**Title**
A note on the identifiability of nonparametric and semiparametric mixtures of GLMs



**Authors**
Wang, Shaoli
Yao, Weixin
Huang, Mian

**Publication Date**
2014

Peer reviewed

# A Note On the Identifiability of Nonparametric and Semiparametric Mixtures of GLMs

Shaoli Wang[a], Weixin Yao[b], Mian Huang[a,*]

[a]*School of Statistics and Management and Key Laboratory of Mathematical Economics at SHUFE, Ministry of Education, Shanghai University of Finance and Economics (SHUFE), Shanghai 200433, P. R. China*

[b]*Department of Statistics, Kansas State University, Manhattan, KS, 66506*

**Abstract**

Recently, many semiparametric and nonparametric finite mixture models have been proposed and investigated, which have widened the scope of finite mixture models. However, these works either lack identifiability results or only give identifiability results on a case-by-case basis. In this article, we first propose a semiparametric mixture of generalized linear models (GLMs) and a nonparametric mixture of GLMs to unify many of the recently proposed nonparametric and semiparametrc mixture models. We then further establish identifiability results for the proposed two models under mild conditions. The new results reveal the identifiability of some recently proposed nonparametric and semiparametrc mixture models, which are not previously established, and thus provide theoretical foundations for the estimation and inference of those mixture models. In addition, the methods can be easily generalized for many other semiparametric and nonparametric mixture models which are not considered in this article.

**Keywords:** Mixture Models; Identifiability; GLMs; Semiparametric and Nonparametric Models.

*Corresponding author. Email:huang.mian@shufe.edu.cn

## 1. Introduction

The finite mixture model is a powerful tool for modeling data sampled from a heterogeneous population which consists of relatively homogeneous subpopulations if the subpopulation identity of each observation is unknown. Please see, for example, Titterington et al. (1985), Lindsay (1995), McLachlan and Peel (2000), and Frühwirth-Schnatter (2006), for a general overview of mixture models.

Identifiability is of fundamental importance and has been one of the important research topics for finite mixture models. Without the identifiability result, the mixture model parameters might not be estimable and some of the related statistical inference might be meaningless. One well known feature for the mixture model is that the identifiability of each component density can not guarantee the identifiability of the corresponding mixture model. See, for example, Titterington et al. (1985), Hennig (2000), and Lindsay (1995). Therefore, one usually needs to investigate the identifiability of mixture models case by case. Classical identifiability results given by Teicher (1963) and Yakowitz and Spragins (1968) provide a foundation for finite mixtures of parametric distributions. For the identifiability results of mixtures of linear regression models for continuous responses, and mixtures of generalized linear models (GLMs) for binary or count responses, please see Hennig (2000), Grün and Leisch (2008a), and Grün and Leisch (2008c) and references therein.

Some generalization efforts have recently been made to relax the usual parametric assumptions on mixtures of linear regression models and mixtures of GLMs. Grün and Leisch (2008b) discussed the parameter estimation for mixtures of GLMs with varying proportions. Cao and Yao (2012) proposed a mixture of binomial regression models with a degenerate component, assuming mixing proportions and a component success probability function are nonparametric functions of time. Young and Hunter (2010) and Huang and Yao (2012) generalized mixtures of linear regression models by allowing varying mixing proportions that depend nonparametrically on a covariate. Huang et al. (2013) further extended mixtures of linear regression models by assuming the conditional mean and variable functions, and the mixing proportions are all nonparametric functions of a covariate. These nonparametric and semiparametric extensions of mixtures of GLMs enable more flexible modeling and widen the scope of applications of mixture models. However, these works either lack of identifiability results, such as Grün and Leisch (2008a), Young and Hunter (2010), and Cao and Yao (2012), or only give identifiability results on a case-by-case basis.

In this article, we propose two unified models for the nonparametric and semiparametric generalizations of finite mixtures of GLMs and prove their general identifiability. The first model, which we will call a semiparametric mixture of GLMs, assumes that the mixing proportions depend on some covariates nonparametrically and the canonical parameters are linear functions of covariates with constant dispersion parameters. The second model, which we will call nonparametric mixture of GLMs, assumes that the mixing proportions, the canonical parameters, and the dispersion parameters are all nonparametric functions of some covariates. The models proposed by Grün and Leisch (2008b), Young and Hunter (2010), Cao and Yao (2012), Huang and Yao (2012), and Huang et al. (2013) are all special cases of the proposed two models. We prove that the proposed two classes of mixture models are identifiable under some mild conditions, and thus build a foundation for estimation, inference, and applications of these models.

The rest of the paper is organized as follows. In Section 2, we give formal formulation of the two classes of mixture models. In Section 3, we prove the identifiability results for the two classes of models. Some discussions are given in Section 4. All proofs are given in the appendix.

## 2. Nonparametric and Semiparametric Mixtures of GLMs

**GLMs and Finite Mixtures of GLMs:** McCullagh and Nelder (1989) introduced generalized linear models (GLMs) as a flexible generalization of ordinary linear regression models, which allows for both discrete and continuous response variables if they are from the exponential family. Assume that $\{(\mathbf{X}_i, Y_i), i = 1, \cdots, n\}$ is a random sample from the population $(\mathbf{X}, Y)$, where $Y$ is a scalar response and $\mathbf{X}$ is a $p$-dimensional vector of covariates. The conditional density of $Y$ given $\mathbf{X} = \mathbf{x}$ is assumed to be from a one-parameter exponential family

$$f\{y; \theta(\mathbf{x}), \phi\} = \exp[\phi^{-1}\{y\theta(\mathbf{x}) + b(\theta(\mathbf{x}))\} + \kappa(y, \phi)],$$

where $b(\cdot)$ and $\kappa(\cdot)$ are known functions, $\theta(\mathbf{x})$ is the natural or canonical parameter, and $\phi$ is the dispersion parameter. Under the canonical link function $g = (b')^{-1}$, $\theta(\mathbf{x}) = g(\mu(\mathbf{x}))$, where $\mu(\mathbf{x}) = \mathrm{E}(Y|\mathbf{X} = \mathbf{x})$ is the conditional mean function of $Y$ given $\mathbf{x}$. In GLMs, the canonical parameter $\theta(\mathbf{x})$ is assumed to be a linear function of $\mathbf{x}$, i.e., $\theta(\mathbf{x}) = \mathbf{x}^T\boldsymbol{\beta}$.

Finite mixtures of GLMs can be viewed as a generalization of GLMs and finite mixtures of ordinary linear regression models. Grün and Leisch (2008a) discussed the identifiability and parameter estimation for finite mixtures of GLMs with fixed covariates. In the mixture of GLMs, there is a latent class variable $\mathcal{C}$ which has a discrete distribution $P(\mathcal{C} = c \mid \mathbf{X} = \mathbf{x}) = \pi_c$, $c = 1, 2, \ldots, C$, where $C$ is the number of components and assumed to be known in this article. Conditioning on $\mathcal{C} = c$ and $\mathbf{X} = \mathbf{x}$, $Y$ has a distribution $f\{y; \theta_c(\mathbf{x}), \phi_c\}$ from the exponential family. Since the latent variable $\mathcal{C}$ is unobservable in general, conditioning on $\mathbf{X} = \mathbf{x}$, the response variable $Y$ follows a mixture distribution:

$$Y|\mathbf{X} = \mathbf{x} \sim \sum_{c=1}^{C} \pi_c f\{y; \theta_c(\mathbf{x}), \phi_c\}, \tag{2.1}$$

where the canonical parameters $\theta_c(\mathbf{x}) = \mathbf{x}^T\boldsymbol{\beta}_c$ are assumed to be linear functions of the covariates, and the proportions $\pi_c$ and the dispersion parameters $\phi_c$ are constant. Note that the Gaussian mixture of linear regressions with normal errors is a special case of (2.1). Therefore, all the results given in this article are applicable to Gaussian mixtures of linear regressions.

Next, we propose two models to relax the linear assumption on the canonical parameters, and the constant assumption on the proportions and the dispersion parameters.

**Semiparametric Mixtures of GLMs:** The first class of mixtures of GLMs is called semiparametric mixtures of GLMs, which assumes that the mixing proportions in model (2.1) are smoothing functions of a covariate $\mathbf{z}$, while the canonical functions remain linear functions of covariates and the dispersion parameters remain constant.

More specifically, conditioning on $\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}$, $\mathcal{C}$ has a discrete distribution $P(\mathcal{C} = c|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) = \pi_c(\mathbf{z})$, $c = 1, 2, \ldots, C$. The conditional distribution of $Y$ given $\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}$ is

$$Y|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z} \sim \sum_{c=1}^{C} \pi_c(\mathbf{z}) f\{y; \mathbf{x}^T \boldsymbol{\beta}_c, \phi_c\}, \tag{2.2}$$

where $\sum_{c=1}^{C} \pi_c(\mathbf{z}) = 1$ and $\mathbf{z}$ can be part of or the same as $\mathbf{x}$. A constant 1 might be included in the vector of covariates to allow for an intercept in each GLM component. Compared to the model (2.1), the new model (2.2) allows to incorporate the covariates information into the mixing proportions and thus can better cluster the data. Note that the models proposed by Grün and Leisch (2008b), Young and Hunter (2010), and Huang and Yao (2012) are special cases of (2.2). The model estimation and inference of (2.2) can be done similarly to Huang and Yao (2012).

In (2.2), similar to $\pi_c$, we can also allow the component dispersion parameter $\phi_c$ to depend on $\mathbf{z}$ nonparametrically. The identifiability results can be established similarly to the model (2.2).

**Nonparametric Mixtures of GLMs:** The second class of mixtures of GLMs is called nonparametric mixtures of GLMs, in which the mixing proportions, the component canonical parameters, and the component dispersion parameters are all nonparametric functions of covariates.

More precisely, we assume that conditioning on $\mathbf{X} = \mathbf{x}$, $\mathcal{C}$ has a discrete distribution $P(\mathcal{C} = c|\mathbf{X} = \mathbf{x}) = \pi_c(\mathbf{x})$, $c = 1, 2, \ldots, C$. Conditioning on $\mathcal{C} = c$ and

$\mathbf{X} = \mathbf{x}$, $Y$ has a distribution $f\{y; \theta_c(\mathbf{x}), \phi_c(\mathbf{x})\}$ with the canonical parameter $\theta_c(\mathbf{x})$ and the dispersion parameter $\phi_c(\mathbf{x})$. We further assume that $\pi_c(\mathbf{x})$, $\theta_c(\mathbf{x})$, and $\phi_c(\mathbf{x})$ are unknown but smooth functions. Therefore, without observing $\mathcal{C}$, conditioning on $\mathbf{X} = \mathbf{x}$, the response variable $Y$ follows a finite mixture of GLMs

$$Y|\mathbf{X} = \mathbf{x} \sim \sum_{c=1}^{C} \pi_c(\mathbf{x}) f\{y; \theta_c(\mathbf{x}), \phi_c(\mathbf{x})\}. \tag{2.3}$$

The model estimation and inference of (2.3) can be performed similarly to Huang et al. (2013). Compared to the models (2.1) and (2.2), the model (2.3) can relax the linear assumption about the canonical parameters $\theta_c$. Note that the models proposed by Cao and Yao (2012) and Huang et al. (2013) are special cases of (2.3).

## 3. Identifiability

Identifiability results for finite mixtures of ordinary linear models and finite mixtures of GLMs are available, see Hennig (2000) and Grün and Leisch (2008a); but these results are not applicable to models (2.2) and (2.3) as some of parameters are nonparametric functions of covariates.

Seminal papers of Teicher (1961, 1963) and Yakowitz and Spragins (1968) established the fundamental identifiability results for finite mixtures models. Lindsay (1995) gave an elegant account of the identifiability of finite mixtures of discrete distributions from one parameter exponential family. For more references, see also Chandra (1977), Titterington et al. (1985), McLachlan and Peel (2000), and Frühwirth-Schnatter (2006). Based on existing results, we know that many finite mixtures of continuous distribution families are identifiable. These include finite mixtures of univariate normal distributions, finite mixtures of multivariate normal distributions, finite mixtures of exponential distributions, finite mixtures of Gamma distributions, etc. As to finite mixtures of discrete one parameter families, we know that finite mixtures of Poisson distributions and finite mixtures of negative binomial distributions are identifiable. In addition, finite mixtures of binomial distributions are identifiable if the number of components is not greater than half of the number of trials of the binomial distributions. In this section, we assume that the one-parameter exponential family distribution $f(y;\theta,\phi)$ generates identifiable finite mixtures models, i.e., the mixture model $\sum_{c=1}^{C}\pi_c f\{y;\theta_c,\phi_c\}$ is identifiable.

Next, we establish the identifiability results of the proposed two models. We first formally define the identifiability for the model (2.2) and model (2.3).

**Definition 3.1.** *Let $\sum_{c=1}^{C}\pi_c(\boldsymbol{z})f\{y;\boldsymbol{x}^T\boldsymbol{\beta}_c,\phi_c\}$ and $\sum_{d=1}^{D}\lambda_d(\boldsymbol{z})f\{y;\boldsymbol{x}^T\gamma_d,\psi_d\}$ be any two semiparametric mixture of GLMs of the form (2.2), where $\pi_c(\boldsymbol{z})>0$, $c=1,\ldots,C$, $\sum_{c=1}^{C}\pi_c(\boldsymbol{z})$=1, $\lambda_d(\boldsymbol{z})>0$, $d=1,\ldots,D$, $\sum_{d=1}^{D}\lambda_d(\boldsymbol{z})=1$. The model (2.2) is said to be identifiable, if*

$$\sum_{c=1}^{C}\pi_c(\boldsymbol{z})f\{y;\boldsymbol{x}^T\boldsymbol{\beta}_c,\phi_c\}=\sum_{d=1}^{D}\lambda_d(\boldsymbol{z})f\{y;\boldsymbol{x}^T\gamma_d,\psi_d\} \tag{3.1}$$

*for all $\boldsymbol{x}\in\mathcal{X}$ and $\boldsymbol{z}\in\mathcal{Z}$ implies that $C=D$ and that the summations in (3.1) can be reordered such that $\pi_c(\boldsymbol{z})=\lambda_c(\boldsymbol{z})$, $\boldsymbol{\beta}_c=\gamma_c$, and $\phi_c=\psi_c$, $c=1,\ldots,C$.*

**Definition 3.2.** *Let $\sum_{c=1}^{C}\pi_c(\boldsymbol{x})f\{y;\theta_c(\boldsymbol{x}),\phi_c(\boldsymbol{x})\}$ and $\sum_{d=1}^{D}\lambda_d(\boldsymbol{x})f\{y;\gamma_d(\boldsymbol{x}),\psi_d(\boldsymbol{x})\}$ be any two nonparametric mixtures of GLMs of the form (2.3), where $\pi_c(\boldsymbol{x})>0$, $c=1,\ldots,C$, $\sum_{c=1}^{C}\pi_c(\boldsymbol{x})$=1, $\lambda_d(\boldsymbol{x})>0$, $d=1,\ldots,D$, $\sum_{d=1}^{D}\lambda_d(\boldsymbol{x})=1$. The model (2.3) is said to be identifiable, if*

$$\sum_{c=1}^{C}\pi_c(\boldsymbol{x})f\{y;\theta_c(\boldsymbol{x}),\phi_c(\boldsymbol{x})\}=\sum_{d=1}^{D}\lambda_d(\boldsymbol{x})f\{y;\gamma_d(\boldsymbol{x}),\psi_d(\boldsymbol{x})\} \tag{3.2}$$

*for all $\boldsymbol{x}\in\mathcal{X}$ implies that $C=D$ and that the summations in (3.2) can be reordered such that $\pi_c(\boldsymbol{x})=\lambda_c(\boldsymbol{x})$, $\theta_c(\boldsymbol{x})=\gamma_c(\boldsymbol{x})$, and $\phi_c(\boldsymbol{x})=\psi_c(\boldsymbol{x})$, $c=1,\ldots,C$.*

The following theorem provides identifiability conditions for the model (2.2) and the proof is given in the appendix.

**Theorem 3.1.** *The model (2.2) is identifiable if the following conditions are satisfied:*

*(i) the domain $\mathcal{X}$ of $\boldsymbol{x}$ contains an open set in $R^p$, and the domain $\mathcal{Z}$ of $\boldsymbol{z}$ has no isolated points.*

*(ii) $\pi_c(\boldsymbol{z})>0$ are continuous functions, $c=1,\ldots,C$, and $(\boldsymbol{\beta}_c,\phi_c)$, $c=1,\ldots,C$, are distinct pairs.*

*(iii) the parametric mixture model $\sum_{c=1}^{C}\pi_c f\{y;\theta_c,\phi_c\}$ is identifiable.*

Based on the above theorem, we can see that the proposed semiparametric mixtures of GLMs are identifiable under some mild conditions. Note that the identifiability result proved in Huang and Yao (2012) is a special case of Theorem 3.1. In addition, our Theorem 3.1 proves that the models considered by Grün and Leisch (2008b) and Young and Hunter (2010) are also identifiable, which are not previously established for their models.

The following theorem provides identifiability conditions for the model (2.3) and the proof is given in the appendix.

**Theorem 3.2.** *The model (2.3) is identifiable if the following conditions are satisfied:*

*(i) The domain $\mathcal{X}$ of $\boldsymbol{x}$ is an open set in $\mathbb{R}^p$.*

*(ii) $\pi_c(\boldsymbol{x}) > 0$ are continuous functions, and $\theta_c(\boldsymbol{x})$ and $\phi_c(\boldsymbol{x})$ have continuous first derivative, $c = 1, \dots, C$.*

*(iii) For any $\boldsymbol{x}$ and $1 \le j \ne k \le C$,*

$$\sum_{l=0}^{1} \|\theta_j^{(l)}(\boldsymbol{x}) - \theta_k^{(l)}(\boldsymbol{x})\|^2 + \sum_{l=0}^{1} \|\phi_j^{(l)}(\boldsymbol{x}) - \phi_k^{(l)}(\boldsymbol{x})\|^2 \ne 0,$$

*where $g^{(l)}$ is the $l$th derivative of $g$ and equal to $g$ if $l = 0$.*

*(iv) the parametric mixture model $\sum_{c=1}^{C} \pi_c f\{y; \theta_c, \phi_c\}$ is identifiable.*

The third condition of Theorem 3.2 requires that the canonical parameter functions and the dispersion parameter functions of any two components can not be tangent to each other at the same $\mathbf{x}$. Based on the above theorem, we can see that the proposed nonparametric mixtures of GLMs are identifiable under some mild conditions. Note that the identifiability result proved in Huang et al. (2013) is a special case of Theorem 3.2. In addition, Theorem 3.2 proves that the model proposed in Cao and Yao (2012) is identifiable, which is not previously proved.

## 4. Discussion

The model identifiability is of great importance for the study of finite mixture models, since it is the necessary theoretical foundation for the estimation and inference. Recently, many semiparametric and nonparametric extensions of existing parametric finite mixture models are proposed and some of them are lack of identifiability results. In this article, we proposed a semiparametric mixture of GLMs and a nonparametric mixture of GLMS, which include many recently proposed nonparametric and semiparametrc mixture models as special cases. We further established identifiability results for the proposed two models under mild conditions. In particular, our result verifies that the semiparametric and nonparametric mixture models proposed by Grün and Leisch (2008a), Young and Hunter (2010), and Cao and Yao (2012) are identifiable under the conditions given in Theorems 3.1 and 3.2.

**Acknowledgements**

The authors are grateful to Dr. Paindaveine and two referees for their insightful comments and suggestions, which greatly improved this article. Huang's research is supported by National Natural Science Foundation of China (NNSFC), grant 11301324.

**Appendix**

**Proof of Theorem 3.1.** The proof is similar to Huang and Yao (2012).

Let $\mathbf{x} = (1, \mathbf{x}_s^T)^T$. Suppose that model (2.2) admits another representation

$$Y|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z} \sim \sum_{d=1}^{D} \lambda_d(\mathbf{z}) f(\mathbf{x}^T \boldsymbol{\gamma}_d, \delta_d),$$

where $\lambda_d(\mathbf{z}) > 0$, $d = 1, \ldots, D$, $\sum_{d=1}^{D} \lambda_d(\mathbf{z}) = 1$, and $(\boldsymbol{\gamma}_d, \delta_d)$, $d = 1, \ldots, D$, are distinct.

For any two distinct pairs of parameters $(\boldsymbol{\beta}_a, \phi_a)$ and $(\boldsymbol{\beta}_b, \phi_b)$, if $\phi_a = \phi_b$, then $\boldsymbol{\beta}_a \neq \boldsymbol{\beta}_b$. When $\boldsymbol{\beta}_a \neq \boldsymbol{\beta}_b$, the set $\{\mathbf{x}_s \in \mathbb{R}^p : \mathbf{x}^T \boldsymbol{\beta}_a = \mathbf{x}^T \boldsymbol{\beta}_b\}$ is a $(p-1)$-dimensional hyperplane in $\mathbb{R}^p$, and thus has zero Lebesgue measure in $\mathbb{R}^p$. This implies that there are at most a finite number of $(p-1)$-dimensional hyperplanes on which $(\mathbf{x}^T \boldsymbol{\beta}_a, \phi_a^2) = (\mathbf{x}^T \boldsymbol{\beta}_b, \phi_b^2)$ for some $a$, $b$. Hence the union of these finite number of hyperplanes has zero Lebesgue measure in $\mathbb{R}^p$. The same thing is true for the set of parameters $(\boldsymbol{\gamma}_d, \delta_d^2)$, $d = 1, \ldots, D$.

Noting that the domain $\mathcal{X}$ of $\mathbf{x}_s$ contains an open set in $R^p$, based condition (iii), for any given $(\mathbf{x}, \mathbf{z})$ such that both sets of parameters $(\mathbf{x}^T\boldsymbol{\beta}_c, \phi_c^2)$, $c = 1, \ldots, C$, and $(\mathbf{x}^T\boldsymbol{\gamma}_d, \delta_d^2)$, $d = 1, \ldots, D$, are distinct pairs, respectively, model (2.3) conditioning on $\mathbf{u} = (\mathbf{x}, \mathbf{z})$ is identifiable. Therefore, $C = D$ and there exists a permutation $\boldsymbol{\omega}_\mathbf{u} = \{\boldsymbol{\omega}_\mathbf{u}(1), \ldots, \boldsymbol{\omega}_\mathbf{u}(C)\}$ of set $\{1, \ldots, C\}$ depending on $\mathbf{u}$, such that

$$\lambda_{\boldsymbol{\omega}_\mathbf{u}(c)}(\mathbf{z}) = \pi_c(\mathbf{z}), \mathbf{x}^T\boldsymbol{\gamma}_{\boldsymbol{\omega}_\mathbf{u}(c)} = \mathbf{x}^T\boldsymbol{\beta}_c, \delta^2_{\boldsymbol{\omega}_\mathbf{u}(c)} = \phi_c^2, c = 1, \ldots, C. \tag{4.1}$$

Since there are only a finite number $(C!)$ of possible permutations of $\{1, 2, \ldots, C\}$ and the domain $\mathcal{X}$ of $\mathbf{x}_s$ contains an open set in $\mathbb{R}^p$, there must exist a permutation $\boldsymbol{\omega}^* = \{\boldsymbol{\omega}^*(1), \ldots, \boldsymbol{\omega}^*(C)\}$, such that (4.1) holds on a subset of $\mathcal{X}$ with nonzero Lebesgue measure. Hence, $\boldsymbol{\gamma}_{\boldsymbol{\omega}^*(c)} = \boldsymbol{\beta}_c, \delta^2_{\boldsymbol{\omega}^*(c)} = \phi_c^2, c = 1, \ldots, C$. Because that $(\boldsymbol{\beta}_c, \phi_c^2)$, $c = 1, \ldots, C$ are distinct and $(\boldsymbol{\gamma}_c, \delta_c^2)$, $c = 1, \ldots, C$ are distinct, it follows that $\boldsymbol{\omega}^*$ is the unique permutation such that (4.1) holds on a subset of $\mathcal{X}$ with nonzero Lebesgue measure. If $\mathbf{z}$ is not from $\mathbf{x}$, then $\lambda_{\boldsymbol{\omega}^*(c)}(\mathbf{z}) = \pi_c(\mathbf{z})$, $c = 1, \ldots, C$ for any $\mathbf{z} \in \mathcal{Z}$. If $\mathbf{z}$ is from $\mathbf{x}$, $\lambda_{\boldsymbol{\omega}^*(c)}(\mathbf{z}) = \pi_c(\mathbf{z})$, $c = 1, \ldots, C$, for all $\mathbf{z} \in \mathcal{Z}$ but a zero Lebesgue measure set. Because $\pi_c(\mathbf{z})$ are continuous and the domain of $\mathbf{z}$ has no isolated point, the values of $\pi_c(\mathbf{z})$ at those zero Lebesgue measure set are also uniquely determined. This completes the proof.

**Proof of Theorem 3.2.** The proof is similar to Huang, et al. (2013). For simplicity, here we only give the proof for the univariate covariate case. The proof for the multivariate covariate case is similar. Let

$$T = \{x \in \mathbb{R} : \ (\theta_j(x), \phi_j(x)) = (\theta_k(x), \phi_k(x)) \text{ for some } 1 \leq j \neq k \leq C\}$$

be the set of points where some component regression curves intersect. Let $\boldsymbol{\eta}_j(x) = (\theta_j(x), \phi_j(x))^T, j = 1, \ldots, C$. Based on Theorem 3.2 (iii), if $\boldsymbol{\eta}_j(x) = \boldsymbol{\eta}_k(x)$ for some $x$ and $1 \leq j \neq k \leq C$, then $\boldsymbol{\eta}'_j(x) \neq \boldsymbol{\eta}'_k(x)$. Therefore, any point in $T$ is an isolated point. This implies that set $T \subset \mathbb{R}$ has no limit point and contains countably many points. Without loss of generality, we assume that $x_l < x_{l+1}$ and $(x_l, x_{l+1}) \cap T = \emptyset$, $l = 0, \pm 1, \pm 2, \ldots$.

Assume that model (2.3) admits another representation

$$Y|X = x \sim \sum_{d=1}^{D} \lambda_d(x) f(y; \gamma_d(x), \delta_d(x)),$$

where $\lambda_d(x) > 0$, $d = 1, \ldots, D$, and $\sum_{d=1}^{D} \lambda_d(x) = 1$.

Based on condition (iv), we know that for any given $x \notin T$, model (2.3) is identifiable. It follows that $C = D$, and there exists a permutation $\boldsymbol{\omega}_x = \{\boldsymbol{\omega}_x(1), \ldots, \boldsymbol{\omega}_x(C)\}$ of set $\{1, \ldots, C\}$ depending on $x$, such that

$$\lambda_{\boldsymbol{\omega}_x(c)}(x) = \pi_c(x),\ \gamma_{\boldsymbol{\omega}_x(c)}(x) = \theta_c(x),\ \delta_{\boldsymbol{\omega}_x(c)}(x) = \phi_c(x),\ \ c = 1, \ldots, C. \tag{4.2}$$

Since all component regression curves $(\theta_j(x), \phi_j(x))$ are continuous, and no pair of component regression curves intersect on any interval $(x_l, x_{l+1})$, the permutation $\boldsymbol{\omega}_x$ must be constant on $(x_l, x_{l+1})$ and will be simply denoted by $\boldsymbol{\omega}_l$.

Next, we will prove that $\boldsymbol{\omega}_l = \boldsymbol{\omega}_{l+1}$ and $\boldsymbol{\omega}_l = \boldsymbol{\omega}_{x_{l+1}}$ for any $l$. If the above statement is true, then $\boldsymbol{\omega}_x$ will be the same over the interval $(x_l, x_{l+2})$ for any $l$. By taking different $l$ values, we can know that $\boldsymbol{\omega}_x$ will be constant over the whole space of $x$.

Note that any pair of component regression curves have different derivatives at $x_{l+1}$ if they intersect at $x_{l+1}$. Since equation (4.2) implies that the identity of derivatives of parameter curves on either side of $x_{l+1}$, $\boldsymbol{\omega}_x$ must be constant on the neighborhood of $x_{l+1}$. Therefore, $\boldsymbol{\omega}_l = \boldsymbol{\omega}_{l+1}$ and there exists a permutation $\boldsymbol{\omega} = \{\boldsymbol{\omega}(1), \ldots, \boldsymbol{\omega}(C)\}$ of set $\{1, \ldots, C\}$ which is independent of $x$ such that

$$\lambda_{\boldsymbol{\omega}(c)}(x) = \pi_c(x),\ \gamma_{\boldsymbol{\omega}(c)}(x) = \theta_c(x),\ \delta_{\boldsymbol{\omega}(c)}(x) = \phi_c(x),\ \ c = 1, \ldots, C.$$

This completes the proof of identifiability.

**Reference**


Cao, J. and Yao, W. (2012). Semiparametric mixture of binomial regression with a degenerate component. *Statistica Sinica*, 22:27–46.

Chandra, S. (1977). On the mixtures of probability distributions. *Scandinavian Journal of Statistics*, pages 105–112.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models: Modeling and Applications to Random Processes*. Springer.

Grün, B. and Leisch, F. (2008a). Finite mixtures of generalized linear regression models. In *Recent advances in linear models and related areas*, pages 205–230. Springer.

Grün, B. and Leisch, F. (2008b). Flexmix version 2: finite mixtures with concomitant variables and varying and constant parameters. *Journal of Statistical Software*, 28(4):1–35.

Grün, B. and Leisch, F. (2008c). Identifiability of finite mixtures of multinomial logit models with varying and fixed effects. *Journal of classification*, 25(2):225–247.

Hennig, C. (2000). Identifiablity of models for clusterwise linear regression. *Journal of Classification*, 17(2):273–296.

Huang, M., Li, R., and Wang, S. (2013). Nonparametric mixture of regression models. *Journal of the American Statistical Association*, 108(503):929–941.

Huang, M. and Yao, W. (2012). Mixture of regression models with varying mixing proportions: A semiparametric approach. *Journal of the American Statistical Association*, 107(498):711–724.

Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry and Applications.* Hayward, Institute of Mathematical Statistics.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models, 2nd Edition.* Chapman and Hall.

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*, volume 299. Wiley-Interscience.

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

Teicher, H. (1963). Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 34:1265C1269.

Titterington, D. M., Smith, A. F. M., and Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*, volume 38. Wiley New York.

Yakowitz, S. J. and Spragins, J. D. (1968). On the identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 39(1):209–214.

Young, D. S. and Hunter, D. R. (2010). Mixtures of regressions with predictor-dependent mixing proportions. *Computational Statistics & Data Analysis*, 54(10):2253–2266.