamma]\varpi_b^2\right\}$$
$$=\cosh(q\gamma)-\frac{(\varphi_1+\gamma-\lambda)k_c-\varphi_1+\lambda}{\gamma}\sinh(q\gamma);$$
$$\exp(-s\eta)\left[1-((\varphi_1+\gamma-\lambda)k_b+\theta_1+(\lambda-\eta))\varpi_a^2\right]=\cosh(s\eta)-\frac{p+\theta_1+\gamma}{\eta}\sinh(s\eta),$$

where $p=(\varphi_1+\gamma-\lambda)(k_c-1)k_b$. Lastly,

$$\exp(-a\gamma)\left[1-[(\varphi_1+\gamma-\lambda)(k_c-1)k_bk_a+(\theta_1+\gamma-\eta)k_a-\theta_1+\eta-\gamma]\varpi^2\right]$$
$$=\cosh(a\gamma)-\frac{(p+\theta_1+\gamma-\eta)k_a-\theta_1+\eta}{\gamma}\sinh(a\gamma).$$

Thus, the MGF is given by

$$M_{a,b,c,d}(\theta_1,\theta_2,\varphi_1,\varphi_2)=E\left\{E\left[E\left(M_{a,b,c,d}(\theta_1,\theta_2,\varphi_1,\varphi_2)|F_0^c\right)|F_0^b\right]|F_0^a\right\}=$$
$$\exp\left(-\frac{\theta_1}{2}s-\frac{\varphi_1}{2}e-\frac{\gamma}{2}d\right)\left[\cosh(a\gamma)-\frac{(p+\theta_1+\gamma-\eta)k_a-\theta_1+\eta}{\gamma}\sinh(a\gamma)\right]^{-1/2}$$
$$\times\left[\cosh(s\eta)-\frac{p+\theta_1+\gamma}{\eta}\sinh(s\eta)\right]^{-1/2}\left[\cosh(e\lambda)-\frac{\varphi_1+\gamma}{\lambda}\sinh(e\lambda)\right]^{-1/2}$$
$$\times\left[\cosh(q\gamma)-\frac{(\varphi_1+\gamma-\lambda)k_c-\varphi_1+\lambda}{\gamma}\sinh(q\gamma)\right]^{-1/2}$$
$$=\exp\left(-\frac{\theta_1}{2}-\frac{\varphi_1}{2}s-\frac{\gamma}{2}d\right)H_{a,b,c,d}(\theta_1,\theta_2,\varphi_1,\varphi_2)^{-1/2},$$

where

$$H_{a,b,c,d}\equiv H=\left[\cosh(a\gamma)-\frac{(p+\theta_1+\gamma-\eta)k_a-\theta_1+\eta}{\gamma}\sinh(a\gamma)\right]$$
$$\left[\cosh(s\eta)-\frac{p+\theta_1+\gamma}{\eta}\sinh(s\eta)\right]\left[\cosh(e\lambda)-\frac{\varphi_1+\gamma}{\lambda}\sinh(e\lambda)\right]$$
$$\left[\cosh(q\gamma)-\frac{(\varphi_1+\gamma-\lambda)k_c-\varphi_1+\lambda}{\gamma}\sinh(q\gamma)\right],$$

$\delta=\zeta=\gamma$, $\lambda=\sqrt{\gamma^2-2\varphi_2}$, $\eta=\sqrt{\gamma^2-2\theta_2}$, $e=d-c$, $q=c-b$, $s=b-a$, $\varpi_c^2=\frac{\exp(2e\lambda)-1}{2\lambda}$, $k_c=[1-(\varphi_1+\gamma-\lambda)\varpi_c^2]^{-1}\exp(2e\lambda)$, $\varpi_b^2=\frac{\exp(2q\gamma)-1}{2\gamma}$, $k_b=[1-(\varphi_1+\gamma-\lambda)(k_c-1)\varpi_b^2]^{-1}\exp(2q\gamma)$, $\varpi_a^2=\frac{\exp(2s\eta)-1}{2\eta}$, $k_a=\left[1-(p+\theta_1+\gamma-\eta)\varpi_a^2\right]^{-1}\exp(2s\eta)$, $\varpi^2=\frac{\exp(2a\gamma)-1}{2\gamma}$, $p=[(\varphi_1+\gamma-\lambda)(k_c-1)k_b$.

By Lemma 2.1 of Chen and Yu (2015), taking the first partial derivative gives

$$\frac{\partial M}{\partial\theta_1}\Big|_{\theta_1=0}=\exp\left(-\frac{\gamma}{2}d\right)\left\{-\frac{1}{2}s\exp\left(-\frac{\varphi_1}{2}e\right)H_0^{-1/2}-\frac{1}{2}\exp\left(-\frac{\varphi_1}{2}e\right)H_0^{-3/2}\frac{\partial H}{\partial\theta_1}\Big|_{\theta_1=0}\right\}.$$

The second partial derivative is given by

$$\frac{\partial\left\{\frac{\partial M}{\partial\theta_1}|_{\theta_1=0}\right\}}{\partial\varphi_1}\Big|_{\varphi_1=0} = \exp\left(-\frac{\gamma}{2}d\right)\left\{\frac{1}{4}esH_{00}^{-1/2} + \frac{3}{4}H_{00}^{-5/2}\left(\frac{\partial H_0}{\partial\varphi_1}|_{\varphi_1=0}\right)\left(\frac{\partial H}{\partial\theta_1}|_{\theta_1=0,\varphi_1=0}\right)\right.$$
$$\left.+ H_{00}^{-3/2}\left[\frac{1}{4}e\left(\frac{\partial H}{\partial\theta_1}|_{\theta_1=0,\varphi_1=0}\right) + \frac{1}{4}s\left(\frac{\partial H_0}{\partial\varphi_1}|_{\varphi_1=0}\right) - \frac{1}{2}\left(\frac{\partial\left(\frac{\partial H}{\partial\theta_1}|_{\theta_1=0}\right)}{\partial\varphi_1}|_{\varphi_1=0}\right)\right]\right\},$$

from which we require

$$H_{00} = \frac{1}{\gamma^2\lambda\eta}[(\phi\eta d - \psi\gamma c)\gamma g - (\psi\gamma d - \phi\eta c)\eta f],$$
$$\frac{\partial H_0}{\partial\varphi_1}|_{\varphi_1=0} = -\frac{1}{\gamma\lambda\eta}[(\gamma c + \eta d)\gamma g + (\gamma d + \eta c)\eta f][v + h]a,$$
$$\frac{\partial H}{\partial\theta_1}|_{\theta_1=0,\varphi_1=0} = -\frac{1}{\gamma\lambda\eta}[\phi g + \psi f]c,$$
$$\frac{\partial\left\{\frac{\partial H}{\partial\theta_1}|_{\theta_1=0}\right\}}{\partial\varphi_1}\Big|_{\varphi_1=0} = \frac{1}{\lambda\eta}[g - f][v + h]ac,$$

where $\phi = \pi\gamma v - \Delta\lambda h$, $\psi = \Delta\lambda v - \pi\gamma h$, $f = \sinh(a\gamma)$, $g = \cosh(a\gamma)$, $h = \sinh(q\gamma)$ and $v = \cosh(q\gamma)$. □

References

[1] Angrist, J.D., Imbens, J.W., and Krueger, A.B. (1999) Jackknife instrumental variables estimation, *Journal of Applied Econometrics*, 14, 57-67.

[2] Bao, Y. (2007) The approximate moments of the least squares estimator for the stationary autoregressive model under a general error distribution, *Econometric Theory*, 23, 1013-1021.

[3] Blomquist, S., Dahlberg, M. (1999) Small sample properties of LIML and jackknife IV estimators: Experiments with weak instruments, *Journal of Applied Econometrics*, 14, 69-88.

[4] Chambers, M.J. (2013) Jackknife estimation of stationary autoregressive models, *Journal of Econometrics*, 172, 142-157.

[5] Chambers, M.J. and Kyriacou, M. (2013) Jackknife estimation with a unit root, *Statistics and Probability Letter*, 83, 1677-1682.

[6] Chambers, M.J. and Kyriacou, M. (2016) Jackknife estimation and inference in the presence of a near-unit root, *Unpublished manuscript.*

[7] Chen, y. and Yu, J. (2015) Optimal jackknife for unit root models, *Statistics and Probability Letters*, 99, 135-142.

[8] Davidson, R., MacKinnon, J.G. (2006) The case against JIVE, *Journal of Applied Econometrics*, 21, 827-833.

[9] Ducan, G.T, (1978) An empirical study of jackknife-constructed confidence regions in nonlinear regression, *Technometrics*, 20, 123-129.

[10] Efron, B. (1979) Bootstrap methods: another look at the jackknife, *Annals of Statistics*, 7, 1–26.

[11] Efron, B. (1982) The jackknife, the boostrap and other resampling plans, *Society for Industrial and Applied Mathematics.*

[12] Gourieroux, C., Renault, E. and Touzi, N. (2000) Calibration by simulation for small sample bias correction, *Simulation-Based Inference in Econometrics: Methods and Applications*, (ed. R. S. Mariano, T. Schuermann, and M. Weeks, eds.,), Cambridge, UK: Cambridge University Press, 2000, 328-58.

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

[14] Hamiltion, J.D. (1994) Time series analysis, Princeton University Press.

[15] Hansen, B.E. (2014) Asymptotic Moments of Autoregressive Estimates with a Near Unit Root and Minimax Risk, *Advances in Econometrics*, 33, 3-21.

[16] Hurwicz, L. (1950) Least squares bias in time series, *Statistical Inference in Dynamic Economic Models*, (ed. T. C. Koopmans), Wiley, New York.

[17] Kim, J. (2003) Forecasting autoregressive time series with bias-corrected parameter estimators, *International Journal of Forecasting*, 19, 493-502.

[18] Kruse, R., Kaufmann, Hendrik (2013) Bias-corrected estimation in mildly explosive autoregressions, *Conference Paper.*

[19] Künsch, H.R. (1989) The jackknife and the bootstrap for general stationary observations, *Annals of Statistics*, 17, 1217–1241.

[20] Le Breton, A., Pham, D.T. (1989) On the bias of the least squares estimator for the first order autoregressive process, *Ann. Inst. Statist. Math.*, 41, 555-563.

[21] MacKinnon, J. G., SmithA. A. (1988) Approximate bias correction in econometrics, *Journal of Econometrics*, 85, 205-230.

[22] Magnus, J. R. (1986) The exact moments of a ratio of quadratic forms in normal variables, *Annales d'Economie et de Statistique*, 4.

[23] Miller, R.G., (1974a) The jackknife-a review, *Biometrika* 61, 1-15.

[24] Miller, R.G., (1974b) An unbalanced jackknife, *Annals of Statistics*, 2, 880-91.

[25] Phillips, P.C.B. (1987) Time series regression with a unit root, *Econometrica*, 55, 277-301.

[26] Phillips, P.C.B. (2012) Folklore theorems, implicit maps and indirect inference, *Econometrica*, 80, 425-454.

[27] Phillips, P.C.B, Magdalinos, T. (2007) Limit theory for moderate deviations from a unit root, *Journal of Econometrics*, 136, 115-130.

[28] Phillips, P.C.B., Yu, J. (2005) Jackknifing bond option prices, *Review of Financial Studies*, 18, 707–742.

[29] Quenouille, M. H. (1956) Notes on bias in estimation, *Biometrika*, 43, 353–360.

[30] Ross, S. (2012) A first course in probability, Pearson, 9th Edition.

[31] Roy, A., Fuller, W. (2001) Estimation for autoregressive time series with a root near 1, *Journal of Business & Economic Statistics*, 19, 482-493.

[32] Shaman, P., Stine, R.A. (1988) The bias of autoregressive coefficient estimators, *Journal of the American Statistical Association*, 83, 842-848.

[33] Shao, J., Wu., C.F.J. (1987) Heteroskedasticity-robustness of jackknife variance estimators in Linear Models, *Annals of Statistics*, 15, 1563-1579.

[34] Stoykov, M.Z. (2017) Least squares bias in time series with moderate deviations from a unit root, *PhD Thesis*.

[35] Tanaka, K. (1996) Time series analysis: Nonstationary and noninvertible distribution theory, New York: John Wiley and Sons.

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

[37] White, J. (1958) The limiting distribution of the serial correlation coefficient in the explosive case, *Annals of Mathematical Statistics*, 29, 1188-1197.

Table 1

Values of asymptotic variances of full and sub-sample estimators and covariances between them.

	$n\hat{\rho}$	$l\hat{\rho}_1$	$l\hat{\rho}_2$	$l\hat{\rho}_3$	$l\hat{\rho}_4$	$l\hat{\rho}_5$	$l\hat{\rho}_6$
$\gamma = -10$							
$n\hat{\rho}$	29.1456	0.8339	0.9043	0.9115	0.9119	0.9136	0.9144
$l\hat{\rho}_1$	0.8339	12.9361	0.6458	0.0189	0.0007	0.0000	0.0000
$l\hat{\rho}_2$	0.9043	0.6458	9.6851	0.5407	0.0162	0.0006	0.0000
$l\hat{\rho}_3$	0.9115	0.0189	0.5407	9.6212	0.5385	0.0162	0.0006
$l\hat{\rho}_4$	0.9119	0.0007	0.0162	0.5385	9.6190	0.5384	0.0162
$l\hat{\rho}_5$	0.9136	0.0000	0.0006	0.0162	0.5384	9.6189	0.5384
$l\hat{\rho}_6$	0.9144	0.0000	0.0000	0.0006	0.0162	0.5384	9.6189
$\gamma = -1$							
$n\hat{\rho}$	11.7605	0.3355	0.4031	0.4551	0.4946	0.5297	0.5141
$l\hat{\rho}_1$	0.3355	10.3767	1.0519	0.3472	0.1763	0.1035	0.0654
$l\hat{\rho}_2$	0.4031	1.0519	5.8032	0.8718	0.3415	0.1847	0.1122
$l\hat{\rho}_3$	0.4551	0.3472	0.8718	4.9206	0.8292	0.3484	0.1948
$l\hat{\rho}_4$	0.4946	0.1763	0.3415	0.8292	4.5116	0.8053	0.3517
$l\hat{\rho}_5$	0.5297	0.1035	0.1847	0.3484	0.8053	4.2807	0.7901
$l\hat{\rho}_6$	0.5141	0.0654	0.1122	0.1948	0.3517	0.7901	4.1378
$\gamma = 0$							
$n\hat{\rho}$	10.1122	0.2816	0.3451	0.4000	0.4457	0.4834	0.4638
$l\hat{\rho}_1$	0.2816	10.1122	1.1053	0.4287	0.2531	0.1738	0.1295
$l\hat{\rho}_2$	0.3451	1.1053	5.3612	0.8980	0.4167	0.2640	0.1886
$l\hat{\rho}_3$	0.4000	0.4287	0.8980	4.2839	0.8248	0.4153	0.2740
$l\hat{\rho}_4$	0.4457	0.2531	0.4167	0.8248	3.7065	0.7717	0.4087
$l\hat{\rho}_5$	0.4834	0.1738	0.2640	0.4153	0.7717	3.3268	0.7294
$l\hat{\rho}_6$	0.4638	0.1295	0.1886	0.2740	0.2740	0.7294	3.0507
$\gamma = 1$							
$n\hat{\rho}$	8.5810	0.2366	0.2945	0.3469	0.3895	0.4183	0.3876
$l\hat{\rho}_1$	0.2366	9.8514	1.1555	0.5054	0.3236	0.2358	0.1832
$l\hat{\rho}_2$	0.2945	1.1555	4.9217	0.9057	0.4700	0.3204	0.2410
$l\hat{\rho}_3$	0.3469	0.5054	0.9057	3.6477	0.7827	0.4393	0.3101
$l\hat{\rho}_4$	0.3895	0.3236	0.4700	0.7827	2.9032	0.6807	0.4008
$l\hat{\rho}_5$	0.4183	0.2358	0.3204	0.4393	0.6807	2.3810	0.5925
$l\hat{\rho}_6$	0.3876	0.1832	0.2410	0.3101	0.4008	0.5925	1.9831

Table 2

Values of weights for the optimal jackknife estimator.

m:	2	4	6	8	12
$\gamma=-10$					
w^*	2.1081	1.4004	1.2568	1.1941	1.1357
w_1^*	−0.5216	−0.0828	−0.0299	−0.0137	−0.0034
w_2^*	−0.5865	−0.1043	−0.0443	−0.0243	−0.0097
w_3^*		−0.1042	−0.0449	−0.0255	−0.0117
w_4^*		−0.1091	−0.0449	−0.0256	−0.0120
w_5^*			−0.0450	−0.0256	−0.0121
w_6^*			−0.0478	−0.0256	−0.0121
w_7^*				−0.0260	−0.0121
w_8^*				−0.0277	−0.0121
w_9^*					−0.0121
w_{10}^*					−0.0122
w_{11}^*					−0.0127
w_{12}^*					−0.0133
$\gamma=-1$					
w^*	2.6143	1.6332	1.4287	1.3370	1.2486
w_1^*	−0.6133	−0.0910	−0.0322	−0.0150	−0.0044
w_2^*	−1.0011	−0.1409	−0.0520	−0.0255	−0.0088
w_3^*		−0.1813	−0.0659	−0.0335	−0.0125
w_4^*		−0.2200	−0.0781	−0.0396	−0.0154
w_5^*			−0.0946	−0.0452	−0.0178
w_6^*			−0.1058	−0.0518	−0.0199
w_7^*				−0.0610	−0.0219
w_8^*				−0.0654	−0.0240
w_9^*					−0.0265
w_{10}^*					−0.0296
w_{11}^*					−0.0333
w_{12}^*					−0.0345
$\gamma=1$					
w^*	3.0311	1.8567	1.5993	1.4791	1.3590
w_1^*	−0.7262	−0.1125	−0.0444	−0.0238	−0.0102
w_2^*	−1.3049	−0.1610	−0.0570	−0.0289	−0.0117
w_3^*		−0.2513	−0.0755	−0.0358	−0.0136
w_4^*		−0.3320	−0.1044	−0.0454	−0.0160
w_5^*			−0.1479	−0.0588	−0.0190
w_6^*			−0.1701	−0.0773	−0.0228
w_7^*				−0.1005	−0.0274
w_8^*				−0.1086	−0.0331
w_9^*					−0.0403
w_{10}^*					−0.0487
w_{11}^*					−0.0569
w_{12}^*					−0.0594

Table 3

Bias of OLS and jackknife estimators.

n	$\hat{\rho}$	$\hat{\rho}_{J,m}$	$\hat{\rho}_{J,CY}$	$\hat{\rho}_{J,CK}$	$\hat{\rho}^*_J$	$\hat{\rho}^*_{J,2S}$	$\hat{\rho}^*_{R;J,2S}$
$\gamma = -10$							
24	-0.0508	-0.0092_2	$\mathbf{0.0007_6}$	-0.0049_2	-0.0052_2	-0.0044_2	-0.0232_6
48	-0.0290	-0.0030_2	0.0090_{12}	-0.0003_2	-0.0003_2	$\mathbf{-0.0001_2}$	-0.0116_{12}
96	-0.0171	-0.0028_2	0.0094_2	-0.0013_2	$\mathbf{-0.0012_2}$	-0.0012_2	-0.0041_{12}
192	-0.0094	-0.0007_2	0.0061_2	0.0002_2	0.0001_2	$\mathbf{0.0001_2}$	-0.0013_{12}
$\gamma = -1$							
24	-0.0671	-0.0283_2	$\mathbf{-0.0064_2}$	-0.0104_2	-0.0130_2	-0.0154_2	-0.0455_8
48	-0.0332	-0.0101_2	0.0013_4	$\mathbf{0.0006_2}$	0.0008_2	-0.0023_2	-0.0158_{12}
96	-0.0184	-0.0081_2	$\mathbf{0.0002_4}$	-0.0025_4	-0.0024_4	-0.0043_4	-0.0070_{12}
192	-0.0091	-0.0026_2	0.0007_8	0.0004_2	$\mathbf{0.0003_2}$	-0.0005_2	-0.0022_{12}
$\gamma = -0.1$							
24	-0.0663	-0.0309_2	-0.0137_2	$\mathbf{-0.0111_2}$	-0.0143_2	-0.0182_2	-0.0481_{12}
48	-0.0324	-0.0108_2	0.0023_2	$\mathbf{0.0013_2}$	0.0019_2	-0.0031_2	-0.0162_{12}
96	-0.0174	-0.0083_2	$\mathbf{-0.0022_4}$	-0.0023_4	-0.0025_4	-0.0047_4	-0.0070_{12}
192	-0.0085	-0.0026_2	$\mathbf{-0.0003_6}$	-0.0005_6	-0.0004_6	-0.0004_2	-0.0023_{12}
$\gamma = 0$							
24	-0.0660	-0.0301_2	-0.0144_2	$\mathbf{-0.0112_2}$	-0.0144_2	-0.0184_2	-0.0482_6
48	-0.0322	-0.0109_2	0.0018_2	$\mathbf{0.0012_2}$	0.0018_2	-0.0033_2	-0.0163_{12}
96	-0.0173	-0.0082_2	-0.0026_4	$\mathbf{-0.0023_4}$	-0.0026_4	-0.0047_4	-0.0070_{12}
192	-0.0084	-0.0025_2	-0.0004_4	-0.0005_6	-0.0004_4	$\mathbf{-0.0004_2}$	-0.0023_{12}
$\gamma = 0.1$							
24	-0.0658	-0.0311_2	-0.0151_2	$\mathbf{-0.0112_2}$	-0.0146_2	-0.0187_2	-0.0396_6
48	-0.0320	-0.0109_2	0.0013_2	$\mathbf{0.0011_2}$	0.0017_2	-0.0035_2	-0.0163_{12}
96	-0.0171	-0.0082_2	-0.0029_4	$\mathbf{-0.0024_4}$	-0.0027_4	-0.0047_4	-0.0069_{12}
192	-0.0084	-0.0025_2	-0.0005_4	-0.0005_6	-0.0004_4	$\mathbf{-0.0004_2}$	-0.0023_{12}
$\gamma = 1$							
24	-0.0615	-0.0291_2	-0.0170_2	$\mathbf{-0.0106_2}$	-0.0132_2	-0.0184_2	-0.0370_6
48	-0.0290	-0.0102_2	-0.0019_2	0.0005_2	$\mathbf{0.0005_2}$	-0.0042_2	-0.0151_{12}
96	-0.0154	-0.0071_2	-0.0039_2	$\mathbf{-0.0021_4}$	-0.0026_4	-0.0043_4	-0.0063_{12}
192	-0.0074	-0.0021_2	$\mathbf{0.0001_2}$	-0.0005_6	-0.0005_4	-0.0001_2	-0.0019_{12}

Subscripts denote the value of m.

Table 4

RMSE of OLS and jackknife estimators.

n	$\hat{\rho}$	$\hat{\rho}_{J,m}$	$\hat{\rho}_{J,CY}$	$\hat{\rho}_{J,CK}$	$\hat{\rho}_J^*$	$\hat{\rho}_{J,2S}^*$
$\gamma = -10$						
24	**0.1772**	0.1786_6	0.1958_6	0.1802_6	0.1800_6	0.1801_6
48	0.1100	0.0988_{12}	0.1015_{12}	0.0986_{12}	$\mathbf{0.0984_{12}}$	0.1003_{12}
96	0.0552	0.0528_{12}	0.0543_{12}	0.0523_{12}	$\mathbf{0.0520_{12}}$	0.0536_{12}
192	0.0286	0.0268_8	0.0292_{12}	0.0266_{12}	$\mathbf{0.0266_{12}}$	0.0273_{12}
$\gamma = -1$						
24	0.1395	0.1344_6	0.1327_8	0.1376_8	$\mathbf{0.1313_8}$	0.1405_8
48	0.0731	0.0674_8	0.0620_{12}	0.0654_{12}	$\mathbf{0.0626_{12}}$	0.0692_{12}
96	0.0386	0.0346_6	$\mathbf{0.0312_{12}}$	0.0325_{12}	0.0313_{12}	0.0352_{12}
192	0.0194	0.0171_8	0.0152_{12}	0.0160_{12}	$\mathbf{0.0151_{12}}$	0.0174_{12}
$\gamma = -0.1$						
24	0.1337	0.1295_6	0.1270_6	0.1348_8	$\mathbf{0.1269_6}$	0.1365_8
48	0.0698	0.0644_8	$\mathbf{0.0590_{12}}$	0.0628_{12}	0.0591_{12}	0.0664_{12}
96	0.0359	0.0322_6	$\mathbf{0.0283_{12}}$	0.0302_{12}	0.0284_{12}	0.0328_{12}
192	0.0180	0.0158_8	0.0137_{12}	0.0148_{12}	$\mathbf{0.0137_{12}}$	0.0162_{12}
$\gamma = 0$						
24	0.1331	0.1290_6	$\mathbf{0.1263_6}$	0.1345_6	$\mathbf{0.1263_6}$	0.1360_6
48	0.0694	0.0641_8	$\mathbf{0.0588_8}$	0.0626_{12}	$\mathbf{0.0588_8}$	0.0661_{12}
96	0.0356	0.0319_6	$\mathbf{0.0281_{12}}$	0.0300_{12}	$\mathbf{0.0281_{12}}$	0.0325_{12}
192	0.0179	0.0157_8	$\mathbf{0.0135_{12}}$	0.0146_{12}	$\mathbf{0.0135_{12}}$	0.0161_{12}
$\gamma = 0.1$						
24	0.1324	0.1284_6	$\mathbf{0.1255_6}$	0.1340_6	0.1256_6	0.1355_6
48	0.0690	0.0637_8	0.0583_8	0.0622_{12}	$\mathbf{0.0582_{12}}$	0.0657_{12}
96	0.0354	0.0317_6	0.0279_{12}	0.0298_{12}	$\mathbf{0.0278_{12}}$	0.0323_{12}
192	0.0177	0.0156_6	$\mathbf{0.0134_{12}}$	0.0145_{12}	0.0134_{12}	0.0159_{12}
$\gamma = 1$						
24	0.1257	0.1209_6	$\mathbf{0.1151_6}$	0.1266_6	0.1157_6	0.1278_6
48	0.0635	0.0588_8	$\mathbf{0.0540_8}$	0.0589_{12}	0.0539_8	0.0610_{12}
96	0.0330	0.0299_6	0.0264_{12}	0.0285_{12}	$\mathbf{0.0260_{12}}$	0.0304_{12}
192	0.0161	0.0144_{12}	0.0120_{12}	0.0135_{12}	$\mathbf{0.0118_{12}}$	0.0147_{12}

Subscripts denote the value of m.