⟨Research Paper⟩

Journal of the Korean
Statistical Society
Vol. 18, No. 2, 1989

Asymptotically Optimal Estimators of the Differences of Two Regression Parameters+

Byeong U. Park*, Woo C. Kim* and Moon S. Song*

ABSTRACT

We consider two semiparametric regression lines where the density of the error terms are unknown. We give simultaneous estimators of the differences of intercepts and slopes which turn out to be asymptotically minimax as well as efficient in semiparametric sense.

1. Introduction

Consider two simple linear regression models;

$$\begin{aligned} Y_1 &= \alpha_1 + \beta_1 X_1 + \varepsilon_1 \\ Y_2 &= \alpha_2 + \beta_2 X_2 + \varepsilon_2 \end{aligned} \qquad (1.1)$$

where α_1, α_2, β_1 and β_2 are regression parameters, X_1 and X_2 are independent random covariates with a common density h and ε_1 and ε_2 are independent random errors with a common density g. Throughout this paper errors are assumed to be independent of covariates. This is just a stochastic version of the usual simple linear regression models,

$$\begin{aligned} Y_{1j} &= \alpha_1 + \beta_1 X_{1j} + \varepsilon_{1j} \\ Y_{2j} &= \alpha_2 + \beta_2 X_{2j} + \varepsilon_{2j} \end{aligned} \qquad (1.2)$$

+ The research was supported by the Ministry of Education, through the Research Institute for Basic Sciences, SNU, 1988−1989.

* Department of Computer Science and Statistics, Seoul National University, Seoul 151−742, Korea.

where X_{1j} and X_{2j} are known constants.

Our interest here is in the estimation of $(\alpha_2-\alpha_1)$ and $(\beta_2-\beta_1)$. The classical approach for this problem is the least squares method which has been shown to be optimal if the errors are i.i.d. normal. But it is well-known that least squares estimates suffer from being sensitive to outlying observations, which often come from long-tailed error distributions, hence losing efficiency if the actual error distribution has long-tail.

The precise error distributions are not known in almost all the cases. Hence it can not be said that an estimator, optimal in an assumed error distribution, is optimal actually. One may release the assumptions on the error distributions and try to find the best possible estimators in those situations. The resulting models are called "semiparametric models" where the error distributions are set to be totally unknown. The first thing one has to do is to find what is the best one can do, mostly in asymptotic sense. Asymptotic efficiency and minimaxity have been used widely in large sample theory as criteria for identifying good estimators. Begun et al. (1983) characterizes the estimators achieving these two optimalities through the representation theorem and the asymptotic minimax theorem in semiparametric sense. Their work follows Hájek (1970, 1972)'s parametric settings and Le Cam (1972)'s general settings. Briefly speaking, estimators which achieve the information bounds and asymptotic minimax bounds of the least favorable regular parametric submodel are shown to be asymptotically efficient and minimax respectively.

The construction of asymptotically minimax and efficient estimators of $(\alpha_2-\alpha_1)$ and $(\beta_2-\beta_1)$ in the model (1.1) with g unknown is our present theme. A simple two-sample generalization of Example 2 in Bickel (1982) and Example 1 in Schick(1987) turns out to yield asymptotically efficient estimators although Bickel's estimator uses only a part of the sample to estimate the unknown density g. Hence the main point of this paper should be addressed on the construction of asymptotically minimax estimators but it turns out that our estimators are asymptotically efficient as well.

In section 2 we will find the information and minimax bounds for our present regression models. In section 3, the estimators are constructed and the proof of the Theorem 3.1 is deferred to section 4.

2. Assumptions, Information and minimax bounds

The model (1.1) can be reparametrized as follows.

$$\begin{aligned} Y_1 &= \alpha+\beta X_1+\varepsilon_1 \\ Y_2 &= (\alpha+\Delta)+(\beta+\delta)X_2+\varepsilon_2. \end{aligned} \tag{2.1}$$

The common density h of covariates is assumed to be known and have a finite second moment. The minimal condition on g is $\int (\dot{g}^2/g)(y)\,dy < \infty$ where $\dot{g}$ is the first derivative of g. The parameters of interest are (Δ,δ) and the nuisance ones are (α,β,g). We observe $(X_1,Y_1,Z_1),\cdots,(X_n,Y_n,Z_n)$ where $Z_i=0$ if (X_i,Y_i) comes from the first sample and $Z_i=1$ otherwise. If we assume that the proportion of $Z=0$ is p $(0<p<1)$, then the density of (X_i,Y_i,Z_i) is

$$f(x,y,z,\theta,g)=\begin{cases} f(x,y,0,\theta,g)=pg(y-\alpha-\beta x)h(x) \\ f(x,y,1,\theta,g)=qg(y-\alpha-\Delta-(\beta+\delta)x)h(x) \end{cases} \tag{2.2}$$

where $\theta=(\alpha,\beta,\Delta,\delta)$ and $q=1-p$.

The information for estimating parameters of interest in semiparametric models is $I_*=\int \ell^*\ell^{*T}dP$ where P is a probability measure having the true density f and ℓ^* is the component of the score function for the parameters of interest which is orthogonal to all nuisance parameter scores. For further details see Begun et al.(1983) or Bickel et al. (1989). For the model (2.2) the score function for (Δ,δ) is

$$\dot{\ell}_2=\begin{bmatrix}(\partial/\partial\Delta)\ell \\ (\partial/\partial\delta)\ell\end{bmatrix}=\begin{bmatrix}-z(\dot{g}/g)(y-\alpha-\Delta-(\beta+\delta)x) \\ -xz(\dot{g}/g)(y-\alpha-\Delta-(\beta+\delta)x)\end{bmatrix}$$

where $\ell=\ell(x,y,z,\theta,g)=\log f(x,y,z,\theta,g)$. We first obtain the orthogonal component of $\dot{\ell}_2$ to the scores of the nuisance parameters (α,β). Observing that the score function of (α,β) is

$$\dot{\ell}_1=\begin{bmatrix}(\partial/\partial\alpha)\ell \\ (\partial/\partial\beta)\ell\end{bmatrix}=\begin{bmatrix}-z(\dot{g}/g)(y-\alpha-\Delta-(\beta+\delta)x)-(1-z)(\dot{g}/g)(y-\alpha-\beta x) \\ -xz(\dot{g}/g)(y-\alpha-\Delta-(\beta+\delta)x)-x(1-z)(\dot{g}/g)(y-\alpha-\beta x)\end{bmatrix},$$

we see that it is

$$\dot{\ell}_2-\pi(\dot{\ell}_2\mid[\dot{\ell}_1])=\dot{\ell}_2-q\dot{\ell}_1=\begin{bmatrix}q(1-z)(\dot{g}/g)(y-\alpha-\beta x)-pz(\dot{g}/g)(y-\alpha-\Delta-(\beta+\delta)x) \\ qx(1-z)(\dot{g}/g)(y-\alpha-\beta x)-pxz(\dot{g}/g)(y-\alpha-\Delta-(\beta+\alpha)x)\end{bmatrix} \tag{2.3}$$

where $\pi(\cdot \mid M)$ is the projection operator on M in $L^2(\mu)$ (μ: Lebesgue measure) and $[\dot{\ell}_1]$ is the linear span of $\dot{\ell}_1$. The space of the scores for g, which is called "tangent space", is given by $\{zt(y-\alpha-\Delta-(\beta+\delta)x)+(1-z)t(y-\alpha-\beta x) \mid \int t(x)g(x)dx=0\}$. See Bickel et al. (1989) for the precise definition of tangent space. The function given in (2.3) is already perpendicular to this tangent space. Hence it is the "efficient score function" ℓ^* for estimating (Δ,δ). This means that we can estimate (Δ,δ) asymptotically as well not knowing g as knowing g, so called "adaptation is possible".

In the sense of Begun et al. (1983) and Bickel et al. (1989) an estimator $\hat{\rho}_n=(\hat{\Delta}_n,\hat{\delta}_n)$ is called "asymptotically efficient" if for every sequence (θ_n,g_n) such that $|n^{1/2}(\theta_n-\theta)-s| \to 0$ and $\| n^{1/2}(g_n^{1/2}-g^{1/2})-t \|_\mu \to 0$ as $n\to\infty$ for some $s\in R^4$ and $t\in L^2(\mu)$

$$n^{1/2}(\hat{\rho}_n-\rho_n) \Rightarrow N(0, I_*^{-1}) \qquad (2.4)$$

as $n\to\infty$ under the probability measure P_{θ_n,g_n} having a density $f(\cdot,\theta_n,g_n)$ where $\rho_n=(\Delta_n,\delta_n)$, $\|\cdot\|_\mu$ is the usual L^2 norm with respect to Lebesgue measure μ and

$$I_*=\int \ell^*\ell^{*T}dP=pqI_g\begin{bmatrix}1 & EX\\ EX & EX^2\end{bmatrix}$$

where $I_g=\int(\dot{g}^2/g)(y)dy$. And an estimator is called "asymptotically minimax" if

$$\lim_{\substack{c_1\to\infty\\ c_2\to\infty}} \lim_n \sup_{\substack{|s|\le c_1\\ \|t\|_\mu\le c_2}} E_{P_{\theta_n,g_n}}\ell(n^{1/2}(\hat{\rho}_n-\rho_n))=E\ell(Z_*) \qquad (2.5)$$

where ℓ is a subconvex loss function (see Begun et al. (1983) for definition) and $Z_*\sim N(0, I_*^{-1})$. If ℓ is restricted to be bounded it suffices to show that

$$n^{1/2}(\hat{\rho}_n-\rho_n) \Rightarrow N(0, I_*^{-1}) \qquad (2.6)$$

under P_{θ_n,g_n} for any sequence (θ_n,g_n) such that $n^{1/2}|\theta_n-\theta| \le M_1$ and $n^{1/2}\|g_n^{1/2}-g^{1/2}\|_\mu\le M_2$ for some $M_1,M_2>0$. A simple argument for this is presented in Beran (1981) and Millar (1984). Note that (2.6) simply implies (2.4). So an estimator $\hat{\rho}_n$ satisfying (2.6) is asymptotically efficient as well as asymptotically minimax. In next section such an estimator will be constructed.

3. Main Results

We now construct an estimator which satisfies (2.6). The idea is to use LeCam's one-step MLE as in the classical estimation of the location problem:

(a) Find a good estimate $\tilde{\theta}_n$ of $\theta = (\rho, n)$ where $\rho = (\Delta, \delta)$ and $n = (\alpha, \beta)$.

(b) Construct a suitable estimate $\hat{\ell}^*(x,y,z,\theta;\ X_1,Y_1,Z_1,\cdots,X_n,Y_n,Z_n)$ of $\ell^*(x,y,z,\theta,g)$ and form

$$\hat{\rho}_n = \bar{\rho}_n + n^{-1}\hat{I}_*^{-1} \sum_{i=1}^{n} \hat{\ell}^*(X_i,Y_i,Z_i,\bar{\theta}_n;\ X_1,Y_1,Z_1,\cdots,X_n,Y_n,Z_n) \tag{3.1}$$

as the proposed estimator where $\hat{I}_* = n^{-1} \sum_{j=1}^{n} \hat{\ell}^*\hat{\ell}^{*T}(X_j,Y_j,Z_j,\theta_n;\ X_1,Y_1,Z_1,\cdots,X_n,Y_n,Z_n)$ and $\bar{\theta}_n = (\bar{\rho}_n, \bar{\eta}_n)$ is a discretized version of $\tilde{\theta}_n$ (see Bickel (1982) or Appendix in Le Cam (1960) for explanation of discretization). For simplicity, we will suppress $X_1,Y_1,Z_1,\cdots,X_n,Y_n,Z_n$ in the expression of $\hat{\ell}^*$ from now on.

For (2.6) it suffices to show that

(A) $n^{1/2}(\tilde{\theta}_n - \theta)$ is tight under the probability measure P_{θ_n,g_n},

(B) $n^{1/2} \int \ell^*(x,y,z,\theta_n)\ f(x,y,z,\theta_n,g)\ dxdy = O_{P_{\theta_n,g_n}}(1)$,

(C) $E_{P_{\theta_n,g_n}} \int \left| \hat{\ell}^*(x,y,z,\theta_n) - \ell^*(x,y,z,\theta_n) \right|^2 f(x,y,z,\theta_n,g)\ dxdy = o(1)$,

(D) $n^{1/2} \left| \hat{\ell}_j^*(x,y,z,\theta) - \hat{\ell}^*(x,y,z,\theta) \right| \leq L_n$,

(E) $n^{-1/2} \left| (\partial / \partial\theta)\hat{\ell}^*(x,y,z,\theta) \right| \leq M_n$,

(F) $n^{-1/4} \left| \hat{\ell}^*(x,y,z,\theta) \right| \leq N_n$

where P_{θ_n,g_n} is specified just below (2.6), $\hat{\ell}_j^*$ is a cross-validated estimator of ℓ^*, i.e., $\hat{\ell}_j^* = \hat{\ell}^*$ computed from the sample deleting (X_j,Y_j,Z_j) and L_n, M_n, N_n are constants converging to zero. The reason for this is given in Park (1988).

Now note that even if we are interested only in $\rho = (\Delta, \delta)$, we need to estimate $\eta = (\alpha, \beta)$ as we can see in (3.1). But obviously η is unidentifiable since a change in α could be hidden by the same location change in g. However, without loss of generality, we can restrict g to the densities satisfying

$$\int \Psi(y) g(y)\ dy = 0 \tag{3.2}$$

for bounded and antisymmetric Ψ with a bounded positive first derivative. For otherwise there exists the unique constant $a(\neq 0)$ such that $\int \Psi(y-a)g(y)dy = 0$ and changing α to $\alpha+a$ and $g(\cdot)$ to $g_a(\cdot) = g(\cdot + a)$ does not alter the density f in (2.2) and satisfies (3.2) with g_a. With g satisfying (3.2), η is identifiable.

PRELIMINARTY ESTIMATOR. Let Ψ be defined as above. Let $\tilde{\eta}_n$ be the M-estimator corresponding to Ψ, i.e., the unique solution of

$$\sum_{\{j:Z_j=0\}} \psi(Y_j-\tilde{\alpha}_n-\tilde{\beta}_n X_j) = 0$$
$$\sum_{\{j:Z_j=0\}} X_j\psi(Y_j-\tilde{\alpha}_n-\tilde{\beta}_n X_j) = 0 \tag{3.3}$$

Then $n_1^{-1/2} \sum_{\{j:Z_j=0\}} \psi(Y_j-\alpha_n-\beta_n X_j)$ and $n_1^{-1/2} \sum_{\{j:Z_j=0\}} X_j\psi(Y_j-\alpha_n-\beta_n X_j)$ converge to normal distributions under P_{θ_n, g_n} (use triangular array Central Limit Theorem) where $n_1 = \#(Z_j = 0)$. Furthermore $\int (\Psi')^2(x)g_n(x)dx$ is asymptotically bounded since

$$\left| \int (\Psi')^2(x)[g_n(x)-g(x)]dx \right| \leq M \| g_n^{1/2} - g^{1/2} \|_\infty$$

for some $M>0$. Now, following the arguments appeared in pp. 805-6 of Huber(1973), we can conclude that $n^{1/2}|\tilde{\eta}_n-\eta_n| = O_{P_{\theta_n,g_n}}(1)$ and together with this and $n^{1/2}|\eta_n-\eta| = O(1)$ we obtain that $n^{1/2}|\tilde{\eta}_n-\eta|$ is tight under P_{θ_n,g_n}. Similar arguments can be applied to the estimator of $(\alpha+\Delta, \beta+\delta)$ defined by the same way as in (3.3) and taking differences between these two estimators yields a $\sqrt{n}$-consistent estimator ρ_n of $\rho = (\Delta, \delta)$ under the probability measure P_{θ_n,g_n}.

OPTIMAL ESTIMATOR. Now we will construct a suitable estimator of ℓ^*. Let K be a logistic density function, i.e., $K(w) = e^{-w}(1+e^{-w})^{-2}$. We estimate the unknown density g by $\hat{g}(\cdot, \bar{\theta}_n)$ where

$$\hat{g}(y,\theta) = b_n + n^{-1}b_n^{-1} \sum_{i=1}^{n} \{Z_i K\{b_n^{-1}(y-Y_i+(\alpha+\Delta)+(\beta+\delta)X_i)\} I(|X_i| \leq c_n)$$
$$+(1-Z_i)K(b_n^{-1}(y-Y_i+\alpha+\beta X_i)\} I(|X_i| \leq c_n)\},$$

$b_n \to 0$ and $c_n \to \infty$ at some rates to be specified later. Define

$$\hat{h}(y,\theta) = (\partial/\partial y)\hat{g}(y,\theta) / \hat{g}(y,\theta).$$

The estimator of $\ell^*(x,y,z,\theta,g)$ is defined by

$$\hat{\ell}^*(x,y,z,\theta)=\begin{bmatrix} q(1-z)\hat{h}(y-\alpha-\beta x,\theta)-pz\hat{h}(y-\alpha-\Delta-(\beta+\delta)x,\theta) \\ \\ q(1-z)x\hat{h}(y-\alpha-\beta x,\theta)I(|x|\leq c_n)-pzx\hat{h}(y-\alpha-\Delta-(\beta+\delta)x,\theta)I(|x|\leq c_n) \end{bmatrix} \quad (3.4)$$

and the asymptotically optimal (efficient and minimax) estimator is given by

$$\hat{\rho}_n=\rho_n+n^{-1}\hat{I}^{-1}\sum_{i=1}^{n}\hat{\ell}^*(X_i,Y_i,Z_i,\bar{\theta}_n)$$

where $\hat{I}_*=n^{-1}\sum_{i=1}^{n}\hat{\ell}^*\hat{\ell}^{*T}(X_i,Y_i,Z_i,\bar{\theta}_n)$. Let $\{b_n\}$ and $\{c_n\}$ be such that

$$nb_n^6c_n^{-4}\rightarrow\infty.$$

Theorem 3.1. Under the assumptions stated in Section 2,

$$n^{1/2}(\hat{\rho}_n-\rho_n)\xrightarrow{P_{\theta_n,g_n}}N(0,I_*^{-1})$$

for all the sequences (θ_n,g_n) such that $n^{1/2}|\theta_n-\theta|=O(1)=n^{1/2}\|g_n^{-1/2}-g^{1/2}\|_\mu$.

As we mentioned it at the end of Section 2, the above theorem implies that $\hat{\rho}_n$ is asymptotically minimax in the sense of (2.5) for bounded subconvex loss functions. The proof of the theorem is given in the next section.

REMARK. We can simply extend the above results to multiple linear regression problems where we are interested in estimating the differences between two corresponding regression coefficients.

4. Proof of Theorem 3.1

We shall have to show (B)～(F) in Section 3. First (B) is obvious since $\int\hat{\ell}^*(x,y,z,\theta_n)f(x,y,z,\theta_n,g)\,dxdy=0$. We will show (D)～(F) first and (C) later. Note that $\hat{\ell}_j^*$ is defined as in (3.4) but replacing $\hat{h}$ by $\hat{h}_j$ where $\hat{h}_j(y,\theta)=(\partial/\partial y)\hat{g}_j(y,\theta)/\hat{g}_j(y,\theta)$ and $\hat{g}_j(y,\theta)=\hat{g}(y,\theta)-n^{-1}b_n^{-1}[Z_jK\{b_n^{-1}(y-Y_j+\alpha+\Delta+(\beta+\delta)X_j)\}I(|X_j|\leq c_n)+(1-Z_j)K\{b_n^{-1}(y-Y_j+\alpha+\beta X_j)\}I(|X_j|\leq c_n)]$. It follows easily (see Schick (1987)) that

$$\left|\dot{\hat{g}}_j/\hat{g}_j-\dot{\hat{g}}/\hat{g}\right|\leq 2n^{-1}b_n^{-2},$$

where $\dot{\hat{g}}(y) = (\partial/\partial y)\hat{g}(y)$. Thus we have

$$\left| \hat{\ell}_j^*(x,y,z,\theta) - \hat{\ell}^*(x,y,z,\theta) \right| \leq B_1 n^{-1} b_n^{-3} \ c_n \tag{4.1}$$

for some $B_1 > 0$. Now (D) is satisfied by (4.1). Note that

$$\dot{\hat{g}}(y,\theta) = n^{-1} b_n^{-2} \sum_{i=1}^{n} \{Z_i K'(b_n^{-1} \ (y - Y_i + \alpha + \Delta + (\beta + \delta) X_i)) I(|X_i| \leq c_n) + (1 - Z_i) K'(b_n^{-1} \ (y - Y_i + \alpha + \beta X_i)) I(|X_i| \leq c_n)\}.$$

For the logistic kernel K, $|K'| \leq K$ and we can see that

$$|\hat{\ell}^*(x,y,z,\theta)| \leq B_2 b_n^{-1} \ c_n$$

for some $B_2 > 0$, satisfying (F). Taking derivatives of $\dot{\hat{g}}$ and $\hat{g}$ with respect to each parameter, for example,

$$(\partial/\partial\beta)\dot{\hat{g}}(y,\theta) = n^{-1} b_n^{-3} \sum_{i=1}^{n} \{X_i Z_i K''(b_n^{-1} \ (y - Y_i + \alpha + \Delta + (\beta + \delta) X_i)) I(|X_i| \leq c_n) + X_i (1 - Z_i) K''(b_n^{-1} \ (y - Y_i + \alpha + \beta X_i)) I(|X_i| \leq c_n)\}$$

and observing that $|K''| \leq 2K$, it is easy to verify that

$$|(\partial/\partial\vartheta)\hat{\ell}^*(x,y,z,\theta)| \leq B_3 b_n^{-2} c_n^2$$

for some $B_3 > 0$. Therefore (E) is also satisfied. Verification of (C) relies on

Lemma. Under the same conditions as in Theorem 3.1,

$$E_{P_{\theta_n, g_n}} \int \int | \hat{h}(y,\theta_n) I(|x| \leq c_n) - (\dot{g}/g)(y)|^2 g(y) h(x) dx dy = o(1)$$

and

$$E_{P_{\theta_n, g_n}} \int \int x^2 | \hat{h}(y,\theta_n) I(|x| \leq c_n) - (\dot{g}/g)(y)|^2 g(y) h(x) dx dy = o(1)$$

By the above lemma, (C) is obvious.

Proof of lemma. It is enough to show that

$$E_{P_{\theta_n, g_n}} \int | \hat{h}(y,\theta_n) - (\dot{g}/g)(y)|^2 g(y) dy \to 0$$

as $n \to \infty$ since h has a finite second moment and $\int \dot{g}^2 / g < \infty$. Let

$$p_n(y,\theta_n) = E_{P_{\theta_n, g_n}} \hat{g}(y,\theta_n) - b_n$$
$$q_n(y,\theta_n) = E_{P_{\theta_n, g}} \hat{g}(y,\theta_n) - b_n.$$

Since $q_n(y,\theta_n)$ does not depend on θ_n, we can fix θ_n and proceed as in Section 6.1 of Bickel (1982) or Section 3 of Schick (1987) to get

$$\int [\dot{q}_n(y,\theta_n) / (q_n(y,\theta_n)+b_n) - (\dot{g} / g)(y)]^2 g(y)dy \to 0 \qquad (4.2)$$

as $n \to \infty$ since $\int_{|x|>c_n} h(x)dx \to 0$ where $\dot{q}_n(y,\theta_n)$ is the first derivative of $q_n(y,\theta_n)$ with respect to y. Also by the same method as in Schick (1987) (the bound obtained in (3.1 6) of Schick's paper comes independent of underlying carrying measures), we can see

$$E_{P_{\theta_n, g_n}} | \hat{h}(y,\theta_n) - \dot{p}_n(y,\theta_n) / (p_n(y,\theta_n)+b_n) |^2 \leq 4n^{-1} b_n^{-6} \qquad (4.3)$$

By (4.2) and (4.3) it is now enough to show that

$$\int | \dot{p}_n(y,\theta_n) / (p_n(y,\theta_n)+b_n) - \dot{q}_n(y,\theta_n) / (q_n(y,\theta_n)+b_n)|^2 g(y) \to 0 \qquad (4.4)$$

as $n \to \infty$. But the integrand of (4.4) is bounded by

$$2b_n^{-6} \ \{ \int | g_n(y) - g(y) | \ dy\}^2 \leq 2b_n^{-6} \ \int | g_n^{1/2}(y) - g^{1/2}(y) |^2 dy \int | g_n^{1/2}(y) + g^{1/2}(y)|^2 dy. \quad (4.5)$$

The above inequality is due to Cauchy-Schwarz. The right hand side of (4.5) is now bounded by $8b_n^{-6} \| g_n^{1/2} - g^{1/2} \|_\mu^2$ which converges to zero since $nb_n^6 \to \infty$. ■

References

(1) Begun, J.M., Hall, W.J., Huang, W. and Wellner, J.A. (1983). Information and asymptotic efficiency in parametric-nonparametric models, *Annals of Statistics,* Vol.11, 432-452.

(2) Beran, R. (1981). Efficient robust estimates in parametric models, *Zeitschrift für Wahrscheinlichkeitstheorie Verw. Gebiete,* Vol. 55, 91-108.

(3) Bickel, P.J.(1982). On adaptive estimation, *Annals of Statistics,* Vol. 10, 647-671.

(4) Bickel P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1989). Efficient and Adaptive Estimation in Semiparametric Models, *Forth coming monograph,* Johns Hopkins University Press, Baltimore.

(5) Hájek, J. (1970). A characterization of limiting distributions of regular estimates, *Zeitschrift für Wahrscheinlichkeitstheorie Verw. Gebiete,* Vol. 14,323-330.

(6) Hájek, J.(1972), Local asymptotic minimax and admissibility in estimation, *Proceedings of Sixth Berkeley Symposium on Mathematical Statistics and Probability.* Vol. 1, 175-194. University of California Press, Berkeley.

(7) Huber, P.J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo, *Annals of Statistics,* Vol.1, 799-821.

(8) Le Cam, L.(1960). Locally asymptotically normal families of distributions. University of California Publications in Statistics, Vol. 3, 37-98.

(9) Le Cam, L. (1972). Limits of experiments. *Proceedings of Sixth Berkeley Symposium on Mathematical Statistics and Probability.* Vol. 1, 245-261. University of California Press, Berkeley.

(10) Millar, P.W. (1984). A general approach to the optimality of minimum distance estimators, *Transactions of the American Mathematical Society,* Vol. 286, 377-418.

(11) Park, B.U.(1988). Asymptotic minimax estimation in semiparmetric models, Institute of Statistics Mimeo Series # 1751, University of North Carolina at Chapel Hill.

(12) Schick, A. (1987). A note on the construction of asymptotically linear estimators, *Journal of Statistical Planning and Inference.* Vol. 16, 89-105.