L_1-*Statistical Procedures and Related Topics*
IMS Lecture Notes – Monograph Series (1997) Volume 31

Exploring data sets using partial residual plots based on robust fits

Joseph W. McKean

Western Michigan University, USA

Simon J. Sheather

University of New South Wales, Australia

Abstract: Partial residual plots are one of the most useful graphical procedures in the exploratory fitting of data sets. They are frequently used in the identification of unknown functions, g, of predictor variables. Traditionally these plots have been based on least squares (LS) fitting. It is well known that LS estimates are sensitive to outlying observations. The examples and sensitivity study discussed in this paper show that this vulnerability to outliers carries over to the LS based partial residual plots. A few outliers in the data set can distort the LS partial residual plot making the identification of g impossible. Furthermore, if g is non-linear, good data points may act as outliers and cause distortion in the plot. Partial residual plots based on highly efficient robust estimates are presented. In the simulated data sets explored in this paper, the robust based partial residual plots are insensitive to the outlying observations leading to a much easier identification of the unknown functions than their LS counterparts. In the sensitivity study presented, these robust based partial residual plots do not become distorted in the presence of outliers but they maintain their focus, enabling the identification of g.

Key words: Linear model, M-estimates, outlier, regression diagnostics, R-estimates, rank based methods.

AMS subject classification: 62J20, 62G35.

1 Introduction

Partial residual plots are one of the most useful graphical procedures in the exploratory fitting of data sets. These plots are quite simple. Consider a model of the form $y_i = \alpha + \boldsymbol{\beta}_1'\mathbf{x}_{i1} + g(x_{i2}) + \epsilon_i$, where the function $g(x)$ is unknown. Then (the first-order) partial residuals are the residuals of the fit of the misspecified model $y_i = \alpha + \boldsymbol{\beta}_1'\mathbf{x}_{i1} + \beta_2 x_{i2} + e_i$ added to the fitted part $\widehat{\beta}_2 x_{i2}$. The plot consists of these partial residuals plotted versus x_{i2}. This plot is often informative in the identification of the unknown function $g(x)$.

Partial residual plots were proposed by Ezekiel (1924) and have been discussed by numerous authors. Larsen and McCleary (1972) gave the name partial residual plot to this procedure. Mallows (1986) extended these first-order plots to higher orders, the so-called augmented partial residual plots; Mansfield and Conerly (1987) considered informative algebraic representations of partial residuals; Cook (1993) obtained further theoretical underpinnings of these plots and proposed an extended class, the CERES plots; and Berk and Booth (1995) compare partial residual plots with several other diagnostic plots in a series of interesting examples. Based on work such as this, partial residual plots have become an important tool in data exploration.

Most of the discussion of partial residual plots is based on the traditional least squares (LS) fitting of models. Partial residuals, though, are simply residuals added to the fit of the misspecified part. Hence, fits other than LS can be considered. McKean and Sheather (1997) developed properties of partial residuals based on robust fitting. They showed that the expected behavior of the resulting robust partial residual plots was similar to that of the LS partial residual plots. Furthermore, they showed that the robust partial residual plots were not as sensitive to outliers as the LS based plots.

To determine which robust estimates to use, note that the function $g(x)$ is often a nonlinear function. Hence the employed fitting criteria should be able to detect and fit curvature. We have selected highly efficient M and R estimates as the basis of our fitting criteria. These estimates and their residuals have been shown to behave similar to their LS counterparts in detecting and fitting curvature on good data, while being much less sensitive to LS procedures on data containing outliers in the Y-space; see McKean, Sheather and Hettmansperger (1990, 1993, and 1994). These fitting criteria are based on minimizing convex functions; hence, the consistency theory developed by Cook (1993) for LS partial residual plots extends to these robust partial residual plots. Also these highly efficient robust fitting criteria are computationally fast and available.

In this paper, we explore several data sets using robust based partial residual plots. In many of these data sets, outliers distort the LS based plots to the point where the identification of the unknown function $g(x)$ is impossible. In one of the data sets, due to the nonlinearity of the function g, good data acted as outliers and distorted the LS based partial residual plot. The robust based partial residual plots, though, are not sensitive to the effect of the outliers. These plots clearly identify the unknown function g. The sensitivity study in Section 5 shows the distortion of the LS based partial residual plot in a sequential fashion as a few points become increasingly outlying. The robust based partial residual plots, however, retain their "focus" under the increasing influence of the outliers.

2 Notation

This paper considers partial residual plots based on robust estimates. As discussed in Section 3, these plots are often used to graphically determine unknown functions of predictors. These functions are often nonlinear so fitting procedures which can detect curvature are of interest. Studies by Cook, Hawkins and Weisberg (1992) and McKean, Sheather and Hettmansperger (1993, 1994) have shown that high breakdown and bounded influence estimates have problems in detecting and fitting curvature, while highly efficient robust estimates are capable of detecting and fitting curvature. Hence, in this article we will focus on highly efficient robust estimates. To keep things simple, we have chosen the Huber M estimate and the Wilcoxon R estimate. Both of these estimates are widely available. But clearly other robust estimates, (other ψ-functions and other score functions), can be used and will produce similar results. Similar to LS-estimates, though, the Huber and Wilcoxon estimates are highly sensitive to outliers in the $\mathbf{x}$-space. This should be considered in exploring any data set prone to outliers in factor space. McKean, Naranjo and Sheather (1996a, 1996b) discuss diagnostic procedures that measure the overall difference between highly efficient and high breakdown robust estimates and determine cases where the fits differ.

Consider the linear regression model $y_i = \alpha + \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i, \quad i = 1, \ldots, n$ where $\mathbf{x}_i'$ is the ith row of the $n \times p$ centered matrix $\mathbf{X}$ of explanatory variables defined here. The least squares estimates $\widehat{\alpha}$ and $\widehat{\boldsymbol{\beta}}_{LS}$ minimize the dispersion

$$D_{LS}(\alpha, \boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i - \alpha - \mathbf{x}_i'\boldsymbol{\beta})^2. \tag{1}$$

Let σ^2 be the common error variance. Under regularity conditions, the asymptotic distribution of the LS estimates is given by

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\boldsymbol{\beta}}_{LS} \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha \\ \boldsymbol{\beta} \end{pmatrix} \sigma^2 \begin{bmatrix} 1/n & \mathbf{0}' \\ \mathbf{0} & (X'X)^{-1} \end{bmatrix} \right) . \tag{2}$$

The regular (Wilcoxon) R-estimate $\widehat{\boldsymbol{\beta}}_W$ minimizes the dispersion

$$D_W(\boldsymbol{\beta}) = \sum_{i=1}^{n} a(R(y_i - \mathbf{x}_i'\boldsymbol{\beta}))(y_i - \mathbf{x}_i'\boldsymbol{\beta}) \ , \tag{3}$$

where $R(y_j - \mathbf{x}_j'\boldsymbol{\beta})$ is the rank of $y_j - \mathbf{x}_j'\boldsymbol{\beta}$ among $y_1 - \mathbf{x}_1'\boldsymbol{\beta}, \ldots, y_n - \mathbf{x}_n'\boldsymbol{\beta}$ and the scores $a(i)$ are generated by the linear function

$$\varphi(u) = \sqrt{12}\left(u - \frac{1}{2}\right) \ , \tag{4}$$

as $a(i) = \varphi(i/(n+1))$. Although we will be using Wilcoxon scores throughout this paper, the φ notation will be useful. The function (3) is a convex function of $\boldsymbol{\beta}$ and Gauss-Newton type algorithms suffice for the minimzation; see Kapenga, McKean and Vidmar (1988). Note that (3) is invariant with respect to an intercept term. We shall estimate α by the median of the Wilcoxon residuals, i.e.,

$$\widehat{\alpha}_W = \text{med}(y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}_W) \ . \tag{5}$$

Estimating the intercept in this way, avoids unnecessary assumptions such as symmetric error distributions; see Hettmansperger, McKean and Sheather (1997). Our aim is to make as few assumptions as possible when concerned with data exploration.

Under regularity conditions, the Wilcoxon estimates have asymptotic distribution

$$\begin{pmatrix} \widehat{\alpha}_W \\ \widehat{\boldsymbol{\beta}}_W \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha \\ \boldsymbol{\beta} \end{pmatrix} \begin{bmatrix} \tau_s^2/n & \mathbf{0}' \\ \mathbf{0} & \tau^2(X'X)^{-1} \end{bmatrix} \right) \ , \tag{6}$$

where $\tau^{-1} = \sqrt{12}\int f^2(t)dt$, (Jaeckel, 1972), $\tau_s = 1/(2f(0))$, and f is the error density. Consistent estimates of τ and τ_s are presented in Koul, Sievers and McKean (1987) and McKean and Schrader (1983), respectively.

Our third estimate will be Huber's M-estimate $\widehat{\boldsymbol{\beta}}_M$ which minimizes the dispersion function

$$D_M(\boldsymbol{\beta}) = \sum_{i=1}^{n} \rho((y_i - \alpha - \mathbf{x}_i'\boldsymbol{\beta})/\sigma_0) \ , \tag{7}$$

where σ_0 is a sacle parameter and ρ is given by

$$\rho(x) = \begin{cases} x^2/2 & \text{if } |x| < h, \\ |x|h - h^2/2 & \text{otherwise.} \end{cases} \tag{8}$$

The bend, the parameter h, must be set. In this paper we will take $h = 1.345$, the default setting used in Splus; see Becker, Chambers and Wilks (1988). Under regularity conditions $\widehat{\boldsymbol{\beta}}_M$ has an asymptotic normal distribution with asymptotic variance $\kappa^2(\mathcal{X}'\mathcal{X})^{-1}$, where $\mathcal{X} = [\mathbf{1}_n \vdots \mathbf{X}]$ and

$$\kappa^2 = \frac{\sigma^2 E[\psi^2(\epsilon_i/\sigma_0)]}{(E[\psi'(\epsilon_i/\sigma_0)])^2} ,$$

where $\psi(t) = \rho'(t)$; see Huber (1981). The constant of proportionality κ^2 can be estimated by the usual moment estimators.

3 Partial residual plots

We will be concerned with models of the form

$$y_i = \alpha + \boldsymbol{\beta}_1' \mathbf{x}_{1i} + g(\mathbf{x}_{2i}) + \epsilon_i , \tag{9}$$

where $\mathbf{x}_{1i}$ and $\mathbf{x}_{2i}$ are $p \times 1$ and $q \times 1$ vectors of regression coefficients, $\mathbf{x}_{1i}'$ is the ith row of the $n \times p$ matrix $\mathbf{X}_1$, and $g(x)$ is an unknown function. The goal is to try to determine the function g as best as possible using simple graphic techniques. The partial residual plot described below is an attempt to obtain this goal. A recent overview can be found in the paper by Cook (1993).

The description of the partial residual plot is the same regardless of what criteria is used to fit a model, so we will describe it generically by dropping the subscripts LS, W and M for the fitting criteria. Hence, let $\widehat{\boldsymbol{\beta}}$ denote an estimate of the parameter $\boldsymbol{\beta}$ in a model. We will use the subscripts when distinctions are necessary.

In this article, we will only be looking at cases where the predictor x_{i2} is univariate. Let $\mathbf{x}_2 = (x_{12}, \ldots, x_{n2})'$. Since g is unknown we begin our exploration by fitting a first order model, (at the end of this section we will discuss fitting higher order models). Consider then fitting the model

$$y_i = \alpha + \boldsymbol{\beta}_1' \mathbf{x}_{1i} + \beta_2 x_{i2} + e_i . \tag{10}$$

Note that, unless $g(x_{i2}) = \beta_2 x_{i2}$, model (10) is a misspecified model because model (9) is the correct model. We have indicated this in Model (10) by using e_i instead of ϵ_i for the random error.

Suppose we have fitted the misspecified model (10). Denote the fit by $\widehat{y}_i$ and let $\widehat{e}_i = y_i - \widehat{y}_i$ denote the residual. The **partial residuals** are defined by

$$\widehat{e}_i^* = \widehat{e}_i + \widehat{\beta}_2 x_{2i} , \tag{11}$$

that is, the fit of the misspecified part is added back to the residuals. The **partial residual plot** is the plot of $\widehat{e}_i^*$ versus x_{i2}.

3.1 Discussion of partial residual plots

As Cook (1993) noted, since $\widehat{e}_i = y_i - \widehat{y}_i$, we can substitute the right side of equation (9) for y_i and obtain

$$\widehat{e}_i^* = (\alpha - \widehat{\alpha}) + (\boldsymbol{\beta}_1 - \widehat{\boldsymbol{\beta}}_1)'\mathbf{x}_1 + g(x_{i2}) + e_i \ . \tag{12}$$

Although these estimates are based on a misspecified model, if they are close to their true values then the partial plot is close to a plot of $g(x_{i2}) + e_i$ versus x_{i2}.

Mansfield and Conerly (1987) considered the expectation properties of LS based partial residual plots by obtaining algebraic representations of the partial residuals using the true model distributional properties. Based on these representations, they showed, among other conclusions, that if the correct model was fitted then the expected partial residual plot should be a linear function of x_{2i}. They also showed that when $\mathbf{x}_2$ and g are both orthogonal to $\mathbf{X}_1$, then we expect the partial residuals to be the unknown function $g(x)$. On the other hand, if $\mathbf{x}_2$ and $\mathbf{X}_1$ are highly collinear then there is little information in the partial residual plot.

Using the first-order approximation theory for robust residuals and fitted values established in McKean, Sheather and Hettmansperger (1990, 1993), McKean and Sheather (1997) obtained representations for the partial residuals when the true model is (9). Based on these representations, the conclusions described above of Mansfield and Conerly (1987) hold for the robust partial residual plots, also.

McKean and Sheather (1997) further developed a measure of efficiency between the robust and LS partial residual plots. If the correct model is fit then as discussed above the partial residual plot is expected to be the linear function $\beta_2\mathbf{x}_2$. Thus the plot of interest would be that of $\widehat{\beta}_{W2}x_{i2}$ versus x_{i2} overlaid on the partial residual plot. Hence, it is the precision in the linear predicted equation $\widehat{\beta}_{W2}x_{i2}$ of $\beta_2 x_{i2}$ which is of interest in terms of efficiency. This relative efficiency measure is given by the usual asymptotic relative efficience between a robust estimate and the corresponding LS estimate. For example, if the Wilcoxon based residual plots are used then this asymptotic relative efficiency is given by

$$e_{W,LS} = \frac{\tau^2}{\sigma^2} \ , \tag{13}$$

where σ^2 is the variance of the errors and τ is defined in expression (6). If the error distribution is normal then $e_{W,LS} = .955$. However, if the error distribution is heavier tailed than the normal distribution then this ratio can be quite large; see Hettmansperger (1991).

In comparing the LS and robust representations of the partial residuals, McKean and Sheather (1997) showed that the random part of the LS partial residuals has unbounded influence while the corresponding part for the robust partial residuals has bounded influence. One bad outlier, say ϵ_i, not only distorts the ith residual but other cases, also, because the representation of the LS partial residual includes the unbounded term $\mathbf{H}\boldsymbol{\epsilon}$, where $\mathbf{H}$ is the projection matrix onto the column space of $[\mathbf{X}_1 \vdots \mathbf{x}_2]$. On the other hand, this is not true of the robust partial residuals because in the respective, representational expansion of the robust partial residuals all terms are bounded. The examples and sensitivity study found in Sections 4 and 5 provide illustrations of the distortion of LS partial residual plots due to outliers.

3.2 Augmented partial residual plots

The misspecified part of model, (10), is a first-order approximation to $g(x)$. We can also crawl up the Taylor series expansion of $g(x)$ to fit higher order polynomials. This was proposed by Mallows (1986) for second-order representations. In this case, we fit the second-order model,

$$y_i = \alpha + \boldsymbol{\beta}_1' \mathbf{x}_{1i} + \beta_2 x_{2i} + \beta_3 x_{2i}^2 + e_i \ . \tag{14}$$

Now, the partial residuals are $\widehat{e}_i^* = \widehat{e}_i + \widehat{\beta}_2 x_{2i} + \widehat{\beta}_3 x_{2i}^2$, where $\widehat{e}_i$ are the residuals from the fit of model (14). Mallows called the resulting plot of $\widehat{e}_i^*$ versus x_{2i} the **augmented partial residual plots**. Another plot of interest here is $\widehat{e}_i^*$ versus $\widehat{\beta}_2 x_{2i} + \widehat{\beta}_3 x_{2i}^2$, because if the quadratic model is correct this later plot will appear linear; see the discussion above on the expected behavior of partial residual plots when the correct model is fit. Augmented plots are shown in Examples 1 and 3. Certainly higher degree polynomial approximations to $g(x)$ can be handled in the same way as these quadratic plots.

4 Examples

In this section, we discuss several examples. We have chosen them to illustrate the exploratory behavior of the partial residual plots based on robust estimates and to show the sensitivity of the LS based partial residual plots to outlying observations. The data for all the examples is simulated, so at all times the correct model is known. There was little difference between the Wilcoxon based and the Huber based partial residual plots, so in a few examples only the results for the Wilcoxon based plots are shown. The Gauss-Newton type algorithm of Kapenga et al. (1988) was used to compute the Huber and Wilcoxon estimates.

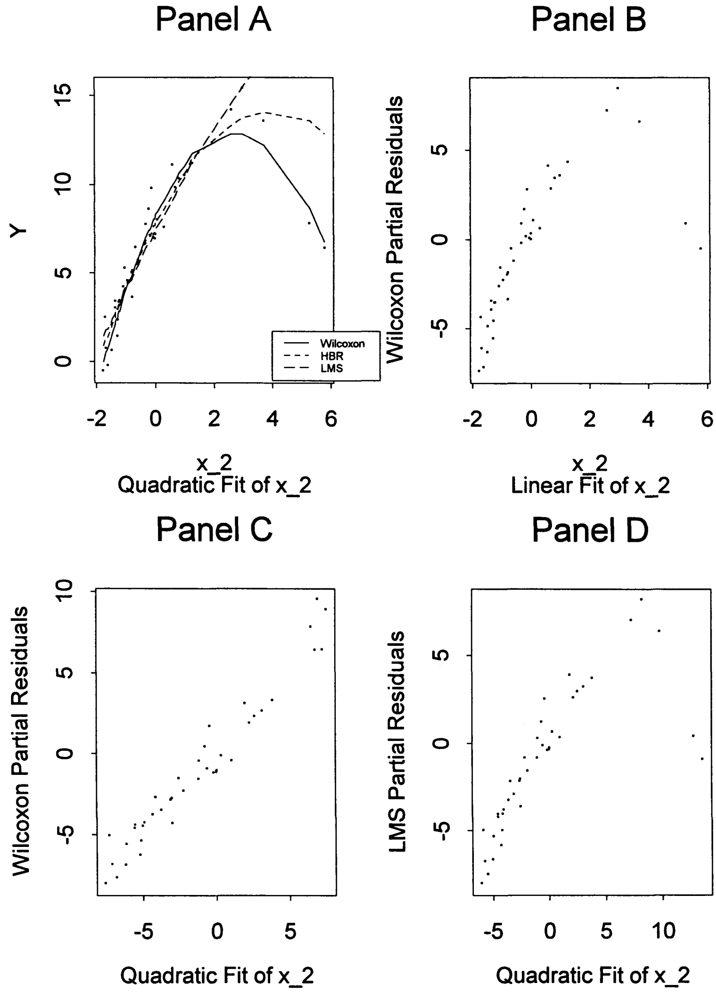

Figure 1: Plots for Example 1: Pane A, Data Overlaid with Wilcoxon, LMS and HBR fits; Panel B, Partial Residual Plot of the Wilcoxon Fit; Panel C, Augmented Partial Residual Plot of the Wilcoxon Fit; Panel D, Augmented Partiel Residual Plot of the LMS Fit.

Example 1 *Quadratic Model*

The purposes of the first example is to show how the partial and augmented partial residual plots based on a robust fit behave for a simple quadratic model. It also shows why caution is necessary when considering partial residual plots based on high breakdown estimates. The generated data follow the model

$$Y_i = 0 \cdot x_{1i} + 5.5|x_{i2}| - .6x_{i2}^2 + \epsilon_i \ , \tag{15}$$

where x_{i1} are iid uniform$(-1,1)$ variates, the ϵ_i's are simulated iid $N(0,1)$ variates and the x_{i2}'s are simulated contaminated normal variates with the contamination proportion set at .25 and the ratio of the variance of the contaminated part to the non-contaminated part set at 16. This was similar to an example discussed in Chang et al. (1997). Panel A of Figure 1 displays the scatterplot of the data overlaid by the Wilcoxon fit and two 50% breakdown fits: least median squares, LMS (Rousseeuw and Leroy, 1987), and a 50% high breakdown R estimate proposed by Chang et al. (1997), HBR. The LMS was computed using Stromberg's (1993) algorithm. Note that the fit based on the Wilcoxon estimates fits the curvature in the data quite well while the 50% breakdown estimates miss the curvature.

For the misspecified model

$$Y_i = \alpha + \beta_1 x_{1i} + \beta_2 |x_{i2}| + e_i \ ,$$

Panel B of Figure 1 displays the partial residual plot based on the Wilcoxon fit. The plot clearly shows the need to fit a quadratic model. Panel C shows the augmented Wilcoxon partial residual plot when a quadratic model was fit. This is a plot of the partial residual versus the fit of the quadratic part. If the correct model has been specified then, as noted above, this plot should show a linear pattern, which it is does. Panel D shows the same plot as Panel C except the fit based on the LMS estimates was used. Instead of a linear pattern, it shows a quadratic pattern, which is not helpful here because a quadratic model was fit.

Example 2 *Cook's (1993) Nonlinear Model*

This is an example discussed in Cook (1993). The observations are generated by

$$y_i = x_{i1} + x_{i2} + \frac{1}{1+\exp(-x_{i3})} \ , \quad i = 1, \ldots, 100 \ , \tag{16}$$

where x_{3i}'s are iid uniform$(1,26)$ random variables, $x_{1i} = x_{3i}^{-1} + z_{1i}$, $x_{2i} = \log x_{3i}^{-1} + z_{2i}$, z_{1i} has a $N(0,.1^2)$ distribution, z_{2i} has a $N(0,.25^2)$ distribution, and the z_{1i}'s and the z_{2i}'s are independent. A plot of the function $g(x_{3i}) = \frac{1}{1+\exp(-x_{i3})}$ versus x_{3i} appears in Panel A of Figure 2.

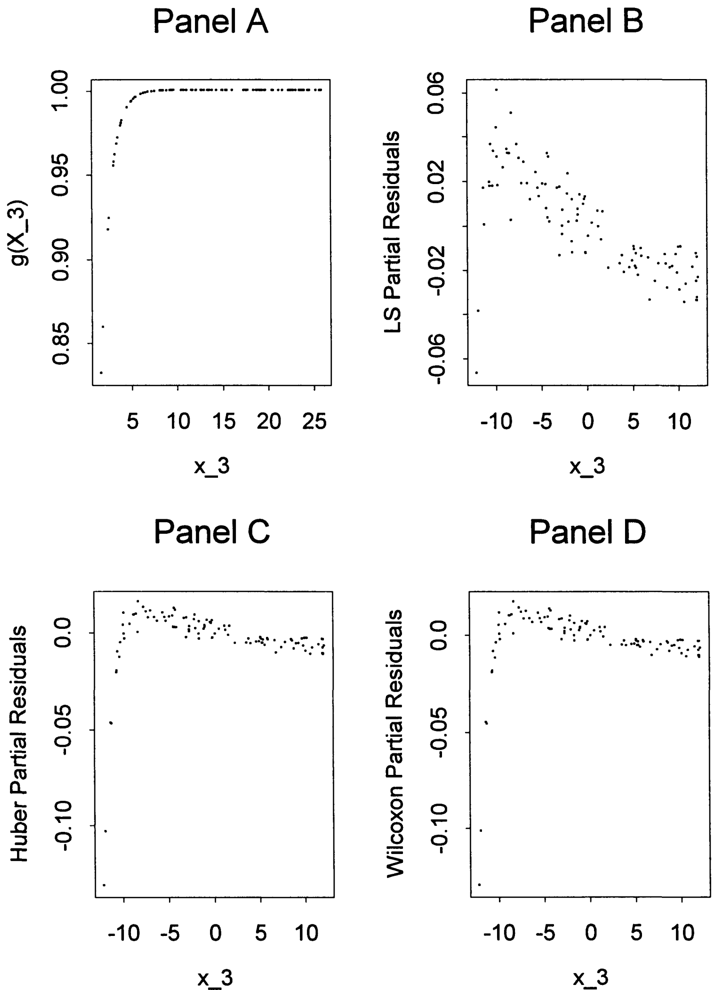

Figure 2: Plots for Example 2: Panel A, Plot of $g(x_3)$ versus x_3; Panel B, Partial Residual Plot of the LS Fit; Panel C, Partial Residual Plot of the Huber Fit; Panel D, Partial Residual Plot of the Wiloxon Fit.

This is the function that the partial residual plots are attempting to identify. Panels B, C, and D display the partial residual plots based on the LS-, Huber and Wilcoxon fits, respectively. Note that the variable x_{i3} has been centered in these plots. The function g is identifiable from both robust residual plots, but g is not identifiable from the LS-plot. The points which distorted the LS partial residual plot are the points corresponding to the low values of x_{i3}. These points acted as outliers in Y-space and corrupted the LS fit, resulting in the poor LS partial residual plot. On the other hand, the robust fits are much less sensitive to outliers in the Y-space than the LS fit; hence, these points did not corrupt the robust partial residual plots. As discussed in Section 3, one way of measuring efficiency in these plots is by the estimates of the constants of proportionality for the fitting procedure. For this example, these estimates are: $\widehat{\sigma} = .0162$, $\widehat{\kappa} = .0048$, and $\widehat{\tau} = .0053$. Hence, the robust estimates are three times more precise than the LS estimates on this data set.

Cook (1993) expands partial residual plots to the larger class of CERES plots. This procedure uses a nonparametric estimate of $E(\mathbf{x}_1 \mid x_2)$ in place of the linear function $\beta_2 x_{i2}$ in the regular partial residual plot in its construction of a partial residual plot. For this data set, as shown in the article by Cook, the procedure worked well with the LS fit. Similar plots could be developed based on robust fits, but for this example they are not needed.

Example 3 *Berk and Booth's Model*

This is an example discussed in Berk and Booth (1995). The first-order partial residual plots fail on this example for all three fits. We include it, to show the importance of the augmented residual plot.

The values for the x_{i2}'s are the 100 values: $-.99, -.97, \ldots, .99$. Then x_{i1} is generated as $x_{i1} = x_{i2}^2 + .05 z_{i1}$, where z_{i1} are iid standard normal variates. The responses are generated by

$$y_i = x_{i2}^2 + .1 z_{i2} \ , \tag{17}$$

where z_{i2}'s are iid standard normal variates and are independent of the z_{i1}'s. In this example, $g(x_{i2}) = x_{i2}^2$ and Panel A of Figure 3 shows a plot of it versus x_{i2}. This is the function that the partial residual plots are attempting to identify. Panels B, C, and D display the partial residual plots based on the LS-, Huber and Wilcoxon fits, respectively. Note that none of them identify the function g. This is hardly surprising. The generating equation for x_{i1} is a strong quadratic in x_{i2}, there is little noise. Fitting x_{i1} stole the "clout" of x_{i2}. Also, the quadratic function g is centered over the region of interest. In its Taylor series expansion about 0, the linear term would not be important; hence, the inclusion of x_{i2} as linear will not help. If we crawl up the Taylor series expansion to include a second-order term then both of these conditions are alleviated and the (augmented) partial residual plot should identify the quadratic function. This is the case as demonstrated by Panels E and F of Figure 3, which are the augmented partial residual plots of the LS and Wilcoxon partial residuals versus the quadratic fit. The linearity of these plots indicate that the appropriate model has been fit.

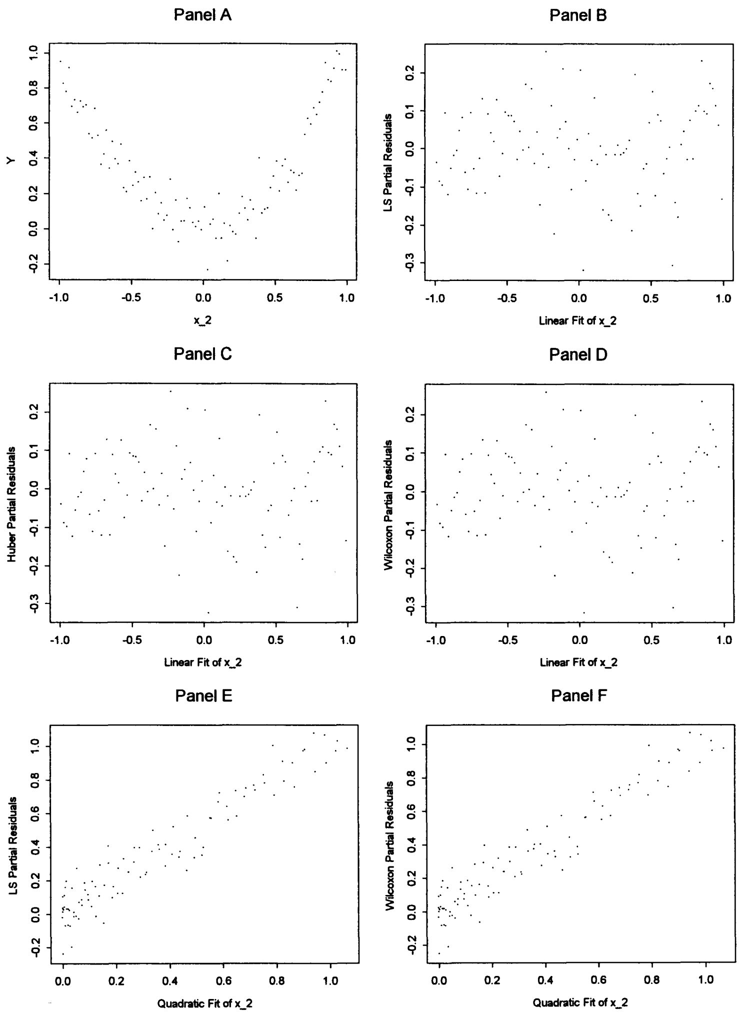

Figure 3: Plots for Example 3: Panel A, Plot of $g(x_3)$ versus x_3; Panel B, Partial Residual Plot of the LS Fit; Panel C, Partial Residual Plot of the Huber Fit; Panel D, Partial Residual Plot of the Wilcoxon Fit; Panel E, Augmented Partial Residual Plot of the LS Fit; Panel F, Augmented Partial Residual Plot of the Wilcoxon Fit.

5 Sensitivity Study

The following sensitivity study serves to illustrate the distortion of LS based partial residual caused by outliers in the Y-space. As our baseline model we consider a cubic polynomial in x_{i2} with normal errors. We chose x_{i1} to be uniform$(-1,1)$ variates. The model is

$$y_i = 0 \cdot x_{i1} + 5.5x_{i2}^2 - .6x_{i2}^3 + \epsilon_i \, ; i = 1, \ldots, 40 \, , \tag{18}$$

where ϵ_i are iid $N(0,1)$ and x_{i2} are generated from a contaminated normal distribution. The misspecified model is

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i \, . \tag{19}$$

In this setting, the partial residual plots should easily show that a cubic needs to be fit. This is confirmed by the top row of Figure 4 which are the partial residual plots based on the LS and the Wilcoxon fits, respectively, when the misspecified model, (19), is fitted.

Next, in a series of four stages we distorted the values of three of the responses, as shown in Table 1, from small to large changes of these values. We then obtained the partial residual plots based on the LS and Wilcoxon fits for each of these stages.

Table 1: Successive changes to the response variable for Model.

Case	Original	Stage 1	Stage 2	Stage 3	Stage 4
3	-.567	5.567	10.56	100.56	1000.56
33	62.44	310.44	-620.44	-620.44	-6200.4
40	67.24	-335.2	-670.2	670.2	-6700.2

Column A of Figure 4 shows the effect these changes had on the LS partial residual plot. Note that limit on the vertical axes were changed so that the bulk of the cases could be plotted. The distortion is obvious. From a clear cubic pattern for the baseline model (the top row of the plots) the LS based partial residual plot becomes more and more distorted as the successive stages are fitted. The clear cubic pattern has been lost even in the first stage (the second row of the plots). By the second stage (third row) the cubic pattern is not identifiable. There is some linear trend in the third stage (fourth row), but in the final stage (last row) there is just noise. On the other hand, the cubic pattern is clearly identifiable in the robust partial residual plots (Column B of Figure 4) for all stages.

6 Conclusion

LS partial residual plots are an important diagnostic tool for exploratory fitting. They are often used to identify unknown functions of the predictors. They are, however, vulnerable to the effect of outliers. One large outlier can severely distort the LS based partial residual plot, making the identification of the unknown function of the predictor difficult to impossible.

Furthermore because the function g can be nonlinear, good data can have the same effect on the LS partial residual plots as outliers; see Example 2.

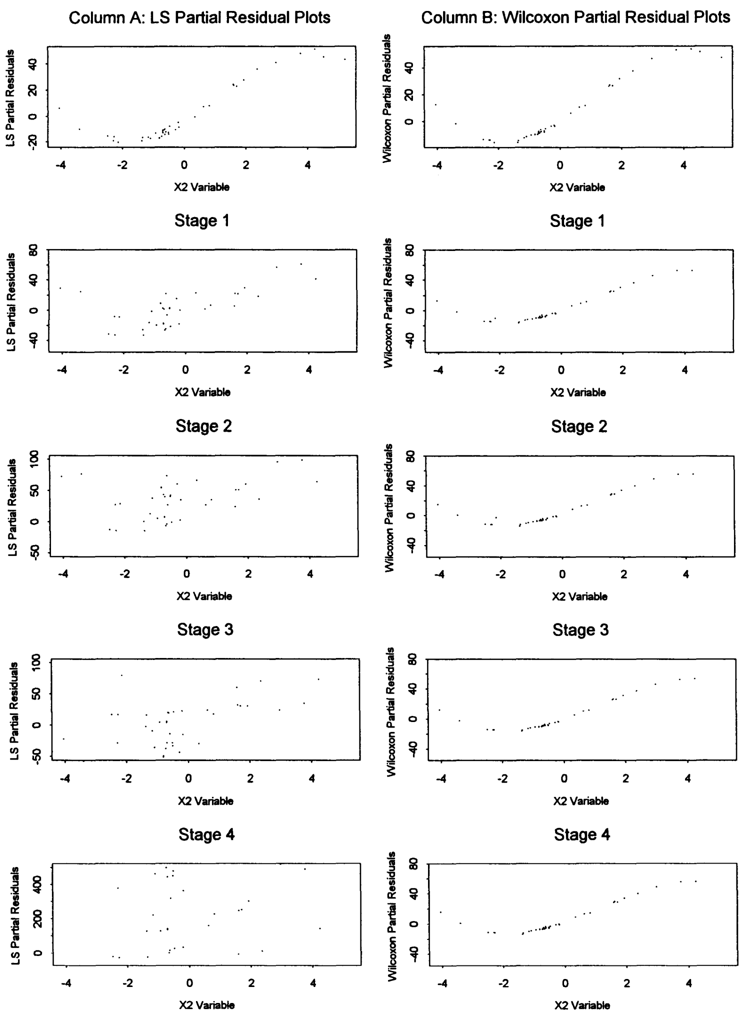

Figure 4: Plots for Sensitivity Study : Column A, LS Partial Residual Plots for Original Data Followed by the LS Plots for Stages 1-4; Column B, Wilcoxon Partial Residual Plots for Original Data Followed by the Wilcoxon Plots for Stages 1-4.

In this paper, we have presented partial residual plots based on robust estimates. As the examples and sensitivity study demonstrated these partial residual plots are effective in exploratory fitting. Furthermore they are not vulnerable to the effect of outliers as their LS counterparts. Also for highly nonlinear situations such as Example 2, they are able to easily identify the unknown function. As the sensitivity study shows, even in the presence of severe outlying observations partial residual plots based on highly efficient robust estimates are able to retain their focus, making the identification of the unknown functions possible.

References

[1] Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988). *The NEW S Language*. New York: Chapman and Hall.

[2] Berk, K. N. and Booth, D. E. (1995). Seeing a curve in multiple regression. *Technometrics* **37** 385-398.

[3] Chang, W., McKean, J. W., Naranjo, J. D. and Sheather, S. J. (1997). High breakdown rank regression. Submitted.

[4] Cook, R. D. (1993). Exploring partial residual plots. *Technometrics* **35** 351-362.

[5] Cook, R. D., Hawkins, D. M. and Weisberg, S. (1992). Comparison of model misspecification diagnostics using residuals from least mean of squares and least median of squares fits. *J. Am. Statist. Assoc.* **87** 419-424.

[6] Ezekiel, M. (1924). A Method for Handling Curvilinear Correlation for any Number of Variables. *J. Am. Statist. Assoc.* 431-453.

[7] Hettmansperger, T. P. (1991). *Statistical Inference Based on Ranks.* New York: Kruger.

[8] Hettmansperger, T. P., McKean, J. W. and Sheather, S. J. (1997). Rank-based analyses of linear models. In *Handbook of Statistics*, Ed. C. R. Rao, Vol. 15. In press.

[9] Huber, P. J. (1981). *Robust Statistics.* New York: Wiley.

[10] Jaeckel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.* **43** 1449-1458.

[11] Kapenga, J. A., McKean, J. W. and Vidmar, T. J. (1988). *RGLM: Users Manual.* Amer. Statist. Assoc. Short course on robust statistical procedures for the analysis of linear and nonlinear models. New Orleans.

[12] Koul, H. L. Sievers, G. L. and McKean, J. W. (1987). An estimator of the scale parameter for the rank analysis of linear models under general score functions. *Scandi. J. Statist.* **14** 131-141.

[13] Larsen, W. A. and McCleary, S. J. (1972). The use of partial residual plots in regression analysis. *Technometrics* **14** 781-790.

[14] Mallows, C. L. (1986). Augmented partial residual plots. *Technometrics* **28** 313-320.

[15] Mansfield, R. E. and Conerly, M. D. (1987). Diagnostics value of residual and partial residual plots. *Am. Statistician* **41** 107-116.

[16] McKean, J. W. Naranjo, J. D. and Sheather, S. J. (1996a). Diagnostics to detect differences in robust fits of linear models. *Comput. Statist.* **11** 223-243.

[17] McKean, J. W. Naranjo, J. D. and Sheather, S. J. (1996b). An efficient and high breakdown procedure for model criticism. *Commun. Statist. Theory and Methods* **25** 2575-2595.

[18] McKean, J. W. and Schrader, R. M. (1984). A comparison of methods for studentizing the sample median. *Commun. Statist.* B **6** 751-773.

[19] McKean, J. W. Sheather, S. J. (1997). Partial residual plots based on robust fitting. Submitted.

[20] McKean, J. W. Sheather, S. J. and Hettmansperger, T. P. (1990). Regression diagnostics for rank-based methods. *J. Am. Statist. Assoc.* **85** 1018-28.

[21] McKean, J. W. Sheather, S. J. and Hettmansperger, T. P. (1993). The use and interpretation of residuals based on robust estimation. *J. Am. Statist. Assoc.* **88** 1254-1263.

[22] McKean, J. W. Sheather, S. J. and Hettmansperger, T. P. (1994). Robust and high breakdown fits of polynomial models. *Technometrics* **36** 409-415.

[23] Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* New York: Wiley.

[24] Stromberg, A. J. (1993). Computing the exact least median of squares estimate and stability diagnostics in multiple linear regression. *SIAM J. Scient. Comput.* **14** 1289-1299.