Three essays on bias, bias reduction and estimation in autoregressive time series models

Marian Zdravkov Stoykov

A thesis submitted for the degree of Doctor of Philosophy

Essex Business School

University of Essex

October 2018

Abstract

This thesis consists of three essays on the subject of autoregressive time series of order one.

The first essay derives an approximate bias of the ordinary least squares estimator (OLS) of the autoregressive parameter for series with moderate deviations from a unit root and for a fixed autoregressive coefficient. The result is used to derive the asymptotic distribution of the indirect inference method for (moderately) stationary, (moderately) explosive and explosive series with a fixed coefficient. The essay also shows how one can construct a jackknife and a simple bias-reduced estimator for stationary series by use of the bias function. A simple Monte Carlo experiment provides evidence that the three estimators outperform OLS in terms of their bias reduction capabilities.

Given the derived discontinuity of the bias function around the vicinity of unity, the second essay proposes an optimal two-step local to unit root jackknife estimator to try and overcome the problem. This particular version of the jackknife requires knowledge of the variances of the full-sample and sub-sample estimators and the covariances between them. Hence, the essay derives their asymptotic counterparts. Via those asymptotic moments, the essay explains analytically why previous findings have found that using more sub-samples in the construction of the jackknife produces smaller variance.

The third essay provides asymptotic theory for local to unit root autoregressive processes with a drift. It is shown that the limiting distribution is a joint normal with a mean zero and variance-covariance matrix which depends on the localising

parameter. An interesting feature of this setup is that a consistent estimator of the localising parameter can be constructed. Hence, one can construct a t-statistic which has a standard normal limiting distribution to test the hypothesis of a unit root by directly testing the null of the localising parameter being equal to zero.

Dedication

To Maya and Zdravko – two unconditionally loving parents.

Acknowledgements

I am indebted to my supervisor Marcus Chambers for the countless hours of support, guidance and patience. His excellent supervision has made this journey easier and more enjoyable. Being supervised by Marcus Chambers is a privilege and I am grateful to have had the opportunity to learn from him.

I thank Neil Kellard, Georgi Grahovski, Robert Taylor and the participants of the International Association of Applied Econometrics and Econometric Society European Meeting conferences for helpful comments and discussions.

This work would not have been possible in this thesis form without the financial support of the Essex Business School, which is gratefully acknowledged.

I thank my family members Georgi, Mariya, Angel, Krasimira and Stanislav for their encouragement during the hardships of the past few years.

Finally, my warmest thanks go to Juliane for her love and patience when I needed them the most.

Contents

List of Figures

List of Tables

List of Symbols and Abbreviations

$C[0,1]$	the space of continuous real-valued functions on the unit interval
$W(r)$	Wiener process with variance r defined on C[0,1]
$\sim$	asymptotic equivalence
$\simeq$	approximately equal to
$=_d$	distributional equivalence
$:=$	definitional equality
$\lfloor x \rfloor$	integer part of x
$\to$	convergence of a real-valued sequence
$\to_p$	convergence in probability
$\Rightarrow_d$	weak convergence
$o(1)$	tends to zero
$o_p(1)$	tends to zero in probability
$O_p(1)$	bounded in probability
iid	independent and identically distributed
$\mathbb{E}$	mathematical expectation

Chapter 1

Introduction

Economics and finance researchers and practitioners rely heavily on autoregressive time series models. The assumption that the value of an asset, or GDP, inflation, and so on, at time period t depends on the value of the same variable at the previous period, $t-1$, seems quite plausible to make. As such, to make any valid inference, the properties of parameter estimators in stochastic difference equation models need to be well understood.

This thesis is going to focus on autoregressive time series of order one. Suppose that the body of data we are interested in analysing is generated in the following way. Let Ω be the sample space consisting of the set of all possible outcomes, F be a σ-algebra of subsets of Ω and P be a function defined on F that satisfies the axioms of probability. Then (Ω, F, P) is a probability space. Then, the observed data are generated as a realisation

$$y_t = \alpha + \rho_n y_{t-1} + u_t, \qquad t = 1, \ldots, n, \tag{1.1}$$

where u_t is a stochastic process defined on (Ω, F, P), $y_0 = O_p(1)$ is also defined on (Ω, F, P), α is a constant, possibly zero, and ρ_n is a parameter allowed to depend on the sample size, n. This setup is quite general and allows for great flexibility. For example, for a fixed $\rho_n = \phi$, with $|\phi| < 1$, and u_t is iid with mean zero and a positive bounded variance, the process is stationary and ergodic. With other specifications, one can construct non-stationary and, even, explosive series. The three essays that follow will impose different restrictions on the parameters, and u_t, in (1.1) and analyse the properties of the resulting series. Since the three chapters are distinct, this introduction does not contain an overall literature review. These are included separately in each essay for the subject at hand.

The thesis is organised as follows. Chapter 2 provides a discussion on finite sample bias in autoregressive parameter estimation. Chapter 3 discusses an "optimal" jackknife procedure for estimation of the autoregressive parameter. Chapter 4 introduces a drift in a local to unit root setup and shows how the localising parameter can be estimated consistently under general conditions. Chapter 5 contains concluding remarks and discusses some future possible areas of research. All estimation and simulations are conducted in Matlab.

Chapter 2

Least Squares Bias in Time Series with Moderate Deviations from a Unit Root

Abstract

This chapter derives the approximate bias of the least squares estimator of the autoregressive coefficient in discrete autoregressive time series where the autoregressive coefficient is given by $\alpha_T = 1 + c/k_T$, with k_T being a deterministic sequence increasing to infinity at a rate slower than T, such that $k_T = o(T)$ as $T \to \infty$. The cases in which $c < 0$, $c = 0$ and $c > 0$ are considered, corresponding to (moderately) stationary, non-stationary and (moderately) explosive series. The result is used to derive the limiting distribution of the indirect inference method for such processes with moderate deviations from a unit root and for explosive series with a fixed coefficient which does not depend on the sample size. Second, the result demonstrates why the jackknife estimator cannot be constructed for explosive time series for values of the autoregressive parameter close to unity in view of the discontinuity of the bias function, which the chapter derives. Lastly, the expression is used to construct a bias-corrected estimator, and simulations are carried out to assess the three estimators' bias-reduction capabilities.

2.1 Introduction

To try and make our task simpler, in most cases, we rely on asymptotic theory as an approximation to finite sample distributions as asymptotic distributions usually have very simple forms. For example, by applying the central limit theorem to the ordinary least squares (OLS) estimator of the autoregressive coefficient in stationary time series, it can be shown that, under a particular set of assumptions, the former is asymptotically normally distributed with mean zero and a well defined variance. From there, it is straightforward to construct confidence intervals and utilise them for inference. However, it does not have to be the case that any asymptotic distribution is shared by its finite sample counterpart. For example, the exact maximum likelihood (MLE) and OLS estimators share the same asymptotic distribution but differ in terms of their finite sample behaviour as they treat the initial condition differently (the initial condition is asymptotically negligible). Thus, having information on their asymptotic behaviour only is not enough to be a guide on which estimator is to be preferred over the other when applied to finite samples settings. In addition, asymptotic theory relies on having an infinitely large sample, something too luxurious to have in practice. It should also be noted that the OLS estimator has different properties when it is applied to non-stationary and explosive series. All of the above-mentioned becomes very important, especially in macroeconomic settings, since observations of GDP, inflation, etc. are very limited, as most of the macroeconomic variables are usually available quarterly. Thus, it would be of help to know how estimators perform in finite samples.

One of the main features of the OLS estimator of the autoregressive parameter is that, on average, it is downward (negatively) biased for any finite sample. However, the bias vanishes asymptotically. This result holds regardless of whether the data-generating process produces stationary, non-stationary or explosive series.

This chapter has been published as Stoykov (2018) and I am grateful to two anonymous referees for useful comments and suggestions.

This characteristic of the OLS method has been demonstrated both theoretically and via simulations, with many authors having contributed to the topic. In terms of stationary series, Hurwicz (1950) and White (1961) derived the result of the bias by means of series expansions up to order $O(T^{-4})$, where T is the sample size. For the non-stationary case, Phillips (1987a) provides an expansion up to $O(T^{-2})$ order. However, fewer authors have focused on the explosive side. Le Breton and Pham (1989) derive the first order term of the bias explicitly, and Phillips (2012) derives only the asymptotic order of the bias. The above mentioned papers deal with AR(1) processes with normally distributed errors. Shaman and Stine (1988) extend the literature by considering AR models of higher order, which are driven by normal errors, and Bao (2007) derives the bias for an AR(1) process where the errors are allowed to follow any distribution.

These setups take the autoregressive coefficient as fixed and to be either smaller, equal or bigger than one. This means one would need to know *a priori* what the data generating process is. This, for example, could have applications for modelling stock prices as returns are stationary. However, it will be misleading to use the same results for near-integrated processes. As such, Phillips (2012) has considered the local to unit root cases where the autoregressive coefficient is given by $\alpha_T = 1 + c/T$ for both c bigger and smaller than zero and by allowing for $c \to -\infty$ and $c \to \infty$.

Asymptotic theory for processes of the above mentioned form, where $T(1 - \alpha_T) = O(1)$, are well-known (Phillips, 1987b, provides results under very general conditions). However, recent advances in the literature show that non-degenerate asymptotic theory exists for models in which α_T is allowed to go to unity at either a higher or lower than the previously considered $O(T^{-1})$ rate. Andrews and Guggenberger (2008) consider stationary AR models with $T(1 - \alpha_T) = o(1)$ and Giraitis and Phillips (2006) analyse stationary models with $T(1-\alpha_T) \to \infty$. More concretely, Phillips and Magdalinos (2007) (hereafter, PM) derive the limiting distribution of the centered OLS estimator $\hat{\alpha}_T - \alpha_T$, where $\alpha_T = 1 + c/k_T$, with

k_T being a sequence which increases to infinity at a rate slower than T, such that $k_T = o(T)$ as $T \to \infty$. PM consider, both, positive and negative values of c, which corresponds to moderately stationary or moderately explosive processes. The authors show that, in this case, the OLS estimator of the autoregressive parameter follows an asymptotic distribution, which is equivalent to the one that is obtained by considering a fixed autoregressive parameter. These results are further augmented by the work of Phillips *et al.* (2010), who derive the Edgeworth expansion of OLS and show that it is equivalent, up to the third term, to that of the fixed case. This framework of moderate deviations from a unit root was utilised by Phillips *et al.* (2011) to test the NASDAQ index for explosive behaviour and by Figuerola-Ferretti *et al.* (2015) to test for bubbles in London Metal Exchange metals prices.

Given those recent advancements in the literature and the interest of researchers in utilising regressions with moderate deviations from a unit root, this chapter aims to derive an approximate expansion of the bias of the OLS estimator for such processes. Its advantage over that of an Edgeworth expansion is that the former can be used to construct estimators which are very effective in reducing the bias. Immediate candidates that could utilise such results are the jackknife and indirect inference (IIE) estimators. The former aims to remove the first term (or, in principle, first two, three, and so on, but usually this is not pursued) from the asymptotic bias of an initial estimator, usually OLS or ML, and the latter utilises a binding function which links the estimator to the true parameter. By definition of the binding function, the bias expression provides the binding function. Phillips (2012) demonstrates how the binding function can be used to derive the limiting distribution of the IIE. This chapter applies the same method to derive the asymptotic distribution of the IIE for, both, (moderately) stationary and (moderately) explosive series and also for explosive series with a fixed autoregressive coefficient. The result of this work also depicts a severe discontinuity in the bias function in the vicinity of unity and, as such, demonstrates why the jackknife cannot be constructed for any explosive series which are relevant to economics and finance. The chapter also proposes a simple bias-corrected estimator for (moderately) stationary series, which is capable of removing the first-order term of the bias but, unfortunately, does that at the expense of a slightly higher variance in comparison with OLS. The chapter finishes with a simple Monte-Carlo exercise, which assesses the bias-reduction capabilities of those three estimators in comparison to OLS. The outcome shows some evidence in favour of the simple bias-corrected estimator.

The chapter is organised as follows: section 2 summarises the main results and provides a discussion. Section 3 concludes, and the technical details, graphs and tables are collected in the Appendix.

For the bigger part of the paper, with the exception of the discussion in section 2, the lower script in α_T is dropped out for notational simplicity and the symbol $\sum$ will be used for summations running from $t = 1$ to T.

2.2 Main Results

Suppose x_t satisfies the following stochastic difference equation

$$x_t = \alpha_T x_{t-1} + u_t, \qquad \alpha_T = 1 + \frac{c}{k_T} \qquad \text{for } t = 1, \ldots, T, \tag{2.1}$$

where u_t is iid as $N(0, \sigma^2)$ and k_T is a sequence that increases to infinity, such that $k_T = o(T)$ as $T \to \infty$. The distribution of $\boldsymbol{x} = (x_1, ..., x_T)$ in (2.1) is uniquely determined by specifying an initial condition for the process. The density function of $\boldsymbol{x}$ for a constant initial condition, including zero, is given by

$$f(\boldsymbol{x}) = \left(2\pi\sigma^2\right)^{-T/2} \exp\left\{ -\frac{\sum (x_t - \alpha x_{t-1})^2}{2\sigma^2} \right\}.$$

If the initial condition is specified not as a constant but as a random variable with a distribution $N\left(0, \frac{\sigma^2}{1-\alpha^2}\right)$, the density of $\boldsymbol{x}$ becomes

$$f^*(\boldsymbol{x}) = \frac{(1-\alpha^2)^{1/2}}{(2\pi\sigma^2)^{T/2}} \exp\left\{-\frac{(1-\alpha^2)x_0 + \sum(x_t - \alpha x_{t-1})^2}{2\sigma^2}\right\}.$$

The model which this chapter will consider is the former. Unfortunately, when $|\alpha| > 1$ the exact MLE for the latter is inconsistent due to the specification of the initial condition. Hamilton (1994, pp. 118-23) provides a discussion on why that is the case. In particular, we will take $x_0 = 0$. The centered MLE for the constant initial condition, which coincides with the OLS estimator, is given by

$$\hat{\alpha} - \alpha = \frac{\sum x_{t-1}u_t}{\sum x_{t-1}^2}.$$

By considering $\hat{\alpha}_T$, the OLS estimator when the autoregressive coefficient is assumed to be α_T, under appropriate normalisations, PM showed, that for negative c the estimator converges to a normal random variable, and for positive c the limiting distribution is the standard Cauchy.

To derive the bias, we will make use of the joint moment generating function (MGF) of the numerator and denominator of the result for $\hat{\alpha}$. We may assume $\sigma^2 = 1$ as $\hat{\alpha}$ is independent of σ^2. Following the procedure of White, define $U = \sum x_t x_{t-1}$ and $V = \sum x_{t-1}^2$ such that their joint MGF is given by

$$\begin{aligned}
&\mathbb{E}\exp(Uu + Vv) \\
&\quad = \int_{-\infty}^{\infty} \exp(Uu + Vv) f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\
&\quad = (2\pi)^{-T/2} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left[-2(\alpha+u)U + (1+\alpha^2-2v)V + x_T^2\right]\right\} \mathrm{d}\boldsymbol{x} \\
&\quad = (2\pi)^{-T/2} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\boldsymbol{x}D_T\boldsymbol{x}'\right\} \mathrm{d}\boldsymbol{x} = |D_T(u,v)|^{-1/2}, \qquad (2.2)
\end{aligned}$$

where the second line follows from the fact that, for $\sigma^2 = 1$, the density becomes $f(\boldsymbol{x}) = (2\pi)^{-T/2}\exp\left\{-\frac{1}{2}\left(V + x_T^2 - 2\alpha U + \alpha^2 V\right)\right\}$. The matrix in the third line is given by

$$D_T(u,v) = \begin{bmatrix} p(v) & q(u) & 0 & 0 & \cdots & 0 \\ q(u) & p(v) & q(u) & 0 & \cdots & 0 \\ 0 & q(u) & p(v) & q(u) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & q(u) & p(v) & q(u) \\ 0 & 0 & 0 & \cdots & q(u) & 1 \end{bmatrix}$$

with $p(v) = 1+\alpha^2 - 2v$ and $q(u) = -(\alpha+u)$. The last equality in (2.2) is a result which can be found in Cramér (1946, pp. 118-20). Shenton and Johnson (1965) showed that

$$\mathbb{E}(\hat{\alpha} - \alpha) = \int_0^\infty \frac{\partial}{\partial \alpha} |D_T(-v)|^{-1/2} \mathrm{d}v, \tag{2.3}$$

where $|D_T(-v)|$ is the determinant of the matrix evaluated at $u = 0$ and $-v$. The authors showed that the integrand and integral in (2.3) exist and are well-defined. Now, the right hand side of (2.3) can be utilised to derive the approximate bias (the details are given in the appendix). The results are summarised in the following theorem.

Theorem 2.2.1. *For the model considered in* (2.1)*, with* $x_0 = 0$*, as* $T \to \infty$*, the bias of the autoregressive coefficient* $\mathbb{E}(\hat{\alpha}_T - \alpha_T)$ *has the following asymptotic expansion*

$$\mathbb{E}(\hat{\alpha}_T - \alpha_T) = \begin{cases} -\frac{2\alpha_T}{T}\left\{1 + O\left(\max\left(\frac{k_T}{T}, k_T^{-1}\right)\right)\right\}\left[1 + O\left(T^{-1}\right)\right], & c < 0, \\ -\frac{1.7814}{T}\left[1 + O\left(T^{-1}\right)\right], & c = 0, \\ -\frac{2\pi^{1/2}c^{3/2}T^{1/2}}{k_T^{3/2}\alpha_T^{T+1}}\left\{1 + O\left(\max\left(\frac{k_T}{T}, k_T^{-1}\right)\right)\right\}\left[1 + O\left(T^{-1}\right)\right], & c > 0. \end{cases}$$

Remark 1. The bias is negative for all values of c. Starting with $c < 0$, which corresponds to $|\alpha| < 1$, the first term of the expansion is $-2\alpha_T/T$. For a fixed α the first order term of the expansion is $-2\alpha/T$ (White (1961)). Thus, up to the first order term the two results are equivalent. However, the results differ in higher order terms. The second order term for the fixed case is $4\alpha/T^2$ (e.g. White), which is different from the result for $c < 0$ due to the presence of the $1 + O\left(\max\left(k_T/T, k_T^{-1}\right)\right)$ term. Furthermore, it is not clear what the second order term of the expansion is until one specifies k_T, as the result in Theorem 2.2.1 is very general. One consistent with the definition of k_T parameterisation would be to set $k_T = T^{\delta}$, with $\delta \in (0,1)$. In this case, the cut-off point is $\delta = 1/2$. Thus, for $\delta \in (0,1/2]$, $\max\left(k_T/T, k_T^{-1}\right) = k_T^{-1} = T^{-\delta}$ and for $\delta \in [1/2,1)$, $\max\left(k_T/T, k_T^{-1}\right) = k_T/T = T^{\delta-1}$. Finally, it is perhaps surprising that the second-order term is not of order $O\big(T^{-1}\big)$ smaller in magnitude than the first in view of the fixed coefficient expansion.

Remark 2. For $c = 0$, corresponding to $\alpha = 1$, a unit root process, the constant -1.7814 is a well known result (e.g. Tanaka (1996), p240). It is an approximation (up to the fourth decimal) of the expectation of the functional $\int_0^1 W(r)dW(r) \Big/ \int_0^1 W(r)^2 dr$, which is the leading term of the asymptotic bias for the unit root process (see Phillips (1987a)).

Remark 3. For $c > 0$, which corresponds to $|\alpha| > 1$, the first term of the expansion is $-2\pi^{1/2}c^{3/2}T^{1/2}k_T^{-3/2}\alpha_T^{-(T+1)}$. From Le Breton and Pham, the respective term from the expansion for a fixed α is $-(\pi/2)^{1/2}(\alpha^2-1)^{3/2}T^{1/2}\alpha^{-(T+1)}$. Note that $\alpha_T^2 - 1 = (1+c/k_T)^2 - 1 = 2c/k_T\left(1 + O\left(k_T^{-1}\right)\right)$. Thus, for the first order term of the expansion the two results are equivalent. The result of Le Breton and Pham does not provide the second order term of the expansion for the fixed case and, as such, comparison between the two cannot be made. Thus, the present paper sheds some light into the higher order terms of the expansion on the explosive side. Taking a closer look into the result, it is perhaps surprising that the second-order term is not of order $O\big(\alpha^{-T}\big)$ smaller in magnitude than the first. One would have expected that the order would be a multiple of α^{-T} as the consecutive terms in the stationary and non-stationary cases are a multiple of T^{-1}.

Remark 4. This formulation, namely $\alpha_T = 1 + c/k_T$, suffers from the same problem as the result of Le Breton and Pham. Their analytical result is discontinuous in that the limits of their bias expressions $\alpha \nearrow 1$ and $\alpha \searrow 1$ are different. This is also the case in the analytical result of Theorem 2.2.1 from the present chapter, namely, the limits $c \nearrow 0$ and $c \searrow 0$ are not identical. This is not surprising as the result from this chapter and that of Le Breton and Pham are asymptotically equivalent up to the first term. The solid line in Figure 2.1 depicts the result with $T = 24$, $k_T = \sqrt{T}$ and $c \in [-\sqrt{T}, 3]$, where the lower limit of c is taken as $-\sqrt{T}$ to match with $\alpha_T = 1 + c/\sqrt{T} = 0$. The lonely dot is the result for $c = 0$. The graph also shows the simulated OLS bias for the autoregressive coefficient of (2.1) with $x_0 = 0$ and $\sigma = 1$ for comparison. The number of replications is $10,000$, and the graph has been smoothed. It can be seen that for negative values of c or high positive values of the parameter the analytical solution provides a good approximation. It is also expected that adding additional terms from the bias would improve the approximation. This, however, is not undertaken in this work as the first term is enough to construct the jackknife and indirect estimations. Those are discussed in subsequent remarks.

Remark 5. PM provide an interesting discussion on the comparison between a fixed $\alpha = 1+c$ and the moderate deviations from a unit root given by $\alpha_T = 1+c/k_T$ frameworks, where k_T was parameterised in remark 1. Their aim is to find out whether taking the limits $\delta \to 0, 1$ produces the correct rates of convergence and asymptotic distributions. They show this to be the case only for the boundary limit $\delta \to 0$. When $k_T = T^{\delta}$ from Theorem 3.2 of PM, as $T \to \infty$, we have

$$T^{1/2+\delta/2}\left(\hat{\alpha}_T - \alpha_T\right) \Rightarrow_d N(0, -2c) \qquad \text{for } c < 0, \tag{2.4}$$

$$\frac{T^{\delta}\alpha_T^T}{2c}\left(\hat{\alpha}_T - \alpha_T\right) \Rightarrow_d C \qquad \text{for } c > 0, \tag{2.5}$$

where C is the standard Cauchy random variable. In comparison, it is well-known that for the model considered in (2.1), and a fixed α

$$T^{1/2}\left(\hat{\alpha}-\alpha\right) \Rightarrow_d N\left(0,1-\alpha^2\right) \qquad \text{for } |\alpha| = |1+c| < 1, \tag{2.6}$$

$$\frac{\alpha^T}{\alpha^2-1}\left(\hat{\alpha}-\alpha\right) \Rightarrow_d C \qquad \text{for } \alpha = 1+c > 1, \tag{2.7}$$

where the result in (2.7) is due to White (1958). From (2.4) and (2.6) there is a discrepancy between the terms $-2c$ and $1-\alpha^2 = -2c-c^2$, and from (2.5) and (2.7) between $2c$ and $\alpha^2-1 = 2c+c^2$. Continuity can be achieved by substituting c with $c+c^2/2T^\delta$ without affecting the asymptotic distributions and moments. This argument does not apply for the case $\delta \to 1$, as discussed by PM.

In the same fashion, it would be interesting to check whether the same arguments hold for the bias as well. From Theorem 3.1 of Le Breton and Pham, as $T \to \infty$

$$T\mathbb{E}(\hat{\alpha}-\alpha) \text{ converges to } -2\alpha \qquad \text{for } |\alpha| = |1+c| < 1 \tag{2.8}$$

$$T^{-1/2}\alpha^{T+1}\mathbb{E}(\hat{\alpha}-\alpha) \text{ converges to } (\pi/2)^{1/2}(\alpha^2-1)^{3/2} \quad \text{for } \alpha = 1+c > 1. \tag{2.9}$$

For $c < 0$, Theorem 3.1 from Le Breton and Pham and Theorem 2.2.1 produce the same result as $\delta \to 0$ without any adjustment. However, for $c > 0$ Theorem 2.2.1 and (2.9) involve an $\alpha^2-1 = 2c+c^2$ term. Thus, continuity can be achieved by the substitution proposed by PM. This result follows immediately by substituting k_T with 1 at the beginning of the integral expressions from the appendix and is stated as a corollary to Theorem 2.2.1.

Corollary 2.2.2. *For model (1) with $x_0 = 0$ and a fixed $\alpha = 1+c$, as $T \to \infty$, the bias of the autoregressive coefficient $\mathbb{E}(\hat{\alpha}-\alpha)$ has the following asymptotic expansion*

$$\mathbb{E}(\hat{\alpha}-\alpha) = \begin{cases} -\frac{2(1+c)}{T}\left[1+O\left(T^{-1}\right)\right], & c<0, \\ -\frac{1.7814}{T}\left[1+O\left(T^{-1}\right)\right], & c=0, \\ -\frac{\pi^{1/2}(2c+c^2)^{3/2}T^{1/2}}{2^{1/2}(1+c)^{T+1}}\left[1+O\left(T^{-1}\right)\right], & c>0. \end{cases}$$

There is nothing surprising regarding this result given the discussion around Theorem 2.2.1. However, it shows the magnitude of the second order term on the explosive side for a fixed autoregressive coefficient, which is new to the literature. Lastly, as in the discussion of PM, taking the limit $\delta \to 1$ does not produce the same asymptotic bias. These results are interesting. One would have perhaps expected that there should be no need for adjustment for the asymptotic bias, only for the asymptotic variance, in view of the fact that only the variances of the asymptotic distributions depend on either $1-\alpha^2$ or α^2-1 for the stationary or explosive cases respectively.

Remark 6. It is also interesting to compare the convergence rates of the centered OLS estimator, stated in (2.4) and (2.5), to the orders of the bias from Theorem 2.2.1. As pointed out by a referee, for $c<0$, from (2.4), with T^δ replaced by k_T, we can see that the order of $\hat{\alpha}_T-\alpha_T$ is $O_p\big(1/\sqrt{Tk_T}\big)$ and the corresponding order of the bias is $O(1/T)$. This can be explained by the fact that the limiting distribution of the centered OLS estimator is a $N(0,-2c)$ variate which has a zero expectation. For $c>0$, the centered OLS estimator converges to a Cauchy variate with a convergence rate $k_T\alpha_T^T$ and the corresponding order of the bias is

$$\mathbb{E}(\hat{\alpha}_T-\alpha_T)=O\left(T^{1/2}k_T^{-3/2}\alpha_T^{-T}\right)=\left(\frac{T}{k_T}\right)^{1/2}O\left(k_T^{-1}\alpha_T^{-T}\right),$$

which is slower than the consistency rate $O_p\left(k_T^{-1}\alpha_T^{-T}\right)$ due to $T/k_T\to\infty$. The bias' slower rate could be explained by the fact that the Cauchy distribution has undefined moments. However, even though its first moment does not exist, the standard central Cauchy distribution is centered around zero which helps in explaining why $\mathbb{E}(\hat{\alpha}_T-\alpha_T)$ still goes to zero in the limit.

Remark 7. It is possible to derive the asymptotic distributions of the IIE for positive and negative c by utilising the results from Theorem 2.2.1. Define the binding function $b_T(\alpha) \coloneqq \mathbb{E}(\hat{\alpha})$. The IIE is then defined as

$$\tilde{\alpha}_T = \arg\min_{\alpha} ||\hat{\alpha} - b_T(\alpha)||,$$

for some metric $||\cdot||$. The idea behind the IIE is similar to that of the jackknife. Even though they do it in different ways, they both aim at subtracting an expression of the bias from the original estimator. This should in theory work as we know that on average OLS would underestimate the true parameter, as depicted in Theorem 2.2.1.

To derive the asymptotic distribution of the IIE we note that if $b_T(\alpha_T)$ is invertible, we have that $\tilde{\alpha}_T = b_T^{-1}(\hat{\alpha}) =: f_T(\hat{\alpha})$. Starting with $c < 0$, the derivative of the binding function with respect to α_T is

$$b_T'(\alpha_T) = 1 + O(T^{-1}). \tag{2.10}$$

For $c < 0$, corresponding to $|\alpha_T| < 1$, $\delta > 0$ and any sequence $s_T \to \infty$ for which $s_T/(\sqrt{k_T T}) \to 0$, we have

$$\sup_{s_T|\alpha_T - r| < \delta} \left| \frac{b_T'(\alpha_T) - b_T'(r)}{b_T'(r)} \right| \to 0.$$

Using the fact that $f_T'(r) = 1/b_T'(r)$ and

$$\frac{f_T'(r) - f_T'(\alpha_T)}{f_T'(\alpha)} = \frac{b_T'(\alpha_T) - b_T'(r)}{b_T'(r)},$$

it follows that

$$\sup_{s_T|r - \alpha_T| < \delta} \left| \frac{f_T'(r) - f_T'(\alpha_T)}{f_T'(\alpha_T)} \right| \to 0.$$

By considering (2.4), with T^δ replaced by k_T, (2.10) and applying Theorem 1 from Phillips (2012), we have as $T \to \infty$

$$\sqrt{k_T T}(\tilde{\alpha}_T - \alpha_T) \sim \frac{1}{b_T'(\alpha)} \sqrt{k_T T}(\hat{\alpha}_T - \alpha_T) \sim \sqrt{k_T T}(\hat{\alpha}_T - \alpha_T) \Rightarrow_d N(0, -2c).$$

Hence, as in the fixed case, the IIE for the autoregressive parameter in moderately stationary series has the same asymptotic distribution as OLS.

The two explosive cases can be tackled in a similar fashion. To facilitate the discussion we will make use of the following lemma which is proven in the appendix.

Lemma 2.2.3. *For $\alpha = 1 + c$ and $\alpha_T = 1 + c/k_T$, define $g(\alpha) \coloneqq (\pi/2)^{1/2}(\alpha^2 - 1)^{3/2}T^{1/2}\alpha^{-(T+1)}$ and $h(\alpha_T) \coloneqq 2(\pi T)^{1/2}(\alpha_T - 1)^{3/2}\alpha_T^{-(T+1)}$. Then, for $c > 0$, as $T \to \infty$*

a) $g'(\alpha) = o(1)$;

b) $h'(\alpha_T) = o(1)$.

Starting with the fixed case, from Corollary 2.2.2, $b_T(\alpha) = \alpha - g(\alpha)(1 + o(1))$ in view of $\alpha = 1 + c$. The derivative of the binding function is then given by

$$b'_T(\alpha_T) = 1 + o(1), \tag{2.11}$$

by Lemma 1. By considering any $\delta > 0$ and any sequence $s_T \to \infty$ for which $s_T/\alpha^T \to 0$, we have

$$\sup_{s_T|r-\alpha|<\delta} \left| \frac{f'_T(r) - f'_T(\alpha)}{f'_T(\alpha)} \right| \to 0.$$

By considering (2.7), (2.11), and applying Theorem 1 from Phillips (2012), we have

$$\frac{\alpha^T}{\alpha^2 - 1}(\tilde{\alpha} - \alpha) \sim \frac{1}{b'_T(\alpha)} \frac{\alpha^T}{(\alpha^2 - 1)}(\hat{\alpha} - \alpha) \sim \frac{\alpha^T}{\alpha^2 - 1}(\hat{\alpha} - \alpha) \Rightarrow_d C,$$

where C is the standard Cauchy distribution.

The case of moderate deviations can be tackled in the same fashion. From Theorem 2.2.1, the binding function is given by $b_T(\alpha_T) = \alpha_T - h(\alpha_T)(1 + o(1))$ in view of $(c/k_T)^{3/2} = (\alpha_T - 1)^{3/2}$ by the definition of α_T. The derivative of the binding function is given by

$$b'_T(\alpha_T) = 1 + o(1), \tag{2.12}$$

by Lemma 1. Consequently, for $\delta > 0$ and any sequence $s_T \to \infty$ for which $s_T/(k_T\alpha_T^T) \to 0$ we have

$$\sup_{s_T|r-\alpha_T|<\delta} \left| \frac{f_T'(r) - f_T'(\alpha_T)}{f_T'(\alpha_T)} \right| \to 0.$$

By considering (2.5), with T^δ replaced by k_T, (2.12) and applying Theorem 1 from Phillips (2012), we have as $T \to \infty$

$$\frac{k_T\alpha_T^T}{2c}(\tilde{\alpha}_T - \alpha_T) \sim \frac{1}{b_T'(\alpha)} \frac{k_T\alpha_T^T}{2c}(\hat{\alpha}_T - \alpha_T) \sim \frac{k_T\alpha_T^T}{2c}(\hat{\alpha}_T - \alpha_T) \Rightarrow_d C.$$

Thus, the IIE shares the same asymptotic distribution as OLS in the explosive and moderately explosive cases as well. Figures 2.2 and 2.3 depicts non-parametric estimates of the densities of the OLS and II estimators for $\alpha_T = 1 + c/\sqrt{T}$, with $c = \{-10, 0.5\}$ and large values of $T = \{3000, 500\}$ for the stationary and explosive sides respectively. We have utilised a normal kernel with fixed bandwidths of $h_{OLS} = 1.06R^{-1/5}\hat{\sigma}_{OLS}$ and $h_{IIE} = 1.06R^{-1/5}\hat{\sigma}_{IIE}$, where $R = 10,000$ is the number of replications and $\hat{\sigma}_{(.)}$ denotes the estimated standard deviation. The number of observations used differ in the two cases due to the fast exponential convergence rate of the explosive side which causes numerical issues. We can observe that it's very difficult to distinguish between the two estimators for large values of T.

Even though the expression of the binding function can be utilised to derive the asymptotic distribution of the estimator, one should proceed with caution when constructing the IIE for values of the autoregressive parameter close to unity. Due to the non-linearity of the binding function, it would be best to use the simulated binding function rather than the analytical expression. This is vital on the explosive side as, otherwise, the estimator will minimise the distance at the wrong value for α and consequently produce undesirable results. Furthermore, once away from unity, using results such as the one from Theorem 1 provides ground for construction of the IIE that does not rely on simulations. If such a result can be obtained for a general distribution of the error term, then simulations can be completely bypassed as there would be no need for data generation which is distribution dependent.

Remark 8. Another candidate which could utilise the result from Theorem 2.2.1 is the jackknife estimator. The idea behind it is to remove the first order bias of an estimator, provided that it exists. The jackknife utilises sub-samples and it requires an asymptotic expression for the bias of the original estimator for the full-sample and a number of sub-samples. Construction of the jackknife involves giving a weight to each of the full-sample and sub-sample estimators such that the first-order term bias of the original estimator is removed. Chambers (2013) and Chambers and Kyriacou (2013) demonstrate how bias reduction can be achieved in stationary and non-stationary autoregressive series respectively. Theorem 2.2.1 shows that the analysis of Chambers can be directly translated to moderately stationary series. To demonstrate this, consider splitting the entire sample T into m non-overlapping sub-samples each of length l, such that $T = m \times l$. The jackknife for stationary series is then defined as

$$\hat{\alpha}_J = w_1 \hat{\alpha} + w_2 \frac{1}{m} \sum_{j=1}^{m} \hat{\alpha}_j, \qquad j = 1, \ldots, m, \tag{2.13}$$

where $\hat{\alpha}_j$ is the OLS estimator within each sub-sample. Chambers assumes that the bias expansion of each sub-sample has the same form as that of the full sample. In the fixed autoregressive coefficient case, this is justified as White (1961) showed that the bias expression is the same for the first term of the bias regardless of whether the initial condition is a constant or a stationary random variable. In the moderate deviations case this would be justified too as the bias expression for the full sample is asymptotically equivalent to that of the fixed case when the initial condition is zero. We choose $k_T = \sqrt{T}$, and under the above assumption and Theorem 2.2.1, we have

$$\mathbb{E}(\hat{\alpha}_T) = \alpha_T - \frac{2\alpha_T}{T} + O\left(T^{-3/2}\right) \qquad \text{and} \qquad \mathbb{E}(\hat{\alpha}_{j;T}) = \alpha_T - \frac{2\alpha_T}{l} + O\left(l^{-3/2}\right),$$

for $j = 1, \ldots, m$. Note that since T and l are of the same magnitude they are interchangeable under the big-oh notation. The optimal weights in (2.13) are chosen such that the the jackknife would be biased of order $O(T^{-3/2})$, which is of a lower order than $O(T^{-1})$, the order of OLS and ML. Consider the expectation of the jackknife estimator for (moderately) stationary series

$$\begin{aligned}
\mathbb{E}(\hat{\alpha}_{J;T}) &= w_1 \mathbb{E}(\hat{\alpha}_T) + w_2 \frac{1}{m} \sum_{j=1}^{m} \mathbb{E}(\hat{\alpha}_{j;T}) \\
&= w_1 \left(\alpha_T - \frac{2\alpha_T}{T} \right) + w_2 \frac{1}{m} \sum_{j=1}^{m} \left(\alpha_T - \frac{2\alpha_T}{l} \right) + O\left(T^{-3/2}\right) \\
&= (w_1 + w_2)\alpha_T - 2\alpha_T \left(\frac{w_1}{T} + \frac{w_2}{l} \right) + O\left(T^{-3/2}\right).
\end{aligned}$$

Setting $w_1 + w_2 = 1$ and $w_1/T + w_2/l = 0$ yields the same optimal weights $w_1^* = m/(m-1)$ and $w_2^* = -1/(m-1)$ as the ones founds in Chambers' work. The author of the present paper has attempted to construct the jackknife estimator for mildly explosive autoregressive series but was unsuccessful due to the severe right-hand discontinuity, which was subsequently derived in this paper. As such, Theorem 2.2.1 depicts why it would be impossible to construct the jackknife for mildly explosive series for positive values of c which are close to zero. The issue can be tackled by either substituting what should be the correct explosive weights by the ones applied to stationary series (Kruse and Kaufmann (2018) provide an interesting comparison between a number of estimators) or by considering local to unit root alternatives (Chambers and Kyriacou, 2018), where the approximate bias function has been shown to be continuous at $c = 0$ (Phillips (2012)).

Remark 9. Another way to try and utilise the result in Theorem 2.2.1 is to get an estimate of the bias and subtract it from the OLS estimator. However, as with the jackknife, this would only work on the stationary side. More formally, construct $\breve{\alpha}_T$ as

$$\breve{\alpha}_T = \hat{\alpha}_T - \widehat{bias} = \hat{\alpha}_T - \left(-\frac{2\hat{\alpha}_T}{T}\right) = \hat{\alpha}_T\left(1 + \frac{2}{T}\right).$$

The expectation of the estimator is given by

$$\mathbb{E}(\breve{\alpha}_T) = \mathbb{E}(\hat{\alpha}_T)\left(1 + \frac{2}{T}\right) = \left(\alpha_T - \frac{2\alpha_T}{T} + O\left(T^{-2}\right)\right)\left(1 + \frac{2}{T}\right) = \alpha_T + O\left(T^{-2}\right)$$

and its variance by

$$VAR(\breve{\alpha}_T) = \left(1 + \frac{2}{T}\right)^2 VAR(\hat{\alpha}_T).$$

Thus, the bias-corrected estimator is biased of order T^{-2}. Unfortunately, this happens at the expense of having a higher variance. MacKinnon and Smith (1998) propose a similar to this estimator to reduce bias in autoregressive series with an intercept. However, they rely on simulations instead of an analytical expression and, as such, their procedure would be more time consuming.

Remark 10. We finish the discussion by providing a simple Monte-Carlo experiment, where a comparison is made between the OLS, II, bias-corrected and jackknife estimators in terms of their bias. The latter three are constructed by means of utilising the result from Theorem 2.2.1. We simulate data from (2.1) with $\alpha_T = 1 + c/\sqrt{T}$, $u_t \sim N(0,1)$ and $x_0 = 0$. The number of replications is set at $10,000$, $n = \{30, 60, 90, 120\}$ and $c = \{-5, -4, -3, -2, -1\}$. The results are gathered in Table 2.1. It can be observed that each of the estimators reduces bias dramatically in comparison to OLS. Furthermore, the simple bias-corrected estimator seems to produce the lowest bias for most of the values of the parameters considered.

2.3 Conclusion

This chapter has had the aim to derive the approximate bias of the autoregressive parameter in the discrete time autoregressive process of order one. The autoregressive coefficient is assumed to be of the form $\alpha_T = 1 + c/k_T$, where c is a constant and k_T is a sequence that is increasing to infinity at a slower rate than T, the sample size, such that $k_T = o(T)$ as $T \to \infty$. The bias is shown to be negative for the three cases considered, namely $c < 0$, $c = 0$ and $c > 0$, corresponding to (moderately) stationary, non-stationary and (moderately) explosive series, respectively. The bias expression is shown to be discontinuous in that the limits $c \nearrow 0$ and $c \searrow 0$ are not identical in the final bias expression.

The results of the chapter are used to derive the limiting distribution of the IIE for (moderately) stationary, (moderately) explosive series and explosive series with a fixed coefficient. It is shown that the asymptotic distribution of the IIE is identical to that of OLS in each of the three cases considered. The result is also used to construct the jackknife estimator for (moderately) stationary series. Furthermore, given the severe discontinuity this paper derives, it also shows why the jackknife cannot be constructed on the explosive side for values of the autoregressive parameter close to one and also explains why construction of the IIE on the same side of unity should be done by simulating the bias rather than by using the analytical expression. Phillips (2012) had already derived the limiting distribution of the IIE in the local to unit root case and shown that it's different from that of OLS. Nevertheless, the bias expression on the stationary side seems satisfactory to utilise (unlike the explosive) even for values of the autoregressive coefficient close to one. The final usage of the result is to construct a simple bias-corrected estimator the idea of which is to just subtract an estimate of the bias from the OLS estimator. The paper finishes by conducting a simple Monte-Carlo exercise, the purpose of which is to compare the bias-reduction capabilities of the II, jackknife and bias-corrected estimators to that of OLS'. The simulations provide some evidence that the simple bias-corrected estimator outperforms the rest for most of the values of the parameters considered.

The results of this chapter could be extended in a couple of directions. First, it would be of interest to derive an expression such as the one found in Theorem 2.2.1 of the present paper for an error term which follows a general distribution. Second, derivation of higher order moments. It would be interesting to observe what form higher order centered moments would have on the explosive side as we know that they can rewritten as a function of the first moment, which is discontinuous at the vicinity of unity. It might also be of interest to derive those on the stationary side as the first moment in the case of moderate deviations from a unit root is equivalent to that of a fixed autoregressive parameter but only up to the first term.

2.4 Appendix

Proof of Theorem 2.2.1. To evaluate the determinant in (2.3) we note that it can be written in the form of a second order difference equation. Define $\kappa := p(-v) = 1 + \alpha^2 + 2v$ such that

$$|D_T(-v)| = \kappa |D_{T-1}(-v)| - \alpha^2 |D_{T-2}(-v)|,$$

with initial conditions $|D_1(-v)| = 1$ and $|D_2(-v)| = \kappa - \alpha^2$. Denote the positive root of the characteristic equation as

$$\lambda = \frac{1}{2}\left(\kappa + \sqrt{\kappa^2 - 4\alpha^2}\right) \tag{2.14}$$

such that the solution to the homogeneous difference equation has the form

$$|D_T(-v)| = C_1 \lambda^T + C_2 \left(\frac{\alpha^2}{\lambda}\right)^T.$$

From the initial conditions the complete solution is given by

$$|D_T(-v)| = \left(\frac{\lambda-\alpha^2}{\lambda^2-\alpha^2}\right)\lambda^T + \left(1-\frac{\lambda-\alpha^2}{\lambda^2-\alpha^2}\right)\alpha^{2T}\lambda^{-T}.$$

Phillips (2012) showed that $|D_T(-v)|$ is positive for all $v > 0$ meaning the integrand in (2.3) is well-defined for all α, and Shenton and Johnson (1965) showed that the improper integral is well-defined. They write that, for v small and positive, the integral converges since $D_T(0,0) = 1$ and for v large and positive

$$\frac{\partial|D_T(-\nu)|^{-1/2}}{\partial\alpha} = O(\nu^{-\omega}),$$

where $2\omega = n-5$. Thus, the integrand in (2.3) converges for a large enough n since it would bounded by a integrable function.

Taking the derivative in (2.3) leads to

$$\mathbb{E}(\hat{\alpha}-\alpha) = -\frac{1}{2}\int_0^\infty |D_T(-v)|^{-3/2}\frac{\partial|D_T(-v)|}{\partial\alpha}\mathrm{d}v.$$

Following Phillips (2012), define $x = 1/\lambda$. It follows that

$$x = \frac{2}{\kappa+\sqrt{\kappa^2-4\alpha^2}} = 2\frac{\kappa-\sqrt{\kappa^2-4\alpha^2}}{\kappa^2-(\kappa^2-4\alpha^2)} = \frac{\kappa-\sqrt{\kappa^2-4\alpha^2}}{2\alpha^2},$$

from which it follows that

$$\kappa-\sqrt{\kappa^2-4\alpha^2} = 2\alpha^2 x. \tag{2.15}$$

From (2.14) and (2.15) we have

$$\kappa-\sqrt{\kappa^2-4\alpha^2} = 2\alpha^2 x \quad \text{and} \quad \kappa+\sqrt{\kappa^2-4\alpha^2} = \frac{2}{x},$$

and by adding them together we get $\kappa = 1/x + \alpha^2 x$. From $\kappa = 1 + \alpha^2 + 2v$ we can solve for v as a function of x

$$v = \frac{1}{2}\left\{\frac{1}{x} + \alpha^2 x - \alpha^2 - 1\right\} = \frac{(1-x)(1-\alpha^2 x)}{2x}.$$

We write

$$v = \frac{(1-x)(1-\alpha^2 x)}{2x} = \begin{cases} \frac{(1-x)(1-\alpha^2 x)}{2x}, \ x \in (0,1], & |\alpha| \le 1, \\ \frac{(x-1)(\alpha^2 x-1)}{2x}, \ x \in [1,\infty), & |\alpha| > 1, \end{cases}$$

with derivative

$$\frac{dv}{dx} = -\frac{(1-\alpha^2 x^2)}{2x^2} \begin{cases} < 0, \ x \in (0,1], & |\alpha| \le 1, \\ > 0, \ x \in [1,\infty), & |\alpha| > 1. \end{cases}$$

As pointed out by Phillips (2012), $v = v(x)$ is a monotonic transformation over the two domains specified and as such the variable of integration in (2.3) can be changed. Applying the change of variable yields

$$\begin{aligned} \frac{\lambda-\alpha^2}{\lambda^2-\alpha^2} &= \frac{\frac{1}{x}-\alpha^2}{\frac{1}{x^2}-\alpha^2} = \frac{x(1-\alpha^2 x)}{1-\alpha^2 x^2}, \\ 1-\frac{\lambda-\alpha^2}{\lambda^2-\alpha^2} &= 1-\frac{x(1-\alpha^2 x)}{1-\alpha^2 x^2} = \frac{1-x}{1-\alpha^2 x^2}, \end{aligned}$$

leading to

$$\begin{aligned} |D_T(-v)| &= \frac{1-\alpha^2 x}{1-\alpha^2 x^2}\frac{1}{x^{T-1}} + \frac{1-x}{1-\alpha^2 x^2}\alpha^{2T} x^T \\ &= \begin{cases} \frac{1-\alpha^2 x + (1-x)\alpha^{2T} x^{2T-1}}{(1-\alpha^2 x^2) x^{T-1}} = \frac{A_T(x;\alpha)}{(1-\alpha^2 x^2) x^{T-1}}, & \text{for } |\alpha| \le 1, \\ \frac{\alpha^2 x - 1 + (x-1)\alpha^{2T} x^{2T-1}}{(\alpha^2 x^2 - 1) x^{T-1}} = \frac{B_T(x;\alpha)}{(\alpha^2 x^2-1) x^{T-1}}, & \text{for } |\alpha| > 1, \end{cases} \end{aligned} \tag{2.16}$$

with

$$\begin{aligned} A_T(x;\alpha) &= 1 - \alpha^2 x + (1-x)\alpha^{2T} x^{2T-1}, \\ B_T(x;\alpha) &= \alpha^2 x - 1 + (x-1)\alpha^{2T} x^{2T-1}. \end{aligned} \tag{2.17}$$

Thus, for $|\alpha| \leq 1$, by changing the variable of integration from v to x and by taking the derivative in (2.3) we have

$$
\begin{aligned}
\mathbb{E}(\hat{\alpha}-\alpha) &= \frac{\partial}{\partial\alpha}\int_0^\infty |D_T(-v)|^{-1/2}\mathrm{d}v \\
&= \frac{\partial}{\partial\alpha}\int_1^0 \left[\frac{A_T(x;\alpha)}{(1-\alpha^2x^2)x^{T-1}}\right]^{-1/2}\left[-\frac{1-\alpha^2x^2}{2x^2}\right]dx \\
&= \frac{1}{2}\frac{\partial}{\partial\alpha}\int_0^1 x^{\frac{T-5}{2}}\left(1-\alpha^2x^2\right)^{3/2}A_T(x;\alpha)^{-1/2}dx.
\end{aligned} \tag{2.18}
$$

Evaluating the derivative in (2.18) gives

$$
\begin{aligned}
&\frac{\partial}{\partial\alpha}\int_0^1 x^{\frac{T-5}{2}}\left(1-\alpha^2x^2\right)^{3/2}A_T(x;\alpha)^{-1/2}\mathrm{d}x \\
&\quad= \frac{3}{2}\int_0^1 x^{\frac{T-5}{2}}\left(1-\alpha^2x^2\right)^{1/2}(-2\alpha x^2)A_T(x;\alpha)^{-1/2}\mathrm{d}x \\
&\quad- \frac{1}{2}\int_0^1 x^{\frac{T-5}{2}}\left(1-\alpha^2x^2\right)^{3/2}A_T(x;\alpha)^{-3/2}\frac{\partial}{\partial\alpha}A_T(x;\alpha)\mathrm{d}x.
\end{aligned} \tag{2.19}
$$

The derivative from the last line in (2.19) is

$$
\begin{aligned}
\frac{\partial}{\partial\alpha}A_T(x;\alpha) &= \frac{\partial}{\partial\alpha}\left(1-\alpha^2x+(1-x)\alpha^{2T}x^{2T-1}\right) \\
&= -2\alpha x + 2T(1-x)\alpha^{2T-1}x^{2T-1}.
\end{aligned} \tag{2.20}
$$

Combining (2.18)-(2.20), for $c<0$, corresponding to $|\alpha|<1$, we have

$$
\begin{aligned}
\mathbb{E}(\hat{\alpha}-\alpha) = &-\frac{3\alpha}{2}\int_0^1 x^{\frac{T-1}{2}}\left(1-\alpha^2x^2\right)^{1/2}A_T(x;\alpha)^{-1/2}\mathrm{d}x \\
&+\frac{\alpha}{2}\int_0^1 x^{\frac{T-3}{2}}\left(1-\alpha^2x^2\right)^{3/2}A_T(x;\alpha)^{-3/2}\mathrm{d}x \\
&-\frac{T\alpha^{2T-1}}{2}\int_0^1 x^{\frac{5T-7}{2}}\left(1-\alpha^2x^2\right)^{3/2}A_T(x;\alpha)^{-3/2}(1-x)\mathrm{d}x.
\end{aligned} \tag{2.21}
$$

To evaluate the expectation in (2.21), we start with the first integral. Setting $y = x^{2T-1}$ gives $dy = (2T-1)x^{2T-2}dx = (2T-1)y^{\frac{2T-2}{2T-1}}dx$. Note that $A_T(x;\alpha) = 1-\alpha^2x+O\left(\alpha^{2T}\right)$. By substituting $\alpha = 1+c/k_T$ in the third line of the following derivations and using the facts that $y^{\frac{b}{2T-a}} = 1+\frac{b}{2T-a}\log y + O\left(T^{-2}\right)$, $k_T = o(T)$

and $\frac{k_T}{T}\alpha^{2T} = o(1)$ as $T \to \infty$, the integral becomes

$$\begin{aligned}
&\int_0^1 x^{\frac{T-1}{2}} \left(1-\alpha^2x^2\right)^{1/2} A_T(x;\alpha)^{-1/2}\mathrm{d}x \\
&= \frac{1}{2T-1}\int_0^1 y^{\frac{-3T+3}{4T-2}} \left\{ \frac{1-\alpha^2 y^{\frac{2}{2T-1}}}{1-\alpha^2 y^{\frac{1}{2T-1}} + (1-y^{\frac{1}{2T-1}})y\alpha^{2T}} \right\}^{1/2} \mathrm{d}y \\
&= \frac{1}{2T}\int_0^1 y^{-\frac{3}{4}} \left\{ \frac{1-\left(1+\frac{2c}{k_T}+\frac{c^2}{k_T^2}\right)\left(1+\frac{2}{2T}\log y + O\left(T^{-2}\right)\right)}{1-\left(1+\frac{2c}{k_T}+\frac{c^2}{k_T^2}\right)\left(1+\frac{1}{2T}\log y + O(T^{-2})\right) + O(\alpha^{2T})} \right\}^{1/2} \mathrm{d}y \\
&\qquad\qquad \times \left[1+O\left(T^{-1}\right)\right] \\
&= \frac{1}{2T}\int_0^1 y^{-\frac{3}{4}} \left\{ \frac{-\frac{2\log y}{2T}-\frac{2c}{k_T}-\frac{c^2}{k_T^2}}{-\frac{\log y}{T}-\frac{2c}{k_T}-\frac{c^2}{k_T^2}} \right\}^{1/2} \mathrm{d}y \left[1+O\left(T^{-1}\right)\right] \\
&= \frac{1}{2T}\int_0^1 y^{-\frac{3}{4}} \left\{ \frac{\frac{k_T}{T}2\log y + 4c + \frac{2c^2}{k_T}}{\frac{k_T}{T}\log y + 4c + \frac{2c^2}{k_T}} \right\}^{1/2} \mathrm{d}y \left[1+O\left(T^{-1}\right)\right] \\
&= \frac{1}{2T}\int_0^1 y^{-\frac{3}{4}} \left\{ \frac{4c\left[1+O\left(\max\left\{\frac{k_T}{T}, k_T^{-1}\right\}\right)\right]}{4c\left[1+O\left(\max\left\{\frac{k_T}{T}, k_T^{-1}\right\}\right)\right]} \right\}^{1/2} \mathrm{d}y \times \left[1+O\left(T^{-1}\right)\right] \\
&= \frac{1}{2T}\int_0^1 y^{-\frac{3}{4}}\mathrm{d}y \left\{1+O\left(\max\left(\frac{k_T}{T}, k_T^{-1}\right)\right)\right\} \left[1+O\left(T^{-1}\right)\right].
\end{aligned}$$

The second integral can be dealt with in the same fashion

$$\begin{aligned}
&\int_0^1 x^{\frac{T-3}{2}} \left(1-\alpha^2x^2\right)^{3/2} A_T(x;\alpha)^{-3/2}\mathrm{d}x \\
&= \frac{1}{2T}\int_0^1 y^{-\frac{3}{4}} \left\{ \frac{\frac{k_T}{T}2\log y + 4c + \frac{2c^2}{k_T}}{\frac{k_T}{T}\log y + 4c + \frac{2c^2}{k_T} + \frac{k_T}{T}\alpha^{2T}y\log y} \right\}^{3/2} \mathrm{d}y \left[1+O\left(T^{-1}\right)\right] \\
&= \frac{1}{2T}\int_0^1 y^{-\frac{3}{4}}\mathrm{d}y \left\{1+O\left(\max\left(\frac{k_T}{T}, k_T^{-1}\right)\right)\right\} \left[1+O\left(T^{-1}\right)\right].
\end{aligned}$$

The third integral becomes exponentially small as $T\alpha^{2T-1} = o(1)$ as $T \to \infty$. By combining the three integrals, for $c < 0$, we have

$$\begin{aligned}
\mathbb{E}(\hat{\alpha}-\alpha) &= \left[-\frac{3\alpha}{2}\frac{1}{2T}\int_0^1 y^{-\frac{3}{4}}dy + \frac{\alpha}{2}\frac{1}{2T}\int_0^1 y^{-\frac{3}{4}}dy\right] \left\{1+O\left(\max\left(\frac{k_T}{T}, k_T^{-1}\right)\right)\right\} \\
&\qquad \times \left[1+O\left(T^{-1}\right)\right] \\
&= -\frac{2\alpha}{T}\left\{1+O\left(\max\left(\frac{k_T}{T}, k_T^{-1}\right)\right)\right\} \left[1+O\left(T^{-1}\right)\right]. \qquad (2.22)
\end{aligned}$$

For $c = 0$, corresponding to $|\alpha| = 1$, the algebra reduces to the one found in Phillips (2012). From (2.19) as $T \to \infty$ we have

$$\begin{aligned}
\mathbb{E}(\hat{\alpha} - 1) = &-\frac{3}{2}\int_0^1 x^{\frac{T-1}{2}}\left\{\frac{1+x}{1+x^{2T-1}}\right\}^{1/2}\mathrm{d}x \\
&+\frac{1}{2}\int_0^1 x^{\frac{T-3}{2}}\left\{\frac{1+x}{1+x^{2T-1}}\right\}^{3/2}\mathrm{d}x \\
&-\frac{T}{2}\int_0^1 x^{\frac{5T-7}{2}}\left\{\frac{1+x}{1+x^{2T-1}}\right\}^{3/2}(1-x)\mathrm{d}x.
\end{aligned}$$

By setting $y = x^{2T-1}$ the first integral becomes

$$\begin{aligned}
\int_0^1 x^{\frac{T-1}{2}}\left\{\frac{1+x}{1+x^{2T-1}}\right\}^{1/2}\mathrm{d}x &= \frac{1}{2T}\int_0^1 y^{-\frac{3}{4}}\left\{\frac{1+y^{\frac{1}{2T-1}}}{1+y}\right\}^{1/2}\mathrm{d}y\left[1+O\left(T^{-1}\right)\right] \\
&= \frac{2^{1/2}}{2T}\int_0^1 y^{-\frac{3}{4}}\{1+y\}^{-1/2}\mathrm{d}y\left[1+O\left(T^{-1}\right)\right].
\end{aligned}$$

The second and third integrals can be dealt in the same fashion. The second becomes

$$\int_0^1 x^{\frac{T-3}{2}}\left\{\frac{1+x}{1+x^{2T-1}}\right\}^{3/2}\mathrm{d}x = \frac{2^{3/2}}{2T}\int_0^1 y^{-3/4}\{1+y\}^{-3/2}\mathrm{d}y\left[1+O\left(T^{-1}\right)\right],$$

and the third is given by

$$\int_0^1 x^{\frac{5T-7}{2}}\left\{\frac{1+x}{1+x^{2T-1}}\right\}^{3/2}(1-x)\mathrm{d}x = -\frac{2^{3/2}}{4T^2}\int_0^1 y^{\frac{1}{4}}\{1+y\}^{-3/2}\log y\mathrm{d}y\left[1+O\left(T^{-1}\right)\right].$$

Combining the three integrals yields

$$\begin{aligned}
\mathbb{E}(\hat{\alpha} - 1) = &-\frac{3}{2}\frac{2^{1/2}}{2T}\int_0^1 y^{-\frac{3}{4}}\{1+y\}^{-1/2}\mathrm{d}y\left[1+O\left(T^{-1}\right)\right] \\
&+\frac{1}{2}\frac{2^{3/2}}{2T}\int_0^1 y^{-3/4}\{1+y\}^{-3/2}\mathrm{d}y\left[1+O\left(T^{-1}\right)\right] \\
&-\frac{1}{2}\left(-\frac{2^{3/2}}{4T}\int_0^1 y^{\frac{1}{4}}\{1+y\}^{-3/2}\log y\mathrm{d}y\left[1+O\left(T^{-1}\right)\right]\right).
\end{aligned}$$

Evaluating the expression numerically leads to the following approximation

$$\mathbb{E}(\hat{\alpha}-1) = -\frac{1.7814}{T} + O\left(T^{-2}\right). \tag{2.23}$$

Finally, consider the case in which $c > 0$, corresponding to $|\alpha| > 1$. From (2.16) and (2.17) we have

$$|D_T(-v)| = \frac{\alpha^2 x - 1 + (x-1)\alpha^{2T} x^{2T-1}}{(\alpha^2 x^2 - 1)x^{T-1}} = \frac{B_T(x;\alpha)}{(\alpha^2 x^2 - 1)x^{T-1}}$$

From (2.3) and (2.16), we have, for $c > 0$,

$$\begin{aligned}
\mathbb{E}(\hat{\alpha}-\alpha) &= \frac{\partial}{\partial\alpha}\int_0^\infty |D_T(-v)|^{-1/2}\mathrm{d}v \\
&= \frac{\partial}{\partial\alpha}\int_1^\infty \left[\frac{B_T(x;\alpha)}{(\alpha^2x^2-1)x^{T-1}}\right]^{-1/2}\left[\frac{\alpha^2x^2-1}{2x^2}\right]\mathrm{d}x \\
&= \frac{1}{2}\frac{\partial}{\partial\alpha}\int_1^\infty x^{\frac{T-5}{2}}\left(\alpha^2x^2-1\right)^{3/2} B_T(x;\alpha)^{-1/2}\mathrm{d}x.
\end{aligned} \tag{2.24}$$

Evaluating the derivative in (2.24) gives

$$\begin{aligned}
&\frac{\partial}{\partial\alpha}\int_1^\infty x^{\frac{T-5}{2}}\left(\alpha^2x^2-1\right)^{3/2} B_T(x;\alpha)^{-1/2}\mathrm{d}x \\
&\quad= \frac{3}{2}\int_1^\infty x^{\frac{T-1}{2}}\left(\alpha^2x^2-1\right)^{3/2} B_T(x;\alpha)^{-1/2}2\alpha x^2\mathrm{d}x \\
&\quad- \frac{1}{2}\int_1^\infty x^{\frac{T-5}{2}}\left(\alpha^2x^2-1\right)^{3/2} B_T(x;\alpha)^{-3/2}\frac{\partial}{\partial\alpha}B_T(x;\alpha)\mathrm{d}x.
\end{aligned} \tag{2.25}$$

The derivative from the last line in (2.25) is

$$\frac{\partial}{\partial\alpha}B_T(x;\alpha) = 2\alpha x + 2T(x-1)\alpha^{2T-1}x^{2T-1}. \tag{2.26}$$

Combining (2.24)-(2.26), for $c > 0$ we have

$$\begin{aligned}
\mathbb{E}(\hat{\alpha}-\alpha) &= \frac{3\alpha}{2}\int_1^\infty x^{\frac{T-1}{2}}\left(\alpha^2x^2-1\right)^{1/2} B_T(x;\alpha)^{-1/2}\mathrm{d}x \\
&- \frac{\alpha}{2}\int_1^\infty x^{\frac{T-3}{2}}\left(\alpha^2x^2-1\right)^{3/2} B_T(x;\alpha)^{-3/2}\mathrm{d}x \\
&- \frac{T\alpha^{2T-1}}{2}\int_1^\infty x^{\frac{5T-7}{2}}\left(\alpha^2x^2-1\right)^{3/2} B_T(x;\alpha)^{-3/2}(x-1)\mathrm{d}x,
\end{aligned} \tag{2.27}$$

The first and second integrals in (2.27) become exponentially small since $\alpha^{2T} = e^{\frac{2cT}{k_T}}\{1+o(1)\}$ explodes as $T \to \infty$. One can show that result with an argument similar to the one used in (A.17). As a result, the bias for $c > 0$ is determined by the third integral only which, as $T \to \infty$, becomes

$$
\begin{aligned}
&-\frac{T\alpha^{2T-1}}{2}\int_1^\infty x^{\frac{5T-7}{2}}\left(\alpha^2x^2-1\right)^{3/2}B_T(x;\alpha)^{-3/2}(x-1)\mathrm{d}x\\
&=-\frac{T\alpha^{2T-1}}{2(2T-1)}\int_1^\infty y^{\frac{5T-7-4(T-1)}{2(2T-1)}}\left\{\frac{\alpha^2y^{\frac{2}{2T-1}}-1}{\alpha^2y^{\frac{1}{2T-1}}-1+\left(y^{\frac{1}{2T-1}}-1\right)y\alpha^{2T}}\right\}^{3/2}\\
&\times\left(y^{\frac{1}{2T-1}}-1\right)\mathrm{d}y\\
&=-\frac{T\alpha^{2T-1}}{2(2T-1)}\int_1^\infty y^{\frac{1}{4}}\left\{\frac{\alpha^2y^{\frac{2}{2T-1}}-1}{\alpha^{2T}\left(\alpha^{-2T}\left(\alpha^2y^{\frac{1}{2T-1}}-1\right)+\left(y^{\frac{1}{2T-1}}-1\right)y\right)}\right\}^{3/2}\\
&\times\left(y^{\frac{1}{2T-1}}-1\right)\mathrm{d}y\left[1+O\left(T^{-1}\right)\right]\\
&=-\frac{T\alpha^{-T-1}}{2(2T-1)^2}\int_1^\infty y^{\frac{1}{4}}\left\{\frac{\left(1+\frac{2c}{k_T}+\frac{c^2}{k_T^2}\right)\left(1+\frac{2}{2T-1}\log y\right)-1}{\frac{1}{2T-1}y\log y}\right\}^{3/2}\\
&\times\log y\mathrm{d}y\left[1+O\left(T^{-1}\right)\right]\\
&=-\frac{T\alpha^{-T-1}}{2(2T-1)^2}\int_1^\infty y^{\frac{1}{4}}\left\{\frac{(2T-1)\left(\frac{2}{2T-1}\log y+\frac{2c}{k_T}+\frac{c^2}{k_T^2}\right)}{y\log y}\right\}^{3/2}\log y\mathrm{d}y\left[1+O\left(T^{-1}\right)\right]\\
&=-\frac{T\alpha^{-T-1}}{2(2T-1)^{1/2}k_T^{3/2}}\int_1^\infty y^{\frac{1}{4}}\left\{\frac{2c\left(1+O\left(\max\left(\frac{k_T}{T},k_T^{-1}\right)\right)\right)}{y\log y}\right\}^{3/2}\\
&\times\log y\mathrm{d}y\left[1+O\left(T^{-1}\right)\right]\\
&=-\frac{T\alpha^{-T-1}(2c)^{3/2}}{2(2T-1)^{1/2}k_T^{3/2}}\int_1^\infty y^{\frac{1}{4}}\{y\log y\}^{-3/2}\log y\mathrm{d}y\left\{1+O\left(\max\left(\frac{k_T}{T},k_T^{-1}\right)\right)\right\}\\
&\times\left[1+O\left(T^{-1}\right)\right]\\
&=-\frac{c^{3/2}T^{1/2}}{k_T^{3/2}\alpha^{T+1}}\int_1^\infty y^{\frac{1}{4}}\{y\log y\}^{-3/2}\log y\mathrm{d}y\left\{1+O\left(\max\left(\frac{k_T}{T},k_T^{-1}\right)\right)\right\}\left[1+O\left(T^{-1}\right)\right].
\end{aligned}
$$

Finally, setting $\log y = w$ such that $dy = e^w dw$ yields

$$
\begin{aligned}
&-\frac{c^{3/2}T^{1/2}}{k_T^{3/2}\alpha^{T+1}}\int_1^\infty y^{\frac{1}{4}}\{y\log y\}^{-3/2}\log y\mathrm{d}y\left\{1+O\left(\max\left(\frac{k_T}{T},k_T^{-1}\right)\right)\right\}\left[1+O\left(T^{-1}\right)\right]\\
&=-\frac{c^{3/2}T^{1/2}}{k_T^{3/2}\alpha^{T+1}}\int_0^\infty e^{\frac{5}{4}w}\{we^w\}^{-3/2}w\mathrm{d}w\left\{1+O\left(\max\left(\frac{k_T}{T},k_T^{-1}\right)\right)\right\}\left[1+O\left(T^{-1}\right)\right]\\
&=-\frac{c^{3/2}T^{1/2}}{k_T^{3/2}\alpha^{T+1}}\int_0^\infty e^{-\frac{1}{4}w}w^{-1/2}\mathrm{d}w\left\{1+O\left(\max\left(\frac{k_T}{T},k_T^{-1}\right)\right)\right\}\left[1+O\left(T^{-1}\right)\right]\\
&=-\frac{2\pi^{1/2}c^{3/2}T^{1/2}}{k_T^{3/2}\alpha^{T+1}}\left\{1+O\left(\max\left(\frac{k_T}{T},k_T^{-1}\right)\right)\right\}\left[1+O\left(T^{-1}\right)\right]. \qquad (2.28)
\end{aligned}
$$

The three results of the bias from (2.22), (2.23) and (2.28) are collected in the Theorem. □

Proof of Lemma 2.2.3 We take the derivative directly yielding

$$
\begin{aligned}
g'(\alpha) &= \left(\frac{\pi T}{2}\right)^{1/2}\frac{(3/2)(\alpha^2-1)^{1/2}2\alpha\alpha^{T+1}-(T+1)\alpha^T(\alpha^2-1)^{3/2}}{\alpha^{2(T+1)}}\\
&= \left(\frac{\pi T}{2}\right)^{1/2}\frac{3\alpha(\alpha^2-1)^{1/2}-(T+1)\alpha^{-1}(\alpha^2-1)^{3/2}}{\alpha^{T+1}}\\
&= O\left(T^{3/2}\alpha^{-(T+2)}\right)=o(1),
\end{aligned}
$$

and

$$
\begin{aligned}
h'(\alpha_T) &= 2(\pi T)^{1/2}\frac{(3/2)(\alpha_T-1)^{1/2}\alpha_T^{T+1}-(T+1)\alpha_T^T(\alpha_T-1)^{3/2}}{\alpha_T^{2(T+1)}}\\
&= (\pi T)^{1/2}\frac{3(\alpha_T-1)^{1/2}-2(\alpha_T-1)^{3/2}(T+1)\alpha_T^{-1}}{\alpha_T^{T+1}}\\
&= (\pi T)^{1/2}\frac{3(c/k_T)^{1/2}\alpha_T-2(c/k_T)^{3/2}(T+1)}{\alpha_T^{T+2}}. \qquad (2.29)
\end{aligned}
$$

Now, the second term in the fraction above dominates the expression. Thus, it would suffice to consider only that term and show it converges to zero, from which it will follow that the entire expression converges to zero. To do this, we first derive an asymptotically equivalent version of $\alpha_T^{-(T+2)}$. By utilising the asymptotic equivalence $\log(1+x) = x + O(x^2)$ as $x \to 0$, we have as $T \to \infty$

$$\begin{aligned}
\alpha_T^{-(T+2)} = \left(1 + \frac{c}{k_T}\right)^{-(T+2)} &= \exp\left(-(T+2)\log\left(1+\frac{c}{k_T}\right)\right) \\
&= \exp\left(-(T+2)\left(\frac{c}{k_T} + O\left(k_T^{-2}\right)\right)\right) \\
&= \exp\left(-\frac{c(T+2)}{k_T}\left(1+o(1)\right)\right) \\
&= \exp\left(-\frac{cT}{k_T}(1+o(1))\right). \qquad (2.30)
\end{aligned}$$

Now, showing that (2.29) converges to zero is a straightforward consequence of (2.30). We write

$$\frac{(\pi T)^{1/2}(c/k_T)^{3/2}(T+1)}{\alpha_T^{T+2}} \sim \frac{2\pi^{1/2}c^{3/2}(T/k_T)^{3/2}}{e^{cT/k_T}} = o(1),$$

due to $T/k_t \to \infty$ completing the proof. $\square$

Bibliography

[1] Andrews, D.W.K. and P. Guggenberger (2008) Asymptotics for stationary very nearly unit root processes, *Journal of Time Series Analysis*, 23, 203-12.

[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, 1013-21.

[3] Chambers, M.J. (2013) Jackknife estimation of stationary autoregressive models, *Journal of Econometrics*, 172, 142-57.

[4] Chambers, M.J. and M. Kyriacou (2013) Jackknife estimation with a unit root, *Statistics and Probability Letter*, 83, 1677-82.

[5] Chambers, M.J. and Kyriacou, M. (2018) Jackknife Bias Reduction in the Presence of a Near-Unit Root, *Econometrics*, 6, 11.

[6] Cramér, H. (1946) *Mathematical methods of statistics*, Princeton University Press.

[7] Figuerola-Ferretti, I., Gilbert, C. and J.R. McCrorie (2015) Testing for mild explosivity and bubbles in LME non-ferrous metals prices, *Journal of Time Series Analysis*, 36, 763-82.

[8] Giraitis, L. and P.C.B. Phillips (2006) Uniform limit theory for stationary autoregression, *Journal of Time Series Analysis*, 27, 51-60.

[9] Hamilton, J.D. (1994) *Time series analysis*, Princeton University Press.

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

[11] Kruse, R. and H. Kaufmann (2015) Bias-corrected estimation for speculative bubbles in stock prices, *Economic Modelling*, 73, 354-364.

[12] Le Breton, A. and D.T. Pham (1989) On the bias of the least squares estimator for the first order autoregressive process, *Annals of the Institute of Statistical Mathematics* 41, 555-63.

[13] MacKinnon, J.G. and A.A. Smith (1998) Approximate bias correction in econometrics, *Journal of Econometrics*, 85, 205-30.

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

[15] Phillips, P.C.B. (1987b) Towards a unified asymptotic theory for autoregression, *Biometrika*, 74, 535-47.

[16] Phillips, P.C.B. (2012) Folklore theorems, implicit maps and indirect inference, *Econometrica* 80, 425-54.

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

[18] Phillips, P.C.B, Magdalinos, T. and L. Giraitis (2010) Smoothing local-to-moderate unit root theory, *Journal of Econometrics* 158, 274-79.

[19] Pillips, P.C.B, W. Yangru and Y. Jun (2011) Explosive behaviour in the 1990s NASDAQ: When did exuberance escalate asset values?, *International Economic Review* 52, 201-26.

[20] Shaman, P. and R.A. Stine (1988) The bias of autoregressive coefficient estimators, *Journal of the American Statistical Association* 83, 842-48.

[21] Shenton, L.D. and W.L. Johnson (1965) Moments of a serial correlation coefficient, *Journal of the Royal Statistical Society, Series B* 27, 308-20.

[22] Stoykov, M.Z. (2018) Least squares bias in autoregressive time series with moderate deviations from a unit root, *Journal of Time Series Analysis*, Forthcoming.

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

[24] White, J.S. (1958) The limiting distribution of the serial correlation coefficient in the explosive case, *Annals of Mathematical Statistics* 29, 1188-97.

[25] White, J.S. (1961) Asymptotic expansion for the mean and variance of the serial correlation coefficient, *Biometrika* 48, 85-94.

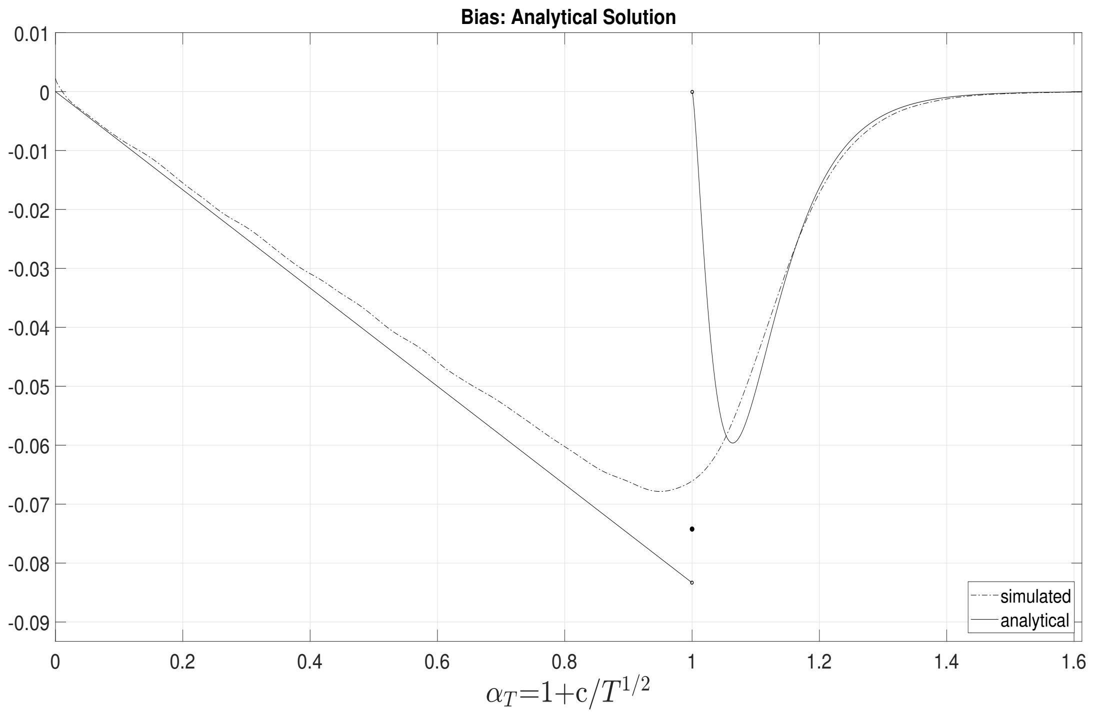

Figure 2.1: Bias: Analytical Solution, $T = 24$.

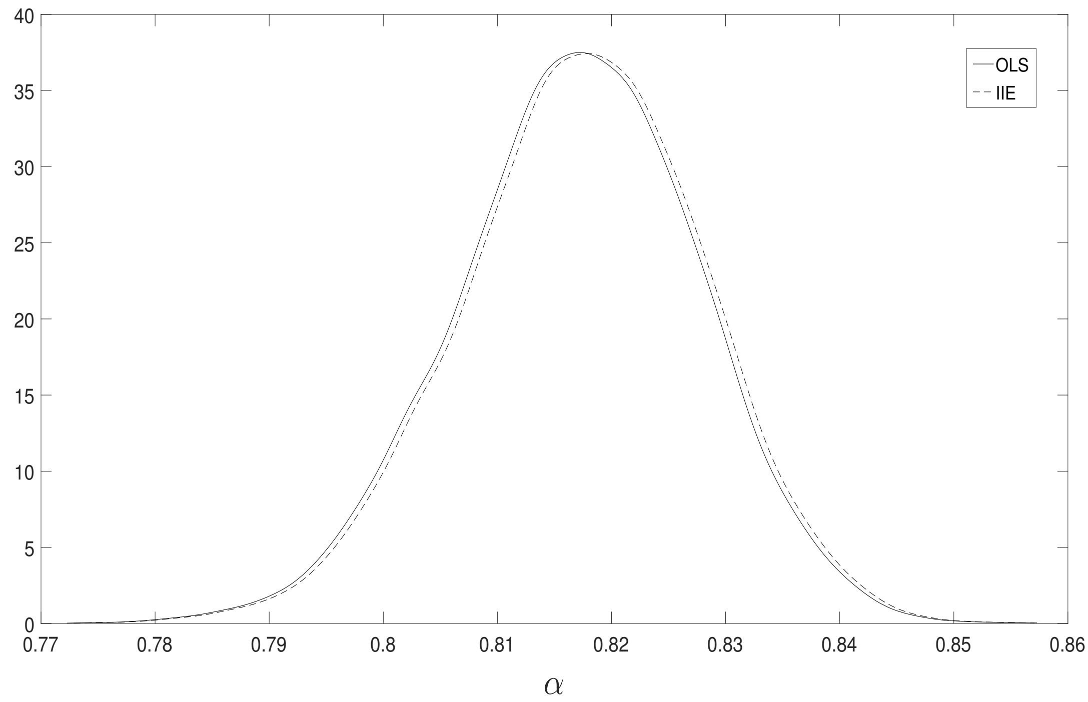

Figure 2.2: Non-parametric estimates of OLS and II estimators, $T = 3000$.

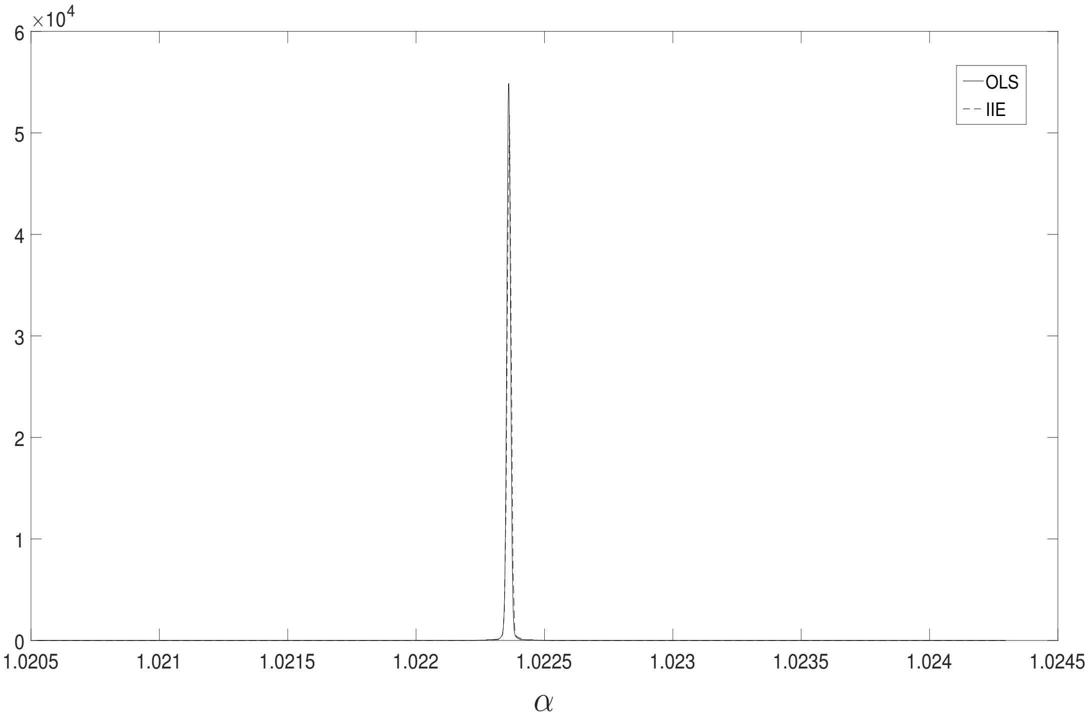

Figure 2.3: Non-parametric estimates of OLS and II estimators, $T = 500$.

T	$\hat{\alpha}_T$	$\tilde{\alpha}_T$	$\hat{\alpha}_J$	$\breve{\alpha}_T$
$c = -5$				
30	-0.0055	0.0003	-0.0009	-0.0001
60	-0.0124	-0.0006	-0.0021	-0.0010
90	-0.0097	0.0008	-0.0005	0.0006
120	-0.0085	0.0005	0.0002	0.0004
$c = -4$				
30	-0.0169	0.0011	-0.0030	-0.0001
60	-0.0163	-0.0001	-0.0024	-0.0007
90	-0.0125	0.0004	-0.0011	0.0001
120	-0.0097	0.0009	0.0004	0.0008
$c = -3$				
30	-0.0286	0.0017	-0.0054	-0.0003
60	-0.0198	0.0007	-0.0022	-0.0000
90	-0.0148	0.0004	-0.0008	0.0001
120	-0.0119	0.0002	-0.0008	-0.0000
$c = -2$				
30	-0.0407	0.0018	-0.0095	-0.0011
60	-0.0234	0.0014	-0.0020	0.0005
90	-0.0170	0.0005	-0.0013	0.0001
120	-0.0134	0.0002	-0.0007	0.0000
$c = -1$				
30	-0.0507	0.0041	-0.0150	0.0005
60	-0.0285	0.0005	-0.0060	-0.0005
90	-0.0191	0.0008	-0.0022	0.0004
120	-0.0148	0.0003	-0.0014	0.0001

Table 2.1: Bias of OLS and jackknife estimators.

Chapter 3

Optimal Jackknife Estimation of Local to Unit Root Models

Abstract

This chapter 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 between each of the estimators have been derived to compute the asymptotic moments. The numerical results from the moments are used to explain why previous simulation studies find that utilising a higher number of sub-samples produces smaller variance. Furthermore, simulation results demonstrate the excellent bias reducing performance of the estimator and its lower variance in comparison with rival jackknives. 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 chapter proposes a two-step "optimal" jackknife estimator to overcome the issue. Simulations outcome is encouraging for the usage of the two-step estimator.

3.1 Introduction

The ordinary least squares (OLS) estimator of the autoregressive coefficient 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 framework that is applicable to all cases is not easy.

One method that could prove useful is the jackknife estimator which is due to Quenouille (1956), with Tukey (1958) showing 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. Essentially, its aim is the removal of a first order bias in the original estimator by a weighted combination between the full-sample and a set of sub-samples estimators. The existence of a first-order bias can be justified by a Nagar-type expansion of the original estimator. Miller (1974a) provides an excellent overview on the jackknife.

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 iid errors when applied to the linear regression model. 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 and found evidence that its performance is superior compared to the two-stage least squares and limited information maximum likelihood estimators, when applied to data with more instruments than endogenous variables. However, the estimator of Angrist et al. has been criticised by Davidson and MacKinnon (2006) who find evidence on its bad performance in the case of weak instruments. Applications of the jackknife estimator to panel data fixed effects have been considered by Hahn and Newey (2004), and Dhaene and Jochmans (2015) extend further to dynamic panel data models. In terms of pure time series, Künsch (1989) used both the bootstrap and the jackknife to estimate standard errors in stationary time series and Phillips and Yu (2005) 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) study the properties of the estimator in non-stationary processes and Chambers and Kyriacou (2018, henceforth CK) consider the local to unit root case. Lastly, Chen and Yu (2015, henceforth CY) propose a more efficient modified version of the CK estimator which applies to unit root processes, which we shall label as an "optimal" jackknife estimator. The aim of this chapter is to extend the literature by constructing an "optimal" jackknife for pure local to unit root processes which could be utilised in applied work. The paper is organised in the following way.

Section 2 introduces the non-overlapping jackknife estimator which Chambers, CK and CY considered, 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 a 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 asymptotic bias which unfortunately has different forms in stationary, non-stationary and explosive series (Le Bretton and Pham (1988) and Theorem 2.2.1). Furthermore, if local to unit root alternatives are considered construction of the weights depends 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 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 a 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 of the numerical results of the moments as the properties of the estimator change in the locally stationary, non-stationary and locally explosive cases. More specifically, in explosive series 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 finite moments. In addition, the asymptotic moments are used to derive the asymptotic variance of the "optimal" jackknife. We provide an analytical derivation of why choosing a higher number of sub-samples produces smaller asymptotic variance, thus explaining findings from previous simulation studies. This feeds through to finite sample settings as subsequent sections show. The results hold regardless of whether the data generated is locally stationary, non-stationary ot locally explosive. These also explain why the "optimal" jackknife produces 10% smaller variance than CK.

Subsequent to all of the preliminaries, section 4 explores the "optimal" local to unit root jackknife estimator and discusses its properties and performance via simulations. The number of replications is set to 10,000, with normally distributed innovations, and comparison is conducted against OLS and rival jackknife estimators. The outcome of the exercise provides evidence in favour of the newly constructed estimator as it provides the same level of bias reduction as the CK bias minimising jackknife but also produces a lower variance.

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 proposes 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 a second stage. Another simulation exercise is performed to consider the features of the newly constructed two-step "optimal" jackknife estimator. For the values of the autoregressive parameter and sample sizes considered the estimator is found to perform excellently. Lastly, section 6 concludes and the appendix depicts all proofs.

For notational simplicity the following notation shall be used throughout. $J_\gamma(r) = \int_0^r e^{(r-s)\gamma} \mathrm{d}W(s)$ denotes the Ornstein-Uhlenbeck (O-U) process which satisfies the following stochastic differential equation $\mathrm{d}J_\gamma(r) = \gamma J_\gamma(r)\mathrm{d}r + \mathrm{d}W(r)$ for some constant parameter γ. The processes $J_\gamma(r)$ and $W(r)$ will be denoted by J and W, respectively, such that functionals of the form $\int_a^b J_\gamma(r)\mathrm{d}J_\gamma(r)$ will be denoted as $\int_a^b J\mathrm{d}J$.

3.2 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 (e.g. Phillips, 2012). Figure 3.1 shows a smoothed graph of the simulated bias for different sample sizes in an autoregressive model without an intercept. The value of the autoregressive coefficient has been plotted on the abscissa and the value of the bias on the ordinate. It can be observed that as the sample size increases the bias diminishes. Furthermore, the bias increases as the value of the autoregressive parameter approaches unity. Figure 3.2 depicts an even more severe bias around unity when an intercept is estimated in the model. Thus, it would be plausible to consider an estimator which could reduce bias for values of the autoregressive coefficient in the vicinity of unity.

The original jackknife (also sometimes referred to as "delete 1" jackknife, see Quenouille, 1956) works only with iid data and is inapplicable in time series settings as removing observations heuristically would affect the correlation structure of the process. One way to try to circumvent this issue is to use non-overlapping blocks of sub-samples. Consider the case in which a researcher is interested in the population parameter θ. Further, assume that they can utilise a full sample estimator which satisfies

$$\mathbb{E}(\hat{\theta}_n) = \theta + \frac{a_1}{n} + \frac{a_2}{n^2} + O\left(n^{-3}\right), \tag{3.1}$$

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

$$\mathbb{E}(\hat{\theta}_j) = \theta + \frac{a_1}{l} + \frac{a_2}{l^2} + O\left(l^{-3}\right). \tag{3.2}$$

Then, the jackknife is constructed as

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

Theorem 1 from Chambers ensures that setting up the weights as $w_1 = m/(m-1)$ and $w_2 = -1/(m-1)$, which we shall label as standard weights and the estimator as C_m, completely removes the first order bias of the estimator in stationary autoregressive settings as long as the expansions in (3.1) and (3.2) exist. Chambers analysed the asymptotic properties of the estimator in stationary AR(p) processes and showed that it has the same asymptotic distribution as OLS. As such, the jackknife is consistent and asymptotically normally distributed. What is more, there is no asymptotic efficiency loss in comparison with OLS. The author also shows, by means of simulations, that the jackknife is robust to different specifications and also performs extremely well in comparison with the median-unbiased estimator and the bootstrap. However, in addition to the excellent bias reduction Chambers also documented the estimator's higher variance when constructed as a bias-minimisation tool.

The above mentioned weights work in stationary cases but not in unit root settings. Chambers and Kyriacou (2013) argue that the constants in (3.2) are no longer identical across each of the sub-samples in the limit and the weights require modifications to take this into account. Consider

$$y_t = \beta y_{t-1} + \epsilon_t, \quad \epsilon_t \sim i.i.d(0, \sigma^2), \quad \beta = 1, \quad t = 1, \ldots, n, \tag{3.3}$$

where $y_0 = 0$. Then, the first-order bias minimising jackknife is defined as

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

with $\kappa^{opt} = -\sum_{j=1}^{m}\mu_j/\bar{\mu}$, $\delta^{opt} = \mu/\bar{\mu}$ and $\bar{\mu} = \mu - \sum_{j=1}^{m}\mu_j$, where μ and μ_j are the respective constants in the asymptotic expansion of the full and each of the sub-sample estimators. Again, simulation results conducted by the authors highlight the outstanding bias reduction performance at the expense of a higher variance for bias-minimising values of m.

To try to mitigate the adverse effects, CY propose to construct an optimisation problem, which aims to minimise the variance of the above mentioned estimator. The new estimator is defined as[1] $\tilde{\beta}_m^{CY} = b_m^{CY}\tilde{\beta}_m - \sum_{j=1}^{m} a_{j,m}^{CY}\tilde{\beta}_j$, where the weights b_m^{CY} and $a_{j,m}^{CY}$ are obtained as the solution to

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

[1]Following CY's notation.

subject to the constraints

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

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 with the first-order bias minimising jackknife.

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 (variances and covariances in addition to the means), which are theoretically incorrect once the autoregressive parameter starts shifting away from unity. To depict the distortions in the CY estimator we have simulated data from (3.3) with normally distributed errors and $10,000$ replications for different values of the autoregressive parameters. The means of the OLS, C_m and CY estimators constructed with $m = 2$ have been plotted on Figures 3.3 and 3.4 for sample sizes $n = 24$ and $n = 192$, respectively. It can be observed that the constant weights of CY do not produced the desired bias reduction across all of the values considered. It is also interesting to note that the constant weights of Chambers' estimator produce the same shape as those of CY in terms of bias reduction. This is not surprising as the jackknife is constructed by utilising weights applied to different estimators. When the weights are constant, the two estimators should have similar shapes. In addition, CY only consider values of m equal to either two or three. It would be interesting to consider the performance of the estimator for bigger values of m as the analytical results from this chapter provide a formal explanation of why choosing a higher number of sub-samples reduces the jackknife's variance.

Lastly, the OLS estimator has the following asymptotic expansion (due to Le Breton and Pham): as $n \to \infty$, for $|\beta| < 1$, $n(\mathbb{E}(\hat{\beta}) - \beta)$ converges to -2β; for $|\beta| = 1$, $n(\mathbb{E}(\hat{\beta}) - \beta)$ converges to -1.7814; for $|\beta| > 1$, $n^{-1/2}|\beta|^n(\mathbb{E}(\hat{\beta}) - \beta)$ 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. In contrast, local to unit root processes have the advantage that their asymptotic expansion satisfies $\mathbb{E}(\hat{\beta}_n) = \beta + O\left(n^{-1}\right)$, as $n \to \infty$ (Phillips (2012)), which is continuous through $\beta = 1$. This provides a framework in which the jackknife could prove useful in locally stationary, non-stationary and locally explosive series.

3.3 Optimal jackknife estimation in local to unit root models

Assume the variable of interest satisfies the following stochastic difference equation

$$y_t = \rho_n y_{t-1} + u_t, \qquad t = 1, \ldots, n, \tag{3.4}$$

where $y_0 = 0$, $\rho_n = e^{\gamma/n} = 1 + \gamma/n + O(n^{-2})$ with γ being a constant and u_t is iid with zero mean and variance σ^2. No assumption will be made regarding the autoregressive coefficient in terms of whether it is smaller, equal or bigger than 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".

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 (1987b) and CK, respectively, and are given by

$$n(\hat{\rho} - \rho_n) \overset{d}{\Rightarrow} \frac{\int_0^1 J\mathrm{d}W}{\int_0^1 J^2},$$
$$l(\hat{\rho}_j - \rho_n) \overset{d}{\Rightarrow} \frac{\int_{(j-1)/m}^{j/m} J\mathrm{d}W}{m\int_{(j-1)/m}^{j/m} J^2}.$$

In this scenario, formulating the usual jackknife with the standard weights fails to completely remove the first order bias as the different sub-samples have different limiting distributions. 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

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

s.t.

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

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 rewrite the optimisation problem in a matrix form, which would also facilitate the simulations which are carried in a subsequent section. We write

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

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

$$\boldsymbol{\Omega} = \begin{bmatrix} \sigma_0^2 & \sigma_{0,1} & \cdots & \sigma_{0,m} \\ \sigma_{0,1} & \sigma_1^2 & \cdots & \sigma_{1,m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{0,m} & \sigma_{1,m} & \cdots & \sigma_m^2 \end{bmatrix},$$

where $\sigma_{i,k}$ $(i, k = 0, \ldots, m)$ denotes the covariance between the ith and kth subsamples, with the subscript 0 denoting the full sample and σ_i^2 denoting their variances. The $(m+1) \times (m+1)$ covariance matrix $\boldsymbol{\Omega}$ is symmetric due to $\sigma_{0,1} = \sigma_{1,0}$. The corresponding Lagrangian is

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

where $\boldsymbol{\lambda} = [\lambda_1, \lambda_2]'$, $\boldsymbol{C} = [\boldsymbol{\nu}'\boldsymbol{\mu}']'$ 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

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

The solution is given by

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

Since $\boldsymbol{\Omega}$ is positive definite and since the constraints are linear, $\boldsymbol{w}^*$ produces a global minimum. It could also be observed that the matrix is partitioned.

It follows that we can decompose its inverse and solve directly for $\boldsymbol{w}^*$ without having to include the lagrange multipliers (the result can be found in Abadir and Magnus (2005), p106). Thus, we can define the weights for the "optimal" jackknife estimator as

$$\begin{aligned}\boldsymbol{w}^* &= \begin{bmatrix}\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\boldsymbol{C}'\left[\boldsymbol{C}\boldsymbol{\Omega}^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C}\boldsymbol{\Omega}^{-1} & \boldsymbol{\Omega}^{-1}\boldsymbol{C}'\left[\boldsymbol{C}\boldsymbol{\Omega}^{-1}\boldsymbol{C}'\right]^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{p}\end{bmatrix}\\ &= \boldsymbol{\Omega}\boldsymbol{C}'[\boldsymbol{C}\boldsymbol{\Omega}^{-1}\boldsymbol{C}']^{-1}\boldsymbol{p}.\end{aligned}$$

We now need to calculate $\boldsymbol{\Omega}$ which contains the variances of each of the estimators and the covariances between them. We employ the procedure of CY and substitute the finite sample expressions with their asymptotic counterparts:

$$\begin{aligned}
n^2Var(\hat{\rho}) &= \mathbb{E}\left(\frac{\int_0^1 J\mathrm{d}W}{\int_0^1 J^2}\right)^2 - \mu_\gamma^2 + o(1);\\
l^2Var(\hat{\rho}_j) &= \mathbb{E}\left(\frac{\int_{(j-1)/m}^{j/m} J\mathrm{d}W}{m\int_{(j-1)/m}^{j/m} J^2}\right)^2 - \mu_{\gamma,j}^2 + o(1),\ \ j = 1,\ldots,m;\\
n^2Cov(\hat{\rho},\hat{\rho}_j) &= \mathbb{E}\left(\frac{\int_0^1 J\mathrm{d}W}{\int_0^1 J^2}\frac{\int_{(j-1)/m}^{j/m} J\mathrm{d}W}{\int_{(j-1)/m}^{j/m} J^2}\right) - m\mu_\gamma\mu_{\gamma,j} + o(1),\ \ 1\le j\le m;\\
n^2Cov(\hat{\rho}_i,\hat{\rho}_j) &= \mathbb{E}\left(\frac{\int_{(i-1)/m}^{i/m} J\mathrm{d}W}{\int_{(i-1)/m}^{i/m} J^2}\frac{\int_{(j-1)/m}^{j/m} J\mathrm{d}W}{\int_{(j-1)/m}^{j/m} J^2}\right) - m^2\mu_{\gamma,i}\mu_{\gamma,j} + o(1),\\
&\qquad 1\le i<j\le m.
\end{aligned}$$

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

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

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

Proposition 3.3.1. *The second derivative of $M_{a,b}(\theta_1,\theta_2)$ with respect to θ_1 is given by*

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

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

The expression above is amenable to 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 $j=1,\ldots,m$. 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 3.3.1 involves terms which are integrated with respect to an O-U process. To overcome this we utilise the procedure of CK. They showed that

$$N(a,b) = \int_a^b J\mathrm{d}J = \int_a^b J\mathrm{d}W + \gamma\int_a^b J^2,$$

from which follows that

$$\int_{(j-1)/m}^{j/m} J\mathrm{d}W = \int_{(j-1)/m}^{j/m} J\mathrm{d}J - \gamma\int_{(j-1)/m}^{j/m} J^2.$$

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

$$
\begin{aligned}
n^2 Var(\hat{\rho}) &= \mathbb{E}\left(\frac{\int_0^1 J\mathrm{d}W}{\int_0^1 J^2}\right)^2 - \mu_\gamma^2 + o(1) \\
&= \mathbb{E}\left(\frac{\int_0^1 J\mathrm{d}J - \gamma\int_0^1 J^2}{\int_0^1 J^2}\right)^2 - \mu_\gamma^2 + o(1) \\
&= \mathbb{E}\left(\frac{\int_0^1 J\mathrm{d}J}{\int_0^1 J^2} - \gamma\right)^2 - \mu_\gamma^2 + o(1) \\
&= \mathbb{E}\left(\frac{\int_0^1 J\mathrm{d}J}{\int_0^1 J^2}\right)^2 - 2\gamma\mathbb{E}\left(\frac{\int_0^1 J\mathrm{d}J}{\int_0^1 J^2}\right) + \gamma^2 - \mu_\gamma^2 + o(1).
\end{aligned}
$$

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

$$
\begin{aligned}
& l^2 Var(\hat{\rho}_j) \\
&= \mathbb{E}\left(\frac{\int_{(j-1)/m}^{j/m} J\mathrm{d}W}{m\int_{(j-1)/m}^{j/m} J^2}\right)^2 - \mu_{\gamma,j}^2 + o(1) \\
&= \mathbb{E}\left(\frac{\int_{(j-1)/m}^{j/m} J\mathrm{d}J}{m\int_{(j-1)/m}^{j/m} J^2}\right)^2 - \frac{2\gamma}{m}\mathbb{E}\left(\frac{\int_{(j-1)/m}^{j/m} J\mathrm{d}J}{m\int_{(j-1)/m}^{j/m} J^2}\right) + \frac{\gamma^2}{m^2} - \mu_{\gamma,j}^2 + o(1),
\end{aligned}
$$

for $j = 1, \ldots, m$. The first terms on the right-hand side can be obtained from Proposition 3.3.1 and the second and fourth from CK. To obtain the covariances, we utilise the same procedure.

Let $M_{a,b,c,d}(\theta_1, \theta_2, \varphi_1, \varphi_2) = \exp\left(\theta_1\int_a^b J\mathrm{d}J + \theta_2\int_a^b J^2 + \varphi_1\int_c^{\mathrm{d}} J\mathrm{d}J + \varphi_2\int_c^{\mathrm{d}} J^2\right)$ 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

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

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

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

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

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

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

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

$$\begin{aligned} n^2 Cov(\hat{\rho}, \hat{\rho}_j) &= \mathbb{E}\left(\frac{\int_0^1 J\mathrm{d}W}{\int_0^1 J^2} \frac{\int_{(j-1)/m}^{j/m} J\mathrm{d}W}{\int_{(j-1)/m}^{j/m} J^2} \right) - m\mu_\gamma \mu_{\gamma,j} + o(1) \\ &= \mathbb{E}\left(\frac{\int_0^1 J\mathrm{d}J}{\int_0^1 J^2} \frac{\int_{(j-1)/m}^{j/m} J\mathrm{d}J}{\int_{(j-1)/m}^{j/m} J^2} \right) - \gamma \mathbb{E}\left(\frac{\int_0^1 J\mathrm{d}J}{\int_0^1 J^2} + \frac{\int_{(j-1)/m}^{j/m} J\mathrm{d}J}{\int_{(j-1)/m}^{j/m} J^2} \right) + \\ &\quad + \gamma^2 - m\mu_\gamma \mu_{\gamma,j} + o(1), \quad j = 1, \ldots, m. \end{aligned}$$

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

$$\begin{aligned} &n^2 Cov(\hat{\rho}_i, \hat{\rho}_j) \\ =&\mathbb{E}\left(\frac{\int_{(i-1)/m}^{i/m} J\mathrm{d}W}{\int_{(i-1)m}^{i/m} J^2} \frac{\int_{(j-1)/m}^{j/m} J\mathrm{d}W}{\int_{(j-1)/m}^{j/m} J^2} \right) - m^2\mu_{\gamma,i} \mu_{\gamma,j} + o(1) \\ &= \mathbb{E}\left(\frac{\int_{(i-1)m}^{i/m} J\mathrm{d}J}{\int_{(i-1)m}^{i/m} J^2} \frac{\int_{(j-1)/m}^{j/m} J\mathrm{d}J}{\int_{(j-1)/m}^{j/m} J^2} \right) - \gamma \mathbb{E}\left(\frac{\int_{(i-1)m}^{i/m} J\mathrm{d}J}{\int_{(i-1)m}^{i/m} J^2} + \frac{\int_{(j-1)/m}^{j/m} J\mathrm{d}J}{\int_{(j-1)/m}^{j/m} J^2} \right) + \\ &\quad + \gamma^2 - m^2\mu_{\gamma,i} \mu_{\gamma,j} + o(1), \quad 1 \leq i < j \leq m. \end{aligned}$$

As in the asymptotic variance cases the first term can be obtained from Theorem 3.3.2 and the second and fourth terms from CK.

Table 3.1 depicts the normalised variances of the full-sample and each of the first six sub-sample estimators (the diagonal entries) and the normalised covariances between them (the off diagonal entries) for different values of γ. 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 sub-sample'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 CK. Furthermore, the variances of each sub-sample are bigger than any of the consecutive sub-samples. In terms of the cross-moments, the covariance between the full-sample and each of the sub-samples increase as the number of the sub-sample used increases, i.e. $\sigma_{0,1} < \sigma_{0,2}$. 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 $\sigma_{1,2} > \sigma_{1,3}$, etc. Some of the entries in the table coincide with results from previous studies, namely CY derived the variances of the full sample and sub-sample estimators for $\gamma = 0$ and $m = 2, 3$, and Hansen (2014) derived the variances for the full-sample for negative values of γ. However, most of the entries are new to the literature.

Now that we have the asymptotic moments, we can construct the weights. Table 3.2 depicts those for different values of γ. As m increases the weight that is applied to the full sample estimator decreases for any γ. For example, when $\gamma = -10$ and $m = 2$, $w^* = 2.1081$, for $m = 12$, $w^* = 1.1357$ and for $m = 1000$, $w^* = 1.0111$. The last result is not reported in Table 3.2. It seems that as m gets large, w^* approaches unity. In other words the jackknife estimator converges to OLS as the number of sub-samples becomes large since the weights sum up to one. Thus the local to unit root optimal jackknife estimator shares some of its feature with the standard jackknife analysed by Chambers.

The results from Proposition 3.3.1 and Theorem 3.3.2 can also be used to calculate an approximation (up to order $O(n^{-1})$) of the asymptotic variance of the jackknife estimator for any fixed m. Table 3.3 contains numerical results of $n^2Var(\hat{\rho}_J)$ and $n^2Var(\hat{\rho}_{CK})$ as $n \to \infty$ by utilising their respective optimal weights. We can observe that the variance decreases as the number of sub-samples increase. What is more, for any number of sub-samples the "optimal" jackknife has smaller variance than CK. For values of γ close to zero the discrepancy is approximately 10%. These asymptotic results translate to finite samples as the following sections depict. Furthermore, Table 3.1 contains the asymptotic variance of OLS and we can make comparison for a number of values of γ. We will focus on the "optimal" jackknife only. For $\gamma = 0$ and $m = 2$ the jackknife has an asymptotic variance bigger than OLS's by almost a factor of two. It takes $m = 6$ to produce an asymptotic variance lower than OLS'. These results are important as previous studies find evidence via simulations that $m = 2$ produces the smallest bias. However, we can see why it also produces the highest variance. Thus, if the researcher is interested in utilising the jackknife as a variance minimiser than choosing a high number of sub-samples seems plausible. Some finite samples simulations not reported here suggest that choosing $m = n/4$, or $l = 4$, produces the smallest variance almost uniformly, which is consistent with those analytical results. Choosing $m = n/2$ would not provide enough observations for estimation within the sub-samples and typically does not perform well. This procedure is going to be utilised for the finite-sample simulations that follow. Under the author's knowledge this is the first attempt to explain analytically why choosing a bigger m produces a smaller variance. Chambers and Kyriacou (2012) provide simulations with similar outcome. Finally, the same analysis cannot currently be applied to explain why $m = 2$ produces the smallest bias as this would require an analytical expression of the second-order term of the bias of OLS as the jackknife removes the first one. Unfortunately, this has not been derived yet for local to unit root models.

3.4 Simulation studies

This section has the aim to investigate the performance of the "optimal" jackknife estimator in different settings via simulations. A comparison will be made with the following estimators: OLS, C_m and CK[2]. Inclusion of the OLS estimator is self-explanatory. The C_m 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

[2]For parsimonious presentation of the results, the two-step estimator, which is the subject of the next section, has also been included in these tables.

(2015) find evidence that those weights perform best in terms of bias reduction for mildly explosive processes in the samples of smallest sizes when compared with 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, C_m is theoretically suboptimal. Comparison with the CK estimator is also self-explanatory since the argument for the "optimal" jackknife estimator is that it performs as well as that of CK's in terms of bias reduction but, in addition, also has a reduced variance. The CY estimator has not been included due to its shortcomings, which were explained in section 2. The comparison will be made in terms of bias and RMSE since the latter has the aim to capture the trade-off between bias reduction and an increase in the variance. Non-jackknife estimators have not been included in the study as Chambers, CK and Kruse and Kaufmann have done extensive simulations to show the superior bias reduction capabilities of the jackknife and we only need be concerned with showing that our estimator produces the same bias reduction and also has a smaller variance in comparison with CK's.

The model considered is the one given by (3.4) with normally distributed error terms with mean zero and a unit variance, i.e. $u_t \sim N(0,1)$, and the number of replications is set at $10,000$. The number of observations considered start with $n = 36$ and are then increased by the same amount for each further scenario considered: $n \in \{36, 72, 108, 144\}$. The sample sizes are 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. The outcome of the simulations is reported in four tables: two for bias-minimising and two for RMSE-minimising values of m. Previous studies find that $m = 2$ produces the smaller bias. As RMSE in these settings is dominated by the variance of the estimators we will choose $m = n/4$ for RMSE-minimisation.

Looking at Tables 3.4 and 3.6, it can be observed that the bias decreases as the sample size increases. Furthermore, it can also be seen that the biases of the CK and "optimal" jackknife are significantly lower in magnitude in comparison to OLS. For the range of values of γ considered, the former two produce practically the same level of bias reduction. Secondly, the CK and "optimal" jackknives outperform C_m on the stationary side. This could be explained by the fact that they remove different first order and leave different second order, and so on, terms from the asymptotic bias. Note that the asymptotic bias on the stationary side is given by $\mathbb{E}(\hat{\rho} - \rho) = -2\rho/n + O(n^{-2})$, which is different from the local to unit root's. Furthermore, the CK and "optimal" jackknives also outperform C_m on the explosive side, which should not come as a surprise as by construction the former two are optimal in this scenario and the latter is not. These result hold regardless of whether we consider the bias or RMSE-minimising values of m.

Tables 3.5 and 3.7 contain a comparison between the RMSEs of the estimators (the subscripts denote the value of m). In each scenario the "optimal" jackknife produces a smaller variance than CK, something the estimator was constructed to do in the first place. In addition, for $m = 2$ the "optimal" jackknife has a slightly higher variance than C_m, which is the cost the estimator pays when constructed as a bias-minimisation tool as explained in section 3. However, for $m = n/4$ it can be observed that for all cases the "optimal" estimator has the smallest RMSE.

3.5 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 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, denoted by $\hat{\gamma}$, 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 $\rho = e^{\gamma/n}$, with γ being a constant, we reverse and solve for γ, which is given by $\gamma = n \log \rho$. The estimator of γ would then be given by $\hat{\gamma} = n \log \hat{\rho}$ and has the following property as $n \to \infty$

$$\begin{aligned}\hat{\gamma} = n \log((\hat{\rho} - \rho_n) + \rho_n) &= n \log\left((\hat{\rho} - \rho_n) + 1 + \gamma/n + O_p(n^{-2})\right) \\ &= n\left(\hat{\rho} - \rho_n\right) + \gamma + O_p(n^{-1}),\end{aligned}$$

Hence, we have that

$$\hat{\gamma} - \gamma = n(\hat{\rho} - \rho_n) + O_p(n^{-1}) \Rightarrow \frac{\int_0^1 J \mathrm{d}W}{\int_0^1 J^2}.$$

Therefore, $\hat{\gamma}$ would underestimate the true γ on average even asymptotically. However, this is not a problem as the purpose of the two-step estimator is to reduce bias in finite sample estimation of ρ rather than γ. We have carried out a simulation with $n = 25^4$, $m = 2$ and $\gamma = 0$ with $10,000$ replications (this sample size is not considered in any of the tables) and the difference between the bias of the absolute terms of the "optimal" and two-stage "optimal" estimators is equal to approximately -0.9×10^{-5}. Thus, even though $\hat{\gamma}$ does not converge in probability to γ as the sample size gets large, we can still utilise the two-step estimator. Furthermore, this construction does not rely on a normality assumption in (3.4).

Tables 3.4-3.7 contain the performance of the two-step "optimal" estimator. Firstly, starting with $m = 2$, for higher absolute values of γ the two-step estimator practically has the same performance as the "optimal" estimator in terms of both bias and RMSE. For absolute values of γ closer to zero, the two-step estimator has higher bias than the "optimal" estimator. It is interesting to note that the two-step estimator produces smaller RMSE than the "optimal" jackknife for some of the values of the parameters considered. Furthermore, the two-step estimator also has smaller bias and slightly higher variance than C_m. For $m = n/4$, the two-step estimator produces bias and RMSE comparable to that of the "optimal" jackknife and always smaller bias than C_m. Furthermore, with the exception of 3 values of n for $\gamma = -10$, and one value of n for $\gamma = -5$, the two-step estimator produces smaller RMSE than C_m. The simulations studies provide evidence that the two-step estimator performs better than C_m for values of the autoregressive coefficient close to unity regardless of whether it is to be used as a bias or RMSE-minimising estimator. In view of Chambers' (2013) work, similar results are expected to hold for departures from normality but have not been pursued here due to the computational time required to conduct the simulations.

3.6 Conclusion

This chapter 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 (2018). The unifying method is applicable to locally stationary, non-stationary and locally explosive series. Construction of the estimator requires the calculation of the variances of the full-sample and each of the sub-samples, and the covariances between all of them. As such, the paper derives their asymptotic counterparts by means of moment generating functions and provides a discussion on some of the moments' features. The results are used to provide a formal explanation of findings from previous simulation studies, namely, why the jackknife has smaller variance when the number of sub-samples utilised increases. 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. To overcome the problem of the weights' dependence on the true parameter generating the data, the chapter also proposes a two-step "optimal" jackknife estimator, the idea of which is to get an estimate of the parameter and use it to construct the weights. Simulation studies show that the there is not much loss in the performance of the two-step estimator in comparison with the theoretical one. What is more, the two step estimator performs better than the standard jackknife for values of the autoregressive coefficient close to unity. The two-step procedure is encouraging and could easily be utilised in applied frameworks. Future areas of research could include unit root testing, with the jackknife developed in this chapter being useful as it offers smaller variance than rival jackknife estimators. Also the estimator has the advantage of being developed under the local-to-unity framework, which is particularly suitable for unit root testing.

3.7 Appendix

Proof of Proposition 3.3.1 Magnus (1986) showed that

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

CK showed that for $\gamma \neq 0$ the MGF of $\frac{\int_a^b J \mathrm{d}J}{\int_a^b J^2}$, 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

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

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 amenable for numerical integration. □

Proof of Theorem 3.3.2. Using the techniques of Chambers and Kyriacou (2012) and Chen and Yu (2015) 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

$$\mathrm{d}J(t) = \gamma J(t)\mathrm{d}t + \mathrm{d}W(t), \quad J(0) = 0$$
$$\mathrm{d}Y(t) = \lambda Y(t)\mathrm{d}t + \mathrm{d}W(t), \quad Y(0) = 0$$

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

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

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

$$\begin{aligned}&M_{0,1,a,b}(\theta_1,\theta_2,\varphi_1,\varphi_2)\\&= \mathbb{E}\left[\exp\left\{\theta_1\int_0^1 J\mathrm{d}J + \theta_2\int_0^1 J^2 + \varphi_1\int_a^b J\mathrm{d}J + \varphi_2\int_a^b J^2\right\}\right].\end{aligned}$$

Then by Girsanov's theorem

$$\begin{aligned}&M_{0,1,a,b}(\theta_1,\theta_2,\varphi_1,\varphi_2)\\&= \mathbb{E}\Bigg[\exp\Bigg\{\theta_1\int_0^1 Y\mathrm{d}Y + \theta_2\int_0^1 Y^2 + \varphi_1\int_a^b Y\mathrm{d}Y + \varphi_2\int_a^b Y^2 + (\gamma-\lambda)\int_0^1 Y\mathrm{d}Y\\&\quad - \frac{\gamma^2-\lambda^2}{2}\int_0^1 Y^2\Bigg\}\Bigg].\end{aligned}$$

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

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

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

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

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)

$$\begin{aligned}
&\mathbb{E}\exp\left(\frac{(\theta_1+\gamma-\lambda)}{2}Y(1)^2\middle| Y(b)\right) \\
&\quad= [1-(\theta_1+\gamma-\lambda)\varpi_b^2]^{-1/2}\exp\left\{\left(\frac{\theta_1+\gamma-\lambda}{2}k_bY(b)^2\right)\right\},
\end{aligned}$$

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

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

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

$$\begin{aligned}
&\mathbb{E}\exp\left\{\left(\frac{\theta_1+\gamma-\lambda}{2}k_b+\frac{\varphi_1}{2}\right)Z(b)^2-\frac{\varphi_1}{2}Z(a)^2+\varphi_2\int_a^b Z^2+(\lambda-\eta)\int_0^b ZdZ\right. \\
&\quad\left.-\frac{\lambda^2-\eta^2}{2}\int_0^b Z^2\right\} \\
&= \exp\left(-\frac{\lambda-\eta}{2}b\right) \\
&\quad\times \mathbb{E}\exp\left\{\left(\frac{\theta_1+\gamma-\lambda}{2}k_b+\frac{\varphi_1}{2}+\frac{\lambda-\eta}{2}\right)Z(b)^2-\frac{\varphi_1}{2}Z(a)^2-\varphi_2\int_0^a Z^2\right\},
\end{aligned}$$

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

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

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

$$\begin{aligned}
&\mathbb{E}\exp\left\{\left(\frac{\theta_1+\gamma-\lambda}{2}k_b+\frac{\varphi_1}{2}+\frac{\lambda-\eta}{2}\right)Z(b)^2\middle|Z(a)\right\} \\
&= \left[1-[(\theta_1+\gamma-\lambda)k_b+\varphi_1+(\lambda-\eta)]\varpi_a^2\right]^{-1/2} \\
&\quad\times\exp\left\{\left[\left(\frac{\theta_1+\gamma-\lambda}{2}\right)k_b+\frac{\varphi_1}{2}+\frac{\lambda-\eta}{2}\right]k_aZ(a)^2\right\},
\end{aligned}$$

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:

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

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

$$
\begin{aligned}
&\mathbb{E}\exp\left\{\left[\left(\frac{\theta_1+\gamma-\lambda}{2}\right)k_b + \frac{\varphi_1}{2} + \frac{\lambda-\eta}{2}\right]k_a X(a)^2 - \varphi_2\int_0^a X^2 \right.\\
&\quad \left. + (\eta-\zeta)\int_0^a X dX - \frac{\eta^2-\zeta^2}{2}\int_0^a X^2\right\}\\
&= \exp\left(-\frac{\eta-\lambda}{2}a\right)\\
&\quad \times \mathbb{E}\exp\left\{\left[\left(\left(\frac{\theta_1+\gamma-\lambda}{2}\right)k_b + \frac{\varphi_1}{2} + \frac{\lambda-\eta}{2}\right)k_a - \frac{\varphi_1}{2} + \frac{\eta-\lambda}{2}\right]X(a)^2\right\},
\end{aligned}
$$

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

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

The MGF thus far is:

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

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

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

Note that

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

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

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

where

$$H_{0,1,a,b}(\theta_1,\theta_2,\varphi_1,\varphi_2) = \left[\cosh(a\lambda) - \frac{p+\eta}{\lambda}\cosh(a\lambda)\right]\left[\cosh(e\lambda) - \frac{\theta_1+\gamma}{\lambda}\sinh(e\lambda)\right]$$
$$\times \left[\cosh(s\eta) - \frac{(\theta_1+\gamma-\lambda)k_b + \varphi_1 + \lambda}{\eta}\sinh(s\eta)\right],$$
$$\zeta = \lambda = \sqrt{\gamma^2 - 2\theta_2}, \eta = \sqrt{\gamma^2 - 2\theta_2 - 2\varphi_2}, e = 1-b, s = b-a,$$
$$\varpi_b^2 = (\exp(2e\lambda)-1)/(2\lambda), \varpi_a^2 = (\exp(2s\eta)-1)/(2\eta), \varpi^2 = (\exp(2a\lambda)-1)/(2\lambda),$$
$$k_a = \left[1 - [(\theta_1+\gamma-\lambda)k_b + \varphi_1 + (\lambda-\eta)]\varpi_a^2\right]^{-1}\exp(2s\eta),$$
$$k_b = \left[1 - (\theta_1+\gamma-\lambda)\varpi_b^2\right]^{-1}\exp(2e\lambda),$$
$$p = [(\theta_1+\gamma-\lambda)k_b + \varphi_1 + (\lambda-\eta)]\,k_a - \varphi_1.$$

Denote $M = M_{0,1,a,b}(\theta_1,\theta_2,\varphi_1,\varphi_2)$ and $H = H_{0,1,a,b}(\theta_1,\theta_2,\varphi_1,\varphi_2)$. Taking the first partial derivative gives

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

where H_0 denotes H evaluated at $\theta=0$. The second partial derivative is given by

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

from which we need

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

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

$$
\begin{aligned}
\mathrm{d}J(t) &= \gamma J(t)\mathrm{d}t + \mathrm{d}W(t), \quad J(0) = 0\\
\mathrm{d}Y(t) &= \lambda Y(t)\mathrm{d}t + \mathrm{d}W(t), \ \ Y(0) = 0
\end{aligned}
$$

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

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

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

$$
\begin{aligned}
&M_{a,b,c,d}(\theta_1, \theta_2, \varphi_1, \varphi_2) \\
&= \exp\left(-\frac{\theta_1}{2}s - \frac{\varphi_1}{2}e - \frac{\gamma - \lambda}{2}d\right) \\
&\quad \times \mathbb{E}\Bigg[\exp\left\{\frac{\theta_1}{2}\left[Y(b)^2 - Y(a)^2\right] + \theta_2 \int_a^b Y^2 + \frac{(\varphi_1 + \gamma - \lambda)}{2}Y(d)^2\right. \\
&\quad \left. - \frac{\varphi_1}{2}Y(c)^2 - \varphi_2 \int_0^c Y^2\right\}\Bigg].
\end{aligned}
$$

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

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

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

$$
\begin{aligned}
&\mathbb{E}\exp\left(\frac{(\theta_1 + \gamma - \lambda)}{2}Y(d)^2 \middle| Y(c)\right) \\
&= \left[1 - (\varphi_1 + \gamma - \lambda)\varpi_c^2\right]^{-1/2} \exp\left(\frac{\varphi_1 + \gamma - \lambda}{2}k_c Y(c)^2\right),
\end{aligned}
$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The MGF becomes

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

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

$$
\begin{aligned}
&\exp(-q\gamma)\left\{[(\varphi_1+\gamma-\lambda)k_c-\varphi_1+\lambda-\gamma]\varpi_b^2\right\}\\
&=\cosh(q\gamma)-\frac{(\varphi_1+\gamma-\lambda)k_c-\varphi_1+\lambda}{\gamma}\sinh(q\gamma)
\end{aligned}
$$

and

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

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

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

Thus, the MGF is given by

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

where

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

Denote $M'=M_{a,b,c,d}(\theta_1,\theta_2,\varphi_1,\varphi_2)$ and $H'=H_{a,b,c,d}(\theta_1,\theta_2,\varphi_1,\varphi_2)$. Taking the first partial derivative gives

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

where H_0' denotes H' evaluated at $\theta = 0$. The second partial derivative is given by

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

from which we require

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

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

Bibliography

[1] Abadir, K.M. and Magnus, J.R. (2005) Matrix algebra, Cambridge University Press.

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

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

[4] Chambers, M.J. and M. Kyriacou (2012) Jackknife Bias Reduction in the Presence of a Near-Unit Root, preprint.

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

[6] Chambers, M.J. and M. Kyriacou (2018) Jackknife Bias Reduction in the Presence of a Near-Unit Root, *Econometrics*, 6, 11.

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

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

[9] Dhaene, G. and K. Jochmans (2015) Split-panel jackknife estimation of fixed-effect models, *The Review of Economic Studies*, 82, 991-1030.

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

[11] Hamilton, J.D. (1994) Time series analysis, Princeton University Press.

[12] Hansen, B.E. (2014) Asymptotic moments of autoregressive estimates with a near unit root and minimax risk, *Advances in Econometrics*, 33, 3-21.

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

[14] Kruse, R. and H. Kaufmann (2015) Bias-corrected estimation in mildly explosive autoregressions, *Conference Paper*.

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

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

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

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

[19] Miller, R.G. (1974b) An unbalanced jackknife, *Annals of Statistics*, 2, 880-891.

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

[21] Phillips, P.C.B. (1987b) Towards a unified asymptotic theory for autoregression, *Biometrika*, 74, 535-547.

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

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

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

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

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

[27] Shao, J. and C.F.J. Wu (1987) Heteroskedasticity-robustness of jackknife variance estimators in linear models, *Annals of Statistics*, 15, 1563-1579.

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

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

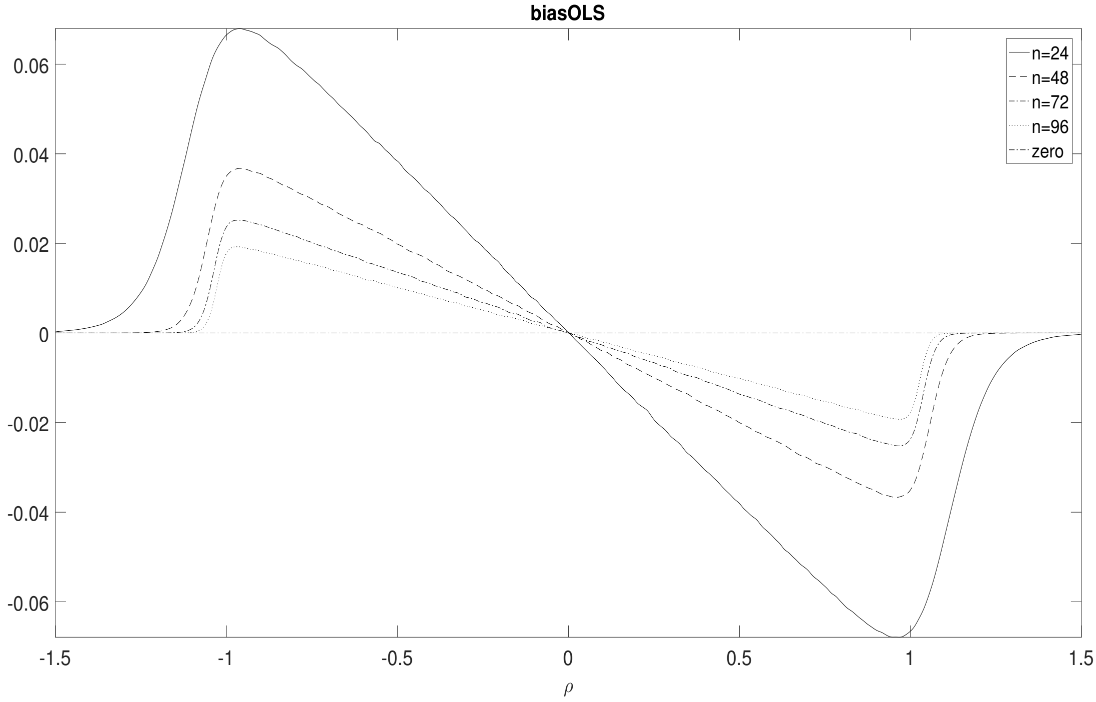

Figure 3.1: Bias of the OLS estimator for different sample sizes in regression without an intercept.

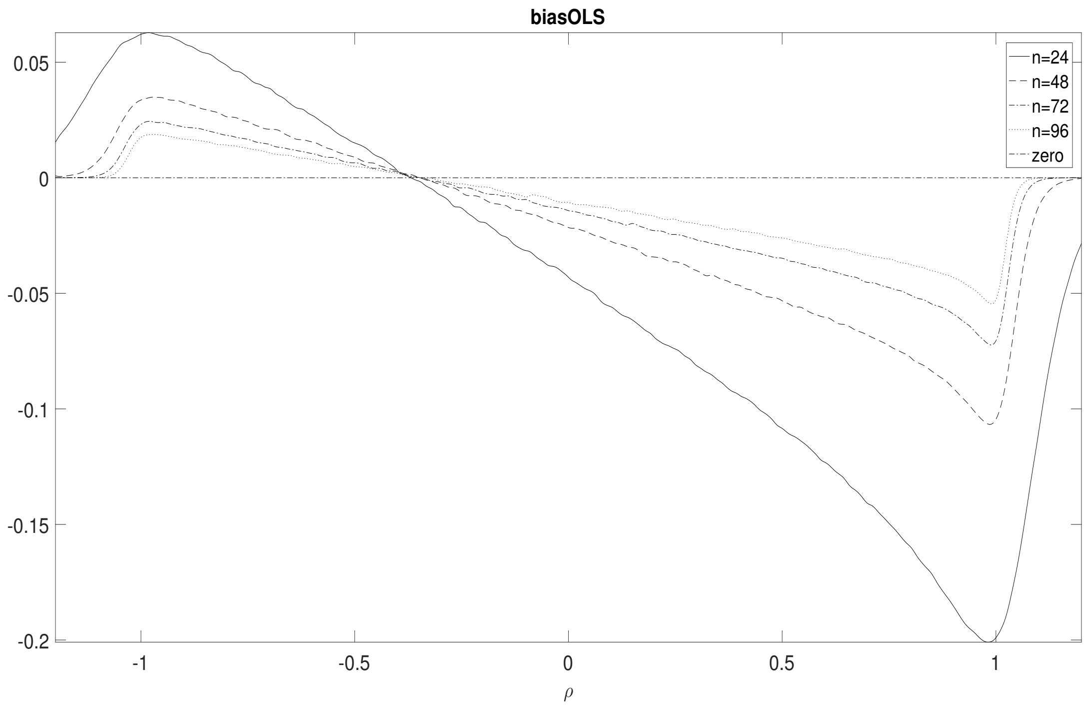

Figure 3.2: Bias of the OLS estimator for different sample sizes in regression with an intercept.

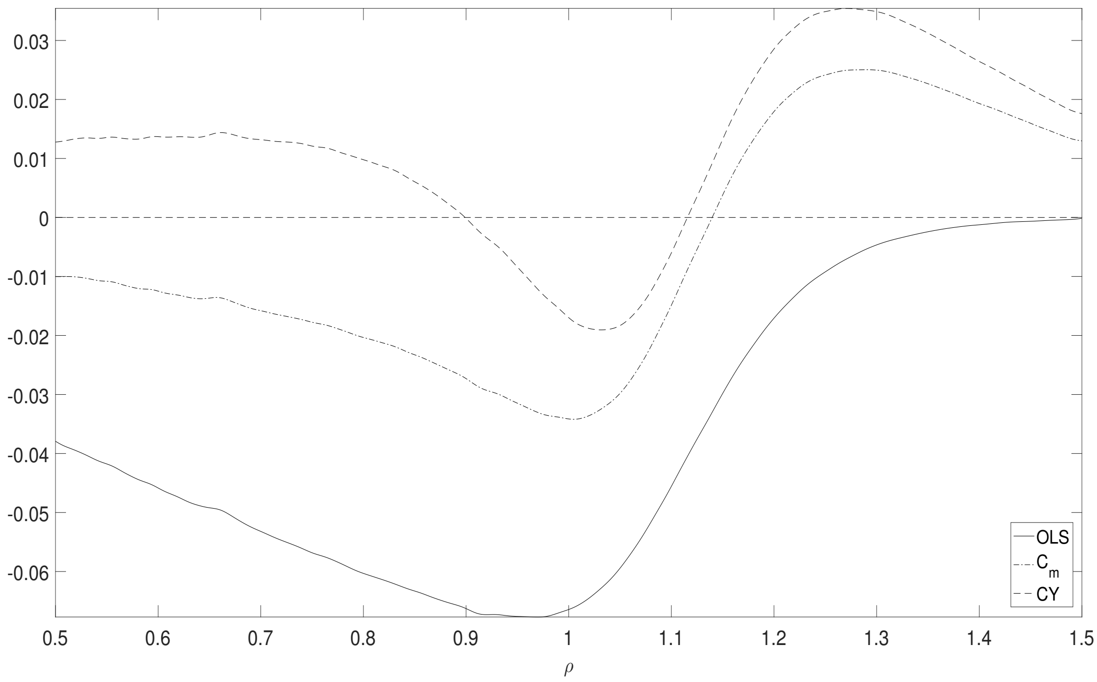

Figure 3.3: Bias of OLS, C_m and CY estimators for $n = 24$.

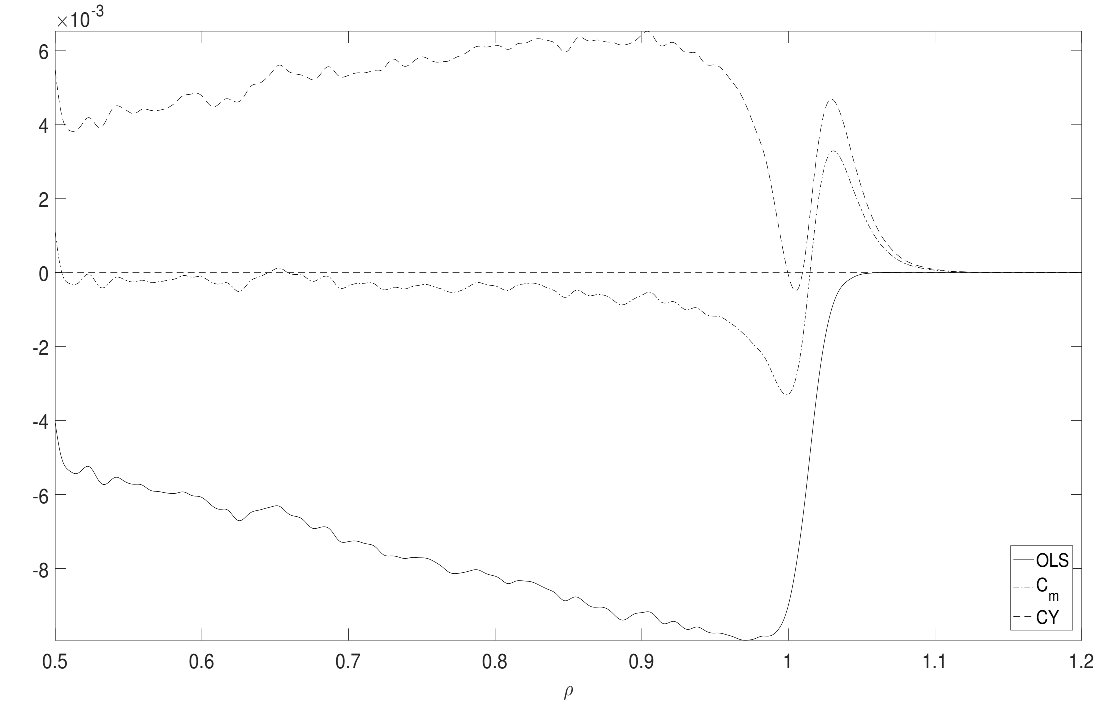

Figure 3.4: Bias of OLS, C_m and CY estimators for $n = 192$.

	$n\hat{\rho}$	$l\hat{\rho}_1$	$l\hat{\rho}_2$	$l\hat{\rho}_3$	$l\hat{\rho}_4$	$l\hat{\rho}_5$	$l\hat{\rho}_6$
$\gamma = -10$							
$n\hat{\rho}$	29.1455	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 3.1: Values of asymptotic variances of full and sub-sample estimators and covariances between them.

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 3.2: Values of weights for the optimal jackknife estimator.

m	$\hat{\rho}_J$	$\hat{\rho}_{CK}$	m	$\hat{\rho}_J$	$\hat{\rho}_{CK}$
$\gamma = -10$					
2	36.9700	37.1197	9	28.4627	28.5664
3	32.5204	32.6191	18	27.2839	27.4429
4	30.9052	30.9944	27	26.7696	26.9859
6	29.4493	29.5396	36	26.5625	26.8512
$\gamma = -5$					
2	27.1401	27.5759	9	18.0669	18.3116
3	22.3865	22.6614	18	16.8454	17.1785
4	20.6593	20.8964	27	16.3183	16.7321
6	19.1088	19.3340	36	16.0590	16.5528
$\gamma = -1$					
2	20.7446	22.6523	9	9.9461	10.8115
3	15.0235	16.1912	18	8.6384	9.5880
4	12.9582	13.9315	27	8.1138	9.1394
6	11.1325	12.0047	36	7.8479	8.9318
$\gamma = 0$					
2	20.1932	22.6009	9	8.3392	9.5570
3	13.8789	15.4483	18	7.0015	8.2754
4	11.5970	12.9499	27	6.4915	7.8184
6	9.6037	10.8407	36	6.2387	7.6003
$\gamma = 1$					
2	19.4208	21.1768	9	7.1144	8.5371
3	12.8999	14.3206	18	5.7850	7.2851
4	10.5046	11.8793	27	5.3006	6.8447
6	8.4167	9.8023	36	5.0620	6.6262

Table 3.3: Values of normalised asymptotic variance for “optimal” jackknife and CK estimators.

n	$\hat{\rho}$	$\hat{\rho}_{J,m}$	$\hat{\rho}_{J,CK}$	$\hat{\rho}^*_J$	$\hat{\rho}^*_{J,2S}$
$\gamma = -10$					
36	-0.0395	-0.0087_2	-0.0055_2	-0.0055_2	-0.0054_2
72	-0.0240	-0.0044_2	-0.0024_2	-0.0025_2	-0.0025_2
108	-0.0162	-0.0019_2	-0.0004_2	-0.0004_2	-0.0004_2
144	-0.0132	-0.0021_2	-0.0010_2	-0.0009_2	-0.0010_2
$\gamma = -5$					
36	-0.0437	-0.0117_2	-0.0055_2	-0.0055_2	-0.0058_2
72	-0.0250	-0.0061_2	-0.0024_2	-0.0026_2	-0.0029_2
108	-0.0166	-0.0033_2	-0.0008_2	-0.0007_2	-0.0009_2
144	-0.0129	-0.0025_2	-0.0005_2	-0.0006_2	-0.0008_2
$\gamma = -1$					
36	-0.0478	-0.0204_2	-0.0078_2	-0.0079_2	-0.0127_2
72	-0.0251	-0.0091_2	-0.0017_2	-0.0018_2	-0.0047_2
108	-0.0169	-0.0060_2	-0.0010_2	-0.0010_2	-0.0030_2
144	-0.0127	-0.0044_2	-0.0006_2	-0.0008_2	-0.0022_2
$\gamma = -0.1$					
36	-0.0464	-0.0220_2	-0.0084_2	-0.0087_2	-0.0152_2
72	-0.0248	-0.0102_2	-0.0022_2	-0.0022_2	-0.0062_2
108	-0.0163	-0.0065_2	-0.0010_2	-0.0010_2	-0.0037_2
144	-0.0125	-0.0051_2	-0.0010_2	-0.0009_2	-0.0030_2
$\gamma = 0$					
36	-0.0468	-0.0229_2	-0.0093_2	-0.0098_2	-0.0162_2
72	-0.0232	-0.0091_2	-0.0011_2	-0.0013_2	-0.0051_2
108	-0.0165	-0.0071_2	-0.0018_2	-0.0015_2	-0.0044_2
144	-0.0116	-0.0041_2	0.0001_2	-0.0000_2	-0.0020_2
$\gamma = 0.1$					
36	-0.0453	-0.0220_2	-0.0086_2	-0.0091_2	-0.0156_2
72	-0.0238	-0.0100_2	-0.0020_2	-0.0023_2	-0.0062_2
108	-0.0159	-0.0064_2	-0.0009_2	-0.0010_2	-0.0036_2
144	-0.0122	-0.0050_2	-0.0008_2	-0.0009_2	-0.0029_2
$\gamma = 1$					
36	-0.0405	-0.0167_2	-0.0032_2	-0.0039_2	-0.0111_2
72	-0.0211	-0.0089_2	-0.0019_2	-0.0020_2	-0.0060_2
108	-0.0140	-0.0053_2	-0.0004_2	-0.0007_2	-0.0033_2
144	-0.0109	-0.0044_2	-0.0008_2	-0.0007_2	-0.0028_2

Subscripts denote the value of m.

Table 3.4: Bias of OLS and bias-minimising jackknife estimators.

n	$\hat{\rho}$	$\hat{\rho}_{J,m}$	$\hat{\rho}_{J,CK}$	$\hat{\rho}^*_J$	$\hat{\rho}^*_{J,2S}$
$\gamma = -10$					
36	0.1302	0.1389_2	0.1413_2	0.1409_2	0.1422_2
72	0.0715	0.0744_2	0.0756_2	0.0754_2	0.0761_2
108	0.0478	0.0510_2	0.0520_2	0.0519_2	0.0524_2
144	0.0382	0.0396_2	0.0403_2	0.0402_2	0.0406_2
$\gamma = -5$					
36	0.1133	0.1204_2	0.1255_2	0.1244_2	0.1262_2
72	0.0620	0.0643_2	0.0670_2	0.0664_2	0.0672_2
108	0.0418	0.0435_2	0.0454_2	0.0450_2	0.0457_2
144	0.0319	0.0335_2	0.0351_2	0.0346_2	0.0351_2
$\gamma = -1$					
36	0.1004	0.1050_2	0.1189_2	0.1138_2	0.1120_2
72	0.0524	0.0552_2	0.0634_2	0.0598_2	0.0591_2
108	0.0351	0.0365_2	0.0419_2	0.0402_2	0.0393_2
144	0.0266	0.0277_2	0.0318_2	0.0303_2	0.0297_2
$\gamma = -0.1$					
36	0.0954	0.1012_2	0.1195_2	0.1116_2	0.1083_2
72	0.0505	0.0534_2	0.0639_2	0.0602_2	0.0574_2
108	0.0332	0.0352_2	0.0424_2	0.0401_2	0.0382_2
144	0.0254	0.0267_2	0.0320_2	0.0302_2	0.0289_2
$\gamma = 0$					
36	0.0944	0.1004_2	0.1189_2	0.1109_2	0.1080_2
72	0.0475	0.0514_2	0.0626_2	0.0586_2	0.0556_2
108	0.0340	0.0355_2	0.0423_2	0.0404_2	0.0382_2
144	0.0241	0.0261_2	0.0320_2	0.0299_2	0.0286_2
$\gamma = 0.1$					
36	0.0927	0.0989_2	0.1176_2	0.1176_2	0.1058_2
72	0.0488	0.0518_2	0.0625_2	0.0625_2	0.0559_2
108	0.0324	0.0348_2	0.0422_2	0.0422_2	0.0377_2
144	0.0250	0.0266_2	0.0266_2	0.0322_2	0.0288_2
$\gamma = 1$					
36	0.0849	0.0968_2	0.1202_2	0.1134_2	0.1056_2
72	0.0444	0.0492_2	0.0607_2	0.0578_2	0.0536_2
108	0.0298	0.0334_2	0.0416_2	0.0391_2	0.0365_2
144	0.0230	0.0249_2	0.0305_2	0.0294_2	0.0271_2

Subscripts denote the value of m.

Table 3.5: RMSE of OLS and bias-minimising jackknife estimators.

n	$\hat{\rho}$	$\hat{\rho}_{J,m}$	$\hat{\rho}_{J,CK}$	$\hat{\rho}^*_J$	$\hat{\rho}^*_{J,2S}$
$\gamma = -10$					
36	−0.0398	-0.0249_{9}	-0.0194_{9}	-0.0192_{9}	-0.0197_{9}
72	−0.0236	-0.0157_{18}	-0.0108_{18}	-0.0107_{18}	-0.0110_{18}
108	−0.0166	-0.0115_{27}	-0.0074_{27}	-0.0072_{27}	-0.0075_{27}
144	−0.0130	-0.0093_{36}	-0.0058_{36}	-0.0056_{36}	-0.0059_{36}
$\gamma = -5$					
36	−0.0445	-0.0299_{9}	-0.0212_{9}	-0.0209_{9}	-0.0217_{9}
72	−0.0252	-0.0183_{18}	-0.0117_{18}	-0.0115_{18}	-0.0124_{18}
108	−0.0168	-0.0125_{27}	-0.0072_{27}	-0.0071_{27}	-0.0078_{27}
144	−0.0131	-0.0101_{36}	-0.0058_{36}	-0.0056_{36}	-0.0062_{36}
$\gamma = -1$					
36	−0.0482	-0.0370_{9}	-0.0243_{9}	-0.0239_{9}	-0.0287_{9}
72	−0.0251	-0.0200_{18}	-0.0115_{18}	-0.0114_{18}	-0.0147_{18}
108	−0.0167	-0.0136_{27}	-0.0073_{27}	-0.0072_{27}	-0.0097_{27}
144	−0.0123	-0.0102_{36}	-0.0052_{36}	-0.0051_{36}	-0.0070_{36}
$\gamma = -0.1$					
36	−0.0474	-0.0374_{9}	-0.0242_{9}	-0.0242_{9}	-0.0307_{9}
72	−0.0243	-0.0200_{18}	-0.0115_{18}	-0.0113_{18}	-0.0153_{18}
108	−0.0161	-0.0135_{27}	-0.0072_{27}	-0.0072_{27}	-0.0101_{27}
144	−0.0124	-0.0106_{36}	-0.0056_{36}	-0.0055_{36}	-0.0079_{36}
$\gamma = 0$					
36	−0.0465	-0.0370_{9}	-0.0242_{9}	-0.0242_{9}	-0.0303_{9}
72	−0.0249	-0.0206_{18}	-0.0121_{18}	-0.0119_{18}	-0.0162_{18}
108	−0.0164	-0.0138_{27}	-0.0075_{27}	-0.0074_{27}	-0.0105_{27}
144	−0.0121	-0.0103_{36}	-0.0053_{36}	-0.0053_{36}	-0.0076_{36}
$\gamma = 0.1$					
36	−0.0453	-0.0358_{9}	-0.0229_{9}	-0.0230_{9}	-0.0292_{9}
72	−0.0234	-0.0193_{18}	-0.0111_{18}	-0.0112_{18}	-0.0151_{18}
108	−0.0165	-0.0140_{27}	-0.0078_{27}	-0.0077_{27}	-0.0108_{27}
144	−0.0115	-0.0098_{36}	-0.0049_{36}	-0.0051_{36}	-0.0073_{36}
$\gamma = 1$					
36	−0.0419	-0.0036_{9}	-0.0215_{9}	-0.0223_{9}	-0.0287_{9}
72	−0.0213	-0.0176_{18}	-0.0099_{18}	-0.0103_{18}	-0.0142_{18}
108	−0.0138	-0.0116_{27}	-0.0061_{27}	-0.0063_{27}	-0.0091_{27}
144	−0.0108	-0.0093_{36}	-0.0049_{36}	-0.0050_{36}	-0.0073_{36}

Subscripts denote the value of m.

Table 3.6: Bias of OLS and RMSE-minimising jackknife estimators.

n	$\hat{\rho}$	$\hat{\rho}_{J,m}$	$\hat{\rho}_{J,CK}$	$\hat{\rho}^*_J$	$\hat{\rho}^*_{J,2S}$
$\gamma = -10$					
36	0.1302	0.1284_9	0.1286_9	0.1283_9	0.1419_9
72	0.0713	0.0690_{18}	0.0682_{18}	0.0681_{18}	0.0696_{18}
108	0.0496	0.0479_{27}	0.0469_{27}	0.0468_{27}	0.0480_{27}
144	0.0380	0.0366_{36}	0.0356_{36}	0.0355_{36}	0.0365_{36}
$\gamma = -5$					
36	0.1149	0.1107_9	0.1099_9	0.1093_9	0.1124_9
72	0.0623	0.0590_{18}	0.0570_{18}	0.0566_{18}	0.0590_{18}
108	0.0422	0.0400_{27}	0.0382_{27}	0.0380_{27}	0.0399_{27}
144	0.0323	0.0307_{36}	0.0290_{36}	0.0287_{36}	0.0304_{36}
$\gamma = -1$					
36	0.1012	0.0955_9	0.0932_9	0.0905_9	0.0951_9
72	0.0525	0.0492_{18}	0.0456_{18}	0.0440_{18}	0.0479_{18}
108	0.0347	0.0325_{27}	0.0294_{27}	0.0282_{27}	0.0313_{27}
144	0.0258	0.0243_{36}	0.0215_{36}	0.0205_{36}	0.0232_{36}
$\gamma = -0.1$					
36	0.0960	0.0903_9	0.0878_9	0.0842_9	0.0891_9
72	0.0505	0.0474_{18}	0.0436_{18}	0.0413_{18}	0.0459_{18}
108	0.0333	0.0312_{27}	0.0280_{27}	0.0263_{27}	0.0299_{27}
144	0.0253	0.0239_{36}	0.0211_{36}	0.0196_{36}	0.0227_{36}
$\gamma = 0$					
36	0.0960	0.0906_9	0.0886_9	0.0843_9	0.0896_9
72	0.0499	0.0467_{18}	0.0426_{18}	0.0399_{18}	0.0449_{18}
108	0.0334	0.0314_{27}	0.0280_{27}	0.0262_{27}	0.0300_{27}
144	0.0246	0.0232_{36}	0.0205_{36}	0.0191_{36}	0.0221_{36}
$\gamma = 0.1$					
36	0.0933	0.0879_9	0.0860_9	0.0816_9	0.0870_9
72	0.0481	0.0451_{18}	0.0417_{18}	0.0392_{18}	0.0435_{18}
108	0.0338	0.0318_{27}	0.0285_{27}	0.0265_{27}	0.0304_{27}
144	0.0241	0.0227_{36}	0.0200_{36}	0.0186_{36}	0.0216_{36}
$\gamma = 1$					
36	0.0876	0.0827_9	0.0814_9	0.0766_9	0.0818_9
72	0.0448	0.0421_{18}	0.0390_{18}	0.0361_{18}	0.0406_{18}
108	0.0292	0.0275_{27}	0.0248_{27}	0.0227_{27}	0.0263_{27}
144	0.0227	0.0215_{36}	0.0191_{36}	0.0172_{36}	0.0205_{36}

Subscripts denote the value of m.

Table 3.7: RMSE of OLS and RMSE-minimising jackknife estimators.

Chapter 4

Ordinary least squares limiting distribution for local to unit root autoregression with a drift driven by weakly dependent errors

Abstract

This essay investigates the asymptotic distribution of the ordinary least squares estimator in discrete time autoregressive series with a drift. The autoregressive coefficient is assumed to be of the local to unit root type and the disturbances which drive the process are allowed to be dependent over time. The convergence rates of the drift and autoregressive parameters are shown to be $n^{1/2}$ and $n^{3/2}$, respectively, where n denotes the sample size. The higher than n rate allows deriving a limiting law for an estimator of the localising parameter given by $\hat{c} = n \log \hat{\rho}$. The centred estimator $\hat{c}$ converges to a normal variate with mean zero. This result permits testing of the null hypothesis of a unit root directly via a t-statistic on $\hat{c}$. The paper also proposes a different estimator of c and a Monte Carlo experiment is run to assess finite sample performance. The limiting result depends on the process possessing a drift, or whether a trend is included as an explanatory variable. Hence, the essay also derives two more limiting distributions. The first one is under the null hypothesis that the drift is zero and the autoregressive parameter is unity, and the second is under the null that the autoregressive parameter is unity and the trend is zero.

4.1 Introduction

The purpose of this chapter is to discuss statistical inference in univariate processes in which the autoregressive coefficient is of the local to unit root type discussed in the previous chapter. However, the model studied here is also assumed to posses a drift and is driven by weakly dependent errors. This is a general setup and permits for autoregressive and moving-average components of the error term. The analysis shows that the vector of estimators of the drift and autoregressive parameter have a joint normal distribution in the limit, where the drift parameter estimator converges at a rate $n^{1/2}$ and the estimator of the autoregressive parameter converges at rate $n^{3/2}$. This permits construction of a t-test of the autoregressive coefficient that has a standard normal distribution asymptotically. Hence, any rejection decision can be based on the standard normal critical values. It is important to note that all of this depends on the assumption that the drift exists in the data generating process. If that assumption is violated, the above-mentioned asymptotic results fail and any t-test constructed about the autoregressive parameter would have a non-standard limiting distribution that can be expressed as a functional of Brownian motions. Similar results have been obtained for pure unit root processes and an excellent text-book treatment is provided in Hamilton (1994). Since it is assumed the data are generated with a drift, and whether or not it exists changes asymptotic behaviour, it is natural to test the joint hypothesis that the drift parameter is equal to zero and the autoregressive parameter is equal to unity. As such, this essay derives the limiting behaviour of the F-test for such a hypothesis. Under the null, the limiting distribution is not the square of a random normal since, when the drift is equal to zero, the process just become a local to unit root and the limiting distribution for such a process is not normal.

The main feature of the model studied here is the rate of convergence of the estimator of the autoregressive parameter. In essence, the explanatory variable y_{t-1} in the regression behaves asymptotically like a trend. The $n^{3/2}$ rate of convergence can then be justified with that argument. However, the higher than n rate permits consistent estimation of the localising parameter. The essay shows that there exists an asymptotic law for an estimator of the localising parameter that has the same asymptotic distribution as the estimator of the autoregressive parameter with a rate of convergence $n^{1/2}$. Consequently, this permits construction of a t-test of the localising parameter that has an asymptotic standard normal distribution and decision for rejection based on the critical values of a standard normal variable. Therefore, one could test the null hypothesis of a unit root via a t-test directly with the estimator of the localising parameter being equal to zero. As far as the author is concerned, this is the first study which documents that the localising parameter is consistently estimable in a process with a drift.

In applied work, when trending data are observed, it is natural to estimate a model which includes an intercept, an autoregressive parameter and a trend. Therefore, this paper also derives the joint distribution of the three estimators under the same data-generating process. Since y_{t-1} behaves like a trend asymptotically including a trend as a separate explanatory variable would induce collinearity in large samples. Thus, deriving an asymptotic law requires rewriting the model and estimating it in a different way. It is shown that the limiting distribution is non-normal and the rate of convergence of the autoregressive parameter reduces to n. Due to the reduced rate of convergence, consistent estimation of the localising parameter is impossible and any inference on that should be conducted via the original regression. Again, since assumption that the trend parameter is equal to zero is imposed, which is an auxiliary assumption, a test for joint significance is a natural way to conduct inference. The essay derives the limiting distribution of such a test as well.

The chapter is organised in the following way: section 2 derives asymptotic laws in regression without a trend, section 3 derives those when a trend is added to the regression, section 4 provides some evidence on the finite sample behaviour of the localising parameter via Monte Carlo experiments, section 5 concludes and the supplementary appendix contains two lemmas which facilitate the proofs from the main text.

The following notation will be employed throughout. The process $J(r) = \int_0^r e^{c(r-s)}\mathrm{d}W(s)$ denotes the Ornstein-Uhlenbeck process which satisfies $dJ(r) = cJ(r)dr + \mathrm{d}W(r)$ for a constant c. Lastly, the essay makes use of the lag operator L which is defined as $L^j y_t = y_{t-j}$.

4.2 Asymptotic results for parameter estimation

We consider a time series process y_t which satisfies the following assumption.

Assumption 4.2 Suppose $y_1, y_2, \ldots, y_n$ are generated according to the following stochastic difference equation

$$y_t = \alpha + \rho_n y_{t-1} + u_t, \qquad t = 1, \ldots, n, \tag{4.1}$$

where the process is initiated at $y_0 = O_p(1)$, $\alpha \neq 0$ and $\rho_n = e^{c/n}$. It is assumed that u_t is a zero mean, stationary and ergodic process with finite autocovariances $\gamma_j = E(u_t u_{t-j})$ such that

(a) $\omega^2 = \sum_{j=-\infty}^{\infty} \gamma_j$ is finite and nonzero;

(b) the scaled partial-sum process $n^{-1/2}\sum_{t=1}^{\lfloor rn \rfloor} u_t \Rightarrow \omega W(r)$, with $r \in [0, 1]$.

The assumptions on u_t cover a wide range of processes including stationary and invertible ARMA models under moment conditions and are standard in the literature (see Elliot *et al.*, 1996). The assumption on the autoregressive coefficient allows for deviations from unity at a rate $O(n^{-1})$ since $e^{c/n} = 1 + c/n + O(n^{-2}) = 1 + O(n^{-1})$. Thus, asymptotically $e^{c/n} \to 1$ but depending on the sign of c, for any finite sample the processes can exhibit (locally) stationary or (locally) explosive behaviour. For $c = 0$, we have unit root process since $e^{c/n}|_{c=0} = 1$. Under those assumptions, the process $\zeta_t = \rho_n \zeta_{t-1} + u_t = \sum_{j=1}^{t} \rho_n^{t-j} u_j$, with $\zeta_0 = 0$, satisfies $n^{-1/2}\zeta_{\lfloor rn \rfloor} \Rightarrow \omega J(r)$ as $n \to \infty$ (Phillips, 1987). For the purposes of this essay, estimation is conducted via the linear ordinary least squares (OLS) method.

4.2.1 Autoregressive coefficient

The OLS estimators of the intercept and autoregressive coefficient in (4.1) are given by

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\rho}_n \end{bmatrix} = \begin{bmatrix} n & \sum_{t=1}^{n} y_{t-1} \\ \sum_{t=1}^{n} y_{t-1} & \sum_{t=1}^{n} y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{n} y_t \\ \sum_{t=1}^{n} y_t y_{t-1} \end{bmatrix}.$$

The deviations from the true parameters are then given by

$$\begin{bmatrix} \hat{\alpha} - \alpha \\ \hat{\rho}_n - \rho_n \end{bmatrix} = \begin{bmatrix} n & \sum_{t=1}^{n} y_{t-1} \\ \sum_{t=1}^{n} y_{t-1} & \sum_{t=1}^{n} y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{n} u_t \\ \sum_{t=1}^{n} y_{t-1} u_t \end{bmatrix}. \tag{4.2}$$

The following theorem summarises the asymptotic distribution of the estimators.

Theorem 4.2.1. *Let $y_1, y_2, \ldots, y_n$ satisfy assumption 4.2. Then as $n \to \infty$*

$$\begin{bmatrix} n^{1/2}(\hat{\alpha} - \alpha) \\ n^{3/2}(\hat{\rho}_n - \rho_n) \end{bmatrix} = \begin{bmatrix} 1 & n^{-2} \sum_{t=1}^{n} y_{t-1} \\ n^{-2} \sum_{t=1}^{n} y_{t-1} & n^{-3} \sum_{t=1}^{n} y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} n^{-1/2} \sum_{t=1}^{n} u_t \\ n^{-3/2} \sum_{t=1}^{n} y_{t-1} u_t \end{bmatrix}$$

$$\Rightarrow N\left(\mathbf{0}, \mathbf{Q^{-1}} \omega^2 \mathbf{Q} \mathbf{Q^{-1}}\right) = N(\mathbf{0}, \omega^2 \mathbf{Q^{-1}}),$$

where

$$\mathbf{Q} = \begin{bmatrix} 1 & c^{-2}\alpha(e^c - c - 1) \\ c^{-2}\alpha(e^c - c - 1) & (2c^3)^{-1}\alpha^2(e^{2c} - 4e^c + 2c + 3) \end{bmatrix}.$$

Proof: The proof of the asymptotic law demonstrates that the first matrix in (4.2) converges in probability to a constant matrix $\mathbf{Q^{-1}}$ which satisfies $\{\mathbf{Q^{-1}}\}^{\mathbf{T}} = \mathbf{Q^{-1}}$. This follows from the fact that the deterministic term in (4.1) is of a higher order than the stochastic. The second matrix is shown to converge to a joint normal random variable with a variance-covariance matrix $\mathbf{Q}$.

We start by writing (4.1) recursively as

$$\begin{aligned}
y_t &= \alpha + \rho_n y_{t-1} + u_t \\
&= \alpha \sum_{i=1}^{t} \rho_n^{i-1} + \rho^t y_0 + \sum_{j=1}^{t} \rho_n^{t-j} u_j \\
&= \alpha \sum_{i=1}^{t} \rho_n^{i-1} + \rho_n^t y_0 + \zeta_t, \qquad (4.3)
\end{aligned}$$

where $\zeta_t = \sum_{j=1}^{t} \rho^{t-j} u_j$. Now, we consider the sums from the first matrix in (4.2) and show their probability limits. For ease of notation we consider y_t instead of y_{t-1} since asymptotically they behave in the same way. For the sum $\sum_{t=1}^{n} y_t$, we have

$$\begin{aligned}
\sum_{t=1}^{n} y_t &= \sum_{t=1}^{n} \left(\alpha \sum_{i=1}^{t} \rho_n^{i-1} + \rho_n^t y_0 + \zeta_t \right) \\
&= \alpha \sum_{t=1}^{n} \sum_{i=1}^{t} \rho_n^{i-1} + y_0 \sum_{t=1}^{n} \rho_n^t + \sum_{t=1}^{n} \zeta_t. \qquad (4.4)
\end{aligned}$$

By Lemma 4.6.1 (a) with $\beta(t/n) = 1$, the first term satisfies

$$\lim_{n\to\infty} n^{-2} \alpha \sum_{t=1}^{n} \sum_{i=1}^{t} \rho_n^{i-1} = \alpha \int_0^1 \int_0^r e^{cs} \mathrm{d}s \mathrm{d}r = \frac{\alpha}{c^2} \left(e^c - c - 1 \right).$$

To deal with the second term we note that

$$\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} \rho_n^t = \int_0^1 e^{cr} \mathrm{d}r,$$

which is just a Riemann sum. For the third, by applying Lemma 4.6.1 (b) with $\beta(t) = e^{ct}$, $\delta(j) = e^{-cj}$ and omitting the $\beta(t)$ outside of ξ_t yields

$$\begin{aligned}
\lim_{n\to\infty} n^{-3/2} \sum_{t=1}^{n} \zeta_t &\Rightarrow \omega \int_0^1 e^{cr} \left(W(r) e^{-cr} + c \int_0^r W(s) e^{-cs} \mathrm{d}s \right) \mathrm{d}r \\
&= \omega \int_0^1 J(r) \mathrm{d}r. \qquad (4.5)
\end{aligned}$$

Therefore, in (4.4) we have that the first term is $O(n^2)$, the second is $O_p(n)$ and the third is $O_p(n^{3/2})$. It follows that

$$n^{-2}\sum_{t=1}^{n} y_t \to_p \frac{\alpha}{c^2}\left(e^c - c - 1\right) =: q_{12}. \tag{4.6}$$

Similarly, we have

$$\begin{aligned}\sum_{t=1}^{n} y_t^2 =& \sum_{t=1}^{n}\left(\alpha\sum_{i=1}^{t}\rho_n^{i-1} + y_0\rho^t + \zeta_t\right)^2\\ =&\alpha^2\sum_{t=1}^{n}\left(\sum_{i=1}^{t}\rho_n^{i-1}\right)^2 + y_0^2\sum_{t=1}^{n}\rho_n^{2t} + \sum_{t=1}^{n}\zeta_t^2 + 2y_0\sum_{t=1}^{n}\rho_n^t\zeta_t\\ &+ 2\alpha y_0\sum_{t=1}^{n}\sum_{i=1}^{t}\rho_n^{t+i-1} + 2\alpha\sum_{t=1}^{n}\sum_{i=1}^{t}\rho_n^{i-1}\zeta_t. \end{aligned}\tag{4.7}$$

By Lemma 4.6.1 (a) the first term from the second line above satisfies

$$\begin{aligned}\lim_{n\to\infty} n^{-3}\alpha^2\sum_{t=1}^{n}\left(\sum_{i=1}^{t}\rho_n^{i-1}\right)^2 &= \alpha^2\int_0^1\left(\int_0^r e^{cs}\mathrm{d}s\right)^2\mathrm{d}r\\ &= \frac{\alpha^2}{2c^3}\left(e^{2c} - 4e^c + 2c + 3\right).\end{aligned}$$

To deal with the second term we note that

$$\lim_{n\to\infty} n^{-1}\sum_{t=1}^{n}\rho_n^{2t} = \int_0^1 e^{2cr}\mathrm{d}r,$$

which is just a Riemann sum. For the third, from Lemma 2.1 of Phillips (1987)

$$\lim_{n\to\infty} n^{-2}\sum_{t=1}^{n}\zeta_t^2 \Rightarrow \omega^2\int_0^1 J(r)^2\mathrm{d}r. \tag{4.8}$$

To deal with the fourth and fifth terms we note that

$$\lim_{n\to\infty} n^{-3/2}\sum_{t=1}^{n}\rho_n^t\zeta_t \Rightarrow \omega\int_0^1 e^{cr}J(r)\mathrm{d}r,$$

by Lemma 4.6.1 (b) and

$$\lim_{n\to\infty} n^{-2}2\alpha\sum_{t=1}^{n}\sum_{i=1}^{t}\rho_n^{t+i-1} = 2a\int_0^1\int_0^r e^{c(r+s)}\mathrm{d}s\mathrm{d}r,$$

by Lemma 4.6.1 (a). Lastly, by Lemma 4.6.1 (c), the sixth satisfies

$$\lim_{n\to\infty} n^{-5/2}2\alpha\sum_{t=1}^{n}\sum_{i=1}^{t}\rho_n^{i-1}\zeta_t \Rightarrow 2a\omega\int_0^1\int_0^r e^{cs}J(r)\mathrm{d}s\mathrm{d}r.$$

Consequently, we have the order the terms in (4.7) as follows: the first is $O(n^3)$, second is $O_p(n)$, third is $O_p(n^2)$, fourth is $O_p(n^{3/2})$, fifth is $O_p(n^2)$ and the sixth is $O_p(n^{5/2})$. It follows that

$$n^{-3}\sum_{t=1}^{n}y_t^2 \to_p \frac{\alpha^2}{2c^3}\left(e^{2c}-4e^c+2c+3\right) =: q_{22}. \tag{4.9}$$

Now, we consider the elements from the second matrix in (4.2). For the first, we have $n^{-1/2}\sum_{t=1}^{n}u_t \Rightarrow \omega\int_0^1 \mathrm{d}W(r)$ by Lemma 4.6.1 (e) with $\beta(t)=1$. The second becomes

$$\begin{aligned}\sum_{t=1}^{n}y_{t-1}u_t &= \sum_{t=1}^{n}\left(\alpha\sum_{i=1}^{t-1}\rho_n^{i-1}+\rho_n^{t-1}y_0+\zeta_{t-1}\right)u_t \\ &= \alpha\sum_{t=1}^{n}\sum_{i=1}^{t-1}\rho_n^{i-1}u_t + y_0\sum_{t=1}^{n}\rho_n^{t-1}u_t+\sum_{t=1}^{n}\zeta_{t-1}u_t.\end{aligned} \tag{4.10}$$

By Lemma 4.6.1 (d) and (e) the first and second terms above satisfy

$$\begin{aligned}\lim_{n\to\infty} n^{-3/2}\alpha\sum_{t=1}^{n}\sum_{i=1}^{t-1}\rho_n^{i-1}u_t &\Rightarrow \alpha\omega\int_0^1\int_0^r e^{cs}\mathrm{d}s\mathrm{d}W(r); \\ \lim_{n\to\infty} n^{-1/2}\sum_{t=1}^{n}\rho_n^{t-1}u_t &\Rightarrow \omega\int_0^1 e^{cr}\mathrm{d}W(r),\end{aligned}$$

respectively. Lastly, by Lemma 2.1 of Phillips (1987) the third term satisfies

$$\lim_{n\to\infty} n^{-1}\sum_{t=1}^{n}\zeta_{t-1}u_t \Rightarrow \omega^2\left(\int_0^1 J(r)\mathrm{d}W(r)+\frac{1}{2}(1-\lambda)\right), \tag{4.11}$$

where $\lambda = \gamma_0/\omega^2$. The order of the terms in (4.10) are as follows: the first is $O_p(n^{3/2})$, the second is $O_p(n^{1/2})$ and the third is $O_p(n)$. Therefore, we have

$$n^{-3/2}\sum_{t=1}^{n} y_{t-1}u_t \Rightarrow \alpha\omega \int_0^1 \int_0^r e^{cs}\mathrm{d}s\mathrm{d}W(r). \tag{4.12}$$

The quantities in (4.2) require appropriate scaling to achieve an asymptotic law. However, unlike in the scalar case, this requires a scaling matrix. The appropriate choice of scaling matrix in this case is

$$\Gamma_n = \begin{bmatrix} n^{1/2} & 0 \\ 0 & n^{3/2} \end{bmatrix}.$$

Then (4.2) can be rewritten as

$$\begin{aligned}\Gamma_n \begin{bmatrix} \hat{\alpha} - \alpha \\ \hat{\rho}_n - \rho_n \end{bmatrix} &= \left\{ \Gamma_n^{-1} \begin{bmatrix} n & \sum_{t=1}^n y_{t-1} \\ \sum_{t=1}^n y_{t-1} & \sum_{t=1}^n y_{t-1}^2 \end{bmatrix} \Gamma_n^{-1} \right\}^{-1} \\ &\times \left\{ \Gamma_n^{-1} \begin{bmatrix} \sum_{t=1}^n u_t \\ \sum_{t=1}^n y_{t-1}u_t \end{bmatrix} \right\}. \end{aligned} \tag{4.13}$$

Or

$$\begin{aligned}\begin{bmatrix} n^{1/2}(\hat{\alpha} - \alpha) \\ n^{3/2}(\hat{\rho}_n - \rho_n) \end{bmatrix} &= \begin{bmatrix} 1 & n^{-2}\sum_{t=1}^n y_{t-1} \\ n^{-2}\sum_{t=1}^n y_{t-1} & n^{-3}\sum_{t=1}^n y_{t-1}^2 \end{bmatrix}^{-1} \\ &\times \begin{bmatrix} n^{-1/2}\sum_{t=1}^n u_t \\ n^{-3/2}\sum_{t=1}^n y_{t-1}u_t \end{bmatrix}. \end{aligned}$$

From (4.6) and (4.9)

$$\begin{bmatrix} 1 & n^{-2}\sum_{t=1}^n y_{t-1} \\ n^{-2}\sum_{t=1}^n y_{t-1} & n^{-3}\sum_{t=1}^n y_{t-1}^2 \end{bmatrix}^{-1} \to_p \begin{bmatrix} 1 & q_{12} \\ q_{12} & q_{22} \end{bmatrix}^{-1} = \mathbf{Q^{-1}} \tag{4.14}$$

and from Lemma 4.6.1 and (4.12)

$$\begin{bmatrix} n^{-1/2}\sum_{t=1}^{n} u_t \\ n^{-3/2}\sum_{t=1}^{n} y_{t-1}u_t \end{bmatrix} \Rightarrow \begin{bmatrix} \omega \int\limits_0^1 \mathrm{d}W(r) \\ a\omega \int\limits_0^1 \int\limits_0^r e^{cs}\mathrm{d}s\mathrm{d}W(r) \end{bmatrix}. \tag{4.15}$$

The variables in (4.15) are both normal with mean zero. Their variances and covariance can be derived by virtue of the Itô isometry. The variance of the first is

$$E\left(\omega \int_0^1 \mathrm{d}W(r)\right)^2 = \omega^2 \int_0^1 \mathrm{d}r = \omega^2$$

and the variance of the second is

$$\begin{aligned} E\left(a\omega \int_0^1 \int_0^r e^{cs}\mathrm{d}s\mathrm{d}W(r)\right)^2 &= \frac{a^2\omega^2}{c^2} E\left(\int_0^1 \left(e^{cr}-1\right)\mathrm{d}W(r)\right)^2 \\ &= \frac{a^2\omega^2}{c^2}\int_0^1 \left(e^{cr}-1\right)^2 \mathrm{d}r \\ &= \omega^2 q_{22}. \end{aligned} \tag{4.16}$$

Their covariance is

$$\begin{aligned} &E\left\{\left(\omega\int_0^1 \mathrm{d}W(r)\right)\left(a\omega \int_0^1\int_0^r e^{cs}\mathrm{d}s\mathrm{d}W(r)\right)\right\} \\ &= \frac{a\omega^2}{c} E\left(\int_0^1 \mathrm{d}W(r)\int_0^1 \left(e^{cr}-1\right)\mathrm{d}W(r)\right) \\ &= \frac{a\omega^2}{c}\int_0^1 \left(e^{cr}-1\right)\mathrm{d}r \\ &= \omega^2 q_{12}. \end{aligned}$$

Therefore, (4.15) becomes

$$\begin{aligned} \begin{bmatrix} n^{-1/2}\sum_{t=1}^{n} u_t \\ n^{-3/2}\sum_{t=1}^{n} y_{t-1}u_t \end{bmatrix} \Rightarrow \begin{bmatrix} \omega \int\limits_0^1 \mathrm{d}W(r) \\ a\omega \int\limits_0^1 \int\limits_0^r e^{cs}\mathrm{d}s\mathrm{d}W(r) \end{bmatrix} &=_d N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \omega^2 \begin{pmatrix} 1 & q_{12} \\ q_{12} & q_{22} \end{pmatrix}\right] \\ &= N(\mathbf{0}, \omega^2 \boldsymbol{Q}). \end{aligned} \tag{4.17}$$

Finally, by virtue of Slutsky's theorem, the asymptotic distribution of the centered OLS estimators can be derived by combining (4.14) and (4.17) and is stated in the theorem. □

Hence, provided that the true α is different from zero, the limiting distribution of the vector of OLS estimates converges to a joint normal distribution. Thus, any t-test or F-test would have the standard critical values.

In passing by the proof of Theorem 4.2.1, we make use of equation (4.16) to prove the following proposition.

Proposition 4.2.2. *Define* $q^2 := q_{22}/\alpha^2 = (1/2c^3)(e^{2c} - 4e^c + 2c + 3)$. *Then, the following distributional equality holds:* $\int_0^1 \int_0^r e^{cs} \mathrm{d}s\mathrm{d}W(r) =_d \int_0^1 J(r)\mathrm{d}r =_d N(0, q^2)$.

Proof: Phillips (1987) showed that $\int_0^1 J(r)\mathrm{d}r =_d N(0, q^2)$ and from (4.16) we have that $\int_0^1 \int_0^r e^{cs}\mathrm{d}s\mathrm{d}W(r) =_d N(0, q^2)$. Furthermore, two normal random variables which have the same first and second moments have identical characteristics functions. The proof follows from the fact that the characteristic function of a random variable defines its cumulative distribution uniquely. □

As a consequence of Theorem 1, we have the following corollary.

Corollary 4.2.3. *Let the assumptions from Theorem 4.2.1 be satisfied. Then, as* $n \to \infty$

$$n^{3/2}(\hat{\rho}_n - \rho_n) \Rightarrow N\left(0, \omega^2 \left(\frac{2c^4}{\alpha^2(c(e^{2c} - 1) - 2(e^c - 1)^2)}\right)\right).$$

Proof: To show the result we only need invert the $\mathbf{Q}$ matrix and pick the respective element. We write

$$\mathbf{Q}^{-1} = \begin{bmatrix} 1 & q_{12} \\ q_{12} & q_{22} \end{bmatrix}^{-1} = \frac{1}{q_{22} - q_{12}^2} \begin{bmatrix} q_{22} & -q_{12} \\ -q_{12} & 1 \end{bmatrix}.$$

Now, we are interested in the last element of the diagonal which after a bit of algebra becomes

$$\frac{1}{q_{22} - q_{12}^2} = \frac{2c^4}{\alpha^2(c(e^{2c} - 1) - 2(e^c - 1)^2)},$$

from which the result follows immediately. □

The numerator above is positive and Lemma 4.6.2 ensures that the denominator is positive as well. It is important to note that for $\alpha = 0$ the asymptotic theory would break as in this case the variance of $\hat{\rho}_n$ would be undefined. This is in line with the discussion from the previous paragraph that the normal limiting distribution breaks.

It would be interesting to consider the case in which $c = 0$. We can do this directly by taking the limit $\lim_{c\to 0} \mathbf{Q}^{-1}$ to derive the limiting distribution of OLS under a unit root process with a drift. Since the result follows from Theorem 1, we state it as a corollary.

Corollary 4.2.4. *Let the assumptions from Theorem 4.2.1 be satisfied with $c = 0$. Then as $n \to \infty$*

$$\begin{bmatrix} n^{1/2}(\hat{\alpha} - \alpha) \\ n^{3/2}(\hat{\rho}_n - 1) \end{bmatrix} \Rightarrow N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \omega^2 \begin{pmatrix} 1 & \alpha/2 \\ \alpha/2 & \alpha^2/3 \end{pmatrix}^{-1} \right].$$

Proof: We first start by taking

$$\lim_{c\to 0} q_{12} = \lim_{c\to 0} \frac{\alpha}{c^2}(e^c - c - 1).$$

Applying the l'Hopital's rule twice to the fraction yields

$$\lim_{c\to 0} q_{12} = \alpha/2.$$

We finish by taking

$$\lim_{c\to 0} q_{22} = \lim_{c\to 0} \frac{\alpha^2}{2c^3}(e^{2c} - 4e^c + 2c + 3).$$

Applying the l'Hopital's rule thrice yields

$$\lim_{c\to 0} q_{22} = \alpha^2/3,$$

completing the proof. □

If the sequence of disturbances is iid, Corollary 4.2.4 reduces to a result which can be found in Hamilton (1994).

It is important to mention that if the true α is in fact equal to zero, then asymptotic behaviour changes and with that any critical values for a t or F-test. If the true value of α is in fact zero, we have

Proposition 4.2.5. *Let $y_1, y_2, \ldots, y_n$ satisfy assumption 4.2 with $\alpha = 0$. Then,*

$$\begin{aligned}
\begin{bmatrix} n^{1/2}\hat{\alpha} \\ n(\hat{\rho}_n - \rho_n) \end{bmatrix} &= \begin{bmatrix} 1 & n^{-3/2}\sum_{t=1}^n y_{t-1} \\ n^{-3/2}\sum_{t=1}^n y_{t-1} & n^{-2}\sum_{t=1}^n y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} n^{-1/2}\sum_{t=1}^n u_t \\ n^{-1}\sum_{t=1}^n y_{t-1}u_t \end{bmatrix} \\
&\Rightarrow \begin{bmatrix} 1 & \omega\int_0^1 J(r)\mathrm{d}r \\ \omega\int_0^1 J(r)\mathrm{d}r & \omega^2\int_0^1 J(r)^2\mathrm{d}r \end{bmatrix}^{-1} \\
&\times \begin{bmatrix} \omega W(1) \\ \omega^2\left(\int_0^1 J(r)\mathrm{d}W(r) + (1/2)(1-\lambda)\right) \end{bmatrix} \\
&\equiv \boldsymbol{A_c^{-1}B_c}\,.
\end{aligned}$$

Proof: The dominating stochastic terms and their orders are given in (4.5), (4.8) and (4.11). Define the scaling matrix as

$$\boldsymbol{\Lambda_n} = \begin{bmatrix} n^{1/2} & 0 \\ 0 & n \end{bmatrix}$$

and the rest of the proof follows the algebraic manipulations used to derive Theorem 4.2.1. □

The resulting distributions are non-normal and any critical values for rejection will be different from the standard normal. Due to this asymptotic behaviour, it is natural to consider an F-test that $\alpha = 0$ and $\rho_n = 1$. The null hypothesis can be written in the form $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, where $\mathbf{R} = I_2$, the identity matrix of dimension two, $\boldsymbol{\beta} = (\alpha, \rho_n)'$ and $\mathbf{r} = (0,1)'$. Define $\mathbf{x}_t := (1, y_{t-1})'$. Let $\mathbf{b} = (\hat{\alpha}, \hat{\rho}_n)$. The F-test is, then, given by

$$\begin{aligned}
F &= (\mathbf{b} - \boldsymbol{\beta})'\boldsymbol{R}' \left\{ \hat{\omega}^2 \boldsymbol{R} \left(\sum_{t=1}^{n} \mathbf{x_t}\mathbf{x_t}' \right)^{-1} \boldsymbol{R} \right\}^{-1} \boldsymbol{R}(\mathbf{b} - \boldsymbol{\beta})/2 \\
&= (\boldsymbol{\Lambda_n}(\mathbf{b} - \boldsymbol{\beta}))' \left\{ \hat{\omega}^2 \boldsymbol{\Lambda_n} \left(\sum_{t=1}^{n} \mathbf{x_t}\mathbf{x_t}' \right)^{-1} \boldsymbol{\Lambda_n} \right\}^{-1} \boldsymbol{\Lambda_n}(\mathbf{b} - \boldsymbol{\beta})/2 \\
&\Rightarrow (2\omega^2)^{-1} (\boldsymbol{A_0^{-1}B_0})'\boldsymbol{A_0}\boldsymbol{A_0^{-1}B_0} \\
&= (2\omega^2)^{-1} \boldsymbol{B_0'}\boldsymbol{A_0^{-1}}\boldsymbol{B_0}.
\end{aligned}$$

Under the null hypothesis, the limiting distribution is not a Chi-squared variable and any testing would require critical values from this non-standard distribution.

4.2.2 Localising parameter

This subsection considers limiting results regarding the localising parameter c. Since $\rho = e^{c/n}$, taking natural logs on both sides and solving for c yields $c = n \log \rho$. Thus, a natural estimator of c would be $\hat{c} = n \log \hat{\rho}$. The following theorem provides an asymptotic law for the estimator.

Theorem 4.2.6. *Let $y_1, y_2, \ldots, y_n$ satisfy assumption 4.1 and let $\hat{c} = n \log \hat{\rho}_n$. Then as $n \to \infty$, we have*

$$n^{1/2}(\hat{c} - c) \Rightarrow N\left(0, \omega^2 \left(\frac{2c^4}{\alpha^2 (c(e^{2c} - 1) - 2(e^c - 1)^2)} \right) \right).$$

Proof: The proof utilises the result from Corollary 1 and Proposition 2. By utilising the asymptotic equivalence $\log(1 + x) = x + O(x^2)$ as $x \to 0$, we have that

$$\begin{aligned}
\hat{c} = n\log\hat{\rho}_n &= n\log((\hat{\rho}_n - \rho_n) + \rho_n) \\
&= n\log\left((\hat{\rho}_n - \rho_n) + 1 + \frac{c}{n} + O\left(n^{-2}\right)\right) \\
&= n\left\{(\hat{\rho}_n - \rho_n) + \frac{c}{n} + O\left(n^{-2}\right) + O\left[\left((\hat{\rho}_n - \rho_n) + \frac{c}{n} + O\left(n^{-2}\right)\right)^2\right]\right\} \\
&= n(\hat{\rho}_n - \rho_n) + c + O(n^{-1}).
\end{aligned}$$

Now, if $\alpha \neq 0$, we can utilise the fast convergence rate of $\hat{\rho}$ yielding

$$\begin{aligned}
n^{1/2}(\hat{c} - c) &= n^{3/2}(\hat{\rho} - \rho) + O(n^{-1/2}) \\
&\Rightarrow N\left(0, \omega^2\left(\frac{2c^4}{\alpha^2(c(e^{2c} - 1) - 2(e^c - 1)^2)}\right)\right).
\end{aligned}$$

as required. □

Thus, consistent estimation of the localising parameter depends on whether α is equal to zero or not. In the former case, consistent estimation is not possible. However, in the latter case it is. Furthermore, the asymptotic distributions of both ρ_n and c are normal and depend on parameters which can be estimated consistently. It is, therefore, straightforward to construct estimators which have $N(0,1)$ limiting distributions and utilise those for inference. The results are gathered in the following corollary to Theorems 4.2.1 and 4.2.6.

Corollary 4.2.7. *Let the assumptions of Theorem 4.2.1 be satisfied. Also, assume that we have $\hat{\omega} \rightarrow_p \omega$, for some $\hat{\omega}$. Furthermore, let $\alpha \neq 0$ and define*

$$f(c, \alpha, \omega) = (\alpha^2(c(e^{2c} - 1) - 2(e^c - 1)^2)^{-1} 2c^4\omega^2.$$

Then, as $n \to \infty$

(a) $t_\rho \equiv f(\hat{c}, \hat{\alpha}, \hat{\omega})^{-1/2} n^{3/2}(\hat{\rho}_n - \rho_n) \Rightarrow N(0,1)$;

(b) $t_c \equiv f(\hat{c}, \hat{\alpha}, \hat{\omega})^{-1/2} n^{1/2}(\hat{c} - c) \Rightarrow N(0,1)$.

Proof: *The proof is a straightforward application of Slutsky's theorem since* $f(\hat{c}, \hat{\alpha}, \hat{\omega})^{-1/2} \to_p f(c, \alpha, \omega)^{-1/2}$. $\square$

Given Corollary 4.2.7, constructing tests for a hypothesised value $\bar{\rho}_n$ or $\bar{c}$ is straightforward. Let our null hypothesis be $\hat{\rho}_n = \bar{\rho}_n$ or $\hat{c} = \bar{c}$. Then, under the null, as $n \to \infty$

$$t_{\bar{\rho}_n} := f(\hat{c}, \hat{\alpha}, \hat{\omega})^{-1} n^{3/2} (\hat{\rho}_n - \bar{\rho}_n) \Rightarrow N(0,1) \tag{4.18}$$

and

$$t_{\bar{c}} := f(\hat{c}, \hat{\alpha}, \hat{\omega})^{-1} n^{1/2} (\hat{c} - \bar{c}) \Rightarrow N(0,1). \tag{4.19}$$

Estimating $f(c, \alpha, \omega)$ in this way is inefficient due to the need of estimating three parameters, taking exponents of those and multiplying them by each other. Further possible loss of efficiency comes from the fact that ρ needs to be estimated first, transformed and then multiplied by n. Thus, any variation in the estimation of the autoregressive parameter would be then magnified by multiplying by n. However, there is a way to estimate c without the need of $\hat{\rho}$. The following subsection derives such an estimator and also considers more efficient t-statistics than the ones constructed in (4.18) and (4.19).

4.2.3 A second estimator of the localising parameter and more efficient t-statistics

Since $e^{c/n} = 1 + c/n + O(n^{-2})$ we can rewrite (4.1) as

$$\begin{aligned} \Delta y_t &= \alpha + c(y_{t-1}/n) + u_t + y_{t-1} O(n^{-2}) \\ &= \alpha + c(y_{t-1}/n) + \eta_t, \end{aligned}$$

where $\Delta y_t = (1-L)y_t = y_t - y_{t-1}$ and $\eta_t = u_t + y_{t-1}O(n^{-2})$. The OLS estimator of α and c is then given by

$$\begin{bmatrix} \tilde{\alpha} \\ \tilde{c} \end{bmatrix} = \begin{bmatrix} n & \sum_{t=1}^{n} y_{t-1}/n \\ \sum_{t=1}^{n} y_{t-1}/n & \sum_{t=1}^{n} y_{t-1}^2/n^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{n} \Delta y_t \\ \sum_{t=1}^{n} (y_{t-1}/n)\Delta y_t \end{bmatrix},$$

and the deviations from the true parameters are given by

$$\begin{bmatrix} \tilde{\alpha}-\alpha \\ \tilde{c}-c \end{bmatrix} = \begin{bmatrix} n & \sum_{t=1}^{n} y_{t-1}/n \\ \sum_{t=1}^{n} y_{t-1}/n & \sum_{t=1}^{n} y_{t-1}^2/n^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{n} \eta_t \\ \sum_{t=1}^{n} (y_{t-1}/n)\eta_t \end{bmatrix}. \tag{4.20}$$

For $\alpha \neq 0$, the correct scaling matrix is

$$\Psi_n = \begin{bmatrix} n^{1/2} & 0 \\ 0 & n^{1/2} \end{bmatrix}.$$

Multiplying both sides of (4.20) by Ψ_n yields

$$\begin{bmatrix} n^{1/2}(\tilde{\alpha}-\alpha) \\ n^{1/2}(\tilde{c}-c) \end{bmatrix} = \begin{bmatrix} 1 & n^{-2}\sum_{t=1}^{n} y_{t-1} \\ n^{-2}\sum_{t=1}^{n} y_{t-1} & n^{-3}\sum_{t=1}^{n} y_{t-1}^2 \end{bmatrix}^{-1} \times \begin{bmatrix} n^{-1/2}\sum_{t=1}^{n} \eta_t \\ n^{-3/2}\sum_{t=1}^{n} y_{t-1}\eta_t \end{bmatrix}. \tag{4.21}$$

For $\alpha = 0$, we have

$$\begin{bmatrix} n^{1/2}\tilde{\alpha} \\ \tilde{c}-c \end{bmatrix} = \begin{bmatrix} 1 & n^{-3/2}\sum_{t=1}^{n} y_{t-1} \\ n^{-3/2}\sum_{t=1}^{n} y_{t-1} & n^{-2}\sum_{t=1}^{n} y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} n^{-1/2}\sum_{t=1}^{n} \eta_t \\ n^{-1}\sum_{t=1}^{n} y_{t-1}\eta_t \end{bmatrix}. \tag{4.22}$$

Note that $n^{-1/2}\sum_{t=1}^{n}\eta_t = n^{-1/2}\sum_{t=1}^{n} u_t + o_p(1)$ irrespective of the value of α. Furthermore, for $\alpha \neq 0$, $n^{-3/2}\sum_{t=1}^{n} y_{t-1}\eta_t = \sum_{t=1}^{n} y_{t-1}u_t + o_p(1)$ and, for $\alpha = 0$, $n^{-1}\sum_{t=1}^{n} y_{t-1}\eta_t = n^{-1}\sum_{t=1}^{n} y_{t-1}u_t + o_p(1)$. Hence, the matrices on the right hand sides of both (4.21) and (4.22) asymptotically behave in the same way as the ones from Theorem 4.2.1 and Proposition 4.2.5, respectively, and $\tilde{\alpha}$ and $\tilde{c}$ have the same limiting distribution as $\hat{\alpha}$ and $\hat{c}$. The result is stated as a theorem.

Theorem 4.2.8. *Let $y_1, y_2, \ldots, y_n$ satisfy assumption (1). Then as $n \to \infty$*

$$(a)\quad \begin{bmatrix} n^{1/2}(\tilde{\alpha}-\alpha) \\ n^{1/2}(\tilde{c}-c) \end{bmatrix} = \begin{bmatrix} 1 & n^{-2}\sum_{t=1}^{n} y_{t-1} \\ n^{-2}\sum_{t=1}^{n} y_{t-1} & n^{-3}\sum_{t=1}^{n} y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} n^{-1/2}\sum_{t=1}^{n} u_t \\ n^{-3/2}\sum_{t=1}^{n} y_{t-1}u_t \end{bmatrix}$$

$$\Rightarrow N\left(\mathbf{0}, \mathbf{Q^{-1}}\omega^2\mathbf{Q}\mathbf{Q^{-1}}\right) = N(\mathbf{0}, \omega^2\mathbf{Q^{-1}}), \qquad \alpha \neq 0;$$

$$(b)\quad \begin{bmatrix} n^{1/2}\tilde{\alpha} \\ \tilde{c}-c \end{bmatrix} = \begin{bmatrix} 1 & n^{-3/2}\sum_{t=1}^{n} y_{t-1} \\ n^{-3/2}\sum_{t=1}^{n} y_{t-1} & n^{-2}\sum_{t=1}^{n} y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} n^{-1/2}\sum_{t=1}^{n} u_t \\ n^{-3/2}\sum_{t=1}^{n} y_{t-1}u_t \end{bmatrix}$$

$$\Rightarrow \boldsymbol{A_c^{-1}B_c}, \qquad \alpha = 0.$$

From Theorems 4.2.6 and 4.2.8, we have $\hat{c} - \tilde{c} \to_p 0$. Thus, the asymptotic behaviour of $\hat{c}$ and $\tilde{c}$ is the same. Section 4 provides a discussion between the finite sample performance of the two estimators.

To construct more efficient t-statistics than those given in (4.18) and (4.19) we recall that $(\sum_{t=1}^{n} y_{t-1}^2, \sum_{t=1}^{n} y_{t-1}^2) \to_p (q_{12}^2, q_{22})$. Hence, we can define a t-statistic as

$$\tilde{t}_{\bar{\rho}_n} := \frac{n^{3/2}(\hat{\rho}_n - \bar{\rho}_n)}{(n^3\hat{\sigma}^2)^{1/2}}, \tag{4.23}$$

where

$$\begin{aligned} n^3\hat{\sigma}^2 &= n^3\hat{\omega}^2 \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} n & \sum_{t=1}^{n} y_{t-1} \\ \sum_{t=1}^{n} y_{t-1} & \sum_{t=1}^{n} y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \hat{\omega}^2 \begin{bmatrix} 0 & 1 \end{bmatrix} \Gamma_n \begin{bmatrix} n & \sum_{t=1}^{n} y_{t-1} \\ \sum_{t=1}^{n} y_{t-1} & \sum_{t=1}^{n} y_{t-1}^2 \end{bmatrix}^{-1} \Gamma_n \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \hat{\omega}^2 \begin{bmatrix} 0 & 1 \end{bmatrix} \left\{ \Gamma_n^{-1} \begin{bmatrix} n & \sum_{t=1}^{n} y_{t-1} \\ \sum_{t=1}^{n} y_{t-1} & \sum_{t=1}^{n} y_{t-1}^2 \end{bmatrix} \Gamma_n^{-1} \right\}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{aligned}$$

$$\to_p \omega^2 \begin{bmatrix} 0 & 1 \end{bmatrix} \mathbf{Q}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= \frac{\omega^2}{q_{22} - q_{12}^2}. \tag{4.24}$$

Combining this with Corollary 4.2.3, it follows that under the null

$$\tilde{t}_{\bar{\rho}_n} = \frac{n^{3/2}(\hat{\rho}_n - \bar{\rho}_n)}{(n^3\hat{\sigma}^2)^{1/2}} \Rightarrow N(0,1). \tag{4.25}$$

By the same line of analysis we obtain

$$\tilde{t}_{\bar{c}} = \frac{n^{1/2}(\hat{c} - \bar{c})}{(n^3\hat{\sigma}^2)^{1/2}} \Rightarrow N(0,1). \tag{4.26}$$

Evidence of the better finite sample performance of the estimators constructed in this section will be provided in section 4 where Monte Carlo experiments are conducted.

4.3 Trend fitted in estimation

Often enough in empirical work, researchers fit a trend in the model when they have evidence that the data is trending. Hence, typically the autoregression utilised has the form

$$y_t = \alpha + \rho_n y_{t-1} + \delta t + u_t, \qquad t = 1, \ldots, n.$$

Under assumption 4.1 $\delta = 0$. In this case, when $\alpha \neq 0$, the component y_{t-1} would be asymtotically equal to a time trend. To see why, consider the first sum in the last line of (4.3) for a fixed t.

$$\sum_{i=1}^{t} \rho^{i-1} = \frac{e^{ct/n} - 1}{e^{c/n} - 1} = t\left(\frac{1 + O(ct/n)}{1 + O(c/n)}\right) \to t,$$

as $n \to \infty$. In fact, for $c = 0$, the above expression reduces to t, the linear trend we get algebraically under a unit root process. Since a time trend is already included as a separate variable, this would induce collinearity between the explanatory variables in large samples. One way to overcome the problem is to rewrite the model as

$$\begin{aligned} y_t &= (1-\rho_n)\alpha + \rho_n(y_{t-1} - \alpha(t-1)) + (\delta + \rho_n\alpha)t + u_t \\ &\equiv \alpha^* + \rho^*\zeta_{t-1} + \delta^* t + u_t, \end{aligned} \tag{4.27}$$

where $\alpha^* \equiv (1-\rho_n)\alpha$, $\rho^* = \rho_n$, $\delta^* \equiv (\delta + \rho_n\alpha)$ and ζ_t being the process utilised in section 2 with $\zeta_0 = \rho_n^t y_0$. Under the hypothesis that $\rho \sim 1 + c/n$ and $\delta = 0$, we have that $\alpha^* \sim (c/n)\alpha$ and $\delta^* = \rho_n\alpha$. The OLS estimates of the parameters in (4.27) satisfy

$$\begin{bmatrix} \hat{\alpha}^* - \alpha^* \\ \hat{\rho}^* - \rho \\ \hat{\delta}^* - \delta^* \end{bmatrix} = \begin{bmatrix} n & \sum_{t=1}^n \zeta_{t-1} & \sum_{t=1}^n t \\ \sum_{t=1}^n \zeta_{t-1} & \sum_{t=1}^n \zeta_{t-1}^2 & \sum_{t=1}^n \zeta_{t-1} t \\ \sum_{t=1}^n t & \sum_{t=1} t\zeta_{t-1} & \sum_{t=1}^n t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^n u_t \\ \sum_{t=1}^n \zeta_{t-1} u_t \\ \sum_{t=1}^n t u_t \end{bmatrix}.$$

In this case, we should define the scaling matrix as

$$\mathbf{\Upsilon_n} \coloneqq \begin{bmatrix} n^{1/2} & 0 & 0 \\ 0 & n & 0 \\ 0 & 0 & n^{3/2} \end{bmatrix}.$$

Applying the same procedure as in (4.13) yields

$$\begin{bmatrix} n^{1/2}(\hat{\alpha}^* - \alpha^*) \\ n(\hat{\rho}^* - \rho) \\ n^{3/2}(\hat{\delta}^* - \delta^*) \end{bmatrix} = \begin{bmatrix} 1 & n^{-3/2}\sum_{t=1}^n \zeta_{t-1} & n^{-2}\sum_{t=1}^n t \\ n^{-3/2}\sum_{t=1}^n \zeta_{t-1} & n^{-2}\sum_{t=1}^n \zeta_{t=1}^2 & n^{-5/2}\sum_{t=1}^n \zeta_{t-1} t \\ n^{-2}\sum_{t=1}^n t & n^{-5/2}\sum_{t=1} t\zeta_{t-1} & n^{-3}\sum_{t=1}^n t^2 \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} n^{-1/2}\sum_{t=1}^n u_t \\ n^{-1}\sum_{t=1}^n \zeta_{t-1} u_t \\ n^{-3/2}\sum_{t=1}^n t u_t \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 1 & \omega\int_0^1 J(r)\mathrm{d}r & \frac{1}{2} \\ \omega\int_0^1 J(r)\mathrm{d}r & \omega^2\int_0^1 J(r)^2\mathrm{d}r & \omega\int_0^1 rJ(r)\mathrm{d}r \\ \frac{1}{2} & \omega\int_0^1 rJ(r)\mathrm{d}r & \frac{1}{3} \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \omega\int_0^1 \mathrm{d}W(r) \\ \omega^2\left(\int_0^1 J(r)\mathrm{d}W(r) + \frac{1}{2}(1-\lambda)\right) \\ \omega\int_0^1 r\mathrm{d}W(r) \end{bmatrix} =: \mathbf{C_c}^{-1}\mathbf{D_c}.$$

Note that the asymptotic distribution in this case is independent from the value of α. However, for $\alpha \neq 0$ we can utilise $\hat{c} \rightarrow_p c$. One can use the above convergence result for hypothesis testing. Define $\mathbf{z_t} = (1, y_{t-1}, t)'$. Then, construct the t-statistic for the autoregressive coefficient as

$$t_{\hat{\rho}^*} = \frac{n(\hat{\rho}_n^* - \bar{\rho}_n)}{(n^2\hat{\sigma}_{\hat{\rho}^*}^2)^{1/2}},$$

where

$$\begin{aligned} n^2\hat{\sigma}_{\hat{\rho}^*}^2 &= n^2\hat{\omega}^2 \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \left(\sum_{t=1}^n \mathbf{z_t}\mathbf{z_t}'\right)^{-1} \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}' \\ &= \hat{\omega}^2 \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \Upsilon_n \left(\sum_{t=1}^n \mathbf{z_t}\mathbf{z_t}'\right)^{-1} \Upsilon_n \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}' \\ &\Rightarrow \omega^2 \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \mathbf{C_c}^{-1} \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}'. \end{aligned}$$

The asymptotic distribution of the t-statistic is then given by

$$
\begin{aligned}
t_{\hat{\rho}^*} &= \frac{n(\hat{\rho}^*_n - \bar{\rho}_n)}{(n^2 \hat{\sigma}^2_{\hat{\rho}^*})^{1/2}} \\
&\Rightarrow \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \mathbf{C_c}^{-1} \mathbf{D_c} \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}' \left(\omega^2 \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \mathbf{C_c}^{-1} \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}' \right)^{-1/2}.
\end{aligned}
$$

Since the asymptotic values are obtained by simulations, the results are not reduced down to scalars and are left in matrix forms for reasons of brevity. It is only important to note that the results are not divergent. Furthermore, since it is assumed that the true value of δ is zero, which is an auxiliary hypothesis which affects the asymptotic properties, it is natural to consider an OLS F test of the joint hypothesis that $\delta = 0$ and $\rho^*_n = 1$. The hypothesis is of the form $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, where

$$
\mathbf{R} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \alpha^* & \rho^* & \delta^* \end{bmatrix}', \mathbf{r} = \begin{bmatrix} 1 & 0 \end{bmatrix}'.
$$

Let $\mathbf{b} = (\hat{\alpha}^*, \hat{\rho}^*, \hat{\delta}^*)'$ and define

$$
\tilde{\boldsymbol{\Upsilon}}_{\boldsymbol{n}} = \begin{bmatrix} n & 0 \\ 0 & n^{3/2} \end{bmatrix}.
$$

The F-test is given by

$$
\begin{aligned}
F &= (\mathbf{b} - \boldsymbol{\beta})' \mathbf{R}' \left\{ \hat{\omega}^2 \mathbf{R} \left(\sum_{t=1}^{n} \mathbf{z_t} \mathbf{z_t}' \right)^{-1} \mathbf{R}' \right\}^{-1} \mathbf{R}(\mathbf{b} - \boldsymbol{\beta})/2 \\
&= (\mathbf{b} - \boldsymbol{\beta})' \mathbf{R}' \tilde{\boldsymbol{\Upsilon}}_{\boldsymbol{n}} \left\{ \hat{\omega}^2 \tilde{\boldsymbol{\Upsilon}}_{\boldsymbol{n}} \mathbf{R} \left(\sum_{t=1}^{n} \mathbf{z_t} \mathbf{z_t}' \right)^{-1} \mathbf{R}' \tilde{\boldsymbol{\Upsilon}}_{\boldsymbol{n}} \right\}^{-1} \tilde{\boldsymbol{\Upsilon}}_{\boldsymbol{n}} \mathbf{R}(\mathbf{b} - \boldsymbol{\beta})/2.
\end{aligned}
$$

Note that $\tilde{\boldsymbol{\Upsilon}}_{\boldsymbol{n}}\mathbf{R} = \mathbf{R}\boldsymbol{\Upsilon}_{\boldsymbol{n}}$, which implies that

$$F = (\mathbf{R}\boldsymbol{\Upsilon}_{\boldsymbol{n}}(\mathbf{b}-\boldsymbol{\beta}))' \left\{ \hat{\omega}^2 \mathbf{R}\boldsymbol{\Upsilon}_{\boldsymbol{n}} \left(\sum_{t=1}^{n} \mathbf{z_t z_t}' \right)^{-1} \boldsymbol{\Upsilon}_{\boldsymbol{n}}\mathbf{R}' \right\}^{-1} \mathbf{R}\boldsymbol{\Upsilon}_{\boldsymbol{n}}(\mathbf{b}-\boldsymbol{\beta})/2$$

$$\Rightarrow (2\omega^2)^{-1}(\mathbf{R}\mathbf{C_0}^{-1}\mathbf{D_0})' \{\mathbf{R}\mathbf{C_0}\mathbf{R}'\} \mathbf{R}\mathbf{C_0}^{-1}\mathbf{D_0}.$$

Just as in the previous case, for reasons of brevity, we will not multiply the matrices out.

4.4 Monte Carlo Simulations

The purpose of this section is to asses finite sample behaviour via simulations. The focus is on the theory presented in section 2. First, an assessment will be carried out between the two estimators $\tilde{c}$ and $\hat{c}$. Consequently, we provide graphs with non-parametric estimates of

$$t_{\tilde{c}} = \frac{n^{1/2}(\tilde{c}-\bar{c})}{(n^3\hat{\sigma}^2)^{1/2}} \Rightarrow N(0,1)$$

for different $\bar{c}$, n and α. We compare this with a random normal variable with mean zero and variance one. We simulate data from (4.1), with u_t being iid as $N(0,1)$, $n = \{100, 200, 300, 400\}$, $c = \{-10, -1, 0, 1\}$ and $\alpha = \{-0.5, 0, 0.5, 1\}$. The number of replications are set at $R = 10,000$. The non-parametric estimator of density that is utilised is

$$\hat{f}(x_j) = \frac{1}{hR} \sum_{i=1}^{R} K\left(\frac{x_j - x_i}{h}\right),$$

for $j = 1, \ldots, R$, where the data has been sorted in ascending order, i.e. $x_{j-1} \leq x_j$. We utilise Silverman's (1986) rule of thumb for the bandwidth, such that $h = 1.06sR^{-1.05}$, where s denotes estimated standard deviation of the data. We also use the normal kernel, such that $K(x) = (2\pi)^{-1/2}e^{-x^2/2}$. Hence, the non-parametric estimate of density becomes

$$\hat{f}(x_j) = \frac{1}{hR\sqrt{2\pi}} \sum_{i=1}^{R} \exp\left(-\frac{(x_j - x_i)^2}{2h^2}\right).$$

Figures 4.1-4.4 compare the non-parametric estimates of densities of the normalised $\tilde{c}$ and that of a normal random variable with mean zero and variance one for the cases in which $\alpha \neq 0$. For the case in which $\alpha = 0$, there is no normalisation as we know that $\tilde{c} = O_p(1)$ by Theorem 4.2.8. The estimator $\tilde{c}$ was preferred over $\hat{c}$ as typically researchers are interested in the hypothesis $\bar{c} = 0$ against $\bar{c} < 0$ and the following paragraph provides evidence that $\tilde{c}$ performs better for such values of c. For $\alpha \neq 0$ and $c \neq -10$, we can see that, pratically, there is no difference between the normalised $\tilde{c}$ and the standard normal random variable. For the case in which $c = -10$, the estimator does not perform as well due to the argument expressed in the previous paragraph - the asymptotic normality depends on how quickly the deterministic part dominates the stochastic. In fact, for $c = -10$, we have $\rho \simeq 0.9$. Thus, any finite sample from this data generating process will behave as trend stationary and the localising parameter estimator would not be able to estimate the true parameter accurately.

Given the discussion in the previous paragraph, we will focus our attention on $\alpha \neq 0$ and values of c such that the generated data resembles series with a drift. Furthermore, following Elliot *et al*, we will parameterise the error term as an MA(1) and an AR(1) process in the following way

1) iid: $u_t \sim N(0,1)$;

2) MA(1): $u_t = \eta_t + \theta\eta_{t-1}$;

3) AR(1): $u_t = \phi u_{t-1} + \eta_t$,

where $\eta_t \sim N(0,1)$, $\theta = \{-0.8, 0.8\}$ and $\phi = \{-0.5, 0.5\}$. The initial conditions for the MA(1) case are $\eta_0 \sim N(0,1)$ and for the AR(1) case $u_0 \sim N(0,1)$. Tables 4.1-4.4 depict the outcome for the $\hat{c}$ and $\tilde{c}$ estimators. Firstly, we note that on average the two estimators, with one exception, always underestimate the true parameter. This is to be expected as we know that on average the estimator of the autoregressive parameter is negatively biased in finite samples. Recall that $n^{1/2}(\hat{c}-c) = n^{3/2}(\hat{\rho}_n-\rho_n)+o_p(1)$ and $n^{1/2}(\tilde{c}-c) = n^{3/2}(\hat{\rho}_n-\rho_n)+o_p(1)$. Secondly, it should be noted that the bias and variance of the estimators diminish as the sample size increases, providing evidence on the consistency of the estimators. Interestingly, for $c \leq 0$, $\tilde{c}$ outperforms $\hat{c}$ uniformly in both bias and variance estimates. However, for $c > 0$ we can observe the converse, again uniformly. One way to explain this result would be to consider the behaviour of $\hat{\rho}$. For a fixed $\rho > 1$, Wang and Yu (2015) showed that $(\hat{\rho} - \rho)$ converges with a rate ρ^n for iid errors. Moreover, the variance of $\hat{\rho}$ on the explosive side decreases quickly and this permits $\hat{c}$ to perform better than $\tilde{c}$.

Lastly, we note that whenever $\alpha = 1$ the estimators perform best. This is to be expected since the asymptotic normality depends on the assumption that the deterministic part in (4.1) dominates the stochastic. The higher $|\alpha|$ is, everything else held constant, the quicker this happens.

4.5 Conclusion

This essay studies local-to-unit root autoregressive time series which have a drift and are driven by errors, which are allowed to be correlated over time. This permits fitting an ARMA process as an error term. The drift and autoregressive parameter estimators are shown to have a joint normal limiting distribution with rates $n^{1/2}$ and $n^{3/2}$, respectively. The variance-covariance matrix of that distribution depends on the localising parameter. It is shown that this model permits for consistent estimation of the localising parameter and an asymptotic law is obtained with a consistency rate of $n^{1/2}$. This result is obtained by taking advantage of the quicker than n consistency rate of the estimator of the autoregressive parameter. The asymptotic distribution of the estimator for c is the same as the one of the autoregressive parameter. Having a consistent estimator of c allows for the construction of t-statistics which converge to a standard normal random variable. Since the consistent estimator of c and its asymptotic normality depend on the assumption that the drift is different from zero, it is natural to consider an F-test of that hypothesis. Hence, the essay also derives the limiting distribution of the F-test under the null that the drift parameter is zero and the autoregressive parameter is unity.

In applied work, when researchers are faced with trending data they typically fit a trend as an explanatory variable in the model. Hence, the essay also derives the limiting distribution of the vector of parameters when the autoregressive parameter is of the local to unit root setup. It is interesting to note that the consistency rate of the estimator of the autoregressive parameter drops down to n. Thus, if any estimation of c is to be conducted, it should be done via the original regression without a trend included as an explanatory variable.

The paper also discusses a second linear estimator of the localising parameter and the features of those are discussed in a Monte Carlo experiment. Simulation results provide evidence that the estimators performs well in finite samples when the model is approximately linear. Since, the asymptotic normality depends on that linearity, for values of the localising parameter that are small we find that the estimators suffer in small samples.

Results such as the ones found in Theorems 4.2.1, 4.2.6 and 4.2.8 could find applications in macroeconomic settings. Empirical data on US log GDP per capita and other developed economies seem to be well approximated by a linear trend. If one is allowed to make a priori assumption of a drift in the data, the testing procedures developed in this chapter would provide as easy to implement unit root testing procedures which follow a normal limiting distribution.

4.6 Supplementary Appendix:

A central role to our proofs will play the partial sums $S_t = \sum_{j=1}^t u_j$ and the functional

$$W_n(r) = \frac{1}{n^{1/2}} S_{\lfloor nr \rfloor} = \frac{1}{n^{1/2}} S_{j-1}, \qquad \frac{j-1}{n} \leq r < \frac{j}{n}.$$

For a continuous function h under assumption 1, we have $h(W_n(r)) \Rightarrow h(\omega W(r))$ by he functional central limit and continuous mapping theorems.

Lemma 4.6.1. *Let β, γ, δ and ϕ be bounded continuous functions on $[0,1]$ and define $\xi_t = \sum_{j=1}^t \beta(t/n)\delta(j/n)u_j$. Then as $n \to \infty$*

$$(a) \quad \lim_{n\to\infty} n^{-2} \sum_{t=1}^{n} \sum_{i=1}^{t} \beta(t/n)^2 \gamma(i/n) = \int_0^1 \int_0^r \beta(r)^2 \gamma(s) ds dr;$$

$$(b) \quad \lim_{n\to\infty} n^{-3/2} \sum_{t=1}^{n} \beta(t/n)\xi_t \Rightarrow \omega \int_0^1 \beta(r)^2 \left(W(r)\delta(r) - \int_0^r W(s)\delta'(s) ds \right) dr;$$

$$(c) \quad \lim_{n\to\infty} n^{-5/2} \sum_{t=1}^{n} \sum_{i=1}^{t} \gamma(i/n)\xi_t$$

$$\Rightarrow \omega \int_0^1 \beta(r) \int_0^r \gamma(q) \left(W(r)\delta(r) - \int_0^r W(s)\delta'(s) ds \right) dq dr;$$

$$(d) \quad \lim_{n\to\infty} n^{-3/2} \sum_{t=1}^{n} \sum_{i=1}^{t} \gamma(i/n) u_t \Rightarrow \omega \int_0^1 \int_0^r \gamma(s) ds dW(r);$$

$$(e) \quad \lim_{n\to\infty} n^{-1/2} \sum_{t=1}^{n} \beta(t/n) u_t \Rightarrow \omega \int_0^1 \beta(r) dW(r).$$

Those results also hold jointly.

Proof: Let $\mu(\beta(t/n)) = \beta(t/n)^2$. Then for (a) we obtain

$$\begin{aligned}
\lim_{n\to\infty} n^{-2} & \sum_{t=1}^{n}\sum_{i=1}^{t} \beta(t/n)^2\gamma(i/n) \\
&= \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n}\mu(\beta(t/n))\sum_{i=1}^{n(t/n)}\gamma(i/n)\int_{(i-1)/n}^{i/n} ds \\
&= \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n}\mu(\beta(t/n))\sum_{i=1}^{n(t/n)}\int_{(i-1)/n}^{i/n}\gamma(s)ds + o(1) \\
&= \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n}\mu(\beta(t/n))\int_{0}^{t/n}\gamma(s)ds + o(1) \\
&= \int_0^1\int_0^r \mu(\beta(r))\gamma(s)dsdr.
\end{aligned}$$

The result follows by the definition of $\mu(\nu)$. For part (b) we have

$$\begin{aligned}
n^{-3/2}\sum_{t=1}^{n}\beta(t/n)\xi_t &= n^{-3/2}\sum_{t=1}^{n}\beta(t/n)\sum_{j=1}^{t}\beta(t/n)\delta(j/n)u_j \\
&= n^{-1}\sum_{t=1}^{n}\beta(t/n)^2\sum_{j=1}^{n(t/n)}\delta(j/n)\int_{(j-1)/n}^{j/n} dW_n(s) \\
&= \int_0^1 \mu(\beta(r))\int_0^r \delta(s)dW_n(s)dr + o_p(1).
\end{aligned}$$

We apply integration by parts to the inner integral, which is permissible since $\delta(s)$ is continuous and $W_n(s)$ is increasing and of bounded variation. This yields

$$\int_0^r \delta(s)dW_n(s) = W_n(r)\delta(r) - \int_0^r W_n(s)\delta'(s)ds. \tag{4.28}$$

By the functional central limit and continuous mapping theorems, from (4.28), we get

$$\lim_{n\to\infty} n^{-3/2}\sum_{t=1}^{n}\beta(t/n)\zeta_t \Rightarrow \omega\int_0^1 \beta(r)^2\left(W(r)\delta(r) - \int_0^r W(s)\delta'(s)ds\right)dr.$$

The proof of (c) is entirely similar to that of (a) and (b). To show (d) we write

$$\begin{aligned} n^{-3/2}\sum_{t=1}^{n}\sum_{i=1}^{t}\gamma(i/n)u_t &= \int_0^1\int_0^r \gamma(s)dsdW_n(r) + o_p(1) \\ &= \int_0^1 \phi(r)dW_n(r) + o_p(1), \end{aligned}$$

where $\phi(r) = \int_0^r \gamma(s)ds$. By integration by parts again we obtain

$$\int_0^1 \phi(r)dW_n(r) = \phi(r)W_n(r) - \int_0^1 W_n(r)\phi'(r)dr.$$

Thus,

$$\lim_{n\to\infty} n^{-3/2}\sum_{t=1}^{n}\sum_{i=1}^{t}\gamma(i/n)u_t \Rightarrow \phi(r)W(r) - \int_0^1 W(r)\phi'(r)dr.$$

Reversing the integration by parts delivers the result. The proof of (e) is entirely similar to that of (d). Joint convergence follows from the fact that every linear combination of variables on the left hand side convergences to the same linear combination of variables on the right hand side. □

Lemma 4.6.2. *Define* $g(c) = c(e^{2c}-1) - 2(e^c-1)^2$. *Then,* $\forall c \in \mathbb{R}$, $g(c) \geq 0$.

Proof: We have $g'(c) = 2ce^{2c} - 3e^c + 4e^c - 1$. At $c^* = 0$, $g'(c^*) = 0$. Consequently we have

$$\begin{aligned} &g''(c)|_{c=0} = 4e^c(ce^c - e^c + 1)|_{c=0} = 0; \\ &g^{(3)}(c)|_{c=0} = 4e^c(2ce^c - e^c + 1)|_{c=0} = 0; \\ &g^{(4)}(c)|_{c=0} = 4e^c(4ce^c + 1)|_{c=0} = 4; \\ &g(0) = 0. \end{aligned}$$

Since c^* is unique and the first nonzero derivative evaluated at c^* is even, $g(c)$ has a global minimum equal to zero showing the function is non-negative, as required. □

Bibliography

[1] Elliot, G., Rothenberg, T.J. and J.H. Stock (1996) Efficient tests for an autoregressive unit root, *Econometrica*, 64, 813-836.

[2] Hamilton, J.D. (1994) Time series analysis, Princeton University Press.

[3] Phillips, P.C.B. (1987) Towards a unified asymptotic theory for autoregression, *Biometrika*, 74, 535-547.

[4] Silvermann, B.W. (1986) Density estimation for statistics and data analysis, Chapman & Hall/CRC, London; New York.

[5] Wang, X. and J. Yu (2015) Limit theory for an explosive autoregressive process, *Economics Letters*, 126, 176-180.

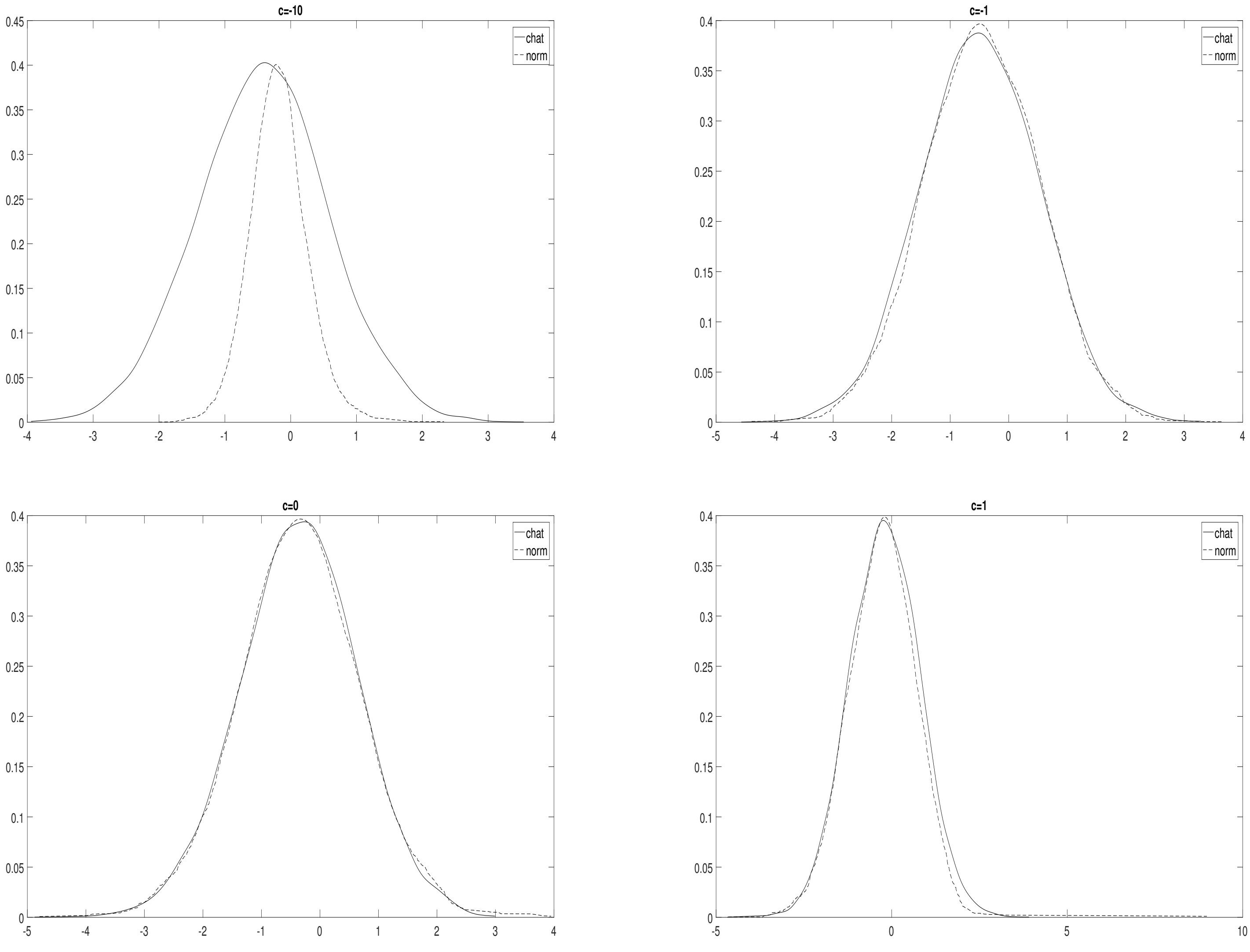

Figure 4.1: Density estimates of standardised $\hat{c}$ vs standard normal for $n = 100$ and $\alpha = -0.5$.

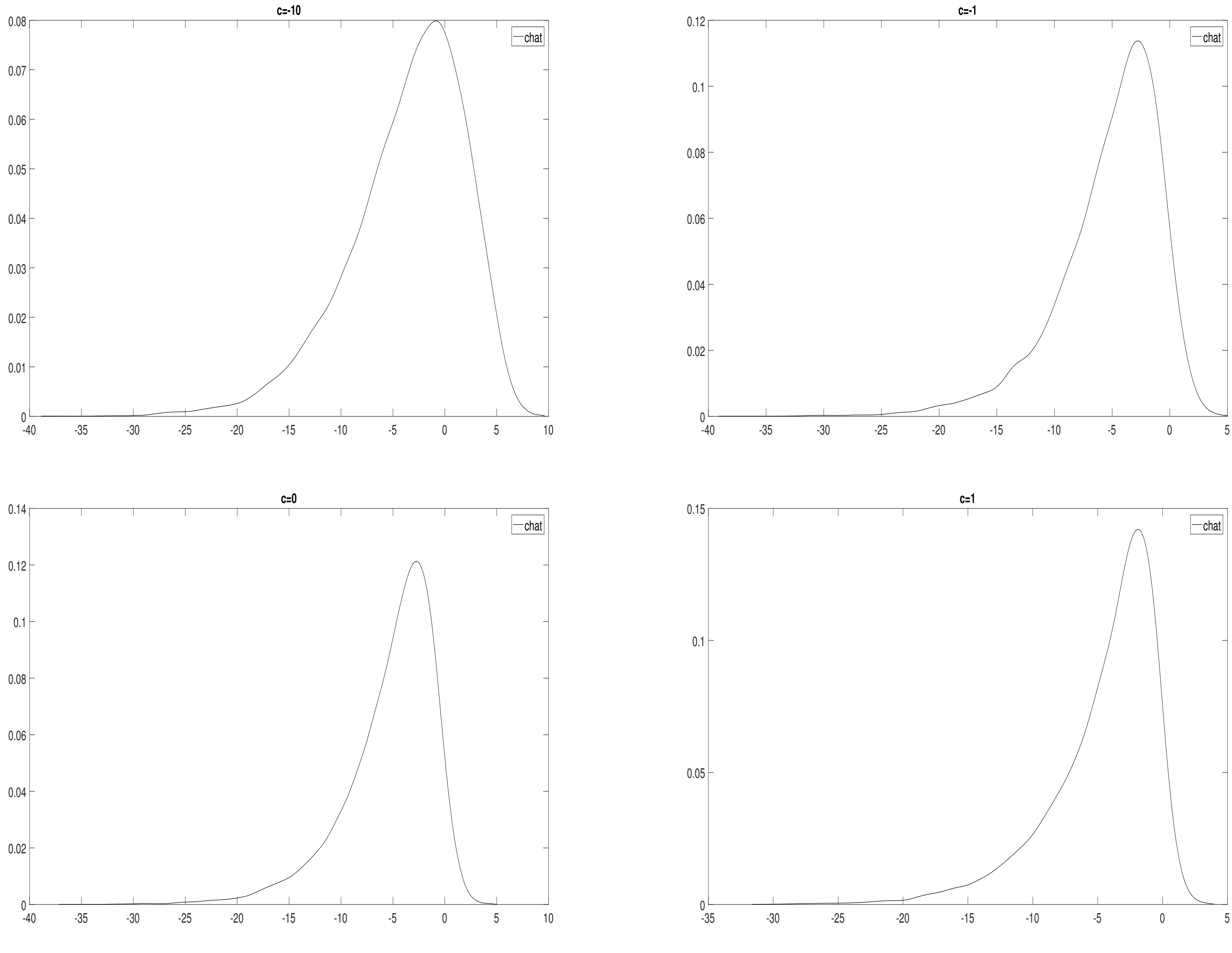

Figure 4.2: Density estimates of $\hat{c}$ for $n = 100$ and $\alpha = 0$.

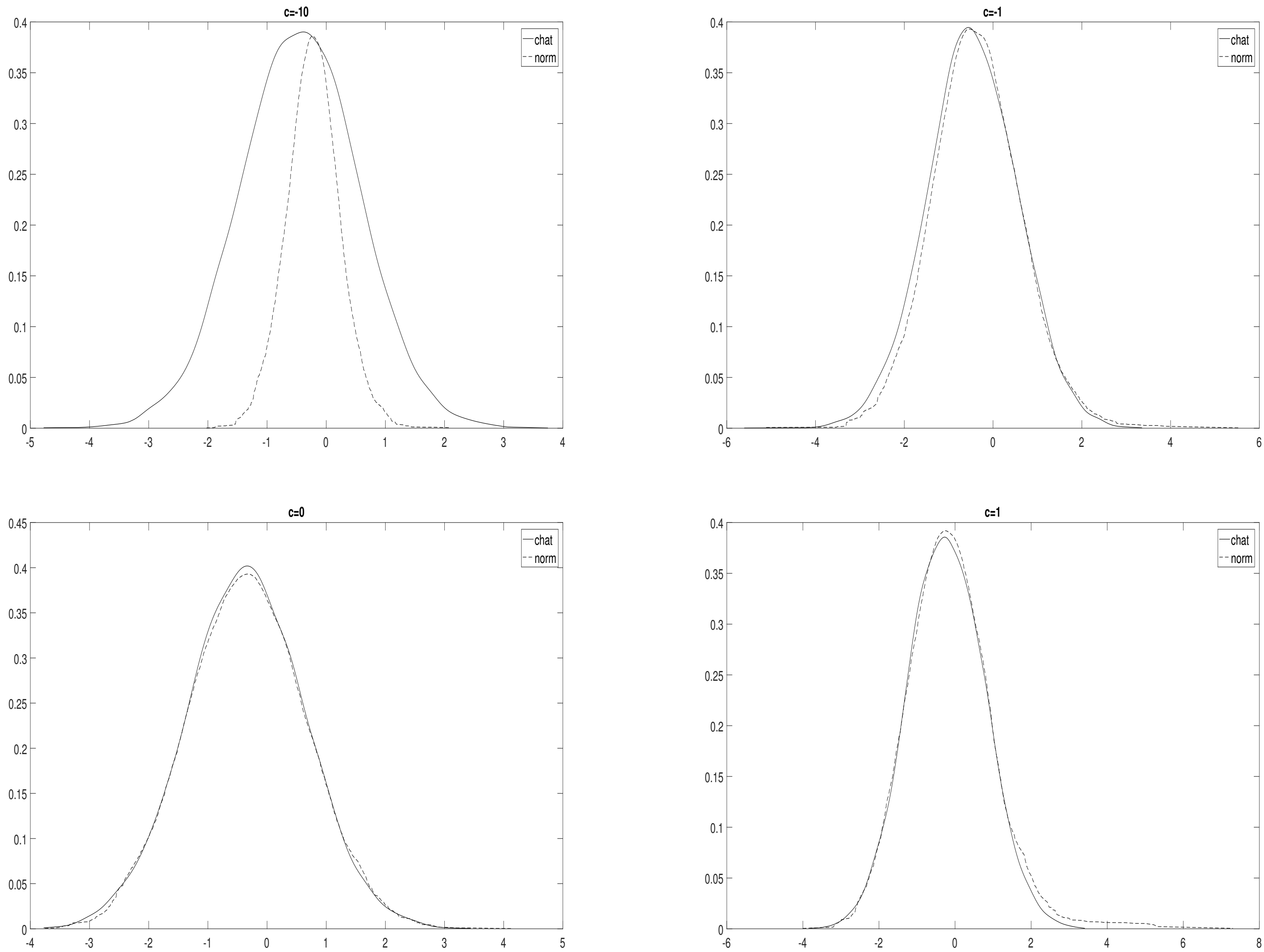

Figure 4.3: Density estimates of standardised $\hat{c}$ vs standard normal for $n = 100$ and $\alpha = 0.5$.

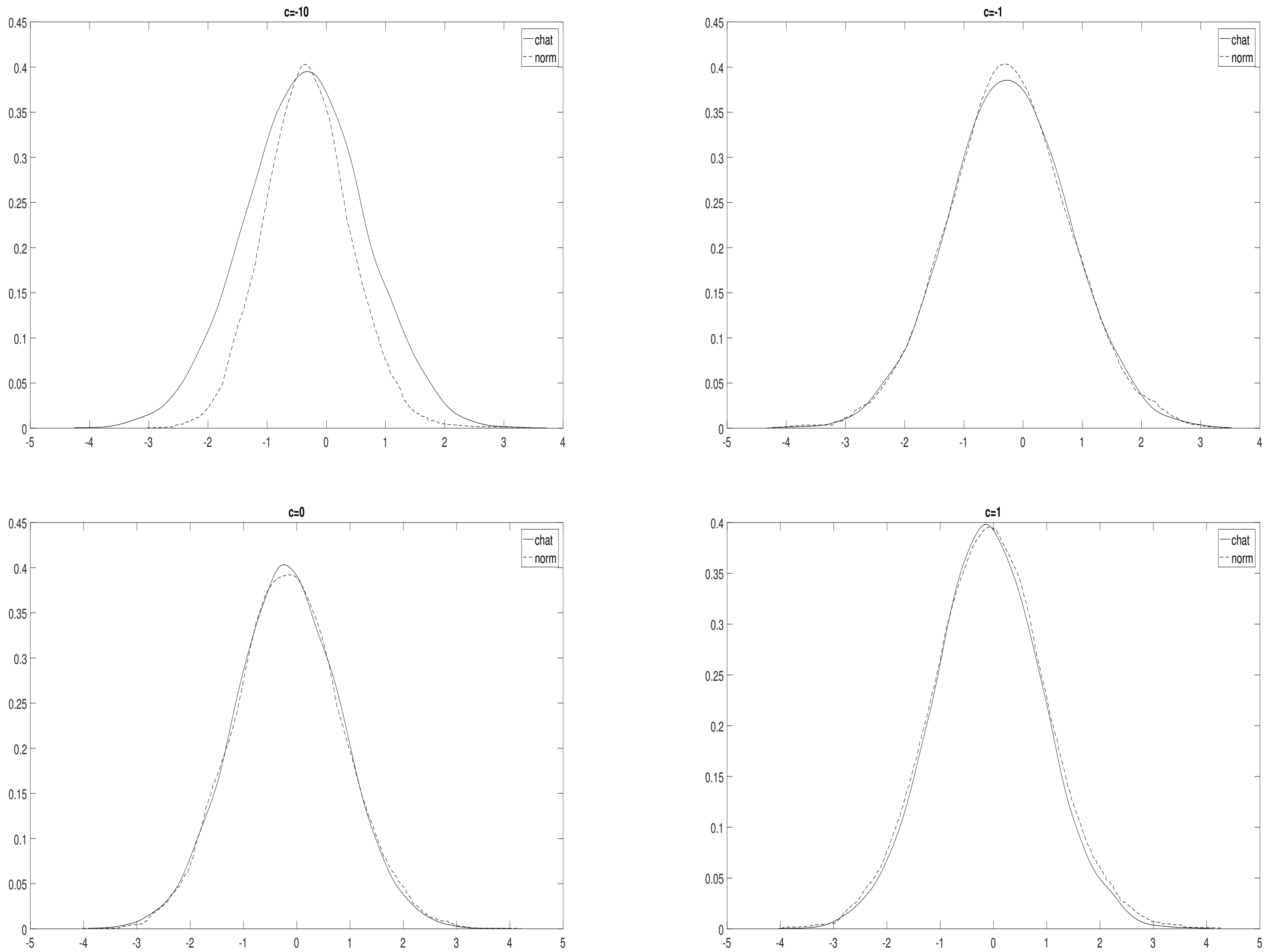

Figure 4.4: Density estimates of standardised $\hat{c}$ vs standard normal for $n = 100$ and $\alpha = 1$.

	iid		MA(1), $\theta =$				AR(1), $\phi =$			
			-0.8		0.8		-0.5		0.5	
$\alpha = -0.5$	$\hat{c}$	$\tilde{c}$	$\hat{c}$	$\tilde{c}$	$\hat{c}$	$\tilde{c}$	$\hat{c}$	$\tilde{c}$	$\hat{c}$	$\tilde{c}$
$c = -3$										
$n = 100$	-4.9949	-4.8336	-7.6218	-7.3256	-3.8616	-3.7552	-5.8002	-5.6094	-3.0437	-2.9712
$n = 200$	-4.1219	-4.0706	-5.2698	-5.1998	-3.6717	-3.6249	-4.5391	-4.4831	-3.1034	-3.0678
$n = 300$	-3.7328	-3.7060	-4.5030	-4.4690	-3.5891	-3.5603	-4.0685	-4.0392	-3.1627	-3.1390
$n = 400$	-3.5594	-3.5416	-4.1249	-4.1036	-3.4969	-3.4771	-3.8211	-3.8019	-3.1585	-3.1413
$c = 0$										
$n = 100$	-0.2602	-0.2571	-0.3958	-0.3948	-0.5632	-0.5503	-0.3310	-0.3292	-0.4720	-0.4588
$n = 200$	-0.1205	-0.1198	-0.1956	-0.1954	-0.2462	-0.2435	-0.1633	-0.1629	-0.2130	-0.2101
$n = 300$	-0.0789	-0.0786	-0.1306	-0.1306	-0.1553	-0.1543	-0.1068	-0.1067	-0.1320	-0.1307
$n = 400$	-0.0647	-0.0645	-0.0974	-0.0974	-0.0967	-0.0961	-0.0829	-0.0829	-0.0866	-0.0859
$c = 1$										
$n = 100$	0.9306	0.9358	0.8588	0.8626	0.7683	0.7762	0.8881	0.8924	0.8141	0.8228
$n = 200$	0.9636	0.9662	0.9312	0.9334	0.9386	0.9416	0.9456	0.9480	0.9519	0.9552
$n = 300$	0.9789	0.9806	0.9547	0.9562	0.9662	0.9682	0.9659	0.9675	0.9794	0.9815
$n = 400$	0.9827	0.9839	0.9654	0.9666	0.9679	0.9692	0.9729	0.9741	0.9823	0.9837

Table 4.1: Means of the estimators for $\alpha = -0.5$.

	iid		MA(1), $\theta =$				AR(1), $\phi =$			
			-0.8		0.8		-0.5		0.5	
$\alpha = -0.5$	$\hat{c}$	$\tilde{c}$	$\hat{c}$	$\tilde{c}$	$\hat{c}$	$\tilde{c}$	$\hat{c}$	$\tilde{c}$	$\hat{c}$	$\tilde{c}$
$c = -3$										
$n = 100$	8.2291	7.1892	2.8266	2.3970	6.8852	6.2223	5.5292	4.7884	5.5323	5.1104
$n = 200$	3.7482	3.5613	0.5027	0.4762	5.4140	5.1596	2.0024	1.9026	4.7622	4.5809
$n = 300$	2.2591	2.1938	0.2206	0.2139	4.5330	4.3910	1.1221	1.0897	4.3361	4.2206
$n = 400$	1.6138	1.5819	0.1276	0.1250	3.6766	3.5982	0.7980	0.7818	3.8080	3.7328
$c = 0$										
$n = 100$	0.5478	0.5437	0.0493	0.0489	2.3004	2.2319	0.2481	0.2462	2.4539	2.3968
$n = 200$	0.2555	0.2552	0.0159	0.0158	1.0060	0.9998	0.1107	0.1105	1.1231	1.1179
$n = 300$	0.1640	0.1639	0.0090	0.0089	0.5740	0.5732	0.0747	0.0746	0.7541	0.7530
$n = 400$	0.1238	0.1237	0.0065	0.0065	0.4299	0.4296	0.0551	0.0551	0.5298	0.5294
$c = 1$										
$n = 100$	0.1646	0.1679	0.0150	0.0152	0.9769	0.9761	0.0763	0.0777	1.0804	1.0822
$n = 200$	0.0844	0.0852	0.0054	0.0054	0.3361	0.3398	0.0378	0.0382	0.4096	0.4140
$n = 300$	0.0579	0.0583	0.0030	0.0031	0.2057	0.2072	0.0250	0.0251	0.2540	0.2558
$n = 400$	0.0424	0.0426	0.0021	0.0021	0.1501	0.1508	0.0186	0.0187	0.1864	0.1874

Table 4.2: Variances of the estimators for $\alpha = -0.5$.

	iid		MA(1), $\theta =$				AR(1), $\phi =$			
			-0.8		0.8		-0.5		0.5	
$\alpha = 1$	$\hat{c}$	$\tilde{c}$	$\hat{c}$	$\tilde{c}$	$\hat{c}$	$\tilde{c}$	$\hat{c}$	$\tilde{c}$	$\hat{c}$	$\tilde{c}$
$c = -3$										
$n = 100$	-3.5702	-3.4988	-4.1687	-4.0817	-3.4985	-3.4202	-3.8186	-3.7425	-3.2173	-3.1465
$n = 200$	-3.2981	-3.2692	-3.5680	-3.5362	-3.3143	-3.2817	-3.4094	-3.3796	-3.1732	-3.1421
$n = 300$	-3.1850	-3.1673	-3.3784	-3.3594	-3.2072	-3.1877	-3.2867	-3.2684	-3.0994	-3.0806
$n = 400$	-3.1405	-3.1278	-3.2811	-3.2676	-3.1808	-3.1668	-3.2127	-3.1996	-3.0839	-3.0704
$c = 0$										
$n = 100$	-0.0557	-0.0551	-0.0557	-0.0551	-0.1052	-0.1031	-0.0829	-0.0826	-0.0864	-0.0838
$n = 200$	-0.0347	-0.0346	-0.0347	-0.0346	-0.0400	-0.0395	-0.0428	-0.0427	-0.0388	-0.0382
$n = 300$	-0.0216	-0.0215	-0.0216	-0.0215	-0.0358	-0.0356	-0.0283	-0.0283	-0.0265	-0.0262
$n = 400$	-0.0160	-0.0159	-0.0160	-0.0159	-0.0242	-0.0241	-0.0204	-0.0203	-0.0168	-0.0166
$c = 1$										
$n = 100$	0.9816	0.9867	0.9662	0.9709	0.9771	0.9826	0.9754	0.9803	0.9872	0.9930
$n = 200$	0.9925	0.9950	0.9825	0.9849	0.9901	0.9927	0.9858	0.9882	0.9954	0.9980
$n = 300$	0.9954	0.9971	0.9889	0.9905	0.9953	0.9970	0.9920	0.9936	0.9994	1.0012
$n = 400$	0.9980	0.9992	0.9916	0.9928	0.9954	0.9967	0.9929	0.9942	0.9973	0.9986

Table 4.3: Means of the estimators for $\alpha = 1$.

	iid		MA(1), $\theta =$				AR(1), $\phi =$			
			-0.8		0.8		-0.5		0.5	
$\alpha = 1$	$\hat{c}$	$\tilde{c}$	$\hat{c}$	$\tilde{c}$	$\hat{c}$	$\tilde{c}$	$\hat{c}$	$\tilde{c}$	$\hat{c}$	$\tilde{c}$
$c = -3$										
$n = 100$	1.7519	1.6171	0.2723	0.2501	3.7111	3.4036	0.8497	0.7836	4.0751	3.7586
$n = 200$	0.7675	0.7412	0.0661	0.0637	2.1719	2.0893	0.3559	0.3436	2.4907	2.3983
$n = 300$	0.4868	0.4762	0.0344	0.0336	1.4785	1.4436	0.2247	0.2197	1.6955	1.6560
$n = 400$	0.3638	0.3580	0.0219	0.0216	1.1117	1.0925	0.1685	0.1658	1.2979	1.2759
$c = 0$										
$n = 100$	0.1268	0.1266	0.1268	0.1266	0.4219	0.4209	0.0572	0.0571	0.5141	0.5128
$n = 200$	0.0598	0.0598	0.0598	0.0598	0.2016	0.2015	0.0274	0.0274	0.2461	0.2460
$n = 300$	0.0407	0.0407	0.0407	0.0407	0.1304	0.1304	0.0177	0.0177	0.1651	0.1651
$n = 400$	0.0308	0.0307	0.0308	0.0307	0.0971	0.0971	0.0132	0.0132	0.1210	0.1210
$c = 1$										
$n = 100$	0.0418	0.0426	0.0038	0.0039	0.1465	0.1496	0.0190	0.0193	0.1838	0.1878
$n = 200$	0.0204	0.0206	0.0013	0.0013	0.0690	0.0697	0.0092	0.0093	0.0840	0.0849
$n = 300$	0.0140	0.0141	0.0008	0.0008	0.0458	0.0461	0.0063	0.0063	0.0587	0.0591
$n = 400$	0.0104	0.0105	0.0005	0.0005	0.0350	0.0351	0.0047	0.0047	0.0421	0.0423

Table 4.4: Variances of the estimators for $\alpha = 1$.

Chapter 5

Conclusion

This thesis has aimed to expand our understanding of autoregressive time series of order one. It consists of three essays: one on autoregressive bias, one on bias reduction and one on parameter estimation in non-stationary series with a drift.

The first essay derives the approximate bias of the OLS estimator for series with moderate deviations from a unit root and a fixed autoregressive coefficient. The result is then utilised to derive the asymptotic distribution of the indirect inference method. Those estimators are of importance as the essay shows that the bias can be substantial in small samples. Even though the mathematical machinery used to derive those results depends on the normality assumption, this is a step in the right direction of coping with the problem of bias in finite samples. Future research could possibly expand the class of models which we can derive those results for and, consequently, reduce the bias' negative effect on parameter estimation. In addition, once the literature is developed enough to cover general models of this kind, the IIE could be constructed in a completely analytical fashion without the need of simulating pseudo-datasets. This would immensely safe computational time and erase, if any, errors due to misspecifying the error distribution in the process of simulating data. This, however, would be possible for values of the autoregressive parameter sufficiently away from unity.

Essay two derives analytically the asymptotic second moments of the full-sample and each of the sub-samples estimators in the local to unit root model.

The result is used to construct an "optimal" jackknife estimator for such models. Furthermore, this permits numerical calculation of the jackknife estimator's asymptotic moments and this is used to formally explain why previous simulations find that using a higher number of sub-samples produces a smaller variance for the jackknife. The two-step estimator developed here finds applications where the analytical IIE is unusable, i.e. for values of the autoregressive parameter close to unity. It would be interesting to know if this estimator can be utilised for unit root testing. Since the localising parameter cannot be estimated consistently, one way, possibly, to circumvent this would be in the spirit of power envelope functions.

Essay three discusses parameter estimation in local to unit root autoregression with a drift. The main feature of the model is that it permits consistent estimation of the localising parameter of the autoregressive parameter. Thus, one could test the null-hypothesis of a unit root by directly constructing a t-statistic of the localising parameter being equal to zero, which is shown to have a standard normal limiting distribution. These results could straightforwardly be utilised in applied work by finding a suitable dataset which satisfies the assumptions of the model. Suitable candidates involve log GDP per capita of developed countries. One way to improve on those results would be to seek more efficient estimates of the parameters and, thus, be able to improve on any unit root testing procedure.

The results from Chapters 2 and 3 are important not only from a theoretical perspective but also from practitioners' point of view. The OLS estimator is the best unbiased linear estimator once the Gauss-Markov assumptions are satisfied. However, the strict exogeneity assumption in autoregressive time series fails which results in OLS being biased. This leads to a failure of the estimator's optimal properties in finite samples. It turns out that it is possible to construct a jackknife estimator which outperforms OLS in mean and variance dimensions in finite samples. This could potentially be employed in models where forecasting is the main goal since estimators with smaller bias and variance would typically be preferred in such settings.

Extending results, such as the ones in this thesis, for autoregressive series of higher order is possible but far from straightforward. Deriving an approximate bias for general autoregressive series requires an immense algebraic effort. In terms of extending the jackknife results from chapter 3 in such fashion, this could possibly be achieved by assuming one root on the unit circle and the rest bigger the unity. Then, asymptotic distributions for the autoregressive parameter could follow in the spirit of the augmented Dickey Fuller test and asymptotic moments would be the same as the ones derived in chapter 3. Those and other questions that stem from this thesis provide exciting opportunities for future research.