ROBUST ANALYSIS OF M-ESTIMATORS OF NONLINEAR MODELS

by
Shawn Patrick Neugebauer

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE
in
Electrical Engineering

Approved:

Dr. Lamine Mili
(Principal Advisor)

Dr. Jeffrey H. Reed

Dr. Clint W. Coakley
Department of Statistics

August 1996
Blacksburg, Virginia

Keywords: nonlinear regression, nonlinear model estimation, robust regression,
influence function, leverage, outlier

Robust Analysis of M-Estimators of Nonlinear Models

by
Shawn Patrick Neugebauer
Committee Chairman: Dr. Lamine Mili
Electrical Engineering

(ABSTRACT)

Estimation of nonlinear models finds applications in every field of engineering and the sciences. Much work has been done to build solid statistical theories for its use and interpretation. However, there has been little analysis of the tolerance of nonlinear model estimators to deviations from assumptions and normality.

We focus on analyzing the robustness properties of M-estimators of nonlinear models by studying the effects of deviations from assumptions and normality on these estimators. We discuss St. Laurent and Cook's Jacobian Leverage and identify the relationship of the technique to the robustness concept of influence. We derive influence functions for M-estimators of nonlinear models and show that influence of position becomes, more generally, influence of model. The result shows that, for M-estimators, we must bound not only influence of residual but also influence of model. Several examples highlight the unique problems of nonlinear model estimation and demonstrate the utility of the influence function.

Acknowledgements

I would like to thank several people that have provided support and guidance during the research and writing of this thesis.

In particular, I would like to thank my advisor Dr. Lamine Mili for his support and patience. His encouragement and critique of my work has been greatly appreciated. I look forward to continued collaboration and friendship.

For many years, I have been privileged to be a friend and colleague of Mr. Thomas E. Shrimpton. Our wide-ranging conversations have been much-enjoyed and a constant source of ideas and excitement. I wish Tom the greatest success and enjoyment in his future studies, employment, and personal life.

Over the past few years, I have watched, with great pride, Ms. Jennifer M. Neugebauer pursue higher education and begin her career. I wish Jennifer the best success in her future studies and employment.

My parents have been a source of continual encouragement. I thank them for providing the atmosphere and instilling the spirit which has allowed me to pursue advanced study. I dedicate this thesis to them.

Contents

List of Tables

List of Figures

Chapter 1

Overview

1.1 Introduction

Estimation of nonlinear models finds applications in every field of engineering and the sciences [1], [2], [3], [4], [5]. Much work has been done to build solid statistical theories on its use and interpretation [6], [7], [8], [9], [10], [11], [12]. Pre-existing statistical techniques, including robust estimators, have been used directly for nonlinear model estimation problems [13], [6], [14]. Over, the past two decades, attention has been increasingly focused on the unique problems of nonlinear model estimation and analysis.

Despite the large body of nonlinear statistical theory, there has been little analysis of the tolerance of these estimators to deviations from assumptions and normality. Surprisingly little of the large volume of existing robust statistical theory has been applied to the problems of nonlinear model estimation. Perhaps owing to increasingly powerful computation environments, the past few years have seen the emergence of a robust theory of nonlinear model estimation [15], [16], [17], [18], [19], [20], [21], [22]. Existing robustness tools, such as the breakdown point, are being tailored to nonlinear model estimation and studied in the context of this very challenging environment [23], [14], [24]. Future work promises to require a diverse collection of mathematical and analytical techniques.

We focus on analyzing the robustness properties of M-estimators of nonlinear models by studying methods for analyzing the effects of deviations from assumptions and normality on these estimators. We discuss St. Laurent and Cook's Jacobian Leverage [25], [26], and describe the relationship of the technique to the robustness concept of influence. We derive influence functions for M-estimators of nonlinear models, including nonlinear regression, and show that influence of position becomes, more generally, influence of model in nonlinear model estimation. The result shows that, for M-estimators, we must bound not only influence of residual but also influence of model.

1.2 Overview

In Chapter 2, we review basic concepts of robust estimation theory. We define the basic nonlinear model, estimators, and functionals. We define tools used to evaluate the robustness of an estimator, such as influence function, gross-error sensitivity, breakdown point, and bias curve.

In Chapter 3, we study St. Laurent and Cook's Jacobian Leverage for nonlinear regression. After discussing the relationship of the technique to the influence function, we discuss several limitations of the technique including its origin in least-squares estimation and its inability to identify multiple leverage points.

In Chapter 4, we derive influence functions for M-estimators of nonlinear models. Specifically, we derive the influence function for M-estimators of nonlinear location and we derive the influence function for an M-estimator of nonlinear regression. We discuss the interpretation of the results and conjecture on techniques for redefining the concept of leverage in nonlinear model estimation.

In Chapter 5, we present several examples which show the unique problems of nonlinear model estimation and demonstrate the utility of the new influence functions. Finally, in Chapter 6, we conclude and offer suggestions for future work.

Chapter 2

Robustness Concepts

2.1 Introduction

The goal of *regression* is to describe the structure best fitting the data. Additionally, we require the identification of deviating data points or sets of data points. Often, a mathematical expression, or *model*, is available (or generated) which best explains the data. When estimating the parameters of the model, we desire that the estimation technique have several properties, including unbiasedness, consistency, efficiency, and robustness. *Robustness* is concerned with evaluating and improving the stability of estimation techniques when data points deviate from assumptions. This requires techniques for measuring both the local and global robustness of an estimator. Using these concepts, the goal of *robust regression estimation* is to estimate the model best fitting the bulk of the data.

In this chapter we define the general nonlinear model estimation problem. We review the tools used for studying the local and global robustness of estimators for linear models. Particularly, we discuss the derivation, definition, and interpretation of the influence function, which measures the local robustness of an estimator. Also, we review the definitions and interpretation of breakdown point and bias curve, which are tools for evaluating the global robustness of an estimator and thus complementary to the influence function.

2.2 Models, Estimators, and Functionals

2.2.1 Definition of Nonlinear Model Estimation

Let $\{(\underline{x}_i, y_i) : i = 1, \ldots, n\}$ be a sequence of i.i.d. random variables such that

$$y_i = \eta(\underline{x}_i, \underline{\theta}) + e_i, \quad i = 1, \ldots, n \tag{1}$$

where $y_i \in \Re$ is the ith observation value and $\underline{x}_i \in \mathcal{X} \subset \Re^p$ is the ith row of the $n \times p$ design matrix. The variables $\underline{x}_i$ are called the *explanatory variables* or the *carriers*; the variable y_i is called the *response variable.* Let $\underline{\theta} \in \Theta \subset \Re^p$ be the p-vector of unknown parameters and $e_i \in \Re$ be the ith *error.* Finally, let $\eta(\underline{x}, \underline{\theta}) : \mathcal{X} \times \Theta \rightarrow \Re$ be the *model function.* Estimates are denoted by a hat.

The purpose of regression is to fit the model (1) to observations. We desire an estimate $\hat{\underline{\theta}}$ of the parameters $\underline{\theta}$ such that some criterion is satisfied, e.g. minimization of the sum of the absolute value of the residuals, $r_i = y_i - \eta(\underline{x}_i, \hat{\underline{\theta}})$, also known as least absolute values or L_1 regression. When the model $\eta(\underline{x}, \underline{\theta})$ is linear with respect to $\underline{x}$ and/or $\underline{\theta}$, e.g. $\eta(\underline{x}, \underline{\theta}) = \underline{x}\,\underline{\theta}$, we call this process *linear regression.* When the model $\eta(\underline{x}, \underline{\theta})$ is nonlinear with respect to $\underline{x}$ and/or $\underline{\theta}$, we call the process *nonlinear regression.*

2.2.2 Definition of Statistical Functional

Let $X_1, \ldots, X_n$ be a sample from a population with distribution function F and let $T_n = T_n(X_1, \ldots, X_n)$ be a statistic. When T_n can be written as a functional T of the empirical distribution function F_n, $T_n = T(F_n)$ where T does not depend on n, then we call T a *statistical functional.* The domain of T is assumed to contain the empirical distribution functions F_n for all $n > 1$ and the population distribution function F. The range of T is assumed to be $\Re$.

2.2.3 Definition of M-estimator

The specific class of statistical functionals we study in this thesis are the M-estimators. Let ψ be a real-valued function and let T_n be defined implicitly by

$$\sum_{i=1}^{n} \psi(X_i, T_n) = 0 \; . \tag{2}$$

The corresponding functional is defined as the solution $T(F) = \theta$ of

$$\int \psi(x, \theta) dF(x) = 0 \ . \tag{3}$$

Functionals of this form are called *M-estimators*. We consider functionals that are Fisher consistent, that is

$$T(F) = \theta \quad \text{for all } \theta \in \Theta \tag{4}$$

meaning that the estimator asymptotically measures the true value. M-estimators are easily extended to nonlinear regression as defined in [13] and [6].

2.3 Influence Function

2.3.1 Definition of the Influence Function

Following [27], [28], and [29], we say that a functional T is Gâteaux differentiable at F if there exists a linear functional L_F such that for all H

$$\lim_{t \to 0} \frac{T(G) - T(F)}{t} = L_F(H - F) \tag{5}$$

where $G = F_t = (1-t)F + tH$. L_F is called the Gâteaux derivative of T at F.

The *influence function* of T is defined as (5) when $H = \Delta_x$, the distribution with unit mass at x, yielding

$$IF(x, T, F) = \lim_{t \to 0} \frac{T((1-t)F + t\Delta_x) - T(F)}{t} \tag{6}$$

for $x \in \mathcal{X}$ where the limit exists. For example, to find the influence function of an M-estimator given implicitly in functional form, we substitute $G = F_t$ into the definition of the estimator, evaluate the derivative with respect to t at $t = 0$, and solve for the influence function. The influence function describes the effect of an infinitesimal contamination at x on the estimator T. We will develop the influence function of M-estimators of nonlinear models and study it for situations which cause unbounded influence.

2.3.2 Hampel's Empirical Influence Function

The influence function (6) is based on asymptotic functionals. We can also define a finite-sample influence function. Hampel [29] defines two versions, one by addition of an observation and one by replacement. Given an estimator T_n, a sample $\{x_1, \ldots, x_{n-1}\}$, and an additional observation x, define the empirical influence function as

$$EIF_n(x; T_n) = T_n(x_1, \ldots, x_{n-1}, x) \ . \tag{7}$$

We can also define the empirical influence function by replacement, as follows. Given a sample of n observations, replace one, say x_n, by an arbitrary x and define the empirical influence function as in (7). The empirical influence function can be a useful method for evaluating an estimator if closed-form derivations of the influence function prove difficult.

2.3.3 Tukey's Sensitivity Curve

Tukey [30] popularized the sensitivity curve as a tool for evaluating the effect on an estimate of perturbing an observation at a finite sample. Tukey defined a version for addition of an observation as follows. Given an estimator T_n and a sample $\{x_1, \ldots, x_{n-1}\}$, define the sensitivity curve as a function of an additional observation x scaled by the sample size n. Formally, we have

$$SC_n(x; T_n) = n\left[T_n(x_1, \ldots, x_{n-1}, x) - T_{n-1}(x_1, \ldots, x_{n-1})\right] \ . \tag{8}$$

It is clear by comparison with (7) that this can be interpreted as a scaled and shifted version of the empirical influence function.

2.4 Gross-Error Sensitivity

The influence function given by (6) may be used to study several robustness properties. One of the simplest and most revealing is the *gross-error sensitivity* of an estimator T at a distribution F. It is defined by

$$\gamma^\star = \sup_x |IF(x; T, F)| \ . \tag{9}$$

By taking the supremum over all x for which the $IF(x; T, F)$ exists, gross-error sensitivity measures the worst possible influence on an estimator by an arbitrary infinitesimal contaminant. If the gross-error sensitivity is unbounded, $\gamma^\star = \infty$, then the estimator is completely intolerant of outliers; a single outlier can ruin the estimator.

2.5 Global Robustness Concepts

2.5.1 Breakdown Point

The *breakdown point* is a measure of the maximum fraction $\epsilon^\star$ of arbitrary gross errors that an estimator can handle. Following [31] and [32], consider an estimator T_m and a sample

$X = \{x_1, \ldots, x_m\}$. Now, consider all possible corrupted samples X' obtained by replacing any fraction $\epsilon = f/m$ of the good samples by arbitrary values. Denote the estimate at the corrupted sample X' by T'_m and the maximum bias caused by the contamination by $b_{max} = \sup |T_m - T'_m|$. Then the breakdown point $\epsilon^\star$ of the estimator T_m at the sample X is given by

$$\epsilon^\star = \max \left\{ \epsilon = \frac{f}{m} \ : \ b_{max} \text{ finite} \right\} \tag{10}$$

The breakdown point is independent of the distribution of the good data as it involves no probability distributions. The largest amount of bad data that any estimator can handle is half the number of redundant observations $f_{max} = [(m-n)/2]$ which gives

$$\epsilon^\star_{max} = \frac{[(m-n)/2]}{m} \tag{11}$$

where $[a]$ denotes the integer part of a and n is the number of parameters to be estimated. As $m \to \infty$, $\epsilon^\star_{max} = 1/2$ or 50%. An estimator with a breakdown point of 50% is called a *high breakdown point estimator*.

2.5.2 Bias Curve

As Rousseeuw and Croux [33] discuss, the robustness of an estimator is not characterized solely by its influence function and its breakdown point. A more general view is given by the *bias curve* which describes the effect of a given fraction of contamination. They define the bias curve for location estimators as follows. Let

$$\mathcal{F}_\epsilon = \{(1-\epsilon)F + \epsilon H\} \tag{12}$$

where H ranges over all distributions and $\epsilon > 0$. The bias curve of an estimator T is defined as

$$B(\epsilon, T, F) = \sup_{G \in \mathcal{F}_\epsilon} |T(G)| \ . \tag{13}$$

For example the bias curve for M-estimators of location as derived in [33] is given by the implicit equation

$$0 = (1-\epsilon) E_F \left[\psi(X - B(\epsilon, T, F)) \right] + \epsilon \psi(\infty) \ . \tag{14}$$

The bias curve is an important complement to both the influence function and the breakdown point as it links the various robustness concepts. The gross-error sensitivity is the slope of the tangent to the bias curve at $\epsilon = 0$ and the breakdown point is the value of ϵ at which the bias curve is infinite.

Chapter 3

St. Laurent and Cook's Jacobian Leverage for Nonlinear Regression

3.1 Generalized and Jacobian Leverage

3.1.1 Definition of Generalized and Jacobian Leverage

St. Laurent and Cook develop [25] and analyze a technique they call Jacobian Leverage for measuring the influence of an observation at the nonlinear least-squares regression estimator. In their development, they assume the general nonlinear regression model. Let $\underline{Y}$ be the vector of response values. Denote the vector of perturbed responses $\underline{Y}_{m,b} = \underline{Y} + b\underline{f}_m$ where $\underline{f}_m$ is the mth standardized basis vector in $\Re^n$. Let $\eta(\underline{\theta})$ be the model function describing the response surface. Values that are evaluated at the least-squares estimate are hatted. Define $\underline{V} = \underline{V}(\underline{\theta})$ as the $n \times p$ matrix with ith row $\partial\eta_i(\underline{\theta})/\partial\underline{\theta}$.

St. Laurent and Cook propose two measures of leverage, namely Generalized Leverage and Jacobian Leverage. *Generalized Leverage* is defined as the ratio of change in response due to perturbation of the mth response by b scaled by the magnitude of the perturbation itself, yielding

$$G(b;m) = \frac{(\underline{\hat{Y}}_{m,b} - \underline{\hat{Y}})}{b} \ . \tag{15}$$

Jacobian Leverage is defined as the limit of Generalized Leverage $G(b;m)$ as the magnitude of the perturbation approaches zero, namely

$$J(m) = \lim_{b \to 0} G(b;m) \ . \tag{16}$$

Jacobian Leverage measures the magnitude of the derivative of each fitted value with respect to the mth fitted response. St. Laurent and Cook derive expressions evaluating the two measures at the least squares regression estimator for nonlinear models.

3.1.2 Overview of Derivation of Jacobian Leverage

We outline the procedure followed by St. Laurent and Cook to derive expressions for Generalized and Jacobian Leverage. First, generalize the perturbation of the response vector $\underline{Y}$ by perturbing it by an arbitrary vector f of unit length, namely $\underline{Y} + b\underline{f}$. Assume there exists a twice continuously differentiable vector-valued function which gives the least-squares estimate as a function of the size and direction of a perturbation; denote the function $\hat{\underline{\theta}}_{\underline{f}}(b)$. Note that when $b = 0$, the function gives the least-squares estimate of the unperturbed data. Expand the function about $b = 0$ via a second-order Taylor series. Solve for the Generalized Leverage $G(b; \underline{f})$. Use the constraints imposed by the normal equations requiring orthogonality of the error vector at the least-squares fit and tangent vector at that fit to solve for the unknown parameter in the expansion. This gives an expression for $G(b; \underline{f})$.

St. Laurent and Cook define the *Jacobian Leverage matrix* as

$$\hat{\underline{J}} = \hat{\underline{V}}(\hat{\underline{V}}^T \hat{\underline{V}} - [\hat{\underline{e}}^T][\hat{\underline{W}}])^{-1}\hat{\underline{V}}^T \tag{17}$$

where [ ] [ ] denotes column multiplication of three-dimensional arrays (see [34]). Taking the limit of the expression for $G(b; \underline{f})$ as the magnitude of the perturbation b approaches zero yields the Jacobian Leverage matrix and, thus, a closed expression for $\underline{J}$. St. Laurent and Cook use the main diagonal of $\underline{J}$, denoted by $\{\underline{J}_{ii}\}$, to study the sensitivity of the ith observation on the ith fitted value. Note that for a linear regression model, $\eta(\underline{\theta}) = \underline{X}\,\underline{\theta}$, Jacobian Leverage reduces to the hat matrix

$$\hat{\underline{H}} = \hat{\underline{V}}(\hat{\underline{V}}^T \hat{\underline{V}})^{-1}\hat{\underline{V}}^T \tag{18}$$

which has been heavily studied in linear regression theory and shown to be unreliable for identifying highly influential values in linear models (e.g. [31]). St. Laurent and Cook [26] provide more detailed analysis of their technique, including the relationship to $\hat{\underline{H}}$, the relationship of $\hat{\underline{J}}$ to curvature, and the difficulty in interpreting the diagonal entries of $\hat{\underline{J}}$.

3.2 Assessment of Jacobian Leverage

3.2.1 Relationship to Influence

Generalized and Jacobian Leverage measure the effect on an estimator of replacing a single response with a perturbed value. Therefore, what St. Laurent and Cook call "Leverage" is really a type of influence (see [28], [29], etc.). Additionally, Jacobian Leverage measures the effect as the magnitude of the perturbation approaches zero. Therefore, it can be viewed as measuring influence only for an infinitesimal amount and an infinitesimal amplitude of contamination. In contrast, the influence function (6) measures the effect of a contaminant of arbitrary size.

3.2.2 Derivation Based on Least-Squares Estimator

The expressions that St. Laurent and Cook derive for the Generalized and Jacobian Leverage are based on the nonlinear least-squares regression estimator. Therefore, while the expressions may be good tools for evaluating the influence of the nonlinear least-squares regression estimator of nonlinear models at $\epsilon = 0$ with contaminants of infinitesimal amplitude, they are not necessarily good tools for evaluating the influence of other estimators of nonlinear models (e.g. M-estimators).

The assumption of least-squares also leads us to conjecture that the technique may suffer from many of the same problems as the least squares estimator, namely difficulty identifying multiple influential points, susceptibility to the masking effect (hiding of influential points by other influential points), and failure to identify the effect of large arbitrary outliers.

In the spirit of the diagnostic school, the technique requires a (potentially computationally difficult) fit before identifying influential points. Assuming that the leverage effect exists in nonlinear model estimation (demonstrated in subsequent chapters), this approach is clearly dangerous as it will enable masking and (potentially) hide highly influential data.

3.2.3 Discussions

There are difficulties interpreting the Jacobian Leverage matrix. St. Laurent and Cook admit that it is not clear how to assess what "large" Jacobian Leverage means since the Jacobian Leverage matrix is not an orthogonal projection matrix with main diagonal entries limited to nonnegative values less than or equal to one. This is not a fault of the technique per se since "superleverage" (more aptly super-linear influence) is still a fundamental problem in nonlinear regression. However, tools must be developed which provide a means for unambiguously assessing the sensitivity of an estimator to contamination.

Jacobian Leverage has several limitations. First, it does not give quantitative information about the effect of large outliers. By analyzing the diagonal entries of the matrix, it is true that we may identify observations that have more influence on the nonlinear least-squares regression estimate relative to other observations in the given design, but the technique does not indicate whether the influence of an outlier is bounded or not.

A second limitation stems from the use of the least squares estimator in the derivation. Due to that, Jacobian Leverage fails when more than one outlying observation is present in the data because one leverage point may hide another one. This is called *masking* (see [31] for examples in the case of linear regression).

3.3 Example Calculations of Jacobian Leverage

3.3.1 Unperturbed Data at St. Laurent and Cook's Model

We study the model

$$y_i = \theta_1 x_{i1} + e^{(\theta_2 x_{i2})} + e_i \tag{19}$$

that St. Laurent and Cook considered in [26] while putting the true parameter $\tilde{\theta} = \{2.0, 0.0\}$, design values

$$\begin{aligned}\underline{x} = \{&(-4.0,-2.5),(-3.0,-2.0),(-2.0,-1.5),(-1.0,-1.0),\\ &(0.0,0.0),(1.0,1.0),(2.0,1.5),(3.0,2.0),(4.0,2.5)\}\end{aligned} \tag{20}$$

and response values

$$\underline{y} = \{-10.0,-5.0,-2.0,-1.0,1.5,4.0,5.0,6.0,7.0\}\,. \tag{21}$$

We perform an L_1 fit at these observations using the El-Attar-Vidyasagar-Dutta algorithm [12]. Starting from $\underline{\theta} = \{1.9, 0.1\}$, we obtain the estimated value $\hat{\theta} = \{2.0, 0.0\}$ as expected. We calculate the Jacobian Leverage matrix $\hat{J}$ and examine the diagonal entries

$$\{0.267, 0.15, 0.067, 0.02, 0, 0.02, 0.067, 0.15, 0.267\} \tag{22}$$

which follow [26] very closely. We check for outliers (i.e. influential observations) by calculating the robust standardized residuals $(z_i - \text{med}\{z_i\})/\text{MADM}$ where

$$\text{MADM} = 1.4826\ \text{med}_i\{\ |z_i - \text{med}_j\{z_j\}|\ \} \tag{23}$$

(note that the factor 1.4826 corrects for Fisher consistency) and find them to be

$$\{2.0134, 0.8356, 0.0, 0.4732, 0.6745, 0.4732, 0.0, 0.8356, 2.0134\}\ . \tag{24}$$

We observe no outliers amidst the standardized residuals using a threshold of 2.5 and thus no influential points are identified by Jacobian Leverage.

3.3.2 Multiple Vertical Outliers at St. Laurent and Cook's Model

To evaluate the ability of Jacobian Leverage to identify multiple influential values, we use the model given by (19), perturb two observations in the response vector $\underline{y}$, refit using the L_1 estimator, and calculate $\hat{J}$. Specifically, we let $y_8 = 6$ and $y_9 = 10$, perform an L_1 fit, and obtain $\hat{\theta} = \{2.1674, 0.1143\}$. We calculate the Jacobian Leverage matrix $\hat{J}$ and examine its diagonal entries, which are given by

$$\{0.5558, 0.2345, 0.0714, \ldots, 0.1295, 0.2279, 0.3940\} \tag{25}$$

Calculating the robust standardized residuals of the diagonal entries, we find

$$\{4.3355, 1.0676, 0.5915, \ldots, 0.0, 1.0, 2.6894\} \tag{26}$$

and observe that although the outlier at $y_9 = 10$ was pinpointed as influential, the first observation was identified as more influential. In addition, the outlier at $y_8 = 6$ was missed completely. Thus, we conclude that Jacobian Leverage is susceptible to the masking effect.

3.3.3 Multiple Vertical Outliers at Contrived Model

As another test of the ability of Jacobian Leverage to identify multiple influential values, we consider the model

$$y_i = \theta_1 + {x_i}^{\theta_2} \tag{27}$$

and study the case where the true parameter is $\tilde{\theta} = \{0.0, 1.0\}$, the design values are

$$\underline{x} = \{1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0\} \tag{28}$$

and the response values are

$$\underline{y} = \{1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0\} \ . \tag{29}$$

This is an *exact fit* analysis, that is, the estimator is studied at observed responses exactly matching the response at the known true parameter. We perturb several responses and calculate the Jacobian Leverage matrix. Specifically, we let $y_6 = y_7 = 198$. We perform an L_1 fit and obtain $\hat{\underline{\theta}} = \{-14.69, 2.75\}$, which follows (x_7, y_7) nearly exactly, as expected since it is a leverage point. The data and the fitted response are shown in Figure 1. We calculate the Jacobian Leverage matrix $\hat{J}$ and the robust standardized residuals of the diagonal entries

$$\{0.3259, 0.0, 0.6745, 1.4891, 1.7428, 0.0095, 6.2443\} \ . \tag{30}$$

We observe in Figure 1 that although the most significant outlier at $x = 7$ was identified as significant, the less extreme outlier was not.

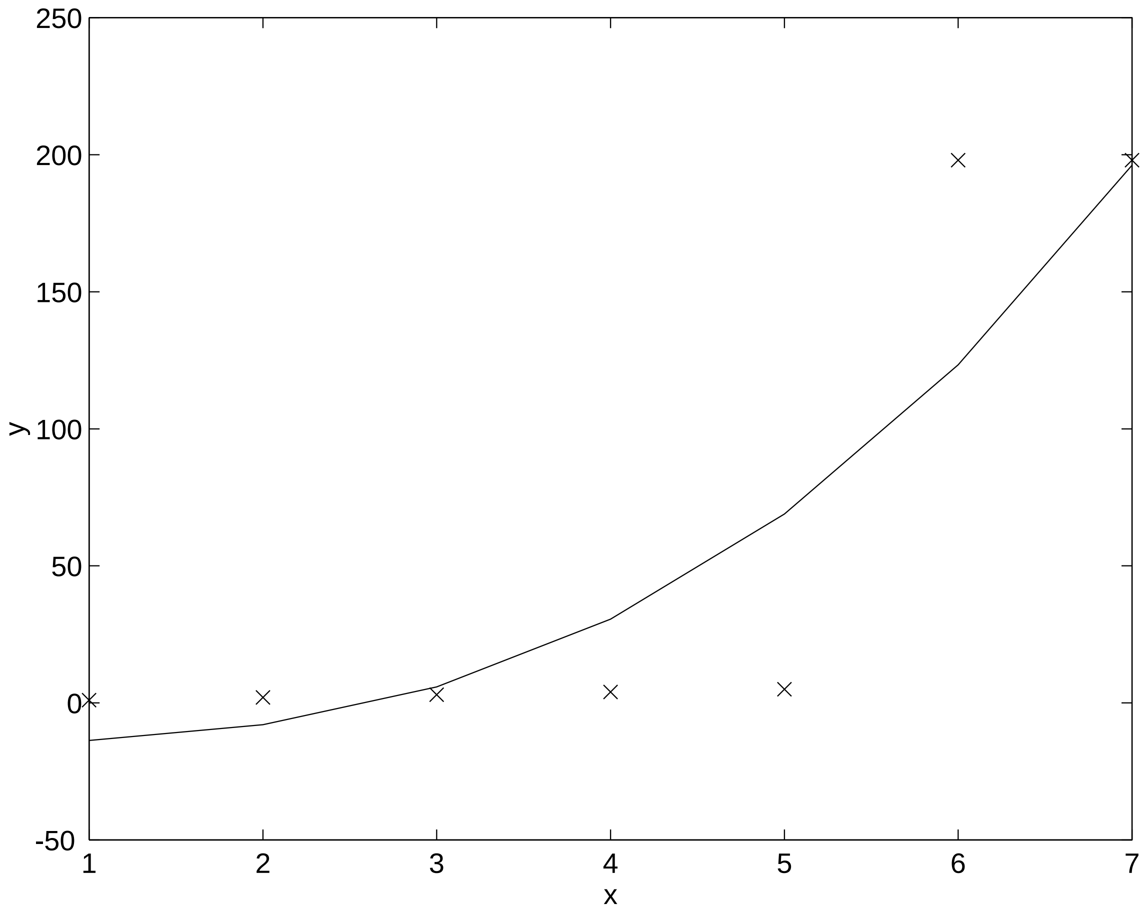

Figure 1: Fit of Exponential Model to Contrived Data Used to Evaluate Jacobian Leverage

Chapter 4

Influence Functions for M-Estimators of Nonlinear Models

In this chapter, we derive the influence function for M-estimators of nonlinear models, first in the nonlinear location case and then in the nonlinear regression case.

4.1 Nonlinear Location Case

4.1.1 Definition of a Location M-Estimator

Let $\{y_i : i = 1, \ldots, n\}$ be a sequence of i.i.d. random variables. Let $y_i = \eta(\theta) + e_i$ for $i = 1, \ldots, n$. Let $y_i \in \Re$ be the ith observation value. Let $\theta \in \Theta \subset \Re$ be a scalar unknown parameter. Let $e_i \in \Re$ be the ith error and let $\eta(\theta) : \Theta \to \Re$ be the (possibly nonlinear) model function.

Assume the error is distributed with distribution $G(e/\sigma)$, $\sigma > 0$, with density g with respect to Lebesgue measure. Let the target distribution be defined as $F_\theta(y)$ with density $f_\theta(y) = \sigma^{-1} g((y - \eta(\theta))/\sigma)$. Specifically, assume normally distributed errors, with density $f_\theta(y) = \phi(y - \eta(\theta))$. We assume a known scale σ; without loss of generality, we put $\sigma = 1$.

We use an M-estimator to estimate θ. These estimators are defined implicitly by

$$\Gamma(T_n) = \min\{\Gamma(\theta) : \theta \in \Theta\} \tag{31}$$

where

$$\Gamma(\theta) = \sum_{i=1}^{n} \rho(r_i) \tag{32}$$

for some $\rho : \Re \rightarrow \Re^+$ and residual $r = y_i - \eta(\theta)$. If the derivative of $\rho(\cdot)$ exists with respect to θ and $\eta(\cdot)$ is twice differentiable with respect to θ, then T_n satisfies the equation

$$\sum_{i=1}^{n} \lambda(r) = 0 \tag{33}$$

where λ is the function

$$\lambda(r) = \frac{\partial}{\partial\theta}\rho(r) = \rho'(r)\frac{\partial r}{\partial\theta} = -\rho'(r)\frac{\partial\eta(\theta)}{\partial\theta} \ . \tag{34}$$

Put $\psi(r) := \rho'(r)$ and define the estimator T_n implicitly as the solution of

$$\sum_{i=1}^{n} \psi(y_i - \eta(T_n)) \left[\frac{\partial\eta(\theta)}{\partial\theta}\right]_{T_n} = 0 \ . \tag{35}$$

Given an empirical c.d.f. G_n, the functional form of a location M-estimator is $T_n = T(G_n)$, where T is the functional

$$\int \psi(y - \eta(T(G))) \left[\frac{\partial\eta(\theta)}{\partial\theta}\right]_{T(G)} dG = 0 \ . \tag{36}$$

4.1.2 Influence Function of a Location M-Estimator

To find the influence function of an M-estimator given implicitly in functional form by (36) we substitute $G = F_t = (1-t)F + tH$ into the definition of the estimator, evaluate the derivative with respect to t at $t = 0$, and solve for the influence function.

Substituting $G = (1-t)F + tH$ into (36) yields

$$0 = \int \psi(y^\star, \eta(T(G))) \left[\frac{\partial\eta(\theta)}{\partial\theta}\right]_{T(G)} d(F + t(H-F)) \ . \tag{37}$$

If the kernel of the integral is bounded and measurable then

$$0 = \int \psi(y^\star, \eta(T(G))) \left[\frac{\partial\eta(\theta)}{\partial\theta}\right]_{T(G)} dF + t\int \psi(y^\star, \eta(T(G))) \left[\frac{\partial\eta(\theta)}{\partial\theta}\right]_{T(G)} d(H-F) \ . \tag{38}$$

Differentiating with respect to t and applying the product rule to the right-hand side yields

$$\begin{aligned} 0 \quad = \quad & \frac{\partial}{\partial t}\int \psi(y^\star, \eta(T(G))) \left[\frac{\partial\eta(\theta)}{\partial\theta}\right]_{T(G)} dF + \\ & \int \psi(y^\star, \eta(T(G))) \left[\frac{\partial\eta(\theta)}{\partial\theta}\right]_{T(G)} d(H-F) + \\ & t\frac{\partial}{\partial t}\left\{\int \psi(y^\star, \eta(T(G))) \left[\frac{\partial\eta(\theta)}{\partial\theta}\right]_{T(G)} d(H-F)\right\} \ . \end{aligned} \tag{39}$$

By [35], we get

$$\frac{d}{dt}\int_E f(t,x)\mu(dx) = \int_E f'(t,x)\mu(dx) \tag{40}$$

if f is continuous for $\forall x \in E$ and measurable on E, and if f' is measurable on E. We add these regularity conditions so that differentiation and integration are interchangeable. Permuting them yields

$$\begin{aligned} 0 &= \int \frac{\partial}{\partial t}\left[\psi(y^\star, \eta(T(G)))\left[\frac{\partial\eta(\theta)}{\partial\theta}\right]_{T(G)}\right] dF + \\ &\quad \int \psi(y^\star, \eta(T(G)))\left[\frac{\partial\eta(\theta)}{\partial\theta}\right]_{T(G)} d(H-F) + \\ &\quad t\int \frac{\partial}{\partial t}\left[\psi(y^\star, \eta(T(G)))\left[\frac{\partial\eta(\theta)}{\partial\theta}\right]_{T(G)}\right] d(H-F)\ . \end{aligned} \tag{41}$$

Let $H = \Delta_y$ be the distribution with unit mass at y and evaluate (41) at $t = 0$. Noting that the last term falls out gives

$$\begin{aligned} 0 &= \int \frac{\partial}{\partial t}\left[\psi(y^\star, \eta(T(G)))\left[\frac{\partial\eta(\theta)}{\partial\theta}\right]_{T(G)}\right]_{t=0} dF + \\ &\quad \int \psi(y^\star, \eta(T(F)))\left[\frac{\partial\eta(\theta)}{\partial\theta}\right]_{T(F)} d\Delta_y - \\ &\quad \int \psi(y^\star, \eta(T(F)))\left[\frac{\partial\eta(\theta)}{\partial\theta}\right]_{T(F)} dF\ . \end{aligned} \tag{42}$$

Assuming Fisher consistency, the third term falls out. The second term simplies to the kernel evaluated at y due to the sifting property of the delta, yielding

$$0 = \int \frac{\partial}{\partial t}\left[\psi(y^\star, \eta(T(G)))\left[\frac{\partial\eta(\theta)}{\partial\theta}\right]_{T(G)}\right]_{t=0} dF + \psi(y, \eta(T(F)))\left[\frac{\partial\eta(\theta)}{\partial\theta}\right]_{T(F)}\ . \tag{43}$$

Let $\eta'(\theta) = \partial\eta(\theta)/\partial\theta$, $\eta''(\theta) = \partial^2\eta(\theta)/\partial\theta^2$, and $\psi'(\cdot) = \partial\psi/\partial\theta$. Applying the product rule to the kernel of the first term and using the chain rule where needed gives

$$\begin{aligned} &\frac{\partial}{\partial t}\left[\psi(y^\star, \eta(T(G)))\left[\frac{\partial\eta(\theta)}{\partial\theta}\right]_{T(G)}\right]_{t=0} \\ &= \left[\frac{\partial}{\partial t}\left[\psi(y^\star, \eta(T(G)))\right]_{t=0}\right]\eta'(T(F)) + \psi(y^\star, \eta(T(F)))\frac{\partial}{\partial t}\left[\eta'(T(G))\right]_{t=0} && (44) \\ &= [\psi'(y^\star, \eta(T(F)))\eta'(T(F))\frac{\partial}{\partial t}[T(G)]_{t=0}]\eta'(T(F)) + \\ &\quad \psi(y^\star, \eta(T(F)))\eta''(T(F))\frac{\partial}{\partial t}[T(G)]_{t=0} && (45) \\ &= \frac{\partial}{\partial t}[T(G)]_{t=0}\left[\psi'(y^\star, \eta(T(F)))\eta'(T(F))\eta'(T(F)) + \psi(y^\star, \eta(T(F)))\eta''(T(F))\right]\ . && (46) \end{aligned}$$

Substituting the simplified kernel (46) and solving for the influence function $IF(y, T, F) = \frac{\partial}{\partial t}[T(G)]_{t=0}$ yields

$$
\begin{aligned}
&IF(y, T, F) = \\
&\quad -\left[\int \left[\psi'(y^\star, \eta(T(F)))(\eta'(T(F)))^2 + \psi(y^\star, \eta(T(F)))\eta''(T(F))\right] dF\right]^{-1} \\
&\qquad\qquad\qquad \psi(y, \eta(T(F)))\eta'(T(F))
\end{aligned}
\tag{47}
$$

when the inverse exists. Note that when $\eta(\theta) = \theta$, (47) reduces to

$$IF(y, T, F) = \frac{\psi(y)}{E[\psi'(y)]} \tag{48}$$

which is the well-known influence function for the one-dimensional M-estimator of location.

4.2 Nonlinear Regression Case

4.2.1 Definition of a Regression M-Estimator

Let $\{(\underline{x}_i, y_i) : i = 1, \ldots, n\}$ be a sequence of i.i.d. random vectors such that

$$y_i = \eta(\underline{x}_i, \underline{\theta}) + e_i, \quad i = 1, \ldots, n \tag{49}$$

where $y_i \in \Re$ is the ith observation value, and $\underline{x}_i \in \mathcal{X} \subset \Re^p$ is the ith row of the $n \times p$ design matrix. Let $\underline{\theta} \in \Theta \subset \Re^p$ be the p-vector of unknown parameters, $e_i \in \Re$ be the ith error, and $\eta(\underline{x}, \underline{\theta}) : \mathcal{X} \times \Theta \to \Re$ be the model function. Assume the model $\eta(\underline{x}, \underline{\theta})$ is nonlinear with respect to $\underline{x}$ and/or $\underline{\theta}$ with both observations $\{y_i\}$ and carriers $\{\underline{x}_i\}$ random. Assuming random carriers allows for the consideration of extreme designs.

Let $K(\underline{x})$ be the distribution of $\underline{x}_i$ with density k with respect to Lebesgue measure. Assume e_i is independent of $\underline{x}_i$ and distributed according to $G(e/\sigma)$, $\sigma > 0$ with density g with respect to Lebesgue measure. Let the model distribution be defined $F_\theta(\underline{x}, y)$ with density $f_{\underline{\theta}}(\underline{x}, y) = \sigma^{-1} g((y - \eta(\underline{x}, \underline{\theta}))/\sigma) k(\underline{x})$, where $f_{\underline{\theta}}(\underline{x}, y)$ is the joint density of $(\underline{x}_i, y_i)$. We will study the model distribution $F_{\underline{\theta}}(\underline{x}, y)$ with density $f_{\underline{\theta}}(\underline{x}, y) = \phi(y - \eta(\underline{x}, \underline{\theta})) k(\underline{x})$ where $\phi(e)$ is the standard normal density. We assume a known scale σ; without loss of generality, we put $\sigma = 1$.

The M-estimator T_n in (33) is, in the more general multiparameter case, the vector solution to the system of equations

$$\sum_{i=1}^{n} \underline{\lambda}(\underline{x}_i, T_n) = \underline{0} \tag{50}$$

where $\lambda : \mathcal{X} \times \Theta \rightarrow \Re^p$. Given an empirical c.d.f. G_n, the functional form of the multiparameter M-estimator is $T_n = T(G_n)$, where T is defined implicitly by the vector-valued functional

$$\int \underline{\lambda}(\underline{x}, T(G))dG(\underline{x}) = \underline{0} \ . \tag{51}$$

Regression M-estimators are a specific case of multiparameter estimation. They are defined implicitly by

$$\Gamma(T_n) = \min\{\Gamma(\underline{\theta}) : \underline{\theta} \in \Theta\} \tag{52}$$

where

$$\Gamma(\underline{\theta}) = \sum_{i=1}^{n} \rho(r) \tag{53}$$

for $\rho : \Re \rightarrow \Re^+$ and residual $r_i = y_i - \eta(\underline{x}_i, \underline{\theta})$. If the derivative of $\rho(\cdot)$ exists with respect to $\underline{\theta}$ and $\eta(\cdot)$ is twice differentiable with respect to $\underline{\theta}$ then the regression M-estimator T_n satisfies the system of equations given by

$$\sum_{i=1}^{n} \underline{\lambda}(r) = \sum_{i=1}^{n} \underline{\lambda}(\underline{x}, y, T_n) = \underline{0} \ . \tag{54}$$

Here $\underline{\lambda}(r)$ is the vector-valued function

$$\underline{\lambda}(r) = \frac{\partial}{\partial \underline{\theta}} \rho(r) = \rho'(r) \frac{\partial r}{\partial \underline{\theta}} = -\rho'(r) \frac{\partial \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta}} \ . \tag{55}$$

Put $\psi(r) := \rho'(r)$. Given an empirical c.d.f. G_n, the functional form of the regression M-estimator is $T_n = T(G_n)$, where T is defined implicitly by the vector-valued functional

$$\int \underline{\lambda}(r)dG(\underline{x}, y) = \underline{0} \ . \tag{56}$$

Expanding the kernel, the functional becomes

$$\int \underline{\lambda}(\underline{x}, y, T(G))dG(\underline{x}, y) = \underline{0} \ . \tag{57}$$

We exploit the similarity of this functional to the multiparameter M-estimator functional (51). After deriving the influence function of the multiparameter M-estimator, we will use it to derive the influence function of the regression M-estimator.

4.2.2 Influence Function of Regression M-Estimators

To find the influence function of regression M-estimators given implicitly in functional form by (57), we first derive the influence function of multiparameter M-estimators defined by (51). To this end, we put $G = F_t = (1-t)F + tH$, evaluate the derivative with respect to t at $t = 0$, and solve for the influence function along the same lines as in [27], [28], and [29].

Into this influence function, we substitute the $\underline{\lambda}(\cdot)$ function of the regression M-estimator and evaluate the influence function at the model distribution $F_{\underline{\theta}}(\underline{x}, y)$. Finally, we solve for the influence function of the regression M-estimator.

We use several definitions from [36]. Let $\underline{\theta}$ be a $p\times 1$ vector and let $f(\underline{\theta})$ be a scalar-valued function of $\underline{\theta}$. Define

$$\frac{\partial f(\underline{\theta})}{\partial \underline{\theta}} := \left[\frac{\partial f(\underline{\theta})}{\partial \theta_1} \cdots\cdots \frac{\partial f(\underline{\theta})}{\partial \theta_p} \right]^T \tag{58}$$

and

$$\frac{\partial^2 f(\underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}} := \frac{\partial}{\partial \underline{\theta}} \left[\frac{\partial f(\underline{\theta})}{\partial \underline{\theta}} \right]^T = \left[\frac{\partial^2 f(\underline{\theta})}{\partial \theta_i \partial \theta_j} \right] \tag{59}$$

which corresponds to the matrix with i,j-th element $\left[\frac{\partial^2 f(\underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}}\right]_{ij}$.

First, we derive the influence function of the multiparameter estimator. Substituting G into the functional form of the multiparameter estimator (51) yields

$$\underline{0} = \int \underline{\lambda}(\underline{x}^\star, T(G))dG = \int \underline{\lambda}(\underline{x}^\star, T(G))d(F + t(H - F)) \ . \tag{60}$$

If $\underline{\lambda}(\cdot)$ is bounded and measurable then

$$\underline{0} = \int \underline{\lambda}(\underline{x}^\star, T(G))dF + \int \underline{\lambda}(\underline{x}^\star, T(G))d(t(H - F)) \ . \tag{61}$$

Simplifying the second term yields

$$\underline{0} = \int \underline{\lambda}(\underline{x}^\star, T(G))dF + t \int \underline{\lambda}(\underline{x}^\star, T(G))d(H - F) \ . \tag{62}$$

Differentiating with respect to t and applying the chain rule to the right hand side gives

$$\underline{0} = \frac{\partial}{\partial t} \int \underline{\lambda}(\underline{x}^\star, T(G))dF + \int \underline{\lambda}(\underline{x}^\star, T(G))d(H - F) + t \frac{\partial}{\partial t} \int \underline{\lambda}(\underline{x}^\star, T(G))d(H - F) \ . \tag{63}$$

By [35], we get

$$\frac{d}{dt} \int_E f(t,x)\mu(dx) = \int_E f'(t,x)\mu(dx) \tag{64}$$

if f is continuous and measurable on E and if f' is measurable on E. We add these regularity conditions so that differentiation and integration are interchangeable, yielding

$$\underline{0} = \int \frac{\partial}{\partial t}\{\underline{\lambda}(\underline{x}^\star, T(G))\}dF + \int \underline{\lambda}(\underline{x}^\star, T(G))d(H - F) + t \int \frac{\partial}{\partial t}\{\underline{\lambda}(\underline{x}^\star, T(G))\}d(H - F) \ . \tag{65}$$

Let $H = \Delta_{\underline{x}}$ and evaluate (65) at $t = 0$. Noting that the last term falls out gives

$$\underline{0} = \int \frac{\partial}{\partial t}\{\underline{\lambda}(\underline{x}^\star, T(G))\}_{t=0} dF + \int \underline{\lambda}(\underline{x}^\star, T(F))d\Delta_{\underline{x}} - \int \underline{\lambda}(\underline{x}^\star, T(F))dF \ . \tag{66}$$

Assuming Fisher consistency, the third term falls out. The second term simplifies to the kernel evaluated at $\underline{x}$ due to the sifting property of the delta, yielding

$$\underline{0} = \int \frac{\partial}{\partial t}\{\underline{\lambda}(\underline{x}^{\star}, T(G))\}_{t=0} dF + \underline{\lambda}(\underline{x}, T(F)) \ . \tag{67}$$

Applying the chain rule to the kernel of the first term, we obtain

$$\underline{0} = \int \frac{\partial}{\partial \underline{\theta}}[\underline{\lambda}(\underline{x}^{\star}, \underline{\theta})]_{T(F)} dF \left(\frac{\partial}{\partial t}[T(G)]_{t=0}\right) + \underline{\lambda}(\underline{x}, T(F)) \ . \tag{68}$$

Solving for the influence function $IF(\underline{x}, T, F) = \frac{\partial}{\partial t}\left[T(G)\right]_{t=0}$, we get

$$\underline{IF}(\underline{x}, \lambda, F) = -\left[\int \frac{\partial}{\partial \underline{\theta}}\left[\underline{\lambda}(\underline{x}^{\star}, \underline{\theta})\right]_{T(F)} dF\right]^{-1} \underline{\lambda}(\underline{x}, T(F)) \ . \tag{69}$$

Note that the first term in the r.h.s of the equation is a $p \times p$ matrix and the second term is a $p \times 1$ vector.

The required regularity conditions are

1. continuity, boundedness, measurability of $\lambda(\cdot)$
2. existence of measurable derivatives of $\lambda(\cdot)$

From the influence function of multiparameter M-estimators given by (69), we can calculate the influence function of regression M-estimators. Substitute

$$\underline{\lambda}(r) = \psi(r)\frac{\partial \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta}} \tag{70}$$

and $F = F_{\underline{\theta}}(\underline{x}, y)$ in (51). This allows us to use the definition of the influence function given by (69). First, rewriting (69) to include y yields

$$\underline{IF}(\underline{x}, y; \lambda, F) = -\left[\int \frac{\partial}{\partial \underline{\theta}}\left[\lambda(\underline{x}^{\star}, y^{\star}, \underline{\theta})\right]_{T(F)} dF(\underline{x}, y)\right]^{-1} \lambda(\underline{x}, y, T(F)) \ . \tag{71}$$

Also, define

$$\underline{M} := -\int \frac{\partial}{\partial \underline{\theta}}\left[\lambda(\underline{x}^{\star}, y^{\star}, \underline{\theta})\right]_{T(F)} dF(\underline{x}, y) \ . \tag{72}$$

Then

$$\underline{IF}(\underline{x}, y; \lambda, F_{\underline{\theta}}) = \underline{M}^{-1}\lambda(\underline{x}, y, T(F_{\underline{\theta}})) \ . \tag{73}$$

Applying the definitions in (58) and (59) to the kernel of (72) yields

$$\frac{\partial}{\partial \underline{\theta}}\underline{\lambda}(\underline{x}, y, \underline{\theta}) = \frac{\partial}{\partial \underline{\theta}}\{\psi(r)\frac{\partial \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta}}\} \ . \tag{74}$$

Expanding the r.h.s. gives

$$\begin{aligned}
\frac{\partial}{\partial \underline{\theta}} \underline{\lambda}(\underline{x}, y, \underline{\theta}) &= \frac{\partial}{\partial \underline{\theta}} \{\psi(r)\} \left[\frac{\partial \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta}}\right]^T + \psi(r) \frac{\partial^2 \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}} && (75) \\
&= \frac{\partial \psi(r)}{\partial r} \frac{\partial r}{\partial \underline{\theta}} \left[\frac{\partial \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta}}\right]^T + \psi(r) \frac{\partial^2 \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}} && (76) \\
&= -\psi'(r) \left[\frac{\partial \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta}}\right] \left[\frac{\partial \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta}}\right]^T + \psi(r) \frac{\partial^2 \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}} \ . && (77) \\
& && (78)
\end{aligned}$$

Substituting the expanded kernel into (72) gives the $p \times p$ matrix

$$\underline{M} = \int \left\{ \psi'(r) \left[\frac{\partial \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta}}\right] \left[\frac{\partial \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta}}\right]^T - \psi(r) \frac{\partial^2 \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}} \right\}_{T(F_{\underline{\theta}})} dF_{\underline{\theta}}(\underline{x}, y) \ . \tag{79}$$

If the inverse of $\underline{M}$ exists, then the influence function is written as

$$\underline{IF}(\underline{x}, y; \psi, F_{\underline{\theta}}) = \psi(r) \underline{M}^{-1} \frac{\partial \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta}} \ . \tag{80}$$

Collecting all terms, we obtain the general influence function for M-estimators of nonlinear models. It is given by

$$\begin{aligned}
&\underline{IF}(\underline{x}, y; \psi, F_{\underline{\theta}}) = \\
&\psi(r) \left\{ \int \left[\psi'(r) \left[\tfrac{\partial \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta}}\right] \left[\tfrac{\partial \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta}}\right]^T - \psi(r) \tfrac{\partial^2 \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}} \right]_{T(F_{\underline{\theta}})} dF_{\underline{\theta}}(\underline{x}, y) \right\}^{-1} \tfrac{\partial \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta}} \ . \qquad (81)
\end{aligned}$$

4.2.3 Factorization of the Influence Function

In the case where e and $\underline{x}$ are independent, we can separate the influence function $\underline{IF}$ clearly into influence of residual IR and influence of model $\underline{IM}$. To this end, we write the kernel of $\underline{M}$ as a product of components due to residual r and components due to design $\underline{x}$ and model $\eta(\cdot)$ as follows. Expanding $\underline{M}$ in (79), we find

$$\begin{aligned}
\underline{M} &= \int \left\{ \psi'(r) \left[\frac{\partial \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta}}\right] \left[\frac{\partial \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta}}\right]^T \right\}_{T(F_{\underline{\theta}})} dF_{\underline{\theta}}(\underline{x}, y) - \\
&\quad \int \left\{ \psi(r) \frac{\partial^2 \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}} \right\}_{T(F_{\underline{\theta}})} dF_{\underline{\theta}}(\underline{x}, y) \ . && (82)
\end{aligned}$$

Due to the assumption of independence of e from $\underline{x}$ we can separate each integral into the product of the expectation of a component due to residual r and the expectation of a component due to design $\underline{x}$ and model $\eta(\cdot)$. Rewriting (82), we get

$$\begin{aligned}
\underline{M} &= \int \psi'(r) dG(r) \int \left\{ \left[\frac{\partial \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta}}\right] \left[\frac{\partial \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta}}\right]^T \right\} dK(\underline{x}) - \\
&\quad \int \psi(r) dG(r) \int \left\{ \frac{\partial^2 \eta(\underline{x}, \underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}} \right\} dK(\underline{x}) && (83)
\end{aligned}$$

We consider several cases for G and ψ. If G is symmetric and ψ is odd then the second term in (83) goes to 0. Specifically, we put $G = \Phi$ in (83) and, noting that $\int \psi(r)d\Phi(r) = 0$, we get

$$\underline{M} = \int \psi'(r)d\Phi(r) \int \left\{ \left[\frac{\partial\eta(\underline{x},\underline{\theta})}{\partial\underline{\theta}}\right] \left[\frac{\partial\eta(\underline{x},\underline{\theta})}{\partial\underline{\theta}}\right]^T \right\} dK(\underline{x}) \; . \tag{84}$$

If we neglect the Hessian completely, namely $\partial^2\eta(\underline{x},\underline{\theta})/\partial\underline{\theta}\partial\underline{\theta}$, we obtain the same $\underline{M}$ without restriction on G or ψ. Without neglecting the Hessian, and if G is not symmetric, then it is not possible to separate $\underline{M}$ into a product of components due to residual and components due to design and model. Therefore, the influence function cannot be separated into influence of residual and influence of model (or position) and we need to take into consideration the Hessian of the model. There has been much work analyzing such model properties, including their relationships to curvature. For example, Murray and Rice [37] define statistical curvature as the Hessian of the log-likelihood. It may be possible to re-interpret such work more generically and use it to analyze influence functions of estimators of nonlinear model.

If the inverse of (84) exists then the influence function is as in (80). Collecting all terms and substituting into (80), we obtain a simplified influence function for M-estimators of nonlinear models, given by

$$\underline{IF}(\underline{x}, r; \psi, F_{\underline{\theta}}) = \frac{\psi(r)}{E_\Phi\left\{\psi'(r)\right\}} \left\{ E_K \left\{ \left[\frac{\partial\eta(\underline{x},\underline{\theta})}{\partial\underline{\theta}}\right] \left[\frac{\partial\eta(\underline{x},\underline{\theta})}{\partial\underline{\theta}}\right]^T \right\} \right\}^{-1} \frac{\partial\eta(\underline{x},\underline{\theta})}{\partial\underline{\theta}} \; . \tag{85}$$

Further simplification is obtained by writing the influence function (85) as a product of two factors, specifically

$$\underline{IF}(\underline{x}, r; \psi, F_{\underline{\theta}}) = IR(r; \psi, \Phi) \cdot \underline{IM}(\underline{x}; \psi, K) \tag{86}$$

where $IR(\cdot)$ is (scalar) *influence of residual* defined as

$$IR(r; \psi, \Phi) := \frac{\psi(r)}{E_\Phi\left\{\psi'(r)\right\}} \tag{87}$$

and $\underline{IM}(\cdot)$ is (vector-valued) *influence of model* defined as

$$\underline{IM}(\underline{x}; \psi, K) := \left\{ E_K \left\{ \left[\frac{\partial\eta(\underline{x},\underline{\theta})}{\partial\underline{\theta}}\right] \left[\frac{\partial\eta(\underline{x},\underline{\theta})}{\partial\underline{\theta}}\right]^T \right\} \right\}^{-1} \frac{\partial\eta(\underline{x},\underline{\theta})}{\partial\underline{\theta}} \; . \tag{88}$$

Note that $\underline{IM}$ depends only on the model, its derivatives, and the design. It is clear that in the linear regression case, influence of model $\underline{IM}(\cdot)$ simplifies to the well-known influence of position $\underline{IP}(\cdot)$.

Assuming a bounded $\psi(\cdot)$, it is clear that $IR(\cdot)$ is bounded. However, without additional conditions on the model $\eta(\cdot)$, a similar conclusion cannot be reached about $\underline{IM}(\cdot)$. The new notation emphasizes that, for M-estimators of nonlinear models, unbounded influence is caused as much by the properties of the model as it is by the properties of the estimator and the design.

Note that when $\frac{\partial \eta(\underline{x}, \theta)}{\partial \underline{\theta}} = \underline{x}^T$, as in linear regression, (81) simplifies to

$$\underline{IF}(\underline{x}, y; \psi, F_{\underline{\theta}}) = \frac{\psi(r)}{E_K\left\{\psi'(r)\right\}} E_\Phi \left\{\underline{x}^T \underline{x}\right\}^{-1} \underline{x}^T \tag{89}$$

which is the well-known influence function for the M-estimator of linear regression. Also, we note that the structure of (81) is the same as the structure of the influence function of the nonlinear location M-estimator (47).

In linear regression, leverage points are defined as extreme values in $\mathcal{X}$ due to their potential for unbounded influence on M-estimators of linear regression (89). As in the linear regression case, outliers in $\mathcal{X}$ may have unbounded influence on the M-estimators of nonlinear regression. In the nonlinear regression case, the influence function is a complicated nonlinear expression involving residual, design, and model and it is not clear how to generally robustify the M-estimators of nonlinear regression.

Chapter 5

Examples and Case Studies

5.1 Introduction

In this chapter, we use the influence function derived in Chapter 4 to study the effect of contaminants on M-estimators of several nonlinear models. Specifically, we study the L_1 nonlinear regression estimator and calculate the estimates using the El-Attar-Vidyasagar-Dutta algorithm [12] and the Broyden-Fletcher-Goldfarb-Shanno approach [12] to unconstrained minimization. We demonstrate the usefulness of the influence function in predicting the weaknesses of nonlinear model estimators. We show that not only are M-estimators of nonlinear models vulnerable to factor space outliers, e.g. leverage points, as in linear regression, but that observations not outlying in the factor space can cause high influence. More generally, we show that influence in nonlinear model estimation using the M-estimators is caused as much by the properties of the model as it is by the properties of the estimator.

5.2 Logarithmic Model

Consider the model

$$\eta(x_i, \underline{\theta}) = \ln(x_i^3 + \theta_1) + \theta_2 \tag{90}$$

with design $x = \{1, 2, 3, 4, 5, 6, 10\}$. We study the exact fit properties of the M-estimators at this model with the true parameter $\underline{\tilde{\theta}} = \{0, 1\}$. The last observation is moved through the design space and its response perturbed to study the effect of outliers on the M-estimator.

Using the starting point $\underline{\theta} = \{0.1, 0.9\}$, we perform an L_1 estimate of the parameters of the model function. The fitted model parameters are $\underline{\hat{\theta}} = \{0.0, 1.0\}$. The data and the fitted response are shown in Figure 2.

5.2.1 Influence due to Outlying Observations

Using the results from Chapter 4, we study the effect of observations outlying in X on M-estimators at the model (90).

Calculating the Jacobian vector of the model function, we find

$$\underline{J} = \frac{\partial g(x_i, \underline{\theta})}{\partial \underline{\theta}} = \begin{bmatrix} \frac{1}{x_i^3+\theta_1} \\ 1 \end{bmatrix} . \tag{91}$$

Simple calculus shows that $\lim_{x\to\infty} \frac{\partial g}{\partial \theta_1} = 0$. Since the Jacobian in the influence function of the M-estimators of nonlinear models is unbounded and directly influenced by the observations, we expect that M-estimators at this model will not be influenced by outlying x_i, as the components of the Jacobian are not unbounded.

Let $\tilde{\underline{\theta}} = \{0.0, 1.0\}$, assume exact fit, and let $y|_{x=10} = 100$. Performing an L_1 estimate of the parameters of the model function, we obtain the fitted model parameters $\hat{\underline{\theta}} = \{0.0, 1.0\}$. Perturbing the outlier further, we move it to $x = 22$ and let $y|_{x=22} = 100$. We again obtain the L_1 estimate $\hat{\underline{\theta}} = \{0.0, 1.0\}$. The data and the fitted response are shown in Figure 3. Other x_i and y_i values for the outlier perform similarly. This confirms our expectations.

5.3 Bent Hyperbola Model

Consider the bent hyperbola model

$$g(x_i, \underline{\theta}) = \theta_1 + \theta_2(x_i - \theta_4) + \theta_3 \left[(x_i - \theta_4)^2 + \theta_5\right]^{1/2} \tag{92}$$

studied by St. Laurent and Cook [25] and Ratkowsky [9]. Ratkowsky fit this model to the data in Table 1 relating the logarithm of the stagnant surface layer height Y to the logarithm of the flow rate of water down an inclined channel X.

Using the starting point $\underline{\theta} = \{0.585, -0.735, -0.359, 0.062, 0.096\}$, we perform an L_1 estimate of the parameters of the model function (92) for the band height data in Table 1. The fitted model parameters were $\hat{\underline{\theta}} = \{0.5893, -0.7284, -0.3582, 0.0573, 0.1151\}$. The data and the fitted response are shown in Figure 4.

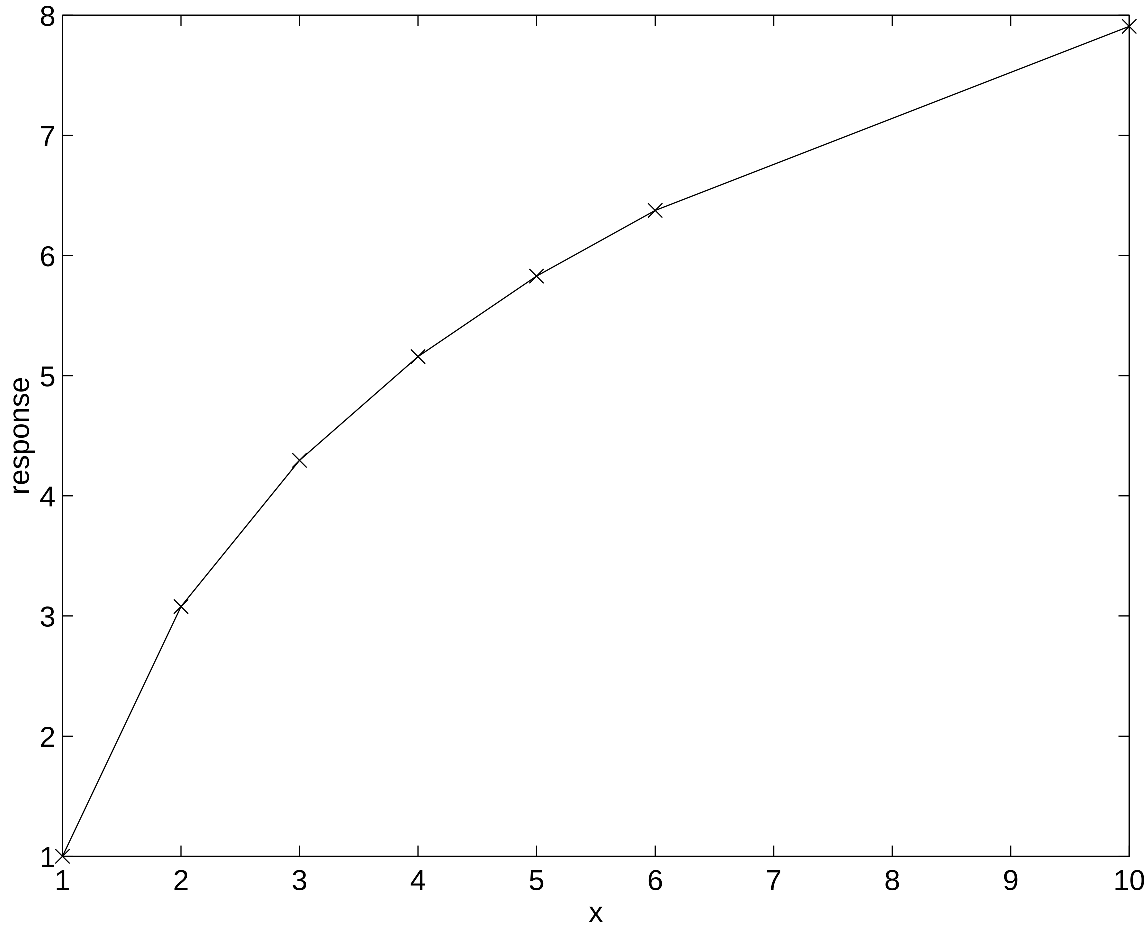

Figure 2: Fit of Logarithmic Model at True Data

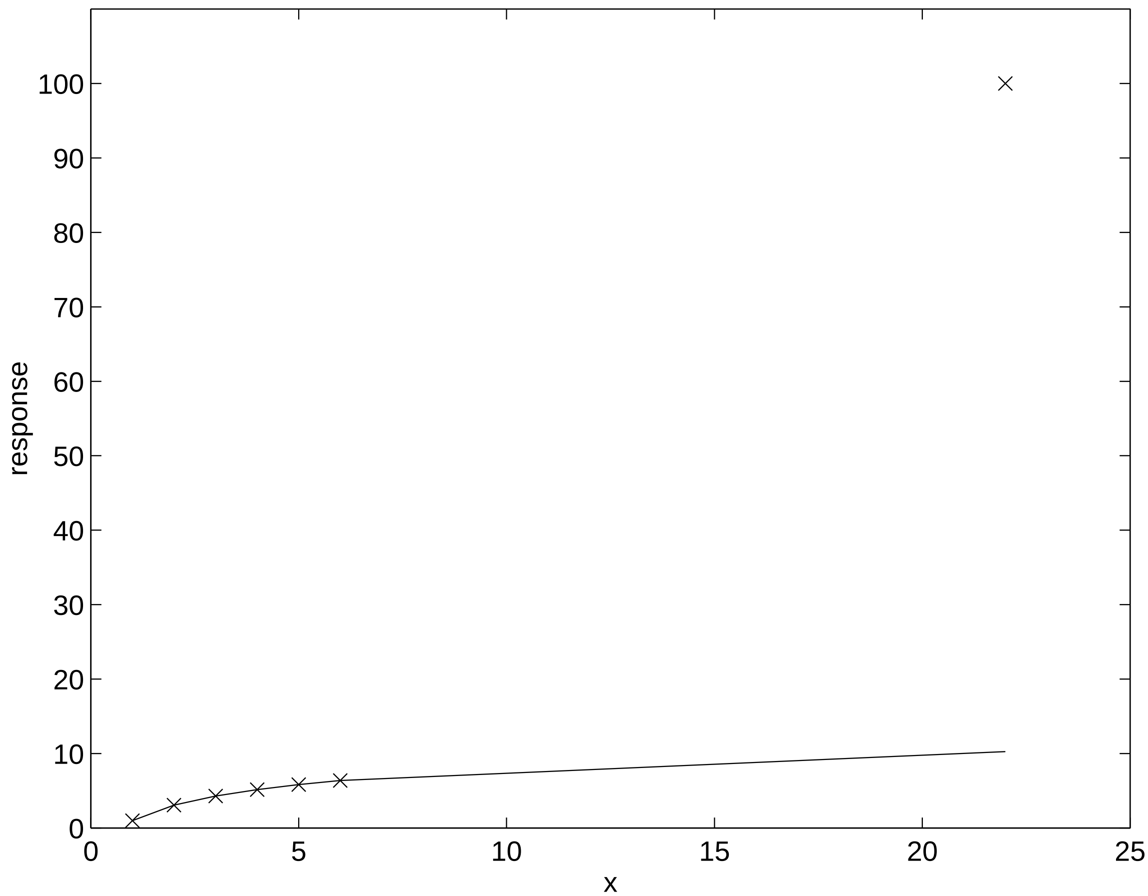

Figure 3: Fit of Logarithmic Model at True Data with Large x_i and y_i Perturbation

Table 1: Band Height Data

X	Y
-1.39	1.12
-1.39	1.12
-1.08	0.99
-1.08	1.08
-0.94	0.92
-0.80	0.90
-0.63	0.81
-0.63	0.83
-0.25	0.65
-0.25	0.67
-0.12	0.60
-0.12	0.59
0.01	0.51
0.11	0.44
0.11	0.43
0.11	0.43
0.25	0.33
0.25	0.30
0.34	0.24
0.34	0.25
0.44	0.13
0.59	-0.01
0.70	-0.13
0.70	-0.14
0.85	-0.30
0.85	-0.33
0.99	-0.46
0.99	-0.43
1.19	-0.65

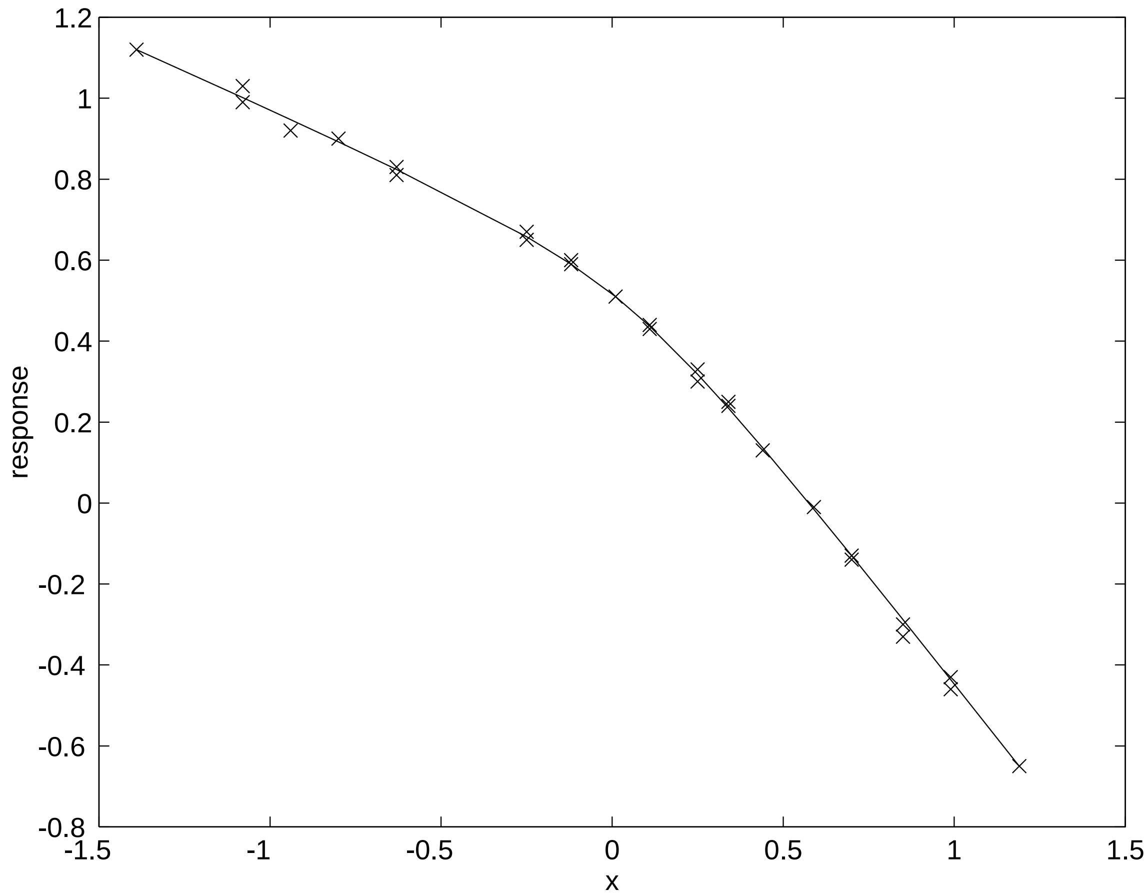

Figure 4: Fit of Bent-Hyperbola Model to Band Height Data

5.3.1 Influence due to Outlying Observations

Using the results from Chapter 4, we study the effect of observations outlying in X on M-estimators at the model (92). Calculating the Jacobian vector of the model function (92), we find

$$\underline{J} = \frac{\partial g(x_i, \underline{\theta})}{\partial \underline{\theta}} = \begin{bmatrix} 1 \\ \dfrac{x_i - \theta_4}{\left[(x_i - \theta_4)^2 + \theta_5\right]^{1/2}} \\ -\theta_2 - \frac{\theta_3(x_i - \theta_4)}{\sqrt{\theta_5 + (x_i - \theta_4)^2}} \\ \frac{\theta_3}{2\sqrt{\theta_5 + (x_i - \theta_4)^2}} \end{bmatrix} . \tag{93}$$

Application of simple calculus shows that $\lim_{x\to\infty} \frac{\partial g}{\partial \theta_2} = \infty$, $\lim_{x\to\infty} \frac{\partial g}{\partial \theta_3} = \infty$, $\lim_{x\to-\infty} \frac{\partial g}{\partial \theta_2} = -\infty$, and $\lim_{x\to-\infty} \frac{\partial g}{\partial \theta_3} = \infty$. Since the Jacobian in the influence function of the M-estimators of nonlinear models is unbounded, and these components of the Jacobian of this model are unbounded in x, we expect that M-estimators at this model will be highly influenced by outlying x_i.

To confirm our expectations, we perturb the x_i component of a single observation from Table 1 such that the observation becomes an outlier in the factor space X. Specifically, we move the first observation $(-1.39, 1.12)$ to $(-10, 1.12)$. Performing an L_1 estimate of the parameters of the model function, we obtain the fitted model parameters $\hat{\underline{\theta}} = \{1.5094, -0.7563, -0.7880, 0.07520, 1.7910\}$. The perturbed data and the fitted response are shown in Figure 5. We observe that the fitted response follows the outlier exactly. This confirms our expectation of the low tolerance of the M-estimators to factor space outliers at this model.

5.3.2 Influence due to Observations in X

Now we study the effect of observations not outlying in X on M-estimators at the model (92). Since the influence function of the M-estimators of nonlinear models is a nonlinear function of the Jacobian of the model, we expect that there will exist x_i that cause unbounded influence. Specifically, we conjecture that observations not outlying in X can have influence on the M-estimators at this model.

To confirm our expectations, we perturb the y component of a single observation from Table 1 without perturbing the x component. Specifically, we move the first observation

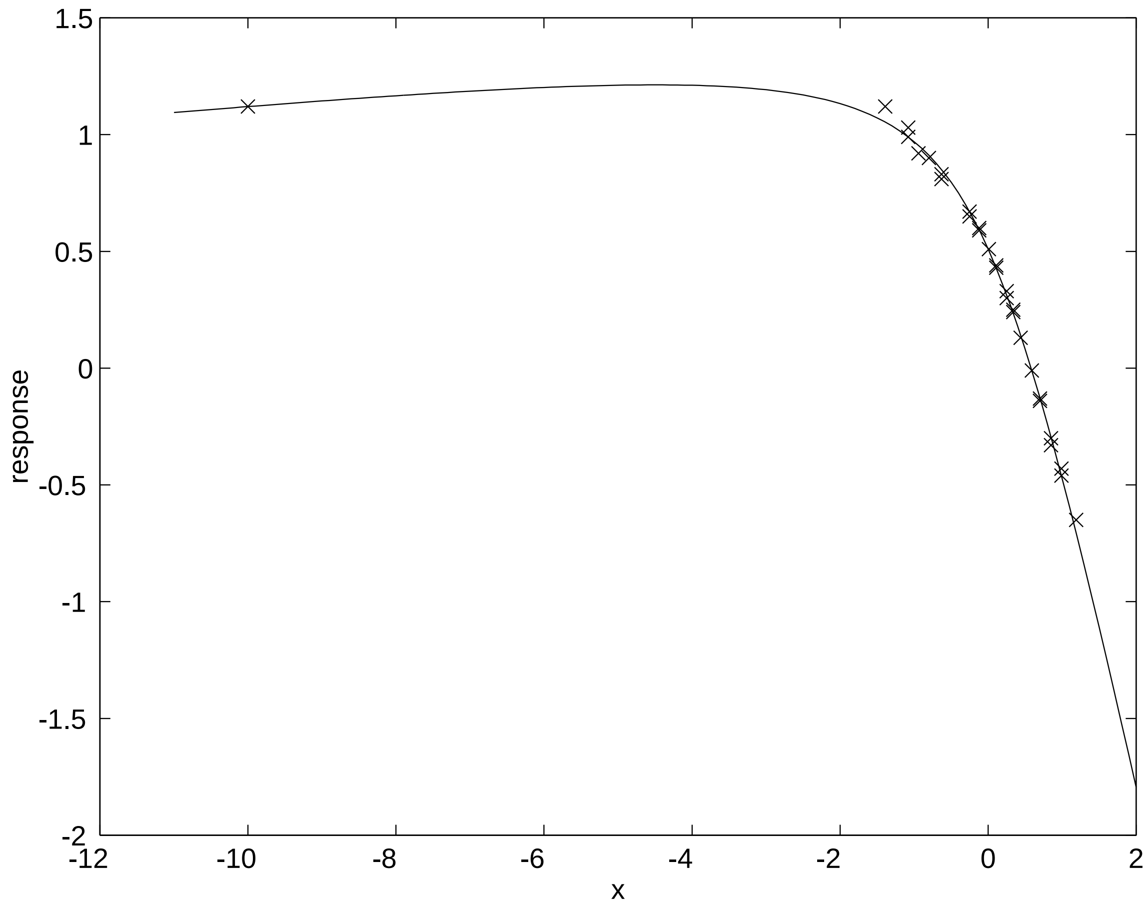

Figure 5: Fit of Bent-Hyperbola Model to Band Height Data with Large x_i Perturbation

$(-1.39, 1.12)$ to $(-1.39, 0.9)$. Performing an L_1 estimate of the parameters of the model function, we obtain the fitted model parameters $\hat{\underline{\theta}} = \{0.6854, -0.7051, -0.3934, 0.0013, 0.1850\}$. The perturbed data and the fitted response are shown in Figure 6. We observe that the fitted response was influenced by the perturbation despite the existence of a duplicate observation at the pre-perturbation location.

Further insight is obtained by perturbing the y components of the first two observations from Table 1 without perturbing the x components. To this end, we move the first two observations (which are identical) from $(-1.39, 1.12)$ to $(-1.39, 0.9)$. Performing an L_1 estimate of the parameters of the model function, we obtain the fitted model parameters $\hat{\underline{\theta}} = \{4.6584, -0.1919, -1.6226, -0.8225, 5.3587\}$. They are *significantly* different that the parameters at the unperturbed data. The perturbed data and the fitted response are shown in Figure 7. We observe that the fitted response follows the perturbations exactly. This example re-confirms our expectation of the low tolerance of the M-estimators to non-outlying, vertically perturbed observations at this model. Also, this example highlights the dangers and non-robustness of the M-estimators applied to nonlinear model estimation.

5.4 Logistic Model

Consider the continuous logistic model

$$g(x_i, \underline{\theta}) = \frac{1}{1 + e^{-(\theta_1 + \theta_2 x_i)}} \tag{94}$$

studied by Stromberg [23] at the design $x_i = \{-4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7, 8\}$. First, we study the exact fit properties of the M-estimators at this model with the true parameter $\tilde{\underline{\theta}} = \{1, 1\}$ (see Table 2). Second, we study the properties of the M-estimators at this model with contrived data due to Stromberg (see Table 3).

Using the starting point $\underline{\theta} = \{0.9, 1.0\}$, we perform an L_1 estimate of the parameters of the model function (94) for the data in Table 2. The fitted model parameters are $\hat{\underline{\theta}} = \{1.0, 1.0\}$. The data and the fitted response are shown in Figure 8. For the contrived data, we use the starting point $\underline{\theta} = \{-0.7, 1.2\}$ and perform an L_1 estimate of the parameters of the model for the data in Table 3. The fitted model parameters are $\hat{\underline{\theta}} = \{-0.7754, 0.7834\}$. The data and the fitted response are shown in Figure 9.

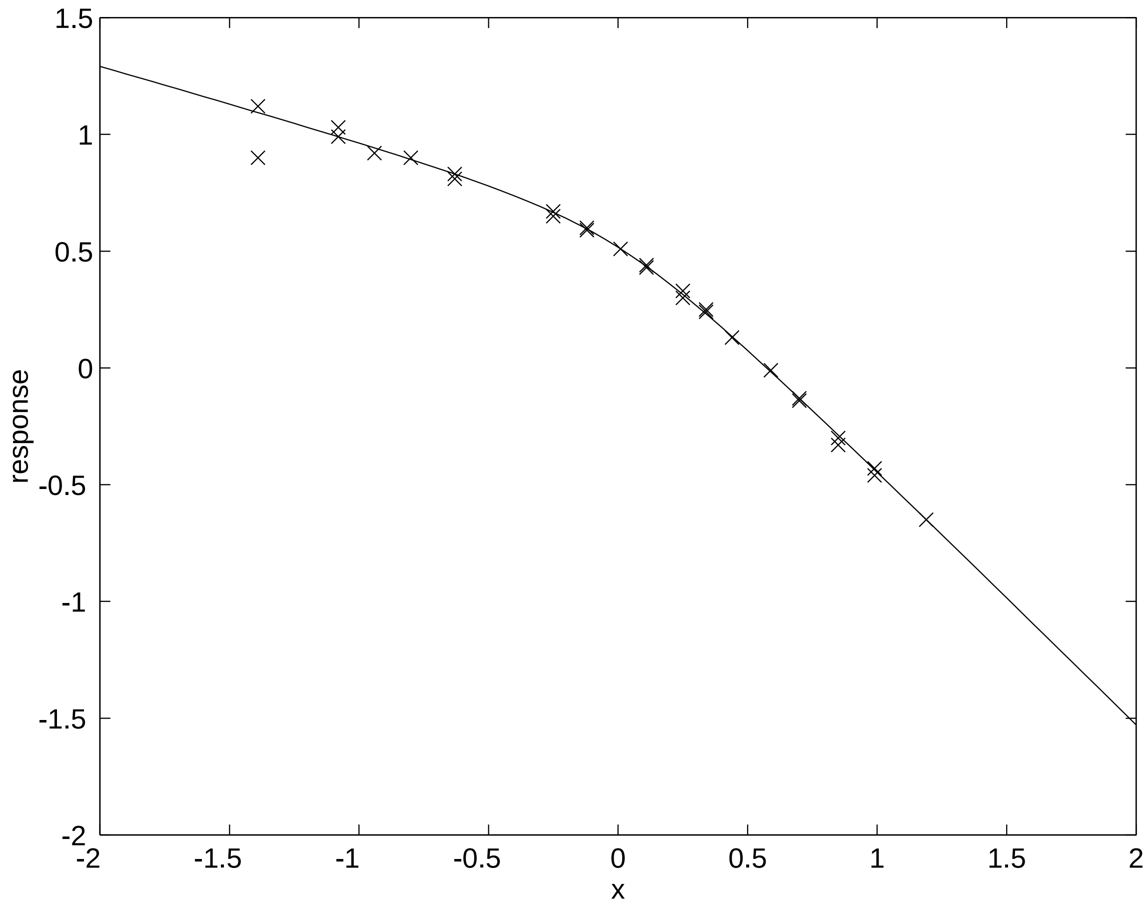

Figure 6: Fit of Bent-Hyperbola Model to Band Height Data with Non-outlying Vertical Perturbation

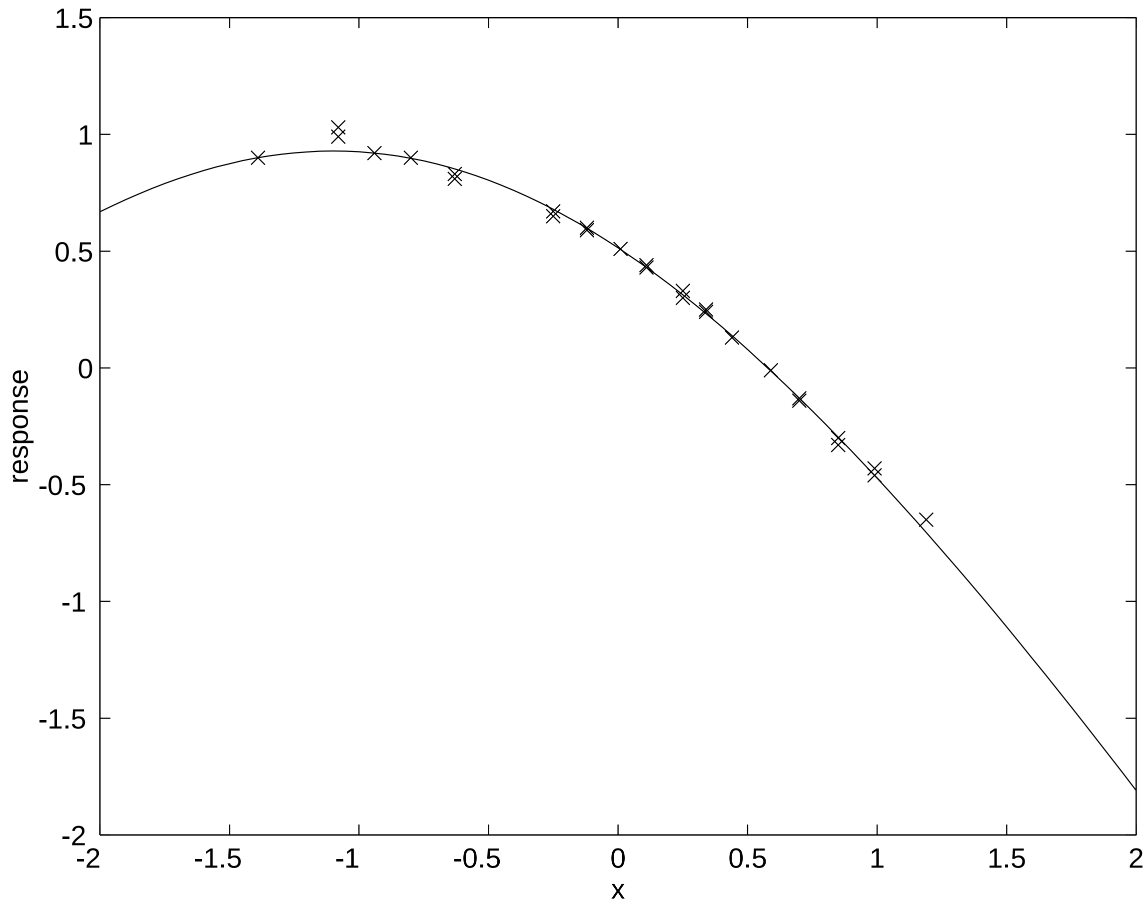

Figure 7: Fit of Bent-Hyperbola Model to Band Height Data with Two Non-outlying Vertical Perturbation

Table 2: Logistic Model Data at True Parameter

X	Y
-4	0.0474
-3	0.1192
-2	0.2689
-1	0.5000
0	0.7311
1	0.8808
2	0.9526
3	0.9820
4	0.9933
5	0.9975
6	0.9991
7	0.9997
8	0.9999

Table 3: Logistic Model Contrived Data due to Stromberg

X	Y
-4	0.0219
-3	0.0001
-2	0.1064
-1	0.1738
0	0.3315
1	0.4893
2	0.6899
3	0.8075
4	0.9136
5	0.9999
6	0.9999
7	0.9803
8	0.9998

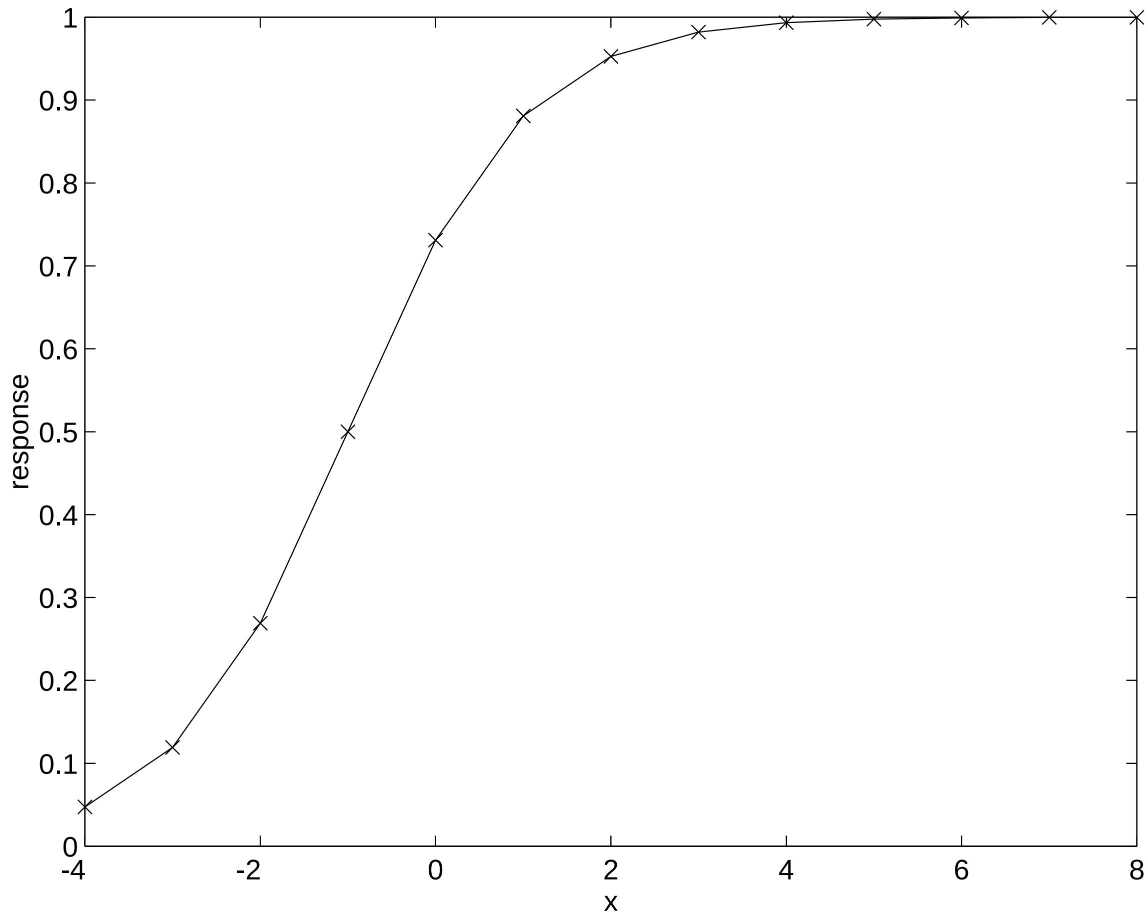

Figure 8: Fit of Logistic Model at True Data

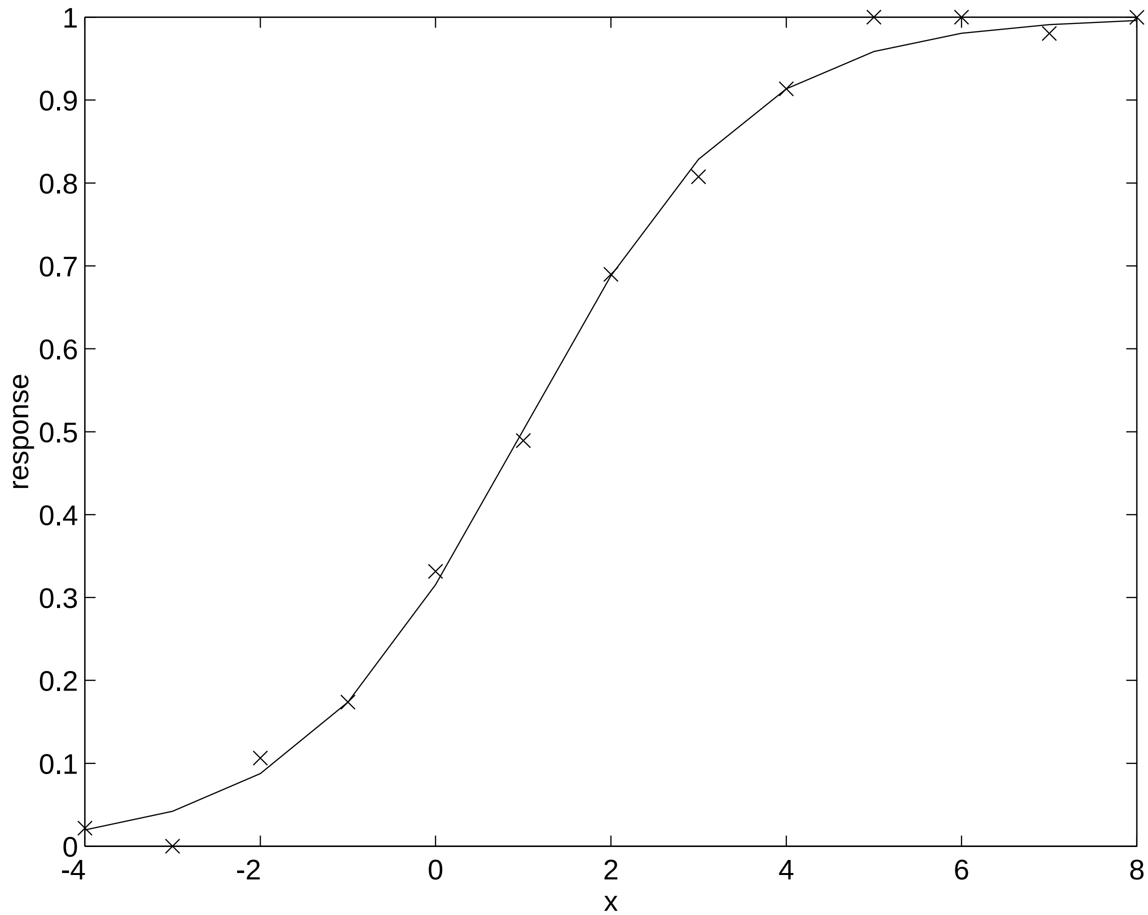

Figure 9: Fit of Logistic Model at Contrived Data

5.4.1 Influence due to Outlying Observations

The effect of observations outlying in X on M-estimators at the model (94) are now analyzed. Calculating the Jacobian vector of the model function, we find

$$\underline{J} = \frac{\partial g(x_i, \underline{\theta})}{\partial \underline{\theta}} = \begin{bmatrix} \frac{e^{-(\theta_1+\theta_2 x_i)}}{(1+e^{-(\theta_1+\theta_2 x_i)})^2} \\ \frac{x_i e^{-(\theta_1+\theta_2 x_i)}}{(1+e^{-(\theta_1+\theta_2 x_i)})^2} \end{bmatrix}. \tag{95}$$

We have $\lim_{x\to\infty} \frac{\partial g}{\partial \theta_1} = 0$, $\lim_{x\to\infty} \frac{\partial g}{\partial \theta_2} = 0$, $\lim_{x\to-\infty} \frac{\partial g}{\partial \theta_1} = 0$, and $\lim_{x\to-\infty} \frac{\partial g}{\partial \theta_2} = 0$. Since the Jacobian in the influence function of the M-estimators of nonlinear models is unbounded and directly influenced by the observations, we expect that M-estimators at this model will not be influenced by outlying x_i as the components of the Jacobian are not unbounded. To confirm this, we perturb the last observation in Table 2 from $(8.0, 0.9999)$ to $(15.0, 2.0)$. Performing an L_1 estimate of the parameters, we obtain the fitted parameters $\hat{\underline{\theta}} = \{1.0, 1.0\}$. The perturbed data and the fitted response are shown in Figure 10. We observe that the fitted response was not influenced by the perturbation.

5.4.2 Influence due to Observations in X

Although outlying x_i are not influential, the influence function of the M-estimators of nonlinear models is a nonlinear function of the Jacobian of the model. Additionally, when we do not ignore the Hessian in the influence function, the observed response can have influence despite the usual resistance of the M-estimators to vertical outliers (e.g. M-estimators of linear models). Therefore, we expect that there will exist non-outlying x_i that will influence the M-estimators at this model. To confirm our expectations, we perform several experiments where we perturb observations vertically.

First, we perturb the y_i component of a single observation from Table 2 without perturbing the x_i component. Specifically, we move the third observation from $(-2, 0.2689)$ to $(-2, 0.1)$. Performing an L_1 estimate of the parameters of the model function, we obtain the fitted model parameters $\hat{\underline{\theta}} = \{1.0, 1.0\}$. The perturbed data and the fitted response are shown in Figure 11. We observe that the fitted response was not influenced by the perturbation.

Second, we perturb the response of several observations from Table 2 without perturbing the x_i component. Specifically, we move $(-4, 0.0474)$ to $(-4, 0.01)$, $(-3, 0.1992)$ to $(-3, 0.03)$, and $(-2, 0.2689)$ to $(-2, 0.1)$. Performing an L_1 estimate of the parameters of the model function, we obtain the parameters $\hat{\underline{\theta}} = \{1.0, 1.3987\}$. The perturbed data and

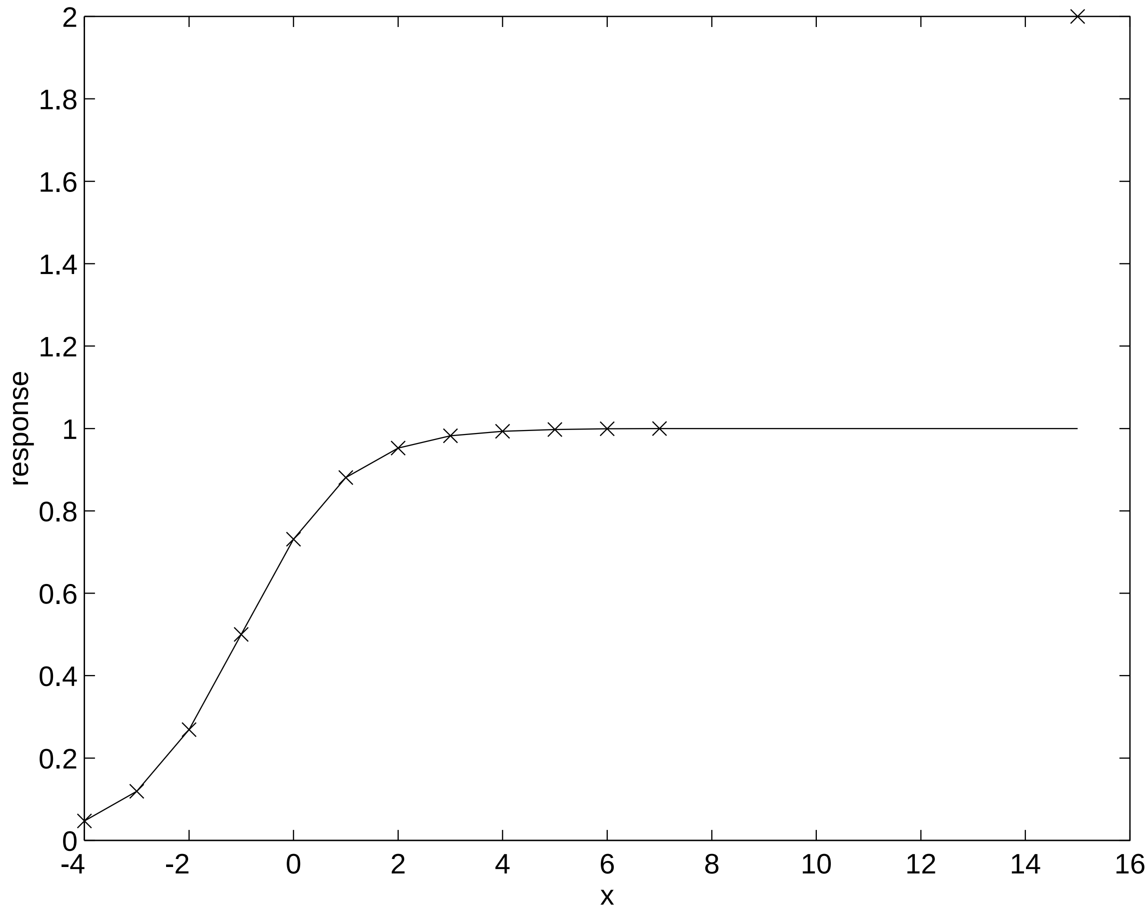

Figure 10: Fit of Logistic Model at True Data with Single Leverage Point

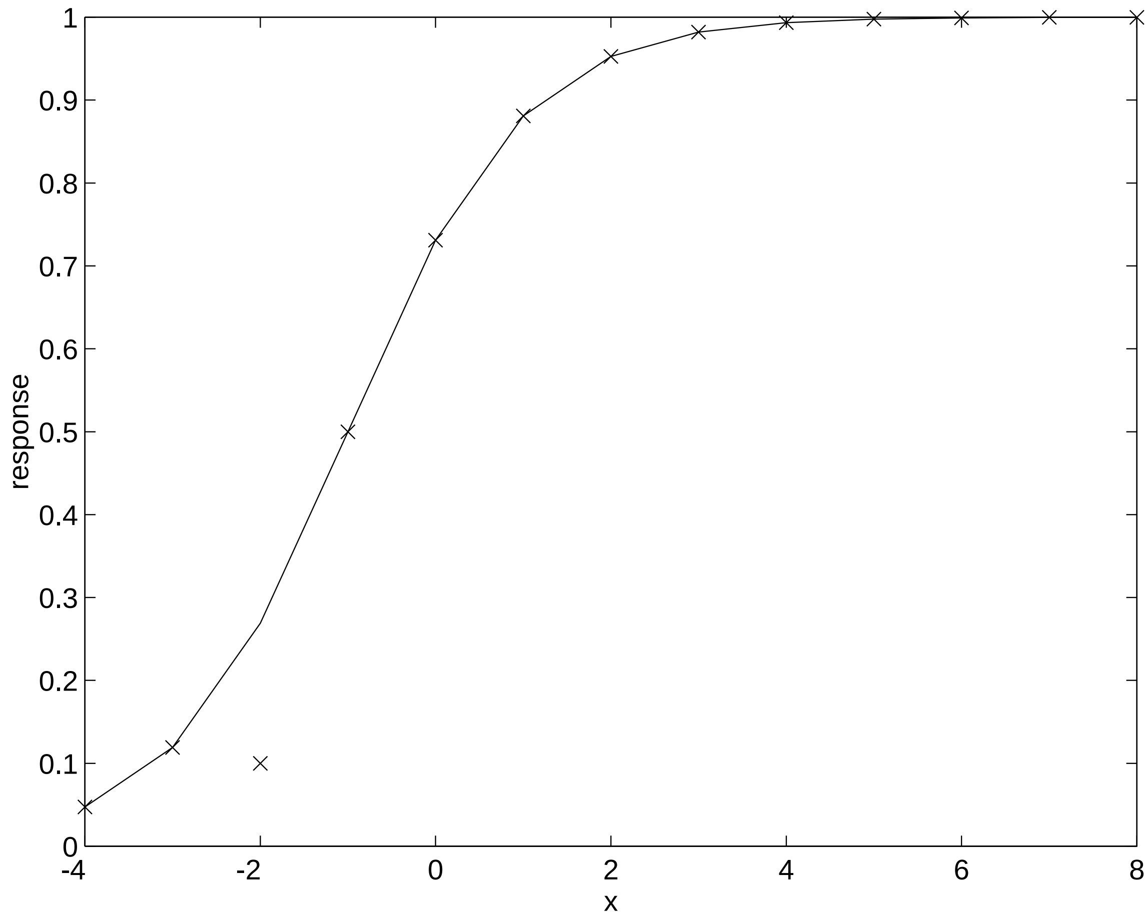

Figure 11: Fit of Logistic Model at True Data with Single Vertical Perturbation

the fitted response are shown in Figure 12. We observe that the estimator was directly influenced by the perturbed response and that the fitted response follows the perturbed observation at $(-4, 0.01)$ *exactly*. This example re-confirms our expectation of the low tolerance of the M-estimators to non-outlying, vertically perturbed observations at this model.

Third, we influence the estimate of the model parameters by perturbing the response of a single observation from Table 3 without perturbing the x_i component. Specifically, we move $(-2, 0.1064)$ to $(-2, 0.1)$. Performing an L_1 estimate of the parameters of the model function, we obtain the fitted model parameters $\hat{\underline{\theta}} = \{-0.8432, 0.8004\}$. The perturbed data and the fitted response are shown in Figure 13. We observe that the fitted response was influenced by the single perturbation, re-confirming our expectations.

Finally, we emphasize the nonlinear nature of the influence of model and the dependence of influence on the magnitude of the response. We attempt to influence the M-estimators at this model by perturbing the response of a single observation from Table 3 without perturbing the x_i component. Specifically, we move $(-2, 0.1064)$ to $(-2, 0.8)$. Performing an L_1 estimate of the parameters of the model function, we obtain the fitted model parameters $\hat{\underline{\theta}} = \{-0.7754, 0.7834\}$, the same fit obtained for the unperturbed contrived data. The perturbed data and the fitted response are shown in Figure 14. We conclude that influence is dependent not only only on the model but also on the observations.

5.5 Pharmacokinetic Model

Consider the model

$$g(x_i, \underline{\theta}) = \frac{\theta_1 \theta_3}{\theta_1 - \theta_2} \left[\exp\{-\theta_2 t\} - \exp\{-\theta_1 t\}\right] \tag{96}$$

studied by Gonin et al. [12]. The model relates a drug concentration Y in the serum to time T. Gonin et al. studied the exact fit properties of this model at the true parameter $\tilde{\underline{\theta}} = \{3, 0.3, 50\}$ for the contrived data in Table 4.

Using the starting point $\underline{\theta} = \{25, 1, 10\}$, we perform an L_1 estimate of the parameters of the model function (96) for the contrived data in Table 4. The fitted model parameters were $\hat{\underline{\theta}} = \{3, 0.3, 50\}$ which are equal to the true parameters. The data and the fitted response are shown in Figure 15.

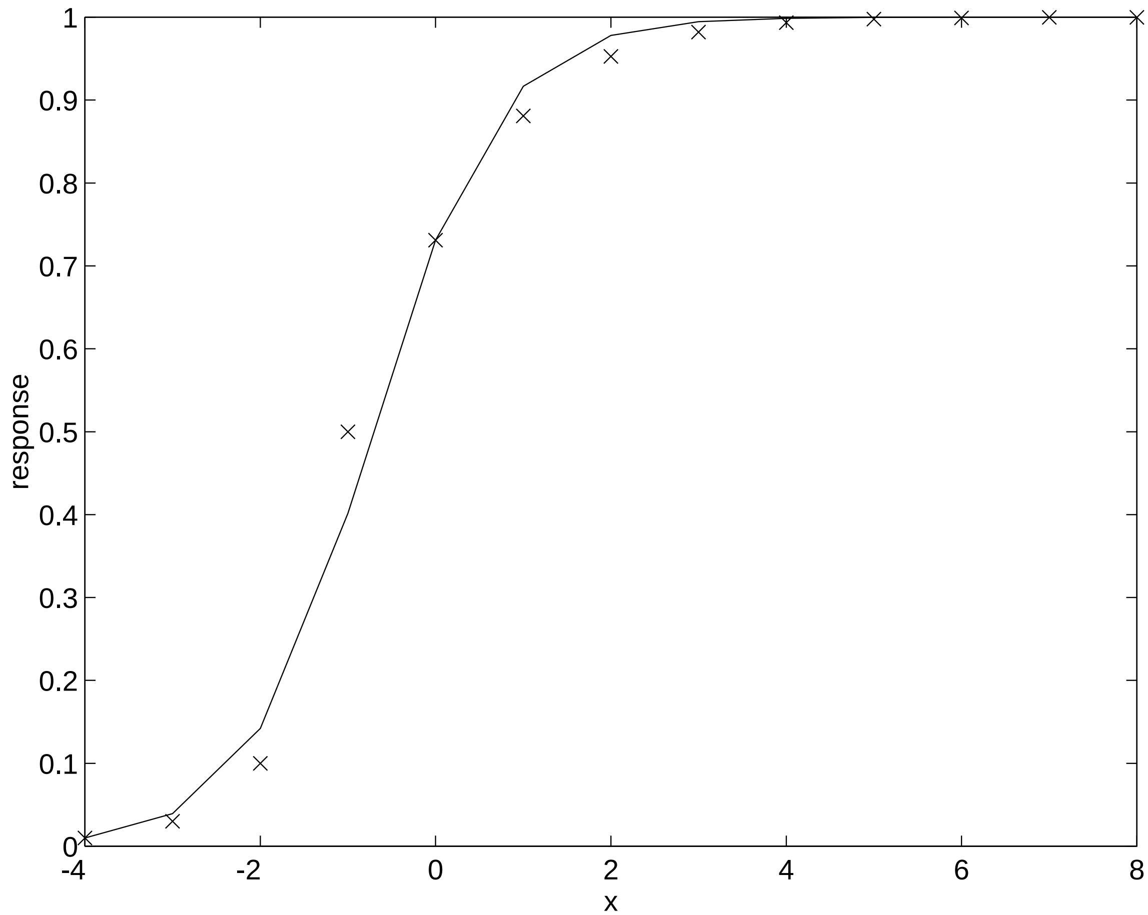

Figure 12: Fit of Logistic Model at True Data with Three Vertical Perturbations

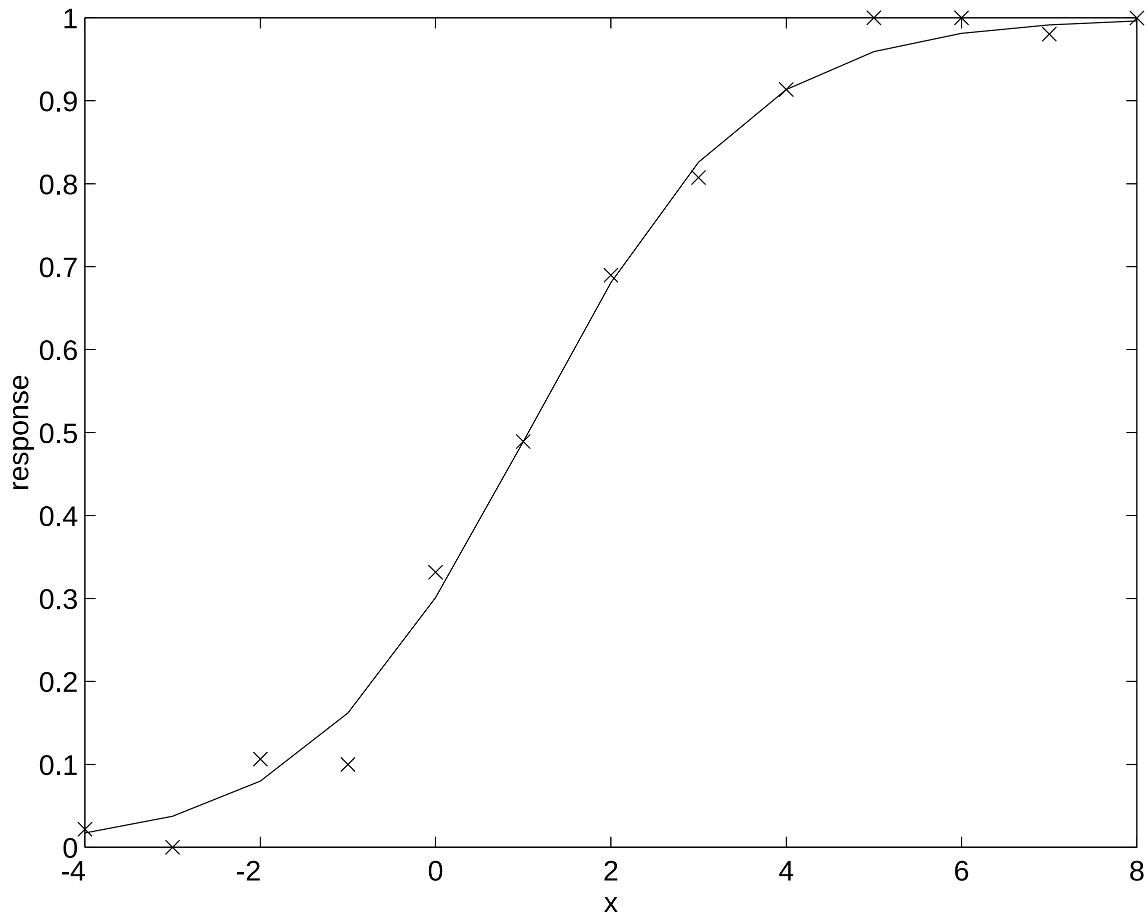

Figure 13: Fit of Logistic Model at Contrived Data with Single Influential Vertical Perturbation

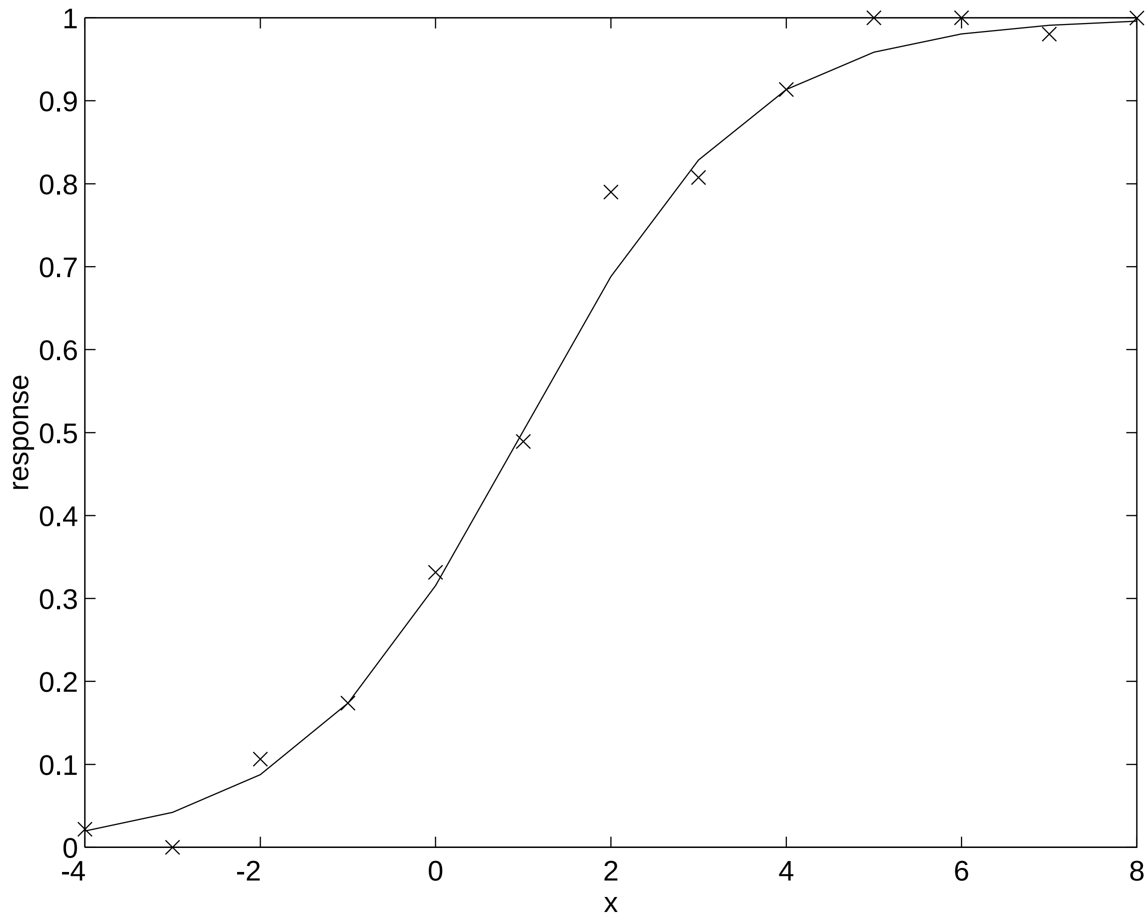

Figure 14: Fit of Logistic Model at Contrived Data with Single Non-Influential Vertical Perturbation

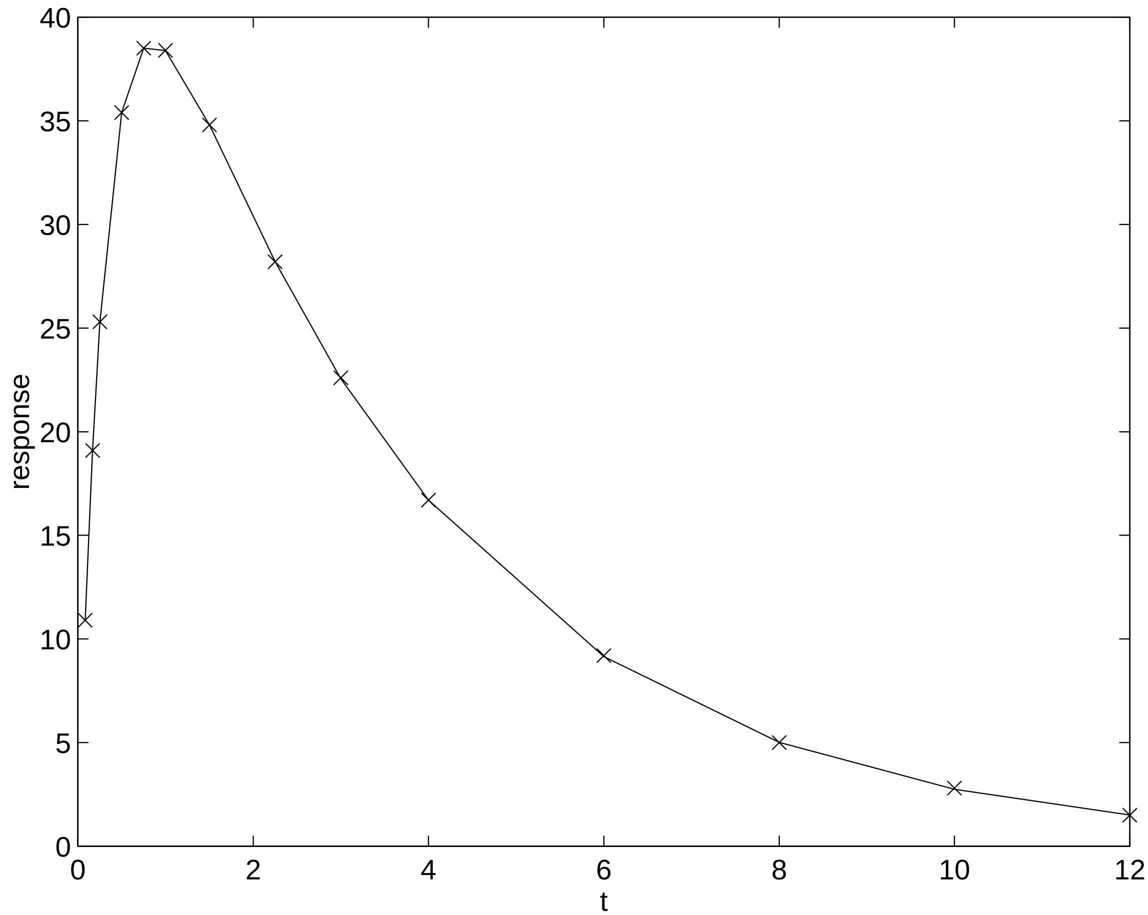

Figure 15: Fit of Pharmacokinetic Model to Contrived Data

Table 4: Pharmacokinetic Model Contrived Data due to Gonin

X	Y
0.083	10.900
0.167	19.100
0.250	25.300
0.500	15.000
0.750	38.500
1.000	38.400
1.500	34.154
2.250	28.200
3.000	22.600
4.000	16.700
6.000	9.200
8.000	5.000
10.000	2.800
12.000	1.500

5.5.1 Influence due to Outlying Observations

We study the effect of observations outlying in X on M-estimators at the model (96).

Calculating the Jacobian vector of the model function, we find

$$\underline{J} = \frac{\partial g(x_i, \underline{\theta})}{\partial \underline{\theta}} = \begin{bmatrix} \frac{\theta_3}{(\theta_2-\theta_1)^2 e^{t(\theta_1+\theta_2)}} \left[t\theta_1^2 e^{t\theta_2} - \theta_2 e^{t\theta_1} + \theta_2 e^{t\theta_2} - t\theta_1\theta_2 e^{t\theta_2} \right] \\ \frac{\theta_1\theta_3(e^{-t\theta_2} - e^{-t\theta_1})}{(\theta_1-\theta_2)^2} - \frac{t\theta_1\theta_3}{e^{t\theta_2}(\theta_1-\theta_2)} \\ \frac{\theta_1(e^{-t\theta_2} - e^{-t\theta_1})}{\theta_1-\theta_2} \end{bmatrix}. \tag{97}$$

We have $\lim_{x\to\infty} \frac{\partial g}{\partial \theta_1} = 0$, $\lim_{x\to\infty} \frac{\partial g}{\partial \theta_2} = 0$, and $\lim_{x\to\infty} \frac{\partial g}{\partial \theta_3} = 0$. Since the Jacobian in the influence function of the M-estimators of nonlinear models is unbounded and directly influenced by the observations, we expect that M-estimators at this model will not be influenced by outlying x_i as they do not cause large Jacobians.

To show this, we perturb the x_i component of a single observation from Table 4 such that the observation becomes an outlier in the factor space X. Specifically, we move the last observation from $(12.0, 1.5)$ to $(100, 1.5)$. Performing an L_1 estimate of the parameters of the model function, we obtain the true model parameters $\tilde{\underline{\theta}}$. The perturbed data and the fitted response are shown in Figure 16. We observe that the estimator is resistant to outlying x_i at this model confirming our expectations.

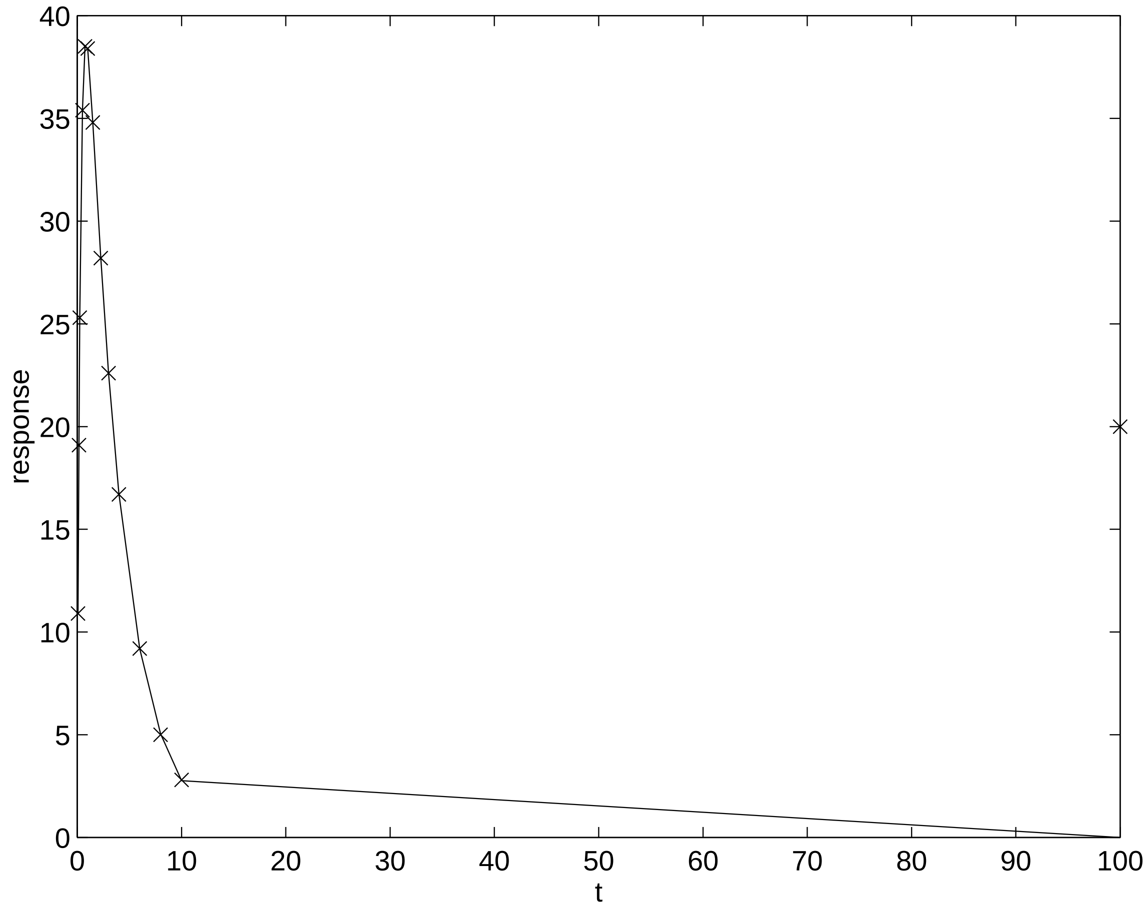

Figure 16: Fit of Pharmacokinetic Model to Contrived Data with Large x_i Perturbation

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this thesis we have focused on studying and improving the robustness properties of M-estimators of nonlinear models. We presented and evaluated a technique due to St. Laurent and Cook called Jacobian Leverage. We pointed out the relationship of the technique to the robustness concept of influence, namely that Jacobian Leverage measures influence but does not measure the influence of an arbitrary contaminant. We discussed several limitations of the technique, including its inability to reveal multiple influential observations. The latter was demonstrated by example.

We derived the influence function for M-estimators of nonlinear models. The result shows that influence of position becomes, more generally, influence of model. However, if the Hessian of the model is significant and assumptions are not made about the properties of the error distribution, the influence function is a nonlinear expression involving residual, position, and model. Regardless, the influence function may be interpreted heuristically to give valuable information about the sensitivity of the M-estimators to arbitrary models and contaminants. Using this approach, we demonstrated the utility of the influence function by analyzing several models and data sets from the literature. The analysis showed the existence of a new type of influential observation that does not exist in linear regression. These are highly influential observations that are not outlying in factor space.

6.2 Future Work

The most important future work will propose and analyze methods for identifying highly influential observations and techniques for robustifying M-estimators of nonlinear models. Complete analysis of the new influence function is the first step in this direction. The influence function will be used to evaluate robust estimators of nonlinear models. The current work has shown a new category of influential observation which must be studied and understood. It is possible that new estimators will be required to handle this new type of influential observation.

A variety of other work is important. Following the interpretation of Hampel et al. [29], it is especially important to develop bias curves for M-estimators of nonlinear models. The bias curve will provide a more complete view of the robustness properties of these estimators. A first step is to extend the work of Rousseeuw and Croux [33] to the M-estimators of nonlinear location and, eventually, to the M-estimators of nonlinear regression.

Long-term work will apply more sophisticated mathematical tools, including higher-order von Mises calculus and differential geometry. Amari [38], Murray and Rice [37], and Pazman [39] are among those who are building a theory of statisical estimation based on differential geometry. Extension of such work to robust statistical theory would be a very powerful tool for analysis of nonlinear model estimation. Higher-order von Mises calculus and more sophisticated functional analysis will provide better and more accurate approximations and analysis of nonlinear models and estimators of nonlinear models.

Bibliography

[1] J. A. Cadzow, "Signal processing via least squares error modeling," *IEEE ASSP Magazine*, vol. 7, pp. 12–31, October 1990.

[2] R. E. Ewing, M. S. Pilant, J. G. Wade, and A. T. Watson, "Estimating parameters in scientific computation: a survey of experience from oil and groundwater modeling," *IEEE Computational Science and Engineering*, pp. 19–31, 1994.

[3] L. Lobo and M. Lobo, "Robust and efficient nonlinear regression of kinetic systems using a direct search method," *Computers & Chemical Engineering*, vol. 15, pp. 141–144, February 1991.

[4] S. Kassam and H. Poor, "Robust techniques for signal processing: a survey," *Proceedings of the IEEE*, vol. 73, pp. 433–481, 1985.

[5] S. Johansen, *Functional Relations, Random Coefficients, and Nonlinear Regression with Application to Kinetic Data*. No. 22 in Lecture Notes in Statistics, New York, New York: Springer-Verlag, 1984.

[6] G. Seber and C. Wild, *Nonlinear Regression*. New York, New York: John Wiley & Sons, 1988.

[7] D. M. Bates and D. G. Watts, *Nonlinear Regression Analysis and its Applications*. New York, New York: John Wiley & Sons, 1988.

[8] D. A. Ratkowsky, *Handbook of Nonlinear Regression Models*. New York, New York: Marcel Dekker, 1990.

[9] D. A. Ratkowsky, *Nonlinear Regression Modeling: A Unified Practical Approach*. Marcel Dekker, 1983.

[10] K. D. Lawrence and J. L. Arthur, eds., *Robust Regression: Analysis and Applications*. New York, New York: Marcel Dekker, 1990.

[11] A. R. Gallant, *Nonlinear Statistical Models.* New York, New York: John Wiley & Sons, 1987.

[12] R. Gonin and A. H. Money, *Nonlinear L_p-norm Estimation.* New York, New York: Marcel-Dekker, 1989.

[13] R. Dutter and P. Huber, "Numerical methods for the nonlinear robust regression problem," *J. Stat. Comput. Simulation*, vol. 13, pp. 79–114, 1981.

[14] A. Stromberg, "Computation of high breakdown nonlinear regression parameters," *Journal of the American Statistical Association*, vol. 88, pp. 237–244, March 1993.

[15] K. Humak, *Statistical Methods of Model Building, Volume II: Nonlinear Regression, Functional Relations and Robust Methods.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, Chichester: John Wiley & Sons, 1989.

[16] D. Birkes and Y. Dodge, *Alternative Methods of Regression.* Wiley Series in Probability and Mathematical Statistics, New York, New York: John Wiley & Sons, 1993.

[17] A. Stromberg, *Robust Efficient Estimation of Nonlinear Regression Parameters.* PhD thesis, University of North Carolina, 1989.

[18] A. Stromberg, "Consistency of the least median of squares estimator in nonlinear regression," *Communications in Statistics–Theory and Methods*, vol. 24, no. 8, pp. 1971–1984, 1995.

[19] A. Tabatabiani, "Robust estimation and testing for general nonlinear regression models," *Applied Mathematics and Computation*, vol. 58, no. 1, pp. 85–101, 1993.

[20] Y. Zhao and A. H. Lee, "Assessment of influence in non-linear measurement error models," *Journal of Applied Statistics*, vol. 22, pp. 215–225, 1995.

[21] E. Ronchetti, "Robust model selection in regression," *Statistics and Probability Letters*, vol. 3, pp. 21–23, 1985.

[22] F. Hampel, "Nonlinear estimation in linear models," in *Statistical Data Analysis and Inference* (Y. Dodge, ed.), pp. 19–32, Elsevier Science Publishers B.V. (North–Holland), 1989.

[23] A. J. Stromberg and D. Rupert, "Breakdown in nonlinear regression," *JASA: Theory and Methods*, vol. 87, no. 420, pp. 991–997, 1992.

[24] S. Sakata and H. White, "An alternative definition of finite-sample breakdown point with applications to regression model estimators," *Journal of the American Statistical Association*, vol. 90, pp. 1099–1106, September 1995.

[25] R. T. St. Laurent and R. D. Cook, "Leverage and superleverage in nonlinear regression," *JASA, Theory and Methods*, vol. 87, pp. 985–990, December 1992.

[26] R. T. St. Laurent and R. D. Cook, "Leverage, local influence, and curvature in nonlinear regression," *Biometrika*, vol. 80, no. 1, pp. 99–106, 1993.

[27] P. J. Huber, *Robust Statistics.* Wiley Series in Probability and Mathematical Statistics, New York, New York: John Wiley & Sons, 1981.

[28] F. R. Hampel, "The influence curve and its role in robust estimation," *JASA: Theory and Methods*, vol. 69, pp. 383–393, June 1974.

[29] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions.* New York, New York: John Wiley & Sons, 1986.

[30] D. C. Hoaglin, F. Mosteller, and J. W. Tukey, *Understanding Robust and Exploratory Data Analysis.* New York, New York: John Wiley & Sons, 1983.

[31] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection.* New York, New York: John Wiley & Sons, 1987.

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

[33] P. J. Rousseeuw and C. Croux, "The bias of k-step m-estimators," Tech. Rep. 93–06, Dept. of Mathematics and Computer Science, University of Antwerp, UIA, Belgium, March 1993.

[34] D. M. Bates and D. G. Watts, "Relative curvature measures of nonlinearity," *Journal of the Royal Statistical Society Series B*, vol. 42, no. 1, pp. 1–25, 1980.

[35] D. W. Stroock, *A Concise Introduction to the Theory of Integration.* Boston: Birkhäuser, second ed., 1994.

[36] A. Graham, *Kronecker Products and Matrix Calculus with Applications.* New York, New York: Ellis Horwood Limited, Division of John Wiley & Sons, 1981.

[37] M. K. Murray and J. W. Rice, *Differential Geometry and Statistics.* No. 48 in Monographs on Statistics and Applied Probability, Chapman and Hall, 1993.

[38] S. Amari, *Differential-Geometric Methods in Statistics*, vol. 28 of *Lecture Notes in Statistics.* New York, New York: Springer-Verlag, 1985.

[39] A. Pazman, *Nonlinear Statistical Models.* Kluwer Academic Publishers, 1993.

Vita

Shawn Patrick Neugebauer was born on March 13, 1971 in Silver Spring, Maryland. Shawn began his undergraduate studies in August 1989 at Virginia Polytechnic Institute and State University. While pursuing his degree in electrical engineering, he worked at Ideas, Inc. in Columbia, Maryland as a participant in the cooperative education program. During his senior year, Shawn became a dual registrant, completed several graduate classes, and performed graduate research as a member of the Mobile and Portable Radio Research Group (MPRG) under the guidance of Dr. Jeffrey H. Reed. The research culminated in the paper "Pseudorandom sequence prediction using neural networks" published and presented at the 1994 Artificial Neural Networks in Engineering conference, St. Louis, Missouri. Shawn received his Bachelor of Science in Electrical Engineering in May 1994.

In May 1994, Shawn joined Booz·Allen & Hamilton, Inc. in Linthicum, Maryland as a Consultant. Shawn has participated in a variety of theoretical and applied research and development efforts involving digital and analog communications, signal classification, signal detection, real-time signal processing, and object-oriented design. While working full-time, Shawn continued to pursue his graduate studies at Virginia Tech by taking classes from the University of Maryland, Johns Hopkins University, and George Mason University. In August 1994, Shawn began research under the guidance of Dr. Lamine Mili on robust statistical theory and nonlinear model estimation. Shawn completed his thesis in August 1996 by making maximum use of facsimile, email, telephone, airport lobbies, and the flexibility of his advisor. Shawn is currently a Senior Consultant at Booz·Allen & Hamilton, Inc.