Conditional median-based Bayesian growth mixture modeling for nonnormal data

Seohyun Kim[1] · Xin Tong[1] [ID] · Jianhui Zhou[2] · Jeffrey P. Boichuk[3]

Abstract

Growth mixture modeling is a common tool for longitudinal data analysis. One of the key assumptions of traditional growth mixture modeling is that repeated measures within each class are normally distributed. When this normality assumption is violated, traditional growth mixture modeling may provide misleading model estimation results and suffer from nonconvergence. In this article, we propose a robust approach to growth mixture modeling based on conditional medians and use Bayesian methods for model estimation and inferences. A simulation study is conducted to evaluate the performance of this approach. It is found that the new approach has a higher convergence rate and less biased parameter estimation than the traditional growth mixture modeling approach when data are skewed or have outliers. An empirical data analysis is also provided to illustrate how the proposed method can be applied in practice.

Keywords Robust methods · Growth mixture modeling · Conditional medians · Asymmetric Laplace distribution

Modeling longitudinal data is an active area of research, especially in the social and behavioral sciences. Longitudinal data consist of repeated measurements from the same individuals that are obtained at different occasions. The main goals of a longitudinal study are to characterize change in responses over time and investigate factors that are associated with the change (Fitzmaurice, Laird, & Ware, 2012; Singer & Willett, 2003). Growth curve modeling (Bollen & Curran, 2006; McArdle, 2009; Meredith & Tisak, 1990) is often used for longitudinal studies because it can directly investigate within-subjects changes over time and between-subjects differences in change. The main objectives of this modeling approach are to describe an overall trajectory that characterizes repeated responses and measure

between-subjects differences around the overall growth trajectory. Growth curve models can be estimated within both a multilevel modeling framework and a structural equation modeling framework and can be easily extended to include more complex structural relationships (Bollen & Curran, 2006; McArdle & Epstein, 1987; Meredith & Tisak, 1990).

Growth curve modeling assumes that all observed growth trajectories are sampled from a population that is characterized by a single overall growth curve (Connell & Frye, 2006; Muthén, 2004). A population, however, may consist of several underlying groups that can be characterized by different growth trajectories. In order to accommodate heterogeneity in a population, researchers have used finite mixture modeling (McLachlan & Peel, 2000) as a technique for finding qualitatively meaningful latent groups. Growth mixture modeling (Muthén, 2004; Muthén & Shedden, 1999), which is an extension of growth curve modeling, assumes that a population consists of multiple latent classes and that each class has a class-specific growth trajectory. Growth mixture models have been used in many research studies over the past few decades. For example, Huang et al. (2011) used a growth mixture model to identify distinct employment patterns over 20 years and investigated the impact of drug use on employment. Guay et al. (2021) used a growth mixture model to identify latent groups of students that follow

✉ Xin Tong
  xt8b@virginia.edu

1 Department of Psychology, University of Virginia, 102 Gilmer Hall, Charlottesville, VA 22903, USA

2 Department of Statistics, University of Virginia, 113 Halsey Hall, Charlottesville, VA 22903, USA

3 McIntire School of Commerce, University of Virginia, 125 Ruppel Dr., Charlottesville, VA 22903, USA

different trajectories of self-motivation over the course of their time in secondary school and tried to characterize the latent groups using different sources of relatedness.

Although growth mixture modeling has been increasingly used in social and behavioral sciences over the past few decades and is flexible in modeling growth trajectories, there are several issues that researchers need to be aware of (e.g., Bauer, 2007; Bauer and Curran, 2003; Hipp & Bauer, 2006). We focus on two major issues in this article: the violation of normality assumptions and nonconvergence issues.

Traditional growth mixture models (GMMs) assume that repeated measures are normally distributed within each class. If data violate the normality assumption, a traditional GMM may result in invalid statistical inferences, such as biased estimation or the over-extraction of latent classes (Bauer & Curran, 2003; Depaoli, Winter, Lai, & Guerra-Peña, 2019; Muthén & Asparouhov, 2015). Numerous robust methods have been proposed to relax the normality assumption in growth mixture modeling. Lu and Zhang (2014) assumed that either measurement errors or latent growth factors (i.e., random effects) of GMM follow a t-distribution to reduce any effects of outliers. Muthén and Asparouhov (2015) imposed a skewed t-distribution on latent growth factors to reduce the risk that GMM detects spurious latent classes due to data with skewed distributions. In addition, Depaoli et al. (2019) and Son et al. (2019) conducted systematic simulation studies to investigate the performance of GMM with latent growth factors following normal, skewed-normal, and skewed t-distributions, and concluded that mis-specifying the distributions in GMM, especially fitting a traditional GMM to nonnormal data can result in biased parameter estimation.

Recently, a quantile regression approach has been actively used as a robust approach to longitudinal data analysis. A quantile, which can be easily found in introductory statistics books, is a value that corresponds to a specified proportion of a sample (Gilchrist, 2000). Conventional regression focuses on the mean of a dependent variable y conditional on the values of given independent variables x. Quantile regression extends this approach by allowing researchers to investigate a conditional distribution of y given x at various locations of y (i.e., quantiles) so that researchers can globally understand the relationship between them (Davino, Furno, & Vistocco, 2014; Koenker, 2005). As a special case of quantile regression, median regression is a method that models the conditional median of a dependent variable as a function of independent variables. The median, which is a quantile at level 0.5, is a natural measure of centrality that can be used as an alternative to the mean and is robust against outliers or skewed distributions. For nonnormally distributed data, median regression can be

used as an alternative method to traditional mean-based regression (Bassett & Koenker, 1978; Koenker & Bassett, 1982). In longitudinal research, He et al. (2003) used median regression for longitudinal data and compared three different estimators for median regression. Tong et al. (2020) proposed a robust Bayesian growth curve model based on conditional medians and showed that the proposed method outperforms traditional growth curve modeling when outlying observations exist. The conditional median-based method has a high breakdown point against the existence of outlying observations. Despite the robustness and increasing popularity of median-based methods, they have not yet been systematically implemented and evaluated in growth mixture modeling.

The second issue we tackle is nonconvergence, which can arise when model is misspecified or data are irregular (Tueller & Lubke, 2010). Research has shown that several factors, including model complexity, sample size, and latent class separation can all prevent GMMs from converging (e.g., Nylund, Asparouhov, & Muthén, 2007; Depaoli, 2013; Depaoli et al., 2019; Tueller & Lubke, 2010; Muthén & Asparouhov, 2015). In a Bayesian framework, employing informative priors can help if sample size and latent class separation are at fault (e.g., Depaoli, 2013). However, even with a large sample size and well-separated latent classes, nonconvergence can still occur if the number of latent classes is over-specified (e.g., Liu & Hancock, 2014; Nylund et al., 2007) or the distributional form of a GMM is ill-specified (e.g., Depaoli, Winter, Lai, & Guerra-Peña, 2019; Guerra-Peña, García-batista, Depaoli, & Garrido, 2020).

To address the nonnormality and nonconvergence issues in growth mixture modeling, we propose a robust approach using conditional medians. Bayesian methods are used for model estimation as a Bayesian approach enables incorporating prior information into model estimation and estimating complex models with advanced sampling algorithms. In this approach, a likelihood needs to be specified for model inferences. For growth mixture modeling using conditional medians, we skillfully used an asymmetric Laplace distribution for measurement errors since it provides a natural way to use Markov chain Monte Carlo methods to estimate a transformed model (Reich, Bondell, & Wang, 2010; Yu & Moyeed, 2001). We conducted a simulation study to systematically investigate the performance of the proposed conditional median-based Bayesian growth mixture modeling approach under various data conditions. Model convergence based on the proposed method is also evaluated.

This paper is structured as follows. We briefly review growth curve modeling, followed by an introduction to traditional growth mixture modeling based on conditional means. We then propose the robust growth mixture model based on conditional medians. Next, we present a simulation

study to evaluate the performance of our proposed robust growth mixture modeling. In the subsequent section, we apply the proposed method using empirical data. Finally, we conclude with discussion and recommendations.

Growth mixture modeling based on conditional medians

A brief review of growth curve models

Growth curve models characterize changes in responses over time and estimate interindividual variability in those changes. Suppose that individual i ($i = 1, \ldots, N$) is assessed at J time points. A typical growth curve model for characterizing these repeated responses can be denoted as

$$\mathbf{y}_i = \mathbf{\Lambda}\mathbf{b}_i + \boldsymbol{\epsilon}_i,$$
$$\mathbf{b}_i = \boldsymbol{\beta} + \mathbf{u}_i,$$

where $\mathbf{y}_i = (y_{i1}, \ldots, y_{iJ})^T$ is a $J \times 1$ vector of repeated measures for individual i, $\mathbf{\Lambda}$ is a $J \times q$ matrix of factor loadings, in which q is the number of latent growth factors, $\boldsymbol{\epsilon}_i$ is a $J \times 1$ vector of measurement errors, $\mathbf{b}_i$ is a $q \times 1$ vector of latent growth factors, $\boldsymbol{\beta}$ is a $q \times 1$ vector of latent growth factor means, and $\mathbf{u}_i$ is a $q \times 1$ vector of random effects that are independent of the measurement errors. The elements of $\mathbf{\Lambda}$ determine the shape of trajectories such as linear or quadratic forms. For example,

$$\mathbf{\Lambda}^T = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ 0 & 1 & \ldots & J-1 \end{pmatrix} \tag{1}$$

is for a linear growth curve. Traditional growth curve models assume that $\boldsymbol{\epsilon}_i$ and $\mathbf{u}_i$ follow a multivariate normal distribution such as

$$\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \mathbf{\Omega}), \quad \mathbf{u}_i \sim N(\mathbf{0}, \mathbf{\Psi}),$$

where $\mathbf{\Omega}$ is a $J \times J$ covariance matrix of measurement errors, and $\mathbf{\Psi}$ is a $q \times q$ covariance matrix of latent growth factors. Growth curve models generally assume $\mathbf{\Omega}$ to be $\sigma^2 \mathbf{I}$, which indicates that measurement errors are independent across J time points, given the latent growth factors, with variances of σ^2. Combining all together, $\mathbf{y}_i$ has the following probability density function

$$f(\mathbf{y}_i|\Theta) = \Phi(\mathbf{\Lambda}\boldsymbol{\beta}, \mathbf{\Lambda}\mathbf{\Psi}\mathbf{\Lambda}^T + \mathbf{\Omega}),$$

where Θ is the set of all parameters, and Φ is the J-dimensional multivariate normal probability density function with mean $\mathbf{\Lambda}\boldsymbol{\beta}$ and covariance matrix $\mathbf{\Lambda}\mathbf{\Psi}\mathbf{\Lambda}^T + \mathbf{\Omega}$.

A brief review of growth mixture models

Growth mixture models (GMMs) assume that a population consists of a number of latent groups, and each group is characterized by a unique growth trajectory. A growth curve model can be seen as a special case of GMM with one latent class.

The probability density function of $\mathbf{y}_i$ in a GMM is expressed as a finite mixture of G probability density functions,

$$f(\mathbf{y}_i|\Theta) = \sum_{g=1}^{G} \pi_g f_g(\mathbf{y}_i|\Theta_g),$$

where π_g is a mixing proportion for latent class g, which indicates a probability that $\mathbf{y}_i$ is drawn from latent class g with probability density function f_g, and Θ_g is a set of class-specific parameters for f_g ($\Theta_g \subset \Theta$). In a GMM, each latent class describes a distinct growth trajectory such as

$$\mathbf{y}_i = \mathbf{\Lambda}_g \mathbf{b}_{ig} + \boldsymbol{\epsilon}_i,$$
$$\mathbf{b}_{ig} = \boldsymbol{\beta}_g + \mathbf{u}_i, \tag{2}$$

where the subscript g indicates that a corresponding parameter or variable is class-specific. Traditional GMMs assume that the measurement error and random effect variables follow a multivariate normal distribution such as $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \mathbf{\Omega}_g)$ and $\mathbf{u}_i \sim N(\mathbf{0}, \mathbf{\Psi}_g)$.

Growth mixture modeling using conditional medians

In this section, we build a growth mixture model based on conditional medians that inherits robust properties from the median. What immediately follows is an introduction to median regression, which we use as a building block.

In median regression, conditional medians are modeled, $Q_{0.5}(y_i|\mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}_{0.5}$, where $Q_{0.5}(y_i|\mathbf{x}_i)$ represents a conditional median of y_i given $\mathbf{x}_i$. It can be rewritten as

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta}_{0.5} + \epsilon_i, \quad Q_{0.5}(\epsilon_i|\mathbf{x}_i) = 0,$$

where ϵ_i is a random error with median 0. The unknown parameter $\boldsymbol{\beta}_{0.5}$ can be estimated by minimizing the following sum of absolute residuals (Koenker & Bassett, 1978):

$$\hat{\boldsymbol{\beta}}_{0.5} = \arg\min_{\beta} \sum_{i=1}^{N} |y_i - \mathbf{x}_i^T \boldsymbol{\beta}_{0.5}|. \tag{3}$$

An asymmetric Laplace distribution (ALD) provides a natural way to minimize the above sum of absolute residuals and link to the maximum likelihood principle by assuming y_i follows an asymmetric Laplace distribution (Koenker & Machado, 1999; Geraci & Bottai, 2007; Yu & Moyeed, 2001). If a variable ω follows an asymmetric Laplace distribution, $\omega \sim ALD(m, \delta, \tau)$, its probability density function is given as

$$f(\omega|m, \delta, \tau) = \frac{\tau(1-\tau)}{\delta} \exp\{-\frac{1}{\delta}\rho_\tau(\omega - m)\}, \tag{4}$$

where m is a location parameter equal to the τth quantile of ω, δ is a scale parameter, τ is an asymmetry parameter, $\rho_\tau(\omega) = \omega(\tau - I(\omega < 0))$, and $I(\cdot)$ is an indicator function. In the case of the median, τ is equal to 0.5, $\rho_\tau(\omega)$ is equal to $\rho_\tau(\omega) = \frac{|\omega|}{2}$, and Eq. 4 becomes a Laplace distribution, which is also referred to as a double exponential distribution (Kotz, Kozubowski, & Podgórski, 2001). A Laplace distribution has a sharp peak around the mean and has thicker tails than a normal distribution. If we assume $y_i \sim ALD(m_i, \delta, \tau)$ $(i = 1, \ldots, N)$ with $m_i = \boldsymbol{x}_i^T \boldsymbol{\beta}_{0.5}$ and $\tau = 0.5$, the likelihood for N independent observations is

$$L(\boldsymbol{\beta}_{0.5}, \delta | \boldsymbol{y}) \propto \delta^{-N} \exp \left\{ -\sum_{i=1}^{N} \frac{|y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}_{0.5}|}{2\delta} \right\}. \quad (5)$$

Minimizing Eq. 3 is equivalent to maximizing the likelihood in Eq. 5 (Geraci & Bottai, 2007; Yu & Moyeed, 2001).

A growth mixture model based on conditional medians can be constructed in a similar way. Given the traditional growth mixture model in Eq. 2, and assuming the jth subject belongs to latent class g, the growth mixture model based on conditional medians can be written as

$$y_{ij} = \boldsymbol{\lambda}_{jg}^T \mathbf{b}_{ig(0.5)} + \epsilon_{ij}, \quad with \quad Q_{0.5}(\epsilon_{ij} | \mathbf{b}_{ig(0.5)}) = 0 \\ \mathbf{b}_{ig(0.5)} = \boldsymbol{\beta}_{g(0.5)} + \mathbf{u}_i \quad (6)$$

where $\boldsymbol{\lambda}_{jg}$ is the jth row of $\boldsymbol{\Lambda}_g$, $\mathbf{b}_{ig(0.5)}$ is a vector of latent growth factors for latent class g based on conditional medians, $\boldsymbol{\beta}_{g(0.5)}$ is a vector of latent growth factor means for latent class g based on conditional medians, and the random effect $\mathbf{u}_i$ is assumed to be independent from ϵ_{ij}, follow a multivariate distribution with a mean of 0, and have a class-specific covariance matrix of $\boldsymbol{\Psi}_g$. In this paper, we assume that $\mathbf{u}_i$ follows a multivariate normal distribution for our simulation study and empirical data analysis. Equation 6 implies that $Q_{0.5}(y_{ij} | \mathbf{b}_{ig(0.5)}) = \boldsymbol{\lambda}_{jg}^T \mathbf{b}_{ig(0.5)}$ for latent class g. That is, a conditional median of y_{ij} given $\mathbf{b}_{ig(0.5)}$ is $\boldsymbol{\lambda}_{jg}^T \mathbf{b}_{ig(0.5)}$ for latent class g. Since an asymmetric Laplace distribution provides a natural way to deal with parameter estimation of median regression, this study also imposes that $\epsilon_{ij} \sim ALD(0, \delta, 0.5)$. Thus, $y_{ij} | \mathbf{b}_{ig(0.5)} \sim ALD(m_{ij}, \delta, 0.5)$, where $m_{ij} = \boldsymbol{\lambda}_{jg}^T \mathbf{b}_{ig(0.5)}$ and δ is a scale parameter for ϵ_{ij}.

In the case of a linear growth mixture model, Equation 6 can be built using the following vector presentations and $\boldsymbol{\Lambda}$ specified in Eq. 1:

$$\boldsymbol{b}_{ig(0.5)} = (b_{Ig(0.5)}, b_{Sg(0.5)})^T, \boldsymbol{\beta}_{g(0.5)} = (\beta_{Ig(0.5)}, \beta_{Sg(0.5)})^T,$$

$$\boldsymbol{\Psi}_{g(0.5)} = \begin{pmatrix} \sigma_{Ig(0.5)}^2 & \sigma_{ISg(0.5)} \\ \sigma_{ISg(0.5)} & \sigma_{Sg(0.5)}^2 \end{pmatrix},$$

where $b_{Ig(0.5)}$ and $b_{Sg(0.5)}$ are a random intercept and random slope, respectively, that vary across individuals in latent class g, and $\beta_{Ig(0.5)}$ and $\beta_{Sg(0.5)}$ are the average of the intercepts and slopes, respectively, for latent class g. In this model, the intercept is the initial level of conditional medians, and the slope is the rate of change in conditional medians over time. The $\sigma_{Ig(0.5)}^2$ and $\sigma_{Sg(0.5)}^2$ in $\boldsymbol{\Psi}_{g(0.5)}$ represent the variability in $b_{Ig(0.5)}$ and $b_{Sg(0.5)}$, respectively, and $\sigma_{ISg(0.5)}$ represents the covariance between $b_{Ig(0.5)}$ and $b_{Sg(0.5)}$.

In this study, we use a Bayesian approach for our model estimation and inference, as Bayesian methods have many advantages (Lee, 2007). For example, prior knowledge on parameters can be incorporated in model estimation, which helps to get accurate parameter estimates. In addition, posterior distributions of parameters can be relatively easily obtained even for complex models by using advanced statistical computing algorithms. In the Bayesian framework, model estimation and inferences are conducted based on a joint posterior distribution of model parameters. In the proposed growth mixture model, direct inferences from a joint posterior distribution are difficult due to the complexity of the model. Thus, we used the data augmentation technique (Tanner & Wong, 1987) and Markov chain Monte Carlo (MCMC) techniques such as Gibbs sampling to simulate samples from the posterior distribution. The observed dataset $\{y_i | i = 1, \ldots, N\}$ is augmented by latent growth factors ($\mathbf{b}_i$), and latent variables for the asymmetric Laplace distribution. Specifically, an asymmetric Laplace distribution, $y \sim ALD(m, \delta, \tau)$, can be specified as the following mixture presentation using $v \sim \exp(\delta)$ with mean δ and $W \sim N(0, 1)$ (Kotz, Kozubowski, & Podgórski, 2001, Kozumi & Kobayashi, 2011)

$$y = m + \zeta v + \xi \sqrt{\delta v} W,$$

where

$$\zeta = \frac{1 - 2\tau}{\tau(1 - \tau)}, \quad \xi^2 = \frac{2}{\tau(1 - \tau)},$$

and $\tau = 0.5$ for the median.

With the aid of the mixture presentation of asymmetric Laplace distribution and augmented data, the likelihood function for observed data is

$$L(\Theta | \mathbf{y}, \mathbf{b}, \boldsymbol{v}) = \prod_{i=1}^{N} \left[\sum_{g=1}^{G} \pi_g \left\{ \prod_{j=1}^{J} f(y_{ij} | \mathbf{b}_{ig(0.5)}, v_{ij}) \right\} \right]$$

$$\propto \prod_{i=1}^{N} \left[\sum_{g=1}^{G} \pi_g \left\{ \prod_{j=1}^{J} \left(\delta_g v_{ij} \right)^{-\frac{1}{2}} \right. \right.$$

$$\left. \left. \exp \left(-\frac{(y_{ij} - \boldsymbol{\lambda}_{jg}^T \mathbf{b}_{ig(0.5)} - \zeta v_{ij})}{2\xi^2 \delta_g v_{ij}} \right) \right\} \right].$$

If conjugate priors are used, Gibbs sampling can be easily constructed for model inferences as we describe in the Appendix.

Model estimations in the following simulation study as well as the empirical example are performed using rjags (Plummer, 2017), which is an R (R Core Team, 2019) package that calls the software "Just Another Gibbs Sampler" or JAGS (Plummer, 2003) in R. JAGS is a program for Bayesian modeling and inferences. Sample JAGS scripts used in the current study are provided in the supplemental materials at https://github.com/CynthiaXinTong/MedianGMM.git.

A simulation study

In this section, we evaluate the performance of the proposed median growth mixture modeling approach and compare that to the performance of traditional growth mixture modeling under various conditions by manipulating sample size, latent class separation, mixing proportion (i.e., latent class relative size), and population distribution. In this simulation study, we evaluate convergence rate and parameter estimates when the number of latent classes is the same as its population model. Note that class enumeration is a complicated issue and has been thoroughly investigated in Kim et al. (2021).

Simulation design

For this simulation study, we generated data using a linear growth mixture model with four measurement occasions and two latent classes. For simplicity, we only allowed latent growth factor means to vary across latent classes, and we kept the variances for the measurement errors at the four measurement occasions constant.[1]

Three different sample sizes ($N = 300$, 500, and 1000), and two levels of class separation were considered. The class separation was measured by the Mahalanobis distance, which is given as $\Delta = \sqrt{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}$, where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are means for the two latent classes and $\boldsymbol{\Sigma}$ is the shared covariance matrix of latent growth factors. One of the two levels was set to be relatively low at 1.7 (condition MD1), which can be seen as medium separation, and the other one was set to be relatively high at 3.6 (condition MD2), which can be seen as high separation (Depaoli, Winter, Lai, & Guerra-Peña, 2019, Lubke & Neale, 2006; Tueller & Lubke, 2010). These two levels of latent class separations were set to compare the performance of the proposed approach and the traditional approach to growth mixture modeling without confounding effects due to poor

---

[1]It is common to assume that variance components are invariant over latent classes in growth mixture modeling (e.g., Bauer and Curran, 2003; Depaoli, 2013; Hipp & Bauer, 2006, Son, Lee, Jang, Yang, & Hong, 2019).

class separation. Latent classes were generated with two different mixing proportions: unbalanced proportions (0.3, 0.7) and balanced proportions (0.5, 0.5). Values for the population parameters were set to be similar to the ones obtained from the empirical data analysis described in the next section. Growth factor means for Class 1 were $(\beta_{I1}, \beta_{S1})^T = (18, 0.8)^T$ for both MD1 and MD2, and the means for Class 2 were $(\beta_{I2}, \beta_{S2})^T = (15, 0.3)^T$ for MD1 and $(\beta_{I2}, \beta_{S2})^T = (10, 0.3)^T$ for MD2. Additionally, $\boldsymbol{\Psi} = \begin{pmatrix} 6 & -0.27 \\ -0.27 & 0.3 \end{pmatrix}$ and $\sigma^2 = 4$ for both MD1 and MD2. As shown in the empirical example in the next section, these population parameter values are realistic in practice.

Four different types of population distributions within each class were considered by manipulating the distribution of the measurement errors (conditions D1 to D4). Data of the D1 condition followed the normality assumption of traditional growth mixture modeling with $\epsilon_{ij} \sim N(0, \sigma^2)$. The D2 and D3 conditions were for evaluating the model performance when samples have outliers. Measurement errors for the D2 and D3 conditions were first generated using $N(0, \sigma^2)$, and 5% and 10% of subjects were randomly selected, respectively, to have outliers for an arbitrarily selected one out of the four measurement occasions. The outliers were generated to be extremely higher than the majority of observations by using $\epsilon_{ij} \sim N(M\sigma, \sigma^2)$, where M was randomly selected from {5, 8, 10} with probabilities 0.2, 0.5, and 0.3, respectively. Lastly, the D4 condition was included to evaluate performance when samples have a skewed distribution. Measurement errors were first generated from $LogNormal(0, 1)$, and then the generated values were rescaled to have a mean of 0 and a variance of σ^2. In total, there were 48 data generation conditions in this simulation study (2 mixing proportions × 2 class separations × 3 sample sizes × 4 error distributions). For each condition, 500 datasets were generated.

Estimation

A linear median-based growth mixture model and a traditional mean-based growth mixture model were fit to each dataset for the purpose of assessing and comparing the performance. For the traditional growth mixture model, we used weakly informative priors (e.g., Lu, Zhang, & Lubke, 2011; Zhang, 2016) as follows: $\beta_{Ig} \sim N(0, 10^2)$, $\beta_{Sg} \sim N(0, 10)$, $\sigma^2 \sim InvGamma(0.1, 0.1)$, $\boldsymbol{\Psi}_g \sim InvWishart(\boldsymbol{I}_2, 3)$, where $\boldsymbol{I}_2$ is a 2×2 identity matrix, and $\boldsymbol{\pi} \sim Dirichlet(15, 25)$ for the unbalanced mixing proportion condition and $\boldsymbol{\pi} \sim Dirichlet(25, 25)$ for the balanced mixing proportion condition. For the median growth mixture model, the scale parameter δ had prior $\delta \sim InvGamma(0.1, 0.1)$, and the rest of the priors were the same as the priors for the traditional growth mixture

model. It is possible for growth mixture models to have several local solutions (Hipp & Bauer, 2006). The informative prior on π was used for our Bayesian estimation to avoid chains converging to local solutions (in real data analyses, multiple chains can be used to avoid getting a local solution) and recover proper latent classes.[2] Additionally, we put constraints on the mean intercept parameters such that $\beta_{I1} > \beta_{I2}$ to avoid within-chain label switching. The number of total iterations for each chain was initially set to 10,000 with the first half of the iterations used for the burn-in period. When a chain did not converge within 10,000 iterations, the total number of chains increased to 100,000. Convergence of chains was assessed using Geweke's convergence test (Geweke, 1991). The chains with Geweke's statistic values less than 2 were deemed converged, and results were summarized using converged chains.

Parameter estimates were evaluated using relative bias, and mean squared error (MSE). Relative bias was obtained using the following equation: $\frac{1}{R}\sum_{r=1}^{R}\left(\frac{\hat{\theta}_r - \theta_p}{\theta_p}\right)$, where R is the number of converged replications, θ_p is the population value, and $\hat{\theta}_r$ is the posterior mean for the parameter obtained from the rth replication. MSE was obtained using the following equation: $\frac{1}{R}\sum_{r=1}^{R}(\hat{\theta}_r - \theta_p)^2$. We additionally reported coverage rates of 95% credible intervals to investigate how well the posteriors reproduced the sample distribution.

Results

We mainly present results from conditions with $N = 500$ in this subsection. Overall patterns for $N = 300$ and $N = 1000$ were similar to the patterns found in $N = 500$, and results for these conditions can be found in our supplementary document at https://github.com/CynthiaXinTong/MedianGMM.git.

Table 1 summarizes convergence rates for conditions with $N = 500$. In general, convergence rates were higher for higher class separation (i.e., MD2) than for lower class separation (i.e., MD1), and higher for balanced mixing proportion conditions than the unbalanced counterpart. The median growth mixture modeling approach (i.e., median GMM) was very stable in terms of convergence and had

---

[2] Using informative priors for mixing proportions is known to be helpful in recovering proper latent classes in growth mixture modeling. For example, Depaoli (2013) found that using informative priors allowed proper latent class recovery in growth mixture modeling especially when mixing proportions were unbalanced. In Depaoli et al. (2017), they conducted a sensitivity analysis on priors and recommended incorporating prior knowledge on class size when possible. When prior information is not available for class size, they recommended using a noninformative prior. More discussion on priors for mixing proportions for growth mixture modeling can be found in Depaoli (2013) and Depaoli, Yang, and Felt (2017).

convergence rates over 99% across different latent class separations, mixing proportions, and distributional conditions. In contrast, the convergence rates for the traditional growth mixture modeling (i.e., mean GMM) varied greatly across conditions. The convergence rates were over 96% when there were balanced mixing proportions or when latent classes were highly separated. Convergence rates for the mean GMM notably decreased under the conditions with unbalanced mixing proportions, lower class separation, and nonnormally distributed measurement errors (i.e., D2-D4). When class separation is low (i.e., MD1) with unbalanced mixing proportions and 10% outlying observations (i.e., D3), the convergence rate can be as low as 46%.

Unbalanced mixing proportions

Figures 1 and 2 show parameter estimation results for $N = 500$ with unbalanced mixing proportions. Note that the y-axes have different scales in the plots. Figure 1 presents relative bias, MSE, and coverage rates for the intercepts and slopes from Class 1 (upper panels in Fig. 1) and Class 2 (lower panels in Fig. 1). Figure 2 presents relative bias, MSE, and coverage rates for the variances of the intercept and slope parameters in Ψ. When data were normally distributed (i.e., D1), both mean GMM and median GMM performed similarly except for the variance parameters. Both methods had low bias and MSE under the MD2 condition, and relatively higher bias and MSE under the MD1 condition. The median GMM tended to have larger bias and MSE for the variance parameters than the mean GMM. This difference was mostly due to the discrepancy between distributions behind the data generation (the measurement errors followed a normal distribution) and estimating model (the measurement errors followed a Laplace distribution for the median GMM). Performance for the other parameters was similar. This pattern can also be found in Geraci (2017)'s simulation study. Coverage rates for the intercepts and slopes were similar for both methods, ranging from 0.92 to 0.98. Coverage rates for the variance parameters were also similar, ranging from 0.92 to 0.99 except for the variance of the slope under MD2 and the covariance under MD1. The median GMM had a relatively low coverage rate for the variance of slope (about 0.85), and the mean GMM had a relatively low coverage rate for the covariance (about 0.81).

When data contained outliers (i.e., D2 and D3), the magnitudes of bias and MSE for the median GMM were considerably lower than those for the mean GMM, and coverage rates for the median GMM were higher than those for the mean GMM. For the mean GMM, bias and MSE increased as the proportion of outliers increased. β_I tended to be overestimated under MD1, which indicated

Table 1 Convergence rates for the median GMM and the mean GMM when *N*=500

| | Unbalanced mixing proportions | | | | Balanced mixing proportions | | | |
| | Median GMM | | Mean GMM | | Median GMM | | Mean GMM | |
	MD1	MD2	MD1	MD2	MD1	MD2	MD1	MD2
D1	0.99	1.00	1.00	1.00	0.99	1.00	0.99	1.00
D2	0.99	1.00	0.74	0.99	0.99	1.00	0.99	0.99
D3	0.99	1.00	0.46	0.98	0.99	1.00	0.96	0.99
D4	1.00	1.00	0.84	1.00	1.00	1.00	1.00	1.00

Median GMM: Growth mixture modeling based on conditional medians; mean GMM: Traditional growth mixture modeling based on conditional means. The total number of replications was 500

that the parameter estimate was influenced by outliers.[3] The bias and MSE of β_I estimate increased substantially as the proportion of outliers increased. The relative bias for β_S did not have a discernible pattern of how the bias was related to the proportion of outliers, although the MSE for β_S appeared to increase as the proportion of outliers increased. Under MD2, the bias and MSE for β_I and β_S were relatively small, but β_I still tended to be overestimated. The magnitudes of bias and MSE for the variance parameter estimates also increased as outliers were included. For the median GMM, although the bias and MSE for β_I and β_S slightly increased as the proportion of outliers increased, the magnitudes were much smaller than those for the mean GMM. The bias and MSE for the variances of the intercept and slope parameters increased as outliers were included, but the MSE values were smaller than those from the mean GMM. The coverage rates for the mean GMM dropped as the proportion of outliers increased. The coverage rates for the median GMM, on the contrary, did not change much for β_I and β_S. The coverage rates of the slope variance and covariance for the median GMM appeared to drop when the proportion of outliers was 0.10.

When data were skewed (i.e., D4), the median GMM had smaller bias and MSE and had higher coverage rates than those for the mean GMM except for β_I under MD2. In the case of MD2, the bias and MSE of β_I for the mean GMM were smaller than those for the median GMM. This is mainly related to the data generating procedure. When measurement errors were generated using the lognormal distribution, the values were rescaled to have mean 0 and a specified standard deviation. Since the distribution was positively skewed, the rescaled measurement errors had a mean of 0, but the median was negative. Therefore, it is natural for the median GMM to have underestimated β_I. Under MD1, the mean GMM overestimated β_I with high MSE values, especially for Class 1. The median GMM

performed better and had smaller bias and MSE for β_I and β_S than the mean GMM. The mean GMM yielded smaller bias for the variance parameter estimates, but the median GMM had smaller MSE. The median GMM had coverage rates higher than or similar to those for the mean GMM, except for β_I and the slope variance under MD2.

Table 2 presents bias and MSE of Class 1 mixing proportion (i.e., π_1) estimates[4] and presents average class membership recovery with its standard deviation. Mixing proportions for both methods were well recovered in general under MD2. In the case of MD1, mixing proportions for the mean GMM were not well recovered when the error distribution was not normally distributed. The median GMM had well recovered mixing proportions regardless of the error distribution. For the membership recovery, the MD2 condition had better recovery than the MD1 condition. Less than 80% of the subjects' class memberships were correctly estimated under MD1, and about 95% of the subjects' memberships were correctly estimated under MD2. While the membership recovery for the mean GMM appeared to be slightly influenced by the normality of the data, the median GMM had similar membership recovery across the four distribution conditions.

Balanced mixing proportions

When the mixing proportions were balanced, the general result patterns for conditions with $N = 500$ were similar to the ones found in the above section with the unbalanced mixing proportions: bias and MSE values were lower for MD2, the median GMM had smaller bias and MSE and had more stable coverage rates than the mean GMM especially when data contained outliers. The pattern of class membership recovery was also similar across the two mixing proportion conditions. The overall performance of the median GMM was consistent across the two different mixing proportion conditions. In contrast, the mean GMM tended to

---

[3]When the subscripts of latent class membership for β_I and β_S are omitted, β_I and β_S indicate parameters for both classes.

[4]Evaluating the performance of π_1 was enough because $\pi_2 = 1 - \pi_1$.
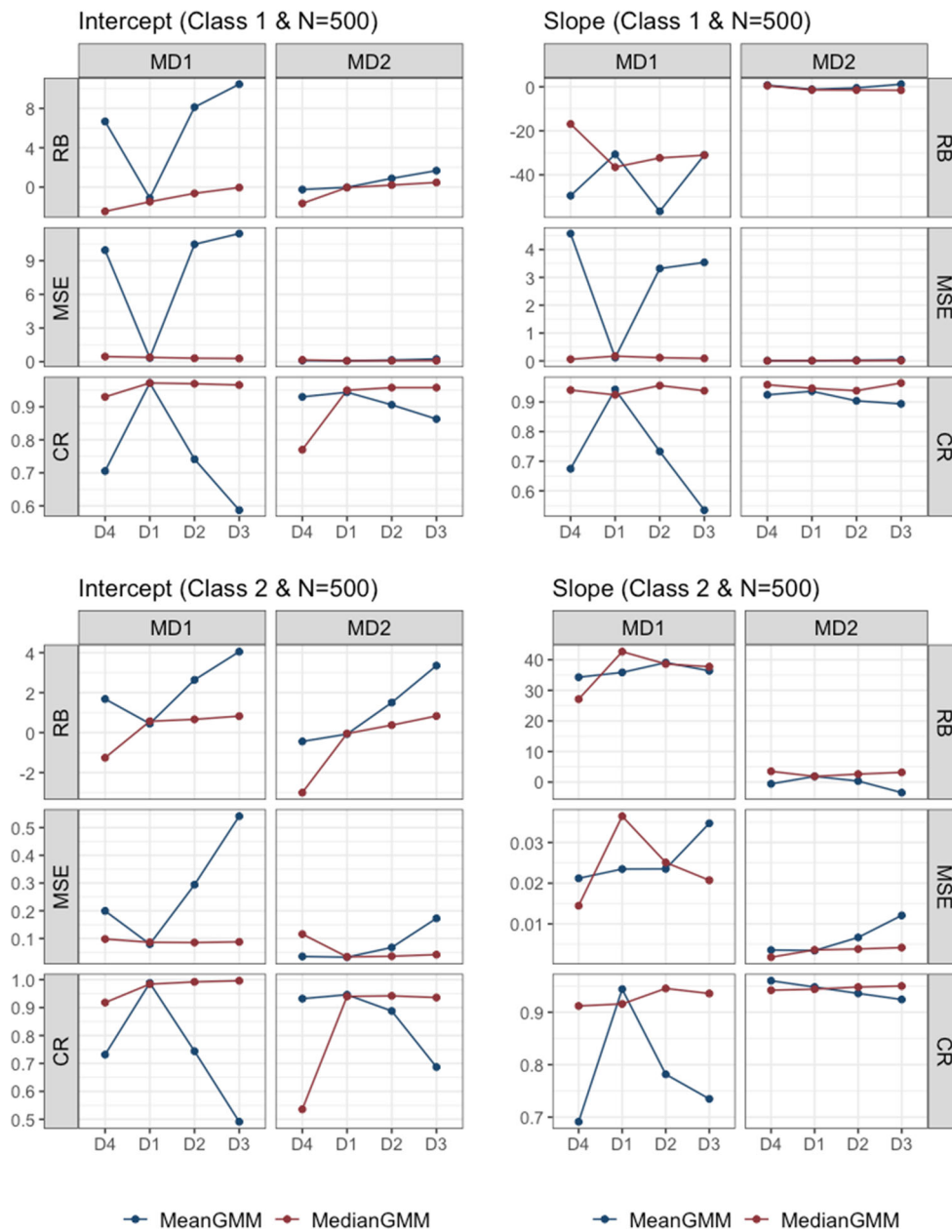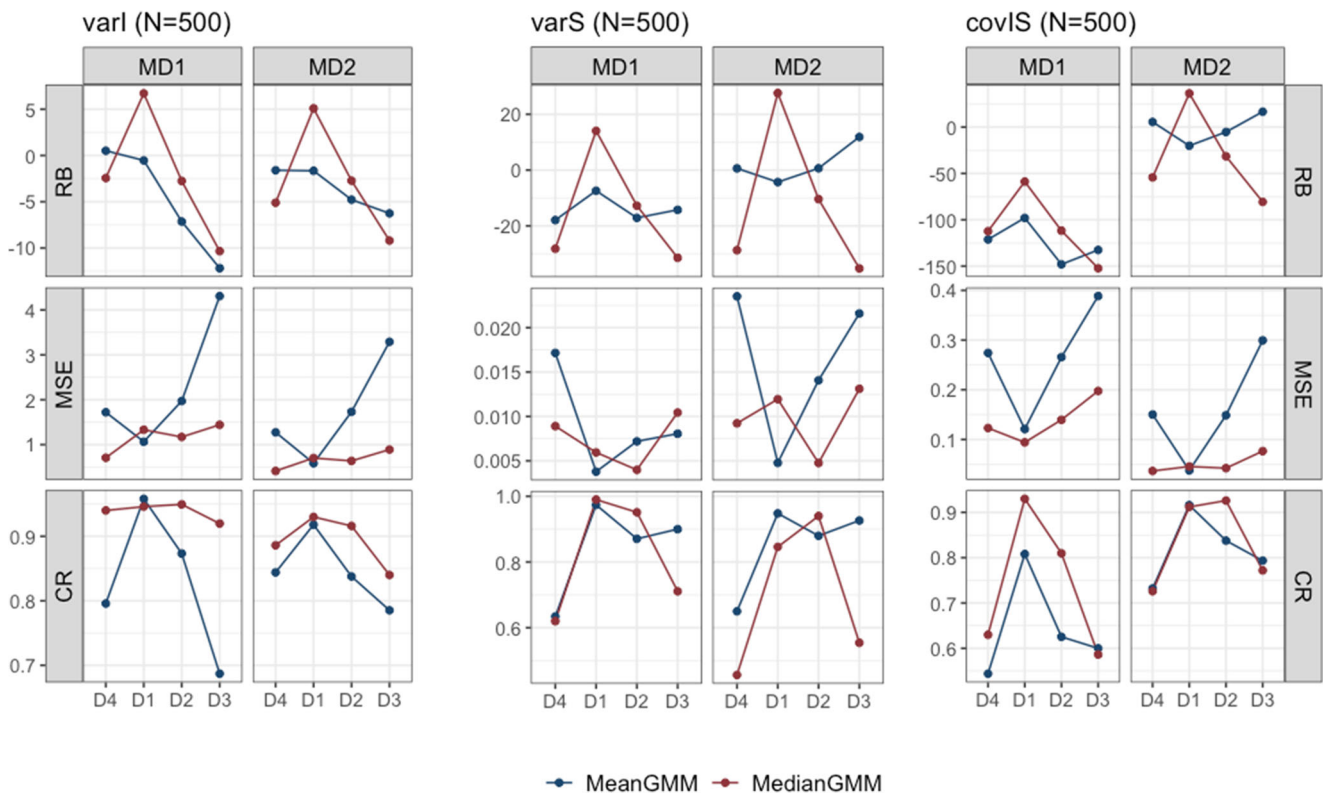
Fig. 1 Estimation results for the intercept and slope parameters when *N*=500 and mixing proportions were unbalanced. RB represents relative bias, and CR represents coverage rate

perform better with balanced mixing proportions, especially under MD1. When data contained outliers, the mean GMM with balanced mixing proportions had better mixing proportion recovery and had lower bias and MSE values, but its performance was still worse than the corresponding median GMM. The results for the balanced mixing proportions can be found in the supplemental materials at https://github. com/CynthiaXinTong/MedianGMM.git.

Conclusions for the simulation study

In this simulation study, the performance of the proposed conditional median-based growth mixture model and traditional mean-based growth mixture model were compared under various data conditions. Overall, both methods performed similarly when the data were normally distributed within each class. When the data contained outliers, how-

Fig. 2 Estimation results for the parameters in Ψ when N=500 and mixing proportions were unbalanced. RB represents relative bias, and CR represents coverage rate. varI shows results for intercept variance estimates, varS shows results for slope variance estimates, and covIS shows results for intercept-slope covariance estimates

ever, the median growth mixture model outperformed the mean growth mixture model in terms of model convergence, estimation bias, MSE, and coverage rate. When the data were skewed, the median growth mixture model performed better than the mean growth mixture model in general, but the median method underestimated the mean intercept parameters due to the data generating procedure.

Table 2 Mixing proportion and membership recovery when N=500 and unbalanced mixing proportions

	Mixing proportion				Membership			
	MD1		MD2		MD1		MD2	
	Median GMM	Mean GMM	Median GMM	Mean GMM	Median GMM	Mean GMM	Median GMM	Mean GMM
D1	0.01	0.00	0.00	0.00	0.77	0.78	0.96	0.96
	(0.02)	(0.02)	(0.02)	(0.02)	(0.05)	(0.03)	(0.01)	(0.01)
D2	0.00	−0.08	0.00	0.01	0.78	0.75	0.96	0.95
	(0.02)	(0.12)	(0.02)	(0.02)	(0.03)	(0.04)	(0.01)	(0.01)
D3	0.00	−0.11	0.00	0.01	0.78	0.74	0.95	0.94
	(0.02)	(0.14)	(0.02)	(0.02)	(0.03)	(0.04)	(0.01)	(0.01)
D4	-0.01	−0.10	0.00	0.00	0.79	0.76	0.96	0.96
	(0.02)	(0.13)	(0.02)	(0.02)	(0.03)	(0.05)	(0.01)	(0.01)

Numbers in the mixing proportion column show bias and MSE (in the parentheses) of π_1. Numbers in the membership column show average membership recovery and its standard deviation (in the parentheses)

Empirical data analysis

In this section, we apply the proposed and traditional methods to growth mixture modeling using a fast-casual restaurant's purchase data, which is currently available at https://github.com/CynthiaXinTong/MedianGMM.git. The restaurant has five locations in three cities on the East Coast of the United States. Its menu consists of a wide variety of organic options, all of which are made exclusively of plant-based ingredients. Because the restaurant uses Square for its point-of-sale system, we were able to use the squareupr R package to connect to Square's API and pull the data into R (Mortimer, 2018).

The data offer a record of how much customers spent, on average, when they visited the restaurant over the course of four 3-month periods (i.e., quarters). Figure 3 plots these four values for each of the 1241 customers in the data. The four measurement occasions in the data have skewness of 5.36, 2.42, 2.97, and 3.20, respectively, and additional descriptive statistics for the data can be found in our supplementary document at https://github.com/CynthiaXinTong/MedianGMM.git. As Fig. 3 displays, each measurement occasion has outliers. For example, the maximum values for Time 1 and Time 4 are more than ten standard deviations away from their means. We recognize that the population of customers is heterogeneous, but these extreme values are notably far away from a majority of the values, which implies that a median-based growth mixture model might be appropriate.

In a typical growth mixture analysis, researchers need to first determine the number of latent classes through model comparison by using model fit statistics (e.g., Nylund, Asparouhov, & Muthén, 2007, Peugh & Fan, 2012; 2015; Tofighi & Enders, 2008) or likelihood-based tests (e.g., Nylund, Asparouhov, & Muthén, 2007, Peugh & Fan, 2012; 2015), and by considering usefulness and interpretability
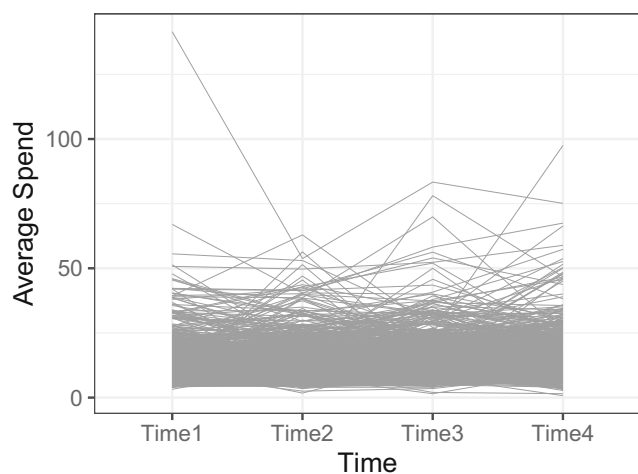
of latent classes (Connell & Frye, 2006; Dziak, Li, Tan, Shiffman, & Shiyko, 2015; Lu & Song, 2012). Since class enumeration is a complicated issue and not the focus of this study, we will not present a thorough model comparison procedure using model fit statistics (e.g., DIC, Spiegelhalter, Best, Carlin, & Van Der Linde, 2002; WAIC, Watanabe, 2010) in this empirical example especially because these model fit statistics have not been discussed in this article. Details about class enumeration for median-based GMM can be found in Kim et al. (2021). For the rest of this section, we will compare models with different latent classes based on model interpretations. Given that Fig. 3 suggests a linear pattern of growth trajectories, we fitted four linear growth mixture models based on conditional medians and varied the number of latent classes from one to four to identify an appropriate number of latent classes. For the purpose of comparison, we also fit four traditional growth mixture models with one to four latent classes. The total length of Markov chains was set to be 60,000 and the first 30,000 iterations were set to be the burn-in period. Parameter constraints were imposed on the mean intercept parameter to prevent within-chain label switching. Geweke's convergence test was used to verify the convergence of chains.

After inspecting the estimated mean trajectories and mixing proportions for the two methods, we found that the four-class solution was not empirically meaningful, because some of the mean trajectories were very close to one another. Additionally, the one-class solution did not capture heterogeneous patterns in growth trajectories. Therefore, in what follows, we present and compare the median GMM and the traditional mean GMM with two- and three-class solutions.

Table 3 presents each method's parameter estimates for the two-class solution. Overall, the characteristics and mixing proportions of the two classes are similar across the two methods. In both cases, Class 1 has a higher initial level β_I and higher rate of change β_S than Class 2, and the probability of belonging to Class 1 (0.32) is lower than the probability of belonging to Class 2 (0.68). Based on these results, the classes represent customers that tend to spend more (less) at the restaurant, respectively. Parameter estimates from the two methods differ to some degree. The β_I and β_S estimates are higher for the mean growth mixture model, indicating that outliers influenced these estimates. For example, in the case of the β_I estimates for Class 1, the mean GMM estimated this parameter to be 18.55, which is higher than what the Medan GMM estimated (i.e., 17.65). Regarding the estimated variance of the intercepts for customers in Class 1 (i.e., $\Psi(1, 1)$), both methods indicated that they vary substantially. This result stems from the fact that some of the customers in Class 1 have extreme values at Time 1. By contrast, the estimated variance of the intercepts for customers in Class

Fig. 3 A collection of individual trajectories

Table 3 Parameter estimates for the two-class solution

	Median GMM		Mean GMM	
	Class 1	Class 2	Class 1	Class 2
β_I	17.65	10.47	18.55	10.75
	(16.66, 18.70)	(10.27, 10.68)	(17.51, 19.62)	(10.53, 10.96)
β_S	0.59	0.16	0.61	0.20
	(0.25, 0.94)	(0.09, 0.23)	(0.24, 0.99)	(0.12, 0.28)
$\Psi(1, 1)$	41.87	3.84	51.31	4.12
	(31.01, 54.42)	(3.12, 4.61)	(38.95, 65.56)	(3.28, 5.08)
$\Psi(1, 2)$	–1.62	–0.01	–0.92	0.04
	(-4.90, 1.11)	(-0.19, 0.15)	(-4.42, 2.06)	(-0.21, 0.27)
$\Psi(2, 2)$	0.79	0.08	0.56	0.17
	(0.15, 2.12)	(0.05, 0.14)	(0.13, 1.72)	(0.08, 0.29)
π	0.32	0.68	0.32	0.68
	(0.29, 0.36)	(0.64, 0.71)	(0.28, 0.35)	(0.65, 0.72)

The numbers in the parentheses represent 95% credible intervals

2 is much lower. The 95% credible intervals for $\Psi(1, 2)$ indicate that the covariances between β_I and β_S for the two methods were not significantly different from 0 for the two latent classes, which suggests that there was no relationship between customers' initial average spend and the rate of change in their average spends.

Table 4 presents each method's parameter estimates for the three-class solution. The purchase trajectories for the three classes can be characterized as high, medium, and low, respectively. Similar to the two-class solution, the β_I and β_S estimates are higher for the mean GMM, especially regarding the estimates for Class 1. The mixing proportion estimates for Class 1, Class 2, and Class 3 were 0.12,

0.44, and 0.44, respectively, for both methods. Across the methods, Class 1 has the highest initial level and rate of change, but these parameters vary considerably across subjects. Class 2 and Class 3 mainly differed in their initial level (i.e., β_I). The 95% credible intervals for $\Psi(1, 2)$ indicated that $\Psi(1, 2)$ are nonsignificant across the two methods, save for the covariance between β_I and β_S for Class 1 in the median GMM.

In sum, the purpose of this empirical data analysis is to demonstrate how researchers can apply the median GMM, and compare it to the mean GMM in practice. Since the one-class and four-class solutions lacked interpretability, we present the parameter estimates for the two- and three-class

Table 4 Parameter estimates for the three-class solution

	Median GMM			Mean GMM		
	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
β_I	23.26	13.09	9.75	24.45	13.70	9.90
	(20.87, 25.68)	(12.51, 13.70)	(9.53, 9.97)	(22.06, 26.97)	(13.20, 14.19)	(9.68, 10.12)
β_S	1.02	0.30	0.14	1.13	0.30	0.15
	(0.24, 1.82)	(0.11, 0.51)	(0.08, 0.21)	(0.21, 2.05)	(0.11, 0.50)	(0.08, 0.23)
$\Psi(1, 1)$	75.64	6.68	2.81	89.10	5.23	2.78
	(45.11, 114.14)	(4.21, 9.59)	(2.24, 3.46)	(49.38, 139.25)	(2.93, 7.90)	(2.15, 3.49)
$\Psi(1, 2)$	–10.19	–0.60	–0.08	–9.54	0.08	–0.06
	(-23.61, -0.26)	(-1.74, 0.28)	(-0.22, 0.04)	(-24.62, 2.02)	(-0.74, 0.66)	(-0.23, 0.09)
$\Psi(2, 2)$	2.14	0.72	0.08	1.88	0.37	0.11
	(0.27, 6.07)	(0.26, 1.36)	(0.04, 0.12)	(0.19, 5.83)	(0.13, 0.82)	(0.06, 0.17)
π	0.12	0.44	0.44	0.12	0.44	0.44
	(0.10, 0.15)	(0.39, 0.49)	(0.38, 0.49)	(0.09, 0.14)	(0.40, 0.49)	(0.40, 0.48)

The numbers in the parentheses represent 95% credible intervals

solutions. The empirical data used in this analysis contained extremely high values at each measurement occasion. For both two- and three-class solutions, the median GMM and the mean GMM provided different parameter estimates for the latent classes that covered the extreme values. Our earlier simulation study suggests that the reason for the difference is that the mean GMM is more sensitive to outliers than the Median GMM, and it led to higher intercept and slope estimates than those from the median GMM. This empirical data analysis also implies that this pattern did not disappear by increasing the number of latent classes (i.e., from two to three).

Discussion

This paper proposed a robust Bayesian approach to estimating growth mixture models based on conditional medians. The median provides a useful measure when data are skewed or have outliers, and it has been used for a variety of models because of its robust properties and interpretability. In this article, we described how the median can be used in the context of growth mixture modeling, and evaluated the performance of the proposed model under various conditions using a Monte Carlo simulation study. Our simulation study showed that growth mixture modeling based on conditional medians performs as well as traditional growth mixture modeling based on conditional means when data are normally distributed, and outperforms the traditional method when data are nonnormal.

More specifically, our simulation study showed that convergence rates for median and traditional growth mixture modeling are different. The two methods had higher convergence rates under the higher class separation condition. The traditional growth mixture model appeared to have convergence problems when the data had outliers or were skewed, particularly under the medium class separation condition. In these nonnormal cases, the convergence rates were notably lower for the unbalanced mixing proportion condition than those for the balanced mixing proportion condition. On the other hand, the median growth mixture model had convergence rates over 99%, regardless of the type of distribution, class separation, or mixing proportion. Stronger priors may provide better convergence rates for the nonnormal conditions of the traditional method. In practice, however, the implementation of strong priors is oftentimes infeasible. The fact that median growth mixture modeling generally had reasonable convergence rates with relatively weak informative priors implies great potential for the proposed method. When informative priors are not available, the median-based GMM is recommended.

The performance of median-based growth mixture modeling depended on class separation. Overall, bias

and MSE were lower for the higher separation condition than those for the medium separation condition. Mixing proportions were generally well recovered for both class separation conditions regardless of the type of distributions. In the case of the traditional growth mixture model, as reported in Depaoli (2013), class separation influenced the parameter estimation in our simulation study. When data had outliers, the traditional growth mixture model recovered mixing proportions well under the higher class separation condition, but it failed to recover them under the medium class separation condition.

In an empirical data analysis, deciding an appropriate number of latent classes is one challenge in growth mixture modeling. In our empirical data analysis, we only considered the interpretability of the results and presented the two- and three-class solutions for the purpose of illustration. In practice, choosing the number of latent classes in a Bayesian framework can be guided by using information criteria such as DIC (Spiegelhalter, Best, Carlin, & Van Der Linde, 2002). Several research studies on traditional growth mixture modeling have made reference to the possibility of over-extraction of latent classes when data do not follow the within-class normality assumption (e.g., Bauer & Curran, 2003; Muthén & Asparouhov, 2015). In a recent study on identifying the number of latent classes using information criteria (Kim, Tong, & Ke, 2021), median-based growth mixture modeling was more resistant to outliers in terms of identifying the number of latent classes, indicating that the median-based method can provide more parsimonious solutions and more interpretable estimation results.

This study assumed that latent growth factors follow a multivariate normal distribution, and the distribution of measurement errors was manipulated to generate various nonnormal data. In practice, there are many other factors that can cause nonnormality. Muthén & Asparouhov (2015) and Depaoli, Winter, Lai, and Guerra-Peña (2019) considered various types of skewed distributions for latent growth factors, and Tong et al. (2020) considered the influence of data with leverage observations, which are caused by extreme scores in random effects. It would be worth evaluating the performance of median-based growth mixture modeling for these various types of nonnormal data and comparing median-based growth mixture modeling to other robust methods such as utilizing a t-distribution or other skewed distributions in growth mixture modeling. In this direction, we also want to point out that outliers can be results of unusual events such as incorrectly recorded data or a measuring instrument failure, or unusual but plausible observations which can control key model properties (Montgomery, Peck, and Vining, 2006). In this study, as shown in our data generation procedure, we assumed that outliers were randomly generated without

having any patterns. Namely, we did not consider the latter cause of outliers. Note that this study also did not consider the case in which data have missing values. This study, however, can be adapted to include missing data as Bayesian methods are flexible in this regard. These topics will be further investigated in our future research.

The present paper shows that median-based growth mixture modeling has many advantages over traditional mean-based growth mixture modeling, especially when data do not satisfy the within-class normality assumption. As discussed in the literature, it is often the case that empirical data are skewed or contaminated by outliers (Micceri, 1989). The overall findings of this paper suggest that median-based growth mixture modeling is more resilient to violations of the normality assumption than traditional growth mixture modeling. The findings also suggest that median-based growth mixture modeling has better convergence rates than the traditional approach. Researchers who need to fit complex models such as growth mixture models are often challenged by nonconvergence. If results from a traditional growth mixture model are deemed to be influenced by outlying observations, the median-based approach that we propose offers a likely solution because it provides a straightforward way to achieve robustness while lessening the chances of encountering nonconvergence issues.

Appendix

In growth mixture modeling based on conditional medians, the jth observation for individual i in class g can be specified as follows

$$y_{ij} = \lambda_j' b_{ig} + \zeta v_{ij} + \xi \sqrt{\delta v_{ij}} W_{ij}, \tag{7}$$

where $b_{ig} = \beta_g + u_i$, $u_i \sim N(\mathbf{0}, \Psi)$, $v_{ij} \sim \exp(\delta)$, $W_{ij} \sim N(0, 1)$, $\zeta = \frac{1-2\tau}{\tau(1-\tau)}$, and $\xi^2 = \frac{2}{\tau(1-\tau)}$. For our median growth mixture model, $\tau = 0.5$, which leads to $\zeta = 0, \xi^2 = 8$. Thus, Equation 7 can be rewritten as

$$y_{ij} = \lambda_j' b_{ig} + \sqrt{8\delta v_{ij}} W_{ij}.$$

In this appendix, we assume that only mean trajectories vary across latent classes. Full conditional distributions for growth mixture models that allow for class-specific covariance matrices can be easily obtained by revising the provided equations.

It is common for mixture modeling to introduce a class membership variable $z_i \in \{1, \ldots, G\}$ with probability $P(z_i = g) = \pi_g$ $(g = 1, \ldots, G)$. Let $y_i = (y_{i1}, \ldots, y_{iJ})'$, $y = \{y_1, \ldots, y_N\}$, $v_i = (v_{i1}, \ldots, v_{iJ})'$, $v = \{v_1, \ldots, v_N\}$, $u = \{u_1, \ldots, u_N\}$, and $z = \{z_1, \ldots, z_N\}$. The complete-data likelihood of observed responses in y and unobserved latent variables in v, u, and z is

$$
\begin{aligned}
&L(\Theta|y, v, u, z) \\
&= f(y|v, u, z, \beta, \delta) f(v|\delta) p(z|\pi) f(u|\Psi) \\
&= \prod_{i=1}^{N} \left[\prod_{j=1}^{J} \left\{ f(y_{ij}|v_{ij}, u_i, \beta_{z_i}, \delta) f(v_{ij}|\delta) \right\} p(z_i|\pi) f(u_i|\Psi) \right].
\end{aligned}
$$

We used the following conjugate prior distributions: $\pi \sim Dirichlet(\alpha)$, $\beta_g \sim N(\beta_{0g}, \Sigma_{0g})$, $\Psi \sim InvWishart(n_0, S_0)$, and $\delta \sim InvGamma(c_0, d_0)$.

1. The full conditional distribution of π is a Dirichlet distribution

   $$\pi \sim Dirichlet(x_1 + \alpha_1, \ldots, x_G + \alpha_G),$$

   where x_g is the number of individuals that belong to class g.

2. The full conditional distribution of z_i $(i = 1, \ldots, N)$ is a multinomial distribution with a probability of belonging to class g with

   $$P(z_i = g|y_i, v_i, u_i, \beta_g, \delta) = \pi_g \prod_{j=1}^{J} f(y_{ij}|v_{ij}, u_i, \beta_g, \delta).$$

3. The full conditional distribution of β_g $(g = 1, \ldots, G)$ is a normal distribution $\beta_g \sim N(\mu_{\beta_g}, \Sigma_{\beta_g})$, where

   $$\mu_{\beta_g} = \Sigma_{\beta_g} \left(\Sigma_{0g}^{-1} \beta_{0g} + \sum_{i \in N_g} \sum_{j=1}^{J} \frac{1}{8\delta v_{ij}} \left(\lambda_j y_{ij} - \lambda_j \lambda_j' u_i \right) \right),$$

   $$\Sigma_{\beta_g} = \left(\Sigma_{0g}^{-1} + \sum_{i \in N_g} \sum_{j=1}^{J} \frac{\lambda_j \lambda_j'}{8\delta v_{ij}} \right)^{-1},$$

   in which N_g represents a group of individuals that belong to class g.

4. The full conditional distribution of u_i $(i = 1, \ldots, N)$ is a normal distribution $u_i \sim N(\mu_{ui}, \Sigma_{ui})$, where

   $$\mu_{ui} = \Sigma_{ui} \left(\sum_{j=1}^{J} \frac{\lambda_j y_{ij} - \lambda_j \lambda_j' \beta_{z_i}}{8\delta v_{ij}} \right) \quad and \quad \Sigma_{ui} = \left(\Psi^{-1} + \sum_{j=1}^{J} \frac{\lambda_j \lambda_j'}{8\delta v_{ij}} \right)^{-1}.$$

5. The full conditional distribution of δ is an inverse gamma distribution $\delta \sim InvGamma(c_p, d_p)$, where

   $$c_p = c_o + \frac{3NJ}{2},$$

   $$d_p = d_0 + \sum_{i=1}^{N} \sum_{j=1}^{J} \left(\frac{(y_{ij} - \lambda_j' \beta_{z_i} - \lambda_j' u_i)^2}{16 v_{ij}} + v_{ij} \right).$$

6. The full conditional distribution of Ψ is an inverse Wishart distribution $\Psi \sim InvWishart(n_p, S_p)$, where $n_p = n_0 + N$ and $S_p = S_0 + \sum_{i=1}^{N} u_i u_i'$.

7. The full conditional distribution of v_{ij} $(i = 1, \ldots, N;$ $j = 1, \ldots, J)$ is a generalized inverse Gaussian

distribution $v_{ij} \sim GIG(\rho_a, \rho_b, \rho_p)$, where $\rho_a = \frac{2}{\delta}$, $\rho_b = \frac{(y_{ij} - \lambda'_j \beta_{z_i} - \lambda'_j u_i)^2}{8\delta}$, and $\rho_p = \frac{1}{2}$.

Open Practices Statements The empirical data and programming code are available on our GitHub site: https://github.com/CynthiaXinTong/MedianGMM.git

References

Bassett, G., & Koenker, R. (1978). Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association*, *73*(363), 618–622. httpd://doi.org/10.1080/01621459.1978.10480065.

Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, *42*(4), 757–786. https://doi.org/10.1080/00273170701710338.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*(3), 338–363. https://doi.org/10.1037/1082-989X.8.3.338.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective* Vol. 467. Hoboken: Wiley. https://doi.org/10.1002/0471746096.

Connell, A. M., & Frye, A. A. (2006). Growth mixture modelling in developmental psychology: Overview and demonstration of heterogeneity in developmental trajectories of adolescent antisocial behaviour. *Infant and Child Development: An International Journal of Research and Practice*, *15*(6), 609–621. https://doi.org/10.1002/icd.481.

R Core Team (2019). R: a language and environment for statistical computing r foundation for statistical computing. Vienna, Austria.

Davino, C., Furno, M., & Vistocco, D. (2014). *Quantile regression: Theory and applications* Vol. 988. Chichester: Wiley. https://doi.org/10.1002/9781118752685.

Depaoli, S. (2013). Mixture class recovery in gmm under varying degrees of class separation: Frequentist versus Bayesian estimation. *Psychological Methods*, *18*(2), 186–219. https://doi.org/10.1037/a0031609.

Depaoli, S., Winter, S. D., Lai, K., & Guerra-Peña, K. (2019). Implementing continuous non-normal skewed distributions in latent growth mixture modeling: an assessment of specification errors and class enumeration. *Multivariate behavioral research*, 1–27. https://doi.org/10.1080/00273171.2019.1593813.

Depaoli, S., Yang, Y., & Felt, J. (2017). Using Bayesian statistics to model uncertainty in mixture models: a sensitivity analysis of priors. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(2), 198–215.

Dziak, J. J., Li, R., Tan, X., Shiffman, S., & Shiyko, M. P. (2015). Modeling intensive longitudinal data with mixtures of nonparametric trajectories and time-varying effects. *Psychological Methods*, *20*(4), 444. https://doi.org/10.1037/met0000048.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied longitudinal analysis*. Hoboken: Wiley. https://doi.org/10.1002/9781119513469.

Geraci, M. (2017). Mixed-effects models using the normal and the Laplace distributions: A 2 × 2 convolution scheme for applied research. arXiv:1712.07216.

Geraci, M., & Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics*, *8*(1), 140–154. https://doi.org/10.1093/biostatistics/kxj039.

Geweke, J. (1991). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bernardo,

J. M., Berger, J. O., Dawid, A. P., & Smith, A. F. M. (Eds.) *Bayesian statistics 4*, (pp. 169–193). Oxford: Clarendon Press.

Gilchrist, W. (2000). *Statistical modelling with quantile functions*. Boca Raton: CRC Press. https://doi.org/10.1201/9781420035919.

Guay, F., Morin, A. J., Litalien, D., Howard, J. L., & Gilbert, W. (2021). Trajectories of self-determined motivation during the secondary school: A growth mixture analysis. *Journal of Educational Psychology*, *113*(2), 390.

Guerra-Peña, K., García-batista, Z. E., Depaoli, S., & Garrido, L. E. (2020). Class enumeration false positive in skew-t family of continuous growth mixture models. *Plos One*, *15*(4), e0231525. https://doi.org/10.1371/journal.pone.0231525.

He, X., Fu, B., & Fung, W. K. (2003). Median regression for longitudinal data. *Statistics in Medicine*, *22*(23), 3655–3669. https://doi.org/10.1002/sim.1581.

Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods*, *11*(1), 36–53. https://doi.org/10.1037/1082-989X.11.1.36.

Huang, D. Y., Evans, E., Hara, M., Weiss, R. E., & Hser, Y.-I. (2011). Employment trajectories: Exploring gender differences and impacts of drug use. *Journal of Vocational Behavior*, *79*(1), 277–289. https://doi.org/10.1016/j.jvb.2010.12.001.

Kim, S., Tong, X., & Ke, Z. (2021). Exploring class enumeration in Bayesian growth mixture modeling based on conditional medians. *Frontiers in Education*, *6*, 624149. https://doi.org/10.3389/feduc.2021.624149.

Koenker, R. (2005). *Quantile regression. Econometric society monographs*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511754098.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50. https://doi.org/10.2307/1913643.

Koenker, R., & Bassett, G. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, *50*(1), 43–61. https://doi.org/10.2307/1912528.

Koenker, R., & Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, *94*(448), 1296–1310. https://doi.org/10.1080/01621459.1999.10473882.

Kotz, S., Kozubowski, T., & Podgórski, K. (2001). *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Boston: Birkhäuser. https://doi.org/10.1007/978-1-4612-0173-1.

Kozumi, H., & Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation*, *81*(11), 1565–1578. https://doi.org/10.1080/00949655.2010.496117.

Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. Chichester: Wiley. https://doi.org/10.1002/9780470024737.

Liu, M., & Hancock, G. R. (2014). Unrestricted mixture models for class identification in growth mixture modeling. *Educational and Psychological Measurement*, *74*(4), 557–584. https://doi.org/10.1177/0013164413519798.

Lu, Z., & Song, X. (2012). Finite mixture varying coefficient models for analyzing longitudinal heterogenous data. *Statistics in Medicine*, *31*(6), 544–560. https://doi.org/10.1002/sim.4420.

Lu, Z. L., & Zhang, Z. (2014). Robust growth mixture models with non-ignorable missingness: Models, estimation, selection, and application. *Computational Statistics & Data Analysis*, *71*, 220–240. https://doi.org/10.1016/j.csda.2013.07.036.

Lu, Z. L., Zhang, Z., & Lubke, G. (2011). Bayesian inference for growth mixture models with latent class dependent missing data. *Multivariate Behavioral Research*, *46*(4), 567–597. https://doi.org/10.1080/00273171.2011.589261.

Lubke, G., & Neale, M. C. (2006). Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood?. *Multivariate Behavioral Research*, *41*(4), 499–532. https://doi.org/10.1207/s15327906mbr4104_4

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, *60*, 577–605.

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, *58*(1), 110–133. https://doi.org/10.2307/1130295.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley. https://doi.org/10.1002/0471721182.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*(1), 107–122. https://doi.org/10.1007/BF02294746.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*(1), 156–166. https://doi.org/10.1037/0033-2909.105.1.156.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2006). *Introduction to linear regression analysis*. Hoboken: Wiley.

Mortimer, S. (2018). squareupr: An implementation of the square platform apis. Available at: https://github.com/StevenMMortimer/squareupr.

Muthén, B. (2004). Latent variable analysis: growth mixture modeling and related techniques for longitudinal data. In Kaplan, D. (Ed.) *The SAGE handbook of quantitative methodology for the social sciences.* https://doi.org/10.4135/9781412986311, (pp. 345–368). Thousand Oaks: SAGE Publications Inc.

Muthén, B., & Asparouhov, T. (2015). Growth mixture modeling with non-normal distributions. *Statistics in Medicine*, *34*(6), 1041–1058. https://doi.org/10.1002/sim.6388.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the em algorithm. *Biometrics*, *55*(2), 463–469. https://doi.org/10.1111/j.0006-341X.1999.00463.x.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569.

Peugh, J., & Fan, X. (2012). How well does growth mixture modeling identify heterogeneous growth trajectories? a simulation study examining gmm's performance characteristics. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*(2), 204–226. https://doi.org/10.1080/10705511.2012.659618.

Peugh, J., & Fan, X. (2015). Enumeration index performance in generalized growth mixture models: a Monte Carlo test of muthén's (2003) hypothesis. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(1), 115–131. https://doi.org/10.1080/10705511.2014.919823.

Plummer, M. (2003). Jags: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing.* Vienna, Austria, (Vol. 124, pp. 1–10).

Plummer, M. (2017). Jags version 4.3. 0 user manual.

Reich, B. J., Bondell, H. D., & Wang, H. J. (2010). Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics*, *11*(2), 337–352. https://doi.org/10.1093/biostatistics/kxp049.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195152968.001.0001.

Son, S., Lee, H., Jang, Y., Yang, J., & Hong, S. (2019). A comparison of different nonnormal distributions in growth mixture models. *Educational and Psychological Measurement*, *79*(3), 577–597. https://doi.org/10.1177/0013164418823865.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series b (Statistical Methodology)*, *64*(4), 583–639. https://doi.org/10.1111/1467-9868.00353.

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, *82*(398), 528–540. https://doi.org/10.1080/01621459.1987.10478458.

Tofighi, D., & Enders, C. K. (2008). Identifying the correct number of classes in growth mixture models. In Hancock, G. R., & Samuelson, K. M. (Eds.) *Advances in latent variable mixture models*, (pp. 317–341). Charlotte: Information Age.

Tong, X., Zhang, T., & Zhou, J. (2020). Robust Bayesian growth curve modelling using conditional medians. British Journal of Mathematical and Statistical Psychology. https://doi.org/10.1111/bmsp.12216.

Tueller, S., & Lubke, G. (2010). Evaluation of structural equation mixture models: Parameter estimates and correct class assignment. *Structural Equation Modeling*, *17*(2), 165–192. https://doi.org/10.1080/10705511003659318.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*(116), 3571–3594.

Yu, K., & Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, *54*(4), 437–447. https://doi.org/10.1016/S0167-7152(01)00124-9.

Zhang, Z. (2016). Modeling error distributions of growth curve models through Bayesian methods. *Behavior Research Methods*, *48*(2), 427–444. https://doi.org/10.3758/s13428-015-0589-9.

Springer