NONPARAMETRIC ANALYSIS OF FINITE MIXTURES

BY YUICHI KITAMURA[1,a] AND LOUISE LAAGE[2,b]

[1]*Cowles Foundation for Research in Economics, Yale University,* [a]*yuichi.kitamura@yale.edu*

[2]*Department of Economics, Georgetown University,* [b]*louise.laage@georgetown.edu*

This paper shows that finite mixture models are nonparametrically identified under weak assumptions that are plausible in a wide array of applications. The key is to utilize the identification power implied by information in covariates variation. First, three identification approaches are presented, under distinct and non-nested sets of sufficient conditions. Observable features of data inform us which of the three approaches is valid. These results apply to general nonparametric switching regressions, as well as more structural models often used in economics, such as auction models with unobserved heterogeneity, an application detailed in this paper. Second, some extensions of the identification results are developed. In particular, a mixture regression where the mixing weights depend on the value of the regressors in a fully unrestricted manner is shown to be nonparametrically identifiable. Third, the paper shows that fully nonparametric estimation of the entire mixture model is possible, by forming a sample analogue of one of the new identification strategies. The estimator is shown to possess a desirable polynomial rate of convergence as in a standard nonparametric estimation problem, despite nonregular features of the model.

1. Introduction. In empirical research it is often crucially important to control for unobserved heterogeneity, and mixture distributions provide convenient ways to deal with it. Likewise, mixtures models are frequently used in the context of clustering, a major agenda in data science. This paper studies identification problems in the presence of unobserved heterogeneity under weak assumptions, by exploring identification in nonparametric finite mixture models. We then propose a fully nonparametric estimation method.

A generic mixture model takes the following form. Consider a probability distribution function $F_\alpha(\cdot)$, indexed by a random variable α that takes values on a sample space $\mathcal{A}$. α is sometimes called a mixing variable or a latent variable. It can be interpreted as a term representing unobserved heterogeneity. Let G denote the probability distribution for α. Define

$$F(z) = \int_{\mathcal{A}} F_\alpha(z) dG(\alpha). \tag{1.1}$$

The researcher observes z distributed according to F. In other words, the mixture distribution $F(\cdot)$ is generated by mixing the component probability measures $F_\alpha(\cdot), \alpha \in \mathcal{A}$ according to the mixing distribution $G(\cdot)$. In an important special case where G is discretely distributed and the space $\mathcal{A}$ is finite, (1.1) becomes

$$F(z) = \sum_{j=1}^{J} \lambda_j F_j(z), \quad \sum_{j=1}^{J} \lambda_j = 1. \tag{1.2}$$

For example, suppose there are J types of economic agents that have type specific distributions $F_j(z), j = 1, ..., J$. If type j is drawn with probability λ_j, the resulting data obeys the

MSC2020 subject classifications: Primary 62E10; secondary 62J05, 62G99.

Keywords and phrases: Auction models, Componentwise shift-restriction, Nonparametric regression, Unobserved heterogeneity.

finite mixture model (1.2). The F defined in (1.2) is called a finite mixture distribution function. This is the main concern of the current paper. Since the paper presents various results with different models, a brief discussion of the overall nature of our contributions might be in order, as summarized in the following three points:

(i) Relation to other identification results. As we mention below, currently available nonparametric identification strategies for finite mixtures often require either (A) multiple observations (a leading example being panel data) or (B) exclusion restriction and/or specific conditions on the shapes of the component distribution functions $F_\alpha, \alpha \in \mathcal{A}$. All the results in this paper concern identification in cross-section settings (i.e. the researcher never observes an individual with a particular realization of the mixing/latent variable α more than once), therefore our identification strategy has little in common with the ones that belong to Category (A). Some papers in Category (B) assume exclusion restrictions, then invoke an identification-at-infinity type argument by focusing on observations at the tails of the component distributions. This paper does not rely on exclusion restrictions (and it even allows the mixture weights to depend on covariates in the model discussed in Section 4). Some other papers in Category (B) rely on symmetry of F_α, or assume that the component distributions themselves have mixture representations, and we do not impose neither restrictions.

(ii) Source of identification. The primal identification power in this paper comes from what may be called "componentwise shift-restriction" when a covariate is observed. That is, under an independence assumption, each component distribution generates a set of cross restrictions over a family indexed by the covariate values. Here the term "shift-restriction" is adopted from [35], who consider semiparametric estimation of ordered response models (hence their paper is not about mixtures) though the identification strategy in the current paper is not directly related to theirs: it is crucial to observe that in our case *for each component distribution* we obtain continuous limit analogues of shift-restrictions defined for a (possibly finite) set of covariate values. These componentwise shift-restrictions — and equally importantly, the fact that after aggregating such latent distributions, the resulting mixture distribution function *lacks* the shift-restriction property under a "non-parallel condition" described later — deliver fully nonparametric identification.

(iii) On identification/estimation strategies. The "componentwise shift-restriction" described above can be usefully exploited after taking Fourier/Laplace transforms of the model. We then take limits in the Fourier/Laplace domains. As noted in *(i)* above, this is quite different from the approach based on exclusion restrictions together with nonparametric estimators with observations at the tails of the component distributions. Moreover, basing identification on the upper and lower tails generally limits the number of identifiable components, typically to the case with $J = 2$, whereas our approach can be used to identify models with arbitrary J (Section 5). The number of components J itself will be identified in our approach as well. Alternatively, if we impose a large support restriction on covariates we can in principle establish identification in a straightforward manner. This would be a variant of the identification-at-infinity argument, and our approach does not share this feature either. As we shall see in Section 8 it is possible to estimate the entire mixture model fully nonparametrically with standard polynomial convergence rates under mild assumptions. This desirable property is achieved without focusing on observations at the tails of the component distributions, nor a large support condition on the covariate.

We now mention some literature on the use of mixture models in general, followed by existing methods of identification for (finite) mixtures. As noted before, mixtures are commonly used in models with unobserved heterogeneity, especially in labor economics and

industrial organization. See, for example, [6], [9] and [37] for applications of finite mixture models in these fields. They are also used extensively in duration models with unobserved heterogeneity; see [23], [24] and [46]. A somewhat different use of mixtures can be found in models of regime changes, which can be viewed as finite mixture models. [41], for example, uses a switching simultaneous equations for an empirical IO model (see also [17] and [36]). Some models with multiple equilibria can be regarded as mixtures as well, e.g., [15]. Finally, contaminated models, as analyzed by [28] and [39], can be formulated as mixture models.

The most common estimation approach for a mixture model is to parametrize it and apply maximum likelihood estimation (MLE). In the notation introduced in (1.1), MLE requires parameterizing $F_\alpha(\cdot)$ and $G(\cdot)$ so that they are known up to a finite number of parameters. The EM algorithm often provides a convenient way to calculate the MLE for a mixture model.

This paper considers nonparametric identification problems in finite mixture models. The goal of the paper is to show that it is possible to treat the component distributions of a mixture model in a flexible manner. It should be noted that [31] and [23] provide important identification results for mixture models in semiparametric settings. Again in the notation in (1.1), these authors treat the component distributions $F_\alpha(\cdot)$ parametrically, (so that it is parameterized as $F_\alpha(\cdot,\theta)$, say, by a finite dimensional parameter θ) while treating G nonparametrically. They develop nonparametric MLE estimators (NPMLE) for this type of models. Note that NPMLE, in actual applications, yields nonparametric estimates for G that are typically discrete distributions with only a few support points. This fact may suggest that considering finite mixture distributions from the outset, as this paper does, is likely to be flexible enough for practical purposes.

Identification problems of finite mixtures have attracted much attention in the statistics literature. Teicher's pioneering work [44, 45] initiated this research area. [43] provides a nice summary of this topic. See, also, [38] for a comprehensive treatment of mixture models including their identification issues. Many results known in this area assume parametric component distributions, see, e.g., [49] for a recent review of semiparametric finite mixture models. On the other hand, a number of papers have appeared on the subject of fully nonparametric mixture models, especially after the first version of the current paper was circulated. These include approaches based on multiple outcomes (e.g., [2], [7], [8], [14], [33], [48], [47]), or identification results based on exclusion restrictions, with/without tail restrictions on component distributions (e.g. [1], [13], [25], [26], [32]), or methods based on symmetry restrictions (e.g. [10], [11], [27]). A paper studying a framework close to ours is [4]. Looking at a finite mixture of nonparametric regressions, its main focus is providing uniformly consistent estimators. The argument does not rely on variations of the covariates, in fact they identify the components' distributions and the corresponding weights using one value of the covariate. They can have parallel functions but impose clusterability restrictions, as introduced in [3], which we do not need.

We note that a large part of the paper focuses on the case $J = 2$; our results are most general for this case and we suggest estimators only for this case. This focus on $J = 2$ is ubiquitous in the literature, see [20], [10], [30], [27]. Moreover as already mentioned, we extend many of our results to arbitrary J mixture components, and show that J is identified.

The main result of the present paper is that a nonparametric treatment of the component distributions of a finite mixture model is possible in a cross-sectional setting, if appropriate covariates are available. Most of our identification results are constructive which enables us to propose natural estimators which we study in Section 8.

2. Mixture Model with Covariates. Consider random vectors z and x. Suppose the conditional distribution of z given x is given by a finite mixture model of the following form:

$$F(z|x) = \sum_{j=1}^{J} \lambda_j F_j(z|x), \quad \lambda_j > 0, \quad j = 1, ..., J, \quad \sum_{j=1}^{J} \lambda_j = 1. \tag{2.1}$$

The main goal is to identify the mixing probability weights $\lambda_j, j = 1, ..., J$ and the conditional component distributions $F_j(z|x)$ from the conditional mixture distribution $F(\cdot|x)$, using nonparametric restrictions. Most sections in the paper (except for Sections 5 and 6) consider the case where $J = 2$. The above expression then becomes:

$$F(z|x) = \lambda F_1(z|x) + (1-\lambda)F_2(z|x), \quad \lambda \in (0, 1]. \tag{2.2}$$

The case with $\lambda = 0$ is ruled out as we seek identification only up to labeling. Section 5 considers an extension to the case with $J \geq 3$.

3. Identification when $J = 2$. This section develops basic nonparametric identification results for (2.2). Suppose z and x reside in $\mathbb{R}$ and $\mathbb{R}^k$, respectively. Define

$$m_j(x) := \int_{\mathbb{R}} z dF_j(z|x), j = 1, 2,$$

i.e., the mean regression functions of the component distributions. Let $F^j_{\epsilon|x}, j = 1, 2$ denote the distribution functions of type specific residuals, that is

$$F^j_{\epsilon|x}(\epsilon|x) := F_j(m_j(x) + \epsilon|x), \quad j = 1, 2. \tag{3.1}$$

Note that by construction $\int \epsilon dF^j_{\epsilon|x}(\epsilon|x) = 0, j = 1, 2$. With this notation $F_j(z|x) = F^j_{\epsilon|x}(z - m_i(x)|x), j = 1, 2$, and the model (2.2) can be written as

$$F(z|x) = \lambda F^1_{\epsilon|x}(z - m_1(x)|x) + (1-\lambda)F^2_{\epsilon|x}(z - m_2(x)|x). \tag{3.2}$$

Our goal in this section is then to identify the elements of the right hand side of (3.2) nonparametrically from the knowledge of $F(\cdot|x)$ evaluated at various x. Note that the model (3.2) is further interpreted as a switching regression model:

$$z = \begin{cases} m_1(x) + \epsilon_1, \epsilon_1|x \sim F^1_{\epsilon|x} & \text{with probability } \lambda \\ m_2(x) + \epsilon_2, \epsilon_2|x \sim F^2_{\epsilon|x} & \text{with probability } 1-\lambda. \end{cases} \tag{3.3}$$

Models as described above are conventionally estimated using parametric MLE. That is, the researcher specifies (1) parametric functions for $m_1(x), m_2(x)$, e.g. $m_1(x) = \beta_1^\top x$, $m_2(x) = \beta_2^\top x$, and (2) parametric distribution functions for $F^1_{\epsilon|x}$ and $F^2_{\epsilon|x}$, e.g. $\epsilon_1|x \sim N(0, \sigma_1^2)$, $\epsilon_2|x \sim N(0, \sigma_2^2)$. Examples of such methods can be found in [42] and [34]; see also [21] for application of MLE in a time series context. The EM algorithm is often used in computing the maximum likelihood estimator.

While the parametric approach is attractive and practical, the consistency of the MLE depends crucially on whether the parametric model is correctly specified or not. For example, even if m_1 and m_2 have the correct form, misspecifications in $F^1_{\epsilon|x}$ and $F^2_{\epsilon|x}$ would result in a failure of consistency. This is quite different from standard (possibly nonlinear) regression models, for which many distribution free estimators are available. This may discourage applied researchers from using mixture models. It also raises a more fundamental question: Is the model (3.3) identified under weaker, non/semi-parametric assumptions? The results in this section provide a positive answer to this question.

Before discussing how nonparametric identification is possible, it may be helpful to see that a certain nonparametric restriction *fails* to generate identification in the model. Arguably the most common identification assumption for the standard regression model (without mixtures) is the conditional mean restriction. In our case, by the construction of $F^1_{\epsilon|x}$ and $F^2_{\epsilon|x}$ we have $\int_{\mathbb{R}} \epsilon dF^1_{\epsilon|x}(\epsilon) = 0$ and $\int_{\mathbb{R}} \epsilon dF^2_{\epsilon|x}(\epsilon) = 0$. The question is whether the knowledge of the conditional mixture distribution $F(z|x)$ at various x, combined with these "restrictions,"

uniquely determine $F^1_{\epsilon|x}$, $F^2_{\epsilon|x}$, m_1, m_2, and λ. The answer is negative; at each x, we can split the mixture distribution $F(z|x)$ into increasing and right continuous $\mathbb{R}_+$-valued functions $a(z)$ and $b(z)$, say, so that $F(z|x) = a(z) + b(z)$. If we let $\lambda = \int da(z)$, $m_1(x) = \frac{1}{\lambda}\int z da(z)$, $m_2(x) = \frac{1}{1-\lambda}\int z db(z)$, $F^1_{\epsilon|x}(\epsilon) = a(\epsilon + m_1(x))/\lambda$ and $F^2_{\epsilon|x}(\epsilon) = b(\epsilon + m_2(x))/(1-\lambda)$ they would satisfy all the available restrictions and information at all x. Even if m_1 and m_2 are completely parameterized, the model is not identified; "splitting" of $F(z|x)$ is not unique.

While it is straightforward to see the above identification failure, it highlights the fact that the conditional mean zero condition allows "too many" ways to split the mixture distribution, thereby failing to deliver identification. Fortunately, however, there exists an alternative nonparametric restriction which identifies the model (3.3). In what follows we focus on independence restrictions, i.e. independence of (ϵ_1, ϵ_2) from x.

REMARK 3.1. Note that it suffices to assume that the independence restriction holds (i) for just one element of the k-vector of covariates (wlog we assume that it is the first element) (ii) over a small subset of the support of the element. The dependence property between ϵ's and the elements of x other than the first is completely left unspecified. In this sense the independence requirement should be interpreted as a conditional independence assumption. With a rich set of controls such a requirement might be regarded reasonable. Note this point applies to all the other identification results in this paper as well.

3.1. *First identification result.* Our first result is concerned with cases where at least one element of the vector of covariates $x = (x^1, ..., x^k)^\top$ is continuous. Assume that the first k^* elements $x^1, ..., x^{k^*}$ are continuous covariates. We establish nonparametric identifiability at $x = x_0$ utilizing local variation in one of the k^* continuous covariates. It is convenient to assume that the first element x^1 is such an element, which is assumed to be prior knowledge both for identification and estimation. The following notation is useful in considering local variations of x^1: for a point $x_0 = (x^1_0, ..., x^k_0)^\top \in \mathbb{R}$, define

$$N^1(x_0, \delta) = \{(x^1, x^2_0, ..., x^k_0)^\top \in \mathbb{R}^k | x^1 \in (x^1_0 - \delta, x^1_0 + \delta)\}.$$

ASSUMPTION 3.1. *For some $\delta > 0$,*

(i) *$\epsilon_1|x \sim F_1$ and $\epsilon_2|x \sim F_2$ at all $x \in N^1(x_0, \delta)$ where F_1 and F_2 do not depend on the value of x,*
(ii) *If $0 < \lambda < 1$, $m_1(x_0) - m_1(x) \neq m_2(x_0) - m_2(x)$, for all $x \in N^1(x_0, \delta)$, $x \neq x_0$,*
(iii) *m_1 and m_2 are continuous in x^1 at x_0.*

With the notation above, (3.2) is written as:

$$F(z|x) = \lambda F_1(z - m_1(x)) + (1-\lambda) F_2(z - m_2(x)). \tag{3.4}$$

Note that the mixing distribution is allowed to be degenerate, i.e. $J = 1$. As a convention let $\lambda = 1$ if the mixing model is degenerate. That is, with degeneration (3.2) becomes

$$F(z|x) = F_1(z - m_1(x)). \tag{3.5}$$

The parameter space of λ is therefore $(0, 1]$.

We first discuss identification of the slopes of the functions $m_1(\cdot), m_2(\cdot)$ in a neighborhood of the point $x_0 \in \mathbb{R}^k$. To this end a set of regularity conditions for nonparametric identification are stated in terms of moment generating functions. Let

$$M_i(t) := \int_{\mathbb{R}} e^{t\epsilon} dF_i(\epsilon), \quad i = 1, 2,$$

for all t such that this integral exists. M_1 and M_2 are the moment generating functions of the disturbance terms ϵ_1 and ϵ_2. Define

$$D(x) := m_2(x) - m_1(x), \quad x \in \mathbb{R}^k,$$

$$h(c,t) := e^{tD(x_0)(1+c)} \frac{M_2(t)}{M_1(t)}, \quad c \in \mathbb{R}, t \in \mathbb{R}.$$

The following imposes a very weak regularity condition on the behavior of these moment generating functions.

ASSUMPTION 3.2. (i) *The domains of $M_1(t)$ and $M_2(t)$ are $(-\infty, \infty)$, and*
(ii) *For some $\varepsilon > 0$ either $h(\pm\varepsilon, t) = O(1)$ or $1/h(\pm\varepsilon, t) = O(1)$, or both hold as $t \to +\infty$. Moreover, the same holds as $t \to -\infty$.*

REMARK 3.2. Note that the requirement (ii) for the asymptotic behavior of the ratio $\frac{M_2(t)}{M_1(t)}$ is very weak and reasonable, as it allows the ratio to grow, decline or remain bounded as t diverges.

Let $M(t|x)$ denote the moment generating function of z conditional on x, that is,

$$M(t|x) := \int_{\mathbb{R}} e^{tz} dF(z|x),$$

whose domain, by (3.2) and Assumption 3.2(i), is $(-\infty, \infty)$, and also let

$$R(t,x) := \frac{M(t|x)}{M(t|x_0)}.$$

Note that these functions are observable. The domain of these functions is $\mathbb{R}^k \times (-\infty, \infty)$ by Assumption 3.2(i). By (3.2) we have

$$M(t|x) = \lambda e^{tm_1(x)} M_1(t) + (1-\lambda) e^{tm_2(x)} M_2(t).$$

To convey intuition on the first step of our identification argument, we rewrite R as follows. Suppose $h(\pm\epsilon, t) = O(1)$ holds, then

$$\begin{aligned}\frac{1}{t}\log R(t,x') &= \frac{1}{t}\log\left(\frac{\lambda e^{tm_1(x')}M_1(t) + (1-\lambda)e^{tm_2(x')}M_2(t)}{\lambda e^{tm_1(x_0)}M_1(t) + (1-\lambda)e^{tm_2(x_0)}M_2(t)}\right) \\ &= m_1(x') - m_1(x_0) + \frac{1}{t}\log\left(\frac{\lambda + (1-\lambda)e^{t[m_2(x')-m_1(x')]}\frac{M_2(t)}{M_1(t)}}{\lambda + (1-\lambda)e^{t[m_2(x_0)-m_1(x_0)]}\frac{M_2(t)}{M_1(t)}}\right),\end{aligned}$$

and for $x' \in N^1(x_0, \delta)$ close enough to x_0, $|m_2(x') - m_1(x')|$ is less than $|D(x_0)|(1+\epsilon)$ by Assumption 3.1. Thus we have

$$\lim_{t\to\infty} \frac{1}{t}\log R(t,x') = m_1(x') - m_1(x_0).$$

If $1/h(\pm\epsilon, t) = O(1)$ instead, then write

$$\frac{1}{t}\log R(t,x') = m_2(x') - m_2(x_0) + \frac{1}{t}\log\left(\frac{\lambda e^{t[m_1(x')-m_2(x')]}\frac{M_1(t)}{M_2(t)} + (1-\lambda)}{\lambda e^{t[m_1(x_0)-m_2(x_0)]}\frac{M_1(t)}{M_2(t)} + (1-\lambda)}\right)$$

and again by Assumption 3.1, we obtain

$$\lim_{t\to\infty} \frac{1}{t}\log R(t,x') = m_2(x') - m_2(x_0).$$

The following lemma formalizes and generalizes this argument.

LEMMA 3.1. *Suppose Assumptions 3.1 and 3.2 hold. Then there exists $\delta' \in (0, \delta)$ such that for every $x' \in N^1(x_0, \delta')$*

(i) $\lim_{t\to\infty} \frac{1}{t} \log R(t, x') = m_1(x') - m_1(x_0)$ *or* $\lim_{t\to\infty} \frac{1}{t} \log R(t, x') = m_2(x') - m_2(x_0)$, *and*

(ii) $\lim_{t\to-\infty} \frac{1}{t} \log R(t, x') = m_1(x') - m_1(x_0)$ *or* $\lim_{t\to-\infty} \frac{1}{t} \log R(t, x') = m_2(x') - m_2(x_0)$ *hold.*

Lemma 3.1 suggests that the slopes of m_1 and m_2 are identified as far as the following condition holds. To state it, define

$$\mathbb{E}[z|x] = \int z dF(z|x)$$

and

$$\lambda_c := \frac{\mathbb{E}[z|x] - \mathbb{E}[z|x_0] - (1+c)\lim_{t\to-\infty} \frac{1}{t} \log R(t, x)}{\lim_{t\to+\infty} \frac{1}{t} \log R(t, x) - (1+c)\lim_{t\to-\infty} \frac{1}{t} \log R(t, x)}.$$

Note these are well defined under Assumption 3.2 (i). Take δ' as defined in Lemma 3.1.

CONDITION 3.1. *Either*

(i) $\lim_{t\to\infty} \frac{1}{t} \log R(t, x) \neq \lim_{t\to-\infty} \frac{1}{t} \log R(t, x)$ *for some* $x \in N^1(x_0, \delta')$ *or*

(ii) $\lim_{c\downarrow 0} \lambda_c = 1$ *holds.*

Under Condition 3.1, we show that λ is also identified as $\lim_{c\downarrow 0} \lambda_c = \lambda$, which gives the following lemma.

LEMMA 3.2. *Suppose Assumptions 3.1, 3.2 and Condition 3.1 hold. Then there exists $\delta' \in (0, \delta)$ such that $F(\cdot|x), x \in N^1(x_0, \delta)$ uniquely determines the value of λ, and moreover,*

$$(m_1(x) - m_1(x_0), m_2(x) - m_2(x_0)) \text{ if } \lambda \in (0, 1)$$

up to labeling and

$$m_1(x) - m_1(x_0) \text{ if } \lambda = 1$$

for all x in $N^1(x_0, \delta')$ as well.

REMARK 3.3. Condition 3.1 is the main regularity restriction for our first identifiability result. *Importantly, it is testable, as both $R(t, x)$ and λ_δ are observable.* Note that Condition 3.1 (i) fails if, for example, F_1 and F_2 are the CDFs of $N(0, \sigma_1^2)$, $N(0, \sigma_2^2)$, $\sigma_1^2 \neq \sigma_2^2$ — the case covered by our second identification result in Section 3.2 (in particular see our discussion right after Assumption 3.4).

To identify the remaining parameters, we proceed as follows. Define

$$\mathbb{E}[z^2|x] = \int z^2 dF(z|x),$$

as well as the identified function

$$C(x) = \left\{\mathbb{E}[z^2|x_0] - \mathbb{E}[z^2|x] + \lambda[m_1(x_0) - m_1(x)]^2 + (1-\lambda)[m_2(x_0) - m_2(x)]^2\right\}/2,$$

and

$$\Lambda = \begin{pmatrix} \lambda & 0 \\ 0 & (1-\lambda) \end{pmatrix}.$$

We obtain

$$\begin{bmatrix} C(x) \\ \mathbb{E}(z|x_0) \end{bmatrix} = \begin{pmatrix} [m_1(x_0) - m_1(x)] & [m_2(x_0) - m_2(x)] \\ 1 & 1 \end{pmatrix} \Lambda \begin{bmatrix} m_1(x_0) \\ m_2(x_0) \end{bmatrix}, \tag{3.6}$$

for all $x \in N^1(x_0, \delta')$. If $\lambda \neq 1$, this equation identifies $m_1(x_0)$ and $m_2(x_0)$ as the two first matrices on the right hand side are identified. Finally, to identify F_1 and F_2, we show that

$$\begin{bmatrix} M(t|x_0) \\ M(t|x) \end{bmatrix} = e^{t[m_1(x_0)+m_2(x_0)]} \begin{pmatrix} e^{-tm_2(x_0)} & e^{-tm_1(x_0)} \\ e^{t\{[m_1(x)-m_1(x_0)]-m_2(x_0)\}} & e^{t\{[m_2(x)-m_2(x_0)]-m_1(x_0)\}} \end{pmatrix} \Lambda \begin{bmatrix} M_1(t) \\ M_2(t) \end{bmatrix}, \tag{3.7}$$

which guarantees that the MGFs of F_1 and F_2, and thus F_1 and F_2, are identified if $\lambda \neq 1$. The following lemma states the full identification result and its proof formalizes the arguments above as well as generalizes to the case $\lambda = 1$.

Let $\mathcal{F}(\mathbb{R}^p)$ denote the space of distribution functions on $\mathbb{R}^p$ for some $p \in \mathbb{N}$. Define

$$\bar{\mathcal{F}}(\mathbb{R}^p) = \{F : \int u F(du) = 0, F \in \mathcal{F}(\mathbb{R}^p)\},$$

the set of distribution functions over $\mathbb{R}^p$ with mean zero. The parameter space of $(F_1(\cdot), F_2(\cdot))$ is given by $\bar{\mathcal{F}}(\mathbb{R})^2$. Also, for a set $\mathcal{C} \subset \mathbb{R}^k$ let $\mathcal{V}(\mathcal{C})$ denote the space of all $\mathbb{R}$-valued functions on $\mathcal{C}$.

LEMMA 3.3. *Suppose Assumptions 3.1, 3.2 and Condition 3.1 hold. Then* $F(\cdot|x), x \in N^1(x_0, \delta)$ *uniquely determines* $(\lambda, F_1(\cdot), F_2(\cdot), m_1(\cdot), m_2(\cdot))$ *in the set* $(0,1] \times \bar{\mathcal{F}}(\mathbb{R})^2 \times \mathcal{V}(N^1(x_0, \delta'))^2$ *up to labeling.*

REMARK 3.4. To show the above lemma, some regularity conditions on the nature of m_1, m_2, F_1 and F_2 (e.g. Assumptions 3.1(ii), 3.2(i)-(ii)) are imposed. Note that such restrictions are *not* imposed on the parameter set $(0,1] \times \bar{\mathcal{F}}(\mathbb{R})^2 \times \mathcal{V}(N^1(x_0, \delta'))^2$. The space of candidate parameters being searched over generally contains parameter values that violate, say, the non-parallel regression function condition, Assumption 3.1(ii). The only restrictions imposed on the parameter space are the independence restriction, which enables us to have $\bar{\mathcal{F}}(\mathbb{R})^2$ as the space of the distributions of ϵ's, and the mean zero property of ϵ's, which holds by construction. Lemmas 3.1 and 3.3 claim that as far as the *true parameter value* $(\lambda, F_1(\cdot), F_2(\cdot), m_1(\cdot), m_2(\cdot))$ satisfies the regularity conditions like Assumptions 3.1(ii), 3.2(i)-(ii)), it is uniquely determined in the unrestricted parameter space $(0,1] \times \bar{\mathcal{F}}(\mathbb{R})^2 \times \mathcal{V}(N^1(x_0, \delta'))^2$. This point should be clear from the proof. It is of course much easier to establish nonparametric identification by restricting the parameter space we search over, for example, by making the parameter space for m_1 and m_2 the space of pairs of functions that are non-parallel. Such a result is not satisfactory from a practical point of view: *imposing* conditions such as Assumption 3.1(ii) in estimation is difficult in practice. This is the reason why this paper considers the more challenging problem which removes unnecessary restrictions on the parameter space.

REMARK 3.5. Note that Lemmas 3.1 and 3.3 do not require $\lambda < 1$. That is, if the true model has $J = 1$, the model is still correctly identified (to be a model with just one "type" of individuals).

REMARK 3.6. Some of the assumptions made above are crucial. The main source of identification is the independence assumption, Assumption 3.1(i), as discussed before. Assumption 3.1(ii) is also essential and tied to our use of covariate variation for identification. If m_1 and m_1 are completely parallel everywhere, it is easy to see that the "shift restriction" implied by independence loses its identifying power. A distinct but related assumption of "transversality" can be found in [29], which also studies a mixture of nonparametric regressions but with normally distributed mixture components.

REMARK 3.7. On the other hand, some of the assumptions made here are "regularity conditions". First, Assumption 3.2(i) imposes a rather strong assumption requiring that the moment generating functions M_1 and M_2 of F_1 and F_2 exist over $\mathbb{R}$. Second, Assumption 3.2(ii) imposes a very mild condition: see Remark 3.2. Assumption 3.1 is important for this result, and as discussed earlier, it is testable. It is satisfied for a large set of parameters, and interestingly, it even includes the case where F_1 and F_2 are completely identical.

3.2. *Second identification result.* This section proposes an alternative approach for identifying (3.2). One advantage of this second identification result is that it is based on characteristic functions, so their existence is not an issue. Like the first identification result, the key sufficient condition, which differs from the MGF based condition in the previous section, is testable. Nonparametric identification holds under the following alternative set of sufficient conditions.

ASSUMPTION 3.3. *There exist three points x_a, x_b, x_c in $\mathbb{R}^k$ such that*

(i) *$\epsilon_1|x \sim F_1$ and $\epsilon_2|x \sim F_2$ at all $x = x_a, x = x_b, x = x_c$, where F_1 and F_2 do not depend on x,*
(ii) *$m_1(x_a) - m_1(x_b) \neq m_2(x_a) - m_2(x_b)$, $m_1(x_a) - m_1(x_c) \neq m_2(x_a) - m_2(x_c)$, and $m_1(x_b) - m_1(x_c) \neq m_2(x_b) - m_2(x_c)$.*

Assumption 3.3 is similar to Assumption 3.1, though here the continuity of m_1 and m_2 is not an issue. The next assumption imposes regularity conditions of the characteristic functions of F_1 and F_2, defined by

$$\phi_i(t) := \int_{\mathbb{R}} e^{it\epsilon} dF_i(\epsilon), \quad i = 1, 2.$$

ASSUMPTION 3.4. $\lim_{t\to\infty} \left|\frac{\phi_1(t)}{\phi_2(t)}\right| \to 0$ *or* $\left|\frac{\phi_2(t)}{\phi_1(t)}\right| \to 0$ *or* $\lambda = 1$.

It is interesting to compare Assumption 3.4 with Condition 3.1. The former gives a sufficient condition in terms of the characteristic function, whereas the latter the moment generating function. It holds, for example, if F_1 and F_2 are the CDFs of $N(0, \sigma_1^2)$, $N(0, \sigma_2^2)$, $\sigma_1^2 \neq \sigma_2^2$. [45] uses an assumption similar to this. Assumption 3.4 rules out the case with $F_1 \equiv F_2$, which is allowed by Assumption 3.1. Fortunately, just like Condition 3.1, the new condition Assumption 3.4 is verifiable through the observables, as is clear from the next lemma. This means which of the two identification strategies to be used can be determined by the observable features of the data. To state this more precisely, let $\phi(t|x)$ denote the characteristic functions of the conditional mixture distribution $F(z|x)$, that is,

$$\phi(t|x) := \int_{\mathbb{R}} e^{itz} dF(z|x),$$

and for $x_0 \in \mathbb{R}^k$ define

$$\rho(x, t) := \frac{\phi(t|x)}{\phi(t|x_0)}, \quad x \in \mathbb{R}^k.$$

CONDITION 3.2. *There exists $\epsilon > 0$ such that*

$$\lim_{t\to\infty} |\rho(x,t)| = 1$$

and

$$\lim_{t\to\infty} \frac{-i}{a}\mathrm{Log}\left(\frac{\rho(x,t+a)}{\rho(x,t)}\right) = \text{const.}$$

for every $x \in N^1(x_0,\epsilon)$ and $a \in (0,\epsilon]$ where the constant in the second condition may depend on x and $\mathrm{Log}(z)$ denotes the principal value of the complex logarithm of $z \in \mathbb{C}$.

LEMMA 3.4. *If m_1 and m_2 are non-parallel on $N^1(x_0,\epsilon)$, Assumption 3.4 and Condition 3.2 are equivalent.*

Finally, assume

ASSUMPTION 3.5. *$\sigma_1^2 := \int \epsilon^2 dF_1(\epsilon)$ and $\sigma_2^2 := \int \epsilon^2 dF_2(\epsilon)$ are finite.*

Note that the next lemma holds if the set of the regressors values includes at least three points. It therefore allows, for example, two-regressor cases where one regressor is binary and the other is continuous.

LEMMA 3.5. *Under Assumption 3.4 (or Condition 3.2), Assumptions 3.3 and 3.5, $F(\cdot|x)$ at $x = x_a, x_b$ and x_c uniquely determine $(\lambda, m_1(x_a), m_1(x_b), m_1(x_c), m_2(x_a), m_2(x_b), m_2(x_c), F_1(\cdot), F_2(\cdot))$ in the set $\mathbb{R}^7 \times \bar{\mathcal{F}}(\mathbb{R})^2$ up to labeling.*

3.3. *Third identification result.* We now propose an identification strategy that takes an approach similar to the first identification result, though differs from it in some important ways. It uses one sided limit (e.g. t tending to *positive* infinity) of MGFs and also characteristic functions. Unlike our first result, it for instance addresses the case where F_1 and F_2 are CDFs of $N(0,\sigma_1^2)$, $N(0,\sigma_2^2)$, $\sigma_1^2 \neq \sigma_2^2$. Moreover, the identification strategy for the distribution functions avoids Laplace inversion, a problematic step in practice. For these reasons, we use this identification strategy to construct our estimator in Section 8.

Recall our definition of the function $h(\cdot,\cdot)$ (see Assumption 3.2) in the statements of the following assumption.

ASSUMPTION 3.6. (i) *The domains of $M_1(t)$ and $M_2(t)$ include $[0,\infty)$ and for some $\varepsilon > 0$ either $h(\pm\varepsilon,t) = O(1)$ or $1/h(\pm\varepsilon,t) = O(1)$ or both hold as $t \to +\infty$,*
or
(ii) *The domains of $M_1(t)$ and $M_2(t)$ include $(-\infty,0]$ and for some $\varepsilon > 0$ either $h(\pm\varepsilon,t) = O(1)$ or $1/h(\pm\varepsilon,t) = O(1)$ or both hold as $t \to -\infty$.*

Note that this assumption does not demand the MGFs M_1 and M_2 to be defined on the whole real line, sometimes a restrictive assumption.

LEMMA 3.6. *Suppose Assumptions 3.1, 3.4 and 3.6 hold. Then there exists $\epsilon \in (0,\delta)$ such that for every $x \in N^1(x_0,\epsilon)$ and $a \in (0,\epsilon]$*

(i) *$\lim_{t\to\infty} \frac{1}{t}\log R(t,x') = m_1(x') - m_1(x_0)$ or $\lim_{t\to\infty} \frac{1}{t}\log R(t,x') = m_2(x') - m_2(x_0)$ if Assumption 3.6*(i) *holds, and*
$\lim_{t\to-\infty} \frac{1}{t}\log R(t,x') = m_1(x') - m_1(x_0)$ or $\lim_{t\to-\infty} \frac{1}{t}\log R(t,x') = m_2(x') - m_2(x_0)$ if Assumption 3.6 (ii) *holds instead.*

(ii) $\lim_{t\to\infty} \frac{-i}{a}\mathrm{Log}\left(\frac{\rho(x,t+a)}{\rho(x,t)}\right) = m_1(x') - m_1(x_0)$ *or* $\lim_{t\to\infty} \frac{-i}{a}\mathrm{Log}\left(\frac{\rho(x,t+a)}{\rho(x,t)}\right) = m_2(x') - m_2(x_0)$.

The constant δ in the following condition is specified in Assumption 3.1 and we modify the definition of λ_c to

$$\lambda_c := \frac{\mathbb{E}[z|x] - \mathbb{E}[z|x_0] - (1+c)\lim_{t\to\infty}\frac{1}{t}\log R(t,x)}{\lim_{t\to+\infty}\frac{1}{t}\frac{-i}{a}\mathrm{Log}\left(\frac{\rho(x,t+a)}{\rho(x,t)}\right) - (1+c)\lim_{t\to\infty}\frac{1}{t}\log R(t,x)}.$$

if Assumption 3.6(i) holds. If (ii) holds instead, the limits of $\log R(t,x)/t$ are taken as $t \to -\infty$.

CONDITION 3.3. *Either*

(i) *there exists* $\epsilon \in (0,\delta)$ *such that* $\lim_{t\to\infty}\frac{1}{t}\log R(t,x) \neq \lim_{s\to\infty}\frac{-i}{a}\mathrm{Log}\left(\frac{\rho(x,s+a)}{\rho(x,s)}\right)$ *for every* $x \in N^1(x_0,\epsilon)$ *and* $a \in (0,\epsilon]$ *if Assumption 3.6*(i) *holds*
or
(ii) *there exists* $\epsilon \in (0,\delta)$ *such that* $\lim_{t\to-\infty}\frac{1}{t}\log R(t,x) \neq \lim_{s\to-\infty}\frac{-i}{a}\mathrm{Log}\left(\frac{\rho(x,s+a)}{\rho(x,s)}\right)$ *for every* $x \in N^1(x_0,\epsilon)$ *and* $a \in (0,\epsilon]$ *if Assumption 3.6*(ii) *holds*
or
(iii) $\lim_{c\downarrow 0}\lambda_c = 1$
holds.

The above condition is verifiable with information in the observables as ρ, R and λ_c are all observed. Under this condition, the slopes are identified as in Lemma 3.2 and we obtain similarly

$$\lim_{c\downarrow 0}\lambda_c = \lambda.$$

These are summarized in the following lemma.

LEMMA 3.7. *Suppose Assumptions 3.1, 3.4, 3.6 and Condition 3.3 hold. Then there exists* $\delta' \in (0,\delta)$ *such that* $F(\cdot|x), x \in N^1(x_0,\delta)$ *uniquely determines the value of* λ*, and moreover,*

$$(m_1(x) - m_1(x_0), m_2(x) - m_2(x_0)) \text{ if } \lambda \in (0,1)$$

up to labeling and

$$m_1(x) - m_1(x_0) \text{ if } \lambda = 1$$

for all x *in* $N^1(x_0,\delta')$ *as well.*

Note that we once again need the non-parallel regression function condition. Once the increments of the regression functions are identified, their levels are obtained using (3.6) as in Lemma 3.3.

LEMMA 3.8. *Suppose Assumptions 3.1, 3.4, 3.6 and Condition 3.3 hold. Then there exists* $\delta' \in (0,\delta)$ *such that* $F(\cdot|x), x \in N^1(x_0,\delta)$ *uniquely determines* $(\lambda, m_1(\cdot), m_2(\cdot))$ *in the set* $(0,1] \times \mathcal{V}(N^1(x_0,\delta'))^2$ *up to labeling.*

To identify the distribution functions $(F_1(.), F_2(.))$, we propose a different identification argument. Its main benefit is that unlike (3.7) it does not resort to empirical MGF inversion, which is hard to handle in terms of obtaining polynomial rates of convergence. Let $g(x) := m_1(x) - m_2(x)$ and $\delta(x, x') := g(x) - g(x')$. We consider two points x_1 and x_0 at which the values of m_1 and m_2 are identified. Then we show in the proof of Lemma 3.9 that if $\delta(x_1, x_0) > 0$,

$$F_2(z) = 1 - \frac{1}{1-\lambda} \sum_{j=0}^{\infty} F(z + j\delta(x_1, x_0) + m_1(x_1) - g(x_0)|x_1) - F(z + j\delta(x_1, x_0) + m_2(x_0)|x_0), \tag{3.8}$$

$$F_1(z) = \frac{1}{\lambda} \left[F(z + m_1(x)) - (1-\lambda) F_2(z + m_1(x) - m_2(x)) \right]. \tag{3.9}$$

We use these closed-form expressions to construct our estimators in Section 8.

LEMMA 3.9. *Suppose Assumptions 3.1, 3.4, 3.6 and Condition 3.3 hold. Then there exists $\delta' \in (0, \delta)$ such that $F(\cdot|x), x \in N^1(x_0, \delta)$ uniquely determines $(F_1(\cdot), F_2(\cdot))$ in the set $\bar{\mathcal{F}}(\mathbb{R})^2$.*

4. A model with "fixed effects". The model we have focused on so far assumes that heterogeneity is exogenously determined. With $J = 2$, a draw (z, x) is generated from the first type of population or from the second with *fixed* probabilities λ and $1 - \lambda$. This section relaxes this assumption. We assume that the binary probability distribution over the two types/population can depend on x in an unspecified, nonparametric manner. In terms of the switching regression formulation, this means:

$$z = \begin{cases} m_1(x) + \epsilon_1, \epsilon_1 | x \sim F_1 & \text{with probability } \lambda(x) \\ m_2(x) + \epsilon_2, \epsilon_2 | x \sim F_2 & \text{with probability } 1 - \lambda(x). \end{cases} \tag{4.1}$$

where x and ϵ_1 (ϵ_2) are, as before, assumed to be independent. Equivalently, we can write

$$F(z|x) = \lambda(x) F_1(z - m_1(x)) + (1 - \lambda(x)) F_2(z - m_2(x)). \tag{4.2}$$

The goals is now to identify the 5-tuple of functions $(\lambda(\cdot), m_1(\cdot), m_2(\cdot), F_1(\cdot), F_2(\cdot))$ from the joint distribution of (z, x).

This model is of a particular interest in terms of its implications. As in the rest of the paper, we often interpret the difference between $(m_1(\cdot), F_1(\cdot))$ and $(m_1(\cdot), F_1(\cdot))$ as a representation of unobserved heterogeneity. In a standard panel data regression model often such heterogeneity is represented by a scalar, and when it is assumed to be independent of the regressor it would be representing random effects, whereas if it is allowed to be correlated with the regressor in an arbitrary manner it becomes a fixed effects model. In certain applications fixed effects models are highly desirable. Panel data often offers approaches to deal with fixed effects, a leading case being a linear model with additive scalar-valued fixed effects. The model (4.1) (or equivalently (4.2)) is in this sense analogous to these panel data fixed effects models. Unobserved heterogeneity in (4.1) is function-valued (i.e. m and F), as opposed to, say, an additive scalar. Its distribution, represented by $\lambda(x)$, is dependent on x in an unspecified manner, accommodating arbitrary correlation between the unobserved heterogeneity and the regressor. We now show that (4.1) is nonparametrically identified, without requiring panel data, when the finite mixture modeling of unobserved heterogeneity is appropriate. Moreover, unlike in the standard panel data fixed effects model, the distribution of unobserved heterogeneity conditional on x is identified fully nonparametrically. This means we identify the entire model, enabling the researcher to calculate desired counterfactuals.

Let

$$h(c,t,x) := e^{tD(x)(1+c)}\frac{M_2(t)}{M_1(t)}, \quad c \in \mathbb{R}, t \in \mathbb{R}, x \in \mathbb{R}^k.$$

We replace Assumptions 3.1 and 3.2 with

ASSUMPTION 4.1. *For some $\delta > 0$,*

(i) *$\epsilon_1|x \sim F_1$ and $\epsilon_2|x \sim F_2$ at all $x \in N^1(x_0,\delta)$ where F_1 and F_2 do not depend on the value of x,*
(ii) *If $0 < \lambda(\underline{x}) < 1$, $m_1(x) - m_1(\underline{x}) \neq m_2(x) - m_2(\underline{x})$, for all $(x,\underline{x}) \in N^1(x_0,\delta) \times N^1(x_0,\delta)$, $x \neq \underline{x}$,*
(iii) *λ, m_1 and m_2 are continuous in x^1 at all $x \in N^1(x_0,\delta)$.*

ASSUMPTION 4.2. (i) *The domains of $M_1(t)$ and $M_2(t)$ are $(-\infty,\infty)$, and*
(ii) *For some $\varepsilon > 0$ and every $\underline{x} \in N^1(x_0,\delta)$ either $h(\pm\varepsilon,t,\underline{x}) = O(1)$ or $1/h(\pm\varepsilon,t,\underline{x}) = O(1)$, or both hold as $t \to +\infty$. Moreover, the same holds as $t \to -\infty$.*

Define

$$R(t,x,\underline{x}) := \frac{M(t|x)}{M(t|\underline{x})},$$

$$K_{+\infty,t}(x,\underline{x}) := R(t,x,\underline{x}) \exp\left(-t \lim_{s\to+\infty} \frac{1}{s}\log(R(s,x,\underline{x}))\right),$$

$$K_{-\infty,t}(x,\underline{x}) := R(t,x,\underline{x}) \exp\left(-t \lim_{s\to-\infty} \frac{1}{s}\log(R(s,x,\underline{x}))\right),$$

$$K_{+\infty}(x,\underline{x}) := \lim_{t\to+\infty} K_{+\infty,t}(x,\underline{x})$$

and

$$K_{-\infty}(x,\underline{x}) := \lim_{t\to-\infty} K_{-\infty,t}(x,\underline{x}).$$

Note that the limits in these definitions are well-defined over a neighborhood of x_0.
We replace Condition 3.1 with:

CONDITION 4.1. *For some $\delta > 0$, either*

(i) $\lim_{t\to\infty} \frac{1}{t}\log R(t,x,\underline{x}) \neq \lim_{t\to-\infty} \frac{1}{t}\log R(t,x,\underline{x})$ *for every $(x,\underline{x}) \in N^1(x_0,\delta) \times N^1(x_0,\delta)$.*
or
(ii) *for some $\delta > 0$, $K_{+\infty,t}(x,\underline{x}) = 1$ holds for every $t \in \mathbb{R}$ and $(x,\underline{x}) \in N^1(x_0,\delta) \times N^1(x_0,\delta)$.*

Like Condition 3.1, Condition 4.1 is testable as the values of $R(\cdot,\cdot,\cdot)$ (hence $K_{+\infty,\cdot}(\cdot,\cdot)$) are observable.

Finally, we impose a condition on $\lambda(\cdot)$ to accommodate its dependency on the covariates:

ASSUMPTION 4.3. *For some $\delta > 0$ it holds that*

(i) *$\lambda(\cdot)$ is weakly monotonic in x^1 over $[x_0^1 - \delta, x_0]$*
and

(ii) $\lambda(\cdot)$ *is weakly monotonic in* x^1 *over* $[x_0, x_0^1 + \delta]$.

Note that Assumption 4.3 allows λ to be non-monotone in an arbitrarily small neighborhood of x_0 (so obviously λ can be globally non-monotone) as it imposes monotonicity separately and independently in Assumptions 4.3(i) and 4.3(ii).

LEMMA 4.1. *Suppose Assumptions 4.1, 4.2, 4.3 and Condition 4.1 hold. Then there exists* $\delta' \in (0, \delta)$ *such that* $F(\cdot|x), x \in N^1(x_0, \delta)$ *uniquely determines*

$$(m_1(x) - m_1(x_0), m_2(x) - m_2(x_0)) \text{ if } \lambda(x)\lambda(x_0) \in (0, 1)$$

up to labeling and

$$m_1(x) - m_1(x_0) \text{ if } \lambda(x)\lambda(x_0) = 1$$

for all x *in* $N^1(x_0, \delta')$ *as well.*

The next result shows that the model that allows λ to be dependent on x nonparametrically is identified. Note that the mixture can be degenerate (i.e. $\lambda(x) = 1$) for some values of x, and this can be also inferred from the observables. As in the previous identification results presented in Lemmas 3.3, 3.5 and 3.8, its main sufficient condition (i.e. Condition 4.1) is verifiable in terms of observables.

LEMMA 4.2. *Suppose Assumptions 4.1, 3.2 and Condition 4.1 hold. Then* $F(\cdot|x), x \in N^1(x_0, \delta)$ *uniquely determines* $(\lambda(x_0), F_1(\cdot), F_2(\cdot), m_1(x_0), m_2(x_0))$ *in the set* $(0, 1] \times \bar{\mathcal{F}}(\mathbb{R})^2 \times \mathbb{R}^2$ *up to labeling.*

5. Mixtures with arbitrary J. Previous sections studied the identifiability for mixtures with $J = 2$. It is desirable, however, to be able to deal with mixtures with many components in some applications, especially when mixtures are used to represent unobserved heterogeneity. This section shows that nonparametric identification can be established for general J, possibly greater than 2, and moreover we show that the number J itself is also identifiable.

The basic setup in this section is analogous to the one considered in Section 3, though the conditional distribution of $z \in \mathbb{R}$ given $x \in \mathbb{R}^k$ consists of J components, $J \in \mathbb{N}$, as in (2.1). As before, define

$$m_j(x) = \int_{\mathbb{R}} z dF_j(z|x), j = 1, 2, ..., J$$

and

$$F^j_{\epsilon|x}(\epsilon|x) := F_j(m_j(x) + \epsilon|x), \quad j = 1, 2, ..., J.$$

Later we impose independence restriction (Assumption 5.1(i) below), under which we define

$$F_j(\epsilon) := F^j_{\epsilon|x}(\epsilon|x), (\epsilon, x) \in \mathbb{R} \times N^1(x_a, \delta), j = 1, 2, ..., J$$

and this enables us to write $F(z|x)$ as

$$F(z|x) = \sum_{j=1}^{J} \lambda_j F_j(z - m_j(x)). \tag{5.1}$$

For later use, define $M_j(t) = \int e^{t\epsilon} F_j(d\epsilon), j = 1, ..., J$. This section shows that the parameter $\left(\{\lambda_j\}_{j=1}^J, \{F_j(\cdot)\}_{j=1}^j, \{m_j(\cdot)\}_{j=1}^J\right)$ is identifiable under suitable conditions.

At an intuitive level, the argument developed in Section 3 still offers a valid picture behind the identifiability result here. The independence of ϵ from x leads to a shift restriction: the shapes of the distributions $\{F_j\}_{j=1}^J$ have to remain invariant along the J regression functions. This restriction, with other conditions, nails down the true parameters uniquely. Moving from $J=2$ to $J \geq 3$, however, involves rather different theoretical arguments as developed subsequently. Recall that Section 3 presented alternative conditions that guarantee the identifiability of two-component mixture models, as summarized by Lemma 3.3, Lemma 3.5 and Lemma 3.7. This section proves the nonparametric identifiability of (5.1) under conditions that are similar to the ones used in Lemma 3.3, which seems least prohibiting of the three to generalize. Even so, this generalization calls for a multistep identification strategy with recursive procedures. The argument is rather involved thus we provide in the Appendix Section A.4.1 an intuitive explanation when $J=3$, that is, for the first iteration of the procedure.

Let us now state the identifiability results of the general model (5.1) for an arbitrary J that is assumed to be known, at a point $x_a \in \mathbb{R}^k$. The case with unknown J will be presented towards the end of this section. The general setting is the same as in Section 3: the first $k*$ elements $x^1, ..., x^{k*}$ of the vector of covariates x are continuous covariates, and we use local variations in x^1.

ASSUMPTION 5.1. *For some $\delta > 0$,*

(i) *$F^j_{\epsilon|x}(\epsilon|x)$ at all $x \in N^1(x_a, \delta)$, $j = 1, ..., J$ do not depend on the value of x;*
(ii) *$m_j, j = 1, ..., J$ are J times differentiable in x^1 at all $x \in N^1(x_a, \delta)$.*

Though Condition (ii) imposes J-th order differentiability in one argument for simplicity of presentation, this is not essential: it is sufficient to assume that there exists at least one multi-index $\alpha := (\alpha_1, ..., \alpha_{k*}) \in \mathbb{Z}^{k*}, \alpha_1 + \cdots \alpha_{k*} = J$ such that the derivative $D^\alpha m(x) = \left(\frac{\partial}{\partial x^1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x^{k*}}\right)^{\alpha_{k*}} m(x)$ is well-defined for every x in the appropriate neighborhood. See Remark 5.1 for further discussion.

The independence assumption (i) enables us to write the observable conditional distribution in the form (5.1). Continuity was assumed in Lemma 3.1. We replace it with a stronger differentiability condition (ii); it may not be essential for the proof of the Lemma, though replacing derivatives in the proof with differences leads to extremely complex case-by-case analysis. We denote with D_{x^1} the differentiation operator with respect to x^1, $m_{k,j}(x) := m_k(x) - m_j(x)$ and for a function f, $\Delta_{ab} f := f(x_a) - f(x_b)$.

ASSUMPTION 5.2. (i) $\min_{k \neq j} |m_{k,j}(x_a)| > \Delta$, *$\Delta > 0$;*
(ii) *$D_{x^1} m_j(x_a), j = 1, ..., J$ takes J distinct values in $\mathbb{R}$;*
(iii) *The domains of $M_j(t), j = 1, 2, ..., J$ are $(-\infty, \infty)$;*
(iv) *For some $\epsilon > 0$, $k = 1, ..., J$ and $J > j > k$,*

$$\lim_{t\to\infty} e^{t(\epsilon-\Delta)} \frac{M_j(t)}{M_k(t)} = 0, \lim_{t\to\infty} e^{t(\epsilon-\Delta)} \frac{\frac{\partial}{\partial t} M_j(t)}{M_k(t)} = 0, \lim_{t\to\infty} e^{t(\epsilon-\Delta)} \frac{M_j(t) \frac{\partial}{\partial t} M_k(t)}{M_k(t)^2} = 0.$$

Part (i) of the assumption is not restrictive. As before, our goal is to establish identification up to labeling, so we can assume that

$$m_1(x_a) > m_2(x_a) > ... > m_J(x_a) \tag{5.2}$$

without loss of generality: this does not impact the validity of Assumption 5.2. Part (ii) is an infinitesimal version of the non-parallel regression function conditions used in the previous sections. Given the results in Section 3, it is possible that the case with arbitrary $J \in \mathbb{N}$ can be

proved under testable conditions. Here we decided to proceed instead with Part (iv), which is useful in keeping the complexity of the proof, which is already quite lengthy, manageable.

Under these assumptions, let us look at identifiability of the slope $\Delta_{ab}m_1$ using the approach developed in Section 3, for all x_b in a chosen neighborhood of x_a. Note that we know $\lambda_1 \neq 0$. By Assumption 5.1 (ii), there exists $\delta' > 0$, $\delta' < \delta$, such that for all $x_b \in N^1(x_a, \delta')$ and for all $j = 1, ..., J$, $|m_j(x_b) - m_j(x_a)| < \frac{\epsilon}{2}$, and $D_{x^1}m_j(x_b), j = 1, ..., J$ take J distinct values. We write, in the case $m_1(x_b) - m_1(x_a) > 0$,

$$\frac{1}{t}\log\left(\frac{M(t|x_b)}{M(t|x_a)}\right) = \frac{1}{t}\log\left(\frac{e^{t[m_1(x_b)-m_1(x_a)]} + \sum_{j=2}^{J}\frac{\lambda_j}{\lambda_1}\frac{M_j(t)}{M_1(t)}e^{t[m_j(x_b)-m_1(x_a)]}}{1 + \sum_{j=2}^{J}\frac{\lambda_j}{\lambda_1}\frac{M_j(t)}{M_1(t)}e^{t[m_j(x_a)-m_1(x_a)]}}\right),$$

and in the case $m_1(x_b) - m_1(x_a) < 0$, we write

$$\frac{1}{t}\log\left(\frac{M(t|x_b)}{M(t|x_a)}\right) = \frac{1}{t}\log\left(\frac{1 + \sum_{j=2}^{J}\frac{\lambda_j}{\lambda_1}\frac{M_j(t)}{M_1(t)}e^{t[m_j(x_b)-m_1(x_b)]}}{e^{t[m_1(x_a)-m_1(x_b)]} + \sum_{j=2}^{J}\frac{\lambda_j}{\lambda_1}\frac{M_j(t)}{M_1(t)}e^{t[m_j(x_a)-m_1(x_b)]}}\right).$$

Similarly, since $m_j(x_b) - m_1(x_a)$, $m_j(x_a) - m_1(x_a)$, $m_j(x_b) - m_1(x_b)$, and $m_j(x_a) - m_1(x_b)$ are less than $\epsilon - \Delta$, this gives in both cases,

$$\forall x_b \in N^1(x_a, \delta'), \lim_{t\to\infty}\frac{1}{t}\log\left(\frac{M(t|x_b)}{M(t|x_a)}\right) = \Delta_{ba}m_1.$$

Hence the slope $\Delta_{ab}m_1$ is identifiable for all $x_b \in N^1(x_a, \delta')$.

We now focus on the identifiability of the slopes $\Delta_{ab}m_j$ for all $j = 2, ..., J$ and x_b in an appropriate neighborhood of x_a. Pick a point $x_b \neq x_a$ in $\mathbb{R}^k$. Define $R_2^j(t, x_a) = tD_{x_a^1}(m_j(x_a) - m_1(x_a)), j = 2, ..., J$, and $R_{k+1}^j(t, x_a) = D_{x_a^1}\frac{R_k^j(t,x_a)}{R_k^k(t,x_a)} + t\frac{R_k^j(t,x_a)}{R_k^k(t,x_a)}D_{x_a^1}[m_j(x_a) - m_k(x_a)]$, $k = 3, ..., J$, $j = k + 1, ..., J$. Let $R_k(t, x_a) = R_k^k(t, x_a), k = 2, ..., J$. Define also the operator $A(x_a, x_b, t, k)$

$$A(x_a, x_b, t, k)(f)(x_a) = \frac{\partial}{\partial x_a^1}\left[\frac{e^{-t[\Delta_{ab}m_k - \Delta_{ab}m_{k-1}]}}{R_k(t, x_a)}f(x_a)\right], k = 2, 3, ..., J. \tag{5.3}$$

for $f : \mathbb{R}^k \to \mathbb{R}$ a function differentiable in its first argument. Our approach is sequential. At the k^{th} step, the slopes $\Delta_{ab}m_j$, $j = 1, ..., k$, are identified and the operator $A(x_a, x_b, t, k)$, which is then identified as well, generalizes the procedure explained in Appendix A.4.1 and eliminates unknown parameters associated to the mixture components $j = 1, ..., k$. Note that the definition of A uses fractions where $R_k(t, x_a)$ appears on the denominator. We obtain in Proposition A.1 a simpler expression for $R_k(t, x_a)$ that guarantees that it is nonzero for t large enough, see also Remark A.1.

Finally, define for $k = 2, ..., J$, the identified function $Q_k(x_a, t)$ such that

$$Q_2(x_a, t) = \frac{\partial}{\partial x_a^1}[e^{-t\Delta_{ab}m_1}M(t|x_a)],$$

$$Q_k(x_a, t) = \tag{5.4}$$

$$A(x_a, x_b, t, k-1)A(x_a, x_b, t, k-2)\cdots A(x_a, x_b, t, 2)D_{x_a^1}[e^{-t\Delta_{ab}m_1}M(t|x_a)].$$

The following lemma identifies $\Delta_{ab}m_k - \Delta_{ab}m_{k-1}$ for all x_b in a neighborhood of x_a. Using the knowledge of $\Delta_{ab}m_1$, $\Delta_{ab}m_k, k = 2, ..., J$ is thus identified for all x_b in a neighborhood of x_a.

LEMMA 5.1. *There exists $\delta' > 0$ with $\delta' \leq \delta$, such that for all $x_b \in N^1(x_a, \delta')$,*

$$\lim_{t\to\infty} \frac{\partial}{\partial t} \log\left(\frac{Q_k(x_a,t)}{Q_k(x_b,t)}\right) = \Delta_{ab}m_k - \Delta_{ab}m_{k-1}, k = 2,3,...,J.$$

The identifiability of the rest of the model (at x_a) is established using the knowledge of $\Delta_{ab}m_k, k = 1,2,...,J$ and conditional moments of z given x_a. To state the complete identification result, we denote the point of identification x_0 instead of x_a as the following assumption relies on several values of x.

ASSUMPTION 5.3. *There exists $X = (x_1,...,x_{J-1}) \in N^1(x_0,\delta')^{J-1}$ such that*

$$\Xi(x_0,X) = \begin{pmatrix} \Delta_{0,1}m_1 - \Delta_{0,1}m_J & \dots & \Delta_{0,1}m_{J-1} - \Delta_{0,1}m_J \\ \vdots & \ddots & \vdots \\ \Delta_{0,J-1}m_1 - \Delta_{0,J-1}m_J & \dots & \Delta_{0,J-1}m_{J-1} - \Delta_{0,J-1}m_J \end{pmatrix},$$

where $\Delta_{0,i}m_j = m_j(x_0) - m_j(x_i)$, is nonsingular.

Note that the matrix $\Xi(x_0,X)$ is observed and thus Assumption 5.3 is testable.

LEMMA 5.2. *Suppose Assumptions 5.1, 5.2 and 5.3 hold. Then $F(\cdot|x)$, $x \in N^1(x_0,\delta')$, uniquely determines $((\lambda_j)_{j=1..J-1}, (F_j(\cdot))_{j=1..J}, (m_j(x_0))_{j=1..J})$ in the set $(0,1)^{J-1} \times \bar{\mathcal{F}}(\mathbb{R})^J \times \mathbb{R}^J$ up to labeling.*

Having shown identification of our model assuming knowledge of J, we now consider the case where J is unknown. We find that J is identified using the observable sequence of functions $(Q_k(x_a,.))_{k=1,...}$. As we see below, the number of mixture components J is equal to the largest j for which the function $Q_j(x_a,.)$ is not identically 0. That is, one can compute $\Delta_{ab}m_j$ using Q_j sequentially and once $Q_{j_0} = 0$ for some j_0, $J = j_0 - 1$.

PROPOSITION 5.1.

$$J = \max\{j \geq 1 | \exists t_0 \in \mathbb{R}, Q_j(x_a,t_0) \neq 0\}.$$

REMARK 5.1. Note that it is not essential for our identification strategy to impose in Assumption 5.1 (ii) that m is J-times differentiable in one argument. The use of the differentiation operator $\frac{\partial}{\partial x^1}$ in the linear operator A is motivated by the fact that it eliminates terms that do not involve x_a, therefore with respect to which argument we differentiate is unimportant. The same identification argument applies if at each application of the operator A in the definition of Q_k, see Equation (5.4), we use $\frac{\partial}{\partial x^\ell}$ with a different $\ell \in \{1,...,k\}$ instead of always differentiating with respect to the same covariate. As noted before, we need m to be differentiable up to a J-th order multi-index. This is less stringent than Assumption 5.1 (ii), though we chose to state the result in the current form for notational simplicity.

REMARK 5.2. We show in Section A.5 of the Appendix how to extend the identification results for Model (5.1) with $J \geq 2$ to the case where the distribution of types may depend on x, i.e., $(\lambda_j)_{j=1..J-1}$ may vary with x. In this setting, we argue that J can be identified as well.

6. Application to Identifiability of Auction Models with Unobserved Heterogeneity. It is of great importance to demonstrate that the preceding identification results potentially apply to nonparametric analysis of auction models with unobserved heterogeneity. As recognized in the recent literature, failing to account for unobserved heterogeneity in empirical auction models can lead to grossly misleading policy implications and incorrect counterfactual analyses. The reader is referred to [19] for various approaches to nonparametric identifiabilty in auction models when unobserved heterogeneity is present. Here we focus on application of the preceding mixture identification results to models with auction-specific unobserved heterogeneity. In particular, we focus on a symmetric affiliated auction model as considered in [40]. Suppose that valuations have the following multiplicative form, with J unknown types of auctions

$$V^k = \Gamma_j(x) U_j^k \quad \text{with probability } \lambda_j, 1 \leq j \leq J \tag{6.1}$$

where V^k is the valuation of bidder k, $1 \leq k \leq I$, who knows the number of bidders I, observed characteristics x, unobserved heterogeneity (i.e. unobserved type of auction) j, and a signal S^k. The function $\Gamma_j(x)$ depends on the two characteristics x and j. The term U_j^k can be interpreted as the "homogenized valuation" for bidder k, as used in [18]. Let B^k denote the bid of bidder k. The observables in this application are $(I, B^1, ..., B^I, x)$.

We maintain that there is a finite number of types in terms of auction heterogeneity. It is then possible to establish identification under weak assumptions. In the following result note that (i) valuations can be affiliated, and (ii) unobserved heterogeneity, represented by $j \in \{1, ..., J\}$, is treated flexibly, as not only it can affect valuations through the index function Γ_j in an unrestricted way, the distribution of the homogenized valuation U_j^k is allowed to depend on j freely. Moreover, as presented in Section 4 and further extended in Appendix A.5, it is possible to allow the distribution of unobserved heterogeneity to depend on the observed covariate x. The latter feature can be highly important when little is known about the dependence between the observed auction heterogeneity x and the unobserved auction heterogeneity j. Property (i) is important, as many preceding nonparametric identification results for auction with unobserved heterogeneity focus on the independent private value (IPV) model, as they tend to impose independence assumptions across valuations, with the exception of [12]. For example, Property (i) implies that the result in this section applies to the common values model.

Assume

$$(U_j^1, \ldots, U_j^I, S^1, \ldots, S^I) \perp\!\!\!\perp x | I \tag{6.2}$$

for every $j \in \{1, ..., J\}$. Note that standard approaches to deal with unobserved heterogeneity do so through the index function Γ_j, and would not allow $(U^1, ..., U^I)$ to depend on j . Define

$$w(S, I, x, j) := \mathrm{E}\left[V^k | S^k = \max_{i \neq k, 1 \leq i \leq I} S^i = S, I, x, j\right]$$

which corresponds to the expected value of a bidder's valuation conditional on I, x, j, and the event that her equilibrium bid is pivotal. This is a quantity sometimes simply called "pivotal expected value". Let $w^k := w(S^k, I, x, j), 1 \leq k \leq I$ denote the pivotal expected value of the k-th bidder (whose signal is S^k) in an auction with characteristics (x, j) and I bidders. Our main goal here is to identify the joint distribution of $(w^1, ..., w^I)$ conditional on in an auction with (x, I, j). Note that such knowledge is sufficient to address important questions often asked in practice: see, for example, footnote 9 of [19] for further discussions.

The above setting, in turn, implies an expression of w of the following form

$$w\left(S, I, x, j\right) = \Gamma_j(x) \omega\left(S, I, j\right), \tag{6.3}$$

where $\omega(S;I,j) = E[U_j^k | S^k = \max_{i\neq k} S^i = S, I, j]$. Like the homogenized valuations $\{\{U_j^k\}_{k=1}^I\}_{j=1}^J$, $\omega_j^k := \omega\left(S^k, I, j\right)$ is interpreted as a homogenized pivotal expected value of bidder k in an auction of unobserved type j. It is well-known that the equilibrium bidding function preserves multiplicative separability in (6.1), hence (6.3), for each bidder k. Thus we obtain

$$B^k = \Gamma_j(x) R_j^k,$$

where R_j^k is the homogenized bid of bidder k in type j auction. Note that the unobserved auction type can affect equilibrium bids through two channels, that is, the index function Γ_j and the homogenized bid R_j^k. Define $b^k = \log B^k$, $\gamma_j(x) := \log \Gamma_j(x)$ and $r_j^k := \log R_j^k$, then we have

$$b^k = \gamma_j(x) + r_j^k, \quad 1 \leq j \leq J, \quad 1 \leq k \leq I. \tag{6.4}$$

Note that (6.2) implies

$$(r_j^1, \ldots, r_j^I) \perp\!\!\!\perp x \quad \text{for every } j \tag{6.5}$$

conditional on I.

We now invoke our general nonparametric theory to establish identification of this model. Let $c = (c_1, ..., c_I)' \in \mathbb{R}^I$, and define $b(c) := \sum_{k=1}^I c_k b_k$, $C(c) := \sum_{k=1}^n c_k$ and $r_j(c) := \sum_{k=1}^I c_i r_j^k$. One of the main objects to be identified is the I-dimensional joint distribution of the pivotal expected values $w^1, ..., w^I$ conditional on (x, j, I), and our identification strategy works for each value of I. Thus in the rest of this section we treat I as being fixed at a value, and suppress the index I unless necessary. By (6.4) and the finite mixture structure of the evaluation in (6.1) we have

$$b(c) = C(c)\gamma_j(x) + r_j(c) \quad \text{ with probability } \lambda_j, 1 \leq j \leq J$$

where $r(c) \perp\!\!\!\perp x$ by (6.5). Let $\left(b(c), \{C(c)\gamma_j(\cdot)\}_{j=1}^J, \{r_j(c)\}_{j=1}^J\right)$ play the role of $\left(z, \{m_j(\cdot)\}_{j=1}^J, \{\epsilon_j\}_{j=1}^J\right)$ in Lemma 5.2, then $(C(c)\gamma_j(\cdot), \lambda_j)$ and the distribution of $r_j(c)$ are all identified for every $c \in \mathbb{R}^n$ and each $j \in \{1, ..., J\}$. Such results extend to the case where $\{\lambda_j\}_{j=1}^J$ depends on x nonparametrically with the approach presented in Appendix A.5, thereby accommodating correlations between the observed and unobserved auction specific characteristics. Moreover, we now know $\gamma_j(\cdot)$, $j \in \{1, ..., J\}$ since $C(c)$ is known, and the number of (unobserved) auction types J is identified as well. Note that for each j, the marginal distribution of every linear combination $r_j(c)$ of the I-vector $r_j := (r_j^1, ..., r_j^I)$ is identified as $c \in \mathbb{R}^I$ can be chosen arbitrarily. Then by Cramér-Wold the joint distribution, say G_j, of the vector r_j is obtained for each j. Apply this and the knowledge of γ_j to equation (6.4) to determine the joint distribution $(b^1, ..., b^I)|x, j, I$ by $G_j(b^1 - \gamma_j(x), ..., b^I - \gamma_j(x))$. Using the first order condition for equilibrium bidding (see, e.g. [18], [5] and Equation (2.4) in [19]) we can now back out the joint distribution of $(w^1, ..., w^I)|x, j, I$ as desired.

7. Instrumental Variables. The identification results developed in the preceding sections can be used to identify nonparametric finite mixture regression with endogenous regressors. Suppose we observe a triple of random variables (y, w, x) taking its value in $\mathcal{Y} \times \mathcal{W} \times \mathcal{X}$ where $\mathcal{Y} \subset \mathbb{R}$, $\mathcal{W} \in \mathbb{R}^p$ and $\mathcal{X} \in \mathbb{R}^k$. Also let

$$z := \begin{pmatrix} w \\ y \end{pmatrix}.$$

For brevity of exposition, we treat the case with $J = 2$. In a manner similar to Section 3, consider a switching regression model:

$$y = \begin{cases} g_1(w) + \eta_1, & (y, w, x, \eta_1) \sim F_1 \qquad \text{with probability } \lambda \\ g_2(w) + \eta_2, & (y, w, x, \eta_2) \sim F_2 \qquad \text{with probability } 1 - \lambda. \end{cases} \tag{7.1}$$

Unlike in the previous sections, however, we no longer assume that η's and w are uncorrelated or independent. Instead, we assume

$$\int \eta dF_1(\eta|x) = \int \eta dF_2(\eta|x) = 0, \tag{7.2}$$

that is

$$\mathbb{E}[\eta_1|x] = \mathbb{E}[\eta_2|x] = 0.$$

Here and thereafter the notation $F_j(\star_1, \star_2, ...)$ and $F_j(\star_1, \star_2, ...|\star)$ denote the joint distribution of $\star_1, \star_2, ...$ and the conditional distribution of $\star_1, \star_2, ...$ given $\star$ when the joint distribution is given by $F_j, j = 1, 2$. Consider linear operators

$$T_1[f](x) = \int f(w) dF_1(w|x), \quad T_2[f](x) = \int f(w) dF_2(w|x) \tag{7.3}$$

and assume that these operators are invertible.

The main goal is to identify g_1 and g_2. Here x plays the role of instrumental variables. As before, define $m_1(x) = \int z dF_1(z|x)$ and $m_2(x) = \int z dF_2(z|x)$. Note that $m_1 : \mathbb{R}^k \to \mathbb{R}^{p+1}$ and $m_2 : \mathbb{R}^k \to \mathbb{R}^{p+1}$. For $h = 1, ..., p+1$, let $m_{h,1}(\cdot)$ and $m_{h,2}(\cdot)$ denote the $h-$th elements of $m_1(\cdot)$ and $m_2(\cdot)$, respectively. Define the $p+1$-dimensional vectors of random variables $\epsilon_j = z - m_j(x), (z, x) \sim F_j(z, x), j = 1, 2$. Consistent with the previous notation let $F_j(\epsilon_j|x), j = 1, 2$, denote the conditional distributions of ϵ_1 and ϵ_2 under F_1 and F_2.

By construction,

$$\int \epsilon dF_j(\epsilon|x) = 0, j = 1, 2.$$

If we further assume that $F_j(\epsilon|x), j = 1, 2$ do not depend on x, a possibly strong condition when p is large, an appropriate extension of the theory developed in Section 3 can be used to identify $m_{p+1,1}(x)$, $m_{p+1,2}(x)$, $F_1(z|x)$ and $F_2(z|x)$, which in turn, also identify the operators T_1 and T_2. By (7.1), (7.2) and (7.3) we have

$$m_{p+1,1}(x) = T_1[g_1](x), \quad m_{p+1,2}(x) = T_2[g_2](x).$$

Then by their invertibility, g_1 and g_2 are identified. To formalize this idea, consider the following assumptions:

ASSUMPTION 7.1. *For some $\delta > 0$,*

(i) *$\epsilon_1|x \sim F_1^\epsilon$ and $\epsilon_2|x \sim F_2^\epsilon$ at all $\mathcal{X}$ where F_1^ϵ and F_1^ϵ do not depend on the value of x;*
(ii) *$m_{h,1}(x_0) - m_{h,1}(x) \neq m_{h,2}(x_0) - m_{h,2}(x)$, for all $x \in N^1(x_0, \delta)$, $x \neq x_0$ and for all h, $h = 1, ..., p+1$;*
(iii) *m_1 and m_2 are continuous at x_0.*

To state a multivariate extension of Assumption 3.2, define the multivariate moment generating function

$$M_i(\mathbf{t}) = \int e^{\mathbf{t}^\top \eta} dF_i(\eta), \quad i = 1, 2, \quad \mathbf{t} \in \mathbb{R}^{p+1}.$$

Let $\mathbf{e}_j$ denote the unit vector whose $j-$th element is 1. Accommodating the identification strategy in Section 3.1 requires some modification as follows. Define $D(x) := m_2(x) - m_1(x)$ as before, though now $D : \mathbb{R}^k \to \mathbb{R}^p$ is vector-valued. Also let

$$h_j(c,t) := e^{t\mathbf{e}_j{}^\top D(x_0)(1-c)} \frac{M_2(t\mathbf{e}_j)}{M_1(t\mathbf{e}_j)}, \quad c \in \mathbb{R}_{++}, t \in \mathbb{R}$$

and

$$R(\mathbf{t}, x) := \frac{M(\mathbf{t}|x)}{M(\mathbf{t}|x_0)}, \qquad \mathbf{t} \in \mathbb{R}^{p+1}.$$

ASSUMPTION 7.2. (i) *The domains of $M_1(\mathbf{t})$ and $M_2(\mathbf{t})$ are $(-\infty, \infty)^{p+1}$;*
(ii) *For some $\varepsilon > 0$ either $h_j(\pm\varepsilon, t) = O(1)$ or $1/h_j(\pm\varepsilon, t) = O(1)$, or both hold as $t \to +\infty$ for each $j \in \{1, ..., p+1\}$. Moreover, the same holds as $t \to -\infty$*

CONDITION 7.1. *Either*

(i) $\lim_{t\to\infty} \frac{1}{t} \log R(t\mathbf{e}_j, x) \neq \lim_{t\to-\infty} \frac{1}{t} \log R(t\mathbf{e}_j, x)$ *for each $j \in \{1, ..., p\}$ and for some $x \in N^1(x_0, \delta)$*
or
(ii) $\lim_{c\downarrow 0} \lambda_c = 1$
holds.

By modifying the proofs of Lemmas 3.1 and 3.3 appropriately to deal with $\mathbb{R}^p$-valued random variables, combined with Cramer-Wald (see Section 6 for a similar step) we can show that $(\lambda, F_1^\epsilon, F_2^\epsilon, \{m_1(x), m_2(x)\}_{x\in N^1(x_0,\delta)})$ is identified under Assumptions 7.1 and 7.2. Therefore each component distribution of z is obtained by

$$F_1(z|x) = F_1^\epsilon(z - m_1(x)), \quad F_2(z|x) = F_2^\epsilon(z - m_2(x)).$$

Assume the aforementioned operators

$$T_j : \mathcal{V}(\mathcal{W}) \mapsto \mathcal{V}(N^1(x_0, \delta)), j = 1, 2$$

are invertible. We now have:

THEOREM 7.1. *Suppose Assumptions 7.1, 7.2 and Condition 7.1 hold. Then $g_1(\cdot)$ and $g_2(\cdot)$ are identified.*

8. Nonparametric estimation for $J = 2$. This section uses the arguments developed in the previous sections, more specifically Section 3.3, to nonparametrically estimate (2.1) with a sample $(Z_p, X_p)_{p=1}^n$. Note that the number of mixture components is two.

We focus on two values x_0 and x_1 and first estimate the slopes of m_1 and m_2 nonparametrically. Define $\Delta_{10} m_1 = m_1(x_1) - m_1(x_0)$, $\Delta_{10} m_2 = m_2(x_1) - m_2(x_0)$ and, for $j = 0, 1$,

$$\phi^j(s) = \mathbb{E}(e^{isZ}|X = x_j),\ \widehat{\phi}^j(s) = \frac{\sum_{p=1}^n e^{isZ_p} K(\frac{X_p - x_j}{b_n})}{\sum_{p=1}^n K(\frac{X_p - x_j}{b_n})},\ \widehat{\rho}(s) = \frac{\widehat{\phi}^1(s)}{\widehat{\phi}^0(s)}$$

$$M^j(t) = \mathbb{E}(e^{tZ}|X = x_j),\ \widehat{M}^j(t) = \frac{\sum_{p=1}^n e^{tZ_p} K(\frac{X_p - x_j}{h_n})}{\sum_{p=1}^n K(\frac{X_p - x_j}{h_n})},\ \widehat{R}(t) = \frac{\widehat{M}^1(t)}{\widehat{M}^0(t)}$$

where h_n and b_n are carefully chosen bandwidths for kernel density estimation and $\widehat{M}^j(t)$ and $\widehat{\phi}^j(s)$ are the Nadaraya-Watson estimators of respectively the conditional moment generating function and conditional characteristic function of Z when $X = x_j$ for $j = 0, 1$. The

kernel function K may have a product form such as $K(X) = \Pi_{l=1}^{k} k(X^{(l)})$. Then our estimators are

$$\widehat{\Delta}_{10} m_1 = \frac{1}{t_n} \log \widehat{R}(t_n), \tag{8.1}$$

$$\widehat{\Delta}_{10} m_2 = \frac{-i}{a_n} \operatorname{Log}\left(\widehat{\rho}(s_n + a_n) \left(\widehat{\rho}(s_n)\right)^{-1} \right), \tag{8.2}$$

where $(a_n)_n, (s_n)_n$ and $(t_n)_n$ are sequences of tuning parameters such that $a_n \to 0$, $s_n \to \infty$ and $t_n \to \infty$ as $n \to \infty$. The notation $\operatorname{Log}(\cdot)$ as before corresponds to the principal value of the logarithm of $\cdot$. Define f_X the density of X and ϵ_i for $i = 1, 2$ as in (3.3). We make the following assumptions.

ASSUMPTION 8.1. (i) *$(Z_p, X_p)_{p=1}^{n}$ is i.i.d across p and distributed according to* (2.2)*,*
(ii) *f_X has continuous second order partial derivatives. f_X and all its first and second order partial derivatives are bounded on $\mathbb{R}^k$. Moreover, $f_X(x_j) > 0$ for $j = 0, 1$,*
(iii) *$\lambda \in (0, 1)$,*
(iv) *m_i, $i = 1, 2$ have continuous second order partial derivatives, and all their first and second order partial derivatives are bounded on $\mathbb{R}^k$,*
(v) *$h_n \underset{n\to\infty}{\to} 0$, $nh_n^k \underset{n\to\infty}{\to} \infty$, and $b_n \underset{n\to\infty}{\to} 0$, $nb_n^k \underset{n\to\infty}{\to} \infty$,*
(vi) *$t_n \underset{n\to\infty}{\to} \infty$, $t_n h_n \underset{n\to\infty}{\to} 0$, and $s_n \underset{n\to\infty}{\to} \infty$, $s_n b_n \underset{n\to\infty}{\to} 0$. Additionally, $a_n \underset{n\to\infty}{\to} 0$.*

ASSUMPTION 8.2. (i) *$\epsilon_1 | x \sim F_1$ and $\epsilon_2 | x \sim F_2$ at all $x \in \mathbb{R}^k$,*
(ii) *The domains of $M_1(t)$ and $M_2(t)$ are $[0, \infty)$,*
(iii) *For all ϵ, $e^{\epsilon t} \frac{M_2(t)}{M_1(t)} \underset{t\to\infty}{=} O(\mu(t))$ for some $\mu(\cdot)$ where $\mu(t) \underset{t\to\infty}{\to} 0$,*
(iv) *$\frac{\phi_1(s)}{\phi_2(s)} \underset{s\to\infty}{=} O(f(s))$ for some $f(\cdot)$ where $f(t) \underset{t\to\infty}{\to} 0$,*
(v) *$\frac{1}{nh_n^k} \frac{M_1(2t_n)}{M_1(t_n)^2} \underset{n\to\infty}{\to} 0$.*

Note that for simplicity, we exclude boundary cases by imposing $\lambda \in (0, 1)$. Moreover, Condition (iii) of Assumption 8.2 is a stronger version of Assumption 3.6 (i); this allows for x_1 and x_2 to be arbitrary and not necessarily close to each other.

We also impose assumptions on the kernel function.

ASSUMPTION 8.3. (i) $\int |K(U)| \, \mathrm{d}U < \infty$, $\int K(U) \, \mathrm{d}U = 1$, $\lim_{||U|| \to \infty} UK(U) \to 0$,
(ii) $\int K(U)^2 \, \mathrm{d}U < \infty$, $\int |K(U)| \, U'U \, \mathrm{d}U < \infty$, $\int K(U) \, U \, \mathrm{d}U = 0$,
(iii) *For some α_0, $\alpha \leq \alpha_0$ implies $\int e^{\alpha ||U||} |K(U)| \, U'U \, \mathrm{d}U < \infty$ and $\int e^{\alpha ||U||} K(U)^2 \, \mathrm{d}U < \infty$.*

LEMMA 8.1. *Suppose Assumptions 8.1, 8.2 and 8.3 hold. Then*

$$\widehat{\Delta}_{10} m_1 - \Delta_{10} m_1 = O_{\mathbb{P}}\left[\frac{\mu(t_n)}{t_n} + \frac{1}{t_n} \left(t_n^4 h_n^4 + \frac{1}{nh_n^k} \frac{M_1(2t_n)}{M_1(t_n)^2} \right)^{\frac{1}{2}} \right] \tag{8.3}$$

$$\widehat{\Delta}_{10} m_2 - \Delta_{10} m_2 = \frac{1}{a_n} O_{\mathbb{P}}\Bigg[f(s_n + a_n) + f(s_n) + \left(b_n^4 s_n^4 + \frac{1}{nb_n^k |\phi_2(s_n + a_n)|^2} \right)^{1/2} + \left(b_n^4 s_n^4 + \frac{1}{nb_n^k |\phi_2(s_n)|^2} \right)^{1/2} \Bigg] \tag{8.4}$$

These rates depend on the distributions of the error terms. Consider the case where these distributions are normal, i.e, $\epsilon_1|x \sim \mathcal{N}(0,\sigma_1^2)$ and $\epsilon_2|x \sim \mathcal{N}(0,\sigma_2^2)$ with $\delta = \sigma_1^2 - \sigma_2^2 > 0$. Then the rates are polynomial in n. To see this, note that Assumptions 8.2 (i)-(iv) are satisfied in this case. Indeed, for all ϵ, $e^{\epsilon t}\frac{M_2(t)}{M_1(t)} = e^{\epsilon t - \frac{\delta}{2}t^2} \underset{t\to\infty}{=} O(\mu(t))$, with $\mu(t) := e^{-(\frac{\delta}{2}-\nu)t^2}$ for any $0 < \nu < \frac{\delta}{2}$. And $\frac{\phi_1(s)}{\phi_2(s)} \underset{s\to\infty}{=} O(f(s))$ with $f(s) := e^{-\frac{1}{2}\delta s^2}$. Assumption 8.2 (v) requires t_n to increase slowly. We take a fixed a in the definition of $\widehat{\Delta}_{10}m_2$ here to simplify the computations, assuming a is small enough. Then applying the results from the estimation proofs, the convergence rates are

$$\widehat{\Delta}_{10}m_1 - \Delta_{10}m_1 = \frac{1}{t_n}O_{\mathbb{P}}\left[e^{-(\frac{\delta}{2}-\nu)t_n^2} + \left(t_n^4 h_n^4 + \frac{1}{nh_n^k}e^{\sigma_1^2 t_n^2}\right)^{\frac{1}{2}}\right],$$

$$\widehat{\Delta}_{10}m_2 - \Delta_{10}m_2 = O_{\mathbb{P}}\left[e^{-\frac{1}{2}\delta s_n^2} + \left(b_n^4 s_n^4 + \frac{1}{nb_n^k}e^{\sigma_2^2(s_n+a)^2}\right)^{\frac{1}{2}}\right].$$

With appropriate choices of the sequences t_h, h_n, s_n and b_n, the rates are polynomial in n. For example if $k = 1$, $h_n = n^{-\frac{1}{5}+\epsilon}$, $t_n = \frac{1}{\sigma_1}(\epsilon\log(n))^{\frac{1}{2}}$, $s_n = \frac{1}{\sigma_2}(\beta\log(n))^{\frac{1}{2}}$ and $b_n = n^{-\frac{1}{5}+\beta}$ for $\epsilon, \beta < \frac{1}{5}$, then $\widehat{\Delta}_{10}m_1 - \Delta_{10}m_1 = O_{\mathbb{P}}\left(n^{-\omega_1}\right)$ and $\widehat{\Delta}_{10}m_2 - \Delta_{10}m_2 = O_{\mathbb{P}}\left(n^{-\omega_2}\right)$ with $\omega_1 = \min\left(2/5 - 2\epsilon,\, (\delta-\nu)\epsilon/(2\sigma_1^2)\right)$ and $\omega_2 = \min\left(2/5 - 2\beta,\, \delta\beta/(2\sigma_2^2)\right)$. As expected, these rates decrease with δ.

We turn to estimation of the remaining parameters. Note that when $\lambda \neq 1$, we can write

$$\lambda = \frac{\mathbb{E}[z|x] - \mathbb{E}[z|x_0] - [m_2(x_0) - m_2(x)]}{[m_1(x_0) - m_1(x)] - [m_2(x_0) - m_2(x)]}.$$

Thus a natural estimator is

$$\widehat{\lambda} = \frac{\widehat{E}(Z|X=x_1) - \widehat{E}(Z|X=x_0) - \widehat{\Delta}_{10}m_2}{\widehat{\Delta}_{10}m_1 - \widehat{\Delta}_{10}m_2}, \tag{8.5}$$

where $\widehat{E}(W|X=x) = \frac{\sum_{p=1}^n W_p K\left(\frac{X_p - x}{d_n}\right)}{\sum_{p=1}^n K\left(\frac{X_p - x_j}{d_n}\right)}$ is the Nadaraya Watson estimator of $\mathbb{E}(W|X = x)$ with bandwidth d_n. To simplify notation, we use the same kernel function as for the previous estimators. Estimators of m_1 and m_2 evaluated at x_0 are obtained using Equation (3.6). Defining

$$\widehat{C} = \frac{1}{2}\left(\widehat{E}(Z^2|X=x_0) - \widehat{E}(Z^2|X=x_1) + \widehat{\lambda}\left(\widehat{\Delta}_{10}m_1\right)^2 + (1-\widehat{\lambda})\left(\widehat{\Delta}_{10}m_2\right)^2\right),$$

our estimators of $m_1(x_0)$ and $m_2(x_0)$ are

$$\begin{bmatrix}\widehat{m}_1(x_0)\\ \widehat{m}_2(x_0)\end{bmatrix} = \begin{pmatrix}\widehat{\lambda}^{-1} & 0\\ 0 & (1-\widehat{\lambda})^{-1}\end{pmatrix}\begin{pmatrix}-\widehat{\Delta}_{10}m_1 & -\widehat{\Delta}_{10}m_2\\ 1 & 1\end{pmatrix}^{-1}\begin{bmatrix}\widehat{C}\\ \widehat{E}(Z|X=x_0)\end{bmatrix}. \tag{8.6}$$

The convergence rate of these estimators can be computed easily. Under standard conditions and imposing $d_n = n^{-\frac{1}{k+4}}$, $\widehat{E}(W|X=x) - \mathbb{E}(W|X=x) = O_{\mathbb{P}}(n^{-\frac{2}{k+4}})$, see, e.g, [22]. Writing $\varepsilon_n = n^{-\frac{2}{k+4}} + \alpha_n + \beta_n$, one obtains $\widehat{\lambda} = \lambda + O_{\mathbb{P}}(\varepsilon_n)$. The estimators of $m_1(x_0)$ and $m_2(x_0)$ are linearizable functions of $\widehat{\lambda}$, $\widehat{\Delta}_{10}m_1$ and $\widehat{\Delta}_{10}m_2$ thus their rates of convergence are similarly bounded. This is summarized in the next proposition.

LEMMA 8.2. *If Assumptions 8.1, 8.2, 8.3 hold, letting $d_n = n^{-\frac{1}{k+4}}$, then*

(i) $\widehat{\lambda} = \lambda + O_{\mathbb{P}}(\varepsilon_n)$,
(ii) $\widehat{m}_i(x_0) = m_i(x_0) + O_{\mathbb{P}}(\varepsilon_n)$, *for* $i = 1, 2$.

To estimate the CDF of ϵ_1 and ϵ_2, we use Equation (3.8) and propose the following estimator

$$\widehat{F}_2(z) = 1 - \frac{1}{1-\widehat{\lambda}} \sum_{j=0}^{p_n} \widehat{F}(z + j\widehat{\delta}(x_1,x_0) + \widehat{m}_1(x_1) - \widehat{g}(x_0)|x_1) - \widehat{F}(z + j\widehat{\delta}(x_1,x_0) + \widehat{m}_2(x_0)|x_0), \tag{8.7}$$

for a truncation parameter $p_n \in \mathbb{N}$ with $p_n \to \infty$ as $n \to \infty$, and where $\widehat{g}(x) = \widehat{m}_1(x) - \widehat{m}_2(x)$, $\widehat{\delta}(x_1, x_0) = \widehat{\Delta}_{10} m_1 - \widehat{\Delta}_{10} m_2$, and

$$\widehat{F}(z|x) = \frac{\sum_{p=1}^n \mathbb{1}(Z_p \leq z) k(\frac{X_p - x}{c_n})}{\sum_{p=1}^n k(\frac{X_p - x}{c_n})}.$$

Note that we assume that $\delta(x_1, x_0) > 0$. The kernel function k may differ from the choice of kernel for our other estimators. As in [16], we let $\mathcal{K} = \left\{ k(\frac{x-\cdot}{c});\ c > 0,\ x \in \mathbb{R}^k \right\}$ and for $\alpha > 0$, $N(\alpha, \mathcal{K}) = \sup_Q N(||k||_\infty \alpha, \mathcal{K}, d_Q)$ where the supremum is over all probability measures Q on $(\mathbb{R}^k, \mathcal{B})$, d_Q is the $L_2(Q)$ metric and $N(., \mathcal{K}, d_Q)$ is the covering number for $(\mathcal{K}, d_Q)$.

ASSUMPTION 8.4. *For $i = 1, 2$, ϵ_i is continuously distributed. Let f_i be its density. We assume that f_i is twice differentiable on $\mathbb{R}$ and f_i, f_i' and f_i'' are continuous and bounded on their support.*

ASSUMPTION 8.5. *Let $k(.)$ satisfy Assumptions 8.3* (i) *and* (ii) *and*

(i) *k is continuous and $||k||_\infty < \infty$,*
(ii) *The kernel function k has support contained in $[-\frac{1}{2}, \frac{1}{2}]^k$,*
(iii) *There exist $A > 0$ and $\nu > 0$ such that for all $\alpha \in (0, 1)$, $N(\alpha, \mathcal{K}) \leq A\alpha^{-\nu}$.*

LEMMA 8.3. *If Assumptions 8.1 - 8.5 hold, letting $d_n = c_n = n^{-\frac{1}{k+4}}$ and $p_n \to \infty$ as $n \to \infty$, then for any small $a > 0$ and for some $\gamma_0 > 0$,*

$$\widehat{F}_1(z) - F_1(z) = O_{\mathbb{P}}\left(p_n n^{\frac{-2}{k+4}+a} + p_n^2 \varepsilon_n + e^{-\gamma_0 p_n} \right),$$

$$\widehat{F}_2(z) - F_2(z) = O_{\mathbb{P}}\left(p_n n^{\frac{-2}{k+4}+a} + p_n^2 \varepsilon_n + e^{-\gamma_0 p_n} \right).$$

9. Simulations. To illustrate our estimation procedure and show the finite sample properties of our estimators, we proceed here to a Monte Carlo exercise. Consider the following model of a mixture of nonparametric regressions with two components,

$$z = \begin{cases} m_1(x) + \epsilon_1, & \text{if } \tau = 1 \\ m_2(x) + \epsilon_2, & \text{if } \tau = 0 \end{cases}$$

$$\text{with: } \epsilon_1|x \sim \mathcal{N}(0, \sigma_1^2), \epsilon_2|x \sim \mathcal{N}(0, \sigma_2^2), \tau \sim \mathcal{B}ern(\lambda) \text{ where } \lambda = 0.3,$$
$$m_1(x) = 2\cos(0.3\,x),\ m_2(x) = 2\sin(0.5\,x),\ x \sim \mathcal{U}[0, 3].$$

We examine the case where $\sigma_1 > \sigma_2$. The assumptions of Section 8 hold[1] since $M_2(t)/M_1(t) = e^{-\sigma_\mu^2 t^2/2}$ and $\phi_1(t)/\phi_2(t) = e^{-\sigma_\mu^2 t^2/2}$, where we write $\sigma_\mu^2 = \sigma_1^2 - \sigma_2^2$. Our parameters of interest are $m_j(x_0)$, $j = 1, 2$ for $x_0 = 0.5$ as well as the mixing weight and the distribution of the error terms. Estimation of these objects relies on a second point on the support of x which we pick to be $x_1 = 1.2$. As explained in the Introduction, identifiability in our model comes from a "shift restriction": the independence assumption on $(\epsilon_1, \epsilon_2, x)$ constrains the shape of the conditional distribution of z given x as x varies. We illustrate this property in Figure 1.

We estimate $\Delta_{10} m_1 = m_1(x_1) - m_1(x_0) = -0.11$ and $\Delta_{10} m_2 = m_2(x_1) - m_2(x_0) = 0.63$ using $\widehat{\Delta}_{10} m_1$ and $\widehat{\Delta}_{10} m_2$ as defined by (8.1) and (8.2). As for the remainder of our simulation results, we report these estimates in Appendix A.1, computed for 500 simulations with sample size $n = 5,000$. Figure 2 displays these estimates for two different values for σ_μ, 0.2 and 0.4, and fixing σ_2 to be 0.1. These results show that the bias of the estimator for $\Delta_{10} m_2$ decreases as σ_μ gets larger. Indeed, the approximation error in the bias of $\widehat{\Delta}_{10} m_2$, see Equation (A.5), depends on the ratio of the characteristic functions $f(t) = \phi_1(t)/\phi_2(t)$ and as σ_μ increases, this ratio converges to 0 as $t \to \infty$ at a faster rate. Note however that the bias of the estimator of $\Delta_{10} m_1$ is much less sensitive to this variation of σ_μ, likely because the approximation term is divided by t, see Equation (A.1). On the other hand, as σ_μ grows (thus σ_1 grows), the variance of $\widehat{\Delta}_{10} m_1$ increases. This is related to the fact that a first order approximation of the variance term depends on the function $M_1(2t)/M_1(t)^2 = e^{\sigma_1^2 t^2}$.

The estimators for λ and $(m_1(x_0), m_2(x_0))$ are given by (8.5) and (8.6), respectively. Figure 3 graphs the obtained density of the estimators. Since $\delta(x_1, x_0) < 0$, the estimator of the cumulative distributive function $F_2(z)$ is not given by (8.7) but by

$$\widehat{F}_2(z) = -\frac{1}{1-\widehat{\lambda}} \sum_{j=0}^{p_n} \widehat{F}(z + j\widehat{\delta}(x_1, x_0) + \widehat{m}_1(x_1) - \widehat{g}(x_0)|x_1) \\ - \widehat{F}(z + j\widehat{\delta}(x_1, x_0) + \widehat{m}_2(x_0)|x_0),$$

where the functions $\widehat{g}$ and $\widehat{\delta}$ are as defined below Equation (8.7). Figure 4 shows the finite sample behavior of the estimator of F_2 evaluated at $z = 0$, compared to the true value $F_2(0) = 0.5$ and for different sample sizes. Convergence is unsurprisingly slower and large sample sizes are required for the estimator to show desirable properties such as validity of a gaussian approximation.

10. Conclusion. This paper develops various fully nonparametric identification results that build on the identifying power generated by covariate variations. Our constructive identification strategies directly lead to nonparametric estimators, which are shown to have polynomial rates of convergence. We demonstrate that our nonparametric identification framework applies to an important class of auction models where flexible treatments of unobserved heterogeneity is highly desirable. Extensions of the basic identification results to models where covariates are endogenous and situations where mixture distributions are correlated with covariates are investigated. Note that the number of mixture components is identified as well; that is, our identification results hold without prior knowledge of the number of mixtures. That said, as well known in the literature, estimation and inference of the number of mixture components remain a hard problem in general. Of course, it is of practical interest to study these issues in fully nonparametric models like ours, so that we can provide guidance as to how to choose the number of mixtures empirically. This is an important agenda for future research.

[1] We assume here that the researcher knows which identification procedure applies. This is likely not prior knowledge but the testable conditions we provide can be used to that effect. Studying asymptotic properties of a formal test is beyond the scope of this paper.

Acknowledgments. This paper supersedes a previously circulated manuscript entitled "Nonparametric identifiability of finite mixtures." The authors thank Werner Ploberger and participants at various seminars for their comments.

Funding. Kitamura acknowledges financial support from the National Science Foundation.

REFERENCES

[1] ADAMS, C. P. (2016). Finite mixture models with one exclusion restriction. *The Econometrics Journal* **19** 150–165.

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

[3] ARAGAM, B., DAN, C., XING, E. P. and RAVIKUMAR, P. (2020). Identifiability of nonparametric mixture models and Bayes optimal clustering. *The Annals of Statistics* **48** 2277 – 2302.

[4] ARAGAM, B. and YANG, R. (2023). Uniform consistency in nonparametric mixture models. *The Annals of Statistics* **51** 362 – 390. https://doi.org/10.1214/22-AOS2255

[5] ATHEY, S. and HAILE, P. A. (2007). Nonparametric approaches to auctions. *Handbook of econometrics* **6** 3847–3965.

[6] BERRY, S., CARNALL, M. and SPILLER, P. T. (1996). Airline hubs: costs, markups and the implications of customer heterogeneity Technical Report, National Bureau of Economic Research.

[7] BONHOMME, S., JOCHMANS, K. and ROBIN, J.-M. (2016). Estimating multivariate latent-structure models. *The Annals of Statistics* **44** 540–563.

[8] BONHOMME, S., JOCHMANS, K. and ROBIN, J.-M. (2016). Non-parametric estimation of finite mixtures from repeated measurements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **78** 211–229.

[9] BONHOMME, S., LAMADON, T. and MANRESA, E. (2019). A distributional framework for matched employer employee data. *Econometrica* **87** 699–739.

[10] BORDES, L., MOTTELET, S. and VANDEKERKHOVE, P. (2006). Semiparametric Estimation of A Two-Component Mixture Model. *The Annals of Statistics* **34** 1204–1232.

[11] BUTUCEA, C. and VANDEKERKHOVE, P. (2014). Semiparametric mixtures of symmetric distributions. *Scandinavian Journal of Statistics* **41** 227–239.

[12] COMPIANI, G., HAILE, P. and SANTÂĂŹANNA, M. (2020). Common values, unobserved heterogeneity, and endogenous entry in US offshore oil lease auctions. *Journal of Political Economy* **128** 3872–3912.

[13] COMPIANI, G. and KITAMURA, Y. (2016). Using mixtures in econometric models: a brief review and some new results. *The Econometrics Journal* **19** C95–C127.

[14] D'HAULTFŒUILLE, X. and FÉVRIER, P. (2015). Identification of mixture models using support variations. *Journal of Econometrics* **189** 70–82.

[15] ECHENIQUE, F. and KOMUNJER, I. (2009). Testing models with multiple equilibria by quantile methods. *Econometrica* **77** 1281–1297.

[16] EINMAHL, U. and MASON, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *The Annals of Statistics* **33** 1380–1403.

[17] ELLISON, G. (1994). Theories of cartel stability and the joint executive committee. *The Rand journal of economics* 37–57.

[18] HAILE, P. A., HONG, H. and SHUM, M. (2003). Nonparametric tests for common values at first-price sealed-bid auctions Technical Report, National Bureau of Economic Research.

[19] HAILE, P. A. and KITAMURA, Y. (2019). Unobserved heterogeneity in auctions. *The Econometrics Journal* **22** C1–C19.

[20] HALL, P. and ZHOU, X.-H. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *The annals of statistics* **31** 201–224.

[21] HAMILTON, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society* 357–384.

[22] HÄRDLE, W. and LINTON, O. (1994). Applied nonparametric methods. *Handbook of econometrics* **4** 2295–2339.

[23] HECKMAN, J. and SINGER, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica: Journal of the Econometric Society* 271–320.

[24] HECKMAN, J. J. and TABER, C. R. (1994). Econometric mixture models and more general models for unobservables in duration analysis. *Statistical Methods in Medical Research* **3** 279–299.

[25] HENRY, M., KITAMURA, Y. and SALANIÉ, B. (2014). Partial identification of finite mixtures in econometric models. *Quantitative Economics* **5** 123–144.
[26] HOHMANN, D. and HOLZMANN, H. (2013). Semiparametric location mixtures with distinct components. *Statistics* **47** 348–362.
[27] HOHMANN, D. and HOLZMANN, H. (2013). Two-component mixtures with independent coordinates as conditional mixtures: Nonparametric identification and estimation. *Electronic Journal of Statistics* **7** 859–880.
[28] HOROWITZ, J. L. and MANSKI, C. F. (1995). Identification and robustness with contaminated and corrupted data. *Econometrica: Journal of the Econometric Society* 281–302.
[29] HUANG, M., LI, R. and WANG, S. (2013). Nonparametric mixture of regression models. *Journal of the American Statistical Association* **108** 929–941.
[30] HUNTER, D. R., WANG, S. and HETTMANSPERGER, T. P. (2007). Inference for mixtures of symmetric distributions. *The Annals of Statistics* **35** 224 – 251. https://doi.org/10.1214/009053606000001118
[31] JEWELL, N. P. (1982). Mixtures of exponential distributions. *The annals of statistics* 479–484.
[32] JOCHMANS, K., HENRY, M. and SALANIÉ, B. (2017). Inference on two-component mixtures under tail restrictions. *Econometric Theory* **33** 610–635.
[33] KASAHARA, H. and SHIMOTSU, K. (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica* **77** 135–175.
[34] KIEFER, N. M. (1978). Discrete parameter variation: Efficient estimation of a switching regression model. *Econometrica: Journal of the Econometric Society* 427–434.
[35] KLEIN, R. W. and SHERMAN, R. P. (2002). Shift restrictions and semiparametric estimation in ordered response models. *Econometrica* **70** 663–691.
[36] LEE, L.-F. and PORTER, R. H. (1984). Switching Regression Models with Imperfect Sample Separation Information–With an Application on Cartel Stability. *Econometrica: Journal of the Econometric Society* 391–418.
[37] LENTZ, R., PIYAPROMDEE, S. and ROBIN, J.-M. (2023). The anatomy of sortingâĂŤevidence from danish data. *Econometrica* **91** 2409–2455.
[38] LINDSAY, B. G. (1995). Mixture models: theory, geometry and applications. In *NSF-CBMS regional conference series in probability and statistics* i–163. JSTOR.
[39] MANSKI, C. F. (2003). *Partial identification of probability distributions*. Springer Science & Business Media.
[40] MILGROM, P. R. and WEBER, R. J. (1982). A theory of auctions and competitive bidding. *Econometrica: Journal of the Econometric Society* 1089–1122.
[41] PORTER, R. H. (1983). A study of cartel stability: the Joint Executive Committee, 1880-1886. *The Bell Journal of Economics* 301–314.
[42] QUANDT, R. E. (1972). A new approach to estimating switching regressions. *Journal of the American statistical association* **67** 306–310.
[43] RAO, B. P. (1992). *Identifiability in stochastic models: characterization of probability distributions*. Academic Press.
[44] TEICHER, H. (1961). Identifiability of mixtures. *The annals of Mathematical statistics* **32** 244–248.
[45] TEICHER, H. (1963). Identifiability of finite mixtures. *The annals of Mathematical statistics* 1265–1269.
[46] VAN DEN BERG, G. J. (2001). Duration models: specification, identification and multiple durations. In *Handbook of econometrics*, **5** 3381–3460. Elsevier.
[47] VANDERMEULEN, R. A. and SAITENMACHER, R. (2024). Generalized identifiability bounds for mixture models with grouped samples. *IEEE Transactions on Information Theory* **70** 2746–2758.
[48] VANDERMEULEN, R. A. and SCOTT, C. D. (2019). An operator theoretic approach to nonparametric mixture models. *The Annals of Statistics* **47** 2704–2733.
[49] XIANG, S., YAO, W. and YANG, G. (2019). An Overview of Semiparametric Extensions of Finite Mixture Models. *Statistical Science* **34** 391–404.

SUPPLEMENTAL APPENDIX FOR "NONPARAMETRIC ANALYSIS OF FINITE MIXTURES"

YUICHI KITAMURA AND LOUISE LAAGE

This Appendix includes simulation results, the proofs of the results presented in the main text as well as additional results.

A.1. Simulation Results

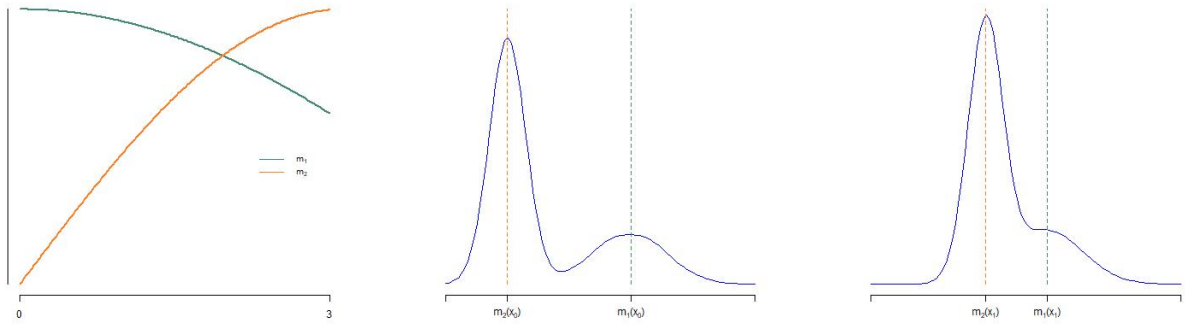

FIGURE 1. *Left: Regression functions m_1 (green) and m_2 (orange) on the support of x. Center and right: Conditional density of z conditional on $x = x_0$ (center) and $x = x_1$ (right).*

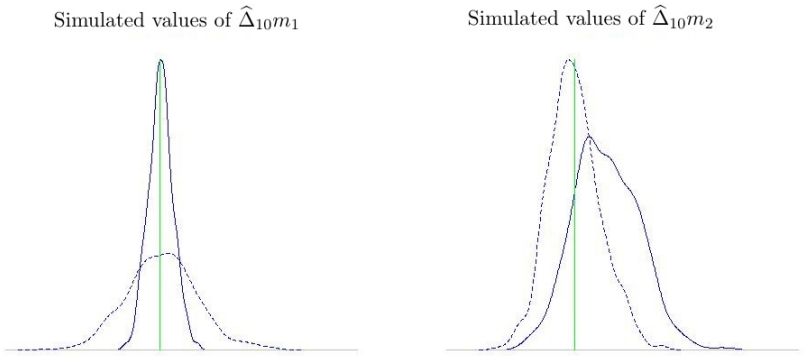

FIGURE 2. *Densities of the simulated estimators* $\widehat{\Delta}_{10}m_1$ *(left) and* $\widehat{\Delta}_{10}m_2$ *(right) for* $\sigma_\mu = 0.4$ *(dashed line) and* $\sigma_\mu = 0.2$ *(solid line). The vertical lines indicate the true values. Sample size:* $n = 5,000$.

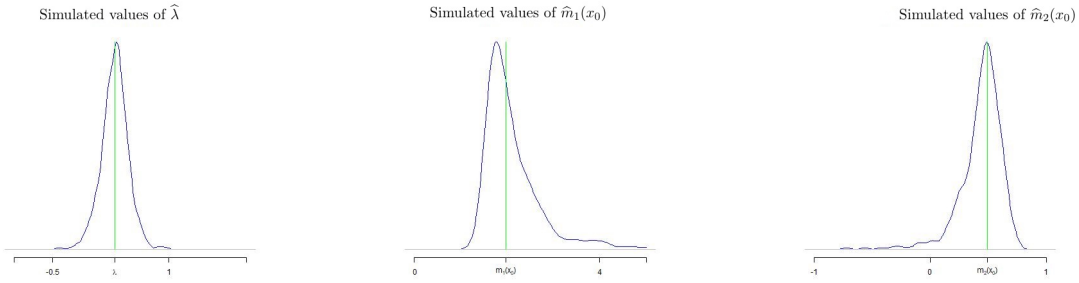

FIGURE 3. *Densities of the simulated estimators* $\widehat{\lambda}$, $\widehat{m}_1(x_0)$ *and* $\widehat{m}_2(x_0)$ *(in that order, from left to right) for* $\sigma_\mu = 0.4$. *The vertical lines indicate the true values. Sample size:* $n = 5,000$.

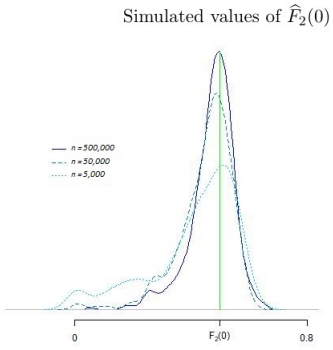

FIGURE 4. *Densities of the simulated estimators* $\widehat{F}_2(0)$ *for different sample sizes and* $\sigma_\mu = 0.4$. *The vertical line indicates the true value. Sample size:* $n = 5,000$.

A.2. Proofs of Results in Section 3

Proof of Lemma 3.1. We first consider the case with $0 < \lambda < 1$ and $D(x_0) \neq 0$. By continuity, i.e., Assumption 3.1(iii), there exist $\delta' \in (0, \delta)$ such that

$$|m_2(x') - m_2(x_0)| < \frac{\epsilon |D(x_0)|}{2} \qquad \text{and} \qquad |m_1(x') - m_1(x_0)| < \frac{\epsilon |D(x_0)|}{2} \tag{A.1}$$

for all $x' \in N^1(x_0, \delta')$. By (3.2) we have

$$M(t|x) = \lambda e^{tm_1(x)} M_1(t) + (1-\lambda) e^{tm_2(x)} M_2(t). \tag{A.2}$$

Now we prove part (i), i.e., the result with $t \to \infty$. Suppose $h(\pm\epsilon, t) = O(1)$ holds. Write

$$\begin{aligned}
\frac{1}{t} \log R(x', t) &= \frac{1}{t} \log \left(\frac{\lambda e^{tm_1(x')} M_1(t) + (1-\lambda) e^{tm_2(x')} M_2(t)}{\lambda e^{tm_1(x_0)} M_1(t) + (1-\lambda) e^{tm_2(x_0)} M_2(t)} \right) \\
&= m_1(x') - m_1(x_0) + \frac{1}{t} \log \left(\frac{\lambda + (1-\lambda) e^{t[m_2(x') - m_1(x')]} \frac{M_2(t)}{M_1(t)}}{\lambda + (1-\lambda) e^{t[m_2(x_0) - m_1(x_0)]} \frac{M_2(t)}{M_1(t)}} \right).
\end{aligned}$$

Note that (A.1) guarantees that $|m_2(x') - m_1(x')|$ is less than $|D(x_0)|(1+\epsilon)$. We have

$$\lim_{t \to \infty} \frac{1}{t} \log R(x', t) = m_1(x') - m_1(x_0).$$

If $1/h(\pm\epsilon, t) = O(1)$ instead, then write

$$\frac{1}{t} \log R(x', t) = m_2(x') - m_2(x_0) + \frac{1}{t} \log \left(\frac{\lambda e^{t[m_1(x') - m_2(x')]} \frac{M_1(t)}{M_2(t)} + (1-\lambda)}{\lambda e^{t[m_1(x_0) - m_2(x_0)]} \frac{M_1(t)}{M_2(t)} + (1-\lambda)} \right)$$

and again by $|m_2(x') - m_1(x')| < |D(x_0)|(1+\epsilon)$ we obtain

$$\lim_{t \to \infty} \frac{1}{t} \log R(x', t) = m_2(x') - m_2(x_0).$$

Note that if both hold, then it has to be the case that $D(x_0) = 0$. Let us now consider the case $0 < \lambda < 1$ and $D(x_0) = 0$ (implying $m_1(x_0) = m_2(x_0)$) and prove the result with $t \to \infty$. Suppose $h(\pm\epsilon, t) = O(1)$ holds. If additionally $D(x') = m_2(x') - m_1(x') > 0$, we write

$$\lim_{t \to \infty} \frac{1}{t} \log R(x', t) = m_2(x') - m_2(x_0) + \lim_{t \to \infty} \frac{1}{t} \log \left(\frac{\lambda e^{t[m_1(x') - m_2(x')]} + (1-\lambda) \frac{M_2(t)}{M_1(t)}}{\lambda + (1-\lambda) \frac{M_2(t)}{M_1(t)}} \right) = m_2(x') - m_2(x_0).$$

On the other hand, if $D(x') = m_2(x') - m_1(x') < 0$, we write

$$\lim_{t \to \infty} \frac{1}{t} \log R(x', t) = m_1(x') - m_1(x_0) + \lim_{t \to \infty} \frac{1}{t} \log \left(\frac{\lambda + (1-\lambda) e^{t[m_2(x') - m_1(x')]} \frac{M_2(t)}{M_1(t)}}{\lambda + (1-\lambda) \frac{M_2(t)}{M_1(t)}} \right) = m_1(x') - m_1(x_0).$$

The analysis of the case with $1/h(\pm\epsilon, t) = O(1)$ is similar. The proof of part (ii) is similar. If $\lambda = 1$ (i.e. the mixing distribution is degenerate) we have

$$\frac{1}{t}\log R(x', t) = m_1(x') - m_1(x_0)$$

thus the claim trivially holds. □

Proof of Lemma 3.2 . First consider the case with $\lambda \in (0, 1)$. Suppose Condition 3.1(i) fails, i.e. $\lim_{t\to\infty} \frac{1}{t}\log R(t, x) = \lim_{t\to-\infty} \frac{1}{t}\log R(t, x)$ for every $x \in N^1(x_0, \delta')$. In view of Lemma 3.1 these limits are either equal to $m_1(x) - m_1(x_0)$ or $m_2(x) - m_2(x_0)$. Wlog suppose it is the former. Note

$$\mathbb{E}[z|x] = \lambda m_1(x) + (1-\lambda)m_2(x), \tag{A.3}$$

therefore

$$\begin{aligned}\mathbb{E}[z|x] - \mathbb{E}[z|x_0] &= \lambda[(m_1(x) - m_1(x_0)) - (m_2(x) - m_2(x_0))] + (m_2(x) - m_2(x_0)) \\ &= (1-\lambda)[(m_2(x) - m_2(x_0)) - (m_1(x) - m_1(x_0))] + (m_1(x) - m_1(x_0)).\end{aligned}$$

Using this

$$\begin{aligned}\lambda_c &= \frac{(1-\lambda)[(m_2(x) - m_2(x_0)) - (m_1(x) - m_1(x_0))] + (m_1(x) - m_1(x_0)) - (1+c)(m_1(x) - m_1(x_0))}{(m_1(x) - m_1(x_0)) - (1+c)(m_1(x) - m_1(x_0))} \\ &= -\frac{(1-\lambda)[(m_2(x) - m_2(x_0)) - (m_1(x) - m_1(x_0))]}{c(m_1(x) - m_1(x_0))} + 1.\end{aligned}$$

Thus Condition 3.1(ii) does not hold either. In sum, if $\lambda \neq 1$ then Condition 3.1 reduces to its first part, i.e. Condition 3.1(i). Lemma 3.1 and Condition 3.1(i) imply either

$$\lim_{t\to+\infty} \frac{1}{t}\log R(t, x) = m_1(x) - m_1(x_0), \quad \lim_{t\to-\infty} \frac{1}{t}\log R(t, x) = m_2(x) - m_2(x_0)$$

or

$$\lim_{t\to+\infty} \frac{1}{t}\log R(t, x) = m_2(x) - m_2(x_0), \quad \lim_{t\to-\infty} \frac{1}{t}\log R(t, x) = m_1(x) - m_1(x_0).$$

Either way the slopes are identified. If the former holds, then

$$\begin{aligned}\lambda_c &= \frac{(1-\lambda)[(m_2(x) - m_2(x_0)) - (m_1(x) - m_1(x_0))] + (m_1(x) - m_1(x_0)) - (1+c)(m_2(x) - m_2(x_0))}{(m_1(x) - m_1(x_0)) - (1+c)(m_2(x) - m_2(x_0))} \\ &\to \lambda\end{aligned}$$

as $c \downarrow 0$, which identifies λ. In the latter case re-labeling delivers the result, with λ replaced by $1 - \lambda$.

Next, consider the case with $\lambda = 1$. Then Condition 3.1(i) cannot hold; as noted before

$$\frac{1}{t}\log R(x', t) = m_1(x') - m_1(x_0)$$

(which identifies the slope). On the other hand

$$\begin{aligned}\lambda_c &= \frac{(m_1(x) - m_1(x_0)) - (1+c)(m_1(x) - m_1(x_0))}{(m_1(x) - m_1(x_0)) - (1+c)(m_1(x) - m_1(x_0))} \\ &= 1,\end{aligned}$$

so indeed Condition 3.1(ii) is consistent with $\lambda = 1$. Moreover this shows that the limit of λ_c once again identifies λ. □

Proof of Lemma 3.3. Define $\dot{M}(t,x) = \frac{\partial}{\partial t}M(t,x)$, $\ddot{M}(t,x) = \frac{\partial^2}{(\partial t)^2}M(t,x)$, $\ddot{M}_i(t) = \frac{\partial^2}{(\partial t)^2}M_i(t), i = 1,2$ and $\ddot{M}_i(t) = \frac{\partial^2}{(\partial t)^2}M_i(t), i = 1,2$, whose existences follow from Assumption 3.2 (i).

Note

$$\dot{M}(0|x) = \int z dF(z|x) = \lambda m_1(x) + (1-\lambda)m_2(x). \tag{A.4}$$

Using this,

$$\begin{aligned}\dot{M}(0|x_0) - \dot{M}(0|x) &= \lambda[(m_1(x_0) - m_1(x)) - (m_2(x_0) - m_2(x))] + (m_2(x_0) - m_2(x)) \\ &= (1-\lambda)[(m_2(x_0) - m_2(x)) - (m_1(x_0) - m_1(x))] + (m_1(x_0) - m_1(x)).\end{aligned}$$

Next, to show that $m_1(x_0)$ and $m_2(x_0)$ are identified, note the basic relationship of the first and second order moments:

$$\ddot{M}(0|x) = \lambda[m_1(x)^2 + \ddot{M}_1(0)] + (1-\lambda)[m_2(x)^2 + \ddot{M}_2(0)].$$

Therefore

$$\begin{aligned}\ddot{M}(0|x_0) - \ddot{M}(0|x) =& \lambda[m_1(x_0)^2 - m_1(x)^2] + (1-\lambda)[m_2(x_0)^2 - m_2(x)^2] \\ =& \lambda(2m_1(x_0) - [m_1(x_0) - m_1(x)])[m_1(x_0) - m_1(x)] \\ &+ (1-\lambda)(2m_2(x_0) - [m_2(x_0) - m_2(x)])[m_2(x_0) - m_2(x)].\end{aligned}$$

By definition,

$$C(x) = \left\{\ddot{M}(0|x_0) - \ddot{M}(0|x) + \lambda[m_1(x_0) - m_1(x)]^2 + (1-\lambda)[m_2(x_0) - m_2(x)]^2\right\}/2,$$

then

$$C(x) = [m_1(x_0) - m_1(x)]\lambda m_1(x_0) + [m_2(x_0) - m_2(x)](1-\lambda)m_2(x_0).$$

Notice that $C(x)$ is already identified over $N^1(x_0, \delta')$ from the above argument and Lemma 3.1. Together with (A.4), we retrieve Equation (3.6), that is,

$$\begin{bmatrix} C(x) \\ \dot{M}(0|x_0) \end{bmatrix} = \begin{pmatrix} [m_1(x_0) - m_1(x)] & [m_2(x_0) - m_2(x)] \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & (1-\lambda) \end{pmatrix} \begin{bmatrix} m_1(x_0) \\ m_2(x_0) \end{bmatrix},$$

for all $x \in N^1(x_0, \delta')$. By Assumptions 3.1(ii), this can be uniquely solved for $m_1(x_0)$ and $m_2(x_0)$ (if $\lambda = 1$, the above equation can be solved directly to determine $m_1(x_0)$; another way to proceed in the degenerate case is to solve (3.6) using the Moore-Penrose generalized inverse, which identifies $m_1(x_0)$ and yields the solution that $m_2(x_0) = 0$). As the slopes are already obtained in Lemma 3.2, the levels of m_1 and m_2 over $N^1(x_0, \delta)$ are also identified. The only components remaining are F_1 and F_2. By evaluating (A.2) at x_0 and $x \in N^1(x_0, \delta')$, $x \neq x_0$, obtain

$$\begin{bmatrix} M(t|x_0) \\ M(t|x) \end{bmatrix} = E(x_0, x, t)\Lambda \begin{bmatrix} M_1(t) \\ M_2(t) \end{bmatrix}, \tag{A.5}$$

where

$$E(x, x', t) = \begin{pmatrix} e^{tm_1(x)} & e^{tm_2(x)} \\ e^{tm_1(x')} & e^{tm_2(x')} \end{pmatrix}, \Lambda = \begin{pmatrix} \lambda & 0 \\ 0 & (1-\lambda) \end{pmatrix}.$$

If the mixing distribution is non-degenerate,

$$\begin{aligned} \mathrm{Det}(E(x_0, x, t)) &= e^{t[m_1(x_0)+m_2(x)]} - e^{t[m_1(x)+m_2(x_0)]} \\ &= e^{t[m_1(x_0)+m_2(x)]} \left(1 - e^{t\{[m_1(x)-m_1(x_0)]-[m_2(x)-m_2(x_0)]\}}\right) \\ &\neq 0 \end{aligned}$$

for all $x \in N^1(x_0, \delta)$, $x \neq x_0$, $t \neq 0$, because of Assumption 3.1(ii), guaranteeing the invertibility of $E(x_0, x', t)$. Moreover,

$$E(x_0, x, t) = e^{t[m_1(x_0)+m_2(x_0)]} \begin{pmatrix} e^{-tm_2(x_0)} & e^{-tm_1(x_0)} \\ e^{t\{[m_1(x)-m_1(x_0)]-m_2(x_0)\}} & e^{t\{[m_2(x)-m_2(x_0)]-m_1(x_0)\}} \end{pmatrix}.$$

Therefore $E(x_0, x, t)$ for all $x \in N^1(x_0, \delta')$ and t are identified from the above argument and Lemma 3.1. Evaluate (A.5) at an arbitrary $x \in N^1(x_0, \delta')$ and solve it to determine $M_1(\cdot)$ and $M_2(\cdot)$. If $\lambda = 1$, solve (A.5) directly to identify M_1 (or, alternatively, use the Moore-Penrose generalized inverse as before). Since distribution functions are uniquely determined by their Laplace transforms (see, for example, Feller (1968), p.233), $F_1(\cdot)$ and $F_2(\cdot)$ are uniquely determined. This completes the proof. □

Proof of Lemma 3.4. Define $\delta(x) := m_2(x) - m_1(x)$. Note

$$\rho(x,t) = e^{it[m_1(x)-m_1(x_0)]}\frac{1+\frac{1-\lambda}{\lambda}e^{it\delta(x)}\frac{\phi_2(t)}{\phi_1(t)}}{1+\frac{1-\lambda}{\lambda}e^{it\delta(x_0)}\frac{\phi_2(t)}{\phi_1(t)}} \tag{A.6}$$

$$= e^{it[m_2(x)-m_2(x_0)]}\frac{\frac{1-\lambda}{\lambda}e^{-it\delta(x)}\frac{\phi_1(t)}{\phi_2(t)}+1}{\frac{1-\lambda}{\lambda}e^{-it\delta(x_0)}\frac{\phi_1(t)}{\phi_2(t)}+1} \tag{A.7}$$

The treatment of the case with $\lambda = 1$ is trivial, thus we maintain that $\lambda \in (0,1)$ in the rest of the proof. It is enough to prove the necessity, since the sufficiency follows from (A.6) and (A.7), with the constant in the second condition being either $m_1(x) - m_1(x_0)$ or $m_2(x) - m_2(x_0)$. So suppose the necessity fails, i.e. Condition 3.2 holds but also

$$\limsup_{t\to\infty}\left|\frac{\phi_1(t)}{\phi_2(t)}\right| = C, C \in (0,\infty] \tag{A.8}$$

and

$$\limsup_{t\to\infty}\left|\frac{\phi_2(t)}{\phi_1(t)}\right| = C', C' \in (0,\infty]. \tag{A.9}$$

hold. Then if either C or C' is finite (so suppose C is) then there exists a sequence $\{t_k\}_{k=1}^{\infty}$ such that $\lim_{k\to\infty} t_k = \infty$ and $\lim_{k\to\infty}\left|\frac{\phi_1(t_k)}{\phi_2(t_k)}\right| = C$. But then with the first part of Condition 3.2 and (A.7) we have to have

$$\lim_{k\to\infty}\left|\frac{\frac{1-\lambda}{\lambda}e^{-it_k\delta(x)}\frac{\phi_1(t_k)}{\phi_2(t_k)}+1}{\frac{1-\lambda}{\lambda}e^{-it_k\delta(x_0)}\frac{\phi_1(t_k)}{\phi_2(t_k)}+1}\right| = 1, x \in N^1(x_0,\epsilon).$$

which holds only if

$$\lim_{k\to\infty}\left[\mathrm{Arg}\left(\left(\frac{\phi_1(t_k)}{\phi_2(t_k)}\right)^2\right) - \left(t_k[\delta(x)-\delta(x_0)] + 2\pi\left\lfloor\frac{1}{2} - \frac{t_k[\delta(x)-\delta(x_0)]}{2\pi}\right\rfloor\right)\right] = 0$$

at every $x \in N^1(x_0,\epsilon)$. Under the non-parallel hypothesis this is impossible. Finally, if both C and C' are infinite, then there exits two sequences $\{t_k\}_{k=1}^{\infty}$ and $\{s_k\}_{k=1}^{\infty}$ such that $\lim_{k\to\infty} t_k = \infty$, $\lim_{k\to\infty} s_k = \infty$, $\lim_{k\to\infty}\left|\frac{\phi_1(t_k)}{\phi_2(t_k)}\right| = \infty$ and $\lim_{k\to\infty}\left|\frac{\phi_2(s_k)}{\phi_1(s_k)}\right| = \infty$. With (A.6) and (A.7), these imply that for sufficiently small a

$$\lim_{k\to\infty}\frac{-i}{a}\mathrm{Log}\left(\frac{\rho(x,t_k+a)}{\rho(x,t_k)}\right) = m_1(x) - m_1(x_0)$$

and

$$\lim_{k\to\infty}\frac{-i}{a}\mathrm{Log}\left(\frac{\rho(x,s_k+a)}{\rho(x,s_k)}\right) = m_2(x) - m_2(x_0)$$

hold simultaneously, which contradicts the second part of Condition 3.2 under the non-parallel assumption. □

Proof of Lemma 3.5. The proof proceeds in two steps. Step 1 considers the slopes of m_1 and m_2. Using the results in Step 1, Step 2 establishes the identification of all the parameters.

(Step 1)

By (3.2)

$$\phi(t|x) = \lambda e^{itm_1(x)}\phi_1(t) + (1-\lambda)e^{itm_2(x)}\phi_2(t). \tag{A.10}$$

Suppose there exists an alternative set of parameters

$$(\lambda^*, m_1^*(x_a), m_1^*(x_b), m_1^*(x_c), m_2^*(x_a), m_2^*(x_b), m_2^*(x_c), F_1^*(\cdot), F_2^*(\cdot))$$

in $\mathbb{R}^7 \times \bar{\mathcal{F}}(\mathbb{R})^2$ such that

$$F(z|x) = \lambda^* F_1^*(z - m_1^*(x)) + (1-\lambda^*)F_2^*(z - m_2^*(x)), \qquad x = x_a, x_b, x_c. \tag{A.11}$$

Let ϕ_1^* and ϕ_2^* denote the characteristic functions of F_1^* and F_2^*. Then

$$\lambda e^{itm_1(x_a)}\phi_1(t) + (1-\lambda)e^{itm_2(x_a)}\phi_2(t) = \lambda^* e^{itm_1^*(x_a)}\phi_1^*(t) + (1-\lambda^*)e^{itm_2^*(x_a)}\phi_2^*(t), \tag{A.12}$$

$$\lambda e^{itm_1(x_b)}\phi_1(t) + (1-\lambda)e^{itm_2(x_b)}\phi_2(t) = \lambda^* e^{itm_1^*(x_b)}\phi_1^*(t) + (1-\lambda^*)e^{itm_2^*(x_b)}\phi_2^*(t), \tag{A.13}$$

$$\lambda e^{itm_1(x_c)}\phi_1(t) + (1-\lambda)e^{itm_2(x_c)}\phi_2(t) = \lambda^* e^{itm_1^*(x_c)}\phi_1^*(t) + (1-\lambda^*)e^{itm_2^*(x_c)}\phi_2^*(t). \tag{A.14}$$

Let α and β be two arbitrary indices from the index set $\{a, b, c\}$. For a function $f : \mathbb{R}^k \to \mathbb{R}$, let $\Delta_{\alpha\beta} f$ denote the differences of the values of f at x_α and x_β, that is, $\Delta_{\alpha\beta} f = f(x_\alpha) - f(x_\beta)$. Define the following function of t that also depends on functions $f_1 : \mathbb{R}^k \to \mathbb{R}$, $f_2 : \mathbb{R}^k \to \mathbb{R}$ and indices α and β:

$$\begin{aligned} H(t; f_1, f_2, \alpha, \beta) &= e^{itf_2(x_\alpha)}\left(1 - e^{it(\Delta_{\alpha\beta}(f_1 - f_2)}\right) \\ &= e^{itf_2(x_\alpha)}\left(1 - e^{it\{[f_1(x_\alpha) - f_1(x_\beta)] - [f_2(x_\alpha) - f_2(x_\beta)]\}}\right). \end{aligned}$$

Now, multiply (A.13) by $e^{it\Delta_{ab}m_2^*}$ then subtract both sides from (A.12) to obtain

$$\lambda H(t; m_2^*, m_1, a, b)\phi_1(t) + (1-\lambda)H(t; m_2^*, m_2, a, b)\phi_2(t) = \lambda^* H(t; m_2^*, m_1^*, a, b)\phi_1^*(t). \tag{A.15}$$

Repeat this with (A.13) and $e^{it\Delta_{ab}m_2^*}$ replaced by (A.14) and $e^{it\Delta_{ac}m_2^*}$:

$$\lambda H(t; m_2^*, m_1, a, c)\phi_1(t) + (1-\lambda)H(t; m_2^*, m_2, a, c)\phi_2(t) = \lambda^* H(t; m_2^*, m_1^*, a, c)\phi_1^*(t). \tag{A.16}$$

(A.15) and (A.16) imply

$$\lambda H(t; m_2^*, m_1, a, b)H(t; m_2^*, m_1^*, a, c)\phi_1(t) + (1-\lambda)H(t; m_2^*, m_2, a, b)H(t; m_2^*, m_1^*, a, c)\phi_2(t) \tag{A.17}$$

$$= \lambda^* H(t; m_2^*, m_1^*, a, b) H(t; m_2^*, m_1^*, a, c) \phi_1^*(t),$$

and

$$\lambda H(t; m_2^*, m_1^*, a, b) H(t; m_2^*, m_1, a, c) \phi_1(t) + (1-\lambda) H(t; m_2^*, m_1^*, a, b) H(t; m_2^*, m_2, a, c) \phi_2(t) \tag{A.18}$$
$$= \lambda^* H(t; m_2^*, m_1^*, a, b) H(t; m_2^*, m_1^*, a, c) \phi_1^*(t),$$

yielding

$$\lambda H(t; m_2^*, m_1, a, b) H(t; m_2^*, m_1^*, a, c) \phi_1(t) + (1-\lambda) H(t; m_2^*, m_2, a, b) H(t; m_2^*, m_1^*, a, c) \phi_2(t)$$
$$= \lambda H(t; m_2^*, m_1^*, a, b) H(t; m_2^*, m_1, a, c) \phi_1(t) + (1-\lambda) H(t; m_2^*, m_1^*, a, b) H(t; m_2^*, m_2, a, c) \phi_2(t),$$

or

$$\lambda \left[H(t; m_2^*, m_1^*, a, b) H(t; m_2^*, m_1, a, c) - H(t; m_2^*, m_1, a, b) H(t; m_2^*, m_1^*, a, c) \right] \phi_1(t) \tag{A.19}$$
$$= (1-\lambda) \left[H(t; m_2^*, m_1^*, a, b) H(t; m_2^*, m_2, a, c) - H(t; m_2^*, m_2, a, b) H(t; m_2^*, m_1^*, a, c) \right] \phi_2(t).$$

Divide both sides of (A.19) by $e^{m_1^*(x_a)}$ and rewriting:

$$\lambda e^{itu_1} \left[(1-e^{itu_{11}})(1-e^{itu_{12}}) - (1-e^{itu_{13}})(1-e^{itu_{14}}) \right] \phi_1(t)$$
$$= (1-\lambda) e^{itu_2} \left[(1-e^{itu_{21}})(1-e^{itu_{22}}) - (1-e^{itu_{23}})(1-e^{itu_{24}}) \right] \phi_2(t) \qquad \text{for all } t$$

where $u_1 = m_1(x_a)$, $u_2 = m_2(x_a)$, $u_{11} = \Delta_{ab}(m_2^* - m_1^*)$, $u_{12} = \Delta_{ac}(m_2^* - m_1)$, $u_{13} = \Delta_{ab}(m_2^* - m_1)$, $u_{14} = \Delta_{ac}(m_2^* - m_1^*)$, $u_{21} = \Delta_{ab}(m_2^* - m_1^*) = u_{11}$, $u_{22} = \Delta_{ac}(m_2^* - m_2)$, $u_{23} = \Delta_{ab}(m_2^* - m_2)$, $u_{24} = \Delta_{ac}(m_2^* - m_1^*) = u_{14}$.

First, consider the non-degenerate case, i.e. $\lambda \neq 1$. Define

$$L_1(t) = (1-e^{itu_{11}})(1-e^{itu_{12}}) - (1-e^{itu_{13}})(1-e^{itu_{14}})$$

and

$$L_2(t) = (1-e^{itu_{21}})(1-e^{itu_{22}}) - (1-e^{itu_{23}})(1-e^{itu_{24}}),$$

then

$$L_1(t) = e^{it(u_2-u_1)} \frac{1-\lambda}{\lambda} \frac{\phi_2(t)}{\phi_1(t)} L_2(t) \qquad \text{for all } t. \tag{A.20}$$

We now use the condition

$$\lim_{t\to\infty} \frac{\phi_2(t)}{\phi_1(t)} = 0 \tag{A.21}$$

from Assumption 3.4 (the treatment of the case with $\lim_{t\to\infty} \frac{\phi_1(t)}{\phi_2(t)} = 0$ is essentially identical). The following argument shows that

$$L_1(t) = 0 \qquad \text{for all } t. \tag{A.22}$$

Suppose (A.22) is false, i.e. suppose the set $A = \{t : L_1(t) \neq 0, \quad t \in \mathbb{R}\}$ is non-empty. Pick an arbitrary point t_0 from A. Then there exists an $\epsilon > 0$ such that $|L_1(t_0)| \geq \epsilon > 0$. But since $\lim_{t\to\infty} e^{it(u_2-u_1)} \frac{1-\lambda}{\lambda} \frac{\phi_2(t)}{\phi_1(t)} L_2(t) = 0$ under (A.21), together with (A.20), there exists $t_1(\epsilon) \in \mathbb{R}$ such that

$$|L_1(t)| < \frac{\epsilon}{2} \qquad \text{for all } t > t_1(\epsilon). \tag{A.23}$$

Because of the definition of t_0, it must be the case that $t_0 \leq t_1(\epsilon)$. Now, since $L_1(\cdot)$ is a sum of periodic functions, it is almost periodic (see, e.g. Dunford and Schwartz (1958))). Therefore there exists a positive number $l(\epsilon)$ such that for all $\tau \in \mathbb{R}$ one can find a $\xi(\tau, \epsilon, l(\epsilon)) \in [\tau, \tau + l(\epsilon)]$ such that

$$|L_1(t) - L_1(t + \xi(\tau, \epsilon, l(\epsilon)))| < \frac{\epsilon}{2} \qquad \text{for all } t \in \mathbb{R}. \tag{A.24}$$

In particular, evaluating (A.24) at $t = t_0$ and $\tau = -t_0 + t_1(\epsilon)$;

$$|L_1(t_0) - L_1(t_0 + \xi^*)| < \frac{\epsilon}{2} \tag{A.25}$$

where $\xi^* = \xi(-t_0 + t_1(\epsilon), \epsilon, l(\epsilon)))$. But $\xi^* \in [-t_0 + t_1(\epsilon), -t_0 + t_1(\epsilon) + l(\epsilon)]$, therefore $t_0 + \xi^* \leq t_0 - t_0 + t_1(\epsilon) = t_1(\epsilon)$. By (A.23),

$$|L_1(t_0 + \xi^*)| < \frac{\epsilon}{2}. \tag{A.26}$$

Using the triangle inequality, (A.25) and (A.26), conclude that

$$\begin{aligned} |L_1(t_0)| &\leq |L_1(t_0) - L_1(t_0 + \xi^*)| + |L_1(t_0 + \xi^*)| \\ &< \epsilon. \end{aligned}$$

But the ϵ was originally defined so that $|L_1(t_0)| \geq \epsilon$, contradicting the last inequality. Since the choice of $t_0 \in A$ was arbitrary, (A.22) is now proved.

Next, as $\lambda \neq 0$, (A.20) and (A.22) imply that

$$\phi_2(t) L_2(t) = 0 \qquad \text{for all } t.$$

But by the basic properties a characteristic function, $\phi_2(\cdot)$ is continuous and $\phi_2(1) = 0$. Therefore for a $d > 0$, $\phi_2(t) \neq 0$ for all $t \in [-d, d]$. It follows that $L_2(t) = 0$ for all $t \in [-d, d]$. Moreover, $L_2(t)$ is analytic on the entire complex plane, and $[-d, d]$ obviously has an accumulation point, therefore by

the identity theorem of analytic functions, $L_2(t) = 0$ for all $t \in \mathbb{R}$. In summary, $L_1(t) = L_2(t) = 0$ for all $t \in \mathbb{R}$, or:

$$(1 - e^{itu_{11}})(1 - e^{itu_{12}}) - (1 - e^{itu_{13}})(1 - e^{itu_{14}}) = 0 \tag{A.27}$$

and

$$(1 - e^{itu_{21}})(1 - e^{itu_{22}}) - (1 - e^{itu_{23}})(1 - e^{itu_{24}}) = 0 \tag{A.28}$$

for all t. These conditions in turn identify the slopes of m_1 and m_2, as shown by the subsequent argument.

Consider the following set of conditions

(C1) $\Delta_{ab}(m_2^* - m_1^*) = \Delta_{ab}(m_2^* - m_1)$ and $\Delta_{ac}(m_2^* - m_1^*) = \Delta_{ac}(m_2^* - m_1)$,

(C2) $\Delta_{ab}(m_2^* - m_1^*) = \Delta_{ac}(m_2^* - m_1^*)$ and $\Delta_{ab}(m_2^* - m_1) = \Delta_{ac}(m_2^* - m_1)$,

(C3) $\Delta_{ab}(m_2^* - m_1^*) = \Delta_{ab}(m_2^* - m_2)$ and $\Delta_{ac}(m_2^* - m_1^*) = \Delta_{ac}(m_2^* - m_2)$,

(C4) $\Delta_{ab}(m_2^* - m_1^*) = \Delta_{ac}(m_2^* - m_1^*)$ and $\Delta_{ab}(m_2^* - m_2) = \Delta_{ac}(m_2^* - m_2)$.

Then by (A.27) and (A.28), if $u_{jk} \neq 0$ for all $j = 1, 2, k = 1, 2, 3, 4$, one of the following four cases has to be true:

(D1): (C1) and (C3) hold;

(D2): (C1) and (C4) hold;

(D3): (C2) and (C3) hold;

(D4): (C2) and (C4) hold.

First, consider (D1). (C1) and (C3) imply $\Delta_{ab}m_1^* = \Delta_{ab}m_1$ and $\Delta_{ab}m_1^* = \Delta_{ab}m_2$, respectively, thereby yielding $\Delta_{ab}m_1 = \Delta_{ab}m_2$, which violates Assumption 3.3(ii). Next, turn to (D2). From (C1) get $\Delta_{ab}m_1 = \Delta_{ab}m_1^*$ and $\Delta_{ac}m_1 = \Delta_{ac}m_1^*$, therefore $\Delta_{bc}m_1 = \Delta_{bc}m_1^*$. But (C4) also implies $\Delta_{bc}m_2^* = \Delta_{bc}m_1^*$ and $\Delta_{bc}m_2^* = \Delta_{bc}m_2$, hence $\Delta_{bc}m_1 = \Delta_{bc}m_2$, violating Assumption 3.3(ii). Since (D3) is identical to (D2) except for the switched roles of m_1 and m_2, it also violates Assumption 3.3(ii). Finally, (D4) also leads to a violation of Assumption 3.3(ii), because the second equations of (C2) and (C4) yield $\Delta_{bc}m_1 = \Delta_{bc}m_2$. As (D1)-(D4) are impossible, some of the u_{jk}'s should be non-zero. To consider the cases with some non-zero u_{jk}, it is useful to introduce the following classification (note that for $i = 1, 2$, if $u_{ij} = 0$ for $j = 1$ or 2 (3 or 4), then $u_{ij} = 0$ for $j = 3$ or 4 (1 or 2),

Case (i): $u_{11} = 0$

Case (ii): $u_{12} = u_{13} = 0$

Case (iii): $u_{14} = 0$

Case (iv): $u_{21} = 0$

Case (v): $u_{22} = u_{23} = 0$

Case (vi): $u_{24} = 0$

First consider Case (i). Then $H(t, m_2^*, m_1^*, a, b) = e^{itm_1^*(x_a)}(1 - e^{it(\Delta(m_2^* - m_1^*))}) = 0$. Therefore (A.15) becomes

$$\lambda H(t; m_2^*, m_1, a, b)\phi_1(t) + (1-\lambda)H(t; m_2^*, m_2, a, b)\phi_2(t) = 0, \tag{A.29}$$

or

$$(1 - e^{it\Delta_{ab}(m_2^* - m_1)}) + \frac{1-\lambda}{\lambda}\frac{\phi_2(t)}{\phi_1(t)}e^{-itm_1(x_a)}H(t; m_2^*, m_2, a, b) = 0. \tag{A.30}$$

Let $t \to \infty$, then again by (A.21), the third term goes to zero. Since the first term is periodic, it must be the case that $u_{13} = 0$ for all $t \in \mathbb{R}$. Since $1 - \lambda \neq 0$ in the current analysis of the non-degenerate case, $H(t; m_2^*, m_2, a, b)\phi_2(t) = 0$ for all t, or,

$$(1 - e^{itu_{23}})\phi_2(t) = 0 \qquad \text{for all } t.$$

As argued before, this means

$$1 - e^{itu_{23}} = 0 \qquad \text{for } t \in [-d, d]$$

for some $d > 0$. But this is possible iff $u_{23} = 0$. In sum, $u_{11} = 0$ automatically implies that $u_{13} = u_{23} = 0$ as well. But the latter condition means $\Delta_{ab}(m_2^* - m_1) = 0$ and $\Delta_{ab}(m_2^* - m_2) = 0$, which in turn imply $\Delta_{ab}(m_1 - m_2) = 0$, thereby violating Assumption 3.3(ii).

Next, consider Case (ii). This case means that

$$\begin{aligned} \Delta_{ab} m_2^* &= \Delta_{ab} m_1, \\ \Delta_{ac} m_2^* &= \Delta_{ac} m_1. \end{aligned} \tag{A.31}$$

On top of this, (A.28) has to hold at the same time. First, suppose all $u_{2k}, k = 1, 2, 3, 4$ in (A.28) are non-zero. Then (C3) and/or (C4) has to hold. Suppose (C3) holds. Then

$$\begin{aligned} \Delta_{ab} m_1^* &= \Delta_{ab} m_2, \\ \Delta_{ac} m_1^* &= \Delta_{ac} m_2. \end{aligned} \tag{A.32}$$

(A.31) and (A.32) imply that slopes of m_1^* and m_2^* have to coincide with those of m_2 and m_1, respectively, proving a part of the identification result. Next, suppose (C4) holds. In particular, the second equation of (C4), together with (A.31) means that $\Delta_{ab}(m_1 - m_2) = \Delta_{ac}(m_1 - m_2)$, or

$\Delta_{bc}m_1 = \Delta_{bc}m_2$, violating Assumption 3.3(ii). To complete the analysis of Case (ii), now suppose some of $u_{2k}, k = 1, 2, 3, 4$ in (A.28) are zero. If $u_{21} = 0$, then $u_{11} = 0$, but we have already shown that the latter condition leads to a violation of Assumption 3.3(ii). Next, suppose $u_{22} = 0$, i.e., $\Delta_{ac}m_2^* = \Delta_{ac}m_2$. But with the second equation of (A.31), $\Delta_{ac}m_1 = \Delta_{ac}m_2$, again violating Assumption 3.3(ii). If $u_{23} = 0$ or $u_{24} = 0$, it means at least one of u_{21} and u_{22} must be zero, so the above argument covers the cases. This completes the analysis of Case (ii); in sum, Case (ii) implies (A.31) and (A.32).

Case (iii) is identical to Case (i), with the roles of the indices b and c switched, therefore it violates Assumption 3.3(ii). Case (iv) is identical to Case (i). Note that case (v) is identical to Case (ii) with the role of the functions m_1 and m_2 reversed. But the treatment of Case (ii) only uses Equations (A.27) and (A.28), which are equivalent to (A.28) and (A.27), respectively, after switching m_1 and m_2. Therefore the above treatment of Case (ii) applies with m_1 and m_2 reversed; that is, Case (v) implies that

$$\Delta_{ab}m_1^* = \Delta_{ab}m_1, \tag{A.33}$$
$$\Delta_{ac}m_1^* = \Delta_{ac}m_1.$$

and

$$\Delta_{ab}m_2^* = \Delta_{ab}m_2, \tag{A.34}$$
$$\Delta_{ac}m_2^* = \Delta_{ac}m_2.$$

Finally, Case (vi) is identical to Case (vi).

The above arguments prove that if the mixture model is non-degenerate, the only possible cases are either (A): (A.31) and (A.32) hold, or (B): (A.33) and (A.34) hold. That is, the slopes of m_1 and m_2 are identified, up to labeling.

Next consider the case where the mixture model is degenerate, i.e. $\lambda = 1$. Then (A.11) is now written as

$$F_1(z - m_1(x)) = \lambda^* F_1^*(z - m_1^*(x)) + (1 - \lambda^*)F_2^*(z - m_2^*(x)). \tag{A.35}$$

Define $\sigma_1^{*2} = \int \epsilon^2 F_1^*(d\epsilon)$ and $\sigma_2^{*2} = \int \epsilon^2 F_2^*(d\epsilon)$. Taking the conditional variance of both sides given x,

$$\begin{aligned}\sigma_1^2 &= \lambda^*(m_1^*(x)^2 + \sigma_1^{*2}) + (1 - \lambda^*)(m_2^*(x)^2 + \sigma_2^{*2}) - [\lambda^* m_1^*(x) + (1 - \lambda^*)m_2^*(x)]^2 \\ &= \lambda^*(1 - \lambda^*)[m_1^*(x) - m_2^*(x)]^2 + \lambda^*\sigma_1^{*2} + (1 - \lambda^*)\sigma_2^{*2} \qquad \text{at } x = x_a, x_b \text{ and } x_c.\end{aligned}$$

This equation is used to establish identification for the degenerate case. In particular, it admits two solutions:

$$\lambda^* = 1, \quad \sigma_1^{*2} = \sigma_1^2, \tag{A.36}$$

$$[m_1^*(x_a) - m_2^*(x_a)]^2 = [m_1^*(x_b) - m_2^*(x_b)]^2 = [m_1^*(x_c) - m_2^*(x_c)]^2. \tag{A.37}$$

(A.36) obviously leads to full identification: integrating both sides of (A.35) gives $m_1^*(x) = m_1(x)$, and this trivially determines $F_1^*(z) = F_1(z)$ for all z. (A.37) implies that, for at least one pair of points, (x, x'), say, out of the three points $\{x_a, x_b, x_c\}$, the following holds:

$$m_1^*(x) - m_2^*(x) = m_1^*(x') - m_2^*(x'). \tag{A.38}$$

Unlike the case with $\lambda < 1$, this does not fully determine the slopes of m_1 and m_2 over $\{x_a, x_b, x_c\}$; it will be done in (Step 2).

(Step 2)

We now argue that λ is identified whether the model is degenerate or not. Let $m_j^*(x), j = 1, 2, x = x_1, x_b, x_c$ be (arbitrary) six numbers that satisfy (A.11). By (Step 1), in the case $\lambda \neq 1$, they have to satisfy (A.31) and (A.32), or, (A.33) and (A.34). Similarly, in the case $\lambda = 1$, they have to satisfy (A.38) (the case with (A.36) is trivial). For an arbitrary pair of points (x, x') from the three support points x_a, x_b, x_c, define

$$\lambda(x, x') = \lim_{\delta \downarrow 0} \frac{\int z F^*(dz|x) - \int z F^*(dz|x') - (1+\delta)(m_2^*(x) - m_2^*(x'))}{(m_1^*(x) - m_1^*(x')) - (1+\delta)(m_2^*(x) - m_2^*(x'))}.$$

Then λ is uniquely determined from the values $m_j^*(x), j = 1, 2, x = x_1, x_b, x_c$ by

$$\max_{(x,x') = (x_a, x_b), (x_a, x_c), (x_b, x_c)} \lambda(x, x'), \tag{A.39}$$

using an argument as in the proof of Lemma 3.3, up to labeling. It holds whether $\lambda < 1$ or not. (Note that the maximization in the line above is unnecessary if $\lambda \neq 1$, since $\lambda(x, x')$ is identical for all pairs (x, x') in that case.) Let $(\bar{x}, \bar{x}')$ be a maximizer of (A.39), which is possibly not unique.

Now, evaluating (3.6) at $(\bar{x}, \bar{x}')$ and $(\bar{x}', \bar{x})$, instead of (x, x_0) and solving for m_1 and m_2, obtain $m_1(\bar{x})$, $m_2(\bar{x})$, $m_1(\bar{x}')$ and $m_2(\bar{x}')$ ($m_1(\bar{x})$ and $m_1(\bar{x}')$ in the degenerate case).

To identify F_1 and F_2, use

$$\begin{bmatrix} \phi(t|\bar{x}) \\ \phi(t|\bar{x}') \end{bmatrix} = G(\bar{x}, \bar{x}', t) \Lambda \begin{bmatrix} \phi_1(t) \\ \phi_2(t) \end{bmatrix}, \tag{A.40}$$

where

$$G(x,x',t)=\begin{pmatrix} e^{itm_1(x)} & e^{itm_2(x)} \\ e^{itm_1(x')} & e^{itm_2(x')} \end{pmatrix}, \Lambda=\begin{pmatrix} \lambda & 0 \\ 0 & (1-\lambda) \end{pmatrix},$$

instead of (A.5) in the proof of Lemma 3.3. Then

$$\begin{aligned} \mathrm{Det}(G(\bar{x},\bar{x},t)) &= e^{it[m_1(\bar{x})+m_2(\bar{x}')]} - e^{it[m_1(\bar{x}')+m_2(\bar{x})]} \\ &= e^{it[m_1(\bar{x})+m_2(\bar{x}')]}\left(1-e^{it\{[m_1(\bar{x})-m_1(\bar{x}')]-[m_2(\bar{x}')-m_2(\bar{x})]}\right) \\ &\neq 0 \quad \text{for all } t \neq \frac{2\pi j}{[m_1(\bar{x})-m_1(\bar{x}')]-[m_2(\bar{x}')-m_2(\bar{x})]}, \quad j\in\mathbb{Z}. \end{aligned}$$

under Assumption 3.3(ii) if $\lambda<1$, therefore $G(\bar{x},\bar{x},t)$ is invertible (and all of its elements are identified). This determines $\phi_1(t)$ and $\phi_2(t)$ for all $t\neq 0$ for all $t\neq \frac{2\pi j}{[m_1(\bar{x})-m_1(\bar{x}')]-[m_2(\bar{x}')-m_2(\bar{x})]}, j\in\mathbb{Z}$ (as before, solve (A.40) directly or using the Moore-Penrose inverse in the degenerate case to determine ϕ_1). Since $\phi_1(t)$ and $\phi_2(t)$ are continuous, they are identified on $\mathbb{R}$. This identifies F_1 and F_2.

The foregoing argument shows that λ, $F_1(\cdot)$, $F_2(\cdot)$ and $m_1(x)$ and $m_2(x)$ evaluated at two points (i.e. $\bar{x}$ and $\bar{x}'$ defined right after (A.39)) out of the three support points $\{x_a,x_b,x_c\}$, are identified Note that m_1 and m_2 at the third point ($=\tilde{x}$, say) is identified by the relation

$$\phi(t|\tilde{x})=\lambda e^{itm_1(\tilde{x})}\phi_1(t)+(1-\lambda)e^{itm_2(\tilde{x})}\phi_2(t) \quad \text{for all } t.$$

Let

$$\begin{aligned} g_\tau(t) &= \frac{\phi(t+\tau|\tilde{x})}{\lambda\phi_1(t+\tau)} \\ &= e^{i(t+\tau)m_1(\tilde{x})}+\frac{1-\lambda}{\lambda}e^{i(t+\tau)m_2(\tilde{x})}\frac{\phi_2(t+\tau)}{\phi_1(t+\tau)}, \end{aligned}$$

Then under (A.21) the second term converges to zero as $\tau\to\infty$, and if we write, for all c,

$$h(c)=\lim_{\tau\to\infty}\frac{g_\tau(t+c)}{g_\tau(t)}=e^{icm_1(\tilde{x})},$$

then $m_1(\tilde{x})$ is uniquely determined by the formula $m_1(\tilde{x})=\frac{-ih'(c)}{h(c)}$.

If the model is non-degenerate, $m_2(\tilde{x})$ is identified from $e^{itm_2(\tilde{x})}=\frac{\phi(t|\tilde{x})-\lambda e^{itm_1(\tilde{x})}\phi_1(t)}{(1-\lambda)\phi_2(t)}$. □

Proof of Lemma 3.6. The proof of Part (i) is essentially in the proof of Lemma (3.1). For Part (ii), let us assume that $\lim_{t\to\infty}\left|\frac{\phi_1(t)}{\phi_2(t)}\right|\to 0$ and $0<\lambda<1$. Let $\dot{m}_2(x):=m_2(x)-m_2(x_0)$ and note that since

$$\rho(s,x)=e^{is\dot{m}_2(x)}\,\frac{\frac{\lambda}{1-\lambda}e^{is(m_1(x)-m_2(x))}\frac{\phi_1(s)}{\phi_2(s)}+1}{\frac{\lambda}{1-\lambda}e^{is(m_1(x_0)-m_2(x_0))}\frac{\phi_1(s)}{\phi_2(s)}+1},$$

the ratio on the right hand side of this equality converges to 1 as $s \to \infty$. Therefore we have

$$\lim_{s\to\infty} \frac{-i}{a} \operatorname{Log}\left(\frac{\rho(x, s+a)}{\rho(x,s)}\right) = \frac{-i}{a} \operatorname{Log}(e^{ia\dot{m}_2(x)}) \\ = \frac{1}{a}\left(a\dot{m}_2(x) + 2\pi\left\lfloor \frac{1}{2} - \frac{a\dot{m}_2(x)}{2\pi}\right\rfloor\right),$$

where Log corresponds to the principal value of the log. This limit is a piecewise continuous function of a, constant equal to $\dot{m}_2(x)$ only when a is small enough to guarantee $a\dot{m}_2(x) \in (-\pi, \pi)$. If $\lim_{t\to\infty}\left|\frac{\phi_2(t)}{\phi_1(t)}\right| \to 0$, we obtain the other equality in Lemma 3.6 (ii). If $\lambda = 1$, $\frac{\phi(s|x)}{\phi(s|x_0)} = e^{is\dot{m}_1(x)}$. Thus

$$\lim_{s\to\infty} \frac{-i}{a} \operatorname{Log}\left(\frac{\phi(s+a|x)}{\phi(s+a|x_0)}\left(\frac{\phi(s|x)}{\phi(s|x_0)}\right)^{-1}\right) = \frac{1}{a}\left(a\dot{m}_1(x) + 2\pi\left\lfloor \frac{1}{2} - \frac{a\dot{m}_1(x)}{2\pi}\right\rfloor\right).$$

□

Proof of Lemma 3.7. Similar to the proof of Lemma 3.2. □

Proof of Lemma 3.9. Recall that $g(x) := m_1(x) - m_2(x)$ and $\delta(x, x') := g(x) - g(x')$ and that we consider the case $\delta(x_1, x_0) > 0$. To obtain Equation (3.8), note that Equation (3.4) gives $\forall (x, z) \in \mathbb{R}^{k+1},\ F(z|x) = \lambda F_1(z - m_1(x)) + (1-\lambda)F_2(z - m_2(x))$ which implies

$$\forall (x, y) \in \mathbb{R}^{k+1},\ F(m_1(x) + y|x) = \lambda F_1(y) + (1-\lambda)F_2(m_1(x) - m_2(x) + y). \tag{A.41}$$

Applying Equation (A.41) to (x_0, y) and (x_1, y) and taking the difference, we obtain

$$F(m_1(x_1) + y|x_1) - F(m_1(x_0) + y|x_0) \\ = (1-\lambda)\left[F_2(m_1(x_1) - m_2(x_1) + y) - F_2(m_1(x_0) - m_2(x_0) + y)\right]. \tag{A.42}$$

This implies that $F_2(m_1(x_1) - m_2(x_1) + y) - F_2(m_1(x_0) - m_2(x_0) + y)$ is identified for all $y \in \mathbb{R}$. We obtain identification of F_2 using these differences and the fact that $F_2(z)$ converges to 1 as $z \to \infty$. For a given $z \in \mathbb{R}$, apply Equation (A.42) to $y = z - g(x_0)$ to obtain

$$F_2(z + \delta(x_1, x_0)) - F_2(z) = \frac{1}{1-\lambda}\left[F(z + m_1(x_1) - g(x_0)|x_1) - F(z + m_2(x_0)|x_0)\right],$$

and to $y = z + j\delta(x_1, x_0) - g(x_0)$ for any $j \in \mathbb{N}$ to obtain

$$F_2(z+(j+1)\delta(x_1, x_0)) - F_2(z + j\delta(x_1, x_0)) \\ = \frac{1}{1-\lambda}\left[F(z + j\delta(x_1, x_0) + m_1(x_1) - g(x_0)|x_1) - F(z + j\delta(x_1, x_0) + m_2(x_0)|x_0))\right].$$

Using $\lim_{j\to\infty} F_2\left(z+(j+1)\delta(x_1,x_0)\right)=1$, the identifying equation for $F_2(.)$ is

$$F_2(z)=1-\frac{1}{1-\lambda}\sum_{j=0}^{\infty}F(z+j\delta(x_1,x_0)+m_1(x_1)-g(x_0)|x_1)-F(z+j\delta(x_1,x_0)+m_2(x_0)|x_0),$$

where the infinite sum is a convergent series of positive terms. Finally (3.4) identifies $F_1(.)$ as

$$F_1(z)=\frac{1}{\lambda}\left[F(z+m_1(x))-(1-\lambda)F_2(z+m_1(x)-m_2(x)\right].$$

□

A.3. Proofs of Results in Section 4

Proof of Lemma 4.1. Take δ small enough so Assumptions 4.1, 4.2, 4.3 and Condition 4.1 hold. Then, by Assumption 4.3 there exists a $\delta'\leq\delta$ such that either

Case (A): $\lambda(x)=1$ for every $x\in N^1(x_0,\delta')$

Case (B): $\lambda(x)<1$ for every $x\in N^1(x_0,\delta')\setminus\{x_0\}$

Case (C): $\lambda(x)\begin{cases}=1 & \text{if } x^1\in(x_0-\delta',x_0], x\in N^1(x_0,\delta')\\ <1 & \text{if } x^1\in(x_0,x_0+\delta'), x\in N^1(x_0,\delta')\end{cases}$

or

Case (D): $\lambda(x)\begin{cases}=1 & \text{if } x^1\in[x_0,x_0+\delta'), x\in N^1(x_0,\delta')\\ <1 & \text{if } x^1\in(x_0-\delta',x_0), x\in N^1(x_0,\delta').\end{cases}$

Let $x\in N^1(x_0,\delta')$. For $q\in(0,1)$, let $x_q:=x_0+q(x-x_0)$. Also define $\dot{m}_1(x,x_q):=m_1(x)-m_1(x_q)$, $\dot{m}_2(x,x_q):=m_2(x)-m_2(x_q)$. Our identification strategy consists of the following steps:

Step 0: Fix $q\in(0,1)$.

Step 1: Calculate

$$r(\infty,x,x_q):=\lim_{t\to\infty}\frac{1}{t}\log R(t,x,x_q)$$

$$r(-\infty,x,x_q):=\lim_{t\to-\infty}\frac{1}{t}\log R(t,x,x_q)$$

This step is used to identify slopes connecting the points x and x_q.

Step 2: Calculate

$$K_{+\infty}(x,x_q):=\lim_{t\to\infty}K_{+\infty,t}(x,x_q),\quad K_{-\infty}(x,x_q):=\lim_{t\to-\infty}K_{-\infty,t}(x,x_q).$$

Check if $K_{+\infty}(x,x_q)=K_{-\infty}(x,x_q)$ or not. If it is, let $\lambda(x_q)=\lim_{c\downarrow0}\lambda_c(x,x_q)$ where

$$\lambda_c(x,x'):=\frac{\mathbb{E}[z|x]-\mathbb{E}[z|x']-(1+c)\lim_{t\to-\infty}\frac{1}{t}\log R(t,x,x')}{\lim_{t\to+\infty}\frac{1}{t}\log R(t,x,x')-(1+c)\lim_{t\to-\infty}\frac{1}{t}\log R(t,x,x')}.$$

If not, $\lambda(x_q) = \tilde{\lambda}(x, x_q)$ where

$$\tilde{\lambda}_c(x, x') := \frac{1 - K_{-\infty}(x, x')}{K_{+\infty}(x, x') - K_{-\infty}(x, x')}.$$

Step 3: Let $q \to 0$.

We first apply Steps 0-2 to Cases (A)-(D).

Case (A): Step 1 yields $r(\infty, x, x_q) = r(-\infty, x, x_q) = \dot{m}_1(x, x_q) := m_1(x) - m_1(x_q)$, as in this case,

$$R(t, x, x_q) = \frac{e^{tm_1(x)}}{e^{tm_1(x_q)}} = e^{t\dot{m}_1(x, x_q)}$$

and

$$\frac{1}{t} \log R(x, t) = \dot{m}_1(x, x_q),$$

identifying the slope $\dot{m}_1(x, x_q)$. We see that Condition 4.1(i) fails in this case. On the other hand this means $K_{+\infty,t}(x, x_q) = K_{-\infty,t}(x, x_q) = 1$ for every (t, x, q), therefore Condition 4.1(ii) holds. Proceeding to Step 2, it holds that $K_{+\infty}(x, x_q) = K_{-\infty}(x, x_q) = e^{t\dot{m}_1(x, x_q)} \exp(-t\dot{m}_1(x, x_q)) = 1$, and as in the proof of Lemma 3.2 we see $\lim_{c \downarrow 0} \lambda_c(x, x_q) = 1$ in this case.

Case (B): First suppose Condition 4.1(i) holds. Applying Step 1, the two slopes $(m_1(x) - m_1(x_q), m_2(x) - m_2(x_q))$ are identified as in the proof of Lemma 3.1. Take $\delta'' < \delta'$ as in the proof of Lemma 3.1. We first consider the case with t tending to $+\infty$. If $h(\pm\epsilon, t, x_q) = O(1)$ holds for t tending to $+\infty$, then according to the proof of Lemma 3.1 we have

$$\lim_{t \to \infty} \frac{1}{t} \log R(t, x, x_q) = r(+\infty, x, x_q) = m_1(x) - m_1(x_q),$$

and with Condition 4.1(i)

$$\lim_{t \to -\infty} \frac{1}{t} \log R(t, x, x_q) = r(-\infty, x, x_q) = m_2(x) - m_2(x_q),$$

for $x \in N^1(x_0, \delta'')$. In Step 2, note

$$K_{+\infty}(x, x_q) = \frac{\lambda(x)}{\lambda(x_q)} \text{ and } K_{-\infty}(x, x_q) = \frac{1 - \lambda(x)}{1 - \lambda(x_q)}.$$

With these results we see $K_{+\infty}(x, x_q) \neq K_{-\infty}(x, x_q)$ iff $\lambda(x) \neq \lambda(x_q)$. If $\lambda(x) \neq \lambda(x_q)$, note that

$$\tilde{\lambda}(x, x_q) = \frac{1 - K_{-\infty}(x, x_q)}{K_{+\infty}(x, x_q) - K_{-\infty}(x, x_q)} = \lambda(x_q),$$

which identifies $\lambda(x_q)$ since the left-hand side is identified. If $K_{+\infty}(x, x_q) = K_{-\infty}(x, x_q)$ we can obtain $\lambda(x_q)$ by $\lim_{c \downarrow 0} \lambda_c(x, x_q)$, as noted in the proof of Lemma 3.2.

If $1/h(\pm\epsilon, t, x_1) = O(1)$, we have

$$\lim_{t\to-\infty} \frac{1}{t} \log R(t, x, x_q) = m_1(x) - m_1(x_q), \quad K_{-\infty}(x, x_q) = \frac{\lambda(x)}{\lambda(x_q)}$$

and

$$\lim_{t\to\infty} \frac{1}{t} \log R(t, x, x_q) = m_2(x) - m_2(x_q), \quad K_{+\infty}(x, x_q) = \frac{1-\lambda(x)}{1-\lambda(x_q)}$$

respectively, and if $K_{+\infty}(x, x_q) \neq K_{-\infty}(x, x_q)$, $\tilde{\lambda}(x, x_q) = \frac{1-K_{-\infty}(x,x_q)}{K_{+\infty}(x,x_q)-K_{-\infty}(x,x_q)} = 1 - \lambda(x_q)$. If $K_{+\infty}(x, x_q) = K_{-\infty}(x, x_q)$, we have $\lim_{c\downarrow 0} \lambda_c(x, x_q) = 1 - \lambda(x_q)$.

If both $h(\pm\epsilon, t, x_q) = O(1)$ and $1/h(\pm\epsilon, t, x_q) = O(1)$ hold, $D(x_q) = 0$. If $D(x) > 0$, for example, then $r(x, x_q, -\infty) = \dot{m}_1(x, x_q)$ and $r(x, x_q, +\infty) = \dot{m}_2(x, x_q)$. In this case, Condition 4.1(i) is automatically satisfied. Moreover, $\lambda(x_q)$ is identified as above.

Next, suppose Condition 4.1(i) fails under Case (B). Wlog suppose $r(+\infty, x, x_q) = \dot{m}_1(x)$, then

$$K_{+\infty,t}(x, x_q) = \frac{\lambda(x) + (1-\lambda(x))e^{tD(x)}\frac{M_2(t)}{M_1(t)}}{\lambda(x_q) + (1-\lambda(x_q))e^{tD(x_q)}\frac{M_2(t)}{M_1(t)}}, \tag{A.1}$$

so for Condition 4.1(ii) to hold we need

$$\lambda(x) + (1-\lambda(x))e^{tD(x)}\frac{M_2(t)}{M_1(t)} = \lambda(x_q) + (1-\lambda(x_q))e^{tD(x_q)}\frac{M_2(t)}{M_1(t)}$$

or

$$\frac{\lambda(x) - \lambda(x_q)}{1-\lambda(x_q)} = \left[e^{tD(x_q)} - \frac{1-\lambda(x)}{1-\lambda(x_q)}e^{tD(x)}\right]\frac{M_2(t)}{M_2(t)}.$$

Note that $\lambda(x) \neq \lambda(x_q)$ as otherwise $K_{+\infty,t}(x, x_q) = 1$ for all t would imply together with Equation (A.1) that $D(x) = D(x_q)$ which does not satisfy Condition 4.1(ii) by the non-parallel assumption.

Thus by Assumption 4.1(iii), there exists $x_1 \neq x_q$ in $N^1(x_0, \delta')$ such that $\lambda(x_1) \neq \lambda(x_q)$. Then

$$\begin{aligned}
\frac{\frac{\lambda(x)-\lambda(x_q)}{1-\lambda(x_q)}}{\frac{\lambda(x_1)-\lambda(x_q)}{1-\lambda(x_q)}} &= \frac{1 - \frac{1-\lambda(x)}{1-\lambda(x_q)}e^{t[D(x)-D(x_q)]}}{1 - \frac{1-\lambda(x_1)}{1-\lambda(x_q)}e^{t[D(x_1)-D(x_q)]}} \\
&= \frac{1 - \frac{1-\lambda(x)}{1-\lambda(x_q)}e^{t[\dot{m}_2(x)-\dot{m}_1(x)]}}{1 - \frac{1-\lambda(x_1)}{1-\lambda(x_q)}e^{t[\dot{m}_2(x_1)-\dot{m}_1(x_1)]}}.
\end{aligned}$$

In view of the non-parallel assumption again, the right hand side does not depend on t only if $\lambda(x) = 1$, which is a contradiction. That is, in Case (B), under Condition 4.1 it has to be the case that Condition 4.1(i) holds, a situation already covered earlier.

Case (C): If $x^1 \in (x_0 - \delta', x_0]$ proceed as in Case (A) to identify $\dot{m}_1(x, x_q)$ and $\lambda(x_q)$ (the latter being 1 in this case). Otherwise, note that then $\lambda(x_1) < 1$ and $\lambda(x_q) < 1$: proceed as in Case (B).

Case (D): If $x^1 \in [x_0, x_0 + \delta')$ proceed as in Case (A) to identify $\dot{m}_1(x, x_q)$ and $\lambda(x_q)$ (the latter being 1 in this case). Otherwise proceed as in Case (B).

Finally apply Step 3, and under the continuity assumption we obtain the desired conclusion.

□

Proof of Lemma 4.2. Given Lemma 4.1 the only remaining task is to identify the levels of m_1 and m_2 at x_0, F_1 and F_2. Using the notation introduced in the proof of Lemma 4.1, with an additional definition

$$\dot{\lambda}(x) = \lambda(x) - \lambda(x_0),$$

write

$$\mathbb{E}[z|x] - \mathbb{E}[z|x_0] = \lambda(x)[\dot{m}_1(x) - \dot{m}_2(x)] + \dot{\lambda}(x)[m_1(x_0) - m_2(x_0)].$$

If $\lambda(x) \neq \lambda(x_0)$ then we can proceed as in the proof of Lemma 3.3 to show that $m_1(x_0)$ and $m_2(x_0)$ are identified. Accordingly, consider the case $\lambda(x) \neq \lambda(x_0)$. Define

$$c(x) := \frac{\mathbb{E}[z|x] - \mathbb{E}[z|x_0] - \lambda(x)[\dot{m}_1(x) - \dot{m}_2(x)]}{\dot{\lambda}(x)},$$

which is observable by Lemma 4.1, then $c(x) = m_1(x_0) - m_2(x_0)$, and we obtain

$$\begin{bmatrix} c(x) \\ \mathbb{E}[z|x] \end{bmatrix} = \begin{pmatrix} 1 & 1 \\ \lambda(x) & 1 - \lambda(x) \end{pmatrix} \begin{bmatrix} m_1(x_0) \\ m_2(x_0) \end{bmatrix}.$$

Since the determinant of the matrix on the right hand side is unity, once again $m_1(x_0)$ and $m_2(x_0)$ are identified. Finally, we proceed as in as in the proof of Lemma 3.3 to identify F_1 and F_2, though here the 2-by-2 matrix in the following display does not factorize:

$$\begin{bmatrix} M(t|x) \\ M(t|x_0) \end{bmatrix} = \begin{pmatrix} \lambda(x)e^{tm_1(x)} & (1-\lambda(x))e^{tm_1(x)} \\ \lambda(x_0)e^{tm_1(x)} & (1-\lambda(x_0))e^{tm_2(x_0)} \end{pmatrix} \begin{bmatrix} m_1(x_0) \\ m_2(x_0) \end{bmatrix}, \quad \text{for every } t \in \mathbb{R}. \tag{A.2}$$

Nevertheless, its determinant is, if $\lambda(x_0) \neq 1$

$$\begin{aligned} \lambda(x)(1-\lambda(x_0))e^{t[m_1(x)+m_2(x_0)]} &- \lambda(x_0)(1-\lambda(x))e^{t[m_1(x_0)+m_2(x)]} \\ &= \lambda(x_0)(1-\lambda(x_0))e^{t[m_1(x_0)+m_2(x)]} \left\{ \frac{\lambda(x)}{\lambda(x_0)} e^{t[\dot{m}_1(x)-\dot{m}_2(x)]} - \frac{1-\lambda(x)}{1-\lambda(x_0)} \right\} \end{aligned}$$

which is non-zero for almost all t under the non-parallel condition. Therefore (A.2) uniquely determines M_1 and M_2, hence F_1 and F_2. The treatment of the case with $\lambda(x_0) = 1$ is straightforward. □

A.4. Proofs of Results in Section 5

A.4.1. **Some intuition when $J = 3$.** To see how the treatment of general mixtures differs from the $J = 2$ case, consider the case $J = 3$. Instead of Equation (A.2), we now have

$$M(t|x) = \lambda_1 e^{tm_1(x)} M_1(t) + \lambda_2 e^{tm_2(x)} M_2(t) + \lambda_3 e^{tm_3(x)} M_3(t), \quad \lambda_1 + \lambda_2 + \lambda_3 = 1. \tag{A.1}$$

Wlog, suppose $m_1(x) > m_2(x) > m_3(x)$ at a point x in $\mathbb{R}^k$. Take a point x' in the neighborhood of x. The method used in the proof of Lemma 3.1 to identify the $J = 2$ model still works for the slopes of m_1 and m_3. Indeed, consider for x and a point in its neighborhood x',

$$\frac{1}{t}\log\left(\frac{M(t|x')}{M(t|x)}\right) = \frac{1}{t}\log\left(\frac{\lambda_1 e^{tm_1(x')} M_1(t) + \lambda_2 e^{tm_2(x')} M_2(t) + \lambda_3 e^{tm_3(x')} M_3(t)}{\lambda_1 e^{tm_1(x)} M_1(t) + \lambda_2 e^{tm_2(x)} M_2(t) + \lambda_3 e^{tm_3(x))} M_3(t)}\right)$$

$$= m_1(x') - m_1(x) + \frac{1}{t}\log\left(\frac{1 + \frac{\lambda_2}{\lambda_1} e^{t[m_2(x') - m_1(x')]} \frac{M_2(t)}{M_1(t)} + \frac{\lambda_3}{\lambda_1} e^{t[m_3(x') - m_1(x')]} \frac{M_3(t)}{M_1(t)}}{1 + \frac{\lambda_2}{\lambda_1} e^{t[m_2(x) - m_1(x)]} \frac{M_2(t)}{M_1(t)} + \frac{\lambda_3}{\lambda_1} e^{t[m_3(x) - m_1(x)]} \frac{M_3(t)}{M_1(t)}}\right). \tag{A.2}$$

Suppose the ratios $M_2(t)/M_1(t)$ and $M_3(t)/M_1(t)$ do not explode exponentially, m_1, m_2 and m_3 are continuous and x' is close enough to x so that $m_2(x') - m_1(x')$ and $m_3(x') - m_1(x')$ are negative. Then as t grows to infinity, the above expression approaches the slope $m_1(x') - m_1(x)$. Similarly, by taking the limit $t \to -\infty$, the slope of m_3 is identified. This argument, however, leaves the slope of the middle term m_2 undetermined. And in the general case of $J \geq 3$, $J - 2$ slopes remain to be determined. The approach in Lemma 3.1 does fall short of achieving its goal when applied to models with $J \geq 3$.

It is, however, possible to identify the slope of m_2 by proceeding as follows. Suppose, evaluated at x, the regression functions satisfy the inequality $m_1(x) > m_2(x) > m_3(x)$. Pick a point y in a neighborhood of x. Multiply (A.1) by $e^{-t[m_1(x) - m_1(y)]}$ to obtain

$$\begin{aligned} e^{-t[m_1(x) - m_1(y)]} M(t|x) =& \lambda_1 e^{tm_1(y)} M_1(t) \\ &+ \lambda_2 e^{t\{m_2(x) - [m_1(x) - m_1(y)]\}} M_2(t) + \lambda_3 e^{t\{m_3(x) - [m_1(x) - m_1(y)]\}} M_3(t). \end{aligned} \tag{A.3}$$

This purges x out of the first term on the right hand side. Note $[m_1(x) - m_1(y)]$ is identified by the above argument. Therefore the left hand side of the above equation is known.

The above step enables us to eliminate all unknown parameters associated with the first mixture component. To see this, suppose $m_j, j = 1, 2, 3$ are differentiable in at least one of first $k*$ elements of $x = (x^1, x^2, ..., x^k)$. In what follows we assume that it is differentiable in the first element x^1 without loss of generality. As before, we assume that this is a prior knowledge. We often write D_x instead of

D_{x^1} to lighten notations. Differentiating both sides of (A.3) with respect to (wrt) x^1 and rearranging,

$$D_x\left[e^{-t[m_1(x)-m_1(y)]}M(t|x)\right] = t\lambda_2[D_xm_2(x)-D_xm_1(x)]e^{t\{m_2(x)-[m_1(x)-m_1(y)]\}}M_2(t) + t\lambda_3[D_xm_3(x)-D_xm_1(x)]e^{t\{m_3(x)-[m_1(x)-m_1(y)]\}}M_3(t). \tag{A.4}$$

Note that applying D_x eliminates the unknown function $M_1(t)$ from of the right hand side of (A.4). We now have

$$\frac{\partial}{\partial t}\log\left(D_x\left[e^{-t[m_1(x)-m_1(y)]}M(t|x)\right]\right) =: \frac{A_1}{A_2},$$

where

$$\begin{aligned}
A_1 =& \frac{1}{t} + \{m_2(x)-[m_1(x)-m_1(y)]\} + \frac{\frac{\partial}{\partial t}M_2(t)}{M_2(t)} \\
&+ \frac{\lambda_3}{\lambda_2}\frac{[D_xm_3(x)-D_xm_1(x)]}{[D_xm_2(x)-D_xm_1(x)]}\left(\frac{1}{t}+\{m_3(x)-[m_1(x)-m_1(y)]\}\right)e^{t[m_3(x)-m_2(x)]}\frac{M_3(t)}{M_2(t)} \\
&+ \frac{\lambda_3}{\lambda_2}\frac{[D_xm_3(x)-D_xm_1(x)]}{[D_xm_2(x)-D_xm_1(x)]}e^{t[m_3(x)-m_2(x)]}\frac{\frac{\partial}{\partial t}M_3(t)}{M_2(t)}, \\
A_2 =& 1 + \frac{\lambda_3}{\lambda_2}\frac{[D_xm_3(x)-D_xm_1(x)]}{[D_xm_2(x)-D_xm_1(x)]}e^{t[m_3(x)-m_2(x)]}\frac{M_3(t)}{M_2(t)}.
\end{aligned}$$

Note that the factor $D_xm_2(x)-D_xm_1(x)$ is non-zero if the two regression functions are not parallel at x. If $\frac{M_3}{M_2}$ and $\frac{D_xM_3}{M_2}$ do not explode exponentially, all the terms above except for the second and third terms of A_1 and the first term of A_2 converge to zero as $t\to\infty$. It follows that

$$\lim_{t\to\infty}\left\{\frac{\partial}{\partial t}\log\left(D_x\left[e^{-t[m_1(x)-m_1(y)]}M(t|x)\right]\right)\right\} = \{m_2(x)-[m_1(x)-m_1(y)]\} + \lim_{t\to\infty}\frac{\frac{\partial}{\partial t}M_2(t)}{M_2(t)}, \tag{A.5}$$

The only unknown component in the above equation is $\lim_{t\to\infty}\frac{\frac{\partial}{\partial t}M_2(t)}{M_2(t)}$, which we assume is well defined here to simplify the exposition. But this term can be differenced out: indeed, repeat the above argument with replacing $x\in\mathbb{R}^k$ with a point $z\in\mathbb{R}^k$ close to x that $m_1(z)>m_2(z)>m_3(z)$. This yields

$$\lim_{t\to\infty}\left\{\frac{\partial}{\partial t}\log\left(D_z\left[e^{-t[m_1(z)-m_1(y)]}M(t|z)\right]\right)\right\} = \{m_2(z)-[m_1(z)-m_1(y)]\} + \lim_{t\to\infty}\frac{\frac{\partial}{\partial t}M_2(t)}{M_2(t)}.$$

The slope of m_2 is

$$m_2(x)-m_2(z) = \lim_{t\to\infty}\frac{\partial}{\partial t}\log\left(\frac{D_x\left[e^{-t[m_1(x)-m_1(y)]}M(t|x)\right]}{D_z\left[e^{-t[m_1(z)-m_1(y)]}M(t|z)\right]}\right) + (m_1(x)-m_1(z)).$$

The term $m_1(x)-m_1(z)$ on the right hand side is identified as noted earlier. Thus the equation above shows the identifiability of the slope of m_2. Note that in the main identification argument we will not assume that $\lim_{t\to\infty}\frac{\partial}{\partial t}M_2(t)/M_2(t)$ exists.

We have already noted that the identifiability of the slope of m_3 basically follows from Lemma 3.1. It is nevertheless instructive to present an alternative way to identify it by carrying on the foregoing analysis one step further. This will illustrate the basic idea behind our general identification theory for $J \in \mathbb{N}$. Let us return to Equation (A.4), changing the notation and writing x_a for x, x_b for y. We will also write x_c for z. The first step is to purge x_a from the first term on the right hand side, as we did in Equation (A.3), as follows:

$$\frac{e^{-t[\Delta_{ab}m_2-\Delta_{ab}m_1]}}{t[D_{x_a}m_2(x_a)-D_{x_a}m_1(x_a)]}D_{x_a}\left[e^{-t\Delta_{ab}m_1}M(t|x_a)\right]$$
$$=\lambda_2 e^{t\{m_2(x_b)\}}M_2(t)+\lambda_3\frac{D_{x_a}m_3(x_a)-D_{x_a}m_1(x_a)}{D_{x_a}m_2(x_a)-D_{x_a}m_1(x_a)}e^{t\{m_3(x_a)-\Delta_{ab}m_2\}}M_3(t),$$

which yields

$$D_{x_a}\left[\frac{e^{-t[\Delta_{ab}m_2-\Delta_{ab}m_1]}}{t[D_{x_a}m_2(x_a)-D_{x_a}m_1(x_a)]}D_{x_a}\left[e^{-t\Delta_{ab}m_1}M(t|x_a)\right]\right] \tag{A.6}$$
$$=\lambda_3\left\{[D_{x_a}+t(D_{x_a}m_3(x_a)-D_{x_a}m_2(x_a))]\frac{D_{x_a}m_3(x_a)-D_{x_a}m_1(x_a)}{D_{x_a}m_2(x_a)-D_{x_a}m_1(x_a)}\right\}e^{t\{m_3(x_a)-\Delta_{ab}m_2\}}M_3(t).$$

Notice that this eliminates the unknown moment generating function $M_2(t)$. Differentiating the above expression with respect to t and following the line of argument presented above, the slope of m_3 is given by

$$\Delta_{ac}m_3=\lim_{t\to\infty}\frac{\partial}{\partial t}\log\left[\left(\frac{\frac{e^{-t[\Delta_{ab}m_2-\Delta_{ab}m_1]}}{t[D_{x_a}m_2(x_a)-D_{x_a}m_1(x_a)]}D_{x_a}\left[e^{-t\Delta_{ab}m_1}M(t|x_a)\right]}{\frac{e^{-t[\Delta_{cb}m_2-\Delta_{cb}m_1]}}{t[D_{x_c}m_2(x_c)-D_{x_c}m_1(x_c)]}D_{x_c}\left[e^{-t\Delta_{cb}m_1}M(t|x_c)\right]}\right)\right]+\Delta_{ac}m_2.$$

A.4.2. **Arbitrary J.** We consider the neighborhood $N^1(x_a,\delta')$ where $\delta'\leq\delta$ and where for all $x\in N^1(x_a,\delta')$, $m_{k,l}(x)\leq\epsilon-\Delta$ if $k>l$ and $(D_{x^1}m_l)_{l=1..J}$ take J distinct values. We start with proving the following result, where $S(x_a)$ is defined in (A.17).

Lemma A.1. *$\forall x_a\in S(x_a)$,*

$$Q_k(x_a,t)=\sum_{j=k}^{J}\lambda_j R_k^j(t,x_a)e^{t[m_j(x_a)-\Delta_{ab}m_{k-1}]}M_j(t),\quad k=2,3,...,J. \tag{A.7}$$

Proof of Lemma A.1. Start from $k=2$. Then

$$Q_2(x_a,t)=\frac{\partial}{\partial x_a^1}[e^{-t\Delta_{ab}m_1}M(t|x_a)]$$
$$=\sum_{j=2}^{J}\lambda_j(tD_{x_a}[m_j(x_a)-m_1(x_a)])e^{t[m_j(x_a)-\Delta_{ab}m_1]}M_j(t).$$

$$= \sum_{j=2}^{J} \lambda_j R_2^j(t, x_a) e^{t[m_j(x_a) - \Delta_{ab} m_1]} M_j(t),$$

yielding the desired representation for the case of $k = 2$. Then $R_2(t, x_a) = R_2^2(t, x_a) = tD_{x_a}[m_2(x_a) - m_1(x_a)]$. Looking at $k = 3$,

$$\begin{aligned}
Q_3(x_a, t) &= A(x_a, x_b, t, 2) Q_2(x_a, t) \\
&= \frac{\partial}{\partial x_a^1} \left[\frac{e^{-t[\Delta_{ab} m_2 - \Delta_{ab} m_1]}}{R_2(t, x_a)} Q_2(x_a, t) \right] \\
&= \sum_{j=3}^{J} \lambda_j \left\{ D_{x_a} \frac{R_2^j(t, x_a)}{R_2(t, x_a)} + t \frac{R_2^j(t, x_a)}{R_2(t, x_a)} D_{x_a}[m_j(x_a) - m_2(x_a)] \right\} e^{t[m_j(x_a) - \Delta_{ab} m_2]} M_j(t) \\
&= \sum_{j=3}^{J} \lambda_j R_3^j(t, x_a) e^{t[m_j(x_a) - \Delta_{ab} m_2]} M_j(t),
\end{aligned}$$

and the $j = 2$ term in the summation drops out. Note that the above step requires that $R_2(t, x_a)$ is non-zero: this issue is discussed in Remark A.1. The fact that $Q_k(x_a, t), k = 4, ..., J$ has a representation as in (A.7) can be shown by induction: suppose (A.7) holds for $k = h$, that is

$$Q_h(x_a, t) = \sum_{j=h}^{J} \lambda_j R_h^j(t, x_a) e^{t[m_j(x_a) - \Delta_{ab} m_{h-1}]} M_j(t).$$

Using the definition of the operator $A(x_a, x_b, t, h)$ in (5.3) and of the R_k^j functions, we obtain

$$\begin{aligned}
Q_{h+1}(x_a, t) &= A(x_a, x_b, t, h) Q_h(x_a, t) \\
&= \sum_{j=h}^{J} \lambda_j A(x_a, x_b, t, h) \left(R_h^j(t, x_a) e^{t[m_j(x_a) - \Delta_{ab} m_{h-1}]} \right) M_j(t) \\
&= D_{x_a} \lambda_h e^{-t m_h(x_b)} M_h(t) + D_{x_a} \sum_{j=h+1}^{J} \lambda_j \frac{R_h^j(t, x_a)}{R_h^h(t, x_a)} e^{t[m_j(x_a) - \Delta_{ab} m_h]} M_j(t) \\
&= \sum_{j=h+1}^{J} \lambda_j \left\{ D_{x_a} \left(\frac{R_h^j(t, x_a)}{R_h^h(t, x_a)} \right) + t \frac{R_h^j(t, x_a)}{R_h^h(t, x_a)} D_{x_a}[m_j(x_a) - m_h(x_a)] \right\} e^{t[m_j(x_a) - \Delta_{ab} m_h]} M_j(t) \\
&= \sum_{j=h+1}^{J} \lambda_j R_{h+1}^j(t, x_a) e^{t[m_j(x_a) - \Delta_{ab} m_h]} M_j(t),
\end{aligned}$$

which is the desired result. □

The two following propositions, Propositions A.1 and A.2, are useful in proving Lemma 5.1. In what follows, $\deg_t(f)$ and $\mathrm{lc}_t(f)$ denote the degree and the leading coefficients of a polynomial $f(t)$

with respect to t. We also sometimes write

$$R_k^j := R_k^j(t,x)$$

$$m_{k,l} := m_k(x) - m_l(x).$$

as short hand.

Proposition A.1. *Suppose $x \in N^1(x_a, \delta')$. Then $R_k(t,x)$ is a rational function of t for sufficiently large t and takes the following form:*

$$R_k(t,x) = \frac{P_k(t,x)}{P_{k-1}(t,x)^2}$$

where $P_k(t,x), k \geq 3$ are polynomials in t such that

$$\deg_t(P_k(t,x)) = 2^{k-2} - 1$$

and

$$\mathrm{lc}_t(P_k(t,x)) = (\Pi_{g=1}^{k-1} D_x(m_k(x) - m_g(x)))\Pi_{j=2}^{k-1}\{(\Pi_{h=1}^{j-1} D_x(m_j(x) - m_h(x)))^{2^{k-j-1}}\}.$$

Proof of Proposition A.1. The recursive formula in Lemma A.1 is

$$R_{k+1}^j = D_{x^1}\left(\frac{R_k^j}{R_k^k}\right) + t\frac{R_k^j}{R_k^k} D_{x^1} m_{j,k}, j = 1, ..., J \tag{A.8}$$

with initial conditions

$$R_2^j = tD_{x^1} m_{j,1}, j = 1, ..., J. \tag{A.9}$$

For $k = 2$,

$$\begin{aligned}
R_3^j &= D_{x^1}\left(\frac{R_2^j}{R_2^2}\right) + t\frac{R_2^j}{R_2^2} D_{x^1} m_{j,2} \\
&= D_{x^1}\left(\frac{D_{x^1} m_{j,1}}{D_{x^1} m_{2,1}}\right) + t\frac{D_{x^1} m_{j1}}{D_{x^1} m_{2,1}} D_{x^1} m_{j,2} \\
&= \frac{D_{x^1}^2 m_{j,1} D_{x^1} m_{2,1} - D_{x^1}^2 m_{2,1} D_{x^1} m_{j,1} + tD_{x^1} m_{j,1} D_{x^1} m_{j,2} D_{x^1} m_{2,1}}{(D_{x^1} m_{2,1})^2} \\
&= \frac{P_3^j}{(D_{x^1} m_{2,1})^2}
\end{aligned}$$

where

$$P_3^j = D_{x^1}^2 m_{j,1} D_{x^1} m_{2,1} - D_{x^1}^2 m_{2,1} D_{x^1} m_{j,1} + tD_{x^1} m_{j,1} D_{x^1} m_{j,2} D_{x^1} m_{2,1}.$$

Note that P_3^j depends on x (where m's are evaluated) and t, so it can be interpreted as shorthand for $P_3^j(x,t)$. Then

$$\begin{aligned} R_4^j &= D_{x^1}\left(\frac{R_3^j}{R_3^3}\right) + t\frac{R_3^j}{R_3^3}D_{x^1}m_{j,3} \\ &= D_{x^1}\left(\frac{P_3^j}{P_3^3}\right) + t\frac{P_3^j}{P_3^3}D_{x^1}m_{j,3} \\ &= \frac{D_{x^1}P_3^jP_3^3 - P_3^jD_{x^1}P_3^3 + tP_3^jP_3^3D_{x^1}m_{j,3}}{(P_3^3)^2} \\ &= \frac{P_4^j}{(P_3^3)^2} \end{aligned}$$

where

$$P_4^j = D_{x^1}P_3^jP_3^3 - P_3^jD_{x^1}P_3^3 + tP_3^jP_3^3D_{x^1}m_{j,3}.$$

Note that $P_3^3 \neq 0$ at least for large t, therefore the above representation of R_4^j is valid. From here we can argue by induction. Suppose $P_{h-1}^{h-1} \neq 0$, which we justify in Remark A.1. Assume as well that for $k = h$, R_h^j can be written as

$$R_h^j = \frac{P_h^j}{(P_{h-1}^{h-1})^2}, \tag{A.10}$$

where P_h^j and P_{h-1}^j, $j = 1, ..., J$ satisfy the following relationship

$$P_h^j = D_{x^1}P_{h-1}^jP_{h-1}^{h-1} - P_{h-1}^jD_{x^1}P_{h-1}^{h-1} + tP_{h-1}^jP_{h-1}^{h-1}D_{x^1}m_{j,h-1}. \tag{A.11}$$

Then as in the case of $h = 4$ above,

$$\begin{aligned} R_{h+1}^j &= D_{x^1}\left(\frac{R_h^j}{R_h^h}\right) + t\frac{R_h^j}{R_h^h}D_{x^1}m_{j,h} \\ &= D_{x^1}\left(\frac{P_h^j}{P_h^h}\right) + t\frac{P_h^j}{P_h^h}D_{x^1}m_{j,h} \\ &= \frac{D_{x^1}P_h^jP_h^h - P_h^jD_{x^1}P_h^h + tP_h^jP_h^hD_{x^1}m_{j,h}}{(P_h^h)^2} \\ &= \frac{P_{h+1}^j}{(P_h^h)^2} \end{aligned}$$

with

$$P_{h+1}^j = D_{x^1}P_h^jP_h^h - P_h^jD_{x^1}P_h^h + tP_h^jP_h^hD_{x^1}m_{j,h},$$

i.e., if (A.10) and (A.11) hold for $k = h$, they also hold for $k = h + 1$. In short, the original system of equations (A.8) and (A.9) that determine R_k^j can be rewritten in terms of P_k^js as follows:

$$P_1^j = 1, \quad P_{h+1}^j = D_{x^1} P_h^j P_h^h - P_h^j D_{x^1} P_h^h + t P_h^j P_h^h D_{x^1} m_{j,h}, \tag{A.12}$$

$$R_k^j = \frac{P_k^j}{(P_{k-1}^{k-1})^2}, \qquad 1 \le k, j \le J. \tag{A.13}$$

The fact that $P_1^j = 1, j = 1, ...J$ are appropriate initial conditions can be easily verified. In particular, (A.12) implies that

$$P_{h+1}^{h+1} = D_{x^1} P_h^{h+1} P_h^h - P_h^{h+1} D_{x^1} P_h^h + t P_h^{h+1} P_h^h D_{x^1} m_{h+1,h}. \tag{A.14}$$

Note that (A.14) with initial values $P_1^1 = P_1^2$ recursively generates expressions of $P_k(\cdot,\cdot) = P_k^j(\cdot,\cdot), k = 2, ..., J, j = k, ..., J$ that have some useful properties including

(Replacement Property of P_h^j): $P_h^j, j = h+1, ..., J$ are obtained by replacing m_h in the expression for P_h^h with $m_j, j = 1, ..., J$.

To see this, first note that $P_2^j = t D_{x^1} m_{j,1}, j = 2, ..., J$ according to (A.12), therefore this claim applies to the case of $k = 2$. But (A.12) also shows that if the claim applies to $k = h$, it holds for $k = h + 1$ as well. The property holds for all k by induction.

Noting this property, it is easy to see that $P_k(\cdot,\cdot) = P_k^k(\cdot,\cdot), k = 2, ..., J$ are polynomials in t where their coefficients are functions of derivatives of m's. First, it trivially holds for $k = 2$ since $P_2^j = t D_{x^1} m_{j,1}, j = 2, ..., J$. Now, suppose the claim holds for $k = h$. Then by (A.14) P_{h+1}^{h+1} is a polynomial with the stated property, and by the replacement property, so are $P_{h+1}^j, j = h+2, ..., J$. That is, the claim holds for $k = h + 1$. By induction, the claim holds for $k = 2, ..., J$. In particular, we now know that $P_k = P_k^k, k = 3, ..., J$ are polynomials in t, as claimed in the Proposition.

It remains to verify the formulae for $\deg_t(P_k)$ and $\mathrm{lc}_t(P_k)$ given in the Proposition. Start with $k = 3$. It implies that

$$P_3^3 = D_{x^1}^2 m_{3,1} D_{x^1} m_{2,1} - D_{x^1}^2 m_{2,1} D_{x^1} m_{3,1} + t D_{x^1} m_{3,1} D_{x^1} m_{3,2} D_{x^1} m_{2,1},$$

therefore $\deg_t P_3 = 1$ and $\mathrm{lc}_t(P_3) = D_{x^1} m_{3,1} D_{x^1} m_{3,2} D_{x^1} m_{2,1}$, which are certainly consistent with the proposition. Now suppose the Proposition holds for $k = l$: P_l is a polynomial with $\deg_t(P_l) = 2^{l-2} - 1$ and $\mathrm{lc}_t(P_l(t,x)) = (\Pi_{g=1}^{l-1} D_{x^1} m_{l,g}) \Pi_{j=2}^{l-1} \{(\Pi_{h=1}^{j-1} D_{x^1} m_{j,h})^{2^{l-j-1}}\}$.

Since $x \in N^1(x_a, \delta')$, $(D_{x^1} m_l)_{l=1..J}$ take J distinct values, and $\mathrm{lc}_t(P_l(t,x)) \neq 0$. Also, the above observation that P_l^l and P_l^j are identical except for the replacement of m_l with m_j, for all $j \ge l$ implies

that

$$\deg_t(P_l^l) = \deg_t(P_l^j),\ \forall j \geq l, \tag{A.15}$$

and

$$\begin{aligned}\mathrm{lc}_t(P_l^{l+1}(t,x)) &= (\Pi_{g=1}^{l-1} D_{x^1} m_{l+1,g})\Pi_{j=2}^{l-1}\{(\Pi_{h=1}^{j-1} D_{x^1} m_{j,h})^{2^{l-j-1}}\} \\ &\neq 0.\end{aligned}$$

Using the recursion formula (A.14) with $h = l$ and noting that $\deg_t(D_{x^1}P_h^{h+1}P_h^h - P_h^{h+1}D_{x^1}P_h^h) \leq \deg_t(P_l^{l+1}P_l^l)$, we have

$$\deg_t(P_{l+1}^{l+1}) = 2\deg_t(P_l^l) + 1 \tag{A.16}$$

and

$$\begin{aligned}\mathrm{lc}_t(P_{l+1}^{l+1}) &= \mathrm{lc}_t(P_l^{l+1})\mathrm{lc}_t(P_l^l) D_{x^1} m_{l+1,l} \\ &= (\Pi_{g=1}^{l} D_{x^1} m_{l,g})\Pi_{j=2}^{l}\{(\Pi_{h=1}^{j-1} D_{x^1} m_{j,h})^{2^{l-j}}\}.\end{aligned}$$

Moreover, solving the difference equation (A.16) under the initial condition $\deg_t(P_3^3) = 1$,

$$\begin{aligned}\deg_t(P_k^k) &= \sum_{j=0}^{k-4} 2^j + 2^{k-3} \\ &= 2^{k-3} - 1 + 2^{k-3} \\ &= 2^{k-2} - 1.\end{aligned}$$

□

Remark A.1. Since $P_k = P_k^k, k = 1, ..., J$ are polynomials, they are nonzero for sufficiently large t. This justifies dividing by P_k in the above proof. Thus by Proposition A.1, $R_k \neq 0$ for sufficiently large t for $k = 2, 3, ..., J$ and the set

$$S(x_a) = \{t | R_k(t, x_a) \neq 0 \text{ for all } k\} \tag{A.17}$$

is non-empty; it includes (for example) the set $[c, \infty)$ for some constant c. This is important in applying Lemma A.1.

Proposition A.2.

$$\lim_{t\to\infty} \frac{\partial}{\partial t} \log R_k(x,t) = 0$$

for all $t \in \mathbb{R}$ *and* $x \in N^1(x_a, \delta')$.

Proof of Proposition A.2. By the expression of $R_k(x,t)$ given in Proposition A.1,

$$\begin{aligned}\lim_{t\to\infty}\frac{\partial}{\partial t}\log R_k(x,t) &= \lim_{t\to\infty}\frac{\partial}{\partial t}\log\frac{P_k(t,x)}{P_{k-1}(t,x)^2}\\ &= \lim_{t\to\infty}\frac{\partial}{\partial t}\log P_k(t,x) - 2\lim_{t\to\infty}\frac{\partial}{\partial t}\log P_{k-1}(t,x).\end{aligned}$$

By Proposition A.1, $R_k(x,t)$, $P_k(x,t)$ and $P_{k-1}(x,t)$ are well defined for large t, and since $P_k(t,x)$ and $P_{k-1}(t,x)$ are polynomials in t with finite degree, so are the above limits which are both zero. □

Proof of Lemma 5.1. By Lemma A.1 and Proposition A.1,

$$Q_k(x_a,t) = \sum_{j=k}^{J}\lambda_j R_k^j(t,x_a)e^{t[m_j(x_a)-\Delta_{ab}m_{k-1}]}M_j(t), \quad k=2,3,...,J, \tag{A.18}$$

holds for sufficiently large t. Then

$$\begin{aligned}\frac{\partial}{\partial t}Q_k(x_a,t) &= \sum_{j=k}^{J}\lambda_j\left(\frac{\partial}{\partial t}R_k^j(t,x_a) + [m_j(x_a)-\Delta_{ab}m_{k-1}]R_k^j(t,x_a)\right)e^{t[m_j(x_a)-\Delta_{ab}m_{k-1}]}M_j(t)\\ &\quad + \sum_{j=k}^{J}\lambda_j R_k^j(t,x_a)e^{t[m_j(x_a)-\Delta_{ab}m_{k-1}]}\frac{\partial}{\partial t}M_j(t).\end{aligned}$$

and, for $k \le J$,

$$\begin{aligned}\frac{\partial}{\partial t}\log(Q_k(x_a,t)) &= \frac{\frac{\partial}{\partial t}Q_k(x_a,t)}{Q_k(x_a,t)}\\ &= \frac{\frac{\frac{\partial}{\partial t}R_k(t,x_a)}{R_k(t,x_a)} + m_k(x_a) - \Delta_{ab}m_{k-1} + \frac{\frac{\partial}{\partial t}M_k(t)}{M_k(t)}}{1+\sum_{j=k+1}^{J}\frac{\lambda_j}{\lambda_k}\frac{R_k^j(t,x_a)}{R_k(t,x_a)}e^{tm_{j,k}(x_a)}\frac{M_j(t)}{M_k(t)}}\\ &\quad + \frac{\sum_{h=k+1}^{J}\left[(m_h(x_a)-\Delta_{ab}m_{k-1})\frac{\lambda_h}{\lambda_k}\frac{R_k^h(t,x_a)}{R_k(t,x_a)}e^{tm_{h,k}(x_a)}\frac{M_h(t)}{M_k(t)}\right]}{1+\sum_{j=k+1}^{J}\frac{\lambda_j}{\lambda_k}\frac{R_k^j(t,x_a)}{R_k(t,x_a)}e^{tm_{j,k}(x_a)}\frac{M_j(t)}{M_k(t)}}\\ &\quad + \frac{\sum_{h=k+1}^{J}\left[\frac{\lambda_h}{\lambda_k}\frac{\frac{\partial}{\partial t}R_k^h(t,x_a)}{R_k(t,x_a)}e^{tm_{h,k}(x_a)}\frac{M_h(t)}{M_k(t)} + \frac{\lambda_h}{\lambda_k}\frac{R_k^h(t,x_a)}{R_k(t,x_a)}e^{tm_{h,k}(x_a)}\frac{\frac{\partial}{\partial t}M_h(t)}{M_k(t)}\right]}{1+\sum_{j=k+1}^{J}\frac{\lambda_j}{\lambda_k}\frac{R_k^j(t,x_a)}{R_k(t,x_a)}e^{tm_{j,k}(x_a)}\frac{M_j(t)}{M_k(t)}}.\end{aligned} \tag{A.19}$$

Using (A.13), for all $h > k$,

$$\frac{R_k^h(t,x_a)}{R_k(t,x_a)} = \frac{P_k^h(t,x_a)/(P_{k-1}^{k-1}(t,x_a))^2}{P_k^k(t,x_a)(P_{k-1}^{k-1}(t,x_a))^2} = \frac{P_k^h(t,x_a)}{P_k^k(t,x_a)}.$$

As noted in the proof of Proposition A.1, both $P_k^h(t,x_a)$ and $P_k^k(t,x_a)$ are polynomials in t, $P_k^k(t,x_a) \neq 0$ for sufficiently large t, and their degrees are equal. Hence their ratio goes to a constant as t goes to

infinity:

$$\lim_{t\to\infty}\frac{R_k^h(t,x_a)}{R_k(t,x_a)}=c_{h,k,x_a}.$$

Thus we also have, as in Proposition A.2,

$$\lim_{t\to\infty}\frac{\frac{\partial}{\partial t}R_k^h(t,x_a)}{R_k(t,x_a)}=\lim_{t\to\infty}\frac{\frac{\partial}{\partial t}R_k^h(t,x_a)}{R_k^h(t,x_a)}\frac{R_k^h(t,x_a)}{R_k(t,x_a)}=0.$$

By Assumption 5.2 (iv), since $m_{h,k}(x_a)<-\Delta$, the second and third lines of (A.19) converge to zero as t goes to $+\infty$. Recall that δ' is such that for all $x\in N^1(x_a,\delta')$, $m_{h,k}(x)\leq\epsilon-\Delta$. Note that (A.19) holds for any $x_b\in\mathbb{R}^k$, thus we take $x_b\in N^1(x_a,\delta')$. Note too that we can write (A.19) replacing x_a with x_b, the $\Delta_{ab}m_h$ terms being equal to 0. Then

$$\begin{aligned}\frac{\partial}{\partial t}\log\left(\frac{Q_k(x_a,t)}{Q_k(x_b,t)}\right)&\underset{t\to+\infty}{=}\sum_{j=k+1}^{J}\frac{\lambda_j}{\lambda_k}\left(\frac{R_k^j(t,x_a)}{R_k(t,x_a)}e^{tm_{j,k}(x_a)}-\frac{R_k^j(t,x_b)}{R_k(t,x_b)}e^{tm_{j,k}(x_b)}\right)\frac{M_j(t)\frac{\partial}{\partial t}M_k(t)}{M_k(t)^2}\left(1+o(1)\right)\\&\quad+\Delta_{ab}m_k-\Delta_{ab}m_{k-1}+o(1)\\&\underset{t\to+\infty}{=}\left[\sum_{j=k+1}^{J}c_je^{(\epsilon-\Delta)t}\ \frac{M_j(t)\frac{\partial}{\partial t}M_k(t)}{M_k(t)^2}\right]O(1)+\Delta_{ab}m_k-\Delta_{ab}m_{k-1}+o(1)\\&\underset{t\to+\infty}{=}\Delta_{ab}m_k-\Delta_{ab}m_{k-1}+o(1)\end{aligned}$$

where in the second equality, c_j, $j=k+1,...,J$ are constant numbers and where the third equality holds by Assumption 5.2 (iv). □

Proof of Lemma 5.2. Lemma 5.1 together with the proof of identifiability of $\Delta_{ab}m_1$ allow, by induction, the identifiability of the slopes $\Delta_{ab}m_k$ for all $x_b\in N^1(x_a,\delta')$ and for all $k=1,...,J$:

$$\Delta_{ab}m_1=\lim_{t\to\infty}\frac{1}{t}\log\left(\frac{M(t|x_a)}{M(t|x_b)}\right),$$

$$\Delta_{ab}m_k=\sum_{j=2}^{k}\lim_{t\to\infty}\frac{\partial}{\partial t}\log\left(\frac{Q_k(x_a,t)}{Q_k(x_b,t)}\right)+\Delta_{ab}m_1.$$

Thus $\Xi(x_0,X)$ is identified. Following arguments in the proof of Lemma 3.3, since

$$\dot{M}(0|x_0)-\dot{M}(0|x)=\sum_{i=1}^{J}\lambda_i[(m_i(x_0)-m_i(x))-(m_J(x_0)-m_J(x))]+(m_J(x_0)-m_J(x)),$$

we can write

$$\begin{pmatrix}\dot{M}(0|x_0)-\dot{M}(0|x_1)\\\vdots\\\dot{M}(0|x_0)-\dot{M}(0|x_{J-1})\end{pmatrix}=\Xi(x_0,X)\begin{pmatrix}\lambda_1\\\vdots\\\lambda_{J-1}\end{pmatrix}+\begin{pmatrix}\Delta_{0,1}m_J\\\vdots\\\Delta_{0,J-1}m_J\end{pmatrix}.$$

By Assumption 5.3, the $(\lambda_j)_{j=1..J-1}$ are identified with the formula

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_{J-1} \end{pmatrix} = \Xi(x_0, X)^{-1} \left[\begin{pmatrix} \dot{M}(0|x_0) - \dot{M}(0|x_1) \\ \vdots \\ \dot{M}(0|x_0) - \dot{M}(0|x_{J-1}) \end{pmatrix} - \begin{pmatrix} \Delta_{0,1} m_J \\ \vdots \\ \Delta_{0,J-1} m_J \end{pmatrix} \right].$$

To identify $(m_j(x_0))_{j=1..J}$, we let

$$C(x) = \left\{ \ddot{M}(0|x_0) - \ddot{M}(0|x) + \sum_{j=1}^{J} \lambda_j [m_j(x_0) - m_j(x)]^2 \right\} /2$$

and one can show that

$$C(x_k) = \sum_{j=1}^{J} \lambda_j \; m_j(x_0) \; \Delta_{0,k} m_j.$$

This gives

$$\begin{pmatrix} C(x_1) \\ \vdots \\ C(x_{J-1}) \\ \dot{M}(0|x_0) \end{pmatrix} = B(x_0, X) \, diag(\lambda_1, ..., \lambda_J) \begin{pmatrix} m_1(x_0) \\ \vdots \\ m_J(x_0) \end{pmatrix},$$

where

$$B(x_0, X) = \begin{pmatrix} \Delta_{0,1} m_1 & \dots & \Delta_{0,1} m_J \\ \vdots & \ddots & \vdots \\ \Delta_{0,J-1} m_1 & \dots & \Delta_{0,J-1} m_J \\ 1 & \dots & 1 \end{pmatrix} \text{ is identified.}$$

Since $\det B(x_0, X) = \det \Xi(x_0, X)$, $B(x_0, X)$ is invertible. Moreover, since λ_j, $j = 1..J$ are assumed to be nonzero, i.e. J is known, we obtain the following identification result:

$$\begin{pmatrix} m_1(x_0) \\ \vdots \\ m_J(x_0) \end{pmatrix} = diag(\lambda_1^{-1}, ..., \lambda_J^{-1}) B(x_0, X)^{-1} \begin{pmatrix} C(x_1) \\ \vdots \\ C(x_{J-1}) \\ \dot{M}(0|x_0) \end{pmatrix}.$$

Note that this identifies $(m_j(x))_{j=1..J}$ for all $x \in N^1(x_a, \delta')$ using knowledge of the slopes. We now identify the distributions $(F_j(\cdot))_{j=1..J}$. For J generic points $(c_k)_{k=1..J} \in N^1(x_0, \delta')^J$, we have

$$\begin{pmatrix} M(t|c_1) \\ \vdots \\ M(t|c_J) \end{pmatrix} = D(t, c_1, ..., c_J) \; diag(\lambda_1, ..., \lambda_J) \begin{pmatrix} M_1(t) \\ \vdots \\ M_J(t) \end{pmatrix},$$

where $D(t, c_1, ..., c_J)$ is the matrix $(e^{tm_j(c_i)})_{1\leq i,j\leq J}$. Note that $D(t, c_1, ..., c_J)$ is identified. We prove below in Proposition A.3 that there exists a vector $X^{(J)} = (x_1^{(J)}, ..., x_{J-1}^{(J)}) \in N^1(x_a, \delta')^{J-1}$ of size $(J-1)$ such that the set $\mathcal{Z} = \left\{t \in \mathbb{R} | \det D(t, x_0, x_1^{(J)}, ..., x_{J-1}^{(J)}) = 0\right\}$ is finite. Hence, we can invert $D(t, x_0, x_1^{(J)}, ..., x_{J-1}^{(J)})$ for all $t \in \mathbb{R}\backslash\mathcal{Z}$, which identifies $(M_j(t))_{j=1..J}$ for all $t \in \mathbb{R}\backslash\mathcal{Z}$. By Assumption 5.3, $(M_j(t))_{j=1..J}$ have domain $(-\infty, +\infty)$ thus they are continuous on $\mathbb{R}$. For each M_j, there is a unique continuous extension on $\mathbb{R}$ of its restriction to $\mathbb{R}\backslash\mathcal{Z}$; the J functions are identified. By uniqueness of the Laplace transform, this leads to the identification of the distributions $(F_j(\cdot))_{j=1..J}$.

□

Proposition A.3. *There exists* $X^{(J)} = (x_1^{(J)}, ..., x_{J-1}^{(J)}) \in N^1(x_0, \delta')^{J-1}$ *such that*

$$\mathcal{Z} = \left\{t \in \mathbb{R} | \det D(t, x_0, x_1^{(J)}, ..., x_{J-1}^{(J)}) = 0\right\}$$

is a finite set.

Proof of Proposition A.3.

$$D(t, c_1, ..., c_J) = (e^{tm_j(c_i)})_{1\leq i,j\leq J}$$

Writing S_n the set of permutations of the first n natural numbers and $sign(\sigma)$ the signature of a permutation σ, we have $\det D(t, c_1, ..., c_J) = \sum_{\sigma\in S_J} sign(\sigma)\ e^{t\sum_{i=1}^J m_{\sigma(i)}(c_i)}$.

Step 1: We call $V(\sigma, c) = \sum_{i=1}^J m_{\sigma(i)}(c_i)$, where $c = (c_1, ..., c_J) \in N^1(x_0, \delta')$, and our goal is now to construct a vector $c^{(J)} = (c_J^{(J)}, ..., c_J^{(J)})$ such that there is a unique permutation maximizing $V(\cdot, c^{(J)})$.

We fix $c^{(1)} = (c_1^{(1)}, ..., c_J^{(1)}) \in N^1(x_0, \delta')^J$, $A_1 = \max_{\sigma\in S_J} V(\sigma, c^{(1)})$, $\Sigma_1 = \left\{\sigma \in S_J | V(\sigma, c^{(1)}) = A_1\right\}$ and $B_1 = \max_{\sigma\in S_J\backslash\Sigma_1} V(\sigma, c^{(1)})$. If B_1 exists, then $B_1 < A_1$. We consider a vector $c^{(2)}$ which differs from $c^{(1)}$ only in the first component. This first component is a point in $\mathbb{R}^k$ and we consider a variation only in its first covariate, with respect to which the $(m_j)_{j=1...J}$ are J times differentiable.

$$\forall\sigma \in S_J, V(\sigma, c^{(2)}) = V(\sigma, c^{(1)}) + m_{\sigma(1)}(c_1^{(2)}) - m_{\sigma(1)}(c_1^{(1)}).$$

For all $x \in N^1(x_0, \delta')$, $(D_{x^1}m_j(x))_{j=1..J}$ take distinct values: $\operatorname{argmax}_{s\in\{\sigma(1)|\sigma\in\Sigma_1\}} D_{x^1}m_s(c_1^{(1)})$ is a singleton set $\{s_1\}$. Hence, since the m_i functions are at least twice differentiable, they are continuously differentiable, we can choose $c_1^{(2)} \in N^1(x_0, \delta')$ close enough from $c_1^{(1)}$ so that

$$m_{s_1}(c_1^{(2)}) - m_{s_1}(c_1^{(1)}) = \max_{\sigma\in\Sigma_1} m_{\sigma(1)}(c_1^{(2)}) - m_{\sigma(1)}(c_1^{(1)}),$$

and if B_1 exists,

$$m_i(c_1^{(2)}) - m_i(c_1^{(1)}) < \frac{A_1 - B_1}{2}, \forall i \leq J.$$

Therefore, constructing $\Sigma_2 = \{\sigma \in \Sigma_1 | \sigma(1) = s_1\}$ ($\Sigma_2 \neq \emptyset$ by construction), $A_2 = \max_{\sigma \in S_J} V(\sigma, c^{(2)})$, and $B_2 = \max_{\sigma \in S_J \setminus \Sigma_2} V(\sigma, c^{(2)})$, we know that B_2 exists and $B_2 < A_2$. We repeat the same process with the second component of $c^{(2)}$ and construct s_2, Σ_3, $c^{(3)}$, A_3 and B_3, and then we repeat it with the third component of $c^{(3)}$ and so on, until $|\Sigma_i| = 1$ for some i. If this is not the case for some $i < J$, then constructing each of the elements until $i = J$, we have

$$\Sigma_J = \{\sigma \in \Sigma_1 | \sigma(1) = s_1, ..., \sigma(J-1) = s_{J-1}\},$$

implying $|\Sigma_J| = 1$. The vector and the permutation eventually obtained are called $c^{(J)}$ and σ_J and are such that

$$V(\sigma_J, c^{(J)}) = \max_{\sigma \in S_J} V(\sigma, c^{(J)}) \text{ and } \forall \sigma \neq \sigma_J, V(\sigma, c^{(J)}) < V(\sigma_J, c^{(J)}).$$

Step 2: Note that in the previous step, the last component of the vector c^1 did not change during the whole process: we impose $c_J^{(1)} = x_0$. The order of those components do not matter; the previous result holds for some $c^{(J)} = (x_0, x_1^{(J)}, ..., x_{J-1}^{(J)})$. Since $\det D(t, x_0, x_1^{(J)}, ..., x_{J-1}^{(J)}) = \sum_{\sigma \in S_J} sign(\sigma)\ e^{tV(\sigma, c^{(J)})}$, and $sign(\sigma) \in \{-1, 1\}$, $\det D(\cdot, x_0, x_1^{(J)}, ..., x_{J-1}^{(J)})$ is a linear combination of distinct exponential functions multiplied by nonzero coefficients; one such coefficient is that associated with $e^{tV(\sigma_J, c^{(J)})}$. This implies that $\det D(\cdot, x_0, x_1^{(J)}, ..., x_{J-1}^{(J)})$ has a finite number of zeros (see, e.g, Tossavainen (2006)).

□

Proof of Proposition 5.1.

$$Q_J = \lambda_J R_J(t, x_a) e^{t(m_J(x_a) - \Delta_{ab} m_{J-1})} M_J(t)$$

$$\Rightarrow Q_{J+1}(x_a, t) = \lambda_J \frac{\partial}{\partial x_a^1} \left[\frac{R_J(t, x_a)}{R_J(t, x_a)} e^{-t m_J(x_b)} M_J(t) \right] = 0 \text{ for all } t \in \mathbb{R}.$$

This proves that $J \geq \max\{j \geq 1 | \exists t_0 \in \mathbb{R}, Q_j(x_a, t_0) \neq 0\}$.

Reciprocally, if $j_0 \leq J$, then for some $t_0 \in \mathbb{R}, Q_{j_0}(x_a, t_0) \neq 0$. Indeed, $j_0 \leq J \Rightarrow \forall j_0 \leq k \leq J$, $\lambda_k > 0$ and we can write

$$Q_{j_0}(x_a, t) = \lambda_{j_0} R_{j_0}(t, x_a) M_{j_0}(t) e^{t m_{j_0}(x_a) - \Delta_{ab} m_{j_0 - 1}} \left(1 + \sum_{j=k_0+1}^{J} \frac{\lambda_j}{\lambda_{j_0}} \frac{R_{j_0}^j(t, x_a)}{R_{j_0}(t, x_a)} e^{t m_{j, j_0}(x_a)} \frac{M_j(t)}{M_{j_0}(t)} \right).$$

By the Proof of Lemma 5.1,

$$\lim_{t \to \infty} \frac{R_{j_0}^j(t, x_a)}{R_{j_0}(t, x_a)} = c_{j, j_0, x_a}.$$

By Assumption 5.2, $m_{j, j_0}(x_a) < -\Delta$: each term in the sum on the right hand side goes to 0 as t goes to ∞, implying that for large enough t, the term in parenthesis is strictly positive, that is, nonzero.

□

A.5. A model with "fixed effects" when $J \geq 2$

We discuss briefly how to extend the identification argument of the model with "fixed effects" to the case where the number of mixture components J can be greater than 2 and even unknown. We do not develop a general and rigorous treatment of identification but rather lay out a simple explanation of how the arguments in Sections 4 and 5 can be combined into a recursive proof of identification.

To that aim, we consider the case $J = 3$ as we did in the beginning of Section A.4.1. We explain how to identify this model using a recursive approach which we argue can be generalized to any J, with J unknown. Our goal is to identify $\lambda_j(y)$, $m_j(y)$, $F_j(\cdot)$ for $1 \leq j \leq J$ and J itself, for $y \in \mathbb{R}^k$. Our argument proceeds by choosing points x and z in $\mathbb{R}^k$ in a neighborhood of y. Instead of Equation (A.1), we now have

$$M(t|y) = \lambda_1(y)e^{tm_1(y)}M_1(t) + \lambda_2(y)e^{tm_2(y)}M_2(t) + \lambda_3(y)e^{tm_3(y)}M_3(t), \tag{A.1}$$

with $\lambda_1(y)+\lambda_2(y)+\lambda_3(y) = 1, y \in \mathbb{R}^k$. Take x in a neighborhood of y. In this section, we only consider the simpler case where $\lambda_j(x) < 1$, $\lambda_j(y) < 1$ and $\lambda_j(x) \neq \lambda_j(y)$ for all $j \leq J$. We assume as in Section 5 that $D_{x^1}m_j(y), j = 1, ..., J$ takes J distinct values in $\mathbb{R}$. Wlog, suppose $m_1(y) > m_2(y) > m_3(y)$.

The argument of the proof of Lemma 3.1 when $J = 2$ applies here to identify the slope of m_1 as long as x is close enough to y. See the discussion around Equation (A.2) in Section A.4.1 which can easily be adapted to $(\lambda_j(y))_{j\leq 3}$ varying with y. However as mentioned in this discussion, this method does not identify the slope of m_2 even when $(\lambda_j)_{j\leq J}$ is not allowed to vary with y.

It is nevertheless possible to identify the slope of m_2 by proceeding as follows. Multiply (A.1) by $e^{-t[m_1(x)-m_1(y)]}$ to obtain

$$\begin{aligned} e^{-t[m_1(x)-m_1(y)]}M(t|x) =& \lambda_1(x)e^{tm_1(y)}M_1(t) + \lambda_2(x)e^{t\{m_2(x)-[m_1(x)-m_1(y)]\}}M_2(t) \\ &+ \lambda_3(x)e^{t\{m_3(x)-[m_1(x)-m_1(y)]\}}M_3(t). \end{aligned}$$

Comparing with (A.3), this multiplication does not purge x out of the first term on the right hand side because $\lambda_1(x)$ is unknown. We thus use ideas from Section 4. Indeed, one can show using arguments in the discussion around Equation (A.2) in Section A.4.1 that $K_{+\infty}(x, y) = \frac{\lambda_1(x)}{\lambda_1(y)}$, where we recall that $K_{+\infty}(x, \underline{x})$ is defined in Section 4. Thus we multiply (A.1) with $e^{-t[m_1(x)-m_1(y)]}/K_{+\infty}(x, y)$ to obtain

$$e^{-t[m_1(x)-m_1(y)]}\frac{M(t|x)}{K_{+\infty}(x, y)} = \lambda_1(y)e^{tm_1(y)}M_1(t) + \lambda_1(y)\frac{\lambda_2(x)}{\lambda_1(x)}e^{t\{m_2(x)-[m_1(x)-m_1(y)]\}}M_2(t)$$

$$+ \lambda_1(y)\frac{\lambda_3(x)}{\lambda_1(x)}e^{t\{m_3(x)-[m_1(x)-m_1(y)]\}}M_3(t). \tag{A.2}$$

The left hand side of the above equation is known. The above step enables us to eliminate all unknown parameters associated with the first mixture component. Suppose as in Section 5 that $m_j, j = 1, 2, 3$ are differentiable in at least one of the $k*$ continuous elements of x. In what follows, we assume that it is differentiable in the first element x^1 without loss of generality and that this is prior knowledge. Let D_x denote the partial differentiation operator with respect to x^1 and $\dot{m}_1(x) := m_1(x) - m_1(y)$. Let $B^1(t,x) := D_x\left[e^{-t\dot{m}_1(x)}\frac{M(t|x)}{K_{+\infty}(x,y)}\right]$, $C_2^1(t,x) := D_x\frac{\lambda_2}{\lambda_1}(x) + t\frac{\lambda_2}{\lambda_1}(x)D_x(m_2 - m_1)(x)$ and $C_3^1(t,x) := D_x\frac{\lambda_3}{\lambda_1}(x) + t\frac{\lambda_3}{\lambda_1}(x)D_x(m_3 - m_1)(x)$. Differentiating both sides of (A.2) wrt x^1 gives

$$B^1(t,x) = \lambda_1(y)\left[C_2^1(t,x)e^{t\{m_2(x)-\dot{m}_1(x)\}}M_2(t) + C_3^1(t,x)e^{t\{m_3(x)-\dot{m}_1(x)\}}M_3(t)\right]. \tag{A.3}$$

Note that operating D_x eliminates the unknown function $M_1(t)$. We now have

$$\frac{\partial}{\partial t}\log B^1(t,x) = \frac{B_1^1(t,x)}{B_2^1(t,x)},$$

where

$$\begin{aligned}
B_1^1(t,x) &:= \frac{\frac{\partial}{\partial t}C_2^1(t,x)}{C_2^1(t,x)} + \{m_2(x) - \dot{m}_1(x)\} + \frac{\frac{\partial}{\partial t}M_2(t)}{M_2(t)} + \frac{C_3^1(t,x)}{C_2^1(t,x)}e^{t[m_3(x)-m_2(x)]}\frac{\frac{\partial}{\partial t}M_3(t)}{M_2(t)} \\
&\quad + \left[\frac{C_3^1(t,x)}{C_2^1(t,x)}(m_3(x) - \dot{m}_1(x)) + \frac{\frac{\partial}{\partial t}C_3^1(t,x)}{C_2^1(t,x)}\right]e^{t[m_3(x)-m_2(x)]}\frac{M_3(t)}{M_2(t)} \\
B_2^1(t,x) &:= 1 + \frac{C_3^1(t,x)}{C_2^1(t,x)}e^{t[m_3(x)-m_2(x)]}\frac{M_3(t)}{M_2(t)}.
\end{aligned}$$

We can then argue as in Section A.4.2. The ratios of the form $\frac{\frac{\partial}{\partial t}C_j^1(t,x)}{C_{j'}^1(t,x)}$ converge to 0 as $t \to \infty$, while the ratios of the form $\frac{C_j^1(t,x)}{C_{j'}^1(t,x)}$ are bounded as $t \to \infty$. Thus under conditions on the moment generating functions similar to those in Section 5, all terms above converge to zero as $t \to \infty$ except for the second and third terms of $B_1^1(t,x)$ and the first term of $B_2^1(t,x)$. This third term does not depend on x. Taking two distinct points x and z and using arguments similar to the proof of Lemma 5.1, one identifies the slope of m_2 with

$$m_2(x) - m_2(z) = \lim_{t\to\infty}\frac{\partial}{\partial t}\log\left(\frac{B^1(t,x)}{B^1(t,z)}\right) + (m_1(x) - m_1(z)).$$

Reproducing the reasoning of Section A.4.1, we could argue for identifiability of the slope of m_3 using Lemma 4.1 but we instead carry on the foregoing analysis one step further. This will illustrate the basic idea behind a generalization of the recursive argument for any $J \in \mathbb{N}$. Let us return to Equation

(A.3). Repeating the step that purges all dependence on x of the second mixture component requires knowledge of $\dot{m}_1(x)$, $\dot{m}_2(x)$, $\lambda_1(x)/\lambda_1(y)$, which are identified, and of $\lambda_2(x)$, which is not.

But identification of $\lambda_2(x)/\lambda_2(y)$ can be obtained as in proof of Lemma 4.1. Indeed, this proof uses the knowledge of the slope of the first mixture regression function to get rid of the dependance on this slope and to obtain the ratio of the first component mixing weights: the same idea can be applied to B^1 where all dependence on the first mixture component disappeared, using the knowledge of the slope of the second component regression function. One can show that

$$\frac{B^1(t,x)}{B^1(t,y)}e^{t[\dot{m}_1(x)-\dot{m}_2(x)]} = \frac{C_2^1(t,x)}{C_2^1(t,y)}\frac{1+\frac{C_3^1(t,x)}{C_2^1(t,x)}e^{t(m_3(x)-m_2(x))}\frac{M_3(t)}{M_2(t)}}{1+\frac{C_3^1(t,y)}{C_2^1(t,y)}e^{t(m_3(y)-m_2(y))}\frac{M_3(t)}{M_2(t)}} \underset{t\to+\infty}{\to} \frac{\lambda_2(x)}{\lambda_2(y)}\frac{\lambda_1(y)}{\lambda_1(x)}\frac{D_x(m_2-m_1)(x)}{D_y(m_2-m_1)(y)}, \tag{A.4}$$

which identifies $\frac{\lambda_2(x)}{\lambda_2(y)}$. Let $\tilde{C}_2^1(t,x) := \frac{\lambda_1(y)}{\lambda_2(y)}C_2^1(t,x)$. Note that it is identified as well. Then

$$e^{t(\dot{m}_1(x)-\dot{m}_2(x))}\frac{B^1(t,x)}{\tilde{C}_2^1(t,x)} = \lambda_2(y)M_2(t)e^{tm_2(y)} + \lambda_2(y)M_3(t)\frac{C_3^1(t,x)}{C_2^1(t,x)}e^{t(m_3(x)-\dot{m}_2(x))}$$

where the first element on the right hand side does not depend on x. We can take once more derivatives wrt x^1, which gives

$$\begin{aligned} B^2(t,x) &= D_x\left[e^{t(\dot{m}_1(x)-\dot{m}_2(x))}\frac{B^1(t,x)}{\tilde{C}_2^1(t,x)}\right] \\ &= \lambda_2(y)M_3(t)C_3^2(t,x)e^{t(m_3(x)-\dot{m}_2(x))}, \end{aligned}$$

where $C_3^2(t,x) = \frac{C_3^1(t,x)}{C_2^1(t,x)}tD_x(m_3-m_2)(x) + D_x\frac{C_3^1}{C_2^1}(t,x)$. Thus,

$$\frac{\partial}{\partial t}\log B^2(t,x) = \frac{\frac{\partial}{\partial t}C_3^2(t,x)}{C_3^2(t,x)} + \{m_3(x)-\dot{m}_2(x)\} + \frac{\frac{\partial}{\partial t}M_3(t)}{M_3(t)}.$$

A closer look at $C_3^2(t,x)$ shows that it is the sum of two ratios of polynomial functions of t, the first one with a numerator of a degree higher than the denominator by one and the second one being the ratio of two polynomials of the same degree. Thus one can show

$$C_3^2(t,x) \underset{t\to+\infty}{\sim} D_x(m_3-m_2)(x)\frac{\lambda_3(x)}{\lambda_2(x)}\frac{D_x(m_3-m_1)(x)}{D_x(m_2-m_1)(x)}\,t,$$

and $\frac{\partial}{\partial t}C_3^2(t,x)$ converges to a constant as $t\to\infty$. This gives, considering again two points x and z and assuming conditions on the moment generating functions similar to Assumption 5.2,

$$m_3(x)-m_3(z) = \lim_{t\to\infty}\frac{\partial}{\partial t}\log\left(\frac{B^2(t,x)}{B^2(t,z)}\right) + (m_2(x)-m_2(z)),$$

which identifies the slope of m_3.

For $J = 3$, we know $\lambda_3(.) = 1 - \lambda_1(.) - \lambda_2(.)$. However note that similarly to (A.4), looking at $\frac{B^2(t,x)}{B^2(t,y)}$ would alternatively identify the ratio $\frac{\lambda_3(x)}{\lambda_3(y)}$. This recursive argument can be generalized to any known J to identify sequentially the slopes of the m_j functions together with the ratios $\frac{\lambda_j(x)}{\lambda_j(y)}$, constructing first $(\tilde{C}_k^{j-1})_{j\leq k\leq J}$ then B^{j-1}.

This procedure also leads to identification of J when it is unknown. Indeed, $B^k(t,x)$ plays the role of $Q_k(t,x)$ in Section 5. It is also a sum of exponential functions of t, multiplied with ratio of rational functions of t and mgfs evaluated at t. Similarly to the case where $(\lambda_j)_j$ are not allowed to vary with x, one can prove as in Proposition 5.1 that if $k \leq J$ then $B^k(t,x)$ is nonzero for some t large enough, while if $k = J+1$ then $B^k(t,x) = 0$ for all values of t.

To identify the values of the mixing weights $(\lambda_j(y), \lambda_j(x))_{j\leq J}$ and of the regression functions $(m_j(y))_{j\leq J}$, we suggest the following procedure. Write $l_j(x) = \frac{\lambda_j(x)}{\lambda_j(y)}$. We know that

$$\begin{aligned}\dot{M}(0|x) &= \sum_j \lambda_j(x) m_j(x) \\ &= \sum_j \lambda_j(y) l_j(x) \dot{m}_j(x) + \sum_j \lambda_j(y) l_j(x) m_j(y),\end{aligned}$$

Evaluating this expression at distinct values $(x_p)_{p\leq P}$ gives

$$\begin{pmatrix} \dot{M}(0|x_1) \\ \vdots \\ \dot{M}(0|x_P) \end{pmatrix} = \begin{pmatrix} l_1(x_1)\dot{m}_1(x_1) & \dots & l_J(x_1)\dot{m}_J(x_1) & l_1(x_1) & \dots & l_J(x_1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ l_1(x_P)\dot{m}_1(x_P) & \dots & l_J(x_P)\dot{m}_J(x_P) & l_1(x_P) & \dots & l_J(x_P) \end{pmatrix} \begin{pmatrix} \lambda_1(y) \\ \vdots \\ \lambda_J(y) \\ \lambda_1(y)m_1(y) \\ \vdots \\ \lambda_J(y)m_J(y) \end{pmatrix}.$$

Taking $P = 2J$ and assuming that the matrix on the right hand side is invertible, which excludes the case where λ_j does not vary[1] with x, then the vector on the right hand side can be identified. This identifies $(\lambda_j(y))_{j\leq J}$, $(m_j(y))_{j\leq J}$ and thus $(\lambda_j(x))_{j\leq J}$. As for the distributions of the disturbances of each of the mixture components, they can be identified using decompositions such as (A.2).

A.6. Proofs of Results in Section 8

In this section, C and C' denote constants that may change from one line to the next.

[1] If $(\lambda_j(x))_{j\leq J}$ do not vary with x, then this is observed. Other possibilities for identification exist, such as arguments in the proof of Lemma 5.2.

Proof of Lemma 8.1; (8.3). The estimator can be decomposed as

$$\widehat{\Delta}_{10}m_1 = \frac{1}{t_n}\log\left(\frac{\widehat{M^1}(t_n)}{\widehat{M^0}(t_n)}\right) = \frac{1}{t_n}\log\left(\frac{M^1(t_n)}{M^0(t_n)}\right) + \frac{1}{t_n}\log\left(\frac{\widehat{M^1}(t_n)}{M^1(t_n)}\right) - \frac{1}{t_n}\log\left(\frac{\widehat{M^0}(t_n)}{M^0(t_n)}\right).$$

The first term in the decomposition is deterministic. This approximation error can be written

$$\frac{M^1(t_n)}{M^0(t_n)} = e^{t_n\Delta_{10}m_1}\frac{1+\frac{1-\lambda}{\lambda}e^{t_n[m_2(x_1)-m_1(x_1)]}\frac{M_2(t_n)}{M_1(t_n)}}{1+\frac{1-\lambda}{\lambda}e^{t_n[m_2(x_0)-m_1(x_0)]}\frac{M_2(t_n)}{M_1(t_n)}} = e^{t_n\Delta_{10}m_1}\left[1+O(\mu(t_n))\right],$$

where the last equality holds using Assumption 8.2 (iii). This gives

$$\frac{1}{t_n}\log\left(\frac{M^1(t_n)}{M^0(t_n)}\right) = \Delta_{10}m_1 + O(\frac{\mu(t_n)}{t_n}). \tag{A.1}$$

Let us now focus on the two estimation errors $\frac{1}{t_n}\log\left(\frac{\widehat{M^j}(t_n)}{M^j(t_n)}\right)$. We write

$$\widehat{M^j}(t_n) = \frac{\frac{1}{nh_n^k}\sum_{p=1}^n e^{t_nZ_p}K(\frac{X_p-x_j}{h_n})}{\frac{1}{nh_n^k}\sum_{p=1}^n K(\frac{X_p-x_j}{h_n})} = \frac{\widehat{N}^j(t_n)}{\widehat{D}^j},$$

and have

$$\frac{\widehat{M^j}(t_n)}{M^j(t_n)} = \frac{\widehat{N}^j(t_n)}{\widehat{D}^jM^j(t_n)} = \frac{f_X(x_j)}{\widehat{D}^j}\frac{\widehat{N}^j(t_n)}{f_X(x_j)M^j(t_n)}. \tag{A.2}$$

In what follows, we treat separately the two ratios appearing in the last equality in (A.2), showing that they both converge to 1. First, under Assumptions 8.1 and 8.3, standard results on Mean Square Error of the kernel density estimator guarantee

$$\frac{\widehat{D}^j}{f_X(x_j)} = 1 + O_{\mathbb{P}}\left(h_n^4 + \frac{1}{nh_n^k}\right)^{\frac{1}{2}}. \tag{A.3}$$

The analysis of the second ratio in the decomposition of (A.2) will differ from that of standard kernel regression as it depends on t_n which varies with sample size. For the bias term, we define $G_n(x) = f_X(x)\mathbb{E}(e^{t_nZ}|X=x)$. Then

$$\mathbb{E}(\widehat{N}^j(t_n)) = \mathbb{E}\left(\frac{1}{nh_n^k}\sum_{p=1}^n e^{t_nZ_p}K(\frac{X_p-x_j}{h_n})\right) = \int_{U\in\mathbb{R}^k} G_n(x_j+h_nU)K(U)\mathrm{d}U.$$

By Assumption 8.1, G_n is twice continuously differentiable. Since the kernel is of order 2 (Assumption 8.3), by virtue of the Mean Value Theorem, we have

$$\mathbb{E}\left(\frac{\widehat{N}^j(t_n)}{f_X(x_j)M^j(t_n)}\right) - 1 = \frac{1}{G_n(x_j)}\int\frac{h_n^2}{2}U'\nabla^2G_n[x_j+h_n\tau_n(U)U]UK(U)\mathrm{d}U$$

where $\tau_n(u) \in [0;1]$ and $\nabla^2 G_n(x)$ is the hessian matrix of the function G_n evaluated at x. Recall that $G_n(x) = f_X(x)[\lambda e^{t_n m_1(x)} M_1(t_n) + (1-\lambda) e^{t_n m_2(x)} M_2(t_n)]$. Thus

$$\begin{aligned}\nabla^2 G_n(x) = &\lambda e^{t_n m_1(x)} M_1(t_n) \ \big\{ t_n^2 f_x(x) \nabla m_1(x) \nabla m_1(x)' + \nabla^2 f_X(x) \\ &+ t_n \left(\nabla m_1(x) \nabla f_X(x)' + \nabla f_X(x) \nabla m_1(x)' + f_X(x) \nabla^2 m_1(x) \right) \big\} \\ &+ (1-\lambda) e^{t_n m_2(x)} M_2(t_n) \ \big\{ t_n^2 f_x(x) \nabla m_2(x) \nabla m_2(x)' + \nabla^2 f_X(x) \\ &+ t_n \left(\nabla m_2(x) \nabla f_X(x)' + \nabla f_X(x) \nabla m_2(x)' + f_X(x) \nabla^2 m_2(x) \right) \big\}\end{aligned}$$

The terms in braces are polynomial functions of t_n. We write $\nabla^2 G_n(x) = \lambda\, e^{t_n m_1(x)} M_1(t_n) \{ t_n^2 a_1(x) + t_n b_1(x) + c_1(x) \} + (1-\lambda)\, e^{t_n m_2(x)} M_2(t_n) \{ t_n^2 a_2(x) + t_n b_2(x) + c_2(x) \}$. By boundedness of the first order partial derivatives of m_i, $i = 1, 2$,

$$\exists \delta \geq 0,\ \forall (x, U) \in \mathbb{R}^k \times \mathbb{R}^k, |m_i(x + h_n \tau_n(U) U) - m_i(x)| \leq \delta h_n ||U||,$$

implying that $e^{t_n m_i(x + h_n \tau_n(U) U) - m_i(x)} \leq e^{\delta t_n h_n ||U||}$. Therefore, as $G_n(x) \geq f_X(x) \lambda e^{t_n m_1(x)} M_1(t_n)$,

$$\frac{\lambda e^{t_n m_1(x_j + h_n \tau_n(U) U)} M_1(t_n)}{G_n(x_j)} \leq \frac{e^{\delta h_n t_n ||U||}}{f_X(x_j)} \leq \frac{e^{C||U||}}{f_X(x_j)},$$

for some $C > 0$ for n large enough, by Assumptions 8.1 (iv) and (vi). The same holds for the $(1-\lambda)$ term. By Assumption 8.1, a_i, b_i and c_i, $i = 1, 2$, are bounded functions. This, together with the previous argument, implies that $\frac{1}{G_n(x_j)} \int U' \nabla^2 G_n[x_j + h_n \tau_n(U) U] U K(U) \mathrm{d}U = O(t_n^2)$. Thus

$$\mathbb{E}\left(\frac{\widehat{N}^j(t_n)}{f_X(x_j) M^j(t_n)} \right) - 1 = O(t_n h_n)^2.$$

For the variance term, an upper bound is

$$\begin{aligned}&\frac{1}{n} \mathbb{E}\left[\left(\frac{1}{h_n^k M^j(t_n) f_X(x_j)} e^{t_n Z} K\left(\frac{X - x_j}{h_n} \right) \right)^2 \right] \\ &= \frac{1}{n [h_n^k M^j(t_n) f_X(x_j)]^2} \int \mathbb{E}(e^{2 t_n Z} | X) K\left(\frac{X - x_j}{h_n} \right)^2 f_X(X) \mathrm{d}X \\ &= \frac{1}{n h_n^k} \int \frac{\mathbb{E}(e^{2 t_n Z} | h_n U + x_j)}{\mathbb{E}(e^{t_n Z} | x_j)^2} \frac{f_X(h_n U + x_j)}{f_X(x_j)^2} K(U)^2 \mathrm{d}U \\ &= \frac{1}{n h_n^k} \int \frac{\lambda e^{2 t_n m_1(h_n U + x_j)} M_1(2 t_n) + (1-\lambda) e^{2 t_n m_2(h_n U + x_j)} M_2(2 t_n)}{\left(\lambda e^{t_n m_1(x_j)} M_1(t_n) + (1-\lambda) e^{t_n m_2(x_j)} M_2(t_n) \right)^2} \ \frac{f_X(h_n U + x_j)}{f_X(x_j)^2} K(U)^2 \mathrm{d}U \\ &\leq \frac{1}{n h_n^k} \frac{M_1(2 t_n)}{M_1(t_n)^2} \int e^{2 \delta t_n h_n ||U||} \frac{\lambda + (1-\lambda) e^{2 t_n (m_2(h_n U + x_j) - m_1(h_n U + x_j)) \frac{M_2(2 t_n)}{M_1(2 t_n)}}}{\left(\lambda + (1-\lambda) e^{t_n (m_2(x_j) - m_1(x_j)) \frac{M_2(t_n)}{M_1(t_n)}} \right)^2} \ \frac{f_X(h_n U + x_j)}{f_X(x_j)^2} K(U)^2 \mathrm{d}U\end{aligned}$$

$$\leq \frac{1}{nh_n^k}\frac{M_1(2t_n)}{M_1(t_n)^2}\int Ce^{C||U||}\left[\lambda+(1-\lambda)e^{2t_n|m_2(x_j)-m_1(x_j)|+4C||U||}\frac{M_2(2t_n)}{M_1(2t_n)}\right]K(U)^2\mathrm{d}U,$$

where we used Assumptions 8.1 (iv) and (vi) for the first inequality, and Assumptions 8.1 (iv) and (vi) as well as Assumption 8.2 (iii) for the second inequality. Together with Assumption 8.2 (v), these assumptions also imply that the integrand is bounded above by $Ce^{C||U||}K(U)^2$, $\forall\, U \in \mathbb{R}^k$,. Note that for n large enough, $C \leq \alpha_0$. Thus Assumption 8.3 (iii) guarantees that the variance is of order $O(\frac{1}{nh_n^k}\frac{M_1(2t_n)}{M_1(t_n)^2})$. Therefore,

$$\frac{\widehat{N}^j(t_n)}{f_X(x_j)M^j(t_n)} = 1 + O_{\mathbb{P}}\left((t_nh_n)^4 + \frac{1}{nh_n^k}\frac{M_1(2t_n)}{M_1(t_n)^2}\right)^{\frac{1}{2}}. \tag{A.4}$$

With (A.2), (A.3) and (A.4), and given that by Jensen's inequality $\frac{M_1(2t_n)}{M_1(t_n)^2} \geq 1$,

$$\frac{\widehat{M^j}(t_n)}{M^j(t_n)} - 1 = O_{\mathbb{P}}\left((t_nh_n)^4 + \frac{1}{nh_n^k}\frac{M_1(2t_n)}{M_1(t_n)^2}\right)^{\frac{1}{2}}.$$

This finally gives

$$\begin{aligned}\widehat{\Delta}_{10}m_1 - \Delta_{10}m_1 &= O\left(\frac{\mu(t_n)}{t_n}\right) + \frac{1}{t_n}\log(\left[\left(1 + O_{\mathbb{P}}\left((t_nh_n)^4 + \frac{1}{nh_n^k}\frac{M_1(2t_n)}{M_1(t_n)^2}\right)^{\frac{1}{2}}\right)^2\right] \\ &= O_{\mathbb{P}}\left[\frac{\mu(t_n)}{t_n} + \frac{1}{t_n}\left((t_nh_n)^4 + \frac{1}{nh_n^k}\frac{M_1(2t_n)}{M_1(t_n)^2}\right)^{\frac{1}{2}}\right].\end{aligned}$$

□

Proof of Lemma 8.1; (8.4). The estimator is $\widehat{\Delta}_{10}m_2 = \frac{-i}{a_n}\,\mathrm{Log}\left(\frac{\widehat{\phi}^1(s_n+a_n)}{\widehat{\phi}^0(s_n+a_n)}\left(\frac{\widehat{\phi}^1(s_n)}{\widehat{\phi}^0(s_n)}\right)^{-1}\right)$. We first compute the rate of convergence of $\frac{\widehat{\phi}^0(s_n)}{\widehat{\phi}^1(s_n)}$, in a fashion similar to the proof above. From the proof of Lemma 3.6, we know

$$\lim_{s\to\infty}\frac{-i}{a}\,\mathrm{Log}\left(\frac{\phi(s+a|x_1)}{\phi(s+a|x_0)}\left(\frac{\phi(s|x_1)}{\phi(s|x_0)}\right)^{-1}\right) = \frac{1}{a}\left(a\Delta_{10}m_2 + 2\pi\left\lfloor\frac{1}{2} - \frac{a\Delta_{10}m_2}{2\pi}\right\rfloor\right).$$

Because we do not know the interval on which the identifying equation will be a constant of a, we plug in a sequence a_n going to zero instead of a fixed a. For the approximation error, we have

$$\begin{aligned}\frac{\phi^1(s_n)}{\phi^0(s_n)} &= \frac{\lambda e^{is_nm_1(x_1)}\phi_1(s_n) + (1-\lambda)e^{is_nm_2(x_1)}\phi_2(s_n)}{\lambda e^{is_nm_1(x_0)}\phi_1(s_n) + (1-\lambda)e^{is_nm_2(x_0)}\phi_2(s_n)} \\ &= e^{is_n\Delta_{10}m_2}\frac{\frac{\lambda}{1-\lambda}e^{is_n(m_1(x_1)-m_2(x_1))}\frac{\phi_1(s_n)}{\phi_2(s_n)} + 1}{\frac{\lambda}{1-\lambda}e^{is_n(m_1(x_0)-m_2(x_0))}\frac{\phi_1(s_n)}{\phi_2(s_n)} + 1} \\ &= e^{is_n\Delta_{10}m_2}(1 + O(f(s_n))).\end{aligned} \tag{A.5}$$

To compute the estimation error, the scheme is initially similar to the previous proof. We write

$$\widehat{\phi}^j(s_n) = \frac{\frac{1}{nh_n^k}\sum_{p=1}^n e^{is_nZ_p}K(\frac{X_p-x_j}{b_n})}{\frac{1}{nh_n^k}\sum_{p=1}^n K(\frac{X_p-x_j}{b_n})} = \frac{\widehat{num}^j(s_n)}{\widehat{denom}^j}$$

and work with an equation similar to (A.2), here

$$\frac{\widehat{\phi}^j(s_n)}{\phi^j(s_n)} = \frac{f_X(x_j)}{\widehat{denom}^j}\,\frac{\widehat{num}^j(s_n)}{f_X(x_j)\phi^j(s_n)}$$

We compute the convergence rate of the ratios in the last equality.

As in (A.3), we know $\frac{\widehat{denom}^j}{f_X(x_j)} = 1 + O_{\mathbb{P}}\left(b_n^4 + \frac{1}{nb_n^k}\right)^{\frac{1}{2}}$. Now, let $A_n = \frac{\widehat{num}^j(s_n)}{f_X(x_j)\phi^j(s_n)}$. Working with complex numbers for this proof, we use $|.|$ to denote a modulus. Let us focus on the bias term of A_n. We write $g_n(x) = f_X(x)\mathbb{E}(e^{is_nZ}|X=x)$ so that $A_n = \frac{\widehat{num}^j(s_n)}{g_n(x)}$.

Since $g_n(x) = f_X(x)(\lambda e^{is_nm_1(x)}\phi_1(s_n)+(1-\lambda)e^{is_nm_2(x)}\phi_2(s_n))$, we denote $G_n^{lc}(x) = \cos(s_nm_l(x))f_X(x)$ and $G_n^{ls}(x) = \sin(s_nm_l(x))f_X(x)$ for $l = 1, 2$. Then we have

$$\begin{aligned}
\mathbb{E}(\widehat{num}^j(s_n)) =& \mathbb{E}\left(\frac{1}{nb_n^k}\sum_{p=1}^n e^{is_nZ_p}K(\frac{X_p-x_j}{b_n})\right) = \int_{U\in\mathbb{R}^k} g_n(x_j+b_nU)K(U)\mathrm{d}U,\\
=&\lambda\phi_1(s_n)\int[\cos(s_nm_1(x_j+b_nU)) + i\sin(s_nm_1(x_j+b_nU))]f_X(x_j+b_nU)K(U)\mathrm{d}U\\
&+(1-\lambda)\phi_2(s_n)\int[\cos(s_nm_2(x_j+b_nU)) + i\sin(s_nm_2(x_j+b_nU))]f_X(x_j+b_nU)K(U)\mathrm{d}U\\
=&\lambda\phi_1(s_n)\int[G_n^{1c}(x_j+b_nU) + iG_n^{1s}(x_j+b_nU)]K(U)\mathrm{d}U\\
&+(1-\lambda)\phi_2(s_n)\int[G_n^{2c}(x_j+b_nU) + iG_n^{2s}(x_j+b_nU)]K(U)\mathrm{d}U
\end{aligned}$$

Using the assumption that the kernel is of order 2 (Assumption 8.3),

$$\int G_n^{1c}(x_j+b_nU)K(U)\mathrm{d}U - G_n^{1c}(x_j) = \int \frac{b_n^2}{2}U'\nabla^2G_n^{1c}[x_j+b_n\tau_n(U)U]\,UK(U)\,\mathrm{d}U,$$

where $\tau_n(U)\in[0;1]$ and $\nabla^2G_n^{1c}(x)$ is the hessian matrix of the function G_n^{1c} evaluated at x. That is,

$$\begin{aligned}
\nabla^2G_n^{1c}(x) =& -s_n^2f_X(x)\cos(s_nm_1(x))\nabla m_1(x)\nabla m_1(x)'\\
&-s_n\sin(s_nm_1(x))[\nabla m_1(x)\nabla f_X(x)' + \nabla f_X(x)\nabla m_1(x)' + f_X(x)\nabla^2m_1(x)]\\
&+\cos(s_nm_1(x))\nabla^2f_X(x).
\end{aligned}$$

Similarly to what is done in the first part of this proof, Assumption 8.1 guarantees that $\int G_n^{1c}(x_j + b_nU)K(U)\mathrm{d}U - G_n^{1c}(x_j) = O(b_ns_n)^2$. The same rate applies for G_n^{1s}, G_n^{2c} and G_n^{2s}, implying

$$\mathbb{E}(\widehat{num}^j(s_n)) = \int g_n(x_j + b_nU)K(U)\mathrm{d}U = g_n(x_j) + O(\,(b_ns_n)^2\,[\lambda|\phi_1(s_n)| + (1-\lambda)|\phi_2(s_n)|\,]\,),$$

which gives, for the bias term,

$$\begin{aligned}
\mathbb{E}(A_n) &= \frac{1}{f_X(x_j)\phi^j(s_n)}\mathbb{E}(\widehat{num}^j(s_n)) \\
&= 1 + O\left((b_ns_n)^2\frac{1}{f_X(x_j)}\frac{\lambda|\phi_1(s_n)| + (1-\lambda)|\phi_2(s_n)|}{\lambda e^{is_nm_1(x_j)}\phi_1(s_n) + (1-\lambda)e^{is_nm_2(x_j)}\phi_2(s_n)}\right) \\
&= 1 + O\left((b_ns_n)^2\frac{1}{e^{is_nm_2(x_j)}}\frac{\lambda\frac{|\phi_1(s_n)|}{|\phi_2(s_n)|} + (1-\lambda)}{\lambda\frac{\phi_1(s_n)}{|\phi_2(s_n)|}e^{is_n(m_1(x_j)-m_2(x_j))} + (1-\lambda)}\right) \\
&= 1 + O(b_ns_n)^2,
\end{aligned}$$

where the last equality comes from Assumption 8.2 (iv). As for the variance term, we write

$$\begin{aligned}
Var(\frac{\widehat{num}^j(s_n)}{f_X(x_j)\phi^j(s_n)}) &= \frac{1}{f_X(x_j)^2|\phi^j(s_n)|^2}\frac{1}{nb_n^{2k}}Var(e^{is_nZ}K(\frac{X-x_j}{b_n})) \\
&\leq \frac{1}{f_X(x_j)^2|\phi^j(s_n)|^2}\frac{1}{nb_n^{2k}}\mathbb{E}(|e^{is_nZ}K(\frac{X-x_j}{b_n})|^2) \\
&\leq \frac{1}{|\phi^j(s_n)|^2}\frac{1}{nb_n^k}\frac{\int f_X(x_j+b_nU)K^2(U)\mathrm{d}U}{f_X(x_j)^2},
\end{aligned}$$

and Assumption 8.1 (v) guarantees that in the last equality, the third term in the product converges to $\frac{\int K^2(U)\mathrm{d}U}{f_X(x_j)}$. Moreover,

$$\begin{aligned}
|\phi^j(s_n)| &= |\lambda e^{is_nm_1(x_j)}\phi_1(s_n) + (1-\lambda)e^{is_nm_2(x_j)}\phi_2(s_n)| \\
&= |\phi_2(s_n)|\ \left|\lambda e^{is_nm_1(x)}\frac{\phi_1(s_n)}{\phi_2(s_n)} + (1-\lambda)e^{is_nm_2(x)}\right| \sim_{n\to\infty} (1-\lambda)|\phi_2(s_n)|,
\end{aligned}$$

therefore implying $Var(A_n) = O(\frac{1}{nb_n^k|\phi_2(s_n)|^2})$.

From these two computations, the following reasoning gives a convergence rate for A_n. First, $Bias(\Re(A_n)) = \Re(Bias(A_n)) = O(b_ns_n)^2$. Similarly $Bias(\Im(A_n)) = O(b_ns_n)^2$. Plus, by definition for a complex random variable $Var(A_n) = Var(\Re(A_n)) + Var(\Im(A_n))$: both the variances of the real part and the imaginary part are smaller than the variance of A_n. An upper bound of the rates of convergence of the Mean Square Error of the real and imaginary parts is therefore obtained, $\Re(A_n) - 1 = O_{\mathbb{P}}((b_ns_n)^4 + \frac{1}{nb_n^k|\phi^j(s_n)|^2})^{1/2}$, and $\Im(A_n) = O_{\mathbb{P}}((b_ns_n)^4 + \frac{1}{nb_n^k|\phi^j(s_n)|^2})^{1/2}$. This gives

$$|A_n - 1| = O_{\mathbb{P}}((b_ns_n)^4 + \frac{1}{nb_n^k|\phi^j(s_n)|^2})^{1/2},$$

so that the estimation error is

$$\begin{aligned}\frac{\widehat{\phi}^j(s_n)}{\phi^j(s_n)} &= 1 + O_{\mathbb{P}}\left[\left(b_n^4 + \frac{1}{nb_n^k}\right)^{\frac{1}{2}} + \left((b_n s_n)^4 + \frac{1}{nb_n^k|\phi^j(s_n)|^2}\right)^{1/2}\right] \\ &= 1 + O_{\mathbb{P}}\left((b_n s_n)^4 + \frac{1}{nb_n^k|\phi^j(s_n)|^2}\right)^{1/2}.\end{aligned}$$

Finally we obtain

$$\begin{aligned}\frac{\widehat{\phi}^1(s_n)}{\widehat{\phi}^0(s_n)} &= \frac{\widehat{\phi}^1(s_n)}{\phi^1(s_n)}\left(\frac{\widehat{\phi}^0(s_n)}{\phi^0(s_n)}\right)^{-1}\frac{\phi^1(s_n)}{\phi^0(s_n)} \\ &= e^{is_n\Delta_{10}m_2}[1 + O_{\mathbb{P}}(f(s_n))]\ \left[1 + O_{\mathbb{P}}\left((b_n s_n)^4 + \frac{1}{nb_n^k|\phi^j(s_n)|^2}\right)^{1/2}\right].\end{aligned}$$

Plugging in this expression in the definition of the estimator, we obtain

$$\begin{aligned}\widehat{\Delta}_{10}m_2 =& \frac{-i}{a_n}\ \mathrm{Log}\left(\frac{\widehat{\phi}^1(s_n+a_n)}{\widehat{\phi}^0(s_n+a_n)}\left(\frac{\widehat{\phi}^0(s_n)}{\widehat{\phi}^1(s_n)}\right)\right) \\ =& \frac{-i}{a_n}\ \mathrm{Log}\left\{e^{i(s_n+a_n)\Delta_{10}m_2}[1 + O_{\mathbb{P}}(f(s_n+a_n))]\left[1 + O_{\mathbb{P}}\left(b_n^4(s_n+a_n)^4 + \frac{1}{nb_n^k|\phi^j(s_n+a_n)|^2}\right)^{1/2}\right]\right. \\ & \left. e^{-is_n\Delta_{10}m_2}[1 + O_{\mathbb{P}}(f(s_n))]\left[1 + O_{\mathbb{P}}\left(b_n^4 s_n^4 + \frac{1}{nb_n^k|\phi^j(s_n)|^2}\right)^{1/2}\right]\right\} \\ =& \frac{-i}{a_n}\ \mathrm{Log}\left\{e^{ia_n\Delta_{10}m_2}\left[1 + O_{\mathbb{P}}(f(s_n+a_n)) + O_{\mathbb{P}}(f(s_n)) + O_{\mathbb{P}}\left(b_n^4 s_n^4 + \frac{1}{nb_n^k|\phi^j(s_n+a_n)|^2}\right)^{1/2}\right.\right. \\ & \left.\left. + O_{\mathbb{P}}\left(b_n^4 s_n^4 + \frac{1}{nb_n^k|\phi^j(s_n)|^2}\right)^{1/2}\right]\right\},\end{aligned}$$

where we used $a_n \to 0$ as $n \to \infty$. As the term multiplying $e^{ia_n\Delta_{10}m_2}$ in the Log converges to 1, and eventually $a_n\Delta_{10}m_2 \in (-\pi;\pi)$, the expression above becomes

$$\begin{aligned}\widehat{\Delta}_{10}m_2 =& \Delta_{10}m_2 + \frac{-i}{a_n}\ \mathrm{Log}\left[1 + O_{\mathbb{P}}(f(s_n+a_n)) + O_{\mathbb{P}}(f(s_n))\right. \\ & \left. + O_{\mathbb{P}}\left(b_n^4 s_n^4 + \frac{1}{nb_n^k|\phi^j(s_n+a_n)|^2}\right)^{1/2} + O_{\mathbb{P}}\left(b_n^4 s_n^4 + \frac{1}{nb_n^k|\phi^j(s_n)|^2}\right)^{1/2}\right],\end{aligned}$$

that is, using the first order approximation of the principal value of the log around 1,

$$\widehat{\Delta}_{10}m_2 = \Delta_{10}m_2 + \frac{1}{a_n}O_{\mathbb{P}}\left(f(s_n+a_n) + f(s_n) + \left(b_n^4 s_n^4 + \frac{1}{nb_n^k|\phi^j(s_n+a_n)|^2}\right)^{1/2} + \left(b_n^4 s_n^4 + \frac{1}{nb_n^k|\phi^j(s_n)|^2}\right)^{1/2}\right).$$

□

Proof of Lemma 8.3. Fix $z \in \mathbb{R}$. Write $\xi_j^0 = z + j\delta(x_1, x_0) + m_2(x_0)$, $\widehat{\xi}_j^0 = z + j\widehat{\delta}(x_1, x_0) + \widehat{m}_2(x_0)$, $\xi_j^1 = z + j\delta(x_1, x_0) + m_1(x_1) - g(x_0)$ and $\widehat{\xi}_j^1 = z + j\widehat{\delta}(x_1, x_0) + \widehat{m}_1(x_1) - \widehat{g}(x_0)$.

$$\begin{aligned}\widehat{F}_2(z) - F_2(z) =& \frac{1}{1-\lambda}\sum_{j=0}^{p_n}\left\{\left[\widehat{F}(\widehat{\xi}_j^1|x_1) - F(\xi_j^1|x_1)\right] - \left[\widehat{F}(\widehat{\xi}_j^0|x_0) - F(\xi_j^0|x_0)\right]\right\} \\ &- \frac{1}{1-\lambda}\sum_{j=p_n+1}^{\infty}\left[F(\xi_j^1|x_1) - F(\xi_j^0|x_0)\right] \\ &+ \left(\frac{1}{1-\widehat{\lambda}} - \frac{1}{1-\lambda}\right)\sum_{j=0}^{p_n}\widehat{F}(\widehat{\xi}_j^1|x_1) - \widehat{F}(\widehat{\xi}_j^0|x_0).\end{aligned} \tag{A.6}$$

We write $\widehat{F}_2(z) - F_2(z) = I_1 - I_2 + I_3$ and compute separately the convergence rate of each part on the right hand side of (A.6). For I_1, we write

$$\begin{aligned}(1-\lambda)\, I_1 =& \sum_{j=0}^{p_n}\left[\widehat{F}(\widehat{\xi}_j^1|x_1) - F(\widehat{\xi}_j^1|x_1)\right] + \sum_{j=0}^{p_n}\left[F(\widehat{\xi}_j^1|x_1) - F(\xi_j^1|x_1)\right] \\ &- \sum_{j=0}^{p_n}\left[\widehat{F}(\widehat{\xi}_j^0|x_0) - F(\widehat{\xi}_j^0|x_0)\right] - \sum_{j=0}^{p_n}\left[F(\widehat{\xi}_j^0|x_0) - F(\xi_j^0|x_0)\right].\end{aligned}$$

Since $\frac{\partial F(z|x)}{\partial z} = f(z|x) = \lambda f_1(z - m_1(x)) + (1-\lambda) f_2(z - m_2(x))$ and by Assumption 8.4 $f(z|x)$ is bounded, $z \mapsto F(z|x)$ is Lipschitz continuous. That is, for $v = 0, 1$, $|F(\widehat{\xi}_j^v|x_v) - F(\xi_j^v|x_v)| \leq C\,|\widehat{\xi}_j^v - \xi_j^v|$ with probability one, implying

$$|\sum_{j=0}^{p_n} F(\widehat{\xi}_j^v|x_v) - F(\xi_j^v|x_v)| = O_{\mathbb{P}}(p_n^2 \varepsilon_n).$$

For the two remaining terms in I_1, we write

$$F_n(.|x_i) = \frac{\mathbb{E}\left(\mathbb{1}(Z \leq z) k\left(\frac{X - x_i}{c_n}\right)\right)}{\mathbb{E}\left(k\left(\frac{X - x_i}{c_n}\right)\right)}.$$

We now apply Theorem 3 of Einmahl and Mason (2005). Assumptions (K-i) and (K-ii) of this paper hold by Assumption 8.5. Assumption (K-iii) holds by Assumption 8.5 (iii). Assumption (K-iv) holds by continuity of k, see Assumption 8.5(i) and the comment below the statement of Assumption (K-iv) in Einmahl and Mason (2005). Note that by $c_n = n^{-\frac{1}{k+4}}$, $c_n \to 0$ as $n \to \infty$ and for any C, $c_n \geq C \log n / n$ for n large enough. Finally, take $I = \{x_0, x_1\}$, then Condition (1.7) of Einmahl and Mason (2005) holds by Assumption 8.1 (ii). Thus we have,

$$\limsup_{n\to\infty} \sup_{x \in I} (n c_n^k)^{1/2} ||\widehat{F}(.|x) - F_n(.|x)||_\infty = O_{\text{a.s}}\left(\max(\log\log n, -\log(c_n))^{1/2}\right)$$

$$= O_{\text{a.s}}\left((-\log c_n)^{1/2}\right). \tag{A.7}$$

We now examine $F_n(.|x) - F(.|x)$. Let $f(.,.)$ be the joint density of (Z, X), and

$$F_X(z,x) = \int_{z' \leq z} f(z',x)\mathrm{d}z' = F(z|x)f_X(x) = [\lambda F_1(z - m_1(x)) + (1-\lambda)F_2(z - m_2(x))]f_X(x).$$

Write for $i = 0, 1$,

$$F_n(z|x_i) - F(z|x_i) = \frac{\frac{1}{c_n^k}\mathbb{E}(\mathbb{1}(Z \leq z)k\left(\frac{X-x_i}{c_n}\right)) - F_X(z,x_i)}{\frac{1}{c_n^k}\mathbb{E}\left(k\left(\frac{X-x_i}{c_n}\right)\right)} + F_X(z,x_i)\left(\frac{1}{\frac{1}{c_n^k}\mathbb{E}\left(k\left(\frac{X-x_i}{c_n}\right)\right)} - \frac{1}{f_X(x_i)}\right).$$

Under Assumption 8.1, 8.4 and 8.5, we know that $\frac{1}{c_n^k}\mathbb{E}\left(k\left(\frac{X-x_i}{c_n}\right)\right) - f_X(x_i) = O(c_n^2)$. Similarly,

$$\frac{1}{c_n^k}\mathbb{E}\left[\mathbb{1}(Z \leq z)\, k\left(\frac{X-x_i}{c_n}\right)\right] - F_X(z,x_i) = \frac{c_n^2}{2}\int_{U\in\mathbb{R}^k} U'\,\nabla_X^2 F_X(z, x_i + b_n\tau_n(U)U)\,U k(U)\mathrm{d}U,$$

with Assumption 8.1, 8.4 and 8.5 guaranteeing that $\nabla_X^2 F_X(.,.)$ exists and is uniformly bounded over $\mathbb{R}^{k+1}$. Therefore

$$\sup_{z\in\mathbb{R}}\left|\ \frac{1}{c_n^k}\mathbb{E}\left[\mathbb{1}(Z \leq z)\, k\left(\frac{X-x_i}{c_n}\right)\right] - F_X(z,x_i)\right| = O(c_n^2).$$

which gives, for $i = 0, 1$,

$$||F_n(.|x_i) - F(.|x_i)||_\infty = O(c_n^2). \tag{A.8}$$

Equations (A.7) and (A.8) give $||\widehat{F}(.|x) - F(.|x)||_\infty = O_{\mathbb{P}}((-\log(c_n))^{1/2}(nc_n^k)^{-1/2} + c_n^2)$. Recall $c_n = n^{-\frac{1}{k+4}}$, thus for any small $a > 0$,

$$\sup_{y\in\mathbb{R}}|\widehat{F}(y|x_i) - F(y|x_i)| = O_{\mathbb{P}}(n^{-\frac{2}{k+4}+a}),\ i = 0, 1.$$

This implies that

$$\sum_{j=0}^{p_n}\left[\widehat{F}(\widehat{\xi}_j^1|x_i) - F(\widehat{\xi}_j^1|x_i)\right] = O_{\mathbb{P}}(p_n n^{-\frac{2}{k+4}+a}),\ i = 0, 1.$$

Therefore,

$$I_1 = O_{\mathbb{P}}\left(pn^{-\frac{2}{k+4}+a} + p_n^2\varepsilon_n\right).$$

Looking at I_2, by construction the second sum appearing in the right hand side of (A.6) simplifies to

$$\frac{1}{1-\lambda}\sum_{j=p_n+1}^{\infty}\left[F(\xi_j^1|x_1) - F(\xi_j^0|x_0)\right] = 1 - F_2(z + (p_n+1)\delta(x_1,x_0)).$$

Using the exponential version of the Chebyshev's inequality, we have $1 - F_2(C) = \mathbb{P}(\epsilon_2 > C) \leq e^{-tC}M_2(t)$ for any $t \geq 0$. For a fixed value of t and letting $\gamma_0 = \delta(x_1,x_0)t \geq 0$,

$$I_2 = O(e^{-\gamma_0\, p_n}).$$

As for I_3, note that the computations above give $\sum_{j=0}^{p_n} \widehat{F}(\widehat{\xi}_j^1|x_1) - \widehat{F}(\widehat{\xi}_j^0|x_0) \xrightarrow[n\to\infty]{} (1-\lambda)F_2(z)$. As $\frac{1}{1-\widehat{\lambda}} - \frac{1}{1-\lambda} = O_{\mathbb{P}}(\varepsilon_n)$, we have

$$I_3 = O_{\mathbb{P}}(\varepsilon_n).$$

Adding these three parts, we obtain

$$\widehat{F}_2(z) - F_2(z) = O_{\mathbb{P}}\left(p_n n^{-\frac{2}{k+4}+a} + p_n^2\varepsilon_n + e^{-\gamma_0 p_n}\right).$$

Using the equation $F(z|x) = \lambda F_1(z - m_1(x)) + (1-\lambda)F_2(z - m_2(x))$, an estimator of $F_1(z)$ is

$$\widehat{F}_1(z) = \frac{1}{\widehat{\lambda}}\left[\widehat{F}(z + \widehat{m}_1(x)) - (1-\widehat{\lambda})\widehat{F}_2(z + \widehat{m}_1(x) - \widehat{m}_2(x))\right],$$

which will converge to $F_1(z)$ at the same rate. □

References

Dunford, N., and J. T. Schwartz (1958): *Linear operators part I: general theory*, vol. 7. Interscience publishers New York.

Einmahl, U., and D. M. Mason (2005): "Uniform in bandwidth consistency of kernel-type function estimators," *The Annals of Statistics*, 33(3), 1380–1403.

Feller, W. (1968): *An introduction to probability theory and its applications*, vol. 1. John Wiley & Sons.

Tossavainen, T. (2006): "On the zeros of finite sums of exponential functions," *Australian Mathematical Society Gazette*, 33(1), 47.

Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520.

E-mail address: yuichi.kitamura@yale.edu

Department of Economics, Georgetown University, Washington, D.C., 20007.

E-mail address: louise.laage@georgetown.edu