When is a nonlinear mixed-effects model identifiable ?

O. G. Nunez and D. Concordet

Universidad Carlos III de Madrid and Ecole Nationale Vétérinaire de Toulouse

Abstract

We consider the identifiability problem in the nonlinear mixed-effects model $Y_i = m(\theta_i) + \varepsilon_i$, with $\varepsilon_i \sim_{\text{iid}} N(0, \sigma^2 \mathbf{I}_n)$ and $\theta_i \sim_{\text{iid}} Q$, $i = 1, \ldots, N$, where σ^2 and Q are unknown variance and probability measure. We give several explicit conditions on m which ensure the identifiability of $\left(Q, \sigma^2\right)$ from the common distribution of the observed vectors Y_i. Remarkably, one of these conditions fits the intuition : the model is identifiable as soon as m is injective and $n \geq \dim(\theta_i) + 1$. Even if the latter condition is necessary for gaussian linear models, it is not for general nonlinear models. Three classical pharmacokinetic models are used to illustrate the three different conditions of their identifiability.

Keywords: Identifiability of measures; Nonlinear mixed-effects models; Mixture of distributions.

1 Introduction.

This paper deals with the identifiability problem in a version of the nonlinear mixed-effects model proposed by Lindstrom and Bates (1990). This kind of models is often used to analyze longitudinal data. Assume that the vector of observations on the ith subject is modeled as

$$Y_i = m(\theta_i) + \varepsilon_i, \ i = 1, \ldots, N \tag{1}$$

where the error terms ε_i are independent and identically distributed (iid) with a common normal distribution $N(0, \sigma^2 \mathbf{I}_n)$. The vector function $m = (m_j)_{1 \leq j \leq n}$, is assumed measurable and known given the experimental design. The mean vector function m is not necessarily linear; its argument is a parameter $\theta_i \in \mathbb{R}^q$ which describes the response curve of the ith subject. In order to analyze the between subjects variability, the θ_i are assumed to be drawn independently from a latent distribution function Q lying in the set $\mathcal{P}(\mathbb{R}^q)$ of probability measures with support in $\mathbb{R}^q$. So Q and σ^2 determine the probability distribution of the observations from (1); the identifiability property guarantees that, in turn, this distribution determines Q and σ^2. This property is necessary to obtain a consistent estimator of the unknown parameter (Q, σ^2). If no parametric assumptions are made about the shape of Q, the most common way to estimate this latent distribution is to use the nonparametric maximum likelihood estimator (Kiefer and Wolfowitz 1956) which has at most N support points (Lindsay 1995).

Liu and Taylor (1989), Stefansky and Carroll (1990), Zhang (1990), Fan (1991), and Cordy and Thomas (1997) addressed the estimation of Q using deconvolution techniques, when the function m is linear. Most of the papers cited above propose kernel estimators for Q and compute their optimal rate of convergence for some specific error distribution. This rate appears to be very slow for normal errors. Fan (1992), however, shows that if σ^2 is not too large, such a method can still be practical.

The identifiability problem of (Q, σ^2) is related to the identifiability of mixtures of distributions which was addressed by many authors including Teicher (1960, 1961, and 1967), Barndorff-Nielsen (1965), Bruni and Koch (1985), Li and Sedransk (1988), and Pfanzagl (1994). Barndorff Nielsen (1965) provided sufficient conditions for the identifiability of mixtures of nat-

ural exponential families, but these conditions are generally too strong for the curve gaussian family considered in this paper. Bruni and Koch (1985) dealt with gaussian mixtures when Q has a compact support in $\mathbb{R}^q$ and the function m is unknown. Li and Sedransk (1988) studied the identifiability of finite mixtures (Q has a finite support) and Pfanzagl (1994) considered a general case of non identifiability for some parameter of the mixture density function when the mixing distribution Q is completely unknown.

This paper is organized as follow. In the next section, several explicit conditions on the functions m ensuring the identifiability of the mixing distribution Q and the variance parameter σ^2, are given. As examples, conditions on the experimental design which ensure the identifiability of three classical pharmacokinetic models are derived in section 3.

2 The result.

First, let us define more precisely the identifiability we want to prove.

Definition 1 *Let $P_{Y|Q,\sigma^2}$ be the common distribution of the vectors Y_i in model (1). The parameter (Q, σ^2) is identifiable from $P_{Y|Q,\sigma^2}$, if for every Q, Q_0 in $\mathcal{P}(\mathbb{R}^q)$ and σ^2, σ_0^2 in $(0, \infty)$,*

$$P_{Y|Q,\sigma^2} = P_{Y|Q_0,\sigma_0^2} \Longleftrightarrow \left(Q, \sigma^2\right) = \left(Q_0, \sigma_0^2\right)$$

To study the identifiability of (Q, σ^2) in the model (1), the problem of identifiability in a simple transformation model is first considered. Suppose that for every draw θ from a probability distribution Q, one observes

$$Z = m(\theta), \tag{2}$$

in which $m : \mathbb{R}^q \to \mathbb{R}^n$ is a fixed and known Borel measurable function. Let $P_{Z|Q}$ be the probability distribution of Z. The following result gives a necessary and sufficient condition on m for Q to be identifiable from $P_{Z|Q}$.

Lemma 2 *The probability distribution Q is identifiable from $P_{Z|Q}$, if and only if the function m is injective in $\mathbb{R}^q$.*

Proof. Clearly it is necessary for m to be injective. If m was not injective, there would exist at least two distinct points θ and θ_0 in $\mathbb{R}^q$, such that $m(\theta) = m(\theta_0)$. So, choosing for instance $Q = \delta_\theta$ and $Q_0 = \delta_{\theta_0}$, where δ_θ is the Dirac point mass at θ would lead to $P_{Z|Q} = P_{Z|Q_0}$ with $Q \neq Q_0$.
The injectivity of m is also sufficient. Let C be a Borel set of $\mathbb{R}^n$. By definition of $P_{Z|Q}$, we have that $P_{Z|Q}(C) = Q\left(\{m(\theta) \in C\}\right)$. Thus, the knowledge of $P_{Z|Q}$ induces the knowledge of Q on the events $\{m(\theta) \in C\}$. Now, according to a theorem of Kuratowski (see Parthasarathy (1967), p. 22), the image of a Borel subset of a complete separable metric space under a one-to-one measurable map of that subset into a separable metric space is a Borel subset of the second space so that the map is an isomorphism between the two Borel subsets. The function m is injective and measurable. Thus, for all Borel B in $\mathbb{R}^q$, the set $m(B)$ is Borel in $\mathbb{R}^n$, and the sets $\{m(\theta) \in m(B)\}$ and B are equal. It follows that Q is known on any Borel B in $\mathbb{R}^q$. ■

This lemma can be slightly weakened with an assumption on the absolute continuity of Q.

Corollary 3 *Assume that Q is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^q$. The probability distribution Q is identifiable from $P_{Z|Q}$, if and only if the function m is injective Lebesgue-almost everywhere.*

Proof. Indeed, if $m\,|_{N^c}$ is injective for a Borel set $N \subset \mathbb{R}^q$ with Lebesgue measure 0, then the Kuratowski theorem still gives unicity of laws on N^c,

because a Borel set of N^c is a Borel set of the whole $\mathbb{R}^q$, which is complete. Unicity of absolute continuous laws on the whole space follows. ■

In the model considered in the introduction, the observed transformations of the θ is contaminated by an error terms ε, which are normally distributed. So, let us assume that for every draw θ from Q, one observes

$$Y = m(\theta) + \varepsilon,$$

where ε is a random vector in $\mathbb{R}^n$ such that θ and ε are independent and the components of ε are iid according to a normal distribution with mean 0 and variance $\sigma^2 I_n$ for some fixed unknown $\sigma^2 > 0$.
Let $P_{Y|Q,\sigma^2}$ be the probability distribution of Y. The following lemma gives a condition that allows identifiability of Q when σ^2 is identifiable or known. Note that σ^2 is identifiable when for every Q, Q_0 in $\mathcal{P}(\mathbb{R}^q)$ and every σ^2, σ_0^2 in $(0, \infty)$, $P_{Y|Q,\sigma^2} = P_{Y|Q_0,\sigma_0^2}$ implies that $\sigma^2 = \sigma_0^2$.

Lemma 4 *Assume σ^2 is identifiable or known. Then, Q is identifiable from $P_{Y|Q,\sigma^2}$ if and only if m is injective.*

Proof. Since $Z = m(\theta)$ and ε are independent, we have the identity $Ee^{i\langle\xi,Y\rangle} = Ee^{i\langle\xi,Z\rangle}Ee^{i\langle\xi,\varepsilon\rangle}$ for the Fourier-Stieltjes (FS) transforms. From the identifiability of σ^2 and the fact that $Ee^{i\langle\xi,\varepsilon\rangle}$ is unequal to 0 for every ξ in $\mathbb{R}^n$, it follows that $\xi \mapsto Ee^{i\langle\xi,Z\rangle}$ is determined by $\xi \mapsto Ee^{i\langle\xi,Y\rangle}$, that is, $P_{Z|Q}$ is identifiable from $P_{Y|Q,\sigma^2}$. Now, according to Lemma **2**, if m is injective (and measurable), Q is identifiable from $P_{Z|Q}$. But, since $P_{Z|Q}$ is itself identifiable from $P_{Y|Q,\sigma^2}$, it follows that Q is identifiable from $P_{Y|Q,\sigma^2}$. ■

The latter lemma shows that when σ^2 is identifiable, the injectivity of m is a necessary and sufficient condition for identifiability of Q. It remains to identify situations where identifiability of σ^2 holds. Actually, σ^2 is identifiable

when the observation of Y allows to separate the conditional mean $m(\theta)$ and the error term ε. Since only the distribution of Y is observed, situations where σ^2 is identifiable occurs when the distributions of $m(\theta)$ and ε do not weight the space in the same way. Three of these situations are described hereafter.

(i) There exist some components of m, $m^{\#} = (m_{j_1}, m_{j_2}, \dots m_{j_r})$ and a (nonempty) open set $O \subsetneq \mathbb{R}^r$ such that $m^{\#}(\theta) \in O$ for all $\theta \in \mathbb{R}^q$.

(ii) There exists a set $\Theta \subset \mathbb{R}^q$ with a null Lebesgue measure in $\mathbb{R}^q$ and such that $Q(\Theta) > 0$.

(iii) The number n of components of m is greater or equal to $q + 1$.

In case (i), the distribution of $m(\theta)$ does not put weight in some area of the space while the distribution of ε does. The case (iii) is an extreme situation of (i) : $m(\theta)$ and ε live in spaces with different dimensions. Case (ii) treats the case where there exists Lebesgue negligible sets (e.g. points) on which Q puts weight (while the gaussian distribution does not).

Theorem 5 *If m is injective and if one of the three conditions (i), (ii) or (iii), holds, (Q, σ^2) is identifiable from $P_{Y|Q,\sigma^2}$.*

Proof. Once the identifiability of σ^2 holds, identifiability of Q is deduced from the injectivity of m and Lemma **4**. The proof of the theorem thus reduces to show the identifiability of σ^2 in the cases (i)-(iii).

Let us consider $(\theta_0, \varepsilon_0) \sim Q_0 \times N(0, \sigma_0^2 \mathbf{I}_n)$ and $(\theta, \varepsilon) \sim Q \times N(0, \sigma^2 \mathbf{I}_n)$, and let us assume that the random vectors $m(\theta_0) + \varepsilon_0$ and $m(\theta) + \varepsilon$ have the same distribution. Then, the following identity holds for the FS transforms

$$Ee^{i\langle \xi, m(\theta)\rangle} Ee^{i\langle \xi, \varepsilon\rangle} = Ee^{i\langle \xi, m(\theta_0)\rangle} Ee^{i\langle \xi, \varepsilon_0\rangle}, \ \forall\, \xi \in \mathbb{R}^n.$$

Without loss of generality, we can assume that $\sigma^2 \leq \sigma_0^2$. Let us then consider $\gamma^2 = \sigma_0^2 - \sigma^2$. Since, for every $\xi \in \mathbb{R}^n$, $Ee^{i\langle \xi, \varepsilon \rangle} = e^{-\frac{1}{2}\langle \xi, \xi \rangle \sigma^2} \neq 0$, the latter identity may thus be rewritten as

$$Ee^{i\langle \xi, m(\theta) \rangle} = Ee^{i\langle \xi, m(\theta_0) \rangle} e^{-\frac{1}{2}\langle \xi, \xi \rangle \gamma^2}. \tag{3}$$

The last display states that if $\widetilde{\varepsilon} \sim N(0, \gamma^2 \mathbf{I}_n)$ then $Z = m(\theta)$ and $Z_0 = m(\theta_0) + \widetilde{\varepsilon}$ have the same distribution.

Assume that condition (i) holds and let us denote $Z^{\#} = m^{\#}(\theta)$ and $Z_0^{\#} = m^{\#}(\theta_0) + \widetilde{\varepsilon}^{\#}$. Then from (3), $P(Z^{\#} \in C) = P(Z_0^{\#} \in C)$ for all borel sets C in $\mathbb{R}^q$. But $P(Z^{\#} \in O) = 1$ and if $\gamma^2 > 0$, $P(Z_0^{\#} \in O) < 1$ which is impossible. It turns that $\gamma^2 = 0$, that is, $\sigma^2 = \sigma_0^2$.

Let us now consider $\Theta \subset \mathbb{R}^q$ satisfying condition (ii) and let us denote $O = m(\Theta)$. By the measurability of m, O is Lebesgue negligible. It follows that $P(Z_0 \in O) = 0$ if $\gamma^2 > 0$. Since $Q(\Theta) > 0$, $P(Z \in O) > 0$. But from (3) $P(Z \in O) = P(Z_0 \in O)$ which is only possible when $\gamma^2 = 0$.

When (iii) holds, $m(\mathbb{R}^q)$ is a surface of $\mathbb{R}^n$ with a dimension at most equal to q. Thus, there exists a (nonempty) open set O in $\mathbb{R}^n$, such that $m(\mathbb{R}^q) \cap O = \emptyset$. It follows that $P(Z \in O) = 0$ and $P(Z_0 \in O) > 0$ if $\gamma^2 > 0$. Since from (3) $P(Z \in O) = P(Z_0 \in O)$, necessarily $\gamma^2 = 0$. ■

A simple illustration of condition (i) is the case when for instance at least one of the components of m is positive (i.e.: $r = 1$ and $O = (0, \infty)$).

Condition (ii) obviously holds when Q is a discrete probability measure. Of course, the injectivity of m is the more difficult condition to check at least when q is large.

Condition (iii) and the injectivity of m are necessary and sufficient conditions for identifiability of the gaussian linear model: $Y_i = m(\theta_i) + \varepsilon_i$ where $m(\theta) = X\theta$ and X is a $n \times q$ matrix. From the theorem 5, identifiability of the parameter (Q, σ^2) holds if X is a full rank matrix and $n \geq q + 1$. Let us

assume now that Q is normal $N_q(\alpha, \Lambda)$. In this simple case, the common distribution of the observed vectors Y_i is normal $N_n(X\alpha,\ X\Lambda X' + \sigma^2\mathbf{I}_n)$. Thus identifiability may be also checked from the first two moments of this distribution. Identifiability holds when $X\alpha = X\alpha_0$ and $X\Lambda X' + \sigma^2\mathbf{I}_n = X\Lambda_0 X' + \sigma_0^2\mathbf{I}_n$ imply $(\alpha, \Lambda, \sigma^2) = (\alpha_0, \Lambda_0, \sigma_0^2)$. By the first equation, $\alpha = \alpha_0$ if and only if X is a full rank matrix. The second equation may be rewritten as $X(\Lambda - \Lambda_0)X' = (\sigma_0^2 - \sigma^2)\mathbf{I}_n$. Since the rank of $X(\Lambda - \Lambda_0)X'$ is at most q and the rank of $\mathbf{I}_n$ is n, it follows that $\sigma_0^2 = \sigma^2$ and $\Lambda = \Lambda_0$ if and only if $n \geq q + 1$. Thus, at least for the gaussian linear model, the condition (iii) is necessary and sufficient. One could think that (iii) is also necessary for nonlinear models. We will see in examples 2 and 3 that it is not the case.

The result given in the previous theorem can be extended to the non iid framework which is the good framework to analyze longitudinal data. Assume that one observes

$$Y_i = m^i(\theta_i) + \varepsilon_i, \qquad i = 1, \ldots, N,$$

where the θ_i are iid random vectors distributed according to the unknown distribution $Q \in \mathcal{P}(\mathbb{R}^q)$, the ε_i are independent gaussian vectors in $\mathbb{R}^{n_i}$ with mean 0 and variance $\sigma^2\mathbf{I}_{n_i}$ for some fixed unknown $\sigma^2 > 0$. The θ_i and ε_i are assumed mutually independent. The vectorial functions m^i are assumed measurable and known given the experimental design.

Let us now consider the following additional condition.

(iv) For some $i \in \{1, \ldots, N\}$ there exist $j, k \in \{1, \ldots, n_i\}$ such that $m_j^i(\theta) = m_k^i(\theta)$.

This last condition can be interpreted as the existence of repetitions on the ith subject. In such a case, two components (j, k) of its conditional mean are equal.

Corollary 6 *Assume that there exists $i \in \{1, \ldots, N\}$ such that m^i is injective. If (ii) holds or if there exists $j \in \{1, \ldots, N\}$ such that (i), (iii) or (iv) holds then (Q, σ^2) is identifiable from $P_{Y_1,\ldots,Y_N|Q,\sigma^2}$.*

Proof. The proof is the same as in the iid framework for conditions (i), (ii) and (iii). It remains to shows that (iv) allows identifiability of σ^2. When (iv) holds there exist k and l with $m_k^j(\theta) = m_l^j(\theta)$. Then, the difference between the kth and lth components of Y_j is $Y_{jk} - Y_{jl} = \varepsilon_{jk} - \varepsilon_{jl}$, so that $(Y_{jk} - Y_{jl})^2$ is an unbiased estimator for $2\sigma^2$. Therefore, σ^2 has a (strongly) consistent estimator which implies its identifiability. ■

In the following section, we study the identifiability of three common models in the analysis of longitudinal data.

3 Examples

Example 1: The Michaelis-Menten model may be written (Pinheiro and Bates, 2000) as

$$Y_{ij} = \frac{\theta_{i1} t_j}{(\theta_{i2} + t_j)} + \varepsilon_{ij}$$

where Y_{ij} is the concentration measured at time t_j on the ith subject. The measurement error terms ε_{ij} are independent and identically normally distributed with 0 mean and variance σ^2. The subject parameter vectors θ_i are iid with a common distribution Q assumed absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^2$. The adopted design should at least ensure that (Q, σ^2) are identifiable.

Let $t_1 \neq t_2$, two different and positive times. Then, one can see with a little algebra, that the function $(\theta_1, \theta_2) \to (\theta_1 t_1 / (\theta_2 + t_1), \theta_1 t_2 / (\theta_2 + t_2))$ is injective on the domain $\{(\theta_1, \theta_2) \in \mathbb{R}^2, \theta_1 \neq 0, \theta_2 \neq -t1, \theta_2 \neq -t2\}$, and therefore Lebesgue-almost everywhere. It follows from the condition (iii) of the the-

orem, that σ^2 and Q are identifiable as soon as the experimental design contains three times, two of these being different and positive.

Example 2: In order to study the kinetic behavior of a substance with an endogen secretion, an experimental design is needed. The plasma concentration of that substance is to be measured at various times. Suppose that (as it is for some substance) the kinetics can be modeled with the following one-compartment mixed model with a baseline:

$$Y_{ij} = a + e^{\theta_{i1}} + e^{\theta_{i2} - t_j \theta_{i3}} + \varepsilon_{ij},$$

where Y_{ij} and ε_{ij} have the same meaning as in example 1. The subject parameter vectors θ_i are iid with a common distribution Q assumed absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^3$. The known number a represents a lower bound for the true concentrations. However, it is possible to observe a concentration lower than a due to the error ε.

The injectivity of $m(\theta) = \left(a + e^{\theta_1} + e^{\theta_2 - t_j \theta_3}\right)_{1 \leq j \leq n}$ is obtained on the domain $M = \{\theta \in \mathbb{R}^3, \theta_3 \neq 0\}$ as soon as $n = 3$ and $t_1 < t_2 < t_3$.

A simple way to show such injectivity is to consider the following function

$$(\theta, \theta') \in M \times M \mapsto m(\theta) - m(\theta') \tag{4}$$

and to show that it cancels on M only when $\theta = \theta'$. But (4) can be seen as a linear application that can be rewritten as Av, where $v = \left(e^{\theta_1} - e^{\theta_1}, e^{\theta_2}, e^{\theta'_2}\right)$ and

$$A = \begin{pmatrix} 1 & e^{-t_1\theta_3} & e^{-t_1\theta'_3} \\ 1 & e^{-t_2\theta_3} & e^{-t_2\theta'_3} \\ 1 & e^{-t_3\theta_3} & e^{-t_3\theta'_3} \end{pmatrix}.$$

Now, let us solve the linear system $Av = 0$. When $\theta_3 \neq \theta'_3$, the matrix A is not singular (see for instance Polya and Szegö, p.46), it follows that the solution is $v = 0$, which is impossible. So, if θ, θ' belong to M, necessarily $\theta_3 = \theta'_3$. Solving $\overline{A}\overline{v} = 0$, where $\overline{v} = \left(e^{\theta_1} - e^{\theta_1}, e^{\theta_2} - e^{\theta'_2}\right)$ and $\overline{A}$ is the

submatrix of A obtained after removing its third column, gives the solution $(\theta_1, \theta_2) = (\theta'_1, \theta'_2)$.

From condition (i) of theorem, we deduce that (Q, σ^2) is identifiable as soon as the experimental design contains three different times. Even if it is not intuitive, the reason for which only 3 different times are sufficient is that when one observation is less than the lower bound of concentration a, the variance is directly observed as the squared difference between the observation and a. This example shows that the condition (iii), which would lead to a design with four times (three of these being different), is not necessary for general nonlinear mixed-effects models.

Example 3: The Batman model is often used to model an oral administration of a drug. When a single administration of the drug is carried out at time $t = 0$, the mean plasma concentration at time t is given by

$$f(t, \theta_1, \theta_2, \theta_3) = e^{\theta_1} \left(e^{-\exp(\theta_2)t} - e^{-(\exp(\theta_3)+\exp(\theta_2))t} \right).$$

When an additional oral administration is planned at time $t^* > 0$, a non null percentage of patients (say $100(1-p)\%$) can forget to take the pill. The measured plasma concentration on the ith patient becomes

$$Y_{ij} = f(t_j, \theta_{i1}, \theta_{i2}, \theta_{i3}) + \theta_{i4} f((t_j - t^*)_+, \theta_{i1}, \theta_{i2}, \theta_{i3}) + \varepsilon_{ij}$$

where θ_{i4} is a Bernoulli random variable with parameter p which takes the value 1 when the ith patient takes the drug at time t^* and 0 if not, $(x)_+ = x$ if $x \geq 0$ and $(x)_+ = 0$ elsewhere. This is the simplest model that can be used to describe the compliance to the treatment. The subject parameter vectors θ_i are iid with a common distribution Q whose support is $\mathbb{R}^3 \times \{0, 1\}$. Assume that the observation times are $0 < t_1 < t_2 < t_3 < t_4$. The proof of injectivity of $m(\theta)$ on the domain $M = \{\theta \in \mathbb{R}^3 \times \{0, 1\}\}$ is obtained if for

instance $0 < t_1 < t_2 < t_3 < t^* < t_4$. The function

$$(\theta, \theta^{'}) \in M \times M \mapsto m(\theta) - m(\theta') \tag{5}$$

can be rewritten as Av, where $v = \left(e^{\theta_1}, -e^{\theta'_1}, \theta_4 e^{\theta_1}, -\theta'_4 e^{\theta'_1}\right)$ and

$$A = \begin{pmatrix} g(\theta, t_1) & g(\theta', t_1) & 0 & 0 \\ g(\theta, t_2) & g(\theta', t_2) & 0 & 0 \\ g(\theta, t_3) & g(\theta', t_3) & 0 & 0 \\ g(\theta, t_4) & g(\theta', t_4) & g(\theta, t_4 - t^*) & g(\theta', t_4 - t^*) \end{pmatrix},$$

with

$$g(\theta, t) = e^{-\exp(\theta_2)t} - e^{-(\exp(\theta_2)+\exp(\theta_3))t}.$$

Let consider the first three equations of the system $Av = 0$. If $\theta_2 \neq \theta'_2$ or $\theta_3 \neq \theta'_3$, these equations are linearly independent because whatever $\alpha \neq 0$, the function $t \mapsto g(\theta, t) - \alpha g(\theta', t)$ has at most 2 positive roots. Since the first two components of v cannot be equal to zero, necessarily $(\theta_2, \theta_3) = (\theta'_2, \theta'_3)$ and it follows that $\theta_1 = \theta'_1$. The last equation of the system allows to conclude that $\theta_4 = \theta'_4$. It follows that m is injective on M and from the condition (ii) of the theorem, we deduce that σ^2 and Q are identifiable for the previous design.

Conclusion: The results given in this paper deals with identifiability of model (1) when the error terms are gaussian. These results can be extended to models for which the error distribution has a Fourier-Stieltjes transform that does not cancel and a support on $\mathbb{R}^n$. As an example, the same results hold for multidimensional Student distributions.

It is not clear that the given criteria are necessary, especially in the non iid framework. More work is needed to investigate such property.

References

Barndorff-Nielsen, O. (1965). Identifiability of mixtures of exponential families. *Journal of Mathematical Analysis and Applications* **12:** 115-121.

Bruni, C. and Koch, G. (1985). Identifiability of continuous mixtures of unknown Gaussian distributions. *The Annals of Probability.* **13**: 1341-1357.

Cordy, C.B. and Thomas, R. (1997). Deconvolution of a distribution function. *Journal of the American Statistical Association.* **92:** 1459-1465.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems, *The Annals of Statistics.* **19:** 1257-1272.

Fan, J. (1992). Deconvolution with supersmooth distributions, *The Canadian Journal of Statistic.* **20:** 155-169.

Lindstrom, M.J. and Bates, D.M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* **46:** 673-687.

Lindsay, B. G. (1995). Mixture Models: Theory, Geometry and Applications. NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 5, Hayward, CA: IMS.

Parthasarathy, K.R. (1967) Probability measures on metric spaces, Academic Press, New York and London.

Polya, G. and Szego, G. (1976) Problems and Theorems in Analysis, vol. II, Springer-Verlag, Berlin, Heidelberg.

Li, L.A. and Sedransk, N. (1988). Mixtures of distributions: A topological approach, *The Annals of Statistics.* **16:** 1623-1634.

Liu, M.C. and Taylor, R.L. (1989). A consistent nonparametric density estimator for the deconvolution problem, *The Canadian Journal of Statistic.* **17:** 427-438.

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *The Annals of Statistics.* **27:** 886-906.

Pinheiro, J. and Bates, D. (2000) Mixed-effects Models in S and S-PLUS.

New York: Springer-Verlag.

PFANZAGL, J. (1994). On the identifiability of structural parameters in mixtures: applications to psychological tests, *Journal of Statistical Planning and Inference.* **38:** 309-326.

STEFANSKY, L. AND CARROLL, R. (1990). Deconvoluting kernel density estimators, *Statistics,* **21:** 169-184.

TEICHER, H. (1960). On the mixture of distribution. *Annals of Mathematical Statistics.* **31:** 55-77.

TEICHER, H. (1961). Identifiability of mixtures. *Annals of Mathematical Statistics.* **32:** 244-248.

TEICHER, H. (1967). Identifiability of mixtures of product measures. *Annals of Mathematical Statistics.* **38:** 1300-1302.

ZHANG, C. (1990). Fourier methods for estimating mixing densities and distributions. *The Annals of Statistics.* **18:** 806-831.