ROBUST ESTIMATORS OF REGRESSION

by

Franco Peracchi*

Department of Economics
University of California
Los Angeles

May 1988
UCLA Working Paper No. 476

Abstract

This paper presents recent results on robust estimation of the linear regression model. The concept of influence function is used to compare and contrast the efficiency and robustness properties of a number of regression estimators. Various computational problems are also discussed.

JEL Classification No. 211

Keywords: Regression, M-estimators, Robust estimators.

* This paper is based on a Chapter of my Ph.D. dissertation at Princeton University. I am very grateful to my advisors, Angus Deaton, Whitney Newey and Elvezio Ronchetti, for their generous guidance.

1. Introduction

In recent years much research has been devoted to developing regression estimators that perform well when the assumed statistical model is correctly specified, while being robust, that is, not too sensitive to small violations of the model assumptions [see e.g. Huber (1981) and Hampel et al. (1986)]. This informal definition of robustness corresponds to Hampel's (1971) definition of qualitative robustness, which formalizes the notion that an estimator is robust when its probability distribution changes little under small changes in the underlying probability distribution of the observations. Because a test inherit the efficiency and robustness properties of the estimator on which it is based, tests based on robust estimators are also robust, that is, their level and power are stable under small departures from the model assumptions. Tests based on robust estimators with good efficiency properties are robust and have, in addition, good power properties.

The purpose of this paper is to present a number of robust regression estimators and discuss their properties in a unified framework. We hope that this will provide valuable information to the applied investigators about the relative merits of the various estimators, and will encourage their use.

The remainder of the paper is organized as follows. Section 2 presents the statistical model and the class of estimators that we consider. Sections 3-7 introduce the various estimators and discuss their efficiency and robustness properties. Section 8 deals mainly with computational issues.

2. The statistical model and the class of M-estimators

Let $z_n = (y_n, x_n')'$ be a vector of observations on k+1 variables. A common way of modelling the relationship between y_n and x_n is to assume that

$$y_n = x_n' \beta_0 + u_n, \qquad n = 1,\dots,N, \tag{1}$$

where u_n is an unobservable disturbance, and $\beta_0 \in \mathbb{R}^k$ is a vector of unknown regression parameters. Further, it is commonly assumed that (i) the disturbances and the regressors $\{x_n\}$ are independent (or at least uncorrelated) (ii) the disturbances are independently and identically distributed (i.i.d) with zero mean and finite variance, (iii) the common distribution of the disturbances is $N(0,\sigma_0^2)$, where $\sigma_0^2 > 0$ is unknown. We shall say that model is correctly specified when the observations satisfy (1), and (i)-(iii) hold. In fact, the assumed model need not be exactly true. For example, the distribution of the disturbances may have somewhat thicker tails than the normal, there may be a few gross-errors in the data, small departures from linearity in the relationship between y and x, etc. Moreover, it may be difficult to detect exactly the nature of the model misspecification. In this kind of situations, which appear to arise frequently in empirical work, it may be sensible to consider estimators of $\theta_0 = (\beta_0, \sigma_0^2)$ that are reasonably efficient if the model is correctly specified, while being robust, that is, not too sensitive to violations of the model assumptions.

All robust estimators of β_0 that we shall discuss are M- (generalized maximum likelihood) estimators, that is, estimators defined as roots of implicit equations of the form

$$\Sigma_{n=1}^{N} \eta_N(z,\theta) = 0,$$

where the function $\eta_N: \mathbb{R}^{k+1} \times \Theta \longrightarrow \mathbb{R}^k$, with $\Theta = \{(\beta,\sigma^2): \beta\in\mathbb{R}^k, \sigma^2>0\}$, is called the score function associated with the given estimator.

It is well known that the efficiency and robustness properties of an M-estimator are closely related to the properties of its influence function (IF) [Hampel (1974)]. The IF is a measure of the asymptotic bias of an estimator when the assumed model is subject to a small amount of contamination by a point mass distribution. Given a parametric model, an M-estimator is efficient if and only if its IF is equal to the likelihood score (up to a linear transformation). On the other hand, an M-estimator is qualitatively robust [Hampel (1971)] if its IF is bounded and continuous. Qualitative robustness is a desirable property, because it ensures that small perturbations of the assumed model can have only small effects on the estimates, but does generally conflict with efficiency. For M-estimators, the sup-norm of the IF, called the estimator's sensitivity, provides a natural quantitative measure of robustness. A bounded influence (or B-robust) estimator is one with a bounded IF or, equivalently, a finite sensitivity.

The IF provides only a local approximation to the behavior of an estimator. A global measure of robustness is given by the estimator's breakdown point [Hampel (1971)], which measures the distance from the assumed model up to which the estimator still gives some relevant information. Donoho and Huber (1983) and Hampel et al. (1986) provide tractable finite-sample versions of Hampel's original asymptotic concept. The practical importance of the breakdown point has been demonstrated by Hampel (1985), who shows that the robustness properties of various location estimators in a number of simulation experiments can accurately be classified on the basis of their breakdown point alone.

All the estimators of β_0 considered in this paper are based on a score function of the form

$$\eta_N(z,\theta) = w(z,\theta) \; x \; (y - x'\beta), \tag{2}$$

where w: $\mathbb{R}^{k+1} \times \Theta \longrightarrow \mathbb{R}_+$. All these estimators can therefore be interpreted as weighted least squares (LS) estimators, and are all consistent and asymptotically normal under general conditions [Maronna and Yohai (1981)]. Although very large, this class of estimators is not completely general. In particular, it excludes some interesting (but usually computationally burdensome) estimators, such as the resistant line estimator [see e.g. Johnstone and Velleman (1985)], resistant instrumental variable estimators [Krasker and Welsch (1985)], adaptive estimators [Bickel (1982)], and high breakdown-point estimators [Siegel (1982), Rousseuw (1984), Rousseuw and Yohai (1984), and Yohai (1987)].

3. Least squares

When the weight function in (2) is identically equal to 1, we obtain the ordinary LS estimator $\hat{\beta}_{LS}$, which is also the maximum likelihood (ML) estimator under the normality assumption. Let $\hat{\beta}_N$ denote the LS estimate based on a sample of N observations, and consider adding to this sample an additonal observation z = (y,x). Let $\hat{\beta}_{N+1}$ denote the LS estimate for the new sample of N + 1 observations. Using well known results we obtain

$$\hat{\beta}_{N+1} - \hat{\beta}_N = (X'X)^{-1} x \frac{y - x'\hat{\beta}_N}{1 + x'(X'X)^{-1}x} ,$$

where X is the N×k design matrix. We assume for simplicity that the regressors are also i.i.d. with common distribution function H. Then, multiplying both sides by N and taking the limit in probability as $N \longrightarrow \infty$ under the assumption that the model is correctly specified, gives the IF of $\hat{\beta}_{LS}$ evaluated at the point z

$$IF(z,\hat{\beta}_{LS}) = [E_H vv']^{-1} x (y - x'\beta_0),$$

where E_H denotes expectations with respect to the marginal distribution of the regressors. If H is not specified, expectations can be replaced by sample averages.

The IF of $\hat{\beta}_{LS}$ can be viewed as the product of two components. One is equal to $y - x'\beta_0$, and is the 'influence of the residual'. The second, $[E_H vv']^{-1} x$, will be called the 'influence of the regressors'. The sensitivity of $\hat{\beta}_{LS}$ is not finite, because the influence of the residual is not bounded. Moreover, unless the regressors take values in a bounded set, the influence of the regressors is also unbounded. As a consequence, the LS estimator is not qualitatively robust and has zero breakdown point, that is, one large disturbance or one gross-error in the data are sufficient to completely spoil the estimates.

Let $\hat{\sigma}^2_{LS}$ denote the unbiased estimate of σ^2_0 under normality. Then, by the same argument used to derive the IF of $\hat{\beta}_{LS}$, one obtains

$$IF(z,\hat{\sigma}^2_{LS}) = (y - x'\beta_0)^2 - \sigma^2_0.$$

This is a quadratic function of the disturbance and is clearly unbounded. Thus large disturbances have an even greater effect on the LS estimates of σ_0. This can seriously affect inference based on the usual t- and F-statistics.

4. Estimators with a bounded influence of the residual

As a first step in obtaining robust estimates of β_0 one might bound the influence of the residual. Several proposals exist in the literature. However, since the influence of position remains unbounded, all these estimators are not qualitatively robust and have zero breakdown point. In particular, all can be very sensitive to gross-errors in the data.

The least absolute deviation (LAD) estimator corresponds to the choice $w(z,\theta) = 1/|y - x'\beta|$, provided that $y - x'\beta \neq 0$. The fact that $y - x'\beta$ may be equal to zero can be taken into account in practice by putting $w(z,\theta) = \max \{1/|y - x'\beta|, 1/\varepsilon\}$, where ε is a small, positive number [Fair (1974)]. The LAD estimator is also a regression quantile estimator [Koenker and Bassett (1978)], and is the ML estimator when the error distribution is Laplace (double-exponential). The influence of the residual is equal to the sign of $y - x'\beta$, and is clearly bounded but not continuous at $y - x'\beta = 0$. This implies that the LAD estimator can be very sensitive to rounding or grouping of the observations.

The Huber estimator of regression $\hat{\beta}_H$ [Huber (1973, 1981)] corresponds to the weight function $w(z,\theta) = \min \{1, c/|r|\}$, where c is a finite positive constant and $r = (y - x'\beta)/\sigma$. The score function of $\hat{\beta}_H$ can be written

$$\eta(y,x,\beta,\sigma) = \psi_c(r)\ x,$$

where the function $\psi_c(r) = \max\ \{-c,\ \min\ (c,\ r)\}$ is called the Huber function. This function plays a key role in robust statistics. It is easy to show that the influence of the residual is proportional to $\psi_c(r)$ and is bounded and continuous. The LS and LAD estimators may be viewed as limiting cases corresponding respectively to $c \longrightarrow \infty$ and $c \longrightarrow 0$. The Huber estimator can be interpreted as a method of moments estimator, obtained by equating to zero the sample covariance between the regressors and the censored residual $\psi_c(r)$. $\hat{\beta}_H$ is also the ML estimator for Huber's least informative distribution. This distribution has minimum Fisher information measure over all symmetric distributions in a given neighborhood of the normal model. It behaves like the normal in the center and like the Laplace in the tails. According to this ML interpretation, the bound c depends on σ_0 and the radius ε of the neighborhood. In particular, c should decrease as ε or σ_0 increase.

Schweppe's estimator [Handschin et al. (1975)] corresponds to the weight function $w(z,\theta) = \min\ \{1,\ c(1 - N^{-1}h(x))^{1/2}/|r|\}$, where $h(x) = x'(E_H\ vv')^{-1}\ x$ is the squared norm of x in the metric of the inverse second moment matrix of the regressors. Schweppe's estimator behaves asymptotically like the Huber estimator, because $c(1 - N^{-1}h(x))^{1/2} \longrightarrow c$ as $N \longrightarrow \infty$.

5. Optimal bounded influence estimators

We now consider estimators whose IF (and not just a part of it) is bounded and continuous. Therefore, they are all qualitatively robust and have a finite breakdown point. There is a trade-off between efficiency and robustness, because bounding the IF generally involves a loss of efficiency at the normal model. A bounded influence estimator that attains the best trade-off between efficiency and robustness is called an optimal bounded influence estimator. Formally, such an estimator has minimum asymptotic mean square error (MSE) at the normal model among all estimators with a given sensitivity [see e.g. Hampel et al. (1986)]. A whole class of estimators may be obtained by varying the MSE criterion and the metric in which the estimator's sensitivity is defined. A drawback of these estimators is the fact that their breakdown point decreases as the number of estimated parameters increases [Maronna, Bustos and Yohai (1979)].

First, we apply results of Peracchi (1987) to derive optimal bounded influence estimators of β_0. An estimator's sensitivity is defined in the metric of some $k \times k$ positive definite (p.d.) matrix B, and the same metric is used for the asymptotic MSE criterion. By varying B one obtains a whole class of estimators. The sensitivity bound is given by c. A method for choosing c will be discussed in Section 8.

An optimal bounded influence estimator of β_0, denoted by $\hat{\beta}$, is based on a weight function of the form

$$w(z,\theta) = \min \{1, c/[|r| \|A\, x\|_B]\}, \tag{3}$$

where $\|z\|_B = (z'B\, z)^{1/2}$ denotes the norm of the vector z in the metric of the p.d. matrix B. The $k \times k$ matrix A is p.d. and satisfies the implicit equation

$$E_H\ [2\ \Phi(c/\|A\ x\|_B) - 1]\ xx' - A^{-1} = 0,$$

where Φ denotes the standard normal distribution function. It can be shown that A (and hence $\hat{\beta}$) exists only if the sensitivity bound c is at least as great as $c^* = (\pi/2)^{1/2}$ (trace B)/($E_H\ \|x\|_B$), that depends on the matrix B and the spread of the x-distribution.

The score function associated with $\hat{\beta}$ can also be written as

$$\eta(y,x,\beta,\sigma) = \psi_{c(x)}(r)\ x$$

where $c(x) = c/\|A\ x\|_B$. Thus $\hat{\beta}$ can be interpreted as a method of moments estimator, obtained by equating to zero the sample covariance between the regressors and the censored residual $\psi_{c(x)}(r)$. Unlike the Huber estimator, the degree of censoring is not constant but depends on x.

Another way of representing the score associated with $\hat{\beta}$ is

$$\eta(z,\theta) = \psi_c(r/\omega(x))\ \omega(x)\ x$$

where $\omega(x) = 1/\|A\ x\|_B$ is a scalar weight that depends only on the norm of the vector of regressors. Thus, when $|r| > c\ \omega(x)$ the residual is censored and the vector of regressors is downweighted by $\omega(x)$.

By varying the matrix B, a whole class of optimal bounded influence estimators is obtained. Each estimator in this class is only optimal for one particular choice of B, and so optimality is only in a weak sense. However, all estimators in this class are admissible, that is, no estimator can be dominated by another for all choices of B. We are now in the position to establish the relationships between several bounded influence estimators proposed in the literature.

The Hampel-Krasker estimator $\hat{\beta}_{HK}$ [Hampel (1978), Krasker (1980)] is the optimal estimator when the sensitivity and the MSE are both defined in the Euclidean metric, i.e. $B = I_k$. The corresponding weight function is equal to $w(z,\theta) = \min \{1, c/[|r| \|A x\|]\}$, where the matrix A satisfies the implicit equation E_H $[2 \Phi(c/\|A x\|) - 1] xx' - A^{-1} = 0$. The necessary lower bound on c is equal to $c^* = k (\pi/2)^{1/2}/(E_H \|x\|)$ and depends on the number of regression parameters.

The Krasker-Welsch estimator $\hat{\beta}_{KW}$ [Krasker and Welsch (1982)] is the optimal estimator when the sensitivity and the MSE are both defined in the metric of the inverse asymptotic variance matrix of $\hat{\beta}_{KW}$. The corresponding weight function is equal to $w(z,\theta) = \min \{1, c/[|r|(x' Q x)^{1/2}]$, where the matrix Q satisfies the implicit equation $E\, w(z,\theta)^2\, r^2 xx' - Q^{-1} = 0$. Notice that $\hat{\beta}_{KW}$ is not optimal for an arbitrary MSE criterion, which contradicts the original claim of Krasker and Welsch.

Both the Hampel-Krasker and Krasker-Welsch estimators require solving an implicit matrix equation in order to compute the weights. This can be computationally burdensome. Peracchi (1987) proposes two bounded influence estimators based on the weight functions $w(z,\theta) = \min \{1, c/[|r|\|x\|]\}$ and $w(z,\theta) = \min \{1, c/[|r|\ h(x)^{1/2}]\}$. These estimators, denoted respectively by BI1 and BI2, are optimal under an appropriate choice of metric, and are computationally simple because they do not require solving an implicit matrix equation.

An optimal bounded influence estimator of σ_0^2 can be based on the score function

$$\chi(z,\theta) = v(z,\theta)\ (r^2 - 1)$$

where $v(z,\theta) = \min \{1, d/|a (r^2 -1)|\}$ and d is the sensitivity bound. The scalar a is a root of the equation $E_\Phi \min \{1, d/|a (r^2 - 1)|\} (r^2 - 1)^2 - a^{-1} = 0$. It can be shown that such a root exists only if the sensitivity bound d is at least as great as $1/(E_\Phi |r^2 - 1|)$. By analogy with the BI1 estimator, a considerable simplication can be obtained by choosing $v(z,\theta) = \min \{1, d/|r^2 -1|\}$.

Tests of model specification may be based on the difference between the $\hat{\beta}$ and the LS estimates. Tests of this type are likely to be quite powerful, since the difference between the two estimators can be very large when the model is misspecified, but $\hat{\beta}$ will be only slightly less efficient than LS if the model is correctly specified. Useful diagnostics for detecting influential observations and outliers can be based on the set of robust weights (3), computed for each observation in the sample. The use of these weights provides an alternative to the traditional methods based on deleting a subset of observations at a time, and then comparing the resulting estimates with the ones based on the full sample. The use of robust weights has several advantages over these methods. Robust weights are jointly computed with the parameter estimates and require no additional calculation, and are easy to interpret, because of the weighted LS nature of an optimal bounded influence estimator.

There have been some interesting applications of optimal bounded influence estimators in empirical econometrics. For example, Krasker, Kuh and Welsch (1983) and Small (1986) estimate hedonic price models for housing. Swartz and Welsch (1986) estimate and forecast energy demand. Thomas (1987) uses a very large data set to estimate Engel curves for food. All these studies indicate that these estimators can lead to significant differences with respect to LS in terms of point estimates, inference and forecasts. This is mainly due to the fact that they are much less sensitive than LS to outliers and aberrant observations. These studies also demonstrate how robust weights can be used as an effective diagnostic tool.

6. Mallows estimators

Optimal bounded-influence estimators place an overall bound on the norm of the IF. Mallows (1975) proposed a class of bounded influence estimators with separate bounds on the influence of the residual and the influence of position. When both bounds are finite, Mallows estimators are also qualitatively robust and have a finite breakdown point that decreases as the number of estimated parameters increases. The weight function of these estimators takes the form $w(z,\theta) = \min\{1, c/|r|\}\ \omega(x)$, for some weight function $\omega(x)$ depending only on the regressors. When $\omega(x)$ is identically equal to 1 we obtain the Huber estimator. The score function associated with these estimators is of the form

$$\eta(z,\theta) = \psi_c(r)\ \omega(x)\ x.$$

The optimal estimators in this class (in the asymptotic MSE sense) correspond to the choice $\omega(x) = \min\{1, \gamma/\|A\, x\|_B\}$ for some matrix B and a finite positive constant γ [Peracchi (1987)]. The matrix A satisfies the implicit equation $E_H \min\{1, \gamma/\|A\, x\|_B\}\ xx' - A^{-1} = 0$, which implies a necessary lower bound on γ given by $\gamma^* = (\text{trace } B)/(E_H\ \|x\|_B)$. When $B = I_k$ we obtain the Mallows-Hampel estimator [Hampel (1978)]. When $\omega(x) = \min\{1, \gamma/(x'Q\, x)^{1/2}\}$, where Q solves the equation $E_H \min\{1, \gamma/(x'Q\, x)^{1/2}\}^2$

$xx' - Q^{-1} = 0$, we obtain the Mallows-Maronna estimator [Maronna and Yohai (1981)]. When $\omega(x) = \min \{1, \gamma/\|x\|\}$ and $\omega(x) = \min \{1, \gamma/h(x)^{1/2}\}$ we obtain the M1 and M2 estimator proposed by Peracchi (1987). These two estimators are analogues of the BI1 and BI2 estimators, and are computationally convenient because they do not require solving an implicit matrix equation.

7. Redescending estimators

For all robust estimators discussed so far the influence of large residual is finite and bounded away from zero. However, for some distributions with heavy tails, such as Sudent's t, the likelihood score tends to zero for large values of the disturbances. This observation suggests that estimators with good robustness properties against thick-tailed distributions may be obtained by requiring the influence of the residual to go to zero, or become zero, for large values of the residual. Estimators with this property are called 'redescending'. Examples include Hampel's (1974) three-part redescending estimator, Andrews's (1974) sine estimator, Tukey's biweight estimator [Beaton and Tukey (1974)] and the hyperbolic tangent estimator [Hampel, Rousseeuw and Ronchetti (1981), Ronchetti and Rousseeuw (1985)]. The available evidence indicates that redescending estimators are very robust but can be computationally quite burdensome because of non-uniqueness problems.

8. Computational aspects

Computation of optimal bounded influence estimators of θ_0 is relatively simple, because the matrix A and the constant a do not depend on θ but only on the distribution functions Φ and H. Moreover, in the case of the BI1 estimator one can simply put $(A, a) = (I_k, 1)$. Given (A, a), optimal bounded influence estimates $\hat{\theta} = (\hat{\beta}, \hat{\sigma}^2)$ can be obtained by an iterative scheme whose i-th iteration is of the form:

$$\hat{\beta}^{(i+1)} = [\Sigma_{n=1}^{N} w_n^{(i)} x_n x_n']^{-1} \Sigma_{n=n}^{N} w_n^{(i)} x_n y_n$$

$$[\hat{\sigma}^{(i+1)}]^2 = [\Sigma_{n=1}^{N} v_n^{(i)}]^{-1} \Sigma_{n=1}^{N} v_n^{(i)} (y_n - x_n'\hat{\beta}^{(i)})^2, \qquad (4)$$

where $w_n^{(i)} = \min\{1, c/[|r_n^{(i)}| \|A x_n\|_B]\}$, $v_n^{(i)} = \min\{1, d/|a([r_n^{(i)}]^2 - 1)|\}$ and $r_n^{(i)} = (y_n - x_n'\hat{\beta}^{(i)})/\hat{\sigma}^{(i)}$. This iterative scheme can easily be implemented on most LS packages. Good starting values for the iterations may be given by the Huber or LAD estimates.

In some cases the investigator may prefer, for breakdown-point reasons, to use a very robust estimator of scale rather than an optimal one. One possibility in these cases is to use the median absolute deviation estimator. This amounts to replacing (4) with

$$\hat{\sigma}^{(i+1)} = [\Phi^{-1}(3/4]^{-1} \text{ med } \{|y_n - x_n'\hat{\beta}^{(i)}|, n = 1,\ldots,N\}$$

where the factor $[\Phi^{-1}(3/4)]^{-1}$ is introduced in order to obtain asymptotic unbiasedness at the normal model [see e.g. Hampel et al. (1986)]. Another possibility is to follow Huber's (1964) Proposal 2, and derive the estimates of θ_0 by solving the problem

$$\text{Min}_{\theta\in\Theta}\ Q(\theta) = \sigma\ \Sigma_{n=1}^{N}\ [\rho_{c(x_n)}((y_n - x_n'\beta)/\sigma) + \nu], \tag{5}$$

where the choice of ν guarantees asymptotic unbiasedness of the estimator of scale at the normal model, and ρ_c is a convex function defined as

$$\rho_{c(x)}(r) = \begin{cases} r^2/2 & \text{if } |r| \le c(x) \\ |r|c(x) - c(x)^2/2 & \text{otherwise.} \end{cases}$$

The first-order conditions for $(\hat{\beta}, \hat{\sigma}^2)$ are of the form

$$\Sigma_{n=1}^{N}\ \psi_{c(x_n)}((y_n - x_n'\hat{\beta})/\hat{\sigma})\ x_n = 0$$

$$\Sigma_{n=1}\ \chi_{c(x_n)}((y_n - x_n'\hat{\beta})/\hat{\sigma}) - N\ \nu = 0$$

where $\chi_c(r) = \psi_c(r)\ r - \rho_c(r) = \psi_c(r)^2/2$. Thus, $\nu = E_H\ [E_\Phi\ \chi_{c(x)}(r)]$. Since ψ_c is an odd function, that is $\psi_c(-r) = -\ \psi_c(r)$, it follows from the results of Andrews (1986), that the exact distribution of $\hat{\beta}$ is symmetric about β_0 provided that the distribution of the disturbances is symmetric about zero. It can also be shown that $\hat{\beta}$ is consistent and asymptotically normal under general conditions [Maronna and Yohai (1981)]. Furthermore, $\hat{\beta}$ and $\hat{\sigma}^2$ are asymptotically independent. The asymptotic variance matrix of $\hat{\beta}$ is of the form $AV(\hat{\beta}) = P^{-1}\ Q\ (P')^{-1}$, where

$$P = E\ [(\partial/\partial\beta')\ \eta(z,\theta_0)]$$

$$Q = E\ [\eta(z,\theta_0)\ \eta(z,\theta_0)'],$$

and can be estimated consistently as suggested in White (1982). When the disturbances are i.i.d. normal, it is easy to see that $P = E_H\ [2\ \Phi(c(x)) - 1]\ xx'$ and $Q = \sigma_0^{\ 2}\ P + 2\sigma_0^{\ 2}\ E_H\ [c(x)^2\ (1 - \Phi(c(x))) - c(x)\ \phi(c(x))]\ xx'$, where $\phi(\cdot)$ denotes the $N(0,1)$ density. Using these expressions, however, leads to covariance estimates that are inconsistent under heteroskedasticity or non-normality.

All regression estimators discussed above can be obtained by suitable restrictions on the function $Q(\theta)$. If $c(x) = \infty$, $\forall$ x, we obtain the LS estimator. In this case $\hat{\sigma}^2$ is equal to the mean squared deviation of the LS residuals. If $c(x) = c$, $\forall$ x, we obtain the Huber estimator. The choice $c(x) = c\ (1 - N^{-1}\ h(x))^{1/2}$, gives Schweppe's original estimator. If $c(x) = c/\|A\ x\|$ for some matrix A we obtain the class of optimal bounded-influence estimators. When $c(x) = c$ and observations are transformed by multiplying by $\omega(x)^{1/2}$, we obtain the class of Mallows' estimators. Finally, by suitably modifying the function $Q(\theta)$ one obtains the class of redescending estimators.

Minimization of $Q(\theta)$ may be carried out by gradient methods. Alternatively, one may solve the first order conditions by iteratively reweighted LS or a Newton-Raphson type algorithm. The FORTRAN subroutine library ROBETH [Marazzi (1980)] and the conversational package ROBSYS [Marazzi and Randriamiharisoa (1986)] are also available.

We conclude by presenting a simple procedure for choosing the sensitivity bound c for an estimator $\hat{\beta}$ obtained by solving (5). Consider any of the commonly used meausures of the asymptotic relative efficiency (ARE) of two estimators. Under i.i.d. normal errors, the ARE of $\hat{\beta}$ with respect to the LS estimator depends only on the sensitivity bound c and the distribution of the regressors, and is a strictly increasing function of c. Given the distribution of the regressors, choosing a sensitivity bound is therefore equivalent to choosing a level of ARE. The choice problem is therefore straightforward given the econometrician's preferences with respect to efficiency and protection against bias. For Mallows estimators one needs to choose separate bounds on the influence of the residual and the influence of the regressors. If one wants to attain an ARE of, say, .95 at the normal model, the above procedure may be modified as follows. One can first choose the bound on the influence of the residual so as to attain an ARE not exceeding, say, .975. One can then choose the bound on the influence of position so as to attain the desired ARE of 95%.

References

Andrews, D.F. (1974), "A robust method for multiple linear regression", *Technometrics*, 16, 523-531.

Andrews, D.W.K. (1986), "A note on the unbiasedness of feasible GLS, quasi-maximum likelihood, robust, adaptive, and spectral estimators of the linear model", *Econometrica*, 54, 687-698.

Beaton, A.E. and Tukey, J.W. (1974), "The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data", *Technometrics*, 16, 147-185.

Bickel, P.J. (1982), "On adaptive estimation", *Annals of Statistics*, 10, 647-671.

Donoho, D.L. and Huber, P.J. (1983), "The notion of breakdown point", in P.J. Bickel, K.A. Doksum and J.L. Hodges (eds.), *Festschrift for Erich L. Lehmann*, Wadsworth, Belmont, Calif.

Fair, R.C. (1974), "On the robust estimation of econometric models", *Annals of Economic and Social Measurement*, 3, 667-677.

Hampel, F.R. (1971), "A general qualitative definition of robustness", *Annals of Mathematical Statistics*, 42, 1887-1096.

Hampel, F.R. (1974), "The influence curve and its role in robust estimation", *Journal of the American Statistical Association*, 69, 383-393.

Hampel, F.R. (1978), "Optimally bounding the gross-error-sensitivity and the influence of position in factor space", *Proceedings of the ASA Statistical Computing Section*, ASA, Washington, D.C.

Hampel, F.R. (1985), "The breakdown points of the mean combined with some rejection rules", *Technometrics*, 27, 95-107.

Hampel, F.R., Rousseeuw, P.J. and Ronchetti, E. (1981), "The change-of-variance curve and optimal redescending M-estimators", *Journal of the American Statistical Association*, 76, 643-648.

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

Handschin, E., Schweppe, F.C., Kohlas, J., and Fiechter, A. (1975), "Bad data analysis for power system state estimation", *IEEE Transactions*, PAS-94, 329-337.

Huber, P.J. (1973), "Robust regression: Asymptotics, conjectures and Monte-Carlo", *Annals of Statistics*, 1, 799-821.

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

Koenker, R. and Bassett, G. (1978), "Regression quantiles", *Econometrica*, 46, 33-50.

Krasker, W.S. (1980), "Estimation in linear regression models with disparate data points", *Econometrica*, 48, 1333-1346.

Krasker, W.S. and Welsch, R.E. (1982), "Efficient bounded-influence regression estimation", *Journal of the American Statistical Association*, 77, 595-604.

Krasker, W.S. and Welsch, R.E. (1985), "Resistant estimation for simultaneous-equations models using weighted instrumental variables", *Econometrica*, 53, 1475-1488.

Krasker, W.S., Kuh, E., and Welsch, R.E. (1983), "Estimation for dirty data and flawed models", in Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, Vol. 1, North-Holland, Amsterdam.

Johnstone, I.M. and Velleman, P.F. (1985), "The resistant line and related regression methods", *Journal of the American Statistical Association*, 80, 1041-1054.

Mallows, C.L. (1975), "On some topics in robustness", Technical Memorandum, Bell Telephone Laboratories, Murray Hill, N.J.

Marazzi, A. (1980), "ROBETH, a subroutine library for robust statistical procedures", *COMPSTAT 1980, Proceedings in Computational Statistics*, Physica, Vienna.

Marazzi, A. and Randriamiharisoa, A. (1986), "ROBSYS: A package for robust statistical computing", in R. Salamon, B.I. Blum and M.J. Jorgensen (eds.), *MEDINFO 86*, North Holland, Amsterdam.

Maronna, R.A., Bustos, O.H., and Yohai, V.J. (1979), "Bias- and efficiency-robustness of general M-estimators for regression with random carriers", in T. Gasser and M. Rosenblatt (eds), *Smoothing Techniques for Curve estimation*, Lecture Notes in Mathematics No. 757, Springer, Berlin.

Maronna, R. and Yohai, V. (1981), " Asymptotic behavior of general M-estimates for regression and scale with random carriers", *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 58, 7-20.

Peracchi, F. (1987), Bounded Influence Methods in Econometrics with an Application to the Censored Regression Model", Ph.D. Thesis, Princeton University.

Ronchetti, E., and Rousseeuw, P.J. (1985), "Change-of-variance sensitivities in regression analysis", *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 68, 503-519.

Rousseeuw, P.J. (1984), "Least median of squares regression", *Journal of the American Statistical Association*, 79, 871-880.

Rousseeuw, P. J. and V Yohai, (1984), "Robust regression by means of S-estimators", in J. Franke, W. Härdle and R. D. Martin, (eds.), *Robust and Nonlinear Time Series Analysis*, Lecture Notes in Statistics, 26, Springer, New York.

Siegel, A.F. (1982), "Robust regression using repeated medians", *Biometrika*, 69, 242-244.

Small, K. (1986), "Effects of the 1979 gasoline shortages on Philadephia housing prices", *Journal of Urban Economics*, 19, 371-381.

Swartz, S. and Welsch, R.E. (1986), "Applications of bounded-influence and diagnostic methods in energy modeling", in D.A. Belsley and E. Kuh (eds.), *Model Reliability*, MIT Press, Cambridge, Mass.

Thomas, D. (1987), "Robust regression of the food share Engel curve", Yale University, mimeo.

White, H. (1982), "Maximum likelihood estimation of misspecified models", *Econometrica*, 50, 1-25.

Yohai, V.J. (1987), "High breakdown-point and high efficiency robust estimates for regression", *Annals of Statistics*, 15, 642-656.