Conditionally Unbiased Bounded Influence Robust Regression, with Applications to Generalized Linear Models

H.R. Kunsch[1]

L.A. Stefanski[2]

R.J. Carroll[3]

[1]Seminar fuer Statistik, ETH-Zentrum, CH-8092, Zurich, Switzerland.

[2]Department of Statistics, North Carolina State University, Raleigh, NC 27695 (USA). Research supported by the National Science Foundation.

[3]Department of Statistics, University of North Carolina, Chapel Hill, NC 27514 (USA). Research supported by the Air Force Office of Scientific Research.

ABSTRACT

We propose a class of bounded influence robust regression estimators with conditionally unbiased estimating functions given the design. Optimal estimators are found within this class. Applications are made to generalized linear models. An example applying logistic regression to food stamp data is discussed.

Key Words and Phrases : Robust regression, Bounded Influence, Asymptotic Bias Generalized Linear Models, Linear Regression.

1. INTRODUCTION

In this paper we study robust estimation in a model with explanatory variables X : (X_i, Y_i) are assumed to be independent and identically distributed with joint distribution $P_\theta(dy|x)F(dx)$. We consider M-estimators defined implicitely by the equation

$$\sum_{i=1}^{n} \psi(Y_i, X_i, \hat{\theta}_n) = 0. \tag{1.1}$$

In (1.1), θ belongs to a subset Ω of $\mathbb{R}^p$ and ψ takes values in $\mathbb{R}^p$. Usually, ψ is required to be Fisher-consistent, i.e.,

$$\iint \psi(y,x,\theta)\ P_\theta(dy|x)F(dx) = 0 \qquad \text{for all } \theta, \tag{1.2}$$

which implies consistency in the usual sense under weak regularity conditions. In this paper we require a stronger form of (1.2):

$$\int \psi(y,x,\theta)P_\theta(dy|x) = 0 \quad \text{for all } x \text{ and all } \theta, \tag{1.3}$$

which we will call <u>conditional Fisher-consistency</u>.

In the linear model with symmetric errors, essentially all Fisher-consistent estimators which are optimal in some sense automatically satisfy (1.3). This is not the case with asymmetric errors or for generalized linear models. Stefanski, et al. (1986) have investigated robust estimators satisfying (1.2) (with F the empirical distribution function of the $\{X_i\}$) but not (1.3). However, conditional Fisher consistency is an appealing concept because it does not involve the distribution of the explanatory variables $\{X_i\}$, which is independent of the parameter of interest. Moreover, these estimators have the advantages of being computationally simpler in certain cases (Section 3) and less affected by the estimation of nuisance parameters (Section 4).

Recall some general results and definitions from robust statistics (see Hampel et. al., 1986). The influence function of an M-estimator is

$$IC_\psi(y,x,\theta) = D_\psi^{-1}(\theta)\ \psi(y,x,\theta); \tag{1.4}$$

where $$D_\psi(\theta) = -\frac{\partial}{\partial\beta} \iint \psi(y,x,\beta)\ P_\theta(dy|x)F(dx)\Big|_{\beta=\theta}. \qquad (1.5)$$

Under regularity conditions, $N^{1/2}$ $(\hat{\theta}_N - \theta)$ is asymptotically normal with covariance matrix

$$D_\psi(\theta)^{-1}\ W_\psi(\theta)\ D_\psi(\theta)^{-T} = V(\psi) = V(\psi,\theta), \qquad (1.6)$$

where $$W_\psi(\theta) = E[\psi(\theta)\psi(\theta)^T]\ . \qquad (1.7)$$

Finally we consider the self-standardized influence

$$s(\psi)^2 = \sup_{y,x}\ \sup_{\lambda\neq 0} \frac{(\lambda^T IC_\psi)^2}{\lambda^T V(\psi)\lambda} = \sup_{y,x}\ \psi^T\ W_\psi^{-1}\psi \qquad (1.8)$$

which measures the maximal influence an observation can have on a linear combination of interest standardized by the standard deviation of this linear combination. Integrating (1.8) shows that $s(\psi)^2 \geq p$. For other measures of influence, see for example Giltinan, et. al (1986). The usual goal of bounded influence estimation is to minimize the asymptotic covariance (1.6) subject to a bound on the measure of influence, in this case (1.8). As in Giltinan, et. al (1986), optimal estimators satisfying this goal depend on the measure of influence.

2. LINEAR REGRESSION WITH ASYMMETRIC ERRORS

The purpose of this section is to show that the differences between (1.2) and (1.3) are relevant also in linear regression if the errors are not symmetric. We consider the model

$$Y_i = \theta_0 + x_i^T\theta + u_i, \qquad (2.1)$$

where $\{u_i\}$ are independent and identically distributed with common density g having no point of symmetry. The usual design matrix is assumed to be of full rank. The vector θ is of length (p-1). The most important proposals for

M-estimators in linear regression are the Mallows and Schweppe forms:

$$\psi_u(y,x,\theta) = w(x)\ \varphi(y-\theta_0-x^T\theta)(1\ x^T)^T; \tag{2.2}$$

$$\psi_s(y,x,\theta) = w(x)\ \varphi((y-\theta_0-x^T\theta)\ /\ w(x))(1\ x^T)^T. \tag{2.3}$$

Here w(x) is a scalar weight function and φ is a scalar function. Without additional assumptions on g, θ_0 is not identifiable, but the following result gives conditions for consistency of θ. For related results, see Carroll (1979).

THEOREM 2.1 :

i) For the Mallows-form (2.2), there is a constant τ such that $\psi(y,x,\theta_0+\tau,\theta_1,\dots,\theta_p)$ satisfies (1.3).

ii) If the x_i are symmetrically distributed with center $c \in \mathbb{R}^p$ and if $w(X-c) \equiv W(-X+c)$, then there is a constant τ such that for the Schweppe form (2.3) $\psi(y,x,\theta_0+\tau,\theta_1,\dots,\theta_p)$ satisfies (1.2), but in general not (1.3).

PROOF: i) is obvious by defining τ as the solution of $\int \psi(u-\tau)g(u)du = 0$. In case ii) define τ as the solution of

$$\iint w(x)\ \varphi((u-\tau)/w(x))\ g(u)\ F(dx)du = 0.$$

This is just (1.2) for the first component of ψ. For the other components we observe that

$$\iint x_j\ w(x)\ \varphi((u-\tau)/w(x))\ g(u)\ F(dx)du =$$
$$\iint (x_j-c_j)w(x)\ \varphi((u-\tau)/w(x))\ g(u)\ F(dx)du = 0$$

by symmetry, so (1.2) holds. Equation (1.3) does not hold in general, because the solution τ of

$$\int \varphi((u-\tau)/w(x))\ g(u)du = 0$$

depends in most cases on x. □

Thus, the Mallows form has the advantage of being a conditionally unbiased estimating equation even when the $\{x_i\}$ or $\{u_i\}$ are not symmetrically distributed.

3. GENERALIZED LINEAR MODELS

We consider the canonical form of generalized linear models, where

$$F_\theta(dy|x) = \exp\{y\, x^T\theta - G(x^T\theta) - s(y)\}\ \mu(dy). \tag{3.1}$$

If g is the derivative of G, the likelihood score function is

$$\ell(y,x,\theta) = \{y-g(x^T\theta)\}x. \tag{3.2}$$

Note that (3.2) satisfies (1.3), so that the score is conditionally unbiased. Because ℓ is proportional to x, the influence is unbounded, i.e., $s^2(\ell) = \infty$.

We are looking here for M-estimators satisfying (1.3) and $s(\psi) \leq b$ which minimize $V(\psi)$ in some sense. In analogy with a general principle for constructing optimal robust estimators (Hampel, et. al (1986), Section 4.3), we consider the following ψ function:

$$\psi_{cond}(y,x,\theta,B) = d(y,x,\theta,B)\ w_b(|d(y,x,\theta,B)|(x^TB^{-1}x)^{\frac{1}{2}})\ x, \tag{3.3}$$

where

$$d(y,x,\theta,B) = y-g(x^T\theta) - c(x^T\theta,\ b/(x^T\ B^{-1}x)^{\frac{1}{2}})$$

and

$$w_b(a) = H_b(a)/a,$$

where H_b is the Huber function $H_b(a) = \min(a,b)$ $(a \geq 0)$.

We work within the context of the Schweppe-form, although related results are obtainable for the Mallows-form as in Stefanski, et al (1986). The major change is that w_b in (3.3) factors into two parts. The first depends only on x and is of the form $w_1((x^TB^{-1}x)^{1/2})$. The other depends only on

$$d(y,x,\theta) = y - g(x^T\theta) - c(x^T\theta, b/(x^TB^{-1}x)^{1/2})$$

and is of the form $w_2(|d(y,x,\theta)|)$.

The function c and the matrix B in (3.3) will be chosen so that the side conditions (1.3) and $s(\psi_{cond}) = b$ are satisfied. By the definition of ψ_{cond}, (1.3) holds if and only if for all β and all $a > 0$,

$$\int (y-g(\beta)-c(\beta,a))\ w_a(|y-g(\beta)-c(\beta,a)|)\exp(y\ \beta-G(\beta)-S(y))\ \mu(dy) = 0. \tag{3.4}$$

First we discuss the existence of a solution to (3.4).

<u>LEMMA 3.1</u> For any $a > 0$ and β, there is a solution $c = c(\beta,a)$ to (3.4).

<u>PROOF</u>: For fixed y, β, a, the function

$$c \longrightarrow (y - g(\beta) - c)\ w_a(|y - g(\beta) - c|)$$

is continuous, bounded and monotone nonincreasing with limits $\pm$ a. Hence the existence follows from dominated convergence and the intermediate value theorem.

□

A practical advantage here is that often the function c can be calculated in closed or almost close form. This is particularly important compared to (1.2), where c is a vector (depending only on θ) whose computation is quite difficult, see Stefanski, et. al (1986), Section 2.4. Here are two examples where $c(\beta,a)$ can be calculated explicitly.

<u>Example 3.1</u> : <u>Logistic Regression</u> Here μ puts equal mass at 0 and 1, $S(y)=0$, and $G(\beta) = \ell n\{1 + \exp(\beta)\}$. Write $p = \exp(\beta)/(1 + \exp(\beta))$ and $q = 1-p$. It is easily checked that

$$c(\beta,a) = \begin{cases} ap/q - p & \text{if } \beta < 0 \text{ and } a < q \\ q - aq/p & \text{if } \beta > 0 \text{ and } a < p \\ 0 & \text{otherwise} \end{cases}$$

satisfies (3.4).

<u>Example 3.2</u> : <u>Negative Exponential Regression</u> Here μ is Lebesgue-measure on $[0,\infty)$, $G(\beta) = -\ell n(-\beta)$, $s(y) = 0$ and $\beta < 0$. Two cases occur.

If the bound a is large, the Huberization in ψ_{cond} is one-sided (for large y's only), for small a's both large and small y's will be Huberized. It can be checked by straightforward calculations that the cutting point between the two cases is given by the equation $e^{2\beta a} = 1 + \beta a$, so that $\beta a \approx -0.797$. In the former case $c(\beta,a) = -\beta^{-1}$ times the smaller solution of $\exp(x+\beta a-1) = x$ and in the latter case $c(\beta,a) = -\beta^{-1}(1 + \log(\beta a/(\exp(\beta a) - \exp(-\beta a))))$.

Turn now to the matrix B. It follows from the definition of ψ_{cond} that $s(\psi_{cond}) = b$ provided

$$E_\theta[\psi_{cond}(y,x,\theta,B)\ \psi_{cond}(y,x,\theta,B)^T] = B. \tag{3.5}$$

Equation (3.5) is used to define $B=B(\theta)$. $B(\theta)$ depends also on the distribution F of the explanatory variables. A necessary condition for (3.5) is $b \geq p$, but we do not know if it is also sufficient.

We have the following optimality result for ψ_{cond}.

<u>THEOREM 3.1</u> : Suppose that for a given b (3.5) has a solution $B(\theta)$. Then ψ_{cond} minimizes $tr(V(\psi)V(\psi_{cond})^{-1})$ among all ψ satisfying (1.3) and

$$\sup_{(y,x)} IC_\psi\ V(\psi_{cond})^{-1}\ IC_\psi \leq b^2.$$

Theorem 3.1 is a corollary of the following analogy to Theorem 1 of Stefanski, et al. (1986). Note that Theorem 3.2 below also applies to any kind of model with explanatory variables.

<u>THEOREM 3.2</u> : Let $\ell(y,x,\theta)$ be the likelihood score function. Define the score function

$$\psi_{cond}(y,x,\theta) = (\ell-c)\min^{\frac{1}{2}}(1,b^2/\{(\ell-c)^T B^{-1}(\ell-c)\}), \tag{3.6}$$

where $c = c(x,\theta)$ and $B = B(\theta)$ are assumed to exist and satisfy

$$E(\psi_{cond}(y,x,\theta)\,|x) = 0$$

$$E\{\psi_{cond}(y,x,\theta)\ \psi_{cond}(y,x,\theta)^T\} = B.$$

Then (3.6) minimizes $tr(V(\psi)V(\psi_{cond})^{-1})$ among all ψ satisfying (1.3) and

$$\sup_{(y,x)} IC_\psi \, V(\psi_{cond})^{-1} \, IC_\psi \le b^2.$$

With the exception of multiplication by a constant matrix, ψ_{cond} is unique almost surely. □

PROOF OF THEOREM 3.2 The proof is almost identical to tha of Theorem 1 in Stefanski, et al. (1986), once one notes that for any conditionally unbiased score function ψ,

$$E \; c(x,\theta) \; \psi(y,x,\theta)$$

$$= E\{c(x,\theta) \; E(\psi(y,x,\theta)|x)\} = 0. \qquad □$$

The computational complexity of the conditional unbiased estimator is not particular to the model (3.1). For instance, if we have a generalized linear model with arbitrary link function h, we have to replace in (3.3) $d(y,x,\theta,B)$ by

$$h'(x^T\theta)\{y - g(h(x^T\theta)) - c(h(x^T\theta),b/((x^TB^{-1}x)^{1/2} \mid h'(x^T\theta)|))\},$$

where $c(\beta,a)$ is still defined by (3.4).

In applications, the distribution F of the $\{x_i\}$ is unknown. It is common to replace F by its empirical distribution. From (3.3) and (3.5), this means that we solve

$$\sum_{i=1}^{N} \psi_{cond}(y_i,x_i,\hat{\theta}_N,\hat{B}_N) = 0, \tag{3.7}$$

$$N^{-1} \sum_{i=1}^{N} x_i x_i^T \, v(x_i^T \hat{\theta}_N, \; b/(x_i^T \hat{B}_N^{-1} x_i)^{1/2}) = \hat{B}_N, \tag{3.8}$$

where

$$v(\beta,a) = \int (y-g(\beta)-c(\beta,a))^2 \, w^2(y,\beta,a)\exp(y\beta-G(\beta)-S(y)) \, \mu(dy), \tag{3.9}$$

$$w(y,\beta,a) = H_a(|y-g(\beta) - c(\beta,a)|) \, / \, |y-g(\beta) - c(\beta,a)| \tag{3.10}$$

In many applications, improved protection against outliers through higher breakdown points can be achieved by the use of redescending ψ functions, see Rousseeuw (1984). In equations (3.3), (3.4) and (3.10) the Huber function H_b could be replaced by any of the redescenders such as Hampel's three part function ψ or the Tukey biweight. The calculation of $c(\beta,a)$ is of the same complexity as with the Huber function. The breakdown properties of such estimates remain to be studied.

4 : THE EFFECT OF ESTIMATING THE MATRIX B

In the last section we derived the estimator defined by (3.7)-(3.8) as an approximation to the optimal estimator which uses $\psi_{cond}(y,x,\theta,B(\theta))$. We may consider (3.7) and (3.8) as an M-estimator for both θ and a nuisance parameter B. The ψ-function defining this M-estimator is

$$(\psi_{cond}(y,x,\theta,B)^T,\ \chi(x,\theta,B)^T)^T \text{ , where}$$
$$\chi(x,\theta,B) = x\,x^T\, v(x^T\theta,\ b/(x^TB^{-1}x)^{1/2}) - B.$$

The influence function of this estimator is (compare (1.4) and (1.5))

$$IC_{\psi,x}(y,x,\theta,B) = D^{-1}_{\psi,x}(\theta)\ (\psi_{cond}(y,x,\theta,B)^T,\ \chi(x,\theta,B)^T)^T, \tag{4.1}$$

where

$$D_{\psi,x} = \begin{bmatrix} -\frac{\partial}{\partial\beta} E_\theta\left[\psi_{cond}(y,x,\beta,B)\right]\Big|_{\beta=\theta} & -\frac{\partial}{\partial A} E_\theta\left[\psi_{cond}(y,x,\theta,A)\right]\Big|_{A=B(\theta)} \\ -\frac{\partial}{\partial\beta} E_\theta\left[\chi(x,\beta,B)\right]\Big|_{\beta=\theta} & -\frac{\partial}{\partial A} E_\theta\left[\chi(x,\theta,A)\right]\Big|_{A=B(\theta)} \end{bmatrix} \tag{4.2}$$

By the definition of ψ_{cond} and $c(\beta,a)$ in (3.3) and (3.4), $\psi_{cond}(y,x,\theta,A)$ satisfies (1.3) for arbitrary A. Hence, $E_\theta[\psi_{cond}(y,x,\theta,A)] = 0$ for all A and the upper right block of $D_{\psi,x}$ is zero. This means that the θ part of the influence function for (3.7) and (3.8) is equal to

$$\{-\frac{\partial}{\partial\beta} E_\theta[\psi_{cond}(y,x,\beta,B)]\ \Big|_{\beta=\theta}\}^{-1}\ \psi_{cond}(y,x,\theta,B). \tag{4.3}$$

On the other hand, the influence function for the optimal $\psi_0(y,x,\theta) = \psi_{cond}(y,x,\theta,B(\theta))$ is also equal to (4.3) because by the same argument,

$$D_{\psi_0}(\theta) = -\frac{\partial}{\partial\beta} E_\theta[\psi_{cond}(y,x,\beta,B)] \Big|_{\beta=\theta} - \frac{\partial}{\partial A} E_\theta[\psi_{cond}(y,x,\beta,A)] \Big|_{A=B} \frac{\partial}{\partial\theta} B(\theta)$$

$$= -\frac{\partial}{\partial\beta} E_\theta[\psi_{cond}(y,x,\beta,B)] \Big|_{\beta=\theta} .$$

We have thus shown

THEOREM 4.1 : The θ part of the influence function in the case that θ and B are simultaneously estimated by (3.7) and (3.8) is the same as the influence function in the case that θ alone is estimated using the optimal $\psi_0(y,x,\theta) = \psi_{cond}(y,x,\theta,B(\theta))$. In particular, the asymptotic covariance matrix of $\hat{\theta}_N$ is the same in both cases. □

REMARKS :

i) $\hat{\theta}_n$ and $\hat{B}_n$ are not asymptotically independent: $E_\theta[\psi_{cond}\, \chi^T] = 0$ by (1.3), but $\frac{\partial}{\partial\beta} E_\theta[\chi(x,\beta,B)]\Big|_{\beta=\theta} \neq 0$ in general.

ii) Because in linear regression with symmetric errors χ does not depend on θ, an analogue to Theorem 4.1 is obvious. In addition, estimation of the scale of the errors does not change the asymptotic covariance either, and $\hat{\theta}_n$ is asymptotically independent of all nuisance parameters.

iii) From the finite sample interpretation of the influence function, (4.3) means the following: to the first order of approximation the change in $\hat{\theta}_n$ caused by adding or deleting an observation at (x,y) is

$$-N\left\{N^{-1} \sum_{i=1}^{N} \frac{\partial}{\partial\beta} E_{\hat{\theta}_N}[\psi_{cond}(y,x_i,\beta,\hat{B}_N) | x_i] \Big|_{\beta=\hat{\theta}_N}\right\}^{-1} \psi_{cond}(y,x,\hat{\theta}_N,\hat{B}_N),$$

i.e. the change in $\hat{B}_N$ has approximately no effect on the change in $\hat{\theta}_N$. In this sense the estimator (3.7)-(3.8) is reasonably stable.

(iv) For the Fisher-consistent estimator (2.12)-(2.13) of Stefanski, et al. (1986), there is no analogue of Theorem 4.1. The θ part of the influence function is in general a linear combination of ψ_{BI}, $E_\theta[\psi_{BI} \mid x]$ and $E_\theta[\psi_{BI}\psi_{BI}^T | x]-B$ because all blocks in D are in general different from zero.

5: EXAMPLES

We illustrate the conditional Fisher-consistent estimators on two examples of logistic regression, the first considered by Stefanski, et al. (1986).

Example 5.1 : The Food Stamp Data

The response is whether or not one participates in the federal food stamp program. Two dichotomous predictors are available, tenancy and supplemental income. An additional predictor is log(1 + monthly income). There were 150 observations, with 24 participating in the program. One observation, #5, is known to be highly influential on the ordinary analysis, with another, #66, slightly less influential. We used the Huber ψ function with two values of b, and in one case contrasted the use of the conditionally unbiased score function and the biased score function with $c(\beta,a) \equiv 0$. The latter choice was used by Stefanski, et al. (1986) because of numerical problems in enforcing unconditional Fisher consistency (1.2). In Table 1, we list the results of this analysis. We also list the analysis one would obtain if one used a Hampel ψ function with bend points (6,14,32), see Hampel, et al. (1986, pages 66-67). For computational convenience, rather than solving (3.4), in defining the function $c(\beta,a)$ we have chosen to use the same formula as in Example 3.1, which should not affect the results too severely.

In this example, the major difference is not among the biased and conditional Fisher-consistent estimates, but rather among the choices of b (maximum likelihood corresponds to $b = \infty$). With decreasing b the importance of supplemental income as well as the weights for cases #5 and #66 decrease and the importance of log(1+monthly income) increases.

Example 5.2 : Skin Vaso-constriction data

These data have appeared elsewhere in the context of robustness, see Pregibon (1981). The response is the occurrence of vaso-constriction in the skin of the digits and is regressed on the logarithms of the rate and volume of air inspired. In our work, we took $Rate_{32} = 0.30$, see Pregibon (1981).

This data set is inherently unstable. As previous authors have shown, once one deletes observations #4 and #18, almost perfect discrimination is possible. Our analyses are listed in Table 2. We used the Huber weight function only, and have selected various values of b. When we tried to use the Hampel ψ function, observations 4 and 18 were immediately given weight 0, in which case even computing the maximum likelihood estimates is delicate.

A biased analysis uses a Huber function with b = 6.408, but with $c(\beta,a) \equiv 0$. When we used this value of b in our conditional Fisher-consistent score, we find that observations 4 and 18 are barely downweighted and the resulting analysis looks very much like the usual likelihood analysis. As we move b to 5.5 and 5.0, observations 4 and 18 are given very little weight, and parameter estimates and standard errors change dramatically. In this example too the value of the sensitivity b seems to be most important, and we recommend to successively decrease b and see how estimates, standard errors and weights change.

6 : CONCLUSIONS

Conditionally unbiased score functions are appealing because their definition does not depend on the distribution of the predictors $\{x_i\}$. In the context of robustness, the resulting estimators have an analogous optimality theory to that already developed for unconditionally unbiased score functions. In addition, conditionally unbiased score functions are often far easier to define. Although ignoring the bias and setting $c \equiv 0$ turned out not to matter much in the examples considered, one can construct situations where this bias is large. With our estimator, we avoid this problem with little additional complexity.

REFERENCES

Carroll, R. J. (1979). Estimating variances of robust estimators when the errors are asymmetric. Journal of the American Statistical Association 74, 674-679.

Giltinan, D. M., Carroll, R. J. & Ruppert, D. (1986). Some new methods for weighted regression when there are possible outliers. Technometrics 28, 219-230.

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

McCullagh, P. & Nelder, J. A. (1983). Generalized Linear Models. Chapman & Hall, New York and London.

Pregibon, D. (1981). Logistic regression diagnostics. Annals of Statistics 9, 705-724.

Stefanski, L. A., Carroll, R. J. & Ruppert, D. (1986). Optimally bounded score functions for generalized linear models with applications to logistic regression. Biometrika 73, 413-425.

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

TABLE 1

This is a reanalysis of the food stamp data. For selected observations, the weights w_b in equation (3.3) are computed.

	MLE	Huber(7) $c(\beta,a)=0$	Huber(7) Conditional Unbiased	Huber(5.5) Conditional Unbiased	Hampel(6,14,32) Conditional Unbiased
Intercept	0.93	4.26	4.51	5.49	6.00
	(1.62)	(2.55)	(2.54)	(2.66)	(2.76)
Tenancy	-1.85	-1.85	-1.78	-1.76	-1.80
	(.53)	(.54)	(.54)	(.51)	(.54)
Supplemental Income	0.90	0.75	0.74	0.62	0.70
	(.50)	(.52)	(.51)	(.52)	(.52)
Log(1+MI), MI=Monthly Income	-0.33	-0.89	-0.93	-1.10	-1.18
	(.27)	(.43)	(.43)	(.45)	(.47)
Weights					
#5		0.21	0.16	0.13	0.0
#66		0.76	0.60	0.41	0.54

TABLE 2

This is a reanalysis of the skin vasoconstriction data. For selected observations, the weights w_b in equation (3.3) are computed.

	MLE	Huber(6.41) $c(\beta,a)=0$	Huber(6.41) Conditional Unbiased	Huber(5.5) Conditional Unbiased
Intercept	-2.92 (1.29)	-5.71 (2.45)	-2.98 (1.35)	-6.41 (2.84)
log(volume)	5.22 (1.93)	9.13 (3.73)	5.27 (1.93)	9.98 (4.38)
log(rate)	4.63 (1.79)	8.09 (3.31)	4.67 (1.86)	8.85 (3.82)
Weights				
#4		0.38	>.80	0.25
#18		0.44	>.80	0.29