8-2017

A Linear-Linear Growth Model with Individual Change Point and its Application to ECLS-K Data

Ping Zhang
University of Arkansas, Fayetteville

A Linear-Linear Growth Model with Individual Change Point and its Application to ECLS-K Data

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Statistics and Analytics

by

Ping Zhang
Wuhan Yangtze Business University
Bachelor of Science in Finance, 2013

August 2017
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

Dr. Qingyang Zhang
Thesis Director

Dr. Giovanni Petris
Committee Member

Dr. Avishek Chakraborty
Committee Member

Dr. Wenjuo Lo
Committee Member

ABSTRACT

The latent growth curve model with piecewise functions is a useful analytics tool to investigate the growth trajectory consisted of distinct phases of development in observed variables. An interesting feature of the growth trajectory is the time point that the trajectory changes from one phase to another one. In this thesis, we propose a simple computational pipeline to locate the change point under the linear-linear piecewise model and apply it to the longitudinal study of reading and math ability in early childhood (from kindergarten to eighth grade). In the first step, we conduct the hypothesis testing to filter out the samples that do not exhibit a change point. For samples with significant change point, we use the maximum likelihood estimation(MLE) to determine the location of a change point. However, a small portion of samples contains abnormal observations, which makes the MLE method fail to identify the change point. To overcome this difficulty, we apply a Bayesian approach to locate the change point for these samples. By comparison of the change point distributions in math and reading, as well as students with different overall performance, we conclude that: (a) most students have change points between Spring-first grade and Spring-third grade; (b) students with overall better performance have change point at earlier stage; (c) compared with math, the change point distribution for reading is more concentrated between Spring-first grade and Spring-third grade.

TABLE OF CONTENTS

1.INTRODUCTION

This chapter reviews the research on longitudinal studies in different fields and the application of piecewise latent growth model. Modeling developmental processes have attracted a great deal of attention since the 1950s. The outcome measurements of interest are often formulated as a function of time. Longitudinal data is this type of data that are measured for same subjects repeatedly across time. And researchers often use Longitudinal data to explore the developmental trajectory.

Longitudinal studies have broad applications in many scientific domains (e.g., in psychology, medicine, and sociology). It allows researchers to investigate how individual's performances, interests, and attitudes change over time. Longitudinal studies can yield valuable information in behavioral science studies. Researchers could explore developmental trends and individual trajectory differences over time. Longitudinal studies can assist researchers to identify whether the growth for one latent class greater than another one is due to the result of treatment effect (Shin, Davison, Long, Chan, & Heistad, 2013), or to decide whether covariates can predict the change process (Grimm, 2008; Jordan, Hanich, & Kaplan, 2003; Miles, & Miles, 1992).

In many psychological and educational research, longitudinal processes exhibit distinct phases of development in observed repeated measurements (Kreisman, 2003; Paris, 2005; Silverman, Speece, Harring, & Ritchey, 2012). For example, Kreisman's study on earlier childhood intervention elucidated that children with pre-intervention (as well as children without pre-intervention) have different academic growth pattern. However, past evaluations have implied that children with pre-intervention have homogenous growth pattern. In another example, where two distinct phases are found in the reading literature longitudinal study (Silverman, Speece, Harring, & Ritchey,2012). This study shows that most students' ability to accurately and automatically decode may increase at the beginning of second grade, then change to a relatively slower rate in the middle of their third grade.

In the past three decades, several statistical methods have been developed to analyze longitudinal data, such as mixed effects model (Laird, & Ware, 1982), multilevel models (Goldstein, 2003), as well as latent growth curve models (LGC model) (Meredith, & Tisak, 1990). There are several advantages of the latent growth model. It allows the investigation of individual difference, as well as the causes and consequences of change. LGC model provides a straightforward method to compare growth difference in group-level. Last but not the least, it is a flexible model that could be adapted to different requirements. Depending on the characteristics of the data, one could specify different types of piecewise Latent Growth Models. For instance, if the trajectories in the first and second phases are both linear, a linear-linear piecewise LGC model could be adapted. Similarly, a quadratic-linear or exponential-linear piecewise LGC model could be specified when the first phase has some curvature. In this paper, we will focus on the linear-linear piecewise LGC model.

Piecewise growth models have become a prevailing tool in many different scientific domains due to its flexibility and computational tractability. Especially, the piecewise growth models have been widely applied in psychology. For instance, the development of cognitive function in old age is often non-linear, and the age when a change occurs may vary among individuals. Several researchers have investigated the properties of the change point with two linear phases for each subject (Dominicus, Ripatti, Pedersen, & Palmgren, 2008). And by fitting piecewise growth model, a study finds that the prevention program has a marginal effect on reducing the prevalence rate of cigarette use (Chou, Yang, Pentz, & Hser, 2004).

When there exist two distinct unknown phases, piecewise LGC models could be utilized to evaluate a specified functional form of the overall change process and to the identification of different phases (Chou, Yang, Pentz, & Hser, 2004; Cudeck, & Harring, 2010). In this respect, the piecewise latent growth model can summarize various functional forms in the different phases of development such that each phase follows its functional form.

The rest of this paper is organized as follows: in Chapter 2, we briefly summarize the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K), and provide some background information through the development of latent growth model. In Chapter 3, we propose the algorithms to filter out samples, as well as identify the location of change point. In Chapter 4, we apply the algorithms to K-8 Public-Use data and interpret the results from the analysis.

2.LITERATURE REVIEW

2.1 Early Childhood Longitudinal Study, Kindergarten Class of 1998-99

The ECLS-K is a large-scale longitudinal study (Tourangeau, Nord, Lê, Sorongon, & Najarian, 2009) and focused on children's early school experiences beginning with kindergarten and following children through eighth grade. This program is conducted by National Center for Education Statistics (NCES). The children participated in ECLS-K come from both public and private schools and attend both full-day and part-day kindergarten programs. The participants are from diverse socioeconomic and racial backgrounds. The children's parents, teachers, and schools across the United States also participate in this study.

The ECLS-K collected information from children and their parents, teachers, and schools. A variety of methods were used to collect information, including one-on-one assessment, computer-assisted telephone interviews, self-administered paper- pencil questionnaires. It collected information about children's reading and mathematics skills in each round of data collection, and their general knowledge (i.e., science and social studies) in kindergarten and first grade, and their science knowledge in third, fifth, and eighth grades (Tourangeau,et al., 2009). A total of 21,409 kindergarteners throughout the nation participated. The data were collected in the fall and the spring of kindergarten (1998-1999), the fall and spring of 1st grade (1999-2000), the spring of 3rd grade (2002), the spring of 5th grade (2004), and the spring of 8th grade (2007), with a total of 7 round measurements (Table 1.1.1).

Data Collection	Date of collection
Fall-kindergarten	Fall 1998
Spring-kindergarten	Spring 1999
Fall-first grade	Fall 1999
Spring-first grade	Spring 2000
Spring-third grade	Spring 2002
Spring-fifth grade	Spring 2004
Spring-eighth grade	Spring 2007

Table 1.1.1

The ECLS-K assessment frameworks were derived from national and state standards, including National Assessment of Educational Progress (NAEP), the National Council of Teacher of Mathematics, National Academy of Science, and some of from the state assessments (Tourangeau, et al., 2009). The ECLS-K assessments also included items that are specially created for ECLS-K study, and some items are from National Center for Education Statistics (NCES).

The K-8 Public-Use file we use in this paper is preprocessed by NCES so that it could be directly analyzed to explore children's growth and development between kindergarten and eighth grade (Tourangeau, et al., 2009). NCES takes steps to minimize the likelihood that an individual school, teacher, parent, or child participating in this study can be identified, to protect the identity of individual respondents. This study was designed to provide comprehensive and reliable data that can be used to understand the children's development and experiences from elementary to middle school. The dataset used in this study is K-8 Public-Use data. The K-8 (from kindergarten [Fall 1998] to 8th grade [Spring 2007]) full sample Public-Use data includes all children with at least one of the seven rounds of the data collection, from fall-kindergarten to spring- 8th grade. In K-8 full sample public use data file, the scores represent underlying ability (which is normally distributed at all rounds). The scores distribution range is approximately from -3 to 3 (Tourangeau, et al., 2009).

2.2 Latent Growth Curve Model (LGC)

Latent growth curve modeling is a statistical technique that is often used in the structural equation modeling (SEM) framework to estimate growth trajectory over a period. Longitudinal models are stemming from the factor analysis tradition. LGC models assume that the overall change process over time in observed repeated measurements can be described by an underlying latent class (Meredith, & Tisak, 1990). Latent growth model permits straightforward examination of intraindividual (within-person) change over time as well as interindividual (between-person) variability in intraindividual change. Latent growth model allows researchers to adjust the model, as well as to investigate into antecedents and consequents of changes. (Preacher, Wichman, MacCallum, & Briggs, 2008).

A regular LGC model contains a set of observed variables and a relatively small set of latent variables. The latent variables, often serving as the regression weights, are related to measured variables in certain forms. The measured variables can represent the latent variables in turn. The intercept factor represents the level of the outcome measurement when time variable equals zero, and the slope factor represents the linear rate of change of the outcome.

A typical application of latent growth models specifies a function describing a linear change process often composed of two latent growth factors: (a) an intercept that describes the initial status and (b) a slope that summarizes change over time. The intercept and slope parameters are assumed to be random, following a specific joint distribution (Duncan, Duncan, & Strycker, 2006). As to a fully specified latent growth curve model, the loadings from the intercept factor to each of the repeated measures are fixed to be 1, which means the intercept factor equally contributes to all repeated measures. For the slope factor, the loadings are either fixed to a particular value under the linear trajectory, or, can be used to dictate the individual growth in an unspecified trajectory latent growth curve model (Hancock, Harring, & Lawrence, 2013; Meredith, & Tisak, 1990).

The basic formulation of an LGC model includes two components: (a) a measurement model to connect the observed indicators and the latent factors and (b) a structural regression model to describe the means and variances of the latent factors (Duncan, et al., 2006):

$$y_{ij} = \eta_{1i} + \eta_{2i} t_j + \varepsilon_{ij} \tag{2.2.1}$$

$$\eta_{1i} = \alpha_1 + \varsigma_{1i} \tag{2.2.2}$$

$$\eta_{2i} = \alpha_2 + \varsigma_{2i} \tag{2.2.3}$$

Eq. (2.2.1) is the measurement model of the latent growth model. Considering a set of repeated measures of a random variable y for individual i at a different time, y_{ij} refers to the measurement of a random variable y (e.g., individual's response to a certain character) for individual i at time t_j. The responses are observed on a set of repeated measurement occasions $t_j = (1, \dots, T_j)'$, where T_j is the total number of observations. And η_{1i} and η_{2i} are the corresponding intercept and slope factors, ε_{ij} refers to the random error or residual for individual i at j^{th} measurement, which is often assumed to be normally distributed with mean zero.

Eq. (2.2.2) and Eq. (2.2.3) are the structural regression models in the latent growth curve model. Structural regression models provide information about the mean as well as covariance for latent variables. In Eq. (2.2.2) and Eq. (2.2.3), α_1, α_2 refer to the mean value of intercept factor η_{1i} and slope factor η_{2i}, and ς_{1i}, ς_{2i} are individual random variation and covariation around these two latent growth components. As the observations in longitudinal data are collected repeatedly over time, these observations are assumed to be correlated, which could be described by the correlation of intercept η_{1i} and slope η_{2i}. That is equivalent to the correlation of ς_{1i} and ς_{2i}, and $\binom{\varsigma_{1i}}{\varsigma_{2i}}$ is often assumed to be a bivariate normal distribution with zero mean vector.

In LGC model, it is often assumed that the change rate over the entire process is constant. However, this assumption can be violated in the presence of two or more multiple growth phases.

2.3 Piecewise Latent Growth Curve Model

The piecewise LGC model is an extension of the LGC model, which allows the specification of different growth phase to conform to a specified functional form of the overall change process (Chou, Yang, Pentz, & Hser, 2004; Cudeck, & Harring, 2010). For instance, in Kreisman's study of evaluating academic outcomes of Head Start program, the analysis showed that there are two distinct developmental reading and mathematics achievement growth patterns for students with Head Start experience, as well as for students with no preschool experience (Kreisman, 2003). If for each phase, assume the development trajectory is a straight line, and the rate of change is different across phases, piecewise linear-linear LGC models could be used to allow the specification of each growth phase (Cudeck, & Harring, 2010).

Compared with LGC models, the piecewise LGC models offer greater flexibility to model different development trajectories with various functional forms. For instance, a piecewise linear-linear LGC model specifies two straight lines for both first and second growth phases, a piecewise quadratic-linear model could define some curvature in the first stage while in the second stage the rate of change is constant.

An interesting feature of piecewise LGC models is the time point at which the response function transit from one phase to another, known as change point (Cudeck, 1996; Cudeck, & Klebe, 2002). Formulation of a piecewise linear–linear LGC model specifies a separate linear function for each of the two phases of development. The functional form at the time point j is:

$$y_i = f_i + \varepsilon_i$$

where

$$f_{ij} = \begin{cases} \eta_{1i} + \eta_{2i}t_j & t_j \leq \gamma. \\ \eta_{3i} + \eta_{4i}t_j & t_j > \gamma. \end{cases} \quad (2.3.1)$$

In Eq. (2.3.1), the two functions joint when $t_j = \gamma$, known as the change point for a repeated measurement of variable y for individual i. In Eq. (2.2.1), y_i refers to a set of measurements of the variable of y for individual i , η_{1i}, η_{2i} refer to the intercept and the slope growth factor of the first phase, respectively, and η_{3i}, η_{4i} refer to the intercept and slope growth factor for the second phase, respectively. When the two functions intersect at t_j, one of the four parameters become redundant; then we have three free parameters. For example, the intercept in the second phase could be written as ($\eta_{1i} + \eta_{2i}\gamma - \eta_{4i}\gamma$), in this case, f_i can be rewritten as (Cudeck, & Harring, 2010):

$$f_i = \begin{cases} \eta_{1i} + \eta_{2i}t_j & t_j \leq \gamma \\ \eta_{1i} + \eta_{2i}\gamma + \eta_{4i}(t_j - \gamma) & t_j > \gamma \end{cases} \quad (2.3.2)$$

In Eq. (2.3.2), $\boldsymbol{\eta_i}$=$(\eta_{1i}, \eta_{2i}, \eta_{4i})$' is used to represent the growth factors for individual i. And vector $\boldsymbol{\eta_i}$ can be thought of including a fixed effect and a random effect, $\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{\varsigma}_i$. The distribution of $\boldsymbol{\varsigma}_i$ and ε_i are often assumed to be multivariate normal:

$$\boldsymbol{\varsigma}_i \sim N(\mathbf{0}, \boldsymbol{\Psi}) \quad (2.3.3)$$

$$\varepsilon_i \sim N(0, \boldsymbol{\Theta}_i) \quad (2.3.4)$$

$$\begin{pmatrix} \eta_{1i} \\ \eta_{2i} \\ \eta_{4i} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_4 \end{pmatrix} + \begin{pmatrix} \varsigma_{1i} \\ \varsigma_{2i} \\ \varsigma_{4i} \end{pmatrix} \tag{2.3.5}$$

Where $\mathbf{\Theta}_i$ denotes the random effect, and $\mathbf{\Psi}$ denotes the variance-covariance matrix for the random effects. Note that there is a subscript for $\mathbf{\Theta}_i$, but no subscript for $\mathbf{\Psi}$. The reason is that for Piecewise linear-linear model, the underlying assumption is that all individuals share the same change point location.

Even though the piecewise LGC model allows more flexibility to capture the character of developmental growth trajectory compared with LGC model, the assumption inherent to them that all individual from the same population has the same functional form of growth is not practical, especially when considering data come from a mixture of unknown subpopulations.

2.4 Piecewise Linear-linear Latent Growth Mixture Models (LGMMs)

Statistical analysis conducted without considering the heterogeneous population structure may fail to reflect the accurate relationship within the subgroups. In response to the demand of analyzing the developmental trajectories of unobserved subgroups, piecewise linear-linear latent growth mixture model was proposed as a more flexible model. LGMMs infuse the latent classes into piecewise linear-linear LGC model. Within each latent class, there is a distinct piecewise linear-linear growth trajectory. Therefore, LGMMs allow researchers to identify distinct growth trajectories for various latent classes.

To formulate the LGMM, suppose the repeated data are collected from K subpopulations (k=1, …, K), where k refers to the latent class, n observed repeated measurements, $j = 1, \dots, n$ for each individual i. The assumption of this model is that the location of unknown change point is fixed to be the same for all subjects, but potentially different across classes. Assuming there are two classes, the model is specified as:

$$y_{ijk} = \begin{cases} \eta_{1ik} + \eta_{2ik} t_j + \varepsilon_{ijk} & t_j \leq \gamma_k \\ \eta_{3ik} + \eta_{4ik} t_j + \varepsilon_{ijk} & t_j > \gamma_k \end{cases} \tag{2.4.1}$$

For i=1, …, N; $j = 1, \dots, n$ and $k = 1,2$.

In Eq. (2.4.1), y_{ijk} is the observed response of the individual i in the k^{th} class at time j. t_j represents the measurement time; η_{1ik} and η_{2ik} represent the intercept and slope for the first developmental phase in the k^{th} class, respectively; $\eta_{3ik}$and η_{4ik} represent intercept growth factor and slope growth factor for the second developmental phase. γ_k represent the location of change point for the k^{th} class, and ε_{ijk} represent the random error, which is assumed to be normally distributed with mean zero and covariance matrix $\boldsymbol{\Theta}_{ik}$. It is assumed that the residuals are independent with a constant variance, σ_ε^2 , across time, $\boldsymbol{\Theta}_{ik} = \sigma_\varepsilon^2 \mathbf{I}_n$. Since LGMMs assume that all observations within a class share the same change point, there is no subscript 'i' for γ_k .

In Eq. (2.4.1), there are four growth factors. However, given each class has its change point, γ_k, one of the four growth factors could be eliminated. For instance, $\eta_{3ik} = \eta_{1ik} + (\eta_{2ik} - \eta_{4ik})\gamma_k$. Therefore, the structural model component can be specified as:

$$\begin{pmatrix} \eta_{1ik} \\ \eta_{2ik} \\ \eta_{4ik} \end{pmatrix} = \begin{pmatrix} \alpha_{1k} \\ \alpha_{2k} \\ \alpha_{4k} \end{pmatrix} + \begin{pmatrix} \varsigma_{1ik} \\ \varsigma_{2ik} \\ \varsigma_{4ik} \end{pmatrix} \qquad (2.4.1)$$

Where $\boldsymbol{\alpha}_k = (\alpha_{1k}, \alpha_{2k}, \alpha_{3k})^{\,T}$ is a vector of growth factor, and vector $\boldsymbol{\varsigma}_{ik}$ represents the random effect on growth factors. Vector $\boldsymbol{\eta}_{ik}$ is often assumed to be normally distributed with a variance-covariance matrix $\boldsymbol{\Psi}_k$,

$$\boldsymbol{\Psi}_k = \begin{pmatrix} \sigma_{\eta 1}^2 & \sigma_{\eta 1 \eta 2}^2 & \sigma_{\eta 1 \eta 4}^2 \\ & \sigma_{\eta 2}^2 & \sigma_{\eta 2 \eta 4}^2 \\ & & \sigma_{\eta 4}^2 \end{pmatrix}.$$

There are three underlying assumptions with linear-linear PGMMs:

(a) residuals and the latent growth factors are uncorrelated, (cov($\varepsilon_{ik}, \boldsymbol{\eta}_{ik}'$)=0);

(b) residuals and the latent factor residuals are uncorrelated, (cov($\varepsilon_{ik}, \boldsymbol{\varsigma}_{ik}'$)=0);

(c) residuals are uncorrelated with residuals, (cov($\varepsilon_{ik}, \varepsilon_{i'k}$)=0 for $\mathrm{i} \neq i'$).

3.METHODOLOGY

In the ECLS-K data, there are totally seven-time points of data collection. In this article, we only consider the samples without missing values at these seven-time points. Out of the 21,409 children participated in this study, 2,145 samples have complete data for reading ability assessment, and 2913 samples have complete data for mathematical ability assessment.

In this study, we assume that each sample can have at most one change point. It is possible that some participants did not have significant change points. For instance, students may have a stable developmental growth trajectory from kindergarten to eighth grade, or students may have a relatively steep developmental growth trajectory at an earlier stage, then the developmental growth trajectories become stable. Figure 1 shows two samples with different growth patterns.

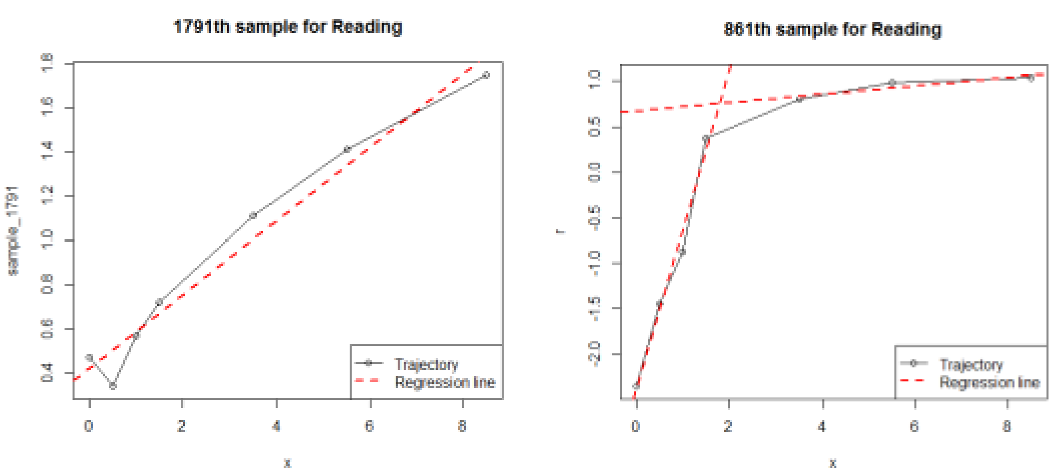
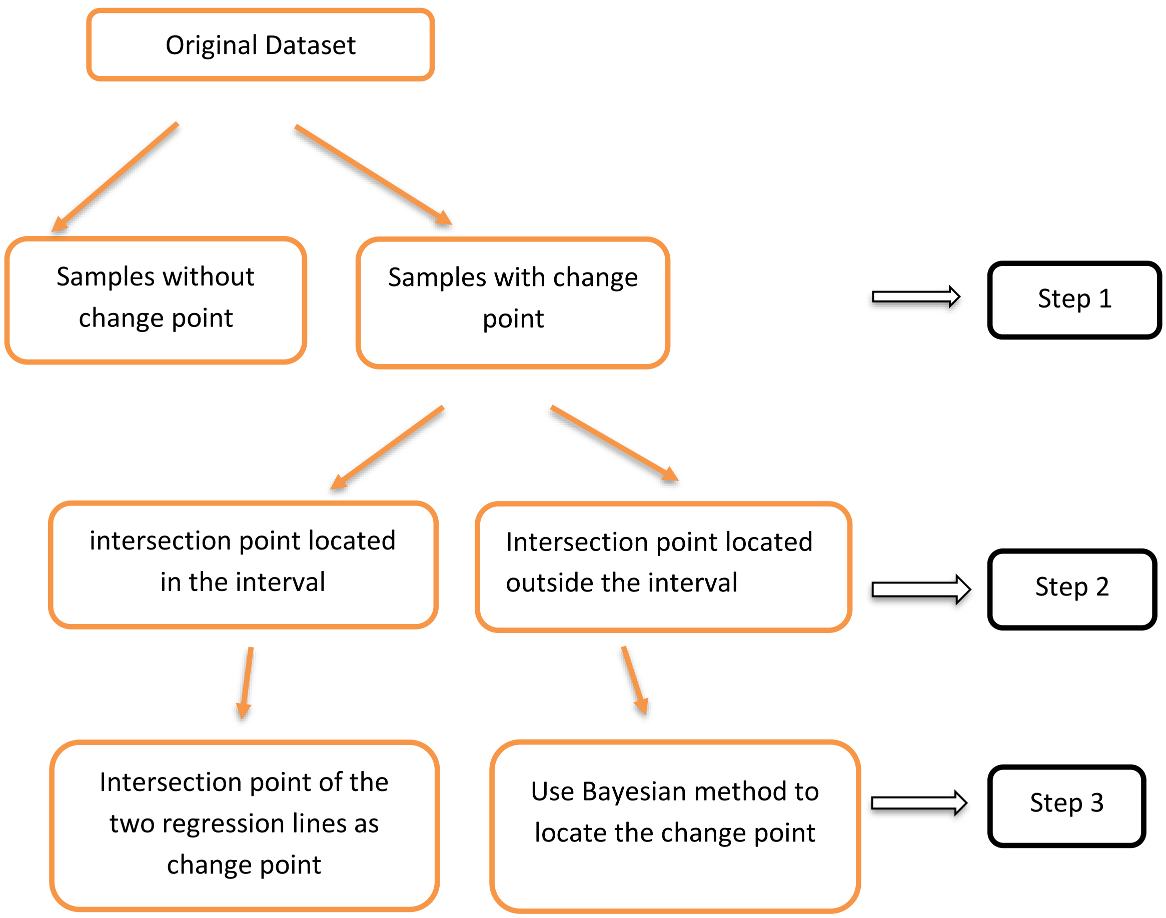


Figure 3.1.1

The workflow of our proposed methods can be illustrated as follows:



3.1 Sample Selection

To examine whether there is a significant change point for a certain child, we fit a linear regression line for sample i over the 7 time points ($t_j = \{t_1, t_2, \cdots, t_7\} = \{0, 0.5, 1, 1.5, 3.5, 5.5, 8.5\}$, where j $= 1, 2, \cdots, 7$). Under the null hypothesis that no change point exists. Besides, we fit two linear regression lines for the first j' time points and the remaining (7-j') time points for each sample i, (where $j' = 2, 3, 4, 5$). For each regression line, calculate the corresponding sum of squared errors of prediction (SSE). Each sample has its distinct j', such that ($SSE_{ij'1} + SSE_{ij'2}$) is minimized. Introducing Δ, as the difference between the SSE of no change point and the SSE with change point. A larger Δ indicates a stronger evidence of change point, and a smaller Δ indicates a weaker evidence of change point.

(a) Under the assumption without change point:

$$\begin{cases} \widehat{y_{ij}} = \widehat{\beta_{io1}} + \widehat{\beta_{io2}} t_j \\ SSE_{io} = \sum_{j=1}^{7} (y_{ij} - \widehat{y_{ij}})^2 \end{cases},$$

where $j = 1, 2, \cdots, 7; i = 1, 2, \cdots N$

In (a), we fit one regression line over all the seven time points, where $\widehat{\beta_{io1}}$ and $\widehat{\beta_{io2}}$ refers to the estimated intercept and slope, respectively, with the assumption that there is no change point for individual i. We use $\widehat{y_{ij}}$ refers to the estimated response for individual i at time point j. And SSE_{io} represents the sum of squared errors for prediction under no change point assumption.

(b) Under the assumption with change point:

$$\begin{cases} \widehat{y_{ij}} = \widehat{\beta_{i11}} + \widehat{\beta_{i12}} t_j \\ SSE_{ij'1} = \sum_{j=1}^{j'} (y_{ij} - \widehat{y_{ij}})^2 \end{cases},$$

where $j = 1, \cdots j'$

$$\begin{cases} \widehat{y_{ij}} = \widehat{\beta_{i13}} + \widehat{\beta_{i14}} t_j \\ SSE_{ij'2} = \sum_{j=j'+1}^{7} (y_{ij} - \widehat{y_{ij}})^2 \end{cases},$$

where $j = j' + 1, \cdots, 7$

$$\min_{j'}(SSE_{ij'1} + SSE_{ij'2}) \tag{3.1.1}$$

$$\Delta_i = SSE_{io} - \min_{j'}(SSE_{ij'1} + SSE_{ij'2}) \tag{3.1.2}$$

Where $i = 1, 2, \cdots N$

In (b), under the assumption with change point, we fit two linear lines over the seven time points, where $\widehat{\beta_{i11}}$ and $\widehat{\beta_{i12}}$ refer to the estimated intercept and slope for individual i in the first developmental phase, respectively. And $\widehat{\beta_{i13}}$, $\widehat{\beta_{i14}}$ refer to the estimated intercept and slope for individual i in the second developmental phase, respectively. And $\widehat{y_{ij}}$ refers to the estimated response for individual i at time j for two distinct growth phases. We use $SSE_{ij'1}$ and $SSE_{ij'2}$, (where $j' = 2, 3, 4, 5$), to represent the sum of squared errors of prediction for the first and second phase. The $\min_{j'}(SSE_{ij'1} + SSE_{ij'2})$ means that for individual i, over all the possible j', choose the one yield minimum $(SSE_{ij'1} + SSE_{ij'2})$. And

Δ_i refers to the difference of SSE under assumption without change point and SSE under assumption with change point for individual i. Δ_i can serve as an indicator of how likely individual i would have a significant change point. Note that $\Delta_i \geq 0$.

(c) Estimate σ^2

$$\begin{cases} \widehat{y_{kj}} = \widehat{\beta_{k11}} + \widehat{\beta_{k12}} t_j \\ SSE_{kj'1} = \sum_{j=1}^{j'} (y_{kj} - \widehat{y_{kj}})^2 \end{cases},$$

where $j = 1, \cdots j'$

$$\begin{cases} \widehat{y_{kj}} = \widehat{\beta_{k13}} + \widehat{\beta_{k14}} t_j \\ SSE_{kj'2} = \sum_{j=j'+1}^{7} (y_{kj} - \widehat{y_{kj}})^2 \end{cases},$$

where $j = j' + 1, \cdots, 7$

$$\min_{j'} (SSE_{kj'1} + SSE_{kj'2})$$

Where $k \in A$.

For each sample i, we calculate the quantity Δ_i. Assume sample yield the highest 30% Δs has a significant change point (shown in Figure 3.1.2). We use the samples produce the highest 30% Δs to estimate the variance $\widehat{\sigma^2}$ with an additional assumption that this variance is the true variance σ^2 for samples with change point. Assume a set of $A = \{ k = 1, \dots , |A|: k \text{ is the sample yield top } 30\% \ \Delta\}$. The sample size of A is $|A|$, where $|.|$stands for the cardinality of the set. In (c), we calculate the minimized sum of squared error for each sample in set A.

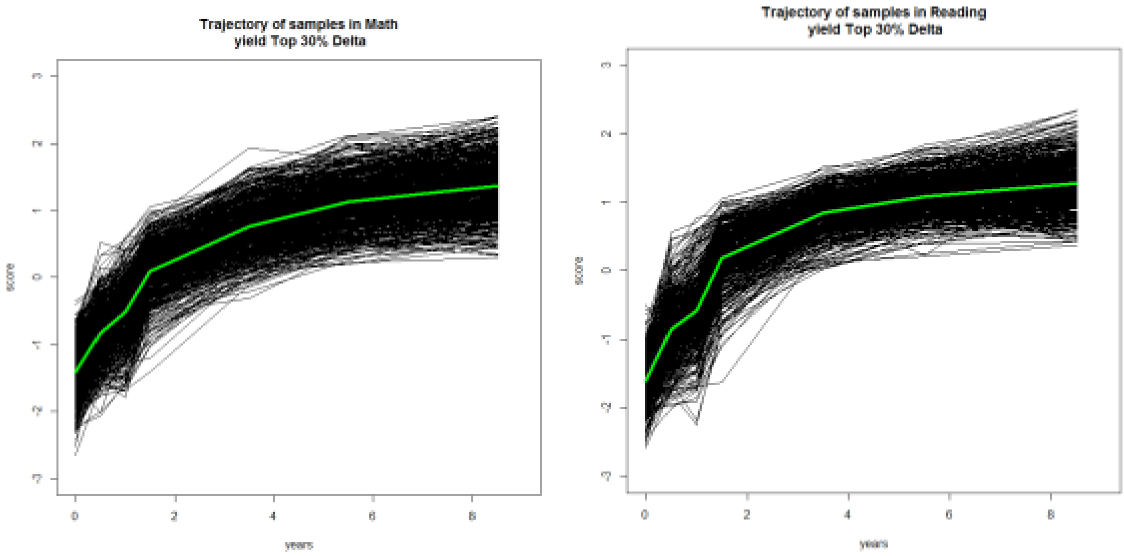


Figure 3.1.2

To filter samples with change point, we conduct a hypothesis testing. The null hypothesis for each sample i is there is no change point, and the alternative hypothesis is there is a change point. The null hypothesis is equivalent to the trajectory of observed measurements is a straight line. Assume y_{ij} follows a normal distribution $y_{ij} \sim N(E(y_{ij}),$ σ^2). However, the population variance σ^2 is unknown. We assume that $\sigma^2 = \widehat{\sigma^2}$, where $\widehat{\sigma^2}$ is the estimated sample variance for the samples in set A. Because the estimated variance $\widehat{\sigma^2}$ is for samples with change points, the sum of SSE is based on the two linear lines. To estimate the $\widehat{y_{ij}}$, we only need four parameters for each sample. For all the samples in set A, there are totally $7|A|$ parameters. Therefore, the variance, $\widehat{\sigma^2}$, is the sum of minimized $(SSE_{ij'1} + SSE_{ij'2})$ for I subsamples divided by $(7|A| - 4|A|$):

$$\widehat{\sigma^2} = \sum_{k \in A}(\min_{j'}(SSE_{kj'1} + SSE_{kj'2}))/(7|A| - 4|A|) \tag{3.1.3}$$

Since we assume $y_{ij} \sim N(E(y_{ij}),\ \sigma^2)$ for each sample, which is equivalent to $(y_{ij} - E(y_{ij}))/_{\sigma} = z_i$ ($z_i \sim N(0,1)$). And $z_i{}^2$ follows a Chi-squared distribution. Because the number of free parameters under the null hypothesis is 7, and the number of free parameters under alternative hypothesis is 2. Therefore, $z_i{}^2$ follows a Chi-squared distribution with (7-2) degree of freedom. We conduct the hypothesis testing at α=0.05 for all the N samples. For each sample, if the p-value is less than 0.05, then there is a significant evidence that this sample has a change point. Otherwise, we fail to reject the null hypothesis that there is no change point for individual i.

(d) Hypothesis Testing

H_0: there is no change point for individual i

H_a: there is a change point for individual i

$$\begin{cases} \widehat{y_{ij}} = \widehat{\beta_{io1}} + \widehat{\beta_{io2}} t_j \\ SSE_{io} = \sum_{j=1}^{7} (y_{ij} - \widehat{y_{ij}})^2 \end{cases}$$

where $j = 1, 2, \cdots, 7;\ i = 1, 2, \cdots N$.

$$z_i{}^2 = \sum_{j=1}^{7} \frac{(y_{ij} - \widehat{y_{ij}})^2}{\sigma^2} = \frac{SSE_{io}}{\sigma^2}$$

Test Statistic: $$T_i = SSE_{io} / \widehat{\sigma^2} \tag{3.1.4}$$

where $T_i \sim \chi^2_{7-2}$.

Another alternative way to conduct the hypothesis testing is based on F-distribution, which could avoid the assumption that the estimated variance, $\widehat{\sigma^2}$, is equal to the true variance σ^2. If there are two independent random variables, U_1 and U_2 followed Chi-squared distributions with d_1 and d_2 degrees of freedom, respectively. Then variable F, $\mathrm{F} = \frac{U_1/d_1}{U_2/d_2}$, follows the F- distribution with degrees of freedom d_1 and d_2.

The null hypothesis for each sample i is that there is no change point, and the alternative hypothesis is that there is a change point. Assume y_{ij} independently follows a normal distribution $y_{ij} \sim N(E(y_{ij}),\ \sigma^2\ )$, where $i = 1, 2, \cdots N$, and $E(y_{ij})$ refers to the estimated response for individual i at time j. The first Chi-squared distribution is for each sample. The SSE_{io} is estimated based on fitting one liner line over 7 time points, then $SSE_{io}/\ \sigma^2$ follows a Chi-squared distribution with (7-2) degree of freedom. The Second Chi-squared distribution is for the samples in set A. The SSE is the sum of minimized $(SSE_{kj'1} + SSE_{kj'2})$ for $|A|$ samples.

(e) Hypothesis Testing

H_0: there is no change point for individual i

H_a: there is a change point for individual i

$$\begin{cases} \widehat{y_{ij}} = \widehat{\beta_{io1}} + \widehat{\beta_{io2}} t_j \\ SSE_{io} = \sum_{j=1}^{7} (y_{ij} - \widehat{y_{ij}})^2 \end{cases}$$

where $j = 1, 2, \cdots, 7;\ i = 1, 2, \cdots N$.

$$T_i = SSE_{io} / \sigma^2 \quad (3.1.5)$$

$$T_i \sim \chi^2_{7-2}$$

and

$$\begin{cases} \widehat{y_{kj}} = \widehat{\beta_{k11}} + \widehat{\beta_{k12}} t_j \\ SSE_{kj'1} = \sum_{j=1}^{j'} (y_{kj} - \widehat{y_{kj}})^2 \end{cases},$$

where $j = 1, \cdots j'$

$$\begin{cases} \widehat{y_{kj}} = \widehat{\beta_{k13}} + \widehat{\beta_{k14}} t_j \\ SSE_{kj'2} = \sum_{j=j'+1}^{7} (y_{kj} - \widehat{y_{kj}})^2 \end{cases},$$

where $j = j' + 1, \cdots, 7$

$$\min_{j'} (SSE_{kj'1} + SSE_{kj'2})$$

Where $k \in A$

$$T_2 = \sum_{k \in A} \min_{j'} (SSE_{kj'1} + SSE_{kj'2}) / \sigma^2 \quad (3.1.6)$$

where $T_2 \sim \chi^2_{7|A|-4|A|}$

Then,

Test Statistic:

$$\frac{T_i}{T_2} = \frac{SSE_{io} / \sigma^2}{\sum_{k \in A} \min_{j'} (SSE_{kj'1} + SSE_{kj'2}) / \sigma^2} = \frac{SSE_{io}}{\sum_{k \in A} \min_{j'} (SSE_{kj'1} + SSE_{kj'2})} \quad (3.1.7)$$

where $F_i = \frac{T_i}{T_2} \sim F(d_1 = 5, d_2 = 3|A|)$

For each sample at α=0.05, if the p-value is less than 0.05, then there is a significant evidence that this sample has a change point. Otherwise, we fail to reject the null hypothesis. Note that the test statistic does not rely on the true variance.

Because $\lim_{d_2 \to \propto} F_{d_1, d_2} = \chi^2{}_{d1}$, when there are sufficient samples in set A, the Chi-squared test and the F test would return similar p-values.

3.2 Inference for Change Point

Through hypothesis testing, the N samples could be separated into two groups: with change point and without change point. For those samples with significant change point, there are two possibilities. Suppose for one sample, the observation j' satisfies $\min_{j'}(SSE_{ij'1} + SSE_{ij'2})$, then fit a linear regression line for the first j' observations, and a linear regression line for $(j'+1)^{th}, \cdots,\ 7^{th}$ observations. Then, two scenarios could appear: (A) the point of intersection falls in $(t_{j'}\ t_{j'+1})$; (B) the point of intersection falls outside $(t_{j'}$, $t_{j'+1})$. The scenario (B) indicates that the observation has outlier(s).

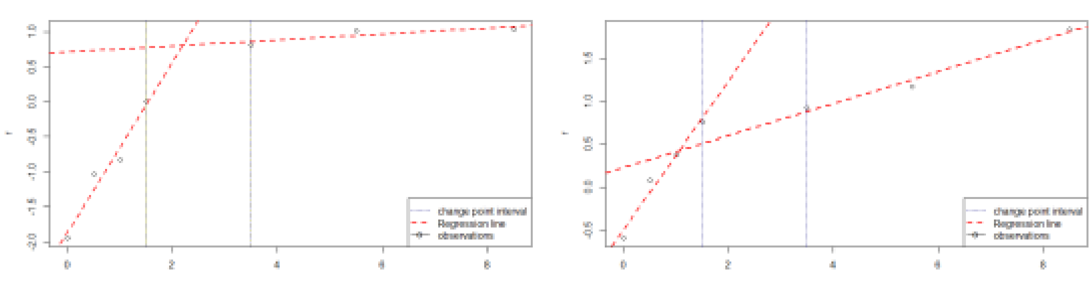


Figure 3.2.1

When scenario (A) happens, the change point could be directly located as the intersection point of the two lines. However, when scenario (B) happens, even though the intersection of the two lines are outside the interval $(t_{j'}, t_{j'+1})$, the intersection point is not necessarily the most likely location for change point. Figure 3.2.1 presents these two cases. Therefore, in this paper, we propose two different algorithms to identify the location for these two cases.

Because algorithms of identifying the location for these two cases are different, we split the samples with change point into two parts: samples without outlier (scenario A) and samples with outlier(s) (scenario B). For every sample has change point, repeat the algorithm in (c) to identify the j'. The developmental trajectories could be modeled as a linear line for the first j' measurement time and a linear line for the last (7- j') measurement time. For each sample, identify the location of the intersection point, (t_{ci}, y_{ci}), where t_{ci} and y_{ci} refer to the coordinates of the intersection point for individual i. Split samples with change point into two parts based on whether $t_{ci} \in (t_{j'}, t_{(j'+1)})$ or not.

For case (A), where the intersection point for a sample within $(t_{j'}, t_{(j'+1)})$, the change point is the intersection point for the two linear regression lines. It is trivial to prove that the intersection point is the maximum likelihood estimate for the change point. For case (B), where the intersection point for a sample falls outside of the interval $\left(t_{j'}, t_{(j'+1)}\right)$, we apply the Bayesian method to understand how likely would a given point be the change point. Given a time point t_c, where $t_c \in (0, 8.5)$, find the probability of the time point t_c be the change point (t_c refers to the time point for change point) for individual i.

When we apply the Bayesian method, there are two parameters of interest t_c and y_c, where y_c refers to the corresponding observed measurement of time t_c. For each sample, the given value for t_c can separate 7 data pairs into two parts, which are the corresponding time point(s) less than t_c and time point(s) large than t_c. For these two parts, we can fit two linear regression lines for each of these two parts, and make the two lines to go through a point (t_c, y_c). When the sum of SSE for these two regression lines become smaller, it is more likely for (t_c, y_c) to be the change point. Since the goal is to identify the change point location, compared with the value for y_c, we are more interested in the value for t_c. Even though, we need y_c to evaluate SSE.

For a certain individual, the observed measurement y_{ij} and measurement time t_j and the fitted two regression lines must go through the point (t_c, y_c). Suppose the slope of the first line and second line are l_1 and l_2, respectively. The two regression lines can be formulated as:

$$\begin{cases} y - y_c = l_1(x - t_c) \\ y - y_c = l_2(x - t_c)' \end{cases}$$

The SSE for the two lines:

$$SSE_{i1} = \sum_{j=1}^{j'}(y_{ij} - \widehat{y_{ij}})^2 \; ; SSE_{i2} = \sum_{j=j'+1}^{7}(y_{ij} - \widehat{y_{ij}})^2,$$

For the first line: $\widehat{y_{ij}} = l_1\left(t_j - t_c\right) + y_c$

For the second line: $\widehat{y_{ij}} = l_2\left(t_j - t_c\right) + y_c.$

For the first line: $SSE_{i1} = \sum_{j=1}^{j'}(y_{ij} - \widehat{y_{ij}})^2$

$$= \sum_{j=1}^{j'} (y_{ij} - l_1(t_j - t_c) - y_c)^2$$

To minimize SSE_{i1}, we take the first derivative of SSE_{i1} with respect to l_1:

$$\frac{d(SSE_{i1})}{dl_1} = 2\sum_{j=1}^{j'} (y_{ij} - l_1(t_j - t_c) - y_c)(t_j - t_c)$$

$$= 2(\sum_{j=1}^{j'} (y_{ij} - y_c)(t_j - t_c) - l_1 \sum_{j=1}^{j'} (t_j - t_c)^2)$$

Set first derivative equal to zero:

$$\frac{d(SSE_1)}{dl_1} = 0$$

$$l_1 = \frac{\sum_{j=1}^{j'} (y_{ij} - y_c)(t_j - t_c)}{\sum_{j=1}^{j'} (t_j - t_c)^2}$$

for second line:

$$l_2 = \frac{\sum_{j=j'+1}^{7} (y_{ij} - y_c)(t_j - t_c)}{\sum_{j=j'+1}^{7} (t_j - t_c)^2}$$

Substitute l_1 and l_2 back to SSE_{i1} and SSE_{i2} to get the minimized summation of (SSE_{i1} + SSE_{i2}):

$$\min_{l_1,l_2}(SSE_{i1} + SSE_{i2}) = \sum_{j=1}^{j'} (y_{ij} - l_1(t_j - t_c) - y_c)^2 + \sum_{j=j'+1}^{7} (y_{ij} - l_2(t_j - t_c) - y_c)^2$$

$$= \sum_{j=1}^{j'}[(y_{ij}-y_c)^2 - 2l_1(y_{ij}-y_c)(t_j-t_c) + {l_1}^2(t_j-t_c)^2] + \sum_{j=j'+1}^{7}[(y_{ij}-y_c)^2 - 2l_2(y_{ij}-y_c)(t_j-t_c) + {l_2}^2(t_j-t_c)^2]$$

$$= \sum_{j=1}^{j'}(y_{ij}-y_c)^2 - 2\frac{\sum_{j=1}^{j'}(y_{ij}-y_c)(t_j-t_c)}{\sum_{j=1}^{j'}(t_j-t_c)^2}\sum_{j=1}^{j'}(y_{ij}-y_c)(t_j-t_c)$$

$$+ \frac{\left(\sum_{j=1}^{j'}(y_{ij}-y_c)(t_j-t_c)\right)^2}{\left(\sum_{j=1}^{j'}(t_j-t_c)^2\right)^2}\sum_{j=1}^{j'}(t_j-t_c)^2 + \sum_{j=j'+1}^{7}(y_{ij}-y_c)^2$$

$$- 2\frac{\sum_{j=j'+1}^{7}(y_{ij}-y_c)(t_j-t_c)}{\sum_{j=j'+1}^{7}(t_j-t_c)^2}\sum_{j=j'+1}^{7}(y_{ij}-y_c)(t_j-t_c)$$

$$+ \frac{\left(\sum_{j=j'+1}^{7}(y_{ij}-y_c)(t_j-t_c)\right)^2}{\left(\sum_{j=j'+1}^{7}(t_j-t_c)^2\right)^2}\sum_{j=j'+1}^{7}(t_j-t_c)^2$$

$$\min_{l_1,l_2}(SSE_{i1} + SSE_{i2})$$

$$= \sum_{j=1}^{7}(y_{ij}-y_c)^2 - \frac{\left(\sum_{j=1}^{j'}(y_{ij}-y_c)(t_j-t_c)\right)^2}{\sum_{j=1}^{j'}(t_j-t_c)^2} - \frac{\left(\sum_{j=j'+1}^{7}(y_{ij}-y_c)(t_j-t_c)\right)^2}{\sum_{j=j'+1}^{7}(t_j-t_c)^2} \qquad (3.2.1)$$

Even though $t_c \in (0, 8.5)$, we have no idea of y_c to evaluate the SSE. For a given value of t_c, $\min_{l_1,l_2}(SSE_{i1} + SSE_{i2})$ could be treated as a function of y_c, Eq. (3.2.1). To find the value for y_c, we use an R package (Nash, 2014), *optimx*. In *optimx* function, set the upper and lower boundary for y_c, and use the 'L-BFGS-B' method. After finding the value for y_c, substitute it back to Eq. 3.2.1 to evaluate the value for $\min_{l_1,l_2}(SSE_{i1} + SSE_{i2})$.

For each given t_c, we can evaluate the corresponding y_c by using *optimx* function. Among all the candidates for t_c, we use the Bayesian method to identify the one that has the highest probability to be the change point, in other words, is to identify the $\min_{l_1,l_2}(SSE_{i1} + SSE_{i2})$.

We use a posterior probability curve to find the most likely change point. The likelihood is the probability of any time point t_c between $(0, 8.5)$ as the change point, $P(data \mid t_c \; as\; the\; change\; point)$. Based on the expert opinion, we could choose a prior such as normal, gamma, etc. Here for illustration purpose, we consider a uniform case. The posterior is the probability of t_c being the change point given the observations for individual i, $P(t_c \; as\; the\; change\; point \mid data)$. Since we assumed $y_{ij} \sim N(E(y_{ij}),\ \sigma^2\ )$. The likelihood for individual i:

$$P(data \mid t_c \; as\; the\; change\; point) = \prod_{j=1}^{7} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_{ij}-\widehat{y_{ij}})^2}{2\sigma^2}}$$

$$\propto e^{-\frac{\sum_{j=1}^{7}(y_{ij}-\widehat{y_{ij}})^2}{2\sigma^2}}$$

Given t_c as the change point, for each sample, we fit two linear regression lines such that t_c would be the intersection time point. The likelihood for individual i can be represented as:

$$P(data \mid t_c \; as\; the\; change\; point) \propto e^{-\frac{\min_{l_1,l_2}(SSE_{i1} + SSE_{i2})}{2\sigma^2}}.$$

And prior is represented as:

$$P(\, t_c \; as \; the \; change \; point) \; \sim Unif(0\;, 8.5).$$

The posterior distribution for individual i is represented as:

$$Posterior \propto likelihood * Prior$$

$$P(t_c \; as \; the \; change \; point \mid data) \; \propto \; e^{-\frac{\min_{l_1, l_2}(SSE_{i1} + SSE_{i2})}{2\sigma^2}}.$$

Here, we assume $\sigma^2 = \widehat{\sigma^2}$. Therefore, every individual has a sequence of possible values for t_c, as well as a sequence of corresponding values for $e^{-\frac{\min_{l_1, l_2}(SSE_{i1} + SSE_{i2})}{2\sigma^2}}$ to reflect the probability of being the change point.

4.RESULTS AND DISCUSSION

4.1 Results

The K-8 Public-Use data contains information about children's reading, mathematics and general knowledge (science and social studies). In this thesis, we only concern the reading and math abilities.

Samples with missing values for reading and mathematics were removed. The sample sizes for math and reading in the final sets are 2,305 and 2,145, respectively. For these two datasets, we first split the data into two parts: samples with change point and samples without change point. We use the samples that yield the top 30% Δs to estimate the variance σ^2 and then conduct a Chi-squared test by treating the estimated σ^2 as true variance. Alternatively, we can use an F-test without the need of true variance. As the denominator degree freedom for F-distribution becomes large, the F-distribution goes close to a Chi-squared distribution, therefore the two methods will generate similar p-values. Due to the large sample size in this study (df2>600), we simply use the Chi-squared distribution to test whether a certain sample contains a change point. For samples with change points, we further split them into two subgroups depending on whether they have outliers in observations. The results we get from reading and mathematics samples are shown in Table 4.1.1.

	Reading	Mathematics
Total sample size	2,145	2,305
Number of samples used to estimate σ^2	644	692
Cutoff for top 30% Δs	1.282	0.863
Estimated variance σ^2	0.040	0.031
Number of samples with change point	2045	2151
Observations with outliers	571	628
Observations without outliers	1,474	1523
Number of samples without change point	100	154

Table 4.1.1

For samples with outliers, we use the Bayesian method. For samples' do not have outliers, we can directly fit two linear regression lines, such that the sum of SSE for the two lines could be minimized. Notice that the Bayesian method could also be applied to locate the change points for samples have no outliers. In Figure 4.1.1 and Figure 4.1.2, we randomly choose one sample has no outliers, and one sample has outliers for Reading and Mathematics, respectively. And fit two linear lines, as well as applying the Bayesian method for each sample. As the Figure 4.1.1 and Figure 4.1.2 show, for samples have no outliers, the Bayesian method works as good as we fit the two linear regression lines.

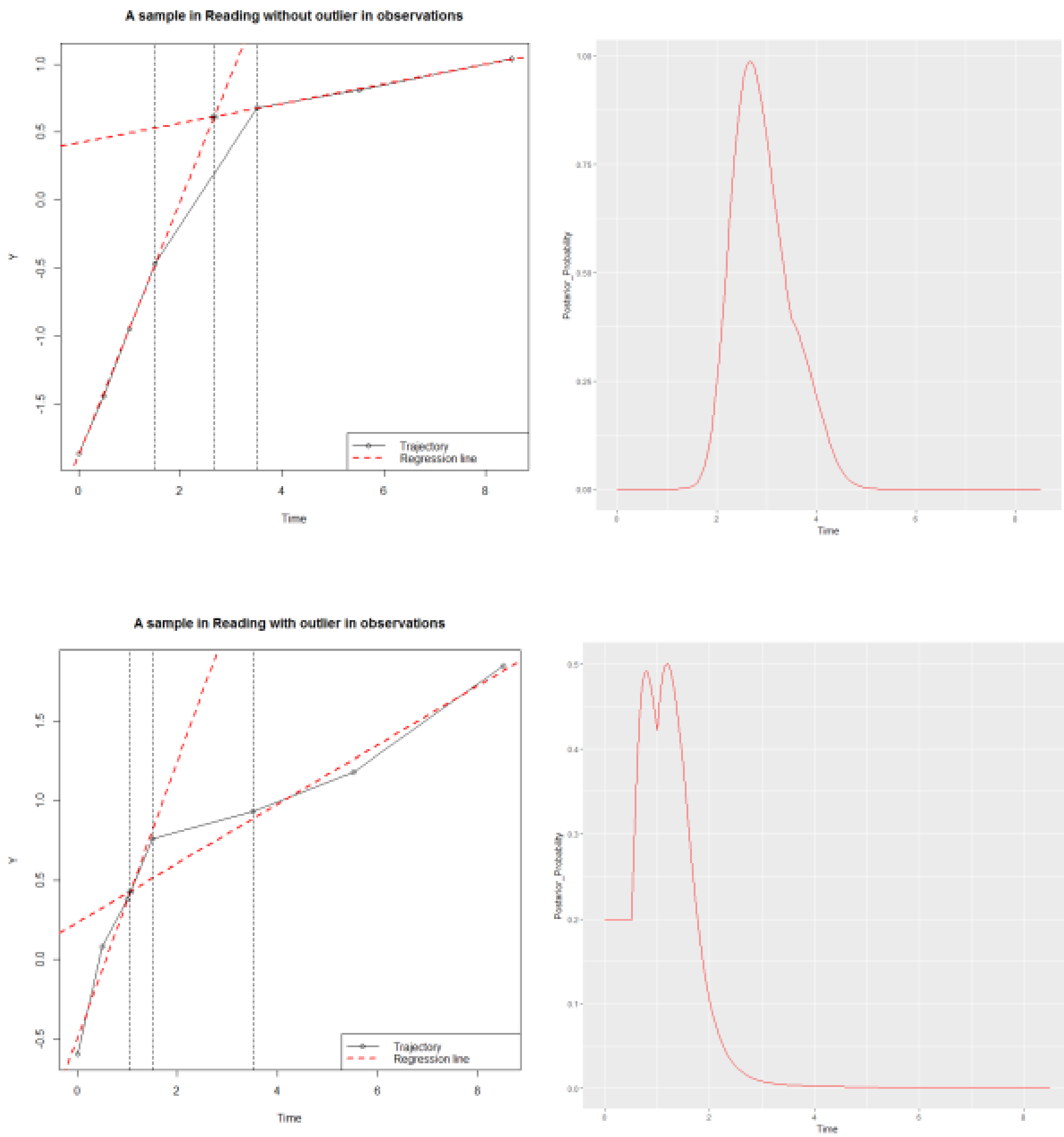

A sample in Reading without outlier in observations
Y
Time
Trajectory
Regression line
Posterior_Probability
Time
A sample in Reading with outlier in observations
Y
Time
Trajectory
Regression line
Posterior_Probability
Time

Figure 4.1.1

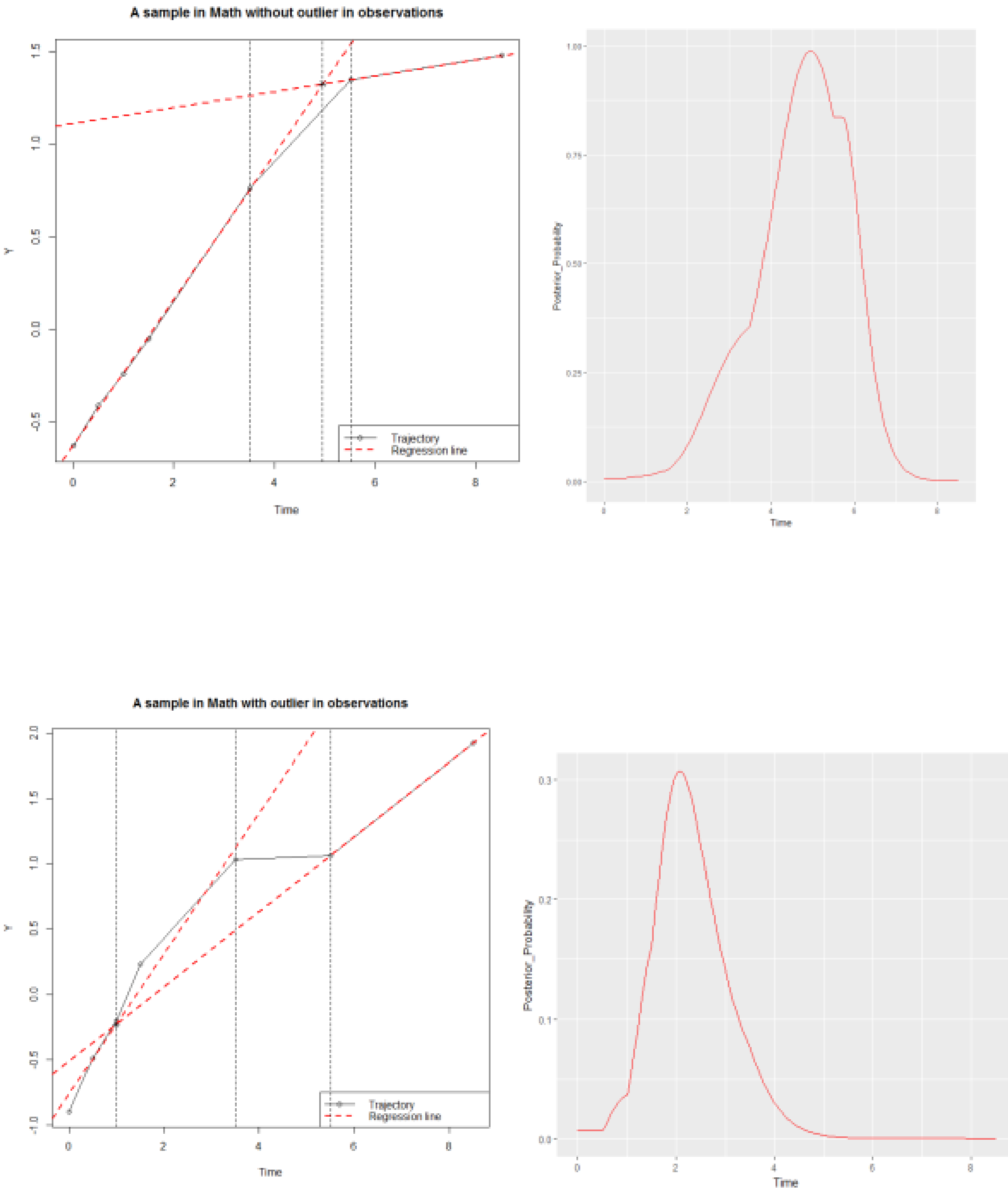

A sample in Math without outlier in observations
Trajectory
Regression line
Time
Y
Posterior_Probability
Time
A sample in Math with outlier in observations
Trajectory
Regression line
Time
Y
Posterior_Probability
Time

Figure 4.1.2

We can find the location of change points for all the samples in Reading and Mathematics. It is interesting to compare the change points distributions for the samples have outliers and for samples have no outliers. In addition, we can compare the change point distributions between reading and mathematics. Figure 4.1.3 shows the change point distribution comparison of the two subjects for samples that have no outliers. Figure 4.1.4 shows the change point distribution comparison of the two subjects for samples with outliers. Figure 4.1.5 shows the change point distribution comparison for all samples.

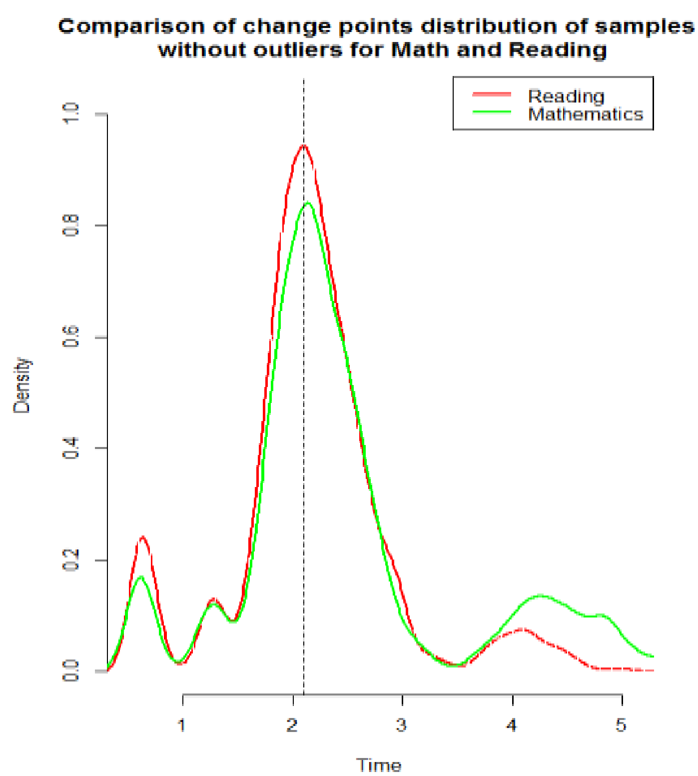


Figure 4.1.3

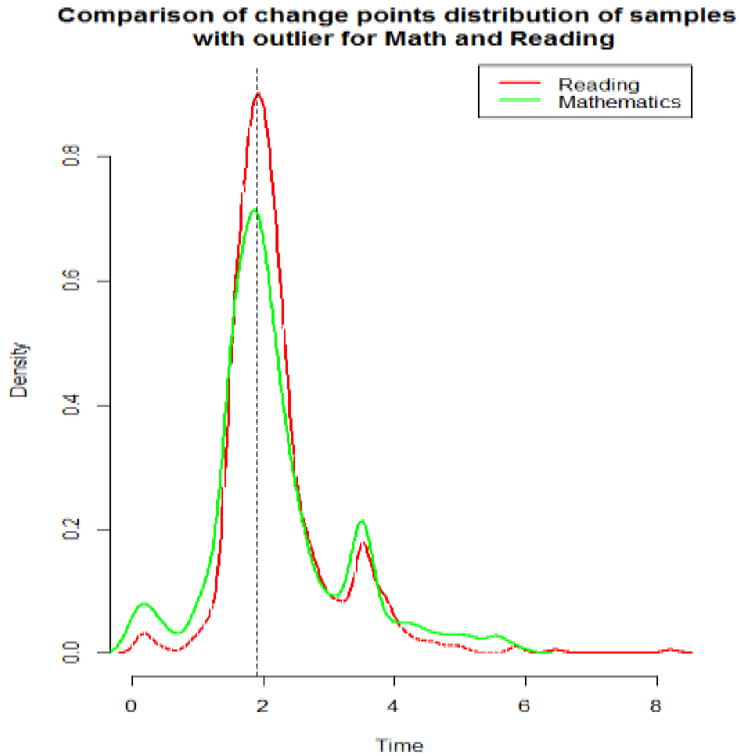

Comparison of change points distribution of samples
with outlier for Math and Reading
Reading
Mathematics
Density
0.0
0.2
0.4
0.6
0.8
0
2
4
6
8
Time

Figure 4.1.4

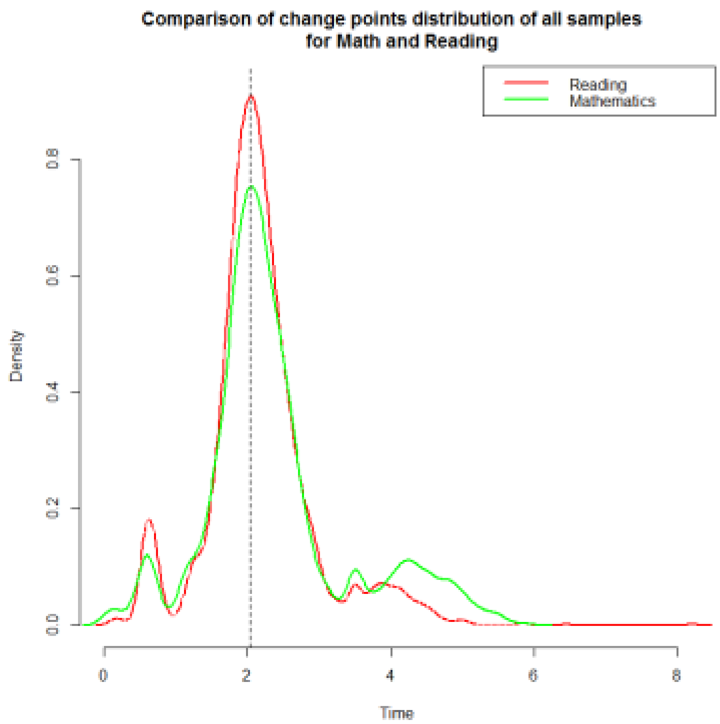


Figure 4.1.5

In Figure 4.1.3, except for the peak appears around 2.1, another significant peak appears around 0.5. While in Figure 4.1.4, except for the peak appears around 1.9, another significant peak appears around 3.5. Compared Figure 4.1.3 and Figure 4.1.4, except for samples have no outliers in math (green curve in Figure 4.1.3), change points for samples have outliers are more likely to happen at the later stage. Because samples have no outliers are approximately three times of the samples have outliers for both subjects, in Figure 4.1.5, the distribution is influenced strongly by samples have no outliers. From these three Figures, we can conclude that compared with math, the change point distribution for reading is more concentrated between spring-first grade and spring third grade.

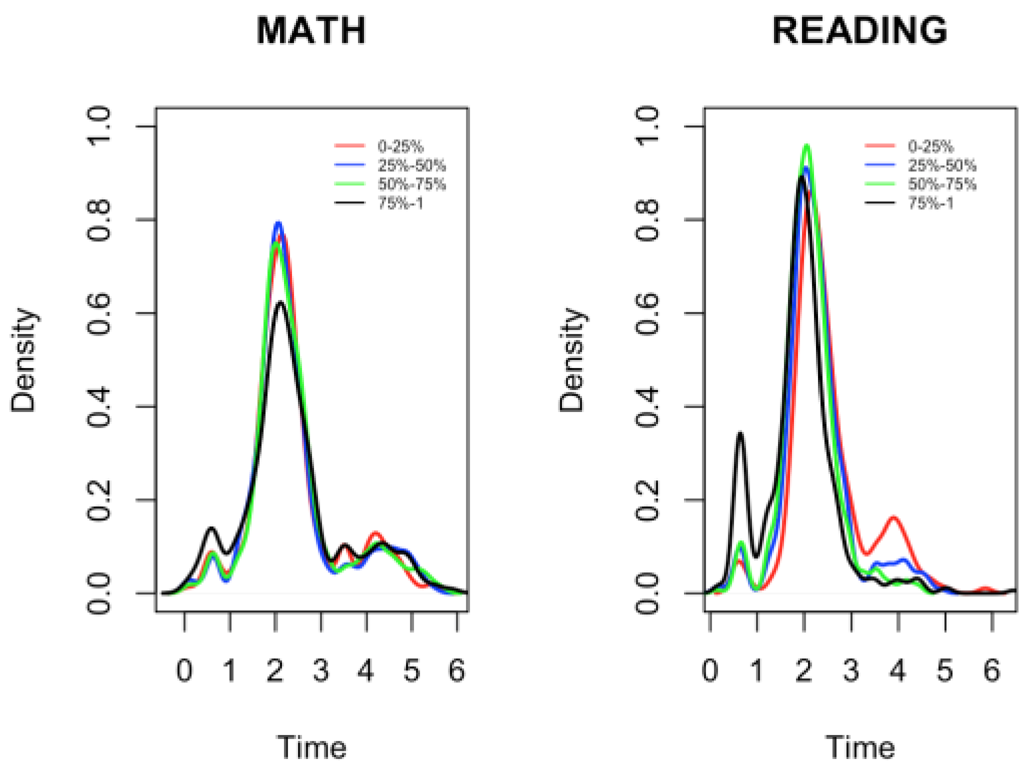


Figure 4.1.6

For most of the samples, children's developmental trajectories jump to another stage at the last three or four rounds. In this paper, we take the average for the last four rounds. Those students with higher average values indicate overall better performance. The change point distribution of students with different overall performance (0-25%, 25-50%, 50-75%, 75-100%) were compared (Figure 4.1.6).

It is shown in Figure 4.1.6 that students with better performance (black curve and green curve) are more likely to have change point at the earlier stage (around 0.5, Spring kindergarten). Besides, most students' change point happens between Spring-first grade and Spring-third grade. Students with poor performance (blue curve and red curve) are more likely to have change point at the later stage (around 4, between spring-third grade and spring-fifth grade).

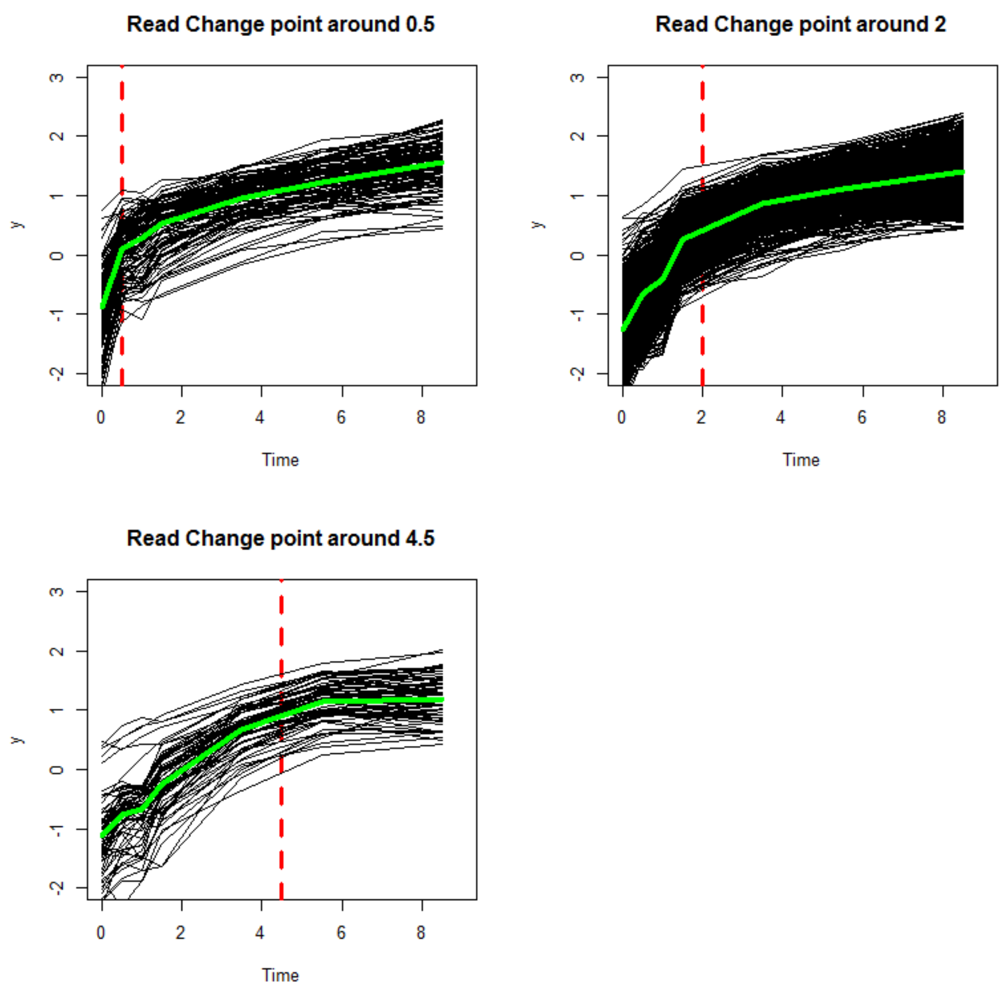


Figure 4.1.7

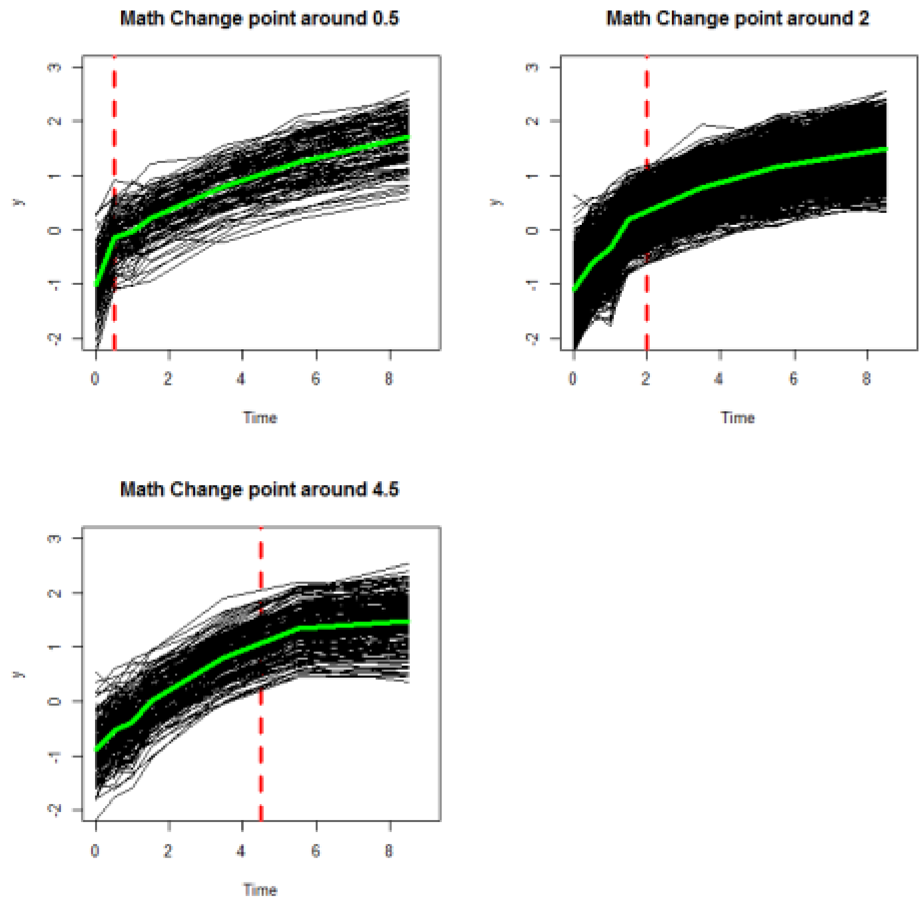


Figure 4.1.8

There are three peaks of the change point distributions: around 0.5, around 2, and around 4. For reading and mathematics dataset, we conduct a subgroup analysis to explore the developmental trajectories for different groups. Figure 4.1.7 and Figure 4.1.8 shows three groups of samples with change points located at 0.5, 2, and 4. The green curve is the average trajectory for each group. These two figures illustrate the underlying developmental trajectories for different groups.

4.2 Discussion

The first step of our proposed computational pipeline is a hypothesis testing to filter out samples without change point. When applying the hypothesis testing to filter out samples do not have change points, to be more precisely, the test statistics should follow the F-distribution. However, the Chi-squared distribution will yield similar results as the F-distribution when the denominator degree of freedom goes to infinity. And it would be easier to implement the Chi-squared test. In this paper, we use the Chi-squared test.

An advantage of using the Bayesian method is that we can estimate the probability of any time point being the change point, and this method can be used for any sample with change point. Because the Bayesian method is more expensive computationally. For the samples with outliers, we directly fit two linear lines, where the SSE for the two lines could be minimized.

In this work, we consider at most one change point for each sample in ECLS-K dataset. However, for a longitudinal study that has more time points, a sophisticated method should be proposed to explore the locations for change points. Moreover, for some other cases where their first phrases have some curvature, instead of linear-linear piecewise models, the quadratic-linear or exponential-linear piecewise model could be considered.

In this thesis, we analyze the dataset for two subjects. The results could provide some insight for researchers in early education. Even though most of the students have change points, for some students, they may have a relatively stable developmental trajectory from kindergarten to eighth-grade. For those students have change points, most of the change points appear between Spring-first grade and Spring third grade (the corresponding time points are 1.5， 3.5). Students have better performance at the later stage are more likely to have change points at the early stage. And students have poor performance at a later stage are more likely to have change points at a later stage. In addition, for those students with similar change point, they share a similar developmental trajectory. The reasons lead to the above phenomena would be worth to do further study.

REFERENCES

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* 213-225.

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

Cudeck, R., & Harring, J. R. 2010. *Statistical methods for modeling human dynamics: An interdisciplinary dialogue.* New York, NY: Routledge.

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

Dominicus, A., Ripatti, S., Pedersen, N. L., & Palmgren, J. 2008. "A random change point model for assessing variability in repeated measures of cognitive function." *Statistics in Medicine* 5786–5798.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. 2006. *An introduction to latent variable.* Mahwah, NJ: Lawrence Erlbaum.

Goldstein, H. 2003. *Multilevel statistical models (3rd ed.).* London, England:: Edward Arnold.

Grimm, K. J. 2008. "Longitudinal associations between reading and mathematics achievement." *Developmental Neuropsychology,* 410-426.

Hancock, G. R., Harring, J. R., & Lawrence, F. R. 2013. *Structural equation modeling: A second course (2nd ed.).* Greenwich, CT: Information Age.

Jordan, N. C., Hanich, L. B., & Kaplan, D. 2003. "A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties." *Child Development* 834–850.

Kreisman, M. B. 2003. "Evaluating academic outcomes of Head Start: An application of general growth mixture modeling." *Early Childhood Research Quarterly.*

Laird, N. M., & Ware, J. H. 1982. "Random-s models for longitudinal data." *Biometrics* 963-974.

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

Miles, T. R., & Miles, E. 1992. *Dyslexia and mathematics.* London: Routledge.

Nash, John C. 2014. "On Best Practice Optimization Methods in {R}." *Journal of Statistical Software* 1--14. http://www.jstatsoft.org/v60/i02/.

Paris, S. G. 2005. "Reinterpreting the development of reading skills." *Reading Research Quarterly* 184–202.

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

Shin, T., Davison, M. L., Long, J., Chan, C.-K., & Heistad, D. 2013. "Exploring gains in reading and mathematics among regular and exceptional students using growth curve modeling." *Learning and Individual Differences* 92-100.

Silverman, R. D., Speece, D. L., Harring, J. R., & Ritchey, K. D. 2012. "Fluency has a role in the simple view of reading." *Scientific Studies of Reading.*

Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., and Najarian, M. . 2009. *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), Combined User's Manual for the ECLS-K Eighth-Grade and K–8 Full Sample Data Files and Electronic Codebooks (NCES 2009–004).* Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.