Amsterdam. University. Institute of actuarial sciences and econometrics

Report AE12/77

Non-detection of serial correlation in least squares regression; frequency and consequences

by Jan F. Kiviet

Instituut voor Actuariaat & Econometrie

Classification: AMS(MOS) subject classification scheme (1970):
62J05, 62M10, 62F05, 62F25, 62M20.

Keywords: multiple regression, serial correlation, confidence regions, hypothesis testing, prediction.

Abstract: The robustness and efficiency of OLS statistical inference is assessed in cases where the disturbances are serially correlated. In addition the frequency of accepting independence of the disturbances (and thus the acceptance of OLS results) is considered. An attainable lower bound for the probability of a type two error in testing for serial correlation is established. Case studies indicate that confidence regions for regression coefficients are very sensitive to departures from independence of the disturbances; OLS prediction appears to be more robust. When the sample is small, non-detection of serial correlation will frequently occur. These results suggest a tentative strategy for the detection of serial correlation.

Free copies are obtainable from
Bibliotheek - Instituut voor Actuariaat en Econometrie
Jodenbreestraat 23 - Amsterdam, 'phone: 020-5254287/17

Universiteit van Amsterdam

NON-DETECTION OF SERIAL CORRELATION IN LEAST SQUARES REGRESSION; FREQUENCY AND CONSEQUENCES.

Jan F. Kiviet
University of Amsterdam,
Amsterdam, The Netherlands.

1. Introduction

In this paper we consider problems arising from the application of Ordinary Least Squares (OLS) in situations where all assumptions of the classical linear regression model hold, except the zero correlation of the disturbances. More precisely, we consider the case where the OLS results are accepted, because the standard assumptions are found to be tenable on a priori grounds or after testing, while in fact the basic assumption of no serial correlation - and only this one - is violated. We will try to outline the relevance of this case and indicate the seriousness of the ensuing problems, viz. the inefficiency and unreliability of OLS statistical analysis in the presence of correlated disturbances. Especially the case of stationary first-order serially correlated errors is considered.

Cochrane and Orcutt [3] and others have shown that the error terms involved in most formulations of economic relations are highly positively autocorrelated. Therefore it is general practice in the analysis of economic time series to test the independence of the disturbances. In order to meet the case we consider here, two conditions have to be satisfied. The first is the presence of serial correlation in the disturbances, the second is a test result that does not contradict the hypothesis of no serial correlation. Together these conditions give rise to what in the statistical theory of hypothesis testing is called a type two error. In the following we shall postulate the distribution function of the disturbances and thus ensure that the first condition is fulfilled. For a specific test procedure, the probability of a type two error can then be calculated for a particular model. Different test procedures go with different probabilities of type two errors. We will show that an attainable lower bound for this probability can be calculated. Numerical results for some particular models indicate that, whatever test procedure is used, OLS results will be accepted frequently when in fact the disturbances are serially correlated.

As the case appears to be relevant, we also investigate the consequences of using the OLS technique in the presence of autocorrelated disturbances. Up to now attention has mainly been paid to the non-optimality and to the bias in the estimation of the variance of both the disturbances and the coefficient estimator in this situation. Here we go on to consider the robustness and efficiency of OLS statistical analysis, and to assess its sensitivity to departures from the assumption of independent disturbances.

In section 2 we present the relevant OLS theory, go through the unreliability of OLS-based inference in section 3, correct this inference in section 4 and indicate the inefficiency of OLS inference in section 5. The probability of non-detection of serial correlation is considered in section 6. In 7 the outcome for some econometric case studies is presented, and in section 8 we come to the conclusions.

2. *Ordinary Least Squares Qualities*

The classical linear normal regression model is

$$(1) \qquad y = X\beta + \varepsilon$$

where y is an n-vector of observations on the dependent variable, X is a non-stochastic n×k regressor matrix of full column rank, β is the coefficient vector and ε a stochastic disturbance vector, n-variate normally distributed with zero expectation and covariance $\sigma^2 I_n$.

The Ordinary Least Squares coefficient vector $b = (X'X)^{-1}X'y$ is the best linear unbiased estimator of β, its variance is $\sigma^2(X'X)^{-1}$, usually estimated by $s^2(X'X)^{-1}$, where $s^2 = \frac{e'e}{n-k}$ is an unbiased estimator of σ^2 with $e = y - Xb$ the OLS residual vector. The normality of the disturbances enables us to construct the following confidence and prediction regions and tests.

Confidence Regions and corresponding Tests

Let C be an r×k transformation matrix of rank $r \leq k$. Then the quadratic form $(Cb-C\beta)'[\sigma^2 C(X'X)^{-1}C']^{-1}(Cb-C\beta)$ has a χ^2_r distribution because the r-vector Cb has expectation $C\beta$ and $\text{cov}(Cb) = \sigma^2 C(X'X)^{-1}C'$. It is well-known that $(n-k)s^2$ is distributed as $\sigma^2\chi^2_{n-k}$ and that b and s^2 are independent. Now a Fisher ratio leads to a confidence region for $C\beta$ of the form[1)]

1) See, for instance, Wonnacott and Wonnacott [15], pp. 248-256.

$$(2) \qquad (C\beta - Cb)'[C(X'X)^{-1}C']^{-1}(C\beta - Cb) \leq rs^2 F^{\alpha}_{r,n-k} .$$

Here $F^{\alpha}_{r,n-k}$ is the upper 100α percent critical value of the Fisher distribution function with r degrees of freedom in the numerator and n-k in the denominator. The region (2) is an r-dimensional ellipsoid with center at Cb. It contains the true $C\beta$ with a probability of $(1-\alpha)\cdot 100\%$, and $\gamma = 1-\alpha$ is the confidence coefficient of the region. The region in the r-dimensional space outside the ellipsoid (2) corresponds with the critical region for the general linear constraints test $H_0: C\beta = c_0$ versus $H_1: C\beta \neq c_0$. The size of the test is α and H_0 is accepted when the ellipsoid contains the r-vector c_0. This test is a likelihood ratio test[2].

For $r = 1$, C is a row vector and the confidence ellipsoid collapses to an interval for a simple linear function of the coefficients. Taking one element of this vector as unity and all other as zero we get a confidence interval for a single coefficient, the same as that of the usual Student procedure. More generally, a joint confidence region for r coefficients simultaneously is derived as follows. Assume, without loss of generality, that a confidence ellipsoid for the r coefficients $\beta_{(r)}$, corresponding with the last r columns of X, has to be constructed. X can be partitioned as $X = \left[X_{k-r} \vdots X_r\right]$ and b as $b' = (b'_{(k-r)}, b'_{(r)})$, while the matrix C equals now $\left[0 : I_r\right]$, where 0 is an $r\times(k-r)$ matrix of zeroes. It follows from (2) that a confidence region for $\beta_{(r)}$ is given by the set

$$(3) \quad \left\{\beta_{(r)} \;\middle|\; (\beta_{(r)}-b_{(r)})'\left[X'_rX_r - X'_rX_{k-r}(X'_{k-r}X_{k-r})^{-1}X'_{k-r}X_r\right](\beta_{(r)}-b_{(r)}) \leq rs^2 F^{\alpha}_{r,n-k}\right\}.$$

When this r-dimensional ellipsoid does not contain any point with a zero coordinate, we accept the hypothesis that the r coefficients $\beta_{(r)}$ simultaneously differ from zero.

Prediction Region

Let the assumptions of the model still hold for m observations not included in the sample. Let the m-vector y_* contain those future values of the dependent variable and the $m\times k$ matrix X_* those of the explanatory variables. Then $y_* = X_*\beta + \varepsilon_*$,

---

2) See Theil [14], pp. 143-144.

with ε_* an m-variate normally distributed m-vector, uncorrelated with ε, and with covariance $\sigma^2 I_m$. Now the predictor X_*b has a prediction error $y_* - X_*b$ with zero expectation and $\text{cov}(y_*-X_*b) = \sigma^2[I_m+X_*(X'X)^{-1}X_*']$. Any other predictor that is linear in y and has zero expected prediction error has a covariance matrix of the prediction error that exceeds $\text{cov}(y_*-X_*b)$ by a positive semidefinite matrix[3]. The statistic $(y_*-X_*b)'[I_m+X_*(X'X)^{-1}X_*']^{-1}(y_*-X_*b)$ is $\sigma^2\chi^2_m$ distributed. Using the independence of b and s^2 and ε and ε_* we get the prediction ellipsoid

$$(4)\quad (y_*-X_*b)'\left[I_m+X_*(X'X)^{-1}X_*'\right]^{-1}(y_*-X_*b) \leq ms^2F^{\alpha}_{m,n-k}$$

with center at X_*b, which contains the true future value vector y_* with a probability $\gamma = 1 - \alpha$.

Non-scalar covariance matrix

When we relax the assumption about the distribution of the disturbances and postulate instead

$$(5)\quad \begin{array}{l} E(\varepsilon) = 0 \quad \text{and} \quad \text{cov}(\varepsilon) = \sigma^2\Omega, \text{ where} \\ \Omega \text{ is an } n\times n \text{ non-singular matrix,} \end{array}$$

the OLS coefficient vector, although still unbiased, will for general X-matrices no longer be best linear. This property is reserved now for the Generalized Least Squares (GLS) estimator, which, however, is only applicable when Ω is known.

Broadly speaking, the consequences of the generalization of the covariance assumption are two-fold: OLS is no longer best linear and OLS statistical inference is no longer reliable. This unreliability is caused by the change in the distribution of the OLS estimates that follows from the change in the distribution of the disturbances. For now

$$(6)\quad \begin{array}{l} \text{cov}(b) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1} \quad \text{and} \\ (n-k)Es^2 = [n-\text{tr}\{(X'X)^{-1}X'\Omega X\}]\sigma^2 . \end{array}$$

Thus s^2 is no longer unbiased and $s^2(X'X)^{-1}$ is no longer an adequate estimator of cov(b). Several textbook writers, among others Malinvaud [13] and Theil [14], investigate the bias of the OLS variance estimators s^2 and

---

3) See Theil [14], pp. 122-124.

$s^2(X'X)^{-1}$ for some typical models. In Kiviet [9] extensions of their results are presented. The bias of these variance estimators, however, is just an intermediate indication of the unreliability which ensues. The change in the distribution of b ans s^2, changes the distribution function of the statistics used in constructing confidence and prediction regions. Neither the numerator nor the denominator of the ratios have χ^2 distributions now, and moreover they are no longer independent. So the probability that the ellipsoids (2) and (4) contain $C\beta$ and y_* respectively may differ from γ. The change in confidence coefficient can be seen as the ultimate indication of the unreliability of OLS-based statistical inference.

3. The confidence coefficient of OLS regions when disturbances are correlated

The confidence and prediction ellipsoids (2) and (4) can be written as quadratic forms in the disturbances. By means of a technique introduced by Imhof [8], the cumulative distribution function of quadratic forms in normal variables can be calculated by numerical integration of an inversion formula. Thus given the distribution of the disturbances, the true confidence coefficient of the regions can be found. In the following we shall distinguish between the null-hypothesis $H_0: \varepsilon \simeq N(0,\sigma^2 I_n)$ and the alternative $H_1: \varepsilon \simeq N(0,\sigma^2\Omega)$. When prediction regions are considered, we have $H_0:(\varepsilon',\varepsilon_*')' \simeq N(0,\sigma^2 I_{n+m})$ versus $H_1:(\varepsilon',\varepsilon_*')' \simeq N(0,\sigma^2\Omega^*)$, with Ω^* an $(n+m)\times(n+m)$ non-singular matrix.

Confidence Regions and corresponding Tests

When the disturbances are not serially correlated, the confidence ellipsoid (2) for $C\beta$ has probability

$$(7)\quad P\left\{(C\beta-Cb)'[C(X'X)^{-1}C']^{-1}(C\beta-Cb) \leq rs^2F^{\alpha}_{r,n-k} \;\middle|\; H_0\right\} = \gamma.$$

The probability of the same ellipsoid under H_1 may differ from γ. Using $Cb - C\beta = C(X'X)^{-1}X'\varepsilon$ and $s^2 = e'e/(n-k) = \varepsilon'M\varepsilon/(n-k)$ with $M = I_n - X(X'X)^{-1}X'$, we can write for the true confidence coefficient $\tilde{\gamma}$:

$$(8)\quad \tilde{\gamma} = P\left\{\varepsilon'X(X'X)^{-1}C'\left[C(X'X)^{-1}C'\right]^{-1}C(X'X)^{-1}X'\varepsilon \leqq rF^{\alpha}_{r,n-k}\varepsilon'M\varepsilon/(n-k) \;\middle|\; H_1\right\}$$

$$= P\left\{\varepsilon'Z\varepsilon \leqq 0 \;\middle|\; H_1\right\}$$

$$= P\left\{\zeta'Z\zeta \leqq 0\right\} \qquad \text{with}$$

$$Z = X(X'X)^{-1}C'\left[C(X'X)^{-1}C'\right]^{-1}C(X'X)^{-1}X' - \frac{r}{n-k}F^{\alpha}_{r,n-k}M \quad \text{and}$$

$$\zeta = \frac{\varepsilon}{\sigma} \quad \text{so} \quad \zeta \simeq N(0,\Omega)\ .$$

This probability can numerically be evaluated by the method of Imhof. Let λ_i $(i=1,\ldots,n)$ be the eigenvalues of $Z\Omega$, then

$$(9)\quad \tilde{\gamma} = \tfrac{1}{2} - \frac{1}{\pi}\int_0^{\infty} \frac{\sin\left[\frac{1}{2}\sum_{i=1}^{n}\arctan(\lambda_i u)\right]}{u\prod_{i=1}^{n}(1+\lambda_i^2u^2)^{\frac{1}{4}}}\,du\ .$$

Formula (8) shows that $\tilde{\gamma}$ depends on C, X, Ω and γ only, and that the value of the true confidence coefficient of a $100\gamma\%$ OLS confidence region for $C\beta$ is independent of β itself and of σ^2. Also the value of the sample y-vector, which is indispensable to calculate the ellipsoid itself, does not influence its true significance. In Appendix A we prove that in the case of a confidence ellipsoid for r coefficients simultaneously, as given in (3), not $X = \left[X_{k-r} : X_r\right]$, but only the spaces spanned by the columns of X_r and X_{k-r} are relevant for the value of $\tilde{\gamma}$.

It is evident that $\tilde{\alpha} = 1 - \tilde{\gamma}$ will be the true significance level of the general linear constraints test on the coefficient vector, H_0: $C\beta = c_0$ versus H_1: $C\beta = c_0$, where the ellipsoid (2) is the region of acceptance.

Prediction Region

When dealing with prediction regions we have the prediction error $y_* - X_*b = \varepsilon_* - X_*(X'X)^{-1}X'\varepsilon$. We define the $m\times(n+m)$ matrix $G = \left[-X_*(X'X)^{-1}X' : I_m\right]$ and the $n+m$ disturbance vector δ, with $\delta' = (\varepsilon', \varepsilon_*')$, and get $y_* - X_*b = G\delta$. Let M^* be the $(n+m)\times(n+m)$ matrix $\begin{bmatrix} M & : & 0 \\ \cdots & & \cdots \\ 0 & : & 0 \end{bmatrix}$, so $\varepsilon'M\varepsilon = \delta'M^*\delta$. The probability of ellipsoid (4) under H_1 can then be written as

$$\tilde{\gamma}_* = P\left\{\delta'G'\left[GG'\right]^{-1}G\delta \leq mF^{\alpha}_{m,n-k}\delta'M^*\delta/(n-k) \;\middle|\; H_1\right\}, \tag{10}$$

where $\delta \simeq N(0,\sigma^2\Omega^*)$.

Here we have a quadratic form in n+m normally distributed variables whose cumulative distribution function at zero can be evaluated when X, X_*, Ω^* and γ are known.

In section 7 the formulas (8) and (10) will be used to illustrate for several X, X_* matrices to what extent OLS confidence fades away when the matrices Ω and Ω^* depart from I.

4. *Correction of OLS confidence and prediction regions when disturbances are correlated*

The robustness of OLS statistical analysis in the presence of correlated disturbances can be established in two ways. In the previous section we presented a method to register the change in confidence of OLS regions. In addition to this approach, we now construct regions with given confidence coefficient γ for $C\beta$ and y_* around Cb and X_*b, and register the changes in the regions caused by departures of Ω^* from I. The first approach reveals the reliability of the statistical inference based on standard OLS procedures. The second approach gives the inferences that should be made on the basis of the OLS coefficient estimator.

Confidence Regions

A confidence region for $C\beta$ based on OLS, given Ω, can be obtained by constructing a statistic that is a function of $C\beta$ and of OLS sample statistics. This statistic must have a distribution fuction that is independent of the parameters β and σ^2. An ellipsoidal region with center at Cb can be obtained from the statistic q, where

$$q = \frac{(C\beta-Cb)'A(C\beta-Cb)}{e'e} = \frac{\zeta'X(X'X)^{-1}C'AC(X'X)^{-1}X'\zeta}{\zeta'M\zeta} \tag{11}$$

with A an r×r matrix independent of β and σ^2, and $\zeta \simeq N(0,\Omega)$. For a critical value Q^{α}, with $P\{q \leq Q^{\alpha}\} = 1 - \alpha = \gamma$, the 100γ percent confidence region is

$$(C\beta-Cb)'A(C\beta-Cb) \leq e'eQ^{\alpha}\ , \tag{12}$$

where A determines the shape of the ellipsoid. Without loss of generality A and Q^{α} can be scaled so that trA = n. The Imhof technique enables the calculation of $P\{q \leq Q_0\}$ for given Q_0. So the root Q^{α} of the equation $P\{q \leq Q\} = \gamma$ can be found numerically.

The matrix A still has to be defined. Because of the non-optimality of OLS, a method for constructing a confidence region in terms of b and e'e that is in some sense optimal is not directly available. When r = 1, the ellipsoid reduces to an interval and A to a scalar equal to unity. The construction of a symmetric confidence interval with confidence coefficient γ for a linear combination of the regression coefficients is straight forward. Here we confine ourselves to the interval for a single coefficient.

Let β_j be the j-th element of β, $\left[(X'X)^{-1}\right]_{jj}$ the j-th diagonal element of $(X'X)^{-1}$ and Q_j^{α} the critical value such that

$$(13)\quad P\{q \leq Q_j^{\alpha}\} = P\left\{\frac{\zeta'X(X'X)^{-1}C'C(X'X)^{-1}X'\zeta}{\zeta'M\zeta} \leq Q_j^{\alpha}\right\} = 1 - \alpha = \gamma$$

where $C'C$ now is a $k \times k$ matrix with its j-th diagonal element equal to unity and all others zero. A 100γ percent confidence interval for β_j is given by $|\beta_j - b_j| \leq \sqrt{e'eQ_j^{\alpha}}$, where b_j is the j-th element of b, j=1,...,k. In comparison with the corresponding $100\tilde{\gamma}$ percent OLS confidence interval, where $|\beta_j - b_j| \leq \sqrt{s^2|(X'X)^{-1}|_{jj}F^{\alpha}_{1,n-k}}$, the length of this interval has changed by a factor

$$(14)\quad \phi_j = \sqrt{(n-k)Q_j^{\alpha}/\left[(X'X)^{-1}\right]_{jj}F^{\alpha}_{1,n-k}} \quad .$$

This multiplicative correction factor can be applied to single coefficient confidence intervals to allow for a covariance matrix $\Omega \neq I$. Apart from γ and j it depends on X and Ω.

Prediction Regions

For y_*, too, an ellipsoid with confidence coefficient γ centered at its OLS estimate can be constructed. The statistic

$$(15)\quad \frac{(y_* - X_* b)'B(y_* - X_* b)}{e'e} = \frac{\delta'G'BG\delta}{\delta'M^*\delta}$$

is distributed independently of β and σ^2, with B an $m \times m$ matrix and δ, G and M^* as in (10). We will just look at the case m=1, where B may be omitted from (15).

The critical value Q_*^α can numerically be found so that

$$P\left\{(y_*-X_*b)'(y_*-X_*b) \leq e'eQ_*^\alpha\right\} = P\left\{\frac{\delta' G' G\delta}{\delta' M_* \delta} \leq Q_*^\alpha\right\} = 1 - \alpha \ .$$

Then a 100γ percent prediction interval for y_* is given by $|X_*b-y_*| \leq \sqrt{e'eQ_*^\alpha}$ and the factor

$$(16) \quad \phi_* = \sqrt{(n-k)Q_*^\alpha / [1+X_*(X'X)^{-1}X_*']F_{1,n-k}^\alpha}$$

indicates the change in length of the prediction interval when Ω^* differs from I.

In section 7 the factors (14) and (16) will be calculated for several X, X_* matrices and specific matrices Ω and Ω^*.

5. *The inefficiency of OLS inferences*

After indicating the unreliability of OLS inferences, we now concentrate on the other consequence of the generalization of the covariance assumption, viz. the inefficiency caused by the fact that OLS is no longer best linear.

A great number of techniques have been devised to improve on OLS estimates, when the scalar covariance assumption is untenable on a priori grounds or after testing. We will not consider these technique here, but just try to indicate the maximum improvement with regard to OLS that can be obtained. We do this by comparing OLS and GLS inference on single coefficients and prediction of one value. The inefficiency of OLS relative to GLS will be expressed in the (relative) length of confidence and prediction intervals. So apart from the covariance of the coefficient estimator, the estimation of σ^2 and the distribution of the relevant statistics are also taken into account.

The GLS coefficient vector $\hat{b} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ has variance $\sigma^2(X'\Omega^{-1}X)^{-1}$, estimated by $\hat{s}^2(X'\Omega^{-1}X)^{-1}$, where $\hat{s}^2 = \hat{e}'\Omega^{-1}\hat{e}/(n-k)$ is an unbiased estimator of σ^2, with $\hat{e} = y - Xb$ the GLS residual vector.

Confidence Regions

The GLS confidence ellipsoid for $C\beta$ is given by

$$(C\beta-C\hat{b})'[C(X'\Omega^{-1}X)^{-1}C']^{-1}(C\beta-C\hat{b}) \leq r\hat{s}^2F_{r,n-k}^\alpha \ .$$

The $(1-\alpha)\cdot 100\%$ confidence interval for the j-th coefficient is then expressed by

$$(17)\quad (\beta_j-\hat{b}_j)^2 \leq \hat{s}^2\left[(X'\Omega^{-1}X)^{-1}\right]_{jj}F^{\alpha}_{1,n-k}\ .$$

The length of this interval, and of the corresponding OLS interval

$$(18)\quad (\beta_j-b_j)^2 \leq s^2\left[(X'X)^{-1}\right]_{jj}F^{\alpha}_{1,n-k}\ ,$$

can only be compared when $\hat{s}^2$ and s^2 are calculated. So apart from X and Ω, the sample y-vector is also needed. In order to eliminate the disturbing influence of the accidental value of this stochastic vector, we take expectations. We now define the factor $\tilde{\psi}_j$ as the square root of the ratio of the expectation of the right sides of (18) and (17). Recalling (6), this leads to

$$(19)\quad \tilde{\psi}_j = \sqrt{\frac{(n-\mathrm{tr}\{(X'X)^{-1}X'\Omega X\})\left[(X'X)^{-1}\right]_{jj}}{(n-k)\left[(X'\Omega^{-1}X)^{-1}\right]_{jj}}}\ .$$

This factor indicates the ratio of the expected lengths of OLS and GLS confidence intervals respectively. But as we have earlier shown, the confidence coefficient of the OLS interval is erroneous, and should be corrected by the factor (14). If we allow for this correction factor, we obtain

$$(20)\qquad \psi_j = \phi_j\tilde{\psi}_j\ .$$

This factor now indicates how much a confidence interval for β_j with confidence coefficient $1-\alpha$, based on the OLS statistics b and s^2, is expected to be wider than the GLS interval with the same confidence coefficient. It just depends on X, Ω and, of course, on α and j.

Prediction Region

The vector $y_* = X_*\beta + \varepsilon_*$ has to be predicted, where the combined disturbance vector $\delta' = (\varepsilon' , \varepsilon_*')$ is supposed to have covariance matrix

$$E\delta\delta' = \sigma^2\Omega^* = \sigma^2\begin{bmatrix}\Omega & \vdots & \Psi\\ \cdots & & \cdots\\ \Psi' & \vdots & \Omega_*\end{bmatrix}\ .$$

The GLS predictor [4] $X_*\hat{b} + \Psi'\Omega^{-1}\hat{e}$ has a prediction error with zero expectation

[4] See Theil [14], pp. 288-289.

and covariance matrix $\sigma^2\Theta$, where

$$(21)\quad \Theta = X_*(X'\Omega^{-1}X)^{-1}X_*' + \Omega_* - \Psi'[\Omega^{-1}-\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}]\Psi$$
$$- X_*(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Psi - \Psi'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X_*' .$$

Any other predictor that is linear in y and has zero expected prediction error has a prediction error covariance matrix that exceeds $\sigma^2\Theta$ by a positive semidefinite matrix.
We can construct the prediction ellipsoid

$$(y_*-X_*\hat{b}-\Psi'\Omega^{-1}\hat{e})'\Theta^{-1}(y_*-X_*\hat{b}-\Psi'\Omega^{-1}\hat{e}) \leq m\hat{s}^2F^{\alpha}_{m,n-k} .$$

In the case where m=1, X_* and Ψ' are row-vectors and y_* and Θ are scalars. Then the GLS (1-α).100% prediction interval is expressed by

$$(y_*-X_*\hat{b}-\Psi'\Omega^{-1}\hat{e})^2 \leq \hat{s}^2\Theta F^{\alpha}_{1,n-k} .$$

Now the factor

$$(22)\quad \tilde{\psi}_* = \sqrt{\frac{(n-\mathrm{tr}\{(X'X)^{-1}X'\Omega X\})[1+X_*(X'X)^{-1}X_*']}{(n-k)\,\Theta}}$$

relates the expected length of the corresponding $100\tilde{\gamma}_*$% OLS interval (10) to this GLS interval. When the difference in confidence coefficient of these intervals is again taken into account, the factor

$$(23)\qquad \psi_* = \phi_*\tilde{\psi}_*$$

indicates the inefficiency of OLS in terms of GLS prediction. This factor depends on X, X_*, Ω^* and α.

In section 7 the factors ψ_j and ψ_* will be calculated for some empirical models. By definition they exceed or equal unity.

6. The probability of non-detection of serial correlation

The calculation of the actual confidence of the OLS ellipsoids and of the necessary change in length of the OLS intervals is proper only when it happens that OLS is erroneously applied in the presence of serially correlated disturbances. The probability that the hypothesis H_0: $\varepsilon \approx N(0,\sigma^2 I_n)$ and the OLS inferences are accepted, while the covariance matrix is not scalar, depends of course on the particular test procedure that is used. In a test of

H_0, the probability that serial correlation is not detected, that is the probability of committing a type two error, equals one minus the power. The best way to guard against the erroneous application of OLS is therefore to use a powerful test.

In Appendix B, we prove that the ratio of $\hat{s}^2$ and s^2, that is the test-statistic

$$(24)\quad t(\Omega) = \frac{y'\left[\Omega^{-1}-\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\right]y}{y'\left[I-X(X'X)^{-1}X'\right]y} = \frac{\hat{e}'\Omega^{-1}\hat{e}}{e'e}\ ,$$

has the attractive property of providing the most powerful invariant test of the hypothesis H_0: $\varepsilon \simeq N(0,\sigma^2 I_n)$ against the simple alternative H_1: $\varepsilon \simeq N(0,\sigma^2\Omega)$. In practice, however, we usually cannot specify a specific simple alternative hypothesis, but a whole class of alternatives has to be covered by H_1. The class of first-order autoregressive processes has attracted much attention in this context. We too adopt the assumption that the disturbances are generated by the scheme

$$(25)\quad \begin{aligned} &\varepsilon_i = \rho\varepsilon_{i-1} + \eta_i \qquad ; i=2,\ldots,n \\ &\text{with} \qquad \varepsilon_1 \simeq N(0,\sigma^2) \\ &\qquad\quad \eta_i \simeq N(0,\sigma_\eta^2) \qquad ; i=2,\ldots,n \qquad \text{where} \\ &\qquad\quad \eta_i \text{ and } \eta_j \text{ are independent for } i \neq j. \end{aligned}$$

One can distinguish the stationary first-order autoregressive process, where

$$|\rho| < 1 \quad \text{and} \quad \sigma_\eta^2 = (1-\rho^2)\sigma^2$$

and the non-stationary first-order autoregressive process where

$$\sigma_\eta^2 = \sigma^2 \quad \text{and} \quad \rho \text{ arbitrary}\ .$$

Now H_1: $\rho \neq 0$ or the one-sided hypothesis H_1: $\rho > 0$ and H_1: $\rho < 0$ are useful composite alternatives.

In the following the non-stationary type of first-order autocorrelation will be left out of consideration as are higher order and moving average types. Attention is only paid to matrices Ω of the type

$$(26)\quad \Omega = (\omega_{ij})\ , \text{ with } \omega_{ij} = \rho^{|i-j|} \text{ and } |\rho| < 1\ .$$

Against this alternative, the exact Durbin and Watson test[5] is locally most powerful invariant for $\rho \to 0$. The Berenblut and Webb test[6], although developed for testing autocorrelation of the non-stationary type, is locally most powerful invariant for $\rho \to 1$, if an overall mean is included in the regression. A test with uniformly most power against a composite alternative for general X matrices does not exist, so no test is preferable in all circumstances.

We now consider the qualities of a test of the type given in (24). Let V be an n×n matrix with typical element v_{ij}, where

$$V = (v_{ij}), \quad \text{with } v_{ij} = \rho_1^{|i-j|} \quad \text{and } |\rho_1| < 1 .$$

It is evident that the test-statistic t(V) is locally most powerful in the neighbourhood of $\rho = \rho_1$ when H_0: $\rho = 0$ is tested against the composite alternative H_1: $\rho \neq 0$. So, for every $\rho_1 \in (-1,+1)$, the maximum attainable power can be calculated in the same way as Koerts ans Abrahamse [11] calculated the power of the exact Durbin and Watson test. First the critical value at significance level α has to be found. This significance point $T^{\alpha}(V)$ is the root of the equation

$$P\left\{\frac{\zeta'\left[V^{-1}-V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}\right]\zeta}{\zeta'\left[I-X(X'X)^{-1}X'\right]\zeta} < T \;\middle|\; \zeta \simeq N(0,I)\right\} = \alpha .$$

Now the lower bound for the probability of an error of the second kind at ρ_1 at significance level α for some regression matrix X is given by

$$(27) \quad P\left\{\frac{\zeta'\left[V^{-1}-V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}\right]\zeta}{\zeta'\left[I-X(X'X)^{-1}X'\right]\zeta} > T^{\alpha}(V) \;\middle|\; \zeta \simeq N(0,V)\right\} .$$

In the neighbourhood of $\rho_1 = 0$, this lower bound is approximated by the Durbin and Watson test, and for $\rho \to 1$ by Berenblut and Webb's. It is easy to prove that the lower bound and the powers of the Durbin and Watson and the Berenblut and Webb test do not depend on X itself, but only on the column space of the X matrix.

---

5) See Durbin and Watson [6], pp. 10.
N.B. Their formula (8), which is to be compared with formula (B.2) of our Appendix B, is inaccurate. However, their theorems can be proved after some slight corrections.

6) See Berenblut and Webb [1], pp. 39-40.

In order to indicate the frequency that OLS inference is erroneously accepted we will calculate the lower bound (27) in section 7 for several X matrices at some values of ρ. Because these values do not result from a single test statistic, they are in practice only attainable when a series of tests is performed. We therefore also calculate the probability of type two errors for the exact Durbin and Watson test. But even then we get a too favourable picture of what is done in practice. Often just the Durbin and Watson bounds test is applied. The actual power and significance then depend on the treatment of the inconclusive region.

7. Numerical Results

The frequency of non-detection of serial correlation and the indicators of its consequences all depend on the X matrix, or on its (sub)spaces(s), and are independent of β, of the sample y-vector and of σ^2. Therefore, to illustrate the features we have considered, we must select some design matrices. The illustrations here will be limited to five data sets that enjoy prodigious popularity in the research on serial correlation. They all include a constant term vector. One consists of four (k=5), the others of two (k=3) exogenous variable vectors. From these five sets we took a subset of 15 successive annual observations (n=15), and from two a subset of 30 successive annual observations (n=30). Thus, seven X matrices are considered. As for prediction, only one period ahead (m=1) is analysed. Here we list the origins of the data sets, their names, the order of the subsets considered and, between brackets, the date of the first observation in the subset(s).

P: Pears data 1925-1940: Henshaw [7], Table 1.
P_2 is supply of California pears,
P_3 is supply of other pears,
P_4 is an index of U.S. nonagricultural income and
P_5 is a trend term.
Subset (1925): 15 × 5.

T: Textile data 1923-1939: Theil [14], Table 3.1.
T_2 is log real income per head and
T_3 is log relative price for the Netherlands.
Subset (1923): 15 × 3.

C: Consumption data 1921-1941: Klein [10], p. 135.

C_2 are profits and

C_3 are wages for the United States.

Subset (1923): 15 × 3.

A: Automobile data 1921-1953: Chow [2], Table 1.

A_2 is log automobile stock per capita and

A_3 log personal money stock per capita for the U.S.

Subsets (1921): 15 × 3 and 30 × 3.

S: Spirits data 1870-1938: Durbin and Watson [5], Table 1.

S_2 is log real income per capita and

S_3 is log relative price of spirits for the U.K.

Subsets (1870): 15 × 3 and 30 × 3.

As already stated, we assume that the disturbances are normally distributed according to a stationary first-order autoregressive process. The conclusions that can be drawn from the figures that follow are limited to the seven adopted X matrices. As we have proved, some of the probabilities do not depend on X itself but only on its column space or on the spaces of partitions of the X matrix. Further, as to the shape of the explanatory variables and to the sample size, we think that these seven X matrices are typical of a great number of design matrices in econometric analysis.

Frequency of non- detection of serial correlation

Table 1 presents the probability of a type two error when testing H_0: $\rho = 0$, evaluated along the lines of section 6. The table shows the lower bounds (27) for arbitrary tests against composite alternatives as well as the results for the exact Durbin and Watson test against specific alternatives; for $\rho < 0$ the alternative hypothesis was H_1: $\rho < 0$, and for $\rho > 0$ we took H_1: $\rho > 0$. All numerical values refer to the significance level $\alpha = 0.05$.

We note the following characteristics:

- The figures suggest that as the number of degrees of freedom increases, probabilities of non-detection of serial correlation decrease, and the power of the exact Durbin and Watson test tends to optimality.
- Acceptable test results are obtained when there are 27 degrees of freedom and $\rho \geq .7$.

Table 1. Lower bound for the probability of accepting H_0: $\rho = 0$ at significance level $\alpha = 0.05$. Between brackets the corresponding probabilities for the exact Durbin and Watson test.

ρ	data set						
	n=15;k=5	n=15;k=3				n=30;k=3	
	P	T	C	A	S	A	S
-.5	.64(.68)	.54(.58)	.53(.57)	.53(.56)	.53(.56)	.19(.21)	.19(.21)
-.3	.83(.83)	.79(.80)	.78(.79)	.78(.79)	.78(.79)	.57(.58)	.57(.58)
.3	.84(.85)	.81(.81)	.79(.80)	.80(.81)	.79(.80)	.59(.59)	.59(.60)
.5	.72(.73)	.67(.67)	.62(.62)	.64(.65)	.62(.63)	.25(.25)	.26(.26)
.7	.58(.59)	.53(.54)	.43(.45)	.47(.51)	.45(.46)	.07(.07)	.08(.08)
.9	.46(.48)	.42(.46)	.27(.31)	.31(.43)	.31(.35)	.02(.02)	.02(.02)

- The results for the case $n-k \leq 12$ show that non-detection of serial correlation may be omnipresent. Even when autocorrelation is considerable (ρ=.7), non-detection will occur in one case out of two.

Dubbelman, Abrahamse and Louter [4] used the same matrices of order 15 × 3 as we do, in comparing the qualities of different autocorrelation tests. They regard the power of the exact Durbin and Watson test as the maximum attainable, although they note that this is only legitimate in the neighbourhood of $\rho = 0$. Table 1 shows the deficiency of their supposition. In view of the meagre optimal power values for X matrices of order 15 × 3, the justification of a test procedure with less power than Durbin and Watson's has, computational costs apart, to be based on an appraisal of the consequences of non-detection of serial correlation. So we reserve our judgment concerning the choice of a test procedure and first establish these consequences.

Confidence of joint confidence regions

The true confidence coefficients of 95% OLS confidence ellipsoids for all k coefficients simultaneously,and for the k-1 coefficients relating to the exogenous variables, are given in Table 2. These values correspond to equation (9) of section 3. Inspection of these $\tilde{\gamma}$ values reveals:

- In all cases considered, positive serial correlation causes OLS regions to be too small, negative autocorrelation leads to overpessimistic inferences.

Table 2. Numerical value of the true confidence coefficient $\tilde{\gamma}$ of the OLS confidence region for all coefficients simultaneously at $\gamma = 0.95$. Between brackets the value for the same region, the constant term excluded.

ρ	data set						
	n=15;k=5	n=15;k=3				n=30;k=3	
	P	T	C	A	S	A	S
-.5	.97(.96)	.998(.997)	.997(.995)	.996(.993)	.997(.995)	.999(.998)	.999(.998)
-.3	.97(.97)	.99(.99)	.99(.99)	.99(.98)	.99(.99)	.99(.99)	.99(.99)
.3	.90(.93)	.80(.83)	.84(.88)	.83(.87)	.82(.86)	.80(.84)	.79(.83)
.5	.77(.86)	.59(.69)	.69(.80)	.66(.76)	.66(.76)	.62(.72)	.59(.69)
.7	.53(.74)	.33(.51)	.48(.71)	.41(.61)	.42(.63)	.38(.57)	.33(.50)
.9	.21(.56)	.10(.32)	.22(.60)	.14(.41)	.16(.47)	.13(.37)	.08(.25)

- An increasing number of degrees of freedom doesnot seem to push the true confidence $\tilde{\gamma}$ towards γ at all.
- We notice the following unfortunate connexion in the four 15 × 3 cases: The smaller the probability to detect serial correlation, the more unreliable OLS inference is.
- At $\rho = .3$ the reliability of these joint confidence regions is seriously affected. To confide in these OLS inferences at higher values of ρ appears to be absurd. At $\rho = .9$ general linear constraint tests on the coefficient vector involve probabilities of type one errors of 80%.

Reliability and inefficiency of single coefficient intervals

Table 3A illustrates the consequences of serial correlation on OLS single coefficient confidence intervals for the exogenous variables of the four 15×3-X matrices. For each coefficient we present $\tilde{\gamma}$ of equation (9) section 3, ϕ of equation (14) section 4 and ψ of equation (20) section 5. We notice:

- $\tilde{\gamma}$ exceeds γ for $\rho < 0$ and $\tilde{\gamma} < \gamma$ for $\rho > 0$ in all cases.
- The deviation of $\tilde{\gamma}$ from γ is less than for the two and three dimensional ellipsoids of Table 2.
- Because of the $\tilde{\gamma}$ values, the ϕ values exceed unity when ρ is positive. The necessary enlargements of the intervals are far from negligible.
- The ψ values show that the difference in efficiency between OLS and GLS is moderate for $|\rho| \leq .5$.

Table 3A. Numerical value of the indicators $\tilde{\gamma}$ (true confidence coefficient of OLS region), ϕ (correction factor to uphold $100\gamma\%$ confidence) and ψ (inefficiency of OLS inference compared with GLS inference) for single coefficient confidence intervals for the exogenous explanatory variables of the data subsets of order 15×3, at $\gamma = 1 - \alpha = .95$.

indicator	exogenous explanatory variable							
ρ	T_2	T_3	C_2	C_3	A_2	A_3	S_2	S_3
$\tilde{\gamma}$								
-.5	.995	.995	.996	.99	.99	.99	.99	.99
-.3	.99	.99	.99	.98	.98	.98	.98	.98
.3	.87	.86	.87	.92	.88	.90	.90	.89
.5	.79	.75	.79	.89	.80	.83	.86	.84
.7	.71	.60	.71	.85	.71	.71	.82	.77
.9	.64	.41	.62	.77	.64	.51	.76	.71
ϕ								
-.5	.63	.62	.61	.73	.64	.77	.68	.66
-.3	.76	.75	.75	.84	.77	.85	.80	.79
.3	1.34	1.38	1.34	1.16	1.31	1.23	1.23	1.26
.5	1.64	1.79	1.62	1.27	1.60	1.46	1.39	1.46
.7	1.97	2.46	1.94	1.39	1.93	1.79	1.55	1.67
.9	2.21	3.59	2.15	1.54	2.21	2.23	1.81	1.86
ψ								
-.5	1.09	1.07	1.07	1.10	1.08	1.17	1.09	1.09
-.3	1.03	1.02	1.02	1.03	1.03	1.06	1.03	1.03
.3	1.02	1.02	1.03	1.04	1.02	1.07	1.03	1.03
.5	1.07	1.08	1.10	1.13	1.09	1.20	1.07	1.11
.7	1.16	1.20	1.28	1.31	1.18	1.47	1.15	1.25
.9	1.26	1.47	1.60	1.68	1.34	1.95	1.30	1.47

In Table 3B we see that the Pears data are distinct from the others because:

- P_3 shows $\tilde{\gamma}$ values with $\tilde{\gamma} > \gamma$ for $\rho > 0$ and the inference towards P_2 is hardly affected by autocorrelation.
- P_1 (the constant) and P_5 (the trend) are the most sensitive to autocorrelation.

Table 3B. Numerical value of the indicators $\tilde{\gamma}$, ϕ and ψ for single coefficient confidence intervals for the five explanatory variables of the Pears data, at $\gamma = 1 - \alpha = 0.95$.

indicator / ρ	P_1	P_2	P_3	P_4	P_5
$\tilde{\gamma}$					
-.5	.99	.96	.89	.91	.99
-.3	.98	.96	.92	.94	.98
.3	.89	.95	.97	.94	.89
.5	.84	.95	.98	.91	.81
.7	.77	.95	.98	.88	.69
.9	.68	.95	.99	.85	.51
ϕ					
-.5	.75	.96	1.29	1.20	.71
-.3	.83	.97	1.16	1.07	.81
.3	1.26	1.02	.89	1.07	1.29
.5	1.48	1.02	.83	1.19	1.60
.7	1.72	1.01	.78	1.32	2.12
.9	2.08	.98	.74	1.40	3.22
ψ					
-.5	1.05	1.20	1.05	1.08	1.10
-.3	1.02	1.06	1.02	1.02	1.03
.3	1.02	1.05	1 03	1.02	1.02
.5	1.06	1.13	1.07	1.06	1.05
.7	1.14	1.24	1.13	1.14	1.09
.9	1.22	1.34	1.21	1.24	1.14

The reason for this will be the different shape of the variables P_2 and P_3. In contrast with the other variables, which are annually slowly changing, these variables, especially P_3, show waves of high frequency.

Table 5 presents outcomes for the two data sets with a sample size of 30. Only $\rho = .5$, where non-detection of serial correlation occurs in one case out of four, is considered. The left part of the table shows that a more extensive sample is no guarantee for more reliable inferences on the coefficients.

Reliability and inefficiency of prediction intervals

Table 4 reveals that prediction is much more robust against autocorrelation, than the inferences concerning confidence regions are. We calculated $\tilde{\gamma}_*$ of equation (10) section 3, ϕ_* of equation (16) section 4 and ψ_* of equation (23) section 5. Only at $\rho = .9$ two $\tilde{\gamma}_*$ values depart from γ rather seriously. Although the necessary corrections of the intervals are limited, they cannot be neglected and so are the losses in efficiency.

Table 4. Numerical value of the indicators $\tilde{\gamma}_*$, ϕ_* and ψ_* for prediction one period ahead for the data subsets of 15 observations, at $\gamma = 1 - \alpha = 0.95$.

indicator / ρ	data subset P	T	C	A	S
$\tilde{\gamma}_*$					
-.5	.95	.96	.95	.95	.96
-.3	.95	.96	.95	.95	.96
.3	.94	.93	.95	.94	.93
.5	.93	.92	.94	.94	.91
.7	.92	.89	.93	.92	.88
.9	.91	.85	.91	.91	.84
ϕ_*					
-.5	1.00	.93	.98	.99	.94
-.3	.99	.95	.99	.99	.95
.3	1.05	1.07	1.02	1.03	1.08
.5	1.09	1.14	1.05	1.07	1.17
.7	1.13	1.23	1.10	1.12	1.27
.9	1.16	1.34	1.15	1.17	1.37
ψ_*					
-.5	1.16	1.14	1.15	1.17	1.17
-.3	1.05	1.04	1.05	1.05	1.05
.3	1.05	1.04	1.04	1.05	1.05
.5	1.11	1.11	1.11	1.13	1.15
.7	1.21	1.22	1.25	1.25	1.32
.9	1.32	1.37	1.49	1.41	1.57

Rather satisfactory prediction results are obtained at sample size $n = 30$ and $\rho = .5$, as the right part of Table 5 shows.

Table 5. Numerical value of the indicators $\tilde{\gamma}$, ϕ and ψ for single coefficient confidence intervals and of $\tilde{\gamma}_*$, ϕ_* and ψ_* for prediction one period ahead for the data subsets of 30 observations at $\rho = .5$ with $\gamma = 1 - \alpha = 0.95$.

indicator	exogenous explanatory variable				indicator	data set	
	A_2	A_3	S_2	S_3		A	S
$\tilde{\gamma}$	.84	.77	.82	.81	$\tilde{\gamma}_*$	.94	.94
ϕ	1.41	1.68	1.50	1.52	ϕ_*	1.03	1.05
ψ	1.07	1.07	1.10	1.10	ψ_*	1.14	1.14

8. Conclusions

Our results show that the statistical analysis of the classical linear regression model can seriously be damaged in the case of unknown serial correlation of the disturbances. OLS prediction appears to be less vulnerable than significance tests of the coefficients. Unfortunately in those cases (given n and k) where OLS inference is most affected by serial correlation, this serial correlation is least detected. For prevention we must turn to the strategy pursued in testing for serial correlation. In view of the maximum attainable power of autocorrelation tests when the sample is small (say n-k<25) and the seriousness of the consequences of non-detected serial correlation, it seems good practice to use a much higher level of significance α than the customary 5%. At what level α should be set depends on the statistical technique that will be applied when autocorrelation is detected. As Ω is unknown, this technique will be haunted by the same problems as OLS: Non-optimality relative to GLS and inaccurate or vague statistical inference. It is not quite clear to what extent OLS can be surpassed. Evidently the possibilities of improvement are best (and most needed) when there is high positive serial correlation. That means that we have to support the view of Berenblut and Webb that their test is preferable to Durbin and Watson's.

When one need not spare computational effort, one can use the following procedure when sample size is small. Compute at some likely value(s) of Ω (not necessary of the first-order autocorrelation type) the locally most powerful invariant test statistic. Then calculate the probability that this test statistic does not exceed this value under the null hypothesis.

Then we have the maximum significance level at which the null hypothesis will be rejected. We suppose one is obliged to a cautious attitude towards OLS inference when this critical level is below 15-25%.

References

[1] Berenblut, I.I. and G.I. Webb (1973). "A New Test for Autocorrelated Errors in the Linear Regression Model", Journal of the Royal Statistical Society B 35, 33-50.

[2] Chow, G.C. (1957). "Demand for Automobiles in the United States". North-Holland, Amsterdam.

[3] Cochrane, D. and G.H. Orcutt (1949). " Applications of Least Squares Regressions to relations containing autocorrelated error terms". J. Amer. Statist. Ass., 44, 32-61.

[4] Dubbelman, C., A.P.J. Abrahamse and A.S. Louter (1976). "On typical characteristics of economic time series and the relative qualities of five autocorrelation tests". Erasmus University Rotterdam, Econometric Institute, Report 7604/S.

[5] Durbin, J. and G.B. Watson (1951). "Testing for serial correlation in least squares regression, II". Biometrika, 38, 159-178.

[6] - - - (1971). "Testing for serial correlation in least squares regression, III". Biometrika, 58, 1-19.

[7] Henshaw, R.C. (1966). "Testing single-equation least squares regression models for autocorrelated disturbances". Econometrica, 34, 646-660.

[8] Imhof, P.J. (1961). "Computing the distribution of quadratic forms in normal variables". Biometrika, 48, 419-426.

[9] Kiviet, J.F. (1976). "The bias of Ordinary Least Squares variance estimators when the disturbances follow a stationary first-order autoregressive scheme". University of Amsterdam, Instituut voor Actuariaat en Econometrie, Report AE10/76.

[10] Klein, L.R. (1950). "Economic Fluctuations in the United States 1921-1941". Wiley, New York.

[11] Koerts, J. and A.P.J. Abrahamse (1969). "On the theory and application of the general linear model". Rotterdam University Press, Rotterdam.

[12] Lehmann, E.L. (1959). "Testing Statistical Hypothesis". Wiley, New York.

[13] Malinvaud, E. (1970). "Statistical Methods of Econometrics", second revised edition. North-Holland, Amsterdam.

[14] Theil, H. (1971). "Principles of Econometrics". North-Holland, Amsterdam.

[15] Wonnacott, R.J. and T.H. Wonnacott (1970). "Econometrics". Wiley, New York.

Appendix A

Theorem

Consider the confidence ellipsoid (3) for r-coefficients simultaneously, where these coefficients relate with the last r columns of $X = [X_{k-r} : X_r]$. Then for $\tilde{\gamma}$, as given in (8) with $C = [0 : I_r]$, not X itself, but just the spaces spanned by the columns of X_r and X_{k-r} are relevant.

Proof

Let R be an n×k matrix so that $R'R = I_k$ and $X = RL$, with L a non-singular k×k matrix that gives the location of X in the space spanned by the orthogonal columns of R. R can be partitioned in $[R_{k-r} : R_r]$ and L can be choosen so that $L = \begin{bmatrix} L_{11} & : & L_{12} \\ \cdot\cdot & & \cdot\cdot \\ L_{21} & : & L_{22} \end{bmatrix}$ with L_{21} an $r \times (k-r)$ matrix of zeroes. Thus $X_{k-r} = R_{k-r}L_{11}$ and $X_r = R_{k-r}L_{12} + R_rL_{22}$. Now we substitute this decomposition of X in the matrix Z of formula (8) with $C = [0 : I_r]$. We have $M = I - X(X'X)^{-1}X' = I - RR'$ $= I - R_{k-r}R'_{k-r} - R_rR'_r$. Partitioned inversion of $X'X$ gives:

$$[C(X'X)^{-1}C']^{-1} = X'_r[I-X_{k-r}(X'_{k-r}X_{k-r})^{-1}X'_{k-r}]X_r = L'_{22}L_{22}$$

$$C(X'X)^{-1}X' = (L'_{22}L_{22})^{-1}X'_r[I-X_{k-r}(X'_{k-r}X_{k-r})^{-1}X'_{k-r}] = L_{22}^{-1}R'_r .$$

So Z equals $R_rR'_r - \frac{r}{n-k}F^{\alpha}_{r,n-k}[I-R_{k-r}R'_{k-r}-R_rR'_r]$ and $\tilde{\gamma}$ appears to be invariant for the location of X_r and X_{k-r} in the spaces spanned by the columns of R_k and R_{k-r} respectively.

Appendix B

Theorem

The test-statistic $\hat{e}'\Omega^{-1}\hat{e}/e'e$ performs the most powerful invariant test of the hypothesis H_0: $y \simeq N(X\beta,\sigma^2I_n)$ against the simple alternative H_1: $y \simeq N(X\beta,\sigma^2\Omega)$.[7]

---

7) A closely similar theorem is indicated in Berenblut and Webb [1], Appendix C.

Proof

The problem is invariant under transformations of the form $L(y) = \ell_0 y + X\ell$, with finite $\ell_0 \neq 0$ and ℓ a vector of k finite elements. The density function associated with H_1 is

$$f(y;X\beta,\sigma^2\Omega) = (2\pi\sigma^2)^{-\frac{n}{2}} \mid \Omega \mid^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(y-X\beta)'\Omega^{-1}(y-X\beta)/\sigma^2\} .$$

Now according to Lehmann [12], the most powerful invariant test is given by the rejection region

$$\text{(B.1)} \qquad \frac{\int_0^\infty \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty \sigma^{-1} f(y;X\beta,\sigma^2\Omega)d\beta_1 \cdots d\beta_k d\sigma}{\int_0^\infty \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty \sigma^{-1} f(y;X\beta,\sigma^2 I_n)d\beta_1 \cdots d\beta_k d\sigma} > T_0 .$$

Like Durbin and Watson[8] we make use of the substitution $y - X\beta = \varepsilon = \hat{e} - X(\beta - \hat{b})$ giving $(y-X\beta)'\Omega^{-1}(y-X\beta) = \hat{e}'\Omega^{-1}\hat{e} + (\beta-\hat{b})'X'\Omega^{-1}X(\beta-\hat{b})$. As

$$\int_{-\infty}^\infty \cdots \int_{-\infty}^\infty \exp\{-\tfrac{1}{2}(\beta-\hat{b})'X'\Omega^{-1}X(\beta-\hat{b})/\sigma^2\}d\beta_1 \cdots d\beta_k$$

$$= (2\pi\sigma^2)^{k/2} \mid X'\Omega^{-1}X \mid^{\frac{1}{2}} ,$$

the numerator of (B.1) is equal to

$$(2\pi)^{-\frac{1}{2}(n-k)} \mid \Omega \mid^{-\frac{1}{2}} \mid X'\Omega^{-1}X \mid^{\frac{1}{2}} \int_0^\infty \sigma^{-n+k-1} \exp\{-\tfrac{1}{2}\hat{e}'\Omega^{-1}\hat{e}/\sigma^2\}d\sigma$$

and after transformation $w = \frac{1}{2}\hat{e}'\Omega^{-1}\hat{e}/\sigma^2$ we get

$$(2\pi)^{-\frac{1}{2}(n-k)} \mid \Omega \mid^{-\frac{1}{2}} \mid X'\Omega^{-1}X \mid^{\frac{1}{2}} (\hat{e}'\Omega^{-1}\hat{e})^{-\frac{1}{2}(n-k)} \int_0^\infty (2w)^{\frac{n-k-2}{2}} \exp(-w)dw .$$

The denominator of (B.1) is found when Ω is replaced by I and $\hat{e}$ by e. The substitution of e is overlooked by Durbin and Watson. Now the rejection region takes the form

$$\frac{|\Omega|^{-\frac{1}{2}}|X'\Omega^{-1}X|^{\frac{1}{2}}(\hat{e}'\Omega^{-1}\hat{e})^{-\frac{1}{2}(n-k)}}{|X'X|^{\frac{1}{2}}(e'e)^{-\frac{1}{2}(n-k)}} > T_0$$

or

$$\text{(B.2)} \qquad \frac{\hat{e}'\Omega^{-1}\hat{e}}{e'e} < T_1 ,$$

which proves the theorem.

---

8) See Durbin and Watson [6], pp. 10, where the same theorem for a less general matrix Ω is considered.