The Annals of Statistics
1982, Vol. 10, No. 2, 429–441

ROBUST ESTIMATION IN HETEROSCEDASTIC LINEAR MODELS

BY RAYMOND J. CARROLL[1] AND DAVID RUPPERT[2]

We consider a heteroscedastic linear model in which the variances are given by a parametric function of the mean responses and a parameter θ. We propose robust estimates for the regression parameter β and show that, as long as a reasonable starting estimate of θ is available, our estimates of β are asymptotically equivalent to the natural estimate obtained with known variances. A particular method for estimating θ is proposed and shown by Monte-Carlo to work quite well, especially in power and exponential models for the variances. We also briefly discuss a "feedback" estimate of β.

1. Introduction. We consider the heteroscedastic linear model

$$Y_i = \tau_i + \sigma_i \varepsilon_i, \qquad \tau_i = x_i \beta, \qquad i = 1, \cdots, N, \tag{1.1}$$

where $\{x_i\}$ and $1 \times p$ design constants, β is a $p \times 1$ regression parameter, $\{\varepsilon_i\}$ are independent and identically distributed with mean zero and unknown symmetric distribution function F, and $\{\sigma_i\}$ are scaling constants which express the possible heteroscedasticity. Our primary interest is in inference about the unknown regression parameter β.

Of course, one could ignore the $\{\sigma_i\}$ and use classical methods such as least squares or M-estimation (Huber, 1981), but such estimates are not efficient. In order to make more efficient inference about β, it is necessary to get information about the $\{\sigma_i\}$. In one approach to the problem, the $\{\sigma_i\}$ are assumed completely unknown, but replication is assumed feasible so that the $\{Y_i\}$ occur in groups of equal variance. Recent results in this direction are due to Fuller and Rao (1978). Their results are complicated, and the delicate calculations involved seem to depend very heavily on an assumption of Gaussian errors, which is undesirable from the viewpoint of efficiency robustness; see Huber (1981) for details and further references.

The second approach to the estimation problem for (1.1) avoids the replication assumption by positing a known form for the error variance, i.e.,

$$\sigma_i = H(x_i, \beta, \theta), \tag{1.2}$$

where θ is an $r \times 1$ vector of unknown coefficients and H is smooth and known. A model such as (1.2) is behind the tests for homoscedasticity developed by Anscombe (1961), Bickel (1978), and Carroll and Ruppert (1981). Of course, in many real problems we suspect a heteroscedastic model because the dispersion of the residuals increases with the magnitude of the fitted values. Thus, it has become quite common to simplify (1.2) by assuming σ_i is a function of τ_i or $|\tau_i|$, e.g.,

$$\begin{aligned} &\sigma_i = \sigma(1 + |\tau_i|)^\lambda; \qquad \sigma_i = \sigma|\tau_i|^\lambda \quad \text{(Box and Hill, 1974);} \\ &\sigma_i = \sigma \exp(\lambda \tau_i) \quad \text{(Bickel, 1978);} \\ &\sigma_i = \sigma(1 + \lambda \tau_i^2)^{1/2} \quad \text{(Jobson and Fuller, 1980).} \end{aligned} \tag{1.3}$$

(See also Dent and Hildreth, 1977.) Following these examples, we will thus assume that for

Received August 1979; revised October 1981.

[1] University of North Carolina and the National Heart, Lung and Blood Institute. Supported by the Air Force Office of Scientific Research under grant AFOSR-80-0080.

[2] University of North Carolina. Supported by National Science Foundation grant MCS78-01240.

AMS 1970 *subject classifications.* Primary 62J05, secondary 62G35.

Key words and phrases. Feedback, M-estimates, non-constant variances, random coefficient models, weighted least squares.

some known H,

$$\sigma_i = \sigma H_*(\tau_i, \lambda) = H(\tau_i, \theta) \quad \text{with} \quad \theta = (\sigma, \lambda). \tag{1.4}$$

Our results can be generalized to the model (1.2), but the statements of results and assumptions then become extremely complicated.

Box and Hill (1974) and Jobson and Fuller (1980) both suggest a form of generalized weighted least squares. One obtains estimates of (θ, β), constructs estimated weights $\hat{\sigma}_i$, and then performs ordinary weighted least squares. Their methods are constructed from a normal error assumption, and their efficiency depends on this assumption. The maximum likelihood estimates for θ under the normality assumption have a quadratic influence curve and may be particularly non-robust. As argued above, the recent literature demonstrates some acceptance to the notion that estimators should be robust against departures from normality. One purpose of this article is to provide a set of such robust estimates.

Implicit in the work of Box and Hill (1974) and Jobson and Fuller (1980) is the notion that this problem is adaptable, i.e., the generalized weighted least squares methods are asymptotically equivalent to the "optimal" weighted least squares estimate for the true $\{\sigma_i\}$. Our second major aim is to show that there is a wide class of robust estimates of β which are adaptable for many distribution functions F and models (1.4).

2. A class of weighted robust estimates.

Suppose we have estimates of (θ, β) which are $N^{1/2}$-consistent, i.e.,

$$N^{1/2}(\hat{\theta} - \theta) = O_p(1), \qquad N^{1/2}(\hat{\beta}_0 - \beta) = O_p(1). \tag{2.1}$$

The existence of such estimates is discussed in the next section. We then form the estimated σ_i as follows,

$$\hat{\sigma}_i = H(t_i, \hat{\theta}), \qquad t_i = x_i\hat{\beta}_0. \tag{2.2}$$

If the $\{\sigma_i\}$ were known, robustness considerations discussed by Huber (1973, 1981) suggest a general class of weighted M-estimates formed by solving the minimization problem in β;

$$\Sigma\rho\left\{\frac{Y_i - x_i\beta}{\sigma_i}\right\} = \text{minimum}. \tag{2.3}$$

Here ρ is taken to be a convex function. If, for example, $\rho(x) = x^2/2$, we get the "optimal" weighted least squares estimate with known weights. In general, the unknown solution to (2.3) is denoted β_{opt}.

The class of estimates we suggest are very simply generated by substituting $\{\hat{\sigma}_i\}$ into (2.3). Taking derivatives, we suggest solving the equation

$$\textstyle\sum_{i=1}^{N} \left(\dfrac{x_i'}{\hat{\sigma}_i}\right)\psi\left\{\dfrac{Y_i - x_i\beta}{\hat{\sigma}_i}\right\} = 0, \tag{2.4}$$

with solution denoted by $\hat{\beta}$. Throughout we take ψ to be an odd, continuous function. The non-robust generalized weighted least squares estimates suggested by Box and Hill (1974) and Jobson and Fuller (1980) fall under the special case of (2.4) when $\psi(x) = x$; both propose possibilities for $\hat{\beta}_0$ and $\hat{\theta}$ of (2.1). As suggested by the literature, choosing a bounded ψ can result in reasonably efficient and robust estimates of β.

Define $d_i = x_i/\sigma_i$ and assume that for a positive definite matrix S,

$$S_N = N^{-1} \textstyle\sum_{i=1}^{N} d_i' \, d_i \to S. \tag{2.5}$$

Then by formal Taylor series arguments, the optimal robust weighted estimate $\hat{\beta}_{\text{opt}}$, which solves (2.3), satisfies

$$N^{1/2}(\hat{\beta}_{\text{opt}} - \beta) = N^{-1/2} \textstyle\sum_{i=1}^{N} S^{-1} \, d_i' \dfrac{\psi(\varepsilon_i)}{E\psi'(\varepsilon_1)} + o_p(1) \to_{\mathscr{L}} N(0, E\psi^2 S^{-1}(E\psi')^{-2}). \tag{2.6}$$

Our main result concerning adaptation is that when (2.1) holds, and hence we have a reasonable estimate of θ, then our estimate $\hat{\beta}$ is asymptotically equivalent to $\hat{\beta}_{\text{opt}}$. In stating assumptions and proofs, we simplify (1.4) to

$$\sigma_i = \exp\{h(\tau_i)\theta\}, \tag{2.7}$$

where h is a function from R to R^r. The model (2.7) includes the first three models in (1.3), but it is not strictly necessary for the validity of our results. Our reason for considering only (2.7) in the formal aspects of this section is to avoid making already cumbersome notation needlessly complicated. Generalizations to the model (1.4) required that $H(\cdot, \cdot)$ be smooth. Formally, we have the following.

THEOREM 1. *Assume* (2.1), (2.5), (2.7), *the smoothness conditions* B6 *through* B8 *listed in Section* 7, *and*

B1. *ψ monotone and odd, F symmetric,* $0 < E\psi^2(\varepsilon_1) < \infty$, $E\psi' > 0$.

B2. $\lim_{N\to\infty}\sup_{i\le N}(\| x_i \| + \| h(\tau_i)\|)N^{-1/2} = 0$.

B3. $\sup_N\{N^{-1} \sum_{i=1}^N (\| x_i \|^2 + \| h(\tau_i)\|^2)\} < \infty$.

B4. *The σ_i are bounded away from zero.*

B5. *On an open interval I (possibly infinite) containing all the $\{\tau_i\}$, h is Lipschitz continuous.*

Then

$$N^{1/2}(\hat{\beta} - \hat{\beta}_{\text{opt}}) \to_p 0. \tag{2.8}$$

That $\hat{\beta}$ is robust against outliers in the errors when ψ is bounded can be seen by combining (2.6) and (2.8). The resulting influence curve is strikingly similar to the unweighted case in homoscedastic models.

The proof is given in Section 7. Conditions B1 through B3 and B6 through B8 are similar to those used by Bickel (1975) in his study of one-step M-estimates in the *homoscedastic* model. Condition B4 ensures that we do not have infinite weights, and condition B5 assures us that when $\sigma_i = H(\tau_i, \theta) = \exp\{h(\tau_i)\theta\}$, the function H is sufficiently smooth.

3. Estimation of θ.

In the previous section we have shown that, except for certain technical conditions, one can construct robust weighted estimates of β as long as one has available estimates of θ and β which satisfy (2.1). Preliminary estimates $\hat{\beta}_0$ satisfying (2.1) are readily available and include (under reasonable assumptions) ordinary least squares estimates and ordinary M-estimates; details of sufficient conditions for this are available from the authors. Bounded influence regression estimates could also be used; see, e.g., Krasker and Welsch (1981). In this section, we propose a class of estimates of θ which are robust and satisfy (2.1). There are, of course, many possible ways to construct such estimates, but our method has the necessary theoretical properties as well as encouraging small sample properties; see the next section for details.

To motivate our estimates, suppose that the $\{\tau_i\}$ were known, that the $\{\sigma_i\}$ satisfy (1.4), and that the density f is proportional to $\exp\{-\rho(x)\}$, where ρ and $\rho' = \psi$ are as in the previous section. This device is common in robustness studies; see Huber (1981), Bickel and Doksum (1981), and Carroll (1980) for examples. In this instance, the log-likelihood for θ is, up to a constant,

$$\ell(\theta) = \textstyle\sum_{i=1}^N \log H(\tau_i, \theta) - \sum_{i=1}^N \rho\left\{\dfrac{Y_i - \tau_i}{H(\tau_i, \theta)}\right\}. \tag{3.1}$$

Taking derivatives in θ suggests that we solve

$$0 = \ell'(\theta) = \textstyle\sum_{i=1}^N [z_i(\theta)\psi\{z_i(\theta)\} - 1]\dfrac{\partial}{\partial\theta} H(\tau_i, \theta)/H(\tau_i, \theta), \tag{3.2}$$

where $z_i(\theta) = \dfrac{(Y_i - \tau_i)}{H(\tau_i, \theta)}$. Because the term in square brackets in (3.2) is not bounded and hence would, in general, lead to an unbounded influence function for the estimated θ and an overall lack of robustness in our estimation procedure, we follow the common device used in the homoscedastic case by Huber (1981) and Bickel and Doksum (1981) of replacing $x\psi(x) - 1$ by a function $\chi(\cdot)$, as well as replacing τ_i by $t_i = x_i\hat{\beta}_0$, thus leading to estimates obtained by solving

$$0 = G_N(\theta) = \textstyle\sum_{i=1}^N \chi\left\{\dfrac{Y_i - t_i}{H(t_i, \theta)}\right\} \dfrac{\partial}{\partial\theta} H(t_i, \theta)/H(t_i, \theta). \tag{3.3}$$

Probably the most common choice of $\chi(\cdot)$ in the homoscedastic case is

$$\chi(y) = \chi^2(y) - \int \psi^2(x)\phi(x)\,dx. \tag{3.4}$$

This choice of $\chi(\cdot)$ gives bounded influence to our estimates of θ, and thus might reasonably be preferred in our problem to $y\psi(y) - 1$, just as it is in the homoscedastic case; see Huber (1981, Section 11.1) for certain optimality properties of this choice. In the case of the special model (2.7), we have

$$G_N(\theta) = \textstyle\sum_{i=1}^N \chi\{(Y_i - t_i)e^{-h(t_i)\theta}\}h(t_i). \tag{3.5}$$

We make the assumptions that $\chi(\cdot)$ is an even function with $\chi(0) < 0$, $\chi(\infty) > 0$. In the model (1.4), σ is a free parameter defined so that

$$E\chi\left(\frac{Y_1 - \tau_1}{\sigma_1}\right) = 0. \tag{3.6}$$

In the first model of (1.3), we have

$$\theta = (\log \sigma, \lambda)^T, \qquad h(\tau) = \log(1 + |\tau|).$$

In many models (such as the first three models in (1.3), the third with $\tau_i > 0$), one can show that solutions to the equation $G_N(\theta) = G_N(\sigma, \lambda) = 0$ exist. We have been unable to show that the solutions are unique, although in all of our examples, unique solutions have been obtained. More importantly, one may not wish to consider all possible values of θ, e.g., in the first three models of (1.3), one may reasonably wish to restrict $|\theta| \leq 1.5$ if one assumes that the variances will be no larger than the cubes of the means. For these reasons, we suggest the following procedure:

(3.7) Minimize $\|G_N(\theta)\| = \|G_N(\sigma, \lambda)\|$ on the interval $\lambda \in J$.
If the solution is not unique, choose the one with smallest $\|\lambda\|$.

The solution to (3.7) is thus well-defined. In all of our examples when θ is unrestricted, the solutions to (3.3) and (3.7) have coincided. In the examples in which we have restricted θ, (3.7) has always had a unique solution even when (3.3) has not had a solution in the restricted space.

An appealing feature of these estimates is that they are natural generalizations of the classical Huber Proposal 2 for the homoscedastic case.

THEOREM 2. *Assume* (2.5), (2.7), (3.6), *and* B2 *through* B5. *Further assume that* $N^{1/2}(\hat{\beta}_0 - \beta) = O_p(1)$. *Finally, make the assumptions*

C1. $0 < E\chi^2(\varepsilon_1) < \infty$, and χ is non-decreasing on $[0, \infty)$.
C2. *As* $r, s \to 0$, *for* $A(\chi) > 0$, $E\chi\{(\varepsilon_1 + r)(1 + s)\} = A(\chi)s + o(|r| + |s|)$.
C3. *Condition* B7 *holds for* χ.
C4. *Condition* B8 *holds for* χ.
C5. *If* λ_N *is the minimal eigenvalue of* $H_N = N^{-1} \sum_{i=1}^N h(\tau_i)^T h(\tau_i)$, *then* $\liminf \lambda_N = \lambda_\infty > 0$.

Then if $\hat{\theta}$ solves (3.7), *we have*

$$\hat{\theta} - \theta = O_p(N^{-1/2}). \tag{3.8}$$

The proof is given in Section 7. The conditions are similar to those of Bickel (1975), with only C5 affected by hetreroscedasticity. Further details of implementation are discussed in the next section.

One can also introduce redescending M-estimates by using ψ redescending to zero. Estimates for θ and β can be obtained by doing one or two steps of Newton-Raphson for (2.4) and (3.3) from any estimate satisfying (2.1). Proofs are similar to those given in the appendices.

4. A Monte-Carlo study.

Because Theorem 1 is an asymptotic result, we performed a small Monte-Carlo study to assess the small sample properties of $\hat{\beta}$. The model was simple linear regression, given by

$$Y_i = \beta_0 + \beta_1 c_i + \sigma_i \epsilon_i = \tau_i + \sigma_i \epsilon_i, \quad i = 1, \cdots, N. \tag{4.1}$$

In the study, the $\{c_i\}$ were equally spaced between -2 and $+2$, and we chose to study the model

$$\sigma_i = \sigma(1 + |\tau_i|)^{\lambda}.$$

The experiments were each repeated two hundred times under the following circumstances:

(a) $N = 21$, $\{\epsilon_i\}$ are $N(0, 1)$, $\sigma = .25$, $\beta_0 = 2$, $\beta_1 = 1$.

(b) $N = 41$, $\{\epsilon_i\}$ are $N(0, 1)$ with probability $p = .90$ and $N(0, 9)$ with $p = .10$, $\sigma = .25$, $\beta_0 = 4$, $\beta_1 = 2$.

We made two choices for ψ. First was $\psi(x) = x$, which yields the usual weighted least squares estimate $\hat{\beta}_L$, and the second was Huber's $\psi(x) = \max\{-2.0, \min(x, 2.0)\}$. This gives a version $\hat{\beta}_R$ of our robust weighted estimates. In constructing $\hat{\sigma}_i$, we defined χ as in equation (3.4).

Both $\hat{\beta}_L$ and $\hat{\beta}_R$ were constructed as follows:

Step (i). Let β_* be the unweighted Huber Proposal 2 estimate ($\lambda = 0$) with χ given by (3.4) and $\psi(x) = \max\{-2.0, \min(x, 2.0)\}$.

Step (ii). Solve (3.7) for (σ_*, λ_*) and form inverse "weights"

$$w_i^2 = (1 + |t_i|)^{2\lambda}, \quad t_i = x_i\beta_*.$$

Step (iii). Solve a weighted Huber Proposal 2 by simultaneously solving (2.4) with the desired function ψ and the part of (3.6) given by

$$\textstyle\sum_{i=1}^N \chi\left(\frac{Y_i - x_i\beta}{\sigma w_i}\right) = 0. \tag{4.2}$$

The result is $\hat{\beta}_0$.

Step (iv). Repeat steps (ii) and (iii) to obtain$t_i = x_i\ddot{\beta}_0$, $\ddot{\lambda}$, $\hat{\sigma}$, $\ddot{\beta}$.

The algorithm given here was chosen so as to reproduce Huber's Proposal 2 in the homoscedastic case $\hat{\lambda} = 0$. Direct application of the results of Section 2 involves only solving (2.4) in Step (iii) and gave results essentially indistinguishable from those reported here. In solving for $(\hat{\lambda}, \hat{\sigma})$, we used the subroutine ZXGSN of the IMSL library.

In Table 1, we list part of the results of the study. The values listed are ratios of mean square errors for estimating β_1 in model (4.1), the ratio being with respect to the "optimal" robust method one would use if $w_i^* = (1 + |\tau_1|)^{\lambda}$ were known, i.e., solve (2.4) and (4.2) simultaneously with the known weights. The study is fairly small, but it does seem to indicate that our robust weighted estimate will work in situations in which heteroscedasticity is suspected.

TABLE 1
Monte-Carlo MSE *ratio for simple linear regression under* (6.1).

	Sample Size N = 21 $\beta_0 = 2.0$, $\beta_1 = 1.0$ Normal Errors			Sample Size N = 41 $\beta_0 = 4.0$, $\beta_1 = 2.0$ Contaminated Errors		
Estimator	**$\lambda = 0.0$**	**$\lambda = .5$**	**$\lambda = 1.0$**	**$\lambda = 0.0$**	**$\lambda = .5$**	**$\lambda = 1.0$**
Unweighted LSE	.98	1.18	1.67	1.24	1.51	2.31
"Optimal" WLSE, known weights	.98	.98	.98	1.24	1.19	1.18
Our WLSE, estimated weights	1.14	1.13	1.11	1.29	1.25	1.26
Unweighted robust estimate	1.00	1.18	1.66	1.00	1.21	1.79
Our weighted robust estimate, estimated weights	1.14	1.13	1.10	1.03	1.04	1.07

It is important to note that our estimate has MSE never more than 15% larger than the unknown estimate formed with the correct weights and seems to do better than unweighted estimates when $\lambda \neq 0$. Note also the robustness feature; the efficiency of the weighted least squares estimates (even the "optimal" one) depends heavily on the normality assumption and is not very high in the contaminated case. All of the results tend to support the applicability of Theorem 1.

We repeated the experiment, but with the model

$$\sigma_i = \sigma \exp(\lambda | \tau_i |),$$

and obtained similar results, which seem to indicate that our theory is applicable for a variety of models for the $\{\sigma_i\}$.

For testing and interval estimation, we use the following generalization of methods suggested by Huber (1973) for the homoscedastic case. Using (2.6) and Theorem 1, we estimate the covariance of$N^{1/2}(\hat{\beta} - \beta)$ by

$$K(\widehat{E\psi^2})\widehat{S}^{-1}(\widehat{E\psi'})^{-2}, \tag{4.2}$$

where

$$\widehat{E\psi'} = N^{-1}\Sigma\psi'\left(\frac{Y_i - x_i\hat{\beta}}{\hat{\sigma}_i}\right), \quad K = 1 + (p+2)\frac{1-\lambda}{N\lambda}, \quad \hat{S} = N^{-1}\,\Sigma x_i' x_i \hat{\sigma}_i^{-2}$$

and $\widehat{E\psi^2}$ is defined similarly to $\widehat{E\psi'}$. In our Monte-Carlo experiment, we constructed confidence intervals for the slope parameter β_1 in (4.1), using (4.3) and t-percentage points with $N - p - r = N - 4$ degrees of freedom. The intended coverage probability was 95%. In none of these cases did the achieved coverage probability fall below 92%, and in the majority of the cases it was at least 94%.

We also attempted to solve equations (2.4) and (3.5) simultaneously using the IMSL routine ZSYSTM. Our experience was much like that of Froehlich (1973) in that the algorithm converged most of the time but not always. Dent and Hildreth (1977) were able to show that the difficulties experienced by Froehlich could be overcome by sophisticated optimization techniques. We suspect that the same holds for our problem.

The particular method for estimating $\theta = (\sigma, \lambda)$ outlined in Section 3 and explored in this section is recommended for models such as the first three in (1.3), which satisfy (2.7). In the fourth model of (1.3), an alternative procedure is preferable because we can exploit the relationship

$$\sigma_i^2 = \alpha_1 + \alpha_2 \tau_i^2.$$

Here one would obtain initial estimates of (α_1, α_2) by robust regression techniques, as long as the lines of Jobson and Fuller (1980), working with the squares of the residuals from a preliminary fit. One would then do one-step of a Newton-Raphson towards solving versions of (3.3) which are obtained by working with (α_1, α_2) and following the line of reasoning in (3.1) through (3.3). Monte-Carlo work, which will be reported elsewhere, indicates that this technique can be quite successful.

5. Feedback. In the case of normal errors, Jobson and Fuller have suggested using the information about β in the terms $\sigma_1 = H(\tau_1, \theta)$. This essentially reduces to maximizing (3.1) jointly in (θ, β). In a very nice result, they show that if the error distribution is exactly normal and if (1.2) is exactly correct, then improvement over the weighted least squares estimate can be achieved. It is clear that such feedback procedures will be adversely affected by outliers or non-normal error distributions, and it is not clear how to robustly modify them.

In cases where using feedback is contemplated, a second form of robustness must also be considered, i.e., robustness against misspecification of the functions H in (3.1). Carroll and Ruppert (1981, unpublished) have shown that as long as H is correctly specified to order $O(N^{-1/2})$, the asymptotic properties of the weighted estimates ((2.4), (3.5)) are the same as if H were correctly specified; in this sense, our weighted estimates are robust against small errors in specifying H. They also show that such robustness is not the case for feedback estimates. In fact, any gain from feedback can be more than offset by slight errors in specifying H. Since our primary interest is in β, and $\sigma_i = H(\tau_i, \theta)$ is at best an approximation, we suggest that feedback should not be automatically preferred in practical use.

6. An example. In Figure 1, we plot the outcomes of 113 observations of Total Esterase $\{C_i\}$ and Radioimmunoassay - RIA $\{Y_i\}$, made available to us by Drs. D. Horowitz and D. Proud of the National Heart, Lung and Blood Institute. The data are clearly heteroscedastic, so we fit the model (4.1) with variance model

$$\sigma_i = \sigma(1 + |\tau_i|)^\lambda \tag{6.1}$$

and estimation done as in the previous section. The results are summarized in Table 2. Since λ appears to be fairly large, the results of the Monte-Carlo indicate that weighting should be of real benefit. The confidence limits on $\hat{\lambda}$ were obtained by bootstrapping (using 60 simulations). In the weighted cases, the standard errors for β_0 and β_1 were obtained from (4.3); similar standard errors not reported here were found by bootstrapping. The weighted results are fairly close together. While our purpose in presenting the numbers is merely illustrative, we note that the values of λ suggest that a logarithmic or square root transformation might stabilize the variances (Box and Hill, 1974). A random coefficient model might also be contemplated (Dent and Hildreth, 1977). We fit a quadratic model to the data with little change.

A program has been written by Neal Thomas to solve equations (2.4) and (3.5) simultaneously when the second model in (1.3) is used. Since the program utilizes the IMSL package's Levenberg-Marquardt algorithm, it can be used on non-linear regression models. The program is now being tested on simulated data and has been used in a study of migration patterns of the Atlantic menhaden, where it was tried on a data set exhibiting heteroscedasticity and numerous outliers. There it produced estimates which, from a biological viewpoint, seemed more credible than estimates from three other procedures: least squares, least squares after a log transformation applied simultaneously to both the dependent variable and the regression function, and Huber's M-estimator with the MAD estimate of scale (Deriso, Reish, Ruppert, and Carroll, manuscript in preparation). Since menhaden are relatively rare in the northern part of their range (New England), catch data from that region exhibit small values but also low variability. Apparently, a weighted estimator is needed in order to obtain reasonable estimates of migration rates to and from northern waters.

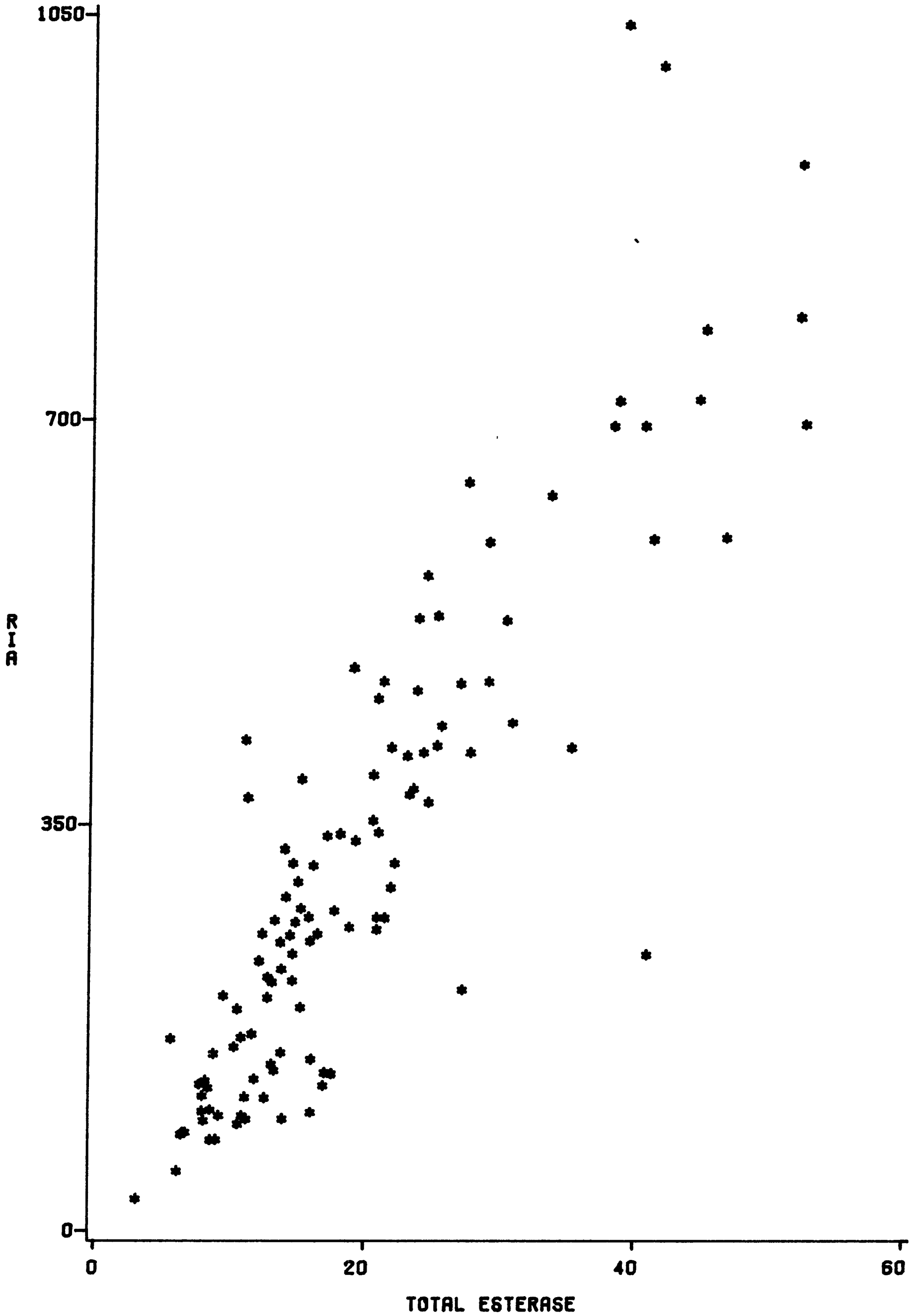

FIG. 1. *Scatterplot of* 113 *observations on* x = *total esterase and* y = *radioimmunoassay RIA.*

TABLE 2
Results of the analysis on the data for Figure 1 *assuming* (6.1).

Method	$\hat{\beta}_0$	Standard Error	$\hat{\beta}$	Standard Error	$\hat{\lambda}$	90% Confidence Limits for λ
Unweighted least squares	−6.30	20.0	16.73	.89	—	—
Our weighted least squares	−19.22	14.1	17.42	.94	.68	(0.4,0.9)
Unweighted robust	−6.54	17.4	16.67	.77	—	—
Our weighted robust	−26.99	11.8	17.73	.88	.85	(0.7,1.1)

Proofs of theorems. The smoothness conditions mentioned in Sections 2 and 3 are as follows:

B6. As $r \to 0$ and $s \to 0$, $E\psi\{(\varepsilon_1 + r)(1+s)\} = rE\psi'(\varepsilon_1) + o(|r| + |s|)$.

B7. There exist $K > 0$ and $C_0 > 0$ such that when $0 < \delta < 1$, $|r| \le K$, and $|s| \le K$,

$$E \sup[|\psi\{(\varepsilon_1 + r)(1+s)\} - \psi\{(\varepsilon_1 + r')(1+s')\}|: |r - r'| \le \delta \text{ and } |s - s'| \le \delta] \le C_0\delta.$$

B8. $\lim_{\delta\to 0} E \sup([\psi\{(\varepsilon_1 + r)(1+s)\} - \psi(\varepsilon_1)]^2: |r|, |s| \le \delta) = 0.$

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

THEOREM 7.1. *Let g_i, k_i, and $A(\phi, i)$ standing for g_{iN}, k_{iN}, and $A(\phi, i, N)$, be sequences of positive constants such that*

(7.1) $$\lim_{N\to\infty}\sup_{i\le N}(k_i + k_i g_i) = 0, \quad \sup_N \sup_{i\le N} A(\phi, i) < \infty,$$

and

(7.2) $$\sup_N \textstyle\sum_{i=1}^N (k_i^2 + k_i^2 g_i^2 + N^{-1/2} g_i k_i) = C_1 < \infty.$$

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

(7.3) $$E\phi_i(\varepsilon_1, 0, 0) = 0 \quad \textit{for all} \quad i.$$

Suppose that there exists $K > 0$ and $C_0 > 0$ such that for all i,

(7.4) $$E \sup\{|\phi_i(\varepsilon_1, r, s) - \phi_i(\varepsilon_1, r', s')|: |r - r'|, |s - s'| \le \delta\} \le C_0\, g_i\delta$$

whenever $0 < \delta < 1$, $|r| \le K$, *and* $|s| \le K$,

(7.5) $$\sup_N \sup_{i\le N} g_i^{-1} E\{\phi_i(\varepsilon_1, r, s) - \phi_i(\varepsilon_1, 0, 0) - A(\phi, i)r\} = o(|r| + |s|) \textit{ as } r, s, \to 0,$$

(7.6) $$\lim_{\delta\to 0} \sup_N \sup_{i\le N} E[\sup_{|r|\le\delta, |s|\le\delta} g_i^{-2}\{\phi_i(\varepsilon_1, r, s) - \phi_i(\varepsilon_1, 0, 0)\}^2] = 0,$$

and $\sup_N \sup_{i\le N} g_i^{-2} E\phi_i^2(\varepsilon_i, 0, 0) < \infty$. *Let $a_i^{(1)}$, $\alpha_i^{(2)}$, and $\boldsymbol{\alpha}_i^{(3)}$ be functions from R^m to R^1, R^1, and R^n respectively, and let $\mathbf{z}_i (= \mathbf{z}_{iN})$ be elements of R^n satisfying*

(7.7) $$\alpha_i^{(\ell)}(0) = 0, \quad \ell = 1, 2, 3,$$

and for each compact set S there exists K such that

(7.8) $$|\alpha_i^{(\ell)}(\mathbf{x}) - \alpha_i^{(\ell)}(\mathbf{y})| \le k_i \|\mathbf{x} - \mathbf{y}\| K, \ \ell = 1, 2,$$

and $\|\boldsymbol{\alpha}_\iota^{(3)}(\mathbf{x}) - \boldsymbol{\alpha}_i^{(3)}(\mathbf{y})\| \le k_i \|\mathbf{z}_i\| \|\mathbf{x} - \mathbf{y}\| K$ *for all x and y in S, $i = 1, \cdots, N$, and*

(7.9) $$N^{-1/2}\|\mathbf{z}_i\| \le k_i.$$

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

$$U_N(\boldsymbol{\Delta}) = N^{-1/2} \textstyle\sum_{i=1}^N \phi_i\{\varepsilon_i, \alpha_i^{(1)}(\boldsymbol{\Delta}), \alpha_i^{(2)}(\boldsymbol{\Delta})\}\{\mathbf{z}_i + \boldsymbol{\alpha}_i^{(3)}(\boldsymbol{\Delta})\}.$$

Then, for all $M > 0$,

$$\sup_{\|\Delta\| \le M} \| U_N(\boldsymbol{\Delta}) - U_N(\mathbf{0}) - N^{-1/2} \textstyle\sum_{i=1}^N A(\phi, i)\alpha_i^{(1)}(\boldsymbol{\Delta})\mathbf{z}_i \| = o_p(1). \tag{7.10}$$

PROOF OF THEOREM 7.1. For convenience, take $M = 1$. For $0 < \delta < 1$, define

$$S_N(\boldsymbol{\Delta}, \delta) = \sup\{\| U_N(\boldsymbol{\Delta}') - U_N(\boldsymbol{\Delta})\| : \|\boldsymbol{\Delta}' - \boldsymbol{\Delta}\| \le \delta\}.$$

We will show that

$$E\{U_N(\boldsymbol{\Delta}) - U_N(\mathbf{0})\} = N^{-1/2} \textstyle\sum_{i=1}^N A(\phi, i)\alpha_i^{(1)}(\boldsymbol{\Delta})\mathbf{z}_i + o(1), \tag{7.11}$$

$$U_N(\boldsymbol{\Delta}) - U_N(\mathbf{0}) - E\{U_N(\boldsymbol{\Delta}) - U_N(\mathbf{0})\} = o_p(1) \tag{7.12}$$

for each fixed $\boldsymbol{\Delta}$, and that there exists K depending upon M but not δ such that for all $0 < \delta < 1$, all N, and all $\|\boldsymbol{\Delta}\| \le 1$,

$$S_N(\boldsymbol{\Delta}, \delta) - ES_N(\boldsymbol{\Delta}, \delta) = o_p(1) \text{ and } ES_N(\boldsymbol{\Delta}, \delta) \le K\delta, \tag{7.13}$$

where K does not depend upon δ. Since for any δ, we can cover the ball of radius 1 in R^m with a finite number of balls of radius δ, (7.11), (7.12) and (7.13) prove the theorem.

To prove (7.11), note that by (7.3),

$$E(U_N(\boldsymbol{\Delta}) - U_N(\mathbf{0})) = N^{-1/2} \textstyle\sum_{i=1}^N E[\phi_i\{\varepsilon_i, \alpha_i^{(1)}(\boldsymbol{\Delta}), \alpha_i^{(2)}(\boldsymbol{\Delta})\} - \phi_i(\varepsilon_i, 0, 0)]\{\mathbf{z}_i + \boldsymbol{\alpha}_i^{(3)}(\boldsymbol{\Delta})\}.$$

We next have by (7.1), (7.7) and (7.8) that, for all large N,

$$\|\mathbf{z}_i + \boldsymbol{\alpha}_i^{(3)}(\boldsymbol{\Delta})\| \le 2\|\mathbf{z}_i\| \tag{7.14}$$

(for simplicity, take $K = 1$ in (7.8)), and also by (7.5), (7.7) and (7.8),

$$E[\phi_i\{\varepsilon_i, \alpha_i^{(1)}(\boldsymbol{\Delta}), \alpha_i^{(2)}(\boldsymbol{\Delta})\} - \phi_i(\varepsilon_i, 0, 0)] = A(\phi, i)\alpha_i^{(1)}(\boldsymbol{\Delta}) + o(g_i k_i)$$

uniformly in i. Therefore,

$$E\{U_N(\boldsymbol{\Delta}) - U_N(\mathbf{0})\} = N^{-1/2} \textstyle\sum_{i=1}^N A(\phi, i)\alpha_i^{(1)}(\boldsymbol{\Delta})\mathbf{z}_i + o\{N^{-1/2} \textstyle\sum_{i=1}^N g_i k_i \|\mathbf{z}_i\|$$
$$+ N^{-1/2} \textstyle\sum_{i=1}^N A(\phi, i)\alpha_i^{(1)}(\boldsymbol{\Delta})\boldsymbol{\alpha}_i^{(3)}(\boldsymbol{\Delta})\}.$$

By (7.1), (7.7), (7.8), and (7.9), the last term on the RHS is $o(1)$. By (7.2), (7.9) and the Cauchy-Schwarz inequality, the second term is bounded by

$$o(\textstyle\sum_{i=1}^N g_i k_i^2) = o(1),$$

so that (7.11) holds. Then, using (7.14), we have that for N large,

$$½ \operatorname{Var}\{U_N(\boldsymbol{\Delta}) - U_N(\mathbf{0})\} \le (2N^{-1} \textstyle\sum_{i=1}^N g_i^2 \|\mathbf{z}_i\|^2) \sup_{i \le N} g_i^{-2} E[\phi_i\{\varepsilon_i, \alpha_i^{(1)}(\boldsymbol{\Delta}), \alpha_i^{(2)}(\boldsymbol{\Delta})\}$$
$$- \phi_i(\varepsilon_1, 0, 0)]^2 + E\| N^{-1/2} \textstyle\sum_{i=1}^N \phi_i(\varepsilon_i, 0, 0)\boldsymbol{\alpha}_i^{(3)}(\boldsymbol{\Delta})\|^2.$$

The second term on the RHS is $o(1)$ by (7.7) and (7.8). It also follows from (7.6) through (7.8) that

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

Therefore, (7.12) is proved by applying (7.2) and (7.9). Finally, by (7.14) $ES_N(\boldsymbol{\Delta}, \delta)$ is less than or equal to

$$2N^{1/2} \textstyle\sum_{i=1}^N E \sup_{\|\Delta - \Delta'\| \le \delta} |\phi_i\{\varepsilon_1, \alpha_i^{(1)}(\boldsymbol{\Delta}), \alpha_i^{(2)}(\boldsymbol{\Delta})\} - \phi_i\{\varepsilon_1, \alpha_i^{(1)}(\boldsymbol{\Delta}'), \alpha_i^{(2)}(\boldsymbol{\Delta}')\}| \|\mathbf{z}_i\|$$
$$+ N^{-1/2} \textstyle\sum_{i=1}^N \sup_{\|\Delta - \Delta'\| \le \delta} \|\boldsymbol{\alpha}_i^{(3)}(\boldsymbol{\Delta}) - \boldsymbol{\alpha}_i^{(3)}(\boldsymbol{\Delta}')\| E|\phi_i\{\varepsilon_1, \alpha_i^{(1)}(\boldsymbol{\Delta}'), \alpha_i^{(2)}(\boldsymbol{\Delta}')\}|.$$

Thus by (7.2), (7.4), (7.6), and (7.9),

$$ES_N(\mathbf{\Delta}, \delta) \le K\delta$$

for some K which is independent of δ. By (7.1), (7.2), (7.6), (7.8), and (7.9),

$$\operatorname{Var} S_N(\mathbf{\Delta}, \delta) \to 0.$$

Therefore, (7.13) is verified.

PROOF OF THEOREM 1. For $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_3$ in R^p, $\mathbf{\Delta}_2$ in R^r, and $\mathbf{\Delta} = (\mathbf{\Delta}_1, \mathbf{\Delta}_2, \mathbf{\Delta}_3)$, define

$$\alpha_i^{(1)}(\mathbf{\Delta}) = N^{-1/2} d_i \mathbf{\Delta}_1$$

$$h_i(\mathbf{\Delta}) = h(\tau_i + x_i \mathbf{\Delta}_3 N^{-1/2})$$

$$\alpha_i^{(2)}(\mathbf{\Delta}) = \exp[-h_i(\mathbf{\Delta})\mathbf{\Delta}_2 N^{-1/2} + \{h_i(\mathbf{0}) - h_i(\mathbf{\Delta})\}\theta] - 1$$

and

$$\boldsymbol{\alpha}_i^{(3)}(\mathbf{\Delta}) = d_i \alpha_i^{(2)}(\mathbf{\Delta}).$$

Define the process

$$U_N(\mathbf{\Delta}) = N^{-1/2} \textstyle\sum_{i=1}^N \psi[\{\varepsilon_i - \alpha_i^{(1)}(\mathbf{\Delta})\}\{1 + \alpha_i^{(2)}(\mathbf{\Delta})\}]\{d_i + \boldsymbol{\alpha}_i^{(3)}(\mathbf{\Delta})\}.$$

Note that (2.4) can be rewritten as

$$U_N(N^{1/2}(\hat{\beta} - \beta), N^{1/2}(\hat{\theta} - \theta), N^{1/2}(\hat{\beta}_0 - \beta)) = 0. \tag{7.15}$$

Letting $g_i \equiv 1$, $k_i = N^{-1/2}\{1 + \|d_i\| + \|h(\tau_i)\|\}$, $\phi_i(\varepsilon_i, r, s) = \psi\{(\varepsilon_i - r)(1 + s)\}$, $d_i = z_i$, and $A(\phi, i) = A(\psi) = E\psi'$, the conditions of Theorem 7.1 are implied by (2.5) and B1 through B8, so for all $M > 0$,

$$\sup_{\|\mathbf{\Delta}\| \le M} \| U_N(\mathbf{\Delta}) - U_N(\mathbf{0}) + A(\psi) S \mathbf{\Delta}_1 \| = o_p(1). \tag{7.16}$$

Now by Chebyshev's theorem, B1 and B2,

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

In proving the theorem, we will not assume that $\hat{\beta}$ actually solves (2.4), but rather that the l.h.s. of (2.4) evaluated at $\hat{\beta}$ is less than twice its infimum over all β. However, as noted by Huber (1981, page 165), (2.4) will have a unique solution if ψ is strictly monotone. From the last equation, we have that if

$$\mathbf{\Delta}_1^* = -\{A(\psi)S\}^{-1} U_N(0) = O_p(1),$$

then by (7.16), $U(\mathbf{\Delta}^*) = o_p(1)$. Consequently, by the equivalence of (2.4) and (7.15),

$$\| U_N(N^{1/2}(\hat{\beta} - \beta), N^{1/2}(\hat{\theta} - \theta), N^{1/2}(\hat{\beta}_0 - \beta)) \| \le 2 \| U_N(\mathbf{\Delta}^*) \| = o_p(1). \tag{7.17}$$

By (2.1), we need only establish that

$$\hat{\beta} - \beta = O_p(N^{-1/2}) \tag{7.18}$$

to conclude from (7.15) and (7.16) that (2.8) holds. But by (7.17), (7.18) holds if for each $\eta > 0$, $\varepsilon > 0$ and M_1, there exists M_2 satisfying

$$P\{\inf_{\|\mathbf{\Delta}_1\| \ge M_2} \inf_{\|\mathbf{\Delta}_2\| \le M_1} \inf_{\|\mathbf{\Delta}_3\| \le M_1} \| U_N(\mathbf{\Delta}) \| > \eta\} > 1 - \varepsilon. \tag{7.19}$$

Now (7.19) follows from (7.16) in a manner quite similar to Jurečková's (1977) proof of her Lemma 5.2. □

PROOF OF THEOREM 2. For $\mathbf{\Delta}_1$ in R^p, $\mathbf{\Delta}_2$ in R^q, and $\mathbf{\Delta}' = (\mathbf{\Delta}_1', \mathbf{\Delta}_2')$, define

$$h_i(\mathbf{\Delta}) = h(\tau_i + x_i \mathbf{\Delta}_1 N^{-1/2}),$$

$$\alpha_i^{(1)}(\mathbf{\Delta}) = \exp[-h_i(\mathbf{\Delta})\mathbf{\Delta}_2 N^{-1/2} + \{h_i(\mathbf{0}) - h_i(\mathbf{\Delta})\}\theta] - 1, \tag{7.20}$$

$$\alpha_i^{(2)}(\mathbf{\Delta}) = N^{-1/2} d_i \mathbf{\Delta}_1, \tag{7.21}$$

and

$$\alpha_i^{(3)}(\mathbf{\Delta}) = h_i(\mathbf{0}) - h_i(\mathbf{\Delta}). \tag{7.22}$$

Then let $\phi(x, y, z) = \chi\{(x - z)(1 + y)\}$ and define the process

$$W_N(\mathbf{\Delta}) = -N^{-1/2} \textstyle\sum_{i=1}^N \phi\{\varepsilon_i, \alpha_i^{(1)}(\mathbf{\Delta}), \alpha_i^{(2)}(\mathbf{\Delta})\}\{h(\tau_i) - \alpha_i^{(3)}(\mathbf{\Delta})\}.$$

Note that (3.7) can be written as

$$\| W_N\{N^{1/2}(\hat{\beta}_0 - \beta), N^{1/2}(\hat{\theta} - \theta)\} \| = \text{minimum}.$$

However, by (3.6), C1 and Chebyshev's inequality,

$$W_N(\mathbf{0}) = O_p(1) \tag{7.23}$$

so that

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

We can therefore prove (3.8) by showing that for each $M_1 > 0$, $\varepsilon > 0$ and $Q > 0$, there exists $M_2 > 0$ such that

$$P[\inf\{\| W_N(\mathbf{\Delta})\| : \|\mathbf{\Delta}_1\| \le M_1, \|\mathbf{\Delta}_2\| \ge M_2\} > Q] \ge 1 - \varepsilon. \tag{7.24}$$

We will prove (7.24) by modifying the proof of Jurečková's (1977) Lemma 5.2. We first apply Theorem 7.1 with $z_i = h_i(\mathbf{0})$, $g_i \equiv 1$, $A(\phi, i) = A(\chi)$, and $k_i = N^{-1/2}\{\| h(\tau_i)\| + \| x_i \| + \| d_i \|\}$. Then

$$\sup_{\|\Delta\| \le M} \| W_N(\mathbf{\Delta}) - W_N(\mathbf{0}) + A(\chi) N^{-1/2} \textstyle\sum_{i=1}^N h(\tau_i)\alpha_i^{(1)}(\mathbf{\Delta}) \| = o_p(1).$$

By a Taylor series expansion,

$$\alpha_i^{(1)}(\mathbf{\Delta}) = -N^{-1/2} h(\tau_i)\mathbf{\Delta}_2 + \{h_i(\mathbf{0}) - h_i(\mathbf{\Delta})\}\theta + o(N^{-1/2}).$$

Thus, by C5 setting

$$G_N(\mathbf{\Delta}) = N^{-1/2} \textstyle\sum_{i=1}^N \{h_i(\mathbf{0}) - h_i(\mathbf{\Delta})\}\theta h(\tau_i),$$

we obtain

$$\sup_{\|\Delta\| \le M} \| W_N(\mathbf{\Delta}) - W_N(\mathbf{0}) - A(\chi)\mathbf{\Delta}_2^T H_N + G_N(\mathbf{\Delta}) \| = o_p(1). \tag{7.25}$$

Now fix $\varepsilon > 0$, $M_1 > 0$, $Q > 0$. Use C1 to choose γ such that

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

Define

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

Then $D < \infty$ (G_N depends only on $\mathbf{\Delta}_1$). Define M_2 by $\{A(\chi)\lambda_\infty M_2/2 - \gamma - D\} = Q$. Using C5 and (7.25), find N_0 such that $\lambda_N \ge \lambda_\infty/2$ and

$$P\{\sup_{\|\Delta_2\| = M, \|\Delta_1\| \le M_1} \| W_N(\mathbf{\Delta}) - W_N(\mathbf{0}) - A(\chi)\mathbf{\Delta}_2^T H_N - G_N(\mathbf{\Delta}) \| < \gamma/2\} \ge 1 - \varepsilon/2 \ (N \ge N_0).$$

If $\| \mathbf{\Delta}_2 \| = M_2$, $\| \mathbf{\Delta}_1 \| \le M_1$, and $N \ge N_0$, then with probability at least $1 - \varepsilon$,

$$\begin{aligned} W_N(\mathbf{\Delta})\mathbf{\Delta}_2 &\ge -M_2 \| W_N(\mathbf{0}) \| + \mathbf{\Delta}_2^T H_N \mathbf{\Delta}_2 A(\chi) - M_2 D - M_2\gamma/2 \\ &\ge \{A(\chi)\lambda_\infty M_2/2 - \gamma - D\} M_2 = QM_2. \end{aligned}$$

Since χ is nondecreasing on $[0, \infty)$ by C1, $W_N(\mathbf{\Delta}_1, \mathbf{\Delta}_2 s)\mathbf{\Delta}_2$ is a nondecreasing function of s. Thus, $\| \mathbf{\Delta}_2 \| \geq M_2$ implies

$$W_N(\mathbf{\Delta})\mathbf{\Delta}_2 \geq (\| \mathbf{\Delta}_2 \| / M_2)\{M_2 \| \mathbf{\Delta}_2 \|^{-1} W_N(\mathbf{\Delta}_1, M_2 \mathbf{\Delta}_2 \| \mathbf{\Delta}_2 \|^{-1})\mathbf{\Delta}_2\} \geq \| \mathbf{\Delta}_2 \| Q.$$

Thus,

$$P\left\{\inf_{\|\Delta_1\| \leq M_1, \|\Delta_2\| \geq M_2} \frac{W_N(\mathbf{\Delta})\mathbf{\Delta}_2}{\| \mathbf{\Delta}_2 \|} \geq Q\right\} \geq 1 - \varepsilon,$$

which with the Cauchy-Schwarz inequality proves (3.8). □

REFERENCES

ANSCOMBE, F. J. (1961). Examination of residuals. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* (J. Neyman, editor) 1-36. University of California Press, Berkeley.

BICKEL, P. J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70** 428–434.

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

BICKEL, P. J. and DOKSUM, K. A. (1981). An analysis of transformations revisited. *J. Amer. Statist. Assoc.* **76** 296–311.

BOX, G. E. P. and HILL, W. J. (1974). Correcting inhomogeneity of variance with power transformation weighting. *Technometrics* **16** 385-389.

CARROLL, R. J. (1980). A robust method for testing transformations to achieve approximate normality. *J. Royal Statist. Soc. Ser. B* 71–78.

CARROLL, R. J. and RUPPERT, D. (1981). On robust tests for heteroscedasticity. *Ann. Statist.* **9** 206–210.

DENT, W. T. and HILDRETH, C. (1977). Maximum likelihood estimation in random coefficient models. *J. Amer. Statist. Assoc.* **72** 69–72.

FROEHLICH, B. R. (1973). Some estimators for a random coefficient regression model. *J. Amer. Statist. Assoc.* **68** 329–335.

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.

HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte-Carlo. *Ann. Statist.* **5** 799–821.

HUBER, P. J. (1981). *Robust Statistics.* Wiley, New York.

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.* **75** 176–181.

JUREČKOVÁ, J. (1977). Asymptotic relations of M-estimates and R-estimates in linear regression model. *Ann. Statist.* **5** 464–472.

KRASKER, W. S. and WELSCH, R. E. (1981). Efficient bounded-influence regression estimation using alternative definitions of sensitivity. To appear in *J. Amer. Statist. Assoc.*

9810 PARKWOOD DRIVE
BETHESDA, MARYLAND 20814

DEPT. OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
AT CHAPEL HILL
321 PHILIPPS HALL 039A
CHAPEL HILL, NORTH CAROLINA 27514