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

Marian Z. Stoykov

University of Essex

March 2017

Abstract

This paper 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 stationary, non-stationary and explosive series.

Keywords: Autoregressive bias; moment generating function.

Address for correspondence: Marian Z. Stoykov, Essex Business School, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, England; e-mail: mzstoy@essex.ac.uk
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, etc., 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.

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 as 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 unlimited samples, 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 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 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 length. For the non-
stationary case, Phillips (1987) 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) only derives 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 as 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 $c \to \infty$.

Recent development in the asymptotic theory of AR processes with an autoregressive coefficient
given by $\alpha_T = 1 + c/k_T$, where $k_T$ is a sequence which increases to infinity at a rate slower
than $T$, such that $k_T = o(T)$ as $T \to \infty$, has been made by Phillips and Magdalinos (2007)
(hereafter, PM). The OLS estimator of the autoregressive parameter in this case of moderate
deviations from a unit root follows an asymptotic distribution which is equivalent to the one
that is obtained by considering a fixed autoregressive parameter. This framework was utilised
by Phillips, Wu and Yu (2011) to test the NASDAQ index for explosive behaviour. Given those
recent advancements in the literature it would be of interest to derive the bias of the OLS
estimator for such processes.

The paper is organised as follows: section 2 summarises the main results and provides a
discussion. Section 3 concludes, and the technical details 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 $\alpha_T$ is dropped out for notational simplicity; the symbol $\sum$ will be used for summations
running from $t = 1$ to $T$; $W(r)$ will be used to denote a Wiener process on $C[0,1]$, the space of
continuous real-valued functions on the unit interval and $\Rightarrow$ denotes convergence in distribution.
2 Main Results

Suppose $x_t$ is given by the following stochastic difference equation

$$x_t = \alpha_T x_{t-1} + u_t, \quad \alpha_T = 1 + \frac{c}{k_T} \quad \text{for } t = 1, \ldots, T, \quad (1)$$

where $u_t$ is identically and independently distributed $N(0, \sigma^2)$ and $k_T$ is a sequence that increases to infinity, such that $k_T = o(T)$ as $T \to \infty$. The distribution of $x = (x_1, \ldots, x_T)$ in (1) is uniquely determined by specifying an initial condition for the process. The density function of $x$ for a constant initial condition, $x_0 = \gamma$ is given by

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

If the initial condition is specified not as a constant but as a random variable given by $x_0 \sim N(0, \frac{\sigma^2}{1-\alpha^2})$ the density of $x$ becomes

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

The model which this paper 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. 118-123) provides an excellent treatment on the subject. In particular, we will take $\gamma = 0$. The MLE for the constant initial condition, which coincides with the OLS estimator, is given by

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

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 $\sigma^2$. Following the procedure of White, define $U = \sum x_t x_{t-1}$ and $V = \sum x_{t-1}^2$ such that their
The joint MGF is given by

\[ \mathbb{E}\exp(Uu + Vv) = m(u, v) = \int_{-\infty}^{\infty} \exp(Uu + Vv)f(x)dx \]

\[ = (2\pi)^{-T/2} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \left[ -2(\alpha + u)U + (1 + \alpha^2 - 2v)V + x^2_T \right] \right\} dx \]

\[ = (2\pi)^{-T/2} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} xDx' \right\} dx = |D_T(u, v)|^{-1/2}, \tag{2} \]

where the second line follows from

\[ f(x) = (2\pi)^{-T/2} \exp \left\{ -\frac{1}{2} \left[ V + x^2_T - 2\alpha U + \alpha^2 V \right] \right\} \]

for \( \sigma^2 = 1 \) and

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

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

\[ \mathbb{E}\left( \hat{\alpha} - \alpha \right) = \int_{0}^{\infty} \frac{\partial}{\partial \alpha} |D_T(-v)|^{-1/2} dv, \tag{3} \]

where \( |D_T(-v)| \) is the determinant of the matrix evaluated at \( u = 0 \) and \(-v\). Now the right hand side of (3) can be utilised to derive the bias (the details are given in the appendix). The results are summarised in the following theorem.

**Theorem 1.** For the model considered in (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

\[ \mathbb{E}(\hat{\alpha}T - \alpha T) = \begin{cases}
  -\frac{2\alpha_T}{T} \left\{ 1 + O \left( \max \left( \frac{k_T}{T}, k_T^{-1} \right) \right) \right\} \left[ 1 + O \left( T^{-1} \right) \right], & c < 0, \\
  -\frac{1.7814}{T} \left[ 1 + O \left( T^{-1} \right) \right], & c = 0, \\
  -\frac{2\pi^{3/2} \sqrt{T}}{k_T^{3/2} \alpha_T T + 1} \left\{ 1 + O \left( \max \left( \frac{k_T}{T}, k_T^{-1} \right) \right) \right\} \left[ 1 + O \left( T^{-1} \right) \right], & c > 0.
\end{cases} \]
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 $\alpha$ the first order term of the expansion is $-2\alpha / T$ (e.g. White). 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 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 $\delta \in [1/2, 1)$, $\max\left(k_T/T, k_T^{-1}\right) = k_T/T = T^{\delta-1}$.

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 the expectation of the functional $\int_0^1 W(r) dW(r) / \int_0^1 W(r)^2 dr$, which is the leading term of the asymptotic bias for the unit root process (see Phillips (1987)).

Remark 3. For $c > 0$, which corresponds to $|\alpha| > 1$, the first term of the expansion is $-2\pi^{1/2} \alpha^{3/2} T^{1/2} k_T^{-3/2} \alpha_T^{-T+1}$. From Le Breton and Pham, the respective term from the expansion for a fixed $\alpha$ is $-2^{-1/2} \pi^{1/2} (\alpha^2 - 1)^{3/2} T^{1/2} \alpha_T^{-T+1}$. Note that $\alpha_T^2 - 1 = (1 + c/k_T)^2 - 1 = 2c/k_T (1 + O(k_T^{-1}))$. 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 $O(\alpha^{-T})$ higher in magnitude than the first. One would have expected that the order would be a multiple of $\alpha^{-T}$ 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 1 from the present paper, namely, the limits $c \nearrow 0$ and $c \searrow 0$ are not identical. This is not surprising as the result from this paper and the of Le Breton and Pham are asymptotically equivalent up to the first term. The solid line in Figure 1 depicts the result with $T = 24$.
Figure 1: Bias: Analytical Solution.

$k_T = \sqrt{T}$ and $c \in [-\sqrt{T}, 3]$, where the lower limit of $c$ was chosen as $-\sqrt{T}$ to match with $\alpha = 1 + c/\sqrt{T} = 0$. The lonely dot is the result for $c = 0$. The graph also shows the simulated OLS bias for the autoregressive coefficient of (1) with $x_0 = 0$ and $\sigma = 1$ for comparison. The number of observations is $T = 24$ 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.

**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 defined 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

\[
T^{1/2+\delta/2} (\hat{\alpha}_T - \alpha_T) \Rightarrow N(0, -2c) \quad \text{for } c < 0, \quad (4)
\]

\[
\frac{T^{\delta} \alpha_T}{2c} (\hat{\alpha}_T - \alpha_T) \Rightarrow C \quad \text{for } c > 0, \quad (5)
\]

where $C$ is the standard Cauchy random variable. In comparison, it is well-known that for the
model considered in (1) and a fixed α

\[ T^{1/2} (\hat{\alpha} - \alpha) \Rightarrow N \left( 0, 1 - \alpha^2 \right) \quad \text{for } |\alpha| = |1 + c| < 1, \tag{6} \]

\[ \frac{\alpha^T}{\alpha^2 - 1} (\hat{\alpha} - \alpha) \Rightarrow C \quad \text{for } \alpha = 1 + c > 1, \tag{7} \]

where the result in (7) is due to White (1958). From (4) and (6) there is a discrepancy between the terms \(-2c\) and \(1 - \alpha^2 = -2c - c^2\), and from (5) and (7) between \(2c\) and \(\alpha^2 - 1 = 2c + c^2\).

Continuity can be achieved by substituting \(c\) by \(c + c^2/2T\delta\) 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 \(T \to \infty\)

\[ T \mathbb{E}(\hat{\alpha} - \alpha) \text{ converges to } -2\alpha \quad \text{for } |\alpha| = |1 + c| < 1 \tag{8} \]

\[ T^{-1/2}\alpha^T \mathbb{E}(\hat{\alpha} - \alpha) \text{ converges to } -2^{-1/2}\pi^{1/2}\alpha^{-1}(\alpha^2 - 1)^{3/2} \quad \text{for } \alpha = 1 + c > 1. \tag{9} \]

For \(c < 0\), Theorem 1 and Theorem 3.1 from Le Breton and Pham produce the same result as \(\delta \to 0\) without any adjustment. However, for \(c > 0\) Theorem 1 and (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 \(kT\) with 1 at the beginning of the integral expressions from the appendix and is stated as a corollary to Theorem 1.

**Corollary 1.** 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

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

There is nothing surprising regarding this result given the discussion around Theorem 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
δ → 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 $\alpha^2 - 1$ for the stationary or explosive cases respectively.

Remark 6. One way to try and utilise the result in Theorem 1 is to get an estimate of the bias and subtract it from the OLS estimator. MacKinnon and Smith (1998) propose a procedure to reduce bias in autoregressive series with an intercept, which utilises a result similar to the one in Theorem 1 for a fixed $\alpha$ parameter. Another estimator which employs a similar idea is the jackknife estimator. Although the jackknife utilises sub-samples, it requires an asymptotic expression for the bias of the original estimator for the full-sample and a number of sub-samples such as the one provided in Theorem 1. Chambers (2013) and Chambers and Kyriacou (2013) show how bias reduction can be achieved in stationary and non-stationary autoregressive series respectively. Theorem 1 shows that the analysis of Chambers can directly be translated to mildly stationary series. 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 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 (Kaufmann and Kruse (2015) provide an interesting comparison between a number of estimators) or by considering local to unit root alternatives (Chambers and Kyriacou (2016) and Stoykov (2016)), where the approximate bias function has been shown to be continuous at $c = 0$ (Phillips (2012)).

3 Conclusion

This paper 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 given by $\alpha_T = 1 + c/k_T$, 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 stationary, non-stationary and explosive series respectively. The result is also discontinuous in
that the limits $c \nearrow 0$ and $c \searrow 0$ are not identical in the final bias expression.

Acknowledgements

I am grateful to my supervisor Marcus Chambers for useful comments and suggestions on this work. I would also like to thank the Essex Business School for scholarship support.

A Appendix

To evaluate the determinant in (3) we note that it can be written in the form of a second order difference equation. Define $\kappa \equiv p(-v) = 1 + \alpha^2 + 2v$ such that

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

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

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

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

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

Phillips (2012) showed that $|D_T(-v)|$ is positive for all $v > 0$ meaning the integral in (3) is well-defined for all $\alpha$. Taking the derivative in (3) leads to

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

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

$$x = \frac{2}{\kappa + \sqrt{\kappa^2 - 4\alpha^2}} = \frac{\kappa - \sqrt{\kappa^2 - 4\alpha^2}}{\kappa^2 - (\kappa^2 - 4\alpha^2)} = \frac{\kappa - \sqrt{\kappa^2 - 4\alpha^2}}{2\alpha^2},$$
from which it follows that
\[ \kappa - \sqrt{\kappa^2 - 4\alpha^2} = 2\alpha^2 x. \quad (A.2) \]

From (A.1) and (A.2) 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 \( \kappa = \frac{1}{x} + \alpha^2 x \). From \( \kappa = 1 + \alpha^2 + 2\nu \) we can solve for \( \nu \) as a function of \( x \)
\[ \nu = \frac{1}{2} \left\{ \frac{1}{x} + \alpha^2 x - \alpha^2 - 1 \right\} = \frac{(1 - x)(1 - \alpha^2 x)}{2x}. \]

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

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

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 (3) can be changed. Applying the change of variable yields
\[ \frac{\lambda - \alpha^2}{\lambda^2 - \alpha^2} = \frac{\frac{1}{x} - \alpha^2}{\frac{1}{x}^2 - \alpha^2} = \frac{x(1 - \alpha^2 x)}{1 - \alpha^2 x^2}, \]
\[ 1 - \frac{\lambda - \alpha^2}{\lambda^2 - \alpha^2} = 1 - \frac{x(1 - \alpha^2 x)}{1 - \alpha^2 x^2} = \frac{1 - x}{1 - \alpha^2 x^2}, \]

leading to
\[ |D_T(-v)| = \frac{1 - \alpha^2 x}{1 - \alpha^2 x^2} \frac{1}{x^{T-1}} + \frac{1 - x}{1 - \alpha^2 x^2} \alpha^{2T} x^T \]
\[ = \begin{cases} \frac{1 - \alpha^2 x + (1 - x)\alpha^{2T} x^{2T-1}}{(1 - \alpha^2 x^2)x^{T-1}} = \frac{A_T(x; \alpha)}{(1 - \alpha^2 x^2)x^{T-1}}, & \text{for } |\alpha| \leq 1, \\ \frac{\alpha^2 x - 1 + (x - 1)\alpha^{2T} x^{2T-1}}{(\alpha^2 x^2 - 1)x^{T-1}} = \frac{B_T(x; \alpha)}{(\alpha^2 x^2 - 1)x^{T-1}}, & \text{for } |\alpha| > 1, \end{cases} \quad (A.3) \]
where \( A_T(x; \alpha) = 1 - \alpha^2 x + (1 - x)\alpha^{2T} x^{2T-1} \) and \( B_T(x; \alpha) = \alpha^2 x - 1 + (x - 1)\alpha^{2T} x^{2T-1} \). Thus, for \(|\alpha| \leq 1\), by changing the variable of integration from \( v \) to \( x \) and by taking the derivative in (3) we have

\[
\mathbb{E}(\hat{\alpha} - \alpha) = \frac{\partial}{\partial \alpha} \int_0^\infty D_T(-v)^{-1/2} dq = \frac{\partial}{\partial \alpha} \int_0^1 \left[ \frac{A_T(x; \alpha)}{(1 - \alpha^2 x^2) x^{1-1}} \right]^{-1/2} \left[ -\frac{1 - \alpha^2 x^2}{2x^2} \right] dx \tag{A.4}
\]

Evaluating the derivative in (A.4) gives

\[
\frac{\partial}{\partial \alpha} \int_0^1 x^{\frac{T-5}{2}} (1 - \alpha^2 x^2)^{3/2} A_T(x; \alpha)^{-1/2} dx = \frac{3}{2} \int_0^1 x^{\frac{T-5}{2}} (1 - \alpha^2 x^2)^{1/2} (-2\alpha x^2) A_T(x; \alpha)^{-1/2} dx \tag{A.5}
\]

The derivative from the last line in (A.5) is

\[
\frac{\partial}{\partial \alpha} A_T(x; \alpha) = \frac{\partial}{\partial \alpha} (1 - \alpha^2 x + (1 - x)\alpha^{2T} x^{2T-1}) = -2\alpha x + 2T(1 - x)\alpha^{2T-1} x^{2T-1}. \tag{A.6}
\]

Combining (A.4), (A.5) and (A.6), for \( c < 0 \), corresponding to \(|\alpha| < 1\), we have

\[
\mathbb{E}(\hat{\alpha} - \alpha) = -\frac{3\alpha}{2} \int_0^1 x^{\frac{T-1}{2}} (1 - \alpha^2 x^2)^{1/2} A_T(x; \alpha)^{-1/2} dx + \frac{\alpha}{2} \int_0^1 x^{\frac{T-3}{2}} (1 - \alpha^2 x^2)^{3/2} A_T(x; \alpha)^{-3/2} dx \tag{A.7}
\]

To evaluate the expectation in (A.7), we start with the first integral. Setting \( y = x^{2T-1} \) gives \( dy = (2T - 1)x^{2T-2} dx = (2T - 1)y^{2T-2} dx \). Note that \( A_T(x; \alpha) = 1 - \alpha^2 x + O(\alpha^{2T}) \). By substituting \( \alpha = 1 + c/k_T \) in the third line of the following derivations and using the facts that \( y^{b/a} = 1 + \frac{b}{2T-a} \log y + O(T^{-2}) \), \( k_T = o(T) \) and \( \frac{k_T}{T} \alpha^{2T} = o(1) \) as \( T \to \infty \) the integral
becomes

$$
\int_0^1 x^{T-1} (1 - \alpha^2 x^2)^{1/2} A_T(x; \alpha)^{-1/2} dx
= \frac{1}{2T - 1} \int_0^1 y^{-3T+1} \left\{ \frac{1 - \alpha^2 y^{2T-1}}{1 - \alpha^2 y^{2T-1} + (1 - y^{2T-1})y\alpha^{2T}} \right\}^{1/2} dy
\approx \frac{1}{2T} \int_0^1 y^{-3} \left\{ \frac{1 - \left(1 + \frac{2c}{k_T} + \frac{c^2}{k_T^2}\right) (1 + \frac{2c}{k_T} \log y + O(T^{-2}))}{1 - \left(1 + \frac{2c}{k_T} + \frac{c^2}{k_T^2}\right) (1 + \frac{1}{2T} \log y + O(T^{-2})) + O(\alpha^{2T})} \right\}^{1/2} dy \left[1 + O(T^{-1})\right]
\approx \frac{1}{2T} \int_0^1 y^{-3} \left\{ \frac{-2\log y - \frac{2c}{k_T} - \frac{c^2}{k_T^2} y\log y}{\frac{2c}{k_T} \log y + \frac{4c}{k_T} + \frac{c^2}{k_T^2} + \frac{k_T^2 y\log y}{\frac{2c}{k_T} \log y + \frac{4c}{k_T} + \frac{c^2}{k_T^2} + \frac{k_T^2 y\log y}} \right\}^{1/2} dy \left[1 + O(T^{-1})\right]
\approx \frac{1}{2T} \int_0^1 y^{-3} \left\{ \frac{4c \left[1 + O\left(\max\left(\frac{k_T}{T}, k_T^{-1}\right)\right)\right]}{4c \left[1 + O\left(\max\left(\frac{k_T}{T}, k_T^{-1}\right)\right)\right]} \right\}^{1/2} dy \left[1 + O\left(\max\left(\frac{k_T}{T}, k_T^{-1}\right)\right)\right] \left[1 + O(T^{-1})\right]
= \frac{1}{2T} \int_0^1 y^{-3} dy \left[1 + O\left(\max\left(\frac{k_T}{T}, k_T^{-1}\right)\right)\right] \left[1 + O(T^{-1})\right].
$$

The second integral can be dealt with in the same fashion

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

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

$$
E(\hat{\alpha} - \alpha) = \left[ -\frac{3\alpha}{2} \frac{1}{2T} \int_0^1 y^{-3} dy + \alpha \frac{1}{2} \frac{1}{2T} \int_0^1 y^{-3} dy \right] \left[1 + O\left(\max\left(\frac{k_T}{T}, k_T^{-1}\right)\right)\right] \left[1 + O(T^{-1})\right]
= -\frac{2\alpha}{T} \left[1 + O\left(\max\left(\frac{k_T}{T}, k_T^{-1}\right)\right)\right] \left[1 + O(T^{-1})\right].
$$

(A.8)
For \( c = 0 \), corresponding to \(|\alpha| = 1\), the algebra reduces to the one found in Phillips (2012). From (A.5) as \( T \to \infty \) we have

\[
\mathbb{E}(\hat{\alpha} - 1) = -\frac{3}{2} \int_0^1 x^\frac{T-1}{2} \left\{ \frac{1 + x}{1 + x^{2T-1}} \right\}^{1/2} dx + \frac{1}{2} \int_0^1 x^\frac{T-3}{2} \left\{ \frac{1 + x}{1 + x^{2T-1}} \right\}^{3/2} dx - \frac{T}{2} \int_0^1 x^\frac{T-7}{2} \left\{ \frac{1 + x}{1 + x^{2T-1}} \right\}^{3/2} (1 - x) dx.
\]

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

\[
\int_0^1 x^\frac{T-1}{2} \left\{ \frac{1 + x}{1 + x^{2T-1}} \right\}^{1/2} dx = \frac{1}{2T} \int_0^1 y^{-\frac{3}{4}} \left\{ \frac{1 + y^{1/2}}{1 + y} \right\}^{1/2} dy [1 + O (T^{-1})] = \frac{2^{1/2}}{2T} \int_0^1 y^{-\frac{3}{4}} \{1 + y\}^{-1/2} dy [1 + O (T^{-1})].
\]

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

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

and the third is given by

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

Combining the three integrals yields

\[
\mathbb{E}(\hat{\alpha} - 1) = -\frac{32^{1/2}}{2T} \int_0^1 y^{-\frac{3}{4}} \{1 + y\}^{-1/2} dy [1 + O (T^{-1})] + \frac{12^{3/2}}{2T} \int_0^1 y^{-3/4} \{1 + y\}^{-3/2} dy [1 + O (T^{-1})] - \frac{1}{2} \left( -\frac{2^{3/2}}{4T} \int_0^1 y^{1/4} \{1 + y\}^{-3/2} \log y dy [1 + O (T^{-1})] \right).
\]

Evaluating the expression numerically yields

\[
\mathbb{E}(\hat{\alpha} - 1) = -\frac{1.7814}{T} + O (T^{-2}).
\]
Lastly, consider the case in which $c > 0$, corresponding to $|\alpha| > 1$. From (A.3) 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 (3) and (A.3), we have that for $c > 0$

$$E(\hat{\alpha} - \alpha) = \frac{\partial}{\partial \alpha} \int_0^\infty D_T(-v)^{-1/2} dq \int_1^\infty \left[ \frac{B_T(x; \alpha)}{(\alpha^2 x^2 - 1)x^{T-1}} \right]^{-1/2} \left[ \frac{\alpha^2 x^2 - 1}{2x^2} \right] dx$$

$$= \frac{1}{2} \frac{\partial}{\partial \alpha} \int_1^\infty x^{T-5} (\alpha^2 x^2 - 1)^{3/2} B_T(x; \alpha)^{-1/2} dx. \tag{A.10}$$

Evaluating the derivative in (A.10) gives

$$\frac{\partial}{\partial \alpha} \int_1^\infty x^{T-5} (\alpha^2 x^2 - 1)^{3/2} B_T(x; \alpha)^{-1/2} dx$$

$$= 3 \int_1^\infty x^{T-1} (\alpha^2 x^2 - 1)^{3/2} B_T(x; \alpha)^{-1/2} 2\alpha x^2 dx$$

$$- \frac{1}{2} \int_1^\infty x^{T-5} (\alpha^2 x^2 - 1)^{3/2} B_T(x; \alpha)^{-3/2} \frac{\partial}{\partial \alpha} B_T(x; \alpha) dx. \tag{A.11}$$

The derivative from the last line in (A.11) is

$$\frac{\partial}{\partial \alpha} B_T(x; \alpha) = 2\alpha x + 2T(x - 1)\alpha^{2T-1}x^{2T-1}. \tag{A.12}$$

Combining (A.10), (A.11) and (A.12), for $c > 0$ we have

$$E(\hat{\alpha} - \alpha) = \frac{3\alpha}{2} \int_1^\infty x^{T-1} (\alpha^2 x^2 - 1)^{1/2} B_T(x; \alpha)^{-1/2} dx$$

$$- \frac{\alpha}{2} \int_1^\infty x^{T-3} (\alpha^2 x^2 - 1)^{3/2} B_T(x; \alpha)^{-3/2} dx$$

$$- \frac{2T\alpha^{2T-1}}{2} \int_1^\infty x^{T-7} (\alpha^2 x^2 - 1)^{3/2} B_T(x; \alpha)^{-3/2} (x - 1) dx. \tag{A.13}$$

Note that $B_T(x; \alpha) = \alpha^2 x - 1 + (x - 1)\alpha^{2T}x^{2T-1}$. Therefore, the first and second integrals in (A.13) become exponentially small due to the $\alpha^{2T} = e^{\frac{2cT}{x^2}} \{1 + O(k^{-1})\}$ term which explodes as $T \to \infty$. As a result, the bias for $c > 0$ is determined by the third integral which, as $T \to \infty$,
b~\frac{\alpha^{2T-1}}{2(2T-1)^2} \int_1^\infty y^{\frac{1}{2T-1}} \left\{ \frac{\alpha^2 y^{\frac{2T-1}{2}} - 1}{(y^{\frac{1}{2T-1}} - 1)(y^{\frac{2T-1}{2}} - 1)} \right\}^{3/2} \log y dy + [O(T^{-1})]
\]
\]
\]
\]
\]
\]
\]
\]
\]
\]
\]
References


