The Annals of Applied Statistics
2013, Vol. 7, No. 1, 126–153
DOI: 10.1214/12-AOAS608

BAYESIAN ANALYSIS OF DYNAMIC ITEM RESPONSE MODELS IN EDUCATIONAL TESTING[1]

BY XIAOJING WANG, JAMES O. BERGER AND DONALD S. BURDICK

University of Connecticut, Duke University and MetaMetrics, Inc.

Item response theory (IRT) models have been widely used in educational measurement testing. When there are repeated observations available for individuals through time, a dynamic structure for the latent trait of ability needs to be incorporated into the model, to accommodate changes in ability. Other complications that often arise in such settings include a violation of the common assumption that test results are conditionally independent, given ability and item difficulty, and that test item difficulties may be partially specified, but subject to uncertainty. Focusing on time series dichotomous response data, a new class of state space models, called Dynamic Item Response (DIR) models, is proposed. The models can be applied either retrospectively to the full data or on-line, in cases where real-time prediction is needed. The models are studied through simulated examples and applied to a large collection of reading test data obtained from MetaMetrics, Inc.

1. Introduction.

1.1. *Background.* Item response theory (IRT) models are frequently used in modeling dichotomous data from educational tests, since they allow separate assessment of the ability of examinees and effectiveness of the test items. A typical one-parameter IRT model is of the form

$$\Pr(X_{il} = 1|\theta_i, d_l) = \mathrm{F}(\theta_i - d_l), \tag{1.1}$$

where θ_i indicates the ability of the ith person; d_l indicates the difficulty of the lth test item; the item response variable X_{il} could be either 0 or 1, corresponding to whether the lth test item taken by the ith person is answered correctly or not; and the item characteristic curve, $\mathrm{F}(\cdot)$, is a cumulative distribution function (c.d.f.) from a continuous distribution. When $\mathrm{F}(\cdot)$ is the standard logistic c.d.f., the one-parameter IRT model (1.1) becomes the famous Rasch model

$$\Pr(X_{il} = 1|\theta_i, d_l) = \frac{\exp(\theta_i - d_l)}{1 + \exp(\theta_i - d_l)}. \tag{1.2}$$

If $\mathrm{F}(\cdot) = \Phi(\cdot)$, where $\Phi(\cdot)$ is the standard normal c.d.f., then

$$\Pr(X_{il} = 1|\theta_i, d_l) = \Phi(\theta_i - d_l) \tag{1.3}$$

Received June 2011; revised October 2012.

[1]Supported by NSF Grant DMS-10-07773 and by MetaMetrics Inc.

Key words and phrases. IRT models, local dependence, random effects, dynamic linear models, Gibbs sampling, forward filtering and backward sampling.

defines the one-parameter Normal Ogive or Probit model. We will focus on the former model in the paper, for reasons to be discussed later, although analysis of the Probit model is actually easier and can be done with a simplified version of the methodology developed here.

The development of item response theory from the classical point of view owes much to the pioneering work of Lord (1953), Rasche (1961) and their colleagues. Among the many noteworthy contributions are Andersen (1970) and Bock and Lieberman (1970).

In classical IRT, it is assumed that the X_{il} are independent, given the person's ability θ_i and the difficulty levels d_l. This is often referred to as the *local independence* assumption. There are situations in which this assumption is violated. One such is computer adaptive testing, wherein the selection of the next test item typically depends specifically on the previous questions and answers.

The situation is less clear with what is studied herein, MetaMetrics' educational assessment program called Computer Adaptive Instruction and Testing (CAIT). With CAIT, a test pool of articles is selected for the student based on an estimate of his/her current ability; the student selects an article from this pool and the test questions (described later) are then generated before reading commences. Thus, in the environment of the CAIT, the possible violation in the local independence would arise from sources such as article selection by the student and test questions related to the same article so that overall understanding of the article could affect all answers; in this paper, such possible effects will be called *test effects*. Other factors that could cause violation of the local independence include health status and emotional status of the student on a given day; these will be referred to as *daily effects*. In the MetaMetrics scenario, there had been no previous demonstration of the violation of the local independence through the presence of test effects or daily effects, and there was a considerable interest in establishing such presence for possible enhancement of current models.

Pioneering papers that addressed the local dependence were Stout (1987, 1990), who introduced the essential dimensionality and the essential independence of a collection of test items, and Gibbons and Hedeker (1992), who considered the conditional dependence within identified subsets of items by allowing random effects in the analysis. More recent work in this direction is testlet response theory modeling, proposed by Bradlow, Wainer and Wang (1999). They defined the testlet as a subset of items; for example, they defined a reading comprehensive section in the SAT as the testlet. They then modified the classic IRT models by including a random effect term to represent the common factor affecting the responses in the testlet. Another approach to handle the local dependence is by the introduction of Markov structure, such as Jannarone (1986) where the conjunctive IRT kernel was introduced. A more recent paper concerned is Andrich and Kreiner (2010), where they modified the Rasch model by allowing the conditional probability of a response to an item to depend on the answer of a previous item.

For the modeling in this paper, the random effect approach will be followed. Indeed, two levels of random effects will be introduced to model the daily effects and test effects, respectively.

Another essential generalization of the IRT model lies in their applicability to analyze longitudinal data, that is, to deal with scenarios in which an individual is tested repeatedly over time; then, the interest typically centers on the growth of an ability of the individual. Embretson (1991) and Marvelde et al. (2006) presented a multidimensional Rasch model to represent the change of an ability as an initial ability and one or more modifiabilities. Based on the belief that a person's ability growth would be increasing over time, Albers et al. (1989), Tan et al. (1999) and Johnson and Raudenbush (2006) used linear or polynomial regression of the time variable to measure the growth of an ability; their analysis required the same time span and testing points for all examinees. Martin and Quinn (2002) modeled the transition of a voting preference as a first-order Markov process, where they assumed voting preference changes from the previous time point to a new point by a random shock; this work did not incorporate a time trend. Park (2011) supposed that changes in a voting preference were subject to discrete agent-specific regime changes and modeled the indicator of the preference regime changes as a first-order Markov process. Bartolucci, Pennoni and Vittadini (2011) analyzed test scores in mathematics observed over 3 years for public and private middle school students by a multilevel latent Markov Rasch model, where they described the dynamic transition of different levels of the individual ability also via a first-order Markov process.

Our approach to the longitudinal issue is based on a new class of dynamic linear models (DLM's) [see West and Harrison (1997) for background on DLM's]. The literature on DLM's or state space models, in the framework considered here of longitudinal binomial data, includes, for example, Carlin and Polson (1992), Fahrmeir (1992) and Czado and Song (2008) and the last three papers mentioned in the previous paragraph. Our models are distinguished from the literature by simultaneously allowing for the following features: (i) observations at variable and irregular time points; (ii) continuously changing ability, but with incorporation of knowledge concerning trends (e.g., increasing ability over time) in a nondogmatic way (thus accommodating, say, a drop in reading ability over a summer vacation); (iii) an analysis that is either individual or hierarchical across a group of individuals, the latter allowing for "borrowing strength" in estimates of certain overall parameters; (iv) either a retrospective analysis based on the full data or a real-time analysis and prediction for an individual based on the data to date.

Moreover, we consider the case in which the test item difficulties are nominally specified, as in CAIT, where the test items are often computer-generated and have theoretically determined difficulties. The actual item difficulties are quite uncertain, however, this uncertainty is also accommodated in our analysis. Previous papers that introduced random effects for item parameters include Sinharay, Johnson and Williamson (2003) and DeBoeck (2008).

1.2. *Testbed application.* The model developed in this paper is motivated by CAIT testing, as developed by MetaMetrics Inc. The main applied goals are as follows:

- The original goal is to assess the appropriateness of the local independence assumption for this type of data. This evolves into the goal of better understanding the nature of the daily and test effects.
- A second goal is to understand the growth in ability of students, by retrospectively producing the estimated growth trajectories of their latent abilities in the study.
- A third goal is to enable on-line prediction of one's ability (based solely on data obtained up to that point), to enable a better assignment of reading materials to match his/her ability and to enable teachers to better assist students.

The data considered is from a school district in Mississippi and consisted of 1983 students who registered over two years in a CAIT reading test program conducted by MetaMetrics Inc. The students were in different grades and entered and left the program at different times between 2007 and 2009. Individuals took tests on different days and had different time lapses between tests. Because of the long periods of testing, a fully adaptive model accommodating continual changes in ability is needed.

The data was generated during sessions in which a student read an article selected from a large bank of available articles. The articles in this bank had been assigned text complexity measured in Lexiles, using the Lexile Receptive Analyzer ®, a software developed by MetaMetrics Inc. to evaluate the semantic and syntactic complexity of a text. The Lexile measure represents either an individual's reading ability or the complexity of a text. The scale for Lexiles ranges from 0 to 1800, with 0 indicating no reading ability and 1800 being the maximum.

A session begins like this: a student selects from a generated list of articles having Lexile complexities in a range targeted to the current estimate of the student's ability. For the selected article, a subset of words from the article are 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. When a blank is encountered while reading the article, the student clicks it and then the true removed word along with three incorrect options called foils is presented. As with the target word, the foils are selected randomly according to a prescribed protocol. The student selects a word to fill in the blank from the four choices and an immediate feedback is provided in the form of the correct answer.

The dichotomous items produced by this procedure are called "Auto-Generated-Cloze" items. They are single-use items generated at the time of an encounter between a student and an article. If another student selects that same article to read, a new set of target words and foils is selected. Although it is not strictly impossible for an individual item to be taken by more than one student, such an occurrence is highly improbable. As a consequence, it is not feasible to obtain data-based estimates of item calibration parameters.

Instead, the difficulties of the items generated for an encounter between a student and an article can be modeled as a sample from an ensemble of item difficulties associated with the article. The text complexity in Lexiles provides a theoretical value for the ensemble mean. An estimated student ability in combination with assumptions about the ensemble allows calculation of a predicted success rate for the encounter. A comparison of the observed success rate with predicted, aggregated over many encounters, provides a basis for assessing the viability of the assumptions incorporated into the model. The predicted success rates in Table 1 in Stenner (2010) include the assumption that the mean of the ensemble of item difficulties for an article is given by its theoretical text complexity. The agreement with observed success rates supports that assumption.

Although MetaMetrics data is typically presented in Lexile units, there is a simple linear transformation from Lexiles to logit units. We will utilize the more common logit units for all data and results in this paper. Note that this also motivates the use of the logistic IRT model in this paper—to preserve compatibility with the MetaMetrics data.

1.3. *Preview.* Because of the complexity of the model considered (and of the testbed data set), as well as the need to incorporate prior information into the model, the analysis will be carried out using Bayesian methodology and Markov chain Monte Carlo (MCMC) computational techniques. A side benefit of using these methodologies is that all uncertainties in all quantities are combined in the overall assessment of inferential uncertainty. The MCMC procedure utilizes a novel combination of Gibbs sampling together with a block sampling scheme involving forward filtering and backward sampling.

In Section 2 we formally describe the proposed models to capture the dynamic changes in a person's ability as well as the local dependence between item responses. Section 3 presents the MCMC strategy to carry out the statistical inference. Section 4 tests the methodology on some simulated examples (where the truth is known). Section 5 applies the proposed models to the MetaMetrics data set. Section 6 draws conclusions from both statistical and psychological sides, and points out some directions for future studies.

2. Dynamic item response (DIR) models. This section formally introduces the proposed one-parameter DIR model. Although the focus is on generalizing one-parameter IRT models, it would be straightforward to similarly generalize two-parameter or three-parameter IRT models.

2.1. *The observation equation in DIR models.* In a typical one-parameter IRT model (1.1), the index of the item response X_{il} indicates the correctness of the ith person's answer to the lth question in a single test. Consider the more involved situation in which the individual completes a series of tests within a given day and over different days. Thus, the item response variable is $X_{i,t,s,l}$, which corresponds to the correctness of the answer of the lth item in the sth test on the tth day taken by the ith person. Here, $i = 1, \dots, n$; $t = 1, \dots, T_i$; $s = 1, \dots, S_{i,t}$; and $l = 1, \dots, K_{i,t,s}$.

Likewise, let $d_{i,t,s,l}$ represent the difficulty level of the lth item in the sth test at the tth day taken by the ith person. As described in the Introduction, we model the test difficulties as being nominally specified, but with uncertainty. Thus, we write

$$d_{i,t,s,l} = a_{i,t,s} + \varepsilon_{i,t,s,l}, \tag{2.1}$$

where $a_{i,t,s}$ indicates the ensemble mean difficulty for the items in the sth test taken by the ith person on the tth day, and $\varepsilon_{i,t,s,l}$ is the random deviation from this ensemble mean difficulty for the lth item within the sth test. In the scenario we consider, the value of $a_{i,t,s}$ is assumed to be known, from the theoretical analysis of text complexity, while it is assumed that $\varepsilon_{i,t,s,l}$ is a normal distribution with zero mean and specified variance σ^2 from the test design in the CAIT testing, which is denoted as $\varepsilon_{i,t,s,l} \sim \mathcal{N}(0, \sigma^2)$.

As mentioned in the Introduction, we will also incorporate a term of daily random effects, $\varphi_{i,t}$, as well as a term of test random effects, $\eta_{i,t,s}$, to account for the possible local dependence factors when person i takes several tests during day t. It is assumed that $\varphi_{i,t} \sim \mathcal{N}(0, \delta_i^{-1})$ and, letting $\eta_{i,t} = (\eta_{i,t,1}, \dots, \eta_{i,t,S_{i,t}})'$ denote the vector of test random effects on day t for individual i, that $\eta_{i,t} \sim \mathcal{N}_{S_{i,t}}(0, \tau_i^{-1}\mathbf{I} | \sum_{s=1}^{S_{i,t}} \eta_{i,t,s} = 0)$, with differing and unknown precision parameters δ_i and τ_i for each individual i. Here $\mathbf{I}$ is an $S_{i,t} \times S_{i,t}$ identity matrix. The multivariate normal distribution for $\eta_{i,t}$ is actually a singular multivariate normal distribution because it is conditioned on the sum of the day's test effects being zero, done to remove any possibility of confounding with the daily random effects. (In analysis and computation, this singular multivariate normal distribution is replaced by the corresponding lower-dimensional nonsingular multivariate normal distribution.)

Finally, at the observation level, the dichotomous test data is modeled as

$$\begin{aligned}
&\Pr(X_{i,t,s,l} = 1 | \theta_{i,t}, a_{i,t,s}, \varphi_{i,t}, \eta_{i,t,s}, \varepsilon_{i,t,s,l}) \\
&\quad = \mathrm{F}(\theta_{i,t} - d_{i,t,s,l} + \varphi_{i,t} + \eta_{i,t,s}) \\
&\quad = \mathrm{F}(\theta_{i,t} - a_{i,t,s} + \varphi_{i,t} + \eta_{i,t,s} + \varepsilon_{i,t,s,l}),
\end{aligned}$$

where $\theta_{i,t}$ represents the ith person's ability on day t; we are thus assuming that a person's ability is constant over a given day, although there could be random fluctuations captured by the $\varphi_{i,t}$ and $\eta_{i,t,s}$. Letting $\mathrm{F}(\cdot)$ be the logistic c.d.f., as previously discussed, results in

$$\begin{aligned}
&\Pr(X_{i,t,s,l} = 1 | \theta_{i,t}, a_{i,t,s}, \varphi_{i,t}, \eta_{i,t,s}, \varepsilon_{i,t,s,l}) \\
&\quad = \frac{\exp(\theta_{i,t} - a_{i,t,s} + \varphi_{i,t} + \eta_{i,t,s} + \varepsilon_{i,t,s,l})}{1 + \exp(\theta_{i,t} - a_{i,t,s} + \varphi_{i,t} + \eta_{i,t,s} + \varepsilon_{i,t,s,l})}.
\end{aligned} \tag{2.2}$$

2.2. *The system equation in DIR models.* As mentioned in the Introduction, both parametric growth models and Markov chain models have been utilized in contexts similar to that of this paper. Here we combine these ideas, through a generalization of dynamic linear models, to model an individual's ability growth trajectory over time. The proposed model is

$$\theta_{i,t} = \theta_{i,t-1} + c_i(1 - \rho\theta_{i,t-1})\Delta^+_{i,t} + w_{i,t}, \tag{2.3}$$

which has three terms, modeling how current ability, $\theta_{i,t}$ for the ith person on the tth day, relates to past ability and other factors. The first term is simply ability at the previous time point, $\theta_{i,t-1}$.

The second term is a parametric growth model. Here c_i can be thought of as the average growth rate of the ith person's ability over time and $\Delta^+_{i,t}$ is the time lapse between the person's tth test day and $(t-1)$th test day but truncated by a pre-specified maximum time interval $\Delta_{T_{\max}}$, that is, $\Delta^+_{i,t} = \min\{\Delta_{i,t}, \Delta_{T_{\max}}\}$; thus, $c_i\Delta^+_{i,t}$ would reflect the ability growth over the given time interval if the growth was indeed linear. However, this growth is truncated at $\Delta_{T_{\max}}$ (chosen herein to be 14 days), reflecting the fact that, when on vacation, the student's ability may not be growing. Furthermore, the growth rate often declines as ability increases (indeed ability typically eventually plateaus), so that a linear growth model is often unsuitable when $\theta_{i,t}$ becomes large. The "correction factor," $-\rho\theta_{i,t-1}$ in (2.3), compensates for this effect, slowing down the linear growth as the ability level becomes larger. ρ is the parameter controlling the rate of this adjustment, and could be known or unknown. In our testbed example, ρ is known, based on experiments conducted at MetaMetrics [Hanlon et al. (2010)]. In principle, ρ should be individual-specific, but it is distinguishable from c_i only as the individual's ability level is reaching maturation; our investigation of ability growth in the testbed data focuses on early age students, so only the c_i are made individual-specific.

As in all dynamic linear models, the third term, $w_{i,t}$ in (2.3), represents the random component of the change in the ith person's ability on the tth day. We assume it is $\mathcal{N}(0, \phi^{-1}\Delta_{i,t})$, where ϕ is unknown. Note that this presumes that the random component of a person's ability change has the variance proportional to the time period between test days. Note, also, that we suppose that ϕ is common across individuals. The reason for this is clear from (2.2), in which $\varphi_{i,t} \sim \mathcal{N}(0, \delta_i^{-1})$ have individual-specific δ_i; there would be a substantial risk of confounding in the likelihood between δ_i's and $\phi^{-1}\Delta_{i,t}$ if the time lapse between tests for the student were equally spaced.

It is possible to rewrite (2.3) as a first-order Markov process, and this is beneficial for computational reasons. Indeed, letting $\lambda_{i,t} = \theta_{i,t} - \rho^{-1}$ and $g_{i,t} = 1 - c_i\rho\Delta^+_{i,t}$, the system equation (2.3) becomes

$$\lambda_{i,t} = g_{i,t}\lambda_{i,t-1} + w_{i,t}, \tag{2.4}$$

where $w_{i,t} \sim \mathcal{N}(0, \phi^{-1}\Delta_{i,t})$, and this is in the form of a standard dynamic linear model. (Note that c_i and ϕ need to be known for this reduction.)

2.3. *DIR model summary.* To sum up, the one-parameter DIR model is constructed in two levels as follows:

$$\begin{aligned}
\text{System equation:} \quad & \theta_{i,t} = \theta_{i,t-1} + c_i(1 - \rho\theta_{i,t-1})\Delta_{i,t}^{+} + w_{i,t}, \\
\text{Observation equation:} \quad & \Pr(X_{i,t,s,l} = 1 | \theta_{i,t}, a_{i,t,s}, \varphi_{i,t}, \eta_{i,t,s}, \varepsilon_{i,t,s,l}) \\
& \quad = \frac{\exp(\theta_{i,t} - a_{i,t,s} + \varphi_{i,t} + \eta_{i,t,s} + \varepsilon_{i,t,s,l})}{1 + \exp(\theta_{i,t} - a_{i,t,s} + \varphi_{i,t} + \eta_{i,t,s} + \varepsilon_{i,t,s,l})},
\end{aligned}$$

where $w_{i,t} \sim \mathcal{N}(0, \phi^{-1}\Delta_{i,t})$, $\varepsilon_{i,t,s,l} \sim \mathcal{N}(0, \sigma^2)$, $\varphi_{i,t} \sim \mathcal{N}(0, \delta_i^{-1})$, $\eta_{i,t} \sim \mathcal{N}_{S_{i,t}}(0, \tau_i^{-1}\mathbf{I} | \sum_{s=1}^{S_{i,t}} \eta_{i,t,s} = 0)$, and $\Delta_{i,t}^{+} = \min\{\Delta_{i,t}, \Delta_{T_{\max}}\}$, with the $a_{i,t,s}$, ρ, $\Delta_{i,t}$, $\Delta_{T_{\max}}$ and σ being known and $\theta_{i,t}$, c_i, ϕ, δ_i and τ_i being unknown.

3. Statistical inference for DIR models. In this section the Bayesian methods that will be used for statistical inference in DIR models are described. Computation is based on a Gibbs sampling scheme, in conjunction with forward filtering and backward sampling.

3.1. *Prior distributions for the unknown parameters.* Prior distributions in a Bayesian analysis must be specified carefully, but they can be either evidence-based priors, reflecting scientific knowledge of the system under study, or they can be objective priors, reflecting a lack of such knowledge but possessing good overall properties—for example, good frequentist properties [see, e.g., Berger (2006)]; a mix of both will be used in the analysis herein. Specification of evidence-based priors is, of course, context dependent and, here, will be done within the context of the MetaMetrics testbed application.

A natural choice of the prior distribution for an individual's initial latent ability, $\theta_{i,0}$, is

$$\theta_{i,0} \sim \mathcal{N}(\mu_{G_{j_i}}, V_{G_{j_i}}),$$

where $\mu_{G_{j_i}}$ and $V_{G_{j_i}}$ are the mean and the variance, on a logit scale, of the population (j) to which the individual i belongs—for instance, the individual's grade in school for the testbed application. For the average growth rate c_i in system equation (2.3), the natural objective prior is a constant prior (since c_i is a linear parameter), but we constrain c_i to be positive, reflecting the belief that there is a positive learning rate; thus, we choose the prior

$$\pi(c_i) \propto I(c_i > 0) \qquad \text{for all } i.$$

Although ϕ is a scale parameter, it occurs at the system-level of the two-stage model and, hence, the usual scale objective prior $(1/\phi)$ would result in an improper posterior; the computationally simplest adjustment is to use $\pi(\phi) = 1/\phi^{3/2}$, which does result in a proper posterior. Similarly, for the scale parameters δ_i and τ_i we utilize the objective priors $\pi(\delta_i) = 1/\delta_i^{3/2}$ and $\pi(\tau_i) = 1/\tau_i^{3/2}$. A natural alternative would be to try to “borrow information” across individuals, by utilizing gamma hyperpriors for the δ_i's and τ_i's. This complicates the computation, however, and does not seem necessary for the testbed application.

3.2. *Posterior distribution.* To facilitate the use of Gibbs sampling techniques in computation, we utilize a mixture of normals representation of the logistic distribution. From Andrews and Mallow (1974), if Y has a logistic distribution with location parameter 0 and scale $\pi^2/3$ ($\mathcal{L}(0, \frac{\pi^2}{3})$), one can write the density as

$$f(y) = \frac{e^{-y}}{(1+e^{-y})^2} = \int_0^\infty \left[\frac{1}{\sqrt{2\pi}}\frac{1}{2\nu}\exp\left\{-\frac{1}{2}\left(\frac{y}{2\nu}\right)^2\right\}\right]\pi(\nu)\,d\nu, \tag{3.1}$$

where ν has the Kolmogorov–Smirnov (K–S) density

$$\pi(\nu) = 8\sum_{\alpha=1}^{\infty}(-1)^{(\alpha+1)}\alpha^2\nu\exp\{-2\alpha^2\nu^2\}, \qquad \nu \geq 0. \tag{3.2}$$

Note that the density in square brackets in (3.1) is $\mathcal{N}(0, 4\nu^2)$. By using the idea of data augmentation from Tanner and Wong (1987), we consider the latent variable $Y_{i,t,s,l}$ for each response variable $X_{i,t,s,l}$, where $Y_{i,t,s,l} \sim \mathcal{N}(\theta_{i,t} - a_{i,t,s} + \varphi_{i,t} + \eta_{i,t,s} + \varepsilon_{i,t,s,l}, 4\nu_{i,t,s,l}^2)$ and define $X_{i,t,s,l} = 1$ if $Y_{i,t,s,l} > 0$ and $X_{i,t,s,l} = 0$ otherwise. It is then easy to show that $\Pr(X_{i,t,s,l} = 1|\theta_{i,t}, a_{i,t,s}, \varphi_{i,t}, \eta_{i,t,s}, \varepsilon_{i,t,s,l}) = \exp(\theta_{i,t} - a_{i,t,s} + \varphi_{i,t} + \eta_{i,t,s} + \varepsilon_{i,t,s,l})/(1 + \exp(\theta_{i,t} - a_{i,t,s} + \varphi_{i,t} + \eta_{i,t,s} + \varepsilon_{i,t,s,l}))$, so that the introduction of the latent variables $Y_{i,t,s,l}$ will not alter the model (except that there are now formally many more unknown parameters).

As $\varepsilon_{i,t,s,l} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$, it can be marginalized out in the distribution of $Y_{i,t,s,l}$, resulting in $Y_{i,t,s,l} \sim \mathcal{N}(\theta_{i,t} - a_{i,t,s} + \varphi_{i,t} + \eta_{i,t,s}, 4\nu_{i,t,s,l}^2 + \sigma^2)$. Therefore, the one-parameter DIR models (2.2) and (2.3) can be rewritten, with latent variables $\{Y_{i,t,s,l}\}$, as

$$\theta_{i,t} = \theta_{i,t-1} + c_i(1 - \rho\theta_{i,t-1})\Delta_{i,t}^+ + w_{i,t}, \tag{3.3}$$

$$Y_{i,t,s,l} = \theta_{i,t} - a_{i,t,s} + \varphi_{i,t} + \eta_{i,t,s} + \xi_{i,t,s,l}, \tag{3.4}$$

$$\nu_{i,t,s,l} \sim \text{K–S distribution}, \tag{3.5}$$

where $w_{i,t} \sim \mathcal{N}(0, \phi^{-1}\Delta_{i,t})$, $\varphi_{i,t} \sim \mathcal{N}(0, \delta_i^{-1})$, $\eta_{i,t} \sim \mathcal{N}_{S_{i,t}}(0, \tau_i^{-1}\mathbf{I}| \sum_{s=1}^{S_{i,t}} \eta_{i,t,s} = 0)$, and $\xi_{i,t,s,l} \sim \mathcal{N}(0, \psi_{i,t,s,l}^{-1})$ with $\psi_{i,t,s,l}^{-1} = 4\nu_{i,t,s,l}^2 + \sigma^2$.

Define $\theta = (\theta_1, \ldots, \theta_n)'$, where $\theta_i = (\theta_{i,0}, \theta_{i,1}, \ldots, \theta_{i,T_i})'$ for $i = 1, \ldots, n$; $c = (c_1, \ldots, c_n)'$ and $\tau = (\tau_1, \ldots, \tau_n)'$; $Y = \{Y_{i,t,s,l}\}$, $\nu = \{\nu_{i,t,s,l}\}$ and $X = \{X_{i,t,s,l}\}$ for $l = 1, \ldots, K_{i,t,s}$, $s = 1, \ldots, S_{i,t}$, $t = 1, \ldots, T_i$ and $i = 1, \ldots, n$; $\varphi = \{\varphi_{i,t}\}$ for $t = 1, \ldots, T_i$, $i = 1, \ldots, n$; $\eta = \{\eta_{i,t,s}\}$ for $s = 1, \ldots, S_{i,t}$, $t = 1, \ldots, T_i$ and $i = 1, \ldots, n$ and $\eta_{i,t}^* = (\eta_{i,t,1}, \ldots, \eta_{i,t,S_{i,t}-1})'$. Then the joint posterior density of θ, Y, c, τ, φ, η, ν and ϕ given the data X, in the one-parameter DIR model, is proportional to

$$
\begin{aligned}
&\pi(\theta, Y, c, \tau, \varphi, \eta, \nu, \phi|X) \\
&\quad \propto \left\{\prod_{i=1}^{n} \pi(\theta_{i,0})\pi(c_i)\pi(\delta_i)\pi(\tau_i)\right\}\pi(\phi)\left\{\prod_{i=1}^{n}\prod_{t=1}^{T_i}\prod_{s=1}^{S_{i,t}}\prod_{l=1}^{K_{i,t,s}} \pi(\nu_{i,t,s,l})\right\} \\
&\qquad \times \left\{\prod_{i=1}^{n}\prod_{t=1}^{T_i}\prod_{s=1}^{S_{i,t}}\prod_{l=1}^{K_{i,t,s}} \big(I\{Y_{i,t,s,l} > 0\}I\{X_{i,t,s,l} = 1\}\right. \\
&\qquad\qquad + I\{Y_{i,t,s,l} \le 0\}I\{X_{i,t,s,l} = 0\}\big) \\
&\qquad\qquad \times \sqrt{\frac{\psi_{i,t,s,l}}{2\pi}} \\
&\qquad\qquad \times \exp\left(-\frac{\psi_{i,t,s,l}(Y_{i,t,s,l} - \theta_{i,t} + a_{i,t,s} - \varphi_{i,t} - \eta_{i,t,s})^2}{2}\right) \\
&\qquad\qquad \left.\times I\left\{\eta_{i,t,S_{i,t}} = -\sum_{s=1}^{S_{i,t}-1} \eta_{i,t,s}\right\}\right\} \\
&\qquad \times \left\{\prod_{i=1}^{n}\prod_{t=1}^{T_i}\left(\frac{\tau_i}{2\pi}\right)^{(S_{i,t}-1)/2} \exp\left(-\frac{\tau_i \eta_{i,t}^{*\prime}\Sigma_{i,t}^{-1}\eta_{i,t}^{*}}{2}\right)\right\} \\
&\qquad \times \left\{\prod_{i-1}^{n}\prod_{t=1}^{T_i}\sqrt{\frac{\delta_i}{2\pi}}\exp\left(-\frac{\delta_i\varphi_{i,t}^2}{2}\right)\right\} \\
&\qquad \times \left\{\prod_{i=1}^{n}\prod_{t=1}^{T_i}\sqrt{\frac{\phi}{2\pi\Delta_{i,t}}}\exp\left(-\frac{\phi\{\theta_{i,t} - \theta_{i,t-1} - c_i(1-\rho\theta_{i,t-1})\Delta_{i,t}^{+}\}^2}{2\Delta_{i,t}}\right)\right\},
\end{aligned}
\tag{3.6}
$$

where

$$
\Sigma_{i,t}^{-1} = \begin{pmatrix} 2 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 2 \end{pmatrix}_{(S_{i,t}-1)\times(S_{i,t}-1)},
$$

and $I(Z \in A)$ is the indicator function equal to 1 if the random variable Z is contained in the set A; $\pi(\theta_{i,0})$, $\pi(c_i)$, $\pi(\delta_i)$, $\pi(\tau_i)$, $\pi(\phi)$ are the priors specified in the previous subsection, and $\pi(\nu_{i,t,s,l})$ is the K–S density defined at the beginning of this subsection. This is a proper posterior under very mild conditions; see Appendix C.

3.3. *Computation*. Computation is done by a MCMC scheme that samples from the posterior (3.6) via a block Gibbs sampling scheme, utilizing the forward filtering and backward sampling algorithm at a key point. The steps of the algorithm are given in Appendix A.

From the MCMC samples, statistical inferences are straightforward. For example, an estimate and 95% credible interval for the latent ability trait $\theta_{i,t}$ can be formed from the median, 2.5%, and 97.5% empirical quantiles of the corresponding MCMC realizations. In examples, these will be graphed as a function of t so that the adaptive nature of the model is apparent.

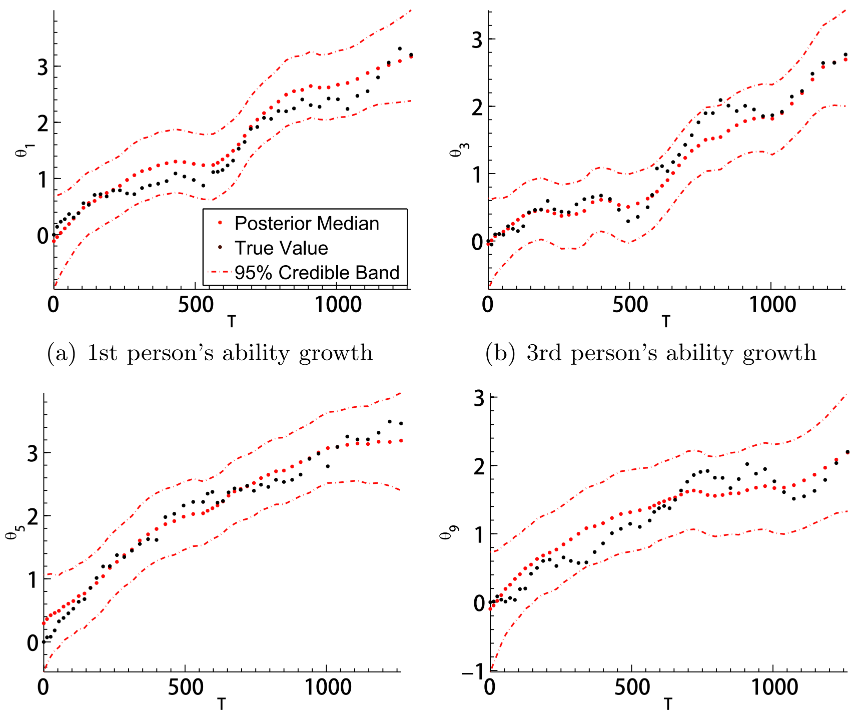
4. Simulated examples. In this section a simulated example is used to illustrate the inferences from the proposed one-parameter DIR models and to study their properties, primarily from a frequentist perspective.

The simulation examines the model's behavior for multiple individuals taking a series of tests that are scheduled during different time periods. In particular, suppose there are 10 individuals and each individual has taken tests on 50 different days. Thus, $n = 10$ and $T_i = 50$, for $i = 1, \dots, 10$. During each distinctive test day, the individual takes four tests; thus, $S_{i,t} = 4$ for $t = 1, \dots, 50$, $i = 1, \dots, 10$. Each test consists of 10 items, so that $K_{i,t,s} = 10$ for $s = 1, \dots, 4$, $t = 1, \dots, 50$ and $i = 1, \dots, 10$. For the ith person, the time lapse between two different tests is assumed to be a function of the tth day, that is, $\Delta_{i,t} = 10 + t$, for $i = 1, \dots, 10$, $t = 1, \dots, T_i/2$ and $\Delta_{i,t} = t - 10$, for $t = T_i/2, \dots, T_i$. Finally, the unknown values of parameters in the models are chosen as follows:

- $\phi = 1/0.0218^2$, and the corresponding standard deviation of the random component $w_{i,t}$ in system equation (2.3) is $0.0218\sqrt{\Delta_{i,t}}$.
- $c = (0.0055, 0.0065, 0.0026, 0.0037, 0.0061, 0.0047, 0.0035, 0.0043, 0.0039, 0.0015)'$, where each element in the vector c corresponds to the ith person's average growth rate, respectively, for $i = 1, \dots, 10$.
- $\delta = (2.0408, 1.3333, 1.8182, 1.2346, 1.5873, 1, 2.2222, 1.0526, 1.1494, 2)'$, where each element in the vector δ corresponds to the precision parameter of daily random effects for the ith person, respectively, $i = 1, \dots, 10$.
- $\tau = (4, 3.1250, 4.3478, 2.7027, 3.7037, 2.8571, 4, 2.2222, 9.0909, 4.5455)'$, where each element in the vector τ corresponds to the precision parameter of test random effects for the ith person, respectively, $i = 1, \dots, 10$.

According to the observation equation (2.2), we then simulated values for the unknown variables and set the test difficulties, $a_{i,t,s}$, to be $\theta_{i,t} + \zeta$, where ζ is a random variable with uniform distribution on $(-0.1, 0.1)$. The values of $\varepsilon_{i,t,s,l}$ were drawn from $\mathcal{N}(0, 0.7333^2)$ and the value of 0.7333 is used in the test design for MetaMetrics. Finally, we chose $\rho = 0.1180$, which is the value estimated by MetaMetrics in their studies [Hanlon et al. (2010)].

From dichotomous data obtained from the simulation, the Bayesian machinery from Section 3 was used in estimating the model parameters in (2.2) and (2.3). Figure 1 shows estimates of the ability trajectory for the 1st, 3rd, 5th and 9th individuals. The red dots in the figures correspond to the estimated posterior median of the



(a) 1st person's ability growth

(b) 3rd person's ability growth

(c) 5th person's ability growth

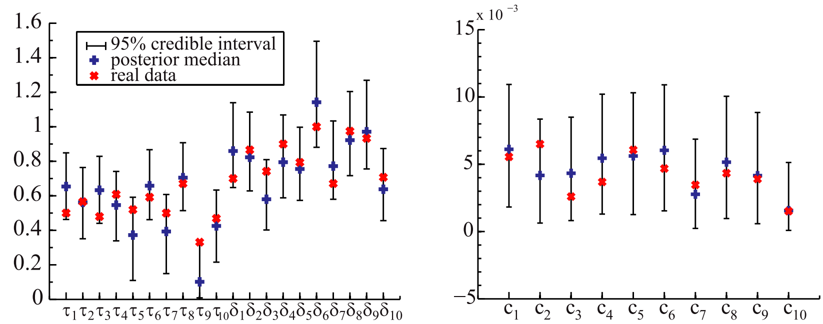
(d) 9th person's ability growth

FIG. 1. *Estimated and actual ability trajectories of* 4 *individuals from the simulated data.*

ability $\theta_{i,t}$ at the tth day for the ith person, and the red dashed lines give the 2.5% and 97.5% quantile trajectories of $\theta_{i,t}$, for $t = 1, \ldots, 50$. The black dots are the real abilities at the tth day for the ith person in the simulation. The third trajectory is typical of what is expected in terms of increasing ability, and is smoothly handled by the Bayesian machinery. The other three trajectories are highly nonmonotonic; the Bayesian estimates err in trying to be increasing (as they are designed to do), but do adapt to the nonmonotonicity when the evidence becomes strong enough.

One method of evaluating the success of the inferential scheme is to evaluate the percentage of time that the true ability, $\theta_{i,t}$, is contained in the 95% credible interval of estimated ability for each individual. For the ten individuals, these estimated coverages were 100%, 100%, 99%, 99%, 100%, 100%, 94%, 100%, 100% and 91%, which produce an overall estimated coverage of 98.3%. Thus, while the inferential method is Bayesian, it seems to be yielding sets that have good frequentist coverage.

To summarize the results for the c_i's, $\tau_i^{-1/2}$'s and $\delta_i^{-1/2}$'s , we compare their true values with the corresponding estimated values in Figure 2. In these plots, the black bar represents the 95% credible interval of the posterior distribution.



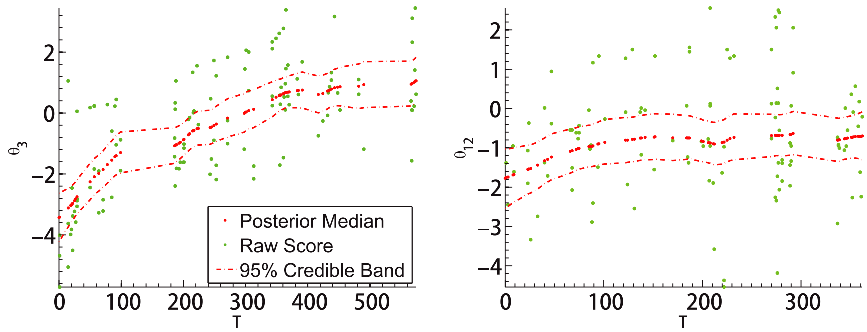
(a) 95% credible intervals of $\frac{1}{\sqrt{\delta_i}}$ and (b) 95% credible intervals of c_i

FIG. 2. 95% *credible intervals of* c_i, $\frac{1}{\sqrt{\tau_i}}$ *and* $\frac{1}{\sqrt{\delta_i}}$, *for* $i = 1, \ldots, 10$ *with the simulated data.*

The blue plus stands for the estimated posterior median and the red cross is the true value in the simulation. Moreover, the estimated posterior median of $\phi^{-1/2}$ is 0.0315 and its 95% credible interval is [0.0148, 0.0484]. Note that the true values of the c_i's, $\tau_i^{-1/2}$'s, $\delta_i^{-1/2}$'s and ϕ are all contained in the 95% credible intervals except $\tau_9^{-1/2}$; thus, the empirical coverage for these parameters is 96.77%.

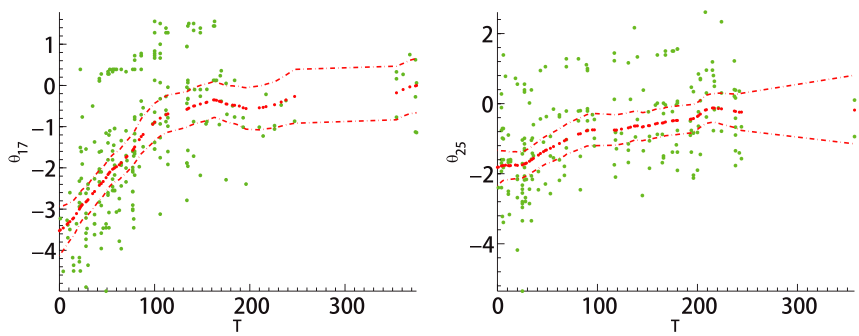
5. MetaMetrics testbed. In this section we apply the DIR model to the testbed MetaMetrics data. A sample of 25 individuals from the data base of students in certain elementary schools in Mississippi is considered here; the differing characteristics of the students are described in Appendix B. The primary focus is study of the goals mentioned in Section 1.2.

5.1. *Retrospective estimation of ability growth.* First consider retrospective estimation of the reading ability for an individual, utilizing all the data recorded for that individual. Figure 3 presents the resulting growth trajectories for the 3rd, 12th, 17th and 25th individuals studied. In Figure 3 the red dots are the posterior median estimates of each individual ability and the red dashes correspond to the 2.5% and 97.5% quantiles of the posterior distributions of the abilities, while the green dots correspond to estimates of an individual's abilities obtained by solving the equation that the expectation of expected score for a person's ability is equivalent to the observed score; these can roughly be thought of as the raw test scores put on the same scale as the $\theta_{i,t}$. The most interesting feature of these growth trajectories is that, while indeed there typically does appear to be overall growth in ability, this growth need not be monotone. In particular, when there is a large time gap between subsequent tests, the ability appears to drop for some individuals. One natural explanation is that, during vacations, a student may not read and could actually lose ability. Another possible explanation is that the student has become less adept at implementation of CAIT after a long break.



(a) 3rd person's ability growth

(b) 12th person's ability growth



(c) 17th person's ability growth

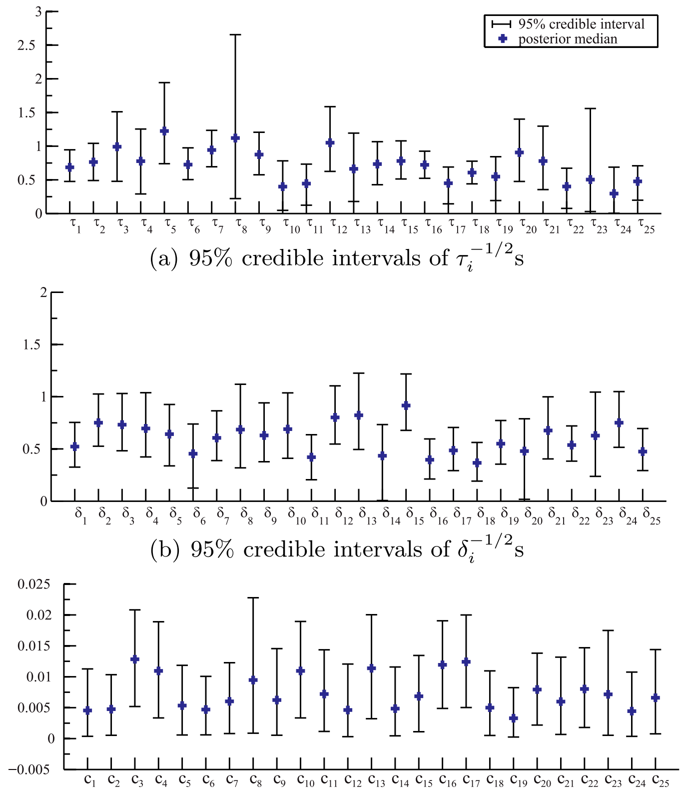
(d) 25th person's ability growth

FIG. 3. *Retrospective estimates of ability trajectories of* 4 *individuals from the MetaMetrics data.*

Figure 4 gives the summaries of the posterior distributions of the standard deviations of test random effects, $\tau_i^{-1/2}$'s, the standard deviations of the daily random effects, $\delta_i^{-1/2}$'s, and the average growth rates, c_i's, for $i = 1, \ldots, 25$. Moreover, the estimated posterior median of $\phi^{-1/2}$ is 0.0612 and its 95% credible interval is [0.0477, 0.0757].

Figures 4(a) and (b) show that the standard deviations of two random effects are almost all quite large with 95% credible intervals well separated from zero. Recall that these were included in the model to account for a possible lack of the local independence; the evidence is thus strong that the local independence is, indeed, not tenable for this data and that both types of random effects are present. The consistency of the standard deviations of the random effects across individuals is somewhat surprising, but lends credence to the notion that random effect modeling of the local dependence is fruitful.

5.2. *On-line estimation of ability growth.* In on-line estimation of reading ability, essentially the same model is used, but, at each time point, only the data up to that time is utilized. Instead of having $\phi^{-1/2}$ unknown, however, we utilize



(a) 95% credible intervals of $\tau_i^{-1/2}$s

(b) 95% credible intervals of $\delta_i^{-1/2}$s

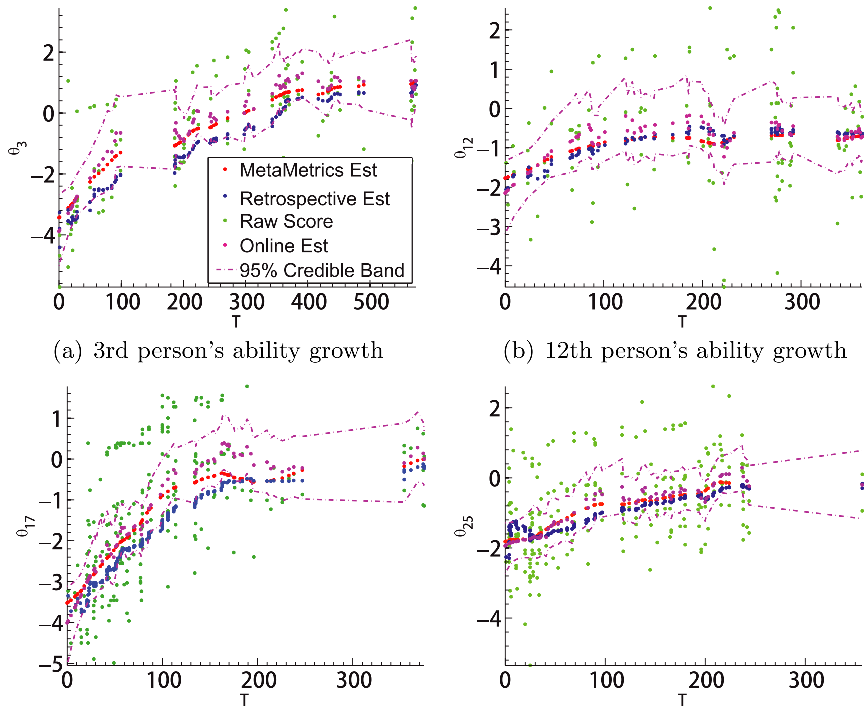
(c) 95% credible intervals of c_is

FIG. 4. *95% credible intervals of the $\tau_i^{-1/2}$s, $\delta_i^{-1/2}$'s and c_i's with the MetaMetrics data set.*

$\phi^{-1/2} = 0.0612$, the estimate arising from the retrospective analysis; $\phi^{-1/2}$ cannot be effectively estimated in an on-line mode.

Applying the Bayesian methodology yields on-line posterior median ability estimates, as well as the 2.5% and 97.5% quantiles of the posterior distribution of abilities for the 25 individuals being studied; these are the purple dots and and dashed purple lines in Figure 5, shown for the 3rd, 12th, 17th and 25th individuals. Again, the green dots show the raw score estimates of each individual ability at each time point, and the red dots are the retrospective estimates discussed earlier. In these figures we also include, as blue dots, the ability estimates obtained from the current methodology of MetaMetrics, which is a partial Bayesian procedure.

As expected, the on-line ability estimates are much more variable than the retrospective estimates. Sometimes, the on-line estimates seem to be somewhat more variable than the current MetaMetrics estimates (the blue dots). This is because at each online estimation point, the current methodology of MetaMetrics uses a very tight prior (arising from the previous data) for the student's ability.



(a) 3rd person's ability growth (b) 12th person's ability growth

(c) 17th person's ability growth (d) 25th person's ability growth

FIG. 5. *On-line estimates of ability trajectories of* 4 *individuals from the MetaMetrics data.*

While we do not know the truth here, it is plausible that the retrospective red dots are our best guesses as to the true abilities, and we can then judge how well the various on-line procedures are doing relative to these best guesses. Our on-line estimates are generally closer to these retrospective estimates than the current MetaMetrics estimates (the 12th individual being the interesting exception). In fact, the average mean squared error of our on-line estimates relative to the retrospective estimates is 0.0851, while the average mean squared error of the current MetaMetrics estimates is 0.1311.

If we do view the retrospective estimates (red dots) as surrogates for the truth, it is interesting to see how often these fall outside the on-line uncertainty bands (purple lines). This happened very rarely; individual 17 in Figure 5 was one case in which this sometimes happened. One final observation from Figure 5 is that the current MetaMetric estimates usually are lower than our on-line estimates of the person's reading ability.

6. Conclusions and generalizations. The evidence of violation of the local dependence assumption in CAIT situations is generally strong, and use of test and daily random effects to model the local dependence seems to be necessary and successful. Embedding a dynamic linear model framework for an individual's ability trajectory within the logistic IRT structure provides a powerful and individually adaptive method for dealing with longitudinal testing data.

The retrospective DIR model analysis seems excellent for assessing actual ability trajectories and, hence, is of considerable use in understanding population behavior, such as the frequently observed drops in ability after a long pause in testing. The on-line DIR analysis provides real-time ability estimates for assignments of material at the right difficulty level and other possible educational goals.

A key advantage of the Bayesian framework adopted is that uncertainty in all unknowns can be built into the model (e.g., uncertainty in the difficulty of the random test items), and uncertainty of the estimates is available for all inferences. Also, prior information (e.g., knowledge about ability distributions over the population and knowledge that general growth in ability is expected) can be built into the analysis, in a nondogmatic fashion that allows the data to overrule the prior.

Many extensions are possible, such as the already mentioned extension to two-parameter and three-parameter IRT models. If one also had data for individuals over a period of many years—including years near the maturation point in one's reading ability—it would be possible to include individual-specific ρ_i in the model.

APPENDIX A: THE MCMC COMPUTATION

The MCMC scheme that will be used to sample from the posterior (3.6) is a block Gibbs sampling scheme, utilizing the forward filtering and backward sampling algorithm at a key point. Because of the block Gibbs sampling scheme, we need only specify the conditional distributions of a block of variables given the data and other unknown variables.

A.1. Sampling Y: Truncated normal distribution sampling. Given θ, φ, η and ν, the latent variables $\{Y_{i,t,s,l}\}$ are sampled from

$$Y_{i,t,s,l} \sim \mathcal{N}_+(\theta_{i,t} - a_{i,t,s} + \varphi_{i,t} + \eta_{i,t,s}, \psi^{-1}_{i,t,s,l}) \qquad \text{if } X_{i,t,s,l} = 1,$$

$$Y_{i,t,s,l} \sim \mathcal{N}_-(\theta_{i,t} - a_{i,t,s} + \varphi_{i,t} + \eta_{i,t,s}, \psi^{-1}_{i,t,s,l}) \qquad \text{if } X_{i,t,s,l} = 0,$$

where $\mathcal{N}_+$ means the normal distribution truncated at the left by zero, while $\mathcal{N}_-$ is the normal distribution truncated at the right by zero and $\psi^{-1}_{i,t,s,l} = 4\nu^2_{i,t,s,l} + \sigma^2$. Sampling from truncated normals is fast and easy.

A.2. Sampling $\boldsymbol{\theta}$: Forward filtering and backward sampling. The latent ability vector $\theta_i = (\theta_{i,0}, \dots, \theta_{i,T_i})$, for each individual, is typically high-dimensional with highly correlated coordinates, so sampling of the variables would appear to be highly challenging. To overcome this roadblock, the proposed model is transformed so that θ_i could be block sampled—within a Gibbs sampling step conditional on the other parameters—by the highly efficient forward filtering and backward sampling algorithm.

To see this, consider ϕ, c, Y, φ, η and ν as given (the Gibbs sampling step). Define $Z_{i,t,s,l} = Y_{i,t,s,l} + a_{i,t,s} - \varphi_{i,t} - \eta_{i,t,s} - \rho^{-1}$ and utilize the formulation of the model in (2.4). Then, the (conditional) one-parameter DIR model fits the framework of dynamic linear models [West and Harrison (1997)], that is,

$$\text{System equation:} \qquad \lambda_{i,t} = g_{i,t}\lambda_{i,t-1} + w_{i,t},$$

$$\text{Observation equation:} \qquad Z_{i,t,s,l} = \lambda_{i,t} + \xi_{i,t,s,l},$$

where $w_{i,t} \sim \mathcal{N}(0, \phi^{-1}\Delta_{i,t})$, $\xi_{i,t,s,l} \sim \mathcal{N}(0, \psi_{i,t,s,l}^{-1})$ with $\psi_{i,t,s,l}^{-1} = 4\nu_{i,t,s,l}^2 + \sigma^2$. As indicated in West and Harrison (1997), the forward filtering and backward sampling algorithm to block update each vector θ_i proceeds as follows.

Since $\lambda_{i,0} = \theta_{i,0} - \rho^{-1}$ and $\theta_{i,0} \sim \mathcal{N}(\mu_{G_j}, V_{G_j})$, the conditional prior for $\lambda_{i,0}$ is $\lambda_{i,0} \sim \mathcal{N}(\mu_{G_j} - \rho^{-1}, V_{G_j})$. Define information available on the tth day for the ith person as

$$D_{i,t} = \{g_{i,q}, \phi, \psi, \varphi, \eta, c, Z_{i,q,1,1}, \dots, Z_{i,q,S_{i,q},K_{i,q,S_{i,q}}}\}_{q=1}^{t}.$$

We claim that the posterior distribution of $\lambda_{i,t}$ is then

$$\lambda_{i,t} | D_{i,t} \sim \mathcal{N}(\mu_{i,t}, V_{i,t}), \tag{A.1}$$

which can be verified by induction as follows. Assume that, on the $(t-1)$th day, the posterior of $\lambda_{i,t-1}$, given $D_{i,t-1}$, is $\mathcal{N}(\mu_{i,t-1}, V_{i,t-1})$. And it is easy to see this assumption is true when $t = 1$. Then, from the system equation, it is easy to establish that $\lambda_{i,t}|D_{i,t-1} \sim \mathcal{N}(d_{i,t}, R_{i,t})$ is a prior for $\lambda_{i,t}$, where $d_{i,t} = g_{i,t}\mu_{i,t-1}$ and $R_{i,t} = g_{i,t}^2 V_{i,t-1} + \phi^{-1}\Delta_{i,t}$. Therefore, we have

$$\begin{aligned}
\Pr(\lambda_{i,t}|D_{i,t}) &\propto \Pr(\lambda_{i,t}|D_{i,t-1}) \prod_{s=1}^{S_{i,t}} \prod_{l=1}^{K_{i,t,s}} \Pr(Z_{i,t,s,l}|\lambda_{i,t}) \\
&\propto \exp\left\{-\frac{R_{i,t}^{-1}(\lambda_{i,t} - d_{i,t})^2}{2}\right\} \\
&\quad \times \left\{\prod_{s=1}^{S_{i,t}} \prod_{l=1}^{K_{i,t,s}} \exp\left\{-\frac{\psi_{i,t,s,l}(Z_{i,t,s,l} - \lambda_{i,t})^2}{2}\right\}\right\}.
\end{aligned}$$

Then, at the tth day, the posterior distribution of $\lambda_{i,t}$ is as (A.1), where $\mu_{i,t} = V_{i,t}(R_{i,t}^{-1}d_{i,t} + \sum_{s=1}^{S_{i,t}} \sum_{l=1}^{K_{i,t,s}} \psi_{i,t,s,l} Z_{i,t,s,l})$ and $V_{i,t} = (\sum_{s=1}^{S_{i,t}} \sum_{s=1}^{K_{i,t,s}} \psi_{i,t,s,l} + R_{i,t}^{-1})^{-1}$.

The above updating procedure is called forward filtering and after it is complete and all quantities, that is, $\mu_{i,t}$ and $V_{i,t}$ are saved, we can begin the backward sampling of $\lambda_{i,t}$. For the time $t = T_i$, we sample $\lambda_{i,t}$ directly from $\mathcal{N}(\mu_{i,T}, V_{i,T})$. As the time from $t = (T_i - 1)$ to 0, at each time we draw $\lambda_{i,t}$ from

$$\lambda_{i,t}|\lambda_{i,t+1}, D_{i,t} \sim \mathcal{N}(h_{i,t}, H_{i,t}),$$

where $h_{i,t} = H_{i,t}(V_{i,t}^{-1}\mu_{i,t} + \phi g_{i,t+1}\Delta_{i,t+1}^{-1}\lambda_{i,t+1})$ and $H_{i,t} = (\phi g_{i,t+1}^2\Delta_{i,t+1}^{-1} + V_{i,t}^{-1})^{-1}$. This follows from

$$\begin{aligned}
\Pr(\lambda_{i,t}|\lambda_{i,t+1}, D_{i,t}) &\propto \Pr(\lambda_{i,t}|D_{i,t})\Pr(\lambda_{i,t+1}|\lambda_{i,t}, D_{i,t}) \\
&\propto \exp\left\{-\frac{V_{i,t}^{-1}(\lambda_{i,t} - \mu_{i,t})^2}{2}\right\} \\
&\quad \times \exp\left\{-\frac{\phi\Delta_{i,t+1}^{-1}(\lambda_{i,t+1} - g_{i,t+1}\lambda_{i,t})^2}{2}\right\}.
\end{aligned}$$

Thus, for $t = 0, \dots, T_i$, we set $\theta_{i,t} = \lambda_{i,t} + \rho^{-1}$ and each vector θ_i is sampled as a whole block, noticing that

$$\Pr(\theta_i|D_{i,T_i}) = \Pr(\theta_{i,T_i}|D_{i,T_i})\Pr(\theta_{i,T_i-1}|\theta_{i,T_i}, D_{i,T-1})\cdots\Pr(\theta_{i,0}|\theta_{i,1}, D_{i,0}).$$

A.3. Sampling c: Truncated normal distribution sampling. When θ and ϕ are given, the full conditional distribution of c_i is the truncated normal distribution

$$c_i \sim \mathcal{N}_+\left(\frac{\sum_{t=1}^{T_i}(1-\rho\theta_{i,t-1})(\theta_{i,t} - \theta_{i,t-1})\Delta_{i,t}^{+}\Delta_{i,t}^{-1}}{\sum_{t=1}^{T_i}(\Delta_{i,t}^{+}(1-\rho\theta_{i,t-1}))^2\Delta_{i,t}^{-1}}, \frac{1}{\phi\sum_{t=1}^{T_i}(\Delta_{i,t}^{+}(1-\rho\theta_{i,t-1}))^2\Delta_{i,t}^{-1}}\right).$$

A.4. Sampling η: Multivariate normal distribution sampling. When θ, φ, τ, Y and ν are given, if $S_{i,t} > 1$, then the full conditional distribution of $\eta_{i,t}^*$ is the multivariate normal distribution

$$\eta_{i,t}^* \sim \mathcal{N}_{S_{i,t}-1}\big((A_{i,t}^T\Sigma_{\psi_{i,t}}^{-1}A_{i,t} + \tau_i\Sigma_{i,t}^{-1})^{-1}A_{i,t}^T\Sigma_{\psi_{i,t}}^{-1}Y_{i,t}^*, (A_{i,t}^T\Sigma_{\psi_{i,t}}^{-1}A_{i,t} + \tau_i\Sigma_{i,t}^{-1})^{-1}\big),$$

where $Y_{i,t}^* = (Y_{i,t,1,1} - \theta_{i,t} + a_{i,t,1} - \varphi_{i,t}, \dots, Y_{i,t,1,K_{i,t,1}} - \theta_{i,t} + a_{i,t,K_{i,t,1}} - \varphi_{i,t}, \dots, Y_{i,t,S_{i,t},K_{i,t,S_{i,t}}} - \theta_{i,t} + a_{i,t,K_{i,t,S_{i,t}}} - \varphi_{i,t})'$, $\Sigma_{\psi_{i,t}}^{-1} = \operatorname{diag}((\psi_{i,t,1,1}, \dots, \psi_{i,t,S_{i,t},K_{i,t,S_{i,t}}})')$,

$$A_{i,t} = \begin{pmatrix} \mathbf{1}_{K_{i,t,1}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{K_{i,t,2}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_{K_{i,t,S_{i,t}-1}} \\ -\mathbf{1}_{K_{i,t,S_{i,t}}} & -\mathbf{1}_{K_{i,t,S_{i,t}}} & \cdots & -\mathbf{1}_{K_{i,t,S_{i,t}}} \end{pmatrix}_{(\sum_{s=1}^{S_{i,t}} K_{i,t,s})\times(S_{i,t}-1)},$$

where $\mathbf{1}_K$ is a K-dimensional column vector with each element being 1 and $\eta_{i,t,S_{i,t}} = -\sum_{s=1}^{S_{i,t}-1}\eta_{i,t,s}$. When $S_{i,t} = 1$, $\eta_{i,t,S_{i,t}} = 0$.

A.5. Sampling τ: Gamma distribution sampling. When η is given, the full conditional distribution of τ_i is the gamma distribution

$$\tau_i \sim \mathcal{G}a\left(\frac{\sum_{t=1}^{T_i} S_{i,t} - (T_i + 1)}{2}, \frac{\sum_{t=1}^{T_i} \eta_{i,t}^{*\prime} \Sigma_{i,t}^{-1} \eta_{i,t}^{*}}{2}\right).$$

A.6. Sampling φ: Normal distribution sampling. When θ, η, δ, Y and ν are given, the full conditional distribution of $\varphi_{i,t}$ is the normal distribution

$$\varphi_{i,t} \sim \mathcal{N}\left(\frac{\sum_{s=1}^{S_{i,t}} \sum_{l=1}^{K_{i,t,s}} \psi_{i,t,s,l}(Y_{i,t,s,l} - \theta_{i,t} + a_{i,t,s} - \eta_{i,t,s})}{\sum_{s=1}^{S_{i,t}} \sum_{l=1}^{K_{i,t,s}} \psi_{i,t,s,l} + \delta_i}, \frac{1}{\sum_{s=1}^{S_{i,t}} \sum_{l=1}^{K_{i,t,s}} \psi_{i,t,s,l} + \delta_i}\right).$$

A.7. Sampling δ: Gamma distribution sampling. When φ is given, the full conditional distribution of δ_i is the gamma distribution

$$\delta_i \sim \mathcal{G}a\left(\frac{T_i - 1}{2}, \frac{\sum_{t=1}^{T_i} \varphi_{i,t}^2}{2}\right).$$

A.8. Sampling ϕ: Gamma distribution sampling. When θ, c is given, the full conditional distribution of ϕ is the gamma distribution

$$\phi \sim \mathcal{G}a\left(\frac{\sum_{i=1}^{n} T_i - 1}{2}, \frac{\sum_{i=1}^{n} \sum_{t=1}^{T_i} \Delta_{i,t}^{-1}(\theta_{i,t} - \theta_{i,t-1} - c_i(1 - \rho\theta_{i,t-1})\Delta_{i,t}^{+})^2}{2}\right).$$

A.9. Sampling ν: Metropolis–Hastings sampling. Given Y, θ, φ and η, the full conditional distribution of $\nu_{i,t,s,l}$ is proportional to

$$\pi(\nu_{i,t,s,l} | Y, \theta, \varphi, \eta) \propto \sqrt{\frac{1}{\sigma^2 + 4\nu_{i,t,s,l}^2}} \times \exp\left\{-\frac{(Y_{i,t,s,l} - \theta_{i,t} + a_{i,t,s} - \varphi_{i,t} - \eta_{i,t,s})^2}{2(\sigma^2 + 4\nu_{i,t,s,l}^2)}\right\},$$

which is not in closed form. So we shall resort to a Metropolis–Hastings scheme to sample this distribution. A suitable proposal for sample ν is the K–S distribution itself. Thus, we first sample ν from the K–S distribution whose density is defined in (3.2). Then, we let

$$\nu_{i,t,s,l}^{(M)} = \begin{cases} \nu^*, & \text{with probability } \min(1, LR), \\ \nu_{i,t,s,l}^{(M-1)}, & \text{otherwise,} \end{cases}$$

where, given Y, θ, φ and η,

$$LR = \sqrt{\frac{\sigma^2 + 4(\nu_{i,t,s,l}^{(M-1)})^2}{\sigma^2 + 4(\nu^*)^2}} \exp\biggl\{-\frac{(Y_{i,t,s,l} - \theta_{i,t} + a_{i,t,s} - \varphi_{i,t} - \eta_{i,t,s})^2}{2} \times \biggl(\frac{1}{\sigma^2 + 4(\nu^*)^2} - \frac{1}{\sigma^2 + 4(\nu_{i,t,s,l}^{(M-1)})^2}\biggr)\biggr\},$$

and M indicates the Mth iteration step in MCMC.

A.10. Implementation. The Gibbs sampling starts at A.1, with initial values for $\theta^{(0)}$, $c^{(0)}$, $\phi^{(0)}$, $\varphi^{(0)}$, $\eta^{(0)}$, $\delta^{(0)}$, $\tau^{(0)}$ and $\nu^{(0)}$, and then loops through A.9 until the MCMC has converged. The initial values chosen in the applications were $\theta^{(0)} = \vec{0}$, $c^{(0)} = \vec{0}$, $\phi^{(0)} = 1$, $\varphi^{(0)} = \vec{0}$, $\eta^{(0)} = \vec{0}$, $\delta^{(0)} = \vec{1}$, $\tau^{(0)} = \vec{1}$ and $\nu^{(0)} = \vec{1}$, where we used "$\vec{a}$" here to indicate that each element of the corresponding vector or set has the same value "a". The convergence was evaluated informally by looking at trace plots, and was found to obtain at most after 30,000 of 50,000 iterations in the examples.

APPENDIX B: CHARACTERISTICS OF 25 STUDIED INDIVIDUALS

Twenty-five individuals from the MetaMetrics data base are studied in detail; the characteristics of the data for these individuals are described in Table 1.

APPENDIX C: POSTERIOR PROPRIETY

THEOREM 1. *Suppose $n \geq 2$ and, for $i = 1, \dots, n$, $T_i \geq 2$ and $S_{i,t} \geq 2$ for at least two days $t \in \{1, \dots, T_i\}$ with at least two of the tests on each of the two days having at least one* 0 *and one* 1 *observation. Then the posterior density of the DIR model is proper.*

We first give some needed lemmas that may be of independent interest for proving posterior propriety in other logistic modeling scenarios. Proofs of these lemmas are given in Appendix A of Wang (2012).

LEMMA 2. *For any three real numbers x, ε_1 and ε_2,*

$$\frac{e^{x+\varepsilon_1}}{1 + e^{x+\varepsilon_1}} \times \frac{1}{1 + e^{x+\varepsilon_2}} \leq e^{-|x|+|\varepsilon_1|+|\varepsilon_2|}.$$

LEMMA 3. *For $\theta_i \in (-\infty, \infty)$, $i = 1, 2$,*

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_0^{\infty} \tau^{-1/2} e^{-\tau(\eta_1^2+\eta_2^2)} e^{-(|\theta_1+\eta_1|+|\theta_1-\eta_1|+|\theta_2+\eta_2|+|\theta_2-\eta_2|)}\, d\tau\, d\eta_1\, d\eta_2 \leq K e^{-(|\theta_1|+|\theta_2|)},$$

with some constant K.

TABLE 1
Characteristics of the 25 considered individuals from the MetaMetrics data

	Total tests	**Days**	**Max. tests/days**	**Range of items/test**	**Max. gap**	**Grade**
No. 1	147	73	8	4–25	105	4
No. 2	162	64	9	3–17	102	2
No. 3	118	77	4	3–21	87	2
No. 4	93	53	4	5–25	147	2
No. 5	114	89	3	6–25	109	2
No. 6	157	57	29	4–20	116	2
No. 7	153	63	7	4–20	97	2
No. 8	60	50	5	3–24	168	6
No. 9	135	53	7	4–24	93	2
No. 10	137	54	6	4–17	219	1
No. 11	214	100	11	3–18	108	2
No. 12	113	76	4	4–16	45	2
No. 13	95	65	4	4–14	113	2
No. 14	116	57	6	5–17	107	2
No. 15	155	71	9	4–20	107	1
No. 16	247	76	13	3–19	113	2
No. 17	254	76	12	3–18	107	2
No. 18	304	53	31	3–12	49	2
No. 19	167	83	5	3–23	58	2
No. 20	101	68	9	4–23	117	2
No. 21	88	58	9	3–23	110	2
No. 22	220	96	8	2–23	104	3
No. 23	80	66	6	2–25	93	6
No. 24	105	60	6	6–24	62	3
No. 25	218	74	12	3–25	113	2

LEMMA 4. *For $\theta_i \in (-\infty, \infty)$, $i = 1, 2$,*

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} \delta^{-1/2} e^{-(\delta/2)(\varphi_1^2+\varphi_2^2)} e^{-(|\theta_1+\varphi_1|+|\theta_2+\varphi_2|)}\, d\delta\, d\varphi_1\, d\varphi_2 \le \frac{K}{1+|\theta_1|},$$

with some constant K.

LEMMA 5. *For $T \ge 2$,*

$$\int_0^{\infty}\int_0^{\infty}\int_0^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{\phi^{3/2}} \cdot \frac{1}{1+|\sqrt{B(c)/\phi}z + A(c)|} e^{-z^2/2} \\ \times \frac{1}{1+|\sqrt{B'(c')/\phi}z' + A'(c')|} e^{-z'^2/2}\, dz\, dz'\, dc\, dc'\, d\phi < \infty,$$

where

$$A(c) = \mu_{G_j} \prod_{t=1}^{T}(1 - c\rho\Delta_t^+) + \sum_{t=1}^{T} c\Delta_t^+ \prod_{i=t+1}^{T}(1 - c\rho\Delta_i^+),$$

$$B(c) = \sum_{t=1}^{T} \Delta_t \prod_{i=t+1}^{T} (1 - c\rho\Delta_i^+)^2 + \phi V_{G_j} \prod_{t=1}^{T} (1 - c\rho\Delta_t^+)^2,$$

$$A'(c') = \mu_{G_j} \prod_{t=1}^{T} (1 - c'\rho\Delta_t^+) + \sum_{t=1}^{T} c'\Delta_t^+ \prod_{i=t+1}^{T} (1 - c'\rho\Delta_i^+),$$

$$B'(c') = \sum_{t=1}^{T} \Delta_t \prod_{i=t+1}^{T} (1 - c'\rho\Delta_i^+)^2 + \phi V_{G_j} \prod_{t=1}^{T} (1 - c'\rho\Delta_t^+)^2,$$

and we have dropped the label i in the subscripts for $\Delta_{i,t}$, c_i, $\mu_{G_{j_i}}$ and $V_{G_{j_i}}$.

LEMMA 6. *For $T \geq 2$,*

$$\int_0^\infty \int_{-\infty}^\infty \frac{1}{1 + |\sqrt{B(c)/\phi}z + A(c)|} \exp\left\{-\frac{z^2}{2}\right\} dz\, dc < \infty,$$

with $A(c)$ and $B(c)$ defined in Lemma 5.

PROOF. In proving posterior propriety, it is easiest to work with the posterior density without the data augmentation, namely,

$$\begin{aligned}
&\pi(\theta, c, \tau, \eta, \varepsilon, \phi|X) \\
&\quad\propto \left\{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} V_{G_{j_i}}} \exp\left(-\frac{(\theta_{i,0} - \mu_{G_{j_i}})^2}{2V_{G_{j_i}}}\right) \mathbf{I}_{\{c_i \geq 0\}} \frac{1}{\tau_i^{3/2}} \frac{1}{\delta_i^{3/2}}\right\} \frac{1}{\phi^{3/2}} \\
&\qquad \times \left\{\prod_{i=1}^{n}\prod_{t=1}^{T_i}\prod_{s=1}^{S_{i,t}}\prod_{l=1}^{K_{i,t,s}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\varepsilon_{i,t,s,l}^2}{2\sigma^2}\right)\right\} \left\{\prod_{i=1}^{n}\prod_{t=1}^{T_i} \sqrt{\frac{\delta_i}{2\pi}} \exp\left(-\frac{\delta_i \varphi_{i,t}^2}{2}\right)\right\} \\
&\qquad \times \left\{\prod_{i=1}^{n}\prod_{t=1}^{T_i} \left(\frac{\tau_i}{2\pi}\right)^{(S_{i,t}-1)/2} \exp\left(-\frac{\tau_i \eta_{i,t}^{*\prime} \Sigma_{i,t}^{-1} \eta_{i,t}^*}{2}\right)\right\} \qquad \text{(C.1)} \\
&\qquad \times \left\{\prod_{i=1}^{n}\prod_{t=1}^{T_i}\prod_{s=1}^{S_{i,t}}\prod_{l=1}^{K_{i,t,s}} \frac{\exp[X_{i,t,s,l}(\theta_{i,t} - a_{i,t,s} + \varphi_{i,t} + \eta_{i,t,s} + \varepsilon_{i,t,s,l})]}{1 + \exp(\theta_{i,t} - a_{i,t,s} + \varphi_{i,t} + \eta_{i,t,s} + \varepsilon_{i,t,s,l})} \right. \\
&\qquad\qquad\qquad \left. \times I\left\{\eta_{i,t,S_{i,t}} = -\sum_{s=1}^{S_{i,t}-1} \eta_{i,t,s}\right\}\right\} \\
&\qquad \times \left\{\prod_{i=1}^{n}\prod_{t=1}^{T_i} \sqrt{\frac{\phi}{2\pi\Delta_{i,t}}} \exp\left(-\frac{\phi\{\theta_{i,t} - \theta_{i,t-1} - c_i(1 - \rho\theta_{i,t-1})\Delta_{i,t}^+\}^2}{2\Delta_{i,t}}\right)\right\}.
\end{aligned}$$

Noting that

$$\frac{\exp[X_{i,t,s,l}(\theta_{i,t} - a_{i,t,s} + \varphi_{i,t} + \eta_{i,t,s} + \varepsilon_{i,t,s,l})]}{1 + \exp(\theta_{i,t} - a_{i,t,s} + \varphi_{i,t} + \eta_{i,t,s} + \varepsilon_{i,t,s,l})} \leq 1,$$

an upper bound on the posterior density can be found by dropping all terms except the 0 and 1 test observations in the assumed tests for each individual. Utilizing Lemma 2 for each pair of observations 0 and 1 then results in the following upper bound on the posterior density (C.1):

$$
\begin{aligned}
&\frac{1}{\phi^{3/2}}\left\{\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}V_{G_{j_i}}}\exp\left(-\frac{(\theta_{i,0}-\mu_{G_{j_i}})^2}{2V_{G_{j_i}}}\right)\mathbf{I}_{\{c_i\geq 0\}}\frac{1}{\tau_i^{3/2}}\frac{1}{\delta_i^{3/2}}\right\}\\
&\quad\times\left\{\prod_{i=1}^{n}\prod_{t=1}^{T_i}\prod_{s=1}^{S_{i,t}}\prod_{\ell=1}^{K_{i,t,s}}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\varepsilon_{i,t,s,l}^2}{2\sigma^2}\right)\right\}\left\{\prod_{i=1}^{n}\prod_{t=1}^{T_i}\sqrt{\frac{\delta_i}{2\pi}}\exp\left(-\frac{\delta_i\varphi_{i,t}^2}{2}\right)\right\}\\
&\quad\times\left\{\prod_{i=1}^{n}\prod_{t=1}^{T_i}\left(\frac{\tau_i}{2\pi}\right)^{(S_{i,t}-1)/2}\exp\left(-\frac{\tau_i\eta_{i,t}^{*\prime}\Sigma_{i,t}^{-1}\eta_{i,t}^{*}}{2}\right)\right\}\\
&\quad\times\left\{\prod_{i=1}^{n}\prod_{t=1}^{T_i}\prod_{s=1}^{S_{i,t}}I\left\{\eta_{i,t,S_{i,t}}=-\sum_{s=1}^{S_{i,t}-1}\eta_{i,t,s}\right\}\right\}\\
&\quad\times\left\{\prod_{i=1}^{n}\exp\bigl(-|\theta_{i,t_i}+\varphi_{i,t_i}+\eta_{i,t_i,m}|+|a_{i,t_i,m}|+|\varepsilon_{i,t_i,m,k}|+|\varepsilon_{i,t_i,m,k'}|\bigr)\right.\\
&\qquad\times\exp\bigl(-|\theta_{i,t_i}+\varphi_{i,t_i}+\eta_{i,t_i,m'}|+|a_{i,t_i,m'}|+|\varepsilon_{i,t_i,m',h}|+|\varepsilon_{i,t_i,m',h'}|\bigr)\\
&\qquad\times\exp\bigl(-|\theta_{i,t_i'}+\varphi_{i,t_i'}+\eta_{i,t_i',r}|+|a_{i,t_i',r}|+|\varepsilon_{i,t_i',r,q}|+|\varepsilon_{i,t_i',r,q'}|\bigr)\\
&\qquad\left.\times\exp\bigl(-|\theta_{i,t_i'}+\varphi_{i,t_i'}+\eta_{i,t_i',r'}|+|a_{i,t_i',r}|+|\varepsilon_{i,t_i',r',g}|+|\varepsilon_{i,t_i',r',g'}|\bigr)\right\}\\
&\quad\times\left\{\prod_{i=1}^{n}\prod_{t=1}^{T_i}\sqrt{\frac{\phi}{2\pi\Delta_{i,t}}}\exp\left(-\frac{\phi\{\theta_{i,t}-\theta_{i,t-1}-c_i(1-\rho\theta_{i,t-1})\Delta_{i,t}^{+}\}^2}{2\Delta_{i,t}}\right)\right\}.
\end{aligned}
\tag{C.2}
$$

Ignoring multiplicative constants, and integrating out all the $\varepsilon_{i,t,s,l}$, (C.2) has an upper bound of

$$
\begin{aligned}
&\frac{1}{\phi^{3/2}}\left\{\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}V_{G_{j_i}}}\exp\left(-\frac{(\theta_{i,0}-\mu_{G_{j_i}})^2}{2V_{G_{j_i}}}\right)\mathbf{I}_{\{c_i\geq 0\}}\frac{1}{\tau_i^{3/2}}\frac{1}{\delta_i^{3/2}}\right\}\\
&\quad\times\left\{\prod_{i=1}^{n}\prod_{t=1}^{T_i}\sqrt{\frac{\delta_i}{2\pi}}\exp\left(-\frac{\delta_i\varphi_{i,t}^2}{2}\right)\right\}\\
&\quad\times\left\{\prod_{i=1}^{n}\prod_{t=1}^{T_i}\left(\frac{\tau_i}{2\pi}\right)^{(S_{i,t}-1)/2}\exp\left(-\frac{\tau_i\eta_{i,t}^{*\prime}\Sigma_{i,t}^{-1}\eta_{i,t}^{*}}{2}\right)\right\}
\end{aligned}
\tag{C.3}
$$

$$\times \left\{ \prod_{i=1}^{n} \prod_{t=1}^{T_i} \prod_{s=1}^{S_{i,t}} I \left\{ \eta_{i,t,S_{i,t}} = - \sum_{s=1}^{S_{i,t}-1} \eta_{i,t,s} \right\} \right\}$$

$$\times \left\{ \prod_{i=1}^{n} \exp\{-|\theta_{i,t_i} + \varphi_{i,t_i} + \eta_{i,t_i,m}|\} \exp\{-|\theta_{i,t_i} + \varphi_{i,t_i} + \eta_{i,t_i,m'}|\} \right.$$

$$\left. \times \exp\{-|\theta_{i,t_i'} + \varphi_{i,t_i'} + \eta_{i,t_i',r}|\} \exp\{-|\theta_{i,t_{i'}} + \varphi_{i,t_{i'}} + \eta_{i,t_i',r'}|\} \right\}$$

$$\times \left\{ \prod_{i=1}^{n} \prod_{t=1}^{T} \sqrt{\frac{\phi}{2\pi \Delta_{i,t}}} \exp\left(- \frac{\phi\{\theta_{i,t} - \theta_{i,t-1} - c_i(1-\rho\theta_{i,t-1})\Delta_{i,t}^{+}\}^2}{2\Delta_{i,t}} \right) \right\}.$$

We only consider here the "least information" case in which $S_{i,t_i} = S_{i,t_{i'}} = 2$; the more general case can be done similarly. Then $\eta_{i,t_i,m} = -\eta_{i,t_i,m'}$, $\eta_{i,t_{i'},r} = -\eta_{i,t_{i'},r'}$, $\exp(-\tau_i \eta_{i,t_i}^{*\prime} \Sigma_{i,t_i}^{-1} \eta_{i,t_i}^{*}/2) = \exp(-\tau_i \eta_{i,t_i,m}^2)$, and $\exp(-\tau_i \eta_{i,t_{i'}}^{*\prime} \times \Sigma_{i,t_{i'}}^{-1} \eta_{i,t_{i'}}^{*}/2) = \exp(-\tau_i \eta_{i,t_{i'},r}^2)$. Using this in (C.3) and integrating out all other η except for $\eta_{i,t_i,m}$ and $\eta_{i,t_{i'},r}$ and all φ except for φ_{i,t_i} and $\varphi_{i,t_i'}$, results in the expression

$$\frac{1}{\phi^{3/2}} \left\{ \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} V_{G_{j_i}}} \exp\left(- \frac{(\theta_{i,0} - \mu_{G_{j_i}})^2}{2 V_{G_{j_i}}} \right) \mathbf{I}_{\{c_i \ge 0\}} \right\}$$

$$\times \left\{ \prod_{i=1}^{n} \frac{1}{\delta_i^{3/2}} \frac{\delta_i}{2\pi} \exp\left(- \frac{\delta_i \varphi_{i,t_i}^2}{2} \right) \exp\left(- \frac{\delta_i \varphi_{i,t_i'}^2}{2} \right) \cdot \frac{1}{\tau_i^{3/2}} \right.$$

$$\times \frac{\tau_i}{2\pi} \exp\left(-\tau_i (\eta_{i,t_i,m}^2 + \eta_{i,t_i',r}^2)\right)$$

$$\times \exp\{-(|\theta_{i,t_i} + \varphi_{i,t_i} + \eta_{i,t_i,m}| + |\theta_{i,t_i} + \varphi_{i,t_i} - \eta_{i,t_i,m}|)\}$$

$$\left. \times \exp\{-(|\theta_{i,t_i'} + \varphi_{i,t_i'} + \eta_{i,t_i',r}| + |\theta_{i,t_i'} + \varphi_{i,t_i'} - \eta_{i,t_i',r}|)\} \right\}$$

$$\times \left\{ \prod_{i=1}^{n} \prod_{t=1}^{T_i} \sqrt{\frac{\phi}{2\pi \Delta_{i,t}}} \exp\left(- \frac{\phi\{\theta_{i,t} - \theta_{i,t-1} - c_i(1-\rho\theta_{i,t-1})\Delta_{i,t}^{+}\}^2}{2\Delta_{i,t}} \right) \right\}.$$

Next integrate out over τ_i, $\eta_{i,t_i,m}$ and $\eta_{i,t_i',r}$ using Lemma 3, resulting in the upper bound (again ignoring multiplicative constants)

$$\frac{1}{\phi^{3/2}} \left\{ \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} V_{G_{j_i}}} \exp\left(- \frac{(\theta_{i,0} - \mu_{G_{j_i}})^2}{2 V_{G_{j_i}}} \right) \mathbf{I}_{\{c_i \ge 0\}} \right\}$$

$$\text{(C.4)} \qquad \times \left\{ \prod_{i=1}^{n} \frac{1}{\delta_i^{3/2}} \frac{\delta_i}{2\pi} \exp\left(- \frac{\delta_i \varphi_{i,t_i}^2}{2} \right) \exp\left(- \frac{\delta_i \varphi_{i,t_i'}^2}{2} \right) \right.$$

$$\times \exp\{-(|\theta_{i,t_i} + \varphi_{i,t_i}| + |\theta_{i,t_i'} + \varphi_{i,t_i'}|)\}\Bigg\}$$

$$\times \left\{ \prod_{i=1}^{n} \prod_{t=1}^{T_i} \sqrt{\frac{\phi}{2\pi \Delta_{i,t}}} \exp\left(- \frac{\phi\{\theta_{i,t} - \theta_{i,t-1} - c_i(1 - \rho\theta_{i,t-1})\Delta_{i,t}^{+}\}^2}{2\Delta_{i,t}} \right) \right\}.$$

Next integrate out δ_i, φ_{i,t_i} and $\varphi_{i,t_i'}$ using Lemma 4. The resulting upper bound on (C.4) is

$$\frac{1}{\phi^{3/2}} \left\{ \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} V_{G_{j_i}}} \exp\left(- \frac{(\theta_{i,0} - \mu_{G_{j_i}})^2}{2V_{G_{j_i}}} \right) \mathbf{I}_{\{c_i \geq 0\}} \cdot \frac{1}{1 + |\theta_{i,t_i'}|} \right\}$$

$$\times \left\{ \prod_{i=1}^{n} \prod_{t=1}^{T_i} \sqrt{\frac{\phi}{2\pi \Delta_{i,t}}} \exp\left(- \frac{\phi\{\theta_{i,t} - \theta_{i,t-1} - c_i(1 - \rho\theta_{i,t-1})\Delta_{i,t}^{+}\}^2}{2\Delta_{i,t}} \right) \right\}.$$

Integrating out all the $\theta_{i,t}$ except the $\theta_{i,t_i'}$ results in the expression

$$\begin{aligned}
&\frac{1}{\phi^{3/2}} \left\{ \prod_{i=1}^{n} \mathbf{I}_{\{c_i \geq 0\}} \cdot \frac{1}{1 + |\theta_{i,t_i'}|} \right\} \\
&\quad \times \Bigg\{ \prod_{i=1}^{n} \sqrt{\frac{\phi}{2\pi V_{G_{j_i}}}} \\
&\qquad \times \Bigg(1 \Bigg/ \Bigg(\sum_{t=1}^{t_i'} \Delta_{i,t} \prod_{i=t+1}^{t_i'} (1 - c_i \rho \Delta_{i,t}^{+})^2 \\
&\qquad\qquad + \phi V_{G_{j_i}} \prod_{t=1}^{t_i'} (1 - c_i \rho \Delta_{i,t}^{+})^2 \Bigg) \Bigg)^{1/2} \\
&\quad \times \exp\Bigg(- \Bigg(\phi \Bigg(\theta_{i,t_i'} - \mu_{G_{j_i}} \prod_{t=1}^{t_i'} (1 - c_i \rho \Delta_{i,t}^{+}) \\
&\qquad\qquad - \sum_{t=1}^{t_i'} c_i \Delta_{i,t}^{+} \prod_{i=t+1}^{t_i'} (1 - c_i \rho \Delta_{i,t}^{+}) \Bigg)^2 \Bigg) \\
&\qquad \Bigg/ \Bigg(2 \Bigg(\sum_{t=1}^{t_i'} \Delta_{i,t} \prod_{i=t+1}^{t_i'} (1 - c_i \rho \Delta_{i,t}^{+})^2 + \phi V_{G_{j_i}} \prod_{t=1}^{t_i'} (1 - c_i \rho \Delta_{i,t}^{+})^2 \Bigg) \Bigg) \Bigg) \Bigg\}.
\end{aligned} \tag{C.5}$$

Finally, defining

$$z_i = \frac{\sqrt{\phi}(\theta_{i,t_i'} - A_i(c_i))}{\sqrt{B_i(c_i)}},$$

using Lemma 6 to integrate out all $\theta_{i,t_i'}$ and c_i, except for two individuals, and then using Lemma 5 for the remaining variables of (C.5), it follows that the integral is finite, completing the proof. □

Acknowledgments. The research of Xiaojing Wang was part of her dissertation at the Department of Statistical Science, Duke University. The authors are grateful to Jack Stenner, Hal Burdick, Carl Swartz and Sean Hanlon at MetaMetrics Inc. for valuable discussions, and to the Editor, Associate Editor and referees for numerous suggestions that significantly improved the paper.

REFERENCES

ALBERS, W., DOES, R. J. M. M., IMBOS, T. and JANSSEN, M. P. E. (1989). A stochastic growth model applied to repeated tests of academic knowledge. *Psychometrika* **54** 451–466. MR1030013

ANDERSEN, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *J. Roy. Statist. Soc. Ser. B* **32** 283–301. MR0273723

ANDREWS, D. F. and MALLOWS, C. L. (1974). Scale mixtures of normal distributions. *J. Roy. Statist. Soc. Ser. B* **36** 99–102. MR0359122

ANDRICH, D. and KREINER, S. (2010). Quantifying response dependence between two dichotomous items using the rasch model. *Appl. Psychol. Meas.* **34** 181–192.

BARTOLUCCI, F., PENNONI, F. and VITTADINI, G. (2011). Assessment of school performance through a multilevel latent Markov Rash model. *Journal of Educational and Behavioral Statistics* **36** 491–522.

BERGER, J. (2006). The case for objective Bayesian analysis. *Bayesian Anal.* **1** 385–402. MR2221271

BOCK, R. D. and LIEBERMAN, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika* **35** 179–197.

BRADLOW, E. T., WAINER, H. and WANG, X. (1999). A Bayesian random effects model for testlets. *Psychometrika* **64** 153–168.

CARLIN, B. P. and POLSON, N. G. (1992). Monte Carlo Bayesian methods for discrete regression models and categorical time series. *Bayesian Statistics* **4** 577–586.

CZADO, C. and SONG, P. X. K. (2008). State space mixed models for longitudinal observations with binary and binomial responses. *Statist. Papers* **49** 691–714. MR2439014

DE BOECK, P. (2008). Random item IRT models. *Psychometrika* **73** 533–559. MR2469780

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

FAHRMEIR, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models. *J. Amer. Statist. Assoc.* **87** 501–509.

GIBBONS, R. D. and HEDEKER, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika* **57** 423–436.

HANLON, S. T., SWARTZ, C. W., STENNER, A. J., BURDICK, H. J. and BURDICK, D. S. (2010). Oasis literacy research platform. **V1**, [Software]. Available from MetaMetrics, Inc.

JANNARONE, R. J. (1986). Conjunctive item response theory kernels. *Psychometrika* **51** 357–373.

JOHNSON, C. and RAUDENBUSH, S. W. (2006). A repeated measures, multilevel rasch model with application to self-reported criminal behavior. In *Methodological Issues in Aging Research* 131–164. Routledge, New York, NY.

LORD, F. M. (1953). The relation of test score to the trait underlying the test. *Educational Psychology Measurement* **13** 517–548.

MARTIN, A. D. and QUINN, K. M. (2002). Dynamic ideal point estimation via Markov chain Monte Carlo for the U.S. supreme count, 1953–1999. *Political Analysis* **10** 134–152.

PARK, J. H. (2011). Modeling preference changes via a hidden Markov item response theory model. In *Handbook of Markov Chain Monte Carlo* 479–491. CRC Press, Boca Raton, FL. MR2858458

RASCHE, G. (1961). On general laws and the meaning of measurement in psychology. In *Proc*. 4*th Berkeley Sympos*. *Math*. *Statist*. *and Prob*., *Vol*. *IV* 321–333. Univ. California Press, Berkeley, CA. MR0135196

SINHARAY, S., JOHNSON, M. S. and WILLIAMSON, D. M. (2003). Calibrating item families and summarizing the results using family expected response functions. *Journal of Educational and Behavioral Statistics* **28** 295–313.

STENNER, A. J. (2010). Using technology to merge assessment and instruction. In *2nd International Conference for Teaching and Learning with Technology*. Available from MetaMetrics, Inc.

STOUT, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika* **52** 589–617. MR0936600

STOUT, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika* **55** 293–325. MR1075575

TAN, E. S., AMBERGEN, A. W., DOES, R. J. M. M. and IMBOS, T. (1999). Approximations of normal IRT models for change. *Journal of Educational and Behavioral Statistics* **24** 208–223.

TANNER, M. A. and WONG, W. H. (1987). The calculation of posterior distributions by data augmentation. *J*. *Amer*. *Statist*. *Assoc*. **82** 528–550. MR0898357

TE MARVELDE, J. M., GLAS, C. A. W., VAN LANDEGHEM, G. and VAN DAMME, J. (2006). Application of multidimensional item response theory models to longitudinal data. *Educ*. *Psychol*. *Meas*. **66** 5–34. MR2247106

WANG, X. (2012). Bayesian modelling using latent structures. Ph.D. thesis, Duke Univ.

WEST, M. and HARRISON, J. (1997). *Bayesian Forecasting and Dynamic Models*, 2nd ed. Springer, New York. MR1482232

X. WANG
DEPARTMENT OF STATISTICS
UNIVERSITY OF CONNECTICUT
215 GLENBROOK ROAD, U-4120
STORRS, CONNECTICUT 06269-4120
USA
E-MAIL: xiaojing.wang@uconn.edu

J. O. BERGER
DEPARTMENT OF STATISTICAL SCIENCE
DUKE UNIVERSITY, P.O. BOX 90251
DURHAM, NORTH CAROLINA 27708-0251
USA
E-MAIL: berger@stat.duke.edu

D. S. BURDICK
METAMETRICS, INC., SUITE 120
1000 PARK FORTY PLAZA DRIVE
DURHAM, NORTH CAROLINA 27713
USA
E-MAIL: dburdick@lexile.com