Mixture of regression models with latent variables and sparse coefficient parameters

Shu Kay Ng, *School of Medicine, Griffith Health Institute, Griffith University, Meadowbrook Q4131, Australia*, s.ng@griffith.edu.au

Geoffrey J McLachlan, *Department of Mathematics, University of Queensland, Brisbane Q4072, Australia*, g.mclachlan@uq.edu.au

Abstract. Mixture models have been widely used in marketing research and epidemiology to capture heterogeneity in endogenous latent variables among individuals. However, when collinearity between endogenous latent variables at the component level is present, some component-specific path coefficients will be zero. In this paper, a systematic computational algorithm is developed to identify parameters that need to be constrained to be zero and to address other issues including the initialization procedure, the provision of standard errors of estimates, and the method to determine the number of components. The proposed algorithm is illustrated using simulated data and a real data set concerning emotional behaviour of preschool children.

Keywords. Mixture models, Latent variables, Regression models, Sparse coefficients, EM algorithm

1 Introduction

Regression models involving latent variables (or constructs) are very common in the marketing research and epidemiology [2, 3]. With this approach, simultaneous regression equations are adopted to model the relationships between multiple dependent (endogenous) latent variables and independent (exogenous) latent variables. Let $\boldsymbol{\eta}_j$ and $\boldsymbol{\xi}_j$ denote the vectors of endogenous and exogenous latent variables for the jth individual ($j = 1, \dots, n$), respectively. The "inner" model is specified in terms of q simultaneous regression equations as

$$\boldsymbol{B}\boldsymbol{\eta}_j + \boldsymbol{\Gamma}\boldsymbol{\xi}_j = \boldsymbol{\zeta}_j, \tag{1}$$

where $\boldsymbol{B}$ is a $q \times q$ matrix with q being the number of endogenous latent variables, $\boldsymbol{\Gamma}$ is a $q \times p$ matrix where p is the number of exogenous latent variables, and $\boldsymbol{\zeta}_j$ is a random vector of residuals. The matrices $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ represent the (path) coefficients relating to the endogenous and exogenous latent variables, respectively, in the inner model. The relationships between the latent variables and the manifest variables, either reflective indicators or formative measures, are specified in the "outer" model [3]. Estimation of model parameters and values for latent variables can be proceed with two different approaches. The structural equation modelling (SEM) approach attempts to reproduce the covariance matrix of the observed measures, while the partial least squares (PLS) approach focuses on maximizing the variance of the endogenous variables explained by the exogenous variables.

In many real problems, the presence of heterogeneity among individuals in terms of different path coefficients is prevalence. Such kind of heterogeneity is due to different individual perception of latent variables and can be captured in the regression modelling via a finite mixture model approach [3, 10]. With the PLS approach to regression models with latent variables, it is assumed that the endogenous latent variables $\boldsymbol{\eta}_j$ $(j = 1, \ldots, n)$ come from a mixture of a finite number, say g of multivariate normal distributions in some unknown proportions $\pi_1, \ldots, \pi_g$ that sum to one:

$$f(\boldsymbol{\eta}_j; \boldsymbol{\Psi}) = \sum_{i=1}^{g} \pi_i \phi(\boldsymbol{\eta}_j; \boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_i) \qquad (j = 1, \ldots, n), \tag{2}$$

where $\boldsymbol{\mu}_{ij} = (\boldsymbol{I} - \boldsymbol{B}_i)\boldsymbol{\eta}_j - \boldsymbol{\Gamma}_i \boldsymbol{\xi}_j$ is the mean vector of the ith component, where $\boldsymbol{I}$ is an identity matrix, and $\boldsymbol{\Sigma}_i = \text{diag}(\boldsymbol{\sigma}_i^2)$ is a diagonal matrix constructed from the vector $\boldsymbol{\sigma}_i^2$, which represents the variance of the random residuals $\boldsymbol{\zeta}_{ij}$ $(i = 1, \ldots, g)$. In (2), $\boldsymbol{\Psi}$ is the vector of all the unknown parameters containing $\pi_1, \ldots, \pi_{g-1}$ and the free parameters in $\boldsymbol{B}_i$, $\boldsymbol{\Gamma}_i$, and $\boldsymbol{\Sigma}_i$ for $i = 1, \ldots, g$. From (1), the conditional multivariate normal density is given by

$$\phi(\boldsymbol{\eta}_j; \boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_i) = \frac{|\boldsymbol{B}_i|}{\sqrt{(2\pi)^q |\boldsymbol{\Sigma}_i|}} \exp\{-\tfrac{1}{2}(\boldsymbol{B}_i\boldsymbol{\eta}_j + \boldsymbol{\Gamma}_i\boldsymbol{\xi}_j)^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{B}_i\boldsymbol{\eta}_j + \boldsymbol{\Gamma}_i\boldsymbol{\xi}_j)\}, \tag{3}$$

where the superscript T denotes vector transpose.

While mixtures of multivariate normal distributions are generically identifiable (that is, the model is unique up to a permutation of the component labels; see [4, 7]), mixtures of regression models with latent variables arisen from (1) and (2) are not identifiable unless some elements of matrices $\boldsymbol{B}_i$ and $\boldsymbol{\Gamma}_i$ $(i = 1, \ldots, g)$ are constrained to zero [3]. In practice, the links between the latent variables represented by simultaneous regression equations in the inner model are usually hypothetical models pre-specified based on a researcher's own experience. When collinearity between endogenous latent variables at the component level is present, some component-specific path coefficients will be zero. However, the setting up of such parameter constraints at present is somewhat arbitrary. There are also issues of initialization procedure, provision of standard errors of parameter estimates for statistical inference, and determination of the number of components g in the mixture model [7]. In this paper, we tackle these issues by developing a systematic computational algorithm for the implementation of mixtures of regression models with latent variables and sparse coefficient parameters as presented in (1) and (2).

The rest of the paper is organized as follows: Section 2 describes the expectation-maximization (EM) algorithm for the iterative computation of maximum likelihood (ML) estimates of the mixture model and the procedure to identify sparse coefficient parameters. Also, we show how to initialize the algorithm, to obtain standard errors using a bootstrap resampling approach, and to determine the value of g. In Section 3, we present simulation studies to illustrate the applicability of the proposed algorithm in terms of the accuracy of the final model derived and the corresponding estimate biases. We show in Section 4 the application of the proposed method to a real data set. Section 5 ends the paper with further discussion.

2 Algorithm for fitting mixture of sparse regression models

The proposed algorithm applies directly to the scores of endogenous and exogenous latent variables, $\boldsymbol{\eta}_j$ and $\boldsymbol{\xi}_j$, calculated using an iterative scheme of standard PLS on the observed manifest variables with specification based on the constraints of $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ for all individuals $(j = 1, \ldots, n)$; see [3, 10]. The "aggregate" predictors of $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ estimated in the PLS procedure may also be used to guide the initial estimates for $\boldsymbol{B}_i$ and $\boldsymbol{\Gamma}_i$ $(i = 1, \ldots, g)$ in the mixture model.

Maximum likelihood estimation and parameter constraint

The fitting of the mixture model (2) to latent variables $\boldsymbol{\eta}_j$ and $\boldsymbol{\xi}_j$ $(j = 1, \ldots, n)$ obtained by PLS can be implemented using ML. An estimate $\hat{\boldsymbol{\Psi}}$ is obtained by solving the log likelihood equation iteratively via the EM algorithm [8]. An appealing property of the EM algorithm is that the likelihood is not decreased after each iteration. Within the EM framework, each individual is conceptualized to have arisen from one of the g components of the mixture model and the unobservable component-indicator vector $\boldsymbol{z}_j$ is treated as missing data. Precisely, the ith element z_{ij} of $\boldsymbol{z}_j$ is taken to be one or zero according as the jth individual does or does not come from the ith component $(i = 1, \ldots, g;\ j = 1, \ldots, n)$. On the $(k+1)$th iteration of the EM algorithm, the E-step computes the so-called Q-function, which is the conditional expectation of the complete-data log likelihood using the current fit for $\boldsymbol{\Psi}$:

$$Q(\boldsymbol{\Psi};\ \boldsymbol{\Psi}^{(k)}) = \sum_{i=1}^{g}\sum_{j=1}^{n} \tau_{ij}^{(k)}\{\log \pi_i + \log \phi(\boldsymbol{\eta}_j; \boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_i)\}, \tag{4}$$

where we simply have to calculate

$$\tau_{ij}^{(k)} = \frac{\pi_i^{(k)}\phi(\boldsymbol{\eta}_j; \boldsymbol{\mu}_{ij}^{(k)}, \boldsymbol{\Sigma}_i^{(k)})}{\sum_{h=1}^{g}\pi_h^{(k)}\phi(\boldsymbol{\eta}_j; \boldsymbol{\mu}_{hj}^{(k)}, \boldsymbol{\Sigma}_h^{(k)})} \qquad (i = 1, \ldots, g;\ j = 1, \ldots, n), \tag{5}$$

which is the posterior probability that the jth individual belongs to the ith component of the mixture; see [7].

The M-step updates the estimate of $\boldsymbol{\Psi}$ by the new value $\boldsymbol{\Psi}^{(k+1)}$ of $\boldsymbol{\Psi}$ that maximizes the Q-function with respect to $\boldsymbol{\Psi}$. It can be seen from (4) that the maximization with respect to the mixing proportions and coefficient parameters can be obtained separately as follows:

$$\pi_i^{(k+1)} = \sum_{j=1}^{n}\tau_{ij}^{(k)}/n \qquad \boldsymbol{\Sigma}_i^{(k+1)} = \frac{\sum_{j=1}^{n}\tau_{ij}^{(k)}(\boldsymbol{B}_i^{(k)}\boldsymbol{\eta}_j + \boldsymbol{\Gamma}_i^{(k)}\boldsymbol{\xi}_j)^T(\boldsymbol{B}_i^{(k)}\boldsymbol{\eta}_j + \boldsymbol{\Gamma}_i^{(k)}\boldsymbol{\xi}_j)}{\sum_{j=1}^{n}\tau_{ij}^{(k)}}$$

$$\boldsymbol{B}_i^{(k+1)} = \sum_{j=1}^{n}\tau_{ij}^{(k)}\boldsymbol{\Gamma}_i^{(k)}\boldsymbol{\xi}_j\boldsymbol{\eta}_j^T\left[\sum_{j=1}^{n}\tau_{ij}^{(k)}\boldsymbol{\eta}_j\boldsymbol{\eta}_j^T\right]^{-1} \qquad \boldsymbol{\Gamma}_i^{(k+1)} = \sum_{j=1}^{n}\tau_{ij}^{(k)}\boldsymbol{B}_i^{(k)}\boldsymbol{\eta}_j\boldsymbol{\xi}_j^T\left[\sum_{j=1}^{n}\tau_{ij}^{(k)}\boldsymbol{\xi}_j\boldsymbol{\xi}_j^T\right]^{-1} \tag{6}$$

In addition to the parameter constraints specified under the hypothetical model in (1) under (2), extra constraints at the component level may be required in the formulation of the final mixture model when collinearity between some endogenous variables is present. In this paper, we propose the following systematic scheme to determine which additional component-parameters in $\boldsymbol{B}_i$ $(i = 1, \ldots, g)$ need to be constrained to be zero:

1. Perform model estimation without any additional constraints;
2. Monitor the log likelihood values at each iteration and the parameter estimates of $\boldsymbol{B}_i$ ($i = 1, \ldots, g$);
3. Determine if the algorithm converges or not (failure to convergence is indicated by either singularity of $\boldsymbol{B}_i$ or decrease of log likelihood values due to estimate in $\boldsymbol{B}_i$, say b_{ilm}, being very close to zero, such as being less than 0.000001 in absolute value[7]);
4. Constrain the parameter b_{ilm}, if convergence fails to achieve in (3), to be zero and then rerun the model estimation;
5. Repeat (2) and (4) to constrain one parameter at a time[8] until convergence of model estimation is achieved.

Initialization, computation of standard errors, and model selection

With applications where the log likelihood equation has multiple local maxima, the EM algorithm should be implemented from a wide choice of initial parameter values in an attempt to search for all local maxima [7, 8]. The proposed algorithm provides three options to initialize the EM algorithm, where the user can either (a) specify initial estimates of the unknown parameters (such as those guided by estimates obtained by the standard PLS); (b) use random groupings of the data to get initial estimates of the unknown parameters; or (c) run the EM algorithm from different random starts as in (b) and use the set of parameter estimates corresponding to the largest likelihood value as initial values for obtaining the final model.

With the proposed algorithm, the standard errors of the estimates of $\boldsymbol{\Psi}$ are obtained using the bootstrap resampling method with replacement, where the number of bootstrap replications is taken to be 100 [7].

In the absence of any prior information as to the number of components present in the data, we can monitor the increase in log likelihood function as the value of g increases in order to determine an appropriate value of g. At any stage, the choice of $g = g_0$ versus $g = g_0 + 1$ can be made by using some information-based criterion, such as the Bayesian Information Criterion (BIC) [9] or by a bootstrap resampling approach to assess the null distribution (and hence the p-value) of the likelihood ratio test statistic [7]; see also [5] and [6]. There is also the integrated classification likelihood (ICL) criterion [1]. Other criteria for the determination of g, including the Akaike Information Criterion (AIC), the consistent AIC (CAIC), and the entropy measure (EN), have been considered specifically within the marketing research [3, 10]. Comparison of these methods in the general context of mixture models has been reported [7].

3 Simulation experiments

In this section, we study the performance of the proposed computational algorithm for fitting mixtures of sparse regression models. We consider a marketing research setting with $p = 5$ exogenous and $q = 7$ endogenous variables. Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_5)^T$ and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_7)^T$ be the scores of exogenous and endogenous variables, respectively, with the subscript j that indicates the jth individual dropped, the 7 simultaneous regression equations that define the path model are given by

[7] Other thresholds close to zero may be used and the choice should not affect the results.
[8] If constraints in multiple parameters are needed, sensitivity analysis may be used to determine the order.

$$\begin{aligned} \eta_1 &= \gamma_{11}\xi_1 + \zeta_1; & \eta_5 &= \gamma_{54}\xi_4 + \zeta_5;\\ \eta_2 &= \gamma_{22}\xi_2 + \zeta_2; & \eta_6 &= \gamma_{65}\xi_5 + \zeta_6;\\ \eta_3 &= b_{32}\eta_2 + \zeta_3; & \eta_7 &= b_{74}\eta_4 + b_{75}\eta_5 + b_{76}\eta_6 + \zeta_7,\\ \eta_4 &= b_{41}\eta_1 + b_{43}\eta_3 + \gamma_{43}\xi_3 + \zeta_4; \end{aligned} \tag{7}$$

which imply that the specifications for $\boldsymbol{B}_i$ and $\boldsymbol{\Gamma}_i$ $(i = 1, \ldots, g)$ are:

$$\boldsymbol{B}_i = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 0 & 0\\ 0 & \text{-}b_{i32} & 1 & 0 & 0 & 0 & 0\\ \text{-}b_{i41} & 0 & \text{-}b_{i43} & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & \text{-}b_{i74} & \text{-}b_{i75} & \text{-}b_{i76} & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Gamma}_i = \begin{bmatrix} \text{-}\gamma_{i11} & 0 & 0 & 0 & 0\\ 0 & \text{-}\gamma_{i22} & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & \text{-}\gamma_{i43} & 0 & 0\\ 0 & 0 & 0 & \text{-}\gamma_{i54} & 0\\ 0 & 0 & 0 & 0 & \text{-}\gamma_{i65}\\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{8}$$

In the simulation experiments, it is assumed that there are $g = 3$ groups of individuals and the total number of individuals is $n = 1000$. Each vector of the exogenous latent variable scores $\boldsymbol{\xi}_j$ $(j = 1, \ldots, 1000)$ was generated independently from a multivariate normal distribution with mean vector and covariance matrix as

$$\text{Mean} = \begin{pmatrix} \text{-}0.063\\ \text{-}0.131\\ \text{-}0.012\\ 0.080\\ \text{-}0.013 \end{pmatrix} \quad \text{and} \quad \text{Cov.} = \begin{bmatrix} 1.14 & 0.66 & 0.72 & 0.45 & 0.57\\ 0.66 & 1.19 & 0.53 & 0.29 & 0.43\\ 0.72 & 0.53 & 1.01 & 0.48 & 0.58\\ 0.45 & 0.29 & 0.48 & 0.99 & 0.47\\ 0.57 & 0.43 & 0.58 & 0.47 & 1.01 \end{bmatrix}. \tag{9}$$

The parameter values for $\boldsymbol{\Psi}$ with reference to (8) are given in Table 1; these parameter values are based on a fitted mixture model we have obtained on a real data set. Realizations of component membership were generated in which an individual has a probability of π_i to belong to the ith component $(i = 1, 2, 3)$. Given the component membership, realizations of $\boldsymbol{\eta}_j$ were then generated from the corresponding component density $\phi(\boldsymbol{\eta}_j|\boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_i)$ as in (2) under (7).

To illustrate the proposed scheme presented in Section 2 for the constraint of additional component-parameters in $\boldsymbol{B}_i$ $(i = 1, 2, 3)$, we consider collinearity between the seventh η_7 and the forth η_4 endogenous latent variables in (7) for the first component. This implies that both parameters b_{175} and b_{176} are zero, with a very small σ_{17}^2; see Table 1. Using a data set of $n = 1000$ scores generated as above, we first consider a mixture model without any additional constraints on parameters in $\boldsymbol{B}_i$ (see Equation (8)). The algorithm fails to converge as the estimate of b_{175} has a value smaller than 0.000001. We then consider a model with an additional constraint of $b_{175} = 0$. The algorithm again fails to converge as the estimate of b_{176} has a value smaller than 0.000001. We thus constrain $b_{176} = 0$ as well. This final model with two additional constraints ($b_{175} = 0$ and $b_{176} = 0$) converges.

Ten independent simulation experiments were conducted to assess the generalization performance of the proposed algorithm for fitting mixtures of sparse regression models. Such evaluation is based on the accuracy of the final model derived, the misclassification rate, and the bias of estimates. In all ten experiments, the algorithm identifies the correct final model with two additional constraints in b_{175} and b_{176} (the rate of correctly identifying sparse coefficients is 100%).

Parameter	$i=1$	$i=2$	$i=3$	Parameter	$i=1$	$i=2$	$i=3$
π_i	0.42	0.46	0.12	γ_{i54}	0.63	0.50	0.17
b_{i32}	0.97	0.64	0.67	γ_{i65}	1.02	0.94	-0.80
b_{i41}	0.60	0.14	0.27	σ^2_{i1}	0.07	0.64	0.39
b_{i43}	0.32	0.40	0.32	σ^2_{i2}	0.09	0.89	0.68
b_{i74}	0.87	0.20	0.49	σ^2_{i3}	0.07	0.60	0.44
b_{i75}	0.00	0.27	0.11	σ^2_{i4}	0.02	0.47	0.35
b_{i76}	0.00	0.24	0.21	σ^2_{i5}	0.58	0.82	0.82
γ_{i11}	0.91	0.67	0.74	σ^2_{i6}	0.01	0.01	0.73
γ_{i22}	1.16	0.49	0.29	σ^2_{i7}	1E-6	0.86	0.87
γ_{i43}	0.03	0.39	0.22				

Table 1: Parameter values for a 3-component mixture model (Simulation experiments).

Parameter	$i=1$	$i=2$	$i=3$	Parameter	$i=1$	$i=2$	$i=3$
π_i	-0.001	0.010	-0.009	γ_{i54}	-0.003	0.008	-0.017
b_{i32}	0.001	0.018	-0.043	γ_{i65}	0.001	0.001	-0.048
b_{i41}	-0.002	0.003	0.003	σ^2_{i1}	-0.001	0.006	0.031
b_{i43}	-0.004	0.002	0.010	σ^2_{i2}	0.003	-0.005	0.022
b_{i74}	-0.001	-0.008	0.045	σ^2_{i3}	0.002	-0.003	-0.018
b_{i75}	—	-0.011	0.016	σ^2_{i4}	0.001	-0.005	-0.011
b_{i76}	—	0.001	0.040	σ^2_{i5}	0.006	-0.001	-0.033
γ_{i11}	0.001	-0.006	0.003	σ^2_{i6}	0.001	0.001	0.009
γ_{i22}	0.002	0.011	0.008	σ^2_{i7}	0.000	-0.006	0.067
γ_{i43}	0.001	-0.004	-0.011				

Table 2: Average bias of estimates for a 3-component mixture model (Simulation experiments).

The average misclassification rate is 0.0137. The average bias of estimates are presented in Table 2. It can be seen that no appreciable bias is observed in the estimation of $\mathbf{\Psi}$.

4 Real example: Emotional behaviour of preschool children

This real example is based on the Early Head Start Research and Evaluation (EHSRE) project conducted from 1996 to 2001. The data set is available from the Inter-University Consortium for Political and Social Research (ICPSR) at http://www.icpsr.umich.edu. It contains data about 2977 children under 3 years who were randomized to receive designed Early Head Start (EHS) services or to seek their own early childhood care in their community; see, for example, [12].

In the current study, we considered $n = 1498$ individuals with complete observations in eight manifest variables and focus on the conceptual model described in [12] for hypothesized relationships among maternal mental health, parenting stress, parent-child routines, and child emotional development. The endogenous and exogenous latent variables of the hypothetical model are presented in Figure 1. In the inner model, there are $p = 1$ exogenous (maternal mental health) and $q = 3$ endogenous (parenting stress, parent-child routine and child emotional development) latent variables. The 3 simultaneous regression equations that define the path model are given by

$$\eta_1 = \gamma_{11}\xi_1 + \zeta_1; \qquad \eta_3 = b_{31}\eta_1 + b_{32}\eta_2 + \zeta_3.$$
$$\eta_2 = b_{21}\eta_1 + \zeta_2; \tag{10}$$

The standard PLS analysis is implemented using the "plspm" package in R [11] to obtain the scores for the 4 latent variables corresponding to the path model presented in Figure 1. The proposed algorithm is then used to fit mixtures of regression models to the scores of the latent variables with $g = 1$ to $g = 5$. No additional parameter constraints are necessary. Using the BIC, we identified two groups of individuals. The larger group ($i = 1$, $n_1 = 1434$) of individuals have all links in the hypothetical inner model significant; see Table 3 for the estimates of the path coefficients. Comparing to the majority, the smaller group ($n_2 = 64$) of individuals have smaller impact from maternal mental health on parenting stress (γ_{211}), and from parenting stress and parent-child routine on child emotional development (b_{231} and b_{232}). A post-hoc analysis finds that these two groups are significantly different in RACE (p-value $= 0.001$; see Table 3), but not in the program allocated, child gender, child overweight indicator, and maternal age at birth.

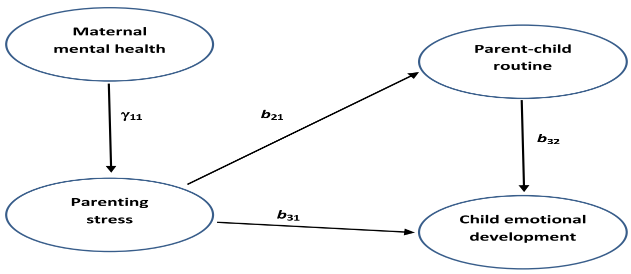

Figure 1: Hypothetical inner model relating maternal mental health, parenting stress, parent-child routines, and child emotional development

Group	b_{i21}	b_{i31}	b_{i32}	γ_{i11}	Race = Hispanic
$i = 1$	-0.178 (0.032)	-0.194 (0.033)	0.142 (0.028)	0.423 (0.021)	312/1357* (23.0%)
$i = 2$	-0.159 (0.039)	-0.074 (0.035)	0.071 (0.062)	0.237 (0.105)	26/61* (42.6%)

Table 3: Estimates (standard errors) of path coefficients for a 2-component mixture model and proportion of Hispanic children (* Missing data exist in both groups).

5 Discussion

We have developed a computational algorithm for fitting mixtures of regression models with latent variables and sparse coefficient parameters. The algorithm adopts a systematic scheme to determine which additional component-parameters in the matrices of path coefficients $\boldsymbol{B}_i$ need to be constrained to be zero. Simulated and real data sets have been used to illustrate the applicability of the proposed algorithm. The method can be readily adopted for component distributions that are not multivariate normal.

Acknowledgement

This work was supported by a grant from the Australian Research Council. The authors are grateful to the Inter-University Consortium for Political and Social Research (ICPSR) for providing the Early Head Start Research and Evaluation (EHSRE) project data (ICPSR ID 3804) for the illustration described in Section 4. The findings and views reported in this paper are those of the authors and should not be attributed to either the ICPSR or the EHSRE project.

Bibliography

[1] Biernacki, C., Celeux, G. and Govaert, G. (1998) *Assessing a mixture model for clustering with the integrated classification likelihood.* IEEE Transactions on Pattern Analysis and Machine Intelligence, **22**, 719–725.

[2] Dollman, J., Ridley, K., Magarey, A., Martin, M. and Hemphill, E. (2007) *Dietary intake, physical activity and TV viewing as mediators of the association of socioeconomic status with body composition: a cross-sectional analysis of Australian youth.* International Journal of Obesity, **31**, 45–52.

[3] Hahn, C., Johnson, M.D., Herrmann, A. and Huber, F. (2002) *Capturing customer heterogeneity using a finite mixture PLS approach.* Schmalenbach Business Review, **54**, 243–269.

[4] Jones, P.N. and McLachlan, G.J. (1992) *Fitting finite mixture models in a regression context.* Australian Journal of Statistics, **34**, 233–240.

[5] Keribin, C. (2000) *Consistent estimation of the order of mixture models.* Sankhyā: The Indian Journal of Statistics, **62**, 49–66.

[6] Leroux, B.G. (1992) *Consistent estimation of a mixing distribution.* Annals of Statistics, **20**, 1350–1360.

[7] McLachlan, G.J. and Peel, D. (2000) *Finite Mixture Models.* New York: Wiley.

[8] Ng, S.K. (2013) *Recent developments in expectation-maximization methods for analyzing complex data.* WIREs Computational Statistics, **5**, 415–431.

[9] Ng, S.K. and McLachlan, G.J. (2014) *Mixture models for clustering multilevel growth trajectories.* Computational Statistics and Data Analysis, **71**, 43–51.

[10] Ringle, C.M., Wende, S. and Will, A. (2010) *Finite mixture partial least squares analysis: Methodology and numerical examples.* In: Handbook of Partial Least Squares, V.E. Vinzi, W.W. Chin, J. Henseler and H. Wang (Eds.). Heidelberg: Springer, pp. 195–218.

[11] Sanchez, G. (2013) *PLS Path Modeling with R.* Berkeley: Trowchez Editions. Available from: `http://www.gastonsanchez.com/PLS_Path_Modeling_with_R.pdf`.

[12] Zajicek-Farber, M.L., Mayer, L.M. and Daughtery, L.G. (2012) *Connections among parental mental health, stress, child routines, and early emotional behavioral regulation of preschool children in low-income families.* Journal of the Society for Social Work and Research, **3**, 31–50.