

A COMPARISON BETWEEN MAXIMUM LIKELIHOOD
AND GENERALIZED LEAST SQUARES IN A HETEROSCEDASTIC LINEAR MODEL

by

R.J. Carroll* and David Ruppert**

Abstract

We consider a linear model with normally distributed but heteroscedastic errors. When the error variances are functionally related to the regression parameter, one can use either maximum likelihood or generalized least squares to estimate the regression parameter. We show that maximum likelihood is much more sensitive to small misspecifications in the functional relationship between the error variances and the regression parameter.

Key Words: Linear Models, Heteroscedasticity, Contiguity, Robustness, Weighted Least Squares, Maximum Likelihood.

Subject Class: Primary 62J05; Secondary 62G35.

*University of North Carolina and National Heart, Lung and Blood Institute. Work supported by the U.S. Air Force Office of Scientific Research Contract AFOSR-80-0080. Part of this work completed at the Universität Heidelberg, with support from the Deutsche Forschungsgemeinschaft.

**University of North Carolina. Work supported by NSF Grant MCS78-01240.

1. Introduction

There has been considerable recent interest in the heteroscedastic linear model

$$(1.1) \quad Y_i = x_i \beta + \epsilon_i [f(x_i, \beta, \theta)]^{-1/2},$$

where $\beta(p \times 1)$ is the regression coefficient, $\{x_i(1 \times p)\}$ are the design vectors, $\{\epsilon_i\}$ are independent and identically distributed with distribution function F , and the function $f(x_i, \beta, \theta)$ expresses the possible heteroscedasticity. Bickel (1978) considers various tests of the hypothesis of homoscedasticity, i.e. tests of

$$(1.2) \quad H_0: f(x_i, \beta, \theta) \equiv \text{constant}.$$

His work has been extended by Carroll and Ruppert (1981a), and the tests have been shown to be locally most powerful by Hammerstrom (1981). Other recent papers are Jobson and Fuller (1980), Carroll and Ruppert (1981b), Box and Hill (1974), and Fuller and Rao (1978).

Box and Hill (1974), Ruppert and Carroll (1979), and Jobson and Fuller (1980) suggest various forms of generalized weighted least squares estimates (GLSE) of β . Basically, the suggestion is to obtain preliminary estimates $(\hat{\beta}_p, \hat{\theta})$ of (β, θ) , compute estimated weights $[f(x_i, \hat{\beta}_p, \hat{\theta})]^{-1}$, and then perform ordinary weighted least squares. Carroll and Ruppert (1981b) emphasize robustness and develop methods which are robust against outliers and non-normal distributions F ; they prove that generalized M-estimates of β , which include GLSE estimates as special cases, are just as good asymptotically as if the weights were really known. The same phenomenon has been found in other models of heteroscedasticity; see Williams (1975) for a review.

Jobson and Fuller (1980) suggest using the information about β in the function f to improve upon the GLSE. They state that their method is asymptotically equivalent to the MLE for β obtained by setting up the normal likelihood based on (1.1) and maximizing it; this likelihood is

$$(1.3) \quad \frac{1}{2} \sum_{i=1}^N \log(f(x_i, \beta, \theta)) - \frac{1}{2} \sum_{i=1}^N (Y_i - x_i \beta)^2 f(x_i, \beta, \theta) .$$

They have a very interesting result suggesting that as long as (1.1) is correct and F is normal, then the MLE will be preferred to the GLSE.

In this heteroscedasticity problem, we have an additional robustness consideration. Besides the usual goal (Huber (1977)) of protecting ourselves against outliers and non-normal error distributions, we also must protect ourselves against slight misspecifications in the functional relationship between $\text{Var}(Y_i)$ and (x_i, β, θ) . Since this functional relationship expressed in (1.1) through f is typically at best an approximation, and since our primary interest is estimating β , we would prefer not to estimate β by a statistic which is adversely affected by slight misspecifications of f .

In this note, we assume that the error distribution F is actually normal. We study the robustness of GLSE and MLE to small specification errors in f by means of simple contiguity techniques. We show that small mistakes in specifying f can easily make GLSE preferable to the MLE.

2. A Contiguous Model

We consider small deviations from (1.1) in the form of

$$(2.1) \quad Y_i = x_i \beta + [g_N(x_i, \beta, \theta)]^{-\frac{1}{2}} \epsilon_i ,$$

where

$$(2.2) \quad g_N(x_i, \beta, \theta) = f(x_i, \beta, \theta) \{1 + 2BN^{-\frac{1}{2}} h(x_i, \beta, \theta)\}$$

$$N^{-1} \sum_{i=1}^N h^2(x_i, \beta, \theta) \rightarrow \mu \quad (0 < \mu < \infty)$$

$\{\varepsilon_i\}$ are i.i.d. standard normal.

One should note that the model (2.1) is very close to the assumed model (1.1). Thus the model (2.1) fits our needs because the variance misspecification error is very small and decreases for larger sample sizes. An estimate of β which is robust against specification errors should have the same asymptotic properties under both models (1.1) and (2.1). Thus the question at hand is to study the sensitivity of the MLE and GLSE when (1.1) is assumed but (2.1) is true. If ℓ_1 denotes the log-likelihood for (1.1), and ℓ_2 is the log-likelihood for (2.1), it is quite simple to show that, when (1.1) is true, to order $o_p(1)$,

$$(2.3) \quad \begin{aligned} \ell_* &= \ell_2 - \ell_1 \\ &\doteq -B^2 \mu + \sum_{i=1}^N (\varepsilon_i^2 - 1) Bh(x_i, \underline{\beta}, \underline{\theta}) N^{-\frac{1}{2}}, \end{aligned}$$

so that

$$(2.4) \quad L(\ell_*) \xrightarrow{L} N(-B^2 \mu, 2B^2 \mu) \text{ when model (1.1) holds,}$$

where $N(a, b)$ is the normal distribution with mean a and variance b .

From Corollary 1.2 of Hájek and Sidák (1967, p. 204), this means that model (2.1) is contiguous to model (1.1).

3. Limit Distributions for GLSE

Suppose that for some positive definite matrix S ,

$$(3.1) \quad N^{-1} \sum_{i=1}^N x_i' x_i f(x_i, \beta, \theta) \rightarrow S .$$

Then, assuming normal errors and smoothness conditions on f , Carroll and Ruppert (1981b) (as well as Jobson and Fuller (1980)) show that when model (1.1) is true, the GLSE $\hat{\beta}_G$ satisfies

$$(3.2) \quad N^{\frac{1}{2}}(\hat{\beta}_G - \beta) - N^{-\frac{1}{2}} \sum_{i=1}^N S^{-1} x_i' f^{\frac{1}{2}}(x_i, \beta, \theta) \varepsilon_i \xrightarrow{P} 0 ,$$

$$(3.3) \quad N^{\frac{1}{2}}(\hat{\beta}_G - \beta) \xrightarrow{L} N(0, S^{-1}) .$$

A formal proof is possible as long as f is smooth, $\{f(x_i, \beta, \theta)\}$ is bounded away from ∞ , and $(\hat{\beta}_p, \hat{\theta})$ satisfy

$$(3.4) \quad \begin{aligned} N^{\frac{1}{2}}(\hat{\beta}_p - \beta) &= o_p(1) \\ N^{\frac{1}{2}}(\hat{\theta} - \theta) &= o_p(1) . \end{aligned}$$

Carroll and Ruppert (1981b) and Jobson and Fuller (1980) verify (3.4) in the normal case under certain technical conditions.

Now, since $\{\varepsilon_i\}$ are symmetric random variables, one uses (2.3) and (3.2) to show that $\lambda_* = \lambda_2 - \lambda_1$ and $N^{\frac{1}{2}}(\hat{\beta}_G - \beta)$ are independent, so that by LeCam's third lemma (Hájek and Šidák (1967, p. 208)),

$$(3.5) \quad L(N^{\frac{1}{2}}(\hat{\beta}_G - \beta)) \rightarrow N(0, S^{-1}) ,$$

and this *under either model (1.1) or (2.1)*. This means that GLSE is

robust against small specification errors of the variance function f .

This encouraging result suggests that one will not go too wrong with GLSE as long as model (1.1) is reasonable.

4. Limit Distribution for the MLE

While GLSE is robust against minor errors in specifying the function f in model (1.1), the same cannot be said for the MLE. Denote this MLE by $\hat{\beta}_M$. Jobson and Fuller (1980) show that for a particular covariance matrix Σ ,

$$(4.1) \quad N^{\frac{1}{2}}(\hat{\beta}_M - \beta) \xrightarrow{L} N(0, \Sigma) .$$

The result of particular interest is that Σ is no larger than S (see (3.1) and (3.3)) in the sense that $S - \Sigma$ is positive semi-definite under the model (1.1). In addition to (4.1), from (2.3) and the proof of Theorem 2 in Jobson and Fuller (1980), $N^{\frac{1}{2}}(\hat{\beta}_M - \beta)$ and ℓ_* are jointly asymptotically normal with mean $(0, -B^2 \mu)$, marginal variances $(\Sigma, 2B^2 \mu)$, and covariances Bq computed below, i.e.

$$(4.2) \quad (N^{\frac{1}{2}}(\hat{\beta}_M - \beta)', \ell_*) \xrightarrow{L} N\left((0, -B^2 \mu), \begin{pmatrix} \Sigma & Bq \\ Bq' & 2B^2 \mu \end{pmatrix}\right) .$$

We now indicate why it is true that the only cases in which the MLE can be expected to be robust against variance specification errors is when $S = \Sigma$ and the MLE is asymptotically equivalent to GLSE. To see this, first consider model (1.1) to hold. Jobson and Fuller (1980) show that $\hat{\beta}_G$ is essentially a linear function of $\{\varepsilon_i\}$ and $\{\varepsilon_i^2 - 1\}$, i.e. for scalars $\{v_i\}$ and $\{w_i\}$,

$$(4.3) \quad N^{\frac{1}{2}}(\hat{\beta}_M - \beta) = N^{-\frac{1}{2}} \sum_{i=1}^N \{v_i \varepsilon_i + w_i (\varepsilon_i^2 - 1)\} + o_p(1) .$$

If we have $w_i \equiv 0$ ($i = 1, \dots, N$), then from (3.2), (4.3), and Gauss-Markov, we have that

$$N^{\frac{1}{2}}(\hat{\beta}_M - \hat{\beta}_G) \xrightarrow{P} 0 ,$$

and the estimates have the same limit distribution. Thus the only way for $\hat{\beta}_M$ to improve on $\hat{\beta}_G$ under (1.1) is for the $\{w_i\}$ to be non-zero. In this case, however, we can perform contiguity calculations based on (2.3) and (4.3), thus showing that under model (2.1),

$$(4.4) \quad N^{\frac{1}{2}}(\hat{\beta}_M - \beta) \xrightarrow{L} N(-2Bq, \Sigma) ,$$

$$q = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N w_i h(x_i, \beta, \theta) .$$

Of course, q will be non-zero in general if the $\{w_i\}$ are.

These results have important consequences for efficiency. Suppose we wish to estimate the linear combination $\alpha' \beta$. Then, under model (1.1),

$$(4.5) \quad N \text{MSE}(\alpha' \hat{\beta}_G) \rightarrow \alpha' S^{-1} \alpha$$

$$N \text{MSE}(\alpha' \hat{\beta}_M) \rightarrow \alpha' \Sigma \alpha .$$

However, under the model (2.1),

$$(4.6) \quad N \text{MSE}(\alpha' \hat{\beta}_G) \rightarrow \alpha' S^{-1} \alpha \quad (\text{no change})$$

$$N \text{MSE}(\alpha' \hat{\beta}_M) \rightarrow \alpha' \Sigma \alpha + 4B^2(\alpha' q)^2 ,$$

and of course $\alpha' \hat{\beta}_M$ will be a rather poor estimate if α is not orthogonal to q and B is large.

5. Monte-Carlo

We performed a small Monte-Carlo study to illustrate the results given above. The model is that of Jobson and Fuller (1980), with a sample size of $N = 40$:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \sigma_i \epsilon_i .$$

The presumed model for variances is

$$(5.1) \quad \sigma_i^2 = \alpha_1 + \alpha_2 \tau_i^2, \quad \tau_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} .$$

We choose the design $\{x_{i1}, x_{i2}\}$ as in Table 1 of Jobson and Fuller (1980), with their choices $(\beta_0, \beta_1, \beta_2) = (10, -4, 2)$ and $(\alpha_1, \alpha_2) = (300, .2)$. All experiments were replicated 200 times.

The first departure from model (4.1) was quite moderate:

$$(5.2) \quad \sigma_i^2 = \alpha_1 + 2\alpha_2 |\tau_i| + \alpha_2 \tau_i^2 .$$

The second departure was quite substantial and reflects severe heteroscedasticity:

$$(5.3) \quad \sigma_i = \sqrt{\alpha_1} \exp(\alpha_2^2 |\tau_i|) .$$

Besides the estimator JLS defined by Jobson and Fuller (1980), three others were considered. The first (LSE) is ordinary least squares. The second (ROBUST) is simply a Huber Proposal 2 regression estimate which

for the model $EY_i = \beta x_i$ simultaneously solves

$$\Sigma \psi((Y_i - \beta x_i)/\sigma) x_i / \sigma = 0$$

(5.4)

$$\frac{1}{N-3} \Sigma [\psi^2((Y_i - \beta x_i)/\sigma)] = \int \frac{\psi^2(v) e^{-v^2/2}}{\sqrt{2\pi}} dv,$$

where $\psi(x) = \max(-2, \min(x, 2))$. Note that the least squares estimate of β can be generated from (5.4) by choosing $\psi(x) = x$. See Huber (1977) for further details about robustness. The third estimate (WEIGHT) is a weighted robust estimate (Carroll and Ruppert (1981b)) defined as follows.

- (i) Let $\hat{\beta}$ = Huber's Proposal 2 estimate.
- (ii) Let $\dot{r} = (Y_i - x_i \hat{\beta})^2$, \dot{M} as in Jobson and Fuller (1980).
- (iii) Perform Proposal 2 for the model $E\dot{r} \approx \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$.
- (iv) Define $\hat{\sigma}_i^2 = \hat{\alpha}_1 + \hat{\alpha}_2 (x_i \hat{\beta})^2$.
- (v) Perform weighted Proposal 2 by solving (5.4) with σ replaced by $\hat{\sigma}_i$. Call the estimate $\hat{\beta}$.
- (vi) Repeat (ii)-(v).

The standard normal distribution $N(\mu = 0, \sigma = 1)$ was one of the error distributions used. The second was a contaminated normal. A uniform (0,1) random variable U and a $N(0, \sigma = 1)$ Z were chosen. We have the error as Z unless $U \geq .90$, in which case the error is $3Z$. We used the IMSL routines GGUBS and GGNPM. The starting seed was 123457. We start the process by generating 50 uniforms from GGUBS and then 50 normals from GGNPM; this we repeated 350 times. We then repeated the experiment 200 times, each time generating 50 uniforms and then 50 normals.

The outcome of our experiment is given in Table 1. The entries are the ratios of mean square errors of different estimates to the LSE. It is clear from the results that the scoring estimate JLS is very sensitive to the assumption of normality and somewhat sensitive to specification errors of the model (5.1).

TABLE 1

Ratio of MSE for different estimates to the MSE of the LSE.

	Standard Normal			Contaminated Normal		
	β_0	β_1	β_2	β_0	β_1	β_2
Correct Model (4.1)						
JLS	.80	.94	1.04	.94	1.13	1.14
ROBUST	.96	.99	1.00	.77	.80	.77
WEIGHT	.86	.94	.99	.69	.73	.71
Incorrect Model (4.2)						
JLS	.80	.94	1.04	.96	1.09	1.09
ROBUST	.96	.99	1.00	.77	.80	.77
WEIGHT	.86	.94	.98	.69	.73	.71
Incorrect Model (4.3)						
JLS	4.08	5.40	7.94	11.21	9.33	5.41
ROBUST	.56	.69	.68	.45	.57	.54
WEIGHT	.42	.46	.56	.39	.40	.43

References

- Bickel, P.J. (1978). Using residuals robustly I: tests for heteroscedasticity, nonlinearity. *Ann. Statist.* 6, 266-291.
- Box, G.E.P. and Hill, W.J. (1974). Correcting inhomogeneity of variances with power transformation weighting. *Technometrics* 16, 385-389.
- Carroll, R.J. and Ruppert, D. (1981a). On robust tests for heteroscedasticity. *Ann. Statist.* 9, 205-209.
- Carroll, R.J. and Ruppert, D. (1981b). Robust estimation in heteroscedastic linear models. Manuscript.
- Fuller, W.A. and Rao, J.N.K. (1978). Estimation for a linear regression model with unknown diagonal covariance matrix. *Ann. Statist.* 6, 1149-1158.
- Hájek, J. and Sidák, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.
- Hammerstrom, T. (1981). Asymptotically optimal tests for heteroscedasticity in the general linear model. To appear in *Ann. Statist.*
- Huber, P.J. (1977). *Robust Statistical Procedures*. SIAM, Philadelphia.
- Jobson, J.D. and Fuller, W.A. (1980). Least squares estimation when the covariance matrix and parameter vector are functionally related. *J. Amer. Statist. Assoc.* 15, 176-181.
- Ruppert, D. and Carroll, R.J. (1979). M-estimates for the heteroscedastic linear model. *Institute of Statistics Mimeo Series #1243*, University of North Carolina.
- Williams, J.S. (1975). Lower bounds on convergence rates of weighted least squares to best linear unbiased estimators. In *A Survey of Statistical Design and Linear Models*. J.N. Srivastava, ed. North-Holland, Amsterdam.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Comparison Between Maximum Likelihood and Generalized Least Squares in a Heteroscedastic Linear Model		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL
		6. PERFORMING ORG. REPORT NUMBER Mimeo Series No. 1334
7. AUTHOR(s) R.J. Carroll and David Ruppert		8. CONTRACT OR GRANT NUMBER(s) AFOSR Contract AFOSR-80-0080 NSF Grant MCS78-01240
9. PERFORMING ORGANIZATION NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Bolling Air Force Base Washington, DC 20332		12. REPORT DATE March 1981
		13. NUMBER OF PAGES 12
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release -- Distribution Unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Linear Models, Heteroscedasticity, Contiguity, Robustness, Weighted Least Squares, Maximum Likelihood		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) We consider a linear model with normally distributed but heteroscedastic errors. When the error variances are functionally related to the regression parameter, one can use either maximum likelihood or generalized least squares to estimate the regression parameter. We show that maximum likelihood is much more sensitive to small misspecifications in the functional relationship between the error variances and the regression parameter.		