THE RIMINI CENTRE FOR ECONOMIC ANALYSIS

WP 24_12

Karim M. Abadir
Imperial College London, UK
The Rimini Centre for Economic Analysis (RCEA), Italy

Rolf Larsson
Uppsala University, Sweden

BIASES OF CORRELOGRAMS AND OF AR REPRESENTATIONS OF STATIONARY SERIES

Biases of correlograms and of AR representations of stationary series

Karim M. Abadir and Rolf Larsson*

Imperial College London and Uppsala University

We derive the relation between the biases of correlograms and of estimates of auto-regressive AR(k) representations of stationary series, and we illustrate it with a simple AR example. The new relation allows for k to vary with the sample size, which is a representation that can be used for most stationary processes. As a result, the biases of the estimators of such processes can now be quantified explicitly and in a unified way.

Short title: Biases of correlograms and of AR representations.

Key words and phrases. Auto-correlation function (ACF) and correlogram, auto-regressive (AR) representation, least-squares bias.

*ESRC grants R000239538, RES000230176, and RES062230790 are gratefully acknowledged.

1 Introduction

Let $\{y_t\}_1^T$ denote a time series. Consider fitting to this data an AR(k) model

$$y_t = \mu + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + ... + \alpha_k y_{t-k} + \varepsilon_t, \tag{1}$$

where the sequence $\{\varepsilon_t\}$ is i.i.d. with mean 0 and variance σ^2. It is not assumed that this AR model generates the data or that k is fixed as $T \to \infty$. However, we will assume that the series is weakly stationary in the sense of having second-order moments that are bounded as T increases, hence ruling out local-to-unity AR representations for example. Because of the Wold decomposition theorem, the data-generating process (DGP) of such a series can be written as an MA(∞) with orthogonal errors, which (1) would be approximating by the invertibility of the MA representation. A few examples of such DGPs include stationary cases of ARMA, fractional I(d), the Gegenbauer ARMA (or GARMA) processes of Gray, Zhang, and Woodward (1989) and their extension by Giraitis and Leipus (1995), processes with spectral singularities away from the origin as in Giraitis, Hidalgo, and Robinson (2001) and Hidalgo (2005), and the cyclical long-memory process CM(ω, d) of Abadir and Talmain (2011).

Let the auto-correlation function (ACF) of $\{y_t\}$ be denoted by ρ_j (where $\rho_0 \equiv 1$). Writing $\boldsymbol{\rho} := (\rho_1, ..., \rho_k)'$ and $\boldsymbol{\alpha} := (\alpha_1, ..., \alpha_k)'$, we have the Yule-Walker equation

$$\mathbf{R}\boldsymbol{\alpha} = \boldsymbol{\rho} \tag{2}$$

where $\mathbf{R}$ is the Toeplitz matrix

$$\mathbf{R} := \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \rho_1 \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & 1 \end{pmatrix}$$

which is assumed to be positive definite (hence invertible). The auto-covariance function can be estimated by $\hat{\gamma}_j := T^{-1} \sum_{t=1}^{T-j} y_t y_{t+j} - T^{-2} \sum_{t=1}^{T-j} y_t \sum_{s=1}^{T-j} y_{s+j}$ and the corresponding correlogram is $\hat{\rho}_j := \hat{\gamma}_j / \hat{\gamma}_0$, yielding $\hat{\boldsymbol{\rho}}$ and $\hat{\mathbf{R}}$ as estimators of $\boldsymbol{\rho}$

and $\mathbf{R}$. The sample auto-covariance matrix is $\hat{\mathbf{\Gamma}} := \hat{\gamma}_0 \hat{\mathbf{R}}$, which is invertible with probability 1. Denote the least squares estimator (LSE) of $\boldsymbol{\alpha}$ by $\hat{\boldsymbol{\alpha}}$ which, apart from the initial conditions, satisfies the same relation as in (2), namely $\hat{\mathbf{R}}\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\rho}}$. The vector $\hat{\boldsymbol{\alpha}}$ can be interpreted as an estimator of the partial auto-correlation function as k increases. Like its counterpart $\mathbf{R}\boldsymbol{\alpha} = \boldsymbol{\rho}$, the relation between $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\rho}}$ is also linear ($\hat{\mathbf{R}}$ is linear in $\hat{\boldsymbol{\rho}}$) and invertible with probability 1, and so is the relation between $(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho})$ and $(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ for any given $\boldsymbol{\rho}$ implied by the DGP. This also holds asymptotically, by the assumption of asymptotic stationarity.

We are going to use these relations to link the biases of the ACF and AR representations. The simple linear structure of the link has surprisingly not been exploited before to derive biases of estimators. We do so in Section 2, then illustrate our approach in Section 3 with a simple AR example where the biases of the two representations are known. When k is not fixed as T varies, the usual expansions in the literature are not valid anymore, and solving this problem is an additional contribution of our formula. Subject to the conditions leading to the consistency of the estimator $\hat{\boldsymbol{\alpha}}$ of the AR representation of a stationary process, our relation provides a new explicit way to quantify the biases of the estimators for this process. Throughout, we use the notation conventions proposed in Abadir and Magnus (2002).

2 Link between correlogram and AR biases

The relation between the bias of $\hat{\boldsymbol{\alpha}}$ and that of $\hat{\boldsymbol{\rho}}$ can be obtained as follows. We have

$$\hat{\boldsymbol{\rho}} - \boldsymbol{\rho} = \hat{\mathbf{R}}\hat{\boldsymbol{\alpha}} - \mathbf{R}\boldsymbol{\alpha} = \left(\hat{\mathbf{R}}\hat{\boldsymbol{\alpha}} - \mathbf{R}\hat{\boldsymbol{\alpha}}\right) + \left(\mathbf{R}\hat{\boldsymbol{\alpha}} - \mathbf{R}\boldsymbol{\alpha}\right) = \left(\hat{\mathbf{R}} - \mathbf{R}\right)\hat{\boldsymbol{\alpha}} + \mathbf{R}\left(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\right),$$

and we now exploit the fact that $(\hat{\mathbf{R}} - \mathbf{R})$ is a linear function of $(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho})$. Let

$$\mathbf{R} := \mathbf{I}_k + \mathbf{R}_{\mathrm{L}} + \mathbf{R}_{\mathrm{L}}'$$

where $\mathbf{R}_{\mathrm{L}}$ is a strictly lower-triangular matrix of correlations having typical j-th column given by $\mathbf{A}^j \boldsymbol{\rho}$, with

$$\mathbf{A} := \begin{pmatrix} \mathbf{0}' & 0 \\ \mathbf{I}_{k-1} & \mathbf{0} \end{pmatrix}.$$

Let $\mathbf{K}_k$ denote the $k^2 \times k^2$ commutation matrix, giving $\mathbf{K}_k \operatorname{vec} \mathbf{R}_{\mathrm{L}}' = \operatorname{vec} \mathbf{R}_{\mathrm{L}}$; see Abadir and Magnus (2005, Chapter 11.1) for the explicit formula of $\mathbf{K}_k$. Let $\mathbf{N}_k := \frac{1}{2}(\mathbf{I}_{k^2} + \mathbf{K}_k)$ be the corresponding symmetrizer matrix. Then,

$$\begin{aligned}
\operatorname{vec}\left(\hat{\mathbf{R}} - \mathbf{R}\right) &= \operatorname{vec}\left(\hat{\mathbf{R}}_{\mathrm{L}} - \mathbf{R}_{\mathrm{L}}\right) + \operatorname{vec}\left(\hat{\mathbf{R}}_{\mathrm{L}}' - \mathbf{R}_{\mathrm{L}}'\right) \\
&= 2\mathbf{N}_k \operatorname{vec}\left(\hat{\mathbf{R}}_{\mathrm{L}} - \mathbf{R}_{\mathrm{L}}\right) = 2\mathbf{N}_k \mathbf{C}\left(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\right)
\end{aligned}$$

by $\operatorname{vec}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}) = \hat{\boldsymbol{\rho}} - \boldsymbol{\rho}$, and

$$\mathbf{C} := \begin{pmatrix} \mathbf{A} \\ \mathbf{A}^2 \\ \vdots \\ \mathbf{A}^{k-1} \\ \mathbf{O} \end{pmatrix} = \sum_{j=1}^{k} \left(\mathbf{e}_j \otimes \mathbf{A}^j\right),$$

where $\mathbf{e}_j$ is the j-th elementary vector (j-th column of $\mathbf{I}_k$). Therefore,

$$\begin{aligned}
\hat{\boldsymbol{\rho}} - \boldsymbol{\rho} &= \operatorname{vec}\left(\mathbf{I}_k \left(\hat{\mathbf{R}} - \mathbf{R}\right) \hat{\boldsymbol{\alpha}}\right) + \mathbf{R}\left(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\right) \\
&= \left(\hat{\boldsymbol{\alpha}}' \otimes \mathbf{I}_k\right) \operatorname{vec}\left(\hat{\mathbf{R}} - \mathbf{R}\right) + \mathbf{R}\left(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\right) \\
&= 2\left(\hat{\boldsymbol{\alpha}}' \otimes \mathbf{I}_k\right) \mathbf{N}_k \mathbf{C}\left(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\right) + \mathbf{R}\left(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\right).
\end{aligned}$$

Define

$$\hat{\mathbf{D}} := \mathbf{I}_k - 2\left(\hat{\boldsymbol{\alpha}}' \otimes \mathbf{I}_k\right) \mathbf{N}_k \mathbf{C} = \mathbf{I}_k - \left[\left(\hat{\boldsymbol{\alpha}}' \otimes \mathbf{I}_k\right) + \left(\mathbf{I}_k \otimes \hat{\boldsymbol{\alpha}}'\right)\right] \mathbf{C},$$

such that $\hat{\mathbf{D}}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}) = \mathbf{R}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ and the invertibility of $\hat{\mathbf{D}}$ (for finite T and asymptotically) is implied by the linear invertibility of the relation of $(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho})$ to $(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ for any given $\boldsymbol{\rho}$.

We will now assume that the LSE $\hat{\boldsymbol{\alpha}}$ is consistent in the sense of converging to the vector $\boldsymbol{\alpha}$ implied by the DGP, and that its third-order moments converge uniformly. Then, writing

$$\hat{\boldsymbol{\beta}} := T^{\delta}\left(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\right) \tag{3}$$

where $\delta > 0$ such that $\hat{\boldsymbol{\beta}}$ is nondegenerate, we have $\hat{\mathbf{D}} = \mathbf{D}_0 - T^{-\delta}\hat{\mathbf{D}}_1$ with

$$\begin{aligned}
&\mathbf{D}_0 := \mathbf{I}_k - 2\left(\boldsymbol{\alpha}' \otimes \mathbf{I}_k\right) \mathbf{N}_k \mathbf{C} = \mathbf{I}_k - 2\left(\boldsymbol{\rho}'\mathbf{R}^{-1} \otimes \mathbf{I}_k\right) \mathbf{N}_k \mathbf{C}, \\
&\hat{\mathbf{D}}_1 := 2\left(\hat{\boldsymbol{\beta}}' \otimes \mathbf{I}_k\right) \mathbf{N}_k \mathbf{C},
\end{aligned}$$

such that $\mathbf{D}_0$ is the plim of $\hat{\mathbf{D}}$. A three-term Taylor theorem expansion of $\hat{\mathbf{D}}^{-1}$ gives

$$\begin{aligned}
\hat{\boldsymbol{\rho}} - \boldsymbol{\rho} &= T^{-\delta}\mathbf{D}_0^{-1}\mathbf{R}\hat{\boldsymbol{\beta}} + T^{-2\delta}\mathbf{D}_0^{-1}\hat{\mathbf{D}}_1\mathbf{D}_0^{-1}\mathbf{R}\hat{\boldsymbol{\beta}} + O_p(T^{-3\delta}) \\
&= T^{-\delta}\operatorname{vec}\left(\mathbf{D}_0^{-1}\mathbf{R}\hat{\boldsymbol{\beta}}\right) + T^{-2\delta}\mathbf{D}_0^{-1}\operatorname{vec}\left(\hat{\mathbf{D}}_1\mathbf{D}_0^{-1}\mathbf{R}\hat{\boldsymbol{\beta}}\right) + O_p(T^{-3\delta}) \\
&= T^{-\delta}\left(\hat{\boldsymbol{\beta}}' \otimes \mathbf{I}_k\right)\operatorname{vec}\left(\mathbf{D}_0^{-1}\mathbf{R}\right) + T^{-2\delta}\mathbf{D}_0^{-1}\left(\hat{\boldsymbol{\beta}}' \otimes \hat{\mathbf{D}}_1\right)\operatorname{vec}\left(\mathbf{D}_0^{-1}\mathbf{R}\right) + O_p(T^{-3\delta}) \\
&= T^{-\delta}\left[\left(\hat{\boldsymbol{\beta}}' \otimes \mathbf{I}_k\right) + 2T^{-\delta}\mathbf{D}_0^{-1}\left(\hat{\boldsymbol{\beta}}' \otimes \left(\hat{\boldsymbol{\beta}}' \otimes \mathbf{I}_k\right)\mathbf{N}_k\mathbf{C}\right)\right]\operatorname{vec}\left(\mathbf{D}_0^{-1}\mathbf{R}\right) + O_p(T^{-3\delta})
\end{aligned}$$

and

$$\begin{aligned}
&(4)\ \operatorname{E}\left(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\right) \\
&= T^{-\delta}\left[\operatorname{E}\left(\hat{\boldsymbol{\beta}}' \otimes \mathbf{I}_k\right) + 2T^{-\delta}\mathbf{D}_0^{-1}\operatorname{E}\left(\hat{\boldsymbol{\beta}}' \otimes \left(\hat{\boldsymbol{\beta}}' \otimes \mathbf{I}_k\right)\mathbf{N}_k\mathbf{C}\right)\right]\operatorname{vec}\left(\mathbf{D}_0^{-1}\mathbf{R}\right) + O(T^{-3\delta}) \\
&\sim T^{-\delta}\left[\operatorname{E}\left(\hat{\boldsymbol{\beta}}' \otimes \mathbf{I}_k\right) + 2T^{-\delta}\mathbf{D}_0^{-1}\operatorname{E}\left(\hat{\boldsymbol{\beta}}' \otimes \left(\hat{\boldsymbol{\beta}}' \otimes \mathbf{I}_k\right)\mathbf{N}_k\mathbf{C}\right)\right]\operatorname{vec}\left(\mathbf{D}_0^{-1}\mathbf{R}\right),
\end{aligned}$$

where the uniform convergence of the third-order moments ensures that the expectation of the $O_p(T^{-3\delta})$ remainder term is finite. The leading term is made up of two components because the variate $\hat{\boldsymbol{\beta}}$ is nondegenerate in the limit: its variance does not tend to zero (hence the second term is of maximal order $T^{-2\delta}$) but its mean may tend to zero (hence the first term can be of order smaller than $T^{-\delta}$). For example, if the data are generated by a stationary AR(k) where k is fixed, we obtain $\sqrt{T}$-consistency of $\hat{\boldsymbol{\alpha}}$ (i.e. $\delta = 1/2$) but $\operatorname{E}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = O(T^{-1})$ because the centering of $\hat{\boldsymbol{\beta}}$ converges to $\mathbf{0}$ at a rate of $O(T^{-1/2})$; e.g. see Abadir (1993).

It is typically hard to derive biases that vanish asymptotically, but much easier to calculate limiting variances. Therefore, by means of (4), we can now freely transform the leading term of the biases of $\hat{\boldsymbol{\rho}}$ into those of $\hat{\boldsymbol{\alpha}}$ (i.e. $\hat{\boldsymbol{\beta}}$), and vice-versa if needed. Notice that none of the results derived so far requires normality of $\{\varepsilon_t\}$.

3 Illustration with an AR process

Kendall (1954) gives explicitly the leading term of $\operatorname{E}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho})$ for general Gaussian stationary series. After correcting some typos and using $1 - j/T \sim 1$, this is

$$(5) \qquad \operatorname{E}\left(\hat{\rho}_j - \rho_j\right) \sim \frac{a_j}{a_0} - \rho_j - \frac{2}{Ta_0^2}\sum_{i=-\infty}^{\infty}\rho_i\left(\rho_{i+j} - \frac{a_j}{a_0}\rho_i\right)$$

where

$$(6)\qquad a_j := \rho_j - \frac{1}{T^2}\left(\sum_{i=1}^{T-1}(T-i)\,\rho_{i+j} + \sum_{i=1}^{T-j-1}(T-j-i)\,\rho_i + \sum_{i=0}^{j}(T-i)\,\rho_{j-i}\right).$$

On the other hand, Bhansali (1981), Shaman and Stine (1988), Kiviet and Phillips (1994) give formulae for $\mathrm{E}\,(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})$ in the AR(k) of (1) for fixed k. In this section, we illustrate the use of the link in (4) in the special case of an AR(2) process. Note that, when k is fixed and one wants to find $\mathrm{E}(\hat{\rho}_m)$ where $m>k$, one should use the formula for $\mathrm{E}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})$ in the overparameterized AR(m). The biases $\mathrm{E}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})$ of the parameters in the true AR(k) and the overparameterized AR(m) are not the same, although our general formula is unaltered, so care needs to be exercised when substituting for $\mathrm{E}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})$.

For an AR(k), we have $\delta = 1/2$ and $\hat{\boldsymbol{\beta}} = T^{1/2}\,(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}) \xrightarrow{d} \mathrm{N}\left(0,\sigma^2\boldsymbol{\Gamma}^{-1}\right)$ as $T\to\infty$, where $\boldsymbol{\Gamma}=\gamma_0\mathbf{R}$; see Brockwell and Davis (1991, p.241). In this section, we let $\sigma^2=1$ without loss of generality. Then,

$$\boldsymbol{\Gamma}\mathrm{E}(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}') = \boldsymbol{\Gamma}\mathrm{var}(\hat{\boldsymbol{\beta}}) + \boldsymbol{\Gamma}\mathrm{E}(\hat{\boldsymbol{\beta}})\mathrm{E}(\hat{\boldsymbol{\beta}})' = \mathbf{I}_k + o\,(1)$$

because $\mathrm{E}(\hat{\boldsymbol{\beta}}) = O\left(T^{-1/2}\right)$. The justification for using the asymptotic variance in place of the finite-sample variance follows from Larsson (1997).

The AR(1) case is straightforward. Let $k=1$. From Shaman and Stine (1988), noting that the coefficients $\boldsymbol{\alpha}$ have opposite signs to the ones we use here,

$$T^{1/2}\mathrm{E}(\hat{\beta}) = T\mathrm{E}\,(\hat{\alpha}-\alpha) = -\,(1+3\alpha) + o\,(1)\,.$$

Using our (4), substituting $\mathbf{C}=0$ and $\mathbf{D}_0 = 1 = \mathbf{R}$ gives $T\mathrm{E}(\hat{\rho}-\rho) \sim -\,(1+3\alpha)$, which is in accord with Kendall's formula. Note that $\alpha=\rho$ here.

Next, we use (4) to translate the bias for an AR(2) into a correlogram bias. We have $k=2$ and Shaman and Stine (1988) give $T^{1/2}\mathrm{E}(\hat{\boldsymbol{\beta}}) = -\,(b_1,b_2)' + o\,(1)$, where $b_1 := 1+\alpha_1+\alpha_2$ and $b_2 := 2+4\alpha_2$. Using $\boldsymbol{\alpha}=\mathbf{R}^{-1}\boldsymbol{\rho}$, we get

$$(7)\qquad \begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix} = \frac{1}{1-\rho_1^2}\begin{pmatrix}1 & -\rho_1\\ -\rho_1 & 1\end{pmatrix}\begin{pmatrix}\rho_1\\ \rho_2\end{pmatrix} = \frac{1}{1-\rho_1^2}\begin{pmatrix}\rho_1\,(1-\rho_2)\\ \rho_2-\rho_1^2\end{pmatrix},$$

hence

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \frac{1+2\rho_1+\rho_2}{1+\rho_1} \\ 2\frac{1+2\rho_2-3\rho_1^2}{1-\rho_1^2} \end{pmatrix}, \quad \frac{1}{1-\alpha_2} = \frac{1-\rho_1^2}{1-\rho_2}, \ \frac{\alpha_1}{1-\alpha_2} = \rho_1,$$

where $|a_2| < 1$ to satisfy our assumption of stationarity. We have

$$2\mathbf{N}_2\mathbf{C} = 2\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}, \tag{8}$$

implying

$$\mathbf{D}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \alpha_1 & 0 & \alpha_2 & 0 \\ 0 & \alpha_1 & 0 & \alpha_2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1-\alpha_2 & 0 \\ -\alpha_1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{1-\alpha_2} & 0 \\ \frac{\alpha_1}{1-\alpha_2} & 1 \end{pmatrix}^{-1}$$

and

$$\begin{aligned} & T^{-1/2}\mathbf{D}_0^{-1}\mathbf{R}\mathrm{E}\left(\hat{\boldsymbol{\beta}}\right) \\ = \ & -T^{-1}\begin{pmatrix} \frac{1-\rho_1^2}{1-\rho_2} & 0 \\ \rho_1 & 1 \end{pmatrix} \begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix} \begin{pmatrix} \frac{1+2\rho_1+\rho_2}{1+\rho_1} \\ 2\frac{1+2\rho_2-3\rho_1^2}{1-\rho_1^2} \end{pmatrix} + o\left(T^{-1}\right) \\ = \ & -T^{-1}\begin{pmatrix} \frac{(1+3\rho_1)\left(1+\rho_2-2\rho_1^2\right)}{1-\rho_2} \\ 2\frac{1+\rho_1-\rho_1^2-2\rho_1^3-3\rho_1^4+2\rho_2+\rho_1\rho_2+\rho_1^2\rho_2}{1-\rho_1^2} \end{pmatrix} + o\left(T^{-1}\right), \end{aligned} \tag{9}$$

which is the first term of the sum in (4).

As shown at the start of this section, $\mathrm{E}(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}') = \boldsymbol{\Gamma}^{-1} + o(1) = \gamma_0^{-1}\mathbf{R}^{-1} + o(1)$ where γ_0 is the long-run variance, given by (e.g. Brockwell and Davis, 1991, p.95)

$$\gamma_0 = \frac{1+\lambda_1^{-1}\lambda_2^{-1}}{\left(1-\lambda_1^{-2}\right)\left(1-\lambda_2^{-2}\right)\left(1-\lambda_1^{-1}\lambda_2^{-1}\right)} \tag{10}$$

where $\lambda_{1,2}$ are the characteristic roots $\lambda_{1,2} := -\alpha_1(1 \pm c)/(2\alpha_2)$, with $c := \sqrt{1+4\alpha_2/\alpha_1^2}$. Repeated roots can be obtained as a limiting case, and are not considered further here. By $\lambda_{1,2}^{-1} = \alpha_1 (1 \mp c)/2$, we have

$$\begin{aligned}\gamma_0 &= \frac{1 + \frac{1}{4}\alpha_1^2 \left(1 - c^2\right)}{\left(1 - \frac{1}{4}\alpha_1^2 (1 + c^2) + \frac{1}{2}\alpha_1^2 c\right)\left(1 - \frac{1}{4}\alpha_1^2 (1 + c^2) - \frac{1}{2}\alpha_1^2 c\right)\left(1 - \frac{1}{4}\alpha_1^2 (1 - c^2)\right)} \\ &= \frac{1 - \alpha_2}{\left[\left(1 - \frac{1}{4}\alpha_1^2 (1 + c^2)\right)^2 - \frac{1}{4}\alpha_1^4 c^2\right](1 + \alpha_2)} = \frac{1 - \rho_1^2}{(1 - \rho_2)\left(1 - 2\rho_1^2 + \rho_2\right)}.\end{aligned}$$

Hence,

$$\mathrm{E}\left(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}'\right) = \frac{(1 - \rho_2)\left(1 - 2\rho_1^2 + \rho_2\right)}{\left(1 - \rho_1^2\right)^2} \begin{pmatrix} 1 & -\rho_1 \\ -\rho_1 & 1 \end{pmatrix} + o(1).$$

For the second term of (4), we work out

$$\begin{aligned} 2\left(\hat{\boldsymbol{\beta}}' \otimes \left(\hat{\boldsymbol{\beta}}' \otimes \mathbf{I}_2\right)\mathbf{N}_2\mathbf{C}\right) &= \begin{pmatrix} \hat{\beta}_1 & \hat{\beta}_2 \end{pmatrix} \otimes \begin{pmatrix} \hat{\beta}_1 & 0 & \hat{\beta}_2 & 0 \\ 0 & \hat{\beta}_1 & 0 & \hat{\beta}_2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \hat{\beta}_1\hat{\beta}_2 & 0 & \hat{\beta}_2^2 & 0 \\ \hat{\beta}_1^2 & 0 & \hat{\beta}_1\hat{\beta}_2 & 0 \end{pmatrix}, \end{aligned}$$

and taking expectations gives the second term of the sum in (4) as

$$\begin{aligned} &\frac{(1 - \rho_2)\left(1 - 2\rho_1^2 + \rho_2\right)}{T\left(1 - \rho_1^2\right)^2}\mathbf{D}_0^{-1} \begin{pmatrix} -\rho_1 & 0 & 1 & 0 \\ 1 & 0 & -\rho_1 & 0 \end{pmatrix} \mathrm{vec}\left(\mathbf{D}_0^{-1}\mathbf{R}\right) \\ &= \frac{(1 - \rho_2)\left(1 - 2\rho_1^2 + \rho_2\right)}{\left(1 - \rho_1^2\right)^2} \begin{pmatrix} \frac{1-\rho_1^2}{1-\rho_2} & 0 \\ \rho_1 & 1 \end{pmatrix} \begin{pmatrix} -\rho_1 & 0 & 1 & 0 \\ 1 & 0 & -\rho_1 & 0 \end{pmatrix} \begin{pmatrix} \frac{1-\rho_1^2}{1-\rho_2} \\ 2\rho_1 \\ \frac{1-\rho_1^2}{1-\rho_2}\rho_1 \\ 1 + \rho_1^2 \end{pmatrix} \\ &= \frac{1 - 2\rho_1^2 + \rho_2}{T} \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{aligned}$$

Together with (9), this yields

$$T\mathrm{E}\left(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\right) = -\begin{pmatrix} \frac{(1+3\rho_1)\left(1+\rho_2-2\rho_1^2\right)}{1-\rho_2} \\ \frac{1+2\rho_1+3\rho_2+\rho_1^2+2\rho_1\rho_2-4\rho_1^3+3\rho_1^2\rho_2-8\rho_1^4}{1-\rho_1^2} \end{pmatrix} + o(1). \tag{11}$$

The first two correlogram biases of Kendall's formula (5) simplify to (11), by using the recursion $\rho_j = \alpha_1\rho_{j-1} + \alpha_2\rho_{j-2}$ for $j = 2, 3, \ldots$.

REFERENCES

Abadir, K.M. (1993). OLS bias in a nonstationary autoregression. *Econometric Theory*, 9, 81-93.

Abadir, K.M. and J.R. Magnus (2002). Notation in econometrics: a proposal for a standard. *Econometrics Journal*, 5, 76-90.

Abadir, K.M. and J.R. Magnus (2005). *Matrix Algebra*. Cambridge University Press, New York.

Abadir, K. M. and G. Talmain (2011). The unconventional dynamics of economic and financial aggregates. Chapter 8 in: A. Ullah and D. E. A. Giles (eds.) *Handbook of Empirical Economics and Finance*, Chapman & Hall/CRC.

Bhansali, R.J. (1981). Effects of not knowing the order of an autoregressive process on the mean squared error of prediction-1. *Journal of the American Statistical Association*, 76, 588-597.

Brockwell, P.J. and R.A. Davis (1991). *Time Series: Theory and Methods*, 2nd ed. Springer-Verlag, New York.

Giraitis, L., J. Hidalgo, and P.M. Robinson (2001). Gaussian estimation of parametric spectral density with unknown pole. *Annals of Statistics*, 29, 987-1023.

Giraitis, L. and R. Leipus (1995). A generalized fractionally differencing approach in long memory modelling. *Lithuanian Mathematical Journal*, 35, 53-65.

Gray, H.L. , N. Zhang, and W.A. Woodward (1989). On generalized fractional processes. *Journal of Time Series Analysis*, 10, 233-257. [Correction (1994), 15, 561-562.]

Hidalgo, J. (2005). Semiparametric estimation for stationary processes whose spectra have an unknown pole. *Annals of Statistics*, 33, 1843-1889.

Kendall, M.G. (1954). Note on bias in the estimation of autocorrelation. *Biometrika*, 41, 403-404.

Kiviet, J.F. and G.D.A. Phillips (1994). Bias assessment and reduction in linear error-correction models. *Journal of Econometrics*, 63, 215-243.

Larsson, R. (1997). On the asymptotic expectations of some unit root tests in a first order autoregressive process in the presence of trend. *Annals of the Institute of Statistical Mathematics,* 49, 585-599.

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