# M ROBUSTIFIED ADDITIVE NONPARAMETRIC REGRESSION

Julien Tamine

*Université de la Méditerranée*

Wolfgang Härdle

*Humboldt-Universität zu Berlin*

Lijian Yang

*Michigan State University*

*Abstract*: Additive modelling has been widely used in nonparametric regression to circumvent the "curse of dimensionality", by reducing the problem of estimating a multivariate regression function to the estimation of its univariate components. Estimation of these univariate functions, however, can suffer inaccuracy if the data set is contaminated with extreme observations. As detection and removal of outliers in high dimension is much more difficult than in one dimension, we propose an M type marginal integration estimator that automatically corrects the extreme influence of outliers. We establish the robustness and obtain the asymptotic distribution of the M estimator through the functional approach. As a consequence, our results are valid for $\beta$-mixing samples under mild constraints. Monte Carlo study confirm our theoretical results.

*Short Running Title.* Robustified Additive Regression

*Key words and phrases*: Frechet differential, kernel estimator, marginal integration, M estimator, outliers, robustness.

## 1 . Introduction

Applications of nonparametric regression techniques has proved to be useful in many scientific disciplines, see for instance, Härdle (1990). Its main drawback is its inability to estimate multivariate regression with accuracy. This limitation known as the curse of dimensionality has led many authors to consider an additive form for the regression function, among other dimension reduction techniques, see Hastie and Tibshirani (1990) for an introduction to additive modeling.

Given a random variable $Y$ and a $d$ dimensional set of stochastic regressors $\mathbf{X}$, additive modelling consists of specifying an additive structure for the regression function $m(\mathbf{x}) = E(Y|\mathbf{X}=\mathbf{x})$ :

$$m(\mathbf{x}) = \mu + \sum_{\alpha=1}^{d} m_\alpha(x_\alpha) \tag{1}$$

where $x_\alpha$ is the $\alpha$-th component of $\mathbf{x}$. This specification supposes that the effect of each regressors can be measured separately by the term $m_\alpha(x_\alpha)$. This effect can be graphically represented by the plot of the univariate function $m_\alpha(x_\alpha)$, with easy interpretation. Estimation methods for additive models include polynomial spline (see Stone 1985), the backfitting procedure introduced by Hastie and Tibshirani (1990) and theoretically developed by Opsomer and Ruppert (1997), the marginal integration procedure independently developed by Linton (1995) and Tjøstheim and Auestad (1994). A more recent development was a nonlinear backfitting algorithm proposed by Mammen, Linton and Nielsen (1999), who also established asymptotic distribution theory for their procedure.

In the existing literature on additive modelling, the robustness of estimation procedure has not been theoretically studied. Nevertheless, the problem of outliers is of particular importance in multidimensional setting, since their removing by visual inspection is impractical. Our first aim in this work is to gain insight into the robustness properties of marginal integration procedure. In that view, we first conduct an asymptotic analysis of marginal integration estimation in a functional setting as in Aït-Sahalia (1995), that allows us to give an analog of Hampel influence function, in the manner of Tamine (2002), which defined a smoothed influence function.

Figure 1 illustrates this non-robustness of marginal integration estimator. One can see from plots (a) and (b) the presence of outliers in the simulated data set. As a consequence, the ordinary marginal integration estimators do not fit their target well as can be seen in plots (c) and (d). At the same time, the robust estimator that we propose gives a much better fit.

Since the Nadaraya Watson estimator (Nadaraya 1964, Watson 1964) and the marginal integration estimator based on it are obtained by solving linear least square problems, a natural way of robustification is to use a flatter tail function for minimization criteria. Such estimators are commonly referred to as M-estimators (see Huber 1981, for an expose of these methods) and have already been proved to be useful in univariate nonparametric regression (see chapter 6 of Härdle 1990). Recently, Bianco and Boente (1998) have used them for additive modelling in the special case where all the components of the regressor vector are independent. Our second aim in this work is to extend these results to the more practical case of dependent components of the regressors vector. Furthermore, we aim at leading a detailed theoretical study of the improvement brought by this estimator in terms of robustness. In this respect, we again benefit from the use of functional analysis concepts.

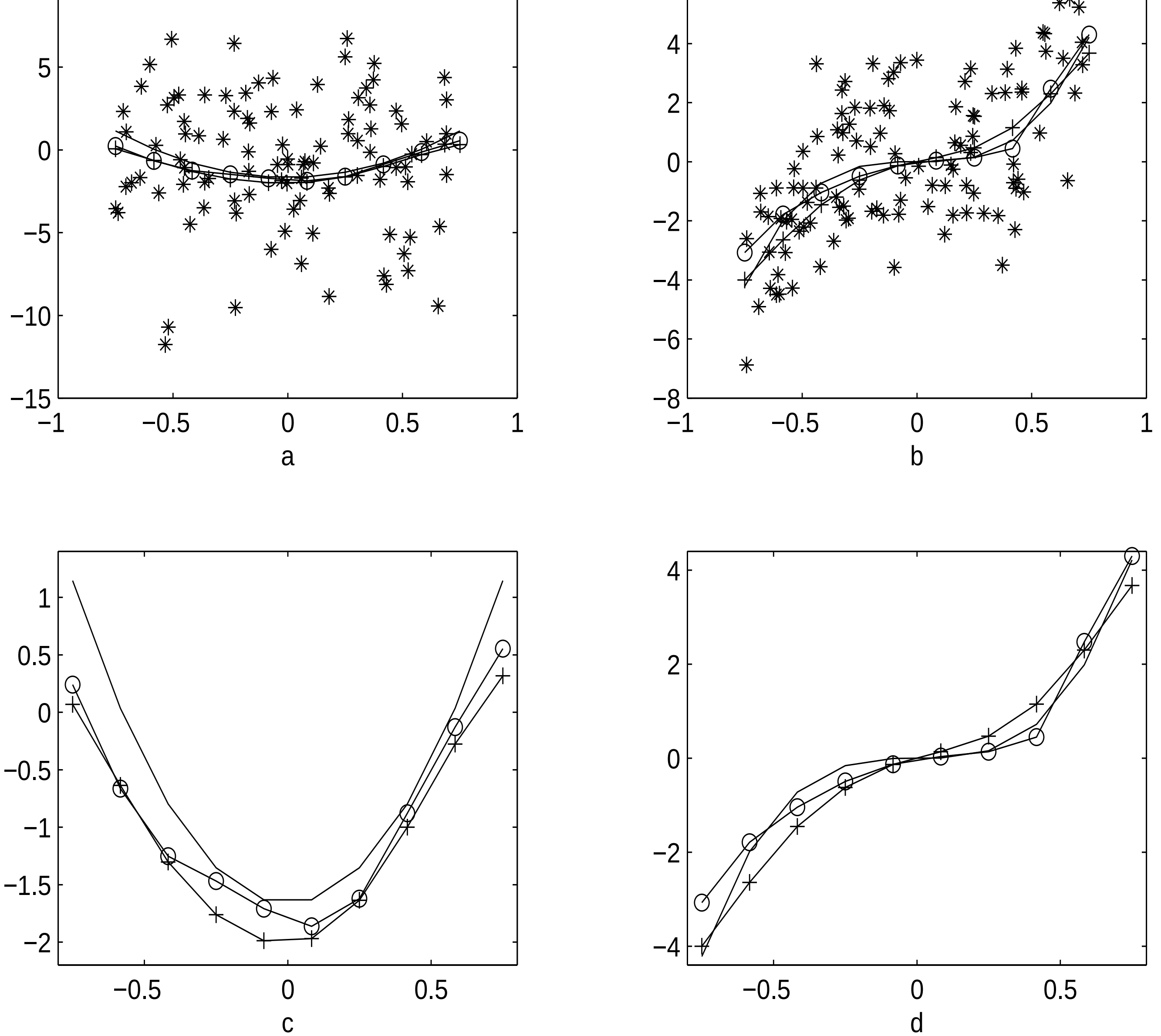



Figure 1: Scatterplots and function estimates for a simulated sample of size 150: $Y = m_1(X_1) + m_2(X_2) + \varepsilon$ where $m_1(t) = 5t^2 - 5/3$, $m_2(t) = 10t^3$ and $\varepsilon$ has a normal mixture distribution (see section 4). (a) plot of $(X_1, Y)$, and $m_1(t)$ — solid, robust estimator — circle, ordinary estimator — cross. (b) is the counterpart of (a) for $(X_2, Y)$ and $m_2$. (c) and (d) are zoomed in copies of (a) and (b).

Our work is organized as follow : in section 2, we study the robustness properties of marginal integration estimator and set the functional framework necessary to calculate its influence function. In section 3, we specify an alternative additive model based on M-estimators, suitable for regression estimation in the presence of outliers in the data. This estimator is shown to be less sensitive to outliers since its influence function is proved to be bounded under suitable conditions on the minimization criterion function. All of our asymptotic results are obtained under mild $\alpha-$mixing assumptions. In section 3, a Monte Carlo simulation illustrates the finite sample behavior of the M-robustified additive estimator.

# 2 . Smooth influence function for marginal integration

## 2.1 Functional Presentation of marginal integration estimator

For the additive model (1), we will suppose without loss of generality that $\mu = 0$. Under the identification condition

$$E_\alpha \left\{ m_{\alpha} \left( X_{\alpha} \right) \right\} = 0, \tag{2}$$

we have

$$m_\alpha(x_\alpha) = \int m\left(\mathbf{x}\right) dF\left(\mathbf{x}_{(-\alpha)}\right) \tag{3}$$

which can also be written as

$$m_\alpha(x_\alpha) = \int m\left(\mathbf{x}\right) f\left(\mathbf{x}_{(-\alpha)}\right) d\mathbf{x}_{(-\alpha)} \tag{4}$$

where $\mathbf{x}_{(-\alpha)}$ denotes the vector $\mathbf{x}$ with the $\alpha$th component removed.

If one uses (3) with $m\left(\mathbf{x}\right)$ estimated by the Nadaraya-Watson (see Nadaraya 1964 and Watson 1964) estimator (or local polynomial estimator), and $F\left(\mathbf{x}_{(-\alpha)}\right)$estimated by its empirical estimator, one obtains the marginal integration estimator

$$\tilde{m}_\alpha(x_\alpha) = \frac{1}{n}\sum_{i=1}^{n} \hat{m}(x_\alpha, \mathbf{X}_{i(-\alpha)}) \tag{5}$$

In this work, we are going to consider a slightly different estimator. We will use (4) with $m\left(\mathbf{x}\right)$ estimated by the Nadaraya-Watson estimator of regression and $f\left(\mathbf{x}_{(-\alpha)}\right)$ estimated by the Parzen-Rosenblatt (see Parzen 1962 and Rosenblatt 1956) kernel density estimator. Here, the estimator we consider is defined by the empirical analog of (4):

$$\hat{m}_\alpha(x_\alpha) = \int \hat{m}\left(\mathbf{x}\right) \hat{f}\left(\mathbf{x}_{(-\alpha)}\right) d\mathbf{x}_{(-\alpha)}$$

where the integral is approximated by any reasonable numerical integration procedure. In order to distinguish our marginal integration $\hat{m}$ from $\tilde{m}$, we will call it the smoothed marginal integration estimator.

Although the primary motivation for using the smoothed marginal integration estimator is theoretical as will be pointed out later, in some cases there is an additional computational advantage for doing so. Indeed, as discussed by Chèze, Poggi and Portier (2000), such an estimator can be costless in computation when the number of observations is large relative to the number of regressors. In that case, it requires less operations to approximate the integral on $\mathbf{x}_{(-\alpha)}$ than to estimate the expectation under $\mathbf{x}_{(-\alpha)}$ by its empirical counterpart.

In order to assess theoretically the non-robustness properties of the smoothed marginal integration estimator, we're going to analyze it from a functional point of view. We define as in Aït-Sahalia (1995) the smoothed cumulative distribution function of the random vector $(\mathbf{X}, Y)$ by

$$\hat{F}_n(\mathbf{x}, y) = \frac{1}{n} \sum_{i=1}^{n} K_I \left( \frac{\mathbf{x} - \mathbf{X}_i}{h} \right) K_I \left( \frac{y - Y_i}{h} \right)$$

where

$$K_I(\mathbf{x}) = \prod_{\alpha=1}^{d} K_I(x_\alpha)$$

and

$$K_I(x_\alpha) = \int_{-\infty}^{x_\alpha} K(t)\, dt$$

with $K$ a univariate kernel whose properties will be made precise in Appendix. The smoothed marginal integration estimator can then be regarded as the plug-in of $\hat{F}_n$ in the functional. It was shown in Aït-Sahalia (1995) that under technical assumptions that this smoothed cumulative distribution function converges to the cumulative distribution function of the random vector $(\mathbf{X}, Y)$ denoted by $F(\mathbf{x}, y)$. Furthermore, under regularity assumptions, Aït-Sahalia (1995) showed that the derivatives of $\hat{F}_n(\mathbf{x}, y)$ which are in fact classical Parzen-Rosenblatt estimators of density converge to corresponding marginal densities of the random vector $(\mathbf{X}, Y)$.

Having said all the above, both $m_\alpha$ and $\hat{m}_\alpha$ admit a tidy functional representation. Indeed, (3) can be written under the form

$$m_\alpha(x_\alpha) = \Gamma_\alpha(F) = \int \left[ \left\{ \frac{\int y \frac{\partial^{d+1} F}{\partial \mathbf{x} \partial y}(\mathbf{x}, y)\, dy}{\int \frac{\partial^{d+1} F}{\partial \mathbf{x} \partial y}(\mathbf{x}, y)\, dy} \right\} \int \frac{\partial^{d+1} F}{\partial \mathbf{x} \partial y}(\mathbf{x}, y)\, dx_\alpha dy \right] d\mathbf{x}_{(-\alpha)}.$$

We then have

$$\hat{m}_\alpha(x_\alpha) = \Gamma_\alpha\left(\hat{F}_n\right)$$

This functional presentation allows us to quantify the (non)robustness properties of the smoothed marginal integration estimator by adapting Hampel's influence function (see Hampel 1994) to this nonparametric framework.

We now define the analog of usual first order Taylor expansion for functionals such as $\Gamma$. Given a normed linear space $(E, \|\circ\|)$, a real valued functional $\Gamma$ defined on an open neighborhood of the point $F \in E$ is said to be strongly Frechet differentiable at the point $F$ if there exists a continuous linear operator $D_F\Gamma : E \to R$ such that

$$\Gamma(F+V) = \Gamma(F) + D_F\Gamma(V) + O\left(\|V\|^2\right)$$

holds for all $V$ satisfying $\|V\| \to 0$. The linear operator $D_F\Gamma$ is called the Frechet differential of the functional $\Gamma$ at the point $F$. In the next section, we will establish the Frechet differentiability of $\Gamma_\alpha$ and then use it to assess the non-robustness of the smoothed marginal integration estimator via an adapted form of Hampel's influence function.

## 2.2 Non-robustness of marginal integration estimator

Let's now state the first lemma which is going to reveal the asymptotic behavior of the global estimator. Let $E$ denote the linear space of bounded real-valued functions defined on $R^{d+1}$ whose partial derivatives up to order $d+1$ are continuous with compact support. For any $V \in E$, define the Sobolev norm of order $d+1$ as

$$\|V\|_{(\infty,d+1)} = \max_{0\le c\le\le d+1} \max_{\Delta,|\Delta|=c} \sup_{u\in R^{d+1}} \left|\partial^\Delta V(u)\right|$$

where $|\Delta| = \sum_{j=1}^{d+1} \Delta_j$ for $\Delta = (\Delta_1, ..., \Delta_{d+1}) \in N^{d+1}$and $\partial^\Delta V(u) = \frac{\partial^{|\Delta|}V(u)}{\partial^{\Delta_1}V(u_1)\cdots\partial^{\Delta_{d+1}}(u_{d+1})}$.

**Lemma 1** *The functional $\Gamma_\alpha$ admits a Frechet differential for the norm $\|\circ\|_{(\infty,d+1)}$ at every cumulative distribution $F \in E$ satisfying assumption (A3). This Frechet differential is given by*

$$D\Gamma_\alpha(V) = \int \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})}\left[\int yv(\mathbf{x},y)dy - v(\mathbf{x})m(\mathbf{x})\right]d\mathbf{x}_{(-\alpha)} + \int m(\mathbf{x})v(\mathbf{x}_{(-\alpha)})d\mathbf{x}_{(-\alpha)}$$

*in which* $v(\mathbf{x},y) = \frac{\partial^{d+1}V}{\partial\mathbf{x}\partial y}(\mathbf{x},y)\,, v(\mathbf{x}) = \int v(\mathbf{x},y)\,dy, v(\mathbf{x}_{(-\alpha)}) = \int v(\mathbf{x})\,dx_\alpha.$

In the case when $V = \hat{F}_n - F, v = \hat{f}_n - f$, we show that the term $\int m(\mathbf{x})v(\mathbf{x}_{(-\alpha)})d\mathbf{x}_{(-\alpha)}$ is of higher order than the first one, and hence one obtains

**Corollary 1** *Under assumptions (A1) to (A5), the following expansion holds :*

$$\hat{m}_\alpha(x_\alpha) - m_\alpha(x_\alpha) = \frac{1}{n}\sum_{i=1}^{n}\Psi(\mathbf{X}_i, Y_i; h) + o_p(n^{-1/2}h^{-1/2}) \qquad (6)$$

*where*

$$\Psi(\mathbf{u_X}, u_Y; h) = \int \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} \left[ \begin{array}{c} \int y \left(K_h(\mathbf{x}-\mathbf{u_X})K_h(y-u_Y) - f(\mathbf{x},y)\right) dy \\ -\left(K_h(\mathbf{x}-\mathbf{u_X}) - f(\mathbf{x})\right) m(\mathbf{x}) \end{array} \right] d\mathbf{x}_{(-\alpha)}$$

Clearly, according to expression (6), each data point $(\mathbf{X}_i, Y_i)$ contributes to the asymptotic error of estimating $m_\alpha(x_\alpha)$ by $\hat{m}_\alpha(x_\alpha)$ through the term $\Psi(\mathbf{X}_i, Y_i; h)$. As pointed in Tamine (2002), $\Psi(\mathbf{X}_i, Y_i; h)$ can be considered as an adaptation of Hampel's influence function to kernel estimators. As a consequence, the analysis of $\Psi(\mathbf{X}_i, Y_i; h)$ is interesting in order to quantify the robustness properties of the estimator $\hat{m}_\alpha(x_\alpha)$ when outliers may be present in the sample. This is done in the following lemma:

**Theorem 1** *Under assumptions* (*A*1) *to* (*A*5), $\Psi(\mathbf{u_X}, u_Y; h)$ *is bounded in the variable* $\mathbf{u_X}$ *(*$u_Y$ *being bounded) and unbounded in the variable* $u_Y$ *(*$\mathbf{u_X}$ *being bounded).*

This lemma shows us that outliers among the regressors $\mathbf{X}_i$ can only have bounded effect on the error of estimation, this is due to the local character of the estimator. On the other hand, outliers among the response variable $Y_i$ can increase the error of estimation without any limit, which in practice can produce a fictitious peak.

The next step of this analysis is to derive the asymptotic distribution of the global estimator. This is done using a central limit theorem. The results are summarized in the following lemma:

**Theorem 2** *Under assumptions* (*A*1) *to* (*A*5) *:*

$$\sqrt{nh}\left\{\hat{m}_\alpha(x_\alpha) - m_\alpha(x_\alpha)\right\} \xrightarrow{\mathcal{L}} N(0, V)$$

*where*

$$V = \int \left\{ \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} \right\}^2 \{y - m(\mathbf{x})\}^2 f(\mathbf{x}, y) d\mathbf{x}_{(-\alpha)} dy \cdot \int K^2(t) dt$$

# 3 . M robustified additive model

## 3.1 Model specification

As discussed in the previous section, the marginal integration estimator is non-robust in the presence of response outliers. In this section, we consider an estimation procedure of the additive components $m_\alpha(x_\alpha)$ using conditional M-estimates.

In what follows, we assume that the conditional cumulative distribution function $F_{\mathbf{x}}(y) = \int_{-\infty}^{y} \frac{f(\mathbf{x}, v)}{f(\mathbf{x})} dv$ is symmetric around $m(\mathbf{x})$.

Let a score function be $\Psi$ which is odd, bounded, twice continuously differentiable on $R$ and satisfies the inequality $\int \frac{\partial \Psi}{\partial t}\left(y - m(\mathbf{x})\right) dF_{\mathbf{x}}(y) > 0.$

It can be easily shown that $m(\mathbf{x})$ is the unique $\theta$ value that solves the equation

$$\int \Psi\left(y-\theta\right)\frac{f(\mathbf{x},y)}{f(\mathbf{x})}dy = 0 \qquad (7)$$

Remark :

- our regularity assumptions on $\Psi$ are stronger than those that are usually required for the study of the M estimator of the regression. This assumption could be relaxed at the cost of increased complexity in the proofs.
- If $F_{\mathbf{x}}(y)$ is not symmetric, equation (7) still admits a unique solution which is no longer $m(\mathbf{x})$, but can nevertheless be useful as a robust location parameter (see Huber 1981 for discussion).

Under the additive specification (1) for $m(\mathbf{x})$ and the identification condition (2) , $m_\alpha(x_\alpha)$ still satisfies the relation

$$m_\alpha(x_\alpha) = \int m(\mathbf{x})\, f\left(\mathbf{x}_{(-\alpha)}\right) d\mathbf{x}_{(-\alpha)}$$

and is still a functional of the cumulative distribution function $F(\mathbf{x},y)$ . This functional that we will denote by $\Gamma_\alpha^M$ doesn't have a closed form since $m(\mathbf{x})$ is implicitly defined. Nevertheless, this functional presentation is still useful in order to establish the robustness properties of the estimator.

## 3.2 Estimation procedure

The estimator we propose is still of marginal integration type. It proceeds in two steps:

- first an estimator of $m(\mathbf{x})$ is computed by solving equation

$$\int \Psi\left(y-\theta\right)\frac{\hat{f}(\mathbf{x},y)}{\hat{f}(\mathbf{x})}dy = 0 \qquad (8)$$

where $\hat{f}(\mathbf{x},y)$ and $\hat{f}(\mathbf{x})$ are the Parzen-Rosenblatt estimators of $f(\mathbf{x},y)$ and $f(\mathbf{x})$. The existence of this estimator for $n$ large enough as well as its convergence properties have discussed in detail in Tamine (2002).This first step estimator will be denoted $\hat{m}^M(\mathbf{x})$

- $m_\alpha(x_\alpha)$ is then estimated by

$$\hat{m}_\alpha^M(x_\alpha) = \int \hat{m}^M(\mathbf{x})\, \hat{f}\left(\mathbf{x}_{(-\alpha)}\right) d\mathbf{x}_{(-\alpha)}$$

- One has to notice that $\hat{m}_\alpha^M(x_\alpha)$ is in fact the plug in estimator of the functional

$$m_\alpha^M(x_\alpha) = \Gamma_\alpha^M(F)$$

at $\hat{F}_n(\mathbf{x}, y)$.

## 3.3 Asymptotic analysis

**Lemma 2** *The functional $\Gamma_\alpha^M$ admits a Frechet differential for the norm $\|\circ\|_{(\infty,d+1)}$ at every cumulative distribution $F \in E$ satisfying assumption $(A3)$. This Frechet differential is given by*

$$D\Gamma_\alpha^M(V) = \int \frac{\int \Psi\{y - m(\mathbf{x})\} v(\mathbf{x}, y) dy}{\int \Psi'\{y - m(\mathbf{x})\} f(\mathbf{x}, y) dy} f(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)} + \int m(\mathbf{x}) v(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)}$$

*in which* $v(\mathbf{x}, y) = \dfrac{\partial^{d+1} V}{\partial \mathbf{x} \partial y}(\mathbf{x}, y)$, $v(\mathbf{x}) = \int v(\mathbf{x}, y)\, dy$, $v(\mathbf{x}_{(-\alpha)}) = \int v(\mathbf{x})\, dx_\alpha$.

As previously, in the case $V = \hat{F}_n - F$ and $v = \hat{f}_n - f$, we show that the term $\int m(\mathbf{x}) v(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)}$ is of higher order than the first one, and hence one obtains

**Corollary 2** *Under assumptions $(A1)$ to $(A5)$*

$$\hat{m}_\alpha^M(x_\alpha) - m_\alpha(x_\alpha) = \frac{1}{n} \sum_{i=1}^{n} \Psi^M(\mathbf{X}_i, Y_i; h) + o_p(n^{-1/2} h^{-1/2})$$

*where*

$$\Psi^M(\mathbf{u}_\mathbf{X}, u_Y; h) = \int \frac{\int \Psi\{y - m(\mathbf{x})\} \{K_h(\mathbf{x} - \mathbf{u}_\mathbf{X}) K_h(y - u_Y) - f(\mathbf{x}, y)\} dy}{\int \Psi'\{y - m(\mathbf{x})\} f(\mathbf{x}, y) dy} f(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)}$$

We can now lead a more careful analysis of $\Psi^M(\mathbf{u}_\mathbf{X}, u_Y; h)$ in order to assess the robustness properties of the estimator $\hat{m}_\alpha^M(x_\alpha)$

**Theorem 3 :** *Under assumptions $(A1)$ to $(A5)$ $\Psi^M(\mathbf{u}_\mathbf{X}, u_Y; h)$ is bounded.*

This theorem shows us how the estimator $\hat{m}_\alpha^M(x_\alpha)$ improves in terms of robustness. Indeed no observations $(\mathbf{X}_i, Y_i)$ can have extreme effects on the asymptotic error of estimation.

We are now going to derive the asymptotic distribution of the robust estimator $m_\alpha^M(x_\alpha)$ :

**Theorem 4 :** *(A1) to (A5) :*

$$\sqrt{nh}\left\{\hat{m}_\alpha^M(x_\alpha) - m_\alpha(x_\alpha)\right\} \xrightarrow{\mathcal{L}} N\left(0, V^M\right)$$

*where*

$$V^M = \int\int\left[\frac{\Psi\left\{y - m(\mathbf{x})\right\}}{\int \Psi'\left\{y - m(\mathbf{x})\right\} f(\mathbf{x}, y)dy} f(\mathbf{x}_{(-\alpha)})\right]^2 f(\mathbf{x}, y)d\mathbf{x}_{(-\alpha)}dy \cdot \int K^2(t)dt$$

# 4 . Monte-Carlo simulation

The data generating process we have chosen is the following :

$$Y = m(\mathbf{X}) + \varepsilon$$

where $m$ is the same bivariate additive function used in Figure 1

$$m(\mathbf{x}) = 5x_1^2 - 5/3 + 10x_2^3$$

$\mathbf{X}$ is uniformly distributed on $[-1, 1]^2$ and $\varepsilon$ is a mixture of normal laws with probability density function

$$f(\varepsilon) = \frac{1-\nu}{\sqrt{2\pi}} exp\left(-\frac{\varepsilon^2}{2}\right) + \frac{\nu}{\sqrt{2\pi k}} exp\left(-\frac{\varepsilon^2}{2k}\right)$$

$\nu$ represents the degree of contamination of the standard normal law. This degree of contamination has been set to 10%, $k$ represents the variance of the contaminating law and has been set equal to 9.

We chose the score function

$$\Psi(e) = \begin{cases} e \text{ if } |e| \leq c_1 \\ \tilde{\Psi}(e) \text{ if } c_1 < |e| \leq c_2 \\ 0 \text{ if } |e| > c_2 \end{cases}$$

where $\tilde{\Psi}$ is a fifth order polynomial chosen in such a way that $\Psi$ is twice continuously differentiable and $c_1$, $c_2$ are trimming constants. Following Lucas (1996), we chose $c_1 = \sqrt{\chi^{-1}(0.99)}$ and $c_2 = \sqrt{\chi^{-1}(0.999)}$. If $Y_i - m(\mathbf{X})$ follows a standardized normal law, such a choice ensures that the observations for which $|Y_i - m(\mathbf{X})| > 3.5$ are discarded and the observations for which $3.5 > |Y_i - m(\mathbf{X})| > 2.8$ are downweighted.

In order to compare the performances of our estimators, we estimated the regression function on a grid $\{t_1, .....t_8\}$ equally spaced on $[-0.7; 0.7]$. This restriction is to avoid the boundary effects. Our criterion of comparison for each $\alpha = 1, 2$ is the sum of squared errors $SSE_\alpha = \frac{1}{8}\sum_{i=1}^{8}\left\{m_\alpha(t_i) - \hat{m}_\alpha(t_i)\right\}^2$ where $\hat{m}_\alpha$ will be either the ordinary marginal integration estimator or the M robustified one of the regression component function $m_\alpha$. Our results are consigned in Figure 2. Clearly, for both components $m_1$ and $m_2$, the robustified estimators have overall smaller SSE than the ordinary estimators. A typical example is seen in Figure 1 in the introduction. The ordinary estimators show greater bias, due to the influence of outliers. This corroborates with our theoretical results, Theorem 1 and Theorem 3.

# 5 Acknowledgements

All three authors received financial support from the Deutsche Forschungsgemeinschaft via Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonomischer Prozesse" at Humboldt-Universität zu Berlin. The first author was also supported by Université de la Méditerranée.

# Appendix

## A.1 Preliminaries

We are first going to give assumptions needed for the following proofs:

- $(A1)$ : The sample $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$ is a strictly stationary and $\beta$ mixing realizations of the vector $(\mathbf{X}, Y)$ satisfying $k^\delta \beta_k \to 0$ for some fixed $\delta > 1$. For definition of mixing, see Doukhan (1990).
- $(A2)$ : The density $f(\mathbf{x}, y)$ is compactly supported and admits continuous derivatives up to order $r$.
- $(A3)$ : The density $f(\mathbf{x})$ admits a strictly positive lower bound, $c$.
- $(A4)$ : The univariate kernel $K$ is an even kernel of order $r$.
- $(A5)$ : The bandwidth satisfies $\lim_{n\to\infty} h = 0$ (the dependence of $h$ on $n$ is left implicit for the simplicity of notations) in such a way that $\lim_{n\to\infty} n^{\frac{1}{2}} h^{2d+\frac{3}{2}} \to \infty$

and $\lim_{n\to\infty} n^{\frac{1}{2}} h^{r+\frac{1}{2}} \to 0$.

Assumption $(A2)$ is stronger than what is usually required due to the use of the $\|\circ\|_{L(\infty,d+1)}$ Sobolev norm for studying the remainder terms. This assumption could easily be weakened by employing sharper bounds on the remainder. In that case, the estimator couldn't be studied in such generality.

We now introduce notations and establish basic results needed in the proofs of main results. In the space $E$ (see section), $U$ will be the neighborhood of zero constituted by the functions $V$ satisfying $\|V\|_{L(\infty,d+1)} < \frac{c}{2}$. For all $V \in U$, and for all $t \in [0,1]$, applying a triangular inequality to $|f(\mathbf{x})+tv(\mathbf{x})|$ gives

$$|f(\mathbf{x})+tv(\mathbf{x})| > \frac{c}{2} > 0 \tag{A.1}$$

so that the real valued function

$$\varphi_{\mathbf{x},y}(t) = \frac{f(\mathbf{x},y)+tv(\mathbf{x},y)}{f(\mathbf{x})+tv(\mathbf{x})}$$

is well defined. Differentiation with respect to $t$ yields

$$\varphi'_{\mathbf{x},y}(t) = \frac{v(\mathbf{x},y)f(\mathbf{x})-f(\mathbf{x},y)v(\mathbf{x})}{\{f(\mathbf{x})+tv(\mathbf{x})\}^2}$$

and

$$\varphi''_{\mathbf{x},y}(t) = -2v(\mathbf{x})\frac{v(\mathbf{x},y)f(\mathbf{x})-f(\mathbf{x},y)v(\mathbf{x})}{\{f(\mathbf{x})+tv(\mathbf{x})\}^3}$$

Using inequality $(A.1)$, it follows immediately that for some constant $C(F,c)$ :

$$\left|\varphi'_{\mathbf{x},y}(t)\right| \leq C(F,c)\,.\,\|V\|_{L(\infty,d+1)} \tag{A.2}$$

and

$$\left|\varphi''_{\mathbf{x},y}(t)\right| \leq C(F,c)\,.\,\|V\|^2_{L(\infty,d+1)} \tag{A.3}$$

## A.2 Proof of Lemma 1

$$\Gamma_\alpha(F) = \int \frac{\int y f(\mathbf{x},y)dy}{f(\mathbf{x})} f(\mathbf{x}_{(-\alpha)})d\mathbf{x}_{(-\alpha)}$$

and for $t \in [0,1]$, let's denote by $\gamma_\alpha$ the function

$$\gamma_\alpha(t) = \Gamma_\alpha(F+tV).$$

where $V \in U$ is as defined in subsection A.1. With notations of subsection A.1, we have

$$\gamma_\alpha(t) = \int \left\{\int y\varphi_{\mathbf{x},y}(t)dy\right\}\left\{f(\mathbf{x}_{(-\alpha)})+tv(\mathbf{x}_{(-\alpha)})\right\}d\mathbf{x}_{(-\alpha)}.$$

Differentiating with respect to $t$ yields:

$$\begin{aligned}\gamma'_\alpha(t) &= \int \left\{\int y\varphi'_{\mathbf{x},y}(t)dy\right\}\left\{f(\mathbf{x}_{(-\alpha)})+tv(\mathbf{x}_{(-\alpha)})\right\}d\mathbf{x}_{(-\alpha)} \\ &\quad + \int\int y\varphi_{\mathbf{x},y}(t)dy v(\mathbf{x}_{(-\alpha)})d\mathbf{x}_{(-\alpha)}\end{aligned}$$

and

$$\gamma_\alpha''(t) = 2\int\int y\varphi_{\mathbf{x},y}'(t)dy v(\mathbf{x}_{(-\alpha)})d\mathbf{x}_{(-\alpha)}+\int\left\{\int y\varphi_{\mathbf{x},y}''(t)dy\right\}\left\{f(\mathbf{x}_{(-\alpha)})+tv(\mathbf{x}_{(-\alpha)})\right\}d\mathbf{x}_{(-\alpha)}. \tag{A.4}$$

A Taylor expansion of $\gamma_\alpha$ between 0 and 1 gives us :

$$\Gamma_\alpha(F+V) = \Gamma_\alpha(F) + \gamma_\alpha'(0) + \gamma''(\bar{t})$$

where $\bar{t} \in ]0,1[$. Using the expression $(A.4)$ , inequalities $(A.2)$ and $(A.3)$ one obtains

$$\gamma''(\bar{t}) = O\left(\|V\|_{L(\infty,d+1)}^2\right).$$

With

$$\gamma_\alpha'(0) = \int\frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})}\left\{\int yv(\mathbf{x},y)dy - v(\mathbf{x})m(\mathbf{x})\right\}d\mathbf{x}_{(-\alpha)} + \int m(\mathbf{x})v(\mathbf{x}_{(-\alpha)})d\mathbf{x}_{(-\alpha)},$$

we get

$$\begin{aligned}\Gamma_\alpha(F+V) &= \Gamma_\alpha(F) + \int\frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})}\left\{\int yv(\mathbf{x},y)dy - v(\mathbf{x})m(\mathbf{x})\right\}d\mathbf{x}_{(-\alpha)} \quad (A.5)\\ &\quad + \int m(\mathbf{x})v(\mathbf{x}_{(-\alpha)})d\mathbf{x}_{(-\alpha)} + O\left(\|V\|_{L(\infty,d+1)}^2\right).\end{aligned}$$

## A.3 Proof of Corollary 1

Using assumptions $(A1)$ to $(A5)$, one has $\left\|\hat{F}_n - F\right\|_{L(\infty,d+1)} \to^{as} 0$, according to Aït-Sahalia (1995). Hence for large enough $n$, $V_n = \hat{F}_n - F \in U$ and one can use equality $(A.5)$ in order to obtain

$$\begin{aligned}&\hat{m}_\alpha(x_\alpha) = m_\alpha(x_\alpha) + \\ &\frac{1}{n}\sum_{i=1}^n\int\frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})}\left[\begin{array}{c}\int y\left(K_h(\mathbf{x}-\mathbf{X}_i)K_h(y-Y_i) - f(\mathbf{x},y)\right)dy \\ -\left(K_h(\mathbf{x}-\mathbf{X}_i) - f(\mathbf{x})\right)m(\mathbf{x})\end{array}\right]d\mathbf{x}_{(-\alpha)} + \\ &\frac{1}{n}\sum_{i=1}^n\int m(\mathbf{x})\left\{K_h(\mathbf{x}_{(-\alpha)} - \mathbf{X}_{(-\alpha)i}) - f\left(\mathbf{x}_{(-\alpha)}\right)\right\}d\mathbf{x}_{(-\alpha)} + O_p\left(\|V_n\|_{L(\infty,d+1)}^2\right). \quad (A.6)\end{aligned}$$

We are now going to study the term

$$\begin{aligned}T_{1n} &= \frac{1}{n}\sum_{i=1}^n\int m(\mathbf{x})\left\{K_h(\mathbf{x}_{(-\alpha)} - \mathbf{X}_{(-\alpha)i}) - f\left(\mathbf{x}_{(-\alpha)}\right)\right\}d\mathbf{x}_{(-\alpha)} = \\ &\int m(\mathbf{x})\left(\hat{f}\left(\mathbf{x}_{(-\alpha)}\right) - f\left(\mathbf{x}_{(-\alpha)}\right)\right)d\mathbf{x}_{(-\alpha)}.\end{aligned}$$

Using an integration by parts, we obtain :

$$|T_{1n}| = \left|\int m'(x)\left\{\hat{F}(\mathbf{x}_{(-\alpha)}) - F(\mathbf{x}_{(-\alpha)})\right\}d\mathbf{x}_{(-\alpha)}\right|$$

so that

$$|T_{1n}| = O_p\left(\sup_{\mathbf{x}_{(-\alpha)}} \left|\hat{F}(\mathbf{x}_{(-\alpha)}) - F(\mathbf{x}_{(-\alpha)})\right|\right).$$

But, using Lemma 1 of Aït-Sahalia (1995), under assumptions $(A1)$ to $(A4)$, we have

$$\sup_{\mathbf{x}_{(-\alpha)}} \left|\hat{F}(\mathbf{x}_{(-\alpha)}) - F(\mathbf{x}_{(-\alpha)})\right| = O_p(n^{-\frac{1}{2}} + h^r)$$

so that

$$T_{1n} = O_p\left(n^{-\frac{1}{2}} + h^r\right)$$

and, under assumption $(A5)$, we obtain

$$(nh)^{\frac{1}{2}} T_{1n} = o_p(1).$$

Now, under assumptions $(A1)$ to$(A5)$, we have

$$\left\|\hat{F}_n - F\right\|^2_{L(\infty,d+1)} = O_p\left(n^{-1}h^{-2(d+1)} + h^{2r}\right)$$

and using expression $(A.6)$, we get

$$\begin{aligned}\hat{m}_\alpha(x_\alpha) &= m_\alpha(x_\alpha) + \frac{1}{n}\sum_{i=1}^{n}\int \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})}\left[\begin{array}{c}\int y\,(K_h(\mathbf{x}-\mathbf{X}_i)K_h(y-Y_i) - f(\mathbf{x},y))\,dy \\ -\,(K_h(\mathbf{x}-\mathbf{X}_i) - f(\mathbf{x}))\,m(\mathbf{x})\end{array}\right] d\mathbf{x}_{(-\alpha)} \\ &\quad + o_p((nh)^{-\frac{1}{2}}) + O_p\left(n^{-1}h^{-2(d+1)} + h^{2r}\right).\end{aligned}$$

But, under assumption $(A5)$, we have $(nh)^{\frac{1}{2}}O_p\left(n^{-1}h^{-2(d+1)} + h^{2r}\right) = o_{p(}(nh)^{-\frac{1}{2}})$ so that

$$\begin{aligned}\hat{m}_\alpha(x_\alpha) &= m_\alpha(x_\alpha) + \frac{1}{n}\sum_{i=1}^{n}\int \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})}\left[\begin{array}{c}\int y\,(K_h(\mathbf{x}-\mathbf{X}_i)K_h(y-Y_i) - f(\mathbf{x},y))\,dy \\ -\,(K_h(\mathbf{x}-\mathbf{X}_i) - f(\mathbf{x}))\,m(\mathbf{x})\end{array}\right] d\mathbf{x}_{(-\alpha)} \\ &\quad + o_p((nh)^{-\frac{1}{2}}).\end{aligned}$$

Thus we have proved the lemma.

## A.4 Proof of Theorem 1

We have:

$$\Psi(\mathbf{u_X}, u_Y; h) = \int \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})}\left[\begin{array}{c}\int y\,(K_h(\mathbf{x}-\mathbf{u_X})K_h(y-u_Y) - f(\mathbf{x},y))\,dy \\ -\,(K_h(\mathbf{x}-\mathbf{u_X}) - f(\mathbf{x}))\,m(\mathbf{x})\end{array}\right] d\mathbf{x}_{(-\alpha)}$$

Let's first restrict $u_Y$ to a compact set. Since the kernel $K$ is compactly supported and all integrations are made over a compact set, the boundedness in $\mathbf{u_X}$ is immediate.

Let's now restrict $\mathbf{u_X}$ to a compact set. The dependence on $u_Y$ only comes from the term

$$\int \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} \left[\int y \left\{K_h(\mathbf{x}-\mathbf{u_X})K_h(y-u_Y)\right\} dy\right] d\mathbf{x}_{(-\alpha)}.$$

In the integration with respect to $y$, we make the change of variable $v = \dfrac{y-u_Y}{h}$, with $\int vK(v)dv = 0$ and $\int K(v)dv = 1$, we obtain

$$\int \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} \left[\int y \left\{K_h(\mathbf{x}-\mathbf{u_X})K_h(y-u_Y)\right\} dy\right] d\mathbf{x}_{(-\alpha)} = u_y \int \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} K_h(\mathbf{x}-\mathbf{u_X}) d\mathbf{x}_{(-\alpha)}$$

which is an unbounded function of $u_Y$. Hence we have completed the proof of the theorem.

## A.5 Proof of Theorem 2

In expansion (6), consider the term

$$T_{2n} = \frac{1}{n}\sum_{i=1}^{n} \int \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} \left[\begin{array}{c} \int y \left\{K_h(\mathbf{x}-\mathbf{X}_i)K_h(y-Y_i) - f(\mathbf{x},y)\right\} dy \\ -\left\{K_h(\mathbf{x}-\mathbf{X}_i) - f(\mathbf{x})\right\} m(\mathbf{x}) \end{array}\right] d\mathbf{x}_{(-\alpha)}.$$

Let's first study the its mean

$$B_{2n} = \int\int\int \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} \left\{\int yK_h(\mathbf{x}-\mathbf{u})K_h(y-v)dy - K_h(\mathbf{x}-\mathbf{u})m(\mathbf{x})\right\} f(\mathbf{u},v)d\mathbf{x}_{(-\alpha)}d\mathbf{u}dv$$

Using the change of variable $\xi = \dfrac{\mathbf{x}-\mathbf{u}}{h}$ in the integration with respect to $\mathbf{u}$ and $\eta = \dfrac{y-v}{h}$ in the integration with respect to $v$, we obtain

$$B_{2n} = \int \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} \left[\int\int\int \left\{yK(\xi)K(\eta) - K(\xi)m(\mathbf{x})\right\} f(\mathbf{x}-h\xi, y-h\eta)d\xi d\eta dy\right] d\mathbf{x}_{(-\alpha)}.$$

Using a Taylor expansion of $f(\mathbf{x}-h\xi, y-h\eta)$ up to order $r$, and using assumption $(A4)$, we obtain

$$B_{2n} = \int \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} \left\{\int yf(\mathbf{x},y)dy - f(\mathbf{x})m(\mathbf{x})\right\} d\mathbf{x}_{(-\alpha)} + O(h^r)$$

so that, under assumption $(A5)$, we have

$$(nh)^{\frac{1}{2}} \left(B_{2n} - \int \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} \left\{\int yf(\mathbf{x},y)dy - f(\mathbf{x})m(\mathbf{x})\right\} d\mathbf{x}_{(-\alpha)}\right) = o(1)$$

i.e.

$$(nh)^{\frac{1}{2}} E\left(T_{2n}\right) = o(1).$$

We now have to study the variance term normalized at rate $(nh)^{\frac{1}{2}}$ :

$$(nh)^{\frac{1}{2}} V_{2n} = \sum_{i=1}^{n} \sqrt{\frac{h}{n}} \left[ \int \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} K_h(\mathbf{x} - \mathbf{X}_i) \left\{ \int y K_h(y - Y_i) dy - m(\mathbf{x}) \right\} d\mathbf{x}_{(-\alpha)} - B_{2n} \right]$$

Let's define $\mathcal{F}_{n,i}$ as the $\sigma$ field generated by $\{\mathbf{X}_j, Y_j\}_{j=1,\dots i}$ and let's define the random variable

$$Z_{n,i} = \sqrt{\frac{h}{n}} \left[ \int \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} K_h(\mathbf{x} - \mathbf{X}_i) \left\{ \int y K_h(y - Y_i) dy - m(\mathbf{x}) \right\} d\mathbf{x}_{(-\alpha)} - B_{2n} \right]$$

so that $(nh)^{\frac{1}{2}} V_{2n} = \sum_{j=1}^{n} Z_{n,j}$. The array $\left\{ \sum_{j=1}^{i} Z_{n,j}, \mathcal{F}_{n,i}, 1 \le i \le n, n \ge 1 \right\}$ is a zero mean, square integrable martingale array. We have

$$\begin{aligned} \sum_{j=1}^{n} E(Z_{n,j}^2) &= h \int \int \left[ \int \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} K_h(\mathbf{x} - \mathbf{u}) \left\{ \int y K_h(y - v) dy - m(\mathbf{x}) \right\} d\mathbf{x}_{(-\alpha)} \right]^2 . \\ & \quad f(\mathbf{u}, v) d\mathbf{u} dv - h(B_{2n})^2. \end{aligned}$$

With the change of variable $\xi_\alpha = \dfrac{x_\alpha - u_\alpha}{h}$ in the integration with respect to $u_\alpha$ and $\xi_{(-\alpha)} = \dfrac{\mathbf{x}_{(-\alpha)} - \mathbf{u}_{(-\alpha)}}{h}$ in the integration with respect to $\mathbf{x}_{(-\alpha)}$, we obtain

$$\begin{aligned} \sum_{j=1}^{i} E(Z_{n,j}^2) &= h \int \int \left[ \begin{array}{c} \int \frac{f(\mathbf{u}_{(-\alpha)} + h\xi_{(-\alpha)})}{f(x_\alpha, \mathbf{u}_{(-\alpha)} + h\xi_{(-\alpha)})} K(\xi_{(-\alpha)}) . \\ \left\{ \int y K_h(y - v) dy - m(x_\alpha, \mathbf{u}_{(-\alpha)} + h\xi_{(-\alpha)}) \right\} d\xi_{(-\alpha)} \end{array} \right]^2 . \\ & \quad K^2(\xi_\alpha) f(x_\alpha - h\xi_\alpha, \mathbf{u}_{(-\alpha)}, v) d\xi_\alpha d\mathbf{u}_{(-\alpha)} dv - h(B_{2n})^2. \end{aligned}$$

Now, we have $\int y K_h(y - v) dy = v$ so that

$$\begin{aligned} \lim_{n \to \infty} E\left( \sum_{j=1}^{n} Z_{n,j}^2 \right) &= \int \int \left( \frac{f(\mathbf{u}_{(-\alpha)})}{f(x_\alpha, \mathbf{u}_{(-\alpha)})} \right)^2 \left( v - m(x_\alpha, \mathbf{u}_{(-\alpha)}) \right)^2 . \\ & \quad f(x_\alpha, \mathbf{u}_{(-\alpha)}, v) d\mathbf{u}_{(-\alpha)} dv \cdot \int K^2(\xi_\alpha) d\xi_\alpha \end{aligned}$$

and

$$\begin{aligned} \sum_{j=1}^{n} E\left( Z_{n,j}^2 / \mathcal{F}_{n,j-1} \right) &\to {}^{p} \int \int \left\{ \frac{f(\mathbf{u}_{(-\alpha)})}{f(x_\alpha, \mathbf{u}_{(-\alpha)})} \right\}^2 \left\{ v - m(x_\alpha, \mathbf{u}_{(-\alpha)}) \right\}^2 . \\ & \quad f(x_\alpha, \mathbf{u}_{(-\alpha)}, v) d\mathbf{u}_{(-\alpha)} dv \cdot \int K^2(\xi_\alpha) d\xi_\alpha \end{aligned}$$

Under assumptions $(A1)$ to $(A5)$, conditions 3.19 and 3.20 of Corollary 3.1 of Hall and Heyde (1981, pp 58) are satisfied, so that we finally get

$$(nh)^{\frac{1}{2}}V_{2n} \to \mathcal{N}(0,V)$$

where

$$V = \int\int\left\{\frac{f(\mathbf{u}_{(-\alpha)})}{f(x_\alpha,\mathbf{u}_{(-\alpha)})}\right\}^2\left\{v - m(x_\alpha,\mathbf{u}_{(-\alpha)})\right\}^2 f(x_\alpha,\mathbf{u}_{(-\alpha)},v)d\mathbf{u}_{(-\alpha)}dv\cdot\int K^2(\xi_\alpha)d\xi_\alpha.$$

## A.6 Proof of Lemma 2

As noticed in Section 3, $m(\mathbf{x})$ is implicitly defined by

$$\int\Psi\left\{y - m(\mathbf{x})\right\}\frac{f(\mathbf{x},y)}{f(\mathbf{x})}dy = 0$$

which is a functional of $F$ that we will denote as $\Gamma^M(F)$. Define also the function $\gamma^M(t) = \Gamma^M(F+tV)$. Applying the implicit function theorem as in Tamine (2002),, we obtain:

$$\gamma^{M\prime}(t) = \frac{\int\Psi\left\{y-\gamma^M(t)\right\}\varphi'_{\mathbf{x},y}(t)dy}{\int\Psi'\left\{y-\gamma^M(t)\right\}\varphi_{\mathbf{x},y}(t)dy}$$

and

$$\begin{aligned}\gamma^{M\prime\prime}(t) \;=\; & \left\{\begin{array}{c}\left[-\gamma^{M\prime}(t)\int\Psi'\left\{y-\gamma^M(t)\right\}\varphi'_{\mathbf{x},y}(t)dy + \int\Psi\left\{y-\gamma^M(t)\right\}\varphi''_{\mathbf{x},y}(t)dy\right]\\ \cdot\int\Psi'\left\{y-\gamma^M(t)\right\}\varphi_{\mathbf{x},y}(t)dy\end{array}\right.\\ & -\left[\begin{array}{c}-\gamma^{M\prime}(t)\int\Psi''\left\{y-\gamma^M(t)\right\}\varphi_{\mathbf{x},y}(t)dy\\ +\int\Psi'\left\{y-\gamma^M(t)\right\}\varphi'_{\mathbf{x},y}(t)dy\end{array}\right]\int\Psi\left\{y-\gamma^M(t)\right\}\varphi'_{\mathbf{x},y}(t)dy\Bigg\}.\\ & \left[\int\Psi'\left\{y-\gamma^M(t)\right\}\varphi_{\mathbf{x},y}(t)dy\right]^{-2}.\end{aligned}$$

Using arguments nearly identical to Theorem 3 of Tamine (2002), one obtains

$$\gamma^{M\prime}(t) = O\left(\|V\|_{L(\infty,d+1)}\right) \tag{A.7}$$

and

$$\gamma^{M\prime\prime}(t) = O\left(\|V\|^2_{L(\infty,d+1)}\right). \tag{A.8}$$

We are now ready to study the term

$$\Gamma^M_\alpha(F) = \int\Gamma^M(F)f(\mathbf{x}_{(-\alpha)})d\mathbf{x}_{(-\alpha)}.$$

First let's define the function $\gamma_\alpha^M(t) = \Gamma_\alpha^M(F + tV)$. We have

$$\gamma_\alpha^M(t) = \int \gamma^M(t) \left\{ f(\mathbf{x}_{(-\alpha)}) + tv(\mathbf{x}_{(-\alpha)} \right\} d\mathbf{x}_{(-\alpha)}$$

so that

$$\gamma_\alpha^{M\prime}(t) = \int \gamma^{M\prime}(t) \left\{ f(\mathbf{x}_{(-\alpha)}) + tv(\mathbf{x}_{(-\alpha)} \right\} d\mathbf{x}_{(-\alpha)} + \int \gamma^M(t) v(\mathbf{x}_{(-\alpha)} d\mathbf{x}_{(-\alpha)}$$

and

$$\gamma_\alpha^{M\prime\prime}(t) = \int \gamma^{M\prime\prime}(t) \left\{ f(\mathbf{x}_{(-\alpha)}) + tv(\mathbf{x}_{(-\alpha)} \right\} d\mathbf{x}_{(-\alpha)} + 2\int \gamma^{M\prime}(t) v(\mathbf{x}_{(-\alpha)} d\mathbf{x}_{(-\alpha)}. \quad \text{(A.9)}$$

A Taylor expansion of $\gamma_\alpha^M$ between 0 and 1 gives us

$$\Gamma_\alpha^M(F + V) = \Gamma_\alpha^M(F) + \gamma_\alpha^{M\prime}(0) + \gamma''(\bar{t})$$

where $\bar{t} \in (0, 1)$. Using the expression $(A.9)$, inequalities $(A.7)$ and $(A.8)$ one obtains

$$\gamma''(\bar{t}) = O\left( \|V\|^2_{L(\infty, d+1)} \right).$$

With

$$\gamma_\alpha^{M\prime}(0) = \int \frac{\int \Psi\left\{y - m(\mathbf{x})\right\} v(\mathbf{x}, y) dy}{\int \Psi'\left\{y - m(\mathbf{x})\right\} f(\mathbf{x}, y) dy} f(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)} + \int m(\mathbf{x}) v(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)}$$

we get

$$\begin{aligned}
\Gamma_\alpha^M(F + V) &= \Gamma_\alpha^M(F) + \int \frac{\int \Psi\left\{y - m(\mathbf{x})\right\} v(\mathbf{x}, y) dy}{\int \Psi'\left\{y - m(\mathbf{x})\right\} f(\mathbf{x}, y) dy} f(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)} \\
&\quad + \int m(\mathbf{x}) v(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)} + O\left( \|V\|^2_{L(\infty, d+1)} \right).
\end{aligned}$$

## A.7 Proof of Corollary 2

In a similar way as in proof of Corollary 1, one readily obtains

$$\begin{aligned}
\hat{m}_\alpha^M(x_\alpha) &= m_\alpha(x_\alpha) + \frac{1}{n} \sum_{i=1}^n \int \frac{\int \Psi\left\{y - m(\mathbf{x})\right\} \left\{ K_h(\mathbf{x} - \mathbf{X}_i) K_h(y - Y_i) - f(\mathbf{x}, y) \right\} dy}{\int \Psi'\left\{y - m(\mathbf{x})\right\} f(\mathbf{x}, y) dy} \\
&\quad f(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)} + o_p((nh)^{-\frac{1}{2}}).
\end{aligned}$$

## A.8 Proof of Theorem 3

We have

$$\Psi^M(\mathbf{u_X}, u_Y; h) = \int \frac{\int \Psi\{y - m(\mathbf{x})\}\{K_h(\mathbf{x} - \mathbf{u_X})K_h(y - u_Y) - f(\mathbf{x}, y)\} dy}{\int \Psi'\{y - m(\mathbf{x})\} f(\mathbf{x}, y) dy}.$$

From the boundedness of the score function $\Psi$, the boundedness in $(\mathbf{u_X}, u_Y)$ is obvious.

## A.9 Proof of Theorem 4

In expansion (2), let's study the term

$$T_{2n}^M = \frac{1}{n}\sum_{i=1}^n \int \frac{\int \Psi\{y - m(\mathbf{x})\}\{K_h(\mathbf{x} - \mathbf{X}_i)K_h(y - Y_i) - f(\mathbf{x}, y)\} dy}{\int \Psi'\{y - m(\mathbf{x})\} f(\mathbf{x}, y) dy} f(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)}$$

Let's first examine the bias term

$$B_{2n}^M = E\left[\int \frac{\int \Psi\{y - m(\mathbf{x})\} K_h(\mathbf{x} - \mathbf{X}_i)K_h(y - Y_i) dy}{\int \Psi'\{y - m(\mathbf{x})\} f(\mathbf{x}, y) dy} f(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)}\right]$$

i.e.

$$B_{2n}^M = \int\int \left[\int \frac{\int \Psi\{y - m(\mathbf{x})\} K_h(\mathbf{x} - \mathbf{u})K_h(y - v) dy}{\int \Psi'\{y - m(\mathbf{x})\} f(\mathbf{x}, y) dy} f(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)}\right] f(u, v) du dv$$

Using the change of variable $\xi = \dfrac{\mathbf{x} - \mathbf{u}}{h}$ in the integration with respect to $\mathbf{u}$ and $\eta = \dfrac{y - v}{h}$ in the integration with respect to $v$, we obtain

$$B_{2n}^M = \int \frac{\int\int\int \Psi\{y - m(\mathbf{x})\} K(\xi)K(\eta) f(\mathbf{x} - h\xi, y - h\eta) d\xi d\eta dy}{\int \Psi'\{y - m(\mathbf{x})\} f(\mathbf{x}, y) dy} f(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)}.$$

Using a Taylor expansion of $f(\mathbf{x} - h\xi, y - h\eta)$ up to order $r$, under assumptions $(A4)$ we obtain

$$B_{2n}^M = \int \frac{\int \Psi\{y - m(\mathbf{x})\} f(\mathbf{x}, y) dy}{\int \Psi'\{y - m(\mathbf{x})\} f(\mathbf{x}, y) dy} f(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)} + O(h^r)$$

so that, under assumption $(A5)$, we have

$$(nh)^{\frac{1}{2}}\left(B_{2n}^M - \int \frac{\int \Psi\left\{y-m(\mathbf{x})\right\} f(\mathbf{x},y)dy}{\int \Psi'\left\{y-m(\mathbf{x})\right\} f(\mathbf{x},y)dy} f(\mathbf{x}_{(-\alpha)})d\mathbf{x}_{(-\alpha)}\right) = o_p(1)$$

i.e.

$$(nh)^{\frac{1}{2}} E\left(T_{2n}^M\right) = o(1).$$

We now have to study the variance term normalized at rate $(nh)^{\frac{1}{2}}$ :

$$(nh)^{\frac{1}{2}} V_{2n}^M = \sum_{i=1}^n \sqrt{\frac{h}{n}}\left[\int \frac{\int \Psi\left\{y-m(\mathbf{x})\right\} K_h(\mathbf{x}-\mathbf{X}_i)K_h(y-Y_i)dy}{\int \Psi'\left\{y-m(\mathbf{x})\right\} f(\mathbf{x},y)dy} f(\mathbf{x}_{(-\alpha)})d\mathbf{x}_{(-\alpha)} - B_{2n}^M\right]$$

Let's define $\mathcal{F}_{n,i}$ the $\sigma$ field generated by $\left\{\mathbf{X}_j, Y_j\right\}_{j=1,\dots i}$ and let's define the random variable

$$Z_{n,i}^M = \sqrt{\frac{h}{n}}\left[\int \frac{\int \Psi\left\{y-m(\mathbf{x})\right\} K_h(\mathbf{x}-\mathbf{X}_i)K_h(y-Y_i)dy}{\int \Psi'\left\{y-m(\mathbf{x})\right\} f(\mathbf{x},y)dy} f(\mathbf{x}_{(-\alpha)})d\mathbf{x}_{(-\alpha)} - B_{2n}^M\right]$$

so that $(nh)^{\frac{1}{2}} V_{2n}^M = \sum_{j=1}^n Z_{n,j}^M$. The array $\left\{\sum_{j=1}^i Z_{n,j}^M, \mathcal{F}_{n,i}, 1 \le i \le n, n \ge 1\right\}$ is a zero mean, square integrable martingale array. We have

$$\sum_{j=1}^n E(Z_{n,j}^{M2}) = h\int\int\left[\int \frac{\int \Psi\left\{y-m(\mathbf{x})\right\} K_h(\mathbf{x}-\mathbf{u})K_h(y-v)dy}{\int \Psi'\left\{y-m(\mathbf{x})\right\} f(\mathbf{x},y)dy} f(\mathbf{x}_{(-\alpha)})d\mathbf{x}_{(-\alpha)}\right]^2 . \\ f(\mathbf{u},v)d\mathbf{u}dv - h(B_{2n}^M)^2.$$

With the change of variable $\xi_\alpha = \dfrac{x_\alpha - u_\alpha}{h}$ in the integration with respect to $u_\alpha$ and $\xi_{(-\alpha)} = \dfrac{\mathbf{x}_{(-\alpha)} - \mathbf{u}_{(-\alpha)}}{h}$ in the integration with respect to $\mathbf{x}_{(-\alpha)}$, we obtain

$$\sum_{j=1}^n E(Z_{n,j}^{M2}) = h\int\int\left[\int \frac{\int \Psi\left(\left\{y-m(x_\alpha,\mathbf{u}_{(-\alpha)}+h\xi_{(-\alpha)})\right\}\right) K_h(y-v)dy}{\int \Psi'\left\{y-m(x_\alpha,\mathbf{u}_{(-\alpha)}+h\xi_{(-\alpha)})\right\} f(x_\alpha,\mathbf{u}_{(-\alpha)}+h\xi_{(-\alpha)},y)dy} \\ K(\xi_{(-\alpha)})f(\mathbf{u}_{(-\alpha)}+h\xi_{(-\alpha)})d\xi_{(-\alpha)}\right]^2 . \\ K^2(\xi_\alpha)f(x_\alpha - h\xi_\alpha, \mathbf{u}_{(-\alpha)}, v)d\xi_\alpha d\mathbf{u}_{(-\alpha)}dv - h(B_{2n}^M)^2.$$

In the integral $\int \Psi\left\{y - m(x_\alpha, \mathbf{u}_{(-\alpha)} + h\xi_{(-\alpha)})\right\} K_h(y-v)dy$, we make the change of variable $w = \dfrac{y-v}{h}$,so that

$$\sum_{j=1}^{n} E(Z_{n,j}^{M2}) = h \int\int \left[ \int \frac{\int \Psi\left\{v + hw - m(x_\alpha, \mathbf{u}_{(-\alpha)} + h\xi_{(-\alpha)})\right\} K(w)dw}{\int \Psi'\left\{y - m(x_\alpha, \mathbf{u}_{(-\alpha)} + h\xi_{(-\alpha)})\right\} f(x_\alpha, \mathbf{u}_{(-\alpha)} + h\xi_{(-\alpha)}, y)dy} K(\xi_{(-\alpha)}) f(\mathbf{u}_{(-\alpha)} + h\xi_{(-\alpha)}) d\xi_{(-\alpha)} \right]^2 .$$
$$K^2(\xi_\alpha) f(x_\alpha - h\xi_\alpha, \mathbf{u}_{(-\alpha)}, v) d\xi_\alpha d\mathbf{u}_{(-\alpha)} dv - h(B_{2n}^M)^2$$

taking the limit for $n \to \infty$, we obtain

$$\lim_{n\to\infty} E\left(\sum_{j=1}^{n} Z_{n,j}^{M2}\right) = \int\int \left[ \frac{\Psi\left\{y - m(x_\alpha, \mathbf{u}_{(-\alpha)})\right\}}{\int \Psi'\left\{y - m(x_\alpha, \mathbf{u}_{(-\alpha)})\right\} f(x_\alpha, \mathbf{u}_{(-\alpha)}, y)dy} f(\mathbf{x}_{(-\alpha)}) \right]^2 f(x_\alpha, \mathbf{u}_{(-\alpha)}, v) d\mathbf{u}_{(-\alpha)} dv \cdot \int K^2(\xi_\alpha) d\xi_\alpha$$

and

$$\sum_{j=1}^{n} E\left(Z_{n,j}^{M2} / \mathcal{F}_{n,j-1}\right) \to^{p} \int\int \left[ \frac{\Psi\left\{y - m(x_\alpha, \mathbf{u}_{(-\alpha)})\right\}}{\int \Psi'\left\{y - m(x_\alpha, \mathbf{u}_{(-\alpha)})\right\} f(x_\alpha, \mathbf{u}_{(-\alpha)}, y)dy} f(\mathbf{x}_{(-\alpha)}) \right]^2 f(x_\alpha, \mathbf{u}_{(-\alpha)}, v) d\mathbf{u}_{(-\alpha)} dv \cdot \int K^2(\xi_\alpha) d\xi_\alpha .$$

Under assumptions $(A1)$ to $(A5)$, conditions 3.19 and 3.20 of Corollary 3.1 of Hall and Heyde (1981, pp 58) are satisfied, so that we finally get

$$(nh)^{\frac{1}{2}} V_{2n}^M \to \mathcal{N}(0, V^M)$$

where

$$V^M = \int\int \left[ \frac{\Psi\left\{y - m(x_\alpha, \mathbf{u}_{(-\alpha)})\right\}}{\int \Psi'\left\{y - m(x_\alpha, \mathbf{u}_{(-\alpha)})\right\} f(x_\alpha, \mathbf{u}_{(-\alpha)}, y)dy} f(\mathbf{x}_{(-\alpha)}) \right]^2 f(x_\alpha, \mathbf{u}_{(-\alpha)}, v) d\mathbf{u}_{(-\alpha)} dv \times \int K^2(\xi_\alpha) d\xi_\alpha .$$

# References


Aït-Sahalia, Y. (1995). The delta method for nonlinear kernel functionals. Ph. D. dissertation, University of Chicago.

Bianco, A., and Boente, G. (1998). Robust kernel estimators for additive models with dependent observations. *The Canadian Journal of Statistics* **26**, 239-255.

Chèze, N., Poggi, J. M., and Portier, B. (2000). Partial and weighted integration estimators for nonlinear additive models. Prépublication Mathématique, Université de Paris sud, No 2000-53.

Doukhan, P. (1990). *Mixing, Properties and Examples.* New York: Springer-Verlag.

Hall, P., and Heyde, C. C. (1981). *Martingale Limit Theory and its Applications.* New York: Academic Press.

Hampel, F. R.. (1994). The influence curve and its role in robust estimation. *Journal of the American Statistical Association* **62**, 1179-1186.

Härdle, W. (1990). *Applied Nonparametric Regression.* Boston: Cambridge University Press.

Hastie, T. J., and Tibshirani, R. J. (1990). *Generalized Additive Models.* Chapman & Hall.

Huber, P. J. (1981). *Robust Statistics.* New-York: Wiley.

Linton, O. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82**, 93-100.

Lucas, A. (1996). Outlier robust unit root analysis. Ph. D. dissertation, Rotterdam, Tinbergen Institute.

Mammen, E. Linton, O. and Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. Annals of Statistics **5**, 1443-1490.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Applications* 10, 186-190.

Opsomer, J., and Ruppert, D. (1997). Fitting bivariate additive model by local polynomial regression. *The Annals of Statistics* **25**, 186-212.

Parzen, E. (1962). On the estimation of a probability density and mode. J*ournal of Mathematical Statistics* **33**, 1065-1076.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics* **27**, 832-837.

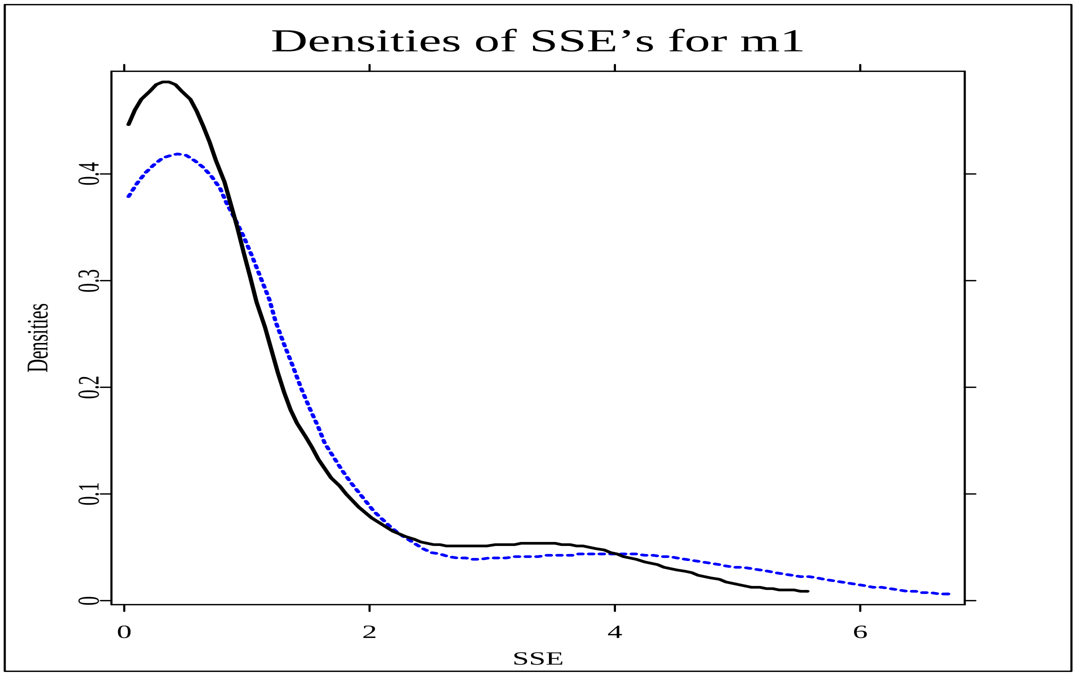
Stone, C. (1985). Additive regression and other nonparametric models. Annals of Statistics **30**, 689-705.

Tjøstheim, D., and Auestad, B. H.. (1994). Nonparametric identification of nonlinear time series: projections. *Journal of the American Statistical Association* **89**, 1398-1409.

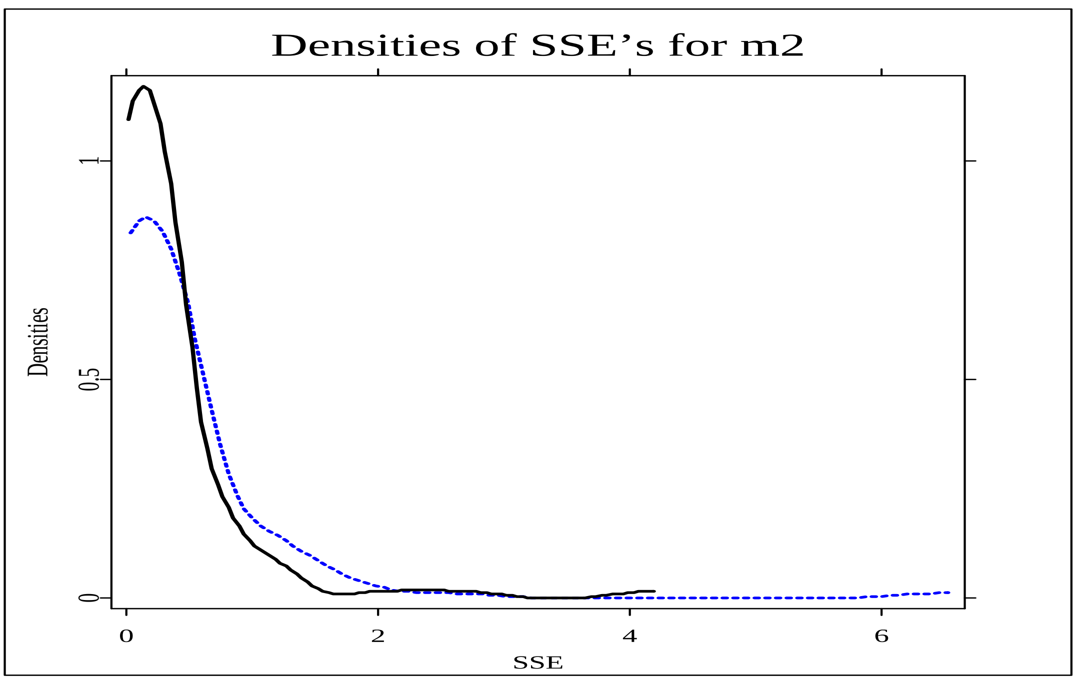
Tamine, J. (2002). Smoothed influence function : another view at robust nonparametric regression. Discussion paper **62**, Sonderforschungsbereich 373, Humboldt-Universität zu Berlin.

Watson, G. S. (1964). Smooth regression analysis. *Sankhya* Ser A., **26**, 359-372.

Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, U.S.A.
E-mail: yang@stt.msu.edu
Fax: 517-432-1405 Phone: 517-353-6369




(a)




(b)

Figure 2: Kernel estimates of density function of the SSE's of estimators of $m_\alpha$, with $\alpha = 1, 2$, based on 100 simulated samples of size 150: (a) SSE's of $m_1$, robustified estimator— solid, ordinary estimator — dash. (b) is the counterpart of (a) for $m_2$