AZI

ROBUST REGRESSION IN THE PRESENCE OF HETEROSCEDASTICITY

Shankar Subramanian and Richard T. Carson

This chapter is motivated by our belief that there are two primary characteristics of cross-sectional economic data that affect regression analyses. The first of these is heteroscedastic error terms, a problem that has been well recognized in microlevel data since the path-breaking work of Prais and Houthakker (1955) on family budgets. In aggregate-level cross-sectional data the problem also exists. For example, average wage rates for men and women across different political jurisdictions (conditional on explanatory variables) are likely to have different variances. This difference may be inherent or due to aggregation over different numbers of workers.[1] The second characteristic is the presence of gross errors, or contamination, in the data.[2] Survey researchers have identified a number of typical mistakes that respondents make in answering both factual and attitudinal questions.[3] Mistakes in coding and data entry of survey data are another all too frequent occurrence. The errors in the microlevel data are translated

Advances in Econometrics, Volume 7, pages 85–138.

ISBN: 0-89232-911-4

into more or less severe errors in aggregated cross-sectional data, which also may suffer from differences in variable definitions used by various reporting subunits. To these two characteristics, we should also add that the assumption of a thick-tailed nonnormal error distribution is frequently more appropriate than the standard assumption of normal error terms. An appropriate statistical technique for handling data with these characteristics is robust regression with a correction for heteroscedasticity.

The first of the chapters eight sections introduces notation and briefly reviews the literature on robust regression. The second section introduces the problem of heteroscedasticity and considers the extensions necessary to obtain robust analogues of generalized least squares. The third section reviews the distribution theory for M-estimators for regression and for the trimmed least-squares estimator. Problems with implementing various suggestions for estimating the standard error of coefficients, and testing hypotheses in the presence of heteroscedasticity using these estimators are discussed. The fourth section sets out our ideas on what typical cross-sectional data looks like and discusses the mechanics of the Monte Carlo experiments reported. The fifth section contains a number of Monte Carlo results when there is a heteroscedasticity problem but no gross errors. The sixth section presents Monte Carlo results under the conditions of heteroscedasticity and gross errors. The seventh section presents the results from the Harrison and Rubinfeld (1978) hedonic pricing equation using various weighted and unweighted robust estimators.[4] The final section provides some guidance for applied work and suggests some directions for future research.

I. A REVIEW OF THE ROBUST REGRESSION LITERATURE AND NOTATION

Consider the linear model $y = X\beta + \mu$, where the unknown parameter β is to be estimated. The observables are y and X while u is an unobservable vector of random deviations from $X\beta$.[5] The researcher obtains a random sample of n observations on y and X, and, to greatly simplify our development, X is taken to be a nonstochastic matrix. Because we have taken a linear model and nonstochastic X, regressing y on X using ordinary least-squares (OLS) yields a consistent, although not necessarily efficient, estimate of all but the first element of the β vector, β_1, which we always take to be the parameter

associated with a vector of ones comprising the first column of the X matrix. If $E(u) = 0$, then β_1 is consistently estimated using OLS. If the u_i are iid with mean zero and finite variance, OLS is the best linear unbiased estimator. If an addition, the error terms are Gaussian, then OLS is the best unbiased estimator, and is equivalent to the maximum likelihood estimator.

In the univariate location problem, there is a whole class of maximum likelihood estimators for the center of unimodal symmetric distributions ranging from averaging the minimum and maximum observations if the distribution is the thin-tailed uniform, to taking the mean if the distribution is normal, to taking the median if the distribution is the thick-tailed, double-exponential. Thus, if one believes in the maximum likelihood principle, for the thin-tailed, uniform distribution, it is optimal to discard all of the observations except for the two extreme values. For the normal, a simple average suffices and for the double exponential, the only order statistic necessary is the median.

The M (maximum likelihood type) robust estimators were developed largely for the situation in which the u_i's are independently and identically distributed (iid) with a symmetric distribution with tails at least as thick as the normal.[6] Three types of M-estimators can be identified: (1) those with monotone pseudo-likelihood functions (i.e., some weight is put on all observations no matter how large the residual); (2) soft redescenders (i.e., the weight on an observation goes to zero as the residual goes to infinity); and (3) hard redescenders (i.e., the weight on an observation with a residual larger than some fixed constant is zero). An M-estimator is the solution to

$$\min_{\beta} \sum_{i=1}^{n} \rho[(y_i - x_i\hat{\beta})/s], \tag{1.1}$$

where $\rho(x)$, the pseudo-likelihood function, is convex and usually even and s is an estimate of scale.[7] An alternative and perhaps more operational form of Eq. (1.1) is known as the estimating equation, where β is the solution to

$$\sum_{i=1}^{n} x_i'\psi\left(\frac{y_i - x_i\beta}{s}\right) = 0, \tag{1.2a}$$

or equivalently,

$$\sum_{i=1}^{n} x_i' w_i(y_i - x_i\beta) = 0, \tag{1.2b}$$

where $w_i = \psi(r_i)/r_i$, the r_i, are the standardized residuals $(y_i - x_i\beta)/s$, and ψ, the derivative of ρ, is taken to be odd and continuous except at a finite number of points.

The most popular M-estimator is Huber's classic 1964 "proposal two" for which

$$\psi(r) = \begin{cases} r & \text{if } |r| < K_h \\ -K_h & \text{if } r < -K_h \\ K_h & \text{if } r > K_h. \end{cases} \tag{1.3}$$

In this case, the weights w_i in (1.2b) are unity when the standardized residuals are smaller in magnitude than K_h, and the weights decrease as the standardized residuals increase in magnitude beyond K_h. The tuning constant K_h is usually chosen to achieve some desired level of efficiency for a particular distribution (e.g., the normal) or to provide some desired degree of protection against outliers. The Welsch estimator (Holland and Welsch, 1977), used later in our Monte Carlo experiments, is an example of a soft redescender, and has the ψ function,

$$\psi_w = r\exp[-(r/K_w)^2] \tag{1.4}$$

The two hard redescenders we use, Andrews' sine and Tukey's bisquare, have the ψ functions,[8]

$$\psi_a = \begin{cases} K_a \sin(r/K_a) & \text{if } |r| < K_a \\ 0 & \text{otherwise} \end{cases} \tag{1.5}$$

and

$$\psi_b = \begin{cases} r[1 - (r/K_b)^2]^2 & \text{if } |r| < K_b \\ 0 & \text{otherwise.} \end{cases} \tag{1.6}$$

There are two major drawbacks in practice to the M-estimators. The first is that they are not scale invariant, that is s must be simultaneously estimated with β or obtained in some independent fashion.[9] Joint optimization is difficult and independent estimates of s dubious. The most commonly used robust regression routine (ROSEPACK, Holland and Welsch, 1977; Coleman et al., 1980) uses

$$s = 1.48[\text{med}|(y_i - x_i\beta^*) - \text{med}(y_i - x_i\beta^*)|], \tag{1.7}$$

where β^* is a consistent preliminary estimate of β, usually LAD or OLS, and the factor of 1.48 makes s an unbiased estimate of σ when the distribution is Gaussian. The second problem is finding suitable starting values for β to use in the iteratively reweighted least-squares routine. The poorer the quality of the data (either in terms of deviation from normality or the number and magnitude of gross errors) the less suitable is least squares for the starting value and the greater the need to use the computationally more burdensome LAD.[10]

The LAD estimator is a member of the class of L-estimators, which are formed by taking linear combinations of order statistics. A familiar L-estimator in the location case is the trimmed mean, the average of the central $[(1 - 2\alpha)n]$ order statistics.[11] While L-estimators have the advantage of not requiring estimates of scale, the development of a regression analogue to the trimmed mean has lagged behind that of M-estimators. Bickel (1973) proposed a regression analogue of the trimmed mean, but it is difficult to calculate, not invariant to some reparametrizations, and dependent to some degree on the initial or preliminary estimator, β_0. The estimator does, however, have a number of desirable properties.

Koenker and Basset (1978) introduced regression quantiles, the analogue for the linear model to sample quantiles, and proposed a trimmed least-square estimator based on these quantiles. Their regression quantiles are solutions, $\beta(\theta)$, to the minimization problem:

$$\min_{\beta} \sum_{i=1}^{n} \rho_\theta(y_i - x_i\beta),$$

where $\rho_\theta(x) = x[\theta - I(x < 0)]$. They show that the regression quantiles have an asymptotically normal distribution with mean $\beta + \eta(\theta)e_1$, where η is the inverse quantile function and $e_1 = (1, 0, 0, \ldots)'$. The LAD estimator is of course $\beta(\theta = 0.5)$. Analogues for the linear model to some other univariate L-estimators can be found directly by using the appropriate combinations of $\beta(\theta)$. The interquartile range is $\beta(\theta = 0.75) - \beta(\theta = 0.25)$, and the Gastwirth estimator is $1/4[\beta(\theta = 0.25)] + 1/2[\beta(\theta = 0.5)] + 1/4[\beta(\theta = 0.75)]$. For their TLS estimator, Koenker and Basset suggest dropping those observations whose residuals from $\beta(\alpha)$ are positive or whose residuals from $\beta(1 - \alpha)$ are negative, and forming the OLS estimator from the remaining observations.

Ruppert and Carroll (1980) study Koenker and Basset's trimmed least-squares estimator and introduce an estimator formed by performing OLS after deleting the $[\alpha n]$th largest and $[\alpha n]$th smallest residuals from a preliminary estimator β_0. They examine the performance of several choices of preliminary estimator including β_{OLS}

and $\hat{\beta}_{LAD}$ as well as the average of $\hat{\beta}(\alpha)$ and $\hat{\beta}(1-\alpha)$.[12] LAD gives an inefficient TLS estimator, unless the error distribution is very thick tailed. Surprisingly, the TLS estimator with OLS as preliminary estimator is inefficient even when α is small and the distribution Gaussian. If the error distribution is symmetric, Ruppert and Carroll show that the trimmed least-square estimator obtained when the preliminary estimator is the average of the $[\alpha n]$th and $[(1-\alpha)n]$th regression quantiles is asymptotically equivalent to the TLS estimator of Koenker and Basset. Ruppert and Carroll's procedure has two main advantages over directly deleting observations in the manner suggested by Koenker and Bassett: (1) the number of observations to be deleted is known a priori and is under direct control of the researcher, and (2) the problem, particularly in smaller data sets, of having to decide what to do with observations whose residuals are equal to zero is avoided.[13]

Before moving to the regression case with heteroscedastic error terms, two observations are in order. First, if the symmetric distribution F is known, for every M-estimator there exists an asymptotically equivalent L-estimator (Jaeckel, 1971). Second, it is necessary to make a distinction between robustness and resistance in the regression case, though they are generally equivalent in the location case. In keeping with recent statistical practice (Mosteller and Tukey, 1977; Huber, 1981), we use the term distributional robustness to refer to estimators that are robust against deviations from the normal distribution. We use resistance to refer to estimators that are robust in the sense that they limit the "influence" that one observation or a group of observations can have in determining the estimated coefficients. The reasons for the divergence between distributional robustness and resistance in the regression case have to do with both the concept of leverage (Belsley et al., 1980) in the design matrix and differences in breakdown bounds between the same estimator in the location and the regression problems. Leverage does not occur in the location problem because the X matrix is simply a vector of ones, but is one of the blessings and curses of the linear regression model. Some points in the X matrix represent relatively sparse areas in the design space and exert a strong influence (leverage) over the slope of the regression line particularly when they occur at the extremes. Krasker and Welsch (1982) have proposed a bounded influence estimator with regard to the elements of the design matrix.[14] The difference between an estimator's breakdown point in the location and regression situations can be clearly seen by examining the median and its regression analogue of minimizing

least absolute deviations (Hoaglin et al., 1983). In the location case, the median has a breakdown point of 50%, that is, 50% of the observations can be changed without changing the median. In the regression case, the breakdown bounds of LAD is zero. This is true, because, while one wild observation will move β from its true value slower than OLS, potentially it can still move $\hat{\beta}$ arbitrarily far away from β. Thus, only the hard redescenders that can put zero weight on some observations can have breakdown bounds of greater than zero in the regression case.

II. ROBUST ESTIMATION WITH HETEROSCEDASTIC ERROR TERMS

In the iid case, robust regression techniques essentially put less weight on observations with large residuals than does OLS. When the errors are independent but not identically distributed, the distributional robustness and resistance of these techniques become obscured.

To show this consider one of the standard ways of representing thick-tailed distributions, that of a mixture of normals.[15] The error terms u_i are considered to come from the distribution $N_1(0,\sigma_1^2)$ with probability $1-\epsilon$ and distribution $N_2(0,\sigma_2^2)$ with probability ϵ, where $0.5 > \epsilon$ and $\sigma_2^2 > \sigma_1^2$. The OLS estimator is consistent but becomes increasingly inefficient as ϵ increases and σ_2^2 becomes large relative to σ_1^2.

Now consider the classic case of heteroscedasticity (Bartlett, 1937) where the variance of u_{mj} depends on which of the m groups it came from, and the indicator variable, $\Lambda = [1, 2, \ldots, m]$, is available. The appropriate procedure in this case (assuming normality and that σ_m^2 is known) is to use generalized least squares where the weights for each of the j observations in the mth group is $1/\sigma_m$. It is important to note that the only difference between this case of heteroscedasticity and the mixture of normals is the availability of Λ and the knowledge that it is an indicator variable. Bickel (1976, 1978) has strongly emphasized that inhomogeneity of scale of the error terms that varies in a systematic fashion (e.g., depending on $X\beta$) is quite different from inhomogeneity of scale that varies randomly.[16]

Application of any of the M-estimators to the above example of heteroscedasticity without considering the indicator variable Λ results in the down-weighting of the observations based upon the magnitude of the residuals u_i without consideration of the variance, σ_m^2, of the group that they belong to.

In general this is undesirable, since what the researcher desires is first to weight each observation by the inverse of its group's standard deviation, and then to down-weight the observations with large *weighted* residuals. If the variances of each of the two groups were known, then the M-estimator of β would be found from

$$\min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{y_i - x_i\beta}{\sigma_m}\right). \tag{2.1}$$

Since σ_m is rarely known, consistent estimates of the σ_m may be used to obtain a feasible estimator from

$$\sum_{i=1}^{n} \frac{x_i'}{\hat{\sigma}_m} \psi\left(\frac{y_i - x_i\beta}{\hat{\sigma}_m}\right) = 0. \tag{2.2}$$

The feasible GLS estimator (FGLS) is of course the special case where $\psi(x) = x$.

The problem with the FGLS estimator, as with all the feasible M-estimators, is the need to obtain consistent estimates of the σ_i. As Carroll and Ruppert (1982b) point out, the usual procedure if a consistent estimate of β is at hand is to assume a known functional form for the σ_i,

$$\sigma_i = H(x_i, \beta, \theta), \tag{2.3}$$

and to obtain a consistent estimate of the parameter vector θ. Then consistent estimates of the σ_i are easily found. There are two common specifications for the σ_i. In one σ_i is a function of $x_i\beta$ alone, i.e., $\sigma_i = H(x_i, \beta, \theta)$. An example is the form used by Anscombe (1961) in a test for heteroscedasticity, $\sigma_i = \sigma|1 + x_i\beta|^\theta$. The second form does not constrain σ_i to be a function of β, i.e., $\sigma_i = G(z_i, \theta)$, where the z_i may be a subset or a superset of x_i and functions hereof. Harvey's (1976) specification $\sigma_i = \sigma_0 \exp(z_i\theta)$ is a case in point.

In many cases, a simple but nonrobust estimate of θ may be found by regressing suitably transformed consistent estimates $\hat{u}_i$ (from a preliminary estimator $\hat{\beta}$) on a suitable function of the x_i or $x_i\hat{\beta}$. With Harvey's specification, for example, regressing $\log(\hat{u}_i^2)$ on the z_i gives a consistent estimate of θ. Another approach is to jointly estimate β and θ by maximum likelihood, assuming the errors to be normally distributed. In the case where σ_i is a function of $x_i\beta$ only and the errors are normal, the MLE of β is, as Jobson and Fuller (1980) have found, more efficient than the GLS estimator formed using consistent estimates of the σ_i. Carroll and Ruppert (1982a) warn that Jobson and Fuller's (1980) "feedback" procedure may not be robust to outliers or non-normality of errors, and that it is not robust to misspecification of the functional form of the heteroscedasticity.

Third, the parameter vector may be estimated robustly, as demonstrated by Carroll and Rupper (1982b) for the model:

$$y_i = x_i\beta + \sigma_i u_i \tag{2.4a}$$

$$\sigma_i = H(x_i\beta, \theta). \tag{2.4b}$$

Let

$$G_n(\theta) = \sum_{i=1}^{n} \chi[\hat{u}_i / H(x_i\hat{\beta}, \theta)] \partial \log H(x_i, \hat{\beta}, \theta) / \partial\theta,$$

where $\chi(0) < 0$, $\chi(\infty) > 0$, and $\hat{\beta}$ is an $n^{1/2}$-consistent estimator of β. Under certain regularity conditions they show that a robust estimate of θ may be found by minimizing $\|G_n(\theta)\|$. Since $\|G_n\|$ may not have a unique minimum, they suggest minimizing it on a subset A of the parameter space. The choice of A is dictated by a priori considerations.

There are two reasons why Carroll and Ruppert's method may be unsuitable for use with survey data. It is often the case with survey data that the variances of individual observations are strongly influenced by one or more of the exogenous variables in the model. This suggests that Harvey's specification, i.e., $\sigma_i = \sigma_0 \exp(z_i\theta)$, would meet the case better than the specification employed by Carroll and Ruppert. Harvey's specification is quite flexible because the z_i could be chosen to be any function of the exogenous variables not involving θ or β. Harvey (1976) used this specification in the context of ML estimation. His MLE procedure is not robust, however, it is not difficult to extend Carroll and Ruppert's robust estimation technique to Harvey's model. But as pointed out above, their technique calls for a grid search and a constrained maximization which could be quite time consuming for large sets of data such as those commonly obtained from surveys. In view of these difficulties we have chosen to estimate the parameter θ by robust regression of $\log \hat{u}_i^2$ on the z_i. Robust regression is superior to OLS in this instance because the distribution of the $\log(u_i^2)$ is decidedly nonnormal even when the u_i are normally distributed.

III. ADAPTABILITY OF ROBUST ESTIMATORS UNDER HETEROSCEDASTICITY

Carroll and Ruppert (1982b) have shown for the heteroscedastic linear model (2.4) that M-estimators are adaptable, i.e., the estimator obtained by using $n^{1/2}$-consistent estimates of the heteroscedasticity parameter θ is asymptotically equivalent to that obtained using the true σ_i. In this section we provide similar results without proof, for Harvey's heteroscedasticity specification, for M-estimators, regression quantiles (including the LAD estimator) and Ruppert and Carroll's trimmed LS estimator. Consider the model

$$y_i = x_i\beta + \sigma_i\epsilon_i, \qquad \sigma_i = \sigma_0 \exp(z_i\theta), \tag{3.1}$$

where the ϵ_i are independent, identically distributed random variables with a symmetric distribution function f. Taking up M-estimators first, let $\hat{\beta}^{opt}$ be an M-estimator of β obtained knowing the weights σ_i, i.e., $\hat{\beta}^{opt}$ satisfies

$$\sum_{i=1}^{n} \frac{x_i'}{\sigma_i} \psi[(y_i - x_i\beta)/\sigma_i] = 0. \tag{3.2}$$

Let $\hat{\theta}$ be an $n^{1/2}$-consistent estimator of θ, i.e.,

$$n^{1/2}(\hat{\theta} - \theta) = O_p(1), \tag{3.3}$$

and let $\hat{\sigma}_i = \sigma_0 \exp(z_i\hat{\theta})$. Let $\hat{\beta}$ be the solution to

$$\sum_{i=1}^{n} \frac{x_i'}{\hat{\sigma}_i} \psi[(y_i - x_i\beta)/\hat{\sigma}_i] = 0. \tag{3.4}$$

Under the following assumptions,

B1. ψ odd, $E\psi' > 0$, $0 < E\psi^2 < \infty$,
B2. $\lim_{n\to\infty} \sup_{i\le n} n^{-1/2}\|x_i/\sigma_i\| = 0$,
B3. $\lim_{n\to\infty} \sup_{i\le n} n^{-1/2}\|z_i\| = 0$,
B4. $\sup_n(n^{-1}\sum_{i=1}^{n} \|x_i/\sigma_i\|\,\|z_i\|) < \infty$,
B5. $n^{-1}\sum_{i=1}^{n} x_i'x_i/\sigma_i^2 \to V$, $n^{-1}\sum_{i=1}^{n} z_i'z_i \to W$, V, W positive definite,

and with certain smoothness conditions on ψ given by Carroll and Ruppert (1982; thm. 1), we have

$$n^{1/2}(\hat{\beta} - \hat{\beta}^{opt}) = o_p(1). \tag{3.5}$$

As a result, $n^{1/2}(\hat{\beta} - \beta)$ and $n^{1/2}(\hat{\beta}^{opt} - \beta)$ have the same limiting distribution, a normal distribution with mean zero and variance $[E(\psi^2)/(E\psi')^2]V^{-1}$.

We now show that regression quantiles are adaptable under heteroscedasticity. Let $\tilde{\beta}^{opt}(\alpha)$ be the αth regression quantile found using the known σ_i for model (3.1) and let $\tilde{\beta}(\alpha)$ be the αth regression quantile formed using an $n^{1/2}$-consistent estimator $\hat{\theta}$ of the heteroscedasticity parameter θ. $\tilde{\beta}^{opt}(\alpha)$ is the solution to

$$\min_{\beta} \sum_{i=1}^{n} \rho_\alpha[(y_i - x_i\beta)/\sigma_i] \tag{3.6}$$

and $\tilde{\beta}(\alpha)$ is the solution to

$$\min_{\beta} \sum_{i=1}^{n} \rho_\alpha[(y_i - x_i\beta)/\hat{\sigma}_i]. \tag{3.7}$$

Then, under assumptions B2–B5, we have

$$n^{1/2}[\tilde{\beta}(\alpha) - \tilde{\beta}^{opt}(\alpha)] = o_p(1). \tag{3.8}$$

This result applies to the LAD estimator as well, which is the special case $\alpha = 0.5$.

Lastly, we take up the trimmed least-square estimator. Let $\hat{\beta}_0^{opt}$ be the preliminary estimator found by taking the average of the αth and $(1-\alpha)$th regression quantiles from (3.6), with the σ_i known beforehand. $\beta_{PE}^{opt}(RQ)$ is the trimmed least-square estimator formed after dropping observations that satisfy

$$(y_i - x_i\hat{\beta}_0^{opt})/\sigma_i \le r_{1n} \text{ or } (y_i - x_i\hat{\beta}_0^{opt})/\sigma_i \ge r_{2n}, \tag{3.9}$$

where r_{1n} and r_{2n} are the $[\alpha n]$th and $[(1-\alpha)n]$th ordered weighted residuals from the preliminary estimator β_0^{opt}. Similarly, $\beta_{PE}(RQ)$ is the trimmed LS estimator formed using the $\hat{\sigma}_i = \sigma_0\exp(z_i\hat{\theta})$. Then, under assumptions B2–B5 we find that

$$n^{1/2}[\beta_{PE}(RQ) - \beta_{PE}^{opt}(RQ)] = o_p(1). \tag{3.10}$$

The proofs of these propositions are similar to those in Carroll and Ruppert (1982b) and in Ruppert and Carroll (1980) and are omitted.

In their derivation of the asymptotic distribution of regression quantiles, Basset and Koenker (1978) assume that the design matrix has a column of ones, as do Ruppert and Carroll (1980) in their study of the TLS estimator. But the heteroscedastic linear model is unlikely to have a constant term after each observation has been weighted to eliminate heteroscedasticity. While the absence of a constant term does not affect the results on adaptability presented above, it substantially alters the asymptotic distributions of regression quantiles and the TLS estimator (Subramanian and Carson; 1987). In particular, if the constant is absent, the asymptotic variance of the TLS estimator differs from its value when the model has an intercept by an indefinite matrix that is difficult to estimate. But if an intercept is added to the transformed model, only to be dropped once the preliminary estimator [the average of the αth and $(1-\alpha)$th regression quantiles] has been found, this modified TLS estimator has the same asymptotic distribution as when there is an intercept. This modified estimator is used throughout this study and its asymptotic variance is $\sigma^2(\alpha,f)V^{-1}$ where $\sigma^2(\alpha,f)$, the asymptotic variance for the trimmed mean in the location model, is given by

$$\sigma^2(\alpha,f) = (1-2\alpha)^{-2}\left[\int_{-\zeta}^{\zeta} x^2 f(x)\,dx + 2\alpha\zeta^2\right],$$

where $\zeta = \eta(1-\alpha)$ is the $(1-\alpha)$th inverse quantile.

A. Standard Errors and Hypothesis Testing

Our results on the adaptability of robust estimators for the linear model with heteroscedasticity show that, using $n^{1/2}$-consistent estimates of the heteroscedasticity parameter θ, we are able to obtain estimators that are asymptotically equivalent, under certain regularity conditions, to the "optimal" robust estimators found using the actual σ_i, i.e., to those found by transforming the model to eliminate heteroscedasticity and then applying a robust estimation procedure. As a result, the asymptotic covariance matrix for a "feasible" robust estimator that uses the consistently estimated $\hat{\sigma}_i$ is identical to that for the corresponding transformed homoscedastic model. In principle, then, a consistent estimate for the asymptotic covariance matrix may be obtained by taking the expression for the asymptotic covariance matrix of the transformed homoscedastic model and replacing all unknown parameters by their consistent estimates.

For the M-estimators we use an expression suggested by Huber (1981, p. 173) for the covariance matrix, which is unbiased to order (p/n) in the case of a balanced design matrix,[17] where p is the number of variables and n the number of observations. Strictly speaking, Huber's result is valid only for a balanced design matrix and for ψ twice differentiable and bounded. His result is

$$K^{-1}(n-p)^{-1}\sum_i \psi(r_i)^2 T^{-1}RT^{-1}. \qquad (3.11)$$

K is a correction factor given by

$$K = 1 + \frac{p}{n}\frac{V(\psi')}{(E\psi')^2}. \qquad (3.12)$$

Here r_i is given by

$$r_i = (y_i - x_i\hat{\beta})/\hat{\sigma}_i, \qquad (3.13)$$

and the matrices T and R are given by

$$T_{jk} = \sum_i \psi'(r_i)x_{ij}x_{jk}/\hat{\sigma}_i^2 \quad \text{and } R = \sum_i x_i'x_i/\hat{\sigma}_i^2. \qquad (3.14)$$

The terms $V(\psi')$ and $E\psi'$ in the expression for K are estimated by the sample variance and sample mean, respectively, of the $\psi'(r_i)$. We have not been able to establish whether the term $(n-p)$ in the above expressions should be replaced with $(n-p-l)$ where l is the number of heteroscedasticity parameters. This has been suggested by Carroll and Ruppert (1982b), among others. However, with large samples usually available with surveys, for example, this correction would be unimportant.

We use the following estimate of the covariance matrix for the LAD estimator

$$[2f(0)]^{-2}R^{-1}, \qquad (3.15)$$

where $f(0)^{-1}$ is estimated by differentiating a smoothed version of the empirical quantile function (Parzen, 1979). The smoothing was performed using Friedman and Stuetzle's (1982) supersmooth algorithm. This was found to be quicker than the method suggested by Koenker and Bassett (1982). Their method involves finding of

$$\hat{R}(\alpha) = \min_\beta \sum_i \rho_\alpha[y_i - x_i\beta(\alpha)] \qquad (3.16)$$

for several values of α, and twice differentiating a smoothed version of $\hat{R}(\alpha)$.

For the trimmed least-squares estimator, we estimate $\sigma^2(\alpha,f)$ along the lines suggested by Ruppert and Carroll (1980) with the modification suggested by Subramanian and Carson (1987) when the transformed model does not have an intercept,

$$\hat{\sigma}^2(\alpha,f) = (1-2\alpha)^{-2}[S/(n-p) + \alpha(1-\alpha)(c_1^2 + c_2^2) - 2\alpha^2 c_1 c_2],$$

where S is the sum of squared weighted residuals for the observations that were not dropped, from the preliminary estimator and c_1 and c_2 are the $[\alpha n]$th and $[(1-\alpha)n]$th ordered weighted residuals from the TLS estimator. When heteroscedasticity is assumed to be absent, we follow Ruppert and Carroll (1980) and take c_1 and c_2 to be $\beta(\alpha)_1 - \hat{\beta}_1$ and $\beta(1-\alpha)_1 - \hat{\beta}_1$, where $\hat{\beta}$ is the TLS estimator.

IV. EXPERIMENTAL DESIGN

The difficulty in designing Monte Carlo experiments lies in depicting those aspects of actual data sets that are relevant to a wide class of applied problems and in choosing what aspects of the data sets to vary across treatments. Failure on either count renders a Monte Carlo exercise close to useless as far as guidance for applied work is concerned.

The first choice we made, and the one we feel the strongest about, is sample size. Cross-sectional analysis, almost by definition, implies a sample of moderate to large size.[18] We have recently estimated regression equations using 50 states, 73 developing countries, 92 consumers, 117 farmers, 210 utilities, 506 census tracts, 564 household heads, 1016 voters, and 3012 counties. The fifty states represent the lower end of the range of samples sizes we wish to consider and the 3012 counties are somewhat outside the upper end of the range we wish to consider. The sample range we are interested in has been described by Tukey as the size where the researcher should be willing to accept small increases in variance in order to reduce the bias. Researchers, particularly those who deal with survey data, have been quite willing to delete a few observations if the responses seem unreasonable but would be very reluctant to drop 25 to 50% of their sample.

Since researchers typically increase the number of parameters as the sample size increases it is the ratio of p to n which matters.[19] We consider p/n ratios of 0.12 to 0.012 concentrating on the case of 100 observations with six parameters including those influencing σ_i.

The second choice made was the form of heteroscedasticity. Here we chose the multiplicative form, $\sigma_i^2 = \exp(Z\theta)$, of Harvey (1976) because (1) it is one of the more popular functional forms estimated in applied work and (2) it contains as special cases, or can closely approximate, most other functional forms including those where the variance is a function of y_i. The equation we estimate in all cases is

$$y_i = 2 + 2x_{i2} + 2x_{i3} + u_i, \qquad (4.1)$$

where

$$u_i = v_i[\exp(0.02x_{i2} + 0.02x_{i3})]^{-1/2}. \qquad (4.2)$$

There are many possible distributions for v. We were tempted here to repeat in its entirety the famous quote about how mathematicians believe in normality because they think it is an empirical fact while applied workers believe in normality because it is a mathematical fact, but this quote from Hotelling (1961, p. 319) seems more apt:

> Practical statisticians have tended to disregard nonnormality, partly for lack of an adequate body of mathematical theory to which an appeal can be made, partly because they think it is too much trouble, and partly because of a hazy tradition that all mathematical ills arising from nonnormality will be cured by sufficiently large numbers. This last idea presumably stems from central limit theorems, or rumors or inaccurate recollections of them.

We believe the presence of heteroscedasticity suggests that v is unlikely to be distributed normally with a constant variance. This is because heteroscedasticity implies that one is no longer dealing with the pure measurement error (using a single instrument) case.

Some of the early work on robust estimators emphasized the stable family of distributions, which included the normal and the very thick-tailed Cauchy. The Cauchy is now considered by most statisticians as an unrealistic situation especially for error terms from a regression equation. Attention has now turned to the Box–Tiao (1973) family of exponential power distributions, whose density with mean μ and shape parameter β is

$$f(x; u; \beta) = [\Gamma((3+\beta)/2)2^{(3+\beta)/2}]^{-1} \exp(-1/2|x-\mu|^{2/(1+\beta)}, \qquad (4.3)$$

where x is finite, and $-1 \leqslant \beta \leqslant 1$, and to the Student t-distribution with varying degrees of freedom.[20] Kurtosis seems to be the primary deciding factor within a family of distributions on the best estimator. We use three members of the exponential power family, the normal, the double exponential, and a distribution with a level of kurtosis halfway in between those two distributions, which we call ZAP after our computer acronym.[21] We also use the contaminated normal, which is an attractive distribution for v because it suggests that different groups (with the grouping variable unobserved) come from distributions with different variances.[22] Heteroscedasticity in this case suggests that, within a group, the variances of observations differs according to measured characteristics.

All of the v_i were drawn so that they had a mean of zero and a standard deviation of 25. This gave our equation an R^2 of approximately 0.4 which is typical in cross section studies.

Belsley et al. (1980) have shown the importance of leverage in the design matrix in a wide variety of situations. Techniques that do not work well in the face of nonnormality or gross errors with low leveraged X matrix tend to get worse when the leverage increases. In particular problems with parameter and standard error calculations appear to shift from the intercept to the slope terms in the presence of high leverage points. In this study, we have created the X matrix by generating the second two columns of the design matrix (the first is a column of 1's) from a uniform distribution on the interval 0 to 100. This will create an X matrix with the minimum possible leverage. Thus our Monte Carlo results are based on a design matrix likely to be more sympathetic to OLS and FGLS than found in actual empirical situations. If we relax this restriction, it is possible to create situations where OLS and GLS give arbitrarily bad results in the face of gross errors.[23] The penultimate section presents estimations using the Boston housing data that are well known for their high leverage data points.

The question of how to generate gross errors also arises. We believe the frequently used method of generating the bad $y_i's$ from a distribution that has the same mean, $X\beta$, but a much larger variance than the correct $y_i's$ to be a poor characterization of the gross errors likely to occur.[24] Keypunch and coding errors are likely to have a uniform distribution having no relationship to $X\beta$. Errors in answering survey questions are much harder to characterize as random draws from some distribution, however, what is clear is that they are not likely to have mean zero. We have chosen to generate random errors as random draws from a uniform distribution on the interval 0 to 1000.[25] Thus the gross errors have an expected value of 500 (while the mean of the true $y_i's$ is 202) and bear no necessary relationship to their corresponding x values.

The regression techniques we examine are OLS, generalized least-squares using estimated weights (FGLS), the LAD and weighted LAD estimator, the weighted and unweighted forms of the HUBER, the WELSCH, the BISQUARE, and the ANDREWS, and the Ruppert and Carroll form of the trimmed least-squares estimator (TRIM) with $\alpha = 0.1$ and 0.2, with and without weights. A "W" in front of the technique, such as WHUBER, denotes the weighted form of the robust estimator. In each case, the mean square error of these estimators is compared with those of GLS with known weights and the bias, standard error, and the number of times out of 200 the true value of each parameter was rejected at the 5% level.[26] In all cases for the M-estimators we used LAD starting values and the robust estimator of scale given by Eq. (1.7). We used the now standard convention of setting K so that 95% efficiency results if the true error terms are iid Gaussian.[27] The weights for the FGLS estimator are obtained by finding the OLS coefficients from the OLS regression of the log of the squared first stage OLS residuals on the X matrix. The weights for the other M-estimators are found in a similar fashion by substituting that particular technique for OLS and GLS. The weighted trimmed least-squares estimators use the weights estimated by trimmed least squares on the residuals from the corresponding unweighted trimmed least-squares equation.

Most of the Monte Carlo experiments were performed with the Bell Lab's statistical language S. The estimation routine used for the M-estimators was the S implementation of ROSEPACK (Coleman et al., 1980).[28] The trimmed least-squares regressions used the modifications of the Barrodale and Roberts (1973) l_1 algorithm suggested by Fulton et al. (1987) and Koenker and D'Ofrey (1987). The method of implementing the standard errors for each of the robust regression procedures is described above. The method of calculating the standard errors for the LAD regressions described above needs a smooth. Friedman and Stuetzle's (1982) supersmooth, which we elected to use, worked well most of the time, particularly if its penalty function was used to keep the smallest smoothing window from getting too much weight. Different penalty functions were optimal with different distributions. We finally settled on use a penalty function of 9 (which greatly reduced the weight put on the small window unless the data were very well behaved since in practice the distribution of the data would be unknown).

Two hundred repetitions of each experiment were run. The experiments can be placed into two major groups: (1) those without gross errors that are to be found in Section V, and (2) those with gross errors that are to be found in Section VI.

V. MONTE CARLO RESULTS (NO GROSS ERRORS)

The Monte Carlo results in this section are for the model

$$y_i = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u_i,$$

where $\beta_1 = 2$, $\beta_2 = 2$, $\beta_3 = 2$. The sample size is 100 in every case. Except for the first table (Table 1), u_i is a heteroscedastic error term defined according to Eq. (4.2). Tables are given for different underlying distributions for the random component of u_i. The tables report the ratio of the mean square error of each estimator to that of GLS with known weights (MRATIO), the bias (BIAS), the standard error (STDERR), and the number of times a technique rejected the

Table 1. Gaussian Errors with No Heteroscedasticity

	MRATIO X1 (BIAS)	MRATIO X2 (BIAS)	MRATIO X3 (BIAS)	STDERR X1 (#REJECT)	STDERR X2 (#REJECT)	STDERR X3 (#REJECT)
OLS	1.0000 (0.0878)	1.0000 (−0.0034)	1.0000 (−0.0031)	6.9767 (11)	0.0916 (13)	0.0845 (8)
FGLS	1.0666 (0.2416)	1.0852 (−0.0065)	1.1716 (−0.0034)	6.9862 (10)	0.0909 (20)	0.0839 (10)
LAD	1.6768 (0.3606)	1.2434 (−0.0088)	1.9276 (0.0028)	7.9265 (23)	0.1041 (21)	0.0960 (21)
WLAD	1.7702 (0.3977)	1.2773 (−0.0094)	2.0471 (0.0028)	7.8837 (23)	0.1039 (20)	0.0956 (25)
HUBER	1.0478 (−0.0230)	1.0350 (−0.0035)	1.0910 (−0.0021)	7.0783 (10)	0.0928 (18)	0.0857 (8)
WHUBER	1.0754 (−0.0296)	1.0708 (−0.0046)	1.1887 (−0.0010)	7.0002 (12)	0.0910 (20)	0.0846 (11)
WELSCH	1.0583 (0.0690)	1.0335 (−0.0056)	1.0998 (−0.0026)	7.1099 (11)	0.0933 (17)	0.0863 (8)
WWELSCH	1.0859 (0.0653)	1.0743 (−0.0067)	1.1921 (−0.0016)	7.0155 (15)	0.0915 (19)	0.0849 (10)
BISQUARE	1.0590 (0.0742)	1.0344 (−0.0055)	1.1072 (−0.0026)	7.1269 (11)	0.0935 (18)	0.0864 (7)
WBISQUARE	1.0854 (0.0804)	1.0759 (−0.0067)	1.1918 (−0.0018)	7.0336 (14)	0.0918 (19)	0.0851 (9)
ANDREWS	1.0542 (0.0289)	1.0345 (−0.0048)	1.1075 (−0.0023)	7.1306 (12)	0.0935 (16)	0.0865 (8)
WANDREWS	1.0855 (0.0810)	1.0761 (−0.0067)	1.1908 (−0.0018)	7.0370 (14)	0.0918 (19)	0.0852 (9)
TRIM(.1)	1.2055 (0.1204)	1.2811 (−0.0076)	1.2263 (−0.0013)	7.4565 (13)	0.0976 (21)	0.0902 (11)
WTRIM(.1)	1.1734 (0.0869)	1.1993 (−0.0062)	1.2125 (−0.0022)	7.1774 (15)	0.0947 (11)	0.0868 (15)
TRIM(.2)	1.2608 (−0.1363)	1.1467 (−0.0072)	1.3418 (0.0020)	7.6613 (20)	0.1002 (14)	0.0930 (15)
WTRIM(.2)	1.1721 (−0.1366)	1.1587 (−0.0052)	1.2938 (−0.0005)	7.4703 (17)	0.0979 (16)	0.0900 (11)
MSE OLS	47.5698	0.0103	0.0066			

parameter value of 2 at the 5% level.[29] These tables can be characterized by the error distribution used, sample size, and the presence or absence of heteroscedasticity:

Table number	*Error distribution*	*Sample size*	*Heteroscedasticity*
1	Gaussian	100	No heteroscedasticity
2	Gaussian	100	Heteroscedasticity
3	ZAP	100	Heteroscedasticity
4	Double exponential	100	Heteroscedasticity
5	Contaminated normal	100	Heteroscedasticity

Table 1 shows results from the classic reference case of normality with no heteroscedasticity, no gross errors, and no leverage points in the X matrix. An interesting feature here is that we went ahead and corrected for the heteroscedasticity suggested by the Harvey equations in the weighted version of each estimator even though the heteroscedasticity parameter estimates in almost all cases were insignificant. The cost of doing this is about a 5% decrease in efficiency when compared to the unweighted form of each estimator. Only the LAD and WLAD estimators are grossly inefficient relative to OLS. The efficiency of the other robust estimators range from the Huber which is about 5% less efficient to the WTRIM (0.2) which is about 20% less efficient. It is important to note that none of the estimators is biased and the number of true parameter values rejected suggests that 5% level t-tests should be interpreted as lying somewhere between 5 and 10%.

The experimental designs for Tables 2–5 are identical except for the increase in the kurtosis of the error terms as one goes from the normal to contaminated normal case. We were surprised at how inefficient OLS was relative to any of the weighted estimators. Dramatic efficiency gains such as this are rarely emphasized in econometrics classes. We took fairly reasonable values for the heteroscedasticity parameters and found OLS to always be less than 50% as efficient as GLS. In contrast, FGLS does not give up much relative to GLS. However, even at the mildly thick-tailed ZAP distribution WHUB is 5–10% more efficient than GLS. At the thicker tailed double exponential and contaminated normal the weighted robust estimators are 20–30% more efficient than GLS and 40% more efficient than FGLS. The results of these tables suggest that the robust estimators are much less sensitive than OLS to heteroscedasticity. A comparison between the weighted and unweighted versions of the robust estimators shows substantial gains from using the weighted forms but nothing like the 100% plus efficiency gain that can be obtained by switching from OLS to FGLS. There is again no appreciable bias in the parameter estimates from any of the estimators although the same statement cannot be made about the estimates of the standard errors. The unweighted LAD and TRIM estimators tend to reject the true parameter value much too frequently. The weighted TRIM's and the weighted and unweighted M-estimators tend to be somewhat conservative rejecting the true parameter value a bit more often than they should. The least-squares estimators are somewhat erratic as far as their rejection pattern.

Moderate changes in simple size do not really change these arguments. In experiments not reported here with sample sizes of 50 and 500 the relative performance of the estimators appears quite similar.[30]

VI. MONTE CARLO RESULTS (GROSS ERRORS)

The tables presented in this section all contain 10% gross errors that were generated by replacing the last 10% of the y values with a draw from a uniform distribution over the interval 0 to 1000. These tables can be characterized by the error distribution used and the presence or absence of heteroscedasticity:

Table number	*Error distribution*	*Sample size*	*Heteroscedasticity*
6	Gaussian	100	No heteroscedasticity
7	Gaussian	100	Heteroscedasticity
8	ZAP	100	Heteroscedasticity
9	Double exponential	100	Heteroscedasticity
10	Contaminated normal	100	Heteroscedasticity

The introduction of gross errors causes a marked deterioration in the performance of the OLS and FGLS estimators. Table 6 (Gaussian errors and no heteroscedasticity), OLS and FGLS are over 1000% less efficient than GLS on the good observations and there is a significant amount of bias in both cases in the intercept parameter.[31] The LAD

(*Text continues on page 124*)

Table 2. Gaussian Errors

	MRATIO X1 (*BIAS*)	*MRATIO X2* (*BIAS*)	*MRATIO X3* (*BIAS*)	*STDERR X1* (*# REJECT*)	*STDERR X2* (*# REJECT*)	*STDERR X3* (*# REJECT*)
GLS	1.0000 (−0.0206)	1.0000 (−0.0071)	1.0000 (−0.0022)	14.2648 (12)	0.2606 (9)	0.2171 (9)
OLS	2.3473 (0.7677)	1.7611 (−0.0124)	1.6802 (−0.0132)	23.8754 (6)	0.3135 (14)	0.2892 (5)
FGLS	1.0485 (−0.1955)	1.0555 (−0.0061)	1.0005 (0.0020)	14.4507 (11)	0.2581 (14)	0.2181 (7)
LAD	1.5882 (−0.3694)	1.6165 (−0.0052)	1.5444 (0.0060)	22.0283 (8)	0.2892 (22)	0.2668 (20)
WLAD	1.4500 (0.4594)	1.5562 (−0.0229)	1.3867 (0.0047)	16.8276 (15)	0.3014 (13)	0.2554 (20)
HUBER	1.6853 (0.5634)	1.4123 (−0.0143)	1.4147 (−0.0096)	20.5663 (4)	0.3013 (12)	0.2619 (7)
WHUBER	1.0268 (0.4143)	1.0361 (−0.0187)	1.0400 (−0.0029)	14.1476 (13)	0.2587 (13)	0.2173 (5)
WELSCH	1.6922 (0.6749)	1.4256 (−0.0149)	1.4661 (−0.0120)	20.8769 (5)	0.3109 (11)	0.2662 (5)
WWELSCH	1.0385 (0.4863)	1.0459 (−0.0193)	1.0404 (−0.0026)	14.4139 (11)	0.2612 (13)	0.2192 (6)
BISQUARE	1.7631 (0.6770)	1.4902 (−0.0147)	1.5326 (−0.0124)	21.3398 (5)	0.3178 (13)	0.2706 (6)
WBISQUARE	1.0431 (0.5418)	1.0590 (−0.0195)	1.0425 (−0.0036)	14.4546 (12)	0.2624 (12)	0.2197 (5)
ANDREWS	1.7675 (0.6842)	1.4970 (−0.0158)	1.5294 (−0.0119)	21.4413 (5)	0.3199 (12)	0.2717 (6)
WANDREWS	1.0460 (0.4923)	1.0619 (−0.0192)	1.0382 (−0.0030)	14.4516 (12)	0.2625 (13)	0.2197 (5)
TRIM(.1)	1.6276 (1.2091)	1.5162 (−0.0082)	1.8712 (−0.0192)	17.1140 (15)	0.2370 (23)	0.2124 (31)
WTRIM(.1)	1.3267 (−0.1160)	1.1784 (−0.0059)	1.5719 (0.0143)	15.1255 (11)	0.2684 (17)	0.2267 (10)
TRIM(.2)	1.6616 (2.2950)	1.5212 (−0.0086)	2.0288 (−0.0382)	15.2735 (20)	0.2227 (31)	0.1961 (37)
WTRIM(.2)	1.2867 (−0.2206)	1.2399 (0.0015)	1.3402 (0.0082)	15.3676 (9)	0.2768 (13)	0.2344 (13)
MSE GLS	205.1405	0.0674	0.0432			

Table 3. Zap Errors

	MRATIO X1 (BIAS)	*MRATIO X2 (BIAS)*	*MRATIO X3 (BIAS)*	*STDERR X1 (#REJECT)*	*STDERR X2 (REJECT)*	*STDERR X3 (#REJECT)*
GLS	1.0000 (1.2004)	1.0000 (−0.0215)	1.0000 (−0.0230)	14.2712 (12)	0.2607 (8)	0.2172 (13)
OLS	2.2603 (3.8254)	1.6460 (−0.0544)	1.7904 (−0.0476)	23.9411 (6)	0.3143 (11)	0.2900 (13)
FGLS	1.0194 (1.5753)	0.9915 (−0.0240)	1.0775 (−0.0298)	14.4038 (10)	0.2586 (10)	0.2169 (13)
LAD	1.4394 (1.9109)	1.3714 (−0.0242)	1.1209 (−0.0282)	16.9586 (12)	0.2226 (27)	0.2054 (21)
WLAD	1.1361 (0.9081)	1.2418 (−0.0094)	0.9432 (−0.0168)	13.2278 (23)	0.2370 (17)	0.2000 (26)
HUBER	1.5717 (2.6228)	1.1928 (−0.0340)	1.2510 (−0.0327)	18.5840 (11)	0.2738 (7)	0.2388 (12)
WHUBER	0.9280 (1.1901)	0.9064 (−0.0180)	0.9511 (−0.0200)	13.2752 (13)	0.2422 (9)	0.2005 (14)
WELSCH	1.7296 (0.6132)	1.2395 (0.0001)	1.4147 (−0.0127)	18.8599 (10)	0.2841 (12)	0.2436 (16)
WWELSCH	1.0904 (0.4950)	1.0300 (−0.0101)	0.9951 (−0.0021)	13.3748 (21)	0.2421 (16)	0.2032 (14)
BISQUARE	1.6919 (2.3171)	1.2920 (−0.0325)	1.3426 (−0.0243)	18.9518 (9)	0.2871 (8)	0.2441 (18)
WBISQUARE	1.0902 (1.2262)	0.9580 (−0.0284)	1.0694 (0.0145)	13.5028 (16)	0.2428 (10)	0.2018 (20)
ANDREWS	1.7223 (2.1433)	1.3141 (−0.0293)	1.3456 (−0.0246)	18.9034 (9)	0.2865 (8)	0.2438 (19)
WANDREWS	0.9998 (1.3114)	0.9981 (−0.0237)	0.9655 (−0.0170)	13.3448 (14)	0.2432 (11)	0.2037 (13)
TRIM(.1)	1.3465 (−0.1637)	1.1265 (−0.0044)	1.1503 (0.0106)	15.8899 (9)	0.2187 (21)	0.1975 (26)
WTRIM(.1)	1.1291 (−0.8614)	1.1055 (0.0188)	0.9968 (0.0150)	13.8099 (10)	0.2491 (12)	0.2098 (7)
TRIM(.2)	1.0830 (−0.2554)	1.0069 (−0.0167)	0.8946 (0.0193)	13.9421 (12)	0.2008 (26)	0.1783 (21)
WTRIM(.2)	0.9792 (−0.7898)	1.0375 (0.0082)	0.9194 (0.0250)	13.4408 (13)	0.2428 (18)	0.2044 (13)
MSE GLS	201.9332	0.0664	0.0521			

Table 4. Double Exponential Errors

	MRATIO X1 (*BIAS*)	*MRATIO X2* (*BIAS*)	*MRATIO X3* (*BIAS*)	*STDERR X1* (# *REJECT*)	*STDERR X2* (# *REJECT*)	*STDERR X3* (# *REJECT*)
GLS	1.0000 (0.2748)	1.0000 (−0.0260)	1.0000 (0.0047)	14.3187 (8)	0.2616 (6)	0.2180 (9)
OLS	2.3378 (0.7655)	1.6018 (−0.0198)	1.6866 (−0.0117)	23.6969 (3)	0.3111 (14)	0.2870 (6)
FGLS	1.1066 (0.4227)	1.0128 (−0.0363)	1.0843 (0.0091)	15.1111 (5)	0.2640 (5)	0.2199 (12)
LAD	1.0055 (0.1710)	0.9736 (−0.0208)	0.6168 (0.0052)	13.5863 (11)	0.1784 (24)	0.1646 (16)
WLAD	0.8378 (0.4284)	0.8008 (−0.0207)	0.6435 (0.0033)	10.7463 (18)	0.1913 (21)	0.1611 (20)
HUBER	1.2237 (0.4965)	0.9119 (−0.0173)	1.0086 (−0.0072)	16.2858 (9)	0.2406 (11)	0.2078 (12)
WHUBER	0.7774 (0.2060)	0.7431 (−0.0239)	0.7925 (0.0036)	12.1923 (9)	0.2175 (12)	0.1807 (15)
WELSCH	1.2635 (0.6758)	0.9374 (−0.0224)	0.9903 (−0.0074)	15.9602 (9)	0.2425 (12)	0.2057 (15)
WWELSCH	0.7727 (−0.0291)	0.7458 (−0.0214)	0.7783 (0.0074)	11.9206 (11)	0.2144 (12)	0.1776 (17)
BISQUARE	1.3211 (0.6666)	0.9909 (−0.0223)	1.0197 (−0.0078)	16.1847 (9)	0.2468 (13)	0.2088 (15)
WBISQUARE	0.8270 (0.1737)	0.7843 (−0.0228)	0.8155 (0.0029)	12.0756 (13)	0.2174 (12)	0.1797 (17)
ANDREWS	1.3249 (0.6657)	0.9958 (−0.0223)	1.0211 (−0.0078)	16.1822 (9)	0.2468 (13)	0.2091 (15)
WANDREWS	0.8312 (0.1663)	0.7857 (−0.0227)	0.8183 (0.0029)	12.0873 (13)	0.2175 (12)	0.1800 (17)
TRIM(.1)	1.1009 (−0.5078)	0.8252 (0.0091)	0.8958 (0.0088)	13.8650 (10)	0.1909 (21)	0.1715 (24)
WTRIM(.1)	0.8434 (−0.0902)	0.8014 (0.0174)	0.7948 (0.0001)	12.9495 (9)	0.2306 (15)	0.1948 (13)
TRIM(.2)	0.8720 (−0.5203)	0.7225 (0.0166)	0.7681 (0.0026)	11.1964 (17)	0.1602 (36)	0.1420 (33)
WTRIM(.2)	0.7231 (0.5803)	0.6810 (0.0048)	0.7108 (−0.0015)	12.3295 (11)	0.2188 (15)	0.1833 (11)
MSE GLS	187.9246	0.0661	0.0506			

Table 5. Contaminated Normal Errors

	MRATIO X1 (*BIAS*)	*MRATIO X2* (*BIAS*)	*MRATIO X3* (*BIAS*)	*STDERR X1* (# *REJECT*)	*STDERR X2* (# *REJECT*)	*STDERR X3* (# *REJECT*)
GLS	1.0000 (−0.0291)	1.0000 (0.0011)	1.0000 (0.0001)	14.5534 (12)	0.2659 (3)	0.2215 (12)
OLS	2.8615 (−0.4699)	2.0580 (0.0035)	1.9321 (0.0070)	24.3326 (8)	0.3195 (16)	0.2947 (14)
FGLS	1.1110 (−0.1686)	1.0602 (0.0108)	1.0725 (−0.0059)	14.8567 (15)	0.2631 (6)	0.2216 (15)
LAD	1.4263 (−0.2892)	1.3532 (0.0195)	1.2896 (−0.0092)	16.8370 (13)	0.2211 (31)	0.2039 (29)
WLAD	0.7080 (0.4355)	0.8524 (−0.0232)	0.5405 (0.0039)	10.2326 (26)	0.1845 (26)	0.1556 (22)
HUBER	1.4208 (0.0130)	1.1814 (0.0167)	1.0768 (−0.0101)	17.3874 (11)	0.2544 (14)	0.2210 (15)
WHUBER	0.7264 (0.9271)	0.7378 (0.0053)	0.7395 (−0.0201)	12.1020 (17)	0.2209 (11)	0.1850 (15)
WELSCH	1.2925 (−0.0181)	1.1162 (0.0214)	1.0220 (−0.0137)	17.2151 (12)	0.2571 (15)	0.2204 (14)
WWELSCH	0.6962 (1.3240)	0.7101 (0.0039)	0.7058 (−0.0255)	11.9232 (18)	0.2813 (12)	0.1829 (14)
BISQUARE	1.3009 (−0.0928)	1.1324 (0.0242)	1.0392 (−0.0139)	17.5891 (11)	0.2620 (15)	0.2239 (15)
WBISQUARE	0.6935 (1.3488)	0.7123 (0.0043)	0.6973 (−0.0258)	11.9458 (15)	0.2183 (12)	0.1831 (14)
ANDREWS	1.3034 (−0.0961)	1.1328 (0.0243)	1.0428 (−0.0139)	17.6408 (11)	0.2629 (14)	0.2244 (15)
WANDREWS	0.6929 (1.3592)	0.7130 (0.0042)	0.6952 (−0.0257)	11.9406 (16)	0.2181 (13)	0.1832 (14)
TRIM(.1)	1.0886 (1.4831)	1.0549 (0.0036)	0.9315 (−0.0276)	13.0436 (24)	0.1800 (43)	0.1612 (33)
WTRIM(.1)	0.8129 (1.2891)	0.8599 (0.0032)	0.7837 (−0.0270)	12.5578 (19)	0.2282 (16)	0.1915 (17)
TRIM(.2)	1.0237 (1.2054)	1.0234 (−0.0051)	0.8438 (−0.0167)	9.8937 (41)	0.1440 (62)	0.1266 (50)
WTRIM(.2)	0.8676 (0.9720)	0.9191 (0.0039)	0.8321 (−0.0202)	12.5487 (22)	0.2287 (14)	0.1921 (19)
MSE GLS	232.4865	0.0656	0.0582			

Table 6. Gaussian Errors with Gross Errors

	MRATIO X1 (*BIAS*)	*MRATIO X2* (*BIAS*)	*MRATIO X3* (*BIAS*)	*STDERR X1* (# *REJECT*)	*STDERR X2* (# *REJECT*)	*STDERR X3* (# *REJECT*)
OLS[1:90,]	1.0000	1.0000	1.0000	7.4782	0.0961	0.0906
	(0.0137)	(0.0015)	(0.0024)	(11)	(12)	(17)
OLS	23.7827	10.2662	13.4302	33.9565	0.4506	0.4160
	(20.8789)	(0.0743)	(0.0021)	(13)	(2)	(5)
FGLS	22.3339	13.4484	10.5225	30.2811	0.4256	0.3911
	(29.1452)	(0.0582)	(−0.1490)	(9)	(6)	(6)
LAD	1.6524	1.6530	1.3641	8.5156	0.1118	0.1031
	(1.4670)	(0.0092)	(−0.0011)	(24)	(23)	(15)
WLAD	1.6002	1.5790	1.4146	8.2036	0.1137	0.1028
	(0.3976)	(−0.0094)	(0.0028)	(23)	(23)	(21)
HUBER	1.2192	1.2244	1.1136	8.7589	0.1110	0.1067
	(1.6532)	(0.0128)	(0.0061)	(9)	(13)	(11)
WHUBER	1.2112	1.2320	1.1557	8.3420	0.1094	0.1043
	(1.8463)	(0.0159)	(0.0005)	(9)	(11)	(11)
WELSCH	1.0369	1.0564	1.0722	7.8651	0.0991	0.0958
	(0.3575)	(0.0010)	(−0.0023)	(11)	(18)	(13)
WWELSCH	1.0625	1.0646	1.1374	7.5508	0.0982	0.0943
	(0.3210)	(0.0027)	(−0.0030)	(12)	(16)	(20)
BISQUARE	1.0380	1.0499	1.0642	7.8404	0.0987	0.0955
	(0.2982)	(0.0001)	(−0.0010)	(9)	(14)	(13)
WBISQUARE	1.0729	1.0641	1.1401	7.5562	0.0979	0.0943
	(0.2216)	(0.0028)	(−0.0021)	(12)	(14)	(16)
ANDREWS	1.0623	1.0469	1.0587	7.8274	0.0986	0.0954
	(0.1753)	(−0.0003)	(0.0003)	(11)	(13)	(16)
WANDREWS	1.0797	1.0794	1.1381	7.5376	0.0979	0.0942
	(0.2103)	(0.0021)	(−0.0014)	(13)	(15)	(15)
TRIM(.1)	1.8610	2.1419	1.7178	7.7844	0.1012	0.0945
	(−2.0455)	(0.0522)	(0.0377)	(31)	(40)	(28)
WTRIM(.1)	1.4950	1.6357	1.5531	9.9863	0.1331	0.1224
	(0.6097)	(0.0301)	(0.0101)	(14)	(7)	(10)
TRIM(.2)	1.3462	1.2644	1.2852	7.6595	0.0987	0.0930
	(1.1032)	(0.0158)	(0.0064)	(21)	(27)	(23)
WTRIM(.2)	1.3089	1.2923	1.2250	8.4897	0.1125	0.1045
	(1.9266)	(0.0109)	(−0.0043)	(16)	(11)	(7)
MSE OLS	57.1537	0.0104	0.0099			

Table 7. Gaussian Errors

	MRATIO X1 (*BIAS*)	*MRATIO X2* (*BIAS*)	*MRATIO X3* (*BIAS*)	*STDERR X1* (*#REJECT*)	*STDERR X2* (*#REJECT*)	*STDERR X3* (*#REJECT*)
GLS[1:90.]	1.0000 (−1.2859)	1.0000 (0.0157)	1.0000 (0.0228)	19.2253 (6)	0.2890 (6)	 (6)
OLS	7.4861 (26.8895)	2.5971 (0.0325)	3.2873 (−0.0620)	40.8507 (12)	0.5363 (5)	0.4948 (5)
FGLS	10.0072 (42.3009)	2.0947 (0.0620)	5.7588 (−0.4005)	33.1230 (27)	0.5537 (3)	0.4689 (16)
LAD	1.7203 (−1.9932)	1.6771 (0.0814)	1.5414 (0.0675)	22.6582 (17)	0.2975 (33)	0.2744 (16)
WLAD	1.3817 (−0.0817)	1.3884 (0.0763)	1.4587 (0.0252)	17.0704 (20)	0.3144 (17)	0.2634 (18)
HUBER	2.1252 (2.5477)	1.8184 (0.0595)	1.4959 (0.0306)	24.6736 (7)	0.3410 (22)	0.3117 (5)
WHUBER	1.0591 (2.6047)	1.0718 (0.0840)	1.0811 (−0.0099)	16.4077 (13)	0.2851 (11)	0.2555 (9)
WELSCH	2.1149 (1.5416)	1.7966 (0.0085)	1.5441 (−0.0065)	23.9647 (9)	0.3342 (19)	0.3041 (5)
WWELSCH	1.0427 (−2.8810)	1.0364 (0.0655)	1.0854 (0.0455)	15.6347 (15)	0.2721 (10)	0.2462 (8)
BISQUARE	2.1829 (1.6379)	1.8591 (0.0018)	1.5810 (−0.0131)	24.4549 (9)	0.3404 (20)	0.3090 (6)
WBISQUARE	1.0534 (−2.4879)	1.0274 (0.0556)	1.0780 (0.0352)	15.5618 (14)	0.2713 (11)	0.2456 (10)
ANDREWS	2.2040 (1.4479)	1.9067 (0.0075)	1.5588 (−0.0151)	24.6325 (9)	0.3433 (21)	0.3109 (6)
WANDREWS	1.0572 (−2.4716)	1.0285 (0.0550)	1.0811 (0.0345)	15.5599 (14)	0.2714 (11)	0.2456 (10)
TRIM(.1)	1.8490 (−9.4064)	2.0404 (0.2118)	1.6625 (0.1306)	18.0435 (20)	0.2416 (45)	0.2221 (28)
WTRIM(.1)	1.4201 (2.5582)	1.2984 (0.1165)	1.5018 (−0.0333)	21.2781 (18)	0.3378 (15)	0.2934 (19)
TRIM(.2)	1.5108 (−3.3181)	1.4997 (0.1156)	1.3491 (0.0753)	13.3167 (34)	0.1883 (61)	0.1696 (55)
WTRIM(.2)	1.2898 (0.9467)	1.2933 (0.0857)	1.2797 (0.0113)	17.4662 (17)	0.3044 (18)	0.2647 (11)
MSE GLS	268.9060	0.0818	0.0563			

Table 8. Zap Errors

	MRATIO X1 (*BIAS*)	*MRATIO X2* (*BIAS*)	*MRATIO X3* (*BIAS*)	*STDERR X1* (*#REJECT*)	*STDERR X2* (*#REJECT*)	*STDERR X3* (*#REJECT*)
GLS[1:90,]	1.0000 (−0.6987)	1.0000 (0.0276)	1.0000 (−0.0013)	15.0772 (11)	0.2670 (9)	(10)
OLS	8.7335 (26.9735)	2.5556 (0.0474)	4.2964 (−0.0701)	40.7582 (13)	0.5351 (3)	0.4936 (10)
FGLS	12.6471 (44.3936)	2.7040 (0.0407)	6.5094 (−0.4176)	33.1596 (33)	0.5549 (1)	0.4682 (13)
LAD	1.6783 (0.4821)	1.5367 (0.0455)	1.4889 (0.0079)	18.5240 (25)	0.2432 (35)	0.2244 (24)
WLAD	1.3240 (1.9742)	1.2383 (0.0368)	1.2684 (−0.0269)	13.7275 (25)	0.2543 (24)	0.2139 (22)
HUBER	2.1441 (3.3737)	1.4196 (0.0624)	1.6749 (0.0062)	22.8846 (6)	0.3165 (7)	0.2897 (8)
WHUBER	1.0943 (3.7846)	1.0476 (0.0767)	1.0958 (−0.0299)	15.5677 (9)	0.2681 (13)	0.2402 (12)
WELSCH	2.0397 (2.6104)	1.3672 (0.0079)	1.6173 (−0.0319)	21.8801 (5)	0.3059 (14)	0.2798 (12)
WWELSCH	1.0074 (−1.2949)	0.9528 (0.0594)	1.0849 (0.0151)	14.8813 (10)	0.2562 (13)	0.2304 (13)
BISQUARE	2.0589 (2.8094)	1.3853 (−0.0004)	1.6553 (−0.0371)	22.3102 (6)	0.3114 (12)	0.2848 (12)
WBISQUARE	1.0170 (−0.9949)	0.9370 (0.0503)	1.1063 (0.0074)	14.8735 (14)	0.2571 (11)	0.2309 (16)
ANDREWS	2.0586 (2.7769)	1.3868 (−0.0005)	1.6577 (−0.0369)	22.4457 (6)	0.3131 (12)	0.2866 (12)
WANDREWS	0.9970 (−0.8186)	0.9219 (0.0470)	1.1018 (0.0058)	14.9014 (12)	0.2574 (13)	0.2310 (14)
TRIM(.1)	2.2284 (−7.4595)	1.9537 (0.1762)	1.9407 (0.1043)	17.0876 (22)	0.2283 (53)	0.2108 (43)
WTRIM(.1)	1.2756 (4.1482)	1.1020 (0.0810)	1.6418 (−0.0530)	20.4987 (18)	0.3274 (15)	0.2848 (21)
TRIM(.2)	1.5746 (−3.4677)	1.3199 (0.1006)	1.5481 (0.0656)	12.1578 (40)	0.1692 (58)	0.1546 (53)
WTRIM(.2)	1.1254 (2.7533)	0.9566 (0.0446)	1.3353 (−0.0150)	15.9283 (10)	0.2773 (12)	0.2389 (13)
MSE GLS	227.8113	0.0727	0.0510			

Table 9. Double Exponential Errors

	MRATIO X1 (*BIAS*)	*MRATIO X2* (*BIAS*)	*MRATIO X3* (*BIAS*)	*STDERR X1* (*#REJECT*)	*STDERR X2* (*#REJECT*)	*STDERR X3* (*#REJECT*)
GLS[1:90,]	1.0000 (−0.4149)	1.0000 (0.0013)	1.0000 (0.0141)	15.1497 (11)	0.2683 (6)	0.2327 (8)
OLS	7.9487 (22.4447)	3.0563 (0.0793)	4.3155 (−0.0199)	40.7147 (10)	0.5345 (6)	0.4931 (6)
FGLS	10.6423 (40.7972)	2.8319 (0.0497)	5.7409 (−0.3674)	32.6186 (29)	0.5446 (6)	0.4597 (11)
LAD	0.9670 (0.0564)	0.8913 (0.0215)	0.8121 (0.0352)	14.9761 (10)	0.1966 (29)	0.1814 (17)
WLAD	0.7683 (2.0717)	0.7355 (0.0072)	0.7585 (−0.0025)	11.2887 (20)	0.2082 (23)	0.1734 (18)
HUBER	1.6941 (2.5736)	1.1459 (0.0479)	1.2568 (0.0182)	20.6269 (8)	0.2849 (8)	0.2617 (7)
WHUBER	1.0105 (2.7940)	0.8456 (0.0579)	0.9291 (−0.0059)	14.2227 (14)	0.2449 (10)	0.2187 (12)
WELSCH	1.4798 (2.0095)	1.0230 (−0.0062)	1.1060 (−0.0163)	19.2372 (8)	0.2701 (7)	0.2468 (6)
WWELSCH	0.883 (−0.9871)	0.6861 (0.0236)	0.9097 (0.0254)	13.3632 (12)	0.2295 (8)	0.2071 (12)
BISQUARE	1.5024 (1.4476)	1.0744 (−0.0027)	1.1132 (−0.0120)	19.4834 (9)	0.2742 (8)	0.2509 (6)
WBISQUARE	0.8912 (−0.5433)	0.6856 (0.0125)	0.8998 (0.0187)	13.4044 (12)	0.2306 (10)	0.2085 (11)
ANDREWS	1.5051 (1.3766)	1.0794 (−0.0021)	1.1147 (−0.0113)	19.4956 (9)	0.2746 (8)	0.2512 (5)
WANDREWS	0.8909 (−0.4732)	0.6851 (0.0113)	0.8983 (0.0177)	13.3905 (11)	0.2305 (11)	0.2088 (11)
TRIM(.1)	1.5586 (−5.1832)	1.5761 (0.1527)	1.5374 (0.0775)	16.1798 (18)	0.2157 (46)	0.1995 (33)
WTRIM(.1)	1.2210 (5.0055)	1.0941 (0.0693)	1.4803 (−0.0615)	19.5393 (14)	0.3119 (8)	0.2702 (20)
TRIM(.2)	1.0068 (−2.2847)	1.0063 (0.0935)	1.0613 (0.0415)	10.9571 (31)	0.1522 (56)	0.1394 (50)
WTRIM(.2)	0.7995 (3.3809)	0.7754 (0.0448)	1.0307 (−0.0323)	14.5129 (14)	0.2495 (10)	0.2170 (9)
MSE GLS	236.8983	0.0762	0.0488			

Table 10. Contaminated Normal Errors

	MRATIO X1 (*BIAS*)	*MRATIO X2* (*BIAS*)	*MRATIO X3* (*BIAS*)	*STDERR X1* (*#REJECT*)	*STDERR X2* (*#REJECT*)	*STDERR X3* (*#REJECT*)
GLS[1:90,]	1.0000 (0.4106)	1.0000 (0.0015)	1.0000 (−0.0041)	15.4454 (13)	0.2735 (6)	0.2373 (12)
OLS	9.7300 (27.3668)	3.4671 (0.0290)	3.8771 (−0.0686)	40.8564 (13)	0.5364 (5)	0.4948 (7)
FGLS	13.3217 (45.4147)	3.1486 (−0.0003)	5.6785 (−0.4071)	33.0883 (41)	0.5478 (3)	0.4662 (15)
LAD	1.7104 (0.7501)	1.7055 (0.0691)	1.4901 (0.0068)	18.3040 (17)	0.2403 (34)	0.2217 (28)
WLAD	1.2470 (3.7224)	1.3036 (0.0456)	1.2131 (−0.0435)	10.2326 (36)	0.1845 (27)	0.1556 (33)
HUBER	2.2230 (2.9597)	1.5362 (0.0545)	1.5077 (0.0133)	21.2072 (13)	0.2925 (16)	0.2673 (15)
WHUBER	1.0867 (4.2041)	0.9539 (0.0564)	0.9857 (−0.0291)	14.0063 (14)	0.2455 (10)	0.2196 (20)
WELSCH	1.9640 (1.7178)	1.4063 (0.0111)	1.4055 (−0.0192)	20.0183 (12)	0.2801 (12)	0.2547 (17)
WWELSCH	0.9116 (−0.2996)	0.8708 (0.0394)	0.9371 (0.0068)	13.1401 (19)	0.2309 (11)	0.2079 (20)
BISQUARE	1.9789 (1.4192)	1.4332 (0.0121)	1.4274 (−0.0202)	20.3081 (11)	0.2835 (13)	0.2574 (19)
WBISQUARE	0.9032 (0.0573)	0.8625 (0.0299)	0.9213 (−0.0010)	13.1031 (18)	0.2301 (11)	0.2076 (20)
ANDREWS	1.9756 (1.3867)	1.4302 (0.0121)	1.4299 (−0.0202)	20.3709 (11)	0.2844 (13)	0.2581 (18)
WANDREWS	0.9026 (0.0817)	0.8632 (0.0293)	0.9175 (−0.0019)	13.1204 (19)	0.2304 (11)	0.2077 (19)
TRIM(.1)	2.0882 (−7.2130)	1.8924 (0.1731)	1.5573 (0.1111)	16.1757 (30)	0.2158 (42)	0.1989 (35)
WTRIM(.1)	1.3664 (5.9778)	1.2607 (0.0708)	1.2003 (−0.0751)	19.7012 (22)	0.3114 (12)	0.2710 (18)
TRIM(.2)	1.2951 (−1.5137)	1.2000 (0.0932)	1.1256 (0.0445)	11.3159 (37)	0.1591 (53)	0.1441 (48)
WTRIM(.2)	1.0948 (3.8142)	0.9993 (0.0539)	1.0219 (−0.0346)	14.5334 (16)	0.2561 (8)	0.2225 (16)
MSE GLS	232.4865	0.0656	0.0582			

and TRIM estimators are relatively inefficient compared to the M-estimators, and particularly the hard redescenders that put zero or low weights on many of the bad observations. Somewhat surprisingly the weighted forms of the trimmed estimators perform better both in terms of MSE and producing standard errors which result in correct t-tests. The same comments in Section V about the actual level of the t-test generally holds here for most of the other estimators. There are stronger indications, however, that the test statistics from OLS and FGLS are unreliable.

When we move to the heteroscedasticity case with gross errors, as Tables 7–10 show, the performance of OLS and FGLS appears to improve (relative to the GLS estimated based the good observations) from what it was in Table 6, but this is largely an illusion due to the larger true variances caused by the heteroscedasticity multiplier. In contrast to the simulations in Section V (no gross errors) where FGLS performed reasonably well, FGLS is now clearly an undesirable estimator due to its bias. The TRIM, HUBER, and the LAD estimators pick up part of the gross errors in the form of bias although to a much smaller degree than FGLS and OLS.[32] The weighted hard redescenders (WBISQUARE and WANDREWS) as expected are best able to deal with the combination of very thick error tails, heteroscedasticity, and a significant number of gross errors. Results for sample sizes of 50 and 500 (not reported here) suggest that increasing the sample size increases the problems with FGLS and OLS if the gross errors increase proportionately.[33] The ability of most of the robust techniques to capture the true parameter estimates in the face of these adversities is truly impressive. Their ability to provide fairly reasonable hypothesis tests concerning those values is even more impressive.[34]

VII. AN EMPIRICAL EXAMPLE: THE BOSTON HOUSING DATA

The Boston housing data set collected by Harrison and Rubinfeld (1978) has been used by a number of authors [e.g., Belsley et al. (1980) and Krasker et al. (1982)] to illustrate different statistical techniques. This data set is characterized by a number of very high leverage points and very influential observations as well as nonnormal error terms.

We estimated the Boston housing equation using each of the techniques employed in the previous two sections. The results are displayed in Table 11.[35]

Table 11. Boston Parameter Estimates and Standard Errors

VAR	*OLS (STDERR)*	*GLS (STDERR)*	*LAD (STDERR)*	*WLAD (STDERR)*	*HUBER (STDERR)*	*WHUBER (STDERR)*
INTERCEPT	9.75629 (.14958)	9.64219 (.11971)	9.47830 (.11782)	9.65357 (.11526)	9.62482 (.12957)	9.62442 (.10491)
CRIME	−.01186 (.00124)	−.01266 (.00237)	−.01179 (.00098)	−.01322 (.00277)	−.01109 (.00325)	−.01356 (.00296)
ZN	.00008 (.00051)	.00016 (.00033)	.00048 (.00039)	.00056 (.00030)	.00005 (.00037)	.00023 (.00026)
INDUS	.00024 (.00237)	.00124 (.00164)	.00218 (.00186)	.00065 (.00143)	.00124 (.00161)	.00106 (.00119)
CHAS	.09140 (.03320)	.07821 (.00119)	.06319 (.02608)	.04499 (.01867)	.07510 (.02438)	.06066 (.02137)
NOXSQ	−.00638 (.00113)	−.00405 (.00098)	−.00443 (.00089)	−.00346 (.00107)	−.00499 (.00094)	−.00383 (.00083)
ROOMS	.00633 (.00131)	.01379 (.00121)	.01399 (.00103)	.01762 (.00098)	0.1181 (.00152)	.01654 (.00098)
AGE	.00009 (.00053)	−.00104 (.00037)	−.00062 (.00041)	−.00159 (.00032)	−.00069 (.00045)	−.00139 (.00034)
DIS	−.19126 (.03339)	−.14404 (.02559)	−.13854 (.02623)	−.15426 (.02270)	−.16309 (.02884)	−.15861 (.02239)
RAD	.09571 (.01913)	.07309 (.01295)	.05649 (.01503)	.05375 (.01181)	.06943 (.01478)	.05493 (.01125)
TAX	−.00042 (.00012)	−.00040 (.00008)	−.00034 (.00010)	−.00038 (.00009)	−.00036 (.00009)	−.00035 (.00007)
PTRATIO	−.03112 (.00501)	−.02914 (.00337)	−.02567 (.00394)	−.02812 (.00320)	−.02881 (.00356)	−.02799 (.00279)
BLACK	.36370 (.10312)	.58695 (.10371)	.63315 (.08098)	.68531 (.13512)	.55271 (.11583)	.67693 (.09973)
LSTAT	−.37116 (.02501)	−.22334 (.02153)	−.24011 (.01964)	−.15065 (.01910)	−.27619 (.02898)	−.18033 (.02014)

Table 11. (continued)

VAR	WELSCH (STDERR)	WWELSCH (STDERR)	BISQUARE (STDERR)	WBISQUARE (STDERR)	ANDREWS (STDERR)	WANDREWS (STDERR)
INTERCEPT	9.55781	9.63427	9.52720	9.58101	9.52677	9.57996
	(.12343)	(.10737)	(.13668)	(.10298)	(.15594)	(.10373)
CRIME	−.00983	−.01316	−.01792	−.01735	−.01799	−.10736
	(.00343)	(.00252)	(.00448)	(.00321)	(.00629)	(.00323)
ZN	−.00004	.00029	.00005	.00035	.00005	.00035
	(.00031)	(.00026)	(.00031)	(.00026)	(.00031)	(.00026)
INDUS	.00182	.00092	.00168	.00112	.00169	.00116
	(.00146)	(.00120)	(.00146)	(.00122)	(.00147)	(.00124)
CHAS	.06687	.05842	.06276	.05955	.06249	.05917
	(.02314)	(.01871)	(.02221)	(.01850)	(.02247)	(.01844)
NOXSQ	−.00399	−.00390	−.00340	−.00355	−.00338	−.00354
	(.00099)	(.00083)	(.00103)	(.00081)	(.00116)	(.00081)
ROOMS	.01542	.01689	.01619	.01709	.01625	.01711
	(.00155)	(.00093)	(.00150)	(.00090)	(.00162)	(.00090)
AGE	−.00129	−.00140	−.00153	−.00145	−.00155	−.00145
	(.00046)	(.00034)	(.00042)	(.00032)	(.00042)	(.00321)
DIS	−.14140	−.16549	−.14451	−.16350	−.14420	−.16262
	(.02778)	(.02313)	(.02925)	(.02225)	(.03076)	(.02210)
RAD	.05650	.05312	.06209	.05622	.06951	.05614
	(.01349)	(.01146)	(.01337)	(.01153)	(.01359)	(.01150)
TAX	−.00037	−.00037	−.00029	−.00031	−.00029	−.00031
	(.00009)	(.00007)	(.00009)	(.00008)	(.00010)	(.00008)
PTRATIO	−.02777	−.02759	−.02674	−.02683	−.02673	−.02690
	(.00322)	(.00271)	(.00314)	(.00265)	(.00321)	(.00262)
BLACK	.69204	.67891	.70327	.69806	.70391	.69727
	(.12149)	(.11141)	(.12857)	(.11154)	(.13626)	(.11160)
LSTAT	−.21073	−.17372	−.19042	−.17066	−.18931	−.17067
	(.03149)	(.01998)	(.02908)	(.01920)	(.02997)	(.01959)

	TRIM(.1) (STDERR)	WTRIM(.1) (STDERR)	TRIM(.2) (STDERR)	WTRIM(.2) (STDERR)
INTERCEPT	9.81525	9.70798	9.67698	9.61341
	(0.16584)	(0.11128)	(0.07322)	(0.11092)
CRIME	−0.01559	−0.01387	−0.01312	−0.01472
	(0.00200)	(0.00274)	(0.00132)	(0.00318)
ZN	0.00044	0.00024	0.00043	0.00054
	(0.00050)	(0.00030)	(0.00022)	(0.00028)
INDUS	0.00080	0.00035	0.00098	0.00109
	(0.00232)	(0.00131)	(0.00102)	(0.00131)
CHAS	0.05767	0.3921	0.05355	0.04008
	(0.03496)	(0.01911)	(0.01628)	(0.01832)
NOXSQ	−0.00532	−0.00422	−0.00496	−0.00390
	(0.00123)	(0.00091)	(0.00059)	(0.00093)
ROOMS	0.01413	0.01676	0.01433	0.01772
	(0.00166)	(0.00096)	(0.00075)	(0.00106)
AGE	−0.00132	−0.00140	−0.00091	−0.00150
	(0.00056)	(0.00033)	(0.00025)	(0.00033)
DIS	−0.20908	−0.17564	−0.18503	−0.16536
	(0.03620)	(0.02303)	(0.01604)	(0.02261)
RAD	0.06933	0.05722	0.06595	0.05933
	(0.01964)	(0.01196)	(0.00894)	(0.01206)
TAX	−0.00038	−0.00038	−0.00039	−0.00038
	(0.00012)	(0.00008)	(0.00006)	(0.00008)
PTRATIO	−0.02734	−0.02746	−0.02916	−0.02632
	(0.00498)	(0.00311)	(0.00219)	(0.00290)
BLACK	0.44888	0.55580	0.61855	0.67597
	(0.13108)	(0.11620)	(0.06183)	(0.11565)
LSTAT	−0.21178	−0.17575	−0.22671	−0.15807
	(0.03151)	(0.01979)	(0.01453)	(0.01925)

The first thing to note is the NOXSQ parameter (obtained using OLS) that was the primary focus of the original Harrison and Rubinfeld (1978) study is over one and a half times that obtained with any of the other techniques including FGLS. Evidently, correcting for heteroscedasticity using the Harvey functional form and $Z = X$ for the heteroscedasticity design matrix is sufficient to bring this parameter in line with that suggested by the robust methods.[36] The FGLS estimation downweights the center city Boston tracts that Belsley et al. identify as causing estimation problems.

The (Harvey) heteroscedasticity equations using several of the robust estimators and OLS are displayed in Table 12. There is agreement among all of the equations that the variance of the error terms

Table 12. Boston Heteroscedasticity Parameters

	OLS (STDERR)	*TRIM(.2) (STDERR)*	*HUBER (STDERR)*	*WELSCH (STDERR)*	*BISQUARE (STDERR)*	*ANDREWS (STDERR)*
INTERCEPT	−2.8522 (1.8026)	−3.0354 (1.277)	−4.1814 (1.6503)	−5.7942 (1.7257)	−5.3491 (1.6981)	−5.3172 (1.6959)
CRIME	0.0250 (0.0150)	0.0144 (0.0091)	0.0256 (0.0120)	0.0202 (0.0129)	0.0489 (0.0131)	0.0488 (0.0131)
ZN	0.0022 (0.0060)	0.0046 (0.0039)	0.0061 (0.0050)	0.0081 (0.0052)	0.0072 (0.0052)	0.0071 (0.0052)
INDUS	−0.0508 (0.0284)	−0.0218 (0.0199)	−0.0472 (0.0242)	−0.0451 (0.0284)	−0.0357 (0.0274)	−0.0355 (0.0272)
CHAS	0.4401 (0.4001)	−0.176 (0.314)	0.2678 (0.3381)	0.1185 (0.3670)	0.1150 (0.4076)	0.1096 (0.4050)
NOXSQ	−0.0066 (0.0136)	−0.0043 (0.0097)	−0.00971 (0.0119)	−0.0032 (0.0125)	−0.0102 (0.0127)	−0.0102 (0.0127)
ROOMS	−0.0012 (0.0158)	−0.0232 (0.0116)	−0.0265 (0.0147)	−0.0338 (0.0157)	−0.0350 (0.0157)	−0.0355 (0.0155)
AGE	−0.0099 (0.0063)	0.0038 (0.0045)	−0.0014 (0.0056)	0.0027 (0.0057)	0.0036 (0.0059)	0.0033 (0.0059)
DIS	−1.5756 (0.4024)	−0.5008 (0.283)	−1.2595 (0.3672)	−1.0649 (0.3712)	−1.0051 (0.3669)	−1.0009 (0.3664)
RAD	0.1010 (0.2305)	0.0699 (0.159)	0.0075 (0.2004)	−0.0430 (0.2061)	−0.0735 (0.2010)	−0.0709 (0.2006)
TAX	0.0022 (0.0014)	0.0017 (0.001)	0.0042 (0.0012)	0.0050 (0.0014)	0.0043 (0.0013)	0.0043 (0.0013)
PTRATIO	−0.393 (0.0604)	0.0099 (0.0399)	−0.0213 (0.0505)	0.0188 (0.0529)	0.0187 (0.0528)	0.0199 (0.0528)
BLACK	1.3183 (1.2426)	0.3687 (0.8821)	1.0071 (1.1314)	0.7049 (1.1971)	−0.0432 (1.2583)	−0.0879 (1.2538)
LSTAT	−0.0800 (0.3013)	−0.3663 (0.2163)	−0.3259 (0.2732)	−0.4526 (0.2757)	−0.4828 (0.2886)	−0.4780 (0.2870)

decreases as the distance to the nearest employment center (DIS) increase. The robust estimators, however, also suggest that the variance of the error terms increases with the crime rate, decreases as the number of rooms goes up, and increases with the tax rate. These results are fairly consistent with the urban economics literature and could be expected in center city areas where some of the most expensive and least expensive census tracts are in close proximity to each other. However, the large change in some of the coefficients indicates that we are probably dealing with other problems in addition to heteroscedasticity.

Belsley et al. (1980, pp. 238–239) using a variety of indicators, including those based on the observation's leverage (the hat matrix) and the magnitude of the studentized residuals, provide a list of 67 observations (out of 506) that are likely to cause problems in estimating the Boston housing equation. Table 13 shows how many of these 67 observations were down-weighted by each of the M-estimators. The 10% trimmed least-squares estimator (20% of the observations deleted) sets 46 (of the 67) observations to zero; the weighted 10% trimmed least squares, 48; the 20% trimmed least squares estimator, 54; and the weighted 20% trimmed least squares estimator, 54.[37] Table 14 shows the total number of times (out of 12) that the M-estimators or trimmed least-squares estimators placed each of the 67 observations among the 20% of those observations with the smallest weights. It is interesting to note that many of the Cambridge tracts [143–172] were not down-weighted while most of the Boston city tracts (including a number of those not in Belsley et al.'s list) were heavily down-weighted by almost all of the techniques. This finding reinforces Belsley et al.'s basic conclusion that it is likely that many of the Boston (proper) tracts belong to a fundamentally different housing market than the majority of tracts included in the Harrison and Rubinfeld study.

VIII. CONCLUDING REMARKS

What we have attempted to do in this chapter is to adhere to the original philosophy behind the development of robust estimators: (1) the estimator should not give up much efficiency relative to the maximum likelihood estimator if the assumptions being made are true and (2) the estimator should be more efficient and more importantly less prone to bias if those assumptions are not met. The standard

Table 13. Number of Troublesome Observation Down-weighted by M-Estimators

	HUBER	*WHUBER*	*WELSCH*	*WWELSCH*	*BISQUARE*	*WBISQUARE*	*ANDREWS*	*WANDREWS*
Lowest 10%	37	25	37	20	35	21	35	21
Lowest 20%	47	37	46	35	46	34	. 46	34
Lowest 40%	48	37	54	50	53	47	52	47

Table 14. Robust Estimators (12) Down-weighting Each Troublesome Observation

OB #	*# LOW*	*OB #*	*# LOW*	*OB #*	*# LOW*	*OB #*	*# LOW*
[8]	12	[163]	0	[369]	12	[410]	12
[124]	0	[164]	0	[370]	12	[411]	9
[127]	0	[215]	12	[371]	9	[412]	8
[143]	8	[258]	3	[372]	12	[413]	12
[144]	0	[284]	0	[373]	12	[414]	10
[148]	0	[285]	2	[381]	8	[415]	6
[149]	12	[343]	12	[386]	10	[416]	12
[151]	0	[358]	2	[388]	9	[417]	12
[152]	1	[359]	9	[392]	12	[419]	9
[153]	0	[360]	12	[398	12	[420]	12
[154]	11	[361]	11	[399]	10	[427]	8
[155]	2	[362]	5	[400]	12	[467]	12
[156]	0	[363]	9	[401]	12	[474]	12
[157]	0	[365]	12	[402]	12	[490]	12
[160]	0	[366]	12	[404]	9	[491]	12
[161]	0	[367]	12	[406]	5	[506]	12
[162]	2	[368]	12	[408]	12		

robust regression estimators fail on the first criteria. In the presence of heteroscedasticity, they can be quite inefficient relative to the maximum likelihood estimator that takes account of that heteroscedasticity. Unfortunately, the cross-sectional regression problems typically considered by economists wil be characterized by heteroscedasticity as well as the other problems that make robust regression look attractive. Robust regression retains all of its good properties in this situation, however, after weighting to correct for the heteroscedasticity. And there appears to be a bonus to this weighting of the robust estimator if the weights are estimated robustly rather than with OLS. This is particularly true if the underlying error distribution is somewhat thick tailed or contaminated with gross errors.

There are obvious extensions to the work we have presented here. These include experimenting with different forms of heteroscedasticity,[38] different methods of generating gross errors, different types of leverage in the design matrix, gross errors in the design matrix, and other possible error distributions including short-tailed and more importantly, asymmetric ones (Carroll, 1979). We devoted much of this chapter to discussing the estimation of regression coefficients. The issue of robust estimation of variances has barely been touched upon and yet the "frail" nature of variance estimates with nonnormal data and the relatively good performance of location estimators were noted long ago in Box's (1953) classic article. One of the reasons for the good performance of the robust estimators in the normal heteroscedastic cases is that they are superior in estimating the weights. The Boston housing regressions touched upon this issue, and we believe that most tests of heteroscedasticity should be carried out using one of the robust techniques.

To some extent the M-estimators outperformed the TRIM estimators. We note here that in our modification of the Barrodale and Roberts algorithm, the trimmed least-squares estimator is extremely fast and has none of the slow convergence problems of the hard redescenders.[39] It is also scale invariant which is an attractive feature. We have real problems with the LAD estimator. The need for a "smooth" to get the proper LAD standard errors is toublesome. For the distributions under investigation, we were able experimentally to set the degree of smoothing used so that the parameter $f(0)^{-1}$ could be estimated reasonably efficiently. In very large samples the appropriate window size for the smooth is less troublesome. This is an area of potentially useful research although the LAD estimators were consistently dominated by the other robust estimators so we do not recommend their use. In general, however, the choice of which robust estimator is less important than simply the use of one of them. We have a slight preference for the WHUBER or the WTRIM(0.1) unless one expects a high percentage of bad observations in which case one of the hard redescenders would be preferred.

We initially thought that confidence interval estimation and hypothesis testing was going to be much more of a problem than it appears to be. We had anticipated having to use Efron's (1979) bootstrap to obtain valid standard errors. That technique is very computer intensive and tends to produce results less accurate than those obtained here.[40] Bootstrapping's best applications are probably in smaller samples where it is unlikely that any asymptotic properties can be relied upon. A related but less computationally intensive method would be to modify White's (1980b) heteroscedasticity-consistent variance matrix for the robust estimators used here. It would be interesting to compare the efficiency of that procedure against direct estimation of the weights when the functional form of the heteroscedasticity equation was misspecified.

As a recommendation for applied work we can see little reason not to use one of the robust estimators. They are quick and produce virtually the same results as OLS and FGLS when the normality and no gross error assumptions are met. This is by now an old

recommendation, but economists are only slowly moving in that direction with their applied work. We have shown that robust methods work well in the presence of heteroscedasticity and gross errors, two of the primary characteristics of the cross-sectional data sets used by economists. The gains from using weighted robust regression estimators appear to be even larger in this environment than in the simple location cases often put forth by statisticians to support the use of robust techniques.

Finally, we hope that we have provided the missing elements that we so often hear as the reason for not using robust regression, i.e., how do you test hypotheses and where are my t-statistics. Our work here suggests that doing this in a robust regression framework is not much more difficult than in the OLS/FGLS framework.[41] We believe that hypothesis tests carried out with robust methods are likely to be much more valid than those based on methods requiring independence, normality, and the absence of gross errors.

NOTES

1. White (1980a) suggests a number of reasons, particularly stratification in sample surveys, why regression estimates of cross-sectional economic data are almost always characterized by heteroscedasticity. Most introductory and advanced econometric books contain good descriptions of how heteroscedasticity comes about. Judge et al. (1985) provides a good overview of the different test and correction methods commonly used in econometrics.

2. Krasker et al. (1982) have emphasized the gross error nature of economic cross-sectional data.

3. Recent work on this topic includes Dijkstra and van der Zouwen (1983), Rossi et al. (1983), and Wright (1983).

4. This data set has become well known through its use in the book by Belsley et al. (1980) on regression diagnostics.

5. We can only hope to present a brief overview of robust regression here. The interested reader should refer to the review article by Koenker (1982), the chapter by Goodall in Hoaglin et al. (1983), and the books by Huber (1981) and Hampel et al. (1986) for more discussion.

6. The two other general classes of estimators are L-estimators, which are based on linear combinations of order statistics, and R-estimators, which are based on ranks. M- and L-estimators are considered in this paper. Huber (1981) and Lehmann (1983) provide general discussions of the different types of estimators.

7. Several M-estimators with different ρ's have been proposed (Andrews et al., 1972; Huber, 1981), most of wheih can be efficiently and easily computed using iteratively reweighted least squares (Coleman et al., 1980). If the p.d.f. f is known, the choice $\rho = -\log(f)$ gives the maximum likelihood estimator, which for the normal density ($\rho = x^2$) is the OLS estimator.

8. The Andrews and the bisquare are smooth redescenders that have some desirable properties that the discontinuous three-part Hampel redescenders do not enjoy (Hoaglin et al., 1983).

9. This is not true of those M-estimators in the l_1 norm (LAD) or the l_2 norm (OLS).

10. The problem of starting values is far more acute for redescending estimators that may have local minima than for the Huber which is well behaved and converges readily.

11. We consider symmetric trimming here. The trimmed means with trimming fraction α between 0 and 0.25 has been shown to perform well in a wide variety of situations including those in which only normal measurement error of a physical constant should have been present (Stigler, 1977).

12. Ruppert and Carroll also show that as α approaches either zero or 0.5 that the trimmed least-squares estimator approaches its preliminary estimator.

13. By recasting the minimization problem for the regression quantile as a linear programming problem, it can be shown that k of the residuals from the regression quantile are zero.

14. Since generalization of this estimator to the heteroscedasticity case appears quite different from the traditional M-estimators, we do not consider it further in this chapter.

15. A mixture of normals is frequently referred to as a contaminated normal, particularly when the percentage of observations coming from the normal with the larger variance is small.

16. Sims (1971) over a decade ago noted the problem and importance of distinguishing between heteroscedasticity and long-tailed error distributions.

17. In the sense that the diagonal elements of the hat matrix are equal.

18. We note that almost all of the Monte Carlo results on robust regression location estimators to date have involved samples sizes between five and fifty observations. Results on moderate and larger size samples are almost nonexistent.

19. Theorems such as those proved by Huber (1981) that are based on the behavior of an estimator as n becomes large with a fixed p/n ratio tend to provide much more guidance for applied work with moderately sized samples than do standard asymptotic results.

20. Note that at fairly high levels of kurtosis (i.e., kurtosis > 4.5, where the kurtosis of the normal distribution is 3) the performance of various robust and nonrobust estimators begins to diverge quite significantly between the members of the exponential power family and the t-distribution with the same level of kurtosis. See D'Agostino and Lee (1977) for some results in the location case.

21. We used the algorithm of Johnson et al. (1980).

22. The ZAP is virtually indistinguishable from a Student's t-distribution with 6 degrees of freedom. The normal has a kurtosis of 3, the ZAP 4.5, the double exponential 6, and the particular contaminated normal we used 8.3. Our contaminated normal draws 10% of the error terms from a normal distribution with a variance 9 times that of the other 90% of the error terms.

23. The researcher should consider Krasker and Welsch's (1982) bounded influence estimator if the $x_i's$ can take on a very wide range of values and the X matrix has very high leverage points that cannot be rejected a priori as bad data.

24. Virtually all of the Monte Carlo simulations we have come across have represented gross errors in this way. The contaminated normal and the "one wild" distribution have been particularly popular (e.g., Hoaglin et al., 1983).

25. Again we feel that $y_i's$ above 1000 are likely to be caught by the careful researcher. $Y_i's$ in the 500 to 1000 range are occasionally generated by the true distribution and hence could not be rejected easily a priori as bad data, although such points might be suspicious.

26. If there is no heteroscedasticity as in Tables 1 and 6, the GLS estimator is simply OLS. In the Monte Carlo experiments reported in Section VI the comparisons are always made to the GLS estimator using only the good observations.

27. For the Huber, K_h = 1.345; for the WELSCH K_w = 2.985; for the BISQUARE K_b = 6.2; for the ANDREWS K_a = 1.339.

28. The potential user should note two flaws in this routine that we had to correct for our experiments: (1) the WELSCH function is miscoded in ROSEPACK and (2) the intrinsic weight function in the S interface (necessary for heteroscedasticity correction) is miscoded.

29. Out of 200 repetitions, 10 rejects is exactly the two-tailed 5% level of rejections. The number of rejections has a binomial distribution and the expected range for most of these distributions is 8 to 12 rejections.

30. Tables for experiments using sample sizes of 50 and 500 are to be found in the discussion paper version of this chapter (No. 87-24, Dept. of Economics, University of California, San Diego).

31. Keep in mind that these results are for a design matrix with virtually no leverage. It is easy to obtain results (not reported here) which show that OLS and FGLS can become arbitrarily bad as gross errors occur in combination with high leverage points.

32. Bias is usually picked up in the intercept coefficient β_1 and in the β_3 coefficient. An examination of the X matrix reveals that what little leverage this matrix has is in the last 10 observations and is caused by a small x_3 value.

33. Tables for these sample sizes are presented in our Discussion Paper 87-24, University of California, San Diego.

34. Only the LAD, WLAD, and unweighted TRIMs consistently produce confidence intervals that are much too short. The properties of this estimator are known to change somewhat in the presence of heteroscedasticity. The work reported in this and the previous section suggests that valid hypothesis testing for this estimator will depend on determining whether or not the heteroscedasticity can be detected. In the experiments presented here this does not seem to be a problem.

35. This data set is reproduced in Belsley et al. (1980).

36. Harrison and Rubinfeld (1978) state that they obtained positive results from a test of heteroscedasticity but that correcting for it did not make much of a change in the NOXSQ coefficient. The test they performed is not described in detail.

37. The bisquare sets the weights of 20 observations to zero, all of which are among the 67 enumerated by Belsley et al. The weighted bisquare sets the weights on 4 observations to zero, 3 of which are among the 67; the Andrew's, 22, all of which are among 67; and the weighted Andrew's, 6, 4 of which are among the 67.

38. Carroll and Ruppert's (1982b) work is the only other that we know of that provides Monte Carlo results on robust estimation with heteroscedasticity. They found that the form of heteroscedasticity had little effect in a comparison between different estimators.

39. In one small experiment of 10 repetitions including calculating the standard errors for the 100 observation case, we found the following times (in seconds) on the VAX 11/750 using the S statistical computing package: OLS (16), LAD (29), TRIM (33), HUBER (46), and ANDREWS (57).

40. Wu (1986) details some of the problems with using the boostrap with the heteroscedastic linear model.

41. Programs for estimators, including the calculation of standard errors, described in this chapter are available from the authors which are easily installed as user supplied S (Becker and Chambers, 1984) functions on a UNIX machine.

REFERENCES

Andrews, D. F., P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, and J. W. Tukey (1972) *Robust Estimates of Location*. Princeton University Press, Princeton, NJ.

Anscombe, F. J. (1961) "Examination of residuals." *Proceedings of the 5th Annual Berkeley Symposium*, 1, 1–36.

Barrodale, I. and F. D. K. Roberts (1973) "An improved algorithm for discrete l_1 linear approximations." *SIAM Journal of Numerical Analysis*, 10, 839–848.

Bartlett, M. S. (1937) "Properties of sufficiency and statistical tests." *Proceedings of the Royal Society, Series A*, 160, 268–282.

Bassett, G. W. and R. W. Koenker (1978) "The asymptotic distribution of the least absolute error estimator." *Journal of the American Statistical Association*, 73, 618–622.

Becker, R. A. and J. M. Chambers (1984) *S: An Interactive Environment for Data Analysis and Graphics*. Wadsworth, Belmont, CA.

Belsley, D. A., E. Kuh, and R. E. Welsch (1980) *Regression Diagnostics: Identifying Influential Data Sources of Collinearity*. John Wiley, New York.

Bickel, P. J. (1973) "On some analogues to linear combinations of order statistics." *Annals of Statistics*. 1, 597–616.

Bickel, P. J. (1976) "Another look at robustness." *Scandinavian Journal of Statistics*, 3, 145–148.

Bickel, P. J. (1978) "Using residuals robustly I." *Annals of Statistics*, 6, 266–291.

Box, G. E. P. (1953) "Non-normality and tests on variances." *Biometrika*, 40, 318–335.

Box, G. E. P. and G. C. Tiao (1973) *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, MA.

Carroll, R. J. (1979) "On estimating the variances of robust estimators when the errors are asymmetric." *Journal of the American Statistical Association*, 74, 674–679.

Carroll, R. J. and D. Ruppert (1982a) "Comparison between maximum likelihood and generalized least squares in a heteroscedastic model." *Journal of the American Statistical Association*, 77, 878–882.

Carroll, R. J. and D. Ruppert (1982b) "Robust estimation in heteroscedastic linear models." *Annals of Statistics*, 10, 429–441.

Coleman, D., P. Holland, N. Kaden, V. Klema, and S. C. Peters (1980) "A system of subroutines for iteratively reweighted least squares computers." *ACM Transactions on Mathematical Software*, 6, 327–336.

D'Agostino, R. B. and A. F. S. Lee (1977) "Robustness of location estimators under changes of population kurtosis." *Journal of the American Statistical Association*, 72, 393–396.

Dijkstra, W. and J. van der Zouwen (1983) *Response Behavior in the Survey Interview*. Academic Press, New York.

Efron, B. (1979) "Bootstrap methods: Another look at the jackknife." *Annals of Statistics*, 7, 1–26.

Friedman, J. and W. Stuetzle (1982) "Smoothing of scatterplots." Technical Report ORION006, Department of Statistics, Stanford University.

Fulton, M., S. Subramanian, and R. Carson (1987) "Fast regression quantiles using a modification of the Barrodale and Roberts' l_1 algorithm." Discussion paper no. 87-8, University of Calfornia, San Diego.

Gross, A. L. (1977) "Confidence intervals for bisquare regression estimates." *Journal of the American Statistical Association*, 72, 341–353.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986) *Robust Statistics: The Approach Based on Influence Functions*. John Wiley, New York.

Harrison, D. and D. Rubinfeld (1978) "Hedonic housing prices and the demand for clean air." *Journal of Environmental Economics and Management*, 5, 81–102.

Harvey, A. C. (1976) "Estimating regression models with multiplicative heteroscedasticity." *Econometrica*, 44, 461–465.

Hoaglin, D. C., F. Mosteller, and J. W. Tukey (eds.) (1983) *Understanding Robust and Exploratory Data Analysis*. John Wiley, New York.

Holland, P. and R. E. Welsch (1977) "Robust regression using iteratively reweighted least squares." *Communications in Statistics*, 9, 813–827.

Hotelling, H. (1961) "The behavior of some standard statistical tests under non-standard conditions." *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 319–360.

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

Jaeckel, L. A. (1971) "Robust estimates of location." *Annals of Mathematical Statistics*, 42, 1020–1034.

Jobson, J. D. and W. A. Fuller (1980) "Least squares estimation when the covariance matrix and parameter vector are functionally related." *Journal of the American Statistical Association*, 75, 176–181.

Johnson, M. E., G. L. Tietjen, and R. J. Beckman (1980) "A new family of probability distributions with applications to Monte Carlo studies." *Journal of the American Statistical Association*, 75, 276–279.

Judge, G. E., W. E. Griffiths, R. C. Hill, H. Lutkepohl, and T. C. Lee (1985) *The Theory and Practice of Econometrics*, 2nd ed. John Wiley, New York.

Koenker, R. W. (1982) "Robust methods in econometrics," with discussions by G. M. Duncan, D. G. Fiebig, H. Theil, J. A. Hausman, and R. E. Welsch. *Econometric Reviews*, 1, 213–289.

Koenker, R. W. and G. W. Basset (1978) "Regression quantiles." *Econometrica*, 46, 33–50.

Koenker, R. W. and G. W. Bassett (1982) "Test of linear hypotheses and l_1 estimation." *Econometrica*, 50, 1577–1583.

Koenker, R. W. and V. D'Orey (1987) "Computing regression quantiles." *Applied Statistics*, 150, 383–393.

Krasker, W. S. and R. E. Welsch (1982) "Efficient bounded-influence regression estimates." *Journal of the American Statistical Association*, 77, 590–604.

Krasker, W. S., E. Kuh, and R. E. Welsch (1982) "Estimation for dirty data and flawed models." In *Handbook of Econometrics*, Vol. I, Z. Griliches and M. D. Intrilligator (eds.) North-Holland, Amsterdam.

Lehmann, E. L. (1983) *The Theory of Point Estimation*. John Wiley, New York.

Mosteller, F. and J. W. Tukey (1977) *Data Analysis and Regression*. Addison-Wesley, Reading, MA.

Parzen, E. (1979) "Nonparametric statistical data modeling." *Journal of the American Statistical Association*, 74, 105–131.

Prais, S. J. and H. S. Houthakker (1955) *The Analysis of Family Budgets*. Cambridge University Press, New York.

Rossi, P. H., J. D. Wright, and A. B. Anderson (eds.) (1983) *Handbook of Survey Research*. Academic Press, New York.

Ruppert, D. and R. J. Carroll (1980) "Trimmed least squares estimation in the linear model." *Journal of the American Statistical Association*, 75, 828–838.

Sims, C. A. (1971) "Linear regression with non-normal errors terms: A comment." *Review of Economics and Statistics*, 53, 204–205.

Stigler, S. M. (1977) "Do robust estimators work with real data." *Annals of Statistics*, 5, 1055–1098.

Subramanian, S. and R. T. Carson (1987) "The use of the trimmed least squares estimator in the presence of heteroscedasticity." Draft manuscript, University of California, San Diego.

White, H. (1980a) "Non-linear regression on cross section data." *Econometrica*, 48, 721–746.

White, H. (1980b) "A heteroscedasticity-consistent estimator and a direct test for heteroscedasticity." *Econometrica*, 48, 817–838.

Wright, T. (ed.) (1983) *Statistical Methods and the Improvement of Data Quality*. Academic Press, New York.

Wu, C. F. J. (1986) "Jackknife, bootstrap and other resampling methods in regression analysis." *Annals of Statistics*, 14, 1261–1295.