



ELSEVIER

Journal of Econometrics 92 (1999) 295–323

JOURNAL OF
Econometrics

www.elsevier.nl/locate/econbase

The sensitivity of OLS when the variance matrix is (partially) unknown

Anurag N. Banerjee^a, Jan R. Magnus^{b,*}

^a *Department of Economics, Queens University, 22 University Square, Belfast BT7 INN, UK*

^b *CentER for Economic Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands*

Received 5 May 1996; received in revised form 1 August 1998; accepted 8 October 1998

Abstract

We consider the standard linear regression model $y = X\beta + u$ with all standard assumptions, except that the variance matrix is assumed to be $\sigma^2\Omega(\theta)$, where Ω depends on m unknown parameters $\theta_1, \dots, \theta_m$. Our interest lies exclusively in the mean parameters β or $X\beta$. We introduce a new sensitivity statistic (B1) which is designed to decide whether \hat{y} (or $\hat{\beta}$) is sensitive to covariance misspecification. We show that the Durbin–Watson test is inappropriate in this context, because it measures the sensitivity of $\hat{\sigma}^2$ to covariance misspecification. Our results demonstrate that the estimator $\hat{\beta}$ and the predictor \hat{y} are not very sensitive to covariance misspecification. The statistic is easy to use and performs well even in cases where it is not strictly applicable. © 1999 Elsevier Science S.A. All rights reserved.

JEL classification: C12; C22; C51; C52

Keywords: Linear regression; Least squares; Autocorrelation; Durbin–Watson test; Sensitivity

1. Introduction

We consider the standard linear regression model $y = X\beta + u$ under all standard assumptions except one. Thus, we assume that X is non-random, has full column-rank k , and that u is normally distributed with mean 0. We assume,

* Corresponding author. E-mail: magnus@kub.nl.

however, that the disturbance covariance matrix is $\sigma^2\Omega(\theta)$, where $\sigma^2 > 0$ and the $m \times 1$ vector θ are unknown. Our parameters of interest are $Ey = X\beta$ or, which amounts to the same, β . The covariance parameters σ^2 and θ are nuisance parameters.

If $\theta = 0$, then $\Omega(\theta) = I_n$ (the identity matrix of order $n \times n$, when n is the number of observations) and the OLS estimator $\hat{\beta}_0$ and the OLS predictor \hat{y}_0 are unbiased and efficient. If $\theta \neq 0$, then $\hat{\beta}_0$ and \hat{y}_0 are, in general, no longer efficient.¹ If we know the structure Ω and the values of the m elements of θ , then GLS is more efficient. If we know the structure Ω but not the value of θ , then estimated GLS is not necessarily more efficient than OLS. But in the most common case, where we do not even know the structure Ω , we have to determine Ω and estimate θ . The question then is whether the resulting estimator for β (or $X\beta$) is 'better' than the OLS estimator $\hat{\beta}_0$.

The first step away from white noise disturbances is an AR(1) process, and the most common test for AR(1) disturbances is the Durbin–Watson (*DW*) test. If the *DW* test tells us that the autocorrelation parameter ϕ_1 is positive rather than 0, then most applied econometricians will assume some more general covariance structure. After fitting this more general structure one often finds that the estimates of the parameters of interest (β or $X\beta$) have not changed much; in other words that the estimates of the parameters of interest are fairly robust against covariance misspecification.

In this paper we do not ask whether the covariance parameters (like ϕ_1) are significantly different from 0 or not. Instead we ask whether the GLS estimators $\hat{\beta}$ and \hat{y} are sensitive to deviations from the white noise assumption. Since this appears to be the question of interest, it seems useful to try and answer this question directly. Thus, this paper is *not* a contribution to the diagnostic testing literature where one asks whether it is true that errors are AR(1) or not, the observations are normally distributed or not, and, more generally, whether a nuisance parameter is zero or not. This paper is a contribution to sensitivity measurement where tools are provided for measuring how sensitive the results of an investigation are to the underlying assumptions.

The statistics proposed in this paper can be justified intuitively as follows. Let $s(\theta)$ be the relevant statistic ($\hat{\beta}$, \hat{y} or $\hat{\sigma}^2$). Developing $s(\theta)$ in a Taylor expansion gives

$$s(\theta) = s(0) + \sum_{j=1}^m \theta_j \left. \frac{\partial s(\theta)}{\partial \theta_j} \right|_{\theta=0} + \dots$$

¹ There are cases where $\hat{\beta}_0$ and \hat{y}_0 are efficient, in spite of the fact that $\theta \neq 0$; see Amemiya (1985), Theorem 6.1.1, and Fiebig et al. (1992).

We would consider $s(\theta)$ and $s(0)$ to be ‘almost equal’ if

$$\sum_{j=1}^m \theta_j \left. \frac{\partial s(\theta)}{\partial \theta_j} \right|_{\theta=0} \approx 0$$

and a sufficient condition for this is that

$$\left. \frac{\partial s(\theta)}{\partial \theta_j} \right|_{\theta=0} = 0 \quad (j = 1, \dots, m).$$

Our statistics are based on this simple observation.

Efficiency is a global property. We, however, ask a local question.² If $\hat{\beta}(\theta)$ denotes the GLS estimator for β , given Ω and θ , and if $\hat{y}(\theta) = X\hat{\beta}(\theta)$ is the GLS predictor, then we ask how far $\hat{y}(\theta)$ is removed from $\hat{y}(0)$. It may be that θ is far away from 0, but still $\hat{y}(\theta)$ close to $\hat{y}(0)$. In fact, we know that this situation occurs frequently.

Let $M = I_n - X(X'X)^{-1}X'$ and $\hat{u} = My$. Also, let $T^{(1)}$ be the $n \times n$ matrix such that $T^{(1)}(i, j) = 1$ when $|i - j| = 1$ and 0 elsewhere. We propose a new sensitivity statistic,

$$B1 = \frac{\hat{u}'C^{(1)}(C^{(1)}C^{(1)})^{-1}C^{(1)}\hat{u}}{\hat{u}'\hat{u}},$$

where

$$C^{(1)} = (I_n - M)T^{(1)}M$$

and A^- denotes a generalized inverse of A . We shall show that $B1$ precisely measures the thing we wish to know, namely the sensitivity (or robustness) of \hat{y} and $\hat{\beta}$. Under the null hypothesis of white noise $B1$ has a Beta distribution (Theorem 2) and hence critical values can be found in standard tables.

As a byproduct we also develop a sensitivity statistic $D1$ which is closely related to the DW test, but has a different interpretation.³ Various other results are obtained as well.

The paper is organized as follows. Section 2 gives some preliminary results and definitions. The sensitivity of the predictor \hat{y} is defined in Section 3 and the main result (Theorem 2) is stated and discussed. In Section 4 we obtain the sensitivity of $\hat{\sigma}^2$ and show its relationship with the DW statistic. This completes

² See Leamer (1984) for a global analysis along different lines.

³ The statistic $D1$ is in fact the ‘alternative’ DW test as developed by King (1981).

the theoretical part of the paper. In Section 4 we show that $B1$ and $D1$ are nearly independent and hence that information through the DW test is almost irrelevant for the sensitivity of \hat{y} . In Sections 6 and 7 we study the behaviour of our main sensitivity statistic $B1$. In Section 6 the disturbances follow an ARMA(1, 1) process so that $B1$ is strictly applicable, while in Section 7 the covariance matrix is AR(2) with $\phi_1 = 0$, so that $B1$ is, strictly speaking, not applicable. We show in both cases that $B1$ can be used with profit and that OLS is very robust again covariance misspecification. After some concluding remarks, we provide two appendices. Appendix A contains the proofs of the four theorems. Appendix B contains two theorems on the limit of a ratio of two quadratic forms.

2. Preliminaries

We consider the standard linear regression model

$$y = X\beta + u, \tag{2.1}$$

where y is an $n \times 1$ random vector of observations, X a non-random $n \times k$ matrix of regressors, β a $k \times 1$ vector of unknown parameters and u an $n \times 1$ vector of random disturbances. We assume that X has full column-rank k and that u follows a normal distribution,

$$u \sim N(0, \sigma^2\Omega(\theta)), \tag{2.2}$$

where $\sigma^2 > 0$ and $\Omega(\theta)$ is a matrix function of the $m \times 1$ parameter vector $\theta = (\theta_1, \dots, \theta_m)'$, positive definite and differentiable at least in a neighbourhood of $\theta = 0$. Without loss of generality, we may assume that

$$\Omega(0) = I_n. \tag{2.3}$$

For $s = 1, \dots, m$ we define the $n \times n$ symmetric matrices

$$A_s = \left. \frac{\partial \Omega(\theta)}{\partial \theta_s} \right|_{\theta=0}, \tag{2.4}$$

and we notice that, in view of Eq. (2.3),

$$\left. \frac{\partial \Omega^{-1}(\theta)}{\partial \theta_s} \right|_{\theta=0} = -A_s. \tag{2.5}$$

We denote by $T^{(h)}$, $0 \leq h \leq n - 1$, the $n \times n$ symmetric Toeplitz matrix with

$$T^{(h)}(i,j) = \begin{cases} 1 & \text{if } |i - j| = h, \\ 0 & \text{otherwise.} \end{cases}$$

For example,

$$T^{(0)} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix},$$

$$T^{(1)} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}.$$

If the y -process is covariance stationary, then $\Omega(\theta)$ can be written as

$$\Omega(\theta) = I_n + \omega_1 T^{(1)} + \dots + \omega_{n-1} T^{(n-1)}, \quad (2.6)$$

where $\omega_1, \dots, \omega_{n-1}$ are real-valued functions of θ satisfying $\omega_h(0) = 0$, $1 \leq h \leq n - 1$. Differentiating both sides of Eq. (2.6) with respect to θ_s then yields

$$A_s = \sum_{h=1}^{n-1} \alpha_s^{(h)} T^{(h)}, \quad \text{where } \alpha_s^{(h)} = \left. \frac{\partial \omega_h(\theta)}{\partial \theta_s} \right|_{\theta=0}. \quad (2.7)$$

In many cases of practical interest the coefficients $\alpha_s^{(h)}$ take a very simple form, namely 0 when $h \neq s$ and 1 when $h = s$. This is the case, for example, in a general ARMA (p, q) process.

Theorem 1. Assume that the disturbances $\{u_t\}$ are generated by a stationary ARMA (p, q) process,

$$u_t = \sum_{i=1}^p \phi_i u_{t-i} + \sum_{j=1}^q \psi_j \varepsilon_{t-j} + \varepsilon_t, \tag{2.8}$$

where the ε_t are i.i.d. $N(0, \sigma^2)$. Let $\theta = (\phi_1, \dots, \phi_p, \psi_1, \dots, \psi_q)'$ and let $\sigma^2 \Omega(\theta)$ be the covariance matrix of u_1, \dots, u_n . Then,

$$\left. \frac{\partial \Omega(\theta)}{\partial \phi_s} \right|_{\theta=0} = \left. \frac{\partial \Omega(\theta)}{\partial \psi_s} \right|_{\theta=0} = T^{(s)}. \tag{2.9}$$

Theorem 1 demonstrates the importance of the Toeplitz matrices $T^{(h)}$. In particular the matrix $T^{(1)}$ will play a central role in this paper.

Let $M = I_n - X(X'X)^{-1}X'$ be the usual idempotent matrix. The $n \times n$ matrix

$$C_s = (I_n - M)A_sM \tag{2.10}$$

will play an important role as well. Letting

$$r_s = \text{rank}(C_s), \tag{2.11}$$

we obtain

$$0 \leq r_s \leq \min(k, n - k). \tag{2.12}$$

3. Sensitivity of the predictor

If θ is known, then the parameters β and σ^2 can be estimated by generalized least squares. Thus,

$$\hat{\beta}(\theta) = (X' \Omega^{-1}(\theta) X)^{-1} X' \Omega^{-1}(\theta) y \tag{3.1}$$

and

$$\hat{\sigma}^2(\theta) = \frac{(y - \hat{y}(\theta))' \Omega^{-1}(\theta) (y - \hat{y}(\theta))}{n - k}, \tag{3.2}$$

where $\hat{y}(\theta)$ denotes the predictor for y , that is,

$$\hat{y}(\theta) = X \hat{\beta}(\theta). \tag{3.3}$$

We denote the OLS estimators by $\hat{\beta}_0 = \hat{\beta}(0)$, $\hat{\sigma}_0^2 = \hat{\sigma}^2(0)$ and $\hat{y}_0 = \hat{y}(0)$. We wish to assess how sensitive (linear combinations of) $\hat{\beta}(\theta)$ are with respect to small changes in θ when θ is close to 0. The predictor is the linear combination most suitable for our analysis. Since any estimable linear combination of $\hat{\beta}(\theta)$ is a linear combination of $\hat{y}(\theta)$, and vice versa, this constitutes no loss of generality. We define *the sensitivity of the predictor* $\hat{y}(\theta)$ (with respect to θ_s) as

$$z_s = \left. \frac{\partial \hat{y}(\theta)}{\partial \theta_s} \right|_{\theta=0}. \quad (3.4)$$

The sensitivity of $\hat{\beta}(\theta)$ (with respect to θ_s) is then

$$\left. \frac{\partial \hat{\beta}(\theta)}{\partial \theta_s} \right|_{\theta=0} = (X'X)^{-1} X' z_s.$$

In order to use the (normally distributed) $n \times 1$ vector z_s as a sensitivity statistic, we transform it into a χ^2 -variable in the usual way. We thus propose

$$B_s = \frac{z_s'(C_s C_s')^{-1} z_s}{(n-k)\hat{\sigma}^2(0)} \quad (3.5)$$

as a statistic to measure the sensitivity of the predictor $\hat{y}(\theta)$ with respect to θ_s . (The notation A^{-} denotes a generalized inverse of A .) Large values of B_s indicate that $\hat{y}(\theta)$ is sensitive to small changes in θ_s when θ is close to 0 and therefore that setting $\theta_s = 0$ is not justified. The statistic B_s depends only on y and X and can therefore be observed. Since the distribution of y depends on θ , so does the distribution of B_s . We now state our main result.

Theorem 2. We have

- (a) $z_s = -C_s y$;
- (b) $B_s = y' W_s y / y' M y$, $W_s = C_s'(C_s C_s')^{-1} C_s$;
- (c) If $0 < r_s < n - k$ and the distribution of y is evaluated at $\theta = 0$, then

$$B_s \sim \text{Beta}(r_s/2, (n - k - r_s)/2).$$

In view of Theorem 1, we shall be particularly interested in the case where A_s is a Toeplitz matrix, that is

$$A_s = T^{(h)} \quad \text{for some } h. \quad (3.6)$$

This is a very common situation for stationary processes and the matrix C_s then becomes

$$C_s = (I_n - M)T^{(h)}M. \tag{3.7}$$

The most important special case in practice is $A_s = T^{(1)}$ and we shall denote the corresponding B_s -statistic as $B1$. We know that $B1$ measures the sensitivity of $\hat{y}(\theta)$ with respect to the AR(1) or MA(1) parameter (see Theorem 1). The statistic $B1$ should be seen as an alternative to the Durbin–Watson statistic. But where the DW statistic answers the question ‘Is ϕ equal to 0?’, our $B1$ statistic answers the question ‘Are \hat{y} and $\hat{\beta}$ sensitive to the fact that ϕ may not be 0?’ In most practical situations the latter question seems more appropriate. In the next section we shall see that DW is essentially the sensitivity of $\hat{\sigma}^2(\theta)$. Hence, we can interpret DW as answering the question ‘Is $\hat{\sigma}^2$ sensitive to ϕ ?’ Thus, DW turns out to be measuring the sensitivity of the estimator for the variance of y , while $B1$ measures the sensitivity of the estimator for its mean. Again, in most practical situations our primary interest lies in the mean of y . $B1$ provides a direct measure for its sensitivity.

Let us return briefly to the conditions in Theorem 2(c). We demand that $0 < r_s < n - k$. From Eq. (2.12) we already know that $0 \leq r_s \leq \min(k, n - k)$. If $r_s = n - k$, then $W_s = M$ (see Magnus and Neudecker, 1999, Theorem 2.8), $B_s = 1$, and $\hat{\sigma}^2(0) = z'_s(C_s C'_s)^- z_s / (n - k)$. The condition $r_s < n - k$ is automatically fulfilled when $n > 2k$. In practice we usually have $r_s = k < n - k$. The condition $r_s > 0$ is more interesting. The situation $r_s = 0$ occurs for example in the two-error components model, where

$$\Omega(\theta) = E + (1 + \theta)(I_n - E)$$

and

$$E = \frac{1}{n} ii', \quad i = (1, 1, \dots, 1)'$$

If the regression contains an intercept, so that $Mi = 0$, then it is easy to see that

$$A_1 = \frac{\partial \Omega(\theta)}{\partial \theta} = I_n - E \quad \text{and} \quad C_1 = (I_n - M)A_1M = 0.$$

In fact, \hat{y} and $\hat{\beta}$ do not depend on θ at all in this case, because the two-error components model (with constant term) is one example where $GLS = OLS$, that is,

$$(X'\Omega^{-1}(\theta)X)^{-1}X'\Omega^{-1}(\theta)y = (X'X)^{-1}X'y \tag{3.8}$$

for every θ .⁴ Apart from such unusual circumstances, the condition $0 < r_s < n - k$ is a very mild one.

In order to compute B_s we need to compute W_s , which involves a generalized inverse. This is most easily accomplished by finding the $n \times r_s$ matrix S_s whose columns are the normalized eigenvectors of $C'_s C_s$, associated with its r_s positive eigenvalues. Then, $W_s = C'_s (C_s C'_s)^- C_s = S_s S'_s$.

We shall study the behaviour of $B1$ and related statistics in detail, but first we develop its counterpart, the sensitivity of $\hat{\sigma}^2$.

4. Sensitivity of the variance estimator

In order to assess the sensitivity of the variance estimator $\hat{\sigma}^2(\theta)$ with respect to small changes in θ , we define the sensitivity of $\hat{\sigma}^2(\theta)$ (with respect to θ_s) as

$$\lambda_s = \left. \frac{\partial \hat{\sigma}^2(\theta)}{\partial \theta_s} \right|_{\theta=0}. \quad (4.1)$$

Upon scaling we find

$$D_s = \frac{\lambda_s}{\hat{\sigma}^2(0)} = \left. \frac{\partial \log \hat{\sigma}^2(\theta)}{\partial \theta_s} \right|_{\theta=0} \quad (4.2)$$

as a suitable statistic for our purpose.

Theorem 3. We have

(a) $\lambda_s = -y' M A_s M y / (n - k)$;

(b) $D_s = -y' M A_s M y / y' M y$;

(c) *If the distribution of y is evaluated at $\theta = 0$, then $D_s = -v' P' A_s P v / v' v$, where P is an $n \times (n - k)$ matrix containing the $n - k$ eigenvectors of M associated with the eigenvalue 1, that is, $M = P P'$, $P' P = I_{n-k}$, and $v \sim N(0, I_{n-k})$.*

Theorem 3 shows that D_s has the same form as the DW statistic. We could obtain upper and lower bounds, using Poincaré's separation theorem, in terms of the eigenvalues of A_s , just as for the DW statistic. Again the most important special case occurs when $A_s = T^{(1)}$ (that is, AR(1) or MA(1)). The corresponding D_s -statistic will be denoted $D1$. This case was considered by Dufour and King (1991, Theorem 1) as a locally best invariant test of $\phi = 0$ against $\phi > 0$.⁵ Not

⁴ It is well-known that GLS = OLS if and only if $M\Omega X = 0$; see Kruskal (1968) and Fiebig et al. (1992).

⁵ King and Evans (1988) show that the DW test is approximately locally best invariant in the case of ARMA(1,1) disturbances.

surprisingly, $D1$ is closely related to the DW statistic, a fact first observed by King (1981).

Theorem 4. In the special case $A_s = T^{(1)}$, we have

$$B_s = B1 = \frac{\hat{u}' W^{(1)} \hat{u}}{\hat{u}' \hat{u}}, \quad \text{where } W^{(1)} = C^{(1)'}(C^{(1)} C^{(1)'})^{-1} C^{(1)},$$

and

$$D_s = D1 = -\frac{\hat{u}' T^{(1)} \hat{u}}{\hat{u}' \hat{u}} = DW - 2 + R/n,$$

where $\hat{u} = My$ is the vector of residuals after fitting OLS, $C^{(1)} = (I - M)T^{(1)}M$, DW denotes the Durbin–Watson statistic,

$$DW = \frac{\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^n \hat{u}_t^2},$$

and $R = (\hat{u}_1^2 + \hat{u}_n^2)/(\sum \hat{u}_t^2/n)$ is a remainder term.

At this point several observations can be made. First, we see from Theorem 1 that $T^{(1)}$ is equally relevant in the AR(1) and MA(1) case (and indeed, the ARMA(1, 1) case). From Theorem 4 we see that $B1$ and $D1$ depend on $T^{(1)}$ and hence are identical for AR(1) and MA(1). This explains, inter alia, the conclusion of Griffiths and Beesley (1984) that a pretest estimator based on an AR and an MA pretest performs essentially the same as a pretest estimator based on only an AR pretest. Secondly, any likelihood-based test (such as Lagrange multiplier, see Breusch and Pagan (1980)) uses the derivatives of the loglikelihood, such as $\partial\Omega(\theta)/\partial\theta_s$. Under the null hypothesis $\theta = 0$ the test thus depends on $A_s = \partial\Omega(\theta)/\partial\theta_s|_{\theta=0}$. This is why the matrix A_s plays such an important role in many test statistics. Any pretest which depends on $A_s = T^{(1)}$ will not be appropriate to distinguish between AR(1) and MA(1). A survey of the DW and $D1$ statistics is given in King (1987).

5. Near independence of $B1$ and $D1$

Before we calculate the sensitivity statistics $B1$ and $D1$ for various distributions, we consider another question. Recall that $D1$ is essentially the DW statistic. The DW statistic is designed to test $\phi_1 = 0$ against $\phi_1 > 0$. The equivalence with $D1$ shows that, in fact, DW measures the sensitivity of $\hat{\sigma}^2$ with

respect to small changes in the AR(1) parameter ϕ_1 when ϕ_1 is close to 0. Our new statistic $B1$, on the other hand, measures the sensitivity of \hat{y} (or $\hat{\beta}$) with respect to small changes in ϕ_1 . Since, as a rule, econometricians tend to be interested in β (or functions thereof) and consider ϕ_1 a nuisance parameter, $B1$ appears to be more appropriate than $D1$ or DW . After all, it measures directly the thing we wish to know: Are our estimates for β (and functions thereof) sensitive to misspecification in the disturbance covariance matrix.

In this section we show that $B1$ and $D1$ are almost independent. This is important because it implies that rejecting $\phi_1 = 0$ using the $D1$ or DW test in favour of $\phi_1 > 0$ gives us very little information on how sensitive $\hat{\beta}$ or \hat{y} are to small changes in ϕ_1 . So it may very well happen that the DW test firmly rejects $\phi_1 = 0$, but that nevertheless the β estimates change very little, a fact all practical econometricians are familiar with.

For this and further experiments we have generated five regressors:

x_1 : 1(constant),

x_2 : 1, 2, ... (time trend),

x_3 : normal distribution, $E x_3 = 0$, $\text{var}(x_3) = 9$,

x_4 : lognormal distribution, $E \log x_4 = 0$, $\text{var}(\log x_4) = 9$,

x_5 : uniform distribution, $-2 \leq x_5 \leq 2$.

These regressors can be combined in various data sets. We consider five datasets with two regressors and five with three regressors (see Table 1). Now consider one of these ten datasets. Let $n = 25$ and assume that the disturbances are generated by white noise. Calculate the critical values $B1^*$ and $D1^*$ such that

$$\Pr(B1 > B1^*) = \alpha = 0.05$$

Table 1

The conditional probability that $B1 \leq B1^*$ given that $D1 \leq D1^*$ for 10 data sets ($n = 25$, $\alpha = 0.05$)

Dataset	Regressors	$\Pr(B1 \leq B1^* D1 \leq D1^*)$
1	Constant, time trend	0.882
2	Constant, normal	0.922
3	Constant, lognormal	0.927
4	Uniform, normal	0.966
5	Time trend, normal	0.924
6	Constant, time trend, normal	0.890
7	Constant, time trend, lognormal	0.894
8	Constant, uniform, lognormal	0.930
9	Uniform, normal, lognormal	0.977
10	Time trend, normal, uniform	0.934

and

$$\Pr(D1 \leq D1^*) = \alpha = 0.05.$$

We define the joint probabilities

$$p_{11} = \Pr(B1 > B1^* \text{ and } D1 \leq D1^*),$$

$$p_{12} = \Pr(B1 > B1^* \text{ and } D1 > D1^*),$$

$$p_{21} = \Pr(B1 \leq B1^* \text{ and } D1 \leq D1^*),$$

$$p_{22} = \Pr(B1 \leq B1^* \text{ and } D1 > D1^*).$$

To simulate the joint probabilities we generate 10,000 replications of 25 i.i.d. $N(0, 1)$ variates. For each of the 10,000 replications we calculate $B1$ and $D1$ and compute the relative frequencies f_{11} , f_{12} , f_{21} and f_{22} . We wish to estimate p_{21} , the probability that $B1 \leq B1^*$ and $D1 \leq D1^*$, that is, the probability that \hat{y} is *not* sensitive while at the same time $\hat{\sigma}^2$ is sensitive to small changes in ϕ close to 0. We could estimate p_{21} by f_{21} , but a more efficient estimate is obtained by taking account of the restrictions

$$p_{11} + p_{12} = \alpha, \quad p_{11} + p_{21} = \alpha, \quad p_{11} + p_{12} + p_{21} + p_{22} = 1.$$

The parameter p_{21} is then estimated from the multinomial distribution, which is proportional to

$$p_{11}^{m_{11}} p_{12}^{m_{12}} p_{21}^{m_{21}} p_{22}^{m_{22}},$$

where $m_{ij} = mf_{ij}$ and $m = 10,000$.⁶ Taking into account the three constraints, the likelihood is maximized when, for $0 < p_{21} < \alpha$,

$$p_{21}^2 - ((1 - \alpha)f_{11} + f_{12} + f_{21} + \alpha f_{22})p_{21} + \alpha(1 - \alpha)(f_{12} + f_{21}) = 0.$$

Solving this quadratic equation gives the ML estimate for p_{21} . Dividing by α gives an estimate of the conditional probability $\Pr(B1 \leq B1^* | D1 \leq D1^*)$.

If the two events $B1 \leq B1^*$ and $D1 \leq D1^*$ were independent, we would find a conditional probability of 0.95 for each of the ten data sets. On the other hand, if the two events were perfectly dependent, then they would never occur together

⁶ Thursby (1981) uses Monte Carlo simulations to test for the independence of DW , RESET and other procedures, but he only uses the relative frequency f_{21} .

and the conditional probability would be 0. Table 1 shows that, while the conditional probability is not equal to 0.95, it is nevertheless very close. The conclusion of the simulation experiment is therefore that the $D1$ or DW statistic tells us almost nothing about the thing we wish to know, namely how sensitive $\hat{\beta}$ and \hat{y} are to misspecification in the disturbance covariance matrix. To know this we must use another statistic, namely $B1$.

Although we have emphasized in this section the complimentary roles of $B1$ and $D1$, there is a basic difference between them. OLS is unbiased (consistent) for β regardless of whether $B1$ shows sensitivity. This is not true in relation to $\hat{\sigma}^2$ when $D1$ indicates sensitivity.

6. Behaviour of $B1$ in the case of ARMA(1, 1) disturbances

We know from Theorem 2 that $B1$ follows a Beta distribution when the disturbances are white noise. The logical next step is to ask how $B1$ behaves when the disturbances follow some more general stationary process. In this section we answer this question for the case where the disturbances follow a stationary ARMA(1, 1) process. The covariance matrix then has two parameters (apart from σ^2): ϕ_1 and ψ_1 , associated with the AR and MA part of the process respectively. Theorem 1 shows that each of the three cases AR(1), MA(1) and ARMA(1, 1) leads to the same B_s -statistic, namely $B1$.⁷ Hence for each of these cases the correct procedure for measuring the sensitivity of \hat{y} (and $\hat{\beta}$) is to use $B1$. Similarly, the correct procedure for measuring the sensitivity of $\hat{\sigma}^2$ is to use $D1$, which is essentially the DW -statistic.

We have 10 data sets; see Table 1. For each dataset we calculate $B1^*$ and $D1^*$ such that

$$\Pr(B1 > B1^*) = \alpha \quad \text{and} \quad \Pr(D1 \leq D1^*) = \alpha, \quad (6.1)$$

where $\alpha = 0.05$ and the disturbances are assumed white noise. In Fig. 1 we have calculated

$$\Pr(B1 > B1^*) \quad \text{and} \quad \Pr(D1 \leq D1^*) \quad (6.2)$$

under the assumption that the disturbances are AR(1) for values of ϕ_1 between 0 and 1. Each line in the figure corresponds to one of the ten different data sets.

As noted before, the $D1$ -statistic is essentially the DW -statistic. As a result, $\Pr(D1 \leq D1^*)$ can be interpreted as the power of $D1$ in testing $\phi_1 = 0$ against

⁷ Even when the AR(1) covariance matrix is based on a fixed start-up, say $u_0 = 0$, as in Berenblut and Webb (1973), the $B1$ and $D1$ statistics are applicable.

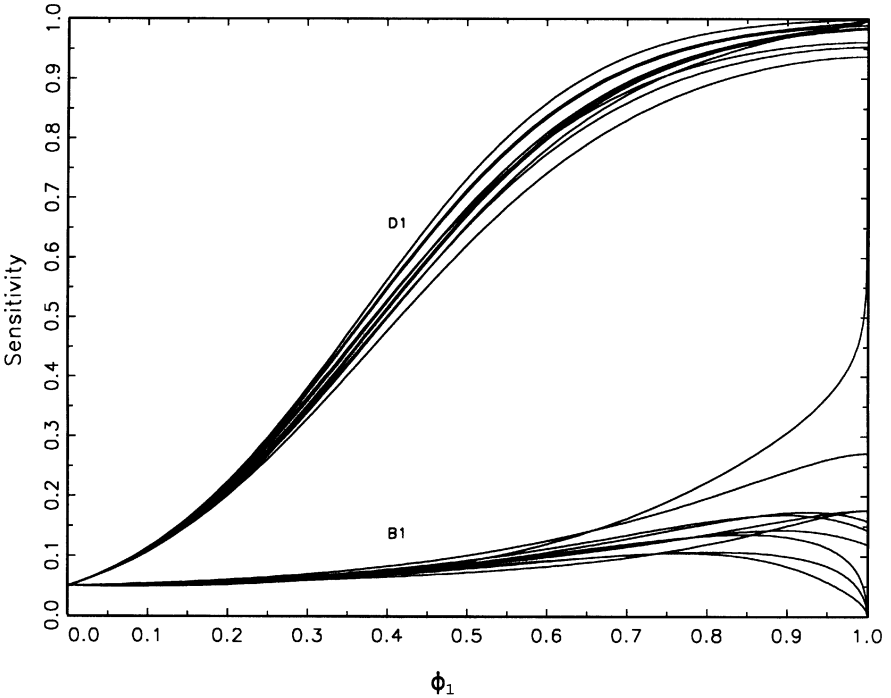


Fig. 1. B1 and D1: AR(1) disturbances, $\alpha = 0.05$.

$\phi_1 > 0$. Alternatively we can interpret $\Pr(D1 \leq D1^*)$ as the sensitivity of $\hat{\sigma}^2$ with respect to ϕ_1 . In the same way, $B1$ measures the sensitivity of \hat{y} (and $\hat{\beta}$) with respect to ϕ_1 . One glance at Fig. 1 shows that $B1$ is quite insensitive, hence robust, with respect to ϕ_1 , even for values of ϕ_1 close to 1. The figure shows the probabilities (6.2) for $n = 25$. The main conclusion is that $D1$ is quite sensitive to ϕ_1 but $B1$ is not. Hence, the $D1$ or DW statistic may indicate that OLS is not appropriate since ϕ_1 is ‘significantly’ different from 0, but the $B1$ statistic shows that the estimates \hat{y} and $\hat{\beta}$ are little effected. This explains and illustrates a phenomenon well-known to all applied econometricians.

The probabilities were all calculated using our own adaptation of Imhof’s (1961) routine which is available in the NAG (1991) library and elsewhere.⁸ If ϕ_1 is close to 1, then the limit (or the limiting distribution) can be calculated from Theorem B1 in Appendix B. If there is no intercept in the regression, then $\Pr(B1 > B1^*)$ either approaches 0 or 1. (This result relates closely to Krämer (1985).) We can see from Fig. 1 that there are three data sets where $\Pr(B1 > B1^*)$

⁸ See also Koerts and Abrahamse (1969) on the computational aspects of these probabilities.

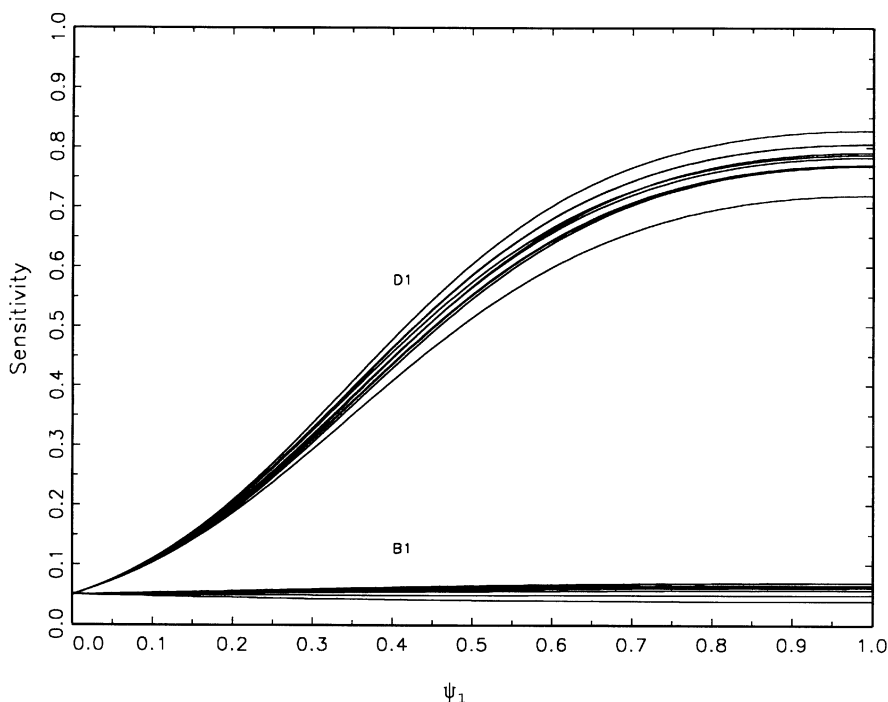


Fig. 2. B1 and D1: MA(1) disturbances, $\alpha = 0.05$.

approaches 0 (numbers 4, 5 and 10) and one where the probability approaches 1 (number 9). If, however, there is an intercept in the regression, then $\Pr(B1 > B1^*)$ approaches some limit between 0 and 1.

The flatness of the $B1$ -curves is, of course, in accordance with the near-independence discussed in the previous section. For $n = 25$ and $\phi_1 = 0.5$ we would decide in only about 7–10% of the cases that \hat{y} is sensitive with respect to ϕ_1 .⁹

In the case of MA(1) disturbances the general conclusions are the same, except that for MA(1) disturbances no difficulties arise close to $\psi_1 = 1$. Fig. 2 is the counterpart to Fig. 1. $D1$ is less sensitive than in the case of AR(1) disturbances, that is, the DW statistic has less power, and the $B1$ statistic is almost flat and hence \hat{y} and $\hat{\beta}$ are extremely robust against MA(1) disturbances.

Fig. 3 shows that \hat{y} and $\hat{\beta}$ are also quite insensitive to ARMA(1,1) disturbances. The figure is based on the same probabilities as before with $\psi_1 = 0.5$ and

⁹ King and Giles (1984) show that the t -test loses power when there is autocorrelation. This is somewhat related to our result, since $B1$ is an F -type statistic.

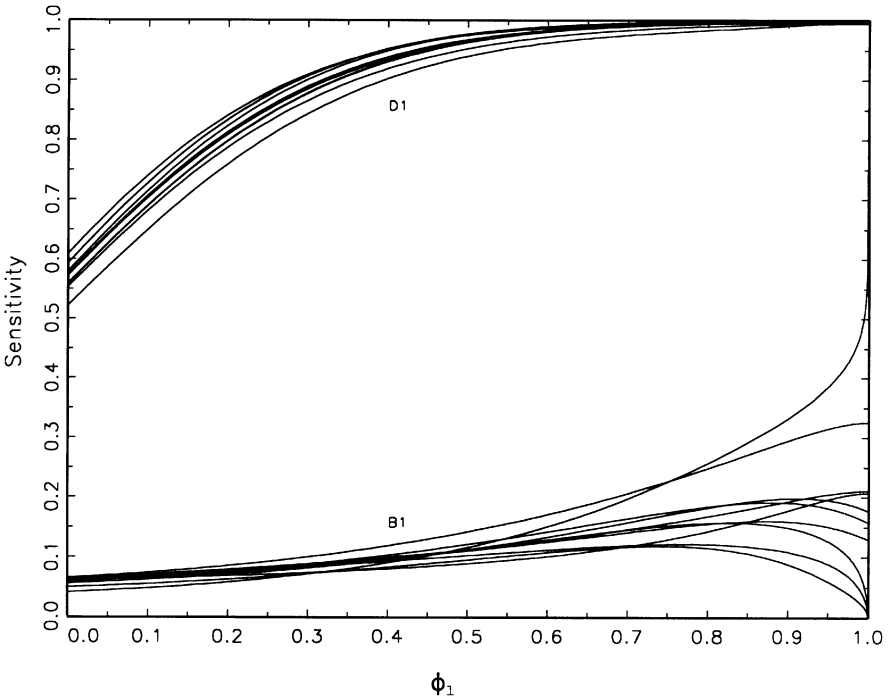


Fig. 3. B1 and D1: ARMA(1,1) disturbances, $\psi_1 = 0.5, \alpha = 0.05$.

$n = 25$. The graph of the $B1$ -statistic closely resembles the graph in Fig. 1. The behaviour close to $\phi_1 = 1$ is given in Theorem B.1 in Appendix B.¹⁰

So far we have shown that our proposed sensitivity statistic $B1$ indicates that \hat{y} (and $\hat{\beta}$) is quite insensitive to covariance misspecification, whereas $D1$ (and DW) indicates substantial sensitivity of $\hat{\sigma}^2$. Could it be true that $B1$ is *never* sensitive to covariance misspecification. If this were the case, then it would clearly be unnecessary to calculate $B1$ for any specific set of observed data. It turns out that this is *not* the case and hence that $B1$ is a useful sensitivity measure. We now provide an example where $B1$ is sensitive. Let $T^{(1)}$ be the $n \times n$ matrix defined in Section 2 and let t_1, \dots, t_n be the eigenvectors associated with the eigenvalues of $T^{(1)}$, ranked in ascending order. Let

$$x_1 = \frac{t_1 + t_{n-2}}{\sqrt{2}}, \quad x_2 = \frac{t_2 + t_{n-3}}{\sqrt{2}}, \quad x_3 = \frac{t_3 + t_{n-4}}{\sqrt{2}}.$$

¹⁰ A lot of work has been done on the power curves of the DW statistic and, to a lesser extent, the $D1$ statistic. See Berenblut and Webb (1973), Tillman (1975) and Bartels (1992).

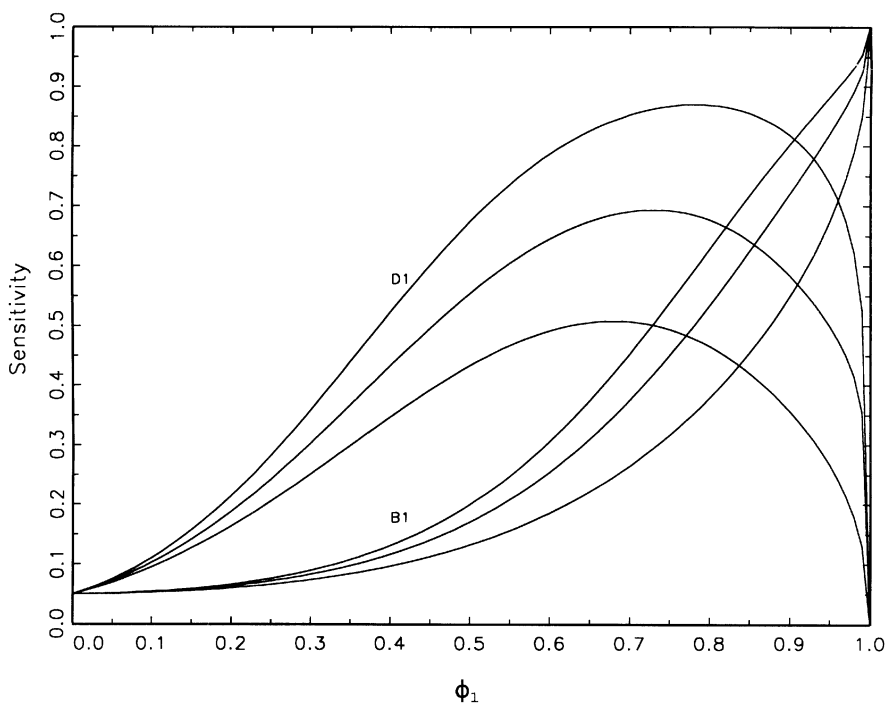


Fig. 4. B1 and D1: AR(1) disturbances, $\alpha = 0.05$.

Tillman (1975) found that the *DW* test has poor power in certain cases, and x_1 , x_2 , and x_3 are slight modifications of Tillman's regressors; see also King and Giles (1984).¹¹ The three data sets used in Fig. 4 are $\{x_1\}$, $\{x_1, x_2\}$ and $\{x_1, x_2, x_3\}$. We see that *B1* is now quite sensitive, especially for values of ϕ_1 close to 1.

Figs. 1–3 give the sensitivities for one value of n , namely $n = 25$. To see how *B1* depends on n we calculate for each of our ten data sets $\Pr(B1 > B1^*)$ for three values of n ($n = 10, 25, 50$) and two covariance specifications (AR(1), MA(1)). The results are given in Table 2.

Table 2 confirms our earlier statements. In only 5–10% of the cases would we conclude that \hat{y} and $\hat{\beta}$ are sensitive to AR(1) or MA(1) disturbances. High values of n are needed to get close to the probability limit and the higher is $\phi_1 > 0$, the higher should be n (see also Sharma, 1987).

In this section we have investigated the sensitivity of the predictor \hat{y} (and estimator $\hat{\beta}$) when the disturbances follow an ARMA(1, 1) process. The

¹¹ Tillman's analysis is in terms of the usual A -matrix, defined by $u' Au = \sum_2^n (u_i - u_{i-1})^2$. Since A is approximately equal to $2I_n - T^{(1)}$, the eigenvectors of $T^{(1)}$ will be approximately equal to those of A .

Table 2

Pr ($B1 > B1^*$), $\alpha = 0.05$, for two covariance specifications and three values of n

Data set	AR(1), $\phi_1 = 0.5$			MA(1), $\psi_1 = 0.5$		
	$n = 10$	$n = 25$	$n = 50$	$n = 10$	$n = 25$	$n = 50$
1	0.078	0.072	0.063	0.070	0.061	0.056
2	0.073	0.087	0.073	0.077	0.062	0.050
3	0.101	0.092	0.089	0.073	0.059	0.055
4	0.073	0.079	0.080	0.069	0.049	0.046
5	0.077	0.082	0.064	0.085	0.062	0.049
6	0.093	0.085	0.069	0.092	0.065	0.053
7	0.092	0.101	0.087	0.082	0.065	0.059
8	0.096	0.088	0.078	0.087	0.056	0.052
9	0.081	0.091	0.096	0.051	0.041	0.043
10	0.104	0.087	0.069	0.099	0.059	0.051

sensitivity was measured using $B1$ which is the correct measure in this case. All calculations indicate that OLS is very robust against ARMA(1, 1) disturbances. In only about 5–10% of the cases does the $B1$ statistic lead us to conclude that OLS is not appropriate for predicting y or estimating β . Our next question is how $B1$ behaves in more general situations.

7. Behaviour of $B1$ in the case of AR(2) disturbances

Let us now consider covariance structures more general than an ARMA(1, 1) process. Almost all stationary processes will have either an AR(1) or an MA(1) component, so that the $B1$ statistic has a justification. In this section we consider the AR(2) process with parameters ϕ_1 and ϕ_2 where $\phi_1 = 0$. In this situation $B1$ is not the correct sensitivity statistic, the correct one being

$$B2 = \frac{\hat{u}'C^{(2)'}(C^{(2)}C^{(2)})^{-1}C^{(2)}\hat{u}}{\hat{u}'\hat{u}}, \quad (7.1)$$

where \hat{u} denotes the vector of OLS residuals and

$$C^{(2)} = (I - M)T^{(2)}M. \quad (7.2)$$

If we know that AR(2) with $\phi_1 = 0$ is the only alternative to white noise, we would use $B2$ to find out whether OLS is still reasonable or not. In most practical situations, however, we do not know this. If we calculate the probabilities $\Pr(B1 > B1^*)$ and $\Pr(B2 > B2^*)$ for $0 < \phi_2 < 1$ we find that $B1$ is more

sensitive than $B2$ with respect to ϕ_2 , even though $B2$ is the correct statistic. This is true for nine of the ten data sets. For $D1$ compared with $D2$ the opposite is the case. $D1$ is less sensitive than $D2$, or, put differently, the DW test is less powerful than the appropriate $AR(2)$ test, which is what we would expect. See Blattberg (1973), Knottnerus (1985), and Harvey (1990, p. 210) for an investigation of the (in)appropriateness of the DW test in this case.

Under the current specification of $AR(2)$ with $\phi_1 = 0$ the correct $B2$ statistic will show sensitivity about 7% of the time, depending of course on the value of ϕ_2 and the data set. The incorrect $B1$ statistic will show sensitivity about 12% of the time. Thus, using $B1$ in this case will lead us to conclude that OLS is sensitive slightly more often than is justified.

We conclude that $B1$ can be usefully employed even in cases for which it was not designed. With 25 observations we will reject OLS slightly more frequently than is necessary, but of course much less frequently than if we were using the DW test.

The behaviour of $B1$ and $B2$ close to $\phi_2 = 1$ is interesting, see Theorem B.2 in Appendix B. In the usual situation when the regression has an intercept, both $B1$ and $B2$ converge to a nonrandom limit and the appropriate probability therefore converges either to 0 or to 1. If the regression does not have an intercept, both $B1$ and $B2$ converge to a random variable. This is just the opposite situation as the behaviour under $AR(1)$.

8. Concluding remarks

In this paper we have introduced a new sensitivity measure, $B1$, which is designed to decide whether the predictor \hat{y} (or the estimator $\hat{\beta}$) is sensitive to covariance misspecification. Many applied statisticians use the Durbin–Watson (DW) test for this purpose, but we show that the DW test can be interpreted as a statistic to decide whether the variance estimator $\hat{\sigma}^2$ is sensitive to covariance misspecification. In most situations we are not interested in the variance parameters themselves, which are nuisance parameters, but rather in the mean parameters β of $X\beta$. Our new sensitivity statistic $B1$ may then provide a useful tool for analysis. The case for a new statistic is strengthened by the fact that the DW test and $B1$ are almost orthogonal to each other (Section 5). That is, we may very well conclude from the DW test that there exists positive autocorrelation, while at the same time $B1$ shows little sensitivity of $\hat{\beta}$ and \hat{y} with respect to the autocorrelation parameter.

Our results show that $\hat{\beta}$ and \hat{y} are not very sensitive to covariance misspecification, a fact well-known to applied statisticians. The $B1$ statistic is easy to use and performs well even in cases where it is not strictly applicable (Section 7).

We note that, even when $\hat{\beta}$ is not sensitive to covariance misspecification, its estimated variance $\widehat{\text{var}}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1}$ may very well be. The $D1$ statistic (or the

DW test) is appropriate to measure the sensitivity of $\hat{\sigma}^2$. Hence, if we are only interested in estimation, then *B1* suffices. But if we are interested in inference, then both *B1* and *D1* are relevant.¹²

The machinery developed in this paper can also be used to find an alternative to the Wallis test. Wallis (1972) introduced the test statistic

$$d_4 = \frac{\sum_{t=5}^n (\hat{u}_t - \hat{u}_{t-4})^2}{\sum_{t=1}^n \hat{u}_t^2}, \quad (8.1)$$

where, as before, \hat{u}_t denotes the t th OLS residual. The Wallis test can be used to test for fourth-order autocorrelation in quarterly regression equations. Clearly the Wallis test is the exact counterpart of the *DW* test. It tests $\phi_4 = 0$ against $\phi_4 > 0$. In this situation we have an AR(4) process with $\phi_1 = \phi_2 = \phi_3 = 0$ and the correct sensitivity statistic should be based on

$$B4 = \frac{\hat{u}' C^{(4)'} (C^{(4)} C^{(4)})^{-1} C^{(4)} \hat{u}}{\hat{u}' \hat{u}}, \quad (8.2)$$

where

$$C^{(4)} = (I - M)T^{(4)}M. \quad (8.3)$$

If we compare *B1* with *B4*, we arrive at the same general conclusions as in Section 7. In particular, *B1* is usually more sensitive than *B4* with respect to ϕ_4 . However, if we have quarterly observations, it is quite sensible to measure directly the impact of possible AR(4) disturbances on the estimates $\hat{\beta}$ and the predictor \hat{y} . Calculations show that $\phi_4 = 0$ might be firmly rejected by the *D4* test (which is essentially the Wallis test), but that, again, the estimates of β will not be much affected. The *B4* statistic can be used as an alternative to the Wallis test, just as the *B1* statistic can be used as an alternative to the *DW* test.

Acknowledgements

Preliminary versions of this paper were presented at the European Doctoral Programme Jamboree, London School of Economics (1994), the NAKE AIO

¹² For known θ , Fiebig et al. (1992) give necessary and sufficient conditions on X and $\Omega(\theta)$ for the complete insensitivity of $\hat{\beta}$ and $\hat{\sigma}^2$.

Presentation Day, Nijmegen (1995), and at CentER for Economic Research, Tilburg University (1995). We are grateful to the workshop participants, to Bob Bartels, Ben van der Genugten and Siem Jan Koopman, and to two referees for their constructive comments.

Appendix A. Proof of theorems

Proof of Theorem 1. We shall show that $\partial\Omega(\theta)/\partial\psi_s = T^{(s)}$ at $\theta = 0$. The second statement is proved similarly. Following Harvey (1993, p. 29) we introduce the autocovariance generating function

$$g(L) = \sum_{h=-\infty}^{\infty} \omega_h L^h,$$

where L is the lag-operator and ω_h is the autocovariance at lag h . For the ARMA(p, q) model we have

$$g(L) = \frac{\psi(L)\psi(L^{-1})}{\phi(L)\phi(L^{-1})} \cdot \sigma^2,$$

where

$$\psi(L) = 1 + \psi_1 L + \dots + \psi_q L^q,$$

$$\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p.$$

Differentiating $g(L)$ with respect to ψ_s gives

$$\frac{\partial g(L)}{\partial \psi_s} = \frac{\sigma^2}{\phi(L)\phi(L^{-1})} (\psi(L^{-1})L^s + \psi(L)L^{-s})$$

and hence, at $\theta = 0$,

$$\left. \frac{\partial g(L)}{\partial \psi_s} \right|_{\theta=0} = \sigma^2(L^s + L^{-s}).$$

For a finite sample (u_1, \dots, u_n) the last equality implies Eq. (2.9). \square

Proof of Theorem 2. Using standard results of differential calculus (see Magnus and Neudecker, 1999) we obtain from Eqs. (3.1) and (3.3)

$$d\hat{y}(\theta) = X(X'\Omega^{-1}(\theta)X)^{-1}X'(d\Omega^{-1}(\theta))(y - X\hat{\beta}(\theta))$$

and hence, at $\theta = 0$,

$$z_s = -X(X'X)^{-1}X'A_sMy = -C_sy.$$

This proves (a). To prove (b) we insert (a) in Eq. (3.5). To prove (c) we notice that $C_sX = 0$ and $MX = 0$. Evaluating the distribution of y at $\theta = 0$ we then find

$$B_s = \frac{v'W_s v}{v'Mv} = \frac{v'W_s v}{v'W_s v + v'(M - W_s)v},$$

where $v \sim N(0, I_n)$. Now, W_s is idempotent with $\text{rank}(W_s) = \text{rank}(C_s) = r_s$. Also, since $MC'_s = C'_s$, we have $MW_s = W_s$. Hence $M - W_s$ is idempotent as well and its rank is $n - k - r_s$. The condition $0 < r_s < n - k$ implies that both W_s and $M - W_s$ have $\text{rank} \geq 1$. It follows that $v'W_s v \sim \chi^2(r_s)$, $v'(M - W_s)v \sim \chi^2(n - k - r_s)$ and the two quadratic forms are independent (because $(M - W_s)W_s = 0$). The result follows. \square

Proof of Theorem 3. Differentiating $\hat{\sigma}^2(\theta)$ in Eq. (3.2) gives

$$(n - k)d\hat{\sigma}^2(\theta) = -2(y - \hat{y}(\theta))'\Omega^{-1}(\theta)d\hat{y}(\theta) + (y - \hat{y}(\theta))'(d\Omega^{-1}(\theta))(y - \hat{y}(\theta))$$

and hence, at $\theta = 0$,

$$(n - k)\lambda_s = 2y'MC_s y - y'MA_s M y = -y'MA_s M y,$$

since $MC_s = 0$. This proves (a). For (b) we simply note that $(n - k)\hat{\sigma}^2(0) = y'My$. To prove (c) we let $v = P'y/\sigma \sim N(0, I_{n-k})$. \square

Proof of Theorem 4. This follows directly from Theorems 2(b) and 3(b) and the fact that

$$\hat{u}'T^{(1)}\hat{u} = 2\sum_{t=2}^n \hat{u}_t \hat{u}_{t-1} = -\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2 + \sum_{t=2}^n \hat{u}_t^2 + \sum_{t=2}^n \hat{u}_{t-1}^2. \quad \square$$

Appendix B. Two results on the limit of a ratio of two quadratic forms

In this appendix we prove two results of independent interest. The first result contains as a special case the result of Sargan and Bhargava (1983), who establish the limit as $\phi \rightarrow 1$ of the *DW* statistic when the process is AR(1) and the model contains a constant term, and also the ‘main theorem’ of Krämer (1985), who shows that the *DW* statistic approaches a certain nonrandom quantity when the process is AR(1) and the model does not contain a constant term. Tables of the *DW* statistic when there is no intercept term were computed by Farebrother (1980). For a survey of the relevant literature, see King (1987). Theorem B1 generalizes both results to the case ARMA(1, 1).

Theorem B.1. Assume that the observations $y = (y_1, \dots, y_n)'$ are generated by a stationary ARMA (1, 1) process,

$$y_t = \phi y_{t-1} + \psi \varepsilon_{t-1} + \varepsilon_t,$$

where the ε_t are i.i.d. $N(0, \sigma^2)$. Let A be a symmetric $n \times n$ matrix and $B \neq 0$ a symmetric positive semidefinite $n \times n$ matrix. Assume that if $Bi = 0$ and $i' Ai = 0$, then $Ai = 0$ (see Note 1 below). Then, as $\phi \rightarrow 1$,

$$\frac{y' Ay}{y' By} \xrightarrow{p} \begin{cases} \frac{v' P' \bar{A} P v}{v' P' \bar{B} P v} & \text{if } Bi = 0, Ai = 0, \\ + \infty & \text{if } Bi = 0, i' Ai > 0, \\ - \infty & \text{if } Bi = 0, i' Ai < 0, \\ \frac{i' Ai}{i' Bi} & \text{if } Bi \neq 0, \end{cases}$$

where \bar{A} is the $(n - 1) \times (n - 1)$ matrix obtained from A by deleting the first row and the first column, \bar{B} is similarly obtained from B , i is an $n \times 1$ vector of ones, $v \sim N(0, I_{n-1})$, P is a lower triangular $(n - 1) \times (n - 1)$ matrix such that

$$(PP')_{ij} = \min(i, j) - \frac{\psi}{(1 + \psi)^2} (1 + \delta_{ij}),$$

δ_{ij} is the Kronecker delta, and \xrightarrow{p} indicates convergence in probability.

Note 1: The only case not covered by the theorem is the case where $Bi = 0$, $i' Ai = 0$, $Ai \neq 0$. The ratio $y' Ay/y' By$ then converges to $\pm \infty$ with probability 1/2 each. In most cases of practical interest A is either positive

semidefinite (in which case $i' Ai = 0$ implies $Ai = 0$) or A can be written as $A = BSB$ for some symmetric matrix S (in which case $Bi = 0$ implies $Ai = 0$), so that the theorem applies.

Note 2: If $Bi \neq 0$, then $\Pr(y'Ay/y'By < c)$ will approach either 0 or 1 depending on the sign of $i'(A - cB)i$. This explains why the *DW* statistic in a regression without intercept can have zero limiting power.

Note 3: For $\psi = 0$ the process is AR(1) and the lower triangular matrix P takes the simple form

$$P_{ij} = \begin{cases} 0 & \text{if } i < j, \\ 1 & \text{if } i \geq j \end{cases}$$

for $i, j = 1, \dots, n - 1$. In the general ARMA (1, 1) case the structure of P is more complicated, but it can always be computed through a standard Choleski separation routine, available in NAG and elsewhere.

Note 4: The normality assumption is much stronger than necessary. All that is required is that the second moments of y are bounded.

Proof of Theorem B.1. Letting $\alpha = \psi/(1 + \psi)^2$, we have

$$\text{cov}(y_t, y_{t-s}) = E y_t y_{t-s} = \frac{\sigma^2(1 + \psi)^2}{1 - \phi^2} \gamma(s), \quad s = 0, 1, \dots,$$

where

$$\gamma(0) = 1 - 2\alpha(1 - \phi),$$

$$\gamma(1) = \phi + \alpha(1 - \phi)^2,$$

$$\gamma(s) = \phi\gamma(s - 1), \quad s \geq 2.$$

Now, let $r = \sqrt{1 - \phi^2}$. Then, for ϕ close to 1,

$$\phi = 1 - \frac{1}{2}r^2 + \mathcal{O}(r^4)$$

and hence,

$$\gamma(0) = 1 - \alpha r^2 + \mathcal{O}(r^4),$$

$$\gamma(s) = 1 - \frac{1}{2}sr^2 + \mathcal{O}(r^4), \quad s \geq 1.$$

The $n \times n$ covariance matrix Ω of y is therefore, apart from an irrelevant factor of proportionality,

$$\Omega = ii' - \frac{1}{2}r^2Q + \mathcal{O}(r^4)$$

with

$$Q = \begin{pmatrix} 2\alpha & 1 & 2 & \dots & n-2 & n-1 \\ 1 & 2\alpha & 1 & \dots & n-3 & n-2 \\ 2 & 1 & 2\alpha & \dots & n-4 & n-3 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ n-2 & n-3 & n-4 & \dots & 2\alpha & 1 \\ n-1 & n-2 & n-3 & \dots & 1 & 2\alpha \end{pmatrix}.$$

Let $\Omega = LL'$, where L is a lower triangular $n \times n$ matrix. We write L as

$$L = L_0 + rL_1 - \frac{1}{2}r^2L_2 + r^3L_3 + \mathcal{O}(r^4),$$

which implies

$$\begin{aligned} \Omega = LL' &= L_0L_0' + r(L_0L_1' + L_1L_0') - \frac{1}{2}r^2(L_0L_2' + L_2L_0' - 2L_1L_1') \\ &+ r^3(L_0L_3' + L_3L_0' - \frac{1}{2}L_1L_2' - \frac{1}{2}L_2L_1') + \mathcal{O}(r^4). \end{aligned}$$

Equating the two expansions for Ω yields the four equations

$$L_0L_0' = ii',$$

$$L_0L_1' + L_1L_0' = 0,$$

$$L_0L_2' + L_2L_0' - 2L_1L_1' = Q,$$

$$L_0L_3' + L_3L_0' - \frac{1}{2}L_1L_2' - \frac{1}{2}L_2L_1' = 0.$$

Recalling that L_0, \dots, L_3 are lower triangular, we find the following solutions:

$$L_0 = \begin{pmatrix} 1 & 0' \\ \bar{i} & 0 \end{pmatrix}, \quad L_1 = \begin{pmatrix} 0 & 0' \\ 0 & P \end{pmatrix}, \quad L_2 = \begin{pmatrix} \alpha & 0' \\ c & 0 \end{pmatrix},$$

$$L_3 = \begin{pmatrix} 0 & 0' \\ 0 & \bar{L}_3 \end{pmatrix},$$

where P is defined in the theorem, \bar{i} is an $(n - 1) \times 1$ vector of ones, c is an $(n - 1) \times 1$ vector with $c_j = j - \alpha$, and \bar{L}_3 is a lower triangular $(n - 1) \times (n - 1)$ matrix. Hence,

$$L'AL = \begin{cases} r^2 \begin{pmatrix} 0 & 0' \\ 0 & P'\bar{A}P \end{pmatrix} + \mathcal{O}(r^3) & \text{if } Ai = 0, \\ r \begin{pmatrix} 0 & i'\tilde{A}P \\ P'\tilde{A}i & 0 \end{pmatrix} + \mathcal{O}(r^2) & \text{if } Ai \neq 0, i' Ai = 0, \\ (i' Ai) \begin{pmatrix} 1 & 0' \\ 0 & 0 \end{pmatrix} + \mathcal{O}(r) & \text{if } i' Ai \neq 0, \end{cases}$$

where \tilde{A} is obtained from A by deleting its first row. For $L'BL$ we find similar expressions except that the second option can not occur since B is positive semidefinite. The result now follows from the boundedness of the second moments of y and the fact that, almost surely,

$$\frac{y' Ay}{y' By} = \frac{\tilde{v}' L' AL \tilde{v}}{\tilde{v}' L' BL \tilde{v}}$$

where $\tilde{v} \sim N(0, I_n)$. \square

Our next theorem considers the general $AR(q)$ process and tells us what happens with a ratio of quadratic forms when the q th autocorrelation parameter ϕ_q converges to 1.

Theorem B.2. Assume that the observations $y = (y_1, \dots, y_n)'$ are generated by a stationary $AR(q)$ process,

$$y_t = \phi y_{t-q} + \varepsilon_t,$$

where the ε_t are i.i.d. $N(0, \sigma^2)$. Let A be a symmetric $n \times n$ matrix and B a symmetric positive semidefinite $n \times n$ matrix. Let m be the smallest integer such that $mq \geq n$ and define the $q \times mq$ matrix

$$\bar{H}'_q = (I_q \cdot I_q \cdot \dots \cdot I_q).$$

Let H'_q be the $q \times n$ matrix containing the first n columns of \bar{H}'_q . If

$$H'_q A H_q \neq 0 \quad \text{and} \quad H'_q B H_q \neq 0,$$

then, as $\phi \rightarrow 1$,

$$\frac{y' Ay}{y' By} \xrightarrow{p} \frac{v' H'_q A H_q v}{v' H'_q B H_q v},$$

where $v \sim N(0, I_q)$.

Note 1: If $q = 1$, we obtain the special case of Theorem B.1 where $i' A i \neq 0$ and $B i \neq 0$, since $H_q = i$ in that case.

Note 2: If $q = 2$ and $A i = 0, B i = 0$ (regression with intercept), then $H_q = (a : b)$, where

$$a' = (1 \ 0 \ 1 \ 0 \ \dots), \quad b' = (0 \ 1 \ 0 \ 1 \ \dots).$$

Since $a + b = i$, we have $A(a + b) = A i = 0$ and hence $Ab = -Aa$. This leads to

$$H'_q A H_q = \begin{pmatrix} a' A a & a' A b \\ b' A a & b' A b \end{pmatrix} = (a' A a) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

and hence

$$v' H'_q A H_q v = (a' A a)(v_1 - v_2)^2.$$

Under the assumption that $A i = B i = 0$ we thus have

$$\frac{y' Ay}{y' By} \xrightarrow{p} \frac{a' A a}{a' B a} \quad \text{as } \phi \rightarrow 1,$$

which is a constant.

Proof of Theorem B.2. Let $\text{var}(y) = \sigma^2 \Omega$ and let $\Omega = LL'$, where L is lower triangular. Then, as $\phi \rightarrow 1$, $L \rightarrow L_q$ and $\Omega \rightarrow L_q L'_q$, where $L_q = (H_q : 0)$. Now write $y = \sigma L \tilde{v}$ where $\tilde{v} \sim N(0, I_n)$. Then, as $\phi \rightarrow 1$,

$$y' Ay = \sigma^2 \tilde{v}' L' A L \tilde{v} \xrightarrow{p} \sigma^2 v' H'_q A H_q v$$

and the result follows. \square

References

- Amemiya, T., 1985. *Advanced Econometrics*. Basil Blackwell, Oxford.
- Bartels, R., 1992. On the power function of the Durbin–Watson test. *Journal of Econometrics* 51, 101–112.
- Berenblut, I.I., Webb, G.I., 1973. A new test for autocorrelated errors in the linear regression model. *Journal of the Royal Statistical Society B35*, 33–50.
- Blattberg, R.C., 1973. Evaluation of the power of the Durbin–Watson statistic for non-first order serial correlation alternatives. *The Review of Economics and Statistics* 55, 508–515.
- Breusch, T.S., Pagan, A.R., 1980. The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies* 47, 239–253.
- Dufour, J.-M., King, M.L., 1991. Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or nonstationary AR(1) errors. *Journal of Econometrics* 47, 115–143.
- Farebrother, R.W., 1980. The Durbin–Watson test for serial correlation when there is no intercept in the regression. *Econometrica* 48, 1553–1563.
- Fiebig, D.G., McAleer, M., Bartels, R., 1992. Properties of ordinary least squares estimators in regression models with nonspherical disturbances. *Journal of Econometrics* 54, 321–334.
- Griffiths, W.E., Beesley, P.A.A., 1984. The small-sample properties of some preliminary test estimators in a linear model with autocorrelated errors. *Journal of Econometrics* 25, 49–61.
- Harvey, A.C., 1990. *The Econometric Analysis of Time Series*, LSE Handbooks in Economics, Phillip Allan, New York.
- Harvey, A.C., 1993. *Time Series Models*. Harvester Wheatsheaf, New York.
- Imhof, J.P., 1961. Computing the distribution of quadratic forms in normal variables. *Biometrika* 48, 419–426.
- King, M.L., 1981. The alternative Durbin–Watson test. *Journal of Econometrics* 17, 51–66.
- King, M.L., 1987. Testing for autocorrelation in linear regression models: a survey. In: King, M.L., Giles, D.E.A. (Eds.), *Specification Analysis in the Linear Model*, Essays in Honour of Donald Cochrane. Routledge and Kegan Paul, London.
- King, M.L., Evans, M.A., 1988. Locally optimal properties of the Durbin–Watson test. *Econometric Theory* 4, 509–516.
- King, M.L., Giles, D.E.A., 1984. Autocorrelation pre-testing in the linear model. *Journal of Econometrics* 25, 35–48.
- Knottnerus, P., 1985. A test strategy for discriminating between autocorrelation and misspecification in regression analysis: a critical note. *The Review of Economics and Statistics* 67, 175–177.
- Koerts, J., Abrahamse, A.P.J., 1969. *On the Theory and Application of the General Linear Model*. Rotterdam University Press, Rotterdam.
- Krämer, W., 1985. The power of the Durbin–Watson test for regressions without an intercept. *Journal of Econometrics* 28, 363–370.
- Kruskal, W., 1968. When are Gauss–Markov and least squares estimators identical? A coordinate-free approach. *Annals of Mathematical Statistics* 39, 70–75.
- Leamer, E.E., 1984. Global sensitivity results for generalized least squares estimates. *Journal of the American Statistical Association* 79, 867–870.
- Magnus, J.R., Neudecker, H., 1999. *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Revised Edition. Wiley, Chichester, New York.
- Numerical Algorithms Group (NAG), 1991. *Fortran Library Manual Mark 15*. Numerical Algorithms Group Ltd., Oxford.
- Sargan, J.D., Bhargava, A., 1983. Testing residuals from least squares regression for being generated by the Gaussian random walk. *Econometrica* 51, 153–174.
- Sharma, S.C., 1987. The effects of autocorrelation among errors on the consistency property of OLS estimator. *Journal of Statistical Computation and Simulation* 28, 43–62.

- Thursby, J.G., 1981. A test strategy for discriminating between autocorrelation and misspecification in regression analysis. *The Review of Economics and Statistics* 63, 117–123.
- Tillman, J.A., 1975. The power of the Durbin–Watson test. *Econometrica* 43, 959–974.
- Wallis, K.F., 1972. Testing for fourth-order autocorrelation in quarterly regression equations. *Econometrica* 40, 617–636.