Statistical Papers Vol. 42, No. 4 (2001), P. 517 to 527

Note: This version has been checked as identical to the printed version. Manfred Mudelsee, 06 December 2001

Note on the bias in the estimation of the serial correlation coefficient of AR(1) processes

Manfred Mudelsee ⋆

Institute of Meteorology, University of Leipzig, Stephanstraße 3, 04103 Leipzig, Germany

Received: November 30, 1999; revised version: July 3, 2000

Abstract We derive approximating formulas for the mean and the variance of an autocorrelation estimator which are of practical use over the entire range of the autocorrelation coefficient ρ. The least-squares estimator $\sum_{i=1}^{n-1} \epsilon_i \epsilon_{i+1} \ / \ \sum_{i=1}^{n-1} \epsilon_i^2$ is studied for a stationary AR(1) process with known mean. We use the second order Taylor expansion of a ratio, and employ the arithmetic–geometric series instead of replacing partial Cesàro sums. In case of the mean we derive Marriott and Pope's (1954) formula, with $(n-1)^{-1}$ instead of $(n)^{-1}$, and an additional term $\propto (n-1)^{-2}$. This new formula produces the expected decline to zero negative bias as ρ approaches unity. In case of the variance Bartlett's (1946) formula results, with $(n-1)^{-1}$ instead of $(n)^{-1}$. The theoretical expressions are corroborated with a simulation experiment. A comparison shows that our formula for the mean is more accurate than the higher-order approximation of White (1961), for $|\rho| > 0.88$ and $n \geq 20$. In principal, the presented method can be used to derive approximating formulas for other estimators and processes.

⋆ *Acknowledgement:* The present study was carried out while the author was Marie Curie Research Fellow (EU-TMR grant ERBFMBICT 971919) at the Institute of Mathematics and Statistics, University of Kent, Canterbury, UK.

1 Introduction

Consider the following well-known least-squares estimator of the serial correlation coefficient ρ of lag 1 from a time series ϵ_i ($i = 1, \ldots, n$) sampled from a process $\mathcal{E}_i$ with known (zero) mean:

$$\widehat{\rho} = \sum_{i=1}^{n-1} \epsilon_i \epsilon_{i+1} \Bigg/ \sum_{i=1}^{n-1} \epsilon_i^2. \tag{1}$$

Let $\mathcal{E}_i$ be the stationary AR(1) process,

$$\mathcal{E}_1 \sim N(0,\ 1), \qquad \mathcal{E}_i = \rho\, \mathcal{E}_{i-1} + \mathcal{U}_i, \qquad i = 2, \ldots, n, \tag{2}$$

with $\mathcal{U}_i$ independent and identically distributed as $N(0,\ 1 - \rho^2)$ and $|\rho| < 1$. Then the variance and the mean of $\widehat{\rho}$ are approximately given by the general formulas (for arbitrary lag) of Bartlett (1946) and Marriott and Pope (1954), respectively:

$$\mathrm{var}(\widehat{\rho}) \simeq \mathrm{var}'(\widehat{\rho})_B = \frac{(1-\rho^2)}{n}, \tag{3}$$

$$\mathrm{E}(\widehat{\rho}) \simeq \mathrm{E}'(\widehat{\rho})_{MP} = \rho - \frac{2\,\rho}{n}. \tag{4}$$

(The indices refer to the two studies.)

An example for an application is fitting an AR(1) model to real data for which the mean is a priori known and can be subtracted. Formulas for $\mathrm{E}(\widehat{\rho})$ can be used for bias correction, whereas $\mathrm{var}(\widehat{\rho})$ gives the uncertainty of the estimated fit parameter $\widehat{\rho}$.

Approximations (3) and (4) are based on the second order Taylor expansion of a ratio. The primes indicate that the formulas are further based on the replacement of partial Cesàro sums, like

$$\sum_{r=-n+1}^{n-1} \left(1 - \frac{|r|}{n}\right) \rho^{|r|}$$

with

$$\sum_{r=-\infty}^{+\infty} \rho^{|r|},$$

see, for example, Priestley (1981, section 5.3). Evidently, the quality of the latter simplification is in doubt when $|\rho|$ is large (Bartlett 1946; Kendall 1954). In particular: as $\rho \to 1$, all time series points become equal and $\widehat{\rho} \to 1$. However, (4) fails to produce that zero bias as $\rho \to 1$. This failure is also shared by published approximations of higher order. For example, the formulas of White (1961) for estimator (1) and process (2) are, to our best knowledge, the most accurate in terms of powers of $(1/n)$:

$$\operatorname{var}(\widehat{\rho}) \simeq \operatorname{var}(\widehat{\rho})_W = \left(\frac{1}{n} - \frac{1}{n^2} + \frac{5}{n^3}\right) - \left(\frac{1}{n} - \frac{9}{n^2} + \frac{53}{n^3}\right)\rho^2 - \frac{12}{n^3}\rho^4, \tag{5}$$

$$\mathrm{E}(\widehat{\rho}) \simeq \mathrm{E}(\widehat{\rho})_W = \left(1 - \frac{2}{n} + \frac{4}{n^2} - \frac{2}{n^3}\right)\rho + \frac{2}{n^2}\rho^3 + \frac{2}{n^2}\rho^5. \tag{6}$$

For $\rho \neq 0$, Eq. (6) cannot produce zero bias.

The true zero variance of $\widehat{\rho}$ as $\rho \to 1$, on the other hand, is produced by Bartlett's formula, (3), whereas White's formula, (5), fails to do so.

The same weaknesses of formulas (4), (5) and (6) exist also for $\rho \to -1$.

Since approximations (3)–(6) are not intended for large $|\rho|$, our focus is to derive approximating formulas for $\operatorname{var}(\widehat{\rho})$ and $\mathrm{E}(\widehat{\rho})$ which can be used over the entire range of ρ, in particular for $\rho \to \pm 1$. This is done in Section 2. Thereby we also use the second order Taylor expansion. However, no further simplification is made: we employ the arithmetic–geometric series instead of replacing partial Cesàro sums. This yields an approximation for $\mathrm{E}(\widehat{\rho})$ which consists of Marriott and Pope's formula, (4), with an additional term; our approximation produces zero bias for $\rho \to \pm 1$. We further confirm Bartlett's result, (3), for $\operatorname{var}(\widehat{\rho})$. The order of the error for the different approximating formulas is considered. The theoretical expressions are examined over the entire positive range of ρ with a simulation experiment (Section 3), for different values of n. As a practical consequence of this experiment, we give the range of ρ for which our formula for $\mathrm{E}(\widehat{\rho})$ should be used instead of (6) or (4).

2 Series expansions

For the derivations we assume $\rho \neq 0$.

2.1 Variance

We write (1) as

$$\begin{aligned}\widehat{\rho} &= \left(\frac{1}{n} \sum_{i=1}^{n-1} \epsilon_i \epsilon_{i+1} \right) \bigg/ \left(\frac{1}{n} \sum_{i=1}^{n-1} \epsilon_i^2 \right) \\ &= c \,/\, v \text{ , say,}\end{aligned}$$

and then we have, up to the second order of deviations from the means of c and v, the variance

$$\begin{aligned}\operatorname{var}(\widehat{\rho}) = \operatorname{var}(c/v) \simeq \frac{\operatorname{var}(c)}{\mathrm{E}^2(v)} - 2\,\mathrm{E}(c) \frac{\operatorname{cov}(v,c)}{\mathrm{E}^3(v)} \\ + \mathrm{E}^2(c) \frac{\operatorname{var}(v)}{\mathrm{E}^4(v)}.\end{aligned} \tag{7}$$

We now apply a standard result for quadravariate standard Gaussian distributions with serial correlations ρ_j (e. g. Priestley 1981, p. 325), namely

$$\operatorname{cov}(\epsilon_a \epsilon_{a+s}, \epsilon_b \epsilon_{b+s+t}) = \rho_{b-a}\, \rho_{b-a+t} \;+\; \rho_{b-a+s+t}\, \rho_{b-a-s},$$

from which we derive, for ϵ_i following process (2):

$$\begin{aligned}&\operatorname{cov}\left(\frac{1}{n} \sum_{a=1}^{n-1} \epsilon_a \epsilon_{a+s}, \frac{1}{n} \sum_{b=1}^{n-1} \epsilon_b \epsilon_{b+s+t} \right) \\ &= \frac{1}{n^2} \sum_{a,b=1}^{n-1} (\rho^{|b-a|}\, \rho^{|b-a+t|} \;+\; \rho^{|b-a+s+t|}\, \rho^{|b-a-s|}),\end{aligned} \tag{8}$$

where the symbol $\sum_{a,b}$ represents a double summation. Without further simplification we can use this equation to directly derive the various constituents of the right-hand side of (7). Firstly, we consider $\operatorname{var}(c)$, which is given by putting $s = 1$ and $t = 0$ in Eq. (8):

$$\begin{aligned}\operatorname{var}(c) &= \operatorname{var}\left(\frac{1}{n} \sum_{i=1}^{n-1} \epsilon_i \epsilon_{i+1} \right) \\ &= \frac{1}{n^2} \left(\sum_{a,b=1}^{n-1} \rho^{2|b-a|} + \sum_{a,b=1}^{n-1} \rho^{|b-a+1|}\, \rho^{|b-a-1|} \right).\end{aligned}$$

We find, by summing over the points of the a-b grid in a 'diagonal' manner, and by using the arithmetic-geometric series, that

$$\sum_{a,b=1}^{n-1} \rho^{2|b-a|} = (n-1) + 2\sum_{i=1}^{n-2} i\,\rho^{2(n-i-1)}$$

$$= (n-1)\left(\frac{2}{1-\rho^2} - 1\right) + 2\,\frac{(\rho^{2n-4} - \rho^{-2})}{(1-\rho^{-2})^2} \quad (9)$$

and

$$\sum_{a,b=1}^{n-1} \rho^{|b-a+1|}\,\rho^{|b-a-1|} = (n-1)\,\rho^2 + 2\sum_{i=1}^{n-2} i\,\rho^{2(n-i-1)}$$

$$= (n-1)\left(\frac{2}{1-\rho^2} + \rho^2 - 2\right)$$

$$+2\,\frac{(\rho^{2n-4} - \rho^{-2})}{(1-\rho^{-2})^2}. \quad (10)$$

Equations (9) and (10) together give $\mathrm{var}(c)$. Further, by putting $s = 0$ and $t = 1$ in Eq. (8),

$$\mathrm{cov}(v,c) = \mathrm{cov}\left(\frac{1}{n}\sum_{i=1}^{n-1}\epsilon_i^2, \frac{1}{n}\sum_{i=1}^{n-1}\epsilon_i\epsilon_{i+1}\right)$$

$$= \frac{2}{n^2}\sum_{a,b=1}^{n-1} \rho^{|b-a|}\,\rho^{|b-a+1|},$$

where

$$\sum_{a,b=1}^{n-1} \rho^{|b-a|}\,\rho^{|b-a+1|} = \rho^{2n-3} + \sum_{i=1}^{n-2}(2\,i+1)\,\rho^{2(n-i)-3}$$

$$= (n-1)\,\frac{2}{\rho}\left(\frac{1}{1-\rho^2} - 1\right)$$

$$+\,2\,\frac{(\rho^{2n-5} - \rho^{-3})}{(1-\rho^{-2})^2}$$

$$+\,\frac{(\rho^{2n-3} - \rho^{-1})}{(1-\rho^{-2})}. \quad (11)$$

Similarly, by putting $s = t = 0$ in Eq. (8),

$$\mathrm{var}(v) = \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n-1}\epsilon_i^2\right) = \frac{2}{n^2}\sum_{a,b=1}^{n-1}\rho^{2|b-a|},$$

which is given by (9). We finally need

$$\mathrm{E}(v) = \frac{n-1}{n} \tag{12}$$

and

$$\mathrm{E}(c) = \frac{\rho(n-1)}{n}. \tag{13}$$

All three terms contributing to the right-hand side of (7) have a part proportional to $(n-1)^{-2}$. However, these parts cancel out to give:

$$\mathrm{var}(\widehat{\rho}) \simeq \mathrm{var}(\widehat{\rho})_M = \frac{(1-\rho^2)}{(n-1)}. \tag{14}$$

(The index refers to this study.) This expression for $\mathrm{var}(\widehat{\rho})_M$ holds also for $\rho = 0$.

2.2 Mean

We have, up to the second order of deviations from the means of c and v, the mean

$$\mathrm{E}(\widehat{\rho}) = \mathrm{E}(c/v) \simeq \frac{\mathrm{E}(c)}{\mathrm{E}(v)} - \frac{\mathrm{cov}(v,c)}{\mathrm{E}^2(v)} + \mathrm{E}(c)\frac{\mathrm{var}(v)}{\mathrm{E}^3(v)}. \tag{15}$$

Making use of previous expressions for $\mathrm{cov}(v, c)$, $\mathrm{var}(v)$, $\mathrm{E}(v)$ and $\mathrm{E}(c)$, we obtain

$$\mathrm{E}(\widehat{\rho}) \simeq \mathrm{E}(\widehat{\rho})_M = \rho - \frac{2\,\rho}{(n-1)} + \frac{2}{(n-1)^2}\frac{(\rho - \rho^{2n-1})}{(1-\rho^2)}. \tag{16}$$

This expression for $\mathrm{E}(\widehat{\rho})_M$ holds also for $\rho = 0$. Letting $\rho \to \pm 1$, $\mathrm{E}(\widehat{\rho})_M$ approaches ± 1.

2.3 Approximation error

Anderson (1995) showed that the error of approximation (3) is of $\mathcal{O}(n^{-2})$ and the error of (4) is of $\mathcal{O}(n^{-1})$. Since we also used the second order Taylor expansion, the same orders are valid also for the approximating formulas derived here, (14) and (16) respectively. Thus, our expression for the variance, (14), cannot be distinguished from Bartlett's result, (3). This is interesting since we have not replaced partial Cesáro sums, while Bartlett has. The first two terms of our expression for the mean, (16), correspond to Marriott and Pope's result, (4).

An additional source of approximation error is due to replacing partial Cesáro sums; this error, for $|\rho|$ large, may be larger than the error due to finite n.

White (1961) derived formulas (5) and (6) in a different manner. He calculated the kth moment of (c/v) via the joint moment generating function $m(c, v)$, with c and v defined as here (Section 2.1) without the factors $1/n$,

$$\mathrm{E}(\widehat{\rho}^k) = \int_{-\infty}^{0} \int_{-\infty}^{v_k} \ldots \int_{-\infty}^{v_2} \left. \frac{\partial^k m(c, v)}{\partial c^k} \right|_{c=0} dv_1\, dv_2\, \ldots\, dv_k.$$

He expanded the integrand to terms of order n^{-3} and ρ^4.

3 Simulation experiment

We generated a set of 250 000 time series from process (2) for each combination of n (20, 100) and ρ (densely distributed over the interval [0; 1[). For each time series $\widehat{\rho}$ was calculated, yielding sample means, $\mathrm{E}(\widehat{\rho})_S$, and sample standard deviations. These can be compared with the theoretical values calculated from formulas (14) and (16), respectively. We include White's (1961) formulas, (5) and (6), in the comparison. Figure 1 shows plots of the negative bias, $\rho - \mathrm{E}(\widehat{\rho})$, and of $\sqrt{\mathrm{var}(\widehat{\rho})}$.

In general, the deviations between the theoretical results and the simulation results decrease with increasing n, as is to be expected.

Up to a certain value of ρ, White's (1961) formulas perform better than ours since they are more accurate with respect to powers of $(1/n)$. For larger ρ, our formulas give results closer to

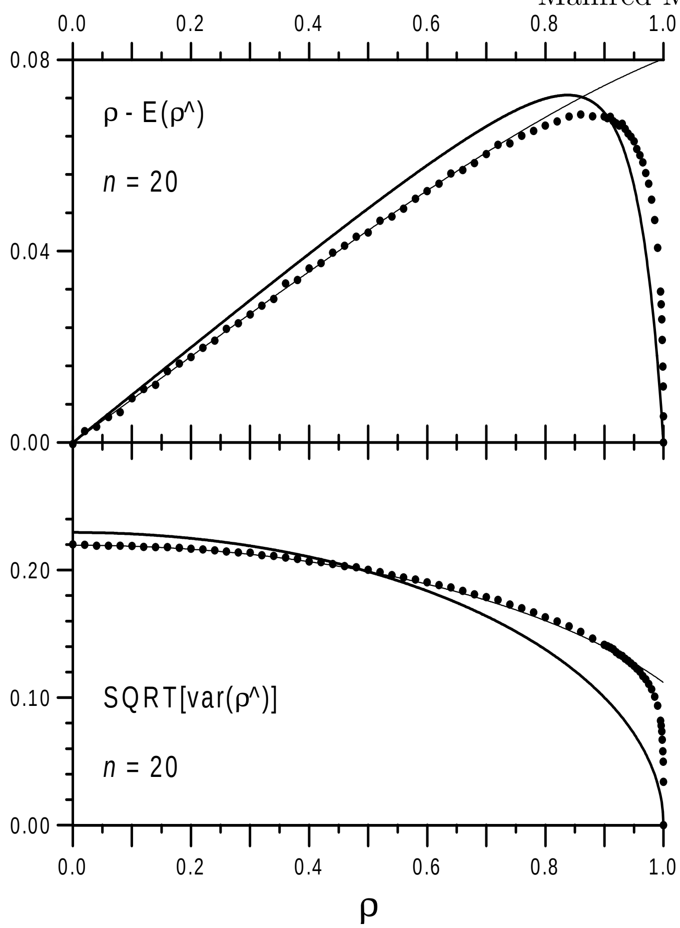

Figure 1: Simulation experiment. Top panel: negative bias, bottom panel: standard deviation. Simulation results (dots). Theoretical result, our formulas, (14) and (16), respectively, (heavy line). Theoretical result, White's (1961) formulas, (5) and (6), respectively, (light line).

the simulation, particularly the decline to zero negative bias as ρ approaches unity. This decline is caused by the term proportional to $(n-1)^{-2}$ in (16). For ρ large, and n small, this term contributes heavier to $\mathrm{E}(\widehat{\rho})$.

Let $\rho_=$ be defined by

$$|\mathrm{E}(\widehat{\rho})_M - \mathrm{E}(\widehat{\rho})_S| < |\mathrm{E}(\widehat{\rho})_W - \mathrm{E}(\widehat{\rho})_S| \text{ for } \rho > \rho_=,$$

that means, $\rho_=$ gives the range for which our formula for the mean, (16), is closer to the simulation than White's formula,

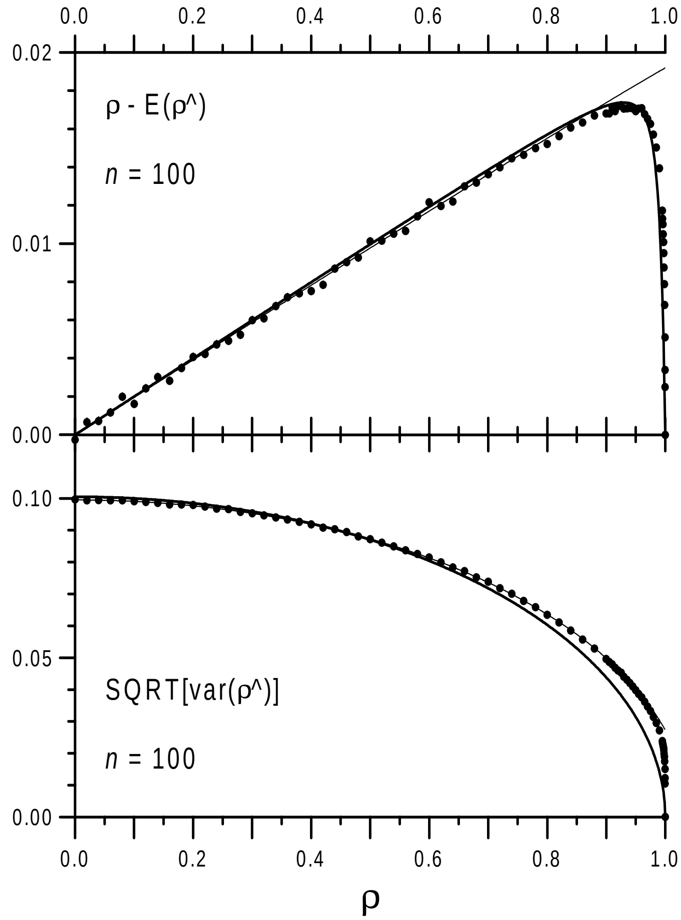

Figure 1: (Continued.)

(6). Since $\rho_=$ should depend on n, we carried out additional simulations as above for n = 10, 50, 200, 500 and 1000 (where necessary with an increased number of time series in order to suppress the simulation noise). It turned out that for $n \geq 20$, $\rho_=(n)$ is given by the intersection between the two theoretical formulas ($\mathrm{E}(\widehat{\rho})_M = \mathrm{E}(\widehat{\rho})_W$) which is plotted in Fig. 2. As can be seen in Fig. 2, $\rho_=(n)$ does not strongly depend on n. The practical consequence is that our formula for the mean, (16), should be used when $|\rho|$ is larger than about 0.88 and $n \geq 20$.

In case of the variance, $\mathrm{var}(\widehat{\rho})_W$ gives results closer to the simulation than $\mathrm{var}(\widehat{\rho})_M$ (and $\mathrm{var}(\widehat{\rho})_B$) for nearly the entire range of ρ (Fig. 1). We consider this is due to the fact that $\mathrm{var}(\widehat{\rho})$ de-

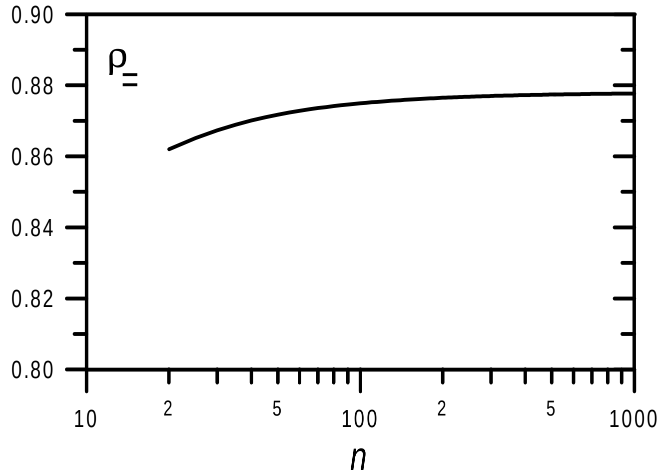

Figure 2: Comparison between White's approximation, (6), and our approximation, (16), for the mean. For $\rho > \rho_=$ our formula is more accurate, and vice versa (see text for details).

creases monotonically with $|\rho|$—a behaviour that is also shown by White's formula. The fact that $\mathrm{var}(\widehat{\rho})_B$ describes the simulation better than $\mathrm{var}(\widehat{\rho})_M$ for $\rho \rightarrow 0$, is regarded as spurious.

4 Extensions

In the same manner as here, approximations can be obtained for other estimators (cf. Box *et al.* 1994, section A7.4) of the serial correlation coefficient. Consider, for example, the Yule–Walker estimator $\widehat{\rho}_{\mathrm{YW}} = \sum_{i=1}^{n-1} \epsilon_i \epsilon_{i+1} \,/ \sum_{i=1}^{n} \epsilon_i^2$. Because here the sums differ in the number of terms, the analogue to Eq. (8) has an additional term, and the approximations for the variance and the mean are more complicated than Eqs. (14) and (16), respectively. However, the principal behaviours for ρ approaching unity are similar in case of $\widehat{\rho}_{\mathrm{YW}}$: the decline of the negative bias (to a value of $1/n$) and the approach of zero variance. An experiment with $n = 20$ and 250 000 simulations, corroborating the approximations, showed that $\widehat{\rho}_{\mathrm{YW}}$ is more biased and has a smaller variance than the least-squares estimator. For $\rho > 0.5$, the mean squared error (MSE, given by bias^2 + variance) of $\widehat{\rho}_{\mathrm{YW}}$ is larger than the MSE of the least-squares estimator, and vice versa for $\rho < 0.5$.

Our approximation formulas apply also to AR(1) processes with constant but unknown variance, and asymptotic stationary AR(1) processes (transient behaviour). In principal, it should be possible to obtain formulas for estimators of arbitrary lag in the same manner as here.

References

Anderson, O.D. (1995). The precision of approximations to moments of sample moments. J. Statist. Comput. Simul. **52,** 163–165.

Bartlett, M.S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. J. Roy. Statist. Soc. Supplement **8,** 27–41 (Corrigenda: (1948) J. Roy. Statist. Soc. Ser. B **10,** No. 1).

Box, G.E.P., Jenkins, G.M., Reinsel, G.C. (1994). Time series analysis: forecasting and control, 3rd edn. Prentice-Hall, Englewood Cliffs, NJ, 598 pp.

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

Marriott, F.H.C., Pope, J.A. (1954). Bias in the estimation of autocorrelations. Biometrika **41,** 390–402.

Priestley, M.B. (1981). Spectral Analysis and Time Series, Academic Press, London.

White, J.S. (1961). Asymptotic expansions for the mean and variance of the serial correlation coefficient. Biometrika **48,** 85–94.