PSYCHOMETRIKA—VOL. 82, NO. 4, 1126-1148
DECEMBER 2017
https://doi.org/10.1007/s11336-016-9537-6

Psychometric Society

CrossMark

MODELLING CONDITIONAL DEPENDENCE BETWEEN RESPONSE TIME AND ACCURACY

Maria Bolsinova
Utrecht University,
Tilburg University

Paul de Boeck
Ohio State University

Jesper Tijmstra
Utrecht University,
Tilburg University

Reporter: Yingshi Huang

Introduction

- The benefit of considering RT
  - provide collateral information for **the estimation of ability**
  - shed further light on **the cognitive processes** that led to the observed response

How to model RT and RA data?

$$\Rightarrow f(\mathbf{X}=\mathbf{x}, \mathbf{T}=\mathbf{t} \mid \boldsymbol{\Theta}=\boldsymbol{\theta}, \mathbf{H}=\boldsymbol{\eta})$$

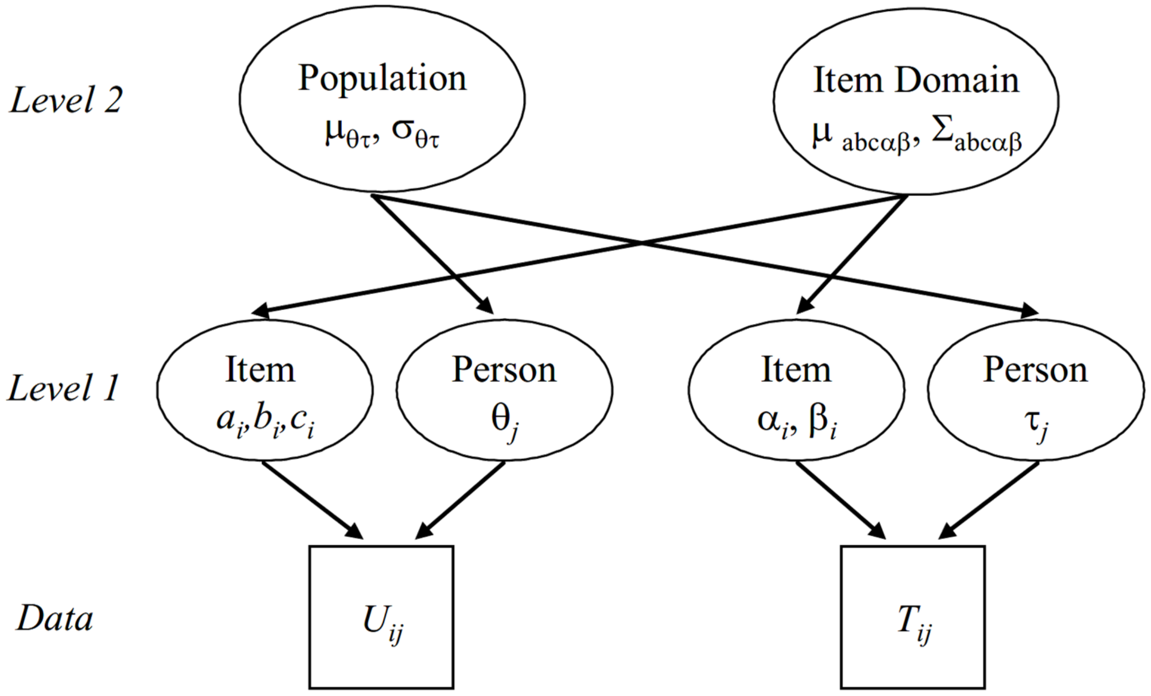
- The assumption of independence
  - standard IRT models: given the ability $\rightarrow$ the RA on different items
  - the lognormal model: given the speed $\rightarrow$ the RT of different items

Introduction

- When considering both RA and RT data

How for each item RT & RA are related ?

$$f(\mathbf{X}, \mathbf{T} \mid \boldsymbol{\theta}, \boldsymbol{\tau}) = \prod_p \prod_i f(t_{pi} \mid \tau_p, \boldsymbol{\gamma}_i) f(x_{pi} \mid \theta_p, \boldsymbol{\delta}_i) \quad \mu_\theta = 0, \quad \sigma_\theta^2 = 1, \quad \mu_\tau = 0$$



- RA model:

$$\Pr(X_{pi} = 1 \mid \theta_p) = \frac{\exp(\alpha_i \theta_p + \beta_i)}{1 + \exp(\alpha_i \theta_p + \beta_i)}$$

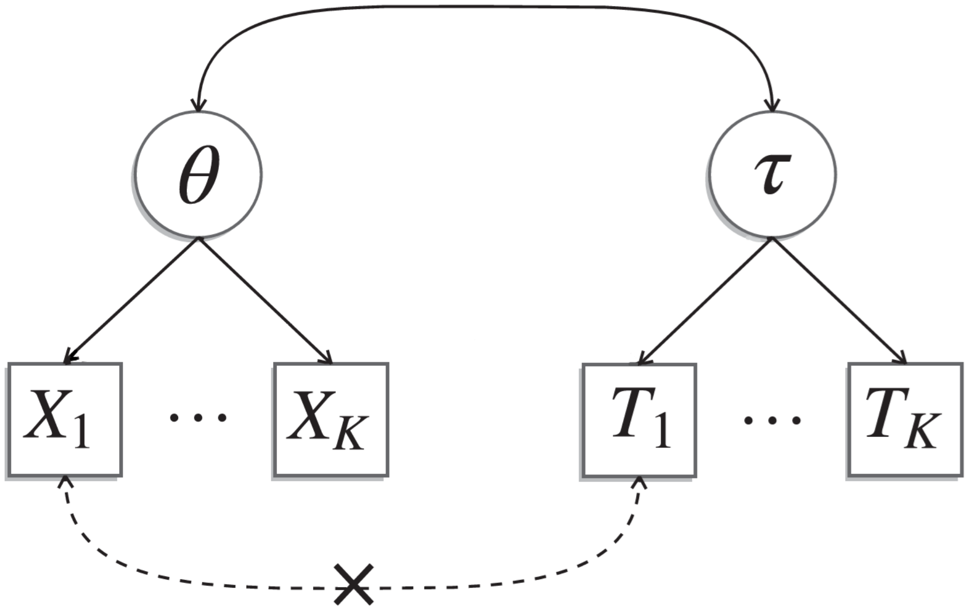
- RT model:

$$T_{pi} \mid \tau_p \sim \ln \mathcal{N}(\xi_i - \tau_p, \sigma_i^2)$$

$$\alpha_i = \frac{1}{\sigma_i^2}$$

Introduction

- Residual associations between RA and RT

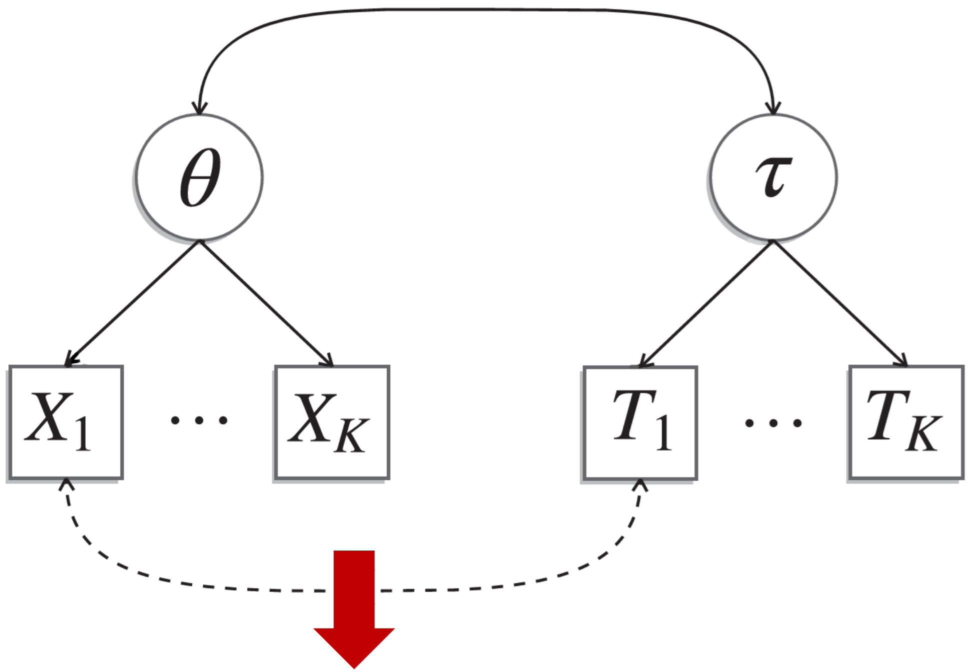


- speed up during the test
- a temporary lapse in concentration
- change problem solving strategies
- …

How to extend the hierarchical modeling framework for RT and RA
to allow for conditional dependence (CD) between the outcome variables?

Introduction

- Conditional Dependence (CD)



1. a bivariate distribution with a nonzero dependence parameter;
2. a marginal distribution of RA and a conditional distribution of RT given RA;
3. a marginal distribution of RT and a conditional distribution of RA given RT.

Introduction

1. A bivariate distribution with a nonzero dependence parameter

$$f(x_i^*, t_i^* \mid \theta, \tau) = \mathcal{N}_2\left(\begin{bmatrix} \alpha_i\theta + \beta_i \\ \xi_i - \tau \end{bmatrix}, \begin{bmatrix} 1 & \rho_i\sigma_i \\ \rho_i\sigma_i & \sigma_i^2 \end{bmatrix}\right)$$

$x_i = \mathcal{I}(x_i^* > 0)$ log-RT

varies across items

2. A marginal distribution of RA and a conditional distribution of RT given RA

$$f(x_i, t_i \mid \boldsymbol{\theta}, \boldsymbol{\eta}) = f(x_i \mid \boldsymbol{\theta}, \boldsymbol{\eta})\, f(t_i \mid x_i, \boldsymbol{\theta}, \boldsymbol{\eta})$$

van der Linden and Glas (2010):
separate time intensity parameters **for the correct and incorrect** responses

Introduction

3. A marginal distribution of RT and a conditional distribution of RA given RT

$$f(x_i, t_i \mid \boldsymbol{\theta}, \boldsymbol{\eta}) = f(t_i \mid \boldsymbol{\theta}, \boldsymbol{\eta}) f(x_i \mid t_i, \boldsymbol{\theta}, \boldsymbol{\eta})$$

depend on whether the response is **relatively fast or slow**

➡ improve the model for response accuracy

investigate the differences in response processes of fast vs slow responses

➡ **Purpose:** model **the effects** of the relative speed of a response **on the parameters of the ICC**

Motivating Example

Why did we choose to model the effect of speed?

How can we get a better model?

- **Source:** the Major Field Test for the Bachelor’s Degree in Business

- **Time limit:** one hour

  (the average time used by the respondents = 42 minutes)

- **The original sample:** 1000 persons to 60 items

  11 items were removed due to low item-rest score correlations (<0.1)

Motivating Example

- To test the assumption of conditional independence:
  - The hierarchical model: $f(x_i, t_i \mid \boldsymbol{\theta}, \boldsymbol{\eta}) = f(x_i \mid \boldsymbol{\theta}, \boldsymbol{\eta}) f(t_i \mid \boldsymbol{\theta}, \boldsymbol{\eta})$

test against (the Lagrange Multiplier test)

  - The approach two: $f(x_i, t_i \mid \boldsymbol{\theta}, \boldsymbol{\eta}) = f(x_i \mid \boldsymbol{\theta}, \boldsymbol{\eta}) \boxed{f(t_i \mid x_i, \boldsymbol{\theta}, \boldsymbol{\eta})}$

$$t_{pi} \sim \ln \mathcal{N}\left(\xi_i + \boxed{\lambda_i (1 - x_{pi})} - \tau_p, \sigma_i^2\right)$$

van der Linden & Glas, 2010 *PSYCHOMETRIKA*

Motivating Example

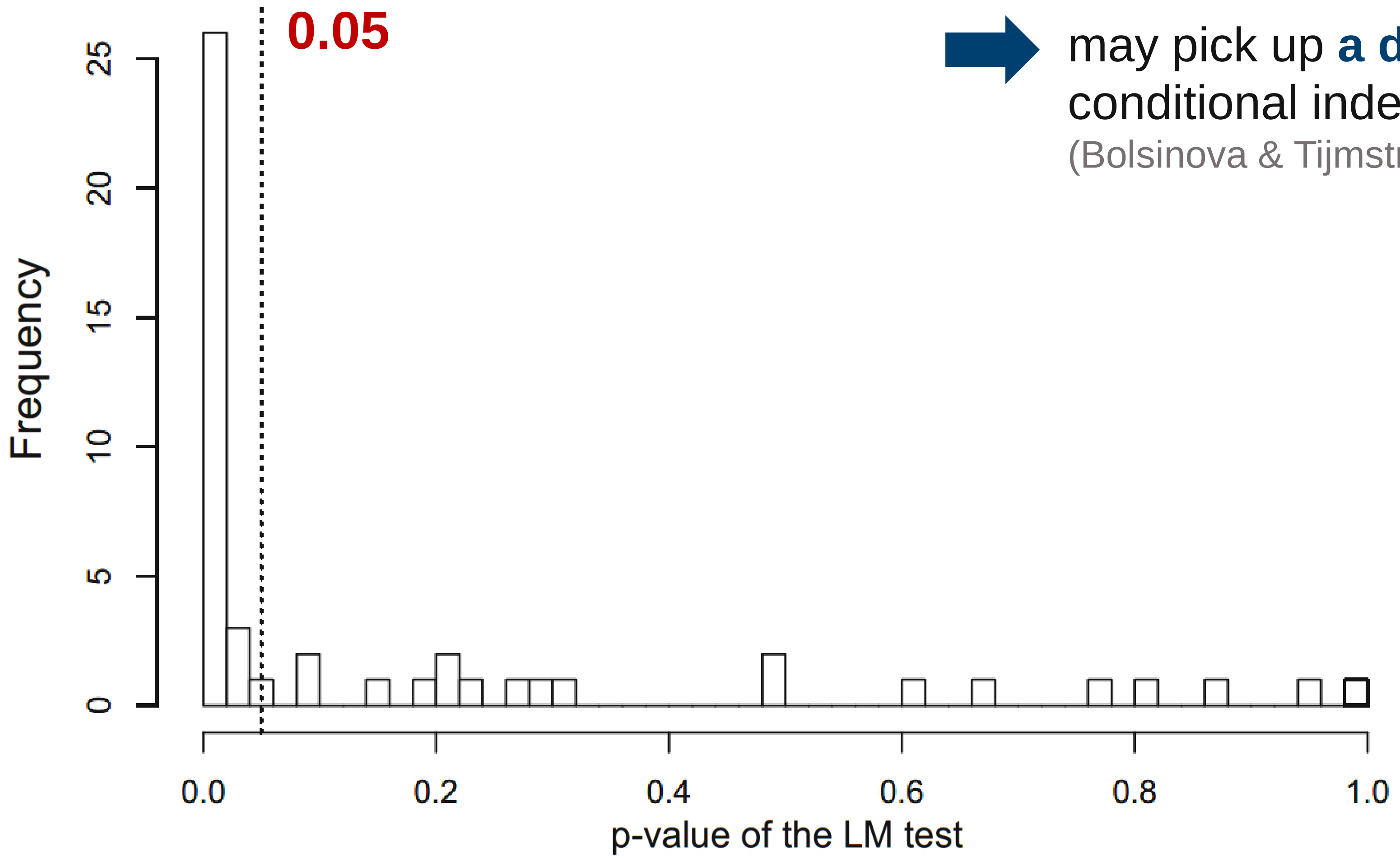


FIGURE 1.
Distribution of the p values of the Lagrange Multiplier test for conditional independence between response time and accuracy. Most of the p values are below .05, indicating that conditional independence is violated.

➡ may pick up **a different type** of conditional independence
(Bolsinova & Tijmstra, 2016)

Motivating Example

Which way conditional independence is violated?

- **difficulty & discriminatory power** between the **slow** and the **fast** responses

➡ can it be observed under the hierarchical model?
(conditional independence)

$$t^*_{pi} = \begin{cases} 1 \text{ if } & t_{pi} \geq t_{med,i} \\ 0 \text{ if } & t_{pi} < t_{med,i} \end{cases}$$

- simple classical test theory statistics
  - difficulty:

$$D_{1i} = \frac{\sum_p x_{pi} t^*_{pi}}{\sum_p t^*_{pi}} - \frac{\sum_p x_{pi}(1 - t^*_{pi})}{\sum_p (1 - t^*_{pi})}$$

  - discriminatory power (item-rest correlation):

$$D_{2i} = Cor\left(\mathbf{x}_{i,slow}, \mathbf{x}^{(i)}_{+,slow}\right) - Cor\left(\mathbf{x}_{i,fast}, \mathbf{x}^{(i)}_{+,fast}\right)$$

Motivating Example

- Posterior predictive check

1. calculated for the **observed data** and for the **hierarchical model**

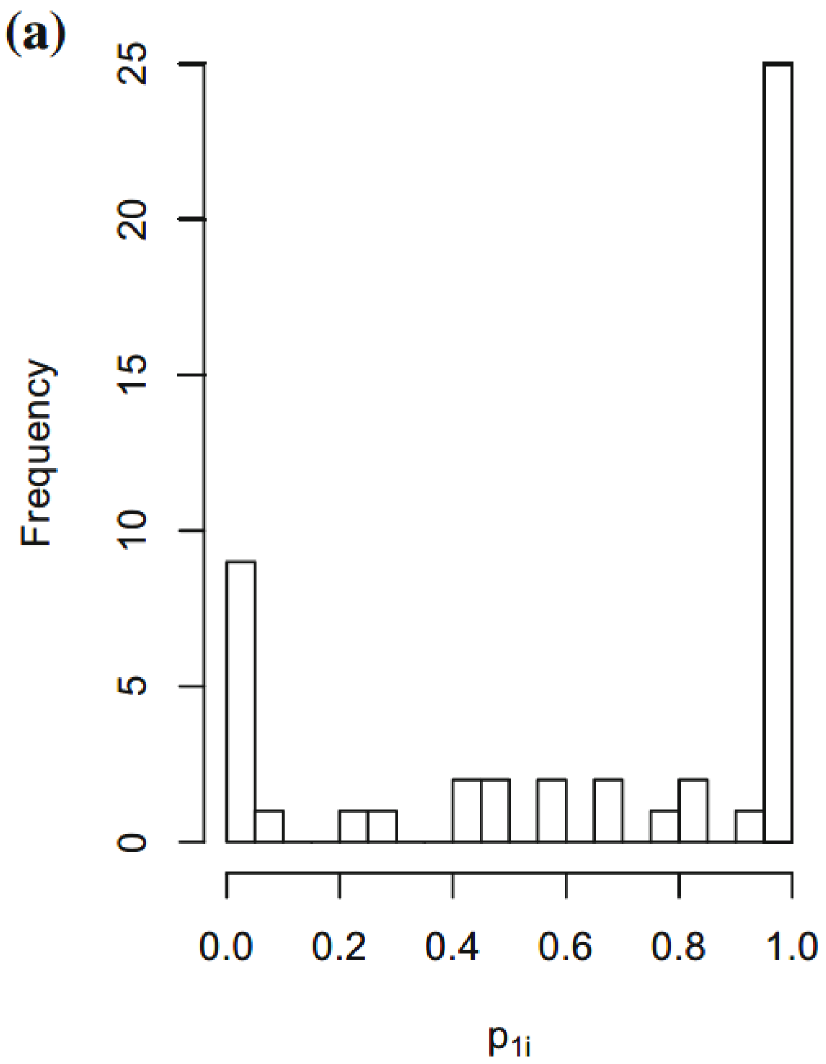
   (G replicated data sets: draws from the posterior distribution)

   $\mathbf{X}_{rep}^{(g)}, \mathbf{T}_{rep}^{(g)} \Rightarrow D_{1i}^{(g)}$ and $D_{2i}^{(g)}$

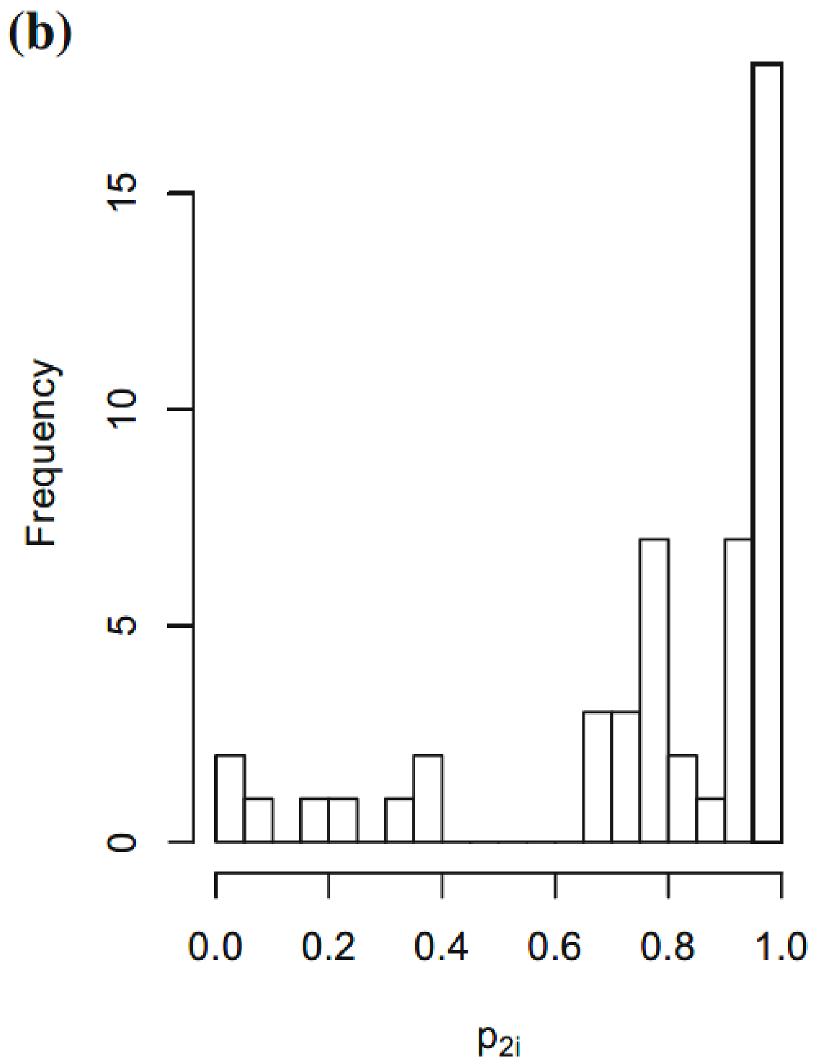
2. p-value: $p_{1i} = \frac{\sum_g \mathcal{I}(D_{1i} > D_{1i}^{(g)})}{2000}$ $\quad p_{2i} = \frac{\sum_g \mathcal{I}(D_{2i} > D_{2i}^{(g)})}{2000}$

3. p-values **close to 0 or close to 1:** not likely under the model

Motivating Example

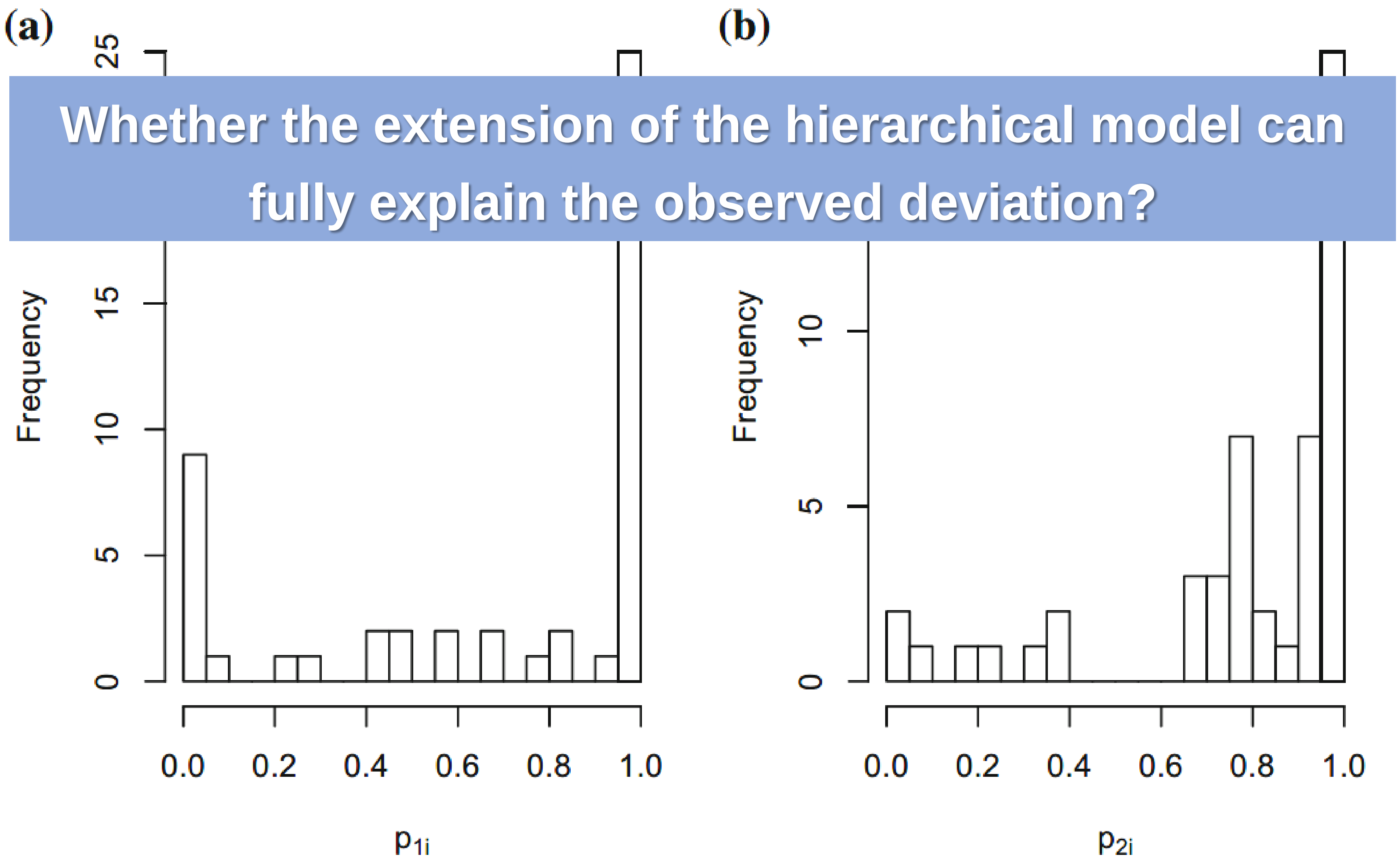


the proportion of correct responses



the item-rest correlations

Motivating Example

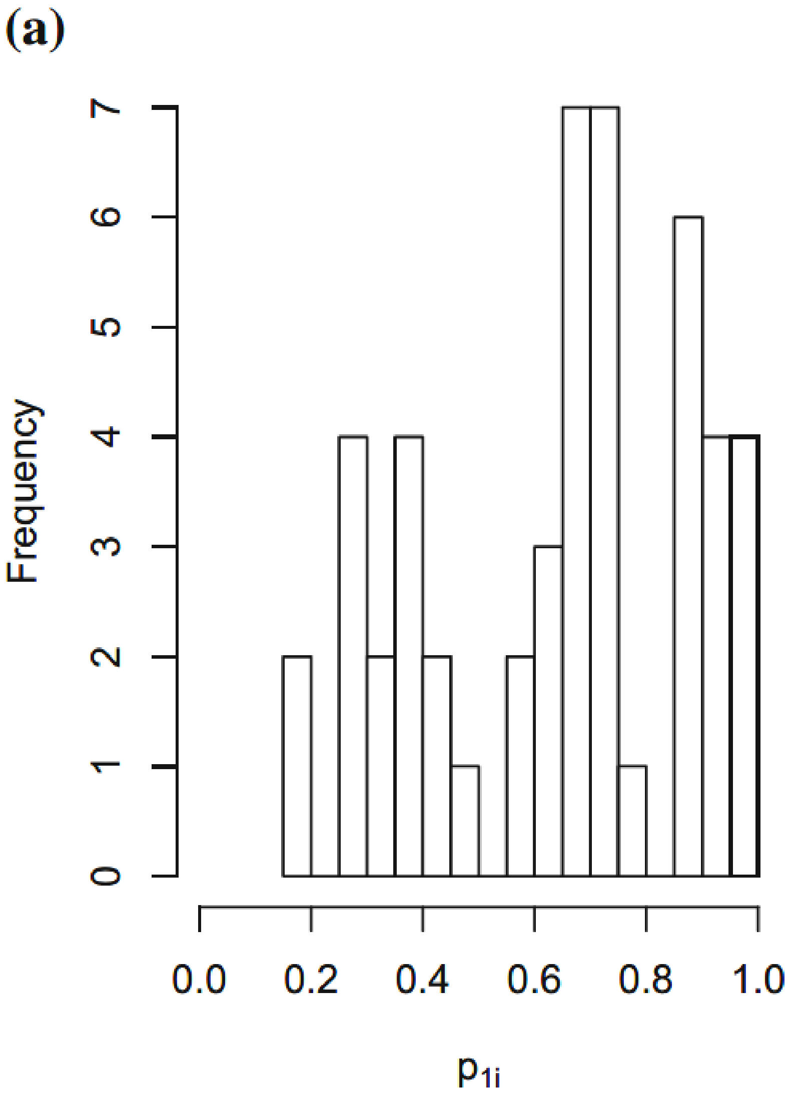


Whether the extension of the hierarchical model can fully explain the observed deviation?

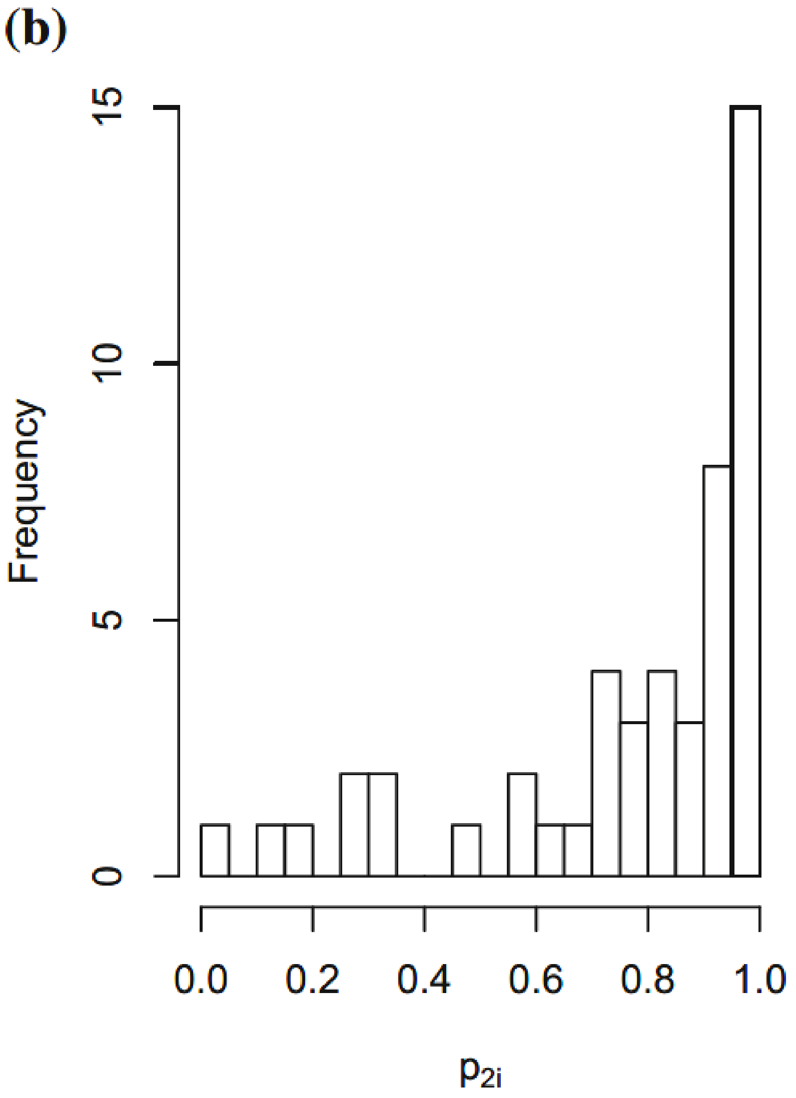
the proportion of correct responses

the item-rest correlations

Motivating Example

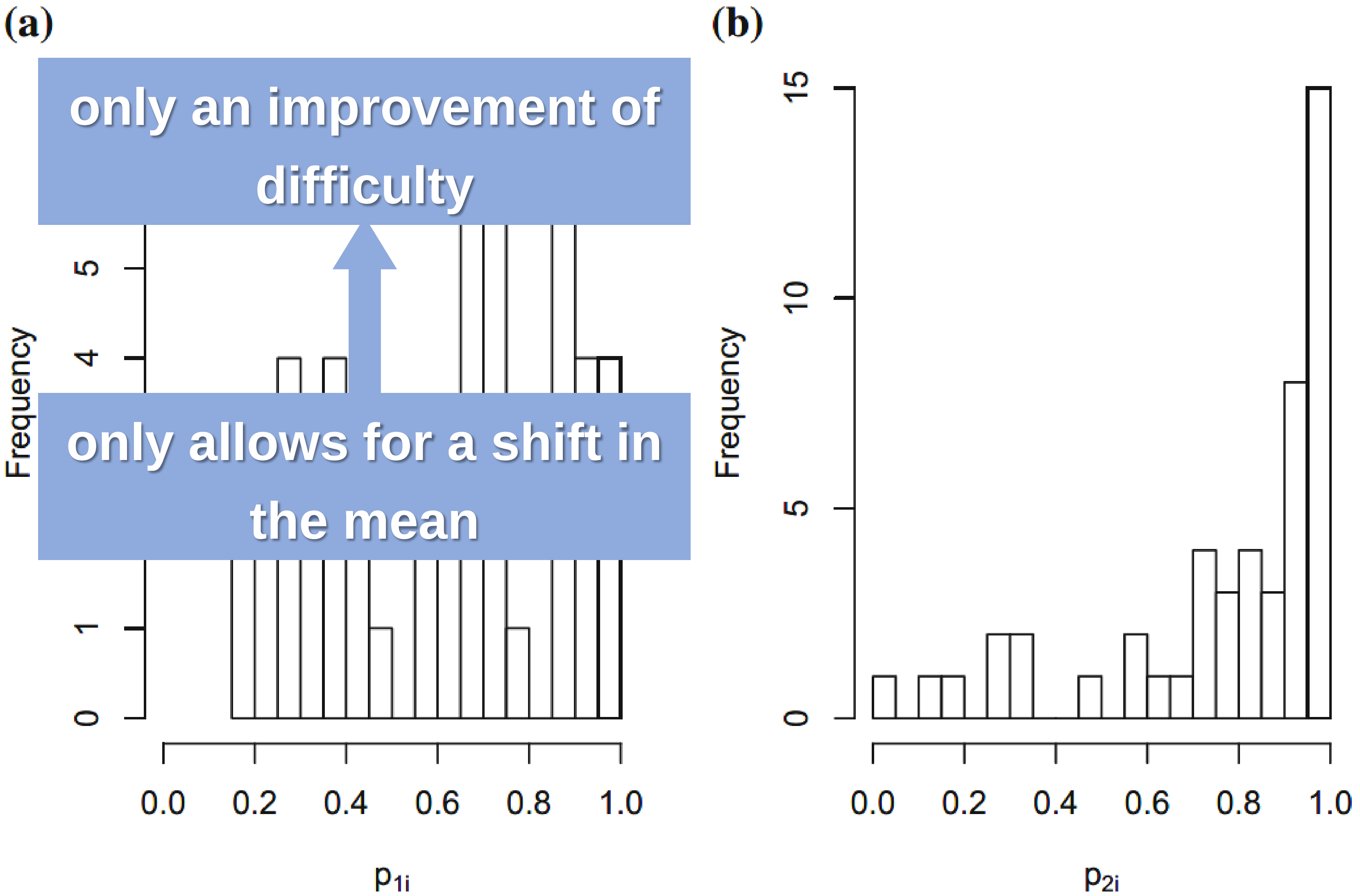


the proportion of correct responses



the item-rest correlations

Motivating Example



the proportion of correct responses

the item-rest correlations

Model Specification

- The goal:
  1. avoid a loss of information: use a **continuous** measure
  2. consider the effect on both **difficulty** and **discriminatory power**

for the first target:

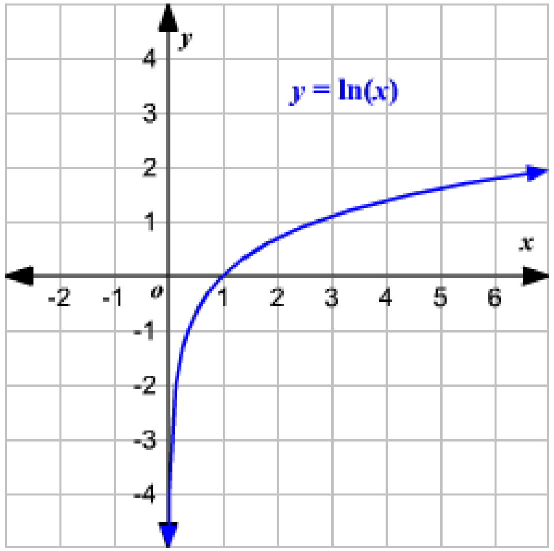
➡ the difference between t_{pi} and the expected response time

$$z_{pi} = \frac{\ln t_{pi} - (\xi_i - \tau_p)}{\sigma_i}$$

for the second target:

➡ a time-related covariate

$$\alpha_{pi} = \alpha_{0i}\alpha_{1i}^{z_{pi}}, \quad \text{or equivalently}$$
$$\ln(\alpha_{pi}) = \ln(\alpha_{0i}) + \ln(\alpha_{1i}) z_{pi}, \quad \text{and}$$
$$\beta_{pi} = \beta_{0i} + \beta_{1i} z_{pi}$$



Model Specification

- The new model for response accuracy:

$$f(x_i|t_i,\theta,\tau) = \frac{\exp\left(\alpha_{0i}\alpha_{1i}^{z_{pi}}\theta + \beta_{0i} + \beta_{1i}z_{pi}\right)}{1+\exp\left(\alpha_{0i}\alpha_{1i}^{z_{pi}}\theta + \beta_{0i} + \beta_{1i}z_{pi}\right)}$$

$$z_{pi} = \frac{\ln t_{pi} - (\xi_i - \tau_p)}{\sigma_i}$$

  - You can define your own **constrained models:**
    - ✓ equal α_{1i} and equal β_{1i} for all items
    - ✓ equal α_{1i} but varying β_{1i} ➡ $f(x_i \mid t_i, \theta, \eta) = \Psi\left(\alpha_i\theta + \beta_{i0} + \beta_{i1}\frac{\ln t_i - (\xi_i - \eta)}{\sigma_i}; x_i\right)$
    - ✓ equal β_{1i} but varying α_{1i}

(Ranger & Ortner, 2012)

Estimation

- sampling from the **joint posterior distribution**

$$f\left(\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\sigma}^2, \boldsymbol{\tau}, \boldsymbol{\mu}_{\mathcal{I}}, \boldsymbol{\Sigma}_{\mathcal{I}}, \sigma_\tau^2, \rho_{\theta\tau} \mid \mathbf{X}, \mathbf{T}\right)$$

✓ point estimate: averages of the sampled values

✓ 95% credible interval: the 2.5% and 97.5% percentiles of the sampled values

– (Posterior)~ (Prior) (Likelihood)

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\xi}, \boldsymbol{\sigma}^2, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\Sigma}_{\mathcal{P}}, \boldsymbol{\mu}_{\mathcal{I}}, \boldsymbol{\Sigma}_{\mathcal{I}} \mid \mathbf{X}, \mathbf{T}) &\propto p(\boldsymbol{\Sigma}_{\mathcal{P}}) p(\boldsymbol{\mu}_{\mathcal{I}}) p(\boldsymbol{\Sigma}_{\mathcal{I}}) \\ &\times \prod_p \mathcal{MVN}(\theta_p, \tau_p; \boldsymbol{\Sigma}_{\mathcal{P}}) \prod_i \frac{1}{\sigma_i^2 \alpha_{0i} \alpha_{1i}} \mathcal{MVN}(\xi_i, \ln \sigma_i^2, \ln \alpha_{0i}, \ln \alpha_{1i}, \beta_{0i}, \beta_{1i}; \boldsymbol{\mu}_{\mathcal{I}}, \boldsymbol{\Sigma}_{\mathcal{I}}) \\ &\times \prod_p \prod_i \frac{1}{t_{pi}\sigma_i} \exp\left(-\frac{(\ln t_{pi} - (\xi_i - \tau_p))^2}{2\sigma_i^2}\right) \frac{\exp(x_{pi}(\alpha_{0i}\alpha_{1i}^{z_{pi}}\theta_p + \beta_{0i} + \beta_{1i} z_{pi}))}{1 + \exp(\alpha_{0i}\alpha_{1i}^{z_{pi}}\theta_p + \beta_{0i} + \beta_{1i} z_{pi})} \end{aligned}$$

Estimation

- sampling from the **joint posterior distribution**

$$f\left(\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\sigma}^2, \boldsymbol{\tau}, \boldsymbol{\mu}_{\mathcal{I}}, \boldsymbol{\Sigma}_{\mathcal{I}}, \sigma_\tau^2, \rho_{\theta\tau} \mid \mathbf{X}, \mathbf{T}\right)$$

✓ point estimate: averages of the sampled values
✓ 95% credible interval: the 2.5% and 97.5% percentiles of the sampled values

– (Posterior)~ (Prior) (Likelihood)

$$p(\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\xi}, \boldsymbol{\sigma}^2, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boxed{\boldsymbol{\Sigma}_{\mathcal{P}}, \boldsymbol{\mu}_{\mathcal{I}}, \boldsymbol{\Sigma}_{\mathcal{I}}} \mid \mathbf{X}, \mathbf{T}) \propto \boxed{p(\boldsymbol{\Sigma}_{\mathcal{P}})\, p(\boldsymbol{\mu}_{\mathcal{I}})\, p(\boldsymbol{\Sigma}_{\mathcal{I}})}$$
$$\times \prod_p \mathcal{MVN}(\theta_p, \tau_p; \boldsymbol{\Sigma}_{\mathcal{P}}) \prod_i \frac{1}{\sigma_i^2 \alpha_{0i} \alpha_{1i}} \mathcal{MVN}(\xi_i, \ln\sigma_i^2, \ln\alpha_{0i}, \ln\alpha_{1i}, \beta_{0i}, \beta_{1i}; \boldsymbol{\mu}_{\mathcal{I}}, \boldsymbol{\Sigma}_{\mathcal{I}})$$
$$\times \prod_p \prod_i \frac{1}{t_{pi}\sigma_i} \exp\left(-\frac{(\ln t_{pi} - (\xi_i - \tau_p))^2}{2\sigma_i^2}\right) \frac{\exp(x_{pi}(\alpha_{0i}\alpha_{1i}^{z_{pi}}\theta_p + \beta_{0i} + \beta_{1i} z_{pi}))}{1 + \exp(\alpha_{0i}\alpha_{1i}^{z_{pi}}\theta_p + \beta_{0i} + \beta_{1i} z_{pi})}$$

Estimation

- sampling from the **joint posterior distribution**

$$f\left(\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\sigma}^2, \boldsymbol{\tau}, \boldsymbol{\mu}_{\mathcal{I}}, \boldsymbol{\Sigma}_{\mathcal{I}}, \sigma_\tau^2, \rho_{\theta\tau} \,|\, \mathbf{X}, \mathbf{T}\right)$$

✓ point estimate: averages of the sampled values
✓ 95% credible interval: the 2.5% and 97.5% percentiles of the sampled values

– (Posterior)~ (Prior) (Likelihood)

$$p(\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\xi}, \boldsymbol{\sigma}^2, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\Sigma}_{\mathcal{P}}, \boldsymbol{\mu}_{\mathcal{I}}, \boldsymbol{\Sigma}_{\mathcal{I}} \,|\, \mathbf{X}, \mathbf{T}) \propto p(\boldsymbol{\Sigma}_{\mathcal{P}})p(\boldsymbol{\mu}_{\mathcal{I}})p(\boldsymbol{\Sigma}_{\mathcal{I}})$$
$$\times \prod_p \mathcal{MVN}(\theta_p, \tau_p; \boldsymbol{\Sigma}_{\mathcal{P}}) \prod_i \frac{1}{\sigma_i^2 \alpha_{0i}\alpha_{1i}} \mathcal{MVN}(\xi_i, \ln\sigma_i^2, \ln\alpha_{0i}, \ln\alpha_{1i}, \beta_{0i}, \beta_{1i}; \boldsymbol{\mu}_{\mathcal{I}}, \boldsymbol{\Sigma}_{\mathcal{I}})$$
$$\times \prod_p \prod_i \frac{1}{t_{pi}\sigma_i} \exp\left(-\frac{(\ln t_{pi} - (\xi_i - \tau_p))^2}{2\sigma_i^2}\right) \frac{\exp(x_{pi}(\alpha_{0i}\alpha_{1i}^{z_{pi}}\theta_p + \beta_{0i} + \beta_{1i}z_{pi}))}{1 + \exp(\alpha_{0i}\alpha_{1i}^{z_{pi}}\theta_p + \beta_{0i} + \beta_{1i}z_{pi})}$$

Estimation

- sampling from the **joint posterior distribution**

$$f\left(\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\sigma}^2, \boldsymbol{\tau}, \boldsymbol{\mu}_{\mathcal{I}}, \boldsymbol{\Sigma}_{\mathcal{I}}, \sigma_\tau^2, \rho_{\theta\tau} \mid \mathbf{X}, \mathbf{T}\right)$$

  ✓ point estimate: averages of the sampled values
  ✓ 95% credible interval: the 2.5% and 97.5% percentiles of the sampled values

– (Posterior)~ (Prior) (Likelihood)

$$p(\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\xi}, \boldsymbol{\sigma}^2, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\Sigma}_{\mathcal{P}}, \boldsymbol{\mu}_{\mathcal{I}}, \boldsymbol{\Sigma}_{\mathcal{I}} \mid \mathbf{X}, \mathbf{T}) \propto p(\boldsymbol{\Sigma}_{\mathcal{P}}) p(\boldsymbol{\mu}_{\mathcal{I}}) p(\boldsymbol{\Sigma}_{\mathcal{I}})$$
$$\times \prod_p \mathcal{MVN}(\theta_p, \tau_p; \boldsymbol{\Sigma}_{\mathcal{P}}) \prod_i \frac{1}{\sigma_i^2 \alpha_{0i} \alpha_{1i}} \mathcal{MVN}(\xi_i, \ln\sigma_i^2, \ln\alpha_{0i}, \ln\alpha_{1i}, \beta_{0i}, \beta_{1i}; \boldsymbol{\mu}_{\mathcal{I}}, \boldsymbol{\Sigma}_{\mathcal{I}})$$
$$\times \prod_p \prod_i \frac{1}{t_{pi}\sigma_i} \exp\left(-\frac{(\ln t_{pi} - (\xi_i - \tau_p))^2}{2\sigma_i^2}\right) \frac{\exp(x_{pi}(\alpha_{0i}\alpha_{1i}^{z_{pi}}\theta_p + \beta_{0i} + \beta_{1i}z_{pi}))}{1 + \exp(\alpha_{0i}\alpha_{1i}^{z_{pi}}\theta_p + \beta_{0i} + \beta_{1i}z_{pi})}$$

Estimation

- sampling from the **joint posterior distribution**

$$f\left(\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\sigma}^2, \boldsymbol{\tau}, \boldsymbol{\mu}_{\mathcal{I}}, \boldsymbol{\Sigma}_{\mathcal{I}}, \sigma_\tau^2, \rho_{\theta\tau} \mid \mathbf{X}, \mathbf{T}\right)$$

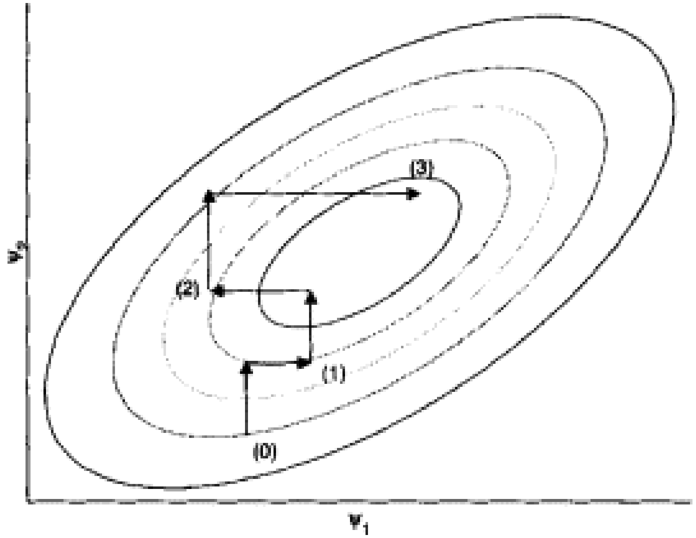
✓ point estimate: averages of the sampled values
✓ 95% credible interval: the 2.5% and 97.5% percentiles of the sampled values

– (Posterior)~ (Prior) (Likelihood)

$$p(\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\xi}, \boldsymbol{\sigma}^2, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\Sigma}_{\mathcal{P}}, \boldsymbol{\mu}_{\mathcal{I}}, \boldsymbol{\Sigma}_{\mathcal{I}} \mid \mathbf{X}, \mathbf{T}) \propto p(\boldsymbol{\Sigma}_{\mathcal{P}})p(\boldsymbol{\mu}_{\mathcal{I}})p(\boldsymbol{\Sigma}_{\mathcal{I}})$$
$$\times \prod_p \mathcal{MVN}(\theta_p, \tau_p; \boldsymbol{\Sigma}_{\mathcal{P}}) \prod_i \frac{1}{\sigma_i^2 \alpha_{0i}\alpha_{1i}} \mathcal{MVN}(\xi_i, \ln\sigma_i^2, \ln\alpha_{0i}, \ln\alpha_{1i}, \beta_{0i}, \beta_{1i}; \boldsymbol{\mu}_{\mathcal{I}}, \boldsymbol{\Sigma}_{\mathcal{I}})$$
$$\times \prod_p \prod_i \frac{1}{t_{pi}\sigma_i} \exp\left(-\frac{(\ln t_{pi} - (\xi_i - \tau_p))^2}{2\sigma_i^2}\right) \frac{\exp(x_{pi}(\alpha_{0i}\alpha_{1i}^{z_{pi}}\theta_p + \beta_{0i} + \beta_{1i}z_{pi}))}{1 + \exp(\alpha_{0i}\alpha_{1i}^{z_{pi}}\theta_p + \beta_{0i} + \beta_{1i}z_{pi})}$$

Estimation

- Metropolis–Hastings algorithm within Gibbs sampler



[Chib, 2001 *Handbook of Econometrics*]

Step 1: speed parameter

$$p(\tau_p \mid \ldots) \propto p(\tau_p \mid \boldsymbol{\Sigma}_{\mathcal{P}}, \theta_p) f(\mathbf{T}_p \mid \tau_p, \ldots) f(\mathbf{X}_p \mid \tau_p, \ldots)$$

- Metropolis–Hastings algorithm
  - ✓ candidate value drawn from

$$\tau^* \sim \mathcal{N}\left(\frac{\sum_i \frac{(\xi_i - \ln t_{pi})}{\sigma_i^2} + \frac{\sigma_\tau \rho_{\theta\tau} \theta_p}{(1-\rho_{\theta\tau}^2)\sigma_\tau^2}}{\sum_i \frac{1}{\sigma_i^2} + \frac{1}{(1-\rho_{\theta\tau}^2)\sigma_\tau^2}}, \frac{1}{\sum_i \frac{1}{\sigma_i^2} + \frac{1}{(1-\rho_{\theta\tau}^2)\sigma_\tau^2}} \right)$$

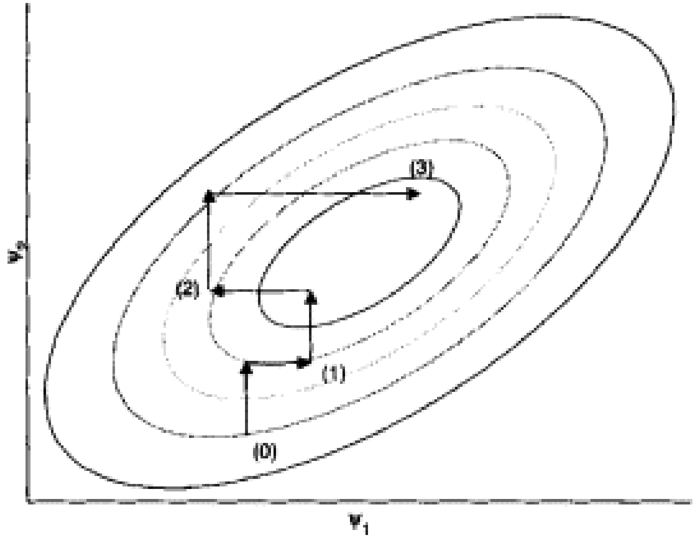
  - ✓ acceptance ratio

$$\Pr(\tau_p \to \tau^*) = \min\left(1, \frac{f(\mathbf{X}_p \mid \tau^*, \ldots)}{f(\mathbf{X}_p \mid \tau_p, \ldots)}\right)$$

Is it the **high density area** of my target distribution?

Estimation

- Metropolis–Hastings algorithm within Gibbs sampler



[Chib, 2001 *Handbook of Econometrics*]

Step 1: speed parameter

$$p(\tau_p \mid \ldots) \propto p(\tau_p \mid \boldsymbol{\Sigma}_{\mathcal{P}}, \theta_p) f(\mathbf{T}_p \mid \tau_p, \ldots) f(\mathbf{X}_p \mid \tau_p, \ldots)$$

Step 2: time intensity parameter

$$p(\xi_i \mid \ldots) \propto p(\xi_i \mid \boldsymbol{\mu}_{\mathcal{I}}, \boldsymbol{\Sigma}_{\mathcal{I}}, \sigma_i^2, \alpha_{0i}, \alpha_{1i}, \beta_{0i}, \beta_{1i}) f(\mathbf{T}_i \mid \xi_i, \ldots) f(\mathbf{X}_i \mid \xi_i, \ldots)$$

$\vdots$

Step 9: re-scale model parameters

$$\begin{aligned} \theta_p &\to \frac{\theta_p}{\sigma_\theta}, & \forall p \in [1:N]; \\ \alpha_{0i} &\to \alpha_{0i}\sigma_\theta, & \forall i \in [1:n]; \\ \mu_{\ln \alpha_0} &\to \mu_{\ln \alpha_0} + \ln \sigma_\theta; & \\ \boldsymbol{\Sigma}_{\mathcal{P}} &\to \begin{bmatrix} 1 & \rho_{\theta\tau} \\ \rho_{\theta\tau} & \sigma_\tau^2 \end{bmatrix}. & \end{aligned}$$

Model Selection and Goodness-of-Fit

- To select the best model:
  - the deviance information criterion [DIC]

1. for each iteration in Gibbs sampling:

$$D^{(g)} = -2\ln\left(f\left(\mathbf{X},\mathbf{T}\,|\,\boldsymbol{\alpha}_0^{(g)},\boldsymbol{\alpha}_1^{(g)},\boldsymbol{\beta}_0^{(g)},\boldsymbol{\beta}_1^{(g)},\boldsymbol{\theta}^{(g)},\boldsymbol{\xi}^{(g)},\boldsymbol{\sigma}^{2(g)},\boldsymbol{\tau}^{(g)}\right)\right)$$

2. for the posterior mean:

$$\hat{D} = -2\ln\left(f\left(\mathbf{X},\mathbf{T}\,|\,\hat{\boldsymbol{\alpha}}_0,\hat{\boldsymbol{\alpha}}_1,\hat{\boldsymbol{\beta}}_0,\hat{\boldsymbol{\beta}}_1,\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\xi}},\hat{\boldsymbol{\sigma}}^2,\hat{\boldsymbol{\tau}}\right)\right)$$

3. the number of effective parameters: $p_D = \left(\frac{\sum_g D^{(g)}}{G} - \hat{D}\right)$

➡ $$DIC = \frac{\sum_g D^{(g)}}{G} + p_D$$

Model Selection and Goodness-of-Fit

- To evaluate the absolute fit:
  - for the **global discrepancy** measure (the log-likelihood)
    - ✓computer for **the observed data**

$$LL_{obs}^{(g)} = \ln\left(f\left(\mathbf{X}, \mathbf{T} \mid \boldsymbol{\alpha_0}^{(g)}, \boldsymbol{\alpha}_1^{(g)}, \boldsymbol{\beta}_0^{(g)}, \boldsymbol{\beta}_1^{(g)}, \boldsymbol{\theta}^{(g)}, \boldsymbol{\xi}^{(g)}, \boldsymbol{\sigma}^{2(g)}, \boldsymbol{\tau}^{(g)}\right)\right)$$

    - ✓computer for a replicated dataset **simulated under the model**

$$LL_{rep}^{(g)} = \ln\left(f\left(\mathbf{X}_{rep}^{(g)}, \mathbf{T}_{rep}^{(g)} \mid \boldsymbol{\alpha_0}^{(g)}, \boldsymbol{\alpha}_1^{(g)}, \boldsymbol{\beta}_0^{(g)}, \boldsymbol{\beta}_1^{(g)}, \boldsymbol{\theta}^{(g)}, \boldsymbol{\xi}^{(g)}, \boldsymbol{\sigma}^{2(g)}, \boldsymbol{\tau}^{(g)}\right)\right)$$

➡ **p value:** the proportion of samples in which observed data are less likely under the model than the replicated data

small p value: the data are unlikely under the model

  - **Posterior predictive checks**: D_{1i} and D_{2i} statistics

Which model is the best fitted model?

- Fitted Models

$\ln \mathcal{N}\left(\xi_i + \lambda_i(1 - x_{pi}) - \tau_p, \sigma_i^2\right)$ → Model with extra λ_i

$\dfrac{\exp\left(\alpha_{0i}\alpha_{1i}^{z_{pi}}\theta + \beta_{0i} + \beta_{1i}z_{pi}\right)}{1 + \exp\left(\alpha_{0i}\alpha_{1i}^{z_{pi}}\theta + \beta_{0i} + \beta_{1i}z_{pi}\right)}$ → z_{pi} as a covariate

Model	
Conditional independence model	
Model with extra λ_i	
z_{pi} as a covariate	Equal α_1 and β_1
	Equal α_1
	Equal β_1
	Full model
$\ln(t_{pi})$ as a covariate	Full model
t_{pi} as a covariate	Full model
t^*_{pi} as a covariate	Full model

- Convergence
  - $\hat{R}$-statistic: the hyper-parameters
  - the multivariate scale reduction factor: overall

→ all fitted models were smaller than 1.1

Which model is the best fitted model?

- Model Selection

TABLE 1.
DIC of the fitted models.

Model		DIC
Conditional independence model		4,66,624.5
Model with extra λ_i		4,65,498.5
z_{pi} as a covariate	Equal α_1 and β_1	4,66,280.4
	Equal α_1	4,65,550.5
	Equal β_1	4,66,100.3
	Full model	4,65,452.7
$\ln(t_{pi})$ as a covariate	Full model	4,65,605.9
t_{pi} as a covariate	Full model	4,65,853.2
t_{pi}^* as a covariate	Full model	4,65,932.4

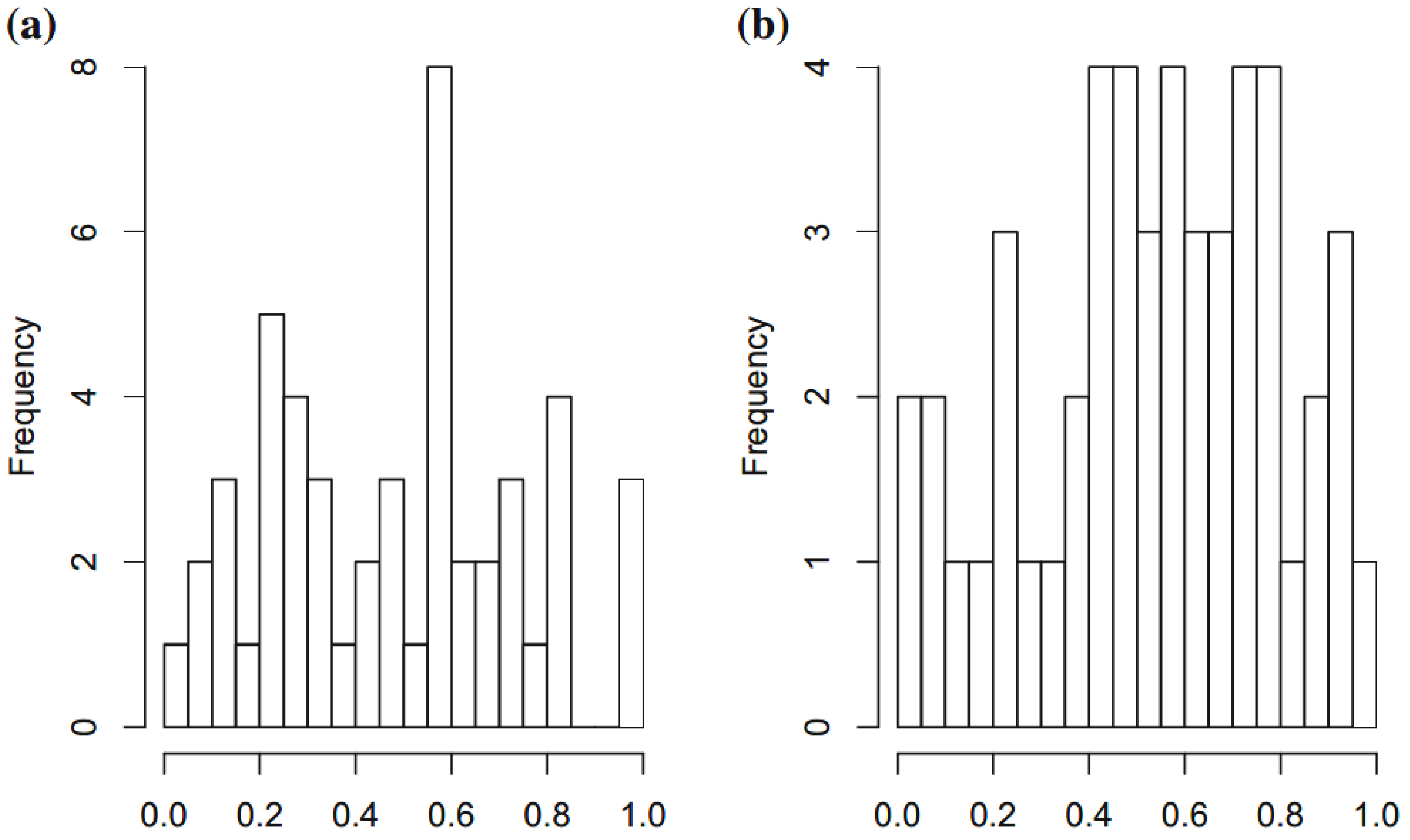
How about its goodness-of-fit?

- for the global discrepancy measure
  - posterior predictive $p = 0.35$

→ providing support for the general fit

How about its goodness-of-fit?

- for the posterior predictive p values



the proportion of correct responses

the item-rest correlations

What kind of effect of residual RT?

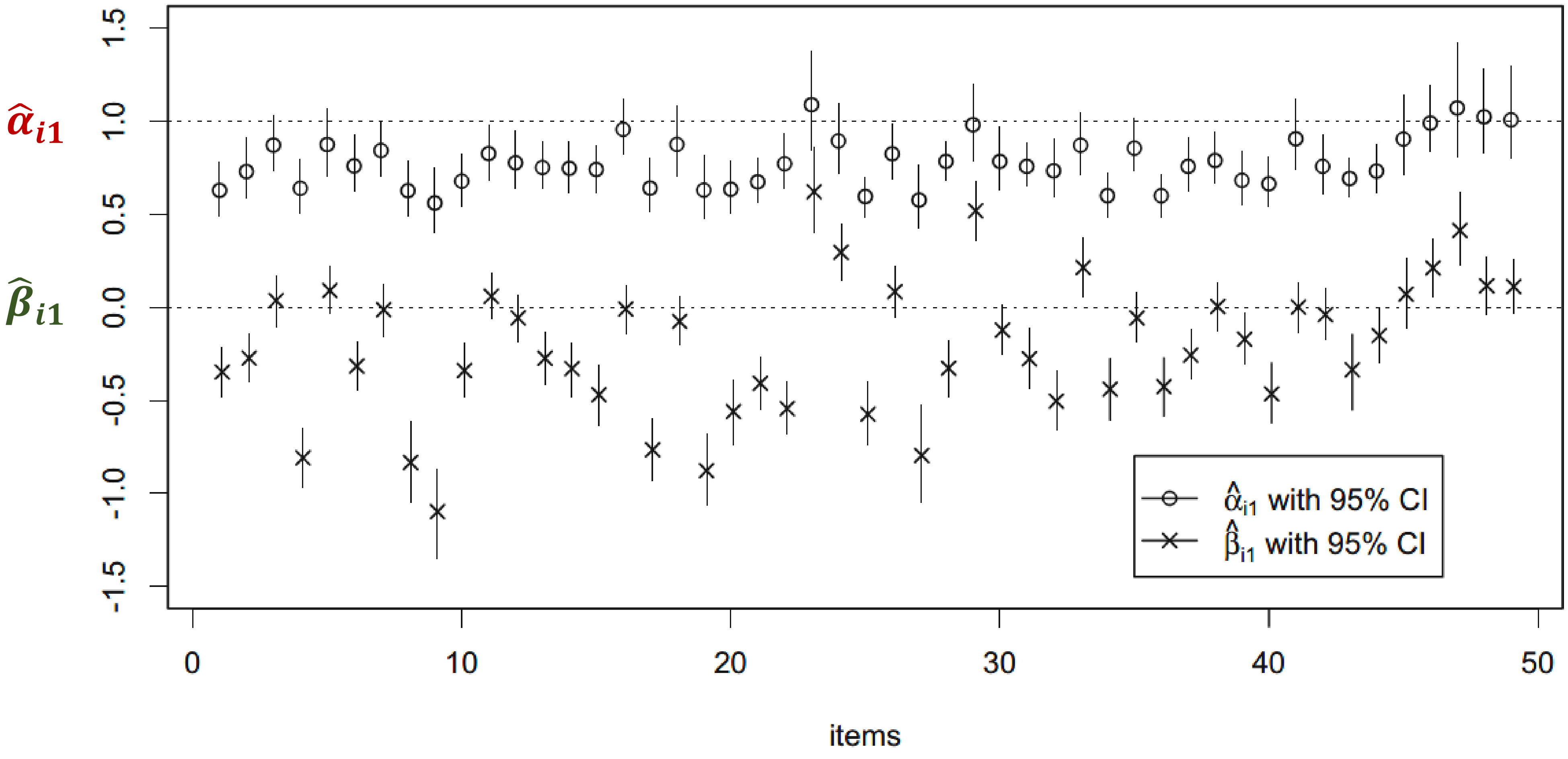


FIGURE 5.
Estimated effects of residual response time on the slope and the intercept of the ICC.

What kind of differences between fast and slow?

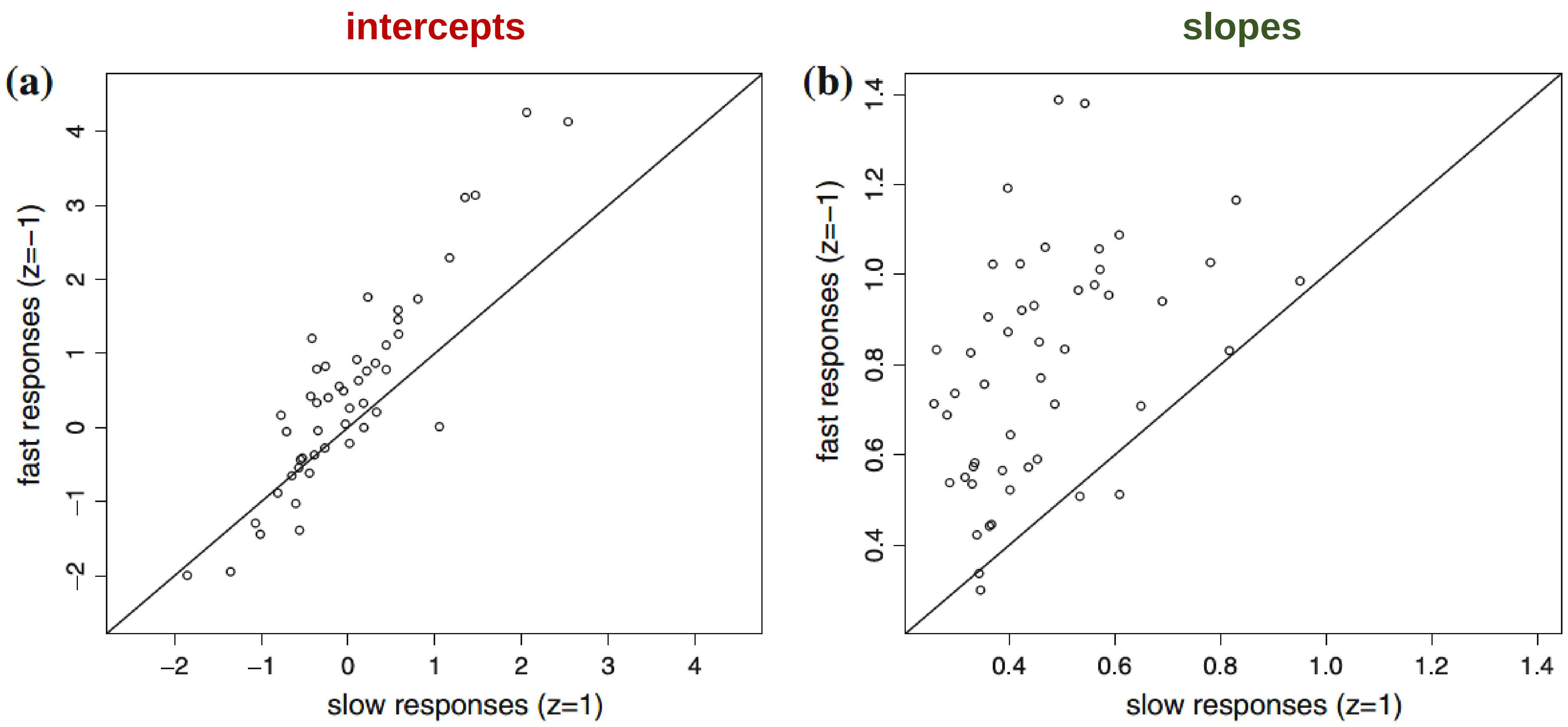


FIGURE 6.

Predicted intercepts (**a**) and slopes (**b**) of the ICC given a slow response ($z_{pi} = 1$) on the x-axis and given a fast response ($z_{pi} = -1$) on the y-axis computed using the estimated baseline intercept (β_0), effect of z_{pi} on the intercept (β_{1i}), baseline slope (α_{0i}) and effect of z_{pi} on the slope (α_{1i}).

Go further…

TABLE 2.
Between-item variances of the item parameters (on the diagonal), correlations between the item parameters (off-diagonal), and the mean vector of the item parameters, with their 95 % credible interval between brackets.

	ξ_i	$\ln(\sigma_i^2)$	$\ln(\alpha_{0i})$	$\ln(\alpha_{1i})$	β_{0i}	β_{1i}
ξ_i	0.20					
	[0.13, 0.30]					
$\ln(\sigma_i^2)$	0.40	0.12				
	[0.14, 0.61]	[0.08, 0.18]				
$\ln(\alpha_{0i})$	0.11	– 0.05	0.12			
	[−0.20, 0.41]	[0−0.34, 0.24]	[0.07, 0.20]			
$\ln(\alpha_{1i})$	0.44	0.33	−0.05	0.04		
	[0.14, 0.69]	[0.02, 0.59]	[−0.43, 0.34]	[0.02, 0.06]		
β_{0i}	−0.44	−0.29	0.13	−0.62	1.39	
	[−0.64, −0.20]	[−0.53,−0.02]	[−0.21, 0.43]	[−0.83,−0.34]	[0.93, 2.09]	
β_{1i}	0.52	0.32	−0.10	0.73	−0.75	0.15
	[0.30, 0.70]	[0.05, 0.55]	[−0.42, 0.23]	[0.48, 0.90]	[−0.85,−0.60]	[0.10, 0.22]
$\boldsymbol{\mu}_{\mathcal{I}}$	3.50	−1.49	−0.57	−0.27	0.17	−0.21
	[3.37, 3.63]	[−1.59,−1.39]	[−0.69,−0.45]	[−0.34,−0.2]	[−0.15, 0.51]	[−0.33,−0.11]

The role of baseline intercept

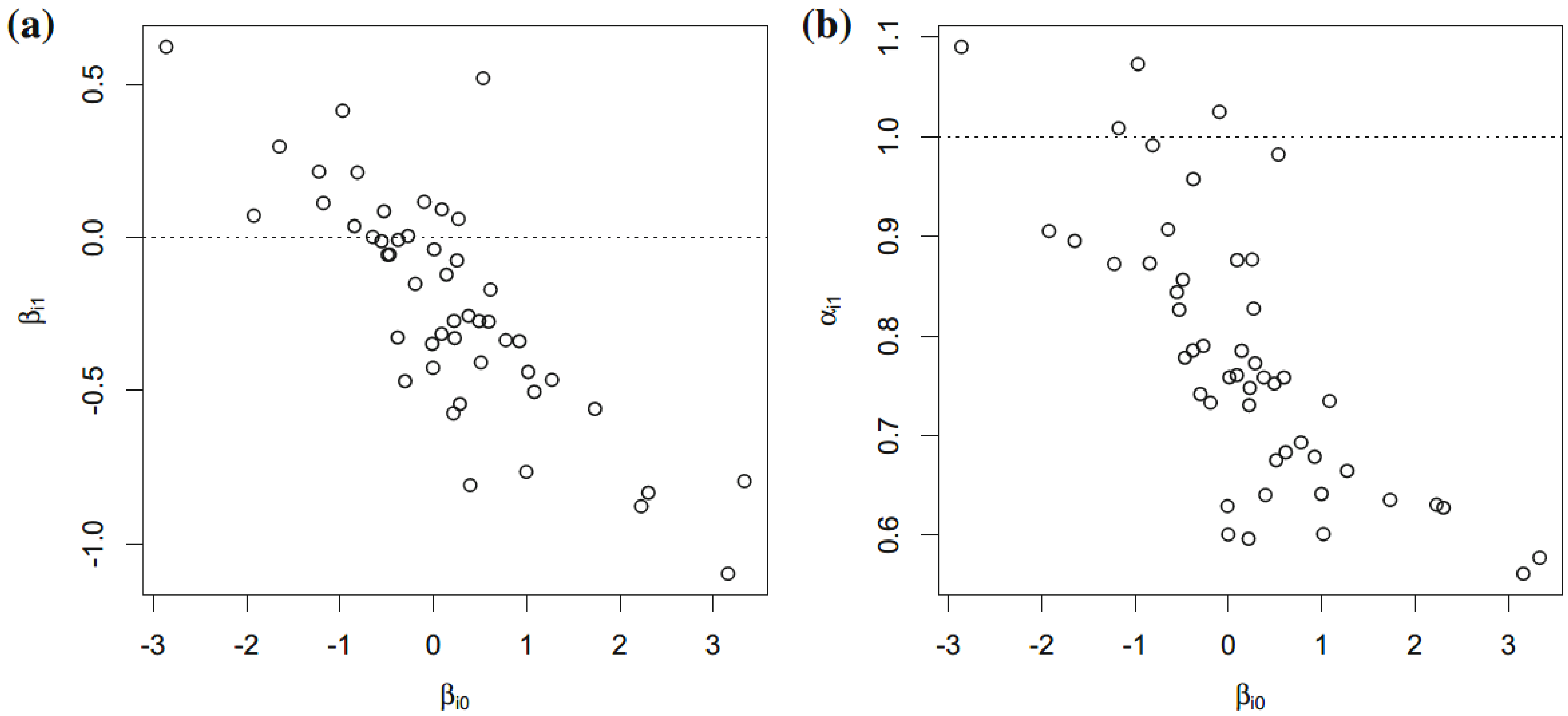


FIGURE 7.
The effects of the residual log-response time on the intercept (**a**) and on the slope (**b**) of the ICC on the y-axis against the baseline intercept of the ICC on the x-axis.

Go further…

TABLE 2.
Between-item variances of the item parameters (on the diagonal), correlations between the item parameters (off-diagonal), and the mean vector of the item parameters, with their 95 % credible interval between brackets.

	ξ_i	$\ln(\sigma_i^2)$	$\ln(\alpha_{0i})$	$\ln(\alpha_{1i})$	β_{0i}	β_{1i}
ξ_i	0.20 [0.13, 0.30]					
$\ln(\sigma_i^2)$	0.40 [0.14, 0.61]	0.12 [0.08, 0.18]				
$\ln(\alpha_{0i})$	0.11 [−0.20, 0.41]	− 0.05 [0−0.34, 0.24]	0.12 [0.07, 0.20]			
$\ln(\alpha_{1i})$	0.44 [0.14, 0.69]	0.33 [0.02, 0.59]	−0.05 [−0.43, 0.34]	0.04 [0.02, 0.06]		
β_{0i}	−0.44 [−0.64, −0.20]	−0.29 [−0.53,−0.02]	0.13 [−0.21, 0.43]	−0.62 [−0.83,−0.34]	1.39 [0.93, 2.09]	
β_{1i}	0.52 [0.30, 0.70]	0.32 [0.05, 0.55]	−0.10 [−0.42, 0.23]	0.73 [0.48, 0.90]	−0.75 [−0.85,−0.60]	0.15 [0.10, 0.22]
$\mu_{\mathcal{I}}$	3.50 [3.37, 3.63]	−1.49 [−1.59,−1.39]	−0.57 [−0.69,−0.45]	−0.27 [−0.34,−0.2]	0.17 [−0.15, 0.51]	−0.21 [−0.33,−0.11]

0.33 [0.06, 0.57]

0.19 [-0.10, 0.48]

after conditioning on β_{0i}: 0.47

The impact of the RT outliers?

- full model with z_{pi} as a covariate without possible outliers
  - **outliers:** *z*-scores below the 0.1-th quantile or above the 99.9-th quantile
  - 514 responses out of the total of 49,000 responses

- effect of the removal:
- **standard deviation of $\boldsymbol{\tau}$:** from **0.33**[0.31, 0.34] to **0.28**[0.27, 0.29]
- **the correlation between $\boldsymbol{\tau}$ and $\boldsymbol{\theta}$:** from **−0.09**[−.16, −.02] to **−0.02**[−.09, .05]

The impact of the RT outliers?

- for the item hyper-parameters

TABLE 3.
Difference between the estimates of the hyper-parameters of the items after the removal of the outliers compared to the original estimates.

	ξ_i	$\ln(\sigma_i^2)$	$\ln(\alpha_{0i})$	$\ln(\alpha_{1i})$	β_{0i}	β_{1i}
ξ_i	0.01					
$\ln(\sigma_i^2)$	−0.10	−0.03				
$\ln(\alpha_{0i})$	−0.02	0.04	0.00			
$\ln(\alpha_{1i})$	−0.06	−0.10	−0.01	0.00		
β_{0i}	0.00	0.02	−0.02	0.03	0.05	
β_{1i}	0.03	−0.05	0.01	0.03	−0.02	0.01
$\mu_{\mathcal{I}}$	0.02	−0.13	0.01	0.00	0.03	−0.01

Simulation study

How parameter recovery is affected by a decrease in **sample size** and **number of items**?

- use the estimates of the item and the person hyper-parameters
  - N = 1000, n = 49 & N = 1000, n = 25;
  - N = 500, n = 49 & N = 500, n = 25;

→ 100 datasets (full model with z_{pi} as a covariate)

- Gibbs Sampler:
  - one chain of 10,000 iterations (including 5000 iterations of burn-in)

Results

TABLE 4.
Results of the simulation study: the expected a posteriori (EAP) estimates of the hyper-parameters averaged across 100 replications and the number of replications in each the true value was within the 95 % credible interval.

	True value	Average EAP				Coverage rate (%)			
N		1000		500		1000		500	
n		49	25	49	25	49	25	49	25
μ_ξ	3.50	3.51	3.50	3.50	3.48	96	95	95	95
$\mu_{\ln(\sigma^2)}$	−1.49	−1.48	−1.48	−1.48	−1.49	95	97	96	98
$\mu_{\ln(\alpha_0)}$	−0.57	−0.57	−0.59	−0.59	−0.59	96	90	91	95
$\mu_{\ln(\alpha_1)}$	−0.27	−0.26	−0.27	−0.27	−0.28	97	96	93	94
μ_{β_0}	0.17	0.17	0.16	0.20	0.18	97	95	96	93
μ_{β_1}	−0.21	−0.21	−0.22	−0.22	−0.22	97	94	94	91
σ^2_ξ	0.20	0.22	0.24	0.24	0.24	97	81	94	94
$\sigma^2_{\ln(\sigma^2)}$	0.12	0.14	0.17	0.16	0.19	94	77	77	79
$\sigma^2_{\ln(\alpha_0)}$	0.12	0.15	0.17	0.14	0.16	95	85	90	88
$\sigma^2_{\ln(\alpha_1)}$	0.04	0.05	0.10	0.05	0.06	92	83	92	91
$\sigma^2_{\beta_0}$	1.39	1.50	1.53	1.50	1.55	95	87	92	92
$\sigma^2_{\beta_1}$	0.15	0.16	0.20	0.16	0.18	95	88	97	93
$\sigma_{\xi,\ln(\sigma^2)}$	0.40	0.33	0.33	0.33	0.29	91	88	97	96
$\sigma_{\xi,\ln(\alpha_0)}$	0.11	0.10	0.02	0.12	0.06	96	84	95	94
$\sigma_{\xi,\ln(\alpha_1)}$	0.44	0.38	0.33	0.36	0.27	93	88	96	94
σ_{ξ,β_0}	−0.44	−0.40	−0.37	−0.39	−0.36	99	84	96	95
σ_{ξ,β_1}	0.53	0.48	0.40	0.47	0.41	96	86	95	93
$\sigma_{\ln(\sigma^2),\ln(\alpha_0)}$	−0.05	−0.05	−0.03	−0.01	−0.02	98	88	94	98
$\sigma_{\ln(\sigma^2),\ln(\alpha_1)}$	0.33	0.27	0.21	0.23	0.16	96	84	96	96
$\sigma_{\ln(\sigma^2),\beta_0}$	−0.30	−0.24	−0.23	−0.24	−0.23	97	89	94	98
$\sigma_{\ln(\sigma^2),\beta_1}$	0.32	0.25	0.24	0.24	0.22	92	87	94	95
$\sigma_{\ln(\alpha_0),\ln(\alpha_1)}$	−0.05	−0.06	−0.02	−0.04	−0.02	98	90	96	98
$\sigma_{\ln(\alpha_0),\beta_0}$	0.13	0.12	0.08	0.09	0.09	94	88	94	96
$\sigma_{\ln(\alpha_0),\beta_1}$	−0.10	−0.10	−0.08	−0.07	−0.11	96	87	95	96
$\sigma_{\ln(\alpha_1),\beta_0}$	−0.62	−0.55	−0.43	−0.52	−0.39	94	86	96	91
$\sigma_{\ln(\alpha_1),\beta_1}$	0.73	0.63	0.49	0.58	0.42	94	82	92	83
σ_{β_0,β_1}	−0.75	−0.69	−0.59	−0.67	−0.63	90	84	90	91
$\rho_{\theta\tau}$	−0.09	−0.09	−0.10	−0.09	−0.09	95	97	93	95
σ_τ	0.33	0.33	0.33	0.33	0.33	96	96	93	94

Discussion

- **model fit:**
  - negative correlation between the baseline item intercept and the effect of the residual response time on the intercept
  - **for difficult items:** slow responses **increase** the probability of a correct response
  - **for easy items:** slow responses **decrease** the probability of the correct response
- **the negative effect on the item slope :**
  - contradict the ‘worst performance rule’: slow responses contain the most information
  - the rule may only apply to the **difficult items**
  - the more time persons take the **more diverse strategies** they may use

Limitations

- **the correlation between ability and speed:**
  - strong and negative
  - or strong and positive

THANKS FOR YOUR ATTENTION!

REPORTER

YINGSHI HUANG