Investigating a Weakly Informative Prior for Item Scale Hyperparameters in Hierarchical 3PNO IRT Models

Yanyan Sheng *

Department of Counseling, Quantitative Methods and Special Education, Southern Illinois University, Carbondale, IL, USA

The half-\(t\) family has been suggested for the scale hyperparameter in Bayesian hierarchical modeling. Two parameters define a half-\(t\) distribution: the scale \(s\) and the degree-of-freedom \(\nu\). When \(s\) is set at a finite value that is slightly larger than the actual standard deviation of the parameters, the half-\(t\) prior density can be vaguely informative. This paper focused on such densities, and applied them to the hierarchical three-parameter item response theory (IRT) model. Monte Carlo simulations were carried out to investigate the performance of such specifications in parameter recovery and model comparisons under situations where the actual variability of item parameters varied, and results suggest that the half-\(t\) family does offer advantages over the commonly adopted uniform or inverse-gamma prior density by allowing the variability for item parameters to be either very small or large. A real data example is also provided to further illustrate this.

Keywords: item response theory, Gibbs sampling, three-parameter models, hyperprior, scale hyperparameter, half-\(t\), half-Cauchy, half-normal

1. INTRODUCTION

With current enhanced computational technology and the emergence of Markov chain Monte Carlo (MCMC) simulation techniques (e.g., Chib and Greenberg, 1995), the methodology for parameter estimation with item response theory (IRT) models has rapidly moved to a fully Bayesian approach. One of the many advantages that this approach offers in the simultaneous estimation of both item and person parameters is the flexibility of setting prior distributions for model parameters or hyperparameters. The existing literature in Bayesian statistics (Gelman et al., 2003) offers two general options in specifying prior distributions by choosing between fully informative priors using application-specific information and non-informative priors. Each of these is adopted depending on the availability of prior information. However, when prior information is desired but not readily available, neither would provide a common solution for various actual situations. The problem with the former, in particular, is that misspecification tends to result in biased estimates and hence incorrect inferences (e.g., Mislevy, 1986). This paper focuses on something in the middle, namely, a somewhat informative prior distribution that can be used in a wide range of applications.

The fully Bayesian estimation procedure has been developed for the three-parameter normal ogive (3PNO; Lord, 1980) model by Sahu (2002; see also Johnson and Albert, 1999) generalizing the approach for the two-parameter model by Albert (1992). The procedure has been further implemented in some applications, e.g., Béguin and Glas (2001) and Glas and Meijer (2003). However, this specification where the hyperparameters take specific values causes problems when prior distributions for the item slope
and intercept parameters are not strongly informative. Specifically, studies have shown that improper non-informative prior densities for the item slope and intercept parameters result in an undefined posterior distribution, which gives rise to unstable parameter estimates (Sheng, 2008, 2010). Even with proper non-informative prior densities, the procedure either fails to converge or requires an enormous number of iterations for the Markov chain to reach convergence (Sheng, 2010). Sheng (2013) shows that if one specifies prior distributions for the hyperparameters of the item parameters, instead of setting values for them, the problem can be resolved. This type of hierarchical modeling allows a more objective approach to inference by estimating the parameters of prior distributions from data rather than specifying them using subjective information.

Research on the effect of prior distributions for hyperparameters in Bayesian hierarchical models advises caution in choosing a prior distribution for the scale hyperparameter, as certain specifications of the hyperpriors may cause problems in inference (Brown and Draper, 2006; Gelman, 2006). In the Bayesian literature and software, various non-informative prior distributions have been suggested for the variance parameter in hierarchical linear models, including an improper uniform prior on \( \log(\sigma) \), and a conjugate inverse-gamma (0.001, 0.001) prior (Spiegelhalter et al., 1994, 2003). Gelman (2006), however, in an attempt to illustrate the performance of these prior distributions for \( \sigma \) near zero, which is where classical and Bayesian inferences differ the most (see Brown and Draper, 2006), pointed out problems with the latter two, especially with the inverse-gamma family of non-informative prior distributions. As Gelman (2006) stated, the problem with the uniform prior distribution on \( \log(\sigma) \) is that it results in an improper posterior distribution (p. 521), and the problem with the inverse-gamma family is that inferences are sensitive to the choice of the hyperparameters when low values of \( \sigma \) are possible (p. 522). With respect to the proper non-informative prior, he recommended the use of a half-t family, which is inherently conjugate (see Gelman, 2006, for a detailed illustration) and is preferred over other parametric family for the hyperprior distributions because flat-tailed distributions allow for robust inference (Berger and Berliner, 1986). When the scale of the half-t distribution is set to a finite value that is slightly larger than the actual variability of the parameters, the resulting prior density can be vaguely informative.

In view of the above, the purpose of this study is to develop Bayesian hierarchical 3PNO IRT models with such half-t densities being the item scale hyperpriors and further investigate their performance in estimating model parameters as well as in providing model-data fit under different test situations where the actual variability of item parameter varies.

The remainder of the paper is organized as follows. Section 2 describes the hierarchical 3PNO IRT model and the Gibbs sampling procedure where half-t hyperpriors are assumed for the scale parameters for item slopes and intercepts. Then, two simulation studies were carried out to evaluate the performance of this model specification in parameter recovery as well as to compare it with other model specifications with uniform or inverse-gamma prior densities. The methodology and results of these simulation studies are presented in Sections 3 and 4, respectively. Section 5 gives an example where the model specification under investigation is implemented on a subset of College Basic Academic Subjects Examination (BASE; Osterlind, 1997) English data. Finally, a few summary remarks are provided in Section 6.

2. MODEL AND THE GIBBS SAMPLING PROCEDURE

Before the study is further described, the hierarchical 3PNO model is briefly illustrated. Suppose a test consists of \( k \) binary response items (e.g., multiple-choice items), each measuring a single unified latent trait, \( \theta \). Let \( y = [y_{ij}]_{n \times k} \) denote a matrix of \( n \) responses to the \( k \) items where \( y_{ij} = 1 \) \( (y_{ij} = 0) \) if the \( i \)-th person answers the \( j \)-th item correctly (incorrectly) for \( i = 1, \ldots, n \) and \( j = 1, \ldots, k \).

The probability of person \( i \) obtaining a correct response to item \( j \) can be defined as

\[
P(y_{ij} = 1 | \theta_i, \alpha_j, \beta_j, \gamma_j) = \gamma_j + (1 - \gamma_j) \Phi(\alpha_j \theta_i - \beta_j), \quad 0 \leq \gamma_j < 1
\] (1)

for the 3PNO IRT model, where \( \Phi \) denotes the normal CDF, \( \theta_i \) is a scalar latent trait parameter, \( \alpha_j \) is a scalar slope parameter describing the item discrimination, \( \beta_j \) is the intercept parameter associated with item difficulty, and \( \gamma_j \) is a pseudo-chance-level parameter indicating that the probability of correct response is \( < \) zero even for those with very low trait levels. This model is applicable for objective items, such as multiple-choice or true-or-false items where an item is too difficult for some examinees.

To implement Gibbs sampling to the 3PNO model defined in (1), two latent variables, \( Z \) and \( W \), are introduced such that \( Z_{ij} \sim N(\eta_{ij}, 1) \) (Albert, 1992; Tanner and Wong, 1987), where \( \eta_{ij} = \alpha_j \theta_i - \beta_j \), and \( W_{ij} = 1 \) \( (W_{ij} = 0) \) if person \( i \) knows (does not know) the correct answer to item \( j \) with a probability density function

\[
P(W_{ij} = w_{ij} | \eta_{ij}) = \Phi(\eta_{ij})^{w_{ij}}(1 - \Phi(\eta_{ij}))^{1-w_{ij}}.
\] (2)

Prior densities \( p(\theta) \) of \( (\xi) \) and \( p(\gamma) \) can be assumed for \( \theta_i, \xi_j \), and \( \gamma_j \), respectively, where \( \xi_j = (\alpha_j, \beta_j)' \). Here we focus on the normal conjugate priors for \( \xi_j \) so that \( \alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2), \beta_j \sim N(\mu_\beta, \sigma_\beta^2) \). Further, with hyperparameters assumed for the hyperparameters \( \mu_\alpha, \mu_\beta, \sigma_\alpha^2, \sigma_\beta^2 \), the joint posterior distribution of \( (\theta, \xi, \gamma, Z, W, \xi, \Sigma) \) is hence

\[
p(\theta, \xi, \gamma, Z, W, \mu_\xi, \Sigma_\xi | y) \propto f(y|W, \gamma)p(W|Z)p(Z|\theta, \xi)p(\theta)
p(\xi|\mu_\xi, \Sigma_\xi)p(\gamma)p(\mu_\xi)p(\Sigma_\xi),
\] (3)

where \( \mu_\xi = (\mu_\alpha, \mu_\beta)' \), \( \Sigma_\xi = \text{diag}(\sigma_\alpha^2, \sigma_\beta^2) \), and \( f(y|W, \gamma) = \prod P_{ij}^{y_{ij}}(1 - p_{ij})^{1-y_{ij}} \) is the likelihood function, with \( p_{ij} \) being the model probability function as defined in (1). Assume a normal prior for \( \theta_i \), a conjugate Beta prior for \( \gamma_j \) so that \( \theta_i \sim N(\mu, \sigma^2), \gamma_j \sim Beta(d, t) \), and conditionally conjugate half-t prior
distributions for $\sigma_\alpha$ and $\sigma_\beta$ with mean 0, degrees-of-freedom $\nu$ and scale $s$, where $\nu$ and $s$ are chosen to provide minimal prior information to constrain the scale hyperparameters to lie in a reasonable range. Hence, with the prior distributions specified this way, the full conditional distributions of all the parameters can be derived in closed forms through a multiplicative reparameterization following Knape et al. (2008) and updated iteratively using the Gibbs sampler (see the Appendix in Supplementary Material).

The half-$t$ distribution considered for $\sigma_\alpha$ and $\sigma_\beta$ takes the form

$$p(\sigma) \propto \left(1 + \frac{1}{\nu} \left(\frac{\sigma}{s}\right)^2\right)^{-\frac{\nu+1}{2}}$$

(Gelman, 2006). Setting $\nu = 1$ results in a special case of half-Cauchy, which has a broad peak at zero and can be weakly informative if $s$ takes large but finite values. On the other hand, setting the scale $s$ to infinity corresponds to a flat prior distribution, setting the degrees of freedom $\nu$ to 100 corresponds to a half-normal distribution, which can be non-informative but proper if the scale $s$ is set to a high value, such as 100. Here, in the study, we considered a half-normal, a half-Cauchy, and a half-$t$ distribution with 4 degrees of freedom ($\nu = 4$).

3. METHODOLOGY OF MONTE CARLO SIMULATIONS

In order to evaluate the performance of the hierarchical 3PNO model as described in Section 2, two simulation studies were conducted where it was compared with other model specifications in parameter recovery and/or model comparisons.

3.1. Simulation Study 1

In the IRT literature, it is well accepted that when prior information is not readily available, large data sizes are needed to estimate IRT parameters (e.g., Swaminathan and Gifford, 1983), as Bayesian estimation with a flat prior is equivalent to a maximum likelihood estimation, and that the accuracy of item (person) parameter estimates is related to the number of subjects (items) (e.g., Natesan et al., 2016; Sheng, 2010). Hence, in this simulation study, we evaluated the effect of sample size, test length, actual variability of item slope and intercept parameters on the accuracy with which model parameters are estimated considering a half-normal, a non-informative uniform or an inverse gamma prior.

Item responses for $k$ items ($k = 10, 20, \text{ and } 40$) and $n$ individuals ($n = 100, 300, 500, \text{ and } 1000$) were generated according to the 3PNO model, as defined in (1). Ability parameters were generated as samples from a standard normal distribution, pseudo-chance-level parameters were generated from a uniform distribution, $\gamma_j \sim U(0,0.05,0.4)$, and item slope and intercept parameters were generated as samples from uniform distributions so that

- Sim1: $a_j \sim U(0,2), \beta_j \sim U(0.5,0.5);$
- Sim2: $a_j \sim U(0,2), \beta_j \sim U(0.5,0.5);$
- Sim3: $a_j \sim U(0.5,1.5), \beta_j \sim U(0,1).$

When implementing the MCMC procedure, a diffuse prior were assumed for $\gamma_j, \mu_\alpha$ and $\mu_\beta$ so that $\gamma_j \sim \text{Beta}(1,1), p(\mu_\alpha) \propto 1$ and $p(\mu_\beta) \propto 1$. In addition, three ways of setting the prior distributions for $a_j$ and $\beta_j$ were considered such that both had

1. non-informative prior distributions that are uniform on $\sigma$, i.e., $p(\sigma_\alpha^2) \propto 1/\sigma_\alpha$ and $p(\sigma_\beta^2) \propto 1/\sigma_\beta$;
2. inverse-gamma (0.001, 0.001) prior distributions;
3. half-$t$ prior distributions with 100 degrees of freedom, which are in practice equivalent to a half-normal.

It is noted that although the inverse-gamma (0.001, 0.001) prior density was not suggested by Gelman (2006), it was considered in this study because of its popularity (see e.g., Xu et al., 2009; O’Brien et al., 2015). With each of the prior specifications considered, the Gibbs sampling procedure was implemented where 10,000–50,000 iterations were obtained with the first half set as burn-in.

3.2. Simulation Study 2

In the first simulation study, only two levels of variability (i.e., $\sigma^2_{\alpha,\beta} = 0.083$ and $\sigma^2_{\alpha,\beta} = 0.333$), and three different hyperpriors (one uniform, one inverse-gamma and one half-normal) were considered. Given that the range for item intercept parameters ($\beta_j$) is generally wider than $[-1,1]$ in practice, it would be interesting to see how the weakly informative half-$t$ family performs when the variance for $\beta$ goes beyond 1. Hence, a second simulation study was conducted where two factors were manipulated, namely, actual variability of the item intercept parameters and specifications of prior distributions for the item scale hyperparameters.

Item responses for 20 items and 1000 individuals were generated according to the 3PNO model, as defined in (1). Ability parameters were generated as samples from a standard normal distribution, item slope and pseudo-chance-level parameters were generated from uniform distributions, $a_j \sim U(0,2)$ and $\gamma_j \sim U(0,0.05,0.4)$, and item intercept parameters were generated as samples from uniform distributions so that

- Sim1: $\beta \sim U(-0.5,0.5);$
- Sim2: $\beta \sim U(-1.1);$
- Sim3: $\beta \sim U(-2.2);$ and
- Sim4: $\beta \sim U(-4.4).$

It is noted that in the four simulations, the uniform distributions from which the intercept parameters were sampled from have an increasingly large $\sigma_\beta$ ranging from 0.29–2.31 (with the corresponding variance ranging from 0.083–5.333).

In addition, six ways of setting the prior distributions for $\sigma^2_\alpha$ and $\sigma^2_\beta$ were considered such that both had

1. non-informative prior distributions that are uniform on $\sigma$, i.e., $p(\sigma^2_\alpha) \propto 1/\sigma_\alpha$ and $p(\sigma^2_\beta) \propto 1/\sigma_\beta$;
2. non-informative inverse-gamma (0.001,0.001) prior distributions for $\sigma^2$;
3. informative inverse-gamma (3, 2) prior distributions for $\sigma^2$;
4. weakly informative half-Cauchy prior distributions for $\sigma$. 
5. weakly informative half-$t$ prior distributions for $\sigma$ with $\nu = 100$, which in practice are equivalent to half-normal distributions;

6. weakly informative half-$t$ prior distributions for $\sigma$ with $\nu = 4$.

The Gibbs sampling procedure was implemented where 20,000 iterations were obtained with the first 10,000 as burn-in.

### 3.3. Evaluation Criteria

For both simulation studies, convergence was evaluated using the Gelman-Rubin R (Gelman and Rubin, 1992) statistic. The usual practice is using multiple Markov chains from different starting points. Alternatively, a single chain can be divided into sub-chains so that convergence is assessed by comparing the between and within sub-chain variance (Fox, 2007). Since a single chain is less wasteful in the number of iterations needed, the latter approach was adopted. For each Markov chain, the initial values were set to be $\alpha_j = 1$, $\beta_j = 0$, and $\gamma_j = 0.2$ for all items $j$ and $\theta_i = 0$ for all persons $i$. After discarding the burn-in samples, the chain was then separated into five sub-chains of equal length and the R statistic was calculated following the procedure by Gelman and Rubin (1992). Convergence can also be monitored visually using time series graphs of the simulated sequence, such as the trace plot, the running mean plot, and the autocorrelation plot shown in Figure 1 for one item. It is observed that for this item, the autocorrelations between successive parameter draws became negligible at lags $> 600$, suggesting burn-in for a single chain should not take longer than that (Geyer, 1992). Indeed, the trace plot and the running mean plot both suggest possible convergence within 10,000 iterations. Inspection of such plots has, however, been criticized for being unreliable and unwieldy in the presence of a large number of model parameters (Gelman et al., 2003; Nylander et al., 2008). The R statistic obtained from using a single chain was hence the major approach for assessing convergence in this study.

For each simulated scenario, 25 replications, as recommended by Harwell et al. (1996), were conducted to avoid erroneous results in estimation due to sampling error. The accuracy of item/person parameter estimates was evaluated using the root mean square error (RMSE) and bias. Let $\tau$ denote the true value of a parameter (e.g., $\alpha_j$, $\beta_j$, $\gamma_j$, or $\theta_i$) and $\hat{\tau}_r$ its estimate in the $r$th replication ($r = 1, \ldots, R$). The RMSE is defined as

$$\text{RMSE}_r = \sqrt{\frac{\sum_{r=1}^{R} (\hat{\tau}_r - \tau)^2}{R}},$$

and the bias is defined as

$$\text{bias}_r = \frac{\sum_{r=1}^{R} (\hat{\tau}_r - \tau)}{R}.$$

These quantities were averaged over items/persons to provide summary indices.

In simulation study 2 where six model specifications were compared, the adequacy of the fit of the hierarchical 3PNO model with a given prior density on the simulated data was evaluated using Bayesian deviance. It should be noted that this measure provides a model comparison criterion. Hence, it evaluates the fit of a model in a relative, not absolute, sense. The Bayesian deviance information criterion (DIC) was introduced by Spiegelhalter et al. (2002) who generalized the classical information criteria (e.g., AIC, BIC) to one that is based on the posterior distribution of the deviance. This criterion is defined as $\text{DIC} = \hat{D} + p_D$, where $\hat{D} = \mathbb{E}[-2\log L(y|\theta)]$ is the posterior expectation of the deviance (with $L$ being the likelihood function), and $p_D = \mathbb{E}_{\theta|y}[D(\theta)] - \hat{D} = D(\hat{\theta})$ is the effective number of parameters (Carlin and Louis, 2000). In addition, let $D(\hat{\theta}) = -2\log L(y|\hat{\theta})$, where $\hat{\theta}$ is the posterior mean. To compute Bayesian DIC, MCMC samples of the parameters, $\theta^{(1)}, \ldots, \theta^{(M)}$, can be drawn from the Gibbs sampler, then $\hat{D}$ is approximated as $\hat{D} = \frac{1}{M} \sum_{m=1}^{M} [-2\log \prod_{k=1}^{K} L(y|\theta^{(m)}_k)]$. Generally more complicated models tend to provide better fit. Hence, penalizing for the number of parameters makes DIC a more reasonable measure to use. In this study, the Bayesian deviance estimate was obtained after each implementation and averaged across the 25 replications for each model specification.

### 4. SIMULATION RESULTS

The results of the two simulation studies are presented in this section and described as follows.

#### 4.1. Simulation Study 1

Each implementation of the Gibbs sampler gave rise to Gelman-Rubin R statistics close to 1, indicating possible convergence of the Markov chains within the simulated number of iterations. Hence, the posterior estimates were obtained as the posterior expectations of the Gibbs samples and the average RMSE and bias values for $\alpha_j$, $\beta_j$, $\gamma_j$, and $\theta_i$ are summarized in Tables 1–4. A close examination of these values leads to the following observations:

1. In Sim1 where the actual variances for $\alpha$ and $\beta$ were both 0.333, the half-normal distribution performed relatively less well in recovering item and person parameters than the uniform or inverse-gamma distributions when data sizes ($n \times k$) were $< 10,000$, although it has to be noted that this distribution consistently resulted in smaller bias in estimating $\alpha$ regardless of sample size or test length. The uniform prior density performed similarly as the inverse-gamma prior density, with a slight advantage for data with small sample sizes (i.e., $n = 100$) especially where $k = 10$.

2. Sim2 differed from Sim1 in that the actual variance for $\beta$ decreased to 0.083, a value closer to zero. Hence, we focus on the results in recovering $\beta$ here. From Table 2, it is obvious that the half-normal distribution consistently performed better in recovering $\beta$ than the uniform or inverse-gamma distribution when $k < 40$ or with large sample sizes ($n > 300$) when $k = 40$. Between the two non-informative priors, the uniform prior density performed less well than the inverse-gamma (0.001, 0.001) prior density.

3. Similar to Sim2, Sim3 differed from Sim1 in the actual variance for $\alpha$ being reduced to 0.083, and consequently, the results in recovering $\alpha$ are discussed here. It is noted from Table 1 that the half-normal distribution consistently resulted in recovering $\alpha$ with a relative advantage.

These quantities were averaged over items/persons to provide summary indices.

In simulation study 2 where six model specifications were compared, the adequacy of the fit of the hierarchical 3PNO model with a given prior density on the simulated data was evaluated using Bayesian deviance. It should be noted that this measure provides a model comparison criterion. Hence, it evaluates the fit of a model in a relative, not absolute, sense. The Bayesian deviance information criterion (DIC) was introduced by Spiegelhalter et al. (2002) who generalized the classical information criteria (e.g., AIC, BIC) to one that is based on the posterior distribution of the deviance. This criterion is defined as $\text{DIC} = \hat{D} + p_D$, where $\hat{D} = \mathbb{E}[-2\log L(y|\theta)]$ is the posterior expectation of the deviance (with $L$ being the likelihood function), and $p_D = \mathbb{E}_{\theta|y}[D(\theta)] - \hat{D} = D(\hat{\theta})$ is the effective number of parameters (Carlin and Louis, 2000). In addition, let $D(\hat{\theta}) = -2\log L(y|\hat{\theta})$, where $\hat{\theta}$ is the posterior mean. To compute Bayesian DIC, MCMC samples of the parameters, $\theta^{(1)}, \ldots, \theta^{(M)}$, can be drawn from the Gibbs sampler, then $\hat{D}$ is approximated as $\hat{D} = \frac{1}{M} \sum_{m=1}^{M} [-2\log \prod_{k=1}^{K} L(y|\theta^{(m)}_k)]$. Generally more complicated models tend to provide better fit. Hence, penalizing for the number of parameters makes DIC a more reasonable measure to use. In this study, the Bayesian deviance estimate was obtained after each implementation and averaged across the 25 replications for each model specification.
in smaller RMSE and bias and hence performed better than the non-informative uniform or inverse-gamma family of prior distributions (except for the situations where \( n < 500 \) and \( k = 40 \)). The uniform prior density tended to result in larger RMSE or bias than the inverse-gamma (0.001, 0.001) prior when \( k < 40 \).

4. It is interesting to note that the half-normal prior density consistently resulted in smaller bias in estimating \( \alpha \) in all the simulated conditions.

5. Further, it is noted that the benefit of using a half-normal prior in Sim2 or Sim3 where it resulted in smaller RMSE and bias in estimating \( \beta \) or \( \alpha \) was not reflected in estimating \( \theta \) unless the data size was over 12,000 in Sim2 or unless \( k \geq 20 \) in Sim3.

From these observations, it is hence noted that the half-normal prior outperforms the uniform or inverse-gamma family of prior distributions in estimating the corresponding parameters in the hierarchical 3PNO model when the variability for item slope or intercept parameters is close to zero. Further, the inverse-gamma (0.001, 0.001) prior tends to perform better in parameter recovery than the uniform prior when the respective item parameters have a small variance. This can be explained by the fact that the non-informative uniform prior density is flat and hence does not restrict \( \sigma \) away from large values, whereas inverse-gamma (0.001, 0.001) and half-normal distributions have a peak around zero and do perform such shrinkage for variances near zero. It is further noted that the estimation error and bias in estimating item parameters (\( \alpha \) or \( \beta \)) reduce with the increase of sample sizes, and the error and bias in estimating person parameters (\( \theta \)) reduce with the increase of test lengths. This is consistent with findings from other studies (e.g., Swaminathan and Gifford, 1983; Sheng, 2010; Natesan et al., 2016).

4.2. Simulation Study 2

Each implementation of the Gibbs sampler gave rise to Gelman-Rubin R statistics close to 1, indicating possible convergence within 20,000 iterations. Hence, the posterior estimates were obtained as the posterior expectations of the Gibbs samples and the average RMSE and bias values for \( \alpha_j \), \( \beta_j \), and \( \gamma_j \) are summarized in Table 5. A close examination of these values leads to the following observations:

1. In all four simulations where the actual variability for the intercept parameters (\( \alpha_j \)) ranged between 0.29–2.31, the three weakly informative half-t prior densities, namely, the half-Cauchy, the half-normal and the half-t with 4 degrees of freedom had consistently smaller RMSE if not bias in recovering item and person parameters, and in particular in recovering \( \beta_j \), compared with the other three prior distributions.

2. It is noted that in Sim2 where \( \sigma_\beta \) was about 0.5, the three weakly informative half-t densities had more bias in estimating \( \beta \) and \( \gamma \) than the two non-informative prior densities. When \( \sigma_\beta \) moved further away from 0.5 (i.e., toward 0 in Sim1 or toward 2.5 in Sim4), their advantages over other model specifications became more obvious in that they resulted in much smaller average RMSE and bias values.

3. Among the three weakly informative half-t densities, the half-normal resulted in relatively smaller RMSE and bias in recovering item and person parameters in Sim1, but larger RMSE and bias in Sim2, Sim3, and Sim4.

4. Between the two non-informative priors, the inverse-gamma (0.001, 0.001) prior distribution performed slightly better in estimating \( \beta \) and \( \gamma \) in Sim1, but worse in estimating other parameters or in other situations. This is consistent with findings from the first simulation study.
5. In Sim1 where $\sigma_\beta$ was close to 0, the informative inverse-gamma (3, 2) prior distribution was correctly specified with little bias and hence had small RMSE in recovering $\beta$. However, when $\sigma_\beta$ moved further away from 0, it became less appropriate, and consequently resulted in an obviously larger bias and RMSE in recovering especially item parameters.

From these observations, it is noted that the weakly informative half-$t$ family works well in a wider range of situations than the non-informative or the informative prior densities in recovering the 3PNO model item parameters. Specifically, when the actual
TABLE 3 | Average RMSE and bias for recovering guessing (\(\gamma\)) parameters in the hierarchical 3PNO model with the three prior specifications under the three test length, four sample size and three actual variance conditions.

|          | RMSE        |           |          | bias        |           |
|----------|-------------|-----------|----------|-------------|-----------|
|          | n           | uniform   | inv-g    | half-n      | uniform   | inv-g    | half-n   |
| Sim1: \(\alpha \sim U(0,2), \beta \sim U(-1,1)\) | k = 10 | 100 | 0.0357 | 0.0371 | 0.0407 | 0.0990 | 0.1026 | 0.0964 |
|          | 300 | 0.0306 | 0.0304 | 0.0359 | 0.0848 | 0.0826 | 0.0831 |
|          | 500 | 0.0294 | 0.0290 | 0.0338 | 0.0752 | 0.0733 | 0.0701 |
|          | 1000 | 0.0278 | 0.0271 | 0.0289 | 0.0712 | 0.0677 | 0.0608 |
| Sim2: \(\alpha \sim U(0,2), \beta \sim U(-0.5,0.5)\) | k = 10 | 100 | 0.0443 | 0.0444 | 0.0509 | 0.1111 | 0.1110 | 0.1192 |
|          | 300 | 0.0301 | 0.0300 | 0.0306 | 0.0664 | 0.0652 | 0.0730 |
|          | 500 | 0.0244 | 0.0245 | 0.0239 | 0.0539 | 0.0536 | 0.0631 |
|          | 1000 | 0.0305 | 0.0305 | 0.0284 | 0.0659 | 0.0651 | 0.0728 |
| Sim3: \(\alpha \sim U(0.5,1.5), \beta \sim U(-1,1)\) | k = 10 | 100 | 0.0298 | 0.0286 | 0.0227 | 0.1128 | 0.1082 | 0.0740 |
|          | 300 | 0.0220 | 0.0217 | 0.0148 | 0.0855 | 0.0831 | 0.0425 |
|          | 500 | 0.0190 | 0.0190 | 0.0159 | 0.0718 | 0.0716 | 0.0452 |
|          | 1000 | 0.0158 | 0.0157 | 0.0125 | 0.0553 | 0.0534 | 0.0301 |
| k = 20 | 100 | 0.0195 | 0.0193 | 0.0186 | 0.0786 | 0.0776 | 0.0641 |
|          | 300 | 0.0158 | 0.0158 | 0.0153 | 0.0599 | 0.0588 | 0.0508 |
|          | 500 | 0.0155 | 0.0151 | 0.0130 | 0.0458 | 0.0450 | 0.0373 |
|          | 1000 | 0.0098 | 0.0102 | 0.0080 | 0.0335 | 0.0341 | 0.0264 |
| k = 40 | 100 | 0.0161 | 0.0160 | 0.0172 | 0.0477 | 0.0469 | 0.0498 |
|          | 300 | 0.0117 | 0.0117 | 0.0119 | 0.0206 | 0.0208 | 0.0324 |
|          | 500 | 0.0117 | 0.0118 | 0.0108 | 0.0204 | 0.0206 | 0.0296 |
|          | 1000 | 0.0103 | 0.0103 | 0.0076 | 0.0153 | 0.0153 | 0.0190 |
| Sim4: \(\alpha \sim U(0.5,0,5), \beta \sim U(-0.5,0.5)\) | k = 10 | 100 | 0.0325 | 0.0328 | 0.0318 | 0.0825 | 0.0838 | 0.0812 |
|          | 300 | 0.0164 | 0.0162 | 0.0227 | 0.0605 | 0.0596 | 0.0604 |
|          | 500 | 0.0160 | 0.0169 | 0.0182 | 0.0480 | 0.0482 | 0.0445 |
|          | 1000 | 0.0374 | 0.0377 | 0.0441 | 0.1129 | 0.1132 | 0.1207 |
| k = 20 | 100 | 0.0325 | 0.0236 | 0.0196 | 0.0698 | 0.0700 | 0.0664 |
|          | 300 | 0.0220 | 0.0221 | 0.0181 | 0.0641 | 0.0639 | 0.0613 |
|          | 500 | 0.0157 | 0.0157 | 0.0118 | 0.0418 | 0.0420 | 0.0382 |
|          | 1000 | 0.0359 | 0.0359 | 0.0328 | 0.0908 | 0.0905 | 0.0874 |
| k = 40 | 100 | 0.0246 | 0.0248 | 0.0164 | 0.0541 | 0.0551 | 0.0505 |
|          | 300 | 0.0223 | 0.0224 | 0.0129 | 0.0485 | 0.0483 | 0.0399 |
|          | 1000 | 0.0178 | 0.0179 | 0.0089 | 0.0330 | 0.0336 | 0.0253 |

|          | RMSE        |           |          | bias        |           |
|----------|-------------|-----------|----------|-------------|-----------|
|          | n           | uniform   | inv-g    | half-n      | uniform   | inv-g    | half-n   |
| Sim1: \(\alpha \sim U(0,2), \beta \sim U(-1,1)\) | k = 10 | 0100 | 0.3618 | 0.3626 | 0.3951 | 0.1188 | 0.1208 | 0.1257 |
|          | 300 | 0.2240 | 0.2329 | 0.3331 | 0.1011 | 0.1009 | 0.1041 |
|          | 500 | 0.3045 | 0.3047 | 0.3098 | 0.0952 | 0.0950 | 0.0959 |
|          | 1000 | 0.3003 | 0.3005 | 0.3024 | 0.0908 | 0.0909 | 0.0902 |
| Sim2: \(\alpha \sim U(0,2), \beta \sim U(-0.5,0.5)\) | k = 10 | 0100 | 0.2072 | 0.2071 | 0.2079 | 0.0882 | 0.0882 | 0.0857 |
|          | 300 | 0.1919 | 0.1922 | 0.1985 | 0.0733 | 0.0737 | 0.0706 |
|          | 500 | 0.1495 | 0.1497 | 0.1927 | 0.0703 | 0.0706 | 0.0686 |
| Sim3: \(\alpha \sim U(0.5,1.5), \beta \sim U(-1,1)\) | k = 10 | 0100 | 0.1827 | 0.1827 | 0.1774 | 0.1450 | 0.1468 | 0.1553 |
|          | 300 | 0.1648 | 0.1650 | 0.1412 | 0.1196 | 0.1201 | 0.0766 |
|          | 500 | 0.1483 | 0.1485 | 0.1236 | 0.1140 | 0.1132 | 0.0602 |

The informative inverse-gamma (3.2) prior density does not outperform the weakly informative half-\(t\) distribution because the latter has a better behavior near 0. Figure 2 graphically illustrates this point. The non-informative inverse-gamma prior distribution, in general, is not recommended for large \(\sigma\) values.

inv-g, inverse-gamma; half-n, half-normal.

scale hyperparameter is close to 0, the non-informative prior density does not work well compared with informative or weakly informative prior densities, as prior information helps to restrict \(\sigma_{\beta}\) away from large values. Even with prior information,

Frontiers in Psychology | www.frontiersin.org

February 2017 | Volume 8 | Article 123

7
because it is not non-informative for these values (see Figure 2). On the other hand, prior information has to be correctly specified, as misspecification leads to large bias and hence large estimation error, e.g., the actual $\sigma_{\beta}$ in Sim 4 was clearly not in the range of the inverse-gamma (3, 2) distribution (see Figure 2). When comparing among the three half-$t$ distributions, the half-Cauchy is more likely to allow for occasionally large values than the half-$t$ density with 4 degrees of freedom, or the half-normal, which has the smallest tail (see Figure 2). Hence, if it is known a priori that the scale hyperparameter might be far away from 0, a half-Cauchy or a half-$t$ distribution with a small degrees of freedom is suggested. Otherwise, a half-normal may be considered.

In addition to parameter recovery, the model comparison results in each simulation were averaged over the 25 replications and are summarized in Table 6, which shows the averaged estimates for the posterior expectation of the deviance ($\hat{D}$), the deviance of the posterior expectation [$D(\hat{\theta})$] values, the effective number of parameters ($p_{D}$), and the Bayesian DIC, respectively. The model with the half-Cauchy or half-$t$ prior density shows consistently smaller $\hat{D}$, $D(\hat{\theta})$, and DIC than those with other prior specifications. Since small deviance values indicate better model fit, models with such prior distributions for the scale hyperparameters are shown to provide a better description of the simulated data when the actual $\sigma_{\beta}$ was especially larger than 0.29, even after penalizing for model complexities, i.e., the effective number of parameters.

After a close examination and comparison of the values shown in the table, a few remarks can be drawn from these results:

- In all four simulations where $\sigma_{\beta}$ ranged between 0.29–2.31, Bayesian deviances consistently preferred the models with half-$t$ prior densities, whose deviance values were increasingly smaller compared with other model specifications when $\sigma_{\beta}$ became larger.
- The two non-informative priors performed similarly in describing the simulated data according to Bayesian DIC, which was slightly larger for the model with the inverse-gamma (0.001, 0.001) prior.
- In all the simulated scenarios, the model with the informative inverse-gamma (3, 2) prior distribution had consistently the largest average deviance values and hence is not preferred. One may further note that as $\sigma_{\beta}$ moved further away from 0, the difference in deviations between this model specification and others was increasingly larger.
- Among the three half-$t$ distributions considered, the Bayesian deviance measures did not favor the half-normal prior density.
- It is interesting to note that when $\sigma_{\beta}$ was small with values of e.g., 0–0.6, models with half-$t$ prior densities tended to have smaller effective number of parameters ($p_{D}$) than those with non-informative prior densities. The opposite is true for situations where $\sigma_{\beta}$ was over 1.

In summary, the results in parameter recovery and model comparisons using Bayesian deviances suggest that the hierarchical 3PNO model with weakly informative half-$t$ densities for the item scale hyperparameters does show advantages compared with those with non-informative or informative prior distributions in various test situations where the actual variability ranges from small to large values. When the actual variability of item parameters is not close to 0, the half-Cauchy or the half-$t$ with small degrees of freedom is preferred over the half-normal distribution because of its flexibility in allowing for occasional large variability.

5. AN EXAMPLE WITH CBASE DATA

As an illustration, the hierarchical 3PNO model with the weakly informative half-$t$ hyperprior was implemented to a subset of CBASE English subject data, and further compared with the other model specifications in describing the data.

### 5.1. Method

The overall CBASE exam contains 25 multiple-choice items on English reading/literature. The data used in this study were from college students who took the LP form of CBASE in years 2001 and 2002. After removing those who attempted the exam multiple times and removing missing responses, a sample of 1,200 examinees was randomly selected. To assess the model
goodness-of-fit, hierarchical 3PNO models with the three half-\(t\) prior distributions were compared with each other and further compared with those with uniform and inverse-gamma prior densities described in Section 4.

### 5.2. Results

Each of the six model specifications was implemented to the CBASE English data using the Gibbs sampling procedure, where 20,000 iterations were obtained with the first 10,000 set as burn-in. The Gelman-Rubin R statistics were used to assess convergence and they were found to be around or close to 1, suggesting that stationarity had been reached within the simulated Markov chains for the model. The Bayesian deviance estimates were subsequently obtained for each model specification and the results are summarized in Table 7. Among the six model specifications considered, the ones with half-\(t\) and half-Cauchy prior densities had the smallest DIC and/or expected posterior deviance (\(\bar{D}\)) values. Therefore, the 3PNO model with a half-\(t\) or half-Cauchy hyperprior provided the best description of the data compared with other model specifications, even after penalizing for a large effective number of parameters (e.g., \(p_D = 987.24\) for the half-\(t\), and \(p_D = 993.80\) for the half-Cauchy). On the other hand, with slightly larger deviances, the model with an informative inverse-gamma (3, 2) hyperprior described the data less adequately than those with other prior specifications. The relatively small differences in deviances suggested the actual variability might be close to 0. Furthermore, the model with a half-normal or half-\(t\) hyperprior had a larger \(p_D\) than that with a non-informative prior density. Given the findings from the second simulation study in Section 4, this indicated that the actual standard deviation for the item slope and/or intercept parameters with the data was not larger than 1.

### 6. CONCLUDING REMARKS

The half-\(t\) family offers a good alternative parametric family for the prior distribution of scale hyperparameters in Bayesian hierarchical modeling. This paper adopts it as the hyperprior for item scale parameters in the hierarchical 3PNO IRT model, with the focus being on the weakly informative half-\(t\) distributions. Their utility in parameter estimation and model comparison has been explored and the results show that they do offer advantages over the commonly adopted uniform or inverse-gamma prior density by allowing the variability for item slope and/or intercept parameters to be either very small or large. The weakly informative half-\(t\), and especially the weakly informative half-Cauchy density provides certain level of prior information while it still allows occasional large values. Hence, it overcomes problems resulting from using either a
to ensure the adequacy of the number of Natesan et al. (2016) given its improved (or posterior predictive model) β

| Sim1 | 1 | 2144.16 | 20595.02 | 846.65 | 22288.31 |
|------|---|---------|----------|--------|-----------|
| 2    | 2144.21 | 20597.86 | 846.34 | 22290.55 |
| 3    | 2149.50 | 20651.26 | 848.23 | 22347.73 |
| 4    | 21420.74 | 20589.93 | 830.81 | 22251.54 |
| 5    | 21435.56 | 20598.11 | 836.45 | 22271.01 |
| 6    | 21421.36 | 20589.28 | 832.08 | 22253.44 |

Sim2

| Sim2 | 1 | 20793.37 | 19947.51 | 845.82 | 21639.23 |
|------|---|---------|----------|--------|-----------|
| 2    | 20795.94 | 19960.44 | 845.50 | 21641.43 |
| 3    | 20881.40 | 20045.02 | 836.39 | 21717.79 |
| 4    | 20730.70 | 19888.92 | 841.78 | 21572.48 |
| 5    | 20752.65 | 19906.82 | 845.83 | 21598.49 |
| 6    | 20731.41 | 19888.26 | 843.15 | 21574.56 |

Sim3

| Sim3 | 1 | 18543.51 | 17741.73 | 801.79 | 19345.30 |
|------|---|---------|----------|--------|-----------|
| 2    | 18547.95 | 17746.12 | 801.82 | 19349.77 |
| 3    | 18686.54 | 17908.49 | 778.04 | 19464.58 |
| 4    | 18310.29 | 17473.68 | 836.60 | 19146.89 |
| 5    | 18347.90 | 17514.27 | 833.60 | 19181.54 |
| 6    | 18313.22 | 17477.57 | 835.65 | 19148.87 |

Sim4

| Sim4 | 1 | 15932.74 | 15262.86 | 669.88 | 16602.62 |
|------|---|---------|----------|--------|-----------|
| 2    | 15940.13 | 15270.32 | 669.82 | 16609.95 |
| 3    | 16109.74 | 15489.93 | 619.81 | 16729.55 |
| 4    | 15510.46 | 14766.48 | 743.98 | 16254.44 |
| 5    | 15581.47 | 14850.04 | 731.43 | 16312.89 |
| 6    | 15512.23 | 14773.10 | 739.13 | 16251.36 |

The use of the half-t prior density for the slope/intercept scale hyperparameter has also an effect on estimating person parameters in 3PNO models. Based on results from the first simulation study, we can see that it tends to result in smaller bias and error in estimating them for data sizes > 10,000. For small data sizes, the use of half-normal prior for item scale hyperparameters may not be suggested over the uniform or inverse-gamma (0.001, 0.001) prior if the focus is primarily on estimating person ability parameters. One may need to further explore the use of other half-t prior densities, such as the half-Cauchy, under these conditions. It would also be interesting to find out the reason why manipulating the hyperpriors for item parameters has such an effect on estimating person parameters in the hierarchical 3PNO model, and/or investigate the effect of different (hyper)priors for person parameters in estimating the model.

Further, results and hence conclusion of the second simulation study are based on tests with k = 20 and n = 1000. Although it is believed that similar results can be obtained with other test lengths (e.g., k = 10 or k > 20) and/or larger sample sizes (n > 1000), additional studies are needed to confirm this and to further investigate the use of half-Cauchy or other half-t prior densities for item scale hyperparameters in smaller sample size conditions (e.g., n < 500).

In this study, due to the computational expense of MCMC procedures, only 25 replications were adopted and hence the simulation results are based on a limited number of replications. Future studies shall follow the procedure illustrated by Koehler et al. (2009) to ensure the adequacy of the number of replications. Alternatively, one may consider using variational Bayes as suggested by Natesan et al. (2016) given its improved computational efficiency and equivalent estimation accuracy when compared with MCMC. In addition, only certain prior or hyperprior densities for item slope and intercept scale hyperparameters were investigated in this paper. Further research may include more prior specifications or adopt non-conjugate priors for them. Finally, in this study, only Bayesian deviance was used to evaluate individual models. Given that DIC may be limited in that it is not invariant to parameterization and sometimes can produce unrealistic results, further studies can adopt other methods for model comparisons, such as Bayes factors (Kass and Raftery, 1995) or posterior predictive model checking (Rubin, 1984).

**AUTHOR CONTRIBUTIONS**

YS designed the study, carried out all the analyses and wrote the manuscript.
REFERENCES

Albert, J. H. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *J. Educ. Stat.*, 17, 251–269. doi: 10.2307/1165149

Béguin, A. A., and Glas, C. A. W. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, 66, 541–562. doi: 10.1007/BF02296195

Berger, J. O., and Berliner, L. M. (1986). Robust Bayes and empirical Bayes analysis with epsilon-contaminated priors. *Ann. Stat.*, 14, 461–486. doi: 10.1214/aos/1176349933

Browne, W. J., and Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Anal.*, 1, 473–514. doi: 10.1214/06-BA117A

Carlin, B. P., and Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis* 2nd Edn. London: Chapman & Hall.

Chib, S., and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *Am. Stat.*, 49, 327–335. doi: 10.1080/00031305.1995.10476177

Fox, J.-P. (2007). Multilevel IRT modeling in practice with the package mlirt. *J. Stat. Softw.*, 20, 1–16. doi: 10.18637/jss.v020.i05

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Anal.*, 1, 515–534. doi: 10.1214/06-BA117A

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis*. 2nd Edn. London: Chapman & Hall.

Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Stat. Sci.*, 7, 457–511. doi: 10.1214/ss/1177011136

Geyer, G. J. (1992). Practical markov chain monte carlo. *Stat. Sci.*, 7, 473–483. doi: 10.1214/ss/1177011137

Glas, C. W., and Meijer, R. R. (2003). A Bayesian approach to person fit analysis in item response theory models. *Appl. Psychol. Meas.*, 27, 217–233. doi: 10.1177/0146621603027003003

Harwell, M., Stone, C. A., Hsu, T.-C., and Kirisci, L. (1996). Monte Carlo studies in item response theory. *Appl. Psychol. Meas.*, 20, 101–125. doi: 10.1177/014662169602000201

Johnson, V. E., and Albert, J. H. (1999). *Ordinal Data Modeling*. New York, NY: Springer-Verlag.

Kass, R. E., and Raftery, A. E. (1995). Bayes factors. *J. Am. Stat. Assoc.*, 90, 773–795. doi: 10.1080/01621459.1995.10476572

Knape, J., Sköld, M., Jonzén, N., Åkesson, M., Bensch, S., Hansson, B., and Hasselquist, D. (2008). An analysis of hatching success in the great reed warbler Acrocephalus arundinaceus. *Oikos* 117, 430–438. doi: 10.1111/j.0030-1299.16266.x

Koehler, E., Brown, E., and Haneuse, S. J. (2009). On the assessment of Monte Carlo error in simulation-based statistical analyses. *Am. Stat.*, 63, 155–162. doi: 10.1198/tast.2009.0030

Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika*, 51, 177–195. doi: 10.1007/bf02293979

Nateson, P., NANDAKUMAR, R., MINKA, T., and RUBRIGHT, J. D. (2016). Bayesian prior choice in IRT estimation using MCMC and variational Bayes. *Front. Psychol.* 7:1422. doi: 10.3389/fpsyg.2016.01422

Nylander, J. A., Wilgenbusch, J. C., Warren, D. L., and Swoford, D. L. (2008). AWTY (Are we there yet?): A system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics* 24, 581–583. doi: 10.1093/bioinformatics/btm388

O'Brien, J., Gunawardena, H., Chen, X., Ibrahim, J., and Qaqish, B. (2015). The Midpoint Mixed Model with a Missingness Mechanism (MS): A Likelihood-Based Framework for Quantification of Mass Spectrometry Proteomics Data (Preprint). arXiv preprint arXiv:1507.06907.

Oosterling, S. (1997). *A national review of scholastic achievement in general education: How are we doing and why should we care?*, ASHE-ERIC Higher Education Report, vol. 25, Washington, DC: George Washington University Graduate School of Education and Human Development.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Stat.*, 12, 1151–1172. doi: 10.1214/aos/1176346785

Sahu, S. K. (2002). Bayesian estimation and model choice in item response models. *J. Stat. Comput. Simul.*, 72, 217–232. doi: 10.1080/00949650212387

Sheng, Y. (2010). A sensitivity analysis of Gibbs sampling for 3PNO IRT models: Effect of priors on parameter estimates. *Behaviormetrika* 37, 87–110. doi: 10.2333/bhmk.37.87

Sheng, Y. (2013). An empirical investigation of Bayesian hierarchical modeling with unidimensional IRT models. *Behaviormetrika* 40, 19–40. doi: 10.1007/s10233-010-0126-4

Spiegelhalter, D. J., Best, N., Carlin, B., and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *J. R. Stat. Soc. B* 64, 583–640. doi: 10.1111/j.1467-9868.2002.00353

Spiegelhalter, D. J., Thomas, A., Best, N. G., Gilks, W. R., and Lunn, D. (1999, 2003). *BUGS: Bayesian Inference Using Gibbs Sampling*. Cambridge, England: MRC Biostatistic Unit.

Swaminathan, H., and Gifford, J. A. (1983). "Estimation of parameters in the three-parameter latent trait model," in *New Horizons in Testing*, ed D. Weiss (New York, NY: Academic Press), 13–30.

Tanner, M. A., and Wong, W. H. (1987). The calculation of posterior distribution by data augmentation (with discussion). *J. Am. Stat. Assoc.*, 82, 528–550. doi: 10.1080/01621459.1987.10478458

Xu, L., Johnson, T. D., Nichols, T. E., and Nee, D. E. (2009). Modeling inter-subject variability in fMRI activation location: A Bayesian hierarchical spatial model. *Biometrics* 65, 1041–1051. doi: 10.1111/j.1541-0420.2008.01190.x

**Conflict of Interest Statement:** The author declares 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 © 2017 Sheng. 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) or licensor 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.