THE INSTITUTE OF STATISTICS

THE CONSOLIDATED UNIVERSITY OF NORTH CAROLINA

A NOTE ON THE EFFECT OF ESTIMATING WEIGHTS IN WEIGHTED LEAST SQUARES

Raymond J. Carroll
Department of Statistics
University of North Carolina
Chapel Hill, NC 27514

David Ruppert
Department of Statistics
University of North Carolina
Chapel Hill, NC 27514

Mimeo Series #1570

January 1985

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

A NOTE ON THE EFFECT OF ESTIMATING WEIGHTS

IN WEIGHTED LEAST SQUARES

Raymond J. Carroll

and

David Ruppert

University of North Carolina at

Chapel Hill

ABSTRACT

We consider heteroscedastic linear models for which the variances are parametric functions of known regressors. Second order expansions are derived for a class of estimators which includes normal theory maximum likelihood and generalized least squares. The result is a fairly precise description of when conventional asymptotic variance formulae are optimistic; i.e., they underestimate the true variances effectively, this optimism persists for all but heavy-tailed error distributions. We find that maximum likelihood and generalized least squares have the same covariance matrix to second order. Our results also indicate the effect of preliminary estimators.

ACKNOWLEDGEMENT

The work of Professor Carroll was supported by the Air Force Office of Scientific Research Contract AFOSR-F49620-82-C-0009, Sonderforschungsbereich 123 at the University of Heidelberg and by The Mathematics Research Center of the University of Wisconsin, Army Research Office Contract DAAG29-80-C-0041. Professor Ruppert was supported by the National Science Foundation Grant MCS-8100748. The authors wish to thank C.F. Jeff Wu for helpful discussions.

Some Key Words and Phrases: Weighted Least Squares, Estimated Weights, Second Order Variance Approximations, Heteroscedasticity, Regression.

AMS Classifications:

Primary: 62F10

Secondary: 62J05

1. Introduction

Consider a heteroscedastic linear model where the variances are parametric functions of known constants. A simple such model which will form the basis of our discussion is

$$Y_i = x_i^T \beta + (\sigma \exp(\theta h_i)) \varepsilon_i, \quad i = 1, \ldots, N. \tag{1.1}$$

Here, $\{x_i\}$ and β are p-vectors, the $\{\varepsilon_i\}$ are independent and identically distributed with mean zero, variance one, fourth moment κ and are symmetrically distributed about zero. The $\{h_i\}$ are known constants, and the major goal is to estimate β, although the parameter θ is also of some interest.

Model (1.1) does not include many models that have been considered in the literature, for example in Carroll & Ruppert (1982a, 1983), but second order calculations are already cumbersome for (1.1). We expect that more general models will exhibit the same qualitative behavior as (1.1).

There are many ways used in practice to estimate the parameters (β, σ, θ) in model (1.1), see for example Fuller and Rao (1978), Carroll & Ruppert (1982 a), Carroll (1982). In this note we restrict ourselves to the following scheme, which we will call Extended Least Squares.

Obtain a preliminary estimate $\hat{\beta}_p$ of $\hat{\beta}$; (1.2a)

Treating β in (1.1) as if it were known and equal to $\hat{\beta}_p$, obtain the normal theory maximum likelihood estimate $\hat{\theta}$ of θ; (1.2b)

Estimate β by $\hat{\beta}$, the weighted least squares estimate with estimated weights $\exp(-2\,\hat{\theta}\,h_i)$. (1.2c)

Both maximum likelihood $\hat{\beta}_{ML}$ and generalized least squares $\hat{\beta}_{GLS}$ fall in the class of extended least squares estimates, the former by starting at the maximum likelihood estimate itself, the latter by starting with least squares $\hat{\beta}_{LS}$.

Adapting the arguments in Carroll & Ruppert (1982 a), one can show that as long as $N^{\frac{1}{2}}(\hat{\beta}_p - \beta) = 0_p(1)$, then $N^{\frac{1}{2}}(\hat{\theta} - \theta) = 0_p(1)$, as well and

$$N^{\frac{1}{2}}(\hat{\beta} - \beta) \xrightarrow{L} N(0, V^{-1}), \text{ where} \tag{1.3}$$

$$V^{-1} = \lim_{N\to\infty} (N-p)^{-1}\sigma^2 \sum_{i=1}^{N} x_i\, x_i^T \exp(-2\,\theta\, h_i). \tag{1.4}$$

Equations (1.3) - (1.4) have a number of consequences, which can be stated informally as

There is no cost incurred for estimating weights; (1.5a)

The limiting normal distribution of extended least squares does not depend on the method used to estimate the weights; (1.5b)

For inference, one can treat the estimated weights as fixed and plug them into standard regression packages. (1.5c)

The result (1.3) - (1.4) suggesting no cost for estimating weights is known to be rather optimistic, see Williams (1975) and Freedman & Peters (1984). Consider fitting (1.1) to data which arise from simple linear regression through the origin based on a sample of size 10. Take $\theta=0$, $\sigma=1$ and $\beta=1$. Choose the design points $\{x_i\}$ as

$\pm 0.0033, \pm 0.0526, \pm 0.2662, \pm 0.8413, \pm 2.0539,$

with $h_i = x_i$. The classical theory (1.3) - (1.4) suggests that the generalized least squares estimate $\hat{\beta}_{GLS}$ should be approximately normally distributed with mean $\beta=1$ and variance 0.11, basically independent of the underlying distribution. We performed a small simulation at the normal model and found that the Monte-Carlo variance of generalized least squares was actually 0.23, a 110% increase over the theoretical value. The average value for the estimated variance

$$(N-1)^{-1} \sum_{i=1}^{N} (Y_i - x_i^T \hat{\beta}_{GLS})^2 \exp(-2\,\hat{\theta} h_i) / \left(\sum_{i=1}^{N} x_i^2 \exp(-2\,\hat{\theta} h_i)\right) \tag{1.6}$$

was 0.15, still a substantial underestimate.

We performed a similar computation for a distribution with heavier tails than the normal. Specifically, with probability 0.90 we generated a normal random error with variance 0.615, while with probability 0.10 we generated a normal random error with variance 4.462. The resultant has variance approximately 1.0, with fourth moments 7.00. Here again the formal large sample theory suggests that the variance of generalized least squares should be 0.11, but this time it is an *overestimate* of more than 50% since the Monte-Carlo estimate is only 0.07.

The major goal of this paper is to obtain some understanding of the simulation phenomenon of extra-variation at the normal model but under-variation at the contaminated normal model, compared to the asymptotic formula (1.4). By expanding the covariance matrix of $N^{\frac{1}{2}}(\hat{\beta} - \beta)$ to second order, we obtain a fairly precise description. If the standardized errors

$\{\varepsilon_i\}$ have zero skewness and fourth moment κ, then generalized least squares and maximum likelihood agree to second order in their covariances. For these two estimators, we show that when $\kappa < 5$, the conventional asymptotic covariance formula (1.4) will be too low, with the opposite occuring if $\kappa > 5$. This result differs from the work of Freedman & Peters in two ways, first because we consider a regression model rather than a one-sample problem, and second that we make explicit that the boundary between extra and under dispersion occurs for $\kappa = 5$. Our results differ from those of Toyooka (1984) in that we have a less general basic model but, we allow non-normal errors and obtain results for estimates other than normal-theory maximum likelihood and generalized least squares. In particular, we are able to study the effect of using other preliminary estimators in (1.2a). Our results also differ from Toyooka (1982) and is considerably more general in that we do not assume that $\hat{\theta}$ is independent of $\{Y_1, \ldots, Y_N\}$. Whether $\hat{\theta}$ depends on $\{Y_1, \ldots, Y_N\}$ or not has a large effect on second order terms.

2. General Results

We assume in discussing extended least squares that $\hat{\theta}$, $\hat{\beta}_p$ are root-N consistent, i.e.,

$$N^{\frac{1}{2}}(\hat{\theta} - \theta) = O_p(1).$$
$$N^{\frac{1}{2}}(\hat{\beta}_p - \beta) = O_p(1). \tag{2.1}$$

We make the following definitions.

$$z_i = x_i/\exp(\theta h_i)$$
$$C_{jN} = N^{-1} \sum_{i=1}^{N} z_i z_i^T (2h_i)^{j-1} \tag{2.2}$$
$$\mu_{kN} = N^{-1} \sum_{i=1}^{N} h_i^k$$

For every N, we are going to normalize so that in (2.1), $\mu_{1N} = 0$ and $\mu_{2N} = 1$. Strictly speaking, this means that the parameters (θ, σ) depend on N, but we will suppress this dependence. The major result is the following: the proof is sketched in the appendix.

Theorem 2.1. Let Σ_N be the covariance matrix of $N^{\frac{1}{2}}(\hat{\beta} - \beta)/\sigma$ for extended least squares. Then

$$C_{1N}\Sigma_N C_{1N} = C_{1N} - (\kappa - 1)[C_{3N} - C_{2N}C_{1N}^{-1}C_{2N}]/4N + (Q_{5N} + Q_{5N}^T)/2N + 0(N^{-3/2}), \tag{2.3}$$

where

$$Q_{5N} = E\ N^{-1} \sum_i \sum_j \sum_k \varepsilon_i \varepsilon_j z_i z_j^T [C_{1N}^{-1}C_{2N} - 2h_j I]h_k$$
$$\times ([z_k^T (\frac{\hat{\beta}_p - \beta}{\sigma})]^2 - 2\varepsilon_k z_k^T (\frac{\beta_p - \beta}{\sigma})) .$$

One of the interesting points about the expansion (2.3) is that it shows the effect of the preliminary estimator $\hat{\beta}_p$ on the covariance matrix, an effect which is suppressed in the first order asymptotic theory (1.4). It is rather difficult to see, but in most instances the term Q_{5N} of (2.3) is rather easy to compute. In fact, if we can write

$$\frac{\hat{\beta}_p - \beta}{\sigma} = N^{-1} \sum_{i=1}^{N} v_i \, \Psi(\varepsilon_i) + O_p(N^{-1}) \tag{2.4}$$
$$= V_N + O_p(N^{-1})$$

for some sequence of vectors $\{v_i\}$ and odd function $\Psi(\cdot)$, then it suffices in (2.3) to replace $(\hat{\beta}_p - \beta)/\sigma$ by V_N and the calculation of Q_{5N} is direct. We examine a few special cases in the next section.

It is possible to obtain a second order expansion for the bias of the usual covariance estimator

$$(N-p)^{-1} \sum_{i=1}^{N} (Y_i - x_i^T\hat{\beta})^2 \exp(-2\,h_i\hat{\theta})\, C_{1N}^{-1}/N ,$$

but the results are not very easy to interpret and will not be presented here.

3. Major Results for Extended Least Squares

It is an easy consequence of the Theorem 2.1 to show that the covariance matrix of generalized least squares $\hat{\beta}_{GLS}$, starting from the ordinary least squares estimate, satisfies

$$\not{Z}_N [\text{for } N^{\frac{1}{2}}(\hat{\beta}_{GLS} - \beta)/\sigma]$$

$$= C_{1N}^{-1} + (\frac{5-\kappa}{4N}) C_{1N}^{-1}(C_{3N} - C_{2N}C_{1N}^{-1}C_{2N}) C_{1N}^{-1} + O(N^{-3/2}) \tag{3.1}$$

$$= C_{1N}^{-1} + (\frac{5-\kappa}{4N}) H_N + O(N^{-3/2}) .$$

Recall that κ is the fourth moment of the standardized errors $\{\varepsilon_i\}$ in (1.1).

It turns out that maximum likelihood has the same expansion for its covariance, i.e.,

$$\not{Z}_N [\text{for } N^{\frac{1}{2}}(\hat{\beta}_{MLE} - \beta)/\sigma]$$

$$= C_{1N}^{-1} + (\frac{5-k}{4N}) H_N + O(N^{-3/2}) . \tag{3.2}$$

The proof is given in the appendix.

Equations (3.1) and (3.2) show that if $\kappa < 5$ as it would be for the normal distribution, then the conventional asymptotic covariance formula (1.4) is conservative, while the opposite occurs for $\kappa > 5$. This phenomenon agrees in spirit with the calculation of Freedman & Peters (1984) in the one-sample problem. In the example of the introduction, the conventional asymptotic formula (1.4) suggested an asymptotic variance for generalized least squares of 0.11 at both the normal and contaminated normal case.

The Monte-Carlo variances were 0.23 and 0.07 respectively. The order in approximation (3.1) suggested variances of 0.173 and 0.027 respectively, while if we make an $N/(N-p)$ adjustment for degrees of freedom these become 0.192 and 0.03. While neither of these is especially accurate, they certainly can be used as a basic indicator of a possible failure in the asymptotic formula.

In various Monte-Carlo studies performed in the last few years, for example Carroll & Ruppert (1983), we have found that the preliminary estimator $\hat{\beta}_p$ can be important, especially for heavier-tailed distributions. To test this, we used the response robust although not bounded influence estimates of Carroll & Ruppert (1982 a,b). Letting $\psi(\cdot)$ be an odd function and $\chi(\cdot)$ an even function, this process involves starting with a preliminary estimate $\hat{\beta}_p$ and then solving

$$\sum_{i=1}^{N} \chi\left(\frac{Y_i - x_i\hat{\beta}_p}{\eta\, e^{\theta h_i}}\right) \binom{1}{h_i} = \binom{0}{0} \text{ for } (\eta,\theta); \quad \sum_{i=1}^{N} \psi\left(\frac{Y_i - x_i^T\hat{\beta}_{CR}}{\hat{\eta}\, e^{\hat{\theta} h_i}}\right) \frac{x_i}{e^{\hat{\theta} h_i}} = 0$$

for $\hat{\beta}_{CR}$.

Here the distribution constant η satisfies

$$E\, \chi(\sigma\varepsilon/\eta) = 0 ,$$

with $\sigma = \eta$ at the normal distribution. We show in the appendix that, starting from $\hat{\beta}_{CR}$ in (1.2a), the extended least squares estimate $\hat{\beta}$ satisfies

$$\not{V}[\text{for } N^{\frac{1}{2}}(\hat{\beta} - \beta)/\sigma] \tag{3.3}$$

$$= C_{1N}^{-1} + (\xi - \kappa)H_N/(4N), \text{ where}$$

$$\xi = 1 + 4(\eta/\sigma)^2\, E\, \psi^2(\sigma\varepsilon/\eta) \quad \{E\, \psi'(\sigma\varepsilon/\eta)\}^{-2} . \tag{3.4}$$

Recalling that by convention the errors $\{\varepsilon_i\}$ in equation (1.1) have variance one, the following general remarks can be made about (3.3) - (3.4). In the normal case, $\xi \approx 5.00$ for most functions $\psi(\cdot)$ used in practice; (3.4) indicates here that there should be little loss in using a robust starting value even in the normal case. For heavier-tailed distributions we will generally have $\xi < 5.00$; comparing (3.4) with (3.1) and (3.2) indicates that starting with a weighted robust estimate can improve the moderate sample size properties of extended least squares. Of course, in this situation one should consider replacing extended least squares by a weighted M-estimator as suggested in Carroll & Ruppert (1982 a).

REFERENCES

Carroll, R.J. (1982). Adapting for heteroscedasticity in linear model. Annals of Statistics 10, 1224-1233.

Carroll, R.J. & Ruppert, D. (1982 a). Robust estimation in heteroscedastic linear models. Annals of Statistics 10, 429-441.

Carroll, R.J. & Ruppert, D. (1982 b). A comparison between maximum likelihood and generalized least squares in a heteroscedastic linear model. Journal of the American Statistical Association 77, 878-882.

Carroll, R.J. & Ruppert, D. (1983). Robust estimators for random coefficient regression models. In Contributions to Statistics, ed. P.K. Sen. Amsterdam: North Holland.

Freedman, D.A. & Peters, S.C. (1984). Bootstrapping a regression equation: some empirical results. Journal of the American Statistical Association 79, 97-106.

Fuller, W.A. & Rao, J.N.K. (1978). Estimation for a linear regression model with unknown diagonal covariance matrix. Annals of Statistics 6, 1149-1158.

Jobson, J.D. & Fuller, W.A. (1980). Least squares estimation when the covariance matrix and parameter vector are functionally related. Journal of the American Statistical Association 77, 595-604.

Toyooka, Y. (1982). Second-order expansion of mean squared error matrix of generalized least squares estimator with estimated parameters. Biometrika 69, 269-273.

Toyooka, Y. (1984). A method for second-order evaluation of risk matrix of GLSE with estimated parameter in trended data. Preprint.

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, ed. J.N. Scivastava. Amsterdam: North Holland.

Appendix

Define

$$B_N = \left\{C_{2N}C_{1N}^{-1}C_{2N} - (\tfrac{1}{2})\ C_{3N}\right\}\ C_{1N}^{-1} .$$

Lemma 1. Under regularity conditions, we have the expansion for extended least squares

$$C_{1N}N^{\frac{1}{2}}(\hat{\beta} - \beta)/\sigma = N^{-\frac{1}{2}} \sum_{i=1}^{N} \varepsilon_i\ z_i + N^{-\frac{1}{2}} \sum_{i=1}^{N} \varepsilon_i\{C_{2N}C_{1N}^{-1} - 2h_i\ I\}\ z_i(\hat{\theta} - \theta)$$

$$+ N^{-\frac{1}{2}} \sum_{i=1}^{N} \varepsilon_i\{B_N - 2h_i C_{2N}C_{1N}^{-1} + 2h_i^2\ I\}\ z_i(\hat{\theta} - \theta)^2 + O_p(N^{-3/2}).$$

Proof of Lemma 1 (Sketch) We have that

$$0 = N^{-\frac{1}{2}} \sum_{i=1}^{N} \left(\frac{Y_i - x_i^T\hat{\beta}}{\sigma\ \exp(\theta h_i)}\right) z_i\ \exp\{-2h_i(\hat{\theta} - \theta)\} .$$

By Taylor series, this means that

$$\{C_{1N} - C_{2N}(\hat{\theta} - \theta) + (\tfrac{1}{2})\ C_{3N}(\hat{\theta} - \theta)^2\}\ N^{\frac{1}{2}}(\hat{\beta} - \beta)/\sigma \tag{A.1}$$

$$= N^{-\frac{1}{2}} \sum_{i=1}^{N} \varepsilon_i z_i\ \{1 - 2h_i(\hat{\theta} - \theta) + 2h_i^2(\hat{\theta} - \theta)^2\}$$

$$+ O_p(N^{-3/2}) .$$

Recalling that $\hat{\theta} - \theta = O_p(N^{-\frac{1}{2}})$ and inverting and expanding the left side of (A.1), we obtain

$$C_{1N}\ N^{\frac{1}{2}}(\hat{\beta} - \beta)/\sigma = \{I_p + C_{2N}C_{1N}^{-1}(\hat{\theta} - \theta) + B_N(\hat{\theta} - \theta)^2\} \tag{A.2}$$

$$\times N^{-\frac{1}{2}} \sum_{i=1}^{N} \varepsilon_i z_i\{1 - 2h_i(\hat{\theta} - \theta) + 2h_i^2(\hat{\theta} - \theta)^2\}$$

$$+ O_p(N^{-3/2}) .$$

Collecting the terms in (A.2) of order $O_p(N^{-1})$ and lower yields Lemma 1.

Lemma 2. Under regularity conditions, for extended least squares the following expansions holds:

$$\hat{\theta} - \theta = N^{-1} \sum_{i=1}^{N} h_i \left\{(\varepsilon_i^2 - 1)/2 - \varepsilon_i z_i^T(\hat{\beta}_p - \beta)/\sigma + [z_i^T(\hat{\beta}_p - \beta)/\sigma]^2/2\right\}$$

$$+ (1/4)\ \mu_{3N} \left\{N^{-1} \sum h_i(\varepsilon_i^2 - 1)\right\}^2 -$$

$$- (\tfrac{1}{2}) \left\{N^{-1} \sum h_i(\varepsilon_i^2 - 1)\right\} \left\{N^{-1} \sum h_i^2(\varepsilon_i^2 - 1)\right\} + O_p(N^{-3/2}) .$$

Proof of Lemma 2 (Sketch)

The maximum likelihood estimate of θ given $\hat{\beta}_p = \beta$ satisfies

$$0 = N^{-1} \sum_{i=1}^{N} h_i \left(\frac{Y_i - x_i^T \hat{\beta}_p}{\sigma \exp(\theta h_i)}\right)^2 \exp\{-2h_i(\hat{\theta} - \theta)\} , \tag{A.3}$$

recalling that $\Sigma h_i = 0$ by convention. Writing $\gamma = (\hat{\beta}_p - \beta)/\sigma$, we have that

$$0 = N^{-1} \sum_{i=1}^{N} h_i(\varepsilon_i - z_i^T\gamma)^2 \exp\{-2h_i(\hat{\theta} - \theta)\} . \tag{A.4}$$

Recalling that $\gamma = O_p(N^{-\frac{1}{2}})$ and $\hat{\theta} - \theta = O_p(N^{-\frac{1}{2}})$, after tedius algebra we arrive at the expansion

$$0 = \mu_{3N}(\hat{\theta} - \theta)^2 - \{1 + A_{3N}/\sqrt{N}\} \quad (\hat{\theta} - \theta) \tag{A.5}$$

$$+ (\tfrac{1}{2})\ \{A_{1N}/\sqrt{N} + A_{2N}/N\} + O_p(N^{-3/2}) ,$$

where

$$A_{1N} = N^{-\frac{1}{2}} \sum_{i=1}^{N} h_i(\varepsilon_i^{\,2} - 1)$$

$$A_{2N} = N^{-\frac{1}{2}} \sum_{i=1}^{N} (-2h_i)\ \varepsilon_i\ z_i^{\,T}\ N^{\frac{1}{2}}\ \gamma$$

$$A_{3N} = N^{-\frac{1}{2}} \sum_{i=1}^{N} h_i^{\,2}(\varepsilon_i^{\,2} - 1)\ .$$

Note that $A_{1N} = O_p(1)$, $A_{2N} = O_p(1)$ and $A_{3N} = O_p(1)$. The hard case is $\mu_{3N} \neq 0$, in which case a quadratic expansion is necessary and we get from equation (A.5)

$$2\ \mu_{3N}(\hat{\theta} - \theta) \tag{A.6}$$

$$= 1 + A_{3N}/\sqrt{N}$$

$$- \left\{1 + \frac{2(A_{3N} - \mu_{3N}\ A_{1N})}{\sqrt{N}} + \frac{A_{3N}^{\ 2} - 2\ \mu_{3N}\ A_{2N}}{N}\right\}^{\frac{1}{2}} + O_p(N^{-3/2}).$$

Further tedious algebra from (A.6) completes the proof.

Proof of (3.2)

If the weights were known a priori, the optimal weighted least squares estimate would satisfy

$$N^{\frac{1}{2}}(\hat{\beta}_{WLS} - \beta)/\sigma = C_{1N}^{-1}\ N^{-\frac{1}{2}} \sum_{i=1}^{N} z_i\ \varepsilon_i\ . \tag{A.7}$$

From Carroll and Ruppert (1982a), we have that

$$\hat{\beta}_{WLS} - \hat{\beta}_{MLE} = O_p(N^{-1}) \ . \qquad \text{(A.8)}$$

In the expression for Q_{5N} in the Theorem, we have here that $\hat{\beta}_p = \hat{\beta}_{MLE}$, and it follows from (A.8) that we can replace $\hat{\beta}_p$ by $\hat{\beta}_{WLS}$ and retain the same order of the expansion. Now using (A.7), it is easy to prove (3.2).

<u>Proof of (3.3)</u>

The argument follows the same lines as that for (3.2), with the exception that by Carroll and Ruppert (1982a),

$$N^{\frac{1}{2}}(\hat{\beta}_{CR} - \beta)/\eta = \{E\ \psi'(\sigma\varepsilon/\eta)\}^{-1}\ C_{1N}^{-1}\ N^{-\frac{1}{2}} \sum_{i=1}^{N} \psi(\sigma\varepsilon_i/\eta) z_i + O_p(N^{-\frac{1}{2}}) \ .$$