Journal of the Korean
Statistical Society
Vol. 27, No. 1, 1998

Minimum Distance Estimation Based On The Kernels For U-Statistics [†]

Hyo-Il Park [1]

ABSTRACT

In this paper, we consider a minimum distance(M.D.) estimation based on kernels for U-statistics. We use *Cramér – von Mises* type distance function which measures the discrepancy between U-empirical distribution function(d.f.) and modeled d.f. of kernel. In the distance function, we allow various integrating measures, which can be finite, σ-finite or discrete. Then we derive the asymptotic normality and study the qualitative robustness of M.D. estimates.

Key Words : Asymptotic bias; gross error model; kernel for U-statistic; qualitatively robust; U-empirical process.

[†] This paper was supported by NON DIRECTED RESEARCH FUND, Korea Research Foundation, 1994.

[1] Department of Applied Statistics, Chong-ju University, Chong-ju, Choong-book, 360-764, Korea

1. INTRODUCTION

For each n, let $X_1, X_2, \ldots, X_n$ be a sample with a modeled d.f. $F_{n:1}$ which may be different from the actual d.f. $G_{n:1}$ with some parameter ξ which is to be estimated. We assume that $F_{n:1}$ and $G_{n:1}$ are continuous and have densities $f_{n:1}$ and $g_{n:1}$, respectively. Here we allow that $F_{n:1}$ and $G_{n:1}$ can be varied with n. Let h be any symmetric kernel of ξ with degree k such as

$$E[h(X_1, X_2, \ldots, X_k)] = \xi \tag{1.1}$$

Also let $F_{n:k}$ and $G_{n:k}$ be the d.f.s of h under $F_{n:1}$ and $G_{n:1}$ with respective densities $f_{n:k}$ and $g_{n:k}$. Furthermore let $< n, k > = \binom{n}{k}$, and

$$S_n(t; u) = \frac{\sqrt{n}}{< n, k >} \sum_{j=1}^{<n,k>} [I(h_j \leq t + u) - F_{n:k}(t)] \tag{1.2}$$

where h_j represents a generic random variable from $< n, k >$ random variables for notational convenience. Finally, let λ be a $\sigma - finite$ (or $finite$) measure and define $Cramér - von\ Mises$ type distance function

$$M_n(\lambda; u) = \int_{-\infty}^{\infty} S_n^2(t; u) d\lambda(t). \tag{1.3}$$

We denote $\hat{\xi}_n(\lambda)$ as a minimizer of (1.3) if it exists. Then $\hat{\xi}_n(\lambda)$ satisfies that

$$\inf_u M_n(\lambda; u) = M_n(\lambda; \hat{\xi}_n(\lambda)) \tag{1.4}$$

and is called an M.D. estimate of ξ.

The M.D. estimation method based on $Cramér - von\ Mises$ type distances, has long been one of the research topics in the theoretical Statistics. Parr(1981) provided an extensive bibliography on the M.D. estimation classified by subject matters up to 1980. However the applications of the M.D. estimation have been confined mainly to location parameters. Koul and DeWet(1983) obtained a class of M.D. estimates of the slope parameter in the linear regression model. Furthermore, very recently, Dhar(1991) applied the M.D. estimation method to the time series data. However, even for the case of the linear regression model or time series data, when $Cramér - von\ Mises$ type distances are constructed, the slope parameter and the autoregressive parameter in the weighted empirical d.f.s play just the role of location parameters. This point makes us to consider using kernels for U-statistics since the interested parameters are just the means of d.f.s of the corresponding kernels and so they become location parameters in the d.f.s of kernels whatever originally they are. Therefore we can apply the M.D. estimation method to other than location parameters such as scale parameter from the nature of the case. For kernels, we consider general forms which were used by Serfling(1984) and Akritas(1986). They obtained the generalized L-estimates from the U-empirical d.f.s which are constructed from the random variables generated by general forms of kernels. In some cases, the generalized L- estimates coincide with Hodges-Lehmann estimates according to choices of kernels. In this vein, we will use general forms of kernels. The integrating measure λ may be chosen along with the level of knowledge for $F_{n:1}$ or $F_{n:k}$. For example, if $F_{n:1}$ or $F_{n:k}$ were fully parameterized, then λ could be chosen as a weight function with Lebesgue measure proposed by Boos(1981) for the efficiency considerations. For more discussions for this subject, refer to Koul(1992). In the sequel, we use F_n and G_n instead of $F_{n:k}$ and $G_{n:k}$ for notational brevity when no confusion arises. Also for densities, we use f_n and g_n instead of $f_{n:k}$ and $g_{n:k}$ in the same situation.

2. EXISTENCE AND ASYMPTOTIC NORMALITY OF M.D. ESTIMATES

In order to discuss the existence and derive the asymptotic normality of M.D. estimates, we begin by introducing some notations and then stating several assumptions

NOTATIONS : For each $i = 0, 1, \ldots, k-1$, let

$$H_n^i(t|x) = \int \cdots \int I(h(x_1, x_2, \ldots, x_i, x, x_{i+2}, \ldots, x_k) - \xi \le t) \prod_{j=1}^{i} dF_{n:1}(x_j) \prod_{j=i+2}^{k} dG_{n:1}(x_j). \qquad (2.1)$$

Especially, when $i = 0$, we note that since

$$H_n^0(t|x) = \int \cdots \int I(h(x, x_2, \ldots, x_k) - \xi \le t) \prod_{j=2}^{k} dG_{n:1}(x_j). \qquad (2.2)$$

$$E_{G_{n:1}}(H_n^0(t|X_1)) = P_{G_{n:1}}(h(X_1, X_2, \dots, X_k) - \xi \le t) = G_n(t). \qquad (2.3)$$

ASSUMPTIONS :

$$(A1) \quad n\int_{-\infty}^{\infty}(G_{n:1}(t) - F_{n:1}(t))^2 d\lambda(t) = O(1). \qquad (2.4)$$

$$(A2) \quad \int_{-\infty}^{\infty}(G_{n:1}(t)(1 - G_{n:1}(t))d\lambda(t) < \infty. \qquad (2.5)$$

(A3) For each $i, i = 1, 2, \dots, k-1$ and for all $t \in \mathbf{R}^1$,

$$\left|\int_{-\infty}^{\infty} H_n^i(t|x)d(G_{n:1}(x) - F_{n:1}(x))\right| \le \alpha_i |G_{n:1}(t) - F_{n:1}(t)|, \qquad (2.6)$$

for some suitable positive constants α_i.

(A4) For $i = 0$, there is a compact subset $\mathbf{M} \subset \mathbf{R}^1$ such that for all $t \in \mathbf{R}^1 \backslash \mathbf{M}$,

$$\int_{-\infty}^{\infty} H_n^0(t|x)dG_{n:1}(x) \le \beta_0 G_{n:1}(t) \qquad (2.7)$$

and

$$\int_{-\infty}^{\infty}(1 - H_n^0(t|x))dG_{n:1}(x) \le \beta_0(1 - G_{n:1}(t)), \qquad (2.8)$$

for some suitable positive constant β_0.

$$(A5) \lim_{z\to 0}\int_{-\infty}^{\infty} g_n(t+z)d\lambda(t) = \int_{-\infty}^{\infty} g_n(t)d\lambda(t) < \infty \qquad (2.9)$$

and

$$\lim_{z\to 0}\int_{-\infty}^{\infty} g_n^2(t+z)d\lambda(t) = \int_{-\infty}^{\infty} g_n^2(t)d\lambda(t) < \infty. \qquad (2.10)$$

Lemma 1. Both assumptions A1 and A3 imply that

$$n\int_{-\infty}^{\infty}(G_n(t)-F_n(t))^2 d\lambda(t)=O(1). \tag{2.11}$$

Proof. First of all, we note that

$$\begin{aligned}
&G_n(t)-F_n(t)\\
=\;& \int\cdots\int I(h(x_1,x_2,\ldots,x_k)-\xi\le t)\{\prod_{i=1}^{k} dG_{n:1}(x_i)-\prod_{i=1}^{k} dF_{n:1}(x_i)\}\\
=\;& \sum_{i=0}^{k-1}\int\cdots\int I(h(x_1,x_2,\ldots,x_k)-\xi\le t)\prod_{j=1}^{i} dF_{n:1}(x_j)\\
& \qquad \prod_{j=i+2}^{k} dG_{n:1}(x_j)\{dG_{n:1}(x_{i+1})-dF_{n:1}(x_{i+1})\}\\
=\;& \sum_{i=0}^{k-1}\int_{-\infty}^{\infty} H_n^i(t|x)(dG_{n:1}(x)-dF_{n:1}(x)).
\end{aligned}$$

Then by applying Minkowski's inequality and using A3, we have

$$\begin{aligned}
&n\int_{-\infty}^{\infty}(G_n(t)-F_n(t))^2 d\lambda(t)\\
=\;& n\int_{-\infty}^{\infty}\left\{\sum_{i=0}^{k-1}\int_{-\infty}^{\infty} H_n^i(t|x)(dG_{n:1}(x)-dF_{n:1}(x))\right\}^2 d\lambda(t)\\
\le\;& n\left\{\sum_{i=0}^{k-1}\left\{\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} H_n^i(t|x)(dG_{n:1}(x)-dF_{n:1}(x))\right\}^2 d\lambda(t)\right\}^{1/2}\right\}^2\\
\le\;& n\left\{\sum_{i=0}^{k-1}\alpha_i\left\{\int_{-\infty}^{\infty}(G_{n:1}(t)-F_{n:1}(t))^2 d\lambda(t)\right\}^{1/2}\right\}^2.
\end{aligned}$$

Finally, by appealing to A1, we see the result.

Lemma 2. Both assumptions A2 and A4 imply that

$$\int_{-\infty}^{\infty} G_n(t)(1 - G_n(t))d\lambda(t) < \infty. \tag{2.12}$$

Proof. From A4, we have

$$\begin{aligned}
&\int_{-\infty}^{\infty} G_n(t)(1 - G_n(t))d\lambda(t) \\
= &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_n^0(t|x)dG_{n:1}(x)\int_{-\infty}^{\infty}(1 - H_n^0(t|x))dG_{n:1}(x)d\lambda(t) \\
= &\int_{\mathbf{M}}\int_{-\infty}^{\infty} H_n^0(t|x)dG_{n:1}(x)\int_{-\infty}^{\infty}(1 - H_n^0(t|x))dG_{n:1}(x)d\lambda(t) \\
&+\int_{\mathbf{R}^1\setminus\mathbf{M}}\int_{-\infty}^{\infty} H_n^0(t|x)dG_{n:1}(x)\int_{-\infty}^{\infty}(1 - H_n^0(t|x))dG_{n:1}(x)d\lambda(t) \\
\leq &\ \lambda(\mathbf{M}) + \beta_0^2\int_{\mathbf{R}^1\setminus\mathbf{M}} G_{n:1}(t)(1 - G_{n:1}(t))d\lambda(t).
\end{aligned}$$

Thus the result follows from the compactness of $\mathbf{M}$ and A2.

Note. As a matter of fact, in case of a convex combination of random variables, Assumptions A3 and A4 are unnecessary and so Lemmas 1 and 2 can be derived from Assumptions A1 and A2 directly. As an example, consider

$$h(X_1, X_2, \cdots, X_k) = (X_1 + X_2 + \ldots + X_k)/k \tag{2.13}$$

Then the variance of $h(X_1, X_2, \ldots, X_k)$ is σ^2/k where σ^2 is the variance of X_1. Thus there exist two points $t_1 < t_2$ such that

$$\begin{cases} G_k(t) < G(t) & \text{if } t < t_1 \\ G_k(t) > G(t) & \text{if } t_1 \leq t \leq t_2 \\ G_k(t) < G(t) & \text{if } t > t_2 \end{cases} \tag{2.14}$$

where G and G_k are the actual d.f.'s of X_1 and $h(X_1, X_2, \ldots, X_k)$, respectively. Thus for $t_1 < t < t_2$, $G_k(t)(1 - G_k(t)) > G(t)(1 - G(t))$ whereas for $t < t_1$ or $t > t_2$, $G_k(t)(1 - G_k(t)) < G(t)(1 - G(t))$ since $G_k(t_1) < 1/2$ and

$G_k(t_2) > 1/2$. Thus Lemma 2 follows directly from A2. Also when the kernel h is of the form (2.13), we note that

$$\begin{aligned}
&\frac{d}{dt}\int_{-\infty}^{\infty} H_n^i(t|x)(dG_{n:1}(x) - dF_{n:1}(x)) \\
&= \underbrace{(f_{n:1} * \cdots * f_{n:1})}_{\text{i times}} * \underbrace{(g_{n:1} * \cdots * g_{n:1})}_{\text{k-i times}}(t) - \underbrace{(f_{n:1} * \cdots * f_{n:1})}_{\text{i+1 times}} * \underbrace{(g_{n:1} * \cdots * g_{n:1})}_{\text{k-i-1 times}}(t) \\
&= (f_{n:1} * \cdots * f_{n:1} * g_{n:1} * \cdots * g_{n:1}) * (g_{n:1} - f_{n:1})(t).
\end{aligned}$$

Thus

$$\begin{aligned}
&\left|\int_{-\infty}^{\infty} H_n^i(t|x)(dG_{n:1}(x) - dF_{n:1}(x))\right| \\
&= \left|\int_{-\infty}^{t} (g_{n:1} * \cdots * g_{n:1} * f_{n:1} * \cdots * f_{n:1}) * (f_{n:1} - g_{n:1})(x)dx\right| \\
&\leq \|g_{n:1}\|_1^{i-1}\|f_{n:1}\|_1^{k-i-1}\left|\int_{-\infty}^{t} (g_{n:1} - f_{n:1})(x)dx\right| \\
&\leq \|g_{n:1}\|_1^{i-1}\|f_{n:1}\|_1^{k-i-1}|G_{n:1}(t) - F_{n:1}(t)|,
\end{aligned}$$

where $\|\ \|_1$ stands for the L_1- norm. Therefore Lemma 1 follows directly from A1. We also can derive A5 from assumptions for density $g_{n:1}$ such as

$$\int_{-\infty}^{\infty} g_{n:1}(t)d\lambda(t) < \infty \ \text{ and } \ \int_{-\infty}^{\infty} g_{n:1}^2(t)d\lambda(t) < \infty,$$

when the kernel h is of the form (2.13).

Now we come to discuss the existence of M.D. estimates. We start with defining a function $L_n(u)$ as follows:

$$L_n(u) = \int_{-\infty}^{\infty} S_n(t;u)g_n(t)d\lambda(t). \tag{2.15}$$

Then we note that $L_n(u)$ is a nondecreasing function of u, which crosses 0. Also we note that by Cauchy-Schwarz inequality,

$$M_n(\lambda;u) \geq L_n^2(u)/q, \tag{2.16}$$

where $q = \int_{-\infty}^{\infty} g_n^2(t)d\lambda(t)$. This implies that $M_n(\lambda; u)$ is bounded below by a nonnegative function which is nonincreasing on $(-\infty, u_0)$ and nondecreasing on (u_0, ∞). This fact guarantees that there exists a minimizer of $M_n(\lambda; u)$ for each n (cf. Koul and DeWet 1983).

In order to show the asymptotic normality of $\sqrt{n}(\hat{\xi}_n(\lambda) - \xi)$, we consider the case that $u - \xi = b$, where $\sqrt{n}|b| \leq B$ for some $0 < B < \infty$. With this notation, we rewrite $S_n(t; u)$ and $M_n(\lambda; u)$ as

$$S_n(t; b) = \frac{\sqrt{n}}{< n, k >} \sum_{j=1}^{<n,k>} [I(h_j \leq t + \xi + b) - F_n(t)] \tag{2.17}$$

and

$$M_n(\lambda; b) = \int_{-\infty}^{\infty} S_n^2(t; b)d\lambda(t). \tag{2.18}$$

Next we define a U-empirical process Y_n as

$$Y_n(t; b) = \frac{\sqrt{n}}{< n, k >} \sum_{j=1}^{<n,k>} [I(h_j \leq t + \xi + b) - G_n(t + b)]. \tag{2.19}$$

Then we can re-express (2.17) using the U-empirical process Y_n as

$$\begin{aligned} S_n(t; b) = (Y_n(t; b) - Y_n(t; 0)) + \sqrt{n}(G_n(t + b) - G_n(t) - bg_n(t)) \\ + \sqrt{n}bg_n(t) + Y_n(t; 0) + \sqrt{n}(G_n(t) - F_n(t)) \end{aligned} \tag{2.20}$$

to obtain the asymptotic quadratic form $\widetilde{M}_n(\lambda; b)$ in b of $M_n(\lambda; b)$, where

$$\widetilde{M}_n(\lambda; b) = \int_{-\infty}^{\infty} (\sqrt{n}bg_n(t) + Y_n(t; 0) + \sqrt{n}(G_n(t) - F_n(t)))^2 d\lambda(t) \tag{2.21}$$

For this purpose, we review some results for U-statistics and U-processes. The proofs are in, for example, Serfling(1980) and Shorack and Wellner(1986).

Lemma 3. For any $t \in \mathbf{R}$,[1] the variance of $Y_n(t; 0)$ is given by

$$Var(Y_n(t; 0)) = \frac{n}{< n, k >} \sum_{j=1}^{k} \binom{k}{j} \binom{n-k}{k-j} \zeta_j(t). \tag{2.22}$$

where $\zeta_j(t) = Cov(I(h(X_1, \dots, X_k) - \xi \le t), I(h(X_1, \dots, X_j, X_{k+1}, \dots, X_{2k-j}) - \xi \le t))$, and satisfies that

$$\mathrm{Var}(Y_n(t;0)) \le k\zeta_k(t) = kG_n(t)(1 - G_n(t)) \tag{2.23}$$

Lemma 4. The U-empirical process $Y_n(t;0)$ converges weakly to a Gaussian process $Y(t)$ with covariance function $\Gamma(s,t)$,

$$\Gamma(s,t) = k^2 \left[\int_{-\infty}^{\infty} H_n^0(s|x) H_n^0(t|x) dG_{n:1}(x) - \int_{-\infty}^{\infty} H_n^0(s|x) dG_{n:1}(x) \int_{-\infty}^{\infty} H_n^0(t|y) dG_{n:1}(y) \right]. \tag{2.24}$$

As a first step for proving the asymptotic normality, we state the following theorem which shows the asymptotic equivalence between $M_n(\lambda; b)$ and $\widetilde{M}_n(\lambda; b)$. The asymptotic quadratic form $\widetilde{M}_n(\lambda; b)$ in b is essential to derive the asymptotic normality of $\sqrt{n}(\hat{\xi}_n(\lambda) - \xi)$. The proof is postponed to Appendix A.

Theorem 1.

For any $0 < B < \infty$ with all the assumptions, we have

$$E \sup_{\{\sqrt{n}|b| \le B\}} |M_n(\lambda; b) - \widetilde{M}_n(\lambda; b)| = o(1) \tag{2.25}$$

The next theorem states that the two minimizers $\hat{\xi}_n(\lambda)$ for $\widehat{M}_n(\lambda; b)$ and $\tilde{\xi}_n(\lambda)$ for $\widetilde{M}_n(\lambda; b)$ have the same limiting distribution. Also the proof is delayed until Appendix B.

Theorem 2.

Let $\tilde{\xi}_n(\lambda)$ be a minimum solution of $\widetilde{M}_n(\lambda; b)$. Then with all the assumptions, we have

$$\sqrt{n}|\hat{\xi}_n(\lambda) - \tilde{\xi}_n(\lambda)| = o_p(1). \tag{2.26}$$

Theorem 2 tells us that

$$\sqrt{n}(\hat{\xi}_n(\lambda) - \xi) = -\frac{\int_{-\infty}^{\infty} g_n(t)\{Y_n(t;0) + \sqrt{n}(G_n(t) - F_n(t))\} d\lambda(t)}{\int_{-\infty}^{\infty} g_n^2(t) d\lambda(t)} + o_p(1). \tag{2.27}$$

Finally, we arrived at the following conclusion.

Theorem 3.

For each n, let

$$\frac{\sqrt{n}\int_{-\infty}^{\infty} g_n(t)(G_n(t) - F_n(t))d\lambda(t)}{\int_{-\infty}^{\infty} g_n^2(t)d\lambda(t)} = \mu_n(\lambda). \tag{2.28}$$

Then with the assumptions introduced up to now,

$$\frac{\sqrt{n}(\hat{\xi}_n(\lambda) - \xi - \mu_n(\lambda))}{\sqrt{\sigma_n^2}} \tag{2.29}$$

converges in distribution to a standard normal random variable, where

$$\sigma_n^2 = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_n(s)g_n(t)\Gamma(s,t)d\lambda(s)d\lambda(t)}{\left[\int_{-\infty}^{\infty} g_n^2(t)d\lambda(t)\right]^2} \tag{2.30}$$

We note that if $\mu_n(\lambda) \to \mu(\lambda)$, then $\mu(\lambda)$ is the asymptotic bias of the M.D. estimate $\hat{\xi}_n(\lambda)$ due to the mis-specified modeled d.f..

3. STUDY OF QUALITATIVE ROBUSTNESS FOR M.D. ESTIMATES

Here we discuss the qualitative robustness(cf. Hampel 1971) for the M.D. estimates against the gross error model. Thus we assume that $F_{n:1} = F_1$ for all n. This means that the modeled d.f. does not vary with n and so $F_{n:k} = F_k$ for all n. Also we denote $\bar{\mathbf{P}}_\mathbf{n}$ and $\bar{\mathbf{Q}}_\mathbf{n}$ as the probability measures according to F_k and $G_{n:k}$, respectively, which are constructed by $< n, k >$ numbers of random variables generated by the kernel h from a sample $X_1, X_2, \ldots, X_n$. To begin with, we state a definition of the qualitative robustness, which is a version due to Koul(1992).

Definition. A sequence of estimates $\{\hat{\xi}\}$ for ξ is said to be qualitatively robust at F_k against $\bar{\mathbf{Q}}_\mathbf{n}$ if it is consistent under $\bar{\mathbf{P}}_\mathbf{n}$ and under those $\bar{\mathbf{Q}}_\mathbf{n}$ that satisfy

$$D_n = \sup_t |F_k(t) - G_{n:k}(t)| \to 0 \qquad \text{as } n \to \infty. \tag{3.1}$$

Since we consider the gross error model, let $G_{n:1} = (1-\delta_n)F_1 + \delta_n G_1$, where G_1 is some continuous d.f. different from F_1 and $\delta_n = O(n^{-1/2})$. Then we note that

$$\sup_{t\in\mathbf{R}^1} |F_1(t) - G_{n:1}(t)| = \delta_n \sup_{t\in\mathbf{R}^1} |F_1(t) - G_1(t)| = O(n^{-1/2}).$$

We begin our task by showing that the transformations of random variables with respect to kernel h maintain the gross error model.

Lemma 5. For any given $G_{n:1} = (1-\delta_n)F_1 + \delta_n G_1$ and a kernel h, we can express $G_{n:k}$ as $G_{n:k} = (1-\delta_n^*)F_k + \delta_n^* G_k^*$, where G_k^* is a d.f., which is a linear combination of convolutions of F_1 and G_1 and $\delta_n^* = O(n^{-1/2})$.

Proof. Now we have for any $t \in \mathbf{R}^1$,

$$\begin{aligned} G_{n:k}(t) &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} I(h(x_1,x_2,\cdots,x_k) - \xi \le t)\prod_{i=1}^{k} dG_{n:1}(x_i) \\ &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} I(h(x_1,x_2,\ldots,x_k) - \xi \le t) \\ &\qquad \prod_{i=1}^{k} d((1-\delta_n)F_1(x_i) + \delta_n G_1(x_i)) \\ &= (1-\delta_n)^k F_k(t) + \sum_{j=0}^{k-1}\binom{k}{j}(1-\delta_n)^j \delta_n^{k-j} H^j(t), \end{aligned}$$

where $H^j(t) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} I(h(x_1,x_2,\ldots,x_k) - \xi \le t)\prod_{i=1}^{j} dF_1(x_i)$ $\prod_{i=j+1}^{k} dG_1(x_i) for j = 0,1,\ldots,k-1$. Let $\delta_n^* = 1-(1-\delta_n)^k$. Then it is easy to see that from the binomial theorem,

$$\delta_n^* = \sum_{j=0}^{k-1}\binom{k}{j}(1-\delta_n)^j\delta_n^{k-j}. \qquad (3.2)$$

This shows that $G_{n:k} = (1-\delta_n^*)F_k + \delta_n^* G_k^*$ and $\delta_n^* = O(n^{-1/2})$.

Now from (2.27), we have

$$Y_n(t;0) + \sqrt{n}(G_{n:k}(t) - F_k(t)) = \frac{\sqrt{n}}{<n,k>}\sum_{j=1}^{<n,k>}[I(h_j - \xi \le t) - F_k(t)]. \quad (3.3)$$

Since (3.3) converges weakly to a zero mean Gaussian process under $\bar{\mathbf{P}}_\mathbf{n}$, it is obvious that $\sqrt{n}(\hat{\xi}_n(\lambda) - \xi)$ converges in distribution to a zero mean normal random variable under $\bar{\mathbf{P}}_\mathbf{n}$. This implies that the sequence $\{\hat{\xi}_n(\lambda)\}$ is consistent under $\bar{\mathbf{P}}_\mathbf{n}$.

Next to show the consistency of the sequence $\{\hat{\xi}_n(\lambda)\}$ under $\bar{\mathbf{Q}}_\mathbf{n}$, it is enough to show that for each n, $\mu_n(\lambda)$, the bias is bounded under $\bar{\mathbf{Q}}_\mathbf{n}$. Straightforward calculations show that

$$\begin{aligned} & \sqrt{n}\left|\int_{-\infty}^{\infty} g_{n:k}(t)(G_{n:k}(t) - F_k(t))d\lambda(t)\right| \\ \leq & \sqrt{n}\delta_n \int_{-\infty}^{\infty} g_{n:k}(t)\left|\sum_{j=0}^{k}\int_{-\infty}^{\infty} H_n^i(t|x)(dG_1(t) - dF_1(t))\right|d\lambda(t) \end{aligned}$$

Then from assumptions A3 and A5 with the fact that $\delta_n = O(n^{-1/2})$, for each n, the bias is bounded. Thus the distribution of $\hat{\xi}_n(\lambda)$ under $\bar{\mathbf{Q}}_\mathbf{n}$ converges weakly to a degenerate distribution, degenerate at ξ. This shows that the sequence $\{\hat{\xi}_n(\lambda)\}$ of M.D. estimates is qualitatively robust against the gross error model. Especially, when the integrating measure, λ is finite, then any M.D. estimates are qualitatively robust (cf. Koul 1992).

4. AN EXAMPLE

In this section, we show an example by taking Lebesque measure for the integrating measure λ. We assume that G is symmetric about ξ. We note that ξ is the mean of G if it exists. Then X_1 is a symmetric kernel for ξ and $\bar{X}_n$, the corresponding U-statistic. It can be easily shown that $\text{med}\{(X_i + X_j)/2, 1 \leq i, j \leq n\}$ is an M.D. estimate for ξ(cf. Koul 1992).

Now we can extend this idea to the more general case. We consider the generalized kernel $(X_1 + X_2 + \ldots + X_k)/k$ for ξ. Since the distribution of $(X_1 + X_2 + \cdots + X_k)/k$ is also symmetric about ξ when the distribution of X_i is symmetric about ξ, an M.D. estimate for ξ would be of the form

$$\text{med}\left[\frac{(X_{i_1} + X_{i_2} + \cdots + X_{i_k}) + (X_{j_1} + X_{j_2} + \cdots + X_{j_k})}{2k}\right] \tag{4.1}$$

from $< n, k >^2$ number of random variables. We note that (4.1) is the generalized L -estimate in the sense of Akritas(1986).

APPENDIX

Appendix A. In this appendix, we prove Theorem 1 with the following lemmas.

Lemma 6. Lemmas 2 and 3 imply that

$$E\left\{\int_{-\infty}^{\infty} Y_n^2(t;0)d\lambda(t)\right\} < \infty. \tag{A.1}$$

Proof. From Fubini's theorem and Lemma 3, we have

$$E\left\{\int_{-\infty}^{\infty} Y_n^2(t;0)d\lambda(t)\right\} \le k\int_{-\infty}^{\infty} \zeta_k(t)d\lambda(t) = k\int_{-\infty}^{\infty} G_n(t)(1-G_n(t))d\lambda(t)$$

Invoking Lemma 2, we see the result.

Lemma 7. Lemma 3 with A5 implies that for every $0 < B < \infty$ with $\sqrt{n}\,|b\,| \le B$,

$$E\left\{\int_{-\infty}^{\infty} (Y_n(t;b) - Y_n(t;0))^2 d\lambda(t)\right\} = o(1). \tag{A.2}$$

Proof. From Fubini's theorem, Lemma 3 and mean value theorem, we have

$$\begin{aligned} E\left\{\int_{-\infty}^{\infty} (Y_n(t;b) - Y_n(t;0))^2 d\lambda(t)\right\} &\le k\int_{-\infty}^{\infty} |G_n(t+b) - G_n(t)|d\lambda(t) \\ &= k\int_{-\infty}^{\infty} |b|g_n(t^*)d\lambda(t), \end{aligned}$$

for some t^*, where t^* is a number between t and $t+b$. Thus the result follows from A5 with the fact that $b = O(n^{-1/2})$.

Lemma 8. From Lemma 7, for every $0 < B < \infty$, we have

$$E\left\{\sup_{\{\sqrt{n}|b|\le B\}} \int_{-\infty}^{\infty} (Y_n(t;b) - Y_n(t;0))^2 d\lambda(t)\right\} = o(1). \tag{A.3}$$

Proof. First of all, we consider a partition of $[-B, B]$ in the following manner. $-B = r_0 < r_1 < \ldots < r_m = B$ and as $n \to \infty$, $\max_j(r_j - r_{j-1}) \to 0$. Then for any $r_{j-1}/\sqrt{n} \le b \le r_j/\sqrt{n}$ with the facts that $a \le b \le c$ implies $b^2 = a^2 + c^2$ and $(a+b)^2 \le 2a^2 + 2b^2$,

$$
\begin{aligned}
&\int_{-\infty}^{\infty} (Y_n(t;b) - Y_n(t;0))^2 d\lambda(t) \\
\le\ & \int_{-\infty}^{\infty} (Y_n(t;r_{j-1}/\sqrt{n}) - Y_n(t;0) + G_n(t + r_{j-1}/\sqrt{n}) - G_n(t+b))^2 d\lambda(t) \\
&+ \int_{-\infty}^{\infty} (Y_n(t;r_j/\sqrt{n}) - Y_n(t;0) + G_n(t + r_j/\sqrt{n}) - G_n(t+b))^2 d\lambda(t) \\
\le\ & 2\int_{-\infty}^{\infty} (Y_n(t;r_{j-1}/\sqrt{n}) - Y_n(t;0))^2 d\lambda(t) \\
&+2\int_{-\infty}^{\infty} (Y_n(t;r_j/\sqrt{n}) - Y_n(t;0))^2 d\lambda(t) \\
&+4\int_{-\infty}^{\infty} (G_n(t + r_j/\sqrt{n}) - G_n(t + r_{j-1}/\sqrt{n}))^2 d\lambda(t).
\end{aligned}
$$

Then we have that

$$
\begin{aligned}
&\sup_{\{\sqrt{n}|b| \le B\}} \int_{-\infty}^{\infty} (Y_n(t;b) - Y_n(t;0))^2 d\lambda(t) \\
\le\ & 4\sum_{j=0}^{m} \int_{-\infty}^{\infty} (Y_n(t;r_j/\sqrt{n}) - Y_n(t;0))^2 d\lambda(t) \\
&+4\int_{-\infty}^{\infty} \max_j (G_n(t + r_j/\sqrt{n}) - G_n(t + r_{j-1}/\sqrt{n}))^2 d\lambda(t).
\end{aligned}
$$

Therefore by letting first $n \to \infty$ and then $m \to \infty$ with Lemma 7, we see the result.

Lemma 9. For every $0 < B < \infty$ with $\sqrt{n}|b| \le B$ and with A5, we have

$$
n\int_{-\infty}^{\infty} (G_n(t+b) - G_n(t) - bg_n(t))^2 d\lambda(t) = o(1). \qquad (A.4)
$$

Proof. From mean value theorem, there is a t^* between t and $t+b$ such as

$$G_n(t+b) - G_n(t) = bg_n(t^*).$$

Then we have

$$n\int_{-\infty}^{\infty}(G_n(t+b) - G_n(t) - bg_n(t))^2 d\lambda(t) = nb^2 \int_{-\infty}^{\infty}(g_n(t^*) - g_n(t))^2 d\lambda(t).$$

Thus by A5, we have the result.

Lemma 10. From Lemma 9, for every $0 < B < \infty$,

$$\sup_{\{\sqrt{n}|b|\leq B\}} n\int_{-\infty}^{\infty}(G_n(t+b) - G_n(t) - bg_n(t))^2 d\lambda(t) = o(1). \qquad (A.5)$$

Proof. With the same partition and arguments used in Lemma 8, we have

$$\begin{aligned} & n\int_{-\infty}^{\infty}(G_n(t+b) - G_n(t) - bg_n(t))^2 d\lambda(t) \\ \leq\ & 2n\int_{-\infty}^{\infty}\left(G_n(t+r_{j-1}/\sqrt{n}) - G_n(t) - r_{j-1}/\sqrt{n}g_n(t)\right)^2 d\lambda(t) \\ & +2n\int_{-\infty}^{\infty}\left(G_n(t+r_j/\sqrt{n}) - G_n(t) - r_j/\sqrt{n}g_n(t)\right)^2 d\lambda(t) \\ & +4(r_j - r_{j-1})^2 \int_{-\infty}^{\infty} g_n^2(t) d\lambda(t). \end{aligned}$$

Thus we have that

$$\begin{aligned} & \sup_{\{\sqrt{n}|b|\leq B\}} n\int_{-\infty}^{\infty}(G_n(t+b) - G_n(t) - bg_n(t))^2 d\lambda(t) \\ \leq\ & 4n\sum_{j=0}^{m}\int_{-\infty}^{\infty}\left(G_n\left(t+r_j/\sqrt{n}\right) - G_n(t) - r_j/\sqrt{n}g_n(t)\right)^2 d\lambda(t) \\ & +4\max(r_j - r_{j-1})^2 \int_{-\infty}^{\infty} g_n^2(t) d\lambda(t). \end{aligned}$$

Then by first letting $n \to \infty$ and then $m \to \infty$ with Lemma 9, we see the result.

Lemma 11. For every $0 < B < \infty$, with A5, we have

$$E\{\sup_{\{\sqrt{n}|b|\le B\}} \sqrt{n}\int_{-\infty}^{\infty} |(Y_n(t;b) - Y_n(t;0)) \\ (G_n(t+b) - G_n(t) - bg_n(t)))|d\lambda(t)\} = o(1). \quad (A.6)$$

Proof. From Cauchy-Schwarz inequality, we have that

$$\begin{aligned} &\sqrt{n}\int_{-\infty}^{\infty} |(Y_n(t;b) - Y_n(t;0))(G_n(t+b) - G_n(t) - bg_n(t))|d\lambda(t) \\ \le\ & \sqrt{n}\left\{\int_{-\infty}^{\infty}(Y_n(t;b) - Y_n(t;0))^2 d\lambda(t)\right\}^{1/2} \\ & \left\{\int_{-\infty}^{\infty}(G_n(t+b) - G_n(t) - bg_n(t))^2 d\lambda(t)\right\}^{1/2}. \end{aligned}$$

Thus from Lemmas 8 and 10, the result follows.

Lemma 12. For every $0 < B < \infty$, we have

$$E\left\{\sup_{\{\sqrt{n}|b|\le B\}} \int_{-\infty}^{\infty} |(\sqrt{n}bg_n(t) + \sqrt{n}(G_n(t+b) - G_n(t)) + Y_n(t;0)) \\ (Y_n(t;b) - Y_n(t;0))\,|d\lambda(t)\right\} = o(1). \quad (A.7)$$

Proof. This also follows from Cauchy-Schwarz inequality and Lemmas 6, 8 and 10.

Lemma 13. For every $0 < B < \infty$, we have

$$E\left\{\sup_{\{\sqrt{n}|b|\le B\}} \sqrt{n}\int_{-\infty}^{\infty} |(\sqrt{n}bg_n(t) + \sqrt{n}(G_n(t+b) - G_n(t)) + Y_n(t;0)) \\ (G_n(t+b) - G_n(t) - bg_n(t))\,|d\lambda(t)\right\} = o(1). \quad (A.8)$$

Proof. We can prove this with the same arguments used for Lemma 12.

Proof Of Theorem 1. First of all, we note that

$$
\begin{aligned}
&|M_n(\lambda;b) - \widetilde{M}_n(\lambda;b)| \\
\leq\ & \int_{-\infty}^{\infty} (Y_n(t;b) - Y_n(t;0))^2\, d\lambda(t) \\
&+n \int_{-\infty}^{\infty} (G_n(t+b) - G_n(t) - bg_n(t)))^2\, d\lambda(t) \\
&+2\sqrt{n} \int_{-\infty}^{\infty} |\,(Y_n(t;b) - Y_n(t;0))\,(G_n(t+b) - G_n(t) - bg_n(t)))\,|d\lambda(t) \\
&+2 \int_{-\infty}^{\infty} |(\sqrt{n}bg_n(t) + \sqrt{n}(G_n(t+b) - G_n(t)) + Y_n(t;0)) \\
&\qquad\qquad\qquad (Y_n(t;b) - Y_n(t;0))\,|d\lambda(t) \\
&+2\sqrt{n} \int_{-\infty}^{\infty} |(\sqrt{n}bg_n(t) + \sqrt{n}(G_n(t+b) - G_n(t)) + Y_n(t;0)) \\
&\qquad\qquad\qquad (G_n(t+b) - G_n(t) - bg_n(t))\,|d\lambda(t).
\end{aligned}
$$

Thus the proof is completed by applying Lemmas 6 to 13.

Appendix B. We begin with the following result for the proof of Theorem2.

Lemma 14. For every $\epsilon > 0$, $0 < z < \infty$, there exists an n_ϵ and a $B(\epsilon, z)$ such that

$$P\{\inf_{\{\sqrt{n}|b|>B(\epsilon,z)\}} M_n(\lambda;b) > z\} \geq 1-\epsilon, \quad \text{for all } n \geq n_\epsilon \qquad (B.1)$$

and

$$P\{\inf_{\{\sqrt{n}|b|>B(\epsilon,z)\}} \widetilde{M}_n(\lambda;b) > z\} \geq 1-\epsilon, \quad \text{for all } n \geq n_\epsilon. \qquad (B.2)$$

Proof. Form (2.16), we have

$$P\{\inf_{\{\sqrt{n}|b|>B(\epsilon,z)\}} M_n(\lambda;b) \geq z\} \geq P\{\inf_{\{\sqrt{n}|b|>B(\epsilon,z)\}} L_n^2(b) \geq zq\},$$

where q was defined in section 2. Also we define a function

$$\widetilde{L}_n(b) = \int_{-\infty}^{\infty} (\sqrt{n}bg_n(t) + Y_n(t;0) + \sqrt{n}(G_n(t) - F_n(t)))g_n(t)d\lambda(t).$$

Then by the same reasons used for Theorem 1,

$$E \sup_{\{\sqrt{n}|b| \le B\}} |L_n(b) - \widetilde{L}_n(b)| = o(1). \tag{B.3}$$

Also we define

$$Q = \int_{-\infty}^{\infty} (Y_n(t;0) + \sqrt{n}(G_n(t) - F_n(t)))g_n(t)d\lambda(t).$$

Then we note that from Lemma 4,

$$E(Q) = \sqrt{n}\int_{-\infty}^{\infty} (G_n(t) - F_n(t))g_n(t)d\lambda(t) = O(1)$$

and

$$V(Q) \le kq \int_{-\infty}^{\infty} G_n(t)(1 - G_n(t))d\lambda(t).$$

Thus for any $\epsilon > 0$, there exists $n_{1\epsilon}$ and K_ϵ such that

$$P\{|Q| \le K_\epsilon\} \ge 1 - \epsilon/2 \quad \text{for all} \quad n \ge n_{1\epsilon}.$$

Let B satisfy that

$$B > (K_\epsilon + \sqrt{zq})/q.$$

Thus

$$\begin{aligned} P\{\inf_{\{\sqrt{n}|b|=B\}} \widetilde{L}_n^2(b) \ge zq\} &= P\{\inf_{\{\sqrt{n}|b|=B\}} |\widetilde{L}_n(b)| \ge \sqrt{zq}\} \\ &= P\{\inf_{\{\sqrt{n}|b|=B\}} |Q + \sqrt{n}bq| \ge \sqrt{zq}\} \\ &\ge P\{|Q + Bq| \ge \sqrt{zq}\} \\ &\ge P\{|Q + K_\epsilon + \sqrt{zq}| \ge \sqrt{zq}\} \\ &\ge P\{|Q| \le K_\epsilon\} \ge 1 - \epsilon/2. \end{aligned}$$

Thus from (B.3), for every $\epsilon > 0$, there exists $n_{2\epsilon}$ such that for all $n \ge n_{2\epsilon}$,

$$P\{\inf_{\{\sqrt{n}|b|=B\}} L_n^2(b) \ge zq\} \ge P\{\inf_{\{\sqrt{n}|b|=B\}} \tilde{L}_n^2(b) \ge zq\} - \epsilon/2. \tag{B.4}$$

Thus we choose $n_\epsilon = \max(n_{1\epsilon}, n_{2\epsilon})$ and use the monotonicity of $L_n(b)$ in b together with (B.4) to see (B.1). Also we can prove (B.2) similarly.

Proof of Theorem 2. From Lemma 14, we obtain that

$$\sqrt{n}|\hat{\xi}_n - \xi| = O_p(1) \quad \text{and} \quad \sqrt{n}|\tilde{\xi}_n - \xi| = O_p(1).$$

Thus from Theorem1, we have that

$$|\widetilde{M}_n(\lambda;\tilde{\xi}_n) - M_n(\lambda;\tilde{\xi}_n)| = o_p(1) \quad \text{and} \quad |M_n(\lambda;\hat{\xi}_n) - \widetilde{M}_n(\lambda;\hat{\xi}_n)| = o_p(1).$$

Therefore by applying the triangle inequality, we have that

$$|\widetilde{M}_n(\lambda;\tilde{\xi}_n) - \widetilde{M}_n(\lambda;\hat{\xi}_n)| = o_p(1).$$

Therefore since

$$\begin{aligned} & \widetilde{M}_n(\lambda;\tilde{\xi}_n) - \widetilde{M}_n(\lambda;\hat{\xi}_n) \\ = \ & n(\tilde{\xi}_n - \hat{\xi}_n)(\tilde{\xi}_n + \hat{\xi}_n) \int_{-\infty}^{\infty} g_n^2(t) d\lambda(t) \\ & +2\sqrt{n}(\tilde{\xi}_n - \hat{\xi}_n) \int_{-\infty}^{\infty} g_n(t) \left(Y_n(t;0) + \sqrt{n}(G_n(t) - F_n(t))\right) d\lambda(t), \end{aligned}$$

we obtain the result from A5 and Lemmas 1 and 6.

REFERENCES

(1) Akritas, M. G. (1986). Empirical processes associated with V-statistics and a class of estimators under random censoring. *The Annals of Statistics,* 14, 619-637.

(2) Boos, D. D. (1981). Minimum distance estimators for location and goodness of fit. *Journal of the American Statistical Association,* 76, 663-670.

(3) Dhar, S. K. (1991). Minimum distance estimation in an additive effects outliers model. *The Annals of Statistics,* 19, 205-228.

(4) Hampel, F. R. (1971). A general qualitative definition of robustness. *The Annals of Mathematical Statistics,* 42, 1887-1896.

(5) Koul, H. L. (1992). *Weighted empiricals and linear models.* IMS Lecture notes- monograph series. vol. 21.

(6) Koul, H. L. and T. DeWet (1983). Minimum distance estimation in a linear regression model. *The Annals of Statistics,* 11, 921-932.

(7) Parr, W. (1981). Minimum distance estimation: a bibliography. *Communications in Statistics-Theory and Methods,* A10(12), 1205-1224.

(8) Serfling, R. J. (1980). *Approximation theorems of mathematical statistics.* Wiley & Sons, Inc.

(9) Serfling, R. J. (1984). Generalized L-, M- and R-statistics. *The Annals of Statistics,* 12, 76-86.

(10) Shorack, G. R. and J. A. Wellner (1986). *Empirical processes with applications to statistics.* Wiley & Sons, Inc.