Globally Robust Inference for Simple Linear Regression Models with Repeated Median Slope Estimator

by

Md Jafar Ahmed Khan

M.Sc., University of Dhaka, 1992

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE STUDIES

(Department of Statistics)

We accept this thesis as conforming

to the required standard

The University of British Columbia

September 2002

Abstract

Globally robust inference takes into account the potential bias of the point estimates (Adrover, Salibian-Barrera and Zamar, 2002). To construct robust confidence intervals for the simple linear regression slope, the authors selected the generalized median of slopes (GMS) as their point estimate, considering its good bias behavior and asymptotic normality. However, GMS has a breakdown point of only 0.25, its asymptotic normality is established under very restrictive conditions, and its bias bound is known only for symmetric carrier distributions.

In this study, we propose the repeated median slope (RMS) estimate as an alternative choice. RMS has a breakdown point of 0.50, its asymptotic normality holds under mild assumptions, and the bias bound for RMS is known for general carrier distributions. The proposed method achieves, more or less, the same observed coverage levels while it constructs intervals of smaller lengths, as compared to the GMS approach.

Contents

Acknowledgements

I would like to acknowledge, warmly and thankfully, the guidance and support of my supervisor, Dr. Ruben H. Zamar, throughout this study. I would like to give many thanks to Dr. Lang Wu for his valuable suggestions as the second reader of this thesis.

Special thanks are to Dr. Matias Salibian-Barrera, who kindly gave me some related codes to start with, and to Dr. Harry Joe, for his timely and valuable suggestions during the computation.

Jafar Ahmed Khan

The University of British Columbia
September 2002

Chapter 1

Introduction

1.1 Motivation

As a major source of uncertainty, sampling variability of an estimator plays an important role in statistical inference, particularly when the sample size is small. As sample size increases, sampling variability becomes less and less important. This is because standard errors are usually of order $O(1/\sqrt{n})$, and tend to zero as sample size n tends to infinity.

This is not the case with the other major source of uncertainty: the bias of an estimator, which is caused by data contamination (e.g., outliers, asymmetric errors, and other departures from the model assumption), gross errors, etc. Since biases are of order $O(1)$, we cannot reduce the bias of an estimator by increasing the sample size. Therefore, for large data sets, the uncertainty due to bias clearly dominates the uncertainty due to sampling variability.

To demonstrate the idea, we perform a simulation. We choose $N(0,1)$ as our central model and $N(10,1)$ as the contaminating distribution. We select three different sample sizes: 25, 100 and 400, and three different levels of contamination: 0%, 10% and 20%. For each combination of sample size and contamination level, we generate 1000 samples, and calculate the median from each sample. The average of the 1000 sample medians (minus the true median which is zero in this case) gives us the bias, and we also calculate the standard error in each case. The following table summarizes the results.

Table 1.1: Bias (standard error) for different sample sizes and contamination levels.

Contamination level	$n = 25$	$n = 100$	$n = 400$
0%	-0.0028 (0.2415)	-0.0039 (0.1238)	0.0028 (0.0610)
10%	0.1355 (0.2853)	0.1435 (0.1383)	0.1381 (0.0679)
20%	0.3278 (0.3390)	0.3188 (0.1645)	0.3194 (0.0831)

As expected, the standard error of the sample median decreases as n increases while the bias remains almost unchanged. Also, the bias increases with increased contamination. This shows the relative importance of bias as a source of uncertainty of an estimator. With the availability of high-speed computers and automated data collection techniques, there are many applications in which data quantity is not a problem, and we should concentrate on data quality. Accordingly, we should have more focus on the bias behavior of an estimator than on its standard error.

Classical inference considers sampling variability to be the only source of uncertainty, and does not address the issue of bias caused by contamination. With one or two exceptions, the few papers that address robust inference also ignore the possible bias of the point estimates. The consequences of this will be explained in Chapter 2 in the context of location and simple linear regression models. In our study, we will consider the bias-uncertainty of an estimator in addition to its standard error.

1.2 Terminology

To introduce some terminology, let us consider the parametric family F_θ and the ϵ-*contamination neighborhood* (Tukey, 1960),

$$\mathcal{F}_\epsilon(F_\theta) = \{F : F = (1-\epsilon)F_\theta + \epsilon F^*,\ F^* \text{ an arbitrary distribution}\},\quad 0 < \epsilon < 1/2. \quad (1.1)$$

According to this model, $100(1-\epsilon)\%$ of the data comes from a distribution F_θ but the remainder $100\epsilon\%$ may come from an unknown arbitrary distribution F^*. Let us consider estimates T_n of θ which, under some mild regularity conditions, converge almost surely to the asymptotic functional $T(F)$. This functional is well-defined on a set of distributions which includes the empirical distribution functions and the family $\mathcal{F}_\epsilon(F_\theta)$. Due to the presence of outliers and other departures from the central parametric model, $T(F)$ is not necessarily equal to θ for all F in $\mathcal{F}_\epsilon(F_\theta)$. For this reason, we must consider the asymptotic bias,

$$b_T(F) = d(T(F), \theta),$$

where d measures the distance between the true value of the parameter and the asymptotic value of the estimate.

Maxbias function

The robustness of an estimate T can be assessed in terms of the maxbias function,

$$B_T(\epsilon) = \sup_{F \in \mathcal{F}_\epsilon(F_\theta)} b_T(F) \quad (1.2)$$

which represents the maximum possible perturbation of $T(F)$ when F varies over the entire neighborhood. Huber (1964 & 1981) introduced the concept of maximum asymptotic bias in the location model setup. Martin and Zamar (1989) and Martin, Yohai and Zamar (1989) defined the maxbias curve which is a plot of $B_T(\epsilon)$ against different values of ϵ, and derived such curves for robust location, scale and regression estimates. It is known that the typical maxbias functions are continuous and increasing from zero to infinity.

Contamination sensitivity

Hampel denoted the supremum of the influence function by γ^*, and called it the gross-error sensitivity. The contamination sensitivity, introduced by He and Simpson (1993), is closely related to the gross-error sensitivity and is defined as

$$\gamma(T) = B_T'(\epsilon)|_{\epsilon=0}. \tag{1.3}$$

The authors showed that in general $\gamma^* \leq \gamma$.

Breakdown point

The breakdown point (BP) of the estimate T is the supremum of the maxbias function domain, i.e., the point at which the maxbias function diverges. Mathematically,

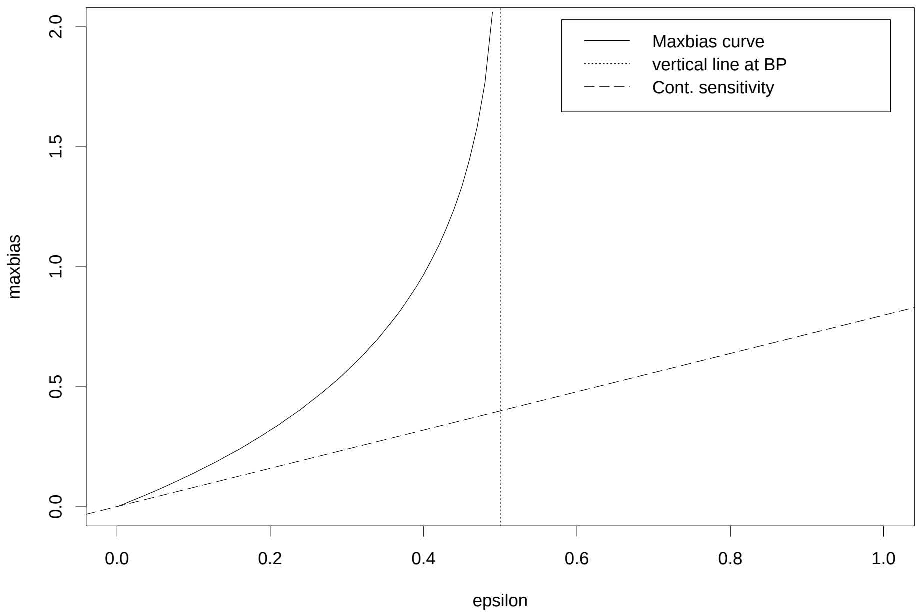
$$\epsilon^* = \sup\{\epsilon : B_T(\epsilon) < \infty\}. \tag{1.4}$$

The breakdown point is the maximum fraction of contamination the estimate can tolerate before its value is completely determined by the contaminating data.

The study of the limiting behavior of $B_T(\epsilon)$ near the ends of its domain is very informative. Hampel (1974) investigated the behavior of $B_T(\epsilon)$ when ϵ is small, focusing on the rate at which $B_T(\epsilon)$ tends to zero as ϵ tends to zero. Berrendero and Zamar (2001) studied the explosion rate of the bias function when ϵ approaches the breakdown point.

Figure 1.1 shows the maxbias curve, breakdown point and contamination sensitivity of median. The dotted vertical line at BP (0.5 in this case) shows how the maxbias

Figure 1.1: Maxbias curve, breakdown point and contamination sensitivity of median.



curve diverges when ϵ approaches BP. The slope of the dashed line is the contamination sensitivity, γ, of median. Notice that the linear approximation is good for small values of ϵ ($\epsilon \leq 0.10$).

Locally and globally robust estimates

An estimate whose maxbias grows slowly near zero is called locally robust and an estimate whose maxbias is relatively small for large fractions of contaminations is called globally robust.

The local and global features of the maxbias function $B_T(\epsilon)$ are summarized by the contamination sensitivity $\gamma(T)$ and the breakdown point ϵ^*. The contamination sensitivity provides an approximation for $B_T(\epsilon)$ near zero, and thus measures the local robustness of the estimate T. We say that T is lacally robust if $\gamma(T)$ is finite.

He and Simpson (1993) showed that $\epsilon^* \leq 0.5$ for all affine-equivariant regression estimates. (Equivariance properties of regression estimates will be discussed in the next section.) If the estimate T attains the maximal breakdown point 0.5, we say that T is globally robust.

Bias bound

The bias bound for the estimate T, introduced by Berrendero and Zamar (2001), highlights the practical potential of maxbias curves. To fix ideas, let us consider the location model and the median functional $M(F)$. Suppose that we want to determine the upper bound for the absolute difference $D_M(F) = |M(F) - \theta|$. Huber (1964) showed that

$$\left| \frac{M(F) - \theta}{\sigma_0} \right| \leq F_0^{-1} \left(\frac{1}{2(1-\epsilon)} \right) \doteq B_M(\epsilon)$$

where σ_0 is the scale of the core distribution. Therefore, $D_M(F)$ is bounded by $\sigma_0 B(\epsilon)$. In practice σ_0 is not usually known and has to be estimated by a robust scale functional $S(F)$, for example, the MAD. However, the quantity $S(F)B_M(\epsilon)$ may not be an upper bound for $D(F)$ because $S(F)$ may underestimate σ_0. For example, let us consider the contaminated distribution $F = 0.90\ N(0,1) + 0.10\ \delta_{0.15}$, then

$$B_M(0.10) = \left| \frac{M(F) - 0}{1} \right| = 0.1397,$$

whereas $MAD(F)B(0.10) = 0.8818 \times 0.1397 = 0.1232 < 0.1397$. A quantity $K_M(\epsilon)$ such that $S(F)K_M(\epsilon)$ is a bound for $D_M(F)$ is called the bias bound for $M(F)$. We will discuss the relation between maxbias functions and bias bounds for the location and the regression estimates in Chapter 2.

A confidence interval is called *globally robust* if it is *stable* (in the sense of keeping coverage at or above the nominal level) and *informative* (in the sense of keeping a reasonable average length), not only at the central model, but also over the entire contamination neighborhood. We will discuss globally robust inference formally in Chapter 2. To construct a globally robust confidence interval, we have to use the bias bound of the point estimate in addition to its standard error.

In this thesis, we will study globally robust inference on the simple linear regression slope. For this, we have to select an appropriate point estimate. We will discuss some robust regression estimates in the following section.

1.3 Simple Linear Regression: Robust estimation of parameters

Let us consider the simple linear regression model,

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad 1 \leq i \leq n,$$

where x_i and ε_i are independent, $\varepsilon_i \sim F_0$, $\text{Median}_{F_0}(\varepsilon_i) = 0$, $x_i \sim G_0$ and (x_i, y_i) are independent. The joint distribution of (x_i, y_i) under this model will be denoted by H_0. To allow for outliers and other departures from this central parametric model we will assume that the joint distribution H of (x_i, y_i) belongs to the ϵ-contamination neighborhood

$$\mathcal{F}_\epsilon(H_0) = \{H = (1-\epsilon)H_0 + \epsilon H^* \, : \, H^* \text{ any distribution on } \Re^2\}, \quad 0 < \epsilon < 1/2. \quad (1.5)$$

Let $T = (T_1, T_2)$ be the intercept and the slope functionals defined on a large class of distributions H on $\Re^2$ which includes $\mathcal{F}_\epsilon(H_0)$ and all the empirical distribution functions H_n. A desirable feature of the regression functional T is the property of Fisher-consistency, i.e.,

$$T_1(H_0) = \beta_0 \, , \; T_2(H_0) = \beta_1 \, .$$

The regression functional T is also expected to have the following equivariance properties. Let $\mathcal{G}(x, y)$ denote the joint distribution of (x, y).

Regression Equivariance:

$$\begin{aligned} T_1(\mathcal{G}(x, y+a+bx)) &= T_1(\mathcal{G}(x,y)) + a, \quad a, b \in \Re. \\ T_2(\mathcal{G}(x, y+a+bx)) &= T_2(\mathcal{G}(x,y)) + a, \quad a, b \in \Re. \end{aligned}$$

Affine Equivariance:

$$\begin{aligned} T_1(\mathcal{G}(cx, y)) &= T_1(\mathcal{G}(x,y)), \quad c > 0. \\ T_2(\mathcal{G}(cx, y)) &= T_2(\mathcal{G}(x,y))/c, \quad c > 0. \end{aligned}$$

Scale Equivariance:

$$\begin{aligned} T_1(\mathcal{G}(x, sy)) &= sT_1(\mathcal{G}(x,y)), \quad s > 0. \\ T_2(\mathcal{G}(x, sy)) &= sT_2(\mathcal{G}(x,y))/c, \quad s > 0. \end{aligned}$$

Location Equivariance:

$$\begin{aligned} T_1(\mathcal{G}(x+a, y)) &= T_1(\mathcal{G}(x,y)) - aT_2(\mathcal{G}(x,y)), \quad a \in \Re. \\ T_2(\mathcal{G}(x+a, y)) &= T_2(\mathcal{G}(x,y)), \quad a \in \Re. \end{aligned}$$

It is easy to show that the least squares (LS) estimates of the slope and the intercept parameters satisfy the above properties, but they have breakdown point zero. Following are some robust estimates, which also satisfy the above equivariance requirements.

1.3.1 Median of Pairwise Slopes (MPS)

This estimator was studied by Theil (1950) and Sen (1968) . For this, we first consider the set $I = \{(i,j) : x_i \neq x_j\}$ and calculate the slopes corresponding to all pair of observations (x_i, y_i) and (x_j, y_j). That is,

$$r(i,j) = \frac{y_j - y_i}{x_j - x_i}, \quad (i,j) \in I, \tag{1.6}$$

with $r(i,j)=0$ if $x_i = x_j$. Then, we define

$$\hat{\beta}_n^{MPS}(H) = \mathrm{Med}_I\, r(i,j). \tag{1.7}$$

The corresponding functional form of this estimator is

$$\hat{\beta}_n^{MPS}(H) = \mathrm{Med}\ \mathcal{G}_H \left(\frac{y_1 - y_2}{x_1 - x_2}\right),$$

where (x_1, y_1) and (x_2, y_2) are independent with common joint distribution H and $\mathcal{G}_H$ stands for the conditional distribution of the ratio given the denominator is different from zero.

Like MPS, many other estimators of the slope parameter β are based on the pairwise slopes $r(i,j)$. Most interestingly, the classical least squares estimator $\widehat{\beta}_{LS}$ may be written as a weighted average of the pairwise slopes,

$$\widehat{\beta}_{LS} = \frac{\sum\limits_{i<j} w_{ij}\ r(i,j)}{\sum\limits_{i<j} w_{ij}}, \tag{1.8}$$

with weights $w_{ij} = (x_i - x_j)^2$. Boscovich (1757) considered a data set with $n = 5$ observations, and computed the unweighted average of the 10 pairwise slopes, as well as a 10% trimmed mean given by the average of 8 of these slopes. Stigler (1986) may be referred to for a more complete historical discussion. Frees (1991) gave a survey of these and related estimators.

In general, given a slope estimate $\hat{\beta}(H)$, a median-based intercept estimate is defined by

$$\hat{\alpha}(H) = \mathrm{Med}_H(y - \hat{\beta}(H)x). \tag{1.9}$$

Thus, the corresponding regression line splits the plane into two halves containing each half of the data. Since the intercept is a location parameter for the difference $y - \hat{\beta}(H)x$, estimating it by the median seems to be appropriate from a bias point of view.

1.3.2 Generalized Median of Slopes (GMS)

The GMS estimate was first proposed by Brown and Mood (1951) and recently studied by Adrover and Zamar (2000). The GMS is defined as the solutions $(\hat{\alpha}_n, \hat{\beta}_n)$ to the equations

$$\frac{1}{n}\sum_{i=1}^{n} \text{sign}\left(y_i - \hat{\alpha}_n - \hat{\beta}_n(x_i - \hat{\mu}_n)\right) \text{sign}(x_i - \hat{\mu}_n) \;=\; 0, \tag{1.10}$$

$$\frac{1}{n}\sum_{i=1}^{n} \text{sign}\left(y_i - \hat{\alpha}_n - \hat{\beta}_n(x_i - \hat{\mu}_n)\right) \;=\; 0, \tag{1.11}$$

where $\hat{\mu}_n = \text{Med}\{x_i\}$. The GMS estimates are defined by the following geometrical property: the slope and the intercept estimates are such that the corresponding regression fit and the vertical line $x = \text{Med}\{x_i\}$ split the plane into four quarters containing the same number of points.

Adrover and Zamar (2000) showed that the slope estimate $\hat{\beta}_n^{GMS}$ satisfies the fixed point equation

$$\hat{\beta}_n^{GMS} = \text{Med}\left\{\frac{y_i - \text{Med}\{y_i - \hat{\beta}_n^{GMS}(x_i - \hat{\mu}_n)\}}{x_i - \hat{\mu}_n}\right\}. \tag{1.12}$$

Notice that this is not really a closed form formula because the right hand side of (1.12) also contains $\hat{\beta}_n^{GMS}$. A solution to this equation can be found with the following iterative algorithm proposed by Adrover and Zamar (2000). Let $\hat{\beta}_n^{(0)}$ be some initial slope estimate. Then

$$\begin{aligned} \hat{\alpha}_n^{(m+1)} &= \text{Med}\{y_i - \hat{\beta}_n^{(m)}(x_i - \hat{\mu}_n)\} \\ \hat{\beta}_n^{(m+1)} &= \text{Med}\left\{(y_i - \hat{\alpha}_n^{(m+1)})/(x_i - \hat{\mu}_n)\right\}. \end{aligned}$$

This algorithm usually converges after a few iterations. However, in some cases it runs into a closed loop. Then the algorithm automatically finds the solution by using a bisection procedure which takes place after the closed loop it detected.

The GMS estimates are a natural generalization of the minimax-bias estimate of regression through the origin. Martin and Zamar (1989) found that in the case of regression through the origin, $y_i = \beta x_i + \varepsilon_i$, the MS estimate

$$\hat{\beta}(H) = \mathrm{Med}_H \left(\frac{y}{x} \right)$$

minimizes the maxbias among all equivariant estimates. It can be verified that $\hat{\beta}(H)$ is a generalized M-estimate (GM) that satisfies the equation

$$\mathrm{E}_H \, \mathrm{sign}(y - bx) \, \mathrm{sign}(x) = 0.$$

For the simple linear regression model, the estimates $(\hat{\alpha}(H), \hat{\beta}(H))$ can be defined implicitly by the equations

$$\mathrm{E}_H \, \mathrm{sign}(y - a - bx) \, \mathrm{sign}(x - m) = 0 \qquad (1.13)$$

$$\mathrm{E}_H \, \mathrm{sign}(y - a - bx) = 0 \qquad (1.14)$$

where $m = \mathrm{Med}_H(x)$. The finite sample versions are obtained by setting $H = H_n$, the empirical distribution function of the data. Clearly, (1.13) and (1.14) are particular cases of GM-estimates. It can be easily seen that the GMS estimates satisfy (1.13) and (1.14) and, therefore, they are a very special type of GM-estimates.

1.3.3 Repeated Median of Slopes (RMS)

Siegel (1982) defined the first slope estimate with breakdown point 0.5, by performing a different median-based operation over the ratios (1.6). In this case, we define the slope estimator as

$$\hat{\beta}_n^{RMS} = \mathrm{Med}_{1 \le i \le n} \, \mathrm{Med}_{j \in J_i} r(i, j),$$

where $J_i = \{j \, : \, (i, j) \in I\}$, for $1 \le i \le n$. The corresponding functional form of this estimate is as follows. First, let us define

$$q_M(a, b, H) = \mathrm{Med} \, \mathcal{G}_H \left(\frac{y_1 - b}{x_1 - a} \right), \qquad (1.15)$$

for fixed numbers a and b. Then, the RMS estimator is defined as

$$\hat{\beta}_n^{RMS} = \text{Med}\, \mathcal{G}_H^* \left(q_M(x_2, y_2, H)\right), \tag{1.16}$$

where (x_1, y_1) and (x_2, y_2) are independent with common joint distribution H and $\mathcal{G}_H$ stands for the conditional distribution of the ratio given the denominator is different from zero. $\mathcal{G}_H^*$ denotes the distribution of $q_M(x_2, y_2, H)$ under H.

1.3.4 S-estimates

The regression estimates defined so far are median-based. Before defining the regression S-, MM- and τ-estimates, we need to define the M-estimates of scale. For this we may consider the parametric scale model

$$F_\sigma(x) = F_0(x/\sigma),$$

where F_0 is the distribution function of a random variable X with a density function symmetric about zero. The M-estimates of scale were defined by Huber (1964) as

$$S(F) = \inf \left\{ s \,:\, \mathrm{E}_F\, \rho\left(\frac{X}{s}\right) < b \right\}, \tag{1.17}$$

where $b = \mathrm{E}_{F_0}\, \rho(X)$ and the score function ρ satisfies the following assumptions:

(a) $\rho(x) : \mathbb{R} \to \mathbb{R}$ is symmetric and nondecreasing on $[0, \infty)$.

(b) $\rho(0) = 0$, and $\rho(\infty) = 1$.

(c) $\rho(x)$ has at most a finite number of discontinuities.

The regression S-estimates were defined by Rousseeuw and Yohai (1984) as

$$\hat{\boldsymbol{\theta}}_S = \underset{\mathbf{t}}{\operatorname{argmin}}\, S_\rho\left(\mathbf{t},\, H\right),$$

where $S_\rho(\mathbf{t},\, H)$ is the M-scale of the absolute residuals $r(\mathbf{t}) = |y - \mathbf{t}^{\mathrm{T}}\mathbf{x}|$, in which $(y, \mathbf{x})$ have joint distribution H. Mathematically,

$$S_\rho(\mathbf{t},\, H) = \inf\left\{ s \,:\, \mathrm{E}_H\, \rho\left(\frac{r(\mathbf{t})}{s}\right) < b \right\}.$$

1.3.5 MM-estimates

The regression MM-estimates were defined by Yohai (1987) as

$$\hat{\boldsymbol{\theta}}_{MM} = \underset{\mathbf{t}}{\operatorname{argmin}}\, \mathrm{E}_H\, \rho_2\left(\frac{r(\mathbf{t})}{s_1}\right),$$

where $s_1 = S_1(H) = \min_{\mathbf{t}} S_1(\mathbf{t},\, H)$ is the scale of the absolute residuals corresponding to a regression S-estimate with score function $\rho_1 \in C_{b_1}$, and $\rho_2 \in C_{b_2}$ is a score function chosen to attain a higher efficiency.

1.3.6 τ-estimates

The regression τ-estimates were defined by Yohai and Zamar (1988) as

$$\hat{\boldsymbol{\theta}}_\tau = \underset{\mathbf{t}}{\operatorname{argmin}}\, \tau(\mathbf{t}, H),$$

where

$$\tau(\mathbf{t}, H) = S^2(\mathbf{t}, H)\, \mathrm{E}_H\, \rho_2\left(\frac{r(\mathbf{t})}{S(\mathbf{t}, H)}\right),$$

and $S(\mathbf{t}, H)$ is the M-scale defined before with $\rho = \rho_1 \in C_{b_2}$. These estimates can attain a high breakdown point, controlled by ρ_1, and a high efficiency, controlled by ρ_2.

The bias behavior of different robust estimates will be discussed in Chapter 3, with a view to selecting a 'good' point estimate for robust inference.

1.4 Purpose of this study

Adrover, Salibian-Barrera and Zamar (2002) developed the idea of globally robust inference for the location and the simple linear regression models. After showing the consequences of ignoring the asymptotic bias of the point estimate, they incorporated the bias bound in the construction of confidence intervals. For robust inference on simple linear regression slope, the authors selected GMS as their point estimate, considering its good bias behavior and asymptotic normality. However, GMS has a breakdown point of only 0.25, and its asymptotic normality is established under very restrictive conditions. Also, the bias bound for the GMS estimate is known only for symmetric carrier distributions, limiting the applications of the method.

In this study, we will consider the RMS as an alternative choice of point estimate for robust inference on slope. RMS has a breakdown point of 0.50, its asymptotic normality holds under very general conditions, and the bias bound for RMS is known for general carrier distributions.

1.5 Organization of subsequent chapters

In Chapter 2, we will explain the idea of globally robust inference for the location and the simple linear regression models. In Chapter 3, we will justify the selection of the RMS estimate as an alternative to the GMS estimate for robust inference on slope. In Chapter 4, we will apply the methods to two different datasets and compare the results. We will present and discuss the results of a Monte Carlo simulation in Chapter 5. In the final chapter, we will conclude by summarizing our study and pointing towards some topics for future research.

Chapter 2

Globally Robust Inference

The vast majority of robustness literature focuses on point estimation. There are a few papers that address robust inference, but they do not take into account the possible bias of the point estimates. One exception is Adrover, Salibian-Barrera and Zamar (2002). Our discussion on globally robust inference will be based mainly on this paper.

In order to highlight the main ideas, let us consider the following *location-scale* model:

$$y = \theta + \sigma\varepsilon. \tag{2.1}$$

Here, θ is an unknown location parameter, σ is a nuisance scale parameter, and ε has a specified distribution F_0. Correspondingly, the distribution of y is $F_\theta(y) = F_0((y-\theta)/\sigma)$. To allow for outliers and other departures from this model, we assume that the actual distribution F belongs to the ϵ-contamination neighborhood ,

$$\mathcal{F}_\epsilon(F_\theta) = \{F = (1-\epsilon)F_\theta + \epsilon F^* : F^* \text{ any arbitrary distribution}\}, \quad 0 < \epsilon < 1/2. \tag{2.2}$$

According to this model, the majority of the data comes from a central parametric model but a minority may come from an unknown arbitrary distribution.

According to Adrover *et al.* (2002), a robust confidence interval should be *stable* and *informative*. The robust confidence interval should be *stable* in the sense of keeping a high coverage level (at or above the nominal level) not only at the central model but also over the contamination neighborhood. The interval should also be *informative* in the sense of keeping a reasonable average length over the entire neighborhood. These two properties are more precisely stated in the following definition by the authors.

Definition 2.1 A confidence interval (L_n, U_n) for θ 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 neighborhood is $(1-\alpha)$:

$$\lim_{n\to\infty} \inf_{F\in\mathcal{F}_\epsilon(F_\theta)} P_F(L_n < \theta < U_n) \geq (1-\alpha);$$

2. (*Informative interval)* The maximum asymptotic length of the interval is bounded over the ϵ-contamination neighborhood:

$$\lim_{n\to\infty} \sup_{F\in\mathcal{F}_\epsilon(F_\theta)} [U_n - L_n] < \infty.$$

2.1 Limitations of the classical confidence intervals

It can be easily shown that classical Student-t confidence intervals,

$$\overline{X}_n \pm t_{(n-1)}(1-\alpha/2)S_n/\sqrt{n},$$

fail Part 1 and Part 2 of Definition 1. Let us consider first the contaminated distribution

$$F^{x_0} = (1-\epsilon)F_0 + \epsilon F^*,$$

where F^* is a point mass distribution at $x_0 > 0$. Then,

$$L_n = \overline{X}_n - t_{(n-1)}(1-\alpha/2)S_n/\sqrt{n} \xrightarrow{a.s.} \epsilon x_0,$$

and

$$U_n = \overline{X}_n + t_{(n-1)}(1-\alpha/2)S_n/\sqrt{n} \xrightarrow{a.s.} \epsilon x_0$$

as n tends to ∞. Therefore,

$$\lim_{n\to\infty} \inf_{F\epsilon\mathcal{F}_\epsilon(F_\theta)} P_F(L_n < \theta < U_n) \leq \lim_{n\to\infty} P_{F^{x_0}}(L_n < \theta < U_n) = 0.$$

Thus, the classical intervals fail Part 1 of Definition 1. Let us now take F^* to be a point mass distribution at $\pm x_0$ (equally weighted). Then,

$$\lim_{n\to\infty} \sup_{F\epsilon\mathcal{F}_\epsilon(F_\theta)} [U_n - L_n] \geq \lim_{n\to\infty} \sup_{x_0>0} 2t_{(n-1)}(\alpha/2)\frac{x_0\sqrt{\epsilon}}{\sqrt{n}} = \infty,$$

failing Part 2 of the definition.

2.2 Naive intervals: consequences of ignoring bias

For robustifying Student's t confidence intervals, it seems natural to replace $\overline{X}_n$ by a robust asymptotically normal point estimate T_n, and $S_n/\sqrt{n}$ by a robust estimate of the standard error of T_n. We will call this the naive procedure. It can be shown that the naive confidence intervals satisfy Part 2 but not Part 1 of Definition 2.1. Consequently, the asymptotic coverage proportion of the naive confidence intervals of any nominal level will invariably tend to zero for all $\epsilon > 0$.

To illustrate this point Adrover *et al.* (2002) conducted a Monte Carlo simulation in which they generated 10,000 normal samples of different sizes, containing various fractions of contamination. The contaminating distribution is a point mass distribution at $x_0 = 4$. For each sample, the authors calculated the location M-estimate with Huber ψ-function

$$\psi(y) = \min\{-c, \max\{c, y\}\},$$

with truncation constant $c = 1.345$, and the corresponding 95% confidence intervals based on the empirical asymptotic variance. The following table summarizes the observed coverage levels and the average lengths of these intervals.

Table 2.1: Percentage of coverage and average length of naive CI for location parameter

ϵ	Sample Size	% of coverage	Average length
0.05	20	92	0.91
	50	92	0.60
	100	88	0.44
	200	82	0.31
0.10	20	91	1.05
	50	84	0.68
	100	67	0.49
	200	39	0.35
0.15	20	88	1.19
	50	72	0.76
	100	35	0.56
	200	5	0.40
0.20	20	82	1.41
	50	45	0.92
	100	8	0.66
	200	0	0.47

The poor coverage levels in the above table are due to the asymptotic bias of the point estimate. Let $\hat{\theta}_n$ be a robust estimate of the location parameter, and $\hat{\theta}(F)$ be its asymptotic value under an asymmetric distribution F belonging to the contamination neighborhood. Usually, $\hat{\theta}(F)$ is different from the actual value of the parameter θ, and the asymptotic bias remains the same even if the sample size increases. However, the standard errors of $\hat{\theta}_n$ are very small for large sample sizes, and most of the probability mass in the distribution of $\hat{\theta}_n$ concentrates on an interval that excludes θ. Ironically, if the data are of uneven quality, large sample sizes are bad for naive confidence intervals.

The problem is more severe in the case of simple linear regression. For the model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, the classical $100(1-\alpha)\%$ confidence intervals for β_1 are of the form $\hat{\beta}_1 \pm z_{\alpha/2}\, \hat{\mathrm{SE}}(\hat{\beta}_1)$ where $\hat{\beta}_1$ is the least squares estimate for β_1 and $\hat{\mathrm{SE}}(\hat{\beta}_1)$ is the estimated standard error of $\hat{\beta}_1$. For robustifying these confidence intervals, the naive approach is to replace $\hat{\beta}_1$ by a robust point estimate, $\hat{\beta}_1^R$, and $\hat{\mathrm{SE}}(\hat{\beta}_1)$ by a robust estimate of the standard error of $\hat{\beta}_1^R$. To show that the naive confidence intervals are not "stable" (do not satisfy Part 1 of Definition 1), Adrover *et al.* generated 600 samples $(x_i,\, y_i)$ of sizes $n =$ 20, 40, 60, 80 and 100 from contaminated normal distributions $(1-\epsilon)N(\mathbf{0}, I)+\epsilon N(\boldsymbol{\mu}, \tau^2 I)$ with $\boldsymbol{\mu}' = (\mu_x, \mu_y)$, $\tau = 0.1$, $\mu_x = 3$ and $\mu_y = 1.5$ (2.0) for $\epsilon = 0.05$ (0.10). The authors refer to this case as the "mild contamination case". In the "strong contamination case" they took $\mu_x = 5$ and $\mu_y = 2.5$ for $\epsilon = 0.05$ and $\epsilon = 0.10$. They calculated the high breakdown point regression MM-estimates (Yohai, 1987) and their asymptotic standard errors. The nominal confidence level in each case is 0.95. The authors reported the coverage for the slope parameter, which is reproduced in the following table.

Clearly, the coverage levels shown in the table are much less than the nominal level, specially for larger sample sizes. As in the case of the location model, the poor coverage levels are due to the asymptotic bias of the point estimate $\hat{\beta}_1^R$ introduced by the contamination in the data. In general, the standard error of $\hat{\beta}_1^R$ is of order $1/\sqrt{n}$ while its bias does not vanish as n goes to infinity.

Therefore, for (globally) robust inference, we have to consider the asymptotic bias of the selected estimator, as well as its standard error. In the following two sections, we will discuss robust inference for the location and the simple linear regression models.

Table 2.2: Coverage proportion of naive CI for the slope parameter

% of Contamination	Sample Size	Mild Contamination	Strong Contamination
5	20	0.88	0.76
	40	0.82	0.57
	60	0.81	0.35
	80	0.62	0.26
	100	0.47	0.23
10	20	0.71	0.50
	40	0.46	0.25
	60	0.27	0.11
	80	0.15	0.07
	100	0.13	0.05

2.3 Robust inference for the location model

First, we should know how to incorporate the asymptotic bias (bias bound) of the point estimate in the construction of confidence intervals, one sided confidence bounds, and p-values. Then, we need to know how to estimate the bias bound of the estimate.

2.3.1 Confidence Intervals

Suppose that we have a robust point estimate $\hat{\theta}_n$ for the parameter θ which satisfies

$$\sqrt{n}(\hat{\theta}_n - \hat{\theta}(F)) \xrightarrow{d} N(0, v^2(F)), \tag{2.3}$$

where $|\hat{\theta}(F) - \theta| < \bar{\theta}$ and $\bar{\theta}$ is an upper bound for the bias of $\hat{\theta}(F)$ as defined in Berrendero and Zamar (2001).

For any fixed $F \in \mathcal{F}_\epsilon(F_\theta)$, an asymptotic confidence interval of level $(1-\alpha)$ is given by

$$\left(\hat{\theta}_n - l_n,\ \hat{\theta}_n + r_n\right),$$

where $l_n = l_n(F)$ and $r_n = r_n(F)$ satisfy the equation

$$P_F\left(-r_n \leq \hat{\theta}_n - \theta \leq l_n\right) = 1-\alpha. \tag{2.4}$$

We can write

$$\begin{aligned} P_F\left(-r_n \leq \hat{\theta}_n - \hat{\theta}(F) + \hat{\theta}(F) - \theta \leq l_n\right) &= \\ P_F\left(\frac{-r_n - b}{v_n} \leq \frac{\hat{\theta}_n - \hat{\theta}(F)}{v_n} \leq \frac{l_n - b}{v_n}\right) &= 1-\alpha, \end{aligned} \tag{2.5}$$

where $\sqrt{n}v_n$ is a consistent estimate for $v(F)$ for all $F \in \mathcal{F}_\epsilon(F_\theta)$, and $b = b(F) = \hat{\theta}(F) - \theta$.

From equation (2.3) the normal distribution can be used to approximate the left-hand side of equation (2.5). Hence, we can obtain estimates $\hat{l}_n$ and $\hat{r}_n$ solving the following equation:

$$\Phi\left(\frac{\hat{l}_n - b}{v_n}\right) + \Phi\left(\frac{\hat{r}_n + b}{v_n}\right) - 1 = 1-\alpha, \tag{2.6}$$

where Φ is the standard normal cumulative distribution function. Since $F \in \mathcal{F}_\epsilon(F_\theta)$ is *unspecified*, the bias $b = b(F)$ in (2.6) is unknown and *cannot* be estimated from the data. Therefore, the endpoints $\hat{l}_n$ and $\hat{r}_n$ may be found in such a way that

$$\Phi\left(\frac{\hat{l}_n - b(F)}{v_n}\right) + \Phi\left(\frac{\hat{r}_n + b(F)}{v_n}\right) - 1 \geq 1-\alpha, \text{ for all } F \in \mathcal{F}_\epsilon(F_\theta).$$

The coverage of the confidence intervals obtained in this way will be $1-\alpha$ not only at the central model, but also over the entire neighborhood.

The endpoint $\hat{l}_n$ may be expressed as a function of $\hat{r}_n$ by using Equation (2.6):

$$\hat{l}_n = v_n \Phi^{-1}\left[2 - \alpha - \Phi\left(\frac{\hat{r}_n + b}{v_n}\right)\right] + b,$$

and $\hat{r}_n$ may be chosen in order to minimize the resulting interval length:

$$\hat{l}_n + \hat{r}_n = v_n \Phi^{-1}\left[2 - \alpha - \Phi\left(\frac{\hat{r}_n + b}{v_n}\right)\right] + b + \hat{r}_n. \tag{2.7}$$

Differentiating the right-hand side of Equation (2.7) with respect to $\hat{r}_n$ and setting the derivative equal to zero, we have

$$-\frac{\varphi\left(\frac{\hat{r}_n+b}{v_n}\right)}{\varphi\left(\Phi^{-1}\left[2-\alpha-\Phi\left(\frac{\hat{r}_n+b}{v_n}\right)\right]\right)} + 1 = 0,$$

where φ is the standard normal density function. That is

$$\varphi\left(\frac{\hat{r}_n + b}{v_n}\right) = \varphi\left(\Phi^{-1}\left[2 - \alpha - \Phi\left(\frac{\hat{r}_n + b}{v_n}\right)\right]\right).$$

From this equation it follows that

$$\frac{\hat{r}_n + b}{v_n} = \pm\,\Phi^{-1}\left[2 - \alpha - \Phi\left(\frac{\hat{r}_n + b}{v_n}\right)\right].$$

The minus sign case is discarded because it can only be satisfied with $\alpha = 1$. From the other case we obtain

$$\Phi\left(\frac{\hat{r}_n + b}{v_n}\right) = 2 - \alpha - \Phi\left(\frac{\hat{r}_n + b}{v_n}\right) \tag{2.8}$$

which yields the solution

$$\begin{aligned} \hat{r}_n &= v_n \Phi^{-1}(1 - \alpha/2) - b = v_n z_{\alpha/2} - b, \\ \hat{l}_n &= v_n z_{\alpha/2} + b. \end{aligned}$$

The corresponding confidence interval for θ of level greater than or equal to $(1-\alpha)$ for fixed $F \in \mathcal{F}_\epsilon(F_\theta)$ is:

$$\hat{I}_n(F) = \left(\hat{\theta}_n - v_n z_{\alpha/2} - b,\ \hat{\theta}_n + v_n z_{\alpha/2} - b\right).$$

Since $\hat{I}_n(F)$ still depends on the unknown F through b, a robust interval of level $(1-\alpha)$ can be constructed as follows:

$$I_n = \bigcup_{F \in \mathcal{F}_\epsilon(F_\theta)} \left(\hat{\theta}_n - v_n z_{\alpha/2} - b,\ \hat{\theta}_n + v_n z_{\alpha/2} - b\right). \tag{2.9}$$

Since $|b| = |b(F)| < \bar{\theta}$ for all $F \in \mathcal{F}_\epsilon(F_\theta)$ we have

$$I_n \subseteq \left(\hat{\theta}_n - v_n z_{\alpha/2} - \bar{\theta},\ \hat{\theta}_n + v_n z_{\alpha/2} + \bar{\theta}\right)$$

with equality if $\bar{\theta}$ is a sharp bias bound, that is if $\bar{\theta} = \sup_{\mathcal{F}_\epsilon(F_\theta)} |b(F)|$. The above robust confidence intervals can be constructed for any estimate $\hat{\theta}_n$ that is asymptotically normal, has a consistent estimate of its standard error, and a known bias bound $\bar{\theta}$. In order to obtain the shortest robust confidence interval we will consider estimates where such a sharp bound is available.

The robust confidence intervals have positive asymptotic lengths, which is the price we have to pay for *robust coverage*.

An alternative approach

An alternative robust interval can be constructed following Fraiman, Yohai and Zamar (2001). The main idea is to find q_n such that

$$P_F(|\hat{\theta}_n - \theta| \leq q_n) = 1 - \alpha.$$

Due to the asymptotic normality of $\hat{\theta}_n$, we can estimate q_n by

$$\Phi\left(\frac{\tilde{q}_n - b}{v_n}\right) + \Phi\left(\frac{\tilde{q}_n + b}{v_n}\right) - 1 = 1 - \alpha. \tag{2.10}$$

Since $\tilde{q}_n$ is a monotone function of the bias, and its largest value is obtained by replacing b by $\bar{\theta}$, we have

$$\Phi\left(\frac{\bar{q}_n - \bar{\theta}}{v_n}\right) + \Phi\left(\frac{\bar{q}_n + \bar{\theta}}{v_n}\right) - 1 = 1 - \alpha, \tag{2.11}$$

which gives the robust confidence interval $\hat{\theta}_n \pm \bar{q}_n$.

The robust interval (2.9) is easier to compute, but slightly longer than that in (2.11).

2.3.2 One sided confidence bound

In this section we will discuss robust lower bounds for the parameter θ. Robust upper bounds can be constructed in a similar way. For any fixed $F \in \mathcal{F}_\epsilon(F_\theta)$, an asymptotic lower bound of level $(1-\alpha)$ for θ is given by

$$\left(\hat{\theta}_n - l_n, \infty\right) \tag{2.12}$$

where $l_n = l_n(F)$ satisfies the equation

$$P_F\left(\hat{\theta}_n - \theta \leq l_n\right) = 1 - \alpha. \tag{2.13}$$

Similarly to (2.5), and recalling that $b = b(F) = \hat{\theta}(F) - \theta$, we have that l_n must satisfy

$$P_F\left(\frac{\hat{\theta}_n - \hat{\theta}(F)}{v_n} \leq \frac{l_n - b}{v_n}\right) = 1 - \alpha \tag{2.14}$$

and, as before, for large n,

$$\hat{l}_n(F) = v_n \Phi^{-1}(1-\alpha) + b. \tag{2.15}$$

A robust lower bound of level $1-\alpha$ can now be defined as

$$(L_n^R, \infty) = \bigcup_{F \in \mathcal{F}_\epsilon(F_\theta)} \left(\hat{\theta}_n - \hat{l}_n(F), \infty\right). \tag{2.16}$$

We have,

$$\left(\hat{\theta}_n - \hat{l}_n(F), \infty\right) = \left(\hat{\theta}_n - v_n \Phi^{-1}(1-\alpha) - b, \infty\right). \tag{2.17}$$

Therefore,

$$(L_n^R, \infty) = \left(\hat{\theta}_n - v_n z_\alpha - \bar{\theta}, +\infty\right).$$

Unlike in the confidence interval case, if we use a construction similar to the one in Fraiman, Yohai and Zamar (2001), we obtain the same lower robust confidence bounds.

2.3.3 P-values

Let us start by considering the following artificial hypothesis testing problem. Let us assume that the contaminating distribution H in $F = (1-\epsilon)F_\theta + \epsilon H$ is known, that is, that F belongs to the following "translated" family

$$\mathcal{F}_\epsilon(F_\theta, H) = \{F = (1-\epsilon)F_\theta + \epsilon H, \text{ fixed } H\} \tag{2.18}$$

Suppose that we are interested in testing

$$H_0 : \theta \le \theta_0 \quad \text{versus} \quad H_1 : \theta > \theta_0 \tag{2.19}$$

for $F \in \mathcal{F}_\epsilon(F_\theta, H)$. Using the lower bounds in (2.17), let us define $\hat{\theta}_n(\alpha, F) = \hat{\theta}_n - v_n\Phi^{-1}(1-\alpha) - b(F)$, and a test for (2.19) is given by the rejection rule:

$$\text{Reject } H_0 \text{ if } \quad \hat{\theta}_n(\alpha, F) > \theta_0.$$

This is a size α test since

$$\begin{aligned} &\sup_{\theta \le \theta_0} P_{(1-\epsilon)F_\theta + \epsilon H}\left(\hat{\theta}_n(\alpha, (1-\epsilon)F_\theta + \epsilon H) > \theta_0\right) \\ &\qquad = \quad P_{(1-\epsilon)F_{\theta_0} + \epsilon H}\left(\hat{\theta}_n(\alpha, (1-\epsilon)F_{\theta_0} + \epsilon H) > \theta_0\right) = \alpha. \end{aligned}$$

In other words, for the fixed family (2.18) the p-value is given by

$$\hat{p}_n(F) = \inf\left\{\alpha : \hat{\theta}_n(\alpha, F) > \theta_0\right\}. \tag{2.20}$$

We can calculate $\hat{p}_n(F)$ explicitly because $\hat{\theta}_n(\alpha, F)$ is increasing in α and hence $\hat{p}_n(F)$ satisfies $\hat{\theta}_n(\hat{p}_n(F), F) = \theta_0$, or equivalently,

$$\hat{p}_n(F) = 1 - \Phi\left(\frac{\hat{\theta}_n - \theta_0 - b}{v_n}\right). \tag{2.21}$$

Similarly, for the case $H_0 : \theta \ge \theta_0$ versus $H_1 : \theta < \theta_0$, we have $\hat{p}_n(F) = \Phi\left(\frac{\hat{\theta}_n - \theta_0 - b}{v_n}\right)$.

Let us consider the 2-sided hypothesis

$$H_0 : \theta = \theta_0 \ \text{ versus } \ H_1 : \theta \ne \theta_0.$$

We construct the p-value for this case based on $\tilde{q}_n(\alpha, F)$ defined by the equation

$$\Phi\left(\frac{\tilde{q}_n(\alpha, F) - b}{v_n}\right) + \Phi\left(\frac{\tilde{q}_n(\alpha, F) + b}{v_n}\right) - 1 = 1 - \alpha. \tag{2.22}$$

The rejection rule would be of the form: reject H_0 if $|\hat{\theta}_n - \theta_0| > \tilde{q}_n(\alpha, F)$. The p-value is

$$\hat{p}_n(F) = \inf\left\{\alpha \,:\, \tilde{q}_n(\alpha, F) < |\hat{\theta}_n - \theta_0|\right\}.$$

Here, $\tilde{q}_n(\alpha, F)$ is a decreasing function of α. Therefore $\hat{p}_n(F)$ solves the equation

$$\tilde{q}_n(\hat{p}_n(F), F) = |\hat{\theta}_n - \theta_0|. \tag{2.23}$$

From (2.22) it is easy to see that the solution $g(x)$ to the equation $\tilde{q}_n(g(x), F) = x$ is given by

$$g(x) = 2 - \Phi\left(\frac{x - b}{v_n}\right) - \Phi\left(\frac{x + b}{v_n}\right). \tag{2.24}$$

We obtain

$$\hat{p}_n(F) = 2 - \Phi\left(\frac{|\hat{\theta}_n - \theta_0| - b}{v_n}\right) - \Phi\left(\frac{|\hat{\theta}_n - \theta_0| + b}{v_n}\right). \tag{2.25}$$

The p-values (2.21) and (2.25) were obtained for $F \in \mathcal{F}_\epsilon(F_\theta, H)$. We can then define a robust p-value for all $F \in \mathcal{F}_\epsilon(F_\theta)$ as

$$\hat{p}_n^R = \sup_{F \in \mathcal{F}_\epsilon(F_\theta)} \hat{p}_n(F). \tag{2.26}$$

It is easy to see that $\hat{p}_n(F)$ is a monotone function of $b = b(F)$ and hence, if the bias bound $\bar{\theta}$ is sharp, then the robust p-values are as shown in Table 2.3.

The rejection rules associated with the p-values in (2.26) have robust Type I Errors at the expense of lower power. In other words, to guarantee a level-α test *with uncertainty in our model* we lose power, in particular for values of the parameter near the null hypothesis. Lower power is a reasonable price to pay to achieve a *robust rejection rule.*

Table 2.3: Robust p-values for location-scale problems.

Hypothesis	Robust p-values
$H_1 : \theta > \theta_0$	$1 - \Phi\left(\frac{\hat{\theta}_n - \theta_0 - \bar{\theta}}{\hat{\sigma}_n}\right)$
$H_1 : \theta < \theta_0$	$\Phi\left(\frac{\hat{\theta}_n - \theta_0 + \bar{\theta}}{\hat{\sigma}_n}\right)$
$H_1 : \theta \neq \theta_0$	$2 - \Phi\left(\frac{\lvert\hat{\theta}_n - \theta_0\rvert - \bar{\theta}}{\hat{\sigma}_n}\right) - \Phi\left(\frac{\lvert\hat{\theta}_n - \theta_0\rvert + \bar{\theta}}{\hat{\sigma}_n}\right)$

2.3.4 Estimation of bias bound

Let T_n be any location estimate with asymptotic value $T(F)$. The *asymptotic bias* is defined in the following invariant way:

$$b(T, F) = |T(F) - \theta|/\sigma, \tag{2.27}$$

where σ is the true error scale parameter. The *maximum asymptotic bias* was defined earlier as

$$B(\epsilon) = \sup_{F \in \mathcal{F}_\epsilon} b(T, F). \tag{2.28}$$

Clearly,

$$|T(F) - \theta| \leq \sigma B(\epsilon) \quad \text{for all} \quad F \in \mathcal{F}_\epsilon.$$

Unfortunately, σ is unknown. Let $\hat{\sigma}(F)$ be the limiting value of the scale estimate $\hat{\sigma}_n$. Then, a more useful bias bound would be

$$\tilde{B}(\epsilon) = \sup_{F \in \mathcal{F}_\epsilon} \tilde{b}(T, F), \tag{2.29}$$

with

$$\tilde{b}(T, F) = \frac{|T(F) - \theta|}{\hat{\sigma}(F)}.$$

This equation is more useful than (2.28) because we have

$$|T(F) - \theta| \leq \hat{\sigma}(F)\tilde{B}(\epsilon) \quad \text{for all} \quad F \in \mathcal{F}_\epsilon,$$

and, for large n, if we replace $\hat{\sigma}(F)$ by $\hat{\sigma}(F_n)$ the above relationship still holds approximately. However, $\tilde{B}(\epsilon)$ is unknown and its theoretical derivation appears to be difficult. A numerical approximation restricting the supremum to point mass contamination would be feasible. However, for the construction of confidence intervals we can choose the following simpler approach (Berrendero and Zamar, 2001). By replacing σ by $\hat{\sigma}(F)$ we do not obtain an upper bound due to the possible underestimation of the scale. An estimated bias bound $\bar{\theta}_n$ is given by

$$\bar{\theta}_n = k\hat{s}_n B(\epsilon), \tag{2.30}$$

with

$$k = \sup_{F\in\mathcal{F}_\epsilon} \frac{\sigma}{\hat{\sigma}(F)} = \frac{1}{s^-(\epsilon)},$$

$$\hat{s}_n = \text{shorth}(y_i),$$

and

$$s^-(\epsilon) = \inf_{F\in\mathcal{F}_\epsilon} \frac{\hat{\sigma}(F)}{\sigma}.$$

The shorth is the standardized length of the shortest half of the sorted data. For the derivation of s^-, Lemma 2.1 (at the end of this chapter) can be used.

2.4 Robust inference for the simple linear regression model

We will discuss the confidence intervals for the slope parameter in the simple linear regression model,

$$y_i = \beta_0 + \beta_1(x_i - \mu) + \varepsilon_i, \quad 1 \le i \le n,$$

where μ is a location parameter, x_i and ε_i are independent, $\varepsilon_i \sim F_0$, $\text{Median}_{F_0}(\varepsilon_i) = 0$, $x_i \sim G_0$ and (x_i, y_i) are independent. The joint distribution of (x_i, y_i) under this model will be denoted by H_0. To allow for outliers and other departures from this central parametric model we will assume that the joint distribution H of (x_i, y_i) belongs to the ϵ-contaminated neighborhood

$$\mathcal{F}_\epsilon(H_0) = \{H = (1-\epsilon)H_0 + \epsilon H^* \,:\, H^* \text{ any distribution on } \Re^2\}, \quad 0 < \epsilon < 1/2.$$

The construction of the robust confidence intervals for the slope follows along the lines of the construction of location confidence intervals discussed earlier. Let us assume that $\hat{\beta}_1$ is an asymptotically normal and robust point estimate for β_1,

$$\sqrt{n}(\hat{\beta}_1 - \hat{\beta}_1(H)) \xrightarrow{d} N(0, v^2(H)),$$

and $\hat{\beta}_1$ has a sharp and known bias bound $\bar{\beta}$. We need to find q_n such that

$$P_H(|\hat{\beta}_1 - \hat{\beta}_1(H)| \leq q_n) = 1 - \alpha.$$

Due to the asymptotic normality of $\hat{\beta}_1$, we can estimate q_n by

$$\Phi\left(\frac{\tilde{q}_n(H) - b}{v_n}\right) + \Phi\left(\frac{\tilde{q}_n(H) + b}{v_n}\right) - 1 = 1 - \alpha. \qquad (2.31)$$

Since $\tilde{q}_n$ is a monotone function of the bias, and its largest value is obtained by replacing b by $\bar{\beta}$, we have

$$\Phi\left(\frac{\bar{q}_n - \bar{\beta}}{v_n}\right) + \Phi\left(\frac{\bar{q}_n + \bar{\beta}}{v_n}\right) - 1 = 1 - \alpha, \qquad (2.32)$$

which gives the robust confidence interval $\hat{\beta}_1 \pm \bar{q}_n$.

2.4.1 Estimation of the bias bound

Let $T_{1,n}$ be estimate for the slope parameter with asymptotic value $T_1(H)$. The invariant asymptotic bias is defined in the following way (Adrover and Zamar, 2000):

$$b(T_1(H)) \;=\; \sigma_x |T_1(H) - \beta_1| / \sigma_\varepsilon,$$

where σ_ε and σ_x are the residual and explanatory variable scales under the central model H_0. These biases are invariant under affine transformations of the data. The maximum asymptotic bias is given by

$$B_1(\epsilon) = \sup_{H \in \mathcal{F}_\epsilon} b(T_1, H).$$

Finally, the bias bound $\bar{\beta}_1$ is defined by

$$\sup_{H \in \mathcal{F}_\epsilon} |T_1(H) - \beta_1| \leq \sup_{H \in \mathcal{F}_\epsilon} \frac{\sigma_\varepsilon(H)}{\sigma_x(H)} B_1(\epsilon) = \bar{\beta}_1.$$

As before, following Berrendero and Zamar (2001) an estimate of the bias bound is given by

$$\bar{\beta}_{1n} = \frac{\sigma_{\varepsilon,n}}{\sigma_{x,n}} k(\epsilon) B_1(\epsilon) = \frac{\sigma_{\varepsilon,n}}{\sigma_{x,n}} B_1^*(\epsilon), \tag{2.33}$$

where $k(\epsilon) = s^+(\epsilon)/s^-(\epsilon)$, and we can use

$$\hat{\sigma}_{\varepsilon,n} = \text{shorth}(y_i - \hat{\alpha}_n - \hat{\beta}_n x_i) \quad \text{and} \quad \hat{\sigma}_{x,n} = \text{shorth}(x_i)$$

to estimate $\hat{\sigma}_\varepsilon(H)$ and $\hat{\sigma}_x(H)$ (Rousseeuw and Leroy, 1987). The shorth is bias-minimax in the class of M-estimates of scale with general location (Martin and Zamar, 1993).

We need to determine $s^+(\epsilon)$ and $s^-(\epsilon)$. Martin and Zamar (1993) showed that

$$\inf_{H \in \mathcal{F}_\epsilon} \frac{\sigma_x}{\hat{\sigma}_x(H)} = \frac{\Phi^{-1}\left(\frac{3}{4}\right)}{\Phi^{-1}\left(\frac{3-2\epsilon}{4(1-\epsilon)}\right)} = \frac{1}{s^+(\epsilon)}.$$

For $s^-(\epsilon)$, we can use the following lemma (Adrover *et al.*, 2002).

Lemma 2.1 *If $(x, y) \sim H \in \mathcal{F}_\epsilon$, then,*

$$\inf_{H \in \mathcal{F}_\epsilon} \frac{\hat{\sigma}_\varepsilon(H)}{\sigma_\varepsilon} = \frac{\Phi^{-1}\left(\frac{3-4\epsilon}{4(1-\epsilon)}\right)}{\Phi^{-1}\left(\frac{3}{4}\right)} = \frac{1}{s^-(\epsilon)},$$

where $\hat{\sigma}_\varepsilon(H) = \text{shorth}\left(y - \hat{\alpha}(H) - \hat{\beta}(H)x\right)$. A similar result follows for the location case by taking $\hat{\beta}(H) = 0$.

The proof of the above lemma is not included here.

To obtain relatively short robust confidence intervals we need to use point estimates with small bias bounds. Adrover, Salibian-Barrera and Zamar (2002) used the GMS estimate for robust inference on the simple linear regression slope. As the simulation study results obtained by the authors suggest, the observed coverage levels of the robust confidence intervals are very satisfactory, and they constitute a major improvement when compared to those of the naive approach.

However, the GMS estimate has a breakdown point of only 0.25, its asymptotic normality is established under very restrictive conditions, and its bias bound is known only for symmetric carrier distributions. Therefore, a better point estimate is needed. In the next chapter, we will look for the best possible point estimate for globally robust inference on the simple linear regression slope.

Chapter 3

The RMS Approach for Robust Inference on Slope

In Chapter 2 we discussed globally robust inference for the location and the simple linear regression models. We explained how the bias bound of an estimate should be used in addition to its standard error to construct a robust confidence interval for the slope. The selection of an appropriate point estimate for the slope parameter is now an issue.

In classical inference, sampling variability of an estimate is considered to be the only source of uncertainty and, therefore, an estimator with a smaller standard error is preferred so that we can construct relatively short confidence intervals of practical relevance. Of course, the normality or asymptotic normality of an estimate is often considered for theoretical reasons.

In robust inference, on the other hand, we consider the bias of an estimator to be at least as important as its standard error (we have shown earlier that the bias of an estimator remains the same while its standard error vanishes as n goes to infinity). Therefore, to construct robust confidence intervals, we should prefer an estimate with a smaller bias bound. We should also consider the asymptotic normality of the estimator because the theory on globally robust inference developed so far is based on this assumption.

3.1 Reasons for the selection of RMS

Most of the available maxbias functions for the slope estimates have been derived assuming that the intercept parameter is known. As an exception, Hennig (1995) derived the maxbiases of MM- and τ-estimates of the intercept and the slope parameters. However, these estimators have very large bias bounds even in the case of known intercepts. The least median of squares (LMS) estimate (Rousseeuw, 1984) has the smallest maxbias in the class of residual admissible estimates with known intercepts.

However, the class of residual admissible estimates has maxbiases larger than the maxbiases of the three median-based estimates MPS, GMS and RMS. In fact, for all residual admissible estimators, the maxbias function can be expressed as follows:

$$B(\epsilon) \approx K_1\sqrt{\epsilon} + O(\sqrt{\epsilon}),$$

whereas, the maxbias function for those three median-based estimators can be expressed as

$$B(\epsilon) \approx K_2\epsilon + O(\epsilon),$$

where K_1 and K_2 are constants (Martin, Yohai & Zamar, 1989, and Berrendero & Zamar, 2001). Since $0 < \epsilon < 1$, we have $\epsilon < \sqrt{\epsilon}$, and the three median-based estimates have maxbiases considerably smaller. Therefore, for globally robust inference, it is reasonable to select one of these three estimates.

Table 3.1 below (extracted from Adrover and Zamar, 2000) displays the maxbiases of these median-based regression estimates assuming normally distributed explanatory variable and regression error under the central model. The values in the second column (labeled MS) are the lowest maximum bias attainable in the class of affine and regression equivariant estimates, when the intercept is equal to zero (Martin *et al.*, 1989). The last column (labeled LMS) gives the lowest maximum bias attainable in the class of residual admissible estimates when the intercept is equal to zero (Yohai and Zamar, 1993).

Table 3.1: Maxbiases for GMS and RMS

ϵ	GMS slope	RMS slope
0.010	0.016	0.019
0.025	0.041	0.046
0.050	0.088	0.096
0.100	0.201	0.198
0.150	0.361	0.339
0.200	0.639	0.505
0.240	1.299	0.668
0.250	∞	0.730

The maxbiases of MPS are larger than those of GMS and RMS. While GMS has slightly smaller biases than RMS for $\epsilon \leq 0.05$, RMS has smaller biases than GMS for $\epsilon \geq 0.10$. Adrover, Salibian-Barrera and Zamar (2002) selected GMS as their point estimate for robust inference on slope for the following reasons:

- GMS shows a good bias performance.
- The asymptotic normality of the GMS estimate can be proved under general conditions (allowing for most distributions in the contamination neighborhood)

We will now discuss some limitations of the GMS approach, and some advantages of RMS over GMS as a point estimate for robust inference.

Some limitations of the GMS approach

The following are some problems that we identified in the GMS approach:

- To prove the consistency and the asymptotic normality of the GMS estimates, one of the regularity conditions used by Adrover, Salibian-Barrera and Zamar (2002) is that the carrier distribution G_0 is symmetric. In practice, this condition may not be satisfied.
- The bias bound of the GMS estimate (Adrover and Zamar, 2000) is valid only when the carrier distribution G_0 is symmetric.
- GMS has a breakdown point (ϵ^*) of 0.25, while other median-based estimates have larger breakdown points.

Advantages of the RMS approach

The RMS method may be preferred to the GMS approach for the following reasons:

- The regularity conditions for the asymptotic normality of the RMS estimate (to be discussed later) are more general than those of the GMS estimate.
- The bias bound of the RMS estimate (Adrover and Zamar, 2000) is valid without the symmetry assumption of the carrier distribution G_0.
- The breakdown point of RMS is 0.50, the maximum for all affine equivariant regression estimates.
- Though for small fractions of contamination ($\epsilon \leq 0.05$) GMS has smaller biases, RMS is a very close competitor. And, for large fractions of contamination RMS

has smaller biases. In our opinion, the overall bias performance of RMS is better than that of GMS.

- RMS is both locally and globally robust.

Based on the above considerations, we decided to use RMS for the construction of robust confidence intervals, and the calculation of robust p-values. We will now discuss the bias behavior and the asymptotic properties of RMS.

3.2 Maxbias of the RMS estimate

Adrover and Zamar (2000) derived the maxbias of RMS. We will start by showing (Huber, 1981) that the maximum bias of the median functional over the ϵ-contamination neighborhood of a general distribution function G_0 (not necessarily symmetric) is attained by placing a point mass contamination at plus or minus infinity. Therefore, the maxbias of the median is given by

$$\bar{m} = \bar{m}(\epsilon) = \max \left\{ G_0^{-1}\left(\frac{1}{2(1-\epsilon)}\right) - G_0^{-1}\left(\frac{1}{2}\right), G_0^{-1}\left(\frac{1}{2}\right) - G_0^{-1}\left(\frac{1-2\epsilon}{2(1-\epsilon)}\right) \right\}. \quad (3.1)$$

To derive the maxbias of the RMS-slope, we need some notation. Let us take $q_M(a,b,H)$ and $\hat{\beta}^{RMS}(H)$ as in the definition of RMS (Chapter 1). In addition, given a general univariate distribution function F, let us define the quantiles

$$q_L(F) = F^{-1}\left(\frac{1-2\epsilon}{2(1-\epsilon)}\right) \quad \text{and} \quad q_U(F) = F^{-1}\left(\frac{1}{2(1-\epsilon)}\right). \quad (3.2)$$

Finally, let us consider

$$Q_L(a,b,H) = q_L\left(\mathcal{G}_H\left(\frac{y_1-b}{x_1-a}\right)\right) \quad \text{and} \quad Q_U(a,b,H) = q_U\left(\mathcal{G}_H\left(\frac{y_1-b}{x_1-a}\right)\right) \quad (3.3)$$

and

$$\bar{Q}_L(H) = q_L\left(\mathcal{G}_H\left(Q_L(x_2,y_2,H)\right)\right) \quad \text{and} \quad \bar{Q}_U(H) = q_U\left(\mathcal{G}_H\left(Q_U(x_2,y_2,H)\right)\right). \quad (3.4)$$

Theorem 3.1 *(Maxbias of RMS-slope) Suppose that F_0 is a symmetric distribution with unimodal density function f_0. Then, the maxbias of RMS slope estimate is*

$$B_2^{RMS}(\epsilon) = \max\left\{|\bar{Q}_L(H_0)|, |\bar{Q}_U(H_0)|\right\},$$

where $\bar{Q}_L$ and $\bar{Q}_U$ are given by (3.4).

Proof: For all H in $\mathcal{F}_\epsilon$, fixed number t and function g, we have

$$(1-\epsilon)P_{H_0}\left(g(x_1,y_1)\le t\right) \le P_H\left(g(x_1,y_1)\le t\right) \le (1-\epsilon)P_{H_0}\left(g(x_1,y_1)\le t\right)+\epsilon. \quad (3.5)$$

In particular, by taking $g(x_1,y_1) = (y_1-b)/(x_1-a)$ we get

$$(1-\epsilon)P_{H_0}\left(\frac{y_1-b}{x_1-a}\le t\right) \le P_H\left(\frac{y_1-b}{x_1-a}\le t\right) \le (1-\epsilon)P_{H_0}\left(\frac{y_1-b}{x_1-a}\le t\right)+\epsilon, \quad (3.6)$$

and

$$Q_L(a,b,H_0) \le q_M(a,b,H) \le Q_U(a,b,H_0), \quad \text{for all } H \text{ in } \mathcal{F}_\epsilon \text{ and all a, b.}$$

Since the median is a monotone operator, we also have

$$\text{Med}\,\mathcal{G}_H^*\,Q_L(x_2,y_2,H_0) \le \text{Med}\,\mathcal{G}_H^*\,q_M(x_2,y_2,H) \le \text{Med}\,\mathcal{G}_H^*\,Q_U(x_2,y_2,H_0),$$
$$\text{for all } H \text{ in } \mathcal{F}_\epsilon. \quad (3.7)$$

Moreover, using (3.7) and taking $g(x_2,y_2) = Q_L(x_2,y_2,H)$ in (3.5) we obtain

$$\bar{Q}_L(H_0) \le \text{Med}\,\mathcal{G}_H^*\,Q_L(x_2,y_2,H_0) \text{ and } \text{Med}\,\mathcal{G}_H^*\,Q_U(x_2,y_2,H_0) \le \bar{Q}_U(H_0) \text{ for all } H \text{ in } \mathcal{F}_\epsilon$$

and, therefore,

$$\bar{Q}_L(H_0) \le \hat{\beta}^{RMS} \le \bar{Q}_U(H_0) \text{ for all } H \text{ in } \mathcal{F}_\epsilon.$$

The theorem follows now because the upper and lower bounds above are attained by taking limit over a sequence of contaminated distributions.

3.3 Asymptotic properties of the RMS estimate

Siegel (1982) showed that when all x_i are distinct (an event with probability 1 if G is continuous), the RMS estimate $\widehat{\beta}_n$ has a finite-sample breakdown point $\varepsilon_n^* = [n/2]/n$, that is, if fewer than $[n/2]$ vectors $\mathbf{z}_i$ are changed, the estimate remains bounded. This yields an asymptotic breakdown point of 0.5. Siegel also showed that $\widehat{\beta}_n$ is a Fisher-consistent estimate of β.

Hössjer, Rousseeuw and Croux (1994) established the asymptotic normality of the RMS slope estimate. The authors assumed the following regularity conditions:

(**F**) The error distribution F is absolutely continuous, $F^{-1}(0.5) = 0$, and the density f is bounded ($\|f\|_\infty < \infty$) and strictly positive.

(**G**) The distribution G of the carriers is continuous, $G^{-1}(0.5) = 0$, and G has a positive and continuous density g around 0 with $g(0) > 0$.

Theorem 3.2 *Let us consider the simple linear regression model $y_i = \alpha + \beta x_i + e_i$, $i = 1, \ldots, n$, with the error and carrier distributions satisfying conditions* (**F**) *and* (**G**)*, respectively. Then*

$$\sqrt{n}(\widehat{\beta}_n - \beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathrm{IF}(x_i, y_i) + O_p(1) \xrightarrow{d} N(0, \sigma^2) \quad \text{as } n \to \infty, \tag{3.8}$$

where

$$\mathrm{IF}(x, y) = \frac{\mathrm{sign}(xy - \beta)}{2f(0)E_G|X|} \tag{3.9}$$

and

$$\sigma = \frac{1}{2f(0)E_G|X|}. \tag{3.10}$$

The authors proved this theorem through a series of lemmas. The proof is extremely long, non-trivial and involved, and is not included here.

Interestingly, the asymptotic variance for RMS (formula 3.10) is the same as the asymptotic variance for GMS derived by Adrover, Salibian-Barrera and Zamar (2002). The difference is that, the regularity conditions for the GMS variance are very restrictive while the RMS variance is derived under more general conditions.

One problem with the asymptotic normality of the RMS estimate is that the convergence to the asymptotic behavior is extremely slow. In order to check whether the asymptotic variance of the RMS slope estimate provides a good approximation to the to its variance at finite samples, Hössjer *et al.* (1994) carried out a Monte Carlo experiment. They considered both G and F to be equal to the standard Gaussian distribution, for which we get the asymptotic variance $\pi^2/4 \approx 2.47$. For each n in Table 3.1 the authors generated $m = 10,000$ samples of size n, computed the corresponding slope estimates $\hat{\beta}_n^{(k)}$ for $k = 1, 2, \ldots, m$, and obtained the n-fold variance,

$$n\ \mathrm{Var}_k(\hat{\beta}_n^{(k)}),$$

which should converge to 2.47 as n tends to ∞.

For $n \leq 40$ the n-fold variances are decreasing with n. After that, for n up to about 1000, the n-fold variances stay around 1.65, after which they slowly increase. For n around 40,000, we get a value of 1.86, which is still much less than 2.47.

Unfortunately, the approximation (to the finite sample variance) provided by the asymptotic variance of the RMS estimate is not very satisfactory. Therefore, in addition to this asymptotic variance formula, we will use the bootstrap distribution of RMS to estimate its finite sample variability in the Monte Carlo simulation in Chapter 5.

Table 3.2: Simulation results of RMS estimate

n	n-fold variance
10	2.62
20	1.88
40	1.67
60	1.67
100	1.63
200	1.63
300	1.66
500	1.64
800	1.62
1000	1.67
2000	1.83
3000	1.80
5000	1.82
10000	1.75
20000	1.85
40000	1.86
∞	2.47

Chapter 4

Application

In Chapter 3, we proposed the RMS method as an alternative to the GMS method for globally robust inference on simple linear regression slope. We will now apply these two methods, along with the classical (LS) approach and a naively robustified approach (using MM), to two real datasets.

4.1 Motorola vs Market

These data were published in Berndt (1994). The Motorola data include ten years of monthly returns of Motorola shares over the time period January 1978 to December 1987 (for 120 months). The Market data include value-weighted composite monthly market returns based on transactions of the New York Stock Exchange and the American Exchange over the same 10-year time span. The returns on 30-day US Treasury bills are also provided.

Adrover, Salibian-Barrera and Zamar (2002) used these data as an example for robust inference on simple linear regression slope with the GMS estimate. The response variable is the difference between the monthly Motorola returns and the returns on 30-day US Treasury bills. The explanatory variable is the difference between the monthly Market returns and the returns on 30-day US Treasury bills. The financial economists fit a straight line to this type of data. The slope measures the riskiness of the stock, the larger the slope the riskier the stock.

We used RMS to estimate the regression parameters. The following table gives RMS slope estimate along with GMS, MM and LS estimates.

Table 4.1: Different slope estimates

Method of Estimation	$\hat{\beta}$
RMS	1.11
GMS	1.21
MM	1.34
LS	0.85

The estimates for the intercept parameter are equal to zero (up to the second decimal place). Figure 4.1 contains a scatter plot of the data and the four fitted lines.

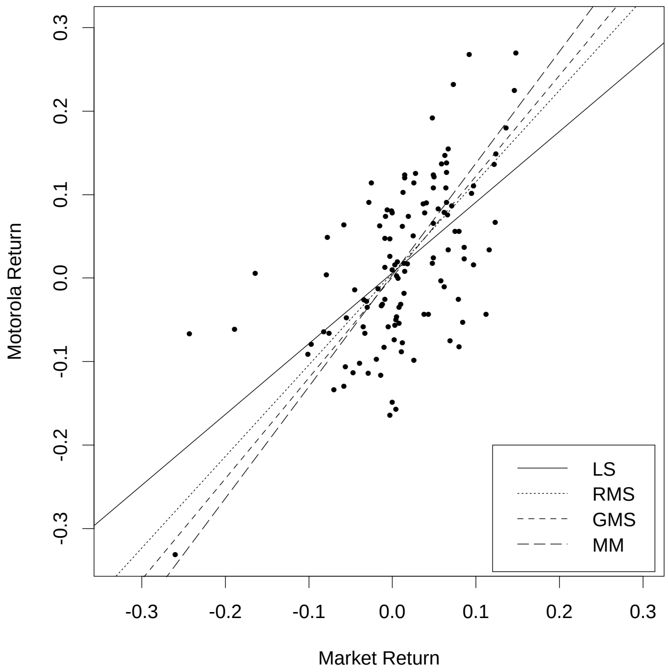
From a financial point of view, the conclusions differ very widely for the four different regression methods. According to the LS method, Motorola's stocks are safer than the market. According to RMS, GMS and MM methods, Motorola's stocks are riskier than the market.

The following hypotheses may be of possible interest in this case

$$H_0 \ : \ \beta \geq 1 \ \text{ versus } \ H_1 \ : \ \beta < 1.$$

If the null hypothesis is rejected, an investor would like to invest on this stock. The conclusions from the analyses using LS, MM, GMS and RMS methods are given below.

Figure 4.1: A scatter plot and four different regression lines



The LS method

The estimated standard error of the LS slope estimate is 0.105. The LS approach rejects H_0 at level $\alpha = 0.1$ (the p-value is 0.074). The corresponding residual plot (not shown here) does not indicate the presence of outliers.

The MM method

Adrover *et al.* used lmRobMM from Splus for this naively robust approach. The estimated standard error of the MM slope estimate is 0.274. Since "The bias is high", this approach is indecisive and the recommendation is not to perform inference based on the final estimate. If we ignore the warning and proceed to test our hypotheses, we get a p-value of 0.890 and the null hypothesis cannot be rejected.

The GMS method

For this globally robust approach, we need to determine a plausible value for ϵ. Adrover *et al.* used the GMS residual plot, which shows one clear outlier out of the 120 observations. They reasonably selected $\epsilon = 0.01$. Since we have

$$\frac{\hat{\sigma}_{\varepsilon}^{GMS}}{\hat{\sigma}_{x}^{GMS}} = 1.41,$$

the bias bound is 0.0225. The standard error of the GMS slope estimate is estimated to be 0.169 by using the shorth of the bootstrap distribution of $\hat{\beta}_n$. Therefore, the robust p-value is

$$\hat{p}^{GMS} = \Phi[(1.21 - 1 + 0.0225)/0.169] = 0.914,$$

and we cannot reject the null hypothesis. The Motorola stocks are not a safe investment.

The RMS method

To estimate the standard error of the RMS slope estimate, we used the shorth of the bootstrap distribution of $\hat{\beta}_n$, and obtained a value of 0.154. Like the GMS case, the RMS residual plot (Figure 5.2) shows one clear outlier out of the 120 observations, and we can use $\epsilon = 0.01$. We have

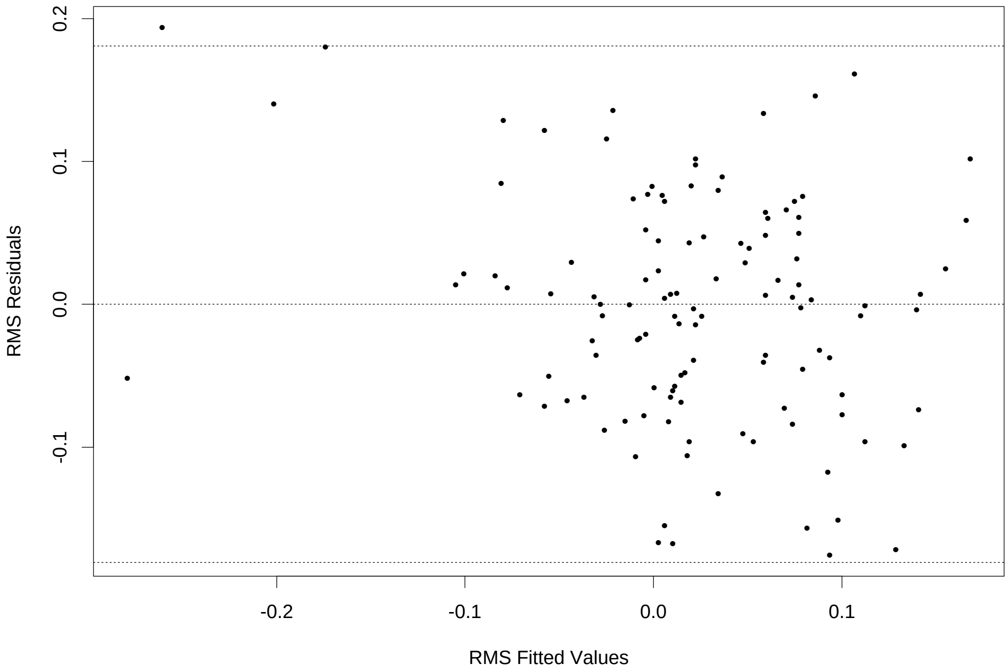
$$\frac{\hat{\sigma}_{\varepsilon}^{RMS}}{\hat{\sigma}_{x}^{RMS}} = 1.429,$$

and the bias bound is 0.0278. Therefore, the robust p-value is

$$\hat{p}^{RMS} = \Phi[(1.11 - 1 + 0.0278)/0.154] = 0.815,$$

and we cannot reject the null hypothesis. Again, the Motorola stocks are not a safe investment.

Figure 4.2: RMS residuals against RMS fitted values



Discussion

According to the LS method, the Motorola stocks are a safe investment. This method considers all the data points, and because of some outlying ‘good’ returns, the LS method leads us to a wrong conclusion.

The other three methods give very large p-values, leaving no room for rejecting the null hypothesis. However, the bias is high for the MM method, and the inference based on this naive method is not recommended.

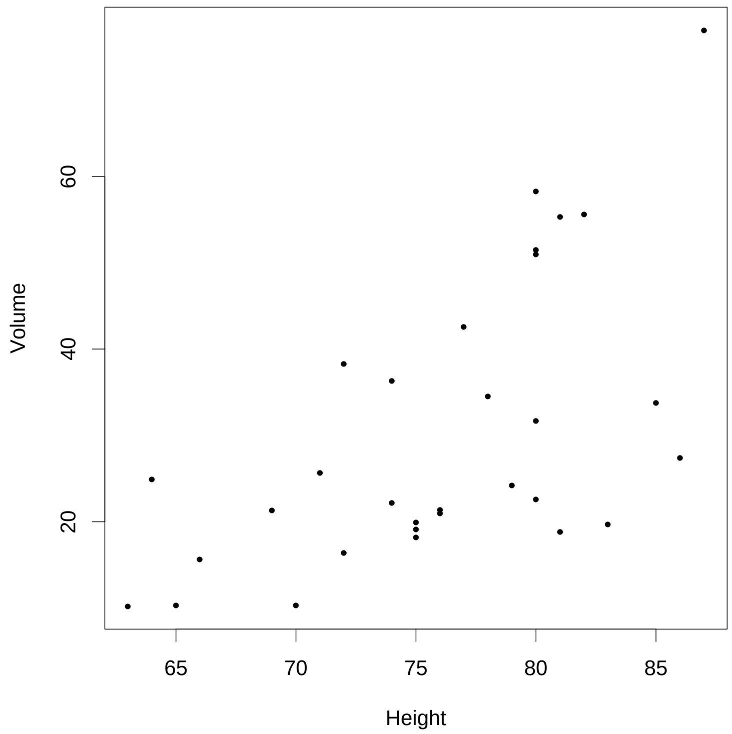
Though the GMS method gives a larger p-value as compared to the RMS method, both of these methods are very conservative, and conclude that Motorola’s stocks are NOT safer than the market.

4.2 Volume vs Height of trees

This dataset is courtesy of Dr. Harry Joe. The data consist of girth, volume and height of each of 31 trees, and are presented in Table 4.2.

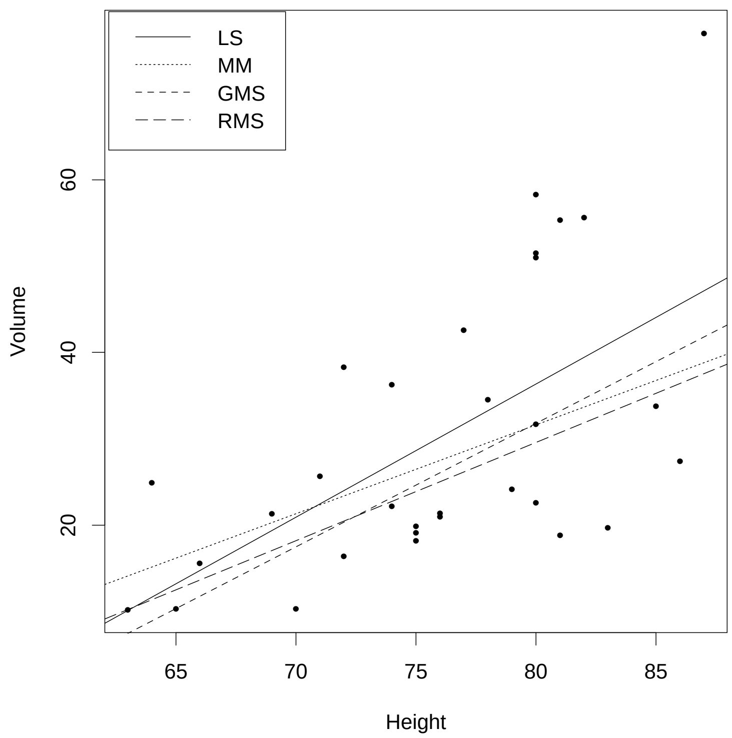
It is reasonable to assume that 'Height' is a good predictor for the response variable 'Volume'. Figure 4.3 contains a scatter plot of these two variables. A straight line seems to be a good fit for the data, except for a group of points.

Figure 4.3: A scatter plot of Volume vs Height



As in the first example, we used four methods for the estimation of the regression parameters in this case: LS, RMS, GMS and MM. Figure 5.3 shows the four fitted lines. The LS slope (1.54) is the largest of all, with the GMS slope (1.43) as the closest competitor. The RMS and the MM slopes are much smaller (1.14 and 1.03, respectively). At a first glance, this seems to be contrary to our expectation, because there is a seemingly evident linearity along the diagonal of the scatter plot and the LS line is closer to this than the robust lines are. The RMS and the MM methods seem to have missed this particular linearity!

Figure 4.4: The four regression lines



To better understand what is going on, let us look at the plot of the residuals against the fitted values for each of the four methods. Figure 5.4 presents the residual plot for the LS method. The zero line and the 2 SE line are shown, the -2 SE line is beyond the plot. One point (tree number 31) is lying outside the 2 SE limit.

Table 4.2: Girth, volume and height of trees

Tree Number	Girth	Height	Volume
1	8.3	70	10.3
2	8.6	65	10.3
3	8.8	63	10.2
4	10.5	72	16.4
5	10.7	81	18.8
6	10.8	83	19.7
7	11.0	66	15.6
8	11.0	75	18.2
9	11.1	80	22.6
10	11.2	75	19.9
11	11.3	79	24.2
12	11.4	76	21.0
13	11.4	76	21.4
14	11.7	69	21.3
15	12.0	75	19.1
16	12.9	74	22.2
17	12.9	85	33.8
18	13.3	86	27.4
19	13.7	71	25.7
20	13.8	64	24.9
21	14.0	78	34.5
22	14.2	80	31.7
23	14.5	74	36.3
24	16.0	72	38.3
25	16.3	77	42.6
26	17.3	81	55.4
27	17.5	82	55.7
28	17.9	80	58.3
29	18.0	80	51.5
30	18.0	80	51.0
31	20.6	87	77.0

Figure 4.5: LS residuals against LS fitted values

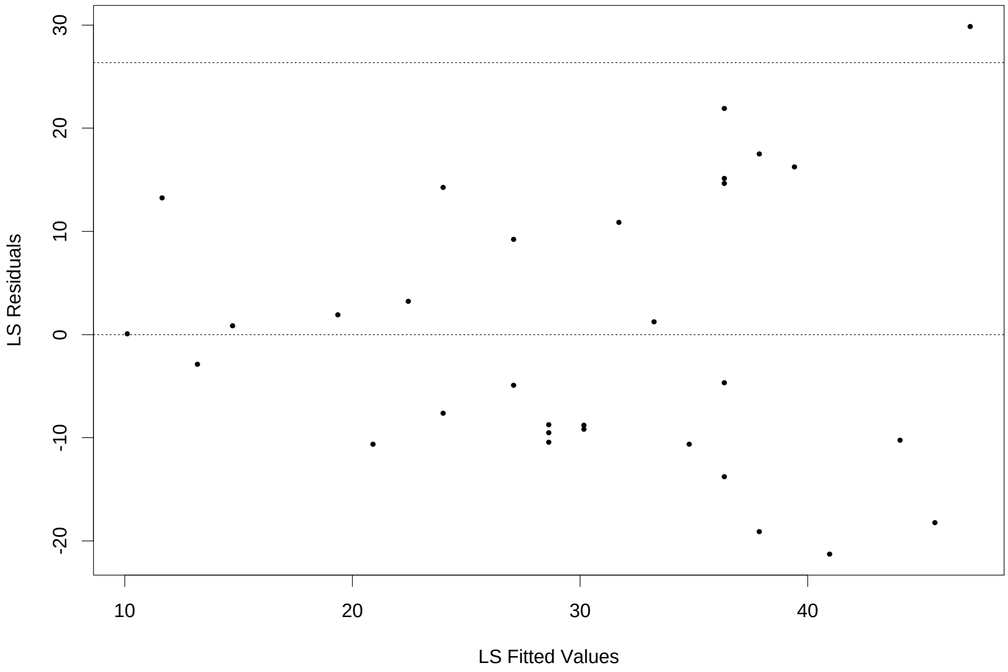


Figure 4.6: GMS residuals against GMS fitted values

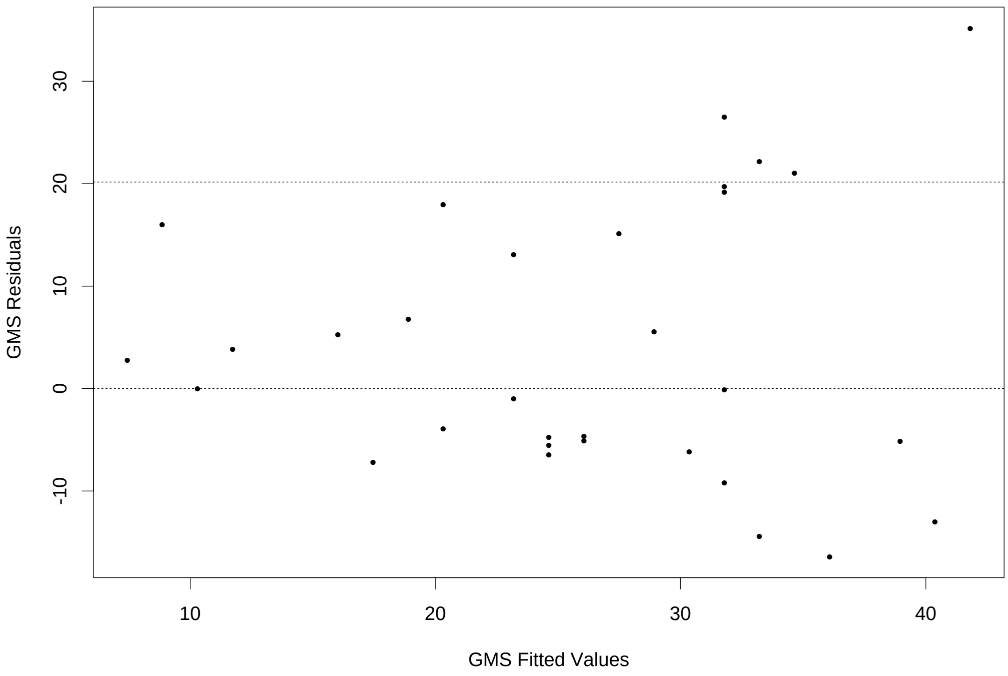


Figure 4.7: RMS residuals against RMS fitted values

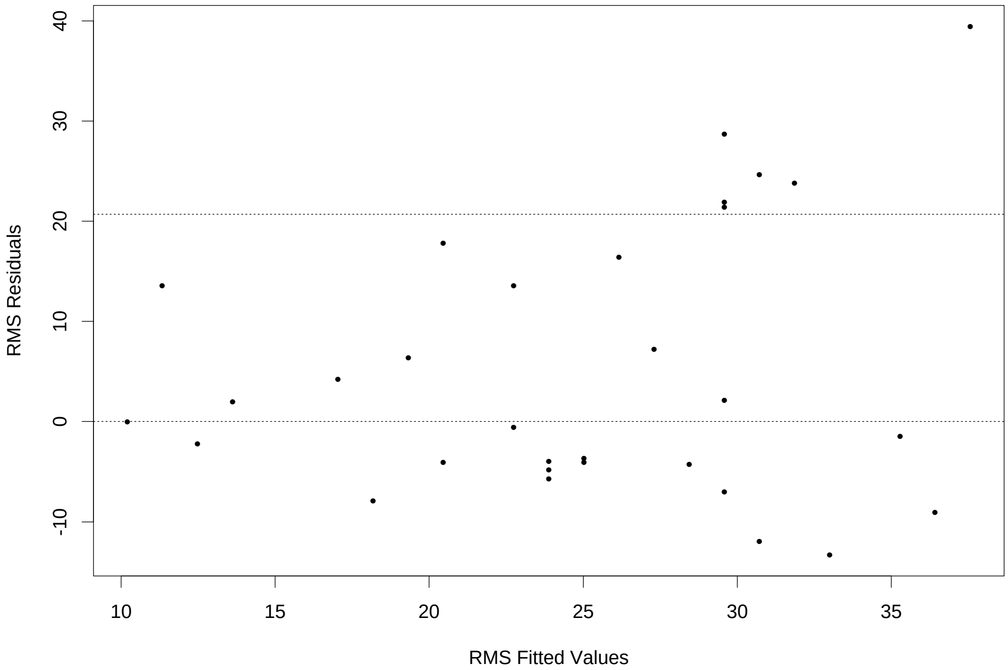
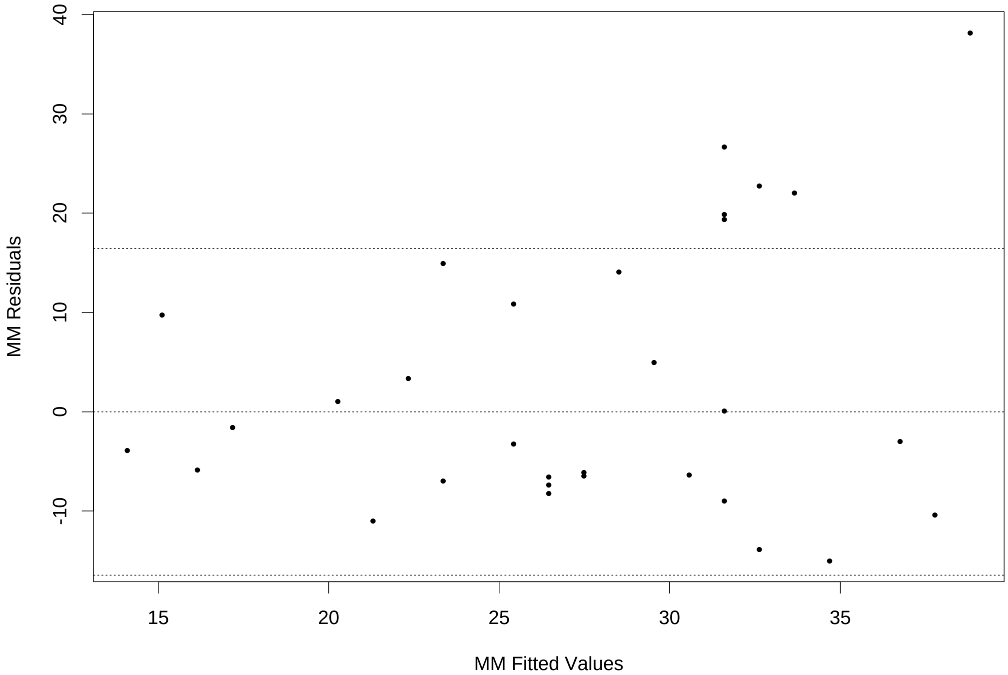


Figure 4.8: MM residuals against MM fitted values



The residual plot for the GMS method is shown in Figure 4.6. The $-2\,\mathrm{SE}$ line is beyond the plot. This time, four points (tree numbers 28, 29, 30 and 31) are lying outside the $2\,\mathrm{SE}$ limit.

Figure 4.7 and Figure 4.8 are the residual plots for the RMS and the MM methods, respectively. For the RMS method, the $-2\,\mathrm{SE}$ line is beyond the plot. The same six points (tree numbers 26, 27, 28, 29, 30 and 31) are lying outside the $2\,\mathrm{SE}$ limit for both the RMS and the MM methods.

To learn more about these six trees, we carefully scrutinized their girths, heights and volumes (Table 4.2). Surprisingly, they have the highest girth values, which means that these six trees are older than the others! It is reasonable to assume that the Height-Volume relationship for the old trees is different from that for the young trees, with a larger slope for the old ones.

Let us look again at the scatter plot (Figure 4.3), more carefully this time. The linearity exhibited along the diagonal does not represent the majority of the data. The six points corresponding to the six 'old' trees are at the end of the diagonal. If we ignore these points, a straight line with a much smaller slope seems appropriate.

There is heterogeneity in the data, and the MM and the RMS methods identify this heterogeneity almost perfectly.

Testing of hypotheses

Buyers might be interested in assessing the volume of wood in a group of trees, and they would not like to invest unless the amount of wood makes it a safe investment. Moreover, since seasoned wood is more useful, age of the trees should also be considered while making a decision. Considering these, the buyers may want to test the following hypotheses

$$H_0 \ : \ \beta \leq 1 \ \text{ versus } \ H_1 \ : \ \beta > 1.$$

If the null hypothesis is rejected, a buyer would like to go for investment considering that there is a good amount of wood and the wood is seasoned (age of the trees is also reflected in the slope parameter). The conclusions from the analyses using LS, MM, GMS and RMS methods are given below.

The LS method

The estimated standard error of the LS slope estimate is 0.3839. The LS approach rejects H_0 at level $\alpha = 0.1$ (the p-value is 0.083).

The MM method

We used lmRob from the library 'robust' of Splus6 for this naively robust approach. The estimated standard error of the MM slope estimate is 0.355. Since "The bias is high", this approach is indecisive and the recommendation is not to perform inference based on the final estimate. If we ignore the warning and proceed to test our hypotheses, we get a p-value of 0.466 and we cannot reject the null hypothesis.

The GMS method

We used the GMS estimate both in the naively robust and in the globally robust approaches:

- The standard error of the GMS slope estimate is estimated to be 0.351 by using the shorth of the bootstrap distribution of $\hat{\beta}_n$. Without any adjustment for the bias (which is equivalent to using $\epsilon = 0$), the p-value is 0.098, and H_0 is rejected at level $\alpha = 0.1$.

- For the globally robust approach, we need to determine a plausible value for ϵ. According to the GMS residual plot, there is one clear outlier out of the 31 observations. It seems reasonable to use $\epsilon = 0.03$. Since
$$\frac{\hat{\sigma}_{\varepsilon}^{GMS}}{\hat{\sigma}_{x}^{GMS}} = 1.82,$$
the bias bound is 0.0986. Therefore, the robust p-value is
$$\hat{p}^{GMS} = 1 - \Phi[(1.45 - 1 - 0.0986)/0.351] = 0.158,$$
and we cannot reject the null hypothesis at the 10% level.

The RMS method

The RMS estimate is also used both in the naively robust and in the globally robust approaches. The standard error of the RMS slope estimate is estimated to be 0.362 by using the shorth of the bootstrap distribution of $\hat{\beta}_n$.

- Without any adjustment for the bias of the point estimate, the p-value is 0.349, and H_0 cannot be rejected at level $\alpha = 0.1$.

- Like in the GMS case, the RMS residual plot shows one clear outlier out of the 31 observations. We can use $\epsilon = 0.03$. In this case,
$$\frac{\hat{\sigma}_{\varepsilon}^{RMS}}{\hat{\sigma}_{x}^{RMS}} = 1.59,$$
and the bias bound is 0.0957. Therefore, the robust p-value is
$$\hat{p}^{RMS} = 1 - \Phi[(1.14 - 1 - 0.0957)/0.362] = 0.451,$$
and we cannot reject the null hypothesis.

Discussion

According to the LS and the naive GMS methods, investment is safe. The LS method is seriously affected by a few 'old' trees – the slope estimate is larger than the 'true' value, and the investment appears to be safer than it really is. The naive GMS approach is misleading too because it does not take into account the possible bias of the estimate.

Since the estimates obtained by the MM and the RMS methods are close to 1, we get large p-values even when we use the methods naively. However, the bias test is significant for the MM method, and the inference based on this method is not recommended by Splus. About the naive RMS approach, though the p-value obtained is very large, we do not recommend this method for the following reasons. In chapter 2, we discussed theoretically the consequences of ignoring the asymptotic bias of the point estimate in robust inference. Moreover, in this particular example, we have seen the consequences of ignoring the bias in the case of naive GMS approach.

Comparing the bias-adjusted approaches based on the GMS and the RMS estimates, we strongly recommend globally robust inference with the RMS estimate.

Chapter 5

A Monte Carlo Study

In this chapter we will present the results of a simulation study that we conducted to investigate the finite sample coverage levels of globally robust confidence intervals for the slope of the simple linear regression model. Adrover, Salibian-Barrera and Zamar (2000) conducted a similar study using the GMS slope estimate. Based on the considerations of Chapter 3 of this thesis, we will use the RMS slope estimate and compare our results with those obtained by Adrover *et al.*

Though Hössjer, Rousseeuw and Croux (1994) established the asymptotic normality of the RMS estimate (and determined its asymptotic variance) without the symmetry assumption of the carrier distribution, we will consider bivariate normal distribution as our central model. We have two reasons for this decision: (1) the maxbiases of the RMS estimate are readily available for normally distributed explanatory variable and regression error under the central model (Table 3.1 extracted from Adrover and Zamar, 2001) and (2) since this is the same model that was considered by Adrover *et al.*, our study will be more comparable to their study.

5.1 Design of the study

We will consider four different choices of v_n, the estimated standard error of the RMS estimate:

Method 1 (*Empirical Asymptotic Variance*): The variability is estimated by formula 3.10. Assuming $F_0 = N(0, \sigma_\varepsilon^2)$ and $G_0 = N(0, \sigma_x^2)$, we have

$$\mathrm{Var}(\hat{\beta}) = \frac{\pi \sigma_\varepsilon^2}{2n\ \mathrm{E}_{G_0}^2 |X|} = \frac{\pi^2 \sigma_\varepsilon^2}{4n\ \sigma_x^2}.$$

Therefore,

$$v_n = \frac{\pi}{2\sqrt{n}} \frac{\hat{\sigma}_{\varepsilon,n}}{\hat{\sigma}_{x,n}} \tag{5.1}$$

Method 2 (*Classical Bootstrap*): The variability is estimated by using the standard deviation of the bootstrap distribution of $\hat{\beta}_n$.

Method 3 (*"Shorth Bootstrap"*): The variability is estimated by using the shorth of the bootstrap distribution of $\hat{\beta}_n$.

Method 4 (*"MAD Bootstrap"*): The variability is estimated by using the MAD of the bootstrap distribution of $\hat{\beta}_n$.

In our Monte Carlo simulation we used $m = 1000$ replicates of each of the following sampling situations: sample sizes n = 20, 40, 60, 80, 100 and 200 from contaminated normal distributions $(1-\epsilon)N(\mathbf{0}, I) + \epsilon N(\boldsymbol{\mu}, \tau^2 I)$ with $\boldsymbol{\mu}' = (\mu_x, \mu_y)$, $\tau = 0.1$, $\mu_x = 3$ and $\mu_y = 1.5$ (2.0) for $\epsilon = 0.05$ (0.10). We will refer to this case as the "mild contamination case". In the "medium contamination case" we took $\mu_x = 5$ and $\mu_y = 2.5$ for $\epsilon = 0.05$ and $\epsilon = 0.10$. We also considered a "strong contamination case" by taking $\mu_x = 5$ and $\mu_y = 15$ for $\epsilon = 0.05$ and 0.10.

Thus, there are three different contamination types (mild, medium and strong) with two different percentages of contamination for each type. Six different sample sizes are considered for each of these six cases. In total, there are 36 different sampling situations. The nominal confidence level in each situation is 0.95.

To estimate the variability of $\hat{\beta}_n$ by the three bootstrap approaches (Method 2, Method 3 and Method 4), we used $B = 1000$ bootstrap samples from each of the 1000 replicates of each sampling situation.

5.2 Numerical results

The observed coverage levels and the median lengths of the robust confidence intervals obtained by the four methods are presented in Tables 5.1, 5.2, 5.3 and 5.4, respectively. At first we will compare the different results in these four tables. Then we will compare our results with those obtained by Adrover *et al.*

Comparison of the four methods

Following are the major observations:

Method 1 (Table 5.1): Most of the observed coverage levels are above 95%. The median lengths reported in this table are the largest of all four methods.

Method 2 (Table 5.2): Most of the coverages are above 95%. The median lengths of this table are larger than the corresponding median lengths for Method 3 and Method 4.

Method 3 (Table 5.3): Except for some of the medium contamination cases, most of the coverages are between 90% and 95%. The median lengths in this table are the smallest.

Table 5.1: Coverage proportion (median length) of robust CI for the slope by Method 1.

% of Contamination	Sample Size	Mild Contamination	Medium Contamination	Strong Contamination
5%	20	0.92 (1.28)	0.96 (1.28)	0.91 (1.37)
	40	0.96 (0.99)	0.94 (1.02)	0.95 (1.03)
	60	0.97 (0.86)	0.96 (0.86)	0.95 (0.89)
	80	0.97 (0.77)	0.96 (0.79)	0.96 (0.80)
	100	0.96 (0.71)	0.96 (0.72)	0.96 (0.74)
	200	0.96 (0.57)	0.98 (0.58)	0.97 (0.58)
10%	20	0.89 (1.39)	0.89 (1.36)	0.91 (1.68)
	40	0.96 (1.16)	0.95 (1.18)	0.97 (1.34)
	60	0.96 (1.09)	0.91 (1.07)	0.95 (1.19)
	80	0.93 (1.04)	0.94 (1.01)	0.97 (1.11)
	100	0.93 (0.97)	0.96 (0.98)	0.97 (1.05)
	200	0.97 (0.85)	0.93 (0.84)	0.98 (0.89)

Method 4 (Table 5.4): Most of the coverages are between 90% and 95%. The median lengths for this method are smaller than those of Method 1 and Method 2, but larger than those of Method 3.

Table 5.1 shows some overcoverages and the lengths of the confidence intervals are also relatively large. An explanation of this may be found in Table 3.1, which shows that the asymptotic variance (formula 3.10) overestimates the variablity of $\hat{\beta}_n$ exhibited in finite samples. Similarly, the overcoverages shown in Table 5.2 and the undercoverages in Table 5.3 are due to the overestimation and the underestimation, respectively, of the 'true' standard error of $\hat{\beta}_n$.

The observed coverage levels for Method 4 are mostly close to the nominal level. This method seems to estimate the standard error of $\hat{\beta}_n$ better than the other methods.

Table 5.2: Coverage proportion (median length) of robust CI for the slope by Method 2.

% of Contamination	Sample Size	Mild Contamination	Medium Contamination	Strong Contamination
5%	20	0.97 (1.51)	0.97 (1.49)	0.98 (1.91)
	40	0.95 (0.95)	0.94 (0.97)	0.97 (1.10)
	60	0.93 (0.80)	0.94 (0.80)	0.97 (0.89)
	80	0.96 (0.71)	0.93 (0.72)	0.96 (0.78)
	100	0.94 (0.65)	0.95 (0.67)	0.96 (0.71)
	200	0.96 (0.53)	0.96 (0.54)	0.97 (0.55)
10%	20	0.94 (1.63)	0.94 (1.55)	0.99 (2.51)
	40	0.97 (1.21)	0.95 (1.14)	0.99 (1.57)
	60	0.96 (1.09)	0.88 (1.02)	0.98 (1.30)
	80	0.94 (1.02)	0.93 (0.96)	0.99 (1.16)
	100	0.93 (0.96)	0.94 (0.93)	0.99 (1.08)
	200	0.96 (0.84)	0.91 (0.81)	0.99 (0.90)

Comparison of RMS and GMS

For the results of the GMS approach, the three tables provided by Adrover *et al.* (2002) are referred to.

The comparison of Method 1 of RMS and that of GMS is interesting. If we focus on the median lengths, we notice that they are more or less equal. However, if we concentrate on the observed coverage levels, we see that for RMS they are mostly over 95% while for GMS they are mostly below 95%.

The explanation is as follows. We mentioned earlier that the asymptotic variance formula for these two estimates are the same (derived under two different sets of regularity conditions, though). For RMS, this formula overestimates the variability (Table 3.1) while

Table 5.3: Coverage proportion (median length) of robust CI for the slope by Method 3.

% of Contamination	Sample Size	Mild Contamination	Medium Contamination	Strong Contamination
5%	20	0.88 (1.19)	0.89 (1.15)	0.94 (1.41)
	40	0.91 (0.86)	0.87 (0.88)	0.94 (0.98)
	60	0.90 (0.76)	0.91 (0.76)	0.94 (0.82)
	80	0.93 (0.68)	0.92 (0.70)	0.95 (0.73)
	100	0.92 (0.62)	0.93 (0.65)	0.94 (0.67)
	200	0.94 (0.52)	0.96 (0.53)	0.96 (0.54)
10%	20	0.85 (1.36)	0.85 (1.25)	0.96 (1.84)
	40	0.90 (1.13)	0.84 (1.03)	0.98 (1.37)
	60	0.90 (1.05)	0.84 (0.97)	0.96 (1.21)
	80	0.94 (1.00)	0.90 (0.93)	0.97 (1.12)
	100	0.93 (0.94)	0.91 (0.91)	0.98 (1.04)
	200	0.94 (0.82)	0.90 (0.80)	0.98 (0.88)

for GMS it underestimates the variability (Adrover *et al.*, 2002). The maxbiases of these two estimates, on the other hand, are very close (slightly smaller for GMS for $\epsilon = 0.05$, and slightly smaller for RMS for $\epsilon = 0.10$). Therefore, for the same lengths, RMS has overcoverage while GMS has undercoverage.

One point is evident from the above comparison. The finite sample variability of RMS is less than that of GMS, which is also reflected in Method 2 and Method 3 results of the two estimates. For each of these methods, the median lengths obtained by RMS are smaller than those obtained by GMS, while the observed coverage levels are more or less equal.

Method 4 is not comparable since Adrover *et al.* did not consider this method for the GMS variability estimation.

Table 5.4: Coverage proportion (median length) of robust CI for the slope by Method 4.

% of Contamination	Sample Size	Mild Contamination	Medium Contamination	Strong Contamination
5%	20	0.91 (1.28)	0.93 (1.26)	0.96 (1.53)
	40	0.92 (0.90)	0.91 (0.93)	0.95 (1.02)
	60	0.92 (0.79)	0.93 (0.79)	0.95 (0.85)
	80	0.94 (0.70)	0.93 (0.72)	0.96 (0.76)
	100	0.93 (0.64)	0.95 (0.67)	0.95 (0.69)
	200	0.95 (0.52)	0.96 (0.54)	0.97 (0.55)
10%	20	0.92 (1.47)	0.88 (1.35)	0.97 (1.99)
	40	0.93 (1.19)	0.90 (1.90)	0.99 (1.43)
	60	0.94 (1.08)	0.88 (1.02)	0.97 (1.24)
	80	0.94 (1.02)	0.93 (0.96)	0.98 (1.14)
	100	0.94 (0.96)	0.95 (0.93)	0.99 (1.06)
	200	0.94 (0.83)	0.91 (0.81)	0.98 (0.89)

5.3 Discussion

The numerical results in the four tables show that the coverage of the robust confidence intervals are in general fairly good (close to the nominal level) and constitute a major improvement when compared to those achieved by the naive procedure (Table 2.2).

Asymptotic variance of RMS overestimates its finite sample variability, so does the standard deviation of the bootstrap distribution of $\hat{\beta}_n$. On the other hand, shorth of the bootstrap distribution underestimates the variability. The performance of MAD bootstrap seems to be better than the others, the coverages for this method are closer to the nominal level than those for the other methods. We would recommend this method for the estimation of the finite sample variability of RMS.

The overall performance of RMS is better than that of GMS. Either the observed coverages of RMS intervals are larger with the same lengths, or the lengths of RMS are smaller while maintaining the same coverages.

Based on these considerations, we recommend RMS for robust inference on the slope of simple linear regression model.

Chapter 6

Conclusion

In this concluding chapter we will summarize our findings and identify some topics for further research.

6.1 Summary

The main results of this thesis may be summarized as follows:

1. In the construction of robust confidence intervals, if we ignore the uncertainty due to the bias of the point estimate, we will get asymptotic coverage level zero.

2. To incorporate the bias bound of an estimate in addition to its standard error in robust inference, the method proposed by Adrover, Salibian-Barrera and Zamar (2002) may be used.

3. A point estimate that has an asymptotically normal distribution and a relatively small bias bound should be preferred.

4. For robust inference on the simple linear regression slope, the problems of the GMS estimate proposed by Adrover *et al.* (2002) are that it has a breakdown point of 0.25, and its asymptotic normality is established under very restrictive conditions.
5. In this study, we proposed the RMS estimate, for which the breakdown point is 0.50, and the asymptotic normality holds under very general conditions.

We applied the RMS method to two real datasets. In the first example, the RMS method performed almost as well as the GMS approach, both of them concluded that the investment was risky while the least squares approach indicated otherwise. The second example was a more challenging problem, and the RMS method performed much better than the GMS. The outlying 'old' trees were identified by RMS almost perfectly, and while GMS was 'shaky' in making a decision (as compared to RMS), RMS was clearly conservative and considered the investment to be NOT good.

In the Monte Carlo study, the RMS method achieved, more or less, the same observed coverage levels while it constructed intervals of smaller lengths, as compared to GMS. Regarding the four methods of estimation of the standard error of RMS, the asymptotic variance formula and the classical bootstrap were overestimating while the shorth bootstarp was underestimating. The MAD bootstrap may be preferred.

Based on our findings, we recommend RMS for globally robust inference on the simple linear regression slope.

6.2 Further study

The following points form some interesting areas for future research:

1. For prediction with simple linear regression models, we may be interested in the

linear combination of the slope and the intercept parameters. The asymptotic bivariate normality of these parameters may be established for this purpose. We will also need the appropriate bias bound for the intercept parameter.

2. The bias bounds of the RMS slope and intercept parameters are available only under the assumptions of normally distributed explanatory variable and regression error under the central model. These bounds may be obtained under more general conditions, for example, without the symmetry assumption of the carrier distribution.

3. RMS can also be used for estimating the slope parameters in multiple linear regression, using kernel functions with more than two arguments (Siegel, 1982). However, these estimators are not affine equivariant when the number of slope parameters is two or more. The asymptotic properties of RMS in higher dimensions constitute another interesting area of study.

Bibliography

Adrover, J. G., Salibian-Barrera, M., and Zamar, R. H. (2002). Globally robust inference for the location and simple linear regression models. *J. Statist. Plann. Inference*: (accepted for publication).

Adrover, J. G. and Zamar, R. H. (2000). Bias robustness of three median-based regression estimates. Technical Report No. 194, Department of Statistics, University of British Columbia, Canada.

Berrendero, J. R. and Zamar, R. H. (2001). Maximum bias curves for robust regression with non-elliptical regressors. *Ann. Statist.*, **29**: 224–251.

Boscovich, R. J. (1757). De litteraria expeditione per pontificiam ditionem, et synopsis amplioris operis. *Bononiensi Scientiarum et Artum Instituto atque Academia Commentarii*, **4**: 353–396.

Brown, G. W. and Mood, A. M. (1951). On median tests for linear hypotheses, in Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Univ. of California Press, Berkeley. pages 159–166.

Fraiman, R., Yohai, V. J., and Zamar, R. (2001). Optimal robust M-estimates of location. *Ann. Statist.*, **29**: 194–223.

Frees, E. (1991). Trimmed slope estimates for simple linear regression. *J. Statist. Plann. Inference*, **27**: 203–221.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.*, **69**: 383–393.

Hössjer, O., Rousseeuw, P. J., and Croux, C. (1994). Asymptotics of the repeated median slope estimator. *Ann. Statist.*, **22**: 1478–1501.

He, X. and Simpson, D. G. (1993). Lower bounds for contamination bias: Globally minimax versus locally linear estimation. *Ann. Statist.*, **21**: 314–337.

Hennig, C. (1995). Efficient high-breakdown point estimators in robust regression: Which function to choose? *Statistics & Decisions*, **13**: 221–241.

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.*, **35**: 73–101.

Huber, P. J. (1981). *Robust Statistics.* Wiley, New York.

Martin, R. D., Yohai, V. J., and Zamar, R. H. (1989). Min-max bias robust regression. *Ann. Statist.*, **17**: 1608–1630.

Martin, R. D. and Zamar, R. H. (1989). Asymptotically min-max bias robust M-estimates of scale for positive random variables. *J. Amer. Statist. Assoc.*, **84**: 494–501.

Martin, R. D. and Zamar, R. H. (1993). Bias-robust estimates of scale. *Ann. Statist.*, **21**: 991–1017.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust regression and outliers detection.* Wiley, New York.

Rousseeuw, P. J. and Yohai, V. J. (1984). Robust regression by means of S-estimators. *Robust and Nonlinear Time Series Analysis* (J. Franke, W. Härdle, and R. D. Martin, *eds.*), Lecture Notes in Statistics **26**, Springer Verlag, New York: 256–272.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Statist. Assoc.*, **63**: 1379–1389.

Siegel, A. F. (1982). Robust regression using repeated medians. *Biometrika*, **69**: 242–244.

Stigler, S. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900.* Harvard Univ. Press.

Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis, I, II and III. *Koninklijke Nederlandse Akademie van Wetenschappen, Proceedings*, **53**: 386–392; 521–525; 1397–1412.

Tukey, J. (1960). A survey of sampling from contaminated distributions, in: *Contributions to Probability and Statistics.* I. Olkin, Ed., Stanford University Press, Stanford.

Yohai, V. J. (1987). High breakdown point and high efficiency robust estimates for regression. *Ann. Statist.*, **15**: 642–656.

Yohai, V. J. and Zamar, R. H. (1993). A minimax bias property of the least α-quantile estimates. *Ann. Statist.*, **20**: 1875–1888.