JEM JOURNAL of EDUCATIONAL MEASUREMENT NCME national council on measurement in education

DOI: 10.1111/jedm.12348

Journal of Educational Measurement
Summer 2023, Vol. 60, No. 2, pp. 202–234

Check for updates

Fully Gibbs Sampling Algorithms for Bayesian Variable Selection in Latent Regression Models

Kazuhiro Yamaguchi
University of Tsukuba
Jihong Zhang
University of Iowa

This study proposed Gibbs sampling algorithms for variable selection in a latent regression model under a unidimensional two-parameter logistic item response theory model. Three types of shrinkage priors were employed to obtain shrinkage estimates: double-exponential (i.e., Laplace), horseshoe, and horseshoe+ priors. These shrinkage priors were compared to a uniform prior case in both simulation and real data analysis. The simulation study revealed that two types of horseshoe priors had a smaller root mean square errors and shorter 95% credible interval lengths than double-exponential or uniform priors. In addition, the horseshoe+ prior was slightly more stable than the horseshoe prior. The real data example successfully proved the utility of horseshoe and horseshoe+ priors in selecting effective predictive covariates for math achievement.

Introduction

One major interest in educational and psychological research is the exploration of factors that can explain latent psychological traits. In various educational contexts, such targeted latent traits are seen as indicators of academic proficiency, such as strong mathematical or linguistic abilities. Unidimensional item response theory models (IRT models; Embretson & Reise, 2000) have been employed to estimate students' academic proficiency in the Programme for International Student Assessment (PISA; OECD, 2019) or Trends in International Mathematics and Science Study (TIMMS; Mullis et al., 2020). In addition, latent proficiencies have been cross-sectionally compared among countries, with their growth tracked over time. Furthermore, students who participate in the educational assessments are usually asked various additional questions, such as their learning attitudes, habit of study, and background information in order to assess which factors influence academic proficiencies. Statistical models explaining latent proficiency with covariates are called *latent regression models* (von Davier & Sinharay, 2007, 2010) or *explanatory item response models* (De Boeck & Wilson, 2004). In this study, we use the term "latent regression models" rather than "explanatory item response models" hereafter. In short, latent regression models are regression models with a latent dependent variable.

The questionnaire for students in PISA or TIMMS has a lot of items (von Davier & Sinharay, 2010) from different aspects, but it is not obvious which variable should be

jihong-zhang@uiowa.edu

 © 2022 by the National Council on Measurement in Education.

included in the model before the data analysis. This is known as the variable selection problem (George, 2000). Usually, researchers choose covariates based on their research interests and theories. However, if important covariates are dropped from the data analysis, this causes the so-called omission variables problem, which results in biased parameter estimates. If too many variables are included, the regression model can be overfitted, and the generalization error becomes larger. This means that the prediction based on the estimated model becomes meaningless for the purpose of predicting new data. Moreover, the number of covariates is also related to the bias-variance trade-off (Bishop, 2006, p. 147; Jacobucci et al., 2019), which indicates that more complex and flexible models have smaller bias but larger variance, while simpler models show larger bias but smaller variance. Furthermore, a dependent variable in latent regression models is latent proficiency. Re-analyzing estimated latent proficiency scores can lead to biased estimates of regression coefficients (Grice, 2001). Therefore, statistical estimation methods that can simultaneously estimate latent proficiency and select covariates that have a large effect on the dependent variable are necessary for latent regression models.

In this study, Bayesian variable selection methods based on the Bayesian lasso (Park & Casella, 2008) and horseshoe shrinkage priors (Carvalho et al., 2009, 2010) were developed to measure data with a set of predictive covariates. We also developed Gibbs sampling algorithms to sample all parameters and latent variables from fully conditional posteriors. Thus, the sampling strategy is more efficient than the rejection sampling type Markov chain Monte Carlo methods (MCMC; Brooks et al., 2011). Another Bayesian way to select an appropriate variable is preparing a possible model set and assigning prior probabilities for each model (Ray & Szabó, 2021). This allow us to calculate posterior probability of each model, and it can be used for model selection. Moreover, information-criteria-based model selection is possible (Zhang et al., 2019). Bayesian method requires priors, which are sometimes determined subjectively (Ames & Smith, 2018), to get posterior. Shrinkage priors reflect the assumption that most coefficients are close to zero and provide the uncertainty of parameter estimates as posterior credible intervals. These credible intervals can be employed to select the coefficients that should be included (e.g., Li and Lin, 2010, p. 157; van Erp et al., 2019). In addition, Bayesian shrinkage estimation methods generally do not require tuning hyperparameters. The hyperparameter tuning is required for maximum likelihood (ML)-based regularization methods (lasso; Tibshiranit, 1996; elastic net; Zou & Hastie, 2005), which are often employed in psychometric research (Helwig, 2017; Jacobucci et al., 2019; McNeish, 2015). ML-based regularization methods are employed for variable selection and dealing with differential item functioning (DIF) in IRT study (Lee, 2020). The proposed Bayesian estimation method can be easily applied to real-life data analysis scenarios without complicated settings.

Bayesian variable selection methods have been actively studied in the field of statistical science. In a review, van Erp et al. (2019) summarized various shrinkage priors in a regression model with observed values and reported that regularized horseshoe and hyper lasso priors showed better performance in terms of the prediction mean squared error. However, van Erp et al. (2019) were limited to regular multiple linear regression models. Meanwhile, Culpepper and Park (2017) devel-

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

oped a Bayesian estimation method for multiple latent variable regression with a generalized asymmetric Laplace prior distribution. They employed a normal-ogive IRT model rather than a logistic IRT model, which required them to tune hyperparameters as the ML-based regularization method (Culpepper & Park, 2017, p. 603), even though their method is Bayesian. Therefore, multiple MCMC runs were applied to cross-validation samples to select appropriate hyper parameters controlling the strength of shrinkage, which minimizes the cross-validation error. This is usually very time consuming, especially in a Bayesian estimation setting when there is a large sample size. For example, if there are ten divided samples, the MCMC procedures are run ten times for each candidate of hyper parameters. This decreases utility of their method. Feng et al. (2017) developed a Bayesian (adaptive) lasso for a generalized latent variable model that could deal with continuous, ordinal, and nominal variables. Their method assumes a classical Laplace prior, which is also known as double-exponential prior, for the regression coefficients among latent proficiencies, where the prior is not theoretically appropriate. The horseshoe prior has a property of variable selection optimality (Bhadra et al., 2017) but a double-exponential prior does not. Specifically, the double-exponential prior is less robust as variable selection than the horseshoe prior. In summary, there is still a lack of effective Bayesian estimation methods with theoretically appropriate shrinkage priors (horseshoe priors) for latent regression models whose measurement model is logistic IRT models without hyperparameter tuning. We will review the theoretical aspects of shrinkage priors later.

To solve previously mentioned problem, the objective of this study is to develop effective Gibbs sampling methods with three types of shrinkage priors' settings: double-exponential (i.e., Laplace), horseshoe, and horseshoe+ priors for a latent variable regression situation in which the measurement model is a two-parameter logistic (2PL) IRT model. The novelty of this study lies in employing theoretically preferable horseshoes shrinkage priors that have not been employed for latent regression models. The contributions of this study are (1) enabling concurrent estimation of latent traits with a logistic item response model and predictive variable selection in a Bayesian manner, (2) providing Gibbs sampling algorithms that can not only sample the regression coefficient but also all parameters in the 2PL IRT model, and (3) comparing both simulation and real data situations to provide evidence of which prior setting is more appropriate for sparse latent regression situations. The horseshoe and horseshoe+ priors employ prespecified hyper parameters but have stronger shrinking effects than usual double-exponential priors, as seen later. For example, the latent regression coefficients of predictive covariates have shrinkage priors with the predetermined hyper parameters, in which no tuning is required (see more details in Sections 2.2.2 and 2.2.3). This point is attractive in time-consuming Bayesian estimation situations in latent variable models because the priors can skip multiple MCMC runs for hyperparameters selection.

The study is structured as follows. Section 2 introduces the latent variable regression model and derives the conditional posteriors for several shrinkage priors. Section 3 provides a Gibbs sampling algorithm using previously introduced conditional posteriors. Sections 4 and 5 discuss the simulation study conducted to compare the three shrinkage priors against a uniform prior setting as a reference. Section

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

6 presents the application of developed methods to real data. Finally, Section 7 provides suggestions about the proposed methods and offers recommendations regarding which shrinkage prior has the best performance in the latent regression situation and discusses the benefits and limitations of the three types of Bayesian variable selection methods.

Bayesian Variable Selection Methods for Latent Regression Models

Formulation of Latent Regression in Two-Parameter Logistic Item Response Model

The item response is represented as $y_{ij} \in \{0,\ 1\}$, which is a realization of a random variable Y_{ij} and if individual $i \in \{1, \dots,\ I\}$ correctly responds to the item $j \in \{1, \dots, J\}$, then y_{ij} takes one otherwise zero. The probability of a correct response of the 2PL IRT model is formulated with a latent proficiency variable $\theta_i \in \mathrm{R}$ and two item parameters, namely the discrimination parameter $a_j \in \mathrm{R}^+$ and difficulty parameter $b_j \in \mathrm{R}$:

$$\mathrm{P}\left(y_{ij} = 1|\theta_i, a_j, b_j\right) = \frac{1}{1 + \exp\left(-a_j\left(\theta_i - b_j\right)\right)}. \tag{1}$$

In the latent regression model, the θ_i is regressed on covariance $\boldsymbol{x}_i = (x_{i1}, \dots, x_{ik}, \dots, x_{ip})^\top$ $\boldsymbol{x}_i = (x_{i1}, \dots, x_{ik}, \dots, x_{ip})^\top$ in which the number of the covariance is p and $x_{ik} \in \mathrm{R},\ k\ =\ 1, 2, \dots, p$:

$$\theta_i = \mu_{\theta_i} + \epsilon_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \tag{2}$$

where $\boldsymbol{\beta}$ is a p -length vector of regression coefficients written as $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k, \dots, \beta_p)^\top$ and error term ϵ_i follows independently normal distribution: $\epsilon_i \sim N(0, \sigma^2)$. Here, $\mu_{\theta_i} = \boldsymbol{x}_i^\top\ \boldsymbol{\beta}$ is the conditional mean of latent proficiency of individual i\$ given covariate vector $\boldsymbol{x}_i$ and regression coefficients $\boldsymbol{\beta}$. This implies that the distribution of individual latent proficiency θ_i is a normal distribution with mean μ_{θ_i} and variance σ^2. Moreover, σ^2 or at least one of a_j should be fixed for identifiability. In this study, we fixed $\sigma^2 =\ 1$.

Before constructing the Bayesian variable selection Gibbs sampling for $\boldsymbol{\beta}$, we first constructed full conditional posteriors for the parameters of 2PL based on the discussion of Jiang and Templin (2019). The prior distributions for a_j and b_j are normal distributions whose means are μ_{a_j} and μ_{b_j} and variances are $\sigma^2_{a_j}$ and $\sigma^2_{b_j}$:

$$p\left(a_j|\mu_{a_j}, \sigma^2_{a_j}\right) = N\left(\mu_{a_j}, \sigma^2_{a_j}\right) I(a_j > 0), \tag{3}$$

$$p\left(b_j|\mu_{b_j}, \sigma^2_{b_j}\right) = N\left(\mu_{b_j}, \sigma^2_{b_j}\right), \tag{4}$$

where $I(\cdot)$ is an indicator function taking one if its argument is true, otherwise zero.

To derive full conditional posteriors for 2PL IRT parameters, we need to introduce an additional auxiliary variable w_{ij} following Pólya-Gamma distribution with parameters 1 and $a_j(\theta_i - b_j)$ (Polson et al., 2013; Scott & Sun, 2013), expressed as

$$w_{ij} \sim \text{PólyaGamma}\left(1, a_j\left(\theta_i - b_j\right)\right). \tag{5}$$

Let $d > 0$ and $c \in \mathrm{R}$ be the parameters of the Pólya-Gamma distribution and PólyaGamma$(d, c) = 1/2\pi^2 \sum_{l=1}^{\infty} \frac{g_l}{(l-1/2)^2 + c^2/(4\pi^2)}$, g_l follows a gamma distribution with parameters d and 1, which is the probability density function of variable x having a gamma distribution defined as $f(x; a, b) = \frac{b^a}{\Gamma(a)} x^{a+1} \exp(-bx)$, where $\Gamma(a)$ is a gamma function. Moreover, item response y_{ij} is converted as

$$k_{ij} = y_{ij} - \frac{1}{2}. \tag{6}$$

Based on Equations 7–9 in Jiang and Templin (2019) and the local independence assumption and exchangeability of individuals, the full conditional distribution of item discrimination parameter a_j is proportional to

$$p\left(a_j|\boldsymbol{y}_j,\ \boldsymbol{\theta}, b_j, \boldsymbol{k}_j, \boldsymbol{w}_j, \mu_{a_j}, \sigma^2_{a_j}\right) \propto p\left(a_j|\mu_{a_j}, \sigma^2_{a_j}\right)$$
$$\exp\left\{-\frac{1}{2}\sum_i w_{ij}\left(\frac{k_{ij}}{w_{ij}} - a_j\left(\theta_i - b_j\right)\right)^2\right\} I\left(a_j > 0\right), \tag{7}$$

where $\boldsymbol{y}_j$ is an I-length item response vector of item j defined as $\boldsymbol{y}_j = (y_{1j}, \ldots, y_{Ij})$, $\boldsymbol{\theta}$ is the Ith length latent proficiency vector $(\theta_1, \ldots, \theta_I)$, $\boldsymbol{k}_j$ is a vector of k_{ij}s defined as $(k_{1j}, \ldots, k_{Ij})$, and $\boldsymbol{w}_j$ is a vector of w_{ij}s written as $(w_{1j}, \ldots, w_{Ij})$. After additional calculation, the posterior distribution of item discrimination parameter a_j becomes a normal distribution:

$$p\left(a_j|\boldsymbol{y}_j,\ \boldsymbol{\theta}, b_j, \boldsymbol{k}_j, \boldsymbol{w}_j, \mu_{a_j}, \sigma^2_{a_j}\right) = N\left(\mu^*_{a_j}, \sigma^{2*}_{a_j}\right) I(a_j > 0), \tag{8}$$

where its mean $\mu^*_{a_j}$ and variance $\sigma^{2*}_{a_j}$ are

$$\begin{cases} \mu^*_{a_j} = \sigma^{2*}_{a_j}\left(\frac{\mu_{a_j}}{\sigma^2_{a_j}} + \sum_i k_{ij}\left(\theta_i - b_j\right)\right), \\ \sigma^{2*}_{a_j} = \left(\frac{1}{\sigma^2_{a_j}} + \sum_i w_{ij}\left(\theta_i - b_j\right)^2\right)^{-1}. \end{cases} \tag{9}$$

The same discussion can be applied to the derivation of posterior of b_j:

$$p\left(b_j|\boldsymbol{y}_j,\ \boldsymbol{\theta}, a_j, \boldsymbol{k}_j, \boldsymbol{w}_j, \mu_{b_j}, \sigma^2_{b_j}\right) \propto p\left(b_j|\mu_{b_j}, \sigma^2_{b_j}\right)$$
$$\exp\left\{-\frac{1}{2}\sum_i w_{ij}\left(\left(\frac{k_{ij}}{w_{ij}} + a_j b_j\right) - a_j\theta_i\right)^2\right\}, \tag{10}$$

where $\boldsymbol{b}$ is a J-length item difficulty vector $(b_1, \ldots, b_J)$. Therefore, the full conditional distribution of b_j is a normal distribution:

$$p\left(b_j|\boldsymbol{y}_j,\ \boldsymbol{\theta}, a_j, \boldsymbol{k}_j, \boldsymbol{w}_j, \mu_{b_j}, \sigma^2_{b_j}\right) = N\left(\mu^*_{b_j}, \sigma^{2*}_{b_j}\right), \tag{11}$$

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where

$$\begin{cases} \mu_{b_j}^* = \sigma_{b_j}^{2*} \left(\frac{\mu_{b_j}}{\sigma_{b_j}^2} + a_j \sum_i \left(k_{ij} - a_j \theta_i w_{ij} \right) \right), \\ \sigma_{b_j}^{2*} = \left(\frac{1}{\sigma_{b_j}^2} + a_j^2 \sum_i w_{ij} \right)^{-1}. \end{cases} \tag{12}$$

Finally, the full conditional distribution of latent proficiency parameter θ_i is

$$p\left(\theta_i | \boldsymbol{y}_i,\ \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{k}_i, \boldsymbol{w}_i, \mu_{\theta_i},\ \sigma^2 =\ 1\right) \propto p\left(\theta_i | \mu_{\theta_i},\ \sigma^2 =\ 1\right)$$
$$\exp\left\{ -\frac{1}{2} \sum_i w_{ij} \left(\left(\frac{k_{ij}}{w_{ij}} + a_j b_j \right) - a_j \theta_i \right)^2 \right\}, \tag{13}$$

where $\boldsymbol{y}_i$ is ith student item response vector expressed by $(y_{i1}, \ldots, y_{iJ})$, $\boldsymbol{k}_i$ is a vector defined as $(k_{i1}, \ldots, k_{iJ})$, and $\boldsymbol{w}_i$ is a vector represented as $(w_{i1}, \ldots, w_{iJ})$ and this becomes a normal distribution:

$$p\left(\theta_i | \boldsymbol{y}_i,\ \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{k}_i, \boldsymbol{w}_i, \mu_{\theta_i},\ \sigma^2 =\ 1\right) = N\left(\mu_{\theta_j}^*, \sigma_{\theta_j}^{2*}\right), \tag{14}$$

with parameters

$$\begin{cases} \mu_{\theta_i}^* = \sigma_{\theta_i}^{2*} \left(\frac{\mu_{\theta_i}}{\sigma^2} + \sum_j a_j \left(a_j b_j w_{ij} + k_{ij} \right) \right), \\ \sigma_{\theta_j}^{2*} = \left(\frac{1}{\sigma_{\theta_i}^2} + \sum_j a_j^2 w_{ij} \right)^{-1}. \end{cases} \tag{15}$$

Note that μ_{θ_i} equals to $\boldsymbol{x}_i^\top \boldsymbol{\beta}$ and σ^2 is fixed to one.

Bayesian Shrinkage Priors

A family of priors that aim to lead parameter estimates toward zero is called shrinkage priors (van Erp et al., 2019). In the Bayesian statistical method, shrinkage priors can be seen as an analog to the regularization term in ML methods. While various Bayesian shrinkage priors have been proposed (Bhadra et al., 2019; van Erp et al., 2019) for various purposes, we selected three famous priors that led to tractable Gibbs sampling algorithms: double-exponential (Laplace; Park & Casella, 2008), horseshoe (Carvalho et al., 2009, 2010), and horseshoe+ priors (Bhadra et al., 2017).

The double-exponential prior can be viewed as a Bayesian version of the ML lasso method (Park & Casella, 2008). This is one of the most famous shrinkage priors and has been employed in various data analyses (e.g., Feng et al., 2017). However, Bayesian variable selection with the double-exponential prior needs to specify hyperparameters similar to the method of Culpepper and Park (2017). We selected this prior as a reference in this study to compare horseshoe and horseshoe+ priors. Horseshoe priors can be classified as global-local priors (Bhadra et al., 2019), while a discrete mixture of point math of zero is known as spike-and-slab priors (Ishwaran & Rao, 2005), which has also been employed in variable selection problems.

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

However, spike-and-slab priors require heavy computation to sample from posteriors in a high-dimensional regression case (Bhadra et al., 2017). That is why the double-exponential and two horseshoe priors that can employ tractable Gibbs sampling algorithms are preferred in prior studies. For example, the sampling method for horseshoe priors has been shown in Makalic and Schmidt (2016a), while Makalic and Schmidt (2016b) provide full conditional posteriors under the horseshoe+ case. Comparing double-exponential, horseshoe, and horseshoe+ priors clarify the difference of shrinkage priors in latent regression model.

Other priors also provide shrinkage estimators (Bhadra et al., 2017) such as generalized Pareto priors (Armagan et al., 2013) or normal-gamma priors (Griffin & Brown, 2010). Compared to these priors, the horseshoe prior has variable selection optimality and is known as a robust variable selection method (Bhadra et al., 2017). This theoretically favorable nature is useful for interpretation when selecting variable selection methods. Furthermore, horseshoe priors are under the umbrella of a normal-scale mixture distribution family (Andrews & Mallows, 1974), which has a strong shrinkage effect for small value regression coefficients but not for large coefficients. In particular, horseshoe+ has stronger shrinkage effects than the usual horseshoe prior. The horseshoe+ prior-based Gibbs sampling algorithm is easily applied with only a small modification of the usual horseshoe prior settings. Other detailed theoretical properties of horseshoe priors have been described by Bhadra et al. (2017, 2019) and Carvalho et al. (2009, 2010). In a nutshell, horseshoe priors are considered cutting-edge priors of Bayesian shrinkage methods.

Double-exponential prior. In this section, we describe the full conditional posterior of regression coeffects under the double-exponential prior, which is the most fundamental Bayesian shrinkage prior. The double-exponential prior case Gibbs sampling algorithm can be compared to the horseshoe priors' case because it is a de facto standard prior for the Bayesian variable selection method. The double-exponential prior for the regression coefficients in Equation 2 is expressed as

$$p(\beta_k|\lambda) = \frac{\lambda}{2}\exp(-\lambda|\beta_k|). \tag{16}$$

However, random sampling from a double-exponential prior is inefficient. Park and Casella (2008) derived a hierarchical representation of the prior of β_k as

$$p(\beta_k|u_k) = N(0, u_k), \tag{17}$$

$$p\left(u_k|\lambda^2\right) = \text{Exp}\left(\frac{\lambda^2}{2}\right), \tag{18}$$

$$p\left(\lambda^2|a, b\right) = \text{Gam}(a_\lambda, b_\lambda), \tag{19}$$

where Exp(ψ) is an exponential distribution with parameter ψ whose probability density function is $f(x; \psi) = \psi\exp(-x\psi)$. Moreover, Gam$(a_\lambda, b_\lambda)$ are gamma distributions with shape parameter a_λ and rate parameter b_λ, which should be specified manually. Note that, in the previous literature formulation, the prior variance of β_k is $\sigma^2 u_k$ rather than u_k. However, in this study, the variance of the latent trait of individual proficiency was fixed at one: $\sigma^2 = 1$, while the prior variance of the regression

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

coefficient was also fixed at one. Under this formulation, if the latent proficiency vector $\boldsymbol{\theta}$ is obtained, the joint conditional distribution of $\boldsymbol{\beta}, \boldsymbol{u} = (u_1, \ldots, u_p)$, and λ^2 given $\boldsymbol{\theta}$ and $\mathbf{X}$ that is $I \times J$ data matrix $(\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_I^\top)^\top$ is proportional to the product of the complete data likelihood and the priors:

$$p\left(\boldsymbol{\beta}, \boldsymbol{u}, \lambda^2 | \boldsymbol{\theta}, \mathbf{X}, a_\lambda, b_\lambda\right) \propto p\left(\boldsymbol{\theta} | \mathbf{X}, \boldsymbol{\beta}\right) p\left(\boldsymbol{\beta} | \boldsymbol{u}\right) p\left(\boldsymbol{u} | \lambda^2\right) p\left(\lambda^2 | a, b\right). \tag{20}$$

The logarithm of the joint conditional distribution is

$$\log p\left(\boldsymbol{\beta}, \boldsymbol{u}, \lambda^2 | \boldsymbol{\theta}, \mathbf{X}, a_\lambda, b_\lambda\right) = -\frac{1}{2}(\boldsymbol{\theta} - \mathbf{X}\beta)^\top (\boldsymbol{\theta} - \mathbf{X}\beta) - \frac{1}{2}\sum_k \frac{\beta_k^2}{u_k} - \frac{\lambda^2}{2}\sum_k u_k$$

$$-\frac{1}{2}\sum_k u_k + (a_\lambda + p)\log\left(\lambda^2\right) - b_\lambda^2 \lambda^2 + C, \tag{21}$$

where C is the normalization constant.

From Equation 21, the full conditional posterior of $\boldsymbol{\beta}$ is a multivariate normal distribution:

$$p\left(\boldsymbol{\beta} | \boldsymbol{\theta}, \mathbf{X}, \boldsymbol{u}\right) = N\left(\boldsymbol{\mu}_\beta, \Sigma_\beta\right), \tag{22}$$

where

$$\begin{cases} \boldsymbol{\mu}_\beta = \Sigma_\beta \mathbf{X}^\top \theta, \\ \Sigma_\beta = \left(\mathbf{X}^\top \mathbf{X} + \mathbf{D}\right)^{-1}, \\ D = diag\left(u_1^{-1}, \ldots, u_p^{-1}\right). \end{cases} \tag{23}$$

The posterior distribution of u_k^{-1} can be expressed as

$$p\left(u_k^{-1} | \lambda^2, \beta_k\right) = \text{InvGauss}\left(\sqrt{\frac{\lambda^2}{\beta_k^2}}, \lambda^2\right), \tag{24}$$

where "InvGauss" is an inverse Gauss distribution with density function $f(x; \mu, \eta) = \sqrt{\frac{\eta}{2\pi}}\, x^{-\frac{3}{2}} \exp\{-\frac{\eta(x-\mu)^2}{2\mu^2 x}\}$, which μ is mean and η is the dispersion parameter (Giner & Smyth, 2016). Finally, the posterior distribution of λ^2 is the gamma distribution.

$$p\ \left(\lambda^2 | \boldsymbol{u}, a_\lambda, b_\lambda\right) = \text{Gam}\left(a_\lambda + p, b_\lambda + \frac{\sum_k u_k}{2}\right). \tag{25}$$

As mentioned above, the full conditional distributions are all tractable distributions. The double-exponential shrinkage prior sometimes overly shrinks large value coefficients (Carvalho et al., 2009) but it still plays an important role as a reference method.

Horseshoe prior. This section introduces the horseshoe prior. The horseshoe prior of the latent regression coefficient is

$$p\left(\beta_k | \lambda_k^2, \tau^2\right) = N\left(0, \lambda_k^2 \tau^2\right), \tag{26}$$

and local shrinkage λ_k^2 and global shrinkage τ^2 parameters with distributions

$$p\left(\lambda_k^2\right) = \mathrm{C}^+\left(0, 1\right), \tag{27}$$

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$p\left(\tau^2\right) = \mathrm{C}^+\left(0, 1\right), \tag{28}$$

where $\mathrm{C}^+(0, 1)$ is the standard half-Cauchy distribution with a probability density function $f\ (x; 0, 1) = \ 2/\pi(1 + x^2)$ for $x > 0$. The local shrinkage parameter λ_k^2 controls the strength of shrinkage of the kth regression coefficient. In addition, the global shrinkage τ^2 parameter controls the overall shrinkage of all regression coefficients. As mentioned before, σ^2 was fixed to one, and the variance term was expressed as $\lambda_k^2\tau^2$ rather than $\lambda_k^2\tau^2\sigma^2$, which was shown in previous sparse Bayesian regression literature. Similar to the double-exponential prior case, the half-Cauchy distribution can be expressed in a hierarchical form with two simple distributions (Makalic & Schmidt, 2016a). The prior distribution of λ_k^2 is decamped as

$$p\left(\lambda_k^2|\nu_k\right) = \mathrm{InvGam}\left(\frac{1}{2}, \frac{1}{\nu_k}\right), \tag{29}$$

$$p\left(\nu_k\right) = \mathrm{InvGam}\left(\frac{1}{2}, 1\right), \tag{30}$$

where “InvGam” is the inverse gamma distribution with parameters a and b: $f\ (x; a, b) = \frac{b^a}{\Gamma(a)}\ (\frac{1}{x})^{a+1}\ \exp(-\frac{b}{x})$. The same expression for τ^2 can be applied, and the result is

$$p\left(\tau^2|\xi\right) = \mathrm{InvGam}\left(\frac{1}{2}, \frac{1}{\xi}\right), \tag{31}$$

$$p\left(\xi\right) = \mathrm{InvGam}\left(\frac{1}{2}, 1\right). \tag{32}$$

These hierarchical representations allow us to derive full conditional posterior with given $\boldsymbol{\theta}$, and posterior is

$$p\left(\boldsymbol{\beta}, \boldsymbol{\lambda}^2, \boldsymbol{\nu}, \tau^2, \xi|\boldsymbol{\theta}, \mathbf{X}\right) \propto p\left(\boldsymbol{\theta}|\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\lambda}^2, \tau^2\right) p\left(\boldsymbol{\lambda}^2|\boldsymbol{\nu}\right) p\left(\boldsymbol{\nu}\right) p\left(\tau^2|\xi\right) p\left(\xi\right), \tag{33}$$

where $\boldsymbol{\lambda}^2 = \ (\lambda_1^2, \ldots, \lambda_p^2)$ and $\boldsymbol{\nu} \ = \ (\nu_1, \ \ldots, \nu_p)$ while independence is assumed among the regression coefficients. The logarithm of the posterior distribution is

$$\log p\left(\boldsymbol{\beta}, \boldsymbol{\lambda}^2, \boldsymbol{\nu}, \tau^2, \xi|\boldsymbol{\theta}, \mathbf{X}\right) = -\frac{1}{2}(\boldsymbol{\theta} - \mathbf{X}\beta)^\top(\boldsymbol{\theta} - \mathbf{X}\beta) - \left(\frac{p+1}{2} + 1\right)\log\left(\tau^2\right)$$
$$-2\sum_k \log\left(\lambda_k^2\right) - \frac{1}{2\tau^2}\sum_k \frac{\beta_k^2}{\lambda_k^2} - \sum_k \frac{1}{\nu_k}\left(\frac{1}{\lambda_k^2} + 1\right) - 2\sum_k \log\nu_k - 2\log\xi - \frac{1}{\xi\tau^2} - \frac{1}{\xi} \tag{34}$$

Sorting out Equation 34, the full conditional distribution of the regression coefficients $\boldsymbol{\beta}$ can be seen as a multivariate normal distribution, similar to the double-exponential prior case:

$$p\left(\boldsymbol{\beta}|\boldsymbol{\theta}, \boldsymbol{\lambda}^2, \tau^2\right) = N\left(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}}\right), \tag{35}$$

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where

$$\begin{cases} \boldsymbol{\mu}_{\beta} = \Sigma_{\beta}\, \mathbf{X}^{\top}\theta, \\ \Sigma_{\beta} = \left(\mathbf{X}^{\top}\mathbf{X} + \mathbf{D}\right)^{-1}, \\ D = diag\left(\lambda_1^{-2}, \ldots, \lambda_p^{-2}\right)/\tau^2. \end{cases} \tag{36}$$

The full conditional distributions of λ_k^2, ν_k, τ^2, and ξ are all inverse gamma distributions with different parameters:

$$p\left(\lambda_k^2|\beta_k, \tau^2, \nu_k\right) = \text{InvGam}\left(1, \frac{\beta_k^2}{2\tau^2} + \frac{1}{\nu_k}\right), \tag{37}$$

$$p\ \left(\nu_k|\lambda_k^2\right) = \text{InvGam}\left(1, 1 + \frac{1}{\lambda_k^2}\right), \tag{38}$$

$$p\ \left(\tau^2|\boldsymbol{\beta}, \boldsymbol{\lambda}^2\right) = \text{InvGam}\left(\frac{p+1}{2}, \frac{1}{\xi} + \frac{1}{2}\sum_k \frac{\beta_k^2}{\lambda_k^2}\right), \tag{39}$$

$$p\ \left(\xi|\tau^2\right) = \text{InvGam}\left(1, 1 + \frac{1}{\tau^2}\right). \tag{40}$$

The full conditional posteriors are a multivariate normal distribution or inverse gamma distribution. Shrinkage estimation can be achieved through a combination of simple well-known distributions. The detailed derivation is shown in Online Supplementary Material A.

The theoretical mechanism of how horseshoe prior shrink coefficients is obtained by the shrinkage weight $\kappa_k = 1/(1 + \lambda_k\tau)$. If $\kappa_k \approx 1$, the kth regression coeffect may be a strong signal, and if $\kappa_k \approx 0$, it indicates that the coefficient has totally shrunk to zero (Carvalho et al., 2010). The shrinkage prior defined in Equations 27 and 28 can be transformed to the shrinkage weight scale, which provides a beta distribution with two 1/2 parameters in the case $\tau = 1$: $p\ (\kappa_k) = \kappa_k^{1/2}\ (1 - \kappa_k)^{1/2}$. The distribution shape of κ_k concentrates on zero and one and looks like a horseshoe, which is the reason for the prior name.

Horseshoe+ prior. The extension of horseshoe prior to horseshoe+ prior is straightforward. We changed $p\ (\lambda_k^2) = \text{C}^+\ (0, 1)$ to

$$p\ \left(\lambda_k^2|\phi_k\right) = \text{C}^+\ (0, \phi_k), \tag{41}$$

$$p\ (\phi_k) = \text{C}^+\ (0, 1). \tag{42}$$

This is a hierarchy of two half-Cauchy distributions. The hierarchical representations of Equations 41 and 42 is

$$p\ \left(\lambda_k^2|\nu_k\right) = \text{InvGam}\left(\frac{1}{2}, \frac{1}{\nu_k}\right), \tag{43}$$

$$p\ \left(\nu_k|\phi_k^2\right) = \text{InvGam}\left(\frac{1}{2}, \frac{1}{\phi_k^2}\right), \tag{44}$$

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$p\left(\phi_k^2|\zeta_k\right) = \text{InvGam}\left(\frac{1}{2}, \frac{1}{\zeta_k}\right), \tag{45}$$

$$p\left(\zeta_k\right) = \text{InvGam}\left(\frac{1}{2}, 1\right). \tag{46}$$

With this hierarchical representation, the full conditional posterior of λ_k^2 is the same as Equation 37, and those of ν_k, ϕ_k^2, and ζ_k are

$$p\left(\nu_k|\lambda_k^2, \phi_k^2\right) = \text{InvGam}\left(1, \frac{1}{\phi_k^2} + \frac{1}{\lambda_k^2}\right), \tag{47}$$

$$p\left(\phi_k^2|\nu_k, \zeta_k\right) = \text{InvGam}\left(1, \frac{1}{\nu_k} + \frac{1}{\zeta_k}\right), \tag{48}$$

$$p\left(\zeta_k|\phi_k^2\right) = \text{InvGam}\left(1, 1 + \frac{1}{\phi_k^2}\right). \tag{49}$$

The full conditional posteriors of the other parameters are the same as in the horseshoe prior setting.

Sampling Algorithms for Shrinkage Priors in Latent Regression Models

We introduced the full conditional posterior distributions of the 2PL IRT model parameters and latent regression coefficients in the previous section. Combining them will prove fully Gibbs sampling algorithms. Introducing the upper script (m) for the MCMC iteration number, the Gibbs sampling algorithm of the double-exponential prior model is as follows:

1. Initialize $\boldsymbol{\theta}$, $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{u}$, and $\boldsymbol{\lambda}$, and set iteration counter $m = 0$.
2. Sample $\boldsymbol{\beta}^{(m+1)}$ from normal distribution in Equation 22.
3. Sample $u_k^{-1(m+1)}, k = 1, \ldots, p$, from inverse Gauss distribution in Equation 24.
4. Sample $\lambda^{2(m+1)}$ from gamma distribution in Equation 25.
5. Sample $w_{ij}^{(m+1)}$, $i = 1, \ldots, I$, $j = 1, \ldots, J$, from Pólya-Gamma distribution in Equation 5.
6. Sample $a_j^{(m+1)}$, $j = 1, \ldots, J$ from normal distribution in Equation 8.
7. Sample $b_j^{(m+1)}$, $j = 1, \ldots, J$ from normal distribution in Equation 11.
8. Sample $\theta_i^{(m+1)}$, $i = 1, \ldots, I$ from normal distribution in Equation 14, and $m = m + 1$.
9. Repeat Steps 2–8 until m reaches the prespecified iteration number.

For Horseshoe prior case,

1. Initialize $\boldsymbol{\theta}$, $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{\lambda}^2$, τ^2, and ξ, and set iteration counter $m = 0$.
2. Sample $\boldsymbol{\beta}^{(m+1)}$ from normal distribution in Equation 35.
3. Sample $\lambda_k^{2(m+1)}, k = 1, \ldots, p$, from inverse gamma distribution in Equation 37.

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4. Sample $\nu_k^{(m+1)}, k = 1, \ldots, p$, from inverse gamma distribution in Equation 38.
5. Sample $\tau^{2(m+1)}$ from inverse gamma distribution in Equation 39).
6. Sample $\xi^{(m+1)}$ from inverse gamma distribution in Equation 40.
7. Sample $w_{ij}^{(m+1)},\ i = 1, \ldots, I, j = 1, \ldots, J$, from Pólya-Gamma distribution in Equation 5.
8. Sample $a_j^{(m+1)},\ j = 1, \ldots, J$ from normal distribution in Equation 8.
9. Sample $b_j^{(m+1)},\ j = 1, \ldots, J$ from normal distribution in Equation 11.
10. Sample $\theta_i^{(m+1)},\ i = 1, \ldots, I$ from normal distribution in Equation 14 and $m = m + 1$.
11. Repeat Steps 2–10 until m reaches the prespecified iteration number.

Finally, for the horseshoe+ prior case, we slightly changed the fourth step in the horseshoe prior case and inserted the following sampling steps:

1. Sample $\nu_k^{2(m+1)}, k = 1, \ldots, p$, from inverse gamma distribution in Equation 47.
2. Sample $\phi_k^{2(m+1)}, k = 1, \ldots, p$, from inverse gamma distribution in Equation 48.
3. Sample $\zeta_k^{(m+1)}, k = 1, \ldots, p$, from inverse gamma distribution in Equation 49.

This small change can provide stronger shrinkage effects on the estimation of $\boldsymbol{\beta}$.

We also developed Gibbs sampling with a uniform prior for the regression coefficient $\boldsymbol{\beta}$ as another reference algorithm. The full conditional distribution given $\boldsymbol{\theta}$ and $\mathbf{X}$ becomes a multivariate normal distribution

$$p\ (\boldsymbol{\beta}|\boldsymbol{\theta}, \mathbf{X}) = N\left(\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\boldsymbol{\theta}, \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\right). \tag{50}$$

We can use this conditional distribution when sampling $\boldsymbol{\beta}$ instead of those with shrinkage priors. The sampling part of the 2PL IRT model parameters did not change.

Simulation Study 1

Simulation Settings

In the simulation study, we compared four types of Gibbs sampling methods introduced in this study: double-exponential, horseshoe, horseshoe+, and uniform priors with varied conditions. We manipulated three factors: (1) number of IRT items (10 and 30), (2) sample size (100 and 1,000), and (3) number of covariates (20 and 40), which leads to eight conditions in total. The true discrimination parameter a_j and difficulty parameter b_j were randomly generated from $Unif(1.3,\ 2.5)$ and $Unif(-2.5,\ 2.5)$, respectively, for each simulation replication, where $Unif(a, b)$ is a uniform distribution from a to b. The difficulty parameters covered a reasonably wide range of latent trait, and discrimination parameters were not particularly small or large. These settings were cleaner situation than real data situation (OECD, 2017; Appendix A). We start a relatively simple case to compare estimation algorithms.

The number of true nonzero regression coefficients was fixed to eight. The first four coefficients generated positive range uniform distributions, which were $\beta_1, \ldots, \beta_4 \sim Unif(1,\ 4)$, while the fifth to eighth coefficients generated a nega-

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

tive range uniform distribution, $\beta_5, \ldots, \beta_8 \sim Unif(-4,\ -1)$. The remaining 12 or 32 regression coefficients were fixed to zero. The covariates were assumed to follow a multivariate normal distribution with zero means and 0.5 correlation for all pairs of covariates and 0.2 standard deviation for all covariates. In the previous variable selection study, the study assumed a simulation setting with regression coefficient $\boldsymbol{\beta} = (3, 1.5, 0, 0, 2, 0, 0, 0)^\top$ and residual variance $\sigma^2 = 9$. Covariate $\mathbf{X}$ was generated from a multivariate normal distribution with 0 means and variance covariance matrix whose diagonal elements were 1 and the other elements were 0.5 (e.g., van Erp et al., 2019). We used $Unif(1,\ 4)$ to simulate strong coefficients in previous simulation studies. Moreover, it was not realistic to assume only positive coefficients, so $Unif(-4,\ -1)$ was employed to extend previous simulation. The range of uniform distribution covered previous nonzero coefficients, 1.5, 2, and 3. Further, we selected an $SD = 0.2$ of covariates to limit the deviation of dependent variable θ. When $SD = 1$, there was a large deviation of θ; so, smaller *SD*s were selected. The correlations were similar to previous simulation study. With the covariate and regression coefficients, true latent proficiencies θ_i were generated from a normal distribution with unit variance using Equation 2. Finally, an item response y_{ij} was randomly generated from the Bernoulli distribution with success probability $(1 + \exp(-a_j(\theta_i - b_j)))^{-1}$ with true parameters a_j, b_j, and θ_i. The four Gibbs sampling algorithms were applied to the same dataset to maintain comparability, and the simulation was replicated 50 times.

Hyper parameters a_λ and b_λ in the double-exponential prior were set to 1. For the MCMC setting, the number of MCMC chains was three, with 10,000 iterations for each chain. The first 5,000 iterations were discarded during the burn-in period. A convergence check was conducted for the regression coefficients, latent proficiency, discrimination, and difficulty parameters. The Gelman-Rubin $\hat{R}$ index (Brooks & Gelman, 1998) was selected as the convergence criterion, and if the value was $\hat{R} \leq 1.10$ then the MCMC chain is considered to have converged. The $\hat{R}$ was calculated using the CODA package (Plummer et al., 2006). The MCMC starting values were randomly selected. The posterior mean was employed for point estimates, except for the double-exponential prior case. The posterior median was used for the point estimate of the double-exponential prior because the posterior median rather than the mean was the shrinkage estimator for the prior setting.

The evaluation criteria of algorithms are bias, root mean square error (RMSE), and an average length of 95% credible interval (95% CI length). In the regression coefficients, bias, RMSE, and 95% CI length were calculated for each parameter, and they were averaged over 50 replications for the first eight coefficients (nonzero coefficients) and the other eight coefficients (zero coefficients). This is because the estimates behave differently for the nonzero and zero coefficients. According to previous studies, it was expected that horseshoe+ and horseshoe priors would show smaller bias and shorter 95% CI length than double-exponential priors. The uniform prior was expected to have the largest RMSE and longest 95% CI length among the four algorithms. The three evaluation indicators were also calculated for the IRT discrimination and difficulty parameters and were averaged over the items. The recovery of the latent trait parameter was evaluated as the average correlation between the estimated and true values. The reason for selecting correlation was as follows.

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The correlation and bias and RMSEs are different evaluation criteria. The individual latent trait recovery was a random effect that depends on the assumption of scale of location. Additionally, in the latent regression context, individual latent trait recovery was not a primary purpose. A high correlation between estimated and true traits was a necessary condition for an appropriate estimation method because if estimation method could not recover the latent traits, it was considered invalid. Biases and RMSEs were important evaluation criteria in our context. The purpose of latent regression model was to assess the effects of covariates on latent traits. Therefore, unbiasedness and smaller RMSEs of regression coefficients were preferable features of estimation methods.

Results

The $\hat{R}$ for all important parameters was less than 1.10, so the MCMC iterations were concluded to have converged. Table 1 lists the simulation results for the regression coefficients. The bias results were not very different for the four Gibbs sampling algorithms. In general, the nonzero coefficient showed a larger bias than the zero coefficients. Increments in sample size and number of items reduced bias. The number of increases in covariates inflated biases for nonzero coeffects. It should be noted that even for 10 items, 100 sample size, and 40 covariate conditions, the bias values were small—the largest bias was less than 0.1.

Explicit differences among the four were shown in RMSE and 95% CI length. In the 100-sample size and 10 item conditions, horseshoe and horseshoe+ priors had smaller RMSEs than double-exponential and uniform priors. Particularly in the 20 covariates case, the RMSE for zero coefficients of two horseshoe priors was half that of the double-exponential priors. Furthermore, in the case of 10 items, 100 sample size, 40 covariate conditions, horseshoe prior, double-exponential prior, and uniform prior indicated 1.978, 3.926, and 7.466, respectively, on the RMSE of nonzero coefficients. Approximately, the horseshoe prior was twice and four times more stable than the double-exponential and uniform priors, respectively, among nonzero coefficients. Under similar condition, horseshoe prior, double-exponential prior, and uniform prior indicated 0.492, 2.636, and 4.814, respectively, on the RMSE of zero coefficients. Again, the horseshoe prior was 5 times and 10 times more stable than the double-exponential and uniform priors, respectively, in zero coefficients. These results clearly indicated that the estimation with the horseshoe prior was stable than the double-exponential or uniform prior cases.

In addition, the 95% CI length for zero-coefficients of two horseshoe priors was approximately one point smaller than the double-exponential prior. The difference was larger in the 40 covariate conditions. Moreover, the 95% CI length for zero coefficients of two horseshoe priors was shorter than half of the uniform prior. The RMSE and 95% CI length for nonzero coefficients of the horseshoe priors in 100 sample size, 10 items, and 40 covariate conditions were much smaller than the double-exponential and uniform prior situation.

In a large-sample size setting, the difference in RMSEs between horseshoe priors and the other ones was smaller. However, even under 1,000 sample size conditions, the RMSEs of the horseshoe priors for zero coefficients were approximately half the

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1
First Simulation Result of Regression Coefficients

# Item	Sample Size	# Covariate	Nonzero/Zero	Double-Exponential			Horseshoe			Horseshoe+			Uniform		
				Bias	RMSE	95%CI Length	Bias	RMSE	95%CI Length	Bias	RMSE	95%CI Length	Bias	RMSE	95%CI Length
10	100	20	Nonzero	.059	2.081	4.056	.056	1.726	3.914	.061	1.835	3.879	.062	3.644	4.439
			Zero	–.021	1.466	3.565	–.024	.742	2.826	–.027	.701	2.576	–.018	2.319	4.346
		40	Nonzero	–.075	3.926	4.881	–.046	1.978	4.084	–.039	2.038	4.020	–.088	7.466	5.647
			Zero	.011	2.636	4.393	.009	.492	2.617	.007	.446	2.357	.011	4.814	5.494
	1,000	20	Nonzero	.008	.140	1.157	.006	.137	1.165	.005	.137	1.165	.009	.148	1.163
			Zero	–.007	.125	1.074	–.005	.080	.926	–.004	.069	.836	–.007	.137	1.139
		40	Nonzero	.023	.158	1.189	.020	.146	1.191	.019	.144	1.184	.024	.169	1.195
			Zero	–.003	.122	1.114	–.003	.049	.814	–.003	.043	.732	–.004	.132	1.168
30	100	20	Nonzero	.013	1.629	3.504	.009	1.378	3.558	.010	1.442	3.544	.022	2.238	3.737
			Zero	.015	1.384	3.133	.015	.785	2.666	.014	.719	2.437	.010	1.964	3.709
		40	Nonzero	–.027	2.427	4.000	–.041	1.817	3.673	–.041	1.911	3.625	–.010	3.396	4.403
			Zero	.007	1.848	3.667	.009	.584	2.411	.009	.552	2.161	.001	2.772	4.387
	1,000	20	Nonzero	–.012	.136	1.041	–.011	.131	1.049	–.010	.132	1.047	–.013	.144	1.047
			Zero	.008	.111	.969	.007	.077	.851	.006	.067	.771	.008	.120	1.027
		40	Nonzero	.013	.158	1.067	.009	.137	1.073	.008	.138	1.068	.014	.170	1.072
			Zero	–.003	.108	1.004	–.002	.047	.745	–.002	.041	.671	–.003	.116	1.053

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

double-exponential case or better. In summary, the horseshoe priors provided more stable estimates than the double-exponential prior or uniform prior setting, especially for small sample sizes. Horseshoe+ was slightly better than the usual horseshoe prior.

The simulation results of the discrimination and difficulty parameters in the IRT model are presented in Tables 2 and 3. In the result of the discrimination parameters (Table 2), the horseshoe priors showed smaller biases and RMSEs than the double-exponential and uniform priors, which was rather unexpected. Previous studies of horseshoe priors only mentioned regression coefficient, but the horseshoe priors indicated similar effects on the discrimination parameters in this study. In particular, biases and RMSEs for two horseshoe priors in the 10 items and 100 sample size conditions were substantially smaller than the other two prior settings. The 95% CI length results of the double-exponential and uniform priors tended to be slightly shorter than the horseshoe priors. However, larger sample sizes and items decreased the estimation difference among the four algorithms.

In the item difficulty parameter results in Table 3, the RMSEs of the horseshoe priors in the 10 items, 100 sample size, and 40 covariate conditions were also smaller than those with the other two prior settings. In addition, the RMSEs of the horseshoe priors were also smaller than the double-exponential or uniform prior in the case of 10 and 30 items, 100 sample size, and 20 and 40 covariates. The other results of the four algorithms were almost the same, but the horseshoe priors were slightly better in some cases. The correlation between true and estimated latent proficiency is shown in Table 4, and all four algorithms indicated sufficiently high correlations.

Simulation Study 2

Simulation Settings

The second simulation aims to check the estimation quality of four shrinkage priors when weak regression coefficients exist. In total, 8 of 30 regression coefficients are weak, 8 coefficients are strong, and the other 16 coefficients have zero effect. Specifically, 8 strong regression coefficients were generated from $Unif(1,\ 4)$ or $Unif(-4,\ -1)$, and the weak regression coefficients $\beta_9, \ldots, \beta_{12}$ and $\beta_{13}, \ldots, \beta_{16}$ were generated from $Unif(0, 1)$ and $Unif(-1, 0)$. Therefore, 16 coefficients were active and the remaining 4 or 24 regression coefficients were fixed to zero. Furthermore, we increased the number of simulation replication up to 200 to reduce sampling error. The rest of simulation settings was similar to the first simulation study.

Results

The MCMC chains satisfied convergence criterion. Most results in the 2PL IRT model were similar to the simulation study 1, so the results tables were shown in Online Supplementary Materials B. We only show the regression coefficients results here. Table 5 shows the summary of simulation result of the regression coefficients. The coefficient types were nonzero strong, nonzero weak, and zero. Biases of horseshoe and horseshoe+ priors were slightly smaller than the double-exponential or uniform priors, but the absolute values were generally small. That is, the four priors were not largely different in terms of bias. However, the biases results indicate that

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2
First Simulation Result of Discrimination Parameter of Two-Parameter Logistic Item Response Theory Model

# Item	Sample Size	# Covariate	Double Exponential			Horseshoe			Horseshoe+			Uniform		
			Bias	RMSE	95%CI Length	Bias	RMSE	95%CI Length	Bias	RMSE	95%CI Length	Bias	RMSE	95%CI Length
10	100	20	–.420	.303	1.392	–.302	.247	1.452	–.298	.245	1.454	–.464	.338	1.335
		40	–.544	.400	1.296	–.296	.258	1.444	–.291	.254	1.447	–.633	.466	1.209
	1,000	20	–.076	.049	.622	–.049	.046	.631	–.049	.045	.631	–.078	.049	.621
		40	–.101	.049	.621	–.047	.045	.635	–.047	.045	.638	–.104	.050	.617
30	100	20	–.363	.247	1.296	–.302	.217	1.322	–.300	.217	1.322	–.364	.250	1.279
		40	–.410	.268	1.266	–.293	.213	1.325	–.290	.211	1.327	–.421	.276	1.241
	1,000	20	–.093	.040	.550	–.077	.038	.553	–.077	.038	.553	–.093	.040	.548
		40	–.124	.053	.534	–.094	.047	.541	–.093	.047	.541	–.125	.054	.531

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3
First Simulation Result of Difficulty Parameter of Two Parameter Logistic Item Response Theory Model

# Item	Sample Size	# Covariate	Double-Exponential			Horseshoe			Horseshoe+			Uniform		
			Bias	RMSE	95%CI Length	Bias	RMSE	95%CI Length	Bias	RMSE	95%CI Length	Bias	RMSE	95%CI Length
10	100	20	.012	.165	1.046	.006	.120	.998	.007	.117	.996	.018	.241	1.100
		40	–.017	.308	1.168	–.017	.125	1.018	–.017	.121	1.015	–.016	.504	1.275
	1,000	20	–.002	.017	.315	–.002	.016	.312	–.002	.015	.312	–.002	.019	.316
		40	–.001	.021	.320	–.001	.015	.312	–.002	.015	.313	–.001	.024	.322
30	100	20	–.023	.135	.985	–.020	.122	.966	–.019	.121	.964	–.024	.160	1.001
		40	–.007	.187	1.027	–.003	.131	.971	–.004	.129	.968	–.012	.232	1.054
	1,000	20	.002	.018	.297	.002	.017	.296	.002	.016	.296	.002	.019	.298
		40	.000	.022	.302	.000	.018	.298	.000	.018	.297	.001	.024	.302

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 4
First Simulation Result of Correlation between Estimated and True Theta

# Item	Sample Size	# Covariate	Double-exponential		Horseshoe		Horseshoe+		Uniform	
			Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)
10	100	20	.926	(.015)	.927	(.015)	.927	(.015)	.922	(.016)
		40	.921	(.019)	.928	(.017)	.928	(.017)	.913	(.021)
	1,000	20	.930	(.011)	.930	(.011)	.930	(.011)	.930	(.011)
		40	.930	(.008)	.931	(.008)	.931	(.008)	.930	(.008)
30	100	20	.973	(.005)	.973	(.005)	.973	(.005)	.972	(.005)
		40	.971	(.005)	.972	(.005)	.972	(.005)	.969	(.005)
	1,000	20	.973	(.003)	.973	(.003)	.973	(.003)	.973	(.003)
		40	.973	(.003)	.973	(.003)	.973	(.003)	.973	(.003)

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 5
Second Simulation Result of Regression Coefficients

# Item	Sample Size	# Covariate	Coefficient Type	Double-Exponential			Horseshoe			Horseshoe+			Uniform		
				Bias	RMSE	95%CI Length	Bias	RMSE	95%CI Length	Bias	RMSE	95%CI Length	Bias	RMSE	95%CI Length
10	100	20	Nonzero strong	.010	2.130	4.057	.006	1.878	3.898	.006	2.007	3.858	.012	3.418	4.443
			Nonzero weak	.005	1.667	3.642	.003	.906	2.939	.002	.881	2.703	.008	2.692	4.364
			Zero	–.028	1.486	3.592	–.018	.665	2.858	–.017	.617	2.607	–.040	2.420	4.360
		40	Nonzero strong	.097	4.095	4.861	.059	2.046	3.991	.055	2.137	3.924	.106	8.416	5.635
			Nonzero weak	.017	3.055	4.455	–.006	.691	2.714	–.008	.662	2.471	.045	5.926	5.527
			Zero	–.034	2.698	4.403	–.014	.611	2.603	–.012	.593	2.341	–.048	5.150	5.525
	1,000	20	Nonzero strong	.009	.151	1.160	.009	.149	1.170	.008	.152	1.173	.009	.163	1.167
			Nonzero weak	–.001	.123	1.107	–.001	.116	1.058	–.002	.123	1.025	.000	.130	1.142
			Zero	–.007	.105	1.075	–.006	.071	.952	–.005	.060	.866	–.008	.117	1.139
		40	Nonzero strong	.003	.170	1.193	.002	.146	1.198	.002	.147	1.195	.003	.185	1.200
			Nonzero weak	.000	.129	1.145	–.002	.113	1.026	–.002	.125	.993	.001	.138	1.174
			Zero	–.002	.122	1.118	–.001	.056	.864	.000	.050	.777	–.002	.132	1.172
30	100	20	Nonzero strong	–.013	1.636	3.478	–.008	1.528	3.481	–.006	1.623	3.456	–.015	2.177	3.729

(*Continued*)

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 5
(Continued)

# Item	Sample Size	# Covariate	Coefficient Type	Double-Exponential			Horseshoe			Horseshoe+			Uniform		
				Bias	RMSE	95%CI Length	Bias	RMSE	95%CI Length	Bias	RMSE	95%CI Length	Bias	RMSE	95%CI Length
			Nonzero weak	.009	1.369	3.168	.010	.891	2.706	.008	.877	2.504	.008	1.900	3.710
			Zero	.019	1.344	3.115	.005	.784	2.622	.004	.750	2.397	.032	1.846	3.696
		40	Nonzero strong	–.059	2.676	4.016	–.041	1.913	3.621	–.039	2.003	3.559	–.065	3.904	4.424
			Nonzero weak	.066	2.180	3.716	.039	.870	2.559	.034	.858	2.346	.087	3.115	4.384
			Zero	.002	1.864	3.689	.003	.579	2.444	.004	.553	2.206	–.001	2.804	4.397
	1,000	20	Nonzero strong	.000	.125	1.040	.001	.119	1.051	.002	.121	1.051	.000	.133	1.045
			Nonzero weak	–.010	.108	1.000	–.010	.100	.972	–.010	.105	.947	–.010	.113	1.030
			Zero	.015	.105	.973	.012	.076	.877	.011	.066	.802	.016	.114	1.029
		40	Nonzero strong	.005	.147	1.065	.004	.131	1.073	.004	.132	1.070	.005	.156	1.069
			Nonzero weak	–.012	.113	1.029	–.009	.102	.945	–.009	.111	.917	–.012	.118	1.055
			Zero	.002	.108	1.006	.002	.055	.796	.002	.049	.718	.002	.115	1.054

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

horseshoe and horseshoe+ priors can correctly recover small magnitude regression coefficients.

The general tendency of RMSE and 95% CI lengths were similar to the first simulation study. The horseshoe and horseshoe+ priors indicated smaller RMSES for nonzero weak covariates or zero covariates than the double-exponential or uniform priors especially in larger number of covariates conditions. For example, the RMSEs of nonzero weak coefficients of the horseshoe, horseshoe+ prior, double-exponential, and uniform priors in 10 items, 100 sample size, and 40 covariates condition were 0.691, 0.662, 3.055, and 5.926, respectively. In this condition, the horseshoe and horseshoe+ priors displayed three to five times better stability than the other prior settings. Furthermore, we found that the horseshoe and horseshoe+ priors worked better in high dimensionality than in relatively lower dimensionality condition. For example, in the 40 covariate conditions, the RMSE of the nonzero weak and zero coefficients with the horseshoe and horseshoe+ priors was generally smaller than the case of 20 covariates. However, the corresponding RMSE of the double-exponential and uniform priors in the 40 covariates conditions was larger than the 20 covariates conditions.

Regarding the uncertainty of Bayesian estimates, the horseshoe and horseshoe+ priors performed better than the other priors. For example, the 95% CI length of the nonzero weak and zero coefficients with the horseshoe and horseshoe+ priors in 100 sample conditions was shorter than the double-exponential and uniform priors. In a lager sample size setting, the RMSE difference among four priors became trivial. In addition, the estimates of nonzero strong coefficients among the horseshoe, horseshoe+, and double-exponential priors were similar. However, the RMSE of nonzero strong coefficients with the three shrinkage priors were smaller than the uniform prior case, indicating that the uniform prior has worst estimation accuracy among all shrinkage methods. Therefore, the horseshoe and horseshoe+ priors can recover regression coefficients, and their RMSE and shorter 95% CI length for the nonzero and zero coefficients are smaller and shorter than the other priors especially in a small sample size setting.

Application to PISA Data

In this section, we demonstrated an application to real data obtained from the PISA 2018 math assessment. The goal was to examine whether Bayesian latent regression modeling with shrinkage prior approaches can select the important predictors of students' mathematics achievement. The construct of mathematical literacy used in this study was intended to describe the capacities of individuals to reason mathematically and use mathematical concepts, procedures, facts, and tools to describe, explain, and predict phenomena. The real data analysis R syntax is available on the Open Science Framework (OSF) web page: https://osf.io/u35z8.

Data Analysis Setting

Since participating students took different mathematics clusters and student questionnaires in the PISA 2018, only the data from a completed cluster of mathematics measures and 19 questions from students' questionnaires answered by students in

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the United States were taken into consideration. After omitting missing values, 422 students with an average age of 15.8 were kept for further analysis. The sample consisted of 206 males (48.8%) and 216 females (51.2%). The data used in this study included nine mathematical items: (I1-I9), and 19 covariates from the student questionnaire. The descriptive statistics of the measurement items and predictive variables are presented in Table 6. It should be noted that Table 6 only shows the raw scales of the variables of the samples. Continuous predictive covariates were standardized before conducting the analysis.

As Table 6 shows, the measurement items were dichotomously scored with 1 = correct and 0 = incorrect, including four difficult items (I1-I4), one moderately difficult item (I5), and four easy items (I6-I9). As for predictive covariates, apart from SEX (0 = female, 1 = male) and REPEAT (0 = no grade repetition, 1 = have grade repetition) as binary variables, other covariates were normalized continuous variables. All continuous covariates were standardized with a mean of zero and a standardized deviation of one for the proposed methods.

The proposed Bayesian variable selection methods with double-exponential, horseshoe, horseshoe+, and uniform prior methods were fitted to the samples using R ver. 4.2 (R Core Team, 2021). Gibbs sampling algorithms were used with four MCMC chains. Each MCMC chain had 10,000 iterations, with the first 5,000 iterations considered as burn-ins. Credible interval-based variable selection criterion was employed in the real data analysis setting, which was proposed in previous research (e.g., Li and Lin 2010, p. 157; van Erp et al., 2019). Li and Lin (2010) mentioned ad hoc treatment for variable selection in Bayesian approaches and suggested the credible interval criterion. The credible interval criterion assumes that a covariate is excluded from the set of covariates if its 95% credible interval covers zero and retained otherwise.

Results

The results showed that all four algorithms converged according to the Gelman-Rubin index ($\hat{R} \leq 1.10$). Table 7 reports the point estimates and 95% CI of the regression coefficients. It was clear that the model with the double-exponential priors provided results with similar magnitude and 95% credit interval (CI) length as the model with the uniform prior. Models with horseshoe and horseshoe+ priors had coefficients with smaller magnitudes and shorter 95% CI lengths in all predictive covariates. Eleven covariates—parents' emotional support, teacher instruction, home education resources, family wealth, students' attitudes, students' sense of belonging to a school, instructional time per week, grade repetition, perception of cooperation at school, subjective well-being, and mastery of goal orientation—negatively affected the latent mathematical proficiency of students. The REPEAT (grade repetition) covariate had the largest negative effect on estimated math proficiency. Conversely, eight covariates—gender, home possessions, parental educational level, the learning time in mathematics, adaptation of instruction, perceived teacher's interest, perception of competitiveness at school, and resilience—had positive coefficients on outcomes in which home possessions had the highest coefficient predicting math ability.

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 6
Descriptive Statistics for Measurement Items and Predictive Covariates

Variables	Contents	Min	Max	Mean	*SD*
Measurement Items					
I1	A View Room—Q01	0	1	.84	.13
I2	Running Time—Q01	0	1	.63	.23
I3	Population Pyramids—Q01	0	1	.77	.18
I4	Population Pyramids—Q04	0	1	.61	.24
I5	Diving—Q01	0	1	.56	.25
I6	Diving—Q02	0	1	.47	.25
I7	Labels—Q01	0	1	.33	.22
I8	Braille—Q02	0	1	.37	.23
I9	Bricks—Q01	0	1	.37	.23
Predictive Covariates					
GENDER	Student (standardized) gender	0	1	.49	.25
EMOSUPS	Parents' emotional support perceived by student	–2.45	1.03	.05	1.01
DIRINS	Teacher-directed instruction	–2.94	1.82	.11	.97
HEDRES	Home educational resources	–4.41	1.18	.03	1.07

(*Continued*)

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 6
(Continued)

Variables	Contents	Min	Max	Mean	*SD*
WEALTH	Family wealth	–2.59	4.26	.56	.95
ATTLNACT	Attitude toward school: learning activities	–2.54	1.08	.29	.98
BELONG	Sense of belonging to school	–3.24	2.76	–.27	1
HOMEPOS	Home possessions	–2.96	3.41	.15	1.04
HISCED	Highest educational level of parents	–4.2	.77	0	1
TMINS	Total minutes of instructional time per week	–4.13	2.53	0	1
MMINS	Learning time in mathematics	–3.1	2.97	0	1
REPEAT	Grade repetition	0	1	.06	.06
ADAPTIVITY	Adaptation of instruction	–2.27	2.01	.13	.91
TEACHINT	Perceived teacher's interest	–2.22	1.82	.29	.97
PERCOMP	Perception of competitiveness at school	–1.99	2.04	.4	.97
PERCOOP	Perception of cooperation at school	–2.14	1.68	–.17	.94
SWBP	Subjective well-being: positive affect	–3.07	1.24	–.17	1.03
RESILIENCE	Resilience	–3.17	2.37	.21	.96
MASTGOAL	Mastery goal orientation	–2.53	1.85	.25	.98

Note: The table only shows the raw scales of the predictive covariates of the samples. The continuous predictive covariates were all standardized before data analysis.

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 7

Estimated Coefficients for and 95% Credible Interval for Bayesian Latent Regression with the Double Exponential, Horseshoe, Horseshoe+, and Uniform Priors

Index	Variables	Label	Double-Exponential			Horseshoe			Horseshoe+			Uniform		
			Estimates	(*SD*)	95% CI [Lower, Upper]	Estimates	(*SD*)	95% CI [Lower, Upper]	Estimates	(*SD*)	95% CI [Lower, Upper]	Estimates	(*SD*)	95% CI [Lower, Upper]
1	GENDER	Student (standardized) gender	**.245**	**(.128)**	**[.001, .502]**	.073	(.098)	[−.060, .308]	.062	(.091)	[−.050, .299]	**.249**	**(.125)**	**[.007, .495]**
2	EMOSUPS	Parents' emotional support perceived by student	–.044	(.072)	[−.187, .092]	–.030	(.051)	[−.151, .053]	–.022	(.045)	[−.143, .046]	–.049	(.072)	[−.189, .093]
3	DIRINS	Teacher-directed instruction	**–.234**	**(.078)**	**[−.388,−.083]**	–.152	(.082)	[−.312, .001]	–.145	(.087)	[−.310, .003]	**–.237**	**(.078)**	**[−.390,−.086]**
4	HEDRES	Home educational resources	**–.211**	**(.094)**	**[−.392,−.026]**	–.073	(.084)	[−.266, .043]	–.068	(.085)	[−.273, .034]	**–.218**	**(.094)**	**[−.404,−.033]**
5	WEALTH	Family wealth	**–.469**	**(.122)**	**[−.713,−.235]**	–.287	(.140)	[−.555, .000]	–.292	(.143)	[−.561, .000]	**–.476**	**(.121)**	**[−.714,−.241]**
6	ATTLNACT	Attitude toward school: learning activities	–.038	(.069)	[−.174, .096]	–.024	(.047)	[−.136, .060]	–.017	(.041)	[−.124, .050]	–.042	(.072)	[−.185, .096]
7	BELONG	Sense of belonging to school	–.057	(.077)	[−.208, .095]	–.014	(.048)	[−.129, .077]	–.010	(.040)	[−.116, .069]	–.056	(.080)	[−.212, .099]
8	HOMEPOS	Home possessions	**.800**	**(.151)**	**[.511, 1.106]**	**.562**	**(.162)**	**[.240, .883]**	**.569**	**(.164)**	**[.253, .898]**	**.814**	**(.151)**	**[.516, 1.109]**
9	HISCED	Highest educational level of parents	.106	(.069)	[−.029, .244]	.061	(.062)	[−.030, .199]	.051	(.061)	[−.028, .197]	.110	(.070)	[−.025, .246]

(Continued)

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 7
(Continued)

Index	Variables	Label	Double-Exponential			Horseshoe			Horseshoe+			Uniform		
			Estimates	(*SD*)	95% CI [Lower, Upper]	Estimates	(*SD*)	95% CI [Lower, Upper]	Estimates	(*SD*)	95% CI [Lower, Upper]	Estimates	(*SD*)	95% CI [Lower, Upper]
10	TMINS	Total minutes of instructional time per week	–.104	(.070)	[−.240, .032]	–.034	(.052)	[−.159, .048]	–.027	(.047)	[−.147, .041]	–.109	(.070)	[−.248, .028]
11	MMINS	Learning time in mathematics	**.270**	**(.070)**	**[.134, .412]**	**.201**	**(.073)**	**[.051, .339]**	**.197**	**(.074)**	**[.029, .339]**	**.274**	**(.070)**	**[.138, .413]**
12	REPEAT	Grade repetition	**–.737**	**(.300)**	**[−1.334,−.155]**	–.365	(.328)	[−1.083, .051]	–.384	(.345)	[−1.099, .039]	**–.761**	**(.302)**	**[−1.353,−.170]**
13	ADAPTIVITY	Adaptation of instruction	.108	(.079)	[−.047, .264]	.043	(.061)	[−.049, .185]	.036	(.057)	[−.042, .186]	.110	(.080)	[−.046, .267]
14	TEACHINT	Perceived teacher's interest	.118	(.076)	[−.027, .270]	.067	(.068)	[−.033, .217]	.056	(.067)	[−.025, .215]	.119	(.076)	[−.030, .270]
15	PERCOMP	Perception of competitiveness at school	.063	(.067)	[−.067, .196]	.043	(.054)	[−.043, .170]	.034	(.052)	[−.034, .163]	.062	(.069)	[−.074, .198]
16	PERCOOP	Perception of cooperation at school	–.021	(.072)	[−.163, .119]	.000	(.045)	[−.096, .094]	.001	(.037)	[−.085, .087]	–.022	(.072)	[−.164, .118]
17	SWBP	Subjective well-being: Positive affect	–.039	(.073)	[−.186, .103]	–.028	(.051)	[−.149, .058]	–.020	(.045)	[−.137, .047]	–.043	(.075)	[−.190, .105]
18	RESILIENCE	Resilience	.160	(.082)	[−.001, .323]	.079	(.072)	[−.026, .235]	.064	(.069)	[−.023, .227]	**.167**	**(.084)**	**[.005, .331]**
19	MASTGOAL	Mastery goal orientation	–.020	(.075)	[−.169, .127]	–.009	(.046)	[−.116, .084]	–.006	(.039)	[−.099, .070]	–.022	(.077)	[−.174, .130]

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

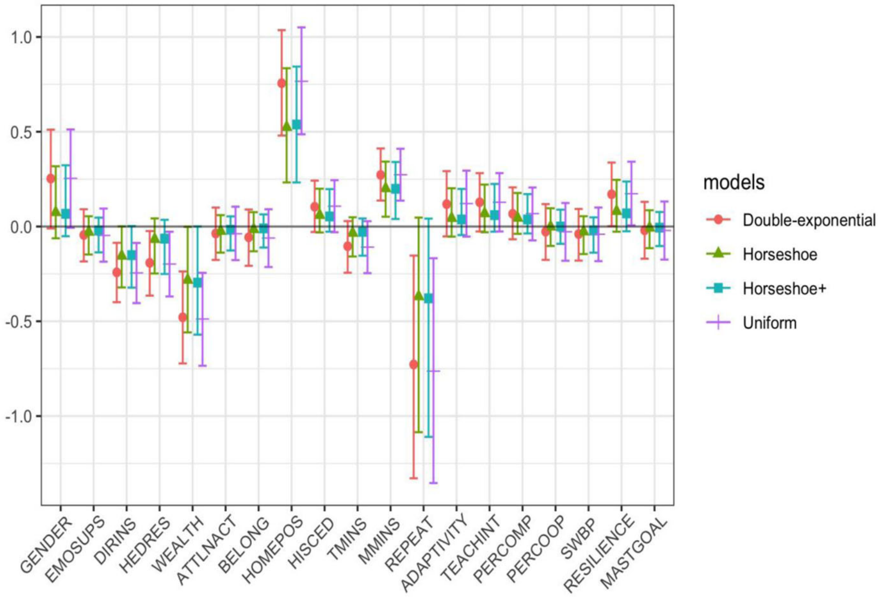

Figure 1. Parameter estimates and 95% credible intervals of covariates in the PISA 2018 math assessment data with four Gibbs sampling results. *Note*: GENDER: Student (Standardized) Gender, EMOSUPS: Parents' emotional support perceived by student, DIRINS: Teacher-directed instruction, HEDRES: Home educational resources, WEALTH: Family wealth, ATTLNACT: Attitude toward school: Learning activities, BELONG: Sense of belonging to school, HOMEPOS: Home possessions, HISCED: Highest educational level of parents, TMINS: Total minutes of instructional time per week, MMINS: Learning time in mathematics, REPEAT: Grade repetition, ADAPTIVITY: Adaptation of instruction, TEACHINT: Perceived teacher's interest, PERCOMP: Perception of competitiveness at school, PERCOOP: Perception of cooperation at school, SWBP: Subjective well-being: Positive affect, RESILIENCE: Resilience, and MASTGOAL: Mastery goal orientation. [Color figure can be viewed at wileyonlinelibrary.com]

Figure 1 shows the point estimates and 95% CIs for all approaches. Compared to the horseshoe and horseshoe+ approaches, double-exponential prior and uniform prior approaches had more nonzero covariates ($p = 7$). The horseshoe+ approach had the most parsimonious feature set among the four approaches, which has only two nonzero covariates, while the horseshoe approach selected three features as nonzero covariates. Home possession (HOMEPOS) as one of surrogate variable of social-economic status (SES; Lee & Stankov [2018] considered parental education and home possessions as the two SES-related variables) and learning time in mathematics (MMINS) were the two covariates selected by all approaches, indicating that the SES status of students and learning time in math had an important impact on the improvement of mathematics capacities. These results are consistent with prior literature (e.g., Gamazo & Martínez-Abad, 2020; Kalaycıoğlu, 2015; Lee & Stankov, 2018), and the horseshoe+ priors identified fewer but stronger predictors to predict mathematical ability.

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Conclusion

In the present study, we proposed fully Gibbs sampling algorithms for Bayesian variable selection under a unidimensional IRT model latent regression model. The shrinkage priors that were double-exponential, horseshoe, and horseshoe+ were compared to the uniform prior case in both the simulation and real data example. The two simulation studies revealed that the horseshoe priors had smaller RMSEs and shorter 95% CI length of regression coefficients than double-exponential or uniform priors. The prior setting difference was drastic when there were fewer samples, lesser measurement items, and several covariates. Moreover, the stability results of the estimation were shown not only in regression coefficients but also in item discrimination parameters. The horseshoe+ prior had slightly better scores than the usual horseshoe. From the simulation study, we recommend using horseshoe+ prior for Bayesian variable selection, even in the latent regression context.

The real data example indicates the utility of Bayesian variable selection with horseshoe priors. Both horseshoe and horseshoe+ priors selected only home possessions (HOMEPOS) and learning time in mathematics (MMINS) as the most predictive variables for students' mathematical proficiency. These findings are consistent with prior studies that SES and self-efficiency are the most important predictors of math achievement. It should be noted that in previous studies, multiple factors (such as home possession and parental education) are considered as SES-related variables, while our approach selects only one most important factor. This may help in big data mining for large-scale assessments when multiple predictors exist.

One limitation of the proposed Gibbs sampling algorithm is that it can be applied to binary-valued item responses without guessing parameters. For future studies, a more flexible algorithm is needed to analyze nominal, ordinal, or partial credit type item responses. Another limitation of this study is that we assumed only one latent trait. In many applications, we assumed unidimensional latent proficiency, so the utility of the proposed algorithms would not be compromised. However, treating multiple latent proficiencies as Culpepper and Park (2017) did may be required for more flexible real data analysis. Bayesian shrinkage priors, especially horseshoe priors, have great flexibility and can easily be expanded to various psychometric models such as multidimensional IRT or diagnostic classification models. This has great implications for future studies.

Our method has several benefits for operational practice in educational measurement as follows. First, when we select several background characteristics of individual such as individual ethics, we can assess DIF of those group variables on the assessment from the estimation, which means that a distinct DIF study is not required. Second, we can efficiently omit noise in large-scale assessment setting with the proposed approach, which benefits data analytics. To be specific, it may be difficult to judge whether the estimated coefficient is large in a large sample size setting because null hypothesis $\mathrm{H}_0 : \beta = 0$ tends to be rejected even if the estimated coefficient is a smaller value. However, shrinkage prior can automatically shrink the coefficient to zero if the value is close to zero. Third, horseshoe priors can provide stable item parameter estimates even when sample size is relatively small, and this feature helps to reduce the number of sample when calibrating item parameters. Item

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

parameter calibration needs large samples but including coefficients can help to reduce the number.

One limitation of proposed method is that the latent trait is unidimensional. In the future study, the method needs to be extended to be suitable for multidimensional latent traits similar to Culpepper and Park (2017). However, unidimensional IRT model is sometimes preferred in real application field (e.g., OECD, 2017, chapter 9) and the prior setting is theoretically appropriate than double-exponential prior that has been employed in many studies. Therefore, the proposed method provided a new way to conduct sparsity induce analysis in IRT model setting.

Focusing on spike-and-slab priors is a future research direction. Spike-and-slab prior is famous in Bayesian variable selection, and variational Bayesian inference algorithm for spike-and-slab priors has already been developed (Ormerod et al., 2017; Ray & Szabó, 2021). Developing variational inference methods for horseshoe priors in latent regression and comparing them with spike-and-slab variational inference can extend the applicability of sparse estimation in latent regression models.

Acknowledgments

The data analysis code is available in the Open Science Framework page: https://osf.io/u35z8. The authors declare no conflicts of interest. This work was supported by JSPS KAKENHI 19H00616, 20H01720, 21H00936, and 22K13810.

References

Ames, A., & Smith, E. (2018). Subjective priors for item response models: Application of elicitation by design. *Journal of Educational Measurement*, *55*(3), 373–402. https://doi.org/10.1111/jedm.12184

Andrews, D. F., & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, *36*(1), 99–102. http://www.jstor.org/stable/2984774

Armagan, A., Dunson, D. B., & Lee, J. (2013). Generalized double Pareto shrinkage. *Statistica Sinica*, *23*(1), 1–7. https://doi.org/10.5705/ss.2011.048

Bhadra, A., Datta, J., Polson, N. G., & Willard, B. (2017). The horseshoe+ estimator of ultra-sparse signals. *Bayesian Analysis*, *12*(4), 1105–1131. https://doi.org/10.1214/16-BA1028

Bhadra, A., Datta, J., Polson, N. G., & Willard, B. (2019). Lasso meets horseshoe: A survey. *Statistical Science*, *34*(3), 405–427. https://doi.org/10.1214/19-STS700

Bishop, M. (2006). *Pattern recognition and machine learning*. Springer. https://doi.org/10.1641/B580519

Brooks, S., Gelman, A., Jones, G. L., & Meng, X.-L. (2011). *Handbook of Markov chain Monte Carlo*. CRC Press.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434–455. https://doi.org/10.2307/1390675

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2009). Handling sparsity via the horseshoe. *Journal of Machine Learning Research*, *5*, 73–80.

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, *97*(2), 465–480. https://doi.org/10.1093/biomet/asq017

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Culpepper, S. A., & Park, T. (2017). Bayesian estimation of multivariate latent regression models: Gauss versus Laplace. *Journal of Educational and Behavioral Statistics*, *42*(5), 591–616. https://doi.org/10.3102/1076998617700598

De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. Springer.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologist*. Lawrence Erlbaum Associates Publishers.

Feng, X. N., Wu, H. T., & Song, X. Y. (2017). Bayesian regularized multivariate generalized latent variable models. *Structural Equation Modeling*, *24*(3), 341–358. https://doi.org/10.1080/10705511.2016.1257353

Gamazo, A., & Martínez-Abad, F. (2020). An exploration of factors linked to academic performance in PISA 2018 through data mining techniques. *Frontiers in Psychology*, *11*(November), 1–17. https://doi.org/10.3389/fpsyg.2020.575167

George, E. I. (2000). The variable selection problem. *Journal of the American Statistical Association*, *95*(452), 1304–1308. https://doi.org/10.2307/2669776

Giner, G., & Smyth, G. K. (2016). statmod: Probability calculations for the inverse Gaussian distribution. *The R Journal*, *8*(1), 339–351. https://doi.org/10.32614/rj-2016-024

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, *6*(4), 430–450. https://doi.org/10.1037/1082-989X.6.4.430

Griffin, J. E., & Brown, P. J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, *5*(1), 171–188. https://doi.org/10.1214/10-BA507

Helwig, N. E. (2017). Adding bias to reduce variance in psychological results: A tutorial on penalized regression. *The Quantitative Methods for Psychology*, *13*(1), 1–19. https://doi.org/10.20982/tqmp.13.1.p001

Ishwaran, H., & Rao, J. S. (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *Annals of Statistics*, *33*(2), 730–773. https://doi.org/10.1214/009053604000001147

Jacobucci, R., Brandmaier, A. M., & Kievit, R. A. (2019). A practical guide to variable selection in structural equation modeling by using regularized multiple-indicators, multiple-causes models. *Advances in Methods and Practices in Psychological Science*, *2*(1), 55–76. https://doi.org/10.1177/2515245919826527

Jiang, Z., & Templin, J. (2019). Gibbs samplers for logistic item response models via the Pólya–Gamma distribution: A computationally efficient data-augmentation strategy. *Psychometrika*, *84*(2), 358–374. https://doi.org/10.1007/s11336-018-9641-x

Kalaycıoğlu, D. B. (2015). The influence of socioeconomic status, self-efficacy, and anxiety on mathematics achievement in England, Greece, Hong Kong, the Netherlands, Turkey, and the USA. *Kuram ve Uygulamada Egitim Bilimleri*, *15*(5), 1391–1401. https://doi.org/10.12738/estp.2015.5.2731

Lee, S. (2020). Logistic regression procedure using penalized maximum likelihood estimation for differential item functioning. *Journal of Educational Measurement*, *57*(3), 443–457. https://doi.org/10.1111/jedm.12253

Lee, J., & Stankov, L. (2018). Non-cognitive predictors of academic achievement: Evidence from TIMSS and PISA. *Learning and Individual Differences*, *65*, 50–64. https://doi.org/10.1016/j.lindif.2018.05.009

Li, Q., & Lin, N. (2010). The Bayesian elastic net. *Bayesian Analysis*, *5*(1), 151–170. https://doi.org/10.1214/10-BA506

Makalic, E., & Schmidt, D. F. (2016a). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, *23*(1), 179–182. https://doi.org/10.1109/LSP.2015.2503725

Makalic, E., & Schmidt, D. F. (2016b). *High-dimensional Bayesian regularized regression with the BayesReg package*. MCMC, pp. 1–18. http://arxiv.org/abs/1611.06649

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

McNeish, D. M. (2015). Using lasso for predictor selection and to assuage overfitting: A method long overlooked in behavioral sciences. *Multivariate Behavioral Research*, *50*(5), 471–484. https://doi.org/10.1080/00273171.2015.1036965

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Boston College, TIMSS & PIRLS International Study Center website. Retrieved from https://timssandpirls.bc.edu/timss2019/international-results/

OECD. (2017). PISA 2015 Assessment and analytical framework: Science, reading, mathematic, financial literacy and collaborative problem solving, *PISA*. Paris: OECD Publishing. https://doi.org/10.1787/9789264281820-en

OECD. (2019), *PISA 2018 Results (Volume I):* What students know and can do, PISA. Paris: OECD Publishing. https://doi.org/10.1787/5f07c754-en

Ormerod, J. T., You, C., & Müller, S. (2017). A variational Bayes approach to variable selection. *Electronic Journal of Statistics*, *11*(2), 3549–3594. https://doi.org/10.1214/17-EJS1332

Park, T., & Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, *103*(482), 681–686. https://doi.org/10.1198/016214508000000337

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, *6*, 7–11. https://journal.r-project.org/archive/

Polson, N. G., Scott, J. G., & Windle, J. (2013). Bayesian inference for logistic models using Pólya-Gamma latent variables. *Journal of the American Statistical Association*, *108*(504), 1339–1349. https://doi.org/10.1080/01621459.2013.829001

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Ray, K., & Szabó, B. (2021). Variational Bayes for high-dimensional linear regression with sparse priors. *Journal of the American Statistical Association*, *117*, 1–31. https://doi.org/10.1080/01621459.2020.1847121

Scott, J. G., & Sun, L. (2013). *Expectation-maximization for logistic regression*. *1*, 1–25. http://arxiv.org/abs/1306.0040

Tibshiranit, B. R. (1996). Shrinkage and Selection via the lasso. *Journal of the Royal Statistical Society Series B*, *58*(1), 267–288. http://www.jstor.org/stable/2346178

van Erp, S., Oberski, D. L., & Mulder, J. (2019). Shrinkage priors for Bayesian penalized regression. *Journal of Mathematical Psychology*, *89*, 31–50. https://doi.org/10.1016/j.jmp.2018.12.004

von Davier, M., & Sinharay, S. (2007). An importance sampling EM algorithm for latent regression models. *Journal of Educational and Behavioral Statistics*, *32*(3), 233–251. https://doi.org/10.3102/1076998607300422

von Davier, M., & Sinharay, S. (2010). Stochastic approximation methods for latent regression item response models. *Journal of Educational and Behavioral Statistics*, *35*(2), 174–193. https://doi.org/10.3102/1076998609346970

Zhang, X., Tao, J., Wang, C., & Shi, N. Z. (2019). Bayesian model selection methods for multilevel IRT models: A comparison of five DIC-based indices. *Journal of Educational Measurement*, *56*(1), 3–27. https://doi.org/10.1111/jedm.12197

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Authors

Kazuhiro Yamaguchi is assistant professor at University of Tsukuba, Institutes of Human Sciences A314, 1-1-1 Tennodai, Tsukuba-shi, Ibaraki-ken 305-0006, Japan; yam-

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

aguchi.kazuhir.ft@u.tsukuba.ac.jp. His primary research interests include Bayesian data analysis, latent variable modeling, diagnostic classification models, and psychometric methods.

Jihong Zhang is doctoral cause student at University of Iowa, 3750 Market Street, S210B Lindquist Center, Iowa City, IA 52242; jihong-zhang@uiowa.edu. His primary research interests include Bayesian analysis and psychometric models.

17453984, 2023, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12348 by University Of Arkansas Library, Wiley Online Library on [06/04/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License