NONPARAMETRIC ANALYSIS OF FINITE MIXTURES

YUICHI KITAMURA AND LOUISE LAAGE

Abstract. Finite mixture models are useful in applied econometrics. They can be used to model unobserved heterogeneity, which plays major roles in labor economics, industrial organization and other fields. Mixtures are also convenient in dealing with contaminated sampling models and models with multiple equilibria. This paper shows that finite mixture models are nonparametrically identified under weak assumptions that are plausible in economic 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 to structural econometric models, such as auction models with unobserved heterogeneity. 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. This means a finite mixture model with function-valued unobserved heterogeneity can be identified in a cross-section setting, without restricting the dependence pattern between the regressor and the unobserved heterogeneity. In this aspect it is akin to fixed effects panel data models which permit unrestricted correlation between unobserved heterogeneity and covariates. 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 economics it is often crucially important to control for unobserved heterogeneity, and mixture models provide convenient ways to deal with it. This paper studies identification problems

---

Date: This Version: November 6, 2018.

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

JEL Classification Number: C14.

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. Kitamura acknowledges financial support from the National Science Foundation.

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 w 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 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 we now summarize 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 econometrician 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, which we do not assume either.

(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 Klein and Sherman (2002), who consider semiparametric estimation of ordered response models (hence their paper is not about mixtures) though the identification strategy in the current paper 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 6). 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, Cameron and Heckman (1998), Keane and Wolpin (1997), Berry, Carnall, and Spiller

(1996), Arcidiacono and Miller (2011), and Aguirregabiria and Mira (2013) for applications of finite mixture models in these fields. They are also used extensively in duration models with unobserved heterogeneity; see Heckman and Singer (1984), Heckman and Taber (1994) and Van den Berg (2001). A somewhat different use of mixtures can be found in models of regime changes, which can be viewed as finite mixture models. Porter (1983), for example, uses a switching simultaneous equations for an empirical IO model (see also Ellison (1994) and Lee and Porter (1984)). Some models with multiple equilibria can be regarded as mixtures as well (e.g. Berry and Tamer (2006), Echenique and Komunjer (2009)). Finally, contaminated models, as analyzed by Horowitz and Manski (1995) and Manski (2003) can be formulated as mixture models.

The most common estimation method for mixture models is parametric maximum likelihood (ML). In the notation introduced in (1.1), ML 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 ML estimator 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 Jewell (1982) and Heckman and Singer (1984) 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 ML 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 (Teicher 1961, 1963) initiated this research area. Rao (1992) provides a nice summary of this topic. See, also, Lindsay (1995) for a comprehensive treatment of mixture models including their identification issues. Many results known in this area assume parametric component distributions. Indeed, as Hall and Zhou (2003) put it, "(v)ery little is known of the potential for consistent nonparametric inference in mixtures without training data." Nevertheless, a number of papers have appeared on this subject, especially after the first version of the current paper was circulated. These include approaches based on multiple outcomes (e.g. Bonhomme, Jochmans, and

Robin (2016b), Bonhomme, Jochmans, and Robin (2016a), D'Haultfœuille and Février (2015), Kasahara and Shimotsu (2009)), or identification results based on exclusion restrictions, with/without tail restrictions on component distributions (e.g. Adams (2016), Compiani and Kitamura (2016), Henry, Kitamura, and Salanié (2014), Henry, Kitamura, and Salanié (2010), Hohmann and Holzmann (2013a), Jochmans, Henry, and Salanié (2017)), or methods based on symmetry restrictions (e.g. Butucea and Vandekerkhove (2014), Hohmann and Holzmann (2013b)).

The main result of the present paper is that nonparametric treatment of the component distributions of a finite mixture model is possible in a cross-sectional setting, if appropriate covariates are available.

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. Sections 3 - 5 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 6 considers an extension to the case with $J \geq 3$.

3. Regression

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 the random variables

$$\epsilon_j = z_j - m_j(x), \quad j = 1, 2.$$

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

$$F(z|x) = \lambda F^1_{\epsilon|x}(z - m_1(x)) + (1 - \lambda)F^2_{\epsilon|x}(z - m_2(x)). \tag{3.1}$$

Our goal in this section is then to identify the elements of the right hand side of (3.1) nonparametrically from the knowledge of $F(\cdot|x)$ evaluated at various x. Note that the model (3.1) is further interpreted as a switching regression model:

$$z = \begin{cases} m_1(x) + \epsilon_1, \epsilon_1|x \sim F^1_{\epsilon|x} & \text{with probability } \lambda \\ m_2(x) + \epsilon_2, \epsilon_2|x \sim F^2_{\epsilon|x} & \text{with probability } 1 - \lambda. \end{cases} \tag{3.2}$$

Models as described above are conventionally estimated using parametric ML. 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 Quandt (1972) and Kiefer (1978); see also Hamilton (1989) for application of ML in a time series context. The EM algorithm is often used in computing the ML estimator.

While the parametric approach is attractive and practical, the consistency of ML 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 (3.2) 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 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 (3.2). 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 $x_0 = (x_0^1, ..., x_0^k)^\top \in \mathbb{R}$, define

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

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, (3.1) is written as:

$$F(z|x) = \lambda F_1(z - m_1(x)) + (1-\lambda) F_2(z - m_2(x)). \tag{3.3}$$

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 (3.1) becomes

$$F(z|x) = F_1(z - m_1(x)). \tag{3.4}$$

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

We first discuss identification 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

$$D(x) := m_2(x) - m_1(x)$$

on $\mathbb{R}^k$ and

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

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 (3.1) 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 are $\mathbb{R}^k \times (-\infty, \infty)$ by Assumption 3.2(i).

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.

Proof of Lemma 3.1. First consider the case with $0 < \lambda < 1$. By the continuity condition (Assumption 3.1(iii)), there exist a $\delta' \in (0, \delta)$ such that

$$|m_2(x') - m_2(x_0)| < \frac{\epsilon|D(x_0)|}{2} \qquad \text{and} \qquad |m_1(x') - m_1(x_0)| < \frac{\epsilon|D(x_0)|}{2} \tag{3.5}$$

for all $x' \in N^1(x_0, \delta')$. By (3.1) we have

$$M(t|x) = \lambda e^{tm_1(x)} M_1(t) + (1 - \lambda)e^{tm_2(x)} M_2(t). \tag{3.6}$$

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 (3.5) 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

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

If both hold, then it has to be the case that $D(x_0) = 0$. If, on top of that, $D(x') = m_2(x') - m_1(x') > 0$ then

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

and

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

The analysis of the case with $D(x') = m_2(x') - m_1(x') < 0$ 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. □

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). The constant δ in the following condition will be specified in the statements of Lemmas 3.1 and 3.3.

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.

With this condition we have:

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.*

Proof. 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

$$\mathbb{E}[z|x] = \lambda m_1(x) + (1-\lambda)m_2(x), \tag{3.7}$$

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

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

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

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.

Remark 3.4. A sufficient condition.

The next Lemma gives a full identification result. Let $\mathcal{F}(\mathbb{R}^p)$ denote the space of distribution functions on $\mathbb{R}^p$ for some $p \in \mathbb{N}$. Define

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

the set of distribution functions 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 real valued functions on $\mathcal{C}$.

Lemma 3.3. *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 $(\lambda, F_1(\cdot), F_2(\cdot), m_1(\cdot), m_2(\cdot))$ in the set $(0,1] \times \bar{\mathcal{F}}(\mathbb{R})^2 \times \mathcal{V}(N^1(x_0,\delta'))^2$ up to labeling.*

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

$$\dot{M}(0|x) = \int zdF(z|x) = \lambda m_1(x) + (1-\lambda)m_2(x). \tag{3.8}$$

Using this,

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

By this and Assumption 3.1(ii), if $0 < \lambda < 1$, λ is identified from

$$\lambda = \frac{[\dot{M}(0|x_0) - \dot{M}(0|x)] - \lim_{t\to-\infty} \frac{1}{t}\log\frac{M(t|x_0)}{M(t|x)}}{\lim_{t\to+\infty} \frac{1}{t}\log\frac{M(t|x_0)}{M(t|x)} - \lim_{t\to-\infty} \frac{1}{t}\log\frac{M(t|x_0)}{M(t|x)}} \tag{3.9}$$

evaluated at an arbitrary $x \in N^1(x_0,\delta')$ (note δ' is defined in Lemma 3.1), since $\lim_{t\to+\infty}\frac{1}{t}\log\frac{M(t|x_0)}{M(t|x)}$ and $\lim_{t\to-\infty}\frac{1}{t}\log\frac{M(t|x_0)}{M(t|x)}$ identify the factors $[m_1(x_0) - m_1(x)]$ and $[m_2(x_0) - m_2(x)]$ by Lemma 3.1 (here and in what follows, we assume that $m_2(x_0) - m_1(x_0) < 0$; if $m_2(x_0) - m_1(x_0) > 0$, λ should be replaced by $(1-\lambda)$). The right hand side of (3.9), however, is not well-defined ($=0/0$) if the mixing distribution is degenerate, i.e. $\lambda = 1$. To avoid the discontinuity, let

$$\lambda_\delta = \frac{[\dot{M}(0|x_0) - \dot{M}(0|x)] - (1+\delta)\lim_{t\to-\infty}\frac{1}{t}\log\frac{M(t|x_0)}{M(t|x)}}{\lim_{t\to+\infty}\frac{1}{t}\log\frac{M(t|x_0)}{M(t|x)} - (1+\delta)\lim_{t\to-\infty}\frac{1}{t}\log\frac{M(t|x_0)}{M(t|x)}},$$

which approaches to λ as $\delta \to 0$ whether $\lambda < 1$ or not. Thus λ is determined by

$$\lambda = \lim_{\delta\to 0}\lambda_\delta.$$

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

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

Let

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

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 (3.8),

$$\begin{bmatrix} C(x) \\ \dot{M}(0|x_0)\end{bmatrix} = \begin{pmatrix}[m_1(x_0) - m_1(x)] & [m_2(x_0) - m_2(x)] \\ 1 & 1\end{pmatrix}\begin{pmatrix}\lambda & 0 \\ 0 & (1-\lambda)\end{pmatrix}\begin{bmatrix} m_1(x_0) \\ m_2(x_0)\end{bmatrix}, \tag{3.10}$$

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 (3.10) 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 (3.6) at x_0 and $x \in N^1(x_0,\delta')$, $x \neq x_0$, obtain

$$\begin{bmatrix} M(t|x_0) \\ M(t|x) \end{bmatrix} = E(x_0,x,t)\Lambda \begin{bmatrix} M_1(t) \\ M_2(t) \end{bmatrix}, \tag{3.11}$$

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 (3.11) 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 (3.11) 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. $\square$

Remark 3.5. 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 as in 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 3.6. 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 3.7. Some of the assumptions made above are crucial. The main source of identification is the independence assumption (Assumption 3.1(i)), as discussed before. Also Assumption 3.1(ii) is essential. If we have m_1 and m_1 that are completely parallel everywhere, it is easy to see that the "shift restriction" implied by independence loses its identifying power.

Remark 3.8. On the other hand, some of the assumption 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 by a large class 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 propose an alternative approach for identifying (3.1). 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. 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 $N(0, \sigma_1^2)$, $N(0, \sigma_2^2)$, $\sigma_1^2 \neq \sigma_2^2$. Teicher (1963) 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.*

Proof. Define $\delta(x) := m_2(x) - m_1(x_0)$. Note

$$\rho(x, t) = e^{it[m_1(x) - m_1(x_0)]} \frac{1 + \frac{1-\lambda}{\lambda} e^{it\delta(x)} \frac{\phi_2(t)}{\phi_1(t)}}{1 + \frac{1-\lambda}{\lambda} e^{it\delta(x_0)} \frac{\phi_2(t)}{\phi_1(t)}} \tag{3.12}$$

$$= e^{it[m_2(x) - m_2(x_0)]} \frac{\frac{1-\lambda}{\lambda} e^{-it\delta(x)} \frac{\phi_1(t)}{\phi_2(t)} + 1}{\frac{1-\lambda}{\lambda} e^{-it\delta(x_0)} \frac{\phi_1(t)}{\phi_2(t)} + 1} \tag{3.13}$$

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 (3.12) and (3.13), 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

$$\limsup_{t\to\infty} \left|\frac{\phi_1(t)}{\phi_2(t)}\right| = C, C \in (0,\infty] \tag{3.14}$$

and

$$\limsup_{t\to\infty} \left|\frac{\phi_2(t)}{\phi_1(t)}\right| = C', C' \in (0,\infty]. \tag{3.15}$$

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 (3.13) 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 (3.12) and (3.13), 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

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

hold simultaneously, which contradicts the second part of Condition 3.2. □

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 regressors cases where one regressor is binary and the other is continuous.

Lemma 3.5. *Under Assumption 3.4 (or Condition 3.2), as well as 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.*

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 (3.1)

$$\phi(t|x) = \lambda e^{itm_1(x)}\phi_1(t) + (1-\lambda)e^{itm_2(x)}\phi_2(t). \tag{3.16}$$

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

$$F(z|x) = \lambda^* F_1^*(z - m_1^*(x)) + (1-\lambda^*)F_2^*(z - m_2^*(x)), \qquad x = x_a, x_b, x_c. \tag{3.17}$$

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

$$\lambda e^{itm_1(x_a)}\phi_1(t) + (1-\lambda)e^{itm_2(x_a)}\phi_2(t) = \lambda^* e^{itm_1^*(x_a)}\phi_1^*(t) + (1-\lambda^*)e^{itm_2^*(x_a)}\phi_2^*(t), \tag{3.18}$$

$$\lambda e^{itm_1(x_b)}\phi_1(t) + (1-\lambda)e^{itm_2(x_b)}\phi_2(t) = \lambda^* e^{itm_1^*(x_b)}\phi_1^*(t) + (1-\lambda^*)e^{itm_2^*(x_b)}\phi_2^*(t), \tag{3.19}$$

$$\lambda e^{itm_1(x_c)}\phi_1(t) + (1-\lambda)e^{itm_2(x_c)}\phi_2(t) = \lambda^* e^{itm_1^*(x_c)}\phi_1^*(t) + (1-\lambda^*)e^{itm_2^*(x_c)}\phi_2^*(t). \tag{3.20}$$

Let α and β be arbitrary two 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 (3.19) by $e^{it\Delta_{ab}m_2^*}$ then subtract both sides from (3.18) to obtain

$$\lambda H(t; m_2^*, m_1, a, b)\phi_1(t) + (1-\lambda)H(t; m_2^*, m_2, a, b)\phi_2(t) = \lambda^* H(t; m_2^*, m_1^*, a, b)\phi_1^*(t). \tag{3.21}$$

Repeat this with (3.19) and $e^{it\Delta_{ab}m_2^*}$ replaced by (3.20) and $e^{it\Delta_{ac}m_2^*}$:

$$\lambda H(t; m_2^*, m_1, a, c)\phi_1(t) + (1-\lambda)H(t; m_2^*, m_2, a, c)\phi_2(t) = \lambda^* H(t; m_2^*, m_1^*, a, c)\phi_1^*(t). \tag{3.22}$$

(3.21) and (3.22) imply

$$\lambda H(t;m_2^*,m_1,a,b)H(t;m_2^*,m_1^*,a,c)\phi_1(t)+(1-\lambda)H(t;m_2^*,m_2,a,b)H(t;m_2^*,m_1^*,a,c)\phi_2(t) = \lambda^* H(t;m_2^*,m_1^*,a,b)H(t;m_2^*,m_1^*,a,c)\phi_1^*(t), \tag{3.23}$$

and

$$\lambda H(t;m_2^*,m_1^*,a,b)H(t;m_2^*,m_1,a,c)\phi_1(t)+(1-\lambda)H(t;m_2^*,m_1^*,a,b)H(t;m_2^*,m_2,a,c)\phi_2(t) = \lambda^* H(t;m_2^*,m_1^*,a,b)H(t;m_2^*,m_1^*,a,c)\phi_1^*(t), \tag{3.24}$$

yielding

$$\begin{aligned}&\lambda H(t;m_2^*,m_1,a,b)H(t;m_2^*,m_1^*,a,c)\phi_1(t)+(1-\lambda)H(t;m_2^*,m_2,a,b)H(t;m_2^*,m_1^*,a,c)\phi_2(t)\\ =\ &\lambda H(t;m_2^*,m_1^*,a,b)H(t;m_2^*,m_1,a,c)\phi_1(t)+(1-\lambda)H(t;m_2^*,m_1^*,a,b)H(t;m_2^*,m_2,a,c)\phi_2(t),\end{aligned}$$

or

$$\begin{aligned}&\lambda\left[H(t;m_2^*,m_1^*,a,b)H(t;m_2^*,m_1,a,c)-H(t;m_2^*,m_1,a,b)H(t;m_2^*,m_1^*,a,c)\right]\phi_1(t)\\ &=(1-\lambda)\left[H(t;m_2^*,m_1^*,a,b)H(t;m_2^*,m_2,a,c)-H(t;m_2^*,m_2,a,b)H(t;m_2^*,m_1^*,a,c)\right]\phi_2(t).\end{aligned} \tag{3.25}$$

Divide both sides of (3.25) by $e^{m_1^*(x_a)}$ and rewriting:

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

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

$$L_1(t)=e^{it(u_2-u_1)}\frac{1-\lambda}{\lambda}\frac{\phi_2(t)}{\phi_1(t)}L_2(t)\qquad\text{for all } t. \tag{3.26}$$

We now use the condition

$$\lim_{t\to\infty}\frac{\phi_2(t)}{\phi_1(t)}=0 \tag{3.27}$$

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

$$L_1(t)=0 \qquad \text{for all } t. \tag{3.28}$$

Suppose (3.28) 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 (3.27), together with (3.26), there exists $t_1(\epsilon)\in\mathbb{R}$ such that

$$|L_1(t)|<\frac{\epsilon}{2} \qquad \text{for all } t>t_1(\epsilon). \tag{3.29}$$

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

$$|L_1(t)-L_1(t+\xi(\tau,\epsilon,l(\epsilon)))|<\frac{\epsilon}{2} \qquad \text{for all } t\in\mathbb{R}. \tag{3.30}$$

In particular, evaluating (3.30) at $t=t_0$ and $\tau=-t_0+t_1(\epsilon)$;

$$|L_1(t_0)-L_1(t_0+\xi^*)|<\frac{\epsilon}{2} \tag{3.31}$$

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 (3.29),

$$|L_1(t_0+\xi^*)|<\frac{\epsilon}{2} \tag{3.32}$$

Using the triangle inequality, (3.31) and (3.32), 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, (3.28) is now proved.

Next, as $\lambda\neq 0$, (3.26) and (3.28) 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:

$$(1 - e^{itu_{11}})(1 - e^{itu_{12}}) - (1 - e^{itu_{13}})(1 - e^{itu_{14}}) = 0 \tag{3.33}$$

and

$$(1 - e^{itu_{21}})(1 - e^{itu_{22}}) - (1 - e^{itu_{23}})(1 - e^{itu_{24}}) = 0 \tag{3.34}$$

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

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

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

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

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

Then by (3.33) and (3.34), 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 (3.21) becomes

$$\lambda H(t; m_2^*, m_1, a, b)\phi_1(t) + (1 - \lambda)H(t; m_2^*, m_2, a, b)\phi_2(t) = 0, \tag{3.35}$$

or

$$(1 - e^{it\Delta_{ab}(m_2^* - m_1)}) + \frac{1 - \lambda}{\lambda}\frac{\phi_2(t)}{\phi_1(t)}e^{-itm_1(x_a)}H(t; m_2^*, m_2, a, b) = 0. \tag{3.36}$$

Let $t \to \infty$, then again by (3.27), 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

$$\begin{aligned} \Delta_{ab}m_2^* &= \Delta_{ab}m_1, \\ \Delta_{ac}m_2^* &= \Delta_{ac}m_1. \end{aligned} \tag{3.37}$$

On top of this, (3.34) has to hold at the same time. First, suppose all $u_{2k}, k = 1,2,3,4$ in (3.34) are non-zero. Then (C3) and/or (C4) has to hold. Suppose (C3) holds. Then

$$\begin{aligned} \Delta_{ab} m_1^* &= \Delta_{ab} m_2, \\ \Delta_{ac} m_1^* &= \Delta_{ac} m_2. \end{aligned} \tag{3.38}$$

(3.37) and (3.38) 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 (3.37) 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 (3.34) 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 (3.37), $\Delta_{ac} m_1 = \Delta_{ac} m_2$, 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 (3.37) and (3.38).

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 (3.33) and (3.34), which are equivalent to (3.34) and (3.33), 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

$$\begin{aligned} \Delta_{ab} m_1^* &= \Delta_{ab} m_1, \\ \Delta_{ac} m_1^* &= \Delta_{ac} m_1. \end{aligned} \tag{3.39}$$

and

$$\begin{aligned} \Delta_{ab} m_2^* &= \Delta_{ab} m_2, \\ \Delta_{ac} m_2^* &= \Delta_{ac} m_2. \end{aligned} \tag{3.40}$$

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): (3.37) and (3.38) hold, or (B): (3.39) and (3.40) 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 (3.17) is now written as

$$F_1(z - m_1(x)) = \lambda^* F_1^*(z - m_1^*(x)) + (1-\lambda^*)F_2^*(z - m_2^*(x)). \tag{3.41}$$

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:

$$\lambda^* = 1, \quad \sigma_1^{*2} = \sigma_1^2, \tag{3.42}$$

$$[m_1^*(x_a) - m_2^*(x_a)]^2 = [m_1^*(x_b) - m_2^*(x_b)]^2 = [m_1^*(x_c) - m_2^*(x_c)]^2. \tag{3.43}$$

(3.42) obviously leads to full identification: integrating both sides of (3.41) gives $m_1^*(x) = m_1(x)$, and this trivially determines $F_1^*(z) = F_1(z)$ for all z. (3.43) 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:

$$m_1^*(x) - m_2^*(x) = m_1^*(x') - m_2^*(x'). \tag{3.44}$$

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 (3.17). By (Step 1), in the case $\lambda \neq 1$, they have to satisfy (3.37) and (3.38), or, (3.39) and (3.40). Similarly, in the case $\lambda = 1$, they have to satisfy (3.44) (the case with (3.42) is trivial). For an arbitrary pair of points (x, x') from the three support points x_a, x_b, x_c, define

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

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

$$\max_{(x,x')=(x_a,x_b),(x_a,x_c),(x_b,x_c)} \lambda(x, x'), \tag{3.45}$$

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 (3.45), which is possibly not unique.

Now, evaluating (3.10) 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

$$\begin{bmatrix} \phi(t|\bar{x}) \\ \phi(t|\bar{x}') \end{bmatrix} = G(\bar{x}, \bar{x}', t)\Lambda \begin{bmatrix} \phi_1(t) \\ \phi_2(t) \end{bmatrix}, \tag{3.46}$$

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 (3.11) 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 (3.46) 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 (3.45)) 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 (3.27) 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 $e^{itm_2(\tilde{x})} = \frac{\phi(t|\tilde{x}) - \lambda e^{itm_1(\tilde{x})}\phi_1(t)}{(1-\lambda)\phi_2(t)}$. □

Remark 3.9. Once identification is achieved at some values of x, as implied by Lemmas 3.3 and 3.5, the complete knowledge of M_1 and M_2 is available. Since the identity for conditional characteristic functions or conditional moment generating functions as in (3.6) holds for all t, it can be used to determine m_1 and m_2 even at points where they fail to satisfy the non-parallel condition (i.e. Assumption 3.1(ii) or 3.3(ii)). Suppose $F(\cdot|x)$ is known on a set $\mathcal{X} \in \mathbb{R}^k$. Assume that, for example, Assumptions 3.1(i) and 3.2 hold. Then $F(\cdot|x), x \in \mathcal{X}$ uniquely determines $(\lambda, F_1(\cdot), F_2(\cdot), m_1(x), m_2(x))$ for all $x \in \mathcal{X}$ up to labeling, unless $\lambda = 1 - \lambda = \frac{1}{2}$ and $F_1(z) = F_2(z)$ for all $z \in \mathbb{R}$.

3.3. **Third identification result.** We now propose an identification strategy that has 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 it is the identification strategy in this section that will be used 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* $\lim_{t\to\infty} \frac{-i}{a}\mathrm{Log}\left(\frac{\rho(x,t+a)}{\rho(x,t)}\right) = m_2(x') - m_2(x_0)$.

Proof. The proof of Part (i) is essentially in the proof of Lemma (3.1). For Part (ii), note that the ratios on the right hand side of (3.12) and by (3.13) converge to 1 as $s \to \infty$. Since

$$\rho(s,x) = e^{is\nabla}\frac{\frac{\lambda}{1-\lambda}e^{is(m_1(x_1)-m_2(x_1))}\frac{\phi_1(s)}{\phi_2(s)}+1}{\frac{\lambda}{1-\lambda}e^{is(m_1(x_0)-m_2(x_0))}\frac{\phi_1(s)}{\phi_2(s)}+1},$$

and under Assumption 3.4 the ratio on the right hand side converges to 1 as $s \to \infty$. Therefore we have

$$\begin{aligned}\lim_{s\to\infty}\frac{-i}{a}\ \mathrm{Log}\left(\frac{\rho(x,s+a)}{\rho(x,s)}\right) &= \frac{-i}{a}\ \mathrm{Log}(e^{ia\nabla}) \\ &= \frac{1}{a}\left(a\nabla + 2\pi\left\lfloor\frac{1}{2} - \frac{a\nabla}{2\pi}\right\rfloor\right),\end{aligned}$$

where Log corresponds to the principal value of the log. This limit is a piecewise continuous function of a, constant equal to ∇ only when a is small enough to guarantee $a\nabla \in (-\pi,\pi)$. And if $\lambda = 1$, $\frac{\phi(s|x_1)}{\phi(s|x_0)} = e^{is\Delta}$ so that $\lim_{s\to\infty}\frac{-i}{a}\ \mathrm{Log}\left(\frac{\phi(s+a|x_1)}{\phi(s+a|x_0)}\left(\frac{\phi(s|x_1)}{\phi(s|x_0)}\right)^{-1}\right) = \frac{1}{a}\left(a\Delta + 2\pi\left\lfloor\frac{1}{2} - \frac{a\Delta}{2\pi}\right\rfloor\right)$. By assumption, $m_1(x_1) - m_1(x_0) \neq m_2(x_1) - m_2(x_0)$ that is, $\Delta \neq \nabla$ therefore if the former limit is equal to Δ, one knows $\lambda = 1$ and there is no m_2. □

The constant δ in the following condition is specified in Assumption 3.1.

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 everyx* $\in N^1(x_0,\epsilon)$ *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*) $\lim_{\delta\downarrow 0}\lambda_\delta = 1$

holds.

The above condition is verifiable with information in the observables as ρ, R and λ_δ are all observed.

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.*

Proof. Similar to the proof of Lemma 3.2.

□

Note that we once again needed the non-parallel regression function condition. Once the increments of the regression functions are identified, their levels as well as the mixture weight λ are obtained using the same procedure as in the first identification result. Thus we have:

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 now propose another method which will be used to construct our estimator and avoids Laplace inversion. The main benefit of this is that it let us nonparametrically estimate the distribution functions without resorting to empirical MGF inversion, which is hard to handle in terms of obtaining polynomial rates of convergence. We will use previous identification of λ and m_1 and m_2 evaluated at two points only, x_1 and x_0.

The idea is the following. Equation (3.3) gives $F(z|x) = \lambda F_1(z - m_1(x)) + (1 - \lambda)F_2(z - m_2(x)), \forall (x, z) \in \mathbb{R}^{k+1}$, implying $\forall (x, y) \in \mathbb{R}^{k+1}$,

$$F(m_1(x) + y|x) = \lambda F_1(y) + (1 - \lambda)F_2(m_1(x) - m_2(x) + y). \tag{3.47}$$

Applying Equation (3.47) to (x_0, y) and (x_1, y) and taking the difference, we obtain

$$\begin{aligned} &F(m_1(x_1) + y|x_1) - F(m_1(x_0) + y|x_0) = \\ &\qquad = (1 - \lambda) \left(F_2(m_1(x_1) - m_2(x_1) + y) - F_2(m_1(x_0) - m_2(x_0) + y)\right), \end{aligned} \tag{3.48}$$

which means that $\forall y \in \mathbb{R}$, $F_2(m_1(x_1) - m_2(x_1) + y) - F_2(m_1(x_0) - m_2(x_0) + y)$ is identified. Using recursively identification of this increment and the fact that the conditional cumulative distribution function F_2 converges to 1 at infinity, we obtain identification of $F_2(z)$, $\forall z \in \mathbb{R}$. Writing $g(x) =$

$m_1(x)-m_2(x)$ and $\delta(x,x')=g(x)-g(x')$, we assume that $\delta(x_1,x_0)>0$. Note that $\delta(x_1,x_0)=\Delta-\nabla$. Now, apply, for a given $z\in\mathbb{R}$, Equation (3.48) to $y=z-g(x_0)$ to obtain

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

and, more generally, $\forall j\in\mathbb{N}$,

$$\begin{aligned}F_2(z+(j+1)\delta(x_1,x_0))-F_2(z+j\delta(x_1,x_0))=\frac{1}{1-\lambda}\{&F(z+j\delta(x_1,x_0)+m_1(x_1)-g(x_0)|x_1)\\ &-F(z+j\delta(x_1,x_0)+m_2(x_0)|x_0))\}.\end{aligned}$$

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

$$\begin{aligned}F_2(z)=1-\frac{1}{1-\lambda}\sum_{j=0}^{\infty}&F(z+j\delta(x_1,x_0)+m_1(x_1)-g(x_0)|x_1)\\ &-F(z+j\delta(x_1,x_0)+m_2(x_0)|x_0)),\end{aligned}\tag{3.49}$$

where the infinite sum is a convergent series of positive terms.

Finally the equation $F(z|x)=\lambda F_1(z-m_1(x))+(1-\lambda)F_2(z-m_2(x))$ identifies $F_1(.)$ as

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

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 a completely unrestricted, 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 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 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 a fully unrestricted way, accommodating arbitrary correlation between the unobserved heterogeneity and the regressor, so it resembles a panel data fixed effects model in this aspect. In this section we 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.

We replace Assumption 3.1 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 $0 < \lambda(x_0) < 1$, $m_1(x_0) - m_1(x) \neq m_2(x_0) - m_2(x)$, for all $x \in N^1(x_0, \delta)$, $x \neq x_0$,

(iii) λ, m_1 and m_2 are continuous in x^1 at x_0.

We maintain Assumption 3.2, which, as noted before, is a weak regularity condition. Define

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

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

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

and

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

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. *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*) $K_{+\infty,t}(x) = 1$ *for every* $t \in \mathbb{R}$ *and* $x \in N^1(x_0,\delta)$

holds for some $\delta > 0$.

Lemma 4.1. *Suppose Assumptions 3.2, 4.1 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* $\lambda(x)$*, and moreover,*

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

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.*

Proof. See appendix. □

The next result shows that the model that allows λ to be arbitrarily dependent on x is nonparametrically 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.*

Proof. 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:

$$\begin{bmatrix} M(t|x) \\ M(t|x_0) \end{bmatrix} = \begin{pmatrix} \lambda(x)e^{tm_1(x)} & (1-\lambda(x))e^{tm_1(x)} \\ \lambda(x_0)e^{tm_1(x)} & (1-\lambda(x_0))e^{tm_2(x_0)} \end{pmatrix} \begin{bmatrix} m_1(x_0) \\ m_2(x_0) \end{bmatrix}, \quad \text{for every } t \in \mathbb{R}. \tag{4.3}$$

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

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

which is non-zero for almost all t under the non-parallel condition. Therefore (4.3) 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. $\square$

5. 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

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

In a manner similar to Section 3.2, consider a switching regression model:

$$y = \begin{cases} g_1(w) + \eta_1, & (y, w, x, \eta_1) \sim F_1 \qquad \text{with probability } \lambda \\ g_2(w) + \eta_2, & (y, w, x, \eta_2) \sim F_2 \qquad \text{with probability } 1-\lambda. \end{cases} \tag{5.1}$$

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

$$\int \eta dF_1(\eta|x) = \int \eta dF_2(\eta|x) = 0, \tag{5.2}$$

that is

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

Here and thereafter the notation $F_i(\star_1, \star_2, ...)$ and $F_i(\star_1, \star_2, ...|\star)$ 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 $F_i, i = 1, 2$. Consider linear operators

$$T_1[f](x) = \int f(w)dF_1(w|x), \quad T_2[f](x) = \int f(w)dF_2(w|x) \tag{5.3}$$

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 $m_1(x) = \int zdF_1(z|x)$ and $m_2(x) = \int zdF_2(z|x)$. Note that $m_1 : \mathbb{R}^k \to \mathbb{R}^{p+1}$ and $m_2 : \mathbb{R}^k \to \mathbb{R}^{p+1}$. For $j = 1, ..., p+1$, let $m_{j,1}(\cdot)$ and $m_{j,2}(\cdot)$ denote the $j-$th elements of $m_1(\cdot)$ and $m_2(\cdot)$, respectively. Define the $p+1$-dimensional vectors of random variables $\epsilon_j = z - m_i(x), (z, x) \sim F_i(z, x), j = 1, 2$. Consistent with the previous notation let $F_i(\epsilon_i|x), i = 1, 2$, denote the conditional distribution of ϵ_1 and ϵ_2 under F_1 and F_2.

By construction,

$$\int \epsilon dF_i(\epsilon|x) = 0, i = 1, 2.$$

If we further assume that $F_i(\epsilon|x), j = 1, 2$ do not depend on x, 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 (5.1), (5.2) and (5.3) 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 as . To formalize this idea, consider the following assumptions:

Assumption 5.1. *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) $m_{j1}(x_0) - m_{j1}(x) \neq m_{j2}(x_0) - m_{j2}(x)$, for all $x \in N^1(x_0, \delta)$, $x \neq x_0$ and for all j, $j = 1, ..., p+1$;
(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 require 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

$$h_j(c,t) := e^{t\mathbf{e}_j{}' D(x_0)(1-c)} \frac{M_2(t\mathbf{e}_j)}{M_1(t\mathbf{e}_j)}, \quad c \in \mathbb{R}_{++}, t \in \mathbb{R}$$

and

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

Assumption 5.2. *(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 5.1. *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, we can show that $(\lambda, F_1^\epsilon, F_2^\epsilon, m_1(x_0), m_2(x_0))$ is identified under Assumptions 5.1 and 5.2. Then, as noted in Remark 3.9, $F(z|x), x \in \mathcal{X}, z \in \mathbb{R}^p$ uniquely determines $(\lambda, F_1^\epsilon, F_2^\epsilon, m_1(x), m_2(x))$ in $\mathbb{R} \times \bar{\mathcal{F}}(\mathbb{R}^p)^2 \times \mathbb{R}^{2p}$ for all $x \in \mathcal{X}$ up to labeling. Therefore each component distribution of z is obtained by

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

We now have:

Theorem 5.1. *Suppose Assumptions 5.1, 5.2 and Condition 5.1 hold. Then $g_1(\cdot)$ and $g_2(\cdot)$ are identified.*

Remark 5.1. It is possible to further introduce flexibility into the model (5.1) by allowing unrestricted dependence between unobserved heterogeneity and the instrument x. This can be achieved by making λ in (5.1) an arbitrary function of x. Applying the results in Section 4 to identify $m_{p+1,1}(x)$, $m_{p+1,2}(x)$, $F_1(z|x)$ and $F_2(z|x)$ and proceeding as above, we recover g_1 and g_2 nonparametrically.

6. 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 $x \in \mathbb{R}^n$ consists of J components, $J \in \mathbb{N}$, as in (2.1). As before, define

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

Define also

$$\epsilon_j = z_j - m_j(x), \quad j = 1, 2, ..., J.$$

Later we impose independence between $\epsilon_j, j = 1, ..., J$ and x, which enables us to write $F(z|x)$ as

$$F(z|x) = \sum_{j=1}^{J} \lambda_j F_j(z - m_j(x)). \tag{6.1}$$

For later use, define $M_j(t) = \int e^{t\epsilon} F_j(d\epsilon), j = 1, ..., J$. This section shows that the parameter $(\{\lambda_j\}_{j=1}^J, \{F_j(\cdot)\}_{j=1}^j, \{m_j(\cdot)\}_{j=1}^J)$ 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 of $\{\epsilon_j\}_{j=1}^J$ 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 (6.1) 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 multistep identification argument with recursive procedures, as will be seen shortly.

To see how the treatment of general mixtures differs from the $J = 2$ case, consider the case $J = 3$. Instead of Equation (3.6), we now have

$$M(t|x) = \lambda_1 e^{tm_1(x)} M_1(t) + \lambda_2 e^{tm_2(x)} M_2(t) + \lambda_3 e^{tm_3(x)} M_3(t), \quad \lambda_1 + \lambda_2 + \lambda_3 = 1. \tag{6.2}$$

Wlog, suppose $m_1(x_0) > m_2(x_0) > m_3(x_0)$ at a point x_0 in $\mathbb{R}^k$. Take a point x' in the neighborhood of x_0 and consider the case $m_1(x') - m_1(x_0) \geq 0$ (if this term is negative, the roles of m_1 and m_3 get interchanged). 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. Following the proof, take the ratio of the conditional moment generating functions at x_0 and a point in its neighborhood, x', say, then take its logarithm followed by a normalization by t:

$$\frac{1}{t}\log\left(\frac{M(t|x')}{M(t|x_0)}\right) = \frac{1}{t}\log\left(\frac{\lambda_1 e^{tm_1(x')}M_1(t) + \lambda_2 e^{tm_2(x')}M_2(t) + \lambda_3 e^{tm_3(x')}M_3(t)}{\lambda_1 e^{tm_1(x_0))}M_1(t) + \lambda_2 e^{tm_2(x_0)}M_2(t) + \lambda_3 e^{tm_3(x_0))}M_3(t)}\right)$$
$$= \frac{1}{t}\log\left(\frac{e^{t[m_1(x')-m_1(x_0)]} + \frac{\lambda_2}{\lambda_1}e^{t[m_2(x')-m_1(x_0)]}\frac{M_2(t)}{M_1(t)} + \frac{\lambda_3}{\lambda_1}e^{t[m_3(x')-m_1(x_0)]}\frac{M_2(t)}{M_1(t)}}{1 + \frac{\lambda_2}{\lambda_1}e^{t[m_2(x_0)-m_1(x_0)]}\frac{M_2(t)}{M_1(t)} + \frac{\lambda_3}{\lambda_1}e^{t[m_3(x_0)-m_1(x_0)]}\frac{M_2(t)}{M_1(t)}}\right).$$

Suppose the ratios of $M_1(t)$, $M_2(t)$ and $M_3(t)$ do not explode exponentially, and m_1, m_2 and m_3 are continuous so that $m_2(x') - m_1(x_0)$ and $m_3(x') - m_1(x_0)$ are negative. Then as t approaches to infinity, the above expression approaches to the slope $m_1(x') - m_1(x_0)$ if it is non-negative (though it yields the identical result if the slope is negative as well, as seen in the proof of Lemma 3.1). 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 (6.2) by $e^{-t[m_1(x)-m_1(y)]}$ to obtain:

(6.3)
$$e^{-t[m_1(x)-m_1(y)]}M(t|x) = \lambda_1 e^{tm_1(y)}M_1(t) + \lambda_2 e^{t\{m_2(x)-[m_1(x)-m_1(y)]\}}M_2(t) + \lambda_3 e^{t\{m_3(x)-[m_1(x)-m_1(y)]\}}M_3(t).$$

This purges x out of the first term on the right hand side. Note $[m_1(x) - m_1(y)]$ can be identified by applying the argument in Lemma 3.1 to the $J = 3$ model (6.2), as demonstrated above. 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 the 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. Let D_x denote the partial differentiation operator with respect to the first component of x, i.e. $D_x f(x) = \frac{\partial}{\partial x^1}f(x)$.

Differentiating both sides of the above equation by x^1 and rearranging,

$$D_x\left[e^{-t[m_1(x)-m_1(y)]}M(t|x)\right] = t\lambda_2[D_x m_2(x) - D_x m_1(x)]e^{t\{m_2(x)-[m_1(x)-m_1(y)]\}}M_2(t) + t\lambda_3[D_x m_3(x) - D_x m_1(x)]e^{t\{m_3(x)-[m_1(x)-m_1(y)]\}}M_3(t). \tag{6.4}$$

Note that operating D_x eliminates the unknown function $M_1(t)$ out of the right hand side of (6.4). We now have

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

say, where

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

and

$$A_2 = 1 + \frac{\lambda_3}{\lambda_2}\frac{[D_x m_3(x) - D_x m_1(x)]}{[D_x m_2(x) - D_x m_1(x)]}e^{t[m_3(x)-m_2(x)]}\frac{M_3(t)}{M_2(t)}.$$

Note that the factor $D_x m_2(x) - D_x m_1(x)$ is non-zero if the two regression functions are not parallel at x, which makes the division by the factor valid. As far as $\frac{M_3}{M_2}$ and $\frac{D_x M_3}{M_2}$ 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

$$\lim_{t\to\infty}\left\{\frac{\partial}{\partial t}\log\left(D_x\left[e^{-t[m_1(x)-m_1(y)]}M(t|x)\right]\right)\right\} = \{m_2(x) - [m_1(x) - m_1(y)]\} + \frac{\frac{\partial}{\partial t}M_2(t)}{M_2(t)}. \tag{6.5}$$

The only unknown component in the above equation is $\frac{\frac{\partial}{\partial t}M_2(t)}{M_2(t)}$, but this term depends only on t, so it can be differenced out: repeat the above argument with replacing $x \in \mathbb{R}^k$ with a point $z \in \mathbb{R}^k$ so close to x that $m_1(z) > m_2(z) > m_3(z)$. This yields

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

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 terms such as $m_1(x) - m_1(z)$ on the right hand side are identified by the method developed in Lemma 3.1, as noted earlier. The equation above shows the identifiability of the slope of m_2.

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 (6.4), changing the notation and writing x_a for x, x_b for y. As before, $\Delta_{ab} f$ stands for $f(x_a) - f(x_b)$. The first step is to purge x_a from the first term on the right hand side, as we did in Equation (6.3), 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

$$D_{x_a}\left[\frac{e^{-t[\Delta_{ab} m_2 - \Delta_{ab} m_1]}}{t[D_{x_a} m_2(x_a) - D_{x_a} m_1(x_a)]} D_{x_a}\left[e^{-t\Delta_{ab} m_1} M(t|x_a)\right]\right] = \lambda_3 \left\{[D_{x_a} + t(D_{x_a} m_3(x_a) - D_{x_a} m_2(x_a))] \frac{D_{x_a} m_3(x_a) - D_{x_a} m_1(x_a)}{D_{x_a} m_2(x_a) - D_{x_a} m_1(x_a)}\right\} e^{t\{m_3(x_a) - \Delta_{ab} m_2\}} M_3(t). \tag{6.6}$$

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

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

Let us now turn to the identifiability of the general model (6.1) for a generic J, at a point $x_a \in \mathbb{R}^k$. 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 will again use local variations in x^1.

Assumption 6.1. *For some $\delta > 0$,*

(i) $\epsilon_j | x \sim F_j, j = 1, ..., J$ *at all* $x \in N^1(x_a, \delta)$ *where* $F_j, j = 1, ..., J$ *do not depend on the value of* x*;*

(ii) $m_j, j = 1, ..., J$ *are continuous in* x^1 *at* x_a*;*

(iii) $m_j, j = 1, ..., J$ *are* J *times differentiable on* $B(x_a, \delta)$ *at least in one of the* k^* *continuous covariates of* x*;*

Though Condition (iii) 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 $B(x_a, \delta)$. See Remark 6.2 for further discussions.

The independence assumption (i) enables us to write the observable conditional distribution in the form (6.1). The continuity assumption (ii) was also assumed in Lemma 3.1. The differentiability condition (iii) 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. Note that differentiability in only one element of x suffices. Without loss of generality in what follows we assume that the $m_j, j = 1, ..., J$ are differentiable in the first element x^1. Recall that D_1 is the differentiation operator with respect to x^1. From now on, we will use the notation

$$m_{k,j}(x) = m_k(x) - m_j(x).$$

Assumption 6.2. *(i)* $\min_{k \neq j} |m_{k,j}(x_a)| > \Delta$*,* $\Delta > 0$*;*
(ii) $D_1 m_j(x_a), j = 1, ..., J$ *takes* J *distinct values in* $\mathbb{R}$*;*
(iii) The domains of $M_1(t)$ *and* $M_2(t)$ *are* $(-\infty, \infty)$*;*
(iv) For some $\epsilon > 0$ *,* $\lim_{t\to\infty} e^{t(\epsilon-\Delta)} \frac{M_j(t)}{M_k(t)} = 0$ *and* $\lim_{t\to\infty} e^{t(\epsilon-\Delta)} \frac{\frac{\partial}{\partial t} M_j(t)}{M_k(t)} = 0$ *for all* $k, j = 1, ..., J$*.*

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

$$m_1(x_a) > m_2(x_a) > ... > m_J(x_a) \tag{6.7}$$

without loss of generality: this does not impact the validity of Assumption 6.2. Part (ii) is an infinitesimal version of the non-parallel regression function conditions used in the previous sections.

Under these assumptions, we first prove identifiability of the slope $\Delta_{ab} m_1$, using the method developed in Section 3.2 , for all x_b in a chosen neighborhood of x_a. Note that we know $\lambda_1 \neq 0$. By the continuity and differentiability assumptions (Assumption 6.1 (ii) and (iii)), 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 $D_1 m_j(x_b), j = 1, ..., J$ take J distinct values. Here we use the fact that twice differentiability of the regression functions implies that they are $\mathcal{C}^1$. Then, as in the proof of Lemma 3.1, in the case $m_1(x_b) - m_1(x_a) > 0$, we write

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

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,

$$\forall x_b \in U, \lim_{t\to\infty}\frac{1}{t}\log\left(\frac{M(t|x_b)}{M(t|x_a)}\right) = \Delta_{ba}m_1.$$

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

Now we 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 $\mathbb{R}^k$. For notational convenience, define the operator $A(x_a, x_b, t, k)$

$$A(x_a, x_b, t, k)(f)(x_a) = \frac{\partial}{\partial x_a^1}\left[\frac{e^{-t[\Delta_{ab}m_k - \Delta_{ab}m_{k-1}]}}{R_k(t, x_a)}f(x_a)\right], k = 2, 3, ..., J. \tag{6.8}$$

where $f : \mathbb{R}^k \to \mathbb{R}$ is a function that is differentiable in its first argument, and $R_k(t, x)$ is a (rational) function in t. Its precise definition will be given shortly. The operator $A(x_a, x_b, t, k)$ generalizes the procedure performed on $D_{x_a}\left[e^{-t\Delta_{ab}m_1}M(t|x_a)\right]$ in Equation (6.6) to eliminate unknown parameters in (6.4). Operate $A(x_a, x_b, t, k)$, $k = 2, 3, ...$ sequentially on $D_{x_a}\left[e^{-t\Delta_{ab}m_1}M(t|x_a)\right]$ to define the expressions

$$Q_k(x_a, t) = A(x_a, x_b, t, k-1)A(x_a, x_b, t, k-2)\cdots A(x_a, x_b, t, 2)\frac{\partial}{\partial x_a^1}[e^{-t\Delta_{ab}m_1}M(t|x_a)], \quad k = 2, 3, ..., J. \tag{6.9}$$

By construction $Q_k(x_a, t)$ satisfies the following recursive formula:

$$Q_{k+1}(x_a, t) = A(x_a, x_b, t, k)Q_k(x_a, t), \quad Q_2(x_a, t) = \frac{\partial}{\partial x_a^1}[e^{-t\Delta_{ab}m_1}M(t|x_a)]. \tag{6.10}$$

The definition of the operator $A(x_a, x_b, t, k)$, as explained further later, is motivated by two facts: (i) the factor $e^{-t[\Delta_{ab}m_k - \Delta_{ab}m_{k-1}]}$ purges x_a out of the exponent in the leading term of $Q_k(x_a, t)$ and (ii) division by the polynomial $R_k(t, x_a)$ then makes the leading term $\lambda_k e^{-tm_k(x_b)}M_k(t)$, which is completely free from x_a and therefore eliminated by D_{x_a}. Once this is done, taking the log-derivative with respect to t as in (6.5) terms and taking the limit $t \to \infty$ yields $\Delta_{ab}m_k$ up to an unknown additive factor $\frac{\frac{\partial}{\partial t}M_k(t)}{M_k(t)}$, which can be differenced out.

Subsequent arguments establish the identifiability of $\Delta_{ab}m_k, k = 2, ..., J$ for all x_b in a neighborhood of x_a. We proceed in two steps. Step 1 shows that, with an appropriate choice of $R_k(t, x_a)$ in (6.8), $Q_k(x_a, t), k = 2, 3, ..., J$ have following representations:

$$Q_k(x_a, t) = \sum_{j=k}^{J} \lambda_j R_k^j(t, x_a) e^{t[m_j(x_a) - \Delta_{ab}m_{k-1}]} M_j(t), \quad k = 2, 3, ..., J, \tag{6.11}$$

where $R_k^j(t, x_a), k = 2, 3..., J, j = k, k+1, ..., J$ are polynomials in t with the property that $R_k^k(t, x_a) = R_k(t, x_a)$; a formal definition of these polynomials are provided later. The representations (6.11) are useful, partly because the unknown functions $M_j(t), j = 1, .., k-1$ do not appear in $Q_k(x_a, t)$. Step 2 uses the representations (6.11) to show that it is possible to identify the slope $\Delta_{ab}m_k, k = 2, ..., J$ using the knowledge of $\Delta_{ab}m_1$, $Q_k(x_a, t)$ and $Q_k(x_b, t)$, $k = 2, ..., J$ for all x_b in a neighborhood of x_a.

The identifiability of the rest of the model (at x_a) is then established using the knowledge of $\Delta_{ab}m_k, k = 1, 2, ..., J$ and conditional moments of z given x_a.

Let us start with Step 1, which derives the representation (6.11) and will be summarized in Lemma 6.1. Note that the definitions of the polynomials $R_k(t, x_a), k = 2, ..., J$ and $R_k^j(t, x_a), k = 2, ..., J, j = k, k+1, ..., J$ are given in the course of our derivation.

Step 1: Start from $k = 2$. Define

$$R_2^j(t, x_a) = tD_{x_a}(m_j(x_a) - m_1(x_a)), j = 2, ..., J,$$

then

$$\begin{aligned} Q_2(x_a, t) &= \frac{\partial}{\partial x_a^1}[e^{-t\Delta_{ab}m_1} M(t|x_a)] \\ &= \sum_{j=2}^{J} \lambda_j (tD_{x_a}[m_j(x_a) - m_1(x_a)]) e^{t[m_j(x_a) - \Delta_{ab}m_1]} M_j(t). \\ &= \sum_{j=2}^{J} \lambda_j R_2^j(t, x_a) e^{t[m_j(x_a) - \Delta_{ab}m_1]} M_j(t), \end{aligned}$$

yielding the desired representation for the case of $k = 2$. Let $R_2(t, x_a)$ (used in the definition of $A(x_a, x_b, t, 2)$) be $R_2^2(t, x_a) = tD_{x_a}[m_2(x_a) - m_1(x_a)]$. With this choice

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

and the $j = 2$ term in the summation drops out. Moreover, this result implies that $R_3(x_a, t)$ should be

$$R_3(x_a, t) = R_3^3(x_a, t) = D_{x_a}\frac{R_2^3(t, x_a)}{R_2(t, x_a)} + t\frac{R_2^3(t, x_a)}{R_2(t, x_a)}D_{x_a}[m_3(x_a) - m_2(x_a)].$$

Note that the above step requires that $R_2(t, x_a)$ is non-zero: this issue will be discussed shortly.

The fact that the rest of $Q_k(x_a, t), k = 4, ..., J$ have the representations as in (6.11) can be shown by induction: suppose (6.11) holds for $k = h$, that is

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

Define

$$R_{h+1}^j(t, x_a) = D_{x_a^1}\left(\frac{R_h^j(t, x_a)}{R_h^h(t, x_a)}\right) + t\frac{R_h^j(t, x_a)}{R_h^h(t, x_a)}D_{x_a^1}[m_j(x_a) - m_h(x_a)], \qquad j = h+1, ..., J.$$

In what follows we sometimes write

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

and

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

as short hand. Let $R_h(t,x_a) = R_h^h(t,x_a)$, then using this and the definition of the operator $A(x_a,x_b,t,h)$ in (6.8), obtain

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

which is the desired result. The next lemma summarizes the foregoing argument. Notice that it relies on the assumption that $R_k(t,x_a) = R_k^k(t,x_a), k = 2,3,...J$ are non-zero, and later we show that the set

$$S(x_a) = \{t|R_k(t,x_a) \neq 0 \text{ for all } k\} \tag{6.12}$$

is non-empty.

Lemma 6.1. *Define $R_2^j(t,x_a) = tD_{x_a}(m_j(x_a)-m_1(x_a)), j = 2,...,J$, 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 $R_k(t,x_a) = R_k^k(t,x_a), k = 2,...,J$ in (6.8). Then $Q_k(x_a,t) = A(x_a,x_b,t,k-1)A(x_a,x_b,t,k-2)\cdots A(x_a,x_b,t,2)D_{x_a^1}[e^{-t\Delta_{ab}m_1}M(t|x_a)], k = 2,...,J$ have the representations (6.11) on $S(x_a)$.*

Step 2: This step shows that the knowledge of the function $Q_k(x,t)$ at $x = x_a$ and $x = x_b$ identifies $\Delta_{ab}m_k - \Delta_{ab}m_{k-1}$. The main result is:

Lemma 6.2. *$\forall x_b \in N^1(x_a,\delta')$,*

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

Lemmas 6.1 and 6.2 will then be useful to prove the identifiability of $\Delta_{ab}m_k, k = 2,...,J$, for all x_b in a neighborhood of x_a, since we already identified $\Delta_{ab}m_1$. The following propositions are useful in proving Lemma 6.2. 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.

Proposition 6.1. *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

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

The proof of the proposition is given in the Appendix.

Remark 6.1. The formula for $R_k(t,x)$ given in Proposition 6.1 and the fact that $P_k(t,x)$ is a polynomial in t imply that $R_k \neq 0$ for sufficiently large t for $k = 2,3,...,J$. Consequently $S(x_a)$ in (6.12) includes (for example) the set $[c, \infty)$ for some constant c and therefore it is not empty. This is important in applying Lemma 6.1.

Proposition 6.2.

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

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

Proof of Proposition 6.2. By the expression of $R_k(x,t)$ given in Proposition 6.1,

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

Since the Proposition shows that $R_k(x,t)$, $P_k(x,t)$ and $P_{k-1}(x,t)$ are well defined for large t, so are the above limits. But Proposition 6.1 also implies that $P_k(t,x)$ and $P_{k-1}(t,x)$ are polynomials in t with finite degree, therefore the two terms are zero. $\square$

Now we are ready to prove the main result in Step 2, that is, Lemma 6.2.

Proof of Lemma 6.2. By Lemma 6.1 and Proposition 6.1,

$$Q_k(x_a,t) = \sum_{j=k}^{J} \lambda_j R_k^j(t,x_a) e^{t[m_j(x_a) - \Delta_{ab} m_{k-1}]} M_j(t), \quad k = 2,3,...,J, \tag{6.13}$$

holds for sufficiently large t. Then

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

and, for $k \leq J$,

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

Using the notation in the proof of Proposition 6.1, for all $h > k$,

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

As noted in the Proof of Proposition 6.1, 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:

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

For a similar reason, using Proposition 6.2,

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

Then, using Assumption (ii) (iv), since $m_{h,k}(x_a) < -\Delta$, we know that the second and third lines of the expression of converge to zero as t goes to $+\infty$, and we have

$$\lim_{t\to\infty}\frac{\partial}{\partial t}\log(Q_k(x_a,t)) = m_k(x_a) - \Delta_{ab}m_{k-1} + \frac{\frac{\partial}{\partial t}M_k(t)}{M_k(t)}. \tag{6.14}$$

Note that 6.14 holds for all $x_b \in \mathbb{R}^k$. Let us take $x_b \in N^1(x_a, \delta')$. Note that we can then also write $\frac{\partial}{\partial t}\log(Q_k(x,t))$ taking $x = x_b$: the $\Delta_{ab} m_h$ terms are equal to 0 and, again since $m_{h,k}(x_b)$ is less than $\epsilon - \Delta$, we have

$$\lim_{t\to\infty} \frac{\partial}{\partial t} \log(Q_k(x_b,t)) = m_k(x_b) + \frac{\frac{\partial}{\partial t} M_k(t)}{M_k(t)},$$

so that, for all $x_b \in N^1(x_a, \delta')$, we have then

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

□

To sum up, Lemma 6.2 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 $k = 1, ..., J$:

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

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

We now state the complete identification result. For the sake of clarity, we name the point of identification x_0 instead of x_a.

Assumption 6.3. *There exists $X = (x_1, ..., x_{J-1}) \in N^1(x_0, \delta')^{J-1}$ such that*

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

is invertible.

In the above assumption, the notation $\Delta_{0,i} m_j$ denotes $m_j(x_0) - m_j(x_i)$.

Lemma 6.3. *Suppose Assumptions 6.1, 6.2 and 6.3 hold. Then $F(\cdot|x), x \in B(x_0, \delta')$ 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.*

Proof of Lemma 6.3. Reproducing what was done 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

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

As Assumption 6.3 guarantees the invertibility of $A(x_0, X)$, and since the slopes of the $(m_j)_{j=1..J}$ were all previously identified, the $(\lambda_j)_{j=1..J-1}$ are identified with the formula

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

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

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

used in the Proof of Lemma 3.3, where we can show that

$$C(x_k) = \sum_{i=1}^{J} \lambda_i \; m_i(x_0) \; \Delta_{0,k} m_i,$$

which 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

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

$diag(\lambda_1, ..., \lambda_J)$ is invertible as λ_j, $j = 1..J$ are assumed to be nonzero. Since $\det B(x_0, X) = \det A(x_0, X)$, $B(x_0, X)$ is invertible. Therefore, we obtain the following identification result:

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

What now remain to be identified are the $(F_j(\cdot))_{j=1..J}$: we will again use a technique similar to what was done in the proof of Lemma 3.3, but using Assumption 6.3. As $M(t|x) = \sum_{i=1}^{J} \lambda_i e^{tm_i(x)} M_i(t)$, considering J generic points $(c_i)_{i=1..J} \in B(x_0, \delta')^J$, we have

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

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

We prove in the appendix (Proposition 10.1) that there is a vector of $(J-1)$ points $X^{(J)} = (x_1^{(J)}, ..., x_{J-1}^{(J)}) \in B(x_0, \delta')^{J-1}$, such that $\mathcal{Z} = \left\{ t \in \mathbb{R} | \det D(t, x_0, x_1^{(J)}, ..., x_{J-1}^{(J)}) = 0 \right\}$ is finite. Hence, we can invert $D(t, x_0, x_1^{(J)}, ..., x_{J-1}^{(J)})$ for all $t \in \mathbb{R} \backslash \mathcal{Z}$. Note that we can write

$$D(t, x_0, x_1^{(J)}, ..., x_{J-1}^{(J)}) = e^{t \sum_{i=1}^{J} m_i(x_0)} \begin{pmatrix} 1 & \dots & 1 \\ e^{-t\left(\Delta_{0,1} m_1 + \sum_{i=2}^{J} m_i(x_0)\right)} & \dots & e^{-t\left(\Delta_{0,1} m_J + \sum_{i=1}^{J-1} m_i(x_0)\right)} \\ \vdots & \ddots & \vdots \\ e^{-t\left(\Delta_{0,J-1} m_1 + \sum_{i=2}^{J} m_i(x_0)\right)} & \dots & e^{-t\left(\Delta_{0,J-1} m_J + \sum_{i=1}^{J-1} m_i(x_0)\right)} \end{pmatrix},$$

and since $(x_1^{(J)}, ..., x_{J-1}^{(J)}) \in B(x_0, \delta')^{J-1}$, by the above result and Lemma 6.2, $D(t, x_0, x_1^{(J)}, ..., x_{J-1}^{(J)})$ is identified. Therefore $(M_i(t))_{i=1..J}$ are identified for all $t \in \mathbb{R} \backslash \mathcal{Z}$ and since the $(M_i(t))_{i=1..J}$ have domain $(-\infty, +\infty)$, we know that they are continuous (see, e.g, Gut (2013) Theorem 8.3 p190) on $\mathbb{R}$. As for each M_i, there is a unique continuous extension on $\mathbb{R}$ of its restriction to $\mathbb{R} \backslash \mathcal{Z}$, the J functions are identified. By the same argument of uniqueness of the Laplace transform for a distribution function, this leads to the identification of the F_i. $\square$

Having showed identification of our model assuming knowledge of J, we now consider the case where J is unknown, and show it is identified, using the observable sequence of functions $(Q_k)_{k=1,...}$. As we see below, the number of mixture components J is equal to the largest j for which the function Q_j not identically 0 in t. Therefore one can sequentially compute the $\Delta_{ab} m_j$ using Q_j, for increasing j. Once there exists j_0 such that $Q_{j_0} = 0$, then $J = j_0 - 1$.

Proposition 6.3.

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

Proof of Proposition 6.3.

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

therefore

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

We actually see that we cannot calculate any $\Delta_{ab} m_{J+1}$ with the method of Lemma 6.2 because of the logarithm: the identification process must be stopped here.

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 proposition 6.1, we know that $\deg_t R_{j_0}^j = 1$, so there is a constant $b_{x_a, x_b, j, j_0} > 0$ such that

$$\frac{R_{j_0}^j(t, x_a)}{R_{j_0}(t, x_a)} \xrightarrow[t \to \infty]{} b_{x_a, x_b, j, j_0}.$$

Using Assumption 6.2 (iv), since $m_{h,k}(x_a) < -\Delta$, 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.

$\square$

Remark 6.2. Note that it is not essential for our identification strategy to assume to impose Assumption 6.1 (iii) m is J-times differentiable in one argument, as stated right after the assumption. Note that 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 recursive formula (6.10) time we use $\frac{\partial}{\partial x^\ell}$ with a different $\ell\{1, ..., k\}$ instead of keeping on using the same differential operator $\frac{\partial}{\partial x^1}$ as in the current proof. What we need is, as noted before, that m can be differentiated up to a J-th order multi-index. This is less stringent than Assumption 6.1 (iii), though we chose to state the result in the current form for notational simplicity.

7. Application to Identifiability of Auction Models with Unobserved Heterogeneity

It is of great interest 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 properly taking account for unobserved heterogeneity in empirical auction models can lead to grossly misleading policy implications and counterfactual analyses. The reader is referred to Haile and Kitamura (2018) 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 Milgrom and Weber (1982). Suppose that valuations have the following multiplicative form, with J unknown types of auctions

$$V^k = \Gamma_j(x) U_j^k \quad \text{with probability } \lambda_j, 1 \leq j \leq J \tag{7.1}$$

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 Haile, Hong, and Shum (2003). Let B^k denote the bid of bidder k. The observables in this application is $(I, B^1, ..., B^I, x)$. The rest remain unobserved.

We maintain that there are finite number of types in terms of auction heterogeneity. It is then possible to establish identification under quite weak assumptions. In the following result note that (i) valuations can be affiliated, and (ii) unobserved heterogeneity 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. 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 Compiani, Haile, and Sant'Anna (2018). For example, Property (i) implies that the result in this section applies to the common values model. Property (ii) about the flexible treatment of homogenized valuations is apparently new.

Assume

$$(U_j^1, \ldots, U_j^I, S^1, \ldots, S^I) \perp\!\!\!\perp x | I \tag{7.2}$$

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 $w^k := w(S^k, I, x, j), 1 \le k \le I$ 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. The goal here is to identify the joint distribution of $(w^1, ..., w^I)$ in an auction with (x, I, j), along with the distribution $(\lambda_1, ..., \lambda_J)$ of the unobserved heterogeneity. Note that such knowledge is sufficient to address important questions often asked in practice: see, for example, footnote 9 of Haile and Kitamura (2018) for further discussions.

The above setting implies an expression of w of the following form

$$w\left(S, I, x, j\right) = \Gamma_j(x)\omega\left(S, I, j\right), \tag{7.3}$$

where $\omega\left(S; I, j\right) = E[V^k|S^k = \max_{i\neq k} S^i = S, I, j]$. Like the homogenized valuation $\{\{U_j^k\}_{k=1}^I\}_{j=1}^J$, $\omega_j^k := \omega\left(S^k, I, j\right)$ 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 (7.1), hence (7.3), for each bidder k. Thus we obtain

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

where R_j^k is the homogenized valuation 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

$$b^k = \gamma_j(x) + r_j^k, \quad 1 \le j \le J, \quad 1 \le k \le I. \tag{7.4}$$

Note that (7.2) implies

$$(r_k^1, \dots, r_k^I) \perp\!\!\!\perp x \quad \text{for every } j \tag{7.5}$$

conditional on I.

We now invoke Lemma 6.3 to establish identification of this model. 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. 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 $r_j(c) := \sum_{k=1}^I c_i r_j^k$. By (7.4) and the finite mixture structure of the evaluation in (7.1) we have

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

where $r(c) \perp\!\!\!\perp x$ by (7.5). 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 6.3, 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 $j \in \{1, ..., J\}$. Moreover, we now know $\gamma_j(\cdot)$, $j \in \{1, ..., J\}$ since $C(c)$ is known. Note that for each j, the marginal distribution of every linear combination $r_j(c)$ of the I-vector $(r_j^1, ..., r_j^I)$ is identified as $c \in \mathbb{R}^I$ can be chosen arbitrarily. Then by Cramér-Wold the joint distribution of $(r_j^1, ..., r_j^I)$ is obtained for each j. Apply this and the knowledge of γ_j to equation (7.4) to determine the joint distribution $(b^i, ..., b^I)|x, j, I$. Using the first order condition for equilibrium bidding (see, e.g. Haile, Hong, and Shum (2003), Athey and Haile (2007) and Equation (2.4) in Haile and Kitamura (2018)) we can now back out the joint distribution of $(w^1, ..., w^I)|x, j, I$ as desired. Note that the number of (unobserved) auction types J is also identified by Proposition 6.3.

8. Nonparametric estimation for $J = 2$

This section develops a fully nonparametric estimation procedure based on our third identification result in Section 3.3 where the number of mixture components is two. We first estimate the slopes of m_1 and m_2 nonparametrically. Define $\Delta = m_1(x_1) - m_1(x_0)$ and $\nabla = m_2(x_1) - m_2(x_0)$. Let us reintroduce notations. We write, for $j = 1, 2$,

$$\phi^j(s) = \mathbb{E}(e^{isZ}|X = x_j),\ \phi_l(s) = \mathbb{E}(e^{is\epsilon_l}) = \int e^{i\epsilon s} \mathrm{d}F_l(\epsilon),\ \hat{\phi}^j(s) = \frac{\sum_{p=1}^n e^{isZ_p} K(\frac{X_p - x_j}{b_n})}{\sum_{p=1}^n K(\frac{X_p - x_j}{b_n})},$$

$$M^j(t) = \mathbb{E}(e^{tZ}|X = x_j),\ M_i(t) = \mathbb{E}(e^{t\epsilon_i}) = \int e^{\epsilon t} \mathrm{d}F_i(\epsilon),\ \hat{M}^j(t) = \frac{\sum_{p=1}^n e^{tZ_p} K(\frac{X_p - x_j}{h_n})}{\sum_{p=1}^n K(\frac{X_p - x_j}{h_n})},$$

where F_i is the cumulative distribution function of ϵ_i, h_n and b_n are carefully chosen bandwidths for kernel density estimation. $\hat{M}^j(t)$ and $\hat{\phi}^j(s)$ are the Nadaraya-Watson regression estimators of respectively the conditional moment generating function and conditional characteristic function of Z, when $X = x_j$. X being a vector, the kernel function K can have a product form such as $K(X) = \Pi_{l=1}^k k(X^{(l)})$.

Our estimators are

$$\hat{\Delta} = \frac{1}{t_n} \log\left(\frac{\hat{M}^1(t_n)}{\hat{M}^0(t_n)}\right),$$

$$\hat{\nabla} = \frac{-i}{a_n} \operatorname{Log}\left(\frac{\hat{\phi}^1(s_n + a_n)}{\hat{\phi}^0(s_n + a_n)} \left(\frac{\hat{\phi}^1(s_n)}{\hat{\phi}^0(s_n)}\right)^{-1}\right),$$

where $(a_n)_n, (s_n)_n$ and $(t_n)_n$ are tuning parameters such that $a_n \to 0$, $s_n \to \infty$ and $t_n \to \infty$. The notation $\operatorname{Log}(\cdot)$ as before corresponds to the principal value of the logarithm of $\cdot$.

We enumerate here the assumptions on the kernel function needed to compute the rates of our estimators.

Assumption 8.1. *The kernel function $K(.)$ must satisfy the following conditions,*

$\int |K(U)|\,\mathrm{d}U < \infty$, $\int K(U)\,\mathrm{d}U = 1$, $\lim_{||U|| \to \infty} UK(U) \to 0$,
$\int K(U)^2\,\mathrm{d}U < \infty$, $\int |K(U)|\,U'U\,\mathrm{d}U < \infty$, $\int K(U)\,U\,\mathrm{d}U = 0$,
$\exists \alpha_0, \alpha \leq \alpha_0 \Rightarrow \int e^{\alpha ||U||}|K(U)|\,U'U\,\mathrm{d}U < \infty, \int e^{\alpha||U||}K(U)^2\,\mathrm{d}U < \infty$.

We need the following assumptions on the model parameters.

Assumption 8.2. *(i) f_X, the density of the random variable X, has continuous second order partial derivatives. f_X and all its first and second order partial derivatives are bounded on $\mathbb{R}^k$. $f_X(x_i) > 0$, for $i = 0, 1$.*

(ii) m_i, $i = 1, 2$ have continuous second order partial derivatives, and all their first and second order partial derivatives are bounded on $\mathbb{R}^k$.

(iii) $h_n \underset{n\to\infty}{\to} 0$, $nh_n^k \underset{n\to\infty}{\to} \infty$, and $b_n \underset{n\to\infty}{\to} 0$, $nb_n^k \underset{n\to\infty}{\to} \infty$,

(iv) $t_n \underset{n\to\infty}{\to} \infty$, $t_n h_n \underset{n\to\infty}{\to} 0$, and $s_n \underset{n\to\infty}{\to} \infty$, $s_n b_n \underset{n\to\infty}{\to} 0$.

Assumption 8.3. *(i) $\epsilon_1 | x \sim F_1$ and $\epsilon_2 | x \sim F_2$ at all $x \in \mathbb{R}^k$ where F_1 and F_2 do not depend on the value of x,*

(ii) The domains of $M_1(t)$ and $M_2(t)$ are $[0, \infty)$,

(iii) $\forall \epsilon > 0$, $e^{\epsilon t} \frac{M_2(t)}{M_1(t)} \underset{t\to\infty}{=} O(\mu(t))$, holds for some $\mu(\cdot)$, where $\mu(t) \xrightarrow[t\to\infty]{} 0$,

(iv) $\frac{\phi_1(s)}{\phi_2(s)} \underset{s\to\infty}{=} O(f(s))$, holds for some $f(\cdot)$, where $f(t) \xrightarrow[t\to\infty]{} 0$.

Proposition 8.1. *Suppose Assumptions 8.1, 8.2 and 8.3 hold.*

Then

(i) $\hat{\Delta} - \Delta = O_{\mathbb{P}}\left[\frac{\mu(t_n)}{t_n} + \frac{1}{t_n}\left((t_n h_n)^4 + \frac{1}{nh_n^k}\frac{M_1(2t_n)}{M_1(t_n)^2}\right)^{\frac{1}{2}}\right]$, where we assume $\frac{1}{nh_n^k}\frac{M_1(2t_n)}{M_1(t_n)^2} \underset{n\to\infty}{\to} 0$

(ii) $\hat{\nabla} - \nabla = \frac{1}{a_n} O_{\mathbb{P}}\left[f(s_n + a_n) + f(s_n) + \left((b_n s_n)^4 + \frac{1}{nb_n^k|\phi_2(s_n+a_n)|^2}\right)^{1/2} + \left((b_n s_n))^4 + \frac{1}{nb_n^k|\phi_2(s_n)|^2}\right)^{1/2}\right]$

Proof 1 part 1.

Proof of Proposition 8.1 i. The estimator can be decomposed as

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

The first term in the decomposition is deterministic. Using the proof of Lemma 3.9, this approximation error can be written

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

where the last equality holds using Assumption 8.3 (iii). This gives

$$\frac{1}{t_n}\log\left(\frac{M^1(t_n)}{M^0(t_n)}\right) = \Delta + O(\frac{\mu(t_n)}{t_n}).$$

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

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

and have

$$\frac{\hat{M}^j(t_n)}{M^j(t_n)} = \frac{\hat{N}^j(t_n)}{\hat{D}^jM^j(t_n)} = \frac{f_X(x_j)}{\hat{D}^j}\,\frac{\hat{N}^j(t_n)}{f_X(x_j)M^j(t_n)}. \tag{8.1}$$

In what follows, we treat separately the two ratios appearing in the last equality in (8.1), showing that they both converge to 1. Part of the reasoning will be different from usual kernel regression. Indeed,for the second ratio, we need to keep the denominator to compute the convergence rate to counterbalance the numerator going to infinity, as the parameter t_n goes to infinity.

Under Assumptions 8.1 and 8.2, we know from usual results on kernel density estimation that when computing the Mean Square Error of the term $\frac{\hat{D}^j}{f_X(x_j)}$, the bias is of order h_n^2 and the variance of order $\frac{1}{nh_n^k}$, so that

$$\frac{\hat{D}^j}{f_X(x_j)} = 1 + O_{\mathbb{P}}\left(h_n^4+\frac{1}{nh_n^k}\right)^{\frac{1}{2}}. \tag{8.2}$$

As for the second ratio in the decomposition of (8.1) , the dependence in t_n requires new assumptions when computing bias and variance. For the bias term we denote $G_n(x) = f_X(x)\mathbb{E}(e^{t_nZ}|X=x)$, then by definition of the estimator,

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

By Assumption 8.2, G_n is twice continuously differentiable. Since the kernel is of order 2 (Assumption 8.1), by virtue of the Mean Value Theorem, we have

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

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. We know that $G_n(x) = f_X(x)[\lambda e^{t_n m_1(x)} M_1(t_n) + (1-\lambda) e^{t_n m_2(x)} M_2(t_n)]$. Twice differentiation gives

$$\begin{aligned}
\nabla^2 G_n(x) = &\lambda e^{t_n m_1(x)} M_1(t_n)\ \{t_n^2 f_x(x) \nabla m_1(x) \nabla m_1(x)' \\
&+ t_n(\nabla m_1(x) \nabla f_X(x)' + \nabla f_X(x) \nabla m_1(x)' + f_X(x) \nabla^2 m_1(x)) \\
&+ \nabla^2 f_X(x)\} \\
&+ (1-\lambda) e^{t_n m_2(x)} M_2(t_n)\ \{t_n^2 f_x(x) \nabla m_2(x) \nabla m_2(x)' \\
&+ t_n(\nabla m_2(x) \nabla f_X(x)' + \nabla f_X(x) \nabla m_2(x)' + f_X(x) \nabla^2 m_2(x)) \\
&+ \nabla^2 f_X(x)\}. \\
= &\lambda\, e^{t_n m_1(x)} M_1(t_n) \{t_n^2 a_1(x) + t_n b_1(x) + c_1(x)\} \\
&+ (1-\lambda)\, e^{t_n m_2(x)} M_2(t_n) \{t_n^2 a_2(x) + t_n b_2(x) + c_2(x)\}.
\end{aligned}$$

By boundedness of the first order partial derivatives of m_i, $i = 1, 2$,

$$\exists \delta, \forall (x, U) \in \mathbb{R}^k \times \mathbb{R}^k, |m_i(x + h_n \tau_n(U) U) - m_i(x)| \leq \delta h_n ||U||,$$

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 $C \leq \alpha_0$, for n large enough, under Assumption 8.2 (iv). The same holds for the $(1-\lambda)$ term. By Assumption 8.2, $a_1(x + h_n \tau_n(U) U)$ is bounded by a constant as well as the other coefficients of the t_n polynomial in the expression of $\nabla^2 G_n[x_j + h_n \tau_n(U) U]$. This, together with the previous argument, implies that $\frac{1}{G_n(x_j)} \int U'.\nabla^2 G_n[x_j + h_n \tau_n(U) U].U K(U) \mathrm{d}U = O(t_n^2)$. The rate of the bias term can therefore be bounded,

$$\mathbb{E}\left(\frac{\hat{N}^j(t_n)}{f_X(x_j) M^j(t_n)}\right) - 1 = O(t_n h_n)^2.$$

For the variance term, an upper bound is

$$
\begin{aligned}
&\frac{1}{n}\mathbb{E}\left[\left(\frac{1}{h_n^k M^j(t_n) f_X(x_j)} e^{t_n Z} K(\frac{X-x_j}{h_n})\right)^2\right] \\
&= \frac{1}{n[h_n^k M^j(t_n) f_X(x_j)]^2}\int \mathbb{E}(e^{2t_n Z}|X) K(\frac{X-x_j}{h_n})^2 f_X(X)\mathrm{d}X \\
&= \frac{1}{nh_n^k}\int \frac{\mathbb{E}(e^{2t_n Z}|h_n U + x_j)}{\mathbb{E}(e^{t_n Z}|x_j)^2}\frac{f_X(h_n U + x_j)}{f_X(x_j)^2} K(U)^2 \mathrm{d}U \\
&= \frac{1}{nh_n^k}\int \frac{\lambda e^{2t_n m_1(h_n U + x_j)} M_1(2t_n) + (1-\lambda) e^{2t_n m_2(h_n U + x_j)} M_2(2t_n)}{\left(\lambda e^{t_n m_1(x_j)} M_1(t_n) + (1-\lambda) e^{t_n m_2(x_j)} M_2(t_n)\right)^2}\ \frac{f_X(h_n U + x_j)}{f_X(x_j)^2} K(U)^2 \mathrm{d}U \\
&\leq \frac{1}{nh_n^k}\frac{M_1(2t_n)}{M_1(t_n)^2}\int e^{2\delta t_n h_n ||U||}\frac{\lambda + (1-\lambda) e^{2t_n (m_2(x_j) - m_1(x_j)}\frac{M_2(2t_n)}{M_1(2t_n)}}{\left(\lambda + (1-\lambda) e^{t_n (m_2(x_j) - m_1(x_j))}\frac{M_2(t_n)}{M_1(t_n)}\right)^2}\ \frac{f_X(h_n U + x_j)}{f_X(x_j)^2} K(U)^2 \mathrm{d}U.
\end{aligned}
$$

Using Assumption 8.2 (iv) and Assumption (8.3) for n large enough, the integrand is bounded above by $C' e^{C||U||} K(U)^2$, $\forall U \in \mathbb{R}^k$, for some C independent of n, $C \leq \alpha_0$, $C' > 0$. Assumption (8.1) and (8.2) guarantee that the variance is of order $O(\frac{1}{nh_n^k}\frac{M_1(2t_n)}{M_1(t_n)^2})$. Therefore,

$$
\frac{\hat{N}^j(t_n)}{f_X(x_j) M^j(t_n)} = 1 + O_{\mathbb{P}}\left((t_n h_n)^4 + \frac{1}{nh_n^k}\frac{M_1(2t_n)}{M_1(t_n)^2}\right)^{\frac{1}{2}}. \tag{8.3}
$$

With (8.1), (8.2) and (8.3), and given that by Jensen's inequality $\frac{M_1(2t_n)}{M_1(t_n)^2} \geq 1$, the second ratio in (8.1) dominates. $\frac{\hat{M}^j(t_n)}{M^j(t_n)} - 1 = O_{\mathbb{P}}\left((t_n h_n)^4 + \frac{1}{nh_n^k}\frac{M_1(2t_n)}{M_1(t_n)^2}\right)^{\frac{1}{2}}$.

This finally gives

$$
\begin{aligned}
\hat{\Delta} - \Delta &= O(\frac{\mu(t_n)}{t_n}) + \frac{1}{t_n}\log(\left(1 + O_{\mathbb{P}}\left((t_n h_n)^4 + \frac{1}{nh_n^k}\frac{M_1(2t_n)}{M_1(t_n)^2}\right)^{\frac{1}{2}}\right)^2) \\
&= O_{\mathbb{P}}\left[\frac{\mu(t_n)}{t_n} + \frac{1}{t_n}\left((t_n h_n)^4 + \frac{1}{nh_n^k}\frac{M_1(2t_n)}{M_1(t_n)^2}\right)^{\frac{1}{2}}\right],
\end{aligned}
$$

since we assumed that $\frac{1}{nh_n^k}\frac{M_1(2t_n)}{M_1(t_n)^2} \underset{n\to\infty}{\to} 0$. □

Proof 1 part 2.

Proof of Proposition 8.1 ii. The estimator is $\hat{\nabla} = \frac{-i}{a}\ \mathrm{Log}\left(\frac{\hat{\phi}^1(s_n+a)}{\hat{\phi}^0(s_n+a)}\left(\frac{\hat{\phi}^1(s_n)}{\hat{\phi}^0(s_n)}\right)^{-1}\right)$. We first compute the rate of convergence of $\frac{\hat{\phi}^0(s_n)}{\hat{\phi}^1(s_n)}$, in a fashion similar to the proof above. From the identification result in Section 3, we know

$$
\lim_{s\to\infty}\frac{-i}{a}\ \mathrm{Log}\left(\frac{\phi(s+a|x_1)}{\phi(s+a|x_0)}\left(\frac{\phi(s|x_1)}{\phi(s|x_0)}\right)^{-1}\right) = \frac{1}{a}\left(a\nabla + 2\pi\left\lfloor\frac{1}{2} - \frac{a\nabla}{2\pi}\right\rfloor\right).
$$

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

$$\begin{aligned}\frac{\phi^1(s_n)}{\phi^0(s_n)} &= \frac{\lambda e^{is_n m_1(x_1)}\phi_1(s_n) + (1-\lambda)e^{s_n m_2(x_1)}\phi_2(s_n)}{\lambda e^{is_n m_1(x_0)}\phi_1(s_n) + (1-\lambda)e^{s_n m_2(x_0)}\phi_2(s_n)} \\ &= e^{is_n\nabla}\frac{\frac{\lambda}{1-\lambda}e^{is_n(m_1(x_1)-m_2(x_1))}\frac{\phi_1(s_n)}{\phi_2(s_n)}+1}{\frac{\lambda}{1-\lambda}e^{is_n(m_1(x_0)-m_2(x_0))}\frac{\phi_1(s_n)}{\phi_2(s_n)}+1} \\ &= e^{is_n\nabla}(1+O(f(s_n))).\end{aligned}$$

To compute the estimation error, the scheme is initially similar to the previous proof. We write

$$\hat{\phi}^j(s_n) = \frac{\frac{1}{nh_n^k}\sum_{p=1}^n e^{is_n Z_p}K(\frac{X_p-x_j}{b_n})}{\frac{1}{nh_n^k}\sum_{p=1}^n K(\frac{X_p-x_j}{b_n})} = \frac{n\hat{u}m^j(s_n)}{de\hat{n}om^j}$$

and work with an equation similar to (8.1), here

$$\frac{\hat{\phi}^j(s_n)}{\phi^j(s_n)} = \frac{f_X(x_j)}{de\hat{n}om^j}\,\frac{n\hat{u}m^j(s_n)}{f_X(x_j)\phi^j(s_n)}$$

We compute the convergence rate of the ratios in the last equality.

As in (8.2), we know $\frac{de\hat{n}om^j}{f_X(x_j)} = 1 + O_{\mathbb{P}}\left(b_n^4 + \frac{1}{nb_n^k}\right)^{\frac{1}{2}}$. Now, let $A_n = \frac{n\hat{u}m^j(s_n)}{f_X(x_j)\phi^j(s_n)} : A_n \in \mathbb{C}$. 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 $g_n(x) = f_X(x)\mathbb{E}(e^{is_n Z}|X=x)$ so that $A_n = \frac{n\hat{u}m^j(s_n)}{g_n(x)}$.

Since $g_n(x) = f_X(x)(\lambda e^{is_n m_1(x)}\phi_1(s_n) + (1-\lambda)e^{is_n m_2(x)}\phi_2(s_n))$, we denote $G_n^{lc}(x) = \cos(s_n m_l(x))f_X(x)$ and $G_n^{ls}(x) = \sin(s_n m_l(x))f_X(x)$ for $l = 1, 2$. Then we have

$$\begin{aligned}\mathbb{E}(n\hat{u}m^j(s_n)) =&\mathbb{E}\left(\frac{1}{nb_n^k}\sum_{p=1}^n e^{is_n Z_p}K(\frac{X_p-x_j}{b_n})\right) = \int_{U\in\mathbb{R}^k} g_n(x_j+b_nU)K(U)\mathrm{d}U, \\ =&\lambda\phi_1(s_n)\int[\cos(s_n m_1(x_j+b_nU)) + i\sin(s_n m_1(x_j+b_nU))]f_X(x_j+b_nU)K(U)\mathrm{d}U \\ &+ (1-\lambda)\phi_2(s_n)\int[\cos(s_n m_2(x_j+b_nU)) + i\sin(s_n m_2(x_j+b_nU))]f_X(x_j+b_nU)K(U)\mathrm{d}U \\ =&\lambda\phi_1(s_n)\int[G_n^{1c}(x_j+b_nU) + iG_n^{1s}(x_j+b_nU)]K(U)\mathrm{d}U \\ &+ (1-\lambda)\phi_2(s_n)\int[G_n^{2c}(x_j+b_nU) + iG_n^{2s}(x_j+b_nU)]K(U)\mathrm{d}U\end{aligned}$$

Using the assumption that the kernel is of order 2 (Assumption 8.1),

$$\int G_n^{1c}(x_j+b_nU)K(U)\mathrm{d}U - G_n^{1c}(x_j),\ = \int \frac{b_n^2}{2}U'\nabla^2 G_n^{1c}[x_j+b_n\tau_n(U)U]\,UK(U)\,\mathrm{d}U,$$

where $\tau_n(U) \in [0;1]$ and $\nabla^2 G_n^{1c}(x)$ is the hessian matrix of the function G_n^{1c} evaluated at x. That is,

$$\begin{aligned}\nabla^2 G_n^{1c}(x) = & - s_n^2 f_X(x)\cos(s_n m_1(x))\nabla m_1(x)\nabla m_1(x)' \\ & - s_n \sin(s_n m_1(x))[\nabla m_1(x)\nabla f_X(x)' + \nabla f_X(x)\nabla m_1(x)' + f_X(x)\nabla^2 m_1(x)] \\ & + \cos(s_n m_1(x))\nabla^2 f_X(x).\end{aligned}$$

Similarly to what is done in the first part of this proof, Assumption 8.2 guarantees that $\int G_n^{1c}(x_j + b_nU)K(U)\mathrm{d}U - G_n^{1c}(x_j) = O(b_n s_n)^2$. The same rate applies for G_n^{1s}, G_n^{2c} and G_n^{2s}, implying

$$\mathbb{E}(n\hat{u}\hat{m}^j(s_n)) = \int g_n(x_j + b_nU)K(U)\mathrm{d}U = g_n(x_j) + O(\,(b_ns_n)^2\,[\lambda|\phi_1(s_n)| + (1-\lambda)|\phi_2(s_n)|\,]\,),$$

which gives, for the bias term,

$$\begin{aligned}\mathbb{E}(A_n) &= \frac{1}{f_X(x_j)\phi^j(s_n)}\mathbb{E}(n\hat{u}\hat{m}^j(s_n)) \\ &= 1 + O\left((b_ns_n)^2 \frac{1}{f_X(x_j)} \frac{\lambda|\phi_1(s_n)| + (1-\lambda)|\phi_2(s_n)|}{\lambda e^{is_nm_1(x_j)}\phi_1(s_n) + (1-\lambda)e^{is_nm_2(x_j)}\phi_2(s_n)}\right) \\ &= 1 + O\left((b_ns_n)^2 \frac{1}{e^{is_nm_2(x_j)}} \frac{\lambda\frac{|\phi_1(s_n)|}{|\phi_2(s_n)|} + (1-\lambda)}{\lambda\frac{\phi_1(s_n)}{|\phi_2(s_n)|}e^{is_n(m_1(x_j)-m_2(x_j))} + (1-\lambda)}\right) \\ &= 1 + O(b_ns_n)^2,\end{aligned}$$

where the last equality comes from Assumption 8.3 (iv). As for the variance term, we write

$$\begin{aligned}Var(\frac{n\hat{u}\hat{m}^j(s_n)}{f_X(x_j)\phi^j(s_n)}) &= \frac{1}{f_X(x_j)^2|\phi^j(s_n)|^2}\frac{1}{nb_n^{2k}}Var(e^{is_nZ}K(\frac{X-x_j}{b_n})) \\ &\leq \frac{1}{f_X(x_j)^2|\phi^j(s_n)|^2}\frac{1}{nb_n^{2k}}\mathbb{E}(|e^{is_nZ}K(\frac{X-x_j}{b_n})|^2) \\ &\leq \frac{1}{|\phi^j(s_n)|^2}\frac{1}{nb_n^k}\frac{\int f_X(x_j+b_nU)K^2(U)\mathrm{d}U}{f_X(x_j)^2},\end{aligned}$$

and Assumption 8.2 (iii) 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 those two computations, the following reasoning gives a convergence rate for A_n : $Bias(\Re(A_n)) = \Re(Bias(A_n)) = O(b_ns_n)^2$. 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, $\Re(A_n) - 1 = O_{\mathbb{P}}((b_n s_n)^4 + \frac{1}{nb_n^k|\phi^j(s_n)|^2})^{1/2}$, and $\Im(A_n) = O_{\mathbb{P}}((b_n s_n)^4 + \frac{1}{nb_n^k|\phi^j(s_n)|^2})^{1/2}$. This gives,

$$|A_n - 1| = O_{\mathbb{P}}((b_n s_n)^4 + \frac{1}{nb_n^k|\phi^j(s_n)|^2})^{1/2},$$

so that the estimation error is

$$\begin{aligned}\frac{\hat{\phi}^j(s_n)}{\phi^j(s_n)} &= 1 + O_{\mathbb{P}}\left[\left(b_n^4 + \frac{1}{nb_n^k}\right)^{\frac{1}{2}} + \left((b_n s_n)^4 + \frac{1}{nb_n^k|\phi^j(s_n)|^2}\right)^{1/2}\right] \\ &= 1 + O_{\mathbb{P}}\left((b_n s_n)^4 + \frac{1}{nb_n^k|\phi^j(s_n)|^2}\right)^{1/2}.\end{aligned}$$

Finally we obtain

$$\begin{aligned}\frac{\hat{\phi}^1(s_n)}{\hat{\phi}^0(s_n)} &= \frac{\hat{\phi}^1(s_n)}{\phi^1(s_n)}\left(\frac{\hat{\phi}^0(s_n)}{\phi^0(s_n)}\right)^{-1}\frac{\phi^1(s_n)}{\phi^0(s_n)} \\ &= e^{is_n\nabla}[1 + O_{\mathbb{P}}(f(s_n))]\ \left[1 + O_{\mathbb{P}}\left((b_n s_n)^4 + \frac{1}{nb_n^k|\phi^j(s_n)|^2}\right)^{1/2}\right].\end{aligned}$$

Plugging in this expression in the definition of the estimator, we obtain

$$\begin{aligned}\hat{\nabla} =& \frac{-i}{a_n}\ \mathrm{Log}\left(\frac{\hat{\phi}^1(s_n + a_n)}{\hat{\phi}^0(s_n + a_n)}\left(\frac{\hat{\phi}^0(s_n)}{\hat{\phi}^1(s_n)}\right)\right) \\ =& \frac{-i}{a_n}\ \mathrm{Log}\{e^{i(s_n+a_n)\nabla}[1 + O_{\mathbb{P}}(f(s_n + a_n))][1 + O_{\mathbb{P}}\left((b_n(s_n + a_n))^4 + \frac{1}{nb_n^k|\phi^j(s_n + a_n)|^2}\right)^{1/2}] \\ & e^{-is_n\nabla}[1 + O_{\mathbb{P}}(f(s_n))][1 + O_{\mathbb{P}}\left((b_n s_n))^4 + \frac{1}{nb_n^k|\phi^j(s_n)|^2}\right)^{1/2}]\} \\ =& \frac{-i}{a_n}\ \mathrm{Log}\{e^{ia_n\nabla}[1 + O_{\mathbb{P}}(f(s_n + a_n)) + O_{\mathbb{P}}(f(s_n)) + O_{\mathbb{P}}\left((b_n s_n)^4 + \frac{1}{nb_n^k|\phi^j(s_n + a_n)|^2}\right)^{1/2} \\ & + O_{\mathbb{P}}\left((b_n s_n)^4 + \frac{1}{nb_n^k|\phi^j(s_n)|^2}\right)^{1/2}]\}.\end{aligned}$$

As the term multiplying $e^{ia_n\nabla}$ in the Log converges to 1, and eventually $a_n\nabla \in (-\pi;\pi)$, the expression above becomes

$$\begin{aligned}\hat{\nabla} = \frac{-i}{a_n}\{ia_n\nabla +\ & \mathrm{Log}[1+O_{\mathbb{P}}(f(s_n + a_n)) + O_{\mathbb{P}}(f(s_n)) \\ & +O_{\mathbb{P}}\left((b_n s_n)^4 + \frac{1}{nb_n^k|\phi^j(s_n + a_n)|^2}\right)^{1/2}\quad +O_{\mathbb{P}}\left((b_n s_n)^4 + \frac{1}{nb_n^k|\phi^j(s_n)|^2}\right)^{1/2}]\},\end{aligned}$$

that is, using the first order approximation of the principal value of the log around 1,

$$\hat{\nabla} = \nabla + \frac{1}{a_n} O_{\mathbb{P}} \left(f(s_n + a_n) + f(s_n) + \left((b_n s_n)^4 + \frac{1}{n b_n^k |\phi^j(s_n + a_n)|^2} \right)^{1/2} + \left((b_n s_n)^4 + \frac{1}{n b_n^k |\phi^j(s_n)|^2} \right)^{1/2} \right).$$

□

The only restriction imposed on the tuning parameter a_n is that it converges to 0.

For the sake of simplicity, we now write $\hat{\Delta} - \Delta = O_{\mathbb{P}}(\alpha_n)$ and $\hat{\nabla} - \nabla = O_{\mathbb{P}}(\beta_n)$. The rates α_n and β_n depend on the distributions of the error terms, and we show here that the rates are polynomial in n if these distributions are normal.

Indeed if $\epsilon_1 | x \sim \mathcal{N}(0, \sigma_1^2)$ and $\epsilon_2 | x \sim \mathcal{N}(0, \sigma_2^2)$, with $\delta = \sigma_1^2 - \sigma_2^2 > 0$, then Assumption 8.3 is satisfied. For the ratio of the mgf, $\forall \epsilon > 0$, $e^{\epsilon t} \frac{M_2(t)}{M_1(t)} = e^{\epsilon t - \frac{\delta}{2} t^2} \underset{t \to \infty}{=} O(\mu(t))$, with $\mu(t) = e^{-(\frac{\delta}{2} - \nu) t^2} \to 0$, as $t \to \infty$, for some $0 < \nu < \frac{\delta}{2}$. And as for the ratio of the characteristic functions, $\frac{\phi_1(s)}{\phi_2(s)} = e^{-\frac{1}{2}\delta s^2} \underset{s \to \infty}{=} O(f(s))$, with $f(s) = e^{-\frac{1}{2}\delta s^2} \xrightarrow[s \to \infty]{} 0$. We take a fixed a in the definition of $\hat{\nabla}$ here to simplify the computations, assuming a is small enough. Applying the results from the estimation proofs, the convergence rates are

(i) $\hat{\Delta} - \Delta = \frac{1}{t_n} O_{\mathbb{P}} \left[e^{-(\frac{\delta}{2} - \nu) t_n^2} + \left((t_n h_n)^4 + \frac{1}{n h_n^k} e^{\sigma_1^2 t_n^2} \right)^{\frac{1}{2}} \right]$,

(ii) $\hat{\nabla} - \nabla = O_{\mathbb{P}} \left[e^{-\frac{1}{2}\delta s_n^2} + \left((b_n s_n)^4 + \frac{1}{n b_n^k} e^{\sigma_2^2 (s_n + a)^2} \right)^{\frac{1}{2}} \right]$.

One can show that with the appropriate choice of the sequences t_h, h_n, s_n, and b_n, the rates are polynomial in n. For example, it is the case 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 $\epsilon, \beta < \frac{1}{5}$.

Proof 2. We now focus on the estimation of the remaining objects. We showed that if $\lambda \in (0,1)$, then $\lambda = \frac{\mathbb{E}(Z|X=x_1) - \mathbb{E}(Z|X=x_0) - \nabla}{\Delta - \nabla}$. A natural estimator is therefore

$$\hat{\lambda} = \frac{\hat{E}(Z|X = x_1) - \hat{E}(Z|X = x_0) - \hat{\nabla}}{\hat{\Delta} - \hat{\nabla}},$$

where $\hat{E}(Z|X = .)$ is the usual multivariate kernel regression estimator, $\hat{E}(Z|X = x) = \frac{\sum_{p=1}^n Z_p K(\frac{X_p - x}{d_n})}{\sum_{p=1}^n K(\frac{X_p - x_j}{d_n})}$ where the kernel does not have to be the one used for the previous estimators but will be written K for the sake of simplicity. Similarly the point estimation of the regression functions m_1 and m_2 is derived from Equation (3.10). Writing $\hat{C} = \frac{1}{2} \left(\hat{E}(Z^2|X = x_0) - \hat{E}(Z^2|X = x_1) + \hat{\lambda} \hat{\Delta}^2 + (1 - \hat{\lambda}) \hat{\nabla}^2 \right)$, our estimators of $m_1(x_0)$ and $m_2(x_0)$ are

$$\begin{bmatrix} \hat{m}_1(x_0) \\ \hat{m}_2(x_0) \end{bmatrix} = \begin{pmatrix} \hat{\lambda}^{-1} & 0 \\ 0 & (1-\hat{\lambda})^{-1} \end{pmatrix} \begin{pmatrix} -\hat{\Delta} & -\hat{\nabla} \\ 1 & 1 \end{pmatrix}^{-1} \begin{bmatrix} \hat{C} \\ \hat{E}(Z|X=x_0) \end{bmatrix}. \tag{8.4}$$

The convergence rate of these estimators can be computed easily. With the usual assumptions for kernel estimation and the appropriate choice of bandwidths $d_n = n^{\frac{-1}{k+4}}$, it is known that $\hat{E}(Z|X = x) - \mathbb{E}(Z|X = x) = O_{\mathbb{P}}(n^{-\frac{2}{k+4}})$ and $\hat{E}(Z^2|X = x) - \mathbb{E}(Z^2|X = x) = O_{\mathbb{P}}(n^{-\frac{2}{k+4}})$, see, e.g, Härdle and Linton (1994). Writing $\epsilon_n = n^{-\frac{2}{k+4}} + \alpha_n + \beta_n$, one obtains $\hat{\lambda} = \lambda + O_{\mathbb{P}}(\epsilon_n)$. The estimators of $m_1(x_0)$ and $m_2(x_0)$, being linearizable functions of $\hat{\lambda}$, $\hat{\Delta}$ and $\hat{\nabla}$, their rates of convergence are similarly bounded. This is summarized in the next proposition.

Proposition 8.2. *Under Assumptions 8.2, 8.3, assuming that $\lambda \in (0,1)$, K satisfies Assumption 8.1, $d_n \to 0$, and $nd_n^k \to 0$,*

(1) $\hat{\lambda} = \lambda + O_{\mathbb{P}}(\epsilon_n)$,

(2) $\hat{m_i}(x_0) = m_i(x_0) + O_{\mathbb{P}}(\epsilon_n), \text{for } i = 1, 2$

Proof 3. To estimate the CDF of ϵ_1 and ϵ_2, we use Equation (3.49) and propose the following estimator

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

In this formula $p(n) \in \mathbb{N}$ will be specified later, $\hat{g}(x) = \hat{m}_1(x) - \hat{m}_2(x)$, $\hat{\delta} = \hat{\Delta} - \hat{\nabla}$, and $\hat{F}(.|.)$ is the kernel regression estimator of the conditional cumulative distribution function,

$$\hat{F}(z|x) = \frac{\sum_{j=1}^n \mathbb{1}(Z_j \leq z) k(\frac{X_j - x}{c_n})}{\sum_{j=1}^n k(\frac{X_j - x}{c_n})}.$$

The kernel function and the bandwidth may differ from the choices for our previous kernel regression estimators, and will be here written as k and c_n respectively.

Assumption 8.4.

(1) The probability distribution functions of ϵ_1 and ϵ_2, f_1 and f_2, are bounded by a constant c,

(2) f_j is twice differentiable on $\mathbb{R}$, and f_j, f_j', f_j'' are continuous and bounded for $j = 1, 2$.

Assumption 8.5. *We assume that $k(.)$ satisfies Assumption 8.1, and in addition impose,*

(1) $||k||_\infty < \infty$,

(2) The kernel function k has support contained in $[-\frac{1}{2}, \frac{1}{2}]^k$,

(3) Assumptions (K-iii) and (K-iv) of Einmahl and Mason (2005) hold for $k(.)$,

(4) $c_n \geq C' \frac{\log(n)}{n}$, $c_n = O(n^{-\gamma_1})$, for some $\gamma_1 < 1$.

The assumptions from Einmahl and Mason (2005) are conditions on the covering and measurability properties of the class of functions $\left\{k(\frac{x-.}{c});\ c > 0,\ x \in \mathbb{R}^k\right\}$.

Proposition 8.3. *Under 8.2, 8.3, 8.4 and 8.5,*

$$\hat{F}_1(z) - F_1(z) = O_{\mathbb{P}}\left((p(n)+1)n^{\frac{-2}{k+4}+a} + p(n)^2\epsilon_n + e^{-\gamma_0 p(n)}\right), \text{ and}$$
$$\hat{F}_2(z) - F_2(z) = O_{\mathbb{P}}\left((p(n)+1)n^{\frac{-2}{k+4}+a} + p(n)^2\epsilon_n + e^{-\gamma_0 p(n)}\right).$$

Proof. Fix $z \in \mathbb{R}$. Write $\xi_j^0 = z + j\delta(x_1, x_0) + m_2(x_0)$, $\hat{\xi}_j^0 = z + j\hat{\delta}(x_1, x_0) + \hat{m}_2(x_0)$, $\xi_j^1 = z + j\delta(x_1, x_0) + m_1(x_1) - g(x_0)$ and $\hat{\xi}_j^1 = z + j\hat{\delta}(x_1, x_0) + \hat{m}_1(x_1) - \hat{g}(x_0)$.

$$\begin{aligned}
\hat{F}_2(z) - F_2(z) =& \frac{1}{1-\lambda} \sum_{j=0}^{p(n)} \left\{ \left[\hat{F}(\hat{\xi}_j^1|x_1) - F(\xi_j^1|x_1)\right] - \left[\hat{F}(\hat{\xi}_j^0|x_0) - F(\xi_j^0|x_0)\right] \right\} \\
& - \frac{1}{1-\lambda} \sum_{j=p(n)+1}^{\infty} \left[F(\xi_j^1|x_1) - F(\xi_j^0|x_0)\right] \\
& + \left(\frac{1}{1-\hat{\lambda}} - \frac{1}{1-\lambda}\right) \sum_{j=0}^{p(n)} \hat{F}(\hat{\xi}_j^1|x_1) - \hat{F}(\hat{\xi}_j^0|x_0).
\end{aligned} \tag{8.5}$$

We write $\hat{F}_2(z) - F_2(z) = I_1 - I_2 + I_3$, and the convergence rate of each part in the right hand side of (8.5) will be computed separately.

For I_1, we write

$$\begin{aligned}
(1-\lambda)\, I_1 =& \sum_{j=0}^{p(n)} \left[\hat{F}(\hat{\xi}_j^1|x_1) - F(\hat{\xi}_j^1|x_1)\right] + \sum_{j=0}^{p(n)} \left[F(\hat{\xi}_j^1|x_1) - F(\xi_j^1|x_1)\right] \\
& - \sum_{j=0}^{p(n)} \left[\hat{F}(\hat{\xi}_j^0|x_0) - F(\hat{\xi}_j^0|x_0)\right] - \sum_{j=0}^{p(n)} \left[F(\hat{\xi}_j^0|x_0) - F(\xi_j^0|x_0)\right].
\end{aligned}$$

We know $\frac{\partial F(z|x)}{\partial z} = f(z|x) = \lambda f_1(z - m_1(x)) + (1-\lambda) f_2(z - m_2(x))$, and Assumption 8.4 guarantees that $f(y|x)$ is bounded by c, $\forall\,(x, y)$ therefore $y \mapsto F(y|x)$ is Lipschitz continuous with constant c. That is, for $v = 0, 1$, $|F(\hat{\xi}_j^v|x_v) - F(\xi_j^v|x_v)| \leq c\,|\hat{\xi}_j^v - \xi_j^v|$ implying

$$|\sum_{j=0}^{p(n)} F(\hat{\xi}_j^v|x_v) - F(\xi_j^v|x_v)| = O_{\mathbb{P}}(p(n)^2\epsilon_n).$$

For the two other terms in I_1, we write

$$F_n(.|x_i) = \frac{\mathbb{E}(\mathbb{1}(Z \le z)k(\frac{X-x_i}{c_n}))}{\mathbb{E}(k(\frac{X-x_i}{c_n}))}.$$

Then under Assumption 8.5, we apply Theorem 3 of Einmahl and Mason (2005), which gives the rate of the supremum of $||\hat{F}(.|x) - F_n(.|x)||_\infty$ over a certain range of bandwidths and over $x \in I$ where I is a compact subset of $\mathbb{R}^k$. For the specific bandwidth b_n and taking $I = \{x_0, x_1\}$ we then have

$$\begin{aligned}\limsup_{n\to\infty}(nc_n^k)^{1/2}||\hat{F}(.|x) - F_n(.|x)||_\infty &= O_{\text{a.s}}\left(\max(\log\log n, -\log(c_n))^{1/2}\right)\\ &= O_{\text{a.s}}\left((-\log c_n)^{1/2}\right). \end{aligned} \tag{8.6}$$

We now examine $F_n(.|x) - F(.|x)$. Write $f(.,.)$ the joint density of (Z, X), then we define

$$F_X(x, z) = \int_{z'\le z} f(z', x)\mathrm{d}z' = F(z|x)f_X(x) = [\lambda F_1(z - m_1(x)) + (1-\lambda)F_2(z - m_2(x))]f_X(x)$$

and write

$$F_n(z|x_i) - F(z|x_i) = \frac{\frac{1}{c_n^k}\mathbb{E}(\mathbb{1}(Z \le z)k(\frac{X-x_i}{c_n})) - F_X(x_i, z)}{\frac{1}{c_n^k}\mathbb{E}(k(\frac{X-x_i}{c_n}))} + F_X(x_i, z)\left(\frac{1}{\frac{1}{c_n^k}\mathbb{E}(k(\frac{X-x_i}{c_n}))} - \frac{1}{f_X(x_i)}\right).$$

Under Assumption 8.2, 8.4 and 8.5, we know that $\frac{1}{c_n^k}\mathbb{E}(k(\frac{X-x_i}{c_n})) - f_X(x_i) = O(c_n^2)$. Similarly,

$$\frac{1}{c_n^k}\mathbb{E}\left[\mathbb{1}(Z \le z)\,k\left(\frac{X - x_i}{c_n}\right)\right] - F_X(x_i, z) = \frac{c_n^2}{2}\int_{U\in\mathbb{R}^k} U'\,\nabla_X^2 F_X(x_i + b_n\tau_n(U)U, z)\,Uk(U)\mathrm{d}U,$$

and Assumption 8.2, 8.4 and 8.5 guarantee that $\nabla_X^2 F_X(.,.)$ is uniformly bounded over $\mathbb{R}^{k+1}$. Therefore

$$\sup_{z\in\mathbb{R}}\left|\frac{1}{c_n^k}\mathbb{E}\left[\mathbb{1}(Z \le z)\,k\left(\frac{X - x_i}{c_n}\right)\right] - F_X(x_i, z)\right| = O(c_n^2).$$

which gives, for $i = 0, 1$,

$$||F_n(.|x_i) - F(.|x_i)||_\infty = O(c_n^2). \tag{8.7}$$

Equations (8.6) and (8.7) give $||\hat{F}(.|x) - F(.|x)||_\infty = O_{\mathbb{P}}((-\log(c_n))^{1/2}(nc_n^k)^{-1/2} + c_n^2)$. For the appropriate choice of γ_1 in Assumption 8.5 and for any small $a > 0$,

$$\sup_{y\in\mathbb{R}}|\hat{F}(y|x_i) - F(y|x_i)| = O_{\mathbb{P}}(n^{-\frac{2}{k+4}+a}),\ i = 0, 1.$$

This implies that

$$\sum_{j=0}^{p(n)}\left[\hat{F}(\hat{\xi}_j^1|x_i) - F(\hat{\xi}_j^1|x_i)\right] = O_{\mathbb{P}}((p(n)+1)n^{-\frac{2}{k+4}+a}),\ i = 0, 1.$$

Therefore,

$$I_1 = O_{\mathbb{P}}\left((p(n)+1)n^{-\frac{2}{k+4}+a} + p(n)^2\epsilon_n\right).$$

Looking at I_2, by construction the second sum appearing in the right hand side of (8.5) 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 $1 - F_2(C) = \mathbb{P}(\epsilon_2 > C) \leq e^{-tC}M_2(t)$ using the assumption that the moment generating functions are finite. Fixing $t_0 \in \mathbb{R}_+$, $1 - F_2[z + (p(n)+1)\delta(x_1, x_0)] \leq e^{t_0 z + (p(n)+1)\delta(x_1,x_0)t_0}$ which guarantees the existence of $\gamma_0 > 0$ such that $I_2 = O(e^{-\gamma_0\, p(n)})$.

As for I_3, we showed in our computation for I_1 that $\sum_{j=0}^{p(n)} \hat{F}(\hat{\xi}_j^1|x_1) - \hat{F}(\hat{\xi}_j^0|x_0) \xrightarrow[n\to\infty]{} F_2(z)$. As $\frac{1}{1-\hat{\lambda}} - \frac{1}{1-\lambda} = O_{\mathbb{P}}(\epsilon_n)$, we have

$$I_3 = O_{\mathbb{P}}(\epsilon_n).$$

Adding these three parts, we obtain

$$\hat{F}_2(z) - F_2(z) = O_{\mathbb{P}}\left((p(n)+1)n^{-\frac{2}{k+4}+a} + p(n)^2\epsilon_n + e^{-\gamma_0 p(n)}\right).$$

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

$$\hat{F}_1(z) = \frac{1}{\hat{\lambda}}\left[\hat{F}(z + \hat{m}_1(x)) - (1-\hat{\lambda})\hat{F}_2(z + \hat{m}_1(x) - \hat{m}_2(x))\right],$$

which will converge to $F_1(z)$ at the same rate.

□

In the case where ϵ_n is slower than $n^{-2\,\frac{1-2a}{k+4}}$, for some a, which happens when for instance the error terms are normally distributed, then p_n is solution to $\epsilon_n p_n = t_0 e^{-t_0 p_n}$.

9. Conclusion

New nonparametric identification results for finite mixture models are developed. These open up the possibility of flexibly modeling economic behavior in the presence of unobserved heterogeneity.

10. Appendix

This Appendix presents the proofs of some of the results presented in the previous sections.

Proof of Lemma 4.1. Define $\delta(x) := m_2(x) - m_1(x)$, $\dot{m}_1(x) := m_1(x) - m_1(x_0)$, $\dot{m}_2(x) := m_2(x) - m_2(x_0)$,

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

and

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

In what follows we show that the slopes of m_1 and m_2 over the interval connecting x and x_0, as well as the values of $\lambda(\cdot)$ at these two points, are all recovered from $r(+\infty, x)$, $r(-\infty, x)$ and $\lim_{c\downarrow 0} \tilde{\lambda}_c(x)$.

Case (1): $\lambda(x) = \lambda(x_0) = 1$.

With the given structure of the model we have $m_1(x) = m_2(x)$, $m_1(x_0) = m_2(x_0)$, and $M_1 \equiv M_2$ in this case. Thus

$$R(x,t) = \frac{e^{tm_1(x)}}{e^{tm_1(x_0)}} = e^{t\dot{m}_1(x)}$$

and

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

therefore Condition 4.1(i) fails. On the other hand this means

$$K_{+\infty}(x) = K_{-\infty}(x) = 1,$$

yielding

$$\tilde{\lambda}_c(x) = \frac{1 - 1 + c}{1 - 1 + c} = 1,$$

therefore Condition 4.1(ii) holds in this case. Moreover, the values of λ are identifiable from $\lim_{c\downarrow 0} \lambda_c(x)$.

Case (2): $\lambda(x) < 1, \lambda(x_0) < 1$. *Condition 4.1*(i) *holds.*

In this case the two slopes $(m_1(x) - m_1(x_0), m_2(x) - m_2(x_0))$ are identified as in the proof of Lemma 3.1.

Take δ' as in the proof of Lemma 3.1. We first consider the case with t tending to $+\infty$. If $h(\pm\epsilon, t) = O(1)$ holds, then according to the proof of Lemma 3.1 we have

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

for $x \in N^1(x_0, \delta')$ and consequently

$$K_{+\infty}(x) = \frac{\lambda(x)}{\lambda(x_0)}.$$

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

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

and then

$$K_{+\infty}(x) = \frac{1-\lambda(x)}{1-\lambda(x_0)}.$$

With these results we see $K_{+\infty}(x) \neq K_{-\infty}(x)$ iff $\lambda(x) \neq \lambda(x_0)$. With

$$\begin{aligned}\lim_{c\downarrow 0} \tilde{\lambda}_c(x) &= \frac{1-K_{-\infty}(x)}{K_{+\infty}(x) - K_{-\infty}(x)} \\ &= \lambda(x).\end{aligned}$$

By continuity $\lambda(x_0)$ is identified as $\lim_{x\to x_0} \lambda(x)$. If $K_{+\infty}(x) = K_{-\infty}(x)$ we can obtain the value of $\lambda(x)$ (and thus $\lambda(x_0)$) as $\lim_{c\downarrow 0} \lambda_c$, as noted in the proof of Lemma 3.2.

Now we let $t \to -\infty$. If $h(\pm\epsilon, t) = O(1)$ and $1/h(\pm\epsilon, t) = O(1)$ as $t \to -\infty$ we have

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

and

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

respectively, so once again we identify $\lambda(x)$ and the two slopes by switching $\lambda(x)$ and $\lambda(x_0)$ and m_1 and m_2.

If both $h(\pm\epsilon, t) = O(1)$ and $1/h(\pm\epsilon, t) = O(1)$ hold, $D(x_0) = 0$. If $D(x) > 0$, for example, then $r(x, -\infty) = \dot{m}_1(x)$ and $r(x, +\infty) = \dot{m}_2(x)$. (In this case Condition 4.1(i) is automatically satisfied.) $\lambda(x)$ is identified, hence $\lambda(x_0)$ too, as above.

Case (3): $\lambda(x) < 1, \lambda(x_0) < 1$. Condition 4.1(i) *fails.*

Wlog suppose $r(x, +\infty) = \dot{m}_1(x)$, then

$$K_{+\infty,t}(x) = \frac{\lambda(x) + (1-\lambda(x))e^{t\delta(x)}\frac{M_2(t)}{M_1(t)}}{\lambda(x_0) + (1-\lambda(x_0))e^{t\delta(x_0)}\frac{M_2(t)}{M_1(t)}},$$

so for Condition 4.1(ii) to hold we need

$$\lambda(x) + (1-\lambda(x))e^{t\delta(x)}\frac{M_2(t)}{M_1(t)} = \lambda(x_0) + (1-\lambda(x_0)))e^{t\delta(x_0)}\frac{M_2(t)}{M_1(t)}$$

or

$$\frac{\lambda(x)-\lambda(x_0)}{1-\lambda(x_0)} = \left[e^{t\delta(x_0)} + \frac{1-\lambda(x)}{1-\lambda(x_0)} e^{t\delta(x)}\right] \frac{M_2(t)}{M_2(t)}.$$

Take $x_1 \neq x_0$ in $N^1(x_0, \delta')$. Since the right hand side of the above equation is positive, we have $\lambda(x_1) \neq \lambda(x_0)$. Then

$$\begin{aligned} \frac{\frac{\lambda(x)-\lambda(x_0)}{1-\lambda(x_0)}}{\frac{\lambda(x_1)-\lambda(x_0)}{1-\lambda(x_0)}} &= \frac{1+\frac{1-\lambda(x)}{1-\lambda(x_0)} e^{t[\delta(x)-\delta(x_0)]}}{1+\frac{1-\lambda(x_1)}{1-\lambda(x_0)} e^{t[\delta(x_1)-\delta(x_0)]}} \\ &= \frac{1+\frac{1-\lambda(x)}{1-\lambda(x_0)} e^{t[\dot{m}_2(x)-\dot{m}_1(x)]}}{1+\frac{1-\lambda(x_1)}{1-\lambda(x_0)} e^{t[\dot{m}_2(x_1)-\dot{m}_1(x_1)]}}. \end{aligned}$$

In view of the non-parallel assumption, the right hand does not depend of t only if $\lambda(x) = 0$, which is a contradiction. Thus Case (3) is (correctly) precluded by Condition 4.1.

Case (4): $\lambda(x) < 1, \lambda(x_0) = 1$. *Condition 4.1*(i) *holds.*

Note that $\lambda(x_0) = 1$ means $m_1(x_0) = m_2(x_0)$, and moreover, with Assumption 4.1, M_1 and M_2 are identical. Then

$$\begin{aligned} R(x,t) &= \frac{\lambda(x)e^{tm_1(x)}M_1(t) + (1-\lambda(x))e^{tm_2(x)}M_1(t)}{e^{tm_1(x_0)}M_1(t)} \\ &= \lambda(x)e^{t\dot{m}_1} + (1-\lambda(x)e^{t\dot{m}_2(x)}. \end{aligned}$$

If, for example, $\dot{m}_1(x) > \dot{m}_2(x)$, $r(x,+\infty) = \dot{m}_1(x)$ and $r(x,-\infty) = \dot{m}_2(x)$, and moreover,

$$\begin{aligned} K_{+\infty,t}(x) &= R(x,t)e^{-t\dot{m}_1(x)} \\ &= \lambda(x) + (1-\lambda(x)e^{t[\dot{m}_2(x)-\dot{m}_1(x)]} \\ &\to \lambda(x) \text{ as } t \to \infty, \end{aligned}$$

that is, $\lambda(x) = K_{+\infty}(x)$. Proceeding analogously, we have $\lambda(x) = K_{-\infty}(x)$. Use these values in the definition of $\tilde{\lambda}$, we see that $\lambda(x)$ is identified from $\tilde{\lambda}(x)$. Analysis of the case with $\dot{m}_1(x) < \dot{m}_2(x)$ is analogous. And of course $\dot{m}_1(x) = \dot{m}_2(x)$ cannot happen.

Case (5): $\lambda(x) < 1, \lambda(x_0) = 1$. *Condition 4.1*(i) *fails.*

As seen in Case (4), in this case we have

$$R(x,t) = \lambda(x)e^{t\dot{m}_1} + (1-\lambda(x))e^{t\dot{m}_2(x)},$$

and if, for example, $\dot{m}_1(x) > \dot{m}_2(x)$

$$K_{+\infty,t}(x) = \lambda(x) + (1-\lambda(x))e^{t[\dot{m}_2(x)-\dot{m}_1(x)]}.$$

Thus Condition 4.1(ii) fails and this case is (correctly) precluded. Analysis of the case with $\dot{m}_1(x) < \dot{m}_2(x)$ is analogous, and $\dot{m}_1(x) \neq \dot{m}_2(x)$ as above.

Finally, note that $\lambda(x_0) < 1$ then by continuity $\lambda(x) < 1$ for every $x \in N^1(x_0, \delta')$ for sufficiently small δ', so this reduces to either Case (2) or (3).

$\square$

Proof of Proposition 6.1. The recursive formula in Lemma 6.1 then becomes (NOTE THE USE OF x, not x_a)

$$R^j_{k+1} = D_{x^1}\left(\frac{R^j_k}{R^k_k}\right) + t\frac{R^j_k}{R^k_k}D_{x^1}m_{j,k}, j = 1, ..., J \tag{10.1}$$

with initial conditions

$$R^j_2 = tD_{x^1}m_{j1}, j = 1, ..., J. \tag{10.2}$$

For $k = 3$,

$$\begin{aligned}
R^j_3 &= D_{x^1}\left(\frac{R^j_2}{R^2_2}\right) + t\frac{R^j_2}{R^2_2}D_{x^1}m_{j,2} \\
&= D_{x^1}\left(\frac{D_{x^1}m_{j1}}{D_{x^1}m_{21}}\right) + t\frac{D_{x^1}m_{j1}}{D_{x^1}m_{21}}D_{x^1}m_{j,2} \\
&= \frac{D^2_{x^1}m_{j,1}D_{x^1}m_{2,1} - D^2_{x^1}m_{2,1}D_{x^1}m_{j,1} + tD_{x^1}m_{j,1}D_{x^1}m_{j,2}D_{x^1}m_{2,1}}{(D_{x^1}m_{2,1})^2} \\
&= \frac{P^j_3}{(D_{x^1}m_{2,1})^2}
\end{aligned}$$

where

$$P^j_3 = D^2_{x^1}m_{j,1}D_{x^1}m_{2,1} - D^2_{x^1}m_{2,1}D_{x^1}m_{j,1} + tD_{x^1}m_{j,1}D_{x^1}m_{j,2}D_{x^1}m_{2,1}.$$

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 will be justified shortly): also assume that for $k = h$, R_h^j can be written as

$$R_h^j = \frac{P_h^j}{(P_{h-1}^{h-1})^2}, \tag{10.3}$$

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

$$P_h^j = D_{x^1}P_{h-1}^jP_{h-1}^{h-1} - P_{h-1}^jD_{x^1}P_{h-1}^{h-1} + tP_{h-1}^jP_{h-1}^{h-1}D_{x^1}m_{j,h-1}. \tag{10.4}$$

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 (10.3) and (10.4) hold for $k = h$, they also hold for $k = h + 1$. In short, the original system of equations (10.1) and (10.2) that determine R_k^j can be rewritten in terms of P_k^js as follows:

$$P_1^j = 1, \quad P_{h+1}^j = D_{x^1} P_h^j P_h^h - P_h^j D_{x^1} P_h^h + t P_h^j P_h^h D_{x^1} m_{j,h},$$
$$R_k^j = \frac{P_k^j}{(P_{k-1}^{k-1})^2}, \qquad 1 \leq k, j \leq J. \tag{10.5}$$

(The fact that $P_1^j = 1, j = 1, ...J$ are appropriate initial conditions can be easily verified.) In particular, (10.5) implies that

$$P_{h+1}^{h+1} = D_{x^1} P_h^{h+1} P_h^h - P_h^{h+1} D_{x^1} P_h^h + t P_h^{h+1} P_h^h D_{x^1} m_{h+1,h}. \tag{10.6}$$

Note that (10.6) 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 $P_2^j = tD_{x^1} m_{j,1}, j = 2, ..., J$ according to (10.5), therefore this claim applies to the case of $k = 2$. But (10.5) 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 $P_2^j = tD_{x^1} m_{j,1}, j = 2, ..., J$. Now, suppose the claim holds for $k = h$. Then by (10.6) 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

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

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 $x \in N^1(x_a, \delta')$, $(D_1 m_l)_{l=1..J}$ 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 $j \geq l$

implies that

$$\deg_t(P_l^l) = \deg_t(P_l^j),\ \forall j \geq l, \tag{10.7}$$

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 (10.6) 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

$$\deg_t(P_{l+1}^{l+1}) = 2\deg_t(P_l^l) + 1 \tag{10.8}$$

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 (10.8) under the initial condition $\deg_t(P_3^3) = 1$,

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

Since $P_k = P_k^k, k = 1, ..., J$ are polynomials, they are nonzero for sufficiently large t. This justifies division by P_k used throughout the current proof for sufficiently large t. $\square$

Proposition 10.1. *There exists* $X^{(J)} = (x_1^{(J)}, ..., x_{J-1}^{(J)}) \in B(x_0, \delta')^{J-1}$ *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 10.1.

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

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 $V(\cdot, c^{(J)})$: what follows explain how to.

We fix $c^{(1)} = (c_1^{(1)}, ..., c_J^{(1)}) \in N^1(x_0, \delta')$, $A_1 = \max_{\sigma \in S_J} V(\sigma, c^{(1)})$, $\Sigma_1 = \left\{\sigma \in S_J | V(\sigma, c^{(1)}) = A_1\right\}$ (and $\Sigma_1 \neq \emptyset$), and $B_1 = \max_{\sigma \in S_J \setminus \Sigma_1} V(\sigma, c^{(1)})$ (if B_1 does exist, then $B_1 < A_1$). We consider a change of the first component of $c^{(1)}$, that is a vector $c^{(2)}$ which differs from $c^{(1)}$ only in the first component: the first component of $c^{(1)}$ is a point in $\mathbb{R}^n$, we consider a variation in its first covariate, with respect to which we know that the $(m_i)_{i=1...J}$ are J times differentiable.

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

We know that for all $x \in N^1(x_0, \delta')$, $(D_1 m_j(x))_{j=1..J}$ take distinct values: $\underset{s \in \{\sigma(1) | \sigma \in \Sigma_1\}}{\operatorname{argmax}} D_1 m_s(c_1^{(1)})$ is a singleton set $\{s_1\}$. Hence, since the m_i functions are at least twice differentiable, they are continuously differentiable, we can choose $c_1^{(2)}$ close enough from $c_1^{(1)}$ so that

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

$$c_1^{(2)} \in N^1(x_0, \delta'),$$

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 obtained at the end that we call $c^{(J)}$ and σ_J whatever the final number of steps is, are such that

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

which is the result we wanted.

Step 2: Note that in the previous step, the last component of the vector c^1 did not change during the whole process: we could have chosen $c_J^{(1)} = x_0$. Since the order of those components do not matter, the previous result hold for some $c^{(J)} = (x_0, x_1^{(J)}, ..., x_{J-1}^{(J)})$. That is,

$$\exists \sigma_J, \forall \sigma \in S_J, \sigma \neq \sigma_J \Rightarrow V(\sigma, c^{(J)}) < V(\sigma_J, c^{(J)}).$$

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 finite sum of exponential functions multiplied by scalars where at least one of the scalars is nonzero. 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)).

$\square$

References

Adams, C. P. (2016): "Finite mixture models with one exclusion restriction," *The Econometrics Journal*, 19(2), 150–165.

Aguirregabiria, V., and P. Mira (2013): "Identification of games of incomplete information with multiple equilibria and common unobserved heterogeneity," *University of Toronto Department of Economics Working Paper*, 474.

Arcidiacono, P., and R. A. Miller (2011): "Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity," *Econometrica*, 79(6), 1823–1867.

Athey, S., and P. A. Haile (2007): "Nonparametric approaches to auctions," *Handbook of econometrics*, 6, 3847–3965.

Berry, S., M. Carnall, and P. T. Spiller (1996): "Airline hubs: costs, markups and the implications of customer heterogeneity," Discussion paper, National Bureau of Economic Research.

Berry, S., and E. Tamer (2006): "Identification in models of oligopoly entry," *Econometric Society Monographs*, 42, 46.

Bonhomme, S., K. Jochmans, and J.-M. Robin (2016a): "Estimating multivariate latent-structure models," *The Annals of Statistics*, 44(2), 540–563.

——— (2016b): "Non-parametric estimation of finite mixtures from repeated measurements," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(1), 211–229.

Butucea, C., and P. Vandekerkhove (2014): "Semiparametric mixtures of symmetric distributions," *Scandinavian Journal of Statistics*, 41(1), 227–239.

Cameron, S. V., and J. J. Heckman (1998): "Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males," *Journal of Political economy*, 106(2), 262–333.

Compiani, G., P. Haile, and M. Sant'Anna (2018): "Common Values, Unobserverd Heterogeneity, and Endogenous Entry in U.S. Offshore Oil Lease Auctions," Discussion paper, Yale University, Cowles Foundation CFDP No. 2137.

Compiani, G., and Y. Kitamura (2016): "Using mixtures in econometric models: a brief review and some new results," *The Econometrics Journal*, 19(3), C95–C127.

D'Haultfœuille, X., and P. Février (2015): "Identification of mixture models using support variations," *Journal of Econometrics*, 189(1), 70–82.

Dunford, N., and J. T. Schwartz (1958): *Linear operators part I: general theory*, vol. 7. Interscience publishers New York.

Echenique, F., and I. Komunjer (2009): "Testing models with multiple equilibria by quantile methods," *Econometrica*, 77(4), 1281–1297.

Einmahl, U., and D. M. Mason (2005): "Uniform in bandwidth consistency of kernel-type function estimators," *The Annals of Statistics*, 33(3), 1380–1403.

Ellison, G. (1994): "Theories of cartel stability and the joint executive committee," *The Rand journal of economics*, pp. 37–57.

Feller, W. (1968): *An introduction to probability theory and its applications*, vol. 1. John Wiley & Sons.

Gut, A. (2013): *Probability: a graduate course*, vol. 75. Springer Science & Business Media.

Haile, P., and Y. Kitamura (2018): "Unobserved Heterogeneity in Auctions," *Econometrics Journal*, forthcoming.

Haile, P. A., H. Hong, and M. Shum (2003): "Nonparametric tests for common values at first-price sealed-bid auctions," Discussion paper, National Bureau of Economic Research.

Hall, P., and X.-H. Zhou (2003): "Nonparametric estimation of component distributions in a multivariate mixture," *The annals of statistics*, 31(1), 201–224.

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*, pp. 357–384.

Härdle, W., and O. Linton (1994): "Applied nonparametric methods," *Handbook of econometrics*, 4, 2295–2339.

Heckman, J., and B. Singer (1984): "A method for minimizing the impact of distributional assumptions in econometric models for duration data," *Econometrica: Journal of the Econometric Society*, pp. 271–320.

Heckman, J. J., and C. R. Taber (1994): "Econometric mixture models and more general models for unobservables in duration analysis," *Statistical Methods in Medical Research*, 3(3), 279–299.

Henry, M., Y. Kitamura, and B. Salanié (2010): "Identifying finite mixtures in econometric models," *Discussion Papers*, pp. 0910–20.

——— (2014): "Partial identification of finite mixtures in econometric models," *Quantitative Economics*, 5(1), 123–144.

Hohmann, D., and H. Holzmann (2013a): "Semiparametric location mixtures with distinct components," *Statistics*, 47(2), 348–362.

——— (2013b): "Two-component mixtures with independent coordinates as conditional mixtures: Nonparametric identification and estimation," *Electronic Journal of Statistics*, 7, 859–880.

Horowitz, J. L., and C. F. Manski (1995): "Identification and robustness with contaminated and corrupted data," *Econometrica: Journal of the Econometric Society*, pp. 281–302.

Jewell, N. P. (1982): "Mixtures of exponential distributions," *The annals of statistics*, pp. 479–484.

Jochmans, K., M. Henry, and B. Salanié (2017): "Inference on two-component mixtures under tail restrictions," *Econometric Theory*, 33(3), 610–635.

Kasahara, H., and K. Shimotsu (2009): "Nonparametric identification of finite mixture models of dynamic discrete choices," *Econometrica*, 77(1), 135–175.

Keane, M. P., and K. I. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105, 473–522.

Kiefer, N. M. (1978): "Discrete parameter variation: Efficient estimation of a switching regression model," *Econometrica: Journal of the Econometric Society*, pp. 427–434.

Klein, R. W., and R. P. Sherman (2002): "Shift restrictions and semiparametric estimation in ordered response models," *Econometrica*, 70(2), 663–691.

Lee, L.-F., and R. H. Porter (1984): "Switching Regression Models with Imperfect Sample Separation Information–With an Application on Cartel Stability," *Econometrica: Journal of the Econometric Society*, pp. 391–418.

Lindsay, B. G. (1995): "Mixture models: theory, geometry and applications," in *NSF-CBMS regional conference series in probability and statistics*, pp. i–163. JSTOR.

Manski, C. F. (2003): *Partial identification of probability distributions*. Springer Science & Business Media.

Milgrom, P. R., and R. J. Weber (1982): "A theory of auctions and competitive bidding," *Econometrica: Journal of the Econometric Society*, pp. 1089–1122.

Porter, R. H. (1983): "A study of cartel stability: the Joint Executive Committee, 1880-1886," *The Bell Journal of Economics*, pp. 301–314.

Quandt, R. E. (1972): "A new approach to estimating switching regressions," *Journal of the American statistical association*, 67(338), 306–310.

Rao, B. P. (1992): *Identifiability in stochastic models: characterization of probability distributions*. Academic Press.

Teicher, H. (1961): "Identifiability of mixtures," *The annals of Mathematical statistics*, 32(1), 244–248.

——— (1963): "Identifiability of finite mixtures," *The annals of Mathematical statistics*, pp. 1265–1269.

Tossavainen, T. (2006): "On the zeros of finite sums of exponential functions," *Australian Mathematical Society Gazette*, 33(1), 47.

Van den Berg, G. J. (2001): "Duration models: specification, identification and multiple durations," in *Handbook of econometrics*. Elsevier, vol. 5, pp. 3381–3460.

Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520.

E-mail address: `yuichi.kitamura@yale.edu`

Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520.

E-mail address: `louise.laage@yale.edu`