The Conditional Breakdown Properties of Robust Local Polynomial Estimators

Avi Giloni
Sy Syms School of Business
Yeshiva University
500 West 185th Street
New York, NY 10033

Jeffrey S. Simonoff
Leonard N. Stern School of Business
New York University
44 West 4th Street
New York, NY 10012

December 14, 2001

Abstract

Nonparametric regression techniques provide an effective way of identifying and examining structure in regression data. The standard approaches to nonparametric regression, such as local polynomial and smoothing spline estimators, are sensitive to unusual observations, and alternatives designed to be resistant to such observations have been proposed as a solution. Unfortunately, there has been little examination of the resistance properties of these proposed estimators. In this paper we examine the breakdown properties of several robust versions of local polynomial estimation. We show that for some estimators the breakdown at any evaluation point depends on the observed distribution of observations and the kernel weight function used. Using synthetic and real data, we show how the breakdown point at an evaluation point provides a useful summary of the resistance of the regression estimator to unusual observations.

Key words: Least absolute values; Least median of squares; Least trimmed squares; M–estimation; nonparametric regression.

1 Introduction

Nonparametric regression techniques have been shown in recent years to be very effective at identifying and estimating structure in regression data, without requiring restrictive assumptions on the form of the relationship between the target and predicting variables. Many different approaches to this problem have been suggested; see Simonoff (1996), chapter 5, for discussion of many of the possibilities. In this paper we focus on local polynomial estimation based on a single predictor variable. Let $\{x_i, y_i\}, i = 1, \ldots, n$, be the data set at hand. The underlying model assumed for these data is

$$y_i = \mu(x_i) + \varepsilon_i,$$

with $E(\varepsilon|X = x) = 0$ and $V(\varepsilon|X = x) = \sigma^2(x)$ not necessarily constant. The goal is to estimate $\mu(x)$, the conditional expectation $E(Y|X = x)$.

Local polynomial estimation proceeds by fitting a polynomial locally over a small neighborhood centered at any evaluation point x, based on weighted least squares. The (pth order) local polynomial regression estimator is based on minimizing

$$\sum_{i=1}^{n} [y_i - \beta_0 - \cdots - \beta_p(x_i - x)^p]^2 K\left(\frac{x_i - x}{h}\right). \qquad (1)$$

Here $K(\cdot)$ is the kernel function, typically a smooth symmetric density function that accomplishes local weighting by downweighting the influence of an observation y_i on $\hat{\mu}(x)$ as

x_i gets farther from x. The estimator $\hat{\mu}(x)$ is then the intercept term $\hat{\beta}_0$ from the weighted least squares regression based on the weight matrix

$$W = h^{-1}\,\mathrm{diag}\left[K\left(\frac{x_1 - x}{h}\right), \ldots, K\left(\frac{x_n - x}{h}\right)\right]. \tag{2}$$

The bandwidth h controls the amount of smoothness of $\hat{\mu}(x)$, and can be fixed for all values of x, or locally varied (based on nearest neighbor distance, for example) to allow different levels of smoothing at different locations. Kernel regression corresponds to $p = 0$, and is known to have inferior performance compared to taking $p \geq 1$ (in terms of bias in the boundary region, for example). Assuming a given amount of smoothness of $\mu(\cdot)$, it can be shown that certain local polynomial estimators, combined with appropriate choice of h, can achieve the best possible asymptotic rate of convergence of the estimator to the true curve $\mu(\cdot)$.

As is the case for any estimator based on least squares, local polynomial estimation based on (1) is susceptible to the effects of observations with unusual response values (outliers). If an observed y_i is sufficiently far from the bulk of observed responses for nearby values of x, $\hat{\mu}(x)$ will be drawn towards the unusual response and away from the majority of the points. This has led to the proposal of the use of criteria other than (1) to fit local polynomials. *Lowess* (Cleveland, 1979), and its successor *loess* (Cleveland and Devlin, 1988), are nearest neighbor–based local polynomial estimators that allow the data analyst to downweight the effect of unusual observations. This is done through an iterative process. An ordinary local polynomial estimate is first calculated. Observations then have weights $\{\delta_1, \ldots, \delta_n\}$ attached to them, where the weights decrease smoothly as the absolute residual from the loess fit increases. The updated estimate is then the local polynomial estimate with weights ΔW, where $\Delta = \text{diag}(\delta_1, \ldots, \delta_n)$. This process is then iterated several times. Unfortunately, as Mächler (1989) noted, since the original residuals are based on the ordinary nonrobust loess fit, the robust version still can be sensitive to outliers.

Several authors have suggested the related approach of using a local version of M–estimation. The M–estimate attempts to achieve robustness by replacing (1) with

$$\sum_{i=1}^{n} \rho[y_i - \beta_0 - \cdots - \beta_p(x_i - x)^p] K\left(\frac{x_i - x}{h}\right), \tag{3}$$

where $\rho(\cdot)$ is chosen to downweight outliers (Tsybakov, 1986; Fan, Hu, and Truong, 1994; Welsh, 1994). This is accomplished by choosing $\rho(\cdot)$ to be symmetric, with a unique minimum at zero, so that its derivative $\psi(\cdot)$ is bounded. A typical choice is Huber's function $\psi(x) = \max\{-c, \min(c, x)\}$ with $c = 1.5$ or $c = 2$. Minimization of (3) requires an iterative procedure, and Fan and Jiang (1999) suggested stopping the iterations after one or two steps (this is effectively what lo(w)ess does). The asymptotic properties of the M–estimator (including one– or two–step versions) are broadly similar to those of the least squares version. Starting the iterations at the least squares local polynomial estimator as is typical, however, implies that the estimator is still potentially sensitive to outliers.

True robustness requires an estimator that is not based on the least squares estimator. Wang and Scott (1994) investigated the least absolute values (ℓ_1) version of (3), estimating $\mu(\cdot)$ by minimizing

$$\sum_{i=1}^{n} |y_i - \beta_0 - \cdots - \beta_p(x_i - x)^p| K\left(\frac{x_i - x}{h}\right).$$

Wang and Scott (1994) showed that the estimator is the solution to a linear program, and derived asymptotic theory under specific conditions. See also Chaudhuri (1991) and Yu and Jones (1998).

An alternative approach to robust nonparametric regression is to use quantile smoothing splines (Koenker, Ng, and Portnoy, 1992, 1994). Let $\rho_p(u)$ be the so–called check function $\rho_p(u) = u[p - I(u < 0)]$ for $p \in [0, 1]$. The quantile smoothing spline is defined as a function $\hat{g}_p(x)$ that minimizes

$$\sum_{i=1}^{n} \rho_p(y_i - g(x_i)) + \lambda \int |g''(u)| du$$

over the space of functions with $\int |g''(u)| du < \infty$, where λ is the smoothing parameter. The median quantile ($p = .5$) corresponds to the minimizer of

$$\sum_{i=1}^{n} \frac{|y_i - g(x_i)|}{2} + \lambda \int |g''(u)| du.$$

Portnoy (1997) and Shen (1998) discuss the asymptotic properties of quantile smoothing splines. Other approaches are also possible; see, for example, White (1990) and Hendricks and Koenker (1992).

A basic difficulty with all of this work is that while the asymptotic properties of the methods have been investigated, the robustness properties have not. Thus, while a primary justification of these methods is their supposed resistance to unusual observations, there are no results that actually quantify this resistance. The breakdown of an estimator is the smallest fraction of outliers that can force the estimator to arbitrary values, and is thus a measure of the resistance of the estimator to unusual values. More specifically, the breakdown point of an estimator τ is defined to be

$$\alpha^* = \min \left[\frac{m}{n}; bias(m; \tau, \mathbf{y}, \mathbf{X}) \text{ is infinite} \right],$$

where $bias(m; \tau, \mathbf{y}, \mathbf{X})$ is the maximum bias that can be caused by replacing any m of the original data points by arbitrary values (Donoho and Huber, 1983). An estimator that is not at all resistant to outliers, such as one based on least squares, thus has breakdown $\frac{1}{n}$. In this paper we propose and investigate a locally varying notion of breakdown that is appropriate for local polynomial estimation. By adapting breakdown results from robust linear regression estimation, we derive the robustness properties of various local polynomial estimators, including ones based on least absolute values, least median of squares (LMS) and least trimmed squares (LTS) (Rousseeuw, 1984), and one–step M–estimators from robust local polynomial starting values. In the next section we propose and discuss the derivation of the various breakdown values. Section 3 provides specific examples of conditional breakdown, demonstrating its dependence in certain instances on the local distribution of predictor values. Artificial and real data sets are used in Section 4 to illustrate the properties of the robust estimators, as well as the connection between breakdown and identification of local curvature. Section 5 concludes the paper.

2 Determining the Conditional Breakdown

Since the local polynomial regression estimate $\hat{\mu}(\cdot)$ is implemented by solving many local regression problems, each centered at an evaluation point x, its breakdown properties are defined on a local level as well. We restrict ourselves to kernel functions $K(\cdot)$ that are positive on a bounded interval (typically $[-1,1]$). When we refer to the conditional breakdown, we are merely reflecting that, unlike for parametric models, the breakdown value changes depending on the evaluation point x. Several key points illuminate how the notion of conditional breakdown at a point x can be defined.

The first point to recognize is that since the local polynomial estimate is based on a weighted regression, the breakdown of $\hat{\mu}(x)$ is simply the breakdown of a weighted version of the linear regression method being used, whether that is least squares, least absolute values, least median of squares, least trimmed squares, or M–estimation.

We must also recognize that if an observation becomes unbounded (i.e., $|x_i| \to \infty$), there is no sensible way to define breakdown (or any robustness properties) in the neighborhood of that x_i. The reason for this is that, unlike in the case of a parametric function μ, it isn't meaningful to talk about the "true" $\mu(x)$ when $x \to \pm\infty$, since μ is only defined by local smoothness ($\mu(\infty)$ is not well–defined). For this reason, we will only treat breakdown at an evaluation point x for bounded x.

Consider now the use of a bandwidth h that is not a function of the local design (a constant bandwidth is an obvious example of this, but h also can vary in ways that do not depend on the observations x_i. In this case, contamination in the predictor variable is no longer relevant, since any value of x_i that goes to $\pm\infty$ eventually has zero weight in the local regression; that is, only observations local to x can have an effect on $\hat{\mu}(x)$. We thus can describe robustness and breakdown in this case by considering the finite sample breakdown point of some regression estimator τ with contamination restricted to the dependent variable, or $\alpha\left(\tau, \mathbf{y}|\mathbf{X}\right)$ as denoted by Giloni and Padberg (2001).

The situation when using a bandwidth that varies as a function of the design is more complicated. Consider the most common bandwidth choice of this type, the nearest neighbor bandwidth chosen at x to yield a fixed proportion s of observations with nonzero weights (the closest observations to x). If $1 - s$ is greater than the proportion of observations with $|x_i| \to \infty$, then once again contamination in the predictor variable is not relevant, since eventually these x_i's will no longer be in the neighborhood of x and will have zero weight. On the other hand, if $1 - s$ is less than or equal to the proportion of observations with predictor contamination, at least one contaminated observation will have nonzero weight. In this case we can appeal to known breakdown results for ℓ_1–, LTS, and LMS regression when there is contamination in the predictor. That is, the breakdown at x of local ℓ_1–regression is $\frac{1}{n}$ (the smallest possible value, indicating no robustness), while that of local LTS/LMS is the same as that described below, since LTS and LMS are as resistant to contamination in the predictor as they are to contamination in the target variable. For these reasons, throughout the rest of this paper we refer to the finite sample breakdown point with contamination restricted to the dependent variable simply as the finite sample breakdown point.

In this section, we provide a discussion of the breakdown properties of local polynomial regression where the regression estimator is either the local ℓ_1–regression estimator, the local LTS/LMS estimator, or either estimate followed by a one–step M–estimate. We first focus on the case of local ℓ_1–regression.

2.1 Local ℓ_1–Regression

In order to describe the breakdown properties of local ℓ_1–regression estimators, we first must consider the breakdown point of weighted ℓ_1–regression. Below, we demonstrate that as long as the weights for weighted ℓ_1–regression remain positive and finite, the breakdown point of weighted ℓ_1–regression can be calculated in the same way as in the case of standard ℓ_1–regression. The weights that are used in each of the local regression problems are determined by the selected kernel function and bandwidth, i.e., $w_i = h^{-1}K\left(\frac{x_i - x}{h}\right)$. In the next section, we show that the presence of weights that are not all constant can cause the breakdown to change.

In our discussion below, we assume that we have n observations on the dependent variable y and some number $p \geq 1$ of independent variables $x_1, \ldots, x_p$, each one also providing n values. We denote

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix}, \ \mathbf{X} = \begin{pmatrix} 1 & x_1^1 & \cdot & \cdot & \cdot & x_p^1 \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ 1 & x_1^n & \cdot & \cdot & \cdot & x_p^n \end{pmatrix} = \begin{pmatrix} \mathbf{x}^1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{x}^n \end{pmatrix} = (\mathbf{1}, \mathbf{x}_1, \ldots, \mathbf{x}_p),$$

where $\mathbf{y} \in \mathbb{R}^n$ is a vector of n observations and $\mathbf{X}$ is a $n \times p + 1$ matrix referred to as the *design matrix*. Furthermore, $\mathbf{1}, \mathbf{x}_1, \ldots, \mathbf{x}_p$ are column vectors with n components and $\mathbf{x}^1, \ldots, \mathbf{x}^n$ are row vectors with $p + 1$ components corresponding to the columns and rows of $\mathbf{X}$ respectively.

We denote the set of indexes corresponding to the rows of $\mathbf{X}$ as N. We denote the cardinality of $Z \subset N$ as $|Z|$. Furthermore, $\mathbf{X}_Z = \left(\mathbf{x}^i\right)_{i \in Z}$, $\mathbf{e}_Z = (1, \ldots, 1)^T$ with $|Z|$

components equal to one and $\mathbf{X}_U$, $\mathbf{e}_U$, $\mathbf{X}_L$ and $\mathbf{e}_L$ are defined similarly.

The properties of the finite sample breakdown point for ℓ_1–regression, when fitting a model $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, were first studied by He et al. (1990). To analyze the finite sample breakdown point of a (finite) weighted ℓ_1–regression estimator, we refer to the definition of a design matrix being q–stable. Giloni and Padberg (2001) defined a design matrix $\mathbf{X}$ to be q–stable if there exists $\mathbf{v} \in \mathbb{R}^{|Z|}$ such that

$$\mathbf{v}\mathbf{X}_Z = -\mathbf{e}_U^T\mathbf{X}_U + \mathbf{e}_L^T\mathbf{X}_L \ , \ -\mathbf{e}_Z^T \le \mathbf{v} \le \mathbf{e}_Z^T$$

is satisfied for all $L, U \subseteq N$ with $L \cap U = \emptyset$ and $|L \cup U| \le q$ where $Z = N - U - L$. q–stability is defined by selecting q, the largest nonnegative integer such that the condition is satisfied.

Using this definition of q–stability of a design matrix, Giloni and Padberg (2001) showed that if a design matrix $\mathbf{X}$ is q–stable for some $q \ge 0$, then the finite sample breakdown point with contamination restricted to the dependent variable of ℓ_1–regression is equal to $\frac{q+1}{n}$. Assuming that the weights are finite and positive, generalizing the above result to weighted ℓ_1–regression requires redefining the design matrix $\mathbf{X}$ as follows.

The weighted ℓ_1–regression problem with positive, finite weights w_i can be formulated and solved as a linear program

$$\min \sum_{i=1}^{n} w_i r_i^+ + w_i r_i^-$$

$s.t.$

$$\mathbf{x}^i \boldsymbol{\beta} + r_i^+ - r_i^- = y_i \quad \text{for } i = 1, \ldots, n$$

$$\boldsymbol{\beta} \text{ free, } \mathbf{r}^+ \geq \mathbf{0}, \mathbf{r}^- \geq \mathbf{0}.$$

Equivalently, the objective function can be taken to be the same as in the case of standard ℓ_1–regression, changing the data by setting $\widetilde{y}_i = w_i y_i$ and setting $\widetilde{\mathbf{x}}^i = w_i \mathbf{x}^i$. Thus, to calculate the breakdown of weighted ℓ_1–regression one just needs to determine the q-stability of $\widetilde{\mathbf{X}}$. In the next section, we give breakdown points for local weighted polynomial regression based upon a tricube kernel function and include results for the case where the weights are all constant for the data points which are in each local problem (that is, a uniform kernel function K).

Note that in the case of local ℓ_1–regression (as opposed to the traditional ℓ_1–regression), we are only concerned with the intercept term, i.e., $\hat{\beta}_0$. In such a case, one would like to ascertain that the breakdown results are the same. It might be the case that the restriction to an intercept increases the breakdown point. This turns out not to be the case, as stated in the following proposition, which is proved in the Appendix.

Proposition 1 *The finite sample breakdown point of $\hat{\beta}_0$ of (weighted) ℓ_1–regression is the same as the finite sample breakdown point of (weighted) ℓ_1–regression.*

Thus, determining the q–stability of a design matrix for each local weighted ℓ_1–regression describes the finite sample breakdown point for local ℓ_1–regression. Giloni and Padberg (2001) demonstrated how to calculate the q–stability of a design matrix through both an enumerative procedure as well as by formulating and solving a suitable mixed–integer program (both of these methods can be very computationally intensive, however). We use this methodology to calculate the finite sample breakdown points locally for local ℓ_1–regression in Section 3.

2.2 Local LTS/LMS Regression

Before discussing the breakdown properties of local LTS/LMS polynomial regression, we first briefly describe the LTS and LMS regression estimators. The LTS regression estimator $\hat{\boldsymbol{\beta}}^{LTS}$ is determined by minimizing

$$\sum_{i=1}^{k} \left(r^2\right)_{i:n},$$

where $r_i = y_i - \hat{\beta}_0 - x_i\hat{\beta}_1 - \cdots - x_p\hat{\beta}_p$ and

$$\left(r^2\right)_{1:n} \leq \left(r^2\right)_{2:n} \leq \cdots \leq \left(r^2\right)_{n:n}.$$

Similarly, $\hat{\beta}^{LMS}$ is determined by minimizing

$$\left(r^2\right)_{k:n}.$$

In the case of local polynomial regression with one predictor, the ith residual is

$$r_i = \left(y_i - \hat{\beta}_0 - (x_i - x)\,\hat{\beta}_1 - \cdots - (x_i - x)^p\,\hat{\beta}_p\right).$$

Thus, each local LTS regression problem evaluated at x requires the minimization of

$$\sum_{i=1}^{k} \left(\widetilde{r}^2\right)_{i:n_x}, \tag{4}$$

where $\widetilde{r}_i = \sqrt{h^{-1}K\left(\frac{x_i - x}{h}\right)}r_i$,

$$\left(\widetilde{r}^2\right)_{1:n_x} \leq \left(\widetilde{r}^2\right)_{2:n_x} \leq \cdots \leq \left(\widetilde{r}^2\right)_{n_x:n_x},$$

and n_x is the number of observations with nonzero weight in the span of the kernel centered at evaluation point x.

Alternatively, one could solve the local LTS problem by minimizing

$$\sum_{i=1}^{k} \left(r^2\right)_{i:n_x},$$

where

$$r_i = \left(\widetilde{y}_i - \hat{\beta}_0 - \widetilde{x}_{i1}\hat{\beta}_1 - \cdots - \widetilde{x}_{ip}\hat{\beta}_p\right),$$

where $\widetilde{x}_{ij} = \sqrt{h^{-1}K\left(\frac{x_i - x}{h}\right)}(x_i - x)^j$ and $\widetilde{y}_i = \sqrt{h^{-1}K\left(\frac{x_i - x}{h}\right)}y_i$. The local LMS regression problem can be formulated similarly. Since each local regression problem can be formulated exactly as a standard LTS or LMS regression problem, it is thus evident that the high breakdown properties of LTS/LMS regression hold in the case of local LTS/LMS polynomial regression. Specifically, if there are n_x observations in the local regression around the value x, the conditional breakdown can be as large as $\{\lfloor (n_x - p)/2 \rfloor + 1\}/n_x$, where $\lfloor \cdot \rfloor$ is the greatest integer function.

2.3 One–Step M–Estimates

In this subsection, we discuss the breakdown properties of local one–step M–estimators with starting estimates of either ℓ_1–, LTS, or LMS regression. In standard linear regression models, one–step M–estimators have been used to improve the efficiency of certain high breakdown regression estimators, for example LMS regression (Rousseeuw and Leroy, 1987, p. 129). In the case of local polynomial regression, the only change is that ρ, and thus ψ, are weighted, where the weights are defined by the kernel function $K\left(\frac{x_i - x}{h}\right)$.

The one–step M–estimator that we discuss here is the Bickel (1975) Type 2 estimator, based on the Huber ψ function, $\psi(x) = \max\{-c, \min(c, x)\}$. An initial robust estimate $\hat{\boldsymbol{\beta}}$ is determined, and residuals $\mathbf{r}$ are calculated. The Huber function is then applied to the residuals, yielding

$$r_i^* = \begin{cases} -c\,\hat{\sigma} & \text{if } r_i < c\,\hat{\sigma} \\ r_i & \text{if } |r_i| < c\,\hat{\sigma}, \\ c\,\hat{\sigma} & \text{if } r_i > c\,\hat{\sigma}. \end{cases}$$

Here $\hat{\sigma}$ is a preliminary robust scale estimate, $\hat{\sigma} = 1.483\,\mathrm{median}|r_i|$ (Rousseeuw and Leroy, 1987, p. 44). Let S_0 be the number of observations where $|r_i| < c\,\hat{\sigma}$. Let

$$\hat{\mathbf{X}} = \begin{pmatrix} 1 & x_1 - x \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n - x \end{pmatrix}$$

The one–step M–estimator is then

$$\hat{\boldsymbol{\beta}} + \frac{n}{S_0}\left(\hat{\mathbf{X}}^T \mathbf{W} \hat{\mathbf{X}}\right)^{-1} \hat{\mathbf{X}}^T \mathbf{W} \mathbf{r}^*, \tag{5}$$

where $\mathbf{W}$ is the weight matrix defined in (2).

Within the context of local regression, since breakdown is only based on observations within the span of the kernel, and we are using a bounded kernel, for h not a function of the local design, predictor value contamination ultimately results in the point having zero weight. Thus, the design matrix, $\hat{\mathbf{X}}$ ultimately used in each local problem is bounded. Furthermore, the modified vector of residuals is also bounded by design, and the elements of $\mathbf{W}$ are bounded. It is natural to restrict ourselves to the situation where $\left(\hat{\mathbf{X}}^T\mathbf{W}\hat{\mathbf{X}}\right)$ is invertible, since otherwise the one–step M–estimate (5) is not defined. Thus the maximal bias between the original estimate $\hat{\boldsymbol{\beta}}$ and the one–step M–estimate defined in (5) is bounded. Therefore, the breakdown of the one–step M–estimator remains the same as that of the original estimator, independent as to whether the original estimate was any one of either ℓ_1–, LTS, or LMS regression.

However, as mentioned previously, when, for example, a nearest neighbor bandwidth is utilized, it is possible that predictor value contamination can result in a point having a positive weight. In such a case, both local ℓ_1–regression as well as a one–step M–estimate based upon it will have a finite sample breakdown point of $\frac{1}{n}$. In order to ensure that one–step M–estimates based upon LTS/LMS regression retains the high breakdown property of LTS/LMS regression, it is sufficient to use a redescending M–estimator such as the biweight ψ–function (see Rousseeuw and Leroy, 1987, p. 129). Whenever it is possible to ascertain that predictor value contamination results in a point having a weight of zero, it is sufficient to use a one–step M–estimator based on the Huber ψ–function as described above. Thus, in our examples and in the figures displayed at the end of the paper, we use a one–step

M–estimator based on the Huber ψ–function.

It is obvious that the conditional breakdown is never larger than roughly one–half of the number of observations within the span of the kernel (that is, $n_x/2$). Since consistency of $\hat{\mu}$ requires that $n_x/n \to 0$ as $n \to \infty$, the asymptotic breakdown of any local polynomial estimator with respect to the total sample size is zero. In other words, if any fixed percentage of the total number of observations, no matter how small, is placed at a particular value x_0, and the associated y values are sent to $\pm\infty$, as $n \to \infty$ eventually the number of outliers will exceed the breakdown point. We do not consider this a meaningful criticism of the idea of conditional breakdown for nonparametric regression, since (as will be seen in the next section) the finite sample conditional breakdown provides a useful summary of meaningful differences between the methods for finite samples.

3 Examples of Conditional Breakdown

In this section we describe the relationship between the conditional breakdown properties of local linear estimators and the distribution of predictor values. More precisely, we describe this relationship for local ℓ_1–regression, since (as was noted in the previous section), the breakdown is not a function of the design distribution for local least squares regression (where the breakdown is always $\frac{1}{n_x}$) or local LTS/LMS regression (where it is always as high as roughly 50%).

Figures 1 through 3 give the "maximum resistant" proportions, which we define as $\frac{1}{n_x}$ less than the breakdown point (that is, this is the maximum proportion of local observations that can be outliers without the estimator breaking down). The ith predictor value satisfies $x_i =$

$F^{-1}[i/(n+1)]$, where $F(\cdot)$ is either the uniform [0, 1], standard Gaussian, or exponential (with mean one) cumulative distribution function (that is, the design density is consistent with either a uniform, Gaussian, or exponential pattern, covering what might be considered typical design patterns), with $n = 100$. In each plot breakdown values at a fine grid of values over the range of the data are connected by lines, with the solid line referring to local ℓ_1–regression based on a tricube kernel,

$$K(x) = \begin{cases} (70/81)(1-|x|^3)^3 & \text{if } -1 \le x \le 1, \\ 0 & \text{otherwise} \end{cases}$$

(this is the kernel used in loess), and the dotted line referring to estimation based on a uniform kernel.

Figure 1 gives proportions for a nearest neighbor version of the estimator, where the local bandwidth is adjusted to guarantee 20% of the observations in the span of the kernel (i.e., $n_x = 20$ for all x). The top plot shows that when the design density is uniform and a uniform kernel is used, the local maximum resistant proportion is exactly 25%, corresponding to a breakdown of 30%. That is, up to five outliers can be accommodated within the span of the kernel at any evaluation point. This can be contrasted with the 50% breakdown value of local linear LTS/LMS, which implies that up to nine outliers can be accommodated by those methods (the latter value is appropriate for any design and any kernel when using a nearest neighbor bandwidth, since it is only dependent on the number of observations within the span of the kernel).

What is also striking is that using a uniform kernel has clear advantages from a robustness point of view. The maximum resistant proportion is no longer constant if a tricube kernel is used, being lower near the endpoints of the interval than in the center. Further, even in the center the tricube kernel leads to an estimator with smaller breakdown than using the uniform kernel.

The maximum resistant proportions (and hence the breakdowns) are not constant for either kernel when the design density is not uniform. The middle (Gaussian design) and bottom (exponential design) plots show that generally speaking the estimator is more resistant where the design is densest (this is not a function of n_x, since that is constant here, but rather the design within the span of the kernel). The breakdown values are at least as high when using the uniform kernel compared to using the tricube kernel, and local perturbations in the breakdown that occur when using the tricube kernel are absent. In the sparser design areas, the uniform–based estimator can accommodate two additional outliers compared to the tricube estimator.

Figure 2 gives the corresponding proportions for a fixed bandwidth version of the estimator. The bandwidths were taken to be $h = .1$ (uniform design), $h = .25$ (Gaussian design), and $h = .12$ (exponential design), which results in roughly 20 observations within the span of the kernel in the densest regions. While the broad patterns are similar (with maximum resistant proportions again peaking at around 25%), the proportions are considerably more unstable for fixed h. The reason for this is that, in addition to the local design changing as the evaluation point changes, the actual number of observations in the span of the kernel also changes. Once again a uniform kernel results in a more robust estimator than using a tricube kernel, although in this case there are a few evaluation locations where the pattern reverses. Note that in regions where the design is sparse, and there are relatively few observations, the maximum resistant number drops to zero, indicating that the local linear ℓ_1–estimator is as nonrobust as local least squares.

The figure also gives corresponding figures for the local LTS/LMS estimator (dashed line). As was the case for nearest neighbor estimation, local LTS/LMS is typically more robust than local ℓ_1, but as the design gets sparser (and the number of observations in the span of the kernel gets smaller), the gap between the two methods becomes smaller. Ultimately, in the sparsest regions, local LTS/LMS is as nonrobust as ℓ_1 and least squares. Figure 3 makes things a bit clearer by plotting the actual number of outliers that can be resisted, rather than the proportion. The patterns are now similar to those in Figure 1, although the maximum resistant values are generally lower than those that would be implied by the nearest neighbor bandwidth, as would be expected from the smaller values of n_x.

4 Application to Real and Synthetic Data

In this section we examine several synthetic data sets, and one real data set, to illustrate the properties of the robust local linear estimators. Figures 4 through 7 refer to synthetic data with $n = 100$, predictors on a uniform grid, and $\mu(x) = \sin(1.5\pi x)$. Figure 4 illustrates performance on clean data, with $y_i = \mu(x_i) + \varepsilon_i$ and $\varepsilon_i \sim N(0, .2^2)$. The top plot gives loess estimates based on a nearest neighbor bandwidth covering 30% of the data, where the solid line is the standard loess estimate and the dotted line is the robust version. The true regression curve is given as the dashed line (the same representations, and the ones given below, are also used in Figures 5 through 9). As would be expected, there is little difference between the robust and nonrobust versions for these data.

The middle and bottom plots give estimates for robust local linear estimation. The middle plot refers to least absolute values ℓ_1–estimates, also based on a nearest neighbor bandwidth covering 30% of the data. The solid line is the ℓ_1–estimate based on a tricube kernel, with the dotted line the one–step M–estimate based on that initial estimate (using $c = 2$). The estimates are similar to the loess estimates, although they are less smooth, and are virtually identical to each other. Increasing the bandwidth would not alleviate this roughness, since it is an inherent property of least absolute values estimation (Ellis, 1998, discusses the tendency for ℓ_1–regression lines to change greatly as a result of a small change in the data in ordinary regression, which results in "jumpiness" in this context as observations move into and out of the span of the kernel). The third (dashed) line is the ℓ_1–estimator based on a uniform kernel. It is very similar to the tricube kernel estimate, except that it levels off a bit at the left end of the data.

This tendency is much more pronounced in the bottom plot, which gives the local LTS estimate (the local LMS estimate was very similar, and is omitted here). The local LTS estimate (solid line) flattens out considerably at both ends. This lack of sensitivity to local curvature arises as a direct result of the high breakdown of LTS (or LMS). The high breakdown estimator, being constructed to resistantly fit a straight line, has trouble distinguishing between a change in the regression line (local curvature) and observations off a straight line that are outliers, particularly at the boundary. The one–step M–estimate corrects for this problem. Note also that the LTS estimate is quite jumpy, since high breakdown methods also can change noticeably from small changes in the data (Hettmansperger and Sheather, 1992).

In Figure 5 three observations have been replaced with outliers. As would be expected, the nonrobust version of loess is affected by these outliers, being drawn towards them. All of the robust estimates, on the other hand, are unaffected by the outliers, looking virtually identical to the estimates in Figure 4. In Figure 6 three more outliers are added. While the uniform kernel–based ℓ_1–estimate, and the LTS estimate (and to a lesser extent the tricube–based ℓ_1–estimate) are relatively unaffected, the M–estimates are now drawn towards the outliers. This reflects an interesting issue in using the M–estimate. When the initial estimate is not robust, the M–step can downweight the effects of the outliers, but when the initial estimate is itself robust, the M–step (in attempting to increase the efficiency of the estimator) actually becomes *more* affected by the outliers, and less robust (although, as shown in Section 2.3, the effect on the M–estimate is limited, since the breakdown point is identical to that of the initial estimate).

The number of outliers is increased to nine in Figure 7. While the robust loess and LTS estimates are unaffected (although the one–step M–estimate from the LTS estimate is drawn towards the outliers), the ℓ_1–estimates break down (consistent with their roughly 30% breakdown point). Interestingly, the ℓ_1–estimate based on a uniform kernel is unaffected by the outliers in the neighborhood of them, but exhibits spurious negative lobes on either side of that region.

It is possible that the location and number of the unusual observations in Figures 6 and 7 might reflect a structural change in μ in that neighborhood, rather than the presence of outliers. Ultimately, this doesn't matter; since nonparametric regression estimation hypothesizes smoothness of μ, these observations represent a violation of the underlying hypothesized relationship. It is important that these observations not unduly affect $\hat{\mu}$ so that they can be identified as discrepant. Note also that while a larger bandwidth (resulting in more local observations) could result in even the nine outliers not causing breakdown, this is not a viable strategy, since it would result in drastic oversmoothing.

Figure 8 shows that the robust loess estimate also can be strongly affected by outliers. The data are on a uniform grid with $n = 100$, with $\mu(x) = \sin(5\pi x)$. The response values satisfy $y_i = \mu(x_i) + \varepsilon_i$ and $\varepsilon_i \sim N(0, .1^2)$, except that four observations are adjusted to be outlying. The figure gives estimates based on a nearest neighbor bandwidth covering 12% of the data. Both versions of loess (top plot) are drawn to the outliers, as is the ℓ_1–estimate. The LTS estimate, on the other hand, is completely resistant to the outliers. As in earlier cases, the M–estimate based on the robust estimate is drawn towards the outliers.

Even one outlier can cause trouble for estimators in a sparse design region. In Figure 9 the data are on a grid generated by a Beta(.16, .16) density, and are denser near zero and one and sparse near .5. The response values are generated as in Figure 8, with two isolated outliers replacing two observations. The nonrobust loess estimator is drawn towards the outliers in the top plot, as expected, but the robust estimator is also affected, in that the estimate in the neighborhood of the outlier in the sparse region is pushed farther away from the outlier, resulting in the peak not being estimated well. The ℓ_1–estimator (middle plot) doesn't have a problem with the single outlier in the sparse region, but the ℓ_1–estimator based on a tricube kernel is affected by the outlier in the asymmetric region around $x = .03$. Locally, the design in this region is similar to the exponential design examined in the previous section, and as was indicated there, a uniform kernel leads to a more robust estimator that is unaffected by the outlier. The local LTS estimate (bottom plot) has no problems with either outlier, but the M–estimates based on the ℓ_1– and LTS estimates are drawn towards the outlier in the sparse data region.

We conclude this section with analysis of a real data set that illustrates the difficulties in robust nonparametric estimation when the design is very asymmetric. The data are from a radioimmunoassay calibration study, and relate counts of radioactivity to the concentration of the dosage of the hormone TSH, in micro units per ml of incubator mixture (Tiede and Pagano, 1979). There is a roughly hyperbolic relationship between counts and concentration, with one clear outlier at $(20, 4478)$. Figure 10 gives local linear estimates for these data, based on a nearest neighbor bandwidth covering 65% of the data. The loess estimates (top plot) are both affected by the outlier. While the nonrobust estimate (solid line) is drawn towards the outlier, the robust estimate (dotted line) is driven away from it, resulting in a spurious dip below the bulk of the points. This dip is not a function of choice of the bandwidth, as bandwidths from the smallest possible value (36% of the data) to one leading to clear oversmoothing (90% of the data) all yield estimates exhibiting it. The ℓ_1–estimates based on a 65% nearest neighbor bandwidth (middle plot) are much more satisfactory, particularly the one based on the tricube kernel (solid line). The estimate based on the uniform kernel (dotted line) is slightly jumpier, but still follows the general pattern of the data.

This cannot be said of the LTS estimate (solid line, bottom plot). The high breakdown estimate is unable to recognize the nonoutlying value at $(20, 2396)$, which occurs as the design becomes sparser, as representing a change of curvature, and tracks the downward trend until $x = 26$, where it suddenly jumps up to the correct level of the data. This is a direct result of the high breakdown, which is illustrated by the dotted line. This is an LTS estimate where k in (4) is taken to result in a 10% breakdown, and it is very similar to the ℓ_1–estimate.

5 Conclusion

In this paper we have discussed and examined the robustness properties of local linear estimates based on ℓ_1–, least trimmed squares, and least median of squares. Although the latter estimates have higher breakdown than the ℓ_1–estimate, this is balanced by the tendency for the high breakdown estimates to be less sensitive to changes in local curvature. While one–step M–estimation improves performance when there are not outliers, outlying observations can have a deleterious effect on the estimate.

The jumpiness of the robust estimates is an issue to be addressed. One simple solution would be to input the estimated regression curve to an ordinary local least squares estimate, thereby smoothing it out. An example of this is given in Figure 11. This is a local linear (least squares) estimate derived from the local ℓ_1–estimate based on a tricube kernel in Figure 10. This estimate preserves the robustness of the underlying ℓ_1–estimate, while exhibiting an intuitively appealing smooth form. The theoretical properties of such post–estimation smoothing are an open question. The apparent connection between breakdown and the ability of a robust estimate to adjust to changes in curvature suggests the possibility of choosing the level of robustness in an adaptive way (based on the curvature in the underlying regression curve), allowing for more robustness when the curve has less complex structure.

We have restricted ourselves to univariate nonparametric regression in this paper, but many problems involve multiple predictors. Local polynomial estimation generalizes to more than one predictor, and it would be interesting to investigate the robustness and estimation properties of the robust local polynomial estimators in that context. Additive models (Hastie and Tibshirani, 1990) provide an alternative to direct nonparametric estimation, fitting models of the form

$$y_i = \mu_1(x_{1i}) + \cdots + \mu_r(x_{ri}) + \varepsilon_i,$$

rather than the more general

$$y_i = \mu(x_{1i}, \ldots, x_{ri}) + \varepsilon_i.$$

Outliers are as much of a problem for additive models as in univariate regression, so being able to assess the breakdown of models fit using robust smoothers would be very informative to the data analyst.

Appendix

Proof of Proposition 1. We prove the proposition by contradiction. Since we are considering the finite sample breakdown point, we assume that the design matrix, $\mathbf{X}$ is known, its entries are bounded, and that it is in general position (all $p+1 \times p+1$ submatrices have full rank). We assume that after a design matrix is suitably contaminated, the maximal bias of the (weighted) ℓ_1–regression is infinite but the maximal bias of its constant term remains bounded. In other words, we assume that we are in a situation in which (weighted) ℓ_1–

regression breaks down but its constant term does not. We show that the above assumption is a contradiction.

It is well-known that an ℓ_1–regression estimate is an exact fit to some $p+1$ observations. Let $B_1 \subset N$, $|B_1| = p+1$, and $r\left(X_{B_1}\right) = p+1$ (i.e., B_1 is a subset of $p+1$ indexes of rows of $\mathbf{X}$ such that these rows of $\mathbf{X}$ are of full rank). Denote the (weighted) ℓ_1–regression estimate for data $(\mathbf{X}, \mathbf{y})$ as $\hat{\beta} = \mathbf{X}_{B_1}^{-1}\mathbf{y}_{B_1}$. Assume that the vector of the dependent variable $\mathbf{y}$ is contaminated by some vector $\mathbf{g} \in \mathbb{R}^n$, i.e., $\mathbf{g} = (g_1, \ldots, g_n)^T$, multiplied by some positive constant θ. Further assume that as $\theta \to \infty$

$$\|\mathbf{X}_{B_1}^{-1}\mathbf{y}_{B_1} - \mathbf{X}_{B_2}^{-1}\left(\mathbf{y}_{B_2} + \theta\mathbf{g}\right)\| \to \infty$$

where B_2 is defined similarly to B_1 in the sense that it is the set of indexes of rows of $\mathbf{X}$ such that it defines the optimal ℓ_1–estimate of the regression problem with contaminated data. In other words, the contamination vector $\mathbf{g}$ has caused the (weighted) ℓ_1–regression estimate to break down. Let $\hat{\beta}_0{}^{B_1}$ be the estimate of the constant term based upon $\hat{\beta} = \mathbf{X}_{B_1}^{-1}\mathbf{y}_{B_1}$, i.e. based upon the rows of $\mathbf{X}$ defined by the set of indexes B_1. Let $\left(\mathbf{X}_{B_1}^{-1}\right)^1$ be the first row of the matrix $\left(\mathbf{X}_{B_1}^{-1}\right)$. Assume that $\hat{\beta}_0$ has not broken down even though $\hat{\beta}$ has, i.e., there exists some $K > 0$ such that

$$\|\hat{\beta}_0{}^{B_1} - \hat{\beta}_0{}^{B_2}\| = \|\left(\mathbf{X}_{B_1}^{-1}\right)^1\mathbf{y}_{B_1} - \left(\mathbf{X}_{B_2}^{-1}\right)^1\left(\mathbf{y}_{B_2} + \theta\mathbf{g}\right)\| = K < \infty$$

as $\theta \to \infty$. This implies that

$$\left(\mathbf{X}_{B_2}^{-1}\right)^1\mathbf{g} = 0. \tag{6}$$

In order that $\hat{\beta}_0$ does not break down, (6) must hold for every $\mathbf{g} \in \mathbb{R}^n$ and for every associated $B_2 \subset N$ where $|B_2| = p+1$ and $r\left(\mathbf{X}_{B_2}\right)$ and B_2 defines an optimal ℓ_1–estimate to some contaminated data set. There are (at most) $T = \binom{n}{p+1}$ candidates for a (weighted) ℓ_1–regression estimator. In the case where the design matrix is in general position, there are exactly T candidates. Let B_{2_k} be the kth subset of indexes of rows of $\mathbf{X}$, where $k = 1, \ldots, T$. In order to show that $\hat{\beta}_0$ will indeed break down it suffices to show that there exists some $\mathbf{g_0} \in \mathbb{R}^n$ such that

$$\left(\mathbf{X}_{B_{2_k}}^{-1}\right)^1 \mathbf{g_0} \neq 0 \quad \text{for } k = 1, \ldots, T. \tag{7}$$

However, as long as $\left(\mathbf{X}_{B_{2_k}}^{-1}\right)^1$ does not consist of only zeros, $\mathbf{g}_0$ as in (7) exists. Since $\left(\mathbf{X}_{B_{2_k}}^{-1}\right)^1$ is a row of an inverse of a matrix, it cannot consist of only zeros, and therefore there exists a case in which $\hat{\beta}_0$ breaks down where $\hat{\boldsymbol{\beta}}$ does as well.

We note that contamination need not have the form described above. However, since the breakdown point is a worst case measure and we have shown that under a particular structure of contamination, β_0 breaks down when $\boldsymbol{\beta}$ does, the proposition follows. $\square$

References

Bickel, P.J. (1975), "One–Step Huber Estimates in the Linear Model," *Journal of the American Statistical Association*, **70**, 428–434.

Chaudhuri, P. (1991), "Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation," *Annals of Statistics*, **19**, 760–777.

Cleveland, W.S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, **74**, 829–836.

Cleveland, W.S. and Devlin, S.J. (1988), "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting," *Journal of the American Statistical Association*, **83**, 596–61 0.

Donoho, D.L., and Huber P.J. (1983), "The Notion of Breakdown Point," in *A Festschrift for Erich Lehmann*, eds. P. Bickel, K. Doksum, and J.L. Hodges, Wadsworth, Belmont, CA, 157–184.

Ellis, S.P. (1998), "Instability of Least Squares, Least Absolute Deviation and Least Median of Squares Linear Regression (with discussion)," *Statistical Science*, **13**, 337–350.

Fan, J., Hu, T.–C., and Truong, Y.K. (1994), "Robust Non–Parametric Function Estimation," *Scandinavian Journal of Statistics*, **21**, 433–446.

Fan, J. and Jiang, J. (1999), "Variable Bandwidth and One–Step Local M–Estimator," *Science in China*, **A29**, 1–15.

Giloni, A., and Padberg, M. (2001), "Calculating the Finite Sample Breakdown Point for ℓ_1–Regression," under review.

Hastie, T.J. and Tibshirani, R.J. (1990), *Generalized Additive Models*, Chapman and Hall, London.

He, X., Jureckova, J., Koenker, R., and Portnoy, S. (1990), "Tail Behavior of Regression Estimators and their Breakdown Points," *Econometrica*, **58**, 1195–1214.

Hendricks, W. and Koenker, R. (1992), "Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity," *Journal of the American Statistical Association*, **87**, 58–68.

Hettmansperger, T.P. and Sheather, S.J. (1992), "A Cautionary Note on the Method of Least Median Squares," *American Statistician*, **46**, 79–83.

Koenker, R., Ng, P.T., and Portnoy, S. (1992), "Nonparametric Estimation of Conditional Quantile Functions," in L_1 *Statistical Analysis and Related Methods*, ed. Y. Dodge, Elsevier, Amsterdam, 217–229.

Koenker, R., Ng, P.T., and Portnoy, S. (1994), "Quantile Smoothing Splines," *Biometrika*, **81**, 673–680.

Mächler, M.B. (1989), *"Parametric" Smoothing Quality in Nonparametric Regression: Shape Control by Penalizing Inflection Points. Ph.D. Thesis*, Swiss Federal Institute of Technology (ETH), Zürich.

Portnoy, S. (1997), "Local Asymptotics for Quantile Smoothing Splines," *Annals of Statistics*, **25**, 414–434.

Rousseeuw, P.J. (1984), "Least Median of Squares Regression," *Journal of the American Statistical Association*, **79**, 871–880.

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

Shen, X. (1998), "On the Method of Penalization," *Statistica Sinica*, **8**, 337–358.

Simonoff, J.S. (1996), *Smoothing Methods in Statistics*, Springer–Verlag, New York.

Tiede, J.J. and Pagano, M. (1979), "The Application of Robust Calibration to Radioimmunoassay," *Biometrics*, **35**, 567–574.

Tsybakov, A.B. (1986), "Robust Reconstruction of Functions by the Local–approximation method," *Problems of Information Transmission*, **22**, 133–146.

Wang, F.T. and Scott, D.W. (1994), "The L_1 Method for Robust Nonparametric Regression," *Journal of the American Statistical Association*, **89**, 65–76.

Welsh, A.H. (1994), "Robust Estimation of Smooth Regression and Spread Functions and Their Derivatives," *Statistica Sinica*, **6**, 347–366.

White, H. (1990), "Nonparametric Estimation of Conditional Quantiles Using Neural Networks," *Proceedings of Computing Science and Statistics: Eighth Annual Symposium on the Interface*, Springer–Verlag, New York, 190–199.

Yu, K. and Jones, M.C. (1998), "Local Linear Quantile Regression," *Journal of the American Statistical Association*, **93**, 228–237.

Figure 1: Local maximum resistant proportion values for nearest neighbor local linear ℓ_1–regression. Top plot refers to uniform design density, middle plot refers to Gaussian design density, and bottom plot refers to exponential design density. The solid line refers to using a tricube kernel and the dotted line refers to using a uniform kernel.

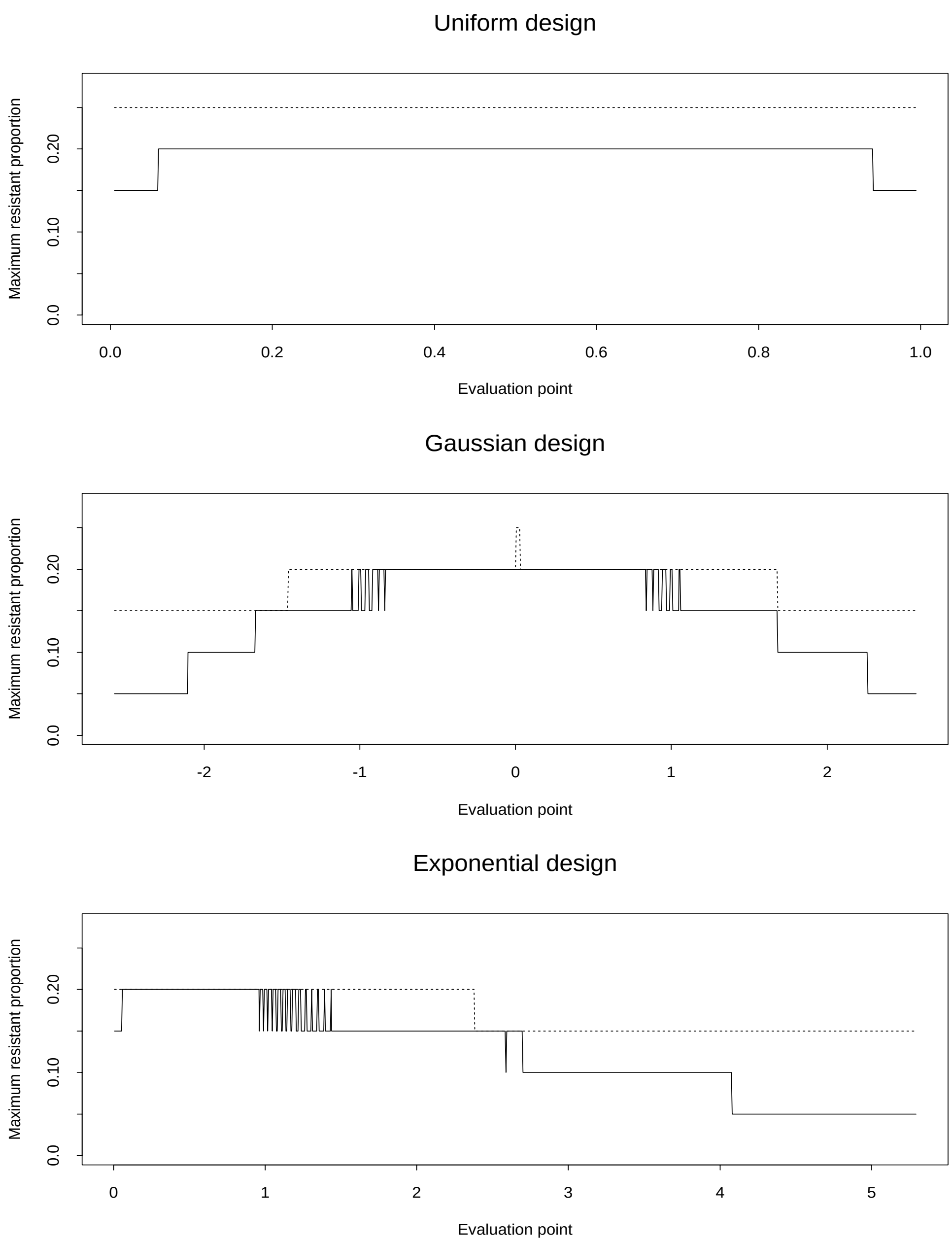

Figure 2: Local maximum resistant proportion values for fixed bandwidth local linear ℓ_1– and LTS/LMS regression. Top plot refers to uniform design density, middle plot refers to Gaussian design density, and bottom plot refers to exponential design density. The solid line refers to ℓ_1 using a tricube kernel, the dotted line refers to using a uniform kernel, and the dashed line refers to local LTS/LMS.

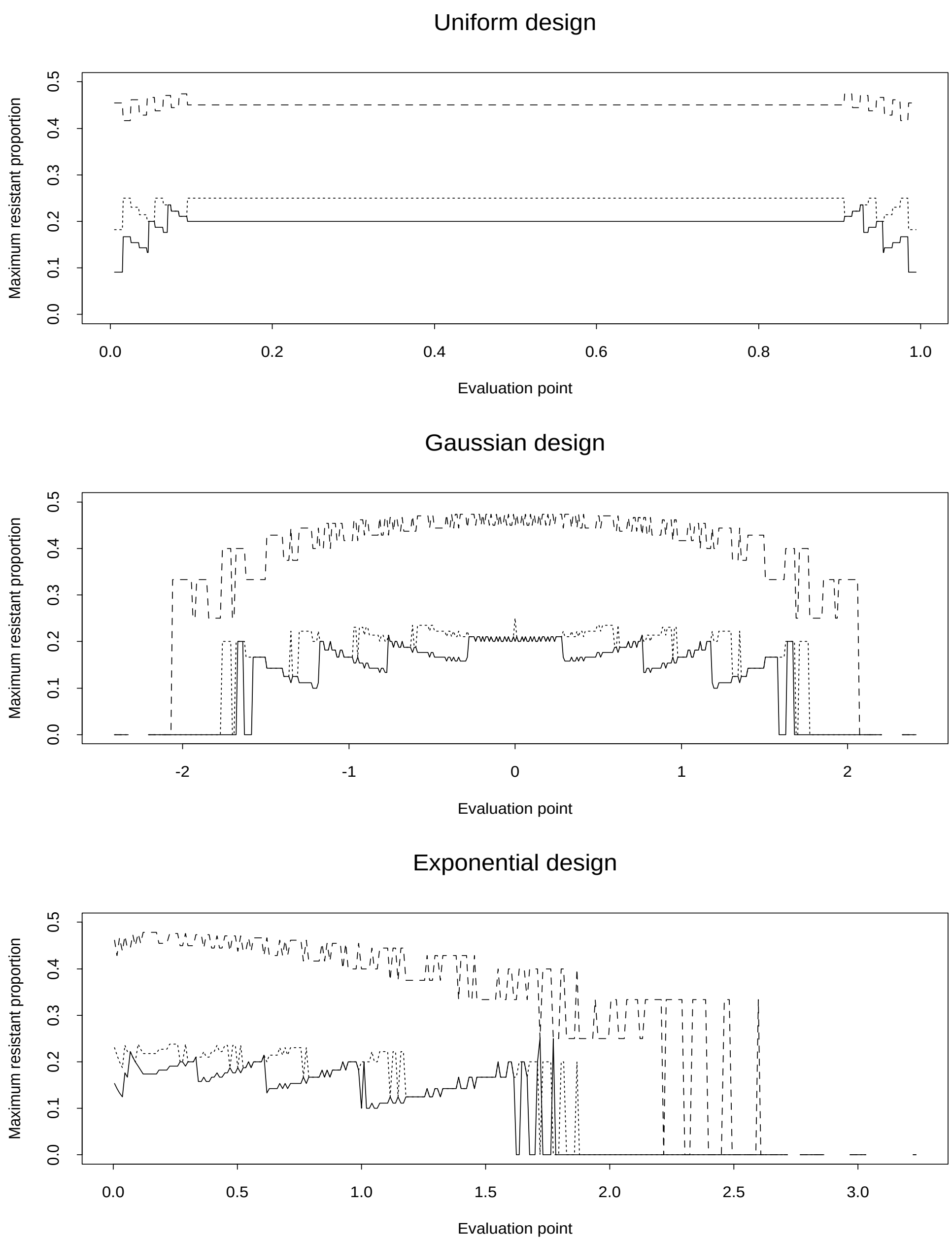

Figure 3: Local maximum resistant number values for fixed bandwidth local linear ℓ_1–regression. Top plot refers to uniform design density, middle plot refers to Gaussian design density, and bottom plot refers to exponential design density. The solid line refers to ℓ_1 using a tricube kernel, the dotted line refers to using a uniform kernel, and the dashed line refers to local LTS/LMS.

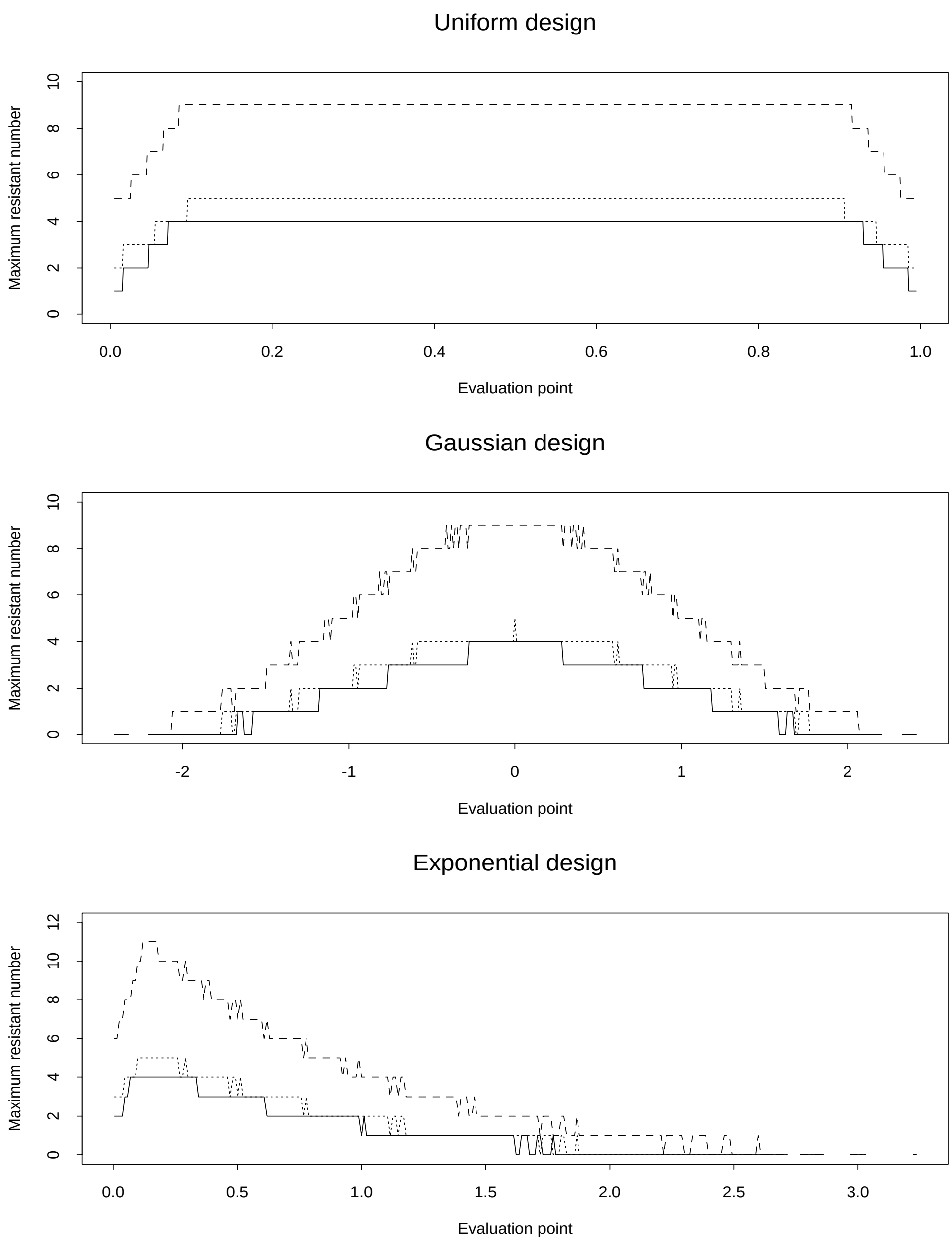

Figure 4: Local regression estimates for clean synthetic data. Top plot refers to nonrobust (solid line) and robust (dotted line) versions of loess, along with true curve (dashed line), middle plot refers to local ℓ_1–estimation based on tricube kernel (solid line) and uniform kernel (dashed line) and one–step M–estimate based on tricube kernel (dotted line), and bottom plot refers to local LTS estimation (solid line) and one–step M–estimate (dotted line).

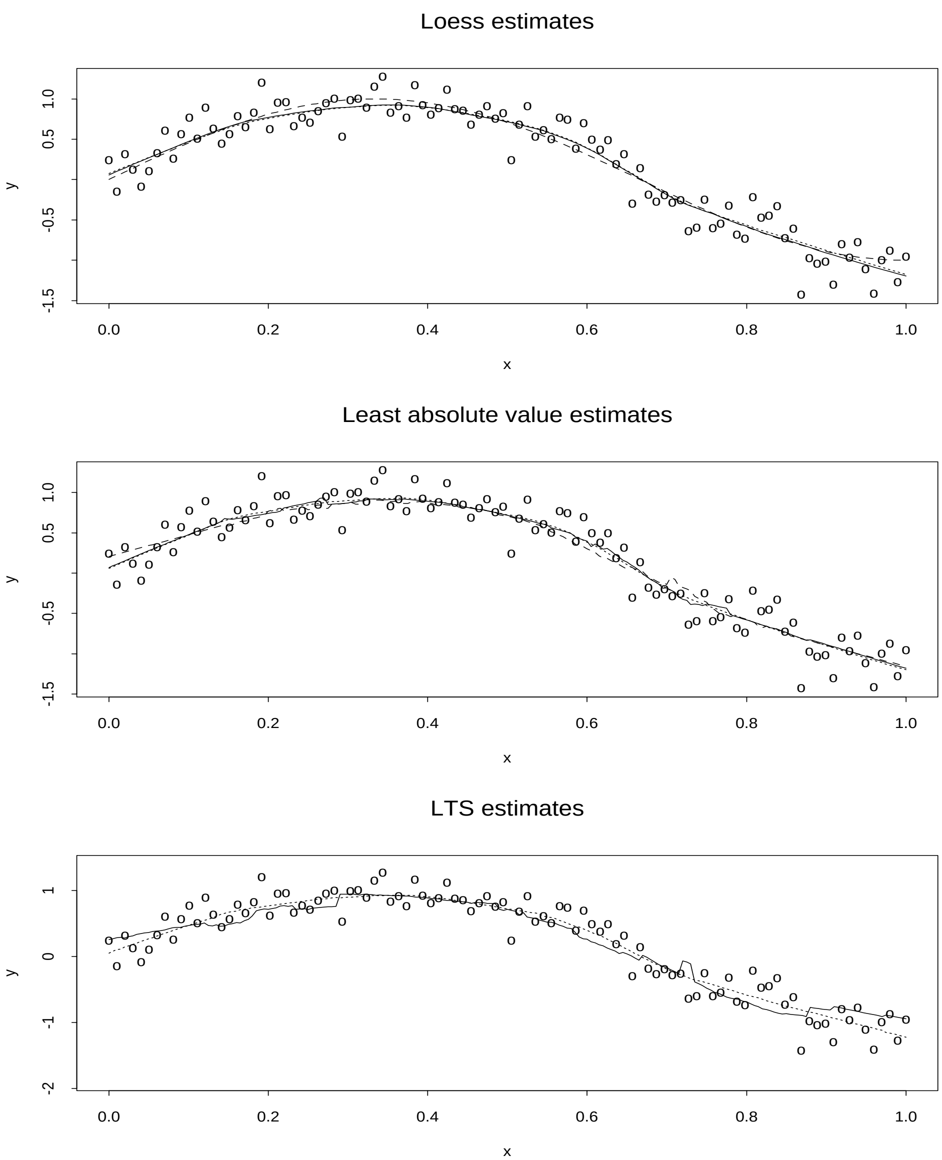

Figure 5: Local regression estimates for synthetic data with three outliers. Plots and curves are as in Figure 4.

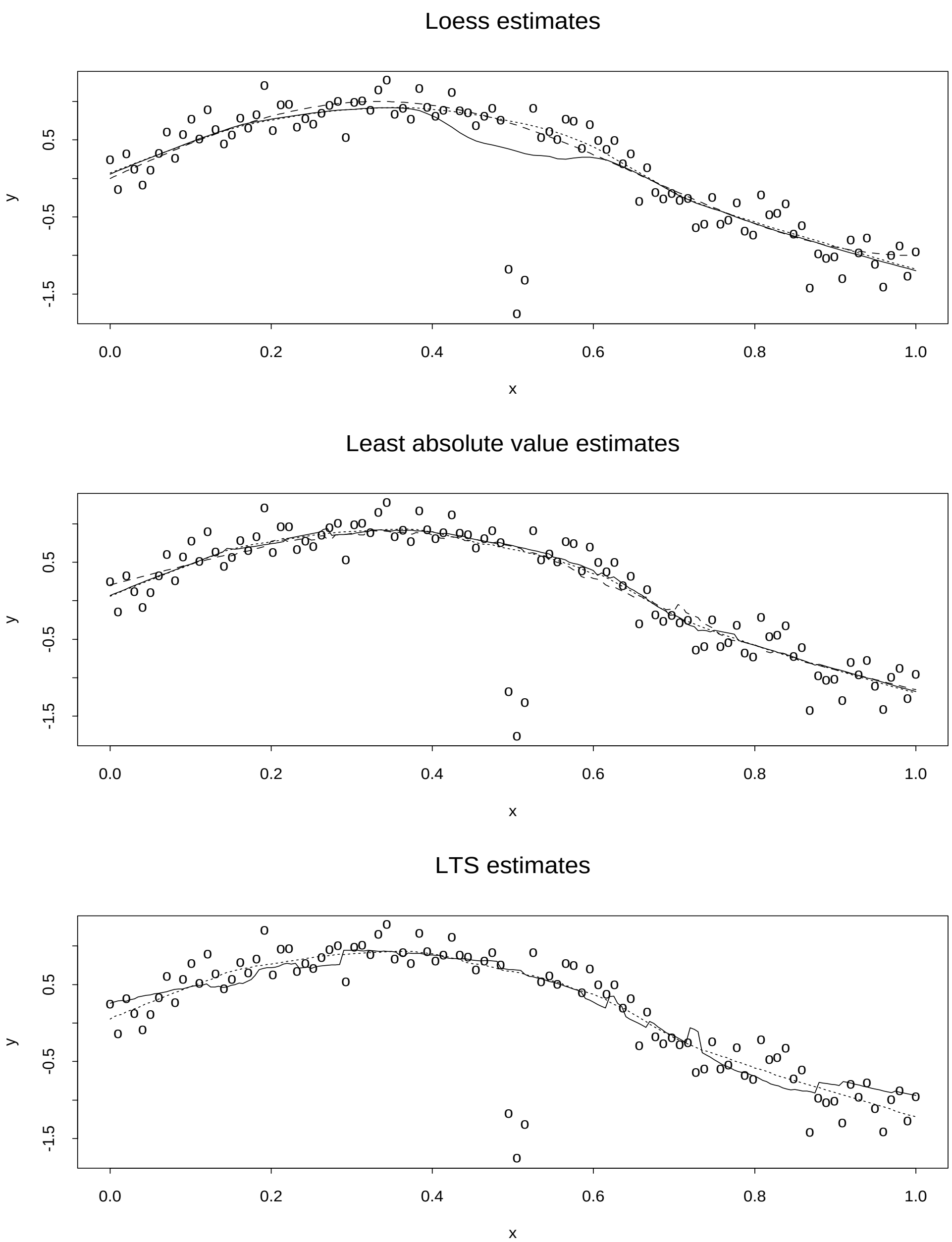

Figure 6: Local regression estimates for synthetic data with six outliers. Plots and curves are as in Figure 4.

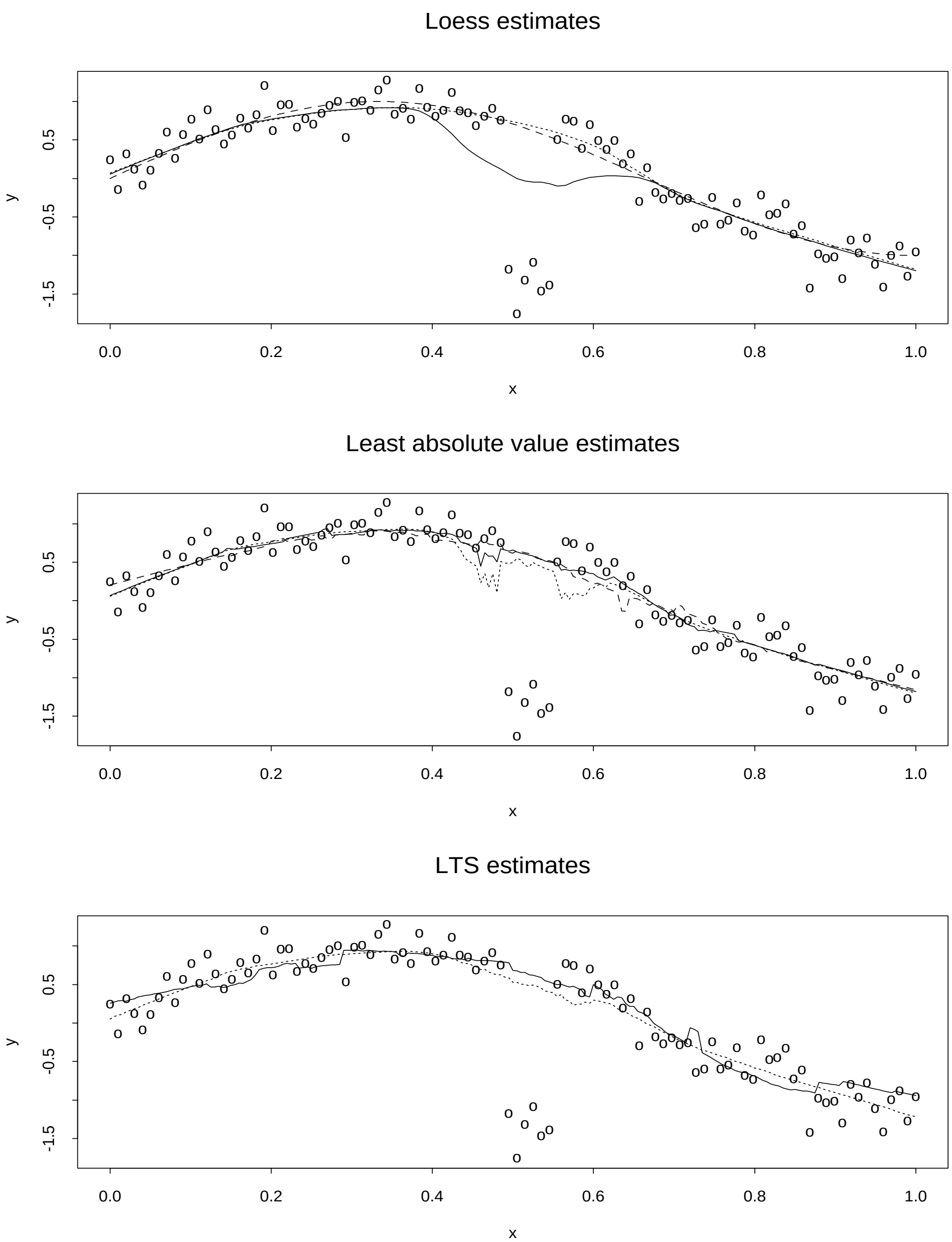

Figure 7: Local regression estimates for synthetic data with nine outliers. Plots and curves are as in Figure 4.

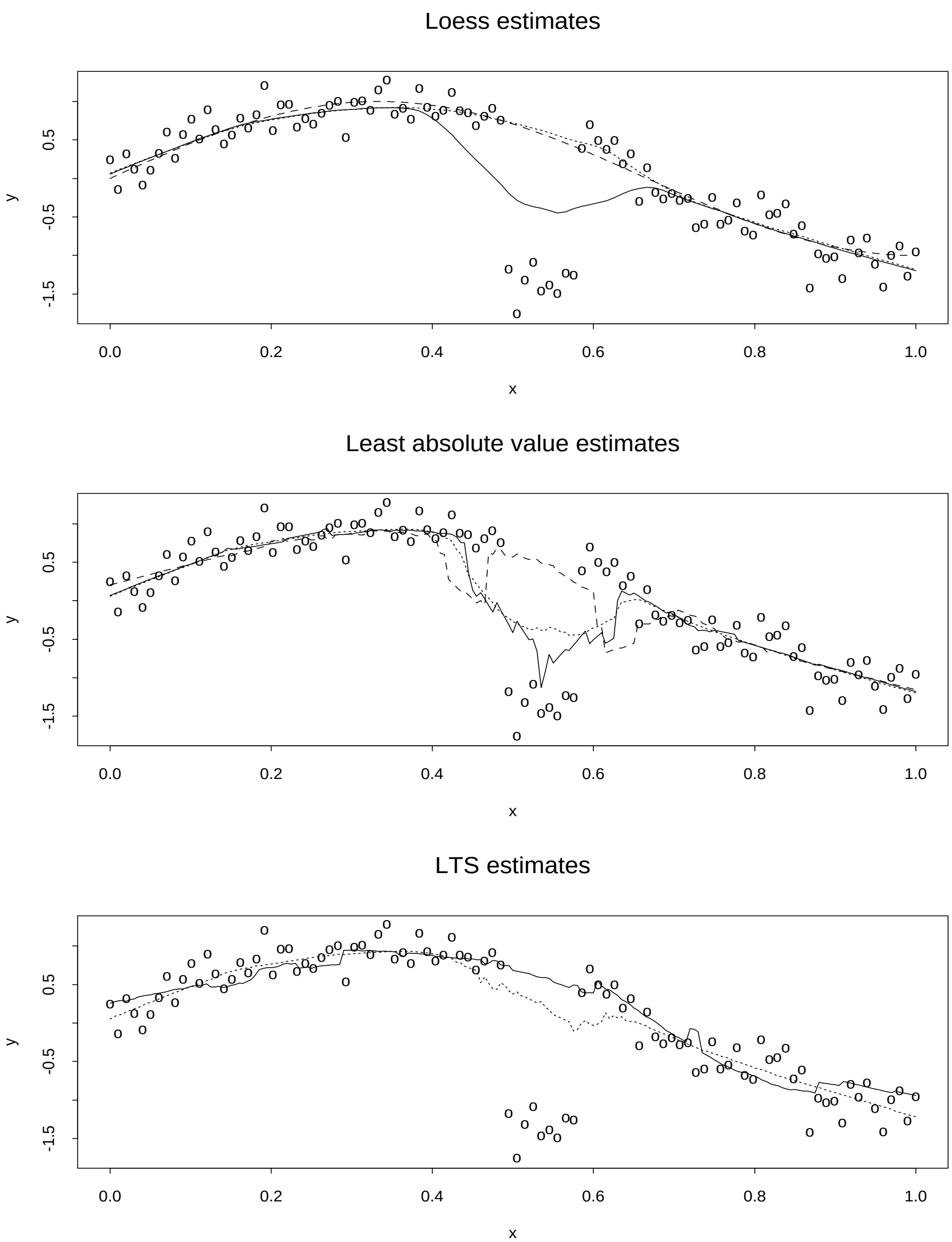

Figure 8: Local regression estimates for synthetic data with four outliers. Plots and curves are as in Figure 4.

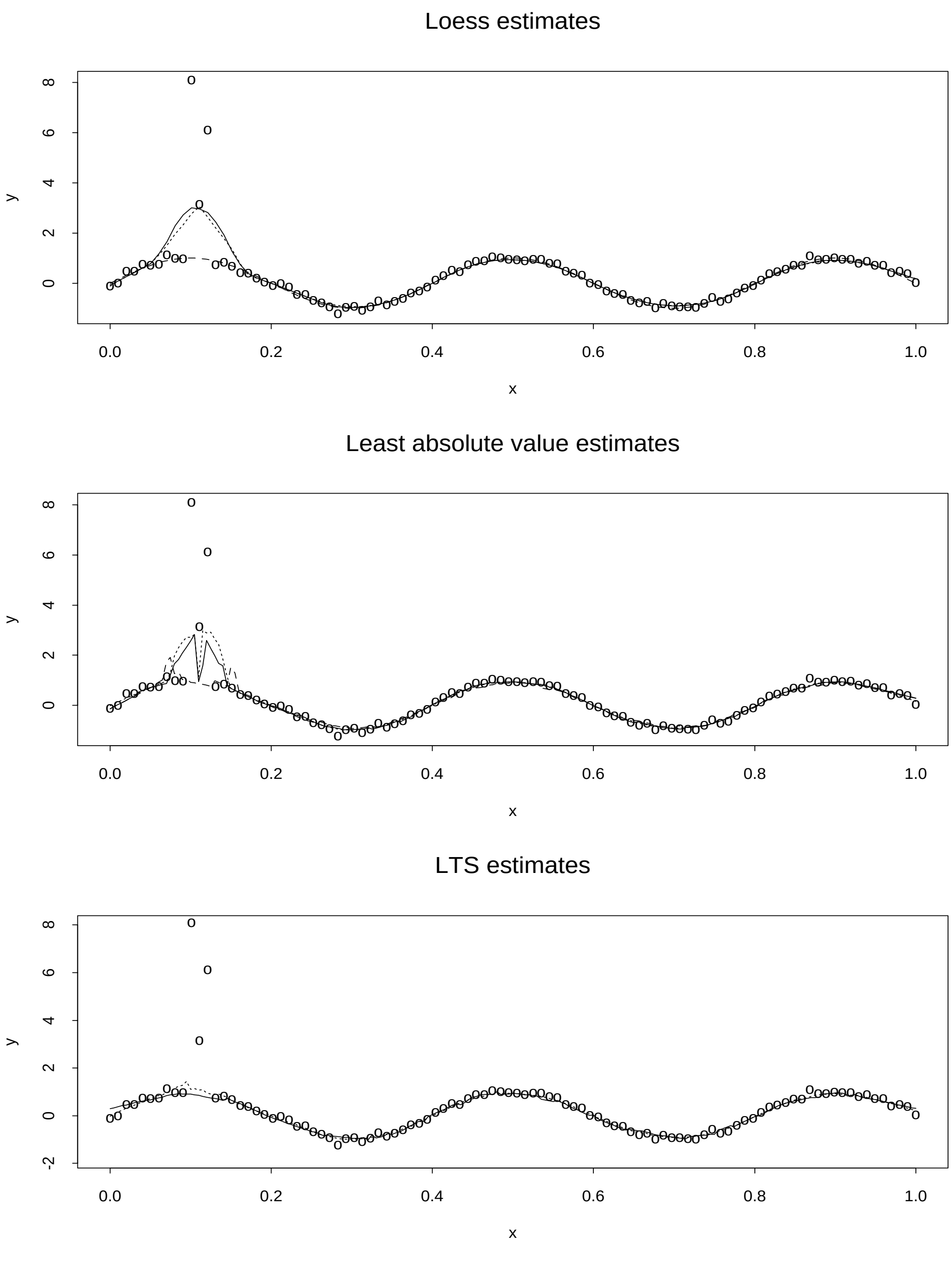

Figure 9: Local regression estimates for synthetic data with nonuniform design and two outliers. Plots and curves are as in Figure 4.

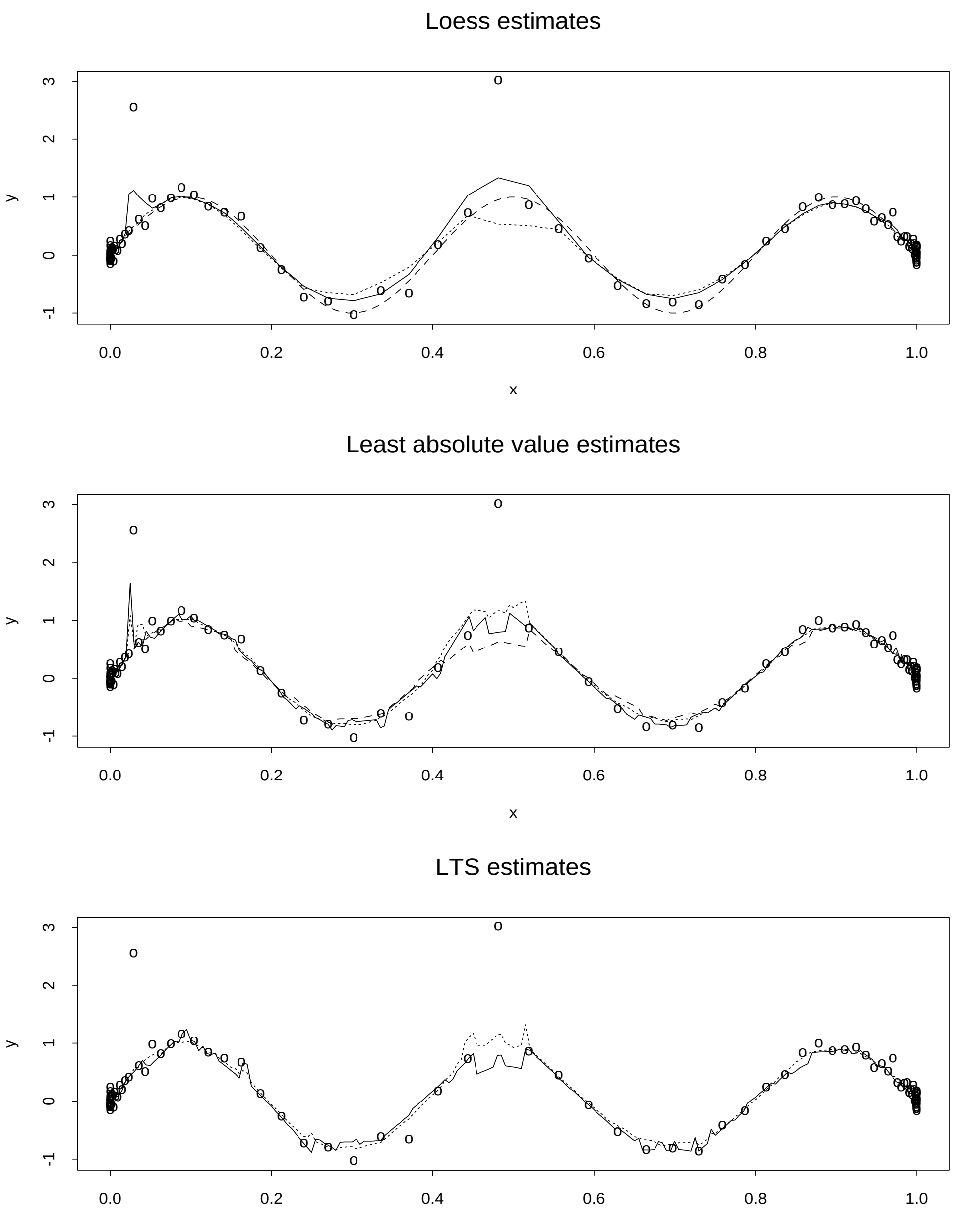

Figure 10: Local regression estimates for calibration data. Top plot refers to nonrobust (solid line) and robust (dotted line) versions of loess, middle plot refers to local ℓ_1–estimation based on tricube (solid line) and uniform (dashed line) kernels, and bottom plot refers to local LTS estimation based on 50% breakdown (solid line) and 10% breakdown (dotted line).

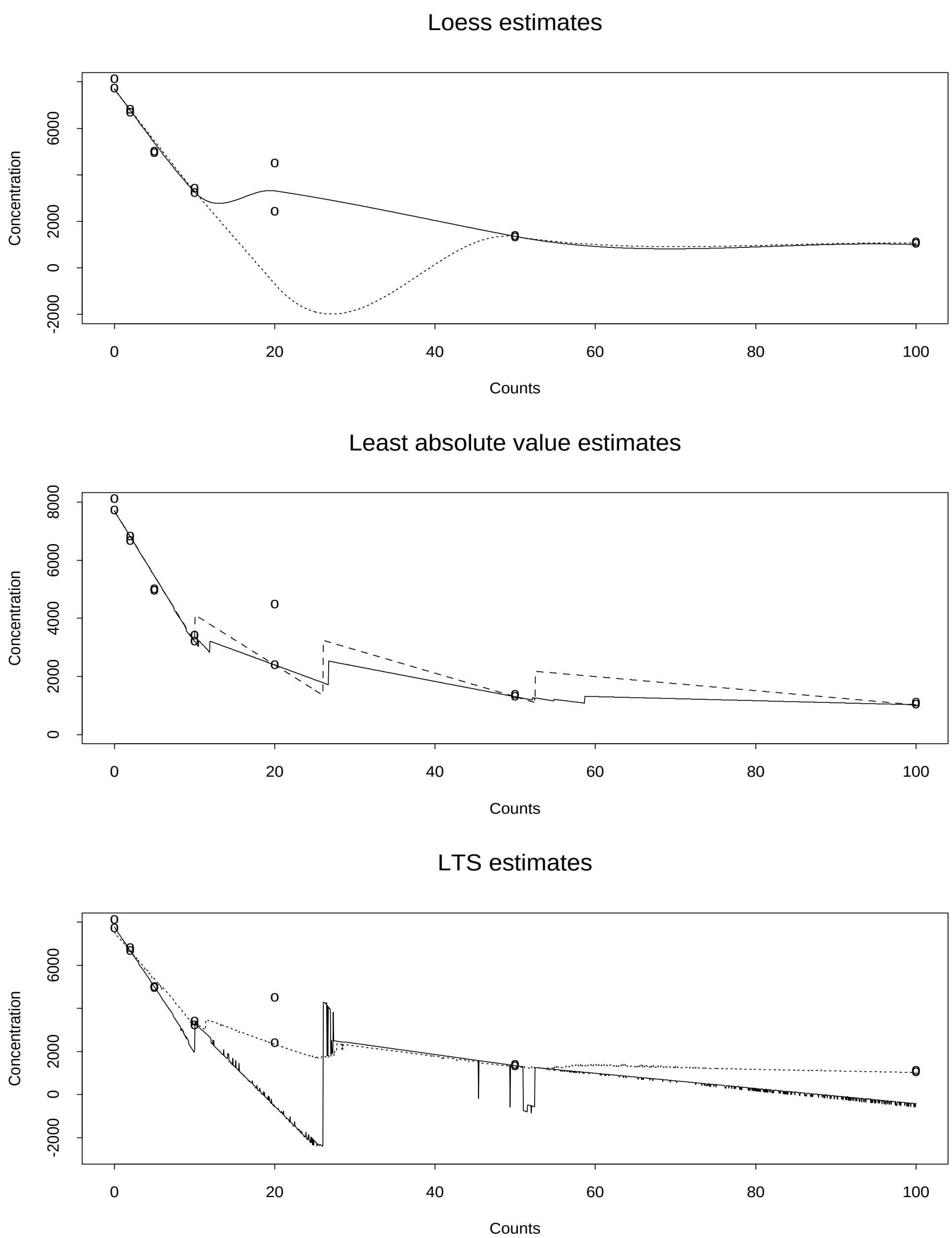

Figure 11: Local ℓ_1–estimate for calibration data, after having curve smoothed using local least squares estimate.