ACTA ASTRONOMICA
Vol. **40** (1990) pp. 379–388

Robust Regression with Huber's Weights in Predictions of Flare Activity

by

Anna Bartkowiak
Institute of Computer Science, University of Wroclaw

Maria Jakimiec
Astronomical Institute, University of Wroclaw, Poland

Received August 10, 1990

ABSTRACT

This work completes our earlier paper (Bartkowiak and Jakimiec 1990). We consider predictions of flare activity by use of linear regression functions. The parameters of these functions are usually estimated by the LSE (least squares of errors) method. It is known that estimates obtained by this method can be strongly influenced by atypical data vectors. In the earlier paper we stated that we can obtain more stable regression estimates when applying the α-trimmed regression method. In this paper we use another, more refined method, called robust regression with Huber's weights. This method permits to obtain more detailed information on the role and impact of data items in the estimated regression.

1. Introduction. Aim of the paper

Predictions of flare activity may be done by linear regression functions (Jakimiec and Wasiucionek 1980, Bartkowiak and Jakimiec 1986, Jakimiec and Jakubowska 1988). Parameters of these functions have to be estimated from sample data. Classically, this is done by the ordinary least squares method. However, this method can be much influenced by atypical data vectors (outliers) which are likely to occur in the data. The problem of detection and identification of influential points in a regression was dealt with by Jakimiec and Bartkowiak (1989).

To obtain more stable regression functions which fit to the bulk of the data and are not so much influenced by outliers – some other methods of estimation were developed. Generally speaking, the new methods aim at providing estimates which are robust (resistant) to outliers. A class of such estimators called L_1-norm, or generally, L_p-norm ($1 \leq p < 2$) estimators, is believed to be more resistant against outliers (Huber 1981, Hampel *et al.* 1986). Other robust estimators are obtained by considering the so called α-trimmed regression as described by Antoch and Jureckova (1985) or Antoch and Bartkowiak (1988). The last method was applied by Bartkowiak and Jakimiec (1990) for solar flare predictions and permitted to identify items with unexpectedly large residuals.

In this paper we consider a more refined method, namely the robust regression with Huber's weights. It gives more detailed informations on the role and impact of individual data items in the considered regression.

In the following we remind briefly the idea of an α-trimmed regression. Next we introduce the robust regression with Huber's weights. In Section 3 we use this regression for construction of short term predictions of flare activity considering the same data which was used by Bartkowiak and Jakimiec (1990).

2. Methods of estimating parameters of a regression function

Consider a predicted variable y and p predictors $x_1, ..., x_p$. We assume the linear regression model

$$y = b_0 + b_1 x_1 + ... + b_p x_p + e \tag{1}$$

where e, the error term, is a random variable with expected value equal to zero and a variance equal to σ^2.

The parameters of the employed regression function have to be estimated from the data. Our data are in the form of a matrix $X = \{xij\}, (i = 1, ..., n, j = 1, ..., p)$, containing the values of the p predictors, and a vector $\mathbf{y} = (\mathbf{y}_1, ..., \mathbf{y}_\mathbf{n})^\mathbf{T}$ of the predicted variable y – collected for n items.

The classical LSE (least squares of errors) estimation method minimizes the sum of squares of residuals

$$\sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y_i - \hat{b}_0 - \hat{b}_1 x_{i1} - \ldots - \hat{b}_p x_{ip})^2 \tag{2}$$

In the above $\hat{b}_0, ..., \hat{b}_p$ are the estimates of $b_0, ..., b_p$ for which expression given by Eq. (2) attains minimum.

The estimation of a regression function can be done in such a way that yields estimators less influenced by outliers. One such method of estimation is the so called α-trimmed regression (Antoch and Jureckova 1985, Antoch and Bartkowiak 1988). This method proceeds in three steps. First it minimizes the sum of weighted absolute residuals r_i

$$\sum_{i=1}^{n} w_i |r_i| = \sum_{i=1}^{n} w_i |y_i - b_0 - b_1 x_{i1} - \ldots - b_0 x_{ip}| \qquad (3)$$

with weights w defined as follows:

$$w_i = \begin{cases} \alpha & \text{if } r_i > 0 \\ \alpha - 1 & \text{if } r_i < 0 \end{cases} \qquad (3a)$$

and rejects the data items with positive weights (*i.e.* yielding positive residuals). In the second step it minimizes the sum of weighted absolute residuals analogous to (3), but with weights

$$w_i{}' = \begin{cases} 1 - \alpha & \text{if } r_i > 0 \\ -\alpha & \text{if } r_i < 0 \end{cases} \qquad (3b)$$

and rejects the data items with negative weights (*i.e.* yielding negative residuals). Finally, in the third step, the classical LSE estimates are evaluated from the remaining data items.

This method proved to be very useful in short-term predictions of flare activity. Its virtue is that it permits to fit the regression function to the bulk of the data and yields estimates of parameters which are not much influenced by the outliers hidden in the data. Its handicap is that it is a crude method: it simply removes a part of the data items and carries out the process of estimation for the diminished data set. The remaining amount of data vectors depends upon α, a declared fraction of the total size n. Moreover, always an α part of the data items yielding the largest positive residuals, and an α part of the data items yielding the largest negative residuals – are removed from the data. So always a fraction 2α of the data items is not accounted for in the process of estimation – and this is done without looking more precisely which was exactly the impact of the removed items on the estimated regression.

The method of robust regression with Huber's weights enables to fit the regression to the bulk of the data items not by removing items with big residuals but by giving them less weight. This method is more refined then α-trimmed method: it permits to reveal more exactly the role played by the individual data vectors in the estimated regression.

Now let us introduce the robust regression with Huber's weights. We seek for estimates $\hat{b}, \hat{b}_1, \ldots, \hat{b}_p$ minimizing the sum of weighted squares of residuals r:

$$\sum_{i=1}^{n} w_i(r_i) r_i{}^2 = \sum_{i=1}^{n} w_i(r_i) \left(yi - \hat{b}_0 - \hat{b}_1 x_{i1} - \ldots - \hat{b}_p x_{ip} \right)^2 \qquad (4)$$

with weights $w_i(r_i)$ depending on the magnitude of the respective residual r_i. Huber's weights are defined as follows :

$$w(t) = \begin{cases} 1 & \text{if } |t| \leq K \\ K/|t| & \text{if } |t| > K \end{cases} \tag{5}$$

with K being an assumed constant called also "tuning" constant (usually we take $K = 1$ or $K = 1.5$).

Often the weights w_i apperaring in (4) are defined as functions of standardized residuals

$$t_i = r_i/\hat{\sigma} \tag{5a}$$

with

$$\hat{\sigma}^2 = \frac{1}{n-p-1}\sum_{i=1}^{n} r_i^2 \tag{5b}$$

Sometimes a more robust estimator for σ^2 is used, *e.g.* σmay be estimated as the median of absolute residuals.

Substituting t_i from (5a) into (5), *i.e.* into the formula for Huber's weights, one can see that: 1. Huber's weights are equal to 1.0 if t_i, the ith standardized residual, is smaller or equal to K the assumed tuning constant; and 2. Huber's weights become smaller than K, inversely proportional to the magnitude of absolute value of r_i, if the residual t is larger than the tuning constant K.

To evaluate the standardized residuals t_i we have to know σ, the expected standard error of the error terms e appearing in the regression equation (1) :

$$\sigma^2 = E(e)$$

Usually σ^2 is not known and as a substitute we use an estimate $\hat{\sigma}^2$ evaluated with Eq. (5b). The estimate $\hat{\sigma}^2$ is evaluated together with the estimates $\hat{b}_0, \hat{b}_1, ..., \hat{b}_p$ of the regression function.

To solve Eq. (4) for $\hat{b}_0, \hat{b}_1, ..., \hat{b}_p$ we used the iterative reweighted least squares method proposed by Beaton and Tukey and described by Li (1985). In the following we shall refer to the estimated regression as the Beaton-Tukey regression with Huber's weights. To obtain estimates of this regression we considered the set of $p+1$ simultaneous equations determined by the matrix equation

$$\mathcal{X}^T W \mathcal{X}\mathbf{b} = \mathcal{X}^T W \mathbf{y} \tag{6}$$

with W being a diagonal matrix, and the matrix $\mathcal{X}$ is obtained by adding to the data matrix X a column of ones: $\mathcal{X} = [1_n, X]$. The obtained equation looks similar as that of a weighted least-squares regression, but now the weights w_i $(i = 1, ..., n)$ are neither equal nor X-determined – they depend on the residuals $r_i = y_i - x_i\mathbf{b}$.

The matrix equation (6) was solved iteratively using the Beaton-Tukey method with Huber's weights.

In each iterative step (no. m) we obtained an estimate $\mathbf{b}^{(m)}$ solving the matrix equation (6) approximately. From the vector $\mathbf{b}^{(m)}$ the residuals ${r_i}^{(m)}$, an estimate of σ and the new weights ${w_i}^{(m)}$ were evaluated:

$$\begin{aligned} {r_i}^{(m)} &= y_i - x_i \mathbf{b}^{(m)}, \\ \hat{\sigma}^{(m)} &= \mathrm{med}({r_1}^{(m)}, \ldots, {r_n}^{(m)}), \\ {w_i}^{(m)} &= \frac{\Psi[{r_i}^{(m)}/\hat{\sigma}^{(m)}]}{{r_i}^{(m)}/\hat{\sigma}^{(m)}} \end{aligned}$$

with the function $\Psi(t)$ defined for the case of Huber's weights as

$$\Psi(t) = \begin{cases} t & \text{for } |t| < K, \\ K\mathrm{sgn}(t) & \text{for } |t| > K. \end{cases} \tag{7}$$

Then a new approximation of $\mathbf{b}$ was evaluated by the formula:

$$\mathbf{b}^{(m+1)} = \mathbf{b}^{(m)} + (\mathcal{X}^T W^{(m)} \mathcal{X})^{-1} \mathcal{X}^T W^{(m)} (\mathbf{y} - \mathcal{X}\mathbf{b}^{(m)}) \tag{8}$$

The parameter σ was estimated before each iterative step as the median of absolute residuals.

The variance-covariance matrix $S_{\hat{\mathbf{b}}}$ of the coefficients $\hat{\mathbf{b}} = \hat{\mathbf{b}}_0, \ldots, \hat{\mathbf{b}}_p$ which solves Eq. (4) is given by

$$S_{\hat{\mathbf{b}}} = \frac{\sum_1^n \Psi(r_i)^2/(n-p)}{[\sum_1^n \Psi'(r_i)/n]^2} (\mathcal{X}^T \mathcal{X})^{-1} \tag{9}$$

where Ψ' is the derivative of Ψ with respect to the argument r_i.

3. Predictions of flare activity by robust regression with Huber's weights

We applied the robust regression with Huber's weights to the data considered formerly by Jakimiec and Bartkowiak (1989) and Bartkowiak and Jakimiec (1990). The data comprises $n = 149$ items for the year 1979 and describes daily characteristics of sunspot groups of D,E,F Zurich classes. The considered predictors are: (1) Mc Intosh sunspot class (McI); (2) Sunspot group area (A); (3) Calcium area (CaA); (4) Calcium plage intensity (CaI); (5) Magnetic class (Mag); (6) Magnetic field strength (H); (7) Magnetic field index (MFI); (8) Maximum value of X-ray flare flux ($maxX$); (9) Number of faint flares (NFF); (10) Number of strong flares (NSF); (11) The total daily sum of the X-ray flare flux for the wavelengths 1–8 Å (Fs); (12) Hardness index (HI); (13) The total daily sum of the X-ray flare flux for the wavelengths 0.5–4 Å (Fh). The predicted characteristics are: Fs and Fh on the next day.

For these data we evaluated the coefficients of the robust regression function satisfying Eq. (4). Together with the estimates $\hat{b}_0, \ldots \hat{b}_p$ we evaluated also their standard deviations by Eq. (9). From these values we evaluated $\tilde{b}_i$, the standardized (robust) coefficients of regression given as

$$\tilde{b}_i = \hat{b}_i / \sqrt{\mathrm{var}(\hat{b}_i)}.$$

The respective values of $\tilde{\mathbf{b}}$ evaluated for the regressions for the predicted variables Fs and Fh are given in Table 1.

Table 1

Standardized values of the regression coefficients evaluated by (a) the classical LSE, (b) robust α-trimmed and (c) robust regression with Huber's weights method.

No. of explanatory variable	Predicted variable: F_s			Predicted variable: F_h		
	LSE (a)	α-trimmed (b)	Huber's (c)	LSE (a)	α-trimmed (b)	Huber's (c)
0	-2.72	-	-2.75	-2.21	-	-2.78
1	0.81	0.64	0.73	0.46	0.93	0.65
2	-0.99	-1.26	-0.84	-1.10	-1.74	-1.13
3	1.95	2.52	2.00	1.70	2.35	1.81
4	2.20	2.43	1.97	2.22	3.11	2.67
5	0.54	1.22	0.88	1.07	0.93	1.10
6	-0.34	-0.05	-0.51	-0.54	-0.97	-0.92
7	1.87	2.04	2.26	1.92	3.47	2.58
8	0.59	1.35	0.60	0.41	0.41	0.42
9	-0.46	-0.18	-0.67	-0.40	-0.29	-0.46
10	2.89	3.44	2.27	3.10	3.19	3.08
11	1.84	2.11	2.15	1.45	1.47	1.59
12	-0.51	-0.95	-0.43	-0.30	-0.44	-0.42
13	-1.47	-2.01	-1.72	-1.13	-1.14	-1.18

One can see that the standardized values of the regression coefficients evaluated by the three methods (*i.e.* the classical least squares method, the robust α-trimmed method and the robust Beaton-Tukey method using Huber's weights) are very similar. For some variables the α-trimmed method gives higher coefficients than the classical LSE methods. The robust Huber's method gives also coefficients similar to those obtained by the LSE method. In principle, the differences between the coefficients obtained by Huber's

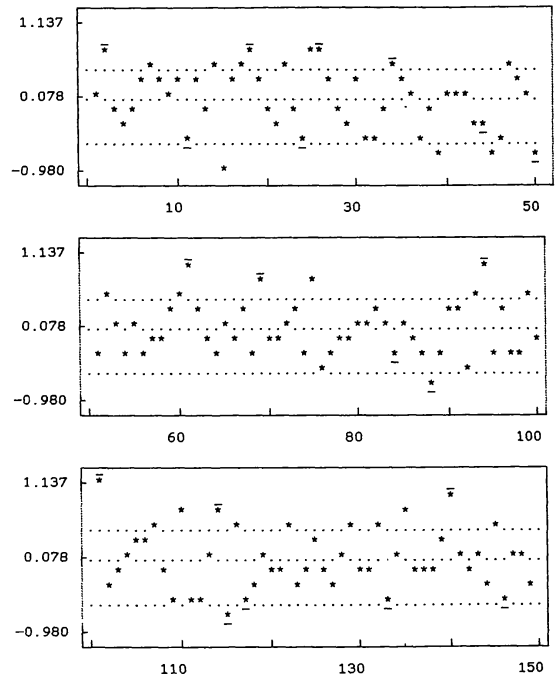

Fig. 1. Residuals for the variable Fs (y-axis) evaluated from robust regression with Huber's weights. The x-axis indicates the item numbers. Items trimmed off by the α-trimmed regression are marked by a dash.

and LSE methods are smaller than the differences between the coefficients obtained by the α-trimmed and the LSE method. We have studied also the residuals evaluated from the robust regression using Huber's weights. The residuals are shown in Fig. (1) and (3) for the predicted variables Fs and Fh, respectively. The residuals r_i in these figures are shown versus i, the no. of the item. The expected value of each residual is equal to zero. The boundaries in the figures marked by dots indicate the value $2\hat{\sigma}$, where $\hat{\sigma}^2$ denotes the estimated value of the variance of the error term appearing in the regression Eq. (1). The value σ was estimated as a median of residuals. We have marked with a dash the residuals of those items, which were trimmed off by the robust α-trimmed regression when using $\alpha = 0.1$. One can see that, generally, the trimmed-off items have large residuals evaluated by the robust regression using Huber's weights.

In Fig. (2) and (4) we show the weights w_i ascribed to subsequent items by the robust regression using Huber's weights for the predicted variables Fs and Fh, respectively. We have marked in these figures by circles the weights for those items, which were trimmed off when applying the α-trimmed regression with $\alpha = 0.1$.

One can see that the robust regression with Huber's weights gives much smaller weights to the items removed by the robust α-trimmed regression.

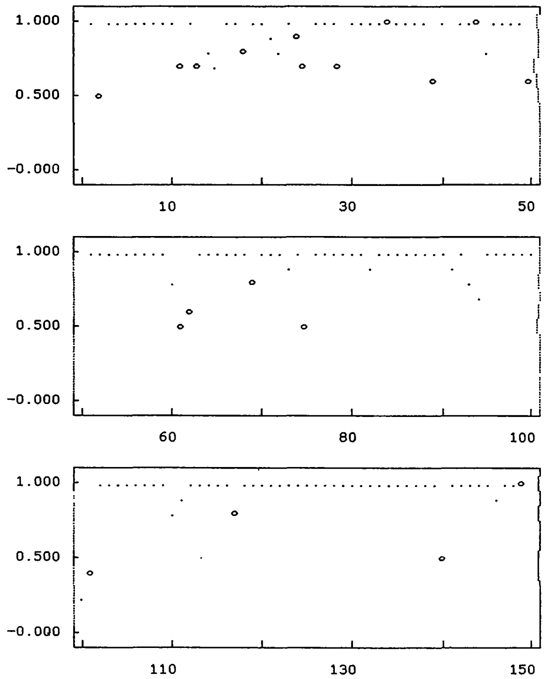

Fig. 2. Predicted variable: Fs. Indexplot of weights (y-axis) given to the considered items by the robust regression with Huber's weights. The item numbers are marked on the x-axis. Items trimmed off by the α-trimmed regression are marked by a circle.

However there are few disaccords: For instance, the item no. 44 was trimmed off by the α-trimmed regression, but it obtained the full weight (equal to 1.0) in the Huber regression. That might be explained by the crudeness of the α-trimmed methods, which assignes data items into two categories only: to be removed or to be left, while the Beaton-Tukey regression assignes to the items values (weights) from the dense interval [0,1].

4. Conclusions

Taking into account the results presented in Figs. (1)–(4) and in Table 1 we may conclude:

The robust regression with Huber's weights is a refined method. It assigns to the considered items various weights according to the magnitude of the residuals obtained from the estimated regression: the items with large residuals obtain smaller weights – what, in turn, makes them less influential in the evaluated regression.

In principle, the items trimmed off by the robust α-trimmed regression obtained smaller weights in the Beaton-Tukey regression with Huber's weights. These items are atypical with respect to some characteristics; this is connected with great changes of flare activity of the sunspot grups from day to day.

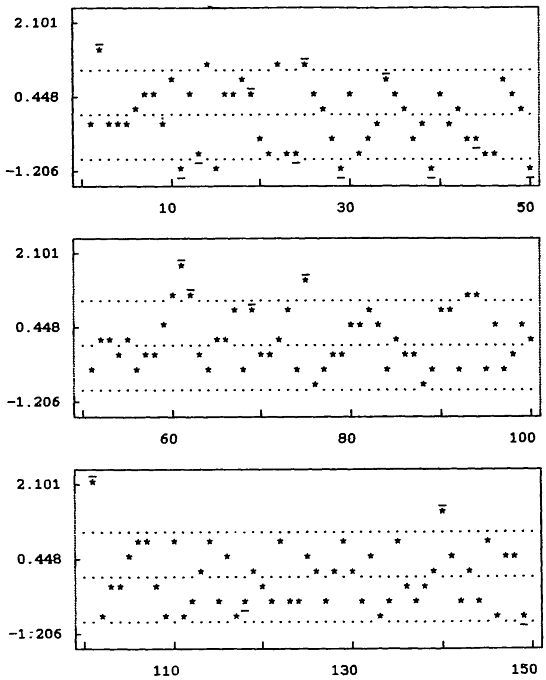

Fig. 3. The same as Fig. 1 but for the predicted variable *Fh*.

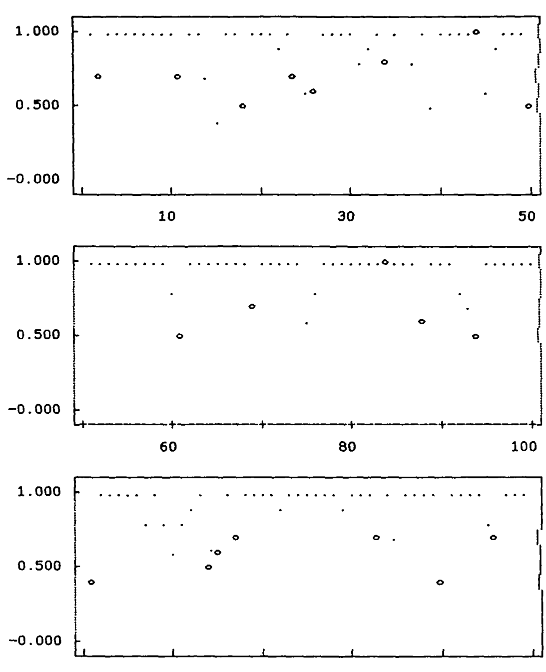

Fig. 4. The same as Fig. 2 but for the predicted variable *Fh*.

What concerns the prediction problem – we can say, that the results obtained by robust regression with Huber's weights confirm the conclusions presented in our earlier paper (Bartkowiak, Jakimiec 1990). In this paper, considering both residuals and weights given to individual items, we got a more detailed information on the items identified earlier as possibly influential in the regression used for predictions.

REFERENCES

Antoch, J., and Bartkowiak, A. 1988, *Listy Biometryczne (Biometrical Letters)*, **35**, 3.

Antoch, J.,and Jurezkova, J. 1985, *Comput.Stat.Quarterly*, **2**, 329.

Bartkowiak, A., and Jakimiec, M. 1986, *Solar-Terrestrial Predictions*, eds. P.A. Simon, G. Heckman, and M.A.Shea, p.285.

Bartkowiak, A., and Jakimiec, M 1989, *Acta Astr.*, **39**, 85.

Bartkowiak, A., and Jakimiec, M 1990, *Acta Astr.*, **40**, 169.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. 1986, *Robust statistics - The approach based on influence functions*, (Wiley, New York), p. 7.

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

Jakimiec, M., and Bartkowiak, A. 1989, *Acta Astr.*, **39**, 257.

Jakimiec, M., and Wanke-Jakubowska, M. 1988, *Acta Astr.*, **38**, 431.

Jakimiec, M., and Wasiucionek, J. 1980, *Solar-Terrestrial Predictions Proceedings*, ed. Donnelly, p. 271.

Li, G. 1985, *Robust regression. In: Exploring Data Tables, Trends and Shapes*, eds. D.C. Hoaglin, F. Mosteller, J.W. Tukey (Wiley), p.281.

Sposito, V., Smith, W., and Mc Cormick, G. 1978, *Minimizing the sum of absolute deviations*, Vanderhoeck and Ruprecht, Gotingen.