當代教育研究季刊　第三十二卷　第四期，2024 年 12 月，頁 073-099
DOI : 10.6151/CERQ.202412_32(4).0003

《Research Note》

ESSENTIAL STEPS FOR LATENT GROWTH CURVE MODELING IN TAIWANESE PANEL STUDY

Chih-Cheng Lee[1] **Yu-Wen Lu**[2] **Man-Kit Lei**[3*]

ABSTRACT

The increasing availability of longitudinal panel datasets in Taiwan, such as the Taiwan Youth Project and the Panel Study of Family Dynamics, provides researchers with the tools to explore complex questions about human development using advanced longitudinal methods. One such statistical approach is Latent Growth Curve Modeling (LGCM), a technique rooted in Structural Equation Modeling that allows for the modeling of initial status and rates of change over time while considering inter-individual differences in trajectories. In this study, we examine recent research titled "Developmental Trajectory of Depressive Symptoms from Adolescence to Early Adulthood," which employed LGCM to investigate the influence of parenting styles and self-esteem on depressive symptom development among Taiwanese adolescents. We provide essential knowledge of LGCM, including its capacity to integrate time-invariant and time-varying co-variates, its ability to specify linear and non-linear growth patterns, and its advanced variations such as piecewise and growth mixture models. This study highlights the potential of LGCM as a powerful analytical tool for leveraging Taiwan's rich longitudinal datasets. By introducing the methodology and its advanced variations, we aim to encourage researchers to utilize LGCM to explore

Chih-Cheng Lee[1], Doctoral Student, Department of Sociology, University of Georgia
E-mail: leonlee@uga.edu
Yu-Wen Lu[2], Doctoral Student, Department of Sociology, University of Georgia
E-mail: winnielu@uga.edu
[*] Man-Kit Lei[3] (corresponding author), Associate Professor, Department of Sociology, University of Georgia
E-mail: karlo@uga.edu
Manuscript received: June 1, 2024; Modified: July 1, 2024; Accepted: July 29, 2024

developmental trajectories, contributing to a deeper understanding of individual growth patterns in the Taiwanese unique context.

Keywords: Latent Growth Curve Modeling, Longitudinal Data Analysis, Structural Equation Modeling, Unspecified Trajectory Model

Introduction

The examination of individual change within the context of macrostructural dynamics constitutes a central focus in the study of human behavior. From a life course perspective, different periods of the lifespan and the transitions between these periods profoundly impact individual behavior. Previous researchers have focused on studying the effects of specific life events and markers of these transitions to understand the processes of socialization and accommodation. However, the increasing uncertainty and diversity in contemporary society have made it challenging to predict human behavior trends, underscoring the importance of studying the variability and sensitivity of these time points across the lifespan (Shanahan, 2000). This focus on individual change is widespread in interdisciplinary research areas such as applied psychology, sociology, and education science, and the study of growth rates in biology (Barnes et al., 2000; Dmitriew, 2011).

This shift in theoretical perspectives has been accompanied by innovations in methodology. First, a significant shift in data collection methods has been witnessed. Longitudinal research has emerged as the preferred approach in empirical studies, replacing traditional cross-sectional research. The maturation and availability of longitudinal panel studies, which observe and collect information form the same groups of individuals over time, have proliferated since the 1980s (Mayer, 2009). This progress has enabled empirical studies to delve into the contextual time factors influencing human lives and track personal trajectories across the lifetime. Longitudinal panel data provide a clear temporal order between variables through different waves of data collection. Additionally, they allow researchers to collect pertinent variables from the same group of observations multiple times, contributing to the ability to predict future trends and capture the effects of causal changes during an individual's life course (Wu, 2008). Compared to cross-sectional studies, which obtain independent and dependent variables simultaneously, this time-lag characteristic of longitudinal studies makes it more feasible to control for common method biases, such as measurement context effects (Podsakoff et al., 2003).

The second aspect of methodological development is the rapid and fruitful innovation in statistical modeling. The key analytical objective in longitudinal panel data is to capture both individual and group-level changes simultaneously. However, traditional analytical approaches, such as repeated-measures ANOVA or multiple regression, can only focus on mean differences or treat individual variability as error variance, failing to account for the heterogeneity in growth trajectories across individuals (Duncan & Duncan, 2004). Latent Growth Curve Modeling (LGCM), rooted in the framework of Structural Equation Modeling (SEM), offers a powerful and flexible approach to analyzing longitudinal panel data. By incorporating latent variables that represent individual growth trajectories, LGCM enables researchers to model both the initial status and the rate of change over time. This approach not only captures the average growth patterns within a population but also accounts for individual differences in these trajectories, providing insights into the factors that influence developmental processes.

Over the past two decades, several longitudinal panel datasets in Taiwan have emerged as significant resources for exploring individual development from a life course perspective. For instance, the Taiwan Youth Project (TYP) and Taiwan Education Panel Survey (TEPS) both target school-age children and teenagers, making them optimal for analyzing students' mental and learning performance development across time and its associated factors. The Panel Study of Family Dynamics (PSFD), targeting teenagers and young adults, can analyze the relationship between family dynamics and the transition from adolescence to adulthood.

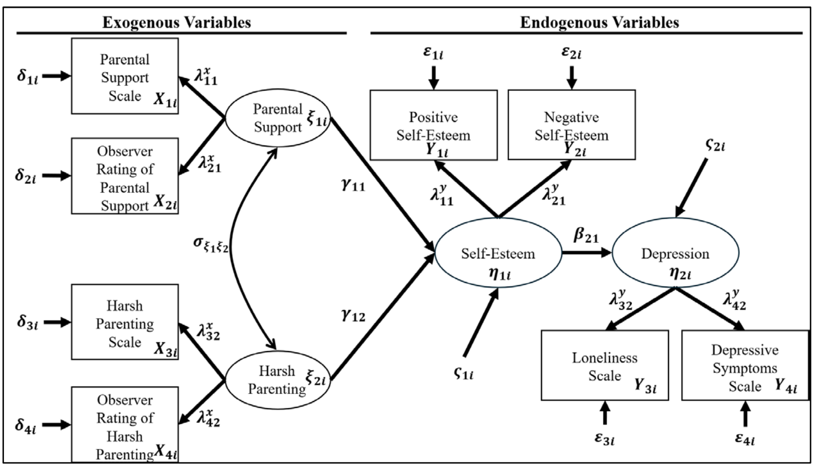
Owing to the availability of these datasets, researchers in Taiwan are now equipped to explore more complex research questions and conduct sophisticated longitu-dinal analyses. A recent study, "Developmental Trajectory of Depressive Symptoms from Adolescence to Early Adulthood," investigated the long-term effects of par-enting styles and self-esteem on depressive symptom development among Taiwanese adolescents. In this article, we aim to introduce the fundamental knowledge of LGCM for understanding the study, providing a possible pathway for further researchers to utilize these valuable datasets and contribute to interdisciplinary knowledge of human behavior studies.

Introduction to Structural Equation Modeling (SEM)

Latent Growth Curve Modeling (LGCM) provides a powerful technique for analyzing longitudinal data, deeply rooted in the Structural Equation Modeling (SEM) framework. SEM is a versatile statistical approach used to examine complex relationships among observed and latent variables within a single point in time (Duncan, 2014). An illustrated example of the SEM is shown in Figure 1. In SEM causal diagrams, latent constructs are usually enclosed in circles, while observed variables are en-closed in squares. SEM encompasses measurement models, which define how observed variables relate to latent constructs, and structural models, which specify the relationships among these latent constructs (Singer & Willett, 2003). There are two different measurement models: The X-measurement model and the Y-measurement model. The distinction between these two models stems from the nature of the latent constructs and variables measured, while the researchers should consider two conditions: exogeneity and endogeneity.

Figure 1

Example for SEM



As in Figure 1, the values of two latent constructs – parental support and harsh parenting – are entirely determined outside the hypothesized causal system, where this type of latent construct or variable is called an exogenous variable. In the SEM, exogeneity indicates that a construct or variable can only act as a predictor within the causal system. This is presented by variables on the left in Figure 1. For an exogenous variable, researchers would apply the X-measurement model to obtain the parameters. Consider the following X-measurement model:

$$\boldsymbol{X} = \boldsymbol{\tau_x} + \boldsymbol{\Lambda_x}\boldsymbol{\xi} + \boldsymbol{\delta}$$

where the $\boldsymbol{X}$ is a vector of the observed indicators. In our example in Figure 1, there are four indicators, so the vector $\boldsymbol{X}$ is $\{X_{1i}, X_{2i}, X_{3i}, X_{4i}\}$, where i indicates the number of observations. The parameter vector $\boldsymbol{\tau_x}$ reflects the mean of the corresponding indicator X_{pi}. The latent constructs are denoted by the vector $\boldsymbol{\xi}$, and in this case, there are two constructs in this case: ξ_{1i} and ξ_{2i}. The matrix $\boldsymbol{\Lambda_x}$ consists of the factor loadings, which rescale the indicators. This allows the indicators to be measured on different scales, making a single construct that can simultaneously influence several indicators. The matrix $\boldsymbol{\Lambda_x}$ has k columns, where k indicates the number of the latent constructs. Each column consists of λ^x_{pk} and 0's. In this case, the matrix $\boldsymbol{\Lambda_x}$ is:

$$\boldsymbol{\Lambda_x} = \begin{bmatrix} \lambda^x_{11} & 0 \\ \lambda^x_{21} & 0 \\ 0 & \lambda^x_{32} \\ 0 & \lambda^x_{42} \end{bmatrix}$$

The last term, $\boldsymbol{\delta}$, is an error vector that contains the measurement error of each observation. In general, the distribution of errors is assumed to have a zero mean, similar to the distribution of residual scores in a regression model. The variance of δ_{pi} is denoted by the variance-covariance matrix Θ_δ:

$$\Theta_{\boldsymbol{\delta}} = cov(\delta_{pi}) = \begin{bmatrix} \sigma^2_{\delta_1} & \cdots & \sigma_{\delta_1\delta_p} \\ \vdots & \ddots & \vdots \\ \sigma_{\delta_p\delta_1} & \cdots & \sigma^2_{\delta_i} \end{bmatrix}$$

It should be noted that although $\boldsymbol{\delta}$ can be seen as the measurement error, it is better considered as part of the indicator X_{pi} that does not depend on the corresponding construct ξ_{ki} (Singer & Willett, 2003).

The Y-measurement model pertains to the endogenous variables, which are always the outcome variables in the SEM. Any variable pointed to by an arrow in the causal diagram is considered to have endogeneity. In Figure 1, the endogenous variables are located on the right. The equation of the Y-measurement model is very similar to the X-measurement model:

$$\boldsymbol{Y} = \boldsymbol{\tau_y} + \boldsymbol{\Lambda_y}\boldsymbol{\eta} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{Y}$, $\boldsymbol{\tau}_y$, and $\boldsymbol{\Lambda}_y$ are the vector of the observed indicators, the mean vector, and the factor loading matrix, respectively. The vector $\boldsymbol{\eta}$ represents the latent constructs, which are self-esteem and depression in our example. The $\boldsymbol{\varepsilon}$ is the error term of the Y-measurement model. Similar to the X-measurement model, the mean of the distribution of $\boldsymbol{\varepsilon}$ is zero, and the variance-covariance matrix Θ_ε is

$$\Theta_\varepsilon = cov(\varepsilon_{qi}) = \begin{bmatrix} \sigma^2_{\delta_1} & \cdots & \sigma_{\delta_1\delta_q} \\ \vdots & \ddots & \vdots \\ \sigma_{\delta_q\delta_1} & \cdots & \sigma^2_{\delta_i} \end{bmatrix}$$

where q indicates the number of latent constructs of Y.

In addition to the measurement models, SEM utilizes the structural model to depict the relationship between latent constructs. As for the case in Figure 1, two exogenous constructs – parental support (ξ_{1i}) and harsh parenting (ξ_{2i}) – influence the endogenous variable, self-esteem (η_{1i}), where the relationship is denoted by γ_{11} and γ_{12}. The relationship of the two endogenous constructs – denoted by β_{21}– can also be observed, in which self-esteem (η_{1i}) can predict depression (η_{2i}). Therefore, a pair of simultaneous structural equations can be obtained from the relationships:

$$\eta_{1i} = \alpha_1 + \gamma_{11}\xi_{1i} + \gamma_{12}\xi_{2i} + \varsigma_{1i}$$
$$\eta_{2i} = \alpha_2 + \beta_{21}\eta_{1i} + \varsigma_{2i}$$

These equations can be written in an abbreviated version:

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \boldsymbol{\Gamma\xi} + \boldsymbol{B\eta} + \boldsymbol{\varsigma},$$

$$where\ \boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}, \boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ 0 & 0 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix}, and\ \boldsymbol{\varsigma} = \begin{bmatrix} \varsigma_{1i} \\ \varsigma_{2i} \end{bmatrix}$$

In the equations, α_1 and α_2 are the population means or intercepts of two endogenous constructs, while the ς_{1i} and ς_{2i} are the residuals. Detailed information and derivation for all the above equations can be found in Duncan (2014) and Singer and Willett (2003).

Introduction to Latent Growth Curve Modeling (LGCM)

The purpose of introducing the LGCM is to model repeated-measured data or panel data. Traditionally, researchers use repeated measures ANOVA and its variations to study longitudinal change (Huck & McLean, 1975). Notably, these approaches do not consider the distinctions among individual subjects; they assume that all individuals share one single trajectory. The distinctions could facilitate a better understanding of the mechanism of change, which should be included in a comprehensive study. Therefore, numerous analytical approaches have been developed to explore heterogeneity within individuals and between individuals, including random coefficient modeling, multilevel modeling, and LGCM (Box, 1950; Dmitriew, 2011; Duncan & Duncan, 2009; Gee, 2014; Podsakoff et al., 2003; Rogosa & Willett, 1985; Singer & Willett, 2003). The basic concept of these methods is to utilize random effects to accommodate the inter-individual variability.

Nevertheless, in recent studies, most approaches fail to investigate multiple repeatedly measured variables simultaneously (Curran et al., 2010). This can be illustrated by the example in Figure 1, where self-esteem is measured using two different indicators: positive and negative self-esteem. Suppose a new research question focuses on the change in self-esteem measured at several time points. In that case, most methods may require creating a single response variable by integrating the indicators first and then putting the variable into the model. In other words, the measurement model is separated from the main model, so the measurement error is assumed to be free in the main model, as in multilevel modeling (Singer & Willett, 2003).

In contrast, LGCM allows for the specification of a measurement model and assumes that the factor structure of the created latent construct is time-invariant (Newsom, 2015). In LGCM, the measurement error is estimated simultaneously, providing statistical power advantages. Consequently, the statistical power advantages and flexibility of LGCM make it one of the preferred modeling options for panel studies (Curran, 2000).

In addition to the built-in measurement model, there are several benefits to applying LGCM. LGCM builds on the foundation of SEM by introducing latent variables that represent growth trajectories over time, such as the initial status (intercept) and rate of change (slope). Due to its relationship with SEM, LGCM inherits the merits of SEM, including the ability to model longitudinal data with complex causal hypotheses and build latent constructs. Compared to multilevel modeling, which is also widely applied in recent studies, the most important characteristic could be its ability to treat variables or model parameters as both predictors and outcomes simultaneously in the same model (Preacher, 2008). For the example above, the change in self-esteem can be the outcome variable that researchers are interested in, while it can also be a predictor of depression at the same time, as hypothesized in Figure 1. This advantage is deeply rooted in its nature as a variety of SEM, where the relationships of numerous variables can be depicted in a path diagram, and mediators can be included in a single model.

Another benefit of applying LGCM is its flexibility in choosing the scale of time, which relates to the decision of factor loadings. In the framework of multilevel modeling, the scale of time can also be decided by researchers, where time is considered a variable with known values (Singer & Willett, 2003). In contrast, in LGCM, time scaling can be either a known variable or an unknown parameter that could be estimated by the algorithm. This unique feature was developed by Meredith and Tisak (1984, 1990) and extended by McArdle (1988), where the model is referred to as the "unspecified latent growth curve model," linear spline model, or latent basis model. This approach is basically a data-driven method for generating a non-linear form in LGCM by transforming the metric of time (Bollen & Curran, 2006; Newsom, 2015; Preacher, 2008). Further details will be discussed later.

Although LGCM possesses distinctive benefits, the disparity between LGCM and multilevel modeling is not markedly substantial (Chou et al., 1998; Preacher, 2008). From a mathematical perspective, Multilevel modeling, also referred to as hierarchical modeling, exhibits expressions identical to LGCM. The main distinction between these methods lies in the assumption regarding the variance inherent in the repeatedly measured variable, which will be discussed later. In fact, two-level multilevel growth curve modeling can be viewed as a specific form of LGCM (Singer & Willett, 2003).

Consider a random variable Y_{ij} that represents a set of repeated measures for individual i at time j, where the random variable is depression in our example, as shown in Figure 2. The metric of time, denoted by T_j, consists of the observed time indicators at each time point, where $T_j=\{t_1,t_2,...,t_j\}$. Based on the interest in the relationship between time and the response variable, we can write an equation in the following form:

$$Y_{ij} = \pi_{0i} + \pi_{1i}(T_j) + \varepsilon_{ij}$$

where π_{0i} represents the initial status of depression, also called the intercept factor or random intercept; π_{1i} is the slope factor, or random slope, that describes the rate of linear change in depression over the two-time points. This expression of the relationship is the standard level-1 equation in multilevel modeling, which also satisfies the standard form of the Y-measurement model in LGCM. Recall the equation of the Y-measurement model in the previous section, where $\boldsymbol{\tau}_y$ is fixed to a zero vector here. We will obtain the equation:

$$Y_{ij} = \boldsymbol{\Lambda_y \eta} + \varepsilon_{ij}$$

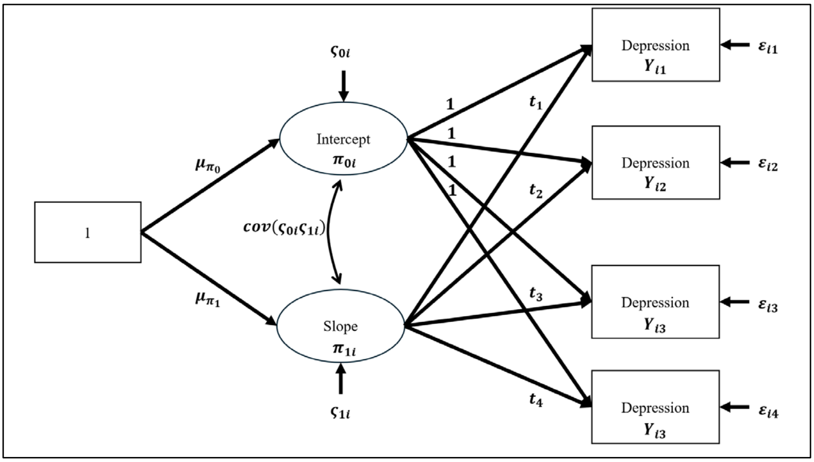
where $\boldsymbol{\Lambda}_y$ is the matrix of factor loadings, and $\boldsymbol{\eta}$ is the latent growth factor. If we let

$$\boldsymbol{\Lambda_y} = \begin{bmatrix} 1 & T_j \end{bmatrix} \; and \; \boldsymbol{\eta} = \begin{bmatrix} \pi_{0i} \\ \pi_{1i} \end{bmatrix},$$

then we can derive the exact same equation in LGCM for multilevel modeling.

Figure 2

Example for unconditional LGCM



However, the main distinction between the two methods can be probed here, which is related to the assumption of the error term ε_{ij}. In the multilevel modeling, the residuals are assumed to follow a normal distribution with a zero mean and variance σ^2_ε (i.e., $\varepsilon_{ij} \sim N(0,\sigma^2_\varepsilon)$). In contrast, the assumption in LGCM is more flexible, where researchers can determine their customized assumption of residuals according to the theories related to the measured variables. The error term in LGCM can be either homoscedastic or heteroscedastic, and it can also be either independent or autocorrelated (Singer & Willett, 2003). This can be achieved by adjusting the covariance matrix of errors since the residuals are assumed to be normally distributed with mean zero and covariance matrix Θ_ε. That is, $\varepsilon_{ij} \sim N(0,\Theta_\varepsilon)$. Typically, researchers will assume the residuals are heteroscedastic and independent over time (Grimm & Widaman, 2010), where the form of the covariance matrix Θ_ε will be defined as:

$$\Theta_\varepsilon = \begin{bmatrix} \sigma^2_{\varepsilon_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma^2_{\varepsilon_j} \end{bmatrix}$$

The presence of zeros in the matrix indicates that the errors are not intercorrelated, implying that the residuals are independent between different time points. Furthermore, the variances along the diagonal are not identical, which suggests the assumption is that residuals are heteroscedastic at distinct times. If the variances along the diagonal are identical here, where $\sigma^2_{\varepsilon 1} = \sigma^2_{\varepsilon 2} = \ldots = \sigma^2_{\varepsilon j} = \sigma^2_\varepsilon$, it will result in the same assumption as in multilevel modeling. Notably, the assumption of error terms in LGCM can always be inspected in the path diagram, while researchers must provide the diagram describing the causal relationships and theories in their studies (Hancock et al., 2010). As for the example in Figure 2, the error term of depression is assumed to be heteroscedastic and independent.

After discussing the error term in the measurement model, another essential element in the equation is the vector of the latent growth factors $\boldsymbol{\eta}$, where it is the intercept and slope factor in Figure 2. The latent growth factors account for the trajectories of the repeated measured variable, which is the purpose of applying the longitudinal analysis. To obtain the trajectories for each individual, consider the following simultaneous equations:

$$\pi_{0i} = \mu_{\pi_0} + \varsigma_{0i} \ and \ \pi_{1i} = \mu_{\pi_1} + \varsigma_{1i}$$

where $\mu_{\pi 0}$ and $\mu_{\pi 1}$ are the mean of intercept and slope factor across all cases. The disturbances, ς_{0i} and ς_{1i}, explain how much the specific individual i deviates from the average of the population. In particular, the simultaneous equations assume that no predictors will influence latent growth factors, so the model is called the "unconditional latent growth curve model." It can be extended to include time-invariant variables as well, where

$$\pi_{0i} = \mu_{\pi_0} + \gamma_{\pi_{0i}} X_i + \varsigma_{0i} \ and \ \pi_{1i} = \mu_{\pi_1} + \gamma_{\pi_{1i}} X_i + \varsigma_{1i}$$

These two pairs of simultaneous equations are the form of level-2 equations in multilevel modeling. It also satisfies the basic formulation of the structural model in SEM:

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \boldsymbol{\Gamma\xi} + \boldsymbol{B\eta} + \boldsymbol{\varsigma},$$

$$where\ \boldsymbol{\eta} = \begin{bmatrix} \pi_{0i} \\ \pi_{1i} \end{bmatrix} = \begin{bmatrix} \mu_{\pi_0} \\ \mu_{\pi_1} \end{bmatrix} + \begin{bmatrix} \varsigma_{0i} \\ \varsigma_{1i} \end{bmatrix}, \boldsymbol{\alpha} = \begin{bmatrix} \mu_{\pi_0} \\ \mu_{\pi_1} \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, and \begin{cases} \boldsymbol{\Gamma} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} for\ unconditional\ LGCM \\ \boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{\pi_{0i}} \\ \gamma_{\pi_{1i}} \end{bmatrix} \ for\ LGCM\ with\ one\ predictor \end{cases}$$

The latent construct $\boldsymbol{\xi}$ is estimated by fixing the parameters in the X-measurement model that $\delta = [0]$, $\tau_x = [0]$, and $\Lambda_x = [1]$. Particularly, the disturbance vector $\boldsymbol{\varsigma}$ has a covariance matrix $\boldsymbol{\Psi}$:

$$\boldsymbol{\Psi} = cov(\boldsymbol{\varsigma}) = \begin{bmatrix} \sigma^2_{\pi_0} & \sigma_{\pi_0 \pi_1} \\ \sigma_{\pi_1 \pi_0} & \sigma^2_{\pi_1} \end{bmatrix}$$

The matrix $\boldsymbol{\Psi}$ contains the variance and covariance of the latent growth factors, which represents the relationship between these factors. This facilitates researchers in querying, as depicted in Figure 2, whether the initial state of depression influences its subsequent trajectory or alteration. Other detailed information and derivation for the above equations about LGCM can be found in Newsom (2015) and Singer and Willett (2003).

Estimating an LGCM

Data Requirement

Although there are few strict requirements for the types of data that might be analyzed using growth models, several general data characteristics are particularly amenable to these methods. LGCM necessitates longitudinal data, meaning data collected at multiple time points from the same subjects. Typically, at least three repeated measures per individual are required to reliably estimate growth trajectories (Duncan & Duncan, 2009; Xitao & Xiaotao, 2005), though more time points (e.g., five or more) are preferred to enhance the model's ability to detect non-linear patterns (Preacher et al., 2008). An adequate sample size is also crucial, though what constitutes "adequate" can vary depending on the complexity of the model and the amount of variance explained. While growth models have been fitted to small samples, sample sizes of at least 100 are often preferred, with the total number of person-by-time observations playing a critical role in model estimation and statistical power (Curran et al., 2010; Lei & Lomax, 2005).

Measurement invariance is essential; the same variables must be measured in the same way across all time points to ensure observed changes reflect true changes in constructs rather than variations in measurement (Hancock et al., 2010; Preacher, 2008). That is, the meaning of the measurement would not change over time (Hancock et al., 2010). LGCM typically assumes continuous and normally distributed repeated measures, but alternative estimation methods can accommodate non-normally distributed, discrete, or ordinal measures (Lee et al., 2018; Mehta et al., 2004; Mehta & West, 2000). Additionally, LGCM can handle partially missing data using techniques like Full Information Maximum Likelihood (FIML) or multiple imputation, provided the data are missing at random (MAR) or missing completely at random (MCAR) (Enders & Bandalos, 2001). These requirements ensure that researchers can effectively utilize LGCM to model individual growth trajectories and explore factors influencing change over time.

Unconditional model

An unconditional model in LGCM serves as a critical foundational step for both baseline understanding and model comparison (Duncan and Duncan, 2009). Baseline understanding is achieved by estimating the initial trajectory of the outcome variable without the influence of covariates. This allows researchers to grasp the general pattern of change over time and identify key characteristics of the growth process, such as the average initial status and the average rate of change within the population. By understanding these baseline dynamics, researchers can establish a clear picture of the underlying growth structure, which is essential for interpreting subsequent analyses.

In terms of model comparison, the unconditional model acts as a benchmark against which more complex models are evaluated. By incrementally adding covariates or testing alternative growth structures (e.g., quadratic or cubic trends), researchers can assess the improvement in model fit (Bollen & Curran, 2006). This is typically done by applying the chi-squared difference test, details of which will be discussed later. Using fit indices such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) is also helpful in determining whether the added complexity justifies the improvement in fit. Comparing these indices across models facilitates a systematic evaluation of how well different specifications capture the underlying data patterns and guides the selection of the most parsimonious yet explanatory model.

Factor Loadings

Determining factor loadings in LGCM is a critical process, as it significantly affects parameter estimations and the subsequent interpretation of underlying growth trajectories. The initial step involves deciding on the number of pairs of factor loadings, which is closely linked to the number of latent growth factors. In general, an LGCM encompasses two primary growth factors: the intercept factor, which represents the initial status, and the slope factor, indicative of the rate of change. However, this does not necessarily mean there are two pairs of factor loadings to be determined.

There are two reasons for this: the first is related to the nature of the intercept factor, and the second concerns the theories or assumptions for the number of slope factors. For the first reason, since the intercept factor represents the initial status of the response variable, determining the factor loadings for the intercept factor is quite intuitive. The intercept factor is a constant value for each individual over time, as depicted in Figure 2, so the values are fixed to 1 to ensure uniformity in the initial value of depression status at each time point for everyone. Due to this characteristic, the factor loadings of the intercept factor are not required to be determined or estimated.

Secondly, the number of slope factors is determined by the theories or assumptions, so it can be greater than 1, especially when modeling polynomial trajectories. Typically, researchers begin by investigating the trajectory in a linear form, so the model will contain only one intercept factor and one slope factor. However, this changes in a non-linear setting. For instance, if the change is assumed to be quadratic, there will be two slope factors to capture the trend: one representing the linear form and the other represents the quadratic form. Overall, the number of pairs of factor loadings is deeply affected by the assumptions from the researchers, which also suggests that theories are essential for determining factor loadings (Preacher, 2008).

The other step in determining factor loadings is to confirm the values of the factor loadings, contingent upon four critical aspects: the time point of most interest, the direction of the trend, mathematical form, and data collection.

First, the time point of most interest is crucial as the chosen loadings can highlight specific periods within the growth trajectory, such as the baseline status or significant developmental phases that are theoretically or practically important. For instance, in clinical studies, the onset of treatment is often a significant phase, and tracking changes from this baseline can inform the efficacy of therapeutic interventions. Consider a set of factor loadings, where $T_j = \{t_1, t_2, t_3, t_4\}$, as shown in Figure 2. In general, T_j will be assigned as $\{0,1,2,3\}$ to express a linear form, emphasizing the baseline status at the first time point (i.e. t_1=0 at time 1) with a positive growth trend. If a study is interested in the last time point, T_j can be assigned as $\{-3,-2,-1,0\}$ to maintain the trend and rescale the baseline status to t_4.

Second, regarding trend direction, whether positive or negative, it can be altered by changing the signs of factor loadings. For example, in the set of factor loadings, $T_j = \{0,1,2,3\}$, which shows a positive trend, it can be switched to a negative direction that $T_j = \{0,-1,-2,-3\}$. It should be noted that the first two aspects will not impact the fit of the model. However, it is crucial to acknowledge that transforming $\mathbf{\Lambda_y}$ bears significant implications for the interpretation of parameters (Mehta & West, 2000).

Third, as previously discussed, the mathematical assumption not only affects the decision on the number of slope factors but also restricts the choice of the values in factor loadings. In particular, the factor loadings of higher-order polynomials are often determined by the linear factor loadings in the polynomial setting. Consider an LGCM with both linear and quadratic slope factors, where the linear factor loadings T_j are {0,1,2,3}. In this situation, the set of quadratic factor loadings is calculated by taking the square of the linear factor loadings, resulting is {0,1,4,9}. Researchers do not need to specify new factor loadings for the quadratic term. Additionally, researchers can utilize orthogonal polynomials to minimize the covariance among latent growth factors (Newsom, 2015).

Lastly, the aspect regarding data collection of the repeated measures is crucial. Generally, the response variables are collected at evenly spaced time points, with linear factor loadings such as {0,1,2,3} assigned to them. This assumes that the response variable changes uniformly over time. Nevertheless, if the data were gathered at irregular intervals, these linear factor loadings would result in non-linear changes. The solution for the violation of uniform change is to adjust the linear factor loadings based on irregular time points. For instance, if longitudinal data were collected in 2000, 2001, 2004, and 2010, the linear slope factor loadings can be set to {0,1,4,10}, which affirms the uniformity of annual change.

Following the discussion on the steps of determining appropriate factor loadings in LGCM for researchers, it should be noted that a more data-driven method exists for deciding the loadings by freely estimating the parameters (Meredith & Tisak, 1990). This method is called the "unspecified trajectory model." One of the primary benefits of the unspecified trajectory model is its ability to adapt to various growth patterns without the constraints of predefined functional forms (Bollen & Curran, 2006; Meredith & Tisak, 1990; Newsom, 2015; Stoolmiller, 1995). This adaptability makes it suitable for analyzing longitudinal data where the growth trajectories are expected to be complex or non-linear. By allowing the data to dictate the trajectory's shape, this model can highlight significant developmental phases and other critical periods, offering richer insights into the processes under study.

The unspecified trajectory can be implemented by only fixing t_1 and t_2 to predetermined values, 0 and 1, respectively. The other factor loadings are estimated by the following equation:

$$t_j = \frac{\overline{y}_j - \overline{y}_1}{\overline{y}_2 - \overline{y}_1}$$

where ȳ represents the observed mean of the repeated measure at the corresponding time point denoted by the subscript. The estimated loading uses the change at the second time point as the reference to assess the difference between the specific time point and the initial value, indicating that the factor loadings act as multipliers of the change between time 1 and time 2. In fact, this approach stretches the unit of time and allows for linearizing the relationship between time and the response variables. McArdle (1988) extends this application by fixing t_1 and t_2 to 0 and 1. This adjustment allows the estimated loadings to accurately depict the average change in relation to the overall differential (McArdle, 1988).

While the unspecified trajectory model offers significant advantages, its application requires careful attention to several practical considerations to ensure reliable and valid results. One of the primary challenges is the need for an adequate sample size. Because the model involves estimating a larger number of parameters, larger sample sizes are essential for model identification to achieve stable and reliable estimates (Bollen & Curran, 2006; Hancock et al., 2010; Newsom, 2015). An insufficient sample size can lead to convergence issues and unreliable parameter estimates, undermining the model's validity. Additionally, researchers must monitor model convergence closely. Non-convergence or improper solutions, such as negative variances, can indicate issues with model specification or data inadequacies. Conducting robustness checks and sensitivity analyses is recommended to ensure that the results are not overly dependent on a particular specification and that the model is robust across different configurations (Hancock et al., 2010).

Another critical aspect is the interpretation of parameters. In unspecified trajectory models, the factor loadings are freely estimated, which can complicate the interpretation of intercepts and slopes compared to models with specified trajectories. This is particularly true for the completely latent trajectory model, which contains only one intercept factor and one slope factor. In this model, the estimated slope factor is presumed to represent both linear and nonlinear change, making it difficult to distinguish between these two trends with a single factor (Stoolmiller, 1995). Any interpretation of this slope factor could be ambiguous before eliminating the linear effect (Hancock et al., 2010). Another significant issue with interpretation is the overfitting of the data, where stochastic fluctuations could influence the shape of the function and lead to spurious findings. Researchers may perceive the estimated factor loadings as random numbers or a nonlinear trajectory, while, in fact, they could be simply explained by a linear trend. This is because the mean value of the slope factor will remain significant even when only a linear trend is observed (Newsom, 2015). A strategic approach to mitigate these issues involves systematically ruling out the linear effect within the slope factor (Newsom, 2015). This process entails incorporating a linear factor into the analysis and conducting a significance test to ascertain whether the original factor yields any additional insights beyond those provided by a linear trend. The test for evaluation can be conducted by the Wald ratio test of mean values of the original slope factor or by employing a likelihood ratio test to compare models—with and without the latent basic factor—while including a linear slope. Additional details about model comparison will be discussed in the next section.

Model fit and comparison

Assessing the fit of an LGCM involves using various statistical indices to gain a comprehensive understanding of how well the model represents the observed data. One of the measures is the Chi-Square Test of Model Fit, which compares the observed covariance matrix with the model-implied covariance matrix. A non-significant chi-square value (*p-value* > 0.05) suggests a good fit, indicating minimal differences between the observed and expected matrices. However, this test is highly sensitive to sample size; in large samples, even minor deviations can result in significant chi-square values, which may not necessarily imply poor model fit. Consequently, in recent years, chi-square values have not been the main metric for evaluation in practice (Lei & Lomax, 2005).

Another important index is the Root Mean Square Error of Approximation (RMSEA), which evaluates the goodness of fit per degree of freedom, considering model complexity (Steiger, 1980). RMSEA values below 0.05 indicate a close fit, while values up to 0.08 are considered reasonable (Browne & Cudeck, 1992). RMSEA adjusts for the sample-size effect, making it a preferred measure for complex models. It also provides a confidence interval, offering a range within which the true RMSEA value is expected to lie, adding to its robustness.

Additionally, the Comparative Fit Index (CFI), Incremental Fit Index (IFI), and Tucker-Lewis Index (TLI) compare the fit of the specified model to a null model (Bentler, 1990; Bollen, 1989; Tucker & Lewis, 1973). CFI and IFI values above 0.90 or 0.95 indicate a good fit. TLI values close to 1 are ideal, while values lower than 0.9 are considered inadequate. TLI values greater than 1.2 suggest an overfitting problem. Unlike the chi-square test, CFI and TLI are less sensitive to sample size, providing a more stable assessment of model fit. The Standardized Root Mean Square Residual (SRMR) measures the difference between observed and predicted correlations, with values below 0.08 typically indicating a good fit. However, SRMR is not sensitive to misfitted mean structures and has relatively insufficient power to identify incorrect functional forms with few time points (Yu, 2002).

Overall, there is no single standard for evaluating an LGCM, and these indices could be inconsistent in some situations. If discrepancies occur, researchers should consider alternative assumptions and models (Felt et al., 2017). The evaluation of the model should not be done in isolation but as part of a comprehensive process that includes comparing alternative models, considering theoretical plausibility, and examining parameter estimates for practical significance (Hancock et al., 2010; Newsom, 2015). This approach helps ensure that the chosen model not only fits the data well statistically but also makes meaningful and theoretically sound contributions to the understanding of the phenomena under study.

There are two different techniques for comparing different LGCMs: information-based metrics and the significance test. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are commonly used metrics that assess model fit while penalizing for the number of parameters, thus discouraging overfitting (Akaike, 1987; Schwarz, 1978). Lower AIC and BIC values indicate a better model, with BIC imposing a larger penalty for model complexity compared to AIC, making it more conservative. The Likelihood Ratio Test, also known as the “chi-square difference test,” is another essential method for comparing models (Chen et al., 2001). This test compares the fit of two nested models by evaluating whether the addition of parameters significantly improves fit. A significant result indicates that the more complex model fits the data significantly better than the simpler model, justifying the inclusion of the additional parameters. Through this comparison, researchers can determine if the more complex model offers substantial advantages.

Covariates

As discussed in the previous section, when an unconditional model, or baseline model, has been confirmed, a more complex model with one or more predictors can be examined with reference to the baseline growth. Predictors, or covariates, in LGCM can broadly be categorized into two types: Time-Invariant Covariates (TICs) and Time-Varying Covariates (TVCs). TICs, as the name suggests, remain constant across time points and typically capture individual differences that persist over the duration of the study. These could include demographic variables like gender, socioeconomic status, or cultural background, as well as stable personality traits or genetic factors. On the other hand, TVCs are variables that may fluctuate over time, influencing the trajectory of growth at each measurement occasion. These could encompass situational factors such as stress levels, environmental conditions, or treatment interventions.

Incorporating covariates into LGCM enables researchers to test hypotheses about the underlying mechanisms driving change. TICs, for example, directly predict individual differences in initial status (intercept) and rates of change (slope) by identifying which characteristics are associated with higher or lower starting points and steeper or flatter slopes over time. TVCs, on the other hand, offer insight into the dynamic interplay between time-varying factors and growth processes. There are two major ways to incorporate TVCs in LGCM: the traditional approach and multivariate latent growth. Each method provides a different interpretation for the corresponding research questions. The traditional approach involves predicting the response variable as a distinctive predictor in the model expression, while the effect of TVCs is the prediction of the response variable after controlling the influence of underlying growth trajectories (Bollen & Curran, 2006). Instead, the multivariate approach considers the trajectories of TVCs as the predictor while answering the question of how the trajectory of one variable can influence the trajectory of the other one (Duncan et al., 1998).

Extensions

LGCM provides a flexible framework that allows researchers to address a wide range of research questions through the development of various extensions. In this section, we will introduce two specific extensions: the piecewise model and the growth mixture model. These models are designed to investigate multiple trajectories, with the former focusing on within-individual trajectories and the latter on between-individual trajectories.

A piecewise latent growth curve model is a statistical approach that divides the overall time span into segments, with each segment having its own growth parameters. This method is particularly useful when the data indicates that the rate of change is not constant throughout the entire observation period, but instead varies at specific points. These points, referred to as "knots" or "breakpoints," mark where the trajectory changes direction or slope and can be determined by either pre-specified criteria or data-driven methods (Harring et al., 2021). Identifying and placing these knots allows the piecewise model to more accurately capture complex growth patterns that a single continuous trajectory might miss. This approach is advantageous in scenarios where distinct phases of growth or decline are evident (Chou et al., 2004). For example, in educational research, a student's learning curve might experience rapid acceleration in the initial stages, slow down during the middle phase, and then pick up again towards the end.

A Growth Mixture Model extends the traditional LGCM by incorporating latent classes, or subgroups, within the population (Berlin et al., 2014; Ram & Grimm, 2009). While traditional LGCMs assume that all individuals in a sample follow a single trajectory with common growth parameters, this assumption may not hold true in heterogeneous populations. Growth Mixture Models (GMMs) address this limitation by allowing for the identification of distinct subgroups within the population, each with its own unique growth trajectory. The model simultaneously estimates the parameters for the growth trajectories within each class and the probabilities of class membership for each individual. This dual focus enables researchers to identify distinct growth patterns and to classify individuals based on their developmental trajectories. This approach provides a more nuanced understanding of developmental processes and can reveal underlying patterns that are not apparent when assuming a homogeneous population. For instance, reading proficiency development is a critical aspect of early education, with substantial variability among students. Some students may start with high proficiency and continue to improve steadily, while others may struggle initially but show significant improvement later. Identifying these subgroups and understanding their unique trajectories can help educators design targeted interventions.

In summation, the piecewise model captures changes within individual trajectories by segmenting the time span into phases with distinct growth parameters, while the growth mixture model identifies distinct subgroups with unique growth patterns in heterogeneous populations. These two extensions can be combined to explore more complicated trajectories (Kohli et al., 2013). These approaches enhance our understanding of developmental processes and inform targeted interventions by revealing nuanced growth trajectories that single continuous models might miss.

Conclusion

Latent Growth Curve Modeling (LGCM) has emerged as a powerful and flexible statistical technique for analyzing longitudinal panel data. By incorporating latent variables that represent individual growth trajectories over time, LGCM enables researchers to model both the initial status and the rate of change, while accounting for individual variability in these trajectories. With the increasing availability of longitudinal panel datasets, LGCM presents a valuable opportunity for researchers across various disciplines to explore individual change within the broader context of macrostructural dynamics.

The key strengths of LGCM lie in its ability to capture heterogeneity in growth processes, accommodate complex causal relationships, and incorporate covariates that influence these developmental pathways. As with any statistical technique, the application of LGCM requires careful consideration of data requirements, model assumptions, and interpretation of results. However, by successfully integrating its theoretical foundations, LGCM offers a robust framework for advancing our understanding of complex developmental processes. Furthermore, LGCM offers significant advantages over traditional analytical approaches by treating individual variability as a critical component of the model rather than merely an error term. By leveraging the flexibility of LGCM and its extensions, such as piecewise models and growth mixture models, researchers can gain nuanced insights into the factors shaping human behavior and development over the lifespan.

The study "Developmental Trajectory of Depressive Symptoms from Adolescence to Early Adulthood," effectively utilized LGCM to analyze the trend of depression among Taiwanese adolescents. The author identified three key stages – freshman year of high school, senior year of high school, and sophomore year of college – by including a slope factor with unspecified loadings. The study provides valuable insights into the role of parent-child relationships and self-esteem in shaping the mental health trajectories of adolescents in Taiwan, enhancing the potential for developing effective preventive strategies and interventions.

Further studies can explore and justify the stage theory among Taiwanese adolescents by applying a piecewise model or a mixture model, allowing researchers to investigate the mechanisms of transitions and differentiate the effect. Essential questions include whether the trajectories in each stage remain linear or whether individuals share the same transition pattern. Additionally, scholars may explore the impacts of transitioning between parenting styles or modifying self-esteem levels on the progression of depressive disorders through the application of the multivariate latent growth model.

For those seeking to expand their understanding of LGCM, numerous researchers have produced helpful tutorials and discussions on the mathematical theories (Berlin et al., 2014; Bollen & Curran, 2006; Curran, 2000; Curran & Bauer, 2011; Duncan & Duncan, 2004; Hedeker & Gibbons, 2006; Newsom, 2015; Preacher, 2008; Singer & Willett, 2003), as well as the software (Klopack & Wickrama, 2020; Mirman, 2017; Rabe-Hesketh & Skrondal, 2008). Several statistical software options are available for LGCM, including Stata, R, and Mplus. It is important to note that LGCM is not always a built-in function in these software programs. For example, in R, users must install the "lavaan" package for SEM and LGCM (Rosseel, 2012). Overall, Latent Growth Curve Modeling represents a valuable addition to the methodological toolkit for longitudinal research, enabling researchers to delve deeper into the intricate patterns of change that characterize human behavior and experience across various domains.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*, 317-332.

Barnes, G. M., Reifman, A. S., Farrell, M. P., & Dintcheff, B. A. (2000). The effects of parenting on the development of adolescent alcohol misuse: A Six-Wave latent growth model. *Journal of Marriage and Family, 62*(1), 175-186.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238.

Berlin, K. S., Parra, G. R., & Williams, N. A. (2014). An introduction to latent variable mixture modeling (part 2): Longitudinal latent class growth analysis and growth mixture models. *Journal of Pediatric Psychology, 39*(2), 188-203.

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research, 17*(3), 303-316.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. John Wiley & Sons.

Box, G. E. (1950). Problems in the analysis of growth and wear curves. *Biometrics, 6*(4), 362-389.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*(2), 230-258.

Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods & Research, 29*(4), 468-508.

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*(2), 213-225.

Chou, C. P., Bentler, P. M., & Pentz, M. A. (1998). Comparisons of two statistical approaches to study growth curves: The multilevel model and the latent curve analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 5*(3), 247-266.

Curran, P. J. (2000). A latent curve framework for the study of developmental trajectories in adolescent substance use. In J. S. Rose, L. Chassin, C. C. Presson, & S. J. Sherman (Eds.), *Multivariate applications in substance use research: New methods for new questions* (pp. 1-42). Lawrence Erlbaum Associates Publishers.

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology, 62*, 583-619.

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development, 11*(2), 121-136.

Dmitriew, C. M. (2011). The evolution of growth trajectories: What limits growth rate? *Biological Reviews, 86*(1), 97-116.

Duncan, O. D. (2014). *Introduction to structural equation models*. Elsevier.

Duncan, S. C., Duncan, T. E., Biglan, A., & Ary, D. (1998). Contributions of the social context to the development of adolescent substance use: A multivariate latent growth modeling approach. *Drug and Alcohol Dependence, 50*(1), 57-71.

Duncan, T. E., & Duncan, S. C. (2004). An introduction to latent growth curve modeling. *Behavior Therapy, 35*(2), 333-363.

Duncan, T. E., & Duncan, S. C. (2009). The ABC's of LGM: An introductory guide to latent variable growth curve modeling. *Social and Personality Psychology Compass, 3*(6), 979-991.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*(3), 430-457.

Felt, J. M., Depaoli, S., & Tiemensma, J. (2017). Latent growth curve models for biomarkers of the stress response. *Frontiers in Neuroscience, 11*, 271623.

Gee, K. A. (2014). Multilevel growth modeling: An introductory approach to analyzing longitudinal data for evaluators. *American Journal of Evaluation, 35*(4), 543-561.

Grimm, K. J., & Widaman, K. F. (2010). Residual structures in latent growth curve modeling. *Structural Equation Modeling, 17*(3), 424-442.

Hancock, G. R., Mueller, R. O., & Stapleton, L. M. (2010). *The reviewer's guide to quantitative methods in the social sciences*. Routledge.

Harring, J. R., Strazzeri, M. M., & Blozis, S. A. (2021). Piecewise latent growth models: beyond modeling linear-linear processes. *Behavior Research Methods, 53*, 593-608.

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. John Wiley & Sons.

Huck, S. W., & McLean, R. A. (1975). Using a repeated measures ANOVA to analyze the data from a pretest-posttest design: A potentially confusing task. *Psychological Bulletin, 82*(4), 511.

Klopack, E. T., & Wickrama, K. (2020). Modeling latent change score analysis and extensions in Mplus: A practical guide for researchers. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(1), 97-110.

Kohli, N., Harring, J. R., & Hancock, G. R. (2013). Piecewise linear-linear latent growth mixture models with unknown knots. *Educational and Psychological Measurement, 73*(6), 935-955.

Lee, T. K., Wickrama, K., & O'Neal, C. W. (2018). Application of latent growth curve analysis with categorical responses in social behavioral research. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(2), 294-306.

Lei, M., & Lomax, R. G. (2005). The effect of varying degrees of nonnormality in structural equation modeling. *Structural Equation Modeling, 12*(1), 1-27.

Mayer, K. U. (2009). New directions in life course research. *Annual Review of Sociology, 35*, 413-433.

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In Nesselroade, J. R. & Cattell, R. B. (Eds.) *Handbook of multivariate experimental psychology* (pp. 561-614). Springer.

Mehta, P. D., Neale, M. C., & Flay, B. R. (2004). Squeezing interval change from ordinal panel data: latent growth curves with ordinal outcomes. *Psychological Methods, 9*(3), 301.

Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods, 5*(1), 23.

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

Mirman, D. (2017). *Growth curve analysis and visualization using R*. Chapman and Hall/CRC.

Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction.* Routledge.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879-903.

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

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata*. STATA press.

Ram, N., & Grimm, K. J. (2009). Methods and measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development, 33*(6), 565-576.

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50*, 203-228.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1-36.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461-464.

Shanahan, M. J. (2000). Pathways to adulthood in changing societies: Variability and mechanisms in life course perspective. *Annual Review of Sociology, 26*(1), 667-692.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford university press.

Steiger, J. H., & Lind, J. C. (1980). *Statistically based tests for the number of common factors*. [Paper presentation]. The Annual Meeting of the Psychometric Society, Iowa City.

Stoolmiller, M. (1995). Using latent growth curve models to study developmental processes. In J. M. Gottman (Ed.), *The analysis of change* (pp. 103-138). Psychology Press.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1-10.

Wu, C. Y., Chang, M. Y., & Chen, Y. C. (2008). Looking for the mechanism and the process: The function of panel study (尋找機制與過程：長期追蹤研究的功用). *αβγ of the Journal for Quantitative Research* (**αβγ 量化研究學刊**)*, 2*(1), 1-26.

Xitao, F., & Xiaotao, F. (2005). Power of latent growth modeling for detecting linear growth: Number of measurements and comparison with other analytic approaches. *The Journal of Experimental Education, 73*(2), 121-139.

Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* [Unpublished doctoral dissertation]. University of California. https://www.statmodel.com/download/Yudissertation.pdf