Inference on Mixtures Under Tail Restrictions

Marc Henry, Koen Jochmans, Bernard Salanié

Department of Economics

Discussion paper 2014-01

INFERENCE ON MIXTURES UNDER TAIL RESTRICTIONS

Marc Henry
Koen Jochmans
Bernard Salanié

Sciences Po Economics Discussion Papers

INFERENCE ON MIXTURES UNDER TAIL RESTRICTIONS

By Marc Henry, Koen Jochmans, and Bernard Salanié

Penn State, Sciences Po, and Columbia

December 31, 2013

Two-component mixtures are nonparametrically identified under tail-dominance conditions on the component distributions if a source of variation is available that affects the mixing proportions but not the component distributions. We motivate these restrictions through several examples. One interesting example is a location model where the location parameter is subject to classical measurement error. The identification analysis suggests very simple closed-form estimators of the component distributions and mixing proportions based on ratios of intermediate quantiles. We derive their asymptotic properties using results on tail empirical processes, and we provide simulation evidence on their finite-sample performance.

1. Introduction. Although finite mixtures are widely used, the dominant mode of inference in such models is parametric; see, e.g., McLachlan and Peel [2000]. Recent work has established sufficient conditions for (nonparametric) identification of multivariate mixtures; see, notably, Allman, Matias and Rhodes [2009], Bonhomme, Jochmans and Robin [2013], Hall and Zhou [2003], and Kasahara and Shimotsu [2009]. Unfortunately, fewer identifying restrictions are available when the model of interest is univariate. Several contributions have established identification of mixtures of location models assuming that the error distribution is symmetric (see Bordes, Mottelet and Vandekerkhove [2006], Butucea and Vandekerkhove [2011], and Hunter, Wang and Hettmansperger [2007]). Here, we provide conditions that identify univariate component distributions and associated mixing proportions, and we propose estimators and analyze their properties. As in most of the recent work (see, e.g., Bordes, Mottelet and Vandekerkhove [2006] and Hall and Zhou [2003]), our analysis here is confined to mixtures of two components, which is arguably the dominant case of interest.

The models we consider are characterized by an exclusion restriction. That is, we assume the existence of a source of variation that shifts the mixing proportions but leaves the component distributions unchanged. Such an exclusion restriction is natural in several important applications, such as in a measurement-error model or a hidden Markov model. It is also an implication of the fundamental conditional-independence restriction underlying the results of Allman, Matias and Rhodes [2009] and Hall and Zhou [2003]. In Henry, Kitamura and Salanié [2013] it was shown that, given such an exclusion restriction, both the component distributions and the mixing proportions can be restricted to lie in a non-trivial set. Here we achieve point-identification by complementing the exclusion restriction with a mild restriction on the relative tail behavior of the component distributions. This restriction is quite natural in location model, for example, but can be motivated more generally, as we will discuss in more detail below.

AMS 2000 subject classifications: 62G05, 62G20, 62G30, 62G32.

Keywords and phrases: mixture model, nonparametric identification and estimation, tail empirical process.

Our identification argument readily suggests plug-in estimators of the mixing proportions and the component distributions. The estimator of the former is based on ratios of intermediate quantiles. The latter is a simple least-squares estimator. Both estimators are available in closed form and are therefore easy to implement. The convergence rate of the estimators is determined by the theory of tail empirical processes (see, e.g., del Barrio, Deheuvels and van de Geer [2007] for a recent account) and is slower than the parametric one, although it can be brought close to it. The intuition for this is the need to rely on the tail behavior of the component densities to infer the mixing proportions. Moreover, if the mixing proportions were known—or could be estimated at the parametric rate—the tail restrictions could be dispensed with and the implied estimator of the component distributions would equally converge at the conventional rate.

Below we first describe the mixture model under study and show identification of its components. These results are next used to motivate the construction of estimators. We provide distribution theory for these estimators and investigate their small-sample performance by means of a small Monte Carlo experiment. All technical details are collected in the Appendix. We omit conditioning variables; the identification analysis extends straightforwardly, and the distribution theory could be extended by using the local empirical-process results in Einmahl and Mason [1997].

2. Mixtures with tail restrictions. Let $(Y, X) \in \mathbb{R} \times \mathcal{X}$ for $\mathcal{X} \subseteq \mathbb{R}$ be a pair of random variables for which $F(y|x) \equiv \mathbb{P}(Y \leq y|X = x)$ decomposes as the two-component mixture

$$F(y|x) = G(y)\,\lambda(x) + H(y)\,(1 - \lambda(x)) \tag{2.1}$$

for distribution functions $G : \mathbb{R} \mapsto [0, 1]$ and $H : \mathbb{R} \mapsto [0, 1]$ and a function $\lambda : \mathcal{X} \mapsto (0, 1)$ that maps values $x \in \mathcal{X}$ into mixing proportions. Everything that follows can be adapted to a situation in which G and H are supported on a subset of the real line with an appropriate change of notation. We complete the mixture model with the following assumption.

Assumption 1. *The functions G and H are absolutely continuous on $\mathbb{R}$, and λ is non-constant and bounded away from zero and one on $\mathcal{X}$.*

The assumption that the component distributions do not depend on X means that X acts like an instrumental variable. Non-constancy of λ implies instrument relevance. Bounding λ away from zero and one ensures that the mixture is irreducible for all x. This could be relaxed at the cost of more complicated notation. However, bounding the weight function in this manner rules out the possibility of achieving identification of G and H via an identification-at-infinity argument, which is an approach that has been studied in a different context; see, e.g., Andrews and Schafgans [1998].

Our setup accommodates many models. We give two leading examples.

Example 1 (Mismeasured treatment). Let T denote a binary treatment indicator. Suppose that T is subject to classification error. That is, rather than observing T, we observe misclassified treatments, say X. The distribution of the outcome variable Y given $X = x$ is

$$F(y|x) = \mathbb{P}(Y \leq y|T = 1, X = x)\,\lambda(x) + \mathbb{P}(Y \leq y|T = 0, X = x)\,(1 - \lambda(x))$$

for $\lambda(x) = \mathbb{P}(T = 1|X = x)$. The conventional ignorability assumption states that X and Y are independent conditionally on $T = t$ for $t \in \{0, 1\}$. Then the decomposition of $F(y|x)$ is as in (2.1) with $G(y) = \mathbb{P}(Y \leq y|T = 1)$ and $H(y) = \mathbb{P}(Y \leq y|T = 0)$. Note also that λ is non-constant unless misclassification in T is completely random.

The problem of identification when the treatment indicator is mismeasured has received considerable attention, especially in the context of regression models (see, e.g., Bollinger [1996], Lewbel [2007], and Mahajan [2006]). The conditional ignorability assumption entertained here is an assumption of non-differential misclassification error and has been routinely made elsewhere (see, e.g., Carroll et al. [2006]).

Example 2 (Hidden Markov model). In a hidden Markov model, outcomes Y are i.i.d. conditional on belonging to a latent state S that follows a Markov process. Suppose that $S \in \{0, 1\}$ has first-order Markov dependence. Then, letting X be the first-order lag of Y, we have that

$$F(y|x) = \mathbb{P}(Y \leq y|S = 1)\,\mathbb{P}(S = 1|X = x) + \mathbb{P}(Y \leq y|S = 0)\,\mathbb{P}(S = 0|X = x),$$

which fits (2.1). Further, $\lambda(x) = \mathbb{P}(S = 1|X = x)$ indeed varies with x unless Y and S are independent or S is independently and identically distributed. The hidden Markov model, or regime-switching model, nests switching regressions as well as several versions of stochastic-volatility models.

Our identification approach complements the recent results on hidden Markov models that are based on the availability of multivariate data (see Allman, Matias and Rhodes [2009] and Gassiat and Rousseau [2013]).

To link our model to the literature on multivariate mixtures we also discuss a third example of a mixture that can be cast into (2.1).

Example 3 (Multivariate mixture). Suppose (Y, X) are independent conditional on a latent binary factor T. Then $\mathbb{P}(X \leq x, Y \leq y)$ factors as a bivariate two-component mixture. Further, because Y is independent of X given T, the conditional distribution of Y given X admits a representation as in (2.1), and so does the distribution of X given Y.

Multivariate mixtures with conditionally-independent components have recently received quite some attention. Hall and Zhou [2003] showed that a two-component mixture is point identified with trivariate data but is set identified when only bivariate data are available. These bounds can equally be established using the results in Henry, Kitamura and Salanié [2013].

We will show below that all of G, H, and λ are identified under a dominance condition on the tails of the component distributions. This condition is given in the next assumption.

Assumption 2 (Tail dominance).
(i) The left tail of G is thinner than the left tail of H, i.e., $\lim_{y\downarrow-\infty} G(y)/H(y) = 0$.
(ii) The right tail of G is thicker than the right tail of H, i.e., $\lim_{y\uparrow+\infty} (1 - H(y))/(1 - G(y)) = 0$.

Assumption 2 can be motivated by a situation in which large values of the outcome are very likely to have been drawn from G, and small values are very likely to have been generated from H. An important

special case where these tail conditions readily follow from the structure of the model is in a location model, as the following example shows.

EXAMPLE 4 (Location models). Suppose that $Y = \mu(T) + U$, where T is a binary treatment indicator, $U \sim F$ independent of T, and $\mu(t) \equiv \mathbb{E}[Y|T = t]$. Then (2.1) yields

$$F(y|x) = F(y - \mu(1))\,\mathbb{P}(T = 1|X = x) + F(y - \mu(0))\,\mathbb{P}(T = 0|X = x).$$

Suppose that $\mu(0) < \mu(1)$, and that the derivatives of $-\ln\big(1 - F(u)\big)$ and $-\ln F(u)$ grow without bound as $u \uparrow +\infty$ and $u \downarrow -\infty$, respectively. Then Assumption 2 is satisfied for the distributions $G(y) = F(y - \mu(1))$ and $H(y) = F(y - \mu(0))$.

Theorem 1 formally states our identification result.

THEOREM 1 (Identification). *Under Assumptions 1 and 2, G, H, and λ are identified.*

The proof is constructive. Fix $x' \in \mathcal{X}$ and choose $x'' \in \mathcal{X}$ so that $\lambda(x') \neq \lambda(x'')$. Then

$$\frac{F(y|x')}{F(y|x'')} = \frac{1 + \lambda(x')\,(G(y)/H(y) - 1)}{1 + \lambda(x'')\,(G(y)/H(y) - 1)},$$
$$\frac{1 - F(y|x')}{1 - F(y|x'')} = \frac{1 + \lambda(x')\,((1 - G(y))/(1 - H(y)) - 1)}{1 + \lambda(x'')\,((1 - G(y))/(1 - H(y)) - 1)},$$

follow from re-arranging (2.1). Taking limits, Assumption 2 further implies that

$$\zeta^-(x', x'') \equiv \lim_{y \downarrow -\infty} \frac{F(y|x')}{F(y|x'')} = \frac{1 - \lambda(x')}{1 - \lambda(x'')}, \qquad \zeta^+(x', x'') \equiv \lim_{y \uparrow +\infty} \frac{1 - F(y|x')}{1 - F(y|x'')} = \frac{\lambda(x')}{\lambda(x'')}, \tag{2.2}$$

which are well-defined. These two equations can be solved for the mixing proportion at x'. This yields

$$\lambda(x') = \frac{1 - \zeta^-(x'', x')}{\zeta^+(x'', x') - \zeta^-(x'', x')}. \tag{2.3}$$

Since λ is non-constant, for any $x' \in \mathcal{X}$ there exists a $x'' \in \mathcal{X}$ for which such a system of equations can be constructed. The function λ is therefore identified on its entire support. To establish identification of G and H, first note that, using (2.1),

$$D(y) \equiv G(y) - H(y) = \frac{F(y|x'') - F(y|x')}{\lambda(x'') - \lambda(x')},$$

which is identified. Further, for any $x \in \mathcal{X}$,

$$H(y) = F(y|x) - \lambda(x)\,D(y), \qquad G(y) = H(y) + D(y),$$

so that both component distributions are identified. This concludes the proof of Theorem 1.

If we only assume one-sided tail dominance, then one of the component distributions remains identified.

COROLLARY 1 (One-sided dominance). *Under Assumption 1, G is identified if Assumption 2(i) holds and H is identified if Assumption 2(ii) holds.*

As an illustration, consider a two-regime stochastic volatility model. Go back to Example 2 and assume that the outcome has conditional mean zero and conditional variance $S\sigma_G^2 + (1-S)\sigma_H^2$ for positive constants σ_G^2, σ_H^2. Suppose that $\sigma_G^2 > \sigma_H^2$. Then G is the distribution associated with a regime that is characterized by higher volatility. In this case, both tails of G dominate those of H and so Assumption 2(i) fails, but the distribution of the lower-volatility regime, H, remains identified.

3. Estimation and inference. The proof of Theorem 1 suggests simple estimators for both the distributions G and H and the mixing function λ. We will characterize their asymptotic distribution under the following sampling conditions.

ASSUMPTION 3. *$\mathcal{X} = \{x_1, x_2, \ldots, x_q\}$ for some finite integer $q \geq 2$ and $\{Y_i, X_i\}_{i=1}^n$ constitutes a random sample on (Y, X).*

The results below can be extended to weakly-dependent sequences using convergence results for tail empirical processes in Rootzén [2009].

Our procedure builds on estimators of the limit quantities in (2.2). These estimators are constructed based on order statistics of the data. For each $x \in \mathcal{X}$, let $n_x \equiv \sum_{i=1}^n 1\{X_i = x\}$, and let

$$F_n(y|x) \equiv n_x^{-1} \sum_{i=1}^n 1\{Y_i \leq y, X_i = x\}.$$

For integers ι and κ, let $\ell_x^{\iota n}$ and $r_x^{\kappa n}$ be the $(\iota+1)$th and $(n-\kappa)$th order statistics of the subsample in which the instrument takes on value x. For each $(x', x'') \in \mathcal{X}$ we estimate $\zeta^-(x', x'')$ and $\zeta^+(x', x'')$ by

$$\zeta_{\iota n}^-(x', x'') \equiv \frac{F_n(\ell_{x''}^{\iota n}|x')}{F_n(\ell_{x''}^{\iota n}|x'')}, \qquad \zeta_{\kappa n}^+(x', x'') \equiv \frac{1 - F_n(r_{x''}^{\kappa n}|x')}{1 - F_n(r_{x''}^{\kappa n}|x'')}. \tag{3.1}$$

We will choose the order statistics so that $\ell_x^{\iota n} \downarrow -\infty$ and $r_x^{\kappa n} \uparrow +\infty$ as $n \uparrow +\infty$ at an appropriate rate.

ASSUMPTION 4 (Order statistics). *For all $x \in \mathcal{X}$, $\iota/n_x \downarrow 0$ and $\iota/\sqrt{n_x \ln\ln n_x} \uparrow +\infty$, and $\kappa/n_x \downarrow 0$ and $\kappa/\sqrt{n_x \ln\ln n_x} \uparrow +\infty$ as $n \uparrow +\infty$.*

Note that $F_n(\ell_{x''}^{\iota n}|x'') > 0$ and $F_n(r_{x''}^{\kappa n}|x'') < 1$ must hold for $\zeta_{\iota n}^-(x', x'')$ and $\zeta_{\kappa n}^+(x', x'')$ to be well defined, and so it is recommended to inspect the empirical distribution functions when selecting the order statistics.

With the estimators in (3.1) at hand we can turn to the estimator of λ. Observe that, because λ need not be strictly monotonic, estimating $\lambda(x)$ by an average of plug-in estimates of (2.3) can be problematic because the denominator of some of the contributions can be zero or be arbitrarily close to it. To avoid such difficulties, we proceed in a different manner. Write $\|\cdot\|$ for the Euclidean norm. Let

$$\boldsymbol{\zeta}^+(x_j) \equiv \big(\zeta^+(x_1, x_j), \zeta^+(x_2, x_j), \ldots, \zeta^+(x_{j-1}, x_j), \zeta^+(x_{j+1}, x_j), \ldots, \zeta^+(x_q, x_j)\big)',$$

and define $\boldsymbol{\zeta}^-(x_j)$ in an analogous fashion. Because λ is non-negative, it follows from (2.3) that it can equally be expressed as $\lambda(x) = \|\mathbf{1} - \boldsymbol{\zeta}^-(x)\|/\|\boldsymbol{\zeta}^+(x) - \boldsymbol{\zeta}^-(x)\|$, where $\mathbf{1}$ is a vector of $(q-1)$ ones. Letting $\boldsymbol{\zeta}^-_{\iota n}(x)$ and $\boldsymbol{\zeta}^+_{\kappa n}(x)$ denote the estimated counterparts of $\boldsymbol{\zeta}^-(x)$ and $\boldsymbol{\zeta}^+(x)$, we estimate $\lambda(x)$ by

$$\lambda_n^{\iota\kappa}(x) = \frac{\|\mathbf{1} - \boldsymbol{\zeta}^-_{\iota n}(x)\|}{\|\boldsymbol{\zeta}^+_{\kappa n}(x) - \boldsymbol{\zeta}^-_{\iota n}(x)\|}.$$

This estimator is immune to the problem just mentioned.

Given an estimator of the mixing proportions, estimators of the distribution functions are readily constructed via minimum distance. Recall that

$$F(y|x) = H(y) + \lambda(x)\, D(y),$$

for each (y, x). Letting $\boldsymbol{F}(y) \equiv (F(y|x_1), F(y|x_2), \ldots, F(y|x_q))'$ and $\boldsymbol{\lambda} \equiv (\lambda(x_1), \lambda(x_2), \ldots, \lambda(x_q))'$, we thus have that, for each fixed y, $(H(y), D(y))' = \arg\min_{\alpha,\beta} \|\boldsymbol{F}(y) - \alpha - \boldsymbol{\lambda}\beta\|^2$. As our estimators of $H(y)$ and $D(y)$, we therefore use

$$(H_n^{\iota\kappa}(y), D_n^{\iota\kappa}(y))' \equiv \arg\min_{\alpha,\beta} \|\boldsymbol{F}_n(y) - \alpha - \boldsymbol{\lambda}_n^{\iota\kappa}\beta\|^2,$$

where $\boldsymbol{F}_n(y)$ and $\boldsymbol{\lambda}_n^{\iota\kappa}$ are defined in an obvious manner. Finally, an estimator of $G(y)$ is then easily constructed as $G_n^{\iota\kappa}(y) \equiv H_n^{\iota\kappa}(y) + D_n^{\iota\kappa}(y)$. These estimators are convenient because they are available in closed form. Indeed, $H_n^{\iota\kappa}(y)$ and $D_n^{\iota\kappa}(y)$ are obtained from the ordinary least-squares regression of $F_n(y|x_1), F_n(y|x_2), \ldots, F_n(y|x_q)$ on the q estimated mixing proportions $\lambda_n^{\iota\kappa}(x_1), \lambda_n^{\iota\kappa}(x_2), \ldots, \lambda_n^{\iota\kappa}(x_q)$ and a q-vector of ones.

To ensure that the limit distributions of our estimators are free of asymptotic bias, we need an additional condition on the relative tails of the component distributions.

ASSUMPTION 5 (Tail rates). *For each $x \in \mathcal{X}$,*

$$G(\ell_x^{\iota n})/H(\ell_x^{\iota n}) = o_p(1/\sqrt{\iota}) \quad \textit{and} \quad (1 - H(r_x^{\kappa n}))/(1 - G(r_x^{\kappa n})) = o_p(1/\sqrt{\kappa})$$

hold.

The following two examples specialize Assumption 5 to densities with log-concave tails and Pareto tails, respectively.

EXAMPLE 5 (Log-concave tails). Suppose that G and H have log-concave tails, i.e.,

$$-\ln(1 - G(y)) \sim \left(\frac{y}{\sigma_G^+}\right)^{\alpha_G^+}, \quad -\ln(1 - H(y)) \sim \left(\frac{y}{\sigma_H^+}\right)^{\alpha_H^+}, \quad \text{as } y \uparrow +\infty,$$

for real numbers $\alpha_G^+, \alpha_H^+ > 1$ and $\sigma_G^+, \sigma_H^+ > 0$, and

$$-\ln G(y) \sim \left(\frac{y}{\sigma_G^-}\right)^{\alpha_G^-}, \quad -\ln H(y) \sim \left(\frac{y}{\sigma_H^-}\right)^{\alpha_H^-}, \quad \text{as } y \downarrow -\infty,$$

for real numbers $\alpha_G^-, \alpha_H^- > 1$ and $\sigma_G^-, \sigma_H^- > 0$. Then Assumption 4 implies Assumption 5 if both

(i) $\alpha_G^+ < \alpha_H^+$, *or* $\alpha_G^+ = \alpha_H^+$ and $\sigma_G^+ > \sigma_H^+$; and

(ii) $\alpha_G^- > \alpha_H^-$, *or* $\alpha_G^- = \alpha_H^-$ and $\sigma_G^- < \sigma_H^-$

hold.

Example 5 does not cover location models with log-concave distributions, for which the α and σ parameters of H equal those of G. In such models, the tails of the error density converge to zero too fast to ensure zero asymptotic bias. To assess the sensitivity of our methods, the simulation experiment below explicitly deals with a mixture of normals, where Assumption 5 fails.

Example 6 (Pareto tails). Let C denote a generic constant. Suppose that G and H have Pareto tails, i.e.,

$$(1 - G(y)) \sim C\, y^{-\alpha_G^+}, \quad (1 - H(y)) \sim C\, y^{-\alpha_H^+}, \quad \text{as } y \uparrow +\infty,$$

for positive real numbers $\alpha_H^+ > \alpha_G^+$ and

$$G(y) \sim C\,(-y)^{-\alpha_G^-}, \quad H(y) \sim C\,(-y)^{-\alpha_H^-}, \quad \text{as } y \downarrow -\infty,$$

for positive real numbers $\alpha_G^- < \alpha_H^-$. Then Assumption 4 implies Assumption 5.

The main step in our asymptotic analysis is summarized in the following lemma.

Lemma 1 (Intermediate quantiles). *Let Assumptions 1–5 hold. Then, as $n \uparrow +\infty$, for each $x', x'' \in \mathcal{X}$,*

$$\begin{aligned}
\sqrt{\kappa}\big(\zeta_{\kappa n}^+(x', x'') - \zeta^+(x', x'')\big) &\overset{A}{\sim} \zeta^+(x', x'')\, Z^+(x'') - \sqrt{\zeta^+(x', x'')}\, Z^+(x'), \\
\sqrt{\iota}\big(\zeta_{\iota n}^-(x', x'') - \zeta^-(x', x'')\big) &\overset{A}{\sim} \zeta^-(x', x'')\, Z^-(x'') - \sqrt{\zeta^-(x', x'')}\, Z^-(x'),
\end{aligned}$$

where the $Z^+(x)$ and $Z^-(x)$ are standard-normal variates that are both mutually independent and independent across x.

The asymptotic behavior of $\lambda_n^{\iota\kappa}$, $G_n^{\iota\kappa}$, and $H_n^{\iota\kappa}$ can readily be deduced from Lemma 1 by an application of the delta method.

Start with $\lambda_n^{\iota\kappa}$. Lemma 1 implies that

$$\sqrt{\iota}\big(\boldsymbol{\zeta}_{\iota n}^-(x) - \boldsymbol{\zeta}^-(x)\big) \xrightarrow{d} \mathcal{N}\big(0, \Sigma(x)\big), \quad \Sigma(x) \equiv \boldsymbol{\zeta}^-(x)\boldsymbol{\zeta}^-(x)' + \operatorname{diag}\{\boldsymbol{\zeta}^-(x)\}.$$

Write the vector of partial derivatives of $\|\mathbf{1} - \boldsymbol{\zeta}^-(x)\|$ with respect to the $\zeta^-(x_i, x)$ as $d(x)$. Then we can state the following theorem.

Theorem 2 (Weights). *Let Assumptions 1–5 hold. Then, for each $x \in \mathcal{X}$,*

$$\sqrt{\iota}\big(\lambda_n^{\iota\kappa}(x) - \lambda(x)\big) \xrightarrow{d} \mathcal{N}(0, \sigma^2(x)), \qquad \sigma^2(x) \equiv \frac{d(x)'\Sigma(x)d(x)}{\|\boldsymbol{\zeta}^+(x) - \boldsymbol{\zeta}^-(x)\|^2},$$

as $n \uparrow +\infty$.

Consistency of a plug-in estimator of the asymptotic variance in this theorem is immediate.

From Lemma 1 we can deduce the asymptotic covariance between $\sqrt{\iota}\zeta^-_{\iota n}(x_l, x_i)$ and $\sqrt{\iota}\zeta^-_{\iota n}(x_k, x_j)$ for $i \neq j$, $l \neq i$, and $k \neq j$. Writing this covariance as $\gamma_{i,j}(l,k)$, we have

$$\gamma_{i,j}(l,k) = \begin{cases} \sqrt{\zeta^-(x_l,x_i)\,\zeta^-(x_l,x_j)} & \text{if } l = k \\ -\zeta^-(x_k,x_j)\sqrt{\zeta^-(x_j,x_i)} & \text{if } l = j, k \neq i \\ -\zeta^-(x_l,x_i)\sqrt{\zeta^-(x_i,x_j)} & \text{if } l \neq j, k = i \\ -\zeta^-(x_j,x_i)\sqrt{\zeta^-(x_i,x_j)} - \zeta^-(x_i,x_j)\sqrt{\zeta^-(x_j,x_i)} & \text{if } l = j, k = i \\ 0 & \text{otherwise} \end{cases}.$$

This gives the asymptotic covariance between $\lambda^{\iota\kappa}_n(x_i)$ and $\lambda^{\iota\kappa}_n(x_j)$ as

$$\upsilon(x_i,x_j) \equiv \frac{d(x_i)'\Omega(x_i,x_j)d(x_j)}{\|\boldsymbol{\zeta}^+(x_i) - \boldsymbol{\zeta}^-(x_i)\|\,\|\boldsymbol{\zeta}^+(x_j) - \boldsymbol{\zeta}^-(x_j)\|},$$

where $\Omega(x_i,x_j)$ is the asymptotic covariance between $\sqrt{\iota}\boldsymbol{\zeta}^-_n(x_i)$ and $\sqrt{\iota}\boldsymbol{\zeta}^-_n(x_j)$, constructed using $\gamma_{i,j}(l,k)$ when $i \neq j$ and using $\Sigma(x_i)$ when $i = j$. Collect the various $\upsilon(x_i,x_j)$ in a $q \times q$ matrix and call this matrix Υ. Then $\sqrt{\iota}(\boldsymbol{\lambda}^{\iota\kappa}_n - \boldsymbol{\lambda}) \xrightarrow{d} \mathcal{N}(0,\Upsilon)$ follows.

For any function A of x, let $\overline{A}(x) \equiv q^{-1}\sum_{i=1}^q A(x_i)$. To be able to state the asymptotic variance of the estimators of $H(y)$ and $G(y)$ for given y we need to introduce the $q \times 1$ vectors

$$\boldsymbol{\xi}_H \equiv -\frac{\overline{\lambda}(x)}{\big(\boldsymbol{\lambda} - \overline{\lambda}(x)\big)'\big(\boldsymbol{\lambda} - \overline{\lambda}(x)\big)}\,\big(\boldsymbol{F}(y) - \overline{F}(y|x)\big)', \quad \boldsymbol{\xi}_G \equiv \frac{1 - \overline{\lambda}(x)}{\big(\boldsymbol{\lambda} - \overline{\lambda}(x)]\big)'\big(\boldsymbol{\lambda} - \overline{\lambda}(x)\big)}\,\big(\boldsymbol{F}(y) - \overline{F}(y|x)\big)',$$

where we leave the dependence on y implicit. Theorem 3 provides the limit distributions of the estimators of the component distributions.

THEOREM 3 (Distributions). *Let Assumptions 1–5 hold. Then, for each $y \in \mathbb{R}$,*

$$\sqrt{\iota}\big(H^{\iota\kappa}_n(y) - H(y)\big) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2(y;H)\right), \qquad \sqrt{\iota}\big(G^{\iota\kappa}_n(y) - G(y)\big) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2(y;G)\right),$$

where $\sigma^2(y;H) \equiv \boldsymbol{\xi}_H \Upsilon \boldsymbol{\xi}_H'$ and $\sigma^2(y;G) \equiv \boldsymbol{\xi}_G \Upsilon \boldsymbol{\xi}_G'$, as $n \uparrow +\infty$.

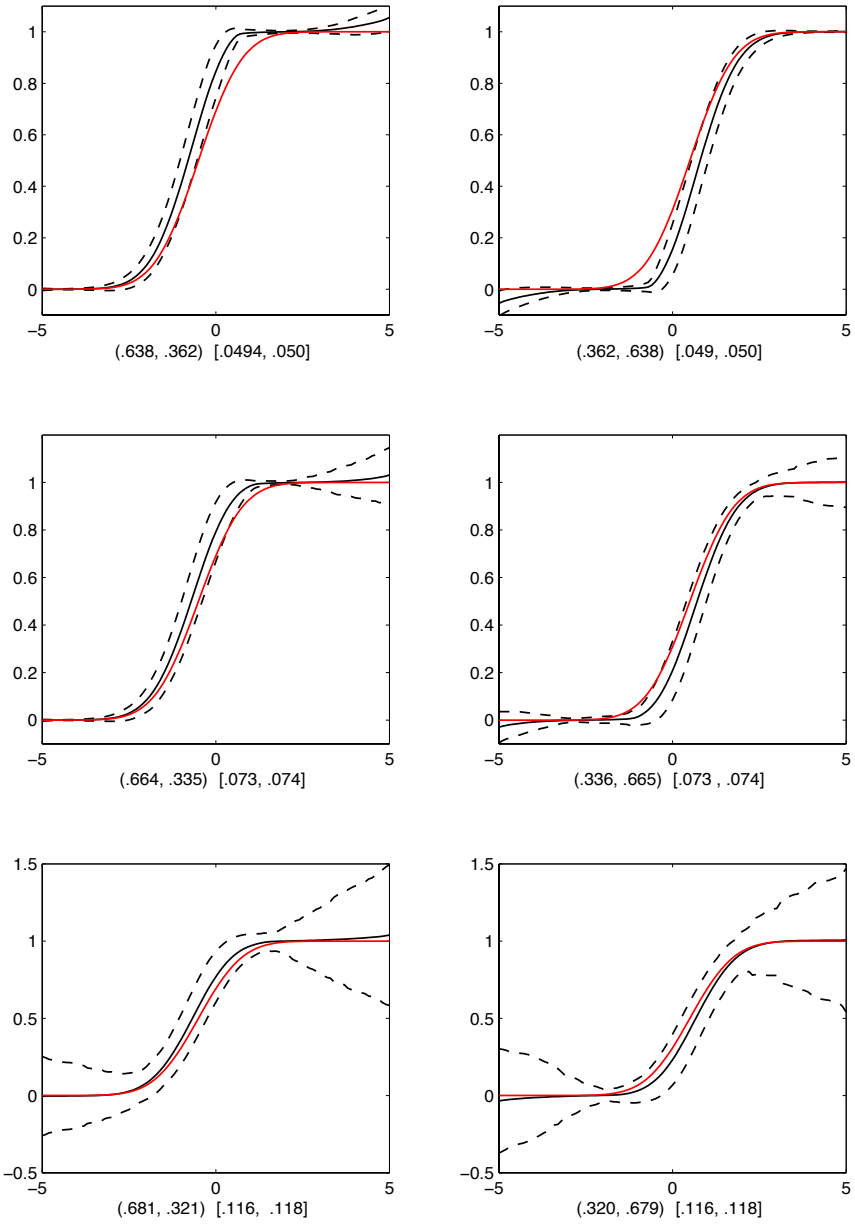
As a corollary to Theorem 3 we obtain uniform confidence bounds for G and H.

COROLLARY 2 (Uniform inference). *Let Assumptions 1–5 hold. Then*

$$\lim_{n\uparrow+\infty} \mathbb{P}\left\{\sup_{y\in\mathbb{R}}\left|\frac{\sqrt{\iota}\,(G^{\iota\kappa}_n(y) - G(y))}{\sigma_n(y;G)}\right| > z_{\tau/2}\right\} = \tau, \quad \lim_{n\uparrow+\infty} \mathbb{P}\left\{\sup_{y\in\mathbb{R}}\left|\frac{\sqrt{\iota}\,(H^{\iota\kappa}_n(y) - H(y))}{\sigma_n(y;H)}\right| > z_{\tau/2}\right\} = \tau,$$

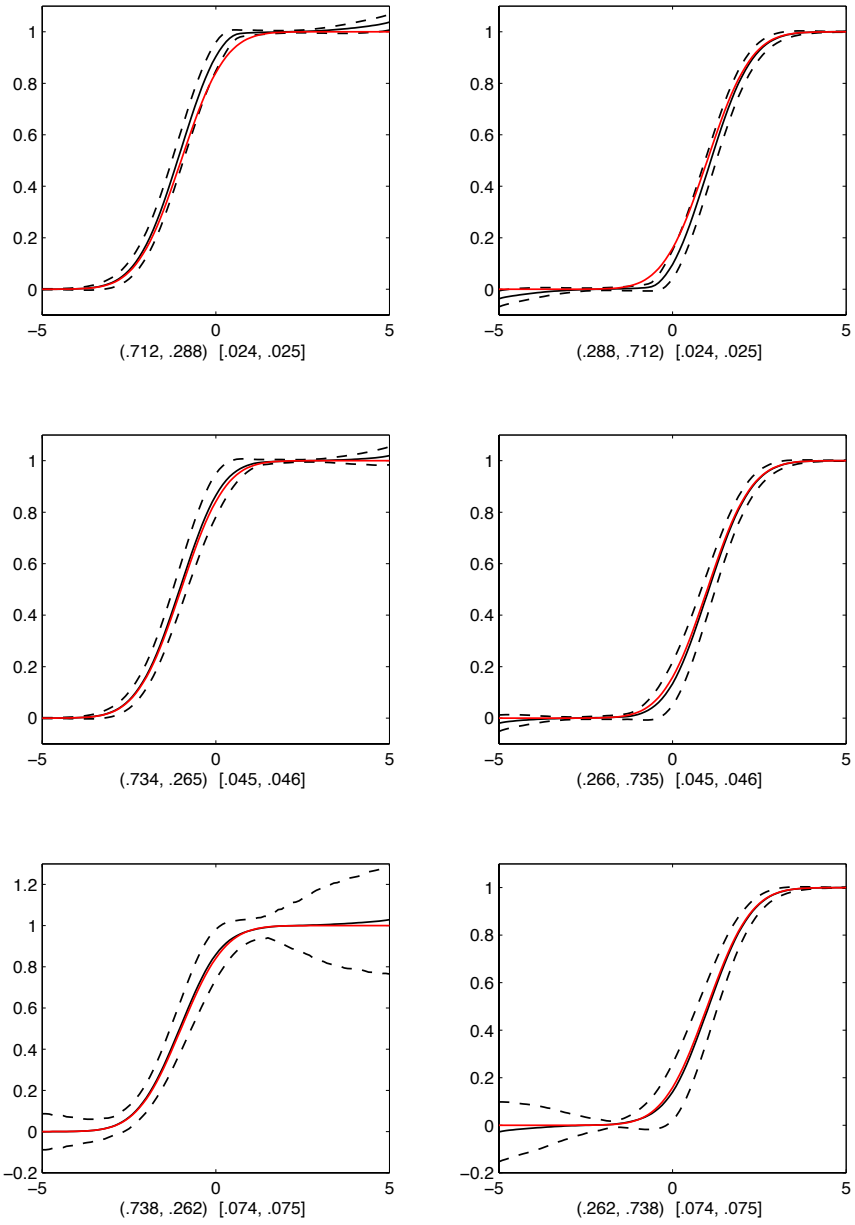
where z_τ denotes the $1-\tau$ quantile of the standard normal distribution, and $\sigma_n(y;G)$ and $\sigma_n(y;H)$ denote plug-in estimators of $\sigma(y;G)$ and $\sigma(y;H)$, respectively.

FIG 1. $\alpha = .5$



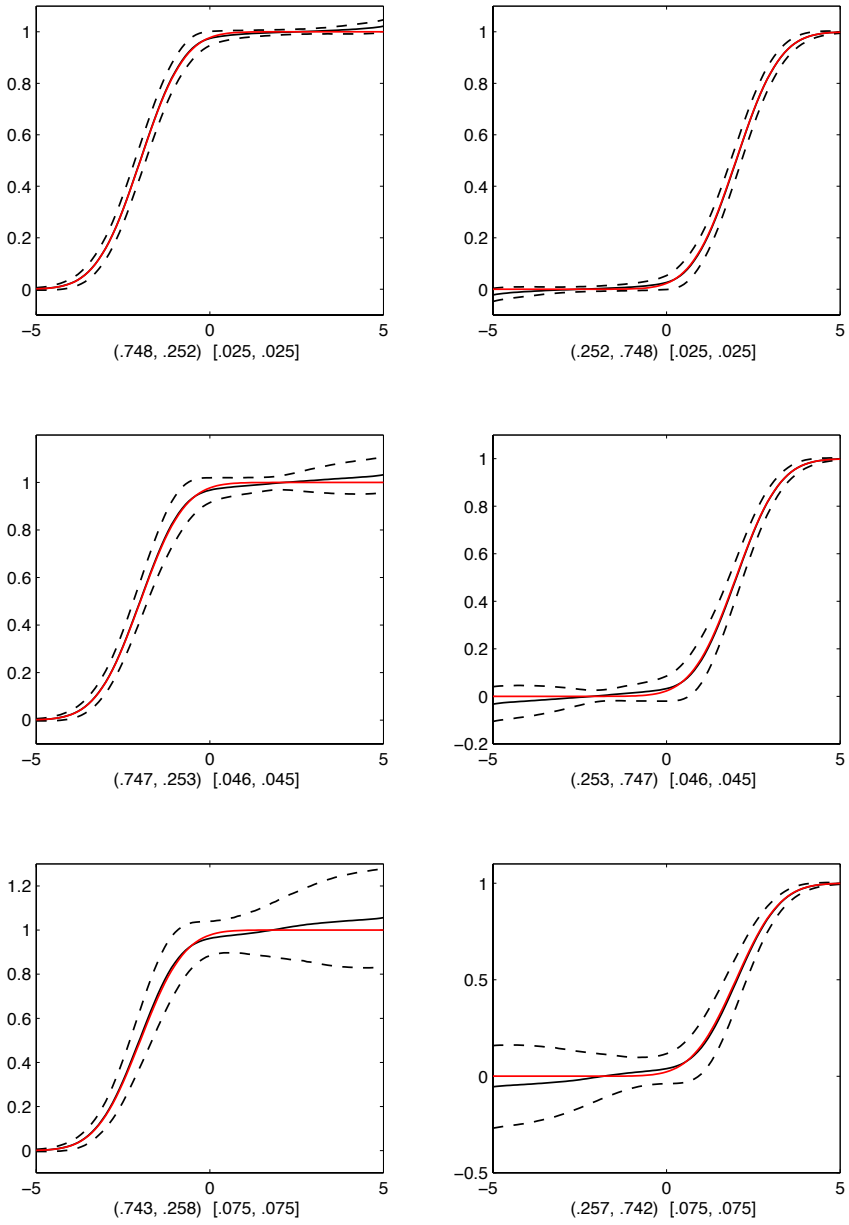
Left: G and $(\lambda(0), \lambda(1))$. Right: H and $(1-\lambda(0), 1-\lambda(1))$. Upper: .80 quantile. Middle: .90 quantile. Lower: .95 quantile. $10,000$ replications and $n = 1,000$.

FIG 2. $\alpha = 1$



Left: G and $(\lambda(0), \lambda(1))$. Right: H and $(1 - \lambda(0), 1 - \lambda(1))$. Upper: .80 quantile. Middle: .90 quantile. Lower: .95 quantile. $10,000$ replications and $n = 1,000$.

FIG 3. $\alpha = 2$



Left: G and $(\lambda(0), \lambda(1))$. Right: H and $(1-\lambda(0), 1-\lambda(1))$. Upper: .80 quantile. Middle: .90 quantile. Lower: .95 quantile. $10,000$ replications and $n = 1,000$.

4. Simulation experiment. As a numerical illustration, we simulate data from a location version of the misclassification problem in Example 1. The observed outcomes were generated from the model

$$Y = \alpha T + \sigma U, \qquad U \sim \mathcal{N}(0,1),$$

where the treatment indicator, T, takes on the values -1 and 1, and $\mathcal{X} = \{0,1\}$. Treatment was assigned according to

$$\lambda(1) = \mathbb{P}(T = 1|X = x) = \frac{1}{4} = 1 - \lambda(0),$$

and X was generated so that $\mathbb{P}(X = 1) = \frac{1}{2}$. Under such a data generating process, the R^2 of the regression of Y on T is

$$R^2 = \frac{\gamma^2}{1+\gamma^2}, \qquad \gamma \equiv \frac{\alpha}{\sigma}.$$

We set $\sigma = 1$ and present results for $\alpha \in \{.5, 1, 2\}$, corresponding to $R^2 \in \{.20, .50, .80\}$, and, therefore,

$$G(y) = \Phi(y - \alpha), \qquad H(y) = \Phi(y + \alpha).$$

Note that larger values of α make the relative thickness of the tails of G and H more pronounced. The sample size was fixed to $n = 1,000$, and 10,000 Monte Carlo replications were performed at each design point. We estimated the distributions G, H and the weights $\lambda(0), \lambda(1)$ using the .80, .90, and .95 quantile of the subsample with $X = 0$ for $r_0^{\kappa n}$ and the .20, .10, and .05 quantile of the empirical distribution function of Y given $X = 0$ for $\ell_0^{\iota n}$. To ensure our estimates of G and H to be monotonic, we first computed $G_n^{\iota\kappa}$ and $H_n^{\iota\kappa}$ via our least squares procedure described above and then applied a re-arrangement algorithm Chernozhukov, Fernández-Val and Galichon [2009] on the estimates.

Figures 1–3 present the results for the three values of α. Each figure provides the mean (solid line) over the point estimates of G and H, the 95% and 5% envelopes of the empirical distribution of the point estimates (dashed lines), as well as the true curves (solid red line). In each figure, the left plots concern G, the right plots concern H. The upper (middle; lower) panels corresponds to the estimates obtained by using the .20th (.10th; .05th) quantile of the empirical conditional distribution of Y given $X = 0$ for $\ell_0^{\iota n}$ and the .80th (.90th; .95th) quantile of the same distribution for $r_0^{\kappa n}$. Below each plot we provide the mean of the estimates of $\lambda(0)$ and $\lambda(1)$ (left) and of $1 - \lambda(0)$ and $1 - \lambda(1)$ (right), together with their standard deviation in square brackets.

For all design points considered the estimates component distributions reflect the shape of the true distributions well. The bias tends to decrease as the cut-off points $\ell_0^{\iota n}$ and $r_0^{\kappa n}$ are chosen further in the tails of the relevant empirical distribution. This decrease in bias is most pronounced for the design with $\alpha = .5$. For the other designs the effect is much smaller because the tail dominance is much stronger in these cases due to the larger values of α. Choosing $\ell_0^{\iota n}$ and $r_0^{\kappa n}$ too far in the tails of the relevant empirical distribution results in larger dispersion of the estimates, however, as is apparent from the plots. This bias-variance trade-off is intuitive. The variance inflation is further concentrated around the tails of G and H, which is again in line with expectations.

Similar conclusions can be drawn for the estimator of the mixing proportions. Indeed, the mixing proportions, too, give roughly correct estimates of the unobserved proportions; recall that $\lambda(0) = 3/4$ and $\lambda(1) = 1/4$. The bias decreases as α increases. A bias-variance trade-off is again visible. Indeed, the standard deviation of the point estimates increases roughly be a factor of two as we go down the plots within a given figure.

Acknowledgements. Some of the results presented here previously circulated as part of Henry, Kitamura and Salanié [2010], whose published version (Henry, Kitamura and Salanié [2013]) only contains results on partial identification. We are grateful to Victor Chernozhukov and Yuichi Kitamura for fruitful discussions. Parts of this paper were written while Henry was visiting the University of Tokyo Graduate School of Economics and while Salanié was visiting the Toulouse School of Economics. The hospitality of both institutions is gratefully acknowledged. Henry's research has received funding from SSHRC Grants 410-2010-242 and 435-2013-0292, and NSERC Grant 356491-2013. Jochmans' research has received funding from the SAB grant "Nonparametric estimation of finite mixtures".

APPENDIX

TECHNICAL DETAILS FOR EXAMPLE 4. Let $\varphi(u) \equiv -\ln(1-F(u))$ and let φ' be its first derivative. A mean-value expansion gives

$$\frac{1-F(y-\mu(0))}{1-F(y-\mu(1))} = \exp\left(\varphi(y-\mu(1)) - \varphi(y-\mu(0))\right) = \exp\left(-\varphi'(*)\left(\mu(1)-\mu(0)\right)\right),$$

where $*$ is some value that lies between $y-\mu(1)$ and $y-\mu(0)$. As φ' increases without bound as $y \uparrow +\infty$ and $\mu(1) > \mu(0)$, the expression on the right-hand side tends to zero as y increases. Assumption 2(i) is verified in the same way. □

PROOF OF THEOREM 1. The proof is given in the main text. □

PROOF OF COROLLARY 1. For brevity we only consider identification of H. Let x', x'' be as in the proof of Theorem 1. Under Assumption 2(ii) we can still determine $\zeta^+(x', x'') = \lambda(x')/\lambda(x'')$, from which we can learn the ratio

$$\frac{\lambda(x'')}{\lambda(x'') - \lambda(x')} = \frac{1}{1-\zeta^+(x', x'')}.$$

Recalling that, for any $x \in \mathcal{X}$,

$$H(y) = F(y|x) - \frac{\lambda(x)}{\lambda(x'') - \lambda(x')}\left(F(y|x'') - F(y|x')\right),$$

the result follows on evaluating this expression at $x = x''$. □

TECHNICAL DETAILS FOR EXAMPLE 5 . We verify the second rate; the first follows in an analogous fashion. Throughout, fix $x \in \mathcal{X}$ and omit it as argument whenever possible. Assumptions 2(ii) and 4 imply that

$$1 - F(r|x) = (1-G(r))\ \lambda(x) + (1-H(r))\ (1-\lambda(x)) = (1-G(r))\ (\lambda(x) + o_p(1)).$$

Further, because $\kappa/n_x = 1 - F_n(r|x)$, adding and subtracting $F(r|x)$ gives

$$\frac{\kappa}{n_x} = (1-F(r|x)) + (F_n(r|x) - F(r|x)) = (1-F(r|x)) + O_{a.s}(\sqrt{(\ln\ln x_x)/n_x}) = (1-F(r|x)) + o_p(1).$$

Put together, we find

$$\frac{\kappa}{n_x} = C\ (1-G(r))\ (1+o_p(1))$$

for some constant C. With G and H having log-concave tails, it then follows from this expression that, asymptotically, r behaves like $\sqrt[\alpha_G^+]{\ln n_x}$. So, because

$$\frac{1-H(r)}{1-G(r)} \sim \exp\left\{\left(\frac{r}{\sigma_G^+}\right)^{\alpha_G^+} - \left(\frac{r}{\sigma_H^+}\right)^{\alpha_H^+}\right\},$$

we have that

$$\frac{1-H(r)}{1-G(r)} = \begin{cases} O_p\left(\exp(-(\ln n_x)^{\alpha_H^+/\alpha_G^+})\right) & \text{if } \alpha_H^+ > \alpha_G^+ \\ O_p\left(1/n_x\right) & \text{if } \alpha_H^+ = \alpha_G^+ \text{ and } \sigma_H^+ < \sigma_G^+ \end{cases},$$

from which the conclusion follows. □

Technical details for Example 6. The argument is very similar to the one that was used to verify Example 5. We treat the right tail only. We have

$$\frac{\kappa}{n_x} = (1-G(r))\,(1+o_p(1)) = C\,r^{-\alpha_G^+}\,(1+o_p(1))\,.$$

Assumption 5 requires that $(1-H(r))/(1-G(r)) = o(1/\sqrt{\kappa})$, that is, that $r^{\alpha_G^+-\alpha_H^+} = o_p(1/\sqrt{\kappa})$. This rate condition is satisfied when

$$\left(\frac{n}{\kappa}\right)^{\frac{\alpha_G^+-\alpha_H^+}{\alpha_G^+}} = o_p(1/\sqrt{\kappa}),$$

which can be achieved by setting $\kappa = o(n^{\gamma^+})$ and $\iota = o(n^{\gamma^-})$ for

$$\gamma^+ \equiv \frac{\alpha_H^+ - \alpha_G^+}{\alpha_H^+ - \alpha_G^+/2}, \qquad \gamma^- \equiv \frac{\alpha_H^- - \alpha_G^-}{\alpha_H^- - \alpha_G^-/2}.$$

Both conditions are weaker than Assumption 4 and therefore implied by it. □

Proof of Lemma 1 . Throughout the proof we fix (x', x'') and omit them as arguments unless doing so would cause confusion. We focus on the limit behavior of $\sqrt{\kappa}(\zeta_{\iota n}^+ - \zeta^+)$; the corresponding result for $\sqrt{\iota}(\zeta_{\iota n}^- - \zeta^-)$ follows in an analogous fashion. First note that

$$\sqrt{\kappa}(\zeta_{\iota n}^+ - \zeta^+) = \sqrt{\kappa}(\zeta_{\iota n}^+ - \zeta^\kappa) + \sqrt{\kappa}(\zeta^\kappa - \zeta^+),$$

for $\zeta^\kappa \equiv (1-F(r|x'))/(1-F(r|x''))$. Using Assumption 5 together with a small calculation shows that

$$\sqrt{\kappa}(\zeta^\kappa - \zeta^+) = \sqrt{\kappa}\left(\frac{\lambda(x') + \frac{1-H(r)}{1-G(r)}(1-\lambda(x'))}{\lambda(x'') + \frac{1-H(r)}{1-G(r)}(1-\lambda(x''))} - \frac{\lambda(x')}{\lambda(x'')}\right) = \sqrt{\kappa}\,O_p\left(\frac{1-H(r)}{1-G(r)}\right) = o_p(1),$$

and so the second term is asymptotically negligible. Moving on to the first term, note that

$$\sqrt{\kappa}(\zeta_{\iota n}^+ - \zeta^\kappa) = \sqrt{\frac{n_{x''}}{\kappa}}\left(\zeta^\kappa \mathbb{G}_{n_{x''}}(r) - \mathbb{G}_{n_{x'}}(r)\right)$$

where $\mathbb{G}_{n_x}(y) \equiv \sqrt{n_x}\big(F_n(y|x) - F(y|x)\big)$. Let $\alpha_n(u) \equiv \sqrt{n}\big(\mathcal{U}_n(u) - u\big)$ where $\mathcal{U}_n$ denotes the empirical measure of an i.i.d. sample of size n from $\mathcal{U}[0,1]$. Then, because $F(\cdot|x)$ is a continuous function for all $x \in \mathcal{X}$ by Assumption 1, $\mathbb{G}_{n_{x'}}(y) = \alpha_{n_x}\big(1-F(y|x')\big)$ and $\mathbb{G}_{n_{x''}}(y) = \alpha_{n_x}\big(1-F(y|x'')\big)$ by an application of the probability integral transform. Thus we may write

$$\sqrt{\kappa}(\zeta^+_{\iota n} - \zeta^\kappa) = \sqrt{\frac{n_{x''}}{\kappa}}\zeta^\kappa\, \alpha_{n_{x''}}\big(1 - F(r|x'')\big) - \sqrt{\frac{n_{x''}}{\kappa}}\, \alpha_{n_{x'}}\big(1 - F(r|x')\big). \tag{A.1}$$

Now, for the first right-hand side term in this expression, by definition of the order statistic r and by adding and subtracting $F_n(r|x'')$ and re-arranging,

$$\alpha_{n_{x''}}\big(1 - F(r|x'')\big) = \alpha_{n_{x''}}\left(\frac{\kappa}{n_{x''}} + \frac{1}{\sqrt{n_{x''}}}\mathbb{G}_{n_{x''}}(r)\right) = \alpha_{n_{x''}}\left(\frac{\kappa}{n_{x''}}\left\{1 + \frac{\sqrt{n_{x''}}}{\kappa}\mathbb{G}_{n_{x''}}(r)\right\}\right).$$

By the law of the iterated logarithm, $\mathbb{G}_{n_{x''}}(r) = O_{a.s}(\sqrt{\ln\ln x''})$, so $\varepsilon \equiv -\sqrt{n_{x''}}/\kappa\, \mathbb{G}_{n_{x''}}(r) = o_{a.s}(1)$. Hence, the first term of $\sqrt{\kappa}(\zeta^+_{\iota n} - \zeta^\kappa)$ equals

$$\sqrt{\frac{n_{x''}}{\kappa}}\zeta^\kappa\, \alpha_{n_{x''}}\big(1 - F(r|x'')\big) = \sqrt{\frac{n_{x''}}{\kappa}}\zeta^\kappa\, \alpha_{n_{x''}}\left(\frac{\kappa}{n_{x''}}\{1-\varepsilon\}\right) = \zeta^\kappa B(1-\varepsilon) + o_p(1) \overset{A}{\sim} \zeta^+ Z^+(x''), \tag{A.2}$$

where $B(1-\varepsilon)$ is a sequence of standard Brownian motions and $Z^+(x'')$ is a standard-normal random variable. Here, the second transition follows from the central limit theorem for tail empirical processes (see Theorem 2.1 in Einmahl [1992]), which states that

$$\sup_{0<t\leq 1}\left|\sqrt{\frac{n_{x''}}{\kappa}}\alpha_{n_{x''}}\left(\frac{\kappa}{n_{x''}}t\right) - B(t)\right| = o_p(1),$$

and the fact that Brownian motions are almost surely continuous. For the second right-hand side term in (A.1), on observing that $1 - F(r|x') = \big(1 - F(r|x'')\big)\zeta^\kappa$ we can proceed in the same fashion to arrive at

$$\sqrt{\frac{n_{x''}}{\kappa}}\, \alpha_{n_{x'}}\big(1 - F(r|x')\big) = \sqrt{\zeta^+}\sqrt{\frac{n_{x'}}{\kappa}\frac{\rho}{\zeta^+}}\, \alpha_{n_{x'}}\left(\frac{\kappa}{n_{x'}}\frac{\zeta^+}{\rho}\{1-\varepsilon\}\right) + o_p(1) \overset{A}{\sim} \sqrt{\zeta^+}Z^+(x'), \tag{A.3}$$

where $\rho \equiv \lim_{n\uparrow+\infty} n_{x''}/n_{x'}$ and $Z^+(x')$ is a standard-normal random variable that by Assumption 3 is independent of $Z^+(x'')$. Combining (A.1) with (A.2) and (A.3) then gives

$$\sqrt{\kappa}(\zeta^+_{\iota n} - \zeta^\kappa) \overset{A}{\sim} \zeta^+ Z^+(x'') - \sqrt{\zeta^+}Z^+(x')$$

as claimed. □

Proof of Theorem 2. By Lemma 1 $\sqrt{\iota}(\zeta^-_{\iota n}(x',x) - \zeta^-(x',x))$ and $\sqrt{\iota}(\zeta^-_{\iota n}(x'',x) - \zeta^-(x'',x))$ have an asymptotic covariance that is equal to $\zeta^-(x',x)\zeta^-(x'',x)$ when $x' \neq x''$ and to $\zeta^-(x',x)(1+\zeta^-(x',x))$ otherwise. The result then follows from an application of the delta method. □

PROOF OF THEOREM 3. Since $\iota/n \downarrow 0$ as $n \uparrow +\infty$, and the empirical distribution function $F_n(y|x)$ is uniformly consistent, the estimators $H_n(y)$ and $D_n(y)$ are asymptotically equivalent to the intercept and slope coefficients from the infeasible regression of $F(y|x_i)$ on $\lambda_n^{\iota\kappa}(x_i)$ $(i = 1, 2, \ldots, q)$. Furthermore,

$$\sqrt{\iota}\,(D_n^{\iota\kappa}(y) - D(y)) = \frac{\sum_{i=1}^q \sqrt{\iota}\,(\lambda_n^{\iota\kappa}(x_i) - \lambda(x_i))\left(F(y|x_i) - \overline{F}(y|x)\right)}{\sum_{i=1}^q \left(\lambda(x_i) - \overline{\lambda}(x)\right)^2} + o_p(1)$$

and $\sqrt{\iota}\,(H_n^{\iota\kappa}(y) - H(y)) = -\overline{\lambda}(x)\sqrt{\iota}\,(D_n^{\iota\kappa}(y) - D(y)) + o_p(1)$, since $|\lambda_n^{\iota\kappa}(x) - \lambda(x)| = o_p(1)$ for all $x \in \mathcal{X}$. Therefore,

$$\sqrt{\iota}\,(D_n(y) - D(y)) = \sqrt{\iota}\,\frac{\left(\boldsymbol{F}(y) - \overline{F}(y|x)]\right)'}{\left(\boldsymbol{\lambda} - \overline{\lambda}(x)\right)'\left(\boldsymbol{\lambda} - \overline{\lambda}(x)\right)}\,(\boldsymbol{\lambda}_n - \boldsymbol{\lambda}) + o_p(1),$$

and so

$$\sqrt{\iota}\,(H_n^{\iota\kappa}(y) - H(y)) = \boldsymbol{\xi}_H\sqrt{\iota}\,(\boldsymbol{\lambda}_n^{\iota\kappa} - \boldsymbol{\lambda}) + o_p(1), \qquad \sqrt{\iota}\,(G_n^{\iota\kappa}(y) - G(y)) = \boldsymbol{\xi}_G\sqrt{\iota}\,(\boldsymbol{\lambda}_n^{\iota\kappa} - \boldsymbol{\lambda}) + o_p(1).$$

The result then follows. □

PROOF OF COROLLARY 2. Follows directly from Theorem 3 as its proof is uniform in y. □

REFERENCES

ALLMAN, E. S., MATIAS, C. and RHODES, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *Annals of Statistics* **37** 3099–3132.

ANDREWS, D. W. K. and SCHAFGANS, M. M. A. (1998). Semiparametric estimation of the intercept of a sample selection model. *Review of Economic Studies* **65** 497–517.

BOLLINGER, C. R. (1996). Bounding mean regressions when a binary regressor is mismeasured. *Journal of Econometrics* **73** 387–399.

BONHOMME, S., JOCHMANS, K. and ROBIN, J.-M. (2013). Nonparametric estimation of finite mixtures. Mimeo.

BORDES, L., MOTTELET, S. and VANDEKERKHOVE, P. (2006). Semiparametric estimation of a two-component mixture model. *Annals of Statistics* **34** 1204–1232.

BUTUCEA, C. and VANDEKERKHOVE, P. (2011). Semiparametric mixtures of symmetric distributions. Mimeo.

CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. and CRAINICEANU, C. (2006). *Measurement error in nonlinear models: a modern perspective*. Chapman and Hall, CRC Press.

CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and GALICHON, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika* **96** 559–575.

DEL BARRIO, E., DEHEUVELS, P. and VAN DE GEER, S. (2007). *Lectures on Empirical Processes*. European Mathematical Society.

EINMAHL, J. (1992). Limit theorems for tail processes with application to intermediate quantile estimation. *Journal of Statistical Planning and Inference* **32** 137–145.

EINMAHL, U. and MASON, D. (1997). Gaussian approximation of local empirical processes indexed by functions. *Probability Theory and Related Fields* **107** 283–311.

GASSIAT, E. and ROUSSEAU, J. (2013). Non parametric finite translation mixtures with dependent regime. Mimeo.

HALL, P. and ZHOU, X. H. (2003). Nonparametric identification of component distributions in a multivariate mixture. *Annals of Statistics* **31** 201–224.

HENRY, M., KITAMURA, Y. and SALANIÉ, B. (2010). Identifying finite mixtures in econometric models. Cowles Foundation Discussion Paper 1767.

HENRY, M., KITAMURA, Y. and SALANIÉ, B. (2013). Partial identification of finite mixtures in econometric models. *Quantitative Economics* (to appear).

HUNTER, D. R., WANG, S. and HETTMANSPERGER, T. P. (2007). Inference for mixtures of symmetric distributions. *Annals of Statistics* **35** 224–251.

Kasahara, H. and Shimotsu, K. (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica* **77** 135–175.

Lewbel, A. (2007). Estimation of average treatment effects with misclassification. *Econometrica* **75** 537–551.

Mahajan, A. (2006). Identification and estimation of regression models with misclassification. *Econometrica* **74** 631–665.

McLachlan, G. J. and Peel, D. (2000). *Finite mixture models*. John Wiley & Sons.

Rootzén, H. (2009). Weak convergence of the tail empirical process for dependent sequences. *Stochastic Processes and their Applications* **119** 468–490.

Department of Economics
The Pennsylvania State University
University Park, PA 16801
U.S.A.
E-mail: marc.henry@psu.edu

Département d'économie
Sciences Po
28 rue des Saints Pères
75007 Paris
France
E-mail: koen.jochmans@sciencespo.fr

Department of Economics
Columbia University
420 West 118th Street
New York, NY 10027
U.S.A.
E-mail: bsalanie@columbia.edu