Running Head: PIECEWISE GROWTH MODELS

Modeling Growth in Latent Variables Using a Piecewise Function

Nidhi Kohli and Jeffrey R. Harring

University of Maryland

Abstract

Latent growth curve (LGC) models with piecewise functions for continuous repeated measures data have become increasingly popular and versatile tools for investigating individual behavior that exhibits distinct phases of development in observed variables. As an extension of this framework, this research study considers a piecewise function for describing segmented change of a latent construct over time where the latent construct is itself measured by multiple indicators gathered at each measurement occasion. The time of transition from one phase to another is not known a priori, and thus, is a parameter to be estimated. Maximum likelihood estimation of the model will be described and Mplus 6.1 will be used to fit the model. An empirical example will be presented to illustrate the utility of the model and annotated M*plus* code is provided in the Appendix to aid in making this class of models accessible to practitioners.

Keywords: Latent growth curve models; piecewise; knot; latent variable

Modeling Growth in Latent Variables Using a Piecewise Function

1. Introduction

A common challenge for researchers and practitioners across different research domains is to understand how certain variables change and develop over time. For instance, when new skills are acquired, or when attitudes and interests develop, people change. Measuring change over time requires a longitudinal perspective where repeated measurements are gathered for collection of individual subjects. To accommodate change over time as an underlying latent process, the latent growth curve model (LGC), a special sub-class within structural equation modeling (SEM), is used to analyze repeated measures data (Meredith & Tisak, 1990). The LGC model is defined for each individual subject; however, the main focus of analyses in many applications is on the change at the population level, that is, average growth trajectory in the population rather than change at the individual-subject level (Cudeck & Harring, 2007).

The LGC model allows us to disentangle the correlational structure of the repeated measures into intra-individual (within-person) variability as well as inter-individual (between-person) variability in individual subjects' growth characteristics across time (Preacher, Wichman, MacCallum, & Briggs, 2008). A classic application of LGC models specifies a function describing a linear change process often comprised of two latent growth factors: (a) an intercept which describes initial level or status at some temporal reference point, and (b) a linear slope of growth which summarizes constant change over time. These two latent growth factors can be characterized by their mean values, individual random variation and covariation around these two latent growth components

(Duncan, Duncan, & Strycker, 2006). Certainly, other functional forms besides one that posits a linear change process for the repeated measures are possible. In lieu of choosing a model on a strictly theoretical basis, summarizing the repeated measures data in this way is typically accomplished via an empirical exploration of the data. For repeated measures data that exhibit curvilinear behavior, the LGC framework is flexible enough to accommodate a variety of nonlinear functions (see e.g., Choi, Harring, & Hancock, 2009; Grimm & Ram, 2009). For example, a quadratic function may be proposed for a developmental process that reaches a peak and then is expected to fall off, perhaps due to fatigue. In other research scenarios, individual performance on a learning task that levels-off toward the end of the study period may suggest choosing an intrinsically nonlinear function that can represent this type of limiting, asymptotic behavior. Another possibility allows the functional form for the repeated measures not to be specified in advance but be estimated (see e.g., Meredith & Tisak, 1990). A more detailed discussion of LGC models, along with a number of extensions, can be found in Duncan, Duncan, and Strycker (2006) as well as Preacher, Wichman, MacCallum, and Briggs (2008).

A LGC model that examines change across time in repeated measurements of observed variables is termed a “first-order” LGC model. An extension of first-order LGC models are “second-order” LGC models that describe change in a latent construct over time, where the latent construct of interest is measured by multiple indicators gathered at each measurement occasion. In second-order LGC models the first-order latent factors are modeled as dependent on one or more second-order latent growth factors, with the latter having only the first-order latent factors as indicator variables. Thus, second-order latent factors explain the means and variances of and covariances among, first-order latent factors (see, e.g., Duncan, Duncan, & Strycker, 2006; Hancock, Kuo, & Lawrence, 2001). Of course, auxiliary variables representing individual attributes, demographic information, or treatment condition can be incorporated to explain why second-order latent growth characteristics differ among individuals. This parallels many applications of first-order LGC models in which investigating treatment effectiveness or attributing differences in growth characteristics to subject-specific explanatory variables is accomplished at a secondary stage of the analysis – typically after the functional form of the repeated measures has been established.

Whether first-order or second-order LGC frameworks are used to investigate longitudinal change, the vast majority of research studies using LGC models regularly presume that the functional form describing the overall change process in the repeated measures data is a smooth, continuous curve with no breaks, elbows, or other irregularities. However, assuming a single uninterrupted functional form underlies the overall change process may be unrealistic for applications where data are comprised of different growth phases. Piecewise latent growth curve (PLGC) models, an extension of LGC models, allows the incorporation of separate growth profiles corresponding to multiple developmental stages from which repeated observations are made (Chou, Yang, Pentz, & Hser, 2004). PLGC models are flexible because each phase can be specified to conform to a particular functional form of the overall change process (Cudeck & Harring, 2010). The term "piecewise" originates from the piecewise regression model, which is a special case of spline regression models (Marsh & Cormier, 2001). To make this idea more concrete, consider a linear-linear piecewise process. In this situation, the formulated model assumes a simple regression line for the dependent variable, but with possibly different parameterizations in different ranges of the predictor (Bates & Watts, 1988; see also Seber & Wild, 1989, Chpt. 9). Figure 1 shows a plot of a linear-linear process.

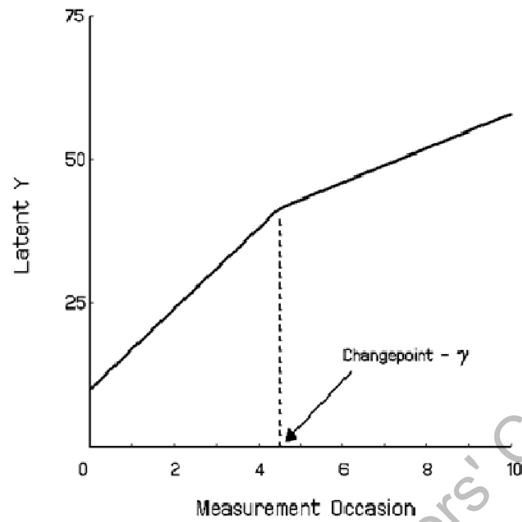

Figure 1. Plot of generic linear-linear process with changepoint at γ.

One of the most interesting features of a piecewise model is the knot or changepoint. The knot is the value of the predictor where the "pieces" from the developmental stages meet and can be known a priori or estimated and is denoted as γ in Figure 1. Harring, Cudeck, and du Toit (2006) demonstrated how a first-order piecewise linear mixed effects model, where the location of the knot was unknown, could be fit as a SEM to data for investigating individual behavior that exhibited distinct phases in observed variables.

The purpose of the current study is to extend the first-order piecewise LGC model to a second-order structure to examine a linear-linear piecewise change process in latent variables with an unknown knot location. In this context, the latent variable is measured by the same multiple indicators gathered at each measurement occasion although this restriction is not necessary to draw valid longitudinal inferences (see e.g., Bollen &

Curran, 2006; Hancock & Buehl, 2008). Although the knot is to be estimated, it is assumed to be the same across individuals. At first glance, constraining the knot to be the same across individuals may seem overly restrictive; yet in many biological or behavioral processes it does not seem unreasonable that some watershed event may occur at roughly the same moment in time for all individuals. For instance, in reading research it is hypothesized that fluency, a measure of accurate and automatic decoding at an appropriate pace, may increase at one rate beginning in second-grade but then changes at a different, slower rate for most students in the middle of their third-grade year. If grade is used as a proxy for the timing of collected observations, the transition between two-phases of fluency development might be expected to be the same for all students, but unknown a priori. Because the knot enters the function in a nonlinear manner but is fixed and does not vary across individuals; this second-order PLGC models turns out not to be much more complicated to set up than a restricted factor analysis with structured mean vector and covariance matrix (see e.g., Harring, 2009; Harring, Kohli, Silverman, & Speece, in review). Thus, SEM software – with all of its features – can be utilized as the platform for estimating model parameters. The estimation of this model is carried out in M*plus* 6.1 (Muthén & Muthén, 1998-2010), a popular SEM program. M*plus* code for the model can be found in the Appendix A.

The remainder of the paper is outlined in the following way. In the next section, the model is developed and the likelihood function specified. Empirical data is introduced and analyzed in a subsequent section. Finally, conclusions are framed in terms of the model's limitations as well as directions for future research.

2. Model Specification

Measurement Model

In a second-order PLGC model the repeated measure to be analyzed is an unobservable construct; hence to fit this model to data the model is augmented to include a measurement model that directly connects the observed variables to the latent factors. This relation is typically operationalized in terms of a measurement model connecting the observed indicators with the corresponding latent variable across time. Consider the $k \times 1$ response vector, $\mathbf{y}_{ij}^{*} = (y_{ij1}, y_{ij2}, \ldots, y_{ijk})'$, for individual i, $i = 1, \ldots, n$, at time *j* with $1 \leq j \leq m$. It is assumed that these *k* observed variables at time *j* measure a single latent variable, η_i, for the *i*th individual. A linear factor model (cf. Lawley & Maxwell, 1971) is specified that characterizes the relation of the observed variables to the latent variable:

$$\dot{\mathbf{y}}_{ij} = \dot{\boldsymbol{\mu}}_j + \dot{\boldsymbol{\lambda}}_j \dot{\eta}_{ij} + \dot{\boldsymbol{\delta}}_{ij} \tag{1}$$

where $\dot{\boldsymbol{\mu}}_j$ is a $k \times 1$ vector of variable intercepts, $\dot{\boldsymbol{\lambda}}_j$ is a $k \times 1$ vector of fixed or unknown factor loadings which describe the linear relation between the latent variable and the manifest variables, $\dot{\eta}_{ij}$ is the latent variable, and $\dot{\boldsymbol{\delta}}_{ij}$ is a $k \times 1$ vector of uniquenesses, or measurement errors. Like standard factor analysis, the common factor is assumed to be independent of the errors (i.e., $\text{cov}(\dot{\delta}'_{ij}, \dot{\eta}_{ij}) = 0$). Furthermore, once the linear dependence among the manifest variables is accounted for, the errors are assumed to be mutually independent (i.e., $\text{cov}(\dot{\delta}_{ij}, \dot{\delta}_{ij'}) = 0$).

In many situations where multiple instruments are used in a longitudinal design, it is not unusual for the same battery to be given repeatedly. In this case, if a complete set of the same *k* variables were obtained at multiple occasions – with a maximum of *m* potential time points ($j = 1, 2, \ldots, m$) then individual i would have a response vector with a total number of $T = mk$ observations – although other design considerations are certainly possible depending on the availability of the same instrumentation (Bollen & Curran, 2006) and whether or not the indicators of the construct shifts over time (Hancock & Buehl, 2008).

Working from the scenario that the observed variable indicators are identical at each time point, let $\mathbf{y}_i' = (\dot{\mathbf{y}}_{i1}', \ldots, \dot{\mathbf{y}}_{im}')$ denote a $T \times 1$ vector of responses for individual *i*, stacked according to *j* across all *m* occasions. Similarly, the linear factor model can be viewed like the stacked response vectors across all *m* measurement occasions can be specified as:

$$\mathbf{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\delta}_i . \qquad (2)$$

In Equation 2, $\boldsymbol{\mu}$ is a $T \times 1$ vector of intercepts,

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_m \end{bmatrix}$$

$\boldsymbol{\Lambda}$ is a $T \times m$ block diagonal matrix of factor loadings,

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\lambda}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \boldsymbol{\lambda}_m \end{bmatrix}$$

$\boldsymbol{\eta}_i$ is a $m \times 1$ vector of latent factors corresponding to individual *i*,

$$\boldsymbol{\eta}_i = \begin{bmatrix} \eta_{i1} \\ \vdots \\ \eta_{im} \end{bmatrix}$$

and $\boldsymbol{\delta}_i$ is a $T \times 1$ vector of measurement errors,

$$\boldsymbol{\delta}_i = \begin{bmatrix} \boldsymbol{\delta}_{i1} \\ \vdots \\ \boldsymbol{\delta}_{im} \end{bmatrix}$$

The distribution of the unique factors is given as:

$$\boldsymbol{\delta}_i \sim N(\mathbf{0}, \boldsymbol{\Theta}(\boldsymbol{\varphi})) \,. \qquad (3)$$

The matrix $\boldsymbol{\Theta}(\boldsymbol{\varphi})$ is a $T \times T$ symmetric covariance matrix in which the diagonal elements contain the variances of the measurement errors corresponding to the linear factor model of the repeated measures while the off-diagonal elements are the covariances of these errors. Unlike conventional factor analysis where the covariance matrix of the unique factors is assumed to be strictly diagonal, specification of off-diagonal elements of $\boldsymbol{\Theta}(\boldsymbol{\varphi})$ under the longitudinal design implied in Equation 2 is commonplace. For example, allowing covariances of temporally adjacent pairs of measurement errors to be freely estimated would seem plausible given that the same indicators are measured repeatedly over time. In some domains, the within-individual variances may actually increase or decrease systematically – a situation in which variances may depend on the mean. Other structures can be tailored to correspond with other design, theoretical, or empirical considerations with the stipulation that this be done as parsimoniously as possible.

Piecewise Model for the Latent Repeated Measures

The structural model for the repeated latent variable is a two-phase linear-linear latent growth process with a piecewise function:

$$\boldsymbol{\eta}_i = g_{ij}(t_j, \boldsymbol{\alpha}_i, \gamma) + \boldsymbol{\zeta}_i \qquad (4)$$

where $\boldsymbol{\zeta}_i$ is a vector of random disturbances in the first-order latent factors, $\boldsymbol{\eta}_i$, that are often assumed to be normally-distributed with mean vector, **0**, and covariance matrix $\boldsymbol{\Delta}$ (i.e., $\boldsymbol{\zeta}_i \sim N(\mathbf{0}, \boldsymbol{\Delta})$) and uncorrelated with $\boldsymbol{\alpha}_i$ and $\boldsymbol{\delta}_i$. Function *g* is defines the linear-linear piecewise model

$$g_{ij} = \begin{cases} \alpha_{i1} + \alpha_{i2} t_j & t_j \le \gamma \\ \alpha_{i3} + \alpha_{i4} t_j & t_j > \gamma \end{cases}. \tag{5}$$

where t_j is the *j*th time point, γ is the unknown knot, α_{i1} and α_{i2} are the intercept and linear slope of the first segment, and α_{i3} and α_{i4} are the intercept and linear slope of the second segment. Note that the regression coefficients have an *i* subscript and therefore vary by individual whereas the knot, γ, is fixed for all subjects. While not universally true, if it is presumed that the functions characterizing the two phases join at the knot, then the function values at γ are equal (i.e., $\alpha_{i1} + \alpha_{i2}\gamma = \alpha_{i3} + \alpha_{i4}\gamma$). This implies that one parameter is unnecessary and can be eliminated. Of the four regression parameters, α_{i3} seems the least interesting as it corresponds to the value of η at $t = 0$ of the second segment – a point not pertinent to the second phase. In the end, the choice is completely arbitrary. The terms of the equality constraint can be rearranged and solved for α_{i3}: $\alpha_{i3} = \alpha_{i1} + \alpha_{i2}\gamma - \alpha_{i4}\gamma$. Equation 6 shows this modification,

$$g_{ij} = \begin{cases} \alpha_{i1} + \alpha_{i2} t_j & t_j \le \gamma \\ \alpha_{i1} + \alpha_{i2}\gamma + \alpha_{i4}(t_j - \gamma) & t_j > \gamma \end{cases}. \tag{6}$$

The number of parameters that must be estimated in the Equation 6 is four, three linear coefficients: $\boldsymbol{\alpha}_i' = (\alpha_{i1}, \alpha_{i2}, \alpha_{i4})$ and one nonlinear coefficient, γ. As a starting point, an individual's regression coefficients, $\boldsymbol{\alpha}_i$, are simply the sum of fixed and random effects

$$\boldsymbol{\alpha}_i = \boldsymbol{\alpha} + \mathbf{a}_i$$

where the random effects are assumed to be multivariate normal such that $\mathbf{a}_i \sim N(\mathbf{0}, \boldsymbol{\Phi})$.

In its current form, the model in Equation 6 together with the measurement model in Equation 2 cannot be directly estimated within SEM software. The difficulty stems from the inability of the software to incorporate executable programming functions, like if-then statements, in the estimation step. In other environments, there have been several solutions put forth to work around this problem including using built-in minimum/maximum functions or user-defined programmable statements within the statistical software module. A parameterization used here was first introduced by Harring, Cudeck, and du Toit (2006), which circumvents this problem by rewriting the function as a polynomial and using the nonlinear constraints feature now pervasive in most SEM software packages. For sake of conserving space, only the final modeling development is reproduced here (the interested reader is encouraged to re-examine Harring et al. (2006) for a comprehensive explanation of the model). Following Harring et al. (2006), the re-parameterized model is

$$g_{ij} = \beta_{i1} + \beta_{i2} t_j + \beta_{i3}\sqrt{(t_j - \gamma)^2}\ . \qquad (7)$$

In Equation 7, where $\beta_{i1} = (\alpha_{i1} + \alpha_{i3})/2$, $\beta_{i2} = (\alpha_{i2} + \alpha_{i4})/2$, and $\beta_{i3} = (\alpha_{i4} - \alpha_{i2})/2$. The newly formed parameters $\boldsymbol{\beta}_i = (\beta_{i1}, \beta_{i2}, \beta_{i3})'$ are assumed to follow a multivariate normal distribution: $\boldsymbol{\beta}_i \sim N(\boldsymbol{\beta}, \boldsymbol{\Omega})$. Upon convergence of the program, the original regression coefficients and their corresponding standard errors can be estimated via the multivariate delta method (Oehlert, 1992). See Appendix B for the delta method transformation.

3. Maximum Likelihood Estimation

All of the parameters on the right side of Equation 7 which have *i* subscripts enter function *g* in a linear fashion. Thus, the model in Equation 7 can be written in matrix form as

$$g_{ij} = \mathbf{\Gamma}(\gamma)\boldsymbol{\beta}_i \,. \tag{7}$$

The coefficient matrix $\mathbf{\Gamma}(\gamma)$ is a function of constants, time, and nonlinear parameter γ with the *j*th row of $\mathbf{\Gamma}(\gamma)$ defined as

$$\mathbf{\Gamma}(\gamma)_j = \begin{bmatrix} 1 & t_j & \sqrt{(t_j - \gamma)^2} \end{bmatrix}$$

In the final staging of formulating the model, the first-order linear factor model for manifest variables in Equation 2 and the model for the first-order factors in Equation 4 with the model for *g* in Equation 7 substituted in Equation 4 can be expressed jointly as:

$$\mathbf{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\left[\mathbf{\Gamma}(\gamma)\boldsymbol{\beta}_i + \boldsymbol{\zeta}_i\right] + \boldsymbol{\delta}_i \,. \tag{8}$$

Given the distributional assumptions of $\boldsymbol{\beta}_i$, $\boldsymbol{\zeta}_i$, and $\boldsymbol{\delta}_i$, the mean value and covariance matrix of the response $\mathbf{y}_i$, are

$$\mathrm{E}[\mathbf{y}_i] = \boldsymbol{\mu}_y = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{\Gamma}(\gamma)\boldsymbol{\beta} \,. \tag{9}$$

$$\mathrm{Var}[\mathbf{y}_i] = \boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}\left(\mathbf{\Gamma}(\gamma)\boldsymbol{\Omega}\mathbf{\Gamma}'(\gamma) + \boldsymbol{\Delta}\right)\boldsymbol{\Lambda}' + \boldsymbol{\Theta} \,. \tag{10}$$

Fitting the Model

A second-order PLGC model imposes structures on the mean vector and covariance matrix, $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta})$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a $z \times 1$ vector whose elements consist of all free parameters of the model. Typically, these models are fitted by minimizing, with respect to $\boldsymbol{\theta}$, a function, $F(\overline{\mathbf{x}}, \mathbf{S}; \boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ that measures the discrepancy between the sample mean vector, $\bar{\mathbf{x}}$, and covariance matrix $\mathbf{S}$ and the mean vector and covariance matrix implied by the model, $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, respectively. For maximum likelihood estimation the discrepancy function to be minimized is

$$F(\bar{\mathbf{x}}, \mathbf{S}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) + \ln|\boldsymbol{\Sigma}| - \ln|\mathbf{S}| + tr\left[(\mathbf{S} - \boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right]. \quad (11)$$

Computations were carried out using M*plus* 6.1(see Appendix A for M*plus* input file).

4. Example

To demonstrate the use of the second-order PLGC model, next we fit the model defined in Equation 7 and Equation 2 to a single artificial data set[1].

Artificial Data Set

The data are modeled after a 2-year longitudinal study investigating the developmental trajectories of vocabulary depth and breadth, and reading comprehension among a group of Spanish-English bilingual and English monolingual students in the northeast United States (Proctor, Silverman, & Harring, 2011). A cohort-sequential, longitudinal design was implemented in which three cohorts of students in second-, third-, and fourth-grades were followed over two years. Conceptually, a cohort-sequential design integrates adjacent segments consisting of limited longitudinal data on a specific grade cohort, which can be linked together with similar segments from other temporally related grade cohorts to determine the existence of a common developmental trend (Duncan, Duncan, & Hops, 1996; Marsh, Craven, & Debus, 1998). As of this writing, the first data collection in the second-year was completed in the late Fall of 2010; however, because of the multi-site, multi-city nature of the design, the data were not yet accessible.

[1] We had fully intended to have a real data set using reading data. The data were not made available before this manuscript was submitted. The characteristics of the artificial data set resemble characteristics of the actual data as described more fully herein.

At the completion of the data collection in the Spring of 2011, there will be 12 time points across four grades. The population model for the data generation is based on sample statistics from the first-year's data collected from cohorts in second-fourth grades (three measurements per grade per year). Following the reading study, vocabulary depth is defined as including measures of morphological awareness, awareness of semantic relations, and syntactic awareness. The Extract the Base test (Anglin, 1993; August, Kenyon, Malabonga, Louguit, & Caglarcan, 2001; Carlisle, 1988) was used to evaluate awareness of derivational morphology. The Clinical Evaluation of Language Fundamentals (CELF; Semel, Wiig, & Secord, 2003) Word Classes 2 subtest was used to measure awareness of semantic relations; while the CELF Formulated Sentences subtest was used to measure syntactic awareness. The CELF formulated sentences subtest will be used to scale the latent variable for vocabulary depth. Higher scores across time represent greater vocabulary depth. The population generating values are given in Table 1.

To give a sense of the data generation, a spaghetti plot is displayed in Figure 2. It is not as straightforward to visualize the "data" in second-order PLGC models as it is in first-order models because the raw data to be modeled is the unobservable, repeated latent trait. One possible solution that addresses this issue is to construct a weighted composite of the multiple indicators at each time point for each individual. Use the subsequent composite score as the ''raw'' data to examine individuals' profiles. A second method, and the one that was employed here, involves fitting the observed variables to a multiple indicator confirmatory factor analysis (CFA) model with correlated factors and allowing the factor means to be estimated. In a subsequent step, predicted factor scores[2] for each individual

Table 1

Population values for the data generation of the PLGM represented in Equations 2, 4 and 6.

Parameter	Value	Distribution
α_1	25	
α_2	5	
α_4	1	
γ	4.75	
μ	0	
λ	0.7	
$\mathbf{\Phi}$	$\mathbf{\Phi} = \begin{pmatrix} 10 & & \\ 0 & 1 & \\ 0 & 0 & .5 \end{pmatrix}$	$\mathbf{a}_i \sim MVN(\mathbf{0}, \mathbf{\Phi})$
ζ	0.5	$\zeta_i \sim N(0,1)$

The errors, δ_i, were generated under $\delta_i \sim N(0,1)$ with variances of the errors chosen so that the reliability of the each indicator was = 0.80. A population of 10,000 cases was generated from which N = 300 cases were randomly selected. The 300 cases represent the approximate sample size in the original study.

on each latent variable are obtained and plotted as the raw data. A random sample of n = 30 artificial cases from the total sample (i.e., N = 300 randomly chosen cases) can be seen in Figure 2 with the trajectory of latent means distinguished by a thicker, darker line. The 9 measurement occasions correspond to grade intervals with the caveat that the time differential between any two occasions is 1/3 of a year. That is, measurement occasion 1 corresponds to early Fall of second-grade while occasion 2 corresponds to mid-Winter of second-grade, and so on. The last measurement occasion corresponds to late Spring of

[2] Factor scores were computed using the regression method (Thomson, 1934; Thurstone, 1935).

fourth-grade. Theory regarding vocabulary depth in children would dictate that children, while still increasing, begin to slow down around the middle of third grade. This corresponds to gains made in *decoding* what they are reading (Proctor, Silverman, Harring, & Montecillo, in review).

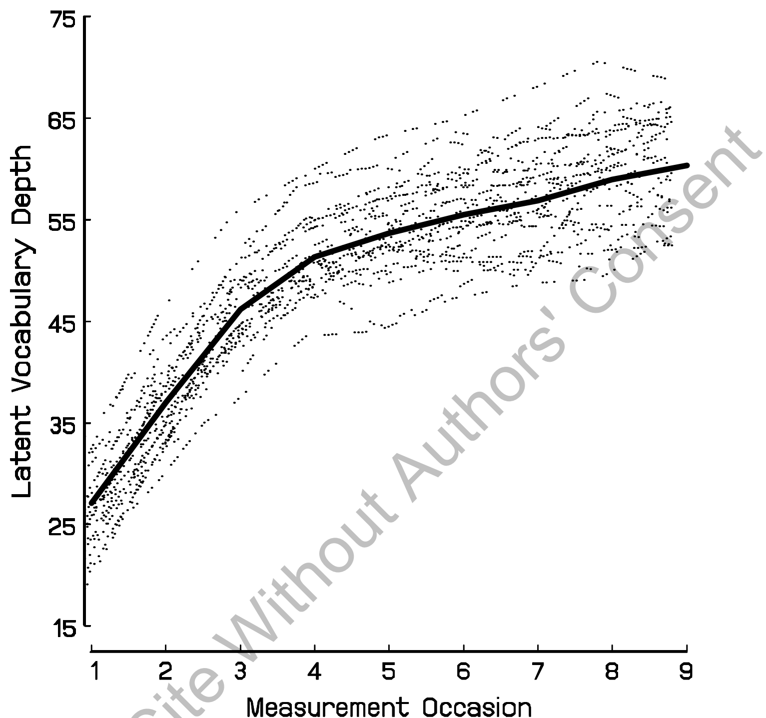

Figure 2. Spaghetti plot of 30 randomly selected cases for the linear-linear piecewise latent growth curve model.

5. Results

When using second-order growth models to investigate longitudinal change, an implicit assumption is that the same latent variable has been measured across time. That is, any change in the latent variable is due to true change in the underlying phenomena or construct and not due to changes that may occur in the measurement model. Thus, the invariance of measurement properties of the latent construct over time must be determined in order to draw valid inferences regarding the change process. In the current study, the data used was artificially generated; hence, measurement invariance of the latent construct is not an issue with respect to this study and will not be discussed further (see e.g., Harring, 2009; for a review of the measurement invariance testing strategy in second-order growth models).

To evaluate the fit of the second-order PLGC model, traditional SEM fit indices: standardized root mean squared residual (SRMR), root mean squared error (RMSEA) or comparative fit index (CFI) can also be computed. SRMR is an index of absolute fit that assess how well a model reproduces the sample data and estimates the amount of variance and covariance accounted for by the model (Hu & Bentler, 1995). A smaller value (≤ 0.08) of SRMR indicates a better model fit (Hu & Bentler, 1999). RMSEA is an index of parsimonious fit that assess the overall discrepancy between the model-implied covariance matrix and sample covariance matrix while taking into account model's simplicity. A smaller value (≤ 0.06) of RMSEA indicates a better model fit (Hu & Bentler, 1999). CFI is an index of incremental fit that assess the relative improvement of the model being tested with the baseline or null model ((Hu & Bentler, 1995, 1999)**.** A higher value (≥ 0.95) indicates a better model fit (Hu & Bentler, 1999). Results of model fit statistics for the second-order PLGC model are summarized in Table 2. Based on model fit indices, the second-order PLGC model seems to fit the data nicely.

Table 2

Model Fit Indices for the second-order PLGC model.

Model	χ^2	*df*	CFI	RMSEA	SRMR
PLGM	385.13*	365	0.99	0.014	0.045

Note: CFI=comparative fit index; RMSEA=root mean squared error of approximation, SRMR=standardized root mean squared residual. *$p < .01$

When model in Equation 7 is estimated directly, the estimated parameters of the original second-order PLGC model can be reconstructed from $\hat{\omega}_{i1}$, $\hat{\omega}_{i2}$, $\hat{\omega}_{i3}$ and $\hat{\gamma}$ as:

$$\hat{\alpha}_{i1} = \hat{\omega}_{i1} + \hat{\omega}_{i3}\hat{\gamma}$$

$$= 34.932 + (-2.053 \times 4.773) = 25.13$$

$$\hat{\alpha}_{i2} = \hat{\omega}_{i2} - \hat{\omega}_{i3}$$

$$= 3.015 - (-2.053) = 5.07$$

$$\hat{\alpha}_{i4} = \hat{\omega}_{i2} + \hat{\omega}_{i3}$$

$$= 3.015 + (-2.053) = 0.96$$

The estimated variances of the original parameters are reconstructed by using the delta method of transformation (See Appendix B for the delta method transformation). The estimates are:

$$\widehat{\mathbf{\Phi}} = \begin{bmatrix} 9.16 & & \\ & 0.89 & \\ & & 0.27 \end{bmatrix}$$

The estimated variance of the random disturbances in the first-order latent factors is: $\hat{\zeta}_i = 0.68$. Although the individual slopes and intercepts are variable between the individual cases, the knot is identical for all: $\hat{\gamma} = 4.77$.

6. Discussion

This research study considers a second-order PLGC model with unknown knot location for describing segmented change in a latent construct across time, where the latent construct is measured by a set of observed variables at each time occasion. To accomplish this model, a measurement model that directly connects the observed variables to the latent factors is augmented to the structural portion of the model. Both the measurement and the structural portion of the model in Equation 2 and Equation 6, respectively, cannot be directly estimated within SEM software, however. Hence, to fit this model the original model in Equation 6 needs to be parameterized. An obvious limitation of reparameterization is that the fit of the model may be affected by the transformation from one version of a model into another form. Harring et al. (2006) mentioned that generally the difference in fit is not great, and any slight loss in fit would seem to be offset by the ease with which the reparameterized model can be estimated.

Overall, second-order PLGC models can be very useful in the area of educational research where most often the interest of researchers is centered on student academic progress or changes in attitude and affect. Second-order PLGC models enable researchers to summarize individual behavior that exhibits distinct phases of development in each segment, and thereby allow researchers to address key questions such as developmental studies or studies seeking to measure the effect of treatment/ intervention, and so forth such as when individuals may need to seek professional services at the timing when mental ability decreases. Additionally, a second-order PLGC model can estimate the unknown location of knot which can enhance the ability of researchers to estimate when the treatment/intervention should be introduced so as to maximize its effects.

7. References

Anglin, J. M. (1993) Vocabulary development: A morphological analysis. *Monographs of Society Research in Child Development*, *58*, (10, Serial No. 238).

August, D., Kenyon, D., Malabonga, V., Louguit, M., & Caglarcan, S. (2001). *Extract the Base.* Washington, DC: Center for Applied Linguistics.

Bates, D. M. & Watts, D. G. (1988). *Nonlinear regression analysis and its applications*. New York, NY: Wiley.

Carlisle, J. F. (1988). Knowledge of derivational morphology and spelling ability in fourth, sixth, and eighth graders. *Applied Psycholinguistics*, *9*(3), 247-266.

Chou, C.-P., Yang, D., Pentz, M. A., & Hser, Y.-I. (2004). Piecewise growth curve modeling approach for longitudinal prevention study. *Computational Statistics & Data Analysis, 46*, 213-225.

Cudeck, R. (1996). Mixed-effects models in the study of individual differences with repeated measures data. *Multivariate Behavioral Research*, *31*, 371-403.

Cudeck, R., & Harring, J. R. (2007). The analysis of nonlinear patterns of change with random coefficient models. *Annual Review of Psychology*, *58*, 615-637.

Cudeck, R., & Harring, J. R. (2010). Developing a random coefficient model for nonlinear repeated measures data. In S.-M. Chow, E. Ferrer, & F. Hsieh (Eds.), *Statistical methods for modeling human dynamics: An interdisciplinary dialogue*. New York: Routledge.

Cudeck, R., & Klebe, K. J. (2002). Multiphase mixed-effects models for repeated measures data. *Psychological Methods*, *7*, 41-63.

Duncan, S. C., Duncan, T. E., & Hops, H. (1996). Analysis of longitudinal data with accelerated longitudinal designs. *Psychological Methods*, *65*, 130-140.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Hancock, G. R., Kuo, W-L., & Lawrence, F. R. (2001). An illustration of second-order latent growth models. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*, 470-483.

Harring, J. R. (2009). A nonlinear mixed effects model for latent variables. *Journal of Educational and Behavioral Statistics*, *34*, 293-318.

Harring, J. R., Cudeck, R., & du Toit, S. H. C. (2006). Fitting partially nonlinear random coefficient models as SEMs. *Multivariate Behavioral Research*, *41*, 579-596.

Marsh, H. W., Craven, R., & Debus, R. (1998). Structure, Stability and Development of Young Children's Self-Concepts: A Multicohort—Multioccasion Study. *Child Development*, *69*(4), 1030-1053.

Marsh, L. C., Cormier, D. R. (2001). *Spline regression models.* Thousand Oaks, CA: Sage.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107-122.

Muthén, L. K. and Muthén, B. O. (1998-2010). Mplus User's Guide (6th ed.). Los Angeles, CA: Muthén & Muthén.

Oehlert, G. W. (1992). A note on the delta method. *The American Statistician*, *46*, 27-29.

Proctor, C. P., Silverman, R. D., Harring, J. R., & Montecillo, C. (in review). The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish-English bilingual children in elementary school. *Reading and writing: An interdisciplinary journal*.

Seber, G. A. F., & Wild, C. J. (1989). *Nonlinear regression*. New York, NY: Wiley.

Semel, E. M., Wiig, E. H., & Secord, W. (2003). *Clinical Evaluation of Language Fundamentals.* San Antonio, TX: The Psychological Corporation.

Seltzer, M., & Svartberg, M. (1998). Studying change during treatment and follow-up periods: The use of piecewise growth models in evaluations of interventions. Los Angeles: Center for Research on Evaluation, Standards and Student Testing (CRESST), UCLA.

Preacher, K. J, Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling*. Thousand Oaks, CA: Sage.

Thomson, G. H. (1934). The meaning of i in the estimate of g. *British Journal of Psychology*, *25*, 92-99.

Thurstone, L. L. (1935). *The vectors of mind.* Chicago, IL: University of Chicago Press.

8. Appendix A

Annotated M*plus* Input for Example

```
TITLE:  2nd-order piecewise model
DATA:   FILE IS dat3.dat;
VARIABLE: NAMES ARE y1-y27;
ANALYSIS: ESTIMATOR = ML;
        ITERATIONS = 10000;
        SDITERATIONS = 500;
        H1ITERATIONS = 10000;
        CONVERGENCE = .001;
        H1CONVERGENCE = .001;
MODEL:
!Measurement Portion of the PLGM
t1 BY
    y1
    y2*.7(1)
    y3*.7(2);
t2 BY
    y4
    y5*.7(1)
    y6*.7(2);
t3 BY
    y7
    y8*.7(1)
    y9*.7(2);
t4 BY
    y10
    y11*.7(1)
    y12*.7(2);
t5 BY
    y13
    y14*.7(1)
    y15*.7(2);
t6 BY
    y16
    y17*.7(1)
    y18*.7(2);
t7 BY
    y19
    y20*.7(1)
    y21*.7(2);
t8 BY
    y22
    y23*.7(1)
```

```
    y24*.7(2);
t9 BY
    y25
    y26*.7(1)
    y27*.7(2);

y1-y27*;
[y1@0 y4@0 y7@0 y10@0 y13@0 y16@0 y19@0 y22@0 y25@0];
[y2 y5 y8 y11 y14 y17 y20 y23 y26](2);
[y3 y6 y9 y12 y15 y18 y21 y24 y27](3);

!Structural Portion of the PLGM
w1 BY t1-t9@1;
w2 BY t1@0 t2@1 t3@2 t4@3 t5@4 t6@5 t7@6 t8@7 t9@8;
w3 BY t1* (p1); !Column 3 of design matrix
w3 BY t2-t9* (p2-p9);
w1*10(v1);
w2*1(v2);
w3*.5(v3);
w1 WITH w2*0;
w1 WITH w3*0;
w2 WITH w3*0;
[w1*56 w2*7.5 w3*-2.4];
[t1-t9@0];
t1-t9*.7(1); !

MODEL CONSTRAINT:
NEW (gam*4.75);
v1 > 0;
v2 > 0;
v3 > 0;

p1 = (sqrt((gam)^2));
p2 = (sqrt((1-gam)^2));
p3 = (sqrt((2-gam)^2));
p4 = (sqrt((3-gam)^2));
p5 = (sqrt((4-gam)^2));
p6 = (sqrt((5-gam)^2));
p7 = (sqrt((6-gam)^2));
p8 = (sqrt((7-gam)^2));
p9 = (sqrt((8-gam)^2));

OUTPUT: SAMPSTAT;
```

9. Appendix B

Multivariate Delta Method

$\hat{\alpha}_{i1} = \hat{\omega}_{i1} + \hat{\omega}_{i3}\hat{\gamma}$ $\hat{\alpha}_{i2} = \hat{\omega}_{i2} - \hat{\omega}_{i3}$ $\hat{\alpha}_{i4} = \hat{\omega}_{i2} + \hat{\omega}_{i3}$

1) $Var(\hat{\alpha}_{i1}) = \mathbf{D^T \Sigma D}$

$$= \begin{bmatrix} \frac{\partial(\hat{\alpha}_{i1})}{\partial(\hat{\omega}_{i1})} & \frac{\partial(\hat{\alpha}_{i1})}{\partial(\hat{\omega}_{i2})} & \frac{\partial(\hat{\alpha}_{i1})}{\partial(\hat{\omega}_{i3})} \end{bmatrix} . \begin{bmatrix} \hat{\sigma}^2_{\omega_{i1}} & \hat{\sigma}_{\omega_{i1},\omega_{i2}} & \hat{\sigma}_{\omega_{i1},\omega_{i3k}} \\ \hat{\sigma}_{\omega_{i2},\omega_{i1}} & \hat{\sigma}^2_{\omega_{i2}} & \hat{\sigma}_{\omega_{i2},\omega_{i3}} \\ \hat{\sigma}_{\omega_{i3k},\omega_{i1}} & \hat{\sigma}_{\omega_{i3},\omega_{i2}} & \hat{\sigma}^2_{\omega_{i3}} \end{bmatrix} . \begin{bmatrix} \frac{\partial(\hat{\alpha}_{i1})}{\partial(\hat{\omega}_{i1})} \\ \frac{\partial(\hat{\alpha}_{i1})}{\partial(\hat{\omega}_{i2})} \\ \frac{\partial(\hat{\alpha}_{i1})}{\partial(\hat{\omega}_{i3})} \end{bmatrix}$$

$$= [1 \quad 0 \quad 4.773] . \begin{bmatrix} 14.407 & 0.700 & -1.220 \\ 0.700 & 0.297 & -0.156 \\ -1.220 & -0.156 & 0.281 \end{bmatrix} . \begin{bmatrix} 1 \\ 0 \\ 4.773 \end{bmatrix} = 9.16$$

2) $Var(\hat{\alpha}_{i2}) = \mathbf{D^T \Sigma D}$

$$= \begin{bmatrix} \frac{\partial(\hat{\alpha}_{i2})}{\partial(\hat{\omega}_{i1})} & \frac{\partial(\hat{\alpha}_{i2})}{\partial(\hat{\omega}_{i2})} & \frac{\partial(\hat{\alpha}_{i2})}{\partial(\hat{\omega}_{i3})} \end{bmatrix} . \begin{bmatrix} \hat{\sigma}^2_{\omega_{i1}} & \hat{\sigma}_{\omega_{i1},\omega_{i2}} & \hat{\sigma}_{\omega_{i1},\omega_{i3k}} \\ \hat{\sigma}_{\omega_{i2},\omega_{i1}} & \hat{\sigma}^2_{\omega_{i2}} & \hat{\sigma}_{\omega_{i2},\omega_{i3}} \\ \hat{\sigma}_{\omega_{i3k},\omega_{i1}} & \hat{\sigma}_{\omega_{i3},\omega_{i2}} & \hat{\sigma}^2_{\omega_{i3}} \end{bmatrix} . \begin{bmatrix} \frac{\partial(\hat{\alpha}_{i2})}{\partial(\hat{\omega}_{i1})} \\ \frac{\partial(\hat{\alpha}_{i2})}{\partial(\hat{\omega}_{i2})} \\ \frac{\partial(\hat{\alpha}_{i2})}{\partial(\hat{\omega}_{i3})} \end{bmatrix}$$

$$= [0 \quad 1 \quad -1] . \begin{bmatrix} 14.407 & 0.700 & -1.220 \\ 0.700 & 0.297 & -0.156 \\ -1.220 & -0.156 & 0.281 \end{bmatrix} . \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} = 0.89$$

3) $Var(\hat{\alpha}_{2k}) = \mathbf{D^T \Sigma D}$

$$= \begin{bmatrix} \frac{\partial(\hat{\alpha}_{i3})}{\partial(\hat{\omega}_{i1})} & \frac{\partial(\hat{\alpha}_{i3})}{\partial(\hat{\omega}_{i2})} & \frac{\partial(\hat{\alpha}_{i3})}{\partial(\hat{\omega}_{i3})} \end{bmatrix} . \begin{bmatrix} \hat{\sigma}^2_{\omega_{i1}} & \hat{\sigma}_{\omega_{i1},\omega_{i2}} & \hat{\sigma}_{\omega_{i1},\omega_{i3k}} \\ \hat{\sigma}_{\omega_{i2},\omega_{i1}} & \hat{\sigma}^2_{\omega_{i2}} & \hat{\sigma}_{\omega_{i2},\omega_{i3}} \\ \hat{\sigma}_{\omega_{i3k},\omega_{i1}} & \hat{\sigma}_{\omega_{i3},\omega_{i2}} & \hat{\sigma}^2_{\omega_{i3}} \end{bmatrix} . \begin{bmatrix} \frac{\partial(\hat{\alpha}_{i3})}{\partial(\hat{\omega}_{i1})} \\ \frac{\partial(\hat{\alpha}_{i3})}{\partial(\hat{\omega}_{i2})} \\ \frac{\partial(\hat{\alpha}_{i3})}{\partial(\hat{\omega}_{i3})} \end{bmatrix}$$

$$= [0 \quad 1 \quad 1] . \begin{bmatrix} 14.407 & 0.700 & -1.220 \\ 0.700 & 0.297 & -0.156 \\ -1.220 & -0.156 & 0.281 \end{bmatrix} . \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = 0.27$$