The Annals of Statistics
1997, Vol. 25, No. 1, 244–262

SEMIPARAMETRIC INFERENCE IN A PARTIAL LINEAR MODEL

BY P. K. BHATTACHARYA AND PENG-LIANG ZHAO

University of California and Merck Research Laboratories

In a partial linear model, the dependence of a response variate Y on covariates (W, X) is given by

$$Y = W\beta + \eta(X) + \mathscr{E},$$

where $\mathscr{E}$ is independent of (W, X) with densities g and f, respectively. In this paper an asymptotically efficient estimator of β is constructed solely under mild smoothness assumptions on the unknown η, f and g, thereby removing the assumption of finite residual variance on which all least-squares-type estimators available in the literature are based.

1. Introduction. In a partial linear model, one observes $Z_i = (W_i, X_i, Y_i)$, $i = 1, \dots, n$, of which the Y_i's are response variates depending on covariates (W_i, X_i) through the relationship

$$(1) \qquad Y_i = W_i\beta + \eta(X_i) + \mathscr{E}_i,$$

where $(W_i, X_i, \mathscr{E}_i)$ are iid as $(W, X, \mathscr{E})$. The covariate (W, X) is $[c, d] \times [0, 1]$-valued with joint pdf $f(w, x)$ with respect to $\nu = \nu_1 \times \nu_2$, where ν_1 is a σ-finite measure on $[c, d]$, ν_2 is the Lebesgue measure on $[0, 1]$ and the residual $\mathscr{E}$ has a Lebesgue density g having a finite and positive Fisher information. Moreover, $\mathscr{E}$ is assumed to be independent of (W, X). The Lebesgue density of the marginal distribution of X is denoted by f_0. If we let $\gamma = (\eta, f, g)$, then model (1) means that $Z_i = (W_i, X_i, Y_i)$ are iid as $Z = (W, X, Y)$ having pdf

$$p(z; \beta, \gamma) = f(w, x)\, g(y - w\beta - \eta(x)), \qquad \beta \in R^1 \text{ and } \gamma \in \Gamma,$$

with respect to $\mu = \nu_1 \times \nu_2 \times$ Lebesgue measure, where Γ is the set of all (η, f, g) satisfying regularity conditions stated in Section 2. The probability measure of Z thus defined will be denoted by $P_{\beta,\gamma}(\cdot)$ or $P(\cdot\,; \beta, \gamma)$. The problem is to estimate the parameter β, treating γ as a nuisance parameter.

Assuming finite variance of the residual $\mathscr{E}$, several estimators of β have been constructed by least-squares methods using spline and kernel smoothing, and properties such as $\sqrt{n}$-consistency, asymptotic normality and asymptotic efficiency in case of Gaussian errors have also been established for some of these estimators. Basically, three approaches have been used to construct these estimators, namely, the penalized least-squares method used by Wahba (1984), Green, Jennison and Scheult (1985), Engle, Granger, Rice and Weiss (1986), Heckman (1986), Rice (1986), Shiau, Wahba and Johnson (1986), and

Received June 1993; revised March 1996.
AMS 1991 *subject classifications*. Primary 62J99, 62F12; secondary 62G07.
Key words and phrases. Partial linear model, semiparametric inference, effective information, efficient influence function, bandwidth-matched M-estimator, M-smoother.

Chen and Shiau (1991), the partial residuals method used by Denby (1984), Robinson (1988), Speckman (1988) and Cuzick (1992a) and the projection method used by Chen (1988). Although some of these estimators of β are asymptotically efficient under Gaussian errors, nothing is known about their properties for non-Gaussian errors. Cuzick (1992b) considered the case of unknown residual distribution and constructed an initial estimator of β from one part of the data and used it to construct an efficient estimator of β from the rest of the data. However, this method also requires finite variance since the initial estimator is of the least-squares type. Another efficient estimator of β for the case of unknown residual distribution with finite variance has been developed by Schick (1993). The issue of achieving the information bound in this and other non- and semiparametric models has been examined by Ritov and Bickel (1990).

In this paper our main goal is to construct an asymptotically efficient estimator of β for arbitrary residual distribution having pdf g with finite and positive Fisher information. This will be achieved solely under mild smoothness conditions and an identifiability condition on η, f and g and the assumption that f_0 is bounded away from 0, but without any moment assumption on the residual distribution g. In the process of constructing such an estimator of β, an estimator of η will also be constructed.

In Section 2, we begin by stating the regularity conditions and the formulas for effective score function, effective information and efficient influence function for estimating β when η, f and g are unknown. We then outline a strategy for constructing an estimator of β which attains the effective information limit, following a general scheme developed by Bickel (1982), Schick (1986) and Klaassen (1987). In order to carry out this program, we shall construct the following:

1. In Section 3, a $\sqrt{n}$-consistent estimator $\widetilde{\beta}_n$ of β will be constructed by a bandwidth-matched M-estimation procedure due to Bhattacharya (1989).
2. In Section 4, an estimator $\widehat{q}_n(x)$ of $q(x) = E(W|X = x)$ with a desired rate of a.s. uniform convergence will be constructed by the standard kernel method.
3. Estimators $\widehat{\eta}_n(x|\beta_n)$ of the location of the conditional distribution of $Y - W\beta_n$ given $X = x$ and $\widehat{\varphi}_n(t|\beta_n)$ of $\varphi(t|\beta_n) = -g'(t|\beta_n)/g(t|\beta_n)$, where $g(\cdot|\beta_n)$ is the pdf of $Y - W\beta_n - \eta(X|\beta_n)$, having suitable convergence properties for $\beta_n = \beta + O(n^{-1/2})$ as $n \to \infty$, will also be constructed in Section 4. Of these, $\widehat{\eta}_n$ will be constructed by the M-smoother technique [see Härdle, Janssen and Serfling (1988) and Hall and Jones (1990)] and $\widehat{\varphi}_n$ by the kernel method.

The desired properties of these estimators and the estimated effective influence function will also be established in Section 4. Finally, all this will be put together in Section 5 to construct an asymptotically efficient estimator of β. The results of a simulation study were reported in Bhattacharya and Zhao (1995). In these simulations, our procedure performed satisfactorily in a variety of situations based on samples of 100 observations.

2. Asymptotic lower bound in semiparametric formulation. We assume that in the model described by (1), the joint pdf f of (W, X), the pdf g of the residual $\mathscr{E}$ and the regression function η satisfy the following regularity conditions.

CONDITION A1. η belongs to $C_{\mathrm{L}}[0, 1]$, the space of Lipschitz-continuous functions on $[0, 1]$.

CONDITION A2. (a) $|f(w, x_1) - f(w, x_2)| \le c_0 |x_1 - x_2|$ for some c_0 and for all $w \in [c, d]$, $x_1, x_2 \in [0, 1]$.
(b) The marginal pdf $f_0(\cdot) = \int f(w, \cdot)\, d\nu_1(w)$ is bounded away from 0.
(c) $E[(W - q(X))^2] > 0$, where $q(x) = E(W|X = x)$.

CONDITION A3. (a) g has bounded derivative g' and finite and positive Fisher information, that is,

$$0 < I(g) = \int \big[\{g'(t)\}^2 / g(t)\big]\, dt < \infty.$$

(b) g is symmetric about 0 and is positive in a neighborhood of 0.

We emphasize that no moment condition has been imposed on the residual distribution.

REMARK 1. Condition A2(a) implies that f_0 is Lipschitz and is therefore bounded.

REMARK 2. Boundedness of g' is used in Lemma 4.2 for the purpose of applying Lemma 3.2 of Härdle, Janssen and Serfling (1988).

REMARK 3. In Condition A3(b), the symmetry of g serves as an identifiability condition to prevent $(\eta + a, g(\cdot + a))$ and (η, g) from resulting in the same model. We have used this condition in proving Lemma 4.2 and part (i) of Lemma 4.5. In Remark 8 following the proof of Lemma 4.2, we have indicated how this proof can be easily modified if Condition A3(b) is replaced by a slightly more general condition A3(b′) suggested by a referee. In the proof of Lemma 4.5(i), the symmetry condition can be avoided altogether, but at the expense of additional requirement that f_0, q and g have bounded second derivatives.

We first define two functions,

$$\rho(y) = -\tfrac{1}{2}\, g'(y)\, g^{-1/2}(y),$$
$$\varphi(y) = -\frac{g'(y)}{g(y)} = 2 g^{-1/2}(y)\, \rho(y), \tag{2}$$

for use throughout the paper. When $\gamma = (\eta, f, g)$ is known, the score function and the information for β are $\rho_\beta(z) = (\partial p^{1/2}/\partial\beta) = f^{1/2}(w, x)\, \rho(y - w\beta -$

$\eta(x))\,w$ and $I_0 = 4\|\rho_\beta\|_\mu^2 = E(W^2)\,I(g)$, respectively. In the following theorem, we state the efficient score function ρ^* when γ is unknown, from which the formulas for the effective information and efficient influence function are obtained. For proof, see Bickel, Klaassen, Ritov and Wellner (1992). We also refer to Bhattacharya and Zhao (1995) for a more easily accessible version of the proof.

THEOREM 1. *For estimating* β, *the efficient score function is*

$$\rho^*(z;\,\beta,\gamma) = f^{1/2}(w,x)\,\rho\big(y - w\beta - \eta(x)\big)\,\big(w - q(x)\big), \tag{3}$$

the effective information is

$$I_*(\beta,\gamma) := 4\|\rho^*(\cdot\,;\,\beta,\gamma)\|_\mu^2 = E\big[\{W - E(W|X)\}^2\big]\,I(g) \tag{4}$$

and the efficient influence function is

$$\begin{aligned} J(z;\,\beta,\gamma) &:= [I_*(\beta,\gamma)]^{-1}\,2\,p^{-1/2}(z;\,\beta,\gamma)\,\rho^*(z;\,\beta,\gamma) \\ &= [I_*(\beta,\gamma)]^{-1}\,[(w - q(x))\,\varphi(y - w\beta - \eta(x))], \end{aligned} \tag{5}$$

where $z = (w, x, y)$, $q(x) = E(W|X = x)$, $I(g) = \int[\{g'(t)\}^2/g(t)]\,dt$ *and* φ *is given by* (2).

REMARK 4. Theorem 1 generalizes in a straightforward manner to the case where W is a k-dimensional random vector taking values in $[c,d]^k$ and the linear part of the regression is $\beta^T W$ where $\beta \in R^k$ is to be estimated. Now $w\beta$ is replaced by $\beta^T w$ in all formulas, $q(x)$ is a vector function and $I_*(\beta,\gamma) = E[\{W - E(W|X)\}\{W - E(W|X)\}^T]I(g)$, the matrix $E[\{\cdots\}\{\cdots\}^T]$ being assumed to be positive definite.

Our goal is to construct an asymptotically linear estimator $\widehat{\beta}_n$, that is, one which approximates

$$\beta_n^* := \beta + n^{-1}\sum_{i=1}^{n} J(Z_i;\,\beta,\gamma) \tag{6}$$

at the $o_p(n^{-1/2})$ rate under $P_{\beta,\gamma}$. Such an estimator is a least dispersed regular estimator of β in the presence of the nuisance parameter γ. All construction schemes developed by Bickel (1982), Schick (1986, 1987, 1993) and Klaassen (1987) [also see Bickel, Klaassen, Ritov and Wellner (1992)] follow this principle. Here we shall use the construction given by Schick (1986) for which the conditions are satisfied within our framework.

For this construction of an asymptotically linear estimator, we need the following condition on the effective score function.

CONDITION B1. The map $\beta \in R^1 \mapsto \rho^*(\cdot\,;\,\beta,\gamma) \in L^2(\mu)$ is continuous for every $\gamma = (\eta, f, g) \in \Gamma$.

We also need an initial estimator $\widetilde{\beta}_n$ of β and an estimator $\widehat{J}_n(\cdot\,;\,\cdot\,,\mathbf{Z}(n)) \equiv \widehat{J}_n(\cdot\,;\,\cdot\,,Z_1,\dots,Z_n)$ of $J(\cdot\,;\,\cdot\,,\gamma)$ satisfying the following conditions under $P_{\beta,\gamma}$.

CONDITION B2. $n^{1/2}(\widetilde{\beta}_n - \beta) = O_p(1)$.

CONDITION B3. For every sequence $\beta_n = \beta + O(n^{-1/2})$,

$$n^{1/2}\int \widehat{J}_n(z;\,\beta_n,\mathbf{Z}(n))\,p(z;\,\beta_n,\gamma)\,d\mu = o_p(1),$$

and

$$\int |\widehat{J}_n(z;\,\beta_n,\mathbf{Z}(n)) - J(z;\,\beta_n,\gamma)|^2\,p(z;\,\beta_n,\gamma)\,d\mu = o_p(1).$$

The construction of an asymptotically linear estimator $\widehat{\beta}_n$ of β essentially involves using $\widetilde{\beta}_n$ for β and $\widehat{J}_n$ for J on the right-hand side of (6). For technical reasons, we actually use a discretized version of $\widetilde{\beta}_n$, namely, the point $\overline{\beta}_n$ in the set $\{jn^{-1/2}\colon\, j \text{ is an integer}\}$ which is closest to $\widetilde{\beta}_n$ [see Bickel (1982)], and use $\widehat{J}_{n2}(Z_i;\,\overline{\beta}_n,\mathbf{Z}(n,2))$ for $i = 1,\dots,[n/2]$ and $\widehat{J}_{n1}(Z_i;\,\overline{\beta}_n,\mathbf{Z}(n,1))$ for $i = [n/2]+1,\dots,n$ in (6), where $\widehat{J}_{n1}$ is based on the first half of the sample $\mathbf{Z}(n,1) = (Z_1,\dots,Z_{[n/2]})$, while $\widehat{J}_{n2}$ is based on the remaining half of the sample $\mathbf{Z}(n,2) = (Z_{[n/2]+1},\dots,Z_n)$. The resulting estimator is

$$\widehat{\beta}_n = \overline{\beta}_n + \frac{1}{n}\left[\sum_{i=1}^{[n/2]} \widehat{J}_{n2}\big(Z_i;\,\overline{\beta}_n,\mathbf{Z}(n,2)\big) + \sum_{i=[n/2]+1}^{n} \widehat{J}_{n1}\big(Z_i;\,\overline{\beta}_n,\mathbf{Z}(n,1)\big)\right]. \tag{7}$$

The following theorem is a restatement of Schick's (1986) Theorem 1 in which the asymptotic linearity of the above estimator was established.

THEOREM 2. *If Condition* B1 *holds and if the estimators* $\widetilde{\beta}_n$ *and* $\widehat{J}_n$ *satisfy Conditions* B2 *and* B3, *respectively*, *then the estimator* $\widehat{\beta}_n$ *given by* (7) *is asymptotically linear*.

We now proceed to construct an initial estimator of β satisfying Condition B2 in Section 3 and an estimator of the efficient influence function J satisfying Condition B3 in Section 4. The latter objective will be accomplished by constructing estimators of the functions η, q and φ with appropriate rates of convergence and then using these estimators in (5).

3. A bandwidth-matched *M*-estimator for β. Due to the smoothness of η in model (1), we should have $Y_i - Y_j \approx (W_i - W_j)\beta + (\mathscr{E}_i - \mathscr{E}_j)$ for any pair with $X_i \approx X_j$. In the spirit of M-estimation, this motivates an estimator $\widetilde{\beta}_n$ which minimizes

$$\{n(n-1)b_n\}^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} K((X_i - X_j)/b_n)\,\tau(Y_i - Y_j - (W_i - W_j)\beta),$$

with respect to β, where τ is a symmetric convex function on R^1 which is minimized at 0, $K\colon R^1 \mapsto R^1$ is a kernel function and $\{b_n\}$ is a positive sequence tending to 0 as $n \to \infty$. For $K(u) = \mathrm{I}(|u| \le 1)$, such an estimator is based on all pairs (i, j) for which X_i, X_j are within a bandwidth b_n of one another. The term bandwidth-matched M-estimation comes from this. Such an estimator has been proposed by Bhattacharya (1989) for the special case when W in model (1) is 0–1-valued.

If the function τ is a.e. differentiable, then taking $\psi = \tau'$ (which is odd and monotone nondecreasing), the estimator $\tilde{\beta}_n$ is equivalently defined as the solution of $S_n(\beta) = 0$, where

$$S_n(\beta) = \{n(n-1)b_n\}^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} K((X_i - X_j)/b_n) \times \psi(Y_i - Y_j - (W_i - W_j)\beta)(W_i - W_j). \tag{8}$$

Since $S_n(\beta)$ may have many zero crossings, we formally define

$$\begin{aligned} \tilde{\beta}_n &= (\beta_n^- + \beta_n^+)/2, \\ \beta_n^- &= \sup\{\beta\colon S_n(\beta) > 0\}, \\ \beta_n^+ &= \inf\{\beta\colon S_n(\beta) < 0\}. \end{aligned} \tag{9}$$

For $\psi(y) = y$ which corresponds to $\tau(y) = y^2$, the above estimator would be another least-squares-type estimator, but since we are not assuming finite variance of the residuals, we shall choose ψ in (8) from a suitable class of bounded functions to make the estimator robust. Let ψ be a bounded, odd and nondecreasing function, and define

$$L(u) = \int \psi(u - y)\, g(y)\, dy, \qquad u \in R.$$

Note that $\int L(u)\, g(u)\, du = E[\psi(\mathscr{E}_1 - \mathscr{E}_2)] = 0$.

We shall choose the kernel K subject to the following condition.

CONDITION K1. The kernel K is a bounded symmetric pdf on $[-1, 1]$.

In the following theorem, we establish the asymptotic normality of $\sqrt{n}\,(\tilde{\beta}_n - \beta)$, thereby showing that $\tilde{\beta}_n$ satisfies Condition B2.

THEOREM 3. *Suppose that Conditions* A1 *and* A2(a) *and* (c) *hold. Assume that L has a bounded continuous derivative with* $\int L'(y)\, g(y)\, dy > 0$ *and that the kernel K satisfies Condition* K1. *If* $nb_n \to \infty$ *and* $nb_n^2 \to 0$, *then the estimator* $\tilde{\beta}_n$ *defined by* (8) *and* (9) *satisfies*

$$\sqrt{n}\,(\tilde{\beta}_n - \beta) \longrightarrow_{\mathscr{L}} \mathscr{N}(0, \zeta_0/(\alpha_o)^2) \quad \text{as } n \to \infty,$$

where

$$\alpha_0 = E[L'(\mathscr{E}_1)]\, E[f_0(X_1)\,\mathrm{Var}(W_1|X_1)] > 0, \tag{10}$$

$$\zeta_0 = E[L^2(\mathscr{E}_1)]\, E[f_0^2(X_1)\,\mathrm{Var}(W_1|X_1)] > 0. \tag{11}$$

REMARK 5. The condition on L is satisfied if ψ has a positive, bounded and continuous derivative, such as $\psi(y) = \arctan(y)$. For $\psi(y) = \mathrm{sign}(y)$, the condition holds if g is continuous, because $L(u) = 2G(u) - 1$. In this case, $E\big[L'(\mathscr{E}_1)\big] = 2\,||g||_\lambda^2$ and $E[L^2(\mathscr{E}_1)] = 1/3$.

PROOF. It suffices to show that, for each $t \in R$,

$$\sqrt{n}\, S_n(\beta + n^{-1/2}t) \to_{\mathscr{L}} \mathscr{N}(-2\,\alpha_0\, t, 4\zeta_0). \tag{12}$$

Indeed, the monotonicity of $\psi(\cdot)$ implies that of $S_n(\cdot)$. Moreover, by Theorem 1 of Hodges and Lehmann (1963), both β_n^- and β_n^+ have continuous cdf's. Hence

$$\begin{aligned} P[n^{1/2}(\beta_n^- - \beta) \le t] &= P[S_n(\beta + n^{-1/2}t) \le 0] \\ &= P\left[\frac{\sqrt{n}S_n(\beta + n^{-1/2}t) + 2\,\alpha_0\, t}{\sqrt{4\zeta_0}} \le \frac{2\,\alpha_0\, t}{\sqrt{4\zeta_0}}\right] \to \Phi\left(\frac{t\,\alpha_0}{\sqrt{\zeta_0}}\right) \end{aligned}$$

by (12), Φ being the cdf of $N(0, 1)$. Since the same holds for $n^{1/2}(\beta_n^+ - \beta)$ and $\beta_n^- \le \widetilde{\beta}_n \le \beta_n^+$ with probability 1, the theorem follows.

To verify (12), fix $t \in R$, let $V_i = (W_i, X_i, \mathscr{E}_i)$ and note that $S_n(\beta + n^{-1/2}t)$ is a U-statistic with symmetric kernel

$$\begin{aligned} U_n(V_i, V_j) &= b_n^{-1}K((X_i - X_j)/b_n)(W_i - W_j) \\ &\quad \times \psi(\eta(X_i) - \eta(X_j) - n^{-1/2}t(W_i - W_j) + \mathscr{E}_i - \mathscr{E}_j). \end{aligned}$$

Standard U-statistic argument now leads to

$$S_n(\beta + n^{-1/2}t) = \mu_n + 2n^{-1}\sum_{i=1}^{n}\{\pi_n(V_i) - \mu_n\} + R_n, \tag{13}$$

where

$$\begin{aligned} \pi_n(V_1) &= E[U_n(V_1, V_2)|V_1] \\ &= b_n^{-1}\int_c^d\int_0^1 K((X_1 - x)/b_n) \\ &\qquad \times (W_1 - w)L(\mathscr{E}_1 + \eta(X_1) - \eta(x) - n^{-1/2}t(W_1 - w)) \\ &\qquad \times f(w, x)\, d\nu_1(w)\, dx, \end{aligned} \tag{14}$$

$$\mu_n = E[U_n(V_1, V_2)] = E[\pi_n(V_1)],$$

and

$$E[R_n^2] \le \binom{n}{2}^{-1} E[U_n^2(V_1, V_2)]$$
$$\le M(nb_n)^{-2} \int_0^1 \int_0^1 I_{[-b_n, b_n]}(x_1 - x_2)\, dx_1\, dx_2$$
$$= O(n^{-2} b_n^{-1}) = o(n^{-1}),$$

since ψ, W_i and K are bounded, and K vanishes outside $[-1, 1]$. Here, as well as in what follows, M is a generic constant. Next let

$$\pi_0(V_1) = L(\mathscr{E}_1) \int (W_1 - w) f(w, X_1)\, d\nu_1(w)$$
$$= L(\mathscr{E}_1) f_0(X_1)\{W_1 - E(W_1|X_1)\},$$
$$\pi_1(V_1) = L'(\mathscr{E}_1) \int (W_1 - w)^2 f(w, X_1)\, d\nu_1(w)$$
$$= L'(\mathscr{E}_1) f_0(X_1)\{W_1^2 - 2W_1 E(W_1|X_1) + E(W_1^2|X_1)\},$$

and note that $E[\pi_0(V_1)] = 0$, $E[\pi_0^2(V_1)] = \zeta_0$ and $E[\pi_1(V_1)] = 2\,\alpha_0$.

The main thing is to show that $\pi_n(V_1) - [\pi_0(V_1) - n^{-1/2} t\, \pi_1(V_1)]$ is small enough for our purpose. To this end, transform x to $u = (X_1 - x)/b_n$ in (14) and let

$$\Delta_{n1} = \eta(X_1) - \eta(X_1 - b_n u),$$
$$\Delta_{n2} = \Delta_{n1} - n^{-1/2} t(W_1 - w),$$
$$\Delta_{n3} = f(w, X_1 - b_n u) - f(w, X_1).$$

Note that u is integrated over $A_n = A_n(X_1) = [b_n^{-1}(X_1 - 1), b_n^{-1} X_1] \cap [-1, 1]$, so that $I_n = \int_{A_n} K(u)\, du \in [0, 1]$ and $I_n = 1$ for $b_n \le X_1 \le 1 - b_n$. Hence $E[(1 - I_n)^r] \le M\{1 - P(b_n \le X_1 \le 1 - b_n)\} \le M b_n$ for $r > 0$. Moreover, $|\Delta_{n1}| \le M b_n$, $|\Delta_{n2}| \le M n^{-1/2}$ and $|\Delta_{n3}| \le M b_n$ a.s., by Conditions A1 and A2(a). We now have

$$\pi_n(V_1) = \iint K(u)(W_1 - w)\, L(\mathscr{E}_1 + \Delta_{n2})\{f(w, X_1) + \Delta_{n3}\}\, d\nu_1(w)\, du$$
$$= \iint K(u)(W_1 - w)\, [L(\mathscr{E}_1) + \Delta_{n2} L'(\mathscr{E}_1) + \Delta_{n2}\{L'(\mathscr{E}_{1n}) - L'(\mathscr{E}_1)\}]$$
$$\times \{f(w, X_1) + \Delta_{n3}\}\, d\nu_1(w)\, du,$$

where $\mathscr{E}_{1n}$ lies between $\mathscr{E}_1$ and $\mathscr{E}_1 + \Delta_{n2}$, so that $L'(\mathscr{E}_{1n}) - L'(\mathscr{E}_1) \to 0$ a.s. Rearrangement of terms in the last expression leads to

$$\pi_n(V_1) = \pi_0(V_1) - n^{-1/2} t\, \pi_1(V_1) + R_{n1} + R_{n2} + R_{n3},$$

where

$$R_{n1} = [\pi_0(V_1) - n^{-1/2}t\,\pi_1(V_1)](1 - I_n),$$

$$R_{n2} = \iint K(u)(W_1 - w)\big[\{L(\mathscr{E}_1) - n^{-1/2}t(W_1 - w)L'(\mathscr{E}_1)\}\Delta_{n3} + \Delta_{n1}L'(\mathscr{E}_{1n})f(w, X_1 - b_n u)\big]\,d\nu_1(w)\,du,$$

$$R_{n3} = -n^{1/2}t \iint K(u)(W_1 - w)^2\{L'(\mathscr{E}_{1n}) - L'(\mathscr{E}_1)\} \times f(w, X_1 - b_n u)\,d\nu_1(w)\,du.$$

Since $\pi_0(V_1)$ and $\pi_1(V_1)$ are bounded, $|E[n^{1/2}R_{n1}]| \le Mn^{1/2}b_n = o(1)$. For R_{n2}, the integrand is bounded by Mb_n, so that $E[n^{1/2}R_{n2}] = o(1)$. Next note that since L' is bounded and continuous, $E[n^{1/2}R_{n3}]$ is a five-fold integral of which the integrand is bounded and converges to 0 a.e. Hence $E[n^{1/2}R_{n3}] = o(1)$ by the Lebesgue dominated convergence theorem. Thus

$$n^{1/2}\mu_n = n^{1/2}E[\pi_0(V_1)] - t\,E[\pi_1(V_1)] + \sum_{j=1}^{3} E[n^{1/2}R_{nj}] = -2\,\alpha_0 t + o(1).$$

Finally, $E[R_{nj}^2] = o(1)$ for $j = 1, 2, 3$ follows by straightforward argument using the various boundedness properties mentioned above. Hence

$$\begin{aligned} &E[\{\pi_n(V_1) - \mu_n - \pi_0(V_1)\}^2] \\ &\quad = E\left[\left\{n^{-1/2}t\,\pi_1(V_1) + \mu_n - \sum_{j=1}^{3} R_{nj}\right\}^2\right] \\ &\quad \le 5\left[n^{-1}t^2E(\pi_1^2(V_1)) + n^{-1}(2\,\alpha_0 t + o(1))^2 + \sum_{j=1}^{3} E(R_{nj}^2)\right] = o(1). \end{aligned}$$

Putting these results in (13), we have

$$\begin{aligned} n^{1/2}S_n(\beta + n^{-1/2}t) &= n^{1/2}\mu_n + 2n^{-1/2}\sum_{i=1}^{n}\{\pi_n(V_i) - \mu_n\} + n^{1/2}R_n \\ &= -2\,\alpha_0 t + 2n^{-1/2}\sum_{i=1}^{n}\pi_0(V_i) + o_p(1), \end{aligned}$$

from which (12) follows. □

REMARK 6. If $\eta \in C^2[0, 1]$ and g is continuous, then Theorem 3 holds for $\psi(y) = \operatorname{sign}(y)$ and b_n chosen so that $nb_n \to \infty$ and $nb_n^4 \to 0$. In this case, the bandwidth h_n used in Section 4 can be taken to be the same as the bandwidth b_n chosen in this section. A simulation study reported in Bhattacharya and Zhao (1995) was carried out like this.

REMARK 7. The bandwidth-matched M-estimation method discussed in this section does not extend to the case of multidimensional W in its full generality. A different method, using a result of Pollard (1991), has been developed for such general multidimensional W by one of the authors [Zhao (1995)]. However, the method discussed here has a limited extension to the case of k-dimensional W having its support on the Cartesian product $W_1 \times \cdots \times W_k$ of finite sets, where $W_r = \{1, \ldots, \alpha_r\}$. This includes the case of 2^k-factorial observational studies in which the rth component β_r of the parameter vector β represents the main effect of the rth factor, ignoring all interactions, and $\eta(\cdot)$ represents the nonparametric regression of the observable Y on a self-selected covariate X. For such an extension, we define $I_r(W_i, W_j)$ to be the indicator of the event $\{W_{is} = W_{js}$ for all $s \neq r\}$ and let

$$S_{nr}(\beta_r) = \{n(n-1)b_n\}^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{X_i - X_j}{b_n}\right) \psi(Y_i - Y_j - (W_{ir} - W_{jr})\beta_r) \times (W_{ir} - W_{jr}) I_r(W_i, W_j)$$

and estimate β_r by the zero crossing $\widetilde{\beta}_{nr}$ of $S_{nr}(\beta_r)$ for $r = 1, \ldots, k$. Then the proof of Theorem 3 extends easily, showing that $\sqrt{n}\,(\widetilde{\beta}_{nr} - \beta_r)$ is asymptotically normal with mean 0 for $r = 1, \ldots, k$, so that $\widetilde{\beta}_n = (\widetilde{\beta}_{n1}, \ldots, \widetilde{\beta}_{nk})^T$ is $\sqrt{n}$-consistent. The joint asymptotic distribution is not needed for our purpose, although the asymptotic covariance matrix can be easily calculated.

4. Estimation of efficient influence function. The efficient influence function $J(z; \beta, \gamma)$ is given by (5). It involves $q(x)$, $\eta(x)$, $\varphi(y - w\beta - \eta(x))$ and $I_*(\beta, \gamma)$, so we have to estimate these quantities. Of these, $q(x) = E(W|X = x)$ can be estimated in a straightforward manner by a kernel estimator

$$\widehat{q}_n(x) = \left[\sum_{i=1}^{n} K((X_i - x)/h_n)\, W_i\right] \Big/ \left[\sum_{i=1}^{n} K((X_i - x)/h_n)\right]. \tag{15}$$

The following is a well-known result in the theory of nonparametric regression, so we state it without proof.

LEMMA 4.1. *Suppose that Condition* A2(a) *and* (b) *holds and that the kernel* K *in* (15) *satisfies Condition* K1. *If* $h_n \to 0$ *and* $nh_n/\log n \to \infty$ *as* $n \to \infty$, *and if we let* $r_n = (nh_n/\log n)^{-1/2} + h_n$, *then*

$$\sup_{x \in [0,1]} \left|\widehat{q}_n(x) - q(x)\right| = O_{\text{a.s.}}(r_n) \quad \text{under } P_{\beta,\gamma}.$$

To estimate the other quantities, first suppose that β is known. Since $\eta(x)$ is the location of $Y - W\beta$ given $X = x$ and $\varphi(t) = -g'(t)/g(t)$, where g is the pdf of $Y - W\beta - \eta(x)$, we could use the M-smoothing technique to estimate $\eta(x)$ by the solution $\widetilde{\eta}_n(x|\beta)$ of

$$0 = D_n(t \mid x; \beta) = (nh_n)^{-1} \sum_{i=1}^{n} K((X_i - x)/h_n)\, \chi_c(Y_i - W_i\beta - t), \tag{16}$$

with the same kernel K as above and the Huber function

$$\chi_c(u) = u\,\mathrm{I}(|u| \le c) + c\{\mathrm{I}(u > c) - \mathrm{I}(u < -c)\}, \qquad c > 0,$$

and then estimate $\varphi(y - w\beta - \eta(x))$ by $\widehat{\varphi}_n(y - w\beta - \widetilde{\eta}_n(x|\beta)|\beta)$, where $\widehat{\varphi}_n(t\,|\,\beta)$ is constructed from the estimated residuals

$$\widehat{\mathscr{E}_i}(\beta) = Y_i - W_i\beta - \widetilde{\eta}_n(X_i\,|\,\beta) \tag{17}$$

by the kernel method as follows. Let

$$\begin{aligned} \widehat{g}_n(t\,|\,\beta) &= a_n + (na_n)^{-1} \sum_{i=1}^{n} K_0((t - \widehat{\mathscr{E}_i}(\beta))/a_n), \\ \widehat{g}_n'(t\,|\,\beta) &= (\partial/\partial t)\widehat{g}_n(t\,|\,\beta), \end{aligned} \tag{18}$$

using a logistic kernel $K_o(u) = \exp(u)/[1+\exp(u)]^2$ and a bandwidth a_n with $a_n \to 0$ and $na_n^8 \to \infty$. Since g is symmetric, $\varphi = -g'/g$ is odd, we would then estimate $\varphi(t)$ by

$$\widehat{\varphi}_n(t\,|\,\beta) = -\frac{1}{2}\left[\frac{\widehat{g}_n'(t\,|\,\beta)}{\widehat{g}_n(t\,|\,\beta)} - \frac{\widehat{g}_n'(-t\,|\,\beta)}{\widehat{g}_n(-t\,|\,\beta)}\right]. \tag{19}$$

Finally, if we let

$$H(z;\, \beta, \gamma) = (w - q(x))\,\varphi(y - w\beta - \eta(x)), \tag{20}$$

then $I_*(\beta, \gamma) = E[H^2(Z;\, \beta, \gamma)]$, so with β known, a natural estimator of $I_*(\beta, \gamma)$ would be

$$\widehat{I}_n(\beta, \mathbf{Z}(n)) = n^{-1} \sum_{i=1}^{n} \widehat{H}_n^2(Z_i;\, \beta, \mathbf{Z}(n)), \tag{21}$$

where

$$\widehat{H}_n(z;\, \beta, \mathbf{Z}(n)) = (w - \widehat{q}_n(x))\,\widehat{\varphi}_n(y - w\beta - \widetilde{\eta}_n(x|\beta)\,|\,\beta). \tag{22}$$

We are now going to implement the above procedures by replacing β throughout by the discretized version $\overline{\beta}_n$ of its estimator $\widetilde{\beta}_n$ obtained in Section 3.

For notational convenience, let β' be arbitrary and let $\widetilde{\eta}_n(x|\beta')$ denote the solution of $D_n(t\,|\,x;\, \beta') = 0$, replacing β by β' in (16). Also replace β by β' in (17), (18), (19), (20), (21) and (22) to define $\widehat{\mathscr{E}_i}(\beta')$, $\widehat{g}_n(t\,|\,\beta')$, $\widehat{g}_n'(t\,|\,\beta')$, $\widehat{\varphi}_n(t\,|\,\beta')$, $H(z;\, \beta', \gamma)$, $\widehat{H}_n(z;\, \beta', \mathbf{Z}(n))$ and $\widehat{I}_n(\beta', \mathbf{Z}(n))$ for arbitrary β'. The estimated efficient influence function is thus given by

$$\widehat{J}_n(z;\, \beta', \mathbf{Z}(n)) = [\widehat{I}_n(\beta', \mathbf{Z}(n))]^{-1}\,\widehat{H}_n(z;\, \beta', \mathbf{Z}(n)). \tag{23}$$

We first establish a rate of a.s. uniform convergence for the estimator $\widetilde{\eta}_n(x|\beta_n)$ of $\eta(x)$ with $\beta_n = \beta + O(n^{-1/2})$.

LEMMA 4.2. *Assume that Conditions* A1, A2(a) *and* (b) *and* A3 *hold, and let the kernel* K *in* (16) *satisfy Condition* K1. *If* $h_n \to 0$ *and* $nh_n/\log n \to \infty$ *as* $n \to \infty$, *then, for* $\beta_n = \beta + O(n^{-1/2})$ *and* $r_n = (nh_n/\log n)^{-1/2} + h_n$,

$$\sup_{x\in[0,1]} \mid \tilde{\eta}_n(x|\beta_n) - \eta(x) \mid = O_{\text{a.s.}}(r_n) \quad \text{under } P_{\beta,\gamma}.$$

PROOF. Let $D(t\mid x) = \int \chi_c(u-t)f_o(x)g(u-\eta(x))\,du$. Then, by Lemma 3.2 of Härdle, Janssen and Serfling (1988),

$$\sup_{t,x} |D_n(t\mid x;\beta) - D(t\mid x)| = O_{\text{a.s.}}(r_n).$$

Moreover, since the W_i's are bounded,

$$|\chi_c(u_1) - \chi_c(u_2)| \le |u_1 - u_2| \quad \text{for all } u_1, u_2$$

and

$$\hat{f}_{0n}(x) = (nh_n)^{-1}\sum_{i=1}^{n} K((X_i - x)/h_n) \to f_0(x) \quad \text{a.s.}$$

which is bounded, we have

$$\sup_{t,x}\left|D_n(t\mid x;\beta_n) - D_n(t\mid x;\beta)\right| \le O(|\beta_n-\beta|)\left[\sup_x \hat{f}_{0n}(x)\right] = O_{\text{a.s.}}(n^{-1/2}) = O_{\text{a.s.}}(r_n). \tag{24}$$

Thus

$$\sup_{t,x}\left|D_n(t\mid x;\beta_n) - D(t\mid x)\right| = O_{\text{a.s.}}(r_n). \tag{25}$$

Next note that $\int \chi_c(u)g(u)\,du = 0$ and that, for all u and ε, $|\chi_c(u+\varepsilon)-\chi_c(u)| \le |\varepsilon|$ to conclude that, as $\varepsilon \to 0$,

$$\left|\varepsilon^{-1}\int \chi_c(u+\varepsilon)g(u)\,du\right| = \left|\varepsilon^{-1}\int[\chi_c(u+\varepsilon)-\chi_c(u)]g(u)\,du\right| \to \left|\int \chi_c{}'(u)g(u)\,du\right| = \left|\int_{-c}^{c} g(u)\,du\right| > 0.$$

Hence there exist $\delta > 0$ and $\delta_0 > 0$ such that for $|\varepsilon| < \delta$,

$$\inf_{x\in[0,1]}\left|\int \chi_c(u-\eta(x)+\varepsilon)g(u-\eta(x))\,du\right| = \left|\int \chi_c(u+\varepsilon)g(u)\,du\right| > \delta_0|\varepsilon|. \tag{26}$$

Since $\tilde{\eta}_n(x|\beta_n)$ is the solution of $D_n(t\mid x;\beta_n) = 0$, the lemma follows from (25) and (26) by arguing as in Theorem 3.4 of Härdle, Janssen and Serfling (1988). □

REMARK 8. If Condition A3(b) is replaced by the condition

CONDITION A3. (b′) $\int \chi(y)\,g(y)\,dy = 0$ and $\int \chi'(y)\,g(y)\,dy > 0$ for a known bounded, nondecreasing, odd function χ with bounded derivative, as suggested by a referee, then in (16), we replace the Huber function χ_c by this function χ. The proof of Lemma 4.2 remains valid if we use χ in place of χ_c throughout and note that, for $\varepsilon \to 0$,

$$\varepsilon^{-1} \int \chi(u+\varepsilon) g(u)\, du = \varepsilon^{-1} \int [\chi(u+\varepsilon) - \chi(u)] g(u)\, du$$

$$\to \int \chi'(u) g(u)\, du > 0,$$

since $\int \chi(u) g(u)\, du = 0$ and χ' is bounded. Condition A3(b$'$) is more general than Condition A3(b) in the sense that Condition A3(b) implies Condition A3(b$'$), but the requirement that a known function χ should make $\int \chi(y) g(y)\, dy$ equal to 0 when g itself is unknown, may seem to be artificial.

In Lemma 4.2, if we replace $\beta_n = \beta + O(n^{-1/2})$ by the estimator $\widetilde{\beta}_n = \beta + O_p(n^{-1/2})$ obtained in Section 3, then everything goes through except that the a.s. rate in (24) and (25) is replaced by an "in probability" rate. This leads to the following theorem, giving a weak uniform convergence rate of the estimator $\widehat{\eta}_n(x) = \widetilde{\eta}_n(x|\widetilde{\beta}_n)$, which is of interest aside from the main issue of estimating β.

THEOREM 4. *Let* $\widehat{\eta}_n(x) = \widetilde{\eta}_n(x|\widetilde{\beta}_n)$. *Then, under the conditions of Theorem* 3 *and Lemma* 4.2 *and with* r_n *as in Lemma* 4.2,

$$\sup_{x \in [0,1]} \left| \widehat{\eta}_n(x) - \eta(x) \right| = O_p(r_n) \quad \text{under } P_{\beta,\gamma}.$$

We now consider two variations of $\widehat{\varphi}_n(t \mid \beta')$, viz., $\varphi_n(t \mid \beta')$ obtained by using $\mathscr{E}_i(\beta') = Y_i - W_i\beta' - \eta(X_i)$ in place of $\widehat{\mathscr{E}_i}(\beta') = Y_i - W_i\beta' - \widetilde{\eta}_n(X_i \mid \beta')$ in (18) and (19), and its leave-one-out version $\varphi_{n(i)}(t \mid \beta')$ based on all samples except Z_i. In the following two lemmas, we state some properties of $\widehat{\varphi}_n$, φ_n and $\varphi_{n(i)}$. The proof of Lemma 4.3 is straightforward and we omit it, while Lemma 4.4 is a restatement of a result from page 100 of Schick (1987).

LEMMA 4.3. *The following hold for all* t, t_1, t_2, β' *and* β'':

(a) $|\widehat{\varphi}_n(t \mid \beta')| \le a_n^{-1}$,
(b) $|\widehat{\varphi}_n(t_1 \mid \beta') - \widehat{\varphi}_n(t_2 \mid \beta')| \le 3\, a_n^{-4}\, |t_1 - t_2|$,
(c) $|\varphi_n(t_1 \mid \beta') - \varphi_n(t_2 \mid \beta')| \le 3\, a_n^{-4} |t_1 - t_2|$,
(d) $|\varphi_n(t \mid \beta') - \varphi_n(t \mid \beta'')| \le 3M a_n^{-4} |\beta' - \beta''|$, *where* $P(|W| \le M) = 1$,
(e) $|\widehat{\varphi}_n(t \mid \beta') - \varphi_n(t \mid \beta')| \le 3\, a_n^{-4} \sup_{x \in [0,1]} |\widetilde{\eta}_n(x|\beta') - \eta(x)|$,
(f) $|\varphi_n(t \mid \beta') - \varphi_{n(i)}(t \mid \beta')| \le 2\, n^{-1}\, a_n^{-3}$.

LEMMA 4.4. *If* $a_n \to 0$ *and* $n a_n^6 \to \infty$ *as* $n \to \infty$, *then* $E_{\beta,\gamma}[\int \{\varphi_n(t \mid \beta) - \varphi(t)\}^2\, g(t)\, dt] \to 0$.

In the Appendix, we use Lemmas 4.1–4.4 to prove the following lemma which leads to the main result of this section.

LEMMA 4.5. *Suppose that Conditions* A1, A2 *and* A3 *hold. If* $a_n \to 0$, $na_n^8 \to \infty$, $h_n \to 0$ *and* $n^{1/2} a_n^{-4} [(nh_n/\log n)^{-1/2} + h_n]^2 \to 0$, *then, for every sequence* $\beta_n = \beta + O(n^{-1/2})$, *the following hold under* $P_{\beta,\gamma}$:

$$\text{(i)} \qquad n^{1/2} \int \widehat{H}_n(z;\, \beta_n, \mathbf{Z}(n))\, p(z;\, \beta_n, \gamma)\, d\mu = o_p(1),$$

$$\text{(ii)} \qquad \int \big|\widehat{H}_n(z;\, \beta_n, \mathbf{Z}(n)) - H(z;\, \beta_n, \gamma)\big|^2\, p(z;\, \beta_n, \gamma)\, d\mu = o_p(1),$$

$$\text{(iii)} \qquad \widehat{I}_n(\beta_n, \mathbf{Z}(n)) = I_*(\beta, \gamma) + o_p(1).$$

We now establish Condition B3 (discussed in Section 2) for our estimator of the efficient influence function.

THEOREM 5. *Suppose that Conditions* A1, A2 *and* A3 *hold, and let* a_n *and* h_n *be chosen as in Lemma* 4.5. *Then the estimator* $\widehat{J}_n(\cdot)$ *of the efficient influence function constructed by* (23) *satisfies Condition* B3.

PROOF. The theorem follows directly from Lemma 4.5 and formulas (5) and (23) for $J(\cdot)$ and $\widehat{J}_n(\cdot)$, respectively. □

5. Efficient estimator of β. We now summarize the developments of Sections 3 and 4 to describe our construction of an asymptotically efficient estimator of β.

First, let $\widetilde{\beta}_n$ be the $\sqrt{n}$-consistent estimator defined by (8) and (9), and consider its discretized version $\overline{\beta}_n$, namely, the point in the set $\{j\, n^{-1/2}\colon j$ is an integer$\}$ which is closest to $\widetilde{\beta}_n$. Next, split the data in two halves: $\{Z_i = (W_i, X_i, Y_i),\ i \in \Lambda_{n1}\}$, $\{Z_i = (W_i, X_i, Y_i),\ i \in \Lambda_{n2}\}$, where $\Lambda_{n1} = \{1, \dots, [n/2]\}$, $\Lambda_{n2} = \{[n/2]+1, \dots, n\}$, and, for $l = 1, 2$, define $\widehat{q}_{nl}(\cdot)$, $\widetilde{\eta}_{nl}(\cdot \,|\, \overline{\beta}_n)$, restricting the sums in (15) and (16) to $i \in \Lambda_{nl}$. Let $\widehat{\varepsilon}_i(\overline{\beta}_n) = Y_i - W_i\overline{\beta}_n - \widetilde{\eta}_{nl}(X_i \,|\, \overline{\beta}_n)$ for $i \in \Lambda_{nl}$ and, with these residuals, define $\widehat{g}_{nl}(\cdot \,|\, \overline{\beta}_n)$ and $\widehat{g}_{nl}'(\cdot \,|\, \overline{\beta}_n)$, again restricting the sums in (18) to $i \in \Lambda_{nl}$. Now use $\widehat{g}_{nl}(\cdot \,|\, \overline{\beta}_n)$ and $\widehat{g}_{nl}'(\cdot \,|\, \overline{\beta}_n)$ in (19) to construct two estimates $\widehat{\varphi}_{nl}(\cdot \,|\, \overline{\beta}_n)$, $l = 1, 2$, of the score function from the two halves of the data, namely, $\mathbf{Z}(n,1) = (Z_1, \dots, Z_{[n/2]})$ and $\mathbf{Z}(n,2) = (Z_{[n/2]+1}, \dots, Z_n)$. Finally, use $\widehat{q}_{nl}(\cdot)$, $\widetilde{\eta}_{nl}(\cdot \,|\, \overline{\beta}_n)$ and $\widehat{\varphi}_{nl}(\cdot \,|\, \overline{\beta}_n)$ in (22), (21) and (23) to obtain

$$\widehat{H}_{nl}(z;\, \overline{\beta}, \mathbf{Z}(n,l)) = (w - \widehat{q}_{nl}(x))\, \widehat{\varphi}_{nl}(y - w\overline{\beta}_n - \widetilde{\eta}_{nl}(x|\overline{\beta}_n) \,|\, \overline{\beta}_n),$$

$$\widehat{I}_{nl}(\overline{\beta}_n, \mathbf{Z}(n,l)) = 2\, n^{-1} \sum_{i \in \Lambda_{nl}} \widehat{H}_{nl}^2(Z_i;\, \overline{\beta}_n, \mathbf{Z}(n,l)),$$

and the two estimates of the efficient influence function

$$\widehat{\mathcal{J}}_{nl}(z; \overline{\beta}_n, \mathbf{Z}(n, l)) = [\widehat{I}_{nl}(\overline{\beta}_n, \mathbf{Z}(n, l))]^{-1} \widehat{H}_{nl}(z; \overline{\beta}_n, \mathbf{Z}(n, l))$$

for $l = 1, 2$. Evaluating $\widehat{\mathcal{J}}_{n2}(\cdot)$ on $Z_1, \ldots, Z_{[n/2]}$ and $\widehat{\mathcal{J}}_{n1}(\cdot)$ on $Z_{[n/2]+1}, \ldots, Z_n$, the efficient estimator $\widehat{\beta}_n$ of β is now given by (7), which we state again:

$$(27)\quad \widehat{\beta}_n = \overline{\beta}_n + \frac{1}{n}\left[\sum_{i=1}^{[n/2]} \widehat{\mathcal{J}}_{n2}(Z_i; \overline{\beta}_n, \mathbf{Z}(n, 2)) + \sum_{i=[n/2]+1}^{n} \widehat{\mathcal{J}}_{n1}(Z_i; \overline{\beta}_n, \mathbf{Z}(n, 1))\right].$$

We have already shown in Theorem 3 that $\widetilde{\beta}_n$ satisfies Condition B2 of being $\sqrt{n}$-consistent, and in Theorem 5 the estimated efficient influence function $\widehat{\mathcal{J}}_n(\cdot)$ has been shown to satisfy Condition B3. Also, by Theorem 9.5 of Rudin (1987), the map $\beta \mapsto \rho^*(\cdot\,; \beta, \gamma)$ is continuous for every $\gamma = (\eta, f, g) \in \Gamma$, so that Condition B1 is also satisfied. By Theorem 2 [due to Schick (1986)], the asymptotic linearity of the estimator $\widehat{\beta}_n$ is thus established. This is stated in the following theorem.

THEOREM 6. *Suppose that Conditions* A1, A2, A3 *and* K1 *hold*, *and let* b_n, a_n, *and* h_n *be chosen to satisfy* $nb_n \to \infty$, $nb_n^2 \to 0$, $a_n \to 0$, $na_n^8 \to \infty$, $h_n \to 0$, $nh_n/\log n \to \infty$ *and* $n^{1/2} a_n^{-4} [(nh_n/\log n)^{-1/2} + h_n]^2 \to 0$. *Then the estimator* $\widehat{\beta}_n$ *of* β *given by* (27) *is asymptotically linear*. *Hence* $\sqrt{n}\,(\widehat{\beta}_n - \beta)$ *is asymptotically* $\mathcal{N}(0, [I_*(\beta, \gamma)]^{-1})$, *thus attaining the smallest possible asymptotic variance among all regular estimators*.

REMARK 9. The condition $nh_n^4 \to 0$ is not stated explicitly in Theorem 6, because it is implied by $n^{1/2} a_n^{-4} [(nh_n/\log n)^{-1/2} + h_n]^2 \to 0$. An example of b_n, a_n, h_n satisfying the conditions of Theorem 6 is $b_n = n^{-2/3}$, $a_n = (n^{-1/8} \log n)^{1/3}$ and $h_n = (n^{-1} \log n)^{1/3}$. However, if $\eta \in C^2[0, 1]$, then in Theorem 6 the condition $nb_n^2 \to 0$ can be replaced by $nb_n^4 \to 0$ as pointed out in Remark 6, so that b_n can be chosen to be the same as h_n. Of course, the choice of b_n, a_n, h_n by a data-driven method to achieve good performance in moderate-sized samples needs to be investigated.

REMARK 10. Schick (1993) has developed a method of estimation in regression models which avoids the sample-splitting technique described above. This approach depends on conditions involving a leave-one-out type of conditional expectation which can be easily verified in partial linear models with finite residual variance, where the nonparametric regression can be effectively estimated by averages over small intervals. However, the problem is much harder without a moment condition on the residuals, because here the robust M-smoother estimate of the nonparametric regression is not in a closed form and the verification of these conditions leads to intractable calculations.

REMARK 11. In this paper, we have mainly discussed the estimation of a real parameter β. We now indicate how these results extend to the multiparameter case. The asymptotic lower bound and the efficient influence function generalize to the multi-parameter case in a straightforward manner as mentioned in Remark 4. The bandwidth-matched M-estimation procedure for constructing a $\sqrt{n}$-consistent initial estimate generalize to the case of W having its support on the cartesian product of finite sets, as described in Remark 7. A $\sqrt{n}$-consistent estimate in the general multidimensional case has been constructed by Zhao (1995) following a different approach. With these multidimensional generalizations of the results of Sections 2 and 3, everything else extends in an obvious manner without any further difficulty to the multiparameter case.

APPENDIX

PROOF OF LEMMA 4.5. We sketch the proof and refer to Bhattacharya and Zhao (1995) for more details. Let $t = y - w\beta_n - \eta(x)$, $\mathscr{E} = y - w\beta - \eta(x)$, $\alpha_n(x) = \tilde{\eta}_n(x|\beta_n) - \eta(x)$ and $\delta_n(x) = q(x) - \hat{q}_n(x)$. Then $H(z) = H(z;\, \beta, \gamma) = (w - q(x))\varphi(\mathscr{E})$, $H_n(z) = H(z;\, \beta_n, \gamma) = (w - q(x))\varphi(t)$, $\widehat{H}_n(z) = \widehat{H}_n(z;\, \beta_n, \mathbf{Z}(n)) = (w - q(x) + \delta_n(x))\,\hat{\varphi}_n(t - \alpha_n(x)\,|\,\beta_n)$, and $p_n(z)\,d\mu = p(z;\, \beta_n, \gamma)\,d\mu = f(w, x)g(t)\,d\nu(w, x)\,dt$. Now let $M_n(x) = \int \hat{\varphi}_n(t - \alpha_n(x)\,|\,\beta_n)\,g(t)\,dt$ and note that $E[|M_n(X)|] < \infty$ by Lemma 4.3(a) so that we can write

$$
\begin{aligned}
\int \widehat{H}_n(z)\, p_n(z)\, d\mu
&= E[M_n(X)\, E(W - q(X)|X)] \\
&\quad + \int \delta_n(x) \left[\int \{\hat{\varphi}_n(t - \alpha_n(x)\,|\,\beta_n) - \hat{\varphi}_n(t\,|\,\beta_n)\}\, g(t)\, dt \right] f_0(x)\, dx \\
&\quad + \left[\int \delta_n(x) f_0(x)\, dx \right] \left[\int \hat{\varphi}_n(t\,|\,\beta_n)\, g(t)\, dt \right].
\end{aligned}
$$

The first term is clearly 0, the second term is $o_{\text{a.s.}}(n^{-1/2})$ by Lemmas 4.1, 4.2, 4.3(b) and the third term is 0 because $\hat{\varphi}_n(t\,|\,\beta_n)$ is an odd function and g is symmetric. If g is not symmetric, we redefine $\hat{\varphi}_n = -\hat{g}_n'/\hat{g}_n$. Now the third term can be shown to be $o_p(n^{-1/2})$ if f_0'', q'' and g'' are bounded. This proves part (i).

Next let $\Delta_{1n}(z) = \delta_n(x)\hat{\varphi}_n(t - \alpha_n(x)\,|\,\beta_n)$, $\Delta_{2n}(z) = (w - q(x))[\hat{\varphi}_n(t - \alpha_n(x)\,|\,\beta_n) - \varphi_n(t\,|\,\beta)]$, $\Delta_{3n}(z) = (w - q(x))[\varphi_n(t\,|\,\beta) - \varphi(t)]$, $\Delta_{0n}(z) = (w - q(x))[\varphi_n(t\,|\,\beta_n) - \varphi_n(\mathscr{E}\,|\,\beta)]$, $B_{1n(i)}(z) = (w - q(x))[\varphi_n(\mathscr{E}\,|\,\beta) - \varphi_{n(i)}(\mathscr{E}\,|\,\beta)]$ and $B_{2n(i)}(z) = (w - q(x))[\varphi_{n(i)}(\mathscr{E}\,|\,\beta) - \varphi(\mathscr{E})]$. Then $\widehat{H}_n(z) - H_n(z) = \sum_{j=1}^{3} \Delta_{jn}(z)$ and $\widehat{H}_n(z) - H(z) = \sum_{j=0}^{2} \Delta_{jn}(z) + \sum_{j=1}^{2} B_{jn(i)}(z)$. We thus have

$$
\int |\widehat{H}_n(z) - H_n(z)|^2\, p_n(z)\, d\mu \le 3 \int \left[\sum_{j=1}^{3} \Delta_{jn}^2(z) \right] p_n(z)\, d\mu,
$$

of which the first two terms are $o_{\text{a.s.}}(1)$ by Lemmas 4.1, 4.2, 4.3(a), (b), (d) and (e), and taking expectation of the third term [with respect to $\mathbf{Z}(n)$], under $P_{\beta,\gamma}$,

$$E\left[\int \Delta_{3n}^2(z)\, p_n(z)\, d\mu\right] = E[\mathrm{Var}(W|X)]E\left[\int \{\varphi_n(t\,|\,\beta) - \varphi(t)\}^2\, g(t)\, dt\right] \to 0$$

by Lemma 4.4, which proves part (ii).

Finally, since $I_*(\beta, \gamma) = E[H^2(Z)]$, part (iii) will follow by the WLLN if we show that

$$n^{-1}\sum_{i=1}^{n}\{\widehat{H}_n^2(Z_i) - H^2(Z_i)\} = o_p(1).$$

Let $\Delta_n(Z_i) = \widehat{H}_n(Z_i) - H(Z_i)$. Then by the Cauchy-Schwarz inequality,

$$\begin{aligned} n^{-1}&\left|\sum_{i=1}^{n}\{\widehat{H}_n^2(Z_i) - H^2(Z_i)\}\right| \\ &= n^{-1}\left|\sum_{i=1}^{n}\{\Delta_n^2(Z_i) + 2H(Z_i)\Delta_n(Z_i)\}\right| \\ &\le n^{-1}\sum_{i=1}^{n}\Delta_n^2(Z_i) + 2\left\{n^{-1}\sum_{i=1}^{n}H^2(Z_i)\right\}^{1/2}\left\{n^{-1}\sum_{i=1}^{n}\Delta_n^2(Z_i)\right\}^{1/2}. \end{aligned}$$

Thus, showing $n^{-1}\sum_{i=1}^{n}\Delta_n^2(Z_i) = o_p(1)$ is enough. But

$$n^{-1}\sum_{i=1}^{n}\Delta_n^2(Z_i) \le 5n^{-1}\left[\sum_{j=0}^{2}\sum_{i=1}^{n}\Delta_{jn}^2(Z_i) + \sum_{j=1}^{2}\sum_{i=1}^{n}B_{jn(i)}^2(Z_i)\right],$$

of which $n^{-1}\sum_{i=1}^{n}\Delta_{jn}^2(Z_i)$, $j = 0, 1, 2$ and $n^{-1}\sum_{i=1}^{n}B_{1n(i)}^2(Z_i)$ are $o_{\text{a.s.}}(1)$ by Lemmas 4.1, 4.2, 4.3, and the expectation of the remaining term [with respect to $\mathbf{Z}(n)$] under $P_{\beta,\gamma}$ is

$$\begin{aligned} E&\left[n^{-1}\sum_{i=1}^{n}B_{2n(i)}^2(Z_i)\right] \\ &= n^{-1}\sum_{i=1}^{n}EE[B_{2n(i)}^2(Z_i)\,|\,Z_i] = EE[B_{2n(1)}^2(Z_1)\,|\,Z_1] \\ &= \iiint (w - q(x))^2\, E[\{\varphi_{n(1)}(\mathscr{E}\,|\,\beta) - \varphi(\mathscr{E})\}^2] f(w, x)\, g(\mathscr{E})\, d\nu(w, x)\, d\mathscr{E} \\ &= E[\mathrm{Var}(W|X)]E\left[\int \{\varphi_{n(1)}(\mathscr{E}\,|\,\beta) - \varphi(\mathscr{E})\}^2\, g(\mathscr{E})\, d\mathscr{E}\right], \end{aligned}$$

which tends to 0 by Lemmas 4.3(f) and 4.4, proving part (iii). □

Acknowledgments. Thanks are due to a referee for some useful suggestions and to Prabir Burman relating to the proof of Lemma 4.5(i) for asymmetric residual distribution.

REFERENCES

BHATTACHARYA, P. K. (1989). Estimation of treatment main effect and treatment-covariate interaction in observational studies using bandwidth-matching. Technical Report 188, Div. Statistics, Univ. California, Davis.

BHATTACHARYA, P. K. and ZHAO, P.-L. (1995). Semiparametric inference in a partial linear model. Technical Report 322, Div. Statistics, Univ. California, Davis.

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 Estimation 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.

CHEN, H. and SHIAU, J. H. (1991). A two-stage spline smoothing method for partially linear models. *J. Statist. Plann. Inference* **27** 187–201.

CUZICK, J. (1992a). Semiparametric additive regression. *J. Roy. Statist. Soc. Ser. B* **54** 831–843.

CUZICK, J. (1992b). Efficient estimates in semiparametric additive regression models with unknown error distribution. *Ann. Statist.* **20** 1129–1136.

DENBY, L. (1984). Smooth regression functions. Ph.D. dissertation, Dept. Statistics, Univ. Michigan.

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

GREEN, P., JENNISON, C. and SEHEULT, A. (1985). Analysis of field experiments by least squares smoothing. *J. Roy. Statist. Soc. Ser. B* **47** 299–314.

HALL, P. and JONES, M. C. (1990). Adaptive M-estimation in nonparametric regression. *Ann. Statist.* **18** 1712–1728.

HÄRDLE, W., JANSSEN, P. and SERFLING, R. (1988). Strong uniform consistency rates for estimators of conditional functions. *Ann. Statist.* **16** 1428–1449.

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

HODGES, J. L., JR. and LEHMANN, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.* **34** 598–611.

KLAASSEN, C. A. J. (1987). Consistent estimation of the influence function of locally asymptotically linear estimators. *Ann. Statist.* **15** 1548–1562.

MÜLLER, H. G. (1988). *Nonparametric Regression Analysis of Longitudinal Data.* Springer, New York.

POLLARD, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7** 186–199.

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

RITOV, Y. and BICKEL, P. J. (1990). Achieving information bounds in non and semiparametric models. *Ann. Statist.* **18** 925–938.

ROBINSON, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica* **56** 931–954.

RUDIN, W. (1987). *Real Complex Analysis*, 3rd ed. McGraw-Hill, New York.

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

SCHICK, A. (1987). A note on the construction of asymptotically linear estimators. *J. Statist. Plann. Inference* **16** 89–105.

SCHICK, A. (1993). On efficient estimation in regression models. *Ann. Statist.* **21** 1486–1521.

SHIAU, J. H., WAHBA, G. and JOHNSON, D. R. (1986). Partial spline models for the inclusion of tropopause and frontal boundary information in otherwise smooth two and three dimensional objective analysis. *J. Atmos. Ocean. Technol.* **3** 714–725.

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

WAHBA, G. (1984). Cross validated spline methods for the estimation of multivariate functions from data on functionals. In *Statistics: An Appraisal. Proceedings 50th Anniversary Conference Iowa State Statistical Laboratory* (H. A. David, ed.). Iowa State Univ. Press.

ZHAO, P.-L. (1992). Semiparametric inference in a partial linear model. Ph.D. dissertation, Div. Statistics, Univ. California, Davis.

ZHAO, P.-L. (1995). Bandwidth-matched M-estimation in a partial linear model. Unpublished manuscript.

DIVISION OF STATISTICS
UNIVERSITY OF CALIFORNIA
DAVIS, CALIFORNIA 95616

MERCK RESEARCH LABORATORIES
P.O. BOX 2000, RY 70–38
RAHWAY, NEW JERSEY 07065