The Annals of Statistics
1993, Vol. 21, No. 3, 1486–1521

ON EFFICIENT ESTIMATION IN REGRESSION MODELS

BY ANTON SCHICK

State University of New York, Binghamton

In this paper we consider the regression model with smooth regression function and smooth error and covariate distributions. We study how well one can estimate functionals of the regression function which may also depend on the distribution of the covariate. This is done by deriving the efficient influence functions of least dispersed regular estimators of such functionals under various assumptions on the parameters of our model. Then we demonstrate how efficient estimates can be constructed. We provide a general procedure for constructing efficient estimates that relies on appropriate auxiliary estimates. We illustrate the usefulness of this procedure by constructing efficient estimates for various parametric, nonparametric and semiparametric models.

1. Introduction. Let Y be a random variable and Z be a random vector taking values in some measurable subset $\mathscr{K}$ of $\mathbb{R}^k$. We consider the regression model defined by the structural relation

$$Y = \varrho(Z, \xi) + \varepsilon, \tag{1.1}$$

where ξ is an unknown parameter in the set Ξ, ϱ is a real valued function on $\mathscr{K} \times \Xi$, and ε is an unobservable random variable that is independent of Z. We denote the distribution of ε by F and call it *the error distribution*. We denote the distribution of Z by G and call it the *covariate distribution*. The parameter of our model is $\theta = (\xi, G, F)$ which we assume to belong to some parameter set $\Theta = \Xi \times \mathfrak{G} \times \mathfrak{F}$, where $\mathfrak{G}$ is a model for the covariate distribution, and $\mathfrak{F}$ a model for the error distribution. Let $P_{(\xi, G, F)}$ denote the distribution of (Y, Z). The goal of this paper is to study the problem of efficiently estimating a characteristic $\chi(\xi, G)$ of the parameter ξ and the covariate G in the presence of the nuisance parameter F. Here χ is a functional from $\Xi \times \mathfrak{G}$ to $\mathbb{R}^m$. We focus on functionals which are estimable at a $\sqrt{n}$-rate. Our treatment is not intended for characteristics such as $\varrho(z_0, \xi)$, the regression function at a fixed point z_0, which cannot be estimated at the $\sqrt{n}$-rate in general.

Our formulation is kept abstract. This allows us to simultaneously consider various regression models. If Ξ is chosen to be a subset of $\mathbb{R}^m$, one obtains parametric regression models. These include linear and nonlinear regression models. In nonparametric regression one takes Ξ to be a set of smooth functions, usually a dense subset of $C_b(\mathscr{K})$, the set of all bounded continuous

Received March 1991; revised June 1992.

AMS 1991 *subject classifications*. 62G20, 62G05.

Key words and phrases. Linear regression, nonparametric regression, semiparametric regression, partly linear additive models, efficient influence function, regular estimator, convolution theorem.

functions from $\mathscr{K}$ to $\mathbb{R}$, and puts

$$\varrho(Z,t) = t(Z), \qquad t \in \Xi.$$

If one takes Ξ to be of the form $\Xi_1 \times \Xi_2$, with Ξ_1 a subset of $\mathbb{R}^m$ and Ξ_2 a subset of some function space, one arrives at semiparametric regression models. Here are some examples of semiparametric regression models.

1. *Partially linear additive regression*. Let $\mathscr{K} = \mathbb{R}^m \times \mathscr{K}_2$, $Z = (U^T, W^T)^T$, $\Xi_1 = \mathbb{R}^m$, and Ξ_2 be a family of smooth functions from $\mathscr{K}_2$ to $\mathbb{R}$, and take
$$\varrho(Z,(b,t)) = b^T U + t(W), \qquad b \in \mathbb{R}^m, t \in \Xi_2.$$
Such models have been studied by Chen (1988), Cuzick (1992a, b), Engle, Granger, Rice and Weiss (1986), Green, Jennison and Seheult (1985), Heckman (1986), Rice (1986), Robinson (1988) and Wahba (1984). They can be viewed as special cases of the additive models considered by Stone (1985).
2. *Semiparametric comparison of regression functions*. Let $\mathscr{K} = \{0,1\} \times [0,1]$, $Z = (U,W)^T$, $\Xi_1 = (0,\infty)$, Ξ_2 be a set of continuous functions from $[0,1]$ to $\mathbb{R}$, and
$$\varrho(Z,(b,t)) = (1-U)t(W) + Ubt(W), \qquad b \in \Xi_1, t \in \Xi_2.$$
This model is a special case of models discussed in Härdle and Marron (1990). It refers to a two sample problem. In the first sample ($U = 0$) one has regression function τ and in the second sample ($U = 1$) one has the regression function $\beta\tau$.

In parametric regression models, a characteristic of great interest is the parameter ξ itself leading us to consider $\chi(\xi, G) = \xi$. In semiparametric regression models, the finite dimensional component β of $\xi = (\beta,\tau)$ is an important characteristics and can be expressed by $\beta = \chi((\beta,\tau), G)$. Other examples of functionals are $\chi(\xi,G) = \int\varrho(z,\xi)\,dG(z)$, the average regression effect, $\chi(\xi,G) = \int\varrho(z,\xi)^2\,dG(z)$, the second moment of the regression function and $\chi(\xi,G) = \int\varrho(z,\xi)^2\,dG(z) - (\int\varrho(z,\xi)\,dG(z))^2$, the variance of the regression function.

Our paper consists of three parts. The first part deals with efficiency considerations. The efficiency criterion we shall be using is that of a least dispersed regular estimator. We briefly review this concept in Section 2. In Section 3 we apply this criterion to our regression model and characterize least dispersed regular estimators of smooth functionals χ. We do this by describing the efficient influence function of such estimates in terms of characteristics derived from the functional and the particular model. In Section 4 we explicitly calculate the efficient influence functions for some specific models. Among others we obtain the efficient influence functions for the two semiparametric regression models mentioned above. The efficient influence function for the second model is new, while the result for the partly linear additive regression model is well known. It appears in Bickel, Klaassen, Ritov and Wellner (1993) and in Cuzick (1992a). Some other results in this first part overlap also with results in the monograph of Bickel, Klaassen, Ritov and Wellner (1993). However, our treatment is more general and provides a unifying approach to the efficiency considerations in regression models.

To carry out this program we shall require that F has a finite Fisher information for location and that ϱ and χ are smooth as made precise in Assumptions 3.1 to 3.3. We also have to be concerned about the identifiability of the regression function and the parameter ξ. Commonly used assumptions to identify the regression function are (i) the error distribution is symmetric about 0; and (ii) the error distribution has mean zero. Of course, other possibilities exist such as requiring that the median of the error distribution is zero or that some other location functional of the error distribution vanishes. In this paper we shall not dwell on the particular conditions that are used to identify the regression. Our results are applicable to any set of identifiability conditions. The only identifiability condition imposed on F in the paper is the condition $J_* > 0$ appearing in Theorem 3.11.

The second part of the paper deals with the abstract problem of constructing efficient estimates. In Section 5 we give a general construction lemma that reduces the problem of constructing efficient estimates to the problem of constructing appropriate preliminary estimates of the parameter ξ and characteristics of the functional χ and the model. We directly estimate the efficient influence function. The work of Schick (1986) and Klaassen (1987) shows that this is the appropriate approach. However, our construction avoids the sample splitting techniques used in Schick (1986) and Klaassen (1987) by using techniques developed in Schick (1987). We have formulated our result to allow for the popular technique of using discrete $\sqrt{n}$-consistent preliminary estimates for finite dimensional parameters in parametric and semiparametric models. But our construction provides also means of avoiding this approach. Some of the proofs of results of Section 5 are deferred to Section 10.

The third part deals with applications of the results in the first two parts. In Section 6 we discuss the construction of efficient estimates in parametric models. We construct efficient estimates of the finite dimensional parameter based on the availability of $\sqrt{n}$-consistent estimates without using the sample splitting technique. In Section 7 we discuss efficient estimation in nonparametric models. In particular, we show that efficient estimates for $\int h\xi\, dG$ and $\int \xi^2\, dG$ can be constructed under mild assumptions on the covariate distribution if the regression function $\xi = \varrho(\cdot, \xi)$ is smooth. The work of Bickel and Ritov (1990) indicates that without smoothness assumptions efficient estimates cannot be constructed. In Section 8 we construct an efficient estimate for the finite dimensional parameter in the partly linear additive regression model. Our estimate improves greatly over an estimate constructed by Cuzick (1992b). His construction uses only a small sample to construct the influence function and imposes stronger assumptions on the model. See Remark 8.3 for more details. In Section 9 we construct an efficient estimate for the finite dimensional parameter in the semiparametric comparison of regression functions model.

NOTATION. Throughout this paper we will use the following notation. If $\{a_n\}$ and $\{b_n\}$ are sequences of positive numbers, we write $a_n \sim b_n$ to indicate that $\{a_n/b_n\}$ is bounded and bounded away from 0. If P is a probability measure and X is a random vector, we denote the distribution of X under P by $\mathfrak{L}(X|P)$ and let $L_*(P) = \{h \in L_2(P)\colon \int h\,dP = 0\}$. Let $P, P_1, P_2, \dots,$ be probability measures defined on the same σ-field. Then we say $\{P_n\}$ has *tangent t with respect to P*, if $t \in L_*(P)$, and if

$$\int \left(\sqrt{n} \left(\sqrt{p_n} - \sqrt{p} \right) - \tfrac{1}{2} t \sqrt{p} \right)^2 d\mu \to 0,$$

for densities $p, p_1, p_2, \dots$ of $P, P_1, P_2, \dots$ with respect to some dominating measure μ. If U is a closed linear subspace of the Hilbert space H and $h \in H$, then we let $\Pi(h|U)$ denote the projection of h onto U. If L is a linear operator on a vector space V into a vector space W, and $v = (v_1, \dots, v_m)$ is an element of V^m, then we let $L(v)$ denote the vector $(L(v_1), \dots, L(v_m))$. This applies in particular to integrals and to the projection operator $\Pi(\cdot|U)$. Also we write LA for the image $\{La\colon a \in A\}$ of the set A under L. In matrix calculations elements in product spaces will be treated as column vectors.

Let A be a subset of a Banach space V with norm $\|\cdot\|_V$ and $a \in A$. We call the set D of all d in V, for which there is a sequence $\{v_n\}$ in A with the property

$$\left\| \sqrt{n} \, (v_n - a) - d \right\|_V \to 0,$$

the *set of directions of A at a*. We refer to such a sequence $\{v_n\}$ as a *local sequence for a in A with direction d*. We say a map γ on A into some Banach space W with norm $\|\cdot\|_W$ is *directionally differentiable at a in the weak* (resp. *strong*) *sense*, if there is a linear operator $\dot{\gamma}$ from V to W such that, for each direction d in D, and some (resp. each) local sequence $\{v_n\}$ for a in A with direction d,

$$\left\| \sqrt{n} \, (\gamma(v_n) - \gamma(a)) - \dot{\gamma} d \right\|_W \to 0.$$

We refer to $\dot{\gamma}$ as the *directional derivative of* γ at a.

2. The efficiency criterion. The efficiency criterion we shall be using in this paper is that of a least dispersed regular estimator as elaborated in Begun, Hall, Huang and Wellner (1983) and Pfanzagl and Wefelmeyer (1982). See also the monograph by Bickel, Klaassen, Ritov and Wellner (1993). In this section we briefly recall this concept as it pertains to our situation.

Let $(S, \mathbf{S})$ be a measurable space, $\mathscr{P} = \{P_\vartheta\colon \vartheta \in \Theta\}$ a family of distributions on $\mathbf{S}$, and κ a functional on Θ into $\mathbb{R}^m$. We focus attention on a fixed point θ in Θ and study how well we can estimate $\kappa(\theta)$ under the product measures $\{P_\theta^n\}$. We require that Θ is a subset of some Banach space $\mathscr{H}$ with norm $\|\cdot\|_{\mathscr{H}}$. Let $\dot{\Theta}$ denote the set of directions of Θ at θ. We impose the following conditions.

2.1 ASSUMPTION. There is a bounded linear operator $\dot{P}$ on $\mathscr{H}$ into $L_*(P_\theta)$ such that, for each δ in $\dot{\Theta}$, and every local sequence $\{\vartheta_n\}$ for θ in Θ with direction δ, P_{ϑ_n} has tangent $\dot{P}\delta$ with respect to P_θ.

2.2 ASSUMPTION. The functional κ is directionaly differentiable at θ in the weak sense, that is, there is a bounded linear operator $\dot{\kappa}$ on $\mathscr{H}$ into $\mathbb{R}^m$, such that, for each $\delta \in \dot{\Theta}$, and some local sequence $\{\vartheta_n\}$ for θ in Θ with direction δ,

$$\sqrt{n}\left(\kappa(\vartheta_n) - \kappa(\theta)\right) \to \dot{\kappa}\delta. \tag{2.1}$$

2.3 DEFINITION. The closed linear subspace generated by $\{\dot{P}\delta\colon \delta \in \dot{\Theta}\}$ is called the *tangent space* of $\mathscr{P}$ and will be denoted by $\mathscr{T}$. A local sequence $\{\vartheta_n\}$ satisfying (2.1) is called *admissible*. By an *estimator* we mean a sequence $\{\kappa_n\}$, where κ_n is a measurable function on S^n into $\mathbb{R}^m$, for each $n = 1, 2, \dots$. We say the estimator $\{\kappa_n\}$ is *regular at* θ if there is a distribution Q such that

$$\mathfrak{L}\left(n^{1/2}(\kappa_n - \kappa(\vartheta_n))|P^n_{\vartheta_n}\right) \quad \Rightarrow \quad Q,$$

for all admissible local sequences $\{\vartheta_n\}$. We say the estimator $\{\kappa_n\}$ has *influence function* ψ *at* θ, if $\psi \in (L_*(P_\theta))^m$ and

$$n^{1/2}\left(\kappa_n - \kappa(\theta) - \tilde{\psi}_n\right) \to 0 \quad \text{in } P^n_\theta\text{-prob.},$$

where $\tilde{\psi}_n$ denotes the map $(s_1, \dots, s_n) \in S^n \to 1/n(\psi(s_1) + \cdots + \psi(s_n))$.

2.4 THEOREM (Convolution theorem). *Let Assumptions* 2.1 *and* 2.2 *hold. Suppose there is a* ψ_θ *in* $\mathscr{T}^m$ *such that*

$$\int \psi_\theta \dot{P}\delta \, dP_\theta = \dot{\kappa}\delta \tag{2.2}$$

for all $\delta \in \dot{\Theta}$, *and*

$$\left\{\alpha^T\psi_\theta\colon \alpha \in \mathbb{R}^m\right\} \textit{ is a subset of the closure of } \left\{\dot{P}\delta\colon \delta \in \dot{\Theta}\right\}. \tag{2.3}$$

Then the following hold, with $\Psi(\theta) = \int \psi_\theta \psi_\theta^T \, dP_\theta$. .

(i) *The limiting distribution* Q *of an estimator regular at* θ *is a convolution of the normal distribution* $N(0, \Psi(\theta))$ *and some distribution* M:

$$Q = N(0, \Psi(\theta)) * M.$$

(ii) *An estimator* $\{\kappa_n\}$ *is regular at* θ *with limiting distribution* $Q = N(0, \Psi(\theta))$ *if and only if* $\{\kappa_n\}$ *has influence function* ψ_θ *at* θ.

2.5 DEFINITION. The map ψ_θ is called the *efficient influence function* (for κ at θ). An estimator is called *efficient* (for κ at θ) if it has influence function ψ_θ at θ.

3. Efficient estimation in the regression model. Let us now apply this theory to our regression model. The parameter set for our regression model is the set $\Theta = \Xi \times \mathfrak{G} \times \mathfrak{F}$, where $\mathfrak{G}$ is a model for the covariate distribution and $\mathfrak{F}$ is a model for the error distribution. We denote the fixed point θ by (ξ, G, F). We are interested in a functional κ of the form

$$\kappa(x, \Gamma, \Phi) = \chi(x, \Gamma), \qquad x \in \Xi, \Gamma \in \mathfrak{G}, \Phi \in \mathfrak{F},$$

where χ is a functional form $\Xi \times \mathfrak{G}$ to $\mathbb{R}^m$.

From now on we view (Y, Z) as the identity map on $\mathbb{R} \times \mathscr{K}$ and assume that Ξ is a subset of some Banach space B. Let $\dot{\mathfrak{F}}$ denote the set of tangents for $\mathfrak{F}$ at F, $\dot{\mathfrak{G}}$ the set of tangents for $\mathfrak{G}$ at G and $\dot{\Xi}$ the set of directions of Ξ at ξ. We may think of $\dot{\mathfrak{F}}$ as the set of directions for $\mathfrak{F}$ at F and of $\dot{\mathfrak{G}}$ as the set of directions for $\mathfrak{G}$ at G. Thus we take $\dot{\Theta}$ to be $\dot{\Xi} \times \dot{\mathfrak{G}} \times \dot{\mathfrak{F}}$. Throughout this paper we always impose the following conditions on F, ϱ and χ.

3.1 Assumption. The error distribution F has finite Fisher information for location, that is, F possesses an absolutely continuous Lebesgue density f and

$$J = \int \ell^2 \, dF < \infty, \quad \text{where } \ell = -\frac{f'}{f} 1_{\{f > 0\}}.$$

3.2 Assumption. For each $x \in \Xi$, $\int \varrho^2(x, x)\, dG(z) < \infty$. The map $x \to \varrho(\cdot, x)$, viewed as a map from Ξ into $L_2(G)$, is directionally differentiable at ξ in the strong sense with derivative $\dot{\varrho}$.

3.3 Assumption. The map χ is directionally differentiable at (ξ, G) in the weak sense with derivative $\dot{\chi}$.

Of course, the differentiability of χ implies that of κ and hence Assumption 2.2. To see that Assumption 2.1 holds, we need the following lemma, which is derived using similar arguments as given in Hájek and Šidák (1967), pages 210–214.

3.4 Lemma. *Let $\{F_n\}$ be a sequence of distributions which has tangent c with respect to F, let $\{G_n\}$ be a sequence of distributions which has tangent b with respect to G and let $\{\xi_n\}$ be a local sequence in Ξ with direction a. Then $\{P_{(\xi_n, G_n, F_n)}\}$ has tangent*

$$\dot{\varrho} a(Z) \ell(\varepsilon) + b(Z) + c(\varepsilon)$$

with respect to $P_{(\xi, G, F)}$. Thus the sequences $\{P^n_{\xi_n, G_n, F_n}\}$ and $\{P^n_{\xi, G, F}\}$ are contiguous.

This shows that our regression model satisfies Assumption 2.1 with $\dot{P}$ given by

$$\dot{P}\delta = \dot{\varrho}\delta_1(Z)\ell(\varepsilon) + \delta_2(Z) + \delta_3(\varepsilon), \qquad \delta = (\delta_1, \delta_2, \delta_3) \in \dot{\Xi} \times \dot{\mathfrak{G}} \times \dot{\mathfrak{F}}.$$

We now impose the following additional conditions.

3.5 CONDITION. $\dot{\mathfrak{F}}$ is a closed linear subspace of $L_*(F)$.

3.6 CONDITION. $\dot{\mathfrak{G}}$ is a closed linear subspace of $L_*(G)$.

3.7 CONDITION. $\mathscr{V}$, the closure of $\{\dot{\varrho} a\colon a \in \dot{\Xi}\}$, is a linear subspace of $L_2(G)$.

3.8 CONDITION. There exist N_1 and N_2 in $L_2(G)^m$ such that

$$\dot{\chi}(a, b) = \int N_1 \dot{\varrho} a \, dG + \int N_2 b \, dG, \qquad a \in \dot{\Xi},\ b \in \dot{\mathfrak{G}}. \tag{3.1}$$

The first three conditions imply that the tangent space $\mathscr{T}$ is given by

$$\mathscr{T} = \left\{ a(Z)\ell(\varepsilon) + b(Z) + c(\varepsilon) \colon a \in \mathscr{V},\ b \in \dot{\mathfrak{G}},\ c \in \dot{\mathfrak{F}} \right\}.$$

Thus (2.3) will be automatically satisifed. Condition 3.8 is an identifiability condition. It restricts the class of functionals. Functionals that do not satisfy Condition 3.8 are not estimable at a $\sqrt{n}$-rate. Note that N_1 and N_2 are not unique. Define now

$$u = \Pi\left(N_2 | \dot{\mathfrak{G}}\right) \quad \text{and} \quad v = \Pi(N_1 | \mathscr{V}).$$

Then u and v are uniquely determined and

$$\dot{\chi}(a, b) = \int v \dot{\varrho} a \, dG + \int u b \, dG, \qquad a \in \dot{\Xi},\ b \in \dot{\mathfrak{G}}.$$

Let

$$\ell_* = \ell - \Pi(\ell | \dot{\mathfrak{F}}), \qquad J_* = \int \ell_*^2 \, dF, \qquad \Delta_* = \frac{J - J_*}{J} \quad \text{and} \quad d = \Pi(1 | \mathscr{V}).$$

For the remainder of this section and throughout the next section we use the following notation.

3.9 NOTATION. If $a \in L_2(G)^s$, then $\bar{a}$ denotes $\int a \, dG$ and a_0 denotes $a - \bar{a}$. This defines $\bar{v}$, v_0, $\bar{d}$ and d_0.

3.10 THEOREM. *If $\bar{v} = \int v \, dG = 0$, then the efficient influence function ψ_θ is given by*

$$\psi_\theta = u(Z) + v(Z) \frac{\ell(\varepsilon)}{J}. \tag{3.2}$$

PROOF. It is easily checked that ψ_θ belongs to $\mathscr{T}^m$ and that

$$E_\theta(\psi_\theta \cdot (a(Z) + b(Z) l(\varepsilon) + c(\varepsilon))) = \int u a \, dG + \int v b \, dG,$$

for all $a \in \dot{\mathfrak{G}}$, $b \in \mathscr{V}$, $c \in \dot{\mathfrak{F}}$. □

3.11 THEOREM. *Suppose* $\bar{d} = \int d\,dG < 1$ *or* $J_* > 0$. *Then*

$$\Delta = \frac{J - J_*}{J - \bar{d}(J - J_*)} = \frac{\Delta_*}{1 - \bar{d}\Delta_*}$$

is well defined, and the efficient influence function ψ_θ *is given by*

$$(3.3)\quad \psi_\theta = u(Z) + \big(v_0(Z) + \bar{v}\Delta d_0(Z)\big)\frac{\ell(\varepsilon)}{J} + \bar{v}\frac{1}{J(1 - \Delta_* \bar{d})}\ell_*(\varepsilon).$$

PROOF. As $\bar{d} \le 1$ and $J_* \le J$, the denominator of Δ is zero if and only if $J_* = 0$ and $\bar{d} = 1$. Thus Δ is well defined under our assumptions. Verify that

$$(3.4)\quad \psi_\theta = u(Z) + w_0(Z)\ell(\varepsilon) + \bar{w}\ell_*(\varepsilon), \quad \text{where } w = \frac{1}{J}(v + \bar{v}\Delta d).$$

As $w \in \mathscr{V}^m$, $\psi_\theta \in \mathscr{T}^m$. Let $a \in \dot{\mathfrak{G}}$, $b \in \mathscr{V}$ and $c \in \dot{\mathfrak{F}}$. It remains to be shown that

$$E_\theta\big(\psi_\theta \cdot (a(Z) + b(Z)\ell(\varepsilon) + c(\varepsilon))\big) = \int ua\,dG + \int vb\,dG.$$

Easy calculations give

$$\begin{aligned} E_\theta\big(\psi_\theta \cdot (a(Z) + b(Z)\ell(\varepsilon) + c(\varepsilon))\big) &= \int ua\,dG + J\int w_0 b\,dG + J_*\bar{w}\bar{b} \\ &= \int ua\,dG + J\int wb\,dG - (J - J_*)\bar{w}\bar{b}. \end{aligned}$$

By the definition of d, $\int db\,dG = \bar{b}$. Thus

$$J\int wb\,dG = \int vb\,dG + \bar{v}\Delta\bar{b} \quad \text{and} \quad (J - J_*)\bar{w} = \bar{v}\Delta.$$

The desired result is now immediate. □

3.12 REMARK. If $\bar{v} = 0$, then (3.3) reduces to (3.2); if $d = 1$, then (3.3) reduces to

$$(3.5)\qquad \psi_\theta = u(Z) + v_0(Z)\frac{\ell(\varepsilon)}{J} + \bar{v}\frac{\ell_*(\varepsilon)}{J_*};$$

and if $\ell = \ell_*$, then (3.3) reduces to

$$(3.6)\qquad \psi_\theta = u(Z) + \big(v(Z) + \bar{v}\Delta d(Z)\big)\frac{\ell(Z)}{J}.$$

3.13 REMARK. We can guarantee the existence of the efficient influence function by selecting an error model which satisfies $J_* > 0$. Let us now discuss such error models.

(a) Let $\mathfrak{F}_0$ be the set of all error distributions that have zero means, finite variances and finite Fisher informations. If $\mathfrak{F} = \mathfrak{F}_0$, then we find

$$\dot{\mathfrak{F}} = \left\{h \in L_2(F)\colon \int h(x)\, dF(x) = 0,\ \int xh(x)\, dF(x) = 0\right\},$$

$$\ell_*(\varepsilon) = \frac{\varepsilon}{\sigma^2} \quad \text{and} \quad J_* = \frac{1}{\sigma^2},$$

where $\sigma^2 = \int x^2\, dF(x)$ is the variance of F. Thus $J_* > 0$, and the efficient influence function is well defined for this error model.

(b) Let ψ be a measurable function from $\mathbb{R}$ to $\mathbb{R}$, and let $\mathfrak{F}_\psi$ be the set of all error distributions which have finite Fisher informations and for which ψ has mean zero and finite positive variance. For $\mathfrak{F} = \mathfrak{F}_\psi$ we calculate

$$\dot{\mathfrak{F}} = \left\{h \in L_2(F)\colon \int h\, dF = 0,\ \int \psi h\, dF = 0\right\},$$

$$\ell_* = \frac{\int \psi \ell\, dF}{\int \psi^2\, dF}\psi \quad \text{and} \quad J_* = \frac{\left(\int \psi \ell\, dF\right)^2}{\int \psi^2\, dF}.$$

Thus, if $\int \psi \ell\, dF \neq 0$, then $J_* > 0$. Note, if ψ has a bounded positive derivative, then $\int \psi \ell\, dF = \int \psi'\, dF > 0$. The choice $\psi(x) = x$ gives $\mathfrak{F}_0$.

(c) let $\mathfrak{F}_S$ be the set of all errror distributions that are symmetric about zero and possess finite Fisher informations. If $\mathfrak{F} = \mathfrak{F}_S$, we find

$$\dot{\mathfrak{F}} = \{h \in L_*(F)\colon h \text{ is symmetric about zero}\}, \qquad \ell_* = \ell \quad \text{and} \quad J_* = J.$$

Thus $J_* > 0$. As $\ell = \ell_*$, (3.3) simplifes to (3.6).

An important tool in the construction of efficient estimates in semiparametric models is the use of discretized $\sqrt{n}$-consistent preliminary estimates for the finite dimensional parameter. Such estimates can be treated as nonstochastic local sequences in the proofs and, combined with contiguity arguments, lead to considerable simplifications in the proofs. See, for example, Bickel (1982) and Schick (1986) for this approach. To apply this technique we need the following result.

3.14 Lemma. *Suppose either $\bar{v} = 0$, $\bar{d} < 1$ or $J_* > 0$, so that $\gamma = (1/J)(v + \bar{v}\Delta d)$ is well defined. For $x \in \Xi$, define $\phi_x \in L_2^m(P_{x,G,F})$ by*

$$\phi_x(y, z) = \chi(x, G) + u(z) + \gamma_0(z)\ell(y - \varrho(z, x)) + \bar{\gamma}\ell_*(y - \varrho(z, x)), \qquad y \in \mathbb{R},\ z \in \mathscr{K},$$

[so that $\phi_\xi = \chi(\xi, G) + \psi_\theta$, c.f. (3.4)] and define maps $\Phi_n(x)$ from $(\mathbb{R} \times \mathscr{K})^n$ to $\mathbb{R}^m$ by

$$\Phi_n(x)(y_1, z_1, \dots, y_n, z_n) = \frac{1}{n}\sum_{j=1}^n \phi_x(y_j, z_j), \qquad y_i \in \mathbb{R},\ z_i \in \mathscr{K}.$$

Let $\{\xi_n\}$ be a local sequence for ξ such that $\{(\xi_n, G, F)\}$ is admissible. Then

$$\sqrt{n}\,(\Phi_n(\xi_n) - \Phi_n(\xi)) \to 0 \quad \text{in } P_\theta^n\text{-prob.} \tag{3.7}$$

PROOF. In view of part (ii) of the convolution theorem, (3.7) follows if we show that $\{\Phi_n(\xi_n)\}$ is regular at θ with limiting distribution $N(0, \Psi(\theta))$. Fix an admissible local sequence $\{\vartheta_n\} = \{(x_n, G_n, F_n)\}$ for θ with direction (a, b, c). We need to show that

$$\mathfrak{L}\left(n^{1/2}(\Phi_n(\xi_n) - \chi(x_n, G_n))|P_{\vartheta_n}^n\right) \quad \Rightarrow \quad N(0, \Psi(\theta)). \tag{3.8}$$

Let a_* be the direction of $\{\xi_n\}$ and $r = \dot{\varrho}(a - a_*)$. Let δ_n be the map from $\mathbb{R} \times \mathscr{K}$ to $\mathbb{R}$ defined by

$$\delta_n(y, z) = r(z)\ell(y - \varrho(z, \xi_n)) + b(z) + c(y - \varrho(z, \xi_n)), \qquad y \in \mathbb{R}, z \in \mathscr{K}.$$

The same argument which yields Lemma 3.4 also yields that

$$\int \left(n^{1/2}(q_n^{1/2} - p_n^{1/2}) - \tfrac{1}{2}\delta_n p_n^{1/2}\right)^2 d\mu_n \to 0,$$

where q_n and p_n are densities of P_{ϑ_n} and $P_{\xi_n, G, F}$ with respect to $\mu_n = P_{\vartheta_n} + P_{\xi_n, G, F}$. This implies that the log-likelihood ratio Λ_n of $P_{\vartheta_n}^n$ with respect to $P_{(\xi_n, G, F)}^n$ satisfies

$$\Lambda_n - \tilde{\delta}_n + \tfrac{1}{2}\int \delta_n^2 \, dP_{(\xi_n, G, F)} \to 0 \quad \text{in } P_{(\xi_n, G, F)}^n\text{-prob.},$$

where $\tilde{\delta}_n$ is the map from $(\mathbb{R} \times \mathscr{K})^n$ to $\mathbb{R}$ defined by $\tilde{\delta}_n(s_1, \dots, s_n) = n^{-1/2}(\delta_n(s_1) + \cdots + \delta_n(s_n))$. An application of Le Cam's third lemma gives now that

$$\mathfrak{L}\left(n^{1/2}(\Phi_n(\xi_n) - \chi(\xi_n, G)) - \varphi_n | P_{(x_n, G_n, F_n)}\right) \quad \Rightarrow \quad N(0, \Psi(\theta))$$

with

$$\begin{aligned}\varphi_n &= \int \phi_{\xi_n}\delta_n \, dP_{(\xi_n, G, F)} = \int \psi_\theta \cdot (r(Z)\ell(\varepsilon) + b(Z) + c(\varepsilon)) \, dP_\theta \\ &= \dot{\chi}(a - a_*, b).\end{aligned}$$

Since $n^{1/2}(\chi(x_n, G_n) - \chi(\xi_n, G)) \to \dot{\chi}(a - a_*, b)$, the desired result (3.8) follows. □

3.15 REMARK. The above result can be generalized as follows. Let $\{\xi_n\}$ be a local sequence for ξ, let $\{u_n\}$ and $\{\gamma_n\}$ be sequences in $L_*^m(G)$ and $L_2^m(G)$, respectively, and let Υ_n be the map from $(\mathbb{R} \times \mathscr{K})^n$ into $\mathbb{R}^m$ defined by

$$\begin{aligned}\Upsilon_n(y_1, z_1, \dots, y_n, z_n) = \frac{1}{n}\sum_{j=1}^n \chi(\xi_n, G) &+ u_n(z_j) + \gamma_{n,0}(z)\ell\left(y_j - \varrho(z_j, \xi_n)\right) \\ &+ \bar{\gamma}_n \ell_*\left(y_j - \varrho(z_j, \xi_n)\right).\end{aligned}$$

Suppose that $\int \|u_n - u\|^2 + \|\gamma_n - \gamma\|^2 \, dG \to 0$. Then

$$\sqrt{n}\,(\Upsilon_n - \Phi_n(\xi_n)) \to 0 \quad \text{in } P_\theta^n\text{-prob.}$$

This follows as $\{P_\theta^n\}$ and $\{P_{(\xi_n, G, F)}^n\}$ are contiguous and

$$n \int |a^T(\Upsilon_n - \Phi_n(\xi_n))|^2 \, dP_{(\xi_n, G, F)}^n \to 0$$

for every $a \in \mathbb{R}^m$.

4. Examples of efficient influence functions. Let us now discuss and illustrate the results of the previous section for parametric, nonparametric and semiparametric regression models.

4.1. *Parametric regression.* Suppose Ξ is an open subset of $B = \mathbb{R}^m$. Then $\dot{\Xi} = \mathbb{R}^m$, and Assumption 3.2 implies that there is a $\phi \in L_2(G)^m$ such that

$$\dot{\varrho} a = \phi^T a, \qquad a \in \mathbb{R}^m.$$

Typically, $\phi(z)$ is the gradient of the map $b \to \varrho(z, b)$ at $b = \xi$. We have $\mathscr{V} = \{\phi^T a \colon a \in \mathbb{R}^m\}$. Suppose now that the matrix

$$\Phi = \int \phi\phi^T \, dG$$

is invertible. Then

$$a = \Phi^{-1}\Phi a = \int \Phi^{-1}\phi\phi^T a \, dG, \qquad a \in \mathbb{R}^m.$$

We are interested in estimating ξ, that is,

$$\chi(x, \Gamma) = x, \qquad x \in \Xi, \Gamma \in \mathfrak{G}.$$

As a linear functional, χ is directionally differentiable,

$$\dot{\chi}(a, b) = a, \qquad a \in \mathbb{R}^m, b \in \dot{\mathfrak{G}},$$

and Condition 3.8 holds with $N_1 = \Phi^{-1}\phi$ and $N_2 = 0$. One finds $u = 0$, $v = \Phi^{-1}\phi$ and $d = \bar{\phi}^T\Phi^{-1}\phi$. Easy calculations show that

$$v + \bar{v}\Delta d = \left(\Phi^{-1} + \Delta\Phi^{-1}\bar{\phi}\bar{\phi}^T\Phi^{-1}\right)\phi = J\left(J\int \phi_0\phi_0^T \, dG + J_*\bar{\phi}\bar{\phi}^T\right)^{-1}\phi.$$

Thus

$$\psi_\theta = \left(J\int \phi_0\phi_0^T \, dG + J_*\bar{\phi}\bar{\phi}^T\right)^{-1}\left(\phi_0(Z)\ell(\varepsilon) + \bar{\phi}\ell_*(\varepsilon)\right). \tag{4.1}$$

4.2 Example. In linear regression, $\Xi = \mathscr{K} = \mathbb{R}^k$ and $\varrho(Z, b) = b^T Z$, $b \in \mathbb{R}^k$. Assume that $\int \|z\|^2 \, dG(z) < \infty$ and $\int zz^T \, dG(z)$ is nonsingular. Then the above applies with $\phi(Z) = Z$.

4.3. *Nonparametric regression*. Suppose that Ξ is a dense subset of $C_b(\mathscr{K})$, the set of bounded continuous functions from $\mathscr{K}$ to $\mathbb{R}$ endowed with the topology of uniform convergence, and

$$\varrho(Z,t) = t(Z), \qquad t \in \Xi.$$

Then Assumption 3.2 holds with $\dot{\varrho}$ the identity map on $C_b(\mathscr{K})$ and $\mathscr{V} = L_2(G)$. Every functional χ from $\Xi \times \mathfrak{G}$ that is weakly directionally differentiable at (ξ, G) has a derivative $\dot{\chi}$ satisfying (3.1). This follows from the Riesz representation theorem. As $\mathscr{V} = L_2(G)$, we find $v = N_1$ and $d = 1$. Assume now that $J_* > 0$ so that the efficient influence function is given by (3.5). In particular, if $\mathfrak{F} = \mathfrak{F}_0$, the set of all error distributions with zero means, finite variances and finite Fisher informations, then

$$\psi_\theta = u(Z) + v_0(Z)\frac{\ell(\varepsilon)}{J} + \bar{v}\varepsilon. \tag{4.2}$$

An example of a weakly differentiable functional χ is given by

$$\chi(t,\Gamma) = \int ht\,d\Gamma, \qquad t \in \Xi, T \in \mathfrak{G},$$

where h is a known bounded measurable function from $\mathscr{K}$ to $\mathbb{R}^m$. It is easy to see that Condition 3.8 holds with $N_1 = h$ and $N_2 = h\xi$. If $h = 1$, then $\chi(\xi, G) = \int \xi\,dG$ is the average regression effect and its efficient influence function is $\xi_0(Z) + \ell_*(\varepsilon)/J_*$ if $J_* > 0$ and $\dot{\mathfrak{G}} = L_*(G)$. In particular, if $\mathfrak{F} = \mathfrak{F}_0$, the set of all error distributions with zero means, finite variances and finite Fisher informations, then the sample average will be an efficient estimate of $\int \xi\,dG$.

Another weakly differentiable functional is given by

$$\chi(t,\Gamma) = \int t^2\,d\Gamma, \qquad t \in \Xi, \Gamma \in \mathfrak{G}.$$

Here Condition 3.8 holds with $N_1 = 2\xi$ and $N_2 = \xi^2$. If $\dot{\mathfrak{G}} = L_*(G)$ and $\mathfrak{F} = \mathfrak{F}_0$, then

$$\psi_\theta = \xi^2(Z) - \int \xi^2\,dG + 2\xi_0(Z)\frac{\ell(\varepsilon)}{J} + \bar{\xi}\varepsilon. \tag{4.3}$$

4.4. *Semiparametric regression*. Suppose that $\Xi = \Xi_1 \times \Xi_2$, where Ξ_1 is an open subset of $\mathbb{R}^m$ and Ξ_2 is a subset of some Banach space B_2. Write $\xi = (\beta, \tau)$, with $\beta \in \Xi_1$. Then $\dot{\Xi} = \mathbb{R}^m$, and Assumption 3.2 implies

$$\dot{\varrho}(x,y) = h^T x + \dot{\varrho}_2 y, \qquad x \in \mathbb{R}^m, y \in B_2,$$

for some $h \in L_2(G)^m$ and some linear operator $\dot{\varrho}_2$ from B_2 to $\mathbb{R}^m$. Let $\mathscr{V}_2$ denote the closure of $\{\dot{\varrho}_2 y \colon y \in \dot{\Xi}_2\}$ and set

$$h_* = h - \Pi(h|\mathscr{V}_2) \quad \text{and} \quad H_* = \int h_* h_*^T\,dG.$$

Assume now that the matrix H_* is nonsingular. Then $\mathscr{V}$ is the sum of the orthogonal subspaces $\mathscr{V}_1 = \{h_*^T a\colon a \in \mathbb{R}^m\}$ and $\mathscr{V}_2$, $\mathscr{V}_1$ has dimension m, and

$$(4.4)\qquad a = H_*^{-1} H_* a = H_*^{-1} \int h_* h^T \, dG a = H_*^{-1} \int h_* \big(h^T a + \dot{\varrho}_2 y \big) \, dG,$$

for all $a \in \mathbb{R}^m$, $y \in \dot{\Xi}_2$. Suppose now we want to estimate β. The corresponding functional is

$$\chi((b,t),\Gamma) = b, \qquad b \in \Xi_1, t \in \Xi_2, \Gamma \in \mathfrak{G}.$$

As a linear map, χ is directionally differentiable and

$$\dot{\chi}((a,b),c) = a, \qquad a \in \mathbb{R}^m, b \in B_2, c \in \dot{\mathfrak{G}}.$$

In view of (4.4), Condition 3.8 holds with $N_1 = H_*^{-1} h_*$ and $N_2 = 0$. Thus $u = 0$ and $v = N_1$. Also $d = d_1 + d_2$, where $d_1 = \Pi(1|\mathscr{V}_1) = \bar{h}_*^T H_*^{-1} h_* = \bar{h}_*^T v$ and $d_2 = \Pi(1|\mathscr{V}_2)$. Let

$$\nu = \frac{1}{J(1 - \Delta_* \bar{d})} \bar{v} = \frac{1}{J - \bar{d}(J - J_*)} H_*^{-1} \bar{h}_*$$

and

$$Q = \frac{1}{J} \big(H_*^{-1} + \Delta H_*^{-1} \bar{h}_* \bar{h}_*^T H_*^{-1} \big).$$

One verifies that $v + \bar{v} \Delta d_1 = JQh_*$. Thus the efficient influence function can be written as

$$(4.5)\qquad \psi_\theta = Q h_{*,0}(Z) \ell(\varepsilon) + \nu \Delta_* d_{2,0}(Z) \ell(\varepsilon) + \nu \ell_*(\varepsilon).$$

If $1 \in \mathscr{V}_2$, then $\bar{h}_* = 0$ and

$$(4.6)\qquad \psi_\theta = H_*^{-1} h_*(Z) \frac{\ell(\varepsilon)}{J}.$$

Let us now consider some special cases.

(a) *Partially linear additive regression.* Suppose $\mathscr{X} = \mathbb{R}^m \times \mathscr{X}_2$, $Z = (U^T, W^T)^T$, $\Xi_1 = \mathbb{R}^m$, Ξ_2 is a dense subset of $B_2 = C_b(\mathscr{X}_2)$, and

$$\varrho(Z,(b,t)) = b^T U + t(W), \qquad b \in \Xi_1, t \in \Xi_2,$$

Suppose that $E_\theta(\|U\|^2) < \infty$. Then $\dot{\Xi} = \mathbb{R}^m \times C_b(\mathscr{X}_2)$ and Assumption 3.2 holds with

$$\dot{\varrho}(x,y)(Z) = U^T x + y(W), \qquad x \in \mathbb{R}^m, y \in C_b(\mathscr{X}_2).$$

Thus $\mathscr{V}_2 = \{a \in L_2(G)\colon\ a(Z) = b(W),\ b \in L_2(G_W)\}$, where $G_W = \mathfrak{L}(W)$, $h(Z) = U$,

$$h_*(Z) = U - E_\theta(U|W),$$

and the efficient influence function is

$$\psi_\theta = H_*^{-1}(U - E_\theta(U|W)) \frac{\ell(\varepsilon)}{J}$$

provided H_* is invertible. This result has already been obtained in Bickel, Klaassen, Ritov and Wellner (1993) and Cuzick (1992a).

(b) *Semiparametric comparison of regression functions*. Let $\mathscr{X} = \{0, 1\} \times [0, 1]$, $Z = (U, W)^T$, $\Xi_1 = (0, \infty)$, Ξ_2 be a dense subset of $B_2 = C([0, 1])$,

$$\varrho(Z, (b, t)) = (1 - U)t(W) + Ubt(W), \qquad b \in \Xi_1, t \in \Xi_2.$$

Assume that U and W are independent and that $\pi = E_\theta(U) = P_\theta(U = 1) \in (0, 1)$. Then Assumption 3.2 holds with

$$\dot{\varrho}(x, y)(Z) = xU\tau(W) + (1 - U)y(W) + U\beta y(W), \qquad x \in \mathbb{R}, y \in C([0, 1]).$$

Thus $h(Z) = U\tau(W)$. Let R denote the distribution of W, and let D denote the bounded linear operator from $L_2(R)$ to $L_2(G)$ defined by

$$Da(Z) = a(W)(1 - U + U\beta), \qquad a \in L_2(R).$$

Then $\mathscr{V}_2 = \{Da\colon a \in L_2(R)\}$. Set

$$h_\# = \frac{\pi\beta\tau}{1 - \pi + \pi\beta^2} \quad \text{and} \quad d_\# = \frac{1 - \pi + \pi\beta}{1 - \pi + \pi\beta^2}.$$

One verifies $Dh_\# = \Pi(h|\mathscr{V}_2)$ and $Dd_\# = d_2 = \Pi(1|\mathscr{V}_2)$. Therefore

$$\begin{aligned} h_*(Z) &= U\tau(W) - Dh_\#(Z) = U\tau(W) - h_\#(W)(1 - U + U\beta) \\ &= \tau(W)\frac{U(1 - \pi) - (1 - U)\beta\pi}{1 - \pi + \pi\beta^2}, \end{aligned}$$

$$\bar{h}_* = \frac{\pi(1 - \pi)(1 - \beta)}{1 - \pi + \pi\beta^2}\int \tau\, dR \quad \text{and} \quad H_* = \frac{\pi(1 - \pi)}{1 - \pi + \pi\beta^2}\int \tau^2\, dR.$$

Suppose now that $\int \tau^2\, dR > 0$. Then $H_* > 0$,

$$d_1(Z) = (1 - \beta)\frac{\int \tau\, dR}{\int \tau^2\, dR} h_*(Z) \quad \text{and} \quad d_2(Z) = \frac{1 - \pi + \pi\beta}{1 - \pi + \pi\beta^2}(1 - U + U\beta).$$

Suppose furthermore that $J_* > 0$. Then the efficient influence function is given by (4.5). Substituting the above expressions and simplifying gives

$$\begin{aligned} \psi_\theta = {} & \frac{\zeta}{J\eta}\left(\frac{U}{\pi} - \beta\frac{1 - U}{1 - \pi}\right)\tau_0(W)\ell(\varepsilon) + \frac{J_*\bar{\tau}(1 - \pi + \beta\pi)}{J\pi(1 - \pi)\eta}(U - \pi)\ell(\varepsilon) \\ & + \frac{(1 - \beta)\bar{\tau}}{\eta}\ell_*(\varepsilon), \end{aligned}$$

where $\tau_0 = \tau - \bar{\tau}$, $\bar{\tau} = \int \tau\, dR$,

$$\zeta = \frac{J\pi(1 - \pi)(1 - \beta)^2 + J_*(1 - \pi + \beta\pi)^2}{1 - \pi + \beta^2\pi}$$

and

$$\eta = \zeta\int \tau_0^2\, dR + \bar{\tau}^2 J_*.$$

5. Construction of efficient estimates. We shall now address the question of how to construct efficient estimates. Our emphasis is on error models of the form $\mathfrak{F}_\psi$ as described in Remark 3.13b, but our results can be easily modified to allow for the symmetric error model $\mathfrak{F}_S$ of Remark 3.13c. Throughout this section we make the following assumption.

5.1 Assumption. Let $\{\xi_n\}$ be a local sequence for ξ. Let ψ be a function from $\mathbb{R}$ to $\mathbb{R}$ which has a bounded first and second derivative and satisfies

$$\int \psi \, dF = 0, \qquad 0 < \int \psi^2 \, dF < \infty \quad \text{and} \quad \int \psi' \, dF \neq 0.$$

Let $q, q_1, q_2, \dots$ be measurable functions from $\mathscr{K}$ to $\mathbb{R}^m$ such that

$$\int \|q\|^2 \, dG < \infty \quad \text{and} \quad \int \|q_n - q\|^2 \, dG \to 0.$$

Let $s, s_1, s_2, \dots$ be measurable functions from $\mathscr{K}$ to $\mathbb{R}^p$ such that

$$\int \|s\|^2 \, dG < \infty \quad \text{and} \quad \int \|s_n - s\|^2 \, dG \to 0.$$

Let $\mu_1, \mu_2, \dots$ be vectors in $\mathbb{R}^m$ such that $\mu_n \to \mu$ and $M, M_1, M_2, \dots$ be $m \times p$ matrices such that $M_n \to M$.

From now on let $(Y_1, Z_1), (Y_2, Z_2), \dots$ denote $\mathbb{R} \times \mathscr{K}$-valued random vectors and let $\{\mathbb{P}_x \colon x \in \Xi\}$ be probability measures such that $(Y_1, Z_1), (Y_2, Z_2), \dots$ are independent and identically distributed with distribution $P_{x, G, F}$. Let $\mathbf{Y}_n = (Y_1, \dots, Y_n)$, $\mathbf{Z}_n = (Z_1, \dots, Z_n)$, and let $\mathbb{E}_n$ denote the conditional expectation given $\mathbf{Z}_n$ calculated under $\mathbb{P}_{\xi_n}$. For $j = 1, \dots, n$, let $\mathbf{Y}_{n,j}$ denote the random vector obtained from $\mathbf{Y}_n$ by deleting its jth component Y_j, $\varepsilon_{n,j} = Y_j - \varrho(Z_j, \xi_n)$, and let $\mathbb{E}_{n,j}$ denote the conditional expectation given $(\mathbf{Y}_{n,j}, \mathbf{Z}_n)$ calculated under $\mathbb{P}_{\xi_n}$. If $\{X_n\}$ is a sequence of random vectors, $\{r_n\}$ is a sequence of positive numbers and $\{\xi_n\}$ is a sequence in Ξ, then we write $X_n = o_{\xi_n}(r_n)$ if $r_n^{-1} X_n$ converges to zero in $\mathbb{P}_{\xi_n}$-probability, that is, $\mathbb{P}_{\xi_n}(\|X_n\| > \delta r_n) \to 0$ for all $\delta > 0$; we write $X_n = \mathscr{O}_{\xi_n}(r_n)$ if $r_n^{-1} X_n$ is bounded in $\mathbb{P}_{\xi_n}$-probability, that is, $\lim_{C \to \infty} \limsup_{n \to \infty} \mathbb{P}_{\xi_n}(\|X_n\| > C r_n) = 0$.

We shall now address the following problem. Construct functions χ_n from $\mathbb{R}^n \times \mathscr{K}^n$ to $\mathbb{R}^m$ such that $\hat{\chi}_n = \chi_n(Y_1, \dots, Y_n, Z_1, \dots, Z_n)$ satisfies

$$\begin{aligned} \hat{\chi}_n = {} & \frac{1}{n} \sum_{j=1}^{n} \Big(q_n(Z_j) + M_n\big(s_n(Z_j) - \bar{s}_n\big) \ell(\varepsilon_{n,j}) + \mu_n \psi(\varepsilon_{n,j}) \Big) \\ & + o_{\xi_n}(n^{-1/2}). \end{aligned} \tag{5.1}$$

The construction of efficient estimates for the error model $\mathfrak{F} = \mathfrak{F}_\psi$ is a special case of this problem. To see this simply choose M, s, q and μ such that

$$Ms = \gamma = \frac{1}{J}(v + \bar{v} \Delta d), \qquad q = \chi(\xi, G) + u \quad \text{and} \quad \mu = \frac{\int \psi' \, dF}{\int \psi^2 \, dF} \bar{\gamma}.$$

Then an estimate satisfying (5.1) will be efficient provided $\int q_n \, dG = \chi(\xi_n, G)$. See Lemma 3.14 and Remark 3.15. A possible choice of s and M satisfying $Ms = \gamma$ is given by $M = (1/J)[I, \bar{v}\Delta]$ and $s = (v^T, d)^T$.

Let us now describe an estimate $\hat{\chi}_n$. Let $\hat{M}_n$ and $\hat{\mu}_n$ be estimates of M_n and μ_n, respectively, and let $\hat{q}_{n,j}$, $\hat{r}_{n,j}$ and $\hat{s}_{n,j}$ be estimates of $q_n(Z_j)$, $\varrho(Z_j, \xi_n)$ and $s_n(Z_j)$, respectively, $j = 1, \dots, n$. Let $\hat{w}_{n,1}, \dots, \hat{w}_{n,n}$ be $\{0,1\}$-valued random weights based on $\mathbf{Z}_n$. We have included these random weights to be able to discard some of the estimates $\hat{r}_{n,j}$ that we know are poor. See Remark 7.3 for such a use. Now set

$$N_n = \sum_{j=1}^{n} \hat{w}_{n,j}, \qquad \hat{s}_{n,*} = \frac{1}{N_n} \sum_{j=1}^{n} \hat{w}_{n,j} \hat{s}_{nj},$$

$$\hat{\varepsilon}_{n,j} = Y_j - \hat{r}_{nj} \quad \text{and} \quad \delta_{n,j} = \hat{r}_{n,j} - \varrho(Z_j, \xi_n), \qquad j = 1, \dots, n.$$

Based on the variables $\hat{\varepsilon}_{n,1}, \dots, \hat{\varepsilon}_{n,n}$ we estimate the function ℓ by

$$\hat{\ell}_n(x) = -\frac{\hat{f}_n'(x)}{\hat{f}_n(x) + b_n}, \qquad x \in \mathbb{R},$$

where $\hat{f}_n$ is the kernel density estimate of f,

$$\hat{f}_n(x) = \frac{1}{N_n} \sum_{j=1}^{n} \hat{w}_{n,j} \frac{1}{a_n} k\left(\frac{x - \hat{\varepsilon}_{n,j}}{a_n}\right), \qquad x \in \mathbb{R},$$

based on the kernel k and positive numbers a_n, b_n converging to 0. We let

$$\hat{J}_n = \frac{1}{N_n} \sum_{j=1}^{n} \hat{w}_{n,j} \hat{\ell}_n'(\hat{\varepsilon}_{n,j}) \quad \text{and} \quad \hat{\Psi}_n = \frac{1}{N_n} \sum_{j=1}^{n} \hat{w}_{n,j} \psi'(\hat{\varepsilon}_{n,j}).$$

We shall now give conditions on the estimates $\hat{q}_{nj}$, $\hat{r}_{nj}$, $\hat{s}_{nj}$, $\hat{M}_n$ and $\hat{\mu}_n$ that imply that (5.1) holds with

$$\hat{\chi}_n = \frac{1}{N_n} \sum_{j=1}^{n} \hat{w}_{n,j} \Big(\hat{q}_{nj} + \hat{M}_n (\hat{s}_{nj} - \hat{s}_{n,*}) \hat{\ell}_n(\hat{\varepsilon}_{n,j}) + \hat{\mu}_n \psi(\hat{\varepsilon}_{n,j}) \Big). \tag{5.2}$$

Condition K. The kernel k is symmetric, three times continuously differentiable, $\int t^2 k(t)\, dt < \infty$, and, for some positive constant C,

$$|k^{(i)}(x)| \le Ck(x), \qquad x \in \mathbb{R}, \ i = 1, 2, 3.$$

A possible choice is the logistic density.

Condition M. $\hat{M}_n - M_n = o_{\xi_n}(1)$ and $\hat{\mu}_n - \mu_n = o_{\xi_n}(1)$.

Under these conditions one can derive the following preliminary result.

5.2 THEOREM. *Suppose Conditions K and M hold and* $\hat{\chi}_n$ *is defined by* (5.2) *with*

$$\hat{q}_{n,j} = q_n(Z_j), \qquad \hat{r}_{n,j} = \varrho(Z_j, \xi_n), \qquad \hat{s}_{n,j} = s_n(Z_j) \quad \text{and} \quad \hat{w}_{n,j} = 1.$$

Suppose that $n^{-1}a_n^{-4}b_n^{-2} \to 0$. *Then* (5.1) *holds.*

The proof is the same as the one given in Schick (1987), Example 2, for linear regression. Thus it will be omitted. We shall show in the next section that this result is quite adequate for the construction of efficient estimates in parametric regression. It is, however, inadequate when dealing with nonparametric or semiparametric models as these models require genuine estimates of q, r and s. To deal with this more general situation we need the following additional conditions.

CONDITION R. There are positive numbers α and α_* such that, with $\hat{r}_{n,j,i} = \mathbb{E}_{n,i}(\hat{r}_{n,j})$,

$$\text{(R1)} \qquad R_{n,1} = \frac{1}{N_n} \sum_{j=1}^{n} \hat{w}_{n,j} \mathbb{E}_n\left(|\hat{r}_{n,j} - \varrho(Z_j, \xi_n)|^2\right) = \mathscr{O}_{\xi_n}(n^{-2\alpha}),$$

$$\text{(R2)} \qquad R_{n,2} = \sum_{j=1}^{n} \hat{w}_{n,j} \mathbb{E}_n\left(|\hat{r}_{n,j} - \hat{r}_{n,j,j}|^2\right) = \mathscr{O}_{\xi_n}(n^{-\alpha_*}),$$

$$\text{(R3)} \qquad R_{n,3} = \frac{1}{N_n} \sum\sum_{i \neq j} \hat{w}_{n,j} \mathbb{E}_n\left(|\hat{r}_{n,j} - \hat{r}_{n,j,i}|^2\right) = \mathscr{O}_{\xi_n}(n^{-2\alpha}).$$

CONDITION S. There are estimates $\tilde{s}_{n,j}$ based on $(\mathbf{Y}_{n,j}, \mathbf{Z}_n)$ and numbers A_n, $A_n \geq 1$, such that

$$\text{(S0)} \qquad S_{n,0} = \max_{1 \leq j \leq n} \|\hat{w}_{n,j} \tilde{s}_{n,j}\| \leq A_n,$$

$$\text{(S1)} \qquad S_{n,1} = \frac{1}{N_n} \sum_{j=1}^{n} \hat{w}_{n,j} \mathbb{E}_n\left(\|\tilde{s}_{n,j} - s_n(Z_j)\|^2\right) = o_{\xi_n}(1),$$

$$\text{(S2)} \qquad S_{n,2} = \sum_{j=1}^{n} \hat{w}_{n,j} \mathbb{E}_n\left(\|\hat{s}_{n,j} - \tilde{s}_{n,j}\|^2\right) = o_{\xi_n}(1),$$

$$\text{(S3)} \qquad S_{n,3} = \frac{1}{N_n} \sum\sum_{i \neq j} \hat{w}_{n,j} \mathbb{E}_n\left(\|\tilde{s}_{n,j} - \tilde{s}_{n,j,i}\|^2\right) = o_{\xi_n}(a_n^2).$$

CONDITION T.

$$T_n = \frac{1}{N_n} \sum_{j=1}^{n} \hat{w}_{n,j}\left(\hat{q}_{n,j} - q_n(Z_j) - \left[\hat{M}_n(\hat{s}_{n,j} - \hat{s}_{n,*})\hat{J}_n + \hat{\mu}_n \hat{\Psi}_n\right] \delta_{n,j}\right)$$
$$= o_{\xi_n}(n^{-1/2}).$$

CONDITION W. The random weights $\hat{w}_{n,j}$ are such that $N_n/n - 1 = o_{\xi_n}(1)$.

5.3 THEOREM. *Suppose Conditions K, M, R, S, T and W hold and*

$$a_n^{-4}n^{-\alpha_*} \to 0, \qquad A_n a_n^{-2} n^{1/2-2\alpha} \to 0,$$
$$A_n^2 n^{-1} a_n^{-4} b_n^{-2} \to 0, \qquad A_n^2 n^{-2\alpha} a_n^{-5} b_n^{-1} \to 0. \tag{5.3}$$

Then $\hat{\chi}_n$ *defined by* (5.2) *satisfies* (5.1).

5.4 REMARK. A necessary condition for (5.3) is that $\alpha > 1/4$. If $\alpha = \alpha_* = 1/3$, then (5.3) is implied by $b_n \geq a_n^3$ and $A_n n^{-1/6} a_n^{-2} \to 0$. Conditions R, S and T will be discussed in more detail in the sections below in the context of concrete examples. Condition S essentially calls for consistent estimates of s, while Condition R calls for estimates of the regression function that possess appropriate rates of convergence. We shall see that, under mild assumptions on the covariate distribution, Condition R is implied by appropriate smoothness conditions on the regression function. Typically, Condition S should be easy to satisfy. Thus the above approach places the difficulty of constructing efficient estimates on verifying Condition T.

We base the proof of Theorem 5.3 on the following propositions.

5.5 PROPOSITION. *Suppose Conditions K, R, S and W hold. If*

$$a_n^{-2} n^{1/2-2\alpha} \to 0,$$

then

$$\frac{1}{N_n}\sum_{j=1}^n \hat{w}_{n,j} \int \left(\hat{\ell}_n'(y-\delta_{n,j}) - \hat{\ell}_n'(y)\right) dF(y) = o_{\xi_n}(n^{\alpha-1/2}),$$

and if $A_n a_n^{-2} n^{1/2-2\alpha} \to 0$, *then*

$$\frac{1}{N_n}\sum_{j=1}^n \hat{w}_{n,j}\hat{s}_{n,j} \int \left(\hat{\ell}_n(y-\delta_{n,j}) - \hat{\ell}_n(y) + \hat{\ell}_n'(y)\delta_{n,j}\right) dF(y) = o_{\xi_n}(n^{-1/2}).$$

5.6 PROPOSITION. *Suppose Conditions K, R, S and W hold,*

$$a_n^{-4} n^{-\alpha_*} \to 0 \quad \text{and} \quad A_n^2\left(n^{-1}a_n^{-4}b_n^{-2} + n^{-2\alpha}a_n^{-5}b_n^{-1}\right) \to 0.$$

Then

$$\frac{1}{N_n}\sum_{j=1}^n \hat{w}_{n,j}\left(\hat{s}_{n,j}\left(\hat{\ell}_n(\hat{\varepsilon}_{n,j}) - \int \hat{\ell}_n(y-\delta_{n,j})\, dF(y)\right) - s_n(Z_j)\ell(\varepsilon_j)\right)$$
$$= o_{\xi_n}(n^{-1/2}).$$

5.7 PROPOSITION. *Suppose Conditions K, R and W hold,*

$$n^{-2\alpha-\alpha_*}a_n^{-6} \to 0, \qquad n^{-2\alpha}a_n^{-4} \to 0,$$
$$n^{-4\alpha}a_n^{-7}b_n^{-1} \to 0 \quad \text{and} \quad n^{-1-2\alpha}a_n^{-6}b_n^{-2} \to 0.$$

Then

$$\frac{1}{N_n}\sum_{j=1}^{n}\hat{w}_{n,j}\Big(\hat{\ell}_n'(\hat{\varepsilon}_{n,j}) - \int\hat{\ell}_n'(y-\delta_{n,j})\,dF(y)\Big) = o_{\xi_n}(n^{\alpha-1/2}).$$

5.8 PROPOSITION. *Suppose Condition R holds and η is a function from $\mathbb{R}$ to $\mathbb{R}$ which is F-square integrable and has a bounded derivative. Then*

$$\frac{1}{N_n}\sum_{j=1}^{n}\hat{w}_{n,j}\eta(\hat{\varepsilon}_{n,j}) = \frac{1}{N_n}\sum_{j=1}^{n}\hat{w}_{n,j}\eta(\varepsilon_{n,j}) + \int\eta(y-\delta_{n,j})\,dF(y) + o_{\xi_n}(n^{-1/2})$$

and

$$\frac{1}{N_n}\sum_{j=1}^{n}\hat{w}_{n,j}\int\eta(y-\delta_{n,j})\,dF(y) = \int\eta\,dF - \frac{1}{N_n}\sum_{j=1}^{n}\hat{w}_{n,j}\delta_{n,j}\int\eta'\,dF + \mathscr{O}_{\xi_n}(n^{-2\alpha}).$$

PROOF OF THEOREM 5.3. Propositions 5.5 and 5.7 yield $\hat{\mathscr{J}}_n = \int\hat{\ell}_n'\,dF + o_{\xi_n}(n^{\alpha-1/2})$. This and Propositions 5.5 and 5.6 give

$$\frac{1}{N_n}\sum_{j=1}^{n}\hat{w}_{n,j}\hat{s}_{n,j}\Big(\hat{\ell}_n(\hat{\varepsilon}_{n,j}) - \int\hat{\ell}_n\,dF - \hat{\mathscr{J}}_n\delta_{n,j}\Big) = \frac{1}{N_n}\sum_{j=1}^{n}\hat{w}_{n,j}s_n(Z_j)\ell(\varepsilon_{n,j}) + o_{\xi_n}(n^{-1/2}).$$

It follows from this, Condition W and Condition M that

$$\frac{1}{N_n}\sum_{j=1}^{n}\hat{w}_{n,j}\hat{M}_n\big[\hat{s}_{n,j} - \hat{s}_{n,*}\big]\Big(\hat{\ell}_n(\hat{\varepsilon}_{n,j}) + \hat{\mathscr{J}}_n\delta_{n,j}\Big) = \frac{1}{n}\sum_{j=1}^{n}M_n\big(s_n(Z_j) - \bar{s}_n\big)\ell(\varepsilon_{n,j}) + o_{\xi_n}(n^{-1/2}). \tag{5.4}$$

It follows from Proposition 5.8 applied with $\eta = \psi$ and $\eta = \psi'$ that

$$\frac{1}{N_n}\sum_{j=1}^{n}\hat{w}_{n,j}\big(\psi(\hat{\varepsilon}_{n,j}) + \hat{\Psi}_n\delta_{n,j}\big) = \frac{1}{N_n}\sum_{j=1}^{n}\hat{w}_{n,j}\psi(\varepsilon_{n,j}) + o_{\xi_n}(n^{-1/2}).$$

This, Condition W and Condition M yield

$$\frac{1}{N_n}\sum_{j=1}^{n}\hat{w}_{n,j}\hat{\mu}_n\big(\psi(\hat{\varepsilon}_{n,j}) + \hat{\Psi}_n\delta_{n,j}\big) = \frac{1}{n}\sum_{j=1}^{n}\mu_n\psi(\varepsilon_{n,j}) + o_{\xi_n}(n^{-1/2}). \tag{5.5}$$

Combining (5.4) and (5.5) shows that

$$\hat{\chi}_n - \frac{1}{n}\sum_{j=1}^{n}\Big(q_n(Z_j) + M_n\big(s_n(Z_j) - \bar{s}_n\big)\ell\big(Y_j - \varrho(Z_j,\xi_n)\big) + \mu_n\psi\big(Y_j - \varrho(Z_j,\xi_n)\big)\Big) = T_n + o_{\xi_n}(n^{-1/2}).$$

The desired result follows now from Condition T. □

5.9 Remark. We have shown in the proof of Theorem 5.3 that $\hat{J}_n = \int\hat{\ell}_n'\,dF + o_{\xi_n}(n^{\alpha-1/2})$. Integration by parts yields $\int\hat{\ell}_n'\,dF = \int\hat{\ell}_n\ell\,dF$. It can be shown that $\int(\hat{\ell}_n - \ell)^2\,dF = o_{\xi_n}(1)$. See Lemma 10.2 and Lemma 10.5 for the relevant arguments. Thus $\hat{J}_n = J + o_{\xi_n}(1)$.

5.10 Remark. If $\bar{s}_n = 0$ and $\hat{s}_{n,*} = \mathscr{O}_{\xi_n}(n^{-1/2})$, then we may replace $\hat{s}_{n,*}$ by 0 in (5.2).

6. Parametric models. Let us now demonstrate how one can apply Theorem 5.2 to obtain efficient estimates in parametric models. We assume that $\mathfrak{F} = \mathfrak{F}_0$ as given in Remark 3.13(a). Thus F has finite Fisher information and satisfies

$$\int x\,dF(x) = 0 \quad\text{and}\quad \int x^2\,dF(x) = \sigma^2 < \infty.$$

Suppose that Ξ is an open subset of $\mathbb{R}^m$ and that there is a known map h from $\mathscr{K}\times\Xi$ to $\mathbb{R}^m$ such that for all $x \in \Xi$, $\int\|h(z,x)\|^2\,dG(z) < \infty$,

$$\int\big(\varrho(z,x+t) - \varrho(z,x) - t^T h(z,x)\big)^2\,dG(z) = o(\|t\|^2)$$

and

$$\int\|h(z,x+t) - h(z,x)\|^2\,dG(z) = o(1),$$

as $\|t\| \to 0$. Suppose also that the matrices

$$H(x) = \int h(z,x)h(z,x)^T\,dG(z), \qquad x \in \Xi,$$

are nonsingular. It was shown in Section 4 [see (4.1) and Remark 3.13(a) for the form of ℓ_* and J_*] that an efficient estimate of ξ is efficient if it satisfies

$$\hat{\chi}_n = \frac{1}{n}\sum_{j=1}^{n} L\big(Y_j - \varrho(Z_j,\xi),Z_j,\xi\big) + o_\xi(n^{-1/2}), \tag{6.1}$$

where

$$L(y,z,x) = x + Q^{-1}(x)\big(h(z,x) - \bar{h}(x)\big)\ell(y) + \frac{1}{\sigma^2}\bar{h}(x)y,$$
$$y \in \mathbb{R},\ z \in \mathscr{K},\ x \in \Xi,$$

$$Q(x) = JH(x) + \left(\frac{1}{\sigma_2} - J\right)\bar{h}(x)\bar{h}(x)^T \quad \text{and}$$

$$\bar{h}(x) = \int h(z,x)\,dG(z), \qquad x \in \Xi.$$

Let us now construct such an estimate. For this let $\hat{\xi}_n$ denote an estimate of ξ. Define $\hat{J}_n$ and $\hat{\ell}_n$ as in the previous section with $\hat{w}_{n,j} = 1$ and $\hat{\varepsilon}_{n,j} = Y_j - \varrho(Z_j, \hat{\xi}_n)$, and set

$$\hat{h}_{n,*} = \frac{1}{n}\sum_{j=1}^{n} h\big(Z_j, \hat{\xi}_n\big), \qquad \hat{H}_n = \frac{1}{n}\sum_{j=1}^{n} h\big(Z_j, \hat{\xi}_n\big)h\big(Z_j, \hat{\xi}_n\big)^T,$$

$$\hat{Q}_n = \hat{J}_n\hat{H}_n + \left(\frac{1}{\hat{\sigma}_n^2} - \hat{J}_n\right)\hat{h}_{n,*}\hat{h}_{n,*} \quad \text{and} \quad \hat{\sigma}_n^2 = \frac{1}{n}\sum_{j=1}^{n}\hat{\varepsilon}_{n,j}^2.$$

6.1 THEOREM. *Suppose the preliminary estimate $\hat{\xi}_n$ is a discrete $\sqrt{n}$-consistent estimate of ξ, and $n^{-1}a_n^{-4}b_n^{-2} \to 0$. Then the estimate*

$$\hat{\chi}_n = \hat{\xi}_n + \hat{Q}_n^{-1}\frac{1}{n}\sum_{j=1}^{n}\left(\big(h\big(Z_j,\hat{\xi}_n\big) - \hat{h}_{n,*}\big)\hat{\ell}_n(\hat{\varepsilon}_{n,j}) + \frac{1}{\hat{\sigma}_n^2}\hat{h}_{n,*}\hat{\varepsilon}_{n,j}\right)$$

satisfies (6.1) *and is therefore efficient.*

PROOF. By the properties of the preliminary estimate $\hat{\xi}_n$ it suffices to prove the result under the assumption that $\hat{\xi}_n$ is a local sequence. Thus assume from now on that $\hat{\xi}_n$ is a local sequence. In view of Lemma 3.14 and Remark 3.15, it suffices to show that

$$\hat{\chi}_n = \frac{1}{n}\sum_{j=1}^{n} L\Big(Y_j - \varrho\big(Z_j, \hat{\xi}_n\big), Z_j, \hat{\xi}_n\Big) + o_{\hat{\xi}_n}(n^{-1/2}).$$

But this follows from Theorem 5.2 as $\hat{Q}_n = Q(\hat{\xi}_n) + o_{\hat{\xi}_n}(1)$ and $(1/\hat{\sigma}_n^2)\hat{h}_{n,*} = (1/\sigma^2)\bar{h}(\hat{\xi}_n) + o_{\hat{\xi}_n}(1)$. □

6.2 EXAMPLE. In linear regression one assumes that

$$\varrho(z,x) = x^T z, \qquad z \in \mathscr{K},\ x \in \mathbb{R}^m,$$

and that $\int zz^T\,dG$ is invertible. In this case $h(z,x) = z$, $z \in \mathscr{K}$, $x \in \Xi$. A $\sqrt{n}$-consistent estimate is given by the least squares estimate. If $\mathscr{K}$ is a compact subset of $\mathbb{R}^m$ and $\hat{\xi}_n$ is the least squares estimate, then discretization can be avoided by using Theorem 5.3 with $\xi_n = \xi$. One verifies Condition R with $\hat{r}_{n,j} = \hat{\xi}_n^T Z_j$ and $\alpha = \alpha_* = 1/2$, Condition S with $A_n = \sup_{z \in \mathscr{K}} \|z\|$ and Condition T with $T_n = 0$.

7. Nonparametric models. Assume now that $\mathscr{K} = [0, 1]$, G has a Lebesgue density g, $\mathfrak{F} = \mathfrak{F}_0$, so that F has mean 0 and finite variance σ^2, and that

$$\varrho(z, t) = t(z), \qquad z \in \mathscr{K}, t \in \Xi,$$

with Ξ a subset of $C([0, 1])$. Let us first address Condition R.

7.1 LEMMA. *Suppose there are positive numbers C, a, b such that $b > 1/2$, $\inf_{0 \le z \le 1} g(z) \ge a$, and*

$$|\xi(z_1) - \xi(z_2)| \le C|z_1 - z_2|^b, \qquad z_1, z_2 \in [0, 1]. \tag{7.1}$$

Then there are estimates $\hat{r}_{n,j}$ which satisfy Condition R with $\hat{w}_{n,j} = 1$, $\xi_n = \xi$, $\alpha = b/(1 + 2b)$ and $\alpha_* = (2b - 1)/(1 + 2b)$.

PROOF. Let m_n be a sequence of positive integers such that $m_n \sim n^{1/(1+2b)}$. Partition $[0, 1]$ into m_n intervals $I_{n,1}, \dots, I_{n,m_n}$ of equal lengths m_n^{-1}. For $l = 1, \dots, m_n$, let $N_{n,l} = \sum_{j=1}^n \mathbf{1}_{I_{n,l}}(Z_j)$ be the number of observations Z_j in the interval $I_{n,l}$ and $Y_{n,l} = \sum_{j=1}^n Y_j \mathbf{1}_{I_{n,l}}(Z_j)/N_{n,l}$ be the average of those Y_j for which Z_j belongs to $I_{n,l}$. Let

$$\hat{\xi}_n(z) = \sum_{l=1}^{m_n} \mathbf{1}_{I_{n,l}}(z) Y_{n,l}, \qquad z \in \mathscr{K}. \tag{7.2}$$

With $\hat{r}_{n,j} = \hat{\xi}_n(Z_j)$ we verify

$$R_{n,1} \le \sigma^2 \sum_{l=1}^{m_n} \frac{I[N_{n,l} > 0]}{N_{n,l}} = \mathscr{O}_\xi(m_n^2 n^{-1}),$$

$$R_{n,2} \le \frac{\sigma^2 m_n}{n} + C^2 m_n^{-2b} \quad \text{and} \quad R_{n,3} \le \frac{\sigma^2 m_n}{n}.$$

We used the fact that $\max_{1 \le l \le n} N_{n,l}^{-1} = \mathscr{O}_\xi(m_n/n)$. The desired result follows now from the choice of m_n. □

7.2 REMARK. Note that $b/(1 + 2b)$ is the optimal rate discussed in Stone (1982).

7.3 REMARK. With $\hat{r}_{n,j}$ as in the above proof and

$$\hat{w}_{n,j} = \sum_{l=1}^{m_n} \mathbf{1}_{I_{n,l}}(Z_j) I[N_{n,l} > d_n],$$

with $d_n \sim n/(m_n \log(1 + n))$, we can verify Condition R with $\alpha = b(1 + 2b)$ and $\alpha_* < (2b - 1)/(1 + 2b)$ without the requirement that g is bounded away from 0 on $[0, 1]$. Simply note that we can now bound $R_{n,1}$ by

$$R_{n,1} \le \sigma^2 \sum_{l=1}^{m_n} \frac{I[N_{n,l} > d_n]}{N_{n,l}} \le \frac{\sigma^2 m_n}{d_n}.$$

Of course, in this case one needs to verify Condition W. However, Condition W can be verified even if g is not bounded away from 0, for example, if g is bounded and bounded away from 0 on each compact subinterval of $(0, 1)$.

In what follows we assume that Ξ is the set of all Lipschitz-continuous functions from $[0, 1]$ to $\mathbb{R}$ and that $\mathfrak{G}$ is the set of all distributions which have Lebesgue densities that are bounded away from 0 on $[0, 1]$. Then $\dot{\Xi} = L_2(G)$ and $\dot{\mathfrak{G}} = L_*(G)$. Thus every weakly differentiable functional χ satisfies Condition 3.8, and an estimator $\hat{\chi}_n$ is efficient if it satisfies

$$(7.3) \qquad \hat{\chi}_n = \frac{1}{n} \sum_{j=1}^{n} q(Z_j) + v_0(Z_j) \frac{\ell(\varepsilon_j)}{J} + \bar{v}\varepsilon_j + o_\xi(n^{-1/2}),$$

where $q(z) = \chi(\xi, G) + N_2(z) - \overline{N}_2$ and $v = N_1$. Let now $\hat{w}_{n,j} = 1$ and $\hat{r}_{n,j}$ be as defined in the above proof with $m_n \sim n^{1/3}$. Now consider the estimate

$$(7.4) \qquad \hat{\chi}_n = \frac{1}{n} \sum_{j=1}^{n} \hat{q}_{n,j} + \left(\hat{v}_{n,j} - \hat{v}_{n,*}\right) \frac{\hat{\ell}_n(\hat{\varepsilon}_{n,j})}{\hat{J}_n} + \hat{v}_{n,*}\hat{\varepsilon}_{n,j},$$

where $\hat{q}_{n,j}$ and $\hat{v}_{n,j}$ are estimates of $q(Z_j)$ and $v(Z_j)$, and $\hat{v}_{n,*} = (1/n)\sum_{j=1}^{n} v_{n,j}$.

7.4 THEOREM. *Suppose Condition S holds with*

$$\hat{s}_{n,j} = \hat{v}_{n,j}, \qquad A_n n^{-1/6} a_n^{-2} \to 0, \qquad b_n \ge a_n^3$$

and

$$(7.5) \qquad T_n = \frac{1}{n} \sum_{j=1}^{n} \hat{q}_{n,j} - q(Z_j) - \hat{v}_{n,j}\left(\hat{r}_{n,j} - \xi(Z_j)\right) = o_\xi(n^{-1/2}).$$

Then the estimate defined in (7.4) *satisfies* (7.3) *and is therefore efficient.*

PROOF. We apply Theorem 5.3 with $\xi_n = \xi$. As $\hat{w}_{n,j} = 1$, Condition W holds. An application of the above lemma with $b = 1$ yields Condition R with $\alpha = \alpha_* = 1/3$. Condition S is assumed. It implies that $\hat{v}_{n,*} = \bar{v} + o_\xi(1)$, and this implies Condition M. Condition T follows from (7.5). Finally, (5.3) follows from the specified rates. Thus Theorem 5.3 applies and gives the desired result. □

7.5 EXAMPLE. Let us now consider the functional χ defined by

$$\chi(x, G) = \int h(z) x(z) \, d\Gamma(z), \qquad x \in \Xi, \Gamma \in \mathfrak{G},$$

where h is a known bounded measurable function from $[0,1]$ to $\mathbb{R}$. For this functional $q = h\xi$ and $v = h$. We take $\hat{q}_{n,j} = h(Z_j)\hat{r}_{n,j}$ and $\hat{v}_{n,j} = h(Z_j)$. Then Condition S holds with $A_n = \sup_{0\le z\le 1}|h(z)|$, and (7.5) holds as $T_n = 0$.

7.6 Example. Now consider the functional χ defined by

$$\chi(x,\Gamma) = \int x^2(z)\,d\Gamma(z), \qquad x \in \Xi, \Gamma \in \mathfrak{G}.$$

We have $q = \xi^2$ and $v = 2\xi$. Let us take $\hat{q}_{n,j} = \hat{r}_{n,j}^2$ and $\hat{v}_{n,j} = 2\hat{r}_{n,j}$. Verify that

$$\max_{1\le j\le n} \left|\hat{r}_{n,j} - \xi(Z_j)\right| = o_\xi(1).$$

It is now easy to see that Condition S holds with $A_n = \sup_{0\le z\le 1}|\xi(z)| + 1$. Finally, (7.5) follows as

$$\begin{aligned} T_n &= \frac{1}{n}\sum_{j=1}^n \hat{r}_{n,j}^2 - \xi^2(Z_j) - 2\hat{r}_{n,j}\big(\hat{r}_{n,j} - \xi(Z_j)\big) \\ &= -\frac{1}{n}\sum_{j=1}^n \big(\hat{r}_{n,j} - \xi(Z_j)\big)^2 = \mathscr{O}_\xi(n^{-2/3}). \end{aligned}$$

8. Efficient estimation in the partly linear additive regression model. Let us now derive an efficient estimate in the partly linear additive model. In this model we take $\mathscr{K} = \mathbb{R}^m \times [0,1]$ and write $Z = (U^T, W)^T$. We take Ξ_2 to be the set of all functions from $[0,1]$ to $\mathbb{R}$ which are Lipschitz continuous. Recall that ϱ is given by

$$\varrho(Z,(b,t)) = b^T U + t(W), \qquad b \in \Xi_1 = \mathbb{R}^m, t \in \Xi_2.$$

We assume from now on that $\mathfrak{F} = \mathfrak{F}_0$, that U is bounded and that the distribution of W has a density with respect to the Lebesgue measure which is bounded away from 0 on $[0,1]$. Let

$$h(Z) = U - E_\theta(U|W) \quad \text{and} \quad H = \int hh^T\,dG.$$

We assume also that H is nonsingular. Then an efficient estimator $\tilde{\beta}_n$ of β must satisfy

$$\tilde{\beta}_n - \beta - H^{-1}\frac{1}{n}\sum_{j=1}^n h(Z_j)\frac{\ell(\varepsilon_j)}{J} = o_\xi(n^{-1/2}).$$

Let $\hat{\beta}_n$ be an estimate of β. We construct estimates of τ and h as follows. Partition $[0,1]$ into m_n^2 intervals $A_{n,1},\dots,A_{n,m_n^2}$ of equal length, so that

$A_{n,1} = [0, m_n^{-2})$, $A_{n,m_n^2} = [1 - m_n^{-2}, 1]$ and $A_{n,j} = [(j-1)m_n^{-2}, jm_n^{-2})$, $j = 2, \dots, m_n^2 - 1$, and let $B_{n,i} = \bigcup_{j=1}^{m_n} A_{n,(i-1)m_n+j}$, $i = 1, \dots, m_n$. Let

$$\hat{\tau}_n(W_j) = \sum_{l=1}^{m_n} 1_{B_{n,l}}(W_j) \frac{\sum_{i=1}^n 1_{B_{n,l}}(W_i)(Y_i - \hat{\beta}_n^T U_i)}{\sum_{i=1}^n 1_{B_{n,l}}(W_i)},$$

$$\hat{h}_{n,j} = U_j - \sum_{l=1}^{m_n^2} \frac{1_{A_{n,l}}(W_j)}{\sum_{a=1}^n 1_{A_{n,l}}(W_a)} \sum_{i=1}^n 1_{A_{n,l}}(W_i) U_i,$$

$$\hat{\varepsilon}_{n,j} = Y_j - \hat{\beta}_n^T U_j - \hat{\tau}_n(W_j) \quad \text{and} \quad \hat{w}_{n,j} = 1,\ j = 1, \dots, n.$$

Define

$$\tilde{\beta}_n = \hat{\beta}_n + \hat{H}_n^{-1} \frac{1}{n} \sum_{j=1}^n \hat{h}_{n,j} \hat{\ell}_n(\hat{\varepsilon}_{n,j}) / \hat{J}_n, \tag{8.1}$$

where

$$\hat{H}_n = \frac{1}{n} \sum_{j=1}^n \hat{h}_{n,j} \hat{h}_{n,j}^T.$$

8.1 THEOREM. *Let $\hat{\beta}_n$ be discrete and $\sqrt{n}$-consistent. Let $m_n \sim n^{1/3}$, $n^{-1/6} a_n^{-2} \to 0$ and $b_n \geq a_n^3$. Then $\tilde{\beta}_n$ defined in* (8.1) *is efficient.*

PROOF. Assume first that $\{\hat{\beta}_n\}$ is a local sequence. Let $\xi_n = (\hat{\beta}_n, \tau)$. It suffices to show that

$$\tilde{\beta}_n = \hat{\beta}_n + \frac{1}{n} \sum_{j=1}^n H^{-1} h(Z_j) \frac{\ell(Y_j - \hat{\beta}_n^T U_j - \tau(W_j))}{J} + o_{\xi_n}(n^{-1/2}).$$

For this we apply Theorem 5.3 with $\hat{r}_{n,j} = \hat{\beta}_n^T U_j + \hat{\tau}_n(W_j)$, $\hat{s}_{n,j} = \hat{h}_{n,j}$, $s_n = s = h$, $\hat{M}_n = \hat{H}_n^{-1}$, $M_n = M = H^{-1}$, $\hat{q}_{n,j} = q_n = \hat{\beta}_n$, and $\hat{\mu}_n = \mu_n = 0$. Clearly, Condition W holds. Condition S holds with A_n a constant as U is assumed to be bounded. Condition M is implied by Condition S. Condition R is verified with $\alpha = \alpha_* = 1/3$. Finally, Condition T holds as

$$T_n = \frac{1}{n} \sum_{j=1}^n \hat{h}_{n,j} (\hat{\tau}_n(W_j) - \tau(W_j)) = -\frac{1}{n} \sum_{j=1}^n \hat{h}_{n,j} \tau(W_j) = o_{\xi_n}(n^{-1/2}).$$

We used $\sum_{j=1}^n \hat{h}_{n,j} 1_{A_{n,l}}(W_j) = 0$, $\hat{\tau}_n$ is constant on each $A_{n,l}$ and $\sup_{x,y \in A_{n,l}} |\tau(x) - \tau(y)| \leq L n^{-2/3}$ for some positive constant L and all n and $l = 1, \dots, m_n^2$. This proves the result in the case that $\{\hat{\beta}_n\}$ is a local sequence. The general case follows as $\{\hat{\beta}_n\}$ is discrete and $\sqrt{n}$-consistent. □

8.2 REMARK. Chen (1988), Cuzick (1992a) and Robinson (1988) have constructed $\sqrt{n}$-consistent estimates under stronger assumptions than imposed here. However, the type of estimate considered by these authors can be shown to be $\sqrt{n}$-consistent under our present conditions as shown next. Let

$$\hat{\beta}_n = \hat{H}_n^{-1} \frac{1}{n} \sum_{j=1}^{n} \hat{h}_{n,j} Y_j.$$

Then $\{\hat{\beta}_n\}$ is a $\sqrt{n}$-consistent estimate of β. Indeed, since

$$\frac{1}{n} \sum_{j=1}^{n} \hat{h}_{n,j} \tau(W_j) = o_\xi(n^{-1/2}) \quad \text{and} \quad \frac{1}{n} \sum_{j=1}^{n} \|\hat{h}_{n,j} - h(Z_j)\|^2 = o_\xi(1),$$

we have

$$\begin{aligned}\frac{1}{n} \sum_{j=1}^{n} \hat{h}_{n,j} Y_j &= \frac{1}{n} \sum_{j=1}^{n} \hat{h}_{n,j} U_j^T \beta + \frac{1}{n} \sum_{j=1}^{n} \hat{h}_{n,j} \tau(W_j) \\ &\quad + \frac{1}{n} \sum_{j=1}^{n} \left(\hat{h}_{n,j} - h(Z_j)\right)\varepsilon_j + \frac{1}{n} \sum_{j=1}^{n} h(Z_j)\varepsilon_j \\ &= \hat{H}_n \beta + \frac{1}{n} \sum_{j=1}^{n} h(Z_j)\varepsilon_j + o_\xi(n^{-1/2}),\end{aligned}$$

and the desired result follows.

8.3 Remark. Cuzick (1992b) has also constructed an efficient estimator of β. While we use all the data to construct an estimate of the score function ℓ, he uses only a small part of the sample to estimate it. In addition, he imposes stronger conditions on the function τ and requires that the conditional expectation of U given W has a twice continuously differentiable version. Our construction shows that this latter condition is not needed.

8.4 Remark. We believe that discretization can be avoided at the expense of a more difficult proof. The requirement that τ is Lipschitz continuous can be relaxed. Of interest are extensions of the above to the case when W is a p-dimensional random vector and τ is additive. These modifications will be carried out somewhere else.

9. Semiparametric comparison of regression functions. Let $\mathscr{K} = \{0,1\} \times [0,1]$, $\Xi_1 = (0,\infty)$, Ξ_2 be the set of all positive Lipschitz continuous functions from $[0,1]$ to $\mathbb{R}$, and let

$$\varrho(Z,(b,t)) = (1-U)t(W) + Ubt(W) = t(W)(1-U+Ub),$$

for $b > 0$, $t \in \Xi_2$. We assume that $\mathfrak{F} = \mathfrak{F}_0$, that U and W are independent, that $\pi = E_\theta(U) = P_\theta(U=1) \in (0,1)$ and that R, the distribution of W, has a Lebesgue density that is bounded and bounded away from zero on $[0,1]$. Let

$$\bar{\tau} = \int \tau \, dR, \qquad \tau_0 = \tau - \bar{\tau}.$$

For $b > 0$, let

$$h(Z, b) = \tau_0(W)\left(\frac{U}{\pi} - b\frac{1-U}{1-\pi}\right),$$

$$\zeta_1(b) = \frac{\sigma^2 J\pi(1-\pi)(1-b)^2 + (1-\pi+b\pi)^2}{1-\pi+b^2\pi}, \qquad \zeta_2(b) = \frac{\bar{\tau}(1-\pi+b\pi)}{\pi(1-\pi)}$$

and

$$\eta(b) = \zeta_1(b)\int \tau_0^2\, dR + \bar{\tau}^2.$$

Then an efficient estimator $\hat{\chi}_n$ of β must satisfy

$$\hat{\chi}_n = \beta + \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\zeta_1(\beta)}{J\eta(\beta)}h(Z_j, \beta)\ell(\varepsilon_j) + \frac{\zeta_2(\beta)}{J\eta(\beta)}(U_j - \pi)\ell(\varepsilon_j) + \frac{(1-\beta)\bar{\tau}}{\eta(\beta)}\varepsilon_j\right) + o_\xi(n^{-1/2}). \tag{9.1}$$

Let $\hat{\beta}_n$ be an estimate of β. We estimate τ as follows. Partition $[0, 1]$ into m_n intervals $B_{n,1}, \dots, B_{n,m_n}$ of equal length m_n^{-1} and set

$$\hat{\tau}_n(w) = \sum_{l=1}^{m_n} \mathbf{1}_{B_{n,l}}(w)\frac{\sum_{j=1}^n\left(1 - U_j + (1/\hat{\beta}_n)U_j\right)Y_j\mathbf{1}_{B_{n,l}}(W_j)}{\sum_{j=1}^n \mathbf{1}_{B_{n,l}}(W_j)}, \qquad w \in [0, 1],$$

Construct the estimates $\hat{\ell}_n$ and $\hat{J}_n$ with

$$\hat{\varepsilon}_{n,j} = Y_j - \left(1 - U_j + \hat{\beta}_n U_j\right)\hat{\tau}_n(W_j) \quad \text{and} \quad \hat{w}_{n,j} = 1,\ j = 1, \dots, n.$$

Let

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{j=1}^n \hat{\varepsilon}_{n,j}^2, \qquad \bar{\tau}_n = \frac{1}{n}\sum_{j=1}^n \hat{\tau}_n(W_j), \qquad \hat{\pi}_n = \frac{1}{n}\sum_{j=1}^n U_j,$$

$$\hat{\zeta}_{1,n} = \frac{\hat{\sigma}_n^2\hat{J}_n\hat{\pi}_n(1-\hat{\pi}_n)\left(1-\hat{\beta}_n\right)^2 + \left(1 - \hat{\pi}_n + \hat{\beta}_n\hat{\pi}_n\right)^2}{1 - \hat{\pi}_n + \hat{\beta}_n^2\hat{\pi}_n},$$

$$\hat{\zeta}_{2,n} = \frac{\bar{\tau}_n\left(1 - \hat{\pi}_n + \hat{\beta}_n\hat{\pi}_n\right)}{\hat{\pi}_n(1-\hat{\pi}_n)},$$

$$\hat{\eta}_n = \hat{\zeta}_{1,n}\frac{1}{n}\sum_{j=1}^n\left(\hat{\tau}_n(W_j) - \bar{\tau}_n\right)^2 + \bar{\tau}_n^2$$

and

$$\hat{h}_{n,j} = \left(\hat{\tau}_n(W_j) - \bar{\tau}_n\right)\left(\frac{U_j}{\hat{\pi}_n} - \hat{\beta}_n\frac{1-U_j}{1-\hat{\pi}_n}\right).$$

Define

$$\hat{\chi}_n = \hat{\beta}_n + \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{\zeta}_{1,n}}{\hat{J}_n\hat{\eta}_n}\left(\hat{h}_{n,j} - \hat{h}_{n,\star}\right)\hat{\ell}_n(\hat{\varepsilon}_{n,j}) + \frac{\hat{\zeta}_{2,n}}{\hat{J}_n\hat{\eta}_n}(U_j - \hat{\pi}_n)\hat{\ell}_n(\hat{\varepsilon}_{n,j}) + \frac{\left(1-\hat{\beta}_n\right)\bar{\tau}_n}{\hat{\eta}_n}\hat{\varepsilon}_{n,j}\right). \tag{9.2}$$

9.1 THEOREM. *Let $\hat{\beta}_n$ be discrete and $\sqrt{n}$-consistent. Suppose $m_n \sim n^{1/3}$, $n^{-1/6}a_n^{-2} \to 0$ and $b_n \geq a_n^3$. Then $\hat{\chi}_n$ defined in* (9.2) *satisfies* (9.1) *and is therefore efficient.*

PROOF. Assume first that $\{\hat{\beta}_n\}$ is a local sequence. Let $\xi_n = (\hat{\beta}_n, \tau)$. It suffices to show that

$$\hat{\chi}_n = \hat{\beta}_n + \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\zeta_1(\hat{\beta}_n)}{J\eta(\hat{\beta}_n)}h\left(Z_j, \hat{\beta}_n\right)\ell(\varepsilon_j) + \frac{\zeta_2(\hat{\beta}_n)}{J\eta(\hat{\beta}_n)}(U_j - \pi)\ell(\varepsilon_j) + \frac{\left(1-\hat{\beta}_n\right)\bar{\tau}}{\eta(\hat{\beta}_n)}\varepsilon_j\right) + o_{\xi_n}(n^{-1/2}).$$

For this we apply Theorem 5.3 with $\hat{q}_{n,j} = q_n = \hat{\beta}_n$,

$$\hat{r}_{n,j} = \left(1 - U_j + U_j\hat{\beta}_n\right)\hat{\tau}_n(W_j), \qquad \hat{s}_{n,j} = \begin{pmatrix}\hat{h}_{n,j}\\ U_j - \hat{\pi}_n\end{pmatrix},$$

$$s_n(Z) = \begin{pmatrix}h\left(Z, \hat{\beta}_n\right)\\ U - \pi\end{pmatrix},$$

$$\hat{M}_n = \left[\frac{\hat{\zeta}_{1,n}}{\hat{J}_n\hat{\eta}_n}, \frac{\hat{\zeta}_{2,n}}{\hat{J}_n\hat{\eta}_n}\right], \qquad M_n = \left[\frac{\zeta_1(\hat{\beta}_n)}{J\eta(\hat{\beta}_n)}, \frac{\zeta_2(\hat{\beta}_n)}{J\eta(\hat{\beta}_n)}\right],$$

$$\hat{\mu}_n = \frac{\left(1-\hat{\beta}_n\right)\bar{\tau}_n}{\hat{\eta}_n} \quad\text{and}\quad \mu_n = \frac{\left(1-\hat{\beta}_n\right)\bar{\tau}}{\eta(\hat{\beta}_n)}.$$

One verifies that Conditions R and S hold with $\alpha = \alpha_* = 1/3$ and all sequences A_n, $A_n \to \infty$. Condition M is easily verified. Condition W is met. We are left to verify Condition T. First note that

$$T_n = \frac{1}{n\hat{\eta}_n}\sum_{j=1}^{n}\left(\hat{\zeta}_{1,n}\left(\hat{h}_{n,j} - \hat{h}_{n,\star}\right) + \hat{\zeta}_{2,n}(U_j - \hat{\pi}_n) + \left(1-\hat{\beta}_n\right)\bar{\tau}_n\right) \times\left(1 - U_j + \hat{\beta}_n U_j\right)\left(\hat{\tau}_n(W_j) - \tau(W_j)\right).$$

Since $(1/n)\sum_{j=1}^n(\hat{\tau}_n(W_j) - \tau(W_j))^2 = \mathscr{O}_{\xi_n}(n^{-2/3})$, we have

$$\frac{1}{n}\sum_{j=1}^{n}\left(\hat{h}_{n,j} - h\left(Z_j, \hat{\beta}_n\right)\right)^2 = \mathscr{O}_{\xi_n}(n^{-2/3}), \qquad \hat{h}_{n,\star} = \mathscr{O}_{\xi_n}(n^{-1/3})$$

and

$$\begin{aligned}&\frac{1}{n}\sum_{j=1}^{n}\left(\hat{h}_{n,j} - \hat{h}_{n,\star}\right)\left(1 - U_j + \hat{\beta}_n U_j\right)\left(\hat{\tau}_n(W_j) - \tau(W_j)\right)\\ &\quad = \frac{1}{n}\sum_{j=1}^{n} h\left(Z_j, \hat{\beta}_n\right)\left(1 - U_j + \hat{\beta}_n U_j\right)\left(\hat{\tau}_n(W_j) - \tau(W_j)\right)\\ &\qquad + o_{\xi_n}(n^{-1/2}).\end{aligned} \tag{9.3}$$

We need the following result. If ϕ is a bounded measurable function from [0, 1] to $\mathbb{R}$, then

$$\begin{aligned}&\frac{1}{n}\sum_{j=1}^{n} U_j\phi(W_j)\left(\hat{\tau}_n(W_j) - \tau(W_j)\right)\\ &\quad = \hat{\pi}_n\int\phi(w)\left(\hat{\tau}_n(w) - \tau(w)\right)dR(w) + o_{\xi_n}(n^{-1/2})\end{aligned}$$

and

$$\begin{aligned}&\frac{1}{n}\sum_{j=1}^{n}(1 - U_j)\phi(W_j)\left(\hat{\tau}_n(W_j) - \tau(W_j)\right)\\ &\quad = (1 - \hat{\pi}_n)\int\phi(w)\left(\hat{\tau}_n(w) - \tau(w)\right)dR(w) + o_{\xi_n}(n^{-1/2}).\end{aligned}$$

This result can be proved using Lemma 3.4.A in Schick (1989). From this we immediately obtain that

$$\begin{aligned}&\frac{1}{n}\sum_{j=1}^{n} h\left(Z_j, \hat{\beta}_n\right)\left(1 - U_j + \hat{\beta}_n U_j\right)\left(\hat{\tau}_n(W_j) - \tau(W_j)\right)\\ &\quad = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{\beta}_n}{\hat{\pi}_n}U_j - \frac{\hat{\beta}_n}{1 - \hat{\pi}_n}(1 - U_j)\right)\tau_0(W_j)\left(\hat{\tau}_n(W_j) - \tau(W_j)\right)\\ &\quad = o_{\xi_n}(n^{-1/2}),\end{aligned} \tag{9.4}$$

$$\begin{aligned}&\frac{1}{n}\sum_{j=1}^{n}\hat{\zeta}_{2,n}(U_j - \hat{\pi}_n)\left(1 - U_j + \hat{\beta}_n U_j\right)\left(\hat{\tau}_n(W_j) - \tau(W_j)\right)\\ &\quad = \hat{\zeta}_{2,n}(1 - \hat{\pi}_n)\hat{\pi}_n\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{\beta}_n}{\hat{\pi}_n}U_j - \frac{1}{1 - \hat{\pi}_n}(1 - U_j)\right)\\ &\qquad\qquad \times\left(\hat{\tau}_n(W_j) - \tau(W_j)\right)\\ &\quad = \bar{\tau}_n\left(1 - \hat{\pi}_n + \hat{\beta}_n\hat{\pi}_n\right)\left(\hat{\beta}_n - 1\right)\int\hat{\tau}_n - \tau\, dR + o_{\xi_n}(n^{-1/2}),\end{aligned} \tag{9.5}$$

and

$$(9.6)\qquad \begin{aligned} &\frac{1}{n}\sum_{j=1}^{n}(1-\hat{\beta}_n)\bar{\tau}_n\big(1-U_j+\hat{\beta}_n U_j\big)\big(\hat{\tau}_n(W_j)-\tau(W_j)\big)\\ &\quad = \big(1-\hat{\beta}_n\big)\bar{\tau}_n\big(1-\hat{\pi}_n+\hat{\beta}_n\hat{\pi}_n\big)\int\hat{\tau}_n-\tau\,dR+o_{\xi_n}\big(n^{-1/2}\big). \end{aligned}$$

Combining (9.3) to (9.6) yields Condition T. This proves the result in the case that $\{\hat{\beta}_n\}$ is a local sequence. The general case follows as $\{\hat{\beta}_n\}$ is discrete and $\sqrt{n}$-consistent. □

9.2 Remark. A $\sqrt{n}$-consistent estimator is given by

$$\hat{\beta}_n=\frac{\Sigma_{j=1}^n(1-U_j)\Sigma_{j=1}^n U_jY_j}{\Sigma_{j=1}^n U_j\Sigma_{j=1}^n(1-U_j)Y_j}.$$

Indeed, one verifies easily that

$$\mathfrak{L}\Big(\sqrt{n}\,\big(\hat{\beta}_n-\beta\big)\,|\,\mathbb{P}_{(\beta,\tau)}\Big)\quad\Rightarrow\quad N\left(0,\frac{\sigma^2(1-\pi+\beta^2\pi)}{\pi(1-\pi)\bar{\tau}^2}\right).$$

See Härdle and Marron (1990) for other choices of $\sqrt{n}$-consistent estimates.

10. Proofs of the propositions. Let $r_n=\varrho(\cdot,\xi_n)$ and set

$$\hat{\mathbf{w}}_n=\begin{pmatrix}\hat{w}_{n,1}\\ \vdots\\ \hat{w}_{n,n}\end{pmatrix},\qquad \boldsymbol{\epsilon}_n=\begin{pmatrix}\varepsilon_{n,1}\\ \vdots\\ \varepsilon_{n,n}\end{pmatrix},$$

$$\hat{\boldsymbol{\epsilon}}_n=\begin{pmatrix}\hat{\varepsilon}_{n,1}\\ \vdots\\ \hat{\varepsilon}_{n,n}\end{pmatrix}\quad\text{and}\quad \hat{\boldsymbol{\epsilon}}_{n,j}=\mathbb{E}_{n,j}(\hat{\boldsymbol{\epsilon}}_n)=\begin{pmatrix}\hat{\varepsilon}_{n,1,j}\\ \vdots\\ \hat{\varepsilon}_{n,n,j}\end{pmatrix}.$$

For each $n\in\mathbb{N}$, let ℓ_n denote the map from $\mathbb{R}\times\mathbb{R}\times\{0,1\}^n$ to $\mathbb{R}$ defined by

$$\ell_n(x,y,w)=-\frac{(1/a_n^2)\Sigma_{j=1}^n w_jk'\big((x-y_j)/a_n\big)}{b_n\Sigma_{j=1}^n w_j+(1/a_n)\Sigma_{j=1}^n w_jk\big((x-y_j)/a_n\big)},$$

$$x\in\mathbb{R},\ y\in\mathbb{R}^n,\ w\in\{0,1\}^n,$$

so that

$$\hat{\ell}_n(x)=\ell_n\big(x,\hat{\boldsymbol{\epsilon}}_n,\hat{\mathbf{w}}_n\big),\qquad x\in\mathbb{R}.$$

Let $\ell_n^{(i)}$ denote the ith-order partial derivative of ℓ_n with respect to its first argument, that is,

$$\ell_n^{(i)}(x,y,w) = \frac{\partial^i}{\partial x^i}\ell_n(x,y,w), \qquad x \in \mathbb{R},\ y \in \mathbb{R}^n,\ w \in \{0,1\}^n.$$

We shall need the following properties of the maps ℓ_n.

10.1 LEMMA. *Suppose Condition K holds. Then there is a positive constant c_0 such that the following inequalities hold for all $x \in \mathbb{R}$, $y, \tilde{y} \in \mathbb{R}^n$, $w \in \{0,1\}^n$ and $i = 0,1,2$:*

$$\text{(L1)} \qquad \left|\ell_n^{(i)}(x,y,w)\right| \le \frac{c_0}{a_n^{i+1}},$$

$$\text{(L2)} \quad \left|\ell_n^{(i)}(x,y,w) - \ell_n^{(i)}(x,\tilde{y},w)\right| \le \frac{c_0}{\sum_{j=1}^n w_j a_n^{3+i} b_n} \sum_{j=1}^n w_j\big(a_n \wedge |\tilde{y}_j - y_j|\big),$$

$$\text{(L3)} \quad \left|\ell_n^{(i)}(x,y,w) - \ell_n^{(i)}(x,\tilde{y},w)\right|^2 \le \frac{c_0}{\sum_{j=1}^n w_j a_n^{5+2i} b_n} \sum_{j=1}^n w_j(\tilde{y}_j - y_j)^2.$$

PROOF. Straightforward. □

10.2 LEMMA. *Let Condition K hold. Let $f_n(x) = \int f(x - a_n t)k(t)\,dt$, $x \in \mathbb{R}$. Then*

$$\Sigma_{n,1} = \mathbb{E}_n\left(\int\left|\ell_n(x,\boldsymbol{\epsilon}_n,\hat{\mathbf{w}}_n) + \frac{f_n'(x)}{f_n(x)+b_n}\right|^2 dF(x)\right) = \mathscr{O}_{\xi_n}(n^{-1}a_n^{-4}b_n^{-2})$$

and

$$\Sigma_{n,2} = \int\left(\frac{f_n'(t)}{b_n + f_n(t)} - \frac{f'(t)}{f(t)}\right)^2 f(t)\,dt \to 0.$$

PROOF. See Schick (1987).

PROOF OF PROPOSITION 5.5. It follows from Condition S that

$$\max_{1\le j\le n}\|\hat{w}_{n,j}\hat{s}_{n,j}\| = \mathscr{O}_{\xi_n}(A_n).$$

As F has finite Fisher information for location,

$$\int|f(x+t) - f(x)|\,dx \le |t|\sqrt{J}.$$

Using this and (L1), one finds

$$\left|\int\left(\hat{\ell}_n(x-t)-\hat{\ell}_n(x)+t\hat{\ell}_n'(x)\right)f(x)\,dx\right|$$

$$\leq |t|\int_0^1\left|\int\left(\hat{\ell}_n'(x-\lambda t)-\hat{\ell}_n'(x)\right)f(x)\,dx\right|d\lambda$$

$$\leq |t|\int_0^1\int\left|\hat{\ell}_n'(x)(f(x+\lambda t)-f(x))\right|dx\,d\lambda \leq c_0 a_n^{-2}t^2\sqrt{J}$$

and

$$\left|\int\left(\hat{\ell}_n'(x-t)-\hat{\ell}_n'(x)\right)f(x)\,dx\right|$$

$$\leq \int\left|\hat{\ell}_n'(x)\right|\left|f(x+t)-f(x)\right|dx \leq c_0 a_n^{-2}|t|\sqrt{J}.$$

The desired results are now immediate. □

The proofs of Propositions 5.6, 5.7 and 5.8 rely on the following modification of Lemma 3.4.A in Schick (1989).

10.3 LEMMA. *For each pair (n, j) of positive integers, $1 \leq j \leq n$, let $h_{n,j}$ be a measurable function from $\mathbb{R}\times\mathbb{R}^n\times\mathscr{K}^n$ to $\mathbb{R}$, let $H_{n,j}$ be a measurable function from $\mathbb{R}\times\mathbb{R}^{n-1}\times\mathscr{K}^n$ to $\mathbb{R}$, let $\hat{h}_{n,j}=h_{n,j}(\cdot,\mathbf{Y}_n,\mathbf{Z}_n)$ and let $\tilde{h}_{n,j}=H_{n,j}(\cdot,\mathbf{Y}_{n,j},\mathbf{Z}_n)$. Suppose that for some $\delta\geq 0$,*

$$C_{n,1}=\frac{1}{n}\sum_{j=1}^n \hat{h}_{n,j}(Y_j)-\tilde{h}_{n,j}(Y_j)=o_{\xi_n}(n^{-\delta}), \tag{10.1}$$

$$\begin{aligned} C_{n,2}&=\frac{1}{n}\sum_{j=1}^n\int\hat{h}_{n,j}\big(y+r_n(Z_j)\big)-\tilde{h}_{n,j}\big(y+r_n(Z_j)\big)\,dF(y)\\ &=o_{\xi_n}(n^{-\delta}),\end{aligned} \tag{10.2}$$

$$C_{n,3}=\frac{1}{n^2}\sum_{j=1}^n\mathbb{E}_n\left(\int\left|\hat{h}_{n,j}\big(y+r_n(Z_j)\big)\right|^2dF(y)\right)=o_{\xi_n}(n^{-2\delta}), \tag{10.3}$$

$$C_{n,4}=\frac{1}{n^2}\sum\sum_{i\neq j}\mathbb{E}_n\left(\left|\tilde{h}_{n,j}(Y_j)-\mathbb{E}_{n,i}\big(\tilde{h}_{n,j}(Y_j)\big)\right|^2\right)=o_{\xi_n}(n^{-2\delta}). \tag{10.4}$$

Then

$$\frac{1}{n}\sum_{j=1}^{n}\hat{h}_{n,j}(Y_j) - \int \hat{h}_{n,j}\big(y + r_n(Z_j)\big)\,dF(y) = o_{\xi_n}(n^{-\delta}).$$

PROOF. Let $U_{n,j} = \tilde{h}_{n,j}(Y_j) - \mathbb{E}_{n,j}(\tilde{h}_{n,j}(Y_j))$ and $U_{n,j,i} = \mathbb{E}_{n,i}(U_{n,j})$. In view of (10.1), (10.2) and the identity $\mathbb{E}_{n,j}(\tilde{h}_{n,j}(Y_j)) = \int \tilde{h}_{n,j}(y + r_n(Z_j))\,dF(y)$, it suffices to verify $(1/n)\sum_{j=1}^{n} U_{n,j} = o_{\xi_n}(n^{-\delta})$. As in Schick (1989) one verifies

$$\begin{aligned}
\mathbb{E}_n&\left(\left(\frac{1}{n}\sum_{j=1}^{n} U_{n,j}\right)^2\right)\\
&= \frac{1}{n^2}\left(\sum_{j=1}^{n}\mathbb{E}_n\big(U_{n,j}^2\big) + \sum\sum_{i\neq j}\mathbb{E}_n\big(U_{n,j}U_{n,i}\big)\right)\\
&= \frac{1}{n^2}\left(\sum_{j=1}^{n}\mathbb{E}_n\big(U_{n,j}^2\big) + \sum\sum_{i\neq j}\mathbb{E}_n\Big(\big(U_{n,j} - U_{n,j,i}\big)\big(U_{n,i} - U_{n,i,j}\big)\Big)\right)\\
&\le \frac{1}{n^2}\left(\sum_{j=1}^{n}\mathbb{E}_n\big(U_{n,j}^2\big) + \sum\sum_{i\neq j}\mathbb{E}_n\Big(\big(U_{n,j} - U_{n,j,i}\big)^2\Big)\right) = o_{\xi_n}(n^{-2\delta}).
\end{aligned}$$

This establishes the desired result. □

The following statements are easy consequences of Condition R.

10.4 LEMMA. *Suppose Condition R holds. Then*

$$U_{n,1} = \frac{1}{N_n^2}\sum\sum_{i\neq j}\hat{w}_{n,j}\mathbb{E}_n\Big(|\varepsilon_{n,j} - \hat{\varepsilon}_{n,j,i}|^2\Big) = \mathscr{O}_{\xi_n}(n^{-2\alpha}),$$

$$U_{n,2} = \frac{1}{N_n}\sum\sum_{i\neq j}\hat{w}_{n,j}\mathbb{E}_n\Big(|\hat{\varepsilon}_{n,j} - \hat{\varepsilon}_{n,j,i}|^2\Big) = \mathscr{O}_{\xi_n}(n^{-2\alpha}),$$

$$U_{n,3} = \frac{1}{N_n^2}\sum\sum\sum_{i\neq j,\, i\neq a,\, j\neq a}\hat{w}_{n,a}\mathbb{E}_n\Big(\big|\hat{\varepsilon}_{n,a,j} - \mathbb{E}_{n,i}\big(\hat{\varepsilon}_{n,a,j}\big)\big|^2\Big) = \mathscr{O}_{\xi_n}(n^{-2\alpha}).$$

PROOF. Note that $U_{n,1} \le (1/N_n)\sum_{j=1}^{n}\hat{w}_{n,j}\mathbb{E}_n(|\varepsilon_{n,j} - \hat{\varepsilon}_{n,j}|^2) \le R_{n,1}$ and $U_{n,3} \le U_{n,2} = R_{n,3}$. □

Let

$$\hat{\ell}_{n,j}^{(\mu)}(y) = \ell_n^{(\mu)}\big(y, \hat{\boldsymbol{\epsilon}}_{n,j}, \hat{\mathbf{w}}_n\big) \quad\text{and}\quad \hat{\ell}_{n,j,i}^{(\mu)}(y) = \ell_n^{(\mu)}\big(y, \mathbb{E}_{n,i}(\hat{\boldsymbol{\epsilon}}_{n,j}), \hat{\mathbf{w}}_n\big).$$

10.5 LEMMA. *Suppose Conditions K, W and R hold and $na_n^4 b_n^2 \to \infty$. Then for $\mu = 0, 1$,*

$$T_{n,1}^{(\mu)} = \frac{1}{N_n} \sum_{j=1}^{n} \hat{w}_{n,j} \mathbb{E}_n \left(\sup_{x \in \mathbb{R}} \left| \hat{\ell}_n^{(\mu)}(x - \hat{r}_{n,j}) - \hat{\ell}_{n,j}^{(\mu)}(x - \hat{r}_{n,j,j}) \right|^2 \right)$$

$$= N_n^{-1} a_n^{-2\mu} \mathscr{O}_{\xi_n}\left(a_n^{-4} n^{-\alpha_*} + N_n^{-1} a_n^{-4} b_n^{-2} + a_n^{-5} b_n^{-1} n^{-2\alpha} \right),$$

$$T_{n,2} = \frac{1}{N_n^2} \sum_{j=1}^{n} \hat{w}_{n,j} \mathbb{E}_n \left(\sup_{x \in \mathbb{R}} \left| \hat{\ell}_{n,j}(x + r_n(Z_j) - \hat{r}_{n,j,j}) - \ell_n(x, \epsilon_n, \hat{\mathbf{w}}_n) \right|^2 \right)$$

$$= N_n^{-1} \mathscr{O}_{\xi_n}\left(N_n^{-2} a_n^{-4} b_n^{-2} + a_n^{-5} b_n^{-1} n^{-2\alpha} \right),$$

$$T_{n,3}^{(\mu)} = \frac{1}{N_n^2} \sum\sum_{i \neq j} \hat{w}_{n,j} \mathbb{E}_n \left(\left| \hat{\ell}_{n,j}^{(\mu)}(Y_j - \hat{r}_{n,j,j}) - \hat{\ell}_{n,j,i}^{(\mu)}(Y_j - \mathbb{E}_{n,i}(\hat{r}_{n,j,j})) \right|^2 \right)$$

$$= N_n^{-1} a_n^{-2\mu} \mathscr{O}_{\xi_n}\left(N_n^{-1} a_n^{-4} b_n^{-2} + a_n^{-5} b_n^{-1} n^{-2\alpha} \right),$$

$$T_{n,4} = \mathbb{E}_n \left(\int \left| \hat{\ell}_n(y) - \ell(y) \right|^2 dF(y) \right) = \mathscr{O}_{\xi_n}\left(a_n^{-5} b_n^{-1} n^{-2\alpha} \right) + o_{\xi_n}(1).$$

PROOF. Using (L1) to (L3), one obtains for some positive constant C:

$$T_{n,1}^{(\mu)} \le C a_n^{-2\mu} \left(N_n^{-1} a_n^{-4} R_{n,2} + N_n^{-2} a_n^{-4} b_n^{-2} + N_n^{-1} a_n^{-5} b_n^{-1} U_{n,2} \right),$$

$$T_{n,2} \le C \left(N_n^{-1} a_n^{-4} R_{n,1} + N_n^{-3} a_n^{-4} b_n^{-2} + N_n^{-1} a_n^{-5} b_n^{-1} U_{n,1} \right),$$

$$T_{n,3}^{(\mu)} \le C a_n^{-2\mu} \left(N_n^{-1} a_n^{-4} R_{n,3} + N_n^{-2} a_n^{-4} b_n^{-2} + N_n^{-1} a_n^{-5} b_n^{-1} U_{n,3} \right),$$

$$T_{n,4} \le C \left(a_n^{-5} b_n^{-1} R_{n,1} + \Sigma_{n,1} + \Sigma_{n,2} \right).$$

The desired results follow from these bounds, Lemma 10.3 and Condition R. □

PROOF OF PROPOSITION 5.6. We shall apply Lemma 10.3 with $\hat{h}_{n,j}(y) = \hat{w}_{n,j}(\hat{s}_{n,j} \hat{\ell}_n(y - \hat{r}_{n,j}) - s_n(Z_j) \ell(y - r_n(Z_j)))$ and $\tilde{h}_{n,j}(y) = \hat{w}_{n,j}(\tilde{s}_{n,j} \hat{\ell}_{n,j}(y - \hat{r}_{n,j,j}) - s_n(Z_j) \ell(y - r_n(Z_j)))$. Thus we have to verify (10.1) to (10.4) for these choices. We have (10.1) in view of the bound

$$\left(\mathbb{E}_n(|C_{n,1}|) \right)^2 \le \left(\frac{1}{N_n} \sum_{j=1}^{n} \hat{w}_{n,j} \mathbb{E}_n \left(\left| \hat{s}_{n,j} \hat{\ell}_n(Y_j - \hat{r}_{n,j}) - \tilde{s}_{n,j} \hat{\ell}_{n,j}(Y_j - \hat{r}_{n,j,j}) \right| \right) \right)^2$$

$$\le 2 S_{n,2} \frac{1}{N_n} \sum_{j=1}^{n} \hat{w}_{n,j} \mathbb{E}_n \left(\hat{\ell}_n^2(Y_j - \hat{r}_{n,j}) \right) + 2 T_{n,1}^{(0)} \frac{1}{N_n} \sum_{j=1}^{n} \hat{w}_{n,j} \mathbb{E}_n \left(\tilde{s}_{n,j}^2 \right)$$

$$= N_n^{-1} \left(o_{\xi_n}(1) + \mathscr{O}_{\xi_n}\left(a_n^{-4} n^{-\alpha_*} + N_n^{-1} a_n^{-4} b_n^{-2} + a_n^{-5} b_n^{-1} n^{-2\alpha} \right) \right)$$

$$= o_{\xi_n}(n^{-1}).$$

The proof of (10.2) is similar. Equation (10.3) follows from Lemma 10.2, Lemma 10.5 and the bound

$$C_{n,3} \le N_n^{-2} \sum_{j=1}^{n} \hat{w}_{n,j} \mathbb{E}_n\Big(\Big| \tilde{s}_{n,j} \hat{\ell}_{n,j}\big(y - (\hat{r}_{n,j,j} - r_n(Z_j))\big) - s_n(Z_j)\ell(y) \Big|^2 dF(y) \Big)$$

$$\le 4\big(A_n^2(T_{n,2} + N_n^{-1}\Sigma_{n,1}) + N_n^{-1} S_{n,1} \int \Big(\frac{f_n'(y)}{b_n + f_n(y)} \Big)^2 dF(y)$$

$$+ \frac{1}{N_n} \sum_{j=1}^{n} \hat{w}_{n,j} s_n^2(Z_j) N_n^{-1} \Sigma_{n,2}$$

$$= N_n^{-1}\big(A_n^2 \mathscr{O}_{\xi_n}(a_n^{-5} b_n^{-1} n^{-2\alpha} + N_n^{-2} a_n^{-4} b_n^{-2} + N_n^{-1} a_n^{-4} b_n^{-2}) + o_{\xi_n}(1) \big)$$

$$= o_{\xi_n}(n^{-1}).$$

Finally, (10.4) follows from the bound

$$C_{n,4} \le \frac{1}{N_n^2} \sum\sum_{i \ne j} \hat{w}_{n,j} \mathbb{E}_n\Big(\Big| \tilde{s}_{n,j} \hat{\ell}_{n,j}(Y_j - \hat{r}_{n,j,j}) - \tilde{s}_{n,j,i} \hat{\ell}_{n,j,i}\big(Y_j - \mathbb{E}_{n,i}(\hat{r}_{n,j,j})\big) \Big|^2 \Big)$$

$$\le 2N_n^{-1}\big(S_{n,3} a_n^{-2} + A_n^2 T_{n,3}^{(0)}\big)$$

$$= N_n^{-1}\big(o_{\xi_n}(1) + \mathscr{O}_{\xi_n}\big(A_n^2(a_n^{-5} b_n^{-1} n^{-2\alpha} + N_n^{-1} a_n^{-4} b_n^{-2})\big)\big) = o_{\xi_n}(n^{-1}).$$

Thus Lemma 10.3 applies and gives the desired result. □

PROOF OF PROPOSITION 5.7. We shall apply Lemma 10.3 with $\hat{h}_{n,j}(y) = \hat{w}_{n,j} \hat{\ell}_n'(y - \hat{r}_{n,j})$ and $\tilde{h}_{n,j}(y) = \hat{w}_{n,j} \hat{\ell}_{n,j}'(y - \hat{r}_{n,j,j})$. Equations (10.1)–(10.4) follow from Lemma 10.5 and the bounds $C_{n,1}^2 + C_{n,2}^2 \le 2T_{n,1}^{(1)}$, $C_{n,3} \le n^{-1} a_n^{-4}$ and $C_{n,4} \le T_{n,3}^{(1)}$. Thus Lemma 10.3 applies and gives the desired result. □

PROOF OF PROPOSITION 5.8. The proof of the first part is similar to the proof of Proposition 5.6, and the second part follows as in the proof of Proposition 5.5. We omit the details. □

REFERENCES

BEGUN, J., HALL, W., HUANG, W.-M. 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.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Univ. Press.

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

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

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.

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.

GREEN, P., JENNISON, C. and SEHEULT, A. (1985). Analysis of field experiments by least squares smoothing. *J. Roy. Statist. Soc. Ser. B* **47** 299–315.

HÄRDLE, W. and MARRON, J. S. (1990). Semiparametric comparison of regression curves. *Ann. Statist.* **18** 63–89.

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academic, New York.

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

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

PFANZAGL, J. and WEFELMEYER, W. (1982). *Contributions to a General Asymptotic Statistical Theory. Lecture Notes in Statist.* **13**. Springer, Berlin.

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

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

SCHICK, A. (1986). On asymptotically efficient estimation in semi-parametric 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. (1989). Correction to "A note on the construction of asymptotically linear estimators." *J. Statist. Plann. Inference* **22** 269–270.

STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **3** 267–284.

WAHBA, G. (1984). Cross validated spline methods for the estimation of multivariate functions from data on functionals. In *Statistics: An Appraisal, Proc. 50th Anniversary Conf. Iowa State Statistical Laboratory* (H. A. David and H. T. David, eds.) 205–235. Iowa State Univ. Press, Ames.

DEPARTMENT OF MATHEMATICAL SCIENCES
SUNY BINGHAMTON
P.O. BOX 6000
BINGHAMTON, NEW YORK 13902-6000