The Annals of Statistics
1992, Vol. 20, No. 2, 1129–1136

EFFICIENT ESTIMATES IN SEMIPARAMETRIC ADDITIVE REGRESSION MODELS WITH UNKNOWN ERROR DISTRIBUTION

BY JACK CUZICK

Imperial Cancer Research Fund

Several authors have shown how to efficiently estimate β in the semiparametric additive model $y = x'\beta + g(t) +$ error, $g(t)$ smooth but unknown when the error distribution is normal. However, the general theory suggests that efficient estimation should be possible for general error distributions with finite Fisher information even when the error distribution is unknown. In this note we construct a sequence of estimators which achieves this goal under technical assumptions.

Several authors [Heckman (1986, 1988), Chen (1988), Speckman (1988) and Cuzick (1992)] have shown to efficiently estimate β in the model $y = x'\beta + g(t) +$ error, $g(t)$ smooth but unknown when the error distribution is normal. However, the general theory of semiparametric models suggests that efficient estimation should be possible for general error distributions with finite Fisher information even when the error distribution is unknown [Wellner (1986), Bickel, Klaasen, Ritov and Wellner (1992) and Cuzick (1992)]. Cuzick (1992) showed this to be possible for general known error distributions, and that approach is extended here to construct a sequence of adaptive estimators which are efficient for a large class of error distributions when certain technical assumptions are satisfied.

1. Statement of the result. We use the notation of Cuzick (1992). Consider the model

$$y_i = x_i\beta + g(t_i) + \varepsilon_i, \qquad i = 1, \dots, n, \tag{1}$$

with $(x_i, t_i, \varepsilon_i)$ iid replicates from (X, T, ε), where ε is independent of (X, T), $E(\varepsilon) = 0$, $E(\varepsilon^2) < \infty$ and (X, T) is some bivariate distribution on $[0, 1]^2$. Assume ε has density f, $\varphi = -(\log f)'$ exists and $E\varphi^2(\varepsilon) < \infty$.

Split the data into two subsamples; the first of size $N \to \infty$, $N = o(n)$ and for each subsample order the data so $t_1 < \cdots < t_N$ and $t_{N+1} < \cdots < t_n$. Define local linear least squares smoothing matrices A_1 and A_2 of orders M_1, M_2, where, for a dataset of size n, the smoothing matrix of order M has

Received January 1991; revised June 1991.
AMS 1980 *subject classifications*. Primary 62G05; secondary 62J05, 62F35.
Key words and phrases. Semiparametric regression, additive models, linear models, Hájek–Le Cam lower bound, efficient estimators.

elements

$$a_{ij} = \begin{cases} \dfrac{1}{2M} + \dfrac{(t_i - \bar{t}_i)(t_j - \bar{t}_i)}{\sum_{1 \le |i-l| \le M}(t_l - \bar{t}_i)^2}, & M < i \le n - M;\ 1 \le |i-j| \le M, \\ \dfrac{1}{2k} + \dfrac{(t_i - \bar{t}_i)(t_j - \bar{t}_i)}{\sum_{1 \le |i-l| \le k}(t_l - \bar{t}_i)^2}, & i = k+1 \text{ or } n-k,\ 1 \le k < M; \\ & \quad 1 \le |i-j| \le k, \\ 1, & (i = 1, j = 2) \\ & \quad \text{or } (i = n, j = n-1), \\ 0, & \text{otherwise}, \end{cases}$$

where

$$\bar{t}_i = (2M)^{-1} \sum_{1 \le |i-l| \le M} t_l, \quad \text{for } M < i < n - M$$

and

$$\bar{t}_i = (2k)^{-1} \sum_{1 \le |i-l| \le k} t_l, \quad \text{for } i = k+1 \text{ or } n-k,\ 1 \le k < M.$$

Using the first subsample define the preliminary estimator

$$\hat{\beta} = \frac{((I - A_1)\mathbf{x})'(I - A_1)\mathbf{y}}{\|(I - A_1)\mathbf{x}\|^2},$$

where $\mathbf{x} = (x_1, \dots, x_N)'$ and $\mathbf{y} = (y_1, \dots, y_N)'$. When A_1 is a spline smoother, this is the least squares estimator originally considered by Rice (1986) and subsequently studied by several other authors. More generally if β is known, then $\mathbf{g}_\beta = A_1(\mathbf{y} - \mathbf{x}\beta)$ can be viewed as a smoother for $\mathbf{g} = (g(t_1), \dots, g(t_n))'$, so that when β is unknown, it is reasonable to estimate it by minimizing $\|\mathbf{y} - \mathbf{x}\beta - \mathbf{g}_\beta\|^2 = \|(I - A_1)(\mathbf{y} - \mathbf{x}\beta)\|^2$, which leads to $\hat{\beta}$. This estimate can also be derived from the general theory of semiparametric models [Cuzick (1992)]. Under weak conditions it can be shown to be $N^{1/2}$-consistent for general errors with finite variance and is fully efficient when the errors are normal.

Also define a smooth estimate φ_N of the influence function based on the first N samples as follows:

Let F_N be the empirical distribution function of the pseudoresiduals $\hat{\varepsilon}_i$ from the first subsample, where $(\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_N)' \equiv \hat{\boldsymbol{\varepsilon}} = (I - A_1)(\mathbf{y} - \hat{\beta}\mathbf{x})$. Let $\nu_\gamma(x) = \gamma^{-1}\nu(x\gamma^{-1})$ where $\nu(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ is the standard normal density, and let $F_{N_\gamma} = F_N * \nu_\gamma$ and $f_{N_\gamma} = F'_{N_\gamma}$, where $*$ denotes convolution. For $\gamma = \gamma_N \downarrow 0$, $\delta = \delta_N \downarrow 0$, $L = L_N \uparrow \infty$ at rates which are specified below, define

$$\varphi_N(x) = (l_N * \nu_\gamma)(x),$$

where

$$l_N(x) = \left\{f'_{N_\gamma}(x)/f_{N_\gamma}(x)\right\}I_{AB},$$
$$A = \left\{x: \left|f'_{N_\gamma}(x)/f_{N_\gamma}(x)\right| \le L\right\},$$
$$B = \left\{x: f_{N_\gamma}(x) \ge \delta\right\}.$$

Now use the remaining data to refine the estimate of β via

$$\hat{\hat{\beta}} = \hat{\beta} + \frac{\sum_{i=N+1}^{n} u_i \varphi_N(\hat{\varepsilon}_i)}{\sum_{i=N+1}^{n} u_i^2 \varphi'_N(\hat{\varepsilon}_i)}, \tag{2}$$

where $u_i = \{(I - A_2)\mathbf{x}\}_i$ and $\hat{\varepsilon}_i = \{(I - A_2)(\mathbf{y} - \hat{\beta}\mathbf{x})\}_i$ are based on samples $i = N + 1, \dots, n$. This can be seen as an approximation to the usual one-step estimator based on a likelihood, but here the influence function has been estimated from the first subsample by φ_N, pseudoresiduals $\hat{\varepsilon}_i$ replace the usual residuals and u_i is used instead of x_i.

THEOREM. *Consider the model* (1) *and the estimator* (2), *where the smoothing matrices A_1 and A_2 have bandwidths M_1 and M_2, respectively. Assume $N \uparrow \infty$, $N = o(n)$, $M_1 = o(N^{3/4})$, $M_2 = o(n^{3/4})$, $M_1^{1/4}\gamma^2\delta \to \infty$, $(N^{-1} + M_2^{-1})n^{1/2}L\gamma^{-2} = o(1)$ and the following conditions hold*:

(i) *T has a density bounded away from zero.*
(ii) *$g(t)$ and $w(t) \equiv E(X|T = t)$ are twice continuously differentiable.*

Then, as $n \to \infty$, $\hat{\hat{\beta}}$ is asymptotically linear, regular and

$$n^{1/2}\left(\hat{\hat{\beta}} - \beta\right) \to N\left(0, \sigma_{\mathrm{LB}}^2\right),$$

where

$$\sigma_{\mathrm{LB}}^2 = \sigma^2 E\left(\mathrm{Var}(X|T)E(\varphi^2(\varepsilon))\right)^{-1}$$

is the lower bound for the variance of any regular estimator.

REMARKS. (i) It is not difficult to exhibit a set of rates for the various constants which satisfy the conditions of the Theorem. For example $N = n^{2/3}$, $M_2 = n^{1/2+\alpha}$, $1/9 < \alpha < 1/4$, $M_1 = N^{2/3} = n^{4/9}$, $\delta = \gamma = M_1^{-1/16} = n^{-1/36}$, $L = n^{1/18}$ gives $M_1^{1/4}\gamma^2\delta = M_1^{1/16} \to \infty$, $M_2^{-1}n^{1/2}L\gamma^{-2} = n^{-\alpha+1/9} \to 0$ and $N^{-1}n^{1/2}L\gamma^{-2} = n^{-1/18} \to 0$. In general, one must always have $Nn^{-1/2} \to \infty$, $M_2 n^{-1/2} \to \infty$, $L = o(n^{1/4})$ and $n^{1/8}\gamma \to \infty$. The most efficient choice of rates or indeed the practical value of this construction remains unclear.

(ii) Bickel (1982) and Schick (1986) have outlined general procedures for constructing efficient estimators in semiparametric models. Because of the smoothing implicit in our approximate score function $u_i\varphi_N(\hat{\varepsilon}_i)$, it does not act on iid triples (y_i, x_i, t_i) so their results cannot be directly applied. Instead we show that our estimator is asymptotically linear by direct means.

(iii) It is not essential to use the particular form stated above for the smoothers A_1 and A_2. What is needed is a boundedness criteria such as $|a_{ij}| < M^{-1}$ and good bias control so that $\{(I - A)\mathbf{g}\}_i = O(t_{i+M} - t_{i-M})^2$.

2. Proof of the theorem. All sums are from $i = N + 1$ to n. The proof is based on two approximations:

$$\left|(\hat{\beta} - \beta) + \frac{\sum u_i \varphi_N(\hat{\varepsilon}_i)}{\sum u_i^2 \varphi_N'(\hat{\varepsilon}_i)} - \frac{\sum u_i \varphi_N(\varepsilon_i)}{\sum u_i^2 \varphi_N'(\varepsilon_i)}\right| = o_p(n^{-1/2}) \tag{3}$$

and

$$\left|\frac{\sum u_i \varphi_N(\varepsilon_i)}{\sum u_i^2 \varphi_N'(\varepsilon_i)} - \frac{\sum u_i \varphi(\varepsilon_i)}{\sum u_i^2 \varphi^2(\varepsilon_i)}\right| = o_p(n^{-1/2}) \tag{4}$$

giving

$$\hat{\hat{\beta}} = \beta + \frac{\sum u_i \varphi(\varepsilon_i)}{\sum u_i^2 \varphi^2(\varepsilon_i)} + o_p(n^{-1/2}),$$

from which the result follows from the methods of Cuzick (1992) when $N = o(n)$. We first show (4) follows from

$$E_1(\varphi_N(\varepsilon) - \varphi(\varepsilon))^2 \to 0 \quad \text{in probability,} \tag{5}$$

where E_1 refers to expectation conditional on the first subsample and convergence in probability is with respect to that subsample.

The left-hand side of (4) is bounded by

$$\left|\frac{\sum u_i(\varphi_N(\varepsilon_i) - \varphi(\varepsilon_i))}{\sum u_i^2 \varphi^2(\varepsilon_i)}\right| \tag{6}$$

$$+\left|\sum u_i \varphi_N(\varepsilon_i)\right|\left|\frac{\sum u_i^2(\varphi^2(\varepsilon_i) - \varphi_N'(\varepsilon_i))}{(\sum u_i^2 \varphi^2(\varepsilon_i))(\sum u_i^2 \varphi_N'(\varepsilon_i))}\right|. \tag{7}$$

Now the denominator of (6) is equal to $Kn(1 + o_p(1))$ for some $K > 0$ by the law of large numbers. To bound the numerator, recall $\mathbf{u} = (I - A_2)\mathbf{x}$ and write $\mathbf{x} = \mathbf{w} + \mathbf{r}$ where $w_i = E(X|T = t_i)$, $i = N + 1, \ldots, n$. Then $\{(I - A_2)\mathbf{w}\}_i = o_p(n^{-1/2})$,

$$E_1\big(\{(I - A_2)\mathbf{r}\}_i | t_k, \varepsilon_k, k = N + 1, \ldots, n\big) = 0$$

and

$$E_1\big(\{(I - A_2)\mathbf{r}\}_i \{(I - A_2)\mathbf{r}\}_j | t_k, \varepsilon_k, k = N + 1, \ldots, n\big) = 0$$

for $|i - j| > 2M_2$, $i, j = N + 1, \ldots, n$, is bounded when $i = j$ and is bounded by a constant times M_2^{-1} for $1 \le |i - j| \le 2M_2$, so

$$E_1\Big(\sum \{(I - A_2)\mathbf{r}\}_i (\varphi_N(\varepsilon_i) - \varphi(\varepsilon_i))\Big)^2 = o(n)$$

since $E_1(\varphi_N(\varepsilon_i) - \varphi(\varepsilon_i))^2 = o(1)$. Thus the numerator in (6) is $o_p(n^{1/2})$, leading to the conclusion that the whole term is $o_p(n^{-1/2})$. A similar argument implies that the first term in (7) is $O_p(n^{1/2})$ and the law of large numbers applied to the denominator establishes that (7) is

$$O_p\Big(n^{-3/2}\Big|\sum u_i^2\big(\varphi^2(\varepsilon_i) - \varphi_N'(\varepsilon_i)\big)\Big|\Big).$$

This is $o_p(n^{-1/2})$ by the law of large numbers and the observation that $E_1(\varphi^2(\varepsilon) - \varphi_N'(\varepsilon)) \to 0$ in probability, which is a consequence of (5) and $E_1\varphi_N'(\varepsilon) = E_1\varphi_N(\varepsilon)\varphi(\varepsilon)$. The truth of (5) is demonstrated in Lemma 1.

To obtain the bound (3), expand $\varphi_N(\hat{\varepsilon}_i)$ about ε_i to two terms in the numerator and its derivative to one term in the denominator, use

$$\sup|\varphi_N''| \le \sup\Big|\Big(\big(f_{N\gamma}'/f_{N\gamma}\big)I_{AB}\Big) * \nu_\gamma''\Big| \le L\gamma^{-2}$$

and

$$\hat{\varepsilon}_i - \varepsilon_i = v_i + \big(\beta - \hat{\beta}\big)u_i - (A_2\boldsymbol{\varepsilon})_i, \tag{8}$$

where $v_i = \{(I - A_2)\mathbf{g}\}_i$, to bound the left side of (3) by

$$O_p\Big[\Big(n^{-1}\Big|\sum u_i(v_i - (A_2\boldsymbol{\varepsilon})_i)\varphi_N'(\varepsilon_i)\Big| + n^{-1}L\gamma^{-2}\sum |u_i|(\varepsilon_i - \hat{\varepsilon}_i)^2\Big) \times\Big(1 + n^{-1}L\gamma^{-2}\sum u_i^2|\varepsilon_i - \hat{\varepsilon}_i|\Big)\Big]. \tag{9}$$

Now $|u_i| \le 1$ and $E|\varphi_N'(\varepsilon_i)| < \infty$, so the first part of the first term of (9) will be $o_p(n^{-1/2})$ if it can be shown that

$$\sum |v_i| = o_p(n^{1/2}). \tag{10}$$

Since $g \in C_2$, a Taylor expansion shows it is enough to bound

$$\sum \big(t_{i-M_2} - t_{i+M_2}\big)^2 \le 2M_2^2\sum (t_{i+1} - t_i)^2,$$

where $t_{i+M_2} = t_n$ if $i + M_2 \ge n$ and $t_{i-M_2} = t_{N+1}$ if $i - M_2 \le N$. But the $\{t_i\}$ have the same law as $F^{-1}(u_{(i-N)})$, where F is the distribution of T and the $\{u_{(j)}\}$ are uniform [0, 1] order statistics. Since we have assumed that F has a density bounded away from zero, this can be used to show that $\sum(t_{i+1} - t_i)^2 = O_p(n^{-1})$, which suffices to establish the result since $M_2 = o(n^{3/4})$.

To bound the second part, compute

$$\begin{aligned} E\Big(\sum_i u_i(A_2\boldsymbol{\varepsilon})_i\varphi_N'(\varepsilon_i)\Big)^2 &= \sum_i\sum_j\sum_k u_iu_ja_{ik}a_{jk}E(\varepsilon^2)E^2(\varphi_N'(\varepsilon)) \\ &\quad + \sum_i\sum_j u_iu_ja_{ij}a_{ji}E^2(\varepsilon\varphi_N'(\varepsilon)) \\ &\quad + \sum_i\sum_k u_i^2a_{ik}^2E(\varepsilon_2)\Big\{E(\varphi_N'(\varepsilon))^2 - E^2\varphi_N'(\varepsilon)\Big\}. \end{aligned} \tag{11}$$

Now it can be shown [Cuzick (1992)] that

$$\sup_{i,j} |a_{ij}| = O(M_2^{-1}) \tag{12}$$

and since $|\varphi_N'| \le L\gamma^{-1}$, the last two terms of (11) are $O((L\gamma^{-1})^2 n M_2^{-1})$, which is adequate to achieve the required bound since $L = O(n^{1/2})$. The first term in (11) is more delicate and requires that we use the probabilistic structure of the $\{x_i\}$. We have

$$\sum_i \sum_j \sum_k u_i u_j a_{ik} a_{jk} = \| A_2'(I - A_2)\mathbf{x} \|^2. \tag{13}$$

Write $\mathbf{x} = \mathbf{r} + \mathbf{w}$, where $w_i = E(X|T = t_i)$, $i = N+1, \dots, n$. Then, since $w(t) \in C_2$ by using (12) and the methods used for bounding $\sum |v_i|$, it can be shown that $\|A_2'(I - A_2)\mathbf{w}\|^2 = O_p(M_2^4 n^{-3}) = o_p(1)$.

Now, since given $\mathbf{t}$ the components of $\mathbf{r}$ are orthogonal and have conditional variances bounded by unity, $E_t\|A_2'(I - A)\mathbf{r}\|^2 \le \operatorname{tr}(I - A_2)'A_2 A_2'(I - A_2)$, which is $O(nM_2^{-1})$ by (12). Thus the first term of (11) is $O_p((L\gamma^{-1})^2 n M_2^{-1})$ as before.

For the other parts of (9), use (8) and the bounds $E(\hat\beta - \beta)^2 = O(N^{-1})$ [see Cuzick (1992)], $\sum v_i^2 = O_p(M_2^4 n^{-3}) = o_p(1)$, which can be obtained in a manner similar to (10), and $E(A\boldsymbol{\varepsilon})_i^2 = O(M_2^{-1})$ to find that

$$E^2\left(n^{-1} \sum |\varepsilon_i - \hat\varepsilon_i|\right) \le n^{-1} \sum E(\varepsilon_i - \hat\varepsilon_i)^2 = O\left(N^{-1} + M_2^{-1}\right),$$

which is adequate to achieve the required bounds. The proof is completed by proving (5), which is done in the following two lemmas.

LEMMA 1. *Assume $M_1 \uparrow \infty$, $\gamma \downarrow 0$, $\delta \downarrow 0$ and $M_1^{1/4}\gamma^2\delta \to \infty$. Then*

$$E_1(\varphi_N(\varepsilon) - \varphi(\varepsilon))^2 \to 0 \quad \textit{in probability},$$

where E_1 refers to expectation conditional on the first subsample and ε is independent of this subsample.

PROOF. In view of the definition of φ_N above (2), one must show that

$$E_1(l_N * \nu_\gamma(\varepsilon) - \varphi(\varepsilon))^2 \to 0 \quad \text{in probability}. \tag{14}$$

Letting f denote the density of ε, the left-hand side of (14) equals

$$\begin{aligned}
&\int \left(\int (l_N(x-y) - \varphi(x))\nu_\gamma(y)\,dy \right)^2 f(x)\,dx \\
&\quad \le \int\int (l_N(x-y) - \varphi(x))^2 \nu_\gamma(y) f(x)\,dy\,dx \\
&\quad = E_1(l_N(\varepsilon + \gamma Z) - \varphi(\varepsilon))^2 \\
&\quad \le 2\left\{ E_1(l_N(\varepsilon + \gamma Z) - \varphi_\gamma(\varepsilon + \gamma Z))^2 + E\left(\varphi_\gamma(\varepsilon + \gamma Z) - \varphi(\varepsilon)\right)^2 \right\},
\end{aligned} \tag{15}$$

where Z is a standard normal variable independent of the first subsample and of ε, $\varphi_\gamma = f'_\gamma/f_\gamma$ and $f_\gamma = f * \nu_\gamma$. Now Bickel [(1982), page 665] has shown $\varphi_\gamma(\varepsilon + \gamma Z)$ is uniformly square integrable, so the second term in (15) tends to zero when $\gamma \to 0$. Now $f' \in L_1(\mu)$ where μ is Lebesgue measure, so $f'_\gamma \to f'$ a.e. and

$$\| f'_{N\gamma} - f'_\gamma \|_\infty \le \gamma^{-2} \| F_N - F \|_\infty \le 2\gamma^{-2} E_1^{1/2} |F_N(\varepsilon) - F(\varepsilon)|$$

by Lemma 2 below. To bound this expression, let $\hat{F}$ be the empirical distribution function for the ε_i, $i = 1, \dots, N$, in the first subsample. Then

$$E(E_1|F_N(\varepsilon) - F(\varepsilon)|) \le E|F_N(\varepsilon) - \hat{F}(\varepsilon)| + E|\hat{F}(\varepsilon) - F(\varepsilon)|.$$

Now

$$E|\hat{F}(\varepsilon) - F(\varepsilon)| \le E^{1/2}(\hat{F}(\varepsilon) - F(\varepsilon))^2 \le N^{-1/2}$$

and

$$E|F_N(\varepsilon) - \hat{F}(\varepsilon)| \le N^{-1} E \left| \sum_{i=1}^{N} I_{\{\varepsilon \in (\varepsilon_i, \hat{\varepsilon}_i)\}} \right|$$
$$\le E|F(\varepsilon_i) - F(\hat{\varepsilon}_i)| \le KE|\varepsilon_i - \hat{\varepsilon}_i| = O(M_1^{-1/2})$$

where $K = \sup_x f(x)$ is bounded because of the assumption that ε has finite Fisher information. Thus

$$\| f'_{N\gamma} - f'_\gamma \|_\infty = O_p(\gamma^{-2} M_1^{-1/4}) \to 0, \tag{16}$$

so $f'_{N\gamma} \to f'$ in probability a.e. Similarly

$$\| f_{N\gamma} - f_\gamma \|_\infty = O_p(\gamma^{-1} M_1^{-1/4}) \to 0, \tag{17}$$

so $f_{N\gamma} \to f$ in probability a.e. and thus $f'_{N\gamma}/f_{N\gamma} \to \varphi$ in probability on $\{f > 0\}$. It follows that $EI_{(AB)^c} \to 0$, so $E_1(\varphi_\gamma(\varepsilon + \gamma Z) I_{(AB)^c})^2 \to 0$ in probability, and to bound the first term in (15), one need only consider

$$E_1\left(I_{AB}\left(\frac{f'_{N\gamma}}{f_{N\gamma}}(\varepsilon + \gamma Z) - \frac{f'_\gamma}{f_\gamma}(\varepsilon + \gamma Z) \right) \right)^2$$
$$\le 2\left\{ E_1\left(I_{AB} \frac{f'_{N\gamma} - f'_\gamma}{f_{N\gamma}} \right)^2 + E_1\left(I_{AB}\left(\frac{f'_\gamma}{f_\gamma} \right)\left(\frac{f_\gamma - f_{N\gamma}}{f_{N\gamma}} \right) \right)^2 \right\}$$
$$\le 2\delta^{-2}\left\{ \| f'_{N\gamma} - f'_\gamma \|_\infty^2 + \| f_{N\gamma} - f_\gamma \|_\infty^2 E(\varphi_\gamma(\varepsilon + \gamma Z))^2 \right\}$$
$$= O_p(\delta\gamma^2 M_1^{1/4})^{-2}$$

by (16), (17) and the uniform square integrability of $\varphi_\gamma(\varepsilon + \gamma Z)$. This last expression tends to zero by assumption. □

LEMMA 2. *Assume F is a continuous distribution function and G is a nondecreasing right continuous function. Then for $p \ge 1$,*

$$\|F - G\|_\infty \le ((p + 1)\|F - G\|_{L_p(F)})^{p/p+1}.$$

In particular for $p = 1$,

$$\|F - G\|_\infty \le \left(2\|F - G\|_{L_1(F)}\right)^{1/2}.$$

PROOF. Let $\Delta = \|F - G\|_\infty$. Given any $\varepsilon > 0$, there exists t such that $\Delta - \varepsilon \le F(t) - G(t^-) \le \Delta$ or $\Delta - \varepsilon \le G(t) - F(t) \le \Delta$. Assume the former and let $\delta = F(t) - G(t^-)$. Then

$$\|F - G\|_{L_p(F)}^p \ge \int_{t_0}^{t} |F(s) - F(t_0)|^p \, dF(s),$$

where t_0 is chosen so that $F(t_0) = F(t) - \delta$. Now let $x = F(s) - F(t_0)$ and change variables to get

$$\|F - G\|_{L_p(F)}^p \ge \int_0^\delta x^p \, dx = (p+1)^{-1}\delta^{p+1},$$

and since this is true for any $\varepsilon > 0$ we can replace δ by Δ and the result follows. □

REFERENCES

BICKEL, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1992). Efficient and adaptive inference in semiparametric models. Johns Hopkins Univ. Press.

CHEN, H. (1988). Convergence rates for parametric components in a partly linear model. *Ann. Statist.* **16** 136–146.

CUZICK, J. (1992). Semiparametric additive regression. *J. Roy. Statist. Soc. Ser. B.* To appear.

HECKMAN, N. (1986). Spline smoothing in partially linear models. *J. Roy. Statist. Soc. Ser. B* **48** 244–248.

HECKMAN, N. (1988). Minimax estimates in a semiparametric model. *J. Amer. Statist. Assoc.* **83** 1090–1096.

RICE, J. (1986). Convergence rates for partially splined models. *Statist. Probab. Lett.* **4** 203–208.

SCHICK, A. (1986). On asymptotically efficient estimation in semiparametric models. *Ann. Statist.* **14** 1139–1151.

SPECKMAN, P. (1988). Kernel smoothing in partial linear models. *J. Roy. Statist. Soc. Ser. B* **50** 413–436.

WELLNER, J. A. (1986). Semiparametric models: Progress and problems. *Bull. Inst. Internat. Statist.* **4** 23.1.1–23.1.20.

DEPARTMENT OF MATHEMATICS,
STATISTICS AND EPIDEMIOLOGY
IMPERIAL CANCER RESEARCH FUND
P.O. BOX NO. 123
LINCOLN'S INN FIELDS
LONDON, WC2A 3PX
ENGLAND