Minimum Hellinger Distance Estimation Of A Regression Function In A Parametric Family With A Random Design Model

by

Hemalika T. K. Abeysundara, M.S.

A Dissertation

in

Statistics

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements for the Degree of

Doctor of Philosophy

Approved

Dr. Frits Ruymgaart
Chair of Committee

Dr. Alex Trindade

Dr. Leif Ellingson

Mark Sheridan
Dean of the Graduate School

August, 2014

ACKNOWLEDGMENTS

First and foremost I would like to express my sincerest gratitude to my advisor, Dr. Frits Ruymgaart, who has enormously supported me throughout my dissertation with his patience and knowledge whilst giving advice during my work. My sincere gratitude goes to my committee members Dr. Alex Trindade and Dr. Leif Ellingson for their important advice. During my time at Texas Tech University I had the opportunity to learn and work with many great teachers in the Department of Mathematics and Statistics and I indebted to them.

I am grateful to all those who helped me to complete this thesis. Especially I would like to thank my husband Sachith and my daughter Miyuru and all my friends in Texas Tech for their help and continuous encouragements.

Last but not all the least, I am deeply indebted to my parents and my sister for their dedication, guidance, neverending advice, and blessings.

TABLE OF CONTENTS

ABSTRACT

There is a vast number of methods to estimate parameters of a statistical model based on maximizers and minimizers. These methods compete each other based on their properties such as unbiasness, robustness and efficiency. Some parameter estimation methods work best for specific models and fail when the underlying distribution undergoes even a slightest change due to the lack of robustness. Beran proposed an estimator based on Minimum Hellinger Distance (MHD) method that turned out to be both efficient and robust. Here we exploit his idea in the context of regression estimation. We consider a regression problem with random design where the regression function is defined on an arbitrary measurable space and is assumed to belong to a parametric family where parameter is a compact subset of the real line. In this random design model, the design variable is drawn form an unknown, completely arbitrary probability distribution on the design space. The error variable is assumed to have a known density with a finite second moment and zero mean. Moreover, we assume that the design variable and the error variable are stochastically independent. Summarizing, the response variable of the regression model turns out to have a density which is a convolution of the error distribution and the distribution of the design variable where the parameter is a compact set. In the estimation procedure, two different estimators for the density of the response variable will play a role. One is entirely nonparametric and automatically adjusts to the specific parameter at hand. The other one, however, is tailored to a specific parametric value. The MHD estimator for the parameter is now obtained as the minimizer of Hellinger distance between these two. After some elementary properties of the proposed MHD estimator, we prove consistency and asymptotic normality.

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Background:Hellinger Distance and MHDE

Hellinger distance can be introduced as the distance between probability distributions which is used to quantify the similarity between them. It is defined in terms of the Hellinger integral, which was introduced in 1909 by Ernst Hellinger. Let P and Q denote two probability measures that are absolutely continuous with respect to a third probability measure λ. In terms of measure theory, the square of the Hellinger distance between P and Q is defined as the quantity

$$H^2(P,Q) = \frac{1}{2}\int \left(\sqrt{\frac{dP}{d\lambda}} - \sqrt{\frac{dQ}{d\lambda}}\right)^2 d\lambda \tag{1.1}$$

Here $\frac{dP}{d\lambda}$and $\frac{dQ}{d\lambda}$ are the Radon-Nikodym derivatives of P and Q respectively. For compactness, the above formula is often written as

$$H^2(P,Q) = \frac{1}{2}\int \left(\sqrt{dP} - \sqrt{dQ}\right)^2 \tag{1.2}$$

Because (1.1) does not depend on λ , and Hellinger distance between P and Q does not change if λ is replaced with a different probability measure with respect to which both P and Q are absolutely continuous. By taking λ to be Lebesgue measure and denoting $\frac{dP}{d\lambda}$and $\frac{dQ}{d\lambda}$ as probability density functions and naming them as f and g, the squared Hellinger distance can be expressed as;

$$H^2(f,g) = \frac{1}{2}\int \left(\sqrt{f(x)} - \sqrt{(g(x)}\right)^2 dx \tag{1.3}$$

Efficiency and robustness are two fundamental concepts in parametric estimation problem but it is known (see Hubers (1972)) that Maximum Likelihood Estimates (MLE) do not in general posses the property of robustness or stability under small perturbation in the underlying model. To remove this instability, Hampbel (1974) suggested replacing MLE of θ with a related M-estimator whose asymptotic mean is θ under the model density f_θ and is close to θ under small changes in the underlying data distribution. But this procedure entails a loss of asymptotic efficiency under the model density f_θ. The Minimum Hellinger Distance Estimate (MHDE) proposed by Beran (1977) which is based on the minimum Hellinger distance is both asymptotically efficient under given parametric model and also stable under small deviations from that model. In fact, Lindsay (1994) has shown that MLE and MHDE are members of a large class of efficient estimators with various second-order efficiency properties, and MHDE has been shown to have excellent robustness properties in parametric models such as resistance to outliers and robustness with respect to model misspecification (Beran (1977)).

The MHDE of θ is defned as the value of the parameter that minimizes the Hellinger distance between the parametric model and a nonparametric density estimator of f. That is, if we use $\hat{\theta}$ to denote the MHDE, then $\hat{\theta}$ is defined by

$$\hat{\theta} = argmin_{\theta \in \Theta} \left\| f_n^{1/2} - f_\theta^{1/2} \right\|, \tag{1.4}$$

where $\|f_1 - f_2\| = (\int [f_1(x) - f_2(x)]^2 dx)^{\frac{1}{2}}$ denotes the L_2-norm and f_n is a nonparametric density estimator of f, such as the kernel density estimator.

From the definition, it is interesting to note that the MHDE $\hat{\theta}$ is related heuristically to the maximum likelihood estimator of θ. When sample size is sufficiently large, the MLE should be close to θ, the true parameter, and the nonparametric density estimator f_n should be close to f_θ. Finding the MLE amounts to maximizing the integral $\int \log f_\theta(x) dF_n(x)$ over $\theta \in \Theta$, where F_n is the empirical distribution function of the data. Note that

$$\begin{aligned} \int f_n(x) log \left[\frac{f_\theta(x)}{f_n(x)} \right] dx &= 2 \int f_n(x) log \left[1 + \left(\frac{f_\theta^{1/2}(x)}{f_n^{1/2}(x)} - 1 \right) \right] dx \\ &\approx 2 \int f_n(x) \left[\left(\frac{f_\theta^{1/2}(x)}{f_n^{1/2}(x)} - 1 \right) - \frac{1}{2} \left(\frac{f_\theta^{1/2}(x)}{f_n^{1/2}(x)} - 1 \right)^2 \right] dx \\ &= -2 \left\| f_\theta^{1/2} - f_n^{1/2} \right\|^2 . \end{aligned}$$

Thus, it is not unreasonable to expect that the MHDE $\hat{\theta}$ is asymptotically efficient under f_θ.

On the other hand, since

$$\left\| f_\theta^{1/2} - f_n^{1/2} \right\|^2 \leq \int \| f_\theta(x) - f_n(x) \| \, dx \leq 2 \left\| f_\theta^{1/2} - f_n^{1/2} \right\|,$$

the topology induced on the space of probability measures by the Hellinger metric is the same as that induced by the L_1-norm. It is known that the L_1-norm induces a robust topology, therefore, the MHDE could be expected to be robust as well.

In the following section we look, in breif, at Beran's estimate and other MHD estimates introduced for few different models.

1.2 Development of Minimum Hellinger Distance (MHD) Estimation

1.2.1 Beran's MHDE for Parametric Model

Beran (1977) defined and studied the minimum Hellinger distance estimator for parametric model, and has shown MHDE to have excellent robustness properties in parametric models such as resistance to outliers and robustness with respect to model misspecification.

Let $X_1, X_2, \ldots, X_n$ be observed random variables and $\{X_i\}$ are iid with density belonging to a specific parametric family $\{f_\theta : \theta \in \Theta\}$. Suppose $\hat{g}_n$ is a suitable non-parametric estimator of the density of X_i. If $\|.\|$ denotes the L_2 norm, then Beran's proposed estimator for θ is that value (or values) $\hat{\theta}_n$ in the parameter space Θ which minimizes the Hellinger distance between $f_{\hat{\theta}_n}$ and $\hat{g}_n$, $\left\| f_{\hat{\theta}_n}^{\frac{1}{2}} - \hat{g}_n^{\frac{1}{2}} \right\|$.

Associated with the MHDE $\hat{\theta}$, a functional T was defined. Let $\mathcal{F}$ denote the set of all densities with respect to Lebesgue measure on the real line. The functional T is defined on $\mathbb{F}$ such that for every $g \in \mathcal{F}$,

$$\left\| f_{T(g)}^{1/2} - g^{1/2} \right\| = min_{\theta \in \Theta} \left\| f_\theta^{1/2} - g^{1/2} \right\|, \tag{1.5}$$

and the MHDE $\hat{\theta}$ is defined as $T(f_n)$.

The continuity and differentiability of functional and the conditions for the existence of MHDE were studied in the following theorem by Beran.

Theorem 1.2.1. (Beran(1977)) Suppose that Θ is a compact subset of $\mathbb{R}^p$, $\theta_1 \neq \theta_2$ implies $f_{\theta_1} \neq f_{\theta_2}$ on a set of positive Lebesgue measure, and for almost every x, $f_\theta(x)$ is continuous in θ. Then

(i) For every $g \in \mathcal{F}$, there exists $T(g) \in \Theta$ satisfying (1.5).

(ii) If $T(g)$ is unique, the functional T is continuous at g in the Hellinger topology.

(iii) $T(f_\theta) = \theta$ uniquely for every $\theta \in \Theta$.

For notational convenience, let $s_t = f_t^{1/2}$. With further assumptions on s_t, the functional T becomes differentiable, a property that is fundamental for further developments. For specified $t \in \Theta \in \mathbb{R}^p$, we will typically assume that there exist a $p \times 1$ vector $\dot{s}(x)$ with components in L_2 and a $p \times p$ matrix $\ddot{s}(x)$ with components in L_2 such that for every $p \times 1$ real vector e of unit Euclidean length and for every scalar α in a neighborhood of zeros,

$$s_{t+\alpha e}(x) = s_t(x) + \alpha e^T \dot{s}_t(x) + \alpha e^T u_\alpha(x) \tag{1.6}$$

$$\dot{s}_{t+\alpha e}(x) = \dot{s}_t(x) + \alpha \ddot{s}_t(x)e + \alpha v_\alpha(x)e \tag{1.7}$$

where $u_\alpha(x)$ is $p \times 1$, $v_\alpha(x)$ is $p \times p$, and the components of u_α and of v_α individually tend to zero in L_2 as $\alpha \to 0$.

Theorem 1.2.2. (Beran(1977)) Suppose that (1.6) and (1.7) hold for every $t \in int(\Theta)$, $T(g)$ exists, is unique and lies in $int(\Theta)$, $\int \ddot{s}_{T(g)} g^{1/2}(x)dx$ is a nonsingular matrix, and the functional T is continuous at g in the Hellinger topology. Then for every sequence of densities g_n converging to g in the Hellinger metric,

$$\begin{aligned} T(g_n) = T(g) \quad &+ \quad \int \rho_g(x)[g_n^{1/2}(x) - g^{1/2}(x)]dx \\ &+ \quad a_n \int \dot{x}_{T(g)}(x)[g_n^{1/2}(x) - g^{1/2}(x)]dx, \end{aligned} \tag{1.8}$$

where

$$\rho_g(x) = -\frac{\dot{s}_{T(g)}(x)}{\int \ddot{s}_{T(g)}(x) g^{1/2}(x) dx}$$

and a_n is a real $p \times p$ matrix which tends to zero as $n \to \infty$. In particular, for $g = f_\theta$,

$$\begin{aligned} \rho_{f_\theta}(x) &= -\frac{\dot{s}_\theta(x)}{\int \ddot{s}_\theta(x) s_\theta(x) dx} \\ &= -\frac{\dot{s}_\theta(x)}{\int \dot{s}_\theta(x) \dot{s}_\theta^T(x) dx} \end{aligned}$$

Next the large sample behavior of $T(f_n)$ is examined, where f_n is a kernel density estimator

$$f_n(x) = \frac{1}{n h_n S_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n S_n}\right), \tag{1.9}$$

where K is a smooth density function, bandwidth h_n are positive constants such that $h_n \to 0$ as $n \to \infty$, and $S_n = S_n(X_1, \ldots X_n)$ is a robust scale estimator. X_i are i.i.d random variables with density f.

With further assumptions on the bandwidths and kernels, the consistency of the MHDE $\hat{\theta}$ follows from the continuity of functionals in the Hellinger topology.

Theorem 1.2.3. (Beran(1977)) Suppose

(i) K is absolutely continuous and has compact support; K' is bounded.

(ii) f is uniformly continuous.

(iii) $\lim_{n\to\infty} h_n = 0. \lim_{n\to\infty} n^{1/2} h_n = \infty$

(iv) As $n \to \infty$, $s_n \xrightarrow{p} s$, a positive finite constant depending on f.

Then $\left\| f_n^{1/2} - f^{1/2} \right\| \xrightarrow{p} 0$ as $n \to \infty$. If T is a functional continuous at f in the Hellinger metric, then $T(f_n) \xrightarrow{p} T(f)$.

In the next theorem, Beran showed that under stronger assumptions, $T(f_n)$ has an asymptotically normal distribution about $T(f)$.

Theorem 1.2.4. (Beran(1977)) Suppose

(i) K is symmetric about 0 and has compact support.

(ii) K is twice absolutely continuous; K" is bounded.

(iii) T satisfy (1.8) and ρ_g has compact support K on which it is continuous.

(iv) $f > 0$ on K; f is twice absolutely continuous and f" is bounded.

(v) $\lim_{n\to\infty} n^{1/2}h_n = \infty$, $\lim_{n\to\infty} n^{1/2}h_n^2 = 0$.

(vi) There exists a positive finite constant s depending on f such that $n^{1/2}(s_n - s)$ is bounded in probability.

Then

$$\sqrt{n}[T(f_n) - T(f)] \xrightarrow{D} N\left(0, \frac{\int \rho_f(x)\rho_f^T(x)dx}{4}\right).$$

In particular, if $f = f_\theta$, then

$$\sqrt{n}[T(f_n) - \theta] \xrightarrow{D} N\left(0, \frac{1}{4\int \dot{s}_\theta(x)\dot{s}_\theta^T(x)dx}\right).$$

Theoretical results showed that the MHDE was minimax robust in a small Hellinger metric neighborhood of the given family, and the local minimax robustness at f_θ entailed asymptotic efficiency at f_θ, but not conversely. On the other hand, the α-influence curve was introduced in order to examine the behavior of T under a mixture model for gross errors.

Let δ_z denote the uniform density on the interval $(z - \epsilon, z + \epsilon)$, where $\epsilon > 0$ is very small, and let $f_{\theta,\alpha,z} = (1-\alpha)f_\theta + \alpha\delta_z$ for $\theta \in \Theta$, $\alpha \in [0, 1)$, and real z. Here, the density $f_{\theta,\alpha,z}$ models an experiment where independent observations distributed according to f_θ are mixed with approximately $100\alpha\%$ gross errors located near z. For every $\alpha \in (0, 1)$, the difference quotient, named α-influence curve

$$IC_{t,\alpha}(z) = \frac{T(f_{\theta,\alpha,z}) - \theta}{\alpha}$$

is a bounded continuous function of z such that

$$\lim_{z\to\infty} \frac{T(f_{\theta,\alpha,z}) - \theta}{\alpha} = 0.$$

Hence, the functional T is robust at f_θ against $100\alpha\%$ contamination by gross errors at arbitrary real z, whether or not the influence function of T is irrelevant to the matter.

This estimator $\hat{\theta}$ is related heuristically to the MLE of θ if the data density is infact some f_{θ_0} and is asymptotically efficient under f_θ.

1.2.2 Mixture of Two Normals

Based on Beran (1977)'s work, Woodward et al. (1995) examined the MHDE in the case of estimation of the mixing proportion in the mixture of two normals. They discussed the practical feasibility of employing the MHDE in this setting and examined empirically its robustness properties. Their results indicated that the MHDE obtained full efficiency at the true model while performing comparably with the minimum distance estimator based on Cramer-von Mises distance under the symmetric departures from component normality considered. Finite Mixture Model has been a hot topic during the past years. The classic paper on mixture models is by the famous biometrician Pearson (1894), where he used a moment based method to fit a mixture of two heteroscedastic normal components in the paper. A few years later, Charlier and Wicksell (1924) extended Pearson's work to the bivariate normal component case and Doetsch (1928) used it in the case of more than two univariate normal components.

The mixture of two normal components has density

$$f_\theta(x) = \frac{p}{\sqrt{2\pi}\sigma_1} exp\left\{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right\} + \frac{1-p}{\sqrt{2\pi}\sigma_2} exp\left\{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right\},$$

where $\theta = (\mu_1, \sigma_1, \mu_2, \sigma_2, p)'$.

At the first step, they considered the case in which $f_\theta(x)$ is a mixture of known densities, which implies that $\theta = p$. Since the kernel density estimator is Hellinger consistent and the Hellinger metric on the probability distributions is equivalent to the Euclidean metric on the parameter space, implying Theorem 1.2.3, the MHDE $\hat{\theta}$ is consistent. Similarly, by implying Theorem 1.2.4, the MHDE $\hat{\theta}$ has an asymptotic normal distribution and is asymptotically fully efficient.

Next, they considered the case in which the five parameters $p, \mu_1, \sigma_1, \mu_2$ and σ_2 are all unknown, meaning that $\theta = (\mu_1, \sigma_1, \mu_2, \sigma_2, p)'$.

Following Beran (1977), minimizing $\left\| f_\theta^{1/2} - f_n^{1/2} \right\|$ is equivalent to maximizing $\int f_\theta^{1/2} f_n^{1/2}$. However, due to convergence issue, Woodward et al. (1995) approximated this integral by the trapezoidal rule to obtain

$$\hat{I} = \Delta t_i \sum_{i=1}^{k} a_i \left(f_\theta^{1/2}(t_i) - f_n^{1/2}(t_i) \right)^2,$$

where $a_1 = a_k = 1/2$ and $a_i = 1$ for $i = 2, 3, \ldots, k-1$ for a partition $t_1, t_2, \ldots, t_k$ of $[a, b]$, a finite interval.

In order to examine the property of MHDE, a simulation study was conducted to compare MHDE and MLE, and the results were based on Bias, MSE, and the relative efficiencies

$$\begin{aligned} B\hat{i}as &= \tfrac{1}{n_s} \textstyle\sum_{i=1}^{n_s} (\hat{p}_i - p) \\ M\hat{S}E &= \tfrac{1}{n_s} \textstyle\sum_{i=1}^{n_s} (\hat{p} - p)^2 \\ \hat{E} &= \tfrac{M\hat{S}E(MLE)}{M\hat{S}E(MHDE)}. \end{aligned}$$

The study showed that the MHDE appeared to obtain full efficiency at the true model as evidenced by $\hat{E}$ near one in all cases. The probability plots indicated that the normality of the MHDE was very similar to that of the MLE. When checking the results for samples which were simulated as mixtures of $t(4)$ component, all of the $\hat{E}'$s were greater than one providing evidence that the MHDE was more robust to the departures from the assumption of normal components than was the MLE. Further study with the component of $t(2)$ showed that the more the mixed models departed from normality, the better the MHDE was.

1.2.3 Poisson Mixture

Finite Poisson mixtures are used to describe data that are overdispersed and hence can't be fitted by a simple Poisson distribution. Based upon Simpson (1987), Karlis and Xekalaki (1998) derived MHDE for finite Poisson mixtures, and proved it to be both efficient and robust. To facilitate computation, they provided an iterative algorithm.

For k-finite Poisson mixtures, the empirical density function is still the most commonly used $f_n(x)$, and

$$f_\theta(x) = \sum_{i=1}^{k} p_i \frac{e^{\lambda_i}\lambda_i^x}{x!}, x = 0, 1, \ldots,$$

where $\theta = (p_1, p_2, \ldots, p_{k-1}, \lambda_1, \lambda_2, \ldots, \lambda_k), \lambda_i > 0, i = 1, 2, \ldots, k$ and $p_i \in (0, 1)$ for $i = 1, 2, \ldots, k$ with $\sum_{i=1}^{k} p_i = 1$.

To compare MHDE and MLE, Karlis and Xekalaki (1998) studied the estimation equations for both estimates. For parameter θ, the estimating equation for MLE is

$$\sum_{x=0}^{\infty} \frac{f_n(x)}{f_\theta(x)} \frac{\partial f_\theta(x)}{\partial \theta_i} = 0,$$

while the estimating equation for MHDE is

$$\sum_{x=0}^{\infty} \left[\frac{f_n(x)}{f_\theta(x)}\right]^{1/2} \frac{\partial f_\theta(x)}{\partial \theta_i} = 0.$$

If the model is well specified and the sample size is large, the square root of $f_n(x)/f_\theta(x)$ should be close to itself, and thus, we would expect MHDE and MLE to behave similarly. On the other hand, in the case of outliers, for values of x for which the ratio is large, the MHDE gives less weight to the estimation by taking the square root, and thus, not so sensitive to outliers.

Simulation study showed that, for contaminated models, MLE usually modeled the contamination with an additional component. Since mixture models were very often not appropriately specified, including the case where the number of components not being assigned prior to analysis, the MHDE was more reliable in such case.

1.2.4 Finite Mixtures of Poisson Regression Models

Lu et al. (2003) extended the MHDE approach from the finite mixtures of Poisson distributions to the finite mixtures of Poisson regressions for count data.

Let $(y_i, t_i, x_i), i = 1, \ldots, n$ denote observations, where y_i is the observation value of the i^{th} response variable Y_i, t_i is a non-negative quantity representing the time or extent of exposure, and x_i is the observed value of random covariate vector of dimension $p + 1$ corresponding to the regression part of the model. A finite mixture of poisson regression model is defined as

$$f_\theta(y_i|x_i) = \sum_{j=1}^{k} \alpha_j g(y_i log(\lambda_{ij}))$$

with

$$g(y, \gamma) = \frac{1}{y!} exp[y\gamma - e^\gamma], y = 0, 1, \ldots,$$

where α_j denotes the proportion of the j^{th} component with $\sum_{j=1}^{k} \alpha_j = 1$, k is the number of components, $g(y, \gamma)$ is the Poisson probability distribution with mean $\lambda = e^\gamma > 0$, and $\lambda_{ij} = t_i \lambda_j(x_i)$ with

$$\log(\lambda_j(x, \beta_j)) = \beta_{j0} + \beta_{j1}x_1 + \ldots + \beta_{jp}x_p = x^T\beta_j, j = 1, \ldots, k.$$

Here, $x = (1, x_1, \ldots, x_p)^T$, $\beta_j = (\beta_{j0}, \beta_{j1}, \ldots, \beta_{jp})^T \in R^{1+p}$, β_{jl} is the regression coefficient for the l^{th} covariate s_l and j^{th} component, and $\theta = (\alpha_1, \ldots, \alpha_{k-1}, \beta_1^T, \ldots, \beta_K^T)^T$.

The same as before, Lu et al. (2003) used the empirical probability function as

f_n

$$f_n(y) = \frac{N_y}{n}, y = 0, 1, 2, \ldots,$$

assuming that the sample size is sufficiently large, and N_y is the frequency of y among $Y_i, \ldots, Y_n$. If

$$f_\theta(y) = \int f_\theta(y|x) f_X(x) dx = \sum_{j=1}^{k} \alpha_j \int g(y; x^T \beta_j) f_X(x) dx \tag{1.10}$$

is known, except for parameter θ, then

$$\hat{\theta} = argmin_\theta \left\| f_\theta^{1/2} - f_n^{1/2} \right\|.$$

If, however, $f_X(x)$ is unknown, or the integration in (1.10) is complex due to the high dimension of X, then replace $f_\theta(y)$ by a consistent estimator

$$f_{\theta,n}(y) = \frac{1}{n} \sum_{i=1}^{n} f_\theta(y|x_i) = \sum_{j=1}^{k} \sum_{i=1}^{n} \frac{\alpha_j}{n} g(y; x_i^T \beta_j)$$

and the MHDE of θ is defined as

$$\hat{\theta} = argmin_\theta \left\| f_{\theta,n}^{1/2} - f_n^{1/2} \right\|.$$

Results from Monte Carlo simulations suggested that MHDE is a viable alternative to the maximum likelihood estimator when the mixture components were not well separated or the model parameters were near zero.

1.2.5 Non-parametric Mixture Model and Two-Sample Semiparametric Model

Wu and Karunamuni (2009) introduced the MHDE for a two-component mixture model $\theta F + (1-\theta)G$ where F and G are two probability distributions and θ a positive real number between 0 and 1 which defines the mixture weights. They observed three independent samples;

$$\begin{aligned} X_1, X_2, \ldots, X_{n_0} &\sim^{iid} F \\ Y_1, Y_2, \ldots, Y_{n_1} &\sim^{iid} G \\ Z_1, Z_2, \ldots, Z_{n_2} &\sim^{iid} \theta F + (1-\theta) G \end{aligned}$$

and then estimated the mixture parameter θ, treating F anf G as nuisance parameters. The result was an extension of Beran's (1977) technique applied to the combined data set $X_1, \ldots, X_{n_0}, Y_1, \ldots, Y_{n_1}, Z_1, \ldots, Z_{n_2}$ which is a collection of independent observations, but not necessarily identically distributed. The proposed estimate was

$$\hat{\theta} = \hat{T}(\hat{h}) \tag{1.11}$$

where,

$$\hat{T}(\phi) = argmin_{t \in [0,1]} \left\| \widetilde{h}_t^{\frac{1}{2}} - \phi^{\frac{1}{2}} \right\| \tag{1.12}$$

Here, $\widetilde{h}_t$ is the estimator of h_t;

$$h_\theta(x) = \theta f(x) + (1-\theta) g(x) \text{ and } \tilde{h}_t(x) = t\hat{f}(x) + (1-t)\hat{g}(x) \tag{1.13}$$

with $\hat{f}$ and $\hat{g}$ which are kernal density estimators of f and g, two different densities of F and G.

Wu and Karunamuni (2009) again introduced MHDE for a two-sample semiparametric model as follows: Let $X_1, \ldots X_{no}$ be iid random variables with density function g. Let $Z_1, \ldots Z_{n1}$ be iid random variables with density function h independent of X_i s. Define

$$\begin{aligned} X_1, \ldots, X_{n0} &\sim^{iid} \quad g(x) \\ Z_1, \ldots, Z_{n1} &\sim^{iid} \quad h_\theta(x) \end{aligned} \tag{1.14}$$

where $h_\theta x = g(x) exp[\alpha + r(x)\beta]$.
Here $r(x) = (r_1(x), \ldots, r_p(x))_{1\times p}$ and $\beta = (\beta_1, \ldots, \beta_p)^T_{p\times 1}$.
Define $\theta = (\alpha, \beta^T)^T$ and the proposed MHDE for θ is

$$\hat{T}(h_{n1}) = T(\left\{\hat{h}_\theta\right\}_{\theta\in\Theta}, h_{n1}) = argmin_{\theta\in\Theta}\left\|\hat{h}_\theta^{\frac{1}{2}} - h_{n1}^{\frac{1}{2}}\right\| \tag{1.15}$$

where $\hat{h}_\theta$ is the estimate for h_θ. The existance, uniqueness and asymptotic properties are discussed for both the estimates 1.11 and 1.15 by Wu and Karunamuni (2009).

1.2.6 MHDE in Regression

Wang (2009) derived the MHDE which is a deviation from Beran's (1977) for the regression model;

$$Y = g_\theta(X) + \epsilon, \theta \in \Theta \tag{1.16}$$

where X is the random design, g_t the regression fuction, Y the response and ϵ the random error variable. The parameter θ was estimated using minimum Hellinger distance as;

$$\hat{\theta}_n = argmin_{t\in\theta}\ \Delta_{\hat{p}\sigma(n)}\ (t) = argmin_{t\in\theta}\left\|\sqrt{\hat{p}_{\sigma(n)}} - \sqrt{f_t}\right\|^2 \tag{1.17}$$

where

$$\sqrt{f_t} \in L^2([0,1]\times\mathbb{R}) \subset L^2(\mathbb{R}\times\mathbb{R}), t \in \theta$$

and $p \in \mathcal{P} = \{$ all densities p with respect to Lebesgue measure on $\mathbb{R}\times\mathbb{R}\}$. Continuity, consistancy and asymptotic normality of the estimator 1.17 are discussed by Wang (2009).

CHAPTER 2
MODEL AND BASIC ASSUMPTIONS

We consider a regression problem, where the regression function is defined on an arbitrary measurable space, and is assumed to belong to a parametric family. In this chapter, we will consider a random design model, where the design variable is drawn from an unknown, completely arbitrary probability distribution on the design space.

More specifically, let $(\Omega, \mathcal{W}, \mathbb{P})$ be a probability space, $(\mathbb{X}, \mathcal{A})$ an arbitrary measurable space and $X : \Omega \to \mathbb{X}$ a random element in X with induced probability distribution Q on $(\mathbb{X}, \mathcal{A})$. Furthermore, let $\epsilon : \Omega \to \mathbb{R}$ be a Borel measurable error variable with density Ψ on $\mathbb{R}$ with respect to Lebesgue measure λ, finite second moment and zero mean. Moreover, we assume X and ϵ to be stochastically independent. Summarizing we have,

$$\mathbb{E}\epsilon = 0, \mathbb{E}\epsilon^2 = \upsilon^2 < \infty, \quad \epsilon \perp X. \tag{2.1}$$

Furthermore we have a regression function in a parametric family:
$r_t : \mathbb{X} \to \mathbb{R}$ measurable, $t \in \Theta \subset \mathbb{R}$ where the parameter set Θ is compact.
We observe n independent copies $(X_1, Y_1), \ldots, (X_n, Y_n)$ of (X, Y), where

$$Y = r_t(X) + \epsilon. \tag{2.2}$$

Because for every $A \in \mathcal{A}$ we have,

$$\begin{aligned} \mathbb{P}\left\{Y \leq y, X \in A\right\} &= \mathbb{E}\mathbb{P}\left\{Y \leq y, X \in A \,|X\right\} \\ &= \int_{\mathbb{X}} 1_A(x)\, \mathbb{P}\left\{\epsilon \leq y - r_t(x)\right\} dQ(x) \\ &= \int_{\mathbb{X}} 1_A(x)\, \Psi\left\{y - r_t(x)\right\} dQ(x)\,. \end{aligned} \tag{2.3}$$

The density of (X, Y) with respect to $\lambda \otimes Q$ equals,

$$p_t(x, y) = \Psi(y - r_t(x)), \quad (x, y) \in \mathbb{X} \times \mathbb{R}. \tag{2.4}$$

The marginal density of Y with respect to λ is given by

$$f_t(y) = \int_{\mathbb{X}} \varphi(y - r_t(x))dQ(x), \quad t \in \Theta. \tag{2.5}$$

Let $\mathcal{F}$ be the class of all densities with respect to λ on $\mathbb{R}$ so that $\{f_t, t \in \Theta\} \subset \mathcal{F}$.

Exploiting the idea of Minimum Hellinger Distance (MHD) eestimation introduced by Beran (1977), an estimator for t will be defined and asymptotic properties for large sample sizes will be derived. Further technical assumptions will be discussed in future chapters.

In order to define the MHD estimator, two estimators for the density f_t in (2.5) will be employed. One is an entirely nonparametric estimator based on the sample $Y_i, \ldots, Y_n$ which is denoted as $\hat{f} = \hat{f}_{\sigma(n)}(y)$. Let us first choose a kernel $\varphi(x)$ such that

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

Then

$$\varphi_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x}{\sigma})^2}.$$

Now we define

$$\hat{f}_{\sigma(n)}(y) = \frac{1}{n}\sum_{i=1}^{n} \varphi_{\sigma(n)}(y - Y_i). \tag{2.6}$$

The second is a partly parametric estimator

$$\hat{f}_t(y) = \frac{1}{n}\sum_{i=1}^{n} \varphi(y - r_t(X_i)), \quad t \in \Theta. \tag{2.7}$$

Let us also introduce the following notation:

$$\sqrt{f_t} = s_t, \quad \sqrt{\hat{f}_t} = \hat{s}_t, \quad \sqrt{\hat{f}} = \hat{s}. \tag{2.8}$$

All functions in (2.8) are in the Hilbert space $\mathbb{L}^2(\mathbb{R})$.

Now we look at the Mean Integrated Squared Error (MISE) of the square root of the density estimators. The MISE of the square root of the density estimator will be partly reduced to the MISE of the density estimator itself and this result can be found in the literature (see Rosenblatt (1956), Silverman (1986) and Scott (1992)) we will not include a derivation.

In particular the asymptotic mean integrated square bias of the density estimator is of order

$$\int \text{bias}^2(\hat{f}_\sigma) = O(\sigma^4), \tag{2.9}$$

and the integrated variance of the density estimator is of order

$$\int \text{var}^2(\hat{f}_\sigma) = O(\frac{1}{n\sigma^4}). \tag{2.10}$$

The MISE of the square root of the density estimator $\sqrt{\hat{f}_\sigma}$, i.e.

$$\mathbb{E}\left\|\sqrt{\hat{f}_\sigma} - \sqrt{f_\theta}\right\|^2, \quad \sigma > 0. \tag{2.11}$$

CHAPTER 3
MHDE OF THE PROPOSED MODEL AND ELEMENTARY PROPERTIES

In this chapter we present the proposed MHD estimator and its elementary propoerties. We use the notation defined in (2.8) throughout the chapter. Also from (2.5) and (2.7) we have,

$$\begin{aligned}\hat{f}_\theta(y) &= \tfrac{1}{n}\textstyle\sum_{i=1}^{n} \varphi(y - r_\theta(X_i)) \\ &= \textstyle\int_{\mathbb{X}} \varphi(y - r_\theta(x)) d\hat{Q}(x).\end{aligned}$$

and

$$f_\theta(y) = \mathbb{E}\hat{f}_\theta(y) = \int_{\mathbb{X}} \varphi(y - r_\theta(x)) dQ(x).$$

Then

$$\begin{aligned}\hat{\Delta}_\theta(y) &= \hat{f}_\theta(y) - f_\theta(y) \\ &= \textstyle\int_{\mathbb{X}} \varphi(y - r_\theta(x)) d\left\{\hat{Q}(x) - Q(x)\right\}.\end{aligned} \tag{3.1}$$

3.1 Preliminaries

In order to define the MHD estimator, we discuss the following notational conventions, assumptions and theorems.

Definition 3.1 For $h = \theta - \hat{\theta}$, for $h \in \mathbb{R}$, we say that

$$f = o_{L^2}(h) \quad if \quad \frac{\|f(.,h)\|}{h} \to 0 \text{ as } h \to 0.$$

Assumption 3.1 For each $\theta \in \Theta$ there exsits $\dot{s_\theta} \in L^2(\mathbb{R})$, such that

$$s_t - s_\theta = (t - \theta)\dot{s_\theta} + o_{L^2}(t - \theta) \tag{3.2}$$

Theorem 3.1 : Differentiability. For each $f \in \mathcal{F}$,the squared Hellinger distance function

$$\Delta_f(t) = \left\| \sqrt{f} - \sqrt{f_t} \right\|^2, t \in \Theta \tag{3.3}$$

is differentiable on Θ with derivative

$$\dot{\Delta}_f(t) = 2\left\langle \sqrt{f}, \dot{s}_\theta \right\rangle. \tag{3.4}$$

Proof. For t, $\theta \in \Theta$ we have

$$\begin{aligned}
\Delta_f(t) - \Delta_f(\theta) &= \left\|\sqrt{f} - \sqrt{f_t}\right\|^2 - \left\|\sqrt{f} - \sqrt{f_\theta}\right\|^2 \\
&= \left\langle \sqrt{f} - \sqrt{f_t}, \sqrt{f} - \sqrt{f_t} \right\rangle - \left\langle \sqrt{f} - \sqrt{f_\theta}, \sqrt{f} - \sqrt{f_\theta} \right\rangle \\
&= \left\langle \sqrt{f}, \sqrt{f} \right\rangle - 2\left\langle \sqrt{f}, \sqrt{f_t} \right\rangle + \left\langle \sqrt{f_t}, \sqrt{f_t} \right\rangle \\
&\quad - \left\langle \sqrt{f}, \sqrt{f} \right\rangle + 2\left\langle \sqrt{f}, \sqrt{f_\theta} \right\rangle - 2\left\langle \sqrt{f_\theta}, \sqrt{f_\theta} \right\rangle \\
&= 2\left\langle \sqrt{f}, \sqrt{f_\theta} - \sqrt{f_t} \right\rangle + \left\|\sqrt{f_t}\right\|^2 - \left\|\sqrt{f_\theta}\right\|^2 \\
&= 2\left\langle \sqrt{f}, (t-\theta)\dot{s}_\theta + o_{L^2}(|t-\theta|) \right\rangle \\
&= 2\left\langle \sqrt{f}, (t-\theta)\dot{s}_\theta \right\rangle + 2\left\langle \sqrt{f}, o_{L^2}(|t-\theta| \right\rangle.
\end{aligned}$$

This entails at once the result in (3.3). □

Assumption 3.2 For each $f \in \mathcal{F}$, the function $t \mapsto \Delta_f(t)$ has a unique minimum at $t \in \Theta$ such that $\Delta_f(\theta_f) = \min_{t\in\Theta} \Delta_f(t)$

Lemma 3.1 Let $A \in \mathbb{R}^d$ be compact, and suppose that $f : A \to \mathbb{R}$ is continuous with a unique minimum at $X_o \in A : f_o = \min_{x\in A} f(x) - f(x_o)$. Let also $g : A \to \mathbb{R}$ be continuous. Then for each $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ such that if $\sup_{x\in A} |g(x) - f(x)| < \delta(\epsilon)$ we have $|x_o - y_o| < \epsilon$, where $y_o = \arg\min_{x\in A} g(x)$.

Proof. Choose an arbitrary $\epsilon > 0$ and set $f_{\epsilon,0} = \min_{\|x-x_0|\geq\epsilon} f(x)$.
Now define $0 < f_{\epsilon,0} - f_o = 2\delta(\epsilon)$.

Note that we have

$$g(x) > f_{\epsilon,0} - \delta(\epsilon) \forall x \in A\backslash N \tag{3.5}$$

where $N = \{x : \|x - x_o\| < \epsilon\}$.

We also have

$$g(x_o) < f_o + \delta(\epsilon) < f_{\epsilon,0} - 2\delta(\epsilon) + \delta(\epsilon) = f_{\epsilon,0} - \delta(\epsilon). \tag{3.6}$$

Hence there exist points in N (of which x_n is one) where g assumes a value smaller than each value of g outside N. This proves that the minimum of g must be in N, and we are done.

□

Theorem 3.2 : Continuity. The function $f \mapsto \theta_f$, $f \in \mathcal{F}$ is continuous on $\sqrt{\mathcal{F}} = \left\{\sqrt{f}, f \in \mathcal{F}\right\}$.

Proof. We have the following figure.

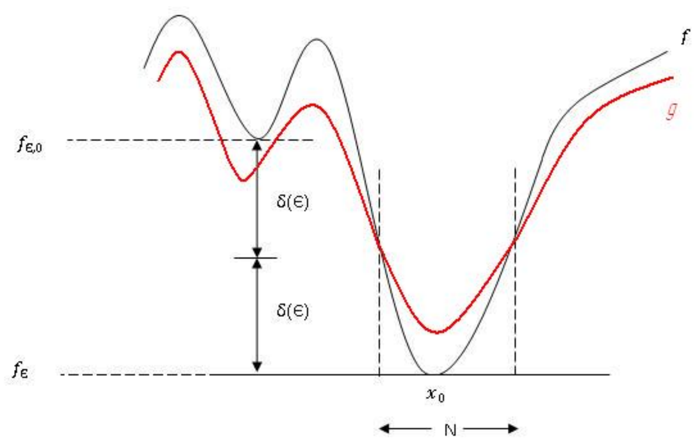

Figure 3.1: Epsilon Band for the Minimum of f.

We have,

$$\sup_{t\in\Theta} |\Delta_f(t) - \Delta_g(t)| \leq 2 \left\|\sqrt{f} - \sqrt{g}\right\|. \tag{3.7}$$

Since we can make $\left\|\sqrt{f} - \sqrt{g}\right\|$ arbitrarily small, Lemma 3.1 applies and yeilds the desired result. □

Theorem 3.3 : Boundedness.

$$\sup_{\theta\in\Theta}\left\|\hat{\Delta}_\theta\right\|^2 \xrightarrow{p} 0$$

Proof. Let the points $\theta_{1(\delta)},\ldots\theta_{j(\delta)}$ be chosen for each $\delta > 0$ such that

$$\bigcup_{j=1}^{J(\delta)} B(\theta_j,\delta) \supset \Theta$$

where Θ is compact.

Choose ϵ as follows.

$$\begin{cases} 0 < \epsilon \leq \psi(y) \leq M < \infty & : a \leq y \leq b \\ \varphi(y) = 0 & : elsewhere \end{cases}$$

with

$$\left|\acute{\Psi}(y)\right| \leq M \quad \forall y \in \mathbb{R}$$

$|r_\theta(x)| \leq M$ **and** $|\dot{r}_\theta(x)| \leq M \quad \forall x \in \mathbb{X}$ **and** $\forall\theta\in\Theta$

We have

$$\begin{aligned} \hat{\Delta}_\theta(y) &= \hat{f}_\theta(y) - f_\theta(y) \\ &= \int_{\mathbb{X}} \Psi(y - r_\theta(x))d\left\{\hat{Q}(x) - Q(x)\right\} \end{aligned}$$

$$\dot{\hat{\Delta}}_\theta(y) = \int \acute{\Psi}(y - r_\theta(x))\dot{r}_\theta(x)d\left\{\hat{Q}(x) - Q(x)\right\}.$$

Note that

$$\begin{aligned} \left|\dot{\hat{\Delta}}_\theta(y)\right| &\leq \int_{\mathbb{X}} \left|\acute{\Psi}(y - r_\theta(x))\right| |\dot{r}_\theta(x)|\, d\left\{\hat{Q}(x) + Q(x)\right\} \\ &\leq \int_{\mathbb{X}} M\cdot M\cdot d\left\{\hat{Q}(x) + Q(x)\right\} \\ &= 2M^2 \quad \forall y\in\mathbb{R} \quad \forall\theta\in\Theta. \end{aligned}$$

We have to show that $\mathbb{P}\left\{\sup_{\theta\in\Theta}\left\|\hat{\Delta}_\theta\right\| \geq \epsilon\right\} \leq \epsilon$ for $n \geq n(\epsilon)$.

We have

$$\begin{aligned}
\sup_{\theta\in\Theta}\left\|\hat{\Delta}_{\theta}\right\|^2 &= \max_j \sup_{\theta\in B(\theta_j,\delta)}\left\|\hat{\Delta}_{\theta}\right\|^2 \\
&\leq 2\max_j \sup_{|h|\leq\delta}\left\{\left\|\hat{\Delta}_{\theta_j}\right\|^2+\left\|\hat{\Delta}_{\theta_{j+h}}-\hat{\Delta}_{\theta_j}\right\|^2\right\} \\
&= 2\max_j\left\|\hat{\Delta}_{\theta_j}\right\|^2+2\max_j\sup_{|h|\leq\delta}\int\left\{h\cdot\hat{\Delta}_{\theta_{j(y,h)}}(y)\right\}^2 dy \\
&\leq 2\max_j\left\|\hat{\Delta}_{\theta_j}\right\|^2+2\max_j\sup_{|h|\leq\delta}h^2\int_{a-B}^{b+B}(2M^2)^2 dy \\
&\leq 2\max_j\left\|\hat{\Delta}_{\theta_j}\right\|^2+2\delta^2 4M^4(b-a+2B).
\end{aligned}$$

Choose $\epsilon > 0$ and $\delta = \delta(\epsilon)$ such that

$$8M^4(b-a+2B)\cdot\delta^2(\epsilon) < \frac{\epsilon}{2}.$$

Then

$$\begin{aligned}
\mathbb{P}\left\{2\max_j\left\|\hat{\Delta}_{\theta_j}\right\|^2+8M^4(b-a+2B)\delta^2(\epsilon)\geq\epsilon\right\} &\leq \mathbb{P}\left\{2\max_j\left\|\hat{\Delta}_{\theta_j}\right\|^2\geq\tfrac{\epsilon}{2}\right\} \\
&= \mathbb{P}\left\{\max_j\left\|\hat{\Delta}_{\theta_j}\right\|^2\geq\tfrac{\epsilon}{4}\right\} \\
&\leq \sum_{j=1}^{N(\delta(\epsilon))}\mathbb{P}\left\{\max_j\left\|\hat{\Delta}_{\theta_j}\right\|^2\geq\tfrac{\epsilon}{4}\right\}
\end{aligned}$$

.

Each of these finite probabilities ($N(\delta(\epsilon)$ in number) is smaller than

$$\frac{\epsilon}{4N(\delta(\epsilon))} \text{ for } n \geq n(\delta(\epsilon)).$$

Then

$$\mathbb{P}\left\{2\max_j \left\|\hat{\Delta}_{\theta_j}\right\|^2 + 8M^4(b-a+2B)\delta^2(\epsilon) \geq \epsilon\right\} \leq N(\delta(\epsilon)) \cdot \tfrac{\epsilon}{N(\delta(\epsilon))}$$

$$= \epsilon \text{ for } n \geq n(\delta(\epsilon)) = n(\epsilon)$$

□

3.2 Proposed Minimum Hellinger Distance Estimator

At this point let us introduce our estimator of $t \in T$. Suppose that the $Y_1, ..., Y_n$ have an arbitrary density f_θ for some $\theta \in \Theta$. Recall the notation in (2.8) , and let $\hat{f}$ be a nonparametric density estimator based on the sample of the Y_i as in (2.6). The second estimator is $\hat{f}_t$ as in (2.7).
We now define

$$\hat{\theta} = argmin_{t\in\Theta} \left\|\sqrt{\hat{f}} - \sqrt{\hat{f}_t}\right\|^2 . \tag{3.8}$$

In the present situation $\hat{f}$ should be close to f_θ, $\hat{f}_\theta$ is also close to f_θ and consequently we may expect $\hat{\theta}$ to be close θ.

We then prove the consistency and asymptotic normality of the proposed estimate in next section and in next chapter.

3.3 Consistency

Now we look at the consistency of the estimator $\hat{\theta}$

Theorem 3.4 If the $Y_1 .....Y_n$ are iid f_θ,$\theta \in \Theta$, then we have

$$\hat{\theta} \xrightarrow{p} \theta \text{ as n } \to \infty \quad \forall \theta \in \Theta. \tag{3.9}$$

Proof. Let us define

$$\hat{\Delta}_{\hat{f}}(t) = \left\| \sqrt{\hat{f}} - \sqrt{\hat{f}_t} \right\|^2 \tag{3.10}$$

and note that

$$\begin{aligned} \sup_{t \in \Theta} \left| \Delta_{\hat{f}}(t) \right| &= \sup_{t \in \Theta} 2 \left| \left\langle \sqrt{\hat{f}}, \sqrt{\hat{f}_t} - \sqrt{f_t} \right\rangle \right| \\ &\leq 2 \sup_{t \in \Theta} \left\| \sqrt{\hat{f}_t} - \sqrt{f_t} \right\| . \end{aligned} \tag{3.11}$$

It is known from the Uniform Law of Large Numbers (ULLN) that for a function $f(Y, \theta)$ defined for $\theta \in \Theta$ and continuous in θ, the sequence $\{f(Y_1, \theta), f(Y_2, \theta) \ldots\}$ will be a sequence of independent and identically distributed random variables such that the sample mean of this sequence converges in probability to $E\left(f(Y, \theta)\right)$.

Then from ULLN and Theorem 3.3 with equation (3.7), we get

$$\sup_{t \in \Theta} \left\| \sqrt{f_t} - \sqrt{\hat{f}_t} \right\| \xrightarrow{p} 0 \text{ as } n \to \infty \tag{3.12}$$

under suitable assumptions.
This means that, for any given $\delta > 0$, the subset

$$\Omega_{1,\delta} = \left\{ \omega \in \Omega : \sup_{t \in \Theta} \left\| \sqrt{\hat{f}_t} - \sqrt{f_t} \right\| \right\} < \frac{\delta}{4} \tag{3.13}$$

has probability

$$\mathbb{P}(\Omega_{1,\delta}) \to 1 \text{ as } n \to \infty \tag{3.14}$$

and therefore, for each $\omega \in \Omega_{1,\delta}$ we have

$$\sup_{t\in\Theta} \left| \Delta_{\hat{f}}(t,\omega) - \hat{\Delta}_{\hat{f}}(t,\omega) \right| < \frac{\delta}{2}. \tag{3.15}$$

Also note that

$$\begin{aligned} \sup_{t\in\Theta} \left| \Delta_{\hat{f}}(t) - \hat{\Delta}_{f_\theta}(t) \right| &\leq 2 \sup_{t\in\Theta} \left| \left\langle \sqrt{f_t}, \sqrt{\hat{f}} - \sqrt{f_\theta} \right\rangle \right| \\ &\leq 2 \left\| \sqrt{\hat{f}} - \sqrt{f_\theta} \right\| \end{aligned} \tag{3.16}$$

and we also know that

$$\left\| \sqrt{\hat{f}} - \sqrt{f_\theta} \right\| \xrightarrow{p} 0 \text{ as } n \to \infty \text{ in } L_2. \tag{3.17}$$

Hence there exist subsets

$$\Omega_{2,\delta} = \left\{ \omega \in \Omega : \left\| \sqrt{\hat{f_{(\omega)}} - \sqrt{f_\theta}} \right\| < \frac{\delta}{4} \right\} \tag{3.18}$$

have probability

$$\mathbb{P}(\Omega_{2,\delta}) \to 1, \text{ as } n \to \infty \tag{3.19}$$

and therefore, for each $\omega \in \Omega_{2,\delta}$ we have

$$\sup_{t\in\Theta} \left| \Delta_{\hat{f}}(t) - \Delta_{f_\theta}(t) \right| < \frac{\delta}{2}. \tag{3.20}$$

Combining (3.15),(3.20), and (3.14) and (3.19), we see that for each

$$\omega \in \Omega_{1,\delta} \cap \Omega_{2,\delta}, \tag{3.21}$$

we have

$$\sup_{t\in\Theta} \left| \Delta_{f_\theta}(t) - \hat{\Delta}_{\hat{f}}(t) \right| < \delta \tag{3.22}$$

and that

$$\mathbb{P}(\Omega_\delta) \to 1, \text{ as } n \to \infty. \tag{3.23}$$

Let us now choose an arbitrary $\epsilon > 0$, and observe that $\min_{t\in\Theta} \Delta_{f_\theta}(t) = \theta$.

In order to apply Lemma 3.1, let us choose

$$\delta(a) = \delta = \min_{t\in\Theta:\|t-\Theta\|>\epsilon} \Delta_{f_\theta}(t) - \theta \tag{3.24}$$

Application of the lemma with this ϵ and δ yeilds that for each $\omega \in \Omega_\delta$

$$\sup_{t\in\Theta} \left| \Delta_{f_\theta}(t) - \hat{\Delta}_{\hat{f}}(\omega, t) \right| < \delta \tag{3.25}$$

and hence $\left|\hat{\theta}(\omega) - \theta\right| < \epsilon$. This means that

$$\mathbb{P}\left\{\omega : \left|\hat{\theta}(\omega) - \theta\right| < \epsilon\right\} \geq \mathbb{P}(\Omega_\delta) \to 1 \text{ as } n \to \infty. \tag{3.26}$$

and the claim of the theorem follows. □

CHAPTER 4
ASYMPTOTIC DISTRIBUTION

4.1 Introduction

In this chapter, we discuss the asymptotic distribution of the proposed estimator $\hat{\theta}$ in (3.8). In addition to the consistency result of Theorem 3.4, we have the following smoothness assumption for s_θ.

Assumption 4.1 For each $\theta \in \Theta$, there exists $\ddot{s}_\theta \in L^2(\mathbb{R})$, such that

$$\dot{s}_t - \dot{s}_\theta = (t-\theta)\ddot{s}_\theta + o_{L^2}(|t-\theta|). \tag{4.1}$$

For $h = \theta - \hat{\theta}$, the notation $o_{L^2}(h)$ is as of the definition 3.1. We also assume $\dot{s}_t$ and $\ddot{s}_t$ to be continuous in $L^2(\mathbb{R})$ as functions of $t \in \Theta$.

Assumption 4.2 Let us write

$$K_\theta(y) = \frac{\dot{s}_t(y)}{2s_t(y)} \qquad y \in \mathbb{R}, t \in \Theta; \tag{4.2}$$

we will assume that $K_\theta(y)$ is bounded and has bounded ordinary derivatives on $y \in \mathbb{R}$ for each t.

In the following sections we will still assume that $Y_1, \ldots Y_n$ are iid with $f_t, t \in \Theta$. Recall that the estimator $\hat{\theta}$ is defined as

$$\begin{aligned}\hat{\theta} = \arg\min_{t\in\Theta} \hat{\Delta}_{\hat{f}}(t) &= \arg\min_{t\in\Theta} \left\| \sqrt{\hat{f}} - \sqrt{\hat{f}_t} \right\|^2 \\ &= \arg\min_{t\in\Theta} \langle \hat{s}, \hat{s}_t \rangle\end{aligned} \tag{4.3}$$

where $\hat{f}$ and $\hat{f}_t$ are defined in (2.6) and (2.7) respectively.

Similar to the Maximum Likelihood Estimator (MLE), MHD estimators also belong to the class of M-estimators, but they are minimizers (of the Hellinger Distance) rather than maximizers like MLE.

Investigating 4.3 in detail, we see

$$\begin{aligned}
0 &= \tfrac{d}{dt}\left(\hat{\Delta}_{\hat{f}}(\hat{\theta})\right) = \tfrac{d}{dt}\left\|\sqrt{\hat{f}} - \sqrt{\hat{f}_{\hat{\theta}}}\right\|^2 \\
&= \tfrac{d}{dt}\left(\left\|\sqrt{\hat{f}}\right\|^2 + \left\|\sqrt{\hat{f}_{\hat{\theta}}}\right\|^2 - 2\left\langle\sqrt{\hat{f}}, \sqrt{\hat{f}_{\hat{\theta}}}\right\rangle\right) \\
&= \tfrac{d}{dt}\left(2 - 2\left\langle\sqrt{\hat{f}}, \sqrt{\hat{f}_{\hat{\theta}}}\right\rangle\right) \\
&= -2\left\langle\sqrt{\hat{f}}, \tfrac{d}{dt}\sqrt{\hat{f}_{\hat{\theta}}}\right\rangle \\
&= -2\left\langle\sqrt{\hat{f}}, \dot{\hat{s}}_{\hat{\theta}}\right\rangle.
\end{aligned}$$

Therefore;

$$\left\langle\sqrt{\hat{f}}, \dot{\hat{s}}_{\hat{\theta}}\right\rangle = \left\langle\hat{s}, \dot{\hat{s}}_{\hat{\theta}}\right\rangle = 0. \tag{4.4}$$

Let us write $\dot{\hat{s}}_t$ and $\ddot{\hat{s}}_t$ for the L_2 derivatives of $\hat{s}_t$. Now using (4.4) , we get

$$\begin{aligned}
0 &= \left\langle\hat{s}, \dot{\hat{s}}_{\hat{\theta}}\right\rangle \\
&= \left\langle\hat{s}, \dot{\hat{s}}_{\theta} + (\hat{\theta}-\theta)\ddot{s}_{\theta} + o_{L^2}(\hat{\theta}-\theta)\right\rangle \\
&= \left\langle\hat{s}, \dot{\hat{s}}_{\theta} + (\hat{\theta}-\theta)\ddot{s}_{\theta} + \frac{(\hat{\theta}-\theta)}{(\hat{\theta}-\theta)} o_{L^2}(\hat{\theta}-\theta)\right\rangle
\end{aligned} \tag{4.5}$$

by expanding $\hat{\theta}$ at θ by Taylor Expansion.

The following Lemma discusses the remainder term of the above expansion.

Lemma 4.1 We have

$$\left\|o_{L^2}(\hat{\theta}-\theta)\right\| = \left|(\hat{\theta}-\theta)\right|.o_p(1) \quad \text{as } n \to \infty. \tag{4.6}$$

Proof. Fix an arbitrary $\epsilon > 0$. According to definition of $o_{L^2}(h)$ there exists $\delta(\epsilon) > 0$ such that if $0 < |h| < \delta(\epsilon)$, then

$$\frac{\|o_{L^2}(h)\|}{|h|} < \epsilon. \tag{4.7}$$

The consistancy in Theorem 3.3 entails that there exists an index $n(\epsilon)$such that

$$\mathbb{P}(\Omega_{n,\epsilon}) = \mathbb{P}\left\{\omega \in \Omega : \left|\hat{\theta}(\omega)\theta\right| < \delta(\epsilon)\right\} \geq 1 - \epsilon. \tag{4.8}$$

Note that the subspace $\Omega_{n,0}$, where $\hat{\theta} = \theta$ satisfies

$$\Omega_{n,0} = \left\{\omega \in \Omega : \hat{\theta}(\omega) = \theta\right\} \subset \Omega_{n,\epsilon},$$

thus for each $\omega \in \Omega_{n,0}$, relation (4.5) holds trivially true. For $\omega \in \Omega_{n,\epsilon} \setminus \Omega_{n,0}$, we have

$$\Omega_{n,\epsilon} \supset \left\{\omega \in \Omega : \left|\hat{\theta}(\omega) - \theta\right| < \delta(\epsilon), \hat{\theta}(\omega) \neq \theta\right\} = \Omega_{n,\epsilon} \setminus \Omega_{n,0}.$$

Thus by (4.7), we have

$$0 < \left|\hat{\theta}(\omega) - \theta\right| < \delta(\epsilon) \Rightarrow \frac{\left\|o_{L^2}\hat{\theta}(\omega) - \theta\right\|}{\left|\hat{\theta}(\omega) - \theta\right|} \leq \epsilon. \tag{4.9}$$

A combination of the above result yeilds that

$$\mathbb{P}\left\{\frac{\left\|o_{L^2}(\hat{\theta} - \theta)\right\|}{\left|\hat{\theta} - \theta\right|} \geq \epsilon\right\} \leq \mathbb{P}\left\{\Omega_{n,\epsilon}^c\right\} \leq \epsilon, \forall n \geq n(\epsilon). \tag{4.10}$$

□

Returning to (4.5) we obtain

$$\hat{\theta} - \theta = -\frac{\left\langle \hat{s}, \hat{\dot{s}}_\theta \right\rangle}{\langle \hat{s}, \ddot{s}_\theta \rangle + \left\langle \hat{s}, \frac{o_{L^2}(\hat{\theta} - \theta)}{\hat{\theta} - \theta} \right\rangle}. \tag{4.11}$$

with help of Lemma 4.1, we get

$$\left|\left\langle \hat{s}, \frac{o_{L^2}(\hat{\theta} - \theta)}{\hat{\theta} - \theta} \right\rangle\right| \leq \|\hat{s}\| .o_p(1) \xrightarrow{P} 0 \text{ as } n \to \infty.$$

It follows from (4.11) that

$$\begin{aligned}\sqrt{n}(\hat{\theta}-\theta) &= -\sqrt{n}\frac{\left\langle \hat{s}, \dot{\hat{s}}_\theta \right\rangle}{\langle \hat{s}, \ddot{s}_\theta\rangle + \left\langle \hat{s}, \frac{o_{L^2}(\hat{\theta}-\theta)}{\hat{\theta}-\theta} \right\rangle} \\ &= -\sqrt{n}\frac{\left\langle \hat{s}, \dot{\hat{s}}_\theta \right\rangle}{\langle \ddot{s}_\theta, \hat{s}-s_\theta+s_\theta\rangle + \left\langle \hat{s}, \frac{o_{L^2}(\hat{\theta}-\theta)}{\hat{\theta}-\theta} \right\rangle} \\ &= -\sqrt{n}\frac{\left\langle \hat{s}, \dot{\hat{s}}_\theta \right\rangle}{\langle \ddot{s}_\theta, s_\theta + o_p(1)\rangle + o_p(1)} \\ &= \sqrt{n}\frac{\left\langle \hat{s}, \dot{\hat{s}}_\theta \right\rangle}{-\langle \ddot{s}_\theta, s_\theta\rangle} + o_p(1).\sqrt{n}\frac{\left\langle \hat{s}, \dot{\hat{s}}_\theta \right\rangle}{-\langle \ddot{s}_\theta, s_\theta\rangle}.\end{aligned} \tag{4.12}$$

by the fact that $\dfrac{a}{b+o_p(1)} = \dfrac{a}{b} + 0_p(1).\dfrac{a}{b}$.

It suffices to prove the asymptotic normality of the first term on the right in the last line of (4.12).

Because $\langle s_\theta, \dot{s}_\theta\rangle = 0$, we have

$$\begin{aligned}\left\langle \hat{s}, \dot{\hat{s}}_\theta \right\rangle &= \left\langle \hat{s}-s_\theta, \dot{\hat{s}}_\theta \right\rangle + \left\langle s_\theta, \dot{\hat{s}}_\theta - \dot{s}_\theta \right\rangle \\ &= \langle \hat{s}-s_\theta, \dot{s}_\theta\rangle + \left\langle \hat{s}-s_\theta, \dot{\hat{s}}_\theta - \dot{s}_\theta \right\rangle + \left\langle s_\theta, \dot{\hat{s}} - \dot{s}_\theta \right\rangle\end{aligned} \tag{4.13}$$

Let us set, for brevity,

$$S_n = \sqrt{n}\left\langle \hat{s} - s_\theta, \dot{s}_\theta \right\rangle. \tag{4.14}$$

$$T_n = \sqrt{n}\left\langle s_\theta, \hat{\dot{s}} - \dot{s}_\theta \right\rangle. \tag{4.15}$$

$$r_n = \sqrt{n}\left\langle \hat{s} - s_\theta, \hat{\dot{s}}_\theta - \dot{s}_\theta \right\rangle. \tag{4.16}$$

We can now write

$$\sqrt{n}(\hat{\theta} - \theta) = -\frac{1}{\langle s_\theta, \ddot{s}_\theta \rangle}\left(S_n + T_n + r_n\right).$$

We can show that the remainder term, r_n attains zero as follows.

$$r_n = \sqrt{n}\left\langle \hat{s} - s_\theta, \hat{\dot{s}}_\theta - \dot{s}_\theta \right\rangle = \sqrt{n}\left\langle \hat{s} - s_\theta, \frac{\hat{\dot{f}}_\theta}{2\hat{s}_\theta} - \frac{\dot{f}_\theta}{2s_\theta} \right\rangle.$$

We have

$$\frac{\hat{\dot{f}}_\theta}{2\hat{s}_\theta} \xrightarrow{P} \frac{\dot{f}_\theta}{2s_\theta} \Rightarrow o_P(1).$$

Now

$$|r_n| = \left| \sqrt{n} \int \hat{s} - s_\theta . \frac{\hat{\dot{f}}_\theta}{2\hat{s}_\theta} - \frac{\dot{f}_\theta}{2s_\theta} \right| \leq \sqrt{n} \sqrt{\int (\hat{s} - s_\theta)^2 . \int \left(\frac{\hat{\dot{f}}_\theta}{2\hat{s}_\theta} - \frac{\dot{f}_\theta}{2s_\theta} \right)^2}$$

by Cauchy-Swartz inequality and

$$|r_n| = O_P(1) o_P(1) \to 0 \text{ as } n \to \infty \tag{4.17}$$

(See Theorem 3.2.1 in Wang, J (2009))

In the following sections, we will prove the main theorems of this thesis, the asymptotic normality of the MHD estimator. This will be obtained in several steps; sections 4.2 and 4.3 discuss the asymptotic normality of the two terms S_n and T_n, then section 4.4 combines these results to obtain the asymptotic normality of the MHD estimator.

4.2 Asymptotic Normality of S_n Term

In this section, the asymptotic normality of S_n term is derived. Recall that from 4.14,

$$S_n = \sqrt{n}\left\langle \hat{s} - s_\theta, \dot{s}_\theta \right\rangle,$$

and from (2.6)

$$\hat{f}_{\sigma(n)}(y) = \frac{1}{n}\sum_{i=1}^{n} \varphi_{\sigma(n)}(y - Y_i)$$

for σ small.

Also

$$\mathbb{E}\hat{f}_\sigma = f_\sigma, \qquad \hat{s}_{\sigma(n)}(y) = \sqrt{\hat{f}_{\sigma(n)}(y)} \tag{4.18}$$

Theorem 4.1 Under assumptions $3.1, 3.2, 4.1$ and 4.2, we have

$$S_n \to^d N(0, \upsilon_{1,\theta}^2),$$

where

$$\upsilon_{1,\theta}^2 = \mathbb{E}\left(\frac{\dot{s}_\theta}{2s_\theta}\right)^2,$$

provided that in 2.6, we have $\sigma = \sigma(n) \quad n^{-\delta}$ as $n \to \infty$ for any $\frac{1}{4} < \delta < \frac{1}{2}$ and $4\left\|\dot{s}_\theta\right\|^2 > 0$.

Proof. Note that $S_n = \sqrt{n}\left\langle \dot{s}_\theta, \hat{s} - s_\theta \right\rangle = \sqrt{n}\left\langle \dot{s}_\theta, \sqrt{\hat{f}_{\sigma(n)}} - s_\theta \right\rangle$

$$
\begin{aligned}
S_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int \dot{s}_\theta(y) \left\{ \sqrt{\varphi_{\sigma(n)}(y - Y_i)} - s_\theta(y) \right\} dy \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int \dot{s}_\theta(y) \frac{\left\{ \sqrt{\varphi_{\sigma(n)}(y - Y_i)} - s_\theta(y) \right\}}{\left\{ \sqrt{\varphi_{\sigma(n)}(y - Y_i)} + s_\theta(y) \right\}} . \left\{ \sqrt{\varphi_{\sigma(n)}(y - Y_i)} + s_\theta(y) \right\} dy \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int \dot{s}_\theta(y) \frac{\left\{ \varphi_{\sigma(n)}(y - Y_i) - f_\theta(y) \right\}}{\left\{ \sqrt{\varphi_{\sigma(n)}(y - Y_i)} + s_\theta(y) \right\}} dy \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int \dot{s}_\theta(y) \frac{\left\{ \varphi_{\sigma(n)}(y - Y_i) - f_\theta(y) \right\}}{2 s_\theta(y)} . \frac{2 s_\theta(y)}{\left\{ \sqrt{\varphi_{\sigma(n)}(y - Y_i)} + s_\theta(y) \right\}} dy \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int \dot{s}_\theta(y) \frac{\left\{ \varphi_{\sigma(n)}(y - Y_i) - f_\theta(y) \right\}}{2 s_\theta(y)} . \\
&\quad \frac{s_\theta(y) + \sqrt{\varphi_{\sigma(n)}(y - Y_i)} + s_\theta(y) - \sqrt{\varphi_{\sigma(n)}(y - Y_i)}}{s_\theta(y) + \sqrt{\varphi_{\sigma(n)}(y - Y_i)}} dy \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int \dot{s}_\theta(y) \frac{\left\{ \varphi_{\sigma(n)}(y - Y_i) - f_\theta(y) \right\}}{2 s_\theta(y)} \\
&\quad . \left(1 + \frac{s_\theta(y) - \sqrt{\varphi_{\sigma(n)}(y - Y_i)}}{s_\theta(y) + \sqrt{\varphi_{\sigma(n)}(y - Y_i)}} \right) dy \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int \dot{s}_\theta(y) \frac{\left\{ \varphi_{\sigma(n)}(y - Y_i) - f_\theta(y) \right\}}{2 s_\theta(y)} dy \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int \dot{s}_\theta(y) \frac{\left\{ \varphi_{\sigma(n)}(y - Y_i) - f_\theta(y) \right\}}{2 s_\theta(y)} . \left(\frac{s_\theta(y) - \sqrt{\varphi_{\sigma(n)}(y - Y_i)}}{\left\{ \sqrt{\varphi_{\sigma(n)}(y - Y_i)} + s_\theta(y) \right\}} \right) dy \\
&= S_{n,1} + S_{n,2}
\end{aligned}
$$

$S_{n,1}$ can be further decomposed into

$$S_{n,1} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int \dot{s}_\theta(y)\frac{\varphi_{\sigma(n)}(y-Y_i) - f_{\sigma(n)}(y)}{2s_\theta(y)}dy + \sqrt{n}\int \dot{s}_\theta(y)\frac{f_{\sigma(n)}(y) - f_\theta(y)}{2s_\theta(y)}dy. \tag{4.19}$$

Then by (2.9) and with the condition $\sigma\ n^{-\delta}$, we have

$$\int (f_{\sigma(n)}(y) - f_\theta(y))^2 = \int bias^2(\hat{f}_{\sigma(n)}) = \mathcal{O}(n^{-\delta}). \tag{4.20}$$

Observe that, by Schwarz inequality and (4.2) in Assumption 4.2, we have

$$\begin{aligned} \left|\sqrt{n}\int \dot{s}_\theta(y)\frac{f_{\sigma(n)}(y) - f_\theta(y)}{2s_\theta(y)}dy\right| &\leq \sqrt{\int (\frac{\dot{s}_\theta}{2s_\theta}(y))^2 . n\int (f_{\sigma(n)} - f_\theta)^2} \\ &= \sqrt{\int K_\theta^2}.\mathcal{O}(n^{(1-4\delta)/2}) \\ &\to 0, \quad \text{as n} \to \infty, \end{aligned} \tag{4.21}$$

since $\delta > 1/4$ by assumption.
The $S_{n,2}$ term which is another remainder term atains zero (4.2.18, Wang), hence it is sufficient to prove the normality of the first term of $S_{n,1}$ i.e,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int \dot{s}_\theta\frac{\varphi_{\sigma(n)}(y-Y_i) - f_{\sigma(n)}(y)}{2s_\theta(y)}dy \xrightarrow{d} N(0, \upsilon_{1,\theta}^2) \text{ as } n \to \infty. \tag{4.22}$$

To show that, let us write the expression on the left as

$$\begin{aligned} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int \dot{s}_\theta\frac{\varphi_{\sigma(n)}(y-Y_i) - f_{\sigma(n)}(y)}{2s_\theta(y)}dy &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int \frac{\dot{s}_\theta}{2s_\theta(y)}\left\{\varphi_{\sigma(n)}(y-Y_i) - f_{\sigma(n)}(y)\right\}dy \\ &= X_n = \tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^{n} Z_{n,i} \end{aligned}$$

.

(4.23)

According to 4.17, $\mathbb{E}(Z_{n,i}) = 0 \; \forall \; i, n$.

The generic term

$$Z_n = \int \frac{\dot{s}_\theta}{2s_\theta}(y)\varphi_{\sigma(n)}(y - Y) - f_{\sigma(n)}(y)dy \tag{4.24}$$

is bounded by Cauchy-Schwarz inequality.

Thus for an arbitrary $\delta > 0$, we have

$$\mathbb{E}\left|Z_n\right|^{2+\delta} \leq 2 \sup \frac{\dot{s}_\theta(y)^{2+\delta}}{2s_\theta(y)} < \infty, \forall \sigma = \sigma(n). \tag{4.25}$$

Let $Z_{n,1}, Zn, 2, \ldots, Z_{n,n}$ be independent random variables with the same distribution as Z_n. It follows from Lyapanov's Central Limit Theorem, the asymptotic normality would be obtained if the variance of X_n in (4.20) satisfies

$$\mathbb{V}ar(X_n) = \mathbb{E}(Z_n^2) \to \mathbb{E}\left(\frac{\dot{s}_\theta}{2s_\theta}\right)^2 (y) \tag{4.26}$$

and if there exists $\delta > 0$ such that

$$\lim_{n\to\infty} \frac{1}{\mathbb{V}ar(X_n)^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}\left|\frac{1}{\sqrt{n}} Z_{n,i}\right|^{2+\delta} = \lim_{n\to\infty} \frac{1}{\mathbb{E}(Z_n^2)^{1+\delta/2}} \frac{1}{n^{1+\delta/2}} n\mathbb{E}\left|Z_n\right|^{2+\delta} = 0 \tag{4.27}$$

(4.26) is true if (4.25) is true. We have

$$\mathbb{V}ar(\frac{\dot{s}_\theta}{2s_\theta})(y) = \mathbb{E}(\frac{\dot{s}_\theta}{2s_\theta}(y))^2 = \upsilon_{1,\theta}^2$$

because $\mathbb{E}(\frac{\dot{s}_\theta}{2s_\theta})(y) = \mathbb{E}K_\theta(y) = 0$. To see this, note that

$$\begin{aligned} \mathbb{E}K_\theta(y) &= \int \frac{\dot{s}_\theta(y)}{2s_\theta(y)} . f_\theta(y)dy \\ &= \frac{1}{2}\int \dot{s}_\theta(y)s_\theta(y)dy \\ &= \frac{1}{2}\left\langle \dot{s}_\theta, s_\theta \right\rangle = 0 \end{aligned}$$

Take

$$\begin{aligned}\mathbb{V}ar(X_n) &= \mathbb{V}ar\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int \frac{\dot{s}_\theta}{2s_\theta}(y)\left\{\varphi_{\sigma(n)}(y-Y_i) - f_{\sigma(n)}(y)\right\}dy \\ &= \frac{1}{n}(n)\mathbb{V}ar\int K_\theta(y)\left\{\varphi_{\sigma(n)}(y-Y_i) - f_{\sigma(n)}(y)\right\}dy\end{aligned}$$

We can then write

$$\begin{aligned}\mathbb{V}ar(X_n) &= \mathbb{E}\left[\int K_\theta(y)\left\{\varphi_{\sigma(n)}(y-Y_i) - f_{\sigma(n)}(y)\right\}dy\right]^2 \\ &= \int_y\int_w\int_z K_\theta(w)K_\theta(z)\left\{\varphi_{\sigma(n)}(w-y) - f_{\sigma(n)}(w)\right\}. \\ &\qquad \left\{\varphi_{\sigma(n)}(z-y) - f_{\sigma(n)}(z)\right\}dwdzf(y)dy \\ &= \int_y\int_w\int_z K_\theta(w)K_\theta(z)\left\{\varphi_{\sigma(n)}(w-y)\right\}\left\{\varphi_{\sigma(n)}(z-y)\right\}dwdzf(y)dy - \\ &\qquad \int_y\int_w\int_z K_\theta(w)K_\theta(z)f_{\sigma(n)}(w)f_{\sigma(n)}(z)dwdzf(y)dy \\ &= \int_y K_\theta^2(y)f(y)dy - \left\{\int_y K_\theta(y)f(y)dy\right\}^2\end{aligned}$$

Since $\mathbb{E}(K_\theta(y)) = 0$, we obtain the desired result.

4.3 Asymptotic Normality of T_n Term

We will now focus on T_n term. Exploiting the notation in (2.8), it is easy to see that

$$\begin{aligned} T_n &= \left\langle \frac{\hat{\dot{f}}_\theta}{2\hat{s}_\theta} - \frac{\dot{f}_\theta}{2s_\theta}, s_\theta \right\rangle \\ &= \sqrt{n}\left\langle \frac{\hat{\dot{f}}_\theta - \dot{f}_\theta}{2s_\theta}, s_\theta \right\rangle + \sqrt{n}\left\langle \hat{\dot{f}}_\theta \frac{s_\theta - \hat{s}_\theta}{2s_\theta\hat{s}_\theta}, s_\theta \right\rangle \\ &= T_{n,1} + T_{n,2} \end{aligned} \tag{4.28}$$

Let us now prove that $f_\theta(y) = 0 \Rightarrow \dot{f}_\theta(y) = 0$, meaning that $\dfrac{\dot{f}_\theta(y)}{\sqrt{f_\theta(y)}}$ cannot be of the form $\dfrac{\neq 0}{0} = \pm\infty$.

To see this, introduce the set

$$D_{\theta,y} = \{x \notin \mathbb{X} : y - b \leq r_\theta(x) \leq y - a\}, \tag{4.29}$$

i.e the set of all x where $\psi(y - r_\theta(x)) > 0$.

Note that $x \notin D_{\theta,y} \Rightarrow \psi(y - r_\theta(x)) > 0$.

In fact, $\psi(y - r_\theta(x)) > \epsilon \forall x \in_{\theta,y}$ by assumption on ψ.

It follows that

$$f_\theta(y) = \int_{\mathbb{X}} \psi(y - r_\theta(x)dQ(x) = 0 \Rightarrow Q(D_{\theta,y}) = 0 \tag{4.30}$$

Because

$$D'_{\theta,y} = \left\{x \in \mathbb{X} : \left|\Psi'(y - r_\theta(x))\dot{r_\theta}(x)\right| > 0\right\} \subset D_{\theta,y}$$

by the assumption Ψ' is bounded away from zero on the same set $[a, b]$ and zero outside this set, we must also have

$$Q(D'_{\theta,y}) = 0 \Rightarrow \int \left| \Psi'(y - r_\theta(x)) \dot{r_\theta}(x) \right| dQ(x) = 0 \Rightarrow \dot{f}_\theta(y) = 0 \tag{4.31}$$

Same reasoning for $\hat{f}_\theta(y)$ and $\dot{\hat{f}}_\theta(y)$ can be employed by replacing Q with $\hat{Q}$. Therefore, we will interpret this quatient always as $\dfrac{\dot{f}_\theta(y)}{\sqrt{f_\theta(y)}} = \phi$ if $f(y) = 0$.

Now we look at the asymptotic distributions of $T_{n,1}$ and $T_{n,2}$ terms in (4.27). First recall that according to (2.7), $T_{n,1}$ can be expressed as,

$$\hat{f}_\theta(y) = \frac{1}{n} \sum_{i=1}^{n} \varphi(y - r_t(x_i))$$

Then the derivative is

$$\dot{\hat{f}}_\theta(y) = \frac{1}{n} \sum_{i=1}^{n} \acute{\varphi}(y - r_t(x_i)) \dot{r}_t(x_i) \tag{4.32}$$

Now we can rewrite $T_{n,1}$ as;

$$\begin{aligned} T_{n,1} &= \int_y \frac{\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \acute{\varphi}(y - r_t(x_i)) \dot{r}_t(x_i) - \sqrt{n} \int_{\mathbb{X}} \acute{\varphi}(y - r_t(x_i)) \dot{r}_t(x) dQ(x) \right\}}{2\sqrt{f_\theta(y)}} \sqrt{f_\theta(y)} dy \\ &= \int_y \left\{ \frac{1}{2\sqrt{n}} \sum_{i=1}^{n} \acute{\varphi}(y - r_t(x_i)) \dot{r}_t(x_i) - \frac{n}{2} \int_{\mathbb{X}} \acute{\varphi}(y - r_t(x_i)) \dot{r}_t(x) dQ(x) \right\} dy \\ &= \frac{1}{2\sqrt{n}} \sum_{i=1}^{n} \int_y \acute{\varphi}(y - r_t(x_i)) \dot{r}_t(x_i) dy - \frac{n}{2} \int_y \int_{\mathbb{X}} \acute{\varphi}(y - r_t(x_i)) \dot{r}_t(x) dQ(x) dy \end{aligned} \tag{4.33}$$

Since both $\hat{f}_\theta(y)$ and $f_\theta(y)$ are densities, the two terms it is clear that $T_{n,1}$ term atains zero.

The next theorem discusses the asymptotic normality of the $T_{n,2}$ term.

Theorem 4.2 Under assumptions 4.1 and 4.2, we have

$$T_{n,2} \to^d N(0, \upsilon_{2,\theta}^2),$$

where

$$\upsilon_{1,\theta}^2 = \mathbb{E}\left(\frac{\dot{s}_\theta \mathbb{G}}{2s_\theta}\right)^2,$$

and $\mathbb{G}$ is a Gaussian process in $L^2(\mathbb{R})$ with zero mean and covariance structure given by

$$\int_{\mathbb{X}} \hat{\dot{f}}_\theta(x,y)\hat{\dot{f}}_\theta(x,z)dQ(x)$$

Proof. Regarding $T_{n,2}$ note that,

$$\begin{aligned}
T_{n,2} &= \sqrt{n}\left\langle \hat{\dot{f}}_\theta \frac{s_\theta - \hat{s}_\theta}{2s_\theta\hat{s}_\theta}, s_\theta \right\rangle = -\frac{1}{2}\left\langle \sqrt{n}(\hat{s}_\theta - s_\theta), \frac{\hat{\dot{f}}_\theta}{\hat{s}_\theta} \right\rangle \\
&= -\frac{1}{2}\left\langle \sqrt{n}\frac{(\hat{s}_\theta^2 - s_\theta^2)}{\hat{s}_\theta + s_\theta}, \frac{\hat{\dot{f}}_\theta}{\hat{s}_\theta} \right\rangle \\
&= -\frac{1}{2}\left\langle \sqrt{n}\frac{\left(\hat{f}_\theta - f_\theta\right)}{\hat{s}_\theta + s_\theta}, \frac{\hat{\dot{f}}_\theta}{\hat{s}_\theta} \right\rangle \\
&= -\frac{1}{2}\left\langle \sqrt{n}(\hat{f}_\theta - f_\theta), \frac{\hat{\dot{f}}_\theta}{\hat{s}_\theta(\hat{s}_\theta + s_\theta)} \right\rangle
\end{aligned} \tag{4.34}$$

Again according to CLT in Hilbert Space,

$$\sqrt{n}(\hat{f}_\theta - f_\theta) \xrightarrow{d} \mathbb{G}, \text{ as } n \to \infty, \text{ in } L^2(\mathbb{R}) \tag{4.35}$$

and

$$\frac{\hat{\dot{f}}_\theta}{\hat{s}_\theta(\hat{s}_\theta + s_\theta)} \xrightarrow{p} \frac{\dot{f}_\theta}{2s_\theta^2} = \frac{1}{2}\frac{\dot{f}_\theta}{f_\theta}, \text{ in } L^2(\mathbb{R}) \tag{4.36}$$

To see this, note that

$$\hat{\dot{f}}_\theta = 2\hat{s}_\theta\hat{\dot{s}}_\theta$$

and

$$\frac{\hat{\dot{f}}_\theta}{\hat{s}_\theta(\hat{s}_\theta + s_\theta)} = \frac{2\hat{\dot{s}}_\theta}{\hat{s}_\theta + s_\theta}$$

and also

$$\dot{f}_\theta = 2s_\theta \dot{s}_\theta$$

and

$$\frac{\dot{f}_\theta}{2s_\theta^2} = \frac{\dot{s}_\theta}{s_\theta}$$

Therefore,

$$\frac{\hat{\dot{f}}_\theta}{\hat{s}_\theta(\hat{s}_\theta + s_\theta)} - \frac{\dot{f}_\theta}{2s_\theta^2} = \hat{\dot{s}}_\theta . \frac{2}{\hat{s}_\theta + s_\theta} - \dot{s}_\theta . \frac{1}{s_\theta}$$

By (4.35), $\sqrt{n}\left(\hat{\dot{f}}_\theta - \dot{f}_\theta\right) \xrightarrow{d} \dot{\mathbb{G}}$ as $n \to \infty$ and hence $\sqrt{n}\left(\hat{\dot{s}}_\theta - \dot{s}_\theta\right) \xrightarrow{d} \dot{\mathbb{G}}^2$ and

$$\left(\hat{\dot{s}}_\theta - \dot{s}_\theta\right) \xrightarrow{P} 0 \text{ as } n \to \infty \tag{4.37}$$

because $\sqrt{n}(\hat{\dot{f}}_\theta - \dot{f}_\theta) = O_p(1)$ and $(\hat{\dot{f}}_\theta - \dot{f}_\theta) = O_p(\frac{1}{\sqrt{n}}) = o_P(1)$

We know if $X \xrightarrow{p} X_n$ and $Y \xrightarrow{p} Y_n$ then $XY \xrightarrow{p} X_nY_n$ for functions. Hence if we can show

$$\left(\frac{2}{\hat{s}_\theta + s_\theta} - \frac{1}{s_\theta}\right) \xrightarrow{P} 0$$

then we are done.

To prove this, let us write

$$\frac{2}{\hat{s}_\theta + s_\theta} - \frac{1}{s_\theta} = \frac{2s_\theta - \hat{s}_\theta - s_\theta}{s_\theta(\hat{s}_\theta + s_\theta)} = \frac{s_\theta - \hat{s}_\theta}{s_\theta(\hat{s}_\theta + s_\theta)}$$

Now

$$\frac{|\hat{s}_\theta - s_\theta|}{s_\theta(\hat{s}_\theta + s_\theta)} . \frac{\hat{s}_\theta + s_\theta}{\hat{s}_\theta + s_\theta} = \frac{\left|\hat{f}_\theta - f_\theta\right|}{s_\theta(\hat{s}_\theta + s_\theta)^2} \leq \frac{\left|\hat{f}_\theta - f_\theta\right|}{s_\theta(\hat{f}_\theta + f_\theta)} \tag{4.38}$$

We know that Ψ has supprt $[a, b]$. Let us now assume that r_θ is bounded (uniformly for all θ); i.e

$$|r_\theta x| \leq M < \infty \forall x \in \mathbb{X}$$

Then we have

$$\varphi(y) = 0 \forall y \notin [a,b] \varphi(y - r_\theta(x)) = 0 \text{ if } y - r_\theta(x) \notin [a,b] \Rightarrow \varphi(y - r_\theta(x)) = 0$$

Certainly

$$\forall x \text{ if } y \notin [a-M, b+M]$$

Indeed

$$y < a - M \Rightarrow y + B < a \Rightarrow y - r_\theta(x) < a \text{ etc.}$$

$$f_\theta(y) = \int_{\mathbb{X}} \Psi(y - r_\theta(x)) dQ(x)$$

hence,

$$f_\theta(y) = 0$$

for $y \notin [a-M, b+M]$
Consequently

$$[a-M, b+M] \supset N_{\theta,\delta} = \{y : 0 < f_\theta(y) < \delta\}$$

Clearly, $N_{\theta,\delta} \downarrow \phi$ as $\delta \downarrow 0$, and hence $\lambda(N_{\theta,\delta}) \downarrow 0$ as $\delta \downarrow 0$ where λ is the Lebesgue measure.

Consider $\dfrac{\left|\hat{f}_\theta - f_\theta\right|}{\hat{f}_\theta + f_\theta}$ in (4.22) and

$$\left\{ \frac{\left|\hat{f}_\theta(y) - f_\theta(y)\right|}{\hat{f}_\theta(y) + f_\theta(y)} \left(1_{N_{\theta,\delta}}(y) + 1^c_{N_{\theta,\delta}}(y) \right) \right\}^2 \leq 2 1_{N_{\theta,\delta}}(y) + \frac{2}{\delta} \left| \hat{f}_\theta(y) - f_\theta(y) \right|$$

Now

$$\mathbb{E} \int \left\{ \frac{\left|\hat{f}_\theta(y) - f_\theta(y)\right|}{\hat{f}_\theta(y) + f_\theta(y)} \right\}^2 dy \leq 2\lambda(N_{\theta,\delta}) + \frac{2}{\delta} \int \mathbb{E} \left\{ \hat{f}_\theta(y) - f_\theta(y) \right\}^2 dy$$

Let $\Delta_n = \int \mathbb{E}\left\{\hat{f}_\theta(y) - f_\theta(y)\right\}^2 dy$

Choose $\delta = \delta(\epsilon) : \lambda(N_{\theta,\delta}) < \dfrac{\epsilon}{2}$

For this $\delta(\epsilon)$, choose $n_1 = n_1(\delta(\epsilon)) : \dfrac{2}{\delta(\epsilon)}\Delta_n < \dfrac{\epsilon}{2} \quad \forall n > n_1$

Considering $\dfrac{1}{s_\theta}\dfrac{\left|\hat{f}_\theta - f_\theta\right|}{\hat{f}_\theta + f_\theta}$ in (4.22),

$$\mathbb{E}\int\left(\frac{1}{s_\theta}\frac{\hat{f}_\theta - f_\theta}{\hat{f}_\theta + f_\theta}\right)^2 dy = \int \frac{1}{f_\theta}\mathbb{E}\left(\frac{\hat{f}_\theta - f_\theta}{\hat{f}_\theta + f_\theta}\right)^2 dy$$

Let $\mathbb{E}\left(\dfrac{\hat{f}_\theta - f_\theta}{\hat{f}_\theta + f_\theta}\right)^2 = g_\theta$

Then

$$\mathbb{E}\int\left(\frac{1}{s_\theta}\frac{\hat{f}_\theta - f_\theta}{\hat{f}_\theta + f_\theta}\right)^2 dy = \int \frac{1}{f_\theta} g_\theta dy = \int_{N_{\theta,\delta}} \frac{1}{f_\theta} g_\theta dy + \int_{N^c_{\theta,\delta}} \frac{1}{f_\theta} g_\theta dy \tag{4.39}$$

Assuming $\int \dfrac{1}{f_\theta y} dy < \infty$, there exists $\delta(\epsilon)$ small enough such that;

$$\int_{N_{\theta,\delta(\epsilon)}} \frac{1}{f_\theta} g_\theta dy \le \int_{N_{\theta,\delta}} \frac{1}{f_\theta} g_\theta dy < \frac{\epsilon}{2}$$

Consequently;

$$T_{n,2} \xrightarrow{d} -\frac{1}{4}\left\langle \mathbb{G}, \frac{\dot{f}_\theta}{f_\theta}\right\rangle = N(0, \upsilon^2_{2,\theta})$$

where

$$\upsilon^2_{2,\theta} = \mathbb{E}\left(\frac{\dot{f}_\theta \mathbb{G}}{2\hat{s}_\theta}\right)^2$$

and we obtain the result in the theorem. □

Now we have reached the point of obtaining the limiting distribution of MHD estimator by combining the results of Theorem 4.1 and Theorem 4.2.

4.4 Asymptotic Normality of the MHD Estimator

This section is reserved for the main result of this thesis, deriving the asymptotic normality of the MHD estimator. The following notations are introduced first as we will be using them in the proof. We write the $T_{n,2}$ term in an alternative way as follows using (2.7) and (4.32).

$$
\begin{aligned}
T_{n,2} &= \sqrt{n}\left\langle \hat{\dot{f}}_\theta \frac{s_\theta - \hat{s}_\theta}{2 s_\theta \hat{s}_\theta}, s_\theta \right\rangle \\
&= \sqrt{n}\int_y \left\{ \frac{1}{n}\sum_{i=1}^{n} \acute{\varphi}(y - r_t(x_i))\dot{r}_t(x_i). \right. \\
&\qquad \frac{\left(\sqrt{\int_{\mathbb{X}} \varphi(y - r_t(x))dQ(x)} - \sqrt{\frac{1}{n}\sum_{i=1}^{n}\varphi(y - r_t(x_i))}\right)}{2 s_\theta \hat{s}_\theta}.s_\theta dy \\
&= \frac{\sqrt{n}}{2}\int_y \left\{ \frac{1}{n}\sum_{i=1}^{n} \acute{\varphi}(y - r_t(x_i))\dot{r}_t(x_i).\frac{s_\theta}{\hat{s}_\theta} \right\} dy - \\
&\qquad \frac{\sqrt{n}}{2}\int_y \left\{ \frac{1}{n}\sum_{i=1}^{n} \acute{\varphi}(y - r_t(x_i))\dot{r}_t(x_i).\frac{\hat{s}_\theta}{\hat{s}_\theta} \right\} dy
\end{aligned}
$$

Finally we have,

$$
T_{n,2} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_y \left(\frac{s_\theta}{2\hat{s}_\theta}(y) - \frac{1}{2} \right) \acute{\varphi}(y - r_t(x_i))\dot{r}_t(x_i)dy. \tag{4.40}
$$

Let us define the term $W_n = S_n + T_n$ as in (4.14) and (4.15). Then using (4.19) and the expansion of $T_{n,2}$ term in (4.22), W_n can be expressed as,

$$
\begin{aligned}
W_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_y \frac{\dot{s}_\theta}{2s_\theta}(y)\left\{\varphi_{\sigma(n)}(y-Y_i)-f_{\sigma(n)}\right\}dy+ \\
&\quad \tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^{n}\int_y \left(\tfrac{s_\theta}{2\hat{s}_\theta}(y)-\tfrac{1}{2}\right)\acute{\varphi}(y-r_t(x_i))\dot{r}_t(x_i)dy \qquad (4.41)\\
&= \tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^{n} Z_{n,i}+\tfrac{1}{\sqrt{n}}\sum_{i=1}^{n} D_i \\
&= \tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^{n}\left(Z_{n,i}+D_i\right)=\tfrac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(U_{n,i}\right).
\end{aligned}
$$

Here, $Z_{n,i}$ is the same notation as (4.23). Now we have,

$$S_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{n,i} \quad \text{and} \quad T_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} D_i$$

Since $\mathbb{E}(U_{n,i})=0$;

$$
\begin{aligned}
\mathbb{V}ar(W_n) &= \mathbb{E}(U_{n,i}^2)=\mathbb{E}\left\{(Z_{n,i}+D_i)^2\right\} \\
&= \mathbb{E}(Z_{n,i}^2)+\mathbb{E}(D_i^2)+2\mathbb{E}(Z_{n,i}D_i)
\end{aligned}
\qquad (4.42)
$$

Let us introduce the generic term,

$$U_n = \int_y \frac{\hat{s}_\theta}{2s_\theta}(y)\left\{\varphi_{\sigma(n)}(y-Y)-f_{\sigma(n)}\right\}dy+\int_y\left(\frac{s_\theta}{2\hat{s}_\theta}(y)-\frac{1}{2}\right)\acute{\varphi}(y-r_t(x))\dot{r}_t(x)dy. \qquad (4.43)$$

U_n is a bounded random variable by the Cauchy-Schwarz inequality, thus for arbitrary $\delta > 0$;

$$\mathbb{E}\left|U_n\right|^{2+\delta} \le 2_Y\left(\frac{\dot{s}_\theta}{2s_\theta}(y)+\frac{s_\theta}{2\hat{s}_\theta}(y)\right)^{2+\delta} < \infty \;\; \forall\sigma=\sigma(n).$$

Now we introduce the following lemma.

Lemma 4.2 Let $U_{n,1}, U_{n,2}, \dots U_{n,n}$ be independent random variable with the same distribution as U_n in(4.43). It follows from the Lyapanov's Central Limit Theorem, the asymptotic normality of W_n would be obtained if the variance of W_n in (4.41) satisfies

$$\begin{aligned} \mathbb{V}ar(W_n) &= \mathbb{V}ar(S_n + T_n) = \mathbb{V}ar(S_n) + \mathbb{V}ar(T_n) + 2Cov(S_n, T_n) \\ &\to \upsilon_{1,\theta}^2 + \upsilon_{2,\theta}^2 + 2\mathbb{E}(Z_n T_n) = \upsilon_\theta^2, \end{aligned} \tag{4.44}$$

and if there exists $\delta > 0$, such that

$$\lim_{n\to\infty} \frac{1}{\mathbb{V}ar(W_n)^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}\left|\frac{1}{\sqrt{n}} U_{n,i}\right|^{2+\delta} = 0$$

which leads to

$$\lim_{n\to\infty} \frac{1}{(\upsilon_\theta)^{1+\frac{\delta}{2}}} \frac{1}{n^{1+\frac{\delta}{2}}} n\mathbb{E}\left|U_n\right|^{2+\delta} = 0 \tag{4.45}$$

Proof. Since (4.45) is true if (4.44) is true, it suffices to prove the limit of $\mathbb{V}ar(W_n)$. We know from Theorem 4.1 and Theorem 4.2, that $\mathbb{V}ar(S_n) \to \upsilon_{1,\theta}^2$ and $\mathbb{V}ar(T_n) \to \upsilon_{2,\theta}^2$ respectively. We then look at the limit of the covariance term.

Since $\mathbb{E}(S_n) = \mathbb{E}(T_n) = 0$;

$$\begin{aligned} Cov(S_n, T_n) &= \mathbb{E}(S_n, T_n) \\ &= \mathbb{E}\left(\tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n Z_{n,i}.\tfrac{1}{\sqrt{n}} \sum_{i=1}^n D_i\right) \\ &= \mathbb{E}(Z_n D_n). \end{aligned}$$

For the T_n term, let us write it in an alternative way as follows;

$$T_n = \frac{1}{2}\left\langle \sqrt{n}(\hat{s}_\theta - s_\theta), \frac{\dot{\hat{f}}_\theta}{\hat{s}_\theta} \right\rangle = \frac{1}{2}\left\langle \sqrt{n}(\hat{f}_\theta - f_\theta), \frac{\dot{\hat{f}}_\theta}{\hat{s}_\theta(\hat{s}_\theta + s_\theta)} \right\rangle.$$

According to (4.36), we get

$$T_n \to \frac{1}{2}\sqrt{n}\int_y (\hat{f}_\theta - f_\theta).\frac{\dot{s}_\theta}{s_\theta}(y)dy.$$

Using the kernel chosen in (2.6),

$$\begin{aligned} T_n &\to \frac{\sqrt{n}}{2}\frac{1}{n}\sum_{i=1}^{n}\int_y \frac{\dot{s}}{s_\theta}\left\{(\varphi_{\sigma(n)}(y-Y_i)) - f_{\sigma(n)}(y)\right\}dy \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_y \frac{\dot{s}_\theta}{2s_\theta}\left\{\varphi_{\sigma(n)}(y-Y_i) - f_{\sigma(n)}(y)\right\}dy. \end{aligned}$$

Therefore;

$$\begin{aligned} \mathbb{E}Z_nT_n &\to \mathbb{E}\int_y \frac{\dot{s}_\theta}{2s_\theta}(y)\left\{\varphi_{\sigma(n)}(y-Y) - f_{\sigma(n)}(y)\right\}dy\times \\ &\qquad \int_y \frac{\dot{s}_\theta}{2s_\theta}(z)\left\{\varphi_{\sigma(n)}(z-Z) - f_{\sigma(n)}(z)\right\}dz \\ &= \int_y\int_w\int_z K_\theta(w)K_\theta(z)\left\{\varphi_{\sigma(n}(w-y) - f_{\sigma(n)}(w)\right\} \\ &\qquad \left\{\varphi_{\sigma(n)}(z-y) - f_{\sigma(n)}(z)\right\}dwdzf(y)dy, \end{aligned}$$

where $K_\theta = \dfrac{\dot{s}_\theta}{2s_\theta}$ as usual notation.
Then,

$$\begin{aligned} \mathbb{E}Z_nT_n &\to \textstyle\int_y\int_w\int_z K_\theta(w)K_\theta(z)\left\{\varphi_{\sigma(n)}(w-y)\varphi_{\sigma(n)}(z-y)\right\}dwdzf(y)dy \\ &\quad + \textstyle\int_y\int_w\int_z K_\theta(w)K_\theta(z)f_{\sigma(n)}(w)f_{\sigma(n)}(z)dwdzf(y)dy \\ &= \textstyle\int_y K_\theta^2(y)f(y)dy + \left\{\int_y K_\theta(y)f(y)dy\right\}^2. \end{aligned}$$

Since $\mathbb{E}(K_\theta(y) = 0$; we wnd up with the first term only.

Then

$$\mathbb{E}Z_nT_n \to \mathbb{E}\left(\frac{\dot{s}_\theta}{2s_\theta}\right)^2 = \upsilon_{\theta,1}^2.$$

Finally,

$$\mathbb{V}(W_n) \to \upsilon_\theta^2 = \upsilon_{\theta,1}^2 + \upsilon_{\theta,2}^2 + 2\upsilon_{\theta,1}^2$$

and the lemma is proven. □

Theorem 4.3 Under stated conditions,

$$\sqrt{n}(\hat{\theta} - \theta) \to^d N(0, \frac{\upsilon_\theta^2}{\langle s_\theta, \ddot{s}_\theta \rangle^2})$$

where $\upsilon_\theta^2 = 3\upsilon_{1,\theta}^2 + \upsilon_{2,\theta}^2$.

Proof.

$$\sqrt{n}(\hat{\theta} - \theta) = -\frac{1}{\left\langle s_\theta, \hat{\ddot{s}}_\theta \right\rangle} \{S_n + T_{n,1} + T_{n,2} + r_n\}$$

by (4.14),(4,16) and (4.28).
From (4.17) and (4.33), $T_{n,1} \to 0$ and $r_n \to 0$.
According to Theorem 4.1,

$$S_n \xrightarrow{d} N\left(0, \upsilon_{1,\theta}^2\right).$$

According to Theorem 4.2,

$$T_n \xrightarrow{d} N\left(0, \upsilon_{2,\theta}^2\right).$$

Since $(\hat{\theta} - \theta)$ can be written as a linear combination of $S_n$and T_n, using Cramer-Wold device,$\upsilon_\theta^2 = \mathbb{V}ar(S_n + T_n)$.
Now using $W_n = S_n + T_n$ with Lemma 4.2, we obtain the desired result. □

CHAPTER 5
SIMULATION RESULTS

In this chapter, we compare the MHD estimator with Maximum Likelihood Estimator (MLE) and Least Squared (LS) method using Empirical Mean Squared Error ($M\hat{S}E$) and empirical bias ($B\hat{i}as$) which are defined as,

$$B\hat{i}as = \frac{1}{m}\sum_{i=1}^{m}(\hat{\theta}_i - \theta), \tag{5.1}$$

and

$$M\hat{S}E = \frac{1}{m}\sum_{i=1}^{m}(\hat{\theta}_i - \theta)^2. \tag{5.2}$$

where $\hat{\theta}$ denotes either the MHD estimator, Maximum Likelihood Estimator or estimator from LS method. We used $m = 100$ replications and $\hat{\theta}_i$ is the estimate of θ for the i^{th} replication.
We look at the following 3 models:

$$\text{Model I} : y = c + \theta x + e$$
$$\text{Model II} : y = \theta + cx + e$$
$$\text{Model III} : y = c + \theta x^2 + e$$

Three sample sizes are used which are $n = 10, 50, 100, 1000$ and 5000 with three different θ values. The Gaussian Kernel is used as the entirely non parametric estimator in the Hellinger diatance function. The results are displayed in Table 5.1 through Table 5.9. The normality of the MHDE is compared using qqplots and the results are displayed in Figure 5.1 through Figure 5.6.

Model I : $y = c + \theta x + e$

	θ = 1, n=10		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	1.194397	0.951866	0.951866
$\hat{Bias}$	0.194397	-0.048134	0.048134
$\hat{MSE}$	0.065069	0.007480	0.007480
	θ = 1, n=50		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	1.0659784	0.9594233	1.040577
$\hat{Bias}$	0.0659784	-0.0405767	0.0405767
$\hat{MSE}$	0.0095721	0.0023714	0.0023714
	θ = 1, n=100		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.9617486	0.9623656	1.038251
$\hat{Bias}$	-0.03825139	-0.0376344	0.03825139
$\hat{MSE}$	0.001863123	0.001994037	0.001863123
	θ = 1, n=1000		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.9616664	0.961412	1.038588
$\hat{Bias}$	-0.03833357	-0.03858799	0.03858799
$\hat{MSE}$	0.001514265	0.001515249	0.001515249
	θ = 1, n=5000		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.9614102	0.9612669	1.038733
$\hat{Bias}$	-0.03858981	-0.03873309	0.03873309
$\hat{MSE}$	0.001502169	0.001506831	0.00150683

Table 5.1: Estimates of Bias and Mean Squared Errors of $\hat{\theta}_{MHD}$,$\hat{\theta}_{LS}$ and $\hat{\theta}_{MLE}$ for Model I,$\theta = 1$

Model I : $y = c + \theta x + e$

	$\theta = 0.5$, **n=10**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.688898	0.485818	0.507092
$\hat{Bias}$	0.188898	-0.014182	0.007092
$\hat{MSE}$	0.005957	0.001841	0.000460
	$\theta = 0.5$, **n=50**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.521310	0.481526	0.509237
$\hat{Bias}$	0.021310	-0.018474	0.009237
$\hat{MSE}$	0.000868	0.000502	0.000126
	$\theta = 0.5$, **n=100**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.508419	0.480865	0.509567
$\hat{Bias}$	0.008419	-0.019135	0.009567
$\hat{MSE}$	0.000072	0.000458	0.000114
	$\theta = 0.5$, **n=1000**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.509980	0.480824	0.518798
$\hat{Bias}$	0.009979	-0.019176	0.018798
$\hat{MSE}$	0.000103	0.000377	0.000361
	$\theta = 0.5$, **n=5000**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.480724	0.480206	0.519344
$\hat{Bias}$	-0.019276	-0.019794	0.019344
$\hat{MSE}$	0.000374	0.000407	0.000376

Table 5.2: Estimates of Bias and Mean Squared Errors of $\hat{\theta}_{MHD}$,$\hat{\theta}_{LS}$ and $\hat{\theta}_{MLE}$ for Model I,$\theta = 0.5$

Model I : $y = c + \theta x + e$

	$\theta = 5$, **n=10**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	4.493976	4.782606	6.086948
$\hat{Bias}$	-0.5060241	-0.2173938	1.086948
$\hat{MSE}$	0.4904423	0.1608879	4.022151
	$\theta = 5$, **n=50**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	4.635304	4.790397	6.048016
$\hat{Bias}$	-0.3646958	-0.2096032	1.048016
$\hat{MSE}$	0.3853441	0.06258919	1.564729
	$\theta = 5$, **n=100**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	4.804771	4.807301	5.963493
$\hat{Bias}$	-0.1952287	-0.1926987	0.9634932
$\hat{MSE}$	0.05265673	0.04732245	1.183061
	$\theta = 5$, **n=1000**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	4.806803	4.803803	5.980983
$\hat{Bias}$	-0.1939093	-0.1961966	0.9809832
$\hat{MSE}$	0.03901075	0.03943447	0.9858618
	$\theta = 5$, **n=5000**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	4.807567	4.807496	5.962164
$\hat{Bias}$	-0.1924329	-0.1925043	0.9621644
$\hat{MSE}$	0.03719587	0.03728366	0.9298968

Table 5.3: Estimates of Bias and Mean Squared Errors of $\hat{\theta}_{MHD}$,$\hat{\theta}_{LS}$ and $\hat{\theta}_{MLE}$ for Model I,$\theta = 5$

Comparing QQ-Plots of MHDE and MLE for Model I.

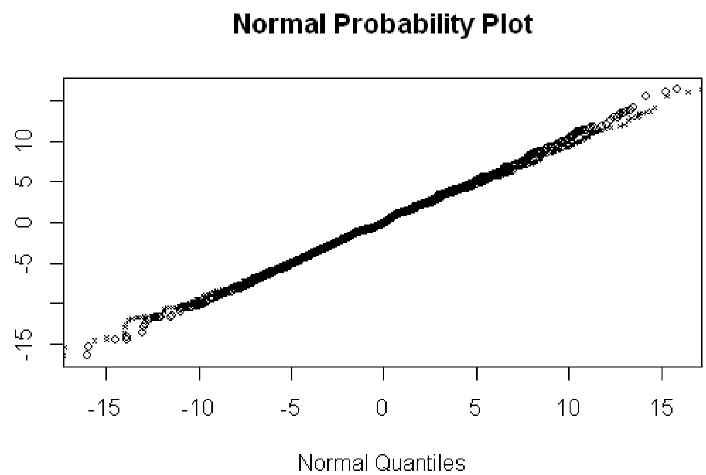

Figure 5.1: Normal Probability Plot for Model I with MHD estimator (o) and MLE (x) for n=1000

Comparing QQ-Plots of MHDE and LS Method for Model I.

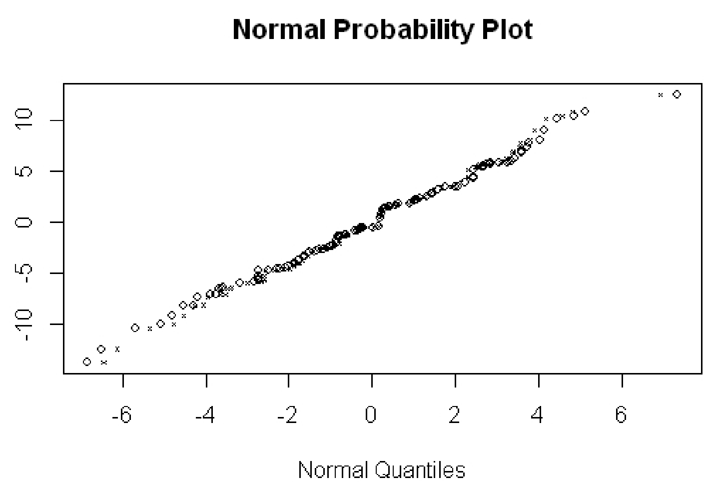

Figure 5.2: Normal Probability Plot for Model I with MHD estimator (o) and LS (x) for n=1000

Model II : $y = \theta + cx + e$

	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\theta = 1$, **n=10**			
$\hat{\theta}$	1.048696	0.935281	0.973549
$\hat{Bias}$	0.048696	-0.064719	-0.026451
$\hat{MSE}$	0.133266	0.151367	0.009155
$\theta = 1$, **n=50**			
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	1.036156	0.948926	1.001668
$\hat{Bias}$	0.036156	-0.051074	0.001668
$\hat{MSE}$	0.030783	0.021599	0.002071
$\theta = 1$, **n=100**			
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.957889	0.906562	1.038323
$\hat{Bias}$	-0.042111	-0.093438	0.038323
$\hat{MSE}$	0.010577	0.012516	0.001859
$\theta = 1$, **n=1000**			
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.962843	0.903656	1.038541
$\hat{Bias}$	-0.037157	-0.096344	0.038541
$\hat{MSE}$	0.001269	0.009442	0.001524
$\theta = 1$, **n=5000**			
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.962726	0.903853	1.038308
$\hat{Bias}$	-0.037274	-0.096147	0.038308
$\hat{MSE}$	0.001474	0.009280	0.001566

Table 5.4: Estimates of Bias and Mean Squared Errors of $\hat{\theta}_{MHD}$,$\hat{\theta}_{LS}$ and $\hat{\theta}_{MLE}$ for Model II,$\theta = 1$

Model II : $y = \theta + cx + e$

$\theta = 0.5$, **n=10**			
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.517347	0.478656	0.495695
$\hat{Bias}$	0.017347	-0.021344	-0.004305
$\hat{MSE}$	0.094393	0.121745	0.001463
$\theta = 0.5$, **n=50**			
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.519431	0.463630	0.491933
$\hat{Bias}$	0.019431	-0.036370	-0.008067
$\hat{MSE}$	0.018175	0.026742	0.000564
$\theta = 0.5$, **n=100**			
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.521400	0.476366	0.518720
$\hat{Bias}$	0.021400	-0.023634	0.018720
$\hat{MSE}$	0.000458	0.010144	0.000453
$\theta = 0.5$, **n=1000**			
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.510334	0.481421	0.519220
$\hat{Bias}$	0.010334	-0.018579	0.019220
$\hat{MSE}$	0.000107	0.000528	0.000379
$\theta = 0.5$, **n=5000**			
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.501241	0.481146	0.519086
$\hat{Bias}$	0.001241	-0.018854	0.019086
$\hat{MSE}$	0.000015	0.001158	0.000366

Table 5.5: Estimates of Bias and Mean Squared Errors of $\hat{\theta}_{MHD}$,$\hat{\theta}_{LS}$ and $\hat{\theta}_{MLE}$ for Model II,$\theta = 0.5$

Model II : $y = \theta + cx + e$

	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\theta = 5$, **n=10**			
$\hat{\theta}$	5.230262	4.814394	5.042623
$\hat{Bias}$	0.230262	-0.185606	0.042623
$\hat{MSE}$	0.187031	0.330489	0.146607
$\theta = 5$, **n=50**			
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	5.183569	4.811362	5.001612
$\hat{Bias}$	0.183569	-0.188638	0.001612
$\hat{MSE}$	0.054391	0.083803	0.028323
$\theta = 5$, **n=100**			
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	5.241935	4.811043	5.188964
$\hat{Bias}$	0.241936	-0.188957	0.188964
$\hat{MSE}$	0.158043	0.054769	0.045228
$\theta = 5$, **n=1000**			
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	5.033250	4.809593	5.196562
$\hat{Bias}$	0.033250	-0.190407	0.196562
$\hat{MSE}$	0.019088	0.037868	0.039446
$\theta = 5$, **n=5000**			
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	4.972282	4.807562	5.193970
$\hat{Bias}$	-0.027718	-0.192438	0.193970
$\hat{MSE}$	0.024744	0.037322	0.037831

Table 5.6: Estimates of Bias and Mean Squared Errors of $\hat{\theta}_{MHD}$,$\hat{\theta}_{LS}$ and $\hat{\theta}_{MLE}$ for Model II,$\theta = 5$

Comparing QQ-Plots of MHDE and MLE for Model II.

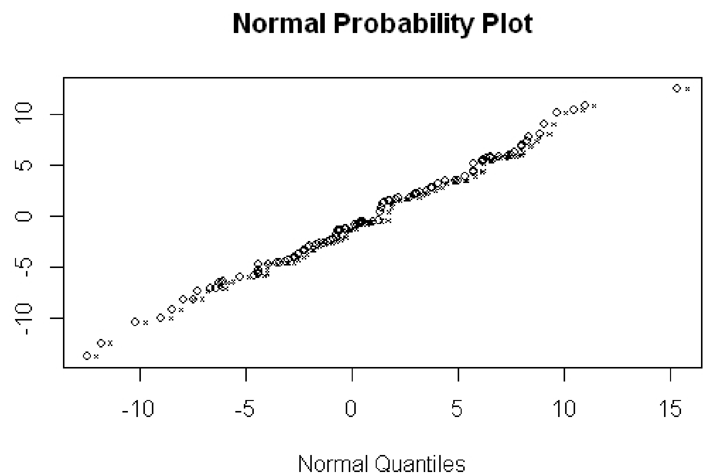

Figure 5.3: Normal Probability Plot for Model II with MHD estimator (o) and MLE (x) for n=1000

Comparing QQ-Plots of MHDE and LS Method for Model II.

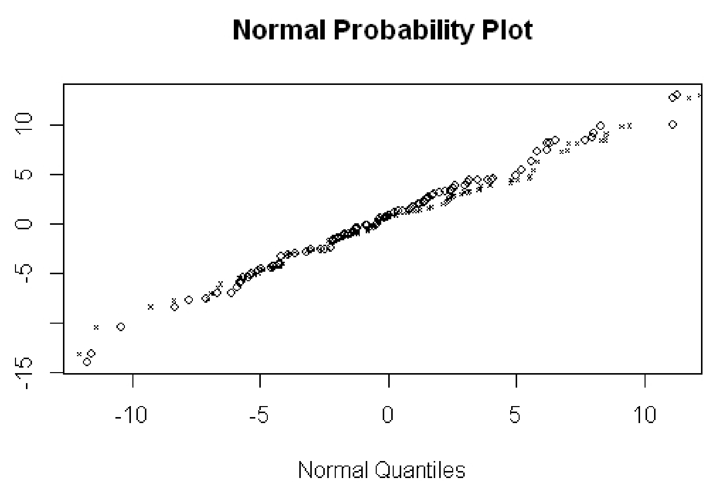

Figure 5.4: Normal Probability Plot for Model II with MHD estimator (o) and LS (x) for n=1000

Model III : $y = c + \theta x^2 + e$

$\theta = 1$, **n=10**			
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.9720602	0.9653553	1.0103330
$\hat{Bias}$	-0.0279398	-0.0346447	0.0103331
$\hat{MSE}$	0.0056610	0.0059347	0.0020234
$\theta = 0.5$, **n=50**			
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.9984462	0.9614209	0.9629046
$\hat{Bias}$	-0.0015538	-0.0385791	-0.0370954
$\hat{MSE}$	0.0021474	0.0033668	0.0024418
$\theta = 0.5$, **n=100**			
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.9948169	0.9596185	0.9623844
$\hat{Bias}$	-0.0051831	-0.0403815	-0.0376157
$\hat{MSE}$	0.0014401	0.0027022	0.0018038
$\theta = 0.5$, **n=1000**			
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.9999368	0.9561030	0.9620872
$\hat{Bias}$	-0.0000632	-0.0438970	-0.0379128
$\hat{MSE}$	0.0001130	0.0020534	0.0014765
$\theta = 0.5$, **n=5000**			
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.9998746	0.9563610	0.9616391
$\hat{Bias}$	-0.0001254	-0.0436390	-0.0383610
$\hat{MSE}$	0.0000229	0.0019299	0.0014795

Table 5.7: Estimates of Bias and Mean Squared Errors of $\hat{\theta}_{MHD}$,$\hat{\theta}_{LS}$ and $\hat{\theta}_{MLE}$ for Model III,$\theta = 1$

Model III : $y = c + \theta x^2 + e$

	$\theta = 0.5$, **n=10**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.4806798	0.4741093	0.4957982
$\hat{Bias}$	-0.0193202	-0.0258907	-0.0042018
$\hat{MSE}$	0.0009790	0.0017082	0.0008315
	$\theta = 1$, **n=50**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.4832484	0.4815595	0.4976804
$\hat{Bias}$	-0.0167516	-0.0184405	-0.0023196
$\hat{MSE}$	0.0005744	0.0005862	0.0001604
	$\theta = 1$, **n=100**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.5002561	0.4769127	0.4813674
$\hat{Bias}$	0.0002561	-0.0230873	-0.0186326
$\hat{MSE}$	0.0000690	0.0007132	0.0004564
	$\theta = 1$, **n=1000**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.4997243	0.4784331	0.4809031
$\hat{Bias}$	-0.0002757	-0.0215669	-0.0190969
$\hat{MSE}$	0.0000076	0.0004939	0.0003727
	$\theta = 1$, **n=5000**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	0.4999451	0.4778982	0.4807382
$\hat{Bias}$	0.0000549	-0.0221018	-0.0192618
$\hat{MSE}$	0.0000013	0.0004941	0.0003729

Table 5.8: Estimates of Bias and Mean Squared Errors of $\hat{\theta}_{MHD}$,$\hat{\theta}_{LS}$ and $\hat{\theta}_{MLE}$ for Model III,$\theta = 0.5$

Model III : $y = c + \theta x^2 + e$

	$\theta = 5$, **n=10**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	4.9444230	4.8508300	4.8860650
$\hat{Bias}$	-0.0555769	-0.1491704	-0.1139359
$\hat{MSE}$	0.0090505	0.1766268	0.0872213
	$\theta = 5$, **n=50**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	4.9786070	4.8079030	5.0626310
$\hat{Bias}$	-0.0213932	-0.1920973	0.0626308
$\hat{MSE}$	0.0093223	0.0600975	0.0102975
	$\theta = 5$, **n=100**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	5.0333350	4.8061620	4.9363430
$\hat{Bias}$	0.0333353	-0.1938378	-0.0636569
$\hat{MSE}$	0.0065847	0.0458472	0.0089111
	$\theta = 5$, **n=1000**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	5.0207740	4.8051110	4.9500440
$\hat{Bias}$	0.0207737	-0.1948890	-0.0499559
$\hat{MSE}$	0.0055847	0.0393487	0.0099012
	$\theta = 5$, **n=5000**		
	$\hat{\theta}_{MHD}$	$\hat{\theta}_{LS}$	$\hat{\theta}_{MLE}$
$\hat{\theta}$	4.9798710	4.8065480	4.9612331
$\hat{Bias}$	-0.0201290	-0.1934517	-0.0487669
$\hat{MSE}$	0.0041289	0.0375711	0.0099000

Table 5.9: Estimates of Bias and Mean Squared Errors of $\hat{\theta}_{MHD}$,$\hat{\theta}_{LS}$ and $\hat{\theta}_{MLE}$ for Model III,$\theta = 5$

Comparing QQ-Plots of MHDE and MLE for Model III.

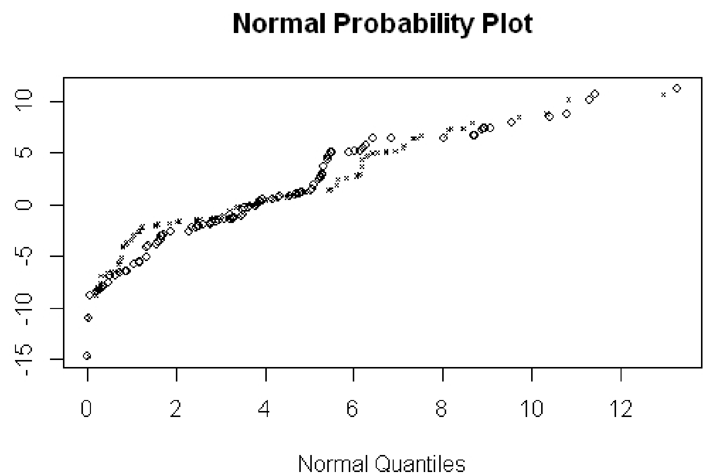

Figure 5.5: Normal Probability Plot for Model III with MHD estimator (o) and MLE (x) for n=1000

Comparing QQ-Plots of MHDE and LS Method for Model III.

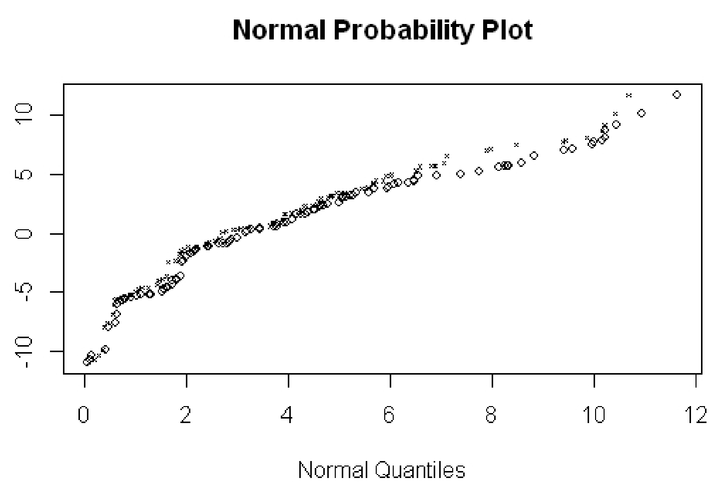

Figure 5.6: Normal Probability Plot for Model III with MHD estimator (o) and LS (x) for n=1000

CHAPTER 6
CONCLUSION AND FUTURE WORK

6.1 Conclusion

In this report, we introduce the Minimum Hellinger Distance estimator and review its history. We introduce a regression problem in a parametric family with random design and propose a minimum Hellinger distance estimator, which has been shown to have good efficiency and robustness properties. We prove the consistancy and the normality of the proposed estimate. Simulation study shows that the MHDE performs comparably to the other estimators. When sample size increases the MHDE performs better compared to other methods.

The work can be extended to parametric vectors and deterministic model as discussed in next sections.

6.2 Vector Parameters

We have discussed the random design model with a single parameter θ in this report. The study can be extended to a model where the parameter space is a vector. In this situation, the properties has to be derived for the following model.

$$Y = r_\theta(x) + \epsilon, \tag{6.1}$$

where

$$\theta = (\theta_1, \theta_2, \ldots, \theta_n).$$

6.3 Deterministic Model

In the case of deterministic design, the data are

$$Y_i = r_t(x_i) + \epsilon_i, \quad i = 1, \ldots, n \ , \ t \in \Theta. \tag{6.2}$$

where the $x_i \in \mathbb{X}$ are given points.

Although the points are known now, they are not independent and identically distributed. Therefore, deriving the consistency and normality of the above model needs further assumptions and more technical details. One suggestion is to normalize all the points and apply the methods used for random design model.

BIBLIOGRAPHY

[1] Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *Annalsof s tatistics*, 5, 445-463.

[2] Cutler, A. and CorderoBrana, O.I. (1998). Minimum Hellinger distance estimation for finite mixture models. *Journal of the American Statistical Association*, 91, 1716-1723.

[2] Huber, P.J (1972). 1972 Wald Lecture Robust Statistics: A review. *The Annals of Mathematics Statistics*, 43, 1041-1067.

[2] Hampbel, F. (1974). The influence and its role in robust estimation. *Jour of de Am. Statistics Assoc*, 69, 383-393.

[3] Karlis, D. and Xekalaki, E. (1998). Minimum Hellinger distance estimation for Poisson mixtures. *Computational Statistics and Data Analysis*, 29, 81-103.

[4] Karunamuni, R.J. and Wu, J. (2009). Minimum Hellinger distance estimation in a nonparametric mixture model. *Journal of Statistical Planning and Inference*, 139, 1118-1133.

[5] Lindsay, B. G. (1977). Efficiency versus robustness: The case for minimum Hellinger distance estimation and related Methods. *Annals of statistics*, 22, 1081-1114.

[6] Lu, Z., Hui, Y. V. and Lee, A. H. (2003). Minimum Hellinger distance estimation for finite mixtures of Poisson regression models and its applications. *Biometrics*, 59, 1016-1026.

[7] Pak, R. (1996). Minimum Hellinger distance estimation in simple linear regression models; distribution and efficiency. *Statistics probability letters*, 26, 263-269.

[8] Silverman, B.W., (1986). *Density Estimation: Theory, Practice and Visualization*. Wiley.

[9] Van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge.

[10] Van der Vaart, A.W., Wellner, Jon A. (1996). *Weak Convergence and Empirical Processes*. Springer.

[11] Xiang, L., Yau, K.K.W., Hui, Y.V. and Lee, A.H. (2008). Minimum Hellinger distance estimation for k-component poisson mixture with random effects. *Biometrics*, 64, 508-518.

[12] Woodward, W. A., Whitney, P. and Eslinger, P. W. (1995). Minimum Hellinger distance estimation of mixture proportions. *Journal of Statistical Planning and Inference*, 48, 303-319.

[13] Wu, J. and Karunamuni, R.J. (2009). On minimum Hellinger distance estimation. *The Canadian Journal of Statistics*, 37, 514-533.

[14] Wu, J., Schick, A. and Karunamuni, R.J. (2011). Profile Hellinger distance estimation. Technical Report.

[15] Wu, J., Karunamuni, R. and Zhang, B. (2010). Minimum Hellinger distance estimation in a two-sample semiparametric model. *Journal of Multivariate Analysis*, 101, 1102-1122.

APPENDIX A

Theorems and Definitions

Theorem 1**(Continuous Mapping Theorem)**
Let $\{X_n\}$, X be random elements defined on a metric space $\mathcal{S}$. Suppose a function $g : \mathcal{S} \to \mathcal{S}'$ has the set of discontinuity points D_g such that $Pr(X \in D_g) = 0$. Then $X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$.

Theorem 2 **(Uniform Law of large Numbers)**
Suppose $f(X, \theta)$ is a function defined for $\theta \in \Theta$, where Θ is compact and $f(X, \theta)$ is continuous at each $\theta \in \Theta$ for almost all X's and measuarable function of X at eachθ. If there exists a dominating fuction $d(x)$ such that $E[d(X)] < \infty$, and

$$\|f(x, \theta)\| \leq d(x) \quad \forall\theta \in \Theta$$

then $E[f(x, \theta)]$ is continuous in θ, and

$$sup_{\theta\in\Theta} \left\| \frac{1}{n} \sum_{i=1}^{n} f(X_i, \theta) - E[f(X, \theta)] \right\| \xrightarrow{a.s.} 0.$$

Theorem 3**(Lyapunov Central Limit Theorem)**
Suppose $X_1, X_2, ...$ is a sequence of independent random variables, each with finite expected value μ and variance σ^2. Define $s_n^2 = \sum_{i=1}^{n} \sigma_i^2$. If for some $\delta > 0$, the Lyapunovs condition

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} \mathrm{E}\left[|X_i - \mu_i|^{2+\delta} \right] = 0$$

is satisfied, then a sum of $(X_i - \mu_i)/s_n$ converges in distribution to a standard normal random variable, as n goes to infinity:

$$\frac{1}{s_n} \sum_{i=1}^{n} (X_i - \mu_i) \ \xrightarrow{d}\ \mathcal{N}(0,\ 1).$$

Theorem 4**(Cramer Wold Device)**
Let $\overline{X}_n = (X_{n1}, \ldots, X_{nk})$ and $\overline{X} = (X_1, \ldots, X_k)$ be random vectors of dimension k. Then $\overline{X}_n$ converges in distribution to $\overline{X}$ if and only if:

$$\sum_{i=1}^{k} t_i X_{ni} \underset{n \to \infty}{\overset{D}{\to}} \sum_{i=1}^{k} t_i X_i$$

for each $(t_1, \ldots, t_k) \in \mathbb{R}^k$, that is, if every fixed linear combination of the coordinates of $\overline{X}_n$ converges in distribution to the correspondent linear combination of coordinates of $\overline{X}$.

Definition 1 **(Probability Space)**
Let Ω be a nonempty set and let $\mathcal{F}$ be a σ -algebra on Ω. Then the pair $(\Omega, \mathcal{F})$ is called a *measuarable space*. If μ is a measure on $(\Omega, \mathcal{F})$, then the triplet $(\Omega, \mathcal{F}, \mu)$ is called a *measure space*. If in addition, μ is a probability measure, then $(\Omega, \mathcal{F}, \mu)$ is called a *probability space*.

APPENDIX B

R Code Used in Simulation

```
Model I:
HellingerDist<-function(theta,x,y)
{
n=length(x)
k=density(y,kernel="gaussian",n=n)#Generate data from Gaussian Kernel.
fhat=k$y
error=y-theta*x
fthat=(1/n)*(dnorm(error,0,1))
diff=sqrt(fhat)-sqrt(fthat)
ans=norm(as.matrix(diff))

return (ans)
}
#MLE Estimation.
regress.ll<-function(int,slope,x,e)
{
predicted.value<-int+slope*(x)+e
-sum(dnorm(x,mean=predicted.value,sd=1,log=T))
 }

MinHellDist<-function(thetaI,theta,meany,meane,sdy,sde,m,n)
{
thetahat=NULL
thetareg=NULL
thetamle=NULL
bias1=NULL
bias2=NULL
bias3=NULL
```

```
#Iterating m times.
for(i in 1:m)
{
y=rnorm(n,mean=meany,sd=sdy)
e=rnorm(n,mean=meane,sd=sde)
x=(y-e)/theta

param <- optim(thetaI,HellingerDist, method= "Brent", control=list(fnscale=-1),low
thetahat[i]=param$par[1]
bias1[i]=(thetahat[i]-theta)

#LS Method
x=(y-e)/thetaI
Regt=lm(y~x)

thetareg[i]=Regt$coef[2]
bias2[i]=(thetareg[i]-theta)

#MLE
MLEt=mle2(regress.ll,start=list(int=0,slope=theta),data=list(x=x,e=e))

thetamle[i]=coef(MLEt)[2]*theta
bias3[i]=(thetamle[i]-theta)
}

mseMHD=mean(bias1^2)
mseReg=mean(bias2^2)
mseMLE=mean(bias3^2)

ans=list(thetahat=mean(thetahat),thetareg=mean(thetareg),thetamle=mean(thetamle),b
return(ans)
}
```

```
Model II:

HellingerDist<-function(theta,x,y)
{
n=length(x)

k=density(y,kernel="gaussian",n=n)
fhat=k$y
error=y-theta-x
fthat=(1/n)*sum(dnorm(error,0,1))
diff=sqrt(fhat)-sqrt(fthat)
ans=norm(as.matrix(diff))

return (ans)
}

MinHellDist<-function(thetaI,theta,meany,meane,sdy,sde,m,n)
{
thetahat=NULL
thetareg=NULL
thetamle=NULL
bias1=NULL
bias2=NULL
bias3=NULL

#Iterating m times.
for(i in 1:m)
{
y=rnorm(n,mean=meany,sd=sdy)
e=rnorm(n,mean=meane,sd=sde)
x=(y-e-theta)
```

```
param <- optim(thetaI,HellingerDist, method= "Brent", control=list(fnscale=-1),low
thetahat[i]=param$par[1]
bias1[i]=(thetahat[i]-theta)

#LS Method
x=(y-e-thetaI)
Regt=lm(y~x)

thetareg[i]=Regt$coef[1]
bias2[i]=(thetareg[i]-theta)

#MLE
MLEt=mle2(regress.ll,start=list(int=theta,slope=1),data=list(x=x,e=e))

thetamle[i]=coef(MLEt)[1]*theta
bias3[i]=(thetamle[i]-theta)
}

mseMHD=mean(bias1^2)
mseReg=mean(bias2^2)
mseMLE=mean(bias3^2)

ans=list(thetahat=mean(thetahat),thetareg=mean(thetareg),thetamle=mean(thetamle),b
return(ans)
}
```

Model III:

```
HellingerDist<-function(theta,x,y)
{
n=length(x)

k=density(y,kernel="gaussian",n=n)
fhat=k$y
error=y-theta*(x^2)
fthat=(1/n)*sum(dnorm(error,0,1))
diff=sqrt(fhat)-sqrt(fthat)
ans=norm(as.matrix(diff))

return (ans)
}

MinHellDist<-function(thetaI,theta,meany,meane,sdy,sde,m,n)
{
thetahat=NULL
thetareg=NULL
thetamle=NULL
bias1=NULL
bias2=NULL
bias3=NULL

#Iterating m times.
for(i in 1:m)
{
y=rnorm(n,mean=meany,sd=sdy)
e=rnorm(n,mean=meane,sd=sde)
x=abs((y-e)/theta)^0.5

param <- optim(thetaI,HellingerDist, method= "Brent", control=list(fnscale=-1),low
```

```
thetahat[i]=param$par[1]
bias1[i]=(thetahat[i]-theta)

#LS Method
x=(y-e)/thetaI
Regt=lm(y~x^2)

thetareg[i]=Regt$coef[2]
bias2[i]=(thetareg[i]-theta)

#MLE
x=abs((y-e)/theta)^0.5
MLEt=mle2(regress.ll,start=list(int=0,slope=theta),data=list(x=x^2,e=e))

thetamle[i]=(coef(MLEt)[2])*theta
bias3[i]=(thetamle[i]-theta)
}

mseMHD=mean(bias1^2)
mseReg=mean(bias2^2)
mseMLE=mean(bias3^2)

ans=list(thetahat=mean(thetahat),thetareg=mean(thetareg),thetamle=mean(thetamle),b
return(ans)
}
```