University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2012   *Paper* 301

The Impact of Covariance Misspecification in Multivariate Gaussian Mixtures on Estimation and Inference: An Application to Longitudinal Modeling

Brianna C. Heggeseth[*]   Nicholas P. Jewell[†]

[*]University of California, Berkeley, brianna.c.heggeseth@williams.edu

[†]Division of Biostatistics, School of Public Health, University of California, Berkeley, jewell@berkeley.edu

The Impact of Covariance Misspecification in Multivariate Gaussian Mixtures on Estimation and Inference: An Application to Longitudinal Modeling

Brianna C. Heggeseth and Nicholas P. Jewell

Abstract

Multivariate Gaussian mixtures are a class of models that provide a flexible parametric approach for the representation of heterogeneous multivariate outcomes. When the outcome is a vector of repeated measurements taken on the same subject, there is often inherent dependence between observations. However, a common covariance assumption is conditional independence—that is, given the mixture component label, the outcomes for subjects are independent. In this paper, we study, through asymptotic bias calculations and simulation, the impact of covariance misspecification in multivariate Gaussian mixtures. Although maximum likelihood estimators of regression and mixing probability parameters are not consistent under misspecification, they have little asymptotic bias when mixture components are well-separated or if the assumed correlation is close to the truth even when the covariance is misspecified. We also present a robust standard error estimator and show that it outperforms conventional estimators in simulations and can indicate the model is misspecified. Body mass index data from a national longitudinal study is used to demonstrate the effects of misspecification on potential inferences made in practice.

1   Introduction

Multivariate Gaussian mixtures are a class of models that provide a flexible parametric approach for the representation of heterogeneous multivariate outcomes potentially originating from distinct subgroups in the population. An overview of finite mixture models is available in many texts [8, 41, 24, 25, 10]. We can estimate covariate effects on the outcome as well as group membership probabilities by extending mixture models to include a regression structure for both the mean and mixing proportions. See De la Cruz-Mesía et al. [5] for a review of finite mixture models with a regression mean structure and Wedel [45] for the history of concomitant variable models that use baseline variables to explain variation in subgroups. These extensions are used in several medical applications [38] including epidemiology, genomics, and pharmacology in addition to other fields including astronomy, biology, economics, and speech recognition. When the multivariate outcome is a vector of repeated measures taken over time, these methods are identified as group-based trajectory modeling [32, 33] or latent-class growth analysis [29, 27]. See Pickles and Croudace [36] and references within for a review of mixture methods applied to longitudinal data. The use of mixture models for multivariate data is increasing due to computational advances that have made maximum likelihood (ML) parameter estimation possible, via the Expectation Maximization (EM) algorithm, through model-specific packages such as Proc Traj in SAS [17], Flexmix [21] and mclust [9] in R, and software such as Mplus [31] and Latent Gold [42].

Despite the increased use of these models, the sensitivity of estimated regression coefficients to model assumptions has only been explored to a limited degree. In a multivariate mixture model, one must specify the component distribution, the form of the mean, the structure of the covariance matrix, and the number of components; therefore, there are many ways to misspecify the model. For example, in practice, the number of components is unknown and model selection procedures based on the Bayesian information criterion are often employed. However, if the specified covariance structure is too restrictive relative to the truth, the estimated number of components will typically be greater than the true number because more components are needed to model the extra variability. The literature in estimating the number of components is vast [34] and continues to debate this unresolved issue. Due to the potential complexity of mixture models, simplifying assumptions are made to reduce the dimension of the parameter space, to make estimation possible, and for computational convenience. In particular, many researchers assume Gaussian components and/or restrict the components to have equal variance, both of which are known to result in asymptotic bias if the assumptions are not met [11, 23]. In this paper, we assume that the number of components, mean structure, and distribution are known and focus on other indeterminacies such as the covariance matrix.

In terms of the covariance matrix, eigenvalue and Cholesky decompositions [1, 26], as well as mixed effects structures [30], are used to impose structure and parsimony. Additionally, one common assumption is conditional independence—given the mixture component label, the outcomes for a subject are assumed independent [35, 28]. Of the available software that estimate regression effects for the mean and mixing probabilities, most of them make this simplifying assumption. This restriction is convenient when the data are unbalanced or if the sample size is small to make estimation of the covariance parameters more stable. Despite the wealth of proposed covariance models, there has been little work done in the area of mixture models with misspecified covariance structures, and the conditional independence assumption is unlikely to hold in many multivariate data settings, specifically in longitudinal applications. If the mixture consists of one component, work done by Liang, Zeger [22] and others suggest that regression estimates are asymptotically unbiased. However, these properties do not hold with additional components since estimation includes mixing proportions as well as component parameters.

Here, we investigate the impact of covariance misspecification on ML estimation of parameters and standard errors in multivariate Gaussian mixture models. In particular, our focus is on the assumption of conditional independence for the covariance structure; therefore, we assume the number of components, the distribution, and the mean structure is known. This paper is organized as follows. Section 2 presents the model specification. Section 3 describes the estimation procedure, issues, and asymptotic properties of the parameter estimators based on the seminal results of White [46]. In Section 4, we present a series of simulations

1

of a simple misspecified example to compare asymptotic and finite-sample bias of parameter and standard error estimates under varying levels of dependence and separation between components. In Section 5, we apply these ideas to body mass index data from a national longitudinal study to demonstrate the effects of misspecification on potential inferences made in practice.

2   Model Specification

In a finite multivariate mixture, the density of a random vector $\mathbf{y}$ takes the form

$$f(\mathbf{y}) = \pi_1 f_1(\mathbf{y}) + \cdots + \pi_K f_K(\mathbf{y})$$

where $\pi_k > 0$ for $k = 1, ..., K$ and $\sum_k^K \pi_k = 1$. The parameters π_k are mixing proportions and the functions $f_1, ..., f_K$ are component densities, assumed multivariate Gaussian here.

We extend the general model to allow other factors to affect the mean as well as the mixing proportions. Let $\mathbf{y} \in \mathbb{R}^m$ be a random vector whose distribution, conditional on regression covariates, $\mathbf{x}$, and concomitant variables, $\mathbf{z}$, is a mixture of K Gaussian densities with mixing proportions: $\pi_1(\mathbf{z}, \boldsymbol{\gamma}), ..., \pi_K(\mathbf{z}, \boldsymbol{\gamma})$. That is, the conditional mixture density for $\mathbf{y}$ is defined by

$$f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) = \sum_{k=1}^K \pi_k(\mathbf{z}, \boldsymbol{\gamma}) f_k(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}_k) \tag{1}$$

where $f_k(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}_k)$ denotes the m-variate Gaussian probability density function with mean $\mathbf{x}\boldsymbol{\beta}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$ $(k = 1, ..., K)$, $\boldsymbol{\theta}_k$ includes both $\boldsymbol{\beta}_k$ and $\boldsymbol{\Sigma}_k$, $\mathbf{x}$ is a $m \times p$ matrix, and $\mathbf{z}$ is a vector of length q. The regression covariates include measures that affect the mean while the concomitant variables influence the mixing proportions. This general structure allows the possibility that some baseline variables could be in both $\mathbf{x}$ and $\mathbf{z}$.

We parameterize the mixing proportions using the multinomial logit with the form

$$\pi_k(\mathbf{z}, \boldsymbol{\gamma}) = \frac{\exp(\mathbf{z}^T \boldsymbol{\gamma}_k)}{\sum_{j=1}^K \exp(\mathbf{z}^T \boldsymbol{\gamma}_j)}$$

for $k = 1, ..., K$ where $\boldsymbol{\gamma}_k \in \mathbb{R}^q$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^T, ..., \boldsymbol{\gamma}_K^T)$ where $\boldsymbol{\gamma}_K = \mathbf{0}$.

Throughout this paper, we generally assume conditional independence with constant variance within a component where $\boldsymbol{\Sigma}_k = \sigma_k^2 I_m$ as the *proposed estimation model*, but it is straightforward to extend the covariance model to include other correlation structures such as exchangeable or exponential. Therefore, the vector of all unknown parameters, $\boldsymbol{\theta}$, consists of the mixing proportion parameters, $\boldsymbol{\gamma}_k$, and the component regression and variance parameters, $\boldsymbol{\theta}_k^T = (\boldsymbol{\beta}_k^T, \sigma_k^2)$, for $k = 1, ..., K$ and could include correlation parameters.

3   Estimation

3.1   EM Algorithm

Under the assumption that $\mathbf{y}_1, ..., \mathbf{y}_n$ are independent realizations from the mixture distribution, $f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \boldsymbol{\theta})$, defined in 1, the log likelihood function for the parameter vector, $\boldsymbol{\theta}$ is given by

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^n \log f(\mathbf{y}_i|\mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}).$$

2

The ML estimate of $\boldsymbol{\theta}$ is obtained by finding an appropriate root of the score equation, $\partial \log L(\boldsymbol{\theta})/\partial \boldsymbol{\theta} = \mathbf{0}$. Solutions of this equation corresponding to local maxima can be found iteratively through the Expectation-Maximization (EM) algorithm [6]. This algorithm is applied in the framework where given $(\mathbf{x}_i, \mathbf{z}_i)$ each $\mathbf{y}_i$ is assumed to have stemmed from one of the components and the indicator denoting its originating component is missing. The complete-data log likelihood is based on these indicator variables as well as the observed data $\{(\mathbf{y}_i, \mathbf{x}_i, \mathbf{z}_i)\}$. The Expectation step (E-step) involves replacing the indicators by current values of the conditional expectation, which is the posterior probability of component membership, written as

$$\alpha_{ik} = \pi_k(\mathbf{z}_i, \boldsymbol{\gamma}) f_k(\mathbf{y}_i | \mathbf{x}_i) / \sum_{j=1}^{K} \pi_j(\mathbf{z}_i, \boldsymbol{\gamma}) f_j(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\theta}_j)$$

for $i = 1, ..., n$ and $k = 1, ..., K$ using current estimates of the parameters. In the Maximization step (M-step), the parameter estimates for the mixing proportions, regression effects, and covariance matrices are updated by maximizing the complete-data log likelihood using the posterior probabilities from the E-step in place of the indicator variables using numerical optimization. The E- and M-steps are alternated repeatedly until convergence. The EM algorithm guarantees convergence to a local maximum; global convergence may be attained through initializing the algorithm by randomly assigning individuals to initial components, running the algorithm multiple times and using the estimates associated with the highest log likelihood.

3.2   Issues

Although an estimation tool exists, there are potential issues of parameter identifiability with mixture models. Frühwirth-Schnatter [10] distinguishes between three types of nonidentifiability: invariance to relabeling of components, potential overfitting, and nonidentifiability due to the family of component distribution and the covariate design matrix. The first two issues are resolved through constraints such as $\boldsymbol{\theta}_k \neq \boldsymbol{\theta}_{k'}$ for all $k, k' = 1, ..., K, \ k \neq k'$. The last concern is solved by assuming Gaussian components since finite mixtures of multivariate Gaussians are identifiable [40, 47]. However, Hennig [14] suggests that the introduction of a regression structure to a Gaussian mixture requires a full rank design matrix as well as a rich covariate domain for regression parameters to be identifiable. On the other hand, mixing proportions parameters from a multinomial logit based on concomitant variables are identifiable by setting the parameters of one component to zero such as $\boldsymbol{\gamma}_K = \mathbf{0}$ [16].

Besides identifiability, there are other known issues with finite mixture models. McLachlan and Peel [25] note that the sample size must be quite large for asymptotic theory to accurately describe the finite sampling properties of the estimator. Also, when component variances are allowed to vary between components, the mixture likelihood function is unbounded and each observation gives rise to a singularity on the boundary of the parameter space [4, 18]. However, Kiefer [19] outlines theory that guarantees that there exists a particular local maximizer of the mixture likelihood function that is consistent, efficient, and asymptotically normal if the mixture is not overfit. To avoid issues of singularities and spurious local modes in the EM algorithm, Hathaway [13] considers constrained maximum likelihood estimation for multivariate Gaussian mixtures based on the following constraint on the smallest eigenvalue of the matrix $\boldsymbol{\Sigma}_h \boldsymbol{\Sigma}_j^{-1}$, denoted as $\lambda_{\min}(\boldsymbol{\Sigma}_h \boldsymbol{\Sigma}_j^{-1})$,

$$\min_{1 \leq h \neq j \leq K} \lambda_{\min}(\boldsymbol{\Sigma}_h \boldsymbol{\Sigma}_j^{-1}) \geq c > 0$$

for some positive constant $c \in [0, 1]$ to ensure a global maximizer.

3.3   Asymptotic Properties

If the true underlying data-generating distribution is a member of the specified model class, then maximum likelihood estimation via the EM algorithm gives parameter estimates that are consistent [43, 20]. However, if the model specified does not contain the true underlying mixture, then the ML estimators potentially

have asymptotic bias [11, 23]. Here, we are interested in the impact of misspecifying the covariance matrix structure on parameter estimation and inference.

General theoretical results for ML estimators are given by White [46]. Our investigation is a special case where the covariance matrices of mixture components are incorrectly specified but the mean structure and distribution are known. Let $f(\mathbf{y}|\boldsymbol{\theta})$ be the assumed estimation model, $g(\mathbf{y})$ be the true density, and C be a compact subset of the parameter space. It follows that the ML estimator, $\hat{\boldsymbol{\theta}}_n$, is consistent for the parameter vector, $\boldsymbol{\theta}^*$, that minimizes the Kullback-Leibler (KL) divergence, $\int \log[g(\mathbf{y})/f(\mathbf{y}|\boldsymbol{\theta})]g(\mathbf{y})d\mathbf{y} = \int \log[g(\mathbf{y})]g(\mathbf{y})d\mathbf{y} - \int \log[f(\mathbf{y}|\boldsymbol{\theta})]g(\mathbf{y})d\mathbf{y}$, under some regularity conditions [46], which is equivalent to maximizing $\int \log[f(\mathbf{y}|\boldsymbol{\theta})]g(\mathbf{y})d\mathbf{y}$ with respect to $\boldsymbol{\theta}$.

In the case of mixture densities, this integral is mathematically intractable. Lo [23] used a modified EM algorithm for univariate data that maximized $\int \log[f(\mathbf{y}|\boldsymbol{\theta})]g(\mathbf{y})d\mathbf{y}$ with respect to $\boldsymbol{\theta}$ in order to estimate $\boldsymbol{\theta}^*$. This procedure could be adapted to bivariate data, but for outcome vectors of larger dimension, this procedure is not as useful. We know that for $\{\mathbf{y}_i\}_{i=1,\dots,n}$ generated from the true density, under suitable regularity conditions [15],

$$\sup_{\boldsymbol{\theta} \in C} \frac{1}{n} \sum_{i=1}^{n} \log f(\mathbf{y}_i|\boldsymbol{\theta}) \xrightarrow{a.s.} \sup_{\boldsymbol{\theta} \in C} \int_{-\infty}^{\infty} \log f(\mathbf{y}|\boldsymbol{\theta})g(\mathbf{y})d\mathbf{y} \quad \text{as } n \to \infty$$

Therefore, to investigate asymptotic bias under a misspecified covariance structure when $g(\mathbf{y})$ is known, we numerically approximate $\boldsymbol{\theta}^*$ using the EM algorithm on a large sample from $g(\mathbf{y})$ of size $n = 100,000$.

In addition to consistency, White [46] also showed that $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) \to N(0, C(\boldsymbol{\theta}^*))$ where the asymptotic covariance matrix is $C(\boldsymbol{\theta}^*) = A(\boldsymbol{\theta}^*)^{-1}B(\boldsymbol{\theta}^*)A(\boldsymbol{\theta}^*)^{-1}$, with

$$A(\boldsymbol{\theta}^*) = \left\{ E\left(\frac{\partial^2 \log f(\mathbf{y}_i|\boldsymbol{\theta}^*)}{\partial \theta_j \partial \theta_l} \right) \right\}, \quad B(\boldsymbol{\theta}^*) = \left\{ E\left(\frac{\partial \log f(\mathbf{y}_i|\boldsymbol{\theta}^*)}{\partial \theta_j} \cdot \frac{\partial \log f(\mathbf{y}_i|\boldsymbol{\theta}^*)}{\partial \theta_l} \right) \right\}.$$

Moreover, $C_n(\hat{\boldsymbol{\theta}}_n) = A_n(\hat{\boldsymbol{\theta}}_n)^{-1}B_n(\hat{\boldsymbol{\theta}}_n)A_n(\hat{\boldsymbol{\theta}}_n)^{-1} \xrightarrow{a.s.} C(\boldsymbol{\theta}^*)$, with

$$A_n(\hat{\boldsymbol{\theta}}_n) = \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log f(\mathbf{y}_i|\hat{\boldsymbol{\theta}}_n)}{\partial \theta_j \partial \theta_l} \right\}, \quad B_n(\hat{\boldsymbol{\theta}}_n) = \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log f(\mathbf{y}_i|\hat{\boldsymbol{\theta}}_n)}{\partial \theta_j} \cdot \frac{\partial \log f(\mathbf{y}_i|\hat{\boldsymbol{\theta}}_n)}{\partial \theta_l} \right\}.$$

Following a similar procedure as Boldea and Magnus [2], we derive the score vector and Hessian needed to calculate A_n and B_n for a multivariate Gaussian mixture model. Derivations are available from the first author.

If the model is correctly specified, then both $-A_n(\hat{\boldsymbol{\theta}}_n)^{-1}$ and $B_n(\hat{\boldsymbol{\theta}}_n)^{-1}$ are consistent estimators of $C(\boldsymbol{\theta}^*)$ [46], and two possible variance-covariance estimates for the parameter estimator are

$$\widehat{\text{Cov}}_1(\hat{\boldsymbol{\theta}}_n) = \widehat{\mathbf{W}}_1 = -(nA_n(\hat{\boldsymbol{\theta}}_n))^{-1}$$

$$\text{and}$$

$$\widehat{\text{Cov}}_2(\hat{\boldsymbol{\theta}}_n) = \widehat{\mathbf{W}}_2 = (nB_n(\hat{\boldsymbol{\theta}}_n))^{-1}$$

On the other hand, $A_n(\hat{\boldsymbol{\theta}}_n)^{-1}B_n(\hat{\boldsymbol{\theta}}_n)A_n(\hat{\boldsymbol{\theta}}_n)^{-1}$ provides a consistent estimator of $C(\boldsymbol{\theta}^*)$ *despite any misspecification*. Therefore, a third and *robust* variance estimate of the parameter estimator is given by

$$\widehat{\text{Cov}}_3(\hat{\boldsymbol{\theta}}_n) = \widehat{\mathbf{W}}_3 = n^{-1}A_n(\hat{\boldsymbol{\theta}}_n)^{-1}B_n(\hat{\boldsymbol{\theta}}_n)A_n(\hat{\boldsymbol{\theta}}_n)^{-1}.$$

We refer to calculated standard error estimates corresponding to these three indexed variance-covariance estimates throughout the rest of the paper.

4
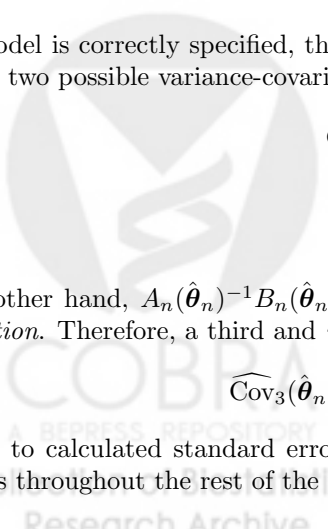
4   Simulations

We carry out two series of simulations to examine the behavior of the maximum likelihood estimators in terms of bias under misspecification of the covariance structure for finite samples from a multivariate mixture. Specifically, we are mainly interested in the impact of dependence in the true error structure on bias in parameter and standard error estimates when the conditional independence is assumed incorrectly and how this is affected by the (i) level of dependence and (ii) the separation between mixture components. Secondly, we investigate the behavior of the bias when the estimation structure gets closer to the true correlation structure by comparing the bias under three correlation structures.

In all of the simulations, the data with sample size n are generated from an m-variate Gaussian mixture model with parameters $(\boldsymbol{\gamma}_k, \boldsymbol{\beta}_k, \sigma_k^2, \mathbf{V}_k)$ for $k = 1, ..., K$ where $\mathbf{V}_k$ is the true correlation structure as follows:

- Fix K.

- For each subject, $i = 1, ..., n$,

    - Fix $\mathbf{x}_i = \mathbf{1}_m$ and $\mathbf{z}_i = 1$.
    - Construct matrices $\mathbf{A}_k$ such that $\mathbf{A}_k \mathbf{A}_k^T = \mathbf{V}_k$ for $k = 1, .., K$ using the Cholesky decomposition.
    - Randomly assign group membership, h_i, by drawing a value from the categorical distribution defined by $P(h = k) = \pi_k(\mathbf{z}_i, \boldsymbol{\gamma})$ for $k = 1, ..., K$.
    - Draw m standard normal random values $\mathbf{e}_i$ and let

$$\mathbf{y}_i = \mathbf{x}_i^T \boldsymbol{\beta}_{h_i} + \sigma_{h_i} \mathbf{A}_{h_i} \mathbf{e}_i$$

Thus, $\mathbf{y}_i \sim N(\mathbf{x}_i^T \boldsymbol{\beta}_{h_i}, \sigma_{h_i}^2 \mathbf{V}_{h_i})$. We then estimate the parameters and standard errors, $\widehat{SE}_1$, $\widehat{SE}_2$, $\widehat{SE}_3$, using constrained maximum likelihood via the EM algorithm [13] doing five random initializations, based on a multivariate mixture model with a specified correlation structure and known design matrix.

For simplicity, we focus on a example of two Gaussian components ($K = 2$) with constant mean vectors (i.e. no relationship between covariates and $\mathbf{y}$), one component with independent errors, the other with some level of dependency in the errors. For the first series, the latter dependence is based on an exchangeable correlation structure where all outcomes in an observational unit are equally correlated, which is mathematically equivalent to a random intercept model.

To investigate the influence of the level of dependence, we set the vector length to $m = 5$, equal mixing proportions ($\gamma_1 = 0$ and baseline variables have no effect), mean of the components to $\beta_1 = 1$ and $\beta_2 = 3$, and the variance of the components to $\sigma_1^2 = 0.25$ and $\sigma_2^2 = 1$. The errors are independent ($\mathbf{V}_1 = I_m$) in component one and we let the level of dependence vary with $\rho = 0, 0.5, 0.99$ within the exchangeable structure ($\mathbf{V}_2 = \rho(J_m - I_m) + I_m$ where J_m is a $m \times m$ matrix of 1's) for component two. We present the bias of parameter estimates and the three standard error estimates under these conditions.

Then, we consider the separation between two component distributions using the concept of c-separation [3]. Two Gaussian distributions, $N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, in $\mathbb{R}^m$ are c-separated if $||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2|| \geq c\sqrt{m \cdot \max(\lambda_{max}(\boldsymbol{\Sigma}_1), \lambda_{max}(\boldsymbol{\Sigma}_2))}$ where $\lambda_{max}(\boldsymbol{\Sigma})$ is the largest eigenvalue of $\boldsymbol{\Sigma}$. Dasgupta [3] notes that two Gaussians are almost completely non-overlapping when $c = 2$. This inequality can be rearranged to establish a measure of separation,

$$S = ||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2|| / \sqrt{m \cdot \max(\lambda_{max}(\boldsymbol{\Sigma}_1), \lambda_{max}(\boldsymbol{\Sigma}_2))},$$

which is a standardized Euclidean distance between mean vectors. In this simulation, we calculate the value of S for data-generating component densities as a measure of the separation between the two components

5

and if $S > 2$, the components do not overlap and are well-separated. For this series of simulations, we again use a strong level of dependence ($\rho = 0.99$) in the exchangeable structure, a vector length of $m = 5$, but vary the mean and variance of the second component ($\beta_2 = 3, 5$ and $\sigma_2^2 = 0.25, 1, 4$) to invoke different degrees of separation between components.

We perform 1000 replications of each simulation for sample sizes $n = 100, 500, 1000$. We approximate the true standard error with the standard deviation of the replicates. To estimate the asymptotic bias of the model parameters ($n = \infty$), we complete one replication with $n = 100,000$.

The two prong simulation described above focuses on the impact of using a conditional independence estimation model under different levels of dependence and separation in the data-generating components. In practice, we can choose correlation structures other than conditional independence. To explore the bias under difference covariance assumptions, we run a short simulation adjusting the data-generating model from above to use an exponential correlation structure, such that the dependence decreases as the time lag increases, rather than the constant dependence from the exchangeable structure. Therefore, for component two, the correlation between two measurements within a subject that are observed d time units apart is $\exp(-d/r)$ where r, the range parameter, determines how quickly the correlation decays to zero. This structure is general enough so that if r is close to zero, the correlation matrix is close to conditional independence and if r is very large, the structure is close to exchangeable correlation.

For this simulation, we continue using the two Gaussian components ($K = 2$) with constant mean vectors ($\beta_1 = 1$ and $\beta_2 = 3$) of length $m = 5$ with observations at times $t = 1, 2, 3, 4$ and 5, one component with independent errors, and the second component with a moderate level of dependence that decays exponentially ($r = 3$). We estimate the mixing proportion, mean, and variance parameters assuming different correlation structures: conditional independence, exchangeable, and exponential correlation. We estimate and compare the finite-sample bias of parameters and standard errors by letting $n = 500$ with 1000 replications. Additionally, we compare the conventional estimate $\widehat{\mathbf{W}}_1$ of the covariance matrix of $\widehat{\boldsymbol{\theta}}$ and the robust estimate $\widehat{\mathbf{W}}_3$ since if the estimation model is close to the true structure, the matrices should be similar and $\mathbf{Q} = \widehat{\mathbf{W}}_1^{-1}\widehat{\mathbf{W}}_3$ should be close to the identity matrix. We calculate $RJ = tr(\mathbf{Q})/\nu$ where ν is the length of $\widehat{\boldsymbol{\theta}}$, which has been termed the RJ criteria and should be close to 1 if the estimation model is close to the truth [39, 37].

4.1   Results

Table 1 lists bias estimates for the dependence-varying simulation study. The estimates range from close to zero when $\rho = 0$ to magnitudes of upwards of 0.3 when $\rho = 0.99$. It is clear from this table that stronger dependence in the errors results in greater finite-sample and asymptotic bias when estimating under the conditional independence assumption. Additionally, the magnitude of bias seems to reach the level of asymptopia at sample sizes of $n = 500$, but it is important to note that the estimates for the asymptotic bias, based on one replication with $n = 100,000$, are only numerically accurate to two decimal places for γ_1, σ_1^2, and σ_2^2 even when using computationally large sample sizes. We see this numerical inaccuracy when $\rho = 0$ since the asymptotic bias should be zero when the conditional independence assumption is met. In terms of standard error estimates, the bias increases with increased dependence with values ranging from 0.001 when $\rho = 0$ to 0.111 when $\rho = 0.99$. We see a divergence between the three variance estimators with $\widehat{SE}_3$, the robust estimator, consistently having the least bias (Table 2). When the model is correctly specified with $\rho = 0$, the three estimators are similar as supported by asymptotic theory.

Figure 1 shows that the relationship between the level of component separation and the magnitude of bias is complex. As in the previous simulation, sample sizes of $n = 500$ and larger produce similar bias estimates so we only present the asymptotic results. When the level of separation is high, $S > 2$, then the magnitude of the bias is small, but when there is some overlap, $S < 2$, there is not a clear, consistent
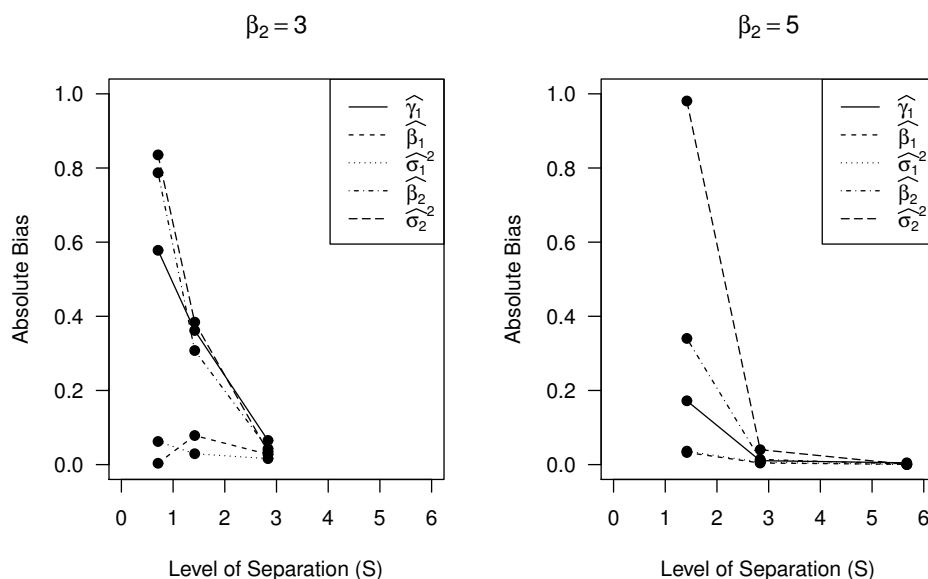
6

Figure 1: Asymptotic bias estimates of maximum likelihood parameter estimators when the covariance structure of a 2-component Gaussian mixture is assumed to be conditionally independent based on 1 replication with $n = 100,000$ under each mixture distribution with $m = 5$, $\gamma_1 = \gamma_2 = 0$, $\beta_1 = 1$, $V_1 = I_m$, $\sigma_1^2 = 0.25$, $V_2 = V(\rho)$, and $\rho = 0.99$ where $V(\rho)$ is the exchangeable correlation matrix. The level of separation (S) is calculated using the true mixture distribution. For $\beta_2 = 3$, variance parameters, $\sigma_2^2 = 0.25, 1, 4$ result in $S = 2.836, 1.418, 0.709$, respectively. For $\beta_2 = 5$, variance parameters, $\sigma_2^2 = 0.25, 1, 4$ result in $S = 5.671, 2.836, 1.418$, respectively. Values of $S \geq 2$ indicate almost completely separated components.

relationship between the value of S and the magnitude of the estimated bias for all parameters. That is, for two sets of parameter values, such as $(\beta_2 = 3, \sigma_2^2 = 0.25)$ and $(\beta_2 = 5, \sigma_2^2 = 1)$, that have the same level of separation, $S = 2.836$, the magnitude of the bias for all parameter estimates is drastically different for the two settings. However, in general, the bias decreases as the level of separation increases for a fixed mean parameter. The only exception is that the estimator for the first component mean $(\widehat{\beta}_1)$ has increased bias when $S = 1.418$ as compared to $S = 0.709$, but the bias then decreases when $S = 2.836$. It appears that when there is high overlap between two components, there is a point at which the bias peaks and then starts to decrease as σ_2 increases even though the amount of overlap continues to increase. Lastly, similar to the parameter estimates, the greater amount of separation results in less bias in the standard errors with biases as large as 1.0 unit in the situation with the most overlap and as little as less than 0.001 when components are well-separated. Again, the robust estimator again has the lowest bias. Tables available upon request. The simulations based on dependence and separation demonstrate the finite-sample and asymptotic bias in the ML estimators when the covariance structure is misspecified as conditional independence and the mixture components overlap. However, if two components are well-separated, the misspecification of the dependence in the errors does not result in large biases and thus any finite-sample bias could be removed potentially conventional techniques such as bootstrapping with careful tracking of component labels [12]. Additionally, when there is no covariance misspecification or when components are well-separated, all of the standard error estimates are similar and have little bias. However, when there is misspecification in the dependence structure, the estimates basely solely on the Hessian matrix or the score vector understate the true variability while the robust estimate has little bias. In cases where the true level of dependence is high, the bias in the Hessian estimator, $\widehat{SE}_1$, and the robust version, $\widehat{SE}_3$, can differ by as much as a relative
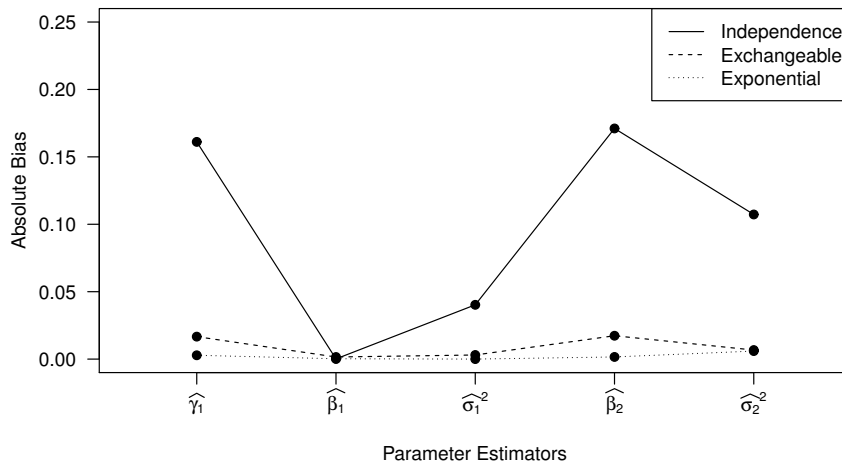
7

Figure 2: Bias estimates of maximum likelihood parameter estimators when the covariance structure of a 2-component Gaussian mixture is assumed to be conditionally independent, exchangeable, and exponential structure based on 1000 replications under each mixture distribution with $n = 500$, $m = 5$, $\gamma_1 = \gamma_2 = 0$, $\beta_1 = 1$, $V_1 = I_m$, $\sigma_1^2 = 0.25$, $\beta_2 = 3$, $V_2 = V(r)$ and $\sigma_2^2 = 2$ where $V(r)$ is the exponential correlation matrix and $r = 3$. Mean values of the RJ criteria are $RJ = 1.97, 1.02, 0.99$ for the three covariance assumptions, respectively.

factor of 2. In simulations not shown, using unequal mixing proportions result in similar conclusions. When the component proportions are unbalanced, the magnitude of bias increases when a majority of observations units originate from the misspecified component (here component two).

Figure 2 shows the absolute bias of parameter estimates under the three different covariance assumptions when the data was generating with the exponential correlation structure for component two. As expected, when the model is correctly specified, there is very little bias. We note that assuming the exchangeable structure, while incorrect, results in less bias than assuming conditional independence. As the RJ criteria gets closer to 1 from 1.97 to 1.02 to 0.99 using independence, exchangeable, and then exponential, the estimation correlation structure gets closer to the true structure resulting in little bias in the parameter estimates.

5   Data Example

To look at the behavior of the parameter and standard error estimates in practice, we use data from the 1979 National Longitudinal Survey of Youth (NLSY79). The NLSY79 is a nationally representative sample of 12,686 young American men and women aged 14-22 years in 1979, that oversampled Hispanic, Black, and economically disadvantaged non-Black/non-Hispanics. The cohort, interviewed annually from 1979 to 1994 and biennially thereafter, provide health and economic data for a total of 23 interviews (until 2008). In particular, the available body weight data for the 1979 cohort span a twenty-five year period [35]. We study body mass index (BMI) over time as it is an important longitudinal measure for public health and elucidating obesity development. Self-reported weight was collected in 17 interviews and height in five of those. BMI [weight (kg)/height (m^2)] was calculated for each interview based on the weight and the average height.

8

For the purposes of this paper, the complex sampling structure is ignored and we randomly sample 500 subjects who were at least 18 years of age in 1981 and reported all 17 weight measurements. Of this sample, 51% are female, 54% are non-Hispanic/non-Black, 29.2% Black and 16.8% Hispanic. To model the BMI outcomes, we allow a quadratic relationship between mean BMI and age and include sex as a baseline concomitant variable. Therefore, for $i = 1, ..., 500$, we assume that the observed data were generated according to

$$BMI = \beta_{k0} + \beta_{k1} \cdot (AGE - 18) + \beta_{k2} \cdot (AGE - 18)^2 + \epsilon_k$$

with probability

$$\pi_k(\mathbf{z}, \boldsymbol{\gamma}) = \frac{e^{\gamma_{k0} + I(male)\gamma_{k1}}}{\sum_j e^{\gamma_{j0} + I(male)\gamma_{j1}}}$$

where $\epsilon_k \sim N(0, \sigma_k^2 \mathbf{V}_k)$ for $k = 1, ..., 4$. The choice of four groups is based on previous research [35]. Using the EM algorithm with five random initializations, we estimate parameters and standard errors and present the estimates that produced the highest log likelihood. For the sake of comparison, we complete the estimation assuming conditional independence ($\mathbf{V}_k = I_m$), and under an exchangeable ($\mathbf{V}_k = \rho_k(J_m - I_m) + I_m$) and exponential ($\mathbf{V}_k = \exp(-D/r_k)$ where D is the Euclidean distance matrix of the ages at interviews for a subject) correlation model.

5.1   Results

Parameter and standard errors are estimated for a four component multivariate Gaussian mixture model assuming conditional independence, exchangeable, and exponential correlation (Table 3). The regression parameter estimates are used to calculate the mean curves for the four groups under all three covariance assumptions and we see that the mean curves differ between the models mainly in terms of the innermost curves (Figure 3). Under exchangeable correlation, one of the middle curves represents little BMI increase over time in contrast to the other groups. Under the exponential correlation assumption, the two lowest groups have a similar pattern over time, but the dependence differs between these groups with the range parameters estimated as $r_1 = 2.973$ and $r_2 = 23.579$ indicating that component 2 has more long range dependence between the BMI outcomes than component 1. Our simulation results suggest the magnitude of bias in the parameter estimator depends on how close the estimation correlation structure is to the truth and the overlap between components. We note there are no well-separated components and we see bias in the mean estimates by comparing the three covariance assumptions.

Given that the repeated outcome is BMI, we expect some dependence in the error structure within individuals. We consider the level of dependence in errors by plotting the estimated autocorrelation function by calculating the empirical variogram of the residuals from the conditional independence model [7] for each estimated component by randomly assigned each individual to a component using posterior probabilities [44]. The estimated autocorrelation function of the residuals shows strong dependence between residuals within 5 to 10 years and the correlation decreases with increasing time lags (Figure 4). This correlation structure is therefore neither consistent with conditional independence nor exchangeable correlation, but rather decreases to zero which is more consistent with the exponential correlation structure. We see that the robust standard estimators are almost twice those of estimates using the standard estimators under conditional independence, and the RJ criteria, which compares the naive and robust estimates of the covariance matrix the parameters, suggests that the exponential correlation structure is the one closest to the truth.

In this data example, we see the influence of the covariance structure on the estimates, especially in terms of the regression parameters. Based on the simulation results and the RJ criteria, we expect the exponential correlation model fits the data the best out of the three structures. However, we note that we fixed the number of components to be four for the sake of consistency and this may not be the optimal number of components. In practice, this value is estimated from the data as mentioned earlier. This data application highlights the impact of covariance misspecification as well as the fact that the mean structure may not be

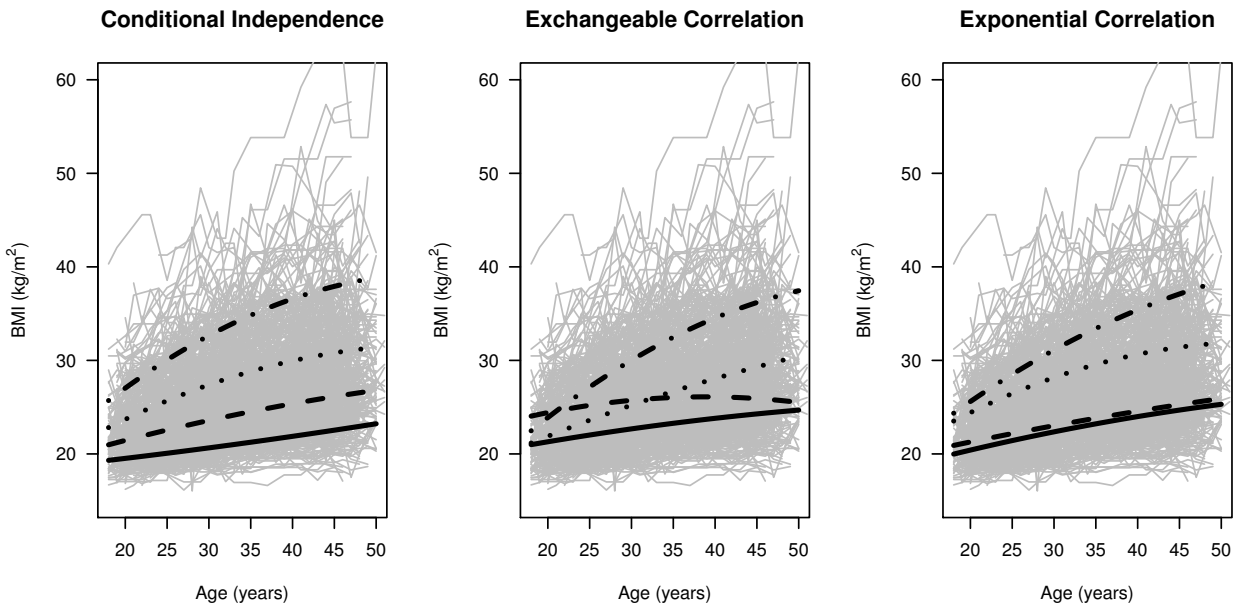Figure 3: Random sample of 500 BMI trajectories from NLSY and mean curves for the four components estimated using a Gaussian mixture model specified with a quadratic mean under the covariance assumptions: conditional independence, exchangeable, and exponential correlation. The labeled are consistent with the tables in the text: component 1 (solid), component 2 (dashed), component 3 (dotted), and component 4 (dashed-dot).

10

the only aspect differing between individuals; the level of dependence and variability also distinguish groups of individuals.

6   Discussion

We have shown that covariance misspecification of a two-component Gaussian mixture may produce very little bias in regression and mixing probability parameter estimates when the components are well-separated. This is well-aligned with Lo's univariate findings [23]. However, when there is some overlap in the component distributions, assuming the wrong correlation structure can produce asymptotically biased parameter estimates, the magnitude dependent on the level of separation and how close the structure is to the truth. With misspecified mixture models, the potential for biased mixing and regression parameters estimates differs from the one component models for which general estimating equations [22] produce unbiased estimates despite dependence present in the errors. Depending on the context and precision of the estimates, the bias may or may not have practical significance, but it is important to note that the ML estimators are inconsistent under covariance misspecification and there may be substantial bias when the components are not well-separated.

In addition to potential biases in the parameter estimates, the simulations provide evidence that conventional standard errors estimates which are based solely on the score equation, or the Hessian, can be extremely biased and underestimate the true variability of the estimators when the covariance structure is misspecified. Therefore, standard errors should be robustly estimated using White's estimator that sandwiches the two conventional estimators. We use the exact formula for this estimator since the numerical approximations to the Hessian matrix and score vector are not by-products produced by the EM algorithm. To the authors' knowledge, very few software programs currently have implemented a robust standard error estimator, but it should be implemented in every mixture model software as the default variance estimator and presented along with standard estimators to allow for comparisons, calculation of the RJ criteria, and the detection of misspecification bias.

Given our results, we recommend three things when estimating parameters in a mixture model. First, count the number of subjects whose maximum posterior probability is less than 0.95. If this count is non-zero, this indirectly indicates that the component distribution are not well-separated, suggesting that specifying the correct correlation structure is important. Second, if the components are not well-separated, fit the mixture model using several correlation structures such as conditional independence, exchangeable, and exponential correlation. For each model, calculate the RJ criteria based on the conventional and robust estimated variance-covariance matrix. Compare the parameter estimates to see if they change under the different assumptions and assess the RJ criteria values to see which structure results in a value closest to one. Choose the most parsimonious model that has an RJ criteria value close to one. Third, if none of these three structure fulfills this requirement, consider a more complex, potentially non-stationary covariance matrix as well as other sources of misspecification such as an incorrect number of components, assumed distribution, or an inflexible mean structure.

Our simulation study is limited, but the results likely apply to more complex mean structures and a larger number of components. In future studies, the impact of bias should be explored for more than two components with all components potentially having a misspecified covariance structure and for non-stationary covariance structures. Additionally, mixture models as specified in this paper group individuals with similar trajectories over time; we are currently investigating methods that distinguish between the shape of the trajectory and the vertical level of the curve when grouping individuals together.
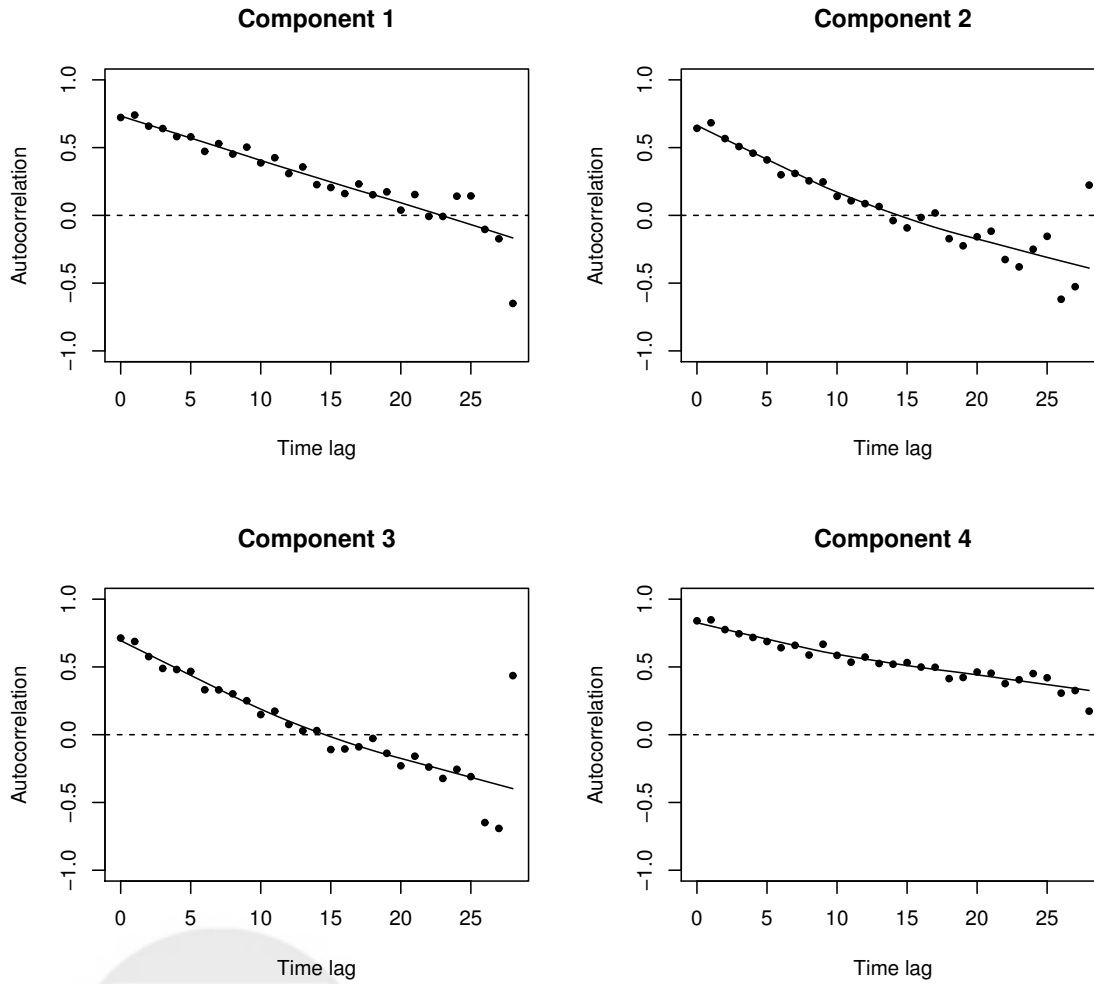
11

Figure 4: Smoothed sample autocorrelation of component residuals of estimated Gaussian mixture model specified with a quadratic mean and conditional independence with a random sample of 500 BMI trajectories from NLSY randomly assigned to components based on estimated posterior probabilities. The dashed line represents the estimated variance for each component.
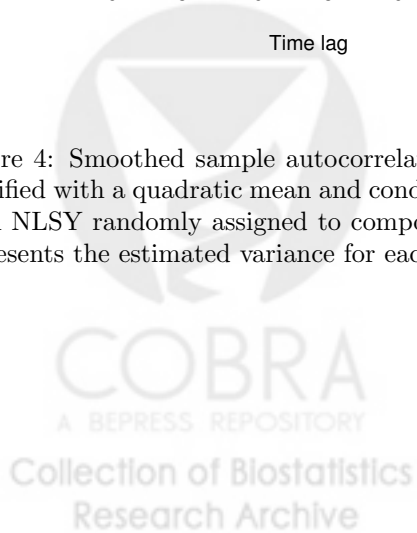
References

[1] J. D. Banfield and A. E. Raftery. Model-based gaussian and non-gaussian clustering. *Biometrics*, 49(3):803–821, 1993.

[2] O. Boldea and J.R. Magnus. Maximum likelihood estimation of the multivariate normal mixture model. *Journal of the American Statistical Association*, 104(488):1539–1549, 2009.

[3] S. Dasgupta. Learning mixtures of gaussians. In *Foundations of Computer Science, 1999. 40th Annual Symposium on*, pages 634–644. IEEE, 1999.

[4] N.E. Day. Estimating the components of a mixture of normal distributions. *Biometrika*, 56(3):463–474, 1969.

[5] R. De la Cruz-Mesía, F. A. Quintana, and G. Marshall. Model-based clustering for longitudinal data. *Computational Statistics & Data Analysis*, 52:1441–1457, January 2008.

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[7] P. Diggle, P. Heagerty, K.Y. Liang, and S.L. Zeger. *Analysis of longitudinal data*. Oxford University Press, USA, 2 edition, 2002.

[8] B.S. Everitt and D.J. Hand. *Finite Mixture Distributions*. Chapman & Hall, London, 1981.

[9] C. Fraley and A.E. Raftery. Mclust: Software for model-based cluster analysis. *Journal of Classification*, 16(2):297–306, 1999.

[10] S. Frühwirth-Schnatter. *Finite Mixture and Markov Switching Models*. Springer, New York, 2006.

[11] G. Gray. Bias in misspecified mixtures. *Biometrics*, 50(2):457–470, 1994.

[12] B. Grün and F. Leisch. Bootstrapping finite mixture models. In *Compstat 2004–Proceeding in Computational Statistics, Antoch J (ed.)*, pages 1115–1122. Heidelberg, Germany: Physica Verlag, 2004.

[13] R.J. Hathaway. A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *The Annals of Statistics*, 13(2):795–800, 1985.

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

[15] R.I. Jennrich. Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 40(2):633–643, 1969.

[16] W. Jiang and M.A. Tanner. On the identifiability of mixtures-of-experts. *Neural Networks*, 12(9):1253–1258, 1999.

[17] B.L. Jones, D.S. Nagin, and K. Roeder. A SAS procedure based on mixture models for estimating developmental trajectories. *Sociological Methods & Research*, 29(3):374, 2001.

[18] J. Kiefer and J. Wolfowitz. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *The Annals of Mathematical Statistics*, 27(4):887–906, 1956.

[19] N.M. Kiefer. Discrete parameter variation: Efficient estimation of a switching regression model. *Econometrica*, 46:427–434, 1978.

[20] L.M. Le Cam. On some asymptotic properties of maximum likelihood estimates and related bayes estimates. *Univeristy of California Publications in Statistics*, 1:277–330, 1953.

[21] F. Leisch. FlexMix: A general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software*, 11(8):1–18, 2004.

[22] Kung-Yee Liang and Scott L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

[23] Y. Lo. Bias from misspecification of the component variances in a normal mixture. *Computational Statistics & Data Analysis*, 55:2739–2747, 2011.

[24] G. J. McLachlan and K. E. Basford. *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker, New York/Basel, 1988.

[25] G. J. McLachlan and D. Peel. *Finite Mixture Models*. Wiley Series in Probability and Statistics. Wiley, New York, 2000.

[26] P. D. McNicholas and T. B. Murphy. Model-based clustering of longitudinal data. *Canadian Journal of Statistics*, 38(1):153–168, 2010.

[27] B. Muthén. Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class–latent growth modeling. In L.M. Collins and A.G. Sayers, editors, *New methods for the analysis of change*, pages 291–322. American Psychological Association, Washington, D.C., 2001.

[28] B. Muthén, H. Brown, A. Leuchter, and A. Hunter. General approaches to analysis of course: Applying growth mixture modeling to randomized trials of depression medication. In P. Shrout and K. Keyes, K.and Ornstein, editors, *Causality and psychopathology: Finding the determinants of disorders and their cures*, pages 159–178. Oxford University Press, New York, 2008.

[29] B. Muthén and L.K. Muthén. Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, 24(6):882–891, 2000.

[30] Bengt Muthén and Kerby Shedden. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55(2):463–469, 1999.

[31] L.K. Muthén and B.O. Muthén. Mplus user's guide, 1998-2010.

[32] D.S. Nagin. Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological methods*, 4(2):139–157, 1999.

[33] D.S. Nagin. *Group-based modeling of development*. Harvard University Press, 2005.

[34] A. Oliveira-Brochado and F. V. Martins. Assessing the number of components in mixture models: a review. FEP Working Papers 194, Universidade do Porto, Faculdade de Economia do Porto, November 2005.

[35] Truls Ostbye, Rahul Malhotra, and Lawrence R Landerman. Body mass trajectories through adulthood: results from the national longitudinal survey of youth 1979 cohort (1981-2006). *International Journal of Epidemiology*, 40(1):240–250, 2011.

[36] A. Pickles and T. Croudace. Latent mixture models for multivariate and longitudinal outcomes. *Statistical Methods in Medical Research*, 19(3):271–289, 2010.

[37] A. Rotnitzky and N.P. Jewell. Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika*, 77(3):485–497, 1990.

[38] Peter Schlattmann. Medical applications of finite mixture models. *International Statistical Review*, 79(2):294–294, 2011.

14

[39] Justine Shults, Wenguang Sun, Xin Tu, Hanjoo Kim, Jay Amsterdam, Joseph M. Hilbe, and Thomas Ten-Have. A comparison of several approaches for choosing between working correlation structures in generalized estimating equation analysis of longitudinal binary data. *Statistics in Medicine*, 28(18):2338–2355, 2009.

[40] H. Teicher. Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 34(4):1265–1269, 1963.

[41] D. M. Titterington, A. F. M. Smith, and U. E. Makov. *Statistical Analysis of Finite Mixture Distributions*. Wiley Series in Probability and Statistics. Wiley, New York, 1985.

[42] J.K. Vermunt and J. Magidson. Latent gold 4.0 user's guide, 2005.

[43] A. Wald. Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics*, 20(4):595–601, 1949.

[44] C.P. Wang, C.H. Brown, and K. Bandeen-Roche. Residual diagnostics for growth mixture models. *Journal of the American Statistical Association*, 100(471):1054–1076, 2005.

[45] M. Wedel. Concomitant variables in finite mixture models. *Statistica Neerlandica*, 56(3):362–375, 2002.

[46] Halbert White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25, 1982.

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

15

		Bias Estimates				
ρ	n	$\widehat{\gamma}_1$	$\widehat{\beta}_1$	$\widehat{\sigma}_1^2$	$\widehat{\beta}_2$	$\widehat{\sigma}_2^2$
0.00	100	-0.004 (0.006)	0.000 (0.001)	-0.000 (0.001)	-0.002 (0.002)	-0.006 (0.003)
	500	0.005 (0.003)	-0.000 (0.000)	-0.000 (0.000)	0.001 (0.001)	-0.000 (0.001)
	1000	0.004 (0.002)	-0.000 (0.000)	-0.000 (0.000)	0.001 (0.001)	-0.000 (0.001)
	∞	-0.001	0.000	0.001	-0.000	0.002
0.50	100	0.125 (0.006)	0.031 (0.001)	0.024 (0.001)	0.106 (0.003)	-0.136 (0.003)
	500	0.125 (0.003)	0.028 (0.001)	0.024 (0.000)	0.101 (0.002)	-0.124 (0.001)
	1000	0.125 (0.002)	0.028 (0.000)	0.024 (0.000)	0.098 (0.001)	-0.125 (0.001)
	∞	0.115	0.027	0.024	0.095	-0.125
0.99	100	0.370 (0.007)	0.087 (0.002)	0.033 (0.001)	0.327 (0.005)	-0.410 (0.005)
	500	0.346 (0.003)	0.078 (0.001)	0.030 (0.001)	0.310 (0.002)	-0.388 (0.002)
	1000	0.350 (0.002)	0.079 (0.001)	0.030 (0.000)	0.309 (0.001)	-0.385 (0.001)
	∞	0.353	0.082	0.031	0.315	-0.383

Table 1: Bias estimates (SE) of maximum likelihood parameter estimators when the covariance structure of a 2-component Gaussian mixture is assumed to be conditionally independent based on 1000 replications under each mixture distribution with $m = 5$, $\gamma_1 = \gamma_2 = 0$, $\beta_1 = 1$, $V_1 = I_m$, $\sigma_1^2 = 0.25$, $\beta_2 = 3$, $V_2 = V(\rho)$ and $\sigma_2^2 = 1$ where $V(\rho)$ is the exchangeable correlation matrix. Asymptotic estimates ($n = \infty$) are based on one replication with $n = 100,000$. Values equal to zero represent values less than 0.001.

					Bias Estimates					
ρ	n	$\widehat{SE}_1(\widehat{\gamma}_1)$	$\widehat{SE}_2(\widehat{\gamma}_1)$	$\widehat{SE}_3(\widehat{\gamma}_1)$	$\widehat{SE}_1(\widehat{\beta}_1)$	$\widehat{SE}_2(\widehat{\beta}_1)$	$\widehat{SE}_3(\widehat{\beta}_1)$	$\widehat{SE}_1(\widehat{\sigma}_1^2)$	$\widehat{SE}_2(\widehat{\sigma}_1^2)$	$\widehat{SE}_3(\widehat{\sigma}_1^2)$
0.00	100	0.007	0.007	0.007	0.000	0.001	-0.000	-0.001	0.001	-0.001
	500	0.002	0.002	0.002	-0.000	0.000	-0.000	0.000	0.000	0.000
	1000	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
0.50	100	0.004	0.004	0.004	-0.006	-0.008	-0.002	-0.007	-0.009	-0.004
	500	0.002	0.002	0.002	-0.003	-0.004	-0.001	-0.002	-0.004	-0.001
	1000	-0.001	-0.001	-0.001	-0.002	-0.002	-0.000	-0.001	-0.002	-0.000
0.99	100	-0.003	-0.003	-0.002	-0.026	-0.035	-0.010	-0.013	-0.016	-0.008
	500	-0.005	-0.005	-0.004	-0.011	-0.015	-0.003	-0.005	-0.007	-0.002
	1000	-0.003	-0.003	-0.002	-0.007	-0.010	-0.001	-0.003	-0.005	-0.002
ρ	n				$\widehat{SE}_1(\widehat{\beta}_2)$	$\widehat{SE}_2(\widehat{\beta}_2)$	$\widehat{SE}_3(\widehat{\beta}_2)$	$\widehat{SE}_1(\widehat{\sigma}_2^2)$	$\widehat{SE}_2(\widehat{\sigma}_2^2)$	$\widehat{SE}_3(\widehat{\sigma}_2^2)$
0.00	100				0.001	0.003	0.001	-0.008	-0.004	-0.010
	500				0.001	0.001	0.001	0.001	0.001	0.000
	1000				0.000	0.000	0.000	0.001	0.001	0.001
0.50	100				-0.046	-0.070	-0.005	-0.019	-0.028	-0.004
	500				-0.021	-0.032	-0.002	-0.010	-0.016	-0.001
	1000				-0.014	-0.021	-0.000	-0.009	-0.013	-0.002
0.99	100				-0.092	-0.122	-0.017	-0.085	-0.111	-0.022
	500				-0.040	-0.054	-0.006	-0.035	-0.049	-0.003
	1000				-0.027	-0.037	-0.002	-0.026	-0.036	-0.004

Table 2: Bias estimates of the three standard error estimators (SE_1, SE_2, SE_3) when the covariance structure of a 2-component Gaussian mixture is assumed to be conditionally independent based on 1000 replications under each mixture distribution with $m = 5$, $\gamma_1 = \gamma_2 = 0$, $\beta_1 = 1$, $V_1 = I_m$, $\sigma_1^2 = 0.25$, $\beta_2 = 3$, $V_2 = V(\rho)$ and $\sigma_2^2 = 1$ where $V(\rho)$ is the exchangeable correlation matrix. Approximate standard error is based on the estimated standard deviation of the simulation distribution. Values equal to zero represent values less than 0.001.

		Conditional Independence			Exchangeable			Exponential		
		Estimate	$\widehat{SE_1}$	$\widehat{SE_3}$	Estimate	$\widehat{SE_1}$	$\widehat{SE_3}$	Estimate	$\widehat{SE_1}$	$\widehat{SE_3}$
Component 1	γ_{10}	0.827	0.201	0.461	0.122	0.182	0.261	0.420	0.235	0.442
	γ_{11}	-1.800	0.380	0.570	0.966	0.300	0.430	-0.888	0.734	2.275
	β_{10}	19.315	0.140	0.178	20.994	0.208	0.225	19.984	0.358	0.467
	β_{11}	0.104	0.020	0.027	0.155	0.010	0.017	0.217	0.049	0.056
	β_{12}	0.001	0.001	0.001	-0.001	0.000	0.001	-0.002	0.001	0.001
	σ_1^2	2.969	0.110	0.277	6.331	0.721	1.002	5.718	0.427	0.976
	ρ_1	-	-	-	0.855	0.017	0.023	-	-	-
	r_1	-	-	-	-	-	-	2.973	0.207	0.280
Component 2	γ_{20}	0.599	0.206	0.435	-0.450	0.216	0.275	0.716	0.221	0.523
	γ_{21}	0.435	0.280	0.385	0.851	0.344	0.464	1.424	0.511	1.687
	β_{20}	20.979	0.120	0.312	24.041	0.520	0.689	20.910	0.241	0.389
	β_{21}	0.238	0.016	0.023	0.198	0.032	0.070	0.187	0.024	0.029
	β_{22}	-0.002	0.000	0.001	-0.005	0.001	0.002	-0.001	0.001	0.001
	σ_2^2	3.150	0.104	0.365	21.113	2.926	4.716	8.147	0.788	2.458
	ρ_2	-	-	-	0.805	0.027	0.022	-	-	-
	r_2	-	-	-	-	-	-	23.579	2.016	4.213
Component 3	γ_{30}	0.279	0.221	0.482	0.129	0.186	0.282	0.174	0.247	0.526
	γ_{31}	0.534	0.296	0.402	1.066	0.304	0.424	1.739	0.514	1.575
	β_{30}	22.818	0.207	0.721	21.165	0.275	0.372	23.516	0.385	0.563
	β_{31}	0.449	0.027	0.051	0.366	0.016	0.031	0.464	0.049	0.085
	β_{32}	-0.006	0.001	0.002	-0.002	0.000	0.001	-0.006	0.001	0.002
	σ_3^2	6.416	0.260	1.162	11.241	1.225	2.230	13.049	1.004	2.851
	ρ_3	-	-	-	0.822	0.018	0.019	-	-	-
	r_3	-	-	-	-	-	-	10.215	0.675	0.826
Component 4	γ_{41}	0	-	-	0	-	-	0	-	-
	γ_{40}	0	-	-	0	-	-	0	-	-
	β_{40}	25.713	0.480	0.672	22.473	0.573	0.480	24.356	1.054	0.833
	β_{41}	0.679	0.069	0.108	0.725	0.042	0.080	0.636	0.138	0.099
	β_{42}	-0.008	0.002	0.003	-0.008	0.001	0.003	-0.006	0.004	0.004
	σ_4^2	26.864	1.154	7.891	32.982	3.484	6.928	44.788	4.093	13.757
	ρ_4	-	-	-	0.681	0.034	0.044	-	-	-
	r_4	-	-	-	-	-	-	8.058	0.654	0.850

Table 3: Parameter and standard error estimates $(\widehat{SE_1}, \widehat{SE_3})$ for a random sample of 500 from NLSY79 assuming a four component mixture model with quadratic mean and the following correlation structures: conditional independence, exchangeable, and exponential correlation. Values equal to zero represent values less than 0.001. Additionally, the RJ criteria was calculated each covariance assumption: $RJ = 7.34, 3.02, 2.22$ under conditional independence, exchangeable, and exponential, respectively.