Appl. Statist. (1994)
43, *No.* 2, *pp.* 357–370

Use of Modified Profile Likelihood for Improved Tests of Constancy of Variance in Regression

By JEFFREY S. SIMONOFF†

New York University, USA

and CHIH-LING TSAI

University of California, Davis, USA

[Received July 1991. Final revision March 1993]

SUMMARY

Non-constant variance (heteroscedasticity) is common in regression data, and many tests have been proposed for detecting it. This paper shows that the properties of likelihood-based tests can be improved by using the modified profile likelihood of Cox and Reid. A modified likelihood ratio test and modified score tests are derived, and both theoretical and intuitive justifications are given for the improved properties of the tests. The results of a Monte Carlo study are mentioned, which show that, whereas the ordinary likelihood ratio test can be very anticonservative, the modified test holds its null size well and is more powerful than the other tests. For non-normal error distributions, Studentized tests hold their size well (without being overconservative), even for long-tailed error distributions. Under short-tailed error distributions, likelihood ratio or Studentized score tests are most powerful, depending on the degree of heteroscedasticity. The modified versions of the score tests consistently outperform the unmodified versions. The use of these tests is demonstrated through analysis of data on the volatility of stock prices.

Keywords: Heteroscedasticity; Profile likelihood ratio test; Risk of a security; Score test; Studentized score test

1. Introduction

A central concept in modern portfolio theory is the systematic risk of a security, i.e. the relative volatility of the return of the security compared with that of the market. The regression coefficient of the return of the security on the return of the market, commonly called β, is an index of this systematic risk. A β-value greater than 1 indicates a security that is more volatile than the market, whereas a value less than 1 is consistent with a security that is less volatile than the market.

A typical approach to estimating β is as follows. First, time series of historical returns of the security (R_i, $i = 1, \ldots, n$, where n might represent 60 months), some market index W_i (such as the value-weighted index of all stocks listed on the New York and American Stock Exchanges) and the return of a 'riskless' asset

†*Address for correspondence*: Department of Statistics and Operations Research, Leonard N. Stern School of Business, New York University, 44 West 4th Street, Room 8-54, New York, NY 10012-1126, USA.

 0035-9254/94/43357

Z_i, such as Treasury bills, are obtained. Then, the following regression model is fitted:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \tag{1}$$

where $Y_i = R_i - Z_i$, $X_i = W_i - Z_i$, ϵ_i is an error term and $\hat{\beta}_1$ is the estimate of β.

Clearly ordinary least squares (OLS) can be used to provide estimates of β. Unfortunately, it is a well-established empirical fact that the error term in model (1) often (but not always) displays heteroscedasticity; specifically, the variability of the error term increases with the value of the predictor X in model (1). There is also some evidence that the value of β for a given security changes over time, which can also manifest itself as heteroscedasticity in model (1). See Fisher and Kamin (1985) for a discussion of these points.

Consider for example, the data given in Table 1. It represents the monthly excess returns over the riskless rate for the market (X) and for the Acme Cleveland Corporation (Y) for the period January 1986–December 1990. Fig. 1 gives a scatter-plot of the company's return *versus* the market return for these data, except that case 22, corresponding to October 1987, has been removed as a clear outlier (reflec-

TABLE 1
Monthly excess returns over riskless rate, January 1986–December 1990: market return, and return for Acme Cleveland Corporation

Month	*Market return*	*Acme return*	*Month*	*Market return*	*Acme return*
January 1986	−0.061134	0.030160	July 1988	−0.061718	−0.110515
February 1986	0.008220	−0.165457	August 1988	−0.101710	−0.168769
March 1986	−0.007381	0.080137	September 1988	−0.032705	−0.135585
April 1986	−0.067561	−0.109917	October 1988	−0.045334	−0.084077
May 1986	−0.006238	−0.114853	November 1988	−0.079288	−0.164550
June 1986	−0.044251	−0.099254	December 1988	−0.036233	0.150269
July 1986	−0.112070	−0.226846	January 1989	−0.011494	−0.015672
August 1986	0.030226	0.073445	February 1989	−0.093729	−0.037860
September 1986	−0.129556	−0.143064	March 1989	−0.065215	−0.074712
October 1986	0.001319	0.034776	April 1989	−0.037113	−0.108530
November 1986	−0.033679	−0.063375	May 1989	−0.044399	−0.036769
December 1986	−0.072795	−0.058735	June 1989	−0.084412	0.023912
January 1987	0.073396	0.050214	July 1989	0.003444	−0.078430
February 1987	−0.011618	0.111165	August 1989	−0.056760	−0.132199
March 1987	−0.026852	−0.127492	September 1989	−0.078970	−0.110141
April 1987	−0.040356	0.054522	October 1989	−0.105367	−0.126302
May 1987	−0.047539	−0.072918	November 1989	−0.038634	−0.095730
June 1987	−0.001732	−0.058979	December 1989	−0.043261	0.065740
July 1987	−0.008899	0.236147	January 1990	−0.139773	−0.120056
August 1987	−0.020837	−0.094778	February 1990	−0.059094	−0.085205
September 1987	−0.084811	−0.135669	March 1990	−0.057736	−0.130433
October 1987	−0.262077	−0.284796	April 1990	−0.102524	−0.116728
November 1987	−0.110167	−0.171494	May 1990	0.023881	−0.078039
December 1987	0.034955	0.242616	June 1990	−0.079116	−0.170322
January 1988	0.012688	−0.063518	July 1990	−0.078965	−0.077727
February 1988	−0.002170	−0.117677	August 1990	−0.161359	−0.277035
March 1988	−0.073462	0.201674	September 1990	−0.119376	−0.207595
April 1988	−0.043419	−0.147728	October 1990	−0.076008	−0.070515
May 1988	−0.054730	−0.170885	November 1990	−0.006444	−0.046274
June 1988	−0.011755	−0.014893	December 1990	−0.026401	−0.190834

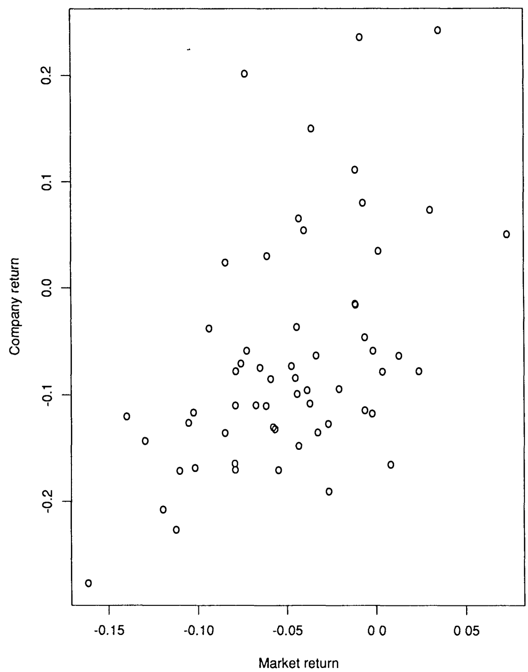

Fig. 1. Scatterplot of monthly excess returns over riskless rate for Acme Cleveland Corporation *versus* excess market returns, with one unusual case removed

ting that the market lost approximately a third of its value over a two-week period during that month).

There is apparently a linear relationship here, with the OLS estimates being $Y = -0.00951 + 1.1724X$ (the F-test for overall significance of the regression is $F = 18.9$, $p = 0.0001$). However, a plot of squared standardized residuals *versus* the market return (Fig. 2) indicates potential (but not definite) heteroscedasticity, along with a possible outlier. This data set is analysed further in Section 3.

Heteroscedasticity would of course imply the need to investigate the use of weighted least squares (WLS) to estimate β (which requires estimating $\sigma_i^2 = \text{var}(\epsilon_i)$ for all i), but the fact that not all securities exhibit such behaviour necessitates the application of a test for heteroscedasticity first. The many tests for heteroscedasticity that have been proposed fall into three broad groups: non-constructive tests, where no specific model for σ_i^2 is hypothesized or estimated, non-likelihood-based constructive tests, where a specific model for σ_i^2 is hypothesized, but use of the likelihood function is not the basis of the test, and likelihood-based constructive tests.

It is tests of the third type that are the focus of this paper. Specifically, it will be demonstrated that both the null and the power properties of likelihood-based constructive tests can be improved by using the modified profile likelihood method of Cox and Reid (1987). In the next section the model being used here is given, and the method of Cox and Reid (1987) is discussed, from both a theoretical and an intuitive point of view. This leads to modified versions of likelihood ratio and score

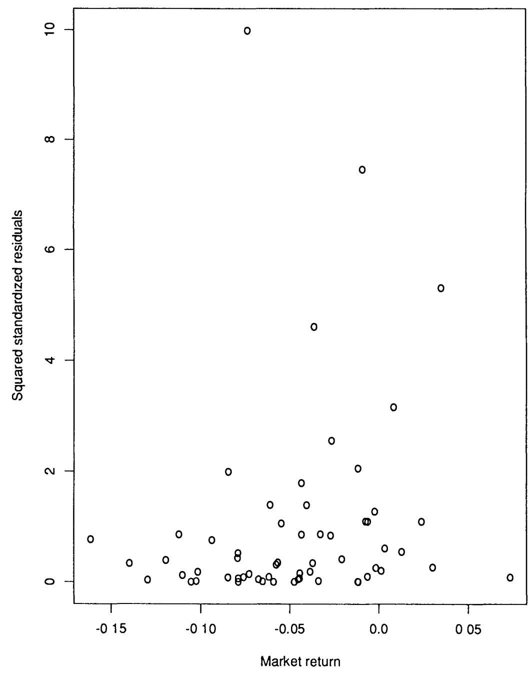

Fig. 2. Squared standardized residuals *versus* market return for the stock volatility regression model, with one unusual case removed

tests. In Section 3 the properties of the tests are examined through discussion of Monte Carlo simulations and application to real data (the previously mentioned stock volatility data). Discussion of the results and of future work comprise Section 4, and two appendixes conclude the paper.

2. Modified Profile Likelihood and Tests for Heteroscedasticity

The model that serves as the basis for analysis here takes the form

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 W), \tag{2}$$

where $\mathbf{y}$ and $\boldsymbol{\epsilon}$ are $n \times 1$, X is $n \times p$ and $\boldsymbol{\beta}$ is $p \times 1$. W is an $n \times n$ diagonal matrix with ith entry $w(z_i, \boldsymbol{\delta})$, where $\mathbf{z}_i'$ is the ith row of the $n \times q$ matrix Z of variance predicting variables and $\boldsymbol{\delta}$ is $q \times 1$. The null hypothesis of homoscedasticity is defined by $\boldsymbol{\delta} = \boldsymbol{\delta}_0$ by requiring that the variance function w satisfies $w(z_i, \boldsymbol{\delta}_0) = 1$ for $i = 1, \ldots, n$.

A natural approach to this problem is through the log-likelihood

$$l(\mathbf{y}; \boldsymbol{\delta}, \sigma^2, \boldsymbol{\beta}) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^{n}\log w(z_i, \boldsymbol{\delta}) - \frac{(\mathbf{y} - X\boldsymbol{\beta})' W^{-1}(\mathbf{y} - X\boldsymbol{\beta})}{2\sigma^2} \tag{3}$$

(ignoring constants). It is easy to demonstrate that the maximum likelihood estimate of $\boldsymbol{\delta}$ can be obtained by maximizing the profile log-likelihood

$$l_{\mathrm{p}}(\mathbf{y}; \boldsymbol{\delta}) = l(\mathbf{y}; \boldsymbol{\delta}, \hat{\sigma}_\delta^2, \hat{\boldsymbol{\beta}}_\delta),$$

where $\hat{\boldsymbol{\beta}}_\delta = (X'W^{-1}X)^{-1}X'W^{-1}\mathbf{y}$ and $\hat{\sigma}_\delta^2 = (\mathbf{y} - X\hat{\boldsymbol{\beta}}_\delta)'W^{-1}(\mathbf{y} - X\hat{\boldsymbol{\beta}}_\delta)/n$. Cox and Reid (1987) showed, however, that inference based on this profile likelihood can be problematic owing to lack of orthogonality between the parameter of interest and a nuisance parameter. They proposed a modification of the profile likelihood such that the nuisance parameter is orthogonal to the parameter of interest, which results in better behaviour; Cruddas *et al.* (1989), Davison (1988), Godambe (1991) and Hinkley (1989) examined the specific problems of first-order autoregressive models, generalized linear models, semiparametric models and transformed linear models respectively. Cox (1988) provides a survey of this, and other, issues in the application of asymptotic inference.

The modified profile likelihood for model (2) is derived in Appendix A. The key idea is to derive transformed parameter estimates with asymptotic covariances that are 0. Consider log-likelihood (3), and assume that $q = 1$ (i.e. δ is a scalar). The expected Fisher information for the vector $\boldsymbol{\theta} = (\delta, \sigma^2, \boldsymbol{\beta}')'$ has the form

$$I = E\left(-\frac{\partial^2 l}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'}\right) = \begin{pmatrix} \frac{1}{2}\sum\left(\frac{\dot{w}_j}{w_j}\right)^2 & \frac{1}{2\sigma^2}\sum\frac{\dot{w}_j}{w_j} & 0 \\ \frac{1}{2\sigma^2}\sum\frac{\dot{w}_j}{w_j} & \frac{n}{2\sigma^4} & 0 \\ 0 & 0 & \frac{X'W^{-1}X}{\sigma^2} \end{pmatrix},$$

using the notation $w_j = w(z_j, \delta)$ and $\dot{w}_j = \mathrm{d}w(z_j, \delta)/\mathrm{d}\delta$. The parameters δ and $\boldsymbol{\beta}$ are already orthogonal (their joint entry in the information matrix is 0), so the modified profile likelihood need only to establish the orthogonality transformation $\boldsymbol{\theta} = (\delta, \sigma^2, \boldsymbol{\beta}')' \to \boldsymbol{\theta}_m = (\delta, \gamma, \boldsymbol{\beta}')'$ with $E(-\partial^2 l/\partial\delta\,\partial\gamma) = 0$. Cox and Reid (1987) show in their equation (4) that the transformation is established by solving

$$i_{\sigma^2\sigma^2}\frac{\partial\sigma^2}{\partial\delta} = -i_{\delta\sigma^2},$$

where the i-values comes from the expected information matrix given above, or

$$\frac{n}{2\sigma^4}\frac{\partial\sigma^2}{\partial\delta} = -\frac{1}{2\sigma^2}\sum\frac{\dot{w}_j}{w_j},$$

which has solution

$$\sigma^2 = \gamma\bigg/\left(\prod_j w_j\right)^{1/n}.$$

A closer examination of $\hat{\sigma}^2$ and $\hat{\gamma}$ illustrates why the transformation is effective here. In the case $w(z_i, \delta) = \exp(z_i\delta)$,

$$\hat{\sigma}_\delta^2 = \frac{1}{n}\sum_{i=1}^n \frac{\hat{e}_i^2}{\exp(z_i\hat{\delta})}, \tag{4a}$$

where $\hat{e}_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}_\delta$, whereas

$$\hat{\gamma}_\delta = \frac{1}{n}\sum_{i=1}^n \frac{\hat{e}_i^2}{\exp\{(z_i - \bar{z})\hat{\delta}\}}. \tag{4b}$$

A comparison of the derivatives of these statistics with respect to $\hat{\delta}$ illustrates their sensitivity to changes in $\hat{\delta}$, and hence correlations. In this case,

$$\frac{\mathrm{d}\hat{\sigma}^2}{\mathrm{d}\hat{\delta}} = -\frac{1}{n}\sum_{i=1}^{n} \frac{\hat{e}_i^2 z_i}{\exp(z_i\hat{\delta})}, \tag{5}$$

whereas

$$\frac{\mathrm{d}\hat{\gamma}}{\mathrm{d}\hat{\delta}} = -\frac{1}{n}\sum_{i=1}^{n} \frac{\hat{e}_i^2(z_i-\bar{z})}{\exp\{(z_i-\bar{z})\hat{\delta}\}}. \tag{6}$$

If $\hat{e}_i^2$ and/or z_i do not vary too much, equation (5) reduces to approximately $-\hat{\sigma}^2\bar{z}$, whereas equation (6) reduces to approximately 0. One way to view what is happening is that $\hat{\sigma}_\delta{}^2$ is inversely related to $\bar{z}$ for given δ (as is apparent from equation (4a)); equation (4b) demonstrates that by automatically centring z this problem is avoided when using $\hat{\gamma}_\delta$. Thus, $\hat{\gamma}_\delta$ is much less sensitive to changes in the data than is $\hat{\sigma}_\delta^2$.

Fig. 3 demonstrates empirically the effectiveness of the transformation, even for small sample sizes. Fig. 3(a) gives a scatterplot of 100 simulated pairs $(\hat{\delta}, \hat{\sigma}_\delta^2)$, based on a simple regression model with $n = 20$, $\delta = 0$, $\sigma = 1$, $\boldsymbol{\beta}' = (0, 1)$ and X and Z being the same vector, generated once as uniform(0, 15). It is apparent that $\hat{\delta}$ and $\hat{\sigma}_\delta^2$ are strongly negatively correlated (in fact, the correlation of these simulated values is -0.787). This pattern can be contrasted with the $(\hat{\delta}, \hat{\gamma}_\delta)$ rela-

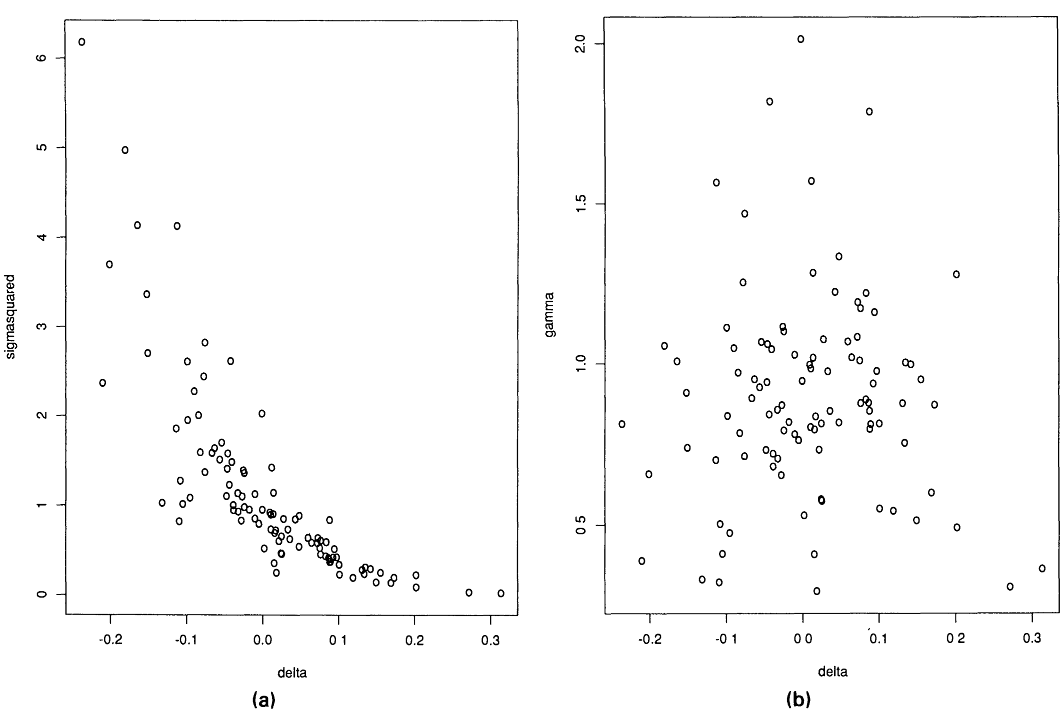

Fig. 3. Scatterplots of simulated values of (a) $\hat{\sigma}^2$ *versus* $\hat{\delta}$ and (b) $\hat{\gamma}$ *versus* $\hat{\delta}$

tionship shown in Fig. 3(b), which clearly demonstrates the benefits of the orthogonality transformation (indeed, $\mathrm{corr}(\hat{\delta}, \hat{\gamma}_\delta) = -0.017$ here).

Use of the modified profile likelihood, based on the orthogonal transformation, leads to new forms of likelihood-based tests, as follows.

2.1. *Likelihood Ratio Statistic*

The standard statistic based on log-likelihood (3) is the likelihood ratio statistic, which has the general form $L = -2\{l_\mathrm{p}(\mathbf{y}; \boldsymbol{\delta}_0) - l_\mathrm{p}(\mathbf{y}; \hat{\boldsymbol{\delta}})\}$. Rutemiller and Bowers (1968) first derived the statistic for model (2) with $w(\mathbf{z}, \boldsymbol{\delta}) = \mathbf{z}'\boldsymbol{\delta}$. This form for w is somewhat problematic, however, since it can be negative. Harvey (1976) derived the statistic for model (2) with $w(\mathbf{z}, \boldsymbol{\delta}) = \exp(\mathbf{z}'\boldsymbol{\delta})$:

$$L = n \log(\hat{\sigma}^2/\hat{\sigma}_\delta^2) - \sum_i \mathbf{z}_i' \hat{\boldsymbol{\delta}}, \tag{7}$$

where $\mathbf{z}_i$ and $\boldsymbol{\delta}$ are $q \times 1$ vectors, $\hat{\sigma}^2 = \hat{\mathbf{e}}'\hat{\mathbf{e}}/n$, $\hat{e}_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$ is the OLS estimate of $\boldsymbol{\beta}$. This statistic, and all statistics discussed in this paper, are referenced to a χ^2-distribution on q degrees of freedom.

The modified profile likelihood substitutes $l_\mathrm{p}(\boldsymbol{\delta}; \hat{\sigma}_\delta^2, \hat{\boldsymbol{\beta}}_\delta) - \log[\det\{j(\boldsymbol{\delta}; \hat{\sigma}_\delta^2, \hat{\boldsymbol{\beta}}_\delta)\}]/2$ for the profile log-likelihood l_p, where j is the observed information for the nuisance parameters $(\sigma^2, \boldsymbol{\beta})$ at a fixed value of the parameter of interest $\boldsymbol{\delta}$. It is shown in Appendix A that the modified profile likelihood ratio statistic has the form

$$L_\mathrm{m} = \frac{n-p-2}{n} L + \log\left\{\frac{\det(X'X)}{\det(\hat{X}_\mathrm{m}'\hat{X}_\mathrm{m})}\right\}, \tag{8}$$

where $\hat{X}_\mathrm{m} = \hat{G}^{-1/2}X$ and $\hat{G}$ is the diagonal matrix with ith entry $w(z_i, \hat{\boldsymbol{\delta}})/\{\Pi_j\, w(z_j, \hat{\boldsymbol{\delta}})\}^{1/n}$. Thus, the standard statistic is modified on the basis of the deviation of the fitted variances from homoscedasticity.

2.2. *Score Statistics*

An important difficulty with the likelihood ratio statistics is that they require estimation of the parameters under both the null and the alternative hypotheses: in this case, both OLS estimation (with $\boldsymbol{\delta} = \boldsymbol{\delta}_0$) and the iterative estimation required to fit model (2). Thus, the statistics are not suitable for routine diagnostic use. Rao (1947) proposed the efficient score statistic as an alternative statistic, which is asymptotically equivalent to the likelihood ratio statistic, but does not require fitting under the alternative hypothesis. The score statistic has the form $S = \mathbf{V}_0' I_0^{-1} \mathbf{V}_0$, where $\mathbf{V}_0 = \partial l/\partial\theta$ is the first-derivative (score) vector and $I_0 = E(-\partial^2 l/\partial\theta\, \partial\theta')$ is the expected information matrix, both evaluated at the null hypothesis. Equivalently, S is a first-order approximation to L evaluated at the null hypothesis. Several researchers (Godfrey, 1978; Breusch and Pagan, 1979; Cook and Weisberg, 1983) have proposed this statistic, apparently independently, for testing homoscedasticity. The statistic has the form

$$S = \frac{\mathbf{u}'\bar{D}(\bar{D}'\bar{D})^{-1}\bar{D}'\mathbf{u}}{2}, \tag{9}$$

where $\mathbf{u}$ is the $n \times 1$ vector with ith entry $u_i = \hat{e}_i^2/\hat{\sigma}^2$, $\bar{D} = (I - \mathbf{11}'/n)D$ and D is the $n \times q$ matrix $(\partial w(z_i, \delta)/\partial \delta_a)$ evaluated at the null hypothesis, i.e. it is half the regression sum of squares of the regression of $\mathbf{u}$ on D, and is thus easily calculated.

A problem with this test (and the likelihood ratio tests (7) and (8)) is that it is crucially dependent on the assumption that ϵ is normally distributed. Koenker (1981), following a suggestion of Bickel (1978) on how to robustify a test for heteroscedasticity when the variance is hypothesized to be a function of the fitted values (Anscombe, 1961), proposed Studentizing the statistic S by substituting $\hat{\phi} = \Sigma_i (\hat{e}_i^2 - \hat{\sigma}^2)^2/n$ for $2\hat{\sigma}^4$ in statistic (9) (note that the term $2\hat{\sigma}^4$ comes from the vector $\mathbf{u}$ and the constant 2 in statistic (9)), i.e.

$$S^* = 2\hat{\sigma}^4 S/\hat{\phi}.$$

Koenker (1981) showed that, unlike S, S^* is asymptotically χ_q^2 for a large class of error distributions.

By taking advantage of the fact that the score statistic is a first-order approximation to the likelihood ratio statistic, it is possible to derive versions of these statistics based on the modified profile likelihood ratio statistic (8). It is shown in Appendix A that these statistics have the form

$$S_{\mathrm{m}} = S + \sum_{a=1}^{q} \left(\sum_{i=1}^{n} h_{ii} t_{ia} \right) \tau_a \tag{10}$$

and

$$S_{\mathrm{m}}^* = S^* + \sum_{a=1}^{q} \left(\sum_{i=1}^{n} h_{ii} t_{ia} \right) \tau_a \tag{11}$$

respectively, where h_{ii} is the ith diagonal element of the prediction matrix $H = X(X'X)^{-1}X'$,

$$t_{ia} = \frac{\partial w(z_i, \delta)}{\partial \delta_a} - \sum_j \frac{\partial w(z_j, \delta)}{\partial \delta_a} \Big/ n$$

evaluated at $\boldsymbol{\delta} = \boldsymbol{\delta}_0$, and $\boldsymbol{\tau} = (\bar{D}'\bar{D})^{-1}\bar{D}'\mathbf{u}$. Thus, the statistics require little additional calculation past the score statistics, except for the determination of the leverage values h_{ii} (which are standard output in most modern regression packages). It is this leverage that determines the difference between the modified and unmodified statistics. In the particular case of $q = 1$, the adjustment to S (or S^*) when $w(z, \delta) = \exp(z\delta)$ has the simple form

$$\left\{ \sum_{i=1}^{n} h_{ii}(z_i - \bar{z}) \right\} \left\{ \sum_{i=1}^{n} (z_i - \bar{z})\hat{e}_i^2 \right\} \Big/ \sum_{i=1}^{n} (z_i - \bar{z})^2 \hat{\sigma}^2.$$

3. Properties of Test Statistics

In this section, the performance of the various test statistics is examined by using Monte Carlo simulations and application to real data. The structure of the simulations performed is described in Appendix B. Representative results are summarized in Figs 4(a)–4(c), which represent analysis for simple regression, with a sample of size 20, under normal, uniform or t_3-errors respectively. As was noted earlier, the

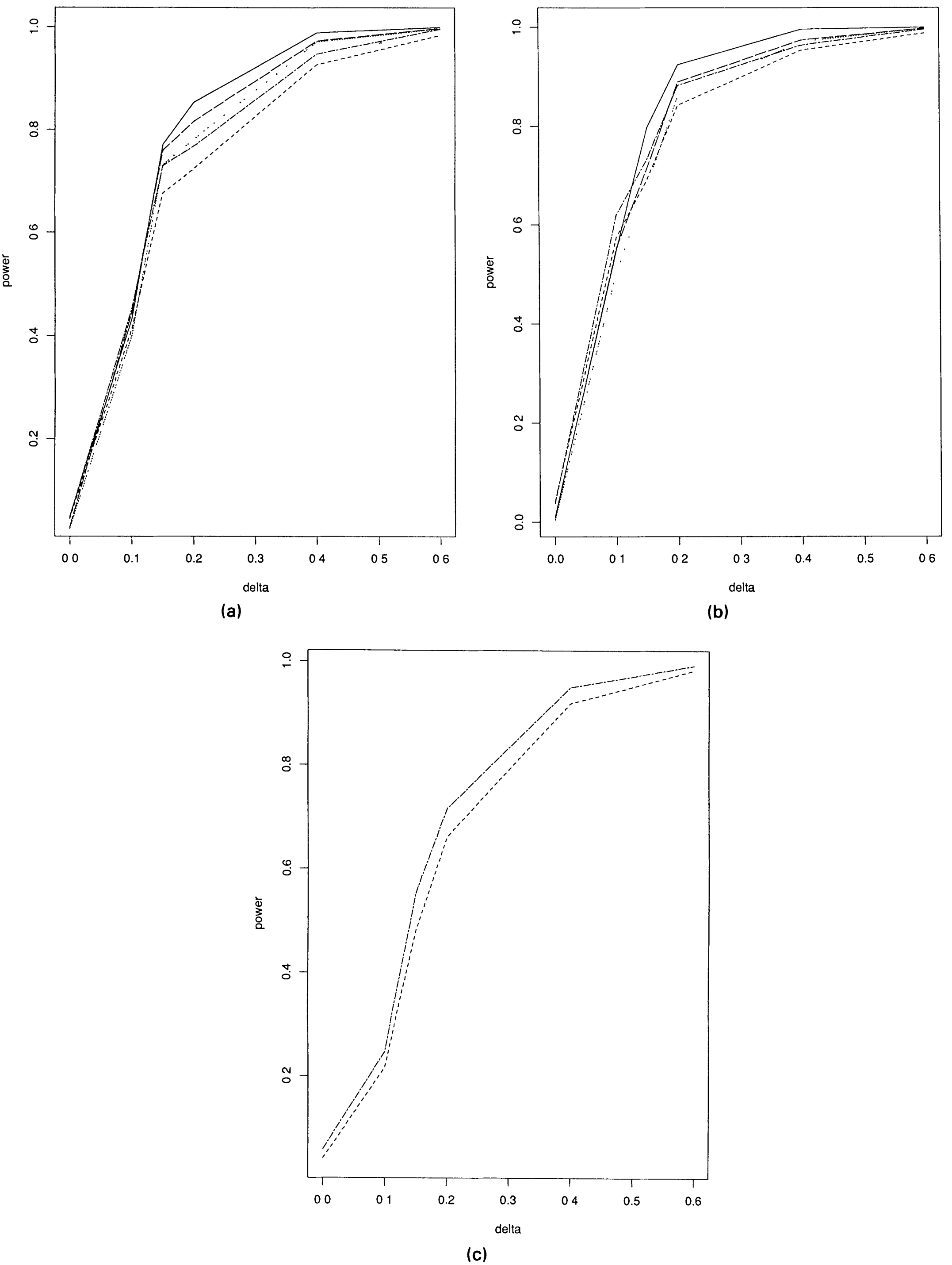

Fig. 4. Simulated power functions of various test statistics: (a) normal errors; (b) uniform errors; (c) t_3-errors (———, modified likelihood ratio test; ·······, score test; — — —, modified score test; --------, Studentized score test; —·—·, modified Studentized score test)

modifications to the tests are larger when the estimated variances are more deviant from homoscedasticity or the leverage values are further from uniform, and these figures reflect that, in that they are based on a predictor that was generated from an exponential distribution (results based on a uniform predictor were similar, but the differences between tests were smaller). Estimated powers at $\delta = 0$, 0.1, 0.15, 0.2, 0.4 and 0.6 are connected by straight lines in the plots. Values are only given for tests that were not significantly anticonservative under the null hypothesis (in the sense that a 95% confidence interval for the true size based on the Monte Carlo runs was too high to include 0.05), since such tests cannot be said to have adequately controlled size, and cannot be recommended. Significantly anticonservative tests included the (unmodified) likelihood ratio statistic L for both normal and uniform errors (thus, the modification of the profile likelihood results in a modified likelihood ratio statistic L_m with much improved small sample null behaviour), and all tests except the Studentized score tests for t_3-errors. It can be noted also that the Studentized tests are not particularly conservative; this is in contrast with Bickel's (1978) Studentized test for heteroscedasticity related to the response (Cook and Weisberg, 1983).

Since the unmodified likelihood ratio test L is the standard test for this problem, it is worth noting that it was often not merely significantly anticonservative, but severely so. For example, when the errors were normally distributed, the estimated sizes of L under the null hypothesis, using a nominal level of 0.05, were 0.218, 0.114 and 0.092 for $n = 15$, $n = 20$ and $n = 35$ respectively. Thus, for small sample sizes, use of the test L is clearly extremely dangerous.

It is apparent from the plots that the properties of any particular test are related to membership in one of the three general types of test (modified likelihood ratio L_m, score S or Studentized score S^*). Under normal errors, the modified likelihood ratio test is most powerful among the tests at almost all levels of δ examined, i.e. the extra effort needed to fit model (2) does give the benefit of a more powerful test for heteroscedasticity. The Studentized score tests lose power as δ increases and are thus inferior choices. It is also apparent that the modification to the score tests is effective, since the modified score tests consistently outperform the unmodified score tests. The pattern is similar for uniform errors, except that the fact that S_m^* is not overconservative makes it the test of choice, until δ exceeds about 0.15; at that point, L_m overcomes its conservativeness and is most powerful. As mentioned earlier, for long-tailed t_3-errors the only choice is between S^* and S_m^*; once again the modified test consistently outperforms the unmodified test. An additional interesting point is that the popular (unmodified) score test S is always beaten by at least one other test under all power situations (although the modified score test S_m is often second best compared with L_m).

Simulations based on multiple-regression data revealed similar, but more extreme, patterns. For example, as p and q grow larger, the anticonservativeness of L becomes larger (for example, for $n = 20$ and $p = q = 2$, a nominal 0.05 level test had true size 0.202); the modified version L_m still controls the null size well, however. Since the χ_q^2-approximation being used here requires n/q to be large, it would be expected that L would fare even worse when using more complicated models for σ_i^2. Once again, for normal errors the computational burden of L_m appears to be worthwhile, since it can be much more powerful than the other tests. For short-tailed errors, only the Studentized tests hold their size, with the other tests being conservative. For long-tailed errors, all tests except the Studentized score tests are hopelessly anticonservative and cannot be used.

These patterns can also be seen in the analysis of real data. Consider again the stock volatility example discussed in Section 1. Recall that Fig. 2 indicated potential heteroscedasticity. Tests for heteroscedasticity using market return as the variable z are as follows:

$$L = 3.329, \quad p = 0.068;$$
$$L_{\mathrm{m}} = 2.963, \quad p = 0.085;$$
$$S = 2.698, \quad p = 0.100;$$
$$S^* = 1.658, \quad p = 0.198;$$
$$S_{\mathrm{m}} = 2.689, \quad p = 0.101;$$
$$S_{\mathrm{m}}^* = 1.650, \quad p = 0.199.$$

Although the unmodified likelihood ratio test L is given, it cannot be recommended on the basis of the simulations. The modified likelihood ratio test and the score tests indicate the marginal heteroscedasticity noted ($\hat{\delta} = 8.092$), with L_{m} closest to significance. The Studentized tests apparently suffer from the lower power seen in the Monte Carlo runs. The associated WLS line is $Y = -0.00492 + 1.2528X$, with overall $F = 24.51$, $p < 0.0001$. It is also worth noting that, when the marginal outlier apparent in Fig. 2 (which corresponds to March 1988) is removed from the data, the tests indicate heteroscedasticity even more strongly (but the modified likelihood ratio test still detected it most strongly among the trustworthy tests, even with the outlier present).

4. Conclusion

In this paper it has been shown that tests for heteroscedasticity based on the modified profile likelihood can be quite effective. If the error distribution appears to be normal, it is worth the effort of fitting the general heteroscedastic model (2), since the modified likelihood ratio test based on it achieves high power, while conrolling the probability of rejection under the null hypothesis. This test statistic is often most powerful among the tests considered here for short-tailed errors as well. If a test statistic was desired that only required fitting the OLS model, the modified score test would be a good choice, as it is sometimes almost as powerful as the modified likelihood ratio test. For long-tailed errors, the modified Studentized score test is a worthy choice, since it controls the null rejection probability far more effecively than the other tests and is more powerful than the modified Studentized test. In particular, the conservativeness of Bickel's Studentized test for heteroscedasticity that is a function of the response does not seem to apply to these tests, where it is a function of some set of predictors.

The ideas discussed here would apply equally well to tests of other regression assumptions. For example, Tsai (1986) derived a score test for simultaneously testing for both heteroscedasticity and autocorrelated errors; a modified version of this test, or the likelihood ratio test, could have improved behaviour. In addition, tests of heteroscedasticity in non-linear regression models could be derived, with the additional possible complications of curvature effects (Bates and Watts, 1988).

Acknowledgements

The authors wish to thank Brad Barber for his assistance in the preparation of this manuscript. Comments by two referees on an earlier draft led to substantial changes in the manuscript.

Appendix A: Derivation of Statistics

A.1. *Derivation of Modified Profile Likelihood Ratio Statistic*

Assume first that $q = 1$. The log-likelihood after the transformation to $\boldsymbol{\theta}_{\mathrm{m}}$ has the form

$$l(\mathbf{y}; \boldsymbol{\theta}_{\mathrm{m}}) = -\frac{n}{2}\log\gamma - \frac{1}{2\gamma}(\mathbf{y} - X\boldsymbol{\beta})' W^{-1}(\mathbf{y} - X\boldsymbol{\beta})\left(\prod_j w_j\right)^{1/n}$$
$$= -\frac{n}{2}\log\gamma - \frac{(\mathbf{y}_{\mathrm{m}} - X_{\mathrm{m}}\boldsymbol{\beta})'(\mathbf{y}_{\mathrm{m}} - X_{\mathrm{m}}\boldsymbol{\beta})}{2\gamma}$$

where $\mathbf{y}_{\mathrm{m}} = G^{-1/2}\mathbf{y}$, and G and X_{m} are defined in the text. Furthermore, the resultant profile log-likelihood function is

$$l_{\mathrm{p}}(\mathbf{y}; \delta) = -\frac{n}{2}\log\hat{\gamma}_\delta - \frac{n}{2},$$

where $\hat{\gamma}_\delta$ is the maximum likelihood estimator of γ for a given value of δ. The adjustment to the profile log-likelihood then has the form $\log\{\det(-\partial^2 l/\partial\boldsymbol{\theta}_{\mathrm{m}2}\,\partial\boldsymbol{\theta}'_{m2})\}|_{\tilde{\boldsymbol{\theta}}_{\mathrm{m}}}/2$, where $\boldsymbol{\theta}_{\mathrm{m}2} = (\gamma, \boldsymbol{\beta}')'$ and $\tilde{\boldsymbol{\theta}}_{\mathrm{m}} = (\delta, \hat{\gamma}_\delta, \boldsymbol{\beta}'_\delta)'$, which is $\log\{(n/2\hat{\gamma}_\delta^{p+2})\det(X'_{\mathrm{m}}X_m)\}/2$; this gives the modified profile log-likelihood

$$l_{\mathrm{mp}}(\mathbf{y}; \delta) = l_{\mathrm{p}}(\mathbf{y}; \delta) - \log\left\{\frac{n}{2\hat{\gamma}_\delta^{p+2}}\det(X'_{\mathrm{m}}X_{\mathrm{m}})\right\}\Big/2.$$

The statistic L_{m} then evaluates this adjustment under both the null and the alternative models:

$$L_{\mathrm{m}} = n\log\left(\frac{\hat{\gamma}_0}{\hat{\gamma}}\right) + \log\left\{\left(\frac{\hat{\gamma}}{\hat{\gamma}_0}\right)^{p+2}\frac{\det(X'X)}{\det(\hat{X}'_{\mathrm{m}}\hat{X}_{\mathrm{m}})}\right\}$$
$$= \left(\frac{n-p-2}{n}\right)L + \log\left\{\frac{\det(X'X)}{\det(\hat{X}'_{\mathrm{m}}\hat{X}_{\mathrm{m}})}\right\},$$

where $\hat{\gamma} = \hat{\gamma}_{\hat{\delta}}$, $\hat{\gamma}_0 = \hat{\gamma}_{\delta_0}$ and $\hat{X}_{\mathrm{m}}$ is X_{m} evaluated at $\hat{\delta}$. Finally, these results are still valid for $q > 1$, since $\partial^2\sigma^2/\partial\delta_a\,\partial\delta_b = \partial^2\sigma^2/\partial\delta_b\,\partial\delta_a$ for $a \neq b$ and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_q)'$ (Cox and Reid (1987), p. 4), with the modification that $\mathbf{w}_j$ is a $q \times 1$ vector and $\Sigma(\dot{w}_j/w_j)^2$ is replaced by the $q \times q$ matrix $\Sigma(\dot{w}_j)(\dot{w}_j)'/w_j^2$.

A.2. *Derivation of Modified Score Statistics*

The score statistic is a first-order approximation to L, so a score statistic based on the modified profile likelihood is simply the standard score statistic (or its Studentized version) modified by a first-order approximation to

$$\log\left[\det\left\{E\left(-\frac{\partial^2 l}{\partial\boldsymbol{\theta}_{\mathrm{m}2}\,\partial\boldsymbol{\theta}'_{\mathrm{m}2}}\right)\Bigg|_{\hat{\boldsymbol{\theta}}_{\mathrm{m}}}\right\}\right] - \log\left[\det\left\{E\left(-\frac{\partial^2 l}{\partial\boldsymbol{\theta}_{\mathrm{m}2}\,\partial\boldsymbol{\theta}'_{\mathrm{m}2}}\right)\Bigg|_{\hat{\boldsymbol{\theta}}_{\mathrm{m}0}}\right\}\right]$$

around $\boldsymbol{\delta} = \boldsymbol{\delta}_0$, with $\hat{\boldsymbol{\theta}}_{\mathrm{m}} = (\hat{\boldsymbol{\delta}}', \hat{\gamma}_{\hat{\delta}}, \hat{\boldsymbol{\beta}}'_{\hat{\delta}})'$ and $\hat{\boldsymbol{\theta}}_{\mathrm{m}0} = (\boldsymbol{\delta}'_0, \hat{\gamma}_{\delta_0}, \hat{\boldsymbol{\beta}}')'$, or

$$\left(\frac{\partial \log[\det\{E(-\partial^2 l/\partial\boldsymbol{\theta}_{\mathrm{m}2}\,\partial\boldsymbol{\theta}'_{\mathrm{m}2})\}]}{\partial\boldsymbol{\delta}} \Bigg|_{\hat{\boldsymbol{\theta}}_{\mathrm{m}0}} \right)' (\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0)$$

$$= \mathrm{tr}\left\{ E\left(-\frac{\partial^2 l}{\partial\boldsymbol{\theta}_{\mathrm{m}2}\,\partial\boldsymbol{\theta}'_{\mathrm{m}2}} \right)^{-1} \frac{\partial E(-\partial^2 l/\partial\boldsymbol{\theta}_{\mathrm{m}2}\,\partial\boldsymbol{\theta}'_{\mathrm{m}2})}{\partial\boldsymbol{\delta}} \Bigg|_{\hat{\boldsymbol{\theta}}_{m0}} \right\}' (\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) \qquad (12)$$

(Cox and Reid (1987), equation (19)). Since $G|_{\hat{\boldsymbol{\theta}}_{\mathrm{m}0}} = I$,

$$E\left(-\frac{\partial^2 l}{\partial\boldsymbol{\theta}_{\mathrm{m}2}\,\partial\boldsymbol{\theta}'_{\mathrm{m}2}} \right)\Bigg|_{\hat{\boldsymbol{\theta}}_{\mathrm{m}0}} = \begin{pmatrix} X'X/\hat{\gamma}_0 & 0 \\ 0 & n/2\hat{\gamma}_0^2 \end{pmatrix}$$

and

$$\frac{\partial E(-\partial^2 l/\partial\boldsymbol{\theta}_{\mathrm{m}2}\,\partial\boldsymbol{\theta}'_{\mathrm{m}2})}{\partial\boldsymbol{\delta}}\Bigg|_{\hat{\boldsymbol{\theta}}_{\mathrm{m}0}} = \frac{\partial}{\partial\boldsymbol{\delta}} \begin{pmatrix} X'G^{-1}X/\gamma & 0 \\ 0 & n/2\gamma^2 \end{pmatrix}\Bigg|_{\hat{\boldsymbol{\theta}}_{\mathrm{m}0}}$$

$$= \begin{pmatrix} -X'\dot{G}X/\hat{\gamma}_0 & 0 \\ 0 & 0 \end{pmatrix},$$

while

$$\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0 \approx \left\{ E\left(-\frac{\partial^2 l}{\partial\delta\,\partial\delta'} \right) \right\}^{-1} \frac{\partial l}{\partial\delta}\Bigg|_{\hat{\boldsymbol{\theta}}_{\mathrm{m}0}} = (\overline{D}'\overline{D})^{-1}\overline{D}'\mathbf{u},$$

equation (12) simplifies to $-\mathrm{tr}(H\dot{G})(\overline{D}'\overline{D})^{-1}\overline{D}'\mathbf{u}$, where $\dot{G}$ is the $q \times n \times n$ matrix whose ath component is the diagonal matrix with entry $\partial G/\partial\delta_a|_{\hat{\boldsymbol{\theta}}_{\mathrm{m}0}}$. Since $\mathrm{tr}(H\dot{G}) = \Sigma_{i=1}^n h_{ii}t_{ia}$ and $\boldsymbol{\tau} = (\overline{D}'\overline{D})^{-1}\overline{D}'\mathbf{u}$, this ultimately gives equations (10) and (11).

Appendix B: Description of Simulation Structure

The simulations were constructed to provide a wide range of potential heteroscedastic situations. Sensitivity to sample size was studied by treating sample sizes of 20, 50 and 100; robustness of the tests was examined by generating errors from normal, uniform or t_3-distributions. Uniform deviates were generated by using a multiplicative congruential generator; these were then transformed to follow a normal distribution, if appropriate, by using the Box–Muller method (Press *et al.*, 1986). The t_3-distributed variates were formed as the ratio of a normal to the square root of a χ_3^2-variate divided by its degrees of freedom, with the χ_3^2-variates generated as the sum of a normal squared and the natural logarithm of a uniform multiplied by -2. The parameter estimates under model (2), with $w(\mathbf{z}, \boldsymbol{\delta}) = \exp(\mathbf{z}'\boldsymbol{\delta})$, were calculated by using the International Mathematical and Statistical Libraries' (1989) routines UVMGS, UMPOL and UMINF. All calculations were performed in double precision. Without loss of generality, the parameters $\boldsymbol{\beta}$ and σ^2 were set such that $\beta_1 = 0$, $\beta_2 = \ldots = \beta_p = 1$ and $\sigma^2 = 1$. The columns of the X-matrix were generated as uniform(0, 15) (except for the first column of all 1s) or exponential with mean 7.5. There were 1000 replications per simulation experiment. All statistics were compared with the appropriate χ_q^2 critical values at an $\alpha = 0.05$ level. In the runs presented here the Z-matrix was taken to be the first q columns of the X-matrix (ignoring the column of 1s), although the results were substantially similar when the Z-matrix was generated independently of X.

References

Anscombe, F. J. (1961) Examination of residuals. In *Proc. 4th Berkeley Symp. Mathematical Statistics*, vol. 1, pp. 1–36. Berkeley: University of California Press.

Bates, D. M. and Watts, D. G. (1988) *Nonlinear Regression Analysis and Its Applications*. New York: Wiley.

Bickel, P. J. (1978) Using residuals robustly: I, Tests for heteroscedasticity, nonlinearity. *Ann. Statist.*, **6**, 266–291.

Breusch, T. S. and Pagan, A. R. (1979) A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, **47**, 1287–1294.

Cook, R. D. and Weisberg, S. (1983) Diagnostics for heteroscedasticity in regression. *Biometrika*, **70**, 1–10.

Cox, D. R. (1988) Some aspects of conditional and asymptotic inference: a review. *Sankhya* A, **50**, 314–337.

Cox, D. R. and Reid, N. (1987) Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Statist. Soc.* B, **49**, 1–39.

Cruddas, A. M., Reid, N. and Cox, D. R. (1989) A time series illustration of approximate conditional likelihood. *Biometrika*, **76**, 231–237.

Davison, A. C. (1988) Approximate conditional inference in generalized linear models. *J. R. Statist. Soc.* B, **50**, 445–461.

Fisher, L. and Kamin, J. H. (1985) Forecasting systematic risk: estimates of "raw" beta that take account of the tendency of beta to change and the heteroskedasticity of residual returns. *J. Fin. Quant. Anal.*, **20**, 127–149.

Godambe, V. P. (1991) Orthogonality of estimating functions and nuisance parameters. *Biometrika*, **78**, 43–51.

Godfrey, L. G. (1978) Testing for multiplicative heteroskedasticity. *J. Econometr.*, **8**, 227–236.

Harvey, A. C. (1976) Estimating regression models with multiplicative heteroscedasticity. *Econometrica*, **44**, 461–466.

Hinkley, D. V. (1989) Modified profile likelihood in transformed linear models. *Appl. Statist.*, **38**, 495–506.

International Mathematical and Statistical Libraries (1989) *MATH/LIBRARY*. Houston: IMSL.

Koenker, R. (1981) A note on studentizing a test for heteroscedasticity. *J. Econometr.*, **17**, 107–112.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1986) *Numerical Recipes*, p. 202. Cambridge: Cambridge University Press.

Rao, C. R. (1947) Large sample tests of statistical hypotheses concerning several parameters with application to problems of testing. *Proc. Camb. Phil. Soc.*, **44**, 50–57.

Rutemiller, H. C. and Bowers, D. A. (1968) Estimation in a heteroscedastic regression model. *J. Am. Statist. Ass.*, **63**, 552–557.

Tsai, C.-L. (1986) Score test for the first-order autoregressive model with heteroscedasticity. *Biometrika*, **73**, 455–460.