Revisiting simple linear regression with autocorrelated errors

By JAECHOUL LEE and ROBERT LUND

Department of Statistics, The University of Georgia, Athens, Georgia 30602-1952, U.S.A.

jaechlee@stat.uga.edu   lund@stat.uga.edu

Summary

This paper studies properties of ordinary and generalised least squares estimators in a simple linear regression with stationary autocorrelated errors. Explicit expressions for the variances of the regression parameter estimators are derived for some common time series autocorrelation structures, including a first-order autoregression and general moving averages. Applications of the results include confidence intervals and an example where the variance of the trend slope estimator does not increase with increasing autocorrelation.

Some key words: Generalised least squares; Ordinary least squares; Simple linear regression; Time series.

1. Introduction

This paper studies the ordinary and generalised least squares parameter estimators in the simple linear regression model

$$X_t = \mu + \alpha t + \varepsilon_t \tag{1.1}$$

when the errors $\{\varepsilon_t\}$ have mean zero and are stationary in time t with autocovariance $\gamma(h) = \text{cov}(\varepsilon_t, \varepsilon_{t+h})$ at lag h. The model in (1.1) has a fundamental role in many statistical analyses (Grenander, 1954; Harvey & Phillips, 1979; Zinde-Walsh & Galbraith, 1991; Choudhury et al., 1999). There is little loss of generality in considering (1.1) over the more general regression $X_i = \mu + \alpha t_i + \varepsilon_i$, and hence we take $t_i = i$ from now on.

Ordinary least squares estimators, which have minimal variance when the $\{\varepsilon_t\}$ are uncorrelated, are frequently computed instead of generalised, weighted, least squares estimators when $\{\varepsilon_t\}$ are in truth autocorrelated. Grenander (1954) shows that ordinary and generalised least squares estimators have the same asymptotic efficiency under very general conditions on a stationary $\{\varepsilon_t\}$. In short, ordinary least squares estimators are attractive in reasonable generality. This paper derives explicit forms for the variances of the ordinary least squares estimators under some autocorrelation structures commonly encountered in time series and draws some conclusions from these forms.

The ordinary least squares estimators of the regression parameters in (1.1), denoted by $\hat{\mu}_{\text{OLS}}$ and $\hat{\alpha}_{\text{OLS}}$, are

$$\hat{\alpha}_{\text{OLS}} = \frac{\sum_{t=1}^n (t - \bar{t})(X_t - \bar{X})}{\sum_{t=1}^n (t - \bar{t})^2}, \quad \hat{\mu}_{\text{OLS}} = \bar{X} - \hat{\alpha}_{\text{OLS}} \bar{t}, \tag{1.2}$$

where $\bar{t} = (n+1)/2$ and $\bar{X} = n^{-1} \sum_{t=1}^n X_t$ are the time and observation averages, and $\sum_{t=1}^n (t - \bar{t})^2 = n(n+1)(n-1)/12$. Alternatively,

$$\begin{pmatrix} \hat{\mu}_{\text{OLS}} \\ \hat{\alpha}_{\text{OLS}} \end{pmatrix} = (D_n' D_n)^{-1} D_n' X^{(n)}, \tag{1.3}$$

where $X^{(n)} = (X_1, \ldots, X_n)'$ and $D_n = (1^{(n)}, i^{(n)})$ is the design matrix. Here, $1^{(n)} = (1, \ldots, 1)'$ and $i^{(n)} = (1, 2, \ldots, n)'$. The estimators $\hat{\mu}_{\text{OLS}}$ and $\hat{\alpha}_{\text{OLS}}$ are unbiased for any zero-mean $\{\varepsilon_t\}$; however, ordinary least squares estimators will not have the smallest variances amongst all unbiased estimators unless $\{\varepsilon_t\}$ are uncorrelated and with a constant variance.

The minimum variance unbiased estimators of μ and α that are linear combinations of $X_1, \ldots, X_n$, denoted by $\hat{\mu}_{\text{GLS}}$ and $\hat{\alpha}_{\text{GLS}}$, are the generalised least squares estimators

$$\begin{pmatrix} \hat{\mu}_{\text{GLS}} \\ \hat{\alpha}_{\text{GLS}} \end{pmatrix} = (D_n' \Gamma_n^{-1} D_n)^{-1} D_n' \Gamma_n^{-1} X^{(n)}. \tag{1.4}$$

In (1·4), Γ_n denotes the covariance matrix of $X^{(n)}$ which is tacitly assumed invertible for each $n \geqslant 1$. In general, generalised least squares estimators do not admit a simple explicit form akin to (1·2).

The generalised least squares estimators are also unbiased for any zero mean $\{\varepsilon_t\}$ and have a smaller variance than the ordinary least squares estimators: $\text{var}(\hat{\mu}_{\text{GLS}}) \leqslant \text{var}(\hat{\mu}_{\text{OLS}})$ and $\text{var}(\hat{\alpha}_{\text{GLS}}) \leqslant \text{var}(\hat{\alpha}_{\text{OLS}})$. The ordinary and generalised estimators are equal if and only if the columns of D_n span the same linear subspace as the columns of $\Gamma_n D_n$ (Bloomfield & Watson, 1975).

2. Variances of the estimators

We have that

$$\text{var}\begin{pmatrix} \hat{\mu}_{\text{OLS}} \\ \hat{\alpha}_{\text{OLS}} \end{pmatrix} = (D_n' D_n)^{-1} (D_n' \Gamma_n D_n)(D_n' D_n)^{-1}, \tag{2.1}$$

$$\text{var}\begin{pmatrix} \hat{\mu}_{\text{GLS}} \\ \hat{\alpha}_{\text{GLS}} \end{pmatrix} = (D_n' \Gamma_n^{-1} D_n)^{-1}. \tag{2.2}$$

There is a wealth of literature comparing the above variances, providing several interesting bounds; see for example Grenander (1954), Gurland (1954), Watson (1955, 1967), Zyskind (1967), Knott (1975) and Chipman (1979) for a historical sample.

Further calculations give

$$\text{var}(\hat{\alpha}_{\text{OLS}}) = \frac{\gamma(0) + 2 \sum_{j=1}^{n-1} w_j \gamma(j)}{\sum_{t=1}^{n} (t - \bar{t})^2}, \tag{2.3}$$

$$\text{var}(\hat{\mu}_{\text{OLS}}) = \frac{1}{n}\left(4\left(\frac{2n+1}{2n-2}\right)\gamma(0) + 2 \sum_{j=1}^{n-1}\left[1 - \frac{j}{n} + \left\{\frac{3(n+1)}{n-1}\right\} w_j \right]\gamma(j) \right), \tag{2.4}$$

where, for $j = 0, \ldots, n-1$, the weights $\{w_j\}$ are

$$w_j = \frac{\sum_{t=1}^{n-j}(t - \bar{t})(t + j - \bar{t})}{\sum_{t=1}^{n}(t - \bar{t})^2}$$

$$= \frac{(1 - j/n)(1 - 2j/n - 2j^2/n^2 - 1/n^2)}{(1 + 1/n)(1 - 1/n)}. \tag{2.5}$$

In addition,

$$\text{cov}(\hat{\mu}_{\text{OLS}}, \hat{\alpha}_{\text{OLS}}) = \frac{-\bar{t}}{\sum_{t=1}^{n}(t - \bar{t})^2}\left\{ \gamma(0) + 2 \sum_{j=1}^{n-1} w_j \gamma(j) \right\}. \tag{2.6}$$

We note here some properties of $\{w_j\}$ for later use. The Cauchy–Schwarz inequality provides $|w_j| \leqslant w_0 = 1$ for each j. Taking a limit in (2·5) shows that $\lim_{n \to \infty} w_j = 1$ for each fixed j. Finally, it is easy to verify algebraically that $\sum_{j=1}^{n-1} w_j = -\frac{1}{2}$ for all $n \geqslant 2$.

3. APPLICATIONS

Example 1. Suppose that $\{\varepsilon_t\}$ is a qth-order moving average, such that

$$\varepsilon_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}, \tag{3.1}$$

where $\{Z_t\}$ is zero-mean white noise with variance σ^2. Then $\gamma(h) = \sigma^2 \sum_{i=h}^{q} \theta_i \theta_{i-h}$ for $h \geqslant 0$, where the convention $\theta_0 = 1$ is adopted. Note in particular that $\gamma(h) = 0$ when $|h| > q$. Using this covariance structure in (2.3) gives

$$\text{var}(\hat{\alpha}_{\text{OLS}}) = \sigma^2 \left\{ \sum_{i=0}^{q} \theta_i^2 + \sum_{j=1}^{q} \frac{2(n-j)}{n(n^2-1)} (n^2 - 2jn - 2j^2 - 1) \sum_{i=j}^{q} \theta_i \theta_{i-j} \right\} \bigg/ \sum_{t=1}^{n} (t - \bar{t})^2. \tag{3.2}$$

An analogous expression for $\text{var}(\hat{\mu}_{\text{OLS}})$ can be obtained by combining the moving-average autocovariance structure with (2.4); we omit the algebraic details.

Example 2. Suppose that $\{\varepsilon_t\}$ is a causal first-order autoregression satisfying

$$\varepsilon_t = \phi \varepsilon_{t-1} + Z_t, \tag{3.3}$$

where $|\phi| < 1$ and $\{Z_t\}$ is zero-mean white noise with variance σ^2. Then $\gamma(h) = \sigma^2 \phi^h (1 - \phi^2)^{-1}$ for $h \geqslant 0$, and (2.3) and (2.4) give the forms

$$\text{var}(\hat{\alpha}_{\text{OLS}}) = \frac{12\sigma^2}{n(n^2-1)(1-\phi)^4(1-\phi^2)} \left(1 + \sum_{i=1}^{n+3} \beta_i \phi^i \right), \tag{3.4}$$

$$\text{var}(\hat{\mu}_{\text{OLS}}) = \frac{2\sigma^2}{n(1-\phi)^4(1-\phi^2)} \left(\frac{2n+1}{n-1} + \sum_{i=1}^{n+3} \xi_i \phi^i \right), \tag{3.5}$$

where the only nonzero coefficients are

$$\beta_1 = -2(n+3)n^{-1}, \quad \beta_2 = 12(n^2+1)n^{-1}(n^2-1)^{-1}, \quad \beta_3 = 2(n-3)n^{-1}, \quad \beta_4 = -1,$$

$$\beta_{n+1} = -6(n+1)n^{-1}(n-1)^{-1}, \quad \beta_{n+2} = 12n^{-1}, \quad \beta_{n+3} = -6(n-1)n^{-1}(n+1)^{-1},$$

and

$$\xi_1 = -4(n+1)(n+2)n^{-1}(n-1)^{-1}, \quad \xi_2 = 4(5n^2 - n + 5)n^{-1}(n-1)^{-2},$$

$$\xi_3 = 4(n^2 - 2n - 2)n^{-1}(n-1)^{-1}, \quad \xi_4 = -(2n+1)(n-1)^{-1},$$

$$\xi_{n+1} = -4(n+2)(2n+1)n^{-1}(n-1)^{-2}, \quad \xi_{n+2} = 4(4n+5)n^{-1}(n-1)^{-1}, \quad \xi_{n+3} = -8n^{-1}.$$

Example 3 (*Confidence intervals*). For known autocovariances in a Gaussian series, confidence intervals take the form $\hat{\mu}_{\text{OLS}} \pm z_{\alpha/2} \text{var}(\hat{\mu}_{\text{OLS}})^{\frac{1}{2}}$ and $\hat{\alpha}_{\text{OLS}} \pm z_{\alpha/2} \text{var}(\hat{\alpha}_{\text{OLS}})^{\frac{1}{2}}$, where the variances are computed from (2.3) and (2.4), and where $z_{\alpha/2}$ is the $(1 - \alpha/2)$th quantile of the standard normal distribution. We focus on α here; confidence intervals for μ are analogously developed.

When autocovariances of $\{\varepsilon_t\}$ are unknown, one can substitute estimates of $\gamma(.)$ into (2.3) and use the interval $\hat{\alpha}_{\text{OLS}} \pm t_{\alpha/2} \text{var}(\hat{\alpha}_{\text{OLS}})^{\frac{1}{2}}$, where $t_{\alpha/2}$ denotes the $(1 - \alpha/2)$th quantile of the t distribution. The appropriate degrees of freedom for the t distribution, however, is an unresolved issue. In an unpublished National Center for Atmospheric Research technical report, D. Nychka, R. Buchberger, T. M. L. Wigley, B. D. Santer, K. E. Taylor and R. H. Jones consider this problem with the first-order autoregressive $\{\varepsilon_t\}$ in Example 2 with unknown $\phi \in (-1, 1)$ and σ^2. For inferences involving α, they suggest replacing the customary $n - 2$ degrees of freedom with the equivalent degrees of freedom $n_e - 2$, where n_e is

$$n_e = n \left\{ \frac{1 - \hat{\rho}(1) - 0.68n^{-\frac{1}{2}}}{1 + \hat{\rho}(1) + 0.68n^{-\frac{1}{2}}} \right\} \tag{3.6}$$

and $\hat{\rho}(1) = \sum_{t=1}^{n-1} R_t R_{t+1} (\sum_{t=1}^{n} R_t^2)^{-1} = \hat{\phi}_{\mathrm{NND}}$ is the nonnegative definite lag-one sample auto-correlation of the residual series $\{R_t\} = \{X_t - \hat{\mu}_{\mathrm{OLS}} - \hat{\alpha}_{\mathrm{OLS}}t\}$. The justification for the general form of (3·6) stems from an asymptotic correction involving (2·3): use $\gamma(h) = \sigma^2 \phi^h (1 - \phi^2)^{-1}$ for $h \geqslant 0$, the dominated convergence theorem, and note that $|w_j| \leqslant 1$ for all j, and $w_j \to 1$ as $n \to \infty$ for each fixed j to obtain

$$\lim_{n \to \infty} \frac{\gamma(0)/\sum_{t=1}^{n}(t-\bar{t})^2}{\gamma(0)(1 + 2\sum_{j=1}^{n-1} w_j \phi^j)/\sum_{t=1}^{n}(t-\bar{t})^2} = \frac{1-\phi}{1+\phi}. \tag{3·7}$$

In the left-hand side of (3·7), the numerator is the variance of $\hat{\alpha}_{\mathrm{OLS}}$ when the $\{\varepsilon_t\}$ are uncorrelated with variance $\gamma(0)$ and the denominator is the variance of $\hat{\alpha}_{\mathrm{OLS}}$ computed from (2·3) under the AR(1) model. Hence, we interpret n_e as the number of independent observations with variance $\gamma(0)$. In their report, Nychka et al. justify the $0 \cdot 68 n^{-\frac{1}{2}}$ term in (3·6) entirely through simulation, stating that the term helps to correct for bias in the sampling distribution of $\hat{\phi}_{\mathrm{NND}}$, and that likelihood methods do not significantly improve upon least squares methods.

For a fixed n, the explicit variances in § 2 can be used to improve accuracy. From (3·7), a better estimate of n_e for a finite sample size n is $n_e = n(1 + 2\sum_{j=1}^{n-1} w_j \phi^j)^{-1}$. Using this in (2·5) and performing a tedious algebraic computation gives

$$n_e = n \left\{ 1 + \frac{2\phi}{n(n+1)(n-1)(1-\phi)^4} \left(\sum_{i=0}^{n+2} \gamma_i \phi^i \right)^{-1} \right\}, \tag{3·8}$$

where the only nonzero γ_i's in (3·8) are $\gamma_0 = (n-3)(n^2-1)$, $\gamma_1 = -3\{n(n^2-1) - 2(n^2+1)\}$, $\gamma_2 = 3(n+1)(n-1)^2$, $\gamma_3 = -n(n^2-1)$, $\gamma_n = -3(n+1)^2$, $\gamma_{n+1} = 6(n^2-1)$ and $\gamma_{n+2} = -3(n-1)^2$.

As noted in the Nychka et al. report, biases in the estimators of ϕ and σ^2 must be taken into account with small n. Tjøstheim & Paulsen (1983) suggest using the bias correction $\hat{\phi}_{\mathrm{BC}} = \hat{\rho}(1) + n^{-1}\{1 + 4\hat{\rho}(1)\}$, rounded to 1 or -1 in cases where this estimate falls outside the range $|\phi| < 1$. Fuller (1996, Ch. 6) discusses bias correction of general autocorrelation estimators.

A simulation was conducted to compare further the above confidence intervals. Table 1 compares the width of the Nychka et al. interval

$$\hat{\alpha}_{\mathrm{OLS}} \pm t_{\alpha/2} \left\{ \frac{\sum_{t=1}^{n} R_t^2/(n_e - 2)}{\sum_{t=1}^{n}(t-\bar{t})^2} \right\}^{\frac{1}{2}} \tag{3·9}$$

with n_e as in (3·6) to a t interval with n_e as in (3·8) and the standard error

$$\mathrm{var}(\hat{\alpha}_{\mathrm{OLS}})^{\frac{1}{2}} = \left\{ \left(\frac{\hat{\sigma}_{\mathrm{BC}}^2}{1 - \hat{\phi}_{\mathrm{BC}}^2} \right) \frac{1 + 2\sum_{j=1}^{n-1} w_j \hat{\phi}_{\mathrm{BC}}^j}{\sum_{t=1}^{n}(t-\bar{t})^2} \right\}^{\frac{1}{2}}, \tag{3·10}$$

where $\hat{\sigma}_{\mathrm{BC}}^2 = (n-3)^{-1} \sum_{t=2}^{n} (R_t - \hat{\phi}_{\mathrm{BC}} R_{t-1})^2$ is a biased-corrected estimator of σ^2. Table 1 compares average confidence interval lengths of the two methods with empirical coverage probabilities in parentheses. To be specific, for the Nychka et al. method, the confidence interval length is

$$2t_{\alpha/2} \left\{ \frac{\sum_{t=1}^{n} R_t^2/(n_e - 2)}{\sum_{t=1}^{n}(t-\bar{t})^2} \right\}^{\frac{1}{2}}. \tag{3·11}$$

A 95% level of confidence was used; the $\{\varepsilon_t\}$ were simulated as Gaussian. One hundred thousand entries were generated for each table entry, so that uncertainty due to simulation is minimal. The parameters chosen for the simulation were $\mu = 0$, $\alpha = 1$ and $\sigma^2 = 1$, but sampling properties are mathematically invariant over the choice of μ and α.

Table 1 shows that the average length of the 'moment-corrected' interval is smaller than that for the Nychka et al. interval, dramatically so when n is small and ϕ is close to unity. The Nychka et al. interval performs poorly in some of the simulations because of the large frequency at which (3·6) returns the physical impossibility $n_e < 2$. In such cases, the simulations set $n_e = 3$, making the

Table 1: *Example 3. Simulation comparison of confidence interval lengths, with empirical coverage probabilities in parentheses*

n	ϕ	Length based on (3·8) and (3·10)		Length based on (3·9)	
25	0·00	0·114	(0·930)	0·134	(0·968)
25	0·25	0·150	(0·920)	0·187	(0·950)
25	0·50	0·225	(0·901)	0·594	(0·937)
25	0·75	0·385	(0·857)	5·423	(0·912)
100	0·00	0·0136	(0·945)	0·0148	(0·963)
100	0·25	0·0182	(0·943)	0·0195	(0·957)
100	0·75	0·0274	(0·940)	0·0301	(0·956)
100	0·00	0·0560	(0·930)	0·0726	(0·958)
500	0·00	0·00121	(0·950)	0·00126	(0·958)
500	0·25	0·00162	(0·948)	0·00167	(0·955)
500	0·50	0·00243	(0·949)	0·00253	(0·957)
500	0·75	0·00488	(0·947)	0·00523	(0·960)

Nychka et al. interval seem better than it really is. Equivalent degrees of freedom computed in (3·8) do not suffer from such structural defects. Whereas the empirical probability of coverage of the (3·9)-based interval is typically slightly closer to its target value of 95%, both methods return reasonable coverages that are more accurate with increasing n and smaller ϕ.

We now consider monotonicity of $\mathrm{var}(\hat{\alpha}_{\mathrm{OLS}})$ in autocovariance. Differentiating the expression derived for $\mathrm{var}(\hat{\alpha}_{\mathrm{OLS}})$ in (2·3) with respect to $\gamma(h)$ for a fixed $h \geqslant 1$ gives

$$\frac{\partial \, \mathrm{var}(\hat{\alpha}_{\mathrm{OLS}})}{\partial \gamma(h)} = 2w_h \left\{ \sum_{t=1}^{n} (t - \bar{t})^2 \right\}^{-1}.$$

Hence, $\mathrm{var}(\hat{\alpha}_{\mathrm{OLS}})$ increases with increasing $\gamma(h)$ when $w_h > 0$. In particular $\mathrm{var}(\hat{\alpha}_{\mathrm{OLS}})$ for the qth-order moving average in Example 3 increases with increasing $\gamma(h)$, for $1 \leqslant h \leqslant q$, whenever $n^2 - 2qn - 2q^2 - 1 \geqslant 0$, which holds whenever $n \geqslant 3q$. As for more general monotonicity, we offer the following perhaps surprising example.

Example 4. In general, $\mathrm{var}(\hat{\alpha}_{\mathrm{OLS}})$ may actually decrease with increasing autocorrelation in $\{\varepsilon_t\}$. To see this, consider the two autocovariances $\gamma^{(1)}(h) = I_{\{0\}}(h)$ and $\gamma^{(2)}(h) = (1 - \rho^2) + \rho^2 I_{\{0\}}(h)$, where $I_A(.)$ denotes the indicator function of the set A and $\rho \in (0, 1)$. Here, $\gamma^{(1)}$ is a white noise auto-covariance under which ordinary least squares estimators have minimal variance, and $\gamma^{(2)}$ is the autocovariance of the stationary series $\{\varepsilon_t\}$, where

$$\varepsilon_t = (1 - \rho^2)^{\frac{1}{2}} S + \rho W_t, \qquad (3·12)$$

and S is a zero-mean unit-variance random shift that is assumed uncorrelated with the zero-mean unit-variance white noise $\{W_t\}$. Note that $\gamma^{(2)} = \gamma^{(1)}$ at the extremum $\rho = 1$. We have scaled both series to make $\gamma^{(1)}(0) = \gamma^{(2)}(0) = 1$ for a common basis of comparison. Note that $\gamma^{(2)}(h) \nrightarrow 0$ as $h \to \infty$, nor is $\gamma^{(2)}$ absolutely summable over all positive lags.

Applying (2·3) to $\gamma^{(2)}$ and using $\sum_{j=1}^{n-1} w_j = -\frac{1}{2}$ gives

$$\mathrm{var}(\hat{\alpha}_{\mathrm{OLS}}) = \frac{\rho^2}{\sum_{t=1}^{n} (t - \bar{t})^2}, \qquad (3·13)$$

and hence $\mathrm{var}(\hat{\alpha}_{\mathrm{OLS}}) < \{\sum_{t=1}^{n} (t - \bar{t})^2\}^{-1}$, the bound being the variance of the ordinary least squares trend estimate with errors having autocovariances $\gamma^{(1)}$. As $\gamma^{(1)}(h) \leqslant \gamma^{(2)}(h)$ for all lags h, we have an example where larger autocorrelations lead to a smaller ordinary least squares variance.

Mathematically, the paradox is explained as follows. The stationary errors $\{\varepsilon_t\}$ in (3·12) give a constant $\hat{\alpha}_{\text{OLS}}$ for each ρ: model (3·12) merely shifts means, albeit randomly, and rescales the noises in (1·1) and $\hat{\alpha}_{\text{OLS}}$ is invariant over mean shifts in $\{X_t\}$ and white noise rescalings. However, the autocovariance of $\{\varepsilon_t\}$ at each fixed lag $h \geqslant 1$ is decreasing in ρ. At the extremum $\rho = 1$, $\{\varepsilon_t\}$ is white noise with a unit variance. The non-monotonicity is not rooted in ordinary least squares methods as $\hat{\alpha}_{\text{OLS}} = \hat{\alpha}_{\text{GLS}}$.

More generally, the above autocovariance structure arises in split-plot designs (Christensen, 2002, Ch. 11). Consider the following split-plot regression with a homogeneous time trend:

$$X_{i,t} = \mu + \alpha t + \varepsilon_{i,t}. \tag{3·14}$$

Here, the indices $1 \leqslant i \leqslant m$ and $1 \leqslant t \leqslant n$ correspond to whole-plot and sub-plot factors, respectively. Consider noise models $\varepsilon_{i,t} = (1 - \rho^2)^{\frac{1}{2}} S_i + \rho W_{i,t}$ where S_i is a random whole-plot effect, independent for each i, with zero mean and unit variance. The $\{W_{i,t}\}$ for each fixed i have zero mean and unit variance; across varying i, they are taken as independent and also independent of $\{S_i\}_{i=1}^m$. Equation (2·3) gives $\text{var}(\hat{\alpha}_{\text{OLS}}) = \rho^2 \{m^2 \sum_{t=1}^n (t - \bar{t})^2\}^{-1}$, which is increasing in ρ. At the extremum $\rho = 1$, the $\{\varepsilon_{i,t}\}$ are unit-variance white noise.

References

BLOOMFIELD, P. & WATSON, G. S. (1975). The inefficiency of least squares. *Biometrika* **62**, 88–116.
CHIPMAN, J. S. (1979). Efficiency of least-squares estimation of linear trends when residuals are autocorrelated. *Econometrica* **47**, 115–28.
CHOUDHURY, A. H., HUBATA, R. & ST. LOUIS, R. D. (1999). Understanding time-series regression estimators. *Am. Statistician* **53**, 342–8.
CHRISTENSEN, R. A. (2002). *Plane Answers to Complex Questions*, 3rd ed. New York: Springer.
FULLER, W. (1996). *Introduction to Statistical Time Series*, 2nd ed. New York: John Wiley & Sons.
GRENANDER, U. (1954). On the estimation of regression coefficients in the case of autocorrelated disturbances. *Ann. Math. Statist.* **25**, 252–72.
GURLAND, J. (1954). An example of autocorrelated disturbances in linear regression. *Econometrica* **22**, 218–27.
HARVEY, A. C. & PHILLIPS, G. D. A. (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika* **66**, 49–58.
KNOTT, M. (1975). On the minimum efficiency of least squares. *Biometrika* **62**, 129–32.
TJØSTHEIM, D. & PAULSEN, J. (1983). Bias of some commonly-used time series estimates. *Biometrika* **70**, 389–99.
WATSON, G. S. (1955). Serial correlation in regression analysis I. *Biometrika* **42**, 327–41.
WATSON, G. S. (1967). Linear least squares regression. *Ann. Math. Statist.* **38**, 1679–99.
ZINDE-WALSH, V. & GALBRAITH, J. W. (1991). Estimators of a linear regression model with stationary ARMA(p,q) errors. *J. Economet.* **47**, 333–57.
ZYSKIND, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimation in linear models. *Ann. Math. Statist.* **38**, 1092–109.

[Received October 2002. *Revised August* 2003*]*