M-ESTIMATES FOR THE HETEROSCEDASTIC LINEAR MODEL

David Ruppert*
and
Raymond J. Carroll**
University of North Carolina at Chapel Hill

We treat the linear model $Y_i = \underline{C}_i^T \underline{\beta} + Z_i$ where $\underline{C}_i$ is a known vector, $\underline{\beta}$ is an unknown parameter, and Var Z_i is a function of $|\underline{C}_i^T \underline{\beta}|$ which is known, except for a parameter $\underline{\theta}$. For simultaneous M-estimates, $\hat{\underline{\beta}}$ and $\hat{\underline{\theta}}$, we show that $(\hat{\underline{\theta}}-\underline{\theta}) = O_p(N^{-\frac{1}{2}})$, and find the limit distribution of $N^{\frac{1}{2}}(\hat{\underline{\beta}}-\underline{\beta})$. For the special case of least squares estimation, this limit distribution is the same as the limit distribution of the weighted least squares using the weights, $w_i = (\text{Var } Z_i)^{-1}$, and in general the distribution is that of a "weighted M-estimate" using these weights. Moreover, the covariance matrix of the limit distribution can be consistently estimated, so large sample confidence ellipsoids and tests of hypotheses concerning $\underline{\beta}$ are feasible.

AMS 1970 subject classifications Primary 62J05; Secondary 62G35.

Key words and phrases. Weighted M-estimates, weighted least squares, estimated weights, linear regression in scale, robustness, asymptotic representation, asymptotically normal, confidence ellipsoids.

*Supported by National Science Foundation Grant NSF MCS78-01240

**Supported by Air Force Office of Scientific Research under Grant AFOSR-75-2796

1. Introduction. Consider the linear model

$$Y_i = \tau_i + \sigma_i e_i$$

and

$$\tau_i = \underline{C}_i^T \underline{\beta}, \text{ for } i = 1,\ldots,N,$$

where $e_1,\ldots,e_N$ are i.i.d. with distribution F, $\underline{C}_i$ $(=\underline{C}_{iN})$ are known elements in R^p, $\underline{\beta}$ is a vector parameter in R^p, and $\sigma_1,\ldots,\sigma_N$ are constants, which express the possible heteroscedasticity of the model. If F were a normal distribution function and the σ_i were known then the minimum variance unbiased estimator of $\underline{\beta}$ would be the weighted least squares estimator $\hat{\underline{\beta}}_W$, which is the solution to

$$\sum_{i=1}^{N} \rho((Y_i - \underline{C}_i^T \hat{\underline{\beta}}_W)/\sigma_i) = \text{minimum}, \tag{1.1}$$

where $\rho(x) = x^2$. If F has heavy tails compared with the Gaussian distribution or gross errors are possibly present in the data, then one may wish to replace $\rho(x) = x^2$ with a function such that $\psi = \rho'$ is bounded. Such estimators have been called M-estimators by Huber (1964, 1973), because they are generalizations of maximum likelihood estimators. If the σ_i are not known it may still be possible to estimate them. For example, Fuller and Rao (1978) consider the case where the Y_i occur in groups for which σ_i is constant, and Box and Hill (1974) assume that

$$\sigma_i(\theta) = \sigma_i = \sigma|\tau_i|^\theta \tag{1.2}$$

for an unknown parameter θ and estimate both θ and $\underline{\beta}$ by Bayesian methods. Both Fuller and Rao and Box and Hill treat only Gaussian errors.

Anscombe (1961) has proposed tests of heteroscedasticity, and Bickel (1978) has developed robust versions of these, but neither has considered modifying the estimate of $\underline{\beta}$ if the null hypothesis of homoscedasticity is rejected.

In many empirical studies, one finds that the dispersion of the residuals increases with the magnitude of the fitted values, so that it is reasonable to assume that $\sigma_i = \sigma(|\tau_i|)$ for some nondecreasing function σ. While the power family (1.2) may be sufficiently rich to model this phenomenon (although it has obvious difficulties in practice when $\tau_i = 0$), we prefer to study the general class,

$$\sigma_i = \exp(\underline{h}(\tau_i)^T \underline{\theta}), \tag{1.3}$$

where $\underline{\theta}$ is a parameter in R^q and $\underline{h}$ is a known function from R^1 to R^q. Notice that choosing $h_1(\tau_i) \equiv 1$ (h_1 is the first coordinate of $\underline{h}$), $\exp(\theta_1)$ becomes a scale parameter ; as a result, our estimate of $\underline{\beta}$ will be scale equivariant regardless of the choice of ρ. Two choices of $\underline{h}$ which have no practical difficulties when $\tau_i = 0$ are $\underline{h}(\tau) = (1, \log(1+|\tau|))$ and $\underline{h}(\tau) = (1, |\tau|)$.

To motivate our method, suppose F is standard normal so that the log-likelihood is

$$L(Y;\underline{\theta},\underline{\beta}) = -\frac{1}{2}\sum_{i=1}^{N}\{\log 2\pi + \log \sigma_i^2 + ((Y_i-\tau_i)/\sigma_i)^2\} .$$

If $\underline{\theta}$ were known, (1.1) would yield the MLE for $\underline{\beta}$, while if $\underline{\beta}$ were known, the MLE of $\underline{\theta}$ would solve

$$\sum_{i=1}^{N}\{((Y_i-\tau_i)/\sigma_i(\theta))^2 - 1\}\, \underline{h}(\tau_i) = 0. \tag{1.4}$$

A reasonable computational alternative to solving (1.1) and (1.4) simultaneously might consist of (i) obtaining a preliminary estimate of $\underline{\beta}$ (such as the least squares estimate) and hence estimates for $\tau_1,\ldots,\tau_n$ (ii) solve (1.4) using these estimates, thus obtaining estimates of $\sigma_1,\ldots,\sigma_n$, which (iii) are used to solve (1.1).

We thus suggest the following procedure. First, a preliminary estimate $\hat{\underline{\beta}}_0$ of $\underline{\beta}$ is calculated and assumed to satisfy

$$(\hat{\underline{\beta}}_0 - \underline{\beta}) = O_p(N^{-\frac{1}{2}}). \tag{1.5}$$

Examples of estimates satisfying (1.5) are given in Section 5. At the second stage, define $t_i = \underline{C}_i^T \hat{\underline{\beta}}_0$ and obtain robustified estimates of $\underline{\theta}$ by solving the following analogue of (1.4):

$$\sum_{i=1}^{N} \{\psi^2((Y_i - t_i) \exp(-\underline{h}(t_i)^T \hat{\underline{\theta}})) - \xi\} \underline{h}(t_i) = 0, \tag{1.6}$$

where ψ is monotone nondecreasing and $\xi = E\psi^2(Z_1)$, the expectation being taken under the standard normal distribution. Clearly, $\psi(x) = x$ leads to (1.4). At the third stage, we now solve a robust version of (1.1):

$$\sum_{i=1}^{N} \psi((Y_i - \underline{C}_i^T \hat{\underline{\beta}})/\hat{\sigma}_i) \underline{C}_i = 0, \tag{1.7}$$

where

$$\hat{\sigma}_i = \exp(-h(t_i)^T \hat{\underline{\theta}}). \tag{1.8}$$

Our main result (Theorem 4.1) lists conditions under which the limit distribution of $\hat{\underline{\beta}}$ defined by (1.5)-(1.8) is the same as that of the estimate which could be found by solving (1.7) *when* $\sigma_1, \ldots, \sigma_N$ *are known*. In Section 2 we introduce notation and assumptions. Section 3 demonstrates that

$$\hat{\theta} - \theta = O_p(N^{-\frac{1}{2}}) . \tag{1.9}$$

In Section 4 we state and prove the main result.

Remarks.

A If one assumes homoscedasticity ($\sigma_i \equiv \sigma$), then $\underline{h}(\tau_i) = 1$. The estimate $\hat{\underline{\beta}}$ becomes an ordinary robust regression estimate with preliminary estimate of scale given by (1.5)-(1.6). See Maronna and Yohai (1979) for further details.

B We do not know if iterating (1.5)-(1.8) will lead to convergence. Further, we do not know whether simultaneous solution of (1.6)-(1.7) is possible.

C Throughout the paper, we will use the following convention. For a real function f on a space $\mathcal{X}$, we say that x_0 solves $f(x) = 0$ if

$$|f(x_0)| < 2 \inf_{\mathcal{X}} |f(x)|. \tag{1.10}$$

With this convention, all our estimators will exist but need not be unique. However, our asymptotic results hold for *every* appropriate sequence of estimators.

D The Box and Cox (1964) transformations form an alternative method for dealing with heteroscedasticity, as well as other deviations from the normal linear model. Defining

$$\begin{aligned} Y^{(\lambda)} &= (Y^{\lambda}-1)/\lambda \qquad && (\lambda \neq 0) \\ &= \log Y && (\lambda = 0), \end{aligned}$$

they postulate that for some λ, $Y^{(\lambda)}$ satisfies a homoscedastic normal linear model. Their methodology is based upon an entirely different model from ours, but a practitioner might consider using both on a given data set. Carroll (1978) and Bickel and Doksum (1978) independently studied both Box-Cox and robustified Box-Cox methods and concluded that variances of the estimated coefficients in the linear model are often much larger when λ is estimated than when it is known. Therefore, confidence intervals and tests of hypotheses which are constructed as if λ were known and not estimated are invalid.

In contrast, we show in Section 4 that, if the errors, e_i, are symmetrically distributed *or* $\psi(x) = x$, then for our method, the variances of estimated coefficients when $\underline{\theta}$ is estimated are similar to those when $\underline{\theta}$ is known, and confidence intervals and tests can be validly constructed as if $\underline{\theta}$ were known.

2. Assumptions and notations. Let ψ be a nondecreasing function from R^1 to R^1 which satisfies

B1. $$E\psi(e_1) = 0,$$

B2. $$0 < E\psi^2(e_1) = \xi < \infty,$$

for some $\alpha > 0$

B3. $$E\psi((e_1+r)(1+s)) = A(\psi)\ r + O(|r|^{1+\alpha} + |s|^{1+\alpha})$$

as $r \to 0$ and $s \to 0$ with $A(\psi) > 0$,

B4. $$E\psi^2((e_1+r)(1+s)) - \xi = A(\psi^2)\ s + O(|r|^{1+\alpha} + |s|^{1+\alpha})$$

as $r \to 0$ and $s \to 0$ with $A(\psi^2) > 0$,

B5. there exists C_0 such that for all $\delta < 1$

$$\limsup_{k\to 0} E\{\sup|\phi((e_1+r)(1+s)) - \phi((e_1+r')(1+s'))|:$$

$$|r|,|r'|,|s|,|s'| \le k \text{ and } |r-r'|,\ |s-s'| \le k\delta\}$$

$$\le C_0\delta$$

for both $\phi = \psi$ and $\phi = \psi^2$, and

B6. $$\lim_{r,s\to 0} E(\phi((e_1+r)(1+s)) - \phi(e_1))^2 = 0$$

for both $\phi = \psi$ and $\phi = \psi^2$.

The function $\underline{h}$ from R^1 to R^q, $\underline{C}_i$ in R^p and $\tau_i = \underline{C}_i^T\ \underline{\beta}$, satisfy

B7. $$\lim_{N\to 0} (\sup_{i\le N} (||\underline{C}_i|| + h(\tau_i))N^{-\frac{1}{2}}) = 0,$$

B8. $$\sup_N (N^{-1} \sum_{i=1}^{N} ||\underline{C}_i||^2 + \underline{h}(\tau_i)^2) < \infty ,$$

B9. Letting λ_n be the minimum eigenvalue of

$$H_N = N^{-1} \sum_{i=1}^{N} \underline{h}(\tau_i)\ \underline{h}(\tau_i)^T$$

then

$$\liminf_{N\to\infty} \lambda_N = \lambda_\infty > 0$$

B10. $\underline{h}$ is Lipschitz continuous on an interval (possibly infinite), I, such that τ_i is in I for all i and N, B11. $\inf_N \inf_{i \le N} \underline{h}(\tau_i)^T \underline{\theta} > -\infty$

and

B12. letting $\underline{d}_i = \underline{C}_i \exp(-\underline{h}(\tau_i)^T \underline{\theta})$,

$$N^{-1} \sum_{i=1}^{N} \underline{d}_i \underline{d}_i^T = S_N \to S \quad \text{(positive definite)}.$$

Remark. The monotonicity of ψ is needed only to prove $\sqrt{N}$ - consistency of $\hat{\underline{\beta}}$ and $\hat{\underline{\theta}}$.

The conditions B1-B6 are notationally complex but widely applicable. They clearly hold for $\rho(x) = x^2/2$ ($\psi(x) = x$) if $Ee_1 = 0$, $Ee_1^4 < \infty$, where $A(\psi) = 1$, $A(\psi^2) = 2Ee_1^2$. If F is symmetric and ψ is odd, B1-B6 can be verified if (i) ψ is constant outside an interval (ii) ψ is Lipschitz continuous and twice boundedly differentiable except possibly at a finite number of points, $a_1, \ldots, a_k$ (iii) F is Lipschitz continuous in neighborhoods of $a_1, \ldots, a_k$ and (iv) $E\psi^2(e_1)$, $E\psi'(e_1)$ and $Ee_1\psi(e_1)\,\psi'(e_1)$ are all positive. Then $A(\psi) = E\psi'(e_1)$ and $A(\psi^2) = 2Ee_1\psi(e_1)\,\psi'(e_1)$.

For the power model (1.2), $\underline{h}(x) = (1, \log |x|)$, $\underline{\theta} = (\log \sigma, \theta)$ and B10, B11 hold if

$$\inf_N \inf_{1 \le i \le N} |\tau_i| > 0. \tag{2.1}$$

Since the power model (1.2) makes little sense when $\tau_i = 0$, (2.1) is reasonable. To avoid this difficulty, in Section 1 we suggested $\underline{h}(x) = (1, \log(1+|x|))$, for which B10, B11 hold. The other example $\underline{h}(x) = (1, |x|)$ will satisfy B10, B11 as long as the $\{\tau_i\}$ are uniformly bounded.

3. Estimation of $\underline{\theta}$.

Theorem 3.1. Suppose B1-B8 hold, $(\hat{\underline{\beta}}_0 - \underline{\beta}) = O_p(N^{-\frac{1}{2}})$, $t_i = \underline{C}_i^T \hat{\underline{\beta}}_0$ and $\hat{\underline{\theta}}$ is *any* solution (see 1.10) to

(3.1) $$\sum_{i=1}^{N} \{\psi^2((Y_i - t_i)\exp(-\underline{h}(t_i)^T \hat{\underline{\theta}})) - \xi\} \underline{h}(t_i) = 0.$$

Then

(3.2) $$\hat{\underline{\theta}} - \underline{\theta} = O_p(N^{-\frac{1}{2}}).$$

Proof. For Δ_1 in R^p, Δ_2 in R^q and

$\underline{\Delta} = (\Delta_1, \Delta_2)$, define $\underline{h}_i(\underline{\Delta}) = \underline{h}(\tau_i + \underline{C}_i^T \Delta, N^{-\frac{1}{2}})$,

(3.3) $$\alpha_i^{(1)}(\underline{\Delta}) = \exp\left\{-\underline{h}_i(\underline{\Delta})^T \Delta_2 N^{-\frac{1}{2}} + (\underline{h}_i(\underline{0}) - \underline{h}_i(\underline{\Delta}))^T \underline{\theta}\right\} - 1$$

(3.4) $$\alpha_i^{(2)}(\underline{\Delta}) = N^{-\frac{1}{2}} \underline{d}_i^T \Delta_1 \qquad \text{(see B12)} .$$

Then let $\phi_i(x,y,z) = \psi^2((x+z)(1+y)) - \xi$ and define the process

$$W_N(\underline{\Delta}) = -N^{-\frac{1}{2}} \sum_{i=1}^{n} \phi(e_i, \alpha_i^{(1)}(\underline{\Delta}), \alpha_i^{(2)}(\underline{\Delta}))\, \underline{h}_i(\underline{\Delta}).$$

Note that (3.1) can be rewritten as

$$W_N(N^{\frac{1}{2}}(\hat{\underline{\beta}}_0 - \underline{\beta}),\ N^{\frac{1}{2}}(\hat{\underline{\theta}} - \underline{\theta})) = 0.$$

By B1 and Chebyshev's inequality,

(3.5) $$W_N(\underline{0}) = O_p(1),$$

so that by our convention (1.10),

$$W_N(N^{\frac{1}{2}}(\hat{\underline{\beta}}_0 - \underline{\beta}),\ N^{\frac{1}{2}}(\hat{\underline{\theta}} - \underline{\theta})) = O_p(1).$$

We can therefore prove (3.2) by showing that for each $M_1>0$, $\varepsilon>0$, $Q>0$ there exists $M_2>0$ such that

(3.6) $$p\left(\inf_{\substack{\|\Delta_1\|\le M_1\\ \|\Delta_2\|\ge M_2}} \|W_N(\underline{\Delta})\| > Q\right) \ge 1 - \varepsilon$$

We will prove (3.6) by modifying the proof of Jurečková's (1977) Lemma 5.2.

We first apply Theorem A.1 of the appendix, with $\underline{x}_i = \underline{h}_i(\underline{0})$ and $\alpha_i^{(3)}(\underline{\Delta}) = \underline{h}_i(\underline{\Delta}) - \underline{h}_i(\underline{0})$. Then B1-B8, B10, B11 imply that the assumptions of Theorem A.1 are met with $g_i \equiv 1$, $A(\phi,i) = A(\psi^2)$, so that for all $M>0$,

$$\sup_{\|\underline{\Delta}\|\le M}\left\|W_N(\underline{\Delta}) - W_N(\underline{0}) - A(\psi^2)N^{-\frac{1}{2}}\sum_{i=1}^{n} \underline{h}(\tau_i)\alpha_i^{(1)}(\underline{\Delta})\right\| = o_p(1).$$

By a Taylor series expansion,

$$\alpha_i^{(1)}(\underline{\Delta}) = -N^{-\frac{1}{2}}\underline{h}(\tau_i)^T\,\Delta_2 + (\underline{h}_i(\underline{0})-\underline{h}_i(\underline{\Delta}))^T\,\underline{\theta} + o_p(N^{-\frac{1}{2}}).$$

Thus, by B9, setting

$$G_N(\underline{\Delta}) = N^{-\frac{1}{2}}\sum_{i=1}^{N} \underline{h}(\tau_i)(\underline{h}_i(\underline{0})-\underline{h}_i(\underline{\Delta}))^T\,\underline{\theta},$$

(3.7) $$\sup_{\|\underline{\Delta}\|\le M}\|W_N(\underline{\Delta}) - W_N(\underline{0}) - A(\psi^2)H\Delta_2 + G_N(\underline{\Delta})\| = o_p(1).$$

Now fix $\varepsilon>0$, $M_1>0$, $Q>0$. Use (3.5) to choose γ such that

$$p(\|W_N(\underline{0})\| \ge \gamma/2) < \varepsilon/2.$$

Define

$$D = \sup_N \sup_{\|\Delta_1\|\le M_1} \|G_N(\underline{\Delta})\|.$$

Then $D<\infty$ (G_N depends only on Δ_1). Define M_2 by

$$[A(\psi^2)\lambda_\infty M_2/2 - \gamma - D] = Q.$$

Using B9 and (3.7), find N_0 such that $\lambda_N \ge \lambda_\infty/2$ and

$$p\left(\sup_{\substack{\|\Delta_2\|=M_2\\ \|\Delta_1\|\le M_1}} \|W_N(\underline{\Delta}) - W_N(\underline{0}) - A(\psi^2)H\Delta_2 - G_N(\underline{\Delta})\| < \frac{\gamma}{2}\right) \ge 1 - \varepsilon/2 \quad (N\ge N_0).$$

If $||\Delta_2|| = M_2$, $||\Delta_1|| \le M_1$, and $N \ge N_0$, then with probability at least $1-\varepsilon$,

$$\Delta_2 W_N(\underline{\Delta})$$

$$\ge -M_2||W_N(\underline{0})|| + \Delta_2^T H\Delta_2 A(\psi^2) - M_2 D - M_2\gamma/2$$

$$\ge [A(\psi^2)\ \lambda_\infty\ M_2/2 - \gamma - D]M_2 = QM_2.$$

Since ψ is nondecreasing, $\Delta_2 W_N(\Delta_1,\Delta_2 s)$ is a nondecreasing function of s. Thus, $||\Delta_2|| \ge M_2$ implies

$$\Delta_2 W_N(\underline{\Delta}) \ge \Delta_2 W_N(\Delta_1, M_2\Delta_2||\Delta_2||^{-1})$$

$$\ge (||\Delta_2||/M_2)(M_2\Delta_2||\Delta_2||^{-1}W_N(\Delta_1, M_2\Delta_2||\Delta_2||^{-1}))$$

$$\ge ||\Delta_2||Q$$

Thus,

$$p\left(\inf_{\substack{||\Delta_1||\le M_1 \\ ||\Delta_2||\ge M_2}} \frac{\Delta_2 W_N(\underline{\Delta})}{||\Delta_2||} \ge Q\right) \ge 1-\varepsilon$$

which with the Cauchy-Schwarz inequality proves (3.2). □

4. Estimation of $\underline{\beta}$. The limiting distribution of $\hat{\underline{\beta}}$ is a simple consequence of the following representation.

Theorem 4.1. Suppose B1 to B12 hold and $A(\psi) > 0$. Then if $\hat{\underline{\beta}}$ is *any* solution (see (1.10)) to

$$\sum_{i=1}^{N} \psi((Y_i - \underline{C}_i^T\hat{\underline{\beta}})\exp(-\underline{h}(t_i)^T\hat{\underline{\theta}})\underline{C}_i \exp(-h(t_i)^T\hat{\underline{\theta}}) = 0 \tag{4.1}$$

with $t_i = \underline{C}_i^T\hat{\underline{\beta}}_0$, $(\hat{\underline{\beta}}_0 - \underline{\beta}) = O_p(N^{-\frac{1}{2}})$, and $\hat{\underline{\theta}}$ satisfying (3.2), then

$$N^{\frac{1}{2}}(\hat{\underline{\beta}} - \underline{\beta}) = N^{-\frac{1}{2}}\sum_{i=1}^{N} S^{-1}d_i\psi(e_i)(A(\psi))^{-1} + o_p(1). \tag{4.2}$$

This result, strengthened so that the remainder is o(1) almost surely, has been established in the homoscedastic case by Carroll and Ruppert (1979).

Proof. For $\underline{\Delta}_1$ and $\underline{\Delta}_3$ in R^p, $\underline{\Delta}_2$ in R^q and $\underline{\Delta} = (\underline{\Delta}_1, \underline{\Delta}_2, \underline{\Delta}_3)$, define

$$\alpha_i^{(1)}(\Delta) = N^{-\frac{1}{2}} \underline{d}_i^T \underline{\Delta}_1$$

$$\underline{h}_i(\underline{\Delta}) = h(\tau_i + \underline{C}_i^T \underline{\Delta}_3 N^{-\frac{1}{2}}),$$

$$\alpha_i^{(2)}(\Delta) = \exp(-\underline{h}_i(\underline{\Delta})^T \underline{\Delta}_2 N^{-\frac{1}{2}} + (\underline{h}_i(0) - h_i(\underline{\Delta}))^T \theta) - 1,$$

and

$$\underline{\alpha}_i^{(3)}(\underline{\Delta}) = \underline{d}_i \alpha_i^{(2)}(\underline{\Delta}).$$

Define the process

$$U_N(\underline{\Delta}) = N^{-\frac{1}{2}} \sum_{i=1}^{N} \psi((e_i + \alpha_i^{(1)}(\underline{\Delta}))(1 + \alpha_i^{(2)}(\underline{\Delta}))(\underline{d}_i + \underline{\alpha}_i^{(3)}(\Delta)).$$

Note that (4.1) can be rewritten as

(4.3) $$U_N(N^{\frac{1}{2}}(\hat{\underline{\beta}} - \underline{\beta}),\ N^{\frac{1}{2}}(\hat{\underline{\theta}} - \underline{\theta}),\ N^{\frac{1}{2}}(\hat{\underline{\beta}}_0 - \underline{\beta})) = 0.$$

Letting $g_i \equiv 1, \phi_i(e_i, r, s) = \psi((e_i + r)(1 + s))$, and $A(\psi, i) = A(\psi)$, the conditions of Theorem A.1 are implied by B1, B3, B5 to B8, B11, and B12, so for all $M>0$

(4.4) $$\sup_{||\Delta|| \le M} ||U_N(\underline{\Delta}) - U_N(\underline{0}) - A(\psi)\, S\underline{\Delta}_1|| = o_p(1).$$

Now by Chebyshev's theorem and B1,

$$U_N(\underline{0}) = O_p(1).$$

Therefore, if we set

$$\underline{\Delta}^* = -(A(\psi)S)^{-1} U_N(\underline{0}), \text{ then}$$

$$\underline{\Delta}^* = O_p(1)$$

and therefore, by (4.4)

$$U(\underline{\Delta}^*) = o_p(1).$$

Consequently, by (4.3) and our convention (1.10),

(4.5) $$U_N(N^{\frac{1}{2}}(\hat{\underline{\beta}}-\underline{\beta}),\ N^{\frac{1}{2}}(\hat{\underline{\theta}}-\underline{\theta}),\ N^{\frac{1}{2}}(\hat{\underline{\beta}}_0-\underline{\beta})) = o_p(1).$$

By Theorem 3.1, $(\hat{\underline{\theta}}-\underline{\theta}) = O_p(N^{-\frac{1}{2}})$, so we need only establish that

(4.6) $$(\hat{\underline{\beta}}-\underline{\beta}) = O_p(N^{-\frac{1}{2}})$$

to conclude from (4.3) and (4.4) that (4.2) holds. But by (4.5), (4.6) holds if for each $\eta>0$, $\varepsilon>0$, and M_1 there exists M_2 satisfying

(4.7) $$P(\inf_{||\underline{\Delta}_1||\geq M_2}\ \inf_{||\underline{\Delta}_2||\leq M_1}\ \inf_{||\underline{\Delta}_3||\leq M_1} ||U_N(\underline{\Delta})||>\eta) > 1-\varepsilon.$$

Now (4.7) follows from (4.4) in a manner quite similar to Jurečková's (1977) proof of her Lemma 5.2. □

Corollary 4.2. If $E\psi^2(e_1) < \infty$ and the assumptions of Theorem 4.1 are met, then

$$N^{\frac{1}{2}}(\hat{\underline{\beta}}-\underline{\beta}) \xrightarrow{D} N(0,\ S^{-1}E\psi^2(e_1)\ A(\psi)^{-2}) .$$

Proof. Use (4.2) and Loève's (1963, p 316) Normal Convergence Criterion to show that $N^{\frac{1}{2}}(\hat{\underline{\beta}}-\underline{\beta})^T\underline{x} \xrightarrow{D} N(0,\ x^TS^{-1}\ x\ E\psi^2(e_1)\ \ A(\psi)^{-2}\ )$ for each x in R^p. Since B7, B11 and B12 show that

$$\lim_{N\to\infty}(\sum_{i=1}^{N}||\underline{d}_i||^2)^{-1}||d_i||^2 = 0,$$

it is easy to see that the criterion is met. □

Recall that $A(\psi) = E\psi'(e_1)$, for commonly used ψ.

Define

$$\hat{\sigma}_i = \exp(\underline{h}(\underline{C}_i^T\ \hat{\underline{\beta}}_0)^T\ \hat{\underline{\theta}})$$

(one could use $\hat{\underline{\beta}}$ instead of $\hat{\underline{\beta}}_0$). Then when $(\hat{\underline{\theta}}-\underline{\theta}) = O_p(N^{-\frac{1}{2}})$ and $(\hat{\underline{\beta}}_0-\underline{\beta}) = O_p(N^{-\frac{1}{2}})$, B7 and B10 imply

$$\sup_{i\leq N} |\hat{\sigma}_i-\sigma_i| = o_p(1).$$

Suppose for both $\phi = \psi$ and $\phi = \psi'$

$$\lim_{\varepsilon \to 0} E \sup\{|\phi((1+x)(e_1+y)) - \phi(e_1)| : |x| \le \varepsilon \text{ and } |y| \le \varepsilon\} = 0.$$

Then it is easy to prove that

$$|N^{-1} \sum_{i=1}^{N} \phi(\hat{\sigma}_i^{-1} (Y_i - \underline{C}_i^T \hat{\underline{\beta}})) - \phi(e_i)| = o_p(1),$$

and by the strong law of large numbers,

$$N^{-1} \sum_{i=1}^{N} \phi(e_i) \to E\phi(e_i).$$

Moreover,

$$||N^{-1} \sum_{i=1}^{N} \hat{\sigma}_i^{-2} \underline{C}_i \underline{C}_i^T - N^{-1} \sum_{i=1}^{N} \underline{d}_i \underline{d}_i^T||$$

$$\le \sup_{i \le N} |\hat{\sigma}_i - \sigma_i| (N^{-1} \sum_{i=1}^{N} ||C_i||^2) = o_p(1)$$

(with $||A||$ equal to, say, the Euclidean norm of the matrix A),

so

$$N^{-1} \sum_{i=1}^{N} \hat{\sigma}_i^{-2} \underline{C}_i \underline{C}_i^T \xrightarrow{P} S.$$

Therefore, $E\psi^2(e_1) A(\psi)^{-2} S$ can be consistently estimated, and large sample confidence ellipsoids and tests of hypotheses for $\underline{\beta}$ can be constructed.

5. $\underline{\sqrt{N}}$ - consistency of preliminary estimators. Since we require that our preliminary estimate satisfy (1.5), we now give conditions which insure that an M-estimate $\hat{\underline{\beta}}_0$ solving

$$\sum_{i=1}^{N} \psi(Y_i - \underline{C}_i^T \hat{\underline{\beta}}_0) \underline{C}_i = 0$$

satisfy (1.5). This $\hat{\underline{\beta}}_0$ is not scale equivariant, but this does not affect the limit distribution of $\hat{\underline{\beta}}$.

B13. ψ is odd

B14. ψ is Lipschitz continuous with Lipschitz constant L_ψ

B15. the Radon-Nikodyn derivative ψ' satisfies $E(\psi(x(e_1+y)) - \psi(xe_1) - \psi'(xe_1)xy) \leq K|xy|^{1+\alpha}$ for some K and $\alpha>0$ and all x and y

B16. $N^{-\frac{1}{2}}(\sigma_i + 1)||\underline{d}_i|| \to 0$

B17. $\sup_N (N^{-1} \sum_{i=1}^{N} (\sigma_i^2 + 1)||\underline{d}_i||^2) < \infty$

B18. the minimum eigenvalue, λ_N, of

$$N^{-1} \sum_{i=1}^{N} \sigma_i \; E\psi^1(\sigma_i e_1) \; \underline{d}_i \underline{d}_i^T$$

satisfies $\liminf_{N\to\infty} \lambda_N > 0$.

<u>Theorem 5.1.</u> Under B13 to B18, (1.9) holds.

<u>Proof</u>. We will apply Theorem A.1 with $\phi_i(e_1,r,s) = \psi(\sigma_i(e_1-r))$(so ϕ_i does not depend on s), $g_i = \sigma_i$, $k_i = N^{-\frac{1}{2}}||\underline{C}_i||$, $A(\psi,i) = \sigma_i \; E\psi^1(\sigma_i e_1)$, $\alpha_i^{(3)} \equiv 0$, and $\alpha_i^{(2)}$ left undefined. By B13 and B14, (A.3) holds. By B15, for all r and r'

$$|\phi_i(e_1,r) - \phi_i(e_1,r')| \leq K \, \sigma|r-r'|$$

so that (A.4) and (A.5) hold. Also, (A.5) is implied by B15, and (A.1), (A.2), (A.7), (A.8), and (A.9) are easily checked. Therefore, for all M>0,

$$\sup_{||\Delta|| \leq M} ||N^{-\frac{1}{2}} \sum_{i=1}^{N} \psi(\sigma_i(e_i - \underline{d}_i^T \Delta) - N^{-\frac{1}{2}} \sum_{i=1}^{N} \psi(\sigma_i e_i)$$

$$+ N^{-1} \sum_{i=1}^{N} \sigma_i \; E\psi^1(\sigma_i e_1) \; \underline{d}_i \underline{d}_i^T \Delta|| = o_p(1).$$

Now (5.1) follows exactly as Jurečková's (1977) Lemma 5.2, since $Y_i - \underline{C}_i^T \hat{\underline{\beta}}_0 = \sigma_i(e_i - \underline{C}_i^T(\hat{\underline{\beta}}_0 - \underline{\beta}))$. □

APPENDIX

The following general theorem will be used when studying $\hat{\underline{\beta}}_0$, $\hat{\underline{\theta}}$ and $\hat{\underline{\beta}}$.

Theorem A.1. Let g_i and k_i ($=g_{iN}$ and k_{iN}) be sequences of positive constants such that

(A.1) $$\lim_{N\to\infty} (\sup_{i\le N} k_i + k_i g_i)) = 0$$

and

(A.2) $$\sup_N \left(\sum_{i=1}^{N} k_i^2 + k_i^2 g_i^2 \right) = c_1 < \infty .$$

Let ϕ_i be a function from R^3 to R^1 satisfying

(A.3) $$E\phi_i(e_1,0,0) = 0 \quad \text{for all } i,$$

(A.4) $$\lim_{k\to 0} \sup E\{\sup|\phi_i(e_1,r,s) - \phi_i(e_i,r',s')|:$$

$$|r|,\ |r'|,\ |s|,\ |s'| \le k \text{ and } |r-r'|,\ |s-s'| \le k\delta\} \le C_0\delta g_i$$

for some C_0 and all $\delta<1$ and all i,

(A.5) $$\sup_{i\le N} g_i^{-1} E(\phi(e_1,r,s) - \phi(e_1,0,0) - A(\phi,i)r) = O(|r|^{1+\alpha} + |s|^{1+\alpha})$$

for positive constants $A(\phi,i)$ and $\alpha>0$, and

(A.6) $$\lim_{r,s\to 0} \sup_{i\le N} E(\phi_i(e_1,r,s) - \phi_i(e_1,0,0))^2 = 0.$$

Let $\alpha_i^{(1)}$, $\alpha_i^{(2)}$, and $\underline{\alpha}_i^{(3)}$ be functions from R^m to R^1, R^1 and R^n, respectively, such that

(A.7) $$\alpha_i^{(\ell)}(0) = 0, \quad \text{and}$$

(A.8) $$||\alpha_i^{(\ell)}(\underline{x}) - \alpha_i^{(\ell)}(y)|| \le k_i||\underline{x}-\underline{y}||$$

for all x and y in R^m, i=1,...,N, and ℓ=1,2, or 3. Let $\underline{x}_i$ ($=\underline{x}_{iN}$) be elements of R^n satisfying

(A.9) $$N^{-\frac{1}{2}}||\underline{x}_i|| \le k_i .$$

For $\underline{\Delta} \in R^m$, define the process

$$U_N(\underline{\Delta}) = N^{-\frac{1}{2}} \sum_{i=1}^{N} \phi_i(e_i, \alpha_i^{(1)}(\underline{\Delta}), \alpha_i^{(2)}(\Delta))(\underline{x}_i + \alpha_i^{(3)}(\underline{\Delta})).$$

Then, for all $M>0$

$$\text{(A.10)} \quad \sup_{||\Delta|| \le M} ||U_N(\Delta) - U_N(0) - N^{-\frac{1}{2}} \sum_{i=1}^{N} A(\phi,i)\alpha_i^{(1)}(\Delta)\underline{x}_i|| = o_p(1).$$

Proof of Theorem A.1. Fix $M>0$. We will show that

$$\text{(A.11)} \quad E(U_N(\underline{\Delta}) - U_N(\underline{0})) = N^{-\frac{1}{2}} \sum_{i=1}^{N} A(\phi,i)\alpha_i^{(1)}(\underline{\Delta})\ \underline{x}_i + o_p(1)$$

and

$$\text{(A.12)} \quad U_N(\underline{\Delta}) - U_N(\underline{0}) - E(U_N(\underline{\Delta}) - U_N(\underline{0})) = o_p(1)$$

for each fixed Δ, and that there exists K depending upon M but not δ such that for all $\delta>0$ and all N

$$\text{(A.13)} \quad E \sup\{||U_N(\underline{\Delta}) - U_N(\underline{\Delta}^*)||:\ ||\Delta|| \le M,\ ||\Delta^*|| \le M,\ ||\Delta-\Delta^*|| \le \delta\} \le K\delta.$$

Since for any δ, we can cover a ball of radius M in R^m with a finite number of balls of radius δ, (A.11), (A.12), and (A.13) prove that for each $\delta>0$,

$$\lim_{N\to\infty} p\left(\sup_{||\Delta|| \le M} ||U_N(\underline{\Delta}) - U_N(\underline{0}) - N^{-\frac{1}{2}} \sum_{i=1}^{N} A(\phi,i)\alpha_i^{(1)}(\underline{\Delta})\underline{x}_i|| \le K\delta\right) = 1,$$

which proves the theorem.

To prove (A.11), note that by (A.3),

$$N^{-\frac{1}{2}} \sum_{i=1}^{N} E(U_N(\underline{\Delta}) - U_N(\underline{0})) =$$

$$N^{-\frac{1}{2}} \sum_{i=1}^{N} E(\phi_i(e_i, \alpha_i^{(1)}(\underline{\Delta}), \alpha_i^{(2)}(\underline{\Delta})) - \phi_i(e_i,0,0))(\underline{x}_i + \underline{\alpha}_i^{(3)}(\underline{\Delta})),$$

by (A.1) and (A.8).

$$\text{(A.14)} \quad ||\underline{x}_i + \underline{\alpha}_i^{(3)}(\Delta)|| \le 2||\underline{x}_i|| \text{ for all large } N, \text{ and by (A.5),}$$

$$E(\phi_i(e_i, \alpha_i^{(1)}(\underline{\Delta}), \alpha_i^{(2)}(\underline{\Delta})) - \phi_i(e_1,0,0)) =$$
$$A(\phi,i)\alpha_i^{(1)}(\underline{\Delta}) + O(g_i k_i ||\Delta||)^{1+\alpha}$$

uniformly in i. Therefore

$$E((U_N(\underline{\Delta}) - U_N(0)) = N^{-\frac{1}{2}} \sum_{i=1}^{N} A(\phi_1 i)\alpha_i^{(1)}(\underline{\Delta})\underline{x}_i$$

$$+ O(N^{-\frac{1}{2}} \sum_{i=1}^{N} (g_i k_i)^{1+\alpha} ||\underline{x}_i||),$$

and by (A.2) and (A.9),

$$N^{-\frac{1}{2}} \sum_{i=1}^{N} (g_i k_i)^{1+\alpha} ||\underline{x}_i|| \le g_i k_i{}^{\alpha} \sum_{i=1}^{N} g_i k_i^2 =$$

$$\le (g_i k_i)^2 \ (\sum_{i=1}^{N} g_i^2 k_i^2)^{\frac{1}{2}} (\sum_{i=1}^{N} k_i^2)^{\frac{1}{2}} = o(1).$$

Thus (A.1) holds.

Next, by (A.14) we have that for N large

$$\text{Var}(U_N(\underline{\Delta}) - U_N(\underline{0})) \le$$

$$(2N^{-1} \sum_{i=1}^{N} g_i^2 ||x_i||^2) \sup_{i\le N} g_i^{-2} E(\phi_i(e_i,\alpha_i^{(1)}(\underline{\Delta}),\ \alpha_i^{(2)}(\underline{\Delta})) - \phi_i(e_1,0,0))^2.$$

It follows from (A.4), (A.5), (A.7), and (A.8) that

$$\sup_{i\le N} g_i^{-2}\ E(\phi_i(e_i,\alpha_i^{(1)}(\underline{\Delta}),\ \alpha_i^{(2)}(\underline{\Delta})) - \phi_i(e_i,0,0))^2 = o(1).$$

Therefore, (A.12) is proved by applying (A.2). Finally, by (A.14) the RHS of (A.13) is less than or equal to

$$2N^{-\frac{1}{2}} \sum_{i=1}^{N} ||\underline{x}_i||\ E\ \sup\{|\phi_i(e_i,\alpha_i^{(1)}(\underline{\Delta}),\ \alpha_i^{(2)}(\underline{\Delta})) -$$

$$\phi_i(e_i,\alpha_i^{(1)}(\underline{\Delta}^*),\ \alpha_i^{(2)}(\underline{\Delta}^*)|:\ ||\underline{\Delta}||\le M,\ ||\underline{\Delta}||\le M,\ ||\underline{\Delta}-\underline{\Delta}^*||\le \delta\}$$

which by (A.2), (A.4) to (A.9), and the Cauchy-Schwarz inequality does not exceed

$$\sup_N 2(\sum_{i=1}^{N} (g_i k_i)^2)^{\frac{1}{2}} C_0\delta = 2C_0 C_1^{\frac{1}{2}}\ \delta.$$

Therefore (A.13) is verified. □

REFERENCES

Anscombe, F. J. (1961), "Examination of residuals", Proc Fourth Berkely Symp Math Statist Prob (J. Neyman, editor), pp 1-36. University of California Press, Berkeley and Los Angeles, California.

Bickel, Peter J. (1978), "Using residuals robustly I: Tests for heteroscedasticity, nonlinearity," Ann Statist 6, pp 266-291.

Bickel, Peter J. and Doksum, Kjell A. (1978), "An analysis of transformations revisited", unpublished manuscript. University of California, Berkeley.

Box, George E. P. and Cox, David R. (1964). "An analysis of transformations", J Roy Statist Soc, Series B 26, pp 211-252

Box, George E. P. and Hill, William J. (1974), "Correcting inhomogeneity of variance with power transformation weighting", Technometrics 16, pp 385-389.

Carroll, Raymond J. (1979), "Robust transformations to achieve approximate normality", to appear in J Roy Statist Soc, Series B.

Carroll, Raymond J. and Ruppert, David (1979), "Almost sure properties of robust regression estimates", unpublished manuscript.

Fuller, Wayne A. and Rao, J. N. K. (1978), "Estimation for a linear regression model with unknown diagonal covariance matrix", Ann Statist 6, pp 1149-1158.

Huber, Peter J. (1964), "Estimation of a location parameter", Ann Math Statist 35, pp 73-101.

Huber, Peter J. (1973), "Robust regression: Asymptotics, conjectures and Monte-Carlo", Ann Statist 5, pp 799-821.

Jurecková, Jana (1977), "Asymptotic relations of M-estimates and R-estimates in linear regression model", Ann Statist 5, pp 464-472.

Loève, Michel (1963), Probability Theory, 3rd Edition. D. Van Nostrand, New York.

Yohai, Victor J. and Maronna, Ricardo A. (1979), "Asymptotic behavior of M-estimates for the linear model", Ann Statist 7, pp 258-268.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) M-Estimates for the Heteroscedastic Linear Model		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER Mimeo Series #1243
7. AUTHOR(s) Raymond J. Carroll and David Ruppert		8. CONTRACT OR GRANT NUMBER(s) AFOSR 75-2796
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of North Carolina at Chapel Hill		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Bolling AFB, DC		12. REPORT DATE July 1979
		13. NUMBER OF PAGES 18
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)

Distribution unlimited - approved for public release

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)

Weighted M-estimates, weighted least squares, estimated weights, linear regression in scale, robustness, asymptotic representation, asymptotically normal, confidence ellipsoids.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We treat the linear model $Y_i = \underline{C}_i^T\underline{\beta} + Z_i$ where $\underline{C}_i$ is a known vector, $\underline{\beta}$ is an unknown parameter, and Var Z_i is a function of $|\underline{C}_i^T\underline{\beta}|$ which is known, except for a parameter $\underline{\theta}$. For simultaneous M-estimates, $\hat{\underline{\beta}}$ and $\hat{\underline{\theta}}$, we show that $(\hat{\underline{\theta}}-\theta) = O_p(N^{-1/2})$, and find the limit distribution of $N^{1/2}(\hat{\underline{\beta}}-\underline{\beta})$. Moreover, the covariance matrix of the limit distribution can be consistently estimated, so large sample confidence ellipsoids and tests of hypotheses concerning $\underline{\beta}$ are feasible.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)