Identification of Mixture Models Using Support Variations*

Xavier D'Haultfœuille † Philippe Février‡

First version: March 2010
This version: May 2015.

Abstract

We consider the issue of identifying nonparametrically continuous mixture models. In these models, all observed variables depend on a common and unobserved component, but are mutually independent conditional on it. Such models are important for instance in the measurement error, matching and auction literatures. Traditional approaches rely on parametric assumptions or strong functional restrictions. We show that these models are actually identified nonparametrically if the supports of the observed variables move with the true value of the unobserved component. Moreover, this "moving support" assumption is testable nonparametrically, using tools from extreme value theory. We develop an appropriate test, derive its asymptotic properties and conduct Monte Carlo simulations. Our approach complements the diagonalization technique introduced by Hu & Schennach (2008), which allows to obtain similar results under different assumptions.

Keywords: mixture models, nonparametric identification, support variations, measurement error, auctions, matching.

JEL classification numbers: C14, D44.

*We would like to thank Jean-Marc Robin, Bernard Salanié, Susanne Schennach, James Stock, as well as the editor and two referees for their useful comments.
†CREST. E-mail address: xavier.dhaultfoeuille@ensae.fr.
‡CREST. E-mail address: fevrier@ensae.fr.

1 Introduction

In this paper, we consider continuous mixture models where all observed variables depend on a common and unobserved component, but are mutually independent conditional on it. Such models have important applications in economics. The main one is probably the measurement error model, in which extensive attention has been devoted to identifying the effect of an unobserved variable when only measures of it are available. Other applications include auctions with unobserved heterogeneity, or matching models. While traditional approaches rely on parametric assumptions or functional restrictions, we introduce a very simple sufficient condition for the model to be identified. More precisely, we suppose that the observed variables have a compact support that moves with the unobserved variable. When this "moving support" assumption is satisfied, and a necessary normalization is imposed, the model is identified without any other restriction.

We believe that our identification result is interesting for several reasons. First, the moving support assumption is naturally satisfied in different economic models. This is for example the case in the matching literature. Building on Becker (1973) result, Shimer & Smith (2000) derive sufficient conditions to extend assortative matching in an environment with search frictions.[1] In this model, at equilibrium, workers match with firms of different qualities. The set of firms with which a worker can match is increasing in the own quality of the worker (see Figure 1 in Shimer & Smith's paper) and the moving support assumption is satisfied. Similarly, in an auction model with a reserve price unobserved by the econometrician, both the lower and upper bounds of the bids vary with the unobserved reserve price (Riley & Samuelson, 1981). Second, our assumption is easy to interpret economically. In the measurement error model, the underlying idea is that the mismeasured variable cannot be too far from the true value of the variable. Finally, we also show how the moving assumption can be tested formally. Indeed, it implies a moving support condition on the observed variables. Such a condition can then be tested using results from the statistical literature on extreme values (see, e.g., Embrechts et al., 1997). We derive the asymptotic properties of our test and conduct Monte Carlo simulations. These simulations indicate that our test discriminates correctly data generating processes satisfying the moving support assumption against other ones.

Our paper is related to results on the identification of measurement error models. While this literature is vast (see, e.g., Carroll et al., 2006 for a survey), most of the papers focus

[1]We thank Jean-Marc Robin for suggesting us this example.

on the case of classical measurement errors, for which errors are either independent of the mismeasured variable or have a zero mean conditional on it (see, e.g., Hausman et al., 1991, Li, 2002, Schennach, 2004 and 2007). Yet, this assumption is likely to fail in many context (see, e.g., Bound & Krueger, 1991). Building on the ideas of Hu (2008), Hu & Schennach (2008) explain how to recover the effect of the true variable in the general case of nonclassical measurement errors with continuous variables. Under an injectivity condition on integral operators, they show that identification can be achieved through an eigenvalue-eigenfunction decomposition. Our approach complements Hu an Schennach's one in the sense that for some models, our condition is satisfied while theirs fails to hold, and conversely. The merits of the moving support assumption, over the injectivity condition, are its simple economic meaning and its testability. In contrast, no empirical test of the injectivity condition has been proposed yet.

Our paper is also connected to the recent literature on the nonparametric identification of finite mixture models. In particular, Allman et al. (2009) show that under the same conditional independence as ours and some rank conditions, nonparametric mixture models with a known number of unobserved components are identifiable from an observed vector of at least three components (see also Hall et al., 2005 and Bonhomme et al., 2014*a*).[2] Henry et al. (2014) derive sharp bounds on the mixture weights and components under a weaker version of the conditional independence condition. Finally, papers more closely related to ours are the one of Shneyerov & Wong (2011) and Hu & Sasaki (2014). Even if their approach is distinct from ours, they also use the fact that the upper or lower bound of the support of observed variables is a strictly increasing function of the unobserved component. All these papers heavily rely on the fact that the mixture component has a finite support, and do not seem to generalize easily to the continuous case on which we focus here.[3]

The paper is organized as follows. Section 2 presents the model, our main identification result. Section 3 develops some extensions where we weaken our initial moving support condition and consider the case of discrete mixtures. Section 4 deals with inference, by first presenting a multistep nonparametric estimator and then developing a test of the moving support condition. Section 5 concludes. All proofs are deferred to the appendix.

[2]Allman et al. (2009), Kasahara & Shimotsu (2009) and Bonhomme et al. (2014*b*) also provide identification results for models with additional dependence between the observed variables, such as hidden Markov models.

[3]Actually, the result of Hu & Sasaki (2014) also applies to continuous mixtures, but their Restriction 1 is unlikely to hold in practice with continuous mixtures.

2 Main identification result

We first define the general mixture model on which we focus. Let us consider K real random variables $(X_1, ...X_K)$ which are observed by the econometrician. All these variables depend on a real continuous variable X^*, which is unobserved. The aim of the econometrician is to recover the distribution of X_k conditional on X^*, and possibly the distribution of X^*. We rely for that purpose on the following conditions. Hereafter, $F_{X_k|X^*}$ denotes the cumulative distribution function (cdf) of X_k conditional on X^*.

Assumption 1 *$K \geq 3$ and $(X_1, ..., X_K)$ are mutually independent conditional on X^*.*

Assumption 2 *X^* has a continuous distribution, with continuous density f_{X^*} on its support. For all $k \in \{1, ..., K\}$ and u in the support of X^*, the support of X_k conditional on $X^* = u$ is an interval and we denote it by $[\underline{X}_k(u), \overline{X}_k(u)]$. Moreover, $u \mapsto F_{X_k|X^*}(x|u)$ is continuous for all x.*

Assumption 3 *(i) $\underline{X}_1(.)$ and $\overline{X}_2(.)$ are strictly increasing, (ii) $\underline{X}_3(.)$ and $\overline{X}_3(.)$ are strictly increasing.*

Assumption 4 $\underline{X}_1(X^*) = X^*$.

Assumption 1 defines the mixture structure. Assumption 2 imposes mild regularity conditions on the distributions. Our main condition is Assumption 3. We refer to it, or sometimes to Assumption 3-(i) only, as the "moving support" condition.[4] The fact that the bounds of the support are increasing functions of X^* reflects the positive link between X^* and the Xs, which would be formally expressed through the monotone likelihood ratio property, for instance.[5] Assumption 3 holds if we reinforce this positive link by stating that higher values of X^* lead to strictly higher values of X_k, i.e., that the support of the observed variables move with the true value of the unobserved component. In a nonseparable model $X_k = \varphi_k(X^*, \varepsilon_k)$ where $\varepsilon_k \in [\underline{\varepsilon}_k, \overline{\varepsilon}_k]$ and is independent of X^* ($k \in \{1, 2, 3\}$), the moving support condition holds for instance if $\varphi_i(., \underline{\varepsilon}_i)$, $\varphi_j(., \overline{\varepsilon}_j)$ and $\varphi_k(X^*, .)$ are strictly increasing, for $(i, j, k) \in \{1, 3\} \times \{2, 3\} \times \{1, 2, 3\}$.

Finally, we fix $\underline{X}_1(X^*)$ to be equal to X^*. This is a normalization that is without loss of generality under a mild additional requirement. Specifically, if X^* satisfies Assumptions

[4] We separate the conditions into two parts, because as we will see in Subsection 3.1, Condition (i) alone suffices for identifying some features of the model.

[5] In our context, this property would amount to supposing that $x \mapsto \frac{f_{X_k|X^*}(x|x_1^*)}{f_{X_k|X^*}(x|x_0^*)}$ is increasing, $f_{X_k|X^*}$ denoting the conditional density of X_k.

1-3 and $\underline{X}_1^{-1}(.)$ is continuously differentiable, Assumptions 1-3 but also 4 hold when using $\widetilde{X}^* = \underline{X}_1(X^*)$ instead of X^*. In the previous case $X_1 = \varphi_1(X^*, \varepsilon_1)$, this would mean replacing X^* by $\widetilde{X}^* = \varphi_1(X^*, \underline{\varepsilon}_1)$. We would then write $X_k = \widetilde{\varphi}_k(\widetilde{X}^*, \varepsilon_k)$, with $\widetilde{\varphi}_k(\widetilde{x}, u) = \varphi_k(g^{-1}(\widetilde{x}), u)$ and $g(x) = \varphi_1(x, \underline{\varepsilon}_1)$. This normalization may be seen as a particular case of the one considered by Hu & Schennach (2008), who impose that a functional of the distribution of, say, X_1 conditional on X^*, is equal to X^*. In our case, the functional is the lower bound of the support of X_1^*. Considering this normalization rather than another one makes sense given Assumption 3-(i). Assumptions 1-3 alone would not guarantee that another functional of $F_{X_1|X^*}$ is strictly increasing in X^*.[6]

Theorem 2.1 *Under Assumptions 1-4,* $F_{X_1|X^*}, ..., F_{X_K|X^*}$ *and* F_{X^*} *are identified.*

While the detailed proof is given in appendix, let us provide some intuition on the result. First, as shown by Figure 1, the range of X^* compatible with an observation $X_1 = x$ is limited, and so is the range of X_2 that one can observe in the data when $X_1 = x$. More formally, observing $X_1 = x$, we know that $X^* \leq x$, and therefore $X_2 \leq \overline{X}_2(X^*) \leq \overline{X}_2(x)$. The shaded area on Figure 1 represents the set of values of (X^*, X_2) compatible with $X_1 = x$.[7]

[6] Of course, if the value of X^* has a precise economic meaning (e.g., the true level of a variable measured with error by X_1), it may be more natural to assume that $\underline{X}_1(X^*) = h(X^*)$, for some known, strictly increasing $h(.)$. Then all our analysis would go through with this normalization instead of $\underline{X}_1(X^*) = h(X^*)$.

[7] In Figure 1, we have imposed the random variables to be positive for illustration purpose only.

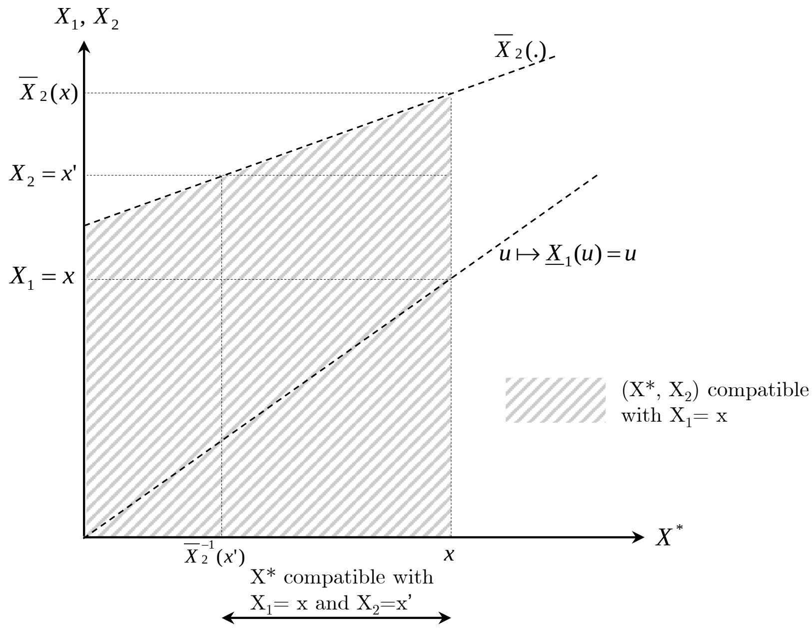

Figure 1: Values of X_2 and X^* compatible with $X_1 = x$

We also represent in Figure 1 the range of possible values for X^* when $X_1 = x$ and $X_2 = x'$, which is $[\overline{X}_2^{-1}(x'), x]$. It turns out that when $X_1 = x$ and $X_2 = \overline{X}_2(x)$, the only possible value for X^* is x. When $K \geq 3$, it is then possible to identify $F_{X_k|X^*}(.|x)$, for all x and $k \geq 3$, by looking at the distribution of X_k conditional on $X_1 = x$ and $X_2 = \overline{X}_2(x)$. Indeed, by the conditional independence assumption,[8]

$$F_{X_3|X_1,X_2}(x_3|x, \overline{X}_2(x)) = F_{X_3|X_1,X_2,X^*}(x_3|x, \overline{X}_2(x), x) = F_{X_3|X^*}(x_3|x). \tag{2.1}$$

By using Assumption 3-(ii), we can identify similarly $F_{X_1|X^*}$ and $F_{X_2|X^*}$. Finally, to recover the distribution of X^*, we use the fact that by conditional independence of X_1 and X_2,[9]

$$f_{X^*}(x) = \frac{f_{X_1,X_2}(x, \overline{X}_2(x))}{f_{X_1|X^*}(x|x) f_{X_2|X^*}(\overline{X}_2(x)|x)}.$$

Because all terms on the right-hand side are identified by the previous steps, the distribution of X^* is identified.

[8]This equality is not rigorous because the density f_{X_1,X_2} is equal to zero at $(x, \overline{X}_2(x))$. To overcome this issue, we have to consider instead the event $(X_1, X_2) \in A_\delta(x_1) = [x_1 - \delta, x_1 + \delta] \times [\overline{X}_2(x_1 - \delta), \overline{X}_2(x_1 + \delta)]$, and let $\delta \to 0$. The formal proof is given in the appendix.

[9]Once again, this equation is not rigorous as the ratio may not be properly defined. The formal proof considers a limit reasoning to circumvent this issue.

Theorem 2.1 is similar to the identification result of Hu & Schennach (2008). However, whereas Hu & Schennach (2008) rely on the injectivity of integral operators (see their Assumption 3), we mostly use the moving support condition. Our results are complementary, as both conditions are distinct. Consider the standard measurement error model $X_k = X^* + \varepsilon_k$, where $(X^*, \varepsilon_1, ..., \varepsilon_K)$ are independent and X^* has a compact support $[\underline{x}, \overline{x}]$. In this setting, we can show that their injectivity condition fails to hold when ε_k is uniform, basically because the characteristic function of ε_k has zeros on the real line. On the contrary, Theorem 2.1 applies in this case. Conversely, their injectivity condition holds when $\varepsilon_k \sim \mathcal{N}(0, \sigma_k^2)$, while we cannot apply Theorem 2.1.

We also believe that the moving support condition is more intuitive than the alternative approach of Hu & Schennach (2008). Not much is known about the injectivity condition. It is closely related to the completeness condition used in additive instrumental nonparametric models to secure identification. This latter condition holds in exponential models (see Newey & Powell, 2003), or in nonlinear models under an additive decomposition and a large support condition, but under restrictive technical conditions (see D'Haultfœuille, 2011). No theoretical result has been obtained otherwise. A merit of our condition is, on the contrary, its simple economic meaning. Moreover, the moving support assumption is testable (see Section 4.2) whereas no empirical test has been proposed for the injectivity condition yet.

Example 1: auction models with unobserved heterogeneity

Let us consider a good which is sold by an auction mechanism. This good has a characteristic X^* which is observed by the K bidders and affects their valuation $(V_1, ..., V_K)$. Conditional on X^*, $(V_1, ..., V_K)$ are independent, but may be non identically distributed if bidders are asymmetric. The econometrician observes the bids $B_k(= X_k) = b_k(V_k, X^*)$ but neither $(V_1, ..., V_K)$ nor X^*. In such a case, $(B_1, ..., B_K)$ are independent conditional on X^*. The ultimate goal in this literature is to recover the distribution of V_k conditional on X^*, and potentially the distribution of X^*. In general, the function b_k is known by the theory and it is thus sufficient to recover the distributions of X^* and B_k conditional on X^*. Such auction models with unobserved heterogeneity have been studied recently by Krasnokutskaya (2011) and Hu et al. (2013), the latter applying Hu & Schennach's methodology. We argue that Assumption 3 is likely to hold in this context, and present two examples supporting this claim.

Suppose first, following Krasnokutskaya (2011), that $V_k = X^* \times \varepsilon_k$, where the ε_k are i.i.d.

with support $[\underline{\varepsilon}, \overline{\varepsilon}]$, with $0 < \underline{\varepsilon} < \overline{\varepsilon} \leq \infty$ (Krasnokutskaya, 2011 actually imposes that $\overline{\varepsilon} < \infty$). The model does not change if we divide ε_k by a positive constant and multiply X^* accordingly, so we can suppose without loss of generality that $\underline{\varepsilon} = 1$. This normalization is convenient and leads to Assumption 4, as we shall see. In such a model, the equilibrium function $b(.,.)$ is given by:

$$b(V_k, X^*) = \left[\frac{(K-1)\int_{\underline{\varepsilon}}^{\varepsilon_k} f_\varepsilon(u) F_\varepsilon^{K-2}(u) u du}{F_\varepsilon^{K-1}(\varepsilon_k)}\right] X^*.$$

The bidding strategy is increasing in the valuation, and some algebra show that

$$X^* \leq b(V_k, X^*) \leq \left[(K-1)\int_{\underline{\varepsilon}}^{\overline{\varepsilon}} u f_\varepsilon(u) F_\varepsilon^{K-2}(u) du\right] X^*.$$

In other words, conditional on X^*, the bids $(X_1, ..., X_K) = (b(V_1, X^*), ..., b(V_K, X^*))$ belong to the set $[X^*, \overline{B}X^*]^n$, with $\overline{B} = (K-1)\int_{\underline{\varepsilon}}^{\overline{\varepsilon}} u f_\varepsilon(u) F_\varepsilon^{K-2}(u) du$. Because $\overline{B} < (K-1)\int_{\underline{\varepsilon}}^{\overline{\varepsilon}} u f_\varepsilon(u) du$, $\overline{B} < \infty$ provided that $E(\varepsilon_1) < +\infty$, both the lower and upper bound are strictly increasing, even if $\overline{\varepsilon} = +\infty$. Assumptions 3 and 4 are satisfied, and one can show that Assumption 2 holds as well. Therefore, Theorem 2.1 applies, and the distribution of $B_k = X_k$ conditional on X^*, for $k \in \{1, ..., K\}$, is identified.

Another example where the moving support condition is satisfied is when X^* is the reserve price, supposed to be unobserved by the econometrician. Suppose that N potential risk neutral and symmetric bidders with valuations $(V_1, ..., V_N) \in [\underline{V}, \overline{V}]^N$ participate to this auction. We denote by F_V (resp. f_V) the cdf (resp. probability distribution function) of V_k. Finally, we suppose that before bidding, the bidders learn the number K of effective bidders i.e. the number of bidders with valuations greater than X^*. In such a case, the equilibrium function $b(.,.,.)$ is given by :

$$b(V, X^*, K) = (K-1)\frac{\int_{X^*}^{V} u f_V(u) F_V^{K-2}(u) du}{F_V^{K-1}(V)}.$$

Hence, for all K and conditional on X^*, the observed bids $X_k = b(V_k, X^*, K)$ belong to the set $[X^*, b(\overline{V}, X^*, K)]^K$. As before, $b(\overline{V}, X^*, K)$ is finite as soon as $E(V_1) < +\infty$. Hence, both bounds are strictly increasing in X^* and Assumptions 3 and 4 are a consequence of the theoretical model.

Example 2: matching models

Let us consider the matching model with search frictions developed by Shimer & Smith (2000). In their model, they assume a continuum of heterogeneous agents. Two agents of type X^* and Y can match to produce $f(X^*, Y)$, where f is strictly increasing in both arguments. At each instant, agents are either matched or unmatched, and nature destroys any match with a positive probability. Unmatched agents constitute the pool of searchers that are trying to form new matches. Shimer & Smith characterize the equilibrium matching sets. They prove in particular that under regularity and supermodularity assumptions, positively assortative matching is ensured.[10] This, in turn, implies that the lower and upper bound functions of the matching set, $\underline{Y}(.)$ and $\underline{Y}(.)$, are nondecreasing,[11] as depicted by Figure 1 in Shimer & Smith's paper.

Let us suppose that the econometrician observes several matches between, for example, firms and workers on the job market, and their corresponding matching outputs.[12] Hence, for a firm of unobserved type X^*, the econometrician observes several wages of different workers, $(X_1, ..., X_K) = (f(X^*, Y_1), ..., f(X^*, Y_K))$. The aim is to recover features of the matching function, to identify for instance what are the relative contributions of workers and firms on the production function.

Given the results of Shimer & Smith (2000) , the support of X_k conditional on X^* is given by $[f(X^*, \underline{Y}(X^*)), f(X^*, \overline{Y}(X^*))]$. Because $x \mapsto f(x, y)$ is strictly increasing whereas $\underline{Y}(.)$ and $\overline{Y}(.)$ are nondecreasing, Assumption 3 is satisfied. Under a normalization such as Assumption 4, the distribution of X_k conditional on X^* and the distribution of X^* are therefore identified.

3 Extensions

We consider in this section various extensions, mostly related to the moving support condition. Theorem 2.1 shows identification of the whole model with four varying bounds on three variables. We investigate below what can be identified if only two or three bounds are varying. We also consider the case of weakly instead of strictly monotonic bounds and situations where the support of X_k conditional on X^* is not compact. Finally, we show

[10]For more details, see their Proposition 6.

[11]For more details, see their Proposition 3.

[12]Wages are usually observed, rather than the production itself. However, in simple models, there is a one-to-one relationship at equilibrium between wages and outputs.

that the same methodology applies if X^* is discrete.

3.1 Two or three varying bounds with asymmetric variables

We first consider a weaker version of Assumption 3, where the support of X_3 may not vary with X^*, or where only one of its bound varies with X^*. Theorem 3.1 shows that Assumption 3-(i) is actually sufficient to recover $F_{X_k|X^*}$, for $k \geq 3$. It also establishes that a single varying bound on X_3 is sufficient to identify f_{X^*}, under regularity conditions related to this varying bound. Henceforth, we denote by $\text{Supp}(U)$ the support of any random variable U.

Assumption 5 $\underline{X}_3(X^*) = c > -\infty$, $\overline{X}_3$ *is strictly increasing and there exists* $m \geq 1$ *such that for all* x^*, $F_{\overline{X}_3^{-1}(X_3)|X^*}(.|x^*)$ *is* m *times differentiable at* x^*, $F^{(m)}_{\overline{X}_3^{-1}(X_3)|X^*}(x^*|x^*) > 0$ *and*

$$(x_3, x^*) \mapsto \left| F^{(m)}_{\overline{X}_3^{-1}(X_3)|X^*}(x_3|x^*) / F^{(m)}_{\overline{X}_3^{-1}(X_3)|X^*}(x^*|x^*) \right|$$

is bounded on Supp(X_3, X^*).

Theorem 3.1 *Under Assumptions 1, 2, 3-(i) and 4,* $F_{X_3|X^*}, ..., F_{X_K|X^*}$ *is identified. If Assumption 5 also holds, then* f_{X^*} *is identified.*

The intuition of the first part of Theorem 3.1 is the same as in Theorem 2.1. A point not emphasized in the theorem is that to identify $F_{X_k|X^*}$, we actually only rely on the conditional independence between (X_1, X_2) and $X_k (k \geq 3)$, not on the mutual conditional independence of $(X_1, ..., X_K)$. In particular, X_1 and X_2 can be correlated conditional on X^*, as long as their joint conditonal support remains of the form $[\underline{X}_1(u), \overline{X}_1(u)] \times [\underline{X}_2(u), \overline{X}_2(u)]$.

The idea of the second part of Theorem 3.1 is that the knowledge of F_{X_3} and of $F_{X_3|X^*}$ allows us, under Assumption 5, to identify f_{X^*}, through the integral equation

$$F_{X_3}(x_3) = \int_{\overline{X}_3^{-1}(x_3)}^{+\infty} F_{X_3|X^*}(x_3|x^*) f_{X^*}(x^*) dx^*, \tag{3.1}$$

which holds for all $x_3 \in \text{Supp}(X_3)$. The fact that the lower bound of the integral moves with x_3 ensures that this integral equation has a unique solution, under Assumption 5. Note that without any restriction between X_3 and X^*, the knowledge of these two distributions may not be sufficient to recover f_{X^*}. If for instance X_3 and X^* are independent, (3.1) does not provide any information on f_{X^*}. Of course, other restrictions than Assumption 5 would be possible to identify f_{X^*} through (3.1).

Example 3: measurement error models with repeated measures

The previous framework is well suited to measurement errors on a covariate. Typically, we seek to measure the effect of a variable X^* on an outcome $Y(= X_3) = f(X^*, \nu)$ but only observe two measures X_1 and X_2 of X^*, so that $X_k = \varphi_k(X^*, \varepsilon_k)$, $k \in \{1, 2\}$ (see, e.g., Hausman et al., 1991 or Schennach, 2004, for papers studying models with repeated measures). In this case, Assumption 1 is satisfied if $(\nu, \varepsilon_1, \varepsilon_2)$ are independent conditional on X^*. As mentioned above, however, conditional independence betwen ε_1 and ε_2 is not necessary for Theorem 3.1 to apply, as long as they remain jointly independent of ν and their joint support is a rectangle.

Denote by $\underline{\varepsilon}_1$ and $\overline{\varepsilon}_2$ the lower and upper bounds of the supports of ε_1 and ε_2, respectively. Assumption 3-(i) is fulfilled if $\varphi_k(x, .)$, $x \mapsto \varphi_1(x, \underline{\varepsilon}_1)$ and $\mapsto \varphi_2(x, \overline{\varepsilon}_2))$ are strictly increasing. These conditions hold in the classical measurement error model or with multiplicative errors, as soon as the error terms are either bounded below or above. If $X_k = X^* + \varepsilon_k$ (resp. $X_k = X^* \times \varepsilon_k$) and $\varepsilon_k \in [\underline{\varepsilon}, \overline{\varepsilon}](k = 1, 2)$, the support of X_k conditional on X^* is $[X^* + \underline{\varepsilon}, X^* + \overline{\varepsilon}]$ (resp. $[X^* \times \underline{\varepsilon}, X^* \times \overline{\varepsilon}]$) and it changes with X^*. Assumption 3-(i) is also satisfied if there is systematic over-reporting for one of the measurement, so that $X_1 \geq X^*$, and systematic under-reporting for the other measurement, so that $X_2 \leq X^*$. Actually, Assumption 3-(i) only requires under-reporting and over-reporting up to some strictly increasing function. Underreporting is plausible when reporting consumption is costly for individuals, because they have to indicate it on a diary for instance (see e.g. Yang et al., 2010). As explained by Hu & Schennach (2008), it may also be the case that tobacco consumption is systematically under-reported by people, for instance. If individuals smoke at most α% (with $\alpha > 0$) of the cigarettes they buy, Assumption 3-(i) would hold with X_1 the number of cigarettes they buy and X_2 the number of cigarettes they report to smoke.[13]

Example 4: measurement error models with instrumental variables

Alternatively, we could only observe one measure of X^*, $X_1 = \varphi(X^*, \varepsilon)$, and an instrument $Z = X_2$ of X^*, such that $X^* = \psi(Z, \eta)$ (see, e.g., Newey, 2001, Schennach, 2007 and Hu & Schennach, 2008, for studies of instrumental models with measurement errors). Assumption 1 is satisfied if $(Z, \nu, \varepsilon, \eta)$ are independent. Assumption 1 is equivalent to Assumption 2

[13] In this example, the normalization $\overline{X}_2(X^*) = X^*$ would be more appropriate than $\underline{X}_1(X^*) = X^*$.

of Hu & Schennach (2008), so that our framework is identical to theirs. In addition to standard instruments, our framework also encompasses the one considered by Schennach (2013) where Z would be a Berkson measurement error, so that $X^* = Z + \eta$, while X_1 would be another variable related to X^*.[14]

Define $\underline{\varphi}(x) = \inf\{\varphi(x,u), u \in \text{Supp}(\varepsilon)\}$ and $\underline{\psi}(z) = \inf\{\psi(z,u), u \in \text{Supp}(\eta)\}$. Because $X_1 \geq \underline{\varphi}(X^*)$ and $\underline{\psi}(Z) \leq X^*$, Assumption 3-(i) holds if $\underline{\varphi}(.)$ and $\underline{\psi}(.)$ are strictly increasing. This is the case for instance if φ and ψ are additively separable, and $\inf \text{Supp}(\varepsilon) > -\infty$ and $\inf \text{Supp}(\eta) > -\infty$. An example where the upper bound of the support of Z moves with X^* is still related to consumption. If X^* denote consumption that we only partially observe, Z could correspond to the part of X^* we observe. For instance X^* may be alcohol, tobacco and drugs consumption, of which we only observe alcohol and tobacco (Z). By construction, we then have $X^* \geq Z$.

3.2 Two or three varying bounds with symmetric variables

In some cases, the moving support assumption may be natural for one bound but not for the other. We show below that it is possible to extend the result of Theorem 3.1, in the symmetric case where X_1 and X_2 are identically distributed conditional on X^*, to situations where only one bound is strictly increasing. We consider here a moving lower bound, but the proof would be similar with a moving upper bound.

Assumption 6 *X_1 and X_2 are identically distributed conditional on X^*, $\overline{X}_1 = \overline{X}$ is constant but $\underline{X}_1(.)$ is strictly increasing, with the normalization $\underline{X}_1(X^*) = X^*$. Moreover, there exists $m > 0$ such that $F_{X_1|X^*}$ is $m+1$ times differentiable, with $F^{(m)}_{X_1|X^*}(\underline{X}_1(x)|x) \neq 0$ and $F^{(j)}_{X_1|X^*}(.|x)$ is bounded for $j \in \{1,...,m+1\}$.*

Theorem 3.2 *Under Assumptions 1 and 6, $F_{X_k|X^*}$ is identified for $k \in \{3,...,K\}$. If Assumption 6 also holds for X_3, then we can identify f_{X^*} as well.*

The proof of Theorem 3.2 is related but a bit different from that of Theorem 2.1. The reason is that it is impossible, with only one bound moving, to pin down directly X^* by choosing appropriately X_1 and X_2. Still, the observations provide some inequality restrictions on X^*. The key insight is that these restrictions entail a relationship between the left and right partial derivatives of the joint density of the data at appropriate points. We are then

[14]A typical example of a Berkson measurement errors is when we observe the group-average exposure to a pollutant instead of the individual exposure.

able to get rid of the integral expression of the mixture, as in Theorem 2.1, to identify $F_{X_k|X^*}$.

Example 1 (continued)

Coming back to auction models with unobserved heterogeneity, the assumption that $\underline{\varepsilon} > 0$ in the multiplicative model $X_k = V_k \times \varepsilon_k$ may be strong in some cases. If $\underline{\varepsilon} = 0$, $\overline{X}_k(.)$ is still strictly increasing, as shown previously. We can therefore still use Theorem 3.2 to identify the model.

Example 5: wage decompositions

This example is related but distinct from the matching models example. Suppose that we are interested in the link between fixed wages and bonuses for, say, salesmen. This question is related to the issue of selection of workers by firms and has received a lot of attention in the personnel literature (see, e.g., Prendergast, 1999 and Lazear & Shaw, 2007, for literature reviews). Suppose that we do not observe fixed wages and bonuses, but only the total wages. If salesmen of the same firm face the same contract, then we can let X^* denote the fixed wage provided by the firm, while $(X_1, ..., X_K)$ correspond to the wages of K employees of the firm. By construction, $X_k \geq X^*$. The symmetry condition between X_k is also natural in this setting. By Theorem 3.2, the joint distribution of bonuses and fixed wages $(X_k - X^*, X^*)$ is then identified, under the boundary condition of Assumption 6 and if $K \geq 3$. This allows one to quantify the importance of bonuses compared to the fixed part of wages, and to identify whether there is a positive or negative dependence between fixed wages and the bonuses.

3.3 Weakly increasing supports

We consider here the case where one bound, say $\overline{X}_2$, is flat on some subset $\mathcal{I}$ of the support of X^*. We still suppose that $\underline{X}_1(.)$ is strictly increasing and that Assumption 4 holds. In such a case, we still identify, for all x_1 in the support of X_1, $\overline{X}_2(x_1)$ as the maximum of the support of X_2 conditional on $X_1 = x_1$. We thus identify $\mathcal{I}$ as the region where $\overline{X}_2(.)$ is flat. Then the same reasoning as in the beginning of the proof of Theorem 2.1 can be applied, and we identify $F_{X_k|X^*}(x|x^*)$ for all $x^* \notin \mathcal{I}$ and $k \geq 3$. On the other hand, it is unclear to us whether we can recover $F_{X_k|X^*}(x|x^*)$ for $x^* \in \mathcal{I}$ or f_{X^*} in this context.

Finally, we may wonder whether the model is still identified without any variation in the support, namely under Assumptions 1 and 2 only. The following counter-example shows that this is not the case. Further restrictions, such as ours or the injectivity assumption of Hu & Schennach (2008), are thus necessary to identify nonparametric mixture models. Perhaps surprisingly, this is true no matter how large is K, the number of variables that we observe.

Example 6: non-identification without additional restrictions

Suppose that $X_k = X^* + \varepsilon_k$ for $k \in \{1, ..., K\}$, with $(X^*, \varepsilon_1, ..., \varepsilon_K)$ mutually independent and $(\varepsilon_1, ..., \varepsilon_K)$ identically distributed. Assume further that X^* has the density function $f_{X^*}(x) = (1 - \cos(x))/(\pi x^2)$ and ε_k have the density function $f_\varepsilon(x) = f_{X^*}(x/2K)/2K$. As a normalization, suppose that the distribution of X^* is known. Then Assumptions 1 and 2 are identified. Yet, the conditional distributions $(F_{X_k|X^*})_{k=1,...,K}$ are not identified, as shown in the appendix. Basically, we prove that the convolution $q(X^*) + \varepsilon_k$, for a well chosen functions q different from the identity function, yields the same joint distribution for $(X_1, ..., X_K)$. This is due to the fact that a distribution which is a convolution may not be decomposed in a unique way, and this remains true even in a multivariate setting. Note that in this example, both the injectivity assumptions of Hu & Schennach (2008) and our moving support conditions fail to hold.

3.4 Non-compact measurement errors

Non-compact measurement errors may be a concern because they affect any empirical strategy relying on the boundaries of the support. Suppose for instance that (X_1, X_2, X_3) satisfy Assumptions 1, 2 and 3-(i), but X_1 is measured with error by $\widetilde{X}_1 = X_1 + \eta$, and η error has a non compact support. In this case, $(\widetilde{X}_1, X_2)$ does not satisfy Assumption 3-(i), since $\inf \text{Supp}(\widetilde{X}_1|X^*) = -\infty$.

However, in such a case, it is still possible to identify the model if we impose restrictions on η. Following Schwarz & Van Bellegem (2010), suppose for instance that $\eta|X_1, X_2 = x_2 \sim \mathcal{N}(0, \sigma^2)$, where σ^2 is unknown. Because the support of X_1 conditional on $X_2 = x_2$ is compact, Theorem 2.1 of Schwarz & Van Bellegem (2010) implies that σ^2 and the distribution of X_1 conditional on $X_2 = x_2$ are identified. We can then proceed as previously to fully identify the model.

3.5 Discrete mixtures

Though we mainly focus on a continuous mixture here, the idea behind our results also applies when X^* has a finite support $\{x_1^*, ..., x_J^*\}$, with $x_1^* < ... < x_J^*$. Specifically, suppose that $\underline{X}_1(.)$ and $\overline{X}_2(.)$ are strictly increasing on $\{x_1^*, ..., x_J^*\}$, and consider as previously the normalization $\underline{X}_1(x_j^*) = x_j^*$. First, we identify the x_j^* and $\overline{X}_2(x_j^*)$ by noting that $x_1 \mapsto \max \text{Supp}(X_2|X_1 = x_1)$ is constant on the J intervals $[x_j^*, x_{j+1}^*)$ (where we let $x_{J+1}^* = +\infty$). The corresponding values $\overline{X}_2(x_j^*)$ and the points x_j^* where this function is discontinuous are therefore identified.

Second, we basically identify $F_{X_3|X^*}$ by conditioning on (X_1, X_2) sufficiently far from each other. Remark that $X_1 < x_{j+1}^*$ and $X_2 > \overline{X}_2(x_{j-1}^*)$ implies $X^* \leq x_j^*$ and $X^* \geq x_j^*$, so that $X^* = x_j^*$. Hence, by conditional independence,

$$F_{X_3|X_1 < x_{j+1}^*, X_2 > \overline{X}_2(x_{j-1}^*)}(x_3) = F_{X_3|X^*}(x_3|x_j^*).$$

Finally, we also recover $P(X^* = x_j^*)$ if the cdfs $(F_{X_3|X^*}(.|x_j^*))_{j=1...J}$ are linearly independent, using

$$F_{X_3}(x_3) = \sum_{j=1}^{J} P(X^* = x_j^*) F_{X_3|X^*}(x_3|x_j^*).$$

This result is related to a recent result of Hu & Sasaki (2014), who show that the distributions of X_1 and X_2 can allow, under in particular some support conditions, to recover the distribution of X^*. An advantage of their approach is that it does not require a third observation. A related result is also the one of Shneyerov & Wong (2011) in first-price auctions with an unknown number of potential bidders (which corresponds to X^* here). Even if their approach is distinct from ours, they also use the fact that the upper bound of the support of observed variables is a strictly increasing function of the unobserved component.

4 Inference

Though the paper is mostly focused on identification, we sketch in this section how inference can be conducted. We first present possible nonparametric estimators of $F_{X_k|X^*}$ and f_{X^*}. We then consider a test of the moving support condition.

4.1 Estimation

We first present a possible multistep nonparametric method, under the standard condition that we observe a sample $(X_{1i}, ..., X_{Ki})_{i=1...n}$ of independent and identically distributed variables. Our method is suitable if Assumption 3-(i) holds, together with either Assumption 3-(ii) or Assumption 5. It does not cover the symmetric case where only $\underline{X}_1(.) = \underline{X}_2(.)$ is strictly increasing.

We focus here on the estimation of X_k conditional on X^* for $k \geq 3$, and on the distribution of X^*. Following the identification strategy, we estimate, in a first step, the bound $\overline{X}_2(.)$. This bound can be obtained by noting that $\overline{X}_2(.)$ is the maximum of the support of X_2 conditional on $X_1 = x$. Because $\overline{X}_2(.)$ is assumed to be strictly increasing, we can use any frontier estimation method. The literature on this topic is large and the statistical properties of the estimators are now well established (see, e.g., Simar & Wilson, 2008, for a survey). We consider here the popular free disposal hull estimator introduced by Deprins et al. (1984). Let us first consider

$$\widehat{q}_\alpha(x_1) = \inf\{x_2 | \widehat{F}_{X_1,X_2}(x_1, x_2)/\widehat{F}_{X_1}(x_1) \geq \alpha\},$$

for any $\alpha \in [0, 1]$, and where $\widehat{F}_{X_1,X_2}$ (resp. $\widehat{F}_{X_1}$) is the empirical cdf of (X_1, X_2) (resp. of X_1). $\widehat{q}_\alpha(x_1)$ is thus the empirical quantile of $X_2|X_1 \leq x_1$. The free disposal hull estimator is simply defined by

$$\widehat{\overline{X}}_2(x_1) = \widehat{q}_1(x_1).$$

Our asymptotic result on this estimator relies on the following regularity condition. For any functions f and g, we write $f(t) \sim g(t)$ whenever $f(t) = g(t) + o(g(t))$ as $t \downarrow 0$.

Assumption 7 *$\overline{X}_2(.)$ is differentiable, with $\overline{X}_2'(.) > 0$. For all x^* in the support of X^*, $f_{X^*}(x^*) > 0$ and there exists $\ell_{1,x^*} > 0$, $\ell_{2,x^*} > 0$, both continuous in x^*, $\beta_1 > 0$ and $\beta_2 > 0$ such that $F_{X_1|X^*}(x^* + t|x^*) \sim \ell_{1,x^*}t^{\beta_1}$ and $1 - F_{X_2|X^*}(\overline{X}_2(x^*) - t|x^*) \sim \ell_{2,x^*}t^{\beta_2}$.*

Assumption 7 holds for instance if we can do a Taylor expansion of $F_{X_1|X^*}(.|x^*)$ and $1 - F_{X_2|X^*}(.|x^*)$ at x^* and $\overline{X}_2(x^*)$, respectively. In this case β_1 (resp. β_2) is the minimal order m such that $F^{(m)}_{X_1|X^*}(x^*|x^*) \neq 0$ (resp. $[1 - F_{X_2|X^*}]^{(m)}(\overline{X}_2(x^*)|x^*) \neq 0$). With a uniform distribution for $F_{X_1|X^*}$, for instance, we get $\beta_1 = 1$, while $\beta_1 = 2$ for a triangular distribution. Under Assumption 7, the free disposal hull estimator has an asymptotic Weibull distribution.[15]

[15] Recall that $W \sim \text{Weibull}(\alpha, \rho)$ if its cdf is $x \mapsto 1 - \exp(-\alpha x^\rho)$ on $\mathbb{R}^+$.

Theorem 4.1 *Under Assumptions 1, 2, 3-(i) and 7, there exists $\alpha_{x_1} > 0$ such that*

$$n^{1/(1+\beta_1+\beta_2)}\left(\overline{X}_2(x_1) - \widehat{\overline{X}}_2(x_1)\right) \xrightarrow{d} Weibull(\alpha_{x_1}, 1+\beta_1+\beta_2).$$

This result is mostly based on previous studies on the asymptotic behavior of the free disposal hull estimator, see in particular Daouia et al. (2010). Importantly, consistency is achieved even if the density $f_{X_2|X_1}(u|x)$ tends to zero as $u \to \overline{X}_2(x)$, at any polynomial rate. The rate of convergence depends on this rate, however. We obtain for instance a rate of $n^{1/3}$ if the densities of X_1 and X_2 do not vanish on the boundary of their conditional support ($\beta_1 = \beta_2 = 1$), as with uniform distributions, but only $n^{1/5}$ if both densities vanish, while their derivatives do not ($\beta_1 = \beta_2 = 2$).

$\widehat{\overline{X}}_2(x_1)$ relies on maxima, and is therefore sensitive to outliers. A more robust estimator, based on high quantiles (namely, using $\widehat{q}_\alpha(x_1)$ with α close to one), has been introduced recently by Daouia et al. (2010), following ideas developed by Dekkers & Haan (1989). Another advantage of this estimator is that it is asymptotically normal, rather than Weibull. On the other hand, it is less efficient, less straightforward to compute and its asymptotic distribution relies on stronger conditions. Nonetheless, its asymptotic normality makes it convenient for inference, and we will consider related estimators for testing the moving support condition in the following subsection.

In a second step, the conditional distribution functions $F_{X_k|X^*}$ ($k \geq 3$) can be estimated by a kernel estimator, using a subsample of "sufficiently extreme" values X_1 and X_2. Specifically, we consider, following Equation (2.1),

$$\widehat{F}_{X_k|X^*}(x_k|x^*) = \frac{\sum_{i=1}^n K\left((X_{1i}-x^*)/h_1\right) K\left((X_{2i}-\widehat{\overline{X}}_2(x^*))/h_2\right) \mathbb{1}\{X_{ik} \leq x_k\}}{\sum_{i=1}^n K\left((X_{1i}-x^*)/h_1\right) K\left((X_{2i}-\widehat{\overline{X}}_2(x^*))/h_2\right)}, \quad (4.1)$$

where $K(.)$ is a kernel function and h_1, h_2 are two bandwidth parameters.

Finally, f_{X^*} can be estimated as well, using the fact that under Assumption 5, it is the unique solution of the integral equation

$$F_{X_3}(x_3) = \int_{\overline{X}_3^{-1}(x_3)}^{+\infty} F_{X_3|X^*}(x_3|x^*) f_{X^*}(x^*) dx^*. \quad (4.2)$$

Alternatively, under Assumption 3-(ii), we can identify and estimate $F_{X_1|X^*}$ and $F_{X_2|X^*}$ as well. Then we can use an integral equation involving F_{X_1,X_2,X_3} and the corresponding conditional cdfs, instead of simply F_{X_3} and $F_{X_3|X^*}$, to achieve accuracy gains. Similarly, if $K > 3$, we can estimate $F_{X_k|X^*}$ for $k \in \{3, ..., K\}$ as in (4.1), and then use the integral equation involving $F_{X_1,...,X_K}$ and the corresponding conditional cdfs to estimate f_{X^*}, resulting also in accuracy gains.[16] In all cases, we face an ill-posed problem as Hu & Schennach (2008) and it is possible to estimate f_{X^*} through any regularization scheme: Tikhonov, spectral cut-off or Landweber-Fridman (see Carrasco et al., 2007). Though all steps of this procedure involve existing estimators, the rate of convergence of the final estimator remains to be established as they incorporate nonparametric first and second-step estimators.

4.2 Tests of the moving support condition

Our identification strategy relies crucially on Assumption 3-(i). We now show that this condition is testable, in the sense that we can test for implications of this condition. The idea is that the maximum of the support of X_2 conditional on $X_1 = x$, namely $\overline{X}_2(x)$, is finite and strictly increasing with x. We investigate below how both points can be formally tested, using a sample $(X_{1i}, X_{2i})_{i=1...n}$ of independent and identically distributed variables. Such formal tests are useful because models for which the moving support condition is satisfied may display, at first glance, similar patterns to others for which this assumption does not hold. Figure 2, for instance, plots X_2 against X_1 in the two models $X_k = X^* + \varepsilon_k$ ($k \in \{1, 2\}$), where X^* is uniformly distributed and $(\varepsilon_1, \varepsilon_2)$ are i.i.d. and follow respectively a normal and a uniform distribution. The moving support is satisfied only in the second model, but the data look very similar. On a related note, if more than two variables are candidates for Assumption 3-(i), our tests below can also be useful to choose the most credible pair among them.

[16] In the same vein, if Assumption 5 holds for X_k, $k \geq 3$, we can estimate f_{X^*} using similar equations as (4.2) but for each $F_{X_k}, k \geq 3$, and then average the corresponding estimators.

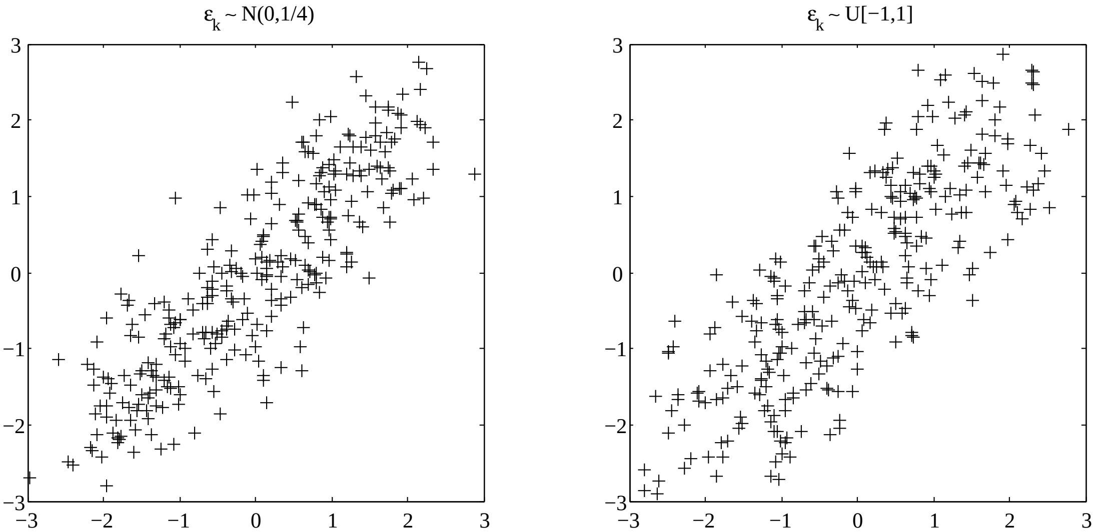

Figure 2: X_2 against X_1 if $X_k = X^* + \varepsilon_k$, with ε_k bounded or not.

4.2.1 Construction of the tests

To test for the fact that the upper bound of X_2 conditional on X_1 is finite, we consider a set A with $\sup A < \sup \text{Supp}(X_1)$. Because $X_1 \in A$ implies that $X_2 \leq \overline{X}_2(X_1) \leq \overline{X}_2(\sup A)$, the upper bound of the support of X_2 conditional on $X_1 \in A$ should also be finite. The idea of the test is to restate this boundary condition in terms of the tail index of $F_{X_2|X_1 \in A}$, a notion borrowed from the statistical extreme value theory. Our test works provided that we can apply the equivalent of the central limit theorem for extremes, the technical condition corresponding to the finite variance in the central limit theorem being Assumption 8 below. We let hereafter $\mathcal{S}$ denote the subsample $\{i : X_{1i} \in A\}$.

Assumption 8 *(Extreme value condition) There exist sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ and a non degenerate distribution H such that $a_n \max_{i\in\mathcal{S}} X_{2i} + b_n \xrightarrow{d} H$.*

This condition is mild and satisfied for all standard families of continuous distributions (see, e.g., Embrechts et al., 1997, chapter 3, for a comprehensive discussion). By the Fisher Tippett Theorem (see e.g., Embrechts et al., 1997, Theorem 3.2.3), H belongs actually to the family of generalized extreme value distributions $(H_\xi)_{\xi\in\mathbb{R}}$, with $H_\xi(x) = \exp(-(1+\xi x)^{-1/\xi})$ for $1 + \xi x > 0$ (and $H_0(x) = \exp(-\exp(-x))$). Moreover, the tail index ξ_F corresponding to F is negative if F has a finite supremum (see Theorem 3.3.12 of Embrechts et al., 1997), zero for distributions with thin tails such as the normal or exponential ones and positive for distributions with heavy tails such as the Pareto or

Student. Hence, within this framework, testing $\sup \text{Supp}(X_2|X_1 \in A) = +\infty$ against $\sup \text{Supp}(X_2|X_1 \in A) < +\infty$ is equivalent to testing for $\xi \geq 0$ against $\xi < 0$ (where $\xi = \xi_{F_{X_2 \,|X_1 \in A}}$). We adapt hereafter a test proposed by Segers & Teugels (2000) because it is simple and consistent without further restrictions, but other choices would be possible (see Neves & Alves, 2008, for a review).

The test works as follows. Let $(k_n)_{n \in \mathbb{N}}$ denote a sequence of integers such that $k_n \to \infty$ and $n/k_n \to \infty$. Split $\mathcal{S}$ into k_n subsamples $(\mathcal{S}_j)_{j=1 \ldots k_n}$ of size $m_n = [n/k_n]$ (where [.] denotes the integer part) and let $X^j_{2(1)} < X^j_{2(2)} < \ldots < X^j_{2(m_n)}$ denote the order statistic of X_2 on subsample $\mathcal{S}_j$. Introducing the ratio

$$G_j = \frac{X^j_{2(m_n)} - X^j_{2(m_n-2)}}{X^j_{2(m_n-1)} - X^j_{2(m_n-2)}},$$

the test statistic is defined by

$$T_{1n} = \sqrt{\frac{5}{k_n}} \sum_{j=1}^{k_n} \left[1 - \frac{6G_j}{(1+G_j)^2} \right].$$

The following proposition, which adapts Segers' result to our context, shows that a consistent test of $\xi \geq 0$ against $\xi < 0$, or, equivalently, of $\sup \text{Supp}(X_2 \,|X_1 \in A) = +\infty$ against $\sup \text{Supp}(X_2 \,|X_1 \in A) < +\infty$, can be obtained using T_{1n}.

Proposition 4.1 *Suppose that Assumption 8 holds. Then the test defined by the critical region $\{T_{1n} < -z_\alpha\}$, where z_α is the α-quantile of a normal random variable, is a consistent test with asymptotic level α of the null hypothesis that* $\sup$ *Supp*$(X_2|X_1 \in A) = +\infty$.

The other issue is to test that $x \mapsto \text{Supp}(X_2|X_1 = x)$ is strictly increasing. We actually test an implication of this hypothesis, by testing $\overline{X}_{21} = \overline{X}_{22}$ against $\overline{X}_{21} < \overline{X}_{22}$, with $\overline{X}_{2j} = \sup \text{Supp}(X_2|X_1 \in A_j)$ $(j \in \{1, 2\})$ and (A_1, A_2) two sets such that $\sup A_1 < \inf A_2$. For that purpose, we compare the two estimators of the upper bounds $\overline{X}_{21}$ and $\overline{X}_{22}$ derived by Dekkers & Haan (1989). Let $(X^j_{2(i)})_{i=1 \ldots n_j}$ denote the order statistic of X_2 on the subsample $\{i : X_{1i} \in A_j\}$ of size n_j. Let also (k_{jn}) denote a sequence of integers. The estimator of Dekkers & Haan (1989) is defined by

$$\widehat{\overline{X}}_{2j} = X^j_{2(n_j - k_{jn}+1)} + \frac{X^j_{2(n_j - k_{jn}+1)} - X^j_{2(n_j - 2k_{jn}+1)}}{2^{-\widehat{\xi}_j} - 1},$$

where $\widehat{\xi}_j$ is the Pickands estimator of $\xi_j = \xi_{F_{X_2|X_1 \in A_j}}$:

$$\widehat{\xi}_j = \frac{1}{\ln 2} \ln \left(\frac{X^j_{2(n_j - k_{jn}+1)} - X^j_{2(n_j - 2k_{jn}+1)}}{X^j_{2(n_j - 2k_{jn}+1)} - X^j_{2(n_j - 4k_{jn}+1)}} \right).$$

Out test statistic is then defined by

$$T_{2n} = \frac{\widehat{\overline{X}}_{22} - \widehat{\overline{X}}_{21}}{\sqrt{\sigma^2_{1n} + \sigma^2_{2n}}},$$

where σ_{jn} satisfies

$$\sigma_{jn} = \sqrt{\frac{3}{k_{jn}} \frac{|\widehat{\xi}_j| 2^{\widehat{\xi}_j - 1}}{\left|2^{\widehat{\xi}_j} - 1\right|^3}} \left[X^j_{2(n_j - k_{jn}+1)} - X^j_{2(n_j - 2k_{jn}+1)} \right].$$

As already mentioned in the previous subsection, the advantage of using $\widehat{\overline{X}}_{2j}$ in this test statistic, rather than the sample maximum $X^j_{2(n_j)}$ is that under the null, the asymptotic distribution of the corresponding test statistic is normal and free of nuisance parameter.

As previously, k_{jn} should tend to infinity at an appropriate rate and mild restrictions on $F_{X_2|X_1 \in A_j}$ have to imposed for our test to be consistent. We let afterwards $\mathcal{F}$ denote the set of differentiable cdfs, and $\mathcal{R}_\alpha$ be the set of regular variation functions with exponent α.[17]

Assumption 9 *(Regular compact distributions) For $j \in \{1, 2\}$, the support of X_2 conditional on $X_1 \in A_j$ is bounded and $F_{X_2|X_1 \in A_j} \in \mathcal{P}$, with*

$$\begin{aligned} \mathcal{P} \quad = \quad & \{F \in \mathcal{F} : \sup(\textit{support}(F)) = \overline{x}_F < \infty \textit{ and } F'(x) = \lambda_F (\overline{x}_F - x)^m (1 + R(1/(\overline{x}_F - x)). \\ & \textit{for some } \lambda_F > 0, m > -1 \textit{ and } R \in \mathcal{R}_\alpha,\ \alpha < 0 \}. \end{aligned}$$

Assumption 10 *(Conditions on $(k_{jn})_{n \in \mathbb{N}}$) For $j \in \{1, 2\}$, $k_{jn} \to \infty$ and $k_{jn} = o(n/g_j^{-1}(n))$, where g_j^{-1} is the generalized inverse of $g_j(t) = t^{3-2\xi_j} [U_j'(t)/(t^{1-\xi_j} U'(t) - \lambda_{F_{X_2|X_1 \in A_j}} |\xi_j|^{1+\xi_j})]^2$, and $U_j(t) = F^{-1}_{X_2|X_1 \in A_j}(1 - 1/t)$.*

The set $\mathcal{P}$ includes all standard distributions with compact support, such as the uniform and the beta. Under Assumptions 9 and 10, a test based on T_{2n} is consistent.

[17] A function R is regularly varying with exponent α if for all $x > 0$, $\lim_{t \to \infty} R(xt)/R(t) = x^\alpha$. We also include the zero function in $\mathcal{R}_\alpha$.

Proposition 4.2 *Suppose that Assumptions 9 and 10 hold. Then the test defined by the critical region* $\{T_{2n} < -z_\alpha\}$*, where* z_α *is the* α*-quantile of a normal random variable, is a consistent test with asymptotic level* α *of the null hypothesis* $\overline{X}_{21} = \overline{X}_{22}$ *against* $\overline{X}_{21} < \overline{X}_{22}$.

An important feature of this test is that it is consistent and has an asymptotically correct level even if the tail indices ξ_1 and ξ_2 differ, or, equivalently, even if the rates of convergence towards 1 of $F_{X_2|X_1 \in A_j}$ (when $x \to \overline{X}_{2j}$) are not the same for $j = 1$ and $j = 2$. In such a case, the estimators $\widehat{\overline{X}}_{21}$ and $\widehat{\overline{X}}_{22}$ have different rates of convergence, but these differences are automatically taken into account by the denominator of the test statistic T_{2n}.

Our approach does not test for strict monotonicity of $\overline{X}_2(.)$ everywhere, but rather between $\sup A_1$ and $\sup A_2$. However, it is possible to conduct several tests using different sets in each test. The rejection of all such tests would strongly support the strict monotonicity condition.[18]

4.2.2 Monte Carlo simulations

Both previous tests rely on the use of data near the boundaries of the support and one might wonder if such tests are useful in practice. To evaluate their power, we perform Monte Carlo simulations.

More precisely, we study the performances of the compact support test in four models of the form $X_k = \rho X^* + \varepsilon_k$ ($k \in \{1, 2\}$), where $(\varepsilon_1, \varepsilon_2)$ are i.i.d. and independent of X^*. In the first one, $\rho = 0$ and $\varepsilon_k \sim U[-1, 1]$. Then $\sup \text{Supp}(X_2|X_1 \in A) < \infty$ and the density of $X_2|X_1 \in A$ is strictly positive at the boundary (because X_1 and X_2 are independent and $X_2 = \varepsilon_2$ is uniform), the tail index being equal to -1. In the second model, $\rho = 1$, $X^* \sim U[-2, 2]$ and $\varepsilon_k \sim U[-1, 1]$. We still have $\sup \text{Supp}(X_2|X_1 \in A) < \infty$ but the density at the boundary is zero. The tail index is still negative as the supremum is finite, but it is smaller in absolute value and equals -0.5. The third model corresponds to the sum of two normal distributions ($\rho = 1$, $X^* \sim \mathcal{N}(0, 1)$ and $\varepsilon_k \sim \mathcal{N}(0, 1/4)$) for which the tail index is equal to 0. In this case $\sup \text{Supp}(X_2|X_1 \in A) = +\infty$, but the distribution of X_2 conditional on $X_1 \in A$ has thin tails. Finally, in the fourth model, $\rho = 0$ and ε_k follows a student distribution with two degrees of freedom. The tail index is positive and equals 0.5, which reflects an infinite upper bound and heavy tails.

[18]Our test could also be generalized to K different subsets $(A_k)_{k=1...K}$. In this case, we can use multivariate constraints tests developed for instance by Wolak (1991).

To test for the finiteness of the upper bound, we simulate, for each model and each sample s, respectively $n = 500$, $n = 1,000$ and $n = 2,000$ observations of the form (X_{1i}^s, X_{2i}^s). We then calculate the test T_{1n}^s and check if T_{1n}^s is less than -1.64.[19] The power of the test is reported in Table 1.

Table 1: Power of the test $\sup \mathrm{Supp}(X_2|X_1) = +\infty$

n	$X_k = \underset{(U)}{\varepsilon_k}$ $(\xi = -1)$	$X_k = \underset{(U)}{X^*} + \underset{(U)}{\varepsilon_k}$ $(\xi = -0.5)$	$X_k = \underset{(N)}{X^*} + \underset{(N)}{\varepsilon_k}$ $(\xi = 0)$	$X_k = \underset{(St)}{\varepsilon_k}$ $(\xi = 0.5)$
500	0.79	0.42	0.06	0.009
1000	0.89	0.49	0.06	0.001
2000	0.97	0.55	0.07	0.002

Note: the results are based on 1,000 simulations.

These simulations are reassuring about the power of our test. When $\sup \mathrm{Supp}(X_2|X_1 \in A) < \infty$, the test is rejected with a rather high probability. In the worst case, i.e. $n = 500$ observations and X_k is the sum of two uniform distributions, so that its density is equal to zero at the boundary, our test is rejected with a probability of 42%. This percentage increases to 79% in the first model in which the distribution is strictly positive at the boundary. On the contrary, when ξ equals 0 or is positive, the test is almost always accepted. The case $\xi = 0$ also shows that the actual level is close to the nominal level of 5%. Hence, even with a rather small number of observations, the nonparametric test of the compact support assumption appears to be quite powerful. It may be visually difficult to distinguish a distribution with compact support against one with thin tails (see Figure 2), but it is relatively easy to compare them empirically, relying on observations near the boundaries.

We then perform monotonicity tests for the first two models, in which $\sup \mathrm{Supp}(X_2|X_1 = x) < \infty$. The results for different sets A_1 and A_2 are reported in Table 2.[20]

[19] In these simulations, $A = [0.05, 0.95]$. As usually in this literature, the choice of the tuning parameter (here k_n) can be delicate and is not discussed in Segers & Teugels (2000). As a rule of thumb, we take $k_n = \sqrt{n}(1 + 10\sqrt{\xi})$.

[20] Once again, the choice of k_{jn} can be delicate in practice. We fix them here to 12 for $n = 500$, 25 for $n = 1,000$ and 50 for $n = 2,000$.

Table 2: Power of the test that $\overline{X}_{21} = \overline{X}_{22}$.

n	Sets A_1 and A_2	$X_k = \underset{(U)}{\varepsilon_k}$	$X_k = \underset{(U)}{X^*} + \underset{(U)}{\varepsilon_k}$
500	$[0.1, 0.2]$, $[0.8, 0.9]$	0	0.89
	$[0.3, 0.4]$, $[0.6, 0.7]$	0.01	0.69
	$[0.4, 0.5]$, $[0.5, 0.6]$	0	0.23
1000	$[0.1, 0.2]$, $[0.8, 0.9]$	0	1
	$[0.3, 0.4]$, $[0.6, 0.7]$	0	0.98
	$[0.4, 0.5]$, $[0.5, 0.6]$	0	0.64
2000	$[0.1, 0.2]$, $[0.8, 0.9]$	0	1
	$[0.3, 0.4]$, $[0.6, 0.7]$	0	1
	$[0.4, 0.5]$, $[0.5, 0.6]$	0.01	0.88

Note: the results are based on 1,000 simulations.

These simulations are also reassuring for the empirical relevance of the monotonicity test. Quite intuitively, the more A_1 and A_2 are separated, the more powerful the test is. Even with only 500 observations, when X_k is the sum of two uniform distributions, the null hypothesis is rejected in 89% of the cases when $A_1 = [0.1, 0.2]$ and $A_2 = [0.8, 0.9]$. On the contrary, when X_k corresponds to a uniform distribution, for which the monotonicity assumption is not satisfied, the test is almost always accepted. When A_1 and A_2 are closer, the test is less powerful except if one has larger data at his disposal. For instance, when $A_1 = [0.4, 0.5]$ and $A_2 = [0.5, 0.6]$, the power of the test for the second model goes from 23% to 88% when the size of the sample increases from 500 to $2,000$ observations.

5 Conclusion

This paper proposes an alternative and complementary approach to Hu & Schennach (2008) to identify continuous mixture models. Our result relies on a moving support assumption that states that the supports of the observed variables strictly change with the underlying unobserved component. We believe that this assumption is economically relevant in many settings such as measurement error, matching and auction models. Finally, it has the advantages of being simple and testable.

Appendix: proofs

Proof of Theorems 2.1 and 3.1

We prove both results here, as their proof are closely related. The proof consists in four steps. The first two steps show that $F_{X_k|X^*}$ is identified for $k \geq 3$ under Assumptions 1, 2, 3-(i) and 4. It therefore applies to both theorems. Then the third step proves that $F_{X_1|X^*}, F_{X_2|X^*}$ and f_{X^*} are identified if we also impose Assumption 3-(ii). This concludes the proof of Theorem 2.1. Finally, Step 4 shows that f_{X^*} is also identified if we impose Assumption 5 instead of Assumption 3-(ii). This proves the second part of Theorem 3.1.

First step: identification of $\overline{X}_2(.)$, $\underline{X}^* \equiv \inf \mathit{Supp}(X^*)$ *and* $\overline{X}^* \equiv \sup \mathit{Supp}(X^*)$.

First, by monotonicity of $\overline{X}_2(.)$ and because $X^* \leq X_1$, we have

$$X_2 \leq \overline{X}_2(X^*) \leq \overline{X}_2(X_1).$$

Therefore, $\overline{X}_2(.)$ is identified by $\overline{X}_2(x) = \max \mathrm{Supp}(X_2|X_1 = x)$.

Second, $\inf \mathrm{Supp}(X^*) = \inf \mathrm{Supp}(X_1)$, so $\underline{X}^*$ is identified. Finally, we have $\sup \mathrm{Supp}(X_2) = \overline{X}_2(\overline{X}^*)$. This implies that $\overline{X}^*$ is identified by $\overline{X}_2^{-1}(\sup \mathrm{Supp}(X_2))$.

Second step: identification of the distribution of $X_k|X^*$ *for* $k \geq 3$.

For all $\eta > 0$ and $x \in (\underline{X}^*, \overline{X}^*)$, let $\underline{x}_\eta = \max(x - \eta, \underline{X}^*)$ and $\overline{x}_\eta = \min(x + \eta, \overline{X}^*)$. We also define the set $A_\eta(x)$ by

$$A_\eta(x) = \left[\underline{x}_\eta; \overline{x}_\eta\right] \times \left[\overline{X}_2(\underline{x}_\eta); \overline{X}_2(\overline{x}_\eta)\right].$$

Remark that $X^* < \underline{x}_\eta$ implies $X_2 < \overline{X}_2(\underline{x}_\eta)$ and $X^* > \overline{x}_\eta$ implies $X_1 > \overline{x}_\eta$. Thus, $(X_1, X_2) \in A_\eta(x)$ implies that X^* belongs to the interval $[\underline{x}_\eta, \overline{x}_\eta]$.

By continuity of $x \mapsto F_{X_k|X^*}(x_k|x)$, there exists, for all $\delta > 0$, $\eta > 0$ such that $|F_{X_k|X^*}(x_k|u) -$

$F_{X_k|X^*}(x_k|x)| < \delta$ for all u such that $|u-x| < \eta$. Hence,

$$\begin{aligned}
& \left|F_{X_k|(X_1,X_2)\in A_\eta(x)}(x_k) - F_{X_k|X^*}(x_k|x)\right| \\
= & \left|\int_{\underline{x}_\eta}^{\overline{x}_\eta} \left(F_{X_k|X^*}(x_k|u) - F_{X_k|X^*}(x_k|x)\right) f_{X^*|(X_1,X_2)\in A_\eta(x)}(x^*)dx^*\right| \\
\leq & \int_{\underline{x}_\eta}^{\overline{x}_\eta} \left|F_{X_k|X^*}(x_k|x^*) - F_{X_k|X^*}(x_k|x)\right| f_{X^*|(X_1,X_2)\in A_\eta(x)}(x^*)dx^* \\
< & \ \delta,
\end{aligned}$$

where the second line stems from the independence between X_k and (X_1, X_2) conditional on X^*. Hence, for all $x \in (\underline{X}^*, \overline{X}^*)$ and all x_k,

$$\lim_{\eta\to 0} F_{X_k|(X_1,X_2)\in A_\eta(x)}(x) = F_{X_k|X^*}(x_k|x). \tag{5.1}$$

As a consequence, the distribution of X_k conditional on X^* is identified. Finally, if $\underline{X}^* > -\infty$ (and similarly for $\overline{X}^*$), $F_{X_k|X^*}(x_k|\underline{X}^*)$ is identified by continuity of $x \mapsto F_{X_k|X^*}(x_k|x)$.

Third step: identification of $F_{X_1|X^*}, F_{X_2|X^*}$ and f_{X^*} under Assumption 3-(ii).

First, applying the previous reasoning with (X_1, X_2) replaced by (X_3, X_2) and (X_1, X_3), we identify the distributions of X_1 and X_2 conditional on X^*. Finally, we prove that we identify f_{X^*} as well. For that purpose, let

$$q_\eta(x) = P(X_1 \in [\underline{x}_\eta, \overline{x}_\eta]|X^* = x)P(X_2 \in [\overline{X}_2(\underline{x}_\eta); \overline{X}_2(\overline{x}_\eta)]|X^* = x).$$

q_η is identified. Moreover, by conditional independence between X_1 and X_2,

$$P\left[(X_1, X_2) \in A_\eta\right] = \int_{\underline{x}_\eta}^{\overline{x}_\eta} q_\eta(x^*) f_{X^*}(x^*)dx^*.$$

Let us consider

$$f_{X^*,\eta}(x) = \frac{P\left[(X_1, X_2) \in A_\eta\right]}{\int_{\underline{x}_\eta}^{\overline{x}_\eta} q_\eta(v)dv}.$$

$f_{X^*,\eta}(x)$ is identified, so the result follows if we prove that $\lim_{\eta\to 0} f_{X^*,\eta}(x) = f_{X^*}(x)$.

By assumption, f_{X^*} is continuous on $(\underline{X}^*, \overline{X}^*)$. Thus, for all $\delta > 0$, there exists η such that $|x^* - x| < \eta$ implies that $|f_{X^*}(x^*) - f_{X^*}(x)| < \delta$. Hence,

$$\begin{aligned}
|f_{X^*,\eta}(x) - f_{X^*}(x)| &= \left| \int_{\underline{x}_\eta}^{\overline{x}_\eta} \frac{q_\eta(x^*)}{\int_{\underline{x}_\eta}^{\overline{x}_\eta} q_\eta(v)dv} \left(f_{X^*}(x^*) - f_{X^*}(x) \right) dx^* \right| \\
&\leq \int_{\underline{x}_\eta}^{\overline{x}_\eta} \frac{q_\eta(x^*)}{\int_{\underline{x}_\eta}^{\overline{x}_\eta} q_\eta(v)dv} \left| f_{X^*}(x^*) - f_{X^*}(x) \right| dx^* \\
&< \delta \int_{\underline{x}_\eta}^{\overline{x}_\eta} \frac{q_\eta(x^*)}{\int_{\underline{x}_\eta}^{\overline{x}_\eta} q_\eta(v)dv} dx^* \\
&< \delta.
\end{aligned}$$

The result follows.

Fourth step: identification of f_{X^*} under Assumption 5.

First, under Assumption 5, we can identify $\overline{X}_3(.)$ by the same reasoning as in the first step. Then the distribution of $\widetilde{X}_3 = \overline{X}_3^{-1}(X_3)$, as well as its distribution conditional on $X^* = x^*$, are identified. Besides, by Assumptions 2 and 5,

$$F_{\widetilde{X}_3}(x_3) = \int_c^{x_3} F_{\widetilde{X}_3|X^*}(x_3|x^*) f_{X^*}(x^*) dx^*.$$

By Assumption 5 and the dominated convergence theorem, $F_{\widetilde{X}_3}$ is differentiable and

$$f_{\widetilde{X}_3}(x_3) = f_{\widetilde{X}_3|X^*}(x_3|x_3) f_{X^*}(x_3) + \int_c^{x_3} f_{\widetilde{X}_3|X^*}(x_3|x^*) f_{X^*}(x^*) dx^*.$$

Suppose first that $m = 1$. Then $f_{\widetilde{X}_3|X^*}(x_3|x_3) > 0$ and

$$\frac{f_{\widetilde{X}_3}(x_3)}{f_{\widetilde{X}_3|X^*}(x_3|x_3)} = f_{X^*}(x_3) + \int_c^{x_3} \frac{f_{\widetilde{X}_3|X^*}(x_3|x^*)}{f_{X_3|X^*}(x_3|x_3)} f_{X^*}(x^*) dx^*.$$

The left hand-side is identified, as well as the fraction in the integral. This equation corresponds to a Volterra equation of the second kind in f_{X^*}. Because $(x_3, x^*) \mapsto f_{\widetilde{X}_3|X^*}(x_3|x^*)/f_{X_3|X^*}(x_3|x_3)$ is bounded by assumption, this equation admits a unique solution (see, e.g., Kress, 1999, Theorem 3.10). Hence, f_{X^*} is identified.

Now, if $m > 1$, we have similarly, by a straightforward induction,

$$\frac{F^{(m)}_{\widetilde{X}_3}(x_3)}{F^{(m)}_{\widetilde{X}_3|X^*}(x_3|x_3)} = f_{X^*}(x_3) + \int_c^{x_3} \frac{F^{(m)}_{\widetilde{X}_3|X^*}(x_3|x^*)}{F^{(m)}_{\widetilde{X}_3|X^*}(x_3|x_3)} f_{X^*}(x^*) dx^*.$$

This equation also corresponds to a Volterra equation of the second kind in f_{X^*}, implying once more that f_{X^*} is identified.

Proof of Theorem 3.2

Identification of $F_{X_3|X^*}, ..., F_{X_K|X^*}$

We suppose, without loss of generality, that $K = 3$ and let (x_1, x_2, x_3) be in the support of (X_1, X_2, X_3), with $x_1 \leq x_2$. Because $X^* \leq \min(X_1, X_2)$, we have,

$$F_{X_1,X_2,X_3}(x_1, x_2, x_3) = \int_{\underline{X}^*}^{x_1} F_{X_1|X^*}(x_1|x^*)F_{X_1|X^*}(x_2|x^*)F_{X_3|X^*}(x_3|x^*)f_{X^*}(x^*)dx^*.$$

By Assumption 6 and the dominated convergence theorem, F_{X_1,X_2,X_3} is twice differentiable with respect to x_1 and x_2 and $Q(x_1, x_2, x_3) = \partial^2 F_{X_1,X_2,X_3}/\partial x_1 \partial x_2(x_1, x_2, x_3)$ satisfies

$$Q(x_1, x_2, x_3) = \int_{\underline{X}^*}^{x_1} f_{X_1|X^*}(x_1|x^*)f_{X_1|X^*}(x_2|x^*)F_{X_3|X^*}(x_3|x^*)f_{X^*}(x^*)dx^*.$$

For $k \in \{1, 2\}$, let $\partial_{kr}Q$ (resp. $\partial_{kl}Q$) denote the right (resp. left) derivative of Q with respect to x_k. We have

$$\begin{aligned}\partial_{1l}Q(x_1, x_2, x_3) =& f_{X_1|X^*}(x_1|x_1)f_{X_1|X^*}(x_2|x_1)F_{X_3|X^*}(x_3|x_1)f_{X^*}(x_1) \\ &+ \int_{\underline{X}^*}^{x_1} \frac{\partial f_{X_1|X^*}}{\partial x}(x_1|x^*)f_{X_1|X^*}(x_2|x^*)F_{X_3|X^*}(x_3|x^*)f_{X^*}(x^*)dx^*\end{aligned}$$

Similarly, taking the derivative with respect to x_2 yields

$$\partial_{2r}Q(x_1, x_2, x_3) = \int_{\underline{X}^*}^{x_1} f_{X_1|X^*}(x_1|x^*)\frac{\partial f_{X_1|X^*}}{\partial x_1}(x_2|x^*)F_{X_3|X^*}(x_3|x^*)f_{X^*}(x^*)dx^*.$$

Hence, by choosing $x_1 = x_2 = x$, we get

$$\partial_{1l}Q(x, x, x_3) - \partial_{2r}Q(x, x, x_3) = f_{X_1|X^*}(x|x)^2 F_{X_3|X^*}(x_3|x)f_{X^*}(x) \qquad (5.2)$$

Suppose that $m = 1$, so that $f_{X_1|X^*}(x|x) > 0$. Then

$$F_{X_3|X^*}(x_3|x) = \frac{\partial_{1l}Q(x, x, x_3) - \partial_{2r}Q(x, x, x_3)}{\lim_{y\to\infty} \partial_{1l}Q(x, x, y) - \partial_{2r}Q(x, x, y)}.$$

Because Q and its derivative can be recovered from the data, the right-hand side is identified, and so is $F_{X_3|X^*}(x_3|x)$.

When $m>1$, some algebra show that the following equation holds:

$$\frac{\partial Q}{\partial x_{1l}^{m}\partial x_{2r}^{m-1}}(x,x,x_3)-\frac{\partial Q}{\partial x_{1l}^{m-1}\partial x_{2r}^{m}}(x,x,x_3)=\left(\frac{\partial f_{X_1|X^*}}{\partial x_1^{m-1}}\right)^2(x|x)F_{X_3|X^*}(x_3|x)f_{X^*}(x). \quad (5.3)$$

Hence, reasoning as previously, we also identify $F_{X_3|X^*}(x_3|x)$ in this case.

Identification of f_{X^*} if Assumption 6 holds for X_3
In this case, $F_{X_1|X^*}=F_{X_3|X^*}$ is identified. Thus, by Equation (5.2), we obtain, when $m=1$,

$$f_{X^*}(x)=\frac{\partial_{1l}Q(x,x,x_3)-\partial_{2r}Q(x,x,x_3)}{f_{X_1|X^*}(x|x)^2F_{X_3|X^*}(x_3|x)},$$

which holds for any x_3. Thus f_{X^*} is identified as well. The same reasoning applies when $m>1$ by using (5.3) instead of (5.2).

Proof of non-identification for Example 6

Let us consider the density function $h(x)=f_{X^*}(x/2)/6+4f_{X^*}(2x)/3$ and let H be the corresponding cdf. H is strictly increasing on the real line since $h(x)>0$ for all $x\in\mathbb{R}$. Then let $q(x)=H^{-1}\circ F_{X^*}(x)$. Because $h\neq f_{X^*}$, q is not the identity function. We show that the data can be rationalized by $\widetilde{f}_{X_k|X^*}(x_k|x^*)\equiv f_\varepsilon(x_k-q(x^*))$, which differs from $f_{X_k|X^*}(x_k|x^*)$. We prove this by showing that $(q(X^*)+\varepsilon_1,...,q(X^*)+\varepsilon_K)$ has the same characteristic function as $(X^*+\varepsilon_1,...,X^*+\varepsilon_K)=(X_1,...,X_K)$. The result follows because by construction, the conditional distribution of $q(X^*)+\varepsilon_k$ is $\widetilde{f}_{X_k|X^*}$.

First, note that the characteristic functions corresponding to f_{X^*}, h and f_ε are respectively $\Psi_{X^*}(t)=(1-|t|)^+$ (where $x^+=\max(x,0)$), $\widetilde{\Psi}(t)=\frac{1}{3}\Psi_{X^*}(2t)+\frac{2}{3}\Psi_{X^*}(t/2)$ and $\Psi_\varepsilon(t)=(1-2K|t|)^+$. Hence, the characteristic function of $(X_1,...,X_K)$ satisfies

$$\begin{aligned}\Psi_{X_1,...,X_K}(t_1,...,t_K) &= \Psi_{X^*}\left(\sum_{k=1}^K t_k\right)\prod_{k=1}^K\Psi_\varepsilon(t_k)\\ &= \left(1-\left|\sum_{k=1}^K t_k\right|\right)^+\prod_{k=1}^K(1-2K|t_k|)^+.\end{aligned}$$

Now, if $\left|\sum_{k=1}^K t_k\right| \geq 1/2$, $|t_k| \geq 1/2K$ is satisfied for at least one $k \in \{1, ..., K\}$. Hence,

$$\Psi_{X_1,...,X_K}(t_1, ..., t_K) = \Psi_{X^*}\left(\sum_{k=1}^K t_k\right) \prod_{k=1}^K \Psi_\varepsilon(t_k) = 0 = \widetilde{\Psi}\left(\sum_{k=1}^K t_k\right) \prod_{k=1}^K \Psi_\varepsilon(t_k).$$

Moreover, $\widetilde{\Psi}$ and Ψ_{X^*} coincide on $[-1/2, 1/2]$. Indeed, if $\left|\sum_{k=1}^K t_k\right| \leq 1/2$,

$$\begin{aligned}
\Psi_{X^*}\left(\sum_{k=1}^K t_k\right) &= \left(1 - \left|\sum_{k=1}^K t_k\right|\right)^+ \\
&= \left(1 - \left|\sum_{k=1}^K t_k\right|\right) \\
&= \frac{1}{3}\left(1 - 2\left|\sum_{k=1}^K t_k\right|\right) + \frac{2}{3}\left(1 - \frac{1}{2}\left|\sum_{k=1}^K t_k\right|\right) \\
&= \frac{1}{3}\left(1 - 2\left|\sum_{k=1}^K t_k\right|\right)^+ + \frac{2}{3}\left(1 - \frac{1}{2}\left|\sum_{k=1}^K t_k\right|\right)^+ \\
&= \widetilde{\Psi}\left(\sum_{k=1}^K t_k\right)
\end{aligned}$$

Hence, for all $(t_1, ..., t_K)$,

$$\Psi_X(t_1, ..., t_K) = \Psi_{X^*}\left(\sum_{k=1}^K t_k\right) \prod_{k=1}^K \Psi_\varepsilon(t_k) = \widetilde{\Psi}\left(\sum_{k=1}^K t_k\right) \prod_{k=1}^K \Psi_\varepsilon(t_k).$$

In other words, $(q(X^*) + \varepsilon_1, ..., q(X^*) + \varepsilon_K)$ has the same characteristic function as $(X^* + \varepsilon_1, ..., X^* + \varepsilon_K) = (X_1, ..., X_K)$. The result follows.

Proof of Theorem 4.1

We show that there exists $\ell_{x_1} > 0$ such that

$$P\left(X_1 \leq x_1, X_2 \geq \overline{X}_2(x_1) - t\right) \sim \ell_{x_1} t^{1+\beta_1+\beta_2}. \tag{5.4}$$

The result follows then directly from Corollary 2.1 of Daouia et al. (2010). As in the proof of Theorem 2.1, we have

$$P\left(X_1 \leq x_1, X_2 \geq \overline{X}_2(x_1) - t\right) = \int_{\overline{X}_2^{-1}(\overline{X}_2(x_1)-t)}^{x_1} F_{X_1|X^*}(x_1|x^*)[1-F_{X_2|X^*}(\overline{X}_2(x_1)-t|x^*)]f_{X^*}(x^*)dx^*.$$

By Assumption 7, we have

$$P\left(X_1 \leq x_1, X_2 \geq \overline{X}_2(x_1) - t\right) \sim \int_{\overline{X}_2^{-1}(\overline{X}_2(x_1)-t)}^{x_1} \ell_{1,x^*}(x_1-x^*)^{\beta_1}\ell_{2,x^*}(\overline{X}_2(x^*)-\overline{X}_2(x_1)+t)^{\beta_2}f_{X^*}(x^*)dx^*.$$

ℓ_{1,x^*}, ℓ_{2,x^*} and $f_{X^*}(x^*)$ are continuous as functions of x^*. Moreover, by Assumption 7, $\overline{X}_2(x^*) - \overline{X}_2(x_1) + t \sim \overline{X}'_2(x_1)(x^* - x_1) + t$. Therefore,

$$P\left(X_1 \leq x_1, X_2 \geq \overline{X}_2(x_1) - t\right) \sim \ell_{1,x_1}\ell_{2,x_1}f_{X^*}(x_1)\int_{\overline{X}_2^{-1}(\overline{X}_2(x_1)-t)}^{x_1} (x_1-x^*)^{\beta_1}(\overline{X}'_2(x_1)(x_1-x^*)+t)^{\beta_2}dx^*.$$

The change of variable $v = (x_1 - x^*)/(x_1 - \overline{X}_2^{-1}(\overline{X}_2(x_1) - t))$ then yields

$$\begin{aligned} P\left(X_1 \leq x_1, X_2 \geq \overline{X}_2(x_1) - t\right) \sim & \ell_{1,x_1}\ell_{2,x_1}f_{X^*}(x_1)(x_1 - \overline{X}_2^{-1}(\overline{X}_2(x_1) - t))^{1+\beta_1} \\ & \int_0^1 v^{\beta_1}\left(\overline{X}'_2(x_1)(x_1 - \overline{X}_2^{-1}(\overline{X}_2(x_1) - t))v + t\right)^{\beta_2} dv. \end{aligned}$$

By Assumption 5, $x_1 - \overline{X}_2^{-1}(\overline{X}_2(x_1) - t) \sim t/\overline{X}'_2(x_1)$. Hence,

$$P\left(X_1 \leq x_1, X_2 \geq \overline{X}_2(x_1) - t\right) \sim \left[\ell_{1,x_1}\ell_{2,x_1}\frac{f_{X^*}(x_1)}{\overline{X}'_2(x_1)^{1+\beta_1}}\int_0^1 v^{\beta_1}(1+v)^{\beta_2}dv\right] t^{1+\beta_1+\beta_2}.$$

Therefore, Equation (5.4) holds, and the result follows.

Proof of Proposition 4.1

We cannot apply directly the result of Segers & Teugels (2000) because the size n_A of $\mathcal{S}$ is random. To overcome this issue, we first prove the result conditional on n_A. We then recover the unconditional result by integration.

Let $(i_n)_{n\in\mathbb{N}}$ denote a sequence of integers such that $\lim_{n\to\infty} i_n/n \in (0,1]$. Conditional on $n_A = i_n$, the sample size is deterministic. Moreover, $i_n/k_n \to \infty$. Hence, we can apply the result of Segers & Teugels (2000):

$$\lim_{n\to\infty} P(T_{1n} \le x | n_A = i_n) = \Phi(x),$$

for all x and where Φ denotes the cdf of a standard normal variable.
By the strong law of large numbers, $P\left(\lim_{n\to\infty} n_A/n \in (0,1]\right) = 1$. Hence, for all x,

$$\lim_{n\to\infty} P(T_{1n} \le x | n_A) = \Phi(x) \text{ almost surely.}$$

Because $P(T_{1n} \le x | n_A)$ is bounded, by Lebesgue's dominated convergence theorem,

$$\lim_{n\to\infty} P(T_{1n} \le x) = \Phi(x).$$

The result follows.

Proof of Proposition 4.2

We prove the result conditional on the size n_j of the subsample $\{i : X_{1i} \in A_j\}(j \in \{1,2\})$, as if n_j were deterministic. To take into account their randomness, we then integrate over these sizes, as in the previous proof.

Let $\alpha_n = \sigma_{1n}/\sqrt{\sigma_{1n}^2 + \sigma_{2n}^2}$ and $U_{jn} = (\widehat{\overline{X}}_{2j} - \overline{X}_{2j})/\sigma_{jn}$ for $j \in \{1,2\}$. Under Assumptions 9 and 10, the conditions of Theorem 3.2 of Dekkers & Haan (1989) are satisfied (with their $c = \lambda_F$ and their $\rho = -\alpha$), so that

$$\sqrt{2k_{jn}} \frac{\widehat{\overline{X}}_{2j} - \overline{X}_{2j}}{X^j_{2(n_j - k_{jn}+1)} - X^j_{2(n_j - 2k_{jn}+1)}} \xrightarrow{d} \mathcal{N}\left(0, \frac{3\xi_j^2 2^{2\xi_j - 1}}{(2^{\xi_j} - 1)^6}\right).$$

Thus, because $\widehat{\xi}_j \xrightarrow{\mathbb{P}} \xi_j$, U_{jn} converges in distribution to a standard normal distribution. U_{1n} and U_{2n} are independent as they are functions of the two independent subsamples $\{i : X_{1i} \in A_j\}, j \in \{1,2\}$. Thus,

$$(U_{1n}, U_{2n}) \xrightarrow{d} \mathcal{N}(0, I_2) \tag{5.5}$$

where I_2 is the 2×2 identity matrix.

Now, applying Lemma 3.1 of Dekkers & Haan (1989) (see also their Remark p.1807), we have

$$\frac{X^j_{2(n_j-k_{jn}+1)} - X^j_{2(n_j-2k_{jn}+1)}}{\frac{n}{k_{jn}}U_j'(n/k_{jn})} \xrightarrow{\mathbb{P}} 1.$$

Using the expression of σ_{jn} and the consistency of $\widehat{\xi}_j$, this implies that there exists a deterministic sequence $(\mu_n)_{n\in\mathbb{N}}$ such that $[\sigma_{1n}/\sigma_{2n}]/\mu_n \xrightarrow{\mathbb{P}} 1$. Then, letting $\lambda_n = 1/\sqrt{1+\mu_n^2}$, we obtain after some algebra $\alpha_n/\lambda_n \xrightarrow{\mathbb{P}} 1$ and $\sqrt{(1-\alpha_n^2)/(1-\lambda_n^2)} \xrightarrow{\mathbb{P}} 1$. Defining $V_{1n} = (\alpha_n/\lambda_n)U_{1n}$ and $V_{2n} = \sqrt{(1-\alpha_n^2)/(1-\lambda_n^2)}U_{2n}$, we thus get, by Slutsky's lemma (see, e.g., van der Vaart, 2000, Lemma 2.8),

$$(V_{1n}, V_{2n}) \xrightarrow{d} \mathcal{N}(0, I_2).$$

Thus, by Skorokhod's representation theorem (see, e.g., van der Vaart, 2000, Theorem 2.19), there exists $(V_1, V_2) \sim \mathcal{N}(0, I_2)$ such that $(V_{1n}, V_{2n}) \xrightarrow{\mathbb{P}} (V_1, V_2)$.[21]

Now, we have

$$\begin{aligned} T_{2n} &= \frac{\widehat{\overline{X}}_{22} - \overline{X}_{22} + \overline{X}_{21} - \widehat{\overline{X}}_{21}}{\sqrt{\sigma_{1n}^2 + \sigma_{2n}^2}} + \frac{\overline{X}_{22} - \overline{X}_{21}}{\sqrt{\sigma_{1n}^2 + \sigma_{2n}^2}} \\ &= \alpha_n U_{1n} + \sqrt{1-\alpha_n^2}U_{2n} + \frac{\overline{X}_{22} - \overline{X}_{21}}{\sqrt{\sigma_{1n}^2 + \sigma_{2n}^2}} \end{aligned} \tag{5.6}$$

Thus, under the null,

$$T_{2n} = \left[\lambda_n(V_{1n} - V_1) + \sqrt{1-\lambda_n^2}(V_{2n} - V_2)\right] + \left[\lambda_n V_1 + \sqrt{1-\lambda_n^2}V_2\right].$$

Because λ_n is bounded, the first term into brackets tends to zero in probability. The second term is a standard normal variable. Thus, under the null, $T_{2n} \xrightarrow{d} \mathcal{N}(0,1)$. This proves that the test has asymptotically level α.

To show consistency, it suffices to prove that the third term in (5.6) tends to infinity under the alternative. Because X_2 has compact support, $X^j_{2(n_j-k_{jn})} - X^j_{2(n_j-2k_{jn})}$ is bounded, and we have $\sigma_{jn} = O_P(1/\sqrt{k_{jn}})$. Moreover, $k_{jn} \to \infty$ by Assumption 10. Hence, $\sqrt{\sigma_{1n}^2 + \sigma_{2n}^2} \xrightarrow{\mathbb{P}} 0$. Besides, $\overline{X}_{22} > \overline{X}_{21}$ under the alternative. The result follows.

[21] Actually, this representation theorem applies on a different probability space but this is not a concern here, as at the end, we are only interested by the convergence in distribution of T_{2n}.

References

Allman, E. S., Matias, C. & Rhodes, J. A. (2009), 'Identifiability of parameters in latent structure models with many observed variables', *Annals of Statistics* **37**, 3099–3132.

Becker, G. (1973), 'A theory of marriage: Part i', *Journal of Political Economy* **81**, 813–846.

Bonhomme, S., Jochmans, K. & Robin, J. M. (2014*a*), Nonparametric estimation of finite mixtures from repeated measurements. Working Paper.

Bonhomme, S., Jochmans, K. & Robin, J. M. (2014*b*), Nonparametric spectral-based estimation of latent structures. Working Paper.

Bound, J. & Krueger, A. B. (1991), 'The extent of measurement error in longitudinal earnings data: Do two wrongs make a right?', *Journal of Labor Economics* **9**, 1–24.

Carrasco, M., Florens, J. & Renault, E. (2007), Linear inverse problems in structural econometrics, *in* J. Heckman & E. Leamer, eds, 'Handbook of Econometrics', Vol. 6, North Holland.

Carroll, R. J., Ruppert, D., Stefanski, L. A. & Crainiceanu, C. M. (2006), *Measurement error in nonlinear models: a modern perspective*, Chapman and Hall.

Daouia, A., Florens, J. & Simar, L. (2010), 'Frontier estimation and extreme value theory', *Bernoulli* **16**, 1039–1063.

Dekkers, A. L. M. & Haan, L. D. (1989), 'On the estimation of the extreme-value index and large quantile estimation', *Annals of Statistics* **17**, 1795–1832.

Deprins, D., Simar, L. & Tulkens, H. (1984), Measuring labor inefficiency in post offices, *in* P. P. M. Marchand & H. Tulkens, eds, 'The Performance of Public Enterprises: Concepts and Measurements', Amsterdam: North-Holland, pp. 243–267.

D'Haultfœuille, X. (2011), 'On the completeness condition in nonparametric instrumental regression', *Econometric Theory* **27**, 460–471.

Embrechts, P., Klueppelberg, C. & Mikosch, T. (1997), *Modelling Extremal Events for insurance and finance*, Springer-Verlag.

Hall, P., Neeman, A., Pakyari, R. & Elmore, R. (2005), 'Nonparametric inference in multivariate mixtures', *Biometrika* **92**, 667–678.

Hausman, J., Newey, W. K., Ichimura, H. & Powell, J. (1991), 'Measurement errors in polynomial regression models', *Journal of Econometrics* **50**, 273–295.

Henry, M., Kitamura, Y. & Salanié, B. (2014), 'Partial identification of finite mixtures in econometric models', *Quantitative Economics* **5**, 123–144.

Hu, Y. (2008), 'Identification and estimation of nonlinear models with misclassification error using instrumental variables: a general solution', *Journal of Econometrics* **144**, 27–61.

Hu, Y., McAdams, D. & Shum, M. (2013), 'Nonparametric identification of auction models with nonseparable unobserved heterogeneity', *Journal of Econometrics* **174**, 186–193.

Hu, Y. & Sasaki, Y. (2014), Recovering the true distribution with paired self reports. Working paper.

Hu, Y. & Schennach, S. (2008), 'Instrumental variable treatment of nonclassical measurement error models', *Econometrica* **76**, 195–216.

Kasahara, H. & Shimotsu, K. (2009), 'Nonparametric identification of finite mixture models of dynamic discrete choices', *Econometrica* **77**, 137–175.

Krasnokutskaya, E. (2011), 'Identification and estimation of auction models with unobserved heterogeneity', *Review of Economic Studies* **78**, 293–327.

Kress, R. (1999), *Linear Integral Equations*, Springer New York.

Lazear, E. P. & Shaw, K. L. (2007), 'Personnel economics: The economist's view of human resources', *Journal of Economic Perspectives* **21**, 91–114.

Li, T. (2002), 'Robust and consistent estimation of nonlinear errors-in-variables models', *Journal of Econometrics* **110**, 1–26.

Neves, C. & Alves, I. F. (2008), 'Testing extreme value conditions - an overview and recent approaches', *Revstat* **6**, 83–100.

Newey, W. K. (2001), 'Flexible simulated moment estimation of nonlinear errors-in-variables models', *Review of Economics and Statistics* **83**, 616–627.

Newey, W. & Powell, J. (2003), 'Instrumental variable estimation of nonparametric models', *Econometrica* **71**, 1565–1578.

Prendergast, C. (1999), 'The provision of incentives in firms', *Journal of Economic Literature* **37**, 7–63.

Riley, J. & Samuelson, W. (1981), 'Optimal auctions', *The American Economic Review* **71**, 381–392.

Schennach, S. (2004), 'Estimation of nonlinear models with measurement error', *Econometrica* **72**, 33–75.

Schennach, S. (2007), 'Instrumental variables estimation of nonlinear errors-in-variables models', *Econometrica* **75**, 201–239.

Schennach, S. (2013), 'Regressions with berkson errors in covariates-a nonparametric approach', *Annals of Statistics* **41**, 1642–1668.

Schwarz, M. & Van Bellegem, S. (2010), 'Consistent density deconvolution under partially know error distribution', *Statistics and Probability Letters* **80**, 236–241.

Segers, J. & Teugels (2000), 'Testing the gumbel hypothesis by galton's ratio', *Extremes* **3**, 291–303.

Shimer, R. & Smith, L. (2000), 'Assortative matching and search', *Econometrica* **68**, 343–369.

Shneyerov, A. & Wong, A. C. L. (2011), 'Identification in first-price and dutch auctions when the number of potential bidders is unobservable', *Games and Economic Behavior* **72**, 574–582.

Simar, L. & Wilson, P. W. (2008), *Statistical Inference in Nonparametric Frontier Models: recent developments and perspectives*, Oxford University Press.

van der Vaart, A. W. (2000), *Asymptotic Statistics*, Cambridge University Press.

Wolak, F. (1991), 'The local nature of hypothesis tests involving inequality constraints in nonlinear models', *Econometrica* **59**, 981–995.

Yang, S., Zhao, Y. & Dhar, R. (2010), 'Modeling the under reporting bias in panel survey data', *Marketing Science* **29**, 525–539.