



Bayesian Analysis of a Quantile Multilevel Item Response Theory Model

Hongyue Zhu¹, Wei Gao¹ and Xue Zhang^{2*}

¹ School of Mathematics and Statistics, Northeast Normal University, Changchun, China, ² China Institute of Rural Education Development, Northeast Normal University, Changchun, China

OPEN ACCESS

Edited by:

Alexander Robitzsch,
IPN - Leibniz Institute for Science
and Mathematics Education,
Germany

Reviewed by:

Matthias Trendtel,
Technical University of Dortmund,
Germany
Peida Zhan,
Zhejiang Normal University, China
Xiangnan Feng,
Fudan University, China

*Correspondence:

Xue Zhang
zhangx815@nenu.edu.cn

Specialty section:

This article was submitted to
Quantitative Psychology
and Measurement,
a section of the journal
Frontiers in Psychology

Received: 18 September 2020

Accepted: 07 December 2020

Published: 08 January 2021

Citation:

Zhu H, Gao W and Zhang X
(2021) Bayesian Analysis of a Quantile
Multilevel Item Response Theory
Model. *Front. Psychol.* 11:607731.
doi: 10.3389/fpsyg.2020.607731

Multilevel item response theory (MLIRT) models are used widely in educational and psychological research. This type of modeling has two or more levels, including an item response theory model as the measurement part and a linear-regression model as the structural part, the aim being to investigate the relation between explanatory variables and latent variables. However, the linear-regression structural model focuses on the relation between explanatory variables and latent variables, which is only from the perspective of the average tendency. When we need to explore the relationship between variables at various locations along the response distribution, quantile regression is more appropriate. To this end, a quantile-regression-type structural model named as the quantile MLIRT (Q-MLIRT) model is introduced under the MLIRT framework. The parameters of the proposed model are estimated using the Gibbs sampling algorithm, and comparison with the original (i.e., linear-regression-type) MLIRT model is conducted via a simulation study. The results show that the parameters of the Q-MLIRT model could be recovered well under different quantiles. Finally, a subset of data from PISA 2018 is analyzed to illustrate the application of the proposed model.

Keywords: multilevel item response theory, quantile regression, Bayesian analysis, Gibbs sampling, non-normality of latent variable

INTRODUCTION

Multilevel or hierarchical linear models are used widely in educational and psychological researches (e.g., Raudenbush, 1988; Goldstein, 1995; Snijders and Bosker, 1999). These models allow data to be collected at different levels; for example, test data are obtained from students, students are nested within schools, and so on. Item response theory (IRT) models can be plugged in multilevel models, known as multilevel IRT (MLIRT) models. In educational research and educational assessment, the MLIRT model is widely applied, where the main objective is to investigate the relationship between covariates (e.g., student's social background, school financial resources) and outcome variables (e.g., student's ability; Fox, 2010).

Adams et al. (1997) noted that a two-level IRT model could be seen as a multilevel perspective on item response modeling. In this formulation, we can divide the two-level IRT model in two components: the measurement part with the respective IRT measurement model and the structural part with the respective regression model. Furthermore, Kamata (2001) and Fox and Glas (2001) extended the two-level IRT model to three levels, with both studies proposing a two-level regression model (i.e., student-level and group-level covariates were analyzed in different levels) on the

ability parameters as the structural model. Because the explanatory variables might not always be measured accurately, Fox and Glas (2003) showed how to model latent explanatory variables with measurement errors within the MLIRT model. Additionally, MLIRT models can also be applied to handle more-complicated factors such as longitudinal response data and response times. For example, Huang (2015); Schmidt et al. (2016) introduced MLIRT models for longitudinal data to assess the changes in students' abilities over time. Also, when jointly modeling the responses and response times, the ability and speed parameters can be considered as outcome variables of a multivariate multilevel model for various analyses (e.g., Entink et al., 2009; Fox, 2010).

Regarding the estimation of MLIRT models, studies have shown that all the model parameters can be estimated simultaneously using a fully Bayesian method, regardless of whether the IRT model is a two-parameter model for dichotomous response data or a graded response model for polytomous response data (Fox, 2005; Natesan et al., 2010). Compared with the traditional two-stage estimation procedure, this estimation procedure in the MLIRT models leads to a proper treatment of the measurement error associated with the ability parameter (Adams et al., 1997; Fox and Glas, 2001, 2003; Fox, 2004, 2005, 2010). See Fox (2010) for more advantages of the MLIRT modeling framework.

One fact that needs to be noted is that the distribution of latent variables in the IRT model is often assumed to be normal (e.g., Bock and Lieberman, 1970; Bock and Aitkin, 1981; Wirth and Edwards, 2007). This assumption may be reasonable when students follow a normally distributed population, but in many cases, it is not satisfied (e.g., Woods, 2006, 2015; Woods and Thissen, 2006; Sass et al., 2008). It has been shown that when a non-normal latent distribution is assumed to be normal, estimates of IRT model parameters are biased (e.g., Stone, 1992; Sass et al., 2008; Woods, 2015). This issue has also been explored regarding multidimensional and graded response models (e.g., Finch, 2010, 2011; Svetina et al., 2017; Wang et al., 2018). However, there has been little attention to the non-normality of latent variables in the MLIRT framework.

On the other hand, the structural models in the MLIRT models are always linear regression (LR) approaches. Herein, LR generally assumes that the errors are normally distributed and should be homoscedastic, which are not often in accordance with practice, because of non-normality of latent variables or heterogeneity of errors (Harvey, 1976; Koenker and Bassett, 1978, 1982; Jarque and Bera, 1980; Tsionas, 2003). It is also well known that a LR model mainly focuses on the relation between explanatory variables and the conditional mean of latent variables, however, it is quite possible that the relationship between variables may vary at different points along the response distribution, researchers also might be interested in the relationships between variables on various locations of the distribution, such as on one or both of the tails of the distribution.

These problems can be addressed by the quantile regression (QR) (Koenker and Bassett, 1978), which is a valuable and robust tool for analyzing the conditional quantile functions of latent variables. When considering the relationship between

explanatory variables and latent variables, QR allows to compute several different regression curves corresponding to the various percentage points of the distribution and thus obtain a more complete relationship between them. In addition, QR is regarded as a robust method in regression because the estimation results are insensitive to outliers or the non-normality of response distributions. Overview of QR methods, one can refer to Yu et al. (2003); Koenker and Ng (2005). As such, QR has been used by many researchers in the field of education (e.g., Wößmann, 2005; Haile and Nguyen, 2008; Chen and Chalhoub-Deville, 2014; Giambona and Porcu, 2015; Gürsakal et al., 2016; Costanzo and Desimoni, 2017). The aim of most studies has been to scrutinize inequality of opportunity in education through investigating the determinants of student achievement, or to analyze individual and/or family background determinants of student achievement with respect to test results.

In recent years, QR has also been applied successfully in latent variable models (e.g., Dunson et al., 2003; Allen and Powell, 2011; Burgette and Reiter, 2012; Wang et al., 2016; Feng et al., 2017; Belloni et al., 2019; Liu et al., 2020). Burgette and Reiter (2012) considered QR in a factor analysis model to analyze the effects of latent variables on the lower quantiles of the response distribution. Wang et al. (2016) introduced QR into a structural equation model (SEM) to assess the conditional quantile of the outcome latent variable given the explanatory latent variables and covariates. Liu et al. (2020) proposed a quantile hidden Markov model to examine the entire conditional distribution of the response given the hidden state and potential covariates. Factor analysis models can be converted to IRT models under possible parameterizations (Kamata and Bauer, 2008; Wang and Zhang, 2019). Therefore, it is reasonable and implementable to apply QR to MLIRT models.

This study is motivated by the MLIRT models focus largely on the relationship between variables at the mean level, which may result in neglecting potential differences across the response distribution. Herein, taking a two-level (i.e., item level and student level) IRT model as a demonstration, we embed the QR approach in a MLIRT model to obtain what we refer to as the quantile MLIRT (Q-MLIRT) model. This approach enables us to investigate the relationship between explanatory variables and latent variables at various locations of the response distribution, as well as having minimal assumptions on error terms, which is a flexible and applicable approach in general. Moreover, the Q-MLIRT model can be estimated in a Bayesian framework with a Markov chain Monte Carlo (MCMC) algorithm. It should be emphasized that the Q-MLIRT model is not a replacement but a supplement to the conventional MLIRT model. If the relationship between explanatory variables and latent variables on the entire distribution is of interest, Q-MLIRT is a good recommendation.

The rest of this paper is organized as follows. The proposed Q-MLIRT model is introduced in Section "Model Description," the MCMC algorithm is presented in Section "Bayesian Estimation," and a simulation study conducted to evaluate the empirical performance of the proposed method is reported in Section "Simulation Study." A real data study to demonstrate the use of Q-MLIRT models is reported in Section "Empirical

Example,” and finally a discussion and suggestions for further research are presented in Section “Discussion.”

MODEL DESCRIPTION

Before describing the Q-MLIRT model considered here, we revisit MLIRT models briefly. We consider only a two-level IRT model, namely an item level and a student level. As presented by Adams et al. (1997), the two-level IRT model consists of two components: the measurement part—an IRT measurement model—which describes the probability of observed responses conditional upon the latent variable, and the structural part—a LR model—which describes the between-student variation in the latent variable. The LR model is on ability parameters, where the ability parameter is dependent on the covariates:

$$\theta_i = \mathbf{X}_i^T \boldsymbol{\beta} + \delta_i, \quad (1)$$

where θ_i is the latent trait (i.e., ability) of student i ($i = 1, \dots, n$), \mathbf{X}_i is a $\mathbf{q} \times 1$ vector of known covariates (e.g., gender, socioeconomic status, and major) for student i , and $\boldsymbol{\beta}$ is the corresponding $\mathbf{q} \times 1$ vector of regression coefficients. We assume that δ_i are independently and identically normally distributed with mean zero and variance α^2 . In this form, a student's achievement or performance is represented as ability parameter θ_i , and the relationship between students' abilities and the involved explanatory variables is reflected in the regression coefficients $\boldsymbol{\beta}$. Note that the coefficients of all covariates are treated as fixed effects here.

LR analysis measures the relationship between explanatory variables and students' abilities by modeling a conditional mean function of θ_i . Given explanatory variables \mathbf{X}_i , then:

$$E(\theta_i | \mathbf{X}_i) = \mathbf{X}_i^T \boldsymbol{\beta}. \quad (2)$$

That is, the conditional distribution of θ_i is assumed to be a linear combination of covariates \mathbf{X}_i with normally distributed errors. In usual LR approaches, the data are used to find out a single regression line that minimizes the sum of squared errors (least squared estimation) to estimate the relationship. Consequently, the focus is on average performance of response variables about covariates. However, as mentioned above, the errors may be non-normally distributed or even heteroscedastic in reality, and when a distribution is asymmetric (e.g., heavy-tailed or skewed), the mean is not the center of the distribution, while the median is likely to be a more appropriate measure of central tendency than the mean (Edgeworth, 1888; Fox, 1997; Koenker and Ng, 2005; Hao and Naiman, 2007; Trafimow et al., 2018). What is more, the relation between explanatory variables and latent variables may be different across the entire distribution. As a result, the conditional mean fails to meet the research needs in many cases. Therefore, we must provide an elaborate or specific description of the inter-relationship among latent variables and explanatory variables.

A more appropriate analysis can be achieved from conditional quantiles of θ_i . QR provides a suitable method for modeling conditional quantile functions (Koenker and Hallock, 2001).

Under a number of different quantiles $\tau \in (0, 1)$, we say the τ th quantile of δ_i is the value q_τ , for which $p(\delta_i < q_\tau) = \tau$, that is $Q_\tau(\delta_i) = q_\tau$. The τ th conditional quantile of θ_i can be then expressed as:

$$Q_\tau(\theta_i | \mathbf{X}_i) = \mathbf{X}_i^T \boldsymbol{\beta}_\tau, \quad (3)$$

where $\boldsymbol{\beta}_\tau$ is a $\mathbf{q} \times 1$ vector of QR coefficients depend on τ , and it reflects the relationship between \mathbf{X}_i and the τ th conditional quantile of θ_i , which means the marginal effects of the explanatory variables may differ over quantiles of the distribution of θ_i . Since numerous quantiles (e.g., 25%, 50%, 75%) are modeled in QR, it is possible to understand the relationship between variables at various locations of the response distribution. For example, when $\tau = 0.5$, $\mathbf{X}_i^T \boldsymbol{\beta}_\tau$ is the conditional median of θ_i , and $\boldsymbol{\beta}_\tau$ reflects the relationship between explanatory variables and the conditional median of θ_i . In QR approaches, the classical inference of $\boldsymbol{\beta}_\tau$ is through a minimization of the weighted sum of absolute residuals for all the observations.

Therefore, θ_i can be assessed in the following QR-type structural model, by the τ th conditional quantile of θ_i plus an error term:

$$\theta_i = \mathbf{X}_i^T \boldsymbol{\beta}_\tau + \delta_i. \quad (4)$$

Unlike in the LR model [i.e., Eq. (1)], here the distribution of δ_i is not specified. The only assumption is that the τ th quantile of δ_i is zero, that is $q_\tau = 0$, so that we have Eq. (3). In other words, the QR structural model does not rely on any parametric specification of the conditional distribution of θ_i . Note that when the error term δ_i in Eq. (4) is normally or symmetrically distributed, results of LR and QR at the median ($\tau = 0.5$) are consistent, in other words, the MLIRT model of Eq. (1) is a special version of the Q-MLIRT model with normality of δ_i and $\tau = 0.5$.

Therefore, the two-level Q-MLIRT model can be defined by combining Eq. (4) with an IRT model, namely:

$$P(Y_{ik} = 1 | \theta_i, a_k, b_k) = \Phi(a_k(\theta_i - b_k)),$$

$$\theta_i = \mathbf{X}_i^T \boldsymbol{\beta}_\tau + \delta_i, \quad (5)$$

where we adopt a two-parameter normal ogive (2PNO) model in the first level. In this article, only dichotomously scored items are considered. Here, Φ denotes the cumulative standard normal distribution function, $Y_{ik} = 1$ means a correct response of student i on item k ($k = 1, \dots, K$), and a_k and b_k are the discrimination and difficulty parameters, respectively, of item k . The second level is QR on the ability parameters with an arbitrary quantile $\tau \in (0, 1)$, and the parameters have the same meanings as those in Eq. (4).

Herein, we fix the item parameters of one item as known (i.e., $a_1 = 1$ and $b_1 = 0$) to identify the Q-MLIRT model as did Fox and Glas (2001). In the subsequent analysis, we show that the proposed model offers a more complete view of the relationship among explanatory variables and latent variables at various locations of the response distribution, including the central tendency, upper tail, and lower tail. In many cases, the error terms are non-normally distributed or heterogeneous, which fails to meet the assumption of LR, but nevertheless QR is also suitable for such cases.

BAYESIAN ESTIMATION

We adopt the Bayesian estimation approach in the QR structural model (e.g., Yu and Moyeed, 2001; Lee and Neocleous, 2010; Reich et al., 2011; Wang et al., 2016). Combined with the Gibbs sampling method for 2PNO models (Albert, 1992), a fully Bayesian approach is presented to estimate the parameters of the Q-MLIRT model. Similar to Wang et al. (2016), we adopt the asymmetric Laplace distribution1 (ALD¹; Yu and Moyeed, 2001) to approximate the distribution of error terms in the QR structural model. More details about the ALD and the Bayesian QR method are presented in Appendix. Specifically, we assume $\delta_i \sim ALD(0, \omega, \tau)$, then θ_i follows $ALD(\mathbf{X}_i^T \boldsymbol{\beta}_\tau, \omega, \tau)$. As shown by Reed and Yu (2009); Kozumi and Kobayashi (2011), if a random variable follows $ALD(\mu, \omega, \tau)$ then it can be represented as a normal distribution, so we obtain:

$$\theta_i | \boldsymbol{\beta}_\tau, \omega, e_i \sim N(\mathbf{X}_i^T \boldsymbol{\beta}_\tau + k_1 e_i, k_2 \omega e_i), \quad (6)$$

where $k_1 = \frac{1-2\tau}{\tau(1-\tau)}$, $k_2 = \frac{2}{\tau(1-\tau)}$, and $\mathbf{e} = \{e_1, e_2, \dots, e_n\}$ follows an exponential distribution with scale parameter ω . That is to say, the parameter space is augmented by a latent variable \mathbf{e} . This mixed representation of $ALD(\mathbf{X}_i^T \boldsymbol{\beta}_\tau, \omega, \tau)$ enables Bayesian inference based on the Gibbs sampling algorithm to estimate the parameters of the QR structural model.

Gibbs Sampling

From Eq. (5), the observations consist of the item responses \mathbf{Y} in the first level and the explanatory variables \mathbf{X} in the second level. As a result, the full posterior distribution of the parameters given the observed data is:

$$p(\mathbf{Z}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \boldsymbol{\beta}_\tau, \mathbf{e}, \omega | \mathbf{Y}, \mathbf{X}) \propto \prod_{i=1}^n \left(\left(\prod_{k=1}^K p(z_{ik} | \theta_i, a_k, b_k, y_{ik}) \right) p(\theta_i | \boldsymbol{\beta}_\tau, e_i, \omega, \mathbf{X}_i) p(\mathbf{e} | \omega) p(\mathbf{a}) p(\mathbf{b}) p(\boldsymbol{\beta}_\tau) p(\omega), \quad (7)$$

where the augmented data z_{ik} follow:

$$p(z_{ik} | \theta_i, a_k, b_k, y_{ik}) \propto \varphi(z_{ik}; a_k(\theta_i - b_k), 1) [I(z_{ik} > 0) I(y_{ik} = 1) + I(z_{ik} \leq 0) I(y_{ik} = 0)],$$

where φ is the normal probability density function (PDF) and $I(\cdot)$ is an indicator function.

The following conjugate prior distributions of the model parameters are used:

$$a_k \sim N(\mu_a, \sigma_a^2) I(a_k > 0),$$

$$b_k \sim N(\mu_b, \sigma_b^2),$$

$$\boldsymbol{\beta}_\tau \sim N_q(\Lambda_{0\boldsymbol{\beta}}, \mathbf{H}_{0\boldsymbol{\beta}}),$$

$$\omega^{-1} \sim \text{Gamma}(\alpha_{0\omega}, \beta_{0\omega}),$$

where $\mu_a, \sigma_a, \mu_b, \sigma_b, \Lambda_{0\boldsymbol{\beta}}, \mathbf{H}_{0\boldsymbol{\beta}}, \alpha_{0\omega}$, and $\beta_{0\omega}$ are hyperparameters. The detailed procedure of the Gibbs sampling algorithm is summarized below.

Step 1: Sampling z_{ik} . Let $\boldsymbol{\xi}$ be the vector of all item parameters, given the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$, the variables z_{ik} are independent. For $i = 1, \dots, n$ and $k = 1, \dots, K$, the fully conditional posterior distribution of z_{ik} can be written as:

$$(z_{ik} | \boldsymbol{\theta}, \boldsymbol{\xi}, y) \sim \begin{cases} N(a_k(\theta_i - b_k), 1) \text{ truncated at the left by } 0 \text{ if } y_{ik} = 1 \\ N(a_k(\theta_i - b_k), 1) \text{ truncated at the right by } 0 \text{ if } y_{ik} = 0 \end{cases} \quad (8)$$

Step 2: Sampling a_k . According to the principle of conjugate distribution and combined with the prior of a , the fully conditional posterior density of a_k follows:

$$(a_k | \mathbf{z}_k, \boldsymbol{\theta}, b_k) \sim N \left(\frac{\frac{\mu_a}{\sigma_a^2} + \sum_{i=1}^n [z_{ik}(\theta_i - b_k)]}{\frac{1}{\sigma_a^2} + \sum_{i=1}^n (\theta_i - b_k)^2}, \frac{1}{\frac{1}{\sigma_a^2} + \sum_{i=1}^n (\theta_i - b_k)^2} \right) \quad (9)$$

Step 3: Sampling b_k . The fully conditional posterior density of b_k is:

$$(b_k | \mathbf{z}_k, \boldsymbol{\theta}, a_k) \sim N \left(\frac{\frac{\mu_b}{\sigma_b^2} - \sum_{i=1}^n (z_{ik} - a_k \theta_i) a_k}{\frac{1}{\sigma_b^2} + n a_k^2}, \frac{1}{\frac{1}{\sigma_b^2} + n a_k^2} \right). \quad (10)$$

Step 4: Sampling θ_i . The ability parameters are independent given other parameters. Using Eqs. (6) and (8), it follows that:

$$f(\theta_i | \mathbf{z}_i, \boldsymbol{\xi}, \boldsymbol{\beta}_\tau, \omega, e_i) \propto f(\mathbf{z}_i | \theta_i, \boldsymbol{\xi}) f(\theta_i | \boldsymbol{\beta}_\tau, \omega, e_i), \quad (11)$$

Eq. (11) is a normal model, where θ_i has a normal prior parameterized by $\boldsymbol{\beta}_\tau, e_i$, and ω . So the fully conditional posterior density of θ_i is given by:

$$(\theta_i | \mathbf{z}_i, \boldsymbol{\xi}, \boldsymbol{\beta}_\tau, \omega, e_i) \sim N \left(\frac{\frac{\mu_\theta}{\sigma_\theta^2} + \sum_{k=1}^K (z_{ik} + a_k b_k) a_k}{\frac{1}{\sigma_\theta^2} + \sum_{k=1}^K a_k^2}, \frac{1}{\frac{1}{\sigma_\theta^2} + \sum_{k=1}^K a_k^2} \right), \quad (12)$$

where $\mu_\theta = \mathbf{X}_i^T \boldsymbol{\beta}_\tau + k_1 e_i, \sigma_\theta^2 = k_2 \omega e_i$.

Step 5: Sampling ω . Letting $\nu = \omega^{-1}$ and using the gamma conjugate prior density for ν and the PDF of $ALD(\mathbf{X}_i^T \boldsymbol{\beta}_\tau, \omega, \tau)$, the fully conditional posterior distribution of ν is given by:

$$p(\nu | \boldsymbol{\theta}, \boldsymbol{\beta}_\tau) \propto \nu^{n+\alpha_{0\omega}-1} \exp\{-[\beta_{0\omega} + \sum_{i=1}^n \rho_\tau (\theta_i - \mathbf{X}_i^T \boldsymbol{\beta}_\tau)] \nu\},$$

from which we have:

$$(\omega^{-1} | \boldsymbol{\theta}, \boldsymbol{\beta}_\tau) \sim \text{Gamma}(n + \alpha_{0\omega}, \beta_{0\omega} + \sum_{i=1}^n \rho_\tau (\theta_i - \mathbf{X}_i^T \boldsymbol{\beta}_\tau)). \quad (13)$$

Step 6: Sampling e_i . From Eq. (6) together with an exponential distribution of e_i with scale parameter ω , the fully conditional

posterior distribution of e_i^{-1} follows an inverse Gaussian distribution.

$$(e_i^{-1} \mid \theta_i, \beta_\tau, \omega) \sim \text{Inverse - Gaussian} \left(\frac{\sqrt{2k_2 + k_1^2}}{|\theta_i - \mathbf{X}_i^T \beta_\tau|}, \frac{2k_2 + k_1^2}{k_2 \omega} \right). \quad (14)$$

Step 7: Sampling β_τ . Because Eq. (6) is a normal distribution conditionally on e_i , it is straightforward to derive the fully conditional posterior distribution of β_τ , which is given by:

$$(\beta_\tau \mid \theta, \mathbf{e}, \omega) \sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta^{-1}), \quad (15)$$

where $\boldsymbol{\mu}_\beta = \sum_{\beta}^{-1} (\mathbf{H}_{0\beta}^{-1} \boldsymbol{\Lambda}_{0\beta} + \sum_{i=1}^n \frac{(\theta_i - k_1 e_i) \mathbf{X}_i}{k_2 \omega e_i})$ and $\boldsymbol{\Sigma}_\beta = \mathbf{H}_{0\beta}^{-1} + \sum_{i=1}^n \frac{\mathbf{X}_i \mathbf{X}_i^T}{k_2 \omega e_i}$.

SIMULATION STUDY

In this section, a simulation study is carried out to evaluate the performance of the Q-MLIRT model. We evaluate the parameter recovery with the Gibbs sampler. The Gibbs sampling algorithm is implemented in MATLAB (MathWorks, 2016), and the source code is available to readers upon request.

Simulation Design

Two covariates X_1 and X_2 are considered, which are independent standard normal variables. The latent variable θ_i ($i = 1, \dots, n$) is generated from three different models:

Case 1: $\theta_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \delta_i, \delta_i \sim N(0, 0.5)$

Case 2: $\theta_i = \beta_1 X_{i1} + \beta_2 X_{i2} - \rho_\tau |\beta_1 X_{i1} + \beta_2 X_{i2}| + |\beta_1 X_{i1} + \beta_2 X_{i2}| \delta_i, \delta_i \sim N(0, 1)$

Case 3: $\theta_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \delta_i, \delta_i \sim \text{Gamma}(0.5, 1) - 0.5$

Parameters β_1 and β_2 are set to 0.5. Case 1 represents a normal and homoscedastic error model, the assumption of which conforms to that of LR models, case 2 represents a heteroscedastic error model, and case 3 represents a skewed error model. Of these, cases 2 and 3 evaluate the Q-MLIRT model with heavy-tailed and non-normal latent variables, respectively. For cases 1 and 3, the conditional τ th quantile of θ takes the form $Q_\tau(\theta \mid \mathbf{X}_1, \mathbf{X}_2) = \beta_{0\tau} + \beta_{1\tau} \mathbf{X}_1 + \beta_{2\tau} \mathbf{X}_2$, where $\beta_{0\tau} = \rho_\tau, \beta_{1\tau} = 0.5, \beta_{2\tau} = 0.5$ for $\tau \in (0, 1)$, and ρ_τ is the τ th quantile of δ_i . For case 2, the conditional quantile function is $Q_\tau(\theta \mid \mathbf{X}_1, \mathbf{X}_2) = \beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2 - \rho_\tau |\beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2| + \rho_\tau |\beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2| = \beta_{1\tau} \mathbf{X}_1 + \beta_{2\tau} \mathbf{X}_2$, where $\beta_{1\tau} = 0.5$ and $\beta_{2\tau} = 0.5$ for $\tau \in (0, 1)$.

In order to estimate the QR coefficients $\beta_{1\tau}$ and $\beta_{2\tau}$ under different quantiles, and compare with their theoretical values 0.5, three quantile levels (i.e., $\tau = 0.25, 0.5$, and 0.75) are chosen, which represent the lower, central, and upper tail of the response distribution, respectively. We consider three sample sizes (i.e., $n = 500, 1000, 2000$) and two test lengths (i.e., $K = 20, 40$). Overall, there are 54 conditions (3 structural models \times 3 quantiles \times 3 sample sizes \times 2 test lengths). Response patterns are generated by the 2PNO model, with the

discrimination parameters generated from $U(0.5, 1.5)$ and the difficulty parameters generated from $N(0, 0.5)$.

For the conjugate normal prior of β_τ , we set the prior mean $\boldsymbol{\Lambda}_{0\beta} = (0, 0, 0)$ and the covariance matrix $\mathbf{H}_{0\beta}$ is 100 times an identity matrix; which is a type of weakly informative prior. For the conjugate inverse gamma prior of ω , we set $\alpha_{0\omega} = 28$ and $\beta_{0\omega} = 4$. For the prior of item parameters a and b, we set $\mu_a = 0, \sigma_a = 200$ and $\mu_b = 0, \sigma_b = 100$, respectively.

The Gibbs sampling estimation procedure of the Q-MLIRT model was iterated 10,000 times. A burn-in period of 5,000 iterations was used, and the parameter estimate was the mean of the posterior distribution of the parameter. The convergence of the Gibbs sampler was checked by the Gelman–Rubin method (Gelman and Rubin, 1992), and the values of the potential scale reduction factor (PSRF) in the burn-in period were less than 1.10 for all parameters. Overall, 100 replications were conducted across all simulation conditions.

A related issue must be considered in the application of Q-MLIRT, namely that the estimates of item parameters differ for different quantiles because of the fact that the fully conditional posterior distribution of item parameters depends on τ . As can be seen from Eq. (6), the location and scale of θ_i are changed in the estimation process of item parameters under different quantiles. In this case, bias may not make sense, a large bias does not mean poor estimates, and the results of different quantiles are also not comparable. We expect the estimates of item parameters under different quantiles are close to each other after some transformation. To this end, we introduce a new measurement that is actually suitable for situations in which item parameters are estimated on different IRT scales, to further illustrate the similarity and comparability between them.

It is known that if θ_i have a transformation relationship on two scales, namely:

$$\theta_i^* = p\theta_i + t,$$

where p and t are constants, then the relationships between the item parameters can be formulated as:

$$a_k^* = \frac{a_k}{p},$$

$$b_k^* = pb_k + t.$$

In this case, we can use similarity functions, such as the cosine function, to measure the similarity of item parameters at different scales. Two vectors are defined to be similar if the distance between them is small, which is measured by the cosine of the angle between them. Let $\mathbf{a} = (a_1, \dots, a_k)$ and $\mathbf{a}^* = (a_1^*, \dots, a_k^*)$ be the discrimination parameter vectors on two different scales, then the cosine similarity between the two vectors is defined as:

$$\cos(\mathbf{a}, \mathbf{a}^*) = \frac{\sum_k a_k a_k^*}{\sqrt{\sum_k (a_k)^2} \sqrt{\sum_k (a_k^*)^2}}. \quad (16)$$

If $a_k^* = \frac{a_k}{p}$, then $\cos(\mathbf{a}, \mathbf{a}^*) = 1$. Similarly, the cosine similarity between two difficulty parameter vectors $\mathbf{b} = (b_1, \dots, b_k)$ and

$\mathbf{b}^* = (b_1^*, \dots, b_k^*)$ can be defined as:

$$\cos(\mathbf{b}, \mathbf{b}^*) = \frac{\sum_k (b_k - E\mathbf{b})(b_k^* - E\mathbf{b}^*)}{\sqrt{\sum_k (b_k - E\mathbf{b})^2} \sqrt{\sum_k (b_k^* - E\mathbf{b}^*)^2}}, \quad (17)$$

where $E\mathbf{b}$ and $E\mathbf{b}^*$ are the means of the elements in vectors \mathbf{b} and \mathbf{b}^* , respectively. If $b_k^* = pb_k + t$, then $\cos(\mathbf{b}, \mathbf{b}^*) = 1$.

In this sense, if the cosine of the angle between the true-parameter vector and the estimated-parameter vector is close to 1, it can be concluded that these two vectors are very similar and close, and the estimates can also be called “good estimates;” conversely, if the cosine is away from 1, it indicates that these two vectors are very different, and the estimates can also be called “bad estimates.” The higher the cosine similarity, the closer they are. Note that the “good” and “bad” estimates are just defined in the context of the transformation relationship that we considered as above, and accuracy of all the item parameters’ estimations is measured by the cosine similarity. The upshot is, when the estimated results of item parameters under each quantile are close to each other and are both good (i.e., the calculated cosine similarity of item parameters under each quantile is close to 1), we say that the item parameters are well recovered and that any group of estimates can be selected. Otherwise, the recovery of item parameters is poor and the results of different quantiles are not comparable to one another to some extent.

Given that there is no existing work on Q-MLIRT in the literature, we compare the estimated results with the two-level structural IRT model denoted in Eq. (1), where the structural model is a LR on θ_i . Hereafter, we refer to it as the mean regression multilevel IRT (M-MLIRT) model. We estimate it through a fully Bayesian estimation procedure that is easily extracted from the estimation procedure in Fox and Glas (2001). We use the average root-mean-square error (RMSE), the average bias, and the average cosine similarity to evaluate parameter recovery.

Simulation Results

Tables 1, 2 summarize the parameter recovery of the item parameters for the Q-MLIRT model in comparison to the M-MLIRT model for small and large test lengths respectively. In general, the item parameters are estimated accurately in all the cases (i.e., cases 1, 2, and 3). For each sample size, estimates of the item parameters of the Q-MLIRT model under different quantiles are close to the true parameter values, reflected mainly in the facts that (i) the cosine similarities between the true-parameter vector and the estimated-parameter vectors under different quantiles are all very close to 1 and (ii) the RMSEs are all small. In addition, when the sample size increases from 500 to 2,000, the cosine similarities increase, the RMSE and the bias in almost cases decrease for the estimates of item parameters. Results show that with increasing test length, the accuracy of almost all item parameters’ estimations also has some improvement, this could be because we can get more accurate information from students as they answer more items.

From Tables 1, 2, it is also indicated that, as for the estimation of item parameters, the item parameters’ estimations of the M-MLIRT model and the Q-MLIRT model are close in each case, mainly from the reflect of the cosine similarities and RMSEs. It should be noted that although for case 2 and case 3, the distribution of ability is non-normal, it seems not to have effects on the estimation of item parameters for the MLIRT model. Specifically, for case 3, the bias of the difficulty parameters for the M-MLIRT model are always larger than the Q-MLIRT model, but from the cosine similarity and RMSE points of view, the estimates of the difficulty parameters for the M-MLIRT model and Q-MLIRT model are close. The relationship between the cosine similarity and the bias, and how they behave in more different situations, are worthy of further study and discussion.

Tables 3, 4 give the estimated results of the structural regression coefficients for the Q-MLIRT model in comparison with those of the M-MLIRT model. First, it can be concluded that the regression coefficients are estimated accurately in all the cases (cases 1, 2, and 3) for the Q-MLIRT model. When the sample size increases from 500 to 2,000, the RMSEs decrease with increasing sample size, which means that the estimates are increasingly stable. The bias of regression coefficients also decreases obviously with increasing sample size, especially when $\tau = 0.5$. The trends are less obvious for quantiles 0.25 and 0.75, mainly because there are few samples at the upper and lower tails ($\tau = 0.75$ and 0.25, respectively) of the conditional distribution of θ_i . Nevertheless, the directions of the bias are increasingly consistent with increasing sample size, especially when $n = 2,000$, which means that the estimates tend to true and stable values. From Table 4, the results show that as larger test length leads to more accurate estimations of the structural regression coefficients, mainly reflecting in the decreased biases and RMSEs in almost all conditions. Together with the estimation of item parameters, it can be concluded that the increase of test length will improve the accuracy of parameter estimation of the Q-MLIRT model, especially for the structural regression coefficients.

When comparing the estimation of regression coefficients of the M-MLIRT model and the Q-MLIRT model, we find that: first, M-MLIRT slightly outperforms Q-MLIRT in case 1. This is because that the error distribution in case 1 is normal, which meets the assumption of the LR model, so we expect the M-MLIRT model to outperform in that case; second, for case 2, the regression coefficients estimated by the M-MLIRT model are worse than those by the Q-MLIRT model to some extent. Specifically, the results are similar at $\tau = 0.5$ when the sample size is small, but with increasing sample size, there are larger bias and RMSE for the estimates of the M-MLIRT model than those of the Q-MLIRT model. This is mainly because the error is heteroscedastic and the distribution of θ_i is heavy-tailed, and LR models are sensitive to heavy-tailed distributions and extreme outliers. For case 3, the results are close to each other, although the error term has a skewed distribution. It is perhaps that the degree of skewness of the distribution in case 3 is insufficient to influence the estimation of the regression coefficients in MLIRT.

TABLE 1 | Summary of parameter recovery of item parameters when $K = 20$.

Case	n	Method	τ	a			b			
				Cos	Bias	RMSE	Cos	Bias	RMSE	
1	500	Q	0.25	0.994	0.069	0.162	0.985	0.010	0.115	
			0.5	0.994	0.124	0.198	0.987	-0.004	0.112	
			0.75	0.994	0.074	0.171	0.986	-0.009	0.116	
	1,000	M	0.25	0.995	0.024	0.158	0.992	-0.002	0.118	
			0.5	0.997	0.042	0.114	0.992	0.007	0.082	
			0.75	0.997	0.032	0.111	0.992	-0.006	0.083	
		2,000	M	0.25	0.997	2E-04	0.112	0.996	0.001	0.071
				0.5	0.998	0.020	0.084	0.996	0.002	0.059
				0.75	0.999	0.030	0.079	0.996	-0.009	0.056
2	500	Q	0.25	0.998	0.020	0.080	0.996	-0.014	0.061	
			0.5	0.999	0.003	0.070	0.997	4E-04	0.047	
			0.75	0.993	0.054	0.163	0.984	0.038	0.102	
	1,000	M	0.25	0.992	0.108	0.194	0.986	-0.003	0.099	
			0.5	0.993	0.056	0.166	0.982	-0.037	0.107	
			0.75	0.992	0.031	0.197	0.985	-4E-04	0.105	
		2,000	Q	0.25	0.996	0.052	0.130	0.991	0.046	0.089
				0.5	0.996	0.064	0.131	0.993	0.004	0.071
				0.75	0.996	0.051	0.123	0.991	-0.039	0.084
3	500	M	0.25	0.997	0.021	0.107	0.992	3E-04	0.070	
			0.5	0.998	0.043	0.097	0.995	0.049	0.075	
			0.75	0.998	0.032	0.089	0.996	0.003	0.054	
	1,000	Q	0.25	0.998	0.045	0.096	0.996	-0.041	0.071	
			0.5	0.998	0.011	0.083	0.996	3E-04	0.053	
			0.75	0.994	0.015	0.139	0.986	-0.003	0.111	
		2,000	M	0.25	0.994	0.067	0.162	0.986	-0.015	0.108
				0.5	0.992	0.062	0.175	0.984	-0.025	0.116
				0.75	0.993	-0.003	0.144	0.987	0.021	0.106
1,000	Q	0.25	0.997	-0.006	0.096	0.993	0.013	0.078		
		0.5	0.997	0.039	0.118	0.993	-0.008	0.082		
		0.75	0.995	0.018	0.124	0.991	-0.007	0.082		
	2,000	M	0.25	0.996	-0.014	0.124	0.995	0.027	0.078	
			0.5	0.999	0.009	0.074	0.996	-0.005	0.055	
			0.75	0.998	0.035	0.089	0.996	-0.016	0.061	
2,000	Q	0.25	0.997	0.020	0.096	0.995	-0.023	0.067		
		0.5	0.998	-0.013	0.098	0.997	0.023	0.061		
		0.75	0.997	0.020	0.096	0.995	-0.023	0.067		

Q, Q-MLIRT model; M, M-MLIRT model.

EMPIRICAL EXAMPLE

The main research question in this section is how the explanatory variables of interest regarding individual characteristics and family background are related to students' achievement in mathematics. It is quite possible that the relation between explanatory variables and students' achievement may differ across different levels of students' achievement, while LR only provides us the information about the conditional mean of students' achievement and ignores the characteristics of the entire distribution of students' achievement. Thus, the proposed Q-MLIRT model is expected to achieve a more appropriate analysis.

Data Source

In this section, a subset of the PISA 2018 mathematics data was analyzed. There were 24 test forms/booklets that contained clusters of items from the mathematics domain. Each test form included four clusters, two clusters of items from the reading domain and one or two clusters of items from the mathematics domain. We chose a group of students from the United States who were administered test forms 1 to 6 for analysis. Each of these forms contained two mathematics clusters. Therefore, there were six clusters (i.e., M01-M06) in which contained 64 mathematical items. Because each student received one of the forms from his or her assigned group, and each form contained different items, so each student answered about 20 items and each item was

TABLE 2 | Summary of parameter recovery of item parameters when $K = 40$.

Case	n	Method	τ	a			b				
				Cos	Bias	RMSE	Cos	Bias	RMSE		
1	500	Q	0.25	0.995	0.051	0.147	0.986	0.019	0.111		
			0.5	0.995	0.096	0.169	0.985	-0.001	0.112		
			0.75	0.995	0.079	0.159	0.987	-0.026	0.111		
	1,000	M	Q	0.25	0.995	0.011	0.146	0.985	0.005	0.114	
				0.5	0.997	0.039	0.113	0.993	0.001	0.078	
				0.75	0.998	0.059	0.121	0.992	0.003	0.082	
		2,000	M	Q	0.25	0.997	0.039	0.113	0.993	-0.012	0.079
					0.5	0.998	-0.003	0.091	0.993	0.002	0.073
					0.75	0.999	0.022	0.075	0.996	0.009	0.060
2	500	Q	0.25	0.999	0.030	0.075	0.997	0.002	0.054		
			0.5	0.999	0.030	0.075	0.997	0.002	0.054		
			0.75	0.999	0.022	0.074	0.996	-0.003	0.058		
	1,000	M	Q	0.25	0.999	0.009	0.070	0.997	-0.002	0.047	
				0.5	0.993	0.054	0.167	0.984	0.029	0.106	
				0.75	0.994	0.097	0.183	0.985	0.005	0.099	
		2,000	M	Q	0.25	0.993	0.067	0.161	0.983	-0.032	0.102
					0.5	0.994	-0.002	0.152	0.984	-0.002	0.099
					0.75	0.997	0.038	0.115	0.991	0.037	0.082
3	500	Q	0.25	0.997	0.042	0.108	0.993	0.001	0.067		
			0.5	0.997	0.047	0.119	0.991	-0.037	0.081		
			0.75	0.997	0.047	0.119	0.991	-0.037	0.081		
	1,000	M	Q	0.25	0.996	-0.001	0.113	0.991	0.003	0.071	
				0.5	0.998	0.030	0.088	0.996	0.040	0.068	
				0.75	0.998	0.030	0.084	0.996	-0.001	0.049	
		2,000	M	Q	0.25	0.998	0.032	0.088	0.996	-0.039	0.067
					0.5	0.998	-0.007	0.080	0.996	-0.002	0.046
					0.75	0.998	-0.007	0.080	0.996	-0.002	0.046
3	500	Q	0.25	0.995	6E-05	0.139	0.986	-0.005	0.116		
			0.5	0.995	0.069	0.160	0.985	-0.010	0.112		
			0.75	0.994	0.064	0.158	0.985	-0.029	0.117		
	1,000	M	Q	0.25	0.995	0.019	0.148	0.985	0.025	0.099	
				0.5	0.997	-0.005	0.101	0.994	2E-05	0.084	
				0.75	0.997	0.047	0.117	0.993	-0.008	0.081	
		2,000	M	Q	0.25	0.997	0.030	0.116	0.993	-0.015	0.084
					0.5	0.997	0.021	0.114	0.994	0.022	0.073
					0.75	0.997	0.021	0.114	0.994	0.022	0.073
3	2,000	Q	0.25	0.999	-0.005	0.070	0.996	1E-04	0.057		
			0.5	0.999	0.015	0.071	0.996	-0.008	0.056		
			0.75	0.998	0.012	0.082	0.996	-0.013	0.061		
	3	M	Q	0.25	0.998	5E-04	-0.082	0.996	0.027	0.057	

Q, Q-MLIRT model; M, M-MLIRT model.

answered about by 375 students. The sample size was $n = 840$ after eliminating students with missing data, and there were $K = 64$ dichotomous items. More information on the PISA 2018 data can be found at www.oecd.org/pisa/data/2018database/.

Three individual and family background variables, that have been used in most studies (e.g., Adams et al., 1997; Fox and Glas, 2001; Fox, 2005; Lu et al., 2018), were chosen as explanatory variables of interest. The student-grade variable (X_1) equaled -1, 0, and 1, which represented grade 9, grade 10, grade 11, respectively. The corresponding number of students in each grade was 59, 630, and 150. Since only one student was in grade 12, so her was removed, hence the final sample size was $n = 839$. The student-gender variable (X_2) equaled 0

(female) or 1 (male), the corresponding number of students was 421 and 418. The ESCS (economic, social, and cultural status of parents) variable (X_3) was normally standardized. Information of these explanatory variables are summarized in **Table 5**.

Method

The PISA 2018 technical report stated that the item parameters were calibrated by the two-parameter logistic model (2PLM) or the generalized partial credit model (GPCM). As only dichotomously scored items are considered, we use the 2PNO model here, and we fix the first discrimination parameter to one and the first difficulty parameter to zero to identify the model. To

TABLE 3 | Bayesian estimates of regression coefficients when $K = 20$.

Case	n	Method	τ	$\beta_{1\tau}$		$\beta_{2\tau}$	
				Bias	RMSE	Bias	RMSE
1	500	Q	0.25	-0.030	0.080	-0.036	0.081
			0.5	-0.049	0.084	-0.046	0.084
			0.75	-0.030	0.081	-0.032	0.083
		M	0.25	0.021	0.089	0.015	0.088
			0.5	-0.029	0.060	-0.026	0.064
			0.75	-0.017	0.055	-0.025	0.056
	1,000	Q	0.25	-0.029	0.060	-0.026	0.064
			0.5	-0.017	0.055	-0.025	0.056
			0.75	-0.028	0.061	-0.026	0.060
		M	0.25	0.007	0.058	0.008	0.058
			0.5	-0.021	0.045	-0.018	0.046
			0.75	-0.016	0.041	-0.015	0.043
2,000	Q	0.25	-0.021	0.045	-0.018	0.046	
		0.5	-0.016	0.041	-0.015	0.043	
		0.75	-0.025	0.047	-0.021	0.046	
	M	0.25	0.004	0.041	0.005	0.039	
		0.5	-0.043	0.080	-0.044	0.081	
		0.75	-0.068	0.095	-0.073	0.096	
2	500	Q	0.25	-0.043	0.080	-0.044	0.081
			0.5	-0.068	0.095	-0.073	0.096
			0.75	-0.046	0.086	-0.046	0.089
		M	0.25	-0.054	0.096	-0.062	0.099
			0.5	-0.046	0.069	-0.049	0.070
			0.75	-0.049	0.071	-0.047	0.070
	1,000	Q	0.25	-0.046	0.069	-0.049	0.070
			0.5	-0.049	0.071	-0.047	0.070
			0.75	-0.047	0.068	-0.047	0.068
		M	0.25	-0.067	0.082	-0.067	0.082
			0.5	-0.042	0.054	-0.046	0.058
			0.75	-0.034	0.051	-0.037	0.052
2,000	Q	0.25	-0.042	0.054	-0.046	0.058	
		0.5	-0.034	0.051	-0.037	0.052	
		0.75	-0.047	0.057	-0.048	0.060	
	M	0.25	-0.069	0.076	-0.070	0.076	
		0.5	-0.016	0.068	-0.010	0.067	
		0.75	-0.032	0.080	-0.030	0.079	
3	500	Q	0.25	0.012	0.068	0.013	0.065
			0.5	-0.016	0.068	-0.010	0.067
			0.75	-0.032	0.080	-0.030	0.079
		M	0.25	0.016	0.082	0.011	0.080
			0.5	-0.012	0.054	-0.010	0.054
			0.75	-0.018	0.057	-0.023	0.061
	1,000	Q	0.25	0.017	0.051	0.013	0.051
			0.5	-0.012	0.054	-0.010	0.054
			0.75	-0.018	0.057	-0.023	0.061
		M	0.25	0.010	0.054	0.010	0.056
			0.5	-0.012	0.039	-0.012	0.039
			0.75	-0.021	0.043	-0.022	0.044
2,000	Q	0.25	0.006	0.036	0.005	0.037	
		0.5	-0.012	0.039	-0.012	0.039	
		0.75	-0.021	0.043	-0.022	0.044	
	M	0.25	0.005	0.041	0.006	0.037	
		0.5	-0.049	0.071	-0.047	0.070	
		0.75	-0.047	0.068	-0.047	0.068	

Q, Q-MLIRT model; M, M-MLIRT model.

assess the effects of the explanatory variables above on students with different ability levels, the QR model is given by:

$$Q_{\tau}(\theta_i) = \mathbf{X}_i^T \beta_{\tau},$$

where $i = 1, \dots, 839$, $\beta_{\tau} = (\beta_{0\tau}, \beta_{1\tau}, \beta_{2\tau}, \beta_{3\tau})^T$ is a 4×1 vector of unknown QR coefficients under the τ th quantile of the ability distribution, $\mathbf{X}_i = (1, X_{1i}, X_{2i}, X_{3i})^T$ is a 4×1 vector that represents the explanatory variables of student i , the distribution of the error term is assumed as $AL(0, \omega, \tau)$. The estimation was conducted for the quantiles of $\tau = 5\%, 10\%, 25\%, 50\%, 75\%, 90\%$, and 95% to obtain a more thorough overview of the relationship between these explanatory variables and students' mathematical achievement.

TABLE 4 | Bayesian estimates of regression coefficients when $K = 40$.

Case	n	Method	τ	$\beta_{1\tau}$		$\beta_{2\tau}$	
				Bias	RMSE	Bias	RMSE
1	500	Q	0.25	-0.021	0.075	-0.024	0.074
			0.5	-0.037	0.075	-0.037	0.078
			0.75	-0.038	0.079	-0.034	0.081
		M	0.25	0.012	0.078	0.013	0.077
			0.5	-0.023	0.057	-0.023	0.056
			0.75	-0.025	0.060	-0.026	0.060
	1,000	Q	0.25	-0.023	0.057	-0.023	0.056
			0.5	-0.025	0.060	-0.026	0.060
			0.75	-0.021	0.058	-0.019	0.055
		M	0.25	0.008	0.054	0.008	0.052
			0.5	-0.011	0.040	-0.014	0.039
			0.75	-0.013	0.040	-0.015	0.041
2	500	Q	0.25	-0.026	0.075	-0.027	0.074
			0.5	-0.047	0.084	-0.040	0.083
			0.75	-0.036	0.073	-0.035	0.072
		M	0.25	-0.042	0.086	-0.037	0.088
			0.5	-0.020	0.057	-0.020	0.056
			0.75	-0.028	0.058	-0.029	0.058
	1,000	Q	0.25	-0.024	0.059	-0.027	0.060
			0.5	-0.020	0.057	-0.020	0.056
			0.75	-0.028	0.058	-0.029	0.058
		M	0.25	-0.047	0.069	-0.044	0.068
			0.5	-0.018	0.042	-0.019	0.043
			0.75	-0.025	0.046	-0.024	0.046
2,000	Q	0.25	-0.024	0.045	-0.023	0.045	
		0.5	-0.018	0.042	-0.019	0.043	
		0.75	-0.025	0.046	-0.024	0.046	
	M	0.25	-0.052	0.061	-0.051	0.060	
		0.5	-0.021	0.068	-0.020	0.066	
		0.75	-0.024	0.073	-0.035	0.076	
3	500	Q	0.25	0.014	0.066	0.015	0.067
			0.5	-0.021	0.068	-0.020	0.066
			0.75	-0.024	0.073	-0.035	0.076
		M	0.25	0.015	0.075	0.014	0.075
			0.5	-0.017	0.053	-0.013	0.051
			0.75	-0.021	0.060	-0.012	0.057
	1,000	Q	0.25	0.013	0.052	0.016	0.051
			0.5	-0.017	0.053	-0.013	0.051
			0.75	-0.021	0.060	-0.012	0.057
		M	0.25	-0.009	0.052	-0.006	0.054
			0.5	0.011	0.035	0.009	0.035
			0.75	-0.003	0.035	-0.005	0.034
2,000	Q	0.25	0.011	0.035	0.009	0.035	
		0.5	-0.003	0.035	-0.005	0.034	
		0.75	-0.008	0.039	-0.009	0.040	
	M	0.25	0.003	0.038	0.001	0.038	
		0.5	-0.049	0.071	-0.047	0.070	
		0.75	-0.047	0.068	-0.047	0.068	

Q, Q-MLIRT model; M, M-MLIRT model.

TABLE 5 | Summary of explanatory variables.

Label	Description	Mean	SD
Grade (X_1)	-1 = grade 9 0 = grade 10 1 = grade 11	0.1085	0.2376
Gender (X_2)	0 = female 1 = male	0.4982	0.2503
ESCS (X_3)	Normally distributed	0.1314	1.0356

In the Bayesian analysis, the hyperparameters of the prior distributions were given as assigned in the simulation study. In the implementation of the Gibbs sampling algorithm, 10,000 iterations were done to estimate the parameters with an initial

burn-in of 5,000 iterations. Convergence of the chains was checked by PSRF, and the PSRF values were less than 1.10 in the burn-in phase for all parameters under each quantile.

The M-MLIRT model was also used to fit the data, and compared with the Q-MLIRT model using deviance information criterion (DIC, Spiegelhalter et al., 2002). The model with smaller DIC was recommended as a better-fitting one. The DIC calculated using the joint likelihood conditioned on parameters at all levels. For the Q-MLIRT model, the joint likelihood is:

$$f(\theta, Z|\xi, \beta_\tau, \omega, \mathbf{e}) = f(Z | \theta, \xi)f(\theta|\beta_\tau, \omega, \mathbf{e}), \tag{18}$$

and the joint DIC for the τ th quantile of the Q-MLIRT model is defined as:

$$DIC_\tau = -2\log \{f(\theta, Z|\xi, \beta_\tau, \omega, \mathbf{e})\}. \tag{19}$$

For the M-MLIRT model, the joint likelihood is:

$$f(\theta, Z|\xi, \beta, \alpha^2) = f(Z | \theta, \xi)f(\theta|\beta, \alpha^2), \tag{20}$$

and the joint DIC is defined as:

$$DIC = -2\log \{f(\theta, Z|\xi, \beta, \alpha^2)\}. \tag{21}$$

Interested readers can refer to Celeux et al. (2006); Wang et al. (2013), Zhang et al. (2019) for more information of the joint DIC.

Results

The parameter estimates of the Q-MLIRT model under various quantiles and the M-MLIRT model (reported in column “M”) are presented in Tables 6, 7. For the limited space, we only present the results of the item parameters of cluster 1 from form 1, there are 9 items. From Table 6, it is showed that the estimates of item

TABLE 6 | Item parameter estimates of PISA data.

	Estimation at different quantiles							
	M	5%	10%	25%	50%	75%	90%	95%
a_1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
a_2	1.660	0.943	1.249	1.579	1.580	1.473	1.278	1.067
a_3	2.831	1.828	2.412	3.006	2.786	2.495	2.083	1.727
a_4	1.319	0.710	0.937	1.196	1.260	1.258	1.080	0.893
a_5	3.011	1.743	2.323	2.896	2.989	2.951	2.527	2.107
a_6	1.922	1.105	1.474	1.812	1.758	1.740	1.511	1.239
a_7	3.302	1.851	2.484	3.083	3.232	3.317	3.038	2.688
a_8	3.480	2.143	2.791	3.371	3.460	3.468	3.190	2.832
a_9	1.923	1.008	1.359	1.700	1.795	1.863	1.701	1.452
b_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
b_2	-0.282	-0.253	-0.307	-0.231	-0.303	-0.312	-0.392	-0.439
b_3	-0.346	-0.327	-0.361	-0.273	-0.360	-0.373	-0.470	-0.539
b_4	-0.029	0.195	0.033	0.042	-0.037	-0.028	-0.066	-0.055
b_5	-0.004	0.211	0.045	0.056	-0.011	0.009	-0.012	0.012
b_6	0.067	0.354	0.151	0.141	0.070	0.083	0.063	0.103
b_7	0.253	0.666	0.388	0.329	0.250	0.269	0.276	0.349
b_8	0.177	0.504	0.271	0.241	0.167	0.194	0.196	0.256
b_9	0.291	0.788	0.477	0.394	0.304	0.308	0.311	0.392

TABLE 7 | Structural regression coefficient estimates of PISA data.

τ	$\beta_{0\tau}$	SD	$\beta_{1\tau}$	SD	$\beta_{2\tau}$	SD	$\beta_{3\tau}$	SD
5%	-0.474	0.053	0.237	0.041	0.033	0.031	0.170	0.024
10%	-0.432	0.042	0.185	0.028	0.033	0.025	0.131	0.018
25%	-0.235	0.052	0.177	0.024	0.049	0.023	0.115	0.013
50%	-0.133	0.058	0.190	0.025	0.077	0.021	0.117	0.013
75%	0.070	0.040	0.196	0.026	0.098	0.023	0.125	0.013
90%	0.210	0.046	0.213	0.029	0.109	0.026	0.144	0.015
95%	0.337	0.044	0.238	0.030	0.131	0.030	0.168	0.016
M	-0.120	0.054	0.181	0.025	0.068	0.023	0.121	0.013

parameters obtained based on the M-MLIRT model are close to the results obtained from the Q-MLIRT model, especially at 50% quantile. We calculated the cosine similarity between each two discrimination-parameter vectors and each two difficulty-parameter vectors estimated under different quantiles, and these values were all very close to 1, that means they are very similar to each other, and we can choose any group of them as the estimations. Figure 1 shows item parameter estimates of the M-MLIRT model and three selected groups of item parameters estimated under $\tau = 25\%$, 50% , and 75% of the Q-MLIRT model, which represent the corresponding results of the lower, middle, and higher quantiles, respectively. As can be seen, all the estimates of item parameters under different quantiles are close and of the same size order.

The relationship between students’ mathematical achievement and the predictors can be deduced from the structural regression coefficients, the estimates of the QR and LR coefficients associate with the posterior standard deviations (SD) are showed in Table 7. The results of the M-MLIRT model show that, on average, the student grade and family background (i.e., ESCS) variables have significantly positive correlations with students’ mathematical achievement. Specifically, students in higher grades perform better, and students whose parents have higher economic, social, and cultural status have better mathematical achievement. But for gender, results of M-MLIRT show that there is a weakly positive correlation between gender and mathematical achievement, the performance of male students is slightly better than female students.

When the results of M-MLIRT and Q-MLIRT are compared, it is argued that there are several differences between these two models. The results of the Q-MLIRT model show that relationships between these explanatory variables and students’ mathematical achievement are different across the achievement distribution. First, we find positive correlations between those explanatory variables and mathematical achievement at various locations along the distribution, and the results of Q-MLIRT at 50% quantile are as close as the M-MLIRT. For student grade and family background variables, it is found that the positive correlations between them and mathematical achievement are more significant at lower and higher quantiles of the distribution, that is the relationships between them and mathematical achievement are stronger in students with high-ability and low-ability. Regarding gender, we find that the QR structural coefficient of gender increases from 5% quantile towards the 95%

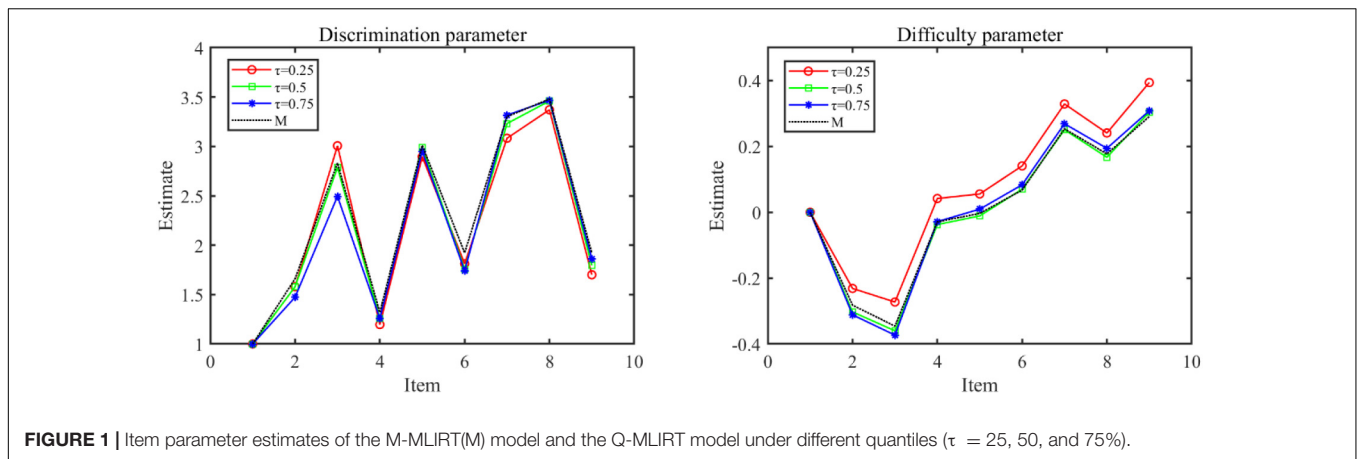


FIGURE 1 | Item parameter estimates of the M-MLIRT(M) model and the Q-MLIRT model under different quantiles ($\tau = 25, 50, \text{ and } 75\%$).

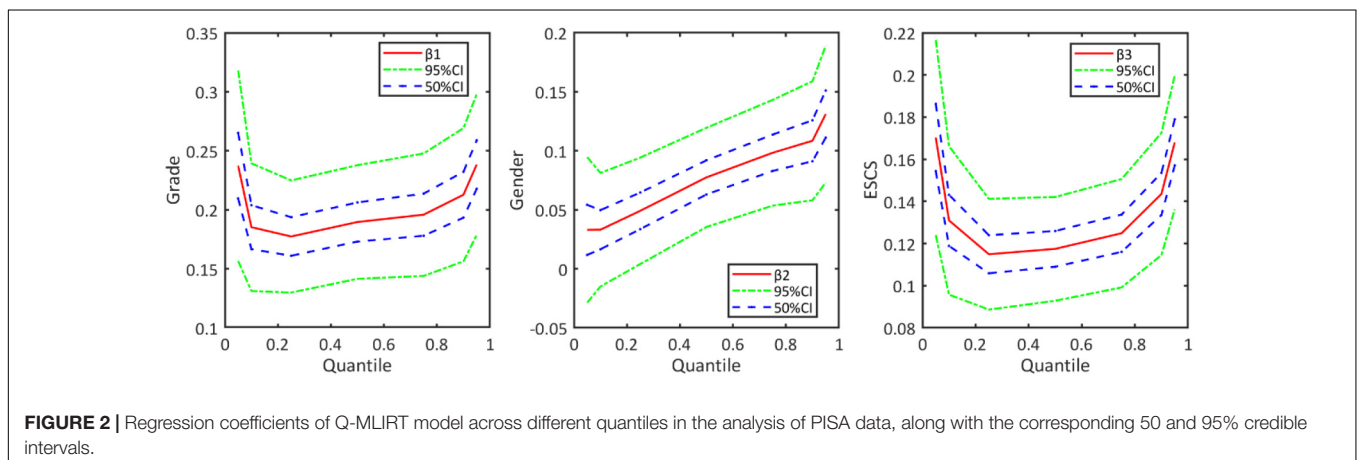


FIGURE 2 | Regression coefficients of Q-MLIRT model across different quantiles in the analysis of PISA data, along with the corresponding 50 and 95% credible intervals.

quantile of the achievement distribution, which means that male students perform better than female students in mathematics, and the difference is exacerbated among high-ability students. This finding is consistent with González de San Román and De La Rica (2012). They analyzed gender gaps in PISA test scores in their study, and showed that in most PISA participating countries, male students outperformed than female students in mathematical achievement, moreover, gender gap in test scores differed significantly in different parts of the test score distribution, where the gap increased from the lower tail to the upper tail. The structural regression coefficients across the various quantiles are illustrated further in **Figure 2**.

In addition, **Table 8** shows a comparison of the Q-MLIRT model and the M-MLIRT model based on the joint DIC. Results indicate that for the Q-MLIRT model, 25% quantile is the best-fitting, followed by quantiles at 75%, 50%, 10%, 5%,

90% and 95% ones. And the fitting situation of M-MLIRT is nearly equally to the Q-MLIRT model when the quantile is 50% or 75%, this is consistent with what is found in the estimation of parameters, that the parameter estimation results of M-MLIRT is mostly close to the results of Q-MLIRT at 50% and 75% quantiles.

To sum up, the analysis results make it clear that the Q-MLIRT model offers a more comprehensive picture of the relationship between students' mathematical achievement and the student-background variables. The results suggest that the relationships between students' mathematical achievement and its predictors differ across different levels of student ability, which can't be achieved with MLIRT. We find that for students with lower and higher mathematical performances in PISA, their achievements are related more closely to grade, and the socioeconomic status of their family, and the gender difference is more pronounced among high-level students. In this respect, Q-MLIRT is a useful research tool to better understand the relationship between students' individual characteristics or family background factors and educational achievement. Note that the results, together with some educational knowledge, would provide a deeper insight for related educators on mathematical learning, and these findings will also help them to develop more-targeted policies for students of different ability levels.

TABLE 8 | Comparing the Q-MLIRT and M-MLIRT using joint DIC for PISA data.

M-MLIRT	DIC of Q-MLIRT at different quantiles						
	5%	10%	25%	50%	75%	90%	95%
51926.1	52012.9	51990.1	51859.7	51980.2	51971.5	52076.0	52189.4

DISCUSSION

In this article, a quantile-regression-type structural model was introduced to MLIRT models. The main advantage of this form is that it offers a richer analysis of the relationship between covariates and latent variables. Meanwhile, Q-MLIRT is more robust and is applicable to various distribution of latent variables. In other words, it is more flexible and applicable in general. We also proposed a fully Bayesian estimation algorithm for the Q-MLIRT model so that the parameters in the model could be estimated simultaneously. In addition, we proposed a new evaluation index based on the cosine similarity function to evaluate further the accuracy of the item parameter estimates under different quantiles.

In the simulation study, we used Gibbs sampling to estimate the Q-MLIRT model. The results showed that the parameters could be recovered well in almost all conditions. The Q-MLIRT model was applied to analyze a PISA data set, and the results provided more specific and in-depth insights into the relationship between the explanatory variables of interest and students' achievement. Also, note that the item parameter estimates differ for different quantiles, although we know they are very close to each other according to the cosine similarity evaluation. An alternative and easy way of applying the Q-MLIRT model in practical problems is to obtain the item parameters at a certain quantile first, such as $\tau = 0.5$ of the model, and then, for analysis with a different value of τ , the item parameters are pre-fixed (Wang et al., 2016).

This study was the first attempt at combining the QR approach with the IRT model in a multilevel modeling framework. However, this study focused only on item response data nested within students. We intend to extend the model to a higher level (school level) and estimate

how the school level affects students' abilities. Moreover, we considered herein only dichotomous response data and one-dimensional ability parameters, whereas Q-MLIRT could be applied to multidimensional and/or polytomous data. In addition, the explanatory variables that we have considered are known covariates, whereas it would make sense to consider latent explanatory variables measured by observations with measurement errors. Finally, variable selection would be of interest applied in Q-MLIRT, especially when some explanatory variables are either insignificant or significant only under some quantiles but not others.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.oecd.org/pisa/data/2018database>.

AUTHOR CONTRIBUTIONS

HZ did the simulation study and analysis on real data, and completed the writing of the article. WG provided key theoretical and technical support. XZ provided original thoughts and article revisions. All authors contributed to the article and approved the submitted version.

FUNDING

This research was partly supported by the National Natural Science Foundation of China (Grant 12001092) and the Key Laboratory of Applied Statistics of MOE.

REFERENCES

- Adams, R. J., Wilson, M., and Wu, M. (1997). Multilevel item response models: an approach to errors in variable regression. *J. Educ. Behav. Stat.* 22, 47–76. doi: 10.2307/1165238
- Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *J. Educ. Stat.* 17, 251–269. doi: 10.3102/10769986017003251
- Allen, D. E., and Powell, S. R. (2011). “Asset pricing, the fama–french factor model and the implications of quantile-regression analysis,” in *Financial Econometrics Modeling: Market Microstructure, Factor Models and Financial Risk Measures*, eds G. N. Gregoriou and R. Pascalau (London: Palgrave Macmillan), 176–193. doi: 10.1057/9780230298101_7
- Belloni, A., Chen, M., and Padilla, O. H. M. (2019). High dimensional latent panel quantile regression with an application to asset pricing. *arXiv [Preprint]*. arXiv:1912.02151.
- Bock, R. D., and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika* 46, 443–459. doi: 10.1007/bf02293801
- Bock, R. D., and Lieberman, M. (1970). Fitting a response model for dichotomously scored items. *Psychometrika* 35, 179–197. doi: 10.1007/bf02291262
- Burgette, L. F., and Reiter, J. P. (2012). Modeling adverse birth outcomes via confirmatory factor quantile regression. *Biometrics* 68, 92–100. doi: 10.1111/j.1541-0420.2011.01639.x
- Celeux, G., Forbes, F., Robert, C. P., and Titterton, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Anal.* 1, 651–673. doi: 10.1214/06-ba122
- Chen, F., and Chalhoub-Deville, M. (2014). Principles of quantile regression and an application. *Lang. Test.* 31, 63–87. doi: 10.1177/0265532213493623
- Costanzo, A., and Desimoni, M. (2017). Beyond the mean estimate: a quantile regression analysis of inequalities in educational outcomes using INVALSI survey data. *Large Scale Assess. Educ.* 5:14. doi: 10.1186/s40536-017-0048-4
- Dunson, D. B., Watson, M., and Taylor, J. A. (2003). Bayesian latent variable models for median regression on multiple outcomes. *Biometrics* 59, 296–304. doi: 10.1111/1541-0420.00036
- Edgeworth, F. Y. (1888). The mathematical theory of banking. *J. R. Statist. Soc.* 51, 113–127.
- Entink, R. K., Fox, J.-P., and van Der Linden, W. J. (2009). A multivariate multilevel approach to the modeling of accuracy and speed of test takers. *Psychometrika* 74, 21–48. doi: 10.1007/s11336-008-9075-y
- Feng, X.-N., Wang, Y., Lu, B., and Song, X.-Y. (2017). Bayesian regularized quantile structural equation models. *J. Multivariate Anal.* 154, 234–248. doi: 10.1016/j.jmva.2016.11.002
- Finch, H. (2010). Item parameter estimation for the MIRT model: bias and precision of confirmatory factor analysis—based models. *Appl. Psychol. Meas.* 34, 10–26. doi: 10.1177/0146621609336112

- Finch, H. (2011). Multidimensional item response theory parameter estimation with nonsimple structure items. *Appl. Psychol. Meas.* 35, 67–82. doi: 10.1177/0146621610367787
- Fox, J. (1997). *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, CA: Sage Publications, Inc.
- Fox, J.-P. (2004). Applications of multilevel IRT modeling. *Sch. Eff. Sch. Improv.* 15, 261–280. doi: 10.1080/09243450512331383212
- Fox, J.-P. (2005). Multilevel IRT using dichotomous and polytomous response data. *Br. J. Math. Stat. Psychol.* 58, 145–172. doi: 10.1348/000711005x38951
- Fox, J.-P. (2010). *Bayesian Item Response Modeling: Theory and Applications*. New York, NY: Springer. doi: 10.1007/978-1-4419-0742-4
- Fox, J.-P., and Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika* 66, 271–288. doi: 10.1007/bf02294839
- Fox, J.-P., and Glas, C. A. W. (2003). Bayesian modeling of measurement error in predictor variables using item response theory. *Psychometrika* 68, 169–191. doi: 10.1007/bf02294796
- Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* 7, 457–472. doi: 10.1214/ss/1177011136
- Giambona, F., and Porcu, M. (2015). Student background determinants of reading achievement in Italy. A quantile regression analysis. *Int. J. Educ. Dev.* 44, 95–107. doi: 10.1016/j.ijedudev.2015.07.005
- Goldstein, H. (1995). *Multilevel Statistical Models*, 2nd Edn. London: Edward Arnold.
- González de San Román, A., and De La Rica, S. (2012). *Gender Gaps in PISA Test Scores: The Impact of Social Norms and the Mother's Transmission of Role Attitudes*. IZA Discussion Paper No. 6338. Bonn: Institute of Labor Economics (IZA).
- Gürsakar, S., Murat, D., and Gürsakar, N. (2016). Assessment of PISA 2012 results with quantile regression analysis within the context of inequality in educational opportunity. *Alphanumeric J.* 4, 41–54. doi: 10.17093/aj.2016.4.2.5000186603
- Haile, G. A., and Nguyen, A. N. (2008). Determinants of academic attainment in the United States: a quantile regression analysis of test scores. *Educ. Econ.* 16, 29–57. doi: 10.1080/09645290701523218
- Hao, L., and Naiman, D. Q. (2007). *Quantile Regression* (No. 149). Thousand Oaks, CA: Sage.
- Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica* 44, 461–465. doi: 10.2307/1913974
- Huang, H. Y. (2015). A multilevel higher order item response theory model for measuring latent growth in longitudinal data. *Appl. Psychol. Meas.* 39, 362–372. doi: 10.1177/0146621614568112
- Jarque, C. M., and Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Econ. Lett.* 6, 255–259. doi: 10.1016/0165-1765(80)90024-5
- Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *J. Educ. Meas.* 38, 79–93. doi: 10.1111/j.1745-3984.2001.tb01117.x
- Kamata, A., and Bauer, D. J. (2008). A note on the relation between factor analytic and item response theory models. *Struct. Equ. Model.* 15, 136–153. doi: 10.1080/10705510701758406
- Koenker, R., and Bassett, G. Jr. (1978). Regression quantiles. *Econometrica* 46, 33–50. doi: 10.2307/1913643
- Koenker, R., and Bassett, G. Jr. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* 50, 43–61. doi: 10.2307/1912528
- Koenker, R., and Hallock, K. (2001). Quantile regression. *J. Econ. Perspect.* 15, 143–156. doi: 10.1257/jep.15.4.143
- Koenker, R., and Ng, P. (2005). Inequality constrained quantile regression. *Sankhya* 67, 418–440. doi: 10.1017/CBO9780511754098
- Kozumi, H., and Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *J. Stat. Comput. Simul.* 81, 1565–1578. doi: 10.1080/00949655.2010.496117
- Lee, D., and Neocleous, T. (2010). Bayesian quantile regression for count data with application to environmental epidemiology. *J. R. Stat. Soc. Ser. C Appl. Stat.* 59, 905–920. doi: 10.1111/j.1467-9876.2010.00725.x
- Liu, H., Song, X., Tang, Y., and Zhang, B. (2020). Bayesian quantile nonhomogeneous hidden markov models. *Statist. Methods Med. Res.* doi: 10.1177/0962280220942802
- Lu, J., Zhang, J., and Tao, J. (2018). Slice-gibbs sampling algorithm for estimating the parameters of a multilevel item response model. *J. Math. Psychol.* 82, 12–25. doi: 10.1016/j.jmp.2017.10.005
- Natesan, P., Limbers, C., and Varni, J. W. (2010). Bayesian estimation of graded response multilevel models using Gibbs sampling: formulation and illustration. *Educ. Psychol. Meas.* 70, 420–439. doi: 10.1177/0013164409355696
- Raudenbush, S. W. (1988). Educational applications of hierarchical linear models: a review. *J. Educ. Stat.* 13, 85–116. doi: 10.3102/10769986013002085
- Reed, C., and Yu, K. (2009). *A Partially Collapsed Gibbs Sampler for Bayesian Quantile Regression*. Technical Report. London: Brunel University.
- Reich, B. J., Fuentes, M., and Dunson, D. B. (2011). Bayesian spatial quantile regression. *J. Am. Stat. Assoc.* 106, 6–20. doi: 10.1198/jasa.2010.ap09237
- Sass, D. A., Schmitt, T. A., and Walker, C. M. (2008). Estimating non-normal latent trait distributions within item response theory using true and estimated item parameters. *Appl. Meas. Educ.* 21, 65–88. doi: 10.1080/0895734070196415
- Schmidt, S., Zlatkin-Troitschanskaia, O., and Fox, J.-P. (2016). Pretest-posttest-posttest multilevel IRT modeling of competence growth of students in higher education in Germany. *J. Educ. Meas.* 53, 332–351. doi: 10.1111/jedm.12115
- Snijders, T. A. B., and Bosker, R. J. (1999). *Multilevel Analysis: An introduction to Basic and Advanced Multilevel Modeling*. London: Sage.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. R. Statist. Soc. Ser. B (Statist. Methodol.)* 64, 583–639. doi: 10.1111/1467-9868.00353
- Stone, C. A. (1992). Recovery of marginal maximum likelihood estimates in the two-parameter logistic response model: an evaluation of MULTLOG. *Appl. Psychol. Meas.* 16, 1–16. doi: 10.1177/014662169201600101
- Svetina, D., Valdivia, A., Underhill, S., Dai, S., and Wang, X. (2017). Parameter recovery in multidimensional item response theory models under complexity and nonnormality. *Appl. Psychol. Meas.* 41, 530–544. doi: 10.1177/0146621617707507
- Trafimow, D., Wang, T., and Wang, C. (2018). Means and standard deviations, or locations and scales? That is the question! *New Ideas Psychol.* 50, 34–37. doi: 10.1016/j.newideapsych.2018.03.001
- Tsionas, E. G. (2003). Bayesian quantile inference. *J. Stat. Comput. Simul.* 73, 659–674. doi: 10.1080/0094965031000064463
- Wang, C., Chang, H. H., and Douglas, J. A. (2013). The linear transformation model with frailties for the analysis of item response times. *Br. J. Math. Statist. Psychol.* 66, 144–168. doi: 10.1111/j.2044-8317.2012.02045.x
- Wang, C., Su, S., and Weiss, D. J. (2018). Robustness of parameter estimation to assumptions of normality in the multidimensional graded response model. *Multivariate Behav. Res.* 53, 403–418. doi: 10.1080/00273171.2018.1455572
- Wang, C., and Zhang, X. (2019). A note on the conversion of item parameters standard errors. *Multivariate Behav. Res.* 54, 307–321. doi: 10.1080/00273171.2018.1513829
- Wang, Y., Feng, X. N., and Song, X. Y. (2016). Bayesian quantile structural equation models. *Struct. Equ. Model.* 23, 246–258. doi: 10.1080/10705511.2015.1033057
- Wirth, R. J., and Edwards, M. C. (2007). Item factor analysis: current approaches and future directions. *Psychol. Methods* 12:58. doi: 10.1037/1082-989x.12.1.58
- Woods, C. M. (2006). Ramsay-curve item response theory (RC-IRT) to detect and correct for nonnormal latent variables. *Psychol. Methods* 11, 253–270. doi: 10.1037/1082-989x.11.3.253
- Woods, C. M. (2015). “Estimating the latent density in unidimensional IRT to permit non-normality,” in *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment*, eds S. P. Reise and D. A. Revicki (Abingdon: Routledge), 60–84.
- Woods, C. M., and Thissen, D. (2006). Item response theory with estimation of the latent population distribution using spline-based densities. *Psychometrika* 71:281. doi: 10.1007/s11336-004-1175-8
- Wößmann, L. (2005). The effect heterogeneity of central examinations: evidence from TIMSS, TIMSS-Repeat and PISA. *Educ. Econ.* 13, 143–169. doi: 10.1080/09645290500031165

- Yu, K., Lu, Z., and Stander, J. (2003). Quantile regression: applications and current research areas. *J. R. Stat. Soc.* 52, 331–350. doi: 10.1111/1467-9884.00363
- Yu, K., and Moyeed, R. A. (2001). Bayesian quantile regression. *Stat. Probabil. Lett.* 54, 437–447. doi: 10.1016/S0167-7152(01)00124-9
- Yu, K., and Zhang, J. (2005). A three-parameter asymmetric Laplace distribution and its extension. *Commun. Stat. Theory Methods* 34, 1867–1879. doi: 10.1080/03610920500199018
- Zhang, X., Tao, J., Wang, C., and Shi, N. Z. (2019). Bayesian model selection methods for multilevel IRT models: a comparison of five DIC-Based indices. *J. Educ. Meas.* 56, 3–27.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Zhu, Gao and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

Here we give more details about the ALD and the Bayesian QR method. We consider a general QR model given by:

$$y_i = \mathbf{d}_i^T \mathbf{r}_\tau + \varepsilon_i, \quad i = 1, \dots, n, \quad (\text{A1})$$

where y_i and \mathbf{d}_i denote response variable and a 4×1 vector of covariates for the i th observation, \mathbf{r}_τ is a 4×1 vector of QR coefficients under the τ th quantile of y_i , and the τ th quantile of ε_i is assumed to be zero. The QR model does not have a specific likelihood function because the distribution of the error term ε_i is not specified. Yu and Moyeed (2001) first proposed Bayesian QR methods, they assigned each error term an asymmetric Laplace distribution (ALD), and considered MCMC methods for posterior inference.

About the main properties of the $ALD(\mu, \omega, \tau)$, parameters μ and τ in ALD satisfy: μ is the τ th quantile of the distribution, and if $X \sim ALD(0, 1, \tau)$, then $G = \mu + \omega X \sim ALD(\mu, \omega, \tau)$. In particular, the mean and variance of g are given by $E(g) = \mu + \frac{\omega(1-2\tau)}{\tau(1-\tau)}$ and $var(g) = \frac{\omega^2(1-2\tau+2\tau^2)}{(1-\tau)^2\tau^2}$ respectively. These important features of ALD have been generally adopted for the quantile inference (Yu et al., 2003). Further details about ALD can be found in Yu and Zhang (2005).

By assuming the error terms in Eq. (A1) distributed as $ALD(0, \omega, \tau)$, then y_i follows $ALD(\mathbf{d}_i^T \mathbf{r}_\tau, \omega, \tau)$, in this case the likelihood function for the model (i.e., Eq. A1) is:

$$L(\mathbf{r}; y, \tau) = \frac{\tau^n (1-\tau)^n}{\omega^n} \exp \left\{ -\frac{\sum_{i=1}^n \rho_\tau(y_i - \mathbf{d}_i^T \mathbf{r}_\tau)}{\omega} \right\}.$$

Yu and Moyeed (2001) noted that the maximum likelihood estimation of \mathbf{r}_τ is equivalent to the solution of the frequentist approach to the estimation of coefficients, which is to solve the following optimization problem as pointed out by Koenker and Bassett (1978):

$$\min_{\mathbf{r}} \sum_{i=1}^n \rho_\tau(y_i - \mathbf{d}_i^T \mathbf{r}_\tau).$$

As discussed in Yu and Moyeed (2001), the use of ALD for the error terms provides a natural way to deal with the Bayesian QR problem.

Since ALD is not a standard distribution so that it lacks a conjugate prior, direct use of the likelihood above is rather inconvenient for Bayesian inference. In this case, Reed and Yu (2009); Kozumi and Kobayashi (2011) used a mixture representation of ALD to develop the Gibbs sampling algorithm to estimate the parameters in QR. Specifically, if a random variable y follows $ALD(\mu, \omega, \tau)$, then it can be represented as the following form:

$$y = \mu + k_1 e + \sqrt{k_2 \omega e} \zeta, \quad (\text{A2})$$

where $k_1 = \frac{1-2\tau}{\tau(1-\tau)}$, $k_2 = \frac{2}{\tau(1-\tau)}$, e follows exponential distribution with scale parameter ω , and ζ is a standard normal distribution. That is to say, y is augmented by a latent variable e , thus the conditional distribution of y is normal with mean $\mu + k_1 e$ and variance $k_2 \omega e$. It has been shown that the mixture representation provided fully conditional posterior densities and simplified the existing estimation procedures for QR models.

¹A random variable g is said to follow $ALD(\mu, \omega, \tau)$ if it has density $f(g|\mu, \omega, \tau) = \frac{\tau(1-\tau)}{\omega} \exp \left\{ -\rho_\tau \left(\frac{g-\mu}{\omega} \right) \right\}$, where μ is the location parameter, ω is the scale parameter, τ is the skewness parameter, and $\rho_\tau(x) = x(\tau - I(x < 0))$ is called the check function.