Estimation of a semiparametric mixture of regressions model

Pierre Vandekerkhove [*]

Université Paris-Est
LAMA (UMR 8050), UPEMLV
F-77454, Marne-la-Vallée, France
and
Georgia Institute of Technology,
UMI Georgia Tech - CNRS 2958,
George W. Woodruff School of Mechanical Engineering
Atlanta, GA 30332-0405, USA

Abstract

We introduce in this paper a new mixture of regressions model which is a generalization of the semiparametric two-component mixture model studied in Bordes et al. (2006b). Namely we consider a two-component mixture of regressions model in which one component is entirely known while the proportion, the slope, the intercept and the error distribution of the other component are unknown. Our model is said to be semiparametric in the sense that the probability density function (pdf) of the error involved in the unknown regression model cannot be modeled adequately by using a parametric density family. When the pdf's of the errors involved in each regression model are supposed to be zero-symmetric, we propose an estimator of the various (Euclidean and functional) parameters of the model, and establish under mild conditions their almost sure rates of convergence. Finally the implementation and numerical performances of our method are discussed using several simulated datasets and one real hight-density array dataset (ChIP-mix model).

Keywords. M-estimator, mixture, regression model, empirical process, semiparametric identifiability, uniform convergence rate.

AMS 2000 subject classifications. 62G05, 62G20, 62G08; secondary 62E10

---

[*]Email: pierre.vandek@univ-mlv.fr

1

1 Introduction

Let $U = (U_i)_{i \geq 1}$ be a label sequence of independent and identically distributed (iid) random variables according to a *Bernoulli* distribution with parameter $p \in (0, 1)$. We consider an iid sample $(Z_1, \ldots, Z_n)$ where for all $i = 1, \ldots, n$, $Z_i = (X_i, Y_i)$ is a bivariate random variable defined, relative to U_i, as follows:

$$\begin{cases} Y_i = a_0 + b_0 X_i + \varepsilon_i^{[0]}, & \text{if} \quad U_i = 0, \\ Y_i = a_1 + b_1 X_i + \varepsilon_i^{[1]}, & \text{if} \quad U_i = 1, \end{cases} \tag{1}$$

where the design sequence $X = (X_i)_{i \geq 1}$, is a sequence of iid random variables with cumulative distribution function (cdf) H and probability density function (pdf) h, when the errors $\varepsilon^{[j]} = (\varepsilon_i^{[j]})_{i \geq 1}$, $j = 0, 1$, are respectively sequences of iid random variables with cdf F_j and pdf f_j. In addition we suppose that the sequences U, X, and $\varepsilon^{[j]}$, $j = 0, 1$, are mutually independent. This model, called the 2-*component mixture of regressions model*, belongs to the wide class of mixture of regressions models which has been studied in Zhu and Zhang (2004); see also Yau et al. (2003) in a LOS (length of stay) medical problem, Bouveyron and Jacques (2010) for prediction, Young and Hunter (2010) in a nonparametric modelling context, or Hunter and Young (2012) for a very recent work on a semiparametric EM-type algorithm for general mixture of regressions models. Recently Martin-Magniette et al. (2008) introduced this model in microarray analysis, specifically for the two color ChIP-chip experiment. In their work, these authors suppose that the error sequences $(\varepsilon_i^{[j]})_{i \geq 1}$, $j = 0, 1$, are such that $\varepsilon_i^{[j]} = \varepsilon_i$ for all $(j, i) \in \{0, 1\} \times \mathbb{N}^*$, where $\mathbb{N}^* := \mathbb{N} \setminus \{0\}$ and ε_i is a Gaussian random variable with mean 0 and variance σ^2 (homoscedaticity with respect to the so-called *probe status* U_i).

In this work, we propose to weaken this last assumption while completely specifying the regression model under the probe standard condition (the parameter $\theta^{[0]} := (a_0, b_0) \in \mathbb{R}^2$ and f_0 are supposed to be entirely known). Note that this kind of assumption arises naturally in microarray analysis. See model Bordes et al. (2006) and references Benjamini and Hochberg (1995), Efron (2007), or Bordes et al. (2006) where analytic expression of f_0, characterizing probe expressivity levels under a certain standard condition, is assumed to be available (generally derived from training data and probabilistic computations). In particular we will suppose that, in model (1), the distribution of the $\varepsilon_i^{[1]}$ is seen as a nuisance parameter (it cannot be modeled using a parametric distribution family), causing model (1) to be a purely semiparametric model. Note that when $\theta^{[0]}$ is known the observations Y_i, for $i = 1, \ldots, n$, can be centered according to $Y_i := Y_i - (a_0 + b_0 X_i)$ which implies the following simplification of model (1):

$$\begin{cases} Y_i = \varepsilon_i^{[0]}, & \text{if} \quad U_i = 0, \\ Y_i = \alpha + \beta X_i + \varepsilon_i^{[1]}, & \text{if} \quad U_i = 1, \end{cases} \tag{2}$$

where $\alpha := a_1 - a_0$ and $\beta := b_1 - b_0$. We suppose that in model (2), which is from now on our model of interest, the $Z_i = (X_i, Y_i)$'s distribution admits a pdf with

respect to the Lebesgue measure on $\mathbb{R}^2$ given by:

$$
\begin{aligned}
g(x,y) &= h(x)g_{Y|X=x}(y) \\
&= h(x)[pf(y-(\alpha+\beta x))+(1-p)f_0(y)], \qquad (x,y)\in\mathbb{R}^2, \qquad (3)
\end{aligned}
$$

where f denotes the unknown pdf of the $\varepsilon_i^{[1]}$, f_0 the known pdf of the $\varepsilon_i^{[0]}$, h the unknown pdf of the X_i, with f and f_0 assumed to be zero-symmetric densities. We will finally denote by $\vartheta := (p,\alpha,\beta) \in (0,1)\times\mathbb{R}^2$ the unknown Euclidean parameter of model (3). Model (2) corresponds exactly to a contaminated version of the semiparametric additive regression model studied in Cuzick (1992a,b) and more recently in Yu et al. (2011). On the other hand, this work extends for the first time to the bivariate case, the asymptotic study (consistency and convergence rates) of the semiparametric class of mixture models introduced by Hall and Zhou (2003) for $\mathbb{R}^s$-valued observations with $s \geq 3$. We remind that semiparametric mixture models were also studied in the univariate case, through two specific models:

$$
g(y) = pf(y-\mu_1)+(1-p)f(y-\mu_2), \qquad y\in\mathbb{R}, \qquad (4)
$$

where $(p,\mu_1,\mu_2) \in (0,1/2)\times\mathbb{R}^2$ and f, supposed to be even, are unknown, see Bordes et al. (2006), Hunter et al (2007), Martin-Magniette et al. (2008), and

$$
g(y) = pf(y-\mu)+(1-p)f_0(y), \qquad y\in\mathbb{R}, \qquad (5)
$$

where $(p,\mu) \in (0,1)\times\mathbb{R}$ and f are unknown, f_0 is known, and the pdfs f and f_0 are supposed to be even, see Bordes et al. (2006), Bordes and Vandekerkhove (2010).

The paper is organized as follows. In Section 2 we present an M-estimating method, inspired by Bordes et al (2006a), Bordes et al. (2006b) and Bordes and Vandekerkhove (2010), that allows us to estimate the Euclidean and the functional parameters of model (2). In Section 3 we address the semiparametric identifiability problem in expression (3) and establish rates of convergence of our estimators. In Section 4 we discuss the performances of our method on simulated examples and also analyze one real high-density array dataset previously considered in Martin-Magniette et al. (2008). Technical results and tedious proofs are relegated to the appendix, Section 5.

2   Estimating method

In this section we generalize the estimating method proposed by Bordes and Vandekerkhove (2010) for model (5) and a fixed unknown location parameter μ, to model (3), where the location parameter, given $\{X = x\}$, is equal to $\alpha + \beta x$ and changes with x due to the non-zero slope parameter β. The main challenge of this paper will consist then in establishing uniform asymptotic results close to those obtained in Bordes and Vandekerkhove (2010) in spite of the introduction of this multiplicative parameter. In particular, the difficulty generated by this change relates to the

statement of minimal conditions on the functional parameters of the model in order to allow a uniform control on the (bigger) classes of functions we will apply to the (X, Y)-empirical process to get convenient approximation results (see, *e.g.*, the treatment of the terms $T_{1,n}^{(i)}$, $i = 1, 2$, in Section 6.6). In the spirit of Bordes et al. (2006a), Bordes et al. (2006b) and Bordes and Vandekerkhove (2010), we will assume that f and f_0 are both zero-symmetric pdfs. To avoid trivial situations or trivial non-identifiability problems (see Remark in Section 3.1), we will impose $p \neq 1$ and $\theta := (\alpha, \beta) \in \Phi \subset \mathbb{R} \times \mathbb{R}^*$, where $\mathbb{R}^* := \mathbb{R} \setminus \{0\}$, which implies that the Euclidean parameter ϑ will be assumed to belong to a parametric compact and convex space

$$\Theta := [\delta, 1 - \delta] \times \Phi \subset (0, 1) \times \{\mathbb{R} \times \mathbb{R}^*\}, \tag{6}$$

where $\delta \in (0, 1)$. For simplicity, we will endow the spaces $\mathbb{R}^s$, $s \geq 1$, with the $\| \cdot \|_s$ norm (for clarity the dimension s is recalled in index) defined for all $v = (v_1, \ldots, v_s)$ by $\|v\|_s = \sum_{j=1}^s |v_j|$ where $|\cdot|$ denotes the absolute value.
Let us introduce now the following notation:

$$\theta \odot x := \alpha + \beta x, \qquad (\theta, x) \in \Phi \times \mathbb{R}.$$

Following the ideas developed by the authors mentioned above, it is possible to use the symmetry of f to identify the true value of the Euclidean parameter. The original idea of this paper consists of noticing that for θ fixed in Φ, the sample $(Y_1^\theta, \ldots, Y_n^\theta)$ obtained by considering the following θ-depending transformation of the original dataset (so-called for simplicity θ-*transformation*),

$$Y_i^\theta := Y_i - \theta \odot X_i, \qquad i = 1, \ldots, n, \tag{7}$$

is distributed according to

$$\Psi_\theta(y) = p_* \int_{\mathbb{R}} f(y + (\theta - \theta_*) \odot x) h(x) dx + (1 - p_*) \int_{\mathbb{R}} f_0(y + \theta \odot x) h(x) dx, \tag{8}$$

where $\vartheta_* = (p_*, \alpha_*, \beta_*) \in \mathring{\Theta}$ denotes the true value of the parameter. In addition, when $\theta = \theta_*$,

$$\Psi_{\theta_*}(y) = p_* f(y) + (1 - p_*) \int_{\mathbb{R}} f_0(y + \theta_* \odot x) h(x) dx, \tag{9}$$

which is very close to model (5) studied in Bordes et al. (2006b) where the location μ is known but the proportion p is unknown. These two major remarks are illustrated and discussed on a simulated example in Section 4.1.

Isolating f in (9) and replacing $\vartheta_* = (p_*, \theta_*)$ by $\vartheta = (p, \theta)$ one can define a new parametric class of functions $\mathcal{F}_\Theta := \{f_\vartheta : \vartheta \in \Theta\}$ where for all $(y, \vartheta) \in \mathbb{R} \times \Theta$

$$f_\vartheta(y) = \frac{1}{p} \Psi_\theta(y) - \frac{1 - p}{p} \int_{\mathbb{R}} f_0(y + \theta \odot x) h(x) dx, \tag{10}$$

that satisfies under $\vartheta = \vartheta_*$, and all $y \in \mathbb{R}$

$$f(y) = f_{\vartheta_*}(y) = f_{\vartheta_*}(-y) = f(-y). \tag{11}$$

4

Note that, on the right hand side of (10), the second integral term is unknown but can be estimated pointwise by a standard integration Monte Carlo method, see expression (16), or a nonparametric Monte Carlo approach, see expression (22). The intuition consists now in claiming that, if we vary ϑ over Θ and that we are able to check that f_ϑ is zero-symmetric for a certain value of ϑ, then we have reached the true value of the Euclidean parameter. A natural idea to detect this kind of situation, and then to estimate $\vartheta = (p, \theta)$, is to consider a contrast function based on the comparison between the cdf version of $f_\vartheta(y)$

$$H_1(y; \vartheta) := H_1(y; p, F_\theta, J_\theta) := \frac{1}{p} F_\theta(y) - \frac{1-p}{p} J_\theta(y), \qquad (y, \theta) \in \mathbb{R} \times \Theta,$$

and the cdf version of $f_\vartheta(-y)$

$$H_2(y; \vartheta) := H_2(y; p, F_\theta, J_\theta) := 1 - \frac{1}{p} F_\theta(-y) + \frac{1-p}{p} J_\theta(-y), \qquad (y, \theta) \in \mathbb{R} \times \Theta,$$

where for all $(y, \theta) \in \mathbb{R} \times \Phi$,

$$J_\theta(y) := \int_{-\infty}^{y} I_\theta(z)dz, \quad y \in \mathbb{R}, \quad \text{with } I_\theta(z) := \int_{\mathbb{R}} f_0(z + \theta \odot x)h(x)dx, \quad z \in \mathbb{R},$$

and $F_\theta(y) := \int_{-\infty}^{y} f_\theta(z)dz$. Notice that for all θ fixed in Φ, $J_\theta(\cdot)$ and $F_\theta(\cdot)$ are the cdfs respectively associated to the θ-transformed known component population (the Y_i such that $U_i = 0$ in (2)) and the θ-transformed whole data (the Y_i such that $U_i = 0$ or $U_i = 1$ in (2)). It is then convenient to define the *comparison*-function

$$H(y; \vartheta) := H_1(y; \vartheta) - H_2(y; \vartheta), \qquad (y, \vartheta) \in \mathbb{R} \times \Theta. \tag{12}$$

and to notice that under ϑ_*, using the symmetry of f,

$$H(y; \vartheta_*) = 0, \qquad y \in \mathbb{R}.$$

To avoid numerical integration in the approximation of an empirical contrast function, based on the comparison function H over $\mathbb{R}$, we proceed as follows. Let Q be an *instrumental* weight probability distribution with pdf q with respect to the Lebesgue measure. We assume that q is strictly positive over $\mathbb{R}$ and easy to simulate. Then we consider

$$d(\vartheta) := \int_{\mathbb{R}} H^2(y, \vartheta)dQ(y), \tag{13}$$

where obviously $d(\vartheta) \geq 0$ for all $\vartheta \in \Theta$ and $d(\vartheta_*) = 0$. Let $(V_1, \ldots, V_n)$ be an iid sample from Q. An empirical version $d_n(\cdot)$ of $d(\cdot)$ can be obtained by considering

$$d_n(\vartheta) := \frac{1}{n} \sum_{i=1}^{n} H^2(V_i; p, \tilde{F}_{n,\theta}, \hat{J}_{n,\theta}), \qquad \vartheta \in \Theta, \tag{14}$$

where

$$\hat{J}_{n,\theta}(y) := \int_{-\infty}^{y} \hat{I}_{n,\theta}(z)dz, \qquad (y, \theta) \in \mathbb{R} \times \Phi, \tag{15}$$

5

with

$$\hat{I}_{n,\theta}(z) := \frac{1}{n}\sum_{i=1}^{n} f_0(z + \theta \odot X_i), \qquad (z,\theta) \in \mathbb{R} \times \Phi, \tag{16}$$

which leads actually to the simple expression for $\hat{J}_{n,\theta}(y)$

$$\hat{J}_{n,\theta}(y) := \frac{1}{n}\sum_{i=1}^{n} F_0(y + \theta \odot X_i), \qquad (y,\theta) \in \mathbb{R} \times \Phi. \tag{17}$$

In (14), the function $\tilde{F}_{n,\theta}$ denotes a smooth version of the empirical cdf

$$\hat{F}_{n,\theta}(y) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}_{Y_i^\theta \leq y}, \qquad (y,\theta) \in \mathbb{R} \times \Phi,$$

defined by

$$\tilde{F}_{n,\theta}(y) := \int_{-\infty}^{y} \hat{\Psi}_{n,\theta}(t)dt, \qquad (y,\theta) \in \mathbb{R} \times \Phi, \tag{18}$$

where

$$\hat{\Psi}_{n,\theta}(t) := \frac{1}{nb_n}\sum_{i=1}^{n} K\left(\frac{t - Y_i^\theta}{b_n}\right), \qquad (t,\theta) \in \mathbb{R} \times \Phi. \tag{19}$$

In (19), we assume the standard condition ensuring, for each $\theta \in \Phi$, the L_1 convergence of $\hat{\Psi}_{n,\theta}$ towards Ψ_θ defined in (8) (see Devroye 1983), namely

$$b_n \to 0, \qquad nb_n \to +\infty, \tag{20}$$

and K is a symmetric kernel density function. Finally we propose to estimate ϑ_* by considering the M-estimator

$$\hat{\vartheta}_n := (\hat{p}_n, \hat{\theta}_n) = \arg\min_{\vartheta \in \Theta} d_n(\vartheta). \tag{21}$$

Once ϑ_* is estimated by $\hat{\vartheta}_n$ a natural way to estimate F and f consistently is then to consider the plug-in empirical versions of $H_1(\cdot; \vartheta)$ and (10), respectively defined for all $y \in \mathbb{R}$ by

$$\begin{aligned} \hat{F}_n(y) &:= H_1(y; \hat{p}_n, \tilde{F}_{n,\hat{\theta}_n}, \tilde{J}_{n,\hat{\theta}_n}), \\ \hat{f}_n(y) &:= \frac{1}{\hat{p}_n}\hat{\Psi}_{n,\hat{\theta}_n}(y) - \frac{1-\hat{p}_n}{\hat{p}_n}\tilde{I}_{n,\hat{\theta}_n}(y), \end{aligned}$$

where, for all $\theta \in \Theta$, $\tilde{I}_{n,\theta}$ and $\tilde{J}_{n,\theta}$ are respectively nonparametric estimators of I_θ and J_θ based on an iid simulated sample $(\tilde{\varepsilon}_1^{[0]}, \ldots, \tilde{\varepsilon}_n^{[0]})$ from f_0 obtained by considering

$$\begin{aligned} \tilde{I}_{n,\theta}(t) &:= \frac{1}{nb_n}\sum_{i=1}^{n} K\left(\frac{t - (-\theta \odot X_i + \tilde{\varepsilon}_i^{[0]})}{b_n}\right), \qquad (t,\theta) \in \mathbb{R} \times \Phi. \\ \tilde{J}_{n,\theta}(y) &:= \int_{-\infty}^{y} \tilde{I}_{n,\theta}(t)dt, \qquad (y,\theta) \in \mathbb{R} \times \Phi, \end{aligned}$$

6

In this second plug-in step, we consider, for the sake of simplicity in our proofs, the nonparametric estimates (22) and (22) instead of (15) and (16). This choice allows us to use similar nonparametric results for both $\hat{f}_{n,\theta}$ and $\tilde{I}_{n,\theta}$ (see the proof of Theorem 3.1 ii) and iii)), but the same results should be obtained, at the price of an additional technical lemma, by considering directly the Monte Carlo estimators (15) and (16).

3   Semiparametric identifiability and consistency

3.1   Identifiability

In this section we recall briefly why model (3) is identifiable under conditions similar to those established in Bordes et al. (2006b) and summarized below. Let us define $\mathcal{F}_s := \{f \in \mathcal{F}; \int_{\mathbb{R}} |x|^s f(x) dx < +\infty\}$ for $s \geq 1$, where $\mathcal{F}$ denotes the set of even pdfs. When $(f, f_0) \in \mathcal{F}_s$ with $s \geq 2$, we denote $m := \int_{\mathbb{R}} x^2 f(x) dx$ and $m_0 := \int_{\mathbb{R}} x^2 f_0(x) dx$. In addition we will denote by $\lambda^{\otimes 2}$ the product Lebesgue measure on $\mathbb{R}^2$.

Definition 3.1 *(Identifiability). Let $(p_1, \theta_1, f_1, h_1)$ and $(p_2, \theta_2, f_2, h_2)$ denote two sets of parameters for model (3). The parameter in model (3) is said to be semi-parametrically identifiable if*

$$(p_1, \theta_1, f_1(y), h_1(x)) = (p_2, \theta_2, f_2(y), h_2(x)),$$

for $\lambda^{\otimes 2}$-almost all $(x, y) \in \mathbb{R}^2$, whenever we have:

$$(p_1 f_1(y - \theta_1 \odot x) + (1 - p_1) f_0(y)) h_1(x)$$
$$= (p_2 f_2(y - \theta_2 \odot x) + (1 - p_2) f_0(y)) h_2(x), \qquad (22)$$

for $\lambda^{\otimes 2}$-almost all $(x, y) \in \mathbb{R}^2$.

Lemma 3.1 *If the Euclidean parameter space Θ is a subset of $\mathbb{R} \times \mathbb{R}^*$, $\mathrm{supp}(f) = \mathrm{supp}(f_0) = \mathbb{R}$, $\mathrm{supp}(h)$ contains at least two intervals respectively in the neighborhood of 0 and $+\infty$ (or $-\infty$), and the pdfs involved in model (3) satisfy $(f_0, f) \in \mathcal{F}_3 \times \mathcal{F}_3$, then the parameter in model (3) is identifiable.*

Proof. Integrating (22) with respect to y over $\mathbb{R}$, we then obtain that $h_1(\cdot) = h_2(\cdot)$ λ-almost everywhere. Let $h(x) := h_1(x)$ for all $x \in \mathrm{supp}(h) := \mathrm{supp}(h_1) \cap \mathrm{supp}(h_2)$. Notice now that, for all $x \in \mathrm{supp}(h)$, (22) coincides with (5) when we replace the location parameter μ by $\theta \odot x$. In our case, the first three conditional moment equations given $\{X = x\}$ associated to (22) lead to

$$\begin{cases} p_1(\theta_1 \odot x) = p_2(\theta_2 \odot x), \\ (1 - p_1)m_0 + p_1((\theta_1 \odot x)^2 + m_1) = (1 - p_2)m_0 + p_2((\theta_2 \odot x)^2 + m_2), \quad (23) \\ p_1(3((\theta_1 \odot x)m_1 + (\theta_1 \odot x)^3)) = p_2(3((\theta_2 \odot x)m_2 + (\theta_2 \odot x)^3)). \end{cases}$$

Replacing in Bordes et al. (2006b) formula (6) the location μ by $\theta_1 \odot x$ and noticing that in our case $\theta = m_1$ and $\theta_0 = m_0$, the solutions are either, for all $x \in \mathrm{supp}(h)$,

$(p_1, \theta_1 \odot x) = (p_2, \theta_2 \odot x)$, which implies $(p_1, \alpha_1, \beta_1) = (p_2, \alpha_2, \beta_2)$, or

$$\begin{cases} p_2 & = p_1 \left(\dfrac{2(\theta_1 \odot x)^2}{3m_1 + (\theta_1 \odot x)^2 - 3m_0} \right), \\[2ex] \theta_2 \odot x & = \theta_1 \odot x + \dfrac{3m_1 - (\theta_1 \odot x)^2 - 3m_0}{2\theta_1 \odot x}, \\[2ex] m_2 & = m_1 + \dfrac{(m_1 + (\theta_1 \odot x)^2 - m_0)(3m_0 + (\theta_1 \odot x)^2 - 3m_1)}{4(\theta_1 \odot x)^2}. \end{cases} \qquad (24)$$

Assume that $\beta_1 \neq 0$ and consider the limit when $x \to +\infty$ in the first row of (24). We then necessarily obtain that $p_2 = 2p_1$ which is only compatible when we take the limit as $x \to 0$, with $m_1 = m_0$. Hence if $m_1 \neq m_0$, model (3) is always identifiable. If we assume $m_1 = m_0$, the second row of (24) leads to $\theta_2 \odot x = (\theta_1 \odot x)/2$. If we introduce this last equation in the third row of (24), we obtain

$$m_2 - m_1 = \frac{1}{4}(\theta_1 \odot x)^2, \qquad x \in \mathbb{R},$$

which is impossible when $x \to +\infty$ and thus provides us the global identifiability of model (3). $\qquad \square$

Remark. In Lemma 3.1 we have considered for simplicity the case where the slope parameter β is supposed to be different than zero. Actually, this condition can be technically relaxed if we allow θ to be equal to $(\alpha, 0)$ with $\alpha \neq 0$. In fact, considering the first row of (24) and taking the limit as $x \to +\infty$, we obtain $\beta_2 = \beta_1 = 0$. To conclude, it is then enough to integrate (22) with respect to x over $\mathbb{R}$ which leads to discuss the same condition as in Bordes et al. (2006b, p. 735, eq. 3). Then Proposition 2 in Bordes et al. (2006b) provides an *almost everywhere*-type identifiability result which unfortunately cannot be strictly compared to the result stated in Lemma 3.1. For this reason we decided to reject $\theta = (\alpha, 0)$, $\alpha \in \mathbb{R}^*$ from the sub-parametric space Φ, see (6).

3.2 Assumptions and asymptotic results

In the following we provide some general conditions ensuring the consistency of our estimation method as well as semiparametric almost sure rates of convergence.

Regularity conditions (R).

i) The pdfs f and f_0 are strictly positive over $\mathbb{R}$ and belong to $\mathcal{F}_3$.

ii) The pdfs f and f_0 are twice differentiable over $\mathbb{R}$ with $\|f^{(j)}\|_\infty < \infty$ and $\|f_0^{(j)}\|_\infty < \infty$, where $f^{(j)}$ and $f_0^{(j)}$ denote respectively the j-th order derivatives of f and f_0, for $j = 1, 2$.

iii) The pdf h satisfies $\int_{\mathbb{R}} |x|^2 h(x)dx < \infty$.

iv) For $i = 0$ or $i = 2$,

$$\int_{\mathbb{R}^2} |x|^i |F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)| h(x) dx dy < \infty,$$

and for $i = 1$ or $i = 3$, and all $u \in \mathbb{R}$,

$$\lim_{y \to \pm \infty} y^i (F_0(y + u) - F_0(y - u)) = 0.$$

v) There exist two collections of functions $\{\ell_{i,j}\}_{0 \leq i \leq j \leq 2}$ and $\left\{\ell_{i,j}^0\right\}_{0 \leq i \leq j \leq 2}$ belonging to $L_1(\mathbb{R}^2)$ and such that, for all $(x, y) \in \mathbb{R}^2$ and all $\theta \in \Theta$

$$|x^i f^{(j)}(y + (\theta - \theta^*) \odot x)| h(x) \leq \ell_{i,j}(x, y),$$

and

$$|x^i f_0^{(j)}(y + \theta \odot x)| h(x) \leq \ell_{i,j}^0(x, y).$$

For all $z \in \mathbb{C}$, let denote respectively $\check{z}$ and $\Im(z)$ the conjugate and imaginary part of z. We will also denote $\bar{f}$, $\bar{f}_0$ the Fourier transforms of f, f_0, and define for all $\kappa = (\kappa_1, \kappa_2) \in \mathbb{R}^2$, $\nu_\kappa(t) := e^{it\kappa_1} \bar{h}(\kappa_2 t)$, where $\bar{h}$ denotes the Fourier transform of h.

The following conditions mainly ensure the contrast property for the function d defined in (13). We point out that these conditions are not equivalent, as it is already the case in Bordes and Vandekerkhove (2010, p. 25), to those established to prove the identifiability property in Lemma 3.1.

Contrast condition (C).

i) The density function h is not non-zero symmetric and the three first moments of X satisfy $E(X)(4E(X)^2 + 3E(X^2)) \neq -E(X^3)$.

ii) The set of parameters $\vartheta = (p, \theta) = (p, \alpha, \beta)$ with $p \neq p_*$ that satisfies

$$p_* \Im(\nu_{\theta - \theta_*}(t)) \bar{f}(t) = (p_* - p) \Im(\nu_\theta(t)) \bar{f}_0(t), \qquad t \in \mathbb{R}, \tag{25}$$

is empty or does not belong to the parametric space Θ.

iii) The second order moments of f and f_0, respectively denoted m and m_0, are assumed to satisfy

$$m \neq m_0 + \frac{\alpha_*^3 + 3\alpha_*^2 \beta_* E(X) + 3\alpha_* \beta_*^2 E(X^2) + \beta_*^3 E(X^3)}{3(\alpha_* + \beta_* E(X))}.$$

Remarks. In order to control the non-symmetry of h required in (C) i), one can use an adequate 95 %-confidence interval $\mathcal{I}_{X,95\%}$ for $E(X)$ and check if $0 \in \mathcal{I}_{X,95\%}$, which should make suspect that $E(X) = 0$. If such a situation happens we advice to add a non-null constant c to the X in order to create a likely non-symmetric design data, *i.e.*, $0 \notin \mathcal{I}_{X+c,95\%}$. This modification is slope-invariant and generates a new

value for the intercept which is then equal to $\rho := \alpha - \beta c$. Obviously, if $(\hat{\rho}_n, \hat{\beta}_n)$ is a strongly consistent estimator of (ρ, β) with a certain rate of convergence, then $\hat{\alpha}_n = \hat{\rho}_n + c\hat{\beta}_n$ inherits the same convergence properties. Note, in addition that if h is symmetric about a non-null constant, then condition (C) i) is satisfied since $E(X)$ and $E(X^3)$ have the same sign.

We also point out that condition (C) ii), which is necessary to get the contrast property for d over Θ, is quite difficult to figure out when nothing is specified on f, f_0 and h. We suggest, in the spirit of conditions C1 and C2 in Hohmann and Holzmann (2012), to consider the sufficient and more intuitive *regularity comparison*-type criterion for C ii)

$$\forall \theta \in \Phi, \quad \left| \frac{\Im(\nu_{\theta-\theta_*}(t))}{\Im(\nu_\theta(t))} \frac{\bar{f}(t)}{\bar{f}_0(t)} \right| \longrightarrow +\infty \ \text{ or } \ 0, \ \text{ as } \ t \to +\infty, \tag{26}$$

which is valid since, according to (25), the term on left hand side of (26) is equal to $|p - p_*|/p_* \in [|p - p_*|, 1/\delta]$ which is in contradiction with (26). However, condition (25) is thoroughly discussed in the Gaussian case as done in the appendix, Section 6.1. We prove, in particular, that there exists at most one spurious solution satisfying (25) that has to be removed from the parametric space so it is not detected by our estimation algorithm as shown in Fig. **??**.

Kernel and Bandwidth conditions (K).

  i)   The even kernel density function K is bounded, uniformly continuous, and square integrable, with bounded variation and second order moment.

  ii) The even kernel density function K is twice differentiable and its first and second derivatives belong to $L_1(\mathbb{R})$.

  iii) The bandwidth b_n satisfies $b_n \searrow 0$, $nb_n \to +\infty$ and $\sqrt{n}b_n^2 = o(1)$.

We state in the following lemma some basic properties of the discrepancy function d and its estimator d_n respectively defined in (13) and (14).

Lemma 3.2   *(i) Under conditions (R) the function d is Lipschitz over Θ.*

  (ii) Under conditions (C) i) and ii) the function d is a contrast function, i.e., for all $\vartheta \in \Theta$, $d(\vartheta) \geq 0$ and $d(\vartheta) = 0$ if and only if $\vartheta = \vartheta_*$.

  (iii) Under condition (C) iii) we have

$$\ddot{d}(\vartheta_*) = 2 \int_{\mathbb{R}} \dot{H}(y, \vartheta_*) \dot{H}^T(y, \vartheta_*) dQ(y) > 0.$$

  (iv) Under conditions (R) and (K) i) and iii), for any $\gamma > 0$, d_n converges to d almost surely with the rate

$$\sup_{\vartheta \in \Theta} |d_n(\vartheta) - d(\vartheta)| = o_{a.s.}(n^{-1/2+\gamma}).$$

10

The proof of this lemma is relegated to the appendix, Section 6.6.

We state in the following theorem the asymptotic properties of our Euclidean and functional estimators defined in (21), (22) and (22).

Theorem 3.1 *i) If assumptions (R), (C) and (K) are satisfied then*

$$\|\hat{\vartheta}_n - \vartheta_*\|_3 = o_{a.s.}(n^{-1/4+\gamma}), \qquad \gamma > 0.$$

ii) The estimator $\hat{f}_n$ of f defined in (22) converges almost surely in the L_1 sense if $n^{-1/4+\gamma}/b_n \to 0$, for all $\gamma > 0$.

iii) For any $\gamma > 0$, the estimator $\hat{F}_n$ of F defined in (22) converges uniformly at the following almost sure rate

$$\|\hat{F}_n - F\|_\infty = O_{a.s.}(n^{-1/4+\gamma}/b_n) + O_{a.s.}(b_n^2), \qquad \gamma > 0. \qquad (27)$$

The above rate is optimized by considering $b_n = n^{-1/12}$, providing the rate of convergence $O_{a.s.}(n^{-1/6+\gamma})$, for all $\gamma > 0$.

The proof of this theorem is relegated to the appendix, Section 6.7.

Comment. Points ii) and iii) reveal the intuitive idea that the bandwidth b_n must not decrease too fast in order to allow the appropriate positioning of the plug-in centered data in the expression of $\hat{\Psi}_{n,\hat{\theta}_n}$.

4   Numerical experiments

4.1   Role of the θ-transformation

We propose in this section to highlight the role played by the θ-transformation, see (7), in our method. For this purpose, we consider an example which corresponds to model (5) taking $p_* = 0.7$, $\alpha_* = 2$, $\beta_* = 1$, $\varepsilon_1^{[j]} \sim \mathcal{N}(0,1)$, $j = 0,1$ and $X_1 \sim \mathcal{N}(2,3)$. In Fig. 1 we plot successively a simulated data set $(X_i, Y_i)_{1 \le i \le n}$, corresponding to the previous description with $n = 200$, and the two θ-transformed datasets obtained with $\theta = (1, 0.5)$ and $\theta = \theta^* = (2, 1)$. These figures are completed by adding their corresponding 2nd-coordinate sample data histograms. Note that these histograms are empirical estimates of the densities f_θ defined in (8), with θ respectively equal to $(0, 0)$, $(1, 0.5)$ and $(2, 1)$. We see clearly through these three situations how a progressive transformation of the data allows one to reach a tractable situation in the sense that it looks strongly like the semiparametric contamination model (5) studied in Bordes et al. (2006b) where a known density is mixed with a symmetric unknown density, which corresponds to the behavior observed in the second row, third column histogram in Fig. 1. Loosely speaking the second idea of our method consists in arguing that once θ is close to θ_* we are allowed to estimate the proportion p according to a Bordes and Vandekerkhove (2010) type-method which corresponds
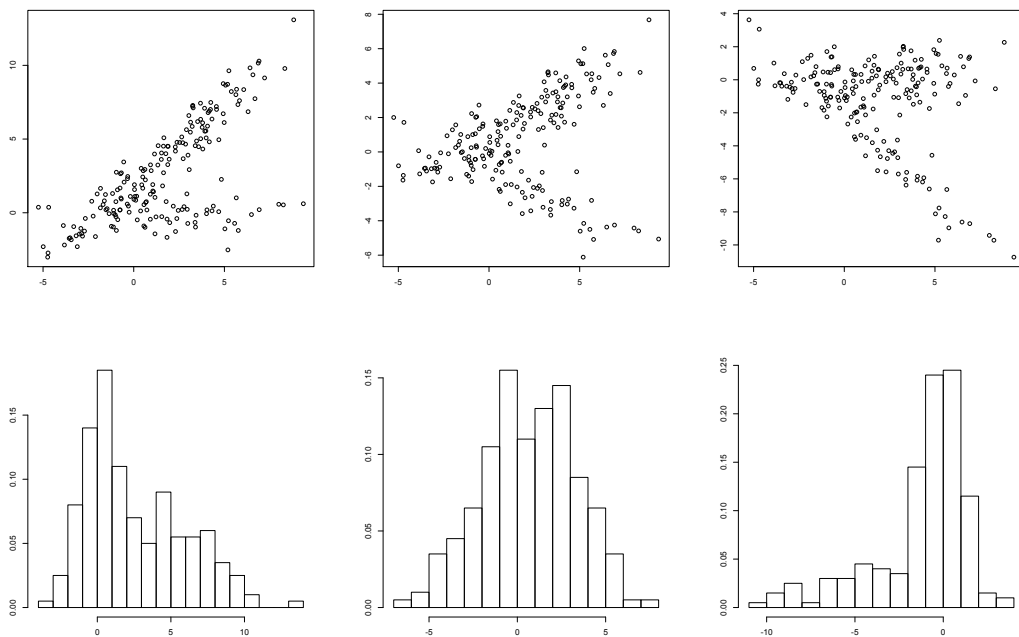
Figure 1: First row: respectively plot of an original data $(X_i, Y_i)_{1 \leq i \leq n}$ according to model (3) with $n = 200$, plot of a wrong θ-tranformation $(\theta \neq \theta^*)$, plot of the true θ^*-tranformation. Second row: respectively histograms of the corresponding first row 2nd-coordinate sample data.

to the minimization step (21). In contrast to this technically satisfying idea, the θ-transformation and the choice of the weight distribution Q introduced in (13) are two sources of difficulty which have to be considered carefully. In fact when β_* is large and the law of the design data has heavy tails with respect to the tails of f, then the θ_* transformation will move the points coming from the F_0-population and located far from the origin to extremely distant positions, which implies intuitively that the integral type density involved in (7) should be extremely heavily tailed. Thus in order to capture the information contained in the tails of the θ-transformed data set it is important to sufficiently weight the empirical index of symmetry $H^2(x; p, \tilde{F}_{n,\theta}, \hat{J}_{n,\theta})$ of expression (14) for large values of x, which reduces to choosing an instrumental distribution Q with non-negligible tails with respect to F_{θ_*}.

4.2 Preliminary discussion

The aim of this section is to illustrate graphically, on two-dimensionnal examples, the behavior of the empirical distance $d_n(p, 0, \beta)$ (the parameter α is assumed to be equal to zero) when p and β lie close to the true value of the parameter. For simplicity the parameter will still be denoted $\vartheta := (p, \beta)$, with $\theta := \beta$ and $d_n(\vartheta) := d_n(p, 0, \beta)$. The interest of this study is to understand closely the influence of the mixing proportion p and the slope parameter β on the shape of the contrast function d (flatness,

12

sharpness, smoothness, etc.). Our models of interest are denoted M1 and M2 and defined according to (2) as follows

M1: $(p_*, \beta_*) = (0.7, 1)$, $V \sim Q = \mathcal{N}(0, 4^2)$, $\varepsilon^{[j]} \sim \mathcal{N}(0, 1)$, $X \sim \mathcal{N}(2, 3^2)$,
M2: $(p_*, \beta_*) = (0.3, 1)$, $V \sim Q = \mathcal{N}(0, 2^2)$, $\varepsilon^{[j]} \sim \mathcal{N}(0, 1)$, $X \sim \mathcal{N}(2, 3^2)$,

where $j = 0, 1$. In Fig. 2 we plot the mapping $(p, \beta) \mapsto d_n(p, \beta)$ obtained from an M1-sample, respectively M2-sample, of size $n = 100$, where $(p, \beta) \in \Theta_1 = [0.5, 0.8] \times [0.9, 1.1]$, respectively $(p, \beta) \in \Theta_2 = [0.1, 0.6] \times [0.6, 1.4]$. Notice that according to discussion (CG) at the end of Section 5.1, since $p_* < 1/2$ and $m = m_0$, model M2 is not necessarily consistently estimated since the parameter space Θ_2 contains the spurious solution $\vartheta_{**} = (2p_*, \beta_*/2)$ (situation we voluntary want to investigate from the numerical point of view). Fig. 2 which plots the level curves of d_n eval-
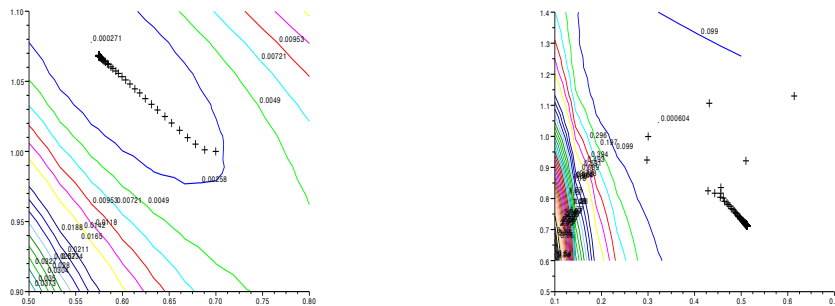

Figure 2: Plot of $(p, \beta) \mapsto d_n(p, 0, \beta)$ with $n = 100$, $\beta_* = 1$, $\varepsilon_1^{[j]} \sim \mathcal{N}(0, 1)$, $j = 0, 1$, $X_1 \sim \mathcal{N}(2, 3^2)$, with the difference that on the left hand side $p_* = 0.7$, $V_1 \sim \mathcal{N}(0, 4^2)$, when on the right hand side $p_* = 0.3$ $V_1 \sim \mathcal{N}(0, 2^2)$.

uated on a homogeneous 10×10 grid of the rectangular domain $[0.5, 0.8] \times [0.9, 1.1]$. Fig. 2 shows that the graph of d_n looks like a sharp valley with a flat trough when β is located near β^* and p ranges $[0.5, 0.8]$. Even if on this simulated example the argmin of d_n is very close to the true value of the parameter, the previous remark suggests that the estimation of the mixing proportion will be less robust than the estimation of slope parameter. The observation of the second plot in Fig. 2 is more unexpected since the graph of d_n does not really look like a contrast function, with its high near $p = 0.1$ and its very large and flat trough that covers most of Θ_2, suggesting more instability in our estimating method in that kind of situation. To validate these thoughts, we propose applying a large sample study of these two examples and a third intermediary one obtained by considering model M2 with $V \sim \mathcal{N}(0, 4^2)$. The results of this study are summarized in Table 1.

To compute our M-estimator $\hat{\vartheta}_n = (\hat{p}_n, \hat{\beta}_n)$ we programmed a constrained quasi-Newton FBGS optimization procedure, see Nocedal and Wright (1999), applied to

the score function $\dot{d}_n := \left(\frac{\partial}{\partial p} d_n, \frac{\partial}{\partial \beta} d_n \right)^T$ and initialized at the true value of the parameter ϑ_*. Note that in practice, intercept and slope parameters deduced from an "imaginary" straight line crossing the two extremities of the cluster likely labeled 1 and a proportion ratio evaluated by guesswork, provide in general a good initial position candidate for the optimization procedure.

The kernel K used to compute (19) is a standard Gaussian kernel and the bandwidth $b_n = \sqrt{1 + 4p(1-p)}(4/(3n))^{1/5}$ (proposed by Bowman and Azzalini (2003) for Gaussian distributions and implemented in R), both obviously satisfying condition (K). Two examples of stabilization/convergence of our gradient optimization procedure are illustrated in Fig 2, where the successive positions (until stabilization) of our algorithm are depicted by cross symbols.

Table 1: Mean and Std. Dev. of 100 estimates of (p_*, β_*).

n	(p_*, β_*, σ_V)	Empirical means	Standard deviation
100	(0.7,1,4)	(0.705,1.005)	[0.037,0.069]
200	(0.7,1,4)	(0.697,0.996)	[0.030,0.059]
500	(0.7,1,4)	(0.695,1.005)	[0.029,0.035]
100	(0.3,1,4)	(0.310,0.958)	[0.057,0.125]
200	(0.3,1,4)	(0.296,0.985)	[0.050,0.085]
500	(0.3,1,4)	(0.297,1.017)	[0.028,0.041]
100	(0.3,1,2)	(0.397,0.858)	[0.094,0.221]
200	(0.3,1,2)	(0.398,0.914)	[0.083,0.190]
500	(0.3,1,2)	(0.331,0.968)	[0.052,0.106]

Comments on Table 1. First of all, it is interesting to compare the performances summarized in rows 1–3 of Table 1 to those obtained in Bordes and Vandekerkhove (2010, p. 35, Table 1) where the model of interest is (5), with $p = 0.7$, $\mu = 3$, and f_0 and f are respectively the pdfs corresponding to the $\mathcal{N}(0,1)$ and $\mathcal{N}(0,(1/2)^2)$ distributions. Even if these two models are not strictly comparable, we think that this comparison is very important to better understand the influence of the θ-transformation and the instrumental distribution Q on the performances of our generalized estimation method. From the numerical point of view, we see that the bias of our estimators, for both models, is negligible. However it also appears that the standard deviation associated to $(\hat{p}_n, \hat{\beta}_n)$ decreases significantly slower than the standard deviation associated to $(\hat{p}_n, \hat{\mu}_n)$ when n grows. The performance summarized in rows 4–6 of Table 1, which corresponds to $p = 0.3$ (which means that the population that will move far from its initial position due to the θ-transformation will be more important), is very highlightning. We observe that for small n ($n = 100, 200$) the standard deviations associated to $(\hat{p}_n, \hat{\beta}_n)$ are dramatically larger compared to those obtained with $p = 0.7$. Let denote by $\mathrm{std}(n, p_*, \beta_*, \sigma_V)$ the couple of standard deviations calculated in the last column of Table 1 under $(n, p_*, \beta_*, \sigma_V)$. If we compute componentwise the ratios $\mathrm{std}(n, 0.3, 1, 4)/\mathrm{std}(n, 0.7, 1, 4)$ respectively for $n = 100, 200, 500$, we obtain

14

approximately $(1.54, 1.8)$, $(1.67, 1.44)$, and $(0.95, 1.17)$, which suggests that when n becomes large the side effect of the θ-transformation vanishes (probably thanks to the size of n, which increases globally the precision of the empirical estimates, and the tails of Q, that allow the algorithm to take these improvements into account efficiently). The performances summarized in rows 7–9 of Table 1, confirms the concerns expressed about model M2. We recall that model M2 is badly affected by the two following drawbacks : smallness of p_* (synonymous with important population shifted far by the θ-transformation and existence of a spurious solution) and a smallness of σ_V which is then clearly not sufficient to counteract the smallness of p_* (and its consequences). In particular we think that, in model M2, the empirical contrast d_n is more easily closer to 0 under $\beta_{**} = \beta_*/2$ since as explained in Section 4.1., this value is then significantly smaller than β_*. This last remark explains why, in spite of the fact that our algorithms were initialized at the true parameter value, our estimates are strongly biased (attracted quite often by the spurious solution ϑ_{**}).

4.3 Large sample study

We propose in this section to investigate the numerical performances of our method in more general situations through the four following models: WO (Weakly Overlapped), MO (Moderately Overlapped), SO (Strongly Overlapped), and MIX (MIXture noise) respectively defined, for all $i \geq 1$, by:

WO: $(p_*\alpha_*, \beta_*) = (0.7, 2, 1)$, $\varepsilon_i^{[1]} \sim \mathcal{N}(0, 1)$, $X_i \sim \mathcal{N}(2, 3^2)$,
MO: $(p_*\alpha_*, \beta_*) = (0.7, 2, 1)$, $\varepsilon_i^{[1]} \sim \mathcal{N}(0, 2^2)$, $X_i \sim \mathcal{N}(2, 3^2)$,
SO: $(p_*, \alpha_*, \beta_*) = (0.7, 1, 0.5)$, $\varepsilon_i^{[1]} \sim \mathcal{N}(0, 2^2)$, $X_i \sim \mathcal{N}(1, 2^2)$,
MIX: $(p_*, \alpha_*, \beta_*) = (0.4, 2, 1)$, $\varepsilon_i^{[1]} \sim 0.5\mathcal{N}(-0.5, 0.5^2) + 0.5\mathcal{N}(0.5, 0.5^2)$, $X_i \sim \mathcal{N}(2, 3^2)$,

where for all the above models $\varepsilon_i^{[0]} \sim \mathcal{N}(0, 1)$. The characteristics of our models of interest are illustrated in Fig. 3 and the performances of our method for various choices of V's distribution, *i.e.*, $V \sim \mathcal{N}(0, \sigma_V^2)$ with $\sigma_V \in \{2\hat{\sigma}, \hat{\sigma}, \hat{\sigma}/2\}$ where $\hat{\sigma}$ denotes the empirical standard deviation of the Y_i, are summarized in Table 2. Let us precise that in Table 2, the empirical mean of our estimators is given within parenthesis and the standard deviation is given within brackets.
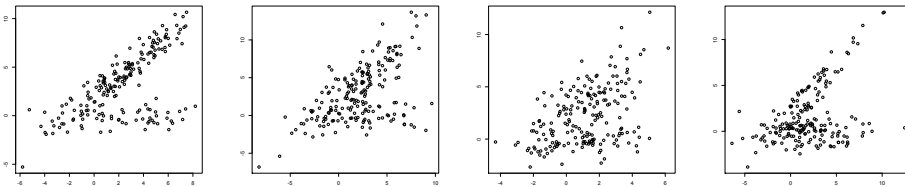

Figure 3: Plots of original data $(X_i, Y_i)_{1 \leq i \leq n}$ with $n = 200$, generated respectively from models WO, MO, SO and MIX.

15

Table 2: Mean and Std. Dev. of 200 estimates of (p_*, α_*, β_*).

Model	n	$2\hat{\sigma}$	$\hat{\sigma}$	$\hat{\sigma}/2$
WO	100	(0.700,1.995,1.006)	(0.696,2.001,1.007)	(0.704,1.990,1.003)
		[0.065,0.089,0.074]	[0.061,0.097,0.081]	[0.065,0.081,0.072]
WO	300	(0.701,1.998,1.005)	(0.696,1.995,0.999)	(0.704,1.996,0.994)
		[0.034,0.063,0.045]	[0.035,0.050,0.042]	[0.031,0.034,0.038]
MO	100	(0.718,1.962,0.971)	(0.732,1.975,0.956)	(0.726,1.980,0.957)
		[0.083,0.148,0.179]	[0.082,0.134,0.178]	[0.088,0.165, 0.171]
MO	300	(0.704,2.022,0.966)	(0.707,1.983,0.994)	(0.717,1.986,0.978)
		[0.050,0.080,0.116]	[0.045,0.074,0.099]	[0.056,0.072,0.114]
SO	100	(0.766,0.958,0.444)	(0.748,0.956,0.465)	(0.728,0.961,0.471)
		[0.104,0.200,0.136]	[0.099,0.219,0.128]	[0.098,0.211,0.134]
SO	300	(0.748,0.970,0.472)	(0.729,0.999,0.471)	(0.723,0.998,0.474)
		[0.079,0.136,0.084]	[0.069,0.156,0.083]	[0.080,0.190,0.111]
MIX	100	(0.415,1.984,0.983)	(0.408,1.964,1.016)	(0.409,1.988,1.011)
		[0.070,0.133,0.187]	[0.073,0.172,0.213]	[0.062,0.189,0.174]
MIX	300	(0.403,1.988,1.009)	(0.407,1.974,1.013)	(0.407,1.999,0.989)
		[0.043,0.115,0.133]	[0.039,0.158,0.138]	[0.037,0.075,0.094]

Comments on Table 2. First of all let us remark, as it is generally expected, that uniformly on the proposed choices for σ_V, the less the test model is overlapped the best our method performs. Secondly it is important to notice that the best performances of our method are achieved in weakly overlapped situations (models WO and MIX) by taking $\sigma_V = \hat{\sigma}/2$ (smallest choice). Contrarily, it also appears that the best performances for model MO, respectively model SO, are achieved when considering respectively $\sigma_V = \hat{\sigma}$ regarding to the estimation of (p_*, β_*) or $\sigma_V = 2\hat{\sigma}$ regarding to the estimation of α_*. These observations actually validate the following intuitive idea: when the populations labelled 0 and 1 are weakly overlapped (small variance of each population with respect to the variance of the design), then the θ-transformation, for θ values near from θ_*, places population 1 almost zero-symmetrically and very separately from population 0 (similarly to the situation illustrated in the last plot of Fig. 1). Hence by taking a small variance σ_V associated to the weight probability distribution Q, we force our method to be focused on symmetry detection, according to (12) and (13), on a domain of $\mathbb{R}$ containing the most relevant information. On the other hand, when populations 0 and 1 are moderately or strongly overlapped, the previous "favorable" situation is no longer true, which obliges us to increase σ_V in order to capture the symmetry information contained in the tails of the comparison-function (12).

4.4 Application: NimbleGen high-density array

In this section we propose to analyze a real dataset of size $n = 30000$ extracted from the ChipMix dataset (which true size is 176 343) considered in Martin-Magniette et al., 2008)Section 3.3, and modeled according to (1). We suggest to consider the estimated parameters provided by Martin-Magniette et al. (2008) as a starting point

16

for our method. Let us recall the value of these estimated parameters respectively obtained under the homoscedastic (Ho) and heteroscedastic (He) assumption:

$\hat{p}^{\texttt{Ho}} = 0.26, \hat{a}_0^{\texttt{Ho}} = 1.47, \hat{b}_0^{\texttt{Ho}} = 0.82, \hat{a}_1^{\texttt{Ho}} = -0.47, \hat{b}_1^{\texttt{Ho}} = 1.16, \hat{\sigma}_0^{\texttt{Ho}} = 0.42, \hat{\sigma}_1^{\texttt{Ho}} = 0.42,$

$\hat{p}^{\texttt{He}} = 0.32, \hat{a}_0^{\texttt{He}} = 1.48, \hat{b}_0^{\texttt{He}} = 0.81, \hat{a}_1^{\texttt{He}} = -0.29, \hat{b}_1^{\texttt{He}} = 1.13, \hat{\sigma}_0^{\texttt{He}} = 0.56, \hat{\sigma}_1^{\texttt{Ho}} = 0.80.$

Considering that $(\hat{a}_0^{\texttt{Ho}}, \hat{b}_0^{\texttt{Ho}}) \simeq (a_0, b_0)$ is a good candidate to recenter our dataset according to $Y = Y - (a_0 + b_0 X)$ (this fact is assessed in Fig. 4), our model of interest (2) is then obtained by considering $(\alpha, \beta) = (a1 - \hat{a}_0^M, b1 - \hat{b}_0^M)$ (so-called *centered version*). We consider in addition that the pdf f_0 is Gaussian with mean 0 and standard deviation $\hat{\sigma}_0 = 0.686$, this value corresponding to the empirical standard deviation of the Y coming from couples (X, Y) such that $X > 14$ (we consider that upon this threshold the size of the population coming from label 1 is negligible with respect to the size of the population coming from label 0). To estimate $\vartheta_* = (p_*, \alpha_*, \beta_*)$, we initialize our algorithm at $\hat{\vartheta}^M = (\hat{p}_0^M, \hat{a}_1^M - \hat{a}_0^M, \hat{b}_1^M - \hat{b}_0^M) = (0.26, -1.94, 0.34)$, which provides the following estimation:

$$\hat{\vartheta} = (\hat{p}, \hat{\alpha}, \hat{\beta}) = (0.31, -1.95, 0.34).$$

In the left side of Fig. 5 we plot our regularized estimation of the density f

$$\tilde{f}_n(x) := \frac{\hat{f}_n(x)\mathbf{I}_{\hat{f}_n(x) \geq 0}}{\int_{\mathbb{R}} f_n(x)\mathbf{I}_{\hat{f}_n(x) \geq 0} dx},$$

where $\hat{f}_n$ is defined in (22) and compare it to the estimation proposed by Martin-Magniette et al. (2008). It is interesting to observe that this two estimations are surprisingly close (except the small bump located between -4 and -2 which insinuate that our method has detected a 2-mixture structure for f). In the right side of Fig. 5 we suggest, in order to check the relevancy of the different methods, to reconstruct the function Ψ_{θ_*} defined in (9) by considering a kernel density estimate based on the $Y - (\hat{\alpha} + \hat{\beta}X)$ (free from mixture model structure), respectively the plug-in estimation obtained by our method and the method of Martin-Magniette et al. (2008), when we substitute (ϑ_*, f_0, f) in (9) respectively by $(\hat{\vartheta}, f_{\mathcal{N}(0, \hat{\sigma}_0)}, \tilde{f}_n)$ and $(\hat{\vartheta}^{\texttt{He}}, f_{\mathcal{N}(0, \hat{\sigma}_0^{\texttt{He}})}, f_{\mathcal{N}(0, \hat{\sigma}_1^{\texttt{He}})})$ (the best reconstruction is actually obtained under the heteroscedastic assumption), the integral term being estimated adequately by using (16). We observe in that case that the semiparametric approach allows a better fitting than the purely parametric approach, except between -3 and -1 where both methods have difficulties in fitting the shape of the target nonparametric estimator. This point should suggest that the true underlying model is in reality more complex than a simple 2-component mixture of regressions model. It would be for example interesting to investigate a generalization of model (1) where the design itself depends on the random label U, *i.e.*, $X := X^{[i]}$ if $U = i$, for $i = 0, 1$, allowing here the possibility to model $X^{[0]}$ by a mixture of Gaussian distributions and $X^{[1]}$ by a simple Gaussian distribution, as it is suggested by the two figures in the center and the right side of Graph. 4 (we can actually observe two small clusters along the first axis and a more spread one on top).
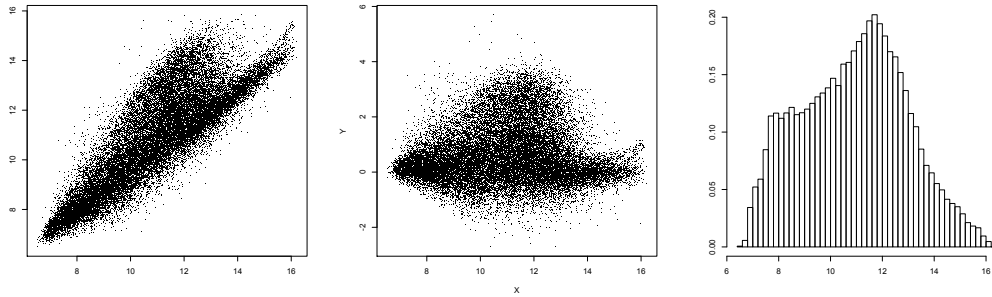
Figure 4: Respectively the plots of an extracted data set of size $n = 30000$ coming from the original NimbleGen high-density array dataset considered in Martin-Magniette et al. (2008), its centered version and the histogram of the design distribution.
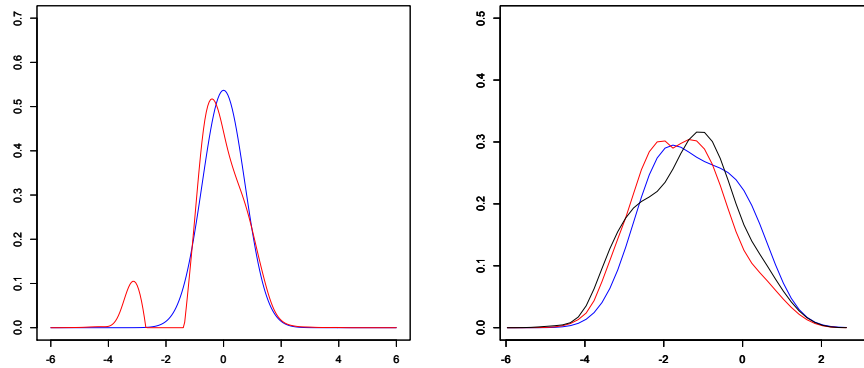

Figure 5: Left side: two estimators of f. In red, respectively in blue, the graph of our semiparametric estimator $\tilde{f}_n$ and the graph of the $f_{\mathcal{N}(0,\hat{\sigma}^{\mathrm{He}})}$ provided by Martin-Magniette et al. (2008). Right side: three estimators of Ψ_{θ_*}. In black a kernel density estimate of the $Y - (\hat{\alpha}_n + \hat{\beta}_n X)$'s pdf, in red, respectively in blue, the graph of the reconstruction of Ψ_{θ_*} obtained with our estimators and the estimators of Martin-Magniette et al. (2008).

5 Conclusion

We generalize in this paper the estimation method proposed in Bordes et al. (2006b) and Bordes and Vandekerkhove (2010), for semiparametric mixture models where one component is known, to mixture of regressions models with one known component, extending consequently, for the first time, the statistical study of semiparametric mixture models to the bivariate case. We prove in particular that our model

18

is semiparametrically identifiable under a set of conditions close to those suggested in Bordes et al. (2006b), and also propose an M-estimator based on the search of a "symmetry-point" when we apply a so-called θ-transformation to the original dataset. Regarding our asymptotic results, we prove that our M-estimator (21) respectively our functional estimator (22), converges towards ϑ_*, respectively the cumulative distribution F, at the rate $o_{a.s.}(n^{-1/4+\gamma})$ and $O_{a.s.}(n^{-1/6+\gamma})$, for all $\gamma > 0$. Let us emphasize that the conditions stated to prove these results are generally satisfied in the Gaussian framework, nevertheless we also prove that there exists, in very particular cases, a spurious solution to our M-estimation problem which can be avoided or rejected by applying some explicit procedures discussed in this paper. We conclude the presentation of our method by discussing, on very simple toy examples, its numerical sensitivity to factors such as: the sample size n, the weight distribution Q involved in (13), the existence of a spurious solution etc. Following the conclusions coming from this preliminary numerical study, we carry out a large sample study in order to highlight the influence of the weight distribution Q (sharpness and tails) on the performances of our method through different kind of situations, *i.e.*, from weakly overlapped to strongly overlapped simulated datasets. The main conclusion we learn from this study concerning the choice of the distribution Q can be summarized as follows: "in weakly overlapped situations let favour a distribution which probability mass is concentrated in a neighborhood of zero, when in strongly overlapped situations rather favour a heavily tailed distribution". Finally we conclude our study by considering a real data set extracted from the NimbleGen high-density array dataset previously analyzed by Martin-Magniette et al. (2008). It turns out from this study that the Euclidean estimator provided by our method is very close to the estimator found in Martin-Magniette et al. (2008), when our respective functional estimators of the pdf f differ slightly. In fact, we observe that our functional estimator places an extra bump on the left side of 0, which symptom brings up the possibility of having, in reality, a more complex link-structure between the ouptut Y and the design X than it is supposed in (1). This last remark indicates that it would be very challenging, in a futur work, to propose a semiparametric approach adapted to a more versatile class of mixture of regressions models, and also having the crucial ability to deal with very large datasets like the more and more popular *high-density arrays*.

6   Appendix

6.1   *Conditions (R) and (C) in the Gaussian Case*

In this section we discuss conditions (R) and (C) when the true underlying model is a contaminated Gaussian regression model with Gaussian design, *i.e.*, f, f_0, h are respectively the pdfs of the $\mathcal{N}(0,m)$, $\mathcal{N}(0,m_0)$, and $\mathcal{N}(E(X),\sigma_h^2)$ distributions.

Comments on condition (R). Conditions (R) i-iii) are standard and easy to verify in the above model. On the other hand, it is interesting to show how conditions (R) iv-v) arise naturally in this case.

Condition (R) iv). We show for simplicity that the first condition in (R) iv) (the same kind of proof works also for the second one) holds when $i = 0$, $\theta^* = (\alpha_*, \beta_*) \in \mathbb{R}^{+*2}$ and $m_0 = 1$. We write the decomposition

$$|F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)| =$$
$$|F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)|\mathbf{I}_{y>1-\theta_* \odot x, \ x<-\frac{\alpha_*}{\beta_*}}$$
$$+ |F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)|\mathbf{I}_{y>1+\theta_* \odot x, \ x\geq-\frac{\alpha_*}{\beta_*}}$$
$$+ |F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)|\mathbf{I}_{y<-1+\theta_* \odot x, \ x<-\frac{\alpha_*}{\beta_*}}$$
$$+ |F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)|\mathbf{I}_{y<-1-\theta_* \odot x, \ x\geq-\frac{\alpha_*}{\beta_*}}$$
$$+ |F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)|\mathbf{I}_{-1+\theta_* \odot x\leq y\leq 1-\theta_* \odot x, \ x<-\frac{\alpha_*}{\beta_*}}$$
$$+ |F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)|\mathbf{I}_{-1-\theta_* \odot x\leq y\leq 1+\theta_* \odot x, \ x\geq-\frac{\alpha_*}{\beta_*}}.$$

Consider the first term on the right hand side of the above decomposition (the three following terms being treated in entirely same way). For all $y > 1 - \theta_* \odot x$ with $x < -\alpha_*/\beta_*$ we have $y - \theta_* \odot x > y + \theta_* \odot x > 1$. Since for $t > 0$ large enough, the following inequality is valid

$$\frac{\exp(-t^2)}{\sqrt{2\pi}} \left(\frac{1}{t} - \frac{1}{t^3} \right) \leq 1 - F_0(t) \leq \frac{\exp(-t^2)}{t\sqrt{2\pi}},$$

we have in particular that for all $t > 1$, $0 \leq 1 - F_0(t) \leq \exp(-t^2)/\sqrt{2\pi}$. Hence it follows that for $y > 1 - \theta_* \odot x$ with $x < -\alpha_*/\beta_*$:

$$|F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)| \leq \frac{\exp(-(y + \theta_* \odot x)^2) + \exp(-(y - \theta_* \odot x)^2)}{\sqrt{2\pi}},$$

which proves that this first term is $h(x)dxdy$ integrable. Let us now sum the last two terms of the above decomposition and notice that

$$|F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)|$$
$$\times \left(\mathbf{I}_{-1+\theta_* \odot x\leq y\leq 1-\theta_* \odot x, \ x<-\frac{\alpha}{\beta_*}} + \mathbf{I}_{-1-\theta_* \odot x\leq y\leq 1+\theta_* \odot x, \ x\geq-\frac{\alpha}{\beta_*}} \right)$$
$$\leq \ 2\mathbf{I}_{-1-|\theta_* \odot x|\leq y\leq 1+|\theta_* \odot x|}.$$

We thus prove that this sum of terms is also $h(x)dxdy$ integrable.

Condition (R) v). We consider for simplicity the construction of the bounding function $\ell_{1,1}^0$ when $(\alpha, \beta) \in \Phi = [\underline{\alpha}, \overline{\alpha}] \times [\underline{\beta}, \overline{\beta}]$, with $(\underline{\alpha}, \underline{\beta}) \in \mathbb{R}^{+*2}$ and $m = m_0 = 1$. Notice first that for all $(x, y) \in \mathbb{R}^2$

$$|f_0^{(1)}(y + \theta \odot x)| \leq \frac{|y| + \overline{\alpha} + \overline{\beta}|x|}{\sqrt{2\pi}} \exp \left(-\frac{(y + \alpha + \beta x)^2}{2} \right). \tag{28}$$

Secondly, noticing that

$$\begin{cases} 0 \leq y + \underline{\beta}x \leq y + \alpha + \beta x & \text{if } x \geq 0, y \geq 0, \\ y + \underline{\alpha} + \overline{\beta}x \leq y + \alpha + \beta x \leq y + \overline{\alpha} + \overline{\beta}x & \text{if } x \geq 0, y \leq 0, \\ y + \underline{\alpha} + \overline{\beta}x \leq y + \alpha + \beta x \leq y + \overline{\alpha} + \underline{\beta}x & \text{if } x \leq 0, y \in \mathbb{R}, \end{cases}$$

20

we then obtain that for all $(x, y) \in \mathbb{R}^2$ and all $\theta \in \Phi$:

$$\exp\left(-\frac{(y + \alpha + \beta x)^2}{2}\right)$$

$$\leq \exp\left(-\frac{(y + \underline{\beta} x)^2}{2}\right) \mathbb{I}_{x \geq 0, y \geq 0}$$

$$+ \exp\left(-\frac{\min(|y + \underline{\alpha} + \underline{\beta} x|, |y + \overline{\alpha} + \overline{\beta} x|)^2}{2}\right) \mathbb{I}_{x \geq 0, y \leq 0}$$

$$+ \exp\left(-\frac{\min(|y + \underline{\alpha} + \overline{\beta} x|, |y + \overline{\alpha} + \underline{\beta} x|)^2}{2}\right) \mathbb{I}_{x \leq 0, y \in \mathbb{R}}$$

$$\leq B_\Phi(x, y),$$

where

$$B_\Phi(x, y) := \exp\left(-\frac{(y + \underline{\beta} x)^2}{2}\right) + \exp\left(-\frac{(y + \underline{\alpha} + \underline{\beta} x)^2}{2}\right)$$

$$+ \exp\left(-\frac{(y + \overline{\alpha} + \overline{\beta} x)^2}{2}\right) + \exp\left(-\frac{(y + \underline{\alpha} + \overline{\beta} x)^2}{2}\right)$$

$$+ \exp\left(-\frac{(y + \overline{\alpha} + \underline{\beta} x)^2}{2}\right).$$

Since the product $B_\Phi(x, y)h(x) \propto B_\Phi(x, y) \exp(-x^2/2)$ has a Gaussian tail, and thus admit any kind of moment, we can propose according to (28) and (R) v), $\ell_{1,1}^0(x, y) = |x|(|y| + \overline{\alpha} + \overline{\beta}|x|)/\sqrt{2\pi}B_\Phi(x, y) \exp(-x^2/2)$ as a candidate for the uniformly bounding function satisfying condition (R) v).

Comments on condition (C) ii). In the whole Gaussian case, *i.e.*, when $h(\cdot) = \mathcal{N}_{E(X),\sigma_h^2}(\cdot)$, $f_0(\cdot) = \mathcal{N}_{0,m_0}(\cdot)$ and $f = \mathcal{N}_{0,m}(\cdot)$, expression of (25) becomes:

$$p_* \sin((\alpha - \alpha_*) + (\beta - \beta_*)E(X))t) \exp\left(-\frac{t^2}{2}(\sigma_h^2(\beta - \beta_*)^2 + m)\right)$$

$$= (p_* - p) \sin((\alpha + \beta E(X))t) \exp\left(-\frac{t^2}{2}(\sigma_h^2\beta^2 + m_0)\right). \qquad (29)$$

When $p = p^*$ we already know that condition (C) i) is satisfied (see Remarks on condition (C) in Section 3.2). When $p \neq p_*$, we denote for convenience $\xi := \alpha + \beta E(X)$, $\xi_* := \alpha_* + \beta_* E(X)$, $\Sigma_{\beta - \beta_*} := \sigma_h^2(\beta - \beta_*)^2 + m$, and $\Sigma_\beta := \sigma_h^2\beta^2 + m_0$. Taking the first and third order derivative of (29) at point $t = 0$ we get the conditions

$$p_* \xi_* = p\xi, \quad \text{and} \quad (p - p_*)\xi(3\Sigma_{\beta - \beta_*} + \xi^2) + p_*(\xi - \xi_*)(3\Sigma_\beta + (\xi - \xi_*)^2) = 0. \quad (30)$$

Introducing the first relation in (30) into the second one, we obtain

$$\frac{p - p_*}{p}\xi_* p_* \left(3[\Sigma_\beta - \Sigma_{\beta - \beta_*}] + \frac{2p_*p - p^2}{p^2}\xi_*^2\right) = 0. \qquad (31)$$

21

Now we observe that, to ensure the validity of expression (29), the factors multiplied by the sin terms on both sides of (29) must be, at least, equivalent as $t \to \infty$. This last remark implies that $\Sigma_\beta = \Sigma_{\beta-\beta_*}$, or equivalently $\beta = \beta_*/2 + (m_0 - m)/2\beta_*\sigma_h^2$, and thus (31) leads to

$$p = 2p_*. \tag{32}$$

Using now the first relation in (29) and (32), we then obtain $\alpha = \alpha_*/2 + E(X)(m_0 - m)/2\beta_*\sigma_h^2$.

The consequences of the previous comments can be presented as follows:

Discussion (CG):

i) If $p_* > 1/2$ then the set of parameters $\vartheta \in \Theta$ satisfying condition (29) is always empty, since $p = 2p^* > 1$ is not an admissible solution.

ii) If $p_* \leq 1/2$ and if, for example, $E(X) = 0$ or $m_0 = m$, then $\vartheta_{**} = (2p^*, \alpha_*/2, \beta_*/2)$. In such a case it would be crucial to build a conveniently constrained parametric space (most of the time a plot of the dataset helps in building reasonable constraints on the intercept and slope parameter spaces) expecting that it contains ϑ_* but not ϑ_{**}.

iii) More generaly one can expect that when the shape of the sample data (see *e.g.* Fig. 1) suggest that $(m_0 - m)/2\beta_*\sigma_h^2$ is negligible with respect to α_* and β_*, which occurs when m_0 is close to m or/and $\sigma_h^2\beta_*$ is very large, then the solution proposed in ii) is loosely speaking still valid. Otherwise, if the situation is too much tricky to define visually a convenient parametric space and that d_n has several local minima (which should make suspect the existence of various spurious solutions) one may select the arg min of d_n that better recreate the density g over a compact set $\mathcal{C}$ large enough of $\mathbb{R}^2$. Namely if $\hat{g}_{n,\vartheta}$ is the ϑ-plugged version of (3) and $\hat{g}_n$ is a kernel density estimate of g, we select the ϑ, among the finite set of arg min values, which minimizes $\int_{\mathcal{C}} |\hat{g}_{n,\vartheta} - \hat{g}_n| d\lambda^{\otimes 2}$. Actually such a methodolgy is asymptotically consistent since model (3) is known to be identifiable over $\mathbb{R}^2$ under conditions established in Lemma 3.1.

6.2 *Explicit formula of $H(\cdot, \vartheta)$ and its derivatives*

In this section all the expressions are valid for all $(\vartheta, y) \in \Theta \times \mathbb{R}$, and the computation of the various derivative functions (under the integral sign) are all allowed according to the Lebesgue Differentiation Theorem and condition (R). According to (8) and

(10), we have

$$
\begin{aligned}
H(y,\vartheta) \;=\; & \frac{p^*}{p}\left(\int_{-\infty}^{y}\int_{\mathbb{R}} f(z+(\theta-\theta^*)\odot x)h(x)dxdz\right.\\
& \left.+\int_{-\infty}^{-y}\int_{\mathbb{R}} f(z+(\theta-\theta^*)\odot x)h(x)dxdz\right)\\
& +\frac{p-p^*}{p}\left(\int_{-\infty}^{y}\int_{\mathbb{R}} f_0(z+\theta\odot x)h(x)dxdz\right.\\
& \left.+\int_{-\infty}^{-y}\int_{\mathbb{R}} f_0(z+\theta\odot x)h(x)dxdz\right)-1.
\end{aligned}
$$

For simplicity we introduce

$$
\begin{aligned}
F^{\theta}(y) \;&=\; \int_{-\infty}^{y}\int_{\mathbb{R}} f(z+(\theta-\theta^*)\odot x)h(x)dxdz,\\
F_0^{\theta}(y) \;&=\; \int_{-\infty}^{y}\int_{\mathbb{R}} f_0(z+\theta\odot x)h(x)dxdz,
\end{aligned}
$$

which leads to

$$
\frac{\partial}{\partial p}H(y,\vartheta) = -\frac{p^*}{p^2}\left[\left(F^{\theta}(y)+F^{\theta}(-y)\right)-\left(F_0^{\theta}(y)+F_0\theta(-y)\right)\right].
$$

Let us denote

$$
\begin{aligned}
\dot{F}^{\alpha}(y) \;&:=\; \frac{\partial}{\partial\alpha}F^{\theta}(y)=\int_{-\infty}^{y}\int_{\mathbb{R}}\dot{f}(z+(\theta-\theta_*)\odot x)h(x)dxdz,\\
\dot{F}_0^{\alpha}(y) \;&:=\; \frac{\partial}{\partial\alpha}F_0^{\theta}(y)=\int_{-\infty}^{y}\int_{\mathbb{R}}\dot{f}_0(z+\theta\odot x)h(x)dxdz,
\end{aligned}
$$

and for

$$
\begin{aligned}
\dot{F}^{\beta}(y) \;&:=\; \frac{\partial}{\partial\beta}F\theta(y)=\int_{-\infty}^{y}\int_{\mathbb{R}}x\dot{f}(z+(\theta-\theta^*)\odot x)h(x)dxdz,\\
\dot{F}_0^{\beta}(y) \;&:=\; \frac{\partial}{\partial\beta}F_0\theta(y)=\int_{-\infty}^{y}\int_{\mathbb{R}}x\dot{f}_0(z+\theta\odot x)h(x)dxdz,
\end{aligned}
$$

we obtain

$$
\frac{\partial}{\partial\alpha}H(y,\vartheta) = \frac{p^*}{p}\left(\dot{F}^{\alpha}(y)+\dot{F}^{\alpha}(-y)\right)+\frac{p-p^*}{p}\left(\dot{F}_0^{\alpha}(y)+\dot{F}_0^{\alpha}(-y)\right).
$$

$$
\frac{\partial}{\partial\beta}H(y,\vartheta) = \frac{p^*}{p}\left(\dot{F}^{\beta}(y)+\dot{F}^{\beta}(-y)\right)+\frac{p-p^*}{p}\left(\dot{F}_0^{\beta}(y)+\dot{F}_0^{\beta}(-y)\right).
$$

At point $\vartheta=\vartheta_*$ the Hessian matrix of $H(\cdot,\vartheta)$ defined in (36) is obtained by considering

$$
\dot{H}(y,\vartheta_*) = \begin{pmatrix} \frac{1}{p}\left(F_0^{\theta_*}(y)+F_0^{\theta_*}(-y)-1\right)\\ 2f(y)\\ 2f(y)E(X) \end{pmatrix}. \tag{33}
$$

23

Let us denote now

$$\ddot{F}^{\beta,\alpha}(y) = \frac{\partial^2}{\partial\beta\partial\alpha}F^\theta(y) = \int_{-\infty}^y \int_{\mathbb{R}} x\ddot{f}(z + (\theta - \theta_*) \odot x)h(x)dxdz,$$

$$\ddot{F}_0^{\beta,\alpha}(y) = \frac{\partial^2}{\partial\beta\partial\alpha}F_0^\theta(y) = \int_{-\infty}^y \int_{\mathbb{R}} x\ddot{f}_0(z + \theta \odot x)h(x)dxdz,$$

$$\ddot{F}^{\alpha,\alpha}(y) = \frac{\partial^2}{\partial\alpha^2}F^\theta(y) = \int_{-\infty}^y \int_{\mathbb{R}} \ddot{f}(z + (\theta - \theta_*) \odot x)h(x)dxdz,$$

$$\ddot{F}_0^{\beta,\beta}(y) = \frac{\partial^2}{\partial\beta^2}F_0^\theta(y) = \int_{-\infty}^y \int_{\mathbb{R}} x^2\ddot{f}_0(z + \theta \odot x)h(x)dxdz.$$

We then obtain

$$\frac{\partial^2}{\partial p^2}H(y,\vartheta) = \frac{p^*}{p^3}\left[(F^\theta(y) + F^\theta(-y)) - (F_0^\theta(y) + F_0\theta(-y))\right],$$

$$\frac{\partial^2}{\partial u \partial p}H(y,\vartheta) = -\frac{p^*}{p^2}\left[(\dot{F}^u(y) + \dot{F}^u(-y)) - (\dot{F}_0^u(y) + \dot{F}_0^u(-y))\right], \quad u = \alpha,\ \beta,$$

$$\frac{\partial^2}{\partial u \partial v}H(y,\vartheta) = \frac{p^*}{p}\left(\ddot{F}^{u,v}(y) + \ddot{F}^{u,v}(-y)\right)$$

$$+ \frac{p - p^*}{p}\left(\ddot{F}_0^{u,v}(y) + \ddot{F}_0^{u,v}(-y)\right), \quad u = \alpha,\ \beta, \quad v = \alpha,\ \beta.$$

6.3 Boundedness

Boundedness of $\Psi_\theta(\cdot)$ and $\Psi_\theta'(\cdot)$. If f and f_0 are supposed to be bounded over $\mathbb{R}$ then we clearly have from (8) that

$$|\Psi_\theta(y)| \leq \|f\|_\infty + \|f_0\|_\infty, \qquad (\theta, y) \in \Phi \times \mathbb{R}.$$

The same kind of argument holds to prove boundedness of $\dot{f}_\theta(\cdot)$ when (R) ii) is supposed.

Boundedness of $H(\cdot, \vartheta)$. Since for all $\theta \in \Phi$ the functions $F_\theta(\cdot)$ and $J_\theta(\cdot)$ are both cdfs, we thus have, since $\delta \leq p \leq 1 - \delta$, from expression (12):

$$H(y,\vartheta) \leq \frac{4}{\delta} + 1, \qquad (y, \vartheta) \in \mathbb{R} \times \Phi. \tag{34}$$

6.4 Integrable Lipschitz property of $\Psi_\theta(\cdot)$

From (8), for all $(y, \theta, \theta') \in \mathbb{R} \times \Phi^2$ we have

$$|\Psi_\theta(y) - f\Psi_{\theta'}(y)| \leq p_* \int_{\mathbb{R}} |f(y + (\theta - \theta_*) \odot x) - f(y + (\theta' - \theta_*) \odot x)|h(x)dx$$

$$+ (1 - p_*) \int_{\mathbb{R}} |f_0(y + \theta \odot x) - f_0(y + \theta' \odot x)|h(x)dx. \tag{35}$$

24

Consider for simplicity the first integral term on the right hand side of (35) (the same argument holding for the second term). According to the Mean Value Theorem there exists, for all $(x, y) \in \mathbb{R}^2$ and $(\theta, \theta') \in \Phi^2$, a value $\gamma := \gamma(x, y, \theta, \theta')$ belonging to the line segment with extremities $y + (\theta - \theta_*) \odot x$ and $y + (\theta' - \theta_*) \odot x$, or equivalently a $\bar{\theta} := \bar{\theta}(x, y, \theta, \theta')$ belonging to the line segment with extremities θ and θ' such that $\gamma = y + (\bar{\theta} - \theta_*) \odot x$ and

$$
\begin{aligned}
|f(y &+ (\theta - \theta_*) \odot x) - f(y + (\theta' - \theta_*) \odot x)| \\
&= |\dot{f}(\gamma)(\alpha - \alpha' + (\beta - \beta')x)| \\
&= |\dot{f}(y + (\bar{\theta} - \theta_*) \odot x)(\alpha - \alpha' + (\beta - \beta')x)| \\
&\leq \sup_{\theta \in \Phi} |\dot{f}(y + (\theta - \theta_*) \odot x)|(|\alpha - \alpha'| + |\beta - \beta'||x|).
\end{aligned}
$$

From condition (R) ii) there thus exists a nonnegative constant c such that

$$
\begin{aligned}
\int_{\mathbb{R}} &|\Psi_\theta(y) - \Psi_{\theta'}(y)|dy \\
&\leq \sum_{j=0,1} (|\alpha - \alpha'|^{1-j} + |\beta - \beta'|^j) \\
&\int_{\mathbb{R} \times \mathbb{R}} |x|^j (\sup_{\theta \in \Phi} |\dot{f}(z + (\theta - \theta_*) \odot x)| + \sup_{\theta \in \Phi} |\dot{f}_0(z + \theta \odot x)|)h(x)dxdz \\
&\leq c\|\theta - \theta'\|_2.
\end{aligned}
$$

6.5   *Uniform Lipschitz property of* $H(\cdot, \vartheta)$

Let us write

$$
\begin{aligned}
H(y, \vartheta) &- H(y, \vartheta') \\
= \frac{1}{p} &(F_\theta(y) - F_{\theta'}(y) + F_\theta(-y) - F_{\theta'}(-y)) + \frac{p - p'}{pp'}(F_{\theta'}(y) + F_{\theta'}(y)) \\
&+ \frac{1 - p}{p} (J_\theta(y) - J_{\theta'}(y) + J_\theta(-y) - J_{\theta'}(-y)) + \frac{p - p'}{pp'}(J_{\theta'}(y) + J_{\theta'}(-y)).
\end{aligned}
$$

To prove the uniform Lipschitz property of $H(\cdot, \vartheta)$ we need to prove it for $J_\theta(\cdot)$ and $F_\theta(\cdot)$. We begin with the simplest term $J_\theta(\cdot)$. According again to the mean value theorem, for all $y \in \mathbb{R}$, all $(x, z) \in \mathbb{R}^2$ with $z \leq y$, and all $(\theta, \theta') \in \Phi^2$ there exists $\bar{\theta} := \bar{\theta}(x, z, \theta, \theta')$ belonging to the line segment with extremities θ and θ' such that

$$
\begin{aligned}
|J_\theta(y) - J_{\theta'}(y)| &\leq \int_{-\infty}^{y} \int_{\mathbb{R}} |f_0(z + \theta \odot x) - f_0(z + (\theta' \odot x)|h(x)dxdz \\
&= \int_{-\infty}^{y} \int_{\mathbb{R}} |\dot{f}_0(z + \bar{\theta} \odot x)(|\alpha - \alpha'| + |\theta - \theta'||x|)h(x)dxdz, \\
&\leq \int_{-\infty}^{y} \int_{\mathbb{R}} \sup_{\theta \in \Phi} |\dot{f}_0(z + \theta \odot x)h(x)dxdz|\alpha - \alpha'| \\
&\quad + \int_{-\infty}^{y} \int_{\mathbb{R}} |x| \sup_{\theta \in \Phi} |\dot{f}_0(z + ux)h(x)dxdz|\beta - \beta'| \\
&\leq c\|\theta - \theta'\|_2,
\end{aligned}
$$

25

where c denotes a nonnegative constant arising from condition (R) ii). Using the same kind of argument we prove that there exists a nonnegative contant c' such that for all $(y, \theta, \theta') \in \mathbb{R} \times \Phi^2$

$$|F_\theta(y) - F_{\theta'}(y)| \leq c'\|\theta - \theta'\|_2.$$

In conclusion, for all $y \in \mathbb{R}$, there exists a nonnegative constant c'' such that for all $(y, \vartheta, \vartheta') \in \mathbb{R} \times \Theta^2$

$$|H(y, \vartheta) - H(y, \vartheta')| \leq \frac{2}{\delta}(c + c')\|\theta - \theta'\|_2 + 4\frac{|p - p'|}{\delta^2} \leq c''\|\vartheta - \vartheta\|_3.$$

6.6 Proof of Lemma 3.2

i) From boundedness and the uniform Lipschitz property of $H(\cdot, \vartheta)$, along with the integrability and the integrable Lipschitz property of $f_\theta(\cdot)$ proved in Sections 6.3, 6.4 and 6.5, there exists a nonnegative constant c such that for all $(\vartheta, \vartheta') \in \Theta^2$

$$\begin{aligned}
&\left| \int_\mathbb{R} H^2(y, \vartheta)dQ(y) - \int_\mathbb{R} H^2(y, \vartheta')dQ(y) \right| \\
&\leq \int_\mathbb{R} \left| H(y, \vartheta) + H(y, \vartheta') \right| \left| H(y, \vartheta) - H(y, \vartheta') \right| q(y)dy \\
&\leq c\|\vartheta - \vartheta'\|_3,
\end{aligned}$$

which concludes the proof of i).

ii) To clarify the similarity between the semiparametric contamination model (5) studied in Bordes et al. (2006b) and the contaminated regression model (3), we can say that $f_\theta(\cdot)$ plays the role of $g(\cdot - \mu)$ and that $I_\theta(\cdot)$ plays the role of $f_0(\cdot - \mu)$.

If $\vartheta = \vartheta_*$ then $d(\vartheta) = 0$. To prove the converse we notice that $d(\vartheta) = 0$, which implies, since $H_1(\cdot, \vartheta)$ and $H_2(\cdot, \vartheta)$ are continuous and $q > 0$ over $\mathbb{R}$, that $H_1(\cdot; \vartheta) = H_2(\cdot; \vartheta)$ which leads, for almost all $y \in \mathbb{R}$, to

$$\Psi_\theta(y) - (1 - p)I_\theta(y) = \Psi_\theta(-y) - (1 - p)I_\theta(-y).$$

Using formula (8), we obtain

$$p_* \int_\mathbb{R} f(y + (\theta - \theta_*) \odot x)h(x)dx + (p - p_*)I_\theta(y)$$
$$= p_0 \int_\mathbb{R} f(-y + (\theta - \theta_*) \odot x)h(x)dx + (p - p_*)I_\theta(-y), \qquad y \in \mathbb{R}.$$

Considering the Fourier transform of the previous equality, using Fubini's Theorem, and noticing that $\bar{f}$ and $\bar{f}_0$ are real-valued functions, it follows that

$$p_* e^{-it(\alpha - \alpha_*)} \bar{f}(t)\breve{\bar{h}}((\beta - \beta_*)t) + (p - p_*)e^{-it\alpha} \bar{f}_0(t)\breve{\bar{h}}(\beta t)$$
$$= p_* e^{it(\alpha - \alpha_*)} \bar{f}(t)\bar{h}((\beta - \beta_*)t) + (p - p_*)e^{it\alpha} \bar{f}_0(t)\bar{h}(\beta t), \qquad t \in \mathbb{R}.$$

Using the notation introduced for the writing of condition (C), the previous equation becomes (25).

Suppose that $p = p_*$ and take the first and third order derivative of (25) at point $t = 0$. We then obtain $\alpha - \alpha_* + (\beta - \beta_*)E(X) = 0$ and

$$3m[\alpha - \alpha_* + (\beta - \beta_*)E(X)] + (\alpha - \alpha_*)^3 + 3(\alpha - \alpha_*)^2(\beta - \beta_*)E(X)$$
$$+ 3(\alpha - \alpha_*)(\beta - \beta_*)^2 E(X^2) + (\beta - \beta_*)^3 E(X^3) = 0,$$

which leads to

$$(\beta - \beta_*)(4E(X)^3 + 3E(X)E(X^2) + E(X^3)) = 0,$$

and thus implies that $\theta = \theta_*$ if $4E(X)^3 + 3E(X)E(X^2) + E(X^3) \neq 0$.

Suppose now that $p \neq p^*$, then condition (C) ii) requires that $\vartheta = \vartheta_*$.

iii) First we have

$$\ddot{d}(\vartheta_*) = 2\int_{\mathbb{R}} \left(\ddot{H}(y, \vartheta_*)H(y, \vartheta_*) + \dot{H}(y, \vartheta_*)\dot{H}^T(y, \vartheta_*) \right) q(y)dy$$
$$= 2\int_{\mathbb{R}} \dot{H}(y, \vartheta_*)\dot{H}^T(y, \vartheta_*)q(y)dy, \tag{36}$$

according to (9) and the fact that $H(\cdot, \vartheta_*) = 0$ on $\mathbb{R}$. Let v be a vector in $\mathbb{R}^3$. We have

$$v^T \ddot{d}(\vartheta_*)v = 2\int_{\mathbb{R}} \left(v^T \dot{H}(y, \vartheta_*) \right)^2 q(y)dy \geq 0.$$

It follows that $\ddot{d}(\vartheta_*)$ is a positive 3×3 real valued matrix. Let us show that it is also definite. If $v \in \mathbb{R}^3$ is a non-null column vector such that $v^T \ddot{d}(\theta_*)v = 0$, then $v^T \dot{H}(y, \vartheta_*) = 0$ for almost all $y \in \mathbb{R}$. According to (33) in the appendix, we have to discuss the proportionality of f and $F_0^{\theta_*}(\cdot) + F_0^{\theta_*}(-\cdot) - 1$. Because f_0 is an even density, we have from Fubini's theorem

$$f(y) = \frac{\int_{\mathbb{R}}[F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)]h(x)dx}{\int_{\mathbb{R}^2}[F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)]h(x)dxdy}. \tag{37}$$

Using integration by parts and assumption (R) iv), the denominator of the right hand side of (37) can be expressed as follows

$$\int_{\mathbb{R}^2} [F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)]h(x)dxdy$$
$$= \int_{\mathbb{R}} \left\{ [y(F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)]_{-\infty}^{\infty} \right\} h(x)dx$$
$$- \int_{\mathbb{R}} \int_{\mathbb{R}} y(f_0(y + \theta_* \odot x) - f_0(y - \theta_* \odot x))dy h(x)dx$$
$$= 2\int_{\mathbb{R}} (\alpha_* + \beta_* x)h(x)dx = 2(\alpha_* + \beta_* E(X)).$$

27

If we calculate now the second-order moment of f we obtain

$$m \quad := \quad \int_{\mathbb{R}} y^2 f(y) dy = \frac{\int_{\mathbb{R}} x^2 [F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)] h(x) dx}{2(\alpha_* + \beta_* E(X))}.$$

Using integration by parts and assumption (R) iv), the numerator of the right hand-side of (37) can be expressed as follows

$$\int_{\mathbb{R}^2} y^2 [F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)] h(x) dx dy$$

$$= \int_{\mathbb{R}} \left\{ \left[\frac{y^3}{3} (F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)) \right]_{-\infty}^{\infty} \right\} h(x) dx$$

$$- \int_{\mathbb{R}^2} \frac{y^3}{3} (f_0(y + \theta_* \odot x) - f_0(y - \theta_* \odot x)) dy h(x) dx$$

$$= 2 \int_{\mathbb{R}^2} \frac{3u^2 \theta \odot x + (\theta \odot x)^3}{3} f_0(u) du h(x) dx$$

$$= 2m_0(\alpha_* + \beta_* E(X)) + \frac{2}{3}(\alpha_*^3 + 3\alpha_*^2 \beta_* E(X) + 3\alpha_* \beta_*^2 E(X^2) + \beta_*^3 E(X^3)),$$

which leads to a contradiction if (C) iii) is assumed.

iv) Let us consider

$$|d_n(\vartheta) - d(\vartheta)| \le T_{1,n}(\vartheta) + T_{2,n}(\vartheta),$$

where

$$T_{1,n}(\vartheta) \quad := \quad \left| \frac{1}{n} \sum_{i=1}^{n} H^2(V_i; \vartheta, \tilde{F}_{n,\theta}, \hat{J}_{n,\theta}) - H^2(V_i; \vartheta, F_\theta, J_\theta) \right|,$$

$$T_{2,n}(\vartheta) \quad := \quad \left| \frac{1}{n} \sum_{i=1}^{n} H^2(V_i; \vartheta, F_\theta, J_\theta) - E\left(H^2(V_1; \vartheta, F_\theta, J_\theta)\right) \right|.$$

Uniform almost sure rate of convergence of $T_{1,n}$. Note first that from boundedness of $H(\cdot; \vartheta, \tilde{F}_{n,\theta}, \hat{J}_{n,\theta})$ and $H(\cdot; \vartheta, F_\theta, J_\theta)$ given by (34), there exist nonnegative constants C and C' such that

$$\begin{aligned} T_{1,n}(\vartheta) \quad &= \quad \left| \frac{1}{n} \sum_{i=1}^{n} (H(V_i; \vartheta, \tilde{F}_{n,\theta}, \hat{J}_{n,\theta}) + H(V_i; \vartheta, F_\theta, J_\theta)) \right. \\ &\qquad \times (H(V_i; \vartheta, \tilde{F}_{n,\theta}, \hat{J}_{n,\theta}) - H(V_i; \vartheta, F_\theta, J_\theta)) \Big| \\ &\le \quad C \sup_{y \in \mathbb{R}} |H(y; \vartheta, \tilde{F}_{n,\theta}, \hat{J}_{n,\theta}) - H(y; \vartheta, F_\theta, J_\theta))| \\ &\le \quad C' \left(\sup_{y \in \mathbb{R}} |\hat{J}_{n,\theta}(y) - J_\theta(y)| + \sup_{y \in \mathbb{R}} |\tilde{F}_{n,\theta}(y) - F_\theta(y)| \right). \end{aligned}$$

Let us now denote

$$
\begin{aligned}
T_{1,n}^{(1)} &:= \sup_{\theta \in \Phi} \sup_{y \in \mathbb{R}} |\hat{J}_{n,\theta}(y) - J_\theta(y)|, \\
T_{1,n}^{(2)} &:= \sup_{\theta \in \Phi} \sup_{y \in \mathbb{R}} |\tilde{F}_{n,\theta}(y) - F_\theta(y)|.
\end{aligned}
$$

Convergence rate of $T_{1,n}^{(1)}$. For simplicity we will suppose that $\mathrm{proj}_2(\Phi) \subset [0, A]$, where A is a nonnegative real number and for all $(x, y) \in \mathbb{R}^2$, $\mathrm{proj}_2 : (x, y) \mapsto y$. Let us introduce $P_n^X = n^{-1} \sum_{i=1}^n \delta_{X_i}$ the empirical measure associated to the iid sample $(X_1, \ldots, X_n)$ with common probability distribution P^X with pdf and cdf respectively denoted by h and H). We use the functional notation $Pf = \int f dP$. Notice now that, according to expression (17), we have for all $y \in \mathbb{R}$:,

$$
\begin{aligned}
\hat{J}_{n,\theta}(y) - J_\theta(y) &= \frac{1}{n} \sum_{i=1}^n F_0(y + \alpha + \beta X_i) - E(F_0(y + \alpha + \beta X)) \\
&= (P_n^X - P^X) F_0(y + \alpha + \beta \cdot).
\end{aligned}
$$

Let us consider the class of functions

$$
\mathcal{F}_0 = \{x \mapsto F_0(u + \beta x); \quad u \in \mathbb{R}, \quad \beta \in [0, A]\}.
$$

Since

$$
(P_n^X - P^X) F_0(u + \beta \cdot) = (P_n^X - P^X) F_0(y + \beta(\cdot \vee 0)) + (P_n^X - P^X) F_0(y + \beta(\cdot \wedge 0)),
$$

it is enough to study the empirical process indexed by the classes of functions

$$
\begin{aligned}
\mathcal{F}_0^+ &= \{x \mapsto F_0(u + \beta(x \vee 0)); \quad u \in \mathbb{R}, \quad \beta \in [0, A]\}, \\
\mathcal{F}_0^- &= \{x \mapsto F_0(u + \beta(x \wedge 0)); \quad u \in \mathbb{R}, \quad \beta \in [0, A]\}.
\end{aligned}
$$

For simplicity we denote $\Gamma_{y,\alpha}(x) = F_0(y + \alpha(x \vee 0))$ and only consider the class $\mathcal{F}_0^+$, the class $\mathcal{F}_0^-$ being treated in a entirely same way. Since F_0 is a cdf, for $\beta_1 \le \beta \le \beta_2$ and $u_1 \le u \le u_2$ we have

$$
\Gamma_{u_1,\beta_1}(x) \le \Gamma_{u,\beta}(x) \le \Gamma_{u_2,\beta_2}(x), \qquad x \in \mathbb{R},
$$

and, since F_0 is supposed to be Lipschitz,

$$
0 \le \Gamma_{u_2,\beta_2}(x) - \Gamma_{u_1,\beta_1}(x) \le C(u_2 - u_1 + (\beta_2 - \beta_1)(x \vee 0)).
$$

Let us consider now $\varepsilon > 0$, and $(\overline{u_\varepsilon}, \underline{u_\varepsilon}) \in \mathbb{R}^2$ such that

$$
F_0(\overline{u_\varepsilon}) \ge 1 - \varepsilon, \qquad \text{and} \qquad F_0(\underline{u_\varepsilon}) \le \varepsilon.
$$

Note that $\overline{u_\varepsilon}$ and $\underline{u_\varepsilon}$ do not depend on β. For all $N \in \mathbb{N}$, define

$$
\underline{u_\varepsilon} = u_{1,\varepsilon} \le u_{2,\varepsilon} \le \ldots \le u_{N,\varepsilon} = \overline{u_\varepsilon},
$$

29

and consider $N(\varepsilon)$ the smallest integer such that $u_{i,\varepsilon} - u_{i-1,\varepsilon} \leq \varepsilon$ for $i = 2, \ldots, N(\varepsilon)$. We denote by $\lceil \cdot \rceil$ the integer part function. For all ε small enough we clearly have

$$N(\varepsilon) \leq \left\lceil \frac{\overline{u_\varepsilon} - \underline{u_\varepsilon}}{\varepsilon} \right\rceil \leq 2\frac{\overline{u_\varepsilon} - \underline{u_\varepsilon}}{\varepsilon}.$$

Let us now define $\alpha_{i,\varepsilon} = \varepsilon(i-1)$, $i = 1, \ldots, M(\varepsilon)$, where $M(\varepsilon) = \lceil \lceil A+1 \rceil / \varepsilon \rceil$ and thus $\alpha_{M(\varepsilon),\varepsilon} > A$. Observe in addition that

$$\begin{aligned}
\|\Gamma_{u_{i+1,\varepsilon},\beta_{j+1,\varepsilon}} - \Gamma_{u_{i,\varepsilon},\beta_{j,\varepsilon}}\|_{2,P^X}^2 &= c^2 E\left((u_{i+1,\varepsilon} - u_{i,\varepsilon} + (\beta_{j+1,\varepsilon} - \beta_{j,\varepsilon})(X_1 \wedge 0))^2\right) \\
&\leq 2c^2\varepsilon^2 + 2C^2\varepsilon E(X_1^2) \\
&= 2c^2\varepsilon^2\left(1 + E(X_1^2)\right).
\end{aligned}$$

Hence the expression

$$[\Gamma_{u_{i+1,\varepsilon},\beta_{j+1,\varepsilon}} - \Gamma_{u_{i,\varepsilon},\beta_{j,\varepsilon}}], \quad 1 \leq i \leq N(\varepsilon), \quad 1 \leq j \leq M(\varepsilon),$$

is a $\left(c\sqrt{2(1+E(X_1^2))}\right)$-covering of $\mathcal{F}_0^+$ in the $L_2(P^X)$-norm sense. Using the standard notation $N_{[]}(\cdot)$ (see van der Vaart and Wellner, 1996) the covering number of the class $\mathcal{F}_0^+$ is bounded as follows

$$N_{[]}\left(\varepsilon, \mathcal{F}_0^+, L_2(P^X)\right) \leq cN(\varepsilon)M(\varepsilon) \leq c'\frac{\overline{u_\varepsilon} - \underline{u_\varepsilon}}{\varepsilon^2}.$$

Thus if there exist constants C and V such that

$$|\overline{u_\varepsilon}| \wedge |\underline{u_\varepsilon}| \leq C/\varepsilon^V, \tag{38}$$

we get $N(\varepsilon)M(\varepsilon) \leq C/\varepsilon^{V+2}$ which allows us to use Theorem 2.14.9, p. 246 in van der Vaart and Wellner (1996) since their Condition (2.14.7), p. 245 is then satisfied after replacing their constant V by $V+2$. Let us discuss condition (38). For ε small enough this condition is true if $\overline{y_\varepsilon} \leq C/\varepsilon^V$ and $\underline{u_\varepsilon} \geq -C/\varepsilon^V$. Denoting by $F_0^\leftarrow$ the quantile function of F_0, condition (38) becomes

$$F_0^\leftarrow(1-\varepsilon) \leq C/\varepsilon^V \qquad \text{and} \qquad F_0^\leftarrow(\varepsilon) \geq -C/\varepsilon^V. \tag{39}$$

We consider for simplicity the first condition in (39) (the second one being treated in the same way); it is equivalent to $F_0(C/\varepsilon^V) \geq 1 - \varepsilon$, and taking $t = C/\varepsilon^V$ this condition turns into

$$F_0(t) \geq 1 - C/t^{1/V}.$$

Thus it suffices to have

$$\liminf_{t\to\infty} \frac{-\log(1 - F_0(t))}{\log(t)} > 0.$$

Finally, using the symmetry of f_0, condition (38) holds if

$$\liminf_{t\to\infty} \frac{-2\log F_0(-t)}{\log(t)} > 0, \tag{40}$$

which is ensured by condition (R) vi). In conclusion if (40) is satisfied and $E(X_1^2) < \infty$ then, according to Theorem 2.14.16, p. 248 in van der Vaart and Wellner (1996), we obtain

$$\sup_{\theta \in \Phi} \| \hat{J}_{n,\theta} - J_\theta \|_\infty \leq \| P_n^X - P^X \|_{\mathcal{F}_0} = o_{a.s}(n^{-1/2+\gamma}), \quad \gamma > 0.$$

Convergence rate of $T_{1,n}^{(2)}$. Recall that F_θ is the cdf of $Y_i - \theta \odot X_i$, i.e.,

$$F_\theta(y) = P(Y_i - \theta \odot X_i \leq y),$$

and

$$F_{n,\theta}(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}_{Y_i - \theta \odot X_i \leq y}.$$

Let K be a kernel satifying (K). The K-regularized versions of F_θ and $F_{n,\theta}$ are

$$\tilde{F}_\theta = K * F_\theta, \qquad \tilde{F}_{n,\theta} = K * F_{n,\theta}.$$

Let us denote by $P_n^{X,Y}$ the empirical measure

$$P_n^{X,Y} = \frac{1}{n} \sum_{i=1}^n \delta_{X_i, Y_i},$$

and by $P^{X,Y}$ the law of (X_1, Y_1).

The set of functions for which $(x, y) \mapsto ax + by + c$ being a 3-dimensionnal vector space, Corollary 2.5 in Kuelbs and Dudley (1980) shows that the class of sets

$$\mathcal{C} = \left\{ \{(u, v) \in \mathbb{R}^2 : au + bv + c < 0 \}; \quad (a, b, c) \in \mathbb{R}^3 \right\},$$

is a Strassen log-log class, which implies that *a.s.*

$$\limsup_{n \to \infty} \sup_{\mathcal{C}} \sqrt{\frac{n}{2 \log \log(n)}} (P_n^Z - P^Z)(C) = \sup_{C \in \mathcal{C}} \sqrt{P^Z(C)(1 - P^Z(C))} \leq 1/2.$$

Since $\mathcal{C}$ contains the class

$$\mathcal{S} := \left\{ \{(u, v) \in \mathbb{R}^2 : v - (\alpha + \beta u) < y \}; \quad (\alpha, \beta; y) \in \Phi \times \mathbb{R} \right\},$$

it follows that, for all set $S \in \mathcal{S}$, $P^Z(S) = \int_S dP^{X,Y}(u,v) = P(Y - (\alpha + \beta X) < y) = F_\theta(y)$ and for the same reason $P_n^Z(S) = F_{n,\theta}(y)$, we have

$$\limsup_{n \to \infty} \sup_{(\theta, y) \in \Phi \times \mathbb{R}} \sqrt{\frac{n}{\log \log(n)}} (F_{n,\theta} - F_\theta)(y) \leq 1/2 \qquad a.s. \tag{41}$$

Now if we replace $F_{n,\theta}$ by its regularized version $\tilde{F}_{n,\theta}$ the approximation is controlled as follows,

$$
\begin{aligned}
\tilde{F}_{n,\theta}(y) &- F_{n,\theta}(y) \\
&= \tilde{F}_{n,\theta}(y) - E(\tilde{F}_{n,\theta}(y)) + E(\tilde{F}_{n,\theta}(y)) - F_\theta(y) + F_\theta(y) - F_{n,\theta}(y) \\
&= \tilde{F}_{n,\theta}(y) - E(\tilde{F}_{n,\theta}(y)) - [F_{n,\theta}(y) - E(F_{n,\theta}(y))] \\
&\quad + E(\tilde{F}_{n,\theta}(y)) - F_\theta(y),
\end{aligned} \tag{42}
$$

recalling that $E(F_{n,\theta}(y)) = F_\theta(y)$. The first term on the right hand side of (42) satisfies

$$
\begin{aligned}
\tilde{F}_{n,\theta}(y) - E(\tilde{F}_{n,\theta}(y)) &= \int_{\mathbb{R}} K(y-u) d(F_{n,\theta} - E(F_{n,\theta}))(u) \\
&= \int_{\mathbb{R}} (F_{n,\theta} - E(F_{n,\theta}))(u) dK(y-u) \\
&= \int_{\mathbb{R}} (F_{n,\theta} - E(F_{n,\theta}))(y-s) dK(s).
\end{aligned}
$$

Thus, if we denote $\Delta_{n,\theta}(y) := F_{n,\theta}(y) - E(F_{n,\theta}(y)) = F_{n,\theta}(y) - F_\theta(y)$, we obtain

$$
\begin{aligned}
\left| \tilde{F}_{n,\theta}(y) - E(\tilde{F}_{n,\theta}(y)) - [F_{n,\theta}(y) - E(F_{n,\theta}(y))] \right| & \\
&\hspace{-6em} \leq \left| \int_{\mathbb{R}} (\Delta_{n,\theta}(y-s) - \Delta_{n,\theta}(y)) dK(s) \right| \\
&\hspace{-6em} \leq \sup_{(\theta,y) \in \Phi \times \mathbb{R}} |\Delta_{n,\theta}(y)| \, \|K\|_{TV}.
\end{aligned}
$$

The last bias-term on the right hand side of (42) can be studied using the R_{2n} bound in Shorack and Wellner (1986, p. 766, eq. e), which establishes that for each $\theta \in \Phi$

$$
\sup_{y \in \mathbb{R}} |E(\tilde{F}_{n,\theta}) - F_\theta|(y) \leq \frac{\|\dot{f}_\theta\|_\infty k_2}{2}. \tag{43}
$$

If K is replaced by $K_n(\cdot) = K(\cdot/b_n)$ and we let $k_{2,n} := \int_{\mathbb{R}} x^2 dK_n(x) = b_n^2 k_2$, then (41–43) lead to

$$
\limsup_{n \to \infty} \sqrt{\frac{n}{\log\log(n)}} \sup_{(\theta,y) \in \Phi \times \mathbb{R}} (F_{n,\theta} - F_\theta)(y) < \infty \qquad a.s., \tag{44}
$$

whenever $\limsup(n/\log\log(n))^{1/2} k_{2,n} < \infty$ which holds when

$$
\limsup_{n \to \infty} \sqrt{\frac{n}{\log\log(n)}} b_n^2 < \infty,
$$

and $\sup_{\theta \in \Phi} \|\dot{f}_\theta\|_\infty < \infty$ which has been proved in Section 5.3 under Condition (R) ii).

Uniform almost sure rate of convergence of $T_{2,n}$. Considering for all $i \geq 0$, the random variable $W_i(\vartheta) := H^2(V_i; \vartheta)$, where $\vartheta \in \Theta$, we see that

$$
\sup_{\vartheta \in \Theta} T_{2,n}(\vartheta) = \sup_{\vartheta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} W_i(\vartheta) - E(W_i(\vartheta)) \right|,
$$

where the right hand term is the supremum of an empirical process indexed by a class of Lipschitz bounded functions, which is known to be $o_{a.s.}(n^{-1/2+\gamma})$ for all $\gamma > 0$, see Bordes et al. (1996b) for details, which concludes the proof.

6.7 Proof of Theorem 3.1

i) The proof follows entirely the proof of Theorem 3.1 in [5] and uses the technical results proved in Lemma 3.2.

ii) Consider the following decomposition:

$$
\begin{aligned}
|\hat{f}_n - f| &= \left| \left[\frac{1}{\hat{p}_n} \hat{\Psi}_{n,\hat{\theta}_n} - \frac{1}{p_*} \Psi_{\theta_*} \right] - \left[\frac{1-\hat{p}_n}{\hat{p}_n} \tilde{I}_{n,\hat{\theta}_n} - \frac{1-p_*}{p_*} I_{\theta_*} \right] \right| \\
&= \left| \left[\frac{1}{\hat{p}_n} (\hat{\Psi}_{n,\hat{\theta}_n} - \hat{\Psi}_{n,\theta_*}) + \frac{1}{\hat{p}_n} \hat{\Psi}_{n,\theta_*} - \frac{1}{p_*} \Psi_{\theta_*} \right] \right. \\
&\quad \left. - \left[\frac{1-\hat{p}_n}{\hat{p}_n} (\tilde{I}_{n,\hat{\theta}_n} - \tilde{I}_{n,\theta_*}) + \frac{1-\hat{p}_n}{\hat{p}_n} \tilde{I}_{n,\theta_*} - \frac{1-p_*}{p_*} I_{\theta_*} \right] \right| \\
&\leq \frac{1}{\hat{p}_n} \left(|\hat{\Psi}_{n,\hat{\theta}_n} - \hat{\Psi}_{n,\theta_*}| + |\hat{\Psi}_{n,\theta_*} - \Psi_{\theta_*}| \right) + \Psi_{\theta_*} \left| \frac{1}{\hat{p}_n} - \frac{1}{p_*} \right| \\
&\quad + \frac{1-\hat{p}_n}{\hat{p}_n} \left(|\tilde{I}_{n,\hat{\theta}_n} - \tilde{I}_{n,\theta_*}| + |\tilde{I}_{n,\theta_*} - I_{\theta_*}| \right) \\
&\quad + I_{\theta_*} \left| \frac{1-\hat{p}_n}{\hat{p}_n} - \frac{1-p_*}{p_*} \right|.
\end{aligned}
\tag{45}
$$

It is now enough to study the behavior of $|\hat{\Psi}_{n,\hat{\theta}_n} - \hat{\Psi}_{n,\theta_*}|$ and $|\tilde{I}_{n,\hat{\theta}_n} - \tilde{I}_{n,\theta_*}|$. For all $t \in \mathbb{R}$, we have

$$
|\hat{\Psi}_{n,\hat{\theta}_n}(t) - \hat{\Psi}_{n,\theta_*}(t)| \leq \frac{1}{nb_n} \sum_{i=1}^{n} \left| K\left(\frac{t - Y_i^{\hat{\theta}_n}}{b_n} \right) - K\left(\frac{t - Y_i^{\theta_*}}{b_n} \right) \right|.
\tag{46}
$$

Consider K a kernel density satisfying (K) ii). We propose to study in a generic way the difference of kernels involved on the right hand side of the above expression. For all $(w, z) \in \mathbb{R}^2$, and letting $h := (z - w)/b$, we write the second-order Taylor expansion with integral remainder term:

$$
K\left(\frac{t - w}{b} \right) - K\left(\frac{t - z}{b} \right) = h\dot{K}\left(\frac{t - z}{b} \right) + \frac{h^2}{2} \int_0^1 (1 - u)\ddot{K}\left(\frac{t - r_u}{b} \right) du
$$

where $r_u := (1 - u)z + uw$. Using the symmetry of K and Fubini's Theorem we obtain

$$
\frac{1}{b} \int_{\mathbb{R}} \left| K\left(\frac{t - w}{b} \right) - K\left(\frac{t - z}{b} \right) \right| dt \leq 2h \int_0^{+\infty} \left| \dot{K}(t) \right| dt + h^2 \int_0^{+\infty} \left| \ddot{K}(t) \right| dt
\tag{47}
$$

Replacing w, z respectively by the $Y_i^{\hat{\theta}_n}$ and $Y_i^{\theta_*}$, and b by b_n in (47) we then obtain from (46) the following bound for the L_1 error:

$$
\|\hat{\Psi}_{n,\hat{\theta}_n}(t) - \hat{\Psi}_{n,\theta_*}(t)\|_{L_1} \leq \frac{C\|\theta_n - \theta_*\|_2}{b_n} \times \frac{1}{n} \sum_{i=1}^{n} \left(|X_i| + |X_i|^2 \right),
\tag{48}
$$

the same kind of bound being available for $\|\tilde{I}_{n,\hat{\theta}_n}(t) - \tilde{I}_{n,\theta_*}(t)\|_{L_1}$. In conclusion, according to the decomposition (45), point i) of Theorem 3.1, the respective L_1 a.s. convergence of $\hat{\Psi}_{n,\theta_*}$ and $\tilde{I}_{n,\theta_*}$ towards Ψ_{θ_*} and I_{θ_*} under (20), we get from (48) and the strong law of large numbers that $\|\hat{f}_n - f\|_{L_1} \to 0$ almost surely as $n \to \infty$ whenever $n^{-1/4+\gamma}/b_n = o(1)$.

iii) The proof uses an integrated version of decomposition (45) and the fact that, for all $y \in \mathbb{R}$, the approximation $|\hat{F}_{n,\hat{\theta}_n} - \hat{F}_{n,\theta_*}|(y)$ is controlled by

$$
\begin{aligned}
|\hat{F}_{n,\hat{\theta}_n} - \hat{F}_{n,\theta_*}|(y) &= \left| \int_{-\infty}^{y} \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{t - Y_i^{\hat{\theta}_n}}{b_n}\right) - K\left(\frac{t - Y_i^{\theta_*}}{b_n}\right) dt \right| \\
&\leq \frac{1}{n} \sum_{i=1}^{n} \int_{\mathbb{R}} \left| K\left(\frac{t - Y_i^{\hat{\theta}_n}}{b_n}\right) - K\left(\frac{t - Y_i^{\theta_*}}{b_n}\right) \right| dt \\
&\leq \frac{C\|\theta_n - \theta_*\|_2}{b_n} \times \frac{1}{n} \sum_{i=1}^{n} \left(|X_i| + |X_i|^2 \right), \quad (49)
\end{aligned}
$$

the last term on the right hand side of above inequality being independent from y. The same bound holds for $|\tilde{I}_{n,\hat{\theta}_n} - \tilde{I}_{n,\theta_*}|(y)$ by a similar argument. To conclude, it is enough to use (49) and Corollary 1 p. 766 in Shorack and Wellner (1986), which allows us to control the terms $\|\hat{F}_{n,\theta_*} - F_{\theta_*}\|_\infty$ and $\|\tilde{I}_{n,\theta_*} - I_{\theta_*}\|_\infty$, to obtain (27). The rate on the right hand side of (27) is optimized by considering $b_n = n^{-1/12}$ which then turns into $O_{a.s.}(n^{-1/6+\gamma})$, for all $\gamma > 0$.

Acknowledgments. The author thanks the referees for their helpful and constructive comments. He is also very gratefull to Philippe Barbe, David Hunter and Marie-Laure Martin Magniettes for their help and good advice during the writing of this manuscript.

References

[1] BENJAMINI, Y., and HOCHBERG, Y. (1995) Controlling the false discovery rate: a pratical and powerful approach to multiple testing. *J. Royal Statist. Soc. Ser. B,* **57**, 289–300.

[2] BORDES, L., MOTTELET, S. and VANDEKERKHOVE, P. (2006a). Semiparametric estimation of a two-component mixture model. *Ann. Statist.,* **34**, 1204–1232.

[3] BORDES, L., DELMAS. C, and VANDEKERKHOVE, P. (2006b). Semiparametric estimation of a two-component mixture model when a component is known. *Scand. J. Statist.,* **33**, 733–752.

[4] BORDES, L., CHAUVEAU, D. and VANDEKERKHOVE, P. (2007). A stochastic EM algorithm for a semiparametric mixture model. *Comput. Statist. Data Anal.,* **51**, 5429–5443.

34

[5] BORDES, L., and VANDEKERKHOVE, P. (2010). Semiparametric Two-Component Mixture Model with a Known Component: an Asymptotically Normal Estimator. *Math. Method. Statist.,* **19**, 22–41.

[6] BOUVEYRON C. AND JACQUES J. (2010). Adaptive mixtures of regressions: Improving predictive inference when population has changed. *Pattern Recognition Letters*, **31**, 2237–2247.

[7] BOWMAN, A.W. and AZZALINI, A. (2003). Computational aspects of nonparametric smoothing with illustrations from the sm library. *Comp. Statist. Data Anal.,* **42**, 545–56.

[8] CRUZ-MEDINA, I. R. and HETTMANSPERGER, T. P. (2004). Nonparametric estimation in semiparametric univariate mixture models. *J. Statist. Comput. Simulation,* **74** ,513–524.

[9] CUZICK, J. (1992) Semiparametric additive regression. *J. Royal Statist. Soc. Ser. B,* **3**, 831–843.

[10] CUZICK, J. (1992) Efficient estimates in semiparametric additive regression models with unknown error distribution. *Ann. Statist.,* **20**, 1129–1136.

[11] DEVROYE, L. (1983) The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates. *Ann. Statist.,* **11**, 896–904.

[12] EFRON, B. (2007). Size, power and false discovery rate. *Ann. Statist.,* **35**, 1351–1377.

[13] HALL, P. and ZHOU, X. H. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Ann. Statist.,* **31**, 201–224.

[14] HENG-YAN LEUNG, D. and QIN, J. (2006). Semi-parametric inference in bivariate (multivariate) mixture model. *Statist. Sinica.,* **16**, 153–163.

[15] HETTMANSPERGER, T. P. and THOMAS, H. (2000). Almost nonparametric inference for repeated measures in mixture models. *J. Royal Statist. Soc. Ser. B,* **62** 811–825.

[16] HOHMANN, D. and HOLZMANN, H. (2012). Semiparametric location mixtures with distinct components. *Statistics*, **1**, 1–15.

[17] HUNTER, D. R. and YOUNG, D. S. (2012). Semiparametric Mixtures of Regressions. *J. Nonparametric Statist.,* **24**, 19–38.

[18] HUNTER, D. R., WANG, S. and HETTMANSPERGER, T. P. (2007). Inference for mixtures of symmetric distributions. *Ann. Statist.,* **35**, 224–251.

[19] KUI W. and YAUA, K. K. W. (2007) Two-component Poisson mixture regression modelling of count data with bivariate random effects. *Math. Comput. Modelling*, **46**, 1468–1476.

[20] KUELBS, J. and DUDLEY, R. M. (1980). Log Log Laws for Empirical Measures *Ann. Probab.*, **8**, 405–418.

[21] MAIBORODA, R. and SUGAKOVA, O. (2010). Adaptive estimating equations for a location parameter constructed by using observations with admixture. *Th. Probab. Math. Statist.*, **80**, 101–110.

[22] MARTIN-MAGNIETTE, M. L., MARY-HUARD, T., BÉRARD C. and ROBIN, S. (2008) ChIPmix: mixture model of regressions for two-color ChIP-chip analysis. *Bioinformatics*, **24**, 181–186.

[23] NOCEDAL, J. and WRIGHT, S. J. (1999). *Numerical Optimization.* Springer, Berlin.

[24] SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

[25] SCHUSTER, E. F. and BARKER, R. C. (1987). Using the bootstrap in testing symmetry versus asymmetry. *Comm. Statist. Simulation Comput.,* **16**, 69–84.

[26] TITTERINGTON, D. M., SMITH, A. F. M. and MAKOV, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*, Wiley, Chichester.

[27] VAN DER VAART. A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer-Verlag, New-York.

[28] YAU, K. K. W. , LEEB, A. H., and NG A. S. K. (2003) Finite mixture regression model with random effects: application to neonatal hospital length of stay. *Comput. Statist. Data Anal.,* **41**, 359–366.

[29] YOUNG, D. S, and HUNTER D. R. (2010). Mixtures of Regressions with Predictor-Dependent Mixing Proportions. *Comput. Statist. Data Anal.,* **54**, 2253–2266.

[30] YU K., MAMMEN E. and PARK B. U. (2011) Semiparametric Regression: Efficiency Gains From Modeling the Nonparametric Part. *Bernoulli*, **17**, 736-748.

[31] ZHU H. T. and ZHANG H. (2004) Hypothesis testing in mixture regression models. *J. Royal Statist. Soc. Ser. B,* **66**, 3–16.