Ann. Inst. Statist. Math.
Vol. 50, No. 3, 503–521 (1998)

MINIMUM DISPARITY ESTIMATION IN LINEAR REGRESSION MODELS: DISTRIBUTION AND EFFICIENCY

RO JIN PAK[1] AND AYANENDRANATH BASU[2]

[1] *Department of Statistics, Taejon University, Taejon 300-716, Korea*
[2] *Applied Statistics Unit, Indian Statistical Institute, 203 B. T. Road, Calcutta 700 035, India*

(Received January 13, 1997, revised July 22, 1997)

Abstract. This paper deals with the minimum disparity estimation in linear regression models. The estimators are defined as statistical quantities which minimize the blended weight Hellinger distance between a weighted kernel density estimator of errors and a smoothed model density of errors. It is shown that the estimators of the regression parameters are asymptotic normally distributed and efficient at the model if the weights of the density estimators are appropriately chosen.

Key words and phrases: Asymptotic efficiency, blended weight Hellinger distance, kernel density estimator, linear regression model.

1. Introduction

Beran (1977) introduced a robust estimation method, called minimum Hellinger distance estimation, which defines an estimator as a statistic minimizing the Hellinger distance between a parametric model density and a non-parametric density estimator. Later, Lindsay (1994) developed minimum disparity estimation, a large subclass of density based minimum distance estimation, of which minimum Hellinger distance estimation is a part. Basu and Lindsay (1994) show that the minimum disparity estimator (MDE) has attractive efficiency and robustness properties among other robust estimators. Above all, Basu and Lindsay (1994) apply the same smoothness to the model and the data, while the conventional methods do not smooth the model (Beran (1977), Simpson (1987) and Tamura and Boos (1986)). As a result, consistency and rate of convergence results for the nonparametric density estimators are no longer required. Also, contrary to one's intuition, it may be possible to choose the kernel, called transparent kernel so that smoothing the model and the data with a transparent kernel does not lead to loss of any information as far as the estimation problem is concerned.

However, since Beran (1977) introduced the minimum Hellinger distance estimation method, no attempts to apply it to regression models were made. In this paper the technique of minimum disparity estimation is applied to the case of linear regression models. The estimators in this case inherit robustness of the minimum disparity estimators which Basu and Lindsay (1994) have investigated. The estimators of regression coefficients, derived by the minimum disparity estimation method, are robust and asymptotically efficient, while some of well-known robust methods produce robust but inefficient estimators (Rousseeuw and Leroy (1987)). Asymptotic distribution and asymptotic efficiency of the minimum disparity estimators of linear regression coefficients are studied. Robustness of estimators and a small sample example were provided by Pak (1995). The minimum Hellinger distance estimation of the simple linear regression coefficients and the scale parameter was studied by Pak (1996).

The rest of the paper is organized as follows: Section 2 gives a review of minimum disparity estimation and Section 3 provides the definition of the minimum disparity estimators of linear regression coefficients. The theoretical framework of the minimum disparity estimators is established in Sections 4 and 5.

2. Review of minimum disparity estimation

Minimum disparity estimation (Lindsay (1994), Basu and Lindsay (1994)) is an efficient and robust estimation method in parametric models. In this section we briefly review minimum disparity estimation when i.i.d. observations are available from a continuous parametric model.

Let $m_\beta(x)$ represent the density of a parametric family of models, completely known except for the parameter vector β. Given a sample of n i.i.d. observations, construct a nonparametric density estimator from the data, say $f^*(x)$. It is usually done using kernel density estimation methods, as

$$f^*(x) = \int k(x; y, h) dF_n(y),$$

where F_n is the empirical distribution function and k is a smooth family of kernel functions like the normal densities with mean y and standard deviation h. The parameter h controls the smoothness of the resulting density. Let $M_\beta(x)$ be the cumulative distribution function (c.d.f.) of the model. Next applying the same smoothness to the model, we get

$$m_\beta^*(x) = \int k(x; y, h) dM_\beta(y).$$

Now we can construct a density based distance between $f^*(x)$ and $m_\beta^*(x)$ like the squared Hellinger distance

$$\int \left[\sqrt{f^*(x)} - \sqrt{m_\beta^*(x)}\right]^2 dx,$$

which may be minimized to obtain the minimum Hellinger distance estimator. The obvious analog of the maximum likelihood estimator in this case is the 'MLE*', which is the value of β that minimizes the distance

$$\int f^*(x) \log[f^*(x)/m_\beta^*(x)] dx.$$

This approach has several advantages over the conventional methods which do not smooth the model before the disparity is constructed, and they are discussed in Basu and Lindsay (1994). In particular, they do not require consistency or rate of convergence results for the nonparametric density estimators. Also, the MDEs are consistent and asymptotically normally distributed for any fixed bandwidth h. We do not have to let h go to zero at an appropriate rate as n tends to infinity.

Define the Pearson residual, a standardized version of the residual as

$$\delta^*(x) = \frac{f^*(x) - m^*_\beta(x)}{m^*_\beta(x)} dx.$$

For an arbitrary real valued twice differentiable convex function G with $G(0) = 0$ define a disparity measure ρ_G between $f^*(x)$ and $m_\beta(x)$ as

$$\rho(f^*, m^*_\beta) = \int G(\delta^*(x)) m^*_\beta(x) dx.$$

If G is strictly convex then the MDE, the value of β which minimizes ρ_G, is a Fisher consistent estimator. $G(\delta^*) = (\sqrt{\delta^* + 1} - 1)^2$ generates the squared Hellinger distance, whereas $G(\delta^*) = (\delta^* + 1)\log(\delta^* + 1)$ generates the likelihood disparity. Pearson's and Neyman's chi-squares and the power weighted divergence measures of Cressie and Read (1984) are other prominent members of the class of disparities.

Let ∇ represent the derivatives with respect to β. Under differentiability if the model, minimization of the disparity measure ρ over β corresponds to solving a set of estimating equations of the form:

$$-\nabla\rho = \int A(\delta^*(x)) \nabla m^*_\beta(x) dx = 0,$$

for $A(\delta^*) = (1 + \delta^*(x))G'(\delta^*(x)) - G(\delta^*)$. The function A is called the residual adjustment function (RAF) of the disparity. It plays an important role in the derivation of the properties of the MDE and helps to describe the finite sample efficiency and robustness properties of the estimator in terms of its treatment of the Pearson residuals. In this respect it is almost exactly like the ψ function of the M-estimation approach.

The curvature parameter $A_2 = A''(0)$, associate with each RAF $A(\cdot)$ plays a very important role in determining the trade-off between the robustness and second-order efficiency properties of the estimator. $A_2 = 0$ implies second-order efficiency in the sense of Rao (1961), while departure from zero corresponds to a loss in that. It has been shown in Lindsay (1994) and Basu and Lindsay (1994) that a large negative value of A_2 corresponds to certain robustness properties of the estimator.

3. Definition of minimum disparity estimators in linear regression

Consider a multiple linear regression model; the model is specified by a linear equation of the observed value of a response variable and the observed values of predictors. Given the l-th observations of the predictors, x_{il}, $i = 0, \ldots, p$, the l-th observed value of a response variable is

$$y_l = b_0 x_{0l} + b_1 x_{1l} + b_2 x_{2l} + \cdots + b_p x_{pl} + e_l, \qquad x_{0l} = 1 \quad \text{for} \quad l = 1, \ldots, n,$$

where e_l has a sufficiently smooth symmetric density of the exponential family with $E[e_l] = 0$, $\mathrm{Var}[e_l] = \sigma^2$(known) and $\mathrm{Cov}(e_l, e_m) = 0$, $l \neq m$. We can express the model in matrix form as follows;

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{b}^T + \boldsymbol{e}, \qquad E[\boldsymbol{e}] = \boldsymbol{0}, \qquad \mathrm{Var}[\boldsymbol{e}] = \sigma^2 \boldsymbol{I}.$$

$\boldsymbol{Y}$ and $\boldsymbol{e}$ are $n \times 1$ vectors of y_l's and e_l's, respectively. $\boldsymbol{X}$ and $\boldsymbol{b}$ are as follows;

$$\text{(3.1)} \qquad \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{bmatrix}, \qquad \boldsymbol{b}^T = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix}.$$

We will denote the vector of the true values of the parameters ($\boldsymbol{b}_o$) and the vector of the estimators of the parameters ($\boldsymbol{b}_n$) such as

$$\boldsymbol{b}_o = (b_{0o}, b_{1o}, \ldots, b_{po}), \qquad \boldsymbol{b}_n = (b_{0n}, b_{1n}, \ldots, b_{pn}).$$

Define the standardized errors such that for $l = 1, \ldots, n$;

$$z_l = (y_l - b_0 - b_1 x_{1l} - b_2 x_{2l} - \cdots - b_p x_{pl})/\sigma,$$

which are symbolic quantities of unknown b_i's but z_l are i.i.d. with a common density at the model (that is, given the true values of the parameters).

Step 1: *Define the density estimator for the model density*
For each i, define a density estimator as a weighted average of kernels;

$$f_i^*(t, \boldsymbol{b}) = \frac{1}{\sum_{l=1}^n x_{il}} \sum_{l=1}^n x_{il} k(z_l; t, h),$$

where $k(z_l; t, h)$ is a smoothed kernel function of z_l with window width h. If we follow Silverman (1986), $k(z_l; t, h) = \frac{1}{h} K((t - z_l)/h)$, where $K(\cdot)$ is a kernel function and h is window width. Since the estimators of the regression coefficients are location invariant, it would be all right without loss of generality, even if we let x_{il}'s be positive.

Step 2: Define the smoothed model density
For each i, the smoothed model density is defined as

$$g_i^*(t) = E_{z_1,\ldots,z_n}[f_i^*(t,\boldsymbol{b})]$$

at the model. In fact, $g_i^*(t)$'s are identical for all i, so let's call them just $g^*(t)$. Also, it is a completely known function of t.

Step 3: Define the Pearson residual
For each i, the Pearson residual is defined as

$$\delta_i^*(t,\boldsymbol{b}) = \frac{f_i^*(t,\boldsymbol{b}) - g_i^*(t)}{g_i^*(t)}.$$

If there exists at least one outlier among observations, then the discrepancy between a density estimator and a smoothed model density at a particular value of t increases, so the Pearson residual does.

Now, let ∇_i, ∇_{ij} and ∇_{ijk} represent the first partial derivative with respect to b_i, the second partial derivative with respect to b_i and b_j and the third partial derivative with respect to b_i, b_j and b_k. Just for convenience, we will drop $\boldsymbol{b}$ in $f_i^*(t,\boldsymbol{b})$ and $\delta_i^*(t,\boldsymbol{b})$, and write $f_i^*(t)$ and $\delta_i^*(t)$ for them, respectively.

Step 4: Define the minimum disparity estimator

DEFINITION 3.1. Suppose

$$y_l = \boldsymbol{x}_l\boldsymbol{b}^T + \boldsymbol{e}_l, \qquad 1 \le l \le n,$$

where $\boldsymbol{x}_l$, a $1 \times (p+1)$ vector, is the l-th row of the design matrix $\boldsymbol{X}$, $\boldsymbol{b}$ is a $1 \times (p+1)$ vector of the parameters and $\boldsymbol{e}_l$ are i.i.d. random errors with mean 0 and variance σ^2 (known). Then for some appropriate disparity ρ, the minimum disparity estimator $\boldsymbol{b}_n$ is defined as the statistic minimizing

$$S(\boldsymbol{b}) = \sum_{i=0}^{p} \rho_i(\boldsymbol{b}) = \sum_{i=0}^{p} \rho(f_i^*(t), g_i^*(t))$$

as a function of $\boldsymbol{b}$. In particular, for the blended weight Hellinger distance (BWHD) (Lindsay (1994); Basu and Lindsay (1994)), the minimum disparity estimator of $\boldsymbol{b}$ will minimize

$$S(\boldsymbol{b}) = \sum_{i=0}^{p} BWHD(f_i^*(t), g^*(t)) = \sum_{i=0}^{p} \int \left(\frac{f_i^*(t) - g_i^*(t)}{\alpha\sqrt{f_i^*(t)} + \bar{\alpha}\sqrt{g_i^*(t)}} \right)^2 dt, \qquad \alpha \in (0,1],$$

where $\bar{\alpha} = 1 - \alpha$.

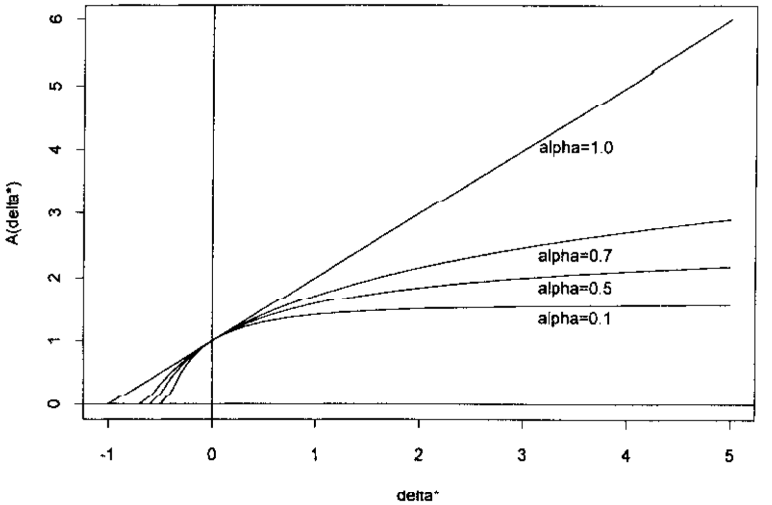

Fig. 1. The residual adjustment functions.

Under differentiability of $f_i^*(t, \boldsymbol{b})$ with respect to the parameter set of interest, the minimum disparity estimators are the solutions to the simultaneous equations of

$$
\begin{aligned}
(3.2) \qquad \nabla_j S(\boldsymbol{b}) &= \sum_{i=1}^{p} \nabla_j \rho_i(\boldsymbol{b}) \\
&= \sum_{i=0}^{p} \nabla_j BWHD(f_i^*(t), g^*(t)) \\
&= \sum_{i=0}^{p} \int A(\delta_i^*(t)) \nabla_j f_i^*(t) dt = 0 \qquad \text{for} \quad j = 0, 1, \ldots, p,
\end{aligned}
$$

where the function $A(\cdot)$ is called the residual adjustment function (RAF) (Lindsay (1994), Basu and Lindsay (1994)). Through the differentiation of $BWHD(f_i^*(t), g^*(t))$ w.r.t. the regression coefficients for the $BWHD$ family of disparities, the residual adjustment functions have the form

$$
A(\delta_i^*(t)) = \delta_i^*(t) \left(\alpha \sqrt{\delta_i^*(t) + 1} + \bar{\alpha} \right)^{-2} - \frac{\alpha}{2} \frac{\delta_i^*(t)}{\sqrt{\delta_i^*(t) + 1}} \left(\alpha \sqrt{\delta_i^*(t) + 1} + \bar{\alpha} \right)^{-3}.
$$

DEFINITION 3.2. The Residual Adjustment Function (RAF), $A(\delta_i^*(t))$, will be called regular if it is twice differentiable, and $A'(\delta_i^*(t))$ and $A''(\delta_i^*(t))(1 + \delta_i^*(t))$ are bounded on $[-1, \infty)$.

Remark 3.1. In Fig. 1 we have plotted RAFs for various alphas. Some useful facts about $A(\delta_i^*(t))$ should be mentioned. In practice, we redefine the function A without changing its estimating properties so that $A(0) = 1$, $A'(0) = 1$ and

$A(\delta^*) = 0$ if $A(\delta^*) < 0$. The RAFs coincide at $(0, 1)$, and the tangent line of the RAFs at $(0, 1)$ is the RAF with $\alpha = 1.0$. Also, we have $|A(\delta^*)| \leq |\delta^*|$. Since $f_i^*(t) \to g^*(t)$ by SSLN as $n \to \infty$, we have $\delta_i^*(t) \to 0$. Therefore it follows that $A(\delta_i^*(t)) \to 1$, $A'(\delta_i^*(t)) = \nabla_t A(\delta_i^*(t)) \to 1$ as $n \to \infty$.

4. Preliminary results

In this section we will briefly present some preliminary results about $f_i^*(t)$ and $g^*(t)$. These results will be very useful to prove consistency, asymptotic normality and asymptotic efficiency of the minimum disparity estimators in a linear regression model. In the following discussion it will be assumed that the z_l's are defined at the true values of the parameters, so that they are i.i.d. with a common density, say $g(z)$. For notational convenience when we discuss about common facts for all z_l's, we will drop a subscript 'l'.

LEMMA 4.1. *Provided it exists,* $\mathrm{Var}[f_i^*(t)] = (\sum_{l=1}^n x_{il}^2/(\sum_{l=1}^n x_{il})^2)\lambda(t)$, *where* $\lambda(t)$ *is given by*

$$\lambda(t) = E_z[(k(z; t, h) - g^*(t))^2].$$

PROOF. Note that $E_z[k(z; t, h)] = g^*(t)$. Since $f_i^*(t) = \sum_{l=1}^n x_{il} k(z_l; t, h)/\sum_{l=1}^n x_{il}$ has the form of a weighted sample mean, the result follows.

We want that the kernel function $k(z; t, h)$ is bounded. From now on, let

$$k(z; t, h) \leq M(h)$$

with $M(h) < \infty$, where $M(h)$ depends on h but not on z. Then

$$\begin{aligned}\lambda(t) &\leq \int [k(z; t, h)]^2 g(z) dy \\ &\leq M(h) \int k(z; t, h) g(z) dz \\ &= M(h) g^*(t).\end{aligned}$$

LEMMA 4.2. $n^{1/4}(f_i^{*1/2}(t) - g^{*1/2}(t)) \to 0$ *with probability* 1 *if* $\lambda(t) < \infty$.

PROOF. Recall that $E[k(z; t, h)] = g^*(t)$ and $\mathrm{Var}[k(z; t, h)] = \lambda(t)$. We can represent $f_i^*(t) - g^*(t)$ as $f_i^*(t) - g^*(t) = \sum_{l=1}^n v_{il}/n$, where

$$v_{il} = \frac{x_{il}}{\sum_{l=1}^n x_{il}/n}[k(z_l; t, h) - g^*(t)].$$

The v_{il}'s are independently distributed for all l with $E[v_{il}] = 0$ and $\mathrm{Var}[v_{il}] < \infty$. Hence, we have by Theorem 2 (Marcinkiewicz-Zygmund) of Chow and Teicher ((1988), p. 125) that

$$n^{1/4}(f_i^*(t) - g^*(t)) = n^{-3/4} \sum_{l=1}^n v_{il} \to 0 \tag{4.1}$$

almost certainly. By a Taylor series expansion of a square root we get

$$n^{1/4}(f_i^{*1/2}(t) - g^{*1/2}(t))$$
$$= n^{1/4}(f_i^*(t) - g^*(t))\frac{1}{2g^*(t)} + o(n^{1/4}|f_i^*(t) - g^*(t)|),$$

so that by using (4.1) we have

$$n^{1/4}(f_i^{*1/2}(t) - g^{*1/2}(t)) \to 0$$

with probability 1.

Remark 4.1. We would like to assume that for $k = 1, 2$, $\sum_{l=1}^n x_{il}^k/n$ and $\sum_{l=1}^n x_{il}^k x_{jl}^k/n$ converge to finite quantities, denoted by $\bar{x_i}^k$ and $\bar{x_{ij}}^k$, in the limit. In other word, for a large n, these are bounded by some finite positive quantities independent of n. These assumptions ensure that the realized values of independent variables should be within certain range. The regression analysis with infinite values on the independent variables has no sense.

LEMMA 4.3. *Define* $\gamma_i^*(t) = (f_i^{*1/2}(t) - g^{*1/2}(t))/g^{*1/2}(t)$. *For any* $k \in [0, 2]$ *and for* $i = 0, 1, \ldots, p$, *we have*

$$\limsup_n E[|n^{1/2}\gamma_i^{*2}(t)|^k] \le \limsup_n n^{k/2}(E[|\delta_i^*(t)|^k]) \tag{4.2}$$
$$\le U_i^{k/2} \limsup_n \left(\frac{\lambda^{1/2}(t)}{g^*(t)}\right)^k,$$

where $U_i = \limsup_n(\sum_{l=1}^n x_{il}^2/n)/(\sum_{l=1}^n x_{il}/n)^2$.

Also,

$$E[|\delta_i^*(t)|] \le \frac{\lambda^{1/2}(t)}{g^*(t)}. \tag{4.3}$$

PROOF. In the following proof, the first inequality holds by the facts that for $a, b \ge 0$, $(a^{1/2} - b^{1/2})^2 \le |a - b|$ and the second inequality holds by using Liapounov's inequality. The forth equality holds due to the independency of $k(z_l; t, h)$'s. That is,

$$E[|n^{1/2}\gamma_i^{*2}(t)|^k] = E\left[n^{1/2}\left|\frac{f_i^{*1/2}(t)}{g^{*1/2}(t)} - 1\right|^{2k}\right]$$
$$\le n^{k/2}E\left[\left|\frac{f_i^*(t)}{g^*(t)} - 1\right|^k\right]$$
$$= n^{k/2}E[|\delta_i^*(t)|^k]$$

$$
\begin{aligned}
&\le n^{k/2}(E[|\delta_i^*(t)|^2])^{k/2}\\
&- \frac{n^{k/2}}{[g^*(t)]^k}\left(E\left[\left|\frac{1}{\sum_{l=1}^n x_{il}}\sum_{l=1}^n x_{il}(k(z_l;t,h)-g^*(t))\right|^2\right]\right)^{k/2}\\
&= \frac{n^{k/2}}{[g^*(t)]^k}\left(\frac{1}{(\sum_{l=1}^n x_{il})^2}\sum_{l=1}^n E[x_{il}^2(k(z_l;t,h)-g^*(t))^2]\right)^{k/2}\\
&= \frac{n^{k/2}}{[g^*(t)]^k}\left(\frac{1}{(\sum_{l=1}^n x_{il})^2}\operatorname{Var}\left[\sum_{l=1}^n x_{il}k(z_l;t,h)\right]\right)^{k/2}\\
&= \left(\frac{\sum_{l=1}^n x_{il}^2/n}{(\sum_{l=1}^n x_{il}/n)^2}\right)^{k/2}\left(\frac{\lambda(t)}{[g^*(t)]^2}\right)^{k/2}.
\end{aligned}
$$

Also,

$$
\begin{aligned}
E[|\delta_i^*(t)|] &= \frac{1}{g^*(t)}\left(E\left[\left|\frac{1}{\sum_{l=1}^n x_{il}}\sum_{l=1}^n x_{il}(k(z_l;t,h)-g^*(t))\right|\right]\right)\\
&\le \frac{1}{g^*(t)}\frac{1}{\sum_{l=1}^n x_{il}}\sum_{l=1}^n x_{il}(E[|k(z_l;t,h)-g^*(t)|^2])^{1/2}.
\end{aligned}
$$

LEMMA 4.4. $\lim_{n\to\infty} E[|n^{1/2}\gamma_i^{*2}(t)|^k] = 0$ *for* $k \in [0,2]$.

PROOF. We know that for any $k \in [0,2]$, $\limsup_n E[|n^{1/2}\gamma_i^{*2}(t)|^k]$ is bounded by Lemma 4.3. By Lemma 4.2, we have

$$
n^{1/2}\gamma_i^{*2} = n^{1/2}\left(\frac{f_i^{*1/2}(t)-g^{*1/2}(t)}{g^{*1/2}(t)}\right)^2 = \left(n^{1/4}\left(\frac{f_i^{*1/2}(t)-g^{*1/2}(t)}{g^{*1/2}(t)}\right)\right)^2 \to 0
$$

in probability as $n \to \infty$. Therefore, the result follows by uniform integrability.

5. Asymptotic properties of the estimators

In order to prove the main asymptotic results, we will first concentrate on the quantity in (3.2):

$$
\nabla_j \rho_i(\boldsymbol{b}) = \int A(\delta_i^*(t))\nabla_j f_i^*(t)dt, \qquad i = 0,\dots,p;\ j = 0,\dots,p.
$$

LEMMA 5.1. (Lindsay (1994)) *Suppose that* $A(\delta_i^*(t))$ *is a regular RAF, then there exists a constant* $B \ge 0$ *such that for all positive* c *and* d

$$
|A(c^2-1)+A(d^2-1)-(c^2-d^2)A'(d^2-1)| \le B(c-d)^2. \tag{5.1}
$$

PROOF. See Lindsay (1994).

LEMMA 5.2. *Suppose that* $\int g^*(t)(\nabla_t \log g^*(t))^2 dt$, $E_z[\int(\nabla_t k(z;t,h))^2 dt]$ *and* $\int(\nabla_t g^*(t))^2 dt$ *exist. For a regular RAF we have*

$$\sqrt{n}\left[\int A(\delta_i^*(t))\nabla_j f_i^*(t)dt - \int \delta_i^*(t)E_z[\nabla_j f_i^*(t)]dt\right] \to 0, \qquad as \quad n \to \infty$$

with probability 1.

PROOF. Consider the following inequality:

$$\begin{aligned}(5.2) \qquad \sqrt{n}&\left|\int A(\delta_i^*(t))\nabla_j f_i^*(t)dt - \int \delta_i^*(t)E_z[\nabla_j f_i^*(t)]dt\right| \\ &\le \sqrt{n}\left|\int A(\delta_i^*(t))\nabla_j f_i^*(t)dt - \int A(\delta_i^*(t))E_z[\nabla_j f_i^*(t)]dt\right| \\ &\quad + \sqrt{n}\left|\int A(\delta_i^*(t))E_z[\nabla_j f_i^*(t)]dt \quad \int \delta_i^*(t)E_z[\nabla_j f_i^*(t)]dt\right|.\end{aligned}$$

where

$$\begin{aligned}E_z[\nabla_j f_i^*(t)] &= E_z\left[\frac{1}{\sum_{l=1}^n x_{il}}\sum_{l-1}^n \frac{x_{il}x_{jl}}{\sigma}\nabla_t k(z_l;t,h)\right] \\ &= \frac{1}{\sigma\sum_{l=1}^n x_{il}}\sum_{l-1}^n x_{il}x_{jl}\nabla_t E_z[k(z;t,h)] \\ &= \frac{\sum_{l=1}^n x_{il}x_{jl}}{\sigma\sum_{l=1}^n x_{il}}\nabla_t g^*(t).\end{aligned}$$

Investigate the first term on the right-hand side of (5.2)

Since $A(\delta_i^*(t)) \to 1$ as $n \to \infty$, if n is large $A(\delta_i^*(t))$ is bounded by, say A, with probability tending to 1. We have

$$\begin{aligned}\sqrt{n}&\left|\int A(\delta_i^*(t))\nabla_j f_i^*(t)dt - \int A(\delta_i^*(t))E_z[\nabla_j f_i^*(t)]dt\right| \\ &\le \sqrt{n}A\int |\nabla_j f_i^*(t) - E_z[\nabla_j f_i^*(t)]|dt.\end{aligned}$$

Consider the limsup of expectation of the term on the right-hand side, and recall that $k(z_l;t,h)$'s are independent at the model. That is,

$$\begin{aligned}&\limsup_n A\int \frac{1}{\sigma\sum_{l=1}^n x_{il}/n}E_z\left[\left|\frac{1}{\sqrt{n}}\sum_{l=1}^n x_{il}x_{jl}(\nabla_t k(z_l;t,h) - \nabla_t g^*(t))\right|\right]dt \\ &\le A\frac{\left(\bar{x_{ij}^2}\right)^{1/2}}{\sigma\bar{x}_i}\int \left(E_z[(\nabla_t k(z;t,h) - \nabla_t g^*(t))^2]\right)^{1/2}dt\end{aligned}$$

$$\leq A\frac{(\bar{x_{ij}^2})^{1/2}}{\sigma\bar{x}_i}\left(\int E_z[(\nabla_t k(z;t,h)-\nabla_t g^*(t))^2]dt\right)^{1/2}$$
$$\leq A\frac{(\bar{x_{ij}^2})^{1/2}}{\sigma\bar{x}_i}\left(E_z\left[\int(\nabla_t k(z;t,h))^2dt\right]-E_z\left[\int(\nabla_t g^*(t))^2dt\right]\right)^{1/2}$$
$$\leq A\frac{(\bar{x_{ij}^2})^{1/2}}{\sigma\bar{x}_i}\left(E_z\left[\int(\nabla_t k(z;t,h))^2dt\right]\right)^{1/2}$$

is bounded due to the assumptions. Hence by the dominated convergence theorem the first term on the right-hand side of (5.2) converges to 0 in probability.

Investigate the second term on the right-hand side of (5.2)

By replacing the c by $(f_i^*(t)/g^*(t))^{1/2}$ and d by 1 in (5.1), we have that

$$\sqrt{n}|A(\delta_i^*(t))-A(0)-\delta_i^*(t)A'(0)|<\sqrt{n}B\gamma_i^{*2}(t),$$

of which the expectation is bounded and it goes to zero as we have shown in Lemma 4.3. We have $\int E_z[\nabla_j f_i^*(t)]dt=E_z[\nabla_j\int f_i^*(t)dt]=0$. Recall that we have $\lambda(t)\leq M(h)g^*(t)$ and that $\limsup_n\sum x_{il}x_{jl}/\sum x_{il}=\limsup_n\sum(x_{il}x_{jl}/n)/(\sum x_{il}/n)$ is bounded, say by K, by Remark 4.1.

Hence we have

$$\sqrt{n}E_z\left[\left|\int A(\delta_i^*(t))E_z[\nabla_j f_i^*(t)]dt-\int\delta_i^*(t)E_z[\nabla_j f_i^*(t)]dt\right|\right]$$
$$\leq\sqrt{n}\int E_z\left[|A(\delta_i^*(t))E_z[\nabla_j f_i^*(t)]-\delta_i^*(t)E_z[\nabla_j f_i^*(t)]|\right]dt$$
$$\leq B\int E_z[|\sqrt{n}\gamma_i^{*2}|]\left|\frac{\sum_{l=1}^n x_{il}x_{jl}}{\sigma\sum_{l=1}^n x_{il}}\right||\nabla_t g^*(t)|dt$$
$$\leq BKU_i^{1/2}\frac{1}{\sigma}\int\frac{\lambda^{1/2}(t)}{g^*(t)}|\nabla_t g^*(t)|dt$$
$$\leq BKU_i^{1/2}\frac{1}{\sigma}\int M(h)\frac{|\nabla_t g^*(t)|}{g^{*1/2}(t)}dt$$
$$\leq BKU_i^{1/2}M(h)\frac{1}{\sigma}\left(\int g^*(t)\left|\frac{\nabla_t g^*(t)}{g^*(t)}\right|^2dt\right)^{1/2},$$

of which limsup is bounded by the assumption.

Therefore, the second term on the right-hand side of (5.2) goes to zero with probability 1. Hence the result holds.

THEOREM 5.3. *Suppose that* $V=\mathrm{Var}[\int k(z;t,h)\nabla_t\log g^*(t)dt]$ *exists. Also suppose that for* $i,j=0,\dots,p$, $\bar{x}_i=\lim_{n\to\infty}(1/n)\sum_{l=1}^n x_{il}$ *and* $\bar{x}_{ij}=\lim_{n\to\infty}(1/n)\sum_{l=1}^n x_{il}x_{jl}$ *exist. Then for a regular RAF*

$$\sqrt{n}\int A(\delta_i^*(t))\nabla_j f_i^*(t)dt\to N(0,a_{ij}V/\sigma^2),$$

where $a_{ij} = (\bar{x_{ij}})^2 \bar{x_{ii}}/(\bar{x}_i)^4$.

PROOF. By Lemma 5.2 and Theorem 4.4.6 of Chung (1974), the asymptotic distribution of

$$\sqrt{n}\int A(\delta_i^*(t))\nabla_j f_i^*(t)dt$$

is as same as that of

$$\begin{aligned}
&\sqrt{n}\int \delta_i^*(t)E_z[\nabla_t g^*(t)]dt \\
&\quad = \sqrt{n}\int \delta_i^*(t)\frac{\sum_{l=1}^n x_{il}x_{jl}}{\sigma\sum_{l=1}^n x_{il}}\nabla_t g^*(t)dt \\
&\quad = \frac{\sum_{l=1}^n x_{il}x_{jl}/n}{\sigma\sum_{l=1}^n x_{il}/n}\frac{1}{\sum_{l=1}^n x_{il}/n}\frac{1}{\sqrt{n}}\sum_{l=1}^n x_{il}\int (k(z_l;t,h)-g^*(t))\nabla_t \log g^*(t)dt.
\end{aligned}$$

By the Central Limit Theorem and by Slutzky's theorem, we have the result.

Now, we will prove the main theorem about the existence and consistency of estimators of the regression coefficients, b_i for $i = 0, \ldots, p$, and we will derive the asymptotic distribution of a vector of estimators:

$$\sqrt{n}(\boldsymbol{b}_n - \boldsymbol{b}_o) = \sqrt{n}\{(b_{0n} - b_{0o}), (b_{1n} - b_{1o}), \ldots, (b_{pn} - b_{po})\}.$$

THEOREM 5.4. *In addition to the conditions of Theorem* 5.3, *assume that* $\int g^*(t)(\nabla_t \log g^*(t))^2 dt$ *exists and that there are functions* $M_{jkl(i)}(t)$ *such that* $|\nabla_{jkl} f_i^*(t)/f_i^*(t)| \le M_{jkl(i)}(t)$, *where* $M_{jkl(i)}(t)$ *for all* j, k, l *and* i, *have finite expectation w.r.t.* $f_i^*(t)$. *Then there exists a consistent sequence of roots* $\boldsymbol{b}_n$ *to the minimum disparity estimating equations*:

$$\nabla_{\boldsymbol{b}} S(\boldsymbol{b}) = \nabla_{\boldsymbol{b}} \sum_{i=0}^{p} \rho_i(\boldsymbol{b}) = 0.$$

PROOF. See the Appendix.

Remark 5.1. We need to consider another important property of a kernel. Suppose $k(z;t,h)$ is transparent (Basu and Lindsay (1994)), that is,

$$S^*(z) = \int k(z;t,h)\nabla_t \log g^*(t)dt = c\nabla_z g(z) = cS(z)$$

for some constant c, where $S(z)$ is $\nabla_z \log g(z)$. Since $E[S^*(z)] = cE_z[S(z)] = 0$, then $\mathrm{Var}[S^*(z)] = E[S^{*2}(z)] = c^2E[S^2(z)] = \; c^2E[\nabla_z S(z)] = \; cE[\nabla_z S^*(z)]$.

Now, we have

$$\begin{aligned}
E[\nabla_z S^*(z)] &= \int g(z) \left(\int \nabla_z k(z;t,h) \nabla_t \log g^*(t) dt \right) dz \\
&= \int \nabla_t \log g^*(t) \left(\int g(z) \nabla_z k(z;t,h) dz \right) dt \\
&= \int \nabla_t \log g^*(t) \left(\int g(z)(-\nabla_t k(z;t,h)) dz \right) dt \\
&= \int \nabla_t \log g^*(t) \left(-\nabla_t \int g(z) k(z;t,h) dz \right) dt \\
&= \int \nabla_t \log g^*(t)(-\nabla_t g^*(t)) dt \\
&= \int \frac{\nabla_t g^*(t)}{g^*(t)} (-\nabla_t g^*(t)) dt \\
&= -\int g^*(t) \left(\frac{\nabla_t g^*(t)}{g^*(t)} \right)^2 dt.
\end{aligned}$$

Furthermore, if z is a properly standardized random variable of a symmetric density in the exponential family as we assumed in the beginning, then $E[\nabla_z S(z)] = E[\nabla_z^2 \log g(z)] = -\operatorname{Var}[z] = -1$. Since $E[\nabla_z S^*(z)] = cE[\nabla_z S(z)] = -c$, it follows that

$$c = -\int g^*(t) \left(\frac{\nabla_t g^*(t)}{g^*(t)} \right)^2 dt.$$

Therefore,

$$V\left[\int k(z;t,h) \nabla_t \log g^*(t) dt \right] = \left(\int g^*(t)(\nabla_t \log g^*(t))^2 dt \right)^2.$$

For example, suppose that $k(z;t,h)$ is a density with mean z and variance h^2, and that z is a normal random variable with mean 0 and variance 1. We have the following:

$$\begin{aligned}
\operatorname{Var}\left[\int k(z;t,h) \nabla_t \log g^*(t) dt \right] &= \operatorname{Var}\left[-\frac{1}{h^2+1} z \right] \\
&= \frac{1}{(h^2+1)^2} \operatorname{Var}[z] \\
&= \frac{1}{(h^2+1)^2},
\end{aligned}$$

and

$$\begin{aligned}
\int g^*(t)(\nabla_t \log g^*(t))^2 dt &= \int g^*(t) \left(\frac{t}{h^2+1} \right)^2 dt \\
&= \frac{1}{h^2+1}.
\end{aligned}$$

THEOREM 5.5. *Under the conditions of Theorem 5.4* $\sqrt{n}(\boldsymbol{b}_n - \boldsymbol{b}_o)$ *has an asymptotic normal distribution with mean zero and variance matrix* $\boldsymbol{\Sigma}_\infty^{-1}\sigma^2$, *where* $\boldsymbol{\Sigma}_\infty = \lim_{n\to\infty}(1/n)\boldsymbol{X}^T\boldsymbol{X}$.

PROOF. Expand $\nabla_j S(\boldsymbol{b})$ about $\boldsymbol{b}_o$ and replace $\nabla_j S(\boldsymbol{b})$ by $\nabla_j S(\boldsymbol{b}_n)$ to obtain

$$\nabla_j S(\boldsymbol{b}_n) = \nabla_j S(\boldsymbol{b}_o) + \sum (b_{kn} - b_{ko})\nabla_{jk} S(\boldsymbol{b}_o) + \frac{1}{2}\sum\sum (b_{kn} - b_{ko})(b_l - b_{lo})\nabla_{jkl} S(\boldsymbol{b}'), \tag{5.3}$$

where $\boldsymbol{b}'$ is a point on the line segment connecting $\boldsymbol{b}$ and $\boldsymbol{b}_o$. The term on the left-hand side is zero, so that the resulting equations can be written as

$$\sqrt{n}\sum (b_{kn} - b_{ko})\left[\nabla_{jk} S(\boldsymbol{b}_o) + \frac{1}{2}\sum (b_{ln} - b_{lo})\nabla_{jkl} S(\boldsymbol{b}')\right] = -\sqrt{n}\nabla_j S(\boldsymbol{b}_o).$$

Now, let

$$\begin{aligned} Y_{kn} &= \sqrt{n}(b_{kn} - b_{ko}) \\ A_{jkn} &= \nabla_{jk} S(\boldsymbol{b}_o) + \frac{1}{2}\sum (b_{ln} - b_{lo})\nabla_{jkl} S(\boldsymbol{b}') \\ T_{jn} &= -\sqrt{n}\nabla_j S(\boldsymbol{b}_o). \end{aligned}$$

As we claim in Theorem 5.4 that the estimators are consistent and that the $\nabla_{jkl} S(\boldsymbol{b}')$ is bounded in probability, and by the (A.4) in the proof of Theorem 5.4, we have

$$A_{jkn} \to \sum_{i=0}^{p} \frac{1}{\sigma^2}\frac{\bar{x_{ij}}\bar{x_{ik}}}{\bar{x}_i\bar{x}_i}\int g^*(t)\left(\frac{\nabla_t g^*(t)}{g^*(t)}\right)^2 dt$$

in probability.

The limiting distribution of the Y_{kn}'s is therefore that of the solution $\boldsymbol{Y}_n = (Y_{0n}, \ldots, Y_{pn})$ of the equations

$$\sum_{k=0}^{p}\left(\sum_{i=0}^{p} \frac{1}{\sigma^2}\frac{\bar{x_{ij}}\bar{x_{ik}}}{\bar{x}_i\bar{x}_i}\int g^*(t)\left(\frac{\nabla_t g^*(t)}{g^*(t)}\right)^2 dt\right) Y_{kn} = T_{jn}.$$

Let's denote $[T_{jn}]$ be a $p \times 1$ vector of T_{jn}, $j = 1, \ldots, p$. Based on the facts of Theorem 5.3, $[T_{jn}]$ has the same asymptotic density as that of $(1/\sigma)[a_{ij}]_{p\times p} \cdot [b_i]_{p\times 1} \cdot [c_i]_{p\times 1}$, where

$$a_{ij} = \frac{\sum_{l=1}^{n} x_{il}x_{jl}}{n}, \qquad b_i = \left(\frac{\sum_{l=1}^{n} x_{il}}{n}\right)^{-2}$$

$$c_i = \frac{1}{\sqrt{n}}\sum_{l=1}^{n} x_{il}\int (k(z_l; t, h) - g^*(t))\nabla_t \log g^*(t) dt$$

for $i, j = 0, 1, \ldots, p$. Hence, $\boldsymbol{T}_n = (T_{0n}, \ldots, T_{pn})$ is asymptotically multivariate normal with mean zero and covariance matrix $(\boldsymbol{A}^T \boldsymbol{B}^{-1})\boldsymbol{A}(\boldsymbol{B}^{-1}\boldsymbol{A})V/\sigma^2$.

$$\boldsymbol{A} = \begin{bmatrix} \bar{x_{00}} & \bar{x_{01}} & \cdots & \bar{x_{0p}} \\ \bar{x_{10}} & \bar{x_{11}} & \cdots & \bar{x_{1p}} \\ \vdots & \vdots & \cdots & \vdots \\ \bar{x_{p0}} & \bar{x_{p1}} & \cdots & \bar{x_{pp}} \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} \bar{x_0}^2 & 0 & \cdots & 0 \\ 0 & \bar{x_1}^2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \bar{x_p}^2 \end{bmatrix}.$$

Therefore, it follows that the distribution of $\boldsymbol{Y}_n$ is that of

$$\left[(\boldsymbol{A}\boldsymbol{B}^{-1}\boldsymbol{A}^T)\frac{1}{\sigma^2} \int g^*(t) \left(\frac{\nabla_t g^*(t)}{g^*(t)} \right)^2 dt \right]^{1} \boldsymbol{T}_n,$$

which is an asymptotically multivariate normal with mean zero and covariance matrix $\sigma^2 \boldsymbol{A}^{-1}$. In fact,

$$\boldsymbol{A} = \boldsymbol{\Sigma}_\infty = \lim_{n\to\infty} \frac{1}{n} \boldsymbol{X}^T \boldsymbol{X},$$

where $\boldsymbol{X}$ is a design matrix of (3.1).

Acknowledgements

This research was supported by the Basic Science Research Institute Program, Ministry of Education, the Republic of Korea, 1997, Project No. BSRI-97-1439.

Appendix

PROOF OF THEOREM 5.4.

1. To prove existence and consistency we will follow the argument of Lehmann ((1991), 430–435). Consider the behavior of $S(\boldsymbol{b})$ on a sphere Q_a which has radius a and center at $\boldsymbol{b}_o$, the vector of true values of parameters. We will show that for a sufficiently small a, the probability tends to 1 that $S(\boldsymbol{b}) > S(\boldsymbol{b}_o)$ and with probability tending to 1 there is a local minimum for the disparity in the interior of Q_a. At a local minimum the estimating equations must be satisfied. Therefore, for any $a > 0$, the minimum disparity estimating equations have a solution $\boldsymbol{b}_n$ within Q_a, with probability tending to 1 as $n \to \infty$.

2. Expand $S(\boldsymbol{b})$ about $\boldsymbol{b}_o$, we get

$$\begin{aligned} S(\boldsymbol{b}_o) - S(\boldsymbol{b}) &= \sum_{i=0}^{p} [\rho_i(\boldsymbol{b}_o) - \rho_i(\boldsymbol{b})] \\ &= \sum_{j=0}^{p} \left[(b_j - b_{jo}) \sum_{i=0}^{p} \nabla_j \rho_i(\boldsymbol{b}) \mid_{b=b_o} \right] \\ &\quad + \frac{1}{2} \sum_{k=0}^{p} \sum_{j=0}^{p} \left[(b_k - b_{ko})(b_j - b_{jo}) \sum_{i=0}^{p} \nabla_{jk} \rho_i(\boldsymbol{b}) \mid_{b=b_o} \right] \\ &\quad + \frac{1}{6} \sum_{l=0}^{p} \sum_{k=0}^{p} \sum_{j=0}^{p} \left[(b_l - b_{lo})(b_k - b_{ko})(b_j - b_{jo}) \sum_{i=0}^{p} \nabla_{jkl} \rho_i(\boldsymbol{b}) \mid_{b=b'} \right] \\ &= S_1 + S_2 + S_3. \end{aligned}$$

$\boldsymbol{b}'$ is a point on the line segment connecting $\boldsymbol{b}$ and $\boldsymbol{b}_o$. We will investigate the linear, quadratic and cubic terms in order and determine their proper limits.

3. Linear term S_1: Since $|A(\delta_i^*)|$ is bounded for a large n and also

$$E[\nabla_j f_i^*(t)] = \frac{\sum_{l=1}^n x_{il}x_{jl}}{\sum_{l=1}^n x_{il}} \frac{\nabla_t g^*(t)}{\sigma},$$

similar to Lemma 5.3, we can show that

$$\begin{aligned}\nabla_j \rho_i(\boldsymbol{b}) \mid_{\boldsymbol{b}=\boldsymbol{b}_o} &- \int A(\delta_i^*) \frac{\sum_{l=1}^n x_{il}x_{jl}}{\sum_{l=1}^n x_{il}} \frac{\nabla_t g^*(t)}{\sigma} dt \\ &= \int A(\delta_i^*) \nabla_j f_i^*(t) dt - \int A(\delta_i^*) \frac{\sum_{l=1}^n x_{il}x_{jl}}{\sum_{l=1}^n x_{il}} \frac{\nabla_t g^*(t)}{\sigma} dt \to 0,\end{aligned}$$

in probability. Next, we will show that

$$\int A(\delta_i^*) \frac{\sum_{l=1}^n x_{il}x_{jl}}{\sum_{l=1}^n x_{il}} \frac{\nabla_t g^*(t)}{\sigma} dt \to 0$$

in probability. Assume that $\int g^{*1/2}(t)|\nabla_t g^*(t)/g^*(t)|dt$ is bounded and show that the expected value of the above integral goes to 0. Since $|A(\delta_i^*(t))| \leq |\delta_i^*(t)|$, and by (4.3), for any constant K, we have that

$$\begin{aligned}(\text{A.1}) \qquad &\limsup_n E\left[\left|\int A(\delta_i^*) \frac{\sum_{l=1}^n x_{il}x_{jl}}{\sum_{l=1}^n x_{il}} \frac{\nabla_t g^*(t)}{\sigma} dt\right|\right] \\ &\leq \limsup_n \int \left|\frac{\sum_{l=1}^n x_{il}x_{jl}}{\sum_{l=1}^n x_{il}}\right| E[|A(\delta_i^*)|] \left|\frac{\nabla_t g^*(t)}{\sigma}\right| dt \\ &\leq \limsup_n \int \left|\frac{\sum_{l=1}^n x_{il}x_{jl}/n}{\sum_{l=1}^n x_{il}/n}\right| E[|\delta_i^*|] \left|\frac{\nabla_t g^*(t)}{\sigma}\right| dt \\ &\leq K \int \lambda^{1/2}(t) \left|\frac{\nabla_t g^*(t)}{\sigma g^*(t)}\right| dt \\ &\leq K M^{1/2}(h) \int g^{*1/2}(t) \left|\frac{\nabla_t g^*(t)}{\sigma g^*(t)}\right| dt < \infty, \\ &\leq K M^{1/2}(h) \left(\int g^*(t) \left|\frac{\nabla_t g^*(t)}{\sigma g^*(t)}\right|^2 dt\right)^{1/2} < \infty,\end{aligned}$$

and thus by the dominated convergence theorem the expectation of the integral (A.1) goes to 0. Once again by the Markov's inequality the integral itself goes to 0 in probability. Therefore,

$$\nabla_j \rho_i(\boldsymbol{b}) \mid_{\boldsymbol{b}=\boldsymbol{b}_o} \to 0$$

in probability. Applying the same procedure as the above to each of the other first-order terms for $i = 0, \ldots, p$, then we have

$$\sum_{i=0}^p \nabla_j \rho_i(\boldsymbol{b}) \mid_{\boldsymbol{b}=\boldsymbol{b}_o} \to 0 \qquad (\text{A.2})$$

in probability as $n \to \infty$. For any given a, it follows from (A.2) that

$$\left| \sum_{i=0}^{p} \nabla_j \rho_i(\boldsymbol{b}) \mid_{\boldsymbol{b}=\boldsymbol{b}_o} \right| < (p+1)a^2,$$

and thus, with probability tending to 1, $|S_1| < (p+1)^2 a^3$, where $p+1$ is the number of regression coefficients and a is the radius of the sphere Q_a.

4. Quadratic term S_2: Since

$$\nabla_{jk}\rho_i(\boldsymbol{b}) \mid_{\boldsymbol{b}=\boldsymbol{b}_o} = \int A'(\delta_i^*) \frac{1}{g_i^*(t)} \nabla_k f_i^*(t) \nabla_j f_i^*(t) dt + \int A(\delta_i^*) \nabla_{jk} f_i^*(t) dt,$$

similar to the proof of Lemma 5.3, we can show that

$$\int A'(\delta_i^*) \frac{1}{g_i^*(t)} \nabla_k f_i^*(t) \nabla_j f_i^*(t) dt + \int A(\delta_i^*) \nabla_{jk} f_i^*(t) dt$$
$$- \left[\int A'(\delta_i^*) \frac{1}{g_i^*(t)} \frac{\sum_{l=1}^n x_{il} x_{kl}}{\sum_{l=1}^n x_{il}} \frac{\sum_{l=1}^n x_{il} x_{jl}}{\sum_{l=1}^n x_{il}} \left(\frac{\nabla_t g^*(t)}{\sigma} \right)^2 dt \right.$$
$$\left. + \int A(\delta_i^*) \frac{\sum_{l=1}^n x_{kl} x_{jl} x_{il}}{\sigma^2 \sum_{l=1}^n x_{il}} \nabla_t^2 g^*(t) dt \right] \to 0$$

in probability. We will first show that

$$\int A'(\delta_i^*) \frac{1}{g_i^*(t)} \frac{\sum_{l=1}^n x_{il} x_{kl}}{\sum_{l=1}^n x_{il}} \frac{\sum_{l=1}^n x_{il} x_{jl}}{\sum_{l=1}^n x_{il}} \left(\frac{\nabla_t g^*(t)}{\sigma} \right)^2 dt$$
$$\to \frac{1}{\sigma^2} \frac{\bar{x_{ij}} \bar{x_{ik}}}{\bar{x}_i \bar{x}_i} \int g^*(t) \left(\frac{\nabla_t g^*(t)}{g^*(t)} \right)^2 dt$$

in probability. Consider the supremum of the absolute value of the integral in left-hand side, which turns out to be bounded due to the boundedness of $A'(\delta_i^*)$ and by the assumption that $\int |g^*(t)(\nabla_t g^*(t)/\sigma g^*(t))^2| dt$ is bounded. For $A'(\delta_i^*) \to 1$, by the dominated convergence theorem we have the result.

Next, we will show that

$$\int A(\delta_i^*) \frac{\sum_{l=1}^n x_{kl} x_{jl} x_{il}}{\sigma^2 \sum_{l=1}^n x_{il}} \nabla_t^2 g^*(t) dt \to 0$$

in probability. Consider that the supremum of the absolute value of the above integral is bounded due to the boundedness of $A(\delta_i^*(t))$ and the integrability of $\nabla_t^2 g^*(t)$. Since $A(\delta_i^*(t)) \to 1$, and $\int \nabla_t^2 g^*(t) dt = \nabla_t^2 \int g^*(t) dt = 0$, by the dominated convergence theorem the result follows.

Hence, we have that

$$\text{(A.3)} \qquad \sum_{i=0}^{p} \nabla_{jk} \rho_i(\boldsymbol{b}) \mid_{\boldsymbol{b}=\boldsymbol{b}_o} \to \sum_{i=0}^{p} \frac{1}{\sigma^2} \frac{\bar{x_{ij}} \bar{x_{ik}}}{\bar{x}_i \bar{x}_i} \int g^*(t) \left(\frac{\nabla_t g^*(t)}{g^*(t)} \right)^2 dt.$$

Let us denote $J^*(\boldsymbol{b}_o)$ as a matrix whose jk-th element, $J^*_{jk}(\boldsymbol{b}_o)$, is the right hand side of (A.3) for $j, k = 0, \ldots, p$.

Therefore,

$$2S_2 = \sum_{k=0}^{p}\sum_{j=0}^{p}\left\{\left[\nabla_k \sum_{i=0}^{p} \nabla_j \rho_i\right] - [J^*_{jk}(\boldsymbol{b}_o)]\right\}(b_j - b_{jo})(b_k - b_{ko})$$
$$+ \sum_{k=0}^{p}\sum_{j=0}^{p}[J^*_{jk}(\boldsymbol{b}_o)](b_j - b_{jo})(b_k - b_{ko}).$$

Similar to the argument of Lehmann ((1991), p. 432) we see that there exists $c > 0$, $a_0 > 0$ such that $a < a_0$, $S_2 < -ca^2$ with probability tending to 1.

5. Cubic term S_3: Since most kernels are bounded and continuously differentiable, we can claim that $\nabla_{jkl}\rho_i(\boldsymbol{b})\mid_{\boldsymbol{b}=\boldsymbol{b}'}$ is bounded by some uniformly bounded function $M_{jkl(i)}(t)$ for all j, k, l. Hence $\sum_{i=0}^{p}\nabla_{jkl}\rho_i(\boldsymbol{b})\mid_{\boldsymbol{b}=\boldsymbol{b}'}$ will be also bounded. Therefore, we have $|S_3| < ba^3$ on Q_a, where

$$b = \frac{(p+1)^3}{3}\sum_{l=1}^{p}\sum_{k=1}^{p}\sum_{j=1}^{p}\sum_{i=1}^{p} E[M_{jkl(i)}(t)].$$

Combining the three inequalities about S_1, S_2 and S_3, then we see that

$$\max(S_1 + S_2 + S_3) < -ca^2 + (b + p + 1)a^3$$

which is less than zero if $a < c/(b + p + 1)$.

Thus, for any sufficiently small a there exists a sequence of roots $\boldsymbol{b}_n$ to the minimum disparity estimating equations such that for $i = 0, \ldots, p$

$$P(\|b_{in} - b_{i0}\| < a) \to 0,$$

where $\|\cdot\|$ represent the L_2 norm. It remains to show that we can determine such a sequence independently of 'a'. Let $\boldsymbol{b}^*_n$ be the root which is the closest to $\boldsymbol{b}_o$. This exists because the limit of a sequence of roots is again a root by the continuity of the disparity as a function of the parameter. This completes the proof of the consistency part.

References

Basu, A. and Lindsay, B. G. (1994). Minimum disparity estimation for continuous models: efficiency, distributions and robustness, *Ann. Inst. Statist. Math.*, **46**, 683–705.

Beran, R. J. (1977). Minimum Hellinger distance estimates for parametric models, *Ann. Statist.*, **5**, 445–463.

Chung, K. L. (1974). *A Course in Probability Theory*, Academic Press, New York.

Chow, Y. S. and Teicher, H. (1988). *Probability Theory; Independence, Interchangeability, Martingales*, 2nd ed., Springer, New York.

Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests, *J. Roy. Statist. Soc. Ser. B*, **46**, 440–464.

Lehmann, E. L. (1991). *Theory of Point Estimation*, Wiley, New York.

Lindsay, B. G. (1994). Efficiency versus robustness: the case for minimum Hellinger distance and related methods, *Ann. Statist.*, **22**, 1081–1114.

Pak, Ro Jin (1995). Robustness of minimum disparity estimation in linear regression models, *Journal of Korean Statistical Association*, **24**, 349–360.

Pak, Ro Jin (1996). Minimum Hellinger distance estimation in simple linear regression models; distribution and efficiency, *Statist. Probab. Lett.*, **26**, 263–269.

Rao, C. R. (1961). Asymptotic efficiency and limiting information, *Proceedings of the Fourth Berkeley Symposium*, Vol. 1, 531–546, University of California Press, Berkeley.

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

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York.

Simpson, D. G. (1987). Minimum Hellinger distance estimation for the analysis of count data, *J. Amer. Statist. Assoc.*, **82**, 802–807.

Tamura, R. N. and Boos, D. D. (1986). Minimum Hellinger distance estimation for multivariate location and covariance, *J. Amer. Statist. Assoc.*, **81**, 223–229.