

Delecroix, Michel; Härdle, Wolfgang; Hristache, Marian

Working Paper

Efficient estimation in single-index regression

SFB 373 Discussion Paper, No. 1997,37

Provided in Cooperation with:

Collaborative Research Center 373: Quantification and Simulation of Economic Processes,
Humboldt University Berlin

Suggested Citation: Delecroix, Michel; Härdle, Wolfgang; Hristache, Marian (1997) : Efficient estimation in single-index regression, SFB 373 Discussion Paper, No. 1997,37, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin,
<https://nbn-resolving.de/urn:nbn:de:kobv:11-10064199>

This Version is available at:

<https://hdl.handle.net/10419/66285>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Efficient Estimation in Single-Index Regression

Michel DELECROIX

ENSAI and CREST
Rue Blaise Pascal
Campus de Ker Lann
35170 Bruz, FRANCE

Wolfgang HÄRDLE

Institut für Statistik und Ökonometrie
Humboldt-Universität zu Berlin
Spandauer Str. 1
D-10178 Berlin, GERMANY

Marian HRISTACHE

ENSAI and CREST
Rue Blaise Pascal
Campus de Ker Lann
35170 Bruz, FRANCE

January 15, 1997

Abstract

Semiparametric single-index regression involves an unknown finite dimensional parameter and an unknown (link) function. We consider estimation of the parameter via the pseudo maximum likelihood method. For this purpose we estimate the conditional density of the response given a candidate index and maximize the obtained likelihood. We show that this technique of adaptation yields an asymptotically efficient estimator : it has minimal variance among all estimators.

⁰Acknowledgements : The second author was supported by Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonomischer Prozesse" Deutsche Forschungsgemeinschaft. This work was substantially completed when the first and third author were visiting at Sonderforschungsbereich 373. The authors thank V. Spokoiny and J. Horowitz for helpful discussions and comments.

1 Introduction

A single-index response model has the form

$$E(Y | X = x) = E(Y | X\theta = x\theta) \quad (1.1)$$

where Y is a scalar dependent variable, X is a d -dimensional vector of explanatory variables, $x\theta$ is the index, the scalar product of x with θ , a vector of parameters whose values are unknown. Many widely used parametric models have this form. Examples are linear regression, binary logit and probit, and tobit models.

These models assume in (1.1) a "link" between the index $x\theta$ and the response. In the linear regression, for example, this link is the identity. In the logit model case it is the conditional distribution function of a logistic distribution. In this paper we consider estimation of the parameter θ in (1.1) without supposing further restrictions on this link. Moreover, we derive the asymptotic normal distribution of this estimator and show that it is efficient in the sense of achieving minimal variance among all estimators for θ .

Adaptation in the linear regression model has been considered by Carroll (1982), Robinson (1987). They consider the case of unknown heteroskedasticity of the error variables. They show that a nonparametric estimate of the unknown heteroskedasticity function gives full adaptation in the sense of yielding the same variance as the Aitken-estimator with known heteroskedasticity function. The model (1.1) is more general since it allows arbitrary relations between the index $X\theta$ and Y .

Several estimators of θ that do not require a fully parametric specification of (1.1) already exist. Ichimura (1993) developed a semiparametric least squares estimator of θ . This estimator is closely related to projection pursuit regression (Friedman and Stuetzle (1981)) since it minimizes a least squares criterion based on nonparametric estimation of the link. Han (1987) and Sherman (1993) describe a maximum rank correlation estimator. Klein and Spady (1993) developed a quasi-maximum likelihood estimator for the case in which Y is a binary response. This estimator achieves the asymptotic efficiency bound of Cosslett (1987) if the link is a conditional distribution function. Horowitz and Härdle (1995) considered fast non-iterative methods for single-index models in the case of discrete covariates. The estimators of Ichimura, Han, Klein and Spady, Sherman, and Horowitz and Härdle are $n^{1/2}$ -consistent, and asymptotically normal under regularity conditions.

The foregoing estimators were designed for specific data situations like *e.g.* discrete covariates or binary response or computational efficiency. The focus on such a particular aspect make them not necessarily efficient. The variance of the direct (computationally efficient) estimator of Horowitz and Härdle (1995) for example is not the best possible one computed in Klein and Spady (1993). The object of this paper is to construct an asymptotically efficient estimator for general single-index response models. It deviates from the ideas of projection pursuit since we use a pseudo maximum likelihood criterion.

Our method will be based on nonparametric estimation of the semiparametric conditional density $f_\theta(y, x\theta)$ of the distribution $\mathcal{L}(Y|X\theta)$. We do not assume a specific structure, like for example a binary response as in Klein and Spady (1993), for this conditional density. Our approach thus covers efficient estimation in linear regression (Bickel (1982)) with unknown error distribution as well as nonlinear response models with single-index structure (see Huet, Jolivet, Messéan (1989)).

Suppose we are given i.i.d. observations $Z_i = (X_i, Y_i) \in \mathbb{R}^d \times \mathbb{R}$, with

$$E(Y_i | X_i = x) = E(Y_i | X_i \theta_0 = x \theta_0), \quad i = 1, \dots, n, \quad (1.2)$$

where $\theta_0 \in \Theta \subset \mathbb{R}^d$ is the true value of the parameter in the model. Assume that for all $x \in \text{supp}X$ the conditional density $f(y|x)$ of Y given $X = x$ with respect to a σ -finite measure exists. This density is supposed to depend upon x through $x\theta_0$.

Thus a positive function f defined on $\text{supp}Y \times \mathcal{M}$, ($\mathcal{M} \subset \mathbb{R}$), is given satisfying :

$$f(y|x) = f(y, x\theta_0) \quad ((x, y) \in \text{supp}Z). \quad (1.3)$$

The main idea of our estimator is to estimate the function f in (1.3) and then to optimize the (estimated) pseudo-likelihood over the parameter vector θ . The technique is called pseudo maximum likelihood estimation (PMLE). We use the kernel estimation method here since it is easy to compute in practice and auxiliary asymptotic results are available in the literature. In order to present our estimator we need some more notation.

Let S be a fixed subset of the support of $Z = (X, Y)$ and let $S_X = \{x : \exists y \text{ s.t. } (x, y) \in S\}$. We assume that for all x in S_X and all θ in Θ , one can define the conditional density $f_\theta(y, x\theta)$ of Y given $X\theta = x\theta$ and $Z \in S$. We will then define n estimators \hat{f}_θ^i of f_θ at the point $(y, x\theta)$, for (x, y) in the fixed subset S , by

$$\hat{f}_\theta^i(y, x\theta) = N_{i,n}(y, x\theta) / D_{i,n}(x\theta), \quad (1.4)$$

with

$$\begin{aligned} N_{i,n}(y, t) &= \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^n K_{h_n}(y - Y_j) \cdot K_{h_n}(t - X_j\theta) \cdot I_{\{Z_j \in S\}}, \\ D_{i,n}(t) &= \frac{1}{(n-1)} \sum_{\substack{j=1 \\ j \neq i}}^n K_{h_n}(t - X_j\theta) \cdot I_{\{Z_j \in S\}}, \end{aligned} \quad (1.5)$$

where h_n is the bandwidth, K is a fixed kernel density, $K_h(\cdot) = K(\cdot/h)/h$.

We define $\hat{\theta}_n$ to be the solution of

$$\hat{L}_n(\hat{\theta}_n) = \max_{\theta \in \Theta} \hat{L}_n(\theta) \quad (1.6)$$

with

$$\hat{L}_n(\theta) = \frac{1}{n} \sum_{i=1}^n \log \hat{f}_\theta^i(Y_i, X_i\theta) \cdot I_{\{Z_i \in S\}}. \quad (1.7)$$

Let L_n be the log-likelihood function defined by

$$L_n(\theta) = \frac{1}{n} \sum_{i=1}^n \log f_\theta(Y_i, X_i\theta) \cdot I_{\{Z_i \in S\}}. \quad (1.8)$$

Define also

$$L(\theta) = E\{\log f_\theta(Y_i, X_i\theta)\} \cdot I_{\{Z_i \in S\}}. \quad (1.9)$$

The idea is to maximize the proxy $\hat{L}_n(\theta)$ for $L_n(\theta)$ which itself is a proxy for $L(\theta)$.

2 Consistency of the semiparametric estimator

First, we show that the estimate $\hat{\theta}_n$ defined in (1.6) converges almost surely to θ_0 as n tends to ∞ . We shall prove the following :

$$\sup_{\theta \in \Theta} \left| \widehat{L}_n(\theta) - L_n(\theta) \right| \xrightarrow[n \rightarrow \infty]{a.s.} 0. \quad (2.1)$$

and

$$\sup_{\theta \in \Theta} |L_n(\theta) - L(\theta)| \xrightarrow[n \rightarrow \infty]{a.s.} 0. \quad (2.2)$$

Then, if $L(\theta)$ has a unique maximum at θ_0 , the *PMLE* estimate $\hat{\theta}_n$ converges almost surely towards this maximum. The precise assumptions are as follows :

(A1) (X_i, Y_i) are i.i.d. random vectors.

(A2) Θ is a compact subset of \mathbb{R}^d .

(A3) The random vectors $(Y_i, X_i\theta)$ have a continuous distribution.

(A4) The compact subset S of the support of $Z_i = (Y_i, X_i)$ is such that:

- i) for all θ in Θ , the conditional density h_θ of $X_i\theta$ given that $Z_i \in S$ verifies

$$\inf_{(x,\theta) \in S_X \times \Theta} h_\theta(x\theta) > 0;$$
- ii) $\inf_{(z,\theta) \in S \times \Theta} f_\theta(y, x\theta) > 0$, where $z = (x, y)$.

(A5) i) h_θ and f_θ are uniformly continuous;

ii) $h_\theta(t)$ and $f_\theta(y, t)$ are three times differentiable with respect to t and the third derivatives satisfy Lipschitz conditions for $t \in T_\theta(S)$, uniformly in $\theta \in \Theta$ and $y \in S_Y$;

iii) $R(x, \theta) \stackrel{def.}{=} f_\theta(y, x\theta)$ and $D(x, \theta) \stackrel{def.}{=} h_\theta(x\theta)$ are twice continuously differentiable with respect to θ on $S \times \Theta$.

(A6) There exists a unique $\theta_0 \in \Theta$ such that relation (1.3) holds.

(A7) i) For all $\theta, \theta' \in \Theta$ and $x \in S_X$, the distributions P_θ and $P_{\theta'}$ defined by the densities $f_\theta(\cdot, x\theta)$ and $f_{\theta'}(\cdot, x\theta')$ are equivalent.

ii) There exists a subset $A \subset S_X$ of positive Lebesgue measure such that X_i is continuous on A .

(A8) The matrix $M = E \left[-\frac{\partial^2}{\partial \theta \partial \theta^T} \log f_\theta(Y_i, X_i\theta) \Big|_{\theta=\theta_0} I_{\{Z_i \in S\}} \right]$ is positive-definite.

(C(δ)) K is a real symmetric fourth order kernel and $h_n = c n^{-\delta}$, with $c, \delta > 0$.

The following preliminary results are shown in the appendix.

Lemma 2.1 Under the assumptions (A1)–(A5) and (C(δ)) with $\delta \in \left(0, \frac{1}{3}\right)$, we have

$$\sup_{z \in S} \sup_{\theta \in \Theta} \left| \log \widehat{f}_\theta(y, x\theta) - \log f_\theta(y, x\theta) \right| \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

Lemma 2.2 Under the assumptions (A1) through (A5) and (C(δ)) with $\delta \in \left(0, \frac{1}{3}\right)$, we have :

$$\sup_{\theta \in \Theta} \left| \widehat{L}_n(\theta) - L_n(\theta) \right| + |L_n(\theta) - L(\theta)| \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

From Lemma 2.2 it follows that $\sup_{\theta \in \Theta} \left| \widehat{L}_n(\theta) - L(\theta) \right| \xrightarrow[n \rightarrow \infty]{a.s.} 0$.

Remark : Inspection of the proofs of Lemma 2.1 and 2.2 show that we do not need (A5) in its full strength. Lipschitz continuity is sufficient. For better exposition we use this stronger smoothing throughout.

Lemma 2.3 Under assumptions (A1), (A3) and (A5)–(A7) the function $L(\theta)$ has a unique maximum at θ_0 .

The proof relies on the properties of Kullback information and can be found in Bonneau, Delecroix and Hristache (1995). Application of Gouriéroux, Monfort (1989, page 431) and Lemmas 2.1 and 2.2, yields the following :

Theorem 2.1 Under assumptions (A1)–(A7) and (C(δ)) with $\delta \in \left(0, \frac{1}{3}\right)$, the estimator $\widehat{\theta}_n$ defined in (1.6) satisfies:

$$\widehat{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta_0.$$

3 Asymptotic distribution of the semiparametric estimator

In order to obtain the asymptotic normality of $\widehat{\theta}_n$ we show uniform convergence of the first and second derivatives of \widehat{f}_θ .

Lemma 3.1 Under assumptions (A1)–(A5) and (C(δ)) with $\delta \in \left(\frac{1}{8}, \frac{1}{6}\right)$,

$$n^{1/4} \sup_{(Z, \theta) \in S \times \Theta} \left| \widehat{f}_\theta(Y, X\theta) - f_\theta(Y, X\theta) \right| \xrightarrow[n \rightarrow \infty]{P} 0, \quad (3.1)$$

$$n^{1/4} \sup_{(Z, \theta) \in S \times \Theta} \left| \frac{\partial \widehat{f}_\theta(Y, X\theta)}{\partial \theta} - \frac{\partial f_\theta(Y, X\theta)}{\partial \theta} \right| \xrightarrow[n \rightarrow \infty]{P} 0, \quad (3.2)$$

$$\sup_{(Z, \theta) \in S \times \Theta} \left| \frac{\partial^2 \widehat{f}_\theta(Y, X\theta)}{\partial \theta \partial \theta^T} - \frac{\partial^2 f_\theta(Y, X\theta)}{\partial \theta \partial \theta^T} \right| \xrightarrow[n \rightarrow \infty]{P} 0. \quad (3.3)$$

We'll show that $-\widehat{L}_n(\theta)$ verifies the assumptions of Lemma 5.1 of Ichimura (1993). This is a consequence of Lemma 3.1 and of the following :

Lemma 3.2 *Under the assumptions (A1)-(A5),*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \frac{\partial}{\partial \theta} \log \widehat{f}_\theta(Y_i, X_i \theta) \Big|_{\theta=\theta_0} - \frac{\partial}{\partial \theta} \log f_\theta(Y_i, X_i \theta) \Big|_{\theta=\theta_0} \right\} I_{\{Z_i \in S\}} \xrightarrow[n \rightarrow \infty]{P} 0.$$

The asymptotic distribution of $\widehat{\theta}_n$ is then given by the following :

Theorem 3.1 *Under assumptions (A1)-(A8) and (C(δ)), $\delta \in (\frac{1}{8}, \frac{1}{6})$ and if $\theta_0 \in \overset{\circ}{\Theta}$, then*

$$\sqrt{n} (\widehat{\theta}_n - \theta_0) \xrightarrow{L} N(0, \Sigma), \quad (3.4)$$

with $\Sigma = M^{-1}VM^{-1}$, where

$$V = E \left\{ \frac{\partial}{\partial \theta} \log f_\theta(Y_i, X_i \theta) \Big|_{\theta=\theta_0} \frac{\partial}{\partial \theta^T} \log f_\theta(Y_i, X_i \theta) \Big|_{\theta=\theta_0} I_{\{Z_i \in S\}} \right\}.$$

and M was defined in (A8).

Proof of Theorem 3.1 : It is sufficient to show that

$$-\frac{1}{n} \sum_{i=1}^n \log \widehat{f}_\theta(Y_i, X_i \theta) I_{\{Z_i \in S\}}$$

verifies the conditions (i)-(iv) of Lemma 5.1 of Ichimura (1993).

(i) $\widehat{\theta}_n$ converges almost surely to θ_0 , by Theorem 2.1.

(ii) $-\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f_\theta(Y_i, X_i \theta) \Big|_{\theta=\theta_0} \cdot I_{\{Z_i \in S\}} \xrightarrow[n \rightarrow \infty]{L} N(0, V)$, since

$$\theta_0 = \arg \max_{\theta \in \Theta} E \left[\log f_\theta(Y_i, X_i \theta) I_{\{Z_i \in S\}} \right]$$

$$\implies E \left[\frac{\partial}{\partial \theta} \log f_\theta(Y_i, X_i \theta) \Big|_{\theta=\theta_0} I_{\{Z_i \in S\}} \right] = 0,$$

and

$$\left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\frac{\partial}{\partial \theta} \log \widehat{f}_\theta(Y_i, X_i \theta) \Big|_{\theta=\theta_0} - \frac{\partial}{\partial \theta} \log f_\theta(Y_i, X_i \theta) \Big|_{\theta=\theta_0} \right] \right| I_{\{Z_i \in S\}}$$

converges to 0 in probability by Lemma 3.2.

(iii) $-\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta \partial \theta^T} \widehat{f}_\theta(Y_i, X_i \theta) I_{\{Z_i \in S\}} \xrightarrow[n \rightarrow \infty]{P} E \left[-\frac{\partial^2}{\partial \theta \partial \theta^T} f_\theta(Y_i, X_i \theta) I_{\{Z_i \in S\}} \right]$ uniformly in $\theta \in \Theta$ (Lemma 3.1 and assumption (A5(iii))).

(iv) $M(\theta_0) = E \left[-\frac{\partial^2}{\partial \theta \partial \theta^T} f_\theta(Y_i, X_i \theta) \Big|_{\theta=\theta_0} I_{\{Z_i \in S\}} \right]$ is a positive-definite matrix by (A8).

4 Efficiency of the semiparametric estimator

Let $\gamma_0(y, x)$ be the density of $Z_i = (Y_i, X_i)$ given that $Z_i \in S$, where S is a subset of the support of Z_i (we assume that (Y_i, X_i) is absolutely continuous with respect to a σ -finite measure μ). Here we do not suppose that S satisfies assumption (A4), so that S may coincide with the support of Z_i , even in the case where Z_i is not compactly supported. According to (1.3) (with f replaced by f_0), for each $z \in S$ we have a decomposition of the form :

$$\gamma_0(y, x) = f_0(y, x\theta_0) \cdot g_0(x), \quad (4.1)$$

where $g_0(x)$ is the marginal density of X_i given that $Z_i \in S$. Hence, our semiparametric model is defined by the family of distributions

$$\mathcal{P} = \left\{ P : \frac{dP}{d\mu} = \gamma(y, x; \theta, f, g), \theta \in \Theta, f \in \mathcal{F}, g \in \mathcal{G} \right\} \quad (4.2)$$

with the densities γ satisfying :

- i) $\gamma(y, x; \theta, f, g) = f(y, x\theta) \cdot g(x)$;
- ii) $\gamma(y, x; \theta_0, f_0, g_0) = \gamma_0(y, x)$.

Following Bickel, Klaassen, Ritov and Wellner (1993), in order to determine the bound of the asymptotic variance of an estimator of θ_0 , we need to calculate the efficient score. For this purpose, we first need to determine the tangent space $\dot{\mathcal{P}}_2$ corresponding to the nonparametric part

$$\mathcal{P}_2 = \left\{ P : \frac{dP}{d\mu} = \gamma(y, x; \theta_0, f, g), f \in \mathcal{F}, g \in \mathcal{G} \right\} \quad (4.3)$$

of the model. This is the closed linear span of the union of tangent spaces corresponding to (one-dimensional) regular parametric submodels $\mathcal{Q} \subset \mathcal{P}_2$. Let

$$\mathcal{Q} = \left\{ P : \frac{dP}{d\mu} = \gamma_\eta(y, x; \theta_0) = \gamma(y, x; \theta_0, f(\cdot, \cdot; \eta), g(\cdot; \eta)), \eta \in \mathcal{H} \subset \mathbb{R} \right\} \quad (4.4)$$

be such a submodel. Thus $\{f(\cdot, \cdot; \eta)\}_{\eta \in \mathcal{H}} \subset \mathcal{F}$, $\{g(\cdot; \eta)\}_{\eta \in \mathcal{H}} \subset \mathcal{G}$ and there exists an element $\eta_0 \in \mathcal{H}$ such that $\gamma_{\eta_0}(y, x; \theta_0) = \gamma_0(y, x)$. The tangent space $\dot{\mathcal{Q}}$ of \mathcal{Q} (at γ_0) is simply the linear subspace of $L^2(P_0) = L^2(\gamma_0 \cdot \mu)$ spanned by the score function $S_\eta = \left. \frac{\partial \ln \gamma_\eta(Y_i, X_i; \theta_0)}{\partial \eta} \right|_{\eta=\eta_0}$. We have :

$$S_\eta = \left. \frac{\partial \ln f(Y_i, X_i\theta_0; \eta)}{\partial \eta} \right|_{\eta=\eta_0} + \left. \frac{\partial \ln g(X_i, \eta)}{\partial \eta} \right|_{\eta=\eta_0} \quad (4.5)$$

so that

$$S_\eta \in \mathcal{S} = \{s_1(Y_i, X_i\theta_0) + s_2(X_i) : E_0[s_1(Y_i, X_i\theta_0) | X_i\theta_0] = 0, E_0[s_2(X_i)] = 0\},$$

where E_0 means that the expectation is taken with respect to the probability measure $P_0 = \gamma_0 \cdot \mu$. This means that the tangent space $\dot{\mathcal{P}}_2$ is a subspace of \mathcal{S} .

Let

$$\begin{aligned} S_\theta &= \left. \frac{\partial \ln \gamma(Y_i, X_i; \theta, f_0, g_0)}{\partial \theta} \right|_{\theta=\theta_0} = \left. \frac{\partial \ln f_0(Y_i, X_i; \theta)}{\partial \theta} \right|_{\theta=\theta_0} \\ &= \partial_2 \ln f_0(Y_i, X_i; \theta_0) \cdot X_i. \end{aligned} \quad (4.6)$$

According to Bickel, Klaassen, Ritov and Wellner (1993), Corollary 3.4.1, the information bound on θ_0 is given by $I_0 = E_0(S_\theta^* S_\theta^{*T})$, where S_θ^* , the efficient score, is the residual of the projection of S_θ on $\dot{\mathcal{P}}_2$. Since $\mathcal{S} \subset \dot{\mathcal{P}}_2$, we have $I_0 \geq E_0(S_1 S_1^T)$, where $S_1 = S_\theta - \text{proj}(S_\theta | \mathcal{S})$ (see Lemma 9 of Bonneau, Delecroix and Hristache (1995)).

On the other hand, if

$$\mathcal{Q} = \left\{ P : \frac{dP}{d\mu} = \gamma(y, x; \theta, f(\cdot, \cdot; \theta), g(\cdot; \theta)), \theta \in \Theta \right\} \quad (4.7)$$

is a regular parametric submodel of \mathcal{P} containing P_0 , then the information bound $I(\theta_0, \mathcal{Q})$ on θ_0 in \mathcal{Q} is such that $I(\theta_0, \mathcal{Q}) \geq I_0$. This means that if we can find a parametric submodel \mathcal{Q} such that $I(\theta_0, \mathcal{Q}) = E_0(S_1 S_1^T)$, we have an explicit formula for I_0 :

$$\begin{aligned} I_0 &= E_0(S_1 S_1^T) = E_0 \left\{ [\partial_2 \ln f_0(Y_i, X_i; \theta_0)]^2 \right. \\ &\quad \left. \times [X_i - E_0(X_i | X_i; \theta_0)] [X_i - E_0(X_i | X_i; \theta_0)]^T \right\}, \end{aligned}$$

since the projection of a vector $s(Y_i, X_i) \in L^2(P_0)$ such that $E_0[s(Y_i, X_i) | X_i] = 0$ on \mathcal{S} is simply $E_0[s(Y_i, X_i) | Y_i, X_i; \theta_0]$. It is not difficult to see that the submodel :

$$\mathcal{Q} = \left\{ P : \frac{dP}{d\mu} = \gamma(y, x; \theta, f_\theta, g_\theta), \theta \in \Theta \right\} \quad (4.8)$$

has the desired property : $I(\theta_0, \mathcal{Q}) = E_0(S_1 S_1^T)$, since

$$\left. \frac{d}{d\theta} f_\theta(Y_i, X_i; \theta) \right|_{\theta=\theta_0} = \left. \frac{\partial}{\partial t} f_{\theta_0}(Y_i, t) \right|_{t=X_i; \theta_0} [X_i - E(X_i | X_i; \theta_0)]. \quad (4.9)$$

If we compare I_0 with the asymptotic variance-covariance matrix of our estimator, we can see that, for the given set S satisfying assumption (A4), we obtain the same thing. This means that the estimator we proposed is efficient, in the model built on the data set $(X_i^*, Y_i^*), i \geq 1$, where (X_i^*, Y_i^*) is the i th among those $(X_j, Y_j) \in S$. If S is the support of Z_i , our estimator is efficient in the initial model.

5 Simulation study

The asymptotic efficiency of an estimator is not always the most important argument for a practitioner to use it instead of one which is easy to compute, even if it is not optimal from

a theoretical point of view. This is why methods like those proposed by Powell, Stock and Stoker (1989) or Horowitz and Härdle (1996) are and will be preferred in practice to an estimator which needs optimization procedures, like the one defined by equations (1.6) and (1.7). A possible solution to this problem¹ would be to use a one-step estimator, as a compromise between asymptotical and computational efficiencies, whenever we dispose of an estimate easy to compute.

This can be done in the following way : if $\hat{\theta}_n$ is defined by (1.6) and (1.7), then $\frac{\partial \hat{L}_n}{\partial \theta}(\hat{\theta}_n) = 0$. If $\tilde{\theta}_n$ is a preliminary \sqrt{n} -consistent estimator of θ_0 (we can take, for example, $\tilde{\theta}_n$ as the weighted average derivative estimator of Powell, Stock and Stoker (1989)), then we have

$$\frac{\partial \hat{L}_n}{\partial \theta}(\hat{\theta}_n) = \frac{\partial \hat{L}_n}{\partial \theta}(\tilde{\theta}_n) + \frac{\partial^2 \hat{L}_n}{\partial \theta \partial \theta^T}(\tilde{\theta}_n)(\hat{\theta}_n - \tilde{\theta}_n) + o_P(\|\hat{\theta}_n - \tilde{\theta}_n\|).$$

By the assumption (A8) and the fact that $\hat{\theta}_n$ and $\tilde{\theta}_n$ are root-n-consistent estimators of θ_0 , we obtain :

$$\hat{\theta}_n = \tilde{\theta}_n - \left(\frac{\partial^2 \hat{L}_n}{\partial \theta \partial \theta^T}(\tilde{\theta}_n) \right)^{-1} \frac{\partial \hat{L}_n}{\partial \theta}(\tilde{\theta}_n) + o_P\left(\frac{1}{\sqrt{n}}\right).$$

If we define

$$\bar{\theta}_n = \tilde{\theta}_n - \left(\frac{\partial^2 \hat{L}_n}{\partial \theta \partial \theta^T}(\tilde{\theta}_n) \right)^{-1} \frac{\partial \hat{L}_n}{\partial \theta}(\tilde{\theta}_n), \quad (5.1)$$

we then obtain an asymptotically efficient estimator, since this one-step estimator has the same asymptotic distribution as $\hat{\theta}_n$.

In order to evaluate the performances of the one-step estimator, which is asymptotically equivalent to $\hat{\theta}_n$ but easier to compute, for small sample sizes, we give here the results of a simulation study. We considered the model

$$Y_i = X_i \theta_0 + \varepsilon_i, \quad i = 1, \dots, n,$$

where $Y_i \in \mathbb{R}$, $X_i = (X_i^{(1)}, X_i^{(2)}) \in \mathbb{R}^2$, $\theta_0 = (-1, 1)$, $X_i^{(1)}$ and $X_i^{(2)}$ are independent and of the same law, a mixture of two normal laws,

$$X_i^{(1)}, X_i^{(2)} \sim 0.2 \mathcal{N}(0, 1) + 0.8 \mathcal{N}(0.25, 2),$$

and the errors are normal of mean zero and variance equal to $(X_i \theta_0)^2 = (X_i^{(1)} - X_i^{(2)})^2$:

$$\varepsilon_i \sim N(0, |X_i \theta_0|).$$

As the initial estimator we used the weighted average derivative estimator defined by :

$$\tilde{\theta}_n = -\frac{2}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{1}{h_n^3} K' \left(\frac{X_i - X_j}{h_n} \right) Y_i,$$

¹As suggested by practical experience of J. Horowitz

where $K' \left(\frac{X_i - X_j}{h_n} \right)$ is a notation for the vector

$$\begin{pmatrix} K' \left(\frac{X_i^{(1)} - X_j^{(1)}}{h_n} \right) & K \left(\frac{X_i^{(2)} - X_j^{(2)}}{h_n} \right) \\ K \left(\frac{X_i^{(1)} - X_j^{(1)}}{h_n} \right) & K' \left(\frac{X_i^{(2)} - X_j^{(2)}}{h_n} \right) \end{pmatrix} \in \mathbb{R}^2,$$

the real valued kernel function is defined by

$$K(u) = \begin{cases} \frac{1}{4}(7 - 31u^2), & |u| \leq \frac{1}{2}, \\ \frac{1}{4}(u^2 - 1), & \frac{1}{2} \leq |u| \leq 1, \\ 0, & |u| \geq 1, \end{cases}$$

and the bandwidth h_n is of the form $h_n = 6 n^{-1/5}$.

For the one-step estimator $\bar{\theta}_n$ given by (5.1), we used the same kernel K in the definition of $\hat{L}_n(\theta)$ and a bandwidth of the form $h_n = 2.5 n^{-1/7.5}$.

As only the direction of θ_0 can be identified and not θ_0 itself, we used for the estimators the same constraint as for θ_0 , that the last component equals 1. The results for the estimation of $\theta_0^{(1)} = -1$ using the weighted average derivative estimator $\tilde{\theta}_n$ and the one-step estimator $\bar{\theta}_n$ with samples sizes $n \in \{50, 100, 200, 400\}$ are summarized in the following table, containing the empirical mean and the empirical mean squared error for each case :

	$n = 50$	$n = 100$	$n = 200$	$n = 400$
$\tilde{\theta}_n$	-1.0406 (0.1067)	-1.0174 (0.0305)	-1.0187 (0.0152)	-1.0030 (78,45 · 10 ⁻⁴)
$\bar{\theta}_n$	-0.9599 (0.0866)	-0.9649 (0.0269)	-0.9808 (0.0138)	-0.9806 (77.81 · 10 ⁻⁴)

As a general conclusion, we can say that the one-step estimator works better than the initial one. However, the rate of improvement of the squared error decreases with the sample size (18.83%, 11.80%, 9.21% and 0.81% respectively), but this may be only a consequence of our bandwidth choice, which is for no reason optimal. Moreover, if we change the constant in the bandwidth used to obtain $\bar{\theta}_n$ from 2.5 to 2.0, taking $h_n = 2.0 n^{-1/7.5}$, this phenomenon disappears but the general conclusion remains the same, namely that $\bar{\theta}_n$ provides better estimates of θ_0 than $\tilde{\theta}_n$ (except for a small "accident" in the case $n = 100$) :

	$n = 50$	$n = 100$	$n = 200$	$n = 400$
$\tilde{\theta}_n$	-1.0406 (0.1067)	-1.0174 (0.0305)	-1.0187 (0.0152)	-1.0030 (78,45 · 10 ⁻⁴)
$\bar{\theta}_n$	-0.9812 (0.0961)	-0.9786 (0.0307)	-0.9886 (0.0134)	-0.9857 (75.37 · 10 ⁻⁴)

6 Appendix

Proof of Lemma 2.1

We first show that

$$\sup_{z \in S} \sup_{\theta \in \Theta} |D_{i,n}(x\theta) - h_\theta(x\theta)\pi_S| \xrightarrow[n \rightarrow \infty]{a.s.} 0, \quad (6.1)$$

$$\sup_{z \in S} \sup_{\theta \in \Theta} |N_{i,n}(y, x\theta) - f_\theta(y, x\theta)h_\theta(x\theta)\pi_S| \xrightarrow[n \rightarrow \infty]{a.s.} 0, \quad (6.2)$$

where $\pi_S = P((X_i, Y_i) \in S)$.

From (6.1) and (6.2) we conclude :

$$\begin{aligned} & \sup_{z \in S} \sup_{\theta \in \Theta} |\log [(N_{i,n}(y, x\theta)/D_{i,n}(x\theta))] - \log f_\theta(y, x\theta)| \\ & \leq \sup_{z \in S} \sup_{\theta \in \Theta} |\log N_{i,n}(y, x\theta) - \log [f_\theta(y, x\theta) h_\theta(x\theta)\pi_S]| \\ & \quad + \sup_{x \in S_X} \sup_{\theta \in \Theta} |\log D_{i,n}(x\theta) - \log [h_\theta(x\theta)\pi_S]| \\ & \leq \left(\inf_{z \in S} \inf_{\theta \in \Theta} [N_{i,n}(y, x\theta), f_\theta(y, x\theta)h_\theta(x\theta)\pi_S] \right)^{-1} \\ & \quad \times \sup_{z \in S} \sup_{\theta \in \Theta} |N_{i,n}(y, x\theta) - f_\theta(y, x\theta)h_\theta(x\theta)\pi_S| \\ & \quad + \left(\inf_{x \in S_X} \inf_{\theta \in \Theta} [D_{i,n}(x\theta), h_\theta(x\theta)\pi_S] \right)^{-1} \sup_{x \in S_X} \sup_{\theta \in \Theta} |D_{i,n}(x\theta) - h_\theta(x\theta)\pi_S|, \end{aligned}$$

since $\inf_{z \in S} \inf_{\theta \in \Theta} D_{i,n}(x\theta)$ can be asymptotically bounded below almost surely by $\frac{1}{2} \inf_{\substack{x \in S_X \\ \theta \in \Theta}} [h_\theta(x\theta)\pi_S]$,

which is positive by (A4(i)) and similarly $N_{i,n}(y, x\theta)$ is bounded below almost surely by $\frac{1}{2} \inf_{\substack{z \in S \\ \theta \in \Theta}} [f_\theta(y, x\theta)h_\theta(x\theta)\pi_S] > 0$.

It remains to show (6.1) and (6.2). We show only (6.1). The argument for (6.2) is similar in character.

Since, for $z = (x, y) \in S$, we have

$$D_{i,n}(x\theta) = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^n K_h(x\theta - X_j\theta) \cdot I_{\{Z_j \in S\}},$$

adapting results of Klein and Spady (1993) or Ichimura (1993) it suffices to find a function $h_\infty(y, x, \theta)$ such that :

$$\sup_{z \in S} \sup_{\theta \in \Theta} \left| E \left\{ K_h(x\theta - X_j\theta) \cdot I_{\{Z_j \in S\}} \right\} - h_\infty(y, x, \theta) \right| \xrightarrow[n \rightarrow \infty]{} 0.$$

We have, with a big enough constants $C', C > 0$,

$$\begin{aligned}
& \sup_{z \in S} \sup_{\theta \in \Theta} \left| E \left\{ K_h(x\theta - X_j\theta) \cdot I_{\{Z_j \in S\}} \right\} - h_\theta(x\theta) \pi_S \right| \\
& \leq \pi_S \sup_{z \in S} \sup_{\theta \in \Theta} \left| \int K_h(x\theta - y) \{h_\theta(y) - h_\theta(x\theta)\} dy \right| \\
& \leq \pi_S \sup_{x \in S_X} \sup_{\theta \in \Theta} \int K(u) \{h_\theta(x\theta - uh_n) - h_\theta(x\theta)\} du \\
& = \pi_S \int K(u) \left\{ \sup_{x \in S_X} \sup_{\theta \in \Theta} |h_\theta(x\theta - uh_n) - h_\theta(x\theta)| \right\} du \\
& \leq C' \left\{ h_n \int |u| K(u) du \right\} \leq Ch_n,
\end{aligned}$$

and then (6.1) is checked.

The proof of Lemma 3.1 is similar and hence omitted. The only difference is that we use a Taylor expansion in probability.

Proof of Lemma 3.2 : We can write

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \partial_\theta \log \hat{f}_{\theta_0}(Y_i, X_i\theta_0) I_{\{Z_i \in S\}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial_\theta \hat{f}_{\theta_0 i}}{\hat{f}_{\theta_0 i}} I_i.$$

Recalling the definition $\hat{f}_\theta(Y_i, X_i\theta) = \frac{N_{i,n}(Y_i, X_i\theta)}{D_{i,n}(X_i\theta)}$, we have

$$\partial_\theta \hat{f}_{\theta_0 i} = \frac{\partial_\theta N_{i,n}(\theta_0)}{D_{i,n}(\theta_0)} - \hat{f}_{\theta_0 i} \frac{\partial_\theta D_{i,n}(\theta_0)}{D_{i,n}(\theta_0)},$$

which gives

$$\frac{\partial_\theta \hat{f}_{\theta_0 i}}{\hat{f}_{\theta_0 i}} = \frac{\partial_\theta N_{i,n}(\theta_0)}{N_{i,n}(\theta_0)} - \frac{\partial_\theta D_{i,n}(\theta_0)}{D_{i,n}(\theta_0)},$$

and finally

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial_\theta \hat{f}_{\theta_0 i}}{\hat{f}_{\theta_0 i}} I_i = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{\partial_\theta N_{i,n}(\theta_0)}{N_{i,n}(\theta_0)} - \frac{\partial_\theta D_{i,n}(\theta_0)}{D_{i,n}(\theta_0)} \right) I_i. \quad (6.3)$$

A decomposition similar to the one used by Ichimura (1993) yields

$$\frac{\partial_\theta N_{i,n}(\theta_0)}{N_{i,n}(\theta_0)} = \frac{\partial_\theta g_{\theta_0 i}}{g_{\theta_0 i}} + \frac{1}{g_{\theta_0 i}} \left(\partial_\theta N_{i,n}(\theta_0) - \frac{\partial_\theta g_{\theta_0 i}}{g_{\theta_0 i}} N_{i,n}(\theta_0) \right) + o_p(n^{-1/2}), \quad (6.4)$$

and

$$\frac{\partial_\theta D_{i,n}(\theta_0)}{D_{i,n}(\theta_0)} = \frac{\partial_\theta h_{\theta_0 i}}{h_{\theta_0 i}} + \frac{1}{h_{\theta_0 i}} \left(\partial_\theta D_{i,n}(\theta_0) - \frac{\partial_\theta h_{\theta_0 i}}{h_{\theta_0 i}} D_{i,n}(\theta_0) \right) + o_p(n^{-1/2}), \quad (6.5)$$

Since

$$\frac{\partial_\theta g_{\theta_0 i}}{g_{\theta_0 i}} = \frac{\partial_\theta (f_{\theta_0 i} h_{\theta_0 i})}{f_{\theta_0 i} h_{\theta_0 i}} = \frac{\partial_\theta f_{\theta_0 i}}{f_{\theta_0 i}} \frac{h_{\theta_0 i}}{h_{\theta_0 i}} + \frac{f_{\theta_0 i}}{f_{\theta_0 i} h_{\theta_0 i}} \frac{\partial_\theta h_{\theta_0 i}}{h_{\theta_0 i}} = \frac{\partial_\theta f_{\theta_0 i}}{f_{\theta_0 i}} + \frac{\partial_\theta h_{\theta_0 i}}{h_{\theta_0 i}},$$

we obtain

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial_\theta \widehat{f}_{\theta_0 i}}{\widehat{f}_{\theta_0 i}} I_i &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial_\theta f_{\theta_0 i}}{f_{\theta_0 i}} I_i + \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial_\theta N_{i,n}(\theta_0)}{g_{\theta_0 i}} I_i}_{\sqrt{n} U_{1n}} \\ &\quad - \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial_\theta g_{\theta_0 i}}{g_{\theta_0 i}^2} N_{i,n}(\theta_0) I_i}_{\sqrt{n} U_{2n}} - \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial_\theta D_{i,n}(\theta_0)}{h_{\theta_0 i}} I_i}_{\sqrt{n} U_{3n}} \\ &\quad + \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial_\theta h_{\theta_0 i}}{h_{\theta_0 i}^2} D_{i,n}(\theta_0) I_i}_{\sqrt{n} U_{4n}} + o_p(1). \end{aligned}$$

Writing U_{1n}, \dots, U_{4n} as second order U -statistics and applying Lemma 3.1 of Powell, Stock and Stoker (1989) yields $\sqrt{n} (U_{kn} - \tilde{U}_{kn}) \xrightarrow[n \rightarrow \infty]{P} 0$ ($1 \leq k \leq 4$), where

$$\begin{aligned} \tilde{U}_{1n} &= \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1}{g_{\theta_0}(Y_i, X_i \theta_0)} [X_i - E(X_i | X_i \theta_0)] \partial_2 g_{\theta_0}(Y_i, X_i \theta_0) I_i \right. \\ &\quad \left. - 2 \frac{d}{dt} E(X_i | X_i \theta_0 = t)|_{t=X_i \theta_0} I_i \right\} + E \left[\frac{d}{dt} E(X_i | X_i \theta_0 = t)|_{t=X_i \theta_0} I_i \right], \end{aligned}$$

$$\tilde{U}_{2n} = \frac{1}{n} \sum_{i=1}^n \frac{\partial_\theta g_{\theta_0}(Y_i, X_i \theta_0)}{g_{\theta_0}(Y_i, X_i \theta_0)} I_i,$$

$$\begin{aligned} \tilde{U}_{3n} &= \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1}{h_{\theta_0}(X_i \theta_0)} [X_i - E(X_i | X_i \theta_0)] \dot{h}_{\theta_0}(X_i \theta_0) I_i \right. \\ &\quad \left. - 2 \frac{d}{dt} E(X_i | X_i \theta_0 = t)|_{t=X_i \theta_0} I_i \right\} + E \left[\frac{d}{dt} E(X_i | X_i \theta_0 = t)|_{t=X_i \theta_0} I_i \right]; \end{aligned}$$

$$\tilde{U}_{4n} = \frac{1}{n} \sum_{i=1}^n \frac{\partial_\theta h_{\theta_0}(X_i \theta_0)}{h_{\theta_0}(X_i \theta_0)} I_i.$$

We thus obtain

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial_\theta \widehat{f}_{\theta_0 i}}{\widehat{f}_{\theta_0 i}} I_i \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial_\theta f_{\theta_0 i}}{f_{\theta_0 i}} I_i + \sqrt{n} (U_{1n} - U_{2n} - U_{3n} + U_{4n}) + o_p(1) \end{aligned}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial_{\theta} f_{\theta_0 i}}{f_{\theta_0 i}} I_i + \sqrt{n} (\tilde{U}_{1n} - \tilde{U}_{2n} - \tilde{U}_{3n} + \tilde{U}_{4n}) + o_p(1),$$

and finally,

$$\left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial_{\theta} \hat{f}_{\theta_0 i}}{\hat{f}_{\theta_0 i}} I_i - \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial_{\theta} f_{\theta_0 i}}{f_{\theta_0 i}} I_i \right| \xrightarrow[n \rightarrow \infty]{P} 0.$$

References

- BICKEL, P.J. (1982): "On adaptive estimation", *Annals of Statistics*, **10**, 647-671.
- BICKEL, P.J., KLAASSEN, C.A.J., RITOV, Y. and WELLNER, J.A. (1993): *Efficient and Adaptive Estimation for Semiparametric Models*, The Johns Hopkins University Press.
- BONNEU, M., DELECROIX, M. and HRISTACHE, M. (1995): "Semiparametric estimation of generalized linear models", preprint, CREST.
- CARROLL, R.J. (1982): "Adapting for heteroscedasticity in linear models", *Annals of Statistics*, **10**, 1224-1233.
- COSSLETT, S.R. (1987): "Efficiency bounds for distribution-free estimators of the binary choice and censored regression models", *Econometrica*, **55**, 559-585.
- FRIEDMAN, J.H. and STUETZLE, W. (1981): "Projection pursuit regression", *Journal of the American Statistical Association*, **76**, 817-823.
- GOURIÉROUX, C. and MONFORT, A. (1989): *Statistique et modèles économétriques*, Economica.
- HAN, A.K. (1987): "Non-parametric analysis of a generalized regression model", *Journal of Econometrics*, **35**, 305-316.
- HOROWITZ, J.L. and HÄRDLE, W. (1996), "Direct semiparametric estimation of single-index models with discrete covariates", *Journal of the American Statistical Association*, to appear.
- HUET, S., JOLIVET, E. and MESSÉAN, A. (1989): "Some simulation results about confidence intervals and bootstrap methods in nonlinear regression", *Statistics*, **21**, 369-432.
- ICHIMURA, H. (1993): "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models", *Journal of Econometrics*, **58**, 71-120.
- KLEIN, R.L. and SPADY, R.H. (1993): "An efficient semiparametric estimator for binary response models", *Econometrica*, **61**, 387-421.

- POWELL, J.L., STOCK, J.M. and STOKER, T.M. (1989): "Semiparametric estimation of index coefficients", *Econometrica*, **57**, 1403-1430.
- ROBINSON, P.M. (1987): "Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form", *Econometrica*, **55**, 875-891.
- SHERMAN, R.P. (1993): "The limiting distribution of the maximum rank correlation estimator", *Econometrica*, **61**, 123-137.