University of Alberta

ON MINIMUM DISTANCE ESTIMATION IN DOSE RESPONSE STUDIES

by

Bangxin Zhao

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

in

Statistics

Department of Mathematical and Statistical Sciences

Fall 2013

Edmonton, Alberta

Abstract

In this thesis, we examine two robust and efficient methods of estimation in dose-response studies context. In particular, we investigate the minimum Hellinger distance estimation and symmetric chi-squared distance methods of estimation. Using these approaches, we obtain estimators which have desirable robustness properties as well as good asymptotic efficiency properties. We support our theoretical results with extensive finite sample simulation studies.

For quantal assay problems, logit and probit analysis are used to analyze binary responses. Based on the minimum Hellinger distance and symmetric chi-squared distance approaches, new estimators of the regression parameters are derived for logistic and probit models. Then their asymptotic properties such as consistency and asymptotic normality are investigated. It is shown that our minimum Hellinger distance estimator is asymptotically equivalent to the traditional estimators derived using the maximum likelihood and weighted least squares approaches. Simulation studies are used to demonstrate that the new estimators work as good as the traditional estimators and most often outperforms them when a contamination occurs in the data.

Further, the proposed methods are used to estimate the critical dose, and the corresponding estimators are again compared with the the maximum likelihood and weighted least squares estimators. This is done only numerically. The final numerical estimates are obtained by performing optimization of the mean value of 1000 replications. The proposed estimators are comparable to the benchmark methods and show good robust properties. A real data set is analyzed as case study to illustrate the performance of the estimators.

Acknowledgements

I would like to express my deepest gratitude to Dr. Rohana Karunamuni, my supervisor, for his support and guidance, especially for his motivation and encouragement. Without his consistent and illuminating instruction, I would not be where I am today. I not only learned new statistical methods from him, but also learned how to be a better person. His enthusiasm to help his students will always be a good example and guide me in my future life.

I would also like to express my sincere thanks to examiners of my oral defense committee, Dr. Linglong Kong and Dr. Yan Yuan, for their time and insightful questions.

Also I would like to thank all the faculty members, staffs and students in our department. Their kindness and help during my two wonderful years at the University of Alberta is appreciated.

Lastly, my thanks would go to my for beloved family for their loving considerations and great confidence in me all through these years.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Parameter Estimation in Dose-Response Studies

1.1.1 Background

In order to obtain a preliminary efficacy and toxicity of a testing drug, wide-ranging quantities of doses are used in the pre-clinical studies. The crucial step is to find the critical dose. Critical dose is a random variable which determines the minimum amount of drug needed to show a response (e.g. cure in an efficacy experiment or death in a dosage-mortality study). Due to variation between individuals in the population, the critical dose is a random variable and the statistical problem concerns with the estimation of the parameters of its distribution.

In the usual dose-response experiments, study subjects are randomized to several subgroups. The outcome of interest is usually measured at several increasing dose levels, denoted as x_j ($j = 1, 2, \ldots, K$, i.e. K different increasing

dosages). In each subgroup, the number of individuals who show a response is observed. Then the ratio $\pi_j = \frac{m_j}{n_j}$ is an estimate and a sufficient statistic of $P(Critical\,Dose \leq x_j)$, where m_j subjects show responses out of n_j. Assume that

$$P(Critical\,Dose \leq x_j) = P(Response\,at\,Dose\,x_j) = F(\alpha + \beta x_j),$$

where F is a known distribution function and α, β are unknown parameters.

One of the most important quantity of dose is the 'median effective dose' (ED_{50}), which is the dose that produces a response in half of the population that takes it. The most common choices of F are the cumulative distribution functions (CDF) of logistic and normal distributions. For instance, the logistic CDF is $F(x) = (1 + e^{-\frac{x-\mu}{\sigma}})^{-1}$, where μ is the location parameter and σ is the scale parameter. The standardized form of the logistic CDF has $\mu = 0$ and $\sigma = 1$ and, as $x \to \infty$, $F(x) \uparrow 1$ when $\sigma > 0$ (we do not consider the case where $\sigma < 0$). The model used here is not formulated in terms of the usual location and scale parameter; the parameters α and β have another interpretation as follows.

The dose-response curve is S-shaped and logistic regression model is one of the most important formulae used to fit this S-shaped pattern:

$$\pi(y) = \frac{e^y}{1 + e^y} \tag{1.1}$$

We solve for y and obtain $y = \log \frac{\pi(y)}{1-\pi(y)}$, and we call it $logit[\pi(y)]$. Since π_j is considered as a sufficient statistic for $F(\alpha + \beta x_j)$, we have $y_j = F^{-1}(\pi_j) = \log \frac{\pi_j}{1-\pi_j} = \alpha + \beta x_j$. The plot of $y_j = F^{-1}(\pi_j)$ against x_j, $j = 1, 2, \ldots, K$,

should be approximately a straight line, if our chosen model is appropriate.

1.1.2 Maximum Likelihood Estimation

Maximum likelihood estimation (MLE) is by far the most popular point estimation method employed in statistics. If model is correct and the observations are not contaminated, then parameters can be estimated using the MLE method effciently. MLEs have many nice properties such as consistency and asymptotic efficiency in most cases. However, they can be highly unstable if the model is not totally correct, and they are not robust if the data is slightly contaminated.

An example of MLE of one-dimensional location parameter is as follows. Let X be a random variable (r.v.) from a probability density function $f(x)$, with location parameter θ. For sample of n independent observations $\{x_1, \ldots, x_n\}$ with the same distribution as X, the likelihood function is the joint probability distribution or density function $f(\mathbf{x}; \theta)$, viewed as a function of the parameter θ given the sample. Then the likelihood function is given by $l(\theta; \mathbf{x}) = \Pi_{i=1}^{n} f(x_i; \theta)$. The estimated parameter in the maximum likelihood sense, say $\hat{\theta}$, is the parameter value that maximizes the likelihood function, and it can be obtained by solving the equation $\frac{\partial l}{\partial \theta} \mid_{\theta=\hat{\theta}} = 0$.

For easier calculation, $\ln l$, which is called the log-likelihood, is commonly used instead of l if any exponential form is shown in the likelihood function. Therefore, the MLE is defined as: $\hat{\theta} = \underset{\theta}{arg\,min} \sum_{i=1}^{n} \rho(x_i; \theta)$, where $\rho = -\ln f(x_i; \theta)$. Our objective is then to minimize $L(\theta) = \sum_{i=1}^{n} \rho(x_i; \theta)$. To solve the optimization problem, the derivative of the objective function was set equal to zero and solve for θ: $\frac{\partial L}{\partial \theta} \mid_{\theta=\hat{\theta}} = 0$. If ρ is an arbitrary function, an implicit

equation $\sum_{i=1}^{n}\psi(x_i;\theta) = 0$ with $\psi(x_i;\theta) = \frac{\partial\rho(x_i;\theta)}{\partial\theta}$ is called an M-estimation equation (M stands for 'maximum likelihood type'). The resulting estimator is called an M-estimator. See Huber (2009) for more details.

In our dose-response problem, let $\boldsymbol{\theta_0} = (\alpha_0, \beta_0)^T$ be the true parameter value, and $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta})^T$ be the MLE of $\boldsymbol{\theta_0} = (\alpha_0, \beta_0)^T$, then

$$(\hat{\alpha}, \hat{\beta})^T = \underset{\alpha,\beta}{arg\,max}\{\Pi_{j=1}^{K}[F(\alpha+\beta x_j)]^{m_j}[1-F(\alpha+\beta x_j)]^{n_j-m_j}\}. \quad (1.2)$$

We typically use an iterative procedure (e.g. Fisher Scoring or Newton-Raphson algorithm) to find the maximum likelihood estimate. If we assume that the following linear logistic model holds

$$logit[\pi_j] = \alpha_0 + \beta_0 x_j$$

where $\pi_j = \text{Prob}(Y_j = 1)$ with $Y_j = 1$ means a positive response is showing, and the F to be the CDF of logistic distribution $F(\alpha_0 + \beta_0 x_j) = \frac{e^{\alpha_0+\beta_0 x_j}}{1+e^{\alpha_0+\beta_0 x_j}}$, then the variance of the predicted probability is given by

$$Var(logit[\hat{\pi_j}]) = Var(\hat{\alpha}) + x_j^2 Var(\hat{\beta}) + x_j Cov(\hat{\alpha}, \hat{\beta}).$$

For maximum likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$, $Var(\hat{\alpha}, \hat{\beta})$ is obtained by inverting the information matrix. Specifically,

$$Var(\hat{\alpha}, \hat{\beta}) = E\begin{bmatrix} \sum_{j=1}^{K}(W_j) & \sum_{j=1}^{K}(W_j x_j) \\ \sum_{j=1}^{K}(W_j x_j) & \sum_{j=1}^{K}(W_j x_j^2) \end{bmatrix}^{-1},$$

where

$$W_j = \frac{n_j exp\{\alpha_0+\beta_0 x_j\}}{(1+exp\{\alpha_0+\beta_0 x_j\})^2}.$$

In general, for a continuous distribution with CDF F and probability density function (PDF) f, $W_j \propto \frac{f^2(\alpha_0+\beta_0 x_j)}{F(\alpha_0+\beta_0 x_j)(1-F(\alpha_0+\beta_0 x_j))}$.

1.1.3 Weighted Least Square Estimation

Suppose $y_j = F^{-1}(\pi_j)$ against x_j, $j = 1, 2, \ldots, K$, is plotted, and if the postulated model is appropriate then the plotted points should approximately lie in a straight line. But due to multiple reasons, such as model contamination or sampling fluctuations in the data, the plotted points usually cannot strictly follow a straight line. The parameters α and β are the y-intercept and the slope of this straight line, respectively. Therefore, the problem can be considered as one of simple linear regression under the assumption that the error variance is non-homogeneous. For known heteroscedasticity, weighted least squares (WLS) is a popular method which is used to obtain efficient unbiased estimates.

In this dose-response problem, consider minimizing the weighting of the individual measurements in the least squares cost function: $V = \sum_{j=1}^{K} n_j w_j (y_j - \alpha - \beta x_j)^2$ with respect to α and β to obtain $\hat{\alpha}$ and $\hat{\beta}$. In matrix form:

$V = (Y - Z\boldsymbol{\theta})^T W (Y - Z\boldsymbol{\theta})$, where $Y = (y_1, y_2, \ldots y_K)^T$, $\boldsymbol{\theta} = (\alpha, \beta)^T$, $\mathbf{Z} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_K \end{pmatrix}$, $\mathbf{W} = \begin{pmatrix} w_1 & & \\ & \ddots & \\ & & w_K \end{pmatrix}$.

Let $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta})^T$ be the WLSE of (α_0, β_0). This involves solving the equation $\frac{\partial}{\partial \theta}[(Y - Z\boldsymbol{\theta})^T W (Y - Z\boldsymbol{\theta})] = 0$ and it results in $\hat{\boldsymbol{\theta}} = (Z^T W Z)^{-1} Z^T W Y$. For calculation in the general form, let f be the density function of F, $w_j =$

$f^2(y_j)/[\pi_j(1-\pi_j)]$ and the minimization gives the following estimates:

$$\hat{\boldsymbol{\theta}} = \Gamma_N^{-1} \sum_{j=1}^{K} n_j w_j y_j \mathbf{Z}_j, \tag{1.3}$$

where $\mathbf{Z}_j = (1, x_j)^T$, $N = \sum_{j=1}^{K} n_j$ and $\Gamma_N = \sum_{j=1}^{K} n_j w_j \mathbf{Z}_j \mathbf{Z}_j^T$.

By examining (1.3) we note that the weighted least squares estimator is not robust and could be greatly affected by many types of errors. Some common errors are:

- errors in the measurement or recording of the x_j values,
- errors in the π_j values as caused, for example, by subjects showing response (e.g. dying) from other causes,
- errors caused by choosing the wrong distribution function,
- errors caused when the inverse function (or matrix) is not exist.

The method of iteratived reweighted least squares (IRLS) is an adjustment of WLSE. One can use this algorithm to minimize $\sum_{j=1}^{K} \log(1+(y_j-\alpha-\beta x_j)^2)$. But as shown in Section 2.3.2 later, the results are not very good when compared with other methods.

1.2 Motivation and Organization of The Thesis

Statisticians stress the importance of robust procedure in statistical inference over the years. Hampel (1968, 1973) and Huber (1972, 1973) are considered

as the landmark papers for finding robust statistics. Although the methods they proposed are good in dealing with outliers, they are easily suffer from a loss of efficiency if the assumed model distribution is actually the real one. Two fundamental ideas in parametric estimation are efficiency and robustness, but there are contradictions between the aims of achieving both, i.e. a robust estimator is usually not efficient and vice versa. Hampel (1968) introduced the influence curve to distinguish these two kinds of estimators. In general, the influence curve of an efficient estimator will show unboundedness, while a robust one will be always bounded. In many statistical inference areas, minimum distance approaches yield statistics that are efficient under the case when the postulated model is true and robust to deviations under the contaminated model. A popular minimum distance approach is the minimum Hellinger distance (MHD) approach introduced by Beran (1977). He also proposed α-influence curve to determine the robustness of an estimator. Various other distances such as chi-squared distance, symmetric chi-squared distance, total-variation distance, etc. have been used in the literature, see Lindsay (2004) for more discussions on these distances and their applications.

This thesis is organized as follows. In Chapter 2, a version of the minimum Hellinger distance estimation (MHDE) method is employed to estimate the regression parameters (α_0, β_0) by comparing an estimate with the postulated parametric distribution. Then the asymptotic properties such as consistency and asymptotic normality of the MHD estimators are studied. Robust properties of the estimators are investigated using a Monte Carlo study.

In Chapter 3, the parameters (α_0, β_0) are estimated using the symmetric chi-squared distance introduced by Lindsay (2004). In particular, in Section 3.1 we propose estimating α_0 and β_0 by the symmetric chi-squared distance

estimation (SCDE) method. In Section 3.2 we give results on the existence, uniqueness, consistency and asymptotic distribution of the SCDE. In Section 3.3 we use some Monte Carlo studies to examine the robustness of the SCDE in comparison with the traditional estimators.

In Chapter 4, the relationship between the predictive dose level and contamination rate is numerically compared using four methods: MLE, WLSE, MHDE and SCDE. Chapter 5 contains some closing remarks. It contains the summary and conclusions on the performance of the methods employed. We also provide some directions for further studies.

Chapter 2

Minimum Hellinger Distance Estimation Method

2.1 Background

The minimum Hellinger distance (MHD) approach was proposed by Beran (1977) for independent and identically distributed (iid) continuous random variables in parametric models. MHD estimators have been shown to have excellent robustness properties in parametric models such as the resistance to outliers and robustness with respect to model misspecification (Beran, 1977; Donoho & Liu, 1988). Since the original work of Beran, MHD estimators have been developed in the literature for various setups and models including discrete random variables, some parametric mixture models, semiparametric models, etc. The literature is too extensive to state a complete listing here. For recent developments in the area and some important references can be seen in the recent articles of Wu and Karunamuni (2009, 2012), Karunamuni and Wu (2011) and Tang and Karunamuni (2013).

There are many versions of mathematical form of Hellinger distance. In the most general form, Hellinger distance between two probability measures P and Q, $D_H(P,Q)$, is defined as

$$D_H^2(P,Q) = \frac{1}{2}\int [\sqrt{p} - \sqrt{q}]^2 d\mu, \tag{2.1}$$

for some dominating measure μ. The following example shows that the choice of μ does not affect the value of $D_H(P,Q)$.

Example 2.1: (Shorack, 2000, p.68). Let P and Q denote probability measures on $(\Omega, \mathcal{A})$. Then the choice of dominating measure μ does not affect the value of $D_H(P,Q)$.

Solution: For any measure μ dominating both P and Q, i.e. $P << \mu$, $Q << \mu$, and for $A \in \mathcal{A}$, $P(A) = \int_A \frac{dP}{d\mu} d\mu$ and $Q(A) = \int_A \frac{dQ}{d\mu} d\mu$, where $p = \frac{dP}{d\mu}$ and $q = \frac{dQ}{d\mu}$ are Randon-Nikodym derivatives. Substituting p and q in (2.1), we get $D_H^2(P,Q) = \frac{1}{2}\int [\sqrt{p} - \sqrt{q}]^2 d\mu = \frac{1}{2}\int \frac{dP}{d\mu} d\mu + \frac{1}{2}\int \frac{dQ}{d\mu} d\mu - \int \sqrt{dP dQ} d\mu = 1 - \int \sqrt{pq}$. The first two integrals are equal to 1 because we are integrating probability density functions. Hence, the choice of dominating measure μ does not affect the value of $D_H(P,Q)$ and, for a discrete case, the counting measure can be used as the dominating measure both P and Q.

From Example 2.1, we find that minimizing the Hellinger distance is equivalent to maximizing the Bhattacharyya (BC) distance, $BC(P,Q) = \int \sqrt{pq}$ since $D_H^2(P,Q) = \frac{1}{2}[2(1 - BC(P,Q))]$.

In a dose-response studies setup, we are dealing with $N(= \sum_{j=1}^{K} n_j)$, $K \geq 2$ independent Bernoulli random variables, but not all identically distributed; for a trial at dose x_j, the probability of success is $F(\alpha + \beta x_j)$ and $\pi_j = \frac{m_j}{n_j}$ is

an estimate of this probability, $j = 1, \ldots, K$.

We define an estimator of (α, β) as the value of $(\hat{\alpha}, \hat{\beta})$ which minimizes the sum of Hellinger distances

$$\Delta_1(\alpha, \beta) = \sum_{j=1}^{K} n_j \{[\sqrt{\pi_j} - \sqrt{F(\alpha + \beta x_j)}]^2 + [\sqrt{1 - \pi_j} - \sqrt{1 - F(\alpha + \beta x_j)}]^2\}.$$

This is equivalent to maximizing

$$\Delta_2(\alpha, \beta) = \sum_{j=1}^{K} n_j \{[\sqrt{\pi_j F(\alpha + \beta x_j)} + \sqrt{(1 - \pi_j)(1 - F(\alpha + \beta x_j))}\}.$$

Now take $\frac{\partial \Delta(\alpha,\beta)}{\partial \alpha}$ and $\frac{\partial \Delta(\alpha,\beta)}{\partial \beta}$ and solve the equations $\frac{\partial \Delta(\alpha,\beta)}{\partial \alpha} = 0$ and $\frac{\partial \Delta(\alpha,\beta)}{\partial \beta} = 0$ to find estimators. However, there are no explicit solutions in this case, only numerical solutions can be obtained. The same situation occurs for maximum likelihood estimation in this context.

Let us denote $I^K = [0, 1] \times [0, 1] \times \ldots \times [0, 1]$ (K copies) and define $G_K = \{(\boldsymbol{\pi}, \mathbf{r}) \in I^K \times I^K : 0 \leq \pi_j \leq 1; \sum_{j=1}^{K} r_j = 1, r_j > 0,\ 1 \leq j \leq K\}$.

Definition 2.1: Let Θ be the parameter space for (α, β); $\Theta \subseteq \mathbb{R} \times (0, \infty)$. A Hellinger distance functional for estimating 'true' unknown parameter value (α_0, β_0) is a functional $T : G_K \to \Theta$ such that $T(\boldsymbol{\pi}, \mathbf{r})$ is a value of (α, β) maximizing

$$\Delta_1(\alpha, \beta) = \sum_{j=1}^{K} n_j \{[\sqrt{\pi_j} - \sqrt{F(\alpha + \beta x_j)}]^2 + [\sqrt{1 - \pi_j} - \sqrt{1 - F(\alpha + \beta x_j)}]^2\}.$$

Note that $\Delta_1(\alpha, \beta)/N$ can be written as:

$$\Delta(\alpha, \beta) = \sum_{j=1}^{K} r_j \{[\sqrt{\pi_j} - \sqrt{F(\alpha + \beta x_j)}]^2 + [\sqrt{1 - \pi_j} - \sqrt{1 - F(\alpha + \beta x_j)}]^2\},$$

where $r_j = \frac{n_j}{N}$, with $N = \sum_{j=1}^{K} n_j$. Thus $\sum_{j=1}^{K} r_j = 1$. This is also equivalent to maximizing

$$H(\alpha, \beta) = \sum_{j=1}^{K} r_j H_j(\alpha, \beta), \tag{2.2}$$

where $H_j(\alpha, \beta) = \{\sqrt{\pi_j F(\alpha + \beta x_j)} + \sqrt{(1 - \pi_j)(1 - F(\alpha + \beta x_j))}\}$.

For example, suppose $F(\alpha + \beta x_j) = e^{\alpha+\beta x_j}/(1 + e^{\alpha+\beta x_j})$, the CDF of the logistic distribution. Then substituting this F in (2.2) we have

$$H = \sum_{j=1}^{K} r_j (1 + e^{\alpha+\beta x_j})^{-\frac{1}{2}} (\sqrt{\pi_j e^{\alpha+\beta x_j}} + \sqrt{1 - \pi_j}).$$

Taking logarithm both sides we obtain:

$$\log H = \sum_{j=1}^{K} [\log r_j - \frac{1}{2} \log(1 + e^{\alpha+\beta x_j}) + \log(\sqrt{\pi_j e^{\alpha+\beta x_j}} + \sqrt{1 - \pi_j})].$$

Then the derivatives with respect to (w.r.t.) α and β give score values

$$\frac{\partial H}{\partial \alpha} = \frac{1}{2} \sum_{j=1}^{K} e^{\alpha+\beta x_j} [\frac{1}{e^{\alpha+\beta x_j} + \sqrt{(1 - \pi)/\pi} e^{(a+bx_j)/2}} - \frac{1}{1 + e^{\alpha+\beta x_j}}],$$

$$\frac{\partial H}{\partial \beta} = \frac{1}{2} \sum_{j=1}^{K} x_j e^{\alpha+\beta x_j} [\frac{1}{e^{\alpha+\beta x_j} + \sqrt{(1 - \pi)/\pi} e^{(\alpha+\beta x_j)/2}} - \frac{1}{1 + e^{\alpha+\beta x_j}}].$$

Let U_i denote the part common to both $\frac{\partial H}{\partial \alpha}$ and $\frac{\partial H}{\partial \beta}$. Then the covariance (Inverse Fisher Information) matrix can be approximated by

$$Var(\hat{\alpha}, \hat{\beta}) = E \begin{bmatrix} \sum_{j=1}^{K}(U_j) & \sum_{j=1}^{K}(U_j x_j) \\ \sum_{j=1}^{K}(U_j x_j) & \sum_{j=1}^{K}(U_j x_j^2) \end{bmatrix}^{-1},$$

2.2 Properties of The MHDE

2.2.1 Consistency

Definition 2.2: An estimator $\hat{\boldsymbol{\theta}}$ is said to be consistent if $\hat{\boldsymbol{\theta}} \to^P \boldsymbol{\theta}$ as $N \to \infty$, where $\boldsymbol{\theta}$ is considered as the true unknown parameter.

It is well known that MLEs are consistent under fairly general conditions (see Casella and Berger, 2002). The existence, continuity and consistency of the MHDE are shown in following three theorems.

Theorem 2.1: Existence

(i) If Θ is compact and F is continuous, then $T(\boldsymbol{\pi}, \mathbf{r})$ exists for all $(\boldsymbol{\pi}, \mathbf{r}) \in G_K$.

(ii) If F is strictly increasing on $\mathbb{R}$ and $\pi_j = F(\alpha + \beta x_j)$, $1 \le j \le K$, with not all x_j's equal, then $T(\boldsymbol{\pi}, \mathbf{r}) = (\alpha, \beta)^T$ uniquely.

Proof:

(i) From (2.2), $H(\alpha, \beta) = \sum_{j=1}^{K} r_j \{\sqrt{\pi_j F(\alpha + \beta x_j)} + \sqrt{(1 - \pi_j)(1 - F(\alpha + \beta x_j))}\}$.

For a sequence $(\alpha_n, \beta_n)_{n\geq 1}$ such that $(\alpha_n, \beta_n) \to (\alpha, \beta)$, we have

$$
\begin{aligned}
&|H(\alpha_n, \beta_n) - H(\alpha, \beta)| \\
=&\sum_{j=1}^{K} r_j |\sqrt{\pi_j F(\alpha_n + \beta_n x_j)} - \sqrt{\pi_j F(\alpha + \beta x_j)}| \\
+&\sum_{j=1}^{K} r_j |\sqrt{(1-\pi_j)(1 - F(\alpha_n + \beta_n x_j))} - \sqrt{(1-\pi_j)(1 - F(\alpha + \beta x_j)}|. \quad (2.3)
\end{aligned}
$$

Since $F(\alpha+\beta x_j)$ is nonnegative and continuous, we have $\sqrt{F(\alpha + \beta x_j)}$ and $\sqrt{1 - F(\alpha + \beta x_j)}$ continuous. Then $|\sqrt{F(\alpha_n + \beta_n x_j)} - \sqrt{F(\alpha + \beta x_j)}| \to 0$ and $|\sqrt{1 - F(\alpha_n + \beta_n x_j)} - \sqrt{1 - F(\alpha + \beta x_j)}| \to 0$ imply that $|H(\alpha_n, \beta_n) - H(\alpha, \beta)| \to 0$. So when π_j is given, $H(\alpha, \beta)$ is continuous with (α, β) on a compact set and attains a maximum there.

(ii) $H(\alpha, \beta)$ is maximized when $F(\alpha + \beta x_j) = \pi_j$, it is obvious from $\Delta(\alpha, \beta) \geq 0$. If there exists $\Delta(a, b) = 0$ for another (a, b), then $F(a+bx_j) = \pi_j$, $1 \leq j \leq K$. Since F is one-to-one this implies that $a+bx_j = \alpha+\beta x_j$, $1 \leq j \leq K$ and $(a, b) = (\alpha, \beta)$. Hence the result follows. The proof of Theorem 2.1 is complete.

Theorem 2.2: Continuity

Suppose Θ is compact, F is continuous and strictly increasing on $\mathbb{R}$ and $(\boldsymbol{\pi}, \mathbf{r})$ is such that $T(\boldsymbol{\pi}, \mathbf{r})$ is unique with for some $0 < \pi_j < 1$, $1 \leq j \leq K$. Then T is continuous in the Hellenger metric at $(\boldsymbol{\pi}, \mathbf{r})$.

Proof: Suppose $\{(\boldsymbol{\pi}_n, \mathbf{r}_n) \in G_K : n \geq 1\}$ is a sequence such that $(\boldsymbol{\pi}_n, \mathbf{r}_n) \to (\boldsymbol{\pi}, \mathbf{r})$ as $n \to \infty$ for some $(\boldsymbol{\pi}, \mathbf{r}) \in G_K$. Denote

$H_j(\alpha, \beta) = \sqrt{\pi_j F(\alpha + \beta x_j)} + \sqrt{(1-\pi_j)(1 - F(\alpha + \beta x_j))}$

and $H(\alpha, \beta) = \sum_{j=1}^{K} r_j H_j(\alpha, \beta)$.

Similarly, we define $H_n(\alpha, \beta) = \sum_{j=1}^{K} r_{j,n} H_{j,n}(\alpha, \beta)$, where $H_{j,n}(\alpha, \beta) =$

$\sqrt{\pi_{j,n}F(\alpha+\beta x_j)}+\sqrt{(1-\pi_{j,n})(1-F(\alpha+\beta x_j))}$ with π_j and r_j replaced by $\pi_{j,n}$ and $r_{j,n}$, respectively, $1\le j\le K$.

For convenience, let $T(\boldsymbol{\pi},\mathbf{r})=(\alpha,\beta)^T$, $T(\boldsymbol{\pi}_n,\mathbf{r}_n)=(\alpha_n,\beta_n)^T$. We want to show that $T(\boldsymbol{\pi}_n,\mathbf{r}_n)\to T(\boldsymbol{\pi},\mathbf{r})$ as $n\to\infty$, which needs

$$\sup\{|H_n(\alpha,\beta)-H(\alpha,\beta)|\colon (\alpha,\beta)^T\in\Theta\}\to 0. \tag{2.4}$$

To obtain (2.4), first note that

$$|H_n(\alpha,\beta)-H(\alpha,\beta)|\le\sum_{j=1}^{K}|H_j(\alpha,\beta)||r_{j,n}-r_j|+\sum_{j=1}^{K}|r_{j,n}||H_{j,n}(\alpha,\beta)-H_j(\alpha,\beta)|. \tag{2.5}$$

Since Θ is compact and F is strictly increasing on $\mathbb{R}$, $H_j(\alpha,\beta)$ is bounded and so $\mathbf{r}_n\to\mathbf{r}$ implies that the supremum of the first term on the RHS of (2.5) converges to zero. Let $F_j=F(\alpha+\beta x_j)$ and consider

$$\begin{aligned}\delta_{j,n}(\alpha,\beta)=&H_{j,n}(\alpha,\beta)-H_j(\alpha,\beta)\\ =&\sqrt{F_j}(\sqrt{\pi_{j,n}}-\sqrt{\pi_j})+\sqrt{1-F_j}(\sqrt{1-\pi_{j,n}}-\sqrt{1-\pi_j})\\ =&\frac{1}{2}\sqrt{\frac{F_j}{\pi_j}}[(\pi_{j,n}-\pi_j)-(\sqrt{\pi_{j,n}}-\sqrt{\pi_j})^2]\\ &-\frac{1}{2}\sqrt{\frac{1-F_j}{1-\pi_j}}[(\pi_{j,n}-\pi_j)+(\sqrt{1-\pi_{j,n}}+\sqrt{1-\pi_j})^2].\end{aligned} \tag{2.6}$$

Since $(\sqrt{b}-\sqrt{a})^2=\frac{(b-a)^2}{(\sqrt{b}+\sqrt{a})^2}\le\frac{(b-a)^2}{a}$ for $b\ge 0$, $a>0$, we have

$$\delta_{j,n}(\alpha,\beta) \leq \frac{1}{2}|\pi_{j,n}-\pi_j|[\sqrt{\frac{F_j}{\pi_j}}+\sqrt{\frac{1-F_j}{1-\pi_j}}]+\frac{1}{2}|\pi_{j,n}-\pi_j|^2[\sqrt{\frac{F_j}{\pi_j^3}}+\sqrt{\frac{1-F_j}{(1-\pi_j)^3}}]$$
$$\leq \frac{1}{2}[\sqrt{\pi_j F_j}+\sqrt{(1-\pi_j)(1-F_j)}]\epsilon_{j,n}, \tag{2.7}$$

where $\epsilon_{j,n} = |\pi_{j,n}-\pi_j|[\frac{1}{\pi_j}+\frac{1}{1-\pi_j}]+|\pi_{j,n}-\pi_j|^2[\frac{1}{\pi_j^2}+\frac{1}{(1-\pi_j)^2}]$.

Now $\pi_{j,n} \to \pi_j$ implies that $\epsilon_{j,n} \to 0$ and hence (2.4) is satisfied. Then it follows that $|\max\{H_n(\alpha,\beta)\} - \max\{H(\alpha,\beta)\}| \to 0$, i.e. $H_n(\alpha,\beta) \to H(\alpha,\beta)$. Further, (2.4) also implies that $|H_n(\alpha,\beta) - H(\alpha,\beta)| \to 0$ and hence

$$H_n(\alpha,\beta) \to H(\alpha,\beta),\ as\, n \to \infty. \tag{2.8}$$

The result follows from (2.8) by standard arguments based on the continuity of function H, compactness of Θ and uniqueness of (α,β). This completes the proof.

In order to study properties of $(\hat{\alpha},\hat{\beta})$ first we introduce some notation. We have $N = \sum_{j=1}^{K} n_j$. For $1 \leq j \leq K$, let $\pi_{j,N} = \frac{m_j}{n_j}$ and $r_{j,N} = \frac{n_j}{N}$ (previously π_j and r_j, respectively). Let $\boldsymbol{\pi}_N$ and $\boldsymbol{r_N}$ be the K dimensional vectors with components $\pi_{j,N}$ and $r_{j,N}$, respectively. Clearly $(\boldsymbol{\pi}_N, \boldsymbol{r_N}) \in G_K$.

Theorem 2.3: Consistency

Suppose Θ is compact, F is continuous and strictly increasing on $\mathbb{R}$ and we select subjects from a population for which the critical dose of a drug is a random variable specified by d.f. F with parameters (α_0,β_0). To each of $n_j\,(1 \leq j \leq K)$ subjects, a dose x_j is applied and m_j of these show a response. If $r_{j,N} \to r_j > 0$, $1 \leq j \leq K$, as $N \to \infty$, then $T(\boldsymbol{\pi}_N, \mathbf{r}_N) \to^P (\alpha_0,\beta_0)^T$ as

$N \to \infty$, i.e. the MHDE is consistent.

Proof: As $N \to \infty$, $\pi_{j,N} \to^P F(\alpha_0 + \beta_0 x_j) = \pi_{j0}$ since for large value of N the sample proportion will be expected to form an approximation to the parent population proportion.

From Theorem 2.2, we have T is continuous at $(\boldsymbol{\pi}, \mathbf{r})$. Let $T(\boldsymbol{\pi_0}, \mathbf{r_0}) = (\alpha_0, \beta_0)^T$. We want to show that $\lim_{N\to\infty} P(|T(\boldsymbol{\pi}_N, \mathbf{r}_N) - T(\boldsymbol{\pi_0}, \mathbf{r_0})| < \varepsilon) = 1$

Use the continuity of T to find $\delta > 0$ such that

$$|(\boldsymbol{\pi}_N, \mathbf{r}_N) - (\boldsymbol{\pi}_0, \mathbf{r_0})| < \delta \Rightarrow |T(\boldsymbol{\pi}_N, \mathbf{r}_N) - T(\boldsymbol{\pi}_0, \mathbf{r_0})| < \varepsilon.$$

Then $P(|(\boldsymbol{\pi}_N, \mathbf{r}_N) - (\boldsymbol{\pi_0}, \mathbf{r_0})| < \delta) \leq P(|T(\boldsymbol{\pi}_N, \mathbf{r}_N) - T(\boldsymbol{\pi_0}, \mathbf{r_0})| < \varepsilon)$. Here we use the fact that if one event can be implied by another, it has a greater probability. Since the first probability goes to 1 as $N \to \infty$, so $T(\boldsymbol{\pi}_N, \mathbf{r}_N) \to^P (\alpha_0, \beta_0)^T$. This completes the proof.

2.2.2 Asymptotic Normality

Definition 2.3: An estimator $\hat{\boldsymbol{\theta}}$ is said to be asymptotically multivariate normal if $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to^{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the asymptotic covariance matrix of the estimate $\hat{\boldsymbol{\theta}}$.

An MLE $\hat{\boldsymbol{\theta}}$ is asymptotically normal under fairly general conditions (see Casella and Beger, 2002) with $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to^{\mathcal{D}} N(\mathbf{0}, I^{-1}(\boldsymbol{\theta}))$, where $I(\boldsymbol{\theta})$ is the Fisher information matrix.

Before giving an asymptotic expansion for our functional T, we introduce some notations first. Define

$$G_j(y) = \frac{\partial}{\partial y}[\sqrt{\pi_j F(y)} + \sqrt{(1-\pi_j)(1-F(y))}] = \frac{f(y)}{2}(\sqrt{\frac{\pi_j}{F(y)}} - \sqrt{\frac{1-\pi_j}{1-F(y)}}),$$

$$G_j'(y) = \frac{f'(y)}{2}\{\sqrt{\frac{\pi_j}{F(y)}} - \sqrt{\frac{1-\pi_j}{1-F(y)}}\} - \frac{f^2(y)}{4}\{\sqrt{\frac{\pi_j}{F(y)^3}} + \sqrt{\frac{1-\pi_j}{[1-F(y)]^3}}\},$$

$$\begin{aligned} G_j''(y) =& \frac{f''(y)}{2}\{\sqrt{\frac{\pi_j}{F(y)}} - \sqrt{\frac{1-\pi_j}{1-F(y)}}\} - \frac{f'(y)[f'(y)+2f(y)]}{4}\{\sqrt{\frac{\pi_j}{F^3(y)}} \\ &+ \sqrt{\frac{1-\pi_j}{[1-F(y)]^3}}\} + \frac{3f^3(y)}{8}\{\sqrt{\frac{\pi_j}{F^5(y)}} - \sqrt{\frac{1-\pi_j}{[1-F(y)]^5}}\}, \end{aligned}$$

for $F(y) \neq 0$ and $F(y) \neq 1$;

$\Sigma = \sum_{j=1}^{K} r_j \mathbf{Z}_j \mathbf{Z}_j^T G_j'(\alpha_0 + \beta_0 x)$ and

$\boldsymbol{\lambda}(\alpha_0, \beta_0, \boldsymbol{\pi}, \mathbf{r}) = \sum_{j=1}^{K} r_j \mathbf{Z}_j G_j(\alpha_0 + \beta_0 x_j)$, with $\mathbf{Z}_j = (1, x_j)^T$.

Theorem 2.4: Suppose Θ is compact, F is a continuous, strictly increasing and thrice differentiable function on $\mathbb{R}$ with derivatives f, f' and f'' bounded on $C = \{\alpha + \beta x_j : (\alpha, \beta)^T \in \Theta, 1 \leq j \leq K\}$ and $F(C) \subseteq [\delta, 1-\delta]$ for some positive δ. Let $(\boldsymbol{\pi}, \mathbf{r}) \in G_K$ be such that $T(\boldsymbol{\pi_0}, \mathbf{r_0}) = (\alpha_0, \beta_0)^T$ is unique and let $(\boldsymbol{\pi}_N, \mathbf{r}_N) \to (\boldsymbol{\pi_0}, \mathbf{r_0})$. Let V_N be a 2×2 matrix whose components converge to 0 as $N \to \infty$. If Σ is non-singular then we have as $N \to \infty$,

$T(\boldsymbol{\pi}_N, \mathbf{r}_N) - T(\boldsymbol{\pi_0}, \mathbf{r_0}) = -\Sigma^{-1}\boldsymbol{\lambda}(\alpha, \beta, \boldsymbol{\pi}_N, \mathbf{r}_N)$.

Proof: Let $T(\boldsymbol{\pi}_N, \mathbf{r}_N) = (\hat{\alpha}_N, \hat{\beta}_N)^T$. Differentiability of F implies that $(\hat{\alpha}_N, \hat{\beta}_N)^T$ satisfies

$$\mathbf{0} = \boldsymbol{\lambda}(\hat{\alpha}_N, \hat{\beta}_N, \boldsymbol{\pi}_N, \mathbf{r}_N) = \sum_{j=1}^{K} r_{j,N} \mathbf{Z}_j G_{j,N}(\hat{\alpha}_N + \hat{\beta}_N x_j), \tag{2.9}$$

where $G_{j,N}$ is defined the same way as G_j with $\pi_{j,N}$ replacing π_j. Expanding

$G_{j,N}(\hat{\alpha}_N + \hat{\beta}_N x_j)$ at $\alpha_0 + \beta_0 x_j$ by Taylor series we obtain

$$G_{j,N}(\hat{\alpha}_N + \hat{\beta}_N x_j) = G_{j,N}(\alpha_0 + \beta_0 x_j) + \mathbf{Z}_j^T \boldsymbol{\gamma}_N G'_{j,N}(\alpha_0 + \beta_0 x_j) + \frac{1}{2}(\mathbf{Z}_j^T \boldsymbol{\gamma}_N)^2 G''_{j,N}(\kappa_j), \tag{2.10}$$

where κ_j is between $\hat{\alpha}_N + \hat{\beta}_N x_j$ and $\alpha_0 + \beta_0 x_j$, $\mathbf{Z}_j = (1, x_j)^T$ and $\boldsymbol{\gamma}_N = (\hat{\alpha}_N - \alpha_0, \hat{\beta}_N - \beta_0)^T$.

Since F has bounded derivatives and is bounded in the interval $[0, 1]$, $G''_{j,N}$ is continuous in a closed interval, $F(\kappa_j) \neq 0$ and $F(\kappa_j) \neq 1$, function $G''_{j,N}$ is bounded. Also, $\pi_{j,N} \to \pi_j$ implies that $G'_{j,N}(y) \to G'_j(y)$ uniformly in y. Then (2.10) becomes

$$G_{j,N}(\hat{\alpha}_N + \hat{\beta}_N x_j) - G_{j,N}(\alpha_0 + \beta_0 x_j) = \mathbf{Z}_j^T \boldsymbol{\gamma}_N G'_j(\alpha_0 + \beta_0 x_j) + o(\mathbf{Z}_j^T \boldsymbol{\gamma}_N). \tag{2.11}$$

Since Theorem 2.2 implies $\boldsymbol{\gamma}_N \to 0$ as $(\boldsymbol{\pi_N}, \mathbf{r}_N) \to (\boldsymbol{\pi_0}, \mathbf{r_0})$. Substituting (2.11) in (2.9) we get

$$\begin{aligned}
\mathbf{0} =& \sum_{j=1}^{K} r_{j,N} \mathbf{Z}_j G_{j,N}(\alpha_0 + \beta_0 x_j) + [\sum_{j=1}^{K} r_{j,N} \mathbf{Z}_j \mathbf{Z}_j^T \{G'_j(\alpha_0 + \beta_0 x_j) + o(1)\}] \boldsymbol{\gamma}_N \\
=& \boldsymbol{\lambda}(\alpha_0, \beta_0, \boldsymbol{\pi}_N, \mathbf{r}_N) + \{\Sigma + V_N\} \boldsymbol{\gamma}_N,
\end{aligned}$$

as $r_{j,N} \to r_j$. Here V_N is also a 2×2 matrix whose component are $o(1)$ as $N \to \infty$. Since $\{\Sigma + V_N\}$ will be non-singular for sufficiently large N,

$\Sigma^{-1} = (\Sigma + V_N)^{-1}$ as $N \to \infty$, the result follows, i.e. for large N,

$$\boldsymbol{\gamma}_N = T(\boldsymbol{\pi}_N, \mathbf{r}_N) - T(\boldsymbol{\pi}, \mathbf{r}) = -\Sigma^{-1}\boldsymbol{\lambda}(\alpha_0, \beta_0, \boldsymbol{\pi}_N, \mathbf{r}_N). \tag{2.12}$$

This completes the proof.

Notes: 1. $G_j(y) = \frac{f(y)}{2}(\sqrt{\frac{\pi_j}{F(y)}} - \sqrt{\frac{1-\pi_j}{1-F(y)}})$, so that $G_j(y) = 0$ when $\pi_j = F(y)$.

2. $G_j'(y) = G_j(y)\frac{f'(y)}{f(y)} - \frac{f^2(y)}{4}[\sqrt{\frac{\pi_j}{F(y)^3}} + \sqrt{\frac{1-\pi_j}{(1-F(y))^3}}]$, so that Σ is very complicated in general. However, when $\pi_j = F(y)$ we have $G_j'(y) = -\frac{f^2(y)}{4F(y)(1-F(y))}$, and we can get the following corollary.

Corollary 2.1: Suppose the conditions of Theorem 2.4 hold with $\pi_{j0} = F(\alpha_0 + \beta_0 x_j)$, $1 \le j \le K$, and let Σ^* be the 2×2 matrix defined by $\Sigma^* = -4\Sigma = \sum_{j=1}^{K} r_j \mathbf{Z}_j \mathbf{Z}_j^T \frac{f^2(\alpha_0 + \beta_0 x_j)}{F(\alpha_0 + \beta_0 x_j)(1 - F(\alpha_0 + \beta_0 x_j))}$.

Then we have

$$\boldsymbol{\gamma}_N = T(\boldsymbol{\pi}_N, \mathbf{r}_N) - T(\boldsymbol{\pi}_\mathbf{0}, \mathbf{r}_\mathbf{0}) = 4\Sigma^{*-1}\boldsymbol{\lambda}(\alpha_0, \beta_0, \boldsymbol{\pi}_N, \mathbf{r}_N), \tag{2.13}$$

as $N \to \infty$. We find that Σ^* is proportional to $\frac{f^2(\alpha_0+\beta_0 x_j)}{F(\alpha_0+\beta_0 x_j)(1-F(\alpha_0+\beta_0 x_j))}$, which has the same structure as the Information matrix of MLE.

Special case: Σ^* is singular only when $f(\alpha_0 + \beta_0 x_j) = 0$ for some j and $x_j = x$ for all other j values. Thus Σ^* is not singular except for this special case.

Finally we state a theorem establishing the asymptotic normality property of MHDE without a proof. A proof can be found in Chapter 3 of '*Advanced Multivariate Statistics with Matrices*' by Kollo and von Rosen (2005).

Theorem 2.5: Asymptotic Normality

Suppose the conditions in Theorem 2.3 hold and the expansion in Theorem 2.4 holds for $T(\boldsymbol{\pi}_N, \mathbf{r}_N)$.

Let $T(\boldsymbol{\pi}_\mathbf{0}, \mathbf{r}_\mathbf{0}) = (\alpha_0, \beta_0)^T$ and $\Sigma_x = \sum_{j=1}^{K} r_j^2 \mathbf{Z}_j \mathbf{Z}_j^T \frac{f^2(\alpha_0 + \beta_0 x_j)}{F(\alpha_0 + \beta_0 x_j)(1 - F(\alpha_0 + \beta_0 x_j))}$.

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

$\sqrt{N}\{T(\boldsymbol{\pi}_N, \mathbf{r}_N) - T(\boldsymbol{\pi}_\mathbf{0}, \mathbf{r}_\mathbf{0})\} \to^{\mathcal{D}} N(\mathbf{0}, \Sigma_H)$, where $\Sigma_H = \Sigma^{*-1}\Sigma_x\Sigma^{*-1}$.

Special case: If all the subgroups are equally weighted, then $r_j = \frac{1}{K}$, $1 \leq j \leq K$, $\Sigma_H = \frac{1}{K}\Sigma^{*-1}$.

2.3 Simulation Study

2.3.1 Influence Function

The idea for the influence function comes from Hampel (1968) and Huber (1972, 1973). Beran (1977) introduced the α-influence curve to measure the robustness of an MHDE. As the name implies, the influence function of an estimator measures the impact that a single observation can have on an estimator. Assume that a sample has n observations $(x_1, \ldots, x_{n-1}, x_n)$, of which the first $n-1$ observations belong to a distribution F and x_n can then take on any value. The influence function $IF_n(x)$ then measure the standard difference between the two estimators: $IF_n(x) = \frac{T(x_1, x_2, \ldots, x_n) - Tn-1(x_1, x_2, \ldots, x_{n-1})}{\varepsilon}$. Lindsay (1994) used a bias plot, which is related to the influence function as $bias \approx \varepsilon * IF_n(x)$.

For measuring the effect of a single contaminated observation in a sample

of size n, so $\varepsilon = 1/n$. This influence function can then be written as

$$IF_n(x) = \frac{T(x_1, x_2, \ldots, x_n) - T(x_1, x_2, \ldots, x_{n-1})}{1/n} = n(\hat{\theta}_n - \hat{\theta}_{n-1}).$$

The above definition applies to finite samples, but it can be generalized to the asymptotic case as follows. Under some regularity conditions, the asymptotic influence function can be derived from the Gâteaux derivative:

$$IF(G) = \lim_{\epsilon \to 0} \frac{T((1-\epsilon)F + \epsilon G) - T(F)}{\epsilon},$$

where F is the appropriate distribution and $G = \delta_x$, a point mass distribution with a mass at x.

Since our responses are binary numbers, either 1 (response) and 0 (no response), it is hard to see the robustness property of estimators if we are trying to put a mass at a point. Thus, we use some Monte Carlo studies to examine the robustness of estimators for departures from the assumed model.

2.3.2 An Algorithm for MHDE Calculation

We discussed algorithms for WLSE and MLE in the introduction. In this section, we focus on how to obtain MHDE numerically. First introduce two $K \times 2$ matrices: $\mathbf{A} = \begin{pmatrix} r_1 p_1 & r_1(1-p_1) \\ \vdots & \vdots \\ r_K p_K & r_K(1-p_K) \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} r_1 q_1 & r_1(1-q_1) \\ \vdots & \vdots \\ r_K q_K & r_K(1-q_K) \end{pmatrix}$, where $p_j = \pi_j$ and $q_j = F(\alpha + \beta x_j)$, $j = 1, \ldots, K$. $vec(A)$ is the $2K \times 1$ vector formed by stringing the first column of A out followed by the second one.

Definition 2.4: Define two discrete probability distributions

$$\mathbf{P} = vec(\mathbf{A}) = \begin{pmatrix} r_1 p_1 \\ \vdots \\ r_K p_K \\ r_1(1-p_1) \\ \vdots \\ r_K(1-p_K) \end{pmatrix} \text{ and } \mathbf{Q} = vec(\mathbf{B}) = \begin{pmatrix} r_1 q_1 \\ \vdots \\ r_K q_K \\ r_1(1-q_1) \\ \vdots \\ r_K(1-q_K) \end{pmatrix},$$

where $p_j = \pi_j$ and $q_j = F(\alpha + \beta x_j)$, $\sum_{j=1}^{K} r_j = 1$, $j = 1, \ldots, K$. Then the Hellinger distance between $\mathbf{P}$ and $\mathbf{Q}$ is defined as

$$D_H(\mathbf{P}, \mathbf{Q}) = \frac{1}{\sqrt{2}} \left\| \sqrt{\mathbf{P}} - \sqrt{\mathbf{Q}} \right\|_2. \tag{2.14}$$

Definition 2.5: A Hellinger distance functional for estimating true parameter value (α_0, β_0) is a functional $T : G_K \to \Theta$ such that $T(\boldsymbol{\pi}, \mathbf{r})$ is the estimator of true parameter value as $(\hat{\alpha}, \hat{\beta})$ which minimizes the square of Hellinger distance:

$$2D_H^2(\mathbf{P}, \mathbf{Q}) = \left\| \sqrt{\mathbf{P}} - \sqrt{\mathbf{Q}} \right\|^2. \tag{2.15}$$

Let $\mathbf{A_1} = \begin{pmatrix} \sqrt{r_1}\sqrt{p_1} & \sqrt{r_1}\sqrt{1-p_1} \\ \vdots & \vdots \\ \sqrt{r_K}\sqrt{p_K} & \sqrt{r_K}\sqrt{(1-p_K)} \end{pmatrix}$ and $\mathbf{B_1} = \begin{pmatrix} \sqrt{r_1}\sqrt{q_1} & \sqrt{r_1}\sqrt{1-q_1} \\ \vdots & \vdots \\ \sqrt{r_K}\sqrt{q_K} & \sqrt{r_K}\sqrt{1-q_K} \end{pmatrix}$,

where $p_j = \pi_j$ and $q_j = F(\alpha+\beta x_j)$, $j = 1, \ldots, K$. $\|vec(\mathbf{A_1})\|_2 = \|vec(\mathbf{B_1})\|_2 = 1$, with $\sum_{j=1}^{K} r_j = 1$.

Note that equation (2.15) is equivalent to:

$$\begin{aligned}\left\|\sqrt{\mathbf{P}}-\sqrt{\mathbf{Q}}\right\|^2 &= \|vec(\mathbf{A_1})-vec(\mathbf{B_1})\|^2 \\ &= \|vec(\mathbf{A_1}-\mathbf{B_1})\|^2 \\ &= tr[(\mathbf{A_1}-\mathbf{B_1})^T(\mathbf{A_1}-\mathbf{B_1})] \\ &= 2-tr(\mathbf{A_1}^T\mathbf{B_1}). \end{aligned} \tag{2.16}$$

So, minimizing (2.14) is also equivalent to maximizing

$$tr(\mathbf{A_1}^T\mathbf{B_1}) = H(\alpha,\beta). \tag{2.17}$$

For a numerical implementation of this method, initially input two matrices $\mathbf{A}$ and $\mathbf{B}$. In many statistical software programs, for example in R, we use $sqrt(\mathbf{A})$ to obtain the square root of all the elements in the matrix, and the result is exactly the $\mathbf{A_1}$. A similar method can be used to obtain $\mathbf{B_1}$. Finally, minimize $tr(\mathbf{A_1}^T\mathbf{B_1})$ to obtain MHDE of the parameter.

2.3.3 Monte Carlo Studies (Logistic Model)

In order to compare the MHDE with MLE and WLSE, a simulation study is conducted. Suppose $x_i,\ i = 1, ..., 10$, represent doses given to 20 subjects. Thus, for each dose x_j, we generate 20 observations from the Bernoulli distribution with probability of success $F(\alpha+\beta x_j)$. We take $\alpha = -2$, $\beta = 0.4$ as the true values and $x_j = j$, $1 \le j \le 10$. Calculations are carried out assuming that F is the CDF from the logistic distribution family first. By using different methods, following four models are used to find the means, variances and covariances of $\hat{\alpha}_n$ and $\hat{\beta}_n$ under 1000 replications.

Model I: $F(y) = \frac{e^y}{1+e^y} = L(y)$,

Model II: $F(y) = 0.9L(y) + 0.1L(2y)$,

Model III: $F(y) = 0.9L(y) + 0.1L(0.5y)$,

Model IV: $F(y) = 0.9L(y) + 0.1$.

Model I is simply the standard logistic model. Model II, III, IV are derived from the classical Tukey-Huber contamination model. Model II represents 10% contamination from a distribution with shorter tails while model III is mixed with a longer tails one. Model IV represents the situation where 10% of the subjects show a response not caused by the stimulus under examination; for example, if subjects recover naturally (similar to the censoring data of survival analysis. We will show robustness by using this model). Simulation results are given in Tables 2.1 and 2.2. These results are based on 1000 replications. $\hat{\alpha}_n(m)$ and $\hat{\beta}_n(m)$, $m = 1, 2, 3, \ldots, 1000$ are estimators of α and β based on the true distribution, then $\hat{\alpha}_n$ and $\hat{\beta}_n$ can be found by averaging the 1000 results of $\hat{\alpha}_n(m)$ and $\hat{\beta}_n(m)$. In the tables, $V(\hat{\alpha}_n)$ and $V(\hat{\beta}_n)$ are used to denote the estimated variance of $\hat{\alpha}_n$ and $\hat{\beta}_n$. Similarly, $Cov(\hat{\alpha}_n, \hat{\beta}_n)$ stands for the estimated covariance between $\hat{\alpha}_n$ and $\hat{\beta}_n$. Also, $MSE(\hat{\alpha}_n)$ and $MSE(\hat{\beta}_n)$ denote estimated the mean squared errors of $\hat{\alpha}_n$ and $\hat{\beta}_n$.

Table 2.1: Iteratively Reweighted Least Squares Estimation

Model	$\hat{\alpha}_n$	$\hat{\beta}_n$	$V(\hat{\alpha}_n)$	$V(\hat{\beta}_n)$	$Cov(\hat{\alpha}_n, \hat{\beta}_n)$	$MSE(\hat{\alpha}_n)$	$MSE(\hat{\beta}_n)$
Model I	-1.8122	0.3619	0.6295	0.0234	-0.1186	0.6642	0.0248
Model II	-1.8106	0.361	0.8596	0.0325	-0.1643	0.8946	0.0339
Model III	-1.7357	0.3465	0.5338	0.0195	-0.0992	0.6031	0.0223
Model IV	-1.2952	0.3019	0.4662	0.0228	-0.1001	0.9625	0.0324

Table 2.2: Results of Four Estimation Methods by Logistic Model

Model	Method	$\hat{\alpha}_n$	$\hat{\beta}_n$	$V(\hat{\alpha}_n)$	$V(\hat{\beta}_n)$	$Cov(\hat{\alpha}_n, \hat{\beta}_n)$	$MSE(\hat{\alpha}_n)$	$MSE(\hat{\beta}_n)$
I	WLSE	−1.9131	0.3817	0.1149	0.0033	−0.0174	0.1223	0.0036
	MLE	−2.046	0.4084	0.1437	0.0040	−0.0215	0.1457	0.0041
	MHDE	−2.1284	0.4277	0.1820	0.0054	−0.0283	0.1983	0.0062
II	WLSE	−2.0093	0.4005	0.1241	0.0037	−0.0193	0.124	0.0037
	MLE	−2.1616	0.4347	0.1426	0.0044	−0.0226	0.1686	0.0056
	MHDE	−2.1053	0.4214	0.1731	0.0054	−0.028	0.1841	0.0058
III	WLSE	−2.1217	0.4291	0.4709	0.0103	−0.0644	0.4852	0.0111
	MLE	−1.8799	0.3776	0.1403	0.0041	−0.0216	0.1546	0.0046
	MHDE	−2.0379	0.4082	0.1652	0.0051	−0.0263	0.1655	0.0052
IV	WLSE	−1.6954	0.4027	0.4128	0.0094	−0.0572	0.5051	0.0094
	MLE	−1.5486	0.3613	0.1242	0.0039	−0.0197	0.3278	0.0054
	MHDE	−2.14	0.4295	0.2320	0.0069	−0.0366	0.2514	0.0077

The results in Table 2.1 show that the IRLS method is not good in this problem; it is far away from the true value and not comparable with others. We focused on the comparison of other three methods. In all cases, MLE is the highest variable while the MHDE is the least variable. WLSE has the smallest variance for Models I and II among the three. If we consider only the biases of the three estimators, we see that the MLE is least biased if the postulated model is correct, i.e. Model I. Further we see that the weighted least squares estimator shows the least bias under Model II, and the MHDE shows the least bias under Models III and IV. These results suggest that the method of weighted least squares has some protection if the true distribution has shorter tail contamination. Also, the method of MHDE has some protection if the true distribution has longer tail contamination. The results for Model IV suggest that the MHDE might be the best if we wish to protect against the possibility of subjects showing a response regardless of what dose they receive, and the next graph shows a comparison of MLE and MHDE by using model IV.

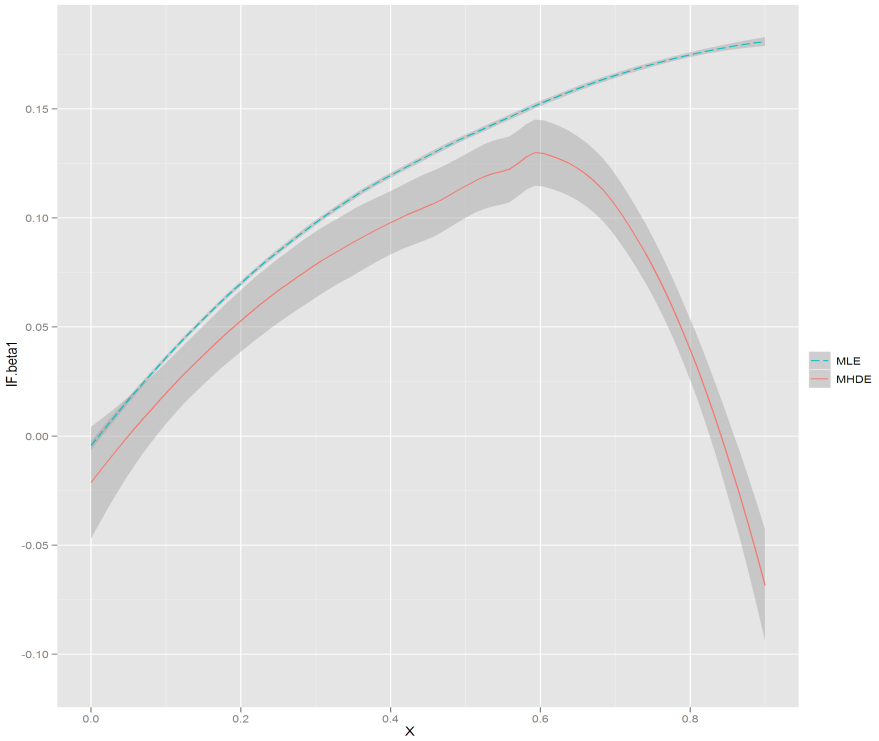


Figure 2.1: Bias Plots between MLE and MHDE for Logistic Model

As shown in Figure 2.1, after increasing the contamination rate (X), the bias plot of MHDE is showing bounded trend but that of MLE keeps increasing. We find that the MHDEs of α and β are comparable to those of the MLE and WLSE in most cases, but the MHDE estimator of α slightly outperforms other two estimators when the data are under contamination. Thus we can conclude that MHD estimator has desirable robustness properties as well as asymptotically efficient properties when using the logistic model.

2.3.4 Monte Carlo Studies (Probit Model)

Although it is computationally more convenient to use the logistic distribution function for binary responses, normal distribution is also frequently used in numerical calculations. For each of the three methods of estimation we estimate the means, variances and covariances of $\hat{\alpha}_n$ and $\hat{\beta}_n$ again under each of the following four models:

Model I: $F(y) = \Phi(y)$, where $\Phi(y)$ is the CDF of standard normal distribution.

Model II: $F(y) = 0.9\Phi(y) + 0.1\Phi(2y)$,

Model III: $F(y) = 0.9\Phi(y) + 0.1\Phi(0.5y)$,

Model IV: $F(y) = 0.9\Phi(y) + 0.1$.

Model I is the case where the chosen model is $N(0, 1)$, the standard normal distribution. Model II represents 10% contamination from $N(0, 0.25)$, while model III is mixed with 10% from $N(0, 4)$; Model IV represents the situation where 10% of the subjects show a response not caused by the stimulus under examination; for example, if subjects recover naturally. Simulation esults are given in Table 2.3, and the values in table are based on 1000 replications.

Again, IRLS estimation is not good for the Probit Model either and the output of this method is not posted.

Table 2.3: Results of Four Estimation Methods by Probit Model

Model	Method	$\hat{\alpha}_n$	$\hat{\beta}_n$	$V(\hat{\alpha}_n)$	$V(\hat{\beta}_n)$	$Cov(\hat{\alpha}_n, \hat{\beta}_n)$	$MSE(\hat{\alpha}_n)$	$MSE(\hat{\beta}_n)$
I	WLSE	−2.0865	0.4175	0.0636	0.0019	−0.0088	0.0706	0.0022
	MLE	−2.0423	0.409	0.0794	0.0025	−0.0130	0.0811	0.0026
	MHDE	−2.217	0.4435	0.1156	0.0039	−0.0197	0.1626	0.0058
II	WLSE	−2.171	0.4325	0.0968	0.0028	−0.0146	0.1253	0.0038
	MLE	−2.1314	0.4268	0.0839	0.0028	−0.0140	0.1011	0.0035
	MHDE	−2.2232	0.4446	0.1183	0.0039	−0.0202	0.168	0.0059
III	WLSE	−1.6825	0.3712	0.1054	0.0028	−0.0151	0.2057	0.0036
	MLE	−1.8858	0.3778	0.0682	0.0022	−0.0111	0.0811	0.0027
	MHDE	−2.173	0.4366	0.1241	0.0042	−0.0211	0.1539	0.0055
IV	WLSE	−1.6814	0.4001	0.4216	0.0096	−0.0586	0.5277	0.0096
	MLE	−1.5747	0.3521	0.0558	0.0018	−0.0091	0.2366	0.0041
	MHDE	−2.2288	0.4483	0.1954	0.0057	−0.0313	0.2476	0.008

The results in the Table 2.3 show that, in all cases, the value of MHDEs are least variable, while the MLEs are the most variable. If we consider only the biases of these estimators, we see that the MLEs are least biased if the postulated model is correct, i.e. Model I. Further we see that the MLEs also show the least bias under Models II and III, while the MHDEs show the least bias under Model IV. These results suggest that the method of MLE has some protection if the true distribution has shorter or longer tails contamination. Also, the results for Model IV suggest that the MHDE might be the best if we wish to protect against the possibility of subjects showing a response regardless of what dose they receive. Figure 2.2 shows a comparison of MLE and MHDE by using Model IV.

Using the same method of bias plot, we also did not see the boundedness of the MLE. But it shows that after increasing the contamination rate (X), the

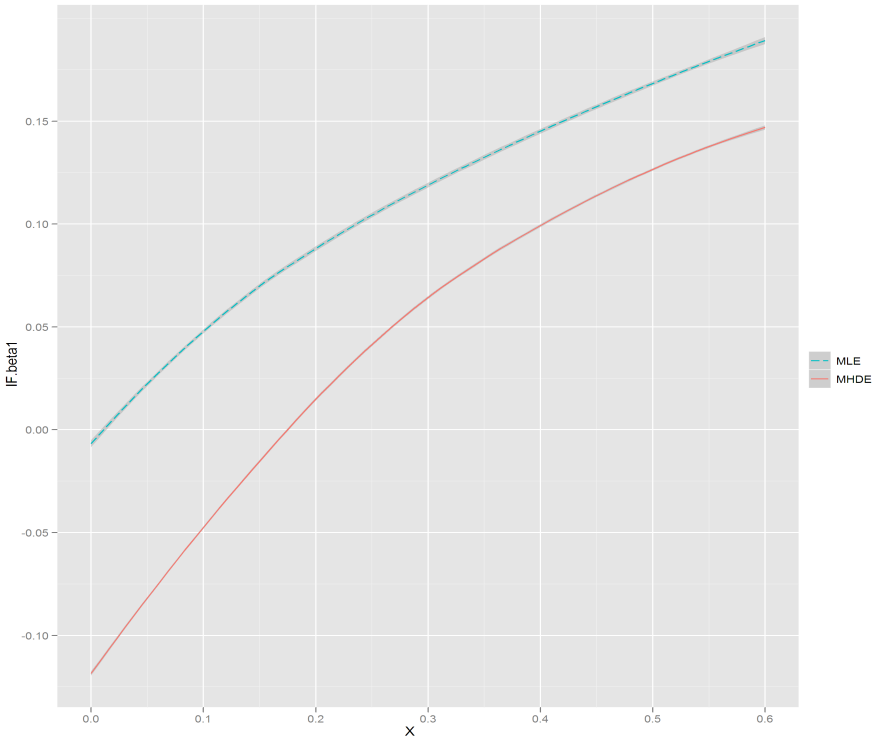


Figure 2.2: Bias Plots between MLE and MHDE for Probit Model

MHDE increases slower than the MLE. Thus, we can reasonably conclude that, when using the Probit model in this dose-response study, Minimum Hellinger distance (MHD) estimation method has desirable robustness properties as well as asymptotically efficiency properties. We find that the MHDEs of α and β are comparable to those of the MLE and WLSE in most cases, but the MHDE estimator of α slightly outperforms the other two estimators when the data are under contamination.

2.4 An Application to Real Data Example

In this Section, we illustrate the above methods for a real data set given in Giltinan et al. (1988). This data are collected from an experiment to investigate the joint activity of two insecticides. Two insecticides are denoted here by A and B. The mixtures are chosen in the ratios 0 : 100, 25 : 75, 75 : 25 and 100 : 0. 30 insects were tested at each of 4 dose levels of each mixture, the insects were exposed for 96 hours to these insecticides and the mortality count were recorded after that. The number of dead insects and total number of insects exposed are presented in Table 2.8.

In this example, we used the model: $F(\alpha + \beta \log x_i) = \frac{e^{\alpha+\beta \log x_i}}{1+e^{\alpha+\beta \log x_i}}$, $i = 1, 2, 3, 4$. From the results of 100% B and 100% A, we used GLM regression, separately, to obtain the parameter values and consider them as the 'true' values: $\alpha_B = -4.4101$, $\beta_B = 1.8056$ and $\alpha_A = -3.1501$, $\beta_A = 1.3699$.

From the simulation section, we conclude that the MHDE might offer some protection if the true distribution has longer tail than the postulated model distribution. We consider B as the postulated model and A as the contaminated model with longer tail. For $75\%B + 25\%A$, we obtained $\hat{\alpha}_{MLE} =$

$-7.7302, \hat{\beta}_{MLE} = 2.6407$; but $\hat{\alpha}_{MHDE} = -4.626892$, $\hat{\beta}_{MHDE} = 1.0608$, which are more robust in dealing with longer tail contamination.

Table 2.4: Mortality in response to mixtures of insecticides

Mixture	Amount of A (ppm)	Amount of B (ppm)	dead insects	insects tested
B	0	30.00	26	30
B	0	15.00	19	30
B	0	7.50	7	30
B	0	3.75	5	30
A25:B75	6.5	19.50	23	30
A25:B75	3.25	9.75	11	30
A25:B75	1.625	4.875	3	30
A25:B75	0.813	2.438	0	30
A75:B25	19.50	6.50	20	30
A75:B25	9.75	3.25	13	30
A75:B25	4.875	1.625	6	29
A75:B25	2.438	0.813	0	30
A	30.00	0	23	30
A	15.00	0	21	30
A	7.50	0	13	30
A	3.75	0	5	30

Chapter 3

Symmetric χ^2 Distance Method

3.1 Background

Consider two discrete probability distributions $P = \{f_i : i \in S\}$ and $Q = \{g_i : i \in S\}$, where S is a discrete set, $\sum f_i = \sum g_i = 1$, $f_i > 0$ and $g_i > 0$. Then the square of Hellinger distance between P and Q is defined as

$$\begin{aligned} D^2(P,Q) &= \sum_{i \in S} (\sqrt{f_i} - \sqrt{g_i})^2 \\ &= \sum_{i \in S} \frac{(f_i - g_i)^2}{(\sqrt{f_i} + \sqrt{g_i})^2}. \end{aligned}$$

On the other hand, the *symmetric chi-squared* distance between P and Q is defined as (Lindsay, 1994)

$$S^2(P,Q) = 2 \sum_{i \in S} \frac{(f_i - g_i)^2}{(f_i + g_i)^2}.$$

Using the inequalities

$$f_i + g_i \leq (\sqrt{f_i} + \sqrt{g_i})^2 \leq 2(f_i + g_i),$$

a little manipulation gives the following near equivalence relationship between Hellinger distance and the symmetric chi-squared distance:

$$\frac{1}{4}S^2(P,Q) \leq D^2(P,Q) \leq \frac{1}{2}S^2(P,Q).$$

Lindsay (1994) noted that, although both are equally robust to outlying observations, Hellinger distance $D^2(P,Q)$ does not behave as well for sampling zeros. (In biological and ecological studies, sampling zeroes typically occur when a species is present but absent in the sample. We borrow the definition from sampling techniques and consider no response at a dose level as sampling zeros.) For this reason, he prefers $S^2(P,Q)$ over $D^2(P,Q)$ for use in statistical inference, especially for discrete distributions. On the other hand, $D^2(P,Q)$ is better than $S^2(P,Q)$ for theoretical calculations.

We apply $S^2(P,Q)$ with

$$P = \begin{pmatrix} r_1 p_1 \\ \vdots \\ r_K p_K \\ r_1(1-p_1) \\ \vdots \\ r_K(1-p_K) \end{pmatrix} \text{ and } Q = \begin{pmatrix} r_1 q_1 \\ \vdots \\ r_K q_K \\ r_1(1-q_1) \\ \vdots \\ r_K(1-q_K) \end{pmatrix},$$

where $p_j = \pi_j$, $q_j = F(\alpha + \beta x_j)$, $\sum_{j=1}^{K} r_j = 1$, $j = 1, \ldots, K$, $\pi_j = \frac{m_j}{n_j}$, $r_j = \frac{n_j}{N}$ and $N = \sum n_i$. Then $S^2(P,Q)$ reduces to

$$S^2(P,Q) = 2\sum_{j=1}^{K}\{\frac{[\pi_j - F(\alpha+\beta x_j)]^2}{[\pi_j + F(\alpha+\beta x_j)]^2} + \frac{[(1-\pi_j) - (1-F(\alpha+\beta x_j))]^2}{[(1-\pi_j) + (1-F(\alpha+\beta x_j))]^2}\}.$$

We define estimators of (α_0, β_0) as $(\hat{\alpha}, \hat{\beta})$ that minimize $S^2(P,Q)$. Then take $\frac{\partial S^2(P,Q)}{\partial \alpha}$ and $\frac{\partial S^2(P,Q)}{\partial \beta}$ and solve the equations $\frac{\partial S^2(P,Q)}{\partial \alpha} = 0$ and $\frac{\partial S^2(P,Q)}{\partial \beta} = 0$ to find estimators. Again they cannot be solved explicitly and only numerical solutions can be obtained.

Definition 3.1: Suppose G_K is as defined in Chapter 2 and Θ is the parameter space for (α, β); $\Theta \subseteq \mathbb{R} \times (0, \infty)$. A symmetric chi-squared distance estimate (SCDE) functional for estimating (α_0, β_0) is a functional $T : G_K \to \Theta$ such that $T(\boldsymbol{\pi}, \mathbf{r})$ is $(\hat{\alpha}, \hat{\beta})$ obtained by minimizing

$$\Delta(\alpha,\beta) = \sum_{j=1}^{K}\{\frac{[\pi_j - F(\alpha+\beta x_j)]^2}{[\pi_j + F(\alpha+\beta x_j)]^2} + \frac{[(1-\pi_j) - (1-F(\alpha+\beta x_j))]^2}{[(1-\pi_j) + (1-F(\alpha+\beta x_j))]^2}\}. \quad (3.1)$$

Note that $\Delta(\alpha, \beta)$ can also be written as

$$\Delta(\alpha,\beta) = \sum_{j=1}^{K}\sum_{i=1}^{2}\{\frac{[\pi_{i,j} - q_{i,j}]^2}{[\pi_{i,j} + q_{i,j}]^2}\}, \quad (3.2)$$

where $\pi_{1,j} = \pi_j$, $\pi_{2,j} = 1 - \pi_{1,j}$; $q_{1,j} = F(\alpha + \beta x_j)$, $q_{2,j} = 1 - F(\alpha + \beta x_j)$.

3.2 Properties of The SCDE

3.2.1 Consistency

Theorem 3.1: Existence

(i) If Θ is compact and F is continuous, then except $\pi_j = 1$, $1 \le j \le K$, $T(\boldsymbol{\pi}, \mathbf{r})$ exists for all $(\boldsymbol{\pi}, \mathbf{r}) \in G_K$.

(ii) If F is strictly increasing on $\mathbb{R}$ and $\pi_j = F(\alpha + \beta x_j)$, $1 \le j \le K$, with not all x_j's equal, then $T(\boldsymbol{\pi}, \mathbf{r}) = (\alpha, \beta)^T$ uniquely for any $\mathbf{r}$.

Proof:

(i) Let $\Delta(\alpha, \beta) = \sum_{j=1}^{K} \{ \frac{[\pi_j - F(\alpha + \beta x_j)]^2}{[\pi_j + F(\alpha + \beta x_j)]^2} + \frac{[(1-\pi_j) - (1 - F(\alpha + \beta x_j))]^2}{[(1-\pi_j) + (1 - F(\alpha + \beta x_j))]^2} \}$, and for a sequence $(\alpha_n, \beta_n)_{n \ge 1} \to (\alpha, \beta)$ as $n \to \infty$, write

$$
\begin{aligned}
&|\Delta(a_n, b_n) - \Delta(a, b)| \\
=&\sum_{j=1}^{K} \{ \frac{[\pi_j - F(\alpha_n + \beta_n x_j)]^2}{[\pi_j + F(\alpha_n + \beta_n x_j)]^2} - \frac{[\pi_j - F(\alpha + \beta x_j)]^2}{[\pi_j + F(\alpha + \beta x_j)]^2} \} \\
+&\sum_{j=1}^{K} \{ \frac{[(1-\pi_j) - (1 - F(\alpha_n + \beta_n x_j))]^2}{[(1-\pi_j) + (1 - F(\alpha_n + \beta_n x_j))]^2} - \frac{[(1-\pi_j) - (1 - F(\alpha + \beta x_j))]^2}{[(1-\pi_j) + (1 - F(\alpha + \beta x_j))]^2} \}.
\end{aligned}
\tag{3.3}
$$

Since $F(\alpha + \beta x_j)$ nonnegative and continuous, we have the function of $F(\alpha + \beta x_j)$ continuous, then $|\Delta(\alpha_n, \beta_n) - \Delta(\alpha, \beta)| \to 0$. So when π_j is given, $\Delta(\alpha, \beta)$ is continuous with (α, β) on compact set and there exist a minimum.

(ii) $\Delta(\alpha, \beta)$ is minimized when $F(\alpha + \beta x_j) = \pi_j$. If there exists $\Delta(a, b) = 0$ for another (a, b), then $F(a + b x_j) = \pi_j$, $1 \le j \le K$. Since F is one-to-one this implies that $a + b x_j = \alpha + \beta x_j$,$1 \le j \le K$. Hence the result follows. The proof of Theorem 3.1 is complete.

Theorem 3.2: Continuity

Suppose Θ is compact, F is continuous and strictly increasing on $\mathbb{R}$ and $(\boldsymbol{\pi}, \mathbf{r})$ is such that $T(\boldsymbol{\pi}, \mathbf{r})$ is unique with $0 < \pi_j < 1$, $1 \le j \le K$. Then T is continuous at $(\boldsymbol{\pi}, \mathbf{r})$.

Proof: Suppose $\{(\boldsymbol{\pi}_n, \mathbf{r}_n) \in G_K : n \ge 1\}$ is a sequence such that $(\boldsymbol{\pi}_n, \mathbf{r}_n) \to$

$(\boldsymbol{\pi}, \mathbf{r})$ as $n \to \infty$ for some $(\boldsymbol{\pi}, \mathbf{r}) \in G_K$. Denote

$\Delta_n(\alpha, \beta) = \sum_{j=1}^{K} \{\frac{[\pi_{j,n} - F(\alpha+\beta x_j)]^2}{[\pi_{j,n} + F(\alpha+\beta x_j)]^2} + \frac{[(1-\pi_{j,n}) - (1-F(\alpha+\beta x_j))]^2}{[(1-\pi_{j,n}) + (1-F(\alpha+\beta x_j))]^2}\}$. For convenience, let $(\alpha, \beta)^T = T(\boldsymbol{\pi}, \mathbf{r})$, $(\alpha_n, \beta_n)^T = T(\boldsymbol{\pi}_n, \mathbf{r}_n)$ (any of the possible values will do in the latter case). It is sufficient to show that

$$\sup\{|\Delta_n(\alpha, \beta) - \Delta(\alpha, \beta)|\colon (\alpha, \beta)^T \in \Theta\} \to 0. \tag{3.4}$$

Let $F_j = F(\alpha + \beta x_j)$,

$$\begin{aligned}
|\Delta_n(\alpha, \beta) - \Delta(\alpha, \beta)| =& \sum_{j=1}^{K} \{\frac{(\pi_{j,n} - F_j)^2}{(\pi_{j,n} + F_j)^2} + \frac{[(1-\pi_{j,n}) - (1-F_j)]^2}{[(1-\pi_{j,n}) + (1-F_j)]^2}\} \\
& - \sum_{j=1}^{K} \{\frac{(\pi_j - F_j)^2}{(\pi_j + F_j)^2} + \frac{[(1-\pi_j) - (1-F_j)]^2}{[(1-\pi_j) + (1-F_j)]^2}\} \\
\leq & \sum_{j=1}^{K} |\frac{(\pi_{j,n} - F_j)]^2}{(\pi_{j,n} + F_j)]^2} - \frac{(\pi_j - F_j)^2}{(\pi_j + F_j)^2}| \\
& + \sum_{j=1}^{K} |\frac{[(1-\pi_{j,n}) - (1-F_j)]^2}{[(1-\pi_{j,n}) + (1-F_j)]^2} - \frac{[(1-\pi_j) - (1-F_j)]^2}{[(1-\pi_j) + (1-F_j)]^2}| \\
= & \sum_{j=1}^{K} |\frac{4F_j(\pi_{j,n}\pi_j - F_j^2)}{[(\pi_{j,n} + F_j)(\pi_j + F_j)]^2}(\pi_{j,n} - \pi_j)| \\
& + \sum_{j=1}^{K} |\frac{4(\pi_{j,n} + \pi_j - \pi_{j,n}\pi_j - 2F_j + F_j^2)}{[(2 - \pi_{j,n} - F_j)(2 - \pi_j - F_j)]^2}(\pi_{j,n} - \pi_j)|.
\end{aligned} \tag{3.5}$$

$|\Delta_n(\alpha, \beta) - \Delta(\alpha, \beta)|$ is bounded since all $\pi_{j,n}$, π_j and F_j are bounded. Also, $\pi_{j,n} \to \pi_j$ implies that $|\Delta_n(\alpha, \beta) - \Delta(\alpha, \beta)| \to 0$ and hence $\sup\{|\Delta_n(\alpha, \beta) - \Delta(\alpha, \beta)|\colon (\alpha, \beta)^T \in \Theta\} \to 0$. The proof of Theorem 3.2 is complete.

In order to study properties of $(\hat{\alpha}, \hat{\beta})$, we again recall that $N = \sum_{j=1}^{K} n_j$, $\pi_{j,N} = \frac{m_j}{n_j}$ and $r_{j,N} = \frac{n_j}{N}$ (previously π_j and r_j, respectively) for $1 \leq j \leq K$. Let $\boldsymbol{\pi}_N$ and $\boldsymbol{r_N}$ be the K dimensional vectors with components $\pi_{j,N}$ and $r_{j,N}$,

respectively.

Theorem 3.3: Consistency

Suppose Θ is compact, F is continuous and strictly increasing on $\mathbb{R}$ and we select subjects from a population for which the critical dose of a drug is a random variable specified by d.f. F with parameter (α_0, β_0). To each of $n_j\,(1 \leq j \leq K)$ subjects, a dose x_j is applied and m_j of these show a response. If $r_{j,N} \to r_j > 0, 1 \leq j \leq K$, as $N \to \infty$, then $T(\boldsymbol{\pi}_N, \mathbf{r}_N) \to^P (\alpha_0, \beta_0)^T$ as $N \to \infty$, i.e. the SCDE is consistent.

Proof: As $N \to \infty$, $\pi_{j,N} \to^P F(\alpha_0 + \beta_0 x_j) = \pi_{j0}$ since for large value of N the sample proportion will be expected to form an approximation to the parent population proportion.

From Theorem 3.2, we have T is continuous at $(\boldsymbol{\pi}, \mathbf{r})$. Let $T(\boldsymbol{\pi_0}, \mathbf{r_0}) = (\alpha_0, \beta_0)^T$, and we want to show that $\lim\limits_{N\to\infty} P(|T(\boldsymbol{\pi}_N, \mathbf{r}_N) - T(\boldsymbol{\pi_0}, \mathbf{r_0})| < \epsilon) = 1$

Use the continuity of T to find $\delta > 0$ such that

$$|(\boldsymbol{\pi}_N, \mathbf{r}_N) - (\boldsymbol{\pi_0}, \mathbf{r_0})| < \delta \Rightarrow |T(\boldsymbol{\pi}_N, \mathbf{r}_N) - T(\boldsymbol{\pi_0}, \mathbf{r_0})| < \varepsilon.$$

Then $P(|(\boldsymbol{\pi}_N, \mathbf{r}_N) - (\boldsymbol{\pi_0}, \mathbf{r_0})| < \delta) \leq P(|T(\boldsymbol{\pi}_N, \mathbf{r}_N) - T(\boldsymbol{\pi_0}, \mathbf{r_0})| < \varepsilon)$. Here we use the fact that if one event implies another, it has a smaller probability. Since the first probability goes to 1 as $N \to \infty$, so $T(\boldsymbol{\pi}_N, \mathbf{r}_N) \to^P (\alpha_0, \beta_0)^T$. The proof of Theorem 3.3 is complete.

3.2.2 Asymptotic Normality

Before giving an asymptotic expansion for our SCDE functional T, we derive the 'score' and the 'information matrix' from (3.1) first.

Let $\pi_{1,j} = \pi_j,\ \pi_{2,j} = 1 - \pi_{1,j},\ q_{1,j} = F_j(y),\ q_{2,j} = 1 - F_j(y),\ \mathbf{Z}_j = (1, x_j)^T$

and $\Delta(y) = \sum_{j=1}^{K} \Delta_j(y)$, where $\Delta_j(y) = \sum_{i=1}^{2} \{\frac{[\pi_{i,j} - q_{i,j}]^2}{[\pi_{i,j} + q_{i,j}]^2}\}$. Let

$$G_j(y) = \frac{\partial \Delta_j(y)}{\partial y} = -4 \sum_{i=1}^{2} \{\frac{\pi_{i,j}(\pi_{i,j} - q_{i,j})}{(\pi_{i,j} + q_{i,j})^3}\} \frac{\partial q_{i,j}}{\partial y},\ j = 1, ..., K,$$

$$\Sigma = \sum_{j=1}^{K} \mathbf{Z}_j \mathbf{Z}_j^T G_j'(\alpha_0 + \beta_0 x) \text{ and } \boldsymbol{\lambda}(\alpha_0, \beta_0, \boldsymbol{\pi}, \mathbf{r}) = \sum_{j=1}^{K} \mathbf{Z}_j G_j(\alpha_0 + \beta_0 x_j).$$

Theorem 3.4: Suppose Θ is compact, F is a continuous, strictly increasing and thrice differentiable function on $\mathbb{R}$ with derivatives f,f' and f'' on $C = \{\alpha + \beta x_j : (\alpha, \beta)^T \in \Theta, 1 \leq j \leq K\}$ and $F(C) \subseteq [\delta, 1 - \delta]$ for some positive number δ. Suppose $(\boldsymbol{\pi}, \mathbf{r}) \in G_K$ be such that $T(\boldsymbol{\pi_0}, \mathbf{r_0}) = (\alpha_0, \beta_0)^T$ is unique and let $(\boldsymbol{\pi}_N, \mathbf{r}_N) \to (\boldsymbol{\pi_0}, \mathbf{r_0})$ as $N \to \infty$. If Σ is non-singular then we have

$T(\boldsymbol{\pi}_N, \mathbf{r}_N) - T(\boldsymbol{\pi_0}, \mathbf{r_0}) = -\Sigma^{-1} \boldsymbol{\lambda}(\alpha_0, \beta_0, \boldsymbol{\pi}_N, \mathbf{r}_N)(1 + o_p(1))$, as $N \to \infty$.

Proof: Let $T(\boldsymbol{\pi}_N, \mathbf{r}_N) = (\hat{\alpha}_N, \hat{\beta}_N)^T$. Differentiability of F implies that $(\hat{\alpha}_N, \hat{\beta}_N)^T$ satisfies

$$\mathbf{0} = \boldsymbol{\lambda}(\hat{\alpha}_N, \hat{\beta}_N, \boldsymbol{\pi}_N, \mathbf{r}_N) = \sum_{j=1}^{K} \mathbf{Z}_j G_{j,N}(\hat{\alpha}_N + \hat{\beta}_N x_j), \tag{3.6}$$

where $G_{j,N}$ is defined as G_j with $\pi_{j,N}$ replaced by π_j. Expanding $G_{j,N}(\hat{\alpha}_N + \hat{\beta}_N x_j)$ at $\alpha_0 + \beta_0 x_j$ by Taylor series we obtain

$$G_{j,N}(\hat{\alpha}_N + \hat{\beta}_N x_j) = G_{j,N}(\alpha_0 + \beta_0 x_j) + \mathbf{Z}_j^T \boldsymbol{\gamma}_N G_{j,N}'(\alpha_0 + \beta_0 x_j) + \frac{1}{2}(\mathbf{Z}_j^T \boldsymbol{\gamma}_N)^2 G_{j,N}''(\kappa_j), \tag{3.7}$$

where κ_j is between $\hat{\alpha}_N + \hat{\beta}_N x_j$ and $\alpha_0 + \beta_0 x_j$, $\mathbf{Z}_j = (1, x_j)^T$ and $\boldsymbol{\gamma}_N = (\hat{\alpha}_N - \alpha_0, \hat{\beta}_N - \beta_0)^T$.

Since F has bounded derivatives and is bounded in the interval $[0,1]$, and $G''_{j,N}$ is continuous in a closed interval, then $G''_{j,N}$ is bounded. Also, $\pi_{j,N} \to \pi_j$ implies that $G'_{j,N}(y) \to G'_j(y)$ uniformly in y, then (3.7) becomes

$$G_{j,N}(\hat{\alpha}_N + \hat{\beta}_N x_j) = G_{j,N}(\alpha_0 + \beta_0 x_j) + \mathbf{Z}_j^T \boldsymbol{\gamma}_N G'_j(\alpha_0 + \beta_0 x_j) + o(\mathbf{Z}_j^T \boldsymbol{\gamma}_N), \tag{3.8}$$

since Theorem 3.2 implies $\boldsymbol{\gamma}_N \to 0$ as $N \to \infty$. Substituting (3.8) in (3.6) we obtain

$$\begin{aligned}\mathbf{0} =& \sum_{j=1}^{K} \mathbf{Z}_j G_{j,N}(\alpha_0 + \beta_0 x_j) + [\sum_{j=1}^{K} \mathbf{Z}_j \mathbf{Z}_j^T \{G'_j(\alpha_0 + \beta_0 x_j) + o(1)\}] \boldsymbol{\gamma}_N \\ =& \boldsymbol{\lambda}(\alpha_0, \beta_0, \boldsymbol{\pi}_N, \mathbf{r}_N) + \{\Sigma + W_N\} \boldsymbol{\gamma}_N,\end{aligned}$$

since $r_{j,N} \to r_j$ as $N \to \infty$. Here W_N is a 2×2 matrix whose component are $o(1)$ as $N \to \infty$. Since $\{\Sigma + W_N\}$ will be non-singular for N sufficiently large, $\Sigma^{-1} = (\Sigma + W_N)^{-1}$ as $N \to \infty$. Hence we have

$$\boldsymbol{\gamma}_N = T(\boldsymbol{\pi}_N, \mathbf{r}_N) - T(\boldsymbol{\pi}_\mathbf{0}, \mathbf{r}_\mathbf{0}) = -\Sigma^{-1} \boldsymbol{\lambda}(\alpha_0, \beta_0, \boldsymbol{\pi}_N, \mathbf{r}_N)[1 + o_p(1)]. \tag{3.9}$$

The proof of Theorem 3.4 is complete.

Special case: If $\pi_{i,j} = q_{i,j} = F_j(y)$, then $\boldsymbol{\gamma}_N = \mathbf{0}$.

Notes: 1. $G_j(y) = -4\sum_{i=1}^{2}\{\frac{\pi_{i,j}(\pi_{i,j}-q_{i,j})}{(\pi_{i,j}+q_{i,j})^3}\}\frac{\partial q_{i,j}}{\partial y}$, so that $G_j(y) = 0$ when $\pi_{i,j} = q_{i,j} = F_j(y)$.

2. $G'_j(y) = 8\sum_{i=1}^{2}\{\frac{\pi_{i,j}(2\pi_{i,j}-q_{i,j})}{(\pi_{i,j}+q_{i,j})^4}\}(\frac{\partial q_{i,j}}{\partial y})^2 - 4\sum_{i=1}^{2}\{\frac{\pi_{i,j}(\pi_{i,j}-q_{i,j})}{(\pi_{i,j}+q_{i,j})^3}\}\frac{\partial^2 q_{i,j}}{\partial y^2}$, so that Σ is very complicated in general. However, when $\pi_{i,j} = q_{i,j}$ we have $G'_j(y) = \sum_{i=1}^{2}(\frac{1}{q_{i,j}}\frac{\partial q_{i,j}}{\partial y})^2 = \sum_{i=1}^{2}(\frac{\partial \log q_{i,j}}{\partial y})^2$, and we can get the following theorem.

Theorem 3.5: Asymptotic Normality

Suppose the conditions in Theorem 3.3 holds and that the expansion in Theorem 3.4 holds for $T(\boldsymbol{\pi}_N, \mathbf{r}_N)$. Let $T(\boldsymbol{\pi}_\mathbf{0}, \mathbf{r}_\mathbf{0}) = (\alpha_0, \beta_0)^T$ and $\Sigma_S = \Sigma^{-1}\Sigma_y\Sigma^{-1}$. Then, we have

$\sqrt{N}\{T(\boldsymbol{\pi}_N, \mathbf{r}_N) - T(\boldsymbol{\pi}_\mathbf{0}, \mathbf{r}_\mathbf{0})\} \to^D N(\mathbf{0}, \Sigma_S)$, as $N \to \infty$.

Proof: Note that

$$\begin{aligned}\Sigma =& \sum_{j=1}^{K}\mathbf{Z}_j\mathbf{Z}_j^T G'_j(\alpha_0+\beta_0 x)\\ =& \sum_{j=1}^{K}\mathbf{Z}_j\mathbf{Z}_j^T\sum_{i=1}^{2}(\frac{\partial \log q_{i,j}}{\partial y})^2\\ =& \sum_{j=1}^{K}\begin{pmatrix} 1 & x_j \\ x_j & x_j^2 \end{pmatrix} A_j, \end{aligned} \tag{3.10}$$

where $A_j = f^2(\alpha+\beta x_j)(\frac{1}{F^2(\alpha+\beta x_j)} + \frac{1}{(1-F(\alpha+\beta x_j))^2})$, $j = 1, \ldots, K$.

Note Σ is singular when $|\Sigma| = (\sum_{j=1}^{K}A_j)(\sum_{j=1}^{K}x_j^2 A_j) - (\sum_{j=1}^{K}x_j Aj)^2 = 0$, i.e. Σ is singular only when $K = 1$ or all x_j equals to each other. Except for special cases, Σ is nonsingular and $\Sigma^{-1} = \sum_{j=1}^{K}\begin{pmatrix} \frac{x_j^2 A_j}{|\Sigma|} & -\frac{x_j A_j}{|\Sigma|} \\ -\frac{x_j A_j}{|\Sigma|} & \frac{A_j}{|\Sigma|} \end{pmatrix}$.

From (3.9), as $N \to \infty$,

$$\begin{aligned}\boldsymbol{\gamma}_n &= T(\boldsymbol{\pi}_N, \mathbf{r}_N) - T(\boldsymbol{\pi_0}, \mathbf{r_0}) = -\Sigma^{-1}\boldsymbol{\lambda}(\alpha_0, \beta_0, \boldsymbol{\pi}_N, \mathbf{r}_N) \\ &= -\Sigma^{-1}\sum_{j=1}^{K}\mathbf{Z}_j G_{jN}(\alpha_0 + \beta_0 x_j). \end{aligned} \tag{3.11}$$

Let $y_0 = \alpha_0 + \beta_0 x$,

$G_{jN}(y_0) = -4\sum_{i=1}^{2}\{\frac{\pi_{i,jN}(\pi_{i,jN} - q_{i,j})}{(\pi_{i,jN} + q_{i,j})^3}\}\frac{\partial q_{i,j}}{\partial y} = -4f_j(y_0)(\pi_{jN} - F_j(y_0))\Lambda$,

where $\Lambda = \frac{\pi_{jN}}{[\pi_{jN}+F_j(y_0)]^3} + \frac{1-\pi_{jN}}{[2-\pi_{jN}-F_j(y_0)]^3}$.

As $N \to \infty$, $\pi_{jN} \to^P F_j(y_0)$, $1 \le j \le K$ and Λ can be considered as a constant.

At each dose level x_j, $\pi_j = \frac{m_j}{n_j}$ and $m_j \sim Binomial(n_j, F_j(y_0))$.

Then since $\sqrt{Nr_j} = \sqrt{n_j}$, $\sqrt{n_j}(\pi_{jN} - F_j(y_0)) \to^{\mathcal{D}} N(0, F_j(y_0)(1 - F_j(y_0))$.

Now let $\Sigma_y = 16\sum_{j=1}^{K} f_j^2(y_0)F_j(y_0)(1 - F_j(y_0))\Lambda^2 Z_j Z_j^T$,

then we have $\sqrt{N}\{T(\boldsymbol{\pi}_N, \mathbf{r}_N) - T(\boldsymbol{\pi_0}, \mathbf{r_0})\} \to^D N(\mathbf{0}, \Sigma_S)$, as $N \to \infty$. The proof of Theorem 3.5 is complete.

3.3 Simulation Study

3.3.1 Logistic Model

It is difficult to establish theoretical results on the robustness of the SCDE because it has a complex form. Thus to explore the robustness properties of SCDE we relied upon on Monte Carlo methods again. To compare the results with MHDE, MLE and WLSE, the same simulation study is conducted with the following four models:

Model I: $F(y) = \frac{e^y}{1+e^y} = L(y)$,

Model II: $F(y) = 0.9L(y) + 0.1L(2y)$,

Model III: $F(y) = 0.9L(y) + 0.1L(0.5y)$,

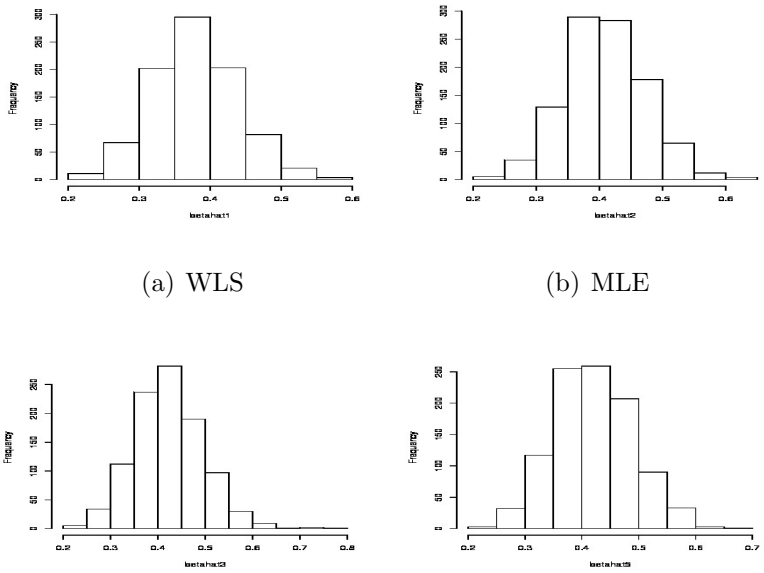
Model IV: $F(y) = 0.9L(y) + 0.1$.

Table 3.1: Symmetric χ^2 Distance Method

Model	$\hat{\alpha}_n$	$\hat{\beta}_n$	$V(\hat{\alpha}_n)$	$V(\hat{\beta}_n)$	$Cov(\hat{\alpha}_n, \hat{\beta}_n)$	$MSE(\hat{\alpha}_n)$	$MSE(\hat{\beta}_n)$
Model I	-2.129	0.4246	0.1754	0.0047	-0.0259	0.1919	0.0053
Model II	-2.1155	0.4224	0.1617	0.0046	-0.0244	0.1749	0.0051
Model III	-2.1436	0.4275	0.1910	0.0055	-0.0293	0.2114	0.0062
Model IV	-2.1046	0.42	0.2261	0.0060	-0.0335	0.2368	0.0064

Simulation results are outlined in Tables 3.1. Again, the results are based on 1000 replications. As in Chapter 2 tables, $V(\hat{\alpha}_n)$ and $V(\hat{\beta}_n)$ are used to denote the estimated variance of $\hat{\alpha}_n$ and $\hat{\beta}_n$. Similarly, $Cov(\hat{\alpha}_n, \hat{\beta}_n)$ stands for the estimated covariance between $\hat{\alpha}_n$ and $\hat{\beta}_n$. Also, $MSE(\hat{\alpha}_n)$ and $MSE(\hat{\beta}_n)$ denote estimated the mean squared errors of $\hat{\alpha}_n$ and $\hat{\beta}_n$. Our results show that MHDE and SCDE are both much more robust to model mispecification than WLSE and MLE. Thus, we focus on comparing the two robust methods, MHDE and SCDE. In all cases considered, values of MHDE are more variable than those of SCDE. SCDE has smaller variances for Models I, II and IV. If we consider only the biases of the two estimators, we see that the SCDE is less biased if the postulated model is correct, i.e. Model I. Further we see that the MHDE shows less bias under Models II and III, while the SCDE shows less bias under Model IV. These results suggest that the MHDE has some protection if the true distribution has small contaminations (10% shorter or longer tails). The results for Model IV suggest that the SCDE might be the best if we wish to protect against the possibility of subjects showing a response

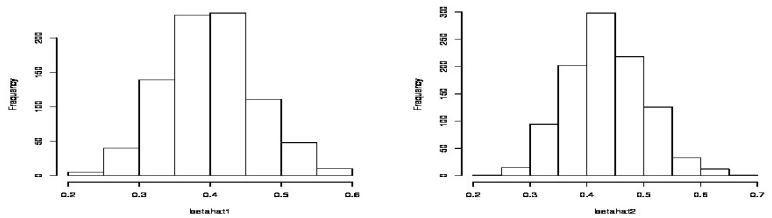
regardless of what dose they receive.



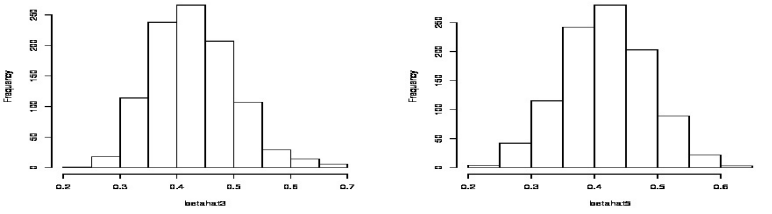
(a) WLS (b) MLE

(c) MHDE (d) SCDE

Figure 3.1: Histograms of 1000 $\hat{\beta}_n$s of Model I by four methods.

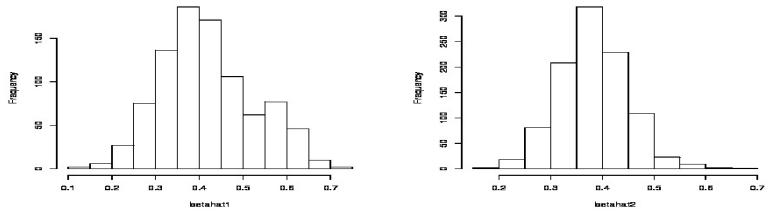


(a) WLS (b) MLE

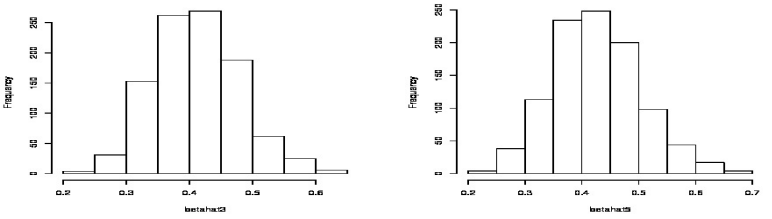


(c) MHDE (d) SCDE

Figure 3.2: Histograms of 1000 $\hat{\beta}_n$s of Model II by four methods.

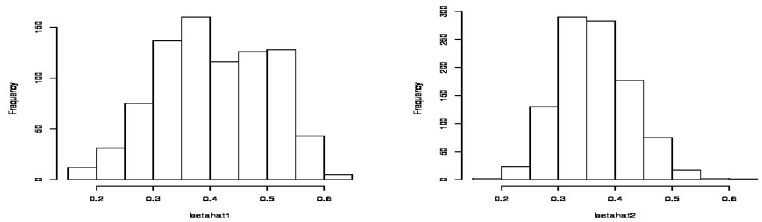


(a) WLS (b) MLE

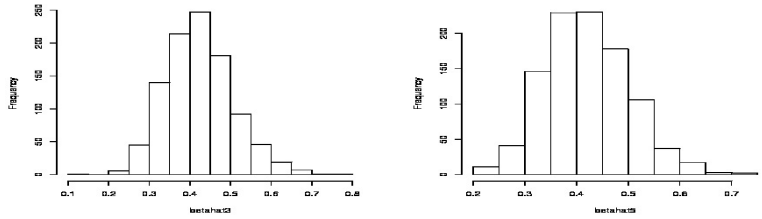


(c) MHDE (d) SCDE

Figure 3.3: Histograms of 1000 $\hat{\beta}_n$s of Model III by four methods.



(a) WLS (b) MLE



(c) MHDE (d) SCDE

Figure 3.4: Histograms of 1000 $\hat{\beta}_n$s of Model IV by four methods.

From Figures 3.1-3.4, we observe that two robust methods, MHDE and SCDE, have less skewed histograms for the estimator $\hat{\beta}$. One can clearly see that the histogram of SCDE is centered at the true parameter value and has the smallest deviation.

We considered six more models based on the Model II by increasing the contamination rate from 10% to 70%. We let the distribution with a short tail as the contaminated one.

Model V: $F(y) = 0.8L(y) + 0.2L(2y)$,

Model VI: $F(y) = 0.7L(y) + 0.3L(2y)$,

Model VII: $F(y) = 0.6L(y) + 0.4L(2y)$,

Model VIII: $F(y) = 0.5L(y) + 0.5L(2y)$,

Model IX: $F(y) = 0.4L(y) + 0.6L(2y)$,

Model X: $F(y) = 0.3L(y) + 0.7L(2y)$,

Table 3.2: WLS Method for the contaminated Logit models

$Model$	$\hat{\alpha}_n$	$\hat{\beta}_n$	$V(\hat{\alpha}_n)$	$V(\hat{\beta}_n)$	$Cov(\hat{\alpha}_n, \hat{\beta}_n)$	$MSE(\hat{\alpha}_n)$	$MSE(\hat{\beta}_n)$
Model II	−2.0093	0.4005	0.1241	0.0037	−0.0193	0.124	0.0037
Model V	−2.1237	0.4238	0.1298	0.0036	−0.0194	0.1449	0.0042
Model VI	−2.2606	0.4488	0.1451	0.0042	−0.0221	0.2128	0.0065
Model VII	−2.4132	0.4789	0.1424	0.0041	−0.0218	0.313	0.0103
Model VIII	−2.5098	0.4989	0.1496	0.0042	−0.0228	0.4092	0.014
Model IX	−2.662	0.5282	0.1600	0.0049	−0.0257	0.5979	0.0214
Model X	−2.833	0.5632	0.1541	0.0050	−0.0255	0.8475	0.0317

Table 3.3: MLE Method for the contaminated Logit models

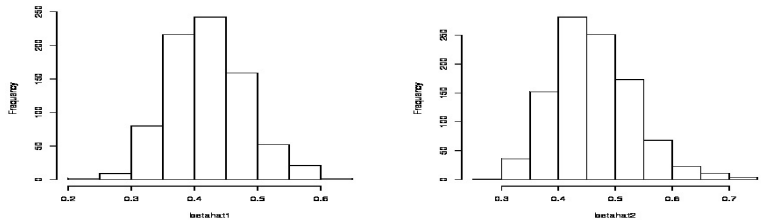
$Model$	$\hat{\alpha}_n$	$\hat{\beta}_n$	$V(\hat{\alpha}_n)$	$V(\hat{\beta}_n)$	$Cov(\hat{\alpha}_n, \hat{\beta}_n)$	$MSE(\hat{\alpha}_n)$	$MSE(\hat{\beta}_n)$
Model II	−2.1616	0.4347	0.1426	0.0044	−0.0226	0.1686	0.0056
Model V	−2.3165	0.4616	0.1690	0.0049	−0.0264	0.269	0.0087
Model VI	−2.4752	0.4943	0.1682	0.0052	−0.0267	0.3939	0.0141
Model VII	−2.6595	0.5311	0.1842	0.0059	−0.0302	0.619	0.0231
Model VIII	−2.8088	0.5611	0.1970	0.0063	−0.0324	0.8509	0.0323
Model IX	−3.0334	0.6043	0.2222	0.0070	−0.0361	1.2898	0.0487
Model X	−3.2183	0.6429	0.2551	0.0082	−0.0425	1.739	0.0672

Table 3.4: Minimum Hellinger Distance Method for the contaminated Logit models

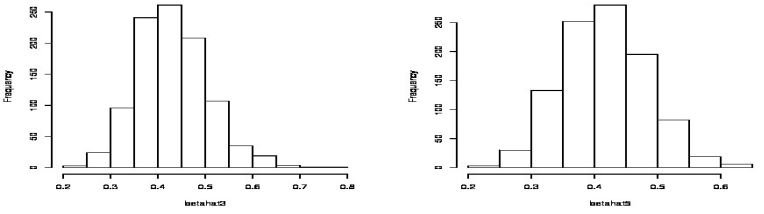
$Model$	$\hat{\alpha}_n$	$\hat{\beta}_n$	$V(\hat{\alpha}_n)$	$V(\hat{\beta}_n)$	$Cov(\hat{\alpha}_n, \hat{\beta}_n)$	$MSE(\hat{\alpha}_n)$	$MSE(\hat{\beta}_n)$
Model II	-2.1053	0.4214	0.1731	0.0054	-0.028	0.1841	0.0058
Model V	-2.121	0.4256	0.1832	0.0059	-0.0302	0.1977	0.0065
Model VI	-2.162	0.4319	0.1675	0.0058	-0.0288	0.1936	0.0068
Model VII	-2.1661	0.4338	0.1698	0.0053	-0.0278	0.1972	0.0065
Model VIII	-2.1726	0.4365	0.1786	0.0059	-0.0304	0.2082	0.0072
Model IX	-2.1841	0.439	0.1611	0.0058	-0.0288	0.1949	0.0073
Model X	-2.2203	0.4443	0.1695	0.0059	-0.0300	0.2179	0.0079

Table 3.5: Symmetric χ^2 Distance Method for the contaminated Logit models

$Model$	$\hat{\alpha}_n$	$\hat{\beta}_n$	$V(\hat{\alpha}_n)$	$V(\hat{\beta}_n)$	$Cov(\hat{\alpha}_n, \hat{\beta}_n)$	$MSE(\hat{\alpha}_n)$	$MSE(\hat{\beta}_n)$
Model II	-2.1155	0.4224	0.1617	0.0046	-0.0244	0.1749	0.0051
Model V	-2.1109	0.4211	0.1604	0.0045	-0.0245	0.1726	0.0049
Model VI	-2.0898	0.4156	0.1550	0.0045	-0.0241	0.1629	0.0048
Model VII	-2.0833	0.4137	0.1312	0.0039	-0.0206	0.138	0.004
Model VIII	-2.0567	0.4082	0.1151	0.0034	-0.0179	0.1182	0.0034
Model IX	-2.0168	0.4007	0.1065	0.0033	-0.0172	0.1067	0.0033
Model X	-1.9816	0.3932	0.0962	0.0029	-0.0152	0.0964	0.0029

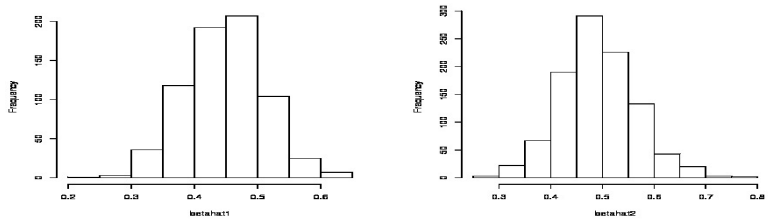


(a) WLS (b) MLE

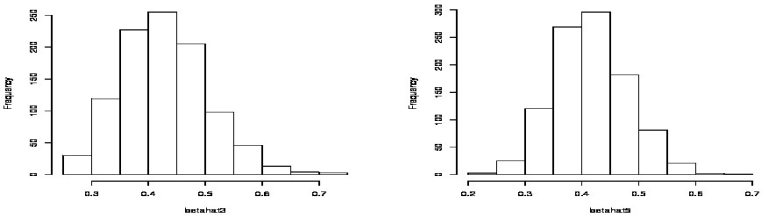


(c) MHDE (d) SCDE

Figure 3.5: Histograms of 1000 $\hat{\beta}_n$s of Model V by four methods.

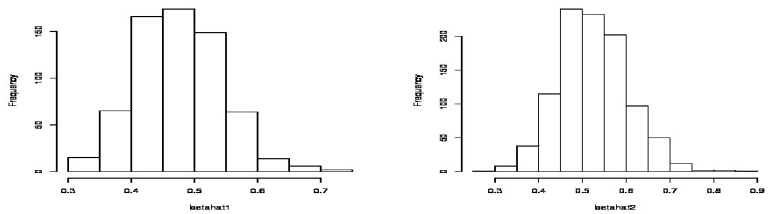


(a) WLS (b) MLE

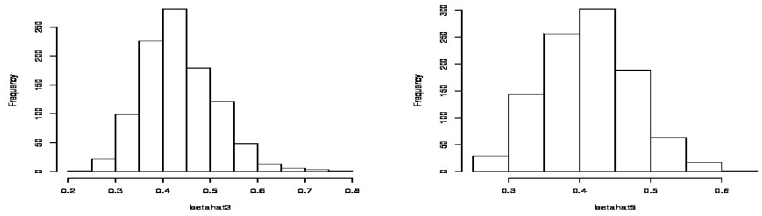


(c) MHDE (d) SCDE

Figure 3.6: Histograms of 1000 $\hat{\beta}_n$s of Model VI by four methods.

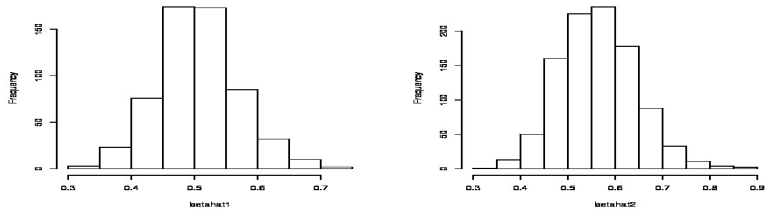


(a) WLS (b) MLE

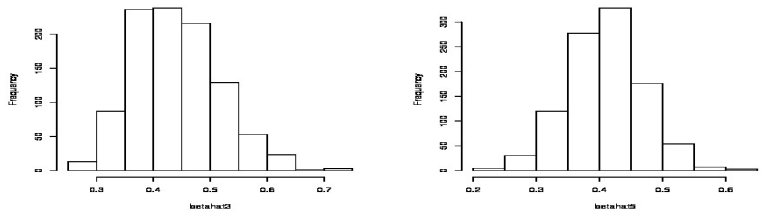


(c) MHDE (d) SCDE

Figure 3.7: Histograms of 1000 $\hat{\beta}_n$s of Model VII by four methods.

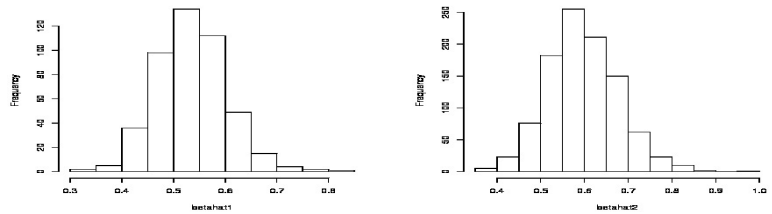


(a) WLS (b) MLE

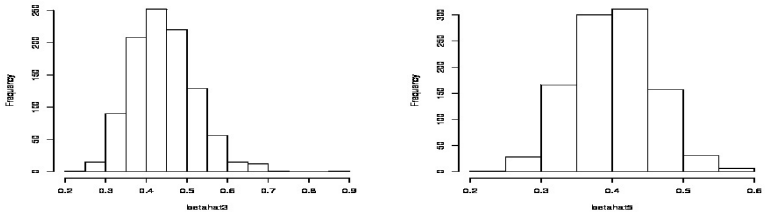


(c) MHDE (d) SCDE

Figure 3.8: Histograms of 1000 $\hat{\beta}_n$s of Model VIII by four methods.

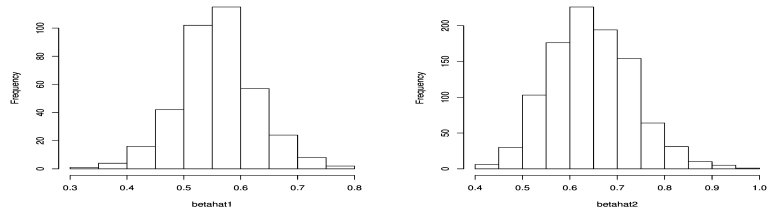


(a) WLS (b) MLE

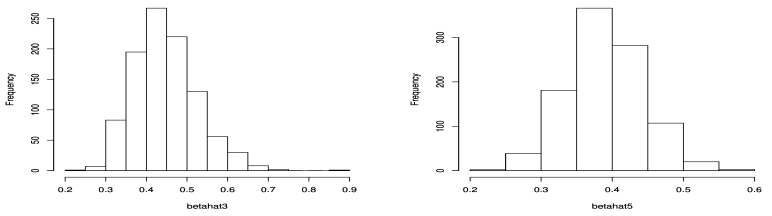


(c) MHDE (d) SCDE

Figure 3.9: Histograms of 1000 $\hat{\beta}_n$s of Model IX by four methods.

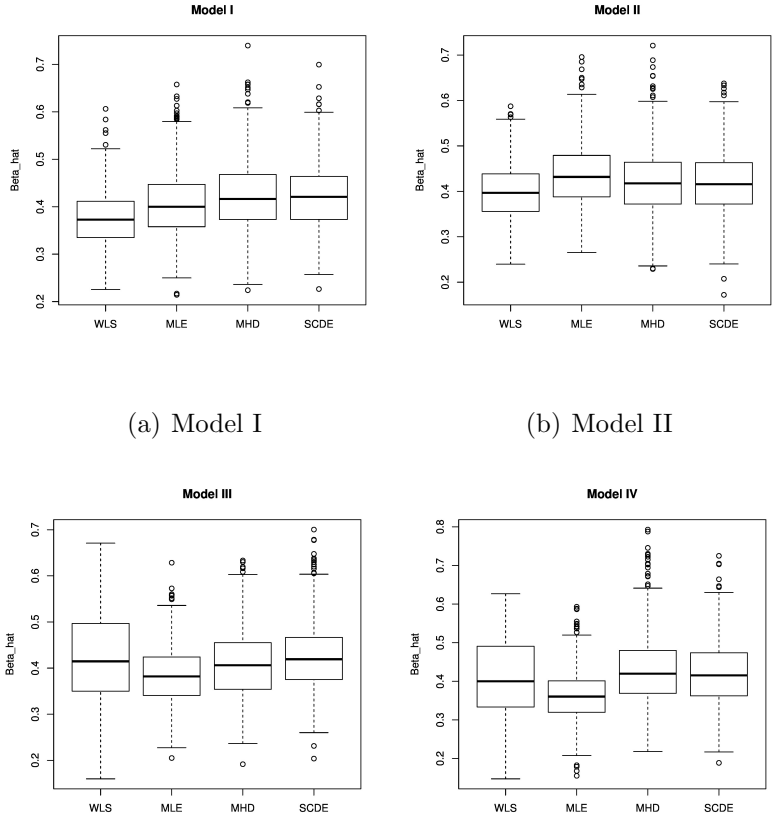


(a) WLS (b) MLE



(c) MHDE (d) SCDE

Figure 3.10: Histograms of 1000 $\hat{\beta}_n$s of Model X by four methods.



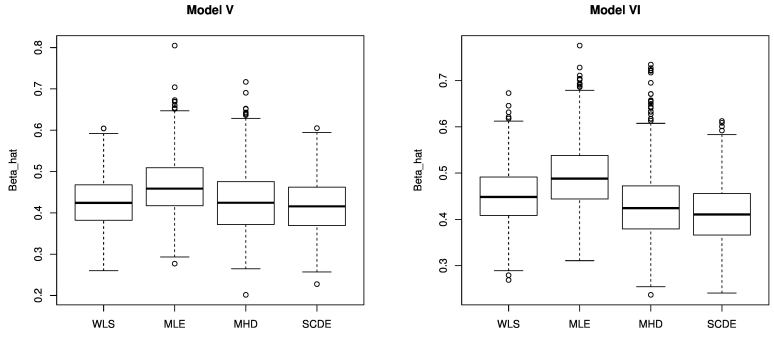
(a) Model I

(b) Model II

(c) Model III

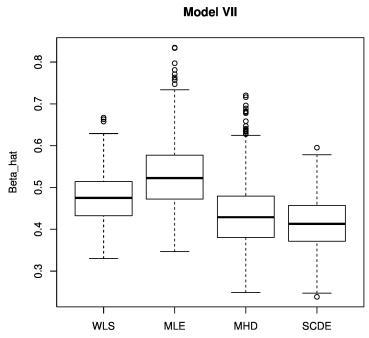
(d) Model IV

Figure 3.11: Boxplot of 1000 $\hat{\beta}_n$'s of four methods for model I-IV

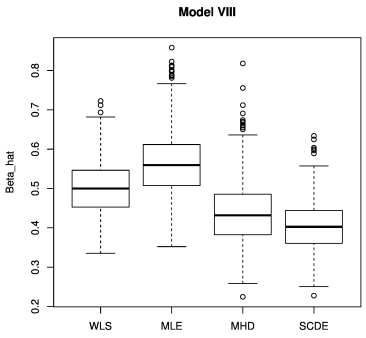


(a) Model V

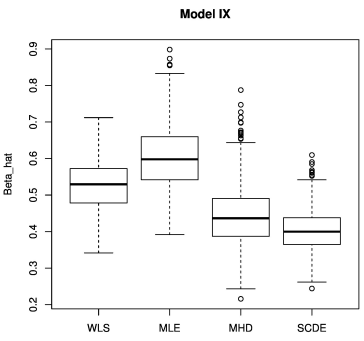
(b) Model VI



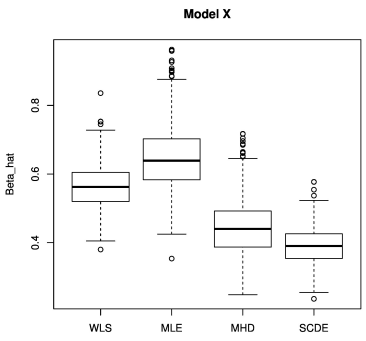
(c) Model VII



(d) Model VIII



(e) Model IX



(f) Model X

Figure 3.12: Boxplot of 1000 $\hat{\beta}_n$'s of four methods for contaminated Logit models

Simulated results are presented in Tables 3.2 to 3.5. These results show that there are monotonic trends (increasing or decreasing) of the values of $E(\hat{\alpha}_n)$ and $E(\hat{\beta}_n)$ as the contamination rate increases. For the non-robust methods, WLSE and MLE method, we see large fluctuations when we increase the contamination rate. When we compared the two robust methods MHDE and SCDE only, we observed following: Although the MHDE has some protection if the true distribution has 10% shorter tails contamination, it lost this advantage immediately when we increase the shorter tails contamination to 20% or higher. The results suggest that the SCDE might be the best if we wish to protect against the true distribution mixed with a higher percentage of shorter tails contamination. Histograms and boxplots of the range of $E(\hat{\beta})$ both show that SCDE is the best one in robustness among the four method.

Huber (2009) used the Asymptotic Relative Efficiency (ARE) to compare two estimators. For $F(y) = (1-\epsilon)L(y) + \epsilon L(2y)$ in this simulation, we use ARE of MHDE relative to MLE defined as

$$ARE(\epsilon) = lim_n \quad \frac{var(\hat{\beta}_2)}{var(\hat{\beta}_3)}.$$

The results are summarized in Table 3.6.

Table 3.6: ARE of MHDE relative to MLE

ϵ	0.1	0.2	0.3	0.4	0.5	0.6	0.7
ARE	0.8148	0.8305	0.8966	1.1132	1.0678	1.2069	1.3898

From Table 3.6, we can see a turning point between 0.3 and 0.4. After this turning point, we should certainly prefer MHDE to MLE. Using the same technique for comparing SCDE and MLE, we can find the turning point is between 0.1 and 0.2.

3.3.2 Probit Model

We continue our discussion of the robustness of SCDE method by using the Probit Models. Specifically, we studied following four models:

Model I: $F(y) = \Phi(y)$, where $\Phi(y)$ is the CDF of standard normal distribution.

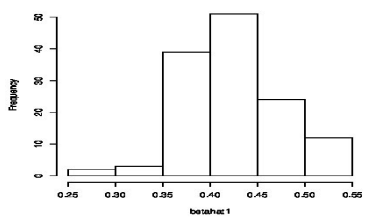
Model II: $F(y) = 0.9\Phi(y) + 0.1\Phi(2y)$,

Model III: $F(y) = 0.9\Phi(y) + 0.1\Phi(0.5y)$,

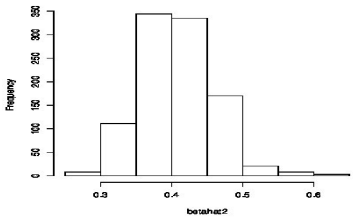
Model IV: $F(y) = 0.9\Phi(y) + 0.1$.

Table 3.7: Symmetric χ^2 Distance Method (Probit Model)

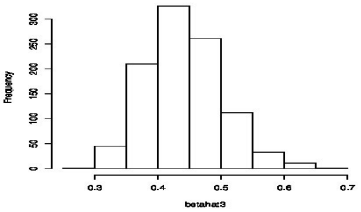
Model	$\hat{\alpha}_n$	$\hat{\beta}_n$	$V(\hat{\alpha}_n)$	$V(\hat{\beta}_n)$	$Cov(\hat{\alpha}_n, \hat{\beta}_n)$	$MSE(\hat{\alpha}_n)$	$MSE(\hat{\beta}_n)$
Model I	-1.883	0.3743	0.0629	0.0019	-0.0096	0.0765	0.0026
Model II	-1.8805	0.372	0.0611	0.0019	-0.0097	0.0753	0.0027
Model III	-2.1424	0.4236	0.0861	0.0027	-0.0136	0.1062	0.0033
Model IV	-2.587	0.4601	0.2024	0.0054	-0.0312	0.5468	0.009



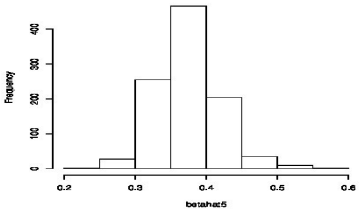
(a) WLS



(b) MLE

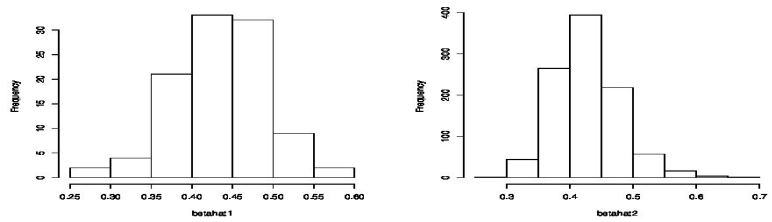


(c) MHDE

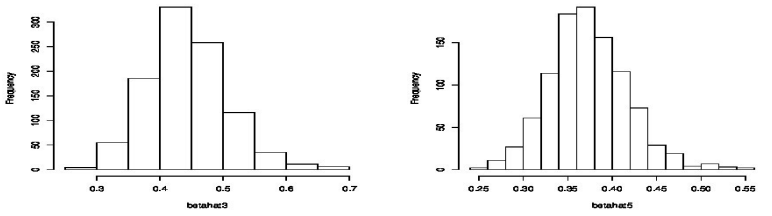


(d) SCDE

Figure 3.13: Histograms of 1000 $\hat{\beta}_n$s of Model I by four methods.

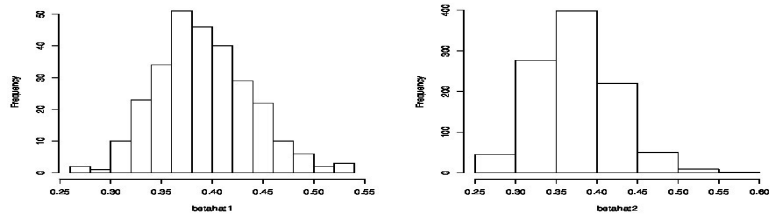


(a) WLS (b) MLE

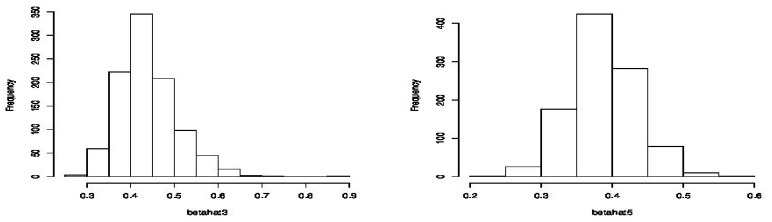


(c) MHDE (d) SCDE

Figure 3.14: Histograms of 1000 $\hat{\beta}_n$s of Model II by four methods.

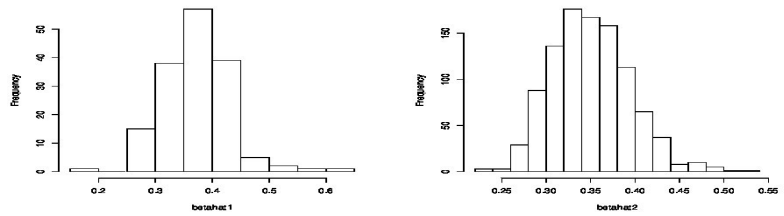


(a) WLS (b) MLE

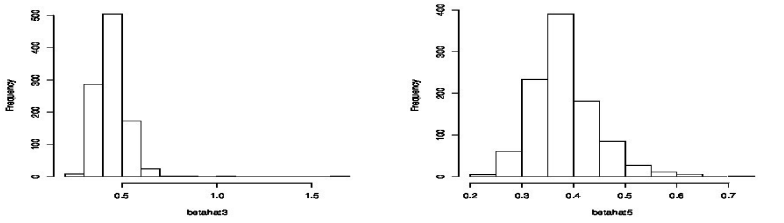


(c) MHDE (d) SCDE

Figure 3.15: Histograms of 1000 $\hat{\beta}_n$s of Model III by four methods.



(a) WLS (b) MLE



(c) MHDE (d) SCDE

Figure 3.16: Histograms of 1000 $\hat{\beta}_n$s of Model IV by four methods.

Again we examined the behavior of the four methods as in the previous section for Models I to IV. Simulation results are displayed in Tables 3.6 to 3.10. These results also show that MHDE and SCDE are both much more robust to model variability than WLSE and MLE. When we compare MHDE and SCDE, we observed that in all four models the MHDE is less variable than the SCDE. However, SCDE has smaller variance for Models I, II and III. If we consider only the biases of these two estimators, we see that the SCDE is less biased if the postulated model is correct, i.e. Model I. Further we see that the SCDE shows less bias under Models II and III, while the MHDE shows less bias under Model IV. These results suggest that the SCDE has some protection if the true distribution has 10% shorter or longer tails contaminations. The results for Model IV suggest that the MHDE might be better if we wish to protect against the possibility of subjects showing a response regardless of

what dose they receive.

From Figures 3.13-3.16, we see that two robust methods, MHDE and SCDE, have less skewed histograms for the estimator $\hat{\beta}$. We also noticed that the histogram of SCDE is centered at the true parameter value and has the smallest variance.

We considered six models based on Model II by increasing the contamination rate from 10% to 70%. We let the distribution with a short tail as the contaminated one.

Model V: $F(y) = 0.8\Phi(y) + 0.2\Phi(2y)$,

Model VI: $F(y) = 0.7\Phi(y) + 0.3\Phi(2y)$,

Model VII: $F(y) = 0.6\Phi(y) + 0.4\Phi(2y)$,

Model VIII: $F(y) = 0.5\Phi(y) + 0.5\Phi(2y)$,

Model IX: $F(y) = 0.4\Phi(y) + 0.6\Phi(2y)$,

Model X: $F(y) = 0.3\Phi(y) + 0.7\Phi(2y)$,

Table 3.8: WLS Method for the contaminated Probit models

Model	$\hat{\alpha}_n$	$\hat{\beta}_n$	$V(\hat{\alpha}_n)$	$V(\hat{\beta}_n)$	$Cov(\hat{\alpha}_n, \hat{\beta}_n)$	$MSE(\hat{\alpha}_n)$	$MSE(\hat{\beta}_n)$
Model II	−2.2084	0.4406	0.0952	0.0026	−0.0134	0.1374	0.0042
Model V	−2.2963	0.4577	0.0841	0.0025	−0.0131	0.1708	0.0058
Model VI	−2.3618	0.4769	0.0699	0.0027	−0.0106	0.1994	0.0086
Model VII	−2.4936	0.4988	0.1444	0.0036	−0.0217	0.3834	0.0132
Model VIII	−2.5648	0.5222	0.1996	0.0061	−0.0316	0.5069	0.0207
Model IX	−2.8476	0.567	0.1046	0.0035	−0.0160	0.8022	0.0307
Model X	−2.8108	0.5951	0.0276	0.002	−0.007	0.6804	0.0397

Table 3.9: MLE Method for the contaminated Probit models

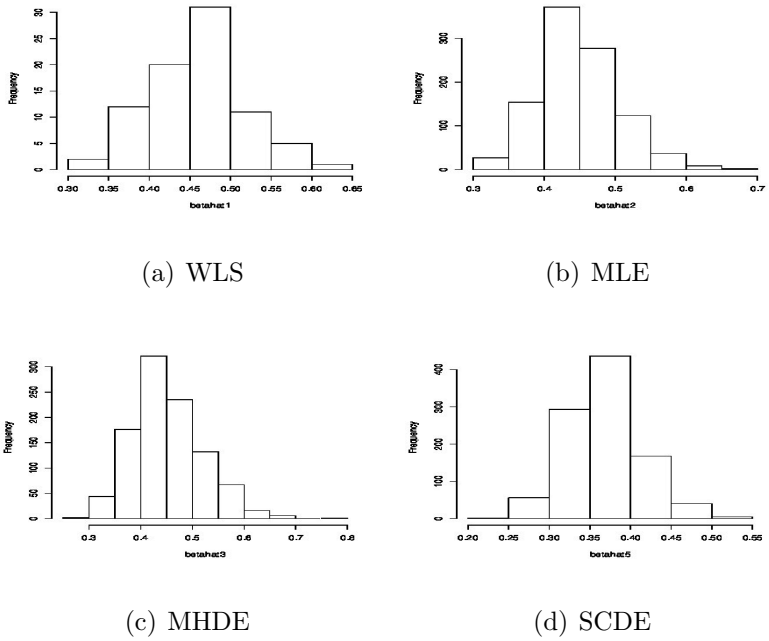
$Model$	$\hat{\alpha}_n$	$\hat{\beta}_n$	$V(\hat{\alpha}_n)$	$V(\hat{\beta}_n)$	$Cov(\hat{\alpha}_n, \hat{\beta}_n)$	$MSE(\hat{\alpha}_n)$	$MSE(\hat{\beta}_n)$
Model II	−2.1477	0.4296	0.0859	0.0028	−0.0142	0.1076	0.0037
Model V	−2.2373	0.4482	0.0929	0.0030	−0.0155	0.1491	0.0054
Model VI	−2.346	0.4683	0.1094	0.0037	−0.0188	0.229	0.0083
Model VII	−2.4944	0.4975	0.1179	0.0040	−0.0204	0.3622	0.0135
Model VIII	−2.6491	0.5292	0.1573	0.0053	−0.0274	0.5784	0.022
Model IX	−2.8211	0.5629	0.1697	0.0062	−0.0309	0.8437	0.0327
Model X	−3.032	0.6053	0.1993	0.0072	−0.0360	1.2641	0.0493

Table 3.10: Minimum Hellinger Distance Method for the contaminated Probit models

$Model$	$\hat{\alpha}_n$	$\hat{\beta}_n$	$V(\hat{\alpha}_n)$	$V(\hat{\beta}_n)$	$Cov(\hat{\alpha}_n, \hat{\beta}_n)$	$MSE(\hat{\alpha}_n)$	$MSE(\hat{\beta}_n)$
Model II	−2.2166	0.4433	0.1263	0.0043	−0.0217	0.1731	0.0061
Model V	−2.2435	0.4489	0.1256	0.0044	−0.0220	0.1848	0.0068
Model VI	−2.2714	0.4557	0.1554	0.0053	−0.0273	0.2289	0.0084
Model VII	−2.2892	0.4599	0.1531	0.0054	−0.0275	0.2366	0.009
Model VIII	−2.3164	0.4646	0.1706	0.0061	−0.0309	0.2705	0.0102
Model IX	−2.3433	0.4696	0.1619	0.0059	−0.0299	0.2796	0.0108
Model X	−2.365	0.4713	0.1680	0.0062	−0.0311	0.301	0.0112

Table 3.11: Symmetric χ^2 Distance Method for the contaminated Probit models

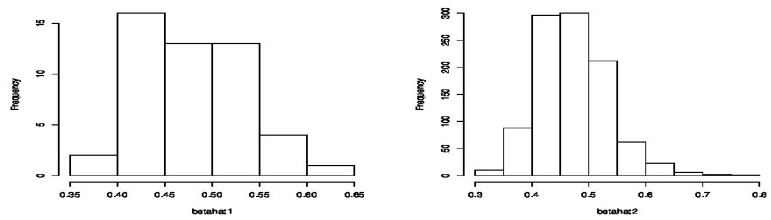
Model	$\hat{\alpha}_n$	$\hat{\beta}_n$	$V(\hat{\alpha}_n)$	$V(\hat{\beta}_n)$	$Cov(\hat{\alpha}_n, \hat{\beta}_n)$	$MSE(\hat{\alpha}_n)$	$MSE(\hat{\beta}_n)$
Model II	−1.8856	0.3741	0.0593	0.0018	−0.0092	0.0724	0.0025
Model V	−1.8588	0.3698	0.0644	0.0021	−0.0106	0.0843	0.003
Model VI	−1.8256	0.3638	0.0709	0.0025	−0.0122	0.1012	0.0038
Model VII	−1.7949	0.3565	0.0703	0.0025	−0.0123	0.1122	0.0044
Model VIII	−1.7737	0.3525	0.0758	0.0027	−0.0134	0.127	0.005
Model IX	−1.7331	0.3453	0.0768	0.0028	−0.0138	0.148	0.0058
Model X	−1.688	0.3367	0.0769	0.0029	−0.0140	0.1742	0.0069



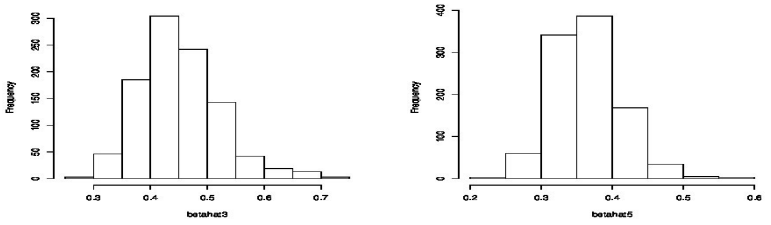
(a) WLS (b) MLE

(c) MHDE (d) SCDE

Figure 3.17: Histograms of 1000 $\hat{\beta}_n$s of Model V by four methods.

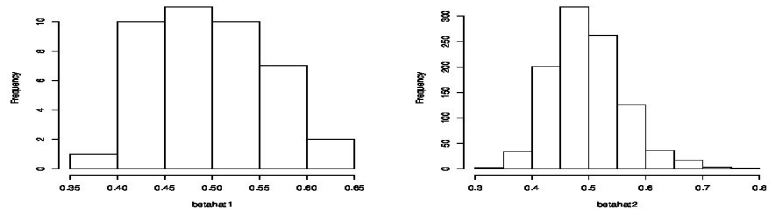


(a) WLS (b) MLE

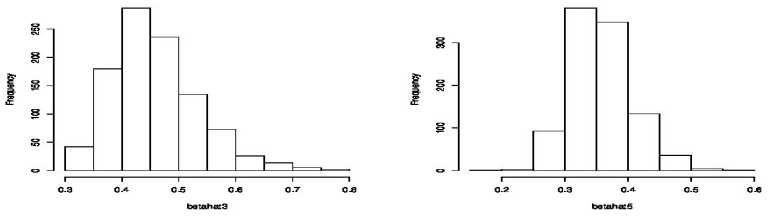


(c) MHDE (d) SCDE

Figure 3.18: Histograms of 1000 $\hat{\beta}_n$s of Model VI by four methods.

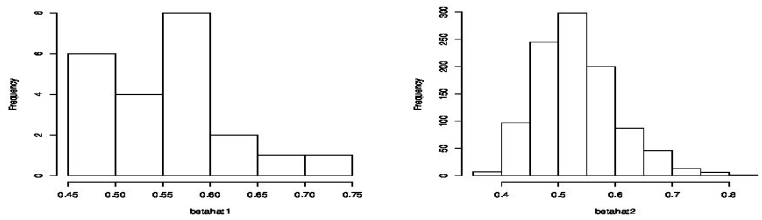


(a) WLS (b) MLE

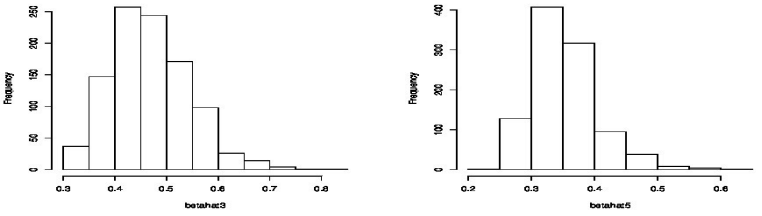


(c) MHDE (d) SCDE

Figure 3.19: Histograms of 1000 $\hat{\beta}_n$s of Model VII by four methods.

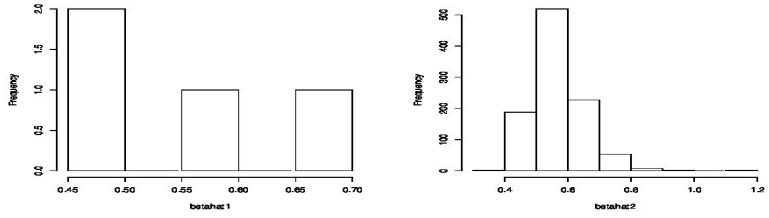


(a) WLS (b) MLE

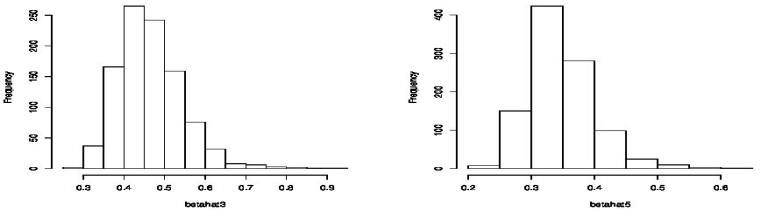


(c) MHDE (d) SCDE

Figure 3.20: Histograms of 1000 $\hat{\beta}_n$s of Model VIII by four methods.

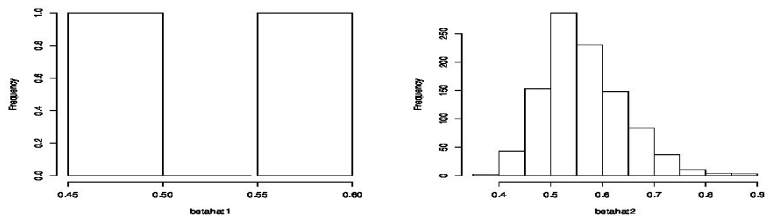


(a) WLS (b) MLE

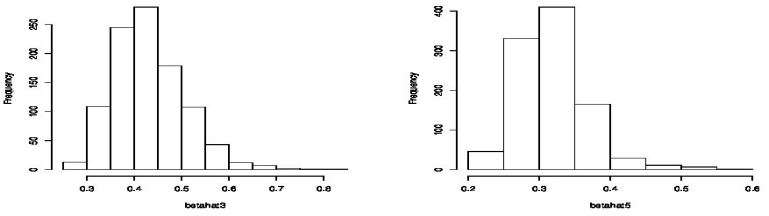


(c) MHDE (d) SCDE

Figure 3.21: Histograms of 1000 $\hat{\beta}_n$s of Model IX by four methods.

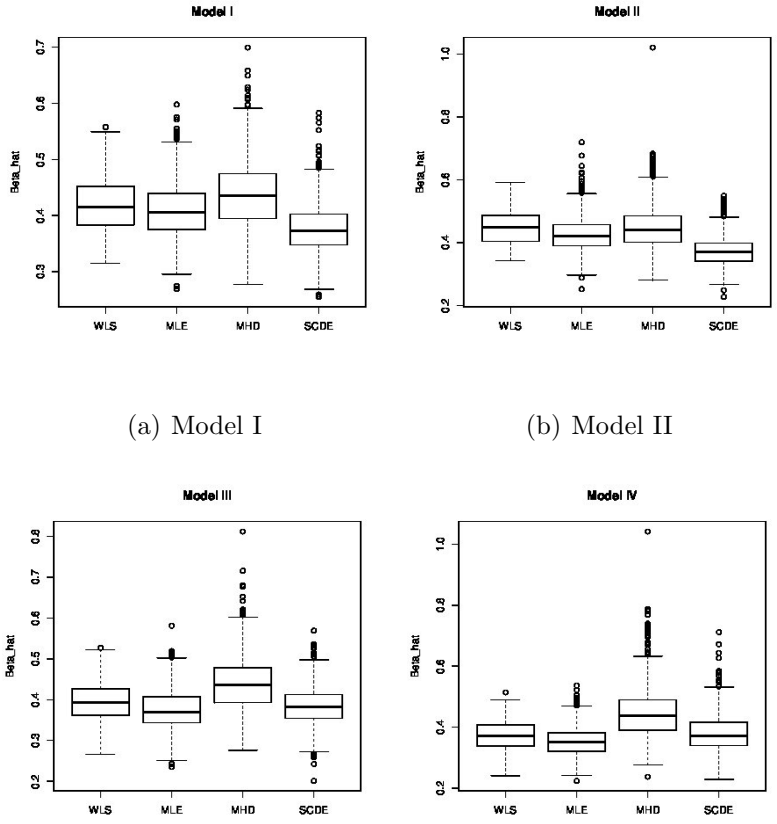


(a) WLS (b) MLE



(c) MHDE (d) SCDE

Figure 3.22: Histograms of 1000 $\hat{\beta}_n$s of Model X by four methods.



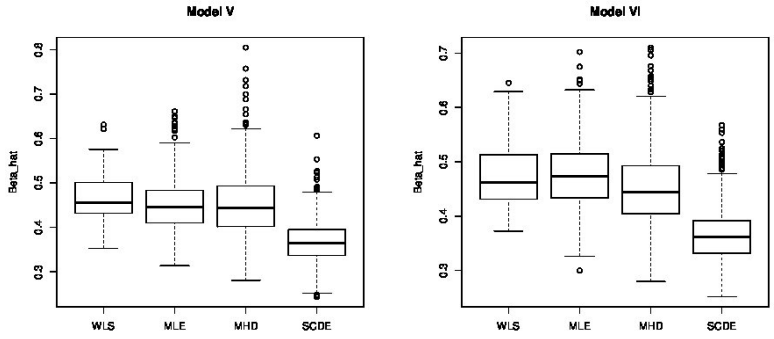
(a) Model I

(b) Model II

(c) Model III

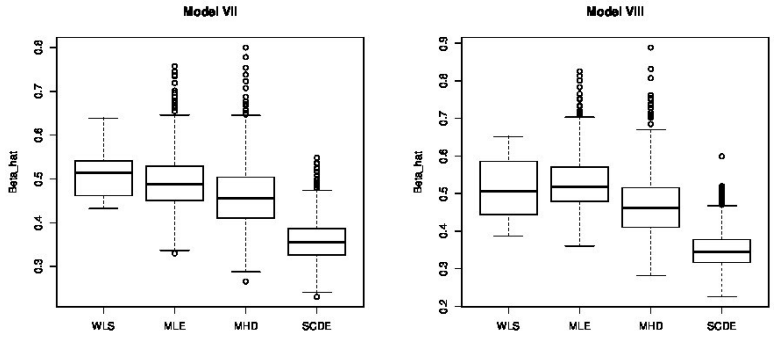
(d) Model IV

Figure 3.23: Boxplot of 1000 $\hat{\beta}_n$'s of four methods for model I-IV



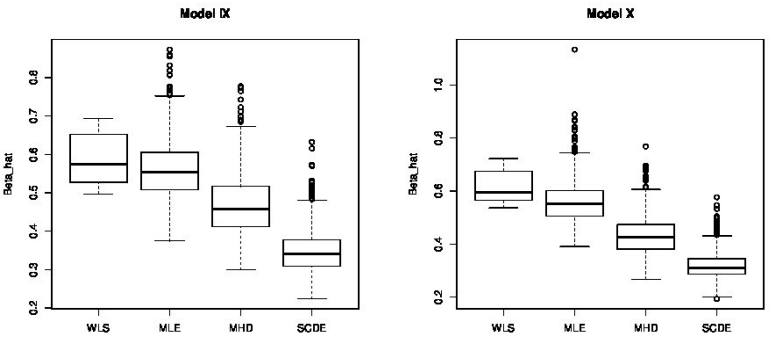
(a) Model V

(b) Model VI



(c) Model VII

(d) Model VIII



(e) Model IX

(f) Model X

Figure 3.24: Boxplot of 1000 $\hat{\beta}_n$'s of four methods for contaminated Probit models

Simulation results are presented in Tables 3.7 to 3.10. These results show that there are monotonic trends (increasing or decreasing) of the values of $E(\hat{\alpha}_n)$ and $E(\hat{\beta}_n)$ as the contamination rate increases. For the non-robust methods, WLSE and MLE, we observe large fluctuations when we increase the contamination rate. When we compare the two robust methods MHDE and SCDE, we find that the SCDE has some protection if the true distribution mixed with shorter tails contamination. Histograms and the boxplots of these estimators shown on Figures 3.17 to 3.24 suggest that the SCDE might be the best if we wish to protect against the true distribution mixed with shorter tails contamination.

3.4 An Application to Real Data Example

In this chapter, we have derived a new robust and efficient estimator, SCDE, by illustrating its properties and by showing numerical advantages in simulations. Now we show this advantage in a real data example. The real data we are going to use are from the second half of Table 2.8.

We consider A (with parameter value $(-3.1501, 1.3699)$) as the postulated model and B (with parameter value $(-4.4101, 1.8056)$) as the contamination. For $75\%A+25\%B$, we obtain $\hat{\alpha}_{MLE} = -5.2679$, $\hat{\beta}_{MLE} = 1.8890$; but $\hat{\alpha}_{SCDE} = -3.9469$, $\hat{\beta}_{SCDE} = 1.4194$, which are more robust in dealing with shorter tail contamination.

From this example, we can reasonably conclude that the SCDE is robust in protecting against the true distribution mixed with shorter tails contamination.

Chapter 4

Stimulus Dose Level Estimation

4.1 Background

In Chapters 2 and 3, binary response experiments are performed to find the critical dose based on the assumption that the probability of response increases monotonically as the stimulus dose level increases. Besides the critical dose, dose level corresponding to a quantile (or quantiles) of a monotonically non-decreasing curve are also important in dose-response studies. One of the most important dose level quantity is the 'median effective dose' (ED_{50}) where response should be shown in half of the population that takes it. Rosenberger and Grill (1997) designed an experiment to show the relationship between stimulus dose level and response by estimating the median, lower and upper quartiles of the dose-response curve. They claimed that logistic and probit analysis would yield similar results in estimating the median. Wiens and Li (2012) gave a robust treatment of the link misspecification and model discrimination. There is a huge literature on this area; more recent work and the relevant reference can be seen in the preceding paper.

4.2 Simulation

In this section we compare four methods, MLE, WLSE, MHDE and SCDE, by estimating the lower quartile (ED_{25}), median (ED_{50}), ED_{60} and upper quartile (ED_{75}) of the dose-response curve.

Under the true models I, II, III,..., X from the previous chapter, let $\Gamma_i = \Gamma(x_i) = F(\alpha + \beta x_i)$.

Let p be a value in $(0, 1)$, and let η be the corresponding quantile, that is, $p = F(\eta)$. Then it is easy to see that $\eta = F^{-1}(p) = \alpha + \beta x(p)$.

Let $x_i(p)$ be the theoretical minimum dose to produce a response for $p\%$ of subjects on the true distribution F_i, $i = 1, 2, 3, \ldots, 10$.

$$x_i(p) = \frac{F_i^{-1}(p) - \alpha}{\beta},\ i = 1, 2, 3, \ldots, 10, \tag{4.1}$$

where $F_i(\alpha + \beta x)$ is the correct model with unknown parameters α and β. As in the previous two chapters, we use $\alpha = -2$ and $\beta = 0.4$ as the 'true' value.

If $\alpha_n(i)$ and $\beta_n(i)$ are estimators of α and β based on the true distribution F_i, then the estimator of dose level $\hat{x}_i(p)$ is

$$\hat{x}_i(p) = \frac{F_i^{-1}(p) - \alpha_n(i)}{\beta_n(i)},\ i = 1, 2, 3, \ldots, 10, \tag{4.2}$$

where $F_i\,, i = 1, 2, 3, \ldots, 10$ denote the models I to X, respectively.

For example, if F follows Model I, then $F^{-1}(p) = \log \frac{p}{1-p}$, the logit, and estimator of the median is $\hat{x}_i(\frac{1}{2}) = -\frac{\alpha_n(i)}{\beta_n(i)}$.

For comparing robustness of the four methods under study, we calculate

the value of $MSE(\hat{x}_i(p))$. Define

$$MSE(\hat{x}_i(p)) = E(\hat{x}_i(p) - x_i(p))^2,\ i = 1, 2, 3, \ldots, 10. \tag{4.3}$$

For $p = 0.25,\ 0.5,\ 0.6,\ 0.75$, Tables 4.1 to 4.4 report the estimated values of $MSE(\hat{x}_i(p))$, $i = 1, 2, 3, \ldots, 10$ for the four methods used to obtained α_n and β_n in Chapters 2 and 3.

4.2.1 Results of Logistic Model

By using the logistic model, the results of the dose level predictions of ED_{25}, ED_{50}, ED_{60} and ED_{75} are presented in Tables 4.1 - 4.4. Again, $\hat{x}(p)$ is the average of 1000 replications.

From Tables 4.1-4.4, we find that in most of cases the values of $MSE(\hat{x}(p))$ using SCDE are the least variable, while the WLSE values are the most variable. If we consider only the $MSE(\hat{x}(p))$ for each model, we see that the MLE is least biased under Model I at ED_{25}, ED_{50}, ED_{60} and ED_{75} levels. Further we see that the MLE shows the least bias under Model II at ED_{25}, ED_{50} and ED_{60} levels, and the WLSE is the least biased under Model II at ED_{75} level. For Model III, the MHDE shows the least bias at ED_{25} and ED_{75} levels, and MLE shows the least bias at ED_{50} and ED_{60} levels. For Model IV, the SCDE shows the least bias at ED_{25}, ED_{50}, ED_{60} and ED_{75} levels. For the contaminated models with shorter tail, i.e. Model V to Model X, SCDE is the least biased at ED_{25}, ED_{60} and ED_{75} levels, but MHDE is smaller in the MSE in most of cases.

We can conclude from the numerical results that the MLE and WLSE have some protection if the true distribution has no contamination. Also, SCDE

Table 4.1: Estimates of dose level $x(p)$ in WLSE Method

Model	p=0.25			p=0.5			p=0.6			p=0.75		
	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$
I	2.2535	2.1607	0.4661	5.00	4.9973	0.1843	6.0137	6.0442	0.1872	7.7465	7.8338	0.3257
II	2.4496	2.3850	0.3636	5.00	4.9951	0.1560	5.9249	5.9416	0.1623	7.5504	7.6051	0.2786
III	2.0901	1.8860	2.0454	5.00	4.8527	0.7027	6.0679	5.9415	0.5342	7.9099	7.8195	0.6527
IV	0.9764	0	5.3628	4.4421	3.5295	2.1290	5.5579	4.7054	1.5703	7.3888	6.6351	1.1619
V	2.6355	2.7107	0.3186	5.00	4.9904	0.1372	5.8491	5.8091	0.1370	7.3644	7.2701	0.2216
VI	2.8084	3.0323	0.2682	5.00	5.0234	0.1251	5.7841	5.7357	0.1346	7.1916	7.0145	0.2317
VII	2.9664	3.2627	0.3009	5.00	5.0133	0.1160	5.7278	5.5608	0.1418	7.0336	6.7640	0.2297
VIII	3.1092	3.4985	0.3150	5.00	5.0241	0.1054	5.6789	5.5718	0.1159	6.8908	6.5497	0.2472
IX	3.2373	3.6773	0.3337	5.00	5.0248	0.0986	5.6359	5.5109	0.1144	6.7627	6.3723	0.2737
X	3.3517	3.8291	0.3215	5.00	5.0190	0.0707	5.5979	5.4507	0.0927	6.6483	6.2090	0.2773

Note: The negative numerical results are recorded as 0.

Table 4.2: Estimates of dose level $x(p)$ in MLE Method

Model	p=0.25			p=0.5			p=0.6			p=0.75		
	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$
I	2.2535	2.3647	0.3834	5.00	4.9979	0.1440	6.0137	5.9698	0.1551	7.7465	7.6312	0.2983
II	2.4496	2.5772	0.3542	5.00	4.9989	0.1527	5.9249	5.8771	0.1611	7.5504	7.4205	0.2809
III	2.0901	1.8728	0.6014	5.00	4.9853	0.1814	6.0679	6.1276	0.2053	7.9099	8.0979	0.4713
IV	0.9764	0	2.8421	4.4421	3.6493	0.9112	5.5579	4.9211	0.6151	6.9308	7.0080	0.4723
V	2.6355	2.8812	0.3327	5.00	4.9631	0.1389	5.8491	5.7108	0.1597	7.3644	7.0451	0.3155
VI	2.8084	3.1794	0.3861	5.00	4.9974	0.1383	5.7841	5.6478	0.1577	7.1916	6.8154	0.3352
VII	2.9664	3.4038	0.3843	5.00	4.9871	0.1234	5.7278	5.4823	0.1834	7.0336	6.5705	0.3735
VIII	3.1092	3.6315	0.4297	5.00	4.9983	0.1065	5.6789	5.4890	0.1438	6.8908	6.3651	0.4117
IX	3.2373	3.8211	0.4723	5.00	5.0058	0.0977	5.6359	5.4331	0.1394	6.7627	6.1904	0.4436
X	3.3517	3.9695	0.4960	5.00	5.0072	0.0877	5.5979	5.3836	0.1350	6.6483	6.0449	0.4699

Note: The negative numerical results are recorded as 0.

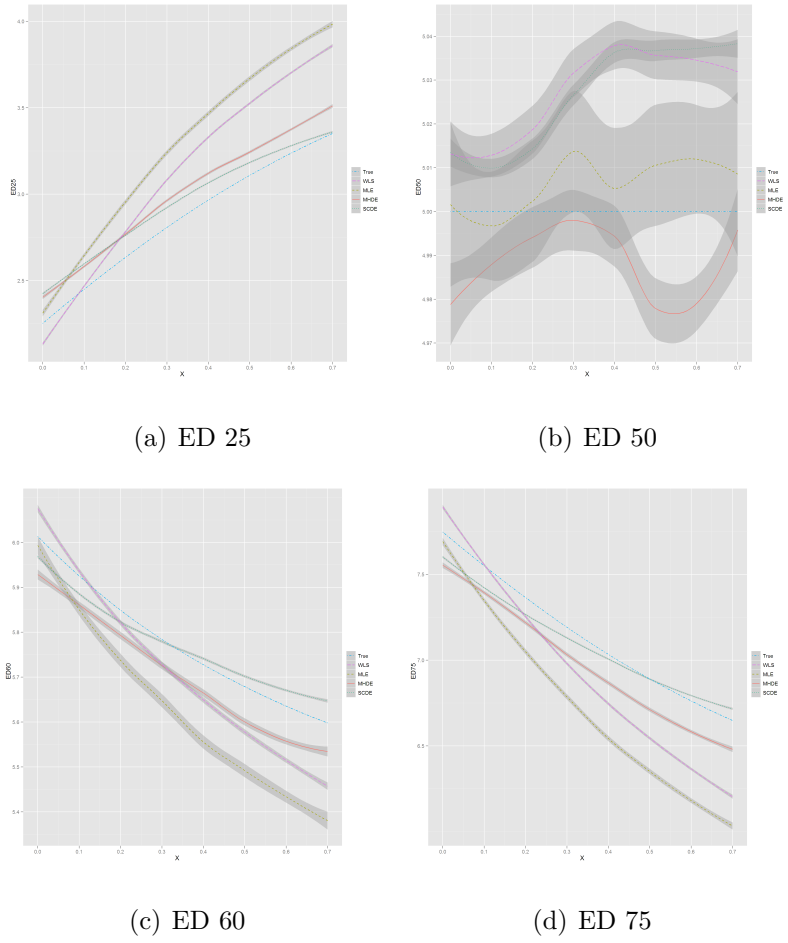
Table 4.3: Estimates of dose level $x(p)$ in MHDE Method

Model	p=0.25			p=0.5			p=0.6			p=0.75		
	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$
I	2.2535	2.4650	0.4287	5.00	4.9832	0.1657	6.0137	5.9126	0.1864	7.7465	7.5015	0.3688
II	2.4496	2.5320	0.3785	5.00	4.9934	0.1561	5.9249	5.8860	0.1713	7.5504	7.4548	0.3216
III	2.0901	2.0449	0.4700	5.00	4.9887	0.1962	6.0679	6.0690	0.2237	7.9099	7.9324	0.4325
IV	0.9764	1.0582	1.0434	4.4421	4.4352	0.2668	5.5579	5.5223	0.2285	7.3888	7.3063	0.3890
V	2.6355	2.7161	0.3273	5.00	4.9909	0.1427	5.8491	5.8078	0.1571	7.3644	7.2656	0.2887
VI	2.8084	2.8880	0.2928	5.00	4.9810	0.1376	5.7841	5.7298	0.1552	7.1916	7.0740	0.2833
VII	2.9664	3.045441	0.2644	5.00	4.9822	0.1235	5.7278	5.5879	0.1476	7.0336	6.9190	0.2308
VIII	3.1092	3.1926	0.2367	5.00	4.9883	0.1221	5.6789	5.6330	0.1328	6.8908	6.7840	0.2201
IX	3.2373	3.3266	0.2041	5.00	4.9776	0.1063	5.6359	5.5732	0.1180	6.7627	6.6286	0.1998
X	3.3517	3.4558	0.1919	5.00	4.9886	0.0998	5.5979	5.5446	0.1104	6.6483	6.5214	0.1856

Table 4.4: Estimates of dose level $x(p)$ in SCDE Method

Model	p=0.25			p=0.5			p=0.6			p=0.75		
	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$
I	2.2535	2.4339	0.4429	5.00	4.9930	0.1921	6.0137	5.9375	0.2054	7.7465	7.5521	0.3603
II	2.4496	2.4858	0.3754	5.00	4.9848	0.1803	5.9249	5.8910	0.1893	7.5504	7.4838	0.3080
III	2.0901	2.1399	0.5614	5.00	4.9997	0.2078	6.0679	6.0493	0.2256	7.9099	7.8595	0.4425
IV	0.9764	1.0206	0.9798	4.4421	4.4295	0.2567	5.5579	5.5269	0.2185	7.3888	7.3278	0.3612
V	2.6355	2.7033	0.3607	5.00	5.0160	0.1731	5.8491	5.8466	0.1764	7.3644	7.3288	0.2751
VI	2.8084	2.8291	0.3182	5.00	4.9979	0.1559	5.7841	5.7738	0.1595	7.1916	7.1667	0.2476
VII	2.9664	2.9853	0.2717	5.00	5.0160	0.1339	5.7278	5.6509	0.1394	7.0336	7.0466	0.2053
VIII	3.1092	3.1408	0.2236	5.00	5.0257	0.1221	5.6789	5.7024	0.1254	6.8908	6.9106	0.1833
IX	3.2373	3.2322	0.1924	5.00	5.0322	0.1172	5.6359	5.6815	0.1209	6.7627	6.8321	0.1678
X	3.3517	3.3116	0.1820	5.00	5.0242	0.1061	5.5979	5.6455	0.1104	6.6483	6.7369	0.1588

and MHDE are better if we wish to protect against contamination models.



(a) ED 25 (b) ED 50

(c) ED 60 (d) ED 75

Figure 4.1: Dose Level Plots of Four Methods for Logistic Model

In Figure 4.1 (a), the dose level plot for ED_{25} shows that after increasing the contamination rate (X), SCDE is getting closer and closer to the true value, while the MLE and WLSE keep increasing. In Figure 4.1 (b), it is not clear which one is the best estimator from the dose level plot for ED_{50}, but the MHDE shows a trend in getting closer to the true value. In Figure 4.1 (c) and (d), the dose level plot for ED_{60} and ED_{60} show that after increasing the contamination rate (X), SCDE is getting closer and closer to the true value, while the MLE and WLSE keep increasing.

4.2.2 Results of Probit Model

By using the Probit model, the simulation results of the dose level prediction of ED_{25}, ED_{50}, ED_{60} and ED_{75} are shown in Tables 4.5-4.8.

From Tables 4.5-4.8, we find by using Probit model, the values of $MSE(\hat{x}(p))$ of MHDE and SCDE are less variable than those of WLSE and MLE. For each model, we see that the MLE is least biased under Models I and II at ED_{25}, ED_{50}, ED_{60} and ED_{75} levels. Further we see that the MHDE is the least biased under Model III at ED_{25} level, while the MLE shows the least bias at levels ED_{50}, ED_{60} and ED_{75}. For Model IV, the MHDE shows the least bias at all levels ED_{25}, ED_{50}, ED_{60} and ED_{75}. For the contaminated models with shorter tail, i.e. Model V to Model X, SCDE and MHDE is smaller in the MSE in most of cases.

We can conclude from the numerical results that the method of MLE has some protection if the true distribution has no contamination. Also, SCDE and MHDE are better if we wish to protect against contamination models.

Table 4.5: Estimates of dose level $x(p)$ in WLSE Method (Probit Model)

Model	p=0.25			p=0.5			p=0.6			p=0.75		
	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$
I	3.3138	3.3406	0.1503	5.00	4.9832	0.1142	5.6334	5.6002	0.1198	6.6862	6.6258	0.1522
II	3.4379	3.5673	0.1851	5.00	5.0172	0.1191	5.5775	5.5532	0.1167	6.5621	6.4672	0.1398
III	3.2158	3.1716	0.1931	5.00	5.0003	0.1158	5.6672	5.6841	0.1160	6.7842	6.8290	0.1519
IV	2.5814	1.8929	0.8279	4.6507	4.1467	0.4190	5.3493	4.9023	0.3312	6.4736	6.1184	0.2535
V	3.5545	3.7271	0.1422	5.00	5.0284	0.0894	5.5300	5.5055	0.0908	6.4455	6.3297	0.1196
VI	3.6617	3.8181	0.1169	5.00	5.0012	0.0606	5.4894	5.4339	0.0661	6.3383	6.1844	0.1086
VII	3.7587	3.8881	0.1096	5.00	4.9858	0.0807	5.4543	5.3875	0.0861	6.2413	6.0835	0.1158
VIII	3.8456	4.0403	0.1811	5.00	4.9098	0.1181	5.4238	5.2289	0.1399	6.1544	5.7792	0.2339
IX	3.9230	4.1944	0.0970	5.00	5.0063	0.0403	5.3970	5.3055	0.0632	6.0770	5.8181	0.1573
X	3.9919	4.2683	0.0765	5.00	4.8561	0.0207	5.3734	5.0738	0.0898	6.0081	5.4439	0.3186

Table 4.6: Estimates of dose level $x(p)$ in MLE Method (Probit Model)

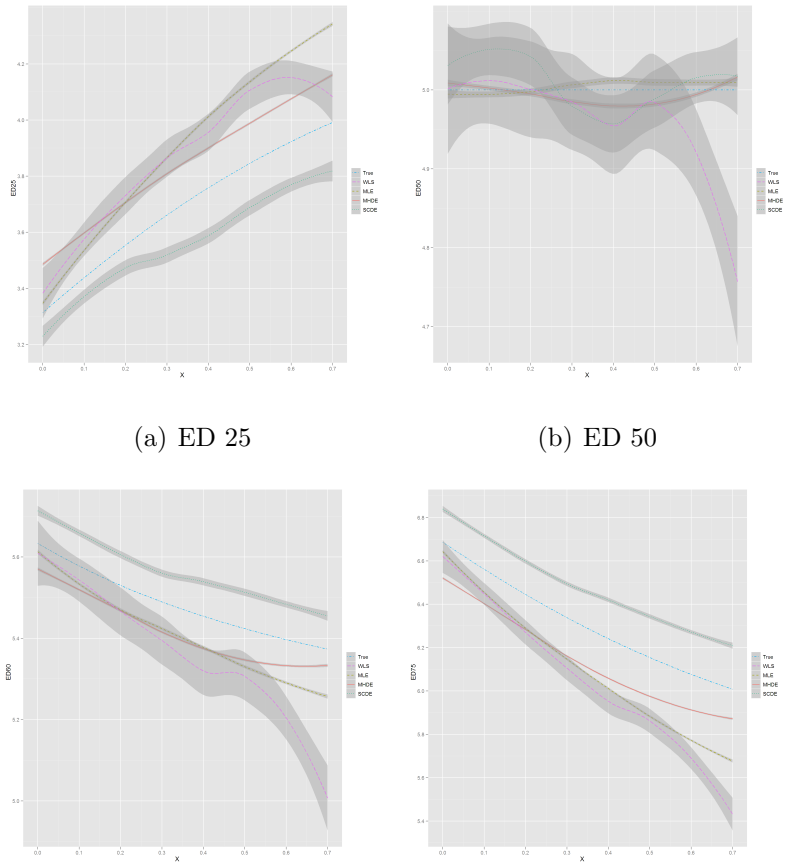
Model	p=0.25			p=0.5			p=0.6			p=0.75		
	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$
I	3.3138	3.3234	0.1162	5.00	5.0045	0.0714	5.6334	5.6359	0.0757	6.6862	6.6855	0.1083
II	3.4379	3.5334	0.1141	5.00	5.0224	0.0663	5.5775	5.5729	0.0688	6.5621	6.5114	0.0983
III	3.2158	3.0867	0.1744	5.00	5.0084	0.0816	5.6672	5.7269	0.0883	6.7842	6.9300	0.1501
IV	2.5814	1.7445	1.0022	4.6507	4.0654	0.4570	5.3493	4.8694	0.3293	6.4736	6.1636	0.2142
V	3.5545	3.6971	0.1151	5.00	4.9983	0.0652	5.5300	5.4754	0.0696	6.4455	6.2995	0.1057
VI	3.6617	3.7928	0.1105	5.00	5.0109	0.0686	5.4894	5.4563	0.0717	6.3383	6.2290	0.1002
VII	3.7587	3.8679	0.1019	5.00	4.9890	0.0698	5.4543	5.3993	0.0743	6.2413	6.1101	0.1028
VIII	3.8456	4.1111	0.1361	5.00	5.0003	0.0538	5.4238	5.3267	0.0653	6.1544	5.8895	0.1376
IX	3.9230	4.2287	0.1538	5.00	5.0122	0.0467	5.3970	5.3010	0.0564	6.0770	5.7956	0.1346
X	3.9919	4.3247	0.1672	5.00	5.0030	0.0498	5.3734	5.2543	0.0656	6.0081	5.6814	0.1662

Table 4.7: Estimates of dose level $x(p)$ in MHDE Method (Probit Model)

Model	p=0.25			p=0.5			p=0.6			p=0.75		
	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$
I	3.3138	3.4159	0.1532	5.00	4.9838	0.0802	5.6334	5.5727	0.0852	6.6862	6.5517	0.1329
II	3.4379	3.5444	0.1448	5.00	4.9823	0.0793	5.5775	5.5139	0.0845	6.5621	6.4203	0.1299
III	3.2158	3.3195	0.1730	5.00	4.9750	0.0940	5.6672	5.5941	0.1036	6.7842	6.6306	0.1672
IV	2.5814	2.7402	0.3354	4.6507	4.6463	0.1234	5.3493	5.2898	0.1080	6.4736	6.3255	0.1427
V	3.5545	3.6756	0.1320	5.00	4.9929	0.0735	5.5300	5.4759	0.0783	6.4455	6.3102	0.1198
VI	3.6617	3.7009	0.1243	5.00	4.9862	0.0764	5.4894	5.4562	0.0790	6.3383	6.2716	0.1090
VII	3.7587	3.7135	0.1262	5.00	4.9726	0.0751	5.4543	5.4334	0.0747	6.2413	6.2317	0.0970
VIII	3.8456	3.9798	0.1072	5.00	4.9968	0.0636	5.4238	5.3702	0.0698	6.1544	6.0139	0.1087
IX	3.9230	4.0413	0.0905	5.00	4.9895	0.0526	5.3970	5.3390	0.0597	6.0770	5.9376	0.0985
X	3.9919	4.1144	0.0845	5.00	4.9915	0.0459	5.3734	5.3164	0.0509	6.0081	5.8687	0.0832

Table 4.8: Estimates of dose level $x(p)$ in SCDE Method (Probit Model)

Model	p=0.25			p=0.5			p=0.6			p=0.75		
	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$	$x(p)$	$\hat{x}(p)$	$MSE(\hat{x}(p))$
I	3.3138	3.2077	0.1596	5.00	5.0331	0.0900	5.6334	5.7187	0.0951	6.6862	6.8584	0.1413
II	3.4379	3.3459	0.1471	5.00	5.0445	0.0959	5.5775	5.6724	0.1058	6.5621	6.7430	0.1586
III	3.2158	3.1578	0.1873	5.00	5.0351	0.1101	5.6672	5.7370	0.1132	6.7842	6.9124	0.1574
IV	2.5814	2.5012	0.3659	4.6507	4.7135	0.1456	5.3493	5.4603	0.1303	6.4736	6.6623	0.1684
V	3.5545	3.4492	0.1421	5.00	5.0416	0.0941	5.5300	5.6254	0.1060	6.4455	6.6339	0.1636
VI	3.6617	3.4923	0.1539	5.00	5.0476	0.0916	5.4894	5.6163	0.1105	6.3383	6.6029	0.1961
VII	3.7587	3.5072	0.1894	5.00	5.0413	0.0903	5.4543	5.6028	0.1160	6.2413	6.5755	0.2392
VIII	3.8456	3.6985	0.1239	5.00	5.0263	0.0727	5.4238	5.5137	0.0853	6.1544	6.3541	0.1469
IX	3.9230	3.7629	0.1259	5.00	5.0278	0.0733	5.3970	5.4941	0.0874	6.0770	6.2927	0.1533
X	3.9919	3.8014	0.1431	5.00	5.0238	0.0724	5.3734	5.4765	0.0887	6.0081	6.2462	0.1691



(a) ED 25 (b) ED 50

(c) ED 60 (d) ED 75

Figure 4.2: Dose Level Plots of Four Methods for Probit Model

We did not see any significant differences among those four methods from Figures 4.2. Therefore, no convincing conclusion can be made by the dose level plots by using Probit model.

Chapter 5

Summary and Future Work

5.1 Summary

In this thesis, we have focused on examining statistical procedures based on some minimum distance methods. We have introduced and analyzed the properties of two estimators, the minimum Hellinger distance estimator and the symmetric Chi-squared distance estimator, and have confirmed that they are both robust and efficient by an example in dose-response studies.

In Chapter 2, we have confirmed previous robustness and efficiency results related to MHDE by illustrating a sequence of theorems and some Monte Carlo studies. Our simulation studies suggests that MHDE works well in the cases of models with contamination, especially good if the true distribution has longer tails than the postulated model distribution. MHDE has also been applied to a real data example of insecticides to show the same results.

In Chapter 3, we suggested using a new method for parameter estimation, namely the symmetric chi-square distance estimation. We investigated a technique for this new parameter estimation method by a similar approach as in

the MHDE case. Our SCDE estimator exhibits a similar level of robustness as the MHDE for the simulation problem described in Chapter 3, but it also shows a better performance in protecting a case with shorter tail contaminations. We have numerically examined the estimator using histograms and boxplots to show that the SCDE is robust.

In Chapter 4, we used four methods to show the relationship between predictive stimulus dose level and model contamination rate at the median and the lower and upper quartile levels. By using the logistic model, we numerically confirmed that MHDE and SCDE are robust in the contamination models.

5.2 Recommendation for Future Work

The following are some recommendations for future research based on this thesis:

* In this thesis, we considered only the case where the contaminated part from the same distribution family with different parameter values. Future research could be conducted on the case where the model density function is of a completely different family than the true probability density of the observed data. This might occur, for example, if the normal distribution was postulated when the logistic distribution was actually true one.

* Up until now, little has been done with symmetric chi-square distance estimators for regression problems. Also, future study may be carried out to develop improved algorithms for the SCDE.

Bibliography

Abdelbasit, K. M., and Plackett, R. L. (1983), "Experimental Design for Binary Data," *Journal of the American Statistical Association*, 78, 90-98.

Agresti, A. (2012), *Categorical Data Analysis, Third Edition.* John Wiley & Sons Inc.

Beran, R.J. (1978), "Robust Location Estimates," *Ann. Statist.*, 5, 431-444.

Beran, R.J. (1977), "Minimum Hellinger Distance Estimates for Parametric Models," *Ann. Statist.*, 5, 445-463.

Bhattacharyya, A. (1943), "On a Measure of Divergence Between Two Statistical Populations Defined by Their Probability Distribution," *Bull. Calcutta Math. Soc.*, 35, 99-109.

Casella, G. and Berger R.L. (2002), *Statistical inference, Second Edition.* Duxbury.

Donoho, D.L. and Liu, R.C. (1988), "The Automatic Robustness of Minimum Distance Functionals," *Ann. Statist.*, 16, 552-586.

Giltinan, D.M., Capizzi, T.P. and Malani, M. (1988), "Diagnostic Tests for

Similar Action of Two Compound," *Applied Statistical Association*, 1, 39-50.

Hampel, F.R. (1968), "Contributions to the Theory of Robust Estimation," *Unpublished Ph.D. thesis*, University of California Berkeley.

Hampel, F.R. (1973), "Robust Estimation: a Condensed Partial Survey," *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 27, 87-104.

Hampel, F.R. (1974), "The Influence Curve and Its Role in Robust Estimation," *J. Amer. Statist. Assoc.*, 69, 383-393.

Hampel, F.R. (1978), "Modern Trends in the Theory of Robustness," *Math. Operationforsch. Statist., Ser. Statistics.*, 9, 425-442.

Huber, P.J. (1972), "Robust Statistics: a Review," *Ann. Math. Statist.*, 43, 1041-1067.

Huber, P.J. (1973), "Robust Regression: Asymptotics, Conjectures and Monte Carlo," *Ann. Statist.*, 1, 799-821.

Huber, P.J. (2009), *Robust Statistics, Second Edition.* John Wiley & Sons Inc.

Karunamuni, R.J. and Wu, J. (2011), "One-Step Minimum Hellinger Distance Estimation," *Comput. Statist. Data Anal.*, 55, 3148-3164.

Kollo, T. and von Rosen, D. (2005), *Advanced Multivariate Statistics with Matrices.* Springer.

Lindsay, B.G. (1994), "Efficiency Versus Robustness: the Case for Minimum Hellinger Distance and Related Methods," *Ann. Statist.*, 22, 1081-1114.

Lindsay, B.G. (2004), "Chapter 14: Statistical Distances as Loss Functions in Assessing Model Adequacy," *The Nature of Scientific Evidence: Statistical, Philosophical, and Empirical Considerations*, 439-464. The University of Chicago Press.

Rosenberger, W. F. and Grill, S. E. (1997), "A Sequential Design for Psychophysical Experiments: An Application to Estimating Timing of Sensory Events," *Statist. Med.* , 16, 2245-2260.

Shorack, G.R. (2000), *Probability for Statisticians*. Springer.

Stather, C.R. (1981), "Robust Statistical Inference," *Unpublished Ph.D. thesis*. La Trobe University.

Tang, Q. and Karunamuni, R.J. (2013), "Minimum Distance Estimation in a Finite Mixture Regression Model," *J. Multivariate Anal.*, 120, 185-204.

Wiens, D.P. and Li, P. (2012), "A Robust Treatment of a Dose-Response Study," *Applied Stochastic Models in Business and Industry*, 28, 164-173.

Wu, J. and Karunamuni, R.J. (2009), "On Minimum Hellinger Distance Estimation," *Canad. J. Statist.*, 37, 514-533.

Wu, J. and Karunamuni, R.J. (2012), "Efficient Hellinger Distance Estimates for Semiparametric Models," *J. Multivariate Anal.*, 107, 1-23.

Appendix

R code to Calculate WLSE, MLE, MHDE, SCDE in Model I and II of logistic model, the others are similar.

```
rm(list = ls()) # clear the memory
#download Rlab for rbern
#WLSE method
wlse = replicate(1000,
p=array(0)
for(i in 1:10)
p[i]=sum(rbern(20,F[i]))/20
w=diag(f(y)^2/(p*(1-p)))
x= matrix(c(1,1,1,1,1,1,1,1,1,1,1,2,3,4,5,6,7,8,9,10),nc=2)
r1=c(0.1,0.1) #or use r=rep(0.1,2) to get a diagonal matrix of Rj.
R=diag(r1)
xtwx=R%*%t(x)%*%w%*%x
xtwxinv=solve(xtwx)
Y=log(p/(1-p))
xtwy=R%*%t(x)%*%w%*%Y
wlse=xtwxinv%*%xtwy)
WLSE=matrix(wlse[is.na(wlse)],nr=2)
```

alphahat1=WLSE[1,]

betahat1=WLSE[2,]

a1=round(mean(alphahat1),4)

b1=round(mean(betahat1),4)

round(cov(cbind(alphahat1,betahat1)),4)

mse.alpha1=mean((alphahat1+2)ˆ2)

round(mse.alpha1,4)

mse.beta1=mean((betahat1-0.4)ˆ2)

round(mse.beta1,4)

hist(betahat1,main = paste())

#MLE Method

mle.logit = replicate(1000,

y=array(0)

for(i in 1:10)

y[i]=sum(rbern(20,F[i]))

my=20-y

out.logit= glm(formula = cbind(y, my)∼ xi, family = binomial(link = "logit"))

mle.logit=out.logit$coefficients)

MLE.logit=matrix(mle.logit[!is.na(mle.logit)],nr=2)

alphahat2=MLE.logit[1,]

betahat2=MLE.logit[2,]

a2=round(mean(alphahat2),4)

b2=round(mean(betahat2),4)

round(cov(cbind(alphahat2,betahat2)),4)

mse.alpha2=mean((alphahat2+2)ˆ2)

```
round(mse.alpha2,4)
mse.beta2=mean((betahat2-0.4)^2)
round(mse.beta2,4)
hist(betahat2,main = paste() )
hist(betahat3,main = paste() )
#MHD Method
mhd.beta = replicate(1000,
p=array(0)
for(i in 1:10)
p[i]=sum(rbern(20,F[i]))/20
A = matrix( c(p, 1-p), nrow=10, ncol=2, byrow = FALSE)
A1=sqrt(A)
mhd.beta¡-function(beta)
FF=array(0)
for(i in 1:10)
FF[i]=(exp(beta[1]+beta[2]*xi[i]))/(1+exp(beta[1]+beta[2]*xi[i]))
B = matrix( c(FF, 1-FF), nrow=10, ncol=2, byrow = FALSE)
B1=sqrt(B)
-sum(diag(t(A1)%*%B1))
mhd.beta=nlm(mhd.beta,c(-2,0.4))$estimate)
alphahat3=mhd.beta[1,]
betahat3=mhd.beta[2,]
round(mean(alphahat3),4)
round(mean(betahat3),4)
round(cov(cbind(alphahat3,betahat3)),4)
mse.alpha3=mean((alphahat3+2)^2)
```

```
round(mse.alpha3,4)
mse.beta3=mean((betahat3-0.4)^2)
round(mse.beta3,4)
#symmetric chi-square method
scs.beta = replicate(1000,
p=array(0)
for(i in 1:10)
p[i]=sum(rbern(20,F[i]))/20
scs.beta=function(beta)
sum((p-exp(beta[1]+beta[2]*xi)/(1+exp(beta[1]+beta[2]*xi)))^2/
(p+exp(beta[1]+beta[2]*xi)/(1+exp(beta[1]+beta[2]*xi)))^2+
(p-exp(beta[1]+beta[2]*xi)/(1+exp(beta[1]+beta[2]*xi)))^2/
(2-p-exp(beta[1]+beta[2]*xi)/(1+exp(beta[1]+beta[2]*xi)))^2)
scs.beta=nlm(scs.beta,c(-2,0.4))$estimate)
alphahat5=scs.beta[1,]
betahat5=scs.beta[2,]
round(mean(alphahat5),4)
round(mean(betahat5),4)
round(cov(cbind(alphahat5,betahat5)),4)
mse.alpha5=mean((alphahat5+2)^2)
round(mse.alpha5,4)
mse.beta5=mean((betahat5-0.4)^2)
round(mse.beta5,4)
hist(betahat5,main = paste() )
rm(list = ls()) # clear the memory
#download Rlab for rbern
```

```
xi=c(1,2,3,4,5,6,7,8,9,10)
alpha=-2
beta=0.4
#Model II, WLSE Method
y=alpha+beta*xi
F=0.9*exp(y)/(1+exp(y))+0.1*exp(2*y)/(1+exp(2*y))
F #cdf
#MLE Method
mle.logit = replicate(1000,
y=array(0)
for(i in 1:10)y[i]=sum(rbern(20,F[i]))
my=20-y
out.logit= glm(formula = cbind(y, my)    xi, family = binomial(link =
"logit"))
mle.logit=out.logit$coefficients)
MLE.logit=matrix(mle.logit[!is.na(mle.logit)],nr=2)
alphahat2=MLE.logit[1,]
betahat2=MLE.logit[2,]
a2=round(mean(alphahat2),4)
b2=round(mean(betahat2),4)
a2
b2
round(cov(cbind(alphahat2,betahat2)),4)
mse.alpha2=mean((alphahat2+2)^2)
round(mse.alpha2,4)
mse.beta2=mean((betahat2-0.4)^2)
```

```
round(mse.beta2,4)
hist(betahat2,main = paste() )
Xp.MSE=function(p)
((log(p/(1-p))-a2)/b2-(log(p/(1-p))+2)/0.4)^2
round(Xp.MSE(0.25),4)
round(Xp.MSE(0.5),4)
#MHD Method
mhd.beta = replicate(1000,
p=array(0)
for(i in 1:10)p[i]=sum(rbern(20,F[i]))/20 #generated from bernoulli function with p=F
r=rep(0.1,10)
mhd.beta=function(beta)
-sum(r*(sqrt(p*(0.99*exp(beta[1]+beta[2]*xi)/(1+exp(beta[1]+beta[2]*xi))+
0.01*exp(2*(beta[1]+beta[2]*xi))/(1+exp(2*(beta[1]+beta[2]*xi)))))+
sqrt((1-p)*(1-0.99*exp(beta[1]+beta[2]*xi)/(1+exp(beta[1]+beta[2]*xi))-
0.01*exp(2*(beta[1]+beta[2]*xi))/(1+exp(2*(beta[1]+beta[2]*xi)))))))
mhd.beta=nlm(mhd.beta,c(-2,0.4))$estimate)
alphahat3=mhd.beta[1,]
betahat3=mhd.beta[2,]
round(mean(alphahat3),4)
round(mean(betahat3),4)
round(cov(cbind(alphahat3,betahat3)),4)
mse.alpha3=mean((alphahat3+2)^2)
round(mse.alpha3,4)
mse.beta3=mean((betahat3-0.4)^2)
```

```
round(mse.beta3,4)
hist(betahat3,main = paste() )
#symmetric chi-square method
scs.beta = replicate(1000, pį-array(0)
for(i in 1:10)
p[i]=sum(rbern(20,F[i]))/20 #generated from bernoulli function with p=F
scs.beta=function(beta)
sum((p-(0.9*exp(beta[1]+beta[2]*xi)/(1+exp(beta[1]+beta[2]*xi))+
0.1*exp(2*(beta[1]+beta[2]*xi))/(1+exp(2*(beta[1]+beta[2]*xi)))))ˆ2/
(p+(0.9*exp(beta[1]+beta[2]*xi)/(1+exp(beta[1]+beta[2]*xi))+
0.1*exp(2*(beta[1]+beta[2]*xi))/(1+exp(2*(beta[1]+beta[2]*xi)))))ˆ2+
(p-(0.9*exp(beta[1]+beta[2]*xi)/(1+exp(beta[1]+beta[2]*xi))+
0.1*exp(2*(beta[1]+beta[2]*xi))/(1+exp(2*(beta[1]+beta[2]*xi)))))ˆ2/
(2-p-(0.9*exp(beta[1]+beta[2]*xi)/(1+exp(beta[1]+beta[2]*xi))+
0.1*exp(2*(beta[1]+beta[2]*xi))/(1+exp(2*(beta[1]+beta[2]*xi)))))ˆ2)
scs.beta=nlm(scs.beta,c(-2,0.4))$estimate)
alphahat5=scs.beta[1,]
betahat5=scs.beta[2,]
round(mean(alphahat5),4)
round(mean(betahat5),4)
round(cov(cbind(alphahat5,betahat5)),4)
mse.alpha5=mean((alphahat5+2)ˆ2)
round(mse.alpha5,4)
mse.beta5=mean((betahat5-0.4)ˆ2)
round(mse.beta5,4)
hist(betahat5,main = paste() )
```

```
par(mfrow=c(2,2))
hist(betahat1,main = paste("WLS method") )
hist(betahat2,main = paste("MLE method") )
hist(betahat3,main = paste("MHDE method") )
hist(betahat5,main = paste("SCDE method") )
boxplot(betahat1,betahat2,betahat3,betahat5,ylab="Beta^hat",main="Model
II")
axis(1, at=1:4, lab=c("WLS", "MLE", "MHD", "SCDE"))
```