ROBUST ESTIMATORS FOR RANDOM COEFFICIENT REGRESSION MODELS

by

R.J. Carroll[1]

and

David Ruppert[2]

Running Title: RANDOM COEFFICIENT REGRESSION

[1]National Heart, Lung, and Blood Institute and the University of North Carolina. Supported by the Air Force Office of Scientific Research Grant AFOSR F49620 82 C 0009.

[2]University of North Carolina. Supported by National Science Foundation Grant MCS 8100748.

Key Words: M-estimation, random coefficients, heteroscedasticity, estimated weights, Monte Carlo.

AMS 1970 Subject Classifications: 62G35, 62J05.

ABSTRACT

Random coefficient regression models have received considerable attention, especially from econometricians. Previous work has assumed that the coefficients have normal distributions. The variances of the coefficients have, in previous papers, been estimated by maximum likelihood or by least squares methodology applied to the squared residuals from a preliminary (unweighted) fit.

Maximum likelihood estimation poses difficult numerical problems. Least squares estimation of the variances is inefficient because the squared residuals have a distribution with a heavy right tail.

In this paper we propose several robust estimators for random coefficients models. We compare them by Monte Carlo with estimators based on least squares applied to the squared residuals. The robust estimators are best overall, even at the normal model.

Among the different robust estimators, none stands out as best. All are rather satisfactory and can be tentatively recommended for routine use.

1. Introduction.

There is now a sizable literature on linear regression models where the regression coefficients are random rather than fixed parameters. See, for example, Dent and Hildreth (1977), Hildreth and Houck (1968), Theil and Mennes (1959), Froehlich (1973), Fisk (1967), Spjotvoll (1977), and Swamy (1971). Such models can be expressed in the form

$$y_i = \underline{x}_i' \underline{\beta}_i \quad , \quad i = 1, \ldots, n \tag{1.1}$$

where $\underline{x}_i' = (x_{i1}, \ldots, x_{ik})$ is a known vector of independent variables, and $\underline{\beta}_i' = (\beta_{i1}, \ldots, \beta_{ik})$ is the unobserved vector of regression coefficients for the ith observation. We will assume, as is usual, that $\underline{\beta}_1, \ldots, \underline{\beta}_n$ are i.i.d. with means $\underline{\beta}$ and covariance matrices $\text{diag}(\alpha_1, \ldots, \alpha_k)$. Moreover, we will be concerned solely with estimation of $\underline{\beta}$ and $\underline{\alpha} = (\alpha_1, \ldots, \alpha_k)'$. We will not consider estimation of the individual coefficient vectors $\underline{\beta}_1, \ldots, \underline{\beta}_k$, and in fact we will not assume that there is enough information in the sample to do so. Allowing $\underline{\alpha}_i$ to have a non-diagonal covariance matrix would not present great theoretical difficulties. However, such models have a considerable number of parameters to estimate. Monte-Carlo studies with diagonal covariance models, such as the work presented here, suggest that rather large sample sizes would be necessary before estimation would be reasonably accurate for models with non-diagonal covariance matrices.

Notice that model (1.1) can be re-expressed as

$$y_i = \underline{x}_i'\underline{\beta} + \varepsilon_i \tag{1.2}$$

where $E\varepsilon_i = 0$ and

$$E\varepsilon_i^2 = \dot{\underline{x}}_i'\underline{\alpha} \ . \tag{1.3}$$

Here and throughout this paper, for any vector or matrix A, $\dot{A}$ is obtained by squaring each entry of A. We see then that for purposes of estimation, the random coefficient model can be treated as a fixed coefficient model with heteroscedasticity.

In particular, it would be possible to ignore $\underline{\alpha}$ and simply estimate $\underline{\beta}$ by an unweighted method. However, there are two reasons for estimating $\underline{\alpha}$: (1) the variances of the random coefficients may be of interest in themselves, or (2) only $\underline{\beta}$ may be of intrinsic interest but one hopes to obtain improved estimates of $\underline{\beta}$ by using weights based upon $\hat{\underline{\alpha}}$, rather than using an unweighted estimate.

Estimators of $\underline{\alpha}$ and $\underline{\beta}$ have been studied by Theil and Mennes (1959), Hildreth and Houck (1968), Froehlich (1973), and Dent and Hildreth (1977). These estimators each fall into one of two categories. The first category consists of various methods which attempt to find the maximum of the likelihood function. The second consists of finding a preliminary (unweighted) least squares estimator of $\underline{\beta}$, then estimating $\underline{\alpha}$ by applying least squares (possibly weighted) to the squared residuals from this fit, and finally re-estimating $\underline{\beta}$ by least squares using weights from the estimate of $\underline{\alpha}$.

Until now it has been assumed in the literature that $\underline{\beta}_i$ has a normal distribution. In this paper we introduce robust estimators for the random coefficients model. These estimators offer three possible advantages over previous estimators:

(i) They should be more efficient when coordinates of $\underline{\beta}_i$ have heavy tailed distributions, and they should be less sensitive to gross errors in the response.

(ii) Even when $\underline{\beta}_i$ is normally distributed, the squared residuals from a preliminary estimator of $\underline{\beta}$ will have a heavy right tail. When estimating $\underline{\alpha}$, robust methods might be superior to least squares.

(iii) When re-estimating $\underline{\beta}$ using estimated weights, robust methods might protect against poor estimates of the weights.

The purpose of this paper is to suggest several robust estimators and to study the extent to which these possible advantages are realized. Points (i) and (ii) can be explored theoretically by asymptotics. Point (iii) requires Monte-Carlo studies. This is because finite sample exact results seem impossible to obtain, and asymptotically any consistent estimate of $\underline{\alpha}$ gives weights which, when estimating $\underline{\beta}$, are just as good as the true weights (in terms of first order asymptotic distributions).

In section 2, we describe those estimates that have performed best in previous Monte Carlo studies. In section 3 we introduce some new methods which are distribution robust.

For the homoscedastic linear model, the so-called bounded influence estimates studied by e.g. Krasker and Welsch (1981) are robust when there are outliers in the design (in the x vectors). The need for bounded influence estimators has been questioned by Huber (1981, section 7.9). Nonetheless, we feel that bounded influence estimators should be developed for heteroscedastic models, including the random coefficients model. In the present paper, however, we do not consider estimators which bound the influence of the design vectors.

In section 4, we present a Monte Carlo study and draw some tentative conclusions from it. We see that all three possible advantages of robust estimation, i, ii, and iii above, are realized for the sampling situations we studied.

2. Previous Estimators.

We will only mention the estimators which proved best in the Monte Carlo studies of Froehlich (1973) and in Dent and Hildreth (1977). Equation (1.2) can be put in matrix form:

$$\underline{y} = X\underline{\beta} + \underline{e} \ . \tag{2.1}$$

If $M = I - X(X'X)^{-1}X'$ and $G = \ddot{M}\ddot{X}$, then as noted by Hildreth and Houck (1968), the vector of squared residuals from a least squares fit to (1.2) is

$$\dot{\underline{r}} = G\underline{\alpha} + \underline{w} \tag{2.2}$$

where $E\underline{w} = 0$. Therefore one can estimate $\underline{\alpha}$ by applying least squares methodology to (2.2). In so doing, one should utilize the constraints: $\alpha_j \geq 0$, $j = 1, \ldots, k$. One could simply truncate the least squares estimators below by 0. Alternatively, Dent and Hildreth (1977) and Froehlich (1973) define $\hat{\underline{\alpha}}$ to be the minimizer of $||\dot{\underline{r}} - G\underline{\alpha}||^2$ subject to these nonnegativity constraints. Froehlich's (1973) Monte Carlo study shows that $\hat{\underline{\alpha}}$ is superior to the unrestricted least squares estimator, either with or without truncation at 0. The calculation of $\hat{\underline{\alpha}}$ is a quadratic programming problem, and Dent and Hildreth (1977) used Lemke's (1965) method, in particular Ravindran's (1972) FORTRAN algorithm, to calculate $\hat{\underline{\alpha}}$.

Froehlich (1973) and Dent and Hildreth (1977) found value in using the weighted estimate of $\underline{\alpha}$ that was suggested by Theil and Mennes (1959). To calculate this estimator one starts with $\hat{\underline{w}} = \dot{\underline{r}} - G\hat{\underline{\alpha}}$, which is the vector of residuals from the restricted least square fit to (2.2). The covariance matrix of w is estimated by $\underline{\hat{w}\hat{w}'}$. Theil and Mennes (1959) have argued that one may ignore the off-diagonal elements, so let us define $\hat{\Sigma}$ to be $\hat{w}\hat{w}'$ with the off-diagonal elements replaced by 0; formally

$$\hat{\Sigma} = I*(\underline{\hat{w}\hat{w}'})$$

where * denotes Hadamard product. The Theil-Mennes estimators, which Dent and Hildreth (1977) call $\hat{\underline{\alpha}}(2)$, is the solution to

$$\text{minimize } (\dot{\underline{r}} - G\underline{\alpha})\Sigma^{-1}(\dot{\underline{r}} - G\underline{\alpha})$$

$$\text{subject to } \alpha_i \geq 0 \quad , \; i = 1,\ldots, k \; .$$

This is, of course, a restricted, weighted least squares estimate.

Note that if $\underline{\beta}_i$ has a multivariate normal distribution, then $r_i^2 \doteq e_i^2$ and e_i^2/E_i^2 has a chi-square distribution with one degree of freedom. Thus, r_i^2 will have a very heavy right tail, and we can expect that least squares estimation applied to (2.2) will not be very efficient. Even restricted and weighted least squares should be inferior to a good robust regression estimator.

If the $\underline{\beta}_i$ are normally distributed, then of course one can obtain asymptotically efficient estimators by maximum likelihood. There are, however, two difficulties with the MLE. As is typical with normal-based likelihood methods, the MLE will be sensitive to heavy-tailed deviations from the normality assumption and to the presence of a few gross outliers.

The second difficulty is in finding the maximum of the likelihood function. Froehlich (1973) tried to find the MLE using the Newton-Raphson algorithm, but this failed to converge in approximately twenty percent of the samples. Dent and Hildreth (1977) tried three methods of maximizing the likelihood function: the Davidson-Fletcher-Powell algorithm, Fisher's method of scoring, and Brent's (1973) PRAXIS algorithm. The three algorithms often gave different optima, and in almost all cases the solution from PRAXIS gave a higher value of the likelihood than the solutions from the two other algorithms. Brent (1973) gave an ALGOL version of PRAXIS. Dent and Hildreth (1977) report using a FORTRAN version. We have been unable to obtain a FORTRAN version of PRAXIS.

3. Robust Estimators.

First we review robust estimators for homoscedastic linear models. Readers unfamiliar with this material are referred to Huber (1981) for further discussion. Consider model (1.2) but with the e_i's i.i.d. with a symmetric distribution F. An M-estimator of $\underline{\beta}$ is a solution to

$$\sum_{i=1}^{n} \psi\left(\frac{y_i - \underline{x}_i' \hat{\underline{\beta}}}{\hat{\sigma}}\right) \underline{x}_i = \underline{0} \tag{3.1}$$

where ψ is an appropriate function and $\hat{\sigma}$ is a scale estimate. Typically, ψ is odd so that $\hat{\underline{\beta}}$ will be consistent, and ψ bounded so that $\hat{\underline{\beta}}$ will not be highly sensitive to outliers in the response. The choice of $\psi_k(x) = \max(-k, \min(x,k))$ for some k between 1 and 2 is common. Typical choices of $\hat{\sigma}$ are the standardized median absolute deviation, that is, MAD/(.6745), and Huber's proposal 2 which solves (3.1) and

$$\sum_{i=1}^{n} \chi\left(\frac{y_i - \underline{x}_i' \underline{\beta}}{\hat{\sigma}}\right) = 0$$

simultaneously. Often, $\chi(x) = \psi^2(x) - \int \psi^2(x)\, d\Phi(x)$, where Φ is the standard normal distribution.

If F is asymmetric, and the model (1.2) has an intercept so that the first (say) coordinate of $\underline{x}_i$ is 1 for all i, then $\hat{\beta}_i$ consistently estimates β_i for i = 2,..., k, but $\hat{\beta}_1$ need not be consistent; see Carroll (1979). Of course, consistency will only be obtained under regularity conditions on the $\underline{x}_i$ and F and/or ψ, but we will not pursue such niceties here.

Notice that if we apply an M-estimate to (2.2), then $\hat{\alpha}_i$ will be consistent for i = 2,..., k. We may, therefore, wish to estimate α_1 in an auxiliary manner. The M-estimate of $\underline{\alpha}$ can be unweighted so that a robust analog of the Hildreth-Houck estimator is obtained. Alternatively, we could use a weighted M-estimator with weights given by $\hat{\Sigma}$ and obtain a robust analog of the

Theil-Mennes estimator. A weighted M-estimator with weights $w_1, \ldots, w_n$ is defined to be an ordinary M-estimator applied to y_i/w_i and $\underline{x}_i/w_i$. If the ordinary M-estimator is scale equivariant because it utilizes a scale estimator $\hat{\sigma}$ as in equation (3.1), then the weighted M-estimator is unaffected by replacing w_i by kw_i, $k > o$, for $i = 1, \ldots, n$.

We will now describe a robust estimator of $\underline{\alpha}$ which, at the normal model, is consistent for α_1 as well as $\alpha_2, \ldots, \alpha_k$. Let $\underline{\tilde{\alpha}}$ and $\underline{\tilde{\beta}}$ be consistent, preliminary estimators of $\underline{\alpha}$ and $\underline{\beta}$. Define

$$\xi = \int \psi(z^2 - 1)\, d\Phi(z) ,$$

(3.2) $$r_i = y_i - \underline{x}_i'\underline{\tilde{\beta}}$$

and

(3.3) $$w_i = \underline{\dot{x}}_i'\underline{\tilde{\alpha}} \quad .$$

Let G_i' be the ith row of G. Then, let $\underline{\hat{\alpha}}_M$ be the solution to

(3.4) $$\sum_{i=1}^{n} \left\{ \psi\left(\frac{r_i^2 - \underline{G}_i'\underline{\hat{\alpha}}_M}{w_i}\right) - \xi \right\} \frac{\underline{G}_i'}{w_i} = 0 \ .$$

Although the r_i^2 have an asymmetric distribution, subtraction of ξ from ψ ensures consistency of $\underline{\hat{\alpha}}_M$ at the normal model.

Needless to say, the small sample properties of robustified Hildreth-Houck and Theil-Mennes estimator, or of $\underline{\hat{\alpha}}_M$, depend upon the particular choice of preliminary estimates. In the next section, we define several specific versions of these robustified estimators, and the study of them by Monte Carlo.

4. Monte Carlo Comparison of the Estimators.

In our simulation study, we used six estimation algorithms. Each algorithm consisted of the following steps:

1) A preliminary unweighted estimate, $\hat{\underline{\beta}}_p$, of $\underline{\beta}$ is computed.

2) An unweighted estimate $\hat{\underline{\alpha}}_{HH}$, of $\underline{\alpha}$ is found by fitting model (2.2).

3) Using weights $\dot{\underline{x}}_i'\underline{\alpha}_{HH}$, one computes a weighted estimate of $\underline{\beta}$ and calls it $\hat{\underline{\beta}}_{HH}$.

4) Let $\sum = I*\hat{\underline{w}}\hat{\underline{w}}'$ where $\hat{\underline{w}}$ is the vector of residuals from step 2. Using these weights, one computes a weighted estimate $\hat{\underline{\alpha}}_{TM}$ of $\underline{\alpha}$.

5) Using weights $\dot{\underline{x}}_i'\underline{\alpha}_{TM}$, one finds a weighted estimate of $\hat{\underline{\beta}}$ and calls it $\hat{\underline{\beta}}_{TM}$.

6) Using $\hat{\underline{\alpha}}_{TM}$ and $\hat{\underline{\beta}}_{TM}$ as preliminary estimates, one computes r_i and w_i from equation (3.2) and (3.3). Then $\underline{\alpha}$ is estimated by solving (3.4). Call the estimator $\hat{\underline{\alpha}}_M$.

7) Using weights $\dot{\underline{x}}_i'\hat{\underline{\alpha}}_M$, one constructs an estimate of $\underline{\beta}$ and calls the result $\hat{\underline{\beta}}_M$.

Note that in order to characterize an algorithm exactly, in steps 1, 2, 3, 4, 5, and 7 we must specify precisely which estimate is used. Table 1 gives this information for each of the six algorithms that we employed. In each case the estimate is either least squares or Huber's proposal 2 with $\psi = \psi_{1.5}$ and $x = \psi^2 - \int \psi^2 \, d\Phi$. When denoting an estimate, we include the specific algorithm used. Thus $\hat{\underline{\alpha}}_{HH}(1)$ is the estimate from step 3 of algorithm 1 and is, in fact, the estimate $\hat{\underline{\alpha}}$ of Froehlich (1973) and of Dent and Hildreth (1977). We also use HH(1) to denote both $\hat{\underline{\alpha}}_{HH}(1)$ and $\hat{\underline{\beta}}_{HH}(1)$. See table 7 for a summary of the acronyms we used.

When estimating $\underline{\alpha}$ by least squares (weighted or unweighted) we used restricted least squares calculated by Ravindran's (1972) algorithm. We did not develop a restricted M-estimate, so to estimate $\underline{\alpha}$ robustly, we truncated the proposal 2 estimator at 0. Because the squared residuals have an asymmetric distribution, in general the M-estimators in steps 2 and 4 will not estimate α_1 consistently. Therefore, in algorithm 5 we tried a separate estimate of α_1. This was the average of

$$\dot{r}_i - (0\ \hat{\alpha}_2\ \hat{\alpha}_3)G_i'$$

where $(\dot{r}_1, \ldots, \dot{r}_n) = \dot{\underline{r}}'$ and G_i is the ith row of G, where G is given in section 2. Also, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ are the proposal 2 estimates truncated at 0.

In algorithms 1 and 3-6, we solve equation (3.4) exactly in step 6. Occasionally, the algorithm for solving (3.4) did not converge, and then we set $\hat{\underline{\alpha}}_M$ equal to $\hat{\underline{\alpha}}_{TM}$. In algorithm 2, we did not try to solve (3.4) exactly, but rather we stopped after two steps towards the solution to (3.4).

Notice that $\hat{\underline{\alpha}}$ and $\hat{\underline{\alpha}}(2)$ of Dent and Hildreth (1977) are the same estimators as HH1 and TM1, respectively. Of course, HH1 = HH2 and TM1 = TM2, so HH2 and TM2 are not included in the report of our results.

In our study, we used the same designs as Dent and Hildreth (1977), i.e., the first column is of ones, the second is harmonic, and the third is random where "harmonic" and "random" are as described by Froehlich (1973). We used sample sizes n = 25 and n = 75, as did Dent and Hildreth, and 300 Monte Carlo iterations.

We used $\underline{\beta} = (1.0, 1.0, 1.0)'$ as did Dent and Hildreth (1977) and Froehlich (1973). For $\underline{\alpha}$ we used $\underline{\alpha} = (1.0, 0.2, 0.5)'$ which we call "mild heteroscedasticity" and which was used by Dent and Hildreth and Froehlich. We also used $\underline{\alpha} = (1.0, 3.0, 3.0)'$, which we called "heavy heteroscedasticity." As we will see, when $\underline{\alpha} = (1.0, 0.2, 0.5)'$ unweighted estimates of $\underline{\beta}$ (which ignore the heteroscedasticity) are rather good, but for $\underline{\alpha} = (1.0, 3.0, 3.0)'$ one always does better by estimating the variance function and using a weighted estimate of $\underline{\beta}$.

The random coefficient error $\beta_{ik} - \beta_i$ was $\sqrt{\alpha_k}\, Z_{ik}$, where the Z_{ik} were either independent $N(0,1)$ variates (normal) or independent $[.9N(0,1) + .1N(0,9)]$ variates (contaminated normal). Notice that the variance of β_{ik} is α_k for the normal sampling situation, but $1.8\alpha_k$ for the contaminated normal sampling situation. A robust scale functional which is Fisher consistent at the normal distribution, for example, the standardized MAD, will be α_k for the normal distribution and close to α_k, not $1.8\alpha_k$, for the contaminated normal distribution. Therefore, for the contaminated distribution, the question arises as to the parameter being estimated. However, the ratios $\sigma(\beta_{ik})/\sigma(\beta_{ik'})$ are independent of the particular scale functional, $\sigma(\cdot)$, that is employed. Moreover, when using the weights $\dot{\underline{x}}_i\underline{\alpha}$ to estimate $\underline{\beta}$ by a weighted proposal 2 (or weighted least squares), only the ratios $(\hat{\alpha}_k/\hat{\alpha}_{k'})$ are relevant. This is because $\underline{\beta}$ is being estimated by a scale equivariant procedure.

To improve the accuracy of comparisons between the various estimators and sampling situations, for each sample size we used the same stream of random numbers for all estimators and sampling situations. The only exception is that for the contaminated normal distribution but not for the normal distribution we needed Bernoulli (.1) random variables in order to decide if Z_{ik} was to be $N(0,1)$ or $N(0,9)$.

Since we used two sample sizes, two choices of $\underline{\alpha}$, and two distributions we had eight sampling situations.

In table 2, we give the mean square error (MSE) for the estimators of $\underline{\beta}$. For each sampling situation and each of β_1, β_2, and β_3 we report the MSE of only the seven estimators with the lowest MSE. In table 3, MSE are given for estimates of $\underline{\alpha}$ in the normal sampling situations. To compare estimates of $\underline{\alpha}$ under the contaminated normal distribution, we use the parameters $\log(\alpha_1/\alpha_2)$, $\log(\alpha_1/\alpha_3)$, and $\log(\alpha_2/\alpha_3)$. Of course, $\log(\hat{\alpha}_i/\hat{\alpha}_j)$ is undefined if $\hat{\alpha}_i = 0$ or

$\hat{\alpha}_j = 0$, which is a common occurence for estimators truncated at zero. Therefore to compare the estimators of $\log(\alpha_i/\alpha_j)$, we computed the median absolute errors (MAE), which are the medians of $|\log(\hat{\alpha}_i/\hat{\alpha}_j) - \log(\alpha_i/\alpha_j)|$. If $\hat{\alpha}_i = 0$ or $\hat{\alpha}_j = 0$, we set $\log(\hat{\alpha}_i/\hat{\alpha}_j) = \infty$. The MAE is finite provided $\hat{\alpha}_i \neq 0$ and $\hat{\alpha}_j \neq 0$ for more than fifty percent of the samples, which was always the case in our study. The advantage of using the parameter $\log(\alpha_i/\alpha_j)$, rather than (α_i/α_j) is that the MAE of $\log(\alpha_i/\alpha_j)$ is symmetric in i and j. The MAE are given in table 4.

From table 2 we can draw some conclusions pertaining to estimation of $\underline{\beta}$.

1) When n = 25 and the heteroscedasticity is mild, then unweighted estimators are competitive with those which use $\underline{\hat{\alpha}}$ to estimate the optimal weights. However, weighted estimates always work about as well or better than unweighted estimates.

2) The robust estimators of $\underline{\beta}$, especially HH4, HH5, TM4, TM5, M4, M5, and M6, are in general quite good and can perhaps be recommended for routine use.

3) Even at the normal model, robust methods typically outperform least squares methods. This may be partly due to the heavy right tail of the distribution of the squared residuals, which makes robust estimators of $\underline{\alpha}$ more efficient than least squares, even at the normal model. However, HH6 and TM6 which utilize robust estimators of $\underline{\alpha}$ and least squares estimators of $\underline{\beta}$, are not uniformly better, at the normal model, than HH4, HH5, TM4, TM5, and M4-M6. These later estimators utilize robust estimators of both $\underline{\alpha}$ and $\underline{\beta}$. This suggests that robust estimators of $\underline{\beta}$ guard against poor choices of weights caused by estimation errors for $\underline{\alpha}$.

4) Whether Hildreth-Houck (HH) type estimators, Theil-Mennes (TH) type estimators, or estimators based on equation (3.4) (which we denote by M) are

best depends upon the sampling situation and the particular coordinate of $\underline{\beta}$ being estimated.

Examining tables 3 and 4, we see that when estimating $\underline{\alpha}$, robust methods outperform non-robust methods, even at the normal model. Moreover, algorithms 4 and 5 which use robust estimators throughout typically outperform algorithms 3 and 6 which use both robust estimates and least squares.

Although HH4 and TM4 do not produce consistent estimators of α_1, in our study, they worked well when estimating α_1, $\log(\alpha_1/\alpha_2)$, and $\log(\alpha_1/\alpha_3)$. This suggests that the biases of HH4 and TM4 for α are typically quite small compared to the standard deviations.

We constructed confidence intervals for β_1, β_2, and β_3 as follows. Define

$$S(\underline{\alpha}) = \{n^{-1} \sum_{i=1}^{n} \underline{x}_i \underline{x}_i' / \dot{\underline{x}}_i \underline{\alpha})\}^{-1}$$

and

$$\sigma^2(\underline{\alpha},\underline{\beta}) = (n-k)^{-1} \sum_{i=1}^{n} (y_i - \underline{x}_i'\underline{\beta})^2 / (\dot{\underline{x}}_i'\underline{\alpha}) .$$

Let $t(\gamma,n)$ be the $(1 - \gamma)$ quantile of the t distribution with n degrees of freedom, and let $S^{(j)}(\underline{\alpha})$ be the jth diagonal element of $S(\underline{\alpha})$. Define

$$\hat{\sigma}_i = (\dot{\underline{x}}_i'\hat{\underline{\alpha}})\ \hat{\sigma}$$

where $\hat{\underline{\alpha}}$ is a given estimate of $\underline{\alpha}$ and $\hat{\sigma}$ is the scale estimate from the proposal 2 estimate of $\underline{\beta}$. Then, let

$$r_i = \frac{y_i - \underline{x}_i'\hat{\underline{\beta}}}{\hat{\sigma}_i} ,$$

$$\hat{\lambda}_1 = n^{-1} \sum_{i=1}^{n} \dot{\psi}(r_i) ,$$

$$\hat{\lambda}_2 = (n-k)^{-1} \sum_{i=1}^{n} \psi^2(r_i) ,$$

and

$$\hat{K} = 1 + \frac{2k(1-\hat{\lambda}_1)}{\hat{\lambda}_1(h-2k)} ,$$

where $\psi = \psi_{1.5}$, $\dot{\psi}(x) = (d/dx)\psi(x)$, and $\hat{\underline{\beta}}$ is a given estimate of $\underline{\beta}$. $\hat{K}$ corresponds to K of Huber (1981, page 174), but we use 2k instead of k as an ad hoc adjustment for having extra parameters ($\underline{\alpha}$) to estimate.

If $\hat{\underline{\beta}}$ is a weighted least squares estimator, then the confidence intervals are

$$\hat{\beta}_j \pm \sigma(\hat{\underline{\alpha}},\hat{\underline{\beta}})t(\gamma/2,n-k)S^{(j)}(\hat{\underline{\alpha}}) \ .$$

If $\hat{\underline{\beta}}$ is the unweighted proposal 2 estimator, then the intervals are

$$\hat{\beta}_j \pm \frac{\hat{\lambda}_2}{\hat{\lambda}^2} \hat{K}\ S^{(j)}(\hat{\underline{\alpha}})t(\gamma/2,n-k) \ .$$

The least squares intervals are obtained by applying standard methodology to $y_i/(\dot{\underline{x}}_i'\hat{\underline{\alpha}})$ and $\underline{x}_i/(\dot{\underline{x}}_i'\hat{\underline{\alpha}})$. The confidence intervals for proposal 2 are based on the asymptotics in, for example, Huber (1981, chapter 7), again applied after weighting by $1/(\dot{\underline{x}}_i'\hat{\underline{\alpha}})$. When constructing these intervals we make no adjustments for the fact that we are using estimated weights, $\dot{\underline{x}}'\hat{\underline{\alpha}}$, not $\dot{\underline{x}}'\underline{\alpha}$. However, by using techniques from other studies of heteroscedastic models, e.g., Carroll and Ruppert (1982), one can show that each of the weighted estimation methods considered here has the following property. The asymptotic distribution of $\hat{\underline{\beta}}$ is the same when the weights $\dot{\underline{x}}_i'\hat{\underline{\alpha}}$ are used as when the weights $\dot{x}_i'\underline{\alpha}$ are used. The adjustment $\hat{K}$ is based on a higher order asymptotic expansion of Huber (1973, 1981), and was found to be essential in a Monte Carlo study by Schrader and Hettmansperger (1980).

Table 5 gives Monte Carlo coverage probabilities for n = 25 and for selected estimators of $\underline{\beta}$. We used $\gamma = .05$. The coverage probabilities are quite close to .95, especially since the standard deviation of these estimated probabilities is 0.0126, if the true probabilities are, in fact, .95. The coverage probabilities for n = 75 and for the other estimators are also close to .95, except that HH1, HH3, TM1, TM3, M1, M2, and M3 typically have coverage probabilities for β_1 between 0.83 and 0.90 when $\underline{\alpha}$ = (1.0, 3.0, 3.0)'.

Among the estimators included in table 5 are the least squares estimator and proposal 2. By examining their coverage probabilities, we can see that ignoring the heteroscedasticity, as these estimators do, does not seriously degrade the validity of the confidence intervals.

5. Conclusions.

The estimators which are robust for both $\underline{\beta}$ and $\underline{\alpha}$, that is HH4, HH5, TM4, TM5, M4, and M6 are very satisfactory overall. We recommend them over the standard Hildreth-Houck and Theil-Mennes estimators, even if there is no worry about normality. We did not consider maximum likelihood estimation in this study, because of numerical difficulties which are involved. As numerical techniques improve and good algorithms become more available, maximum likelihood estimation will become a feasible technique. Even then, however, maximum likelihood estimators of variance parameters have quadratic influence functions and are particularly nonrobust.

Thus we believe that robust estimators, either the ones studied here, or perhaps ones which will be introduced in the future (e.g. estimators with bounded influence), should be standard techniques and at the very least should be computed along with non-robust methods.

6. Further work.

As mentioned previously, one can show that the distribution of $\hat{\underline{\beta}}$ for each of the weighted estimation methods considered here is the same asymptotically when the estimated weights $\dot{\underline{x}}_i'\hat{\underline{\alpha}}$ are used as when the weights $\dot{\underline{x}}_i'\underline{\alpha}$ based on the true value of $\underline{\alpha}$. This conclusion may be established using techniques developed in Carroll and Ruppert (1982), though the results there are not directly applicable since the random variables

$$\frac{y_i - \underline{x}_i'\underline{\beta}}{(\dot{x}_i\alpha)^{\frac{1}{2}}}$$

are generally not identically distributed. The result may not have been widely known previously. Froehlich (1973) states that he "had limited success so far establishing the asymptotic distribution of the resulting estimators of $\underline{\beta}$." Dent and Hildreth (1977) make no mention of asymptotic properties.

The asymptotic distribution of $\hat{\underline{\alpha}}$ remains an open problem. Influence functions for the estimators, conditions for consistency, and further results on asymptotic distributions, particularly for $\hat{\underline{\alpha}}$, are the focus of current research which will be reported at a later date.

REFERENCES

Brent, Richard P. (1973). *Algorithms for Minimization Without Derivatives.* Englewood Cliffs, N.J.: Prentice-Hall.

Carroll, R.J. (1979). On Estimating Variances When the Errors Are Asymmetric. *J. Am. Statist. Assoc., 74.* 674-679.

Carroll, R.J. and David Ruppert. (1982). Robust Estimation in Heteroscedastic Linear Models. To appear in *Ann. Stat.*

Dent, Warren T. and Hildreth, Clifford. (1977). Maximum Likelihood Estimation in Random Coefficient Models. *J. of Am. Statist. Assoc., 72.* 69-72.

Fisk, P.R. (1967). Models of the Second Kind in Regression Analysis. *J. Roy. Statist. Soc. B, 29.* 266-281.

Froehlich, B.R. (1973). Some Estimators for a Random Coefficient Regression Model. *J. Am. Statist. Assoc., 68.* 329-335.

Hildreth, Clifford and Houck, James. (1968). Some Estimators for a Linear Model with Random Coefficients. *J. Am. Statist. Assoc., 63.* 584-595.

Huber, Peter J. (1973). Robust Regression: Asymptotics, Conjectures and Monte Carlo. *Ann. Statist., 5.* 799-821.

Huber, Peter J. (1981). *Robust Statistics.* John Wiley and Sons, N.Y.

Lemke, C.E. (1965). Bimatrix Equilibrium Points and Mathematical Programming. *Management Science, 11.* 681-689.

Krasker, William S. and Welsch, Roy E. (1981). Efficient Bounded-Influence Regression Estimation. Manuscript. (To appear in *J. Am. Statist. Assoc.*).

Ravindran, Arunachalan. (1972). A Computer Routine for Quadratic and Linear Programming Problems [H]. *Communications of the ACM, 5.* 818-820.

Spjotvoll, E. (1977). Random Coefficients Regression Models. A Review. *Math. Operationsforsch. Statist., Ser. Statist., 8.* 69-93.

Swamy, P.A.V.B. (1971). *Statistical Inference in Random Coefficient Regression Models.* Springer-Verlag, Berlin.

Theil, Henri and Mennes, L.B.M. (1959). Conception Stochastique de Coefficients Multiplicateurs dons L'ajustement Lineaire des Series Temporelles. *Publications de l'Institut de Statistique de L'University de Paris, 8.* 211-227.

Algorithm		1,2	3	4	5	6
Step, parameter being estimated and notation for estimation	1 $\underline{\beta}$	Least squares	Proposal 2	Proposal 2	Proposal 2	Least squares
Step = Parameter = Notation =	2 $\underline{\alpha}$ HH	Least squares -restricted	Least squares -restricted	Proposal 2, truncate at 0	Proposal 2, truncate at 0, separate estimate of α_1	Proposal 2, truncate at 0
Step = Parameter = Notation =	3 $\underline{\beta}$ HH	Weighted least squares	Weighted Proposal 2	Weighted Proposal 2	Weighted Proposal 2	Weighted least squares
Step = Parameter = Notation =	4 $\underline{\alpha}$ TM	Weighted least squares-restricted	Weighted least squares-restricted	Weighted Proposal 2	Weighted Proposal 2, truncate at 0, separate estimate of α_1	Weighted Proposal 2, truncate at 0
Step = Parameter = Notation =	5 $\underline{\beta}$ TM	Weighted least squares	Weighted Proposal 2	Weighted Proposal 2	Weighted Proposal 2	Weighted least squares
Step = Parameter = Notation =	7 $\underline{\beta}$ M	Weighted least squares	Weighted Proposal 2	Weighted Proposal 2	Weighted Proposal 2	Weighted Proposal 2

Table 1. Estimates used in the six algorithms used in our Monte Carlo study. Weighted Proposal 2 with weights $w_1, \ldots, w_n$ is obtained by applying Proposal 2 to $y_1/w_1, \ldots, y_n/w_n$, $\underline{x}_1/w_1, \ldots, \underline{x}_n/w_n$. The separate estimate of α_1 is described in the text. In algorithms 1, 3, 4, 5, and 6, step 6 solves equation (3.4) with preliminary estimates coming from steps 4 and 5. Algorithm 2 differs in that in step 6 only two steps towards the solution of (3.4) are used. Any estimator can be identified by giving the step at which it was produced and the algorithm number used. Thus TM4 denotes $\hat{\underline{\alpha}}$ from step 4 and $\hat{\underline{\beta}}$ from step 5, both from algorithm 4. The estimator of $\underline{\beta}$ from step 6 is denoted by M.

Sampling Situation

		1		2		3		4		5		6		7		8	
Sample size		25		25		75		75		25		25		75		75	
Distribution		N		N		N		N		CN		CN		CN		CN	
Heteroscedasticity		M		H		M		H		M		H		M		H	
	1	PT	0.0673	M5	0.2323	HH5	0.0217	M5	0.0734	HH5	0.0993	M5	0.3175	M5	0.0319	M3	0.1009
	2	LS	0.0675	M4	0.2388	M2	0.0217	M6	0.0735	HH4	0.1048	HH5	0.3255	HH5	0.0321	M2	0.1029
Rank for	3	HH5	0.0689	M6	0.2395	M3	0.0218	M4	0.0752	M5	0.1050	TM5	0.3279	TH5	0.0321	M4	0.1032
MSE of $\hat{\beta}_1$	4	M5	0.0690	M2	0.2452	M5	0.0219	TM5	0.0786	TM5	0.1058	M4	0.3348	M4	0.0325	M6	0.1047
	5	HH4	0.0702	M3	0.2460	TM5	0.0220	HH5	0.0791	PT	0.1070	M6	0.3402	M6	0.0326	M5	0.1056
	6	M1	0.0715	HH4	0.2467	M6	0.0220	M2	0.0794	M6	0.1089	HH4	0.3536	HH4	0.0331	HH4	0.1103
	7	M2	0.0715	TM3	0.2768	HH4, M4	0.0222	M3	0.0810	M4	0.1095	TM4	0.3537	PT	0.0336	TM5	0.1140
	1	LS	0.0827	HH5	0.3566	M3	0.0205	M4	0.0867	M5	0.1221	HH5	0.5265	M4	0.0291	M4	0.1217
	2	PT	0.0836	HH6	0.3566	M2	0.0206	M5	0.0868	M3	0.1224	HH4	0.5277	M5	0.0291	M1	0.1218
Rank for	3	HH5	0.0853	HH4	0.3599	M4	0.0206	M6	0.0868	PT	0.1234	TM3	0.5387	M6	0.0291	M3	0.1220
MSE of $\hat{\beta}_2$	4	M5	0.0856	TM5	0.3623	M5	0.0206	M2	0.0872	HH5	0.1236	M5	0.5406	HH4	0.0296	M2	0.1221
	5	HH4	0.0868	M5	0.3642	M6	0.0206	M1	0.0873	HH4	0.1237	M4	0.5435	TM4	0.0296	M6	0.1225
	6	M4	0.0871	M4	0.3706	HH4, HH5	0.0211	M3	0.0884	M4	0.1243	TM4	0.5463	HH5	0.0297	M5	0.1226
	7	M6	0.0871	M6	0.3725	TM4, TM5	0.0211	HH	0.0886	M2	0.1250	M6	0.5535	TM5	0.0297	HH4	0.1251
	1	HH5	0.1004	HH5	0.4110	HH5	0.0301	M1	0.1182	HH5	0.1223	HH1	0.5440	HH5	0.0373	M5	0.1544
	2	LS	0.1006	HH6	0.4137	TM5	0.0303	M4	0.1184	HH4	0.1263	HH5	0.5453	TM5	0.0374	HH4	0.1551
Rank for	3	PT	0.1009	HH4	0.4143	HH4	0.0304	M6	0.1187	M5	0.1271	M4	0.5583	HH4	0.0382	TM5	0.1555
MSE of $\hat{\beta}_3$	4	M5	0.1019	M5	0.4143	TM4	0.0306	M1	0.1195	TM5	0.1283	M6	0.5658	M5	0.0382	HH5	0.1557
	5	HH4	0.1036	M2	0.4196	HH6	0.0307	M2	0.1197	PT	0.1289	TM4	0.5844	TM4	0.0289	TM4	0.1560
	6	M4	0.1037	M4	0.4214	M5	0.0308	HH4	0.1201	M4	0.1292	PT	0.5849	M4	0.0391	M4	0.1567
	7	M6	0.1070	M3	0.4257	M6	0.0308	HH5	0.1204	M6	0.1303	M2	0.5889	M6	0.0392	M6	0.1577

Table 2. MSE of estimators of $\underline{\beta}$. PT = Proposal 2 (unweighted). LS = least squares (unweighted). HH, TM, and M are respectively Hildreth-Houck type estimates, Theil-Mennes type estimates, and M-estimates which solve equation (3.4). Numbers after HH, TM, and M indicate the particular algorithm used. Ties are indicated by round brackets. Heteroscedasticity is M (mild) and H (heavy) for $\underline{\alpha}$ = (1.0, 0.2, 0.5) and $\underline{\alpha}$ = (1.0, 3.0, 3.0)', respectively. N is the normal distribution and CN is the contaminated normal distribution.

Sampling Situation		1		2		3		4	
Sample size		25		25		75		75	
Heteroscedasticity		M		H		M		H	
		Estimator	MSE	Estimator	MSE	Estimator	MSE	Estimator	MSE
Rank for MSE of $\hat{\alpha}_1$	1	HH5	0.288	TM3	2.45	HH5	0.128	M2	1.09
	2	TM5	0.339	TM1	2.50	TM5	0.129	M1	1.13
	3	HH1	0.393	M2	2.82	HH4	0.169	M3	1.13
	4	HH3	0.397	TM4	2.98	HH6	0.171	M6	1.27
	5	TM1	0.399	M3	3.01	M5	0.176	M4	1.31
	6	TM3	0.406	TM6	3.02	TM4	0.182	M5	1.64
	7	M5	0.510	M5	3.09	M6	0.182	TM4	1.70
Rank for MSE of $\hat{\alpha}_2$	1	HH4	0.183	M5	4.53	HH4	0.049	M5	1.93
	2	HH5	0.183	M4	4.70	HH5	0.049	M2	1.96
	3	M5	0.192	M6	4.70	HH6	0.050	M6	1.96
	4	M4	0.199	HH4	4.72	TM5	0.053	M1	1.97
	5	M6	0.200	HH5	4.72	TM4	0.054	M4	1.99
	6	M2	0.213	HH6	4.82	TM6	0.055	M3	2.00
	7	M1	0.214	M2	5.06	M4,M5	0.066	TM5	2.30
Rank for MSE of $\hat{\alpha}_3$	1	HH4	0.289	HH4	6.59	HH4	0.093	M5	2.15
	2	HH5	0.289	HH5	6.59	HH5	0.093	M6	2.19
	3	M6	0.311	HH6	7.03	HH6	0.093	M4	2.25
	4	M5	0.336	TM1	8.20	M5	0.096	M1	2.26
	5	M4	0.343	M5	8.90	M4	0.097	M2	2.28
	6	M2	0.360	M4	9.07	M6	0.097	M3	2.37
	7	M1	0.362	M6	9.22	TM5	0.098	TM5	2.77

Table 3. MSE of estimators of $\underline{\alpha}$ at the norm of model. For notation see Table 2.

Sampling Situation	1		2		3		4	
Sample size	25		25		75		75	
Heteroscedasticity	M		M		M		M	
	Estimator	MAE	Estimator	MAE	Estimator	MAE	Estimator	MAE
Rank for MAE of $\log(\alpha_1/\alpha_2)$	TM5	0.634	M4	0.451	TM5	0.383	M4	0.305
	TM4	0.713	M5	0.582	TM1	0.590	M1	0.690
	TM3	0.835	M2	0.906	TM3	0.605	M2	0.697
	TM1	0.827	M1	0.907	TM4	0.735	M3	0.742
	M1	0.881	M3	0.935	M4	0.861	M5	0.801
	M2	0.881	HH5	1.154	M1	0.964	HH5	1.116
	M4	0.898	HH4	1.399	M3	0.965	HH4	1.196
Rank for MAE of $\log(\alpha_1/\alpha_3)$	TM4	1.565	M4	0.263	M5	1.004	M4	0.206
	M4	1.680	M5	0.313	TM4	1.036	M5	0.269
	M2	1.713	M1	0.366	M4	1.066	M2	0.295
	M1	1.727	M2	0.367	TM1	1.078	M1	0.301
	M3	1.754	M3	0.378	TM5	1.150	M3	0.343
	M5	1.877	HH4	1.059	TM3	1.153	TM5	0.844
	HH5	2.057	HH5	1.071	M1	1.583	TM4	0.933
Rank for MAE of $\log(\alpha_2/\alpha_3)$	TM4	1.031	M4	0.367	TM3	0.693	M4	0.275
	TM5	1.058	M5	0.392	TM1	0.706	M5	0.310
	TM1	1.108	M3	0.397	M5	0.718	M2	0.313
	TM3	1.108	M1	0.409	M4	0.783	M3	0.313
	M2	1.174	M2	0.411	TM4	0.990	M1	0.314
	M1	1.200	HH4	1.006	TM5	1.015	TM5	0.769
	M4	1.202	HH5	1.028	M1	1.060	TM4	0.814

Table 4. MAE for estimators of $\log(\alpha_i/\alpha_j)$. For notation see Table 2. MAE is the median absolute error, which is defined in the text.

Sampling situation	Estimator	Coverage Probabilities β_1	β_2	β_3
1 (normal distribution, mild heteroscedasticity)	LS	0.977	0.967	0.930
	PT	0.973	0.967	0.937
	HH4	0.943	0.950	0.953
	HH5	0.953	0.957	0.953
	TM4	0.920	0.953	0.953
	TM5	0.940	0.953	0.960
	M4	0.930	0.947	0.950
	M5	0.940	0.947	0.953
	M6	0.937	0.947	0.950
2 (normal distribution, heavy heteroscedasticity)	LS	0.983	0.943	0.923
	PT	0.983	0.947	0.927
	HH4	0.913	0.933	0.960
	HH5	0.937	0.923	0.963
	TM4	0.903	0.913	0.957
	TM5	0.923	0.917	0.953
	M4	0.920	0.917	0.953
	M5	0.920	0.913	0.950
	M6	0.920	0.913	0.953
5 (contaminated normal, mild heteroscedasticity)	LS	0.977	0.973	0.963
	PT	0.980	0.983	0.963
	HH4	0.950	0.973	0.970
	HH5	0.967	0.977	0.970
	TM4	0.943	0.973	0.973
	TM5	0.950	0.970	0.967
	M4	0.940	0.977	0.973
	M5	0.953	0.973	0.970
	M6	0.940	0.977	0.973
6 (contaminated normal, heavy heteroscedasticity)	LS	0.980	0.950	0.943
	PT	0.980	0.960	0.947
	HH4	0.933	0.933	0.953
	HH5	0.947	0.943	0.960
	TM4	0.910	0.933	0.947
	TM5	0.947	0.940	0.957
	M4	0.930	0.923	0.957
	M5	0.950	0.927	0.960
	M6	0.927	0.923	0.957

Table 5. Coverage probabilities for n = 25 and selected estimators of $\underline{\beta}$. For notation see Table 2. Nominal coverage probabilities, based upon large sample approximations, are 0.95.