Biometrika (1983), **70**, 1, *pp.* 1–10
Printed in Great Britain

Diagnostics for heteroscedasticity in regression

By R. DENNIS COOK and SANFORD WEISBERG
Department of Applied Statistics, University of Minnesota, St. Paul, Minnesota, U.S.A.

Summary

For the usual regression model without replication, we provide a diagnostic test for heteroscedasticity based on the score statistic. A graphical procedure to complement the score test is also presented.

Some key words: Influence; Linear model; Residual; Score test.

1. Introduction

Diagnostic methods in linear regression are used to examine the appropriateness of assumptions underlying the modelling process and to locate unusual characteristics of the data that may influence conclusions. The recent literature on diagnostics is dominated by studies of methods for the detection of influential observations. Cook & Weisberg (1982) provide a review. Diagnostics for the relevance of specific assumptions, however, have not received the same degree of attention, even though these may be of equal importance. Our purpose here is to provide appropriate diagnostic techniques to aid in an assessment of the validity of the usual assumption of homoscedasticity when little or no replication is present.

Available methods for studying this assumption include both graphical and nongraphical procedures. The usual graphical procedure consists of plotting the ordinary least squares residuals against fitted values or an explanatory variable. A megaphone shaped pattern is taken as evidence that the variance depends on the quantity plotted on the abscissa (Weisberg, 1980, Chapter 6). In § 3 we suggest several ways in which this standard graphical method may be improved, particularly in small to moderate sized data sets.

A number of tests for homoscedasticity have been proposed, some based on a specific alternative model for heteroscedasticity (Anscombe, 1961; Bickel, 1978) and others on plausible, but *ad hoc* grounds (Goldfeld & Quandt, 1965; Glejser, 1969; Harrison & McCabe, 1979; Horn, 1981). Robust tests for homoscedasticity have been proposed by Bickel (1978); see also Hammerstrom (1981) and Carroll & Ruppert (1981). While some of these tests are appropriate for diagnostic use, others are not. Anscombe's test, for example, requires the calculation of all elements of the matrix that projects onto the column space of the explanatory variables, which rules out its routine use. Bickel's test, on the other hand, depends only on the ordinary residuals, fitted values and the diagonal elements of the projection matrix, and is therefore easier to compute and suitable for use as a diagnostic method.

When heteroscedasticity occurs, the variance may often depend on the values of one or more of the explanatory variables or on additional relevant quantities such as time or spatial ordering. The model used by Anscombe and Bickel assumes that variance is a function of expected response. In § 2, we develop a model that allows for dependence of the variance on an arbitrary set of variables, and thus includes Anscombe and Bickel's model as a special case. The tests that we propose are based on the score statistic, are easily computed using standard regression software and, in the special model considered by Anscombe and Bickel, reduce to a form similar to Bickel's test.

2. TESTS CONCERNING HETEROSCEDASTICITY

2·1. *Models*

The usual linear regression model can be written in the form

$$Y = \beta_0 1 + X\beta + \varepsilon, \tag{1}$$

where Y is an $n \times 1$ vector of observable responses, X is an $n \times p$ matrix of known constants, 1 is an $n \times 1$ vector of ones, β is a $p \times 1$ vector of unknown parameters, and ε is an $n \times 1$ vector of unobservable random errors. For convenience, we assume that the augmented matrix $X_0 = (1, X)$ has rank $p+1$. We consider only models with the intercept β_0 included, but modification to remove this assumption is straightforward.

As a foundation for tests concerning the error variances, we assume that ε follows a multivariate normal distribution with mean 0 and covariance matrix $\sigma^2 W$, where W is a diagonal matrix with diagonal entries $w_1, \ldots, w_n$, with all $w_i > 0$. We next assume that the w_i follow the functional form

$$w_i = w(z_i, \lambda) \quad (i = 1, \ldots, n); \tag{2}$$

thus w_i depends on a known $q_z \times 1$ vector $z_i = \{z_{ij}\}$ and a $z_\lambda \times 1$ vector of unknown parameters λ. We further assume that w is a twice differentiable function of λ and there is a unique value λ^* of λ such that $w(z, \lambda^*) = 1$ for all z. Thus tests concerning heteroscedasticity are equivalent to tests of the hypothesis $\lambda = \lambda^*$.

For further progress, an explicit form for w must be chosen. While there are many ways in which this can be done, we will consider two specific families given by

$$w(z_i, \lambda) = \exp(\Sigma_j \lambda_j z_{ij}), \tag{3}$$

$$w(z_i, \lambda) = \prod_{j=1}^{n} z_{ij}^{\lambda_j} = \exp\left(\sum_{j=1}^{n} \lambda_j \log z_{ij}\right). \tag{4}$$

For each family, $q_z = q_\lambda = q$. Of course, (4) requires that the z_{ij} be strictly positive, while no such restriction is needed for (3). Both (3) and (4) can be imbedded in a single more general family,

$$w(z_i, \lambda) = \exp\left(\sum_{j=1}^{q} \lambda_j z_{ij}^{a_j}\right). \tag{5}$$

If, by convention, we take $z^a = \log z$ when $a = 0$, then (3) corresponds to $a_1 = \ldots = a_q = 1$ and (4) corresponds to $a_1 = \ldots = a_q = 0$.

The class of weight functions (5) requires some modification when it is desirable to constrain the variance to depend on the expected response. In this situation, we take w_i to be of the form

$$w_i = w(\lambda x_i^{\mathrm{T}} \beta), \tag{6}$$

where x_i^{T} is the ith row of X, and $q_\lambda = 1$. Further, we take $\lambda^* = 0$ so that w is constant for all x_i under the null hypothesis. For example the special case

$$w(\lambda x_i^{\mathrm{T}} \beta) = \exp(\lambda x_i^{\mathrm{T}} \beta) \tag{7}$$

was considered by Anscombe (1961). The intercept β_0 is not included in (6) or (7) because its inclusion may lead to an overparameterized model. In any event, the score test that follows is the same regardless of whether β_0 is included or not.

2·2. *Score tests*

For routine diagnostic work, it is desirable to have available a test of the hypothesis $\lambda = \lambda^*$ that can be easily constructed using standard regression software. Methods that are based on the maximum likelihood estimator of λ, for example, require special and often complicated programs, and are not well suited for this purpose. The score statistic (Cox & Hinkley, 1974, p. 324) for the hypothesis $\lambda = \lambda^*$ furnishes a suitable diagnostic test. We first describe the score test for the weight function given by (2), and later comment on the analogous approach when the variance is constrained to be a function of the expected response. In other applications, diagnostics based on the score have been proposed by Atkinson (1981, 1982), Box (1980), and by Pregibon (1981).

Let U be an n vector with elements $e_i^2/\hat{\sigma}^2$, where $e_i = (y_i - \hat{\beta}_0 - x_i^{\mathrm{T}} \hat{\beta})$, $\hat{\sigma}^2 = \Sigma e_i^2/n$ and $\hat{\beta}_0$ and $\hat{\beta}$ are maximum likelihood estimators of β_0 and β, respectively, under model (1) with $W = I$. Next, define $w'(z_i, \lambda^*)$ to be the $q_\lambda \times 1$ vector with jth element $\partial w(z_i, \lambda)/\partial \lambda_j$ evaluated at $\lambda = \lambda^*$, and let D be the $n \times q_\lambda$ matrix with ith row $[w'(z_i, \lambda^*)]^{\mathrm{T}}$. Finally, let $\bar{D} = D - 11^{\mathrm{T}} D/n$, the $n \times q_\lambda$ matrix obtained from D by subtracting column averages. Then the score test statistic S for the hypothesis $\lambda = \lambda^*$ when the weight function is given by (2) is, as shown in the Appendix,

$$S = \tfrac{1}{2} U^{\mathrm{T}} \bar{D} (\bar{D}^{\mathrm{T}} \bar{D})^{-1} \bar{D}^{\mathrm{T}} U \tag{8}$$

if we assume, of course, that $\bar{D}$ is of full rank. Computationally, S is one-half of the sum of squares for the regression of U on D in the constructed model $U = \gamma_0 1 + D\gamma + \varepsilon_u$, and thus can be easily obtained using standard regression software.

If (3) is chosen to be the weight function, then $w'(z_i, \lambda^*) = z_i$, while if (4) is used then $w'(z_i, \lambda^*)$ has jth element $\log z_{ij}$. In practice, it will often be appropriate to choose the variables included in z_i from the columns of X, and S is then obtained from the regression of U on the selected columns for weight function (3), or on their elementwise logarithms for weight function (4).

The nominal asymptotic distribution of S when $\lambda = \lambda^*$ is central chi-squared with q_λ degrees of freedom, $\chi^2(q_\lambda)$. When $q_\lambda = 1$, S can be written as

$$S = \frac{[\Sigma_i \{w'(z_i, \lambda^*) - \bar{w}'(z_i, \lambda^*)\} (e_i^2/\hat{\sigma}^2 - 1)]^2}{2 \Sigma \{w'(z_i, \lambda^*) - \bar{w}'(z_i, \lambda^*)\}^2}, \tag{9}$$

where $\bar{w}'(z_i, \lambda^*) = \Sigma w'(z_i, \lambda^*)/n$. To investigate the small-sample behaviour of (9), it is sufficient to consider the numerator, since the denominator is nonstochastic. Apart from an unimportant constant, the signed square root of the numerator can be written as

$$S' = \{Y^{\mathrm{T}}(I - V) A (I - V) Y\}/\{Y^{\mathrm{T}}(I - V) Y\} = \{\varepsilon^{\mathrm{T}}(I - V) A (I - V) \varepsilon\}/\{\varepsilon^{\mathrm{T}}(I - V) \varepsilon\}, \tag{10}$$

where $V = ((v_{ij})) = X_0 (X_0^{\mathrm{T}} X_0)^{-1} X_0^{\mathrm{T}}$ and $A = \operatorname{diag}\{w'(z_i, \lambda^*)\}$. The inner product matrices in the numerator and denominator of (10) commute and thus can be simultaneously diagonalized with an orthogonal transformation:

$$S' = \sum_{i=1}^{n-p-1} \eta_i \chi_i^2 \Big/ \sum_{i=1}^{n-p-1} \chi_i^2, \tag{11}$$

where $\chi_1^2, \ldots, \chi_{n-p-1}^2$ are independent $\chi^2(1)$ random variables and $\eta_1, \ldots, \eta_{n-p-1}$ are the, at most, $n-p-1$ nonzero eigenvalues of $(I-V)\,A(I-V)$. Durbin & Watson (1971) give an account of methods for approximating and finding the exact null distribution of (11); see also Harrison & McCabe (1979).

When the variance is constrained to depend on the expected response and the weight function is of the form $w_i = w(\lambda x_i^{\mathrm{T}} \beta)$, the score statistic S_f for the hypothesis $\lambda = 0$ is given by (9) with $w'(z_i, \lambda^*)$ replaced by

$$\left[\frac{\partial w(\lambda x_i^{\mathrm{T}} \beta)}{\partial \lambda}\right]_{\lambda=0} \propto x_i^{\mathrm{T}} \hat{\beta},$$

or equivalently by $\hat{y}_i$, the ith fitted value from model (1) with $W = I$. In contrast to the previous development, this test does not depend on the particular choice of w as long as it is of the form (6).

The relationship between S_f and Bickel's (1978) studentized version of Anscombe's (1961) test statistic can be seen by writing S_f in the form

$$S_f = \frac{\{\Sigma\,(\hat{y}_i - \tilde{y})\, e_i^2\}^2}{2\hat{\sigma}^4\, \Sigma\,(\hat{y}_i - \tilde{y})^2}, \tag{12}$$

where $\tilde{y} = \Sigma\, \hat{y}_i/n$. Bickel's test statistic is obtained by replacing $\tilde{y}$ by $y^* = \Sigma\,(1-v_{ii})\,\hat{y}_i/(n-p)$ and $2\hat{\sigma}^4$ by $\Sigma\,(e_i^2 - \hat{\sigma}^2)^2/(n-p)$.

We have carried out a small simulation to investigate the χ^2 approximation to the null distributions of the score test and Bickel's test. Five sampling situations were considered, each defined by specification of the matrix X: (i) X was set equal to the 24×10 matrix from the cloudseeding data reproduced by Cook & Weisberg (1980) to represent a problem with widely varying diagonals of V; (ii) for a situation with all v_{ii} nearly equal, X was set equal to a 24×10 matrix with all entries generated as pseudorandom standard normal deviates; (iii) X was set equal to the first three columns of the matrix generated in (2); (iv) X is a 50×10 matrix obtained as in (ii); (v) X is a 50×3 matrix consisting of the first 3 columns of the matrix used in (iv). For each replication a pseudorandom $n \times 1$ vector of independent, standard normal deviates was generated and this vector was taken to be the response. Uniform pseudorandom deviates were produced on a CDC Cyber 720 computer by a congruential generator with multiplier 5^{17} and modulus 2^{48}. Normal deviates were produced from the uniform stream by the method of Marsaglia & Bray (1964).

For each of the 999 replications, Bickel's test, S_f, and three versions of S were computed. The different score tests correspond to alternative choices for Z. In the first version, Z is taken to be the full X matrix for situation (i), the X matrix for situation (ii) is used for both (ii) and (iii), and the X matrix for situation (iv) is used for both situations (iv) and (v). These score tests have an asymptotic $\chi^2(10)$ distribution. The second version uses the first three columns of X for all situations, and has an asymptotic $\chi^2(3)$ distribution. Finally, Z was taken to be the first column of X, leading to a $\chi^2(1)$ distribution. Weight function (3) was used for S and weight function (7) was used for S_f and Bickel's test.

Table 1 gives the 0·90, 0·95 and 0·975 points of the sample distributions. The nominal values from the appropriate chi-squared distributions are also given. The results shown are representative of other percentage points.

Table 1. *Simulated percentage points from the small-sample null distribution of the score statistic*

Situation	Level	Score-10	Score-3	Score-1	S_f	Bickel
(i)	0·90	16·48	5·01	2·54	2·66	1·43
	0·95	19·48	6·19	3·01	3·63	1·98
	0·975	22·92	7·93	3·36	4·54	2·37
(ii)	0·90	14·52	5·04	2·59	2·20	1·42
	0·95	17·66	6·43	3·50	3·32	2·04
	0·975	19·42	8·31	4·42	4·42	2·66
(iii)	0·90	15·20	4·90	2·45	2·31	2·27
	0·95	17·73	6·16	3·48	3·15	3·24
	0·975	20·63	7·51	4·46	3·94	3·76
(iv)	0·90	15·11	6·13	2·38	2·23	1·90
	0·95	17·97	7·61	3·36	3·46	2·80
	0·975	20·92	9·16	4·28	4·54	3·59
(v)	0·90	15·83	5·85	2·63	2·55	2·41
	0·95	19·47	7·65	3·45	3·77	3·49
	0·975	22·40	9·41	4·91	4·70	4·37
χ^2	0·90	15·99	6·25	2·71	2·71	2·71
	0·95	18·31	7·82	3·84	3·84	3·84
	0·975	20·55	9·36	5·02	5·02	5·02

Generally, use of the chi-squared approximation leads to a conservative test. For diagnostic purposes, the approximation appears to be adequate. Bickel's test, however, does not fare so well, as it seems to be much too conservative if $n/(n-p)$ is far from 1. Empirically, $n/(n-p)$ times Bickel's test is closely approximated by the chi-squared percentage points. The correlation between S_f and Bickel's test is typically very high: in situation (v), the observed correlation is 0·94.

3. Graphical methods

The score tests developed in the previous section are based on the often unwarranted assumption that, except for the possibility of heteroscedasticity, the usual normal theory regression model is appropriate. Most residual based tests for specific departures from the standard model are sensitive to several alternatives. Outliers, for example, will affect all residual based tests, including those described here. Graphical procedures provide a degree of robustness by helping the investigator distinguish between various alternatives, and thus are an important part of the diagnostic phase of any analysis.

The graphical version of the score statistic with $q_z = q_\lambda = 1$ given by (9) is simply a plot of $e_i^2/\hat{\sigma}^2$ versus $w'(z_i, \lambda^*)$. When the weight function is $w(z_i, \lambda) = \exp(\lambda z_i)$ and $z_i = x_{ij}$ for some fixed j, this procedure is similar to the usual practice of plotting $e_i/\hat{\sigma}$ against the jth column of X. On the other hand, if $w(x_{ij}, \lambda) = x_{ij}^{\lambda}$, the implied plot is of $e_i^2/\hat{\sigma}^2$ versus the log transform of elements of the jth column of X. We have found both plots useful in practice. When a weight function of the form $w(\lambda x_i^{\mathrm{T}} \beta)$ is used, a graphical version of S_f is a plot of $e_i^2/\hat{\sigma}^2$ versus the fitted values, $\hat{y}_i$.

With small to moderate sample sizes, the usual plots using $e_i/\hat{\sigma}$ are often sparse and difficult to interpret, particularly when the positive and negative residuals do not appear to exhibit the same general pattern. This difficulty is at least partially removed by plotting $e_i^2/\hat{\sigma}^2$ rather than $e_i/\hat{\sigma}$, and thus visually doubling the sample size. This is accomplished without loss of information since, as implied by the score tests, the signs of the residuals are unimportant for the study of heteroscedasticity. Nonconstant variance will be reflected by a wedge-shaped pattern in a plot using $e_i^2/\hat{\sigma}^2$.

Depending on the structure of V, further improvement of the standard graphical methods is possible. It is easily verified that $E(e) = 0$ and

$$\operatorname{var}(e_i) = \sigma^2\{(1-v_{ii})^2 w_i + \sum_{k \neq i} w_k v_{ik}^2\} \quad (i = 1, \ldots, n). \tag{13}$$

Thus, even if $W = I$, a suggestive pattern may be seen if, for example, w_i is chosen to be proportional to $(1-v_{ii})$. To correct for this possibility, we suggest the replacement of e_i by $b_i = e_i/(1-v_{ii})^{\frac{1}{2}}$. Clearly

$$E(b_i^2) = \sigma^2\{(1-v_{ii}) w_i + \sum_{k \neq i} w_k v_{ik}^2/(1-v_{ii})\} \tag{14}$$

and $W = I$ now implies that $E(b_i^2)$ is constant, since V is idempotent.

Next, suppose we expand $w(z_i, \lambda)$ in a linear Taylor series about $\lambda = \lambda^*$,

$$w(z_i, \lambda) \simeq 1 + (\lambda - \lambda^*)\, w'(z_i, \lambda^*). \tag{15}$$

Substitution of (15) into (14) gives the approximate relationship

$$\frac{E(b_i^2)}{\sigma^2} \simeq 1 + (\lambda-\lambda^*)\,\{(1-v_{ii})\, w'(z_i, \lambda^*) + \sum_{k \neq i} w'(z_k, \lambda^*)\, v_{ik}^2/(1-v_{ii})\}. \tag{16}$$

This last form implies that a plot of b_i^2 against the known quantity in braces in (16) will have nonzero slope if heteroscedasticity is present. Each b_i^2/σ^2 is distributed as a multiple of a $\chi^2(1)$ random variable, the multiplier being approximately equal to the right-hand side of (16). Thus this plot will be wedge-shaped rather than clustered about the line $E(b_i^2)/\sigma^2$. Finally, since the second term in square brackets in (16) requires computation of the v_{ij}, a further approximation is desirable for routine application. The sum on the right-hand side of (16) will be unimportant if all the v_{ij}^2 are small; Cook & Weisberg (1982, pp. 11 14, 211–13) discuss the magnitudes of the v_{ij}. If this term is ignored, the suggested abscissa for the plot is $(1-v_{ii})\, w'(z_i, \lambda^*)$. We have found little difference between this latter choice and using the entire term in square brackets in (16). In short we suggest plotting $b_i^2 = e_i^2/(1-v_{ii})$, or equivalently the squared studentized residuals $r_i^2 = b_i^2/s^2$, with $s^2 = n\hat{\sigma}^2/(n-p-1)$, against $(1-v_{ii})\, w'(z_i, \lambda^*)$, where $w'(z_i, \lambda^*)$ is z_i or $\log z_i$, for models (3) and (4), respectively. If the variance is suspected to be a function of expected response, we suggest plotting r_i^2 versus $(1-v_{ii})\, \hat{y}_i$.

4. ILLUSTRATIONS

4·1. *Cherry trees*

Ryan, Joiner & Ryan (1976, p. 278) present data on 31 black cherry trees with Y, tree volume, to be predicted from H, tree height, and D, tree diameter. Previous analyses (Atkinson, 1982; Cook & Weisberg, 1982, §2·4) suggest that the model $Y^{1/3} = \beta_0 + \beta_1 H + \beta_2 D + \varepsilon$ is reasonable. Using this model, we now consider the possibility that the variances are nonconstant.

The usual plot of e_i versus fitted values given in Fig. 1 shows little evidence of heteroscedasticity. The score tests for both weight functions (3) and (4) are given in

Table 2(a). The value $S_f = 0{\cdot}87$ agrees with our visual impression of Fig. 1. However, the score statistics for $z_i = H_i$ under both weight functions approach the 0·95 point of $\chi^2(1)$, which suggests that the variance may be a function of H. The plot of r_i^2 versus $(1-v_{ii})H$, given as Fig. 2 does display an obvious wedge shape, suggesting that variance is increasing with H. This example illustrates that the usual practice of looking for heteroscedasticity as a function of the expected response is not always sufficient.

Table 2. *Score tests* (a) *tree data;* (b) *gas vapour data*

(a) Tree data

Weight function	Z	Score	d.f.
(3)	D	0·47	1
(3)	H	3·24	1
(3)	D, H	3·32	2
(4)	D	0·83	1
(4)	H	3·23	1
(4)	D, H	3·23	2
(6)	$\hat{y}$	0·87	1

(b) Gas vapour data

Weight function	Z	Score	d.f.
(3)	X_1, X_2, X_3, X_4	10·299	4
(3)	X_1, X_4	9·283	2
(3)	X_1	2·791	1
(3)	X_4	0·010	1
(6)	$\hat{y}$	0·000	1

4·2. *Gas vapours*

Weisberg (1980, Table 6.7) presents a set of experimental data relating Y, quantity of hydrocarbons recovered, in grams, to 4 predictors, X_1, initial tank temperature, in degrees Fahrenheit; X_2, temperature of gasoline, °F; X_3, initial vapour pressure, psi; and X_4, vapour pressure of dispensed gasoline, psi, for a series of $n = 32$ fillings of a tank with gasoline. The data were collected to study a device for capturing emitted hydrocarbons. For our purposes, we study the linear regression model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon,$$

and examine the assumption of homoscedasticity.

The fitted regression has $R^2 = 0{\cdot}93$ a usual, though often irrelevant indicator of a good fit. The scatter plot of r_i versus $\hat{y}_i$ is shown in Fig. 3. Although this plot is not entirely satisfactory, it is certainly not suggestive of nonconstant variance. We entertained the possibility that transformations may improve the model; Atkinson's (1973) score test for a transformation of the response in the power family has value $t_D = 0{\cdot}21$, so that no transformation of the response is suggested. Methods for transforming the X's (Cook & Weisberg, 1982, § 2·4·4) indicate only a possible need to transform X_1: however, since the range of values for X_1 is narrow, an appropriate transformation is poorly determined. We choose not to transform the explanatory variables. Influence analysis (Cook & Weisberg, 1982, Chapter 3) finds no unduly influential cases.

We next consider heteroscedasticity. The score tests for weight functions (3) and (6) are shown in Table 2(b); results are similar for weight function (4). The score statistic for (6) is very small. The score statistic with $Z = (X_1, X_2, X_3, X_4)$, however, is large, giving definite evidence of heteroscedasticity. Further investigation indicates that the variance is a function of both X_1 and X_4. Graphical support for these conclusions is given in Figs 4–6. Figures 4 and 5 are plots of r_i^2 versus $(1-v_{ii})$ times $\hat{y}$ and X_1, respectively. The wedge shape is absent in each of these plots; the plot of r_i^2 versus $(1-v_{ii})X_4$ is similar to Fig. 5. Figure 6 is a plot of r_i^2 versus $(1-v_{ii})g_i$, where the $g_i = 0{\cdot}778 + 0{\cdot}110X_1 - 1{\cdot}432X_4$ are the fitted values from the regression of $e_i^2/\hat{\sigma}$ on X_1, X_4; the coefficients of X_1 and X_4

Fig. 1

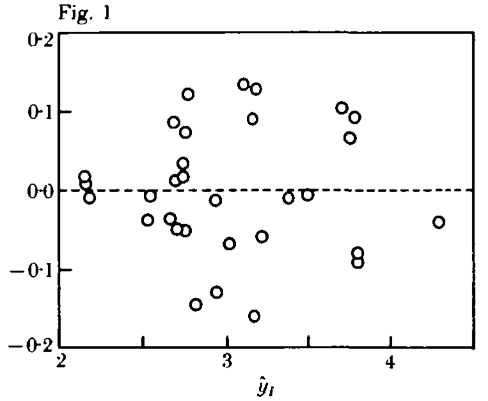

Fig. 2.

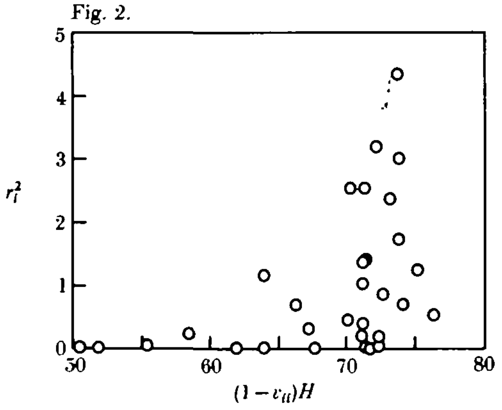

Fig. 3

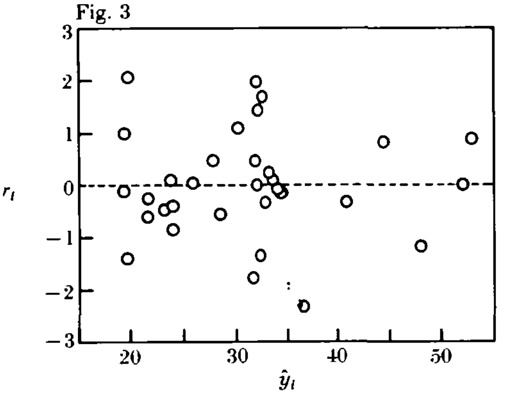

Fig. 4

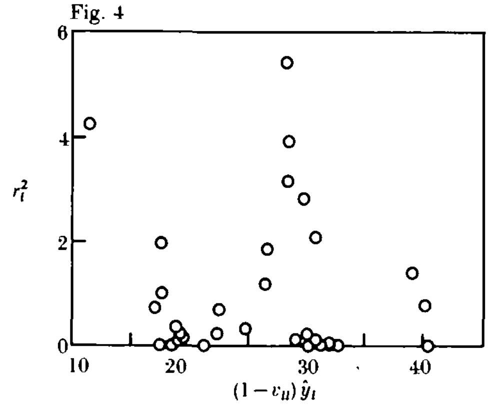

Fig. 5

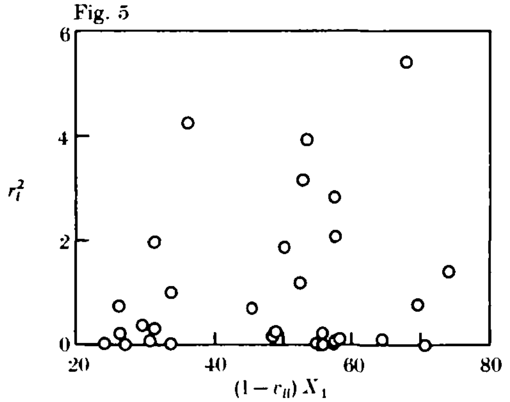

Fig. 6

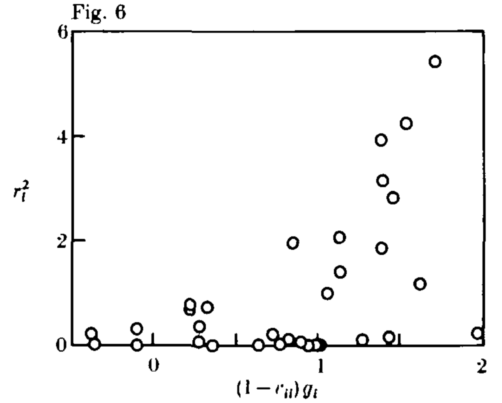

FIG. 1. e_i versus fitted values; tree data. FIG. 2. r_i^2 versus $(1-v_{ii})H$; tree data. FIG. 3. r_i versus fitted values: gas vapour data. FIG. 4. r_i^2 versus $(1-v_{ii})\hat{y}_i$; gas vapour data. FIG. 5. r_i^2 versus $(1-v_{ii})X_1$; gas vapour data. FIG. 6. r_i^2 versus $(1-v_{ii})g_i$, where the g_i are the fitted values from the regression of $e_i^2/\sigma\hat{\sigma}^2$ on X_1 and X_4: gas vapour data.

are quick estimates of the corresponding λ's. In this plot, the wedge shape is clear and may be informally viewed as the graphical counterpart of the score statistic.

At first glance, the finding that the variance is an increasing function of the g_i may seem unusual. Geometrically, however, this finding is not unreasonable. The score tests suggest that the residual variance is monotonic in some direction in the observation space. Usually, only certain special directions, such as those determined by the columns of X and the fitted values, are considered, but we have no reason to limit ourselves to these directions; see Cook & Weisberg (1982, §2.3.1) for more discussion. The direction determined by the g_i is an empirical estimate of the direction in which the variance is increasing.

5. Comments

The methods for diagnosing heteroscedasticity proposed here can be easily carried out with only minor or no modification of existing software for linear regression. We view the two procedures of computing the score test and the graphical method as complementary, and recommend the use of both. The use of the graphical methods alone can be misleading if the density of the plotted points along the x axis is uneven, since areas of higher density will tend on the average to have a greater spread in the y direction even if heteroscedasticity is not present. The score statistic serves to calibrate the plot. Similarly, the score test alone can be an inadequate indicator in the presence of outliers or influential cases. The graph can confirm the indications from the test.

Although we have presented our procedure in the context of the linear model, it is clear that the method will generalize to other regression situations. An analogous score test can be derived for other functional forms or for other error distributions, such as those in the family of generalized linear models (Nelder & Wedderburn, 1972). Approximate graphical methods similar to the one proposed here can also be derived for these models in much the same way that linear model influence methods are applied to the larger class (Cook & Weisberg, 1982, Chapter 5).

The method described in this paper is designed to aid the analyst in finding heteroscedasticity. The problem of what to do when it is found is a topic of current research: see, for example, Box & Hill (1974) and Jobson & Fuller (1980).

We are grateful to Christopher Bingham for his useful comments on an earlier draft of this paper.

Appendix

Derivation of equation (8)

Let $L(\beta, \sigma^2, \lambda)$ denote the log likelihood function for model (1) with the weight function given by (2), and let $L_\lambda(\lambda^* \mid \beta, \sigma^2)$ be the $q_\lambda \times 1$ score vector with jth component

$$\left[\partial L(\beta, \sigma^2, \lambda)/\partial \lambda_j\right]_{\lambda = \lambda^*}.$$

Also, let

$$J(\beta, \sigma^2, \lambda) = \begin{bmatrix} A & B^{\mathrm{T}} \\ B & C \end{bmatrix}$$

be the expected information matrix partitioned so that A corresponds to second partial derivatives with respect to (β, σ^2) and C corresponds to the second partial derivatives with respect to elements of λ. Denote the inverse of this matrix by

$$\{J(\beta, \sigma^2, \lambda)\}^{-1} = \begin{bmatrix} A_* & B_*^{\mathrm{T}} \\ B_* & C_* \end{bmatrix}.$$

Then the score statistic for hypothesis $\lambda = \lambda^*$ is $S = \{L_\lambda(\lambda^* \mid \hat{\beta}, \hat{\sigma}^2)\}^T \hat{C}_* \{L_\lambda(\lambda^* \mid \hat{\beta}, \hat{\sigma}^2)\}$, where $\hat{\beta}$ and $\hat{\sigma}^2$ are the maximum likelihood estimates of β and σ^2 when $\lambda = \lambda^*$ and $\hat{C}_* = C_*(\hat{\beta}, \hat{\sigma}^2, \lambda^*)$. Thus, to evaluate S we need to obtain the log likelihood, the score vector, and $\hat{C}_*$. The log likelihood function is

$$L = -\tfrac{1}{2}n \log(2\pi) - \tfrac{1}{2}n \log \sigma^2 - \tfrac{1}{2}\log |W| - \frac{1}{2\sigma^2}(Y - X\beta)^T W^{-1}(Y - X\beta).$$

Differentiating with respect to λ, we obtain $L_\lambda(\lambda^* \mid \hat{\beta}, \hat{\sigma}^2) = \tfrac{1}{2}D^T(U - 1)$, where D and U are defined near (8) and 1 is an $n \times 1$ vector of ones.

For the submatrices of $J(\hat{\beta}, \hat{\sigma}^2, \lambda^*)$ we obtain

$$A = \begin{bmatrix} X^T X/\hat{\sigma}^2 & 0 \\ 0^T & \tfrac{1}{2}n/\hat{\sigma}^4 \end{bmatrix}, \quad C = \tfrac{1}{2}D^T D, \quad B = (0, \tfrac{1}{2}D^T 1/\hat{\sigma}^2).$$

Using the usual relations to calculate the inverse of a partitioned matrix, we find that $\hat{C}_* = 2(\bar{D}^T \bar{D})^{-1}$, and thus $S = \tfrac{1}{2}(U - 1)^T D(\bar{D}^T \bar{D})^{-1} D^T(U - 1) = \tfrac{1}{2}U^T \bar{D}(\bar{D}^T \bar{D})^{-1} \bar{D}^T U$.

REFERENCES

ANSCOMBE, F. (1961). Examination of residuals. *Proc. 4th Berkeley Symp.* **1**, 1-36.

ATKINSON, A. C. (1973). Testing transformations to normality. *J. R. Statist. Soc.* B **35**, 473-9.

ATKINSON, A. C. (1981). Two graphical displays for outlying and influential observations in regression. *Biometrika* **68**, 13-20.

ATKINSON, A. C. (1982). Regression diagnostics, transformations and constructed variables (with discussion). *J. R. Statist. Soc.* B **44**, 1-35.

BICKEL, P. (1978). Using residuals robustly I: Tests for heteroscedasticity, nonlinearity. *Ann. Statist.* **6**, 266-91.

BOX, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness (with discussion). *J. R. Statist. Soc.* A **143**, 383-430.

BOX, G. E. P. & HILL, W. J. (1974). Correcting inhomogeneity of variances with power transformation weighting. *Technometrics* **16**, 385-9.

CARROLL, R. J. & RUPPERT, D. (1981). On robust tests for heteroscedasticity. *Ann. Statist.* **9**, 205-9.

COOK, R. D. & WEISBERG, S. (1980). Characterizations of an empirical influence function for detecting influential cases in regression. *Technometrics* **22**, 495-508.

COOK, R. D. & WEISBERG, S. (1982). *Residuals and Influence in Regression.* London: Chapman & Hall.

COX, D. R. & HINKLEY, D. V. (1974). *Theoretical Statistics.* London: Chapman & Hall.

DURBIN, J. & WATSON, G. S. (1971). Testing for serial correlation in least squares regression, III. *Biometrika* **58**, 1-19.

GLEJSER, H. (1969). A new test for heteroscedasticity. *J. Am. Statist. Assoc.* **64**, 316-23.

GOLDFELD, S. M. & QUANDT, R. E. (1965). Some tests for heteroscedasticity. *J. Am. Statist. Assoc.* **60**, 539-47.

HAMMERSTROM, T. (1981). Asymptotically optimal tests for heteroscedasticity in the general linear model. *Ann. Statist.* **9**, 368-80.

HARRISON, M. J. & McCABE, B. P. M. (1979). A test for heteroscedasticity based on ordinary least squares residuals. *J. Am. Statist. Assoc.* **74**, 494-500.

HORN, P. (1981). Heteroscedasticity of residuals: A non-parametric alternative to the Goldfeld-Quandt peak test. *Comm. Statist.* A **10**, 795-808.

JOBSON, J. D. & FULLER, W. A. (1980). Least squares estimation when the covariance matrix and parameter vector are functionally related. *J. Am. Statist. Assoc.* **75**, 176-81.

MARSAGLIA, G. & BRAY, T. A. (1964). A convenient method for generating normal variables. *S.I.A.M. Rev.* **6**, 260-4.

NELDER, J. A. & WEDDERBURN, R. W. M. (1972). Generalized linear models. *J. R. Statist. Soc.* A **35**, 370-84.

PREGIBON, D. (1981). Logistic regression diagnostics. *Ann. Statist.* **9**, 705-24.

RYAN, T., JOINER, B. & RYAN, B. (1976). *Minitab Student Handbook.* North Scituate, Mass: Duxbury.

WEISBERG, S. (1980). *Applied Linear Regression.* New York: Wiley.

[*Received May* 1982. *Revised August* 1982]