

# Bayesian Nonparametric Monotone Regression of Dynamic Latent Traits in Item Response Theory Models

Yang Liu  
Xiaojing Wang  
University of Connecticut

Parametric methods, such as autoregressive models or latent growth modeling, are usually inflexible to model the dependence and nonlinear effects among the changes of latent traits whenever the time gap is irregular and the recorded time points are individually varying. Often in practice, the growth trend of latent traits is subject to certain monotone and smooth conditions. To incorporate such conditions and to alleviate the strong parametric assumption on regressing latent trajectories, a flexible nonparametric prior has been introduced to model the dynamic changes of latent traits for item response theory models over the study period. Suitable Bayesian computation schemes are developed for such analysis of the longitudinal and dichotomous item responses. Simulation studies and a real data example from educational testing have been used to illustrate our proposed methods.

**Keywords:** *Bayesian nonparametric; monotonic regression; dynamic changes; item response theory model; Markov chain Monte Carlo*

## 1. Introduction

Longitudinal studies play a prominent role in investigating temporal changes in a construct of interest, which are often referred to as growth curve analysis in social and behavior science. The advent of computerized testing and online rating brings an entirely new way for social and behavior researchers to collect longitudinal data. Test takers have much more freedom to choose their test time than before. Then, the responses collected are often observed at variable and irregular time points across individuals. The randomness of responses may further create sparsity in certain period, such as in the summer or winter holidays.

Traditionally, there are two models of wide usage for studying individual changes. One is called latent growth curve modeling (Bollen & Curran, 2006) and the other is multilevel modeling or hierarchical linear modeling (Raudenbush & Bryk, 2002). However, when outcomes are observed at individually varying

and irregularly spaced time points, the inferences from these two traditional models for studying individual changes may become problematic due to the uneasy adjustment of parametric structures in the models (Geiser, Bishop, Lockhart, Shiffman, & Grenard, 2013). For instance, their analysis typically required the same time span of the study and the same testing points for all examiners. Furthermore, for computerized testing/survey in education, the manifest responses are usually dichotomous, ordinal, or nominal, while latent traits needed to be inferred are often continuous. These make the inference even more difficult because of the information loss in the discretization procedure of the underlying latent variables. In this article, we mainly focus on the extension of classic item response theory (IRT) model framework to model longitudinal dichotomous data collected at irregular and variable time points.

### *1.1. Review of Relevant Literature*

There are two major approaches, that is, the multidimensional and multilevel approach, available in the current literature of IRT models for the analysis of longitudinal data. First, for the multidimensional approach, a multidimensional IRT model is used to represent the change of an ability as an initial ability and one or more modified ability in unidimensional or multidimensional tests (Cho, Athay, & Preacher, 2013; Embretson, 1991; te Marvelde, Glas, Landeghem, & Damme, 2006). However, this approach allows little variation of items on different occasions, and it often requires the individuals to take the same tests. These drawbacks prevent us from extending their methods to analyze a time series of computerized testing data.

Second, for the multilevel approach, the first level is often assumed to follow a classic IRT model, while in the higher level, there are two common ideas to model the growth. One idea is to assume the growth of a latent trait be parametric function of time, such as a linear or polynomial regression of the time variable with fixed or random coefficients. This idea is a variation of the latent growth curve modeling in the analysis of binary/categorical longitudinal data (Albers, Does, Imbos, & Janssen, 1989; Hsieh, von Eye, Maier, Hsieh, & Chen, 2013; Johnson & Raudenbush, 2006; Tan, Amberg, Does, & Imbos, 1999; Verhagen & Fox, 2012). Another idea is to employ Markov chain models, where the changes of a latent trait over time are assumed to be dependent on its previous value or status (Bartolucci, Pennoni, & Vittadini, 2011; Kim & Camilli, 2014; Martin & Quinn, 2002; Park, 2011). However, there are many instances in which neither of two ideas would be enough to describe the growth (Bollen & Curran, 2004). One of such instances is computerized testing, especially when the time lapses between tests are unequally spaced across individuals as well as within individuals.

To tackle the challenges in computerized testing for modeling the growth of latent traits, Wang, Berger, and Burdick (2013) proposed a dynamic model by

combining the ideas of parametric functions of time as well as Markov chain models to describe the growth. They imbedded IRT models into a new class of state space models for analyzing longitudinal data located at individually varying and irregularly spaced time points. Nevertheless, assuming a particular functional relationship for the growth in general is restrictive and usually difficult to justify. Instead, a nonparametric model could be much more flexible to describe changes of latent traits and avoid the errors of model misspecification.

As further investigation of the results shown in Wang et al. (2013), we found the trajectory of one's reading ability grows more quickly in the initial period but slows down when it approaches maturation. Overall, the ability exhibits an increasing trend but often has a flat region at the end. Such discovery of the shape for the growth trajectory is consistent with prior beliefs and experiences from practitioners. In social or behavior science, prior knowledge about the shape of the trajectory, such as monotonicity, convexity, or concavity, may be available ahead of the analysis to aid in the modeling process and enhance interpretability. This calls our attention to incorporate shape constraints as a prior information to nonparametric modeling. It is expected that the usage of shape information may improve the efficiency and accuracy of the nonparametric estimates.

From the Bayesian perspective, nonparametric regression with monotonicity constraints has already been considered in the literature. For instance, Gelfand and Kuo (1991) used an ordered Dirichlet process prior to impose the monotone constraint. Neelon and Dunson (2004) imposed a piecewise linear regression, with an autoregressive prior for the parameters of basis functions. Shively, Sager, and Walker (2009) as well as Brezger and Steiner (2012) implemented restricted splines to ensure monotonicity. McKay Curtis and Ghosh (2011) used Bernstein polynomials with restrictions on parameter space, while Choi, Kim, and Jo (2016) extended this idea by allowing incorporation of uncertainty of the constraints through the prior. Lin and Dunson (2014) proposed Gaussian process projection to perform shape-constrained inference and applied the method to multivariate surface estimation. Wang and Berger (2016) imposed the constraints on the derivative process of the original Gaussian process to estimate shape-constrained functions.

However, typical nonparametric models are often less interpretable. In this article, enlightened by the idea of Bornkamp and Ickstadt (2009), we imbed a flexible Bayesian nonparametric monotone regression of latent traits into the IRT models with easy interpretation. The monotone regression can be written as the sum of two parts: (1) an intercept parameter (interpretable as one's initial ability) and (2) the product of a scale parameter (interpretable as the maximum ability that one can grow during the study period) with a continuous function of the time variable (which is monotonically increasing over the study period and can easily capture the plateaus effect of one's ability at the end). Another advantage of our proposed approach is that the parameters for the underlying base functions which constitute the monotone function do not have constraints on the domain of

parameter space. Lacking of constraints on the parameter space could make the model more flexible and reduce computational burden. Additionally, the base functions themselves can be directly learned from the data.

### 1.2. *EdSphere Test Bed Application*

We will apply our proposed method to the EdSphere data set provided by Highroad Learning Company. EdSphere is a personalized literacy learning platform that continuously collects data about student performance and strategic behaviors each time when he or she reads an article. During a typical reading test, a student selects from a system-generated list of articles having the test difficulty level in a range targeted to the current estimate of the student's ability. Then, for the selected article, a subset of words from the article is eligible to be clozed, that is, removed and replaced by a blank. The computer, following a prescribed protocol, randomly selects a sample of the eligible words to be clozed and presents the article to the student with these words clozed. The question items produced by this procedure are *randomized items*. They are single-use items generated at the time of an encounter between a student and an article. If another student selects the same article to read, a new set of clozed words is selected. As a consequence, the occurrence of individual items among students is highly improbable, so obtaining empirical estimates of item parameters is not feasible.

The difficulty levels of the items in the reading test are provided by Meta-Metrics using proprietary data and methods. The ensemble mean and the variance of difficulty level for the items in a test are known due to the test design of EdSphere learning platform.

Currently, the EdSphere data set consists of 16,949 students from a school district in Mississippi, who participated over 5 years (2007–2011) in EdSphere learning platform. A snapshot of the data sets is included in Table B1 in the online version of the article. The students were in different grades and could enter and leave the program at different times. They were free to take tests on different days and had different time lapses between tests. This design yields *longitudinal observations* located at individually varying and irregularly spaced time points and suggests that we need to model the changes of latent traits with a dynamic structure. Further, in the spirit of EdSphere test design, we could imagine the factors, such as overall comprehension, emotional status, and others, would exist and undermine the *local independence* assumption of IRT models as mentioned in Wang et al. (2013). Therefore, we aim to extend the classic IRT models to accommodate the modern computerized (adaptive) testing (not merely EdSphere data sets), which have the described distinctive features, that is, *randomized items*, *longitudinal observations*, and *local dependence*.

## 2. Nonparametric Monotone Regression

In this article, we keep the discussion focused on the one-parameter IRT model that links latent ability and item difficulty to the correctness of each item. The idea could be similarly implemented on two-parameter/three-parameter IRT models or other continuous, ordinal, and nominal latent variable models. In the classic one-parameter IRT model, the latent ability of each individual and the difficulty level for each item are the two key components and often they assume to be static. But in the EdSphere data set, the item responses are longitudinal, and thus, the latent ability of individuals and item difficulty levels are both varying with time. In addition, each test taker is allowed to take tests at any time they wish in the computerized testing. Moreover, they can take more than one exams per day. Therefore, following the discussion of Wang et al. (2013), we need to extend the classic one-parameter IRT model as below.

The proposed shape-constrained IRT model involves two levels, that is,

$$\text{Level 1 : } \Pr(X_{i,j,s,k} = 1 | \theta_{i,j}, \eta_{i,j,s}, d_{i,j,s,k}) = F(\theta_{i,j} - d_{i,j,s,k} + \eta_{i,j,s}), \quad (1)$$

$$\text{Level 2 : } \theta_{i,j} = f_i(t_{i,j}) + \omega_{i,j}. \quad (2)$$

In the first level, Equation 1 extends the classic one-parameter IRT models to the scenario of computerized testing, where  $\theta_{i,j}$  represents the  $i$ th person's ability (latent trait) on  $j$ th day with assuming a person's ability is constant over a given day;  $d_{i,j,s,k}$  is the difficulty of the  $k$ th item in the  $s$ th test on the  $j$ th day taken by the subject  $i$ ;  $\eta_{i,j,s}$  takes account of the random effects that cannot explained by person's ability and the item difficulty in the  $s$ th test on the  $j$ th day for the  $i$ th person;  $F(\cdot)$  is a cumulative distribution function (CDF) for continuous random variables; and  $i = 1, \dots, n, j = 1, \dots, T_i, s = 1, \dots, S_{i,j}$  and  $k = 1, \dots, K_{i,j,s}$ . For the EdSphere data set, since the ensemble mean and variance of the item difficulties in a test are known quantities due to the test design, we assume

$$d_{i,j,s,k} = a_{i,j,s} + v_{i,j,s,k}, \quad (3)$$

where  $v_{i,j,s,k} \sim \mathcal{N}(0, \sigma^2)$ , and  $a_{i,j,s}$  and  $\sigma$  are known. Further, we presume for each individual  $i$ , the random effects  $\eta_{i,j,s} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \tau_i^{-1})$  for  $j = 1, \dots, T_i$  and  $s = 1, \dots, S_{i,j}$ , where  $\tau_i$  is a precision parameter and changes according to individuals. For the link function  $F^{-1}(\cdot)$ , we will use  $F^{-1}(\cdot) = \Phi^{-1}(\cdot)$  (called the Normal Ogive or Probit link), where  $\Phi^{-1}(\cdot)$  is the inverse function of the standard normal CDF, and this link can ease the computation for Bayesian analysis.

In the second level (Equation 2),  $w_{i,j}$  represents the random residuals,  $w_{i,j} \sim \mathcal{N}(0, \phi^{-1} \Delta_{i,j})$ , where  $\phi$  is an unknown proportion of the precision of  $w_{i,j}$  and  $\Delta_{i,j}$  is the time lapse between the  $j$ th test day and the  $(j-1)$ th test day of the subject  $i, j = 1, \dots, T_i$ . We assume the variance of  $w_{i,j}$  is proportional to the time lapse because it implies the uncertainty about one's ability would

become larger when he or she does not take the tests for a while.  $f_i(\cdot)$  is the  $i$ th individual latent trajectory and assumed to be a continuous function, where  $t_{i,j}$  is the time location, that is, the actual test day for an examinee to take a test in the study period. Often when we model  $f_i(\cdot)$ , there is some prior information available. For example, psychologists and educators can assume that the mean trend for a student's reading ability would be growing or at least not decreasing during the study period. We choose to impose such prior beliefs on the modeling of  $f_i(\cdot)$  since it can be used for a large part of the potential applications, and we may be able to check from the data fit about the reasonableness of this assumption.

We will utilize the idea of Bornkamp and Ickstadt (2009) to model the unknown latent trajectory  $f_i(\cdot)$  flexibly and conveniently with a monotone shape constraint. First, for each individual  $i$ ,  $i = 1, \dots, n$ , we rescale the original time units into  $[0, 1]$ , by subtracting the original time with the minimum time value of the  $i$ th individual and then divided by the time range of individual  $i$ . After such rescaling, we continue to use  $t_{i,j}$  for the notation of time for convenience and assume

$$f_i(t_{i,j}) = \beta_{i,0} + \beta_{i,1}f_i^0(t_{i,j}), i = 1, \dots, n, j = 1, \dots, T_i, \quad (4)$$

where  $f_i^0(\cdot)$  is the CDF of a bounded and continuous random variable on  $[0, 1]$ . Thus,  $f_i(\cdot)$  will be increasing if  $\beta_{i,1} \geq 0$  and decreasing otherwise. Moreover,  $f_i(\cdot)$  can accommodate flat regions if chosen  $f_i^0(\cdot)$  properly. This is because a CDF of a random variable will reach its plateau when the variable approaches its bound of the domain. Under the current time rescaling mechanism and since  $0 \leq f_i^0(t_{ij}) \leq 1$  for each individual  $i$ , the intercept  $\beta_{i,0}$  can be interpreted as the initial level of the latent ability for the  $i$ th individual, while  $\beta_{i,0} + \beta_{i,1}$  can be treated as the maximum level that an individual can reach during the study period. Second, to make the modeling of  $f_i^0(\cdot)$  much flexible, we introduce the nonparametric ideas,

$$f_i^0(t_{i,j}) = \int_{\Xi} F(t_{i,j}, \xi) \mathcal{G}_i(d\xi) = \sum_{\ell=1}^{L_i} \pi_{i,\ell} F(t_{i,j}, \xi_{i,\ell}), \xi_{i,\ell} \stackrel{\text{i.i.d.}}{\sim} P_0, \quad (5)$$

where we model  $f_i^0(\cdot)$  as a discrete mixing of parametric CDFs. In Equation 5,  $F(\cdot, \xi)$  is a CDF with parameters  $\xi$  belong to the parameter space  $\Xi$ ;  $\mathcal{G}_i$  is a discrete probability measure on  $\Xi$  with assigning a general discrete random measure prior introduced by Ongaro and Cattaneo (2004), that is,  $\mathcal{G}_i(d\xi) = \sum_{\ell=1}^{L_i} \pi_{i,\ell} \delta_{\xi_{i,\ell}}(d\xi)$ , where  $\delta$  is a Dirac delta function;  $\xi_{i,\ell}$ 's are independent and identically distributed realizations from a continuous distribution  $P_0$  on  $\Xi$  and are assumed to be independent with  $\pi_{i,\ell}$ 's and  $L_i$ 's;  $\pi_{i,\ell}$  satisfies  $\sum_{\ell=1}^{L_i} \pi_{i,\ell} = 1$ , and  $L_i$  requires its support on positive integers. Also, note the common choices of  $P_0$  in Equation 5 will help us to borrow strength among individual latent trajectories. The construction of Equation 5 actually contains many popular discrete random probability measures in the current literature as a special case, such as the Dirichlet process, general

stick-breaking processes, and so on. In addition, this construction is very flexible in modeling monotone function since it has not imposed any restricted structure on parameters (i.e., for the weights  $\pi_{i,\ell}$ 's and the parameters  $\xi_{i,\ell}$ 's) in comparison to other methods mentioned in the Introduction section.

In Equation 5, proper choice of the base distribution function  $F(\cdot, \xi)$  is the key for the success of modeling  $f_i(\cdot)$ . A typical requirement is that we need a convex combination of functions  $F(\cdot, \xi_{i,1}), F(\cdot, \xi_{i,2}), \dots$ , for any  $i = 1, \dots, n$ , which can approximate any arbitrary continuous CDF on  $[0, 1]$ . The beta distribution functions (i.e., the regularized incomplete beta functions) will satisfy this requirement, however, there is usually heavy computational burden associated with the beta distribution functions since they have no closed form. To balance the computation burden and the adequacy of approximation, we consider the CDF of two-sided power (TSP) distribution (cf. Van Dorp & Kotz, 2002) for  $F(\cdot, \xi)$ :

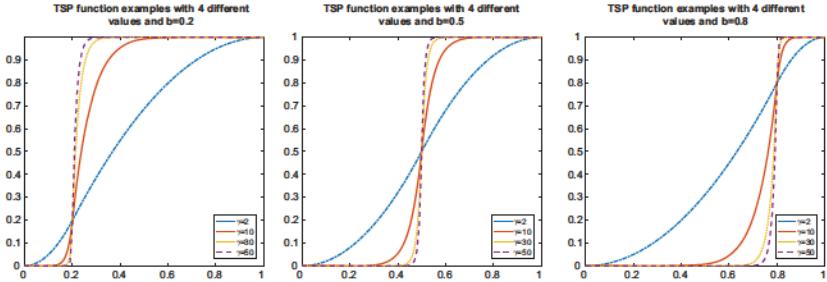
$$F(t_{i,j}, \xi) = \begin{cases} b \left( \frac{t_{i,j}}{b} \right)^\gamma, & \text{if } 0 \leq t_{i,j} \leq b, \\ 1 - (1-b) \left( \frac{1-t_{i,j}}{1-b} \right)^\gamma, & \text{if } b \leq t_{i,j} \leq 1. \end{cases}, \quad (b, \gamma) \in [0, 1] \times \mathbf{R}_+,$$

which has two key parameters  $(b, \gamma)$  and is a viable alternative of beta distribution functions. Here, define  $\xi = (b, \gamma)$ . Bornkamp and Ickstadt (2009) proved the convex combination of the CDF of TSP functions can be capable of approximating any continuous CDF on  $[0, 1]$ .

We illustrate several examples of TSP functions in Figure 1. From Figure 1, we can see that  $\gamma$  controls the steepness of the curve. When  $\gamma$  is small, the pace of increasing is comparatively slow, and when  $\gamma$  becomes larger, the increasing trend becomes steeper. The parameter  $b$  is the unique mode of TSP function when  $\gamma > 1$ . A TSP function with a small  $b$  and a large  $\gamma$  describes a curve increasing steeply in the beginning and stabilizing afterward, which could be viewed as a “fast learner” growth curve, while a TSP function with a large  $b$  and small  $\gamma$  can be viewed as a growth trend for an individual who improves steadily with a slower pace. Other scenarios could be interpreted accordingly, and they will be varying with different values of  $b$  and  $\gamma$ .

### 3. Bayesian Computation Scheme

The hierarchical model of Equations 1 and 2 can accommodate the complex structure of computerized testing (e.g., EdSphere data sets), and it also allows the incorporation of prior information. However, because of the complexity of the model considered, we have to resort to Markov chain Monte Carlo (MCMC) computational techniques for the analysis. A by-product of Bayesian inference is that all uncertainties in all quantities are combined in the overall assessment of inferential uncertainty.

FIGURE 1. Illustration of several two-sided power functions with different  $\gamma$  and  $b$  values.

### 3.1. Prior Specification and Posterior Distribution

Before starting the Bayesian inference, first, we have to specify the prior distributions of unknowns in the model. For parameters  $\phi$ ,  $\tau_i$ 's,  $\beta_{i,0}$ 's, and  $\beta_{i,1}$ 's in Equation 1, 2, and 3, there is a lack of scientific knowledge, so we use the following objective priors for them:  $\pi(\phi) \propto \phi^{-\frac{3}{2}}$ ,  $\pi(\tau_i) \propto \tau_i^{-\frac{3}{2}}$ ,  $\pi(\beta_{i,0}) \propto 1$ , and  $\pi(\beta_{i,1}) \propto 1$ , for  $i = 1, \dots, n$ . The objective priors used for  $\pi(\phi)$  and  $\pi(\tau_i)$  are recommended in (Wang, Berger, & Burdick, 2013).

According to Lemma 1 in Bornkamp and Ickstadt (2009), the prior distributions of  $\xi_{i,\ell}$ 's determine the mean and prior correlation structure of  $f_i^0(\cdot)$ . Without expert's information, a uniform distribution on a finite subset of parameter space  $\Xi$  for  $\xi_{i,\ell}$ 's would be a reasonable choice to start with. Theoretically, we would like to elicit an unbounded prior for  $L_i$ . One viable choice is the zero-truncated Poisson distribution with the rate parameter  $\lambda > 0$ , and thus, its prior mean is  $\frac{\lambda e^\lambda}{e^\lambda - 1}$ . The larger  $\lambda$  value is, the more components are in TSP mixture. For the prior of  $\boldsymbol{\pi}_i = (\pi_{i1}, \dots, \pi_{iL})'$ , a natural choice is a symmetric Dirichlet distribution with common parameter  $\rho > 0$ . Notice the prior variability of  $f_i^0(\cdot)$  is increasing when  $L_i$  or  $\rho$  gets smaller (see Lemma 1 in Bornkamp & Ickstadt, 2009). Hence, in practice, the values of  $\lambda$  and  $\rho$  are chosen according to the desired prior variability for  $f_i^0(\cdot)$  and the expected number of jumps in the model response.

Using the priors specified above, we can derive the posteriors of unknowns in the proposed model as shown in Equation 8 of online Appendix A. We can show this posterior is proper (see details in online Appendix A). Then, the statistical inferences based on the sampling from this posterior is legitimate. To facilitate the computation for the posterior of unknowns, we implement the idea of data augmentation (Albert & Chib, 1993) by introducing a latent variable  $Z_{i,j,s,k}$  for each dichotomous response  $X_{i,j,s,k}$ , that is, defining  $Z_{i,j,s,k} \sim N(\theta_{i,j} - a_{i,j,s} + \eta_{i,j,s}, 1 + \sigma^2)I(Z_{i,j,s,k} > 0)$  if  $X_{i,j,s,k} = 1$  and  $Z_{i,j,s,k} \sim N(\theta_{i,j} - a_{i,j,s} + \eta_{i,j,s}, 1 + \sigma^2)I(Z_{i,j,s,k} \leq 0)$  if  $X_{i,j,s,k} = 0$ .

$\sigma^2)I(Z_{i,j,s,k} \leq 0)$  if  $X_{i,j,s,k} = 0$ . Therefore, the two-level hierarchical model (Equations 1 and 2) can be simplified as

$$Z_{i,j,s,k} = f_i(t_{i,j}) + \omega_{i,j} - d_{i,j,s,k} + \eta_{i,j,s} + \varepsilon_{i,j,s,k}, \quad (6)$$

with  $\varepsilon_{i,j,s,k} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ . Then, our computation schemes for drawing samples from the joint posterior distribution of unknowns will derive from this data augmentation model.

Let us denote  $\Lambda_i = (L_i, \pi_{i,1}, \dots, \pi_{i,L_i}, \xi_{i,1}, \dots, \xi_{i,L_i})'$  with each  $\xi_{i,\ell} = (b_{i,\ell}, \gamma_{i,\ell})$  being the corresponding parameters of the  $\ell$ th TSP mixture component for the  $i$ th subject,  $\pi_{i,\ell}$  is the corresponding assigned weight of the  $\ell$ th mixture component, with  $i = 1, \dots, n$  and  $\ell = 1, \dots, L_i$ . Thus,  $\Lambda_i$  contains all information about the TSP mixture for the  $i$ th individual, and we denote the notations  $\boldsymbol{\Lambda} = \{\Lambda_1, \dots, \Lambda_n\}$  and  $\boldsymbol{\xi} = \{\xi_{1,1}, \dots, \xi_{1,L_1}, \dots, \xi_{n,1}, \dots, \xi_{n,L_n}\}$  to represent the sets of variables for all  $n$  subjects.

Similarly, we use bold notations  $\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{Z}, \boldsymbol{\eta}, \boldsymbol{\tau}$  to define the sets of corresponding variables introduced in Section 2, over all indices, that is,  $\boldsymbol{\beta} = \{\beta_{1,0}, \beta_{1,1}, \dots, \beta_{n,0}, \beta_{n,1}\}$ ,  $\boldsymbol{\eta} = \{\eta_{i,j,s} : i = 1, \dots, n, j = 1, \dots, T_i, s = 1, \dots, S_{i,j}\}$ ,  $\boldsymbol{\tau} = \{\tau_1, \dots, \tau_n\}$ ,  $\mathbf{Z} = \{Z_{i,j,s,k} : i = 1, \dots, n, j = 1, \dots, T_i, s = 1, \dots, S_{i,j}, k = 1, \dots, K_{i,j,s}\}$ ,  $\boldsymbol{\theta} = \{\theta_{1,1}, \dots, \theta_{1,T_1}, \dots, \theta_{n,1}, \dots, \theta_{n,T_n}\}$ . Similarly, define  $\mathbf{X} = \{X_{i,j,s,k} : i = 1, \dots, n, j = 1, \dots, T_i, s = 1, \dots, S_{i,j}, k = 1, \dots, K_{i,j,s}\}$ . Then, the joint posterior distribution of parameters  $\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\Lambda}, \mathbf{Z}, \boldsymbol{\eta}, \boldsymbol{\tau}, \phi$  given the data  $\mathbf{X}$  is derived as

$$\begin{aligned} & f(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\Lambda}, \mathbf{Z}, \boldsymbol{\eta}, \boldsymbol{\tau}, \phi | \mathbf{X}) \\ & \propto f(\mathbf{X} | \mathbf{Z}) f(\mathbf{Z} | \boldsymbol{\theta}, \boldsymbol{\eta}) f(\boldsymbol{\theta} | \boldsymbol{\beta}, \boldsymbol{\Lambda}, \phi) f(\boldsymbol{\eta} | \boldsymbol{\tau}) \pi(\boldsymbol{\beta}) \pi(\boldsymbol{\Lambda}) \pi(\phi) \pi(\boldsymbol{\tau}) \\ & \propto \prod_{i=1}^n \prod_{j=1}^{T_i} \prod_{s=1}^{S_{i,j}} \prod_{k=1}^{K_{i,j,s}} \left\{ (I(Z_{i,j,s,k} > 0) I(X_{i,j,s,k} = 1) + I(Z_{i,j,s,k} \leq 0) I(X_{i,j,s,k} = 0)) \right. \\ & \quad \times \sqrt{\frac{\psi_{i,j,s,k}}{2\pi}} \exp \left\{ -\frac{\psi_{i,j,s,k} (Z_{i,j,s,k} - \theta_{i,j} + a_{i,j,s} - \eta_{i,j,s})^2}{2} \right\} \Big\} \\ & \quad \times \prod_{i=1}^n \prod_{j=1}^{T_i} \sqrt{\frac{\phi}{2\pi \Delta_{i,j}}} \exp \left\{ -\frac{\phi [\theta_{i,j} - \beta_{i,0} - \beta_{i,1} f_i^0(t_{i,j})]^2}{2 \Delta_{i,j}} \right\} \\ & \quad \times \prod_{i=1}^n \prod_{j=1}^{T_i} \left( \frac{\tau_i}{2\pi} \right)^{\frac{S_{i,j}}{2}} \exp \left\{ -\frac{\tau_i \sum_{s=1}^{S_{i,j}} \eta_{i,j,s}^2}{2} \right\} \left[ \prod_{i=1}^n \pi(\tau_i) \pi(\beta_i) \pi(\Lambda_i) \right] \pi(\phi), \end{aligned} \quad (7)$$

where  $\psi_{i,j,s,k} = (1 + \sigma^2)^{-1}$  and  $\pi(\tau_i)$ ,  $\pi(\beta_i)$ ,  $\pi(\Lambda_i)$ , and  $\pi(\phi)$  denote the priors specified in the beginning of this subsection.

### 3.2. The MCMC Sampling Schemes

The key part in developing MCMC algorithm to draw samples from the joint posterior distribution (Equation 7) is to estimate the latent trajectory  $f_i^0(t_{i,j})$ s in Equation 4. Notice that once  $\Lambda$  is known,  $f_i^0(t_{i,j})$ s are fully determined. Since the distribution of  $\Lambda$  can vary in dimension, we employ a reversible jump Markov chain Monte Carlo (RJMCMC) sampling scheme (Green & Hastie, 2009). Further, to reduce the correlation as well as to achieve faster convergence of the MCMC samples of parameters, we implement the idea of partially collapsed Gibbs sampling (Van Dyk & Park, 2008). Thus, at the  $q$ th iteration, we perform the sampling procedure of unknowns in the order below:

1. Sample  $\mathbf{Z}^{(q)}$  from  $f(\mathbf{Z}^{(q)} | \boldsymbol{\theta}^{(q-1)}, \boldsymbol{\eta}^{(q-1)})$ , which is a truncated normal distribution for the full conditional distribution of each individual  $Z_{i,j,s,k}$  given the rest;
2. Sample  $\boldsymbol{\eta}^{(q)}$  from  $f(\boldsymbol{\eta}^{(q)} | \boldsymbol{\theta}^{(q-1)}, \boldsymbol{\tau}^{(q-1)}, \mathbf{Z}^{(q)})$ , which is a normal distribution for the full conditional distribution of each  $\eta_{i,j,s}$  given the rest;
3. Sample  $\boldsymbol{\tau}^{(q)}$  from  $f(\boldsymbol{\tau}^{(q)} | \boldsymbol{\eta}^{(q)})$ , which is a gamma distribution for the full conditional distribution of each  $\tau_i$  given the rest;
4. Sample  $\boldsymbol{\Lambda}^{(q)}$  from  $f(\boldsymbol{\Lambda}^{(q)} | \boldsymbol{\theta}^{(q-1)})$ , which has no closed form; moreover, the dimension of  $\boldsymbol{\Lambda}$  is varying for each iteration, and thus, we employ the Metropolis-Hastings algorithm within the RJ-MCMC to sample the full conditional distribution of each  $\Lambda_i$  given the rest;
5. Sample  $\phi^{(q)}$  from  $f(\phi^{(q)} | \boldsymbol{\theta}^{(q-1)}, \boldsymbol{\Lambda}^{(q)})$ , which is a gamma distribution for the full conditional distribution of  $\phi$  given the rest;
6. Sample  $\boldsymbol{\beta}^{(q)}$  from  $f(\boldsymbol{\beta}^{(q)} | \boldsymbol{\theta}^{(q-1)}, \boldsymbol{\Lambda}^{(q)}, \phi^{(q)})$ , which is a multivariate normal distribution for the full conditional distribution of  $\beta$  given the rest;
7. Sample  $\boldsymbol{\theta}^{(q)}$  from  $f(\boldsymbol{\theta}^{(q)} | \boldsymbol{\beta}^{(q)}, \boldsymbol{\Lambda}^{(q)}, \mathbf{Z}^{(q)}, \boldsymbol{\eta}^{(q)}, \phi^{(q)})$ , which is a normal distribution for the full conditional distribution of each individual  $\theta_{i,j}$  given the rest.

The details of each sampling step are described in online Appendix A. The MCMC sampling loops through Steps 1 through 7 and repeats until the MCMC is converged. The initial values (i.e., when  $(q-1)$ th iteration = (0)th iteration) of parameters chosen in simulations and application are:  $\theta_{i,j}^{(0)}$ 's drawing from  $\mathcal{N}(0, 1)$ ,  $\beta_{i,1}^{(0)}$ 's and  $\beta_{i,2}^{(0)}$ 's being 0,  $\eta_{i,j,s}^{(0)}$ 's being 0,  $\phi^{(0)} = 200$ , and  $\tau_i^{(0)}$ 's being 6 for  $i = 1, \dots, n$ . While, we will specifically discuss how to choose the initial values of  $\boldsymbol{\Lambda}_i^{(0)} = (L_i^{(0)}, \pi_i^{(0)}, b_i^{(0)}, \gamma_i^{(0)})$  for  $i = 1, \dots, n$  in different examples. The convergence is evaluated informally by looking at trace plots.

Then, statistical inferences are made straightforward from the MCMC samples. For example, an estimate and 95% credible interval (CI) for the latent trajectory of one's ability  $\boldsymbol{\theta}_i = (\theta_{i,1}, \dots, \theta_{i,T_i})'$  can be plot from the median, 2.5%, and 97.5% empirical quantiles of the corresponding MCMC realizations of each  $\theta_{i,j}$ , for  $j = 1, \dots, T_i$ . In examples, these will be graphed as a function of time  $t_{ij}$ , so that the dynamic changes of an examinee are apparent.

## 4. Simulation Examples

To validate the inference procedure of MCMC schemes and show the success of using monotone shape constraints in the nonparametric modeling, two simulation studies are conducted. For the first simulated example, we know the true underlying curve  $f_i^0(\cdot)$ 's that generate the latent trajectory of one's ability, while for the second simulated example, we have no information about the true underlying curve  $f_i^0(\cdot)$ 's.

### 4.1. An Example Using the Mixture of TSP Functions as the Latent Trajectory

In this section, we apply our proposed method to a simulated data set that uses the mixture of TSP functions as the true latent trajectory. We consider 10 test takers, that is,  $n = 10$ ; each of them is examined at 60 different test days, that is,  $T_i = 60, i = 1, \dots, 10$ . During each distinctive test day, there are four examinations for each individual, thus  $S_{i,j} = 4$  for  $i = 1, \dots, 10, j = 1, \dots, T_i$ , and there are 10 questions (or items) in each test, that is,  $K_{i,j,s} = 10$  for  $i = 1, \dots, 10, j = 1, \dots, T_i$  and  $s = 1, \dots, S_{i,j}$ . For each person  $i$ , we assume the time lapse between two consecutive tests is a function of  $j$ , which is set to be  $\Delta_{i,j} = j + 10$  for  $j = 1, \dots, T_i/2$  and  $\Delta_{i,j} = j - 10$  for  $j = T_i/2 + 1, \dots, T_i$ , with  $i = 1, \dots, 10$ .

We set the true values for model parameters as below:  $\phi^{-1/2} = 0.05$ , and thus, the corresponding standard deviation of the random component  $w_{i,j}$  in Equation 2 is  $0.05\sqrt{\Delta_{i,j}}\tau^{-1/2} = (0.361, 0.286, 0.322, 0.362, 0.359, 0.302, 0.347, 0.325, 0.360, 0.378)'$ , where each element represents the standard deviation of random effects  $\eta_{i,j,s}$  for the  $i$ th person, respectively.  $\sigma^2 = .7333^2$ , which is chosen based on the test design of EdSphere data. The parameters specified for the mixture components of TSP functions are addressed in Table 1.

We use the prior specified in Section 3.1 for the unknown parameters of the proposed model. Particularly, for this simulated example, we use a zero-truncated Poisson for each  $L_i$  with  $\lambda = 2$ , which implies there are expecting no more than three jumps since each of the mixture of TSP functions in the simulation consists at most two components. Without available scientific information, we employ a symmetric Dirichlet distribution with the parameter  $\rho = 1$  as the prior for  $\pi_{i,\ell}$ 's, assign a *Uniform*(0, 1) for the prior of each  $b_{i,\ell}$ ; and specify a *Uniform*(1, 50) for the prior of each  $\gamma_{i,\ell}$ .

Model fitting is done with the MCMC algorithm described in the Section 3.2, based on 50,000 iterations in total. The first 10,000 samples are burned in, and only every 20th value is taken for our inference as to reduce dependence. After this burn-in process, the mixing of MCMC samples looks pretty good, and the trace plots of MCMC samples of parameters are convergent.

TABLE 1.

Simulated Parameters for Each Individual of Two-Sided Power Functions

	$L$	$\pi$	$b$	$\gamma$
$\Lambda_1$	2	[0.5, 0.5]	[0.5, 0.1]	[3, 5]
$\Lambda_2$	2	[0.2, 0.8]	[0.4, 0.7]	[2, 8]
$\Lambda_3$	1	1	0.15	10
$\Lambda_4$	2	[0.3, 0.7]	[0.3, 0.5]	[5, 13]
$\Lambda_5$	1	1	0.3	4
$\Lambda_6$	1	1	0.5	17
$\Lambda_7$	2	[0.1, 0.9]	[0.2, 0.3]	[3, 15]
$\Lambda_8$	1	1	0.3	5
$\Lambda_9$	2	[0.6, 0.4]	[0.1, 0.5]	[8, 5]
$\Lambda_{10}$	2	[0.7, 0.3]	[0.2, 0.45]	[4, 10]

In Figure 2, we display results of 4 represented individuals among the 10 simulated individuals in the example. Those four individuals have noticeable differences in the shape of their corresponding trajectories. The growth curve for the first subject is steadily increasing without obvious transition. While for the second individual in Figure 2, he or she has obvious turning points. He or she grows slowly during the initial period, then rapidly grows in the middle and reaches the plateaus in the end. For third and sixth subjects in Figure 2, they have similar phenomena as the second individual but with different lengths of the three stages in the study.

In each subfigure of Figure 2, the dash line represents the true underlying function  $f_i(\cdot)$ , the bullet dots are simulated values of  $\theta_{i,j}$ 's for  $j = 1, \dots, T_i$ , and the star dotted line is the estimate of posterior median of  $f_i(\cdot)$ , along with the solid lines representing the corresponding 2.5% and 97.5% credible bands. We could see that the fitted trajectory captures the trend pretty well even though the number of different test date is comparatively small, that is, 60. Simulations for larger sample sizes have been tested, and the results were generally better. Those results are not addressed here as to save the space.

We also calculate the posterior estimates of all other unknowns. The posterior median of  $\phi^{-1/2}$  is 0.0561, with a 95% credible interval (CI) being [0.0491, 0.0634], where the true value of  $\phi^{-1/2}$  (i.e., 0.05) is included. The posterior median as well as 95% CIs of other parameters are displayed in Figures 3 and 4. We could see that their respective true values, which are marked by cross, are all included in the 95% CIs.

To take account of randomness in the simulations, another 100 independent data sets are simulated to check frequentist coverage probabilities, with same parameter setup but different random seeds. For each data set, we also run the MCMC sampling for 50,000 iterations with the first 10,000 samples being

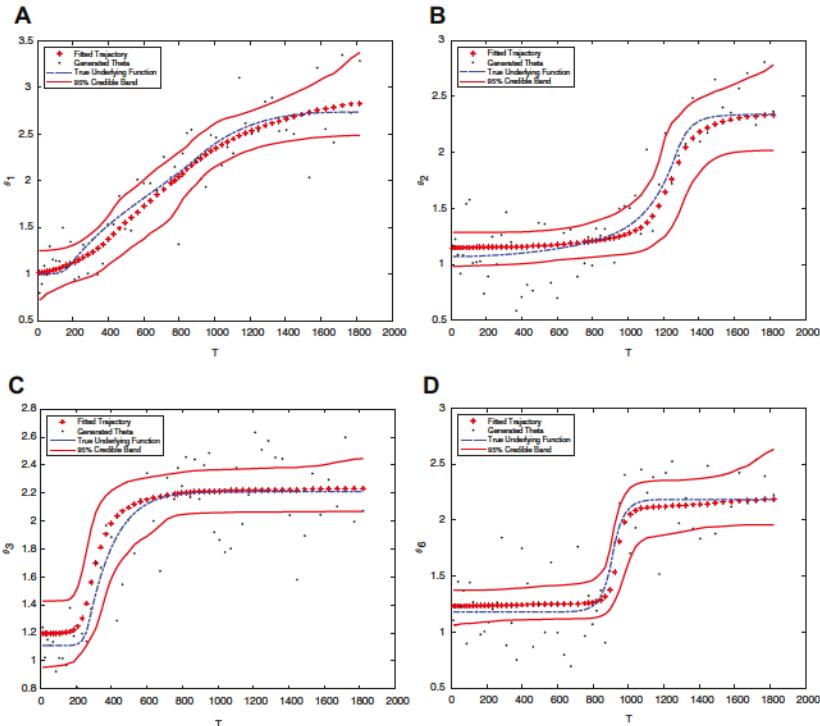


FIGURE 2. The estimation of latent trajectories for the first, second, third, and sixth subjects. (A) First individual. (B) Second individual. (C) Third individual. (D) Sixth individual.

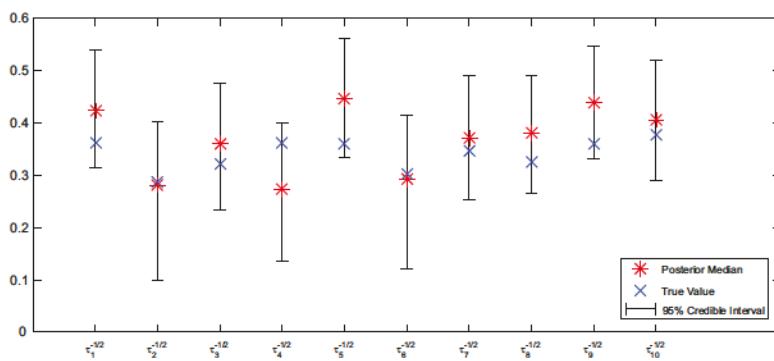


FIGURE 3. Posterior median and 95% CIs of  $\tau_i$ 's in the mixture of two-sided power function based simulation.

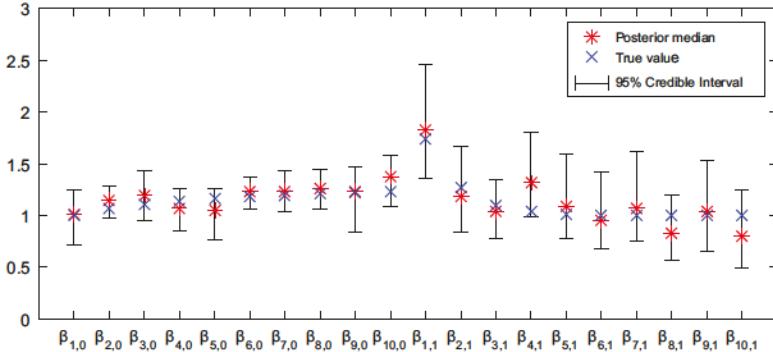


FIGURE 4. The results of posterior median and 95 % CIs of  $\beta_{i,0}$ 's and  $\beta_{i,1}$ 's in the mixture of two-sided power function based simulation.

burned in. In addition, the MCMC chains are thinned by only using every 20th sample. The results are shown in Table 2. We can see that the coverage probabilities of the 95% CIs of all model parameters including the truth are equal or very close to the nominal level 95%. Thus, while the inferential method is Bayesian, it seems to yield sets that have good frequentist coverage.

#### 4.2. An Example of the Logistic Curve as the Latent Trajectory

In this section, we apply our proposed method to some non-TSP-based true latent trajectories. The setup is the same as the previous simulation except the true trajectories  $f_i(\cdot)$ ,  $i = 1, \dots, 10$  become the logistic curves as below:

$$\begin{aligned}
 f_1(t) &= \frac{-0.5 + 2}{(\exp(-10t + 5) + 1)}; f_2(t) = \frac{-0.5 + 2}{(\exp(-5t + 2) + 1)}; \\
 f_3(t) &= \frac{0.5 + 1}{(\exp(-10t + 2.5) + 1)}; f_4(t) = \frac{0.5 + 2}{(\exp(-12t + 3) + 1)}; \\
 f_5(t) &= \frac{2.5 + 1}{(\exp(-5t + 3) + 1)}; f_6(t) = \frac{-0.5 + 2}{(\exp(-10t + 2) + 1)}; \\
 f_7(t) &= \frac{1}{(\exp(-10t + 2) + 1)}; f_8(t) = \frac{0.5 + 2}{(\exp(-4t + 1) + 1)}; \\
 f_9(t) &= \frac{-1.5 + 2}{(\exp(-5t + 1) + 1)}; f_{10}(t) = \frac{1}{(\exp(-2t + 1) + 1)}.
 \end{aligned}$$

The motivation for using logistic curves as the true latent trajectories in the simulation is that the logistic curves have been widely applied in the growth curve analysis. Similarly, our model fitting is done based on 200,000 iterations in

TABLE 2.

Coverage Probabilities of  $\phi$ ,  $\theta_i$ 's,  $\tau_i$ 's,  $\beta_{i,0}$ 's, and  $\beta_{i,1}$ 's by 100 Independent Simulations

Parameter	Coverage Probability	Parameter	Coverage Probability
$\tau_1$	0.940	$\tau_2$	0.940
$\tau_3$	0.960	$\tau_4$	0.920
$\tau_5$	0.940	$\tau_6$	0.900
$\tau_7$	0.900	$\tau_8$	0.920
$\tau_9$	0.880	$\tau_{10}$	0.940
$\theta_1$	0.944	$\theta_2$	0.932
$\theta_3$	0.950	$\theta_4$	0.975
$\theta_5$	0.969	$\theta_6$	0.965
$\theta_7$	0.971	$\theta_8$	0.931
$\theta_9$	0.945	$\theta_{10}$	0.963
$\beta_{1,0}$	0.960	$\beta_{2,0}$	0.920
$\beta_{3,0}$	0.960	$\beta_{4,0}$	1.000
$\beta_{5,0}$	0.980	$\beta_{6,0}$	0.980
$\beta_{7,0}$	1.000	$\beta_{8,0}$	0.940
$\beta_{9,0}$	0.960	$\beta_{10,0}$	0.960
$\beta_{1,1}$	0.960	$\beta_{2,1}$	0.960
$\beta_{3,1}$	0.960	$\beta_{4,1}$	1.000
$\beta_{5,1}$	0.980	$\beta_{6,1}$	0.960
$\beta_{7,1}$	0.960	$\beta_{8,1}$	1.000
$\beta_{9,1}$	0.980	$\beta_{10,1}$	1.000
$\phi$	0.940		

total. The first 40,000 samples are burned in, and every 20th value of MCMC samples is taken to reduce the dependence of samples. We use a zero-truncated Poisson for each  $L_i$  with  $\lambda = 3$ , expecting a few more TSP components are needed than the previous example to fit the logistic curve. Similarly as before, we use a symmetric Dirichlet distribution with the parameter  $\rho = 1$  for  $\pi_{i,\ell}$ 's and assign  $\text{uniform}(0, 1)$  for the prior of each  $b_{i,\ell}$  as well as specify  $\text{uniform}(1, 50)$  for the prior of each  $\gamma_{i,\ell}$ . The priors for the rest unknowns are specified the same as Section 4.1.

Figure 5 displays the results of four selected individuals, where the dash lines represent the truth of the underlying growth curve, the bullet dots are simulated values of  $\theta_{i,j}$ 's for  $j = 1, \dots, T_i$ , the star dots correspond to the posterior median estimates of one's ability, and the solid dash lines indicate the 95% credible band of the estimates. Figure 5(B) and 5(C) show that the true growth curves of fourth and sixth examiners both have a steadily increasing growth curve at the beginning and reach a plateau after half of the study period, whereas our estimated growth curves (star dotted line) capture the trend of the truth (dash lines) very well, and all true values (dash lines) are within the 95% CIs of our estimation

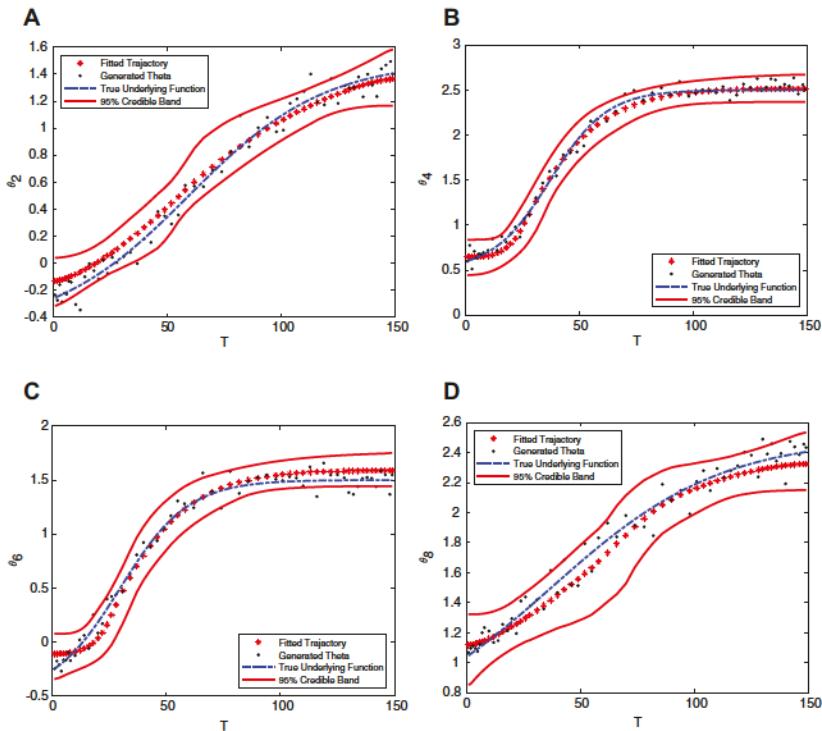
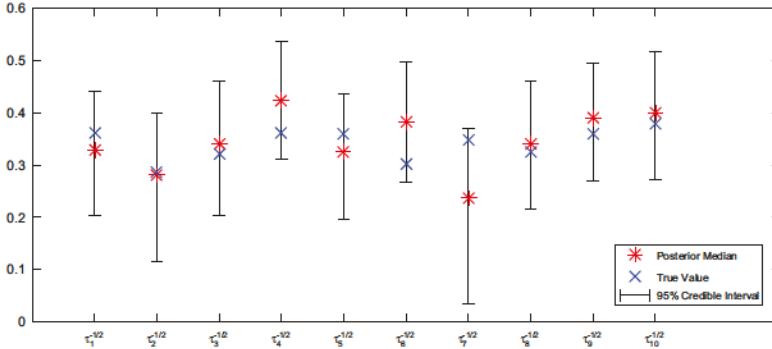
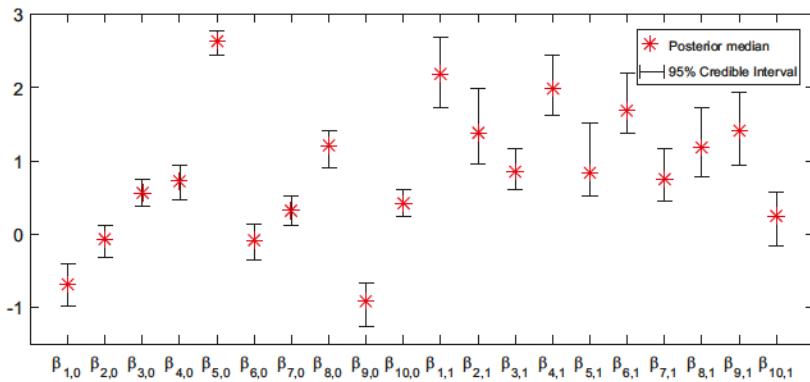


FIGURE 5. The estimation of latent trajectories for the fourth, sixth, seventh, and eighth subjects. (A) Second individual. (B) Fourth individual. (C) Sixth individual. (D) Eighth individual.

(solid lines). For the second and eighth subjects, seen from Figure 5(A) and 5(D), their growth curves are strictly increasing over the study period, and similarly, our proposed method can capture the underlying trend well under these situations.

In addition, we calculate the posterior estimates of all other unknowns. The posterior median of  $\phi^{-1/2}$  is 0.0539, with the 95% CI being [0.0471, 0.0610], which includes its true value 0.05. The posterior median as well as 95% CIs of  $\tau_i$ 's,  $\beta_{i,0}$ 's, and  $\beta_{i,1}$ 's are displayed in Figures 6 and 7, where we can see that the true-simulated values of  $\tau_i$ 's are all inside their corresponding 95% CIs. Notice that in the logistic curve simulations, the true values of  $\beta_{i,0}$ 's and  $\beta_{i,1}$ 's are unknown, and thus, we are not able to compare the truth relative to the corresponding 95% CIs for  $\beta_{i,0}$ 's and  $\beta_{i,1}$ 's.

FIGURE 6. Posterior median and 95 % CIs of  $\tau_i$ 's in logistic curve simulation.FIGURE 7. Posterior median and 95 % CIs of  $\beta_{i,0}$ 's and  $\beta_{i,1}$ 's in logistic curve simulation.

## 5. Application to EdSphere Data

Since our approach has been successfully applied to estimate the trend of  $f_i(\cdot)$  from simulated data and recovered the true values of parameters well, we will employ our two-level hierarchical models to the EdSphere data. Due to the limitation of our computer's RAM (8 GB), a sample of 10 individuals from the EdSphere database was randomly chosen for illustration purpose. The characteristics of the individuals are described in online Appendix B. There are two goals for the analysis of EdSphere data sets. One is to assess the appropriateness of the local independence assumption for this type of data, and the other is to understand the growth in ability of students, by retrospectively producing the estimated growth trajectories of their latent abilities, incorporating the monotone assumption.

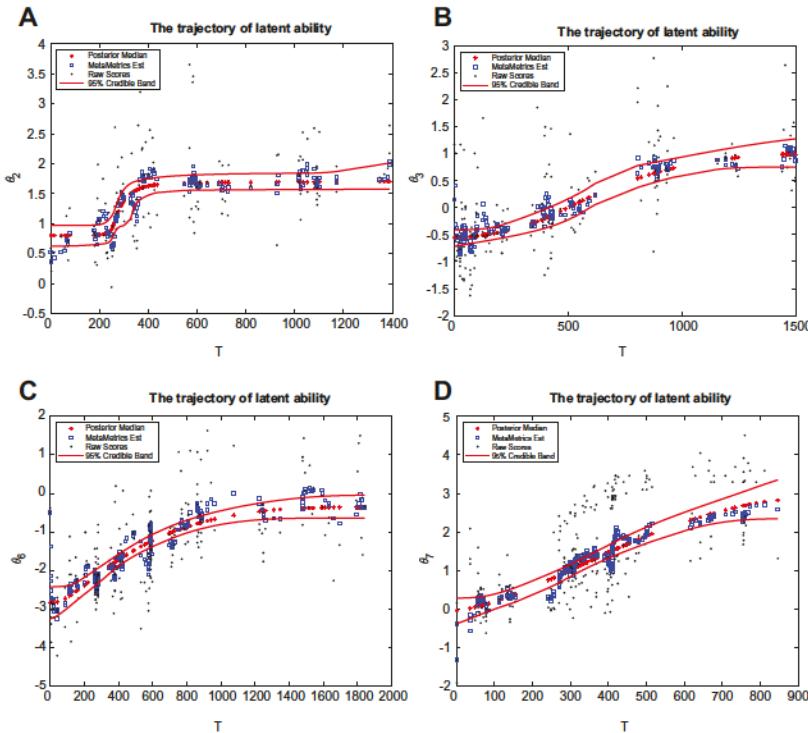
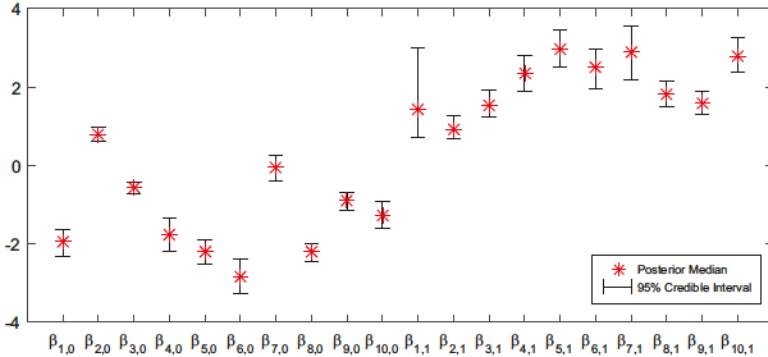


FIGURE 8. The estimation of latent trajectories for the second, third, sixth, and seventh subjects. (A) Second individual. (B) Third individual. (C) Sixth individual. (D) Seventh individual.

The prior specification is the same as the aforementioned simulation examples except we use a zero-truncated Poisson prior for each  $L_i$  with  $\lambda = 1$ , which corresponds to a prior belief that using about two different TSP functions, we can explain the changes of the response trajectory in the data. But such prior assumption can be washed out by the data if our data have strong information to indicate that we need more mixture components of TSP functions to explain one's ability growth. In addition, to examine the sensitivity of the prior specification for  $L_i$ , we have tried other  $\lambda$  values; the yielded estimation of latent trajectories are almost the same as those shown in Figure 8.

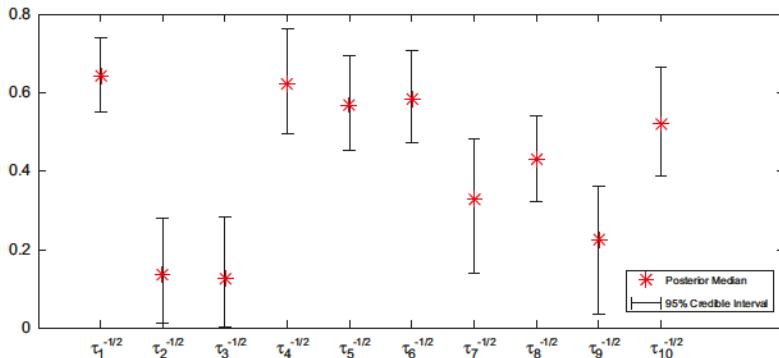
We have run in total 500,000 iterations and burn in the first one fifth of the samples, and to reduce dependence of MCMC samples, we have taken only every 20th value of MCMC samples. We have checked the trace plots of model parameters to access the convergence for MCMC samples, and we found a good mixing is observed after our burn-in and thinning process. Figure 8 shows the

FIGURE 9. Posterior median and 95% CIs of  $\beta_{i,0}$ 's and  $\beta_{i,1}$ 's.

estimated trajectory of four individuals, which represent four types of growth curves we typically observed from the data.

In Figure 8, from our proposed method, the star dotted lines denote the corresponding 95% credible band of one's latent trajectory, the square dots represent the estimated trajectory used in current EdSphere learning platform (where they assume AR(1) model for  $\theta_{i,j}$  and consider an observation equation as Model 1 but without the random effect  $\eta_{i,j,s}$ ), and the bullet dots correspond to the estimates of one's ability obtained by solving the equation that the expectation of expected score for a person's ability is equivalent to the observed score (which can roughly be thought of as the raw test scores put on the same scale as the  $\theta_{i,j}$ 's). In Figure 8, we can see that our estimated latent trajectory of one's ability growth (i.e., the red dots for the posterior median of  $\theta_{i,t}$ 's) displays a much smooth monotone increasing trend of one's ability in comparison to current EdSphere's estimation, where current EdSphere's approach shows a continuously up-and-down oscillation in the estimation of one's ability.

Noticeably, in Figure 8, the latent growth curve of the second individual (i.e., see Figure 8[A]), a six-grader, increases sharply at about the 300th day and reaches the plateau before the 450th day, while the sixth individual (in Figure 8[C]) experiences a moderate growth during most of the time before stabilizing by the end of the study. For the third and the seventh students in Figure 8(B) and 8(D), respectively, they both are in Grade 2 and have a similar type of steadily increasing shape of the growth over the entire study period. However, seen from Figure 8, the growth rate (or learning speed) of the seventh individual is much faster than that of the third individual. Since the study period of seventh individual is much shorter than that of the third individual, clearly, we can also compare the estimation of posterior median as well as the corresponding 95% CIs of  $\beta_{7,1}$  and  $\beta_{3,1}$  (they represent the magnitude of one's ability growth during the study period), respectively, in Figure 9 to validate the differences between the growth rate of these two

FIGURE 10. Posterior median and 95% CIs of  $\tau_i$ 's.

individuals. The results of Figure 8 inform us that the timing for students reaching the “learning ceiling” (i.e., plateau) as well as their “learning speed” differs among different individuals. Further investigation or clustering of students based on the shape of the growth curves might help teachers to tailor education practice or assignments for each individual student.

Moreover, we are able to summarize the results of other parameters in the model. The posterior median of  $\phi^{-1/2}$  is 0.0566, and its 95% CI is [0.0301, 0.0797]. The posterior estimates and 95% CIs for  $\tau$  and  $\beta$  are summarized in Figures 10 and 9, respectively. In Figure 10, all  $\tau_i$  values and their corresponding CIs are far away from 0 except  $\tau_2^{-1/2}$  and  $\tau_3^{-1/2}$ , which suggests the local dependence indeed exists in the EdSphere data sets. In Figure 10, all  $\beta_{i,1}$ 's values as well as their corresponding 95% CIs are above 0, which shows the data support the belief that the growth curve of one's reading ability is always increasing. Notice our method does not require any additional restriction on the parameters in the model; thus, the values of parameters are fully determined by the data. Also, we could see the values of  $\beta_{i,0}$ 's and  $\beta_{i,1}$ 's are varying a lot according to different individuals.

## 6. Conclusion

In this article, we proposed a Bayesian nonparametric two-level hierarchical model for the analysis of one's latent ability growth in educational testing for longitudinal scenarios. The advantage of our method is able to incorporate the monotonic shape constraints into the estimation of latent trajectories. Due to the flexibility of our nonparametric method, we are able to fit any monotonic continuous curve without further restrictions on the estimation of parameters. Therefore, the estimation of the slopes  $\beta_{i,1}$ 's without including 0 in their corresponding 95% CI indicates the monotonicity is supported by EdSphere data sets.

The latent trajectories of ability growth estimated from our approach can help educators or practitioners to better understand the behaviors of students in the study, such as the growth patterns (continuous increasing, sharp increasing, and etc.) and the timing that a student reaches the ceiling of increasing for his or her learning ability (i.e., the timing of reaching the plateau). Further studies on clustering those behavioral patterns of individuals will guide us to design education practice or teaching tailored to individuals. In addition, since the evidence of the local dependence assumption is generally strong from the analysis of EdSphere data sets, we can conclude that the use of random effects to model the local dependence seems to be necessary and successful.

As the MCMC computation is time-consuming and resource-demanding for our current approach, our next goal is to improve the efficiency of our computation by developing big data schemes, such as similar ideas of Wei, Wang, and Conlon (2017), to make the parallel computing possible so as to conveniently apply our approach to the entire data sets. With improvement of computation efficiency, we might be able to develop a computable evaluation criterion to assess the model fit of the proposed nonparametric model using the predictive score ideas (Gneiting & Raftery, 2007). Another potential extension is to explore covariates including the grade and others with their relationships to the growth of one's ability trajectory. This direction will facilitate us to group individuals based on their similar characteristics and encourage the development of personalized education.

### **Acknowledgments**

The authors are grateful to Dr. Carl Swartz from Highroad Learning company for generously sharing their data with us and to the Editor, Dr. Li Cai, and referees for numerous suggestions that significantly improved this article.

### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research of Dr. Yang Liu was part of his dissertation at the Department of Statistics, University of Connecticut, and was supported by start-up fund from Dr. Xiaojing Wang. Also, Dr. Xiaojing Wang wants to acknowledge the receipt of the National Science Foundation award (#1848451) for support for her research, authorship, and/or publication of this article.

### **References**

Albers, W., Does, R., Imbos, T., & Janssen, M. (1989). A stochastic growth model applied to repeated tests of academic knowledge. *Psychometrika*, 54, 451–466.

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, 88, 669–679.

Bartolucci, F., Pennoni, F., & Vittadini, G. (2011). Assessment of school performance through a multilevel latent Markov Rasch model. *Journal of Educational and Behavioral Statistics*, 36, 491–522.

Bollen, K. A., & Curran, P. J. (2004). Autoregressive latent trajectory (ALT) models a synthesis of two traditions. *Sociological Methods and Research*, 32, 336–383.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective* (Vol. 467). Hoboken, NJ: John Wiley.

Bornkamp, B., & Ickstadt, K. (2009). Bayesian nonparametric estimation of continuous monotone functions with applications to dose–response analysis. *Biometrics*, 65, 198–205.

Brezger, A., & Steiner, W. J. (2012). Monotonic regression based on Bayesian P-splines. *Journal of Business & Economic Statistics*, 26, 90–104.

Cho, S.-J., Athay, M., & Preacher, K. J. (2013). Measuring change for a multidimensional test using a generalized explanatory longitudinal item response model. *British Journal of Mathematical and Statistical Psychology*, 66, 353–381.

Choi, T., Kim, H.-J., & Jo, S. (2016). Bayesian variable selection approach to a Bernstein polynomial regression model with stochastic constraints. *Journal of Applied Statistics*, 23, 1–21.

Embretson, S. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika*, 56, 495–515.

Geiser, C., Bishop, J., Lockhart, G., Shiffman, S., & Grenard, J. L. (2013). Analyzing latent state-trait and multiple-indicator latent growth curve models as multilevel structural equation models. *Frontiers in Psychology*, 4, 1–23.

Gelfand, A. E., & Kuo, L. (1991). Nonparametric Bayesian bioassay including ordered polytomous response. *Biometrika*, 78, 657–666.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102, 359–378.

Green, P. J., & Hastie, D. I. (2009). Reversible jump MCMC. *Genetics*, 155, 1391–1403.

Hsieh, C.-A., von Eye, A., Maier, K., Hsieh, H.-J., & Chen, S.-H. (2013). A unified latent growth curve model. *Structural Equation Modeling: A Multidisciplinary Journal*, 20, 592–615.

Johnson, C., & Raudenbush, S. W. (2006). A repeated measures, multilevel Rasch model with application to self-reported criminal behavior. *Methodological Issues in Aging Research*, 5, 131–164.

Kim, S., & Camilli, G. (2014). An item response theory approach to longitudinal analysis with application to summer setback in preschool language/literacy. *Large-Scale Assessments in Education*, 2, 1.

Lin, L., & Dunson, D. B. (2014). Bayesian monotone regression using Gaussian process projection. *Biometrika*, 101, 303–317.

Martin, A. D., & Quinn, K. M. (2002). Dynamic ideal point estimation via Markov chain Monte Carlo for the U.S. Supreme Court, 1953–1999. *Political Analysis*, 10, 134–153.

McKay Curtis, S., & Ghosh, S. K. (2011). A variable selection approach to monotonic regression with Bernstein polynomials. *Journal of Applied Statistics*, 38, 961–976.

Neelon, B., & Dunson, D. B. (2004). Bayesian isotonic regression and trend analysis. *Biometrics*, 60, 398–406.

Ongaro, A., & Cattaneo, C. (2004). Discrete random probability measures: A general framework for nonparametric Bayesian inference. *Statistics & Probability Letters*, 67, 33–45.

Park, J. H. (2011). Modeling preference changes via a hidden Markov item response theory model. In G. L. Jones, S. Brooks, A. Gelman, & X.-L. Meng (Eds.), *Handbook of Markov chain Monte Carlo* (pp. 479–491). Boca Raton, FL: CRC Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Thousand Oaks, CA: Sage.

Shively, T. S., Sager, T. W., & Walker, S. G. (2009). A Bayesian approach to non-parametric monotone function estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71, 159–175.

Tan, E. S., Amberg, A. W., Does, R. J. M. M., & Imbos, T. (1999). Approximations of normal IRT models for change. *Journal of Educational and Behavioral Statistics*, 24, 208–223.

te Marvelde, J. M., Glas, C. A. W., Landeghem, G. V., & Damme, J. V. (2006). Application of multidimensional item response theory models to longitudinal data. *Educational and Psychological Measurement*, 66, 5–34.

Van Dorp, J. R., & Kotz, S. (2002). The standard two-sided power distribution and its properties: With applications in financial engineering. *The American Statistician*, 56, 90–99.

Van Dyk, D. A., & Park, T. (2008). Partially collapsed Gibbs samplers: Theory and methods. *Journal of the American Statistical Association*, 103, 790–796.

Verhagen, J., & Fox, J.-P. (2012). Longitudinal measurement in health-related surveys. A Bayesian joint growth model for multivariate ordinal responses. *Statistics in Medicine*, 32, 2988–3005.

Wang, X., & Berger, J. O. (2016). Estimating shape constrained functions using Gaussian processes. *SIAM/ASA Journal on Uncertainty Quantification*, 4, 1–25.

Wang, X., Berger, J. O., & Burdick, D. S. (2013). Bayesian analysis of dynamic item response models in educational testing. *The Annals of Applied Statistics*, 7, 126–153.

Wei, Z., Wang, X., & Conlon, E. M. (2017). Parallel Markov chain Monte Carlo for Bayesian dynamic item response models in educational testing. *Stat*, 6, 420–433.

## Authors

YANG LIU is a PhD student at Department of Statistics, University of Connecticut and is now a data scientist at Upstart Network, Inc., 711 N. High St., 7th Floor, Columbus, OH 43215; email: yang.5.liu@uconn.edu. His research interests include Bayesian item response theory models, model selection, and subgroup analysis.

XIAOJING WANG is an associate professor at Department of Statistics, University of Connecticut, Storrs, CT 06269; email: xiaojing.wang@uconn.edu. Her current research interests include Bayesian modeling, latent variable models and item response models, state space models and time series data analysis, Gaussian processes, subgroup analysis, and model selection.

Manuscript received November 28, 2016

First revision received April 9, 2018

Second revision received December 9, 2018

Third revision received July 1, 2019

Accepted September 18, 2019