*The Annals of Statistics*
1986, Vol. 14, No. 3, 1139–1151

# ON ASYMPTOTICALLY EFFICIENT ESTIMATION IN SEMIPARAMETRIC MODELS

BY ANTON SCHICK

*State University of New York-Binghamton*

A general method for the construction of asymptotically efficient estimates in semiparametric models is presented. It improves and modifies Bickel's (1982) construction of adaptive estimates and obtains asymptotically efficient estimates under conditions weaker than those in Bickel.

**1. Introduction.** In this paper we give a general method for the construction of asymptotically efficient estimates in semiparametric models. More specifically, our estimates are regular with smallest possible asymptotic variance as discussed in Begun, Hall, Huang, and Wellner (1983) and are LAM-adaptive in the sense of Fabian and Hannan (1982) and adaptive in the sense of Begun, Hall, Huang, and Wellner (1983) if Stein's (1956) necessary condition for adaptive estimation holds. Our construction improves and generalizes Bickel's (1982) method of constructing adaptive estimates: We obtain asymptotically efficient estimates under weaker conditions than in Bickel and use the entire sample to construct estimates of the score function or the nuisance parameter and not just a small fraction as Bickel does. Our construction compares also favorably with a construction given by Huang (1982) in a thesis.

We show that Bickel's condition $S^*$ which he reasons is "heuristically necessary" for the existence of adaptive estimates in convex models and which motivates Bickel's construction is not necessary for adaptive estimation in general. We replace it by a weaker condition and show that this condition is necessary for our construction. It is seen that if $S^*$ does not hold adaptive estimates are more difficult to construct in that a certain rate of convergence is required for the estimate of the nuisance parameter.

Our paper is organized as follows. In Section 2 we present the construction of asymptotically efficient estimates. In Section 3 we present examples and show that condition $S^*$ is not necessary for the construction of adaptive estimates.

Some notation will be introduced next. $\langle\ \rangle$ will be used to denote finite or infinite sequences, and, in particular, points in $\mathbb{R}^k$. In matrix calculations, points in $\mathbb{R}^k$ are columns.

If $\langle A_n, \underline{A}_n, P_n \rangle$ is a probability space and $g_n$ is a measurable function on $A_n$ to $\mathbb{R}^k$ for each $n = 1, 2, \ldots,$ and if $c \in \mathbb{R}^k$, then (i) we write $g_n \to c$ in $\langle P_n \rangle$-prob. if $P_n(\|g_n - c\| > \varepsilon) \to 0$ for every $\varepsilon > 0$ and (ii) we say $\langle g_n \rangle$ is bounded in $\langle P_n \rangle$-prob. if $\langle F_n \rangle$ is tight, where $F_n$ is the distribution of $g_n$ under $P_n$.

Received September 1984; revised October 1985.
*AMS* 1980 *subject classifications*. 62G05, 62G20.
*Key words and phrases*. Efficient estimation, adaptation, semiparametric model, regression model.

**2. The construction of asymptotically efficient estimates.** Throughout this section we assume that $\{f_\theta,\ \theta \in \Theta\}$ is a family of probability densities with respect to a sigma-finite measure $\nu$ on a measurable space $(S, \underline{S})$, that the index set $\Theta = \Theta_1 \times \Theta_2$ for some open subset $\Theta_1$ of $\mathbb{R}^p$ and some arbitrary nonempty set $\Theta_2$, and that, for every $\theta \in \Theta$, there is a $\rho_\theta$ in $L_2^p(\nu)$ such that

$$\left\| f^{1/2}_{\langle \theta_1 + a, \theta_2 \rangle} - f_\theta^{1/2} - a^T \rho_\theta \right\|_\nu = o(\|a\|), \tag{2.1}$$

where $\|\cdot\|_\nu$ denotes the $L_2(\nu)$-norm. We also assume that the matrix

$$I_*(\theta) = 4 \int \rho_\theta^* \rho_\theta^{*T}\, d\nu$$

is nonsingular for all $\theta \in \Theta$, where $\rho_\theta^*$ is the vector whose components are the projections of the components of $\rho_\theta$ onto the orthogonal complement of $\mathcal{J}_2(\theta)$, the set of all $\psi$ in $L_2(\nu)$ for which there is a map $\eta$ on $(-1,1)$ into $\Theta_2$ such that $\eta(0) = \theta_2$ and

$$\left\| f^{1/2}_{\langle \theta_1 + a, \eta(b) \rangle} - f_\theta^{1/2} - a^T \rho_\theta - b\psi \right\|_\nu^2 = o\left(\|a\|^2 + b^2\right).$$

The above conditions generalize some of the concepts in Begun, Hall, Huang, and Wellner (1983) to the case of an arbitrary nuisance parameter set $\Theta_2$. The reader familiar with their paper will recognize that $\mathcal{J}_2(\theta)$ plays the role of their tangent space $\{A\beta\colon \beta \in B\}$ and $\rho_\theta^*$ generalizes what they call the effective score function. Observe also that Stein's (1956) necessary condition for adaptive estimation as reformulated by Bickel (1982) and Fabian and Hannan (1982) can be stated, in the present context, as

$$\rho_\theta^* = \rho_\theta \quad \text{for all } \theta \in \Theta. \tag{S}$$

For convenience in notation we shall often write $f(\cdot, t, v)$ instead of $f_{\langle t, v \rangle}$ and similarly for other functions $g_\theta$, $\theta \in \Theta$.

Now consider probability measures $\{P_\theta,\ \theta \in \Theta\}$ and $S$-valued random variables $X_1, X_2, \ldots,$ such that under each $P_\theta$, $X_1, X_2, \ldots$ are independent and identically distributed with density $f_\theta$. Our goal is to construct an estimate $\langle Z_n \rangle = \langle z_n(X_1, \ldots, X_n) \rangle$ which satisfies

$$n^{1/2}\left(Z_n - Z_n(\theta)\right) \to 0 \quad \text{in } P_\theta\text{-prob.} \tag{2.2}$$

for each $\theta \in \Theta$, where

$$Z_n(\theta) = \theta_1 + \frac{1}{n} \sum_{j=1}^{n} l_*(X_j, \theta)$$

and

$$l_*(\cdot, \theta) = I_*^{-1}(\theta) 2 f_\theta^{-1/2} \rho_\theta^* \chi_{\{f_\theta > 0\}}.$$

We call an estimate $\langle Z_n \rangle$ that satisfies (2.2) *asymptotically linear at* $\theta$ and write $\langle Z_n \rangle$ is AL($\theta$). An estimate that is AL($\theta$) for all $\theta \in \Theta$ is said to be *asymptotically linear*. Our interest in asymptotically linear estimates is based on the following results.

Suppose $\langle Z_n \rangle$ is AL($\theta$), then $\langle Z_n \rangle$ is regular at $\theta$, i.e., for every sequence $\langle t_n, b_n \rangle$ in $\Theta_1 \times (-1,1)$ such that $\langle n^{1/2}(t_n - \theta_1, b_n) \rangle$ is bounded and every map $\eta$ as described in the definition of $\mathcal{J}_2(\theta)$

$$\mathscr{L}\left(n^{1/2}(Z_n - t_n)|P_{t_n, \eta(b_n)}\right) \Rightarrow \mathscr{N}\left(0, I_*^{-1}(\theta)\right).$$

This follows from a straightforward contiguity argument. Moreover, under mild additional assumptions $\langle Z_n \rangle$ has the smallest possible asymptotic variance among all estimates regular at $\theta$ [see Begun, Hall, Huang, and Wellner (1983), Theorem 3.1], and if the necessary condition for adaptive estimation $\rho_\theta^* = \rho_\theta$ holds, then $\langle Z_n \rangle$ is LAM-adaptive at $\theta$ in the sense of Fabian and Hannan (1982) and adaptive at $\theta$ in the sense of Begun, Hall, Huang, and Wellner (1983).

REMARK 1. Bickel (1982) defines adaptivity at $\theta$ for an estimate $\langle Z_n \rangle$ by

(i) For every sequence $\langle t_n \rangle$ in $\Theta_1$ such that $\langle n^{1/2}(t_n - \theta_1) \rangle$ is bounded

$$\mathscr{L}\left(n^{1/2}(Z_n - t_n)|P_{t_n, \theta_2}\right) \Rightarrow \mathscr{N}\left(0, I^{-1}(\theta)\right),$$

where $I(\theta) = 4\int \rho_\theta \rho_\theta^T \, d\nu$.

Condition (i) is equivalent to

$$\text{(ii)} \qquad n^{1/2}\left( Z_n - \theta_1 - \frac{1}{n} \sum_{j=1}^{n} l(X_j; \theta) \right) \to 0 \quad \text{in } P_\theta\text{-prob.},$$

where $l(\cdot, \theta) = I^{-1}(\theta) 2 f_\theta^{-1/2} \rho_\theta \chi_{\{f_\theta > 0\}}$.

This follows from Theorem 6.3 in Fabian and Hannan (1982) and Theorem 6.1 in Bickel (1982) and the note thereafter. Thus an estimate adaptive in Bickel's sense is AL($\theta$) if and only if $\rho_\theta^* = \rho_\theta$. Bickel claims that the existence of estimates adaptive in his sense implies the necessary condition for adaptive estimation which would imply that such estimates are automatically asymptotically linear. But the proof of this claim is incorrect due to an inappropriate reference to Hájek (1972): Bickel considers only local alternatives of the first component of the parameter $\theta$ and not local alternatives of both components as needed in Hájek's Theorem 4.2.

We shall now consider the construction of asymptotically linear estimates. We begin by introducing the following assumptions.

(A.1) The map $t \in \Theta_1 \rightsquigarrow \rho_{\langle t, v \rangle}^*$ is continuous for all $v \in \Theta_2$.

(A.2) $\langle U_n \rangle$ is a $\Theta_1$-valued estimate such that $\langle n^{1/2}(U_n - \theta_1) \rangle$ is bounded in $P_\theta$-prob. for all $\theta \in \Theta$.

(A.3) For every $n = 1, 2, \ldots, \hat{l}_n$ is a measurable map on $S \times \Theta_1 \times S^n$ into $\mathbb{R}^p$ such that for each $\theta \in \Theta$ and every sequence $\langle t_n \rangle$ in $\Theta_1$ for which $\langle n^{1/2}(t_n - \theta_1) \rangle$ is bounded

$$\text{(2.3)} \qquad n^{1/2} \int \hat{l}_n(\cdot, t_n, X_1, \ldots, X_n) f(\cdot, t_n, \theta_2) \, d\nu \to 0 \quad \text{in } P_\theta\text{-prob.}$$

and

$$\text{(2.4)} \qquad \int \left\| \hat{l}_n(\cdot, t_n, X_1, \ldots, X_n) - l_*(\cdot, t_n, \theta_2) \right\|^2 f(\cdot, t_n, \theta_2) \, d\nu \to 0 \quad \text{in } P_\theta\text{-prob.}$$

Assumption (A.1) corresponds to Bickel's condition UR(iii). It allows us to conclude that

$$n^{1/2}(Z_n(t_n, \theta_2) - Z_n(\theta)) \to 0 \quad \text{in } P_\theta\text{-prob.} \tag{2.5}$$

for every $\theta \in \Theta$ and every sequence $\langle t_n \rangle$ in $\Theta_1$ for which $\langle n^{1/2}(t_n - \theta_1) \rangle$ is bounded.

Assumption (A.2) is Bickel's condition GR(iv) and is obviously necessary for the existence of asymptotically linear estimates.

Assumption (A.3) generalizes Bickel's condition $H$. We replace his requirement that $\hat{l}_n(\cdot, \cdot, X_1, \dots, X_n)$ is $\mathscr{H}$-valued, i.e.,

$$\int \hat{l}_n(\cdot, \theta_1, X_1, \dots, X_n) f(\cdot, \theta)\, d\nu = 0 \quad \text{for all } \theta \in \Theta \tag{2.6}$$

by the weaker (2.3). Bickel argues that his condition $S^*$ which suggests (2.6) is "heuristically necessary" for adaptive estimation in convex models. Condition $S^*$, however, is not necessary in general, i.e., there exist nonconvex models for which $S^*$ does not hold but adaptive estimates exist. For an example see Section 3. If $S^*$ does not hold the set

$$\mathscr{H} = \left\{ h\colon h \text{ is a map on } S \times \Theta_1 \text{ into } \mathbb{R}^p \text{ such that } \int h(\cdot, \theta_1) f(\cdot, \theta)\, d\nu = 0 \text{ for all } \theta \in \Theta \right\}$$

may not be large enough for Bickel's condition $H$ to hold. In view of this (2.3) appears to be a necessary improvement.

Next we introduce our estimate. For technical reasons we adopt Bickel's idea of splitting the sample, but modify it to obtain better estimates of the score function. Bickel splits the sample in two unequal parts, estimates the score function based on the observations in the smaller subsample, and evaluates the estimate of the score function only at observations of the larger part. We divide the ample in two equal parts, obtain an estimate of the score function from each part, and evaluate the estimate of the score function obtained from the first part only with observations from the second part and vice versa. Thus our estimates of the score function are based on half of the sample and not just on a small proportion of the sample. Our estimate is formally defined by

$$\tilde{Z}_n = \overline{U}_n + \frac{1}{n}\left( \sum_{j=1}^{k_n} L_{n2}(X_j, \overline{U}_n) + \sum_{j=k_n+1}^{n} L_{n1}(X_j; \overline{U}_n) \right), \tag{2.7}$$

where $k_n$ is the integer part of $n/2$, $L_{n1} = \hat{l}_{k_n}(\cdot, \cdot, X_1, \dots, X_{k_n})$, $L_{n2} = \hat{l}_{n-k_n}(\cdot, \cdot, X_{k_n+1}, \dots, X_n)$, and $\langle \overline{U}_n \rangle$ is a discretized version of $\langle U_n \rangle$. For a discussion and the use of discrete estimates we refer to Fabian and Hannan (1982) and Bickel (1982).

THEOREM 1. *If assumptions* (A.1), (A.2), *and* (A.3) *hold, then* $\langle \tilde{Z}_n \rangle$ *is asymptotically linear.*

PROOF. Let $\theta \in \Theta$. We must show that $\langle \tilde{Z}_n \rangle$ is AL($\theta$). Since $\langle \overline{U}_n \rangle$ is discrete it suffices to show that

$$\left\langle t_n + \frac{1}{n}\left( \sum_{j=1}^{k_n} L_{n2}(X_j, t_n) + \sum_{j=k_n+1}^{n} L_{n1}(X_j, t_n) \right) \right\rangle$$

is AL($\theta$) for any sequence $\langle t_n \rangle$ in $\Theta_1$, such that $\langle n^{1/2}(t_n - \theta_1) \rangle$ is bounded. Fix now such a sequence $\langle t_n \rangle$. In view of (2.3) and (2.5) it suffices to show

$$n^{-1/2} \sum_{j=1}^{k_n} \left( \overline{L}_{n2}(X_j, t_n) - l_*(X_j, t_n, \theta_2) \right) \to 0 \quad \text{in } P_\theta\text{-prob.}$$

and

$$n^{-1/2} \sum_{j=k_n+1}^{n} \left( \overline{L}_{n1}(X_j, t_n) - l_*(X_j, t_n, \theta_2) \right) \to 0 \quad \text{in } P_\theta\text{-prob.},$$

where $\overline{L}_{ni}(\cdot, t_n) = L_{ni}(\cdot, t_n) - \int L_{ni}(\cdot, t_n) f(\cdot, t_n, \theta_2)\, d\nu$, $i = 1, 2$.

These two statements are proved exactly as is (3.7) in Bickel (1982). We omit the details. This concludes the proof. □

REMARK 2. Actually, more can be shown. Suppose (A.1), (A.2), and the conditions of (A.3) except possibly (2.3) hold and

$$Z_n(t) = t + \frac{1}{n}\left( \sum_{j=1}^{k_n} L_{n2}(X_j, t) + \sum_{j=k_n+1}^{k} L_{n1}(X_j, t) \right).$$

Then the following are equivalent for each $\theta \in \Theta$.

(a) For every sequence $\langle t_n \rangle$ in $\Theta_1$ such that $\langle n^{1/2}(t_n - \theta_1) \rangle$ is bounded $\langle Z_n(t_n) \rangle$ is AL($\theta$).

(b) For every sequence $\langle t_n \rangle$ in $\Theta_1$ such that $\langle n^{1/2}(t_n - \theta_1) \rangle$ is bounded

$$n^{1/2} \int L_n(\cdot, t_n) f(\cdot, t_n, \theta_2)\, d\nu \to 0 \quad \text{in } P_\theta\text{-prob.}$$

The proof is easy. Fix $\theta \in \Theta$. Let

$$R_n(t) = \frac{k_n}{n} \int L_{n2}(\cdot, t) f(\cdot, t, \theta_2)\, d\nu + \frac{n - k_n}{n} \int L_{n1}(\cdot, t) f(\cdot, t, \theta_2)\, d\nu.$$

In the proof of Theorem 1 we have shown that $\langle Z_n(t_n) - R_n(t_n) \rangle$ satisfies (2.2) for every sequence $\langle t_n \rangle$ in $\Theta_1$ such that $\langle n^{1/2}(t_n - \theta_1) \rangle$ is bounded. This did not require (2.3). Thus (a) is equivalent to

(c) For every sequence $\langle t_n \rangle$ in $\Theta_1$ such that $n^{1/2}(t_n - \theta_1) \rangle$ is bounded

$$n^{1/2} R_n(t_n) \to 0 \quad \text{in } P_\theta\text{-prob.}$$

And this is easily seen to be equivalent to (b).

Obviously we do not want the asymptotic linearity of $\langle \tilde{Z}_n \rangle$ to depend on the way we discretize, i.e., we want $\langle \tilde{Z}_n \rangle$ to be asymptotically linear for every discretized version $\langle \bar{U}_n \rangle$ of $\langle U_n \rangle$. This is equivalent to requiring (a) for all $\theta \in \Theta$.

The above shows that, in the presence of the other assumptions, (2.3) is necessary for $\langle \tilde{Z}_n \rangle$ to be asymptotically linear for every discretized version $\langle \bar{U}_n \rangle$ of $\langle U_n \rangle$.

We shall now discuss (2.3) in more detail. For this discussion we suppose $\Theta_2$ is a topological space, $l_*$ is measurable and satisfies, for every $\theta \in \Theta$,

$$\int \| l_*(\cdot, t, v) - l_*(\cdot, t, \theta_2) \|^2 f(\cdot, t, \theta_2)\, d\nu \to 0$$

as $t \to \theta_1$ and $v \to \theta_2$, and

(C.1) $\langle V_n \rangle$ is a $\Theta_2$-valued estimate which satisfies $V_n \to \theta_2$ in $P_\theta$-prob. for all $\theta \in \Theta$ and

(C.2) For every $\theta \in \Theta$ and every sequence $\langle t_n \rangle$ in $\Theta_1$ such that $\langle n^{1/2}(t_n - \theta_1) \rangle$ is bounded

$$n^{1/2} Q(t_n, V_n, \theta_2) \to 0 \quad \text{in } P_\theta\text{-prob.,} \tag{2.8}$$

where $Q$ is the map on $\Theta_1 \times \Theta_2 \times \Theta_2$ to $\mathbb{R}^p$ defined by

$$Q(t, v, w) = \int l_*(\cdot, t, v) f(\cdot, t, w)\, d\nu \tag{2.9}$$

if the integral is well defined and 0 otherwise.

It is now easily seen that (A.3) holds with $\hat{l}_n(\cdot, \cdot, X_1, \dots, X_n) = l_*(\cdot, \cdot, V_n)$. Note that (2.8) corresponds to (2.3).

Condition (C.2) implies a certain rate of convergence for the estimate $\langle V_n \rangle$. $Q$ measures how difficult the construction of asymptotically linear estimates is by specifying this rate. In particular, if $Q = 0$, no specific rate is necessary. Bickel's condition $S^*$ implies $Q = 0$ and this explains why Bickel is able to construct adaptive estimates using only a small fraction of the sample to estimate the nuisance parameter.

REMARK 3. In a thesis Huang (1982) considers a different method of constructing asymptotically linear estimates. His estimate is a solution of the equations

$$\int \rho^*(x, t, V_n) \hat{f}_n^{1/2}(x)\, d\nu(x) = 0, \tag{$*$}$$

where $\langle V_n \rangle$ is an appropriate estimate of the nuisance parameter $\theta_2$ and $\langle \hat{f}_n \rangle$ is an appropriate estimate of the density $f_\theta$. He proves that this estimate is asymptotically linear if it is consistent and if strong additional regularity conditions hold. These regularity conditions severely limit the use of his estimate and proving consistency of his estimate may pose difficult mathematical problems.

We feel also that our estimate is easier to calculate, since a solution of $(*)$ may require much more extensive calculations.

**3. Examples.** This section serves two purposes. It illustrates the above results and provides an example of a model for which adaptive estimates exist but condition $S^*$ does not hold.

Throughout $\mathscr{V}$ denotes the set of all real valued functions $v$ on $[0,1]$ which are absolutely continuous with square integrable derivatives and satisfy $\int_0^1 v(t)\,dt = 0$, and $g$ denotes a Lebesgue density which satisfies

$$\int yg(y)\,dy = 0, \tag{3.1}$$

$$\int y^2 g(y)\,dy = \sigma^2 < \infty, \tag{3.2}$$

$g$ is absolute continuous with finite Fisher information

$$J_g = \int \frac{(g'(y))^2}{g(y)}\,dy, \tag{3.3}$$

and

$$\int (L(y-t) - L(y))^2 g(y)\,dy = o(t), \tag{3.4}$$

where $L = -J_g^{-1}(g'/g)\chi_{\{g>0\}}$.

Now consider the regression model

$$Y = \theta_1 + \theta_2(T) + \varepsilon, \tag{3.5}$$

where $\varepsilon$ and $T$ are independent random variables, $\varepsilon$ has density $g$, $T$ has uniform distribution on $[0,1]$, $\theta_1$ is an unknown real number, and $\theta_2$ is an unknown function in $\mathscr{V}$.

This model belongs to a class of models which has been recently proposed by Engle, Granger, Rice, and Weiss (1986) and is of considerable practical interest. For generalizations and related models see also Wahba (1984).

For our model, $\Theta_1 = \mathbb{R}$, $\Theta_2 = \mathscr{V}$, $S = \mathbb{R} \times [0,1]$, $\nu$ is the Lebesgue measure on the Borel field of $S$ and the densities $f_\theta$ are given by

$$f_\theta(x) = g(x_1 - \theta_1 - \theta_2(x_2)), \qquad x \in S.$$

From (3.3) we obtain that (2.1) holds with

$$\rho_\theta(x) = -\frac{g'}{2g^{1/2}}(x_1 - \theta_1 - \theta_2(x_2)), \qquad x \in S.$$

See Hájek (1972) for details. From Theorem 9.5 in Rudin (1974) we can derive that $\theta_1 \in \mathbb{R} \rightsquigarrow \rho_\theta$ is continuous in $L_2(\nu)$.

The tangent space $J_2(\theta)$ is the $L_2(\nu)$-closure of the set of functions $\psi$ in $L_2(\nu)$ of the form

$$\psi(x) = v(x_2)\rho_\theta(x), \qquad x \in S,$$

for some $v \in \mathscr{V}$. Thus $\rho_\theta$ is orthogonal to $J_2(\theta)$, $\rho_\theta^* = \rho_\theta$, and $l_*(x,\theta) =$

$L(x_1 - \theta_1 - \theta_2(x_2))$, $x \in S$. Consequently, the necessary condition for adaptive estimation holds. Also, Bickel's condition $S^*$ holds if $g$ is the standard normal density. But Bickel's condition $S^*$ does not hold in general; e.g., if $g$ is the double exponential density, $g(y) = \frac{1}{2}e^{-|y|}$, then $L(y) = \operatorname{sign}(y)$ and with $Q$ as given in (2.9)

$$\begin{aligned} Q(t, \theta_2 + v, \theta_2) &= Q(0, v, 0) \\ &= \int_0^1 \int_{-\infty}^{\infty} \operatorname{sign}(y - v(t)) \tfrac{1}{2} e^{-|y|}\, dy\, dt \\ &= \int_0^1 \operatorname{sign}(v(t))(e^{-|v(t)|} - 1)\, dt \\ &= \int_0^1 \operatorname{sign}(v(t))\big(e^{-|v(t)|} - 1 - |v(t)|\big)\, dt \\ &= 0\left(\int v^2(t)\, dt\right), \end{aligned} \tag{3.6}$$

but $Q \neq 0$.

Suppose now that $\langle Y_1, T_1\rangle, \langle Y_2, T_2\rangle, \dots$ are independent copies of $\langle Y, T\rangle$. We have already seen that (A.1) is satisfied, and it is easy to verify that (A.2) holds with $\langle U_n\rangle = \langle \overline{Y}_n\rangle = \langle 1/n \Sigma_{j=1}^n Y_j\rangle$. Thus we are left to show (A.3). We shall construct an estimate $\langle V_n\rangle$ which satisfies

$$\int_0^1 V_n(t)\, dt = 0 \quad \text{and} \quad E_\theta \int_0^1 (V_n(t) - \theta_2(t))^2\, dt = O(n^{-2/3}). \tag{3.7}$$

This implies (A.3) since there are functions $\langle L_n\rangle$ such that

$$\int_0^1 \int (L_n(y - v_n(t)) - L(y))^2 g(y)\, dy\, dt \to 0 \tag{3.8}$$

and

$$n^{1/2} \int_0^1 \int L_n(y - v_n(t)) g(y)\, dy\, dt \to 0 \tag{3.9}$$

whenever $\int_0^1 v_n(t)\, dt = 0$ and $\int_0^1 v_n^2(t)\, dt = O(n^{-2/3})$. For many important examples of $g$, such as the normal or the double exponential density, we can choose $L_n = L$. If this choice is not possible the functions $\langle L_n\rangle$ can be constructed as follows. For a sequence $\langle \tau_n\rangle$ of positive numbers such that $\tau_n \to \infty$ and $n\tau_n^{-18} \to \infty$, set

$$\bar{\lambda}_n = \int \lambda_n(\cdot - \tau_n^{-1} x)\psi(x)\, dx,$$

where $\lambda_n = (-\tau_n) \vee L \wedge \tau_n$ and $\psi$ is the standard normal density, and define

$$L_n = \bar{\lambda}_n - \int \bar{\lambda}_n(y) g(y)\, dy.$$

It is easily verified that the sequence $\langle L_n \rangle$ satisfies

$$\int L_n(y)g(y)\,dy = 0,$$

$$\int (L_n(y) - L(y))^2 g(y)\,dy \to 0,$$

and

$$\sup_y |L_n^{(i)}(y)| = O(\tau_n^{i+1}), \qquad i = 0,1,2.$$

The statements (3.8) and (3.9) are now readily derived from this, e.g.,

$$\begin{aligned} n^{1/2} &\int\int L_n(y - v_n(t))g(y)\,dy\,dt \\ &= n^{1/2} \int\int \left(L_n(y - v_n(t)) - L_n(y) + v_n(t)L_n^{(1)}(y)\right)g(y)\,dy\,dt \\ &= O\left(n^{1/2}\tau_n^3 \int_0^1 v_n^2(t)\,dt\right) = o(1) \end{aligned}$$

if $\int_0^1 v_n(t)\,dt = 0$ and $\int_0^1 v_n^2(t)\,dt = O(n^{-2/3})$.

We shall now construct the estimate $\langle V_n \rangle$. For a related construction, under slightly stronger assumptions, see Stone (1985). His results show also that better rates of convergence are possible under additional smoothness conditions on $\theta_2$.

Let $\langle a_n \rangle$ denote a sequence of positive integers and set $b_n = a_n^{-1}$. For each $n = 1, 2, \dots,$ partition the unit interval $[0,1]$ in $a_n$ intervals $I_{ni}$, $i = 1, \dots, a_n$, of equal length $b_n$ and let $m_{ni}$ denote the midpoint of $I_{ni}$ and $\chi_{ni}$ the indicator of $I_{ni}$. We assume that the intervals $I_{ni}$ are numbered in such a way that $m_{nj} < m_{nk}$ for $1 \le j < k \le a_n$. Next set

$$U_n = n^{-1} \sum_{j=1}^{n} Y_j$$

and

$$Y_{ni} = (nb_n)^{-1} \sum_{j=1}^{n} Y_j \chi_{ni}(T_j), \qquad i = 1, \dots, a_n,$$

and define $V_n$ by

$$(3.10) \quad V_n(t) = \begin{cases} Y_{n1} - U_n, & 0 \le t \le m_{n1}, \\ Y_{ni} - U_n + \dfrac{t - m_{ni}}{b_n}(Y_{ni+1} - Y_{ni}), & m_{ni} \le t < m_{ni+1}, \\ Y_{na_n} - U_n, & m_{na_n} \le t \le 1. \end{cases}$$

It is easily verified that

$$\int_0^1 V_n(t)\,dt = \sum_{i=1}^{a_n} b_n Y_{ni} - U_n = 0.$$

LEMMA. *If the sequence $\langle a_n \rangle$ is chosen such that*

$$\frac{a_n^3}{n} \to 1, \tag{1}$$

*then*

$$E_\theta \int_0^1 (V_n(t) - \theta_2(t))^2\, dt = O(n^{-2/3}).$$

PROOF. For $i = 1, 2, \ldots, a_n$ set

$$C_{ni} = a_n \int_0^1 \chi_{ni}(u)\theta_2(u)\, du \tag{2}$$

and note that

$$E_\theta Y_{ni} = \theta_1 + C_{ni}. \tag{3}$$

Easy calculations show that for some constant $c$

$$E_\theta(Y_{ni} - \theta_1 - C_{ni})^2 \le \frac{c}{nb_n} \tag{4}$$

and

$$E_\theta(U_n - \theta_1)^2 \le \frac{c}{n}. \tag{5}$$

Next note that by the Schwarz inequality for $0 \le t < u \le 1$

$$(\theta_2(t) - \theta_2(u))^2 = \left(\int_t^u \theta_2'(x)\, dx\right)^2 \le (u - t)\int_t^u (\theta_2'(x))^2\, dx. \tag{6}$$

Using (2), (6), and Jensen's inequality we obtain

$$\begin{aligned} &\int (\theta_2(t) - C_{ni})^2 \chi_{ni}(t)\, dt \\ &\quad = \int \left(a_n \int (\theta_2(t) - \theta_2(u))\chi_{ni}(u)\, du\right)^2 \chi_{ni}(t)\, dt \\ &\quad \le a_n \int\int (\theta_2(t) - \theta_2(u))^2 \chi_{ni}(u)\, du\, \chi_{ni}(t)\, dt \\ &\quad \le b_n^2 \int (\theta_2'(x))^2 \chi_{ni}(x)\, dx \end{aligned} \tag{7}$$

and

$$\begin{aligned} (C_{ni+1} - C_{ni})^2 &= \left(a_n^2 \int\int (\theta_2(t) - \theta_2(u))\chi_{ni+1}(u)\chi_{ni}(t)\, du\, dt\right)^2 \\ &\le a_n^2 \int\int (\theta_2(t) - \theta_2(u))^2 \chi_{ni+1}(u)\chi_{ni}(t)\, du\, dt \\ &\le 2b_n \int (\theta_2'(x))^2 (\chi_{ni+1}(x) + \chi_{ni}(x))\, dx. \end{aligned} \tag{8}$$

Combining (4) and (8) shows that for some constant $C$

$$\sum_{i=1}^{a_n-1} E_\theta(Y_{ni+1} - Y_{ni})^2 \le C(n^{-1}b_n^{-2} + b_n). \tag{9}$$

Next define

$$\overline{V}_n = \sum_{i=1}^{a_n} Y_{ni}\chi_{ni} - U_n \tag{10}$$

and

$$\tilde{V}_n = \sum_{i=1}^{a_n} C_{ni}\chi_{ni}. \tag{11}$$

It follows from (1) and (9) that

$$\begin{aligned} E_\theta \int (V_n(t) - \overline{V}_n(t))^2\,dt &\le \sum_{i=1}^{a_n-1} E_\theta(Y_{ni+1} - Y_{ni})^2 b_n \\ &\le C(n^{-1}b_n^{-1} + b_n^2) = O(n^{-2/3}) \end{aligned} \tag{12}$$

and from (1), (4), and (5) that

$$E_\theta \int (\overline{V}_n(t) - \tilde{V}_n(t))^2\,dt = O(n^{-2/3}). \tag{13}$$

Furthermore by (1) and (7)

$$\int (\tilde{V}_n(t) - \theta_2(t))^2\,dt = O(n^{-2/3}). \tag{14}$$

Combining (12) to (14) gives the desired result. □

It follows from the above that

$$\tilde{Z}_n = \overline{U}_n + \frac{1}{n}\left(\sum_{j=1}^{k_n} L_n(Y_j - \overline{U}_n - V_{n2}(T_j)) + \sum_{j=k_n+1}^{n} L_n(Y_j - \overline{U}_n - V_{n1}(T_j))\right)$$

is asymptotically linear and adaptive, where $\langle \overline{U}_n \rangle$ is a discrete version of $\langle U_n \rangle = \langle 1/n\Sigma_{j=1}^n Y_j \rangle$, $\langle L_n \rangle$ is as described in (3.8) and (3.9), $k_n$ is the integer part of $n/2$, and $\langle V_{n1} \rangle$ and $\langle V_{n2} \rangle$ are the versions of $\langle V_n \rangle$ based on the first $k_n$ and the second $n - k_n$ observations, respectively. In particular, we obtain

$$\tilde{Z}_n = \overline{U}_n + \frac{1}{n}\left(\sum_{j=1}^{k_n} \operatorname{sign}\left(Y_j - \overline{U}_n - V_{n2}(T_j)\right) + \sum_{j=k_n+1}^{n} \operatorname{sign}\left(Y_j - \overline{U}_n - V_{n1}(T_j)\right)\right)$$

is an adaptive estimate, if $g$ is the double exponential density. This provides an example of a model for which Bickel's condition $S^*$ does not hold (see (3.6)) but adaptive estimates exist. Of course, Bickel's convexity condition $C$ does not hold for this example.

We conclude this section with remarks dealing with extensions of our model (3.5).

REMARK 4. The above results are easily extended to cover the model

$$Y = \sum_{i=1}^{p} \beta_i h_i(T) + w(T) + \varepsilon, \tag{3.11}$$

where $\varepsilon$ and $T$ are as in (3.5), $\beta_1, \dots, \beta_p$ are unknown real numbers, $h_1, \dots, h_p$ are square integrable functions on $[0,1]$ such that the matrix $H = [\int h_i(t) h_j(t)\, dt]$ is the identity matrix, and $w$ is an unknown function on $[0,1]$ that has a square integrable derivative and satisfies $\int w(t) h_i(t)\, dt = 0$, $i = 1, \dots, p$. For this model $\theta_1 = \langle \beta_1, \dots, \beta_p \rangle$ and $\theta_2 = w$. Note that (3.5) is the special case $p = 1$ and $h_1 = 1$. Let $h = \langle h_1, \dots, h_p \rangle$. Easy calculations show that the necessary condition for adaptive estimation holds and that

$$l_*(x, \theta) = h(x_2) L\big(x_1 - \theta_1^T h(x_2) - \theta_2(x_2)\big), \qquad x \in S.$$

Also, verify that

$$E_\theta \left\| \frac{1}{n} \sum_{j=1}^{n} Y_j h(T_j) - \theta_1 \right\|^2 = O\left(\frac{1}{n}\right)$$

and

$$E_\theta \int_0^1 (W_n(t) - \theta_2(t))^2\, dt = O(n^{-2/3}),$$

where $W_n = \tilde{V}_n - \sum_{i=1}^{p} \int_0^1 \tilde{V}_n(t) h_i(t)\, dt\, h_i$ with $\tilde{V}_n = V_n + 1/n \sum_{j=1}^{n} Y_j$ and $V_n$ as in (3.10). It follows as above that

$$\overline{U}_n + \frac{1}{n} \left[ \sum_{j=1}^{k_n} h(X_j) L_n\big(Y_j - \overline{U}_n - W_{n2}(T_j)\big) + \sum_{j=k_n+1}^{n} h(X_j) L_n\big(Y_j - \overline{U}_n - W_{n1}(T_j)\big) \right]$$

is an adaptive estimate, where $\langle \overline{U}_n \rangle$ is a discrete version of $\langle 1/n \sum_{j=1}^{n} Y_j h(T_j) \rangle$, $\langle L_n \rangle$ is an appropriate modification of $L$ in the spirit of (3.8) and (3.9), and $\langle W_{n1} \rangle$ and $\langle W_{n2} \rangle$ are the versions of $\langle W_n \rangle$ which are based on the first $k_n$ and second $n - k_n$ observations, respectively.

REMARK 5. A more realistic version of (3.11) is to assume that also the density $g$ is unknown, but satisfies (3.1) to (3.3). The nuisance parameter in this case is $\theta_2 = \langle w, g \rangle$ and the necessary condition for adaptive estimation holds if and only if $\int h_i(t)\, dt = 0$ for all $i = 1, \dots, p$. This model deserves further investigation. We believe that asymptotically linear estimates can be constructed for this more general model.

**Acknowledgments.** This paper is based on my doctoral dissertation. I am grateful to my advisor, Professor Václav Fabian, for his generous guidance and encouragement. I am also grateful to Professor P. J. Bickel for the Engle, Granger, Rice, and Weiss reference and to the Associate Editor for providing several references.

## REFERENCES

BEGUN, J., HALL, W., HUANG, W. and WELLNER, J. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* **11** 432–452.

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

ENGLE, R. F., GRANGER, C. W. J., RICE, J. and WEISS, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.* **81** 310–320.

FABIAN, V. and HANNAN, J. (1982). On estimation and adaptive estimation for locally asymptotically normal families. *Z. Wahrsch. verw. Gebiete* **59** 459–478.

HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 175–194. Univ. California Press.

HUANG, W.-M. (1982). Parameter estimation when there are nuisance functions. Ph.D. dissertation, Univ. Rochester.

RUDIN, W. (1974). *Real and Complex Analysis*. 2nd ed. McGraw-Hill, New York.

STEIN, C. (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 187–195. Univ. California Press.

STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13** 689–705.

WAHBA, G. (1984). Partial spline models for the semiparametric estimation of functions of several variables. *Proceedings of the Seminar on Statistical Analysis of Time Series*. Japan–U.S. Joint Seminar.

DEPARTMENT OF MATHEMATICAL SCIENCES
STATE UNIVERSITY OF NEW YORK
BINGHAMTON, NEW YORK 13901