On the Effect of Deterministic Terms on the Bias in Stable AR Models

Noud P.A. van Giersbergen

On the Effect of Deterministic Terms on the Bias in Stable AR Models

Noud P.A. van Giersbergen*
Department of Quantitative Economics,
Universiteit van Amsterdam
Roetersstraat 11
1018 WB Amsterdam
The Netherlands
E-mail: N.P.A.vanGiersbergen@uva.nl

June 30, 2004

Abstract

This paper compares the first-order bias approximation for the autoregressive (AR) coefficients in stable AR models in the presence of deterministic terms. It is shown that the bias due to inclusion of an intercept and trend is twice as large as the bias due to an intercept. For the AR(1) model, the accuracy of this approximation is investigated by simulation.

Keywords: Autoregressive models, estimation bias, large sample asymptotics, Nagar expansions
JEL classification: C13; C22.

1 Introduction

Autoregressive (AR) models are widely used to model the dynamic properties of time series data. Their popularity stems from the fact that (i) they are easy to estimate and (ii) they have the flexibility to accurately approximate autoregressive moving average processes. Since many economic time series exhibit trending behavior, deterministic terms like a trend are often incorporated in the estimation. In this paper, we investigate the effect on the estimation bias for the AR coefficients when an intercept and trend are added to the estimation model. Although the literature on bias in AR models has a relatively long history going back to Barlett (1946), to the best of our knowledge, only the recent paper by Kang *et al.* (2003) has focused on the bias effect of a linear time trend in stable AR models.

Consider the following stable AR(k) models

$$\mathcal{M}^p \quad : \quad y_t = \rho_1 y_{t-1} + \ldots + \rho_k y_{t-k} + \varepsilon_t, \tag{1a}$$

$$\mathcal{M}^c \quad : \quad y_t = \rho_1 y_{t-1} + \ldots + \rho_k y_{t-k} + \mu + \varepsilon_t \tag{1b}$$

$$\mathcal{M}^{ct} \quad : \quad y_t = \rho_1 y_{t-1} + \ldots + \rho_k y_{t-k} + \mu + \gamma t + \varepsilon_t \tag{1c}$$

where $\varepsilon_t \sim \mathrm{NID}(0, \sigma_\varepsilon^2)$ for $t = 1, ..., T$. The roots of the characteristic polynomial

$$\phi(z) \equiv 1 - \rho_1 z - \ldots - \rho_k z^k \tag{2}$$

are assumed to lie outside the unit circle. The parameters of interest are the ρ_i's or functions thereof. In model $\mathcal{M}^c$, the AR process is allowed to have a non-zero mean, while in model $\mathcal{M}^{ct}$ the AR process is stationary around a linear trend. Although normality of the innovations is stronger than what is necessary for establishing the results, the assumption is made for ease of exposition; see for instance Bhansali (1981) for a set of less restrictive assumptions. The starting values are assumed to come from the asymptotic stationary distribution, although this is not a critical assumption since the starting values do not influence the first-order bias (provided that they are finite).

As is well known, the estimation bias aggravates when an intercept is included to the estimation model. For instance, in the AR(1) case, Kendall (1954) and Marriot and Pope (1954) have shown that the first-order bias in the pure AR(1) model is equal to

$$\mathbb{E}[\hat{\rho}_1 - \rho_1 | \mathcal{M}^p] \overset{1}{=} \frac{-2\rho_1}{T}, \tag{3}$$

while if an intercept is present the bias becomes

$$\mathbb{E}[\hat{\rho}_1 - \rho_1 | \mathcal{M}^c] \overset{1}{=} \frac{-2\rho_1}{T} - \frac{(1+\rho_1)}{T}; \tag{4}$$

here and elsewhere in the paper $\overset{1}{=}$ indicates equality up to order $o_p(T^{-1})$. We shall show that when the models includes an intercept and linear trend, the bias is equal to

$$\mathbb{E}[\hat{\rho}_1 - \rho_1 | \mathcal{M}^{ct}] \overset{1}{=} \frac{-2\rho_1}{T} - 2\frac{(1+\rho_1)}{T}. \tag{5}$$

By comparing formula (5) to (4), we see that the estimation bias due to deterministic terms has doubled by adding a trend to the estimation model. In Section 3 of this paper, it is shown that this result carries over to higher-order AR models. The section after this introduction contains some notation and the basic Nagar-type bias equation. Section 4 concludes, while all the proofs are provided in Appendix A.

2 Notation and Preliminary Analysis

The models in (1) can be written in matrix notation as

$$\mathbf{y} = Z\beta + \varepsilon, \tag{6}$$

where $\mathbf{y} = (y_1, \ldots, y_T)$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)$ and the content of the matrix Z and vector β depends on the estimation model being used. The bias of the OLS estimator $\hat{\beta} = (Z'Z)^{-1}Z'\mathbf{y}$ is given by

$$\mathbb{E}[(\hat{\beta} - \beta)] = \mathbb{E}[(Z'Z)^{-1}Z'\varepsilon]. \tag{7}$$

Since $\hat{\beta}$ is invariant to σ_ε^2, the variance is normalized to 1 without loss of generality. Furthermore, we shall assume that $\mu = 0$ (and $\gamma = 0$) in the analysis, since the elements of $\hat{\beta}$ that refer to $\hat{\rho} = (\hat{\rho}_1, ..., \hat{\rho}_k)'$ are invariant with respect to μ (and γ) in model $\mathcal{M}^c$ ($\mathcal{M}^{ct}$).

To distinguish the various models, Z will be indexed as Z^m for $m \in \{c, ct\}$. Let Z^c denote the regressor matrix in model $\mathcal{M}^c$, so that $\hat{\beta}^c = (Z^{c\prime}Z^c)^{-1}Z^{c\prime}\mathbf{y}$, while Z^{ct} contains the regressors for model $\mathcal{M}^{ct}$, so that $\hat{\beta}^{ct} = (Z^{ct\prime}Z^{ct})^{-1}Z^{ct\prime}\mathbf{y}$.

In order to distinguish the deterministic and stochastic part of the matrix of regressors Z^m, decompose $Z^m = \bar{Z}^m + \tilde{Z}^m$, where $\bar{Z}^m$ is defined as the mathematical expectation of Z^m, so that

$$
\begin{aligned}
Z^m &= \mathbb{E}[Z^m] + (Z^m - \mathbb{E}[Z^m]) \\
&= \bar{Z}^m + \tilde{Z}^m.
\end{aligned}
\tag{8}
$$

For model $\mathcal{M}^c$, the non-stochastic matrix $\bar{Z}^m$ and stochastic matrix $\tilde{Z}^m$ are given by

$$\bar{Z}^c = (0_{T \times k} : \iota) \qquad \text{and} \qquad \tilde{Z}^c = (\mathbf{y}_{-1} : ... : \mathbf{y}_{-k} : 0_T), \tag{9}$$

where $\mathbf{y}_{-i} = (y_{1-i}, \ldots, y_{T-i})'$ for $1 \leq i \leq k$ and $\iota = (1, \ldots, 1)'$. For model $\mathcal{M}^{ct}$, we have

$$\bar{Z}^{ct} = (0_{T \times k} : \iota : \tau) \qquad \text{and} \qquad \tilde{Z}^{ct} = (\mathbf{y}_{-1} : ... : \mathbf{y}_{-k} : 0_T : 0_T), \tag{10}$$

where $\tau = (1/T, 2/T, ..., 1)'$. Note that the linear trend is divided by T, so that all the regressors are of the same 'magnitude'. This can be done without loss of generality since the elements of $\hat{\beta}$ that refer to $\hat{\rho}$ are invariant to this transformation. The inverse of $\mathbb{E}[Z'Z] = \bar{Z}'\bar{Z} + \mathbb{E}[\tilde{Z}'\tilde{Z}]$ is denoted by Q, *i.e.* $Q = (\bar{Z}'\bar{Z} + \mathbb{E}[\tilde{Z}'\tilde{Z}])^{-1}$. Due to the scaling of the linear trend, we have $Q = O(T^{-1})$ in both models; see formulas (A.1) and (A.6) in the Appendix. The Nagar-type expansion, named after Nagar (1959), that is utilized in this paper follows from the identity

$$(Z'Z)^{-1} = Q \left[I + (\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q + (\tilde{Z}'\tilde{Z} - \mathbb{E}[\tilde{Z}'\tilde{Z}])Q \right]^{-1}, \tag{11}$$

3

where the stochastic terms $(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q$ and $(\tilde{Z}'\tilde{Z} - \mathbb{E}[\tilde{Z}'\tilde{Z}])Q$ both are $O_p(T^{-1/2})$. The inverse of the form $(I + A)^{-1}$ with $A = O_p(T^{-1/2})$ can be approximated by

$$(I + A)^{-1} = I - A + A^2 - A^3..., \qquad (12)$$

see for instance Kiviet and Phillips (1993, p. 77). The bias in (7) together with formula (11) and the approximation $(I + A)^{-1} \approx (I - A)$ gives

$$\mathbb{E}[(\hat{\beta} - \beta)] = \mathbb{E}[Q\tilde{Z}'\varepsilon - Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q\tilde{Z}'\varepsilon - Q(\tilde{Z}'\tilde{Z} - \mathbb{E}[\tilde{Z}'\tilde{Z}])Q\tilde{Z}'\varepsilon] + o(T^{-1}). \qquad (13)$$

Of course, $(\bar{Z}, \tilde{Z}, Q)$ should be replaced by $(\bar{Z}^c, \tilde{Z}^c, Q^c)$ for model $\mathcal{M}^c$ and by $(\bar{Z}^{ct}, \tilde{Z}^{ct}, Q^{ct})$ for model $\mathcal{M}^{ct}$ respectively. Note that $\tilde{Z}'\varepsilon$ has zeros at the elements referring to $\hat{\mu}$ (and $\hat{\beta}$), while $\bar{Z}'\varepsilon$ has zeros at all the elements referring to $\hat{\rho}$. Hence, the first-order bias can be decomposed into two parts: (i) the bias due to the lagged-dependent regressors (indicated by $\mathbb{E}_r$), which is given by

$$\mathbb{E}_r[(\hat{\beta} - \beta)] \overset{1}{=} \mathbb{E}[Q\tilde{Z}'\varepsilon - Q(\tilde{Z}'\tilde{Z} - \mathbb{E}[\tilde{Z}'\tilde{Z}])Q\tilde{Z}'\varepsilon], \qquad (14)$$

and (ii) the bias due to inclusion of deterministic term (indicated by $\mathbb{E}_d$), which is equal to

$$\mathbb{E}_d[(\hat{\beta} - \beta)] \overset{1}{=} -\mathbb{E}[Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q\bar{Z}'\varepsilon]. \qquad (15)$$

Such a decomposition of the bias seems to hold more generally, see e.g. Cordeiro and Klein (1994).

3 Bias Approximations

In order to derive an explicit expression for the bias term, let Ω denote the covariance matrix of $(y_{t-1}, ..., y_{t-k})$ and Ω^{-1} its inverse. The (i, j)-th element of Ω is denoted by Ω_{ij}, while Ω^{ij} denotes the (i, j)-th element of Ω^{-1}. Theorem 1 shows the first-order approximation of the estimation bias when the model contains an intercept.

Theorem 1 *The bias of $\hat{\rho}_i^c$ for $1 \leq i \leq k$ due to the inclusion of an intercept in the stable AR(k) model as shown in (1b) is given by*

$$\mathbb{E}_d[(\hat{\rho}_i^c - \rho_i)|\mathcal{M}^c] \overset{1}{=} -\frac{\Omega^{i1} + ... + \Omega^{ik}}{1 - \rho_1 - ... - \rho_k}T^{-1}. \qquad (16)$$

For the AR(1) model, we have $\Omega^{11} = 1 - \rho_1^2$, so that $\mathbb{E}_d[(\hat{\rho}_1^c - \rho_1)|k = 1] \overset{1}{=} -(1 + \rho_1)T^{-1}$, which is in line with formula (4). Furthermore, we have the following approximations

$$\mathbb{E}_d[(\hat{\rho}^c - \rho)|k = 2] \overset{1}{=} -(1 + \rho_2, 1 + \rho_2)'T^{-1}. \qquad (17a)$$

$$\mathbb{E}_d[(\hat{\rho}^c - \rho)|k = 3] \overset{1}{=} -(1 + \rho_3, 1 - \rho_1 + \rho_2 + \rho_3, 1 + \rho_3)'T^{-1}. \qquad (17b)$$

$$\mathbb{E}_d[(\hat{\rho}^c - \rho)|k = 4] \overset{1}{=} -(1 + \rho_4, 1 - \rho_1 + \rho_3 + \rho_4, 1 - \rho_1 + \rho_3 + \rho_4, 1 + \rho_4)'T^{-1}. \quad (17c)$$

These results are in line with Table 1 of Shaman and Stine (1988, p. 846).

The next theorem shows that the bias due to the deterministic terms in the AR(k) model with an intercept and trend is twice the magnitude of the bias in the model with only an intercept.

Theorem 2 *In the stable AR(k) model with intercept and trend as shown in (1c), we have for* $1 \leq i \leq k$

$$\frac{\mathbb{E}_d[(\hat{\rho}_i^{ct} - \rho_i)|\mathcal{M}^{ct}]}{\mathbb{E}_d[(\hat{\rho}_i^{c} - \rho_i)|\mathcal{M}^{c}]} \stackrel{1}{=} 2. \qquad (18)$$

To assess the quality of the asymptotic result shown in (18), a small simulation study has been carried out. All the simulations were done on a PC using Matlab. Observations were generated according to an AR(1) process. All results are based on $R = 50,000$ replications and three sample sizes were considered: $T \in \{25, 50, 100\}$. The AR(1) parameter was taken as $\rho_1 \in \{-0.99, 0.98, ..., 0.99\}$. Since the bias is invariant with respect to σ_ε, μ in models $\mathcal{M}^c$ and $\mathcal{M}^{ct}$ and γ in model $\mathcal{M}^{ct}$, we take $\sigma_\varepsilon = 1$ and $\mu = \gamma = 0$ without loss of generality. The starting value was drawn from the stationary distribution, i.e. $y_0 \sim \mathcal{N}(0, 1/(1 - \rho_1^2))$. The bias due to inclusion of deterministic terms was approximated by

$$\frac{\sum_{r=1}^{R}(\hat{\rho}_{1,r}^{ct} - \hat{\rho}_{1,r}^{p})}{\sum_{r=1}^{R}(\hat{\rho}_{1,r}^{c} - \hat{\rho}_{1,r}^{p})}, \qquad (19)$$

where the subindex r denotes the r-th replication in the simulation. The estimated relative additional bias due to the trend given in (19) is shown in Figure 1. From this figure, we conclude that the approximation is reasonably accurate even for values of ρ_1 close to 1. This is probably due to the fact that we consider a ratio, since it is well known that the absolute bias shown in (4) and (5) can deviate substantially from the actual bias as ρ_1 approaches the value of 1. So, although the derivation is based on a stable model, the approximation seems to be useful in the unstable case as well. To illustrate this point, suppose that the shape of the distribution remains constant when adding deterministic terms. Then the shift in the quantiles from $\mathcal{M}^p$ to $\mathcal{M}^{ct}$ should be twice as large as the shift from $\mathcal{M}^p$ to $\mathcal{M}^c$. In case of a unit root and $T = 100$, the 5% quantiles of $T(\hat{\rho}_1 - 1)$ are approximately -7.9 ($\mathcal{M}^p$), -13.7 ($\mathcal{M}^c$) and -20.7 ($\mathcal{M}^{ct}$), see Fuller (1976, p. 371), so that

$$\frac{-20.7 - (-7.9)}{-13.7 - (-7.9)} = 2.21,$$

which differs from 2 by 10% only .

Insert Figure 1 about here.

5

4 Conclusion

In this paper, we have shown that the first-order bias for the AR coefficients due to deterministic terms doubles when a linear trend is added to a stable AR model with intercept. The simulation results show that the asymptotic approximation is relevant for sample sizes encountered in practice, at least for the AR(1) model. Although outside the scope of this paper, it can be shown (using the same techniques) that the bias due to deterministic terms triples when a quadratic trend is added to an AR model with intercept and linear trend.

A Proofs

Proof of Theorem 1. For model $\mathcal{M}^c$, we obtain (due to the block structure)

$$
\begin{aligned}
Q^c &= \left(\bar{Z}^{c\prime}\bar{Z}^c + \mathbb{E}[\tilde{Z}^{c\prime}\tilde{Z}^c]\right)^{-1} \\[2mm]
&= \begin{pmatrix} \Omega_{11} & \cdots & \Omega_{1k} & 0 \\ \vdots & & \vdots & \vdots \\ \Omega_{k1} & \cdots & \Omega_{kk} & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}^{-1} T^{-1} = \begin{pmatrix} \Omega^{11} & \cdots & \Omega^{1k} & 0 \\ \vdots & & \vdots & \vdots \\ \Omega^{k1} & \cdots & \Omega^{kk} & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} T^{-1}.
\end{aligned}
\tag{A.1}
$$

since $\mathbb{E}[\mathbf{y}'_{-i}\mathbf{y}_{-j}] = T\Omega_{ij}$ for $i, j \in \{1, \ldots, k\}$. Furthermore,

$$
(\bar{Z}^{c\prime}\tilde{Z}^c + \tilde{Z}^{c\prime}\bar{Z}^c) = \begin{pmatrix} 0 & \cdots & 0 & S_{y_{-1}} \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & S_{y_{-k}} \\ S_{y_{-1}} & \cdots & S_{y_{-k}} & 0 \end{pmatrix} \quad \text{and} \quad Q^c \bar{Z}^{c\prime}\varepsilon = \begin{pmatrix} 0_k \\ T^{-1}S_\varepsilon \end{pmatrix}, \tag{A.2}
$$

where $S_{y_{-i}} \equiv \sum y_{t-i}$ and $S_\varepsilon \equiv \sum \varepsilon_t$. Combining (A.1) and (A.2) leads to

$$
\mathbb{E}[Q^c(\bar{Z}^{c\prime}\tilde{Z}^c + \tilde{Z}^{c\prime}\bar{Z}^c)Q^c\bar{Z}^{c\prime}\varepsilon] \stackrel{1}{=} \mathbb{E}\left[\begin{pmatrix} S_\varepsilon(\Omega^{11}S_{y_{-1}} + \ldots + \Omega^{1k}S_{y_{-k}}) \\ \vdots \\ S_\varepsilon(\Omega^{k1}S_{y_{-1}} + \ldots + \Omega^{kk}S_{y_{-k}}) \\ 0 \end{pmatrix} T^{-2}\right], \tag{A.3}
$$

so that the bias due to the inclusion of the intercept for $\hat{\rho}_i$ is given by

$$
\mathbb{E}_d[(\hat{\rho}_i^c - \rho_i)|\mathcal{M}^c] \stackrel{1}{=} \mathbb{E}[-S_\varepsilon(\Omega^{i1}S_{y_{-1}} + \ldots + \Omega^{ik}S_{y_{-k}})T^{-2}]. \tag{A.4}
$$

Next, we make use of the fact that (for $1 \leq i \leq k$)

$$
\mathbb{E}[T^{-1}S_\varepsilon S_{y_{-i}}] = \mathbb{E}[(T^{-1/2}\sum \varepsilon_t)(T^{-1/2}\sum y_{t-i})] \stackrel{1}{=} \frac{1}{1 - \rho_1 - \ldots - \rho_k}, \tag{A.5}
$$

6

see formulas (A.11) and (A.12) for $m = n = 0$. Substitution of (A.5) into (A.4) give the desired result. $\qquad\square$

Proof of Theorem 2. For model $\mathcal{M}^{ct}$, the matrix Q^{ct} becomes

$$Q^{ct} = \left(\bar{Z}^{ct\prime}\bar{Z}^{ct} + \mathbb{E}[\tilde{Z}^{ct\prime}\tilde{Z}^{ct}]\right)^{-1}$$

$$\stackrel{1}{=} \begin{pmatrix} \Omega_{11} & \cdots & \Omega_{1k} & 0 & 0 \\ \vdots & & \vdots & 0 & 0 \\ \Omega_{k1} & \cdots & \Omega_{kk} & 0 & 0 \\ 0 & 0 & 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{3} \end{pmatrix}^{-1} T^{-1} = \begin{pmatrix} \Omega^{11} & \cdots & \Omega^{1k} & 0 & 0 \\ \vdots & & \vdots & 0 & 0 \\ \Omega^{1k} & \cdots & \Omega^{kk} & 0 & 0 \\ 0 & 0 & 0 & 4 & -6 \\ 0 & 0 & 0 & -6 & 12 \end{pmatrix} T^{-1}. \quad \text{(A.6)}$$

The analogue of (A.2) becomes

$$(\bar{Z}^{ct\prime}\tilde{Z}^{ct} + \tilde{Z}^{ct\prime}\bar{Z}^{ct}) = \begin{pmatrix} 0 & \cdots & 0 & S_{y_{-1}} & S_{\tau y_{-1}} \\ \vdots & & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & S_{y_{-k}} & S_{\tau y_{-k}} \\ S_{y_{-1}} & \cdots & S_{y_{-k}} & 0 & 0 \\ S_{\tau y_{-1}} & \cdots & S_{\tau y_{-k}} & 0 & 0 \end{pmatrix} \quad \text{(A.7)}$$

and

$$Q^{ct}\bar{Z}^{ct\prime}\varepsilon = \begin{pmatrix} 0_k \\ \frac{4}{T}S_\varepsilon - \frac{6}{T^2}S_{\tau\varepsilon} \\ -\frac{6}{T^2}S_\varepsilon + \frac{12}{T^3}S_{\tau\varepsilon} \end{pmatrix}, \quad \text{(A.8)}$$

where $S_{\tau y_{-i}} \equiv \sum(t/T)y_{t-i}$ and $S_{\tau\varepsilon} \equiv \sum(t/T)\varepsilon_t$. After carrying out the multiplication, the bias of $\hat{\rho}_i^{ct}$ due to the intercept and trend is equal to

$$\mathbb{E}_d[(\hat{\rho}_i^{ct} - \rho_i)|\mathcal{M}^{ct}] \stackrel{1}{=} -\mathbb{E}[(\tfrac{4}{T}S_\varepsilon - \tfrac{6}{T^2}S_{\tau\varepsilon})(\Omega^{i1}S_{y_{-1}} + ... + \Omega^{ik}S_{y_{-k}})]T^{-1} +$$
$$-\mathbb{E}[(-\tfrac{6}{T^2}S_\varepsilon + \tfrac{12}{T^3}S_{\tau\varepsilon})(\Omega^{i1}S_{\tau y_{-1}} + ... + \Omega^{ik}S_{\tau y_{-k}})]T^{-1}. \quad \text{(A.9)}$$

For approximating the bias, the following relationships are useful

$$\mathbb{E}[(T^{-3/2}S_{\tau\varepsilon})(T^{-1/2}S_{y_{-i}})] \stackrel{1}{=} \tfrac{1}{2}\mathbb{E}[(T^{-1/2}S_\varepsilon)(T^{-1/2}S_{y_{-i}})], \quad \text{(A.10a)}$$

$$\mathbb{E}[(T^{-1/2}S_\varepsilon)(T^{-3/2}S_{\tau y_{-i}})] \stackrel{1}{=} \tfrac{1}{2}\mathbb{E}[(T^{-1/2}S_\varepsilon)(T^{-1/2}S_{y_{-i}})], \quad \text{(A.10b)}$$

$$\mathbb{E}[(T^{-3/2}S_{\tau\varepsilon})(T^{-3/2}S_{\tau y_{-i}})] \stackrel{1}{=} \tfrac{1}{3}\mathbb{E}[(T^{-1/2}S_\varepsilon)(T^{-1/2}S_{y_{-i}})]. \quad \text{(A.10c)}$$

These relationships follow from the observation that (for $m, n \in \mathbb{N}_0$)

$$\left(T^{-\frac{1}{2}}\sum(t/T)^m\varepsilon_t, T^{-\frac{1}{2}}\sum(t/T)^n y_{t-i}\right) \xrightarrow{d} \left(\int_0^1 r^m dW(r), \sigma_\infty \int_0^1 r^n dW(r)\right), \quad \text{(A.11)}$$

where $\sigma_\infty = \phi^{-1}(1)$ is the 'long run' standard deviation and W denotes a standard Wiener process (note that σ_ε was normalized to 1). Since r^m and r^n are of finite variation on the unit interval, we obtain

$$\mathbb{E}\left[\sigma_\infty \int_0^1 r^m \mathrm{d}W(r) \int_0^1 r^n \mathrm{d}W(r)\right] = \sigma_\infty \int_0^1 r^{m+n+1}\mathrm{d}r = \frac{1}{(m+n+1)\phi(1)}. \qquad (A.12)$$

Using the relationships shown in (A.10), we finally get for $i \in \{1, \ldots, k\}$

$$\mathbb{E}_d[(\hat{\rho}_i^{ct} - \rho_i)|\mathcal{M}^{ct}] \stackrel{1}{=} 2\,\mathbb{E}\left[-S_\varepsilon\left(\Omega^{i1}S_{y_{-1}} + \ldots + \Omega^{ik}S_{y_{-k}}\right)T^{-2}\right]$$

$$\stackrel{1}{=} 2\,\mathbb{E}_d[(\hat{\rho}_i^c - \rho_i)|\mathcal{M}^c], \qquad (A.13)$$

see formula (A.4) for the last equality. So, $\mathbb{E}_d[(\hat{\rho}^{ct} - \rho)]$ is approximately twice the magnitude of $\mathbb{E}_d[(\hat{\rho}^c - \rho)]$. $\qquad \square$

References

Barlett, M. S. (1946). On the Theoretical Specification and Sampling Properties of Autocorrelated Time-Series. *Journal of the Royal Statistical Society*, **8**, 27–41.

Bhansali, R. J. (1981). Effects of Not Knowing the Order of an Autoregressive Process on the Mean Squared Error of Prediction–I. *Journal of the American Statistical Association*, **76**, 588–597.

Cordeiro, G. and R. Klein (1994). Bias correction in ARMA models. *Statistics and Probability Letters*, **19**, 169–176.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*. New York: John Wiley.

Kang, W., D. W. Shin, and Y. Lee (2003). Biases of the Restricted Maximum Likelihood Estimators for ARMA Processes with Polynomial Time Trend. *Journal of Statistical Planning and Inference*, **116**, 163–176.

Kendall, M. G. (1954). Note on the Bias in the Estimation of Auto Correlation. *Biometrika*, **61**, 403–404.

Kiviet, J. F. and G. D. A. Phillips (1993). Alternative Bias Approximations in Regressions with a Lagged Dependent Variable. *Econometric Theory*, **9**, 62–80.

Marriot, F. H. C. and J. A. Pope (1954). Bias in the Estimation of Autocorrelations. *Biometrika*, **61**, 393–403.

Nagar, A. L. (1959). The Bias and Moment Matrix of the General k-class Estimators of the Parameters in Simultaneous Equations. *Econometrica*, **27**, 575–595.

Shaman, P. and R. A. Stine (1988). The Bias of Autoregressive Coefficient Estimators. *Journal of the American Statistical Association*, **83**, 842–848.

Figure 1: The bias due to the deterministic terms in the AR(1) model with trend and intercept relative to bias in the AR(1) model with intercept; see equation (19).