Semiparametric estimation of a mixture of two linear regressions in which one component is known

Laurent Bordes, Ivan Kojadinovic, Pierre Vandekerkhove

Semiparametric estimation of a mixture of two linear regressions in which one component is known

L. Bordes[a], I. Kojadinovic[a] and P. Vandekerkhove[b]

[a]Laboratoire de mathématiques et applications, UMR CNRS 5142
Université de Pau et des Pays de l'Adour
B.P. 1155, 64013 Pau Cedex, France
{laurent.bordes,ivan.kojadinovic}@univ-pau.fr

[b]Laboratoire d'analyse et de mathématiques appliquées
UMR CNRS 8050, Université Paris-Est, Marne-la-Vallée, France,
and Georgia Institute of Technology, UMI Georgia Tech CNRS 2958
G.W. Woodruff School of Mechanical Engineering, Atlanta, USA
pierre.vandekerkhove@univ-mlv.fr

January 15, 2013

Abstract

A new estimation method for the two-component mixture model introduced in Vandekerkhove (2012) is proposed. This model, which consists of a two-component mixture of linear regressions in which one component is entirely known while the proportion, the slope, the intercept and the error distribution of the other component are unknown, seems to be of interest for the analysis of large datasets produced from two-color ChIP-chip high-density microarrays. In spite of good performance for datasets of reasonable size, the method proposed in Vandekerkhove (2012) suffers from a serious drawback when the sample size becomes large, as it is based on the optimization of a contrast function whose pointwise computation requires $O(n^2)$ operations. The range of applicability of the method derived in this work is substantially larger as it is based on a method-of-moment estimator whose computation only requires $O(n)$ operations. From a theoretical perspective, the asymptotic normality of both the estimator of the Euclidean parameter vector and of the semiparametric estimator of the c.d.f. of the error is proved under weak conditions not involving the zero-symmetry assumption typically used this last decade. The finite-sample performance of the latter estimators is studied under various scenarios through Monte Carlo experiments. From a more practical perspective, the proposed method is applied to the tone data analyzed, among others, by Hunter and Young (2012), and to the ChIPmix data studied by Martin-Magniette et al. (2008). An extension of the considered model involving an unknown scale parameter for the first component is discussed in the final section.

1 Introduction

Practitioners are frequently interested in modeling the relationship between a random response variable Y and a d-dimensional random explanatory vector X by means of a linear regression model estimated from a random sample $(X_i, Y_i)_{1 \leq i \leq n}$ of (X, Y). Quite often, the homogeneity assumption claiming that the linear regression coefficients are the same for all the observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ is inadequate. To allow different parameters for different groups of observations, a Finite Mixture of Regressions (FMR) can be considered; see Leisch (2004) and Grün and Leisch (2006) for a nice overview.

Statistical inference for the fully parametric FMR model was first considered by Quandt and Ramsey (1978) who proposed an estimation method based on the moment generating function. An EM estimating approach was proposed by De Veaux (1989) in the case of two components. Variations of the latter approach were also considered in Jones and McLachlan (1992) and Turner (2000). Hawkins et al. (2001) studied the problem of determining the number of components in the parametric FMR model using methods derived from the likelihood equation. In Hurn et al. (2003), the authors proposed a Bayesian approach to estimate the regression coefficients and also considered an extension of the model in which the number of components is unspecified. Zhu and Zhang (2004) established the asymptotic theory for maximum likelihood estimators in parametric FMR models. More recently, Städler et al. (2010) proposed an ℓ_1-penalized method based on a Lasso-type estimator for a high-dimensional FMR model with $d \gg n$.

As an alternative to parametric approaches to the estimation of a FMR model, some authors suggested the use of more flexible semiparametric approaches. This research direction finds its origin in the work of Hall and Zhou (2003) in which d-variate semiparametric mixture models of random vectors with independent components were considered. These authors showed in particular that, for $d \geq 3$, it is possible to identify a two-component model without parametrizing the distributions of the component random vectors. To the best of our knowledge, Leung and Qin (2006) were the first to estimate a FMR model semiparametrically. In the two-component case, they studied the situation in which the components are related by Anderson (1979)'s exponential tilt model. Hunter and Young (2012) studied the identifiability of an m-component semiparametric FMR model and numerically investigated an EM algorithm for estimating its parameters. Vandekerkhove (2012) proposed an M-estimation method for a two-component semiparametric mixture of regressions with symmetric errors in which one component is known. The latter approach was applied to data extracted from a high-density microarray and modeled in Martin-Magniette et al. (2008) by means of a parametric FMR. The semiparametric approach of Vandekerkhove (2012) is of interest for two main reasons. Due to its semiparametric nature, the method allows to detect complex structures in the error of the unknown regression component. It can additionally be regarded as a tool to assess the relevance of the usual EM-type Euclidean parameter estimation. Its main drawbacks however are that it is not theoretically valid when the errors are not symmetric and that its use is very computationally expensive for large datasets as it requires the optimization of a contrast function whose pointwise evaluation requires $O(n^2)$ operations.

The object of interest of this paper is the two-component FMR model studied by Vandekerkhove (2012) in which one component is entirely known while the proportion, the slope, the intercept and the error distribution of the other component are unknown. The estimation of the Euclidean parameter vector is achieved through a method of moments. Semiparametric estimators of the c.d.f. and the p.d.f. of the error of the unknown component are proposed. The proof of the asymptotic normality of the Euclidean and functional estimators is not based on zero-symmetry-like assumptions frequently found in the literature but only involves finite moments of order eight for the explanatory variable and the boundness of the p.d.f.s of the errors and their derivatives. The almost sure uniform consistency of the estimator of the p.d.f. of the unknown error is obtained under similar conditions. A consequence of these theoretical results is that, unlike for EM-type approaches, the estimation uncertainty can be assessed through large-sample standard errors for the Euclidean parameters and by means of an approximate confidence band for the c.d.f. of the unknown error. The latter is computed using an unconditional weighted bootstrap whose asymptotic validity is proved.

From a practical perspective, it is worth mentioning that the range of applicability of the resulting semiparametric estimation procedure is substantially larger than the one of Vandekerkhove (2012) as its computation only requires $O(n)$ operations. As a consequence, very large datasets can be easily processed. For instance, as shall be seen in Section 6, the estimation of the parameters of the model from the ChIPmix data considered in Martin-Magniette et al. (2008) consisting of $n = 176,343$ observations took less than 30 seconds on one 2.4 GHz processor. The estimation of the same model from a subset of $n = 30,000$ observations using the method of Vandekerkhove (2012) took more than two days on a similar processor.

The paper is organized as follows. Section 2 is devoted to a detailed description of the model, while Section 3 is concerned with its identifiability through the moment method. The estimators of the Euclidean parameter vector and of the functional parameter are described in detail in Section 4. The finite-sample performance of the proposed estimation method is studied for various scenarios through Monte Carlo experiments in Section 5. In Section 6, the proposed method is applied to the tone data analyzed, among others, by Hunter and Young (2012), and to the ChIPmix data considered in Martin-Magniette et al. (2008). An extension of the FMR model under consideration involving an unknown scale parameter for the first component is discussed in the final section.

2 Problem and notation

Let Z be a Bernoulli random variable with unknown parameter $\pi_0 \in [0,1]$, let X be an $\mathcal{X}$-valued random variable with $\mathcal{X} \subset \mathbb{R}$, and let $\varepsilon^*, \varepsilon^{**}$ be two absolutely continuous centered real valued random variables with finite variances and independent of X. Assume additionally that Z is independent of X, ε^* and ε^{**}. Furthermore, for fixed $\alpha_0^*, \beta_0^*, \alpha_0^{**}, \beta_0^{**} \in \mathbb{R}$, let $\tilde{Y}$ be the random variable defined by

$$\tilde{Y} = (1-Z)(\alpha_0^* + \beta_0^* X + \varepsilon^*) + Z(\alpha_0^{**} + \beta_0^{**} X + \varepsilon^{**}),$$

i.e.,

$$\tilde{Y} = \begin{cases} \alpha_0^* + \beta_0^* X + \varepsilon^* & \text{if} \quad Z = 0, \\ \alpha_0^{**} + \beta_0^{**} X + \varepsilon^{**} & \text{if} \quad Z = 1. \end{cases}$$

The above display is the equation of a mixture of two linear regressions with Z as mixing variable.

Let F^* and F^{**} denote the c.d.f.s of ε^* and ε^{**}, respectively. Furthermore, α_0^*, β_0^* and F^* are assumed known while α_0^{**}, β_0^{**}, π_0 and F^{**} are assumed unknown. The aim of this work is to propose and study an estimator of $(\alpha_0^{**}, \beta_0^{**}, \pi_0, F^{**})$ based on n i.i.d. copies $(X_i, \tilde{Y}_i)_{1 \leq i \leq n}$ of $(X, \tilde{Y})$. Now, define $Y = \tilde{Y} - \alpha_0^* - \beta_0^* X$, $\alpha_0 = \alpha_0^{**} - \alpha_0^*$ and $\beta_0 = \beta_0^{**} - \beta_0^*$, and notice that

$$Y = \begin{cases} \varepsilon^* & \text{if} \quad Z = 0, \\ \alpha_0 + \beta_0 X + \varepsilon & \text{if} \quad Z = 1, \end{cases} \tag{1}$$

where, to simplify the notation, $\varepsilon = \varepsilon^{**}$ and $F = F^{**}$. It follows that the previous estimation problem is equivalent to the problem of estimating $(\alpha_0, \beta_0, \pi_0, F)$ from the observation of n i.i.d. copies $(X_i, Y_i)_{1 \leq i \leq n}$ of (X, Y).

As we continue, the unknown c.d.f.s of X and Y will be denoted by F_X and F_Y, respectively. Also, for any $x \in \mathcal{X}$, the conditional c.d.f. of Y given $X = x$ will be denoted by $F_{Y|X}(\cdot|x)$, and we have

$$F_{Y|X}(y|x) = (1-\pi_0) F^*(y) + \pi_0 F(y - \alpha_0 - \beta_0 x), \qquad y \in \mathbb{R}. \tag{2}$$

It follows that, for any $x \in \mathcal{X}$, $f_{Y|X}(\cdot|x)$, the conditional p.d.f. of Y given $X = x$, can be expressed as

$$f_{Y|X}(y|x) = (1-\pi_0) f^*(y) + \pi_0 f(y - \alpha_0 - \beta_0 x), \qquad y \in \mathbb{R}, \tag{3}$$

where f^* and f are the p.d.f.s of ε^* and ε, assuming that they exist on $\mathbb{R}$.

Note that, as shall be discussed in Section 7, it is possible to consider a slightly more general version of this model involving an unknown scale parameter for the first component. This more elaborate model remains identifiable and estimation through the moment method is theoretically possible. However, from a practical perspective, estimation of this scale parameter through the moment method seems quite unstable insomuch as that an alternative estimation method appears required.

3 Identifiability through the moment method

Since (1) is clearly equivalent to

$$Y = (1-Z)\varepsilon^* + Z(\alpha_0 + \beta_0 X + \varepsilon), \tag{4}$$

we immediately obtain that

$$\mathbb{E}(Y|X) = \pi_0\alpha_0 + \pi_0\beta_0 X \qquad \text{a.s.} \tag{5}$$

It follows that the coefficients $\gamma_{0,1} = \pi_0\alpha_0$ and $\gamma_{0,2} = \pi_0\beta_0$ can be identified from (5) if $\mathcal{X}$ is not reduced to a singleton. In addition, we have

$$\begin{aligned} \mathbb{E}(Y^2|X) &= \mathbb{E}[\{(1-Z)\varepsilon^* + Z(\alpha_0+\beta_0 X+\varepsilon)\}^2|X] \quad \text{a.s.} \\ &= \mathbb{E}(1-Z)\mathbb{E}\{(\varepsilon^*)^2\} + \mathbb{E}(Z)\mathbb{E}\{(\alpha_0+\beta_0X)^2+\varepsilon^2|X\} \quad \text{a.s.} \\ &= (1-\pi_0)(\sigma_0^*)^2 + \pi_0\left(\alpha_0^2 + 2\alpha_0\beta_0 X + \beta_0^2 X^2 + \sigma_0^2\right) \quad \text{a.s.} \\ &= (1-\pi_0)(\sigma_0^*)^2 + \pi_0(\alpha_0^2+\sigma_0^2) + 2\pi_0\alpha_0\beta_0 X + \pi_0\beta_0^2X^2 \quad \text{a.s.,} \end{aligned} \tag{6}$$

where σ_0^* and σ_0 are the standard deviations of ε^* and ε, respectively. If $\mathcal{X}$ contains three points x_1, x_2, x_3 such that the vectors $\{(1,x_1,x_1^2),(1,x_2,x_2^2),(1,x_3,x_3^2)\}$ are linearly independent then, from (6), we can identify the coefficients $\gamma_{0,3} = (1-\pi_0)(\sigma_0^*)^2 + \pi_0(\alpha_0^2+\sigma_0^2)$, $\gamma_{0,4} = 2\pi_0\alpha_0\beta_0$ and $\gamma_{0,5} = \pi_0\beta_0^2$. In other words, under the aforementioned conditions on $\mathcal{X}$, we have

$$\left\{ \begin{aligned} \gamma_{0,1} &= \pi_0\alpha_0 \\ \gamma_{0,2} &= \pi_0\beta_0 \\ \gamma_{0,3} &= (1-\pi_0)(\sigma_0^*)^2 + \pi_0(\alpha_0^2+\sigma_0^2) \\ \gamma_{0,4} &= 2\pi_0\alpha_0\beta_0 = 2\alpha_0\gamma_{0,2} \\ \gamma_{0,5} &= \pi_0\beta_0^2 = \beta_0\gamma_{0,2}. \end{aligned} \right. \tag{7}$$

From the above system of equations, we see that α_0, β_0 and π_0 can be identified provided $\pi_0\beta_0 \neq 0$, that is, provided the unknown component actually exists and its slope is non zero. The latter condition will be assumed to hold in the rest of the paper.

Let us now consider the functional part F of the model. For any $\boldsymbol{\eta} = (\alpha,\beta) \in \mathbb{R}^2$, denote by $J(\cdot,\boldsymbol{\eta})$ the c.d.f. defined by

$$J(t,\boldsymbol{\eta}) = \Pr(Y \leq t + \alpha + \beta X), \qquad t \in \mathbb{R}. \tag{8}$$

For any $t \in \mathbb{R}$, this can be rewritten as

$$\begin{aligned} J(t,\boldsymbol{\eta}) = &\int_{\mathbb{R}} F_{Y|X}(t+\alpha+\beta x|x)\mathrm{d}F_X(x) \\ =&(1-\pi_0)\int_{\mathbb{R}} F^*(t+\alpha+\beta x)\mathrm{d}F_X(x) + \pi_0\int_{\mathbb{R}} F\{t+(\alpha-\alpha_0)+(\beta-\beta_0)x\}\mathrm{d}F_X(x). \end{aligned}$$

For $\boldsymbol{\eta} = \boldsymbol{\eta}_0 = (\alpha_0, \beta_0)$, we then obtain

$$J(t, \boldsymbol{\eta}_0) = (1 - \pi_0) \int_{\mathbb{R}} F^*(t + \alpha_0 + \beta_0 x) \mathrm{d}F_X(x) + \pi_0 F(t), \qquad t \in \mathbb{R}.$$

Now, for any $\boldsymbol{\eta} \in \mathbb{R}^2$, let $K(\cdot, \boldsymbol{\eta})$ be defined by

$$K(t, \boldsymbol{\eta}) = \int_{\mathbb{R}} F^*(t + \alpha + \beta x) \mathrm{d}F_X(x), \qquad t \in \mathbb{R}. \tag{9}$$

It follows that F is identified since

$$F(t) = \frac{1}{\pi_0} \left\{ J(t, \boldsymbol{\eta}_0) - (1 - \pi_0) K(t, \boldsymbol{\eta}_0) \right\}, \qquad t \in \mathbb{R}. \tag{10}$$

The above equation is at the root of the derivation of an estimator for F.

4 Estimation

Let P be the probability distribution of (X, Y). For ease of exposition, we will frequently use the notation adopted in the theory of empirical processes in the sense of van der Vaart and Wellner (2000) or Kosorok (2008) for instance. Given a measurable function $f : \mathbb{R}^2 \to \mathbb{R}^k$, for some integer $k \geq 1$, Pf will denote the integral $\int f \mathrm{d}P$. Also, the empirical measure obtained from the random sample $(X_i, Y_i)_{1 \leq i \leq n}$ will be denoted by $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i, Y_i}$, where $\delta_{x,y}$ is the probability distribution that assigns a mass of 1 at (x, y). The expectation of f under the empirical measure is then $\mathbb{P}_n f = n^{-1} \sum_{i=1}^n f(X_i, Y_i)$ and the quantity $\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n f - Pf)$ is the *empirical process* evaluated at f. The arrow '$\leadsto$' will be used to denote weak convergence in the sense of Definition 1.3.3 of van der Vaart and Wellner (2000) and, for any set S, $\ell^\infty(S)$ will stand for the space of all bounded real-valued functions on S equipped with the uniform metric. Key results and more details can be found for instance in van der Vaart (1998), van der Vaart and Wellner (2000) and Kosorok (2008).

4.1 Estimation of the Euclidean parameter vector

To estimate the Euclidean parameter vector $(\alpha_0, \beta_0, \pi_0) \in \mathbb{R} \times \mathbb{R} \setminus \{0\} \times (0, 1]$, we first need to estimate the vector $\boldsymbol{\gamma}_0 = (\gamma_{0,1}, \ldots, \gamma_{0,5}) \in \mathbb{R}^5$ whose components were expressed in terms of α_0, β_0 and π_0 in the previous section. From (5) and (6), it is natural to consider the regression function

$$d_n(\boldsymbol{\gamma}) = \mathbb{P}_n \varphi_{\boldsymbol{\gamma}}, \qquad \boldsymbol{\gamma} \in \mathbb{R}^5,$$

where, for any $\boldsymbol{\gamma} \in \mathbb{R}^5$, $\varphi_{\boldsymbol{\gamma}} : \mathbb{R}^2 \to \mathbb{R}$ is defined by

$$\varphi_{\boldsymbol{\gamma}}(x, y) = (y - \gamma_1 - \gamma_2 x)^2 + (y^2 - \gamma_3 - \gamma_4 x - \gamma_5 x^2)^2, \qquad x, y \in \mathbb{R}.$$

As an estimator of $\boldsymbol{\gamma}_0$, we then naturally consider $\boldsymbol{\gamma}_n = \arg\min_{\boldsymbol{\gamma}} d_n(\boldsymbol{\gamma})$ that satisfies

$$\dot{d}_n(\boldsymbol{\gamma}_n) = \mathbb{P}_n \dot{\varphi}_{\boldsymbol{\gamma}_n} = 0,$$

where $\dot{\varphi}_{\boldsymbol{\gamma}}$, the gradient of $\varphi_{\boldsymbol{\gamma}}$ with respect to $\boldsymbol{\gamma}$, is given by

$$\dot{\varphi}_{\boldsymbol{\gamma}}(x,y) = -2 \begin{pmatrix} y - \gamma_1 - \gamma_2 x \\ x(y - \gamma_1 - \gamma_2 x) \\ y^2 - \gamma_3 - \gamma_4 x - \gamma_5 x^2 \\ x(y^2 - \gamma_3 - \gamma_4 x - \gamma_5 x^2) \\ x^2(y^2 - \gamma_3 - \gamma_4 x - \gamma_5 x^2) \end{pmatrix}, \qquad x, y \in \mathbb{R}.$$

Now, for any integers $p, q \geq 1$, define

$$\overline{X^p Y^q} = \frac{1}{n} \sum_{i=1}^{n} X_i^p Y_i^q,$$

and let

$$\Gamma_n = 2 \begin{pmatrix} 1 & \overline{X} & 0 & 0 & 0 \\ \overline{X} & \overline{X^2} & 0 & 0 & 0 \\ 0 & 0 & 1 & \overline{X} & \overline{X^2} \\ 0 & 0 & \overline{X} & \overline{X^2} & \overline{X^3} \\ 0 & 0 & \overline{X^2} & \overline{X^3} & \overline{X^4} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\theta}_n = 2 \begin{pmatrix} \overline{Y} \\ \overline{XY} \\ \overline{Y^2} \\ \overline{XY^2} \\ \overline{X^2Y^2} \end{pmatrix},$$

which respectively estimate

$$\Gamma_0 = 2 \begin{pmatrix} 1 & \mathbb{E}(X) & 0 & 0 & 0 \\ \mathbb{E}(X) & \mathbb{E}(X^2) & 0 & 0 & 0 \\ 0 & 0 & 1 & \mathbb{E}(X) & \mathbb{E}(X^2) \\ 0 & 0 & \mathbb{E}(X) & \mathbb{E}(X^2) & \mathbb{E}(X^3) \\ 0 & 0 & \mathbb{E}(X^2) & \mathbb{E}(X^3) & \mathbb{E}(X^4) \end{pmatrix} \quad \text{and} \quad \boldsymbol{\theta}_0 = 2 \begin{pmatrix} \mathbb{E}(Y) \\ \mathbb{E}(XY) \\ \mathbb{E}(Y^2) \\ \mathbb{E}(XY^2) \\ \mathbb{E}(X^2Y^2) \end{pmatrix}.$$

The linear equation $\mathbb{P}_n \dot{\varphi}_{\boldsymbol{\gamma}_n} = 0$ can then equivalently be rewritten as $\Gamma_n \boldsymbol{\gamma}_n = \boldsymbol{\theta}_n$. Provided the matrices Γ_n and Γ_0 are invertible, we can write $\boldsymbol{\gamma}_n = \Gamma_n^{-1} \boldsymbol{\theta}_n$ and $\boldsymbol{\gamma}_0 = \Gamma_0^{-1} \boldsymbol{\theta}_0$.

To obtain an estimator of $(\alpha_0, \beta_0, \pi_0)$, we use the relationships induced by (5) and (6) and recalled in (7). Leaving the third equation aside because it involves the unknown standard deviation σ_0 of ε, we obtain three possible estimators of α_0:

$$\alpha_n = \frac{\gamma_{n,1}\gamma_{n,5}}{\gamma_{n,2}^2}, \qquad \alpha_n = \frac{\gamma_{n,4}}{2\gamma_{n,2}}, \qquad \text{or} \qquad \alpha_n = \frac{\gamma_{n,4}^2}{4\gamma_{n,1}\gamma_{n,5}},$$

three possibles estimators of β_0:

$$\beta_n = \frac{\gamma_{n,5}}{\gamma_{n,2}}, \qquad \beta_n = \frac{\gamma_{n,4}}{2\gamma_{n,1}}, \qquad \text{or} \qquad \beta_n = \frac{\gamma_{n,2}\gamma_{n,4}^2}{4\gamma_{n,5}\gamma_{n,1}^2},$$

and, three possibles estimators of π_0:

$$\pi_n = \frac{\gamma_{n,2}^2}{\gamma_{n,5}}, \qquad \pi_n = \frac{2\gamma_{n,1}\gamma_{n,2}}{\gamma_{n,4}}, \qquad \text{or} \qquad \pi_n = \frac{4\gamma_{n,1}^2\gamma_{n,5}}{\gamma_{n,4}^2}.$$

There are therefore 27 possible estimators of $(\alpha_0, \beta_0, \pi_0)$. Their asymptotics can be obtained under very reasonable conditions. Unfortunately, all 27 estimators turned out to behave quite poorly in small samples. This prompted us to look for alternative estimators within the "same class".

We now describe an estimator of $(\alpha_0, \beta_0, \pi_0)$ that was obtained empirically and that behaves significantly better for small samples than the aforementioned ones. The new regression function under consideration is $d_n(\boldsymbol{\gamma}) = \mathbb{P}_n \varphi_{\boldsymbol{\gamma}}$, $\boldsymbol{\gamma} \in \mathbb{R}^8$, where, for any $(x, y) \in \mathbb{R}^2$,

$$\varphi_{\boldsymbol{\gamma}}(x, y) = (y - \gamma_1 - \gamma_2 x)^2 + (y^2 - \gamma_3 - \gamma_4 x^2)^2 + (x - \gamma_5)^2 + (x^2 - \gamma_6)^2 + (x^3 - \gamma_7)^2 + (x^4 - \gamma_8)^2.$$

Now, let

$$\Gamma_n = 2 \begin{pmatrix} 1 & \overline{X} & 0 & 0 & 0 & 0 & 0 & 0 \\ \overline{X} & \overline{X^2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \overline{X^2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \overline{X^2} & \overline{X^4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\theta}_n = 2 \begin{pmatrix} \overline{Y} \\ \overline{XY} \\ \overline{Y^2} \\ \overline{X^2Y^2} \\ \overline{X} \\ \overline{X^2} \\ \overline{X^3} \\ \overline{X^4} \end{pmatrix},$$

which respectively estimate

$$\Gamma_0 = 2 \begin{pmatrix} 1 & \mathbb{E}(X) & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbb{E}(X) & \mathbb{E}(X^2) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \mathbb{E}(X^2) & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbb{E}(X^2) & \mathbb{E}(X^4) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\theta}_0 = 2 \begin{pmatrix} \mathbb{E}(Y) \\ \mathbb{E}(XY) \\ \mathbb{E}(Y^2) \\ \mathbb{E}(X^2Y^2) \\ \mathbb{E}(X) \\ \mathbb{E}(X^2) \\ \mathbb{E}(X^3) \\ \mathbb{E}(X^4) \end{pmatrix}.$$

Then, proceeding as previously, provided the matrices Γ_n and Γ_0 are invertible, the estimator $\boldsymbol{\gamma}_n = \arg\min_\gamma d_n(\boldsymbol{\gamma})$ of $\boldsymbol{\gamma}_0 = \Gamma_0^{-1}\boldsymbol{\theta}_0$ is given by $\boldsymbol{\gamma}_n = \Gamma_n^{-1}\boldsymbol{\theta}_n$. To obtain an estimator of $(\alpha_0, \beta_0, \pi_0)$, we have, from the second term of the regression function, that

$$\gamma_{0,4} = \frac{\operatorname{cov}(X^2, Y^2)}{\operatorname{V}(X^2)} = \frac{\operatorname{cov}(X^2, Y^2)}{\gamma_{0,8} - \gamma_{0,6}^2},$$

where the second equality comes from the fact that $\gamma_{0,6} = \mathbb{E}(X^2)$ and $\gamma_{0,8} = \mathbb{E}(X^4)$. Now, using (4), we find

$$\mathrm{cov}(X^2, Y^2) = \pi_0 \beta_0^2 \mathrm{V}(X^2) + 2\pi_0\alpha_0\beta_0 \mathrm{cov}(X^2, X),$$

which, combined with the fact that $\gamma_{0,1} = \pi_0\alpha_0$ and $\gamma_{0,2} = \pi_0\beta_0$, gives

$$\mathrm{cov}(X^2, Y^2) = \gamma_{0,2}\beta_0(\gamma_{0,8} - \gamma_{0,6}^2) + 2\gamma_{0,1}\beta_0(\gamma_{0,7} - \gamma_{0,5}\gamma_{0,6}).$$

This leads to the following estimator of $(\alpha_0, \beta_0, \pi_0)$:

$$\begin{aligned}
\beta_n &= g^\beta(\boldsymbol{\gamma}_n) = \frac{\gamma_{n,4}}{\gamma_{n,2} + 2\gamma_{n,1}(\gamma_{n,7} - \gamma_{n,5}\gamma_{n,6})/(\gamma_{n,8} - \gamma_{n,6}^2)}, \\
\pi_n &= g^\pi(\boldsymbol{\gamma}_n) = \frac{\gamma_{n,2}}{\beta_n}, \\
\alpha_n &= g^\alpha(\boldsymbol{\gamma}_n) = \frac{\gamma_{n,1}}{\pi_n}.
\end{aligned}$$

As we continue, the subsets of $\mathbb{R}^8$ on which the functions g^α, g^β and g^π exist and are differentiable will be denoted by $\mathcal{D}^\alpha$, $\mathcal{D}^\beta$ and $\mathcal{D}^\pi$, respectively, and $\mathcal{D}^{\alpha,\beta,\pi}$ will stand for $\mathcal{D}^\alpha \cap \mathcal{D}^\beta \cap \mathcal{D}^\pi$.

To derive the asymptotic behavior of the estimator $(\alpha_n, \beta_n, \pi_n) = (g^\alpha(\boldsymbol{\gamma}_n), g^\beta(\boldsymbol{\gamma}_n), g^\pi(\boldsymbol{\gamma}_n))$, we consider the following assumptions:

A1. (i) X has a finite fourth order moment; (ii) X has a finite eighth order moment.

A2. $\mathrm{V}(X) > 0$ and $\mathrm{V}(X^2) > 0$.

Clearly, Assumption A1 (ii) implies Assumption A1 (i), and Assumption A2 implies that the matrix Γ_0 is invertible.

The following result, proved in Appendix A, characterizes the asymptotic behavior of the estimator $(\alpha_n, \beta_n, \pi_n)$.

Proposition 4.1. *Assume that $\boldsymbol{\gamma}_0 \in \mathcal{D}^{\alpha,\beta,\pi}$.*

(i) Under Assumptions A1 (i) and A2, $(\alpha_n, \beta_n, \pi_n) \xrightarrow{a.s.} (\alpha_0, \beta_0, \pi_0)$.

(ii) Suppose that Assumptions A1 (ii) and A2 are satisfied and let $\Psi_{\boldsymbol{\gamma}}$ be the 3 by 8 matrix defined by

$$\Psi_{\boldsymbol{\gamma}} = \begin{pmatrix} \frac{\partial g^\alpha}{\partial \gamma_1} & \cdots & \frac{\partial g^\alpha}{\partial \gamma_8} \\ \frac{\partial g^\beta}{\partial \gamma_1} & \cdots & \frac{\partial g^\beta}{\partial \gamma_8} \\ \frac{\partial g^\pi}{\partial \gamma_1} & \cdots & \frac{\partial g^\pi}{\partial \gamma_8} \end{pmatrix} (\boldsymbol{\gamma}), \qquad \boldsymbol{\gamma} \in \mathcal{D}^{\alpha,\beta,\pi}.$$

Then,

$$\sqrt{n}(\alpha_n - \alpha_0, \beta_n - \beta_0, \pi_n - \pi_0) = -\mathbb{G}_n(\Psi_{\boldsymbol{\gamma}_0}\Gamma_0^{-1}\dot{\varphi}_{\boldsymbol{\gamma}_0}) + o_P(1).$$

As a consequence, $\sqrt{n}(\alpha_n - \alpha_0, \beta_n - \beta_0, \pi_n - \pi_0)$ converges in distribution to a centered normal random vector with covariance matrix $\Sigma = \Psi_{\boldsymbol{\gamma}_0} \Gamma_0^{-1} P(\dot{\varphi}_{\boldsymbol{\gamma}_0} \dot{\varphi}_{\boldsymbol{\gamma}_0}^\top) \Gamma_0^{-1} \Psi_{\boldsymbol{\gamma}_0}^\top$, which can be consistently estimated by $\Sigma_n = \Psi_{\boldsymbol{\gamma}_n} \Gamma_n^{-1} \mathbb{P}_n(\dot{\varphi}_{\boldsymbol{\gamma}_n} \dot{\varphi}_{\boldsymbol{\gamma}_n}^\top) \Gamma_n^{-1} \Psi_{\boldsymbol{\gamma}_n}^\top$ in the sense that $\Sigma_n \xrightarrow{a.s.} \Sigma$.

An immediate consequence of the previous result is that large-sample standard errors of α_n, β_n and π_n are given by the square root of the diagonal elements of the matrix $n^{-1}\Sigma_n$. The finite-sample performance of these estimators is investigated in Section 5 and they are used in the illustrations of Section 6.

4.2 Estimation of the functional parameter

To estimate the unknown c.d.f. F of ε, it is natural to start from (10). For a known $\boldsymbol{\eta} = (\alpha, \beta) \in \mathbb{R}^2$, the term $J(\cdot, \boldsymbol{\eta})$ defined in (8) may be estimated by the empirical c.d.f. of the random sample $(Y_i - \alpha - \beta X_i)_{1 \leq i \leq n}$, i.e.,

$$J_n(t, \boldsymbol{\eta}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(Y_i - \alpha - \beta X_i \leq t), \qquad t \in \mathbb{R}.$$

Similarly, since F^* (the c.d.f. of ε^*) is known, a natural estimator of the term $K(t, \boldsymbol{\eta})$ defined in (9) is given by the empirical mean of the random sample $\{F^*(t + \alpha + \beta X_i)\}_{1 \leq i \leq n}$, i.e.,

$$K_n(t, \boldsymbol{\eta}) = \frac{1}{n} \sum_{i=1}^{n} F^*(t + \alpha + \beta X_i), \qquad t \in \mathbb{R}.$$

To obtain estimators of $J(\cdot, \boldsymbol{\eta}_0)$ and $K(\cdot, \boldsymbol{\eta}_0)$, it is then natural to consider the plug-in estimators $J_n(\cdot, \boldsymbol{\eta}_n)$ and $K_n(\cdot, \boldsymbol{\eta}_n)$, respectively, based on the estimator $\boldsymbol{\eta}_n = (\alpha_n, \beta_n) = (g^\alpha, g^\beta)(\boldsymbol{\gamma}_n)$ of $\boldsymbol{\eta}_0$ proposed in the previous subsection.

We shall therefore consider the following nonparametric estimator of F :

$$F_n(t) = \frac{1}{\pi_n} \left\{ J_n(t, \boldsymbol{\eta}_n) - (1 - \pi_n) K_n(t, \boldsymbol{\eta}_n) \right\}, \qquad t \in \mathbb{R}. \tag{11}$$

Note that F_n is not necessarily a c.d.f. as it is not necessarily increasing and can be smaller than zero or greater than one. In practice, we shall consider the partially corrected estimator $(F_n \vee 0) \wedge 1$, where $\vee$ and $\wedge$ denote the maximum and minimum, respectively.

To derive the asymptotic behavior of the previous estimator, we consider the following additional assumptions on the p.d.f.s f^* and f of ε^* and ε, respectively:

A3. (i) f^* and f exist and are bounded on $\mathbb{R}$; (ii) $(f^*)'$ and f' exist and are bounded on $\mathbb{R}$.

Before stating one of our main results, let us first define some additional notation. Let $\mathcal{F}^J$ and $\mathcal{F}^K$ be two classes of measurable functions from $\mathbb{R}^2$ to $\mathbb{R}$ defined respectively by

$$\mathcal{F}^J = \left\{(x,y) \mapsto \psi^J_{t,\boldsymbol{\eta}}(x,y) = \mathbf{1}(y - \alpha - \beta x \leq t) : t \in \mathbb{R}, \boldsymbol{\eta} = (\alpha,\beta) \in \mathbb{R}^2\right\}$$

and

$$\mathcal{F}^K = \left\{(x,y) \mapsto \psi^K_{t,\boldsymbol{\eta}}(x,y) = F^*(t + \alpha + \beta x) : t \in \mathbb{R}, \boldsymbol{\eta} = (\alpha,\beta) \in \mathbb{R}^2\right\}.$$

Furthermore, let $\mathcal{D}^{\alpha,\beta,\pi}_{\boldsymbol{\gamma}_0}$ be a bounded subset of $\mathcal{D}^{\alpha,\beta,\pi}$ containing $\boldsymbol{\gamma}_0$, and let $\mathcal{F}^{\alpha,\beta,\pi}$ be the class of measurable functions from $\mathbb{R}^2$ to $\mathbb{R}^3$ defined by

$$\mathcal{F}^{\alpha,\beta,\pi} = \left\{(x,y) \mapsto -\Psi_{\boldsymbol{\gamma}}\Gamma_0^{-1}\dot{\varphi}_{\boldsymbol{\gamma}}(x,y) = \left(\psi^\alpha_{\boldsymbol{\gamma}}(x,y), \psi^\beta_{\boldsymbol{\gamma}}(x,y), \psi^\pi_{\boldsymbol{\gamma}}(x,y)\right) : \boldsymbol{\gamma} \in \mathcal{D}^{\alpha,\beta,\pi}_{\boldsymbol{\gamma}_0}\right\}.$$

With the previous notation, notice that, for any $t \in \mathbb{R}$,

$$\sqrt{n}\{J_n(t,\boldsymbol{\eta}_0) - J(t,\boldsymbol{\eta}_0)\} = \mathbb{G}_n\psi^J_{t,\boldsymbol{\eta}_0} \qquad \text{and} \qquad \sqrt{n}\{K_n(t,\boldsymbol{\eta}_0) - K(t,\boldsymbol{\eta}_0)\} = \mathbb{G}_n\psi^K_{t,\boldsymbol{\eta}_0},$$

and that, under Assumptions A1 (ii) and A2, Proposition 4.1 states that

$$\sqrt{n}\,(\alpha_n - \alpha_0, \beta_n - \beta_0, \pi_n - \pi_0) = \mathbb{G}_n\left(\psi^\alpha_{\boldsymbol{\gamma}_0}, \psi^\beta_{\boldsymbol{\gamma}_0}, \psi^\pi_{\boldsymbol{\gamma}_0}\right) + o_P(1).$$

Next, for any $\boldsymbol{\gamma} \in \mathcal{D}^{\alpha,\beta,\pi}_{\boldsymbol{\gamma}_0}$, let

$$\psi^F_{t,\boldsymbol{\gamma}} = \frac{1}{\pi}\psi^J_{t,\boldsymbol{\eta}} + f(t)\psi^\alpha_{\boldsymbol{\gamma}} + f(t)\mathbb{E}(X)\psi^\beta_{\boldsymbol{\gamma}} - \frac{1-\pi}{\pi}\psi^K_{t,\boldsymbol{\eta}} + \frac{P\psi^K_{t,\boldsymbol{\eta}} - P\psi^J_{t,\boldsymbol{\eta}}}{\pi^2}\psi^\pi_{\boldsymbol{\gamma}}, \tag{12}$$

with $\boldsymbol{\eta} = (\alpha,\beta) = (g^\alpha, g^\beta)(\boldsymbol{\gamma})$ and $\pi = g^\pi(\boldsymbol{\gamma})$.

The following result, proved in Appendix B, gives the weak limit of the empirical process $\sqrt{n}(F_n - F)$.

Proposition 4.2. *Assume that $\boldsymbol{\gamma}_0 \in \mathcal{D}^{\alpha,\beta,\pi}$ and that Assumptions A1, A2 and A3 hold. Then, for any $t \in \mathbb{R}$,*

$$\sqrt{n}\{F_n(t) - F(t)\} = \mathbb{G}_n\psi^F_{t,\boldsymbol{\gamma}_0} + Q_{n,t},$$

where $\sup_{t\in\mathbb{R}}|Q_{n,t}| = o_P(1)$, and the empirical process $t \mapsto \mathbb{G}_n\psi^F_{t,\boldsymbol{\gamma}_0}$ converges weakly to $t \mapsto \mathbb{G}\psi^F_{t,\boldsymbol{\gamma}_0}$ in $\ell^\infty(\overline{\mathbb{R}})$.

Let us now discuss the estimation of the p.d.f. f of ε. Starting from (10) and after differentiation, it seems sensible to estimate the expectation $\mathbb{E}\{f^*(t + \alpha_0 + \beta_0 X)\}$, $t \in \mathbb{R}$, by the empirical mean of the observable sample $\{f^*(t + \alpha_n + \beta_n X_i)\}_{1\leq i\leq n}$. Hence, a natural estimator of f can be defined, for any $t \in \mathbb{R}$, by

$$f_n(t) = \frac{1}{\pi_n}\left\{\frac{1}{nh_n}\sum_{i=1}^n \kappa\left(\frac{t - Y_i + \alpha_n + \beta_n X_i}{h_n}\right) - \frac{(1-\pi_n)}{n}\sum_{i=1}^n f^*(t + \alpha_n + \beta_n X_i)\right\}, \tag{13}$$

where κ is a kernel function on $\mathbb{R}$ and $(h_n)_{n\geq 1}$ is a sequence of bandwidths converging to zero.

In the same way that F_n is not necessarily a c.d.f., f_n is not necessarily a p.d.f. In practice, we shall use the partially corrected estimator $f_n \vee 0$. A fully corrected estimator can be obtained from the work of Glad et al. (2003).

Consider the following additional assumptions on $(h_n)_{n\geq 1}$, κ and f^* :

A4. (i) $h_n = cn^{-\alpha}$ with $\alpha \in (0, 1/2)$ and $c > 0$ a constant; (ii) κ is a p.d.f. with bounded variations on $\mathbb{R}$ and a finite first order moment; (iii) the p.d.f. f^* has bounded variations on $\mathbb{R}$.

The following result is proved in Appendix C.

Proposition 4.3. *If $\boldsymbol{\gamma}_0 \in \mathcal{D}^{\alpha,\beta,\pi}$, and under Assumptions A1 (i), A2, A3 and A4,*

$$\sup_{t\in\mathbb{R}} |f_n(t) - f(t)| \xrightarrow{a.s.} 0.$$

Finally, note that, in all our numerical experiments, the kernel part of f_n was computed using the excellent `ks` R package (Duong, 2012) in which the univariate plug-in selector of Wand and Jones (1994) was used for the bandwidth h_n.

4.3 An unconditional weighted bootstrap for $\sqrt{n}(F_n - F)$ with application to confidence bands for F

In applications, it may be of interest to carry out inference on F. The result stated in this section can be used for this purpose. It is based on the unconditional multiplier central limit theorem for empirical processes (see e.g. Kosorok, 2008, Theorem 10.1 and Corollary 10.3) and can be used to obtain approximate independent copies of $\sqrt{n}(F_n - F)$.

Given i.i.d. random variables $\xi_1, \ldots, \xi_n$ with mean 0, variance 1, satisfying $\int_0^\infty \{\Pr(|\xi_1| > x)\}^{1/2} \mathrm{d}x < \infty$, and independent of the random sample $(X_i, Y_i)_{1\leq i\leq n}$, let

$$\mathbb{G}'_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\xi_i - \bar{\xi}) \delta_{X_i,Y_i},$$

where $\bar{\xi} = n^{-1} \sum_{i=1}^{n} \xi_i$. Also, let $\Psi_{\boldsymbol{\gamma}_n} \Gamma_n^{-1} \dot{\varphi}_{\boldsymbol{\gamma}_n} = -\left(\hat{\psi}^{\alpha}_{\boldsymbol{\gamma}_n}, \hat{\psi}^{\beta}_{\boldsymbol{\gamma}_n}, \hat{\psi}^{\pi}_{\boldsymbol{\gamma}_n}\right)$ and, for any $t \in \mathbb{R}$, let

$$\hat{\psi}^{F}_{t,\boldsymbol{\gamma}_n} = \frac{1}{\pi_n} \psi^{J}_{t,\boldsymbol{\eta}_n} + f_n(t) \hat{\psi}^{\alpha}_{\boldsymbol{\gamma}_n} + f_n(t) \bar{X} \hat{\psi}^{\beta}_{\boldsymbol{\gamma}_n} - \frac{1-\pi_n}{\pi_n} \psi^{K}_{t,\boldsymbol{\eta}_n} + \frac{\mathbb{P}_n \psi^{K}_{t,\boldsymbol{\eta}_n} - \mathbb{P}_n \psi^{J}_{t,\boldsymbol{\eta}_n}}{\pi_n^2} \hat{\psi}^{\pi}_{\boldsymbol{\gamma}_n} \tag{14}$$

be an estimated version of the influence function $\psi^{F}_{t,\boldsymbol{\gamma}_0}$ arising in Proposition 4.2, where $\boldsymbol{\eta}_n = (\alpha_n, \beta_n) = (g^{\alpha}, g^{\beta})(\boldsymbol{\gamma}_n)$ and $\pi_n = g^{\pi}(\boldsymbol{\gamma}_n)$.

The following proposition, proved in Appendix D, suggests, when n is large, to interpret $t \mapsto \mathbb{G}_n' \hat{\psi}^F_{t,\gamma_n}$ as an independent copy of $\sqrt{n}(F_n - F)$.

Proposition 4.4. *Assume that $\gamma_0 \in \mathcal{D}^{\alpha,\beta,\pi}$, and that Assumptions A1, A2, A3 and A4 hold. Then, the process $(t \mapsto \mathbb{G}_n \psi^F_{t,\gamma_0}, t \mapsto \mathbb{G}_n' \hat{\psi}^F_{t,\gamma_n})$ converges weakly to $(t \mapsto \mathbb{G}\psi^F_{t,\gamma_0}, t \mapsto \mathbb{G}'\psi^F_{t,\gamma_0})$ in $\{\ell^\infty(\overline{\mathbb{R}})\}^2$, where $t \mapsto \mathbb{G}'\psi^F_{t,\gamma_0}$ is an independent copy of $t \mapsto \mathbb{G}\psi^F_{t,\gamma_0}$.*

Let us now explain how the latter result can be used in practice to obtain an approximate confidence band for F. Let N be a large integer and let $\xi_i^{(j)}$, $i \in \{1,\dots,n\}$, $j \in \{1,\dots,N\}$, be i.i.d. random variables with mean 0, variance 1, satisfying $\int_0^\infty \{\Pr(|\xi_i^{(j)}| > x)\}^{1/2} \mathrm{d}x < \infty$, and independent of the data $(X_i, Y_i)_{1 \le i \le n}$. For any $j \in \{1,\dots,N\}$, let $\mathbb{G}_n^{(j)} = n^{-1/2} \sum_{i=1}^n (\xi_i^{(j)} - \bar{\xi}^{(j)}) \delta_{X_i,Y_i}$, where $\bar{\xi}^{(j)} = n^{-1} \sum_{i=1}^n \xi_i^{(j)}$. Then, a consequence of Propositions 4.2 and 4.4 is that

$$\left(\sqrt{n}(F_n - F), t \mapsto \mathbb{G}_n^{(1)} \hat{\psi}^F_{t,\gamma_n}, \dots, t \mapsto \mathbb{G}_n^{(N)} \hat{\psi}^F_{t,\gamma_n}\right) \\ \rightsquigarrow \left(t \mapsto \mathbb{G}\psi^F_{t,\gamma_0}, t \mapsto \mathbb{G}^{(1)}\psi^F_{t,\gamma_0}, \dots, t \mapsto \mathbb{G}^{(N)}\psi^F_{t,\gamma_0}\right)$$

in $\{\ell^\infty(\overline{\mathbb{R}})\}^{N+1}$, where $\mathbb{G}^{(1)}, \dots, \mathbb{G}^{(N)}$ are independent copies of the P-Brownian bridge $\mathbb{G}$. From the continuous mapping theorem, it follows that

$$\left(\sup_{t\in\mathbb{R}} |\sqrt{n}(F_n - F)|, \sup_{t\in\mathbb{R}} |\mathbb{G}_n^{(1)} \hat{\psi}^F_{t,\gamma_n}|, \dots, \sup_{t\in\mathbb{R}} |\mathbb{G}_n^{(N)} \hat{\psi}^F_{t,\gamma_n}|\right) \\ \rightsquigarrow \left(\sup_{t\in\mathbb{R}} |\mathbb{G}\psi^F_{t,\gamma_0}|, \sup_{t\in\mathbb{R}} |\mathbb{G}^{(1)}\psi^F_{t,\gamma_0}|, \dots, \sup_{t\in\mathbb{R}} |\mathbb{G}^{(N)}\psi^F_{t,\gamma_0}|\right)$$

in $[0,\infty)^{N+1}$. The previous result suggests to estimate quantiles of $\sup_{t\in\mathbb{R}} |\sqrt{n}(F_n - F)|$ using the generalized inverse of the empirical c.d.f.

$$G_{n,N}(x) = \frac{1}{N} \sum_{j=1}^n \mathbf{1}\left\{\sup_{t\in\mathbb{R}} |\mathbb{G}_n^{(j)} \hat{\psi}^F_{t,\gamma_n}| \le x\right\}. \tag{15}$$

A large-sample confidence band of level $1-p$ for F is thus given by $F_n \pm G_{n,N}^{-1}(1-p)/\sqrt{n}$. Examples of such confidence bands are given in Figures 1 and 2, and the finite-sample properties of the above construction are empirically investigated in Section 5. Note that in all our numerical experiments, the multipliers $\xi_i^{(j)}$ were taken from the standard normal distribution, and that the supremum in the previous display was replaced by a maximum over 100 points $U_1, \dots, U_{100}$ uniformly spaced over the interval $[\min_{1\le i\le n}(Y_i - \alpha_n - \beta_n X_i), \max_{1\le i\le n}(Y_i - \alpha_n - \beta_n X_i)]$.

Finally, notice that Proposition 4.4 also suggests to estimate the standard error of $F_n(t)$ for some fixed $t \in \mathbb{R}$ by $n^{-1/2}\{\mathbb{P}_n(\hat{\psi}^F_{t,\gamma_n})^2\}^{1/2}$. The finite-sample performance of this estimator is investigated in Section 5 for different values of t.

5 Monte Carlo experiments

A large number of Monte Carlo experiments was carried out to investigate the influence on the estimators of various factors such as the degree of overlap of the mixed populations, the proportion of the unknown component π_0, or the shape of the noise ε involved in the unknown regression model. Starting from (1), the following generic data generating models were considered:

$$\begin{aligned}
\text{WO} : \varepsilon^* \sim \mathcal{N}(0,1),\ (\alpha_0,\beta_0) = (2,1),\ X \sim \mathcal{N}(2,3^2),\ \mathbb{E}(\varepsilon^2) = 1,\\
\text{MO} : \varepsilon^* \sim \mathcal{N}(0,1),\ (\alpha_0,\beta_0) = (2,1),\ X \sim \mathcal{N}(2,3^2),\ \mathbb{E}(\varepsilon^2) = 4,\\
\text{SO} : \varepsilon^* \sim \mathcal{N}(0,1),\ (\alpha_0,\beta_0) = (1,0.5),\ X \sim \mathcal{N}(1,2^2),\ \mathbb{E}(\varepsilon^2) = 4.
\end{aligned}$$

The abbreviations WO, MO and SO stand respectively for "Weak Overlap", "Medium Overlap" and "Strong Overlap". Three possibilities were considered for the distribution of ε: the centered normal (the corresponding data generating models will be abbreviated by WOn, MOn and SOn), a gamma distribution with shape parameter equal to two and rate parameter equal to a half shifted to have mean zero (the corresponding models will be abbreviated by WOg, MOg and SOg) and a standard exponential shifted to have mean zero (the corresponding models will be abbreviated by WOe, MOe and SOe). Depending on the model they are used in, all three error distributions are scaled so that ε has the desired variance.

Examples of datasets generated from WOn, MOg and SOe with $n = 500$ and $\pi_0 = 0.7$ are represented in the first column of graphs of Figure 1. The solid (resp. dashed) lines represent the true (resp. estimated) regression lines. The graphs of the second column represent, for each of WOn, MOg and SOe, the true c.d.f. F of ε (solid line) and its estimate F_n (dashed line) defined in (11). The dotted lines represent approximate confidence bands of level 0.95 for F computed as explained in Subsection 4.3 with $N = 10,000$. Finally, the graphs of the third column represent, for each of WOn, MOg and SOe, the true p.d.f. f of ε (solid line) and its estimate f_n (dashed line) defined in (13).

[Figure 1 about here.]

For each of the three groups of data generating models, {WOn, MOn, SOn}, {WOg, MOg, SOg} and {WOe, MOe, SOe}, the values 0.4 and 0.7 were considered for π_0, and the values 100, 300, 1000 and 5000 were considered for n. For each of the nine data generating scenarios, each value of π_0, and each value of n, $M = 1000$ random samples were generated. Tables 1, 2 and 3 report the number m of samples out of M for which $\pi_n \notin (0,1]$, as well as the estimated bias and standard deviation of α_n, β_n, π_n, $F_n\{F^{-1}(0.1)\}$, $F_n\{F^{-1}(0.5)\}$ and $F_n\{F^{-1}(0.9)\}$ computed from the $M - m$ valid estimates.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

A first general comment concerning the results reported in Tables 1, 2 and 3 is that the number m of samples for which $\pi_n \notin (0,1]$ is the highest for the SO scenarios followed by the MO scenarios and then the WO scenarios. Also, for a fixed amount of overlap between the two mixed populations, it is when the distribution of ε is exponential that m tends to be the highest followed by the gamma and the normal cases. Hence, as expected, the SO scenarios are the hardest and, for a given degree of overlap, the most difficult problems are those involving exponential errors for the unknown regression component.

Influence of the shape of the p.d.f. of ε. A surprising result, when observing Tables 1, 2 and 3, is that the nature of the distribution of ε appears to have very little influence on the performance of the estimators α_n, β_n and π_n. Under weak and moderate overlap in particular, the estimated bias and standard deviations of the estimators are almost unaffected by the distribution of the error of the unknown component.

The effect of the degree of overlap. As expected, the performance of the estimators α_n, β_n and π_n is strongly affected by the degree of overlap. Notice however that the results obtained under the WO and MO data generating scenarios are rather comparable, while the performance of the estimators gets significantly worse when switching to the SO scenarios, especially for π_n. Notice also that, overall, the biases of α_n and β_n are negative under WO and MO and positive under SO.

The influence of π_0. For a given degree of overlap and sample size, the parameter that seems to affect the most the performance of the estimators is the proportion π_0 of the unknown component. On one hand, the number of samples for which $\pi_n \notin (0,1]$ is lower for $\pi_0 = 0.4$ than for $\pi_0 = 0.7$. On the other hand, when considering the samples for which $\pi_n \in (0,1]$, the finite-sample behavior of α_n and β_n improves very clearly when π_0 switches from 0.4 to 0.7.

Performance of the functional estimator. The study of $F_n\{F^{-1}(p)\}$ for $p \in \{0.1, 0.5, 0.9\}$ clearly shows that, for a given degree of overlap between the two mixed population, the performance of the functional estimator is the best when the distribution of ε is normal followed by the gamma and the exponential settings. In addition, it appears that $F_n\{F^{-1}(p)\}$, $p \in \{0.1, 0.5\}$, behaves the best under the MO scenarios, and that, somehow surprisingly, $F_n\{F^{-1}(0.9)\}$ achieves its best results under the SO scenarios.

Asymptotics. The results reported in Tables 1, 2 and 3 are in accordance with the asymptotic theory stated in the previous section. In particular, as expected, the estimated biases and standard deviations of all the estimators tend to zero as n increases. Notice for instance that under SOg and SOe with $\pi_0 = 0.4$ (two of the most difficult scenarios), the estimated standard deviation of α_n is greater than 7 for $n = 100$, drops below 0.7 for $n = 300$, and becomes very reasonable for $n = 1000$ and 5000.

Let us now present the results of the Monte Carlo experiments used to investigate the finite-sample performance of the estimators of the standard errors of α_n, β_n, π_n and

$F_n\{F^{-1}(p)\}$, $p \in \{0.1, 0.5, 0.9\}$, mentioned below Proposition 4.1 and at the end of Subsection 4.3, respectively. The setting is the same as previously with the exception that $n \in \{100, 300, 1000, 5000, 25000\}$. The results are partially reported in Table 4 which gives, for scenarios WOn, MOg and SOe and each of the aforementioned estimators, the standard deviation of the estimates multiplied by $\sqrt{n}$ and the mean of the estimated standard errors multiplied by $\sqrt{n}$. As can be seen, for all estimators and all scenarios, the standard deviation of the estimates and the mean of the estimated standard errors are always very close for $n = 25,000$. The convergence to zero of the difference between these two quantities appears however slower for $F_n\{F^{-1}(p)\}$, $p \in \{0.1, 0.5, 0.9\}$, than for α_n, β_n and π_n, the worse results being obtained for $F_n\{F^{-1}(0.1)\}$. The results also confirm that the SO scenarios are the hardest. Notice finally that the estimated standard errors of α_n and β_n seem to underestimate on average the variability of α_n and β_n, and that the variability of π_n and $F_n\{F^{-1}(p)\}$, $p \in \{0.1, 0.5, 0.9\}$ appears to be underestimated on average for the WO scenarios, and overestimated on average for the SO scenarios.

[Table 4 about here.]

We end this section by an investigation of the finite-sample properties of the confidence band construction proposed in Subsection 4.3. Table 5 reports the proportion of samples for which

$$\max_{t \in \{U_1, \dots, U_{100}\}} |F_n(t) - F(t)| > n^{-1/2} G_{n,N}^{-1}(0.95),$$

where $G_{n,N}$ is defined as in (15) with $N = 1000$, and $U_1, \dots, U_n$ are uniformly spaced over the interval $[\min_{1 \leq i \leq n}(Y_i - \alpha_n - \beta_n X_i), \max_{1 \leq i \leq n}(Y_i - \alpha_n - \beta_n X_i)]$. As could have been partly expected from the results reported in Table 4, the confidence bands are too narrow on average for the WO and MO scenarios, the worse results being obtained when the error of the unknown component is exponential. The results are, overall, more satisfactory for the SO scenarios. In all cases, the estimated coverage probability appears to converge to 0.95, although the convergence appears to be slow.

[Table 5 about here.]

6 Illustrations

We first applied the proposed method to a dataset initially reported in Cohen (1980) and subsequently analyzed by De Veaux (1989) and Hunter and Young (2012), among others. The dataset consists of $n = 150$ observations $(x_i, \tilde{y}_i)$ where the x_i are actual tones and the $\tilde{y}_i$ are the corresponding perceived tones by a trained musician. To apply the proposed semiparametric approach, we make the assumption that the equation of the tilted component is $y = x$. Such an hypothesis seems to be in accordance with the detailed description of the dataset given in Hunter and Young (2012). The transformation $y_i = \tilde{y}_i - x_i$ was then applied to obtain a dataset (x_i, y_i) that fits into the setting considered in this work. The original dataset and the transformed dataset are represented in the upper left and upper right plots of Figure 2.

[Figure 2 about here.]

The approach proposed in this paper was applied under the assumption that the distribution of ε^* in (1) is normal with standard deviation 0.079. The latter value was obtained by considering the upper right plot of Figure 2 and by computing the sample standard deviation of the y_i such that $y_i \in (-0.25, 0.25)$ and $x_i < 1.75$ or $x_i > 2.25$.

The estimate $(1.652, -0.817, 0.790)$ was obtained for $(\alpha_0, \beta_0, \pi_0)$ with $(0.217, 0.108, 0.104)$ as vector of estimated standard errors. The corresponding estimated regression line is represented by a solid line in the upper right plot of Figure 2. The estimate $(F_n \vee 0) \wedge 1$ (resp. $f_n \vee 0$) of the unknown c.d.f. F (resp. p.d.f. f) of ε is represented in the lower left (resp. right) plot of Figure 2. The dotted lines in the lower left plot represent an approximate confidence band of level 0.95 for F computed as explained in Subsection 4.3 using $N = 10,000$. Note that, from the results of the previous section, the later is probably too narrow. Numerical integration using the R function `integrate` (R Development Core Team, 2012) gave $\int_{-1}^{1} (f_n \vee 0) \approx 1.01$. The results reported in Figure 2 suggest that a normal assumption for the error of the second component might not be appropriate.

As a second application, we considered the NimbleGen high density array dataset analyzed by Martin-Magniette et al. (2008). The dataset, produced by a two color ChIP-chip experiment, consists of $n = 176,343$ observations $(x_i, \tilde{y}_i)$. A parametric mixture of linear regressions with two unknown components was fitted to the data by Martin-Magniette et al. (2008) under the assumption of normal errors using an EM approach. More details can be found in Vandekerkhove (2012, Section 4.4). The latter author suggested to consider that the intercept and the slope of the first component were precisely estimated by the values 1.47 and 0.82, respectively, obtained by Martin-Magniette et al. (2008), and applied the transformation $y_i = \tilde{y}_i - (1.47 + 0.82x_i)$ to obtain a dataset (x_i, y_i) that fits into the setting considered in this work. The original dataset of Martin-Magniette et al. (2008) and the transformed dataset are represented in the upper left and upper right plots of Figure 3.

[Figure 3 about here.]

The approach proposed in this work was applied under the hypothesis that the distribution of ε^* in (1) is normal with standard deviation 0.492. The latter value comes from the consideration of the upper right plot of Figure 3 and is the sample standard deviation of the y_i for which $x_i < 8.5$ or $x_i > 14$.

The estimate $(0.483, 0.075, 0.351)$ was obtained for $(\alpha_0, \beta_0, \pi_0)$ with $(0.037, 0.002, 0.008)$ as vector of estimated standard errors. The corresponding estimated regression line is represented by a solid line in the upper right plot of Figure 3 while the dashed line represents the (transformed) regression line estimated by Martin-Magniette et al. (2008) under the assumption of normal errors. The estimate $(F_n \vee 0) \wedge 1$ (resp. $f_n \vee 0$) of the unknown c.d.f. F (resp. p.d.f. f) of ε is represented in the lower left (resp. right) plot of Figure 3. Numerical integration using the R function `integrate` gave $\int_{-6}^{6}(f_n \vee 0) \approx 1.03$. The estimation of $(\alpha_0, \beta_0, \pi_0, f, F)$, implemented in R, took less than 30 seconds on one 2.4 GHz processor. The lower right plot of Figure 2 clearly confirms that a normal assumption for the error of the second component is not appropriate.

7 Extension of the model and discussion

From the two illustrations presented in the previous section, we see that the price to pay for no parametric constraints on the second component is a complete specification of the first component. As mentioned in Section 2, from a theoretical perspective, it is possible to improve this situation by introducing an unknown scale parameter for the first component. Using the notation previously defined, the extended model that we have in mind can be written as

$$Y = \begin{cases} \sigma_0^* \bar{\varepsilon}^* & \text{if} \quad Z = 0, \\ \alpha_0 + \beta_0 X + \varepsilon & \text{if} \quad Z = 1, \end{cases} \tag{16}$$

where $\bar{\varepsilon}^*$ is assumed to have variance one and known c.d.f. $\bar{F}$ while σ_0^* is unknown. With respect to the model given in (1), this simply amounts to writing ε^* as $\sigma_0^* \bar{\varepsilon}^*$ and the c.d.f. F^* of ε^* as $F^* = \bar{F}(\cdot/\sigma_0^*)$. The Euclidean parameter vector of this extended model is therefore $(\alpha_0, \beta_0, \pi_0, \sigma_0^*)$ and the functional parameter is F, the c.d.f. of ε.

The model given in (16) is identifiable provided $\mathcal{X}$, the set of possible values of X, contains four points x_1, x_2, x_3, x_4 such that the vectors $\{(1, x_i, x_i^2, x_i^3)\}_{1 \le i \le 4}$ are linearly independent. This can be verified by using, in addition to (5) and (6), the fact that

$$\mathbb{E}(Y^3|X) = \pi_0\alpha_0(\alpha_0^2 + 3\sigma_0^2) + 3\pi_0\beta_0(\alpha_0^2 + \sigma_0^2)X + 3\pi_0\alpha_0\beta_0^2 X^2 + \pi_0\beta_0^3 X^3 \quad \text{a.s.} \tag{17}$$

By proceeding as in Section 3, one can for instance show that

$$(\sigma_0^*)^2 = \frac{\gamma_{0,3}\gamma_{0,5} - \gamma_{0,7}\gamma_{0,2}}{\gamma_{0,5} - \gamma_{0,2}^2}, \tag{18}$$

where $\gamma_{0,2}$ is the coefficient of X in (5), $\gamma_{0,3}$ and $\gamma_{0,5}$ are the coefficients of 1 and X^2, respectively, in (6), and $\gamma_{0,7}$ is the coefficient of X^2 in (17).

From a practical perspective however, using relationship (18) for estimation (or a similar equation resulting from (5), (6) and (17)) turned out to be highly unstable. The reason why estimation of σ_0^* by the moment method does not work satisfactorily seems to be due to the fact that $(\sigma_0^*)^2$ is always the difference of two positive quantities. The estimation of each quantity is not precise enough to ensure that their difference is close to $(\sigma_0^*)^2$, and the difference is often negative. As an alternative estimation method, an iterative EM-type algorithm could be used to estimate all the unknown parameters of the extended model. Unfortunately, a weakness of such algorithms is that, up to now, the asymptotics of the resulting estimators are not known.

A Proof of Proposition 4.1

Proof. Let us prove (i). From Assumption A1 (i) and (4), we have that $\mathbb{E}(X^p Y^q)$ is finite for all integers $p, q \in \{0, 1, 2\}$. It follows that all the components of the vector of expectations $\mathbb{E}\{\dot\varphi_{\boldsymbol\gamma_0}(X, Y)\} = P\dot\varphi_{\boldsymbol\gamma_0}$ are finite. The strong law of large numbers then implies that $\mathbb{P}_n\dot\varphi_{\boldsymbol\gamma_0} \xrightarrow{\text{a.s.}} P\dot\varphi_{\boldsymbol\gamma_0}$. Using the fact that $\boldsymbol\gamma_0$ is a zero of $\boldsymbol\gamma \mapsto P\dot\varphi_{\boldsymbol\gamma}$, that $\mathbb{P}_n\dot\varphi_{\boldsymbol\gamma_0} = \Gamma_n\boldsymbol\gamma_0 - \boldsymbol\theta_n$, and that $\mathbb{P}_n\dot\varphi_{\boldsymbol\gamma_n} = \Gamma_n\boldsymbol\gamma_n - \boldsymbol\theta_n = 0$, we obtain that $\Gamma_n(\boldsymbol\gamma_n - \boldsymbol\gamma_0) \xrightarrow{\text{a.s.}} 0$. The strong law of large numbers also implies that $\Gamma_n \xrightarrow{\text{a.s.}} \Gamma_0$. Matrix inversion being continuous with respect to any usual topology on the space of square matrices, Assumption A2 implies that $\Gamma_n^{-1} \xrightarrow{\text{a.s.}} \Gamma_0^{-1}$. The continuous mapping theorem then implies that $\Gamma_n^{-1}\Gamma_n(\boldsymbol\gamma_n - \boldsymbol\gamma_0) = \boldsymbol\gamma_n - \boldsymbol\gamma_0 \xrightarrow{\text{a.s.}} 0$. Since $\boldsymbol\gamma_0 \in \mathcal{D}^{\alpha,\beta,\pi}$, the strong consistency of $(\alpha_n, \beta_n, \pi_n)$ is finally again a consequence of the continuous mapping theorem as the function

$$\boldsymbol\gamma \mapsto \left(g^\alpha, g^\beta, g^\pi\right)(\boldsymbol\gamma) = (\alpha, \beta, \pi) \tag{19}$$

from $\mathbb{R}^8$ to $\mathbb{R}^3$ is continuous on $\mathcal{D}^{\alpha,\beta,\pi}$.

Let us now prove (ii). Using the fact that $P\dot\varphi_{\boldsymbol\gamma_0} = 0$ and $\mathbb{P}_n\dot\varphi_{\boldsymbol\gamma_n} = 0$, we have

$$\mathbb{P}_n\dot\varphi_{\boldsymbol\gamma_0} - P\dot\varphi_{\boldsymbol\gamma_0} = -(\mathbb{P}_n\dot\varphi_{\boldsymbol\gamma_n} - \mathbb{P}_n\dot\varphi_{\boldsymbol\gamma_0}) = -\mathbb{P}_n(\dot\varphi_{\boldsymbol\gamma_n} - \dot\varphi_{\boldsymbol\gamma_0}) = -\Gamma_n(\boldsymbol\gamma_n - \boldsymbol\gamma_0),$$

which implies that $\mathbb{G}_n\dot\varphi_{\boldsymbol\gamma_0} = -\Gamma_n\sqrt{n}(\boldsymbol\gamma_n - \boldsymbol\gamma_0)$. From Assumption A1 (ii) and (4), we have that the covariance matrix of the random vector $\dot\varphi_{\boldsymbol\gamma_0}(X, Y)$ is finite. The multivariate central limit theorem then implies that $\mathbb{G}_n\dot\varphi_{\boldsymbol\gamma_0}$ converges in distribution to a centered multivariate normal random vector $\mathbb{G}\dot\varphi_{\boldsymbol\gamma_0}$ with covariance matrix $P\dot\varphi_{\boldsymbol\gamma_0}\dot\varphi_{\boldsymbol\gamma_0}^\top$. Since $(\mathbb{G}_n\dot\varphi_{\boldsymbol\gamma_0}, \Gamma_n) \rightsquigarrow (\mathbb{G}\dot\varphi_{\boldsymbol\gamma_0}, \Gamma_0)$ and under Assumption A2, we obtain, from the continuous mapping theorem, that

$$\sqrt{n}(\boldsymbol\gamma_n - \boldsymbol\gamma_0) = -\Gamma_n^{-1}\mathbb{G}_n\dot\varphi_{\boldsymbol\gamma_0} \rightsquigarrow -\Gamma_0^{-1}\mathbb{G}\dot\varphi_{\boldsymbol\gamma_0}.$$

The map defined in (19) is differentiable at $\boldsymbol\gamma_0$ since $\boldsymbol\gamma_0 \in \mathcal{D}^{\alpha,\beta,\pi}$. We can thus apply the delta method with that map to obtain that

$$\sqrt{n}(\alpha_n - \alpha_0, \beta_n - \beta_0, \pi_n - \pi_0) = -\Psi_{\boldsymbol\gamma_0}\Gamma_n^{-1}\mathbb{G}_n\dot\varphi_{\boldsymbol\gamma_0} + o_P(1),$$

Since $\Gamma_n^{-1} \xrightarrow{\text{a.s.}} \Gamma_0^{-1}$ under Assumption A2, we obtain that

$$\sqrt{n}(\alpha_n - \alpha_0, \beta_n - \beta_0, \pi_n - \pi_0) = -\Psi_{\boldsymbol\gamma_0}\Gamma_0^{-1}\mathbb{G}_n\dot\varphi_{\boldsymbol\gamma_0} + o_P(1).$$

It remains to prove that $\Sigma_n \xrightarrow{\text{a.s.}} \Sigma$. Under Assumption A1 (ii), the strong law of large numbers implies that $\mathbb{P}_n\dot\varphi_{\boldsymbol\gamma_0}\dot\varphi_{\boldsymbol\gamma_0}^\top \xrightarrow{\text{a.s.}} P\dot\varphi_{\boldsymbol\gamma_0}\dot\varphi_{\boldsymbol\gamma_0}^\top$. The fact that $\mathbb{P}_n\dot\varphi_{\boldsymbol\gamma_n}\dot\varphi_{\boldsymbol\gamma_n}^\top = \mathbb{P}_n\dot\varphi_{\boldsymbol\gamma_0}\dot\varphi_{\boldsymbol\gamma_0}^\top +$

$\mathbb{P}_n(\dot\varphi_{\boldsymbol{\gamma}_n}\dot\varphi_{\boldsymbol{\gamma}_n}^\top - \dot\varphi_{\boldsymbol{\gamma}_0}\dot\varphi_{\boldsymbol{\gamma}_0}^\top) \xrightarrow{\text{a.s.}} P\dot\varphi_{\boldsymbol{\gamma}_0}\dot\varphi_{\boldsymbol{\gamma}_0}^\top$ is then a consequence of the fact that $\boldsymbol{\gamma}_n \xrightarrow{\text{a.s.}} \boldsymbol{\gamma}_0$ and the continuous mapping theorem. Similarly, since $\boldsymbol{\gamma}_0 \in \mathcal{D}^{\alpha,\beta,\gamma}$, we additionally have that $\Psi_{\boldsymbol{\gamma}_n} \xrightarrow{\text{a.s.}} \Psi_{\boldsymbol{\gamma}_0}$. Combined with the fact that, under Assumption A2, $\Gamma_n^{-1} \xrightarrow{\text{a.s.}} \Gamma_0^{-1}$, we obtain that $\Sigma_n \xrightarrow{\text{a.s.}} \Sigma$ from the continuous mapping theorem. □

B Proof of Proposition 4.2

The proof of Proposition 4.2 is based on three lemmas.

Lemma B.1. *The classes of functions $\mathcal{F}^J$ and $\mathcal{F}^K$ are P-Donsker. So is the class $\mathcal{F}^{\alpha,\beta,\pi}$ provided Assumptions A1 (ii) and A2 hold, and $\boldsymbol{\gamma}_0 \in \mathcal{D}^{\alpha,\beta,\pi}$.*

Proof. The class $\mathcal{F}^J$ is the class of indicator functions $(x,y) \mapsto \mathbf{1}\{(x,y) \in C_{t,\boldsymbol{\eta}}\}$, where $C_{t,\boldsymbol{\eta}} = \{(x,y) \in \mathbb{R}^2 : y \leq t + \alpha + \beta x\}$. The collection $\mathcal{C} = \{C_{t,\boldsymbol{\eta}} : t \in \mathbb{R}, \boldsymbol{\eta} = (\alpha,\beta) \in \mathbb{R}^2\}$ is the set of all half-spaces in $\mathbb{R}^2$. From van der Vaart and Wellner (2000, Exercise 14, p 152), it is a VC class with VC dimension 4. By Lemma 9.8 of Kosorok (2008), $\mathcal{F}^J$ has the same VC dimension as $\mathcal{C}$. Being a set of indicator functions, $\mathcal{F}^J$ clearly possesses a square integrable envelope function and is therefore P-Donsker.

The class $\mathcal{F}^K$ is a collection of monotone functions, and it is easy to verify that it has VC dimension 1. Furthermore, it clearly possesses a square integrable envelope function because the elements of $\mathcal{F}^K$ are bounded. It is therefore P-Donsker.

The components classes of class $\mathcal{F}^{\alpha,\beta,\pi}$ are well defined since Assumption A2 holds and $\boldsymbol{\gamma}_0 \in \mathcal{D}^{\alpha,\beta,\pi}$. It is easy to see that they are linear combinations of a finite collection of functions that, from Assumption A1 (ii), is P-Donsker. The components classes of $\mathcal{F}^{\alpha,\beta,\pi}$ are therefore VC classes. They possess square integrable envelope functions because $\mathcal{D}^{\alpha,\beta,\pi}_{\boldsymbol{\gamma}_0}$ is a bounded set. The class $\mathcal{F}^{\alpha,\beta,\pi}$ is therefore P-Donsker. □

Lemma B.2. *Under Assumptions A1 (i) and A3 (i),*

$$\sup_{t\in\mathbb{R}} P(\psi^J_{t,\boldsymbol{\eta}} - \psi^J_{t,\boldsymbol{\eta}_0})^2 \to 0 \qquad \textit{and} \qquad \sup_{t\in\mathbb{R}} P(\psi^K_{t,\boldsymbol{\eta}} - \psi^K_{t,\boldsymbol{\eta}_0})^2 \to 0 \qquad \textit{as} \qquad \boldsymbol{\eta} \to \boldsymbol{\eta}_0.$$

Proof. For class $\mathcal{F}^J$, for any $t \in \mathbb{R}$, we have

$$\begin{aligned}
P(\psi^J_{t,\boldsymbol{\eta}} - \psi^J_{t,\boldsymbol{\eta}_0})^2 &= |P(\psi^J_{t,\boldsymbol{\eta}} + \psi^J_{t,\boldsymbol{\eta}_0} - 2\psi^J_{t,\boldsymbol{\eta}}\psi^J_{t,\boldsymbol{\eta}_0})| \\
&= P\{(\psi^J_{t,\boldsymbol{\eta}} - \psi^J_{t,\boldsymbol{\eta}_0})\mathbf{1}(\alpha_0 + \beta_0 x < \alpha + \beta x)\} + P\{(\psi^J_{t,\boldsymbol{\eta}_0} - \psi^J_{t,\boldsymbol{\eta}})\mathbf{1}(\alpha_0 + \beta_0 x > \alpha + \beta x)\} \\
&= \int_{\mathbb{R}} \left\{F_{Y|X}(t + \alpha + \beta x|x) - F_{Y|X}(t + \alpha_0 + \beta_0 x|x)\right\} \mathbf{1}(\alpha_0 + \beta_0 x < \alpha + \beta x)\mathrm{d}F_X(x) \\
&\quad + \int_{\mathbb{R}} \left\{F_{Y|X}(t + \alpha_0 + \beta_0 x|x) - F_{Y|X}(t + \alpha + \beta x|x)\right\} \mathbf{1}(\alpha_0 + \beta_0 x > \alpha + \beta x)\mathrm{d}F_X(x) \\
&\leq \int_{\mathbb{R}} \left|F_{Y|X}(t + \alpha_0 + \beta_0 x|x) - F_{Y|X}(t + \alpha + \beta x|x)\right| \mathrm{d}F_X(x),
\end{aligned}$$

where $F_{Y|X}$ is defined in (2). Since $f_{Y|X}(\cdot|x)$ defined in (3) exists for all $x \in \mathcal{X}$, the mean value theorem enables us to write, for any $t \in \mathbb{R}$ and $x \in \mathcal{X}$,

$$F_{Y|X}(t+\alpha+\beta x|x) - F_{Y|X}(t+\alpha_0+\beta_0 x|x) = f_{Y|X}(t+\tilde{\alpha}_{x,t}+\tilde{\beta}_{x,t}x|x)\left\{(\alpha-\alpha_0)+x(\beta-\beta_0)\right\},$$

where $\tilde{\alpha}_{x,t}+\tilde{\beta}_{x,t}x$ is between $\alpha+\beta x$ and $\alpha_0+\beta_0 x$. It follows that

$$\begin{aligned}\sup_{t\in\mathbb{R}} P(\psi^J_{t,\boldsymbol{\eta}} - \psi^J_{t,\boldsymbol{\eta}_0})^2 &\leq \sup_{t\in\mathbb{R}} \int_{\mathbb{R}} f_{Y|X}(t+\tilde{\alpha}_{x,t}+\tilde{\beta}_{x,t}x|x)\,|(\alpha-\alpha_0)+x(\beta-\beta_0)|\,\mathrm{d}F_X(x) \\ &\leq \left\{\sup_{t\in\mathbb{R}} f^*(t) + \sup_{t\in\mathbb{R}} f(t)\right\}\left\{|\alpha-\alpha_0| + \mathbb{E}(|X|)|\beta-\beta_0|\right\}.\end{aligned}$$

Under Assumption A3 (i), the supremum on the right of the previous display is finite and, under Assumption A1 (i), so is $\mathbb{E}(|X|)$. We therefore obtain the desired result.

For class $\mathcal{F}^K$, we have

$$\begin{aligned}\sup_{t\in\mathbb{R}} P(\psi^K_{t,\boldsymbol{\eta}} - \psi^K_{t,\boldsymbol{\eta}_0})^2 &= \int_{\mathbb{R}} \{F^*(t+\alpha+\beta x) - F^*(t+\alpha_0+\beta_0 x)\}^2 \mathrm{d}F_X(x) \\ &\leq \int_{\mathbb{R}} |F^*(t+\alpha+\beta x) - F^*(t+\alpha_0+\beta_0 x)| \mathrm{d}F_X(x),\end{aligned}$$

from the convexity of $x \mapsto x^2$ on $[0,1]$. Proceeding as previously, by the mean value theorem, we obtain that

$$\sup_{t\in\mathbb{R}} P(\psi^K_{t,\boldsymbol{\eta}} - \psi^K_{t,\boldsymbol{\eta}_0})^2 \leq \left\{\sup_{t\in\mathbb{R}} f^*(t)\right\}\left\{|\alpha-\alpha_0| + \mathbb{E}(|X|)|\beta-\beta_0|\right\}.$$

Under Assumptions A1 (i) and A3 (i), the right-hand side of the previous inequality tends to zero as $\boldsymbol{\eta} \to \boldsymbol{\eta}_0$. □

Lemma B.3. *Under Assumptions A1 (ii), A2 and A3 (ii), for any $t \in \mathbb{R}$,*

$$\begin{aligned}\sqrt{n}\{J_n(\boldsymbol{\eta}_n, t) - J(\boldsymbol{\eta}_0, t)\} &= \sqrt{n}\left(\mathbb{P}_n \psi^J_{t,\boldsymbol{\eta}_n} - P\psi^J_{t,\boldsymbol{\eta}_0}\right) \\ &= \mathbb{G}_n\Big(\psi^J_{t,\boldsymbol{\eta}_0} + [(1-\pi_0)\mathbb{E}\{f^*(t+\alpha_0+\beta_0 X)\} + \pi_0 f(t)]\,\psi^\alpha_{\boldsymbol{\gamma}_0} \\ &\qquad + [(1-\pi_0)\mathbb{E}\{X f^*(t+\alpha_0+\beta_0 X)\} + \pi_0 f(t)\mathbb{E}(X)]\,\psi^\beta_{\boldsymbol{\gamma}_0}\Big) + R^J_{n,t},\end{aligned}$$

and

$$\begin{aligned}\sqrt{n}\{K_n(\boldsymbol{\eta}_n, t) - K(\boldsymbol{\eta}_0, t)\} &= \sqrt{n}\left(\mathbb{P}_n \psi^K_{t,\boldsymbol{\eta}_n} - P\psi^K_{t,\boldsymbol{\eta}_0}\right) \\ &= \mathbb{G}_n\left(\psi^K_{t,\boldsymbol{\eta}_0} + \mathbb{E}\{f^*(t+\alpha_0+\beta_0 X)\}\psi^\alpha_{\boldsymbol{\gamma}_0} + \mathbb{E}\{X f^*(t+\alpha_0+\beta_0 X)\}\psi^\beta_{\boldsymbol{\gamma}_0}\right) + R^K_{n,t},\end{aligned}$$

where $\sup_{t\in\mathbb{R}} |R^J_{n,t}| \to_p 0$ *and* $\sup_{t\in\mathbb{R}} |R^K_{n,t}| \to_p 0$.

Proof. We only prove the first statement as the proof of the second statement is similar. We have

$$\sqrt{n}\left(\mathbb{P}_n\psi^J_{t,\boldsymbol{\eta}_n} - P\psi^J_{t,\boldsymbol{\eta}_0}\right) = \mathbb{G}_n\left(\psi^J_{t,\boldsymbol{\eta}_n} - \psi^J_{t,\boldsymbol{\eta}_0}\right) + \mathbb{G}_n\psi^J_{t,\boldsymbol{\eta}_0} + \sqrt{n}P\left(\psi^J_{t,\boldsymbol{\eta}_n} - \psi^J_{t,\boldsymbol{\eta}_0}\right), \qquad t \in \mathbb{R}.$$

Using the fact that $\boldsymbol{\eta}_n \xrightarrow{\text{a.s.}} \boldsymbol{\eta}_0$, Lemma B.1, and Lemma B.2, we can apply Theorem 2.1 in van der Vaart and Wellner (2007) to obtain that

$$\sup_{t\in\mathbb{R}}\left|\mathbb{G}_n\left(\psi^J_{t,\boldsymbol{\eta}_n} - \psi^J_{t,\boldsymbol{\eta}_0}\right)\right| \to_p 0.$$

Furthermore, for any $t \in \mathbb{R}$, we have

$$\sqrt{n}P\left(\psi^J_{t,\boldsymbol{\eta}_n} - \psi^J_{t,\boldsymbol{\eta}_0}\right) = \sqrt{n}\int_{\mathbb{R}}\left\{F_{Y|X}(t+\alpha_n+\beta_n x|x) - F_{Y|X}(t+\alpha_0+\beta_0 x|x)\right\}\mathrm{d}F_X(x),$$

where $F_{Y|X}$ is defined in (2). Since $f'_{Y|X}(\cdot|x)$, the derivative of $f_{Y|X}(\cdot|x)$, exists for all $x \in \mathcal{X}$ from Assumption A3 (ii) and (3), we can apply the second-order mean value theorem to obtain

$$\sqrt{n}P\left(\psi^J_{t,\boldsymbol{\eta}_n} - \psi^J_{t,\boldsymbol{\eta}_0}\right) = \sqrt{n}\int_{\mathbb{R}} f_{Y|X}(t+\alpha_0+\beta_0 x|x)\{(\alpha_n-\alpha_0)+(\beta_n-\beta_0)x\}\mathrm{d}F_X(x) + R^J_{n,t},$$

where

$$R^J_{n,t} = \frac{\sqrt{n}}{2}\int_{\mathbb{R}} f'_{Y|X}(t+\tilde{\alpha}_{x,t,n}+\tilde{\beta}_{x,t,n}x|x)\{(\alpha_n-\alpha_0)+(\beta_n-\beta_0)x\}^2\mathrm{d}F_X(x),$$

and $\tilde{\alpha}_{x,t,n}+\tilde{\beta}_{x,t,n}x$ is between $\alpha_0+\beta_0 x$ and $\alpha_n+\beta_n x$. Now, from (3),

$$\begin{aligned}\sup_{t\in\mathbb{R}}|R^J_{n,t}| \le \sqrt{n}&\left\{\sup_{t\in\mathbb{R}}(f^*)'(t) + \sup_{t\in\mathbb{R}} f'(t)\right\} \\ &\times\left\{(\alpha_n-\alpha_0)^2 + (\beta_n-\beta_0)^2\mathbb{E}(X^2) + 2|\alpha_n-\alpha_0||\beta_n-\beta_0|\mathbb{E}(|X|)\right\}.\end{aligned}$$

The supremum on the right of the previous inequality is finite from Assumption A3 (ii), and so are $\mathbb{E}(|X|)$ and $\mathbb{E}(X^2)$ from Assumption A1 (ii). Furthermore, under Assumptions A1 (ii) and A2, we know from Proposition 4.1 that $\sqrt{n}(\alpha_n-\alpha_0,\beta_n-\beta_0)$ converges in distribution while $(\alpha_n,\beta_n) \xrightarrow{\text{a.s.}} (\alpha_0,\beta_0)$. It follows that $\sup_{t\in\mathbb{R}}|R^J_{n,t}| \to_p 0$. Hence, we obtain that

$$\begin{aligned}\sqrt{n}P\left(\psi^J_{t,\boldsymbol{\eta}_n} - \psi^J_{t,\boldsymbol{\eta}_0}\right) = {}&\mathbb{E}\{f_{Y|X}(t+\alpha_0+\beta_0 X|X)\}\sqrt{n}(\alpha_n-\alpha_0) \\ &+ \mathbb{E}\{Xf_{Y|X}(t+\alpha_0+\beta_0 X|X)\}\sqrt{n}(\beta_n-\beta_0) + R^J_{n,t}, \qquad t\in\mathbb{R}.\end{aligned}$$

The desired result finally follows from the expression of $f_{Y|X}$ given in (3) and Proposition 4.1.
$\square$

Proof of Proposition 4.2. Under Assumptions A1 (ii) and A2, and since $\boldsymbol{\gamma}_0 \in \mathcal{D}^{\alpha,\beta,\pi}$, we know, from Lemma B.1, that the classes $\mathcal{F}^J$, $\mathcal{F}^K$ and $\mathcal{F}^{\alpha,\beta,\pi}$ are P-Donsker. It follows that

$$\left(t \mapsto \mathbb{G}_n \psi^J_{t,\boldsymbol{\eta}_0}, t \mapsto \mathbb{G}_n \psi^K_{t,\boldsymbol{\eta}_0}, \mathbb{G}_n \psi^\alpha_{\boldsymbol{\gamma}_0}, \mathbb{G}_n \psi^\beta_{\boldsymbol{\gamma}_0}, \mathbb{G}_n \psi^\pi_{\boldsymbol{\gamma}_0}\right)$$

converges weakly in $\{\ell^\infty(\overline{\mathbb{R}})\}^2 \times \mathbb{R}^3$. Assumption A3 (i) then implies that the functions $t \mapsto \mathbb{E}\{f_{Y|X}(t + \alpha_0 + \beta_0 X|X)\}$, $t \mapsto \mathbb{E}\{X f_{Y|X}(t + \alpha_0 + \beta_0 X|X)\}$, $t \mapsto \mathbb{E}\{f^*(t + \alpha_0 + \beta_0 X)\}$, and $t \mapsto \mathbb{E}\{X f^*(t + \alpha_0 + \beta_0 X)\}$ are bounded. By the continuous mapping theorem, we thus obtain that

$$\begin{pmatrix} t \mapsto \mathbb{G}_n \left(\psi^J_{t,\boldsymbol{\eta}_0} + \mathbb{E}\{f_{Y|X}(t + \alpha_0 + \beta_0 X|X)\}\psi^\alpha_{\boldsymbol{\gamma}_0} + \mathbb{E}\{X f_{Y|X}(t + \alpha_0 + \beta_0 X|X)\}\psi^\beta_{\boldsymbol{\gamma}_0}\right) \\ t \mapsto \mathbb{G}_n \left(\psi^K_{t,\boldsymbol{\eta}_0} + \mathbb{E}\{f^*(t + \alpha_0 + \beta_0 X)\}\psi^\alpha_{\boldsymbol{\gamma}_0} + \mathbb{E}\{X f^*(t + \alpha_0 + \beta_0 X)\}\psi^\beta_{\boldsymbol{\gamma}_0}\right) \\ \mathbb{G}_n \psi^\pi_{\boldsymbol{\gamma}_0} \end{pmatrix}$$

converges weakly in $\{\ell^\infty(\overline{\mathbb{R}})\}^2 \times \mathbb{R}$. It follows from Proposition 4.1 and Lemma B.3 that

$$\sqrt{n}\left(J_n(\boldsymbol{\eta}_n, \cdot) - J(\boldsymbol{\eta}_0, \cdot), K_n(\boldsymbol{\eta}_n, \cdot) - K(\boldsymbol{\eta}_0, \cdot), \pi_n - \pi_0\right),$$

converges weakly in $\{\ell^\infty(\overline{\mathbb{R}})\}^2 \times \mathbb{R}$. The desired result is finally a consequence of (11) and the functional delta method applied with the map $(J, K, \pi) \mapsto \{J - (1 - \pi)K\}/\pi$. $\square$

C Proof of Proposition 4.3

Proof. The assumptions of Proposition 4.1 being verified, we have that $\pi_n \xrightarrow{\text{a.s.}} \pi_0 \neq 0$. Then, as can be verified from (13), to show the desired result, it suffices to show that

$$\sup_{t \in \mathbb{R}} \left| \frac{1}{nh_n} \sum_{i=1}^n \kappa\left(\frac{t - Y_i + \alpha_n + \beta_n X_i}{h_n}\right) - \frac{(1 - \pi_0)}{n} \sum_{i=1}^n f^*(t + \alpha_n + \beta_n X_i) - \pi_0 f(t) \right| \xrightarrow{\text{a.s.}} 0.$$

The previous supremum is smaller than $I_n + (1 - \pi_0) II_n$, where

$$I_n = \sup_{t \in \mathbb{R}} \left| \frac{1}{nh_n} \sum_{i=1}^n \kappa\left(\frac{t - Y_i + \alpha_n + \beta_n X_i}{h_n}\right) - (1 - \pi_0) \int_{\mathbb{R}} f^*(t + \alpha_0 + \beta_0 x) f_X(x) \mathrm{d}x - \pi_0 f(t) \right|,$$

and

$$II_n = \sup_{t \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n f^*(t + \alpha_n + \beta_n X_i) - \int_{\mathbb{R}} f^*(t + \alpha_0 + \beta_0 x) f_X(x) \mathrm{d}x \right|.$$

Let us first show that $I_n \xrightarrow{\text{a.s.}} 0$. Consider the class $\mathcal{F}$ of measurable functions from $\mathbb{R}^2$ to $\mathbb{R}$ defined by

$$\mathcal{F} = \left\{(x, y) \mapsto \psi_{\boldsymbol{\eta},t,h}(x) = \kappa\left(\frac{t - y + \alpha + \beta x}{h}\right) : \boldsymbol{\eta} = (\alpha, \beta) \in \mathbb{R}^2, t \in \mathbb{R}, h \in (0, \infty)\right\},$$

and notice that

$$\mathbb{P}_n\psi_{\boldsymbol{\eta}_n,t,h_n} = \frac{1}{n}\sum_{i=1}^n \kappa\left(\frac{t - Y_i + \alpha_n + \beta_n X_i}{h_n}\right), \qquad t \in \mathbb{R},$$

where $\boldsymbol{\eta}_n = (\alpha_n, \beta_n)$. Then, $I_n \leq I_n' + I_n''$, where

$$I_n' = \frac{1}{h_n}\sup_{t\in\mathbb{R}}\left|\mathbb{P}_n\psi_{\boldsymbol{\eta}_n,t,h_n} - P\psi_{\boldsymbol{\eta}_n,t,h_n}\right| = \frac{1}{h_n\sqrt{n}}\sup_{t\in\mathbb{R}}\left|\mathbb{G}_n\psi_{\boldsymbol{\eta}_n,t,h_n}\right|, \tag{20}$$

and

$$I_n'' = \sup_{t\in\mathbb{R}}\left|\frac{1}{h_n}P\psi_{\boldsymbol{\eta}_n,t,h_n} - g(t)\right|,$$

with

$$g(t) = (1-\pi_0)\int_{\mathbb{R}} f^*(t + \alpha_0 + \beta_0 x) f_X(x)\mathrm{d}x + \pi_0 f(t), \qquad t \in \mathbb{R}.$$

Let us first deal with I_n''. From (3), notice that

$$g(t) = \int_{\mathbb{R}} f_{Y|X}(t + \alpha_0 + \beta_0 x|x) f_X(x)\mathrm{d}x, \qquad t \in \mathbb{R}.$$

Also, for any $t \in \mathbb{R}$,

$$P\psi_{\boldsymbol{\eta}_n,t,h_n} = \int_{\mathbb{R}}\left\{\int_{\mathbb{R}}\kappa\left(\frac{t - y + \alpha_n + \beta_n x}{h_n}\right) f_{Y|X}(y|x)\mathrm{d}y\right\} f_X(x)\mathrm{d}x,$$

which, using the change of variable $u = (t - y + \alpha_n + \beta_n x)/h_n$ in the inner integral, can be rewritten as

$$P\psi_{\boldsymbol{\eta}_n,t,h_n} = h_n\int_{\mathbb{R}}\left\{\int_{\mathbb{R}}\kappa(u) f_{Y|X}(t + \alpha_n + \beta_n x - u h_n|x)\mathrm{d}u\right\} f_X(x)\mathrm{d}x.$$

Since κ is a p.d.f. from Assumption A4 (ii), it follows that, for any $t \in \mathbb{R}$,

$$\frac{1}{h_n}P\psi_{\boldsymbol{\eta}_n,t,h_n} - g(t) =$$
$$\int_{\mathbb{R}}\left[\int_{\mathbb{R}}\kappa(u)\left\{f_{Y|X}(t + \alpha_n + \beta_n x - u h_n|x) - f_{Y|X}(t + \alpha_0 + \beta_0 x|x)\right\}\mathrm{d}u\right] f_X(x)\mathrm{d}x.$$

As $f_{Y|X}'(\cdot|x)$, the derivative of $f_{Y|X}(\cdot|x)$, exists for all $x \in \mathcal{X}$ under Assumption A3 (ii), the mean value theorem enables us to write

$$I_n'' \leq \left\{\sup_{t\in\mathbb{R}}(f^*)'(t) + \sup_{t\in\mathbb{R}} f'(t)\right\}\int_{\mathbb{R}}\left[\int_{\mathbb{R}}\kappa(u)\left\{|\alpha_n - \alpha_0| + |\beta_n - \beta_0||x| + |u|h_n\right\}\mathrm{d}u\right] f_X(x)\mathrm{d}x.$$

Hence,

$$I_n'' \leq \left\{\sup_{t\in\mathbb{R}}(f^*)'(t) + \sup_{t\in\mathbb{R}} f'(t)\right\}\left\{|\alpha_n - \alpha_0| + |\beta_n - \beta_0|\mathbb{E}(|X|) + h_n\int_{\mathbb{R}}|u|\kappa(u)\mathrm{d}u\right\},$$

which, from Assumptions A1 (i), A3 (ii), A4 (ii), and Proposition 4.1 (i), implies that $I_n'' \xrightarrow{\text{a.s.}} 0$.

Let us now show that $I_n' \xrightarrow{\text{a.s.}} 0$. Since κ has bounded variations from Assumption A4 (ii), it can be written as $\kappa_1 - \kappa_2$, where both κ_1 and κ_2 are bounded nondecreasing functions on $\mathbb{R}$. Without loss of generality, we shall assume that κ, κ_1 and κ_2 are bounded by 1. Then, for $j = 1, 2$, we define

$$\mathcal{F}_j = \left\{ (x, y) \mapsto \kappa_j \left(\frac{t - y + \alpha + \beta x}{h} \right) : (\alpha, \beta, t) \in \mathbb{R}^3, h \in (0, \infty) \right\}.$$

Proceeding as in Nolan and Pollard (1987, proof of Lemma 22), let us first show that $\mathcal{F}_j$ is a VC class for $j = 1, 2$. Let κ_j^- be the generalized inverse of κ_j defined by $\kappa_j^-(c) = \inf\{x \in \mathbb{R} : \kappa_j(x) \geq c\}$, $c \in \mathbb{R}$. We consider the partition $\{C_1, C_2\}$ of $\mathbb{R}$ defined by

$$\{x \in \mathbb{R} : \kappa_j(x) > c\} = \begin{cases} (\kappa_j^-(c), \infty) & \text{if} \quad c \in C_1, \\ [\kappa_j^-(c), \infty) & \text{if} \quad c \in C_2. \end{cases}$$

Given $(\alpha, \beta, t) \in \mathbb{R}^3$ and $h \in (0, \infty)$, the set

$$\left\{ (x, y, c) \in \mathbb{R}^3 : \kappa_j \left(\frac{t - y + \alpha + \beta x}{h} \right) > c \right\} \tag{21}$$

can therefore be written as the union of

$$\left\{ (x, y, c) \in \mathbb{R}^2 \times C_1 : t - y + \alpha + \beta x - h\kappa_j^-(c) > 0 \right\}$$

and

$$\left\{ (x, y, c) \in \mathbb{R}^2 \times C_2 : t - y + \alpha + \beta x - h\kappa_j^-(c) \geq 0 \right\}.$$

Now, let $f_{\alpha,\beta,t,h}(x, y, c) = t - y + \alpha + \beta x - h\kappa_j^-(c)$. The functions $f_{\alpha,\beta,t,h}$, with $(\alpha, \beta, t) \in \mathbb{R}^3$ and $h \in (0, \infty)$, span a finite-dimensional vector space. Hence, from Lemma 18 (ii) in Nolan and Pollard (1987), the collections of all sets $\{(x, y, c) \in \mathbb{R}^2 \times C_1 : f_{\alpha,\beta,t,h}(x, y, c) > 0\}$ and $\{(x, y, c) \in \mathbb{R}^2 \times C_2 : f_{\alpha,\beta,t,h}(x, y, c) \geq 0\}$ are VC classes. It follows that the collection of subgraphs of $\mathcal{F}_j$ defined by (21), and indexed by $(\alpha, \beta, t) \in \mathbb{R}^3$ and $h \in (0, \infty)$, is also VC, which implies that $\mathcal{F}_j$ is a VC class of functions.

Given a probability distribution Q on $\mathbb{R}^2$, recall that $L_2(Q)$ is the norm defined by $(Qf^2)^{1/2}$, with f a measurable function from $\mathbb{R}^2$ to $\mathbb{R}$. Given a class $\mathcal{G}$ of measurable functions from $\mathbb{R}^2$ to $\mathbb{R}$, the *covering number* $N(\varepsilon, \mathcal{G}, L_2(Q))$ is the minimal number of $L_2(Q)$-balls of radius $\varepsilon > 0$ needed to cover the set $\mathcal{G}$. From Lemma 16 in Nolan and Pollard (1987), since $\mathcal{F} = \mathcal{F}_1 - \mathcal{F}_2$, and since $\mathcal{F}_1$ and $\mathcal{F}_2$ have for envelope the constant function 1 on $\mathbb{R}^2$, we have

$$\sup_Q N(2\varepsilon, \mathcal{F}, L_2(Q)) \leq \sup_Q N(\varepsilon, \mathcal{F}_1, L_2(Q)) \times \sup_Q N(\varepsilon, \mathcal{F}_2, L_2(Q)),$$

for probability measures Q on $\mathbb{R}^2$. Using the fact that both $\mathcal{F}_1$ and $\mathcal{F}_2$ are VC classes of functions with constant envelope 1, from Theorem 2.6.7 in van der Vaart and Wellner (2000) (see also the discussion on the top of page 246), we obtain that there exist constants $u > 0$ and $v > 0$ that depend on $\mathcal{F}_1$ and $\mathcal{F}_2$ such that

$$\sup_Q N(\varepsilon, \mathcal{F}, L_2(Q)) \leq \left(\frac{u}{\varepsilon}\right)^v, \qquad \text{for every } 0 < \varepsilon < u.$$

Then, by Theorem 2.14.9 in van der Vaart and Wellner (2000), there exists constants $c_1 > 0$ and $c_2 > 0$ such that, for every $\varepsilon > 0$,

$$\Pr{}^*\left(\sup_{f \in \mathcal{F}} |\mathbb{G}_n f| > \varepsilon\right) \leq c_1 \varepsilon^{c_2} \exp(-2\varepsilon^2).$$

Starting from (20), we thus obtain that, for every $\varepsilon > 0$,

$$\begin{aligned} \Pr{}^*(I_n' > \varepsilon) = \Pr{}^*\left(\sup_{t \in \mathbb{R}} \left|\mathbb{G}_n \psi_{\boldsymbol{\eta}_n, t, h_n}\right| > \sqrt{n} h_n \varepsilon\right) \\ \leq \Pr{}^*\left(\sup_{f \in \mathcal{F}} |\mathbb{G}_n f| > \sqrt{n} h_n \varepsilon\right) \leq c_1 (\sqrt{n} h_n \varepsilon)^{c_2} \exp(-2n h_n^2 \varepsilon^2) = a_n. \end{aligned}$$

From Assumption A4 (i), it can be verified that $a_{n+1}/a_n \to 1$ and that $n(a_{n+1}/a_n - 1) \to -\infty$. It follows from Raabe's rule that the series with general term a_n converges. The Borel-Cantelli lemma enables us to conclude that $I_n' \xrightarrow{\text{a.s.}} 0$, and we therefore obtain that $I_n \xrightarrow{\text{a.s.}} 0$.

Since f^* has bounded variations from Assumption A4 (iii), one can proceed along the same lines to show that $II_n \xrightarrow{\text{a.s.}} 0$. $\square$

D Proof of Proposition 4.4

The proof of Proposition 4.4 is based on the following lemma.

Lemma D.1. *Let $\Theta \subset \mathbb{R}^p$ and $H_0 \subset \mathbb{R}^q$ for some integers $p, q > 0$, let $\mathcal{F} = \{f_{\theta,\zeta} : \theta \in \Theta, \zeta \in H_0\}$ be a class of measurable functions from $\mathbb{R}^2$ to $\mathbb{R}$, and let ζ_n be an estimator of $\zeta_0 \in H_0$ such that $\Pr(\zeta_n \in H_0) \to 1$. If $\mathcal{F}$ is P-Donsker and*

$$\sup_{\theta \in \Theta} P(f_{\theta,\zeta_n} - f_{\theta,\zeta_0})^2 \to_p 0,$$

then,

$$\sup_{\theta \in \Theta} |\mathbb{G}_n'(f_{\theta_n,\zeta} - f_{\theta,\zeta_0})| \to_p 0.$$

Proof. The result is the analogue of Theorem 2.1 of van der Vaart and Wellner (2007) in which $\mathbb{G}_n$ is replaced by $\mathbb{G}_n'$. The proof of Theorem 2.1 relies on the fact that $\mathbb{G}_n \rightsquigarrow \mathbb{G}$ in $\ell^\infty(\mathcal{F})$ and on the uniform continuity of the sample paths of the P-Brownian bridge $\mathbb{G}$; see van der Vaart (1998, proof of Theorem 19.26) and van der Vaart (2002). From the functional multiplier central limit theorem (see e.g. Kosorok, 2008, Theorem 10.1), we know that $(\mathbb{G}_n, \mathbb{G}'_n)$ converges weakly in $\{\ell^\infty(\mathcal{F})\}^2$ to $(\mathbb{G}, \mathbb{G}')$, where $\mathbb{G}'$ is an independent copy of the $\mathbb{G}$. The desired result therefore follows from a straightforward adaptation of the proof of Theorem 2.1 of van der Vaart and Wellner (2007). $\square$

Proof of Proposition 4.4. Since Assumptions A1 (ii) and A2 hold, we have from Lemma B.1 that $\mathcal{F}^J$, $\mathcal{F}^K$ and $\mathcal{F}^{\alpha,\beta,\pi}$ are P-Donsker. Furthermore, $\mathbb{E}(X)$ is finite from Assumption A1 (i), the function f is bounded from Assumption A3 (i), and so is the function $t \mapsto P(\psi^K_{t,\boldsymbol{\eta}_0} - \psi^J_{t,\boldsymbol{\eta}_0})$ from the definitions of J and K given in (8) and (9). Hence, from the functional multiplier central limit theorem (see e.g. Kosorok, 2008, Theorem 10.1) and the continuous mapping theorem, we obtain that

$$\left(t \mapsto \mathbb{G}_n \psi^F_{t,\boldsymbol{\gamma}_0}, t \mapsto \mathbb{G}'_n \psi^F_{t,\boldsymbol{\gamma}_0}\right) \rightsquigarrow \left(t \mapsto \mathbb{G} \psi^F_{t,\boldsymbol{\gamma}_0}, t \mapsto \mathbb{G}' \psi^F_{t,\boldsymbol{\gamma}_0}\right)$$

in $\{\ell^\infty(\overline{\mathbb{R}})\}^2$, where $\psi^F_{t,\boldsymbol{\gamma}_0}$ is defined in (12) and $t \mapsto \mathbb{G}'\psi^F_{t,\boldsymbol{\gamma}_0}$ is an independent copy of $t \mapsto \mathbb{G}\psi^F_{t,\boldsymbol{\gamma}_0}$. It remains to show that

$$\sup_{t\in\mathbb{R}} \left| \mathbb{G}'_n \left(\hat\psi^F_{t,\boldsymbol{\gamma}_n} - \psi^F_{t,\boldsymbol{\gamma}_0} \right) \right| \to_p 0.$$

From (12) and (14), for any $t \in \mathbb{R}$, we can write

$$\begin{aligned}
\left| \mathbb{G}'_n \left(\hat\psi^F_{t,\boldsymbol{\gamma}_n} - \psi^F_{t,\boldsymbol{\gamma}_0} \right) \right| &\le \left| \mathbb{G}'_n \left(\frac{1}{\pi_n} \psi^J_{t,\boldsymbol{\eta}_n} - \frac{1}{\pi_0} \psi^J_{t,\boldsymbol{\eta}_0} \right) \right| + \left| \mathbb{G}'_n \left(f_n(t) \hat\psi^\alpha_{\boldsymbol{\gamma}_n} - f(t) \psi^\alpha_{\boldsymbol{\gamma}_0} \right) \right| \\
&\quad + \left| \mathbb{G}'_n \left(f_n(t) \bar X \hat\psi^\beta_{\boldsymbol{\gamma}_n} - f(t)\mathbb{E}(X) \psi^\beta_{\boldsymbol{\gamma}_0} \right) \right| + \left| \mathbb{G}'_n \left(\frac{1-\pi_n}{\pi_n} \psi^K_{t,\boldsymbol{\eta}_n} - \frac{1-\pi_0}{\pi_0} \psi^K_{t,\boldsymbol{\eta}_0} \right) \right| \\
&\quad + \left| \mathbb{G}'_n \left(\frac{\mathbb{P}_n \psi^K_{t,\boldsymbol{\eta}_n} - \mathbb{P}_n \psi^J_{t,\boldsymbol{\eta}_n}}{\pi_n^2} \hat\psi^\pi_{\boldsymbol{\gamma}_n} - \frac{P\psi^K_{t,\boldsymbol{\eta}_0} - P\psi^J_{t,\boldsymbol{\eta}_0}}{\pi_0^2} \psi^\pi_{\boldsymbol{\gamma}_0} \right) \right|. \qquad (22)
\end{aligned}$$

The last absolute value on the right of the previous display is smaller than

$$\left| \frac{\mathbb{P}_n \psi^K_{t,\boldsymbol{\eta}_n} - \mathbb{P}_n \psi^J_{t,\boldsymbol{\eta}_n}}{\pi_n^2} - \frac{P\psi^K_{t,\boldsymbol{\eta}_0} - P\psi^J_{t,\boldsymbol{\eta}_0}}{\pi_0^2} \right| \left| \mathbb{G}'_n \psi^\pi_{\boldsymbol{\gamma}_0} \right| + \left| \frac{P\psi^K_{t,\boldsymbol{\eta}_0} - P\psi^J_{t,\boldsymbol{\eta}_0}}{\pi_0^2} \right| \left| \mathbb{G}'_n \left(\hat\psi^\pi_{\boldsymbol{\gamma}_n} - \psi^\pi_{\boldsymbol{\gamma}_0} \right) \right|. \qquad (23)$$

Now,

$$\begin{aligned}
\sup_{t\in\mathbb{R}} \left| \mathbb{P}_n \psi^K_{t,\boldsymbol{\eta}_n} - \mathbb{P}_n \psi^J_{t,\boldsymbol{\eta}_n} - P\psi^K_{t,\boldsymbol{\eta}_0} + P\psi^J_{t,\boldsymbol{\eta}_0} \right| &\le n^{-1/2} \sup_{t\in\mathbb{R}} \left| \mathbb{G}_n \left(\psi^K_{t,\boldsymbol{\eta}_n} - \psi^J_{t,\boldsymbol{\eta}_n} - \psi^K_{t,\boldsymbol{\eta}_0} + \psi^J_{t,\boldsymbol{\eta}_0} \right) \right| \\
&\quad + n^{-1/2} \sup_{t\in\mathbb{R}} \left| \mathbb{G}_n \left(\psi^K_{t,\boldsymbol{\eta}_0} - \psi^J_{t,\boldsymbol{\eta}_0} \right) \right| + \sup_{t\in\mathbb{R}} \left| P \left(\psi^K_{t,\boldsymbol{\eta}_n} - \psi^J_{t,\boldsymbol{\eta}_n} - \psi^K_{t,\boldsymbol{\eta}_0} + \psi^J_{t,\boldsymbol{\eta}_0} \right) \right|. \qquad (24)
\end{aligned}$$

Applying the mean value theorem as in the proof of Lemma B.2, we obtain that,

$$\sup_{t\in\mathbb{R}} \left| P \left(\psi^K_{t,\boldsymbol{\eta}} - \psi^J_{t,\boldsymbol{\eta}} - \psi^K_{t,\boldsymbol{\eta}_0} + \psi^J_{t,\boldsymbol{\eta}_0} \right) \right| \to 0 \qquad \text{as} \qquad \boldsymbol{\eta} \to \boldsymbol{\eta}_0,$$

which, combined with the fact that $\boldsymbol{\eta}_n \xrightarrow{\text{a.s.}} \boldsymbol{\eta}_0$ implies that the last term on the right of (24) converges to zero in probability. From Lemma B.2 and Theorem 2.1 of van der Vaart and Wellner (2007), we obtain that the first term on the right of (24) converges to zero in probability. The second term on the right of (24) converges to zero in probability because the classes $\mathcal{F}^J$ and $\mathcal{F}^K$ are P-Donsker. The convergence to zero in probability of the term on the left of (24) combined with the fact that $\pi_n \xrightarrow{\text{a.s.}} \pi_0$ and that $|\mathbb{G}'_n \psi^{\pi}_{\boldsymbol{\gamma}_0}|$ is bounded in probability implies that the first product in (23) converges to zero in probability uniformly in $t \in \mathbb{R}$. Furthermore, $\mathcal{F}^{\alpha,\beta,\pi}$ being P-Donsker, and since $P\|\Psi_{\boldsymbol{\gamma}_n}\Gamma_n^{-1}\dot{\varphi}_{\boldsymbol{\gamma}_n} - \Psi_{\boldsymbol{\gamma}_0}\Gamma_0^{-1}\dot{\varphi}_{\boldsymbol{\gamma}_0}\|^2 \to_p 0$ under Assumptions A1 (ii) and A2, we have from Lemma D.1 that $\mathbb{G}'_n(\hat{\psi}^{\pi}_{\boldsymbol{\gamma}_n} - \psi^{\pi}_{\boldsymbol{\gamma}_0}) \to_p 0$, which implies that the second product in (23) converges to zero in probability uniformly in $t \in \mathbb{R}$.

One can similarly show that the other terms on the right of (22) converge to zero in probability uniformly in $t \in \mathbb{R}$ using, among other arguments, the fact that, from Lemma D.1,

$$\sup_{t\in\mathbb{R}} \left|\mathbb{G}'_n\left(\psi^J_{t,\boldsymbol{\eta}_n} - \psi^J_{t,\boldsymbol{\eta}_0}\right)\right|, \sup_{t\in\mathbb{R}} \left|\mathbb{G}'_n\left(\psi^K_{t,\boldsymbol{\eta}_n} - \psi^K_{t,\boldsymbol{\eta}_0}\right)\right|, \mathbb{G}'_n(\hat{\psi}^{\alpha}_{\boldsymbol{\gamma}_n} - \psi^{\alpha}_{\boldsymbol{\gamma}_0}), \text{ and } \mathbb{G}'_n(\hat{\psi}^{\beta}_{\boldsymbol{\gamma}_n} - \psi^{\beta}_{\boldsymbol{\gamma}_0})$$

converge to zero in probability, as well as $\sup_{t\in\mathbb{R}} |f_n(t) - f(t)|$ since the assumptions of Proposition 4.3 are satisfied. □

References

J.A. Anderson. Multivariate logistic compounds. *Biometrika*, pages 17–26, 1979.

E.A. Cohen. *Inharmonic Tone Perception*. PhD thesis, Stanford University, 1980.

R.D. De Veaux. Mixtures of linear regressions. *Computational Statistics and Data Analysis*, 8:227–245, 1989.

T. Duong. *ks: Kernel smoothing*, 2012. URL `http://CRAN.R-project.org/package=ks`. R package version 1.8.8.

I.K. Glad, N.L. Hjort, and N.G. Ushakov. Correction of density estimators that are not densities. *Scandinavian Journal of Statistics*, 30:415–427, 2003.

B. Grün and F. Leisch. Fitting finite mixtures of linear regression models with varying ajd fixed effects in `R`. In A. Rizzi and M. Vichi, editors, *Compstat 2006, Proceedings in Computational Statistics*, pages 853–860. Physica Verlag, Heidelberg, Germany, 2006.

P. Hall and X-H. Zhou. Nonparametric estimation of component distributions in a multivariate mixture. *Annals of Statistics*, 31:201–224, 2003.

D.S. Hawkins, D.M. Allen, and A.J. Stomber. Determining the number of components in mixtures of linear models. *Computational Statistics and Data Analysis*, 38:15–48, 2001.

D.R. Hunter and D.S. Young. Semiparametric mixtures of regressions. *Journal of Nonparametric Statistics*, pages 19–38, 2012.

M. Hurn, A. Justel, and C.P. Robert. Estimating mixtures of regressions. *Journal of Computational and Graphical Statistiscs*, 12:1–25, 2003.

P.N. Jones and G.J. McLachlan. Fitting finite mixture models in a regression context. *Australian Journal of Statistics*, 34:233–240, 1992.

M.R. Kosorok. *Introduction to empirical processes and semiparametric inference*. Springer, New York, 2008.

F. Leisch. Flexmix: A general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software*, 2004. http://www.jstatsoft. org/v11/i08/.

D.H-Y. Leung and J. Qin. Semi-parametric inference in a bivariate (multivariate) mixture model. *Statistica Sinica*, 16:153–163, 2006.

M-L. Martin-Magniette, T. Mary-Huard, C. Bérard, and S. Robin. ChIPmix: mixture model of regressions for two-color ChIP-chip analysis. *Bioinformatics*, 24:181–186, 2008.

D. Nolan and D. Pollard. *U*-processes: Rates of convergence. *Annals of Statistics*, 15: 780–799, 1987.

R. Quandt and J. Ramsey. Estimating mixtures of normal distributions and switching regression. *Journal of the American Statistical Association*, 73:730–738, 1978.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

N. Städler, P. Bühlmann, and S. van de Geer. ℓ_1-penalization for mixture of regression models. *Test*, 19:209–256, 2010.

T.R. Turner. Estimating the propagation rate of a viral infection of potato plants via mixtures of regressions. *Applied Statistics*, 49:371–384, 2000.

A.W. van der Vaart. *Asymptotic statistics*. Cambridge University Press, 1998.

A.W. van der Vaart. Semiparametric statistics. In *École d'été de Saint-Flour 1999*, pages 331–457. Springer, New-York, 2002.

A.W. van der Vaart and J.A. Wellner. *Weak convergence and empirical processes*. Springer, New York, 2000. Second edition.

A.W. van der Vaart and J.A. Wellner. Empirical processes indexed by estimated functions. In *Asymptotics: Particles, Processes and Inverse Problems*, pages 234–252. Institute of Mathematical Statistics, 2007.

P. Vandekerkhove. Estimation of a semiparametric mixture of regressions model. *Journal of Nonparametric Statistics*, page in press, 2012.

M.P. Wand and M.C. Jones. Multivariate plugin bandwidth selection. *Computational Statistics*, 9:97–116, 1994.

H. Zhu and H. Zhang. Hypothesis testing in mixture regression models. *Journal of the Royal Statistical Society Series B*, 66:3–16, 2004.

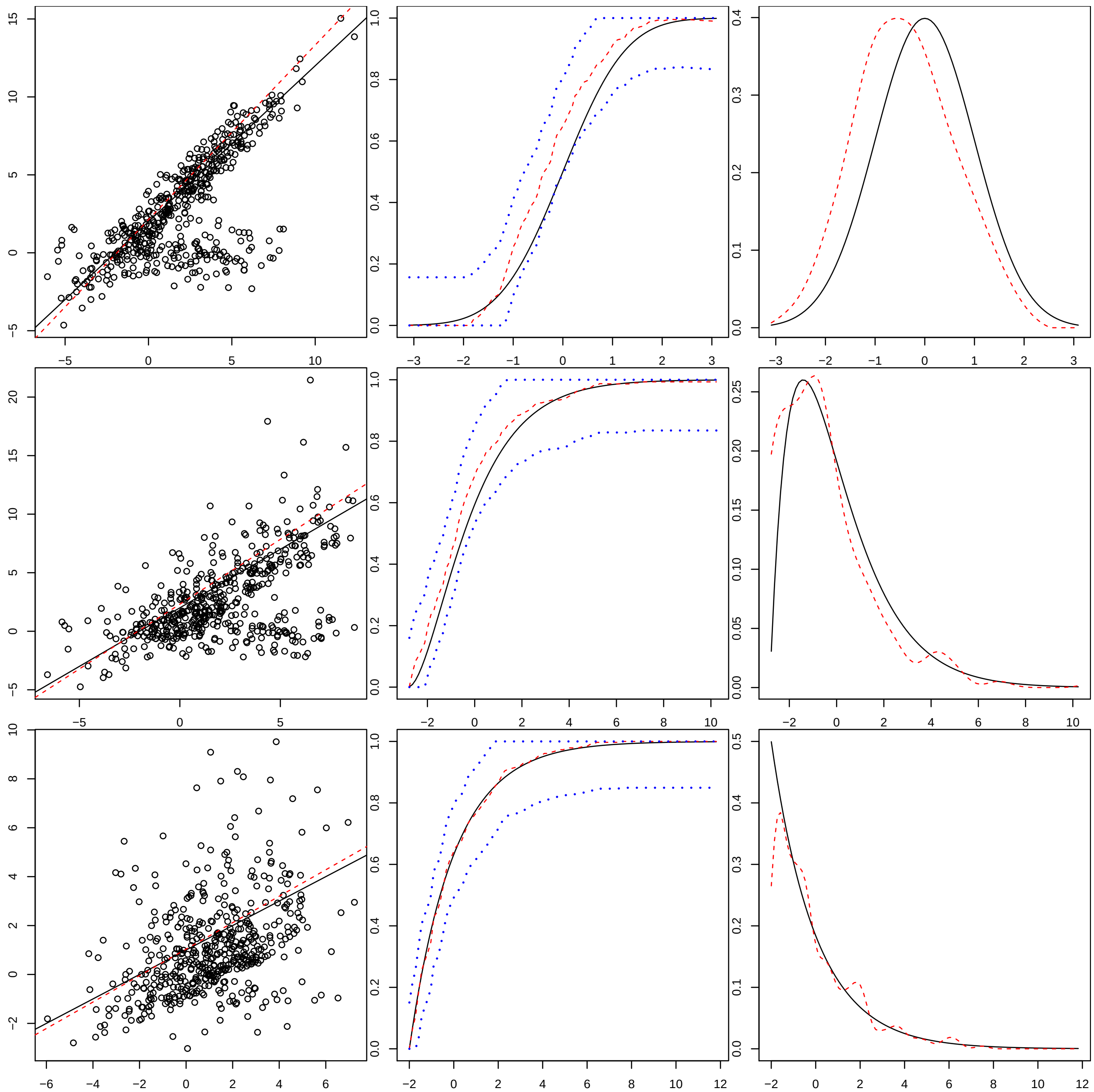

Figure 1: First column, from top to bottom: datasets generated from WOn, MOg and SOe, respectively, with $n = 500$ and $\pi_0 = 0.7$; the solid (resp. dashed) lines represent the true (resp. estimated) regression lines. Second column, from top to bottom: for WOn, MOg and SOe, respectively, the true c.d.f. F of ε (solid line) and its estimate F_n (dashed line) defined in (11). The dotted lines represent approximate confidence bands of level 0.95 for F computed as explained in Subsection 4.3 with $N = 10,000$. Third column, from top to bottom: for WOn, MOg and SOe, respectively, the true p.d.f. f of ε (solid line) and its estimate f_n defined in (13) (dashed line).

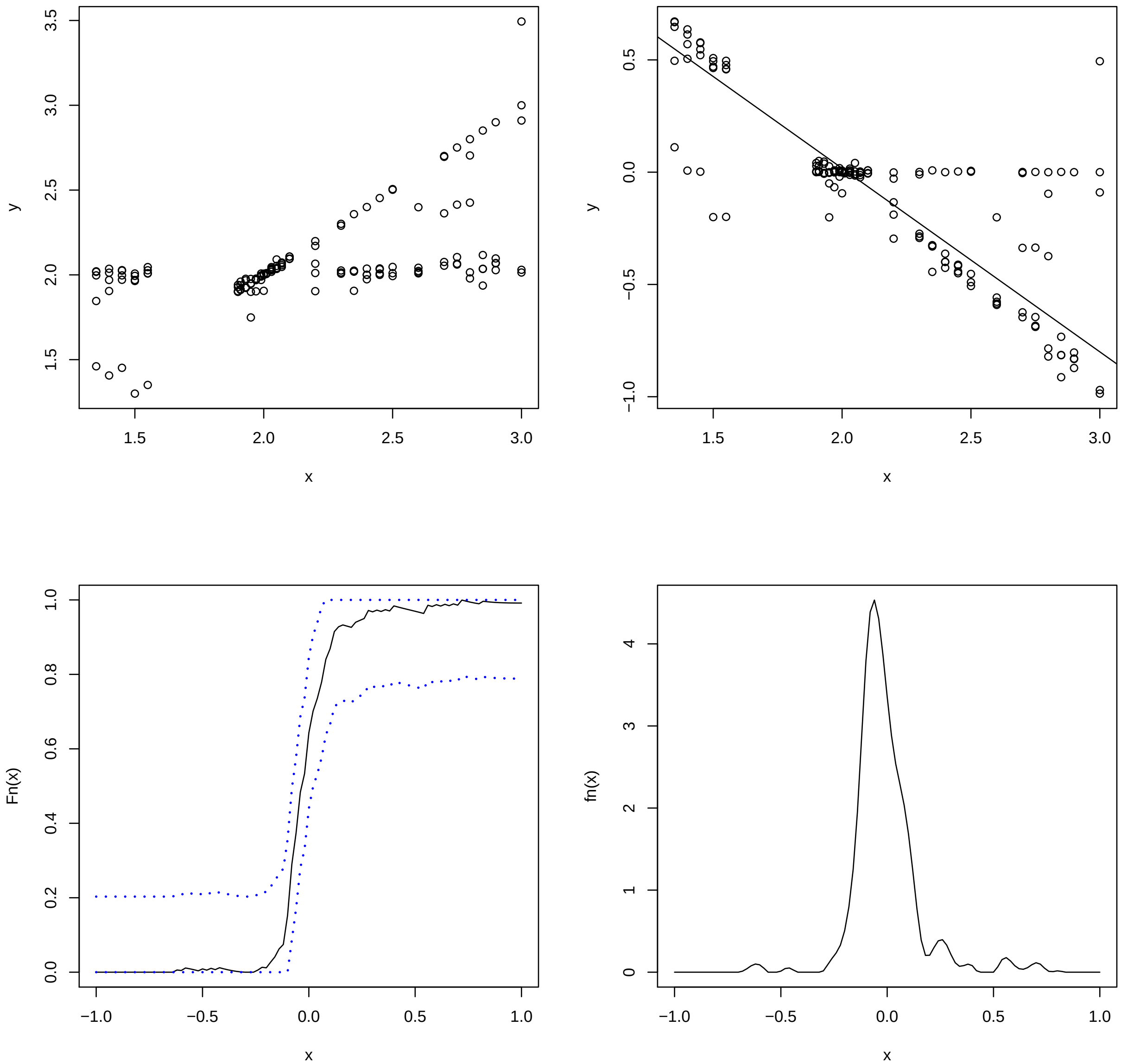

Figure 2: Upper left plot: the original tone data. Upper right plot: the transformed data; the solid line represents the estimated regression line. Lower left plot: the estimate $(F_n \vee 0) \wedge 1$ (solid line) of the unknown c.d.f. F of ε as well as well as an approximate confidence band (dotted lines) of level 0.95 for F computed as explained in Subsection 4.3 with $N = 10,000$. Lower right plot: the estimate $f_n \vee 0$ of the unknown p.d.f. f of ε.

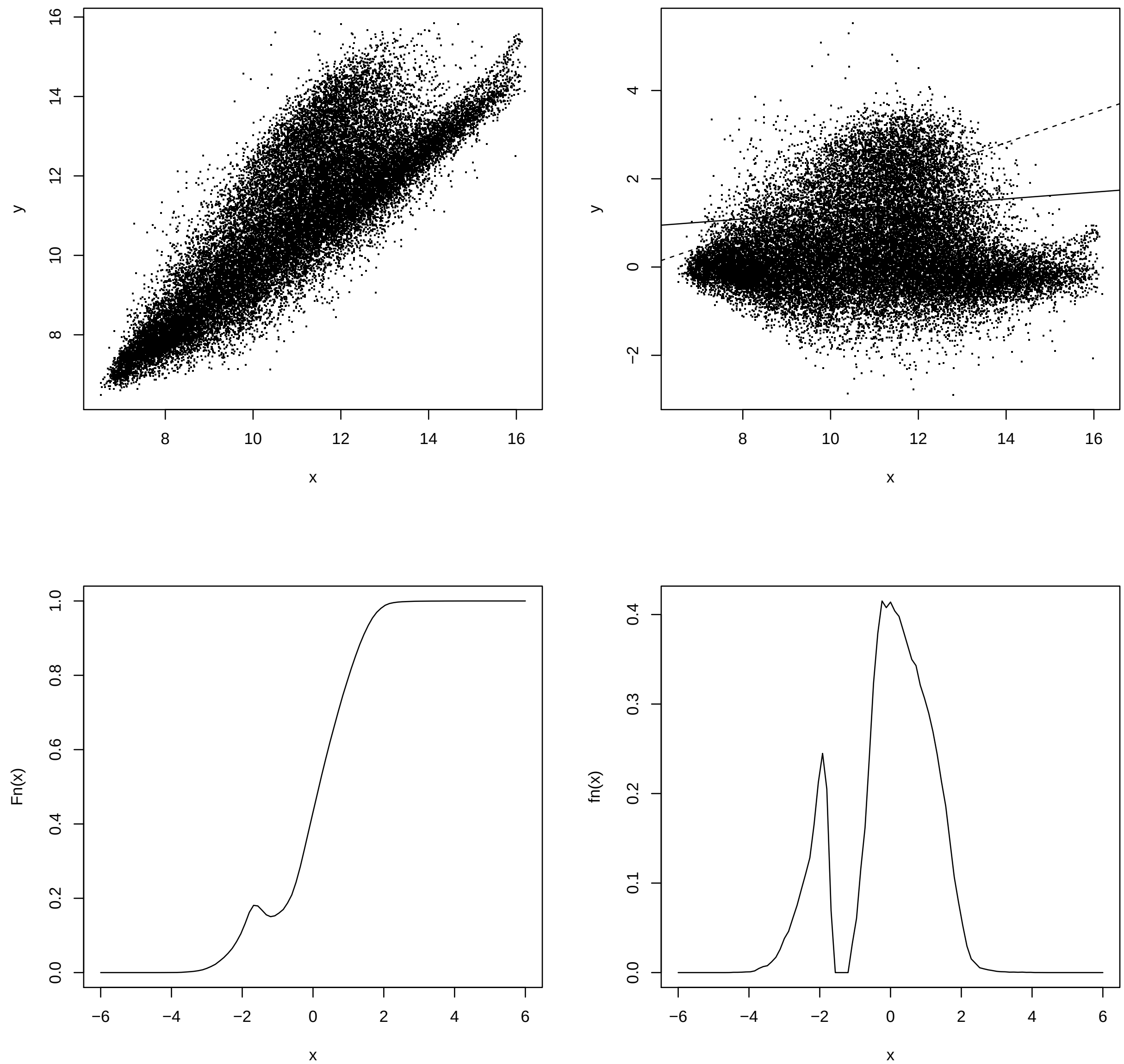

Figure 3: Upper left plot: the original ChIPmix data analyzed by Martin-Magniette et al. (2008). Upper right plot: the ChIPmix data transformed as in Vandekerkhove (2012); the solid line represents the regression line estimated by the method in this work, while the dashed line is the regression line estimated by Martin-Magniette et al. (2008). Lower left plot: the estimate $(F_n \vee 0) \wedge 1$ of the unknown c.d.f. F of ε. Lower right plot: the estimate $f_n \vee 0$ of the unknown p.d.f. f of ε.

Table 1: For $M = 1000$ random samples generated under scenarios WOn, MOn and SOn, number m of samples out of M for which $\pi_n \notin (0, 1]$, as well as estimated bias and standard deviation of α_n, β_n, π_n, $F_n\{F^{-1}(0.1)\}$, $F_n\{F^{-1}(0.5)\}$ and $F_n\{F^{-1}(0.9)\}$ computed from the $M - m$ valid estimates.

				α_n		β_n		π_n		$F_n\{F^{-1}(0.1)\}$		$F_n\{F^{-1}(0.5)\}$		$F_n\{F^{-1}(0.9)\}$	
Scenario	π_0	n	m	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd
WOn	0.4	100	15	-0.049	0.689	-0.008	0.340	0.038	0.139	0.140	0.144	0.051	0.160	-0.070	0.119
		300	0	-0.032	0.392	-0.008	0.220	0.015	0.079	0.078	0.092	0.022	0.129	-0.048	0.098
		1000	0	-0.010	0.213	-0.006	0.125	0.005	0.040	0.030	0.044	0.007	0.092	-0.022	0.062
		5000	0	-0.005	0.096	-0.002	0.058	0.002	0.019	0.008	0.014	0.000	0.049	-0.007	0.030
	0.7	100	38	0.015	0.357	0.019	0.181	0.003	0.101	0.060	0.080	0.035	0.122	-0.024	0.084
		300	2	-0.011	0.205	-0.002	0.118	0.010	0.065	0.025	0.039	0.009	0.086	-0.018	0.061
		1000	0	-0.002	0.112	0.000	0.067	0.001	0.036	0.009	0.018	0.003	0.054	-0.006	0.034
		5000	0	-0.003	0.050	-0.001	0.030	0.001	0.017	0.002	0.006	-0.001	0.027	-0.002	0.015
MOn	0.4	100	34	-0.095	0.827	-0.020	0.376	0.056	0.153	0.054	0.088	0.039	0.099	-0.022	0.068
		300	0	-0.008	0.456	-0.005	0.237	0.018	0.089	0.026	0.054	0.020	0.068	-0.011	0.049
		1000	0	-0.014	0.264	-0.003	0.135	0.006	0.045	0.010	0.030	0.006	0.044	-0.005	0.030
		5000	0	-0.004	0.115	-0.004	0.061	0.002	0.019	0.002	0.013	0.001	0.020	-0.002	0.014
	0.7	100	64	-0.008	0.473	0.020	0.224	0.008	0.119	0.018	0.051	0.023	0.074	-0.005	0.048
		300	4	-0.014	0.274	-0.005	0.147	0.012	0.082	0.011	0.031	0.006	0.046	-0.005	0.034
		1000	0	-0.007	0.155	-0.002	0.084	0.005	0.046	0.004	0.018	0.002	0.027	-0.002	0.020
		5000	0	-0.004	0.069	-0.001	0.038	0.001	0.021	0.001	0.007	0.000	0.012	-0.001	0.009
SOn	0.4	100	251	0.666	3.963	0.110	0.393	0.013	0.222	0.006	0.153	0.057	0.122	0.019	0.053
		300	90	0.042	0.522	0.022	0.230	0.048	0.183	-0.018	0.047	0.021	0.051	0.007	0.028
		1000	2	-0.009	0.279	0.003	0.139	0.026	0.116	-0.012	0.025	0.010	0.028	0.003	0.015
		5000	0	0.005	0.122	0.002	0.063	0.003	0.046	-0.002	0.011	0.002	0.012	0.001	0.007
	0.7	100	310	0.199	0.627	0.112	0.222	-0.057	0.192	-0.016	0.051	0.021	0.067	0.014	0.036
		300	166	0.090	0.346	0.040	0.149	-0.019	0.152	-0.011	0.028	0.008	0.033	0.006	0.020
		1000	36	0.005	0.177	0.006	0.090	0.008	0.106	-0.004	0.014	0.003	0.016	0.002	0.010
		5000	0	0.000	0.084	0.000	0.043	0.005	0.053	-0.001	0.006	0.001	0.007	0.000	0.005

Table 2: For $M = 1000$ random samples generated under scenarios WOg, MOg and SOg, number m of samples out of M for which $\pi_n \notin (0, 1]$, as well as estimated bias and standard deviation of α_n, β_n, π_n, $F_n\{F^{-1}(0.1)\}$, $F_n\{F^{-1}(0.5)\}$ and $F_n\{F^{-1}(0.9)\}$ computed from the $M - m$ valid estimates.

				α_n		β_n		π_n		$F_n\{F^{-1}(0.1)\}$		$F_n\{F^{-1}(0.5)\}$		$F_n\{F^{-1}(0.9)\}$	
Scenario	π_0	n	m	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd
WOg	0.4	100	21	-0.083	0.651	-0.022	0.342	0.044	0.134	0.186	0.167	0.004	0.164	-0.065	0.108
		300	0	-0.053	0.381	-0.007	0.225	0.018	0.082	0.119	0.127	-0.008	0.134	-0.035	0.083
		1000	0	-0.007	0.208	-0.003	0.128	0.005	0.040	0.058	0.087	-0.011	0.103	-0.012	0.043
		5000	0	-0.004	0.094	-0.002	0.056	0.002	0.017	0.016	0.041	-0.006	0.055	-0.003	0.018
	0.7	100	36	-0.014	0.360	-0.009	0.186	0.018	0.106	0.098	0.115	-0.008	0.132	-0.024	0.072
		300	4	-0.009	0.211	-0.005	0.119	0.008	0.069	0.056	0.080	-0.010	0.100	-0.013	0.047
		1000	0	-0.004	0.117	-0.000	0.069	0.002	0.038	0.025	0.050	-0.005	0.068	-0.003	0.024
		5000	0	-0.002	0.051	-0.002	0.031	0.001	0.017	0.004	0.023	-0.003	0.031	-0.001	0.010
MOg	0.4	100	45	-0.067	0.846	0.002	0.400	0.047	0.156	0.106	0.122	0.008	0.112	-0.008	0.056
		300	0	-0.049	0.458	-0.015	0.249	0.024	0.095	0.061	0.079	-0.001	0.079	-0.006	0.035
		1000	0	-0.025	0.248	-0.012	0.141	0.008	0.045	0.024	0.044	-0.008	0.052	-0.003	0.020
		5000	0	-0.006	0.115	-0.002	0.064	0.002	0.020	0.006	0.019	-0.002	0.026	-0.000	0.009
	0.7	100	69	-0.011	0.511	0.007	0.222	0.018	0.124	0.049	0.081	-0.001	0.084	0.000	0.037
		300	7	-0.031	0.299	-0.004	0.153	0.016	0.089	0.029	0.049	-0.005	0.059	-0.002	0.023
		1000	0	-0.008	0.163	-0.003	0.087	0.006	0.049	0.011	0.027	-0.003	0.036	-0.001	0.012
		5000	0	0.002	0.071	0.001	0.040	0.000	0.022	0.003	0.011	-0.000	0.017	0.000	0.006
SOg	0.4	100	305	1.339	12.672	0.155	0.455	0.012	0.224	0.062	0.190	0.024	0.138	0.021	0.049
		300	145	0.076	0.619	0.055	0.274	0.041	0.182	0.018	0.087	0.001	0.060	0.010	0.024
		1000	21	-0.011	0.314	-0.000	0.168	0.035	0.132	0.005	0.042	-0.000	0.032	0.003	0.013
		5000	0	-0.004	0.152	-0.000	0.079	0.011	0.062	0.002	0.018	-0.000	0.014	0.001	0.006
	0.7	100	386	1.222	22.682	0.169	0.326	-0.085	0.207	0.043	0.117	0.020	0.079	0.009	0.036
		300	244	0.101	0.379	0.069	0.189	-0.028	0.167	0.017	0.051	0.005	0.037	0.003	0.017
		1000	75	0.021	0.206	0.018	0.117	0.003	0.126	0.005	0.028	0.001	0.021	0.002	0.010
		5000	0	-0.003	0.100	-0.000	0.055	0.007	0.067	0.001	0.012	0.000	0.009	0.000	0.004

Table 3: For $M = 1000$ random samples generated under scenarios WOe, MOe and SOe, number m of samples out of M for which $\pi_n \notin (0, 1]$, as well as estimated bias and standard deviation of α_n, β_n, π_n, $F_n\{F^{-1}(0.1)\}$, $F_n\{F^{-1}(0.5)\}$ and $F_n\{F^{-1}(0.9)\}$ computed from the $M - m$ valid estimates.

				α_n		β_n		π_n		$F_n\{F^{-1}(0.1)\}$		$F_n\{F^{-1}(0.5)\}$		$F_n\{F^{-1}(0.9)\}$	
Scenario	π_0	n	m	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd	bias	sd
WOe	0.4	100	26	-0.040	0.715	-0.027	0.336	0.045	0.138	0.224	0.185	-0.008	0.179	-0.060	0.106
		300	0	-0.017	0.380	-0.005	0.218	0.013	0.074	0.154	0.152	-0.021	0.151	-0.031	0.077
		1000	0	-0.009	0.215	-0.003	0.125	0.004	0.040	0.084	0.115	-0.025	0.118	-0.011	0.041
		5000	0	-0.003	0.092	0.001	0.055	0.001	0.017	0.028	0.073	-0.010	0.066	-0.002	0.015
	0.7	100	47	0.000	0.372	0.007	0.189	0.013	0.108	0.145	0.149	-0.017	0.149	-0.021	0.071
		300	1	-0.017	0.203	-0.001	0.126	0.010	0.071	0.085	0.113	-0.021	0.116	-0.011	0.046
		1000	0	-0.006	0.111	-0.004	0.070	0.003	0.037	0.036	0.079	-0.017	0.079	-0.004	0.022
		5000	0	-0.002	0.051	0.000	0.031	0.001	0.017	0.009	0.049	-0.004	0.039	-0.000	0.009
MOe	0.4	100	44	-0.020	1.104	-0.005	0.390	0.047	0.153	0.148	0.146	-0.008	0.128	-0.011	0.052
		300	0	-0.040	0.463	-0.005	0.259	0.019	0.090	0.092	0.109	-0.017	0.097	-0.005	0.034
		1000	0	-0.012	0.255	-0.005	0.146	0.007	0.046	0.043	0.073	-0.013	0.067	-0.001	0.019
		5000	0	-0.005	0.115	-0.003	0.065	0.002	0.021	0.010	0.042	-0.004	0.034	-0.001	0.008
	0.7	100	82	-0.021	0.498	0.014	0.242	0.015	0.127	0.081	0.120	-0.018	0.100	-0.000	0.036
		300	4	-0.012	0.289	-0.002	0.155	0.012	0.086	0.048	0.082	-0.013	0.073	-0.001	0.022
		1000	0	-0.002	0.162	-0.001	0.090	0.004	0.050	0.022	0.057	-0.006	0.048	-0.001	0.012
		5000	0	-0.002	0.069	-0.002	0.040	0.001	0.022	0.002	0.030	-0.002	0.021	-0.000	0.006
SOe	0.4	100	325	0.972	7.133	0.191	0.533	0.008	0.220	0.104	0.205	-0.000	0.146	0.015	0.053
		300	194	0.049	0.600	0.044	0.276	0.051	0.192	0.047	0.109	-0.013	0.074	0.007	0.027
		1000	36	-0.014	0.342	0.005	0.177	0.045	0.147	0.029	0.074	-0.011	0.050	0.004	0.015
		5000	0	-0.001	0.160	0.002	0.087	0.009	0.066	0.010	0.042	-0.002	0.025	0.001	0.007
	0.7	100	399	0.432	1.880	0.213	0.437	-0.097	0.211	0.090	0.155	0.016	0.096	0.006	0.036
		300	299	0.133	0.398	0.091	0.213	-0.043	0.170	0.048	0.094	0.007	0.054	0.001	0.018
		1000	97	0.031	0.230	0.019	0.121	0.004	0.135	0.021	0.061	0.000	0.034	0.001	0.010
		5000	1	-0.004	0.110	-0.001	0.061	0.011	0.077	0.004	0.031	-0.001	0.016	0.001	0.005

Table 4: For $M = 1000$ random samples generated under scenarios WOn, MOg and SOe, number m of samples out of M for which $\pi_n \notin (0, 1]$, and, for each of the estimators α_n, β_n, π_n, $F_n\{F^{-1}(0.1)\}$, $F_n\{F^{-1}(0.5)\}$ and $F_n\{F^{-1}(0.9)\}$, standard deviation of the $M - m$ valid estimates times $\sqrt{n}$, and mean of the estimated standard errors times $\sqrt{n}$. The quantities t_1, t_2 and t_3 in the table are equal to $F^{-1}(0.1)$, $F^{-1}(0.5)$ and $F^{-1}(0.9)$, respectively.

				α_n		β_n		π_n		$F_n(t_1)$		$F_n(t_2)$		$F_n(t_3)$	
Scenario	π_0	n	m	sd	$\overline{\text{se}}$	sd	$\overline{\text{se}}$	sd	$\overline{\text{se}}$	sd	$\overline{\text{se}}$	sd	$\overline{\text{se}}$	sd	$\overline{\text{se}}$
WOn	0.4	100	16	6.66	6.67	3.51	2.92	1.37	1.23	1.43	1.15	1.57	1.36	1.18	1.11
		300	0	7.10	6.49	3.88	3.43	1.42	1.23	1.55	1.18	2.22	1.90	1.72	1.50
		1000	0	6.63	6.56	4.09	3.79	1.30	1.22	1.46	1.09	2.88	2.62	1.97	1.81
		5000	0	6.42	6.61	4.00	3.92	1.19	1.24	0.95	0.86	3.31	3.23	1.88	1.93
		25000	0	6.74	6.62	3.98	3.96	1.25	1.24	0.78	0.75	3.55	3.44	1.94	1.92
	0.7	100	33	3.49	3.50	1.86	1.61	1.04	1.05	0.73	0.60	1.16	1.02	0.87	0.75
		300	2	3.56	3.54	2.08	1.89	1.19	1.12	0.71	0.56	1.49	1.34	1.07	0.93
		1000	0	3.77	3.58	2.17	2.08	1.23	1.17	0.56	0.50	1.82	1.65	1.17	1.05
		5000	0	3.60	3.63	2.16	2.18	1.18	1.20	0.45	0.43	1.89	1.88	1.08	1.09
		25000	0	3.60	3.61	2.12	2.17	1.18	1.19	0.41	0.41	1.94	1.92	1.04	1.07
MOg	0.4	100	57	7.96	7.91	3.92	3.33	1.53	1.46	1.15	1.03	1.11	1.08	0.54	0.62
		300	2	7.99	7.69	4.41	3.93	1.60	1.39	1.43	1.09	1.38	1.32	0.61	0.64
		1000	0	8.37	7.83	4.64	4.34	1.50	1.40	1.46	1.10	1.74	1.63	0.64	0.65
		5000	0	8.39	8.04	4.69	4.54	1.52	1.43	1.38	1.13	1.96	1.86	0.65	0.64
		25000	0	8.30	8.04	4.57	4.58	1.52	1.44	1.28	1.19	1.96	1.91	0.65	0.64
	0.7	100	66	4.55	4.70	2.47	2.07	1.27	1.26	0.86	0.65	0.82	0.77	0.37	0.39
		300	8	5.06	4.80	2.71	2.42	1.51	1.40	0.89	0.70	1.03	0.95	0.41	0.41
		1000	0	5.05	4.95	2.73	2.64	1.57	1.48	0.86	0.70	1.15	1.10	0.43	0.42
		5000	0	5.00	5.01	2.72	2.73	1.55	1.52	0.79	0.73	1.17	1.17	0.41	0.42
		25000	0	4.93	5.03	2.71	2.76	1.52	1.53	0.79	0.78	1.19	1.19	0.42	0.42
SOe	0.4	100	294	76.74	60.97	6.19	4.65	2.24	3.59	1.94	2.30	1.36	1.94	0.51	0.80
		300	171	11.91	10.92	5.13	4.92	3.40	4.35	2.13	1.64	1.40	1.55	0.46	0.60
		1000	31	11.20	10.24	6.05	5.52	4.65	4.65	2.47	1.79	1.62	1.58	0.49	0.53
		5000	0	11.47	10.87	6.17	5.93	4.64	4.38	2.91	2.47	1.70	1.68	0.48	0.48
		25000	0	10.96	11.23	6.06	6.16	4.27	4.37	3.68	3.49	1.64	1.72	0.46	0.47
	0.7	100	410	8.91	8.82	3.37	3.43	2.06	3.00	1.48	1.19	0.87	1.11	0.36	0.44
		300	262	7.58	7.51	4.07	4.00	3.06	4.02	1.75	1.36	0.96	1.13	0.33	0.39
		1000	121	7.41	7.55	4.09	4.23	4.44	5.04	1.92	1.54	1.07	1.19	0.31	0.36
		5000	1	8.06	7.83	4.38	4.35	5.58	5.43	2.33	2.11	1.20	1.19	0.34	0.34
		25000	0	8.00	8.00	4.36	4.45	5.44	5.50	2.80	2.76	1.22	1.22	0.33	0.34

Table 5: For $M = 1000$ random samples generated under each of the nine scenarios considered in Section 5, number m of samples out of M for which $\pi_n \notin (0, 1]$, and proportion p out of the $M - m$ remaining samples for which F_n is not in the approximate confidence band computed as explained in Subsection 4.3.

Generic			$\varepsilon \sim$ Normal		$\varepsilon \sim$ Gamma		$\varepsilon \sim$ Exp	
scenario	π_0	n	m	p	m	p	m	p
WO	0.4	100	22	0.306	27	0.362	24	0.444
		300	0	0.238	0	0.251	2	0.334
		1000	0	0.126	0	0.182	0	0.226
		5000	0	0.082	0	0.080	0	0.133
		25000	0	0.064	0	0.055	0	0.092
	0.7	100	32	0.169	32	0.195	24	0.290
		300	2	0.138	5	0.160	3	0.231
		1000	0	0.092	0	0.108	0	0.168
		5000	0	0.073	0	0.074	0	0.090
		25000	0	0.056	0	0.041	0	0.081
MO	0.4	100	45	0.088	42	0.177	48	0.334
		300	0	0.114	2	0.205	1	0.296
		1000	0	0.103	0	0.127	0	0.207
		5000	0	0.073	0	0.095	0	0.126
		25000	0	0.050	0	0.073	0	0.085
	0.7	100	76	0.088	60	0.117	67	0.247
		300	7	0.102	13	0.146	12	0.215
		1000	0	0.084	0	0.082	0	0.140
		5000	0	0.054	0	0.067	0	0.096
		25000	0	0.049	0	0.065	0	0.070
SO	0.4	100	259	0.003	327	0.030	316	0.072
		300	103	0.006	128	0.057	182	0.117
		1000	4	0.027	14	0.067	29	0.142
		5000	0	0.029	0	0.077	0	0.123
		25000	0	0.042	0	0.045	0	0.087
	0.7	100	328	0.001	413	0.036	405	0.099
		300	166	0.005	249	0.037	280	0.094
		1000	32	0.028	91	0.043	119	0.083
		5000	0	0.036	2	0.062	2	0.088
		25000	0	0.044	0	0.061	0	0.071