Globally robust confidence intervals for simple linear regression

Jorge Adrover [a], Matias Salibian-Barrera [b,*]

[a] *Universidad Nacional de Córdoba, CONICET and CIEM, research partially supported by Grants PICT 21407 from ANPCYT and PIP 5505 (CONICET), Argentina*

[b] *University of British Columbia, research supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada*

Abstract

It is well known that when the data may contain outliers or other departures from the assumed model, classical inference methods can be seriously affected and yield confidence levels much lower than the nominal ones. This paper proposes robust confidence intervals and tests for the parameters of the simple linear regression model that maintain their coverage and significance level, respectively, over whole contamination neighbourhoods. This approach can be used with any consistent regression estimator for which maximum bias curves are tabulated, and thus it is more widely applicable than previous proposals in the literature. Although the results regarding the coverage level of these confidence intervals are asymptotic in nature, simulation studies suggest that these robust inference procedures work well for small samples, and compare very favourably with earlier proposals in the literature.

Key words: Robustness, Robust inference, Linear regression, Robust confidence intervals, Robust tests

* Corresponding Author
Email addresses: adrover@mate.uncor.edu (Jorge Adrover), matias@stat.ubc.ca (Matias Salibian-Barrera).

1 Introduction

Consider the simple linear regression model where we observe a bivariate random sample $(Y_1, X_1), \ldots, (Y_n, X_n)$ satisfying

$$Y_i = \beta_0 + \beta_1 (X_i - \mu_X) + \sigma_0 \epsilon_i \,, \; i = 1, \ldots, n \,, \tag{1}$$

where X_i are univariate explanatory variables, $\mu_X = \text{med}(X_i)$, the errors ϵ_i follow a known distribution F_0 and satisfy $\text{med}(\epsilon_i | X_i) = 0$, $i = 1, \ldots, n$. In general, one assumes that the data are generated by a distribution $H_{\boldsymbol{\theta}}$ belonging to a parametric family of distributions $\{H_{\boldsymbol{\theta}}\}$, with $\boldsymbol{\theta} \in \mathbb{R}^2$. To allow for outliers and other departures from the model, we will assume that the data follow a distribution H in an ϵ-contamination neighbourhood $\mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})$ of the true underlying parametric model. More specifically,

$$\mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}}) = \left\{ H = (1-\varepsilon)\, H_{\boldsymbol{\theta}} + \varepsilon H^* \,, \; H^* \text{ an arbitrary distribution on } \mathbb{R}^2 \right\} \,, \tag{2}$$

where $0 < \varepsilon < 0.5$.

Confidence intervals based on maximum likelihood estimators may be seriously affected by a small proportion of atypical observations (see, e.g. [17], [6], [11], [12], [4], [8], and [1]). We will say that a confidence interval is robust if it is able to maintain a high coverage level and a reasonable length when the data comes from any distribution in the contamination neighbourhood (2). Formally, we have the following

Definition 1 *A confidence interval (L_n, U_n) for $\theta \in \mathbb{R}$ is called* globally robust of level $(1-\alpha)$ *if it satisfies the following conditions:*

(1) (Stable interval) *The minimum asymptotic coverage over the ε-contamination neighbourhood is $1-\alpha$, i.e.*

$$\lim_{n\to\infty} \inf_{H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})} P_H \left(L_n < \theta < U_n \right) \geq 1 - \alpha \,.$$

(2) (Informative interval) *The maximum asymptotic length of the interval is*

bounded over the ε-contamination neighbourhood, i.e.

$$\lim_{n\to\infty} \sup_{H\in\mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})} [U_n - L_n] < \infty\,.$$

It is easy to see that, for the location model, confidence intervals of the form $\overline{X}_n \pm t_{(n-1)}(\alpha/2)S_n/\sqrt{n}$ do not satisfy either Parts 1 or 2 of Definition 1 above. The problem with the above confidence intervals is not solely due to the lack of robustness of the estimators $\overline{X}_n$ and S_n. It can be shown that even if we replace the sample mean and standard deviation by robust counterparts $\hat{\theta}_n$ and $\hat{\sigma}_n$, the resulting confidence interval only satisfies Part 2 of the above Definition.

The failure of intervals of the form $\hat{\theta}_n \pm t_{(n-1)}(\alpha/2)\hat{\sigma}_n/\sqrt{n}$ to satisfy Part 1 above is due to the fact that while the length of the interval converges to zero as $n \to \infty$, its center $\hat{\theta}_n$ may converge to a value different from the parameter of interest θ. This problem can be fixed taking into account the largest possible difference between $\hat{\theta}(H)$, the limiting value of $\hat{\theta}_n$, and the parameter of interest θ, across distributions H in the contamination neighbourhood $\mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})$. This quantity is related to the maximum asymptotic bias of the estimator $\hat{\theta}_n$ (e.g. see [10]).

For the location model $Y_i = \theta + \sigma_0\,\epsilon_i$, the maximum asymptotic bias of $\hat{\theta}_n$ is

$$B(\varepsilon) = \sup_{H\in\mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})} \frac{\left|\hat{\theta}(H) - \theta\right|}{\sigma_0}\,,$$

and thus, $\left|\hat{\theta}(H) - \theta\right| \leq B(\varepsilon)\,\sigma_0$ for all $H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})$. Let $\hat{\sigma}_n$ be an estimator of σ_0 with limit $\hat{\sigma}(H)$, which in principle may be different from σ_0. For each $H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})$ we have

$$\left|\hat{\theta}(H) - \theta\right| \leq B(\varepsilon)\,\sigma_0 = B(\varepsilon)\,\frac{\sigma_0}{\hat{\sigma}(H)}\,\hat{\sigma}(H) \leq B(\varepsilon)\,K(\varepsilon)\,\hat{\sigma}(H)\,, \tag{3}$$

where $K(\varepsilon) = \sup_{H\in\mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})} \sigma_0/\hat{\sigma}(H)$. Tabulated values of $K(\varepsilon)$ for different scale estimators are available in [3]. Hence, we can estimate the largest difference $\left|\hat{\theta}(H) - \theta\right|$ using $B(\varepsilon)\,K(\varepsilon)\,\hat{\sigma}_n$.

In the linear regression model, the maximum asymptotic bias for the slope β_1 is

$$B(\varepsilon) = \sup_{H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})} \left|\hat{\beta}_1(H) - \beta_1\right| \frac{\sigma_X}{\sigma_0}\,,$$

where $\hat{\beta}_1(H)$ is the limit of the slope estimator $\hat{\beta}_{1,n}$ when the data have distribution H, and σ_X is the scale of the covariates under model (1). If the estimator is equivariant under affine transformations, $B(\varepsilon)$ above does not depend on the value of the parameters under the central model (see [13]).

Similarly to (3), we have

$$\left|\hat{\beta}_1(H) - \beta_1\right| \leq B(\varepsilon)\,\frac{\sigma_0}{\sigma_X} \leq B(\varepsilon)\,K(\varepsilon)\,\frac{\hat{\sigma}(H)}{\hat{\sigma}_X(H)}\,, \tag{4}$$

for each $H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})$, where

$$K(\varepsilon) = \sup_{H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})} [\sigma_0/\hat{\sigma}(H)] \sup_{H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})} [\hat{\sigma}_X(H)/\sigma_X]\,,$$

and $\hat{\sigma}_X(H)$ is the limit of the scale estimator $\hat{\sigma}_{X,n}$. The quantities $B(\varepsilon)$ and $K(\varepsilon)$ on the right-hand side of (4) are available for some estimators (see [14]) and the ratio $\hat{\sigma}(H)/\hat{\sigma}_X(H)$ can be estimated for a given sample by $\hat{\sigma}_n/\hat{\sigma}_{X,n}$.

In [1] the authors proposed robust confidence intervals of the form $\hat{\beta}_1 \pm q_n$, where: q_n satisfies

$$\Phi\left(\frac{q_n - \bar{\beta}_1}{v_n}\right) + \Phi\left(\frac{q_n + \bar{\beta}_1}{v_n}\right) - 1 = 1 - \alpha\,; \tag{5}$$

$\sqrt{n}\,v_n$ is a consistent estimator of the asymptotic variance of $\hat{\beta}_1$; and $\bar{\beta}_1$ is the estimated bias bound in (4):

$$\bar{\beta}_1 = B(\epsilon)\,K(\epsilon)\,\frac{\hat{\sigma}_n}{\hat{\sigma}_{X,n}}\,.$$

It follows from (5) that this approach to constructing robust confidence intervals will typically produce shorter intervals when applied to estimators with relatively small maximum asymptotic bias. Furthermore, note that this methodology requires using an estimator that is asymptotically normal over the whole contamination neighbourhood $\mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})$ with estimable asymptotic

variance, and the estimation of the bias bound in (4). Unfortunately, the sample variability of the ratio $\hat{\sigma}_n/\hat{\sigma}_{X,n}$ may affect the estimated upper bound in (4), which could result in confidence intervals with finite sample coverage different from the nominal level (see [1] for extensive simulation studies).

In this paper we discuss a new approach to construct robust confidence intervals based on estimators that are consistent over $\mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})$, and have a known maximum asymptotic bias. They are constructed using a non-parametric confidence interval for the location model that only assumes errors with median zero. Note that this proposal does not require the asymptotic distribution of the estimator to be known. Furthermore, to construct a confidence interval for the slope β_1 in (1), only the maximum asymptotic bias of the estimator for β_0 is needed, and viceversa. As a result, this approach can be applied more widely than previous proposals, and, furthermore, its finite-sample properties are better than those of other methods in the literature.

The rest of the paper is organized as follows. Section 2 introduces non-parametric confidence intervals for the simple linear regression model. Section 3 shows how to correct these intervals to account for the damaging effect of the asymptotic bias of the estimators on which they are based. Section 4 reports the result of a simulation study on the finite-sample properties of this approach. Section 5 derives robust hypothesis tests based on these confidence intervals, while Section 6 illustrates their application on real data sets. Some comments on future work are included in Section 7, and Section 8 presents concluding remarks. Finally, all technical proofs can be found in the Appendix.

2 Non-parametric confidence intervals for the slope and intercept

As before, assume that $(Y_1, X_1), \ldots, (Y_n, X_n)$ are a bivariate random sample following the simple regression model (1). The basic idea of our approach is as follows. For each $a, b \in \mathbb{R}$, consider pseudo-observations $Z_i(b) = Y_i - b\,(X_i - \mu_X)$, and $W_i(a) = (Y_i - a)\,/\,(X_i - \mu_X)$, for $i = 1, \ldots, n$. Under the model, we

have

$$Z_i(b) \;= \beta_0 + (\beta_1 - b)\,(X_i - \mu_X) + \sigma_0\epsilon_i = \beta_0 + \tilde{\epsilon}_i\,(b)\;,$$

$$W_i(a) = \beta_1 + \frac{\beta_0 - \;a}{X_i - \mu_X} + \frac{\sigma_0\epsilon_i}{X_i - \mu_X} \qquad = \beta_1 + \ddot{\epsilon}_i\,(a)\;, \tag{6}$$

with

$$\tilde{\epsilon}_i\,(b) = (\beta_1 - b)\;(X_i - \mu_X) + \sigma_0\epsilon_i$$

$$\ddot{\epsilon}_i\,(a) = \frac{\beta_0 - a + \sigma_0\epsilon_i}{X_i - \mu_X}\,.$$

Note that the above approach transforms the linear regression model (1) into two location models for the pseudo-observations $Z_i(b)$ and $W_i(a)$, $i = 1, \ldots, n$, respectively. Furthermore, note that $\tilde{\epsilon}_i(\beta_1) = \sigma_0\epsilon_i$ and $\ddot{\epsilon}_i(\beta_0) = \sigma_0\epsilon_i/(X_i - \mu_X)$, which satisfy $\text{med}\,(\tilde{\epsilon}_i(\beta_1)) = \text{med}\,(\ddot{\epsilon}_i\,(\beta_0)) = 0$. This follows from the fact that $\text{med}(\epsilon_i) = 0$ implies $\text{med}(\epsilon_i/(X_i - \mu_X)) = 0$ (see Lemma 4 in the Appendix).

Hence, if μ_X and β_0 were known, one could construct a robust confidence interval for β_1 using a robust confidence interval for location applied to the pseudo-observations $W_i(\beta_0)$ in (6). In this paper we consider robust non-parametric confidence intervals as proposed in [18]. Given independent observations $U_1, \ldots, U_n$ with $\text{med}(U_i) = \theta$, $i = 1, \ldots, n$ and distribution function H_θ, $\theta \in \mathbb{R}$, a non-parametric confidence interval for θ of approximate level $(1 - \alpha)$ is given by

$$\left[U_{(k+1)}, U_{(n-k)}\right)\;, \tag{7}$$

where k satisfies

$$k = \max\left\{j \;:\; P\left(j < Bi(n, p) < n - j\right) \geq 1 - \alpha\right\}\;, \tag{8}$$

with $p = 1/2$. The probability of success $p = 1/2$ in (8) is associated with the fact that

$$p = P_{H_\theta}(U_i < \theta) = 1/2\,. \tag{9}$$

Following the proof of Theorem 1 in [18], we have that when the observations are contaminated, for $H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})$,

$$\begin{aligned}(1-\varepsilon)/2 = (1-\varepsilon)P_{H_\theta}\left(U_i < \theta\right) & \\ \leq p = P_H\left(U_i < \theta\right) & \\ \leq (1-\varepsilon)P_{H_\theta}\left(U_i < \theta\right) + \varepsilon = (1+\varepsilon)/2\,, & \quad (10)\end{aligned}$$

and p ranges between $(1-\varepsilon)/2$ and $(1+\varepsilon)/2$. Hence, the actual coverage level of the intervals in (7) for the contamination neighbourhood $\mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})$ will be

$$\inf_{(1-\varepsilon)/2 \,\leq\, p \,\leq\, (1+\varepsilon)/2} h(n,k,p)\ ,$$

where

$$h(n,k,p) = P\left(k < Bi(n,p) < n-k\right)\ , \qquad (11)$$

Lemma 1 in the same paper shows that $h(n,k,p) = h(n,k,1-p)$ and that $h(n,k,p)$ is non-decreasing in $0 \leq p \leq 0.5$ for $k = 0,\ 1,\ \ldots,\ [n/2]$, and thus the infimum is attained at $p = (1-\varepsilon)/2$. In other words,

$$\inf_{H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})} P_H\left(\theta \in \left[U_{(k+1)}, U_{(n-k)}\right)\right) = P(k < Bi(n,(1-\varepsilon)/2) < n-k)\,,$$

and hence, if k satisfies

$$P(k < Bi(n,(1-\varepsilon)/2) < n-k) \geq 1-\alpha\,, \qquad (12)$$

then the level of the confidence interval in (7) is at least $1-\alpha$ over the whole contamination neighbourhood $\mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})$.

This discussion shows that if μ_X and β_0 are known, and k satisfies (12), then

$$\left[W_{(k+1)}(\beta_0), W_{(n-k)}(\beta_0)\right)\ ,$$

is a robust non-parametric confidence interval for the slope β_1 of model (1), of level $(1-\alpha)$ over the whole contamination neighbourhood $\mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})$.

Since typically neither μ_X nor β_0 are known, the natural procedure is to replace them with robust estimators, $\hat{\mu}_{X,n}$ and $\hat{\beta}_{0,n}$, respectively, and consider

confidence intervals of the form

$$\left[W_{(k+1)}(\hat{\beta}_{0,n}), W_{(n-k)}(\hat{\beta}_{0,n})\right) . \tag{13}$$

Similarly, given a robust estimator $\hat{\beta}_{1,n}$ a robust confidence interval for the intercept β_0 in model (1) is

$$\left[Z_{(k+1)}(\hat{\beta}_{1,n}), Z_{(n-k)}(\hat{\beta}_{1,n})\right) .$$

We ran a small simulation experiment to study the empirical coverage of the intervals in (13) using $\hat{\mu}_{X,n} = \operatorname{med}(X_1, \ldots, X_n)$ and

$$\hat{\beta}_{0,n} = \operatorname*{med}_{1\le i\le n}\left(Y_i - \hat{\beta}_{1,n}\ (X_i - \hat{\mu}_{X,n})\right) , \tag{14}$$

where $\hat{\beta}_{1,n}$ is the "repeated medians of slopes" estimator [16]:

$$\hat{\beta}_{1,n} = \operatorname*{med}_{1\le j\le n}\ \operatorname*{med}_{i\ne j}\frac{Y_i - Y_j}{X_i - X_j} .$$

We generated 1000 random samples following model (1) with $\beta_0 = \beta_1 = 0$, $X_i \sim \mathcal{N}(0,1)$, $\epsilon_i \sim \mathcal{N}(0,1)$, $\sigma_0 = 1$ and ϵ_i independent from X_i, $i = 1, \ldots, n$. We computed confidence intervals of level 95% as defined in (13), with k as in (12) with $\varepsilon = 0$. Table 1 shows the empirical coverages and median lengths for different sample sizes. Since these confidence intervals correspond to inverting the classical sign test for location, and the samples do not contain outliers, it is not surprising that the empirical coverages are close to the nominal level, and their lengths converge to zero as the sample size increases.

Next we contaminated these samples with 5, 10 and 20% of outliers with distribution $\mathcal{N}_2\left((x_0, y_0)', 0.01\ \mathbf{I}\right)$, where (x_0, y_0) were $(3, 1.5)$, $(5, 2.5)$ and $(5, 15)$. These outliers correspond to three leverage levels: "mild", "medium" and "strong", respectively. The results in Table 2 show that the empirical coverage decreases to zero as the sample size and the level of contamination increase. This means that the intervals are not able to include the slope parameter in the presence of contamination and large sample size, although their lengths are not vanishing. The concealed fact behind this behavior is that, although

very robust, when the data contain outliers, the estimator for β_1 is biased and does not converge to the slope parameter under the model.

It is worth noting that this problem is not due to the choice of the robust estimator for β_0 (or β_1): when the data do not follow the central model all estimators may be biased, in the sense of converging to a different value of the parameter of interest. Furthermore, note that the repeated medians is among the best options available to deal with this bias problem. The rest of the paper is concerned with resolving this issue.

Sample size	Coverage and median length
20	0.94 (1.75)
40	0.93 (1.07)
60	0.94 (0.81)
80	0.92 (0.67)
100	0.92 (0.59)
200	0.95 (0.42)
500	0.94 (0.27)
2000	0.95 (0.14)
10000	0.96 (0.06)

Table 1
Empirical coverages and median lengths of 1000 confidence intervals of level 95% as in (13) for the slope β_1 under the linear model (1).

3 Improving the coverage levels by correcting for maximum asymptotic bias

Note that the asymptotic version of (13) is constructed using $\hat{\beta}_0$ instead of $\hat{\beta}_{0,n}$ and $\hat{\mu}_X$ instead of $\hat{\mu}_{X,n}$. We have

$$W_i(\hat{\beta}_0) = \beta_1 + \frac{\sigma_0\epsilon_i - \left(\hat{\beta}_0 - \beta_0\right)}{X_i - \hat{\mu}_X} = \beta_1 + \frac{\sigma_0\epsilon_i - b(\hat{\beta}_0)}{X_i - \mu_X - b(\hat{\mu}_X)}\,, \qquad (15)$$

Perc. of cont.	Sample size	Mild	Medium	Strong
5%	20	0.95 (1.61)	0.95 (1.62)	0.95 (1.91)
	40	0.94 (1.00)	0.94 (1.00)	0.94 (1.16)
	60	0.95 (0.88)	0.95 (0.88)	0.95 (1.01)
	80	0.95 (0.73)	0.95 (0.74)	0.95 (0.80)
	100	0.93 (0.66)	0.93 (0.66)	0.93 (0.71)
	200	0.94 (0.54)	0.94 (0.54)	0.95 (0.55)
	500	0.93 (0.40)	0.93 (0.40)	0.93 (0.40)
	2000	0.94 (0.29)	0.95 (0.29)	0.95 (0.29)
	10000	0.90 (0.21)	0.90 (0.22)	0.90 (0.22)
10%	20	0.92 (1.50)	0.92 (1.49)	0.91 (2.19)
	40	0.96 (1.10)	0.96 (1.10)	0.95 (1.57)
	60	0.94 (0.91)	0.95 (0.91)	0.94 (1.31)
	80	0.95 (0.80)	0.95 (0.81)	0.95 (1.12)
	100	0.94 (0.76)	0.94 (0.77)	0.94 (1.05)
	200	0.94 (0.63)	0.94 (0.64)	0.94 (0.80)
	500	0.92 (0.54)	0.92 (0.56)	0.92 (0.65)
	2000	0.85 (0.45)	0.85 (0.47)	0.85 (0.50)
	10000	0.65 (0.41)	0.64 (0.42)	0.64 (0.43)
20%	20	0.94 (1.78)	0.94 (1.77)	0.92 (3.77)
	40	0.96 (1.22)	0.95 (1.23)	0.92 (3.33)
	60	0.93 (0.97)	0.93 (0.97)	0.88 (3.04)
	80	0.91 (0.81)	0.91 (0.81)	0.86 (2.79)
	100	0.91 (0.77)	0.90 (0.77)	0.85 (2.59)
	200	0.86 (0.59)	0.84 (0.59)	0.71 (1.95)
	500	0.61 (0.47)	0.54 (0.47)	0.37 (1.55)
	2000	0.11 (0.40)	0.04 (0.40)	0.02 (1.29)
	10000	0.00 (0.36)	0.00 (0.35)	0.00 (1.17)

Table 2

Empirical coverages and median lengths of 1000 confidence intervals of the form (13) for the slope parameter β_1. The nominal level is 95% and different proportion of outliers were placed at: $x_0 = 3, y_0 = 1.5$ (Mild), $x_0 = 5, y_0 = 2.5$ (Medium), and $x_0 = 5, y_0 = 15$ (Strong).

with $b(\hat{\mu}_X) = \hat{\mu}_X - \mu_X$, $b(\hat{\beta}_0) = \hat{\beta}_0 - \beta_0$. Observe that neither $\sigma_0\epsilon_i - b(\hat{\beta}_0)$ nor $X_i - \mu_X - b(\hat{\mu}_X)$ have median zero, and thus, in general, med $\left(W_i(\hat{\beta}_0)\right) \neq \beta_1$.

Hence, we need to revisit the bounds in (10) taking into account the fact that

$$P_{H_0}\left(W_i(\hat{\beta}_0) < \beta_1\right) \neq 1/2\,,$$

where H_0 is the distribution of the data undel model (1). Assume that

A1: X_i and ϵ_i, $i = 1, \ldots, n$ are independent under the model (1);
A2: ϵ_i, $i = 1, \ldots, n$ have a continuous cdf F_0 under the model (1);

and let $G_0(x)$ be the cdf of $(X_i - \mu_X)/\sigma_X$ under the model (1). Denote $G_0^-(a) = \lim_{x \nearrow a} G_0(x)$. The following quantities account for the positive and negative contribution of the bias of the estimate:

$$B_+(\hat{\beta}_0) = \sup_{H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})} \frac{\widehat{\beta}_0(H) - \beta_0}{\sigma_0}\,, \tag{16}$$

$$B_-(\hat{\beta}_0) = \inf_{H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})} \frac{\widehat{\beta}_0(H) - \beta_0}{\sigma_0}\,. \tag{17}$$

and similarly for $B_+(\hat{\mu}_X)$ and $B_-(\hat{\mu}_X)$.

The following theorem shows how to modify the bounds in (10) taking into account the asymptotic bias of $\hat{\beta}_0$ to obtain globally robust confidence intervals for the slope parameter in model (1).

Theorem 1 *Assume that A1 and A2 hold. Let $\hat{\beta}_0$ and $\hat{\mu}_X$ be the asymptotic versions of $\hat{\beta}_{0,n}$ and $\hat{\mu}_{X,n}$, respectively. Assume also that*

$$\begin{aligned} B_+(\hat{\beta}_0) &= -B_-(\hat{\beta}_0) = B(\hat{\beta}_0) \\ B_+(\hat{\mu}_X) &= -B_-(\hat{\mu}_X) = B(\hat{\mu}_X) \end{aligned} \tag{18}$$

Then a robust confidence interval for the slope β_1 in model (1), based on $\hat{\beta}_{0,n}$ and $\hat{\mu}_{X,n}$, is

$$\left[W(\hat{\beta}_{0,n})_{(k+1)}, W(\hat{\beta}_{0,n})_{(n-k)}\right)\,, \tag{19}$$

where k satisfies

$$h\left(n, k, \tilde{p}_S\right) \geq 1 - \alpha\,, \tag{20}$$

$h\,(n,k,p)$ is defined in (11) *and $\tilde{p}_S$ is taken to be*

$$\tilde{p}_S = (1-\epsilon)\,\left[F_0\left(B(\hat{\beta}_0)\right) + G_0^-\left(B(\hat{\mu}_X)\right) - 2\,F_0\left(B(\hat{\beta}_0)\right)\,G_0^-\left(B(\hat{\mu}_X)\right)\right]\,. \quad (21)$$

Remark 1 *A sufficient condition for assumption* (18) *is that both $F_0(x)$ and $G_0(x)$ be symmetric and unimodal (see Theorem 8 in [2]).*

Remark 2 *From the proof of Theorem 8 in [2] it can be shown that when $\hat{\mu}_{X,n} = \mathrm{med}(X_1,\ldots,X_n)$, $\hat{\beta}_{0,n}$ is as in* (14)*, and X is symmetrically distributed, the sequence of point contaminations at $(x_0,y_0) = \pm(n,n^2)$ attain the bounds* (16) *and* (17) *respectively, simultaneously for $\hat{\mu}_X$ and $\hat{\beta}_0$. Similarly, in the case of contaminations at $(x_0,y_0) = \pm(n,-n^2)$ attain these bounds. Both facts together entail that the value $\tilde{p}_S$ is the smallest p to be used over the whole neighbourhood to keep the global coverage.*

We now turn our attention to confidence intervals for β_0 in (1). Recall that

$$Z(\hat{\beta}_1) = \beta_0 - (\beta_1 - \hat{\beta}_1)\,(X - \mu_X) + \sigma_0\epsilon\,.$$

As before, we have that

$$P_{H_0}\left(Z_i(\hat{\beta}_1) < \beta_0\right) \neq 1/2$$

and we need to find appropriate bounds for $P_H\left(Z_i(\hat{\beta}_1) < \beta_0\right)$. To our assumptions A1 and A2 we now need to add:

A3: X_i, $i = 1, \ldots, n$, have a continuous distribution G_0 with symmetric and unimodal density g_0 under model (1).

Similarly to Theorem 1 above, the following result shows how to incorporate in (10) the asymptotic bias of the slope estimator to obtain globally robust confidence intervals for the intercept parameter in model (1).

Theorem 2 *Assume that A1, A2 and A3 hold. Let $\hat{\beta}_1$ and $\hat{\mu}_X$ be the asymptotic versions of $\hat{\beta}_{1,n}$ and $\hat{\mu}_{X,n}$, respectively. Assume also that*

$$B_+(\hat{\beta}_1) = -B_-(\hat{\beta}_1) = B(\hat{\beta}_1)\,,$$
$$B_+(\hat{\mu}_X) = -B_-(\hat{\mu}_X) = B(\hat{\mu}_X)\,, \tag{22}$$

Let $r(m,v) = \int_{-\infty}^{+\infty} F_0(-m\,v + v\,z) g_0(z) dz$. *Then, a robust confidence interval for* β_0 *is*

$$\left[Z_{(k+1)}(\hat{\beta}_{1,n}), Z_{(n-k)}(\hat{\beta}_{1,n})\right)$$

where k *satisfies* $h\,(n,k,\tilde{p}_C) \geq 1-\alpha$, $h(n,k,p)$ *is given in* (11) *and* $\tilde{p}_C = (1-\epsilon)\,r(B_+(\hat{\mu}_X), B_+(\hat{\beta}_1))$.

4 Simulation study

In this section we present the results of a Monte Carlo study to investigate the finite sample coverage level and lengths of the robust confidence intervals for β_1 introduced in Section 3. We will also compare their performance with that of the intervals proposed in [1].

We generated 1000 random samples following model (1) with $\beta_0 = \beta_1 = 0$, $X_i \sim \mathcal{N}(0,1)$, $\epsilon_i \sim \mathcal{N}(0,1)$, and ϵ_i independent from X_i, $i = 1, \ldots, n$. We computed confidence intervals for β_1 of level 95% as defined in (19), with $\hat{\beta}_0$ as in (14). These samples were contaminated with 5, 10 and 20% of outliers with distribution $\mathcal{N}_2\left((x_0,y_0)', 0.01\mathbf{I}\right)$, where $(x_0,y_0) = (3,1.5)$, $(5,2.5)$ and $(5,15)$, which we call "mild", "medium" and "strong" contamination cases respectively.

For each contamination level, the maximum asymptotic bias of $\hat{\beta}_0$ can be computed from Table 1 and Theorem 7 of [2]. We include them in Table 3. The maximum asymptotic bias of $\hat{\mu}_{X,n} = \text{med}(X_1, \ldots, X_n)$ is

$$B(\hat{\mu}_X) = \max\left(G_0^{-1}\left(\frac{1}{2\,(1-\varepsilon)}\right) - G_0^{-1}\left(\frac{1}{2}\right), G_0^{-1}\left(\frac{1}{2}\right) - G_0^{-1}\left(\frac{1-2\varepsilon}{2\,(1-\varepsilon)}\right)\right)$$

When $G_0(x) = \Phi(x)$ the standard normal distribution, we have

$$B(\hat{\mu}_X) = \Phi^{-1}\left(\frac{1}{2\,(1-\varepsilon)}\right)$$

ϵ	0.01	0.025	0.05	0.10	0.15	0.20
$B(\hat{\beta}_0)$	0.013	0.033	0.073	0.170	0.311	0.518

Table 3
Maximum asymptotic biases for $\hat{\beta}_0$ based on the Repeated Median of Slopes estimator, when $(Y_i, X_i) \sim \mathcal{N}(0, I)$ under the model

Table 4 contains the coverages and median lengths for $\varepsilon = 0.05$, 0.10 and 0.20, for "mild", "medium" and "strong" contaminations. Comparing these results with those of Table 2, we can see that for $\varepsilon = 0.05$ the effect of the bias correction discussed in Section 3 is more visible for very large samples ($n = 10000$), while the median lengths are comparable for all n. When $\varepsilon = 0.10$ the interval in (19) has noticeably higher coverage than those in Table 2 for $n \geq 500$. Also in this case the median lengths are similar in both tables. Finally, while for $\varepsilon = 0.20$ the coverages in Table 2 are clearly falling below the nominal level for samples of size $n \geq 80$, those in Table 4 remain high for all sample sizes. Note that the correction in Section 3 is based on the maximum asymptotic bias, and hence, when the contamination present in the data do not correspond to the worst case scenario this robust confidence interval may be conservative. This is why the empirical coverages for the "mild" and "medium" cases in Table 4 are higher than the nominal level, while those for "strong" contaminations are closer to 95%. Furthermore, the impressive gains in coverage in Table 4 only seem to require a modest increase in median length. For example, for the "strong" case, with $n = 2000$, the empirical coverages increased by a factor of 48 from 0.02 to 0.97, while the median length changed by a factor of 1.5 (1.29 to 1.94).

Note that the median lengths of the robust confidence interval defined in (19) decrease as the sample size n increases, as expected, and also increase as the contamination ranges from "mild" to "strong". Finally, note that the median lengths of the robust confidence interval do not tend to zero as n increases. It has been well established in other proposals for robust confidence intervals which maintain the nominal level over the whole contamination neighbourhood, that their length remains bounded away from zero as the sample size increases (e.g. see [8] and [18]).

Tables 5 and 6 compare the performance of the robust confidence interval proposed here with those reported in Table 8 of [1]. For $\varepsilon = 0.05$ we see that the confidence interval defined in (19) has coverage levels much closer to the nominal one, with comparable or favourable median lengths. For $\varepsilon = 0.10$ the new proposal is both much closer to the nominal level and noticeably more stable than that of [1], while keeping the median lengths smaller in almost all cases.

5 Robust hypothesis tests

As in Section 2, consider $U_1, \ldots, U_n$ independent random variables satisfying $\text{med}(U_i) = \theta$, $i = 1, \ldots, n$, where $\theta \in \mathbb{R}$. We are interested in testing

$$H_0 : \theta \geq \theta_0 \qquad H_a : \theta < \theta_0$$

The classical sign test of approximate level α rejects H_0 if

$$\sum_{i=1}^{n} I\left(U_i > \theta_0\right) < k\,,$$

where $I(A) = 1$ if the event A is true, and 0 otherwise, and k satisfies

$$P\left(Bi(n, 1/2) < k\right) \approx \alpha\,.$$

This test rejects H_0 if

$$\theta_0 \notin \left(-\infty, U_{(n-k)}\right]\,. \tag{23}$$

When the observations may be contaminated, we need to control the level of the test over the whole contamination neighbourhood. In other words, we need k to satisfy

$$\sup_{H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})\,, \theta \geq \theta_0} P\left(\, Bi(n, p_{H,\theta}) < k \,\right) \;\leq\; \alpha\,,$$

where $p_{H,\theta} = P_H\left(U_i > \theta_0\right)$. By Lemma 1, this is equivalent to using k such that

$$P\left(Bi(n, \tilde{p}) < k\right) \leq \alpha\,, \tag{24}$$

where

$$\tilde{p} = \inf_{H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}}), \theta \geq \theta_0} p_{H\,\theta} \quad = \inf_{H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}}), \theta \geq \theta_0} P_H\left(U_i > \theta_0\right) .$$

It is easy to see that

$$\inf_{H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}}), \theta \geq \theta_0} P_H\left(U_i > \theta_0\right) = \inf_{H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})} P_H\left(U_i > \theta_0\right) .$$

Using the notation of Section 3, and under assumptions A1 and A2, a robust test for the hypothesis

$$H_0 : \beta_1 \geq \beta \qquad \text{versus} \qquad H_a : \beta_1 < \beta\,,$$

where β is a fixed constant and β_1 is the slope in the linear model (1), rejects H_0 if

$$\beta \notin \left(-\infty, W_{(n-k)}(\hat{\beta}_{0,n})\right] , \tag{25}$$

where $\hat{\beta}_{0,n}$ is an estimator of β_0 in (1), and k satisfies (24) with $\tilde{p} = \tilde{p}_S$ as in (21). This follows because, by (33), we have

$$\begin{aligned} \inf_{H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})} P_H\left(W(\hat{\beta}_0) > \beta\right) &= 1 - \sup_{H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})} P_H\left(W(\hat{\beta}_0) < \beta\right) \\ &= \inf_{H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})} P_H\left(W(\hat{\beta}_0) < \beta\right) = \tilde{p}_S\,. \end{aligned} \tag{26}$$

By Lemma 3 in the Appendix, for any $\alpha \leq 1/2$, we can find k such that (24) holds with $\tilde{p} = \tilde{p}_S$.

Consider now hypotheses of the form

$$H_0 : \theta \leq \theta_0 \qquad \text{versus} \qquad H_a : \theta > \theta_0\,.$$

The classical sign test of approximate level α rejects H_0 if

$$\sum_{i=1}^{n} I\left(U_i > \theta_0\right) > n - k\,,$$

and $n - k$ satisfies

$$P\left(Bi(n, 1/2) > n - k\right) \approx \alpha\,.$$

Equivalently, this test rejects H_0 if

$$\theta_0 \notin \left[U_{(k)}, +\infty\right) . \tag{27}$$

As before, under assumptions A1 and A2, a robust test for the slope β_1 in model (1) for the hypothesis

$$H_0 : \beta_1 \leq \beta \qquad \text{versus} \qquad H_a : \beta_1 > \beta \,,$$

rejects H_0 if

$$\beta \notin \left[W_{(k)}(\hat{\beta}_{0,n}), +\infty \right) \,, \tag{28}$$

where, by Lemma 3 and (26), it is enough that k satisfies

$$P\left(Bi(n, \tilde{p}) > n - k\right) \approx \alpha \,,$$

with $\tilde{p} = 1 - \tilde{p}_S$ in (21). This test has level α over the contamination neighbourhood.

6 Examples

6.1 Motorola

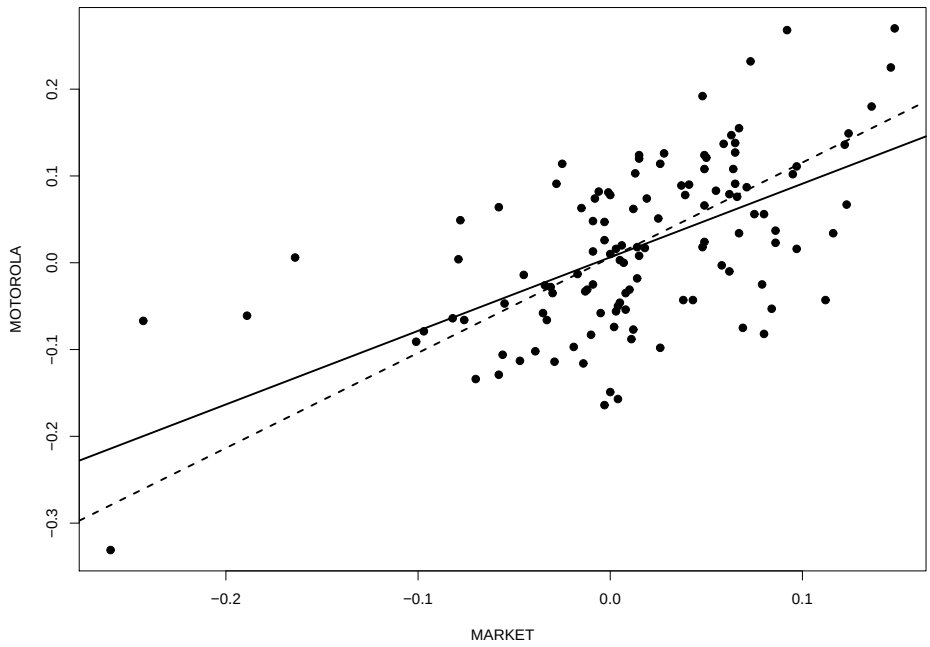
Financial economists measure the risk associated with investing in a particular stock comparing the returns of the stock with an index of the market return. In this example, let M_i, $i = 1, \ldots, 120$, be the monthly returns of Motorola shares between January 1978 and December 1987, and let m_i be the corresponding monthly market returns based on transactions of the New York Stock Exchange and the American Exchange. The regression model used to measure the risk of Motorola shares is

$$M_i - B_i = \beta_0 + \beta_1 \, (m_i - B_i) + \epsilon_i \,,$$

where B_i is the monthly return of 30-day US Treasury bills. These data were first published in [5]. The larger the slope β_1 in the above model, the riskier the stock. If one considers the point of view of a cautious investor that is interested in checking whether Motorola stocks are a safe investment, a hypothesis of interest is

$$H_0 : \beta_1 \leq 1 \qquad \text{versus} \qquad H_a : \beta_1 > 1 \,.$$

Fig. 1. Motorola stock returns. The solid line is the least squares fit, while the dotted line represents the repeated medians fit.



The least squares estimator $\hat{\beta}_n^{LS} = 0.85$ and the corresponding p-value is 0.925. The diagnostic plots do not reveal the presence of any outliers or departures from the model. Moreover, the Theil non-parametric test for the above hypothesis (e.g. see [9]) yields an approximate p-value of 0.16. In other words, these tests suggest that there is not enough evidence to claim that investing in Motorola's shares is riskier than the reference index. However, the fit based on a robust estimator such as the Repeated Median of Slopes identifies one observation with large standardized residual. The data together with the fitted lines are displayed in Figure 1. Hence, we are interested in testing the above hypotheses with a robust procedure to avoid the distortion that could be introduced by this outlier. We will use the rejection region in (28). Since the diagnostic plots suggest that there is one possible outlier out of 120 observations, we take $\varepsilon = 0.01$. From the corresponding entry in Table 3, and the fact that the maximum bias of the median is $\Phi^{-1}(1/[2\,(1-\varepsilon)])$, we obtain $\tilde{p} = 1 - \tilde{p}_S = 1 - 0.99(\Phi(0.013) + 0.50/0.99 - 2\Phi(0.013)\,0.50/0.99\,) = 0.5050519$.

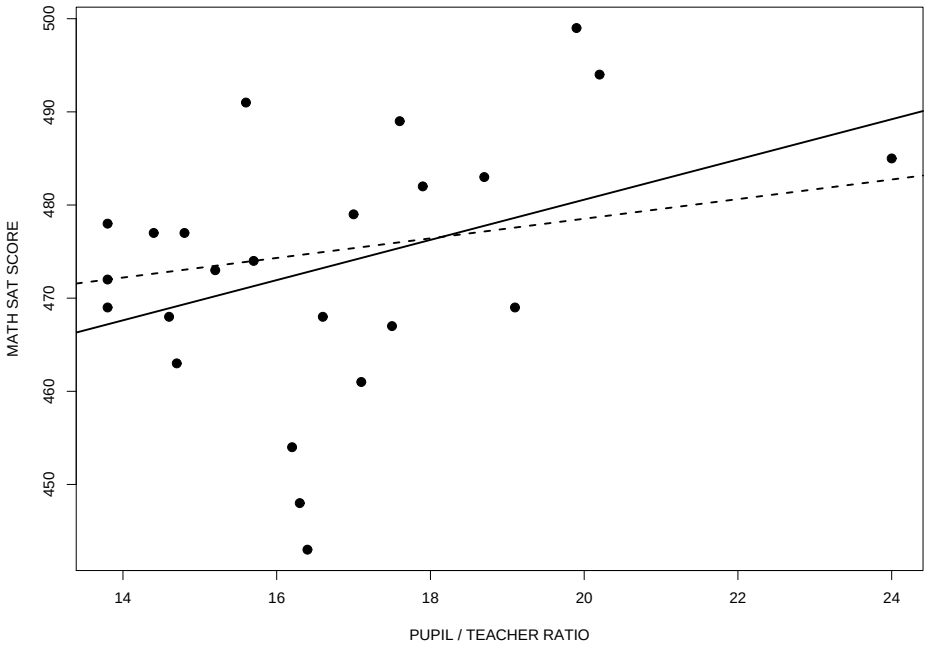
With an approximate level of $\alpha = 0.10$, we have $W_{(53)}(\hat{\beta}_{0,n}) = 1.0003$ and the rejection region $[1.0003, +\infty)$ barely misses 1. We conclude that the robust inferential procedure seems to indicate that there is some evidence showing that the Motorola stock is riskier than the market. This is in contrast with the result obtained by the classical test based on the least squares estimator. The robust procedure has been able to reveal the evidence against the null hypothesis that was hidden to the classical method by the outlier.

6.2 SAT scores versus student / teacher ratio

The SAT is a standardized test used for college admissions in the US. These scores are sometimes used as evidence for or against the effectiveness of public education policy measures at the state level. We consider data on average mathematics SAT scores and average student / teacher ratio in public elementary and secondary schools. These figures correspond to the academic year 1994-1995. It is well known that the average SAT score depends very strongly on the percentage of students taking the exam, with lower participation rates generally resulting in higher average scores. We restrict ourselves to the 24 states that reported participation rates of 30% or higher.

A simple regression of the average SAT scores in mathematics as a function of the average student / teacher ratio for each of the 50 US states yields an estimated slope of 2.16 with an associated standard error of 1.12. The robust repeated medians of slopes point estimator is notably smaller at 1.05. Both estimators seem to indicate that, against our intuition, higher pupil / teacher ratios result in higher mathematics SAT scores. Figure 2 shows the data and the LS and RMS regression estimates. We are interested in determining whether the linear association between these two measurements is indeed positive. The usual 90% and 95% confidence intervals for the slope based on the least squares estimator are (0.32, 3.99) and (-0.03, 4.35), respectively, indicating that there is some evidence supporting a rather counter-intuitive positive relationship between these variables. A residual analysis based on the

Fig. 2. Mathematics SAT scores versus student / teachers ratios. The solid line is the least squares fit, while the dashed line represents the repeated medians fit.



RMS estimate shows one potential outlier (with residual larger than 3 Median Absolute Deviation (MAD) of the residuals). Using a robust non-parametric confidence interval of the form (19) with $\alpha = 0.10$ and $\epsilon = 0.05$ yields the interval (-1.30, 4.72), which indicates that the slope coefficient may also be negative. In fact, a closer look at the data reveals that the LS fit may be affected by one state (California) with large SAT score and a particularly large student / teacher ratio. If we remove this state's observations from our analysis, the estimated SD of the LS slope estimator increases to 1.48, resulting in wider confidence intervals, and consequently yielding weaker evidence of a positive linear association between the variables. The 90% and 95% confidence intervals change to (0.13, 4.98) and (-0.34, 5.45), respectively. Our robust analysis was able to indicate that the apparent positive association was not significant once we take into account the possible model deviations of some of these observations.

7 Prospective work

Constructing robust prediction intervals is still considered an open problem with very few approaches appearing in the literature. One such proposal is [7]. The authors, however, do not take into account the lack of coverage caused by potentially asymmetric contamination. Rather, their proposal is akin to using robust prediction intervals of the form $\hat{\theta}_n \pm t_{(n-2)}(\alpha/2)\sqrt{1+1/n}\hat{\sigma}_n$, with robust location and scale estimates $\hat{\theta}_n$ and $\hat{\sigma}_n$. In what follows we sketch a different proposal to derive a robust prediction interval for a future observation of the response variable.

In the same spirit of model (1), we will focus on building a robust confidence interval for $\eta(x) = med(Y|\, X = x) = \beta_0 + \beta_1(x - \mu_X)$. Let I_1, I_2 and I_3 be robust confidence intervals for for β_0, β_1 and μ_X and consider

$$I_M = \left[\inf_{a,b,m \in I_1 \times I_2 \times I_3} a + b(x-m), \sup_{a,b,m \in I_1 \times I_2 \times I_3} a + b(x-m) \right].$$

Using the Bonferroni inequality we obtain

$$\begin{aligned} P(\eta(x) \in I_M) &\geq P(\beta_0 \in I_1, \beta_1 \in I_2, \mu_X \in I_3) \\ &\geq P(\beta_0 \in I_1) + P(\beta_1 \in I_2) + P(\mu_X \in I_2) - 2 \geq 1 - 3\alpha'. \end{aligned}$$

Recall that under the central model the errors satisfy $\sigma_0 \epsilon \sim F_0$ and let $Y_N(x)$ be the unknown response we want to predict at the point x. Note that $Y_N(x) = (1 - Z_N)Y_G + Z_N Y_B$ with $Z_N \sim Bi(1, \epsilon)$, Z_N is independent of both Y_G and Y_B.Then,

$$\begin{aligned} P(|Y_N(x) - \eta(x)| \leq t) &= P(|(1 - Z_N)\,(Y_G - \eta(x)) + Z_N\,(Y_B - \eta(x))| \leq t) \\ &= (1-\epsilon)P(|Y_G - \eta(x)| \leq t) + \epsilon P(|Y_B - \eta(x)| \leq t) \\ &\geq (1-\epsilon)\left[F_0\left(\frac{t}{\sigma_0}\right) - F_0\left(\frac{-t}{\sigma_0}\right)\right] \\ &= (1-\epsilon)\left[2F_0\left(\frac{t}{\sigma_0}\right) - 1\right] = (1-\epsilon)(1-\alpha''). \end{aligned}$$

Since $\sigma_0 F_0^{-1}(1 - \alpha''/2) \leq K(\epsilon)\widehat{\sigma}_n F_0^{-1}(1 - \alpha''/2) = t_n^*$ ($K(\epsilon)$ is defined in 3)

and

$$Y_N(x) = \eta(x) + Y_N(x) - \eta(x)$$

we have that

$$\begin{aligned} -t_n^* \leq Y_N(x) - \eta(x) \leq t_n^* \\ \inf_{a,b,m \in I_1 \times I_2 \times I_3} a + b(x-m) - t_n^* \leq \eta(x) - t_n^* \leq Y_N(x) \\ \leq t_n^* + \eta(x) \leq t_n^* + \sup_{a,b,m \in I_1 \times I_2 \times I_3} a + b(x-m). \end{aligned}$$

Thus, we take the robust prediction interval at $x = x_0$ as

$$I_P = \left[\inf_{a,b,m \in I_1 \times I_2 \times I_3} a + b(x_0 - m) - t_n^*, t_n^* + \sup_{a,b,m \in I_1 \times I_2 \times I_3} a + b(x_0 - m) \right],$$

and the coverage level is given by

$$\begin{aligned} P(Y_N(x) \in I_P) &\geq P(\eta \in I_M, Y_N(x) - \eta(x) \in [-t_n^*, t_n^*]) \\ &= P(\eta \in I_M) P(Y_N(x) - \eta(x) \in [-t_n^*, t_n^*]) = \\ &= (1 - 3\alpha')(1 - \epsilon)(1 - \alpha''). \end{aligned}$$

This interval will, with high probability, contain future observations at the point x that come from the central regression model. However, future response values that do not follow the model may not fall in the interval. In other words, the interval is meant to predict only "good" future observations. Further work along these initial steps is still necessary.

8 Conclusion

It is easy to see that when the data may contain outliers or other departures from the assumed model, classical inference methods can be seriously affected and might yield confidence levels much lower than the nominal values. In this paper we propose robust confidence intervals for the slope of simple linear regression models. These intervals combine robust non-parametric confidence intervals for location models with bias corrections to control the minimum coverage level even in the case of contaminated samples.

Our approach can be applied to any consistent estimator of the slope and intercept for which maximum bias curves are tabulated. Earlier proposals in the literature (see [1]) required estimators that are $\sqrt{n}$-normal over the entire contamination neighbourhood, and also involved the estimation of bias bounds, which introduces further variability in the confidence interval, affecting their coverage levels. In addition, note that to use the approach discussed in this paper one does not need to estimate neither the scale parameter of the errors σ_0, nor the asymptotic standard deviation of the regression estimator.

Although our derivation is asymptotic in nature, our simulation studies suggest that this approach works well for small samples. In particular, note from Tables 5 and 6 that these new robust confidence intervals maintain coverage levels much closer to the nominal one than previous proposals without sacrificing length. Furthermore, in most cases in these tables, the new approach yields higher coverage levels with shorter intervals.

Finally, we extend these ideas to the hypothesis testing setup and derive robust procedures that maintain the level of the test over the whole contamination neighbourhood.

Acknowledgements: We would like to thank two anonymous referees and the Associate Editor for their comments suggestions that have resulted in a much improved paper.

9 Appendix

Proof of Theorem 1:

$$
\begin{aligned}
P_{H_0}\left(\frac{Y-\hat{\beta}_0}{X-\hat{\mu}_X}-\beta_1<0\right) &= P_{H_0}\left(\frac{\sigma_0\epsilon-b(\hat{\beta}_0)}{X-\mu_X-b(\hat{\mu}_X)}<0\right)\\
&= P_{H_0}\left(\sigma_0\epsilon-b(\hat{\beta}_0)<0 \text{ and } X-\mu_X-b(\hat{\mu}_X)\geq 0\right)\\
&\quad + P_{H_0}\left(\sigma_0\epsilon-b(\hat{\beta}_0)\geq 0 \text{ and } X-\mu_X-b(\hat{\mu}_X)<0\right)\\
&= P_{F_0}\left(\epsilon<b(\hat{\beta}_0)/\sigma_0\right)P_{G_0}\left((X-\mu_X)/\sigma_X\geq b(\hat{\mu}_X)/\sigma_X\right)\\
&\quad + P_{F_0}\left(\epsilon\geq b(\hat{\beta}_0)/\sigma_0\right)P_{G_0}\left((X-\mu_X)/\sigma_X< b(\hat{\mu}_X)/\sigma_X\right)\\
&= F_0\left(b(\hat{\beta}_0)/\sigma_0\right)\left[1-G_0^-\left(b(\hat{\mu}_X)/\sigma_X\right)\right]+\left[1-F_0\left(b(\hat{\beta}_0)/\sigma_0\right)\right]G_0^-\left(b(\hat{\mu}_X)/\sigma_X\right)\\
&= F_0\left(b(\hat{\beta}_0)/\sigma_0\right)+G_0^-\left(b(\hat{\mu}_X)/\sigma_X\right)-2\,F_0\left(b(\hat{\beta}_0)/\sigma_0\right)G_0^-\left(b(\hat{\mu}_X)/\sigma_X\right)\\
&= \left(1-2\,G_0^-\left(b(\hat{\mu}_X)/\sigma_X\right)\right)F_0\left(b(\hat{\beta}_0)/\sigma_0\right)+G_0^-\left(b(\hat{\mu}_X)/\sigma_X\right).
\end{aligned}
$$

Note that when $G_0^-\left(b(\hat{\mu}_X)/\sigma_X\right) > 1/2$ the above quantity is non-increasing in $b(\hat{\beta}_0)$, and if $G_0^-\left(b(\hat{\mu}_X)/\sigma_X\right) < 1/2$ it is non-decreasing in $b(\hat{\beta}_0)$. Thus, we have that if the bias of $\hat{\mu}_X$ is positive then

$$
\begin{aligned}
P_{H_0}\left(\frac{Y-\hat{\beta}_0}{X-\hat{\mu}_X}-\beta_1<0\right) \geq\ & F_0\left(B_+(\hat{\beta}_0)\right)+G_0^-\left(B_+(\hat{\mu}_X)\right)\\
& -2\,F_0\left(B_+(\hat{\beta}_0)\right)G_0^-\left(B_+(\hat{\mu}_X)\right). \quad (29)
\end{aligned}
$$

If the bias of $\hat{\mu}_X$ is negative

$$
\begin{aligned}
P_{H_0}\left(\frac{Y-\hat{\beta}_0}{X-\hat{\mu}_X}-\beta_1<0\right) \geq\ & F_0\left(B_-(\hat{\beta}_0)\right)+G_0^-\left(B_-(\hat{\mu}_X)\right)\\
& -2\,F_0\left(B_-(\hat{\beta}_0)\right)G_0^-\left(B_-(\hat{\mu}_X)\right), \quad (30)
\end{aligned}
$$

A similar argument shows that if the bias of $\hat{\mu}_X$ is positive

$$
\begin{aligned}
P_{H_0}\left(\frac{Y-\hat{\beta}_0}{X-\hat{\mu}_X}-\beta_1<0\right) \leq\ & F_0\left(B_-(\hat{\beta}_0)\right)+G_0^-\left(B_+(\hat{\mu}_X)\right)\\
& -2\,F_0\left(B_-(\hat{\beta}_0)\right)G_0^-\left(B_+(\hat{\mu}_X)\right), \quad (31)
\end{aligned}
$$

while if the bias of $\hat{\mu}_X$ is negative

$$P_{H_0}\left(\frac{Y-\hat{\beta}_0}{X-\hat{\mu}_X}-\beta_1<0\right)\leq F_0\left(B_+(\hat{\beta}_0)\right)+G_0^-\left(B_-(\hat{\mu}_X)\right) \\ -2\,F_0\left(B_+(\hat{\beta}_0)\right)\,G_0^-\left(B_-(\hat{\mu}_X)\right)\,. \quad (32)$$

Since the lower bounds in (29) and (30) are equal, and so are the upper bounds in (31) and (32) we have

$$\begin{aligned}&(1-\epsilon)\,\left[F_0\left(B(\hat{\beta}_0)\right)+G_0^-\left(B(\hat{\mu}_X)\right)-2\,F_0\left(B(\hat{\beta}_0)\right)\,G_0^-\left(B(\hat{\mu}_X)\right)\right]\\ \leq\, &P_H\left(\frac{Y-\hat{\beta}_0}{X-\hat{\mu}_X}-\beta_1<0\right)\\ \leq\, &(1-\epsilon)\left[1-F_0(B(\hat{\beta}_0))-G_0^-\left(B(\hat{\mu}_X)\right)+2\,F_0(B(\hat{\beta}_0))\,G_0^-\left(B(\hat{\mu}_X)\right)\right]+\epsilon\,.\end{aligned}$$

Note that the lower bound in the above equation is 1 minus the upper bound, in other words, in this case we have

$$\tilde{p}_S\leq P_H\left(\frac{Y-\hat{\beta}_0}{X-\hat{\mu}_X}-\beta_1<0\right)\leq 1-\tilde{p}_S\quad\text{for all } H\in\mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})\,, \qquad (33)$$

with $\tilde{p}_S$ as in 21.

Proof of Theorem 2: We have

$$\begin{aligned}P_{H_0}\left((\beta_1-\hat{\beta}_1)\,(X-\mu_X)+\sigma_0\epsilon<0\right)&=P_{H_0}\left(-b(\hat{\beta}_1)\,(X-\mu_X-b(\hat{\mu}_X))+\sigma_0\epsilon<0\right)\\ &=P_{H_0}\left(\sigma_0\epsilon<-b(\hat{\mu}_X)\,b(\hat{\beta}_1)+b(\hat{\beta}_1)\,(X-\mu_X)\right)\\ &=P_{H_0}\left(\epsilon<-\frac{b(\hat{\mu}_X)}{\sigma_X}\,\frac{b(\hat{\beta}_1)\,\sigma_X}{\sigma_0}+\frac{b(\hat{\beta}_1)\,\sigma_X}{\sigma_0}\,\frac{(X-\mu_X)}{\sigma_X}\right)\\ &=P_{H_0}\left(\epsilon<-m\,v+v\,\frac{(X-\mu_X)}{\sigma_X}\right)\\ &=\int_{-\infty}^{+\infty}F_0(-m\,v+v\,z)g_0(z)dz=r(m,v)\,,\end{aligned}$$

where $m=b(\hat{\mu}_X)/\sigma_X$ and $v=b(\hat{\beta}_1)\,\sigma_X/\sigma_0$. The partial derivatives of $r(m,v)$ are

$$\frac{\partial r}{\partial m}(m,v) = -\int f_0(v\,z - m\,v)\,v\,g_0(z)dz\,,$$

$$\frac{\partial r}{\partial v}(m,v) = \int f_0(v\,z - m\,v)\,(z-m)\,g_0(z)dz\,.$$

It is easy to see that if $m > 0$ ($m < 0$) then $r(m,v)$ is decreasing (increasing) in v. Also, $r(m,v)$ is decreasing (increasing) in m if $v > 0$ ($v < 0$). It follows that

$$\begin{aligned}\min\Big(r(B_+(\hat{\mu}_X), B_+(\hat{\beta}_1)), r(B_-(\hat{\mu}_X), B_-(\hat{\beta}_1)),\Big) \\ \leq P_{H_0}\Big((\beta_1 - \hat{\beta}_1)\,(X - \mu_X) + \epsilon < 0\Big) \\ \leq \max\Big(r(B_+(\hat{\mu}_X), B_-(\hat{\beta}_1)), r(B_-(\hat{\mu}_X), B_+(\hat{\beta}_1)),\Big)\,.\end{aligned}$$

then it holds that

$$r(B_+(\hat{\mu}_X), B_+(\hat{\beta}_1)) = r(B_-(\hat{\mu}_X), B_-(\hat{\beta}_1))$$

$$r(B_+(\hat{\mu}_X), B_-(\hat{\beta}_1)) = r(B_-(\hat{\mu}_X), B_+(\hat{\beta}_1))\,.$$

It is easy to see that

$$1 - r(B_+(\hat{\mu}_X), B_+(\hat{\beta}_1)) = r(B_-(\hat{\mu}_X), B_+(\hat{\beta}_1))\,,$$

which implies that

$$1 - \Big[(1-\epsilon)\,r(B_+(\hat{\mu}_X), B_+(\hat{\beta}_1))\Big] = (1-\epsilon)\,r(B_-(\hat{\mu}_X), B_+(\hat{\beta}_1)) + \epsilon\,.$$

Thus,

$$\tilde{p}_C \leq P_H\left(Z_i(\hat{\beta}_1) < \beta_0\right) \leq 1 - \tilde{p}_C \quad \text{for all } H \in \mathcal{H}_\varepsilon(H_{\boldsymbol{\theta}})\,,$$

where $\tilde{p}_C = (1-\epsilon)\,r(B_+(\hat{\mu}_X), B_+(\hat{\beta}_1))$.

Lemma 3 *Let $Z \sim Bi(n,p)$ with $0 \leq p \leq 1$ and $n \geq 1$. Then, for all $k = 0, 1, \ldots, n$, $h_k(p) = P(Z \leq k)$ is non-increasing in p.*

Proof: If $X \sim Bi(n, p_1)$ and $Y \sim Bi(n, p_2)$, with $p_1 < p_2$, then Y is larger than X in the usual stochastic order (see Example 1.A.25, p.14 in [15]).

Lemma 4 *Let U and V be independent random variables defined on the probability space $(\Omega, \mathcal{A}, P)$. Take $Q : \Omega \to \mathbb{R}$ such that*

$$Q = \begin{cases} U/V \ \textit{if} \ V \neq 0 \\ \\ Z \quad \textit{if} \, V = 0, \end{cases}$$

with Z any random variable such that 0 is a median of Z and Z is independent of V. (i) If 0 is a median of U then it is also a median of Q. (ii) If $Z = 0$ and 0 is a median of V then it is also a median of Q,

Proof: (i) Note that

$$\begin{aligned} & P\left(Q \leq 0\right) \\ & = P\left(\left[U \leq 0\right] \cap \left[V > 0\right]\right) + P\left(\left[U \geq 0\right] \cap \left[V < 0\right]\right) + P\left(\left[Z \leq 0\right] \cap \left[V = 0\right]\right) \\ & \qquad = E_V\Big[I_{(0,\infty)}(V)\, P\left(U \leq 0 \middle| \, V\right)\Big] + E_V\Big[I_{(-\infty,0)}(V)\, P\left(U \geq 0 \middle| \, V\right)\Big] \\ & \qquad\qquad + P\left(Z \leq 0 \middle| \, V = 0\right) P(V = 0)\,. \end{aligned}$$

$$\begin{aligned} & P\left(Q \geq 0\right) \\ & = P\left(\left[U \geq 0\right] \cap \left[V > 0\right]\right) + P\left(\left[U \leq 0\right] \cap \left[V < 0\right]\right) + P\left(\left[Z \geq 0\right] \cap \left[V = 0\right]\right) \\ & \qquad = E_V\Big[I_{(0,\infty)}(V)\, P\left(U \geq 0 \middle| \, V\right)\Big] + E_V\Big[I_{(-\infty,0)}(V)\, P\left(U \leq 0 \middle| \, V\right)\Big] \\ & \qquad\qquad + P\left(Z \geq 0 \middle| \, V = 0\right) P(V = 0)\,. \end{aligned}$$

If 0 is a median of U, we have $P(U \leq 0) \geq 1/2$ and $P(U \geq 0) \geq 1/2$. This together with the independence between U and V and Z and V imply the desired result.

(ii) We have

$$\begin{aligned} & P\left(Q > 0\right) \\ & = P\left(\left[V > 0\right] \cap \left[U > 0\right]\right) + P\left(\left[V < 0\right] \cap \left[U < 0\right]\right) + P\left(\left[V = 0\right] \cap \left[Z > 0\right]\right) \\ & \qquad = E_U\Big[I_{(0,\infty)}(U)\, P\left(V > 0 \middle| \, U\right)\Big] + E_U\Big[I_{(-\infty,0)}(U)\, P\left(V < 0 \middle| \, U\right)\Big]\,. \end{aligned}$$

$$
\begin{aligned}
&P\left(Q<0\right)\\
&\quad=P\left(\left[U<0\right]\cap\left[V>0\right]\right)+P\left(\left[U>0\right]\cap\left[V<0\right]\right)+P\left(\left[V=0\right]\cap\left[Z<0\right]\right)\\
&\qquad=E_U\Big[I_{(-\infty,0)}(U)\,P\left(V>0|\,U\right)\Big]+E_U\Big[I_{(0,\infty)}(U)\,P\left(V<0|\,U\right)\Big].
\end{aligned}
$$

If 0 is a median of V, we have $P(V<0)\leq 1/2$ and $P(V>0)\leq 1/2$. This together with the independence between U and V imply the result.

References

[1] J. Adrover, M. Salibian-Barrera, and R. Zamar. Globally robust inference for the location and simple linear regression models. *Journal of Statistical Planning and Inference*, 119:353–375, 2004.

[2] J. Adrover and R. Zamar. Bias robustness of three median-based regression estimates. Technical Report 194, Department of Statistics, University of British Columbia, Canada, 2000.

[3] J. Adrover and R. Zamar. Bias robustness of three median-based regression estimates. *Journal of Statistical Planning and Inference*, 122:203–227, 2004.

[4] B. Barnett and T. Lewis. *Outliers in statistical data*. Wiley & Sons, New York, 1994.

[5] E. Berndt. *The practice of econometrics: classic and contemporary*. Addison-Wesley Publishing Company, New York, 1994.

[6] W. Dixon and J. Tukey. Approximate behavior of the distribution of winsorized t (trimming/winsorization). *Technometrics*, 10:83–98, 1968.

[7] A. Fisher and P. Horn. Robust prediction intervals in a regression setting. *Computational Statistics and Data Analysis*, 17:129–140, 1994.

[8] R. Fraiman, V. Yohai, and R. Zamar. Optimal robust m–estimates of location. *Annals of Statistics*, 29, 2001.

[9] M. Hollander and D. Wolfe. *Nonparametric statistical methods*. John Wiley and Sons, New York, 1999.

[10] P. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

[11] P. Huber. Robust confidence limits. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 10:269–278, 1968.

[12] P. Huber. Studentizing robust estimates. In M. L. Puri, editor, *Nonparametric Techniques in Statistical Inference*, Cambridge, England, 1970. Cambridge University Press.

[13] R. D. Martin, V. J. Yohai, and R. H. Zamar. Min–max bias robust regression. *Annals of Statistics*, 17:1608–1630, 1989.

[14] R. D. Martin and R. H. Zamar. Bias-robust estimates of scale. *Annals of Statistics*, 21:991–1017, 1993.

[15] M. Shaked and J. Shanthikumar. *Stochastic Orders*. Springer, New York, 2007.

[16] A. Siegel. Robust regression using repeated medians. *Biometrika*, 69(1):242–244, 1982.

[17] J. Tukey and D. McLaughlin. Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/winsorization. *Sankhyã*, A 25:331–352, 1963.

[18] V. Yohai and R. Zamar. Robust non-parametric inference for the median. *Annals of Statistics*, 32(5):1841–1857, 2004.

Perc. of cont.	Sample size	Mild cont.	Medium cont.	Strong cont.
5%	20	0.95 (1.61)	0.95 (1.62)	0.95 (1.91)
	40	0.94 (1.00)	0.94 (1.00)	0.94 (1.16)
	60	0.95 (0.88)	0.95 (0.88)	0.95 (1.01)
	80	0.95 (0.73)	0.95 (0.74)	0.95 (0.80)
	100	0.95 (0.72)	0.96 (0.72)	0.95 (0.80)
	200	0.94 (0.54)	0.94 (0.54)	0.95 (0.55)
	500	0.94 (0.42)	0.94 (0.42)	0.94 (0.42)
	2000	0.96 (0.30)	0.96 (0.30)	0.96 (0.30)
	10000	0.94 (0.23)	0.94 (0.23)	0.94 (0.23)
10%	20	0.92 (1.50)	0.92 (1.49)	0.95 (2.19)
	40	0.96 (1.10)	0.96 (1.10)	0.95 (1.57)
	60	0.94 (0.91)	0.95 (0.91)	0.94 (1.31)
	80	0.95 (0.80)	0.95 (0.81)	0.95 (1.12)
	100	0.94 (0.76)	0.94 (0.77)	0.94 (1.05)
	200	0.95 (0.65)	0.95 (0.67)	0.95 (0.85)
	500	0.96 (0.57)	0.95 (0.58)	0.95 (0.69)
	2000	0.96 (0.50)	0.96 (0.51)	0.96 (0.55)
	10000	0.96 (0.44)	0.96 (0.46)	0.96 (0.48)
20%	20	0.98 (2.78)	0.98 (2.75)	0.98 (4.38)
	40	0.98 (1.52)	0.98 (1.53)	0.96 (3.63)
	60	0.98 (1.29)	0.98 (1.29)	0.97 (3.47)
	80	0.99 (1.14)	0.99 (1.14)	0.97 (3.36)
	100	0.97 (1.01)	0.98 (1.03)	0.96 (3.21)
	200	1.00 (0.84)	0.99 (0.85)	0.98 (2.95)
	500	1.00 (0.68)	0.99 (0.67)	0.96 (2.32)
	2000	1.00 (0.58)	1.00 (0.57)	0.97 (1.94)
	10000	1.00 (0.54)	1.00 (0.53)	0.96 (1.76)

Table 4
Empirical coverages and median lengths of 1000 confidence intervals of the form (19) for the slope parameter β_1. The nominal level is 95% and different proportion of outliers were placed at: $x_0 = 3, y_0 = 1.5$ (Mild), $x_0 = 5, y_0 = 2.5$ (Medium), and $x_0 = 5, y_0 = 15$ (Strong).

Perc. of cont.	Sample size	Mild cont.	Medium cont.	Strong cont.
5%	20	0.94 (1.41)	0.94 (1.42)	0.91 (1.42)
	40	0.92 (1.01)	0.92 (1.01)	0.93 (1.04)
	60	0.92 (0.86)	0.92 (0.86)	0.94 (0.90)
BC	80	0.92 (0.76)	0.92 (0.77)	0.94 (0.79)
	100	0.92 (0.71)	0.92 (0.71)	0.94 (0.73)
	200	0.95 (0.57)	0.95 (0.58)	0.96 (0.58)
5%	20	0.95 (1.61)	0.95 (1.62)	0.95 (1.91)
	40	0.94 (1.00)	0.94 (1.00)	0.95 (1.16)
	60	0.95 (0.88)	0.95 (0.88)	0.95 (1.01)
PI	80	0.95 (0.73)	0.95 (0.74)	0.95 (0.80)
	100	0.95 (0.72)	0.95 (0.72)	0.95 (0.80)
	200	0.94 (0.54)	0.94 (0.54)	0.95 (0.55)

Table 5

Comparison of the robust confidence intervals for the slope in [1] ("BC") with those of the form (19) ("PI"). The entries in the table are empirical coverages and median lengths of 1000 confidence intervals for the slope parameter β_1. The nominal level is 95% and 5% of outliers were placed at: $x_0 = 3, y_0 = 1.5$ (Mild), $x_0 = 5, y_0 = 2.5$ (Medium), and $x_0 = 5, y_0 = 15$ (Strong).

Perc. of cont.	Sample size	Mild cont.	Medium cont.	Strong cont.
10%	20	0.95 (1.54)	0.95 (1.56)	0.95 (1.79)
	40	0.89 (1.17)	0.87 (1.18)	0.96 (1.43)
	60	0.87 (1.05)	0.85 (1.08)	0.97 (1.28)
BC	80	0.86 (1.00)	0.85 (1.04)	0.98 (1.18)
	100	0.87 (0.95)	0.86 (1.00)	0.98 (1.10)
	200	0.91 (0.83)	0.92 (0.89)	0.99 (0.95)
10%	20	0.92 (1.50)	0.92 (1.49)	0.95 (2.19)
	40	0.95 (1.10)	0.95 (1.10)	0.95 (1.57)
	60	0.94 (0.91)	0.95 (0.91)	0.94 (1.31)
PI	80	0.95 (0.80)	0.95 (0.81)	0.95 (1.12)
	100	0.94 (0.76)	0.94 (0.77)	0.94 (1.05)
	200	0.95 (0.65)	0.95 (0.67)	0.95 (0.85)

Table 6

Comparison of the robust confidence intervals for the slope in [1] ("BC") with those of the form (19) ("PI"). The entries in the table are empirical coverages and median lengths of 1000 confidence intervals for the slope parameter β_1. The nominal level is 95% and 10% of outliers were placed at: $x_0 = 3, y_0 = 1.5$ (Mild), $x_0 = 5, y_0 = 2.5$ (Medium), and $x_0 = 5, y_0 = 15$ (Strong).