

R robustified additive nonparametric regression

BY JULIEN TAMINE

*Université de la Méditerranée, GREQAM Vieille Charité, 2 rue de la Charité,
13002 Marseille, France*

WOLFGANG HÄRDLE

*Center for Applied Statistics and Economics, Institut für Statistik und
Ökonometrie, Humboldt Universität zu Berlin, Spandauer Str.1, D-10178 Berlin,
Germany*

LIJIAN YANG

*Department of Statistics and Probability, Michigan State University, East Lansing,
MI 48824, U. S. A.*

SUMMARY

Additive modelling is known to be useful for multivariate nonparametric regression as it reduces the complexity of problem to the level of univariate regression. This usefulness could be compromised if the data set was contaminated by outliers whose detection and removal are particularly difficult to achieve in high dimension. We propose an estimation procedure for the additive component of the regression function, less sensitive to possible outliers in the sample. Our procedure is based on marginal integration of conditional R-estimators. In addition to univariate rate of convergence and asymptotic distribution, we also obtain robustness results for our estimator. All of our results are valid for a broad class of β mixing processes. Monte Carlo findings confirm the theoretical results in finite sample.

Some key words: R-estimator; Additive model; Kernel estimator; Marginal integration; Robustness.

1 INTRODUCTION

Regression analysis is a powerful tool for investigating the links between a set of regressors and a response variable for which prediction is needed. Classical linear and nonlinear parametric regression models have proved to be inadequate for many data sets one encounters in practical problems (see for instance Härdle 1990). As a remedy, non and semi-parametric models have received greater attention in recent years from both practitioners and theoreticians. These estimators are free from the constraint of any parametric specification, allowing a much more flexible adaption to the structure of the data.

Due to the lack of any parametric formula, the object of nonparametric modelling is a function unconstrained to lie in a finite dimensional space, and as a consequence,

much more data are needed for an accurate estimation. This limitation of nonparametric estimators becomes acute as the dimension of the data becomes higher. This phenomenon is commonly referred to as the “curse of dimensionality”. There exists various tools under the general term dimension reduction techniques that seek a compromise between flexibility and accuracy. Among these, additive modelling is one of the most natural.

Consider the problem of predicting the value of the random variable Y using a d dimensional set of predictor variables \mathbf{X} . The best predictor one can use in the sense of the mean square error is the conditional expectation $m(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$. Due to the “curse of dimensionality”, the reconstruction of the function m based on a set of realizations $\{(\mathbf{X}_i, Y_i)\}_{i=1\dots n}$ of the random vector (\mathbf{X}, Y) is unsatisfactory for high dimension d unless some simplifying can be imposed on the form of m . Additive modelling consists of assuming that the function $m(\mathbf{x})$ is of the form

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

where x_{α} is the α -th component of \mathbf{x} . Under suitable assumptions many techniques are available for estimating the additive components $m_{\alpha}(x_{\alpha})$ with the univariate rate of convergence. Stone (1985) proposed to estimate $m_{\alpha}(x_{\alpha})$ using polynomial spline and gave bounds on the optimal rate of convergence that can be obtained. Hastie and Tibshirani (1990) developed the backfitting algorithm on heuristic basis. Linton (1995) and Tjøstheim and Auestad (1994) independently used the marginal integration procedure and gave asymptotic distributions. More recently, Mammen, Linton and Nielsen (1999) proposed and established asymptotic distribution theory for a variation of the backfitting algorithm.

Since all the procedures mentioned above are based on least squares criterion, they are vulnerable to the influence from a few extreme observations. For example, if Y_n is extremely large, the estimators using the sample $\{(\mathbf{X}_i, Y_i)\}_{i=1\dots n}$ and $\{(\mathbf{X}_i, Y_i)\}_{i=1\dots n-1}$ may be quite different. In univariate regression estimation such extreme observations can often be removed by a careful visual inspection. In high dimension, their identification and removing become much more problematic since the data set no longer has simple graphical representation. As a consequence, it is necessary to develop estimation procedure of additive model more resistant to outliers.

Figure 1 illustrates the non-robustness of ordinary marginal integration estimator based on least squares. 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.

In this work, we propose an R-type marginal integration estimator. The estimator is shown to be robust in the sense that the contribution of any observation to the error of estimation is bounded. As a consequence, a large observation can not result in an extreme change in the value of the estimator. Hence, there is no need to look for and remove any outliers when using this procedure. We will show that the proposed estimator enjoys the univariate rate of convergence and we derive its asymptotic distribution. Our results are valid under mild β -mixing conditions. Our Monte-Carlo study corroborates the theoretical results. In other words, our R robustified procedure outperforms the ordinary marginal integration estimator when the sample is contaminated with outliers.

Our work is organized as follows. In Section 2, we describe the additive model setup and define an R estimator of $m_\alpha(x_\alpha)$ based on the idea of marginal integration. Its asymptotic and robustness properties are studied in details. In Section 3, a Monte Carlo example illustrates the improvements of the R-robustified estimator when the data are contaminated with outliers. Proofs and important assumptions are given in Appendix.

2 R-ESTIMATOR OF ADDITIVE REGRESSION

We consider an additive form for the regression function $m(\mathbf{x})$ as given in (1). In order to identify the additive components $m_\alpha(x_\alpha)$, we will assume that

$$E \{m_\alpha(X_\alpha)\} = 0, 1 \leq \alpha \leq d.$$

Under these assumptions we have $\mu = E(Y)$. In order to simplify the notations, we will assume in the following that $\mu = 0$. The function $m_\alpha(x_\alpha)$ then satisfies the equality

$$m_\alpha(x_\alpha) = E \left\{ m \left(x_\alpha, \mathbf{X}_{(-\alpha)} \right) \right\}, 1 \leq \alpha \leq d, \quad (1)$$

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

Relation (1) gives a way of estimating $m_\alpha(x_\alpha)$ based on plug-ins

$$\hat{m}_\alpha(x_\alpha) = \int \hat{m}(\mathbf{x}) d\hat{F}(\mathbf{x}_{(-\alpha)}),$$

where $\hat{F}(\mathbf{x}_{(-\alpha)})$ is an estimator of the marginal distribution of $\mathbf{X}_{(-\alpha)}$ and $\hat{m}(\mathbf{x})$ is an estimator of the function $m(\mathbf{x})$. The marginal integration estimator of Linton (1995) consists of using the Nadaraya-Watson estimator for $\hat{m}(\mathbf{x})$ and the empirical cumulative distribution function for $\hat{F}(\mathbf{x}_{(-\alpha)})$.

In the following work, we propose an alternative plug-in version of $\hat{m}_\alpha(x_\alpha)$. In our scheme, $\hat{F}(\mathbf{x}_{(-\alpha)})$ is the integration of the Parzen Rosenblatt estimator of density

$$\hat{F}(\mathbf{x}_{(-\alpha)}) = \int_{-\infty}^{\mathbf{x}_{(-\alpha)}} \hat{f}(\mathbf{t}_{(-\alpha)}) d\mathbf{t}_{(-\alpha)},$$

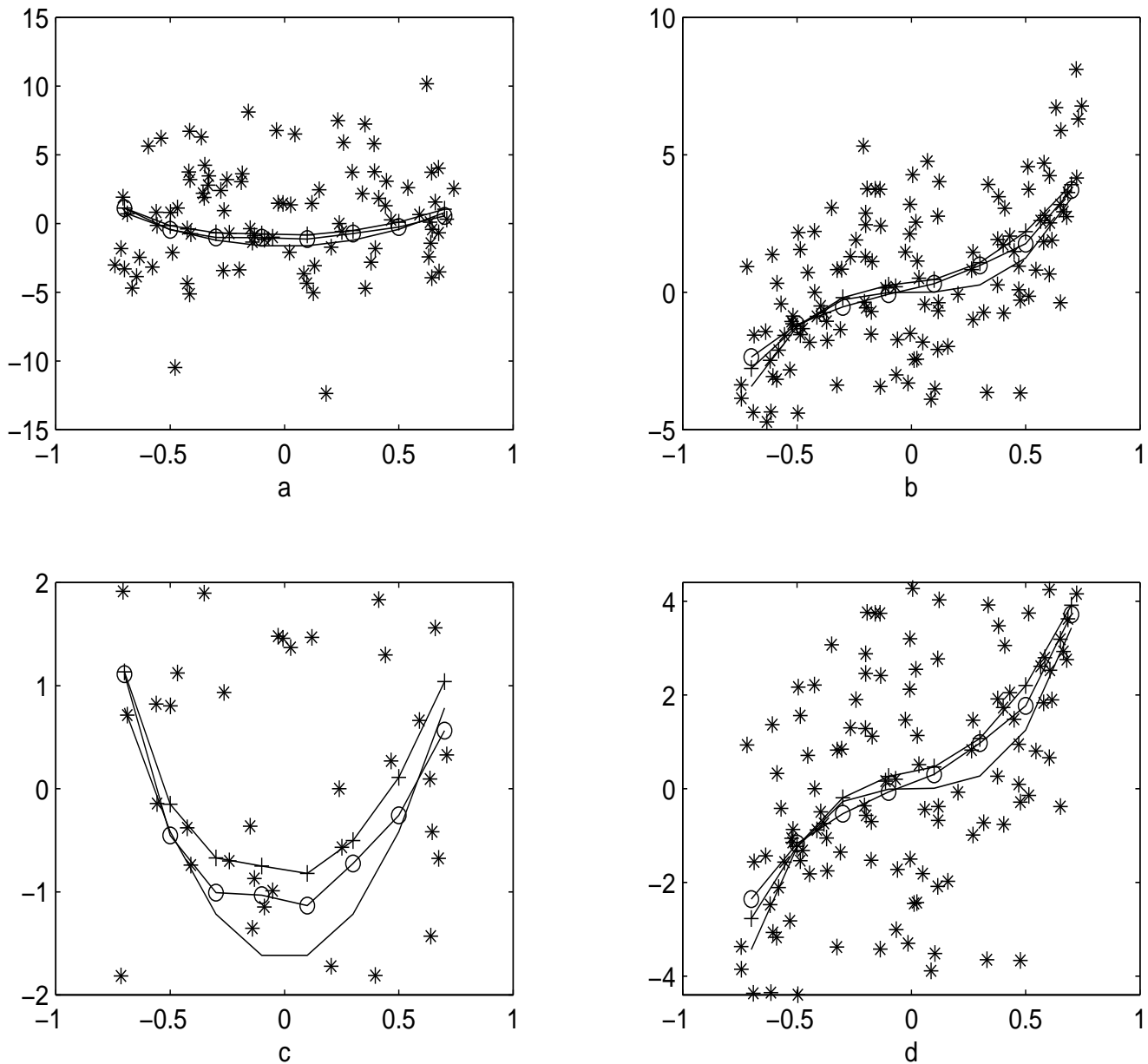


Figure 1: Scatterplots and function estimates for a simulated sample of size 150: $Y = m_1(X_1) + m_2(X_2) + \varepsilon$ where $m_1(t) = 5t^2 - 5/3$, $m_2(t) = 10t^3$ and ε has a normal mixture distribution (see Section 3). (a) plot of (X_1, Y) , and $m_1(t)$ — solid, robust estimator — circle, ordinary 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).

$$\hat{f}(\mathbf{t}_{(-\alpha)}) = \frac{1}{nh^{d-1}} \sum_{i=1}^n K\left(\frac{\mathbf{t}_{(-\alpha)} - \mathbf{X}_{i(-\alpha)}}{h}\right),$$

with K being the product of a univariate kernel describe in assumption (A4). Our robustification process consists of using an estimator for $\hat{m}(\mathbf{x})$ more robust than the Nadaraya-Watson. To be more precise, we will 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 round $m(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$. Under this assumption, adapting rank test theory to conditional setting as in Cheng and Cheng (1990), the regression function $m(\mathbf{x})$ is the unique solution to the equation

$$\int_{-\infty}^{+\infty} J\left[2^{-1}\{F_{\mathbf{x}}(y) + 1 - F_{\mathbf{x}}(2m(\mathbf{x}) - y)\}\right] dF_{\mathbf{x}}(y) = 0, \quad (2)$$

provided that the following assumptions (R1) and (R2) hold.

(R1) The function J is twice continuously differentiable on $[0, 1]$ and satisfies

$$\int_{-\infty}^{+\infty} J'(F_{\mathbf{x}}(y)) (F'_{\mathbf{x}}(y))^2 dy > 0.$$

(R2) The function J is increasing on $[0, 1]$ with $J(t) = -J(1 - t)$.

These assumptions are common in the literature on R-estimation, see, for instance, Huber (1981). A commonly used choice for the score function J is the Hodges-Lehmann score $J(t) = t - 1/2$ which corresponds to Wilcoxon rank test (see section 9.1.3 in Serfling 1980).

The function $m(\mathbf{x})$ can then be estimated by solving the equation (2) where $F_{\mathbf{x}}(y)$ is replaced by

$$\hat{F}_{\mathbf{x}}(y) = \int_{-\infty}^y \frac{\hat{f}(\mathbf{x}, v)}{\hat{f}(\mathbf{x})} dv,$$

with $\hat{f}(\mathbf{x}, v)$ and $\hat{f}(\mathbf{x})$ being kernel Parzen Rosenblatt density estimators. The existence and the convergence properties of the solution to this plug in equation have been studied in details in Tamine (2002). The asymptotic robustness properties of $\hat{m}_{\alpha}(x_{\alpha})$ are summarized in the following theorem:

Theorem 1 *Under assumptions (A1) to (A5), we have*

(i)

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

where

$$\Delta(\mathbf{u}_X, u_Y, h) = \int \frac{J\{F_{\mathbf{x}}(y)\} \frac{K_h(\mathbf{x} - \mathbf{u}_X)K_h(y - u_Y)}{f(\mathbf{x})} dy}{\int_{-\infty}^{+\infty} J'\{F_{\mathbf{x}}(y)\} \{F'_{\mathbf{x}}(y)\}^2 dy} f(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)}.$$

(ii) $\Delta(\mathbf{u}_X, u_Y, h)$ is bounded in (\mathbf{u}_X, u_Y) .

The first part of this theorem allows us to distinguish the leading order error terms in estimating $m_\alpha(x_\alpha)$ by $\hat{m}_\alpha(x_\alpha)$. This is useful not only for studying the asymptotic distribution of the estimator as is done in corollary 1, but also for quantifying its robustness properties. Indeed, as shown by point (ii) of Theorem 1, the contribution of any observation toward the estimation error can't become arbitrarily large. As pointed by Hampel (1994) through the concept of influence function of which $\Delta(\mathbf{u}_X, u_Y, h)$ is a smoothed analog, this robustness property is particularly useful in case the sample would contain outliers.

Corollary 1 *Under assumptions (A1) to (A5), we have*

$$\sqrt{nh} \{\hat{m}_\alpha(x_\alpha) - m_\alpha(x_\alpha)\} \rightarrow N(0, V),$$

where

$$V = \int K^2(t) dt \int_0^1 J^2(t) dt \int \frac{f^2(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x}) \left[\int_{-\infty}^{+\infty} J'\{F_{\mathbf{x}}(y)\} \{F'_{\mathbf{x}}(y)\}^2 dy \right]^2} d\mathbf{x}_{(-\alpha)}.$$

3 MONTE-CARLO SIMULATION

For our simulation study, we use data generated from the following equation

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

where m is the following bivariate additive function, whose components have been plotted in Figure 1

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

The predictor variable \mathbf{X} is uniformly distributed on $[-1, 1]^2$ and ε has a density function which is a mixture of normal densities

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

where ν is the degree that the standard normal law has been contaminated, and k is the variance of the contaminating law. In our study, we set $\nu = 10$ and $k = 9$.

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 Ψ 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 $\{t_1, \dots, t_8\}$ 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 $SSE_\alpha = \frac{1}{8} \sum_{i=1}^8 \{m_\alpha(t_i) - \hat{m}_\alpha(t_i)\}^2$ where \hat{m}_α will be either the ordinary marginal integration estimator of the regression component function m_α or the R robustified estimator. 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 is consistent with Theorem 1.

ACKNOWLEDGEMENT

All three authors have been financially supported by the Deutsche Forschungsgemeinschaft via Sonderforschungsbereich 373 “Quantifikation und Simulation Ökonomischer Prozesse” at Humboldt-Universität zu Berlin. The first author has also been supported by Université de la Méditerranée.

APPENDIX

We need the following assumptions for the proofs

- (A1) The sequence $\{(\mathbf{X}_i, Y_i)\}_{i=1..n}$ is a sequence of strictly stationary and β mixing realizations of the vector (\mathbf{X}, Y) satisfying $k^\delta \beta_k \rightarrow 0$ for some fixed $\delta > 1$. Here $\beta_k = E \sup \left\{ \left| P(A | \mathcal{F}_m^l) - P(A) \right| : A \in \mathcal{F}_{k+l}^\infty \right\}$ where \mathcal{F}_t^l is the σ algebra generated by $(\mathbf{X}_t, Y_t), \dots, (\mathbf{X}_{t+l}, Y_{t+l})$
- (A2) The density $f(\mathbf{x}, y)$ is compactly supported and has continuous derivatives up to order r .

(A3) The density $f(\mathbf{x})$ admits a strictly positive lower bound, c .

(A4) The univariate function K is a symmetric, compactly supported kernel of order r .

(A5) The bandwidth satisfies $\lim_{n \rightarrow \infty} h = 0$ (the dependance of h on n is left implicit for the simplicity of notations) in such a way that $\lim_{n \rightarrow \infty} n^{\frac{1}{2}} h^{2d + \frac{3}{2}} \rightarrow \infty$ and $\lim_{n \rightarrow \infty} n^{\frac{1}{2}} h^{r + \frac{1}{2}} \rightarrow 0$.

A.1 Proof of Theorem 1

Rewriting equation (2) with

$$F_{\mathbf{x}}(y) = \int_{-\infty}^y \frac{f(\mathbf{x}, v)}{f(\mathbf{x})} dv, \quad (\text{A.1})$$

$m(\mathbf{x})$ is implicitly defined by the equation

$$\int_{-\infty}^{+\infty} J \left[2^{-1} \{ F_{\mathbf{x}}(y) + 1 - F_{\mathbf{x}}(2m(\mathbf{x}) - y) \} \right] dF_{\mathbf{x}}(y) = 0. \quad (\text{A.2})$$

As is shown by equation (A.1), $F_{\mathbf{x}}(y)$ is completely determined by $f(\mathbf{x}, y)$. Therefore, using (A.2), $m(\mathbf{x})$ is a functional $f(\mathbf{x}, y)$ that we will denote as $R(f)$. In the following, the norm $\|g(\mathbf{x}, y)\|_{\infty}$ is used for the topology on the function space. Using Tamine (2002), under assumptions (A2), (A3), (R1) and (R2), for g being compactly supported and lying in a neighborhood of zero that satisfies $|g(\mathbf{x})| \leq c/2$, the following expansion holds

$$R(f + g) = R(f) + \frac{\int_{-\infty}^{+\infty} J \{ F_{\mathbf{x}}(y) \} \frac{g(\mathbf{x}, y)}{f(\mathbf{x})} dy}{\int_{-\infty}^{+\infty} J' \{ F_{\mathbf{x}}(y) \} \{ F'_{\mathbf{x}}(y) \}^2 dy} + O \left(\|g(\mathbf{x}, y)\|_{\infty}^2 \right). \quad (\text{A.3})$$

Using the identification equation

$$m_{\alpha}(x_{\alpha}) = \int m(\mathbf{x}) f(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)}, \quad (\text{A.4})$$

$m_{\alpha}(x_{\alpha})$ is also a functional of $f(\mathbf{x}, y)$ which we denote as $R_{\alpha}(f)$. Equation (A.4) can be written using functional forms as

$$R_{\alpha}(f) = \int R(f) f(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)},$$

so that

$$R_\alpha(f+g) = \int R(f+g) \{f(\mathbf{x}_{(-\alpha)}) + g(\mathbf{x}_{(-\alpha)})\} d\mathbf{x}_{(-\alpha)}.$$

Using expansion (A.3), we obtain

$$\begin{aligned} R_\alpha(f+g) &= \int R(f) f(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)} + \int \frac{\int_{-\infty}^{+\infty} J\{F_{\mathbf{x}}(y)\} g(\mathbf{x}, y) dy}{\int_{-\infty}^{+\infty} J'\{F_{\mathbf{x}}(y)\} \{F'_{\mathbf{x}}(y)\}^2 dy} \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} d\mathbf{x}_{(-\alpha)} \\ &\quad + \int R(f) g(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)} + \int \frac{\int_{-\infty}^{+\infty} J\{F_{\mathbf{x}}(y)\} g(\mathbf{x}, y) dy}{\int_{-\infty}^{+\infty} J'\{F_{\mathbf{x}}(y)\} \{F'_{\mathbf{x}}(y)\}^2 dy} \frac{g(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} d\mathbf{x}_{(-\alpha)} \\ &\quad + O(\|g(\mathbf{x}, y)\|_\infty^2) \int \{f(\mathbf{x}_{(-\alpha)}) + g(\mathbf{x}_{(-\alpha)})\} d\mathbf{x}_{(-\alpha)}. \end{aligned}$$

The term

$$\int \frac{\int_{-\infty}^{+\infty} J\{F_{\mathbf{x}}(y)\} g(\mathbf{x}, y) dy}{\int_{-\infty}^{+\infty} J'\{F_{\mathbf{x}}(y)\} \{F'_{\mathbf{x}}(y)\}^2 dy} \frac{g(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} d\mathbf{x}_{(-\alpha)}$$

is easily shown to be $O(\|g(\mathbf{x}, y)\|_\infty^2)$ and the term

$$\int \{f(\mathbf{x}_{(-\alpha)}) + g(\mathbf{x}_{(-\alpha)})\} d\mathbf{x}_{(-\alpha)}$$

is bounded, so we finally obtain

$$\begin{aligned} R_\alpha(f+g) &= \int R(f) f(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)} + \int \frac{\int_{-\infty}^{+\infty} J\{F_{\mathbf{x}}(y)\} g(\mathbf{x}, y) dy}{\int_{-\infty}^{+\infty} J'\{F_{\mathbf{x}}(y)\} \{F'_{\mathbf{x}}(y)\}^2 dy} \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} d\mathbf{x}_{(-\alpha)} \\ &\quad + \int R(f) g(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)} + O(\|g(\mathbf{x}, y)\|_\infty^2). \end{aligned} \tag{A.5}$$

Under assumptions (A1) to (A5), for $g(\mathbf{x}, y) = \hat{f}(\mathbf{x}, y) - f(\mathbf{x}, y)$, Aït-Sahalia (1995) has established that $\|g(\mathbf{x}, y)\|_\infty \rightarrow^{as} 0$, so that we can use expression (A.5) for this

particular g function. We obtain

$$\begin{aligned}
R_\alpha(\hat{f}) &= \int R(f) f(\mathbf{x}_{(-\alpha)}) d\mathbf{x}_{(-\alpha)} + \int \frac{\int_{-\infty}^{+\infty} J\{F_{\mathbf{x}}(y)\} \{\hat{f}(\mathbf{x}, y) - f(\mathbf{x}, y)\} dy}{\int_{-\infty}^{+\infty} J'\{F_{\mathbf{x}}(y)\} \{F'_{\mathbf{x}}(y)\}^2 dy} \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} d\mathbf{x}_{(-\alpha)} \\
&\quad + \int R(f) \{\hat{f}(\mathbf{x}_{(-\alpha)}) - f(\mathbf{x}_{(-\alpha)})\} d\mathbf{x}_{(-\alpha)} + O\left(\|\hat{f}(\mathbf{x}, y) - f(\mathbf{x}, y)\|_\infty^2\right),
\end{aligned}$$

which can also be written as

$$\begin{aligned}
\hat{m}_\alpha(x_\alpha) &= m_\alpha(x_\alpha) + \int \frac{\int_{-\infty}^{+\infty} J\{F_{\mathbf{x}}(y)\} \hat{f}(\mathbf{x}, y) dy}{\int_{-\infty}^{+\infty} J'\{F_{\mathbf{x}}(y)\} \{F'_{\mathbf{x}}(y)\}^2 dy} \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} d\mathbf{x}_{(-\alpha)} \quad (\text{A.6}) \\
&\quad + \int m(\mathbf{x}) \{\hat{f}(\mathbf{x}_{(-\alpha)}) - f(\mathbf{x}_{(-\alpha)})\} d\mathbf{x}_{(-\alpha)} + O\left(\|\hat{f}(\mathbf{x}, y) - f(\mathbf{x}, y)\|_\infty^2\right).
\end{aligned}$$

Here, we have taken advantage of the fact that $\int J[F_{\mathbf{x}}(y)] \frac{f(\mathbf{x}, y)}{f(\mathbf{x})} dy = 0$, which follows directly from the assumption (R2) that $J(t) = -J(1-t)$.

We now show the term

$$T_{1n} = \int m(\mathbf{x}) \{\hat{f}(\mathbf{x}_{(-\alpha)}) - f(\mathbf{x}_{(-\alpha)})\} d\mathbf{x}_{(-\alpha)} \quad (\text{A.7})$$

to be $O_p(n^{-\frac{1}{2}} + h^r)$. Using an integration by parts in expression (A.7) we obtain

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

where $\hat{F}(\mathbf{x}_{(-\alpha)}) = \int_{-\infty}^{\mathbf{x}_{(-\alpha)}} \hat{f}(\mathbf{u}_{(-\alpha)}) d\mathbf{u}_{(-\alpha)}$ and $F(\mathbf{x}_{(-\alpha)}) = \int_{-\infty}^{\mathbf{x}_{(-\alpha)}} f(\mathbf{u}_{(-\alpha)}) d\mathbf{u}_{(-\alpha)}$.

According to Lemma 1 of Aït-Sahalia (1995), under assumptions (A1) to (A5), we have

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

so that

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

Using the same lemma from Aït-Sahalia, we also have

$$\|\hat{f}(\mathbf{x}, y) - f(\mathbf{x}, y)\|_\infty^2 = O_p(n^{-1}h^{-2(d+1)} + h^{2r}),$$

so that (A.6) becomes

$$\begin{aligned} \hat{m}_\alpha(x_\alpha) &= m_\alpha(x_\alpha) + \int \frac{\int_{-\infty}^{+\infty} J\{F_{\mathbf{x}}(y)\} \hat{f}(\mathbf{x}, y) dy}{\int_{-\infty}^{+\infty} J'\{F_{\mathbf{x}}(y)\} \{F'_{\mathbf{x}}(y)\}^2 dy} \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} d\mathbf{x}_{(-\alpha)} \\ &\quad + O_p(n^{-\frac{1}{2}} + h^r) + O_p(n^{-1}h^{-2(d+1)} + h^{2r}). \end{aligned}$$

Under assumption (A5),

$$O_p(n^{-\frac{1}{2}} + h^r) + O_p(n^{-1}h^{-2(d+1)} + h^{2r}) = o_p(n^{-1/2}h^{-1/2}),$$

so that using the expression

$$\hat{f}(\mathbf{x}, y) = \frac{1}{n} \sum_{i=1}^n K_h(\mathbf{x} - \mathbf{X}_i) K_h(y - Y_i)$$

we obtain

$$\begin{aligned} \hat{m}_\alpha(x_\alpha) &= m_\alpha(x_\alpha) + \frac{1}{n} \sum_{i=1}^n \int \frac{\int_{-\infty}^{+\infty} J\{F_{\mathbf{x}}(y)\} K_h(\mathbf{x} - \mathbf{X}_i) K_h(y - Y_i) dy}{\int_{-\infty}^{+\infty} J'\{F_{\mathbf{x}}(y)\} \{F'_{\mathbf{x}}(y)\}^2 dy} \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} d\mathbf{x}_{(-\alpha)} \\ &\quad + o_p(n^{-1/2}h^{-1/2}). \end{aligned}$$

Point (ii) follows immediately from assumptions (A4), (R1) and (R2).

A.2 Proof of Corollary 1

Let's now study the leading term

$$T_{2n} = \frac{1}{n} \sum_{i=1}^n \int \frac{\int_{-\infty}^{+\infty} J\{F_{\mathbf{x}}(y)\} K_h(\mathbf{x} - \mathbf{X}_i) K_h(y - Y_i) dy}{\int_{-\infty}^{+\infty} J'\{F_{\mathbf{x}}(y)\} \{F'_{\mathbf{x}}(y)\}^2 dy} \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} d\mathbf{x}_{(-\alpha)}.$$

We first examine the bias term

$$B_{2n} = E \left(\int \frac{\int_{-\infty}^{+\infty} J\{F_{\mathbf{x}}(y)\} K_h(\mathbf{x} - \mathbf{X}_i) K_h(y - Y_i) dy}{\int_{-\infty}^{+\infty} J'\{F_{\mathbf{x}}(y)\} \{F'_{\mathbf{x}}(y)\}^2 dy} \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} d\mathbf{x}_{(-\alpha)} \right),$$

or

$$B_{2n} = \int \int \left\{ \frac{\int_{-\infty}^{+\infty} J \{F_{\mathbf{x}}(y)\} K_h(\mathbf{x} - \mathbf{u}) K_h(y - v) dy}{\int_{-\infty}^{+\infty} J' \{F_{\mathbf{x}}(y)\} \{F'_{\mathbf{x}}(y)\}^2 dy} \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} d\mathbf{x}_{(-\alpha)} \right\} f(\mathbf{u}, v) d\mathbf{u} dv.$$

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

$$B_{2n} = \int \frac{\int \int \int J \{F_{\mathbf{x}}(y)\} K(\xi) K(\eta) f(\mathbf{x} - h\xi, y - h\eta) d\xi d\eta dy}{\int_{-\infty}^{+\infty} J' \{F_{\mathbf{x}}(y)\} \{F'_{\mathbf{x}}(y)\}^2 dy} \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} d\mathbf{x}_{(-\alpha)}.$$

Using a Taylor expansion of $f(\mathbf{x} - h\xi, y - h\eta)$ up to order r , under assumptions (A2) and (A4), we obtain

$$B_{2n} = \int \frac{\int_{+\infty} J \{F_{\mathbf{x}}(y)\} f(\mathbf{x}, y) dy}{\int_{-\infty} J' \{F_{\mathbf{x}}(y)\} \{F'_{\mathbf{x}}(y)\}^2 dy} \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} d\mathbf{x}_{(-\alpha)} + O(h^r).$$

Nest, use again the fact that $\int J [F_{\mathbf{x}}(y)] \frac{f(\mathbf{x}, y)}{f(\mathbf{x})} dy = 0$, implied by assumption (R2), we obtain

$$B_{2n} = O(h^r),$$

so that, under assumption (A5), we have

$$n^{1/2} h^{1/2} B_{2n} = o_p(1).$$

We now have to study the variance term normalized at rate $n^{1/2} h^{1/2}$

$$n^{1/2} h^{1/2} V_{2n} = \sum_{i=1}^n \sqrt{\frac{h}{n}} \left[\int \frac{\int_{-\infty}^{+\infty} J \{F_{\mathbf{x}}(y)\} K_h(\mathbf{x} - \mathbf{X}_i) K_h(y - Y_i) dy}{\int_{-\infty}^{+\infty} J' \{F_{\mathbf{x}}(y)\} \{F'_{\mathbf{x}}(y)\}^2 dy} \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} d\mathbf{x}_{(-\alpha)} - B_{2n} \right].$$

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

$$Z_{n,i} = \sqrt{\frac{h}{n}} \left[\frac{\int_{-\infty}^{+\infty} J \{F_{\mathbf{x}}(y)\} K_h(\mathbf{x} - \mathbf{X}_i) K_h(y - Y_i) dy}{\int_{-\infty}^{+\infty} J' \{F_{\mathbf{x}}(y)\} \{F'_{\mathbf{x}}(y)\}^2 dy} \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} d\mathbf{x}_{(-\alpha)} - B_{2n} \right],$$

so that $n^{1/2}h^{1/2}V_{2n} = \sum_{j=1}^n Z_{n,j}$. The array $\left\{ \sum_{j=1}^i Z_{n,j}, \mathcal{F}_{n,i}, 1 \leq i \leq n, n \geq 1 \right\}$ is a zero mean, square integrable martingale array. We have

$$\sum_{j=1}^n E(Z_{n,j}^2) = h \int \int \left[\frac{\int_{-\infty}^{+\infty} J \{F_{\mathbf{x}}(y)\} K_h(\mathbf{x} - \mathbf{u}) K_h(y - v) dy}{\int_{-\infty}^{+\infty} J' \{F_{\mathbf{x}}(y)\} \{F'_{\mathbf{x}}(y)\}^2 dy} \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} d\mathbf{x}_{(-\alpha)} \right]^2 f(\mathbf{u}, v) d\mathbf{u} dv - h(B_{2n})^2.$$

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

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

Now making the change of variable $w = \frac{y - v}{h}$ in the expression $K_h(y - v)$ and letting $n \rightarrow \infty$ we have

$$\lim_{n \rightarrow \infty} \int J \left\{ F_{(x_{\alpha}, \mathbf{u}_{(-\alpha)} + h\xi_{(-\alpha)})}(y) \right\} K_h(y - v) dy = J \left\{ F_{(x_{\alpha}, \mathbf{u}_{(-\alpha)})}(v) \right\}$$

so that

$$\lim_{n \rightarrow \infty} E \left(\sum_{j=1}^n Z_{n,j}^2 \right) = \int \int \left[\frac{J \left\{ F_{(x_\alpha, \mathbf{u}_{(-\alpha)})}(v) \right\}}{\int_{-\infty}^{+\infty} J' \left\{ F_{(x_\alpha, \mathbf{u}_{(-\alpha)})}(y) \right\} \left\{ F'_{(x_\alpha, \mathbf{u}_{(-\alpha)})}(y) \right\}^2 dy} \frac{f(\mathbf{u}_{(-\alpha)})}{f(x_\alpha, \mathbf{u}_{(-\alpha)})} \right]^2 f(x_\alpha, \mathbf{u}_{(-\alpha)}, v) d\mathbf{u}_{(-\alpha)} dv \int K^2(\xi_\alpha) d\xi_\alpha,$$

and

$$\sum_{j=1}^n E \left(Z_{n,j}^2 / \mathcal{F}_{n,j-1} \right) \rightarrow \int \int \left[\frac{J \left\{ F_{(x_\alpha, \mathbf{u}_{(-\alpha)})}(v) \right\}}{\int_{-\infty}^{+\infty} J' \left\{ F_{(x_\alpha, \mathbf{u}_{(-\alpha)})}(y) \right\} \left\{ F'_{(x_\alpha, \mathbf{u}_{(-\alpha)})}(y) \right\}^2 dy} \frac{f(\mathbf{u}_{(-\alpha)})}{f(x_\alpha, \mathbf{u}_{(-\alpha)})} \right]^2 f(x_\alpha, \mathbf{u}_{(-\alpha)}, v) d\mathbf{u}_{(-\alpha)} dv \int K^2(\xi_\alpha) d\xi_\alpha.$$

Under assumptions (A1) to (A4), conditions 3.19 and 3.20 of Corollary 3.1 of Hall and Heyde (1981, pp 58) are satisfied, so that we finally get

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

where

$$V = \int \int \left[\frac{J \left\{ F_{(x_\alpha, \mathbf{u}_{(-\alpha)})}(v) \right\}}{\int_{-\infty}^{+\infty} J' \left\{ F_{(x_\alpha, \mathbf{u}_{(-\alpha)})}(y) \right\} \left\{ F'_{(x_\alpha, \mathbf{u}_{(-\alpha)})}(y) \right\}^2 dy} \frac{f(\mathbf{u}_{(-\alpha)})}{f(x_\alpha, \mathbf{u}_{(-\alpha)})} \right]^2 f(x_\alpha, \mathbf{u}_{(-\alpha)}, v) d\mathbf{u}_{(-\alpha)} dv \int K^2(\xi_\alpha) d\xi_\alpha.$$

This can be written with simpler variable notations

$$V = \int \int \left[\frac{J \{ F_{\mathbf{x}}(v) \}}{\int_{-\infty}^{+\infty} J' \{ F_{\mathbf{x}}(y) \} \{ F'_{\mathbf{x}}(y) \}^2 dy} \frac{f(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x})} \right]^2 f(\mathbf{x}, v) d\mathbf{x}_{(-\alpha)} dv \int K^2(t) dt,$$

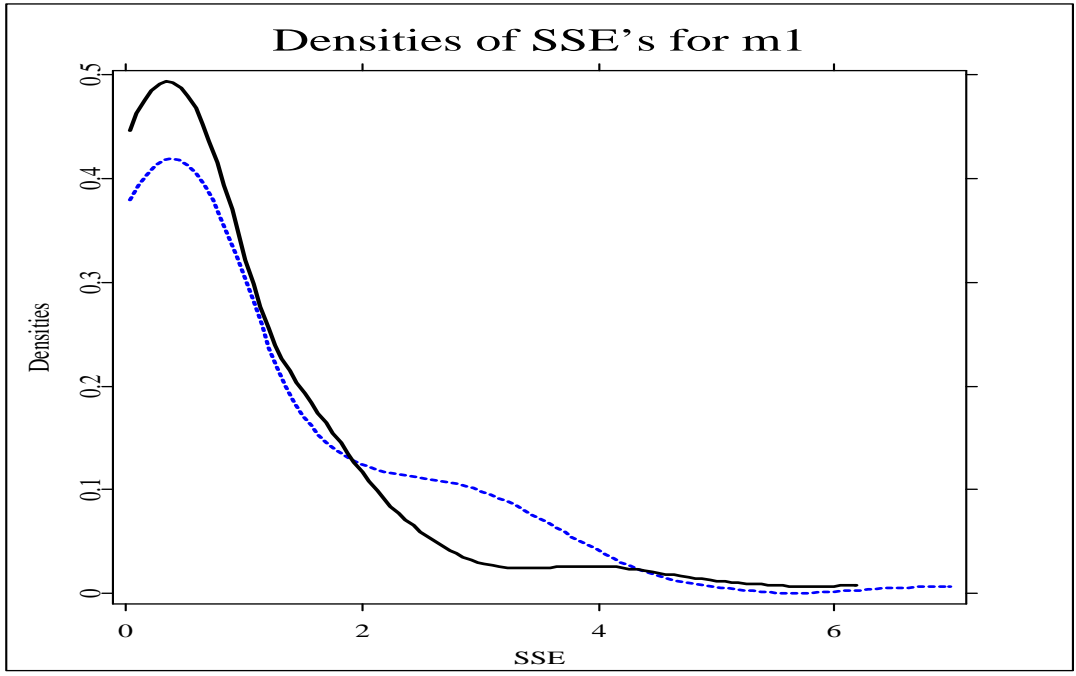
i.e.

$$V = \int K^2(t) dt \int_0^1 J^2(t) dt \int \frac{f^2(\mathbf{x}_{(-\alpha)})}{f(\mathbf{x}) \left[\int_{-\infty}^{+\infty} J' \{ F_{\mathbf{x}}(y) \} \{ F'_{\mathbf{x}}(y) \}^2 dy \right]^2} d\mathbf{x}_{(-\alpha)}.$$

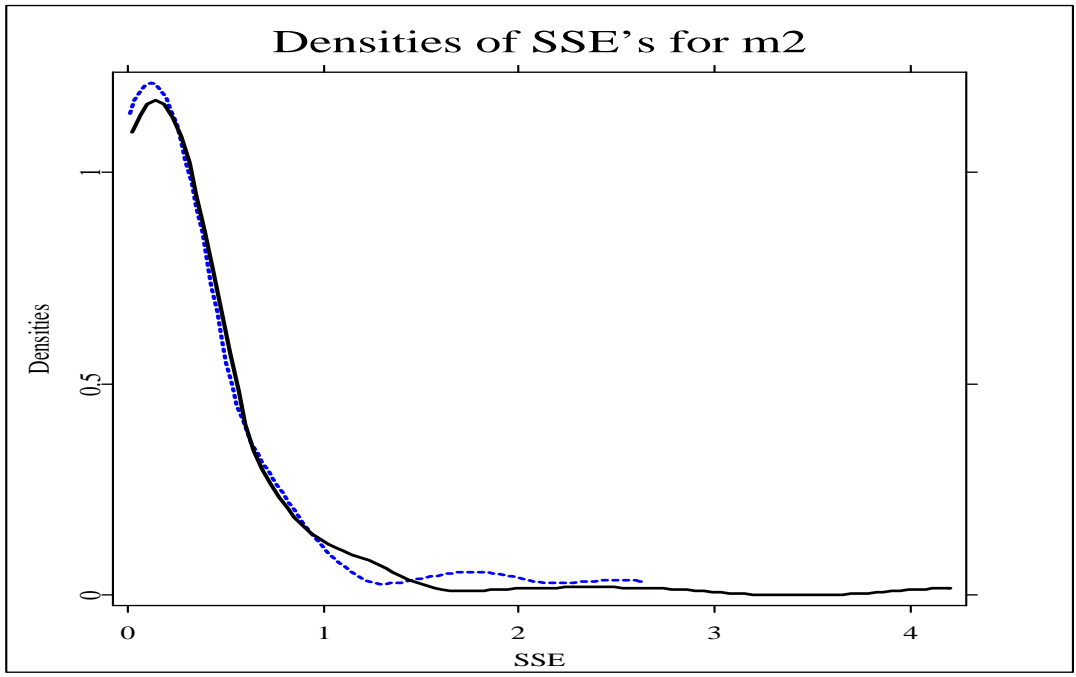
REFERENCES

- AÏT-SAHALIA, Y. (1995). The delta method for nonlinear kernel functionals. Ph. D. dissertation, University of Chicago, USA.
- BIANCO, A. & BOENTE, G. (1998). Robust kernel estimators for additive models with dependent observations. *The Canadian Journal of Statistics* **26**, 239-255.
- CHENG, K. F. & CHENG P. E. (1990). Asymptotic normality for robust R-estimators of regression function. *Journal of Statistical Planning and Inference* **24**, 137-149.
- CHÈZE, N., POGGI, J. M., & 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. & 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. & TIBSHIRANI, R. J. (1990). *Generalized Additive Models*. London: Chapman and 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. & 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. & 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. *Journal 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.
- SERFLING R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York : Wiley.
- STONE, C. (1985). Additive regression and other nonparametric models. *Annals of Statistics* **30**, 689-705.
- TJØSTHEIM, D. & AUESTAD, B. H. (1994). Nonparametric identification of non-linear time series : projections. *Journal of the American Statistical Association* **89**, 1398-1409.
- TAMINE, J. (2002). Smoothed influence function : another view at robust non-parametric regression. Discussion paper 62, SFB 373, Humboldt Universität zu Berlin.
- WATSON, G. S. (1964). Smooth regression analysis. *Sankhya Ser A*. **26**, 359-372.



(a)



(b)

Figure 2: Kernel estimates of density function of the SSE's of estimators of m_α , 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