entropy

MDPI

Article

Robust Inference after Random Projections via Hellinger Distance for Location-Scale Family

Lei Li [1], Anand N. Vidyashankar [1,*], Guoqing Diao [1] and Ejaz Ahmed [2]

1 Department of Statistics, George Mason University, Fairfax, VA 22030, USA; lli14@masonlive.gmu.edu (L.L.); gdiao@gmu.edu (G.D.)
2 Department of Mathematics and Statistics, Brock University, St. Catharines, ON L2S 3A1, Canada; sahmed5@brocku.ca
* Correspondence: avidyash@gmu.edu

Received: 6 February 2019; Accepted: 24 March 2019; Published: 29 March 2019

Abstract: Big data and streaming data are encountered in a variety of contemporary applications in business and industry. In such cases, it is common to use random projections to reduce the dimension of the data yielding compressed data. These data however possess various anomalies such as heterogeneity, outliers, and round-off errors which are hard to detect due to volume and processing challenges. This paper describes a new robust and efficient methodology, using Hellinger distance, to analyze the compressed data. Using large sample methods and numerical experiments, it is demonstrated that a routine use of robust estimation procedure is feasible. The role of double limits in understanding the efficiency and robustness is brought out, which is of independent interest.

Keywords: compressed data; Hellinger distance; representation formula; iterated limits; influence function; consistency; asymptotic normality; location-scale family

1. Introduction

Streaming data are commonly encountered in several business and industrial applications leading to the so-called Big Data. These are commonly characterized using four V's: velocity, volume, variety, and veracity. Velocity refers to the speed of data processing while volume refers to the amount of data. Variety refers to various types of data while veracity refers to uncertainty and imprecision in data. It is believed that veracity is due to data inconsistencies, incompleteness, and approximations. Whatever be the real cause, it is hard to identify and pre-process data for veracity in a big data setting. The issues are even more complicated when the data are streaming.

A consequence of the data veracity is that statistical assumptions used for analytics tend to be inaccurate. Specifically, considerations such as model misspecification, statistical efficiency, robustness, and uncertainty assessment-which are standard part of a statistical toolkit-cannot be routinely carried out due to storage limitations. Statistical methods that facilitate simultaneous addressal of twin problems of volume and veracity would enhance the value of the big data. While health care industry and financial industries would be the prime benefactors of this technology, the methods can be routinely applied in a variety of problems that use big data for decision making.

We consider a collection of n (n is of the order of at least 10^6) observations, assumed to be independent and identically distributed (i.i.d.), from a probability distribution $f(\cdot)$ belonging to a location-scale family; that is,

$$f(x;\mu,\sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right), \ \mu \in \mathbb{R}, \sigma > 0.$$

We denote by Θ the parameter space and without loss of generality take it as compact since otherwise it can be re-parametrized in a such a way that the resulting parameter space is compact (see [1]).

The purpose of this paper is to describe a methodology for joint robust and efficient estimation of μ and σ^2 that takes into account (i) storage issues, (ii) potential model misspecifications, and (iii) presence of aberrant outliers. These issues-which are more likely to occur when dealing with massive amounts of data-if not appropriately accounted in the methodological development, can lead to inaccurate inference and misleading conclusions. On the other hand, incorporating them in the existing methodology may not be feasible due to a computational burden.

Hellinger distance-based methods have long been used to handle the dual issue of robustness and statistical efficiency. Since the work of [1,2] statistical methods that invoke alternative objective functions which converge to the objective function under the posited model have been developed and the methods have been shown to possess efficiency and robustness. However, their routine use in the context of big data problems is not feasible due to the complexity in the computations and other statistical challenges. Recently, a class of algorithms-referred to as *Divide and Conquer*—have been developed to address some of these issues in the context of likelihood. These algorithms consist in distributing the data across multiple processors and, in the context of the problem under consideration, estimating the parameters from each processor separately and then combining them to obtain an overall estimate. The algorithm assumes availability of several processors, *with substantial processing power,* to solve the complex problem at hand. Since robust procedures involve complex iterative computations-invoking the increased demand for several high-speed processors and enhanced memory-routine use of available analytical methods in a big data setting is challenging. Maximum likelihood method of estimation in the context of location-scale family of distributions has received much attention in the literature ([3–7]). It is well-known that the maximum likelihood estimators (MLE) of location-scale families may not exist unless the defining function $f(\cdot)$ satisfies certain regularity conditions. Hence, it is natural to ask if other methods of estimation such as minimum Hellinger distance estimator(MHDE) under weaker regularity conditions. This manuscript provides a first step towards addressing this question. Random projections and sparse random projections are being increasingly used to "compress data" and then use the resulting compressed data for inference. The methodology, primarily developed by computer scientists, is increasingly gaining attention among the statistical community and is investigated in a variety of recent work ([8–12]). In this manuscript, we describe a Hellinger distance-based methodology for robust and efficient estimation after the use of random projections for compressing i.i.d data belonging to the location-scale family. The proposed method consists in reducing the dimension of the data to facilitate the ease of computations and simultaneously maintain robustness and efficiency when the posited model is correct. While primarily developed to handle big and streaming data, the approach can also be used to handle privacy issues in a variety of applications [13].

The rest of the paper is organized as follows: Section 2 provides background on minimum Hellinger distance estimation; Section 3 is concerned with the development of Hellinger distance-based methods for compressed data obtained after using random projections; additionally, it contains the main results and their proofs. Section 4 contains results of the numerical experiments and also describes an algorithm for implementation of the proposed methods. Section 5 contains a real data example from financial analytics. Section 6 is concerned with discussions and extensions. Section 7 contains some concluding remarks.

2. Background on Minimum Hellinger Distance Estimation

Ref. [1] proposed minimum Hellinger distance (MHD) estimation for i.i.d. observations and established that MHD estimators (MHDE) are simultaneously robust and first-order efficient under the true model. Other researchers have investigated related estimators, for example, [14–20]. These authors establish that when the model is correct, the MHDE is asymptotically equivalent to the

maximum likelihood estimator (MLE) in a variety of independent and dependent data settings. For a comprehensive discussion of minimum divergence theory see [21].

We begin by recalling that the Hellinger distance between two probability densities is the L^2 distance between the square root of the densities. Specifically, let, for $p \geq 1$, $||\cdot||_p$ denote the L^p norm defined by

$$||h||_p = \left\{ \int |h|^p \right\}^{1/p}.$$

The Hellinger distance between the densities $f(\cdot)$ and $g(\cdot)$ is given by

$$H^2(f(\cdot), g(\cdot)) = ||f^{1/2}(\cdot) - g^{1/2}(\cdot)||_2^2.$$

Let $f(\cdot|\boldsymbol{\theta})$ denote the density of $\mathbb{R}^d$ valued independent and identically distributed random variables $X_1, \cdots, X_n$, where $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$; let $g_n(\cdot)$ be a nonparametric density estimate (typically a kernel density estimator). The Hellinger distance between $f(\cdot|\boldsymbol{\theta})$ and $g_n(\cdot)$ is then

$$H^2\left(f(\cdot|\boldsymbol{\theta}), g_n(\cdot)\right) = ||f^{1/2}(\cdot|\boldsymbol{\theta}) - g_n^{1/2}(\cdot)||_2^2.$$

The MHDE is a mapping $T(\cdot)$ from the set of all densities to $\mathbb{R}^p$ defined as follows:

$$\boldsymbol{\theta}_g = T(g) = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmin}}\, H^2\left(f(\cdot|\boldsymbol{\theta}), g(\cdot)\right). \tag{1}$$

Please note that the above minimization problem is equivalent to maximizing $\mathcal{A}\left(f(\cdot|\boldsymbol{\theta}), g(\cdot)\right) = \int f^{1/2}(x|\boldsymbol{\theta}) g^{1/2}(x) dx$. Hence MHDE can alternatively be defined as

$$\boldsymbol{\theta}_g = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmax}}\, \mathcal{A}\left(f(\cdot|\boldsymbol{\theta}), g(\cdot)\right).$$

To study the robustness of MHDE, ref. [1] showed that to assess the robustness of a functional with respect to the gross-error model it is necessary to examine the α-influence curve rather than the influence curve, except when the influence curve provides a uniform approximation to the α-influence curve. Specifically, the α-influence function ($\mathrm{IF}_\alpha(\boldsymbol{\theta}, z)$) is defined as follows: for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, let $f_{\alpha,\boldsymbol{\theta},z} = (1-\alpha) f(\cdot|\boldsymbol{\theta}) + \alpha \eta_z$, where η_z denotes the uniform density on the interval $(z-\epsilon, z+\epsilon)$, where $\epsilon > 0$ is small, $\alpha \in (0,1)$, $z \in \mathbb{R}$; the α-influence function is then defined to be

$$\mathrm{IF}_\alpha(\boldsymbol{\theta}, z) = \frac{T(f_{\alpha,\boldsymbol{\theta},z}) - \boldsymbol{\theta}}{\alpha}, \tag{2}$$

where $T(f_{\alpha,\boldsymbol{\theta},z})$ is the functional for the model with density $f_{\alpha,\boldsymbol{\theta},z}(\cdot)$. Equation (2) represents a complete description of the behavior of the estimator in the presence of contamination, up to the shape of the contaminating density. If $\mathrm{IF}_\alpha(\boldsymbol{\theta}, z)$ is a bounded function of z such that $\lim_{z\to\infty} \mathrm{IF}_\alpha(\boldsymbol{\theta}, z) = 0$, for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, then the functional T is robust at $f(\cdot|\boldsymbol{\theta})$ against $100\%\alpha$ contamination by gross errors at arbitrary large value z. The influence function can be obtained by letting $\alpha \to 0$. Under standard regularity conditions, the minimum divergence estimators (MDE) are first order efficient and have the same influence function as the MLE under the model, which is often unbounded. Hence the robustness of these estimators cannot be explained through their influence functions. In contrast, the α-influence function of the estimators are often bounded, continuous functions of the contaminating point. Finally, this approach often leads to high breakdown points in parametric estimation. Other explanations can also be found in [22,23].

Ref. [1] showed that the MHDE of location has a breakdown point equal to 50%. Roughly speaking, the breakdown point is the smallest fraction of data that, when strategically placed, can cause an estimator to take arbitrary values. Ref. [24] obtained breakdown results for MHDE of multivariate

location and covariance. They showed that the affine-invariant MHDE for multivariate location and covariance has a breakdown point of at least 25%. Ref. [18] showed that the MHDE has 50% breakdown in some discrete models.

3. Hellinger Distance Methodology for Compressed Data

In this section we describe the Hellinger distance-based methodology as applied to the compressed data. Since we are seeking to model the streaming independent and identically distributed data, we denote by J the number of observations in a fixed time-interval (for instance, every ten minutes, or every half-hour, or every three hours). Let B denote the total number of time intervals. Alternatively, B could also represent the number of sources from which the data are collected. Then, the incoming data can be expressed as $\{X_{jl}, 1 \leq j \leq J; 1 \leq l \leq B\}$. Throughout this paper, we assume that the density of X_{jl} belongs to a location-scale family and is given by $f(x;\boldsymbol{\theta}^*) = \frac{1}{\sigma^*} f(\frac{x-\mu^*}{\sigma^*})$, where $\boldsymbol{\theta}^* = (\mu^*, \sigma^*)$. A typical example is a data store receiving data from multiple sources, for instance financial or healthcare organizations, where information from multiple sources across several hours are used to monitor events of interest such as cumulative usage of certain financial instruments or drugs.

3.1. Random Projections

Let $\boldsymbol{R}_l = (r_{ijl})$ be a $S \times J$ matrix, where S is the number of compressed observations in each time interval, $S \ll J$, and r_{ijl}'s are independent and identically distributed random variables and assumed to be independent of $\{X_{jl}, j = 1, 2, \cdots, J; 1 \leq l \leq B\}$. Let

$$\widetilde{Y}_{il} = \sum_{j=1}^{J} r_{ijl} X_{jl}$$

and set $\widetilde{\boldsymbol{Y}}_l = (\widetilde{Y}_{1l}, \cdots, \widetilde{Y}_{Sl})'$; in matrix form this can be expressed as $\widetilde{\boldsymbol{Y}}_l = \boldsymbol{R}_l \boldsymbol{X}_l$. The matrix $\boldsymbol{R}_l$ is referred to as the *sensing matrix* and $\{\widetilde{Y}_{il}, i = 1, 2 \cdots, S; l = 1, 2, \cdots, B\}$ is referred to as the *compressed data*. The total number of compressed observations $m = SB$ is much smaller than the number of original observations $n = JB$. We notice here that $\boldsymbol{R}_l$'s are independent and identically distributed random matrices of order $S \times J$. Referring to each time interval or a source as a group, the following Table 1 is a tabular representation of the compressed data.

Table 1. Illustration of Data Reduction Mechanism, Here $\boldsymbol{r}^*_{il} = (r_{i \cdot l}, \omega_{il})$.

	Grp 1	**Grp 2**	$\cdots$	**Grp B**		**Grp 1**	**Grp 2**	$\cdots$	**Grp B**
Original	X_{11}	X_{12}	$\cdots$	X_{1B}	Compressed	$(\widetilde{Y}_{11}, \boldsymbol{r}^*_{11})$	$(\widetilde{Y}_{12}, \boldsymbol{r}^*_{12})$	$\cdots$	$(\widetilde{Y}_{1B}, \boldsymbol{r}^*_{1B})$
Data	X_{21}	X_{22}	$\cdots$	X_{2B}	Data	$(\widetilde{Y}_{21}, \boldsymbol{r}^*_{21})$	$(\widetilde{Y}_{22}, \boldsymbol{r}^*_{22})$	$\cdots$	$(\widetilde{Y}_{2B}, \boldsymbol{r}^*_{2B})$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\overset{S \ll J}{\Longrightarrow}$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	X_{J1}	X_{J2}	$\cdots$	X_{JB}		$(\widetilde{Y}_{S1}, \boldsymbol{r}^*_{S1})$	$(\widetilde{Y}_{S2}, \boldsymbol{r}^*_{S2})$	$\cdots$	$(\widetilde{Y}_{SB}, \boldsymbol{r}^*_{SB})$

In random projections literature, the distribution of r_{ijl} is typically taken to be Gaussian; but other distributions such as Rademacher distribution, exponential distribution and extreme value distributions are also used (for instance, see [25]). In this paper, we do not make any strong distributional assumptions on r_{ijl}. We only assume that $\boldsymbol{E}\left[r_{ijl}\right] = 1$ and $\boldsymbol{Var}\left[r_{ijl}\right] = \gamma_0^2$, where $\boldsymbol{E}[\cdot]$ represents the expectation of the random variable and $\boldsymbol{Var}[\cdot]$ represents the variance of the random variable. Additionally, we denote the density of r_{ijl} by $q(\cdot)$.

We next return to the storage issue. When $S = 1$ and $r_{ijl} = 1$, $\widetilde{Y}_{il}$ is a sum of J random variables. In this case, one retains (stores) only the sum of J observations and robust estimates of $\boldsymbol{\theta}^*$ are sought using the sum of observations. In other situations, that is when r_{ijl} are not degenerate

at 1, the distribution of $\widetilde{Y}_{il}$ is complicated. Indeed, even if r_{ijl} are assumed to be normally distributed, the marginal distribution of $\widetilde{Y}_{il}$ is complicated. The conditional distribution is $\widetilde{Y}_{il}$ (given r_{ijl}) is a weighted sum of location scale distributions and does not have a useful closed form expression. Hence, in general for these problems the MLE method is *not feasible*. We denote by $\omega_{il}^2 = \sum_{j=1}^{J} r_{ijl}^2$ and work with the random variables $Y_{il} \equiv \omega_{il}^{-1}\widetilde{Y}_{il}$. We denote the true density of Y_{il} to be $h_J(\cdot|\boldsymbol{\theta}^*, \gamma_0)$. Also, when $\gamma_0 = 0$ (which implies $r_{ijl} \equiv 1$) we denote the true density of Y_{il} by $h^{*J}(\cdot|\boldsymbol{\theta}^*)$ to emphasize that the true density is a convolution of J independent and identically distributed random variables.

3.2. Hellinger Distance Method for Compressed Data

In this section, we describe the Hellinger distance-based method for estimating the parameters of the location scale family using the compressed data. As described in the last section, let $\{X_{jl}, j = 1, 2, \cdots, J; l = 1, 2, \cdots, B\}$ be a doubly indexed collection of independent and identically distributed random variables with true density $\frac{1}{\sigma^*} f\left(\frac{\cdot - \mu^*}{\sigma^*}\right)$. Our goal is to estimate $\boldsymbol{\theta}^* = (\mu^*, \sigma^{2*})$ using the compressed data $\{Y_{il}, i = 1, 2, \cdots, S; l = 1, 2, \cdots, B\}$. We re-emphasize here that the density of Y_{il} depends additionally on γ_0, the variance of the sensing random variables r_{ijl}.

To formulate the Hellinger-distance estimation method, let $\mathcal{G}$ be a class of densities metrized by the $\boldsymbol{L}_1$ distance. Let $\{h_J(\cdot|\boldsymbol{\theta}, \gamma_0); \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ be a parametric family of densities. The Hellinger distance functional T is a measurable mapping mapping from $\mathcal{G}$ to $\boldsymbol{\Theta}$, defined as follows:

$$\begin{aligned} T(g) &\equiv \arg\min_{\boldsymbol{\theta}} \int_{\mathbb{R}} \left(g^{\frac{1}{2}}(y) - h_J^{\frac{1}{2}}(y|\boldsymbol{\theta}, \gamma_0) \right)^2 dy \\ &= \arg\min_{\boldsymbol{\theta}} HD^2\left(g, h_J(\cdot|\boldsymbol{\theta}, \gamma_0)\right) = \boldsymbol{\theta}_g^*(\gamma_0). \end{aligned}$$

When $g(\cdot) = h_J(\cdot|\boldsymbol{\theta}^*, \gamma_0)$, then under additional assumptions $\boldsymbol{\theta}_g^*(\gamma_0) = \boldsymbol{\theta}^*(\gamma_0)$. Since minimizing the Hellinger-distance is equivalent to maximizing the affinity, it follows that

$$T(g) = \arg\max_{\boldsymbol{\theta}} \mathcal{A}\left(g, h_J(\cdot|\boldsymbol{\theta}, \gamma_0)\right), \text{ where}$$

$$\mathcal{A}(g, h_J(\cdot|\boldsymbol{\theta}, \gamma_0)) \equiv \int_{\mathbb{R}} g^{\frac{1}{2}}(y) h_J^{\frac{1}{2}}(y|\boldsymbol{\theta}, \gamma_0) dy.$$

It is worth noticing here that

$$\mathcal{A}(g, h_J(\cdot|\boldsymbol{\theta}, \gamma_0)) = 1 - \frac{1}{2} HD^2(g, h_J(\cdot|\boldsymbol{\theta}, \gamma_0)). \tag{3}$$

To obtain the Hellinger distance estimator of the true unknown parameters $\boldsymbol{\theta}^*$, expectedly we choose the parametric family $h_J(\cdot|\boldsymbol{\theta}, \gamma_0)$ to be density of Y_{il} and $g(\cdot)$ to be a non-parametric $\boldsymbol{L}_1$ consistent estimator $g_B(\cdot)$ of $h_J(\cdot|\boldsymbol{\theta}, \gamma_0)$. Thus, the MHDE of $\boldsymbol{\theta}_B^*$ is given by

$$\hat{\boldsymbol{\theta}}_B(\gamma_0) = \arg\max_{\boldsymbol{\theta}} \mathcal{A}\left(g_B, h_J(\cdot|\boldsymbol{\theta}, \gamma_0)\right) = T(g_B).$$

In the notation above, we emphasize the dependence of the estimator on the variance of the projecting random variables. We notice here that the solution to (1) may not be unique. In such cases, we choose one of the solutions in a measurable manner.

The choice of the density estimate, typically employed in the literature is the kernel density estimate. However, in the setting of the compressed data investigated here, there are S observations per group. These S observations are, conditioned on r_{ijl} independent; however they are marginally

dependent (if $S > 1$). In the case when $S > 1$, we propose the following formula for $g_B(\cdot)$. First, we consider the estimator

$$g_B^{(i)}(y) = \frac{1}{Bc_B} \sum_{l=1}^{B} K\left(\frac{y - Y_{il}}{c_B}\right), \; i = 1, 2, \cdots, S.$$

With this choice, the MHDE of $\boldsymbol{\theta}_B^*$ is given by, for $1 \leq i \leq S$,

$$\hat{\boldsymbol{\theta}}_{i,B}(\gamma_0) = \arg\max_{\boldsymbol{\theta}} \mathcal{A}\left(g_B^{(i)}, h_J(\cdot|\boldsymbol{\theta}, \gamma_0)\right). \tag{4}$$

The above estimate of the density chooses i^{th} observation from each group and obtains the kernel density estimator using the B independent and identically distributed compressed observations. This is one choice for the estimator. Of course, alternatively, one could obtain S^B different estimators by choosing different combinations of observations from each group.

It is well-known that the estimator is almost surely $\boldsymbol{L}_1$ consistent for $h_J(\cdot|\boldsymbol{\theta}^*, \gamma_0)$ as long as $c_B \to 0$ and $Bc_B \to \infty$ as $B \to \infty$. Hence, under additional regularity and identifiability conditions and further conditions on the bandwidth c_B, existence, uniqueness, consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}_{i,B}(\gamma_0)$, for fixed γ_0, follows from the existing results in the literature.

When $\gamma_0 = 0$ and $r_{ijl} \equiv 1$, as explained previously, the true density is a $J-$fold convolution of $f(\cdot|\boldsymbol{\theta}^*)$, it is natural to ask the following question: if one lets $\gamma_0 \to 0$, will the asymptotic results converge to what one would obtain by taking $\gamma_0 = 0$. We refer to this property as a *continuity property in* γ_0 of the procedure. Furthermore, it is natural to wonder if these asymptotic properties can be established uniformly in γ_0. If that is the case, then one can also allow γ_0 to depend on B. This idea has an intuitive appeal since one can choose the parameters of the sensing random variables to achieve an optimal inferential scheme. We address some of these issues in the next subsection.

Finally, we emphasize here that while we do not require $S > 1$, in applications involving streaming data and privacy problems S tends to greater than one. In problems where the variance of sensing variables are large, one can obtain an overall estimator by averaging $\hat{\boldsymbol{\theta}}_{i,B}(\gamma_0)$ over various choices of $1 \leq i \leq S$; that is,

$$\hat{\boldsymbol{\theta}}_B(\gamma_0) = \frac{1}{S} \sum_{i=1}^{S} \hat{\boldsymbol{\theta}}_{i,B}(\gamma_0). \tag{5}$$

The averaging improves the accuracy of the estimator in small compressed samples (data not presented). For this reason, we provide results for this general case, even though our simulation and theoretical results demonstrate that for some problems considered in this paper, S can be taken to be one. We now turn to our main results which are presented in the next subsection. The following Figure 1 provides a overview of our work.

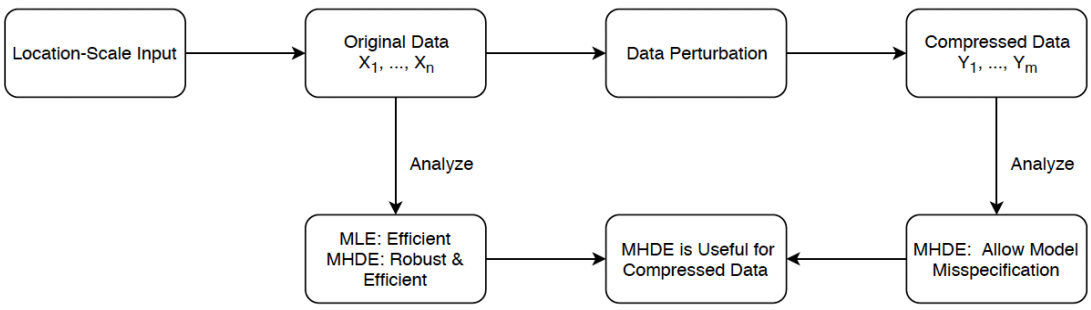


Figure 1. MLE vs. MHDE after Data Compression.

3.3. Main Results

In this section we state our main results concerning the asymptotic properties of the MHDE of compressed data Y_{il}. We emphasize here that we only store $\{(\widetilde{Y}_{il}, r_{i\cdot l}, \omega_{il}^2) : i = 1, 2, \cdots, S; l = 1, 2, \cdots, B\}$. Specifically, we establish the continuity property in γ_0 of the proposed methods by establishing the existence of the iterated limits. This provides a first step in establishing the double limit. The first proposition is well-known and is concerned with the existence and uniqueness of MHDE for the location-scale family defined in (4) using compressed data.

Proposition 1. *Assume that $h_J(\cdot|\boldsymbol{\theta}, \gamma_0)$ is a continuous density function. Assume further that if $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$. Then for every $\gamma_0 \geq 0$, $h_J(y|\boldsymbol{\theta}_1, \gamma_0) \neq h_J(y|\boldsymbol{\theta}_2, \gamma_0)$ on a set of positive Lebesgue measure, the MHDE in (4) exists and is unique.*

Proof. The proof follows from Theorem 2.2 of [20] since, without loss of generality, $\boldsymbol{\Theta}$ is taken to be compact and the density function $h_J(\cdot|\boldsymbol{\theta}, \gamma_0)$ is continuous in $\boldsymbol{\theta}$. □

Consistency: We next turn our attention to consistency. As explained previously, under regularity conditions for each fixed γ_0, the MHDE $\hat{\boldsymbol{\theta}}_{i,B}(\gamma_0)$ is consistent for $\boldsymbol{\theta}^*(\gamma_0)$. The next result says that under additional conditions, the consistency property of MHDE is continuous in γ_0.

Proposition 2. *Let $h_J(\cdot|\boldsymbol{\theta}, \gamma_0)$ be a continuous probability density function satisfying the conditions of Proposition 1. Assume that*

$$\lim_{\gamma_0 \to 0} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \int_{\mathbb{R}} |h_J(y|\boldsymbol{\theta}, \gamma_0) - h^{*J}(y|\boldsymbol{\theta})| dy = 0. \tag{6}$$

Then, with probability one (wp1) the iterated limits also exist and equals $\boldsymbol{\theta}^*$; that is, for $1 \leq i \leq S$,

$$\lim_{B \to \infty} \lim_{\gamma_0 \to 0} \hat{\boldsymbol{\theta}}_{i,B}(\gamma_0) = \lim_{\gamma_0 \to 0} \lim_{B \to \infty} \hat{\boldsymbol{\theta}}_{i,B}(\gamma_0) = \boldsymbol{\theta}^*.$$

Proof. Without loss of generality let $\boldsymbol{\Theta}$ be compact since otherwise it can be embedded into a compact set as described in [1]. Since $f(\cdot)$ is continuous in $\boldsymbol{\theta}$ and $g(\cdot)$ is continuous in γ_0, it follows that $h_J(\cdot|\boldsymbol{\theta}, \gamma_0)$ is continuous in $\boldsymbol{\theta}$ and γ_0. Hence by Theorem 1 of [1] for every fixed $\gamma_0 \geq 0$ and $1 \leq i \leq S$,

$$\lim_{B \to \infty} \hat{\boldsymbol{\theta}}_{i,B}(\gamma_0) = \boldsymbol{\theta}^*(\gamma_0).$$

Thus, to verify the convergence of $\boldsymbol{\theta}^*(\gamma_0)$ to $\boldsymbol{\theta}^*$ as $\gamma_0 \to 0$, we first establish, using (6), that

$$\lim_{\gamma_0 \to 0} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |\mathcal{A}(h_J(\cdot|\boldsymbol{\theta}, \gamma_0), h^{*J}(\cdot|\boldsymbol{\theta})) - 1| = 0.$$

To this end, we first notice that

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} HD^2(h_J(\cdot|\boldsymbol{\theta}, \gamma_0), h^{*J}(\cdot|\boldsymbol{\theta})) \leq \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \int_{\mathbb{R}} |(h_J(y|\boldsymbol{\theta}, \gamma_0) - h^{*J}(y|\boldsymbol{\theta})| dy.$$

Hence, using (3),

$$\begin{aligned} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |\mathcal{A}(h_J(\cdot|\boldsymbol{\theta}, \gamma_0), h^{*J}(\cdot|\boldsymbol{\theta})) - 1| &= \frac{1}{2} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} HD^2(h_J(\cdot|\boldsymbol{\theta}, \gamma_0), h^{*J}(\cdot|\boldsymbol{\theta})) \\ &\to 0 \text{ as } \gamma_0 \to 0. \end{aligned}$$

Hence,

$$\lim_{\gamma_0 \to 0} \mathcal{A}(h_J(\cdot|\boldsymbol{\theta}^*(\gamma_0), \gamma_0), h^{*J}(\cdot|\boldsymbol{\theta}^*(\gamma_0))) = 1.$$

Also, by continuity,

$$\lim_{\gamma_0 \to 0} \mathcal{A}(h^{*J}(\cdot|\boldsymbol{\theta}^*(\gamma_0), \gamma_0), h^{*J}(\cdot|\boldsymbol{\theta}^*)) = 1,$$

which, in turn implies that

$$\lim_{\gamma_0 \to 0} \mathcal{A}(h_J(\cdot|\boldsymbol{\theta}^*(\gamma_0), \gamma_0), h^{*J}(\cdot|\boldsymbol{\theta}^*)) = 1.$$

Thus existence of the iterated limit first as $B \to \infty$ and then $\gamma_0 \to 0$ follows using compactness of $\boldsymbol{\Theta}$ and the identifiability of the model. As for the other iterated limit, again notice notice that for each $1 \leq i \leq S$, $\mathcal{A}(g_B^{(i)}, h_J(\cdot|\boldsymbol{\theta}, \gamma_0))$ converges to $\mathcal{A}(g_B^{(i)}, h^{*J}(\cdot|\boldsymbol{\theta}))$ with probability one as γ_0 converges to 0. The result then follows again by an application of Theorem 1 of [20]. □

Remark 1. *Verification of condition (6) seems to be involved even in the case of standard Gaussian random variables and standard Gaussian sensing random variables. Indeed in this case, the density of $h_J(\cdot|\boldsymbol{\theta}, \gamma_0)$ is a $J-$fold convolution of a Bessel function of second kind. It may be possible to verify the condition (6) using the properties of these functions and compactness of the parameter space $\boldsymbol{\Theta}$. However, if one is focused only on weak-consistency, it is an immediate consequence of Theorems 1 and 2 below and condition (6) is not required. Finally, it is worth mentioning here that the convergence in (6) without uniformity over $\boldsymbol{\Theta}$ is a consequence of convergence in probability of r_{ijl} to 1 and Glick's Theorem.*

Asymptotic limit distribution: We now proceed to investigate the limit distribution of $\boldsymbol{\theta}_B^*(\gamma_0)$ as $B \to \infty$ and $\gamma_0 \to 0$. It is well-known that for fixed $\gamma_0 \geq 0$, after centering and scaling, $\boldsymbol{\theta}_B^*(\gamma_0)$ has a limiting Gaussian distribution, under appropriate regularity conditions (see for instance [20]). However to evaluate the iterated limits as $\gamma_0 \to 0$ and $B \to \infty$, additional refinements of the techniques in [20] are required. To this end, we start with additional notations. Let $s_J(\cdot|\boldsymbol{\theta}, \gamma_0) = h_J^{\frac{1}{2}}(\cdot|\boldsymbol{\theta}, \gamma_0)$ and let the score function be denoted by $u_J(\cdot|\boldsymbol{\theta}, \gamma_0) \equiv \nabla \log h_J(\cdot|\boldsymbol{\theta}, \gamma_0) = \left(\frac{\partial \log h_J(\cdot|\boldsymbol{\theta}, \gamma_0)}{\partial \mu}, \frac{\partial \log h_J(\cdot|\boldsymbol{\theta}, \gamma_0)}{\partial \sigma}\right)'$. Also, the Fisher information $I(\boldsymbol{\theta}(\gamma_0))$ is given by

$$I(\boldsymbol{\theta}(\gamma_0)) = \int_{\mathbb{R}} u_J(y|\boldsymbol{\theta}, \gamma_0) u_J'(y|\boldsymbol{\theta}, \gamma_0) h_J(y|\boldsymbol{\theta}, \gamma_0) dy.$$

In addition, let $\dot{s}_J(\cdot|\boldsymbol{\theta}, \gamma_0)$ be the gradient of $s_J(\cdot|\boldsymbol{\theta}, \gamma_0)$ with respect to $\boldsymbol{\theta}$, and $\ddot{s}_J(\cdot|\boldsymbol{\theta}, \gamma_0)$ is the second derivative matrix of $s_J(\cdot|\boldsymbol{\theta}, \gamma_0)$ with respect to $\boldsymbol{\theta}$. In addition, let $t_J(\cdot|\boldsymbol{\theta}) = h^{*J\frac{1}{2}}(\cdot|\boldsymbol{\theta})$ and $v_J(\cdot|\boldsymbol{\theta}) = \nabla \log h^{*J}(\cdot|\boldsymbol{\theta})$. Furthermore, let Y_{il}^* denote Y_{il} when $\gamma_0 \equiv 0$. Please note that in this case, $Y_{il} = Y_{1l}$ for all $i = 1, 2, \cdots, S$. The corresponding kernel density estimate of Y_{il}^* is given by

$$g_B^*(y) = \frac{1}{Bc_B} \sum_{l=1}^{B} K\left(\frac{y - Y_{il}^*}{c_B}\right). \tag{7}$$

We emphasize here that we suppress i on the LHS of the above equation since $g_B^{(i)*}(\cdot)$ are equal for all $1 \leq i \leq S$.

The iterated limit distribution involves additional regularity conditions which are stated in the Appendix. The first step towards this aim is a representation formula which expresses the quantity of interest, *viz.*, $\sqrt{B}\left(\hat{\boldsymbol{\theta}}_{i,B}(\gamma_0) - \boldsymbol{\theta}^*(\gamma_0)\right)$ as a sum of two terms, one involving sums of *compressed* i.i.d. random variables and the other involving remainder terms that converge to 0 at a specific rate.

This expression will appear in different guises in the rest of the manuscript and will play a critical role in the proofs.

3.4. Representation Formula

Before we state the lemma, we first provide two crucial assumptions that allow differentiating the objective function and interchanging the differentiation and integration:

Model assumptions on $h_J(\cdot|\boldsymbol{\theta},\gamma_0)$

(D1) $h_J(\cdot|\boldsymbol{\theta},\gamma_0)$ is twice continuously differentiable in $\boldsymbol{\theta}$.
(D2) Assume further that $||\nabla s_J(\cdot|\boldsymbol{\theta},\gamma_0)||_2$ is continuous and bounded.

Lemma 1. *Assume that the conditions (**D1**) and (**D2**) hold. Then for every $1 \leq i \leq S$ and $\gamma_0 \geq 0$, the following holds:*

$$B^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}}_{i,B}(\gamma_0) - \boldsymbol{\theta}^*(\gamma_0)\right)' = A_{1B}(\gamma_0) + A_{2B}(\gamma_0), \quad \textit{where} \tag{8}$$

$$A_{1B}(\gamma_0) = B^{\frac{1}{2}} D_B^{-1}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) T_B(\gamma_0), \quad A_{2B}(\gamma_0) = B^{\frac{1}{2}} D_B^{-1}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) R_B(\gamma_0), \tag{9}$$

$$\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0) \in U_B(\boldsymbol{\theta}'(\gamma_0)), \quad U_B(\boldsymbol{\theta}'(\gamma_0)) = \{\boldsymbol{\theta}' : \boldsymbol{\theta}'(\gamma_0) = t\boldsymbol{\theta}^*(\gamma_0) + (1-t)\hat{\boldsymbol{\theta}}_{i,B}(\gamma_0), t \in [0,1]\}, \tag{10}$$

$$\begin{aligned} D_B(\boldsymbol{\theta}(\gamma_0)) &= -\frac{1}{2}\int_{\mathbb{R}} \dot{u}_J(y|\boldsymbol{\theta},\gamma_0) s_J(y|\boldsymbol{\theta},\gamma_0) g_B^{(i)\frac{1}{2}}(y)dy \\ &\quad -\frac{1}{4}\int_{\mathbb{R}} u_J(y|\boldsymbol{\theta},\gamma_0) u_J'(y|\boldsymbol{\theta},\gamma_0) s_J(y|\boldsymbol{\theta},\gamma_0) g_B^{(i)\frac{1}{2}}(y)dy \\ &\equiv D_{1B}(\boldsymbol{\theta}(\gamma_0)) + D_{2B}(\boldsymbol{\theta}(\gamma_0)), \end{aligned} \tag{11}$$

$$T_B(\gamma_0) \equiv \frac{1}{4}\int_{\mathbb{R}} u_J(y|\boldsymbol{\theta}^*,\gamma_0)\left(h_J(y|\boldsymbol{\theta}^*,\gamma_0) - g_B^{(i)}(y)\right)dy, \quad \textit{and} \tag{12}$$

$$R_B(\gamma_0) \equiv \frac{1}{4}\int_{\mathbb{R}} u_J(y|\boldsymbol{\theta}^*,\gamma_0)\left(h_J^{\frac{1}{2}}(y|\boldsymbol{\theta}^*,\gamma_0) - g_B^{(i)\frac{1}{2}}(y)\right)^2 dy. \tag{13}$$

Proof. By algebra, note that $\dot{s}_J(y|\boldsymbol{\theta},\gamma_0) = \frac{1}{2}u_J(y|\boldsymbol{\theta},\gamma_0)s_J(y|\boldsymbol{\theta},\gamma_0)$. Furthermore, the second partial derivative of $s_J(\cdot|\boldsymbol{\theta},\gamma_0)$ is given by $\ddot{s}_J(y|\boldsymbol{\theta},\gamma_0) = \frac{1}{2}\dot{u}_J(y|\boldsymbol{\theta},\gamma_0)s_J(y|\boldsymbol{\theta},\gamma_0) + \frac{1}{4}u_J(y|\boldsymbol{\theta},\gamma_0)u_J'(y|\boldsymbol{\theta},\gamma_0)s_J(y|\boldsymbol{\theta},\gamma_0)$. Now using **(D1)** and **(D2)** and partially differentiating $HD_B^2(\boldsymbol{\theta}(\gamma_0)) \equiv HD^2(g_B^{(i)}(\cdot), h_J(\cdot|\boldsymbol{\theta},\gamma_0))$ with respect to $\boldsymbol{\theta}$ and setting it equal to 0, the estimating equations for $\boldsymbol{\theta}^*(\gamma_0)$ is

$$\nabla HD_B^2(\boldsymbol{\theta}^*(\gamma_0)) = 0. \tag{14}$$

Let $\hat{\boldsymbol{\theta}}_{i,B}(\gamma_0)$ be the solution to (14). Now applying first order Taylor expansion of (14) we get

$$\nabla HD_B^2(\boldsymbol{\theta}^*(\gamma_0)) = \nabla HD_B^2(\hat{\boldsymbol{\theta}}_{i,B}(\gamma_0)) + D_B(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0))\left(\hat{\boldsymbol{\theta}}_{i,B}(\gamma_0) - \boldsymbol{\theta}^*(\gamma_0)\right),$$

where $\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)$ is defined in (10), and $D_B(\cdot)$ is given by

$$\begin{aligned} D_B(\boldsymbol{\theta}(\gamma_0)) &= -\frac{1}{2}\int_{\mathbb{R}} \dot{u}_J(y|\boldsymbol{\theta},\gamma_0)s_J(y|\boldsymbol{\theta},\gamma_0)g_B^{(i)\frac{1}{2}}(y)dy \\ &\quad -\frac{1}{4}\int_{\mathbb{R}} u_J(y|\boldsymbol{\theta},\gamma_0)u_J'(y|\boldsymbol{\theta},\gamma_0)s_J(y|\boldsymbol{\theta},\gamma_0)g_B^{(i)\frac{1}{2}}(y)dy \\ &\equiv D_{1B}(\boldsymbol{\theta}(\gamma_0)) + D_{2B}(\boldsymbol{\theta}(\gamma_0)), \end{aligned}$$

and $\nabla HD_B^2(\cdot)$ is given by

$$\nabla HD_B^2(\boldsymbol{\theta}(\gamma_0)) = -\frac{1}{2}\int_{\mathbb{R}} u_J(y|\boldsymbol{\theta},\gamma_0)s_J(y|\boldsymbol{\theta},\gamma_0)\left(h_J^{\frac{1}{2}}(y|\boldsymbol{\theta}^*,\gamma_0) - g_B^{(i)\frac{1}{2}}(y)\right)dy.$$

Thus,

$$\left(\hat{\boldsymbol{\theta}}_{i,B}(\gamma_0) - \boldsymbol{\theta}^*(\gamma_0)\right)' = D_B^{-1}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0))\nabla HD_B^2(\boldsymbol{\theta}^*(\gamma_0)).$$

By using the identity, $b^{\frac{1}{2}} - a^{\frac{1}{2}} = (2a^{\frac{1}{2}})^{-1}\left((b-a) - (b^{\frac{1}{2}} - a^{\frac{1}{2}})^2\right)$, $\nabla HD_B^2(\boldsymbol{\theta}^*(\gamma_0))$ can be expressed as the difference of $T_B(\gamma_0)$ and $R_B(\gamma_0)$, where

$$T_B(\gamma_0) \equiv \frac{1}{4}\int_{\mathbb{R}} u_J(y|\boldsymbol{\theta}^*,\gamma_0)\left(h_J(y|\boldsymbol{\theta}^*,\gamma_0) - g_B^{(i)}(y)\right)dy,$$

and

$$R_B(\gamma_0) \equiv \frac{1}{4}\int_{\mathbb{R}} u_J(y|\boldsymbol{\theta}^*,\gamma_0)\left(h_J^{\frac{1}{2}}(y|\boldsymbol{\theta}^*,\gamma_0) - g_B^{(i)\frac{1}{2}}(y)\right)^2 dy.$$

Hence,

$$B^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}}_{i,B}(\gamma_0) - \boldsymbol{\theta}^*(\gamma_0)\right)' = A_{1B}(\gamma_0) + A_{2B}(\gamma_0),$$

where $A_{1B}(\gamma_0)$ and $A_{2B}(\gamma_0)$ are given in (9). □

Remark 2. *In the rest of the manuscript, we will refer to $A_{2B}(\gamma_0)$ as the remainder term in the representation formula.*

We now turn to the first main result of the manuscript, namely a central limit theorem for $\hat{\boldsymbol{\theta}}_{i,B}(\gamma_0)$ as first $B \to \infty$ and then $\gamma_0 \to 0$. As a first step, we note that the Fisher information of the density $h^{*J}(\cdot|\boldsymbol{\theta})$ is given by

$$I(\boldsymbol{\theta}) = \int_{\mathbb{R}} v_J(y|\boldsymbol{\theta})v_J'(y|\boldsymbol{\theta})h^{*J}(y|\boldsymbol{\theta})dy. \tag{15}$$

Next we state the assumptions needed in the proof of Theorem 1. We separate these conditions as (i) model assumptions, (ii) kernel assumptions, (iii) regularity conditions, (iV) conditions that allow comparison of original data and compressed data.

Model assumptions on $h^{*J}(\cdot|\boldsymbol{\theta})$

(D1′) $h^{*J}(\cdot|\boldsymbol{\theta})$ is twice continuously differentiable in $\boldsymbol{\theta}$.
(D2′) Assume further that $||\nabla t_J(\cdot|\boldsymbol{\theta})||_2$ is continuous and bounded.

Kernel assumptions

(B1) $K(\cdot)$ is symmetric about 0 on a compact support and bounded in $\boldsymbol{L}_2$. We denote the support of $K(\cdot)$ by $Supp(K)$.

(**B2**) The bandwidth c_B satisfies $c_B \to 0$, $B^{\frac{1}{2}} c_B \to \infty$, $B^{\frac{1}{2}} c_B^2 \to 0$.

Regularity conditions

(**M1**) The function $u_J(\cdot|\boldsymbol{\theta}, \gamma_0) s_J(\cdot|\boldsymbol{\theta}, \gamma_0)$ is continuously differentiable and bounded in $\boldsymbol{L}_2$ at $\boldsymbol{\theta}^*$.
(**M2**) The function $\dot{u}_J(\cdot|\boldsymbol{\theta}, \gamma_0) s_J(\cdot|\boldsymbol{\theta}, \gamma_0)$ is continuous and bounded in $\boldsymbol{L}_2$ at $\boldsymbol{\theta}^*$. In addition, assume that

$$\lim_{B\to\infty} \int_{\mathbb{R}} \left(\dot{u}_J(y|\boldsymbol{\theta}_{i,B}, \gamma_0) s_J(y|\boldsymbol{\theta}_{i,B}, \gamma_0) - \dot{u}_J(y|\boldsymbol{\theta}^*, \gamma_0) s_J(y|\boldsymbol{\theta}^*, \gamma_0)\right)^2 dy = 0.$$

(**M3**) The function $u_J(\cdot|\boldsymbol{\theta}, \gamma_0) u_J'(\cdot|\boldsymbol{\theta}, \gamma_0) s_J(\cdot|\boldsymbol{\theta}, \gamma_0)$ is continuous and bounded in $\boldsymbol{L}_2$ at $\boldsymbol{\theta}^*$; also,

$$\lim_{B\to\infty} \int_{\mathbb{R}} \left(u_J(y|\hat{\boldsymbol{\theta}}_{i,B}, \gamma_0) u_J'(y|\hat{\boldsymbol{\theta}}_{i,B}, \gamma_0) s_J(y|\boldsymbol{\theta}_{i,B}, \gamma_0) - u_J(y|\boldsymbol{\theta}^*, \gamma_0) u_J'(y|\boldsymbol{\theta}^*, \gamma_0) s_J(y|\boldsymbol{\theta}^*, \gamma_0)\right)^2 dy = 0.$$

(**M4**) Let $\{\alpha_B : B \geq 1\}$ be a sequence diverging to infinity. Assume that

$$\lim_{B\to\infty} B \sup_{t\in Supp(K)} \boldsymbol{P}_{\boldsymbol{\theta}^*(\gamma_0)} \left(|\Delta - c_B t| > \alpha_B\right) = 0,$$

where $Supp(\boldsymbol{K})$ is the support of the kernel density $K(\cdot)$ and Δ is a generic random variable with density $h_J(\cdot|\boldsymbol{\theta}^*, \gamma_0)$.
(**M5**) Let

$$M_B = \sup_{|y|\leq\alpha_B} \sup_{t\in Supp(K)} \left|\frac{h_J(y - tc_B|\boldsymbol{\theta}^*, \gamma_0)}{h_J(y|\boldsymbol{\theta}^*, \gamma_0)}\right|.$$

Assume $\sup_{B\geq 1} M_B < \infty$.
(**M6**) The score function has a regular central behavior relative to the smoothing constants, i.e.,

$$\lim_{B\to\infty} (B^{\frac{1}{2}} c_B)^{-1} \int_{-\alpha_B}^{\alpha_B} u_J(y|\boldsymbol{\theta}^*, \gamma_0) dy = 0.$$

Furthermore,

$$\lim_{B\to\infty} (B^{\frac{1}{2}} c_B^4) \int_{-\alpha_B}^{\alpha_B} u_J(y|\boldsymbol{\theta}^*, \gamma_0) dy = 0.$$

(**M7**) The density functions are smooth in an $\boldsymbol{L}_2$ sense; i.e.,

$$\lim_{B\to\infty} \sup_{t\in Supp(K)} \int_{\mathbb{R}} \left(u_J(y + c_B t|\boldsymbol{\theta}^*, \gamma_0) - u_J(y|\boldsymbol{\theta}^*, \gamma_0)\right)^2 h_J(y|\boldsymbol{\theta}^*, \gamma_0) dy = 0.$$

(**M1′**) The function $v_J(\cdot|\boldsymbol{\theta}) t_J(\cdot|\boldsymbol{\theta})$ is continuously differentiable and bounded in $\boldsymbol{L}_2$ at $\boldsymbol{\theta}^*$.
(**M2′**) The function $\dot{v}_J(\cdot|\boldsymbol{\theta}) t_J(\cdot|\boldsymbol{\theta})$ is continuous and bounded in $\boldsymbol{L}_2$ at $\boldsymbol{\theta}^*$. In addition, assume that

$$\lim_{B\to\infty} \int_{\mathbb{R}} \left(\dot{v}_J(y|\boldsymbol{\theta}_B) t_J(y|\boldsymbol{\theta}_B) - \dot{v}_J(y|\boldsymbol{\theta}^*) t_J(y|\boldsymbol{\theta}^*)\right)^2 dy = 0.$$

(**M3′**) The function $v_J(\cdot|\boldsymbol{\theta}) v_J'(\cdot|\boldsymbol{\theta}) t_J(\cdot|\boldsymbol{\theta})$ is continuous and bounded in $\boldsymbol{L}_2$ at $\boldsymbol{\theta}^*$. also,

$$\lim_{B\to\infty} \int_{\mathbb{R}} \left(v_J(y|\hat{\boldsymbol{\theta}}_{i,B}) v_J'(y|\hat{\boldsymbol{\theta}}_{i,B}) t_J(y|\hat{\boldsymbol{\theta}}_{i,B}) - v_J(y|\boldsymbol{\theta}^*) v_J'(y|\boldsymbol{\theta}^*) t_J(y|\boldsymbol{\theta}^*)\right)^2 dy = 0.$$

Assumptions comparing models for original and compressed data

(O1) For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\lim_{\gamma_0 \to 0} \int_{\mathbb{R}} \left(u_J(y|\boldsymbol{\theta}, \gamma_0) u_J'(y|\boldsymbol{\theta}, \gamma_0) s_J(y|\boldsymbol{\theta}, \gamma_0) - v_J(y|\boldsymbol{\theta}) v_J'(y|\boldsymbol{\theta}) t_J(y|\boldsymbol{\theta}) \right)^2 dy = 0.$$

(O2) For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\lim_{\gamma_0 \to 0} \int_{\mathbb{R}} \left(\dot{u}_J(y|\boldsymbol{\theta}, \gamma_0) s_J(y|\boldsymbol{\theta}, \gamma_0) - \dot{v}_J(y|\boldsymbol{\theta}) t_J(y|\boldsymbol{\theta}) \right)^2 dy = 0.$$

Theorem 1. *Assume that the conditions (**B1**)–(**B2**), (**D1**)–(**D2**) , (**D1′**)–(**D2′**), (**M1**)–(**M7**), (**M1′**)–(**M3′**), and (**O1**)–(**O2**) hold. Then, for every $1 \leq i \leq S$, the following holds:*

$$\lim_{\gamma_0 \to 0} \lim_{B \to \infty} \boldsymbol{P} \left(\sqrt{B} \left(\hat{\boldsymbol{\theta}}_{i,B}(\gamma_0) - \boldsymbol{\theta}^*(\gamma_0) \right) \leq x \right) = \boldsymbol{P} \left(G \leq x \right),$$

where G is a bivariate Gaussian random variable with mean 0 and variance $I^{-1}(\boldsymbol{\theta}^*)$, where $I(\boldsymbol{\theta})$ is defined in (15).

Before we embark on the proof of Theorem 1, we first discuss the assumptions. Assumptions (**B1**) and (**B2**) are standard assumptions on the kernel and the bandwidth and are typically employed when investigating the asymptotic behavior of divergence-based estimators (see for instance [1]). Assumptions (**M1**)–(**M7**) and (**M1′**)–(**M3′**) are regularity conditions which are concerned essentially with $\boldsymbol{L}_2$ continuity and boundedness of the scores and their derivatives. Assumptions (**O1**)–(**O2**) allow for comparison of $u_J(\cdot|\boldsymbol{\theta}, \gamma_0)$ and $v_J(\cdot|\boldsymbol{\theta})$. Returning to the proof of Theorem 1, using representation formula, we will first show that $\lim_{\gamma_0 \to 0} \lim_{B \to \infty} \boldsymbol{P} \left(A_{1B}(\gamma_0) \leq x \right) = \boldsymbol{P} \left(G \leq x \right)$, and then prove that $\lim_{\gamma_0 \to 0} \lim_{B \to \infty} A_{2B}(\gamma_0) = 0$ in probability. We start with the following proposition.

Proposition 3. *Assume that the conditions (**B1**), (**D1**)–(**D2**), (**M1**)–(**M3**), (**M1′**)–(**M3′**), (**M7**) and (**O1**)–(**O2**) hold. Then,*

$$\lim_{\gamma_0 \to 0} \lim_{B \to \infty} \boldsymbol{P} \left(A_{1B}(\gamma_0) \leq x \right) = \boldsymbol{P} \left(G \leq x \right),$$

where G is given in Theorem 1.

We divide the proof of Proposition 3 into two lemmas. In the first lemma we will show that

$$\lim_{\gamma_0 \to 0} \lim_{B \to \infty} D_B(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) = \frac{1}{4} I(\boldsymbol{\theta}^*).$$

Next in the second lemma we will show that first letting $B \to \infty$ and then allowing $\gamma_0 \to 0$,

$$4B^{\frac{1}{2}} T_B(\gamma_0) \xrightarrow{d} N\left(0, I(\boldsymbol{\theta}^*)\right).$$

We start with the first part.

Lemma 2. *Assume that the conditions (**D1**)–(**D2**), (**D1′**)–(**D2′**), (**M1**)–(**M3**), (**M1′**)–(**M3′**) and (**O1**)–(**O2**) hold. Then, with probability one, the following prevails:*

$$\lim_{\gamma_0 \to 0} \lim_{B \to \infty} D_B(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) = \frac{1}{4} I(\boldsymbol{\theta}^*).$$

Proof. Using representation formula in Lemma 1. First fix $\gamma_0 > 0$. It suffices to show

$$\lim_{B\to\infty} D_{1B}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) = \frac{1}{2}I(\boldsymbol{\theta}^*(\gamma_0)), \quad \text{and} \quad \lim_{B\to\infty} D_{2B}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) = -\frac{1}{4}I(\boldsymbol{\theta}^*(\gamma_0)).$$

We begin with $D_{1B}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0))$. By algebra, $D_{1B}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0))$ can be expressed as

$$D_{1B}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) = D_{1B}^{(1)}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) + D_{1B}^{(2)}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) + D_{1B}^{(3)}(\boldsymbol{\theta}^*(\gamma_0)), \quad \text{where}$$

$$D_{1B}^{(1)}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) = -\frac{1}{2}\int_{\mathbb{R}} \dot{u}_J(y|\tilde{\boldsymbol{\theta}}_{i,B},\gamma_0)s_J(y|\tilde{\boldsymbol{\theta}}_{i,B},\gamma_0)\left(g_B^{(i)\frac{1}{2}}(y) - s_J(y|\boldsymbol{\theta}^*,\gamma_0)\right)dy,$$

$$D_{1B}^{(2)}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) = -\frac{1}{2}\int_{\mathbb{R}} \left(\dot{u}_J(y|\tilde{\boldsymbol{\theta}}_{i,B},\gamma_0)s_J(y|\tilde{\boldsymbol{\theta}}_{i,B},\gamma_0) - \dot{u}_J(y|\boldsymbol{\theta}^*,\gamma_0)s_J(y|\boldsymbol{\theta}^*,\gamma_0)\right)h_J^{\frac{1}{2}}(y|\boldsymbol{\theta}^*,\gamma_0)dy,$$

$$\text{and} \quad D_{1B}^{(3)}(\boldsymbol{\theta}_B^*(\gamma_0)) = -\frac{1}{2}\int_{\mathbb{R}} \dot{u}_J(y|\boldsymbol{\theta}^*,\gamma_0)h_J(y|\boldsymbol{\theta}^*,\gamma_0)dy = \frac{1}{2}I(\boldsymbol{\theta}^*(\gamma_0)).$$

It suffices to show that as $B \to \infty$, $D_{1B}^{(1)}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) \to 0$, and $D_{1B}^{(2)}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) \to 0$. We first consider $D_{1B}^{(1)}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0))$. By Cauchy-Schwarz inequality and assumption (**M2**), it follows that there exists $0 < C_1 < \infty$,

$$\begin{aligned}\left|D_{1B}^{(1)}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0))\right| &\leq \frac{1}{2}\left\{\int_{\mathbb{R}} \left(\dot{u}_J(y|\tilde{\boldsymbol{\theta}}_{i,B},\gamma_0)s_J(y|\tilde{\boldsymbol{\theta}}_{i,B},\gamma_0)\right)^2 dy\right\}^{\frac{1}{2}} \left\{\int_{\mathbb{R}} \left(g_B^{(i)\frac{1}{2}}(y) - s_J(y|\boldsymbol{\theta}^*,\gamma_0)\right)^2 dy\right\}^{\frac{1}{2}} \\ &\leq C_1\left\{\int_{\mathbb{R}} \left(g_B^{(i)\frac{1}{2}}(y) - s_J(y|\boldsymbol{\theta}^*,\gamma_0)\right)^2 dy\right\}^{\frac{1}{2}} \to 0,\end{aligned}$$

where the last convergence follows from the $\boldsymbol{L}_1$ convergence of $g_B^{(i)}(\cdot)$ and $h_J(\cdot|\boldsymbol{\theta}^*,\gamma_0)$. Hence, as $B \to \infty$, $D_{1B}^{(1)}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) \to 0$. Next we consider $D_{1B}^{(2)}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0))$. Again, by Cauchy-Schwarz inequality and assumption (**M2**), it follows that $D_{1B}^{(2)}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) \to 0$. Hence $D_{1B}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) \to \frac{1}{2}I(\boldsymbol{\theta}^*(\gamma_0))$. Turning to $D_{2B}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0))$, by similar argument, using Cauchy-Schwarz inequality and assumption (**M3**), it follows that $D_{2B}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) \to -\frac{1}{4}I(\boldsymbol{\theta}^*(\gamma_0))$. Thus, to complete the proof, it is enough to show that

$$\lim_{\gamma_0\to 0}\lim_{B\to\infty} D_{1B}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) = \frac{1}{2}I(\boldsymbol{\theta}^*) \quad \text{and} \quad \lim_{\gamma_0\to 0}\lim_{B\to\infty} D_{2B}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) = -\frac{1}{4}I(\boldsymbol{\theta}^*). \tag{16}$$

We start with the first term of (16). Let

$$\mathcal{J}_1(\gamma_0) = \int_{\mathbb{R}} \dot{u}_J(y|\boldsymbol{\theta}^*,\gamma_0)h_J(y|\boldsymbol{\theta}^*,\gamma_0)dy - \int_{\mathbb{R}} \dot{v}_J(y|\boldsymbol{\theta}^*)h^{*J}(y|\boldsymbol{\theta}^*)dy.$$

We will show that $\lim_{\gamma_0\to 0}\mathcal{J}_1(\gamma_0) = 0$. By algebra, the difference of the above two terms can be expressed as the sum of $\mathcal{J}_{11}(\gamma_0)$ and $\mathcal{J}_{12}(\gamma_0)$, where

$$\mathcal{J}_{11}(\gamma_0) = \int_{\mathbb{R}} \left(\dot{u}_J(y|\boldsymbol{\theta}^*,\gamma_0)s_J(y|\boldsymbol{\theta}^*,\gamma_0) - \dot{v}_J(y|\boldsymbol{\theta}^*)t_J(y|\boldsymbol{\theta}^*)\right)s_J(y|\boldsymbol{\theta}^*,\gamma_0)dy, \quad \text{and}$$

$$\mathcal{J}_{12}(\gamma_0) = \int_{\mathbb{R}} \dot{v}_J(y|\boldsymbol{\theta}^*)t_J(y|\boldsymbol{\theta}^*)\left(s_J(y|\boldsymbol{\theta}^*,\gamma_0) - t_J(y|\boldsymbol{\theta}^*)\right)dy.$$

$\mathcal{J}_{11}(\gamma_0)$ converges to zero by Cauchy-Schwarz inequality and assumption (**O2**), and $\mathcal{J}_{12}(\gamma_0)$ converges to zero by Cauchy-Schwarz inequality, assumption (**M2′**) and Scheffe's theorem. Next we consider the second term of (16). Let

$$\mathcal{J}_2(\gamma_0) = \int_{\mathbb{R}} u_J(y|\boldsymbol{\theta}^*, \gamma_0) u_J'(y|\boldsymbol{\theta}^*, \gamma_0) h_J(y|\boldsymbol{\theta}^*, \gamma_0) dy - \int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*) v_J'(y|\boldsymbol{\theta}^*) h^{*J}(y|\boldsymbol{\theta}^*) dy.$$

We will show that $\lim_{\gamma_0 \to 0} \mathcal{J}_2(\gamma_0) = 0$. By algebra, the difference of the above two terms can be expressed as the sum of $\mathcal{J}_{21}(\gamma_0)$ and $\mathcal{J}_{22}(\gamma_0)$, where

$$\mathcal{J}_{21}(\gamma_0) = \int_{\mathbb{R}} \left(u_J(y|\boldsymbol{\theta}^*, \gamma_0) u_J'(y|\boldsymbol{\theta}^*, \gamma_0) s_J(y|\boldsymbol{\theta}^*, \gamma_0) - v_J(y|\boldsymbol{\theta}^*) v_J'(y|\boldsymbol{\theta}^*) t_J(y|\boldsymbol{\theta}^*) \right) s_J(y|\boldsymbol{\theta}^*, \gamma_0) dy,$$

$$\text{and} \quad \mathcal{J}_{22}(\gamma_0) = \int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*) v_J'(y|\boldsymbol{\theta}^*) t_J(y|\boldsymbol{\theta}^*) \left(s_J(y|\boldsymbol{\theta}^*, \gamma_0) - t_J(y|\boldsymbol{\theta}^*) \right) dy.$$

$\mathcal{J}_{11}(\gamma_0)$ converges to zero by Cauchy-Schwarz inequality and assumption (**O1**), and $\mathcal{J}_{12}(\gamma_0)$ converges to zero by Cauchy-Schwarz inequality, assumption (**M3′**) and Scheffe's theorem. Therefore the lemma holds. □

Lemma 3. *Assume that the conditions (**B1**), (**D1**)–(**D2**), (**D1′**)–(**D2′**), (**M1**)–(**M3**), (**M3′**), (**M7**) and (**O1**)–(**O2**) hold. Then, first letting $B \to \infty$, and then $\gamma_0 \to 0$,*

$$4B^{\frac{1}{2}} T_B(\gamma_0) \xrightarrow{d} N\left(0, I(\boldsymbol{\theta}^*)\right).$$

Proof. First fix $\gamma_0 > 0$. Please note that using $\int_{\mathbb{R}} u_J(y|\boldsymbol{\theta}^*, \gamma_0) h_J(y|\boldsymbol{\theta}^*, \gamma_0) dy = 0$, we have that

$$\begin{aligned} 4B^{\frac{1}{2}} T_B(\gamma_0) &= B^{\frac{1}{2}} \int_{\mathbb{R}} u_J(y|\boldsymbol{\theta}^*, \gamma_0) g_B^{(i)}(y) dy \\ &= B^{\frac{1}{2}} \int_{\mathbb{R}} u_J(y|\boldsymbol{\theta}^*, \gamma_0) \frac{1}{B} \sum_{l=1}^{B} \frac{1}{c_B} K\left(\frac{y - Y_{il}}{c_B}\right) dy \\ &= B^{\frac{1}{2}} \frac{1}{B} \sum_{l=1}^{B} \int_{\mathbb{R}} u_J(Y_{il} + c_B t|\boldsymbol{\theta}^*, \gamma_0) K(t) dt. \end{aligned}$$

Therefore,

$$4B^{\frac{1}{2}} T_B(\gamma_0) - B^{\frac{1}{2}} \frac{1}{B} \sum_{l=1}^{B} u_J(Y_{il}|\boldsymbol{\theta}^*, \gamma_0) = B^{\frac{1}{2}} \frac{1}{B} \sum_{l=1}^{B} \int_{\mathbb{R}} \left(u_J(Y_{il} + c_B t|\boldsymbol{\theta}^*, \gamma_0) - u_J(Y_{il}|\boldsymbol{\theta}^*, \gamma_0) \right) K(t) dt.$$

Since Y_{il}'s are i.i.d. across l, using Cauchy-Schwarz inequality and assumption (**B1**), we can show that there exists $0 < C < \infty$,

$$\begin{aligned} \boldsymbol{E}\left[4B^{\frac{1}{2}} T_B - B^{\frac{1}{2}} \frac{1}{B} \sum_{l=1}^{B} u_J(Y_{il}|\boldsymbol{\theta}^*, \gamma_0) \right]^2 &= \boldsymbol{E}\left[\int_{\mathbb{R}} \left(u_J(Y_{i1} + c_B t|\boldsymbol{\theta}^*, \gamma_0) - u_J(Y_{i1}|\boldsymbol{\theta}^*, \gamma_0) \right) K(t) dt \right]^2 \\ &\leq C\boldsymbol{E}\left[\left\{ \int_{\mathbb{R}} \left(u_J(Y_{i1} + c_B t|\boldsymbol{\theta}^*, \gamma_0) - u_J(Y_{i1}|\boldsymbol{\theta}^*, \gamma_0) \right)^2 dt \right\}^{\frac{1}{2}} \right]^2 \\ &\leq C\boldsymbol{E}\left[\int_{\mathbb{R}} \left(u_J(Y_{i1} + c_B t|\boldsymbol{\theta}^*, \gamma_0) - u_J(Y_{i1}|\boldsymbol{\theta}^*, \gamma_0) \right)^2 dt \right] \\ &= C \int_{\mathbb{R}} \int_{\mathbb{R}} \left(u_J(y + c_B t|\boldsymbol{\theta}^*, \gamma_0) - u_J(y|\boldsymbol{\theta}^*, \gamma_0) \right)^2 h_J(y|\boldsymbol{\theta}^*, \gamma_0) dy dt, \end{aligned}$$

converging to zero as $B \to \infty$ by assumption (**M7**). Also, the limiting distribution of $4B^{\frac{1}{2}}T_B(\gamma_0)$ is $N(0, I(\boldsymbol{\theta}^*(\gamma_0)))$ as $B \to \infty$. Now let $\gamma_0 \to 0$. It is enough to show that as $\gamma_0 \to 0$ the density of $N(0, I(\boldsymbol{\theta}^*(\gamma_0)))$ converges to the density of $N(0, I(\boldsymbol{\theta}^*))$. To this end, it suffices to show that $\lim_{\gamma_0 \to 0} I(\boldsymbol{\theta}^*(\gamma_0)) = I(\boldsymbol{\theta}^*)$. However, this is established in Lemma 2. Combining the results, the lemma follows. □

Proof of Proposition 3. The proof of Proposition 3 follows immediately by combining Lemmas 2 and 3. □

We now turn to establishing that the remainder term in the representation formula converges to zero.

Lemma 4. *Assume that the assumptions (**B1**)–(**B2**), (**M1**)–(**M6**) hold. Then*

$$\lim_{\gamma_0 \to 0} \lim_{B \to \infty} A_{2B}(\gamma_0) = 0 \quad \textit{in probability.}$$

Proof. Using Lemma 2, it is sufficient to show that $B^{\frac{1}{2}}R_B$ converges to 0 in probability as $B \to \infty$. Let

$$d_J(y|\boldsymbol{\theta}^*(\gamma_0)) = g_B^{(i)\frac{1}{2}}(y) - s_J(y|\boldsymbol{\theta}^*, \gamma_0).$$

Please note that

$$d_J^2(y|\boldsymbol{\theta}^*(\gamma_0)) \leq 2\left\{\left(h_J(y|\boldsymbol{\theta}^*, \gamma_0) - \boldsymbol{E}\left[g_B^{(i)}(y)\right]\right)^2 + \left(\boldsymbol{E}\left[g_B^{(i)}(y)\right] - g_B^{(i)}(y)\right)^2\right\} h_J^{-1}(y|\boldsymbol{\theta}^*, \gamma_0).$$

Then

$$\begin{aligned} |R_B(\gamma_0)| &\leq \frac{1}{2}\int_{\mathbb{R}} |u_J(y|\boldsymbol{\theta}^*, \gamma_0)| d_J^2(y|\boldsymbol{\theta}^*(\gamma_0))dy \\ &\leq \frac{1}{2}\int_{-\alpha_B}^{\alpha_B} |u_J(y|\boldsymbol{\theta}^*, \gamma_0)| d_J^2(y|\boldsymbol{\theta}^*(\gamma_0))dy + \frac{1}{2}\int_{|y| \geq \alpha_B} |u_J(y|\boldsymbol{\theta}^*, \gamma_0)| d_J^2(y|\boldsymbol{\theta}^*(\gamma_0))dy \\ &\equiv R_{1B}(\gamma_0) + R_{2B}(\gamma_0). \end{aligned}$$

We first deal with $R_{1B}(\gamma_0)$, which can be expressed as the sum of $R_{1B}(\gamma_0)$ and $R_{2B}(\gamma_0)$, where

$$R_{1B}^{(1)}(\gamma_0) = \int_{-\alpha_B}^{\alpha_B} |u_J(y|\boldsymbol{\theta}^*, \gamma_0)| \left(h_J(y|\boldsymbol{\theta}^*, \gamma_0) - \boldsymbol{E}\left[g_B^{(i)}(y)\right]\right)^2 h_J^{-1}(y|\boldsymbol{\theta}^*, \gamma_0)dy, \tag{17}$$

$$\text{and} \quad R_{1B}^{(2)}(\gamma_0) = \int_{-\alpha_B}^{\alpha_B} |u(y|\boldsymbol{\theta}^*, \gamma_0)| \left(\boldsymbol{E}\left[g_B^{(i)}(y)\right] - g_B^{(i)}(y)\right)^2 h_J^{-1}(y|\boldsymbol{\theta}^*, \gamma_0)dy.$$

Now consider $R_{1B}^{(2)}$. Let $\epsilon > 0$ be arbitrary but fixed. Then, by Markov's inequality,

$$\begin{aligned} \boldsymbol{P}\left(B^{\frac{1}{2}}R_{1B}^{(2)} > \epsilon\right) &\leq \epsilon^{-1}B^{\frac{1}{2}}\boldsymbol{E}\left[R_{1B}^{(2)}\right] \\ &\leq \epsilon^{-1}B^{\frac{1}{2}}\int_{\alpha_B}^{\alpha_B} |u_J(y|\boldsymbol{\theta}^*, \gamma_0)| \left(\boldsymbol{Var}\left[g_B^{(i)}(y)\right]\right) h_J^{-1}(y|\boldsymbol{\theta}^*, \gamma_0)dy. \end{aligned} \tag{18}$$

Now since $Y_{il}'s$ are independent and identically distributed across l, it follows that

$$\boldsymbol{Var}\left[g_B^{(i)}(y)\right] \leq \frac{1}{Bc_B}\int_{\mathbb{R}} K^2(t)h_J(y - tc_B|\boldsymbol{\theta}^*, \gamma_0)dt. \tag{19}$$

Now plugging (19) into (18), interchanging the order of integration (using Tonelli's Theorem), we get

$$\boldsymbol{P}\left(B^{\frac{1}{2}}R_{1B}^{(2)} > \epsilon\right) \leq C\left(B^{\frac{1}{2}}c_B\right)^{-1}\int_{-\alpha_B}^{\alpha_B}|u_J(y|\boldsymbol{\theta}^*,\gamma_0)|dy \to 0,$$

where C is a universal constant, and the last convergence follows from conditions (**M5**)–(**M6**). We now deal with $R_{1B}^{(1)}$. To this end, we need to calculate $\left(\boldsymbol{E}\left[g_B^{(i)}(y)\right] - h_J(y|\boldsymbol{\theta}^*,\gamma_0)\right)^2$. Using change of variables, two-step Taylor approximation, and assumption (**B1**), we get

$$\begin{aligned}\boldsymbol{E}\left[g_B^{(i)}(y)\right] - h_J(y|\boldsymbol{\theta}^*,\gamma_0) &= \int_{\mathbb{R}} K(t)\left(h_J(y-tc_B|\boldsymbol{\theta}^*,\gamma_0) - h_J(y|\boldsymbol{\theta}^*,\gamma_0)\right)dt \\ &= \int_{\mathbb{R}} K(t)\frac{(tc_B)^2}{2}h_J''(y_B^*(t)|\boldsymbol{\theta}^*,\gamma_0)dt. \end{aligned} \tag{20}$$

Now plugging in (20) into (17) and using conditions **(M3)** and **(M6)**, we get

$$B^{\frac{1}{2}}R_{1B}^{(1)}(\gamma_0) \leq CB^{\frac{1}{2}}c_B^4\int_{-\alpha_B}^{\alpha_B}|u_J(y|\boldsymbol{\theta}^*,\gamma_0)|dy. \tag{21}$$

Convergence of (21) to 0 now follows from condition **(M6)**. We next deal with $R_{2B}(\gamma_0)$. To this end, by writing our the square term of $d_J(\cdot|\boldsymbol{\theta}^*(\gamma_0))$, we have

$$B^{\frac{1}{2}}R_{2B}(\gamma_0) = \int_{|y|\geq\alpha_B}|u_J(y|\boldsymbol{\theta}^*,\gamma_0)|\left(h_J(y|\boldsymbol{\theta}^*,\gamma_0) + g_B^{(i)}(y) - s_J(y|\boldsymbol{\theta}^*,\gamma_0)g_B^{(i)\frac{1}{2}}(y)\right)dy. \tag{22}$$

We will show that the RHS of (22) converges to 0 as $B \to \infty$. We begin with the first term. Please note that by Cauchy-Schwarz inequality,

$$\begin{aligned}B\left(\int_{|y|\geq\alpha_B}|u_J(y|\boldsymbol{\theta}^*,\gamma_0)|h_J(y|\boldsymbol{\theta}^*,\gamma_0)dy\right)^2 &\leq \left\{\int_{\mathbb{R}}u_J(y|\boldsymbol{\theta}^*,\gamma_0)u_J'(y|\boldsymbol{\theta}^*,\gamma_0)h_J(y|\boldsymbol{\theta}^*,\gamma_0)dy\right\} \\ &\quad \{BP_{\boldsymbol{\theta}^*(\gamma_0)}(|\Delta| \geq \alpha_B)\},\end{aligned}$$

the last term converges to 0 by (**M4**). As for the second term, note that, a.s., by Cauchy-Schwarz inequality,

$$\left(\int_{|y|\geq\alpha_B}|u_J(y|\boldsymbol{\theta}^*,\gamma_0)|g_B^{(i)}(y)dy\right)^2 \leq \int_{|y|\geq\alpha_B}u_J(y|\boldsymbol{\theta}^*,\gamma_0)u_J'(y|\boldsymbol{\theta}^*,\gamma_0)g_B^{(i)}(y)dy.$$

Now taking the expectation and using Cauchy-Schwarz inequality, one can show that

$$B\boldsymbol{E}\left[\int_{|y|\geq\alpha_m}|u_J(y|\boldsymbol{\theta}^*,\gamma_0)|g_B^{(i)}(y)dy\right]^2 \leq a_B\int_{\mathbb{R}}K(t)\int_{\mathbb{R}}u_J(y|\boldsymbol{\theta}^*,\gamma_0)u_J'(y|\boldsymbol{\theta}^*,\gamma_0)h_J(y-c_Bt|\boldsymbol{\theta}^*,\gamma_0)dydt,$$

where $a_B = B\sup_{z\in Supp(K)}P_{\boldsymbol{\theta}^*}\left(|\Delta - c_Bz| > \alpha_B\right)$. The convergence to 0 of the RHS of above inequality now follows from condition (**M4**). Finally, by another application of the Cauchy-Schwarz inequality,

$$B\boldsymbol{E}\left[\int_{|y|\geq\alpha_m}|u_J(y|\boldsymbol{\theta}^*,\gamma_0)|g_B^{(i)\frac{1}{2}}(y)s_J(y|\boldsymbol{\theta}^*,\gamma_0)dy\right] \leq a_B\int_{\mathbb{R}}u_J(y-c_Bt|\boldsymbol{\theta}^*,\gamma_0)u_J'(y-c_Bt|\boldsymbol{\theta}^*,\gamma_0)h_J(y|\boldsymbol{\theta}^*,\gamma_0)dy.$$

The convergence of RHS of above inequality to zero follows from (**M4**). Now the lemma follows. □

Proof of Theorem 1. Recall that

$$B^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}}_{i,B}(\gamma_0)-\boldsymbol{\theta}^*(\gamma_0)\right)' = A_{1B}(\gamma_0)+A_{2B}(\gamma_0),$$

where $A_{1B}(\gamma_0)$ and $A_{2B}(\gamma_0)$ are given in (9). Proposition 3 shows that $\lim_{\gamma_0\to 0}\lim_{B\to\infty} A_{1B}(\gamma_0) = N(0, I^{-1}(\boldsymbol{\theta}^*))$; while Lemma 4 shows that $\lim_{\gamma_0\to 0}\lim_{B\to\infty} A_{2B}(\gamma_0) = 0$ in probability. The result follows from Slutsky's theorem. □

We next show that by interchanging the limits, namely first allowing γ_0 to converge to 0 and then letting $B\to\infty$ the limit distribution of $\hat{\boldsymbol{\theta}}_{i,B}(\gamma_0)$ is Gaussian with the same covariance matrix as Theorem 1. We begin with additional assumptions required in the proof of the theorem.

Regularity conditions

(M4′) Let $\{\alpha_B : B\geq 1\}$ be a sequence diverging to infinity. Assume that

$$\lim_{B\to\infty} B \sup_{t\in Supp(K)} \boldsymbol{P}_{\boldsymbol{\theta}^*}\left(|\Delta - c_B t| > \alpha_B\right) = 0,$$

where $Supp(\boldsymbol{K})$ is the support of the kernel density $K(\cdot)$ and Δ is a generic random variable with density $h^{*J}(\cdot|\boldsymbol{\theta}^*)$.

(M5′) Let

$$M_B = \sup_{|y|\leq\alpha_B}\sup_{t\in Supp(K)}\left|\frac{h^{*J}(y-tc_B|\boldsymbol{\theta}^*)}{h^{*J}(y|\boldsymbol{\theta}^*)}\right|.$$

Assume that $\sup_{B\geq 1} M_B < \infty$.

(M6′) The score function has a regular central behavior relative to the smoothing constants, i.e.,

$$\lim_{B\to\infty}(B^{\frac{1}{2}}c_B)^{-1}\int_{-\alpha_B}^{\alpha_B} v_J(y|\boldsymbol{\theta}^*)dy = 0.$$

Furthermore,

$$\lim_{B\to\infty}(B^{\frac{1}{2}}c_B^4)\int_{-\alpha_B}^{\alpha_B} v_J(y|\boldsymbol{\theta}^*)dy = 0.$$

(M7′) The density functions are smooth in an $\boldsymbol{L}_2$ sense; i.e.,

$$\lim_{B\to\infty}\sup_{t\in Supp(K)}\int_{\mathbb{R}}\left(v_J(y+c_Bt|\boldsymbol{\theta}^*)-v_J(y|\boldsymbol{\theta}^*)\right)^2 h^{*J}(y|\boldsymbol{\theta}^*)dy = 0.$$

Assumptions comparing models for original and compressed data

(V1) Assume that $\lim_{\gamma_0\to 0}\sup_y |u_J(y|\boldsymbol{\theta}^*,\gamma_0) - v_J(y|\boldsymbol{\theta}^*)| = 0$.

(V2) $v_J(\cdot|\boldsymbol{\theta})$ is $\boldsymbol{L}_1$ continuous in the sense that $X_n \xrightarrow{p} X$ implies that $\boldsymbol{E}\left[v_J(X_n|\boldsymbol{\theta}) - v_J(X|\boldsymbol{\theta})\right] = 0$, where the expectation is with respect to distribution $K(\cdot)$.

(V3) Assume that for all $\boldsymbol{\theta}\in\boldsymbol{\Theta}$, $\int_{\mathbb{R}}\nabla h^{*J}(y|\boldsymbol{\theta})dy < \infty$.

(V4) Assume that for all $\boldsymbol{\theta}\in\boldsymbol{\Theta}$, $\lim_{\gamma_0\to 0}\sup_y\left|\frac{s_J(y|\boldsymbol{\theta},\gamma_0)}{t_J(y|\boldsymbol{\theta})} - 1\right| = 0$.

Theorem 2. *Assume that the conditions* ***(B1)–(B2)***, ***(D1′)–(D2′)***, ***(M1′)–(M7′)***, ***(O1)–(O2)*** *and* ***(V1)–(V4)*** *hold. Then,*

$$\lim_{B\to\infty}\lim_{\gamma_0\to 0}\boldsymbol{P}\left(\sqrt{B}\left(\hat{\boldsymbol{\theta}}_{i,B}(\gamma_0)-\boldsymbol{\theta}^*(\gamma_0)\right)\leq x\right) = \boldsymbol{P}\left(G\leq x\right),$$

where G is a bivariate Gaussian random variable with mean 0 *and variance* $I^{-1}(\boldsymbol{\theta}^*)$.

We notice that in the above Theorem 2 that we use conditions (**V2**)–(**V4**) which are regularity conditions on the scores of the $J-$ fold convolution of $f(\cdot)$ while (**V1**) facilitates comparison of the scores of the densities of the compressed data and that of the $J-$fold convolution. As before, we will first establish (a):

$$\lim_{B\to\infty}\lim_{\gamma_0\to 0}\boldsymbol{P}\left(A_{1B}(\gamma_0)\le x\right)=\boldsymbol{P}\left(G\le x\right),$$

and then (b): $\lim_{B\to\infty}\lim_{\gamma_0\to 0}A_{2B}(\gamma_0)=0$ in probability. We start with the proof of (a).

Proposition 4. *Assume that the conditions* (***B1***)*–*(***B2***)*,* (***D1′***)*–*(***D2′***)*,* (***M1′***)*–*(***M3′***)*,* (***M7′***)*,* (***O1***)*–*(***O2***)*, and* (***V1***)*–*(***V2***)* *hold. Then,*

$$\lim_{B\to\infty}\lim_{\gamma_0\to 0}\boldsymbol{P}\left(A_{1B}(\gamma_0)\le x\right)=\boldsymbol{P}\left(G\le x\right).$$

We divide the proof of Proposition 4 into two lemmas. In the first lemma, we will show that

$$\lim_{B\to\infty}\lim_{\gamma_0\to 0}D_B(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0))=\frac{1}{4}I(\boldsymbol{\theta}^*).$$

In the second lemma, we will show that first let $\gamma_0\to 0$, then let $B\to\infty$,

$$4B^{\frac{1}{2}}T_B(\gamma_0)\xrightarrow{d}N\left(0,I(\boldsymbol{\theta}^*)\right).$$

Lemma 5. *Assume that the conditions* (***B1***)*–*(***B2***)*,* (***D1′***)*–*(***D2′***)*,* (***M1′***)*–*(***M3′***)*,* (***O1***)*–*(***O2***)*, and* (***V1***)*–*(***V2***)* *hold. Then,*

$$\lim_{B\to\infty}\lim_{\gamma_0\to 0}D_B(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0))=\frac{1}{4}I(\boldsymbol{\theta}^*). \tag{23}$$

Proof. First fix B. Recall that

$$\begin{aligned}D_B(\boldsymbol{\theta}(\gamma_0)) &= -\frac{1}{2}\int_{\mathbb{R}}\dot{u}_J(y|\boldsymbol{\theta},\gamma_0)s_J(y|\boldsymbol{\theta},\gamma_0)(g_B^{(i)}(y))^{\frac{1}{2}}dy\\ &\quad -\frac{1}{4}\int_{\mathbb{R}}u_J(y|\boldsymbol{\theta},\gamma_0)u_J'(y|\boldsymbol{\theta},\gamma_0)s_J(y|\boldsymbol{\theta},\gamma_0)(g_B^{(i)}(y))^{\frac{1}{2}}dy\\ &\equiv D_{1B}(\boldsymbol{\theta}(\gamma_0))+D_{2B}(\boldsymbol{\theta}(\gamma_0)).\end{aligned}$$

By algebra, $D_{1B}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0))$ can be expressed as the sum of $H_{1B}^{(1)}$, $H_{1B}^{(2)}$, $H_{1B}^{(3)}$, $H_{1B}^{(4)}$ and $H_{1B}^{(5)}$, where

$$H_{1B}^{(1)}=-\frac{1}{2}\int_{\mathbb{R}}\left[\dot{u}_J(y|\tilde{\boldsymbol{\theta}}_{i,B},\gamma_0)s_J(y|\tilde{\boldsymbol{\theta}}_{i,B},\gamma_0)-\dot{v}_J(y|\tilde{\boldsymbol{\theta}}_{i,B})t_J(y|\tilde{\boldsymbol{\theta}}_{i,B})\right]g_B^{(i)\frac{1}{2}}(y)dy,$$

$$H_{1B}^{(2)}=-\frac{1}{2}\int_{\mathbb{R}}\left[\dot{v}_J(y|\tilde{\boldsymbol{\theta}}_{i,B})t_J(y|\tilde{\boldsymbol{\theta}}_{i,B})-\dot{v}_J(y|\boldsymbol{\theta}^*)t_J(y|\boldsymbol{\theta}^*)\right]g_B^{(i)\frac{1}{2}}(y)dy,$$

$$H_{1B}^{(3)}=-\frac{1}{2}\int_{\mathbb{R}}\dot{v}_J(y|\boldsymbol{\theta}^*)t_J(y|\boldsymbol{\theta}^*)\left[g_B^{(i)\frac{1}{2}}(y)-h_J^{\frac{1}{2}}(y|\boldsymbol{\theta}^*,\gamma_0)\right]dy,$$

$$H_{1B}^{(4)}=-\frac{1}{2}\int_{\mathbb{R}}\dot{v}_J(y|\boldsymbol{\theta}^*)t_J(y|\boldsymbol{\theta}^*)\left[s_J(y|\boldsymbol{\theta}^*,\gamma_0)-t_J(y|\boldsymbol{\theta}^*)\right]dy,\quad\text{and}\quad H_{1B}^{(5)}=\frac{1}{2}I(\boldsymbol{\theta}^*).$$

We will show that

$$\lim_{\gamma_0\to 0} D_{1B}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) = H_{1B}^{(2)} + \lim_{\gamma_0\to 0} H_{1B}^{(3)} + H_{1B}^{(5)}, \tag{24}$$

where

$$\lim_{\gamma_0\to 0} H_{1B}^{(3)} = -\frac{1}{2}\int_{\mathbb{R}} \dot{v}_J(y|\boldsymbol{\theta}^*)t_J(y|\boldsymbol{\theta}^*)\left[g_B^{*\frac{1}{2}}(y) - t_J(y|\boldsymbol{\theta}^*)\right]dy \quad \text{and} \tag{25}$$

$g_B^*(\cdot)$ is given in (7). First consider $H_{1B}^{(1)}$. it converges to zero as $\gamma_0 \to 0$ by Cauchy-Schwarz inequality and assumption (**O2**). Next we consider $H_{1B}^{(3)}$. We will first show that

$$\lim_{\gamma_0\to 0} -\frac{1}{2}\int_{\mathbb{R}} \dot{v}_J(y|\boldsymbol{\theta}^*)t_J(y|\boldsymbol{\theta}^*)g_B^{(i)\frac{1}{2}}(y)dy = -\frac{1}{2}\int_{\mathbb{R}} \dot{v}_J(y|\boldsymbol{\theta}^*)t_J(y|\boldsymbol{\theta}^*)g_B^{*\frac{1}{2}}(y)dy.$$

To this end, notice that by Cauchy-Schwarz inequality and boundedness of $\dot{v}_J(y|\boldsymbol{\theta}^*)t_J(y|\boldsymbol{\theta}^*)$ in $\boldsymbol{L}_2$, it follows that there exists a constant C such that

$$\begin{aligned}\left|\int_{\mathbb{R}} \dot{v}_J(y|\boldsymbol{\theta}^*)t_J(y|\boldsymbol{\theta}^*)\left[g_B^{(i)\frac{1}{2}}(y) - g_B^{*\frac{1}{2}}(y)\right]dy\right| &\leq C\left\{\int_{\mathbb{R}}\left(g_B^{(i)\frac{1}{2}}(y) - g_B^{*\frac{1}{2}}(y)\right)^2 dy\right\}^{\frac{1}{2}} \\ &\leq C\left\{\int_{\mathbb{R}}\left|g_B^{(i)}(y) - g_B^*(y)\right|dy\right\}^{\frac{1}{2}}.\end{aligned}$$

It suffices to show that $g_B^{(i)}(\cdot)$ converges to $g_B^*(\cdot)$ in $\boldsymbol{L}_1$. Since

$$\int_{\mathbb{R}} |g_B^{(i)}(y) - g_B^*(y)|dy = 2 - 2\int_{\mathbb{R}} \min\left\{g_B^{(i)}(y), g_B^*(y)\right\}dy,$$

and $\min\left\{g_B^{(i)}(y), g_B^*(y)\right\} \leq g_B^*(y)$, by dominated convergence theorem, $g_B^{(i)}(\cdot) \xrightarrow{\boldsymbol{L}_1} g_B^*(\cdot)$. Next we will show that

$$\lim_{\gamma_0\to 0} -\frac{1}{2}\int_{\mathbb{R}} \dot{v}_J(y|\boldsymbol{\theta}^*)t_J(y|\boldsymbol{\theta}^*)s_J(y|\boldsymbol{\theta}^*,\gamma_0)dy = -\frac{1}{2}\int_{\mathbb{R}} \dot{v}_J(y|\boldsymbol{\theta}^*)t_J(y|\boldsymbol{\theta}^*)t_J(y|\boldsymbol{\theta}^*)dy.$$

In addition, by Cauchy-Schwarz inequality, boundedness of $\dot{v}_J(y|\boldsymbol{\theta}^*)t_J(y|\boldsymbol{\theta}^*)$ in $\boldsymbol{L}_2$ and Scheffe's theorem, we have that $\int_{\mathbb{R}} \dot{v}_J(y|\boldsymbol{\theta}^*)h_J^{\frac{1}{2}}(y|\boldsymbol{\theta}^*,\gamma_0)\left(s_J(y|\boldsymbol{\theta}^*|\gamma_0) - t_J(y|\boldsymbol{\theta}^*)\right)dy$ converges to zero as $\gamma_0 \to 0$. Next we consider $H_{1B}^{(4)}$. it converges to zero by Cauchy-Schwarz inequality and assumption (**M2′**). Thus (24) holds. Now let $B \to \infty$, we will show that $\lim_{B\to\infty} H_{1B}^{(2)} = 0$ and $\lim_{B\to\infty}\lim_{\gamma_0\to 0} H_{1B}^{(3)} = 0$. First consider $\lim_{B\to\infty} H_{1B}^{(2)}$. It converges to zero by Cauchy-Schwarz inequality and assumption (**M2′**). Next we consider $\lim_{B\to\infty}\lim_{\gamma_0\to 0} H_{1B}^{(3)}$. It converges to zero by Cauchy-Schwarz inequality and $\boldsymbol{L}_1$ convergence of $g_B^*(\cdot)$ and $h^{*J}(\cdot|\boldsymbol{\theta}^*)$. Therefore $\lim_{B\to\infty}\lim_{\gamma_0\to 0} D_{1B}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) = \frac{1}{2}I(\boldsymbol{\theta}^*)$.

We now turn to show that $\lim_{B\to\infty}\lim_{\gamma_0\to 0} D_{2B}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) = -\frac{1}{4}I(\boldsymbol{\theta}^*)$. First fix B and express $D_{2B}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0))$ as the sum of $H_{2B}^{(1)}$, $H_{2B}^{(2)}$, $H_{2B}^{(3)}$, $H_{2B}^{(4)}$, and $H_{2B}^{(5)}$, where

$$H_{2B}^{(1)} = -\frac{1}{4}\int_{\mathbb{R}}\left[u_J(y|\tilde{\boldsymbol{\theta}}_{i,B},\gamma_0)u_J'(y|\tilde{\boldsymbol{\theta}}_{i,B},\gamma_0)s_J(y|\tilde{\boldsymbol{\theta}}_{i,B},\gamma_0) - v_J(y|\tilde{\boldsymbol{\theta}}_{i,B})v_J'(y|\tilde{\boldsymbol{\theta}}_{i,B})t_J(y|\tilde{\boldsymbol{\theta}}_{i,B})\right]g_B^{(i)\frac{1}{2}}(y)dy,$$

$$H_{2B}^{(2)} = -\frac{1}{4}\int_{\mathbb{R}}\left[v_J(y|\tilde{\boldsymbol{\theta}}_{i,B})v_J'(y|\tilde{\boldsymbol{\theta}}_{i,B})t_J(y|\tilde{\boldsymbol{\theta}}_{i,B}) - v_J(y|\boldsymbol{\theta}^*)v_J'(y|\boldsymbol{\theta}^*)t_J(y|\boldsymbol{\theta}^*)\right]g_B^{(i)\frac{1}{2}}(y)dy,$$

$$H_{2B}^{(3)} = -\frac{1}{4}\int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*)v_J'(y|\boldsymbol{\theta}^*)t_J(y|\boldsymbol{\theta}^*)\left[g_B^{(i)\frac{1}{2}}(y) - h_J^{\frac{1}{2}}(y|\boldsymbol{\theta}^*,\gamma_0)\right]dy,$$

$$H_{2B}^{(4)} = -\frac{1}{4}\int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*)v_J'(y|\boldsymbol{\theta}^*)t_J(y|\boldsymbol{\theta}^*)\left[s_J(y|\boldsymbol{\theta}^*,\gamma_0) - t_J(y|\boldsymbol{\theta}^*)\right]dy, \quad \text{and} \quad H_{2B}^{(5)} = -\frac{1}{4}I(\boldsymbol{\theta}^*).$$

We will show that

$$\lim_{\gamma_0\to 0} D_{2B}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) = H_{2B}^{(2)} + \lim_{\gamma_0\to 0} H_{2B}^{(3)} + H_{2B}^{(5)}, \quad \text{where} \tag{26}$$

$$\lim_{\gamma_0\to 0} H_{2B}^{(3)} = -\frac{1}{2}\int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*)v_J'(y|\boldsymbol{\theta}^*)t_J(y|\boldsymbol{\theta}^*)\left[g_B^{*\frac{1}{2}}(y) - t_J(y|\boldsymbol{\theta}^*)\right]dy. \tag{27}$$

First consider $H_{2B}^{(1)}$. It converges to zero as $\gamma_0 \to 0$ by Cauchy-Schwarz inequality and assumption (**O1**). Next consider $H_{2B}^{(3)}$. By similar argument as above and boundedness of $v_J^2(y|\boldsymbol{\theta}^*)t_J(y|\boldsymbol{\theta}^*)$, it follows that (27) holds. Next consider $H_{2B}^{(4)}$. It converges to zero as $\gamma_0 \to 0$ by Cauchy-Schwarz inequality and assumption (**M3′**). Now let $B \to \infty$, we will show that $\lim_{B\to\infty} H_{2B}^{(2)} = 0$ and $\lim_{B\to\infty}\lim_{\gamma_0\to 0} H_{2B}^{(3)} = 0$. First consider $H_{2B}^{(2)}$. It converges to zero by Cauchy-Schwarz inequality and assumption (**M3′**) as $B \to \infty$. Finally consider $\lim_{B\to\infty}\lim_{\gamma_0\to 0} H_{2B}^{(3)}$. It converges to zero by Cauchy-Schwarz inequality and $\boldsymbol{L}_1$ convergence of $g_B^*(\cdot)$ and $h^{*J}(\cdot|\boldsymbol{\theta}^*)$. Thus $\lim_{B\to\infty}\lim_{\gamma_0\to 0} D_{2B}(\tilde{\boldsymbol{\theta}}_{i,B}(\gamma_0)) = -\frac{1}{4}I(\boldsymbol{\theta}^*)$. Now letting $B \to \infty$, the proof of (23) follows using arguments similar to the one in Lemma 2. □

Lemma 6. *Assume that the conditions* ***(B1)–(B2),(D1′)–(D2′), (M1′)–(M3′), (M7′), (O1)–(O2),*** *and* ***(V1)–(V2)*** *hold. Then, first letting $B \to \infty$, and then letting $\gamma_0 \to 0$,*

$$4B^{\frac{1}{2}}T_B(\gamma_0) \xrightarrow{d} N\left(0, I(\boldsymbol{\theta}^*)\right). \tag{28}$$

Proof. First fix B. We will show that as $\gamma_0 \to 0$,

$$4B^{\frac{1}{2}}T_B(\gamma_0) \xrightarrow{d} \int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*)g_B^*(y)dy.$$

First observe that

$$4B^{\frac{1}{2}}T_B(\gamma_0) - \int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*)g_B^*(y)dy = \int_{\mathbb{R}}\left[u_J(y|\boldsymbol{\theta}^*,\gamma_0) - v_J(y|\boldsymbol{\theta}^*)\right]g_B^{(i)}(y)dy \tag{29}$$

$$+ \int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*)\left[g_B^{(i)}(y) - g_B^*(y)\right]dy. \tag{30}$$

We will show that the RHS of (29) converges to zero as $\gamma_0 \to 0$ and the RHS of (30) converges to zero in probability as $\gamma_0 \to 0$. First consider the RHS of (29). Since

$$\int_{\mathbb{R}}\left[u_J(y|\boldsymbol{\theta}^*,\gamma_0) - v_J(y|\boldsymbol{\theta}^*)\right]g_B^{(i)}(y)dy \le \int_{\mathbb{R}}\sup_y |u_J(y|\boldsymbol{\theta}^*,\gamma_0) - v_J(y|\boldsymbol{\theta}^*)|g_B^{(i)}(y)dy,$$

which converges to zero as $\gamma_0 \to 0$ by assumption (**V1**). Next consider the RHS of (30). Since

$$\int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*)\left[g_B^{(i)}(y) - g_B^*(y)\right]dy = \frac{1}{B}\sum_{l=1}^{B}\int_{\mathbb{R}}\left[v_J(Y_{il} + uc_B) - v_J(Y_{il}^* + uc_B)\right]K(u)du.$$

By assumption (**V2**), it follows that as $\gamma_0 \to 0$, (30) converges to zero in probability. Now letting $B \to \infty$, we have

$$B^{\frac{1}{2}}\int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*)g_B^*(y)dy - B^{\frac{1}{2}}\frac{1}{B}\sum_{l=1}^{B} v_J(Y_{il}^*|\boldsymbol{\theta}^*) = B^{\frac{1}{2}}\frac{1}{B}\sum_{l=1}^{B}\int_{\mathbb{R}}\left(v_J(Y_{il}^* + c_B t|\boldsymbol{\theta}^*) - v_J(Y_{il}^*|\boldsymbol{\theta}^*)\right)K(t)dt,$$

and

$$\begin{aligned}
\boldsymbol{E}\left[B^{\frac{1}{2}}\int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*)g_B^*(y)dy - B^{\frac{1}{2}}\frac{1}{B}\sum_{l=1}^{B} v_J(Y_{il}^*|\boldsymbol{\theta}^*)\right]^2 &= \boldsymbol{E}\left[B^{\frac{1}{2}}\frac{1}{B}\sum_{l=1}^{B}\int_{\mathbb{R}}\left(v_J(Y_{il}^*+c_Bt|\boldsymbol{\theta}^*)-v_J(Y_{il}^*|\boldsymbol{\theta}^*)\right)K(t)dt\right]^2 \\
&\leq C\boldsymbol{E}\left[\int_{\mathbb{R}}\left(v_J(Y_{i1}^*+c_Bt|\boldsymbol{\theta}^*)-v_J(Y_{i1}^*|\boldsymbol{\theta}^*)\right)^2 dt\right] \\
&= C\int_{\mathbb{R}}\int_{\mathbb{R}}\left(v_J(y+c_Bt|\boldsymbol{\theta}^*)-v_J(y|\boldsymbol{\theta}^*)\right)^2 h^{*J}(y|\boldsymbol{\theta}^*)dydt \\
&\to 0 \quad \text{as} \quad B\to\infty,
\end{aligned}$$

where the last convergence follows by assumption (**M7′**). Hence, using the Central limit theorem for independent and identically distributed random variables it follows that the limiting distribution of $B^{\frac{1}{2}}\int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*)g_B^*(y)dy$ is $N(0, I(\boldsymbol{\theta}^*))$, proving the lemma. □

Proof of Proposition 4. The proof of Proposition 4 follows by combining Lemmas 5 and 6. □

Lemma 7. *Assume that the conditions (**M1′**)–(**M6′**) and (**V1**)–(**V4**) hold. Then,*

$$\lim_{B\to\infty}\lim_{\gamma_0\to 0} A_{2B}(\gamma_0) = 0 \quad \textit{in probability.}$$

Proof. First fix B. Let

$$\mathcal{H}_B(\gamma_0) = \int_{\mathbb{R}} u_J(y|\boldsymbol{\theta}^*,\gamma_0)\left[h_J^{\frac{1}{2}}(y|\boldsymbol{\theta}^*,\gamma_0) - g_B^{(i)}(y)\right]^2 dy - \int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*)\left[t_J(y|\boldsymbol{\theta}^*) - g_B^*(y)\right]^2 dy.$$

we will show that as $\gamma_0 \to 0$, $\mathcal{H}_B(\gamma_0) \to 0$. By algebra, $\mathcal{H}_B(\gamma_0)$ can be written as the sum of $\mathcal{H}_{1B}(\gamma_0)$ and $\mathcal{H}_{2B}(\gamma_0)$, where

$$\mathcal{H}_{1B}(\gamma_0) = \int_{\mathbb{R}}\left(u_J(y|\boldsymbol{\theta}^*,\gamma_0) - v_J(y|\boldsymbol{\theta}^*)\right)\left[h_J^{\frac{1}{2}}(y|\boldsymbol{\theta}^*,\gamma_0) - g_B^{(i)}(y)\right]^2 dy, \quad \text{and}$$

$$\mathcal{H}_{2B}(\gamma_0) = \int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*)\left[h_J^{\frac{1}{2}}(y|\boldsymbol{\theta}^*,\gamma_0) - g_B^{(i)}(y)\right]^2 dy.$$

First consider $\mathcal{H}_{1B}(\gamma_0)$. It is bounded above by $C\sup_y |u_J(y|\boldsymbol{\theta}^*,\gamma_0) - v_J(y|\boldsymbol{\theta}^*)|$, which converges to zero as $\gamma_0 \to 0$ by assumption (**V1**), where C is a constant. Next consider $\mathcal{H}_{2B}(\gamma_0)$. We will show that $\mathcal{H}_{2B}(\gamma_0)$ converges to

$$\int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*)\left[t_J(y|\boldsymbol{\theta}^*,\gamma_0) - g_B^*(y)\right]^2 dy.$$

In fact, the difference of $\mathcal{H}_{2B}(\gamma_0)$ and the above formula can be expressed as the sum of $\mathcal{H}_{2B}^{(1)}(\gamma_0)$, $\mathcal{H}_{2B}^{(2)}(\gamma_0)$, and $\mathcal{H}_{2B}^{(3)}(\gamma_0)$, where

$$\mathcal{H}_{2B}^{(1)}(\gamma_0) = \int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*)\left(h_J(y|\boldsymbol{\theta}^*,\gamma_0) - h^{*J}(y|\boldsymbol{\theta}^*)\right)dy,$$

$$\mathcal{H}_{2B}^{(2)}(\gamma_0) = \int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*)\left(g_B^{(i)}(y) - g_B^*(y)\right)dy, \quad \text{and}$$

$$\mathcal{H}_{2B}^{(3)}(\gamma_0) = \int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*)\left(h_J^{\frac{1}{2}}(y|\boldsymbol{\theta}^*,\gamma_0)g_B^{(i)}(y) - t_J(y|\boldsymbol{\theta}^*,\gamma_0)g_B^*(y)\right)dy.$$

First consider $\mathcal{H}_{2B}^{(1)}(\gamma_0)$. Please note that

$$\begin{aligned}\left|\mathcal{H}_{2B}^{(1)}(\gamma_0)\right| &\leq \int_{\mathbb{R}} |\nabla h^{*J}(y|\boldsymbol{\theta}^*)| \left|\frac{h_J(y|\boldsymbol{\theta}^*,\gamma_0)}{h^{*J}(y|\boldsymbol{\theta}^*)} - 1\right| dy \\ &\leq \left\{\left(\sup_y \left|\frac{s_J(y|\boldsymbol{\theta},\gamma_0)}{t_J(y|\boldsymbol{\theta})} - 1\right|\right)^2 + 2\sup_y \left|\frac{s_J(y|\boldsymbol{\theta},\gamma_0)}{t_J(y|\boldsymbol{\theta})} - 1\right|\right\} \int_{\mathbb{R}} |\nabla h^{*J}(y|\boldsymbol{\theta}^*)| dy,\end{aligned}$$

which converges to 0 as $\gamma_0 \to 0$ by assumptions (**V3**) and (**V4**). Next we consider $\mathcal{H}_{2B}^{(2)}(\gamma_0)$. Since

$$\mathcal{H}_{2B}^{(2)}(\gamma_0) = \frac{1}{B}\sum_{l=1}^{B} \int_{\mathbb{R}} \left(v_J(Y_{il} + uc_B|\boldsymbol{\theta}^*) - v_J(Y_{il}^* + uc_B|\boldsymbol{\theta}^*)\right) K(u) du,$$

which converges to zero as $\gamma_0 \to 0$ due to assumption (**V2**). Finally consider $\mathcal{H}_{2B}^{(3)}(\gamma_0)$, which can be expressed as the sum of $\mathcal{L}_{1B}(\gamma_0)$ and $\mathcal{L}_{2B}$, where

$$\mathcal{L}_{1B}(\gamma_0) = \int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*) \left(h_J^{\frac{1}{2}}(y|\boldsymbol{\theta}^*,\gamma_0) - t_J(y|\boldsymbol{\theta}^*)\right) g_B^{(i)\frac{1}{2}}(y) dy, \quad \text{and}$$

$$\mathcal{L}_{2B} = \int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*) t_J(y|\boldsymbol{\theta}^*) \left(g_B^{(i)\frac{1}{2}}(y) - g_B^{*\frac{1}{2}}(y)\right) dy.$$

First consider $\mathcal{L}_{1B}(\gamma_0)$. Notice that

$$|\mathcal{L}_{1B}(\gamma_0)| \leq \sup_y \left|\frac{s_J(y|\boldsymbol{\theta},\gamma_0)}{t_J(y|\boldsymbol{\theta})} - 1\right| \int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*) t_J(y|\boldsymbol{\theta}^*) g_B^{(i)\frac{1}{2}}(y) dy \to 0,$$

where the last convergence follows by Cauchy-Schwarz inequality and assumption (**V4**). Next we consider $\mathcal{L}_{2B}$. By Cauchy-Schwarz inequality, it is bounded above by

$$\left\{\int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*) v_J'(y|\boldsymbol{\theta}^*) h^{*J}(y|\boldsymbol{\theta}^*) dy\right\}^{\frac{1}{2}} \left\{\int_{\mathbb{R}} \left(g_B^{(i)\frac{1}{2}}(y) - g_B^{*\frac{1}{2}}(y)\right)^2 dy\right\}^{\frac{1}{2}}. \tag{31}$$

Equation (31) converges to zero as $\gamma_0 \to 0$ by boundedness of $\int_{\mathbb{R}} v_J(y|\boldsymbol{\theta}^*) v_J'(y|\boldsymbol{\theta}^*) h^{*J}(y|\boldsymbol{\theta}^*) dy$ and $\boldsymbol{L}_1$ convergence between $g_B^{(i)}(\cdot)$ and $g_B^*(\cdot)$, where the $\boldsymbol{L}_1$ convergence has already been established in Lemma 5. Now letting $B \to \infty$, following similar argument as Lemma 4 and assumptions (**M1'**)–(**M6'**), the lemma follows. □

Proof of Theorem 2. Recall that

$$B^{\frac{1}{2}} \left(\hat{\boldsymbol{\theta}}_{i,B}(\gamma_0) - \boldsymbol{\theta}^*(\gamma_0)\right)' = A_{1B}(\gamma_0) + A_{2B}(\gamma_0).$$

Proposition 4 shows that first letting $\gamma_0 \to 0$, then $B \to \infty$, $A_{1B}(\gamma_0) \xrightarrow{d} N(0, I^{-1}(\boldsymbol{\theta}^*))$; while Lemma 7 shows that $\lim_{B\to\infty} \lim_{\gamma_0\to 0} A_{2B}(\gamma_0) = 0$ in probability. The theorem follows from Slutsky's theorem. □

Remark 3. *The above two theorems (Theorems 1 and 2) do not immediately imply the double limit exists. This requires stronger conditions and more delicate calculations and will be considered elsewhere.*

3.5. Robustness of MHDE

In this section, we describe the robustness properties of MHDE for compressed data. Accordingly, let $h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0) \equiv (1-\alpha)h_J(\cdot|\boldsymbol{\theta},\gamma_0) + \alpha\eta_z$, where η_z denotes the uniform density on the interval $(z-\epsilon, z+\epsilon)$, where $\epsilon > 0$ is small, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\alpha \in (0,1)$, and $z \in \mathbb{R}$. Also, let $s_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) = h^{\frac{1}{2}}_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)$, $u_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) = \nabla \log h_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)$, $h^{*J}_{\alpha,z}(\cdot|\boldsymbol{\theta}) \equiv (1-\alpha)h^{*J}(\cdot|\boldsymbol{\theta}) + \alpha\eta_z$, $s^{*J}_{\alpha,z}(\cdot|\boldsymbol{\theta}) = h^{*J\frac{1}{2}}_{\alpha,z}(\cdot|\boldsymbol{\theta})$, and $u^{*J}_{\alpha,z} = \nabla \log h^{*J}_{\alpha,z}(\cdot|\boldsymbol{\theta})$. Before we state the theorem, we describe certain additional assumptions-which are essentially $\boldsymbol{L}_2-$ continuity conditions-that are needed in the proof.

Model assumptions for robustness analysis

(O3) For $\alpha \in [0,1]$ and all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\lim_{\gamma_0 \to 0} \int_{\mathbb{R}} \left(\dot{u}_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) s_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) - \dot{u}^{*J}_{\alpha,z}(y|\boldsymbol{\theta}) s^{*J}_{\alpha,z}(y|\boldsymbol{\theta}) \right)^2 dy = 0.$$

(O4) For $\alpha \in [0,1]$ and all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\lim_{\gamma_0 \to 0} \int_{\mathbb{R}} \left(u_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) u'_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) s_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) - u^{*J}_{\alpha,z}(y|\boldsymbol{\theta}) u^{*J'}_{\alpha,z}(y|\boldsymbol{\theta}) s^{*J}_{\alpha,z}(y|\boldsymbol{\theta}) \right)^2 dy = 0.$$

Theorem 3. *(i) Let $\alpha \in (0,1)$, and assume that for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and assume that the assumptions of Proposition 1 hold, also assume that $T(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0))$ is unique for all z. Then, $T(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0))$ is a bounded, continuous function of z and*

$$\lim_{\gamma_0 \to 0} \lim_{|z| \to \infty} T(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0)) = \boldsymbol{\theta}; \tag{32}$$

*(ii) Assume further that the conditions (**V1**), (**M2**)-(**M3**), and (**O3**)-(**O4**) hold. Then,*

$$\lim_{\gamma_0 \to 0} \lim_{\alpha \to 0} \alpha^{-1} \left[T(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0)) - \boldsymbol{\theta} \right] = [I(\boldsymbol{\theta})]^{-1} \int_{\mathbb{R}} \left[\eta_z(y) v_J(y|\boldsymbol{\theta}) \right] dy,$$

Proof. Let $\boldsymbol{\theta}_z(\gamma_0)$ denote $T(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0))$ and let $\boldsymbol{\theta}_z$ denote $T(h^{*J}_{\alpha,z}(\cdot|\boldsymbol{\theta}))$ We first show that (32) holds. Let $\gamma_0 \geq 0$ be fixed. Then, by triangle inequality,

$$\lim_{|z| \to \infty} |\boldsymbol{\theta}_z(\gamma_0) - \boldsymbol{\theta}| \leq \lim_{|z| \to \infty} |\boldsymbol{\theta}_z(\gamma_0) - \boldsymbol{\theta}(\gamma_0)| + \lim_{|z| \to \infty} |\boldsymbol{\theta}(\gamma_0) - \boldsymbol{\theta}|. \tag{33}$$

We will show that the first term of RHS of (33) is equal to zero. Suppose that it is not zero, without loss of generality, by going to a subsequence if necessary, we may assume that $\boldsymbol{\theta}_z \to \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}$ as $|z| \to \infty$. Since $\boldsymbol{\theta}_z(\gamma_0)$ minimizes $HD^2(h_J(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta},\gamma_0))$, it follows that

$$HD^2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}_z,\gamma_0)) \leq HD^2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}',\gamma_0)) \tag{34}$$

for every $\boldsymbol{\theta}' \in \boldsymbol{\Theta}$. We now show that as $|z| \to \infty$,

$$HD^2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}_z,\gamma_0)) \to HD^2((1-\alpha)h_J(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}_1,\gamma_0)). \tag{35}$$

To this end, note that as $|z| \to \infty$, for every y,

$$h_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) \to (1-\alpha)h_J(y|\boldsymbol{\theta},\gamma_0), \quad \text{and} \quad h_J(y|\boldsymbol{\theta}_z,\gamma_0) \to h_J(y|\boldsymbol{\theta}_1,\gamma_0)$$

Therefore, as $|z| \to \infty$,

$$\left| HD^2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}_z,\gamma_0)) - HD^2((1-\alpha)h_J(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}_1,\gamma_0)) \right| \leq 2(Q_1 + Q_2),$$

where

$$Q_1 = \int_{\mathbb{R}} \left| h_{J,\alpha,z}^{\frac{1}{2}}(y|\boldsymbol{\theta},\gamma_0) - \left((1-\alpha)h_J(y|\boldsymbol{\theta},\gamma_0)\right)^{\frac{1}{2}} \right| \left(h_J(y|\boldsymbol{\theta}_z,\gamma_0))\right)^{\frac{1}{2}} dy,$$

$$Q_2 = \int_{\mathbb{R}} \left| h_J^{\frac{1}{2}}(y|\boldsymbol{\theta}_z,\gamma_0) - \left(h_J(y|\boldsymbol{\theta}_1,\gamma_0)\right)^{\frac{1}{2}} \right| \left((1-\alpha)h_J(y|\boldsymbol{\theta},\gamma_0)\right)^{\frac{1}{2}} dy.$$

Now, by Cauchy-Schwarz inequality and Scheffe's theorem, it follows that as $|z| \to \infty$, $Q_1 \to 0$ and $Q_2 \to 0$. Therefore, (35) holds. By Equations (34) and (35), we have

$$HD^2((1-\alpha)h_J(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}_1,\gamma_0)) \leq HD^2((1-\alpha)h_J(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}',\gamma_0)) \tag{36}$$

for every $\boldsymbol{\theta}' \in \boldsymbol{\Theta}$. Now consider

$$HIF(\alpha, h_J(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}',\gamma_0)) \equiv \int_{\mathbb{R}} \left(\left[(1-\alpha)\delta(h_J(\cdot|\boldsymbol{\theta},\gamma_0), h_J(y|\boldsymbol{\theta}',\gamma_0)) + 1\right]^{\frac{1}{2}} - 1 \right)^2 h_J(y|\boldsymbol{\theta}',\gamma_0)dy,$$

where $\delta(h_J(\cdot|\boldsymbol{\theta},\gamma_0), h_J(y|\boldsymbol{\theta}',\gamma_0)) = \frac{h_J(y|\boldsymbol{\theta},\gamma_0)}{h_J(y|\boldsymbol{\theta}',\gamma_0)} - 1$. Since $G^*(\delta) = \left[((1-\alpha)\delta+1)^{\frac{1}{2}} - 1\right]^2$ is a non-negative and strictly convex function with $\delta = 0$ as the unique point of minimum. Hence $HIF(\alpha, h_J(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}',\gamma_0)) > 0$ unless $\delta(h_J(\cdot|\boldsymbol{\theta},\gamma_0), h_J(y|\boldsymbol{\theta}',\gamma_0)) = 0$ on a set of Lebesgue measure zero, which by the model identifiability assumption , is true if and only if $\boldsymbol{\theta}' = \boldsymbol{\theta}$. Since $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}$, it follows that

$$HIF(\alpha, h_J(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}_1,\gamma_0)) > HIF(\alpha, h_J(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta},\gamma_0)).$$

Since $HIF(\alpha, h_J(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}',\gamma_0)) = HD^2((1-\alpha)h_J(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}',\gamma_0)) - \alpha$. This implies that

$$HD^2((1-\alpha)h_J(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}_1,\gamma_0)) > HD^2((1-\alpha)h_J(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}',\gamma_0)),$$

which contradicts (36). The continuity of $\boldsymbol{\theta}_z$ follows from Proposition 2 and the boundedness follows from the compactness of $\boldsymbol{\Theta}$. Now let $\gamma_0 \to 0$, the second term of RHS of (33) converges to zero by Proposition 2.

We now turn to part (ii) of the Theorem. First fix $\gamma_0 \geq 0$. Since $\boldsymbol{\theta}_z$ minimizes $H^2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), t(\gamma_0))$ over $\boldsymbol{\Theta}$. By Taylor expansion of $HD^2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}_z,\gamma_0))$ around $\boldsymbol{\theta}$, we get

$$\begin{aligned} 0 = \nabla HD^2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}_z,\gamma_0)) &= HD^2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta},\gamma_0)) \\ &\quad + (\boldsymbol{\theta}_z - \boldsymbol{\theta}) D(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}_z^*,\gamma_0)), \end{aligned}$$

where $\boldsymbol{\theta}_z^*(\gamma_0)$ is a point between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_z$,

$$\nabla HD^2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta},\gamma_0)) = \frac{1}{2}\int_{\mathbb{R}} u_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) s_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) \left(s_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) - s_J(y|\boldsymbol{\theta},\gamma_0)\right) dy, \tag{37}$$

and $D(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}',\gamma_0))$ can be expressed the sum of $D_1(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}',\gamma_0))$ and $D_2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}',\gamma_0))$, where

$$D_1(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}',\gamma_0)) = \frac{1}{2}\int_{\mathbb{R}} \dot{u}_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) s_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) s_J(y|\boldsymbol{\theta}',\gamma_0) dy \quad \text{and} \tag{38}$$

$$D_2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}',\gamma_0)) = \frac{1}{4}\int_{\mathbb{R}} u_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) u'_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) s_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) s_J(y|\boldsymbol{\theta}',\gamma_0) dy. \tag{39}$$

Therefore,

$$\alpha^{-1}\left(\boldsymbol{\theta}_z-\boldsymbol{\theta}\right)=-\alpha^{-1}D^{-1}(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0),h_J(\cdot|\boldsymbol{\theta}_z^*,\gamma_0))\nabla HD^2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0),h_J(\cdot|\boldsymbol{\theta},\gamma_0)).$$

We will show that

$$\lim_{\alpha\to 0}D(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0),h_J(\cdot|\boldsymbol{\theta}_z^*,\gamma_0))=-\frac{1}{4}I(\boldsymbol{\theta}(\gamma_0)),\quad\text{and}\tag{40}$$

$$\lim_{\alpha\to 0}\alpha^{-1}\nabla HD^2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0),h_J(\cdot|\boldsymbol{\theta},\gamma_0))=\frac{1}{4}\int_{\mathbb{R}}\left[\eta_z(y)u_J(y|\boldsymbol{\theta},\gamma_0)\right]dy.\tag{41}$$

We will first establish (40). Please note that as $\alpha\to 0$, by definition $\boldsymbol{\theta}_z(\alpha)\to\boldsymbol{\theta}$. Thus, $\lim_{\alpha\to 0}\boldsymbol{\theta}_z^*(\alpha)=\boldsymbol{\theta}$. In addition, by assumptions (**O3**) and (**O4**), $D(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0),h_J(\cdot|\boldsymbol{\theta}_z,\gamma_0))$ is continuous in $\boldsymbol{\theta}_z$. Therefore, to prove (40), it suffices to show that

$$\lim_{\alpha\to 0}D_1(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0),h_J(\cdot|\boldsymbol{\theta},\gamma_0))=-\frac{1}{2}\int_{\mathbb{R}}\dot{u}_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)h_J(y|\boldsymbol{\theta},\gamma_0)dy=-\frac{1}{2}I(\boldsymbol{\theta}(\gamma_0)),\quad\text{and}\tag{42}$$

$$\lim_{\alpha\to 0}D_2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0),h_J(\cdot|\boldsymbol{\theta},\gamma_0))=\frac{1}{4}\int_{\mathbb{R}}u_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)u'_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)h_J(y|\boldsymbol{\theta},\gamma_0)dy=\frac{1}{4}I(\boldsymbol{\theta}(\gamma_0)).\tag{43}$$

We begin with $D_2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0),h_J(\cdot|\boldsymbol{\theta},\gamma_0))$. Notice that

$$\lim_{\alpha\to 0}s_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)=s_J(y|\boldsymbol{\theta},\gamma_0),\quad\lim_{\alpha\to 0}u_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)=u_J(y|\boldsymbol{\theta},\gamma_0),\quad\text{and}$$

$$\lim_{\alpha\to 0}\dot{u}_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)=\dot{u}_J(y|\boldsymbol{\theta},\gamma_0).$$

Thus,

$$\lim_{\alpha\to 0}\dot{u}_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)s_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)s_J(y|\boldsymbol{\theta},\gamma_0)=\dot{u}_J(y|\boldsymbol{\theta},\gamma_0)h_J(y|\boldsymbol{\theta},\gamma_0).$$

In addition, in order to pass the limit inside the integral, note that, for every component of matrix $u_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0)u'_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0)$, we have

$$\begin{aligned}\left|u_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)u'_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)\right| &= \left|\left(\frac{(1-\alpha)\nabla h_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)}{(1-\alpha)h_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)+\alpha\eta_z(y)}\right)\left(\frac{(1-\alpha)\nabla h_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)}{(1-\alpha)h_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)+\alpha\eta_z(y)}\right)'\right|\\ &= \left|\left(\frac{\nabla h_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)}{h_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)+\frac{\alpha}{1-\alpha}\eta_z(y)}\right)\left(\frac{\nabla h_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)}{h_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)+\frac{\alpha}{1-\alpha}\eta_z(y)}\right)'\right|\\ &\le \left|\left(\frac{\nabla h_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)}{h_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)}\right)\left(\frac{\nabla h_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)}{h_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)}\right)'\right| = \left|u_J(y|\boldsymbol{\theta},\gamma_0)u'_J(y|\boldsymbol{\theta},\gamma_0)\right|,\end{aligned}$$

where $|\cdot|$ represents the absolute function for each component of the matrix, and

$$\left|s_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)\right|\le\left[h_J(y|\boldsymbol{\theta},\gamma_0)+\eta_z(y)\right]^{\frac{1}{2}}.$$

Now choosing the dominating function

$$m_J^{(1)}(y|\boldsymbol{\theta},\gamma_0)=\left|u_J(y|\boldsymbol{\theta},\gamma_0)u'_J(y|\boldsymbol{\theta},\gamma_0)\right|\left[h_J(y|\boldsymbol{\theta},\gamma_0)+\eta_z(y)\right]^{\frac{1}{2}}s_J(y|\boldsymbol{\theta},\gamma_0)$$

and applying Cauchy-Schwarz inequality, we obtain that there exists a constant C such that

$$\int_{\mathbb{R}} \left| m_J^{(1)}(y|\boldsymbol{\theta},\gamma_0) \right| dy \quad \leq \quad C \left\{ \int_{\mathbb{R}} \left(u_J(y|\boldsymbol{\theta},\gamma_0) u_J'(y|\boldsymbol{\theta},\gamma_0) s_J(y|\boldsymbol{\theta},\gamma_0) \right)^2 dy \right\}^{\frac{1}{2}},$$

which is finite by assumption (**M2**). Hence, by the dominated convergence theorem, (43) holds. Turning to (42), notice that for each component of the matrix $\dot{u}_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)$,

$$\begin{aligned} \left| \dot{u}_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) \right| &= \left| \frac{\ddot{h}_J(y|\boldsymbol{\theta},\gamma_0) \left[h_J(y|\boldsymbol{\theta},\gamma_0) + \frac{\alpha}{1-\alpha}\eta_z(y) \right] - \left(\nabla h_J(y|\boldsymbol{\theta},\gamma_0) \right) \left(\nabla h_J(y|\boldsymbol{\theta},\gamma_0) \right)'}{\left(h_J(y|\boldsymbol{\theta},\gamma_0) + \frac{\alpha}{1-\alpha}\eta_z(y) \right)^2} \right| \\ &\leq \left| \frac{\ddot{h}_J(y|\boldsymbol{\theta},\gamma_0)}{h_J(y|\boldsymbol{\theta},\gamma_0)} \right| + \left| \frac{\left(\nabla h_J(y|\boldsymbol{\theta},\gamma_0) \right) \left(\nabla h_J(y|\boldsymbol{\theta},\gamma_0) \right)'}{h_J^2(y|\boldsymbol{\theta},\gamma_0)} \right|, \end{aligned}$$

where $|\cdot|$ denotes the absolute function for each component. Now choosing the dominating function

$$m_J^{(2)}(y|\boldsymbol{\theta},\gamma_0) = \left(\left| \frac{\ddot{h}_J(y|\boldsymbol{\theta},\gamma_0)}{h_J(y|\boldsymbol{\theta},\gamma_0)} \right| + \left| \frac{\left(\nabla h_J(y|\boldsymbol{\theta},\gamma_0) \right) \left(\nabla h_J(y|\boldsymbol{\theta},\gamma_0) \right)'}{h_J^2(y|\boldsymbol{\theta},\gamma_0)} \right| \right) \left[h_J(y|\boldsymbol{\theta},\gamma_0) + \eta_z(y) \right]^{\frac{1}{2}} s_J(y|\boldsymbol{\theta},\gamma_0),$$

and applying the Cauchy-Schwarz inequality it follows, using (**M3**), that

$$\int_{\mathbb{R}} \left| m_J^{(2)}(y|\boldsymbol{\theta},\gamma_0) \right| dy < \infty.$$

Finally, by the dominated convergence theorem, it follows that

$$\lim_{\alpha \to 0} D_1(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta}_z^*,\gamma_0)) = -\frac{1}{2} I(\boldsymbol{\theta}(\gamma_0)).$$

Therefore (40) follows. It remains to show that (41) holds. To this end, note that

$$\nabla HD^2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta},\gamma_0)) \quad = \quad -\frac{1}{2} \int_{\mathbb{R}} s_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) u_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) s_J(y|\boldsymbol{\theta},\gamma_0) dy.$$

Now taking partial derivative of $HD^2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta},\gamma_0))$ with respect to α, it can be expressed as the sum of $\mathcal{U}_1, \mathcal{U}_2$ and $\mathcal{U}_3$, where

$$\mathcal{U}_1 = -\frac{1}{4} \int_{\mathbb{R}} \frac{-h_J(y|\boldsymbol{\theta},\gamma_0) + \eta_z(y)}{s_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)} u_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) s_J(y|\boldsymbol{\theta},\gamma_0) dy,$$

$$\mathcal{U}_2 = -\frac{1}{2} \int_{\mathbb{R}} s_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) \frac{-\nabla h_J(y|\boldsymbol{\theta},\gamma_0) h_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)}{h_{J,\alpha,z}^2(y|\boldsymbol{\theta},\gamma_0)} s_J(y|\boldsymbol{\theta},\gamma_0) dy, \quad \text{and}$$

$$\mathcal{U}_3 = -\frac{1}{2} \int_{\mathbb{R}} s_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) \frac{-(1-\alpha) \nabla h_J(y|\boldsymbol{\theta},\gamma_0) (-h_J(y|\boldsymbol{\theta},\gamma_0) + \eta_z(y))}{h_{J,\alpha,z}^2(y|\boldsymbol{\theta},\gamma_0)} s_J(y|\boldsymbol{\theta},\gamma_0) dy.$$

By dominated convergence theorem (using similar idea as above to find dominating functions), we have

$$\lim_{\alpha \to 0} \frac{\partial \nabla HD^2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta},\gamma_0))}{\partial \alpha} = \frac{1}{4} \int_{\mathbb{R}} u_J(y|\boldsymbol{\theta},\gamma_0) \eta_z(y) dy.$$

Hence, by L'Hospital rule, (41) holds. It remains to show that

$$\lim_{\gamma_0\to 0}\lim_{\alpha\to 0} D(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta},\gamma_0)) = -\frac{1}{4}I(\boldsymbol{\theta}), \quad \text{and} \tag{44}$$

$$\lim_{\gamma_0\to 0}\lim_{\alpha\to 0} \alpha^{-1}\nabla HD^2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta},\gamma_0)) = \frac{1}{4}\int_{\mathbb{R}} \left[\eta_z(y)v_J(y|\boldsymbol{\theta})\right] dy. \tag{45}$$

We start with (44). Since for fixed $\gamma_0 \geq 0$, by the above argument, it follows that

$$\lim_{\alpha\to 0} D(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta},\gamma_0)) = -\frac{1}{4}I(\boldsymbol{\theta}(\gamma_0)) = -\frac{1}{4}\int_{\mathbb{R}} u_J(y|\boldsymbol{\theta},\gamma_0)u_J'(y|\boldsymbol{\theta},\gamma_0)h_J(y|\boldsymbol{\theta},\gamma_0)dy,$$

it is enough to show

$$\lim_{\gamma_0\to 0}\int_{\mathbb{R}} u_J(y|\boldsymbol{\theta},\gamma_0)u_J'(y|\boldsymbol{\theta},\gamma_0)h_J(y|\boldsymbol{\theta},\gamma_0)dy = \int_{\mathbb{R}} v_J(y|\boldsymbol{\theta})v_J'(y|\boldsymbol{\theta})h^{*J}(y|\boldsymbol{\theta})dy,$$

which is proved in Lemma 2. Hence (44) holds. Next we prove (45). By the argument used to establish (40), it is enough to show that

$$\lim_{\gamma_0\to 0}\int_{\mathbb{R}} \left[\eta_z(y)u_J(y|\boldsymbol{\theta},\gamma_0)\right] dy = \int_{\mathbb{R}} \left[\eta_z(y)v_J(y|\boldsymbol{\theta})\right] dy. \tag{46}$$

However,

$$\int_{\mathbb{R}} \eta_z(y)\left[u_J(y|\boldsymbol{\theta},\gamma_0) - v_J(y|\boldsymbol{\theta})\right] dy \leq \sup_y \left|u_J(y|\boldsymbol{\theta},\gamma_0) - v_J(y|\boldsymbol{\theta})\right|,$$

and the RHS of the above inequality converges to zero as $\gamma_0 \to 0$ from assumption (**V1**). Hence (46) holds. This completes the proof. □

Our next result is concerned with the behavior of the $\alpha-$influence function when $\gamma_0 \to 0$ first and then $|z| \to \infty$ or $\alpha \to 0$. The following three additional assumptions will be used in the proof of part (ii) of Theorem 4.

Model assumptions for robustness analysis

(**O5**) For $\alpha \in [0,1]$ and all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\dot{u}^{*J}_{\alpha,z}(y|\boldsymbol{\theta})s^{*J}_{\alpha,z}(y|\boldsymbol{\theta})$ is bounded in $\boldsymbol{L}_2$.
(**O6**) For $\alpha \in [0,1]$ and all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $u^{*J}_{\alpha,z}(y|\boldsymbol{\theta})u^{*J'}_{\alpha,z}(y|\boldsymbol{\theta})s^{*J}_{\alpha,z}(y|\boldsymbol{\theta})$ is bounded in $\boldsymbol{L}_2$.
(**O7**) For $\alpha \in [0,1]$ and all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\lim_{\gamma_0\to 0}\int_{\mathbb{R}} \left(s_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)u_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) - s^{*J}_{\alpha,z}(y|\boldsymbol{\theta})u^{*J}_{\alpha,z}(y|\boldsymbol{\theta})\right)^2 dy = 0.$$

Theorem 4. *(i) Let $\alpha \in (0,1)$, and assume that for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, assume that the assumptions of Proposition 1 hold, also assume that $T(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0))$ is unique for all z. Then, $T(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0))$ is a bounded, continuous function of z such that*

$$\lim_{|z|\to\infty}\lim_{\gamma_0\to 0} T(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0)) = \boldsymbol{\theta};$$

*(ii) Assume further that the conditions (**O3**)–(**O7**) hold. Then,*

$$\lim_{\alpha\to 0}\lim_{\gamma_0\to 0} \alpha^{-1}\left[T(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0)) - \boldsymbol{\theta}\right] = \left[I(\boldsymbol{\theta})\right]^{-1}\int_{\mathbb{R}} \left[\eta_z(y)v_J(y|\boldsymbol{\theta})\right] dy.$$

Proof. Let $\boldsymbol{\theta}_z(\gamma_0)$ denote $T(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0))$ and let $\boldsymbol{\theta}_z$ denote $T(h^{*J}_{\alpha,z}(\cdot|\boldsymbol{\theta}))$. First fix $z \in \mathbb{R}$; then by the triangular inequality,

$$\lim_{\gamma_0\to 0} |\boldsymbol{\theta}_z(\gamma_0) - \boldsymbol{\theta}| \leq \lim_{\gamma_0\to 0} |\boldsymbol{\theta}_z(\gamma_0) - \boldsymbol{\theta}_z| + \lim_{\gamma_0\to 0} |\boldsymbol{\theta}_z - \boldsymbol{\theta}|. \tag{47}$$

The first term of RHS of (47) is equal to zero due to proposition 2. Now let $|z| \to \infty$, then the second term on the RHS of (47) converges to zero using similar argument as Theorem 3 with density converging to $h^{*J}_{\alpha,z}(\cdot|\boldsymbol{\theta})$. This completes the proof of I). Turning to (ii), we will prove that

$$\lim_{\alpha\to 0}\lim_{\gamma_0\to 0} D(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta},\gamma_0)) = -\frac{1}{4}I(\boldsymbol{\theta}), \tag{48}$$

$$\lim_{\alpha\to 0}\lim_{\gamma_0\to 0} \alpha^{-1}\nabla HD^2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta},\gamma_0)) = \frac{1}{4}\int_{\mathbb{R}} \left[\eta_z(y)v_J(y|\boldsymbol{\theta})\right] dy. \tag{49}$$

Recall from the proof of part (ii) of Theorem 3 that

$$\begin{aligned} D(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta},\gamma_0)) &= \frac{1}{2}\int_{\mathbb{R}} \dot{u}_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)s_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)s_J(y|\boldsymbol{\theta},\gamma_0) \\ &\quad + \frac{1}{4}\int_{\mathbb{R}} u_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)u'_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)s_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)s_J(y|\boldsymbol{\theta},\gamma_0) \\ &\equiv D_1(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta},\gamma_0)) + D_2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta},\gamma_0)). \end{aligned}$$

We will now show that for fixed $\alpha \in (0,1)$

$$\lim_{\gamma_0\to 0} D_1(h_{J,\alpha,z}(\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta},\gamma_0)) = \frac{1}{2}\int_{\mathbb{R}} \dot{u}^{*J}_{\alpha,z}(y|\boldsymbol{\theta})s^{*J}_{\alpha,z}(y|\boldsymbol{\theta})t_J(y|\boldsymbol{\theta})dy, \quad \text{and} \tag{50}$$

$$\lim_{\gamma_0\to 0} D_2(h_{J,\alpha,z}(\boldsymbol{\theta},\gamma_0), h_J(\cdot|\boldsymbol{\theta},\gamma_0)) = \frac{1}{4}\int_{\mathbb{R}} u^{*J}_{\alpha,z}(y|\boldsymbol{\theta})u^{*J'}_{\alpha,z}(y|\boldsymbol{\theta})s^{*J}_{\alpha,z}(y|\boldsymbol{\theta})t_J(y|\boldsymbol{\theta})dy. \tag{51}$$

We begin with (50). A standard calculation shows that $D_1(h_{J,\alpha,z}(\boldsymbol{\theta},\gamma_0), u^{*J}_{\alpha,z}(y|\boldsymbol{\theta}))$ can be expressed as the sum of $\mathcal{D}_{11}$, $\mathcal{D}_{12}$ and $\mathcal{D}_{13}$, where

$$\mathcal{D}_{11} = \frac{1}{2}\int_{\mathbb{R}} \left(\dot{u}_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)s_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) - \dot{u}^{*J}_{\alpha,z}(y|\boldsymbol{\theta})s^{*J}_{\alpha,z}(y|\boldsymbol{\theta})\right) s_J(y|\boldsymbol{\theta},\gamma_0)dy,$$

$$\mathcal{D}_{12} = \frac{1}{2}\int_{\mathbb{R}} \dot{u}^{*J}_{\alpha,z}(y|\boldsymbol{\theta})s^{*J}_{\alpha,z}(y|\boldsymbol{\theta}) \left(s_J(y|\boldsymbol{\theta},\gamma_0) - t_J(y|\boldsymbol{\theta})\right) dy, \quad \text{and}$$

$$\mathcal{D}_{13} = \frac{1}{2}\int_{\mathbb{R}} \dot{u}^{*J}_{\alpha,z}(y|\boldsymbol{\theta})s^{*J}_{\alpha,z}(y|\boldsymbol{\theta})t_J(y|\boldsymbol{\theta})dy.$$

It can be seen that $\mathcal{D}_{11}$ converges to zero as $\gamma_0 \to 0$ by Cauchy-Schwarz inequality and assumption (**O3**); also, $\mathcal{D}_{12}$ converges to zero as $\gamma_0 \to 0$ by Cauchy-Schwarz inequality, assumption (**O5**) and Scheffe's theorem. Hence (50) follows. Similarly (51) follows as $\gamma_0 \to 0$ by Cauchy-Schwarz inequality, assumption (**O4**), assumption (**O6**) and Scheffe's theorem.

Now let $\alpha \to 0$. Using the same idea as in Theorem 3 to find dominating functions, one can apply the dominated convergence Theorem to establish that

$$\lim_{\alpha\to 0} \frac{1}{2}\int_{\mathbb{R}} \dot{u}^{*J}_{\alpha,z}(y|\boldsymbol{\theta})s^{*J}_{\alpha,z}(y|\boldsymbol{\theta})t_J(y|\boldsymbol{\theta})dy = -\frac{1}{2}I(\boldsymbol{\theta}), \quad \text{and}$$

$$\lim_{\alpha\to 0}\frac{1}{4}\int_{\mathbb{R}} u^{*J}_{\alpha,z}(y|\boldsymbol{\theta})u^{*J'}_{\alpha,z}(y|\boldsymbol{\theta})s^{*J}_{\alpha,z}(y|\boldsymbol{\theta})t_J(y|\boldsymbol{\theta})dy = \frac{1}{4}I(\boldsymbol{\theta}).$$

Hence (48) follows. Finally, it remains to establish (49). First fix $\alpha \in (0,1)$; we will show that

$$\lim_{\gamma_0\to 0}\nabla HD^2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0),h_J(\cdot|\boldsymbol{\theta},\gamma_0)) = -\frac{1}{2}\int_{\mathbb{R}} s^{*J}_{\alpha,z}(y|\boldsymbol{\theta})u^{*J}_{\alpha,z}(y|\boldsymbol{\theta})t_J(y|\boldsymbol{\theta})dy. \tag{52}$$

Please $\nabla HD^2(h_{J,\alpha,z}(\cdot|\boldsymbol{\theta},\gamma_0),h_J(\cdot|\boldsymbol{\theta},\gamma_0))$ can be expressed as the sum of $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_3$, where

$$\mathcal{T}_1 = -\frac{1}{2}\int_{\mathbb{R}}\left(s_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0)u_{J,\alpha,z}(y|\boldsymbol{\theta},\gamma_0) - s^{*J}_{\alpha,z}(y|\boldsymbol{\theta})u^{*J}_{\alpha,z}(y|\boldsymbol{\theta})\right)s_J(y|\boldsymbol{\theta},\gamma_0)dy,$$

$$\mathcal{T}_2 = -\frac{1}{2}\int_{\mathbb{R}} s^{*J}_{\alpha,z}(y|\boldsymbol{\theta})u^{*J}_{\alpha,z}(y|\boldsymbol{\theta})\left(s_J(y|\boldsymbol{\theta},\gamma_0) - t_J(y|\boldsymbol{\theta})\right)dy, \quad \text{and}$$

$$\mathcal{T}_3 = -\frac{1}{2}\int_{\mathbb{R}} s^{*J}_{\alpha,z}(y|\boldsymbol{\theta})u^{*J}_{\alpha,z}(y|\boldsymbol{\theta})t_J(y|\boldsymbol{\theta})dy.$$

It can be seen thatr $\mathcal{T}_1$ converges to zero as $\gamma_0 \to 0$ by Cauchy-Schwarz inequality and assumption (**O7**); $\mathcal{T}_2$ converges to zero as $\gamma_0 \to 0$ by Cauchy-Schwarz inequality, boundedness of $u^{*J}_{\alpha,z}(\cdot)s^{*J}_{\alpha,z}(\cdot)$ in $\boldsymbol{L}_2$, and Scheffe's theorem. Therefore, (52) holds. Finally, letting $\alpha \to 0$ and using the same idea as in Theorem 3 to find the dominating function, it follows by the dominated convergence theorem and L'Hospital rule that (49) holds. This completes the proof of the Theorem. □

Remark 4. *Theorems 3 and 4 do not imply that the double limit exists. This is beyond the scope of this paper.*

In the next section, we describe the implementation details and provide several simulation results in support of our methodology.

4. Implementation and Numerical Results

In this section, we apply the proposed MHD based methods to estimate the unknown parameters $\boldsymbol{\theta} = (\mu, \sigma^2)$ using the compressed data. We set J = 10,000 and $B = 100$. All simulations are based on 5000 replications. We consider the Gaussian kernel and Epanechnikov kernel for the nonparametric density estimation. The Gaussian kernel is given by

$$K(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right),$$

and the Epanechnikov kernel is given by

$$K(x) = \frac{3}{4}\left(1 - x^2\right)1_{(|x|\le 1)}.$$

We generate $\boldsymbol{X}$ and uncontaminated compressed data $\widetilde{\boldsymbol{Y}}$ in the following way:

- Step 1. Generate $\boldsymbol{X}_l$, where $X_{jl} \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$.
- Step 2. Generate $\boldsymbol{R}_l$, where $r_{ijl} \overset{i.i.d.}{\sim} N(1, \gamma_0^2)$.
- Step 3. Generate the uncontaminated $\widetilde{\boldsymbol{Y}}_l$ by calculating $\widetilde{\boldsymbol{Y}}_l = \boldsymbol{R}_l \boldsymbol{X}_l$.

4.1. Objective Function

In practice, we store the compressed data $(\widetilde{\boldsymbol{Y}}_l, \boldsymbol{r}_{\cdot l}, \boldsymbol{\omega}_l)$ for all $1 \leq l \leq B$. Hence if X_{jl} follows Normal distribution with mean μ and variance σ^2, the form of the marginal density of the compressed data, *viz.*, Y_{il} is complicated and does not have a closed form expression. However, for large J, using the local limit theorem its density can be approximated by Gaussian density with mean $\sqrt{J}\mu$ and variance $\sigma^2 + \gamma_0^2(\mu^2 + \sigma^2)$. Hence, we work with U_{il}, where $U_{il} = \frac{\widetilde{Y}_{il} - \mu r_{i \cdot l}}{\omega_{il}}$. Please note that with this transformation, $\boldsymbol{E}[U_{il}] = 0$ and $\boldsymbol{Var}[U_{il}] = \sigma^2$. Hence, the kernel density estimate of the unknown true density is given by

$$g_B^{(i)}(y|\mu) = \frac{1}{Bc_B} \sum_{l=1}^{B} K\left(\frac{y - U_{il}}{c_B}\right).$$

The difference between the kernel density estimate and the one proposed here is that we include the unknown parameter μ in the kernel. Additionally, this allows one to incorporate $(\boldsymbol{r}_{\cdot r}, \boldsymbol{\omega}_l)$ into the kernel. Consequently, only the scale parameter σ is part of the parametric model. Using the local limit theorem, we approximate the true parametric model by $\phi(\cdot|\sigma)$, where $\phi(\cdot|\sigma)$ is the density of $N(0, \sigma^2)$. Hence, the objective function is

$$\Psi(i, \boldsymbol{\theta}) \equiv \mathcal{A}(g_B^{(i)}(\cdot|\mu), \phi(\cdot|\sigma)) = \int_{\mathbb{R}} g_B^{(i)\frac{1}{2}}(y|\mu)\phi^{\frac{1}{2}}(y|\sigma)dy;$$

and, the estimator is given by

$$\hat{\boldsymbol{\theta}}_B(\gamma_0) = \frac{1}{S}\sum_{i=1}^{S} \hat{\boldsymbol{\theta}}_{i,B}(\gamma_0), \quad \text{where} \quad \hat{\boldsymbol{\theta}}_{i,B}(\gamma_0) = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmax}}\, \Psi(i, \boldsymbol{\theta}).$$

It is clear that $\hat{\boldsymbol{\theta}}_B(\gamma_0)$ is a consistent estimator of $\boldsymbol{\theta}^*$. In the next subsection, we use Quasi-Newton method with Broyden-Fletcher-Goldfarb-Shanno (BFGS) update to estimate $\boldsymbol{\theta}$. Quasi-Newton method is appealing since (i) it replaces the complicated calculation of the Hessian matrix with an approximation which is easier to compute ($\Delta_k(\boldsymbol{\theta})$ given in the next subsection) and (ii) gives more flexible step size t (compared to the Newton-Raphson method), ensuring that it does not "jump" too far at every step and hence guaranteeing convergence of the estimating equation. The BFGS update (H_k) is a popular method for approximating the Hessian matrix via gradient evaluations. The step size t is determined using Backtracking line search algorithm described in Algorithm 2. The algorithms are given in detail in the next subsection. Our analysis also includes the case where $S \equiv 1$ and $r_{ijl} \equiv 1$. In this case, as explained previously, one obtains significant reduction in storage and computational complexity. Finally, we emphasize here that the density estimate contains μ and is not parameter free as is typical in classical MHDE analysis. In the next subsection, we describe an algorithm to implement our method.

4.2. Algorithm

As explained previously, we use the Quasi-Newton Algorithm with BFGS update to obtain $\hat{\boldsymbol{\theta}}_{\text{MHDE}}$. To describe this method, consider the objective function (suppressing i) $\Psi(\boldsymbol{\theta})$, which is twice continuously differentiable. Let the initial value of $\boldsymbol{\theta}$ be $\boldsymbol{\theta}^{(0)} = \left(\mu^{(0)}, \sigma^{(0)}\right)$ and $H_0 = I$, where I is the identity matrix.

Algorithm 1: The Quasi-Newton Algorithm.

Set k = 1.

repeat

1. Calculate $\Delta_k(\boldsymbol{\theta}) = -H_{k-1}^{-1}\nabla\Psi(\boldsymbol{\theta}^{(k-1)})$, where $\nabla\Psi(y;\boldsymbol{\theta}^{k-1})$ is the first derivative of $\Psi(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ at $(k-1)$th step.

2. Determine the step length parameter t via backtracking line search.

3. Compute $\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}^{(k-1)} + t\Delta_k(\boldsymbol{\theta})$.

4. Compute H_k, where the BFGS update is

$$H_k = H_{k-1} + \frac{q_{k-1}q_{k-1}^T}{q_{k-1}^T d_{k-1}} - \frac{H_{k-1}d_{k-1}d_{k-1}^T H_{k-1}^T}{d_{k-1}^T H_{k-1} d_{k-1}},$$

where

$$d_{k-1} = \boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}^{(k-1)},$$
$$q_{k-1} = \nabla\Psi(\boldsymbol{\theta}^{(k)}) - \nabla\Psi(\boldsymbol{\theta}^{(k-1)}).$$

5. Compute $e_k = |\Psi(\boldsymbol{\theta}^{(k)}) - \Psi(\boldsymbol{\theta}^{(k-1)})|$.

6. Set $k = k+1$.

until $(e_k) <$ threshold.

Remark 5. *In step 1, one can directly use the Inverse update for H_k^{-1} as follows:*

$$H_k^{-1} = \left(I - \frac{d_{k-1}q_{k-1}^T}{q_{k-1}^T d_{k-1}}\right) H_{k-1}^{-1} \left(I - \frac{q_{k-1}d_{k-1}^T}{q_{k-1}^T d_{k-1}}\right) + \frac{d_{k-1}d_{k-1}^T}{q_{k-1}^T d_{k-1}}.$$

Remark 6. *In step 2, the step size t should satisfy the Wolfe conditions:*

$$\Psi\left(y;\boldsymbol{\theta}^{(k)} + t\Delta_k\right) \leq \Psi\left(\boldsymbol{\theta}^{(k)}\right) + u_1 t\nabla\Psi^T\left(\boldsymbol{\theta}^{(k)}\right)\Delta_k,$$
$$\nabla\Psi\left(\boldsymbol{\theta}^{(k)} + t\Delta_k\right) \geq u_2\nabla\Psi^T\left(\boldsymbol{\theta}^{(k)}\right)\Delta_k,$$

where u_1 and u_2 are constants with $0 < u_1 < u_2 < 1$. The first condition requires that t sufficiently decrease the objective function. The second condition ensures that the step size is not too small. The Backtracking line search algorithm proceeds as follows (see [26]):

Algorithm 2: The Backtracking Line Search Algorithm.

Given a descent direction $\Delta(\boldsymbol{\theta})$ for Ψ at $\boldsymbol{\theta}$, $\zeta \in (0, 0.5), \kappa \in (0,1)$. $t := 1$.

while $\Psi(\boldsymbol{\theta} + t\Delta\boldsymbol{\theta}) > \Psi(\boldsymbol{\theta}) + \zeta t\nabla\Psi(\boldsymbol{\theta})^T\Delta\boldsymbol{\theta}$,
do
$t := \kappa t$.

end while

4.3. Initial Values

The initial value for $\boldsymbol{\theta}$ are taken to be

$$\mu^{(0)} = \text{median}\left(\widetilde{Y}_{il}\right)/J,$$
$$\sigma^{(0)} = 1.48 \times \text{median}\left(|\widetilde{Y}_{il} - \text{median}(\widetilde{Y}_{il})|\right)/B.$$

Another choice of the initial value for σ is:

$$\hat{\sigma}^{(0)} = \sqrt{\frac{\left(\frac{\widehat{\boldsymbol{Var}[\widetilde{Y}_{il}]}}{J} - \gamma_0^2\mu\right)}{\gamma_0^2 + \mu_0^2}}, \tag{53}$$

where $\widehat{\boldsymbol{Var}[\widetilde{Y}_{il}]}$ is an empirical estimate of the variance of $\widetilde{Y}_1$.

Bandwidth Selection: A key issue in implementing the above method of estimation is the choice of the bandwidth. We express the bandwidth in the form $h_B = c_B s_B$, where $c_B \in \{0.3, 0.4, 0.5, 0.7, 0.9\}$, and s_B is set equal to $1.48\times$ median $\left(|\widetilde{Y}_{il} - \text{median}(\widetilde{Y}_{il})|\right)/B$.

In all the tables below, we report the average (Ave), standard deviation (StD) and mean square error (MSE) to assess the performance of the proposed methods.

4.4. Analyses Without Contamination

From Tables 2–5, we let true $\mu = 2, \sigma = 1$, and take the kernel to be Gaussian kernel. In Table 2, we compare the estimates of the parameters as the dimension of the compressed data S increases. In this table, we allow S to take values in the set $\{1, 2, 5, 10\}$. Also, we let the number of groups $B = 100$, the bandwidth is chosen to be $c_B = 0.3$, and $\gamma_0 = 0.1$. In addition, in Table 2, $S^* = 1$ means that $S = 1$ with $\gamma_0 \equiv 0$.

Table 2. MHDE as the dimension S changes for compressed data $\widetilde{\boldsymbol{Y}}$ using Gaussian kernel.

	$\hat{\mu}$			$\hat{\sigma}$		
	Ave	**StD**$\times 10^3$	**MSE**$\times 10^3$	**Ave**	**StD**$\times 10^3$	**MSE**$\times 10^3$
$S^* = 1$	2.000	1.010	0.001	1.016	74.03	5.722
$S = 1$	2.000	1.014	0.001	1.018	74.22	5.844
$S = 2$	2.000	1.005	0.001	1.019	73.81	5.832
$S = 5$	2.000	0.987	0.001	1.017	74.16	5.798
$S = 10$	2.000	0.995	0.001	1.019	71.87	5.525

From Table 2 we observe that as S increases, the estimates for μ and σ remain stable. The case $S^* = 1$ is interesting, since even by storing the sum we are able to obtain point estimates which are close to the true value. In Table 3, we choose $S = 1, B = 100$ and $c_B = 0.3$ and compare the estimates as γ_0 changes from 0.01 to 1.00. We can see that as γ_0 increases, the estimate for μ remains stable, whereas the bias, standard deviation and MSE for σ increase.

Table 3. MHDE as γ_0 changes for compressed data $\widetilde{\boldsymbol{Y}}$ using Gaussian kernel.

	$\hat{\mu}$			$\hat{\sigma}$		
	Ave	**StD**$\times 10^3$	**MSE**$\times 10^3$	**Ave**	**StD**$\times 10^3$	**MSE**$\times 10^3$
$\gamma_0 = 0.00$	2.000	1.010	0.001	1.016	74.03	5.722
$\gamma_0 = 0.01$	2.000	1.017	0.001	1.015	74.83	5.814
$\gamma_0 = 0.10$	2.000	1.023	0.001	1.021	72.80	5.717
$\gamma_0 = 0.50$	2.000	1.119	0.001	1.076	72.59	11.08
$\gamma_0 = 1.00$	2.000	1.399	0.002	1.226	82.21	57.75

In Table 4, we fix $S = 1, B = 100$ and $\gamma_0 = 0.1$ and allow the bandwidth c_B to increase. Also, $c_B^* = 0.30$ means that the bandwidth is chosen as 0.30 with $\gamma_0 \equiv 0$. Notice that in this case when $c_B = 0.9$ $B^{\frac{1}{2}} c_B = 9$ while $B^{\frac{1}{2}} c_B^2 = 8.1$ which is not small as is required in assumption (**B2**). We notice again that as c_B decreases, the estimates of μ and σ are close to the true value with small MSE and StD.

Table 4. MHDE as the bandwidth c_B changes for compressed data $\widetilde{\boldsymbol{Y}}$ using Gaussian kernel.

	$\hat{\mu}$			$\hat{\sigma}$		
	Ave	**StD**$\times 10^3$	**MSE**$\times 10^3$	**Ave**	**StD**$\times 10^3$	**MSE**$\times 10^3$
$c_B^* = 0.30$	2.000	1.010	0.001	1.016	74.03	5.722
$c_B = 0.30$	2.000	1.014	0.001	1.018	74.22	5.844
$c_B = 0.40$	2.000	1.015	0.001	1.063	79.68	10.26
$c_B = 0.50$	2.000	1.014	0.001	1.108	82.33	18.33
$c_B = 0.70$	2.000	1.004	0.001	1.212	93.96	53.64
$c_B = 0.90$	2.000	1.009	0.001	1.346	110.5	132.2

In Table 5, we let $S = 1, c_B = 0.3$ and $\gamma_0 = 0.1$ and let the number of groups B increase. This table implies that as B increases, the estimate performs better in terms of bias, standard deviation and MSE.

Table 5. MHDE as B changes for compressed data $\widetilde{\boldsymbol{Y}}$ using Gaussian kernel with $\gamma_0 = 0.1$.

	$\hat{\mu}$			$\hat{\sigma}$		
	Ave	**StD**$\times 10^3$	**MSE**$\times 10^3$	**Ave**	**StD**$\times 10^3$	**MSE**$\times 10^3$
$B = 20$	2.000	2.205	0.005	1.739	378.5	688.6
$B = 50$	2.000	1.409	0.002	1.136	125.2	34.17
$B = 100$	2.000	1.010	0.001	1.016	74.03	5.722
$B = 500$	2.000	0.455	0.000	0.972	32.63	1.873

In Table 6, we set $\gamma_0 \equiv 0$ and keep other settings same as Table 5. This table implies that as B increases, the estimate performs better in terms of bias, standard deviation and MSE. Furthermore, the standard deviation and MSE are slightly smaller than the results in Table 5.

Table 6. MHDE as B changes for compressed data $\widetilde{\boldsymbol{Y}}$ using Gaussian kernel with $\gamma_0 = 0$.

	$\hat{\mu}$			$\hat{\sigma}$		
	Ave	**StD**$\times 10^3$	**MSE**$\times 10^3$	**Ave**	**StD**$\times 10^3$	**MSE**$\times 10^3$
$B = 20$	2.000	2.282	0.005	1.749	381.4	706.0
$B = 50$	2.000	1.440	0.002	1.148	125.2	37.42
$B = 100$	2.000	1.014	0.001	1.018	74.22	5.844
$B = 500$	2.000	0.465	0.000	0.973	31.33	1.692

We next move on to investigating the effect of other sensing variables. In the following table, we use Gamma model to generate the additive matrix $\boldsymbol{R}_l$. Specifically, the mean of Gamma random variable is set as $\alpha_0 \beta_0 = 1$, and the variance $var \equiv \alpha_0 \beta_0^2$ is chosen from the set $\{0, 0.01^2, 0.01, 0.25, 1.00\}$ which are also the variances in Table 3.

From Table 7, notice that using Gamma sensing variable yields similar results as Gaussian sensing variable. Our next example considers the case when the mean of the sensing variable is not equal to one and the sensing variable is taken to have a discrete distribution.Specifically, we use Bernoulli sensing variables with parameter p. Moreover, we fix $S = 1$ and let $pJ = S$. Therefore $p = 1/J$. Hence as J

increases, the variance decreases. Now notice that in this case the mean of sensing variable is p instead of 1. In addition, $\boldsymbol{E}[\widetilde{Y}_{il}] = \mu$ and $\boldsymbol{Var}[\widetilde{Y}_{il}] = \sigma^2 + \mu^2(1 - \frac{1}{J})$. Hence we set the initial value as

$$\mu^{(0)} = \text{median}\left(\widetilde{Y}_{il}\right),$$
$$\sigma^{(0)} = 1.48 \times \text{median}\left(|\widetilde{Y}_{il} - \text{median}(\widetilde{Y}_{il})|\right).$$

Additionally, we take $B = 100$, $c_B = 0.30$ and s_B to be $1.48 \times \text{median}\left(|\widetilde{Y}_{il} - \text{median}(\widetilde{Y}_{il})|\right)$.

Table 7. MHDE as variance changes for compressed data $\widetilde{\boldsymbol{Y}}$ using Gaussian kernel under Gamma sensing variable.

	$\hat{\mu}$			$\hat{\sigma}$		
	Ave	**StD**$\times 10^3$	**MSE**$\times 10^3$	**Ave**	**StD**$\times 10^3$	**MSE**$\times 10^3$
$var = 0.00$	2.000	1.010	0.001	1.016	74.03	5.722
$var = 0.01^2$	2.000	1.005	0.001	1.016	74.56	5.806
$var = 0.01$	2.000	1.006	0.001	1.018	73.70	5.762
$var = 0.25$	2.000	1.120	0.001	1.078	73.70	11.56
$var = 1.00$	2.000	1.438	0.001	1.228	81.94	58.48

Table 8 shows that MHD method also performs well with Bernoulli sensing variable, although the bias of σ, standard deviation and mean squre error for both estimates are larger than those using Gaussian sensing variable and Gamma sensing variable.

Table 8. MHDE as J changes for compressed data $\widetilde{\boldsymbol{Y}}$ using Gaussian kernel under Bernoulli sensing variable.

	$\hat{\mu}$			$\hat{\sigma}$		
	Ave	**StD**$\times 10^3$	**MSE**$\times 10^3$	**Ave**	**StD**$\times 10^3$	**MSE**$\times 10^3$
$J = 10$	2.000	104.9	11.01	1.215	97.78	55.79
$J = 100$	1.998	104.5	10.93	1.201	104.5	51.26
$J = 1000$	1.998	104.7	10.96	1.195	106.6	49.36
$J = 5000$	2.001	103.9	10.80	1.200	105.7	51.20
$J = 10000$	1.996	105.1	11.07	1.196	104.4	49.16

4.5. Robustness and Model Misspecification

In this section, we provide a numerical assessment of the robustness of the proposed methodology. To this end, let

$$f_{\alpha,\eta}(x|\boldsymbol{\theta}) = (1-\alpha)f(x|\boldsymbol{\theta}) + \alpha\eta(x),$$

where $\eta(x)$ is a contaminating component, $\alpha \in [0, 1)$. We generate the contaminated reduced data $\boldsymbol{Y}$ in the following way:

- Step 1. Generate $\boldsymbol{X}_l$, where $X_{jl} \overset{i.i.d.}{\sim} N(2, 1)$.
- Step 2. Generate $\boldsymbol{R}_l$, where $r_{ijl} \overset{i.i.d.}{\sim} N(1, \gamma_0^2)$.
- Step 3. Generate uncontaminated $\widetilde{\boldsymbol{Y}}_l$ by calculating $\widetilde{\boldsymbol{Y}}_l = \boldsymbol{R}_l\boldsymbol{X}_l$.
- Step 4. Generate contaminated $\widetilde{Y}_{il}^c$, where $\widetilde{Y}_{il}^c = \widetilde{Y}_{il} + \eta(x)$ with probability α, and $\widetilde{Y}_{il}^c = \widetilde{Y}_{il}$ with probability $1 - \alpha$.

In the above description, the contamination with outliers is within blocks. A conceptual issue that one encounters is the meaning of outliers in this setting. Specifically, a data point which is an

outlier in the original data set may not remain an outlier in the reduced data and vice-versa. Hence the concepts such as breakdown point and influence function need to be carefully studied. The tables below present one version of the robustness exhibited by the proposed method. In Tables 9 and 10, we set $J = 10^4, B = 100, S = 1, \gamma_0 = 0.1, c_B = 0.3, \eta = 1000$. In addition, $\alpha^* = 0$ means that $\alpha = 0$ with $\gamma_0 \equiv 0$.

Table 9. MHDE as α changes for contaminated data $\widetilde{Y}$ using Gaussian kernel.

	$\hat{\mu}$			$\hat{\sigma}$		
	Ave	**StD**$\times 10^3$	**MSE**$\times 10^3$	**Ave**	**StD**$\times 10^3$	**MSE**$\times 10^3$
$\alpha^* = 0.00$	2.000	1.010	0.001	1.016	74.03	5.722
$\alpha = 0.00$	2.000	1.014	0.001	1.018	74.22	5.844
$\alpha = 0.01$	2.000	1.002	0.001	1.022	74.89	6.079
$\alpha = 0.05$	2.000	1.053	0.001	1.023	77.86	6.599
$\alpha = 0.10$	2.000	1.086	0.001	1.034	79.30	7.350
$\alpha = 0.20$	2.000	1.146	0.001	1.073	93.45	14.06
$\alpha = 0.30$	2.001	7.205	0.054	1.264	688.2	542.5
$\alpha = 0.40$	2.026	21.60	1.100	3.454	1861	9480
$\alpha = 0.50$	2.051	14.00	2.600	4.809	1005	15513

Table 10. MHDE as α changes for contaminated data $\widetilde{Y}$ using Epanechnikov kernel.

	$\hat{\mu}$			$\hat{\sigma}$		
	Ave	**StD**$\times 10^3$	**MSE**$\times 10^3$	**Ave**	**StD**$\times 10^3$	**MSE**$\times 10^3$
$\alpha^* = 0.00$	2.000	0.972	0.001	1.008	73.22	5.425
$\alpha = 0.00$	2.000	1.014	0.001	1.018	74.22	5.844
$\alpha = 0.01$	2.000	0.978	0.001	1.028	107.4	12.19
$\alpha = 0.05$	2.000	1.264	0.002	1.025	108.7	12.35
$\alpha = 0.10$	2.000	1.202	0.001	1.008	114.7	13.09
$\alpha = 0.20$	2.000	1.263	0.002	1.046	129.8	18.76
$\alpha = 0.30$	2.001	5.098	0.026	1.104	557.8	318.9
$\alpha = 0.40$	2.021	21.80	0.900	3.004	1973	7870
$\alpha = 0.50$	2.051	10.21	3.000	4.893	720.4	15669

From the above Table we observe that, even under 50% contamination the estimate of the mean remains stable; however, the estimate of the variance is affected at high-levels of contamination (beyond 30%). An interesting and important issue is to investigate the role of γ_0 on the breakdown point of the estimator.

Finally, we investigate the bias in MHDE as a function of the values of the outlier. The graphs below (Figure 2) describe the changes to MHDE when outlier values (η) increase. Here we set $S = 1, B = 100, \gamma_0 = 0.1$. In addition, we let $\alpha = 0.2$, and η to take values from $\{100, 200, 300, 400, 500, 600, 700, 800, 900, 1000\}$. We can see that as η increases, both $\hat{\mu}$ and $\hat{\sigma}$ increase up to $\eta = 500$ then decrease, although $\hat{\mu}$ does not change too much. This phenomenon is because when the outlier value is small (or closer to the observations), then it may not be considered as an "outlier" by the MHD method. However, as the outlier values move "far enough" from other values, then the estimate for μ and σ remain the stable.

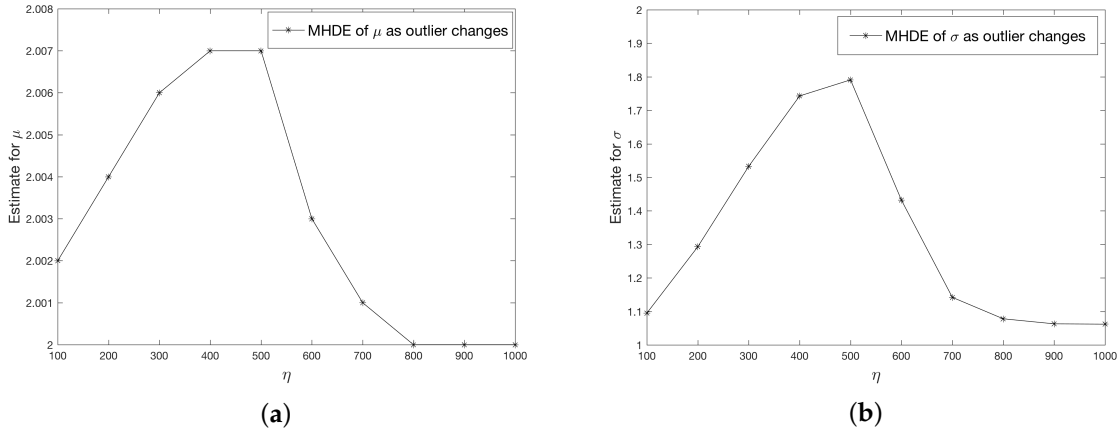


Figure 2. Comparison of estimates of μ (**a**) and σ (**b**) as outlier changes.

5. Example

In this section we describe an analysis of data from financial analytics, using the proposed methods. The data are from a bank (a cash and credit card issuer) in Taiwan and the targets of analyses were credit card holders of the bank. The research focused on the case of customers' default payments. The data set (see [27] for details) contains 180,000 observations and includes information on twenty five variables such as default payments, demographic factors, credit data, history of payment, and billing statements of credit card clients from April 2005 to September 2005. Ref. [28] study machine learning methods for evaluating the probability of default. Here, we work with the first three months of data containing 90,000 observations concerning bill payments. For our analyses we remove zero payments and negative payment from the data set and perform a logarithmic transformation of the bill payments . Since the log-transformed data was multi-modal and exhibited features of a mixture of normal distributions, we work with the log-transformed data with values in the range (6.1, 13). Next, we performed the Box-Cox transformation to the log-transformed data. This transformation identifies the best transformation that yields approximately normal distribution (which belongs to the location-scale family). Specifically, let $\boldsymbol{L}$ denote the log-transformed data in range (6.1, 13), then the data after Box-Cox transformation is given by $\boldsymbol{X} = \left(\boldsymbol{L}^2 - \mathbf{1}\right)/19.9091$. The histogram for $\boldsymbol{X}$ is given in Figure 3. The number of observations at the end of data processing was 70,000.

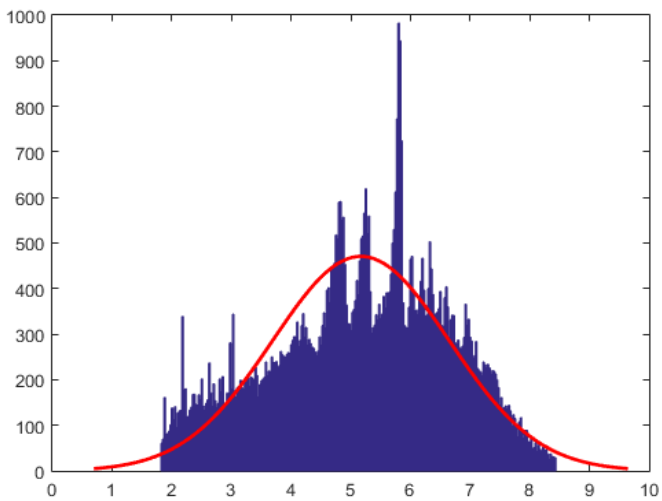


Figure 3. The histogram of credit payment data after Box-Cox transformation to Normality.

Our goal is to estimate the average bill payment during the first three months. For this, we will apply the proposed method. In this analysis, we assume that the target model for $\boldsymbol{X}$ is Gaussian and split the data, randomly, into $B = 100$ blocks yielding $J = 700$ observations per block.

In Table 11, "est" represents the estimator, "95% CI" stands for 95% confidence interval for the estimator. When analyzing the whole data and choosing bandwidth as $c_n = 0.30$, we get the MHDE of μ to be $\hat{\mu} = 5.183$ with 95% confidence interval $(5.171, 5.194)$, and the MHDE of σ as $\hat{\sigma} = 1.425$ with confidence interval $(1.418, 1.433)$.

In Table 11, we choose the bandwidth as $c_B = 0.30$. Also, $S^* = 1$ represents the case where $S = 1$ and $\gamma_0 \equiv 0$. In all other settings, we keep $\gamma_0 = 0.1$. We observe that all estimates are similar as S changes.

Table 11. MHDE from the real data analysis.

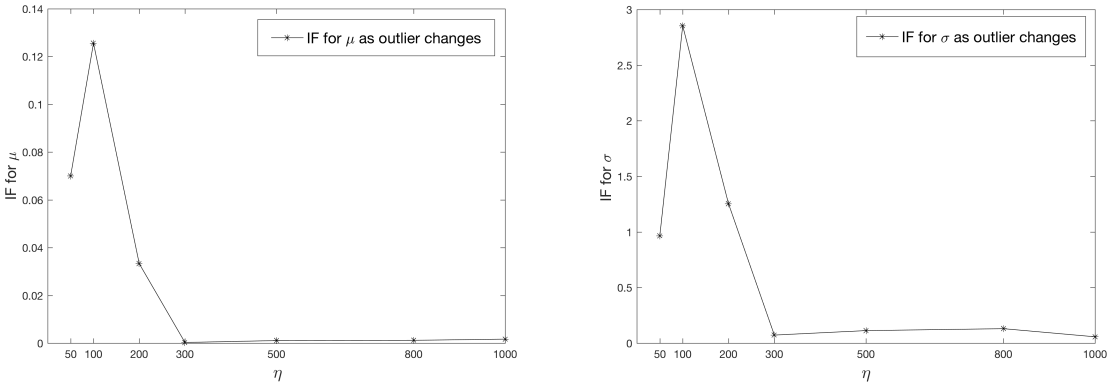
		$\hat{\mu}$	$\hat{\sigma}$
$S^* = 1$	est	5.171	1.362
	95% CI	(4.904, 5.438)	(1.158, 1.540)
$S = 1$	est	5.171	1.391
	95% CI	(4.898, 5.443)	(1.183, 1.572)
$S = 5$	est	5.172	1.359
	95% CI	(4.905, 5.438)	(1.155, 1.535)
$S = 10$	est	5.171	1.372
	95% CI	(4.902, 5.440)	(1.167, 1.551)
$S = 20$	est	5.171	1.388
	95% CI	(4.899, 5.443)	(1.180, 1.569)

Next we study the robustness of MHDE for this data by investigating the relative bias and studying the influence function. Specifically, we first reduce the dimension from $J = 700$ to $S = 1$ for each of the $B = 100$ blocks and obtain the compressed data $\widetilde{\boldsymbol{Y}}$; next, we generate the contaminated reduced data $\widetilde{Y}^c_{il}$ from step 4 in Section 4.5. Also, we set $\alpha = 0.20, \gamma_0 = 0.20$; the kernel is taken to be to be Epanechnikov density with bandwidth $c_B = 0.30$. $\eta(x)$ is assumed to takes values in $\{50, 100, 200, 300, 500, 800, 1000\}$ (note that the approximate mean of $\widetilde{\boldsymbol{Y}}$ is around 3600). Let T_{MHD} be the Hellinger distance functional. The influence function given by

$$\text{IF}(\alpha; T, \widetilde{\boldsymbol{Y}}) = \frac{T_{\text{MHD}}(\widetilde{\boldsymbol{Y}}^c) - T_{\text{MHD}}(\widetilde{\boldsymbol{Y}})}{\alpha},$$

which we use to assess the robustness. The graphs shown below (Figure 4) illustrate how the influence function changes as the outlier values increase. We observe that for both estimates ($\hat{\mu}$ and $\hat{\sigma}$), the influence function first increase and then decrease fast. From $\eta(x) = 300$, the influence functions remain stable and are close to zero, which clearly indicate that MHDE is stable.

Additional Analyses: The histogram in Figure 3 suggests that, may be a mixture of normal distributions may fit the log and Box-Cox transformed data better than the normal distribution. For this reason, we calculated the Hellinger distance between four component mixture (chosen using BIC criteria) and the normal distribution and this was determined to be 0.0237, approximately. Thus, the normal distribution (which belongs to the location-scale family) can be viewed as a misspecified target distribution; admittedly, one does lose information about the components of the mixture distribution due to model misspecification. However, since our goal was to estimate the overall mean and variance the proposed estimate seems to possess the properties described in the manuscript.



(**a**) Influence Function of $\hat{\mu}$ (**b**) Influence Function of $\hat{\sigma}$

Figure 4. Influence function of $\hat{\mu}$ (**a**) and $\hat{\sigma}$ (**b**) for MHDE.

6. Discussion and Extensions

The results in the manuscript focus on the iterated limit theory for MHDE of the compressed data obtained from a location-scale family. Two pertinent questions arise: (i) is it easy to extend this theory to MHDE of compressed data arising from *non location-scale* family of distributions? and (ii) is it possible to extend the theory from iterated limits to a double limit? Turning to (i), we note that the heuristic for considering the location-scale family comes from the fact that the first and the second moment are consistently estimable for partially observed random walks (see [29,30]). This is related to the size of J and can be of exponential order. For such large J, other moments may not be consistently estimable. Hence, the entire theory goes through as long as one is considering parametric models $f(\cdot|\boldsymbol{\theta})$, where $\boldsymbol{\theta} = \mathcal{W}(\mu, \sigma^2)$, for a known function $\mathcal{W}(\cdot, \cdot)$. The case in point is the Gamma distribution which can be re-parametrized in terms of the first two moments.

As for (ii), it is well-known that existence and equality of iterated limits for real sequences does not imply the existence of the double limit unless additional uniformity of convergence holds (see [31] for instance). Extension of this notion for distributional convergence requires additional assumptions and are investigated in a different manuscript wherein more general divergences are also considered.

7. Concluding Remarks

In this paper we proposed the Hellinger distance-based method to obtain robust estimates for mean and variance in a location-scale model using compressed data. Our extensive theoretical investigations and simulations show the usefulness of the methodology and hence can be applied in a variety of scientific settings. Several theoretical and practical questions concerning robustness in a big data setting arise. For instance, the effect of the variability in the $\boldsymbol{R}$ matrix and its effect on outliers are important issues that need further investigation. Furthermore, statistical properties such as uniform consistency and uniform asymptotic normality under different choices for the distribution of $\boldsymbol{R}$ would be useful. These are under investigation by the authors.

Author Contributions: The problem was conceived by E.A., A.N.V. and G.D. L.L. is a student of A.N.V., and worked on theoretical and simulation details with inputs from all members at different stages.

Funding: The authors thank George Mason University Libraries for support with the article processing fees; Ahmed's research is supported by a grant from NSERC.

Acknowledgments: The authors thank the anonymous reviewers for a careful reading of the manuscript and several useful suggestions that improved the readability of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MHDE	Minimum Hellinger Distance Estimator
MHD	Minimum Hellinger Distance
i.i.d.	independent and identically distributed
MLE	Maximum Likelihood Estimator
CI	Confidence Interval
IF	Influence Function
RHS	Right Hand Side
LHS	Left Hand Side
BFGS	Broyden-Fletcher-Goldfarb-Shanno
var	Variance
StD	Standard Deviation
MSE	Mean Square Error

References

1. Beran, R. Minimum Hellinger distance estimates for parametric models. *Ann. Stat.* **1977**, *5*, 445–463. [CrossRef]
2. Lindsay, B.G. Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *Ann. Stat.* **1994**, *22*, 1081–1114. [CrossRef]
3. Fisher, R.A. Two new properties of mathematical likelihood. *Proc. R. Soc. Lond. Ser. A* **1934**, *144*, 285–307. [CrossRef]
4. Pitman, E.J.G. The estimation of the location and scale parameters of a continuous population of any given form. *Biometrika* **1939**, *30*, 391–421. [CrossRef]
5. Gupta, A.; Székely, G. On location and scale maximum likelihood estimators. *Proc. Am. Math. Soc.* **1994**, *120*, 585–589. [CrossRef]
6. Duerinckx, M.; Ley, C.; Swan, Y. Maximum likelihood characterization of distributions. *Bernoulli* **2014**, *20*, 775–802. [CrossRef]
7. Teicher, H. Maximum likelihood characterization of distributions. *Ann. Math. Stat.* **1961**, *32*, 1214–1222. [CrossRef]
8. Thanei, G.A.; Heinze, C.; Meinshausen, N. Random projections for large-scale regression. In *Big and Complex Data Analysis*; Springer: Berlin, Germany, 2017; pp. 51–68.
9. Slawski, M. Compressed least squares regression revisited. In *Artificial Intelligence and Statistics*; Addison-Wesley: Boston, MA, USA, 2017; pp. 1207–1215.
10. Slawski, M. On principal components regression, random projections, and column subsampling. *Electron. J. Stat.* **2018**, *12*, 3673–3712. [CrossRef]
11. Raskutti, G.; Mahoney, M.W. A statistical perspective on randomized sketching for ordinary least-squares. *J. Mach. Learn. Res.* **2016**, *17*, 7508–7538.
12. Ahfock, D.; Astle, W.J.; Richardson, S. Statistical properties of sketching algorithms. *arXiv* **2017**, arXiv:1706.03665.
13. Vidyashankar, A.; Hanlon, B.; Lei, L.; Doyle, L. Anonymized Data: Trade off between Efficiency and Privacy. **2018**, preprint.
14. Woodward, W.A.; Whitney, P.; Eslinger, P.W. Minimum Hellinger distance estimation of mixture proportions. *J. Stat. Plan. Inference* **1995**, *48*, 303–319. [CrossRef]
15. Basu, A.; Harris, I.R.; Basu, S. Minimum distance estimation: The approach using density-based distances. In *Robust Inference, Handbook of Statistics*; Elsevier: Amsterdam, The Netherlands, 1997; Volume 15, pp. 21–48.
16. Hooker, G.; Vidyashankar, A.N. Bayesian model robustness via disparities. *Test* **2014**, *23*, 556–584. [CrossRef]
17. Sriram, T.; Vidyashankar, A. Minimum Hellinger distance estimation for supercritical Galton–Watson processes. *Stat. Probab. Lett.* **2000**, *50*, 331–342. [CrossRef]
18. Simpson, D.G. Minimum Hellinger distance estimation for the analysis of count data. *J. Am. Stat. Assoc.* **1987**, *82*, 802–807. [CrossRef]

19. Simpson, D.G. Hellinger deviance tests: Efficiency, breakdown points, and examples. *J. Am. Stat. Assoc.* **1989**, *84*, 107–113. [CrossRef]
20. Cheng, A.; Vidyashankar, A.N. Minimum Hellinger distance estimation for randomized play the winner design. *J. Stat. Plan. Inference* **2006**, *136*, 1875–1910. [CrossRef]
21. Basu, A.; Shioya, H.; Park, C. *Statistical Inference: The Minimum Distance Approach*; Chapman and Hall/CRC: London, UK, 2011.
22. Bhandari, S.K.; Basu, A.; Sarkar, S. Robust inference in parametric models using the family of generalized negative exponential dispatches. *Aust. N. Z. J. Stat.* **2006**, *48*, 95–114. [CrossRef]
23. Ghosh, A.; Harris, I.R.; Maji, A.; Basu, A.; Pardo, L. A generalized divergence for statistical inference. *Bernoulli* **2017**, *23*, 2746–2783. [CrossRef]
24. Tamura, R.N.; Boos, D.D. Minimum Hellinger distance estimation for multivariate location and covariance. *J. Am. Stat. Assoc.* **1986**, *81*, 223–229. [CrossRef]
25. Li, P. Estimators and tail bounds for dimension reduction in l α (0< $\alpha \leq$ 2) using stable random projections. In Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms, San Francisco, CA, USA, 20–22 January 2008; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2008; pp. 10–19.
26. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
27. Lichman, M. UCI Machine Learning Repository. Available online: https://archive.ics.uci.edu/ml/index.php (accessed on 29 March 2019).
28. Yeh, I.C.; Lien, C.H. The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Syst. Appl.* **2009**, *36*, 2473–2480. [CrossRef]
29. Guttorp, P.; Lockhart, R.A. Estimation in sparsely sampled random walks. *Stoch. Process. Appl.* **1989**, *31*, 315–320, doi:10.1016/0304-4149(89)90095-1. [CrossRef]
30. Guttorp, P.; Siegel, A.F. Consistent estimation in partially observed random walks. *Ann. Stat.* **1985**, *13*, 958–969. doi:10.1214/aos/1176349649. [CrossRef]
31. Apostol, T.M. *Mathematical Analysis*; Addison Wesley Publishing Company: Boston, MA, USA, 1974.

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).