THE UNIVERSITY OF BRITISH COLUMBIA
DEPARTMENT OF STATISTICS
TECHNICAL REPORT # 218

ROBUST ESTIMATION OF ERROR SCALE IN NONPARAMETRIC REGRESSION MODELS

BY

ISABELLA GHEMENT, MARCELO RUIZ AND RUBEN ZAMAR

MARCH 2006

ROBUST ESTIMATION OF ERROR SCALE IN NONPARAMETRIC REGRESSION MODELS

ISABELLA GHEMENT*, MARCELO RUIZ† AND RUBEN ZAMAR‡

ABSTRACT

When the data used to fit a nonparametric regression model are contaminated with outliers, we need to use a robust estimator of scale in order to make robust estimation of the regression function possible. We develop a family of M-estimators of scale constructed from consecutive differences of regression responses. Estimators in our family robustify the estimator proposed by Rice (1984). Under appropriate conditions, we establish the weak consistency and asymptotic normality of all estimators in our family. Estimators in our family vary in terms of their robustness properties. We quantify the robustness of each estimator via a quantity called maxbias. We use the maxbias as a basis for deriving the breakdown point of the estimator. Our theoretical results allow us to specify conditions for estimators in our family to achieve maximum breakdown point of 1/2. We conduct a simulation study to compare the finite sample performance of our preferred M-estimator with that of three other estimators.

KEYWORDS AND PHRASES: Breakdown point, consecutive differences, error scale, fixed design, maxbias, M-scale estimator, M-scale functional, nonparametric regression, outliers, robust.

*Department of Statistics, University of British Columbia, Canada.
†Departamento de Matemática, Facultad de Ciencias Exactas, Universidad Nacional de Río IV, Río IV, Argentina.
‡Department of Statistics, University of British Columbia,Canada.

1. Introduction

Robust estimators of error scale are widely used in nonparametric regression with outliers. In particular, such estimators are needed to compute robust M-estimators of the regression curve (see Härdle and Gasser, 1984, Härdle and Tsybakov, 1988, Boente and Fraiman, 1989, and references therein). Other applications include outliers detection (see Hannig and Lee, 2006), robust bandwidth selection for accurate estimation of the regression curve (see, for example, Boente, Fraiman and Meloche, 1997, Leung, Marriott and Wu, 1993, Cantoni and Ronchetti, 2001, and Leung, 2005) and robust inference about the regression curve.

Several authors considered the problem of error scale estimation in the context of outlier-free nonparametric regression. Dette, Munk and Wagner (1998) give an exhaustive discussion of the various estimators of error scale available in the literature and note that the most popular estimators are based on differences of regression responses. Since they do not rely on preliminary estimation of the regression curve itself, these difference-based estimators have fast $\sqrt{n}$-convergence rate and are computationally convenient. However, such estimators do not perform well in the presence of outliers.

In this paper, we introduce a family of robust M-estimators of error scale constructed from consecutive differences of regression responses. Our family includes the estimator proposed by Rice (1984) and its robustified version proposed by Boente, Fraiman and Meloche (1997) as particular cases. Under appropriate regularity conditions, we establish the weak consistency and asymptotic normality of all estimators in our family. These estimators differ in terms of their robustness properties. We quantify the robustness of each estimator via its maxbias. We rely on maxbias to derive the breakdown point of each estimator. As far as we are aware, all the proposed estimators of error scale based on differences fail to achieve maximum breakdown point of 1/2. Using our theoretical results we are able to specify conditions for estimators in our family to achieve a breakdown point of 1/2.

The rest of the paper is organized as follows. In Section 2, we introduce the nonparametric regression model of interest in this paper and identify the relevant scale parameter for this model. In Section 3, we define a family of M-estimators for this scale parameter. In Section 4, we introduce the family of M-scale functionals associated with the family of M-estimators. In Section 5, we show that each M-estimator in our family is weakly consistent to its corresponding M-functional and has an asymptotically normal distribution. In Section 6, we investigate the robustness properties of the estimators in our family. In Section 7, we conduct a simulation study to compare the finite sample performance of our preferred estimator against three competing estimators, including the estimators of Rice (1984) and Boente, Fraiman and Meloche (1997). Section 8 provides some concluding remarks. All the proofs are collected in the Appendix.

2. Nonparametric Regression Model

The nonparametric regression model of interest in this paper can be expressed as

$$Y_i = g(x_i) + U_i,\ i = 1, \ldots, n, \tag{1}$$

where the Y_i's are observed responses, the x_i's are fixed design points, $g(\cdot)$ is an unknown, smooth regression curve and the U_i's are independent, identically distributed unobservable random errors. We assume that the majority of the observations Y_i in model (1) is of good quality and has constant variability about the regression curve $g(\cdot)$, but a fraction ϵ is possibly of bad quality. We formalize this assumption below.

Let G denote the distribution function of the U_i's. Then G belongs to the ϵ-contaminated neighbourghood:

$$\mathcal{F}_\epsilon = \{G \in \mathcal{D} :\ G(y) = (1-\epsilon)F(y) + \epsilon H(y)\}. \tag{2}$$

Here, $\mathcal{D}$ denotes the set of all distribution functions, F belongs to the scale family associated with some fixed distribution function F_0, that is, $F(y) = F_0(y/\sigma)$ for an unspecified scale parameter $\sigma > 0$, H is an arbitrary distribution function in $\mathcal{D}$ and $\epsilon \in [0, 1/2]$ denotes the amount of contamination. Throughout, we assume that F_0 admits a symmetric, strictly positive and unimodal density f_0 and H is absolutely continuous.

Our interest is in robustly estimating the scale parameter σ from the sequence of consecutive differences $Y_{i+1} - Y_i, i = 1, \ldots, n-1$. Note that σ is not only unambiguously defined, but also fixed across the neighbourhood $\mathcal{F}_\epsilon$.

3. M-Estimator of Error Scale

To estimate the scale parameter σ of the central distribution F in (2), we propose using a regression-free estimator, constructed from the sequence of consecutive differences $Y_{i+1} - Y_i, i = 1, \dots, n-1$. This estimator, referred to as an *M-scale estimator*, is defined as

$$\widehat{\sigma}_n = inf\left\{ s > 0 : \frac{1}{n-1}\sum_{i=1}^{n-1} \chi\left(\frac{Y_{i+1} - Y_i}{as}\right) \leq b \right\}. \tag{3}$$

The score function $\chi : \mathcal{R} \to [0, \infty)$ must be chosen by the user and the constants $b \in (0, 1)$ and $a \in (0, \infty)$ are tuning constants that satisfy

$$E[\chi(Z_1)] = b \tag{4}$$

and

$$E\left[\chi\left(\frac{Z_2 - Z_1}{a}\right)\right] = b, \tag{5}$$

where Z_1, Z_2 are independent random variables with common distribution F_0.

The infimum in (3) is needed to handle situations where the score function χ is discontinuous. If χ is continuous, then $\widehat{\sigma}_n$ satisfies:

$$\frac{1}{n-1}\sum_{i=1}^{n-1} \chi\left(\frac{Y_{i+1} - Y_i}{a\widehat{\sigma}_n}\right) = b. \tag{6}$$

If, in addition, χ is strictly increasing on $\{x : \chi(x) < \underset{x}{sup}\, \chi(x)\}$, then $\widehat{\sigma}_n$ is uniquely defined by (6).

A property that is natural to expect from an estimator of scale is that of scale equivariance. It is easy to see that the M-estimator $\widehat{\sigma}_n = \widehat{\sigma}_n(Y_1, \dots, Y_n)$ is scale equivariant, in the sense that, for any $c \in R$, it satisfies

$$\widehat{\sigma}_n(cY_1, \dots, cY_n) = |c|\, \widehat{\sigma}_n(Y_1, \dots, Y_n).$$

Note that $\widehat{\sigma}_n$ is a generic member of a family of M-estimators, whose particular members correspond to different choices of the score function χ and the tuning constants a and b. In this paper, we show that the choice of χ is not very crucial to ensuring that $\widehat{\sigma}_n$ achieves the desired robustness properties, as long as χ is bounded and smooth, but the choice of b is (see Section 6). Given b, a is chosen so that $\widehat{\sigma}_n$ is Fisher-consistent when there is no contamination in the data (see Section 5).

The examples below illustrate various choices of χ, b and a for the case when $F_0 = \Phi$, where Φ is the standard normal distribution function.

EXAMPLE 1. *Choosing $\chi(x) = x^2$, $b = 1$ and $a = \sqrt{2}$ in (3) yields the (non-robust) estimator of scale proposed by Rice (1984):*

$$\widehat{\sigma}_n^{(1)} = \sqrt{\frac{1}{2(n-1)} \sum_{i=1}^{n-1} (Y_{i+1} - Y_i)^2}.$$

EXAMPLE 2. *Choosing $\chi(x) = I(|x| > \Phi^{-1}(3/4))$, $b = 1/2$ and $a = \sqrt{2}$ in (3) yields the (robust) estimator of scale proposed by Boente, Fraiman and Meloche (1997):*

$$\widehat{\sigma}_n^{(2)} = \frac{Q(0.50)}{\sqrt{2}\Phi^{-1}(3/4)},$$

where $Q(0.50)$ is the 50^{th} quantile of the absolute differences $|Y_{i+1} - Y_i|, i = 1, \ldots, n$.

EXAMPLE 3. *Choosing $\chi(x) = I(|x| > \Phi^{-1}(5/8))$, $b = 3/4$ and $a = \sqrt{2}$ in (3) yields the (robust) estimator of scale*

$$\widehat{\sigma}_n^{(3)} = \frac{Q(0.25)}{\sqrt{2}\Phi^{-1}(5/8)},$$

where $Q(0.25)$ is the 25^{th} quantile of the absolute differences $|Y_{i+1} - Y_i|, i = 1, \ldots, n$. This estimator is a modification of the estimator in Example 2.

EXAMPLE 4. *For $c > 0$ fixed, let*

$$\chi_c(x) = \begin{cases} 3\,(x/c)^2 - 3\,(x/c)^4 + (x/c)^6 & \text{if} \quad |x| \leq c \\ 1 & \text{if} \quad |x| > c \end{cases} \tag{7}$$

be the score function introduced by Beaton and Tukey (1974). Choosing $\chi(x) = \chi_c(x)$, with $c = 0.70417$, $b = 3/4$ and $a = \sqrt{2}$ in (3) yields the (robust) estimator of scale:

$$\widehat{\sigma}_n^{(4)} = inf\left\{ s > 0 : \frac{1}{n-1} \sum_{i=1}^{n-1} \chi_c\left(\frac{Y_{i+1} - Y_i}{\sqrt{2}s}\right) \leq \frac{3}{4} \right\}.$$

4. M-Scale Functional

The *M-scale functional* $\sigma(G)$ corresponding to $\widehat{\sigma}_n$ is defined as

$$\sigma(G) = inf \left\{ s > 0 : \ E\left[\chi\left(\frac{U_2 - U_1}{as}\right)\right] \leq b \right\}, \tag{8}$$

where we recall that U_1 and U_2 are independent error terms with common distribution G.

As we shall see in Section 5, $\sigma(G)$ is the asymptotic value of $\widehat{\sigma}_n$ since, under suitable assumptions, $\widehat{\sigma}_n$ converges in probability to $\sigma(G)$ as $n \to \infty$. Note that $\sigma(G)$ is scale equivariant - just like $\widehat{\sigma}_n$. Also, note that, if χ is continuous, $\sigma(G)$ satisfies

$$E\left[\chi\left(\frac{U_2 - U_1}{a\sigma(G)}\right)\right] = b. \tag{9}$$

If, in addition, χ is strictly increasing on $\{x : \ \chi(x) < \underset{x}{sup}\ \chi(x)\}$, then $\sigma(G)$ is uniquely defined by (9).

5. Asymptotics

In this section, we investigate the asymptotic behavior of the M-scale estimator $\widehat{\sigma}_n$ under the following assumptions:

(A1) The regression curve $g : [0,1] \to R$ is Lipschitz continuous, that is, there exists a constant $\mathcal{C}_g > 0$ such that $|g(x) - g(y)| \leq \mathcal{C}_g |x - y|$ for any $x, y \in [0,1]$.

(A2) The fixed design points $x_i, i = 1, \ldots, n$, satisfy the conditions $0 \leq x_1 \leq \cdots \leq x_n \leq 1$ and $\underset{1 \leq i \leq n-1}{max} \{|x_{i+1} - x_i|\} = \mathcal{O}(n^{-1})$.

(A3) The score function χ is such that $\chi(u) = 1$ for $|u| \geq c$ and $\chi(u) < 1$ for $|u| < c$ for some user-chosen constant $0 < c < \infty$. Furthermore, χ is even, satisfies $\chi(0) = 0$, is strictly increasing on $(0, c)$ and is twice continuously differentiable.

Remark 1. *A wide class of continuous score functions χ proposed in the robustness literature satisfy assumption* (A3). *In particular, the score function in* (7) *satisfies this assumption.*

Remark 2. *By assumption* (A3), *we have that $\chi' \equiv 0 \equiv \chi''$ outside the interval $[-c, c]$.*

The following theorem shows that the M-estimator $\widehat{\sigma}_n$ is weakly consistent for the M-functional $\sigma(G)$.

THEOREM 1. *Suppose that assumptions* (A1) - (A3) *hold. Then, as* $n \to \infty$,

$$\widehat{\sigma}_n \xrightarrow{P} \sigma(G), \tag{10}$$

with $\xrightarrow{P}$ *denoting convergence in probability.*

The above result suggests that $\widehat{\sigma}_n$ is asymptotically biased when the data are contaminated, as it converges in probability to $\sigma(G)$ instead of σ as $n \to \infty$. When there is no contamination in the data, that is, when $\epsilon = 0$, the distribution function of the errors in model (1) is F. In this case, we would like to be able to estimate σ, the scale parameter of F, without bias. This leads to the notion of Fisher-consistency. We say that $\sigma(G)$ is Fisher-consistent for $G = F$ if $\sigma(F) = \sigma$. It is easy to see that the choices of b and a given in (4) and (5), respectively, ensure the Fisher-consistency of $\sigma(G)$ for $G = F$.

Having shown the weak consistency of the M-estimator $\widehat{\sigma}_n$, we turn our attention to deriving its asymptotic distribution. The next theorem establishes that $\widehat{\sigma}_n$ has an asymptotically normal distribution.

THEOREM 2. *Suppose that assumptions* (A1) - (A3) *hold. Set*

$$V_1(G) = Var\left[\chi\left(\frac{U_2 - U_1}{a\sigma(G)}\right)\right] \tag{11}$$

$$V_2(G) = 2Cov\left[\chi\left(\frac{U_2 - U_1}{a\sigma(G)}\right), \chi\left(\frac{U_3 - U_2}{a\sigma(G)}\right)\right] \tag{12}$$

$$V_3(G) = E\left[\chi'\left(\frac{U_2 - U_1}{a\sigma(G)}\right)\left(\frac{U_2 - U_1}{a\sigma(G)^2}\right)\right] \tag{13}$$

and let $V(G) = (V_1(G) + V_2(G))/V_3^2(G)$. *Then, as* $n \to \infty$, *we have*

$$\sqrt{n}\left(\widehat{\sigma}_n - \sigma(G)\right) \xrightarrow{d} N(0, V(G)), \tag{14}$$

with $\xrightarrow{d}$ *denoting convergence in distribution.*

6. Robustness Properties

In this section, we introduce the maximum generalized asymptotic bias (or maxbias) of $\widehat{\sigma}_n$ as the most complete and accurate measure for assessing the robustness of $\widehat{\sigma}_n$. We then use this measure as a basis for our breakdown point considerations regarding $\widehat{\sigma}_n$.

6.1. Generalized Asymptotic Bias

We have established in Section 5 that the M-estimator $\widehat{\sigma}_n$ converges in probability to the M-scale functional $\sigma(G)$ as $n \to \infty$. If $G = F$, then $\sigma(G) = \sigma$ provided (4) and (5) hold. However, in general, if $G \neq F$, then $\sigma(G) \neq \sigma$. In other words, $\widehat{\sigma}_n$ is generally asymptotically biased for $G \in \mathcal{F}_\epsilon$.

The *raw asymptotic bias* of $\widehat{\sigma}_n$ quantifies the distance between $\sigma(G)$, the asymptotic value of $\widehat{\sigma}_n$, and σ, the scale parameter of interest, and is defined as:

$$B_r(\sigma(G)) = \frac{\sigma(G)}{\sigma} - 1. \tag{15}$$

If G is an outliers generating distribution, the raw asymptotic bias is likely positive. If G is an inliers generating distribution, the raw asymptotic bias is likely negative.

A more useful measure for assessing the asymptotic bias of $\widehat{\sigma}_n$ is the *generalized asymptotic bias* of this estimator, defined as

$$B_g(\sigma(G)) = \begin{cases} L_1\left(\frac{\sigma(G)}{\sigma}\right), & \text{if } 0 < \sigma(G) \leq \sigma, \\ L_2\left(\frac{\sigma(G)}{\sigma}\right), & \text{if } \ \sigma < \sigma(G) < \infty. \end{cases}$$

The functions L_1 and L_2 allow the user to penalize under-estimation and over-estimation of σ in different ways. Both functions are assumed to be non-negative, continuous, monotone and to satisfy the conditions:

$$L_1(1) = L_2(1) = 0 \quad \text{and} \quad \lim_{s \searrow 0} L_1(s) = \lim_{s \to \infty} L_2(s) = \infty.$$

A robust estimator $\widehat{\sigma}_n$ can be expected to have a relatively small and stable generalized asymptotic bias $B_g(\sigma(G))$ as G ranges over $\mathcal{F}_\epsilon$. The overall bias performance of $\widehat{\sigma}_n$ on the neighbourhood $\mathcal{F}_\epsilon$ can thus be measured by the maximum generalized asymptotic bias (maxbias):

$$\overline{B}_g(\epsilon) = \sup_{G \in \mathcal{F}_\epsilon} B_g(\sigma(G)). \tag{16}$$

Note that $\overline{B}_g(\epsilon)$ is scale invariant since the M-scale functional $\sigma(G)$ is scale equivariant. Also, note that the maxbias is a function that depends on ϵ, the fraction of contamination in the data. The maxbias curve, obtained by plotting $\overline{B}_g(\epsilon)$ versus ϵ, can be used to visually assess the robustness properties of $\widehat{\sigma}_n$. We consider $\widehat{\sigma}_n$ to be robust if $\overline{B}_g(\epsilon) < \infty$ for some $\epsilon \in (0, 1/2]$.

To derive an explicit expression for $\overline{B}_g(\epsilon)$, let

$$S^+(\epsilon) = \sup_{G \in \mathcal{F}_\epsilon} \sigma(G) \tag{17}$$

and

$$S^-(\epsilon) = \inf_{G \in \mathcal{F}_\epsilon} \sigma(G) \tag{18}$$

be the maximum and minimum values of the M-scale functional $\sigma(G)$ over $\mathcal{F}_\epsilon$. Then, using the monotonicity of L_1 and L_2, $\overline{B}_g(\epsilon)$ can be expressed as:

$$\overline{B}_g(\epsilon) = \max\left\{ L_1\left(\frac{S^-(\epsilon)}{\sigma}\right), L_2\left(\frac{S^+(\epsilon)}{\sigma}\right) \right\}. \tag{19}$$

Figure 1 displays a plot of the functions $L_1(s) = -\ln(s), s \in (0, 1]$ and $L_2(s) = \ln(s), s \in [1, \infty)$. For the situation depicted in this figure, $\overline{B}_g(\epsilon) = -\ln(S^-(\epsilon)/\sigma)$.

6.2. Breakdown Point Considerations

If the amount of contamination in the data is too large, $\widehat{\sigma}_n$ can suffer two types of breakdown: it can either explode, in the sense of taking on arbitrarily large aberrant values, or implode, in the sense of taking on arbitrarily small aberrant values.

The *asymptotic explosion breakdown point* of $\widehat{\sigma}_n$ is defined as

$$\epsilon^\infty = \inf\{\epsilon \in (0, 1/2] : S^+(\epsilon) = \infty\}, \tag{20}$$

whereas its *asymptotic implosion breakdown point* is defined as

$$\epsilon^0 = \inf\{\epsilon \in (0, 1/2] : S^-(\epsilon) = 0\}.$$

The *overall asymptotic breakdown point* of $\widehat{\sigma}_n$ is defined as the minimum of the asymptotic implosion and explosion breakdown points

$$\epsilon^* = \min\{\epsilon^0, \epsilon^\infty\}. \tag{21}$$

Clearly, if the amount of contamination in the data exceeds the overall asymptotic breakdown point of $\widehat{\sigma}_n$, then $\widehat{\sigma}_n$ ceases to provide a useful summary for the scale of the uncontaminated errors. Note that

$$\epsilon^* = \inf\left\{\epsilon \in (0, 1/2] : \quad \overline{B}_g(\epsilon) = \infty\right\} \tag{22}$$

since $\overline{B}_g(\epsilon) = \infty$ if and only if $S^-(\epsilon) = 0$ or $S^+(\epsilon) = \infty$.

The overall asymptotic breakdown point of $\widehat{\sigma}_n$ depends on the value of the tuning constant b in (3). What is the maximum overall asymptotic breakdown point that can be achieved by $\widehat{\sigma}_n$ as the value of b varies? Based on (22), to answer this question we must first derive an explicit expression for $\overline{B}_g(\epsilon)$.

In view of (19), to obtain an explicit expression for $\overline{B}_g(\epsilon)$ it suffices to obtain explicit expressions for $S^+(\epsilon)$ and $S^-(\epsilon)$. Such expressions are provided in Propositions 1 and 2 below, whose proofs can be found in the Appendix. For Propositions 1 and 2 and the subsequent results in this section, we assume without loss of generality that $\sigma = 1$.

Proposition 1. *Let $S^+(\epsilon)$ be as in* (17), *with $\epsilon \in (0, 1/2]$ fixed. Then, provided assumption* (A3) *holds, we have:*

$$S^+(\epsilon) = \begin{cases} s^+(\epsilon) & \text{if } \epsilon(2-\epsilon) < b \\ \infty & \text{if } \epsilon(2-\epsilon) \geq b \end{cases}$$

where $s^+(\epsilon)$ is implicitly defined by

$$\lambda_+(s^+(\epsilon)) = 0. \tag{23}$$

Here,

$$\lambda_+(s) = (1-\epsilon)^2 E\left[\chi\left(\frac{Z_2 - Z_1}{as}\right)\right] + \epsilon(2-\epsilon) - b, \tag{24}$$

and Z_1, Z_2 are independent random variables with common distribution F_0.

REMARK 3. *By (iii) of Lemma 6 in the Appendix, the equation $\lambda_+(s) = 0$ admits a unique, strictly positive solution for those $\epsilon \in (0, 1/2]$ with $\epsilon(2-\epsilon) < b$. Therefore, the quantity $s^+(\epsilon)$ satisfying (23) exists and is uniquely defined.*

PROPOSITION 2. *Let $S^-(\epsilon)$ be as in (18), with $\epsilon \in (0, 1/2]$ fixed. If assumption (A3) holds, then*

$$S^-(\epsilon) = \begin{cases} s^-(\epsilon) & \text{if } 1-\epsilon^2 > b \\ 0 & \text{if } 1-\epsilon^2 \le b \end{cases}$$

where $s^-(\epsilon)$ is implicitly defined by

$$\lambda_-(s^-(\epsilon)) = 0. \tag{25}$$

Here,

$$\lambda_-(s) = (1-\epsilon)^2 E\left[\chi\left(\frac{Z_2 - Z_1}{as}\right)\right] + 2\epsilon(1-\epsilon)E\left[\chi\left(\frac{Z_1}{as}\right)\right] - b, \tag{26}$$

and Z_1, Z_2 are as in Proposition 1.

REMARK 4. *By (iii) of Lemma 7, the equation $\lambda_-(s) = 0$ admits a unique, strictly positive solution for those $\epsilon \in (0, 1/2]$ for which $1-\epsilon^2 > b$, so the quantity $s^-(\epsilon)$ satisfying (25) exists and is uniquely defined.*

The next theorem provides an explicit expression for $\overline{B}_g(\epsilon)$, the maxbias of $\widehat{\sigma}_n$ over $\mathcal{F}_\epsilon$. This theorem is proven in the Appendix.

THEOREM 3. *Suppose the notation and assumptions in Propositions 1 and 2 hold. For* $\epsilon \in (0, 1/2]$, *let* $\overline{B}_g(\epsilon)$ *be as in* (16). *Also, let* $b \in (0, 1)$ *be the tuning constant in* (3). *The following facts hold.*

(i) If $b = 3/4$, *then*

$$\overline{B}_g(\epsilon) = \begin{cases} max\{L_2(s^+(\epsilon)), L_1(s^-(\epsilon))\} & \text{if} \quad \epsilon < 1/2, \\ \infty & \text{if} \quad \epsilon = 1/2. \end{cases}$$

(ii) If $b \in (0, 3/4)$, *then*

$$\overline{B}_g(\epsilon) = \begin{cases} max\{L_2(s^+(\epsilon)), L_1(s^-(\epsilon))\} & \text{if} \quad \epsilon < 1 - \sqrt{1-b}, \\ \infty & \text{if} \quad 1 - \sqrt{1-b} \leq \epsilon. \end{cases}$$

(iii) If $b \in (3/4, 1)$, *then*

$$\overline{B}_g(\epsilon) = \begin{cases} max\{L_2(s^+(\epsilon)), L_1(s^-(\epsilon))\} & \text{if} \quad \epsilon < \sqrt{1-b}, \\ \infty & \text{if} \quad \sqrt{1-b} \leq \epsilon. \end{cases}$$

As an immediate consequence of the above theorem, we derive an explicit expression for the overall asymptotic breakdown point of $\widehat{\sigma}_n$ as a function of b:

THEOREM 4. *Let* ϵ^* *be the overall asymptotic breakdown point of* $\widehat{\sigma}_n$ *defined by* (22). *Also, let* $b \in (0, 1)$ *be the tuning constant in* (3).

(i) If $b = 3/4$, *then* $\epsilon^* = 1/2$.

(ii) If $b \in (0, 3/4)$, *then* $\epsilon^* = 1 - \sqrt{1-b}$.

(iii) If $b \in (3/4, 1)$, *then* $\epsilon^* = \sqrt{1-b}$.

COROLLARY 1. *The maximum overall asymptotic breakdown point that can be achieved by* $\widehat{\sigma}_n$ *as the value of the tuning constant* b *in* (3) *varies in the interval* $(0, 1)$ *is* $\epsilon^*_{opt} = 1/2$*; this optimal breakdown point is attained for* $b = 3/4$.

So far, we have investigated the robustness properties of the M-estimator $\widehat{\sigma}_n$ for a general score function χ satisfying assumption (A3). In practice, χ must be specified by the user. One particular choice of χ that we recommend and that satisfies assumption (A3) is the score function χ_c in (7). The tuning constant $c > 0$ should be chosen to ensure that: (i) $\widehat{\sigma}_n$ achieves the optimal overall asymptotic breakdown point and (ii) $\widehat{\sigma}_n$'s limiting value, $\sigma(G)$, is Fisher-consistent when $G = F$, that is, when there is no contamination in the data. In what follows, we explain how to choose c for the case when F_0 is the standard normal distribution function Φ.

Recall from Corollary 1 that we should choose $b = 3/4$ to ensure that $\widehat{\sigma}_n$ achieves the optimal breakdown point of 1/2. According to the Fisher consistency considerations in Section 4, to ensure that $\sigma(G)$ is Fisher-consistent when $G = F$, we must choose c so that

$$E\left[\chi_c(Z_1)\right] = b.$$

Here, Z_1 is a standard normal random variable. One can easily see that $c = 0.70417$ satisfies the above equality.

7. Simulations

In this section, we report the results of a Monte Carlo simulation study on the finite sample properties of the estimators $\widehat{\sigma}_n^{(1)}$, $\widehat{\sigma}_n^{(2)}$, $\widehat{\sigma}_n^{(3)}$ and $\widehat{\sigma}_n^{(4)}$ introduced in Examples 1 to 4. The main goals of the study are to: (i) investigate the efficiency properties of $\widehat{\sigma}_n^{(2)}$, $\widehat{\sigma}_n^{(3)}$ and $\widehat{\sigma}_n^{(4)}$ relative to $\widehat{\sigma}_n^{(1)}$ in the absence of outlier contamination and (ii) compare the mean squared error performance of the four estimators in the presence of outlier contamination.

For our simulation study, we generate data from model (1) as follows. We take $n = 20$, 50 and 100. We consider $g(x) = sin(4\pi x)$. We take $x_i = (i-1)/(n-1), i = 1, \ldots, n$. We assume the U_i's to be independent with common distribution $G = (1-\epsilon)F + \epsilon H$, where $F(\cdot) = \Phi(\cdot/\sigma)$ and $\sigma = 1$. We allow ϵ to take the values $0, 0.05, 0.10, 0.20, 0.30$ and 0.40. Further, we use $H(y) = \Phi(y/10)$ to model symmetric outlier contamination and $H(y) = \Phi(y-10)$ to model asymmetric outlier contamination. For each model configuration, we generate $10,000$ data sets.

Figure 2 displays data generated for simulation settings with $n = 100$ and $H(y) = \Phi(y/10)$. Figure 3 provides the same display for simulation settings with $n = 100$ and $H(y) = \Phi(y-10)$. As expected, the two figures reveal that the larger the amount of contamination ϵ, the more outliers are present in the data. When the contamination is symmetric, the outliers tend to be located both below and above the true regression curve. However, when the contamination is asymmetric, the outliers are concentrated exclusively above the regression curve. We suspect that the estimation of σ will be more difficult when the contamination is asymmetric.

Before studying the finite sample properties of the estimators $\widehat{\sigma}_n^{(1)}$, $\widehat{\sigma}_n^{(2)}$, $\widehat{\sigma}_n^{(3)}$ and $\widehat{\sigma}_n^{(4)}$, we make some considerations regarding their overall asymptotic breakdown points. The overall asymptotic breakdown point of $\widehat{\sigma}_n^{(1)}$ is 0, as this estimator uses an unbounded score function. The overall asymptotic breakdown point of $\widehat{\sigma}_n^{(4)}$ was determined in Section 6 to be 1/2. We do not have theoretical results concerning the exact value of the optimal overall asymptotic breakdown point for $\widehat{\sigma}_n^{(2)}$ and $\widehat{\sigma}_n^{(3)}$. The reason for this is that, unlike $\widehat{\sigma}_n^{(4)}$, both of these estimators are computed with discontinuous score functions. Nevertheless, given that these score functions can be easily adjusted to become twice continuously differentiable, we expect the breakdown point considerations in Section 6 to hold, at least approximately, for $\widehat{\sigma}_n^{(2)}$ and $\widehat{\sigma}_n^{(3)}$. Therefore, we conjecture that $\widehat{\sigma}_n^{(2)}$'s overall asymptotic breakdown point is roughly 0.29 (use Theorem 4 with $b = 1/2$), while $\widehat{\sigma}_n^{(3)}$'s is roughly 1/2 (use Theorem 4 with $b = 3/4$). Our conjecture is supported by the simulation results reported in this section.

We now assess the efficiency of the robust estimators $\widehat{\sigma}_n^{(2)}$, $\widehat{\sigma}_n^{(3)}$ and $\widehat{\sigma}_n^{(4)}$ relative to the non-robust estimator $\widehat{\sigma}_n^{(1)}$ for those simulation settings with $\epsilon = 0$. For $j = 2, 3, 4$ fixed, we evaluate the efficiency of $\widehat{\sigma}_n^{(j)}$ relative to $\widehat{\sigma}_n^{(1)}$ by computing the ratio $RE(\widehat{\sigma}_n^{(j)}, \widehat{\sigma}_n^{(1)}) = \widehat{Var}(\widehat{\sigma}_n^{(j)})/\widehat{Var}(\widehat{\sigma}_n^{(1)})$, where

$$\widehat{Var}(\widehat{\sigma}_n^{(j)}) = \frac{1}{10,000} \sum_{i=1}^{10,000} \left(\widehat{\sigma}_{n,i}^{(j)} - \overline{\widehat{\sigma}_n}^{(j)} \right)^2 .$$

Here, $\widehat{\sigma}_{n,i}^{(j)}$ is the value of $\widehat{\sigma}_n^{(j)}$ corresponding to the ith sample generated from the model configuration of interest and $\overline{\widehat{\sigma}_n}^{(j)} = \sum_{i=1}^{10,000} \widehat{\sigma}_{n,i}^{(j)}/10,000$. Notice that both $\widehat{\sigma}_n^{(3)}$ and $\widehat{\sigma}_n^{(4)}$ have roughly the same overall asymptotic breakdown point, so comparing their relative efficiencies is appropriate. Comparing the relative efficiency of $\widehat{\sigma}_n^{(2)}$ against that of $\widehat{\sigma}_n^{(3)}$ and $\widehat{\sigma}_n^{(4)}$ may however not be appropriate as $\widehat{\sigma}_n^{(2)}$ has a much smaller overall asymptotic breakdown point than both $\widehat{\sigma}_n^{(3)}$ and $\widehat{\sigma}_n^{(4)}$.

Table 1 displays the values of $RE(\widehat{\sigma}_n^{(j)}, \widehat{\sigma}_n^{(1)}), j = 2, 3, 4$, for the simulation settings with $\epsilon = 0$. From this table, we see that $\widehat{\sigma}_n^{(2)}$ attains slightly better relative efficiency than $\widehat{\sigma}_n^{(3)}$ and $\widehat{\sigma}_n^{(4)}$ at the expense of robustness by achieving only 25% overall asymptotic breakdown point instead of 50%. However, $\widehat{\sigma}_n^{(3)}$ and $\widehat{\sigma}_n^{(4)}$ have much better robustness properties and not much worse relative efficiencies than $\widehat{\sigma}_n^{(2)}$, so we prefer them to $\widehat{\sigma}_n^{(2)}$.

Next, we compare the mean squared error performance of the estimators $\widehat{\sigma}_n^{(1)}$,$\widehat{\sigma}_n^{(2)}$, $\widehat{\sigma}_n^{(3)}$ and $\widehat{\sigma}_n^{(4)}$ under outlier contamination. For each simulation setting, we estimate the mean squared error of these estimators as:

$$\widehat{MSE}(\widehat{\sigma}_n^{(j)}) = \frac{1}{10,000} \sum_{i=1}^{10,000} \left(\widehat{\sigma}_{n,i}^{(j)} - \sigma \right)^2, \; j = 1, 2, 3, 4.$$

Estimators with small mean squared error are preferred.

Table 2 shows the estimated mean squared errors of $\widehat{\sigma}_n^{(1)}$, $\widehat{\sigma}_n^{(2)}$, $\widehat{\sigma}_n^{(3)}$ and $\widehat{\sigma}_n^{(4)}$ for the simulation settings with symmetric contamination. Table 3 displays similar quantities for the simulation settings with asymmetric contamination. Based on these tables, we conclude the following.

Regardless of the sample size or the amount of contamination, $\widehat{\sigma}_n^{(1)}$ has the poorest mean squared error performance amongst all estimators, both for symmetric and asymmetric contamination. In particular, note that the larger the amount of contamination in the data, the less accurate $\widehat{\sigma}_n^{(1)}$ becomes. This is not surprising, given that $\widehat{\sigma}_n^{(1)}$ is non-robust and therefore expected to break down in the presence of outliers.

For all sample sizes considered and for both types of contamination, the mean squared error performance of $\widehat{\sigma}_n^{(3)}$ and $\widehat{\sigma}_n^{(4)}$ is slightly worse than that of $\widehat{\sigma}_n^{(2)}$ when the amount of contamination is small, that is, when $\epsilon = 0.05$ or 0.10. However, as the amount of contamination becomes larger, the mean squared error performance of $\widehat{\sigma}_n^{(3)}$ and $\widehat{\sigma}_n^{(4)}$ becomes marginally better than that of $\widehat{\sigma}_n^{(2)}$ for symmetric contamination and significantly better than that of $\widehat{\sigma}_n^{(2)}$ for asymmetric contamination. This behaviour is in line with the fact that $\widehat{\sigma}_n^{(2)}$ has an (approximate) overall asymptotic breakdown point of 0.29; we would therefore expect $\widehat{\sigma}_n^{(2)}$ to perform poorly for amounts of contamination exceeding its breakdown point. In fact, one can see that the mean squared error performance of $\widehat{\sigma}_n^{(2)}$ shows signs of deterioration even when the amount of contamination is $\epsilon = 0.20$. On the other hand, both $\widehat{\sigma}_n^{(3)}$ and $\widehat{\sigma}_n^{(4)}$ have an overall asymptotic breakdown point of 1/2, so they are expected to perform reasonably well for amounts of contamination smaller than their breakdown point.

In summary, for practical use, we recommend using the estimators $\widehat{\sigma}_n^{(3)}$ and $\widehat{\sigma}_n^{(4)}$.

8. Concluding Remarks

In this paper, we introduced a family of robust M-estimators for estimating the error scale in nonparametric regression models with outliers. The estimators in our family are regression-free, being constructed from consecutive differences of regression responses. Under appropriate conditions, we established the weak consistency and asymptotic normality of all estimators in our family. To quantify the robustness of each M-estimator in the family in a complete and accurate way, we introduced a quantity called maxbias. We obtained explicit expressions for this maxbias as a function of the amount of contamination in the errors, and used these expressions to derive the breakdown point of the estimators in our family. Our theoretical results allowed us to specify conditions for estimators in our family to achieve maximum breakdown point of 1/2. We conducted a simulation study to investigate the finite sample performance of our preferred M-estimator. For the settings considered in this study, we found that this estimator outperformed the (non-robust) estimator introduced by Rice (1984) as well as the (robust) estimator proposed by Boente, Fraiman and Meloche (1997). We also found that, when modified to achieve an overall asymptotic breakdown point close to 1/2, the latter estimator performed almost as well as our preferred M-estimator.

Acknowledgements

This research was partially supported by the National Science and Engineering Research Council of Canada and the Secretaría de Ciencia y Tecnología de la Universidad Nacional de Río Cuarto, Argentina.

References

Beaton, A.E. and Tukey, J.W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics.* **16** 147–85.

Boente, G. and Fraiman, R. (1989). Robust Nonparametric Regression Estimation for Dependent Observations. **17** 1242–1256.

Boente, G., Fraiman, R. and Meloche, J. (1997). Robust plug-in bandwidth estimators in nonparametric regression. *Journal of Statistical Planning and Inference.* **57** 109–142.

Cantoni, E. and Ronchetti, E. (2001). Resistant selection of the smoothing parameter for smoothing splines. *Statistics and Computing.* **11** 141–146.

Chung, K.L. (1974). *A course in Probability Theory.* Academic Press.

Dette, H., Munk, A. and Wagner, T. (1998). Estimating the variance in nonparametric regression - What is a reasonable choice? *Journal of the Royal Statistics Society.* **60** 751–764.

Hannig, J. and Lee, T.C.M. (2006). Robust SiZer for Exploration of Regression Structures and Outlier Detection. *Journal of Computational and Graphical Statistics.* **15** 101–117.

Härdle, W. and Gasser, T. (1984). Robust nonparametric function fitting. *Journal of the Royal Statistical Society, Series B.* **46** 42–51.

Härdle, W. and Tsybakov, A. B. (1988). Robust Nonparametric Regression with Simultaneous Scale Curve Estimation. *The Annals of Statistics.* **16** 120–135.

Leung, D.H.-Y. (2005). Cross-validation in nonparametric regression with outliers. *Journal of Nonparametric Statistics.* **2** 333–339.

Leung, D.H.-Y., Marriott, F.H.C. and Wu, E.K.H. (1993). Bandwidth selection in robust smoothing. *Journal of Nonparametric Statistics.* **2** 333–339.

Salibian-Barrera, M. (2000). Contributions to the Theory of Robust Inference. *Ph.D. Thesis.* Department of Statistics, University of British Columbia, Vancouver, Canada.

Rice, J. (1984). Bandwidth Choice for Nonparametric Regression. *The Annals of Statistics.* **12** 1215–1230.

Appendix

This appendix collects the proofs of the theoretical results introduced in Sections 5 and 6. Throughout the appendix, we set

$$S_n(s) = \frac{1}{n-1}\sum_{i=1}^{n-1}\chi\left(\frac{Y_{i+1}-Y_i}{as}\right) - b$$

for $s > 0$ and express $\widehat{\sigma}_n$ as

$$\widehat{\sigma}_n = \inf\{s > 0 : \; S_n(s) \leq 0\}. \tag{27}$$

The next lemmas are used for proving Theorem 1.

Lemma 1. *Let $\{Z_i\}_{i\geq 1}$ be a sequence of m-dependent, identically distributed random variables, with $Var(Z_1) < \infty$. Then, as $n \to \infty$,*

$$\frac{\sum_{i=1}^n Z_i}{n} \xrightarrow{P} E(Z_1).$$

Proof: By Chebyshev's inequality, for any $\epsilon > 0$ we have:

$$P\left(\left\{\left|\frac{\sum_{i=1}^n Z_i}{n} - E(Z_1))\right| \geq \epsilon\right\}\right) \leq \frac{1}{\epsilon^2}Var\left(\frac{\sum_{i=1}^n Z_i}{n}\right).$$

Therefore, it suffices to show that $Var(\sum_{i=1}^n Z_i/n)$ converges to zero as $n \to \infty$. Using the m-dependence of the sequence $\{Z_i\}_{i\geq 1}$ together with the Cauchy-Schwartz inequality, we obtain:

$$\begin{aligned} Var\left(\frac{\sum_{i=1}^n Z_i}{n}\right) &= \frac{1}{n^2}\sum_{i=1}^n Var(Z_i) + \frac{2}{n^2}\sum_{i<j} Cov(Z_i, Z_j) = \frac{\sigma^2}{n} + \frac{2}{n^2}\sum_{i=1}^{n-m}\sum_{j=1}^{m} Cov(Z_i, Z_{i+j}) \\ &\leq \frac{\sigma^2}{n} + \frac{2(n-m)m\sigma^2}{n^2}. \end{aligned}$$

Clearly, $Var(\sum_{i=1}^n Z_i/n)$ converges to zero as $n \to \infty$.

Lemma 2. *Let $\tilde{G}$ be an arbitrary absolutely continuous distribution function with strictly positive density $\tilde{g}$. For $s > 0$, define*

$$\lambda_{\tilde{G}}(s) = E\left[\chi\left(\frac{\tilde{U}_2 - \tilde{U}_1}{as}\right)\right] - b,$$

where χ is a score function satisfying assumption (A3), a and b are tuning constants satisfying equations (4)-(5), and $\tilde{U}_1, \tilde{U}_2$ are independent random variables with common distribution $\tilde{G}$. Then the function $\lambda_{\tilde{G}}$ is continuous, strictly decreasing and admits the limits:

$$\underset{s\to\infty}{lim}\,\lambda_{\tilde{G}}(s) = -b \text{ and } \underset{s\searrow 0}{lim}\,\lambda_{\tilde{G}}(s) = 1 - b.$$

Proof: The continuity of $\lambda_{\tilde{G}}$ is an immediate consequence of the continuity of χ. To show $\lambda_{\tilde{G}}$ is strictly decreasing, we use reduction to the absurd. Specifically, we assume that there exist two positive real numbers $s_1 < s_2$ such that $\lambda_{\tilde{G}}(s_1) = \lambda_{\tilde{G}}(s_2)$. Then:

$$0 = \lambda_{\tilde{G}}(s_1) - \lambda_{\tilde{G}}(s_2) = \int_{-\infty}^{\infty}\left[\chi\left(\frac{v}{as_1}\right) - \chi\left(\frac{v}{as_2}\right)\right]\tilde{g}^*(v)dv, \tag{28}$$

where $\tilde{g}^*$ is the strictly positive density function of the random variable $V_2 - V_1$, given by

$$\tilde{g}^*(x) = \int_{-\infty}^{+\infty}\tilde{g}(x-y)\tilde{g}(-y)dy.$$

Since, by assumption (A3), $\chi(v/(as_1)) \geq \chi(v/(as_2))$ for any real number v, (28) implies that

$$\chi\left(\frac{v}{as_1}\right) - \chi\left(\frac{v}{as_2}\right) = 0 \; a.e. \; (v).$$

But this contradicts assumption (A3), which states that χ is strictly increasing on the set $\{v \geq 0 : \chi(v) < 1\}$. In conclusion, $\lambda_{\tilde{G}}$ is strictly increasing. The derivation of the two limits is straightforward.

Lemma 3. *Let U_1, U_2 be error terms in model* (1) *and let $G = (1-\epsilon)F + \epsilon H \in \mathcal{F}_\epsilon$ be their common distribution. For $s > 0$, define*

$$\lambda_G(s) = E\left[\chi\left(\frac{U_2 - U_1}{as}\right)\right] - b,$$

where χ is a score function satisfying assumption (A3), and a and b are tuning constants satisfying equations (4) - (5)*. Then:*

(i) The function λ_G is continuous, strictly decreasing and admits the limits $\lim_{s\to\infty} \lambda_G(s) = -b$ and $\lim_{s\searrow 0} \lambda_G(s) = 1 - b$.

(ii) The equation $\lambda_G(s) = 0$ admits a unique solution, namely the M-scale functional $\sigma(G)$ defined in (8)*.*

(iii) For any $s > 0$, $\lambda_G(s)$ can be decomposed as

$$\lambda_G(s) = (1-\epsilon)^2 E\left[\chi\left(\frac{V_2 - V_1}{as}\right)\right] + 2\epsilon(1-\epsilon) E\left[\chi\left(\frac{V_2 - W_1}{as}\right)\right] + \epsilon^2 E\left[\chi\left(\frac{W_2 - W_1}{as}\right)\right] - b,$$

where V_1, V_2 are independent random variables with common distribution F, W_1, W_2 are independent random variables with common distribution H, and (V_i, W_i), $i = 1, 2$, are independent.

Proof: The proof of (i) follows from Lemma 2 with $\tilde{G} = G$; the proof of (ii) follows immediately from (i).

For (iii), let B_1 and B_2 be independent, identically distributed random variables having a Bernoulli distribution with parameter ϵ. If B_i is independent of $\{V_i, W_i\}$, then we can write U_i as $U_i = (1 - B_i) V_i + B_i W_i$, where $i = 1, 2$. Therefore:

$$\begin{aligned} E\left[\chi\left(\frac{U_2 - U_1}{as}\right)\right] &= E\left[\chi\left(\frac{(1-B_2) V_2 + B_2 W_2 - (1-B_1) V_1 - B_1 W_1}{as}\right)\right] \\ &= P(B_1 = 0, B_2 = 0)\, E\left[\chi\left(\frac{V_2 - V_1}{as}\right)\right] + P(B_1 = 1, B_2 = 0)\, E\left[\chi\left(\frac{V_2 - W_1}{as}\right)\right] \\ &\quad + P(B_1 = 0, B_2 = 1)\, E\left[\chi\left(\frac{W_2 - V_1}{as}\right)\right] + P(B_1 = 1, B_2 = 1)\, E\left[\chi\left(\frac{W_2 - W_1}{as}\right)\right]. \end{aligned}$$

Result (iii) follows easily from the above by using the independence of B_1 and B_2 as well as the properties of the Bernoulli distribution.

Proof of Theorem 1:

To prove the theorem it suffices to show that, for any $\delta > 0$, $\lim_{n\to\infty} P\{\widehat{\sigma}_n \leq \sigma(G) + \delta\} = 1$ and $\lim_{n\to\infty} P\{\widehat{\sigma}_n < \sigma(G) - \delta\} = 0$. The former result is proven below. The latter result can be established by a similar argument.

Fix $\delta > 0$ and note that the inclusion $\{\widehat{\sigma}_n \leq \sigma(G) + \delta\} \supseteq \{S_n(\sigma(G) + \delta) \leq 0\}$ holds since, by (27), $\{\widehat{\sigma}_n \leq s\} \supseteq \{S_n(s) \leq 0\}$ for any $s > 0$. Therefore, it is enough to prove:

$$\lim_{n\to\infty} P\{S_n(\sigma(G) + \delta) \leq 0\} = 1.$$

We prove this by showing that $S_n(\sigma(G) + \delta)$ converges in probability to a strictly negative quantity.

Using the Mean Value Theorem, we have

$$\begin{aligned} S_n(\sigma(G) + \delta) = &\left[\frac{1}{n-1}\sum_{i=1}^{n-1} \chi\left(\frac{U_{i+1} - U_i}{a(\sigma(G) + \delta)}\right) - b\right] \\ &+ \frac{1}{n-1}\sum_{i=1}^{n-1} \chi'(W_i) \cdot (g(x_{i+1}) - g(x_i)) \cdot \frac{1}{a(\sigma(G) + \delta)} \end{aligned}$$

with W_i being an intermediate value between $(Y_{i+1} - Y_i)/[a(\sigma(G) + \delta)]$ and $(U_{i+1} - U_i)/[a(\sigma(G) + \delta)]$. The first term converges in probability to $\lambda_G(\sigma(G) + \delta)$ by Lemma 1. The second term converges in probability to zero as it is bounded by $O(1/n)$. Combining these results yields that $S_n(\sigma(G) + \delta)$ converges in probability to $\lambda_G(\sigma(G) + \delta)$. Clearly, $\lambda_G(\sigma(G) + \delta)$ is strictly negative as λ_G is strictly decreasing by (i) of Lemma 3, and satisfies $\lambda_G(\sigma(G)) = 0$ by (ii) of Lemma 3.

Lemmas 4 and 5 below are needed for proving Theorem 2. The proof of Lemma 4 can be found in Chung (1974, pp. 214–215).

Lemma 4. *Suppose $\{Z_i\}_{i\geq 1}$ is a sequence of m-dependent, uniformly bounded random variables such that*

$$\lim_{n\to\infty} \frac{\sqrt{Var(\sum_{i=1}^n Z_i)}}{n^{1/3}} = \infty.$$

Then, as $n \to \infty$, we have

$$\frac{\sum_{i=1}^n Z_i - E(\sum_{i=1}^n Z_i)}{\sqrt{Var(\sum_{i=1}^n Z_i)}} \xrightarrow{d} N(0, 1).$$

Lemma 5. *Let χ be a score function satisfying assumption (A3) and let $K = [s_1, s_2] \subset (0, \infty)$ be a compact interval. For y an arbitrary real number and $s > 0$, set $h(y,s) = \chi'(y/s)(y/s^2)$. Then, for each $s_0 \in K$, h is continuous in s_0 uniformly in y.*

Proof: The proof is similar to that of Lemma 7.7 in Salibian-Barrera (2000). First note that, as the support of χ' is the interval $[-c, c]$, then

$$h(y,s) = 0, \forall s \in [s_1, s_2], \forall y \notin T = [-cs_2, cs_2].$$

Thus, we need to consider only the points (y, s) with $y \in T$.

Let $s_o \in K$ and $\epsilon > 0$. We have

$$\begin{aligned} |h(y,s) - h(y,s_o)| &= \left|\chi'(y/s)y/s^2 - \chi'(y/s_o)y/s_o^2\right| \\ &\leq |\chi'(y/s)| \left|\frac{y}{s^2} - \frac{y}{s_o^2}\right| + \left|\frac{y}{s_o^2}\right| |\chi'(y/s) - \chi'(y/s_o)| \\ &\leq 2cc_{\chi'} \left[\left|\frac{1}{s^2} - \frac{1}{s_o^2}\right| + \frac{1}{s_1^2}|\chi'(y/s) - \chi'(y/s_o)|\right]. \end{aligned}$$

Using the uniform continuity of $u(s) = 1/s^2$ and χ' on the compact set K and the fact that $|y| \leq 2cs_2$ yields the desired result.

Proof of Theorem 2:

Using the Mean Value Theorem, together with the fact that $S_n(\widehat{\sigma}_n) = 0$ by equation (6), we obtain

$$\sqrt{n}(\widehat{\sigma}_n - \sigma(G)) = \frac{\sqrt{n}}{\sqrt{n-1}} \cdot \frac{\sqrt{n-1}S_n(\sigma(G))}{-S_n'(\tilde{\sigma}_n)},$$

with $\tilde{\sigma}_n$ being an intermediate point between $\widehat{\sigma}_n$ and $\sigma(G)$. The desired asymptotic normality result will follow from Slutsky's Theorem, provided

$$\sqrt{n-1}S_n(\sigma(G)) \xrightarrow{d} N(0, V_1(G) + V_2(G)) \tag{29}$$

and

$$-S_n'(\tilde{\sigma}_n) = \frac{1}{n-1}\sum_{i=1}^{n-1} \chi'\left(\frac{Y_{i+1} - Y_i}{a\tilde{\sigma}_n}\right)\left(\frac{Y_{i+1} - Y_i}{a\tilde{\sigma}_n^2}\right) \xrightarrow{P} V_3(G) \tag{30}$$

as $n \to \infty$.

To prove (29), set

$$T_n(G) = \sum_{i=1}^{n-1} \left[\chi\left(\frac{Y_{i+1} - Y_i}{a\sigma(G)}\right) - b \right] \equiv \sum_{i=1}^{n-1} Z_i$$

and write:

$$\sqrt{n-1} S_n(\sigma(G)) = \frac{\sqrt{Var(T_n(G))}}{\sqrt{n-1}} \cdot \frac{T_n(G) - E(T_n(G))}{\sqrt{Var(T_n(G))}} + \frac{E(T_n(G))}{\sqrt{n-1}}.$$

By Slutsky's Theorem, it suffices to show:

$$\lim_{n\to\infty} \frac{\sqrt{Var(T_n(G))}}{\sqrt{n-1}} = \sqrt{V_1(G) + V_2(G)}, \tag{31}$$

$$\frac{T_n(G) - E(T_n(G))}{\sqrt{Var(T_n(G))}} \xrightarrow{d} N(0,1) \tag{32}$$

and

$$\lim_{n\to\infty} \frac{E(T_n(G))}{\sqrt{n-1}} = 0. \tag{33}$$

Result (31) is proven below. Result (32) follows directly from Lemma 4. Clearly, the Z_i's are uniformly bounded and, if (31) holds,

$$\lim_{n\to\infty} \frac{\sqrt{Var(T_n(G))}}{n^{1/3}} = \lim_{n\to\infty} \frac{\sqrt{Var(T_n(G))}}{\sqrt{n-1}} \cdot \lim_{n\to\infty} \frac{\sqrt{n-1}}{n^{1/3}} = \infty.$$

Result (33) is straightforward.

Consider (31). By the definition of $T_n(G)$ and the one-dependence of $Y_{i+1} - Y_i$, $i = 1, \ldots, n-1$, we have:

$$Var(T_n(G)) = \sum_{i=1}^{n-1} Var\left[\chi\left(\frac{Y_{i+1} - Y_i}{a\sigma(G)}\right)\right] + 2\sum_{i=1}^{n-2} Cov\left[\chi\left(\frac{Y_{i+1} - Y_i}{a\sigma(G)}\right), \chi\left(\frac{Y_{i+2} - Y_{i+1}}{a\sigma(G)}\right)\right]. \tag{34}$$

Using a Mean Value Theorem argument, the first sum in (34) can be written as

$$\sum_{i=1}^{n-1} Var\left[\chi\left(\frac{Y_{i+1}-Y_i}{a\sigma(G)}\right)\right] = \sum_{i=1}^{n-1} Var\left[\chi\left(\frac{U_{i+1}-U_i}{a\sigma(G)}\right) + \frac{g(x_{i+1})-g(x_i)}{a\sigma(G)} \cdot \chi'(\xi_i)\right]$$
$$= \sum_{i=1}^{n-1} Var\left[\chi\left(\frac{U_{i+1}-U_i}{a\sigma(G)}\right)\right] + \sum_{i=1}^{n-1}\left[\frac{g(x_{i+1})-g(x_i)}{a\sigma(G)}\right]^2 \cdot Var\left[\chi'(\xi_i)\right]$$
$$+ 2Cov\left[\chi\left(\frac{U_{i+1}-U_i}{a\sigma(G)}\right), \frac{g(x_{i+1})-g(x_i)}{a\sigma(G)} \cdot \chi'(\xi_i)\right],$$

where ξ_i is an intermediate point between $U_{i+1} - U_i$ and $g(x_{i+1}) - g(x_i)$. The first term equals $(n-1)V_1(G)$ since $U_{i+1} - U_i, i = 1, \ldots, n-1$, are identically distributed random variables. The second term is bounded by $\mathcal{O}(1/n) = o(1)$. An application of the Cauchy-Schwartz inequality shows that the third term is bounded by $\mathcal{O}(1)$. Combining these results we obtain that the first sum in (34) is $(n-1)V_1(G)+\mathcal{O}(1)$. A similar argument yields that the second sum in (34) is $(n-2)V_2(G)+\mathcal{O}(1)$. Thus, $Var(T_n(G)) = (n-1)V_1(G) + (n-2)V_2(G) + \mathcal{O}(1)$, and result (31) is immediate.

To complete the proof of the theorem, we must show that result (30) holds. The left hand side of (30) can be written as

$$-S_n'(\tilde{\sigma}_n) = \frac{1}{n-1}\sum_{i=1}^{n-1}\left[h\left(\frac{Y_{i+1}-Y_i}{a}, \tilde{\sigma}_n\right) - h\left(\frac{Y_{i+1}-Y_i}{a}, \sigma(G)\right)\right]$$
$$+ \frac{1}{n-1}\sum_{i=1}^{n-1} h\left(\frac{Y_{i+1}-Y_i}{a}, \sigma(G)\right),$$

where $h(y, s) = \chi'(y/s)(y/s^2)$. The first term converges to zero in probability by Theorem 1 and Lemma 5. Using the Mean Value Theorem, the second term can be expressed as:

$$\frac{1}{n-1}\sum_{i=1}^{n-1} h\left(\frac{Y_{i+1}-Y_i}{a}, \sigma(G)\right) = \frac{1}{n-1}\sum_{i=1}^{n-1} h\left(\frac{U_{i+1}-U_i}{a}, \sigma(G)\right)$$
$$+ \frac{1}{n-1}\sum_{i=1}^{n-1}\left[\chi''(\xi_i)\cdot\xi_i + \chi'(\xi_i)\right]\left[\frac{g(x_{i+1})-g(x_i)}{a\sigma(G)^2}\right],$$

where ξ_i is an intermediate point between $(Y_{i+1} - Y_i)/(a\sigma(G))$ and $(U_{i+1} - U_i)/(a\sigma(G))$. By Lemma 1, the first term converges in probability to $V_3(G)$. The second term converges in probability to zero as it is bounded by $O(1/n)$. Combining these results yields (30).

Lemmas 6 and 7 below are needed for proving Propositions 1 and 2. The proof of Lemma 6 is given below. The proof of Lemma 7 is similar to that of Lemma 6, so we omit it.

LEMMA 6. *For $n \geq 1$ and $\epsilon \in (0, 1/2]$, let $G_n = (1-\epsilon)F_0 + \epsilon H_n$ be a contaminated distribution, where F_0 is the nominal distribution of the ϵ-contaminated neighborhood in* (2) *and $H_n(y) = \Phi(y/n)$. Moreover, for $s > 0$, set*

$$\lambda_{G_n}(s) = E\left[\chi\left(\frac{U_{2,n} - U_{1,n}}{as}\right)\right] - b,$$

where $U_{1,n}, U_{2,n}$ are independent random variables with common distribution G_n, χ is a score function satisfying assumption (A3) and a and b are tuning constants satisfying equations (4)-(5)*. Then the following facts hold.*

(i) For any $s > 0$, we have:

$$\lim_{n\to\infty} \lambda_{G_n}(s) = \lambda_+(s),$$

with $\lambda_+(s)$ as in (24)*.*

(ii) The function λ_+ is continuous, strictly decreasing and admits the limits:

$$\lim_{s\searrow 0} \lambda_+(s) = 1 - b \text{ and } \lim_{s\to\infty} \lambda_+(s) = \epsilon(2-\epsilon) - b.$$

(iii) If $\epsilon(2-\epsilon) < b$, the equation $\lambda_+(s) = 0$ has a unique finite, strictly positive solution.

PROOF: For (i), fix $s > 0$ and use (iii) of Lemma 3 with $G = G_n$ (hence $F = F_0$ and $H = H_n$) to write:

$$\lambda_{G_n}(s) = (1-\epsilon^2)E\left[\chi\left(\frac{Z_2 - Z_1}{as}\right)\right] + 2\epsilon(1-\epsilon)E\left[\chi\left(\frac{Z_2 - W_1^{(n)}}{as}\right)\right] + \epsilon^2 E\left[\left(\frac{W_2^{(n)} - W_1^{(n)}}{as}\right)\right] - b, \tag{35}$$

where Z_1, Z_2 are independent random variables with common distribution F_0, and $W_1^{(n)}$, $W_2^{(n)}$ are independent random variables with common distribution H_n. Furthermore, $(Z_i, W_i^{(n)}), i = 1, 2$, are independent.

To analyze the second term in (35), note that the density of $Z_2 - W_1^{(n)}$ is given by

$$g(x) = \int_{-\infty}^{\infty} \frac{1}{n}\phi\left(\frac{x-y}{n}\right) f_0(y)dy,$$

where $\phi = \Phi'$ and $f_0 = F_0'$. Therefore,

$$E\left[\chi\left(\frac{Z_2 - W_1^{(n)}}{as}\right)\right] = \int_{-\infty}^{\infty} \chi\left(\frac{x}{as}\right) g(x)dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \chi\left(\frac{x}{as}\right)\frac{1}{n}\phi\left(\frac{x-y}{n}\right) f_0(y)dydx$$

or, equivalently,

$$E\left[\chi\left(\frac{Z_2 - W_1^{(n)}}{as}\right)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \chi\left(\frac{z+nu}{as}\right) f_0(z)\phi(u)dzdu.$$

Taking limit as n goes to infinity in both sides of the above and using the Dominated Convergence Theorem together with the fact that $\chi(\infty) = 1$ (assumption (A3)) yields that the second term in (35) equals 1. A similar argument yields that the third term in (35) equals 1. In conclusion, (35) can be re-written as

$$\begin{aligned}\lambda_{G_n}(s) &= (1-\epsilon^2)E\left[\chi\left(\frac{Z_2 - Z_1}{as}\right)\right] + 2\epsilon(1-\epsilon) + \epsilon^2 - b \\ &= (1-\epsilon^2)E\left[\chi\left(\frac{Z_2 - Z_1}{as}\right)\right] + \epsilon(2-\epsilon) - b \equiv \lambda_+(s)\end{aligned}$$

by the definition of $\lambda_+(s)$ in (24). Thus, (i) holds.

For (ii) and (iii), write:

$$\begin{aligned}\lambda_+(s) &= (1-\epsilon^2)\left\{E\left[\chi\left(\frac{Z_2 - Z_1}{as}\right)\right] - b + b\right\} + \epsilon(2-\epsilon) - b \\ &\equiv (1-\epsilon)^2\left\{\lambda_{F_0}(s) + b\right\} + \epsilon(2-\epsilon) - b.\end{aligned}$$

and use Lemma 3 with $\tilde{G} = F_0$.

Lemma 7. *For $n \geq 1$ and $\epsilon \in (0, 1/2]$, let $G_n = (1-\epsilon)F_0 + \epsilon H_n$ be a contaminated distribution, where F_0 is the nominal distribution of the ϵ-contaminated neighborhood in* (2) *and $H_n(y) = \Phi(ny)$. Moreover, for $s > 0$, set*

$$\lambda_{G_n}(s) = E\left[\chi\left(\frac{U_{2,n} - U_{1,n}}{as}\right)\right] - b,$$

where $U_{1,n}, U_{2,n}$ are independent random variables with common distribution G_n, χ is a score function satisfying assumption (A3) and a and b are tuning constants satisfying equations (4)-(5)*. Then the following facts hold.*

(i) For any $s > 0$, we have:

$$\lim_{n\to\infty} \lambda_{G_n}(s) = \lambda_-(s),$$

with $\lambda_-(s)$ as in (26)*.*

(ii) The function λ_- is continuous, strictly decreasing and admits the limits:

$$\lim_{s\searrow 0} \lambda_-(s) = 1 - \epsilon^2 - b \text{ and } \lim_{s\to\infty} \lambda_-(s) = -b.$$

(iii) If $1 - \epsilon^2 > b$, the equation $\lambda_-(s) = 0$ has a unique finite, strictly positive solution.

Proof of Proposition 1:

Fix $\epsilon \in (0, 1/2]$ such that $\epsilon(2-\epsilon) < b$. By (17), to prove that $S^+(\epsilon) = s^+(\epsilon)$, it suffices to show that the following facts hold: (i) $\sigma(G) \leq s^+(\epsilon)$ for any $G \in \mathcal{F}_\epsilon$ and (ii) there exists a sequence of distributions $\{G_n\}_{n\geq 1} \subseteq \mathcal{F}_\epsilon$ such that $\lim_{n\to\infty} \sigma(G_n) = s^+(\epsilon)$.

For (i), fix $G \in \mathcal{F}_\epsilon$. If the inclusion

$$\{s > 0 : \ s > s^+(\epsilon)\} \subseteq \{s > 0 : \lambda_G(s) \leq 0\} \tag{36}$$

holds, then the proof of (i) follows by taking infimum in both sides of (36) and using the definition of $\sigma(G)$ in (8). To prove (36), take $s > s^+(\epsilon)$ and note that $\lambda_G(s) < \lambda_G(s^+(\epsilon))$ since λ_G is strictly decreasing by (*i*) of Lemma 3. Thus, it is enough to show $\lambda_G(s^+(\epsilon)) \leq 0$. Using (*iii*) of Lemma 3 with $s = s^+(\epsilon)$, we write:

$$\lambda_G(s^+(\epsilon)) = (1-\epsilon)^2 E\left[\chi\left(\frac{Z_2 - Z_1}{as^+(\epsilon)}\right)\right] + \epsilon(1-\epsilon)E\left[\chi\left(\frac{Z_2 - W_1}{as^+(\epsilon)}\right)\right]$$
$$+ \epsilon(1-\epsilon)E\left[\chi\left(\frac{W_2 - Z_1}{as^+(\epsilon)}\right)\right] + \epsilon^2 E\left[\chi\left(\frac{W_2 - W_1}{as^+(\epsilon)}\right)\right] - b,$$

where Z_1, Z_2 are independent random variables with common distribution F_0, W_1, W_2 are independent random variables with common distribution H and (Z_i, W_i), $i = 1, 2$, are independent. Using that $||\chi||_\infty = 1$ (assumption (A3)) together with equation (24), we get

$$\lambda_G(s^+(\epsilon)) \leq (1-\epsilon)^2 E\left[\chi\left(\frac{Z_2 - Z_1}{as^+(\epsilon)}\right)\right] + \epsilon(2-\epsilon) - b = \lambda_+(s^+(\epsilon)) = 0.$$

For (ii), define the sequence of distributions $\{G_n\}_{n\geq 1} \subseteq \mathcal{F}_\epsilon$ such that $G_n = (1-\epsilon)F_0 + \epsilon H_n$, with $H_n(y) = \Phi(y/n)$. Then proceed as follows.

Fix $0 < \delta < s^+(\epsilon)$. Set $d = s^+(\epsilon) - \delta$ and $\delta_1 = \lambda_+(d) - \lambda_+(s^+(\epsilon))$ and note that $\delta_1 > 0$ since, by *(ii)* of Lemma 6, λ_+ is strictly decreasing. Given that $\lim_{n\to\infty} \lambda_{G_n}(d) = \lambda_+(d)$ by *(i)* of Lemma 6 with $s = d$, there exists $N_0 \geq 1$ such that, for any $n \geq N_0$, $|\lambda_{G_n}(d) - \lambda_+(d)| < \delta_1$, hence $\lambda_{G_n}(d) > \lambda_+(d) - \delta_1 = \lambda_+(s^+(\epsilon)) = 0$. By Lemma 2 with $\tilde{G} = G_n$, the equation $\lambda_{G_n}(s) = 0$ admits a unique finite, strictly positive solution. If we denote this solution by $\sigma(G_n)$, then $\lambda_{G_n}(\sigma(G_n)) = 0$ and the above yields that $\lambda_{G_n}(d) > \lambda_{G_n}(\sigma(G_n))$ for any $n \geq N_0$. But λ_{G_n} is strictly decreasing by Lemma 2 with $\tilde{G} = G_n$, so $\sigma(G_n) > d$ or, equivalently, $\sigma(G_n) > s^+(\epsilon) - \delta$ for any $n \geq N_0$. Also, considering that for each $n \geq 1$, $\sigma(G_n) \leq s^+(\epsilon)$, we conclude that $|\sigma(G_n) - s^+(\epsilon)| < \delta$, for each $n \geq N_0$, and, as δ was chosen arbitrarily, $\underset{n\to\infty}{lim} \sigma(G_n) = s^+(\epsilon)$. Thus, (ii) holds.

To complete the proof of Proposition 1, it remains to show that, for $\epsilon \in (0, 1/2]$ fixed such that $\epsilon(2-\epsilon) \geq b$, $S^+(\epsilon) = \infty$. This result follows if we show that there exists a sequence of distributions $\{G_n\}_{n\geq 1} \subseteq \mathcal{F}_\epsilon$ satisfying $\underset{n\to\infty}{lim} \sigma(G_n) = \infty$.

Consider the sequence of distributions $\{G_n\}_{n\geq 1} \subseteq \mathcal{F}_\epsilon$, where $G_n = (1-\epsilon)F_0 + \epsilon H_n$ and $H_n(y) = \Phi(y/n)$. Let $\sigma(G_n)$ be the solution to the equation $\lambda_{G_n}(s) = 0$; by Lemma 2 with $\tilde{G} = G_n$, λ_{G_n} is strictly decreasing hence $\sigma(G_n)$ is uniquely defined, finite and strictly positive. Suppose, by contradiction, that there exists $K > 0$ such that $\sigma(G_n) \leq K$ for any $n \geq 1$. Then, using the monotonicity of λ_{G_n}, we have $\lambda_{G_n}(\sigma(G_n)) > \lambda_{G_n}(K)$ for any $n \geq 1$. Further, using that $\lambda_{G_n}(\sigma(G_n)) = 0$ for any $n \geq 1$, we get $\lambda_{G_n}(K) < 0$ for any $n \geq 1$. We now show that $\underset{n\to\infty}{lim} \lambda_{G_n}(K) \geq 0$, which contradicts the above.

By (i) of Lemma 6 with $s = K$, $\underset{n\to\infty}{lim} \lambda_{G_n}(K) = \lambda_+(K)$, so it suffices to show that $\lambda_+(K) \geq 0$. Using (ii) of Lemma 6, we obtain that $\lambda_+(K) \geq \epsilon(2-\epsilon) - b$. Since $\epsilon(2-\epsilon) \geq b$, we conclude that $\lambda_+(K) \geq 0$.

Proof of Proposition 2:

Fix $\epsilon \in (0, 1/2]$ such that $1 - \epsilon^2 > b$. In view of (18), to prove that $S^-(\epsilon) = s^-(\epsilon)$, it is enough to show the following: (i) $s^-(\epsilon) \leq \sigma(G)$ for any $G \in \mathcal{F}_\epsilon$ and (ii) there exists a sequence of distributions $\{G_n\}_{n\geq 1} \subseteq \mathcal{F}_\epsilon$ such that $\lim_{n\to\infty} \sigma(G_n) = s^-(\epsilon)$.

For (i), fix $G \in \mathcal{F}_\epsilon$ and note that, if the inclusion

$$\{s > 0 : \ s < s^-(\epsilon)\} \subseteq \{s > 0 : \lambda_G(s) > 0\} \tag{37}$$

holds, then the proof follows by taking infimum in both sides of (37) and using the definition of $\sigma(G)$ in (8). To prove (37), take $0 < s < s^-(\epsilon)$ and note that $\lambda_G(s) > \lambda_G(s^-(\epsilon))$ since, by (i) of Lemma 3, λ_G is strictly decreasing. To show $\lambda_G(s) > 0$ it therefore suffices to show $\lambda_G(s^-(\epsilon)) \geq 0$. This fact is proven below.

Using (iii) of Lemma 3 with $s = s^-(\epsilon)$, we express $\lambda_G(s^-(\epsilon))$ as:

$$\begin{aligned}\lambda_G(s^-(\epsilon)) = (1-\epsilon)^2 E\left[\chi\left(\frac{Z_2 - Z_1}{as^-(\epsilon)}\right)\right] + 2\epsilon(1-\epsilon) E\left[\chi\left(\frac{Z_2 - W_1}{as^-(\epsilon)}\right)\right] \\ + \epsilon^2 E\left[\chi\left(\frac{W_2 - W_1}{as^-(\epsilon)}\right)\right] - b.\end{aligned} \tag{38}$$

Here, Z_1 and Z_2 are independent random variables with common distribution F_0. Also, W_1 and W_2 are independent random variables with common distribution H. Finally, Z_2 and W_1 are independent.

To analyze the second term in (38), use that $Z_2 - W_1$ has density $g^*(x) = \int_{-\infty}^{\infty} h(x) f_0(x-t)dt$, with $h = H'$ and $f_0 = F_0'$. Then, using the symmetry and unimodality of f_0 together with the fact that χ is even and increasing, we have:

$$
\begin{aligned}
E\left[\chi\left(\frac{Z_2 - W_1}{as^-(\epsilon)}\right)\right] &= \int_{-\infty}^{\infty} \chi\left(\frac{x}{as^-(\epsilon)}\right) g^*(x)dx \\
&= \int_{-\infty}^{\infty} h(t)\left[\int_{-\infty}^{\infty} \chi\left(\frac{x}{as^-(\epsilon)}\right) f_0(x-t)dx\right] dt \\
&\geq \left[\int_{-\infty}^{\infty} h(t)dt\right]\left[\int_{-\infty}^{\infty} \chi\left(\frac{x}{as^-(\epsilon)}\right) f_0(x)dx\right] \\
&= E\left[\chi\left(\frac{Z_2}{as^-(\epsilon)}\right)\right].
\end{aligned}
$$

The third term in (38) is clearly positive as χ itself is positive. Therefore:

$$
\begin{aligned}
\lambda_G(s^-(\epsilon)) &\geq (1-\epsilon)^2 E\left[\chi\left(\frac{Z_2 - Z_1}{as^-(\epsilon)}\right)\right] + 2\epsilon(1-\epsilon)E\left[\chi\left(\frac{Z_2}{as^-(\epsilon)}\right)\right] - b \\
&= \lambda_-(s^-(\epsilon)) = 0.
\end{aligned}
$$

The first equality holds by (26) with $s = s^-(\epsilon)$, while the second equality holds by (25).

For (ii), define the sequence of distributions $\{G_n\}_{n\geq 1} \subseteq \mathcal{F}_\epsilon$ such that $G_n = (1-\epsilon)F_0 + \epsilon H_n$, where $H_n(y) = \Phi(ny)$. Then show that $\lim_{n\to\infty} \sigma(G_n) = s^-(\epsilon)$ using the same technique as in the proof of Proposition 1.

The proof will be completed once we show that $S^-(\epsilon) = 0$ for any $\epsilon \in (0, 1/2]$ for which $1 - \epsilon^2 \leq b$. This fact follows by showing that, for any such ϵ, there exists a sequence of distributions $\{G_n\}_{n\geq 1} \subseteq \mathcal{F}_\epsilon$ satisfying $\lim\limits_{n\to\infty} \sigma(G_n) = 0$. This is established using an argument by contradiction as in the proof of Proposition 1.

Proof of Theorem 3:

Let $\epsilon \in (0, 1/2]$. Using the definition of $\overline{B}_g(\epsilon)$ in (16), the explicit expressions for $S^+(\epsilon)$ and $S^-(\epsilon)$ provided in Propositions 1 and 2 and the fact that $\sigma = 1$, we obtain:

$$
\overline{B}_g(\epsilon) = \begin{cases} max\{L_2(s^+(\epsilon)), L_1(s^+(\epsilon))\} & \text{if} \quad \epsilon(2-\epsilon) < b < 1 - \epsilon^2 \\ \infty & \text{else.} \end{cases}
$$

To prove the theorem, it therefore suffices to solve the system of inequalities below with respect to ϵ

$$\begin{cases} 0 < \epsilon \leq 1/2 \\ \epsilon(2-\epsilon) < b \\ 1-\epsilon^2 > b. \end{cases}$$

One can easily see that the ϵ's that solve this system must satisfy

$$\begin{cases} \epsilon \in (0, 1/2] \\ \epsilon \in (-\infty, 1-\sqrt{1-b}) \cup (1+\sqrt{1-b}, +\infty) \\ \epsilon \in (-\sqrt{1-b}, \sqrt{1-b}). \end{cases}$$

In particular, if $b = 3/4$, then $\epsilon \in (0, 1/2)$. If $b \in (0, 3/4)$, then $\epsilon \in (0, 1-\sqrt{1-b})$. Finally, if $b \in (3/4, 1)$, then $\epsilon \in (0, \sqrt{1-b})$.

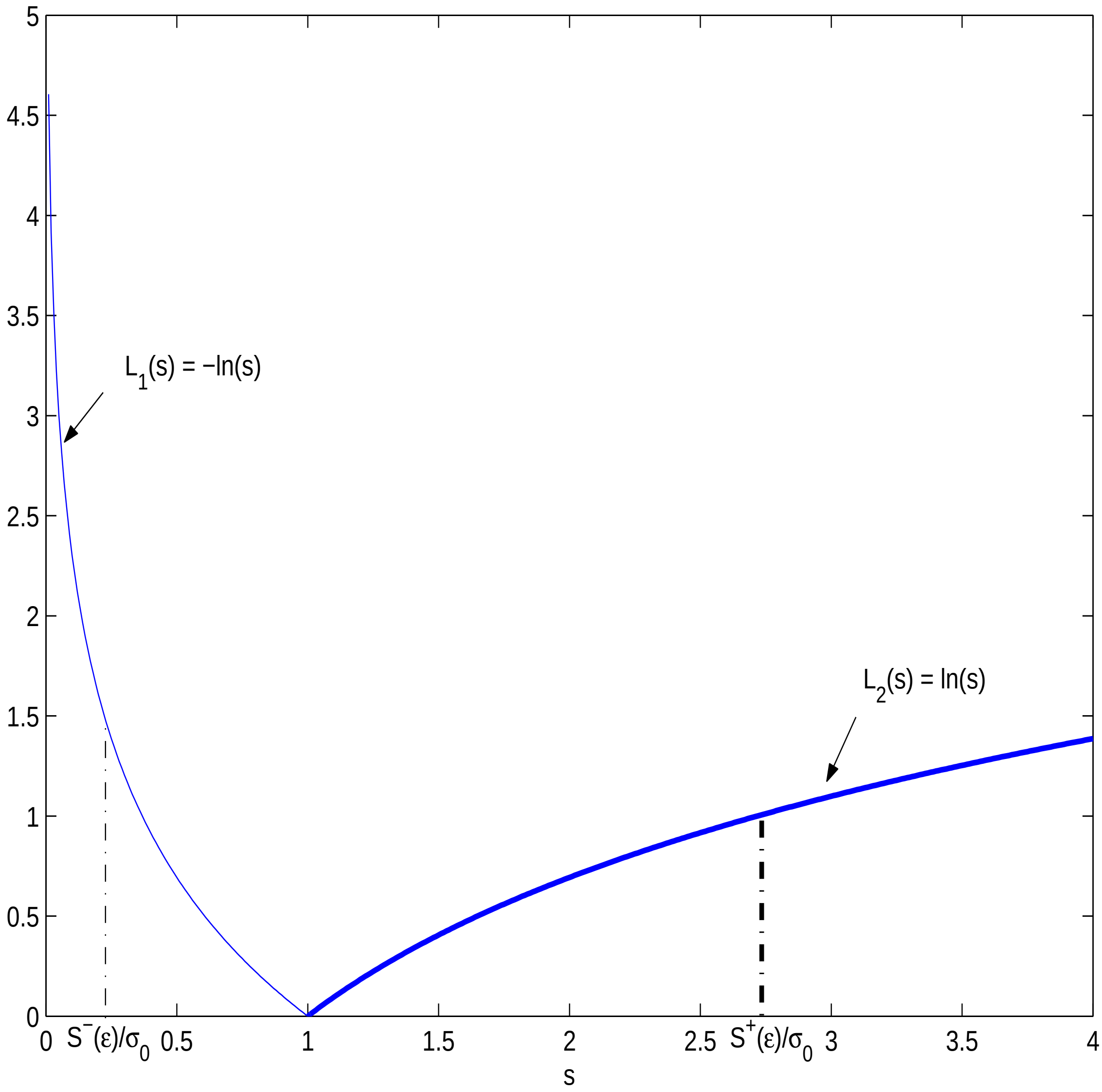

Figure 1: Plot of the functions $L_1(s) = -\ln(s), 0 < s \leq 1$, and $L_2(s) = \ln(s), s \geq 1$. For the situation depicted in this figure, the maxbias is $\overline{B}_g(\epsilon) = -\ln(S^-(\epsilon)/\sigma)$.

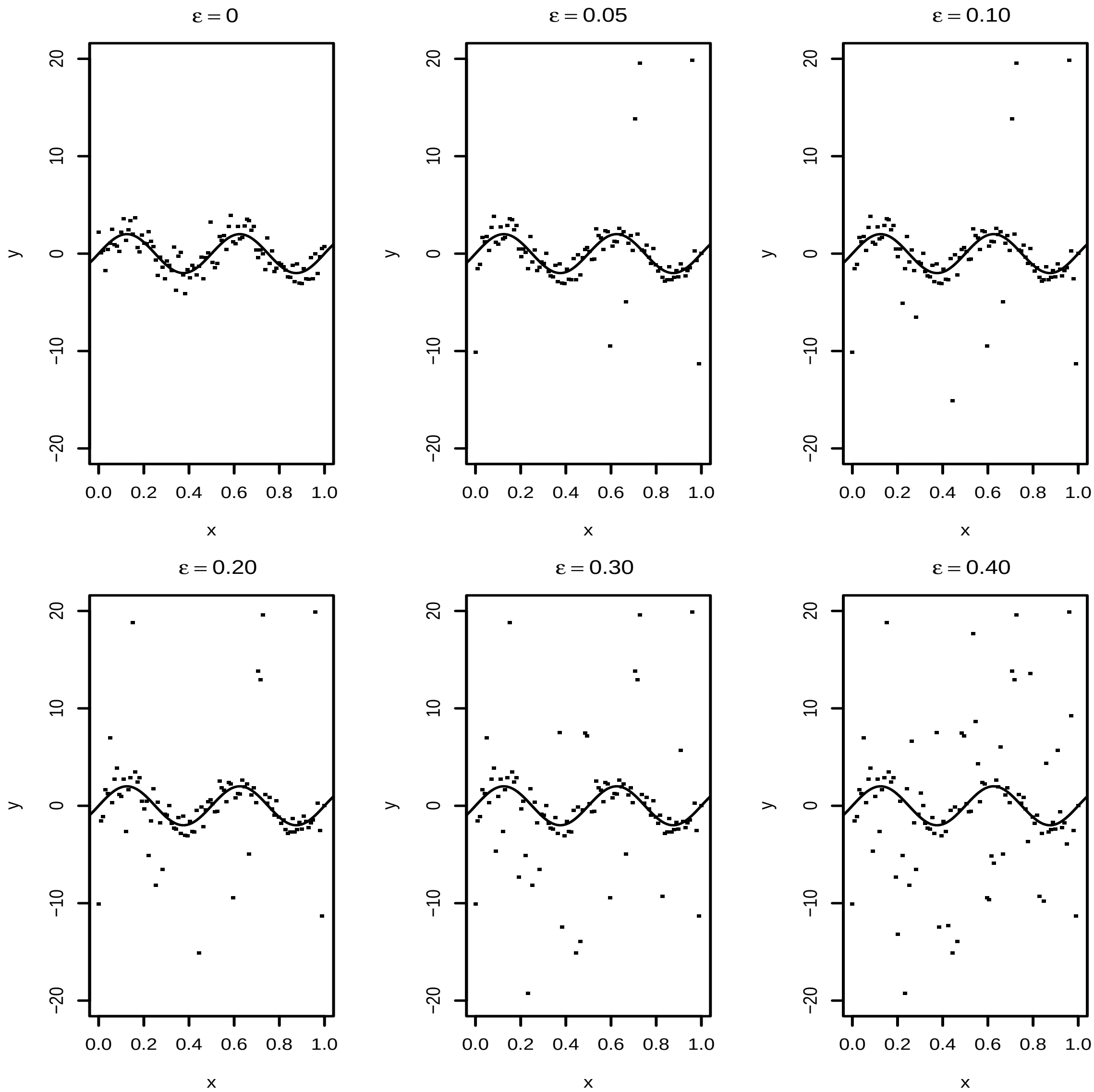

Figure 2: Data simulated from model (1) for the simulation settings with $n = 100$ and $H(y) = \Phi(y/10)$. The six panels show data corresponding to different amounts of contamination. The true regression function is superimposed.

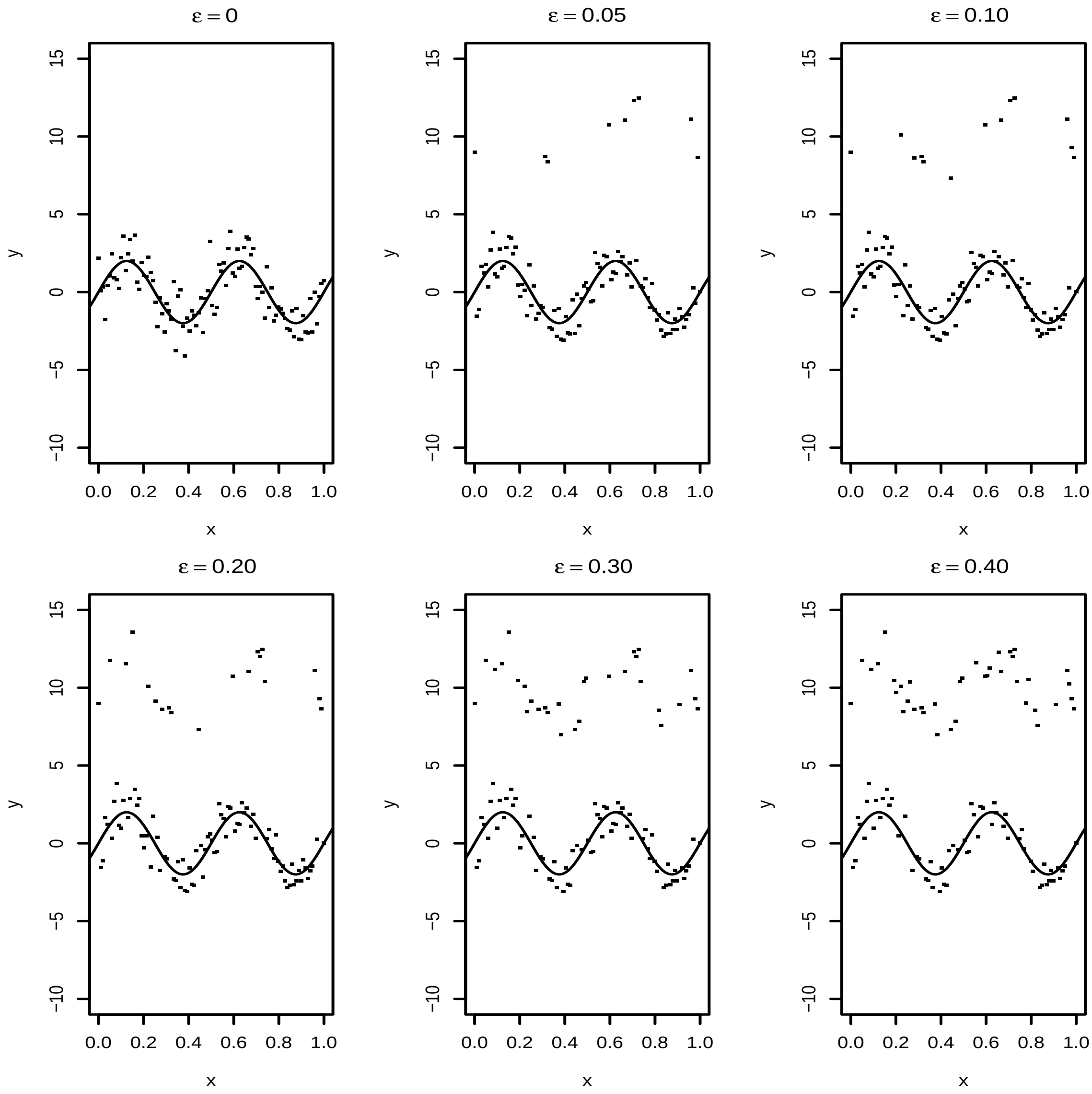

Figure 3: Data simulated from model (2) for the simulation settings with $n = 100$ and $H(y) = \Phi(y - 10)$. The six panels show data corresponding to different amounts of contamination. The true regression curve is superimposed.

TABLE 1: Efficiencies of $\widehat{\sigma}_n^{(2)}$, $\widehat{\sigma}_n^{(3)}$ and $\widehat{\sigma}_n^{(4)}$ relative to $\widehat{\sigma}_n^{(1)}$ for the simulation settings with $\epsilon = 0$.

n	$RE(\widehat{\sigma}_n^{(2)}, \widehat{\sigma}_n^{(1)})$	$RE(\widehat{\sigma}_n^{(3)}, \widehat{\sigma}_n^{(1)})$	$RE(\widehat{\sigma}_n^{(4)}, \widehat{\sigma}_n^{(1)})$
20	0.592	0.373	0.447
50	0.653	0.438	0.525
100	0.672	0.454	0.535

TABLE 2: Estimates for the mean squared error of the scale estimators $\widehat{\sigma}_n^{(1)}$, $\widehat{\sigma}_n^{(2)}$, $\widehat{\sigma}_n^{(3)}$ and $\widehat{\sigma}_n^{(4)}$. The contaminating distribution is symmetric; its distribution function is $H(y) = \Phi(y/10)$.

	Estimator	$\epsilon = 0.05$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$	$\epsilon = 0.40$
$n = 20$	$\widehat{\sigma}_n^{(1)}$	3.012	6.385	13.805	21.660	29.930
	$\widehat{\sigma}_n^{(2)}$	0.273	0.549	2.349	7.346	15.860
	$\widehat{\sigma}_n^{(3)}$	0.519	0.805	1.816	4.394	10.060
	$\widehat{\sigma}_n^{(4)}$	0.352	0.578	1.449	3.720	8.544
$n = 50$	$\widehat{\sigma}_n^{(1)}$	2.505	5.661	13.027	20.900	29.220
	$\widehat{\sigma}_n^{(2)}$	0.072	0.163	0.730	3.181	10.060
	$\widehat{\sigma}_n^{(3)}$	0.129	0.217	0.602	1.520	3.808
	$\widehat{\sigma}_n^{(4)}$	0.092	0.168	0.526	1.455	3.867
$n = 100$	$\widehat{\sigma}_n^{(1)}$	2.319	5.452	12.78	20.770	29.020
	$\widehat{\sigma}_n^{(2)}$	0.038	0.100	0.485	2.106	7.781
	$\widehat{\sigma}_n^{(3)}$	0.063	0.118	0.382	1.050	2.672
	$\widehat{\sigma}_n^{(4)}$	0.047	0.098	0.356	1.055	2.849

TABLE 3: Estimates for the mean squared error of the scale estimators $\widehat{\sigma}_n^{(1)}$, $\widehat{\sigma}_n^{(2)}$, $\widehat{\sigma}_n^{(3)}$ and $\widehat{\sigma}_n^{(4)}$. The contaminating distribution is asymmetric; its distribution function is $H(y) = \Phi(y - 10)$.

	Estimator	$\epsilon = 0.05$	$\epsilon = 0.10$	$\epsilon = 0.20$	$\epsilon = 0.30$	$\epsilon = 0.40$
$n = 20$	$\widehat{\sigma}_n^{(1)}$	2.564	5.281	10.32	14.120	16.440
	$\widehat{\sigma}_n^{(2)}$	0.329	1.049	6.691	17.160	26.900
	$\widehat{\sigma}_n^{(3)}$	0.560	0.917	2.274	5.011	8.280
	$\widehat{\sigma}_n^{(4)}$	0.386	0.678	1.842	4.054	6.591
$n = 50$	$\widehat{\sigma}_n^{(1)}$	2.160	4.816	9.882	13.720	16.090
	$\widehat{\sigma}_n^{(2)}$	0.082	0.204	1.588	9.970	23.120
	$\widehat{\sigma}_n^{(3)}$	0.140	0.248	0.645	1.306	2.053
	$\widehat{\sigma}_n^{(4)}$	0.102	0.196	0.577	1.276	2.119
$n = 100$	$\widehat{\sigma}_n^{(1)}$	2.037	4.703	9.778	13.670	16.050
	$\widehat{\sigma}_n^{(2)}$	0.045	0.126	0.668	5.461	20.050
	$\widehat{\sigma}_n^{(3)}$	0.069	0.135	0.412	0.897	1.434
	$\widehat{\sigma}_n^{(4)}$	0.052	0.115	0.393	0.913	1.528