Robust Second-order Least Squares Estimation for Linear Regression Models

by

Xin Chen

B.Math., University of Waterloo, 2007

A Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Mathematics and Statistics

University of Victoria

Robust Second-order Least Squares Estimation for Linear Regression Models

by

Xin Chen

B.Math., University of Waterloo, 2007

Supervisory Committee

Dr. Julie Zhou, Co-supervisor

(Department of Mathematics and Statistics)

Dr. Min Tsao, Co-supervisor

(Department of Mathematics and Statistics)

Dr. Farouk Nathoo, Departmental Member

(Department of Mathematics and Statistics)

Supervisory Committee

Dr. Julie Zhou, Co-supervisor

(Department of Mathematics and Statistics)

Dr. Min Tsao, Co-supervisor

(Department of Mathematics and Statistics)

Dr. Farouk Nathoo, Departmental Member

(Department of Mathematics and Statistics)

ABSTRACT

The second-order least-squares estimator (SLSE), which was proposed by Wang (2003), is asymptotically more efficient than the least-squares estimator (LSE) if the third moment of the error distribution is nonzero. However, it is not robust against outliers. In this paper, we propose two robust second-order least-squares estimators (RSLSE) for linear regression models, RSLSE-I and RSLSE-II, where RSLSE-I is robust against X-outliers and RSLSE-II is robust against X-outliers and Y-outliers. The basic idea is to choose proper weight matrices, which give a zero weight to an outlier. The RSLSEs are asymptotically normally distributed and are highly efficient with high breakdown point. Moreover, we compare the RSLSEs with the LSE, the SLSE and the robust MM-estimator through simulation studies and real data examples. The results show that they perform very well and are competitive to other robust regression estimators.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Dr. Julie Zhou and Dr. Min Tsao who have guided me into the field of statistics, provided many ideas and offered infinite help and patience during my study at the University of Victoria. I would like to acknowledge my gratitude to my committee member, Dr. Farouk Nathoo who has been very helpful in assisting in the completion of the thesis. I also like to thank all my family members for supporting me all the time.

Chapter 1

Introduction

Regression analysis is an important statistical tool with many applications in sciences. There are different methods for parameter estimation and inference for regression models but the least-squares (LS) method is by far the most important and widely used method. The LS method was developed in the early eighteenth century and statistical properties of the least-squares estimator (LS estimator) are well-known (Rao, Toutenburg, Fieger, Heumann, Nittner and Scheid, 1999). The LS method estimates the parameters by minimizing the sum of the squared differences between the response variable and its conditional mean given the explanatory variables. When the random errors are normally distributed, the LS estimator has the smallest variance and is the most efficient (Bain and Engelhardt, 1992). However, in practice the normality assumption of the random errors is not always valid. If the random error term is not normally distributed, in particular, if the error distribution is skewed, the LS estimator is not the most efficient.

To find more efficient estimators, it is necessary to consider higher order conditional moments of the response variable. Wang (2003) proposed a new estimation method, the *second-order least-squares (SLS) method* which makes use of the first two conditional moments of the response variable. The SLS method was first used to deal with measurement error problems in nonlinear regression models. It extends the LS method by also taking into consideration the differences between the squared response variable and its conditional moment given the explanatory variables. Wang and Leblanc (2008) compared the second-order least-squares estimator (SLS estimator) with the LS estimator for linear and nonlinear models and proved that the SLS estimator is asymptotically more efficient than the LS estimator when the third moment of the random error is not zero.

Both the LS and SLS methods are concerned with the ideal setting where all data points are generated by the underlying regression model. In particular, there are no outliers in the data and hence the robustness of the LS and SLS estimators is not an issue in this ideal setting. Nevertheless, outliers do occur in real data and they may have substantial influence on these estimators. Moreover, when using the LS method the outliers cannot always be detected through the LS residual plots. Hence one cannot always rely on simple plots to spot outliers. To protect against the presence and influence of the outliers, many robust statistical methods have been developed which can handle the outliers automatically without the need to identify and remove them individually. Robust alternatives to the LS estimator include the S-estimator of Rousseeuw and Yohai (1984), the least trimmed squares estimator of Rousseeuw (1984) and the MM-estimator of Yohai (1987). There are other studies using parametric regression to deal with outliers in the errors, for example, Azzalini and Capitanio (2003). However the parametric regression may not be robust against X-outliers. The SLS method is a promising new method for fitting regression models but its robustness has not been previously studied. As an extension of the LS method, it inherits the non-robustness of the latter. The objective of this thesis is to address the non-robustness of the SLS method for the case of linear regression models. To this end, we demonstrate the non-robustness of the SLS method through the concept of an influence function. More importantly, we propose two robust extensions of the SLS method for linear models and demonstrate through a comprehensive simulation study that such robust SLS methods perform very well in the presence of outliers. We also investigate the asymptotic properties of the robust SLS estimators.

The rest of this chapter is organized as follows. In Section 1.1, we define the general linear regression model and describe the classical LS method. In Section 1.2, we introduce the SLS method and demonstrate its close connection with the LS method through one real data example. In Section 1.3, we demonstrate the non-robustness of these two methods through one real life data set which contains an outlier. We conclude this chapter with an overview of the thesis in Section 1.4.

1.1 Method of ordinary least-squares

A linear regression model describes a linear relationship between a dependent variable and one or more independent variables. The dependent variable is also called the response variable, and the independent variables are also referred to as the explanatory variables. Consider a multiple linear regression model,

$$Y_i = \theta_0 + \theta_1 X_{i1} + \theta_2 X_{i2} + \cdots + \theta_{p-1} X_{i(p-1)} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

where Y_i is the i^{th} response, $X_{i1}, \ldots, X_{i(p-1)}$ are the associated explanatory variables, ε_i is a random error with mean $E(\varepsilon_i) = 0$ and variance $Var(\varepsilon_i) = \sigma^2$. The $\theta_0, \theta_1, \ldots, \theta_{p-1}$ are the unknown regression parameters. For convenience, we will use $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_{p-1})^T$ to represent the vector of regression parameters, and use $X_i = (1, X_{i1}, \ldots, X_{i(p-1)})^T$ to represent the vector of all independent variables. Denote by $g(X_i, \boldsymbol{\theta})$ the conditional expectation of Y_i on X_i. By (1.1), we have

$$g(X_i, \boldsymbol{\theta}) = E(Y_i|X_i) = \theta_0 + \theta_1 X_{i1} + \ldots + \theta_{p-1} X_{i(p-1)},$$

which is a linear function of $\theta_0, \ldots, \theta_{p-1}$. Once we replace these parameters with their estimated values, we have an estimated model and the goal of a regression analysis is to find the estimated model that will fit the data the best according to some criterion. There are various criteria that can be used to measure the fit of a model. In the popular LS method, the criterion used is the sum of squared errors, $S_n(\boldsymbol{\theta})$, given by

$$S_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - g(X_i; \boldsymbol{\theta}))^2. \tag{1.2}$$

Function $S_n(\boldsymbol{\theta})$ is the summation of the squared random errors, which is also equivalent to the summation of the squared differences between the response variable and its conditional mean given the explanatory variable. The LS estimator $\hat{\boldsymbol{\theta}}_{LS}$ for $\boldsymbol{\theta}$ is defined as the value that minimizes $S_n(\boldsymbol{\theta})$.

This estimator is the most efficient when the distribution of the error term is normal. However, in many practical applications, the error distribution is not normal. It may even be asymmetric. In such cases, $\hat{\boldsymbol{\theta}}_{LS}$ is not efficient and we need to use more information from the data to improve the efficiency. One method that uses higher moments of the data to produce more efficient estimates is the SLS method of Wang (2003), which we now discuss.

1.2 Second-order least-squares method

The SLS method was initially proposed by Wang (2003) to handle the measurement error in the independent variables. In Wang and Leblanc (2008) this method was extended to handle the estimation and inference for linear and non-linear regression models. In order to apply the SLS method to (1.1), we assume that Y and ε each has a finite fourth moment and $(Y_i, X_i^T)^T$, $i = 1, \ldots, n$, are identically and independently distributed (i.i.d.). The first two conditional moments of Y given X are, respectively,

$$E(Y|X) = g(X; \boldsymbol{\theta}) \quad \text{and} \quad E(Y^2|X) = g^2(X; \boldsymbol{\theta}) + \sigma^2.$$

For convenience, we will use γ to represent the vector of all parameters, that is,

$$\gamma = (\boldsymbol{\theta}^T, \sigma^2)^T = (\theta_0, \theta_1, \ldots, \theta_{p-1}, \sigma^2)^T. \tag{1.3}$$

It follows that the parameter space $\Gamma = \Theta \times \Sigma \subset R^{p+1}$.

Wang and Leblanc (2008) defined the SLS estimator $\hat{\gamma}_{SLS}$ for γ as the measurable function that minimizes objective function

$$Q_n(\gamma) = \sum_{i=1}^{n} \rho_i^T(\gamma) W_i \rho_i(\gamma), \tag{1.4}$$

where

$$\rho_i(\gamma) = (Y_i - g(X_i; \boldsymbol{\theta}), Y_i^2 - g^2(X_i; \boldsymbol{\theta}) - \sigma^2)^T, \tag{1.5}$$

and $W_i = W(X_i)$ is a 2×2 non-negative definite weight matrix which may depend on X_i. By minimizing $Q_n(\gamma)$ in (1.4), the SLS method computes the estimates for parameters $\boldsymbol{\theta}$ and σ^2 simultaneously. Since $\rho_i(\gamma)$ in (1.5) contains the difference between the response variable and its first conditional moment *as well as* the difference of the squared response variable and its second conditional moment, the resulting objective function $Q_n(\gamma)$ is clearly an extension of the least-squares objective function $S_n(\boldsymbol{\theta})$ in (1.2), which involves only the first difference. In particular, when the weight matrix W_i takes the following special form,

$$W_i = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

the objective function $Q_n(\gamma)$ reduces to $S_n(\boldsymbol{\theta})$ and the SLS method reduces to the LS method. Of course this weight matrix is never used in the SLS method, and much of the theory of the SLS method centres on finding an optimal weight matrix which will minimize the variance of the SLS estimator. See Chapter 3. When the third moment of the random error is non-zero, Wang and Leblanc (2008) showed that the optimal SLS estimator (the SLS estimator based on the optimal weight matrix) is asymptotically more efficient than the LS estimator. Further, when the third moment of the random error is zero (which is the case when the error distribution is normal, favouring the LS estimator), the variance of the optimal SLS estimator is asymptotically equivalent to that of the LS estimator. These are very appealing properties as they showed that the SLS estimator is more efficient than the LS estimator in general and is just as efficient for cases where the LS estimator is the optimal.

It is difficult to quantify the improvement in efficiency of the SLS estimator over the LS estimator for finite sample cases, which will depend on the extend to which the random error deviates from the normality assumption underlying the LS method. But it is not difficult to see clear evidence of such improvement through simulations. Wang and Leblanc (2008) gave some numerical examples which demonstrate the superiority of the SLS estimator. Instead of repeating such numerical examples, here we choose a real life data set to show that under the normality assumption, the SLS estimator is very close to the LS estimator and hence little may be lost for using the SLS method in general.

Example 1.1. *Table 1.1 contains the shelf-stocking data set from Montgomery, Peck and Vining (2006, P45-56), consisting of 15 observations of the time required for a merchandiser to stock a shelf with a soft drink product (Y) and the number of cases of the product stocked (X).*

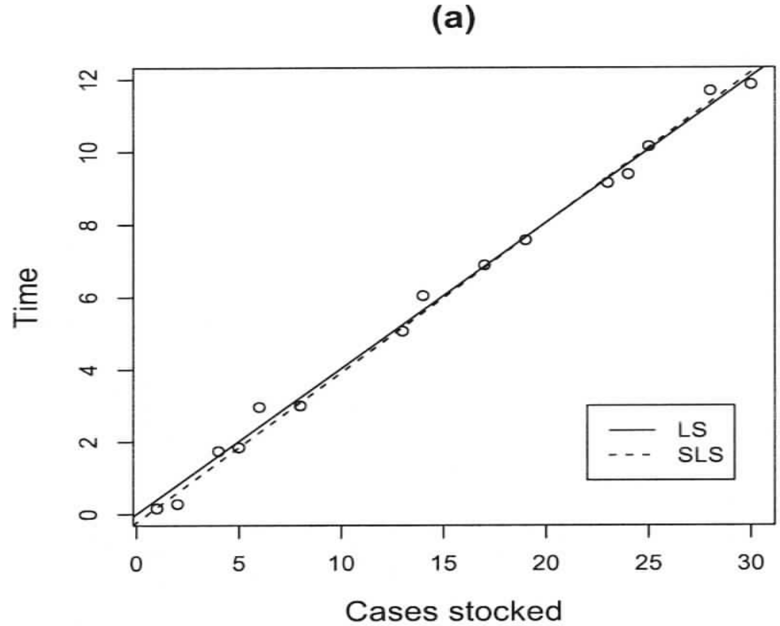
Table 1.1: Shelf-stocking data for Example 1.1

Time, Y(minutes)	Cases Stocked, X
10.15	25
2.96	6
3.00	8
6.88	17
0.28	2
5.06	13
9.14	23
11.86	30
11.69	28
6.04	14
7.57	19
1.74	4
9.38	24
0.16	1
1.84	5

Table 1.2: Estimated linear model parameters with standard errors (in brackets) for the shelf-stocking data in Example 1.1: LSE – least-squares estimates; SLSE – second-order least-squares estimates.

	Intercept(θ_0)	Slope(θ_1)
LSE	-0.0938	0.4071
	(0.1436)	(0.0082)
SLSE	-0.2318	0.4142
	(0.1259)	(0.0074)

The scatter plot in Figure 1.1(a) shows a clear linear relationship between Y and X, supporting a linear model for the data set. To compare the SLS method with LS



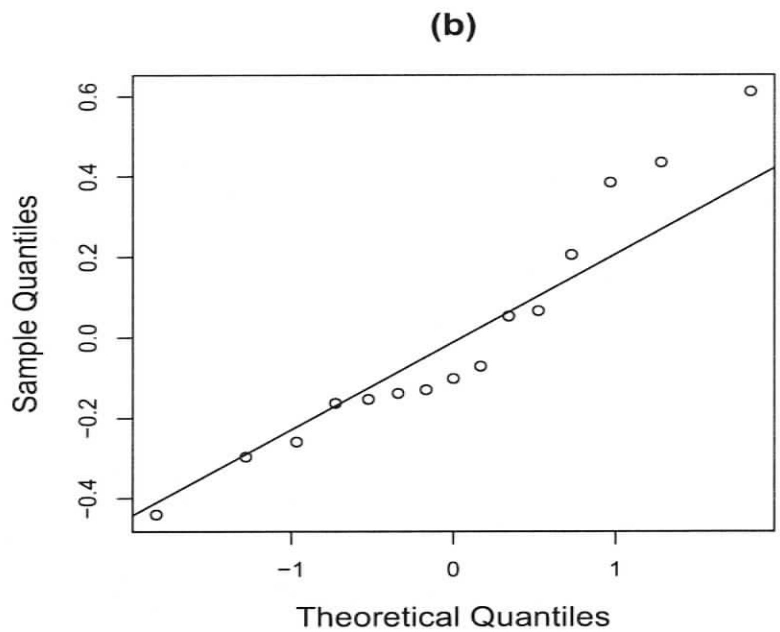


Figure 1.1: Analysis for the shelf-stocking data in Example 1.1: (a) scatter diagram of the data points and the fitted lines, (b) Q-Q plot of the LS residuals.

method, we used both to fit the linear model to this data set. The solid line passing through most of the data points is the LS fitted line. The SLS fitted line is the dashed line, which also provides a good fit and agrees with the LS line very well. The Q-Q plot of the residuals from the LS fit in Figure 1.1(b) does not show obvious departures from normality. A Shapiro's normality test on these residuals was conducted, yielding a p-value of 0.2937. Both the residual plot and the p-value suggest that the normality assumption of the linear model is reasonable. Table 1.2 lists the estimates given by LS and SLS estimators, which are similar. Noting that this is one of the cases where the normality assumption is reasonable and the LS method is expected to be the optimal, the observation that the SLS results are close to the LS results supports Wang and Leblanc's (2008) theoretical finding that SLS is competitive even in cases when the normality assumption hold.

Nevertheless, as we have previously noted that the LS method and SLS method are not robust in that one single bad outlier will render the LS and SLS estimates useless. We now illustrate this point with an numerical example in the next section.

1.3 Impact of an outlier on the least-squares and second-order least-squares estimators

To illustrate the non-robust nature of LS and SLS methods, we apply both methods to a well known data set in Example 1.2, which contains one outlier. The outlier is shown to have a dramatic impact on the LS and SLS estimates, leading to extremely poor fit. As a contrast, we also apply the robust MM-method (Yohai, 1987) to this example and show that the MM fitted line accurately captures the trend of the data. The focus of this section is the numerical comparison of the methods. Discussions about the MM-method will be given in Chapter 2.

Example 1.2. *The water flow measurements through the towns of Libby (X) and Newgate (Y) on the Kootenay river in January for years 1931-1943 are given in Table 1.3. The X and Y variables are highly positively correlated. The original data came from Ezekiel and Fox (1959, p.57 - 58). In Hampel et al. (1986), the 1934 measurement for Newgate was changed to 15.7 to create an outlier. This altered data set has since been used by other authors,* e.g., *Rousseeuw and Leroy (1987).*

Figure 1.2(a) shows the plot of water flow measurements for Libby verses the corresponding same year measurements for Newgate. The overall relationship is roughly linear but there is a clear outlier $(77.6, 15.7)$ in the lower-right conner. The solid line is the LS fitted line, which has failed to capture the clear increasing trend of the majority of the data points. Rather, it is heavily influenced by the outlier at the lower-right corner and it shows a decreasing trend. The dashed line is the SLS fitted line. Like the LS line, it too fails to capture the increasing trend of the data. It looks worse than the LS fitted line in that its decreasing trend is even more pronounced than that of the LS line. On the other hand, fitting the linear model using the robust MM-method produced the dotted line which provides the good fit to the data set. Figure 1.2(b) shows the plot of the residuals verses the fitted values for the MM-estimated model. It reveals the existence of one outlier as indicated by the large residual at the lower-right conner. However, the plot of residuals verses the fitted values for the SLS estimator shown in Figure 1.2(c) fails to identify the outlier. Finally, the LS estimates, the SLS estimates and MM-estimates are listed in Table 1.4. Comparing MM-estimates with the LS and SLS estimates, the LS and SLS estimates are close to each other and both of them have large intercepts and negative slopes, whereas the MM-estimates have a smaller intercept and a positive slope, accurately capturing the trend of the data data set.

Table 1.3: Water flow measurements in Libby and Newgate on the kootenay river in January for the years 1931 - 1943

Year	Libby (x_i)	Newgate (y_i)
1931	27.1	19.7
1932	20.9	18.0
1933	33.4	26.1
1934	77.6	15.7
1935	37.0	26.1
1936	21.6	19.9
1937	17.6	15.7
1938	35.1	27.6
1939	32.6	24.9
1940	26.0	23.4
1941	27.6	23.1
1942	38.7	31.3
1943	27.8	23.8

Table 1.4: Estimated linear model parameters with standard errors (in brackets) for water flow data in Example 1.2

	Intercept	Slope
LSE	23.1649	-0.0138
	(3.3600)	(0.0944)
SLSE	25.6273	-0.1052
	(1.5597)	(0.0264)
MM-est.	5.4668	0.6209
	(1.7344)	(0.0649)

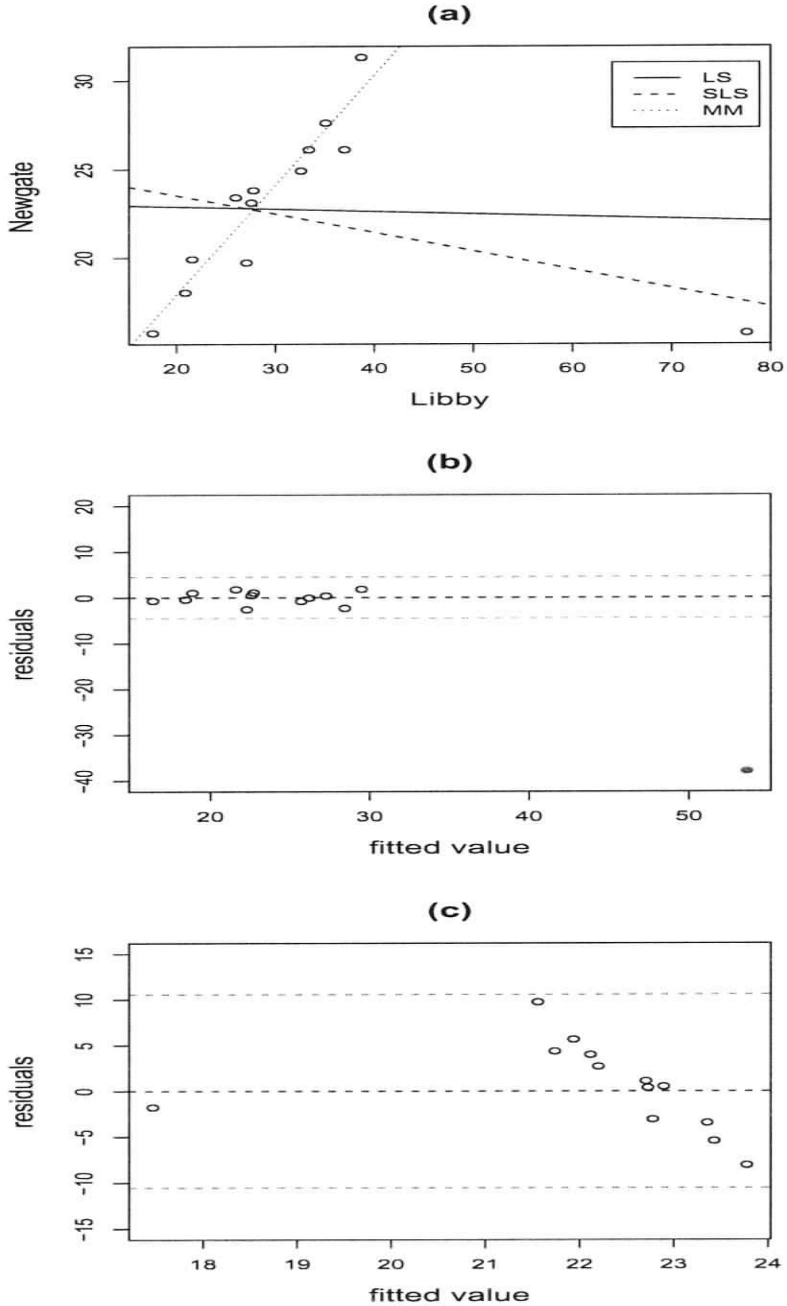


Figure 1.2: Analysis for water flow data in Example 1.2: (a) fitted lines for LS, SLS and MM-estimate, (b) residuals vs. fitted values for MM-estimate, (c) residuals vs. fitted values for the SLS estimator.

1.4 Overview of the thesis

We now provide an overview of the thesis.

In Chapter 2, we first discuss two important quantitative measures of robustness of an estimator: the influence function and the breakdown point. We then discuss robust estimation of the location and scale parameters of a univariate distribution. This is followed by a discussion on the more general M-estimator and the related MM-estimator for linear models. Numerical examples illustrating the use and effectiveness of some of these robust methods are also presented in this chapter.

In Chapter 3, we first explore the robustness of the SLS estimator through its empirical influence function and demonstrate that the SLS method is not robust. We also describe a few key formulae related to the SLS estimator defined in (1.4) and (1.5), including the formula for the optimal weight matrix W_{opt}. We then discuss robust extensions of the SLS estimator. The basic idea is to modify the weight matrix W in order to control the influence of potential outliers. We present two different versions of such modified weight matrices which give rise to two robust versions of the SLS method. One of these is designed to guard against X-outliers and another against both X- and Y-outliers. We present a computation algorithm for these robust SLS methods and demonstrate their robustness through their empirical influence functions. We also investigate theoretical properties of these robust SLS estimators such as their theoretical influence functions, their finite sample breakdown points and their asymptotic distributions.

In Chapter 4, we compare the robust SLS estimators with the LS estimator, SLS estimator and MM-estimator through an extensive simulation study. Specifically, we compare the biases and the standard errors of these estimators. We also compare the corresponding confidence intervals in terms of their coverage probabilities and average interval lengths. It turns out that the robust SLS estimators are highly efficient with high breakdown points. One of the robust SLS estimators is also asymptotically unbiased. We then apply these estimators to two well known real examples. The SLS estimates for these examples are consistent with estimates given by the robust MM-method. Further, for the purpose of outlier detection, the SLS methods are seen to have done a better job than the MM-method.

Finally, we provide some general discussions on the robust SLS methods as well as related future research problems in Chapter 5.

Chapter 2

Robust Methods

Qualitatively, an estimator is robust if outliers, however extreme, have only small impact on the estimates. Various quantitative means have been devised to measure the robustness of estimators. Two of the most important quantitative robustness measures are the influence function and breakdown point. The influence function is used to measure the impact of a small change in the underlying distribution on an estimator, whereas the breakdown point is used to measure the percentage of outliers in the data that an estimator can handle before it completely breaks down.

This chapter is organized as follows. In Section 2.1, we give the definitions of the influence function and the breakdown point of an estimator. In Section 2.2, we demonstrate the robustness of several robust estimators for location and scale using the tools of influence function and breakdown point. We also compare them to the popular but non-robust estimators for location and scale. Then in Section 2.3, we introduce the robust M-estimator and MM-estimator for linear regression models and examine their robustness through one real data example.

2.1 Measures of robustness

In this section, we will use θ to represent an unknown (scaler) parameter for a distribution of interest F. Let $x_1, \ldots, x_n$ be a random sample of size n from F. Consider an estimator for θ, $\hat{\theta}_n = \hat{\theta}_n(x_1, \ldots, x_n)$ such that $\hat{\theta}_n \to \hat{\theta}_\infty(F)$ in probability as $n \to \infty$. Here, $\hat{\theta}_\infty(F)$ is called the *asymptotic value* of the estimate at F. In practice, we can only assume that F is approximately known, say F is in some well defined set. We are interested in the behavior of $\hat{\theta}_\infty(F)$ when F belongs to $\mathcal{F}(F_0, \epsilon)$, a *contamination neighbourhood* centred around a distribution F_0 given below:

$$\mathcal{F}(F_0, \epsilon) = \{(1-\epsilon)F_0 + \epsilon G : G \in \mathcal{G}\}, \tag{2.1}$$

where $\mathcal{G}$ is a set of all distributions including point mass distributions. The "point mass" δ_{x_0} is the distribution such that $P(x = x_0) = 1$.

To quantify the robustness of estimators, the most commonly used measures are the breakdown point and influence function (Maronna, Martin and Yohai, 2006). Roughly speaking, an influence function measures the change of an estimator when the distribution underlying the data changes slightly. The ***influence function (IF)*** of an estimator (Hampel, 1974) is the change of $\hat{\theta}_\infty$ when the sample contains a small fraction ϵ of outliers. It is defined by

$$\begin{aligned} \mathrm{IF}_{\hat{\theta}}(x_0, F_0) &= \lim_{\epsilon \downarrow 0} \frac{\hat{\theta}_\infty((1-\epsilon)F_0 + \epsilon\delta_{x_0}) - \hat{\theta}_\infty(F_0)}{\epsilon} \\ &= \frac{\partial}{\partial \epsilon} \hat{\theta}_\infty((1-\epsilon)F_0 + \epsilon\delta_0)|_{\epsilon \downarrow 0} \end{aligned} \tag{2.2}$$

where "$\downarrow$" stands for "limit from the right-hand side". Here $\hat{\theta}_\infty((1-\epsilon)F_0 + \epsilon\delta_0)$ represents the asymptotic estimate when the underlying distribution of interest is F_0 but there is a small fraction ϵ of outliers at x_0. When the influence function of an estimator is bounded, we say that the estimator is robust as it indicates that a small change in the underlying distribution will only have a limited impact on the estimator. On the other hand, if the influence function of an estimator is unbounded, we say that the estimator is not robust as the impact of a small change in the underlying distribution can be arbitrarily large. In situations where θ is a p-dimensional vector, the asymptotic estimate $\hat{\theta}_\infty$ is also a p-dimensional vector. In such cases, there are p influence functions, one for each parameter in the parameter vector. We can define a p-dimensional influence function for the whole parameter vector (see Maronna, Martin and Yohai, 2006). Later in Section 3.4, we will consider one such example of multidimensional influence function.

In robustness studies, the distribution of interest F_0 is sometimes unknown. Hence the theoretical influence function (2.2) may not be available. For such cases, an *empirical influence function* (EIF) can be used instead to measure the sensitivity of an estimator to a "small change" in the sample. The empirical influence function at observation i is defined by

$$EIF_i(x) = \hat{\theta}_n(x_1, x_2, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n). \tag{2.3}$$

The rationale of empirical influence function is the following. The replacement of just one observation x_i by some arbitrary value x (an outlier) represents a "small change" to the sample in the sense of the percentage of observations being changed. Hence the empirical influence function measures the impact of such a small change on the estimator. The estimator is not robust if the empirical influence function is an unbounded function of x, meaning that a small change or one single outlier in the data set has unlimited impact on the estimator. On the other hand, if the empirical influence function is bounded, then the estimator is robust against one outlier.

Another widely used tool to measure robustness is the ***breakdown point (BP)***.

Roughly speaking, the breakdown point of an estimator is the largest proportion of outliers in the sample that an estimator can handle before "breaking down". Theoretically, the asymptotic breakdown point of an estimator $\hat{\theta}$ at distribution F_0, which is denoted by $\epsilon^*(\hat{\theta}, F_0)$, is the largest $\epsilon^* \in (0,1)$ such that $\hat{\theta}_\infty((1-\epsilon)F_0 + \epsilon G)$ remains bounded and bounded away from the boundary of Θ for $\epsilon < \epsilon^*$ and any distribution G (Maronna, Martin and Yohai, 2006). Here Θ is the parameter space of θ and the boundedness of $\hat{\theta}_\infty$ is taken to mean that it is not "breaking down". The higher the breakdown point of an estimator, the more robust the estimator. However, the breakdown point cannot exceed 0.5, which means for finite sample situations the maximum amount of outliers that a robust estimator can handle is 50% of the sample.

The asymptotic breakdown point $\epsilon^*(\hat{\theta}, F_0)$ is a conceptually simple and attractive measure for the robustness of an estimator, but it may be difficult to obtain due to technical difficulties involved in working with $\hat{\theta}_\infty((1-\epsilon)F_0 + \epsilon G)$. When this is the case, the *finite-sample breakdown point (FBP)* can be used instead to directly measure the maximum proportion of outliers in a sample that the estimator can handle before breaking down. The finite-sample breakdown point is defined as the largest proportion $\epsilon^*(\hat{\theta})$ of data points that can be replaced by arbitrary values with $\hat{\theta}$ remaining bounded and also bounded away from the boundary of Θ (Donoho and Huber, 1983). To make precise the definition of the finite-sample breakdown point, we now use $\mathbf{x}$ and $\mathbf{y}$ to represent sets of n observations, for example, $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$. We will also use $\partial\Theta$ to represent the boundary of Θ. For a fixed $\mathbf{x}$, denote by X_m the set of all $\mathbf{y}$ which have and only have $n - m$ elements in common with $\mathbf{x}$, that is,

$$X_m = \{\mathbf{y} : \#(\mathbf{y}) = n, \#(\mathbf{x} \cap \mathbf{y}) = n - m\}.$$

Then, the finite-sample breakdown point is

$$\epsilon_n^*(\hat{\theta}) = \frac{m^*}{n}, \tag{2.4}$$

where

$$m^* = max\left\{m \geq 0 : \hat{\theta}(y) \text{ bounded and also bounded away from } \partial\Theta \ \forall y \in X_m\right\}.$$

In most cases, when $n \to \infty$, ϵ_n^* tends to the asymptotic breakdown point ϵ^*.

2.2 Robust method for location and scale

We now discuss examples of important robust estimators. We begin with the simple case of robust estimation of location and scale, and then the more challenging case of robust estimation for linear models.

2.2.1 M-estimation for location

Consider a location model,

$$x_i = \mu + \varepsilon_i, \ i = 1, 2, \dots, n, \tag{2.5}$$

where μ is the unknown location parameter and ε_i are random errors with mean 0 and variance σ^2. If ε_i are i.i.d. with distribution function F_0, the observations $x_1, x_2, \dots, x_n$ are also i.i.d. and have distribution function

$$F(x) = F_0(x - \mu). \tag{2.6}$$

The ***M-estimator for location*** parameter μ is given by

$$\hat{\mu} = arg \ \min_{\mu} \sum_{i=1}^{n} \rho(x_i - \mu), \tag{2.7}$$

where the M stands for "maximum likelihood type" and function ρ satisfies the following conditions:

C1: $\rho(x)$ is a nondecreasing function of $|x|$,

C2: $\rho(0) = 0$.

C3: $\rho(x)$ is increasing for $x > 0$ such that $\rho(x) < \rho(\infty)$.

C4: If ρ is bounded, it is also assumed that $\rho(\infty) = 1$.

It is clear that when $\rho = - \ log \ f_0$, where f_0 is the density function corresponding to distribution function F_0, the M-estimator (2.7) is just the maximum likelihood estimator (MLE). Thus the M-estimator is a generalization of the MLE. Indeed, finding a robust generalization of the MLE was part of the motivation behind the M-estimator. See Huber (1977). Note that the MLE is in general not robust. Hence not every M-estimator is robust and a careful selection of the ρ function is necessary to ensure the robustness of the M-estimator. We will revisit this issue later in this section.

Suppose ρ is differentiable and $\psi = \rho' = \partial\rho(x - \mu)/\partial\mu$. Then the M-estimator (2.7) is seen as equal to a solution of the following estimating equation:

$$\sum_{i=1}^{n} \psi(x_i - \hat{\mu}) = 0. \tag{2.8}$$

The existence and uniqueness of solution to equation (2.8) depends on the ψ function involved. Maronna, Martin and Yohai (2006) gave the following summary regarding the existence and uniqueness:

1. If ψ is discontinuous, equation (2.8) may not have solutions.

2. If ψ is monotone nondecreasing and $\psi(-\infty) < 0 < \psi(\infty)$, equation (2.8) always has solutions.

3. If ψ is continuous and increasing, the solution for (2.8) is unique.

As we have noted before that the selection of the ρ and ψ functions are crucial for the robustness of the corresponding M-estimator. Among the many possible choices of such functions, an important choice is the family of *Huber functions* given below:

$$\rho_h(x) = \begin{cases} \frac{1}{2}x^2, & \text{if } |x| \leq h, \\ h\,|x| - \frac{h^2}{2}, & \text{if } |x| > h, \end{cases} \tag{2.9}$$

and

$$\psi_h(x) = \begin{cases} x, & \text{if } |x| \leq h, \\ sgn(x)h, & \text{if } |x| > h, \end{cases} \tag{2.10}$$

where h is a constant which determines the asymptotic variance of the M-estimator and hence the relative asymptotic efficiency (as measured by the ratio of the asymptotic variances) to the MLE. It follows that the value of h may be set according the desired level of relative efficiency. For example, suppose we wish to have 95% relative efficiency to the MLE for the case where the error distribution is normal and there are no outliers in the sample, then we need to have $h = 1.345$.

Another popular choice is *Tukey's bisquare function* (Beaton and Tukey, 1974) given below:

$$\rho_b(x) = \begin{cases} 1 - [1 - \left(\frac{x}{b}\right)^2]^3, & \text{if } |x| \leq b, \\ 1, & \text{if } |x| > b, \end{cases} \tag{2.11}$$

and

$$\psi_b(x) = \frac{6}{b}x[1 - \left(\frac{x}{b}\right)^2]^2\, I(|x| \leq b), \tag{2.12}$$

where the constant b is positive. Here we notice that the ψ function vanishes outside $[-b, b]$ and it is everywhere differentiable. When $b = 2.3849$, the M-estimator is 95% efficient for normal errors. Besides its application in M-estimation of location, the bisquare function is also used for MM-estimation for linear regression models. We will discuss this application in the next section.

Huber (1981) contains several other good choices of the ρ and ψ functions. M-estimators based on these functions are all robust with small bias and high efficiency relative to the classical estimators. Among these, the most popular choices for ψ function in practice are the bisquare ψ_b function and the redescending ψ function (Maronna, Martin and Yohai, 2006).

Finally, we present the influence function and breakdown point of the location M-estimate $\hat{\mu}$ from Maronna, Martin and Yohai (2006). Following (2.2), it can be shown that the influence function with respect to an underlying distribution F is

$$\mathrm{IF}_{\hat{\mu}}(x_0, F) = \frac{\psi(x_0 - \hat{\mu}_\infty)}{E_F \psi'(x - \hat{\mu}_\infty)}, \tag{2.13}$$

which is proportional to its ψ function.

The breakdown point for a general M-estimator is rather complicated. Here we only consider the special case where ψ is monotonic and

$$k_1 = -\psi(-\infty) \text{ and } k_2 = \psi(\infty).$$

One example of such a ψ function is Huber's ψ_h function in (2.10). M-estimator $\hat{\mu}$ given by such a ψ function has breakdown point

$$\epsilon^*(\hat{\mu}) = \frac{min(k_1, k_2)}{k_1 + k_2}. \tag{2.14}$$

If $k_1 = k_2$, then $\epsilon^*(\hat{\mu}) = 0.50$.

2.2.2 M-estimation for scale

M-estimator of scale parameter σ is similar to M-estimator of location. We assume x_i are i.i.d. with a distribution belonging to a scale family with density function

$$\frac{1}{\sigma} f_0\left(\frac{x}{\sigma}\right), \tag{2.15}$$

where $\sigma > 0$ is the unknown parameter. The ***M-estimator of scale*** $\hat{\sigma}$ is the solution to the following equation

$$\frac{1}{n}\sum_{i=1}^{n} \rho_{scale}\left(\frac{x_i}{\hat{\sigma}}\right) = \delta, \tag{2.16}$$

where δ is a positive constant and $\rho_{scale}(x) = x\psi(x)$. Maronna, Martin and Yohai (2006) noted the following important facts about the $\rho_{scale}(x)$ function:

1. If $0 < \delta < \rho_{scale}(\infty)$, then (2.16) has a solution.

2. If ρ_{scale} is bounded, then we can assume $\rho(\infty) = 1$, which means $\delta \in (0, 1)$.

For example, if we use Tukey's bisquare ρ_b function in (2.11) with $b = 1$, then

$$\rho_{scale} = \min\left\{1 - (1 - x^2)^3, 1\right\} \tag{2.17}$$

and $\delta = 0.5$. This ρ_{scale} function also satisfies conditions (C1)-(C4).

The influence function for the scale M-estimate $\hat{\sigma}$ is

$$\text{IF}_{\hat{\sigma}}(x_0, F) = \hat{\sigma}_\infty \frac{\rho(x_0/\hat{\sigma}_\infty) - \delta}{E_F\left[(x/\hat{\sigma}_\infty)\rho'(x/\hat{\sigma}_\infty)\right]}. \tag{2.18}$$

We can see that the influence function of $\hat{\sigma}$ is proportional to the ρ function.

As a related problem to the M-estimation of scale, note that the location M-estimator given by (2.7) is not scale equivariant. That is,

$$\hat{\mu}(cx_1, \ldots, cx_n) \neq c\hat{\mu}(x_1, \ldots, x_n)$$

for any fixed $c > 0$. This is not desirable but it can be fixed by modifying the definition in (2.7) with a scale parameter as follows:

1. If the scale parameter σ is known, (2.7) can be modified into

$$\hat{\mu} = arg\ \min_{\mu} \sum_{i=1}^{n} \rho\left(\frac{x_i - \mu}{\sigma}\right), \tag{2.19}$$

   and the resulting M-estimator for location is then scale equivariant.

2. If σ in (2.19) is unknown, we need to estimate it first with a robust estimator. For this purpose, normally we use the normalized median absolute deviation about the median (MADN), which is a robust alternative to the sample standard deviation (SD). The MADN of data $\mathbf{x} = (x_1, \ldots, x_n)$ is defined as

$$MADN(\mathbf{x}) = \frac{Med|\mathbf{x} - Med(\mathbf{x})|}{0.6745}, \tag{2.20}$$

   where "Med" is the sample median. The value of the normalizing constant (denominator) in (2.20) is set to 0.6745 so that $MADN(x)$ is an unbiased estimator for σ when the distribution for x_i is normal.

2.2.3 An example of M-estimation for location and scale

We now examine the robustness of various location and scale estimators through their empirical influence functions and the finite-sample breakdown points. The numerical example used here is based the well known passage times of light data from from Stigler (1977).

Example 2.1. *The following are transformed times t_i (in microseconds) needed for light to travel 7442 m shown in data set 9 in Stigler (1977). The actual times are* $t_i \times 0.001 + 24.8$.

28	26	33	24	34	-44	27	16	40	-2
29	22	24	21	25	30	23	29	31	19

Figure 2.1(a) shows the Q-Q plot of the data set. It is clear that the two smallest observations, $t_6 = -44$ and $t_{10} = -2$, deviate from the trend of the majority of the data points represented by the line. In particular, the smallest observation $t_6 = -44$ deviates substantially from that trend. Because of this, these two observations, especially the smallest one t_6, are viewed as outliers in various papers in the literature, and they will affect the performance of location and scale estimators.

We now compare the empirical influence functions and breakdown points of various location and scale estimators by using this example. Since t_6 is a recognized outlier, to compare the empirical influence functions we will use $EIF_6(t)$ in (2.3) obtained by keeping all other observations unchanged but replacing t_6 with an arbitrary value t. For this comparison, we allow the value of t to vary from -44 to 150. Figure 2.1(b) shows the $EIF_6(t)$ for the following location estimators: the mean, the median and the location M-estimator. Figure 2.1(c) shows the $EIF_6(t)$ for the following scale estimators: the SD, the MADN and the scale M-estimator. For both the location and scale M-estimators, we use Huber's ρ function. In Figure 2.1(b), the influence function of the sample mean is a straight line which is unbounded whereas that for the sample median and location M-estimator are almost horizontal and bounded. Similarly, in Figure 2.1(c), the influence function of the SD is unbounded, but that for the MADN and scale M-estimator are also almost horizontal and bounded. Recall that estimator

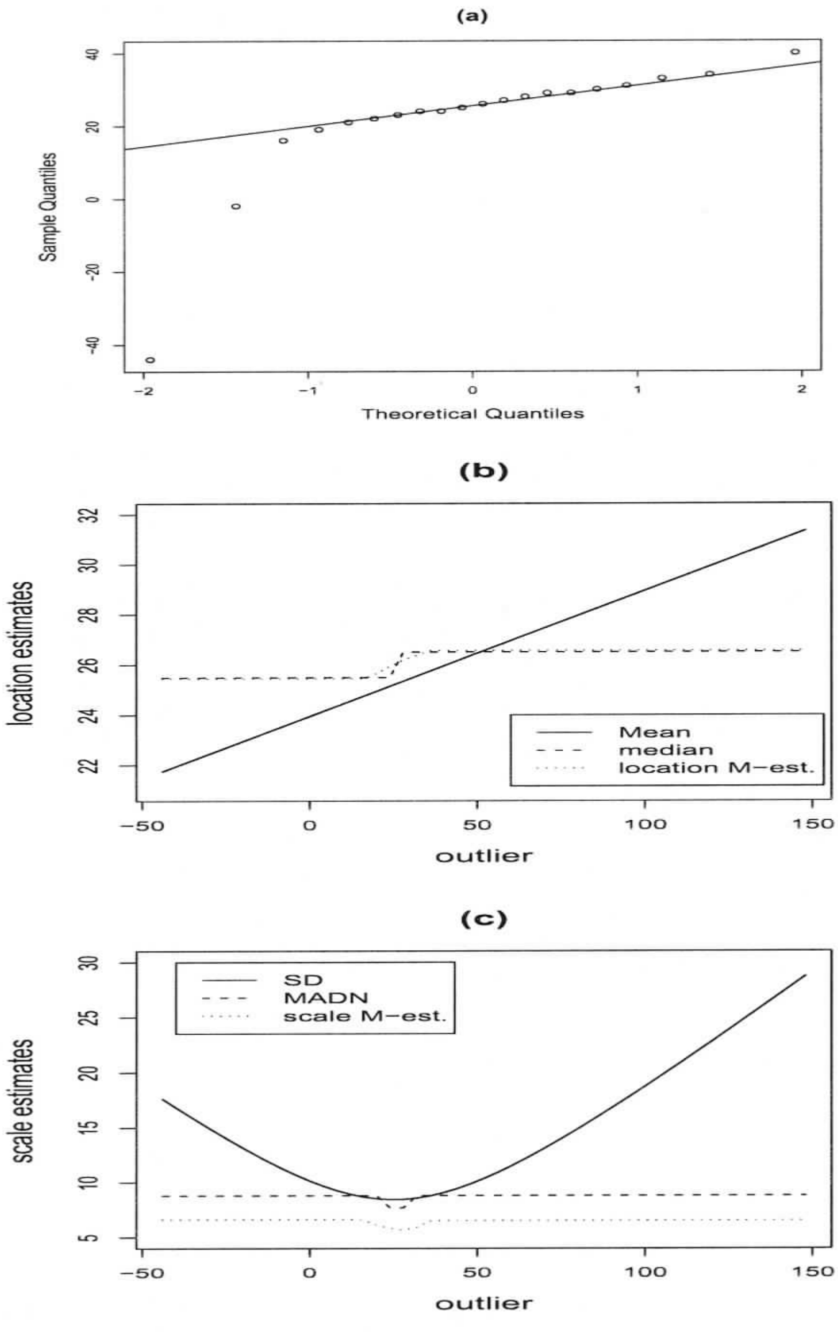


Figure 2.1: Analysis for light data in Example 2.1: (a) Q-Q plot of observed times, (b) the EIFs for the sample mean, sample median and location M-estimator for changing the value of the 6th observation, (c) the EIFs for SD, MADN and scale M-estimator for changing the value of the 6th observation.

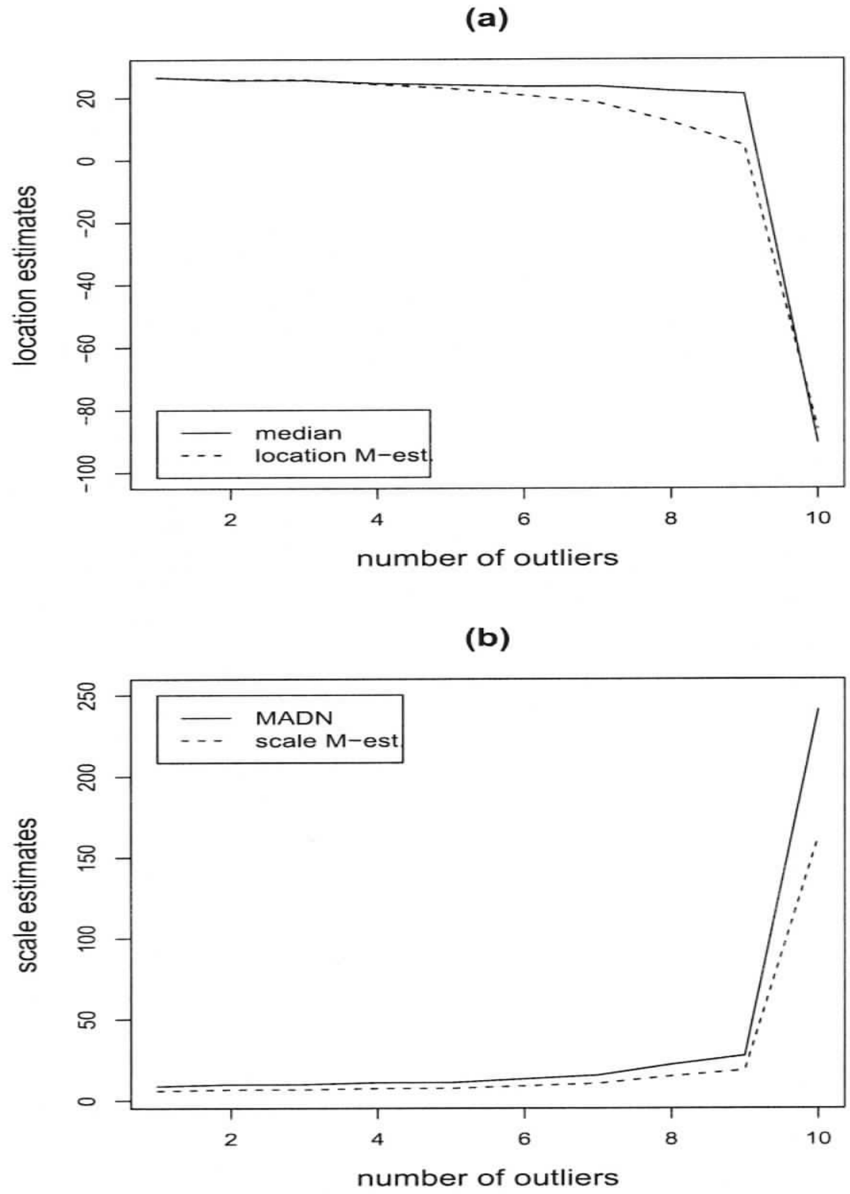


Figure 2.2: The FBPs for Example 2.1: (a) location estimators, (b) scale estimator.

with unbounded influence functions are not robust. Hence this example shows that the sample median, MADN, location M-estimator and scale M-estimator are robust, but the simple sample mean and standard deviation are not robust.

To discuss the robustness of the above estimators in terms of their finite-sample breakdown points, we select m ($m = 1, 2, \ldots, 10$) observations from the data set and replace each of them with -200, resulting in a data set containing m outliers. For each m value, we compute the estimates using the estimators and then plot the estimates against the m values to see at what value of m the estimators will break down. Figure 2.2 shows the plots of the estimates against m. The breakdown of each estimator occurs at the point of a steep drop. From Figure 2.2(a), there are big jumps in both median estimate and location M-estimate from $m = 9$ to $m = 10$. Thus the finite-sample breakdown points for these two estimators are both 9/20. Similarly Figure 2.2(b) shows both MADN and scale M-estimate have big jumps from $m = 9$ to $m = 10$ outliers, and their breakdown points are also 9/20. We did not show the plots for the sample mean and standard deviation as their breakdown points are both zero. This is clear from their empirical influence functions which are unbounded when there is only one outlier.

2.3 M-estimator and MM-estimator for linear regression models

Huber (1973) first considered M-estimation for regression. Consider linear model in (1.1). Similar to our discussion on M-estimators for location and scale in Section 2.2, we assume that the random error ε_i in the linear regression model has a density:

$$\varepsilon \sim \frac{1}{\sigma} f_0 \left(\frac{\varepsilon}{\sigma} \right),$$

where σ is a scale parameter. It follows that the conditional density of Y given X is

$$Y|X \sim \frac{1}{\sigma} f_0 \left(\frac{Y - X^T \boldsymbol{\theta}}{\sigma} \right).$$

Assume σ is a constant independent of X. Based on the above density function of Y, the likelihood function for $\boldsymbol{\theta}$ is

$$L(\boldsymbol{\theta}) = \frac{1}{\sigma^n} \prod_{i=1}^{n} f_0 \left(\frac{y_i - x_i^T \boldsymbol{\theta}}{\sigma} \right).$$

To find the MLE is to find the value $\hat{\boldsymbol{\theta}}$ which maximizes the likelihood function $L(\boldsymbol{\theta})$. This is also equivalent to finding the $\hat{\boldsymbol{\theta}}$ that minimizes the following function

$$\frac{1}{n} \sum_{i=1}^{n} \rho_0 \left(\frac{r_i(\boldsymbol{\theta})}{\sigma} \right) + log\sigma, \tag{2.21}$$

where $r_i(\boldsymbol{\theta}) = Y_i - X_i^T \boldsymbol{\theta}$ and $\rho_0 = -log f_0$. Assuming that the function in (2.21) is differentiable with respect to $\boldsymbol{\theta}$, we have the likelihood equation

$$\sum_{i=1}^{n} \psi_0 \left(\frac{r_i(\boldsymbol{\theta})}{\sigma} \right) X_i = 0, \tag{2.22}$$

where $\psi_0 = \rho_0' = -f_0'/f_0$. The MLE is then the solution to (2.22).

The MLE is efficient when the model assumptions are true but it is not robust. Having seen how it may be obtained through equations (2.21) and (2.22), we now generalize it to obtain a class of M-estimators for the regression parameters, which contain many robust special cases. To generalize through (2.21), we define the ***regression M-estimator*** $\hat{\boldsymbol{\theta}}$ as the minimizer for function

$$\sum_{i=1}^{n} \rho \left(\frac{r_i(\boldsymbol{\theta})}{\hat{\sigma}} \right), \tag{2.23}$$

where $\hat{\sigma}$ is a scale estimate and function ρ is no longer necessarily $-log f_0$ but any function satisfying conditions (C1)-(C4). Taking derivative of the function in (2.23) with respect to $\boldsymbol{\theta}$ yields the equation

$$\sum_{i=1}^{n} \psi\left(\frac{r_i(\boldsymbol{\theta})}{\hat{\sigma}}\right) X_i = 0, \tag{2.24}$$

where $\psi = \rho'$. Hence the M-estimator is also the solution to equation (2.24), which is a generalization of (2.22) for MLE.

As was previously mentioned that redescending ψ functions such as Tukey's bisquare function are good choices for constructing M-estimators. However, the resulting M-estimators for regression parameters may not be robust in all cases. Specifically, such estimators are robust against outliers in the response variable but not robust against outliers in the explanatory variables. Because of this, several alternative robust estimators have been proposed to overcome this problem. Among these, we mention the MM-estimator of Yohai (1987) for robust estimation of linear models. MM-estimator is highly robust against outliers in the response variable as well as outliers in the explanatory variables. It is also highly efficient. The MM-estimator can achieve simultaneous high breakdown point (50%) and high efficiency (95% asymptotic efficiency for normal errors) over a contamination neighbourhood.

To introduce MM-method, we now discuss S-estimator first. Let $\hat{\sigma} = \hat{\sigma}(\mathbf{r}(\boldsymbol{\theta}))$ be a scale estimate based on a vector of residuals

$$\mathbf{r}(\boldsymbol{\theta}) = (r_1(\boldsymbol{\theta}), \ldots, r_n(\boldsymbol{\theta})).$$

Then *S-estimate of scale* $\hat{\sigma}$ (Rousseeuw and Yohai, 1984) is defined as

$$\hat{\sigma} = \min_{\boldsymbol{\theta}} \hat{\sigma}(\mathbf{r}(\boldsymbol{\theta})), \tag{2.25}$$

where, for each $\mathbf{r}(\boldsymbol{\theta})$, $\hat{\sigma}(\mathbf{r}(\boldsymbol{\theta}))$ is the solution of

$$\frac{1}{n}\sum_{i=1}^{n} \rho_{scale}\left(\frac{r_i}{\hat{\sigma}}\right) = \delta. \tag{2.26}$$

Here, ρ_{scale} is bounded, continuous and an even function with $\rho_{scale}(0) = 0$ such as Tukey's bisquare ρ_b function. The corresponding *S-estimator* of $\boldsymbol{\theta}$ is then given by

$$\hat{\boldsymbol{\theta}} = arg\ \min_{\boldsymbol{\theta}} \hat{\sigma}(\mathbf{r}(\boldsymbol{\theta})). \tag{2.27}$$

The popular ***MM-estimator*** for parameter vector $\boldsymbol{\theta}$ of a linear model is defined as the point of convergence of the following iterative process:

1. Compute an initial consistent high breakdown estimate $\hat{\boldsymbol{\theta}}_0$ which may be of low efficiency.

2. Compute the residuals $r_i(\hat{\boldsymbol{\theta}}) = Y_i - X_i^T\hat{\boldsymbol{\theta}}$ where $\hat{\boldsymbol{\theta}}$ is the most current estimate of $\boldsymbol{\theta}$ for $i = 1, 2 \ldots, n$.

3. Compute an S-estimate of scale $\hat{\sigma}(\hat{\boldsymbol{\theta}})$ using residuals $r_i(\hat{\boldsymbol{\theta}})$.

4. Find a solution $\hat{\boldsymbol{\theta}}$ of (2.24). Here $\hat{\boldsymbol{\theta}}$ is the MM-estimate of $\boldsymbol{\theta}$.

There are software for computing MM-estimates available in the public domain. In R, for example, we may use "lmrob" from R library "Robust" to compute MM-estimates for linear regression models . In Example 1.2 that we discussed previously, the MM-method provides the best fit to the data set. In Figure 1.2(b), the MM-estimation flags the outlier with its largest residual. This is one of many examples where the MM-method outperforms the classical methods when there are outliers in the data. We now examine the influence functions of the M-estimator and the MM-estimator with an example.

Example 1.1 continued: To verify the robustness of the regression M-estimator and MM-estimator, the empirical influence functions are computed using the data set in Example 1.1. We vary the x value of the first observation continuously from 50 to 300 to create an x outlier, and compute the empirical influence functions, which are just the estimated parameter values corresponding to different values of x, for both the regression M-estimators (2.23) and the MM-estimators. The resulting influence functions are shown in Figure 2.3. We see from the plots in Figure 2.3 that the empirical influence functions for the M-estimators are not bounded, which indicate that the M-estimator is not robust against x-outlier. On the other hand, the MM-estimator is robust against x-outlier as its influence functions are bounded. To create a y-outlier, we change the y value of the first observation continuously from 30 to 200, and the resulting influence functions are plotted in Figure 2.4. The influence functions for the M-estimators and the MM-estimators are all horizontal lines and hence bounded, implying that both estimators are robust against a y-outlier.

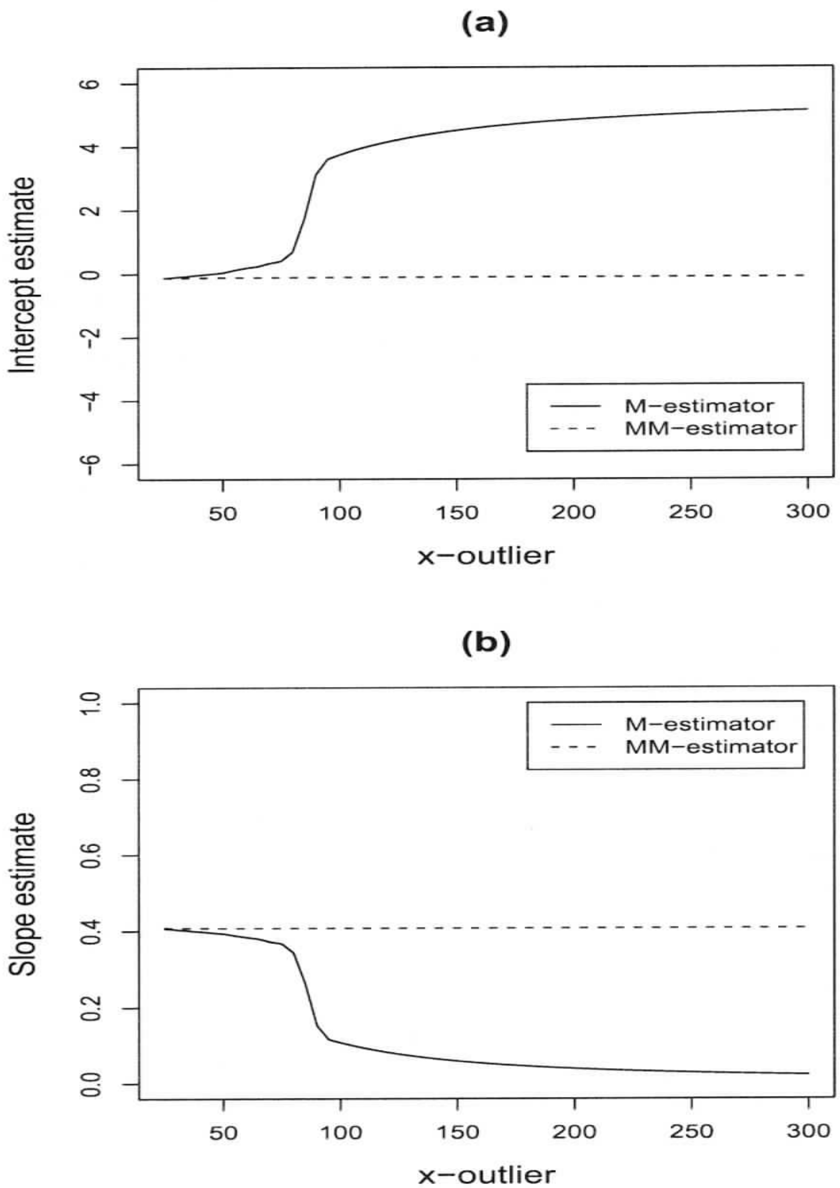


Figure 2.3: The empirical influence functions (EIFs) for Example 1.1 with an x-outlier: (a) EIFs for the intercept, (b) EIFs for the slope.

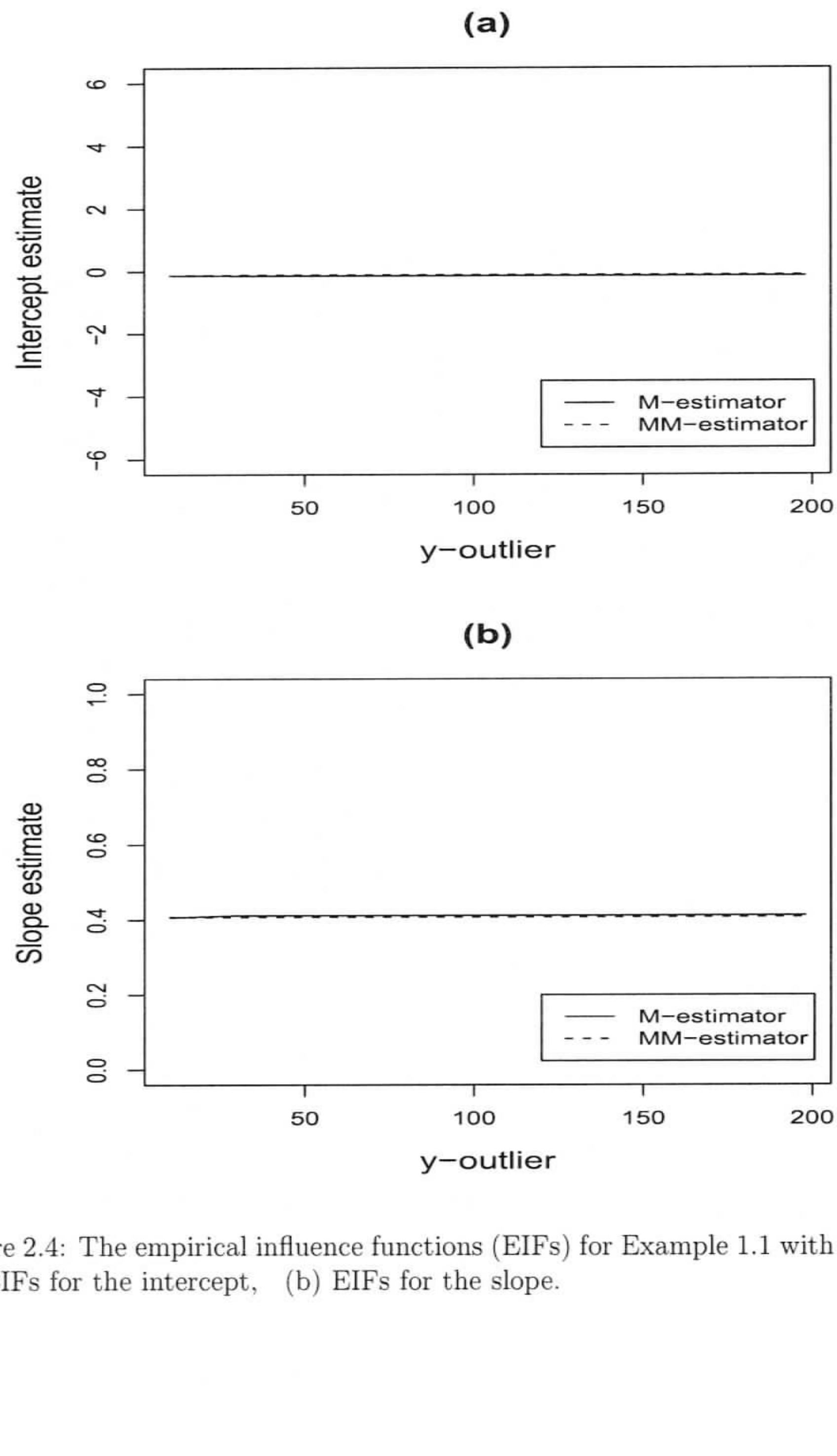


Figure 2.4: The empirical influence functions (EIFs) for Example 1.1 with a y-outlier: (a) EIFs for the intercept, (b) EIFs for the slope.

Chapter 3

Robust Second-order Least-squares Estimators

The second-order least-squares (SLS) estimator is asymptotically more efficient than the least-squares (LS) estimator when the error distribution has non-zero third moment. It is asymptotically as efficient as the LS estimator when the error distribution is symmetric (Wang and Leblanc, 2008). Nevertheless, Example 1.2 in Chapter 1 shows that just like the LS estimator, the SLS estimator is also not robust against outliers. In Section 3.1, we further examine the non-robustness of the SLS estimator using the tool of empirical influence function. The main objective of this chapter and hence this thesis, however, is to develop robust extensions of the SLS method. To this end, we propose two robust second-order least-squares (RSLS) estimators in Section 3.2 and examine their robustness using the empirical influence function. Both of these RSLS estimators are based on the observation that the SLS estimators can be made robust by proper selections of their weight matrices, and the two RSLS estimators that we propose are the results of such careful selections of weight matrices. In Section 3.3, we discuss the computation of the RSLS estimators and give one numerical example about the computation. In Section 3.4, we investigate the asymptotic properties of the RSLS estimators; we study their asymptotic influence functions, breakdown points as well as their asymptotic distributions.

3.1 Robustness of the second-order least-squares estimator

We have seen from Figure 1.2(a) for Example 1.2 that the SLS fitted line provides a poor fit to the data when there is even one outlier present. This suggests that the SLS estimator is not robust against outliers. We now use one of the robustness measures described in Chapter 2, the empirical influence function, to further explore the non-robustness of the SLS estimator. To do so, we use an example concerning rocket propellant from Montgomery, Peck and Vining (2006, p15).

Example 3.1. *A rocket motor is composed by bonding a sustainer propellant and an igniter propellant together. The shear strength of the bond is an important quality characteristic. It is suspected that the shear strength depends on the age (in weeks) of the batch of sustainer propellant. Table 3.1 contains twenty observations of shear strength and the age of the corresponding batch of sustainer propellant.*

In Figure 3.1, the scatter diagram exhibits a strong linear relationship between shear strength (y) and the age of the propellant (x), supporting the linear model

$$y_i = \theta_0 + \theta_1 x_i + \varepsilon_i.$$

Hence we may apply the SLS method to estimate the model parameters θ_0 and θ_1. To construct the empirical influence functions for the SLS estimators $\hat{\theta}_0$ and $\hat{\theta}_1$, the value of one observation is modified to create an outlier. The largest value of propellant

Table 3.1: The rocket propellant data for Example 3.1

Observation (i)	Shear Strength(psi) (y_i)	Age of Propellant(weeks) (x_i)
1	2158.70	15.50
2	1678.15	23.75
3	2316.00	8.00
4	2061.30	17.00
5	2207.50	5.50
6	1708.30	19.00
7	1784.70	24.00
8	2575.00	2.50
9	2357.90	7.50
10	2256.70	11.00
11	2165.20	13.00
12	2399.55	3.75
13	1779.80	25.00
14	2336.75	9.75
15	1765.30	22.00
16	2053.50	18.00
17	2414.40	6.00
18	2200.50	12.50
19	2654.20	2.00
20	1753.70	21.50

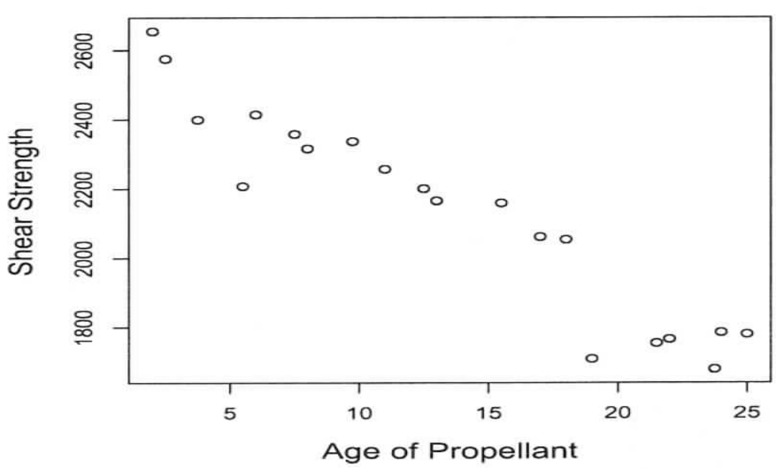


Figure 3.1: Scatter plot of shear strength verses propellant age for Example 3.1

age is $x_{13} = 25$ weeks and the smallest is $x_{19} = 2$. To create an x-outlier, we replace x_{19} with an outlier, x^*, whose value varies continuously from 56 to 300 while keeping the other x_i values unchanged. Figure 3.2 shows the empirical influence functions $\hat{\theta}_0(x^*)$ and $\hat{\theta}_1(x^*)$ as x^* increases from 56 to 300. For sufficiently large x^* values, the empirical influence function of the intercept $\hat{\theta}_0(x^*)$ decreases as x^* increases, while that for the slope $\hat{\theta}_1(x^*)$ increases. Both influence functions are unbounded if we are to allow x^* to go to infinity.

To create a y-outlier, we replace y_1 with a y-outlier y^* and allow its value to increase from 3500 to 10000. The corresponding empirical influence functions $\hat{\theta}_0(y^*)$ and $\hat{\theta}_1(y^*)$ are plotted in Figure 3.3. Figure 3.3(a) shows that the empirical influence function of the intercept estimate $\hat{\theta}_0(y^*)$ increases as y^* increases, and Figure 3.3(b) shows that the influence function of the slope estimate $\hat{\theta}_1(y^*)$ decreases as y^* increases. Again, both influence functions are unbounded if we are to let y^* to go to infinity. Since the empirical influence functions for the SLS estimator are unbounded in all of the above cases, the SLS estimator is not robust against either an X-outlier or a Y-outlier.

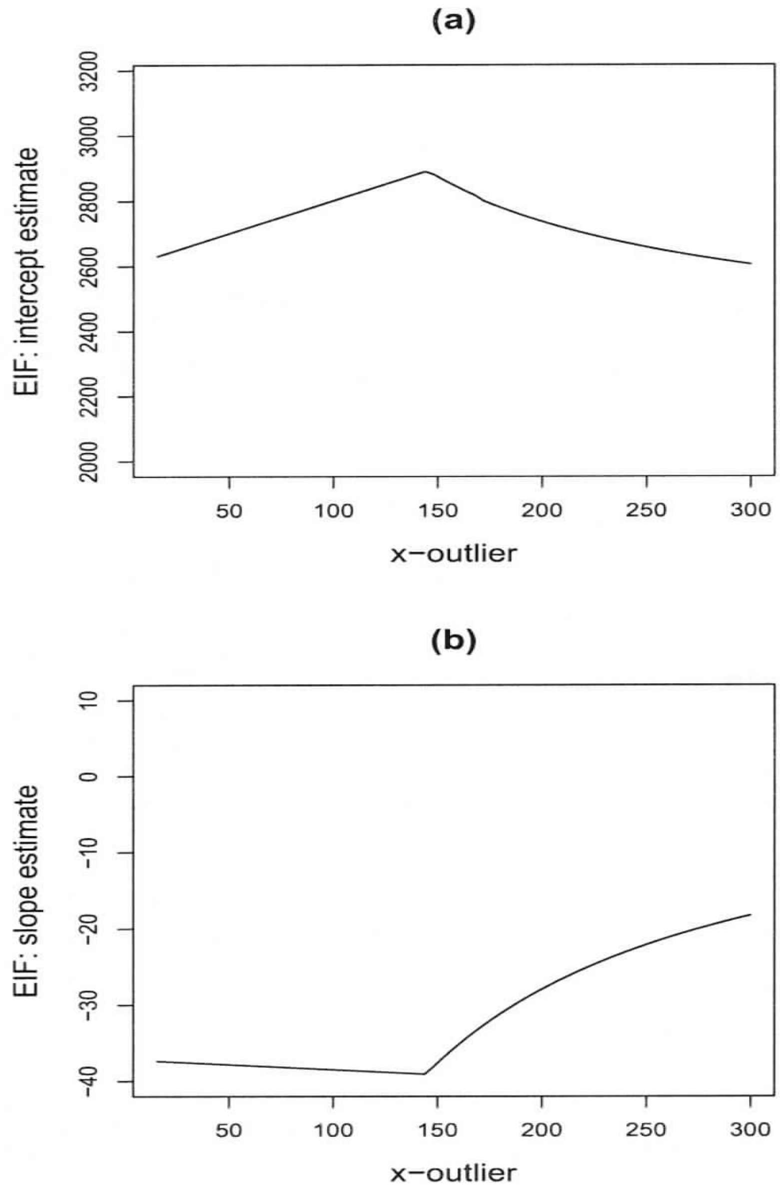


Figure 3.2: The empirical influence function (EIF) for the SLS estimator when there is an x-outlier in Example 3.1: (a) EIF for intercept estimate $\hat{\theta}_0$, (b) EIF for slope estimate $\hat{\theta}_1$.

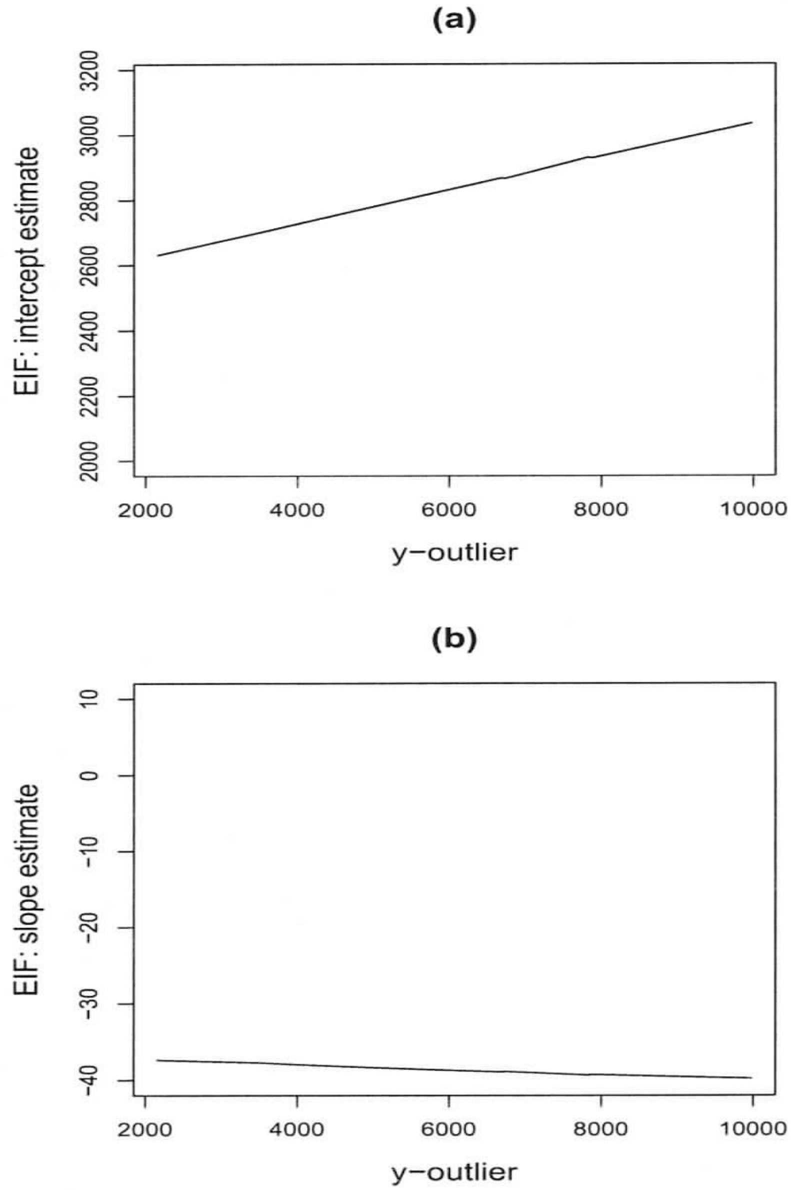


Figure 3.3: The empirical influence function (EIF) for the SLS estimator when there is a y-outlier in Example 3.1: (a) EIF for intercept estimate $\hat{\theta}_0$, (b) EIF for slope estimate $\hat{\theta}_1$.

3.2 Robust second-order least-squares estimators

Recall that the second-order least-squares (SLS) estimator for the parameter vector of a linear model γ defined in (1.3) is $\hat{\gamma}_{SLS}$, which minimizes function $Q_n(\gamma)$ in (1.4). The $\hat{\gamma}_{SLS}$ satisfies the following equation

$$\frac{\partial\, Q_n(\gamma)}{\partial\gamma} = 0, \tag{3.1}$$

which is equivalent to

$$\sum_{i=1}^{n} \frac{\partial}{\partial\gamma}\rho_i^T(\gamma) W\left(X_i\right)\rho_i(\gamma) = 0. \tag{3.2}$$

The two diagonal elements of the weight matrix $W(X_i)$ are roughly weights assigned to the distance of the response to its first conditional moment and the distance of the squared response to its second conditional moment.

Wang and Leblanc (2008) proved that under some regularity conditions, the SLS estimator is consistent and has an asymptotic normal distribution as $n \to \infty$. That is,

$$\sqrt{n}(\hat{\gamma}_{SLS} - \gamma_0) \xrightarrow{\mathcal{D}} N(0, A^{-1}BA^{-1}), \tag{3.3}$$

where $\gamma_0 = (\boldsymbol{\theta}_0^T, \sigma_0^2)^T$ is the true parameter value for the linear model (1.1),

$$A = E\left[\frac{\partial\rho^T(\gamma_0)}{\partial\gamma} W(X) \frac{\partial\rho(\gamma_0)}{\partial\gamma^T}\right], \tag{3.4}$$

$$B = E\left[\frac{\partial\rho^T(\gamma_0)}{\partial\gamma} W(X)\rho(\gamma_0)\rho^T(\gamma_0) W(X) \frac{\partial\rho(\gamma_0)}{\partial\gamma^T}\right], \tag{3.5}$$

and

$$\frac{\partial\rho^T(\gamma_0)}{\partial\gamma} = -\begin{bmatrix} \frac{\partial g(X;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} & 2g(X;\boldsymbol{\theta}_0)\frac{\partial g(X;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} \\ 0 & 1 \end{bmatrix}. \tag{3.6}$$

It is clear from (3.4) to (3.6) that the asymptotic covariance matrix $A^{-1}BA^{-1}$ in (3.3) depends on the weight matrix $W(X)$ and γ_0. Wang and Leblanc (2008) proved that the optimal choice for the weight matrix $W(X)$ is $W_{opt}(X) = U^{-1}$ where

$$U = U(X) = E\left[\rho(\gamma_0)\rho^T(\gamma_0)|X\right].$$

This choice of $W(X)$ is the optimal since the resulting asymptotic covariance matrix is the "smallest" in the sense that

$$A^{-1}BA^{-1} \geq \left(E\left[\frac{\partial\rho^T(\gamma_0)}{\partial\gamma}U^{-1}\frac{\partial\rho(\gamma_0)}{\partial\gamma^T}\right]\right)^{-1}, \tag{3.7}$$

where the "$\geq$" sign indicates that the difference matrix of the left-hand side of the above inequality minus the right-hand side is always non-negative definite. It is easy to verify that the determinant of U, $\det(U)$, is

$$\det(U) = \sigma_0^2(\mu_4 - \sigma_0^4) - \mu_3^2,$$

where $\mu_3 = E(\varepsilon^3|X)$ and $\mu_4 = E(\varepsilon^4|X)$. If $det\,(U) \neq 0$, then

$$\begin{aligned} W_{opt}(X) &= U^{-1} \\ &= \frac{1}{\det(U)}\begin{pmatrix} \mu_4 + 4\mu_3 g(X;\boldsymbol{\theta}_0) + 4\sigma_0^2 g^2(X;\boldsymbol{\theta}_0) - \sigma_0^4 & -\mu_3 - 2\sigma_0^2 g(X;\boldsymbol{\theta}_0) \\ -\mu_3 - 2\sigma_0^2 g(X;\boldsymbol{\theta}_0) & \sigma_0^2 \end{pmatrix}. \end{aligned}$$

Different SLS estimators may be constructed by using different weight matrices. In our subsequent discussions of SLS estimators, we assume they are the optimal SLS estimators given by the optimal weight matrix above, unless specified otherwise. Obviously, the exact optimal weight matrix is unavailable because it depends on the unknown parameters. In practice we can only construct approximate or asymptotically optimal estimators. Details about such optimal estimators and their computation are given in the next section.

The lack of robustness of the SLS estimator analogous to that of the LS estimator in that the weighting involved allows one outlier in the data set to dominate the objective function $Q_n(\gamma)$ in (1.4). The resulting SLS estimator, which is the minimizer of this objective function, will thus be heavily influenced by the outlier. Motivated by this observation, to robustify the SLS estimator we consider modifications of the weight matrix to weight down or even to eliminate the influence of outliers. To do so, let

$$W_R(X_i) = W_{opt}(X_i) \cdot V(X_i),$$

where $V(X_i)$ is a non-negative scalar function that we use to incorporate information about the outlyingness of X_i relative to the other X values. We will generalize function V later by allowing it to depend on Y_i as well. A small $V(X_i)$ weights down the contribution of the i^{th} observation to the objective function $Q_n(\gamma)$, and a $V(X_i)$ of value zero eliminates the contribution. Because of these, we will refer to the modified weight matrix $W_R(X_i)$ as the *robust weight matrix* and refer to the resulting SLS estimator for γ as the ***robust second-order least-squares (RSLS) estimator***. Denote by $\hat{\gamma}_R$ the RSLS estimator. Then $\hat{\gamma}_R$ is the measurable function that minimizes the following modified objective function

$$Q_n^*(\gamma) = \sum_{i=1}^{n} \rho_i^T(\gamma) W_R(X_i) \rho_i(\gamma). \tag{3.8}$$

Equivalently, the RSLS estimator $\hat{\gamma}_R$ is the solution to

$$\sum_{i=1}^{n} \frac{\partial}{\partial \gamma} \rho_i^T(\gamma) W_R(X_i) \rho_i(\gamma) = 0. \tag{3.9}$$

There are different choices of the V function, leading to different robust weight matrices $W_R(X_i)$ and hence different RSLS methods. We have experimented with several choices of V functions. It turned out that two simple choices have worked very well. In our subsequent discussions, we will be concerned with only these two choices and the corresponding RSLS methods. We now describe these two V functions.

1. Version one of robust second-order least-squares estimator (RSLSE-I)

Consider the robust weight matrix $W_{R1}(X_i) = W_{opt}(X_i) \cdot V_1(X_i)$ given

$$V_1(X_i) = \begin{cases} 1, & \text{if } X_i \text{ is not an outlier,} \\ 0, & \text{if } X_i \text{ is an outlier.} \end{cases} \tag{3.10}$$

Clearly, such a weight matrix will eliminate the influence of observations whose X-values are outliers. To assign values to $V_1(X_i)$, we need to decide whether X_i is an outlier. There are different methods in the literature for detecting X-outliers. The one that we found particularly suitable for our purpose is based on the *minimum volume ellipsoid estimator* (MVE) for location introduced by Rousseeuw (1984). We now describe this approach for constructing the $V_1(X_i)$ function.

The MVE gives two estimates:

$$\begin{aligned} T(X) &= \text{center of the minimal volume ellipsoid} \\ & \text{covering (at least) } h \text{ points of } X_1, \ldots, X_n, \\ C(X) &= \text{the covariance matrix for } X_1, \ldots, X_n, \end{aligned}$$

where h can equal to $[n/2] + 1$, and function $[t]$ is the integer part of t. Rousseeuw proved that for any q-dimensional sample, the breakdown point of the MVE is

$$\epsilon_n^*(MVE) = ([n/2] - q + 1)/n, \tag{3.11}$$

which converges to 0.50 as $n \to \infty$. The MVE has high breakdown point. Once $T(X)$ and $C(X)$ are computed, we can construct function $V_1(X_i)$ in (3.10) as follows:

$$V_1(X_i) = \begin{cases} 1, & \text{if } (X_i - T(X))^T C(X)^{-1}(X_i - T(X)) \leq c, \\ 0, & \text{otherwise,} \end{cases} \tag{3.12}$$

where the cut-off value c can be $\chi^2_{q,0.975}$. An simple generalization of $V_1(X_i)$ in (3.12) is to allow it to depend continuously on the outlyingness of X_i as measure by

$$(X_i - T(X))^T C(X)^{-1}(X_i - T(X)).$$

But this makes the resulting RSLS estimator more complicated both theoretically and computationally. Hence we have left this possibility as a future research problem and will not further pursue it in this thesis.

2. Version two of robust second-order least-squares estimator (RSLSE-II)

The above robust second-order least-squares estimator RSLSE-I takes into consideration only the outlyingness of X_i. In practice, however, there may be Y-outliers or simply (X, Y)-outliers which deviate substantially from the underlying linear model. To handle more general outliers in the data, we construct a new RSLS estimator with robust weight matrix $W_{R2}(X_i) = W_{opt}(X_i) \cdot V_2(X_i)$ where

$$V_2(X_i) = V_1(X_i) \cdot V_{ei}, \tag{3.13}$$

and V_{ei} is the following indicator function of the residual $\varepsilon_i = Y_i - g(X_i, \hat{\boldsymbol{\theta}}_0)$ calculated using, say the MM-method:

$$V_{ei} = \begin{cases} 1, & \text{if } \varepsilon_i \text{ is not an outlier,} \\ 0, & \text{if } \varepsilon_i \text{ is an outlier.} \end{cases} \tag{3.14}$$

In our subsequent discussions, we will assign a 1 to V_{ei} if $|\varepsilon_i| < 2.5\hat{\sigma}_0$ where $\hat{\sigma}_0$ is MM-estimate of the error standard deviation.

It is clear by the construction of $V_2(X_i)$ in (3.13) that it has two layers of robustness built-in; (1) through the component $V_1(X_i)$ it is able to handle the X-outliers and (2) through the component V_{ei} it is able to handle outliers in a more general sense. Indeed, all outliers (regardless of the nature of their outlyingness) identified by the MM-method will be given a weight of zero. The two-layer robustness structure is necessary as the first and the second layers are complimentary to each other in that in general $V_2(X_i)$ cannot be simplified to either $V_1(X_i)$ or V_{ei}.

Example 3.1 continued: To check the robustness of the RSLSE-I and RSLSE-II, we compute their empirical influence functions using the same data set for Example 3.1. Figure 3.4(a) contains the two empirical influence functions of the two RSLS estimators for the intercept θ_0 in the case of an X-outlier. There we see that both influence functions are bounded, implying that the RSLS estimators for the intercept are robust against an X-outlier. Further, RSLS estimators for the slope θ_1 are also robust against an X-outlier as their empirical influence functions are also bounded. See Figure 3.4(b).

To examine their performance against a Y-outlier, Figure 3.5(a) shows the empirical influence functions of the two RSLS estimators for the intercept θ_0 when there is one Y-outlier, y^*. We see that the empirical influence function for RSLSE-I increases as y^* increases. But for RSLSE-II, the empirical influence function is a horizontal line and hence bounded. In Figure 3.5(b), the empirical influence function of RSLSE-I for the slope θ_1 decreases as y^* increases, but that of RSLSE-II is bounded. Thus only RSLSE-II is robust against a Y-outlier. This is not surprising as RSLSE-I only contains robust weights for explanatory variables X_i and hence is only robust against X-outliers. The RSLSE-II, on the other hand, contains robust weights for both explanatory variables and the errors, which makes it also robust against other types of outliers, including the Y-outliers.

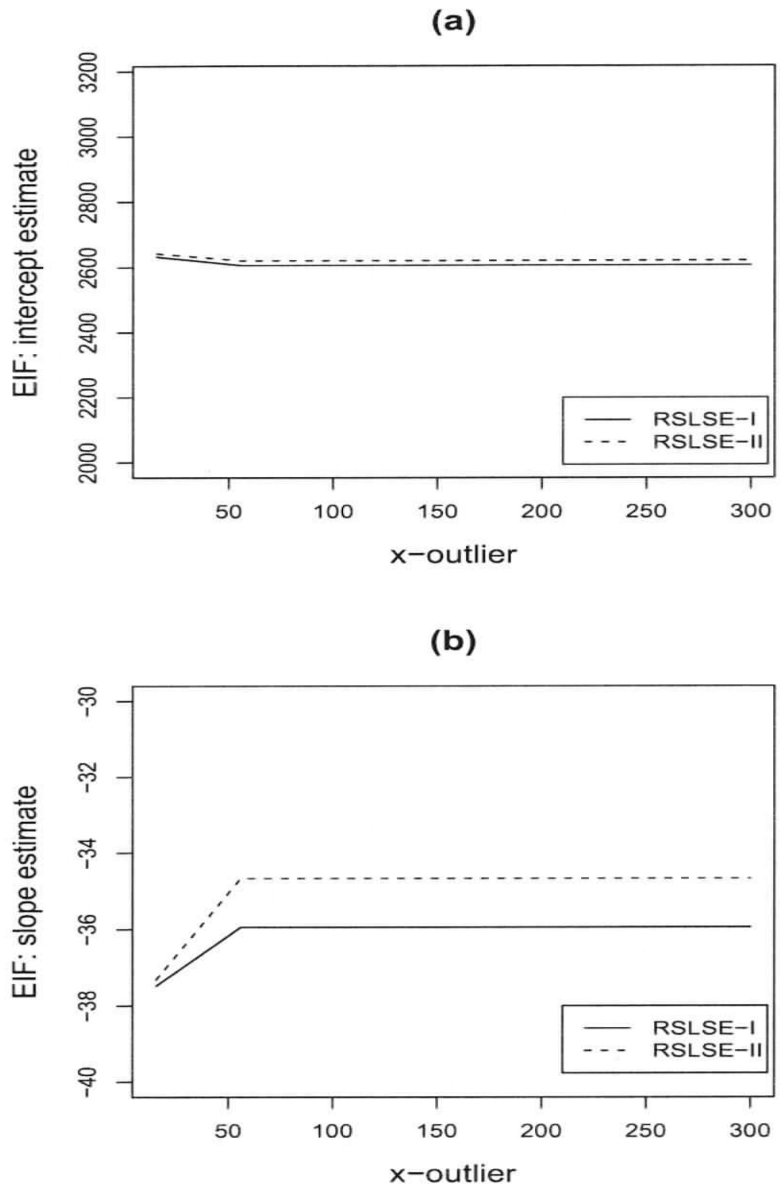


Figure 3.4: The empirical influence functions (EIFs) for the RSLSEs for one X-outlier: (a) EIFs for intercept estimate $\hat{\theta}_0$, (b) EIFs for slope estimate $\hat{\theta}_1$.

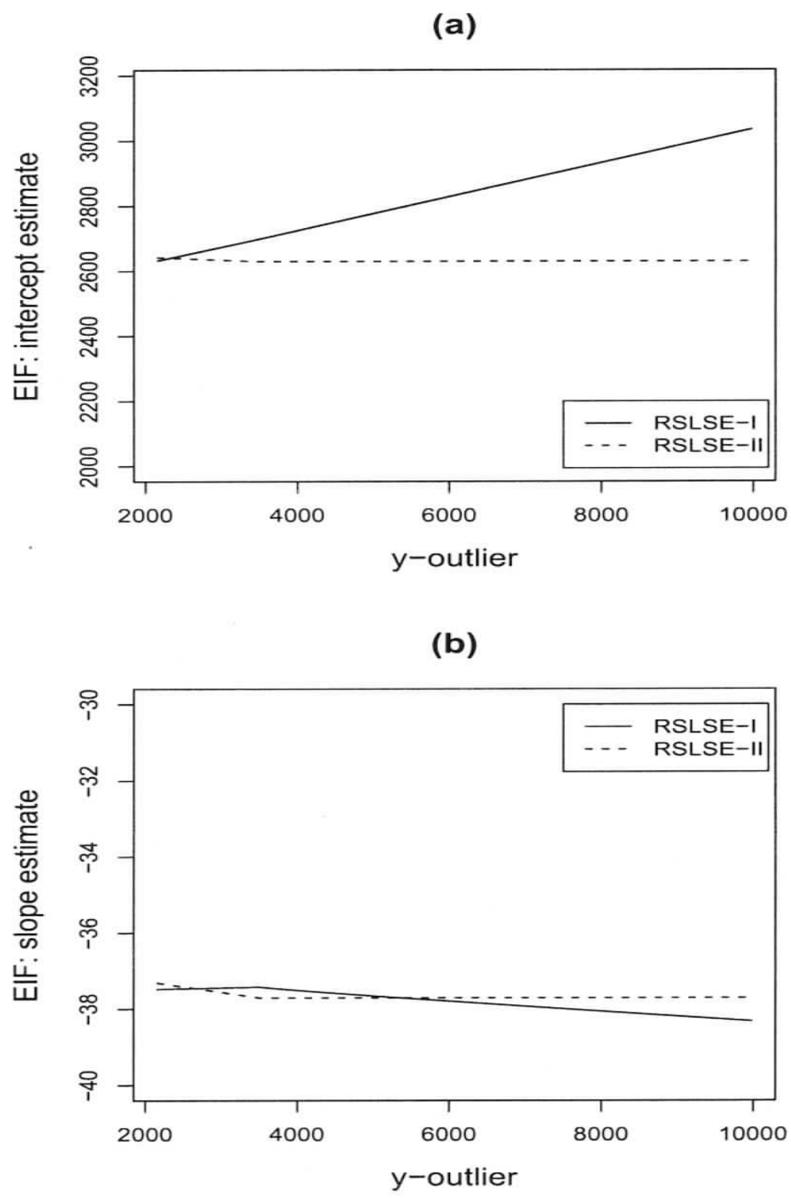


Figure 3.5: The empirical influence functions (EIFs) for the RSLSEs for one Y-outlier: (a) EIFs for intercept estimate $\hat{\theta}_0$, (b) EIFs for slope estimate $\hat{\theta}_1$.

3.3 Computation and example

We first discuss the computation of the (non-robust) SLS estimate. Wang and Leblanc (2008) suggested the following two-stage procedure:

Stage 1. Using the identity weight $W = I_{2\times 2}$ to minimize $Q_n(\gamma)$ and obtain the first-stage estimate of γ, $\hat{\gamma}^{(1)} = (\hat{\boldsymbol{\theta}}^{(1)}, \hat{\sigma}^{2(1)})$.

Stage 2. This stage consists of two steps:

(a) Evaluate the elements of $W_{opt}(X)$ with estimated values for μ_3 and μ_4 from the residuals, i.e., $\hat{\mu}_3 = \sum_{i=1}^n r_i^3/n$, $\hat{\mu}_4 = \sum_{i=1}^n r_{i=1}^4/n$ where $r_i = Y_i - g(X_i, \hat{\boldsymbol{\theta}}_0^{(1)})$. Denote by $\hat{W}_{opt}(X)$ the estimated weight matrix.

(b) Compute the second-stage estimate $\hat{\gamma}^{(2)}$ by minimizing $Q_n(\gamma)$ defined with the estimated weight matrix $\hat{W}_{opt}(X)$.

The second-stage estimate $\hat{\gamma}^{(2)}$ is then the optimal SLS estimate of γ.

We now present an iterative algorithm for computing the robust SLS (RSLS) estimates:

Step 1. Compute the MVE for the center and the covariance matrix of $X_1, \ldots, X_n$, and evaluate $V_1(X_i)$ for $i = 1, \ldots, n$.

Step 2. Find a robust estimate $\hat{\gamma}^{(0)} = (\hat{\boldsymbol{\theta}}^{(0)}, \hat{\sigma}^{2(0)})$ for γ as an initial estimate. For linear regression, we use the MM-estimate.

Step 3. Denote $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}^{(0)}$, $\sigma_0^2 = \hat{\sigma}^{2(0)}$ and $r_i = Y_i - g(X_i, \boldsymbol{\theta}_0)$. Evaluate V_{ei} in (3.14) and compute $V_2(X_i) = V_1(X_i) \cdot V_{ei}$.

Step 4. Estimate the weight matrix $\hat{W}_{R1}(X_i) = \hat{W}_{opt}(X_i) \cdot V_1(X_i)$ or $\hat{W}_{R2}(X_i) = \hat{W}_{opt}(X_i) \cdot V_2(X_i)$ (depending on the estimator used), where

$\hat{W}_{opt}(X_i) = \frac{1}{\sum_{j=1}^{n} V(X_j)} \cdot$

$$\begin{pmatrix} \sum_{j=1}^{n} V(X_j)(r_j^2 + 2r_j g(X_i, \boldsymbol{\theta}_0) - \sigma_0^2)^2 & -\sum_{j=1}^{n} V(X_j) r_j (r_j^2 + 2r_j g(X_i, \boldsymbol{\theta}_0) - \sigma_0^2) \\ -\sum_{j=1}^{n} V(X_j) r_j (r_j^2 + 2r_j g(X_i, \boldsymbol{\theta}_0) - \sigma_0^2) & \sum_{j=1}^{n} V(X_j) r_j^2 \end{pmatrix}$$

with $V(X_j) = V_1(X_j)$ for $\hat{W}_{R1}$ and $V(X_j) = V_2(X_j)$ for $\hat{W}_{R2}$.

Step 5. Minimize function $Q_n^*(\gamma)$ with $\hat{W}_{R1}(X_i)$ or $\hat{W}_{R2}(X_i)$ to get $\hat{\gamma}^{(1)}$. Set $\hat{\gamma}^{(0)} = \hat{\gamma}^{(1)}$, and go back to Step 3 until the convergence of $\hat{\gamma}^{(1)}$ is reached.

The number of iterations required to reach convergence depends on the data set involved and the initial value $\hat{\gamma}^{(0)}$. In practice, when we use the MM-estimate as the initial value $\hat{\gamma}^{(0)}$ for fitting linear model with RSLS methods, the number of iterations required are usually quite small; often one or two iterations are sufficient. To simplify the computation, in our simulation studies in Chapter 4, we will use just one iteration for the computation of all RSLS estimates.

Note that in our subsequent discussions, the RSLS estimate refers to the final result given by the five-step algorithm. Consequently, the associated weight matrix W_R in estimating equation (3.9) is the weight matrix in the final step of the iteration. It is clear from the algorithm that a RSLS estimator derives its robustness from that of the methods involved in the various steps of the algorithm. The RSLSE-I, for example, builds its robustness on that of the MVE method and MM-method. Hence its robustness properties are tied to that of the methods involved. We will further illustrate this point in Section 3.4 when we derive the breakdown point for RSLSE-I. We now give one numerical example on the use of the above algorithms involving the Pilot-Plant data from Daniel and Wood (1971).

Example 3.2. *The response variable corresponding to the acid content is determined by titration (y_i), and the explanatory variable organic acid content is determined by* *extraction (x_i) and weighing. They are listed in Table 3.2. This data set is also analysed by Rousseeuw and Leroy (1987), and Yale and Forsythe (1976) with the simple linear regression model.*

Table 3.2: Pilot-plant data for Example 3.2

Observation (i)	Extraction (x_i)	Titration (y_i)
1	123	76
2	109	70
3	62	55
4	104	71
5	57	55
6	37	48
7	44	50
8	100	66
9	16	41
10	28	43
11	138	82
12	105	68
13	159	88
14	75	58
15	88	64
16	164	88
17	169	89
18	167	88
19	149	84
20	167	88

The scatterplot in Figure 3.6(a) shows that there is no obvious outlier in this data, and the non-robust estimates from the LS estimator and SLS estimator are almost the same as robust estimates from the RSLSE-I, RSLSE-II and MM-estimator. Also, see Table 3.3 for the estimated parameter values. Now suppose the x-value of the 6^{th} observation has been wrongly recorded as 370 instead of (the correct value of) 37. This error produces an outlier, which is indicated by a solid circle point in Figure 3.6(b). The robust estimators, the RSLSE-I, RSLSE-II and MM-estimator, yield good fits to the majority of the data and these fits are quite close to those to the

Table 3.3: Estimated linear model parameters with standard errors (in brackets) for pilot-plant data without outlier in Example 3.2

	Intercept (θ_0)	Slope (θ_1)
LSE	35.4583	0.3216
	(0.6350)	(0.0056)
SLSE	35.5800	0.3205
	(0.5592)	(0.0047)
RSLSE-I	35.5321	0.3209
	(0.5370)	(0.0046)
RSLSE-II	35.5321	0.3209
	(0.5370)	(0.0046)
MM-est.	35.4972	0.3210
	(0.5254)	(0.0050)

Table 3.4: Estimated linear model parameters with standard errors (in brackets) for pilot-plant data with an outlier in Example 3.2

	Intercept (θ_0)	Slope (θ_1)
LSE	58.9388	0.0807
	(6.6142)	(0.0469)
SLSE	54.4945	0.1425
	(3.3831)	(0.0137)
RSLSE-I	35.3174	0.3226
	(0.5503)	(0.0048)
RSLSE-II	35.3174	0.3226
	(0.5503)	(0.0048)
MM-est.	35.3359	0.3222
	(0.5761)	(0.0051)

original data (with the correct value of 37). On the other hand, the LS and SLS estimators are pulled downward by the outlier, and they do not provide good fits to the majority of the data. Moreover, the parameter estimates of the LS method and SLS method are quite different from those of the robust estimators in Table 3.4. The RSLSE-I and RSLSE-II perform very well in this example, and they are similar to the MM-estimator.

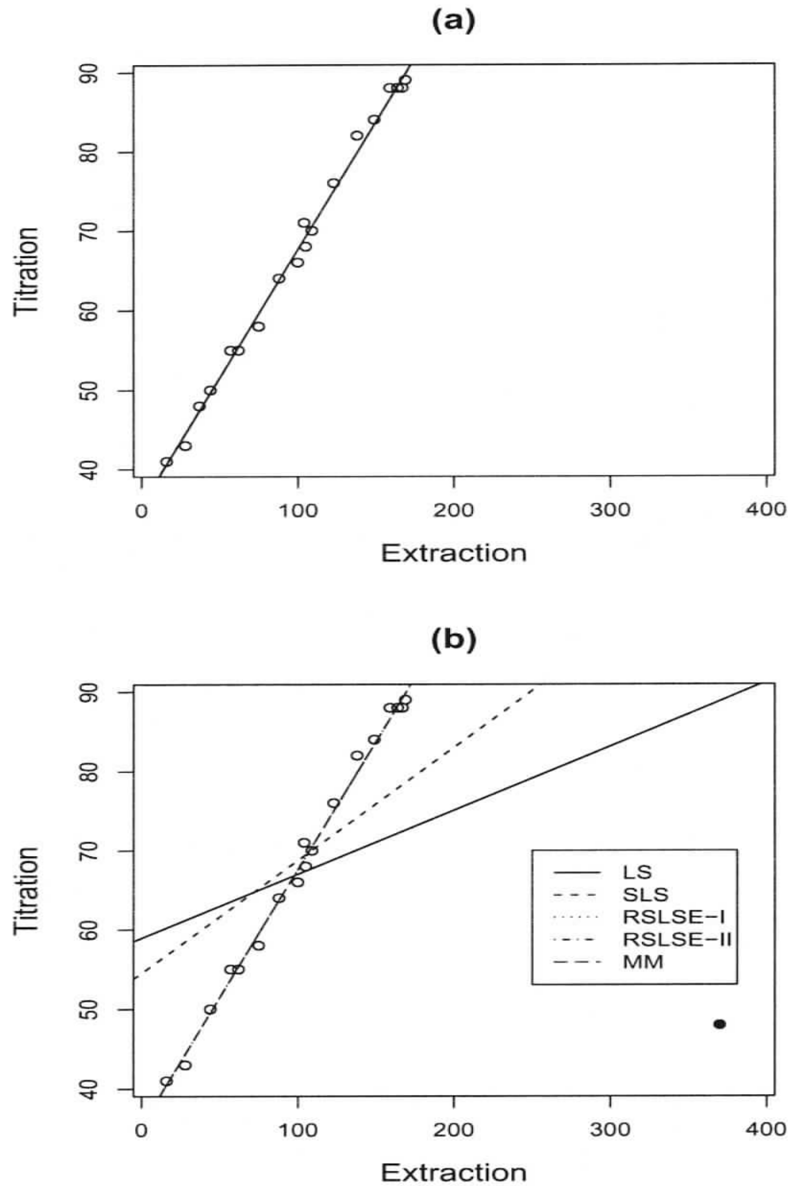


Figure 3.6: The LS, SLS, RSLSE-I, RSLSE-II and MM fitted lines for pilot-plant data in Example 3.2: (a) fitted lines for data without outlier, (b) fitted lines for data with one X-outlier. The solid line is the LS fit, dotted line is the SLS fit, dashed lines are the RSLSE-I, RSLSE-II and MM-estimator fits.

3.4 Properties of robust second-order least-squares estimators

In this section, we will derive the asymptotic influence functions, breakdown points and asymptotic distributions for the RSLS estimators.

3.4.1 Influence function

In Section 2.1, we gave the definition for the influence functions of an estimator for a scaler parameter in (2.2). Most M-estimators do not have simple analytic expressions and can only be expressed as a solution of estimating equations. It may seem from (2.2) that their influence functions are difficult to evaluate due to the lack of analytic expressions. But it turns out that their influence functions can be conveniently expressed in terms of the estimating functions involved, Ψ. We now give such expressions as shown in Maronna, Martin and Yohai (2006).

Consider a multidimensional parameter vector $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_{p-1})^T$. Let $\hat{\boldsymbol{\theta}}_\infty = (\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_{p-1})^T$ be the asymptotic value of an estimator $\hat{\boldsymbol{\theta}}$ satisfying

$$E_F \boldsymbol{\Psi}\left(Z, \hat{\boldsymbol{\theta}}\right) = \mathbf{0},$$

where F is the distribution of $Z = (X, Y)$, $\boldsymbol{\Psi} = (\Psi_0, \Psi_1, \ldots, \Psi_{p-1})$ and Ψ_j could be any ψ function previously described in Chapter 2. Then the influence function of $\hat{\boldsymbol{\theta}}$ at a fixed point $Z_0 = (X_0, Y_0)$ is given by (Maronna, Martin and Yohai, 2006)

$$IF_{\hat{\boldsymbol{\theta}}}(Z_0, F) = -M^{-1}\boldsymbol{\Psi}(Z_0, \hat{\boldsymbol{\theta}}_\infty), \tag{3.15}$$

where the matrix M has elements

$$M_{jk} = E\left\{ \frac{\partial \Psi_j(Z_0, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} \bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_\infty} \right\}, \; j, k = 1, 2, \ldots, p.$$

Since the RSLS estimator $\hat{\gamma}_R$ for γ is the solution to equation (3.9), it is an M-estimator with asymptotic value $\hat{\gamma}_\infty(F)$ satisfying

$$E_F\left[\frac{\partial \rho^T(\hat{\gamma}_\infty(F))}{\partial \gamma} W(X)\rho(\hat{\gamma}_\infty(F))\right] = \mathbf{0}. \tag{3.16}$$

It follows that the influence function of the RSLS estimator $\hat{\gamma}_R$ is given by (3.15) where the $\boldsymbol{\Psi}$ function is (implicitly) specified in (3.16). We now demonstrate the validity of (3.15) by proving that the influence function of RSLS estimator $\hat{\gamma}_R$ is indeed (3.15).

Theorem 1. *The influence function for the RSLS estimator $\hat{\gamma}_R$ at point $Z_0 = (X_0, Y_0)$ under the distribution F is*

$$IF_{\hat{\gamma}_R}(Z_0, F) = -M^{-1}(\hat{\gamma}_\infty(F))\boldsymbol{\Psi}\left(Z_0, \hat{\gamma}_\infty(F)\right), \tag{3.17}$$

where

$$\boldsymbol{\Psi}(Z, \gamma) = \tfrac{\partial \rho^T(\gamma)}{\partial \gamma} W_R(X)\rho(\gamma),$$
$$M(\hat{\gamma}_\infty(F)) = E_F\left[\tfrac{\partial \boldsymbol{\Psi}(Z, \hat{\gamma}_\infty(F))}{\partial \gamma}\right] = E_F\left[\tfrac{\partial \rho^T(\hat{\gamma}_\infty(F))}{\partial \gamma} W_R(X) \tfrac{\partial \rho(\hat{\gamma}(F))}{\partial \gamma^T}\right].$$

Proof: Let $F_\epsilon = (1-\epsilon)F + \epsilon\delta_{Z_0}$ and $\gamma_\epsilon = \hat{\gamma}_\infty(F_\epsilon)$. Then γ_ϵ satisfies

$$E_{F_\epsilon}\boldsymbol{\Psi}(Z, \gamma_\epsilon) = \mathbf{0},$$

which implies

$$E_{F_\epsilon}\boldsymbol{\Psi}(Z, \gamma_\epsilon) = (1-\epsilon)E_F\boldsymbol{\Psi}(Z, \gamma_\epsilon) + \epsilon\boldsymbol{\Psi}(Z_0, \gamma_\epsilon) = \mathbf{0}. \tag{3.18}$$

Taking partial derivative of (3.18) with respect to ϵ yields

$$-E_F\Psi(Z,\gamma_\epsilon)+(1-\epsilon)E_F\left(\frac{\partial\Psi(Z,\gamma_\epsilon)}{\partial\gamma}\right)\frac{\partial\gamma_\epsilon}{\partial\epsilon}+\Psi(Z_0,\gamma_\epsilon)+\epsilon\frac{\partial\Psi(Z_0,\gamma_\epsilon)}{\partial\gamma}\frac{\partial\gamma_\epsilon}{\partial\epsilon}=\mathbf{0}. \quad (3.19)$$

The first and last term in (3.19) approach to 0 and $\gamma_\epsilon \to \hat{\gamma}_\infty(F)$ as $\epsilon \downarrow 0$. Noting that the influence function of $\hat{\gamma}_R$ in this case is just $\partial\gamma_\epsilon/\partial\epsilon$. Hence upon taking ϵ to zero, (3.19) becomes

$$M(\hat{\gamma}_\infty(F))IF_{\hat{\gamma}_R}(Z_0,F)+\Psi(Z_0,\hat{\gamma}_\infty(F))=\mathbf{0}.$$

It follows that

$$IF_{\hat{\gamma}_R}(Z_0,F)=-M^{-1}(\hat{\gamma}_\infty(F))\Psi(Z_0,\hat{\gamma}_\infty(F)),$$

where

$$M(\hat{\gamma}_\infty(F))=E_F\left[\frac{\partial\Psi\left(Z,\hat{\gamma}_\infty(F)\right)}{\partial\gamma}\right]=E_F\left[\frac{\partial\rho^T(\hat{\gamma}_\infty(F))}{\partial\gamma}W_R(X)\frac{\partial\rho(\hat{\gamma}(F))}{\partial\gamma^T}\right].$$

□

The first term $-M^{-1}$ in the influence function (3.17) is a constant which does not depend on Z_0. Hence if the second term, function $|\Psi(Z,\gamma)|$, is bounded, then the influence function for the RSLS estimator is bounded and $\hat{\gamma}_R$ is robust. For linear regression model with $g(X,\boldsymbol{\theta})=X^T\boldsymbol{\theta}$, we have

$$\begin{aligned}\Psi(Z,\gamma) &= -\begin{pmatrix} X & 2(X^T\boldsymbol{\theta})X \\ 0 & 1\end{pmatrix} W_R(X)\begin{pmatrix} Y-X^T\boldsymbol{\theta} \\ Y^2-(X^T\boldsymbol{\theta})^2-\sigma^2\end{pmatrix}\\ &= -\begin{pmatrix} X & 2(X^T\boldsymbol{\theta})X \\ 0 & 1\end{pmatrix} W_R(X)\begin{pmatrix} \varepsilon \\ \varepsilon^2+2\varepsilon(X^T\boldsymbol{\theta})-\sigma^2\end{pmatrix}.\end{aligned}$$

For RSLSE-I, the weight matrix $W_{R1}(X)$ is zero when X is an outlier and it is

$W_{opt}(X)$ otherwise. Hence the corresponding $|\Psi(Z, \gamma)|$ is bounded in X as it equals to zero whenever X is an outlier. However, when X is not an outlier but Y is, ε can be very large. This and a non-zero weight matrix $W_{R1}(X) = W_{opt}(X)$ lead to an unbounded $|\Psi(Z, \gamma)|$. Hence the RSLSE-I is robust against X-outliers only but not Y-outliers. For RSLSE-II with $W_{R2}(X)$, $|\Psi(Z, \gamma)|$ is bounded both in X and ε, so RSLSE-II is robust against both X-outliers and Y-outliers.

A more detailed analytic proof the boundedness of the influence functions for RSLSE-I and RSLSE-II would be desirable but this is not available because the outlier indicator function in Ψ depends on the outlier detection criterion used which does not have a simple analytic expression. The above explanation of the boundedness assumes the indicator function is 100% accurate, that is, it will assign a value of zero to an outlier. This is a reasonable assumption since for influence functions we are only concerned with one single extreme outlier approaching infinity, and any reasonable outlier detection criterion should be able to produce such an indicator function.

3.4.2 Breakdown point

The breakdown point of a RSLS estimator is also tied directly to the outlier detection criterion involved. Recall from the definition of RSLS-I in Section 3.2 that we have used the minimum volume ellipsoid estimator (MVE) for outlier detection. We now derive the breakdown point of the RSLS-I estimator through that of the MVE.

We have noted previously in (3.11) that the breakdown point for the MVE, $\epsilon_n^*(MVE)$ is given by (Rousseeuw and Leroy, 1987):

$$\epsilon_n^*(MVE) = (\left[\frac{n}{2}\right] - q + 1)/n \rightarrow 0.5$$

as $n \rightarrow \infty$. For the MM-estimator, the breakdown point depends on the breakdown point of the initial estimator in the computation. Yohai (1987) showed that

$$\epsilon_n^*(MM) \leq \min\left\{\epsilon_{n0}^*, \frac{n-2q}{2n}\right\}, \tag{3.20}$$

where ϵ_{n0}^* is the breakdown point for the initial estimator. If $\epsilon_{n0}^* \to 0.5$ as $n \to \infty$, then $\epsilon_n^*(MM) \to 0.5$ as well. The breakdown point for the RSLS estimators also depend on the initial estimator in the computation in Step 2. The breakdown point for the RSLS estimator against X-outliers is shown in following theorem.

Theorem 2. *For linear regression, if the MM-estimator is used as the initial estimator for the RSLSE-I and $n > 2q$, then the breakdown point for the RSLSE-I is*

$$\epsilon_n^*(\text{RSLSE-I}) \geq min\left\{\epsilon_n^*(MVE), \epsilon_n^*(MM)\right\}.$$

Proof: For brevity, the V, W and other related components of a RSLS estimator refer to that of the RSLSE-I. Suppose there are m X-outliers in the data such that

$$m = n \cdot \min\left\{\epsilon_n^*(MVE), \epsilon_n^*(MM)\right\}. \tag{3.21}$$

Without loss of generality, assume that $X_{n-m+1}, \ldots, X_n$ are the outliers and that $V(X_i) = 0$ for $i = n-m+1, \ldots, n$. The latter assumption is justified by the fact that $m \leq n \cdot \epsilon_n^*(MVE)$, and hence the MVE based outlier detection (3.12) can identify all m outliers. Because $m \leq n \cdot \epsilon_n^*(MM)$, the MM-estimator is also bounded. If we use the MM-estimator as the initial estimate $\hat{\gamma}^{(0)}$, the elements of $\hat{W}_{opt}(X_i)$ are bounded for $i = 1, \ldots, n-m$. Therefore, the elements of $\hat{W}_R(X_i)$ are bounded and $\hat{W}_R(X_i) \geq 0$ for all $i = 1, \ldots, n$. Recall that the RSLS estimator $\hat{\gamma}_R = (\hat{\boldsymbol{\theta}}_R^T, \hat{\sigma}_R^2)^T$ minimizes the objective function

$$Q_n^*(\gamma) = \sum_{n}^{i=1} \rho_i^T(\gamma)\hat{W}_R(X_i)\rho_i(\gamma). \tag{3.22}$$

The boundedness of the weight matrices $\hat{W}_{opt}(X_i)$ implies that

$$Q_n^*((\mathbf{0}^T, 1)^T) < \infty.$$

It follows that

$$Q_n^*(\hat{\gamma}_R) = \min_{\gamma} Q_n^*(\gamma) \leq Q_n^*((\mathbf{0}^T, 1)^T) < \infty. \tag{3.23}$$

From the definition of the ρ function in (1.5), we have for linear models

$$\rho_i(\gamma) = (Y_i - X_i^T\boldsymbol{\theta}, Y_i^2 - (X_i^T\boldsymbol{\theta})^2 - \sigma^2)^T. \tag{3.24}$$

Since the weight matrices $\hat{W}_R(X_i)$ in (3.22) are non-negative definite, and for non-trivial cases some of these are strictly positive definite, (3.23) and (3.24) imply that $Q_n^*(\gamma)$ is bounded if and only if $\boldsymbol{\theta}$ and σ^2 are both bounded. It then follows from the boundedness of $Q_n^*(\hat{\gamma}_R)$ in (3.23) that the RSLSE-I $\hat{\gamma}_R = (\hat{\boldsymbol{\theta}}_R^T, \hat{\sigma}^2)^T$ is bounded for m satisfying (3.21). Hence its breakdown point ϵ_n^*(RSLSE-I) is at least m/n, which is $\min\{\epsilon_n^*(MVE), \epsilon_n^*(MM)\}$. □

Since both the MVE and MM-estimator have high breakdown points, RSLSE-I also has a high breakdown point. The same arguments used above may also be applied to show that the RSLSE-II involving the MVE in Step 1 and MM-estimator in Step 2 of the computation algorithm also has a high breakdown point bounded below by $\min\{\epsilon_n^*(MVE), \epsilon_n^*(MM)\}$. We have seen that the breakdown points of the RSLS estimators depend critically on that of the methods involved in the two steps. The above result may not be valid if other outlier detection methods (instead of MVE) are used in Step 1 and/or other methods (instead of MM-method) are used for the initial value estimation in Step 2.

3.4.3 Asymptotic distribution

We have obtained the asymptotic influence functions for the RSLS estimators in Theorem 1. Huber (1981) gave the relationship between the asymptotic distribution of an estimator and its influence function. Using this result and Theorem 1, We now give the asymptotic distributions of the RSLS estimators in Theorem 3.

Theorem 3. *Under some regularity conditions, as* $n \to \infty$,

$$\sqrt{n}(\hat{\gamma}_R - \gamma_0) \xrightarrow{L} N(\boldsymbol{0}, (A^*)^{-1}B^*(A^*)^{-1}),$$

where

$$\begin{aligned} A^* &= E\left[\frac{\partial \rho^T(\gamma_0)}{\partial \gamma} W_R(X) \frac{\partial \rho(\gamma_0)}{\partial \gamma^T}\right], \\ B^* &= E\left[\frac{\partial \rho^T(\gamma_0)}{\partial \gamma} W_R(X)\rho(\gamma_0)\rho^T(\gamma_0) W_R(X) \frac{\partial \rho(\gamma_0)}{\partial \gamma^T}\right], \end{aligned}$$

γ_0 is the true value and A^* is the M matrix in (3.17).

Proof: From Huber (1981), the asymptotic variance for $\sqrt{n}\hat{\gamma}_R$ is

$$\begin{aligned} V &= E\left[IF_{\hat{\gamma}_R}(Z,F) IF_{\hat{\gamma}_R}(Z,F)^T\right] \\ &= E\left[M^{-1}(\hat{\gamma}_\infty(F))\Psi(Z,\hat{\gamma}_\infty(F))\Psi^T(Z,\hat{\gamma}_\infty(F))M^{-1}(\hat{\gamma}_\infty(F))\right] \\ &= M^{-1}(\hat{\gamma}_\infty(F)) E\left[\Psi(Z,\hat{\gamma}_\infty(F))\Psi^T(Z,\hat{\gamma}_\infty(F))\right] M^{-1}(\hat{\gamma}_\infty(F)) \\ &= (A^*)^{-1} E\left[\frac{\partial \rho^T(\gamma_0)}{\partial \gamma} W_R(X)\rho(\gamma_0)\rho^T(\gamma_0) W_R(X) \frac{\partial \rho(\gamma_0)}{\partial \gamma^T}\right] (A^*)^{-1} \\ &= (A^*)^{-1} B^* (A^*)^{-1}. \end{aligned}$$

□

The result in Theorem 3 is consistent with that in Wang and Leblance (2008). In practice matrices A^* and B^* can be estimated by

$$\hat{A}^* = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial \rho_i^T(\hat{\gamma}_R)}{\partial \gamma}\hat{W}_R(X_i)\frac{\partial \rho_i(\hat{\gamma}_R)}{\partial \gamma^T}\right], \tag{3.25}$$

$$\hat{B}^* = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial \rho_i^T(\hat{\gamma}_R)}{\partial \gamma}\hat{W}_R(X_i)\rho_i(\hat{\gamma}_R)\rho_i^T(\hat{\gamma}_R)\hat{W}_R(X_i)\frac{\partial \rho_i(\hat{\gamma}_R)}{\partial \gamma^T}\right]. \tag{3.26}$$

For linear regression models with $g(X,\boldsymbol{\theta}) = X^T\boldsymbol{\theta}$, (3.25) and (3.26) can be expressed as the following,

$$\hat{A}^* = \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix}\frac{\partial X^T\hat{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} & 2X^T\hat{\boldsymbol{\theta}}\frac{\partial X^T\hat{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} \\ 0 & 1\end{bmatrix}\hat{W}_R(X_i)\begin{bmatrix}\frac{\partial X^T\hat{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}^T} & 0 \\ 2X^T\hat{\boldsymbol{\theta}}\frac{\partial X^T\hat{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}^T} & 1\end{bmatrix},$$

$$\hat{B}^* = \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix}\frac{\partial X^T\hat{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} & 2X^T\hat{\boldsymbol{\theta}}\frac{\partial X^T\hat{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} \\ 0 & 1\end{bmatrix}\hat{W}_R(X_i)\rho_i(\hat{\gamma}_R)\rho_i^T(\hat{\gamma}_R)\hat{W}_R(X_i)\begin{bmatrix}\frac{\partial X^T\hat{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}^T} & 0 \\ 2X^T\hat{\boldsymbol{\theta}}\frac{\partial X^T\hat{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}^T} & 1\end{bmatrix}.$$

3.5 Summary

In this chapter, we presented two robust second-order least-squares (RSLS) estimators with focus on their properties. The empirical influence function of the original second-order least-squares (SLS) estimator shows that it is not robust against either X-outliers or Y-outliers. Hence, we proposed two RSLS estimators and the empirical influence functions of these two estimators show that the RSLSE-I is robust against X-outliers and the RSLSE-II is also robust against other type outliers including Y-outliers. The iterative computing procedure of the RSLS estimators is also studied. We applied the RSLS estimators to one example of simple linear regression. The two RSLS estimators work very well for both data with outliers and data without outliers. We presented and proved the asymptotic influence function and breakdown point of the RSLS estimators. The RSLS estimators have high breakdown point if the initial estimator has high breakdown point. We also gave the asymptotic normal distributions for the RSLS estimators. The finite sample size behaviour of the RSLS estimators will be investigated through simulation studies in next chapter.

Chapter 4

Simulation Studies and Applications

We have proposed two versions of the RSLS estimator in Chapter 3. In this chapter, we examine the performance of the RSLS estimators through simulation studies involving a simple linear regression model and a multiple linear regression model. In Section 4.1, we compare the bias, standard error, confidence interval coverage and the length of the confidence interval of the RSLSE-I and RSLSE-II with those of the LS estimator, SLS estimator and MM-estimator for the simple linear regression model. In Section 4.2, we compare the RSLSE-I and RSLSE-II with the LS estimator, SLS estimator and MM-estimator for a multiple linear regression model. In Section 4.3, we apply RSLSE-I and RSLSE-II to analyse two real data sets.

4.1 Simulation study I: simple linear regression model

In this section, we use a simulation study to investigate the finite sample behaviour of the RSLS estimators and compare them with the LS estimator, SLS estimator and MM-estimator for the simple linear regression. We will examine the consistency and efficiency of the estimators through their biases and standard errors, respectively. We will also examine the coverage levels of the corresponding confidence intervals and the lengths of the confidence intervals.

The simple linear regression model that we use for this simulation study is

$$y_i = \theta_0 + \theta_1 x_i + \varepsilon_i,\ i = 1, \ldots, n,$$

where true parameter values are fixed at $\theta_0 = 5$ and $\theta_1 = 3$. The choices of these true values are rather arbitrary and have little impact on the conclusions. To simulate situations with and without X-outliers, we use the following two distributions to generate the x_i values:

(D1) $x_i \sim \text{Unif}(5, 10)$,

(D2) $x_i \sim 0.9\,\text{Unif}(5, 10) + 0.1\delta_{30}$.

Distribution (D1) does not generate X-outliers, but distribution (D2) generates 10% X-outliers at value 30. We also generate the random errors ε_i using three different distributions:

(E1) $\varepsilon_i \sim N(0, \sigma^2)$, with $\sigma^2 = 1$,

(E2) $\varepsilon_i \sim \frac{1}{\sqrt{3}} t_3$, with the mean $E(\varepsilon_i) = 0$ and variance $\sigma^2 = Var(\varepsilon_i) = 1$,

(E3) $\varepsilon_i \sim \frac{\chi_5^2 - 5}{\sqrt{10}}$, with the mean $E(\varepsilon_i) = 0$ and variance $\sigma^2 = Var(\varepsilon_i) = 1$.

Distributions (E1) and (E2) are symmetric, and (E3) is asymmetric.

We consider three sample sizes $n = 20, 30$ and 50. For each sample size, we generate 1000 random samples of (x_i, y_i) using a combination of the x and ε distributions, say (D1,E1). For each of these 1000 samples, we compute the RSLS estimate, LS estimate, SLS estimate, and MM-estimate for θ_0 and θ_1, as well as the associated confidence intervals. We then compute the bias, standard error (s.e.), confidence interval coverage level and length of the confidence interval associated with each estimator using the corresponding 1000 estimates and confidence intervals. Two confidence levels, 90% and 95%, are used to construct the confidence intervals. We use the statistical software R for the computation. The LS estimates and MM-estimates are computed directly by using R-functions "lm" and "lmrob", respectively. To avoid potential optimization problems involved in the computation of the SLS estimates, we use the MM-estimates as the initial estimate to estimate the weight matrix. To save time, one-step estimate is used for the RSLS estimators.

Let $\hat{\theta}_i^{(j)}$ be the estimate of θ_i based on the j^{th} simulated sample where $j = 1, 2, \ldots, 1000$. The bias and s.e. of $\hat{\theta}_i$ are computed by

$$\text{bias}(\hat{\theta}_i) = \frac{1}{1000} \sum_{j=1}^{1000} \left(\hat{\theta}_i^{(j)} - \theta_i \right),$$

$$\text{s.e.}(\hat{\theta}_i) = \left[\frac{1}{1000 - 1} \sum_{j=1}^{1000} (\hat{\theta}_i^{(j)} - \bar{\hat{\theta}}_i)^2 \right]^{\frac{1}{2}},$$

where $\bar{\hat{\theta}}_i = \sum_{j=1}^{1000} \hat{\theta}_i^{(j)}/1000$. The $100(1-\alpha)\%$ confidence interval for θ_i is computed for each simulation run as follows:

$$\hat{\theta}_i^{(j)} \pm t_{\alpha/2, n-p} S_i^{(j)},$$

where $S_i^{(j)}$ is the standard error estimated from the asymptotic variance of $\hat{\theta}_i$ using the j^{th} simulated sample. The length of the confidence interval is

$$L_i^{(j)} = 2t_{\alpha/2,n-p}S_i^{(j)},$$

and the average length over 1000 simulation runs is $\text{Ave}(L_i) = \sum_{j=1}^{1000} L_i^{(j)}/1000$. To compute the coverage level of the confidence interval, denote

$$c_i^{(j)} = \begin{cases} 1, & \text{if } \theta_i \in \left(\theta_i^{(j)} - t_{\alpha/2,n-p}S_i^{(j)}, \theta_i^{(j)} + t_{\alpha/2,n-p}S_i^{(j)}\right), \\ 0, & \text{otherwise.} \end{cases}$$

Then the empirical coverage level is computed by

$$EC_i = \frac{1}{1000}\sum_{j=1}^{1000} c_i^{(j)}.$$

We now compare the estimators using bias($\hat{\theta}_i$), s.e.($\hat{\theta}_i$), $Ave(L_i)$ and EC_i based on 1000 simulated random samples. Tables 4.1 - 4.14 contain the simulated values of these quantities for all 6 combinations of x-distribution and ε-distribution. We first look at Tables 4.1 - 4.3. The explanatory variable distribution corresponding to these three tables is (D1). Hence there are no outliers in the samples. From the simulation results in these tables, we obtain the following observations:

1. The LS estimator for θ_0 and θ_1 is the most efficient for error distribution (E1), as its standard errors are the smallest among that of the five estimators. Its bias is very small in all cases, reflecting the unbiasedness of the LS estimator.

2. The SLS estimator and RSLSE-I are similar in terms of bias and they both tend to have smaller biases at larger sample sizes for all error distributions, suggesting that they are asymptotically unbiased for both symmetric and asymmetric error

distributions.

3. The SLS estimator and RSLSE-I have smaller standard errors than the LS estimator for sample size $n = 50$ for the non-normal error distributions (E2) and (E3), which is consistent with the result from Leblanc and Wang (2008).

4. The RSLSE-II and MM-estimator seem to underestimate the standard deviation σ, especially for non-normal error distributions. The estimates of σ from the SLS estimator and RSLSE-I seem to be asymptotically unbiased.

5. The MM-estimator is asymptotically unbiased for symmetric error distributions, but it is asymptotically biased for asymmetric error distribution. In particular, the MM-estimator for θ_0 is biased for asymmetric error distribution.

6. The RSLSE-II is asymptotically biased for asymmetric error distributions, but is asymptotically unbiased for symmetric error distributions.

7. For most cases, the RSLSE-II has slightly larger bias and standard errors than that of the RSLSE-I.

We now examine Tables 4.4 - 4.6 which contain the simulation results of the RSLSE-I, RSLSE-II and MM-estimator for samples with 10% X-outliers generated by X-distribution (D2). Since the LS estimator and SLS estimator do not perform well in the presence of outliers, we did not include them in this comparison. The following are observations based on these three tables:

8. For symmetric error distributions, the biases of RSLSE-I and RSLSE-II are small and comparable. Further, they are also comparable to that of the MM-estimator. For error distribution (E1), the RSLSE-I has the smallest standard errors.

9. The RSLSE-II and the MM-estimator seem to be biased for asymmetric error distribution (E3). Their biases for estimating θ_0 are particularly noticeable at all sample sizes.

10. MM-estimator performs very well for error distribution (E2), a heavy tailed symmetric distribution.

Table 4.1: Simulation results for the simple linear regression model for distributions (D1) and (E1): the bias and the standard error.

Sample size and parameter	LSE bias (s.e.)	SLSE bias (s.e.)	RSLSE-I bias (s.e.)	RSLSE-II bias (s.e.)	MM bias (s.e.)
n = 20					
$\hat{\theta}_0$	-0.061	-0.070	-0.064	-0.074	-0.068
	(1.028)	(1.143)	(1.077)	(1.113)	(1.059)
$\hat{\theta}_1$	0.009	0.010	0.009	0.010	0.010
	(0.139)	(0.155)	(0.146)	(0.151)	(0.143)
$\hat{\sigma}$	-0.014	-0.077	-0.069	-0.114	-0.021
	(0.164)	(0.156)	(0.156)	(0.181)	(0.223)
n = 30					
$\hat{\theta}_0$	0.025	0.006	0.010	0.010	0.021
	(0.901)	(0.957)	(0.930)	(0.953)	(0.925)
$\hat{\theta}_1$	-0.003	-0.001	-0.001	-0.001	-0.003
	(0.121)	(0.130)	(0.126)	(0.128)	(0.124)
$\hat{\sigma}$	-0.013	-0.051	-0.048	-0.088	-0.024
	(0.132)	(0.128)	(0.128)	(0.148)	(0.180)
n = 50					
$\hat{\theta}_0$	0.031	0.037	0.036	0.045	0.040
	(0.723)	(0.763)	(0.752)	(0.774)	(0.749)
$\hat{\theta}_1$	-0.004	-0.005	-0.005	-0.006	-0.005
	(0.095)	(0.101)	(0.099)	(0.102)	(0.099)
$\hat{\sigma}$	-0.006	-0.027	-0.027	-0.053	-0.017
	(0.098)	(0.096)	(0.096)	(0.104)	(0.133)

Table 4.2: Simulation results for the simple linear regression model for distributions (D1) and (E2): the bias and the standard error.

Sample size and parameter	LSE bias (s.e.)	SLSE bias (s.e.)	RSLSE-I bias (s.e.)	RSLSE-II bias (s.e.)	MM bias (s.e.)
n = 20					
$\hat{\theta}_0$	-0.057	-0.044	-0.046	-0.044	-0.047
	(1.007)	(0.926)	(0.890)	(0.853)	(0.790)
$\hat{\theta}_1$	0.008	0.007	0.007	0.006	0.007
	(0.137)	(0.125)	(0.119)	(0.116)	(0.108)
$\hat{\sigma}$	-0.102	-0.150	-0.144	-0.387	-0.325
	(0.380)	(0.370)	(0.371)	(0.170)	(0.173)
n = 30					
$\hat{\theta}_0$	-0.045	-0.020	-0.019	-0.022	-0.027
	(0.866)	(0.778)	(0.767)	(0.734)	(0.690)
$\hat{\theta}_1$	0.005	0.002	0.002	0.003	0.003
	(0.114)	(0.104)	(0.102)	(0.098)	(0.093)
$\hat{\sigma}$	-0.082	-0.112	-0.111	-0.354	-0.313
	(0.388)	(0.378)	(0.378)	(0.140)	(0.142)
n = 50					
$\hat{\theta}_0$	0.007	-0.019	-0.017	-0.025	-0.018
	(0.756)	(0.640)	(0.636)	(0.588)	(0.551)
$\hat{\theta}_1$	-0.001	0.002	0.002	0.003	0.002
	(0.098)	(0.084)	(0.083)	(0.077)	(0.072)
$\hat{\sigma}$	-0.052	-0.070	-0.069	-0.346	-0.328
	(0.324)	(0.323)	(0.323)	(0.110)	(0.108)

Table 4.3: Simulation results for the simple linear regression model for distributions (D1) and (E3): the bias and the standard error.

Sample size and parameter	LSE bias (s.e.)	SLSE bias (s.e.)	RSLSE-I bias (s.e.)	RSLSE-II bias (s.e.)	MM bias (s.e.)
n = 20					
$\hat{\theta}_0$	0.028	-0.043	-0.001	-0.108	-0.115
	(1.069)	(0.917)	(0.885)	(0.926)	(1.018)
$\hat{\theta}_1$	-0.004	0.002	-0.000	0.001	0.001
	(0.143)	(0.124)	(0.118)	(0.123)	(0.136)
$\hat{\sigma}$	-0.028	-0.093	-0.081	-0.229	-0.157
	(0.217)	(0.207)	(0.210)	(0.204)	(0.211)
n = 30					
$\hat{\theta}_0$	0.089	0.038	0.053	-0.034	-0.048
	(0.896)	(0.729)	(0.715)	(0.726)	(0.818)
$\hat{\theta}_1$	-0.012	-0.007	-0.008	-0.007	-0.009
	(0.121)	(0.098)	(0.096)	(0.097)	(0.110)
$\hat{\sigma}$	-0.022	-0.063	-0.057	-0.193	-0.152
	(0.185)	(0.180)	(0.181)	(0.161)	(0.160)
n = 50					
$\hat{\theta}_0$	-0.024	-0.020	-0.017	-0.113	-0.149
	(0.734)	(0.590)	(0.575)	(0.576)	(0.669)
$\hat{\theta}_1$	0.004	0.003	0.003	0.004	0.004
	(0.097)	(0.077)	(0.075)	(0.075)	(0.088)
$\hat{\sigma}$	-0.003	-0.026	-0.023	-0.174	-0.155
	(0.151)	(0.150)	(0.150)	(0.127)	(0.128)

Table 4.4: Simulation results for the simple linear regression model for distributions (D2) and (E1): the bias and the standard error.

Sample size and parameter	RSLSE-I bias (s.e.)	RSLSE-II bias (s.e.)	MM bias (s.e.)
n = 20			
$\hat{\theta}_0$	-0.018	-0.012	-0.016
	(1.065)	(1.104)	(1.081)
$\hat{\theta}_1$	0.002	0.001	0.002
	(0.150)	(0.155)	(0.152)
$\hat{\sigma}$	-0.075	-0.102	0.153
	(0.163)	(0.178)	(0.255)
n = 30			
$\hat{\theta}_0$	-0.015	-0.015	-0.012
	(0.930)	(0.952)	(0.944)
$\hat{\theta}_1$	0.003	0.003	0.002
	(0.126)	(0.129)	(0.128)
$\hat{\sigma}$	-0.055	-0.072	0.140
	(0.133)	(0.143)	(0.205)
n = 50			
$\hat{\theta}_0$	-0.031	-0.025	-0.027
	(0.763)	(0.771)	(0.771)
$\hat{\theta}_1$	0.005	0.004	0.004
	(0.101)	(0.102)	(0.102)
$\hat{\sigma}$	-0.025	-0.036	0.158
	(0.105)	(0.111)	(0.158)

Table 4.5: Simulation results for the simple linear regression model for distributions (D2) and (E2): the bias and the standard error.

Sample size and parameter	RSLSE-I bias (s.e.)	RSLSE-II bias (s.e.)	MM bias (s.e.)
n = 20			
$\hat{\theta}_0$	0.020	-0.010	-0.007
	(1.048)	(0.882)	(0.836)
$\hat{\theta}_1$	-0.001	0.002	0.002
	(0.147)	(0.123)	(0.117)
$\hat{\sigma}$	-0.146	-0.347	-0.179
	(0.351)	(0.204)	(0.232)
n = 30			
$\hat{\theta}_0$	0.010	0.007	0.009
	(0.900)	(0.756)	(0.703)
$\hat{\theta}_1$	-0.002	-0.001	-0.001
	(0.121)	(0.103)	(0.095)
$\hat{\sigma}$	-0.117	-0.310	-0.179
	(0.342)	(0.156)	(0.169)
n = 50			
$\hat{\theta}_0$	0.043	0.019	0.025
	(0.776)	(0.628)	(0.584)
$\hat{\theta}_1$	-0.006	-0.003	-0.004
	(0.103)	(0.083)	(0.077)
$\hat{\sigma}$	-0.060	-0.295	-0.188
	(0.308)	(0.124)	(0.134)

Table 4.6: Simulation results for the simple linear regression model for distributions (D2) and (E3): the bias and the standard error.

Sample size and parameter	RSLSE-I bias (s.e.)	RSLSE-II bias (s.e.)	MM bias (s.e.)
n = 20			
$\hat{\theta}_0$	-0.032	-0.097	-0.116
	(1.090)	(1.065)	(1.040)
$\hat{\theta}_1$	0.005	0.004	0.005
	(0.155)	(0.150)	(0.147)
$\hat{\sigma}$	-0.094	-0.212	0.006
	(0.215)	(0.208)	(0.253)
n = 30			
$\hat{\theta}_0$	-0.021	-0.096	-0.118
	(0.925)	(0.891)	(0.865)
$\hat{\theta}_1$	0.002	0.003	0.003
	(0.127)	(0.121)	(0.118)
$\hat{\sigma}$	-0.062	-0.177	-0.006
	(0.186)	(0.174)	(0.199)
n = 50			
$\hat{\theta}_0$	0.009	-0.062	-0.097
	(0.759)	(0.720)	(0.702)
$\hat{\theta}_1$	0.000	0.000	0.001
	(0.102)	(0.096)	(0.094)
$\hat{\sigma}$	-0.034	-0.146	-0.012
	(0.148)	(0.129)	(0.143)

To examine the asymptotic normality of the RSLSE-I and RSLSE-II, we constructed histograms for estimates $\hat{\theta}_0$ and $\hat{\theta}_1$ given by these two estimators. Representative histograms are given in Figures 4.1 - 4.4. Figures 4.1 and 4.2 are for the case with error distribution (E1), no X-outliers and sample size $n = 30$. Figures 4.3 and 4.4 are for the case with error distribution (E3), 10% X-outliers and sample size $n = 30$. All the histograms are bell shaped and the corresponding normal Q-Q plots are close to straight lines. These suggest that both the RSLSE-I and RSLSE-II are approximately normally distributed, even for sample size as small as $n = 30$. Similar observations have been obtained based on the histograms and Q-Q plots for other combinations of sample sizes (not included), X-distribution and ε-distribution. These suggest that normal distribution based inference is justified when using these estimators. In particular, we may use the estimated value and its standard error to construct t-based or z-based confidence intervals.

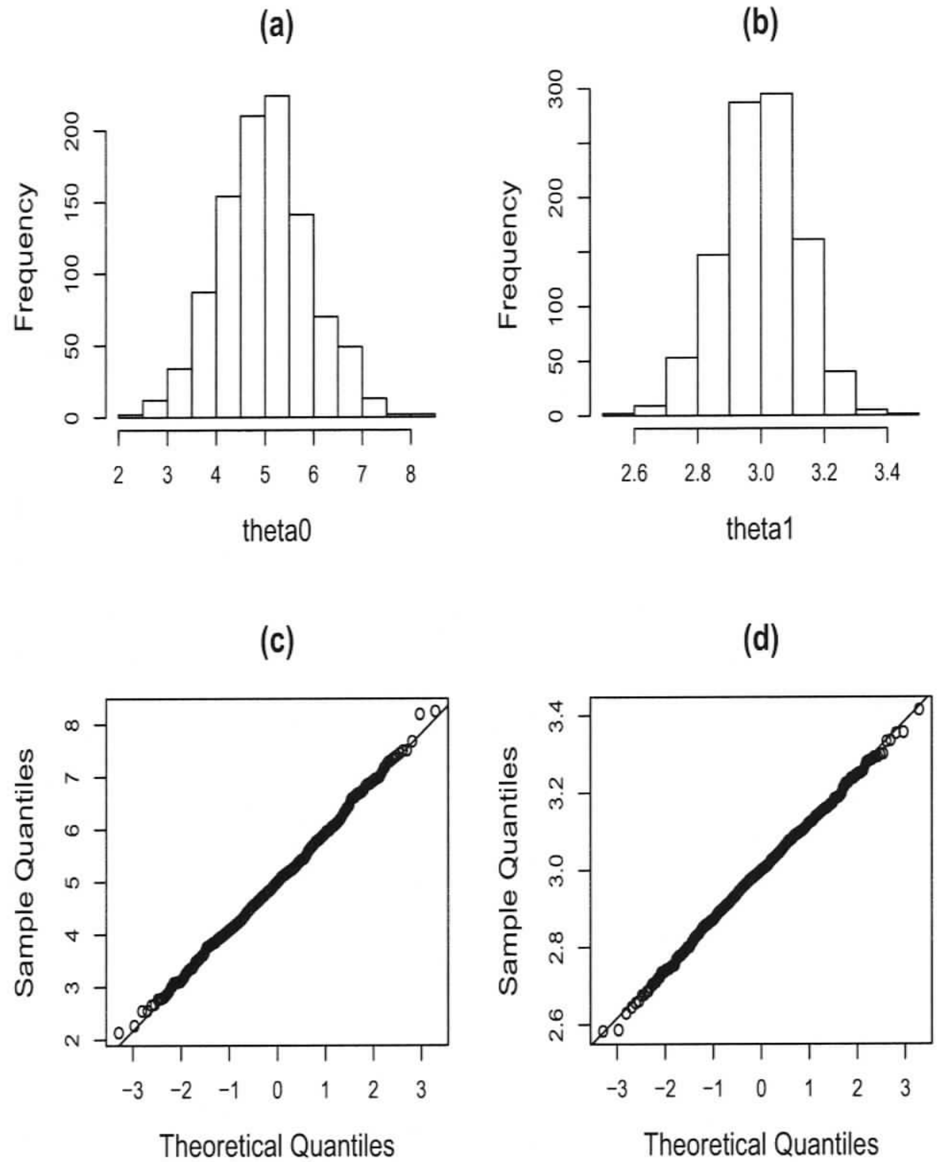


Figure 4.1: Histograms and Q-Q plots for RSLSE-I for the case with error distribution (E1), no X-outliers and sample size $n = 30$: (a) histogram for $\hat{\theta}_0$, (b) histogram for $\hat{\theta}_1$, (c) Q-Q plot for $\hat{\theta}_0$, (d) Q-Q plot for $\hat{\theta}_1$.

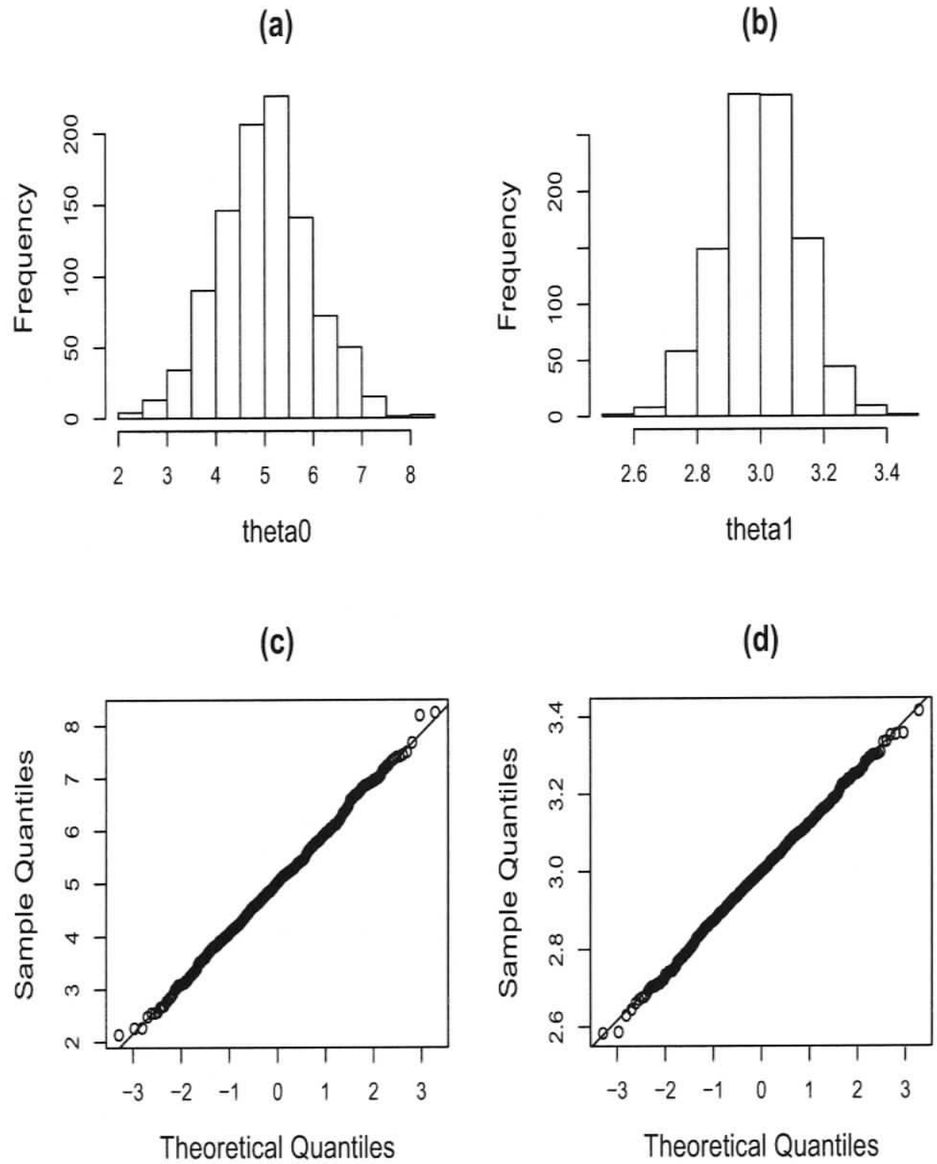


Figure 4.2: Histograms and Q-Q plots for RSLSE-II for the case with error distribution (E1), no X-outliers and sample size $n = 30$: (a) histogram for $\hat{\theta}_0$, (b) histogram for $\hat{\theta}_1$, (c) Q-Q plot for $\hat{\theta}_0$, (d) Q-Q plot for $\hat{\theta}_1$.

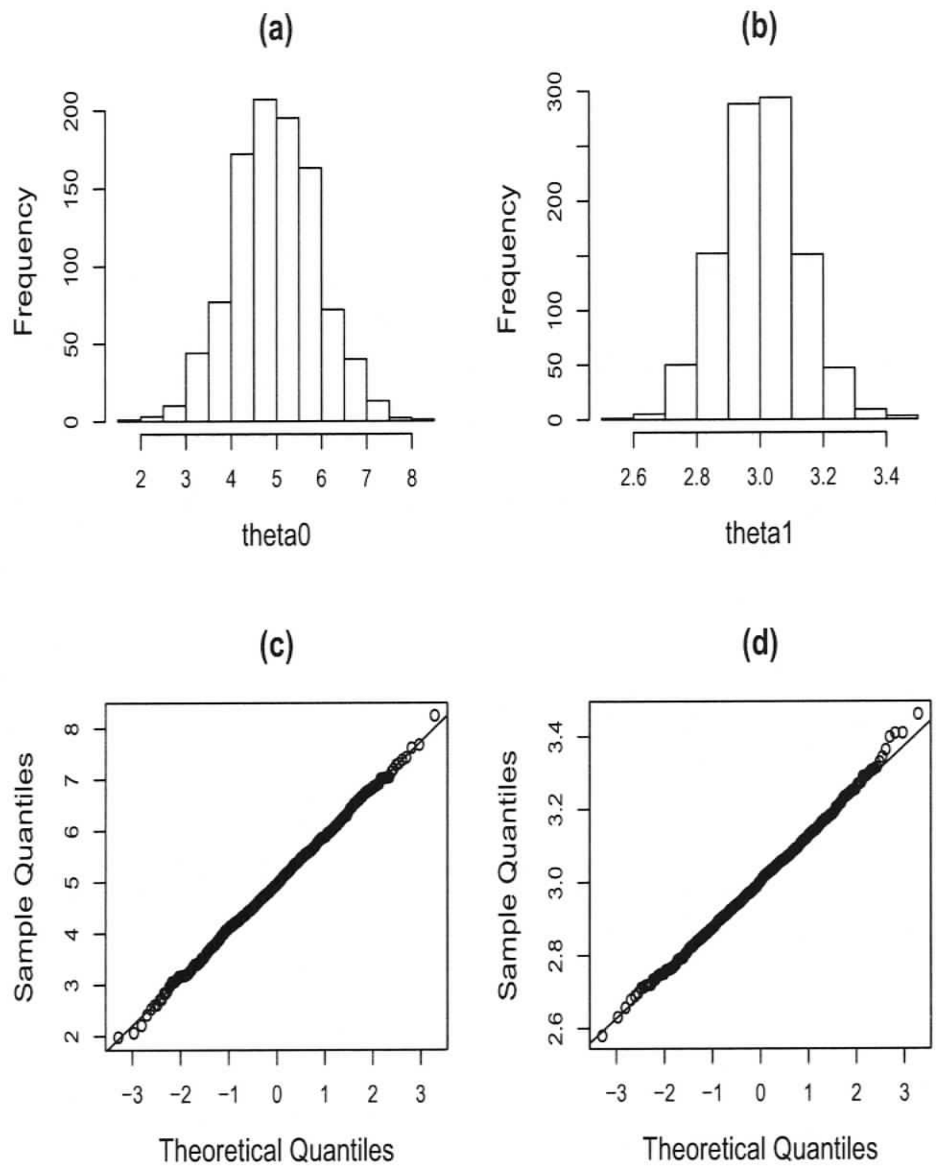


Figure 4.3: Histograms and Q-Q plots for RSLSE-I for the case with error distribution (E3), 10% X-outliers and sample size $n = 30$: (a) histogram for $\hat{\theta}_0$, (b) histogram for $\hat{\theta}_1$, (c) Q-Q plot for $\hat{\theta}_0$, (d) Q-Q plot for $\hat{\theta}_1$.

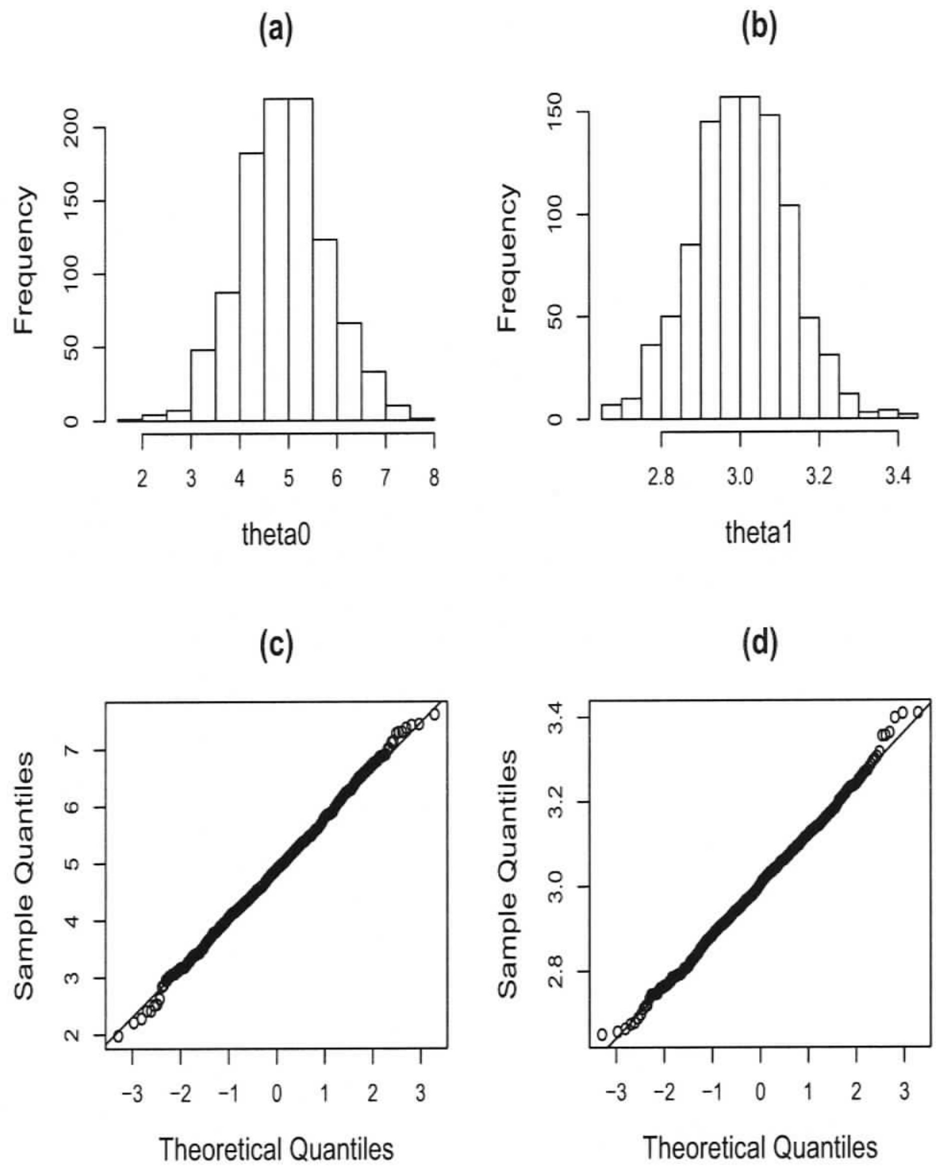


Figure 4.4: Histograms and Q-Q plots for RSLSE-II for the case with error distribution (E3), 10% X-outliers and sample size $n = 30$: (a) histogram for $\hat{\theta}_0$, (b) histogram for $\hat{\theta}_1$, (c) Q-Q plot for $\hat{\theta}_0$, (d) Q-Q plot for $\hat{\theta}_1$.

For the 90% and 95% nominal levels, the (empirical) coverage level and the average length of the confidence intervals for θ_0 and θ_1 are presented in Tables 4.7 - 4.14. In terms the coverage level, a method is superior if its coverage level is the closest to the nominal level. In terms of the average length, a method is better if it has the smallest average length at a fixed nominal level. We now make the following observations for cases without X-outliers (Tables 4.7, 4.8, 4.11 and 4.12).

1. The LS estimator outperforms all others in terms of the coverage level.
2. Compared with the LS estimator and MM-estimator, the RSLSE-I has a slightly lower coverage level but it has a shorter average length. Also, the RSLSE-I performs better than the SLS estimator in that it has significantly better coverage level than the latter but only slightly larger average length.
3. The RSLSE-II has lower coverage and shorter average interval length than the RSLSE-I.
4. For large sample size, the coverage levels of intervals based on all the estimators are quite close to the nominal level.

For cases with X-outliers (Tables 4.9, 4.10, 4.13 and 4.14), we observe that

5. The RSLSE-I and MM-estimator based confidence intervals are very similar. The coverage levels of the MM-estimator based intervals are always closer to the nominal levels, but they have longer interval lengths.
6. Compared to the RSLSE-I, the RSLSE-II has lower coverage level but has slightly shorter interval length.
7. The coverage levels for three robust estimators based confidence intervals are all lower then the nominal levels.

Table 4.7: Simulation results for the simple linear regression model: the average length and coverage of 90% confidence intervals for θ_0.

Distributions and sample size	LSE coverage (length)	SLSE coverage (length)	RSLSE-I coverage (length)	RSLSE-II coverage (length)	MM coverage (length)
(D1) and (E1)					
n = 20	0.896	0.796	0.819	0.801	0.887
	(3.534)	(3.055)	(3.082)	(3.005)	(3.641)
n = 30	0.890	0.830	0.842	0.822	0.878
	(2.984)	(2.717)	(2.723)	(2.671)	(3.052)
n = 50	0.903	0.860	0.866	0.847	0.890
	(2.464)	(2.336)	(2.338)	(2.304)	(2.529)
(D1) and (E3)					
n = 20	0.885	0.829	0.852	0.799	0.866
	(3.491)	(2.558)	(2.585)	(2.440)	(3.425)
n = 30	0.882	0.873	0.874	0.850	0.884
	(2.958)	(2.285)	(2.268)	(2.171)	(2.833)
n = 50	0.901	0.870	0.890	0.859	0.899
	(2.526)	(1.993)	(1.955)	(1.871)	(2.358)

Table 4.8: Simulation results for the simple linear regression model: the average length and coverage of 90% confidence intervals for θ_1.

Distributions and sample size	LSE coverage (length)	SLSE coverage (length)	RSLSE-I coverage (length)	RSLSE-II coverage (length)	MM coverage (length)
(D1) and (E1)					
n = 20	0.897	0.788	0.813	0.784	0.882
	(0.478)	(0.410)	(0.414)	(0.404)	(0.492)
n = 30	0.890	0.833	0.840	0.824	0.877
	(0.403)	(0.365)	(0.366)	(0.360)	(0.412)
n = 50	0.892	0.857	0.858	0.845	0.896
	(0.325)	(0.308)	(0.308)	(0.304)	(0.334)
(D1) and (E3)					
n = 20	0.887	0.829	0.848	0.821	0.886
	(0.473)	(0.342)	(0.343)	(0.327)	(0.464)
n = 30	0.882	0.876	0.870	0.854	0.901
	(0.400)	(0.305)	(0.301)	(0.290)	(0.378)
n = 50	0.907	0.890	0.891	0.872	0.906
	(0.335)	(0.262)	(0.256)	(0.247)	(0.311)

Table 4.9: Simulation results for the simple linear regression model: the average length and coverage of 90% confidence intervals for θ_0.

Distributions and sample size	RSLSE-I coverage (length)	RSLSE-II coverage (length)	MM coverage (length)
(D2) and (E1)			
n = 20	0.877	0.852	0.889
	(3.468)	(3.376)	(3.742)
n = 30	0.879	0.862	0.886
	(3.005)	(2.958)	(3.171)
n = 50	0.888	0.881	0.898
	(2.509)	(2.486)	(2.603)
(D2) and (E3)			
n = 20	0.860	0.809	0.876
	(3.349)	(2.960)	(3.468)
n = 30	0.878	0.841	0.882
	(2.950)	(2.638)	(2.948)
n = 50	0.895	0.865	0.889
	(2.475)	(2.222)	(2.395)

Table 4.10: Simulation results for the simple linear regression model: the average length and coverage of 90% confidence intervals for θ_1.

Distributions and sample size	RSLSE-I coverage (length)	RSLSE-II coverage (length)	MM coverage (length)
(D2) and (E1)			
n = 20	0.867	0.843	0.875
	(0.480)	(0.467)	(0.520)
n = 30	0.874	0.861	0.885
	(0.411)	(0.404)	(0.434)
n = 50	0.889	0.884	0.898
	(0.334)	(0.331)	(0.346)
(D2) and (E3)			
n = 20	0.860	0.818	0.864
	(0.465)	(0.410)	(0.481)
n = 30	0.874	0.842	0.891
	(0.403)	(0.361)	(0.401)
n = 50	0.885	0.860	0.898
	(0.328)	(0.295)	(0.317)

Table 4.11: Simulation results for the simple linear regression model: the average length and coverage of 95% confidence intervals for θ_0.

Distributions and sample size	LSE coverage (length)	SLSE coverage (length)	RSLSE-I coverage (length)	RSLSE-II coverage (length)	MM coverage (length)
(D1) and (E1)					
n = 20	0.948	0.869	0.889	0.868	0.939
	(4.282)	(3.701)	(3.734)	(3.640)	(4.412)
n = 30	0.943	0.891	0.905	0.884	0.935
	(3.594)	(3.271)	(3.280)	(3.216)	(3.676)
n = 50	0.963	0.917	0.920	0.904	0.945
	(2.954)	(2.801)	(2.802)	(2.763)	(3.032)
(D1) and (E3)					
n = 20	0.945	0.902	0.904	0.863	0.920
	(4.230)	(3.099)	(3.132)	(2.956)	(4.150)
n = 30	0.940	0.945	0.944	0.921	0.937
	(3.562)	(2.752)	(2.732)	(2.614)	(3.412)
n = 50	0.956	0.942	0.948	0.920	0.945
	(3.030)	(2.369)	(2.345)	(2.244)	(2.829)

Table 4.12: Simulation results for the simple linear regression model: the average length and coverage of 95% confidence intervals for θ_1.

Distributions and sample size	LSE coverage (length)	SLSE coverage (length)	RSLSE-I coverage (length)	RSLSE-II coverage (length)	MM coverage (length)
(D1) and (E1)					
n = 20	0.947	0.856	0.881	0.861	0.932
	(0.580)	(0.497)	(0.501)	(0.489)	(0.596)
n = 30	0.948	0.886	0.906	0.893	0.935
	(0.485)	(0.440)	(0.441)	(0.433)	(0.497)
n = 50	0.964	0.921	0.926	0.909	0.945
	(0.390)	(0.369)	(0.369)	(0.364)	(0.400)
(D1) and (E3)					
n = 20	0.951	0.902	0.903	0.877	0.936
	(0.573)	(0.414)	(0.416)	(0.396)	(0.562)
n = 30	0.941	0.925	0.927	0.916	0.943
	(0.482)	(0.368)	(0.363)	(0.349)	(0.455)
n = 50	0.960	0.937	0.940	0.926	0.956
	(0.402)	(0.311)	(0.307)	(0.296)	(0.374)

Table 4.13: Simulation results for the simple linear regression model: the average length and coverage of 95% confidence intervals for θ_0.

Distributions and sample size	RSLSE-I coverage (length)	RSLSE-II coverage (length)	MM coverage (length)
(D2) and (E1)			
n = 20	0.929	0.908	0.936
	(4.202)	(4.090)	(4.534)
n = 30	0.928	0.920	0.936
	(3.619)	(3.562)	(3.818)
n = 50	0.941	0.937	0.949
	(3.008)	(2.980)	(3.121)
(D2) and (E3)			
n = 20	0.926	0.894	0.934
	(4.058)	(3.586)	(4.201)
n = 30	0.930	0.903	0.937
	(3.552)	(3.177)	(3.550)
n = 50	0.930	0.913	0.936
	(2.967)	(2.664)	(2.871)

Table 4.14: Simulation results for the simple linear regression model: the average length and coverage of 95% confidence intervals for θ_1.

Distributions and sample size	RSLSE-I coverage (length)	RSLSE-II coverage (length)	MM coverage (length)
(D2) and (E1)			
n = 20	0.922	0.901	0.932
	(0.581)	(0.566)	(0.630)
n = 30	0.934	0.922	0.941
	(0.495)	(0.487)	(0.522)
n = 50	0.946	0.944	0.950
	(0.400)	(0.396)	(0.415)
(D2) and (E3)			
n = 20	0.917	0.885	0.920
	(0.563)	(0.497)	(0.582)
n = 30	0.932	0.908	0.943
	(0.485)	(0.434)	(0.483)
n = 50	0.938	0.927	0.939
	(0.394)	(0.354)	(0.380)

4.2 Simulation study II: multiple linear regression model

In this section, we continue our simulation study to investigate the finite sample behaviour of the RSLS estimators and compare them with the LS estimator, SLS estimator and MM-estimator with a more complicated example, a multiple linear regression model given below:

$$y_i = \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} + \theta_3 x_{i3} + \varepsilon_i,\ i = 1, \ldots, n,$$

where the true parameter values are $\theta_0 = 10$, $\theta_1 = -1$, $\theta_2 = 5$ and $\theta_3 = 2$. The explanatory variables x_{i1}, x_{i2}, x_{i3} are generated as follows:

$$\textbf{(D3)} \quad \begin{pmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \end{pmatrix} \sim N(\boldsymbol{\mu_1}, \Sigma_1), \text{ with } \boldsymbol{\mu_1} = \begin{pmatrix} 2 \\ 3 \\ 6 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{pmatrix}.$$

$$\textbf{(D4)} \quad \begin{pmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \end{pmatrix} \sim 0.9\, N(\boldsymbol{\mu_1}, \Sigma_1) + 0.1\, N(\boldsymbol{\mu_2}, \Sigma_2), \text{ with } \boldsymbol{\mu_2} = \begin{pmatrix} 20 \\ 20 \\ 30 \end{pmatrix}, \Sigma_2 = I_3.$$

Distribution (D3) does not generate X-outliers, but (D4) generates 10% X-outliers. The random errors ε_i's are again generated by the three error distributions (E1) - (E3) from the previous section. Three sample sizes $n = 20, 30$, and 50 are also used. For each fixed combination of sample size, x-distribution and error distribution, we carried out 1000 simulation runs, resulting in 1000 estimates and 1000 corresponding confidence intervals for each estimator.

The simulation results are reported in Tables 4.15 - 4.22. The biases and standard errors of the regression parameters are presented in Tables 4.15 - 4.18, while Tables

4.19 - 4.22 contain observed coverage levels of the confidence intervals as well as the average lengths of the confidence intervals. Observations on the relative performance of the estimators based on these results are similar to those obtained for the simple linear regression case. In particular, the LS estimator performs the best for error distribution (E1). The MM-estimator is asymptotically biased for asymmetric error distribution. For case without X-outlier, $\hat{\sigma}$ of the RSLSE-II and MM-estimator are underestimated for error distribution (E3).

The relative performance of the associated confidence intervals for individual model parameters θ_i ($i = 0, 1, 2, 3$) are also similar to that observed for the simple linear model case. Since we now have twice as many parameters than before, we would like to also examine the relative performance of confidence intervals for linear combinations of the true parameters. We first consider confidence intervals for the linear combination $\theta_0 + 2\theta_1 + 3\theta_2 + 6\theta_3$. Tables 4.19 - 4.20 contain the 90% and 95% confidence intervals corresponding to the LSE, SLSE, RSLSE-I, RSLSE-II and MM-estimator for this linear combination. We make the following observations based on these results:

1. The LSE based confidence interval performs best in terms of coverage level for distribution (D3), but it has poor coverage levels for distribution (D4).

2. The MM-estimator based interval outperforms the RSLSE-I and RSLSE-II in terms of coverage level for error distribution (E1).

3. For error distribution (E3), the RSLSE-I based interval performs better than the MM-estimator based interval in both coverage level and average interval length. The coverage of the RSLSE-II is low relative to that of RSLSE-I and MM-estimator based intervals. The poor performance of the RSLSE-II and MM-estimator based intervals are likely due to the biases in the underlying es-

timators for θ_0 which affect the proper centering or location of the corresponding confidence intervals.

4. The RSLSE-I interval is better than the RSLSE-II interval in terms of coverage level for all cases.

5. For distribution (D3), the SLS estimator based interval performs better than the RSLSE-II based interval in terms of coverage level. For distribution (D4), the SLS estimator has very low coverage for small sample sizes, while for larger sample size the SLS estimator has higher coverage levels but with very long interval lengths.

Next we consider confidence intervals for linear combination $\theta_1 + \theta_2 + \theta_3$ which does not involve the intercept term θ_0. The reason to exclude the intercept term is that the RSLSE-II and MM-estimator have the most biases for estimating this parameter, and the biases are seen as the primary cause of the poor performance of their corresponding confidence intervals for the above example involving a linear combination containing the intercept. Hence with this new linear combination without the intercept, we anticipate that the RSLSE-II and MM-estimator based intervals will perform better. Tables 4.21 - 4.22 contain 90% and 95% confidence intervals based on all five estimators. Not surprisingly, the RSLSE-II and MM-estimator based intervals now perform better than before. In particular, the performances of RSLSE-I and MM-estimator based intervals are comparable in many cases, whereas that of the RSLSE-II based intervals, while it has improved substantially over the previous example, is still relatively poor when compared to the other two robust intervals. Further, the performance of the LSE and SLSE based intervals have similar performance as before.

Finally, all confidence intervals discussed here are approximate confidence intervals based on the asymptotic normality of the underlying estimators. We have provided the histograms of the estimates (see Figures 4.1 - 4.4) for the simple linear model which all have normal shape, supporting the use of the asymptotic normality for constructing confidence intervals for parameters of that model. For the present multiple linear regression example, Figures 4.5 - 4.8 show the histograms of the RSLSE-I and RSLSE-II for the four regression parameters θ_0, θ_1, θ_2 and θ_3 with distributions (D3, E1) and (D4, E3). Again, they all have typical normal shape. Further, the Q-Q plots for the RSLS estimates (not included here) all follow straight lines. These support the use of the asymptotic normality of the RSLSE-I and RSLSE-II for constructing confidence intervals. Histograms and Q-Q plots of estimates given by other three estimators were also examined and they exhibit similar normal behaviour, supporting the asymptotic normality based confidence intervals.

Table 4.15: Simulation results for the multiple linear regression model for distributions (D3) and (E1): the bias and the standard error.

Sample size and parameter	LSE bias (s.e.)	SLSE bias (s.e.)	RSLSE-I bias (s.e.)	RSLSE-II bias (s.e.)	MM bias (s.e.)
n = 20					
$\hat{\theta}_0$	-0.043	0.013	-0.006	-0.038	-0.032
	(0.973)	(1.226)	(2.117)	(1.987)	(1.089)
$\hat{\theta}_1$	0.025	0.033	-0.018	-0.030	0.012
	(0.260)	(0.353)	(0.375)	(0.385)	(0.297)
$\hat{\theta}_2$	-0.002	-0.021	0.008	0.017	0.004
	(0.131)	(0.159)	(0.244)	(0.225)	(0.139)
$\hat{\theta}_3$	0.001	-0.006	-0.001	0.001	-0.002
	(0.084)	(0.098)	(0.142)	(0.137)	(0.090)
$\hat{\sigma}$	0.009	-0.181	-0.160	-0.253	-0.061
	(0.176)	(0.166)	(0.183)	(0.201)	(0.246)
n = 30					
$\hat{\theta}_0$	-0.021	-0.014	-0.029	-0.025	-0.023
	(0.809)	(1.017)	(0.968)	(0.955)	(0.836)
$\hat{\theta}_1$	0.005	0.007	0.010	0.005	0.002
	(0.231)	(0.292)	(0.282)	(0.280)	(0.238)
$\hat{\theta}_2$	0.002	-0.003	0.001	0.000	0.003
	(0.084)	(0.111)	(0.117)	(0.115)	(0.087)
$\hat{\theta}_3$	0.002	0.003	0.002	0.003	0.002
	(0.070)	(0.090)	(0.083)	(0.085)	(0.072)
$\hat{\sigma}$	-0.008	-0.121	-0.101	-0.150	-0.054
	(0.140)	(0.133)	(0.135)	(0.155)	(0.182)
n = 50					
$\hat{\theta}_0$	0.012	-0.005	0.024	0.026	0.008
	(0.492)	(0.552)	(0.847)	(0.840)	(0.505)
$\hat{\theta}_1$	-0.001	0.002	-0.006	-0.008	-0.002
	(0.152)	(0.172)	(0.197)	(0.201)	(0.158)
$\hat{\theta}_2$	-0.005	-0.003	-0.005	-0.003	-0.004
	(0.078)	(0.090)	(0.101)	(0.104)	(0.080)
$\hat{\theta}_3$	0.001	0.001	0.001	-0.000	0.001
	(0.044)	(0.051)	(0.093)	(0.091)	(0.045)
$\hat{\sigma}$	-0.008	-0.062	-0.067	-0.104	-0.038
	(0.103)	(0.102)	(0.112)	(0.120)	(0.137)

Table 4.16: Simulation results for the multiple linear regression model for distributions (D3) and (E3): the bias and the standard error.

Sample size and parameter	LSE bias (s.e.)	SLSE bias (s.e.)	RSLSE-I bias (s.e.)	RSLSE-II bias (s.e.)	MM bias (s.e.)
n = 20					
$\hat{\theta}_0$	-0.017	0.060	0.062	-0.120	-0.072
	(1.100)	(1.215)	(1.850)	(1.727)	(1.154)
$\hat{\theta}_1$	-0.004	-0.036	-0.035	-0.025	-0.022
	(0.269)	(0.265)	(0.301)	(0.302)	(0.253)
$\hat{\theta}_2$	-0.002	-0.026	-0.018	-0.013	-0.009
	(0.155)	(0.193)	(0.221)	(0.212)	(0.156)
$\hat{\theta}_3$	0.001	0.003	0.011	0.012	0.001
	(0.079)	(0.090)	(0.124)	(0.118)	(0.090)
$\hat{\sigma}$	-0.053	-0.228	-0.182	-0.388	-0.259
	(0.243)	(0.247)	(0.261)	(0.195)	(0.201)
n = 30					
$\hat{\theta}_0$	0.000	-0.050	0.001	-0.093	-0.106
	(0.741)	(0.723)	(0.743)	(0.733)	(0.737)
$\hat{\theta}_1$	0.012	0.015	0.022	0.013	0.015
	(0.227)	(0.203)	(0.215)	(0.215)	(0.220)
$\hat{\theta}_2$	0.000	0.002	0.001	0.000	0.001
	(0.077)	(0.070)	(0.087)	(0.085)	(0.075)
$\hat{\theta}_3$	-0.005	-0.006	-0.011	-0.009	-0.007
	(0.067)	(0.064)	(0.062)	(0.064)	(0.063)
$\hat{\sigma}$	-0.039	-0.133	-0.117	-0.265	-0.193
	(0.186)	(0.185)	(0.194)	(0.165)	(0.164)
n = 50					
$\hat{\theta}_0$	0.006	-0.014	-0.030	-0.114	-0.109
	(0.476)	(0.439)	(0.659)	(0.612)	(0.454)
$\hat{\theta}_1$	-0.008	-0.008	-0.007	-0.007	-0.008
	(0.147)	(0.128)	(0.146)	(0.140)	(0.137)
$\hat{\theta}_2$	0.005	0.004	0.005	0.004	0.004
	(0.078)	(0.068)	(0.078)	(0.075)	(0.073)
$\hat{\theta}_3$	-0.001	-0.001	0.004	0.003	-0.001
	(0.044)	(0.041)	(0.074)	(0.068)	(0.041)
$\hat{\sigma}$	-0.014	-0.066	-0.067	-0.216	-0.175
	(0.147)	(0.147)	(0.158)	(0.126)	(0.123)

Table 4.17: Simulation results for the multiple linear regression model for distributions (D4) and (E1): the bias and the standard error.

Sample size and parameter	RSLSE-I bias (s.e.)	RSLSE-II bias (s.e.)	MM bias (s.e.)
n = 20			
$\hat{\theta}_0$	0.048	0.014	-0.026
	(1.564)	(1.579)	(1.049)
$\hat{\theta}_1$	0.006	0.007	0.013
	(0.305)	(0.318)	(0.271)
$\hat{\theta}_2$	-0.006	-0.003	0.000
	(0.198)	(0.200)	(0.142)
$\hat{\theta}_3$	-0.005	-0.002	0.000
	(0.097)	(0.100)	(0.080)
$\hat{\sigma}$	-0.151	-0.190	0.154
	(0.174)	(0.195)	(0.278)
n = 30			
$\hat{\theta}_0$	0.115	0.115	0.116
	(0.898)	(0.910)	(0.873)
$\hat{\theta}_1$	-0.035	-0.038	-0.037
	0.281	0.288	0.270
$\hat{\theta}_2$	-0.007	-0.008	-0.007
	(0.106)	(0.105)	(0.094)
$\hat{\theta}_3$	-0.003	-0.002	-0.003
	(0.071)	(0.075)	(0.071)
$\hat{\sigma}$	-0.074	-0.103	0.159
	(0.147)	(0.161)	(0.232)
n = 50			
$\hat{\theta}_0$	0.006	0.013	0.023
	(0.859)	(0.872)	(0.622)
$\hat{\theta}_1$	0.001	0.000	-0.004
	(0.202)	(0.200)	(0.163)
$\hat{\theta}_2$	-0.009	-0.008	-0.006
	(0.094)	(0.097)	(0.088)
$\hat{\theta}_3$	0.004	0.003	0.001
	(0.092)	(0.093)	(0.055)
$\hat{\sigma}$	-0.065	-0.082	0.138
	(0.114)	(0.120)	(0.159)

Table 4.18: Simulation results for the multiple linear regression model for distributions (D4) and (E3): the bias and the standard error.

Sample size and parameter	RSLSE-I bias (s.e.)	RSLSE-II bias (s.e.)	MM bias (s.e.)
n = 20			
$\hat{\theta}_0$	-0.054	-0.145	-0.079
	(1.458)	(1.331)	(0.927)
$\hat{\theta}_1$	0.006	0.015	0.005
	(0.297)	(0.281)	(0.244)
$\hat{\theta}_2$	0.004	0.005	-0.001
	(0.189)	(0.175)	(0.133)
$\hat{\theta}_3$	0.005	0.005	0.002
	(0.091)	(0.087)	(0.072)
$\hat{\sigma}$	-0.159	-0.273	0.043
	(0.214)	(0.217)	(0.280)
n = 30			
$\hat{\theta}_0$	0.052	-0.030	-0.021
	(0.879)	(0.868)	(0.829)
$\hat{\theta}_1$	-0.012	-0.014	-0.022
	(0.281)	(0.274)	(0.248)
$\hat{\theta}_2$	0.002	0.001	-0.002
	(0.098)	(0.094)	(0.085)
$\hat{\theta}_3$	-0.006	-0.006	-0.004
	(0.076)	(0.072)	(0.067)
$\hat{\sigma}$	-0.078	-0.230	-0.010
	(0.194)	(0.182)	(0.207)
n = 50			
$\hat{\theta}_0$	0.026	-0.059	-0.066
	(0.885)	(0.807)	(0.589)
$\hat{\theta}_1$	-0.019	-0.014	-0.012
	(0.196)	(0.180)	(0.149)
$\hat{\theta}_2$	0.007	0.004	0.005
	(0.095)	(0.090)	(0.083)
$\hat{\theta}_3$	-0.001	-0.000	-0.003
	(0.097)	(0.086)	(0.051)
$\hat{\sigma}$	-0.064	-0.198	-0.019
	(0.160)	(0.133)	(0.151)

Table 4.19: Simulation results for the multiple linear regression model: the average length and coverage of 90% confidence intervals for $\theta_0 + 2\theta_1 + 3\theta_2 + 6\theta_3$.

Distributions and sample size	LSE coverage (length)	SLSE coverage (length)	RSLSE-I coverage (length)	RSLSE-II coverage (length)	MM coverage (length)
(D3) and (E1)					
n = 20	0.911	0.693	0.709	0.697	0.852
	(0.871)	(0.617)	(1.066)	(1.001)	(0.860)
n = 30	0.918	0.797	0.839	0.793	0.885
	(0.639)	(0.563)	(0.620)	(0.583)	(0.660)
n = 50	0.901	0.847	0.856	0.832	0.888
	(0.481)	(0.460)	(0.521)	(0.496)	(0.493)
(D3) and (E3)					
n = 20	0.931	0.651	0.699	0.645	0.802
	(0.871)	(0.582)	(1.040)	(0.895)	(0.908)
n = 30	0.923	0.787	0.825	0.651	0.799
	(0.639)	(0.540)	(0.599)	(0.500)	(0.659)
n = 50	0.908	0.851	0.856	0.677	0.774
	(0.481)	(0.441)	(0.500)	(0.426)	(0.504)
(D4) and (E1)					
n = 20	0.859	0.000	0.801	0.776	0.860
	(0.906)	(6.105)	(0.878)	(0.844)	(0.839)
n = 30	0.567	0.000	0.867	0.844	0.880
	(0.673)	(16.819)	(0.628)	(0.615)	(0.667)
n = 50	0.778	0.000	0.869	0.856	0.906
	(0.532)	(17.072)	(0.554)	(0.549)	(0.532)
(D4) and (E3)					
n = 20	0.842	0.000	0.779	0.693	0.833
	(0.906)	(6.082)	(0.881)	(0.770)	(0.857)
n = 30	0.516	0.002	0.848	0.730	0.823
	(0.673)	(16.874)	(0.612)	(0.542)	(0.667)
n = 50	0.780	0.000	0.883	0.767	0.839
	(0.532)	(16.867)	(0.554)	(0.492)	(0.544)

Table 4.20: Simulation results for the multiple linear regression model: the average length and coverage of 95% confidence intervals for $\theta_0 + 2\theta_1 + 3\theta_2 + 6\theta_3$.

Distributions and sample size	LSE coverage (length)	SLSE coverage (length)	RSLSE-I coverage (length)	RSLSE-II coverage (length)	MM coverage (length)
(D3) and (E1)					
n = 20	0.968	0.787	0.781	0.760	0.899
	(1.057)	(0.749)	(1.295)	(1.216)	(1.044)
n = 30	0.959	0.858	0.903	0.858	0.942
	(0.771)	(0.678)	(0.747)	(0.703)	(0.796)
n = 50	0.955	0.922	0.919	0.895	0.951
	(0.577)	(0.551)	(0.625)	(0.595)	(0.592)
(D3) and (E3)					
n = 20	0.973	0.698	0.781	0.731	0.864
	(1.057)	(0.707)	(1.263)	(1.087)	(1.102)
n = 30	0.966	0.844	0.889	0.739	0.862
	(0.771)	(0.650)	(0.721)	(0.603)	(0.794)
n = 50	0.953	0.909	0.918	0.762	0.847
	(0.577)	(0.529)	(0.600)	(0.511)	(0.605)
(D4) and (E1)					
n = 20	0.926	0.000	0.874	0.856	0.916
	(1.100)	(7.413)	(1.066)	(1.024)	(1.018)
n = 30	0.696	0.757	0.922	0.907	0.940
	(0.811)	(20.270)	(0.757)	(0.741)	(0.804)
n = 50	0.867	0.803	0.931	0.918	0.948
	(0.638)	(20.491)	(0.665)	(0.659)	(0.638)
(D4) and (E3)					
n = 20	0.922	0.000	0.866	0.785	0.887
	(1.100)	(7.385)	(1.070)	(0.935)	(1.041)
n = 30	0.653	0.743	0.917	0.801	0.886
	(0.811)	(20.336)	(0.738)	(0.653)	(0.804)
n = 50	0.877	0.821	0.942	0.844	0.907
	(0.638)	(20.241)	(0.664)	(0.590)	(0.652)

Table 4.21: Simulation results for the multiple linear regression model: the average length and coverage of 90% confidence intervals for $\theta_1 + \theta_2 + \theta_3$.

Distributions and sample size	LSE coverage (length)	SLSE coverage (length)	RSLSE-I coverage (length)	RSLSE-II coverage (length)	MM coverage (length)
(D3) and (E1)					
n = 20	0.924	0.560	0.676	0.655	0.835
	(1.092)	(0.786)	(1.261)	(1.200)	(1.077)
n = 30	0.908	0.700	0.758	0.739	0.864
	(0.925)	(0.762)	(0.832)	(0.812)	(0.948)
n = 50	0.896	0.797	0.823	0.808	0.877
	(0.556)	(0.496)	(0.660)	(0.643)	(0.558)
(D3) and (E3)					
n = 20	0.919	0.605	0.693	0.694	0.852
	(1.092)	(0.625)	(1.123)	(1.032)	(1.055)
n = 30	0.925	0.812	0.804	0.786	0.875
	(0.925)	(0.648)	(0.692)	(0.655)	(0.860)
n = 50	0.913	0.827	0.837	0.835	0.867
	(0.556)	(0.413)	(0.537)	(0.512)	(0.508)
(D4) and (E1)					
n = 20	0.000	0.000	0.776	0.752	0.835
	(0.167)	(0.781)	(1.302)	(1.250)	(1.096)
n = 30	0.000	0.001	0.840	0.831	0.870
	(0.144)	(0.472)	(1.034)	(1.012)	(1.068)
n = 50	0.000	0.914	0.866	0.849	0.899
	(0.113)	(1.664)	(0.620)	(0.612)	(0.581)
(D4) and (E3)					
n = 20	0.000	0.000	0.802	0.779	0.853
	(0.167)	(0.782)	(1.237)	(1.102)	(1.052)
n = 30	0.000	0.001	0.862	0.834	0.879
	(0.144)	(0.458)	(0.993)	(0.886)	(0.985)
n = 50	0.000	0.935	0.885	0.846	0.873
	(0.113)	(1.656)	(0.626)	(0.554)	(0.542)

Table 4.22: Simulation results for the multiple linear regression model: the average length and coverage of 95% confidence intervals for $\theta_1 + \theta_2 + \theta_3$.

Distributions and sample size	LSE coverage (length)	SLSE coverage (length)	RSLSE-I coverage (length)	RSLSE-II coverage (length)	MM coverage (length)
(D3) and (E1)					
n = 20	0.977	0.653	0.760	0.746	0.902
	(1.326)	(0.954)	(1.531)	(1.457)	(1.308)
n = 30	0.955	0.791	0.834	0.825	0.931
	(1.114)	(0.918)	(1.003)	(0.978)	(1.143)
n = 50	0.955	0.868	0.899	0.884	0.931
	(0.667)	(0.594)	(0.792)	(0.771)	(0.669)
(D3) and (E3)					
n = 20	0.958	0.674	0.793	0.783	0.909
	(1.326)	(0.759)	(1.364)	(1.253)	(1.281)
n = 30	0.964	0.882	0.867	0.859	0.933
	(1.114)	(0.781)	(0.834)	(0.789)	(1.037)
n = 50	0.954	0.891	0.908	0.909	0.934
	(0.667)	(0.496)	(0.644)	(0.614)	(0.609)
(D4) and (E1)					
n = 20	0.000	0.000	0.844	0.819	0.892
	(0.203)	(0.949)	(1.581)	(1.518)	(1.331)
n = 30	0.000	0.007	0.910	0.897	0.928
	(0.173)	(0.569)	(1.246)	(1.219)	(1.287)
n = 50	0.000	0.987	0.923	0.923	0.951
	(0.136)	(1.998)	(0.744)	(0.734)	(0.697)
(D4) and (E3)					
n = 20	0.000	0.000	0.863	0.838	0.910
	(0.203)	(0.949)	(1.502)	(1.338)	(1.278)
n = 30	0.000	0.008	0.922	0.891	0.931
	(0.173)	(0.552)	(1.197)	(1.067)	(1.187)
n = 50	0.000	0.994	0.934	0.913	0.933
	(0.136)	(1.988)	(0.751)	(0.665)	(0.650)

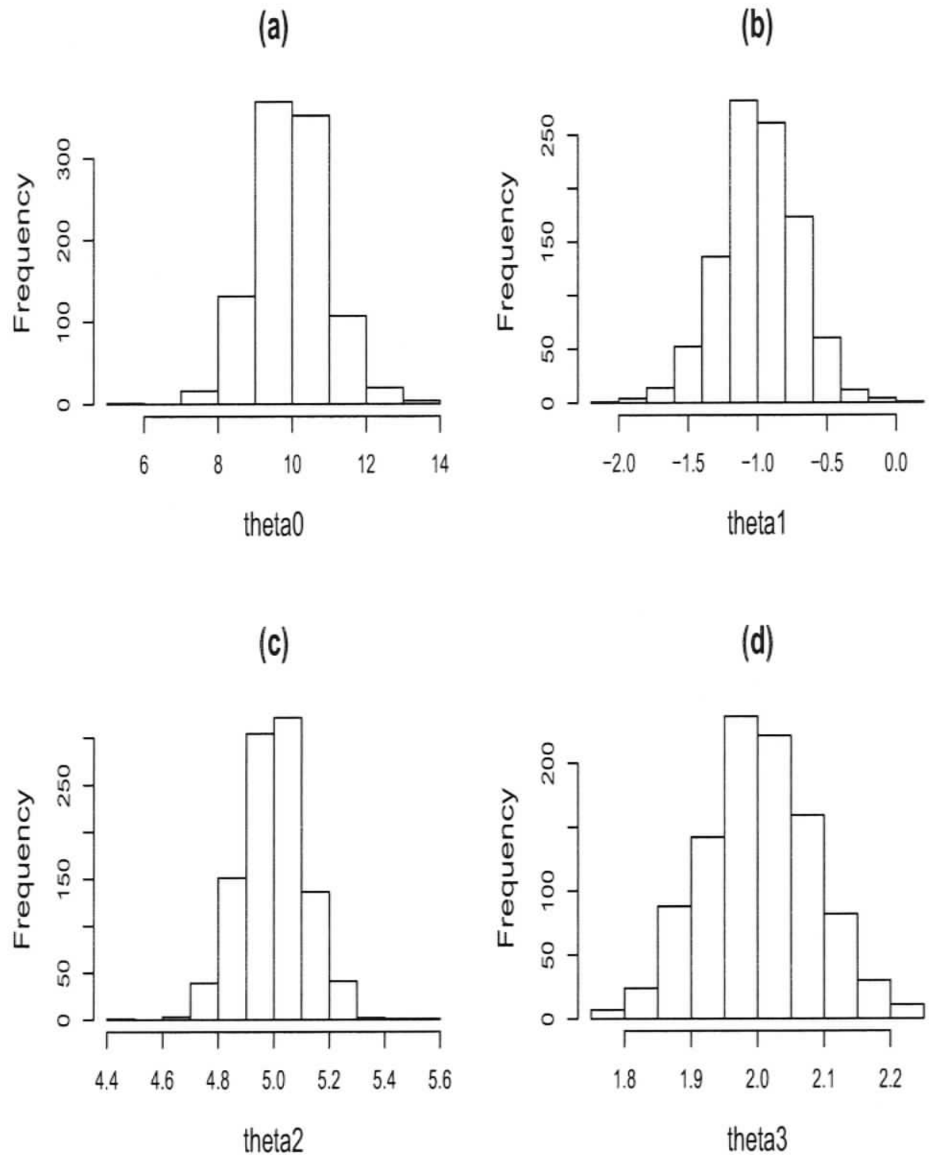


Figure 4.5: Histograms for RSLSE-I for the case with error distribution (E1), no X-outliers and sample size $n = 30$: (a) histogram for $\hat{\theta}_0$, (b) histogram for $\hat{\theta}_1$, (c) histogram for $\hat{\theta}_2$, (d) histogram for $\hat{\theta}_3$.

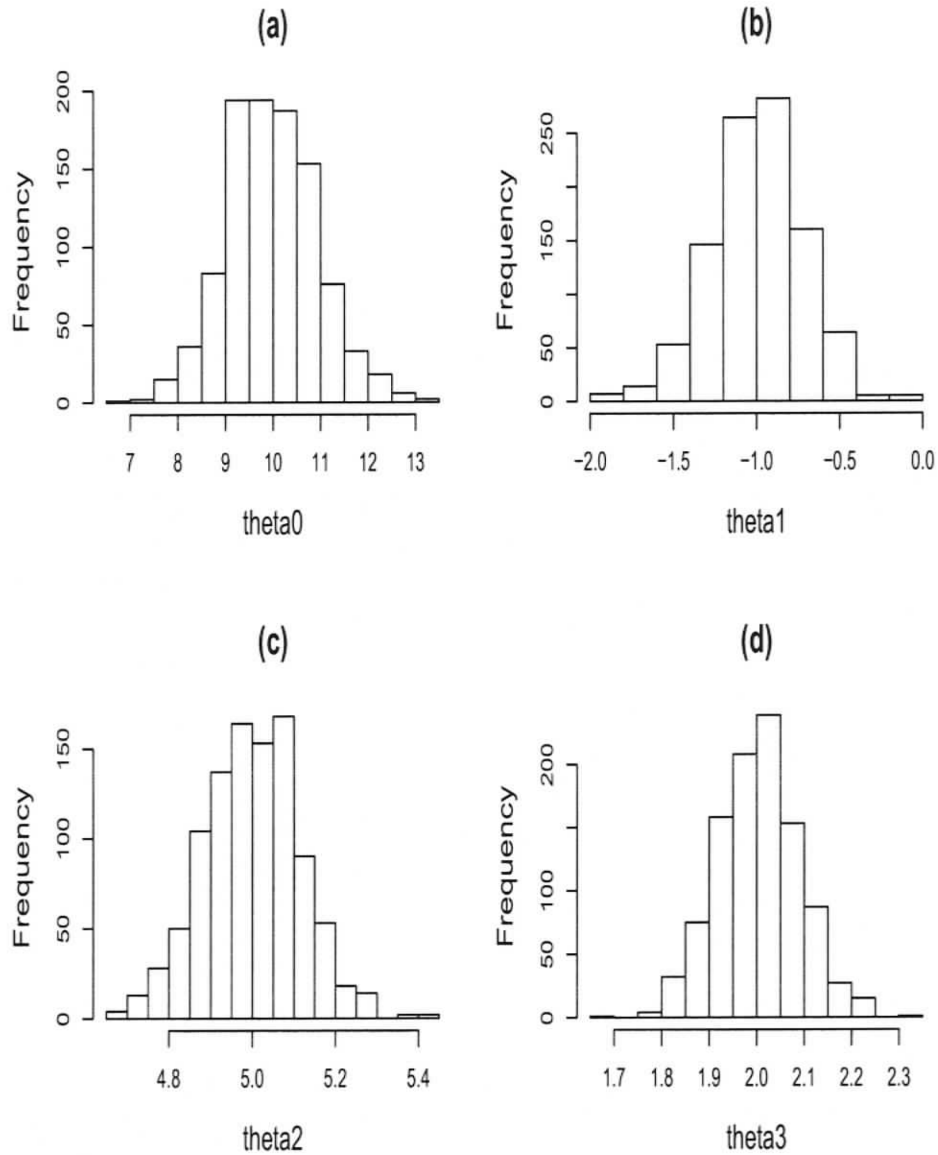


Figure 4.6: Histograms for RSLSE-II for the case with error distribution (E1), no X-outliers and sample size $n = 30$: (a) histogram for $\hat{\theta}_0$, (b) histogram for $\hat{\theta}_1$, (c) histogram for $\hat{\theta}_2$, (d) histogram for $\hat{\theta}_3$.

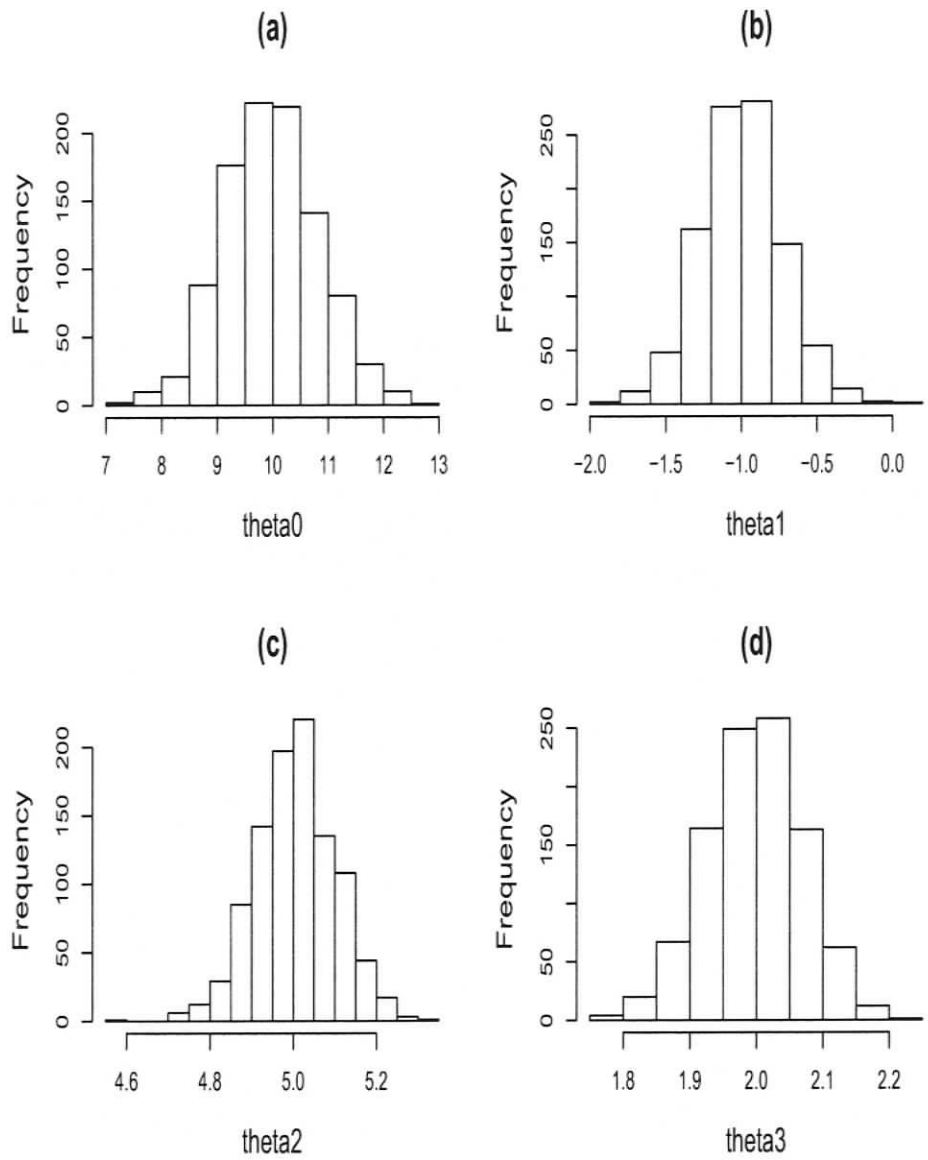


Figure 4.7: Histograms for RSLSE-I for the case with error distribution (E3), 10% X-outliers and sample size $n = 30$: (a) histogram for $\hat{\theta}_0$, (b) histogram for $\hat{\theta}_1$, (c) histogram for $\hat{\theta}_2$, (d) histogram for $\hat{\theta}_3$.

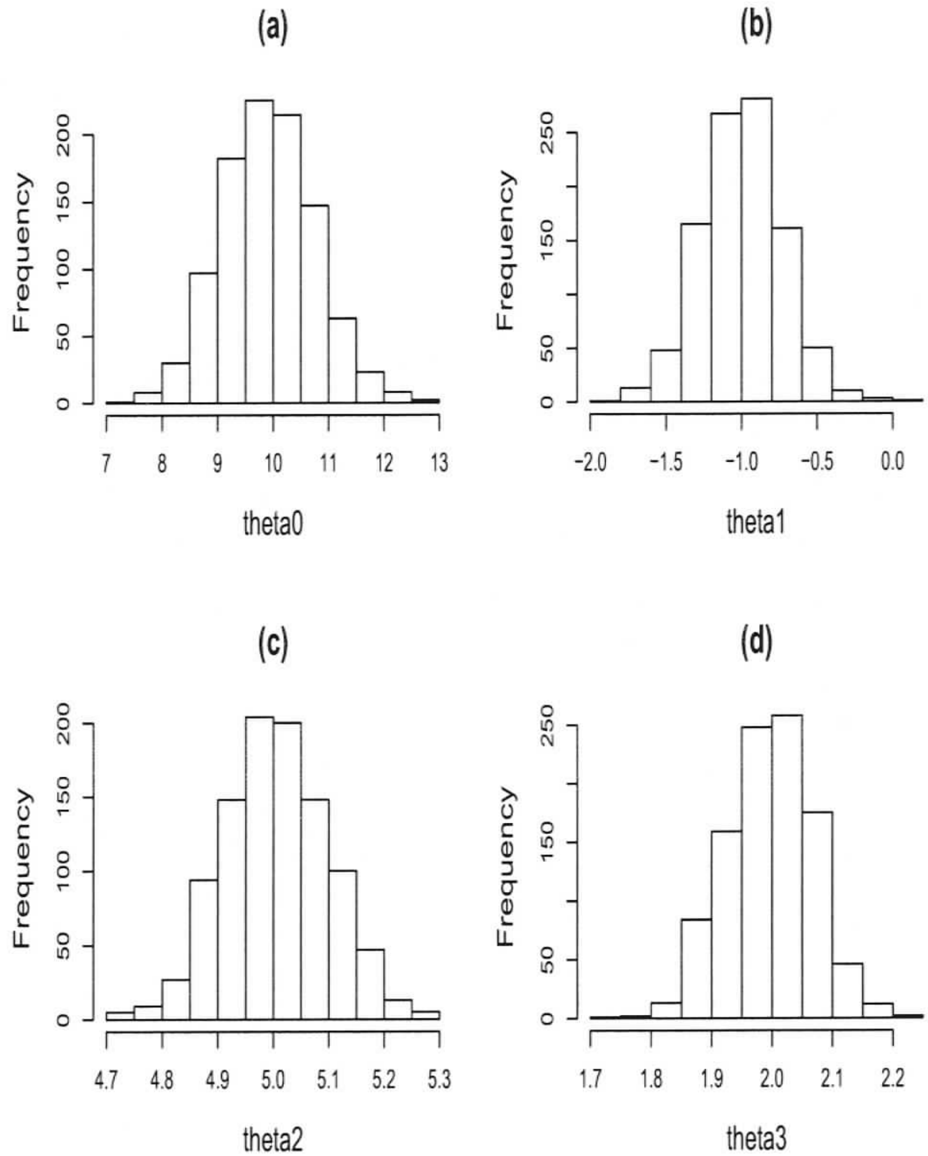


Figure 4.8: Histograms for RSLSE-II for the case with error distribution (E3), 10% X-outliers and sample size $n = 30$: (a) histogram for $\hat{\theta}_0$, (b) histogram for $\hat{\theta}_1$, (c) histogram for $\hat{\theta}_2$, (d) histogram for $\hat{\theta}_3$.

4.3 Applications

In this section, we apply the RSLS estimators to two real data sets and compare them with the LS estimator, SLS estimator and MM-estimator.

Example 4.1. *Consider the stack loss data, a well-known real data set first studied by Brownlee (1965) and subsequently examined by a number of statisticians using several methods. Table 4.23 contains this data set with 21 four-dimensional observations which describe the operation of a plant for the oxidation of ammonia to nitric acid; the rate of operation (x_1), the cooling water inlet temperature (x_2) and the acid concentration (x_3) are the explanatory variables and the stack loss (y) is the dependent variable. Most statisticians who examined this data identified observations 1, 3, 4, and 21 as outliers, and some of them also regard observation 2 as an outlier.*

A multiple linear regression model involving all three explanatory variables

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \varepsilon$$

is fitted to this data set. The estimated parameter values given by the RSLS estimators, the LS estimator, the SLS estimator and the MM-estimator are presented in Table 4.24. The estimated standard errors of the RSLSE-I and RSLSE-II are the two smallest among that of all estimators. The residual plots with lines at 0 and $\pm 2.5\hat{\sigma}$ lines are used to check for outliers in Figure 4.9. Observations whose residuals are outside $\pm 2.5\hat{\sigma}$ lines are outliers. In Figure 4.9(a) and (b), the residual plots for the LS estimator and SLS estimator show that there is no outlier in this data. However, in Figure 4.9(c), (d) and (e) for the RSLSE-I, RSLSE-II and MM-estimator, the residual plots clearly indicate there are outliers in the stack loss data. The RSLSE-I identifies 4 outliers, which are observations 1, 3, 4 and 21. The RSLSE-II identifies 5 outliers, which include the previous four observations plus observation 2. The MM-estimator identifies only 2 outliers. Since most statisticians reported the observations 1, 3, 4 and 21 as outliers and some also reported observation 2 as an outlier, from the standpoint of identifying outliers, both the RSLSE-I and RSLSE-II worked very well for this example. The MM-estimator, on the other hand, identified only two very extreme outliers and is seen as less effective in identifying the outliers. This indicates that the MM-estimator may have been influenced by the mild outliers to a larger degree than the RSLSE-I and RSLSE-II.

Table 4.23: Stack loss data for Example 4.1

Index (i)	Stackloss (y)	Rate (x_1)	Temperature (x_2)	Acid Concentration (x_3)
1	42	80	27	89
2	37	80	27	88
3	37	75	25	90
4	28	62	24	87
5	18	62	22	87
6	18	62	23	87
7	19	62	24	93
8	20	62	24	93
9	15	58	23	87
10	14	58	18	80
11	14	58	18	89
12	13	58	17	88
13	11	58	18	82
14	12	58	19	93
15	8	50	18	89
16	7	50	18	86
17	8	50	19	72
18	8	50	19	79
19	9	50	20	80
20	15	56	20	82
21	15	70	20	91

Table 4.24: Estimated linear model parameters with standard errors (in brackets) for stack loss data in Example 4.1

	LSE	SLSE	RSLSE-I	RSLSE-II	MM-estimator
$\hat{\theta}_0$	-39.9197	-38.1501	-36.9291	-35.8267	-41.5246
	(11.8960)	(8.5158)	(3.2811)	(2.8111)	(5.2978)
$\hat{\theta}_1$	0.7156	0.5485	0.7018	0.6682	0.9390
	(0.1349)	(0.1982)	(0.0869)	(0.0716)	(0.1174)
$\hat{\theta}_2$	1.2953	1.7287	0.7058	0.5917	0.5796
	(0.3680)	(0.5333)	(0.1873)	(0.1374)	(0.2630)
$\hat{\theta}_3$	-0.1521	-0.1700	-0.0376	-0.0073	-0.1129
	(0.1563)	(0.1122)	(0.0613)	(0.0533)	(0.0699)
$\hat{\sigma}$	3.2430	2.4678	2.1734	0.9223	1.9120
	(2.9902)	(2.6626)	(3.3144)	(0.3471)	(*)

Example 4.2. *(Montgomery and Peck, 2006, P70). In order to predict the amount of time (delivery time y) required by the route driver to service the vending machines, a soft drink bottler collected a data set of the delivery time and related variables affecting the delivery time. Among such variables affecting the delivery time, the two most important are the number of cases of the product stocked (x_1) and the distance walked by the route driver (x_2). Table 4.25 shows this data set which consists of 25 sets of observations of these three variables.*

To guard against possible outliers, we fit the following multiple regression model

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \varepsilon$$

to this data set using different methods. Table 4.26 contains the estimated parameter values given by the LS estimator, SLS estimator, RSLSE-I, RSLSE-II and MM-estimator. The estimates for the RSLSE-I and RSLSE-II are exactly same in this example. This is due to the absence of Y-outliers in the data since in such cases the weight matrices of RSLSE-I and RSLSE-II are the same and hence they produce the same estimates. Further, the estimates given by the RSLSE-I and RSLSE-II are

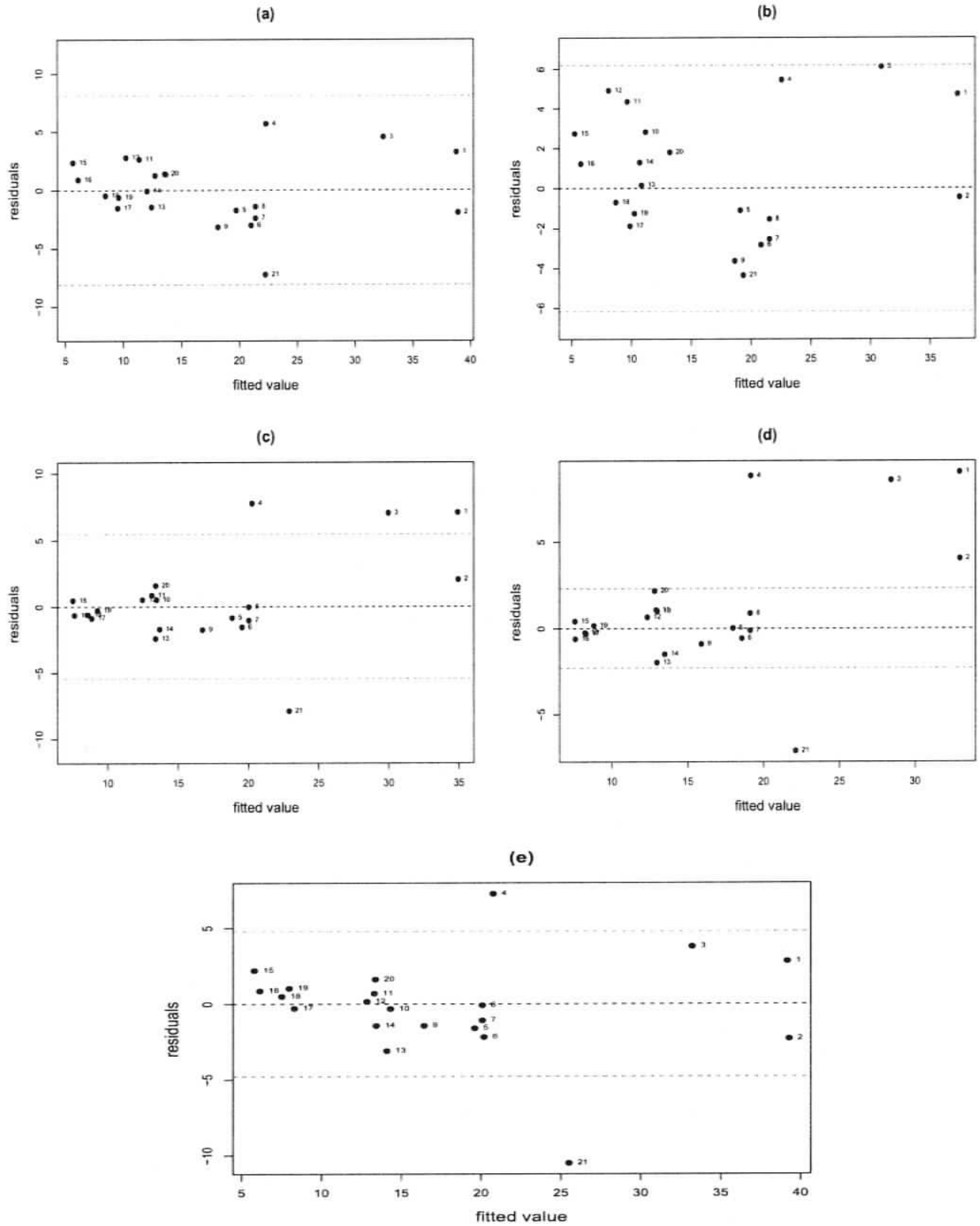


Figure 4.9: Plots of residuals verses fitted values with lines at 0 and $\pm 2.5\hat{\sigma}$ for stack loss data in Example 4.1: (a) residuals verses fitted values for LS estimator, (b) residuals verses fitted values for SLS estimator, (c) residuals verses fitted values for RSLSE-I, (d) residuals verses fitted values for RSLSE-II, (e) residuals verses fitted values for MM-estimator.

Table 4.25: Delivery data for Example 4.2

Index (i)	Delivery Time (y)	Number of Cases (x_1)	Distance (x_2)
1	16.68	7	560
2	11.50	3	220
3	12.03	3	340
4	14.88	4	80
5	13.75	6	150
6	18.11	7	330
7	8.00	2	110
8	17.83	7	210
9	79.24	30	1460
10	21.50	5	605
11	40.33	16	688
12	21.00	10	215
13	13.50	4	255
14	19.75	6	462
15	24.00	9	448
16	29.00	10	776
17	15.35	6	200
18	19.00	7	132
19	9.50	3	36
20	35.10	17	770
21	17.90	10	140
22	52.32	26	810
23	18.75	9	450
24	19.83	8	635
25	10.75	4	150

quite close to that of the MM-estimator, while the estimates given by LS estimator and SLS estimator are quite different from that of the three robust estimators. The residual plots are presented in Figure 4.10. In Figure 4.10(a) and (b), residual plots for the LS estimator and SLS estimator do not identify any outliers. The RSLSE-I and RSLSE-II plots in Figure 4.10(c) and (d) find two outliers, observation 9 and 11, whereas the MM-estimator plot in Figure 4.10(e) identifies only one very extreme outlier, observation 9. Overall, the three robust estimators are more consistent and are seen to be less affected by the outliers. But the two non-robust estimators, LS and SLS, are rather seriously affected by the outliers as their estimated values differ substantially from that of the robust estimates. This is also reflected by the large estimated σ values given by these non-robust estimators as the influence of the outliers are not properly discounted but rather accommodated through an inflated error variance.

Table 4.26: Estimated linear model parameters with standard errors (in brackets) for delivery data in Example 4.2

	LSE	SLSE	RSLSE-I	RSLSE-II	MM-estimator
$\hat{\theta}_0$	2.3412	2.7404	5.5480	5.5480	4.4718
	(1.0967)	(0.8418)	(0.9856)	(0.9856)	(0.6979)
$\hat{\theta}_1$	1.6159	1.6811	1.2821	1.2821	1.4718
	(0.1707)	(0.1940)	(0.1710)	(0.1710)	(0.1401)
$\hat{\theta}_2$	0.0144	0.0157	0.0106	0.0106	0.0108
	(0.0036)	(0.0042)	(0.0030)	(0.0030)	(0.0044)
$\hat{\sigma}$	3.2590	3.7412	2.0343	2.0343	2.0230
	(3.1207)	(4.3432)	(0.9541)	(0.9541)	(*)

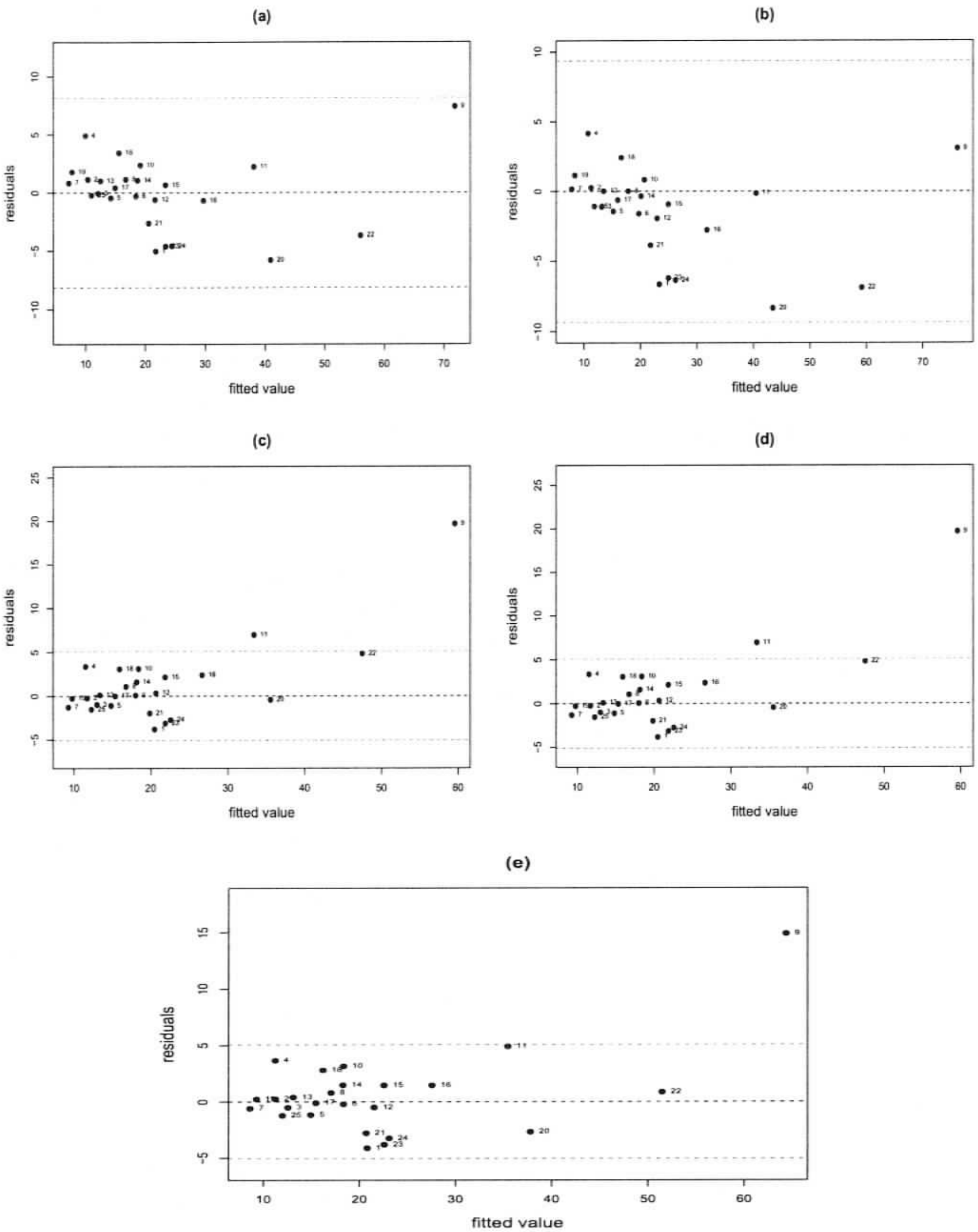


Figure 4.10: Plots of residuals verses fitted values with lines at 0 and $\pm 2.5\hat{\sigma}$ for delivery data: (a) residuals verses fitted values for LS estimator, (b) residuals verses fitted values for SLS estimator, (c) residuals verses fitted values for RSLSE-I, (d) residuals verses fitted values for RSLSE-II, (e) residuals verses fitted values for MM-estimator.

4.4 Concluding remarks

In this chapter, we investigate the performance of the RSLS estimators through simulation studies and compared them with the LS estimator, SLS estimator and MM-estimator. We applied the RSLS estimators to two real-life examples involving multiple linear regression models. Both the RSLSE-I and RSLSE-II work very well and are more effective at detecting and discounting the influences of outliers than the MM-estimator in these two examples.

Through the simulation studies, we have demonstrated that both the RSLSE-I and RSLSE-II are valuable robust methods for linear regression models. The RSLSE-I, in particular, is a competitive robust estimator for linear regression models to the highly efficient and highly robust MM-estimator. The advantage of the RSLSE-I is that it is asymptotically unbiased for both symmetric and asymmetric error distributions. Because of this, the RSLSE-I in general outperforms the MM-estimators when the error distributions are asymmetric. Also, when the error distributions are symmetric, it lags the MM-estimator only slightly in performance. Hence for real applications where the true nature of the error distribution is unknown, the RSLSE-I is preferred. The RSLSE-II is also asymptotically unbiased when the error distributions are symmetric but is biased if the error distributions are not symmetric. The efficiency and coverage level of the RSLSE-II are lower than that of the RSLSE-I, but it is still comparable to the RSLSE-I and MM-estimator in many cases. One advantage of the RSLSE-II that we did not elaborate previously is that the RSLSE-II is robust against both X-outliers and Y-outliers.

Chapter 5

Conclusion

In this thesis, we proposed two robust methods to robustify the promising SLS estimator for linear regression models. We investigated their properties through theoretical and numerical approaches. These include the asymptotic breakdown point, influence function and asymptotic distributions of the RSLS estimators. Through simulation studies, we also examined the performance of the RSLS estimators for finite sample sizes. We also applied the proposed RSLS estimators to several real data examples.

The proposed RSLS estimators have several advantages. The RSLSE-I is robust against X-outliers and the RSLSE-II is robust against both X-outliers and Y-outliers. Furthermore, the RSLS estimators have high breakdown point if the initial estimator has high breakdown point. The simulation studies indicate that the RSLSE-I is asymptotically unbiased in all situations and it is highly efficient. On the other hand, the MM-estimator is biased for estimating the intercept of a linear model with an asymmetric error distribution. Both the theory and the simulation studies show that the RSLS estimators are asymptotically normally distributed. The numerical results also show that the RSLS estimators are comparable to the MM-estimator in many cases. The real data analysis indicate both the RSLSE-I and RSLSE-II provide good outlier detection and parameter estimation.

The real data examples that we have studied also indicate that outliers often exist in real-life applications and the non-robust estimators cannot identify outliers, especially when there are more than one outliers. Therefore, we recommend that both non-robust estimators and robust estimators be used for data analysis. If their estimates are similar, then we can use the non-robust estimators for inference. Otherwise, we should use the robust estimators to do the analysis as when there are likely outliers in the data and the estimates given by the non-robust estimators are not reliable. For linear regressions, our RSLS estimators are effective robust estimators and should be used together with the SLS estimators whenever the latter is applied.

On future extensions and improvements for our proposed RSLS estimators, we first note that in some situations the coverage levels of the corresponding confidence intervals may be lower than the nominal level. This is likely due to an underestimation of the error variance σ^2 which will lead to narrower confidence intervals and hence reduce the coverage level. To resolve this problem, further research is needed to study the construction of the weight matrices. The ideal weight matrices are those which will give rise to robust estimators that can discount the influences of the outliers but not overly sensitive as to discount the influence of good data points. The latter is typically the cause for the underestimation of the error variance. In this thesis, we have focused on the RSLS estimators for linear regression models but the underlying SLS method can be applied to other types of regression models. Therefore it is of interest to extend the RSLS estimators to other regression models, such as logistic models and non-linear regression models where the SLS method has been applied. The basic idea that we have used to construct RSLS for linear models may be applied to robustify SLS estimator for other models; that is, we may seek to construct robust SLS estimators through careful selection of the weight matrix $W_R(X)$. However, in order to construct a proper weight matrix $W_R(X)$, we need to find a good $V(X)$ function in (3.10) and this is more challenging for more complicated regression models. Research is ongoing to find such $V(X)$ functions and to construct the proper weight matrices for more complicated regression models.

References

Abarin, T. and Wang, L. (2006), Comparison of GMM with second-order least squares estimator in nonlinear models, *Far East Journal of Theoretical Statistics*, **20**, 179-196.

Azzalini, A. and Capitania, A. (2003), Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t-Distribution, *Journal of the Royal Statistical Society*, **65**, 367-389.

Bain, L.J. and Engelhardt, M. (1992), *Introduction to Probability and Mathematical Statistics*, Second Edition, Duxbury, Pacific Grove.

Beaton, A.E. and Tukey, J.W. (1974), The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data, *Technometrics*, **16**, 147-185.

Brownlee, K.A. (1965), *Statistical Theory and Methodology in Science and Engineering*, Second Edition, Wiley, New York.

Daniel, C. and Wood, F.S. (1971), *Fitting Equations to Data*, Wiley, New York.

Donoho, D.L. and Huber, P.J. (1983), *A Festschrift for Erich L. Lehmann*, Wadsworth, Belmont, 157-184.

Ezekiel, M. and Fox, K.A. (1959), *Methods of Correlation and Regression Analysis*, Wiley, New York.

Hampel, F.R. (1974), Then influence curve and its role in robust estimation, *The Annals of Statistics*, **69**, 383-393.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986), *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York.

Huber, P.J. (1973), Robust regression: asymptotics, conjectures and Monte Carlo, *The Annals of Statistics*, **1**, 799-821.

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

Leblanc, A. and Wang, L. (2008), Second-order nonlinear least squares estimation, *Annals of the Institute of Statistical Mathematics*, **4**, 883-900.

Maronna, R. A., Martin, D. and Yohai V.J. (2006), *Robust Statistics Theory and Methods.* Wiley, Chichester.

Montgomery, D.C., Peck, E.A. and Vining G.G. (2006), *Introduction to Linear Regression Analysis.* Fourth Edition, Wiley, Hoboken.

Rao, C.R., Toutenburg, H., Fieger, A., Heumann, C., Nittner, T. and Scheid, S. (1999), *Linear Models: Least Squares and Alternatives*, Springer, New York.

Rousseeuw, P.J. (1984), Least median of squares regression, *Journal of the American Statistical Association*, **79**, 871-880.

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

Rousseeuw, P.J. and Yohai, V.J. (1984), Robust regression by mean of S-estimators, *Robust and Nonlinear Time Series*, **26**, 256-272.

Stigler, S.M. (1977), Do robust estimators deal with real data? *The Annals of Statistics*, **5**, 1055-1098.

Wang, L. (2003), Estimation of nonlinear berkson-type measurement error models, *Statistica Sinica*, **13**, 1201-1210.

Yale, C. and Forsythe, A.B. (1976), Winsorized regression, *Technometrics*, **18**, 291-300.

Yohai, V.J. (1987), High breakdown-point and high efficiency estimates for regression, *The Annals of Statistics*, **15**, 642-656.