A A

Identifiability of Finite

Linear Regression Mixtures

Christian Hennig

Preprint–No. 96-06 März 1996

INSTITUT FÜR MATHEMATISCHE STOCHASTIK

BUNDESSTR. 55, D – 20146 HAMBURG

IDENTIFIABILITY OF FINITE LINEAR REGRESSION MIXTURES [1]

Christian Hennig

Abstract

Identifiability is a necessary condition for the existence of consistent estimates for the parameters of mixture models. In this paper the identifiability of finite mixtures of linear regression models with Normal errors is investigated. Three different models are treated: Mixture models with random and fixed independent variables and a model with fixed partition of the data to the mixture components. Sometimes only parts of the unknown parameter values are of interest. "Partial identifiability" is introduced for this purpose.

It turns out that identifiability of finite linear regression mixtures depends on the number of $p-1$-dimensional hyperplanes which one needs to cover the independent variables. Counterexamples and sufficient conditions for identifiability are given for all models.

1 Introduction

In general, a stochastic identifiability problem can be explained as follows:

Definition 1.1 (Identifiability) *Let Ω be an arbitrary parameter space, $\mathcal{P}$ be some space of distributions,*

$$C := (F_\omega)_{\omega\in\Omega} \subseteq \mathcal{P}^\Omega, \tag{1}$$

"$\sim$" be an equivalence relation on Ω. Then C is identifiable w.r.t. "$\sim$" if

$$\forall \omega, \hat\omega \in \Omega : \qquad F_\omega = F_{\hat\omega} \Leftrightarrow \omega \sim \hat\omega. \tag{2}$$

In the context of mixture models the concept of identifiability goes back to Teicher (1961). It means that the same model cannot be constructed by different mixing distributions of members from a given class of distributions. This paper deals with the case of finite mixtures on $I\!R^d$, i.e. the mixing distribution has finite support:

Remark 1.2 (Identifiability: Finite Mixture Models) *Let T be some parameter space with $\sigma-$algebra $\mathcal{T}$ containing all singletons of T, $\Omega = \mathcal{J}(T)$*

[1] Keywords and phrases: Identifiability, Linear Regression, Mixture Model, Fixed Partition Model, Random Carriers Model

the space of distributions on T with finite support, $\mathcal{F} := \{F(\bullet, \theta) : \theta \in T\}$ a set of distribution functions on $I\!R^d$,

$$C = \left(H_J : H_J(x) = \textstyle\int_T F(x,\theta)dJ(\theta), \quad x \in I\!R^d\right)_{J\in\Omega},$$
$$J \sim \hat{J} :\Leftrightarrow J = \hat{J} \quad \forall J, \hat{J} \in \Omega \Leftrightarrow$$
$$\Leftrightarrow \forall J, \hat{J} \in \Omega : \text{supp}(J) = \text{supp}(\hat{J}), \quad J\{\theta\} = \hat{J}\{\theta\} \quad \forall \theta \in \text{supp}(J).$$

Some familiar classes of distributions have been shown to generate identifiable mixtures, e.g. multivariate normal distributions (Yakowitz and Spragins (1968)). It is clear that if C is a parametric class of distributions, the parameters of the mixture, i.e. the mixing distribution J characterized by supp(J) and $J\{\theta\}$ $\forall \theta \in$ supp(J) (the members of supp(J) are called the "mixture components"), can only be estimated consistently if C is identifiable. Note, however, that proofs of the existence of consistent estimators need stronger assumptions than simple identifiability (Yakowitz (1969), Chen (1995)).

There is very few literature on the estimation of finite mixtures of linear regression models. As a reference see DeSarbo and Cron (1988). Especially, there are no proofs of identifiability of the model and, consequently, of consistency of the proposed estimates.

For the classical linear regression case where all errors are independently Normal distributed, some possibilities to model a situation where different components are mixed, are introduced (section 2) and investigated with respect to identifiability: Random independent variables ("carriers"), fixed carriers with Normal mixture for the dependent variable and fixed carriers with fixed but unknown classification of the carriers to the components (fixed partition model).

Identifiability for finite mixtures given in remark 1.2 is only applicable to the case of random carriers. In this setup, the distribution of the carriers is a nuisance parameter. The model is not identifiable if the class of distributions of the carriers is not identifiable. This is especially the case if it contains the empirical distributions which would be used if one wanted to treat a fixed carriers situation by a random carriers model.

So, to treat fixed carriers, the general definition 1.1 will be used. For the case that some parameters are not of interest (e.g. the carrier distribution), we define

Definition 1.3 (Partial identifiability) *With the notation of definition 1.1, C is partially identifiable w.r.t. "$\sim$" if*

$$\forall \omega, \hat{\omega} \in \Omega : \qquad F_\omega = F_{\hat{\omega}} \Rightarrow \omega \sim \hat{\omega}. \tag{3}$$

Remark 1.4 *The various identifiability problems treated in Prakasa Rao (1992) could be covered with the definitions 1.1 and 1.3 if they would be exten-*

ded to the identifiability of single equivalence classes. Especially the concept of "partial identifiability" is a generalization of that given by Prakasa Rao on p. 149.

Li and Sedransk (1988) introduced "P-" and "B-identifiability" for finite mixtures. The former is covered by definition 1.1, the latter additionally demands a bicontinuous correspondence between parameter and distribution.

Section 3 gives some counterexamples: None of the models is identifiable in general, nor even partial identifiable. The crucial point is that the number of mixture components must not exceed the number of hyperplanes one needs to cover the carriers of a single component. In the fixed partition model, the component membership of the data points is modelled as parameter. This parameter is never identifiable.

In section 4, some identifiability results are shown: The random carrier model is identifiable if the carrier distributions give probability 0 on $p-1$-dimensional hyperplanes and the regression parameters are partially identifiable if no finite union of such hyperplanes gets probability 1. The fixed carrier mixture model is identifiable if the number of components is smaller than the number of hyperplanes one needs to cover the carriers, and the regression parameters of the fixed partition model are partially identifiable if the same condition holds for the carriers of each single component.

The estimation theory for mixture models usually assumes the number of components as known. Nevertheless identifiability theory is also relevant for this case, because if identifiability holds for unknown number of components, this of course carries over to the easier case of known component number and on the other hand some of the counterexamples presented below also hold for fixed number of components.

Notation: Φ_{0,σ^2} denotes the c.d.f. of the Normal distribution with mean 0 and variance σ^2, δ_a denotes the c.d.f. of the Dirac measure in a, "$t \leq x$", where t, x are d-dimensional vectors, means componentwise $\leq$. Distributions will be notated with the same letters as their c.d.f., depending on the argument (except Φ and $\mathcal{N}$). $\mathcal{P}_d$ denotes the space of distributions on $\mathbb{R}^d$.

If nothing else is said explicitly, the letter x will be used for the independent variables, which are $p+1$-dimensional points with $p+1$st component equal to 1 (intercept). x^- denotes the first p components of x. This means especially that $\dim\langle\{x_i : i \in I\}\rangle < p+1 \Leftrightarrow x_i^-, i \in I$ lie on a common $p-1$-dimensional hyperplane

$$H := \{x^- \in \mathbb{R}^p : \alpha' x^- = a\}, \quad \mathbb{R}^p \ni \alpha \neq 0, \tag{4}$$

where $\langle A \rangle$ denotes the linear hull of A. Let $\mathcal{H}$ denote the space of $p-1$-dimensional hyperplanes of $\mathbb{R}^p$, $\mathcal{J}(T)$ the space of distributions with finite support on the parameter space T, $S(J) := \text{supp}(J)$ for $J \in \mathcal{J}(T)$.

2 Models for Linear Regression Mixtures

In this section three models for very similar situations are introduced. In all cases we treat mixtures of distributions of y (or (x, y)) where

$$y_i = x_i'\beta + u_i, \qquad u_i \sim \mathcal{N}_{0,\sigma^2} \text{ i.i.d.},$$
$$y_i \in I\!R, \qquad x_i = (x_{i1}, \ldots, x_{ip}, 1) \in I\!R^p \times \{1\},$$
$$\beta \in I\!R^{p+1}, \qquad i = 1, \ldots, n,$$

u_i st. independent of x_i. Here, the $p+1$st component of β denotes the intercept parameter. For most of the results it would be sufficient to assume the $u_i \sim F_\theta, \theta \in T$, where $E(F_\theta) = 0$, all F_θ coming from some distribution family $\mathcal{F}$ which generates identifiable mixtures and contains the associated location families $\{F_\theta(\bullet - t) : t \in I\!R\}$.

The easiest case for identifiability considerations is to assume x_i i.i.d. $\sim$ some c.d.f. G (random carriers). Then it suffices to model a single random vector (x_i, y_i), and the subscript i can be left out.

Model 2.1 (Random carriers)

$$F_J(x, y) = \textstyle\int_T F(x, y, \theta) dJ(\theta), \text{ where}$$
$$F(x, y, \theta) = \textstyle\int 1(t \leq x) \Phi_{0,\sigma^2}(y - t'\beta) dG(t),$$
$$\theta := (\beta, \sigma^2, G) \in T := I\!R^{p+1} \times I\!R_0^+ \times \mathcal{G}, J \in \mathcal{J}(T),$$

$\mathcal{G}$ being some set of $p+1$-dimensional distribution functions where $x_{p+1} \sim \delta_1$. For the sake of simplicity in the following we consider $G \in \mathcal{G}$ as p-dimensional distribution function and $x_{p+1} = 1$ as fixed.

Remark 2.2 *Here, the distribution G of the carriers is treated as parameter. If a $\sigma-$algebra $\mathcal{B}$ on $\mathcal{P}_p$ is defined as follows:*

$$\mathcal{B} := \sigma(\{\{P \in \mathcal{P}_p : P(A) \in B\} : A \in I\!B^p, B \in I\!B\}), \tag{5}$$

one gets from section 12 of Hinderer (1970) that

- *$(\mathcal{P}_p, \mathcal{B})$ is a Standard Borel Space (Theorem 12.13). From this:*
- *$\mathcal{B}$ contains all singletons of $\mathcal{P}_p$ (p. 87),*
- *the mapping*

$$t : (x, P) \mapsto \int u(\bullet, x) dP, \quad (x, P) \in I\!R^q \times \mathcal{P}_p \tag{6}$$

is $I\!B^q \otimes \mathcal{B}$-measurable, if $u : I\!R^{q+p} \mapsto I\!R$ is bounded from above or below (Lemma 12.2). This, of course, is the case if u denotes probabilities or distribution functions.

This means that $\mathcal{J}(T)$ and $\int F dJ$ can be proper defined on $(T, I\!B^{p+1} \otimes \mathcal{B})$.

Definition 2.3 (Equivalence relations for model 2.1)
Let $C := (F_J)_{J \in \Omega}$, F_J defined as in model 2.1, T defined as in model 2.1, Ω and "$\sim$" defined as in remark 1.2.

If one only wants to identify the regression, scale and proportion parameters, and the carrier distribution is not of interest, so that components with the same linear relationship between x and y shall always be treated as one component, Ω is replaced by

$$\Omega^* := \{J \in \mathcal{J}(T) : (\beta, \sigma^2, G) \in S(J), G \neq \hat{G} \Rightarrow (\beta, \sigma^2, \hat{G}) \notin S(J)\}. \tag{7}$$

In this case $\mathcal{G}$ should be assumed to contain all finite mixtures of its members so that replacing Ω by Ω^* does not restrict C. Again,

$$J \sim_* \hat{J} :\Leftrightarrow J = \hat{J} \quad \forall J, \hat{J} \in \Omega^*. \tag{8}$$

If the carrier distributions are not of interest, it seems to be natural to define for partial identifiability on Ω^*:

$$J^{(\beta,\sigma^2,G)} \sim_0 \hat{J}^{(\beta,\sigma^2,G)} :\Leftrightarrow J^{(\beta,\sigma^2)} = \hat{J}^{(\beta,\sigma^2)}. \tag{9}$$

But it will turn out (see example 3.2) that we need an even weaker equivalence relation "$\sim_1$" on Ω^* to prove a partial identifiability result:

$$J \sim_1 \hat{J} :\Leftrightarrow$$
$$\Leftrightarrow \{(\beta, \sigma^2) : (\beta, \sigma^2, G) \in S(J)\} = \{(\beta, \sigma^2) : (\beta, \sigma^2, G) \in S(\hat{J})\}. \tag{10}$$

Random carriers model the situation where the independent variables cannot be determined by the designer of the experiment. This is adequate e.g. in some applications in economics or social sciences where relations in individual properties are investigated and the individuals are chosen by random.

The regression model with fixed carriers is used more frequently, because in most technical applications the independent variables can be controlled. On the contrary to the random carriers case, here we have different distributions for every data point y_i, so that the model has to include their joint distribution for $i \in I$. Here $I = \{1, \ldots, n\}$ is allowed, where n can be interpreted as sample size or as number of pairwise distinct x_i if we follow the idea that the experiment can be repeated arbitrarily frequently with the same x_i. This can also be modelled by $I = I\!N$ (the x_i not assumed to be pairwise distinct) if we imagine that the experiment is also avialable with arbitrary many different x_i. Note that the concept of identifiability does not concern a concrete sample but the theoretical idea of where the sample comes from. So for identifiability it is adequate to model all information which we imagine to be available by infinite sampling. With this consideration we could also allow all possible x_i $(I = I\!R^p)$. We will see that identifiability is given in this case (corollaries

4.5 and 4.7), but on the other hand this is useless if one wants to specify a sample design.

DeSarbo and Cron (1988) use a mixture model for the dependent variable:

Model 2.4 (Fixed carriers, mixture model)

$$(y_i)_{i\in I} \sim \bigotimes_{i\in I} F_{x_i}, \ \ \textit{where}$$

$$F_{x,J}(y) = \int_T \Phi_{0,\sigma^2}(y - x'\beta)dJ(\beta,\sigma^2), \quad T := I\!R^{p+1} \times I\!R_0^+,$$

$I = \{1,\ldots,n\}$ or $I = I\!N$, x, y as in model 2.1, $J \in \mathcal{J}(T)$.

Definition 2.5 (Equivalence relation for model 2.4) *For given $\tilde{x} = (x_i)_{i\in I}$ define*

$$C := \left(F_{\tilde{x},J} : F_{\tilde{x},J} = \bigotimes_{i\in I} F_{x_i,J}\right)_{J\in\Omega}, \tag{11}$$

where $F_{x_i,J}$ is defined as in model 2.4. Further let $\Omega := \mathcal{J}(T)$, T as in model 2.4,

$$J \sim \hat{J} :\Leftrightarrow J = \hat{J} \quad \forall J, \hat{J} \in \Omega. \tag{12}$$

The mixture model can be interpreted as a model for y generated by distribution $\mathcal{N}_{x'\beta,\sigma^2}$ with probability $J\{(\beta,\sigma^2)\}$, so that the probability for y_i to be generated by the mixture component with a certain parameter $(\beta_0,\sigma_0^2) \in S(J)$ is independent of i and x_i. This is not necessarily reasonable.

In changepoint or switching regression situations e.g. the y_i come from different distributions dependent on time or the x-values. This can be modelled in the random carrier situation by choosing a different c.d.f. G for each component. With fixed carriers it can be handlede with the fixed partition model. Here the distribution J is replaced by (unknown) parameters $\gamma(i)$ which give the component membership for every (x_i, y_i). With this formalization we have no specific assumptions about the (in-)dependency between the x_i and the distribution which generates the y_i.

Model 2.6 (Fixed carriers, fixed partition)

$$(y_i)_{i\in I} \sim \bigotimes_{i\in I} F_{x_i,\gamma(i)}, \ \ \textit{where}$$

$$\gamma : I \mapsto I\!R^{p+1} \times I\!R_0^+, \quad |\gamma(I)| < \infty,$$

$$F_{x,\beta,\sigma^2}(y) = \Phi_{0,\sigma^2}(y - x'\beta) \quad \forall (\beta,\sigma^2) \in \gamma(I),$$

$I = \{1,\ldots,n\}$ or $I = I\!N$, x, y as in model 2.1.

Definition 2.7 (Equivalence relations for model 2.6) *For given* $\tilde{x} = (x_i)_{i \in I}$ *let*

$$\Omega := \left\{ \gamma : I \mapsto I\!R^{p+1} \times I\!R_0^+, \quad |\gamma(I)| < \infty \right\}, \tag{13}$$

$$C := \left(F_{\tilde{x},\gamma} : F_{\tilde{x},\gamma} = \bigotimes_{i \in I} F_{x_i,\gamma(i)} \right)_{\gamma \in \Omega}, \tag{14}$$

where the $F_{x_i,\gamma(i)}$ *are defined as in model 2.6. Let* $\gamma_1(i)$ *be the vector of the first* $p+1$ *components (i.e. the regression parameters) and* $\gamma_2(i)$ *be the* $p+2nd$ *component (i.e. the scale parameter) of* $\gamma(i), i \in I$.

Further, for $\gamma, \hat{\gamma} \in \Omega$,

$$\gamma \sim \hat{\gamma} :\Leftrightarrow \gamma = \hat{\gamma}, \tag{15}$$

and for partial identifiability,

$$\gamma \sim_1 \hat{\gamma} :\Leftrightarrow \gamma(I) = \hat{\gamma}(I). \tag{16}$$

3 Examples for Non-identifiability

Titterington, Smith and Makov (1985) write on p. 42: "These positive findings confirm the overall picture that, apart from special cases..., identifiability of classes of finite mixtures is generally assured.".

In linear regression situations one can find such special cases. There are two main reasons for linear regression mixtures to be non-identifiable:

- Non-identifiability because of problems with the carriers (non-identifiability of the carrier distribution in model 2.1 or no unique classification of carriers in model 2.6).
- Non-identifiability because of too much components for too few carrier points.

Example 3.1 (Carriers not identifiable) *If* $\mathcal{G}$ *in definition 2.3 contains non-identifiable finite mixtures of its members, i.e.*

$$G = \int_{\mathcal{G}} P dJ(P) = \int_{\mathcal{G}} P d\hat{J}(P) = \hat{G}, \tag{17}$$

where $J \neq \hat{J} \in \mathcal{J}(\mathcal{G})$, *then* C *is not identifiable w.r.t. "*$\sim$*", since if* $G = \hat{G}$, *then also their products with the conditional distribution of the dependent variables are equal. This happens especially if* $\mathcal{G}$ *contains the set of empirical distributions. The problem does not appear if* Ω *is replaced by* Ω^*.

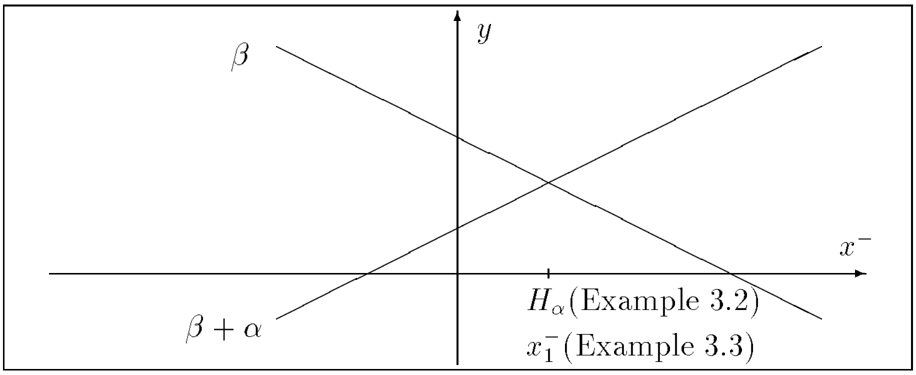
Example 3.2 (Carriers not uniquely assignable) *Replace* Ω *by* Ω^* *in definition 2.3. Then* C *is not identifiable in general, if* $\mathcal{G}$ *contains distributions which give positive probability to a* $p-1$*-dimensional hyperplane*

$$H_\alpha := \{x^- \in I\!R^p : x'\alpha = 0, \quad \alpha \in I\!R^{p+1}\} \tag{18}$$

as in the following picture. (In $I\!R^1$, H_α is a single point.) Then for $x \in H_\alpha$, $x'\beta = x'(\beta+\alpha)$. Let $G_1 \in \mathcal{G}$, where

$$\begin{gathered}\epsilon := G_1(H_\alpha) > 0, \quad G_H(B) := G_1(B|H_\alpha) \forall \text{ Borel sets } B \in I\!B^p, \\ G_2 := \tfrac{1}{1-\epsilon}(G_1 - \epsilon G_H) \text{ (supposed to be } \in \mathcal{G}), \\ J := \tfrac{1}{2-\epsilon}\delta_{(\beta,\sigma^2,G_1)} + \tfrac{1-\epsilon}{2-\epsilon}\delta_{(\beta+\alpha,\sigma^2,G_2)}, \\ \hat{J} := \tfrac{1-\epsilon}{2-\epsilon}\delta_{(\beta,\sigma^2,G_2)} + \tfrac{1}{2-\epsilon}\delta_{(\beta+\alpha,\sigma^2,G_1)}. \text{ Then,} \\ F_J^{(X,Y)} := \tfrac{1}{2-\epsilon}(G_1^X \otimes \mathcal{N}^Y_{X'\beta,\sigma^2}) + \tfrac{1-\epsilon}{2-\epsilon}(G_2^X \otimes \mathcal{N}^Y_{X'(\beta+\alpha),\sigma^2}) = \\ = F_{\hat{J}}^{(X,Y)} := \tfrac{1-\epsilon}{2-\epsilon}(G_2^X \otimes \mathcal{N}^Y_{X'\beta,\sigma^2}) + \tfrac{1}{2-\epsilon}(G_1^X \otimes \mathcal{N}^Y_{X'(\beta+\alpha),\sigma^2}), \\ J \not\sim_* \hat{J}, \quad J \not\sim_0 \hat{J}.\end{gathered}$$

Therefore C is not identifiable w.r.t. "$\sim_*$" and not even partial identifiable w.r.t. "$\sim_0$".



An analogous problem appears in model 2.6:

Example 3.3 (Assignment problems in model 2.6) *C in definition 2.7 is not identifiable w.r.t. "$\sim$". Choose $\alpha \in I\!R^{p+1}$ such that $x_1'\alpha = 0$. Let*

$$\gamma(I) = \{(\beta,\sigma^2),(\beta+\alpha,\sigma^2)\}. \tag{19}$$

Then, $F_{(x_i)_{i\in I},\gamma}$ is equal for $\gamma(1) = (\beta,\sigma^2)$ and $\gamma(1) = (\beta+\alpha,\sigma^2)$. Therefore, the component membership of the x_i is not identifiable.

Now examples will be given where not even the regression parameters are identifiable.

Example 3.4 (Too few carriers) *In model 2.6 one has to demand enough points per component: If $\dim\langle\{x_i : \gamma(i) = (\beta,\sigma^2\}\rangle < p+1$, then of course β_j is not unique. If $|\gamma^{-1}\{(\beta,\sigma^2)\}| = p+1$ is allowed, then if we have $p+1$ mixture components defined by $\gamma(I) = \{(\beta_1,\sigma_0^2),\ldots,(\beta_{p+1},\sigma_0^2)\}$, each containing more than $p+1$ points, a $p+2$nd component can be constructed by changing $\gamma(i)$ to (β_{p+2},σ_0) for one x_i from each component such that the distribution of y_i for any x_i remains unchanged.*

Example 3.5 (Too much components) *Non-identifiability is also possible if the number of carrier points (and even the number of $p-1$-dimensional hyperplanes which one needs to cover the x_i^-) per component is arbitrarily high: Let $p=1, n=s^2, s:=|\gamma(I)|$, choose $x_1,\ldots,x_n$ equidistant (which would not be necessary), say $x_i=(i,1)$. Then the joint distribution of the $(y_i)_{i\in I}$ is defined by*

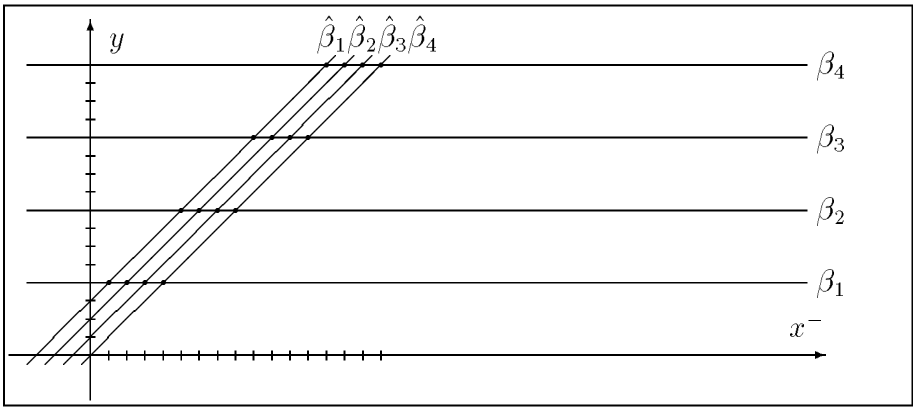
$$F_{x_i}(y_i)=\Phi_{0,\sigma^2}(y_i-js), \quad i\in\{(j-1)s+1,\ldots,js\} \tag{20}$$

simultaneously for the following parameter choices:

$$\gamma(i)=(0,js,\sigma^2) :\Leftrightarrow i\in\{(j-1)s+1,\ldots,js\}, \tag{21}$$

or

$$\hat\gamma(i)=(1,s-j,\sigma^2) :\Leftrightarrow i\in\{ks+j : k\in\{0,\ldots,s-1\}\}. \tag{22}$$



In higher dimensions one can replace the x_i^- by arbitrarily many points on a common $p-1$-dimensional hyperplane to construct examples with s components and arbitrarily more than s^2 points.

To construct examples in model 2.4 is more difficult, because one cannot easy change the assignments of the carrier points to the mixture components. Every mixture component is represented in the distribution of y_i for each carrier x_i. Consequently here the number of points n is simultaneously what we called "the number of points per mixture component" in the previous examples. Nevertheless a counterexample is again possible:

Example 3.6 (Too much components in model 2.4) *Let $p=1, n=3, |S(J)|=3$. Examples with larger $n, |S(J)|$ are possible but much more difficult to construct. For larger p the same remark as in the previous example holds. The carrier points again are equidistant, say $x_1=(0,1), x_2=(1,1), x_3=(2,1)$. Then the joint distribution of the (y_i) is the independent product of the distribution functions*

$$F_{x_1}(y_1)=(\pi\Phi_{0,\sigma^2}+\pi\Phi_{1,\sigma^2}+\pi\Phi_{2,\sigma^2})(y_1), \tag{23}$$

$$F_{x_2}(y_2)=(\pi\Phi_{1,\sigma^2}+\pi\Phi_{\frac{3}{2},\sigma^2}+\pi\Phi_{2,\sigma^2})(y_2), \tag{24}$$

$$F_{x_3}(y_3)=(\pi\Phi_{1,\sigma^2}+\pi\Phi_{2,\sigma^2}+\pi\Phi_{3,\sigma^2})(y_3), \tag{25}$$

$$\sigma^2\geq 0, \pi:=\tfrac{1}{3} \tag{26}$$

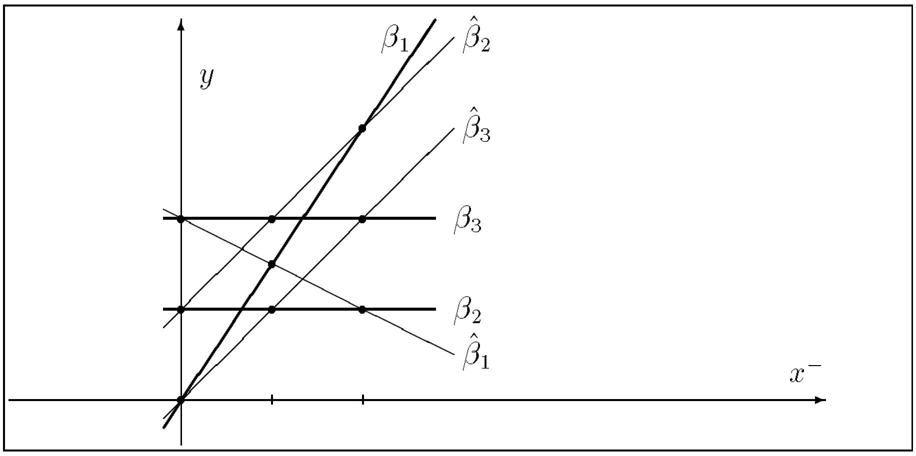
for the following different parameter choices:

$$j = 1, 2, 3 : \sigma_j^2 = \sigma^2, \beta_1 = (\frac{3}{2}, 0), \beta_2 = (0, 1), \beta_3 = (0, 2) \tag{27}$$

and $J\{(\beta_j, \sigma_j^2)\} = \pi$, *or*

$$j = 1, 2, 3 : \hat{\sigma}_j^2 = \sigma^2, \quad \hat{\beta}_1 = (-\frac{1}{2}, 2), \hat{\beta}_2 = (1, 1), \hat{\beta}_3 = (1, 0) \tag{28}$$

and $\hat{J}\{(\beta_j, \sigma_j^2)\} = \pi$.



Remark 3.7 *As one should expect, the random carriers model with empirical distributions is linked to the fixed partition model: From counterexamples against partial identifiability of* C *from definition 2.7 w.r.t. "*$\sim_1$*" one can easily construct counterexamples against partial identifiability of* C *from definition 2.3 w.r.t. "*$\sim_1$*", if*

$$|I| = n < \infty, \tag{29}$$

$$\mathcal{G} \supset \left\{ \sum_{i \in I} \epsilon_i \delta_{x_i} : \epsilon_i \geq 0 \quad \forall i \in I, \sum_{i \in I} \epsilon_i = 1 \right\} \tag{30}$$

and all restrictions to $|\gamma(I)|$ *in* Ω *from definition 2.7 are placed on* $|S(J)|$ *in* Ω *from definition 2.3.*

4 Identifiability results

In this section, some conditions for identifiability and partial identifiability will be derived. The crucial assumption for (partial) identifiability is always that the number of $p-1$-dimensional hyperplanes, which one needs to cover all carriers belonging to a single component, is larger than the number of components.

Firstly, model 2.1 will be investigated where the carrier distributions have infinite support. In the case of finite support (empirical distributions) there is a strong relation to the fixed carriers case illustrated by the remarks at the end of this and the previous section.

The easiest case where identifiability (w.r.t. "$\sim$") is ensured, occurs if we assume the independent variables x^- to be Normal distributed. More generally, we have identifiability w.r.t. "$\sim_*$" if we assume the carrier distributions to give probability 0 to every $p-1$-dimensional hyperplane. Partial identifiability w.r.t. "$\sim_1$" can be obtained if the carrier distributions are not concentrated on a finite number of hyperplanes.

Theorem 4.1 (Normal distributed carriers) *In model 2.1 with the notations of definition 2.3, if*

$$\mathcal{G} = \{\mathcal{N}_{(a,SS')} : a \in I\!R^p, S \in I\!R^{p^2}\} \tag{31}$$

then C is identifiable w.r.t. "$\sim$".

Proof: Here, model 2.1 is a reparametrization of a finite mixture model of $p+1$-variate Normal distributions. The latter is identifiable by Yakowitz and Spragins (1968). (Note that $\mathcal{G}$ does not contain any finite mixture of its members with more than one component, so that we are not in the situation of example 3.1.)

Theorem 4.2 (No mass on lower dimensional hyperplanes) *In model 2.1 with the notations of definition 2.3, if*

$$\mathcal{G} \subseteq \{P : \mathcal{P}_p \ni P, P(H) = 0 \quad \forall H \in \mathcal{H}\} \tag{32}$$

then C is identifiable w.r.t. "$\sim_*$".

Proof: Only "$\Rightarrow$" (i.e. partial identifiability) is to show, since J contains all information to define of F_J.

We use: If $P^{(X,Y)} = P^X \otimes P^{(Y|X)}$, $P^{(X,Z)}$ analogously, $P^{(Y|X=x)}$ and $P^{(Z|X=x)}$ defined on σ-algebras $\mathcal{B}$, where

$$\exists E \text{ countable: } \sigma(E) = \mathcal{B}, \tag{33}$$

then

$$P^{(X,Y)} = P^{(X,Z)} \Leftrightarrow \exists A \subseteq \{x : P^{(Y|X=x)} = P^{(Z|X=x)}\} : P^X(A) = 1 \tag{34}$$

(by e.g. Gänssler / Stute (1977), p. 197).

Now let $s := |S(J)|, \hat{s} := |S(\hat{J})|$, where

$$S(J) = \{(\beta_i, \sigma_i^2, G_i), i = 1, \ldots, s\}, \quad \pi_i := J\{(\beta_i, \sigma_i^2, G_i)\}, \tag{35}$$
$$S(\hat{J}) = \{(\beta_i, \sigma_i^2, G_i), i = 1, \ldots, \hat{s}\}, \quad \hat{\pi}_i := \hat{J}\{(\hat{\beta}_i, \hat{\sigma}_i^2, \hat{G}_i)\}. \tag{36}$$

With this notation, we have

$$F_J = \sum_{i=1}^{s} \pi_i F(\bullet, \beta_i, \sigma_i^2, G_i), \tag{37}$$

F_J defined as in model 2.1, analogously $F_{\hat{J}}$. Now,

$$F_J = F_{\hat{J}} \Rightarrow J = \hat{J} \tag{38}$$

has to be shown. Define

$$\mu := \sum_{i=1}^{s} G_i + \sum_{i=1}^{\hat{s}} \hat{G}_i, \tag{39}$$
$$F^{(X,Y)} = F_J = F_{\hat{J}}, \quad g_i(x^-) := \tfrac{dG_i}{d\mu}(x^-), \quad \hat{g}_i(x^-) := \tfrac{d\hat{G}_i}{d\mu}(x^-). \tag{40}$$

With that,

$$F^{Y|X=x} = \sum_{i=1}^{s} \pi_i \frac{g_i(x^-)}{\sum_{j=1}^{s} \pi_j g_j(x^-)} \mathcal{N}_{(x'\beta_i, \sigma_i^2)}, \tag{41}$$

and with $F^{(X,Z)} := F_{\hat{J}}$, by analogy,

$$F^{Z|X=x} = \sum_{i=1}^{\hat{s}} \hat{\pi}_i \frac{\hat{g}_i(x^-)}{\sum_{j=1}^{\hat{s}} \hat{\pi}_j \hat{g}_j(x^-)} \mathcal{N}_{(x'\hat{\beta}_i, \hat{\sigma}_i^2)}. \tag{42}$$

Let

$$\begin{gathered} M := \{x : \forall j, k \in \{1, \ldots, s\}, l, m \in \{1, \ldots, \hat{s}\} : \\ x'\beta_j = x'\beta_k \Rightarrow \beta_j = \beta_k, \quad x'\beta_j = x'\hat{\beta}_l \Rightarrow \beta_j = \hat{\beta}_l, \\ x'\hat{\beta}_l = x'\hat{\beta}_m \Rightarrow \hat{\beta}_l = \hat{\beta}_m\}. \end{gathered} \tag{43}$$

M is complement of a finite union of elements of $\mathcal{H}$, therefore from the assumption to $\mathcal{G}$,

$$F^X = \sum_{i=1}^{s} \pi_i G_i \Rightarrow F^X(M) = 1. \tag{44}$$

For $x \in M$, all $(x'\beta_i, \sigma_i^2)$, $i = 1, \ldots, s$, are pairwise distinct, since all (β_i, σ_i^2), $i = 1, \ldots, s$, are pairwise distinct by $J \in \Omega^*$ as in definition 2.3. For $x \in M$, let J_x be an empirical distribution on $I\!R \times I\!R_0^+$ defined by

$$J_x\{(x'\beta_i, \sigma_i^2)\} := \pi_i \tfrac{g_i(x^-)}{\sum_{j=1}^{s} \pi_j g_j(x^-)}, \quad i = 1, \ldots, s, \tag{45}$$
$$S(J_x) = \{(x'\beta_i, \sigma_i^2) : i = 1, \ldots, s, \quad g_i(x^-) > 0\} \tag{46}$$

such that with (41),

$$F^{Y|X=x} = \int_{\boldsymbol{R}\times\boldsymbol{R}_0^+} \mathcal{N}_\theta dJ_x(\theta), \tag{47}$$

and, with $\hat{J}_x$ defined analogously to (45),

$$F^{Z|X=x} = \int_{\boldsymbol{R}\times\boldsymbol{R}_0^+} \mathcal{N}_\theta d\hat{J}_x(\theta) \tag{48}$$

by (42). Thus, by $F_J = F_{\hat{J}}$ and (34),

$$\begin{aligned} \exists N \subseteq M, \quad F^X(N) = 1: \quad x \in N \Rightarrow F^{Y|X=x} = F^{Z|X=x} \\ \text{such that } J_x = \hat{J}_x, \end{aligned} \tag{49}$$

since Normal distributions are identifiable w.r.t. "$\sim$" as defined in remark 1.2 (see e.g. Titterington, Smith and Makov (1985)). Now,

$$\begin{aligned} &\forall i \in \{1,\ldots,s\}, j \in \{1,\ldots,\hat{s}\}: \quad G_i(N) = \hat{G}_j(N) = 1 \Rightarrow \\ &\Rightarrow \exists x(i) \in N, \hat{x}(j) \in N: \quad g_i(x(i)^-) > 0, \quad \hat{g}_j(\hat{x}(j)^-) > 0. \end{aligned} \tag{50}$$

For given $i \in \{1,\ldots,s\}$, (49) implies

$$\exists j \in \{1,\ldots,\hat{s}\}: x(i)'\beta_i = x(i)'\hat{\beta}_j, \quad \sigma_i^2 = \hat{\sigma}_j^2. \tag{51}$$

The definition of M yields:

$$(\beta_i, \sigma_i^2) \in \{(\hat{\beta}_j, \hat{\sigma}_j^2) : j \in \{1,\ldots,\hat{s}\}\}, \tag{52}$$

and by the same argument applied to $\hat{x}(j) \quad \forall j \in \{1,\ldots,\hat{s}\}$:

$$(\hat{\beta}_j, \hat{\sigma}_j^2) \in \{(\beta_i, \sigma_i^2) : i \in \{1,\ldots,s\}\}. \tag{53}$$

Therefore

$$\{(\beta_i, \sigma_i^2) : i \in \{1,\ldots,s\}\} = \{(\hat{\beta}_j, \hat{\sigma}_j^2) : j \in \{1,\ldots,\hat{s}\}\}. \tag{54}$$

The definition of Ω^* in definition 2.3 yields pairwise distinctness of the elements, therefore $s = \hat{s}$ and w.l.o.g we can assume $(\beta_i, \sigma_i^2) = (\hat{\beta}_i, \hat{\sigma}_i^2), \quad i = 1,\ldots,s$. Further, for $i = 1,\ldots,s$, there are unique G_i and $\pi_i = J\{(\beta_i, \sigma_i^2, G_i)\}$ for given (β_i, σ_i^2) such that

$$(\beta_i, \sigma_i^2, G_i) \in S(J), \tag{55}$$

analogously $\hat{G}_i$, $\hat{\pi}_i$. Define for $x \in N, i = 1,\ldots,s$:

$$\epsilon_i(x) := J_x\{(x'\beta_i, \sigma_i^2)\}, \quad \hat{\epsilon}_i(x) := \hat{J}_x\{(x'\beta_i, \sigma_i^2)\} \tag{56}$$

as defined in (45). Then, by (49),

$$\forall x \in N : \epsilon_j(x) = \hat{\epsilon}_j(x). \tag{57}$$

With this and $F^X(N) = 1$,

$$\begin{aligned}
\forall i = 1, \ldots, s: \int_N g_i(x^-) d\mu(x^-) = 1 \Rightarrow \\
\Rightarrow \pi_i = \int_N \pi_i g_i(x^-) d\mu(x^-) = \int_N \epsilon_i(x) \sum_{j=1}^{s} \pi_j g_j(x^-) d\mu(x^-) = \\
= \int_N \epsilon_i(x) dF^X(x) = \int_N \hat{\epsilon}_i(x) dF^X(x) = \hat{\pi}_i. \\
\text{Last, for } x \in N: g_i(x^-) = \tfrac{\epsilon_i(x)}{\pi_i} \sum_{j=1}^{s} \pi_j g_j(x^-) = \\
= \tfrac{\hat{\epsilon}_i(x)}{\hat{\pi}_i} \tfrac{dF^X}{d\mu}(x^-) = \hat{g}_i(x^-) \Rightarrow G_i = \hat{G}_i \Rightarrow \\
\Rightarrow J = \hat{J}.
\end{aligned}$$

Theorem 4.3 (Random carriers, partial identifiability) *In model 2.1 with the notations of definition 2.3, if*

$$\begin{aligned}
\mathcal{G} \subseteq \mathcal{Q}, \quad \mathcal{Q} \subset \mathcal{P}_p, \ \textit{where} \\
\forall Q \in \mathcal{Q}, m \in I\!N, A = \bigcup_{i=1}^{m} H_i: \quad Q(A) < 1 \qquad (58)
\end{aligned}$$

for arbitrary $p-1$-dimensional hyperplanes $H_i \in \mathcal{H}, i = 1, \ldots, m$, then C is partially identifiable w.r.t. "$\sim_1$".

Proof: The proof is almost equal to the proof of Theorem 4.2. The only difference is that here the assumptions only imply $F^X(M) > 0$ with M defined in (43). Therefore, instead of (49),

$$\begin{aligned}
\exists N \subseteq M, \quad F^X(N) = F^X(M) > 0: x \in N \Rightarrow F^{Y|X=x} = F^{Z|X=x} \\
\text{such that } J_x = \hat{J}_x \qquad (59)
\end{aligned}$$

and, replacing (50),

$$\begin{aligned}
\forall i \in \{1, \ldots, s\}, j \in \{1, \ldots, \hat{s}\}: G_i, \hat{G}_j \in \mathcal{Q} \Rightarrow \\
\Rightarrow 0 < G_i(M) = G_i(N), \quad 0 < \hat{G}_j(M) = \hat{G}_j(N) \Rightarrow \\
\Rightarrow \exists x(i) \in N, \hat{x}(j) \in N: \quad g_i(x(i)^-) > 0, \quad \hat{g}_j(\hat{x}(j)^-) > 0. \qquad (60)
\end{aligned}$$

This only suffices to show

$$\{(\beta_i, \sigma_i^2): i \in \{1, \ldots, s\}\} = \{(\hat{\beta}_j, \hat{\sigma}_j^2): j \in \{1, \ldots, \hat{s}\}\}, \qquad (61)$$

that is $J \sim_1 \hat{J}$.

Now we turn to the fixed carriers models 2.4 and 2.6. In the case of finite I we have to place restrictions to the parameter space to get identifiability results. The following theorems need only restrictions to the number of components and in the case of model 2.6 to the component assignment of the carriers. Note that the conditions below implicitly exclude the case that there are components with less than p points, since otherwise the values h ($h(j)$ respectively) defined below would be equal to 1.

For an infinite number of carriers $|I|$, the restrictions on the number of components vanish. This is formulated in the corollaries. The results are familiar to the random carriers case: (Partial) identifiability holds, if the domain $\{x_i : i \in I\}$ of the carriers is a set which has probability 1 under the carrier distributions of the previous theorems.

Theorem 4.4 (Model 2.4) *In model 2.4 with the notations of definition 2.5, if $|S(J)| < h$, where*

$$h := \min_q \left\{ \{x_i^- : i \in I\} \subseteq \bigcup_{i=1}^q H_i : H_i \in \mathcal{H} \right\}, \tag{62}$$

then C is identifiable w.r.t. "$\sim$".

Proof: Only "$\Rightarrow$" (i.e. partial identifiability) is to show, since J contains all information to define $F_{x,J}$. *Let* $F_{x,J} = F_{x,\hat{J}}$ $J \neq \hat{J}$, w.l.o.g.

$$|S(J)| \geq |S(\hat{J})|, \quad J\{(\beta_1, \sigma_1^2)\} \neq \hat{J}\{(\beta_1, \sigma_1^2)\}. \tag{63}$$

We have

$$\begin{aligned} \forall(\beta, \sigma^2) \in S(J) \quad \exists i(\beta) \in I \quad \forall(\hat{\beta}, \hat{\sigma}^2) \in S(\hat{J}) : \\ x'_{i(\beta)}\beta = x'_{i(\beta)}\hat{\beta} \Rightarrow \beta = \hat{\beta}, \end{aligned} \tag{64}$$

since if this were not the case for some $(\beta, \sigma^2) \in S(J)$ then

$$\bigcup_{(\hat{\beta}, \hat{\sigma}^2) \in S(\hat{J}) : \hat{\beta} \neq \beta} \{x^- : x'\beta = x'\hat{\beta}\} \supset \{x_i^- : i \in I\} \Rightarrow h \leq |S(\hat{J})| \tag{65}$$

in contradiction to $|S(J)| < h$. Let $t := x_{i(\beta_1)}$.

Now, $F_{x,J} = F_{x,\hat{J}}$ implies

$$\begin{aligned} F_{t,J} = \int_T \mathcal{N}_{(t'\beta, \sigma^2)} dJ(\beta, \sigma^2) = \\ = F_{t,\hat{J}} = \int_T \mathcal{N}_{(t'\beta, \sigma^2)} d\hat{J}(\beta, \sigma^2). \end{aligned} \tag{66}$$

The identifiability of finite Normal mixtures yields

$$(t'\beta_1, \sigma_1^2) \in \{(t'\hat{\beta}, \hat{\sigma}^2) : (\hat{\beta}, \hat{\sigma}^2) \in S(\hat{J})\}. \tag{67}$$

Because of the definiton of t, by (64)

$$(\beta_1, \sigma_1^2) \in S(\hat{J}). \tag{68}$$

Further, the definition of t implies that

$$\forall S(\hat{J}) \ni (\hat{\beta}, \hat{\sigma}^2) \neq (\beta_1, \sigma_1^2) : (t'\hat{\beta}, \hat{\sigma}^2) \neq (t'\beta_1, \sigma_1^2), \tag{69}$$

so that, with the identifiability of the Normal mixture at t,

$$\begin{aligned} \hat{J}\{(\beta_1, \sigma_1^2)\} = \hat{J}\{(\hat{\beta}, \hat{\sigma}^2) : (t'\hat{\beta}, \hat{\sigma}^2) = (t'\beta_1, \sigma_1^2)\} = \\ = J\{(\beta, \sigma^2) : (t'\beta, \sigma^2) = (t'\beta_1, \sigma_1^2)\}. \end{aligned} \tag{70}$$

By (63), $\hat{J}\{(\beta_1, \sigma_1^2)\} \neq J\{(\beta_1, \sigma_1^2)\}$, therefore

$$\exists S(J) \ni (\beta_2, \sigma_2^2) \neq (\beta_1, \sigma_1^2) : (t'\beta_2, \sigma_2^2) = (t'\beta_1, \sigma_1^2). \tag{71}$$

Applying the same argument to $x_{i(\beta_2)}$, we get

$$(\beta_2, \sigma_2^2) \in S(\hat{J}). \tag{72}$$

This is in contradiction to (69) and (71), so that $J = \hat{J}$.

Corollary 4.5 (Model 2.4, no restrictions) *In model 2.4 with the notations of definition 2.5, if*

$$\forall m \in I\!N, A = \bigcup_{i=1}^{m} H_i : \{x_i^- : i \in I\} \not\subset A \tag{73}$$

for arbitrary $p-1$-dimensional hyperplanes $H_i \in \mathcal{H}, i = 1, \ldots, m$, then C is identifiable w.r.t. "$\sim$".

Proof: (73) $\Leftrightarrow h = \infty$.

Theorem 4.6 (Model 2.6) *In model 2.6 with the notations of definition 2.7, if $|\gamma(I)| < \min_{(\beta,\sigma^2)\in\gamma(I)} h(\beta, \sigma^2)$, where*

$$h(\beta, \sigma^2) := \min_q \left\{ \{x_i^- : i \in I, \gamma(i) = (\beta, \sigma^2)\} \subseteq \bigcup_{i=1}^{q} H_i : H_i \in \mathcal{H} \right\}, \tag{74}$$

then C is partially identifiable w.r.t. "$\sim_1$".

Proof: Let

$$\gamma \not\sim_1 \hat{\gamma} \in \Omega, \quad F_{x,\gamma} = F_{x,\hat{\gamma}}, \tag{75}$$

$$\text{w.l.o.g. } |\gamma(I)| \geq |\hat{\gamma}(I)|, \quad \gamma(I) \ni (\beta_0, \sigma_0^2) \notin \hat{\gamma}(I). \tag{76}$$

Then,

$$\forall (\hat{\beta}, \hat{\sigma}^2) \in \hat{\gamma}(I): \quad \dim\langle X_{(\hat{\beta},\hat{\sigma}^2)} \rangle < p+1, \tag{77}$$

where $X_{(\hat{\beta},\hat{\sigma}^2)} := \{x_i : i \in I, \gamma(i) = (\beta_0, \sigma_0^2), \hat{\gamma}(i) = (\hat{\beta}, \hat{\sigma}^2)\}$, will be shown. This means that

$$\begin{aligned} \forall (\hat{\beta}, \hat{\sigma}^2) \in \hat{\gamma}(I) \quad \exists H_{(\hat{\beta},\hat{\sigma}^2)} \in \mathcal{H}: \quad X_{(\hat{\beta},\hat{\sigma}^2)} \subset H_{(\hat{\beta},\hat{\sigma}^2)} \Rightarrow \\ \Rightarrow h(\beta_0, \sigma_0^2) \leq |\hat{\gamma}(I)| \leq |\gamma(I)| \end{aligned} \tag{78}$$

in contradiction to $|\gamma(I)| < \min\{h(\beta, \sigma^2)\}$. Thus,

$$F_{x,\gamma} = F_{x,\hat{\gamma}} \Rightarrow \gamma \sim_1 \hat{\gamma}. \tag{79}$$

Proof of (77):

$$\begin{aligned} &[x_i \in X_{(\hat{\beta},\hat{\sigma}^2)} \Leftrightarrow F_{x_i,\gamma(i)} = F_{x_i,\hat{\gamma}(i)} \Leftrightarrow \\ &\Leftrightarrow \mathcal{N}_{(x_i'\beta_0,\sigma_0^2)} = \mathcal{N}_{(x_i'\hat{\beta},\hat{\sigma}^2)}] \Rightarrow \\ &\Rightarrow \forall x_i \in X_{(\hat{\beta},\hat{\sigma}^2)}: x_i'\beta_0 = x_i'\hat{\beta}, \sigma_0^2 = \hat{\sigma}^2 \Rightarrow \\ &\Rightarrow (\beta_0, \sigma_0^2) = (\hat{\beta}, \hat{\sigma}^2) \text{ or } \dim\langle X_{(\hat{\beta},\hat{\sigma}^2)} \rangle < p+1. \end{aligned}$$

Corollary 4.7 (Model 2.6, no restrictions on $|\gamma(I)|$) *In model 2.6 with the notations of definition 2.7, if*

$$\forall m \in I\!N, (\beta, \sigma^2) \in \gamma(I), A = \bigcup_{i=1}^{m} H_i : \{x_i^- : i \in I, \gamma(i) = (\beta, \sigma^2)\} \not\subset A \tag{80}$$

for arbitrary $p-1$-dimensional hyperplanes $H_i \in \mathcal{H}, i = 1, \ldots, m$, then C is partially identifiable w.r.t. "$\sim_1$".

Proof: (80) $\Leftrightarrow \min_{(\beta,\sigma^2)\in\gamma(I)} h(\beta, \sigma^2) = \infty$.

Remark 4.8 *Note that the assumption (80) is a restriction to the possible parameters, which cannot be verified by considering only the theoretical design $\{x_i : i \in I\}$. This is contrary to corollary 4.5.*

Remark 4.9 *Also the results for finite I can be written in terms of the random carrier model. For example one can show straightforward:*

Identifiability of model 2.4 w.r.t. "$\sim$" as in definition 2.5 is equivalent to identifiability of model 2.1 w.r.t. "$\sim$" as in definition 2.3, where

$$\mathcal{G} = \left\{ \sum_{i \in I} \epsilon_i \delta_{x_i} \right\} \tag{81}$$

for arbitrary but fixed $\epsilon_i > 0, \sum_{i \in I} \epsilon_i = 1$. (That is, $\mathcal{G}$ contains only one element here.)

References

[1995] Chen, J.: *Optimal rate of convergence for finite mixture models,* Ann. Statist. 23 (1995), p. 221-233

[1988] DeSarbo, W. S. and Cron, W. L.: *A maximum likelihood methodology for clusterwise linear regression,* Journal of Classification 5 (1988), p. 249-282

[1977] Gänssler, P. and Stute, W.: *Wahrscheinlichkeitstheorie,* Springer, Berlin 1977

[1970] Hinderer, K.: *Foundations of Non-stationary Dynamic Programming with Discrete Time Parameter,* Springer, Berlin 1970

[1988] Li, L. A. and Sedransk, N.: *Mixtures of distributions: A topological approach,* Ann. Statist. 16 (1988), p. 1623-1634

[1992] Prakasa Rao, B. L. S.: *Identifiability in Stochastic Models: Characterizations of Probability Distributions,* Academic Press, Boston 1992

[1961] Teicher, H.: *Identifiability of mixtures,* Ann. Math. Statist. 34 (1961), p. 244-248

[1985] Titterington, D. M., Smith, A. F. M., Makov, U. E.: *Statistical Analysis of Finite Mixture Distributions,* John Wiley and Sons, Chichester 1985

[1969] Yakowitz, S. J.: *A consistent estimator for the identification of finite mixtures,* Ann. Math. Statist. 40, p. 1728-1735

[1968] Yakowitz, S. J. and Spragins, J. D.: *On the identifiability of finite mixtures,* Ann. Math. Statist. 39 (1968), p. 209-214

Christian Hennig
Institut für Mathematische Stochastik
Bundesstr. 55
D-20146 Hamburg
Germany
e-mail: hennig@math.uni-hamburg.de