Optimal smoothing in semiparametric index approximation of regression functions

M. Delecroix[1], M. Hristache[1] and V. Patilea[2]

[1]ENSAI and CREST, Campus de Ker Lann, 35170 Bruz, France
[2]Université d'Orléans, Rue de Blois, BP 6739, 45067 Orléans Cedex 2, France

Abstract

The problem of approximating a general regression function $m(x) = E(Y|X = x)$ is addressed. As in the case of the classical L_2-type projection pursuit regression considered by Hall (1989), we propose to approximate $m(x)$ through a regression of Y given an index, that is a unidimensional projection of X. The orientation vector defining the projection of X is taken to be the optimum of a Kullback-Leibler type criterion. The first step of the classical projection pursuit regression and the single-index models (SIM) are obtained as particular cases. We define a kernel-based estimator of the 'optimal' orientation vector and we suggest a simple empirical bandwidth selection rule. Finally, the true regression function $m(\cdot)$ is approximated through a kernel regression of Y given the estimated index. Our procedure extends the idea of Härdle, Hall and Ichimura (1993) which propose, in the case of SIM, to minimize an empirical L_2-type criterion simultaneously with respect to the orientation vector and the bandwidth. We show that a same bandwidth of order $n^{-1/5}$ can be used for the root-n estimation of the orientation and for the kernel approximation of the true regression function. Our methodology could be extended to more accurate multi-index approximations.

1 Introduction

The statistical problem of estimating a regression function $m(x) = E(Y|X = x)$ from independent copies $(Y_1, X_1), \ldots, (Y_n, X_n)$ of a random vector $(Y, X) \in I\!R \times I\!R^d$ has been extensively studied. The classical approach remains the linear regression model which assumes that the conditional law of Y given $X = x$ is normal of mean $m(x) = x\theta_0$, $\theta_0 \in I\!R^d$, and variance σ^2 (herein $x\theta$ denotes the usual inner product of x and θ in $I\!R^d$). This model is a particular case of the generalized linear models (GLM) as considered, *e.g.*, in McCullagh and Nelder (1989). GLM are defined by :

1. $m(x) = r_0(x\theta_0)$ with r_0 known (r_0 is the inverse of the so-called link function);
2. the conditional density $f_{Y|X=x}$ of Y given $X = x$ belongs to the linear exponential family, *i.e.*,

$$f_{Y|X=x}(y) = \exp\left[B(r_0(x\theta_0)) + C(r_0(x\theta_0))\, y + D(y)\right],$$

where B, C and D are known functions.

A natural extension of GLM is provided by the semiparametric single-index models (SIM), where one only assumes that there exists some $\theta_0 \in I\!R^d$ such that

$$E\left(Y|X\right) = E\left(Y|X\theta_0\right), \tag{1}$$

that is $m\left(x\right) = r_0\left(x\theta_0\right)$, with unknown r_0. In this framework, both θ_0 and r_0 are to be estimated. Root-n-consistent estimation of θ_0 has been obtained, for example, by Ichimura (1993), Sherman (1994b), Delecroix and Hristache (1999). The M-estimators of θ_0 proposed by these authors can be written as

$$\widehat{\theta} = \arg\max_{\theta} \ \frac{1}{n}\sum_{i=1}^{n} \psi\left(Y_i, \widehat{r}_{\theta,h}\left(X_i\theta\right)\right), \tag{2}$$

where $\widehat{r}_{\theta,h}\left(t\right)$ is a kernel estimator (with bandwidth h) of $r_\theta\left(t\right) = E\left(Y|X\theta = t\right)$ and $-\psi$ a contrast function. When $\exp\psi$ is linear exponential, that is

$$\psi(y,r) = B(r) + C(r)y + D(y), \tag{3}$$

such semiparametric $M-$estimators are justified by the fact that θ_0 of the SIM assumption verifies

$$\theta_0 = \arg\max_{\theta} \ E\left[\psi\left(Y,\ E\left(Y|X\theta\right)\right)\right] \tag{4}$$

(see Delecroix and Hristache (1999); see also Gouriéroux, Monfort and Trognon (1984)). Finally, the regression function $m\left(x\right)$ is estimated by $\widehat{r}_{\widehat{\theta},h}\left(x\widehat{\theta}\right)$.

When the SIM condition does not seem to be valid, one may nevertheless approximate $m\left(x\right)$ by $\widehat{r}_{\widehat{\theta},h}\left(x\widehat{\theta}\right)$ defined above. This is an extension of the first step of the so-called projection pursuit method, as considered by Hall (1989); see also Friedman and Stuetzle (1981). Hall suggested to approximate $E\left(Y|X\right)$ by the unidimensional regression $r_{\theta_1}\left(\cdot\,\theta_1\right) = E\left(Y|X\theta_1 = \cdot\,\theta_1\right)$, with θ_1 an orientation vector. The vector θ_1, also called the first projective direction, is defined as the minimizer of a L_2-type distance :

$$\theta_1 = \arg\min_{\theta} \ E\left\{\left[E\left(Y|X\right) - E\left(Y|X\theta\right)\right]^2\right\}. \tag{5}$$

More generally, herein we propose to approximate $m(\cdot) = E\left(Y|X = \cdot\right)$ by an unidimensional regression

$$r_{\theta_\psi}\left(\cdot\,\theta_\psi\right) = E\left(Y|X\theta_\psi = \cdot\,\theta_\psi\right) = E\left(m(X)|X\theta_\psi = \cdot\,\theta_\psi\right), \tag{6}$$

where θ_ψ, that we call the first $\psi-$projective direction, is defined by

$$\theta_\psi = \arg\max_{\theta} \ E\left[\psi\left(E\left(Y|X\right),\ E\left(Y|X\theta\right)\right)\right], \tag{7}$$

with ψ as in (3). The vector θ_ψ and the function r_{θ_ψ} are to be estimated. Note that θ_ψ is determined only up to a scaling factor. When $\psi\left(y,r\right) = -\left(y-r\right)^2$, we recover the framework considered by Hall (1989). Moreover, if $\exp\psi$ is linear exponential, θ_ψ is also the maximizer of $E\left[\psi\left(Y,\ E\left(Y|X\theta\right)\right)\right]$ since in this case

$$E\left[\psi\left(E\left(Y|X\right),\ E\left(Y|X\theta\right)\right)\right] = E\left[\psi\left(Y,\ E\left(Y|X\theta\right)\right)\right]. \tag{8}$$

If, in addition, the SIM assumption (1) holds, θ_ψ coincides with θ_0, since they are both the maximizers with respect to θ of the same quantity (see (4)). In other cases, approximating the regression function $m(\cdot)$ by $r_{\theta_\psi}(\cdot\,\theta_\psi)$ rather than $r_{\theta_1}(\cdot\,\theta_1)$ is a well-adapted statistical solution when, for instance, some information on the conditional distribution of Y given X is available. As an example, let us consider the case where the conditional distribution of Y given $X = x$ is Bernoulli of parameter $m(x)$. It seems natural to choose θ_ψ by minimizing with respect to θ the Kullback-Leibler contrast between the Bernoulli laws of parameters $m(X)$ and $E(Y|X\theta)$, respectively :

$$E\left[\log\ \frac{m(X)^Y(1-m(X))^{1-Y}}{E(Y|X\theta)^Y(1-E(Y|X\theta))^{1-Y}}\right].$$

This leads to take $\psi(y,r) = y\log r + (1-y)\log(1-r)$ in (7).

In order to estimate $r_{\theta_\psi}(\cdot\,\theta_\psi)$, two bandwidths seem to be necessary. First, after choosing a primary bandwidth h, the estimator $\widehat{\theta}$ is computed as in (2). Afterwards, $r_{\theta_\psi}(x\theta_\psi)$ is estimated by $\widehat{r}_{\widehat{\theta},h^*}\left(x\widehat{\theta}\right)$, a kernel estimator, with bandwidth h^*, of the expectation of Y given $x\widehat{\theta}$. The rates of decay for the two bandwidths should verify some conditions. In a SIM framework, Härdle, Hall and Ichimura (1993) defined more directly

$$\left(\widehat{\theta},\widehat{h}\right) = \arg\max_{\theta,h}\ \frac{1}{n}\sum_{i=1}^{n}\psi\left(Y_i,\widehat{r}_{\theta,h}(X_i\theta)\right) \tag{9}$$

for $\psi(y,r) = -(y-r)^2$. In this paper we extend their idea to more general ψ and we estimate $r_{\theta_\psi}(\cdot\,\theta_\psi)$ by $\widehat{r}_{\widehat{\theta},\widehat{h}}\left(\cdot\,\widehat{\theta}\right)$. Asymptotic properties of $\left(\widehat{\theta},\widehat{h}\right)$ and $\widehat{r}_{\widehat{\theta},\widehat{h}}\left(\cdot\,\widehat{\theta}\right)$ are obtained without necessarily assuming the SIM condition. For $\widehat{\theta}$, the estimator of θ_ψ, we prove convergence and asymptotic normality, while for $\widehat{h}$ we obtain an asymptotic equivalence with the theoretical optimal bandwidth minimizing what we call the $\psi-$MISE. When $\psi(y,r) = -(y-r)^2$ our $\psi-$MISE coincides with the usual MISE from nonparametric smoothing. The asymptotic normality of $\widehat{r}_{\widehat{\theta},\widehat{h}}\left(\cdot\,\widehat{\theta}\right)$ is easily obtained after proving its asymptotic equivalence with a classic kernel estimator of $r_{\theta_\psi}(\cdot\,\theta_\psi)$.

Since the 'optimal' bandwidth selection rule we propose herein is applicable whether the SIM condition is verified or not, our findings contradict an intuitive argument provided by Hall (1989), page 583, who suggests that two quite different bandwidths may be necessary in order to construct good estimators of θ_1 and r_{θ_1} in the case of the classical projection pursuit regression.

Our methodology could be extended to a multi-index framework. More precisely, consider a semiparametric multi-index model, where it is assumed that there exists $p > 1$, but smaller than d, the dimension of X, and $\theta_0^1, ..., \theta_0^p \in I\!R^d$, such that

$$E(Y|X) = E\left(Y|X\theta_0^1, ..., X\theta_0^p\right). \tag{10}$$

Under suitable identification conditions (see Ichimura and Lee (1991)), the estimation methodology that we propose herein could be applied to such multi-index models. That is, for ψ as in (3), we may follow (9) and define

$$\left(\widehat{\theta}^1, ..., \widehat{\theta}^p, \widehat{h}\right) = \arg\max_{\theta^1,...,\theta^p,h}\ \frac{1}{n}\sum_{i=1}^{n}\psi\left(Y_i, \widehat{r}_{\left(\theta^1,...,\theta^p\right),h}\left(X_i\theta^1, ..., X_i\theta^p\right)\right) \tag{11}$$

with $\widehat{r}_{(\theta^1,...,\theta^p),h}(t_1,...,t_p)$ a kernel estimator of $E\left(Y \mid X\theta^1 = t_1,...,X\theta^p = t_p\right)$. Moreover, let us note that even if (10) is not verified, one may still want to approximate the regression function $m(\cdot)$ by a multi-index regression more accurate than a single-index regression. A multi-index $\left(X\theta_\psi^1,...,X\theta_\psi^p\right)$ can be defined through

$$\left(\theta_\psi^1,...,\theta_\psi^p\right) = \underset{(\theta^1,...,\theta^p)}{\arg\max}\ E\left[\psi\left(E\left(Y|X\right),\ E\left(Y|X\theta^1,...,X\theta^p\right)\right)\right]. \tag{12}$$

The regression $m(\cdot)$ is then approximated by

$$E\left(Y|X\theta_\psi^1 = \cdot\,\theta_\psi^1,...,X\theta_\psi^p = \cdot\,\theta_\psi^p\right) = E\left(m(X)|X\theta_\psi^1 = \cdot\,\theta_\psi^1,...,X\theta_\psi^p = \cdot\,\theta_\psi^p\right),$$

which, in turn, is estimated through a kernel smoother $\widehat{r}_{\left(\widehat{\theta}^1,...,\widehat{\theta}^p\right),\widehat{h}}\left(\cdot\,\widehat{\theta}^1,...,\cdot\,\widehat{\theta}^p\right)$, with $\left(\widehat{\theta}^1,...,\widehat{\theta}^p,\widehat{h}\right)$ defined as in (11). Asymptotic results for the parametric and the non-parametric parts of the semiparametric multi-index regression could be obtained without assuming condition (10). The arguments are similar to those used in the unidimensional case. For the sake of simplicity, herein we confine ourselves to the case $p = 1$.

Section 2 states the assumptions and gives the simultaneous definition of $\widehat{\theta}$, the estimator of θ_ψ, and of the 'optimal' bandwidth $\widehat{h}$. Section 3 contains the asymptotic results : convergence and asymptotic normality for $\widehat{\theta}$, asymptotic rate for $\widehat{h}$ and asymptotic normality of $\widehat{r}_{\widehat{\theta},\widehat{h}}\left(\cdot\,\widehat{\theta}\right)$. The first part of the appendix recalls definitions and results of Sherman (1994a) on rates of convergence of degenerate $U-$statistics. Sherman's results represent main tools for our proofs presented in the second part of the appendix.

2 Definitions and assumptions

2.1 Model assumptions

As announced in the introduction, the observations $(Y_1,X_1),\ldots,(Y_n,X_n)$ are n independent copies of a random vector $(Y,X) \in I\!R \times I\!R^d$. In order to define and estimate θ_ψ, we consider a criterion

$$\psi(y,r) = B(r) + C(r)\,y,$$

where :

1. B,$C : I \subseteq I\!R \longrightarrow I\!R$ are twice differentiable, with B'' and C'' Lipschitzian on any compact subset of I;
2. $B'(r) + C'(r)\,r \equiv 0$;
3. $C'(r) > 0$.

For the sake of brevity we finally omit the D function in the definition of ψ since D has no influence in the optimization problem (7). The last two conditions above simply mean that $\exp\psi$ is proportional to a linear exponential density of expectation r and variance $[C'(r)]^{-1}$.

In order to ensure the identifiability of the $d-$dimensional parameter θ, we have to impose some restrictions on its components. Two approaches has been usually used :

either the norm of θ and the sign of one of its components are fixed (see, *e.g.* Härdle, Hall and Ichimura (1993) and Hall (1989)), or one of its components is fixed (see, *e.g.* Sherman (1994b)). For the sake of simplicity, herein we adopt the latter solution and we fix the first component of θ to one.

Suppose that there exists a compact subset $\widetilde{\Theta}$ of $\mathbb{R}^{d-1}$ with non empty interior such that the assumptions below are verified. In fact, in the sequel of the paper we rather use a compact

$$\Theta = \left\{ \theta \ : \theta = \ {}^t\left(1, {}^t\widetilde{\theta}\right) \ , \ \widetilde{\theta} \in \widetilde{\Theta} \right\} \subset \mathbb{R}^d$$

which we identify with $\widetilde{\Theta}$ via the application $\widetilde{\theta} \mapsto {}^t\left(1, \ {}^t\widetilde{\theta}\right)$.

Assumption 1 For every $\theta \in \Theta$, the random variable $X\theta$ admits a density f_θ with respect to the Lebesgue measure on $\mathbb{R}$.

In order to guard against working with denominators close to zero, we consider a so-called trimming on the X variables : fix arbitrarily A, a compact subset of $\mathbb{R}^d$, and consider only the observations (Y_i, X_i) with $X_i \in A$. Hereafter $I_A(X)$ denotes a function which equals one if X belongs to A and zero otherwise. For any $\theta \in \Theta$, let us define

$$A(\theta) = \{t : t = x\theta, \ x \in A\}$$

which it assumed to be a finite union of proper intervals. For any $t \in A(\theta)$, let

$$\begin{aligned} N_\theta(t) &= E(Y \ I_A(X) \ | \ X\theta = t) \cdot f_\theta(t), \\ D_\theta(t) &= E(I_A(X) \ | \ X\theta = t) \cdot f_\theta(t). \end{aligned} \tag{13}$$

Assumption 2

1. The functions $(\theta, x) \to N_\theta(x\theta)$ and $(\theta, x) \to D_\theta(x\theta)$ are continuous on $\Theta \times A$.
2. $\inf_{\substack{\theta \in \Theta \\ x \in A}} D_\theta(x\theta) > 0$.
3. For some $k \geq 1$, the function $r_\theta(x\theta)$ is k times differentiable with respect to the last $d-1$ components of θ and the kth derivatives are continous on $A \times \Theta$ and, for any $x \in A$, Lipschitzian on Θ.
4. For any θ in Θ, the functions $N_\theta(\cdot)$ and $D_\theta(\cdot)$ are twice differentiable on $A(\theta)$, with Lipschitzian second derivatives and corresponding Lipschitz constants independent of θ.

Let P^A be the conditional probability measure defined by

$$P^A(\Lambda) = \frac{P(\Lambda \cap \{X \in A\})}{P(\{X \in A\})}$$

for any measurable set Λ. Hereafter, we consider the following notation : for any $\theta \in \Theta$ and $t \in A(\theta)$,

$$r_\theta(t) = E^A(Y | X\theta = t), \tag{14}$$

denotes the conditional expectation of Y given $X\theta = t$, while

$$v_\theta(t) = var^A(Y|X\theta = t), \tag{15}$$

stands for the conditional variance of Y given $X\theta = t$, both defined with respect to P^A. We also call $r_\theta(\cdot)$ a trimmed regression function.

Assumption 3 There exists $\delta > 0$ such that

$$\{r \in I\!R : \exists\, \theta \in \Theta \text{ and } t \in A(\theta) \quad \text{ such that } \quad |r - r_\theta(t)| \leq \delta\ \} \subset I,$$

where I is the domain of the functions B and C defining the contrast function ψ.

Assumption 4 $\sup_{\substack{\theta\in\Theta \\ x\in A}} v_\theta(x\theta) < \infty$.

Assumption 5 There exists a unique θ_ψ interior point of Θ, *i.e.*, $\theta_\psi = {}^t\left(1,\ {}^t\widetilde{\theta}_\psi\right)$ and $\widetilde{\theta}_\psi$ interior point of $\widetilde{\Theta}$, such that

$$E\left[\psi\left(Y, E^A(Y|X\theta_\psi)\right) I_A(X)\right] = \max_{\theta\in\Theta} E\left[\psi\left(Y, E^A(Y|X\theta)\right) I_A(X)\right]. \tag{16}$$

Let us make some comments on the previous assumptions. The $U-$processes techniques we use in the proofs allow us to relax an embarrassing usual constraint, that is the random vector X admits a density (see Härdle, Hall and Ichimura (1993) and Hall (1989)), and replace it with Assumption 1. Note that $X\theta$ may have a density even if X has discrete components, which is often the case in applications. A trimming procedure is quite usual when a nonparametric regression estimator is involved. Even if one assumes X having a density bounded away from zero, this does not yet ensure the second part of Assumption 2. The restriction introduced by the set A could be quite straightforwardly eliminated by considering a sequence of nonrandom or data-driven sets A_n, $n \geq 1$, growing at a suitable speed to the whole set of values of X (see, *e.g.*, Sherman (1994b)). Since this implies longer proofs, for the sake of simplicity, we confine ourselves of a fixed trimming. Assumption 3 ensures that the estimator proposed below is well defined, at least for a large enough sample size. Assumption 4 is used to control the envelopes of certain classes of functions appearing in the proof of the asymptotic normality of our estimator. Assumption 5 identifies the $\psi-$first projective direction θ_ψ.

It is to be stressed that two trimming devices appear in Assumption 5. First, a trimmed criterion ψI_A is convenient in order to avoid denominators close to zero. On the other hand, the trimmed regression $E^A(Y|X\theta)$, instead of $E(Y|X\theta)$, appears as argument of ψ. This facilitates the proofs because it allows us to recover *degenerate* $U-$statistics when decomposing the empirical counterpart of $E\left[\psi\left(Y, E^A(Y|X\theta)\right) I_A(X)\right]$ that we use for defining the estimator of θ_ψ.

In the following lemma we give an useful interpretation of the trimmed regression. Moreover, we show that if the SIM condition holds (see (1)), θ_ψ of Assumption 5 coincides with θ_0.

Lemma 1

1. *If r_θ, $\theta \in \Theta$, denotes the trimmed regression defined above and N_θ and D_θ are defined as in* (13)*, then, for any $t \in A(\theta)$,*

$$r_\theta(t) = \frac{N_\theta(t)}{D_\theta(t)}.$$

2. *If there exists $\theta_0 \in \Theta$ such that $E(Y|X) = E(Y|X\theta_0)$, then, for any $x \in A$,*

$$E^A(Y|X\theta_0 = x\theta_0) = m(x\theta_0).$$

Moreover, if θ_ψ is defined as in Assumption 5, $\theta_\psi = \theta_0$.

Proof :

1. It suffices to remark that

$$E^A(Y|X\theta)\, E(I_A(X) \mid X\theta) = E(Y\ I_A(X) \mid X\theta),$$

since, for any $B \in \sigma(X\theta)$,

$$\begin{aligned}
&\int_{X\theta\in B} E^A(Y|X\theta)\ E(I_A(X) \mid X\theta)\ dP \\
&= \int_{X\theta\in B} E\left(E^A(Y|X\theta)\ I_A(X) \mid X\theta\right) dP = \int_{X\theta\in B} E^A(Y|X\theta)\ I_A(X)\ dP \\
&= P(X \in A) \int_{X\theta\in B} E^A(Y|X\theta)\ dP^A = P(X \in A) \int_{X\theta\in B} Y\ dP^A \\
&= \int_{X\theta\in B} Y\ I_A(X)\ dP = \int_{X\theta\in B} E(Y\ I_A(X) \mid X\theta)\ dP.
\end{aligned}$$

2. From the definition of ψ we deduce that, for any y, the function $r \longmapsto \psi(y,r)$ is maximized for $r = y$. Thus, in order to deduce that θ_0 satisfies the optimum condition of Assumption 5, it suffices to show the first part, that is, for any $x \in A$,

$$E^A(Y|X\theta_0 = x\theta_0) = m(x\theta_0). \tag{17}$$

From 1, we have

$$E(Y\ I_A(X) \mid X\theta_0) = E^A(Y|X\theta_0)\, E(I_A(X) \mid X\theta_0).$$

On the other hand, under the SIM condition,

$$\begin{aligned}
E(Y\ I_A(X) \mid X\theta_0) &= E(E(Y\ I_A(X) \mid X) \mid X\theta_0) \\
&= E(m(X)\ I_A(X) \mid X \mid X\theta_0) \\
&= m(X\theta_0)\, E(I_A(X) \mid X\theta_0).
\end{aligned}$$

and thus, for any $x \in A$, we get (17). ■

2.2 The estimator of θ_ψ

In order to perform our semiparametric approximation of the regression function $m(\cdot)$ we estimate the nonparametric part by a classical kernel estimator.

Condition K The kernel $K : I\!R \longrightarrow I\!R$ is a differentiable symmetric positive function. Moreover, K and K' are of bounded variation and :

1. $\lim_{|u|\to\infty} |u\ K(u)| = 0$.
2. $\int |u|^3 K(u)\, du < \infty$; let $K_1 = \int u^2 K(u)\, du$.
3. $K_2 = \int K^2(u)\, du < \infty$.

For $a_n < b_n$, $n \geq 1$, two sequences of positive real numbers decreasing to zero, let us define $\mathcal{H}_n = \{h : a_n \leq h \leq b_n\}$ and

$$\left(\widehat{\theta}, \widehat{h}\right) = \underset{\theta\in\Theta, h\in\mathcal{H}_n}{\arg\max} \frac{1}{n}\sum_{i=1}^{n} \psi\left(Y_i, \hat{r}^i_{\theta,h}(X_i\theta)\right)\ I_A(X_i) \overset{not}{=} \underset{\theta\in\Theta, h\in\mathcal{H}_n}{\arg\max}\ \widehat{S}(\theta, h), \qquad (18)$$

where

$$\hat{r}^i_{\theta,h}(t) = \frac{\frac{1}{n-1}\sum_{j\neq i} Y_j\ K_h(t - X_j\theta)\ I_A(X_j)}{\frac{1}{n-1}\sum_{j\neq i} K_h(t - X_j\theta)\ I_A(X_j)} \overset{not}{=} \frac{\widehat{N}^i_{\theta,h}(t)}{\widehat{D}^i_{\theta,h}(t)}, \qquad t \in A(\theta),$$

denotes the one-leave-out version of the classic Nadaraya-Watson estimator of the trimmed regression function $r_\theta(\cdot) = E^A(Y|X\theta = \cdot)$. Here $K_h(\cdot)$ stands for $K(\cdot/h)/h$. The trimmed one-leave-one nonparametric estimators $\hat{r}^i_{\theta,h}$ allow us to decompose $\widehat{S}(\theta, h)$ in a sum of $U-$statistics. As mentioned above, the definition of $r_\theta(\cdot)$ renders these $U-$statistics degenerate.

Let us note that an adjustment of $\psi\left(Y_i, \hat{r}^i_{\theta,h}(X_i\theta)\right)$ could be necessary when $\hat{r}^i_{\theta,h}(X_i\theta)$ is outside the domain of $\psi(y, \cdot)$ (see, *e.g.*, Klein and Spady (1993) for the case of a binary response model). Under the assumptions we consider herein an asymptotically negligeable adjustment could be built and thus the convergence results do not change. For the sake of simplicity we do not consider this issue below.

3 Convergence results

3.1 The parametric part

Theorem 1 *Consider that Assumptions 1, 2.1, 2.2, 2.3 (with $k = 1$), 3, and 5 and Condition K are satisfied.*

a) Assume that $E(Y^2) < \infty$ and that $\widehat{\theta}$ is defined with $a_n = Cn^{-\delta}$, where $0 < \delta < \frac{1}{2}$ and $C > 0$, and $b_n \to 0$. Then $\widehat{\theta} \longrightarrow \theta_\psi$, in probability, as n tends to infinity.

b) Assume that $E(Y^{2m}) < \infty$, for some $m \geq 8$, and that $\widehat{\theta}$ is defined with $a_n = Cn^{-\delta}$, where $0 < \delta < \frac{1}{2} - \frac{2}{m}$ and $C > 0$, and $b_n \to 0$. Then $\widehat{\theta} \longrightarrow \theta_\psi$, almost surely, as n tends to infinity.

Proof : See appendix ■

Let us define

$$M_\psi = E\left[\partial_\theta \psi\left(Y_i, r_{\theta_\psi}\left(X_i\theta_\psi\right)\right)\ \left(\partial_\theta \psi\left(Y_i, r_{\theta_\psi}\left(X_i\theta_\psi\right)\right)\right)^T\ I_A\left(X_i\right)\right]$$

and

$$W_\psi = E\left[-\partial^2_{\theta\theta}\psi\left(Y_i, r_{\theta_\psi}\left(X_i\theta_\psi\right)\right)\ I_A\left(X_i\right)\right].$$

Note that

$$M_\psi = E\left[\left[C'\left(r_{\theta_\psi}\left(X_i\theta_\psi\right)\right)\right]^2 var\left(Y_i \mid X_i\right)\ \partial_\theta r_{\theta_\psi}\left(X_i\theta_\psi\right)\ \partial_\theta r_{\theta_\psi}\left(X_i\theta_\psi\right)^T\ I_A\left(X_i\right)\right]$$

and

$$W_\psi = E\left[C'\left(r_{\theta_\psi}\left(X_i\theta_\psi\right)\right)\ \partial_\theta r_{\theta_\psi}\left(X_i\theta_\psi\right)\ \partial_\theta r_{\theta_\psi}\left(X_i\theta_\psi\right)^T\ I_A\left(X_i\right)\right].$$

Let r'_{θ_ψ}, r''_{θ_ψ} and D'_{θ_ψ} denote the derivatives of the functions $t \to r_{\theta_\psi}(t)$ and $t \to D_{\theta_\psi}(t)$, respectively. Define

$$\begin{aligned} A_1 &= \frac{K_1^2}{4}\, E\left\{\frac{1}{2}\,\partial^2_{22}\psi\left(r_{\theta_\psi}\left(X_i\theta_\psi\right), r_{\theta_\psi}\left(X_i\theta_\psi\right)\right)\right. \\ &\qquad \left.\times\left[r''_{\theta_\psi}\left(X_i\theta_\psi\right) + \frac{2\ r'_{\theta_\psi}\left(X_i\theta_\psi\right)\ D'_{\theta_\psi}\left(X_i\theta_\psi\right)}{D_{\theta_\psi}\left(X_i\theta_\psi\right)}\right]^2\ I_A\left(X_i\right)\right\} \qquad (19) \\ A_2 &= K_2\, E\left\{\frac{1}{2}\,\partial^2_{22}\psi\left(r_{\theta_\psi}\left(X_i\theta_\psi\right), r_{\theta_\psi}\left(X_i\theta_\psi\right)\right)\ \frac{1}{D_{\theta_\psi}\left(X_i\theta_\psi\right)}\ v_{\theta_\psi}\left(X_i\theta_\psi\right)\ I_A\left(X_i\right)\right\}, \end{aligned}$$

and

$$h_n = \arg\max_h \left(A_1 h^4 + A_2 \frac{1}{nh}\right) = \left(A_2/4A_1\right)^{1/5} n^{-1/5}.$$

Note that in our framewok $\partial^2_{22}\psi(r,r) = -C'(r) < 0$. The usual L_2-MISE corresponds to $\partial^2_{22}\psi(r,r) = -2$.

In the proof of the following theorem, we need to extend the definition of A_1 and A_2 to θ in a neighborhood of θ_ψ. More precisely, define $A_1(\theta)$ and $A_2(\theta)$ as in (19) with θ_ψ replaced by θ.

Theorem 2 *Consider that Assumptions 1 to 5 (Assumption 2.3 with $k = 2$) and Condition K are satisfied. Moreover, $A_1(\cdot)$ and $A_2(\cdot)$ are Lipschitz in a neighborhood of θ_ψ. Assume that $E(Y^4) < \infty$. Let $\left(\widehat{\theta}, \widehat{h}\right)$ be defined as in (18) for $a_n = Cn^{-1/2+\varepsilon}$ with $C > 0$ and $0 < \varepsilon < 3/10$ and $b_n = 1/\ln n$. Then*

$$\frac{\widehat{h}}{h_n} \xrightarrow{P} 1$$

and

$$\sqrt{n}\left(\widehat{\theta} - \theta_\psi\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, W_\psi^- M_\psi W_\psi^-\right)$$

where W_ψ^- denotes a generalized inverse of W_ψ .

Sketch of the proof : The complete arguments are given in the appendix. Let us present below the lines of the proof. The idea is to generalize the decomposition used by Härdle, Hall and Ichimura (1993). We write

$$\begin{aligned}\widehat{S}(\theta,h) &= \frac{1}{n}\sum_{i=1}^{n}\psi\left(Y_i,\hat{r}^i_{\theta,h}(X_i\theta)\right)\ I_A(X_i)\\ &= \widetilde{S}(\theta)+T(h)+R_1(\theta,h)+R_2(h),\end{aligned}$$

where

$$\begin{aligned}\widetilde{S}(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\underline{\psi\left(Y_i,r_\theta(X_i\theta)\right)}\ I_A(X_i)-\frac{1}{n}\sum_{i=1}^{n}\underline{\underline{\psi\left(Y_i,r_{\theta_\psi}(X_i\theta_\psi)\right)}}\ I_A(X_i),\\ T(h) &= \frac{1}{n}\sum_{i=1}^{n}\underline{\underline{\psi\left(r_{\theta_\psi}(X_i\theta_\psi),\hat{r}^i_{\theta_\psi,h}(X_i\theta_\psi)\right)}}\ I_A(X_i),\\ R_1(\theta,h) &= \frac{1}{n}\sum_{i=1}^{n}\left[\psi\left(Y_i,\hat{r}^i_{\theta,h}(X_i\theta)\right)-\underline{\psi\left(Y_i,r_\theta(X_i\theta)\right)}\right]\ I_A(X_i)\\ &\quad -\frac{1}{n}\sum_{i=1}^{n}\left[\underset{\longleftrightarrow}{\psi\left(Y_i,\hat{r}^i_{\theta_\psi,h}(X_i\theta_\psi)\right)}-\underline{\underline{\psi\left(Y_i,r_{\theta_\psi}(X_i\theta_\psi)\right)}}\right]\ I_A(X_i),\\ R_2(h) &= \frac{1}{n}\sum_{i=1}^{n}\left[\underset{\longleftrightarrow}{\psi\left(Y_i,\hat{r}^i_{\theta_\psi,h}(X_i\theta_\psi)\right)}-\underline{\underline{\psi\left(r_{\theta_\psi}(X_i\theta_\psi),\hat{r}^i_{\theta_\psi,h}(X_i\theta_\psi)\right)}}\right]\ I_A(X_i).\end{aligned}$$

Next we show that maximizing $\widehat{S}(\theta,h)$ simultaneously with respect to both θ and h is asymptotically equivalent with separately maximizing $\widetilde{S}(\theta)$ with respect to θ and $T(h)$ with respect to h. More precisely, in our proof we follow the steps described below.

Step 1 Show that

$$\begin{aligned}T(h) &= A_1h^4+A_2n^{-1}h^{-1}+O_P\left(n^{-1/2}h^2\right)+o_P\left(n^{-1}h^{-1}\right)+o_P\left(h^4\right)\\ &\quad +\{\text{terms independent of } h\}.\end{aligned}$$

In our framework $-T(h)$ is an estimator of $-J(h)$ that we call the $\psi-$MISE and which is defined as

$$-J(h)=E\left[-\psi\left(r_{\theta_\psi}(X_i\theta_\psi),\hat{r}^i_{\theta\psi,h}(X_i\theta_\psi)\right)I_A(X_i)\right]=\int-\psi\left(r_{\theta\psi}(t),\hat{r}^i_{\theta_\psi,h}(t)\right)D_{\theta\psi}(t)\ dt.$$

We show in the appendix that

$$J(h)=h^4A_1+n^{-1}h^{-1}A_2+o_P\left(h^4A_1+n^{-1}h^{-1}A_2\right)$$

and thus h_n is nothing else the theoretical optimal bandwidth obtained by minimizing the theoretical $\psi-$MISE criterion $-J(h)$.

Step 2 In order to deduce $\widehat{h}/h_n\to 1$ in probability, it remains to show that the other two terms of our decomposition containing h are asymptotically negligeable when compared with $T(h)$. Therefore, we show that

$$R_2(h) = o_P\left(n^{-1}h^{-1}\right) + O_P\left(n^{-1/2}h^2\right) + \{\text{terms independent of } h\}$$

and

$$R_1(\theta, h) = o_P\left(h^4\right) + o_P\left(n^{-1}h^{-1}\right)$$

uniformly over $o_P(1)$ neighborhoods of θ_ψ.

Step 3 Using a classical method for proving asymptotic normality as presented, *e.g.,* in Sherman (1994a), p 453, Theorem 1 and 2, we deduce first, the $\sqrt{n}-$convergence of $\widehat{\theta}$ using a two steps argument, and afterwards its limit distribution. With Sherman's notation, consider

$$\Gamma_n(\theta) = \widetilde{S}(\theta) + R_1(\theta, h)$$

and

$$\Gamma(\theta) = -\frac{1}{2}(\theta - \theta\psi)^T W_\psi(\theta - \theta\psi).$$

From Step 2 and a version of Sherman's Theorem 1 we obtain a preliminary rate of $O_P(n^{-2/5})$ for $\widehat{\theta}$. This allows us to restrict θ to $O_P(n^{-2/5})$ neighborhoods of θ_ψ. Next, we show that

$$R_1(\theta, h) = o_P\left(\|\theta - \theta_\psi\| / \sqrt{n}\right) + o_P\left(n^{-1}\right),$$

uniformly over $O_p(n^{-2/5})$ neighborhoods of θ_ψ and with respect to $O_p(n^{-1/5})$ ranges for the bandwidth. Finally, the second part of our results is a consequence of Sherman's Theorem 1 and 2. ■

Note that the asymptotic distribution of $\widehat{\theta}$ does not depend on the choice of the nonparametric estimator of the regression function $r_{\theta_\psi}(\cdot)$. We argue that our methodology could be applied for other linear smoothers than the classic kernel estimator. For instance, in the definition of $\left(\widehat{\theta}, \widehat{h}\right)$, one may replace $\widehat{r}^{\,i}_{\theta,\,h}(x\theta)$ by the local linear estimator

$$\widetilde{r}^{\,i}_{\theta,\,h}(x\theta) \;=\; \widetilde{a}^i$$

with $\widetilde{a}$ verifying

$$\left(\widetilde{a}^i, \widetilde{b}^i\right) \;= \arg\min_{a,b} \sum_{j \neq i} K_h\left(X_j\theta - x\theta\right)\left[Y_j - a - b\left(X_j\theta - x\theta\right)\right]^2.$$

The same kind of decompositions could be used and the corresponding asymptotic results could be deduced.

3.2 The nonparametric part

For any $x \in A$, consider

$$\widehat{r}_{\widehat{\theta},\,\widehat{h}}\left(x\widehat{\theta}\right) = \frac{\sum_{i=1}^{n} Y_i \; K_{\widehat{h}}\left(x\widehat{\theta} - X_i\widehat{\theta}\right) \; I_A(X_i)}{\sum_{i=1}^{n} K_{\widehat{h}}\left(x\widehat{\theta} - X_i\widehat{\theta}\right) \; I_A(X_i)}$$

the kernel-based estimator of

$$r_{\theta_\psi}(x\theta_\psi) = E^A(Y \mid X\theta_\psi = x\,\theta_\psi).$$

The asymptotic bias and variance of this estimator are obtained from the fact that $\widehat{r}_{\widehat{\theta},\widehat{h}}\left(x\widehat{\theta}\right)$ behaves very much like $\widehat{r}_{\theta_\psi,\widehat{h}}(x\theta_\psi)$, the theoretical kernel estimator (with bandwidth $\widehat{h}$) of $r_{\theta_\psi}(x\theta_\psi)$. More precisely,

$$\widehat{r}_{\widehat{\theta},\widehat{h}}\left(x\widehat{\theta}_n\right) - \widehat{r}_{\theta_\psi,\widehat{h}}(x\theta_\psi) = \partial_\theta \widehat{r}_{\theta_\psi,\widehat{h}}(x\theta_\psi)\left(\widehat{\theta}-\theta_\psi\right) + O_P\left(\left\|\widehat{\theta}-\theta_\psi\right\|^2\right)$$

with $\partial_\theta \widehat{r}_{\theta_\psi,\widehat{h}}(x\theta_\psi) = O_P(1)$ and $\widehat{\theta}-\theta_\psi = O_P\left(n^{-1/2}\right)$. Classic results on the asymptotic distribution of the kernel estimators (see, *e.g.* Bosq and Lecoutre (1987), Chapter 5) allows us to state the following result.

Corollary 1 *Assume that the conditions of Theorem 3 are fulfilled and that $v_{\theta_\psi}(\cdot)$ defined in (15) is continuous. Then, for any $x \in A$,*

$$\sqrt{n\widehat{h}}\left(\widehat{r}_{\widehat{\theta},\widehat{h}}\left(x\widehat{\theta}\right) - r_{\theta_\psi}(x\theta_\psi) - \widehat{h}^2\beta(x\theta_\psi)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, V(x\theta_\psi))$$

where

$$\beta(t) = (K_1/2)\left[r''_{\theta_\psi}(t) + 2r'_{\theta_\psi}(t)\, f'_{\theta_\psi}(t) \,/\, f_{\theta_\psi}(t)\right]$$

and

$$V(t) = K_2\, v_{\theta_\psi}(t) \,/\, f_{\theta_\psi}(t).$$

4 Appendix

4.1 Appendix 1 : rates of uniform convergence for degenerate $U-$processes

In this section we briefly recall definitions and results from Sherman (1994a) (see also Serfling (1980)) which we use as main tools for our proofs.

Let $Z_1, ..., Z_n$ be independent copies of a random vector Z with distribution P on a set S. Let k be a positive integer, Λ a subset of $\mathbb{R}^m$ $(m \geq 1)$ and S^k the product space $S \otimes ... \otimes S$ (k factors). For each positive integer n and each $\lambda \in \Lambda$, let $f_n(\cdot;\lambda)$ denote a real-valued function on S^k, also called the kernel. Define the corresponding $U-$statistic of order k

$$U_n^k f_n(\cdot;\lambda) = (n)_k^{-1} \sum_{i_j \neq i_l,\, j \neq l} f_n(Z_{i_1}, ..., Z_{i_k};\lambda),$$

where $(n)_k = n(n-1)...(n-k+1)$. In the present framework, allowing the function $f_n(\cdot;\lambda)$ to depend on the sample size is crucial when we have to consider θ in a neighborhood shrinking to a limit point at a suitable speed. Moreover, this dependence could be also useful if one considers a sequence of growing trimming sets $A_n, n \geq 1$, in the space of the explanatory variables. For a simplicity, we omit to write the dependence of this function on n but we consider it implicitly.

The collection $\left\{U_n^k f(\cdot;\lambda),\ \lambda \in \Lambda\right\}$ is called a $U-$process indexed by Λ. Note that U_n^1 is nothing else than the empirical measure P_n that places mass n^{-1} at each Z_i; therefore below we rather write P_n^1 instead of U_n^1. Let $s^k = (s_1, ..., s_k)$ denote an element of S^k. If for each $\lambda \in \Lambda$ and each $s^k \in S^k$,

$$E\left[f\left(s_1, ..., s_{i-1}, Z, s_{i+1}, ..., s_k; \lambda\right)\right] \equiv 0, \qquad i = 1, ..., k,$$

then $\mathcal{F} = \{f(\cdot;\lambda),\ \lambda \in \Lambda\}$ is called a $P-$degenerate (or simply degenerate) class of functions on S^k. Moreover, $U_n^k f(\cdot;\lambda)$ is called a degenerate $U-$statistics of order k and $\left\{U_n^k f(\cdot;\lambda),\ \lambda \in \Lambda\right\}$ is called a degenerate $U-$process of order k.

We say that $F(\cdot)$, defined on S^k, is a (pointwise) envelope for $\mathcal{F} = \{f(\cdot;\lambda),\ \lambda \in \Lambda\}$ if

$$\sup_{\lambda \in \Lambda} |f(\cdot;\lambda)| \leq F(\cdot).$$

Sherman (1994a) states uniform convergence results for degenerate $U-$processes corresponding to classes of functions $\mathcal{F}$ satisfying the so-called *Euclidean condition.*

DEFINITION (*see Sherman (1994a), p 447*) Let $\mathcal{F}$ be a class of real-valued functions on a set $\mathcal{X}$. Call $\mathcal{F}$ Euclidean for the envelope F if there exists positive constants A and V with the following property : if μ is a measure for which $\int F^2 d\mu < \infty$, then

$$D(x, d_\mu, \mathcal{F}) \leq Ax^{-V}, \qquad 0 < x \leq 1,$$

where, for $f, g \in \mathcal{F}$,

$$d_\mu(f,g) = \left[\int |f-g|^2 \, d\mu \,/ \int F^2 d\mu\right]^{1/2}$$

and $D(x, d_\mu, \mathcal{F})$ denotes the packing number, that is the largest number D for which there exists $f_1, ..., f_D$ in $\mathcal{F}$ such that

$$d_\mu(f_i, f_j) > x \qquad \text{for } i \neq j.$$

Note that the constants A and V do not depend on μ. The criteria for determinating the Euclidean property we use in the proofs are summarized in the following proposition.

Proposition 1

1. *(see Nolan and Pollard (1987), Corollary 21) Let $\mathcal{F}$ be a uniformly bounded Euclidean class of functions on $\mathcal{X} \otimes \mathcal{X}$. Then, for each finite measure ν on $\mathcal{X}$, the class $\left\{\int f(x,\cdot)\, d\nu(x),\ f \in \mathcal{F}\right\}$ is Euclidean (for a constant envelope).*

2. *(see Nolan and Pollard (1987), Lemma 22(ii)) Let $g(\cdot)$ be a real-valued function of bounded variation on $\mathbb{R}$. The class of all functions on $\mathbb{R}^d$ of the form $z \to g(\alpha z + \beta)$, with α ranging over $\mathbb{R}^d$ and β ranging over $\mathbb{R}$, is Euclidean for a constant envelope (αz denoed the usual inner product of α and z in $\mathbb{R}^d$).*

3. *(see Pakes and Pollard (1989), Lemma 2.13) Let $\mathcal{F} = \{f(\cdot;\lambda),\ \lambda \in \Lambda\}$ be a class of functions on $\mathcal{X}$ indexed by a bounded subset Λ of $\mathbb{R}^m$. If there exists an $\alpha > 0$ and a non*negative *function $\phi(\cdot)$ such that*

$$|f(x,\lambda) - f(x,\lambda')| \leq \phi(x)\,\|\lambda - \lambda'\|^{\alpha} \qquad \text{for } x \in \mathcal{X} \quad \text{and } \lambda, \lambda' \in \Lambda,$$

then $\mathcal{F}$ is Euclidean for the envelope $|f(\cdot,\lambda_0)| + M\phi(\cdot)$, where λ_0 is an arbitrary point of Λ and $M = (2\sqrt{m}\sup_{\Lambda}\|\lambda - \lambda_0\|)^{\alpha}$.

4. *(see Pakes and Pollard (1989), Lemma 2.14(i) and (ii)) If $\mathcal{F}$ is Euclidean for the envelope F, and $\mathcal{G}$ is Euclidean for the envelope G, then $\{f+g,\ f \in \mathcal{F},\ g \in \mathcal{G}\}$ is Euclidean for the envelope $F+G$ and $\{fg,\ f \in \mathcal{F},\ g \in \mathcal{G}\}$ is Euclidean for the envelope FG.*

Finally, we recall the results on the rates of convergence of degenerate $U-$processes that are used in the proofs.

Proposition 2 *(rates of convergence in probability) Let $\mathcal{F}$ be a class of $P-$degenerate functions on S^k, $k \geq 1$, and $P^k = P \otimes ... \otimes P$ (k factors). Suppose $\mathcal{F}$ is Euclidean for a squared integrable envelope F, that is $\int F^2 dP^k < \infty$.*

1. *(see Sherman (1994a), Corollary 4(ii)) Then*

$$\sup_{f \in \mathcal{F}} \left| n^{k/2} U_n^k f \right| = O_P(1).$$

2. *(see Sherman (1994a), see Corollary 8) Suppose $\mathcal{F} = \{f(\cdot;\lambda),\ \lambda \in \Lambda\}$, and that λ_0 is a point of Λ for which $f(\cdot;\lambda_0) \equiv 0$. If $\int |f(\cdot;\lambda)|\, dP^k \to 0$ as $\lambda \to \lambda_0$, then uniformly over $o_P(1)$ neighborhoods of λ_0,*

$$U_n^k f(\cdot;\lambda) = o_P(n^{-k/2}).$$

Proposition 3 *(rates of almost-sure convergence; see Sherman (1994a), Corollary 9) Let $\mathcal{F}$ be a class of $P-$degenerate functions on S^k, $k \geq 1$, Euclidean for an envelope F. For real numbers $\delta > 0$ and $\beta > 1$, let p be a positive integer satisfying $p \geq \beta/\delta$. If $\int F^{4p} dP^k < \infty$, then*

$$\sup_{f \in \mathcal{F}} \left| n^{k/2-\delta} U_n^k f \right| \to 0,$$

almost surely as n tends to infinity.

4.2 Appendix 2 : proofs

Proof of Theorem 2 :

For any $\theta \in \Theta$ and $t \in A(\Theta)$, let

$$r_{\theta,h}(t) = \frac{N_{\theta,h}(t)}{D_{\theta,h}(t)},$$

where

$$\begin{aligned} N_{\theta,h}(t) &= E\left[Y\ K_h(X\theta - t)\, I_A(X)\right], \\ D_{\theta,h}(t) &= E\left[K_h(X\theta - t)\, I_A(X)\right]. \end{aligned}$$

Asking $b_n \to 0$ ensure that θ_ψ is the unique optimum of the limit problem. We only have to prove that, under the stated assumptions,

$$\sup_{\theta\in\Theta, h\in\mathcal{H}_n} \left|\widehat{S}(\theta,h) - E\left(\psi\left(Y, r_\theta(X\theta)\right) I_A(X)\right)\right| \to 0, \tag{20}$$

in probability and almost surely, respectively. For any $\theta \in \Theta$ and $h \in \mathcal{H}_n$, we can write

$$\begin{aligned} & \widehat{S}(\theta,h) - E\left(\psi\left(Y, r_\theta(X\theta)\right) I_A(X)\right) \\ = & \frac{1}{n}\sum_{i=1}^{n}\left[\psi\left(Y_i, \hat{r}^i_{\theta,h}(X_i\theta)\right) I_A(X_i) \ - \psi\left(Y_i, r_{\theta,h}(X_i\theta)\right) I_A(X_i)\right] \\ & + \frac{1}{n}\sum_{i=1}^{n}\left[\psi\left(Y_i, r_{\theta,h}(X_i\theta)\right) I_A(X_i) \ - \psi\left(Y_i, r_\theta(X_i\theta)\right) I_A(X_i)\ \right] \\ & + \frac{1}{n}\sum_{i=1}^{n}\left[\psi\left(Y_i, r_\theta(X_i\theta)\right) I_A(X_i) \ - E\left(\psi\left(Y_i, r_\theta(X_i\theta)\right) I_A(X_i)\right)\right] \\ & \overset{not}{=} \widehat{S}_1(\theta,h) + \widehat{S}_2(\theta,h) + \widehat{S}_3(\theta). \end{aligned}$$

The almost sure convergence of $\widehat{S}_3(\theta)$ uniformly in θ, can be obtained from a strong uniform law of large numbers as in Pollard (1984), chapter II; see also Pakes and Pollard (1989), Lemma 2.8. Consider the family

$$\mathcal{F}_3 = \left\{(x,y) \to \psi\left(y, r_\theta(x\theta)\right) I_A(x), \quad \theta \in \Theta\right\}$$

which, under the stated assumptions, admits an envelope $F_3(x,y) = C_3(|y|+1)$, for some $C_3 > 0$. Moreover, it is easy to see that $\mathcal{F}_3$ is Euclidean for the envelope F (see Lemma 2.13 Pakes and Pollard (1989)). Thus,

$$\sup_{\theta\in\Theta}\left|\widehat{S}_3(\theta)\right| \to 0,$$

almost surely.

For the uniform convergence of $\widehat{S}_2(\theta,h)$, let us note that

$$\sup_{\theta\in\Theta,\, t\in A(\theta),\, h\in\mathcal{H}_n} \left|N_{\theta,h}(t) - N_\theta(t)\right| \to 0$$

$$\sup_{\theta\in\Theta,\, t\in A(\theta),\, h\in\mathcal{H}_n} \left|D_{\theta,h}(t) - D_\theta(t)\right| \to 0.$$

and thus

$$\sup_{\theta\in\Theta,\, t\in A(\theta),\, h\in\mathcal{H}_n} \left|r_{\theta,h}(t) - r_\theta(t)\right| \to 0.$$

Indeed,

$$\begin{aligned}
\sup_{\theta,\,t,\,h} \left|N_{\theta,h}(t) - N_\theta(t)\right| &= \sup_{\theta,\,t,\,h} \left|\int K(u)\left[N_\theta(t+uh) - N_\theta(t)\right] du\right| \\
&\leq \int K(u)\left(\sup_{\theta,\,t,\,h} \left|N_\theta(t+uh) - N_\theta(t)\right|\right) du
\end{aligned}$$

and the last integral converges to zero due to the uniform continuity of $N_\theta(t)$ as function of θ and t. The same arguments prove the uniform convergence of $D_{\theta,h}(t)$. Since

$$\begin{aligned}
\left|\widehat{S}_2(\theta,h)\right| &= \left|\frac{1}{n}\sum_{i=1}^{n}\left[\psi\left(Y_i, r_{\theta,h}(X_i\theta)\right) - \psi\left(Y_i, r_\theta(X_i\theta)\right)\right] I_A(X_i)\right| \\
&= \left|\frac{1}{n}\sum_{i=1}^{n}\partial_2\psi\left(Y_i, \overline{r}^i_{\theta,h}\right)\left[r_{\theta,h}(X_i\theta) - r_\theta(X_i\theta)\right] I_A(X_i)\right| \\
&\leq \left(\sup_{\theta,\,t,\,h}\left|r_{\theta,h}(t) - r_\theta(t)\right|\right)\frac{1}{n}\sum_{i=1}^{n}\left|\partial_2\psi\left(Y_i, \overline{r}_{\theta,h}\right)\right| I_A(X_i) \\
&\leq \left(\sup_{\theta,\,t,\,h}\left|r_{\theta,h}(t) - r_\theta(t)\right|\right)\frac{1}{n}\sum_{i=1}^{n} C_2\left(|Y_i|+1\right)
\end{aligned}$$

where $\overline{r}^i_{\theta,h}$ lies between $r_{\theta,h}(X_i\theta)$ and $r_\theta(X_i\theta)$ and C_2 is some positive constant, we deduce

$$\sup_{\theta\in\Theta,\, h\in\mathcal{H}_n}\left|\widehat{S}_2(\theta,h)\right| \to 0,$$

almost surely.

Finally, we may write

$$\begin{aligned}
\left|\widehat{S}_1(\theta,h)\right| &= \left|\frac{1}{n}\sum_{i=1}^{n}\left[\psi\left(Y_i, \hat{r}^i_{\theta,h}(X_i\theta)\right) - \psi\left(Y_i, r_{\theta,h}(X_i\theta)\right)\right] I_A(X_i)\right| \\
&= \left|\frac{1}{n}\sum_{i=1}^{n}\partial_2\psi\left(Y_i, \overline{\hat{r}}^i_{\theta,h}\right)\left[\hat{r}^i_{\theta,h}(X_i\theta) - r_{\theta,h}(X_i\theta)\right] I_A(X_i)\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\left|\partial_2\psi\left(Y_i, \overline{\hat{r}}^i_{\theta,h}\right)\right|\left(\sup_{\theta,\,t,\,h}\left|\widehat{r}^i_{\theta,h}(t) - r_{\theta,h}(t)\right|\right) I_A(X_i)
\end{aligned}$$

where $\overline{\hat{r}}^i_{\theta,h}$ is between $\hat{r}^i_{\theta,h}(X_i\theta)$ and $r_{\theta,h}(X_i\theta)$. Now, the idea is to show the uniform convergence of the one-leave-out kernel estimator $\widehat{r}^i_{\theta,h}$. Since

$$\widehat{r}^i_{\theta,h}(X_i\theta) - r_{\theta,h}(X_i\theta) = \frac{\widehat{N}^i_{\theta,h}(X_i\theta)}{\widehat{D}^i_{\theta,h}(X_i\theta)} - \frac{N_{\theta,h}(X_i\theta)}{D_{\theta,h}(X_i\theta)}$$

$$= \frac{1}{\widehat{D}^i_{\theta,h}(X_i\theta)} \left\{ \left[\widehat{N}^i_{\theta,h}(X_i\theta) - N_{\theta,h}(X_i\theta) \right] - r_{\theta,h}(X_i\theta) \left[\widehat{D}^i_{\theta,h}(X_i\theta) - D_{\theta,h}(X_i\theta) \right] \right\},$$

it is sufficient to prove that

$$\sup_{\substack{\theta,\,t,\,h \\ 1\le i\le n}} \left| \widehat{D}^i_{\theta,h}(t) - D_{\theta,h}(t) \right| \to 0 \tag{21}$$

and

$$\sup_{\substack{\theta,\,t,\,h \\ 1\le i\le n}} \left| \widehat{N}^i_{\theta,h}(t) - N_{\theta,h}(t) \right| \to 0, \tag{22}$$

in probability and almost surely, respectively. We have

$$\begin{aligned}
& \frac{n-1}{n} \left| \widehat{N}^i_{\theta,h}(t) - N_{\theta,h}(t) \right| \\
= & \left| \frac{1}{n} \sum_{j\neq i} \left\{ Y_j \ \frac{1}{h} \ K\left(\frac{t - X_j\theta}{h} \right) \ I_A(X_j) - E\left[Y_j \ K_h(t - X_j\theta) \ I_A(X_j) \right] \right\} \right| \\
\le & \frac{1}{h} \left| \int f_{\theta,h,t}\, dP_n - \int f_{\theta,h,t}\, dP \right| \\
& + \frac{1}{h}\frac{1}{n} \left| Y_i \ K\left(\frac{t - X_i\theta}{h} \right) \ I_A(X_i) - E\left[Y_i \ K\left((t - X_i\theta)/h \right) \ I_A(X_i) \right] \right| \\
\le & \frac{1}{a_n} \left| \int f_{\theta,h,t}\, dP_n - \int f_{\theta,h,t}\, dP \right| + \frac{1}{n a_n} g(X_i, Y_i),
\end{aligned}$$

where P_n denotes the empirical probability which places mass n^{-1} to each observation (X, Y) sampled from P,

$$f_{\theta,h,t}(x,y) = y \ K\left((t - x\theta)/h \right) \ I_A(x), \quad \theta \in \Theta, \ t \in A(\theta), \ h \in \mathcal{H}_n,$$

and $g(\cdot,\cdot)$ is a real-valued function with the absolute value bounded by $C(|Y_i| + 1)$, for some $C > 0$. Consider the family

$$\mathcal{F}_1 = \{ f_{\theta,h,t}(\cdot,\cdot), \quad \theta \in \Theta, \ t \in A(\theta) \ \ h \in (0,1] \}$$

and note that

$$\sup_{\theta,\,t,\,h} \left| \int f_{\theta,h}\, dP_n - \int f_{\theta,h}\, dP \right| \le \sup_{\mathcal{F}_1} \left| \int f_{\theta,h,t}\, dP_n - \int f_{\theta,h,t}\, dP \right|$$

and that

$$\sup_{\mathcal{F}_1} |f_{\theta,h,t}\ (x,y)| \le F_1(x,y) = C_1 \ \ (|y| + 1),$$

for some $C_1 > 0$. Under the assumptions of part a), the envelope F_1 has a finite moment of order two. It is easy to see that $\mathcal{F}_1$ is Euclidean for the envelope F_1 (use the bounded variation of K and apply Lemma 22(ii) of Nolan and Pollard (1987) for $\alpha = h^{-1}\theta$ and $\beta = h^{-1}t$). From Sherman (1994a), Corollary 4(ii) with $k = 1$, we get

$$\begin{aligned}
\sup_{\substack{\theta,\,t,\,h \\ 1\le i\le n}} \left| \widehat{N}^i_{\theta,h}(t) - N_{\theta,h}(t) \right| & \le \frac{n}{a_n(n-1)} \sup_{\mathcal{F}_1} \left| \int f_{\theta,h,t}\, dP_n - \int f_{\theta,h,t}\, dP \right| + O_P\left(a_n^{-1} n^{-1} \right) \\
& = o_P\left(a_n^{-1} n^{-(1/2-\gamma)} \right),
\end{aligned}$$

for any $\gamma > 0$. On the other hand, under the assumptions of b) and for some $m \geq 8$, we apply Corollary 9 of Sherman (1994a) with $m = 2p$, $\delta = 2\beta/m$, $\beta > 1$, and we obtain,

$$\sup_{\substack{\theta,\, t,\, h \\ 1 \leq i \leq n}} \left| \widehat{N}^i_{\theta,h}(t) - N_{\theta,h}(t) \right| = o\left(a_n^{-1} n^{-\left(\frac{1}{2} - \frac{2}{m} - \gamma\right)} \right),$$

almost surely, where $\gamma = 2(\beta - 1)/m > 0$. For $Y \equiv 1$ we obtain (21). Thus

$$\sup_{\theta \in \Theta,\, h \in \mathcal{H}_n} \left| \widehat{S}_2(\theta, h) \right| \to 0,$$

in probability and almost surely, respectively. Finally, we may conclude that (20) is verified and the proof is closed. ■

Proof of Theorem 3 :

STEP 1 : the order of $T(h)$ Let us introduce the following simplified notation: I_i stands for $I(X_i)$ while r and $\hat{r}^i$ replace r_{θ_ψ} and $\hat{r}^i_{\theta_\psi,h}$, respectively. Moreover, $V_i = X_i \theta_\psi$. We have

$$\begin{aligned}
T(h) &= \frac{1}{n}\sum_{i=1}^{n} \psi\left(r(V_i), \hat{r}^i(V_i)\right) I_i \\
&= \frac{1}{n}\sum_{i=1}^{n} \psi\left(r(V_i), r(V_i)\right) I_i \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} \partial_2\psi\left(r(V_i), r(V_i)\right)\left[\hat{r}^i(V_i) - r(V_i)\right] I_i \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2}\,\partial^2_{22}\psi\left(r(V_i), r(V_i)\right)\left[\hat{r}^i(V_i) - r(V_i)\right]^2 I_i \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2}\left[\partial^2_{22}\psi\left(r(V_i), \overline{r}^i\right) - \partial^2_{22}\psi\left(r(V_i), r(V_i)\right)\right] \\
&\qquad\qquad \times \left[\hat{r}^i(V_i) - r(V_i)\right]^2 I_i \\
&\quad \stackrel{not}{=} T_0 + T_1(h) + T_2(h) + T_3(h),
\end{aligned}$$

with $\overline{r}^i$ between $r(V_i)$ and $\hat{r}^i(V_i)$. Since $\partial_2\psi(r, r) \equiv 0$, we have $T_1(h) \equiv 0$. Using the Lipschitz condition verified by $r \to \partial^2_{22}\psi(\cdot, r)$, it is easy to see that $T_3(h)$ has a smaller order than $T_2(h)$.

The order of $T_2(h)$ Let us further simplify our notation : when there is no possible confusion we drop the argument V_i; we write $\widehat{r}^i = \widehat{N}^i / \widehat{D}^i$ and $r = N/D$. We have

$$T_2(h) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2}\,\partial^2_{22}\psi\left(\widehat{r}^i - r\right)^2 I_i$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2}\,\partial_{22}^{2}\psi\ \left(\widehat{N}^{\,i} - r\widehat{D}^{\,i}\right)^{2}\ \frac{1}{\widehat{D}^{\,i\,2}}\ I_i \\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2}\,\partial_{22}^{2}\psi\ \left(\widehat{N}^{\,i} - r\widehat{D}^{\,i}\right)^{2}\ \left[\frac{1}{D^2} - \frac{2}{\overline{D}^{\,i\,3}}\left(\widehat{D}^{\,i} - D\right)\right]\ I_i \\
&= \frac{1}{n}\sum_{i=1}^{n} a_i\ \left(\widehat{N}^{\,i} - r\widehat{D}^{\,i}\right)^{2}\ I_i - \frac{1}{n}\sum_{i=1}^{n} a_i\ \left(\widehat{N}^{\,i} - r\widehat{D}^{\,i}\right)^{2}\ b_{in}\ I_i,
\end{aligned}
$$

with $\overline{D}^{i}$ between $\widehat{D}^{\,i}$ and D, $a_i = 2\,D^{-2}\partial_{22}^{2}\psi$ and $b_{in} = 2\left(\overline{D}^{i}\right)^{-3}\left(\widehat{D}^{\,i} - D\right)$. It is clear that $|b_{in}| \leq B_n$ with $B_n = o_P(1)$ (see also the proof of Theorem 2). Using the definition of $\widehat{N}^{\,i}$ and $\widehat{D}^{\,i}$ we get

$$
T_2(h) = \frac{n-2}{n-1}\,T_{21}(h) + \frac{1}{n-1}\,T_{22}(h) + \{\text{terms of smaller order}\}
$$

where

$$
\begin{aligned}
T_{21}(h) &= \frac{1}{n(n-1)(n-2)}\sum_{i\neq j\neq l} a_i\ [Y_j - r(V_i)]\ [Y_l - r(V_i)] \\
&\qquad \times K_h(V_i - V_j)\ K_h(V_i - V_l)\ I_i\ I_j\ I_l \\
T_{22}(h) &= \frac{1}{n(n-1)}\sum_{i\neq j} a_i\ [Y_j - r(V_i)]^2\ K_h^2(V_i - V_j)\ I_i\, I_j\ .
\end{aligned}
$$

The order of $T_{21}(h)$ Using an usual decomposition of a $U-$statistics of order k in *degenerate* $U-$statistics of order $s \leq k$ (see Serfling(1980), pages 177-178), we may write

$$
\begin{aligned}
T_{21}(h) &= U_n^3 f_{21}(\cdot,\cdot,\cdot) \\
&= (n)_3^{-1}\sum_{i\neq j\neq l}[f_{21}(Z_i,Z_j,Z_l) - E(f_{21}|Z_i,Z_j) - E(f_{21}|Z_i,Z_l) - E(f_{21}|Z_j,Z_l) \\
&\qquad +E(f_{21}|Z_i) + E(f_{21}|Z_j) + E(f_{21}|Z_l) - E(f_{21})] \\
&\quad + \left\{(n)_2^{-1}\left[\sum_{i\neq j} E(f_{21}|Z_i,Z_j) + \sum_{i\neq l} E(f_{21}|Z_i,Z_l) + \sum_{j\neq l} E(f_{21}|Z_j,Z_l)\right]\right. \\
&\qquad \left. - (n)_1^{-1}\left(\sum_{i=1}^{n} E(f_{21}|Z_i) + \sum_{j=1}^{n} E(f_{21}|Z_j) + \sum_{l=1}^{n} E(f_{21}|Z_l)\right) + 3E(f_{21})\right\} \\
&\quad + \left\{(n)_1^{-1}\left[\sum_{i=1}^{n} E(f_{21}|Z_i) + \sum_{j=1}^{n} E(f_{21}|Z_j) + \sum_{l=1}^{n} E(f_{21}|Z_l)\right] - 3E(f_{21})\right\} \\
&\quad + E(f_{21}) \\
&\quad \overset{not}{=} U_n^3 f_{21}^3(\cdot,\cdot,\cdot;h) + U_n^2 f_{21}^2(\cdot,\cdot;h) + P_n^1 f_{21}^1(\cdot;h) + E(f_{21})
\end{aligned}
$$

where $(n)_k = n(n-1)...(n-k+1)$, $Z_i = (Y_i, X_i)$ and

$$
\begin{aligned}
f_{21}(z_i,z_j,z_l) &= f_{21}(z_i,z_j,z_l;h) \\
&= a_i\ [y_j - r(x_i\theta_\psi)]\ [y_l - r(x_i\theta_\psi)] \\
&\qquad \times K_h(x_i\theta_\psi - x_j\theta_\psi)\ K_h(x_j\theta_\psi - x_l\theta_\psi)\ I_i\ I_j\ I_l\,.
\end{aligned}
$$

We show below that $E(f_{21})$ is the dominant term in the decomposition of $T_{21}(h)$.

Since $E(Y^4)$ is finite, the family $\{h^2 f_{21}^3(\cdot,\cdot,\cdot;h)\,;\,h\in(0,1]\}$ is Euclidean for a squared integrable envelope (see Lemma 22(ii) and a straightforward generalization of Corollary 21 of Nolan and Pollard (1987); see also Lemma 2.14(i,ii) of Pakes and Pollard (1989)). Hereafter NP87 stands for a short of Nolan and Pollard (1987) while PP89 abbreviates Pakes and Pollard (1989). Apply Corollary 4 of Sherman (1994a) and deduce

$$U_n^3 f_{21}^3(\cdot,\cdot,\cdot;h) = h^{-2}O_P(n^{-3/2}) = O_P(h^{-2}n^{-3/2}). \tag{23}$$

Next remark that

$$\begin{aligned}
E(f_{21}|Z_i,Z_j) &= a_i\ [Y_j - r(V_i)]\ K_h(V_i - V_j) \\
&\quad \times E\{[Y_l - r(V_i)]\ K_h(V_i - V_l)\ I_l|Z_j,Z_i\}\ I_i\ I_j \\
&= a_i\ [Y_j - r(V_i)]\ K_h(V_i - V_j) \\
&\quad \times E\{E\{[Y_l - r(V_i)]\ K_h(V_i - V_l)\ I_l|Z_i,V_l\}|Z_i\}\ I_i\ I_j \\
&= a_i\ [Y_j - r(V_i)]\ K_h(V_i - V_j) \\
&\quad \times E\{[r(V_l) - r(V_i)]\ E(I_l|V_l)\ K_h(V_i - V_l)|V_i\}\ I_i\ I_j.
\end{aligned}$$

On the other hand, the definition of r, the symmetry of K, a simple change of variables and a second order Taylor expansion give

$$\begin{aligned}
g_{12}(v;h) &= E\{[r(V_l) - r(v)]\ E(I_l|V_l)\ K_h(v - V_l)\} \qquad (24)\\
&= \int N(u)K_h(v-u)\,du - r(v)\int D(u)K_h(v-u)\,du \\
&= \int N(v-wh)\,I_{A(\theta_\psi)}(v-wh)\,K(w)dw \\
&\quad -r(v)\int D(v-wh)\,I_{A(\theta_\psi)}(v-wh)\,K(w)dw \\
&= h^2\int s_N(v,w;h)\,I_{A(\theta_\psi)}(v-wh)\ w^2K(w)\,dw \\
&\quad -h^2\int s_D(v,w;h)\,I_{A(\theta_\psi)}(v-wh)\ w^2K(w)\,dw,
\end{aligned}$$

where $s_N(v,w;h)$ and $s_D(v,w;h)$ are the reminders under integral form of the Taylor expansions of N and D, respectively. That is, for $v - wh \in A(\theta_\psi)$,

$$s_L(v,w;h) = \int_v^{v-wh} (v - wh - t)\ L''(t)\,dt$$

where L is either N or D. Little algebra shows that the families $\{s_L(\cdot,\cdot;h);\,h\in(0,1]\}$ verifies the conditions of Lemma 2.13 of PP89 for $t = h$ and $x = (v,w)$. We deduce that these families are Euclidean for a constant envelope. Next apply Lemma 22(ii) of NP87 and Lemma 2.14(ii) of PP89 in order to deduce that $\left\{s_L(\cdot,\cdot;h)\,I_{A(\theta_\psi)}(\cdot - \cdot\, h)\,;\,h\in(0,1]\right\}$ is Euclidean for a constant envelope. Finally, we interpret the last two integrals in (24) as expectations with respect to the finite measure $w^2K(w)\,dw$ and we apply Corollary 21 of NP87. We further deduce that the family $\{h^{-2}g_{12}(\cdot;h);\,h\in(0,1]\}$ is Euclidean for a constant envelope and, consequently, that $\{h^{-1}E[f_{21}(z_i,z_j,\cdot\,;h)]\,;\,h\in(0,1]\}$ is Euclidean for the squared integrable envelope $C|Y_j - r(V_i)|\ I_i\ I_j$, for some $C > 0$ (use K and r bounded and $E(Y_j^2 I_j) < \infty$). Similar arguments can be used in order to deduce that

$\{ h^{-1} E [f_{21} (z_i, \cdot , z_l; h)] ; \ h \in (0,1]\}$ is Euclidean for the corresponding squared integrable envelope. The last term of $U_n^2 f_{21}^2 (\cdot, \cdot; h)$ to be studied is

$$\begin{aligned} E (f_{21}|Z_j, Z_l) &= E [f_{21}| (Y_j, V_j) , (Y_l, V_l)] \\ &= E \{a_i \ [Y_j - r (V_i)] [Y_l - r (V_i)] \\ &\quad \times K_h (V_i - V_j) K_h (V_i - V_l) \ I_i \ I_j \ I_l | (Y_j, V_j) , (Y_l, V_l)\} \\ &= E \{a_i \ [Y_j - r (V_i)] [Y_l - r (V_i)] \\ &\quad \times K_h (V_i - V_j) K_h (V_i - V_l) E (I_i|V_i) \ | (Y_j, V_j) , (Y_l, V_l)\} I_j \ I_l. \end{aligned}$$

Recall that a_i is a bounded function of V_i. By a change of variable we have

$$\begin{aligned} & E [f_{21}| (Y_j, V_j) = (y_j, v_j) , (Y_l, V_l) = (y_l, v_l)] \\ & = \int a(u) \ [y_j - r (u)] \ [y_l - r (u)] \ K_h (u - v_j) \ K_h (u - v_l) \ D(u) \, du \ I_j \ I_l \\ & = h^{-1} \int a(v_j + wh) \ [y_j - r (v_j + wh)] \ [y_l - r (v_j + wh)] \\ & \qquad \times K (w) \ K [(v_j - v_l) / h + w] \ D(v_j + wh) \ I_{A(\theta_\psi)} (v_j + wh) \, dw \ I_j \ I_l. \end{aligned}$$

Inspired from Example 11, p 798, NP87, we consider the last integral an expectation with respect to the probability defined by $K (w) \, dw$. Note that $a(\cdot)$, $r (\cdot)$, $K (\cdot)$, $D(\cdot)$ and $I_{A(\theta_\psi)} (\cdot)$ are all functions with bounded variation. Using Lemma 22(ii) and Corollary 21 of NP87 and Lemma 2.14(ii) of PP89, deduce that $\{ h E [f_{21} (\cdot, z_j, z_l ; h)] ; \ h \in (0,1]\}$ is Euclidean for a squared integrable envelope. Finally, $\{ h \ f_{21}^2 (\cdot, \cdot ; h) ; \ h \in (0,1]\}$ is Euclidean for a squared integrable envelope. Moreover, from the previous calculations and a dominated convergence argument we get

$$\lim_{h \to 0} h \ f_{21}^2 (z_{i_1}, z_{i_2}; h) = 0,$$

provided that $x_{i_1} \theta_\psi \neq x_{i_2} \theta_\psi$, and thus we may extend $h \ f_{21}^2 (\cdot, \cdot, \cdot ; h)$ by continuity, except for a negligeable set, and verify Corollary 8(ii) of Sherman (1994a) for $U_n^2 (h \, f_{21}^2 (\cdot, \cdot; h))$ (in that corollary take $\theta = h$ and $\theta_0 = 0$). We deduce

$$U_n^2 f_{21}^2 (\cdot, \cdot; h) = o_P(h^{-1} n^{-1}). \tag{25}$$

For the order of $P_n^1 f_{21}^1 (\cdot; h)$ let us write

$$\begin{aligned} E (f_{21}|Z_i) &= E [E (f_{21}|Z_i, V_j, V_l) \ | \ Z_i] \ I_i \\ &= a_i E \{ \ E [(Y_j - r (V_i)) \ I_j \ | \ Z_i, V_j] \ E [(Y_l - r (V_i)) \ I_l \ | \ Z_i, V_l] \\ &\qquad \times K_h (V_i - V_j) \ K_h (V_i - V_l) \ | \ Z_i\} \ I_i \\ &= a_i E \{ \ [r (V_j) - r (V_i)] \ [r (V_l) - r (V_i)] \ E (I_j|V_j) \ E (I_l|V_l) \\ &\qquad \times K_h (V_i - V_j) \ K_h (V_i - V_l) \ | \ V_i\} \ I_i \\ &= a_i \ E \{[r (V_j) - r (V_i)] \ K_h (V_j - V_i) \ E (I_j|V_j) \ | \ V_i\} \\ &\qquad \times E \{[r (V_l) - r (V_i)] \ K_h (V_l - V_i) \ E (I_l|V_l) \ | \ V_i\} \ I_i \\ &= a_i \ \{E \{[r (V_j) - r (V_i)] \ K_h (V_j - V_i) \ E (I_j|V_j) \ | \ V_i\}\}^2 \ I_i \\ &= a_i \ [g_{12}(V_i; h)]^2 \ I_i, \end{aligned}$$

with $g_{12}(\cdot; h)$ defined as in (24). Deduce that $\{h^{-4} E [f_{21} (z_i, \cdot, \cdot; h)] ; \ h \in (0,1]\}$ is Eu-

clidean for a constant envelope. On the other hand,

$$
\begin{aligned}
&E\left(f_{21}|Z_j\right) = E\left[E\left(f_{21}|Z_j,Z_i\right)|Z_j\right] \\
&= E\left\{E\left\{a\left(V_i\right)\ \left[Y_j - r\left(V_i\right)\right]\ \left[Y_l - r\left(V_i\right)\right]\ K_h\left(V_i - V_j\right)\ K_h\left(V_i - V_l\right)\ I_i\ I_j\ I_l|Z_j,Z_i\right\}|Z_j\right\} \\
&= E\left\{a\left(V_i\right)\ \left[Y_j - r\left(V_i\right)\right]\ E\left\{\left[Y_l - r\left(V_i\right)\right]\ K_h\left(V_i - V_l\right)\ I_l|\, Z_j,Z_i\right\}\ K_h\left(V_i - V_j\right)\ I_i\,|Z_j\right\}\ I_j \\
&= E\left\{a\left(V_i\right)\ \left[Y_j - r\left(V_i\right)\right]\ E\left\{\left[Y_l - r\left(V_i\right)\right]\ K_h\left(V_i - V_l\right)\ I_l\ I_i\,|\,V_i\right\}\ K_h\left(V_i - V_j\right)|\,Z_j\right\}\ I_j \\
&= E\left\{a\left(V_i\right)\ \left[Y_j - r\left(V_i\right)\right]\right. \\
&\qquad \left.\times\ E\left\{E\left[\left(Y_l - r\left(V_i\right)\right)\ I_l|V_i,V_l\right]\ K_h\left(V_i - V_l\right)|V_i\right\}\ K_h\left(V_i - V_j\right)\ I_i|Z_j\right\}\ I_j \\
&= E\left\{a\left(V_i\right)\ \left[Y_j - r\left(V_i\right)\right]\right. \\
&\qquad \left.\times\ E\left\{\left[r\left(V_l\right) - r\left(V_i\right)\right]\ E\left(I_l|V_l\right)\ K_h\left(V_i - V_l\right)|V_i\right\}\ K_h\left(V_i - V_j\right)\ I_i|Z_j\right\}\ I_j \\
&= E\left\{a\left(V_i\right)\ \left[Y_j - r\left(V_i\right)\right]\ g_{12}(V_i\,;h)\ K_h\left(V_i - V_j\right)\ E\left(I_i|V_i\right)\ |\,Y_j,V_j\right\}\ I_j.
\end{aligned}
$$

Moreover, by a change of variables we have

$$
\begin{aligned}
& E\left\{a\left(V_i\right)\ \left[Y_j - r\left(V_i\right)\right]\ g_{12}(V_i\,;h)\ K_h\left(V_i - V_j\right)\ E\left(I_i|V_i\right)\ |\,Y_j = y, V_j = v\right\} \\
= & \int a(t)\ \left[y - r\left(t\right)\right]\ g_{12}(t;h)\ K_h\left(t - v\right)\ D(t)\,dt \\
= & \int a(uh+v)\ \left[y - r\left(uh+v\right)\right]\ g_{12}(uh+v;h)\ K\left(u\right)\ D(uh+v)\,du \\
= & \int a(u+vh)\ \left[y - r\left(uh+v\right)\right]\ \widetilde{g}_{12}(u,v;h)\ K\left(u\right)\ D(uh+v)\,du,
\end{aligned}
$$

where

$$
\begin{aligned}
\widetilde{g}_{12}(u,v;h) \;=\; & h^2 \int \widetilde{s}_N(u,w;h)\, I_{A(\theta_\psi)}\left(uh+v-wh\right)\ w^2 K(w)\,dw \\
& -h^2 \int \widetilde{s}_D(u,w;h)\, I_{A(\theta_\psi)}\left(uh+v-wh\right)\ w^2 K(w)\,dw,
\end{aligned}
$$

with,

$$
\widetilde{s}_L(u,w;h) = s_L(uh+v,w;h) = \int_{uh+v}^{uh+v-wh} \left(uh+v-wh-t\right)\ L''(t)\,dt
$$

for $uh+v-wh \in A(\theta_\psi)$ and L equal to N or D. With the same arguments as above it can be shown that the family $\{\widetilde{s}_L(\cdot,\cdot;h);\ h\in(0,1]\}$ is bounded and verifies the conditions of Lemma 2.13 of PP89 for $t = h$ and $x = (u,w)$. Now it is easy to see that $\{\,h^{-2}\,E\left[\,f_{21}\left(\cdot,z_j,\cdot\,;h\right)\right];\ h\in(0,1]\}$ is Euclidean for a constant envelope. The same arguments apply for $E\left(f_{21}|Z_l\right)$. Use Corollary 4 of Sherman (1994a) to see that

$$
P_n^1 f_{21}^1\left(\cdot;h\right) = O_P\left(h^2 n^{-1/2}\right). \tag{26}
$$

The order of $T_{22}\left(h\right)$: we consider the following decomposition of $T_{22}\left(h\right)$ as a sum of mean and two *degenerate* $U-$statistics

$$
\begin{aligned}
T_{22}\left(h\right) \;=\; & U_n^2 f_{22}\left(\cdot,\cdot\right) \\
= \; & (n)_2^{-1} \sum_{i\neq j} \left[f_{22}\left(Z_i,Z_j\right) - E\left(f_{22}|Z_i\right) - E\left(f_{22}|Z_j\right) + E\left(f_{22}\right)\right] \\
& + (n)_1^{-1} \sum_{i=1}^{n} \left[E\left(f_{22}|Z_i\right) + E\left(f_{22}|Z_j\right) - 2E\left(f_{22}\right)\right] \\
& + E\left(f_{22}\right) \overset{not}{=} U_n^2 f_{22}^2\left(\cdot,\cdot;h\right) + P_n^1 f_{22}^1\left(\cdot;h\right) + E\left(f_{22}\right)
\end{aligned}
$$

where

$$f_{22}(z_i, z_j) = f_{21}(z_i, z_j; h) = a_i \ [y_j - r(x_i\theta_\psi)]^2 \ K_h^2(x_i\theta_\psi - x_j\theta_\psi) \ I_i \ I_j.$$

Using the same kind of arguments as those used for the order of $U_n^3 f_{21}^3(\cdot; h)$ we deduce that

$$U_n^2 f_{22}^2(\cdot, \cdot; h) = h^{-2} O_P\left(n^{-1}\right) = O_P\left(h^{-2} n^{-1}\right). \tag{27}$$

For the order $P_n^1 f_{22}^1(\cdot, \cdot; h)$ we write

$$\begin{aligned}
&E(f_{22}|Z_i) = E(f_{22} \mid V_i) \\
&= a_i \ E\left\{E\left\{[Y_j - r(V_i)]^2 \ I_j|V_i, V_j\right\} \ K_h^2(V_i - V_j) \mid V_i\right\} \ I_i \\
&= a_i \ E\left\{E\left\{[Y_j - r(V_j) + r(V_j) - r(V_i)]^2 \ I_j|V_i, V_j\right\} \ K_h^2(V_i - V_j) \mid V_i\right\} \ I_i \\
&= a_i \ E\left\{E\left\{[Y_j - r(V_j)]^2 \ I_j \mid V_j\right\} \ K_h^2(V_i - V_j) \mid V_i\right\} \ I_i \\
&\quad + a_i \ E\left\{E\left\{[r(V_j) - r(V_i)]^2 \ I_j \mid V_i, V_j\right\} \ K_h^2(V_i - V_j) \mid V_i\right\} \ I_i \\
&= a_i \ E\left[v_{\theta_\psi}(V_j) \ E(I_i \mid V_i) \ K_h^2(V_i - V_j) \mid V_i\right] \ I_i \\
&\quad + a_i \ E\left\{[r(V_j) - r(V_i)]^2 \ E(I_i \mid V_i) \ K_h^2(V_i - V_j) \,|V_i\right\} \ I_i,
\end{aligned}$$

where $v_{\theta_\psi}(t) = var^A(Y|X\theta_\psi = t)$. Using again a change of variable and the results of NP87 and PP89 we deduce that the family $\{h\, E[f_{22}(z_i, \cdot\,; h)]\,;\ h \in (0, 1]\}$ is Euclidean for a constant envelope. Similar arguments apply to the family $\{h\, E[f_{22}(\cdot, z_j\,; h)]\,;\ h \in (0, 1]\}$ which is thus Euclidean for a squared integrable envelope. Corollary 4 of Sherman (1994a) gives

$$P_n^1 f_{22}(\cdot; h) = O_P\left(h^{-1} n^{-1/2}\right). \tag{28}$$

From (23) and (25) to (28) we obtain

$$\begin{aligned}
T_2(h) &= O_P(h^{-2}n^{-3/2}) + o_P(h^{-1}n^{-1}) + O_P\left(h^2 n^{-1/2}\right) + \ E(f_{21}) \\
&\quad + O_P\left(h^{-2}n^{-2}\right) + O_P\left(h^{-1}n^{-3/2}\right) + n^{-1}E(f_{22}) \\
&= o_P(h^{-1}n^{-1}) + O_P\left(h^2 n^{-1/2}\right) + \ E(f_{21}) + n^{-1}E(f_{22})
\end{aligned}$$

From the previous formulae, suitable change of variables, Taylor expansions and symmetry of K we have

$$\begin{aligned}
E(f_{21}|Z_i) &= a_i \ \{E\{[r(V_j) - r(V_i)] \ K_h(V_j - V_i) \ E(I_j|V_j) \mid V_i\}\}^2 \ I_i \\
&= a_i \ \left\{r(V_i) \ D(V_i) + \frac{K_1}{2} \ h^2 \ (rD)''(V_i) + O_P\left(h^3\right)\right. \\
&\quad \left. - r(V_i) \ D(V_i) - \frac{K_1}{2} \ h^2 \ r(V_i) \ D''(V_i) + O_P\left(h^3\right)\right\}^2 \\
&= h^4 \ a_i \ \frac{K_1^2}{4} \ \left[(rD)''(V_i) - r(V_i) \ D''(V_i)\right]^2 \ I_i + O_P\left(h^5\right)
\end{aligned}$$

and

$$E(f_{21}) = E[E(f_{21}|Z_i)] = h^4 A_1 + O_P\left(h^5\right).$$

On the other hand,

$$\begin{aligned}
E(f_{22}|Z_i) &= a_i \ E\left[v_{\theta_\psi}(V_j) \ E(I_i \mid V_i) \ K_h^2(V_i - V_j) \mid V_i\right] \ I_i \\
&\quad + a_i \ E\left\{[r(V_j) - r(V_i)]^2 \ E(I_i \mid V_i) \ K_h^2(V_i - V_j) \,|V_i\right\} \ I_i, \\
&= h^{-1} K_2 \ a_i \ v_{\theta_\psi}(V_i) \ D(V_i) \ I_i + O_P(1),
\end{aligned}$$

and

$$n^{-1}E\left(f_{22}\right)=n^{-1}E\left[E\left(f_{22}|Z_i\right)\right]=n^{-1}h^{-1}A_2+O_P\left(n^{-1}\right).$$

We may conclude that

$$T\left(h\right)=J\left(h\right)+\left\{\text{terms of smaller order in }h\right\}.$$

STEP 2

2A : The order of $R_2\left(h\right)$ We further decompose $R_2\left(h\right)$:

$$\begin{aligned}
R_2\left(h\right) &= \frac{1}{n}\sum_{i=1}^{n}\left[\psi\left(Y_i,\hat{r}^i\left(V_i\right)\right)-\psi\left(r_{\theta_\psi}\left(X_i\theta_\psi\right),\hat{r}^i\left(V_i\right)\right)\right]\ I_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\partial_1\psi\left(r\left(V_i\right),\hat{r}^i\left(V_i\right)\right)\ \left[Y_i-r\left(V_i\right)\right]\ I_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\{\partial_1\psi\left(r\left(V_i\right),r\left(V_i\right)\right)+\partial_{12}^2\psi\left(r\left(V_i\right),r\left(V_i\right)\right)\ \left[\hat{r}^i\left(V_i\right)-r\left(V_i\right)\right]\right\} \\
&\qquad\qquad \times\left[Y_i-r\left(V_i\right)\right]\ I_i \\
&\quad +\left\{\text{terms of smaller order in }h\right\}
\end{aligned}$$

The second equality is due to $\partial_{11}^2\psi\left(s,r\right)\equiv 0$. We only retain the dominating term containing h :

$$\begin{aligned}
\widetilde{R}_2\left(h\right) &= \frac{1}{n}\sum_{i=1}^{n}\partial_{12}^2\psi\left(r\left(V_i\right),r\left(V_i\right)\right)\ \left[Y_i-r\left(V_i\right)\right]\ \left[\hat{r}^i\left(V_i\right)-r\left(V_i\right)\right]\ I_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\alpha_i\ \left[\hat{r}^i\left(V_i\right)-r\left(V_i\right)\right]\ I_i
\end{aligned}$$

with $\alpha_i=\alpha(Z_i)=\partial_{12}^2\psi\left(r\left(V_i\right),r\left(V_i\right)\right)\ \left[Y_i-r\left(V_i\right)\right]$. Note that $E\left(\alpha\left(Z_i\right)I_i\,|\,V_i\right)=0$. We have

$$\begin{aligned}
\widetilde{R}_2\left(h\right) &= \frac{1}{n}\sum_{i=1}^{n}\alpha_i\ \left(\frac{\widehat{N}^i}{\widehat{D}^i}-\frac{N}{D}\right)\ I_i=\frac{1}{n}\sum_{i=1}^{n}\left[\alpha_i\ \frac{1}{D}\right]\ \left(\widehat{N}^iD-N\widehat{D}^i\right)\ \frac{1}{\widehat{D}^i}\ I_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\beta_i\ \left(\widehat{N}^iD-N\widehat{D}^i\right)\ \left[\frac{1}{D}\ -\ \frac{1}{D^2}\ \left(\widehat{D}^i-D\right)+\ldots\right]\ I_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\beta_i\ \left(\widehat{N}^i-r\widehat{D}^i\right)\ I_i-\frac{1}{n}\sum_{i=1}^{n}\gamma_i\ \left(\widehat{N}^i-r\widehat{D}^i\right)\left(\widehat{D}^i-D\right)\ I_i \\
&\quad +\left\{\text{terms of smaller order in }h\right\} \\
&= \widetilde{R}_{21}\left(h\right)+\widetilde{R}_{22}\left(h\right)+\left\{\text{terms of smaller order in }h\right\}
\end{aligned}$$

with $\beta_i=\beta(Z_i)=\alpha(Z_i)D^{-1}$ and $\gamma_i=\beta_iD^{-1}$. Moreover,

$$\widetilde{R}_{21}\left(h\right) = \frac{1}{n}\sum_{i=1}^{n}\beta_i\left[\frac{1}{n-1}\sum_{j\neq i}Y_j\ K_h\left(V_i-V_j\right)\ I_j-r\left(V_i\right)\frac{1}{n-1}\sum_{j\neq i}K_h\left(V_i-V_j\right)\ I_j\right]\ I_i$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n}\beta_i\ \left\{\frac{1}{n-1}\sum_{j\neq i}\left[Y_j - r\left(V_i\right)\right]\ K_h\left(V_i - V_j\right)\ I_j\right\}\ I_i \\
&= (n)_2^{-1}\sum_{i\neq j}\ \beta_i\ \left[Y_j - r\left(V_i\right)\right]\ K_h\left(V_i - V_j\right)\ I_j\ I_i \\
&= U_n^2 g_2\left(\cdot,\cdot;h\right),
\end{aligned}
$$

As above, we may decompose

$$
\begin{aligned}
\widetilde{R}_{21}\left(h\right) &= U_n^2 g_2^2\left(\cdot,\cdot;h\right) + P_n^1 g_2^1\left(\cdot;h\right) + E\left(g_2\right) \\
&= (n)_2^{-1}\sum_{i\neq j}\left[g_2\left(Z_i,Z_j;h\right) - E\left(g_2|Z_i\right) - E\left(g_2|Z_j\right) + E\left(g_2\right)\right] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left[E\left(g_2|Z_i\right) + E\left(g_2|Z_j\right) - 2E\left(g_2\right)\right] \\
&\quad + E\left(g_2\right) \\
&= (n)_2^{-1}\sum_{i\neq j}\left[g_2\left(Z_i,Z_j\right) - E\left(g_2|Z_i\right)\right] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} E\left(g_2|Z_i\right),
\end{aligned}
$$

since $E\left(g_2|Z_j\right) = 0$. Since $\beta(\cdot)$ is bounded we deduce that $\{\, h\, g_2\left(\cdot,\cdot\,;h\right);\ h\in(0,1]\}$ is Euclidean for a squared integrable envelope. Similar arguments as those used for $h \to h\, f_{21}^2\left(\cdot,\cdot,\cdot;h\right)$ allows us to extend $h \to h\, g_2^2\left(\cdot,\cdot\,;h\right)$ by continuity and to apply Corollary 8 of Sherman (1994a) in order to deduce

$$
U_n^2 g_2^2\left(\cdot,\cdot;h\right) = o_P(h^{-1}n^{-1}).
$$

Same arguments as used for $\{h^{-2}g_{12}(\cdot;h);\ h\in(0,1]\}$ (see (24)) allows to write

$$
P_n^1 g_2^1\left(\cdot;h\right) = O_P\left(h^2 n^{-1/2}\right).
$$

On the other hand, for the other term retained from $R_2\left(h\right)$ we have

$$
\begin{aligned}
\widetilde{R}_{22}\left(h\right) &= \frac{1}{n}\sum_{i=1}^{n}\gamma_i\ \left[\frac{1}{n-1}\sum_{j\neq i}\left[Y_j - r\left(V_i\right)\right]\ K_h\left(V_i - V_j\right)\ I_j\right] \\
&\qquad\qquad \times\left[\frac{1}{n-1}\sum_{l\neq i}\ K_h\left(V_i - V_l\right)\ I_l - D\left(V_i\right)\right]\ I_i \\
&= \frac{n-1}{n-2}\,\widetilde{R}_{221}\left(h\right) - \frac{1}{n-1}\,\widetilde{R}_{222}\left(h\right) + \frac{1}{n-1}\,\widetilde{R}_{223}\left(h\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\widetilde{R}_{221}\left(h\right) &= (n)_3^{-1}\sum_{i\neq j\neq l}\gamma_i\ \left[Y_j - r\left(V_i\right)\right]\ K_h\left(V_i - V_j\right)\ K_h\left(V_i - V_l\right)\ I_i\ I_j\ I_l \\
\widetilde{R}_{222}\left(h\right) &= (n)_2^{-1}\sum_{i\neq j}\beta_i\ \left[Y_j - r\left(V_i\right)\right]\ K_h\left(V_i - V_j\right)\ I_i I_j
\end{aligned}
$$

and

$$\widetilde{R}_{223}(h) = (n)_2^{-1} \sum_{j \neq i} \gamma_i \ [Y_j - r(V_i)] \ K_h^2(V_i - V_j) \ I_i I_j.$$

Using the same techniques as above we obtain

$$\widetilde{R}_{22}(h) = O_P(h^{-2} n^{-3/2}).$$

Consequently,

$$R_2(h) = o_P\left(n^{-1}h^{-1}\right) + O_P\left(n^{-1/2}h^2\right) + \{\text{terms independent of } h\}.$$

2B : The order of $R_1(\theta, h)$

$$\begin{aligned} R_1(\theta,h) &= \frac{1}{n}\sum_{i=1}^{n}\left[\psi\left(Y_i, \hat{r}^i_{\theta,h}(X_i\theta)\right) - \psi\left(Y_i, r_\theta(X_i\theta)\right)\right] \ I_i \\ &\quad -\frac{1}{n}\sum_{i=1}^{n}\left[\psi\left(Y_i, \hat{r}^i_{\theta_\psi,h}(X_i\theta_\psi)\right) - \psi\left(Y_i, r_{\theta_\psi}(X_i\theta_\psi)\right)\right] \ I_i \\ &= \widetilde{R}_1(\theta,h) - \widetilde{R}_1(\theta_\psi,h). \end{aligned}$$

Since $\partial^2_{22}\psi(y,r) = C''(r)(y-r) - C'(r) = C''(r)(y-r) + \partial^2_{22}\psi(r,r)$, we have

$$\begin{aligned} \widetilde{R}_1(\theta,h) &= \frac{1}{n}\sum_{i=1}^{n}\left[\psi\left(Y_i, \hat{r}^i_{\theta,h}(X_i\theta)\right) - \psi\left(Y_i, r_\theta(X_i\theta)\right)\right] \ I_i \qquad (29) \\ &= \frac{1}{n}\sum_{i=1}^{n}\partial_2\psi\left(Y_i, r_\theta(X_i\theta)\right) \ \left[\hat{r}^i_{\theta,h}(X_i\theta) - r_\theta(X_i\theta)\right] \ I_i \\ &\quad +\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\partial^2_{22}\psi\left(Y_i, r_\theta(X_i\theta)\right) \ \left[\hat{r}^i_{\theta,h}(X_i\theta) - r_\theta(X_i\theta)\right]^2 \ I_i \\ &\quad +\{\text{terms of smaller order in } h\} \\ &= \frac{1}{n}\sum_{i=1}^{n}\delta(Z_i;\theta) \ \left[\hat{r}^i_{\theta,h}(X_i\theta) - r_\theta(X_i\theta)\right] \ I_i \\ &\quad +\frac{1}{n}\sum_{i=1}^{n}\zeta(Z_i;\theta)\left[\hat{r}^i_{\theta,h}(X_i\theta) - r_\theta(X_i\theta)\right]^2 \ I_i \\ &\quad +\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\partial^2_{22}\psi\left(r_\theta(X_i\theta), r_\theta(X_i\theta)\right) \ \left[\hat{r}^i_{\theta,h}(X_i\theta) - r_\theta(X_i\theta)\right]^2 \ I_i \\ &\quad +\{\text{terms of smaller order in } h\} \\ &\stackrel{not}{=} \widetilde{R}_{11}(\theta,h) + \widetilde{R}_{12}(\theta,h) + \widetilde{R}_{13}(\theta,h) \\ &\quad +\{\text{terms of smaller order in } h\}, \end{aligned}$$

where $\delta(z;\theta) = \partial_2\psi(y, r_\theta(x\theta))$ and $\zeta(z;\theta) = C''(r_\theta(x\theta))(y - r_\theta(x\theta))/2$. Note that

$$E\left(\delta(Z_i;\theta) I_i \mid X_i\theta\right) = E\left(\zeta(Z_i;\theta) I_i \mid X_i\theta\right) = 0. \qquad (30)$$

For $\widetilde{R}_{11}(\theta,h)$ we may use the same arguments as for $\widetilde{R}_{21}(h)$ in order to deduce

$$\widetilde{R}_{11}(\theta,h) = o_P\left(n^{-1}h^{-1}\right) + O_P\left(n^{-1/2}h^2\right),$$

uniformly over $o_P(1)$ neighborhoods of θ_ψ. It suffices to apply the results of Sherman (1994a) as follows : Corollary 8 for $(\theta,h) \to (\theta_\psi, 0)$ and Corollary 4 for supremum with respect to (θ,h). On the other hand, $\widetilde{R}_{12}(\theta,h)$ and $\widetilde{R}_{13}(\theta,h)$ can be analyzed in the same way as $T_2(h)$ (note that $T_2(h) = \widetilde{R}_{13}(\theta_\psi,h)$). Each of these two quantities can be written as a sum between their expectations and terms of smaller order than $T_2(h)$, uniformly over $o_P(1)$ neighborhoods of θ_ψ. From (30) we deduce $E\left[\widetilde{R}_{12}(\theta,h)\right] = 0$, for any $\theta \in \Theta$. For the expectation of $\widetilde{R}_{13}$ we get

$$E\left[\widetilde{R}_{13}(\theta,h)\right] = A_1(\theta)h^4 + A_2(\theta)n^{-1}h^{-1}$$

with $A_1(\theta)$ and $A_2(\theta)$ defined as in (19) with θ_ψ replaced by $\theta \in \Theta$. Since $A_1(\cdot)$ and $A_2(\cdot)$ are continuous in θ_ψ, we deduce that

$$\begin{aligned} R_1(\theta,h) &= \widetilde{R}_1(\theta,h) - \widetilde{R}_1(\theta_\psi,h) \\ &= \left[A_1(\theta) - A_1(\theta_\psi)\right]h^4 + \left[A_2(\theta) - A_2(\theta_\psi)\right]n^{-1}h^{-1} \\ &\quad + \{\text{terms of smaller order in } h\} \\ &= o_P(T_2(h)), \end{aligned} \tag{31}$$

uniformly over $o_P(1)$ neighborhoods of θ_ψ. We may conclude that the optimal bandwidth $\widehat{h}$ is of order $n^{-1/5}$.

STEP 3 : $\sqrt{n}$−convergence of $\widehat{\theta}_n$

3A. The order of $R_1(\theta,h)$ revised Let us remark (31) together with an obvious bound for $\widetilde{S}(\theta)$ (see the end of Step 3) allows us already to consider θ in a $O_P(h^2) = O_P(n^{-2/5})$ neighborhood of θ_ψ. Indeed, apply a version of Theorem 1 of Sherman (1994a) with $\sqrt{n}$ and n^{-1} replaced by $n^{-2/5}$ and $n^{-4/5}$, respectively (see also Theorem 1, Sherman (1994b), page 376). Using this preliminary rate of $\widehat{\theta}$, we refine the decomposition of $R_1(\theta,h)$ in order to obtain

$$R_1(\theta,h) = o_P\left(\frac{\|\theta - \theta_\psi\|}{\sqrt{n}}\right) + o_P\left(n^{-1}\right),$$

uniformly over $O_P(n^{-2/5})$ neighborhoods of θ_ψ and with respect to domains of order $O_P\left(n^{-1/5}\right)$ for the bandwidth. Assuming a Lipschitz condition on the functions $A_1(\cdot)$ and $A_2(\cdot)$, we remark from (31) that we only have to examine what we called "terms of smaler order in h". More precisely, we have

$$\begin{aligned} R_1(\theta,h) &= \left[\widetilde{R}_{11}(\theta,h) - \widetilde{R}_{11}(\theta_\psi,h)\right] \\ &\quad + \left[\widetilde{R}_{12}(\theta,h) - \widetilde{R}_{12}(\theta_\psi,h)\right] \\ &\quad + \left[\left(\widetilde{R}_{13}(\theta,h) - E\left[\widetilde{R}_{13}(\theta,h)\right]\right) - \left(\widetilde{R}_{13}(\theta_\psi,h) - E\left[\widetilde{R}_{13}(\theta_\psi,h)\right]\right)\right] \\ &\quad + \left[A_1(\theta) - A_1(\theta_\psi)\right]h^4 + \left[A_2(\theta) - A_2(\theta_\psi)\right]n^{-1}h^{-1} \\ &\quad + O_P(\left(\left|\hat{r}^i_{\theta,h}(X_i\theta) - r_\theta(X_i\theta)\right|^3\right). \end{aligned}$$

It can be easily remarked that the last term is of order $o_P(n^{-1})$ and thus we only have to examine the first three differences. Let us consider only the first one, the other two are to be treated in the same manner. Fix $\theta_n \in \Theta$, $n > 1$, an arbitrary sequence such that $\theta_n - \theta_\psi = O(n^{-2/5})$. Moreover, let $H_n = \left(0, C_1 n^{-1/5}\right)$ with $0 < C_1$ arbitrarily fixed. We may use the usual decomposition in degenerate $U-$statistics (see $\widetilde{R}_{21}(h)$) and write

$$\begin{aligned}
& \widetilde{R}_{11}(\theta_n, h) - \widetilde{R}_{11}(\theta_\psi, h) \\
= \; & (n)_2^{-1} \sum_{i \neq j} \delta(Z_i; \theta_n) D^{-1}(X_i\theta_n) \; [Y_j - r(X_i\theta_n)] \; K_h(X_i\theta_n - X_j\theta_n) \; I_j \; I_i \\
& - (n)_2^{-1} \sum_{i \neq j} \delta(Z_i; \theta_\psi) D^{-1}(X_i\theta_\psi) \; [Y_j - r(X_i\theta_\psi)] \; K_h(X_i\theta_\psi - X_j\theta_\psi) \; I_j \; I_i \\
& + o_P(n^{-1}) \\
= \; & (n)_2^{-1} \sum_{i \neq j} \left[\widetilde{f}_{11}^2(Z_i, Z_j; \theta_n, h) - \widetilde{f}_{11}^2(Z_i, Z_j; \theta_\psi, h) \right] \\
& + (n)_1^{-1} \sum_{i=1}^{n} \left[\widetilde{f}_{11}^1(Z_i; \theta_n, h) - \widetilde{f}_{11}^1(Z_i; \theta_\psi, h) \right] \\
& + o_P(n^{-1}) \\
& \overset{not}{=} \widetilde{R}_{111} + \widetilde{R}_{112} + o_P(n^{-1}).
\end{aligned}$$

In order to justify the rate $o_P(n^{-1})$ of the reminder term, see the decomposition of $\widetilde{R}_2(h)$ and remark that the order of $\widetilde{R}_{22}(h)$ is still valid uniformly with respect to $\theta \in \Theta$. Denote $\widetilde{g}_n^2(Z_i, Z_j; h) = \widetilde{f}_{11}^2(Z_i, Z_j; \theta_n, h) - \widetilde{f}_{11}^2(Z_i, Z_j; \theta_\psi, h)$ and remark that

$$E\left[\widetilde{f}_{11}^2(Z_i, Z_j; \theta, h)\right] = E\left[\widetilde{f}_{11}^2(Z_i, Z_j; \theta_\psi, h)\right] = E\left[\widetilde{g}_n^2(Z_i, Z_j; h)\right] = 0.$$

By continuity $h\widetilde{g}_n^2(\cdot, \cdot; h) \equiv 0$, for $h = 0$. Given the Lipschitz conditions verified by the functions contained in $\widetilde{f}_{11}$, it can be shown that

$$\sup_{h \in H_n} \left| h\widetilde{g}_n^2(Z_i, Z_j; h) \right| \leq (\theta_n - \theta_\psi) \Psi(Z_i, Z_j),$$

with Ψ squared integrable. Apply Theorem 3, Sherman (1994b) for $f_n(\cdot, \theta) = h\widetilde{g}_n^2(\cdot, \cdot; h)$, $\delta_n = \gamma_n = n^{-1/5}$, $k = 2$, and deduce

$$\widetilde{R}_{111} = O_P\left(\left(n^{-2/5}\right)^\alpha h^{-1} n^{-1}\right),$$

where $0 < \alpha < 1$, uniformly over $O_P\left(n^{-1/5}\right)$ bandwidths. Next, use the same arguments for $h^{-2}\widetilde{g}_n^1(\cdot; h)$ with $\widetilde{g}_n^1(\cdot; h) = \widetilde{f}_{11}^1(\cdot; \theta_n, h) - \widetilde{f}_{11}^1(\cdot; \theta_\psi, h)$ and apply again Sherman's result for $f_n(\cdot, \theta) = h^{-2}\widetilde{g}_n^1(\cdot; h)$, $\delta_n = \gamma_n = n^{-1/5}$, $k = 1$, in order to deduce

$$\widetilde{R}_{112} = O_P\left(\left(n^{-2/5}\right)^\alpha h^2 n^{-1/2}\right), \qquad 0 < \alpha < 1,$$

uniformly over $O_P\left(n^{-1/5}\right)$ bandwidths. Thus we get

$$\widetilde{R}_{11}(\theta_n, h) - \widetilde{R}_{11}(\theta_\psi, h) = o_P(n^{-1}),$$

uniformly with respect to h of order $O_P\left(n^{-1/5}\right)$. Using the same arguments for the other term appearing in the decomposition of $R_1(\theta, h)$, we may conclude

$$R_1(\theta, h) = o_P\left(\frac{\|\theta - \theta_\psi\|}{\sqrt{n}}\right) + o_P\left(n^{-1}\right), \tag{32}$$

uniformly over $O\left(n^{-2/5}\right)$ neighborhoods of θ_ψ (thus uniformly over $O_P\left(n^{-2/5}\right)$ neighborhoods of θ_ψ) and domains of order $O_P\left(n^{-1/5}\right)$ for the bandwidth.

3B :$\sqrt{n}$−convergence and asymptotic normality for $\widehat{\theta}_n$ Consider

$$\Gamma_n(\theta) = \widetilde{S}(\theta) + R_1(\theta, h)$$

$$\Gamma(\theta) = -\frac{1}{2}(\theta - \theta\psi)^T W_\psi(\theta - \theta\psi)$$

where, recall,

$$\widetilde{S}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \psi\left(Y_i, r_\theta(X_i\theta)\right)\ I_i - \frac{1}{n}\sum_{i=1}^{n} \psi\left(Y_i, r_{\theta_\psi}(X_i\theta_\psi)\right)\ I_i.$$

Let

$$V_n = \frac{1}{n}\sum_{i=1}^{n} \partial_\theta \psi\left(Y_i, r_{\theta_\psi}(X_i\theta_\psi)\right)\ I_i$$

and

$$W_n = -\frac{1}{n}\sum_{i=1}^{n} \partial^2_{\theta\theta} \psi\left(Y_i, r_{\theta_\psi}(X_i\theta_\psi)\right)\ I_i$$

Since by classic asymptotic results we have $W_n \to W_\psi$, almost surely, and $V_n = O_P\left(n^{-1/2}\right)$, we may write

$$\begin{aligned}
\widetilde{S}(\theta) &= (\theta - \theta_\psi)^T V_n - \frac{1}{2}(\theta - \theta_\psi)^T W_n (\theta - \theta_\psi) + o_P\left(\|\theta - \theta_\psi\|^2\right) \\
&= O_P\left(\|\theta - \theta_\psi\| / \sqrt{n}\right) - \frac{1}{2}(\theta - \theta_\psi)^T W_\psi (\theta - \theta_\psi) + o_P\left(\|\theta - \theta_\psi\|^2\right),
\end{aligned}$$

uniformly over $o_P(1)$ neighborhoods of θ_ψ. This and (32) ensure (ii) of Theorem 1 of Sherman (1994a), uniformly over $O_P\left(n^{-2/5}\right)$ neighborhoods of θ_ψ and with respect to h of order $O_P\left(n^{-1/5}\right)$. Since (i) of that theorem is obviously verified, we may deduce that

$$\left\|\widehat{\theta} - \theta_\psi\right\| = O_P\left(n^{-1/2}\right).$$

Finally, apply Theorem 2 of Sherman (1994a) in order to obtain the asymptotic distribution of $\widehat{\theta}$. ■

REFERENCES

Bosq, D. and Lecoutre, J-P. (1987). *Théorie de l'estimation fonctionnelle*, Economica, Paris.

DELECROIX, M. and HRISTACHE, M. (1999). M-estimateurs semi-paramétriques dans les modèles à direction révélatrice unique. *Bull. Belg. Math. Soc.*, **6**, 161-185.

FRIEDMAN, J.H. and STUETZLE, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.*, **76**, 817-823.

GOURIEROUX, C. MONFORT, A. and TROGNON, A. (1984). Pseudo maximum likelihood methods: theory. *Econometrica*, **52**, 681-700.

HARDLE, W., HALL, P and ICHIMURA, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.*, **21**, 157-178.

HALL, P (1989). On projection pursuit regression. *Ann. Statist.*, **17**, 573-588.

ICHIMURA, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics*, **58**, 71-120.

ICHIMURA, H and LEE, L.-F. (1991). Semiparametric least squares estimation of multiple index models: single equation estimation. In *Nonparametric and Semiparametric Methods in Statistics and Econometrics,* W. A. Barnett, J. Powell and G. Tauchen, eds., Cambridge University Press, Ch. 1.

KLEIN, R.W. and SPADY, R.H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica*, **61**, 387-421.

MCCULLAGH, P and NELDER, J.A. (1989). *Generalized Linear Models*, 2nd ed., Chapman and Hall, London.

NOLAN, D AND POLLARD, D. (1987). $U-$processes : Rates of convergence. *Ann. Statist.*, **15**, 780-799.

PAKES, A. AND POLLARD, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, **57**, 1027-1057.

SERFLING, R.J. (1980). *Aproximation Theorems of Mathematical Statistics.* Wiley, New-York.

SHERMAN, R.P. (1994a). Maximal inequalities for degenerate $U-$processes with applications to optimization estimators. *Ann. Statist.*, **22**, 439-459.

SHERMAN, R.P. (1994b). U-processes in the analysis of a generalized semiparametric regression estimator. *Econometric Theory*, **10**, 372-395.