

Extending Latent Basis Growth Model to Explore Joint Development in the Framework of Individual Measurement Occasions

Jin Liu¹[0000–0001–5922–6643]

Data Sciences Institute, Takeda Pharmaceuticals, Boston, MA 02142, USA
Veronica.Liu0206@gmail.com

Abstract. Longitudinal processes often exhibit nonlinear change patterns. Latent basis growth models (LBGMs) provide a versatile solution without requiring specific functional forms. Building on the LBGM specification for unequally-spaced waves and individual measurement occasions proposed by [Liu and Perera \(2024\)](#), we extend LBGMs to multivariate longitudinal outcomes. The extended models enable the analysis of nonlinear parallel longitudinal processes with unequally-spaced study waves in the framework of individual measurement occasions. We present the proposed models by simulation studies and real-world data analyses. Simulation studies demonstrate that the proposed model can provide unbiased and accurate estimates with target coverage probabilities for the parameters of interest. Real-world analyses of reading and mathematics scores demonstrate its effectiveness in analyzing joint developmental processes that vary in temporal patterns. Computational code is included.

Keywords: Latent Basis Growth Model · Parallel Nonlinear Longitudinal Processes · Individual Measurement Occasions · Simulation Studies

1 Introduction

In longitudinal studies, researchers often gather measurements on multiple outcomes to decipher how each evolves over time. While the focus has traditionally been on univariate outcomes, the inter-correlated nature of processes in domains such as development ([Liu & Perera, 2022](#); [Peralta, Kohli, Lock, & Davison, 2022](#); [Shin, Davison, Long, Chan, & Heistad, 2013](#)), behavioral sciences ([Duncan & Duncan, 1994, 1996](#)), and biomedicine ([Dumenci et al., 2019](#)) demands a multifaceted analysis. Recent research reflects a growing interest in exploring how these interconnected outcomes influence one another over time. Developmental studies, for example, often track achievement scores across multiple subjects ([Liu & Perera, 2022, 2023](#); [Peralta et al., 2022](#); [Shin et al., 2013](#)), facilitating an in-depth analysis of correlated growth in multiple domains. Similarly, clinical trials might collect multiple endpoints ([Dumenci et al., 2019](#)) to provide a

holistic evaluation of treatment effects. This complexity underscores the need for advanced modeling techniques that accurately capture the correlation between multiple longitudinal processes.

Another scenario highlighting the complexity of longitudinal studies involves the reconciliation of data from diverse sources. For instance, observational studies may utilize both child and parent reports to assess a child’s health-related quality of life (Rajmil, López, López-Aguilà, & Alonso, 2013). In clinical trials, a single endpoint is often measured using different equipments, adding complexity to data interpretation. Additionally, analyzing repeated outcomes from individuals nested within pairs or small groups (Lyons et al., 2017; McNulty, Wenner, & Fisher, 2016) presents unique statistical challenges. These situations underscore the need for a model capable of describing the joint longitudinal processes, with the aim of elucidating the associations between varied data sources and outcomes. The objective of our study is to develop such models within the Structural Equation Modeling (SEM) framework, as SEM provides a flexible and comprehensive approach for capturing complex relationships and dependencies between variables.

Research in developmental psychology has provided insights into the joint development of cognitive abilities, such as the studies by Robitaille, Muniz, Piccinin, Johansson, and Hofer (2012), revealed complex nonlinear intercept and slope associations in the progression of visuospatial ability and processing speed, using multivariate growth models (MGMs) with linear growth curves. However, a model with linear function often falls short in capturing the full complexity of real-world longitudinal processes, which frequently exhibit nonlinearity and thus necessitate more sophisticated analytical approaches. To better model such complexity, Blozis (2004) developed MGMs with nonlinear parametric functions, such as polynomial and exponential forms, to capture nonlinear parallel growth, which were implemented using LISREL in Blozis, Harring, and Mels (2008). Despite these advancements, parametric models with predetermined nonlinear functional forms may not adequately represent actual change patterns that do not conform to the prespecified functional forms. Furthermore, while these MGMs can estimate correlations between growth factors, such as intercept-intercept and linear/quadratic slope-slope relationships for quadratic functions, they often fail to provide insights into the relationship between two nonlinear longitudinal processes directly.

As the field has evolved, there has been a notable shift toward semi-parametric methods to allow for more flexible analysis in multivariate nonlinear longitudinal processes. Liu and Perera (2022) exemplified this shift with their linear-linear piecewise function within the SEM framework. A longitudinal model with a linear-linear piecewise function divides the growth trajectory into two linear segments, each with its own slope, joined at a specific point (i.e., the knot). By breaking down the growth curve into two distinct phases, the model allows for the assessment of slope-slope correlation at each stage. More importantly, by estimating the knots and their variances, the model also examines knot-knot correlations. This approach provides a unique perspective on the developmental

process by helping to understand how the correlation changes in each stage and when the transition from one stage to the next occurs.

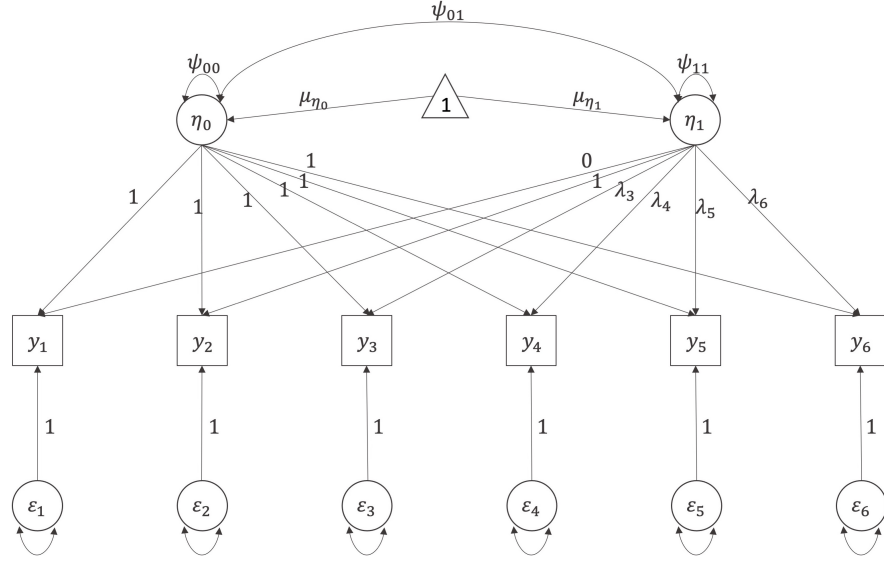
However, while semi-parametric functional forms like those introduced by Liu and Perera (2022) significantly enhance modeling flexibility, they inherently impose constraints by limiting the change patterns to only two distinct phases. Such two-piece functional forms may not adequately capture more complex developmental patterns that exhibit multiple phases over time, particularly in exploratory research stages where the underlying change patterns are not well-defined. Herein lies the advantage of latent basis growth models (LBGMs), which provide greater flexibility by allowing for the determination of the optimal curve shape without the constraints of prior assumptions (McArdle & Epstein, 1987; Meredith & Tisak, 1990). Our work builds on this flexibility to facilitate explorations of multiple longitudinal processes, addressing the need for adaptable analytical tools capable of handling the complexities of real-world challenges.

1.1 Traditional Specification of Latent Basis Growth Model

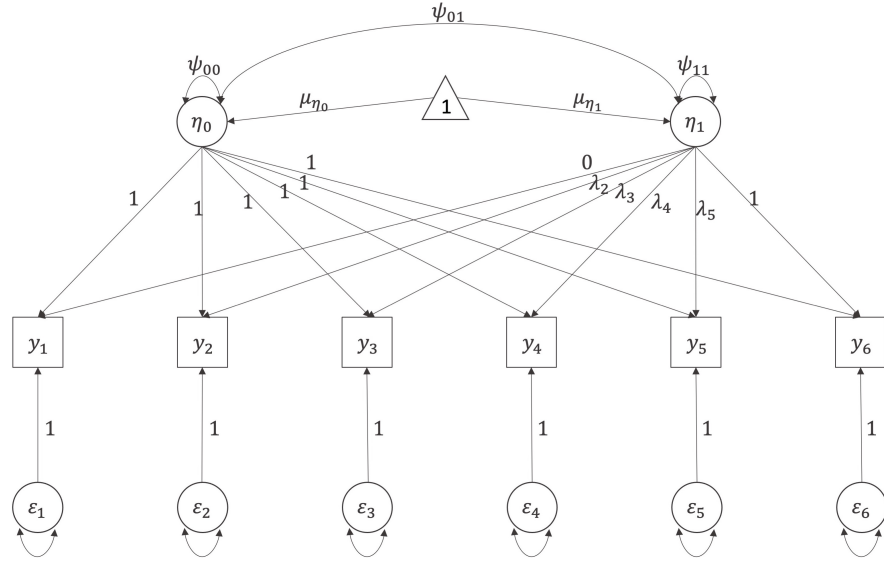
Grimm, Ram, and Estabrook (2016, Chapter 11) demonstrated that LBGMs can be constructed using both the Latent Growth Curve Modeling (LGCN) framework, a subset of the SEM framework, and the mixed-effects modeling framework. While LBGMs were not explicitly discussed, existing literature suggests that, for a majority of longitudinal models, these two frameworks are mathematically equivalent in evaluating between-individual differences in within-individual changes (Bauer, 2003; Curran, 2003). This study focuses on the SEM framework due to its greater modeling flexibility and widespread recognition within the social science research community.

Similar to other latent growth curve models, a LBGM can be expressed as $\mathbf{y}_i = \mathbf{A}\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i$, where \mathbf{y}_i represents the vector of repeated measurements for individual i , $\boldsymbol{\eta}_i$ is the vector of latent growth factors for individual i , \mathbf{A} is the matrix of factor loadings, and $\boldsymbol{\epsilon}_i$ is the residual vector of individual i . Simply put, this equation captures how an individual's change patterns are represented by latent growth factors and measurement occasions. LBGMs typically consist of two growth factors: an intercept and a shape factor.

The factor loading matrix \mathbf{A} is partially constrained for model identification. Specifically, in a setting with J measurements, factor loadings for the intercept are fixed at 1, while two factor loadings for the shape factor are also fixed, and the remaining $J - 2$ are estimated. Figures 1a and 1b illustrate two common specifications of LBGM with six repeated measurements. In Figure 1a, the shape factor is scaled as the change during the initial time interval. In Figure 1b, the shape factor is scaled as the total change over the study duration. These methods allow for the flexible estimation of \mathbf{A} , thus freeing LBGM from being restricted to a specific functional form. This flexibility in specification allows LBGMs to adapt to different research questions and datasets, making them a powerful tool for longitudinal data analysis.



(a) Specification 1



(b) Specification 2

Figure 1: Path Diagram of Traditional Latent Basis Growth Models

1.2 Novel Specification of Latent Basis Growth Model

Although the LBGM described in Section 1.1 is a flexible statistical tool to explore trajectories, it does not specify whether nonlinearity exists in the growth patterns (Grimm, Steele, Ram, & Nesselroade, 2013) nor does it detail the nature of such nonlinearity (Wood, Steinley, & Jackson, 2015); it still has limitations. According to Grimm et al. (2016, Chapter 11), discrete time points are required when specifying an LBGM, and therefore, it cannot be fit in the framework of individual measurement occasions. One approximate method for such continuous measurement time is the time-bins approach, also known as the time-windows method. The time-bins approach involves dividing the assessment period into several intervals (time-bins), where each individual can have up to one response per bin. If a subject does not contribute data to a specific time window, it is treated as a missing record (Sterba, 2014).

However, several studies highlight the limitations of this approach. For example, Blozis and Cho (2008) demonstrated that using the time-bins approach may lead to inadmissible estimation, such as overestimating within-individual changes and underestimating between-individual differences, though these effects can be negligible if individual differences are not substantial. Moreover, Coulombe, Selig, and Delaney (2015) concluded that neglecting time differences often leads to undesirable outcomes, such as biased parameter estimates. Their evaluation of bias, efficiency, and Type I error rate under various conditions—different combinations of sample size, degree of heterogeneity, distribution of time, rate of change, and number of repeated measurements—showed that ignoring time differences can significantly affect the results.

Two parallel but distinct methods for accounting for individual measurement occasions have been developed by Sterba (2014) and Liu and Perera (2024). Sterba (2014) introduces an innovative approach by incorporating two growth factors—the intercept and shape factor. This model defines the loadings of the shape factor as a function of the specific timing of each individual’s measurements, accounting for deviations from a linear progression.

In contrast, the framework by Liu and Perera (2024) specifies the latent basis growth model by incorporating linear piecewise functional forms, which effectively capture the dynamics across J measurements segmented into $J - 1$ intervals. This model is designed to estimate interval-specific slopes and allows for an extension to derive both interval-specific changes and changes from the baseline. In particular, as illustrated in Figure 2a, the interval-specific change is quantified using the area under the curve (AUC) for the corresponding time interval, effectively representing the integral of the growth rate over that period. For example, consider the change from $t = 1$ to $t = 2$ calculated as: $0.8 \times (2 - 1) = 0.8$. This calculation is depicted in Figure 2b, where the change in growth is shown to increase by 0.8, from 21 to 21.8. Using AUC to represent interval-specific change relaxes the traditional constraints of LBGMs and allows for unequally-spaced study waves. For example, in Figure 2a, even if no measurement is taken at $t = 5$, the change from $t = 4$ to $t = 6$ can still be calculated

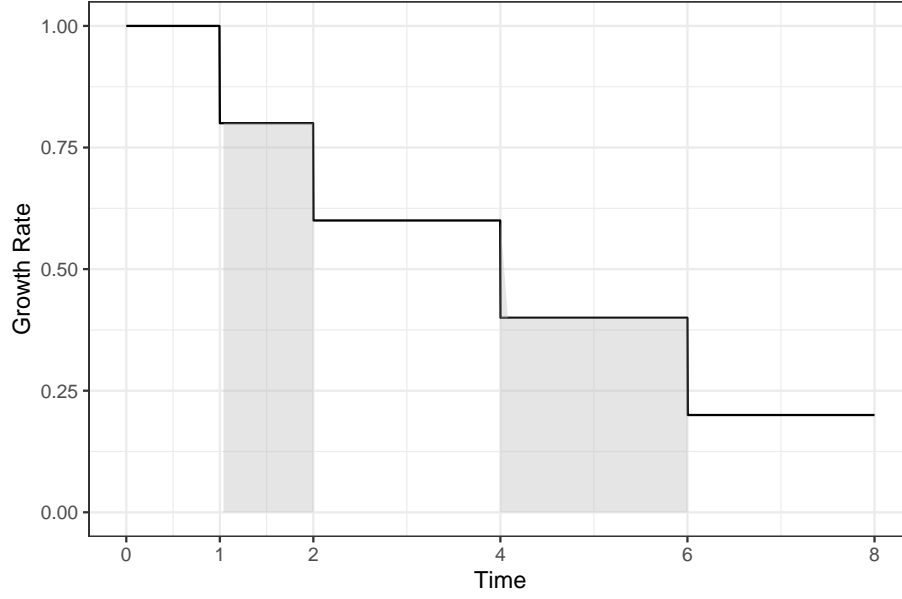
as $0.4 \times (6 - 4) = 0.8$. Similarly, the change from the baseline at any specific time is quantified using the AUC from the baseline to that particular time point.

The path diagram of the LBGM with six measurement occasions, as proposed by Liu and Perera (2024), is illustrated in Figure 3a. This model features two growth factors: the initial status, denoted as η_0 , and the slope during the first interval, denoted as η_1 . As depicted in Figure 3a, η_1 along with the relative rate γ_{j-1} , defines the interval-specific slopes (dy_{ij}). These slopes are then utilized, along with the length of each interval, to derive interval-specific changes. Each interval is enclosed in a diamond shape in the diagram, indicating that these intervals are allowed to vary among individuals. Such flexibility addresses the challenge of individual measurement occasions (which further lead to individual intervals), thus providing a more accurate representation of their growth trajectories. These time intervals are, therefore, considered ‘definition variables’, allowing the model to account for individual differences (Mehta & Neale, 2005; Mehta & West, 2000; Sterba, 2014).

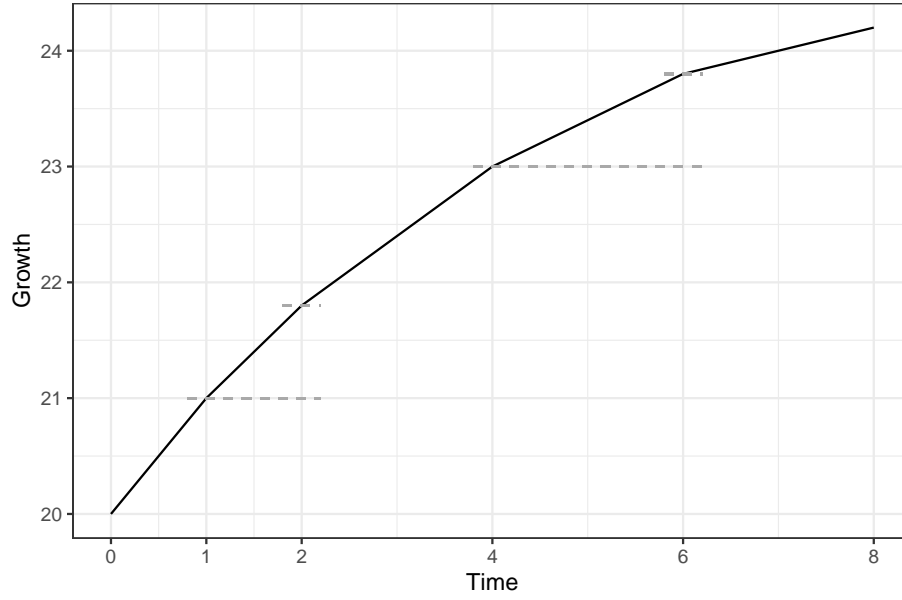
In addition to allowing for unequally-spaced study waves and individual measurement occasions in LBGM, this framework provides flexibility in scaling the growth rate factor, η_1 . Instead of constraining η_1 to the first time interval, it can be adapted to represent growth rate during any selected time frame, such as the last time interval, as demonstrated in Figure 3b. Here, γ_{j-1} still serve as the relative growth rate in relation to η_1 for each $(j - 1)^{th}$ time interval. Note that the models with different scalings of η_1 are mathematically equivalent. With such novel specifications, the shape factor’s loading at each measurement occasion t_j is calculated by dividing the change-from-baseline (the difference between the current value and the initial value at t_1) at t_j by η_1 . The setup of such factor loadings will be further explained in Section 2.1.

1.3 Parallel Latent Basis Growth Model

In the study of joint longitudinal processes, researchers frequently utilize MGMs, also known as parallel process and correlated growth models, which are thoroughly discussed in Grimm et al. (2016, Chapter 8) and McArdle (1988). MGMs are generally utilized to estimate three main types of associations based on the interactions they analyze: (1) within-process growth factors, (2) between-process growth factors, and (3) between-process residuals. Existing research, including studies by Robitaille et al. (2012), who investigated the co-evolution of processing speed and visuospatial ability using linear growth curves, and Blozis (2004); Blozis et al. (2008), who incorporated parametric nonlinear functional forms like polynomial and exponential curves, has significantly contributed to the understanding of these relationships. More recently, Peralta et al. (2022) and Liu and Perera (2022) have advanced this area by developing MGMs with linear-linear functional forms with unknown random knots, in the Bayesian mixed-effects and frequentist structural equation frameworks, respectively. Although effective for theory-driven research, these models sometimes lack the flexibility needed during the exploratory phases of research, especially in the absence of a guiding domain-specific theory for functional form selection.

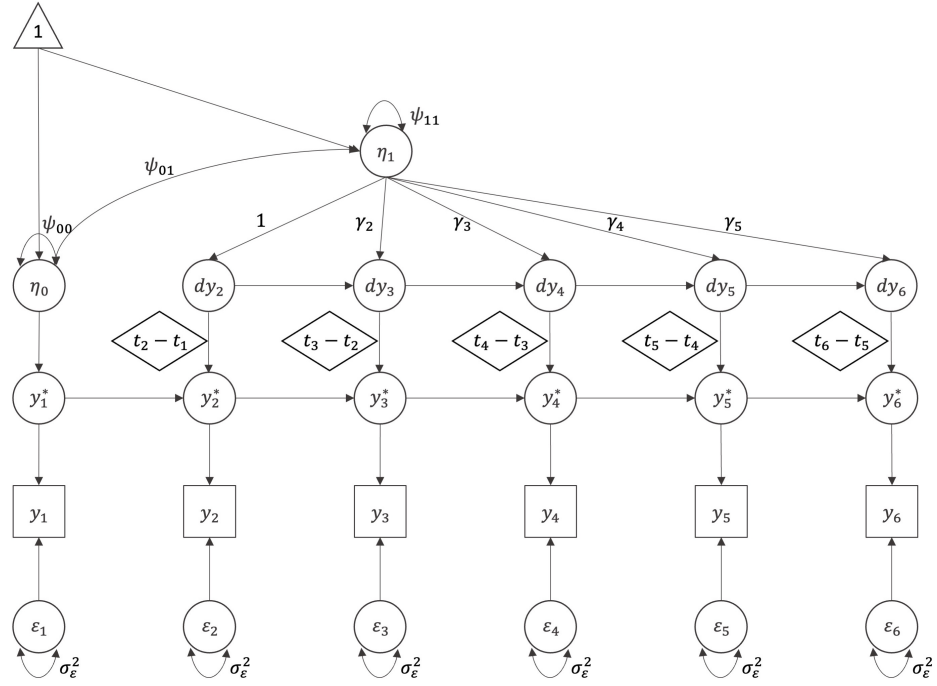


(a) Piecewise Linear Growth Rate

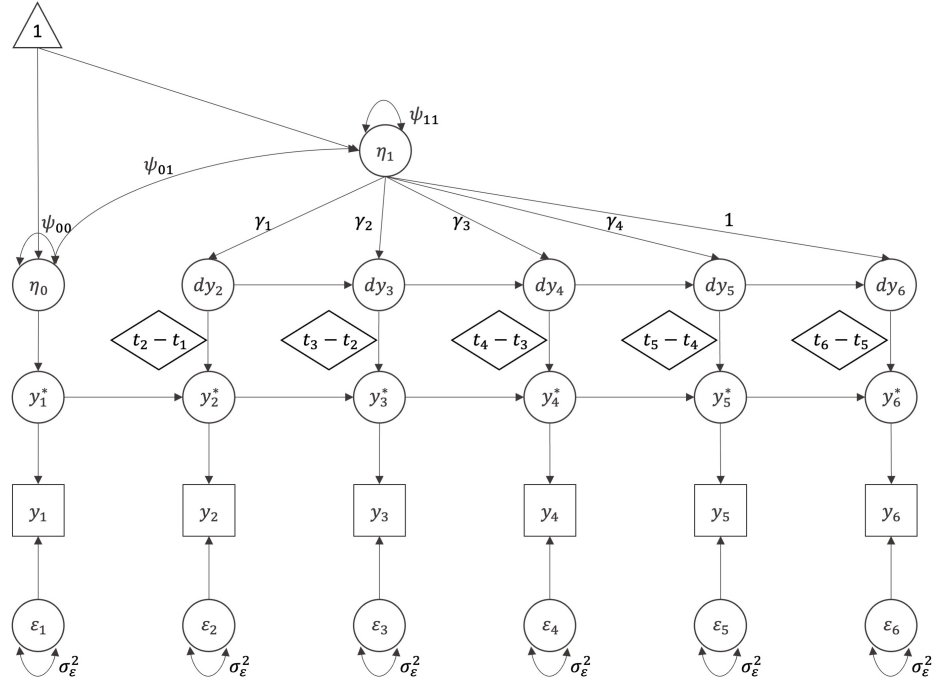


(b) Piecewise Linear Growth Curve

Figure 2: Piecewise Linear Growth Curve and Growth Rate (Values of the Intercept and Slope of Each Time Interval: $\eta_0 = 20$; $\gamma_1 = 1.0$; $\gamma_2 = 0.8$; $\gamma_3 = 0.6$; $\gamma_4 = 0.4$; $\gamma_5 = 0.2$)



(a) Specification 1



(b) Specification 2

Figure 3: Path Diagram of Novel Latent Basis Growth Models

This article aims to advance the field of joint longitudinal process modeling by extending the LBGM with the novel specification detailed in Section 1.2 to the MGM framework. The proposed model addresses existing gaps by demonstrating how to implement a parallel LBGM tailored to unequally-spaced study waves and individual measurement occasions. The structure of this article is organized as follows: We begin with a description of a LBGM for a univariate longitudinal process, incorporating our novel specification in the methods section. This model is then extended to a parallel growth curve framework, where we detail the model specification and estimation procedures. Subsequently, we evaluate the model’s performance using a Monte Carlo simulation study, focusing on metrics such as relative bias, empirical standard error (SE), relative root-mean-square error (RMSE), and the coverage probability (CP) of a 95% confidence interval. We also illustrate the practical application of our model by analyzing a real-world dataset of longitudinal reading and mathematics scores from the Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K: 2011). In the application section, we explore how to derive and interpret insights from the model output. Finally, we conclude with discussions on practical and methodological considerations and directions for future research.

2 Method

2.1 Latent Basis Growth Model in the Framework of Individual Measurement Occasions

This section introduces the novel LBGM specification developed by Liu and Perera (2024) for univariate nonlinear developmental trajectories, applicable to analyzing univariate longitudinal outcomes such as reading or mathematics development. For individual i , the model can be specified as

$$y_{ij} = y_{ij}^* + \epsilon_{ij}^{[y]}, \quad (1)$$

$$y_{ij}^* = \begin{cases} \eta_{0i}^{[y]}, & \text{if } j = 1 \\ y_{i(j-1)}^* + dy_{ij} \times (t_{ij} - t_{i(j-1)}), & \text{if } j = 2, \dots, J \end{cases}, \quad (2)$$

$$dy_{ij} = \eta_{1i}^{[y]} \times \gamma_{j-1}^{[y]} \quad (j = 2, \dots, J). \quad (3)$$

Equations 1 and 2 together specify a LBGM, where y_{ij} , y_{ij}^* , and $\epsilon_{ij}^{[y]}$ are the observed measurement, latent true score, and residual for the i^{th} individual at time j , respectively. At the baseline measurement (i.e., $j = 1$), the true score corresponds to the initial status growth factor ($\eta_{0i}^{[y]}$). For subsequent measurements (i.e., $j \geq 2$), the true score at time j is calculated as a linear combination of the score at the preceding time point $j - 1$ and the true change from time $j - 1$ to j . This true change is further defined as the product of the time interval ($t_{ij} - t_{i(j-1)}$) and the interval-specific slope (dy_{ij}). Equation 3 further represents the interval-specific slope, dy_{ij} , with a shape factor $\eta_{1i}^{[y]}$ and $\gamma_{j-1}^{[y]}$, where γ_{j-1}

($j = 2, \dots, J$) can be interpreted as the relative growth rate in relation to $\eta_{1i}^{[y]}$ during the $(j-1)^{th}$ time interval. In [Liu and Perera \(2024\)](#), the term $\eta_{1i}^{[y]}$ is scaled to represent the growth rate during the first time interval (i.e., $\gamma_{2-1} = 1$), as illustrated in [Figure 3a](#). As discussed in [Section 1.2](#), this term can also be scaled to correspond with the growth rate during any other time interval, such as the last one (i.e., $\gamma_{J-1} = 1$), as depicted in [Figure 3b](#).

The model specified in [Equations 1-3](#) can also be written in a matrix form:

$$\mathbf{y}_i = \mathbf{A}_i^{[y]} \times \boldsymbol{\eta}_i^{[y]} + \boldsymbol{\epsilon}_i^{[y]}, \quad (4)$$

$$\boldsymbol{\eta}_i^{[y]} = \boldsymbol{\mu}_\eta^{[y]} + \boldsymbol{\zeta}_i^{[y]}, \quad (5)$$

where \mathbf{y}_i is a $J \times 1$ vector representing the i^{th} individual's repeated measurements (with J denoting the number of such measurements). The vector $\boldsymbol{\eta}_i^{[y]}$ is a 2×1 vector of growth factors, where the first element (η_{0i}) signifies the initial status and the second element (η_{1i}) indicates the growth rate within a specified time interval. The $J \times 2$ matrix $\mathbf{A}_i^{[y]}$ consists of associated factor loadings. Finally, $\boldsymbol{\epsilon}_i^{[y]}$ is a $J \times 1$ vector of the i^{th} individual's residuals. Equation (5) expresses $\boldsymbol{\eta}_i^{[y]}$ as deviations ($\boldsymbol{\zeta}_i^{[y]}$) from the mean values of the growth factors ($\boldsymbol{\mu}_\eta^{[y]}$).

While the scaling of η_{1i} affects its interpretation, the general form of the factor loading matrix, $\mathbf{A}_i^{[y]}$, remains consistent. The general form is given as:

$$\mathbf{A}_i^{[y]} = \begin{pmatrix} 1 & 0 \\ 1 & \gamma_{2-1}^{[y]} \times (t_{i2} - t_{i1}) \\ 1 & \sum_{j=2}^3 \gamma_{j-1}^{[y]} \times (t_{ij} - t_{i(j-1)}) \\ \dots & \dots \\ 1 & \sum_{j=2}^J \gamma_{j-1}^{[y]} \times (t_{ij} - t_{i(j-1)}) \end{pmatrix}, \quad (6)$$

where the j^{th} element in the second column represents the cumulative value of the relative rate (i.e., $\gamma^{[y]}$) over time up to time j , so the product of this element and $\eta_{1i}^{[y]}$ represents the change from the initial status of the i^{th} individual ([Liu & Perera, 2024](#)). In addition, the subscript i in $\mathbf{A}_i^{[y]}$ emphasizes that the model accommodates individual measurement occasions.

2.2 Model Specification of Parallel Latent Basis Growth Model

In this section, we extend the univariate Latent Basis Growth Model (LBGM) to its parallel version. This parallel version enables the joint analysis of multiple repeated outcomes, such as the joint development of reading and mathematics ability. The necessity for this extension arises from the various compelling reasons that have been discussed in [Section 1](#). We describe the parallel LBGM in the context of individual measurement occasions, extending the univariate model given in [Equation \(4\)](#). Assume that we have bivariate growth trajectories for repeated outcomes, the parallel LBGM can then be formally defined as follows:

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{z}_i \end{pmatrix} = \begin{pmatrix} \mathbf{A}_i^{[y]} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_i^{[z]} \end{pmatrix} \times \begin{pmatrix} \boldsymbol{\eta}_i^{[y]} \\ \boldsymbol{\eta}_i^{[z]} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_i^{[y]} \\ \boldsymbol{\epsilon}_i^{[z]} \end{pmatrix}, \quad (7)$$

where \mathbf{z}_i is also a $J \times 1$ vector of the repeated measurements for individual i , $\boldsymbol{\eta}_i^{[z]}$, $\boldsymbol{\Lambda}_i^{[z]}$ and $\boldsymbol{\epsilon}_i^{[z]}$ are its growth factors (a 2×1 vector), the corresponding factor loadings (a $J \times 2$ matrix), and the residuals of person i (a $J \times 1$ vector), respectively. Similar to $\boldsymbol{\Lambda}_i^{[y]}$, $\boldsymbol{\Lambda}_i^{[z]}$ has a general expression but with one fixed relative growth rate γ_{j-1} , corresponding to the growth rate of $(j-1)^{th}$ time interval that $\eta_{1i}^{[z]}$ represents. We then write the outcome-specific growth factors $\boldsymbol{\eta}_i^{[u]}$ ($u = y, z$) as deviations from the corresponding outcome-specific growth factor means.

$$\begin{pmatrix} \boldsymbol{\eta}_i^{[y]} \\ \boldsymbol{\eta}_i^{[z]} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_\eta^{[y]} \\ \boldsymbol{\mu}_\eta^{[z]} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\zeta}_i^{[y]} \\ \boldsymbol{\zeta}_i^{[z]} \end{pmatrix}, \quad (8)$$

where $\boldsymbol{\mu}_\eta^{[u]}$ is a 2×1 vector of outcome-specific growth factor means, and $\boldsymbol{\zeta}_i^{[u]}$ is a 2×1 vector of deviations of the i^{th} individual from the means. To simplify model, we assume that $(\boldsymbol{\zeta}_i^{[y]} \boldsymbol{\zeta}_i^{[z]})^T$ follows a multivariate normal distribution

$$\begin{pmatrix} \boldsymbol{\zeta}_i^{[y]} \\ \boldsymbol{\zeta}_i^{[z]} \end{pmatrix} \sim \text{MVN}\left(\mathbf{0}, \begin{pmatrix} \boldsymbol{\Psi}_\eta^{[y]} & \boldsymbol{\Psi}_\eta^{[yz]} \\ & \boldsymbol{\Psi}_\eta^{[z]} \end{pmatrix}\right),$$

where both $\boldsymbol{\Psi}_\eta^{[u]}$ and $\boldsymbol{\Psi}_\eta^{[yz]}$ are 2×2 matrices: $\boldsymbol{\Psi}_\eta^{[u]}$ is the variance-covariance matrix of the outcome-specific growth factors while $\boldsymbol{\Psi}_\eta^{[yz]}$ is the covariances between the growth factors of \mathbf{y}_i and \mathbf{z}_i . To simplify the model, we also assume that the individual outcome-specific residual variances are identical and independent normal distributions over time, while the residual covariances are homogeneous over time, that is,

$$\begin{pmatrix} \boldsymbol{\epsilon}_i^{[y]} \\ \boldsymbol{\epsilon}_i^{[z]} \end{pmatrix} \sim \text{MVN}\left(\mathbf{0}, \begin{pmatrix} \theta_\epsilon^{[y]} \mathbf{I} & \theta_\epsilon^{[yz]} \mathbf{I} \\ & \theta_\epsilon^{[z]} \mathbf{I} \end{pmatrix}\right),$$

where \mathbf{I} is a $J \times J$ identity matrix.

2.3 Model Estimation

We then write the expected mean vector and variance-covariance matrix of the bivariate repeated outcome \mathbf{y}_i and \mathbf{z}_i in the parallel LBGM specified in Equations (7) and (8) as

$$\boldsymbol{\mu}_i = \begin{pmatrix} \boldsymbol{\mu}_i^{[y]} \\ \boldsymbol{\mu}_i^{[z]} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Lambda}_i^{[y]} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_i^{[z]} \end{pmatrix} \times \begin{pmatrix} \boldsymbol{\mu}_\eta^{[y]} \\ \boldsymbol{\mu}_\eta^{[z]} \end{pmatrix} \quad (9)$$

and

$$\begin{aligned}
\Sigma_i &= \begin{pmatrix} \Sigma_i^{[y]} & \Sigma_i^{[yz]} \\ \Sigma_i^{[z]} & \Sigma_i^{[z]} \end{pmatrix} \\
&= \begin{pmatrix} \Lambda_i^{[y]} & \mathbf{0} \\ \mathbf{0} & \Lambda_i^{[z]} \end{pmatrix} \times \begin{pmatrix} \Psi_\eta^{[y]} & \Psi_\eta^{[yz]} \\ \Psi_\eta^{[z]} & \Psi_\eta^{[z]} \end{pmatrix} \times \begin{pmatrix} \Lambda_i^{[y]} & \mathbf{0} \\ \mathbf{0} & \Lambda_i^{[z]} \end{pmatrix}^T \\
&\quad + \begin{pmatrix} \theta_\epsilon^{[y]} \mathbf{I} & \theta_\epsilon^{[yz]} \mathbf{I} \\ \theta_\epsilon^{[z]} \mathbf{I} & \theta_\epsilon^{[z]} \mathbf{I} \end{pmatrix}.
\end{aligned} \tag{10}$$

The parameters in the parallel LBGM specified in Equations (7) and (8) include the mean vector and variance-covariance matrix of the growth factors, the outcome-specific relative growth rate, the variance-covariance matrix of the residuals. Accordingly, we define

$$\begin{aligned}
\Theta &= \{\mu_\eta^{[u]}, \Psi_\eta^{[u]}, \Psi_\eta^{[yz]}, \gamma^{[u]}, \theta_\epsilon^{[u]}, \theta_\epsilon^{[yz]}\} \\
&= \{\mu_{\eta_0}^{[u]}, \mu_{\eta_1}^{[u]}, \psi_{00}^{[u]}, \psi_{01}^{[u]}, \psi_{11}^{[u]}, \psi_{00}^{[yz]}, \psi_{01}^{[yz]}, \psi_{10}^{[yz]}, \psi_{11}^{[yz]}, \gamma_{j-1}^{[u]}, \\
&\quad \theta_\epsilon^{[u]}, \theta_\epsilon^{[yz]}\}, \quad u = y, z \\
j &= \begin{cases} 3, \dots, J & \text{Model specification in Figure 3a} \\ 2, \dots, J-1 & \text{Model specification in Figure 3b} \end{cases}
\end{aligned} \tag{11}$$

to list the parameters that we need to estimated in the proposed model.

We estimate Θ using full information maximum likelihood (FIML) to account for the individual measurement occasions and potential heterogeneity of individual contributions to the likelihood function. In this present study, the proposed model is built using the R package *OpenMx* with CSOLNP optimizer (Boker et al., 2020; Hunter, 2018; Neale et al., 2016; Pritikin, Hunter, & Boker, 2015). We provide *OpenMx* code of the proposed parallel LBGM and a demonstration in the online appendix (https://github.com/Veronica0206/LCSM_projects). We also provide *Mplus* 8 code of the proposed model for researchers who are interested in using *Mplus*.

3 Model Evaluation

We aim to assess the effectiveness of the proposed parallel LBGM by employing Monte Carlo simulation studies. Specifically, we examine the model's performance using several metrics: the relative bias, the empirical standard error (SE), the relative root-mean-square error (RMSE), and the empirical coverage probability for a nominal 95% confidence interval for each parameter. These metrics are commonly used in simulation studies to evaluate the performance of statistical methodologies or models. The definitions and estimates for these metrics are presented in Table 1.

Following practices in simulation studies as suggested by Morris, White, and Crowther (2019), we empirically determined the number of replications

Table 1: Performance Metrics: Definitions and Estimates

| Criteria | Definition | Estimate |
|----------------|--|--|
| Relative Bias | $E_{\hat{\theta}}(\hat{\theta} - \theta)/\theta$ | $\sum_{s=1}^S (\hat{\theta}_s - \theta)/S\theta$ |
| Empirical SE | $\sqrt{\text{Var}(\hat{\theta})}$ | $\sqrt{\sum_{s=1}^S (\hat{\theta}_s - \bar{\theta})^2/(S-1)}$ |
| Relative RMSE | $\sqrt{E_{\hat{\theta}}(\hat{\theta} - \theta)^2/\theta}$ | $\sqrt{\sum_{s=1}^S (\hat{\theta}_s - \theta)^2/S/\theta}$ |
| Coverage Prob. | $Pr(\hat{\theta}_{\text{lower}} \leq \theta \leq \hat{\theta}_{\text{upper}})$ | $\sum_{s=1}^S I(\hat{\theta}_{\text{lower},s} \leq \theta \leq \hat{\theta}_{\text{upper},s})/S$ |

^a θ : the population value of the parameter of interest^b $\hat{\theta}$: the estimate of θ ^c S : the number of replications and set as 1,000 in our simulation study^d $s = 1, \dots, S$: indexes the replications of the simulation^e $\hat{\theta}_s$: the estimate of θ from the s^{th} replication^f $\bar{\theta}$: the mean of $\hat{\theta}_s$'s across replications^g $I(\cdot)$: an indicator function^h Coverage Prob.: coverage probability

to be $S = 1,000$. The pilot simulation study was conducted to ensure that the chosen number of replications would provide reliable performance metrics. Among the four performance metrics, the (relative) bias is of utmost importance. The pilot simulation revealed that the standard errors of bias, calculated as Monte Carlo $\text{SE}(\text{Bias}) = \sqrt{\frac{\text{Var}(\hat{\theta})}{S}}$, were less than 0.15 across all parameters, except for $\psi_{00}^{[u]}$ and $\psi_{00}^{[yz]}$. To maintain the Monte Carlo standard error of bias below 0.05, at least 900 replications are needed. Thus, we decided to proceed with $S = 1,000$ replications to account for variability and ensure a more robust evaluation.

3.1 Design of Simulation Study

To thoroughly evaluate the proposed parallel LBGM, we designed a comprehensive set of simulation studies, the conditions of which are outlined in Table 2. A key factor in the effectiveness of a model designed for longitudinal data is the number of repeated measures. We hypothesize that the proposed model's performance will improve with an increasing number of repeated measurements. To test this hypothesis, we considered two levels for the number of repeated measures: six and ten. For conditions with ten repeated measures, we investigated whether equally-placed study waves or unequally-placed waves affect model performance, assuming that the study duration remains constant across conditions. This consideration reflects real-world longitudinal study practices, where measurement waves are typically not equally spaced, often occurring more frequently at the beginning. We aimed to determine if such setups impact model performance. In scenarios with six repeated measures, we examined the model's performance under the more challenging condition of a shorter study duration with the hypothesis that a shorter duration poses greater challenges for the

model due to less available data to accurately capture the underlying growth trajectory. Measurement occasions are individuated by a ‘medium’-width time window, $(-0.25, +0.25)$ around each wave (Coulombe et al., 2015).

Another key variable of interest is the correlation between the two trajectories, as the proposed model is designed for analyzing joint longitudinal processes. Three correlation levels for the between-construct growth factors are considered: ± 0.3 and 0. We are interested in how model over-specification affects performance in zero-correlation conditions, and whether the sign of the correlation (± 0.3) has any impact on model performance. Additionally, we explore the influence of varying trajectory shapes, quantified by the relative growth rate in each time interval. As specified in Table 2, the change patterns considered include both increasing and decreasing growth rates. Moreover, we evaluate the model’s performance across different sample sizes ($n = 200$ and $n = 500$) and levels of outcome-specific residual variances ($\theta_\epsilon^{[u]} = 1$ or $\theta_\epsilon^{[u]} = 2$) to gauge the effects of sample size and measurement precision. In the simulation design, factors considered less critical to the proposed model’s performance, such as the distribution of growth factors and the correlation of between-construct residuals, were held constant.

3.2 Data Generation and Simulation Step

To evaluate the performance of the proposed parallel LBGMs, we conducted a simulation study according to the design presented in Table 2. Each condition was replicated 1,000 times to ensure a robust assessment. The steps for the simulation are outlined as follows:

1. **Growth Factor Generation:** Utilizing the *MASS* R package (Venables & Ripley, 2002), generate the growth factors for both longitudinal processes based on the pre-defined mean vector and variance-covariance matrix as specified in Table 2. The *MASS* package is used for its reliability in generating multivariate Gaussian samples.
2. **Time Structure:** Generate the time structure with J waves t_j as defined in Table 2. Add a uniform disturbance following $U(t_j - \Delta, t_j + \Delta)$ around each wave to obtain individual measurement occasions t_{ij} .
3. **Factor Loadings Calculation:** Compute the factor loadings for each individual of each longitudinal process as Equation 6, using the relative growth rates and individual measurement intervals.
4. **Measurement Value Computation:** Calculate the values of bivariate repeated measurements, incorporating growth factors, factor loadings, and the pre-defined residual variance-covariance structure.
5. **LBGM Implementation:** Execute the proposed LBGM models on the generated dataset, estimating the model parameters and constructing 95% Wald confidence intervals.
6. **Replication:** Repeat steps 1-5 until 1,000 convergent solutions are obtained, as this number of replications provides a stable estimate of performance metrics such as bias and coverage probability.

Table 2: Simulation Design for Parallel Latent Basis Growth Model in the Framework of Individual Measurement Occasions

| Fixed Conditions | |
|---|---|
| Variables | Conditions |
| Distribution of the Intercept | $\eta_{0i}^{[y]} \sim N(50, 5^2); \eta_{0i}^{[z]} \sim N(30, 5^2)$ (i.e., $\mu_{\eta 0}^{[y]} = 50, \mu_{\eta 0}^{[z]} = 25; \psi_{00}^{[y]} = 30, \psi_{00}^{[z]} = 25$) |
| Distribution of the Shape Factor | $\eta_{1i}^{[y]} \sim N(4, 1^2); \eta_{1i}^{[z]} \sim N(5, 1^2)$ (i.e., $\mu_{\eta 1}^{[y]} = 4, \mu_{\eta 1}^{[z]} = 5; \psi_{11}^{[y]} = 1, \psi_{11}^{[z]} = 1$) |
| Correlations of Within-construct GFs | $\rho^{[u]} = 0.3$ ($u = y, z$) |
| Correlation between Residuals | $\rho_e = 0.3$ |
| Manipulated Conditions (Full Factorial) | |
| Variables | Conditions |
| Sample Size | $n = 200, 500$ |
| Time (t_j) | 6 equally-spaced: $t_j = 0, 1.00, 2.00, 3.00, 4.00, 5.00$ 10 equally-spaced: $t_j = 0, 1.00, 2.00, 3.00, 4.00, 5.00, 6.00, 7.00, 8.00, 9.00$ 10 unequally-spaced: $t_j = 0, 0.75, 1.50, 2.55, 3.00, 3.75, 4.50, 6.00, 7.50, 9.00$ |
| Individual t_{ij} | $t_{ij} \sim U(t_j - \Delta, t_j + \Delta)$ ($\Delta = 0.25$) |
| Relative Growth Rate in Each Time Interval ^a | 6 waves: $r^{[u]} = 1.0, 0.8, 0.6, 0.4, 0.2$ ($u = y, z$) |
| | 6 waves: $r^{[u]} = 0.2, 0.4, 0.6, 0.8, 1.0$ ($u = y, z$) |
| | 10 waves: $r^{[u]} = 1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2$ ($u = y, z$) |
| | 10 waves: $r^{[u]} = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$ ($u = y, z$) |
| Correlation of Between-construct GFs | $\rho = 0, \pm 0.3$ |
| Residual Variance | $\theta_e^{[u]} = 1, 2$ ($u = y, z$) |

^a Growth rate is the relative growth rate, which is defined as the absolute growth rate over the value of shape factor.

4 Result

4.1 Model Convergence

Before assessing the four performance measures of the proposed parallel LBGM, we first examined its convergence rate¹. The model exhibited excellent convergence, as evidenced by a 100% rate across all simulation conditions listed in Table 2.

4.2 Performance Measures

This section summarizes the simulation results for four key performance metrics: relative bias, empirical SE, relative RMSE, and empirical coverage probability for a nominal 95% confidence interval. We calculated these metrics for each parameter across 1,000 repetitions under each condition, and summarized the median and range values for all conditions given the scale of parameters and simulation setups. The proposed model generally yielded unbiased and accurate point estimates with target coverage probabilities. Further details for each performance metric are provided in the Online Supplementary Document.

The proposed model produced unbiased and accurate point estimates. Specifically, the magnitudes of the relative biases for outcome-specific growth factor means, variances, and relative growth rates were below 0.004, 0.013, and 0.012, respectively². The magnitudes of the relative RMSEs for outcome-specific growth factor means, variances, and relative growth rates were below 0.05, 0.15, and 0.23, respectively³. Moreover, the model demonstrated excellent empirical coverage probabilities, with median values approximating 0.95. Given these consistently strong performance metrics, further investigations into the effect of different simulation conditions were deemed unnecessary.

5 Application

In this section, we demonstrate how to employ the proposed parallel LBGM to analyze real-world data. This application section includes two examples. In the first example, we illustrate the recommended steps to construct the proposed model in practice. In the second example, we demonstrate how to apply the proposed model to analyze joint longitudinal processes with a more complicated

¹ In this study, we define convergence rate as the achievement of an *OpenMx* status code of 0, indicating successful optimization, in up to 10 runs with varied initial values (Neale et al., 2016).

² Previous simulations have often regarded relative bias in regression coefficients as acceptable if it was below 10%, which is commonly considered a guideline when assessing relative bias (Leite, 2017; Poon & Wang, 2010).

³ Regarding relative RMSE, while there is no universally accepted benchmark for simulation studies, model accuracy is generally considered excellent when the score is below 10%, good when it ranges from 10% to 20%, fair when it falls between 20% and 30%, and poor when it exceeds 30% (Jadon, Patil, & Jadon, 2022).

data structure, where two repeated outcomes have different time frames. We randomly selected 400 students from the Early Childhood Longitudinal Study Kindergarten Cohort of 2010-2011 (ECLS-K: 2011), all of whom had complete records of repeated reading and mathematics scores based on Item Response Theory (IRT), as well as their age in months at each wave⁴.

ECLS-K: 2011 is a national longitudinal study of US children enrolled in around 900 kindergarten programs beginning in the 2010-2011 school year. In ECLS-K: 2011, children's reading and mathematics abilities were assessed in nine waves: fall and spring of kindergarten (2010-2011), first (2011-2012) and second (2012-2013) grade, respectively, as well as spring of the 3rd (2014), 4th (2015), and 5th (2016) grade. Only about 30% of students were assessed in the fall semesters of 2011 and 2012 (Lê, Norman, Tourangeau, Brick, & Mulligan, 2011). In the first example, we used all nine waves of reading and mathematics IRT scores to demonstrate how to apply the proposed model. In the second example, we utilized all nine waves of reading IRT scores but only the mathematics scores obtained in the spring semesters to mimic one possible complex time structure in practice. Note that the initial status and the number of measurement occasions of the two abilities are different in the second example. Additionally, we employed children's age in months rather than their grade-in-school to have individual measurement occasions. The subsample included 41.50% White, 7.25% Black, 37.00% Latinx, 8.25% Asian, and 6.00% of other ethnicity.

5.1 Analyze Joint Longitudinal Records with The Same Time Structure

Following Blozis et al. (2008) and Liu and Perera (2022), we first constructed a latent growth curve model to examine each longitudinal process in isolation before analyzing joint development. Specifically, we employed a LBGM to explore the univariate development of either reading or mathematics from Grade K to 5. Figure 4 illustrates the model-implied curves superimposed on the smooth lines for each ability. For each ability, the estimates from the LBGM produced model-implied trajectories that closely align with the smooth lines representing the observed individual data.

We then applied the proposed parallel LBGM to analyze the joint development of reading and mathematics abilities. Figure 5 illustrates the model-implied curves superimposed on the smooth lines for each ability obtained from the parallel model. From the figure, it is evident that the model-implied curves of the parallel models did not differ from those of the univariate growth models shown in Figure 4. Table 3 presents the parameter estimates of interest for joint development.

Note that we defined $\eta_1^{[u]}$ as the growth rate in the final time interval for each ability's longitudinal process (i.e., the model specification in Figure 3b).

⁴ The total sample size of ECLS-K: 2011 is $n = 18174$. The number of rows after removing records with missing values (i.e., entries with any of NaN/-9/-8/-7/-1) is $n = 2290$.

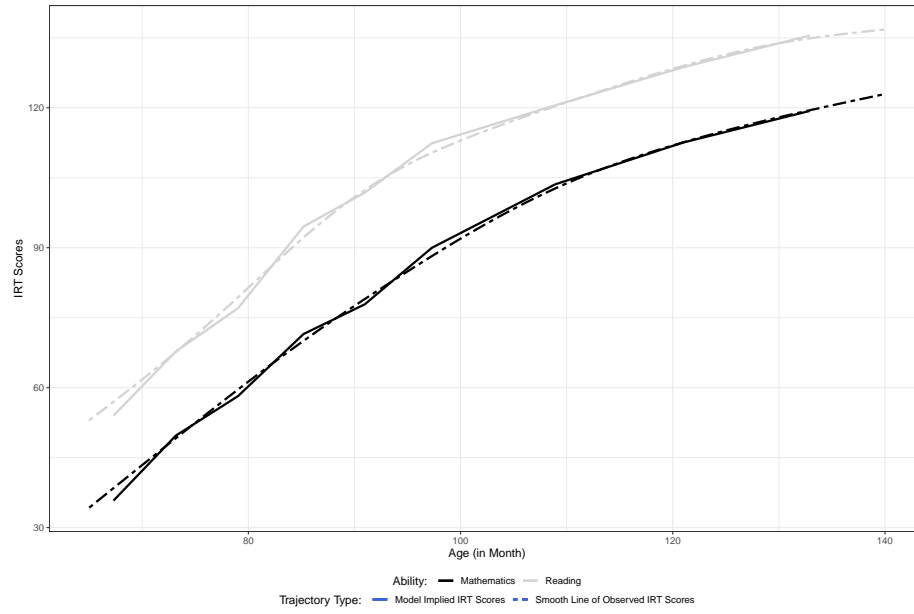


Figure 4: Model Implied Trajectory and Smooth Line of Univariate Development

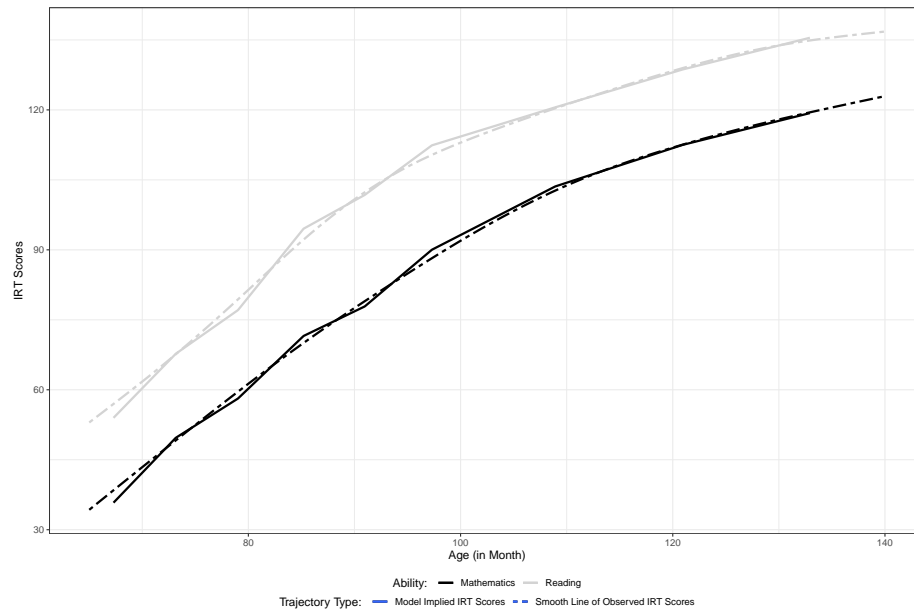


Figure 5: Model Implied Trajectory and Smooth Line of Bivariate Development with The Same Time Structures

Table 3: Estimates of Parallel Latent Basis Growth Model for Reading and Mathematics Ability with the Same Time Structures

| Mean | Reading IRT Scores | | | Math IRT Scores | | | Covariance | | |
|---------------------------------|--------------------|-----------------------|--|------------------|-----------------------|--|------------------|-----------------------|--|
| | Estimate (SE) | P value | | Estimate (SE) | P value | | Estimate (SE) | P value | |
| Initial Status ^a | 53.984 (0.724) | < 0.0001 ^b | | 35.797 (0.659) | < 0.0001 [*] | | — ^c | — | |
| Rate ^d of Interval 1 | 2.341 (0.080) | < 0.0001 [*] | | 2.369 (0.071) | < 0.0001 [*] | | — | — | |
| Rate of Interval 2 | 1.596 (0.077) | < 0.0001 [*] | | 1.437 (0.068) | < 0.0001 [*] | | — | — | |
| Rate of Interval 3 | 2.823 (0.077) | < 0.0001 [*] | | 2.169 (0.066) | < 0.0001 [*] | | — | — | |
| Rate of Interval 4 | 1.256 (0.080) | < 0.0001 [*] | | 1.098 (0.070) | < 0.0001 [*] | | — | — | |
| Rate of Interval 5 | 1.679 (0.074) | < 0.0001 [*] | | 1.920 (0.066) | < 0.0001 [*] | | — | — | |
| Rate of Interval 6 | 0.701 (0.041) | < 0.0001 [*] | | 1.167 (0.036) | < 0.0001 [*] | | — | — | |
| Rate of Interval 7 | 0.676 (0.039) | < 0.0001 [*] | | 0.742 (0.034) | < 0.0001 [*] | | — | — | |
| Rate of Interval 8 | 0.566 (0.040) | < 0.0001 [*] | | 0.568 (0.034) | < 0.0001 [*] | | — | — | |
| Variance | | | | | | | | | |
| Initial Status | 164.926 (13.085) | < 0.0001 [*] | | 139.878 (10.876) | < 0.0001 [*] | | 126.794 (10.650) | < 0.0001 [*] | |
| Rate of Interval 1 | 0.111 (0.013) | < 0.0001 [*] | | 0.106 (0.011) | < 0.0001 [*] | | 0.064 (0.009) | < 0.0001 [*] | |
| Rate of Interval 2 | 0.051 (0.007) | < 0.0001 [*] | | 0.039 (0.005) | < 0.0001 [*] | | 0.026 (0.004) | < 0.0001 [*] | |
| Rate of Interval 3 | 0.161 (0.018) | < 0.0001 [*] | | 0.089 (0.010) | < 0.0001 [*] | | 0.070 (0.009) | < 0.0001 [*] | |
| Rate of Interval 4 | 0.032 (0.005) | < 0.0001 [*] | | 0.023 (0.003) | < 0.0001 [*] | | 0.016 (0.002) | < 0.0001 [*] | |
| Rate of Interval 5 | 0.057 (0.007) | < 0.0001 [*] | | 0.070 (0.008) | < 0.0001 [*] | | 0.037 (0.005) | < 0.0001 [*] | |
| Rate of Interval 6 | 0.010 (0.001) | < 0.0001 [*] | | 0.026 (0.003) | < 0.0001 [*] | | 0.009 (0.001) | < 0.0001 [*] | |
| Rate of Interval 7 | 0.009 (0.001) | < 0.0001 [*] | | 0.010 (0.001) | < 0.0001 [*] | | 0.006 (0.001) | < 0.0001 [*] | |
| Rate of Interval 8 | 0.006 (0.001) | < 0.0001 [*] | | 0.006 (0.001) | < 0.0001 [*] | | 0.004 (0.001) | < 0.0001 [*] | |

^a The initial Status was defined as 60 months old in this case.^b * indicates statistical significance at 0.05 level.^c — indicates that the metric was not available in the model.^d The mean, variance, and covariance of rate in each interval were the corresponding value of absolute growth rate, which can be obtained by *R* function *mzAlgebra()* from estimated shape factor and relative growth rate.

Consequently, in Table 3, the parameters related to 'initial status' and 'rate of Interval 8' were directly estimated from the proposed model, while others were obtained using the function *mxAlgebra()*⁵ in the *R* package *OpenMx*. It is important to note that the estimated initial status and interval-specific slopes remain unaffected if $\eta_1^{[u]}$ is defined as the growth rate in the first time interval for each ability's longitudinal process (i.e., the model specification in Figure 3a). This difference in specification only affects the interpretation of $\eta_1^{[u]}$. Specifically, when $\eta_1^{[u]}$ is scaled to be the slope in the first time interval, the correlation of the two $\eta_1^{[u]}$ for the two outcomes indicates how the growth rates are related in the first interval. In contrast, when $\eta_1^{[u]}$ is scaled to be the slope in the last time interval, the correlation reflects how the growth rates are related during the final interval.

From Figure 5 and Table 3, we observed that the development of both reading and mathematics abilities generally slowed down after Grade 3, which aligns with earlier studies (Liu & Perera, 2022; Peralta et al., 2022). Additionally, there was a positive association between the development of reading and mathematics abilities, indicated by statistically significant intercept-intercept and slope-slope covariances in each time interval. After standardizing the covariances, the intercept-intercept correlation and each interval-specific slope-slope correlation were found to be 0.83 and 0.58, respectively. This suggests that, on average, a child who performed better in reading tests at Grade K also tended to perform better in mathematics examinations, and vice versa. Moreover, on average, children who showed more rapid gains in reading ability also tended to exhibit faster improvement in mathematics, and vice versa.

5.2 Analyze Joint Longitudinal Records with Different Time Structures

In this section, we use the proposed parallel LBGM to investigate the joint development trajectories of reading and mathematics abilities. We retained all nine measurement occasions for reading ability but included only the spring semester measurements for mathematics ability (i.e., Waves 2, 4, 6, 7, 8, and 9). In this configuration, both the initial statuses and the number of measurement occasions differ between the two abilities. Figure 6 illustrates the model-implied curves superimposed on the smooth lines representing each ability in this model. The figure reveals that the model-implied trajectories vary only minimally from those presented in Figure 5 due to fewer measurement occasions for mathematics ability, but they still sufficiently capture the smooth lines of the observed individual data.

Table 4 presents the estimated parameters of interest for the joint model with differing time structures. Note that there are 8 time intervals for the development of reading ability (corresponding to 9 measurement occasions) but only

⁵ By using *mxAlgebra()*, we specify algebraic expressions for new parameters, enabling *OpenMx* to estimate their point values along with standard errors.

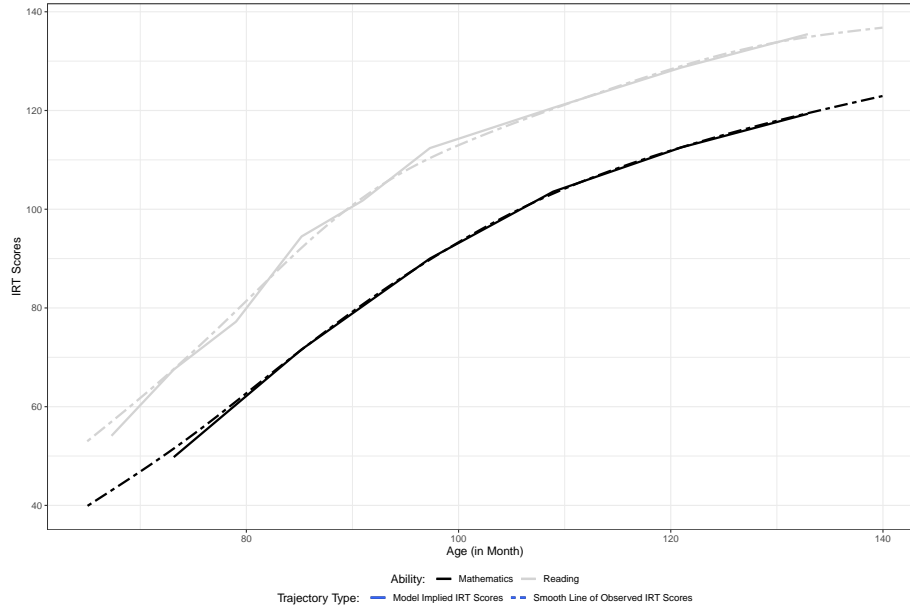


Figure 6: Model Implied Trajectory and Smooth Line of Bivariate Development with Different Time Structures

5 time intervals for the development of mathematics ability because three measurements from the fall semesters were excluded. During the first time interval for mathematics, which corresponds to Intervals 2 and 3 for reading ability (as detailed in Table 3), the estimated growth rate was 1.811. This value represents an average of the growth rates 1.437 and 2.169 from Interval 2 and Interval 3, respectively, in Table 3. These findings suggest that our proposed model effectively captures the underlying patterns of growth trajectories, even with fewer measurements.

Table 4: Estimates of Parallel Latent Basis Growth Model for Reading and Mathematics Ability with Different Time Structures^a

| Mean | Reading IRT Scores | | Math IRT Scores | | Covariance | |
|--|--------------------|-----------------------|-----------------|-----------------------|------------------|-----------------------|
| | Estimate (SE) | P value | Estimate (SE) | P value | Estimate (SE) | P value |
| Initial Status ^b | 54.100 (0.724) | < 0.0001 ^c | 49.782 (0.694) | < 0.0001 [*] | — ^d | — |
| Rate ^e of Interval ^f 1 | 2.288 (0.080) | < 0.0001 [*] | 1.811 (0.037) | < 0.0001 [*] | — | — |
| Rate of Interval 2 | 1.653 (0.077) | < 0.0001 [*] | 1.811 (0.037) | < 0.0001 [*] | — | — |
| Rate of Interval 3 | 2.793 (0.077) | < 0.0001 [*] | 1.523 (0.036) | < 0.0001 [*] | — | — |
| Rate of Interval 4 | 1.263 (0.080) | < 0.0001 [*] | 1.523 (0.036) | < 0.0001 [*] | — | — |
| Rate of Interval 5 | 1.679 (0.074) | < 0.0001 [*] | 1.523 (0.036) | < 0.0001 [*] | — | — |
| Rate of Interval 6 | 0.702 (0.041) | < 0.0001 [*] | 1.167 (0.037) | < 0.0001 [*] | — | — |
| Rate of Interval 7 | 0.676 (0.039) | < 0.0001 [*] | 0.742 (0.035) | < 0.0001 [*] | — | — |
| Rate of Interval 8 | 0.567 (0.039) | < 0.0001 [*] | 0.572 (0.035) | < 0.0001 [*] | — | — |
| Variance | Estimate (SE) | P value | Estimate (SE) | P value | Estimate (SE) | P value |
| Initial Status | 164.755 (13.095) | < 0.0001 [*] | 156.78 (12.804) | < 0.0001 [*] | 130.957 (11.353) | < 0.0001 [*] |
| Rate of Interval 1 | 0.105 (0.012) | < 0.0001 [*] | — | — | — | — |
| Rate of Interval 2 | 0.055 (0.007) | < 0.0001 [*] | 0.060 (0.007) | < 0.0001 [*] | 0.032 (0.005) | < 0.0001 [*] |
| Rate of Interval 3 | 0.157 (0.018) | < 0.0001 [*] | 0.060 (0.007) | < 0.0001 [*] | 0.055 (0.008) | < 0.0001 [*] |
| Rate of Interval 4 | 0.032 (0.005) | < 0.0001 [*] | 0.042 (0.005) | < 0.0001 [*] | 0.021 (0.003) | < 0.0001 [*] |
| Rate of Interval 5 | 0.057 (0.007) | < 0.0001 [*] | 0.042 (0.005) | < 0.0001 [*] | 0.028 (0.004) | < 0.0001 [*] |
| Rate of Interval 6 | 0.010 (0.001) | < 0.0001 [*] | 0.025 (0.003) | < 0.0001 [*] | 0.009 (0.001) | < 0.0001 [*] |
| Rate of Interval 7 | 0.009 (0.001) | < 0.0001 [*] | 0.010 (0.001) | < 0.0001 [*] | 0.005 (0.001) | < 0.0001 [*] |
| Rate of Interval 8 | 0.006 (0.001) | < 0.0001 [*] | 0.006 (0.001) | < 0.0001 [*] | 0.004 (0.001) | < 0.0001 [*] |

^a In this joint model, we included the measurements of reading ability at all nine waves, but we only included the measures of mathematics ability at Wave 2, 4, 6, 7, 8, and 9.

^b The initial Status of reading ability was defined as the measurement at 60 months old, while that of mathematics ability was the measurement half a year later.

^c * indicates statistical significance at 0.05 level.

^d — indicates that the metric was not available in the model.

^e The mean, variance, and covariance of rate in each interval were the corresponding value of absolute growth rate, which can be obtained by R function *mxAlgebra()* from estimated shape factor and relative growth rate.

^f Each ‘interval’ was defined as the interval between any two consecutive measurement occasions of reading ability. The estimates of mathematics ability during the first interval are not applicable because the first measure of mathematics ability was Wave 2. The estimated means and variances of mathematics ability in Interval 2 (Interval 4) and Interval 3 (Interval 5) were the same because we took out its measurement at Wave 3 (Wave 5).

6 Discussion

This article extends the latent basis growth model with the novel specification proposed by Liu and Perera (2024) to explore joint nonlinear longitudinal processes in the framework of individual measurement occasions. This framework is particularly advantageous when investigating parallel development because it helps avoid inadmissible estimation and allows for different time structures across outcomes. Additionally, the proposed model allows scaling the second growth factor as the growth rate during any time interval. In the present study, we specify the second growth factor as the growth rate during either the first or last time interval and estimate the relative rates for each of the other intervals for each repeated outcome. We demonstrate that the proposed parallel LBGM can provide unbiased and accurate point estimates with target coverage probabilities through simulation studies. Additionally, we apply the proposed model to analyze the joint development of reading and mathematics abilities, using the same or different time structures. Our analysis relies on a subsample of $n = 400$ from ECLS-K: 2011.

6.1 Practical Considerations

In this section, we provide recommendations for empirical researchers based on both the simulation study and real-world data analyses. First, although we scale the shape factor η_1 as the growth rate in the first or last time interval of the study duration, it can be specified as the growth rate in any time interval. Note that the interpretation of γ_{j-1} remains as the relative growth rate to η_1 during the $(j - 1)^{th}$ time interval. From the proposed parallel LBGM, we obtain the estimates of the mean and variance of shape factor and the fixed effects of relative growth rates for each construct. Using the `mxAlgebra()` function from the `OpenMx` R package, we derive both fixed and random effects for the absolute growth rate of each time interval, as detailed in the Application section.

In addition, the proposed model is capable of estimating the covariance of between-construct intercepts and that of between-construct shape factors directly. We can derive the covariance of between-construct growth rates for each interval by using the function `mxAlgebra()`. Note that the correlation of the between-construct growth rates is constant because we only estimate fixed effects of relative growth rates.

Third, as the latent basis growth model serves primarily as an exploratory tool, allowing trajectory characteristics to emerge from the data rather than being specified *a priori*, researchers may also be interested in exploring other aspects, such as the change-from-baseline values at each measurement wave for each repeated outcome. We can also derive these features with the function `mxAlgebra()`. In the online appendix (https://github.com/Veronica0206/LCSM_projects), we also provide code to demonstrate how to derive the values of change-from-baseline.

6.2 Methodological Considerations and Future Directions

There are several directions to consider for future studies. First, similar to the standard implementation of latent basis growth models, the proposed model requires a strict proportionality assumption (McNeish, 2020; Wu & Lang, 2016). Wu and Lang (2016) showed that this assumption might potentially result in biased estimates by simulation studies. McNeish (2020) demonstrated that this assumption could be relaxed by specifying random factor loadings of the shape factor. In the same way, we can also relax the proportionality assumption for the proposed parallel LBGM. Note that the extended model, where both the shape factor and relative growth rates are random coefficients, cannot be specified in a frequentist SEM software because these random coefficients enter the model in a multiplicative fashion (i.e., a nonlinear fashion). Similar to McNeish (2020), the extended model can be constructed in Bayesian software such as *jags* or *stan*.

Second, it is not our intention to show that the proposed parallel LBGM is better than any other parallel growth models with parametric or semi-parametric functional forms. The proposed model is a versatile tool for exploratory analyses; it should perform well to detect the trends of trajectories or whether a spike exists over the study duration. However, the insights directly related to research questions might be limited. Accordingly, subsequent analyses may need to be based on the estimates generated by the proposed model. For instance, if we obtain evidence suggesting that developmental processes can generally be divided into two stages, we may employ the parallel bilinear spline growth model (Liu & Perera, 2022) to further estimate the individual transition time to the stage with a slower growth rate. Alternatively, we can constrain the relative growth rates of multiple time intervals to be the same to have a more parsimonious model. Therefore, statistical methods for comparing the full model to a more parsimonious one need to be proposed and tested.

Third, as in any latent growth curve model, baseline covariates can be added to predict the intercept or the growth rate. Additionally, a time-varying covariate can also be added to estimate its effect on the measurements while simultaneously modeling parallel change patterns in these measurements.

6.3 Concluding Remarks

In this article, we propose a novel expression of latent basis growth models to allow for individual measurement occasions and further extend the model to analyze joint longitudinal processes. The results of both the simulation studies and real-world data analyses underscore the model’s valuable capabilities for exploring parallel nonlinear change patterns. As discussed above, the proposed method offers avenues for both practical extensions and further methodological examination.

References

- Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics*, 28(2), 135–167. doi: <https://doi.org/10.3102/10769986028002135>
- Blozis, S. A. (2004). Structured latent curve models for the study of change in multivariate repeated measures. *Psychological Methods*, 9(3), 334–353. doi: <https://doi.org/10.1037/1082-989x.9.3.334>
- Blozis, S. A., & Cho, Y. (2008). Coding and centering of time in latent curve models in the presence of interindividual time heterogeneity. *Structural Equation Modeling: A Multidisciplinary Journal*, 15(3), 413–433. doi: <https://doi.org/10.1080/10705510802154299>
- Blozis, S. A., Harring, J. R., & Mels, G. (2008). Using lisrel to fit nonlinear latent curve models. *Structural Equation Modeling: A Multidisciplinary Journal*, 15(2), 346–369. doi: <https://doi.org/10.1080/10705510801922639>
- Boker, S. M., Neale, M. C., Maes, H. H., Wilde, M. J., Spiegel, M., Brick, T. R., ... Kirkpatrick, R. M. (2020). Openmx 2.17.2 user guide [Computer software manual]. Retrieved from <https://vipbg.vcu.edu/vipbg/OpenMx2/docs/OpenMx/2.17.2/OpenMxUserGuide.pdf>
- Coulombe, P., Selig, J. P., & Delaney, H. D. (2015). Ignoring individual differences in times of assessment in growth curve modeling. *International Journal of Behavioral Development*, 40(1), 76–86. doi: <https://doi.org/10.1177/0165025415577684>
- Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, 38(4), 529–569. doi: https://doi.org/10.1207/s15327906mbr3804_5
- Dumenci, L., Perera, R. A., Keefe, F. J., Ang, D. C., J., S., Jensen, M. P., & Riddle, D. L. (2019). Model-based pain and function outcome trajectory types for patients undergoing knee arthroplasty: A secondary analysis from a randomized clinical trial. *Osteoarthritis and cartilage*, 27(6), 878–884. doi: <https://doi.org/10.1016/j.joca.2019.01.004>
- Duncan, S. C., & Duncan, T. E. (1994). Modeling incomplete longitudinal substance use data using latent variable growth curve methodology. *Multivariate behavioral research*, 29(4), 313–338. doi: https://doi.org/10.1207/s15327906mbr2904_1
- Duncan, S. C., & Duncan, T. E. (1996). A multivariate latent growth curve analysis of adolescent substance use. *Structural Equation Modeling: A Multidisciplinary Journal*, 3(4), 323–347. doi: <https://doi.org/10.1080/10705519609540050>
- Grimm, K. J., Ram, N., & Estabrook, R. (2016). *Growth modeling: Structural equation and multilevel modeling approaches*. Guilford Press.
- Grimm, K. J., Steele, J. S., Ram, N., & Nesselroade, J. R. (2013). Exploratory latent growth models in the structural equation modeling framework. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(4), 568–591. doi: <https://doi.org/10.1080/10705511.2013.824775>

- Hunter, M. D. (2018). State space modeling in an open source, modular, structural equation modeling environment. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(2), 307-324. doi: <https://doi.org/10.1080/10705511.2017.1369354>
- Jadon, A., Patil, A., & Jadon, S. (2022). A comprehensive survey of regression based loss functions for time series forecasting. doi: https://doi.org/10.1007/978-981-97-3245-6_9
- Lê, T., Norman, G., Tourangeau, K., Brick, J. M., & Mulligan, G. (2011). Early childhood longitudinal study: Kindergarten class of 2010-2011 - sample design issues. *JSM Proceedings*, 1629-1639.
- Leite, W. (2017). *Practical propensity score methods using r*. Thousand Oaks, CA: Sage Publications. doi: <https://doi.org/10.4135/9781071802854>
- Liu, J., & Perera, R. A. (2022). Estimating knots and their association in parallel bilinear spline growth curve models in the framework of individual measurement occasions. *Psychological Methods*, 27(5), 703-729. doi: <https://doi.org/10.1037/met0000309>
- Liu, J., & Perera, R. A. (2023). Extending growth mixture model to assess heterogeneity in joint development with piecewise linear trajectories in the framework of individual measurement occasions. *Psychological Methods*, 28(5), 1029-1051. doi: <https://doi.org/10.1037/met0000500>
- Liu, J., & Perera, R. A. (2024). Estimating rate of change for nonlinear trajectories in the framework of individual measurement occasions: A new perspective on growth curves. *Behavior Research Methods*, 56, 1349-1375. doi: <https://doi.org/10.3758/s13428-023-02097-2>
- Lyons, M. J., Panizzon, M. S., Liu, W., McKenzie, R., Bluestone, N. J., Grant, M. D., . . . Xian, H. (2017). A longitudinal twin study of general cognitive ability over four decades. *Developmental psychology*, 53(6), 1170-1177. doi: <https://doi.org/10.1037/dev0000303>
- McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. Nesselroade & R. Cattell (Eds.), *Handbook of multivariate experimental psychology* (p. 561-614). Boston, MA: Springer. doi: https://doi.org/10.1007/978-1-4613-0893-5_17
- McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, 58(1), 110-133. doi: <https://doi.org/10.2307/1130295>
- McNeish, D. (2020). Relaxing the proportionality assumption in latent basis models for nonlinear growth. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(5), 817-824. doi: <https://doi.org/10.1080/10705511.2019.1696201>
- McNulty, J. K., Wenner, C. A., & Fisher, T. D. (2016). Longitudinal associations among relationship satisfaction, sexual satisfaction, and frequency of sex in early marriage. *Archives of sexual behavior*, 45(1), 85-97. doi: <https://doi.org/10.1007/s10508-014-0444-6>
- Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, 10(3), 259-284. doi:

- <https://doi.org/10.1037/1082-989x.10.3.259>
- Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods*, 5(1), 23-43. doi: <https://doi.org/10.1037/1082-989x.5.1.23>
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55, 107-122. doi: <https://doi.org/10.1007/bf02294746>
- Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11), 2074-2102. doi: <https://doi.org/10.1002/sim.8086>
- Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., ... Boker, S. M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, 81(2), 535-549. doi: <https://doi.org/10.1007/s11336-014-9435-8>
- Peralta, Y., Kohli, N., Lock, E. F., & Davison, M. L. (2022). Bayesian modeling of associations in bivariate piecewise linear mixed-effects models. *Psychological Methods (Advance online publication)*, 27(1), 44-64. doi: <https://doi.org/10.1037/met0000358>
- Poon, W.-Y., & Wang, H.-B. (2010). Analysis of a two-level structural equation model with missing data. *Sociological Methods & Research*, 39(1), 25-55. doi: <https://doi.org/10.1177/0049124110371312>
- Pritikin, J. N., Hunter, M. D., & Boker, S. M. (2015). Modular open-source software for Item Factor Analysis. *Educational and Psychological Measurement*, 75(3), 458-474. doi: <https://doi.org/10.1177/0013164414554615>
- Rajmil, L., López, A. R., López-Aguilà, S., & Alonso, J. (2013). Parent-child agreement on health-related quality of life (hrqol): a longitudinal study. *Health Qual Life Outcomes*, 11(101). doi: <https://doi.org/10.1186/1477-7525-11-101>
- Robitaille, A., Muniz, G., Piccinin, A. M., Johansson, B., & Hofer, S. M. (2012). Multivariate longitudinal modeling of cognitive aging: Associations among change and variation in processing speed and visuospatial ability. *GeroPsych*, 25(1), 15-24. doi: <https://doi.org/10.1024/1662-9647/a000051>
- Shin, T., Davison, M. L., Long, J. D., Chan, C., & Heistad, D. (2013). Exploring gains in reading and mathematics achievement among regular and exceptional students using growth curve modeling. *Learning and Individual Differences*, 23(4), 92-100. doi: <https://doi.org/10.1016/j.lindif.2012.10.002>
- Sterba, S. K. (2014). Fitting nonlinear latent growth curve models with individually varying time points. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4), 630-647. doi: <https://doi.org/10.1080/10705511.2014.919828>
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth ed.). New York: Springer. doi: <https://doi.org/10.1007/978-0-387-21706-2>
- Wood, P. K., Steinley, D., & Jackson, K. M. (2015). Right-sizing statistical models for longitudinal data. *Psychological Methods*, 20(4), 470-488. doi: <https://doi.org/10.1037/met0000037>

- Wu, W., & Lang, K. M. (2016). Proportionality assumption in latent basis curve models: A cautionary note. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(1), 140-154. doi: <https://doi.org/10.1080/10705511.2014.938578>