PSYCHOMETRIKA—VOL. 84, NO. 4, 1018–1046
DECEMBER 2019
https://doi.org/10.1007/s11336-019-09682-5

MODELING DIFFERENCES BETWEEN RESPONSE TIMES OF CORRECT AND INCORRECT RESPONSES

MARIA BOLSINOVA

ACTNEXT

JESPER TIJMSTRA

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?

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

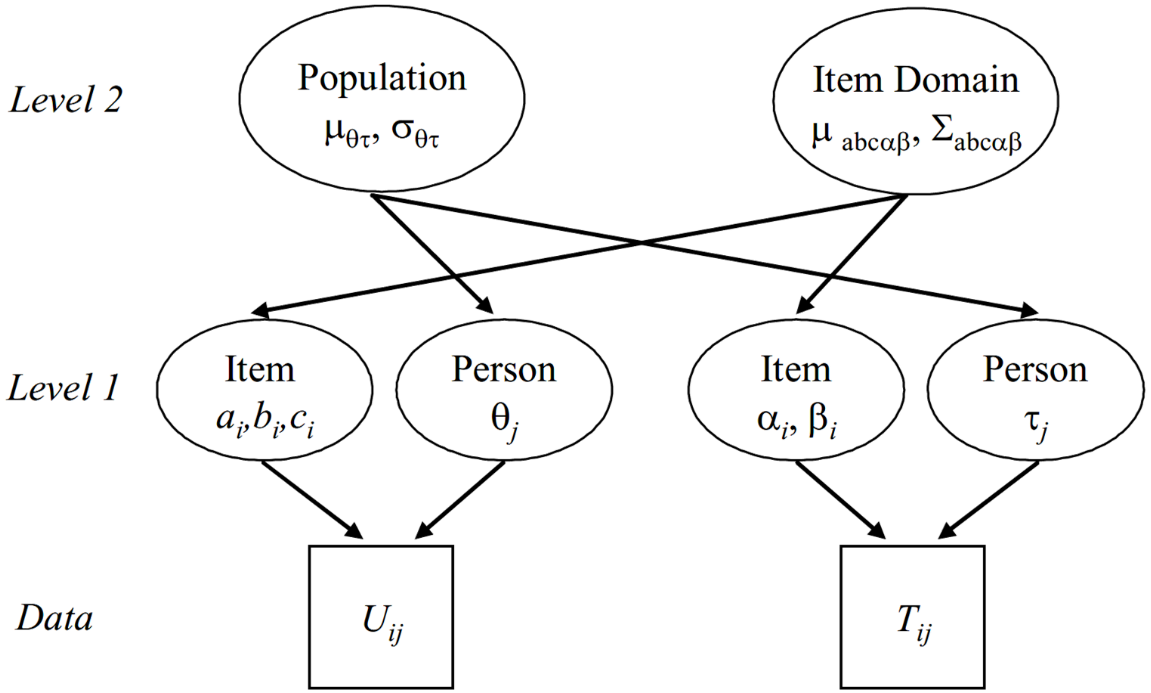
- The assumption of independence
  - standard IRT models: given the ability $\rightarrow$ the RA on different items
  - the lognormal model: given the speed $\rightarrow$ the RT of different items

Introduction

- When considering both RA and RT data

How for each item RT & RA are related ?

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



- RA model:

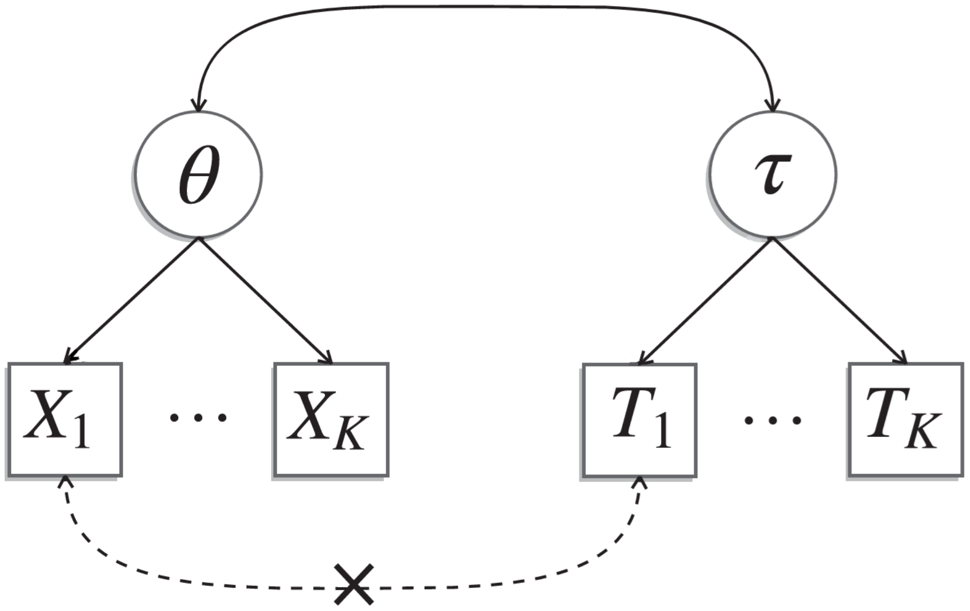
$$\Phi(\alpha_i \theta + \beta_i)^{x_i} (1 - \Phi(\alpha_i \theta + \beta_i))^{1-x_i}$$

- RT model:

$$\ln \mathcal{N}(t_i; \xi_i - \lambda_i \eta, \sigma_i^2)$$

Introduction

- Residual associations between RA and RT

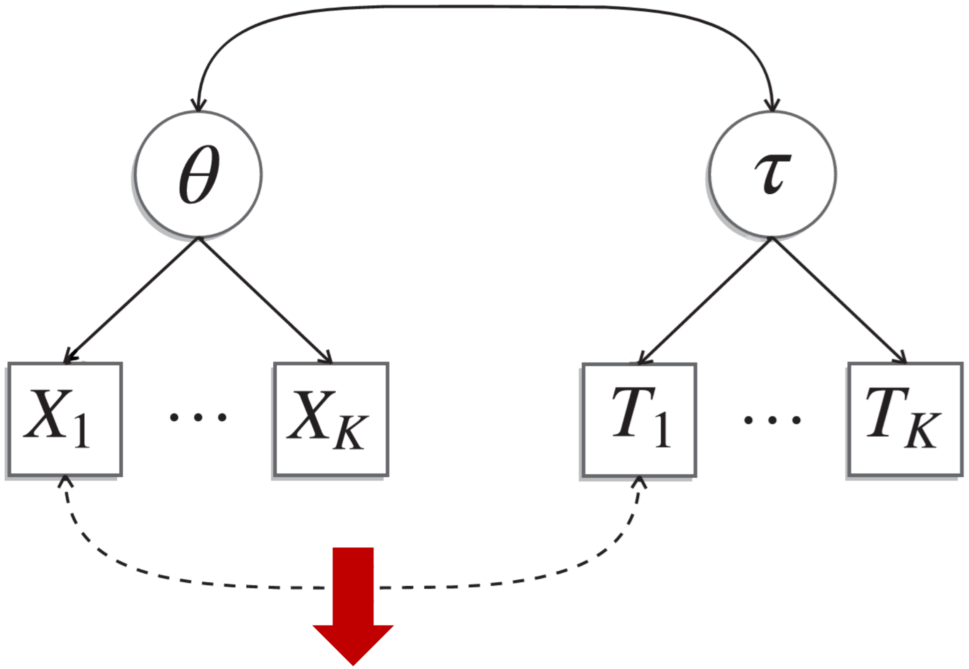


- speed up during the test
- a temporary lapse in concentration
- differential item functioning
- change problem solving strategies

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 RT and a conditional distribution of RA given RT;
3. the marginal distribution of RA and the conditional distribution of RT given RA.

Introduction

1. A bivariate distribution with a nonzero dependence parameter

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

varies across items

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

log-RT

$$E(z_g \mid \theta; \beta_{0g}, \beta_{1g}) = \beta_{0g} + \beta_{1g}\theta \qquad P(x_g = 1 \mid \theta; \beta_{0g}, \beta_{1g}) = \int_0^\infty f(z_g \mid \theta; \beta_{0g}, \beta_{1g})\,dz_g = \Phi(\beta_{0g} + \beta_{1g}\theta)$$

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

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

across person

varies across items

$$f(x_i \mid t_i, \theta, \eta) = \Psi\left(\alpha_i\theta + \beta_{i0} + \beta_{i1}\frac{\ln t_i - (\xi_i - \eta)}{\sigma_i}; x_i\right)$$

$$\Psi(\cdot; x_i) = \Phi(\cdot)^{x_i}(1 - \Phi(\cdot))^{1-x_i}$$

Introduction

3. The marginal distribution of RA and the conditional distribution of RT given RA

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

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

➡ Why is it important to consider correct responses separately from incorrect responses?

➡ What are the benefits of doing so?

Introduction

- Correct and incorrect responses are likely the result of different response processes
  - **a correct response:**

    successfully following the intended solution strategy
  - **an incorrect response:**

    following the intended solution strategy unsuccessfully

    following a different solution strategy than the one intended

    giving up on the item after trying one's best

    failing to attempt to solve the item (e.g., skipping)

Introduction

- **Different residual variances**
  - RTs of correct responses: show more structural patterns
  - incorrect responses: may have larger residual variances

- **Different factor loadings**
  - RTs of correct responses: are more strongly related to the speed

- **Different time intensities**
  - same ability and speed levels: $RT_{correct} > RT_{incorrect}$ or $RT_{correct} < RT_{incorrect}$

- **Different speed latent variables**
  - facing with difficult items: long time or little time

Introduction

- Empirical support
  - Semmes, Davidson and Close (2011)

    correlations between ability and median RT:
    no correlation for correct RTs & positive correlation for incorrect RTs

  - van der Maas and Wagenmakers (2005)

    ability is negatively correlated with the average correct RTs
    not correlated with the average incorrect RTs

⬇

Purpose: propose a modeling framework in line with the third approach
model parameters are allowed to **differ depending on the RA**

Modeling the Differences Between Correct and Incorrect Responses

- The full model

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

- RA model:

$$\Phi(\alpha_i\theta + \beta_i)^{x_i}(1 - \Phi(\alpha_i\theta + \beta_i))^{1-x_i}$$

- RT model:

$$\ln\mathcal{N}(t_i; \xi_i - \lambda_i\eta, \sigma_i^2)$$

How exactly the dependence of t_i on x_i is specified?

$$\ln\mathcal{N}(t_i; \xi_{ix_i} - \lambda_{ix_i}\eta_{x_i}, \sigma_{ix_i}^2) = f(t_i \mid x_i, \boldsymbol{\theta}, \boldsymbol{\eta})$$

only one speed (two-dim): $\eta_0 = \eta_1$

two speed (three-dim): $\eta_0 \neq \eta_1$

		$\sigma_{i0}^2 = \sigma_{i1}^2$	$\sigma_{i0}^2 \neq \sigma_{i1}^2$
$\xi_{i0} = \xi_{i1}$	$\lambda_{i0} = \lambda_{i1}$	the standard HM → $\mathcal{M}_1$	$\mathcal{M}_{3b}$
	$\lambda_{i0} \neq \lambda_{i1}$	-	-
$\xi_{i0} \neq \xi_{i1}$	$\lambda_{i0} = \lambda_{i1}$	van der Linden and Glas (2010) → $\mathcal{M}_2$	$\mathcal{M}_{4b}$
	$\lambda_{i0} \neq \lambda_{i1}$	$\mathcal{M}_{3a}$	$\mathcal{M}_{4a}$

Modeling the Differences Between Correct and Incorrect Responses

- The full model

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

  - RA model:

$$\Phi(\alpha_i\theta + \beta_i)^{x_i}(1 - \Phi(\alpha_i\theta + \beta_i))^{1-x_i}$$

  - RT model:

$$\ln \mathcal{N}(t_i; \xi_i - \lambda_i\eta, \sigma_i^2)$$

$$\ln \mathcal{N}(t_i; \xi_{ix_i} - \lambda_{ix_i}\eta_{x_i}, \sigma_{ix_i}^2) = f(t_i \mid x_i, \boldsymbol{\theta}, \boldsymbol{\eta})$$

only one speed (two-dim): $\eta_0 = \eta_1$

two speed (three-dim): $\eta_0 \neq \eta_1$

- The joint distribution
  - $(\theta, \eta) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & \Sigma_{12} \\ \Sigma_{12} & 1 \end{bmatrix}$$

  - $(\theta, \eta_0, \eta_1) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \\ \mu_3 \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{12} & 1 & \Sigma_{23} \\ \Sigma_{13} & \Sigma_{23} & \Sigma_{33} \end{bmatrix}$$

$\mu_3 \equiv 0$ only when $\xi_{i0} \neq \xi_{i1}$

$\Sigma_{33} \equiv 1$ only when $\lambda_{i0} \neq \lambda_{i1}$

Modeling the Differences Between Correct and Incorrect Responses

- Estimation
  - estimated by sampling from **the joint posterior distribution** of the model parameters
    - ✓ point estimate: averages of the sampled values
    - ✓ 95% credible interval: the 2.5% and 97.5% percentiles of the sampled values
  - (Posterior)~(Likelihood) (Prior)

Modeling the Differences Between Correct and Incorrect Responses

- Likelihood: for the two-dimensional model

$$f(\mathbf{x}, \mathbf{t} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\sigma}^2, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{p=1}^{N} \int \int \prod_{i=1}^{K} \Psi\left(\alpha_i \theta + \beta_i; x_{pi}\right) \ln \mathcal{N}\left(t_{pi}; \xi_{i x_i} - \lambda_{i x_i} \eta_{x_i}, \sigma_{i x_i}^2\right) \mathcal{N}_2(\theta, \eta; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \mathrm{d}\theta \mathrm{d}\eta$$

- Likelihood: for the two-dimensional model

$$f(\mathbf{x}, \mathbf{t} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\sigma}^2, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{p=1}^{N} \int \int \int \prod_{i=1}^{K} \Psi\left(\alpha_i \theta + \beta_i; x_{pi}\right) \ln \mathcal{N}\left(t_{pi}; \xi_{i x_i} - \lambda_{i x_i} \eta_{x_i}, \sigma_{i x_i}^2\right) \mathcal{N}_3(\theta, \eta_0, \eta_1; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \mathrm{d}\theta \mathrm{d}\eta_0 \mathrm{d}\eta_1$$

Modeling the Differences Between Correct and Incorrect Responses

- Prior: for the item parameters
  - independent semi-conjugate low-informative priors

$$f(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\sigma}^2) = \prod_{i=1}^{K} \mathcal{N}(\alpha_i; 0, 100^2)\mathcal{N}(\beta_i; 0, 100^2) \prod_{k=\{0,1\}} \mathcal{N}(\xi_{ik}; 0, 100^2)\mathcal{N}(\lambda_{ik}; 0, 100^2)\mathcal{IG}(\sigma_{ik}^2; 0.001, 0.001)$$

- Prior: for the person parameters
  - $(\theta, \eta_0, \eta_1) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

➡ Gibbs Sampler

✓ the mean vector and the covariance matrix are (partially) constrained

✓ sample them freely but for each sample from the posterior rescale all the parameters

- Model Selection
  - Akaike information criterion (AIC)

    AIC = -2ln(likelihood function) + 2(numbers of parameters)

  - Bayesian information criterion (BIC)

    BIC = -2ln(likelihood function) + ln(n)(numbers of parameters)

↓ Bayesian estimation procedure

at the posterior mean of the parameters

↓

mAIC & mBIC

Modeling the Differences Between Correct and Incorrect Responses

- The values of log-likelihood
  - two-dimensional models

$$v_{ghj1} = \sqrt{2}y_g,$$
$$v_{ghj2} = \sqrt{2(1-\hat{\Sigma}_{12}^2)}y_h + \hat{\Sigma}_{12}v_{ghj1},$$
$$v_{ghj3} = \sqrt{2\left(\hat{\Sigma}_{33} - \frac{\hat{\Sigma}_{13}^2+\hat{\Sigma}_{23}^2-2\hat{\Sigma}_{12}\hat{\Sigma}_{13}\hat{\Sigma}_{23}}{1-\hat{\Sigma}_{12}^2}\right)}y_j + \hat{\mu}_3 + \frac{\hat{\Sigma}_{13}-\hat{\Sigma}_{12}\hat{\Sigma}_{23}}{1-\hat{\Sigma}_{12}^2}v_{ghj1} + \frac{\hat{\Sigma}_{23}-\hat{\Sigma}_{12}\hat{\Sigma}_{13}}{1-\hat{\Sigma}_{12}^2}v_{ghj2}$$

$$\ln \mathcal{L}_{2\text{dim}}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\sigma}}^2, \hat{\Sigma}_{12}; \mathbf{x}, \mathbf{t})$$
$$\approx \sum_{p=1}^{N}\sum_{g=1}^{10}\sum_{h=1}^{10}\frac{w_g}{\sqrt{\pi}}\frac{w_h}{\sqrt{\pi}}\prod_{i=1}^{K}\Psi\left(\hat{\alpha}_i v_{ghj1} + \hat{\beta}_i; x_{pi}\right)\ln\mathcal{N}\left(t_{pi}; \hat{\xi}_{ix_i} - \hat{\lambda}_{ix_i}v_{ghj2}, \hat{\sigma}_{ix_i}^2\right)$$

  - three-dimensional models

$$\ln \mathcal{L}_{3\text{dim}}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\sigma}}^2, \hat{\Sigma}_{12}, \hat{\Sigma}_{13}, \hat{\Sigma}_{23}, \hat{\Sigma}_{33}, \hat{\mu}_3; \mathbf{x}, \mathbf{t})$$
$$\approx \sum_{p=1}^{N}\sum_{g=1}^{10}\sum_{h=1}^{10}\sum_{j=1}^{10}\frac{w_g}{\sqrt{\pi}}\frac{w_h}{\sqrt{\pi}}\frac{w_j}{\sqrt{\pi}}\prod_{i=1}^{K}\Psi\left(\hat{\alpha}_i v_{ghj1} + \hat{\beta}_i; x_{pi}\right)\ln\mathcal{N}\left(t_{pi}; \hat{\xi}_{ix_i} - \hat{\lambda}_{ix_i}v_{ghj(2+x_i)}, \hat{\sigma}_{ix_i}^2\right)$$

Modeling the Differences Between Correct and Incorrect Responses

- The number of parameters

$$\ln \mathcal{L}_{2\text{dim}}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\sigma}}^2, \hat{\Sigma}_{12}; \mathbf{x}, \mathbf{t})$$

$$\ln \mathcal{L}_{3\text{dim}}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\sigma}}^2, \hat{\Sigma}_{12}, \hat{\Sigma}_{13}, \hat{\Sigma}_{23}, \hat{\Sigma}_{33}, \hat{\mu}_3; \mathbf{x}, \mathbf{t})$$

− on the item side

5K parameters

+ K with $\xi_{i0} \neq \xi_{i1}$

+ K with $\lambda_{i0} \neq \lambda_{i1}$

+ K with $\sigma^2_{i0} \neq \sigma^2_{i1}$

− on the population side

one covariance with $\eta_0 = \eta_1$

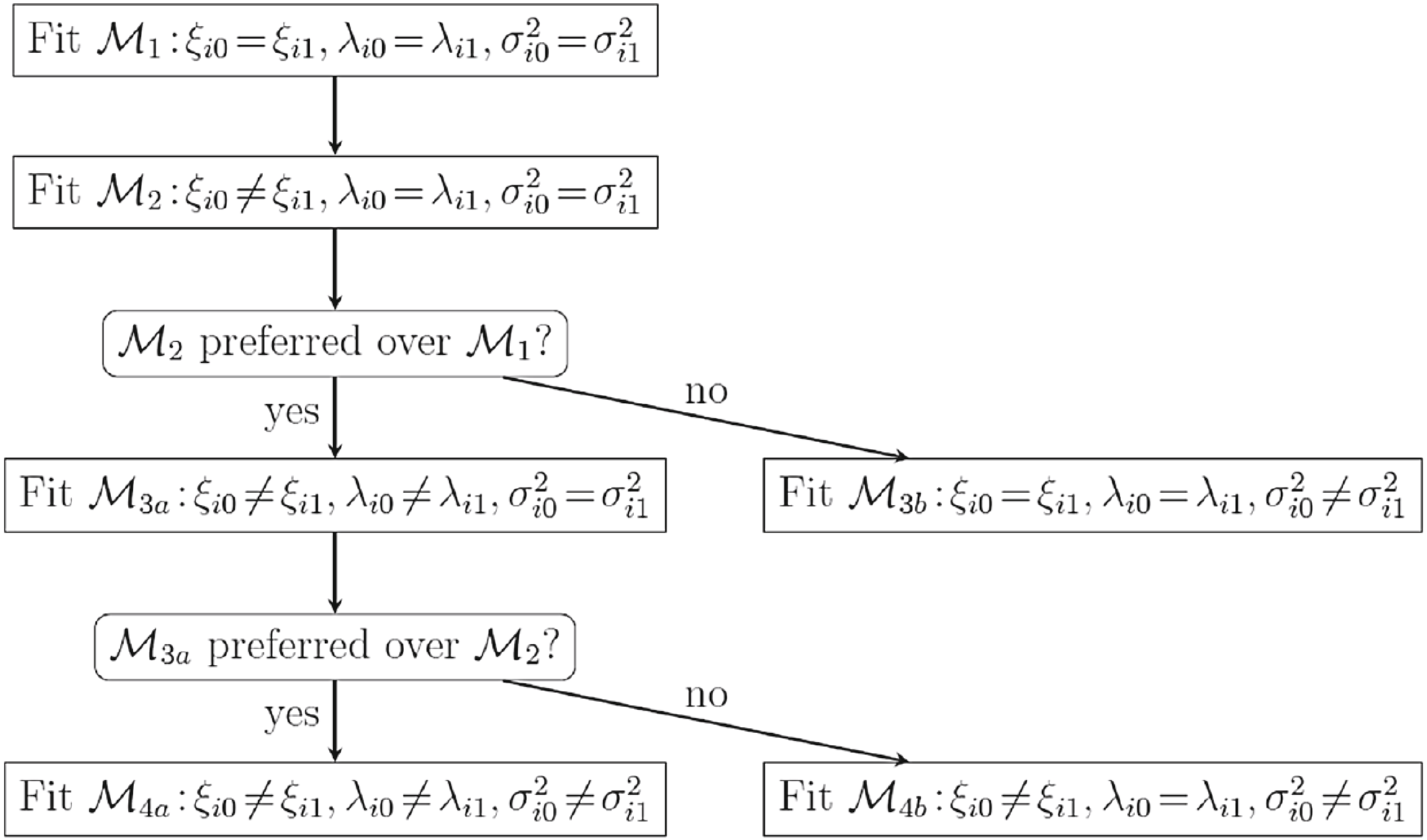
three covariances with $\eta_0 \neq \eta_1$

one freely estimated mean with $\eta_0 \neq \eta_1$ & $\xi_{i0} = \xi_{i1}$

one freely estimated variance with $\eta_0 \neq \eta_1$ & $\lambda_{i0} = \lambda_{i1}$

Modeling the Differences Between Correct and Incorrect Responses

- Stepwise model selection



Modeling the Differences Between Correct and Incorrect Responses

- Posterior Predictive Check (PPC)

Whether the best two-dimensional model adequately captures the relevant patterns?

− the correlation between persons' $M_{\text{correct log-RT}}$ and $M_{\text{incorrect log-RT}}$

1. calculated for the observed data and for G replicated data sets
2. p-value: the proportion of data sets in which the replicated statistic is **larger than** the observed statistic
3. p-values **close to 1** indicate model misfit: **three-dim** is needed
4. p-value is **below a certain threshold** (e.g., 0.95) indicate model **fit well**

Method

- Simulation Study 1: Parameter Recovery
  - for the two-dimensional and three-dimensional models
- Simulation Study 2: Model Selection
  - generate data under all twelve models
- Empirical Example: PIAAC Problem Solving
- Simulation Study Based on the Empirical Example

Simulation Study 1: Parameter Recovery

- the baseline condition
  - sample size = 1000
  - number of items = 16
  - correlation(s) between speed and ability = 0
  - In the case of the three-dimensional model: correlation between the two speed = 0.7

- extra conditions
  - twice as large (2000) and twice as small (500)
  - 32 items
  - correlation of 0.5
  - larger correlation between the two speed (0.9)

- the item parameters
  - α_i values: 0.5 and 1
  - $\{\lambda_{i0}, \lambda_{i1}\}$: {0.3, 0.4} and {0.4, 0.3}
  - $\{\xi_{i0}, \xi_{i1}\}$: {4, 4.1} and {4.1, 4}
  - $\{\sigma^2_{i0}, \sigma^2_{i1}\}$: {0.3, 0.2} and {0.2, 0.3}
  - item intercept parameters β_i: equally spaced between − 1.5 and 1.5

➡ **16 unique item parameter combinations**
the same item parameters were used twice with 32 items

Simulation Study 1: Parameter Recovery

- data sets
  - the first five conditions:
    500 data sets (both the two-dimensional and three-dimensional version of $\mathcal{M}_{4a}$)
  - the last condition: 500 data sets (the three-dimensional $\mathcal{M}_{4a}$)
  - RA data: $\Phi(\alpha_i \theta + \beta_i)^{x_i} (1 - \Phi(\alpha_i \theta + \beta_i))^{1-x_i}$
  - RT data: $\ln \mathcal{N}(t_i ; \xi_{i x_i} - \lambda_{i x_i} \eta_{x_i}, \sigma^2_{i x_i})$

- person parameters
  - $\mathcal{N}_2 \left(\mathbf{0}, \begin{bmatrix} 1 & \Sigma_{12} \\ \Sigma_{12} & 1 \end{bmatrix} \right)$ for the conditions with $\eta_0 = \eta_1$
  - $\mathcal{N}_3 \left(\mathbf{0}, \begin{bmatrix} 1 & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{12} & 1 & \Sigma_{23} \\ \Sigma_{13} & \Sigma_{23} & 1 \end{bmatrix} \right)$ for the conditions with $\eta_0 \neq \eta_1$

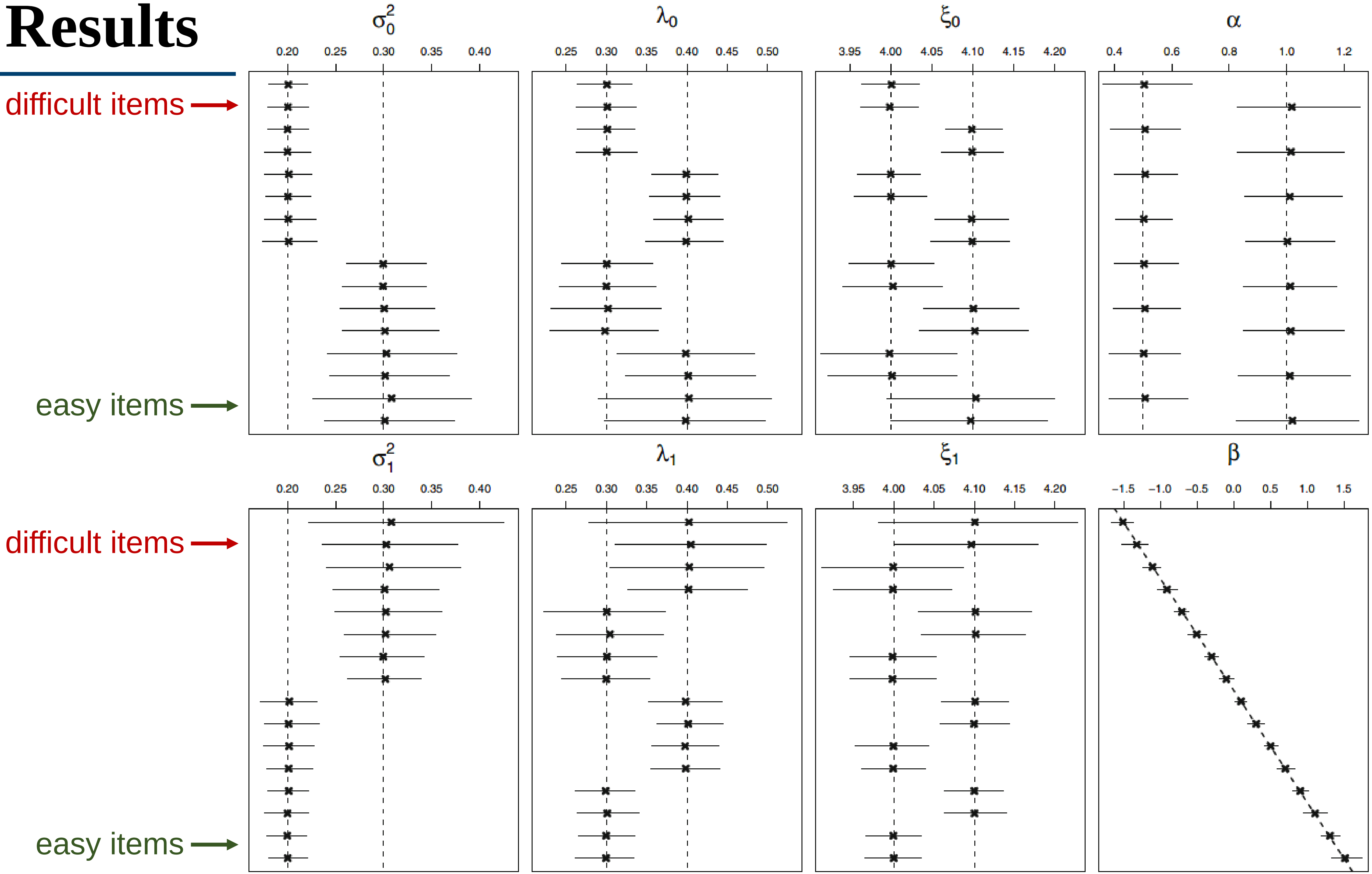
- Estimation
  - Gibbs Sampler with 6000 iterations
  - burn-in: first 1000 iterations

- Evaluation
  - the (average) absolute bias
  - variance
  - mean squared error

Results

Condition		α	β	ξ	λ	σ^2	Σ_{12}	Σ_{13}	Σ_{23}
		Bias							
$\eta_0 = \eta_1$	Baseline	0.008	0.007	0.001	0.001	0.002	0.001	–	–
	$N = 500$	0.021	0.014	0.002	0.002	0.003	0.001	–	–
	$N = 2000$	0.005	0.003	0.001	0.000	0.001	0.001	–	–
	$K = 32$	0.008	0.006	0.001	0.001	0.002	0.000	–	–
	$\Sigma_{12} = .5$	0.009	0.007	0.001	0.001	0.002	0.004	–	–
$\eta_0 \neq \eta_1$	Baseline	0.008	0.006	0.001	0.001	0.001	0.001	0.001	0.003
	$N = 500$	0.018	0.014	0.002	0.002	0.004	0.000	0.000	0.007
	$N = 2000$	0.004	0.003	0.001	0.001	0.001	0.002	0.000	0.002
	$K = 32$	0.009	0.006	0.001	0.001	0.002	0.001	0.000	0.002
	$\Sigma_{12} = \Sigma_{13} = .5$	0.010	0.007	0.001	0.001	0.002	0.002	0.001	0.001
	$\Sigma_{23} = .9$	0.011	0.008	0.001	0.001	0.001	0.000	0.002	0.011
		Variance							
$\eta_0 = \eta_1$	Baseline	0.006	0.004	0.001	0.001	0.001	0.001	–	–
	$N = 500$	0.013	0.009	0.002	0.002	0.001	0.003	–	–
	$N = 2000$	0.003	0.002	0.000	0.001	0.000	0.001	–	–
	$K = 32$	0.005	0.004	0.001	0.001	0.001	0.001	–	–
	$\Sigma_{12} = .5$	0.007	0.004	0.001	0.001	0.001	0.001	–	–
$\eta_0 \neq \eta_1$	Baseline	0.007	0.004	0.001	0.001	0.001	0.002	0.002	0.001
	$N = 500$	0.014	0.009	0.002	0.002	0.001	0.003	0.004	0.001
	$N = 2000$	0.003	0.002	0.001	0.001	0.000	0.001	0.001	0.000
	$K = 32$	0.005	0.004	0.001	0.001	0.001	0.001	0.001	0.000
	$\Sigma_{12} = \Sigma_{13} = .5$	0.006	0.004	0.001	0.001	0.001	0.001	0.001	0.001
	$\Sigma_{23} = .9$	0.007	0.004	0.001	0.001	0.001	0.002	0.002	0.000
		Mean squared error							
$\eta_0 = \eta_1$	Baseline	0.006	0.004	0.001	0.001	0.001	0.001	–	–
	$N = 500$	0.014	0.009	0.002	0.002	0.001	0.003	–	–
	$N = 2000$	0.003	0.002	0.000	0.001	0.000	0.001	–	–
	$K = 32$	0.005	0.004	0.001	0.001	0.001	0.001	–	–
	$\Sigma_{12} = .5$	0.007	0.004	0.001	0.001	0.001	0.001	–	–
$\eta_0 \neq \eta_1$	Baseline	0.007	0.004	0.001	0.001	0.001	0.002	0.002	0.001
	$N = 500$	0.014	0.010	0.002	0.002	0.001	0.003	0.004	0.001
	$N = 2000$	0.003	0.002	0.001	0.001	0.000	0.001	0.001	0.000
	$K = 32$	0.005	0.004	0.001	0.001	0.001	0.001	0.001	0.000
	$\Sigma_{12} = \Sigma_{13} = .5$	0.006	0.004	0.001	0.001	0.001	0.001	0.001	0.001
	$\Sigma_{23} = .9$	0.007	0.005	0.001	0.001	0.001	0.002	0.002	0.000

Results



difficult items
easy items
difficult items
easy items
σ_0^2
λ_0
ξ_0
α
σ_1^2
λ_1
ξ_1
β
0.20 0.25 0.30 0.35 0.40
0.25 0.30 0.35 0.40 0.45 0.50
3.95 4.00 4.05 4.10 4.15 4.20
0.4 0.6 0.8 1.0 1.2
−1.5 −1.0 −0.5 0.0 0.5 1.0 1.5

Simulation Study 2: Model Selection

- model selection
  - mAIC
  - mBIC
  - mAIC in combination with the posterior predictive check
  - mBIC in combination with the posterior predictive check

Condition	N	K	Σ_{23}
A (baseline)	1000	20	.7
B	1000	**10**	.7
C	1000	**40**	.7
D	**500**	20	.7
E	**2000**	20	.7
F	1000	20	**.9**

Note: For each non-baseline condition the factor that differentiates it from the baseline condition is in bold. Condition F was used only for the three-dimensional models.

Simulation Study 2: Model Selection

- data sets
  - each condition 50 data sets were generated under each of the 12 models

- item parameters
  - $\alpha_i \sim \mathcal{N}(1, 0.2^2)$
  - $\beta_i \sim \mathcal{N}(0, 0.5^2)$
  - equal time intensities condition, otherwise:
    $\xi_i \sim \mathcal{N}(4, 0.5^2)$ $\quad [\xi_{i0}\ \xi_{i1}]^T \sim \mathcal{N}_2\left(\begin{bmatrix}4\\4.1\end{bmatrix}, 0.25\begin{bmatrix}1 & 0.7\\0.7 & 1\end{bmatrix}\right)$
  - equal factor loadings condition, otherwise:
    $\lambda_i \sim \mathcal{N}(0.4, 0.1^2)$ $\quad [\lambda_{i0}\ \lambda_{i1}]^T \sim \mathcal{N}_2\left(\begin{bmatrix}0.4\\0.4\sqrt{0.8}\end{bmatrix}, 0.01\begin{bmatrix}1 & 0.7\\0.7 & 1\end{bmatrix}\right)$
  - equal residual variances condition, otherwise:
    $\sigma_i^2 \sim \mathcal{U}(0.2, 0.3)$ $\quad \sigma_{i0}^2 \sim \mathcal{U}(0.2, 0.3)$ and $\sigma_{i1}^2 \sim \mathcal{U}(0.15, 0.25)$

Simulation Study 2: Model Selection

- person parameters
  - for the two-dimensional models: $\mathcal{N}_2\left(\mathbf{0}, \mathbf{I}_2\right)$
  - for the three-dimensional models: $\mathcal{N}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$\boldsymbol{\mu} = \mathbf{0}$ when $\xi_{i0} \neq \xi_{i1}$

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \\ 0.1 \end{bmatrix} \text{ when } \xi_{i0} = \xi_{i1}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \Sigma_{23}/\sqrt{\Sigma_{33}} \\ 0 & \Sigma_{23}/\sqrt{\Sigma_{33}} & 1 \end{bmatrix} \text{ when } \lambda_{i0} \neq \lambda_{i1}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \Sigma_{23}/\sqrt{\Sigma_{33}}\sqrt{0.8} \\ 0 & \Sigma_{23}/\sqrt{\Sigma_{33}}\sqrt{0.8} & 0.8 \end{bmatrix} \text{ when } \lambda_{i0} = \lambda_{i1}$$

- Estimation
  - Gibbs Sampler with 6000 iterations
  - burn-in: first 1000 iterations
  - each second iteration after the burn-in was used

Results

True model		P	mAIC Condition A	B	C	D	E	F	mAIC and PPC Condition A	B	C	D	E	F
$\eta_0 = \eta_1$	$\mathcal{M}_1$	$5K+1$	1	41	6	0	0	–	49	48	47	47	48	–
	$\mathcal{M}_2$	$6K+1$	0	39	4	0	0	–	49	45	48	50	49	–
	$\mathcal{M}_{3a}$	$7K+1$	5	42	13	0	0	–	49	47	45	49	48	–
	$\mathcal{M}_{3b}$	$6K+1$	0	28	2	0	0	–	47	46	47	47	49	–
	$\mathcal{M}_{4a}$	$8K+1$	0	20	1	0	0	–	48	48	49	49	49	–
	$\mathcal{M}_{4b}$	$7K+1$	0	32	4	0	0	–	49	45	47	50	49	–
$\eta_0 \neq \eta_1$	$\mathcal{M}_1$	$5K+5$	49	46	49	50	50	50	49	46	49	50	50	42
	$\mathcal{M}_2$	$6K+4$	50	47	48	49	48	48	50	47	48	49	48	41
	$\mathcal{M}_{3a}$	$7K+3$	50	50	50	50	49	48	50	49	50	50	49	43
	$\mathcal{M}_{3b}$	$6K+5$	50	49	49	50	50	50	50	47	49	50	50	47
	$\mathcal{M}_{4a}$	$8K+3$	49	48	50	50	50	50	50	45	45	50	50	43
	$\mathcal{M}_{4b}$	$7K+4$	50	46	45	50	50	50	49	47	50	50	50	45

True model		P	mBIC Condition A	B	C	D	E	F	mBIC and PPC Condition A	B	C	D	E	F
$\eta_0 = \eta_1$	$\mathcal{M}_1$	$5K+1$	14	50	0	24	0	–	49	50	49	48	50	–
	$\mathcal{M}_2$	$6K+1$	3	50	0	20	0	–	49	50	48	50	50	–
	$\mathcal{M}_{3a}$	$7K+1$	3	16	0	1	2	–	24	17	1	29	48	–
	$\mathcal{M}_{3b}$	$6K+1$	3	33	0	2	0	–	40	33	12	42	49	–
	$\mathcal{M}_{4a}$	$8K+1$	0	15	0	0	0	–	27	19	3	31	47	–
	$\mathcal{M}_{4b}$	$7K+1$	0	32	0	2	0	–	42	38	9	49	49	–
$\eta_0 \neq \eta_1$	$\mathcal{M}_1$	$5K+5$	50	50	50	50	50	50	43	37	48	41	48	45
	$\mathcal{M}_2$	$6K+4$	50	50	50	50	50	50	39	38	33	38	45	43
	$\mathcal{M}_{3a}$	$7K+3$	12	10	10	3	41	16	7	8	3	7	20	14
	$\mathcal{M}_{3b}$	$6K+5$	43	38	50	17	49	43	43	38	17	50	49	41
	$\mathcal{M}_{4a}$	$8K+3$	19	9	17	0	44	40	21	9	2	15	44	36
	$\mathcal{M}_{4b}$	$7K+4$	45	42	50	18	50	45	44	42	16	50	50	42

Note: P denotes the number of free parameters in the true model.

Empirical Example

- the Programme of International Assessment of Adult Competences (PIAAC)
  - **the problem solving** in technology-based environments domain
  - items are interactive and require a constructed response (**no guessing parameter**)
  - two computer-based problem solving modules each consisting of 7 items (7 + 7 intotal)
  - the problem solving modules + a module from a different domain
  - both problem solving modules

    (overall time limit of 30 minutes / module)
  - **data files:** 12th of June 2018, Canada (the largest number of respondents, 10315)
- the RA scores
  - the items were coded as correct/incorrect
  - eigenvalues of the correlation matrix: **one dimension** should be sufficient

Empirical Example

1. the CI-HM ($\mathcal{M}_1$, $\eta_0 = \eta_1$)

   − Gibbs Sampler with 20,000 iterations

   (including 10,000 burn-in, and a thinning of 2 was applied)

whether the model adequately captured the differences between the RTs of correct and incorrect responses?

   − Posterior predictive checks (100 replicated data sets)

   D1: differences between $M_{\text{correct log-RT}}$ and $M_{\text{incorrect log-RT}}$

   D2: the ratio between $S^2_{\text{correct log-RT}}$ and $S^2_{\text{incorrect log-RT}}$

   D3: the ratio between the first eigenvalues of the correlation matrices of log-RTs computed separately for correct and incorrect responses

Empirical Example

1. Results

   - the observed ones,

     D1 = 0.338

     D2 = 0.537

     D3 = 1.285

➡ in all of the 100 generated data sets:

- ✓ D1 & D3: smaller than the observed ones
- ✓ D2: larger than the observed one

➡ there is likely CD between RA and RT

Empirical Example

2. fitted two CD models

- the first model:

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

- the second model:

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

- Gibbs Samplers:

  (20,000 iterations, including 10,000 burn-in, and a thinning of 2)

Empirical Example

2. Results

Model	P	mAIC	$\bar{D}_1$	$\bar{D}_2$	$\bar{D}_3$	p_1	p_2	p_3
CI model ($\mathcal{M}_1, \eta_0 = \eta_1$)	71	234,633.3	0.215	0.779	0.811	.00	1.00	.00
CD model from approach 1	85	229,347.3	0.323	0.887	0.847	.00	1.00	.00
CD model from approach 2	99	228,351.8	0.322	0.819	0.837	.01	1.00	.00

Empirical Example

3. fitted the set of two-dim models and the set of three-dim models
   - Gibbs Samplers:
     (20,000 iterations, including 10,000 burn-in, and a thinning of 2)

Empirical Example

3. Results

➡ the observed ones: D1 = 0.338, D2 = 0.537, D3 = 1.285

Model	P	mAIC	$\bar{D}_1$	$\bar{D}_2$	$\bar{D}_3$	p_1	p_2	p_3
CI model ($\mathcal{M}_1, \eta_0 = \eta_1$)	71	234,633.3	0.215	0.779	0.811	.00	1.00	.00
CD model from approach 1	85	229,347.3	0.323	0.887	0.847	.00	1.00	.00
CD model from approach 2	99	228,351.8	0.322	0.819	0.837	.01	1.00	.00
CD models from approach 3								
$\mathcal{M}_2, \eta_0 = \eta_1$	85	227,699.7	0.336	0.859	0.832	.36	1.00	.00
$\mathcal{M}_{3a}, \eta_0 = \eta_1$	99	225,809.6	0.335	0.707	0.650	.32	1.00	.00
$\mathcal{M}_{4a}, \eta_0 = \eta_1$	113	218,993.7	0.336	0.549	1.240	.34	.94	.08
$\mathcal{M}_1, \eta_0 \neq \eta_1$	75	229,599.2	0.377	0.652	0.663	1.00	1.00	.00
$\mathcal{M}_2, \eta_0 \neq \eta_1$	88	225,194.5	0.336	0.733	0.667	.34	1.00	.00
$\mathcal{M}_{3a}, \eta_0 \neq \eta_1$	101	224,114.3	0.336	0.693	0.796	.29	1.00	0.00
$\mathcal{M}_{4a}, \eta_0 \neq \eta_1$	105	216,824.9	0.337	0.541	1.250	.40	.64	.15

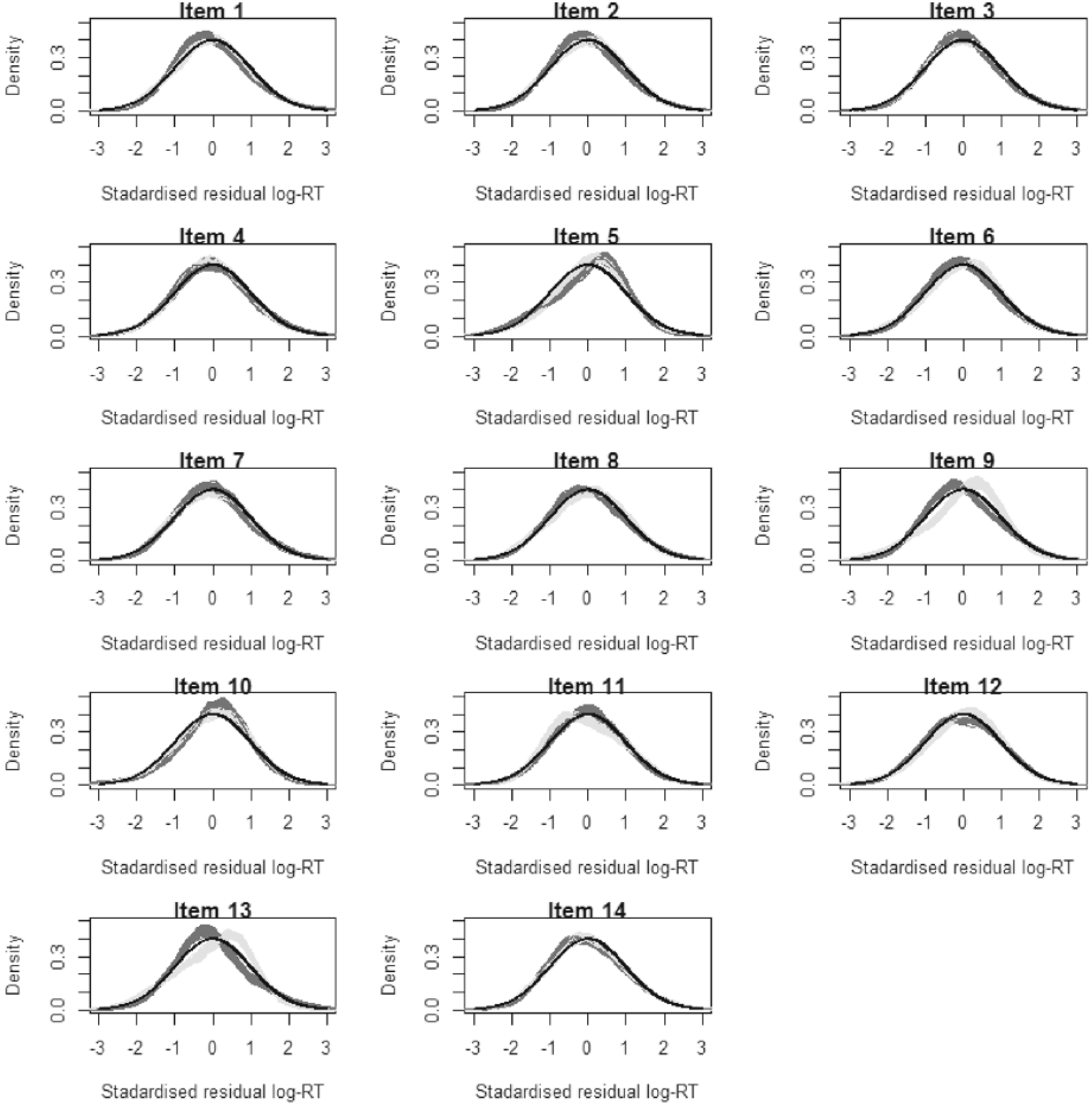
➡ PPC (100 data sets): *p*-value of 1

Distributional assumptions check

- assumes a separate lognormal distribution for RT for the two RA outcomes
- examined the posterior distribution of the standardized residuals of log-RTs

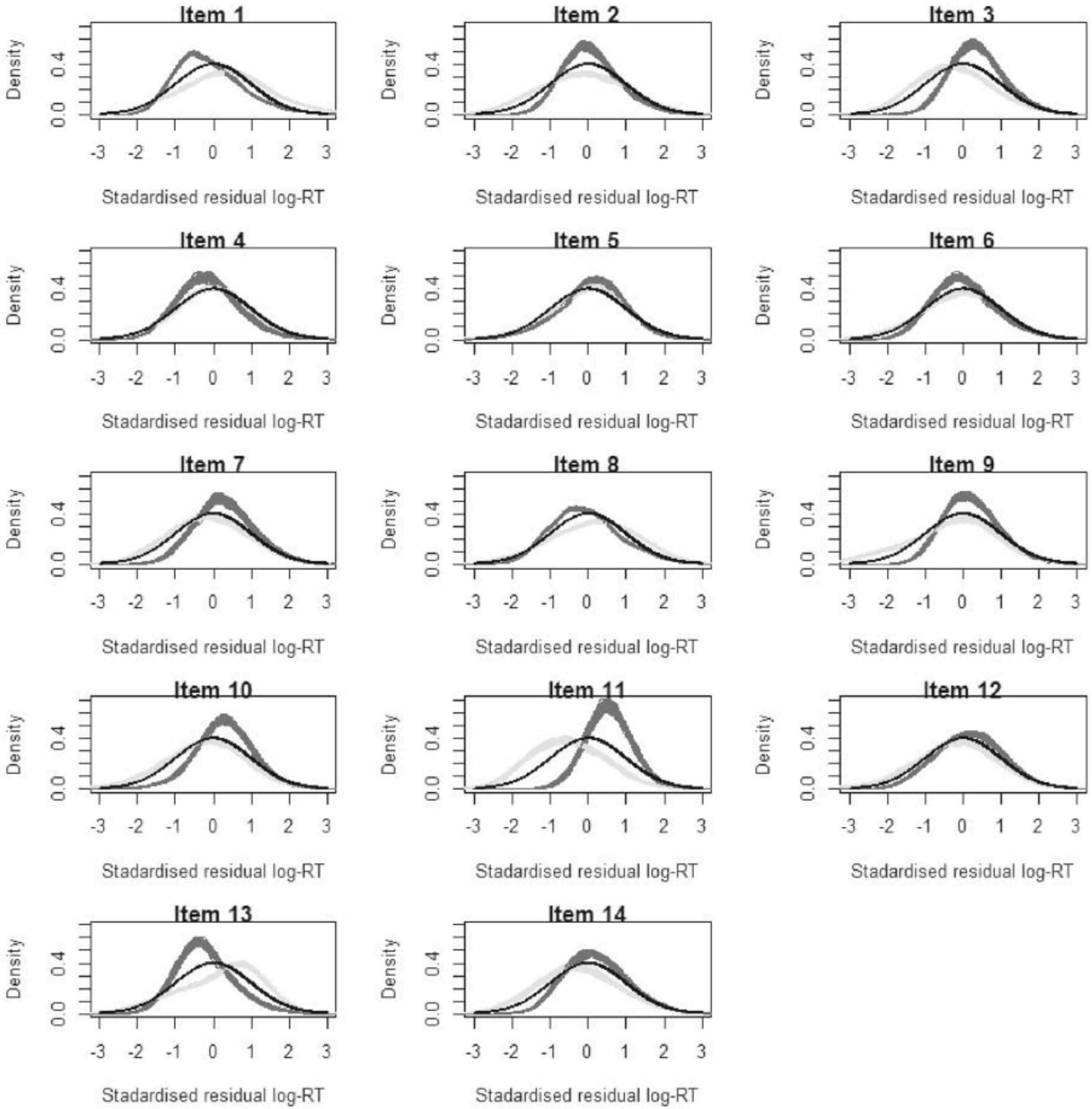
Distributional assu

- three-dimensional $\mathcal{M}_{4a}$



Distributional assu

- the CI model



Empirical Example

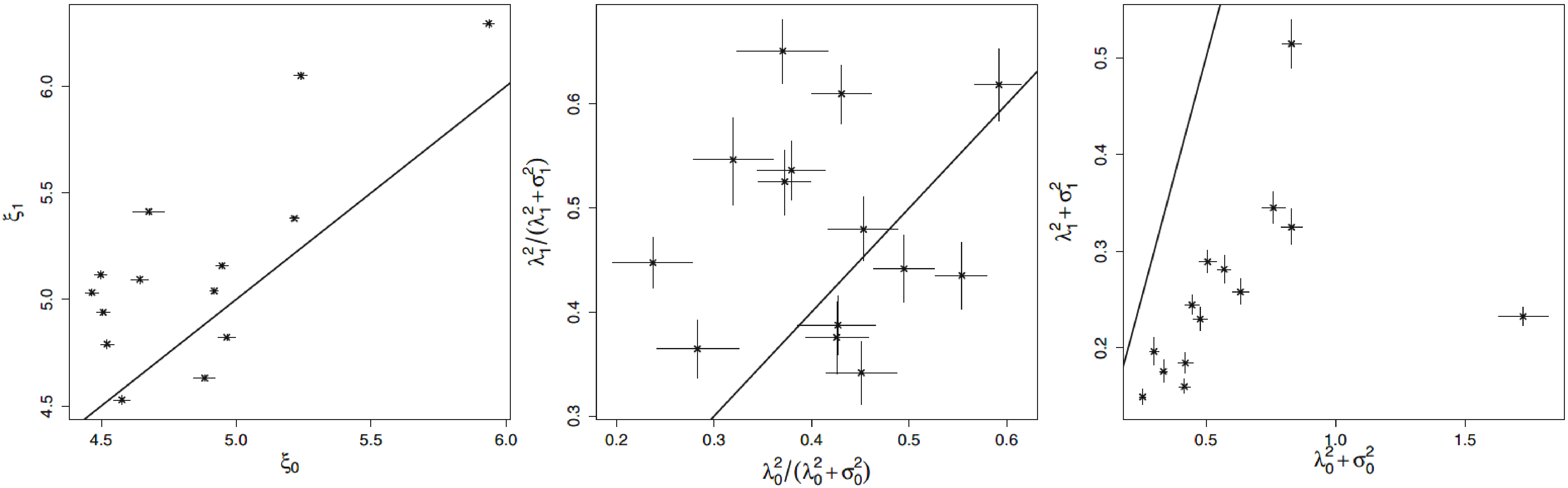
- three correlations between the person parameters
  - **θ and η_0:** -0.661 [-.642,-.679]
  - persons who give fast incorrect responses generally having a lower ability level

  - **θ and η_1:** 0.038 [.005,.072]
  - response speed and ability is much weaker

  - **η_0 and η_1:** 0.689 [.662,.714]
  - the two speed latent variables are strongly associated but still only share less than 50% of their variance

Empirical Example

- the 95% credible intervals for relevant item properties



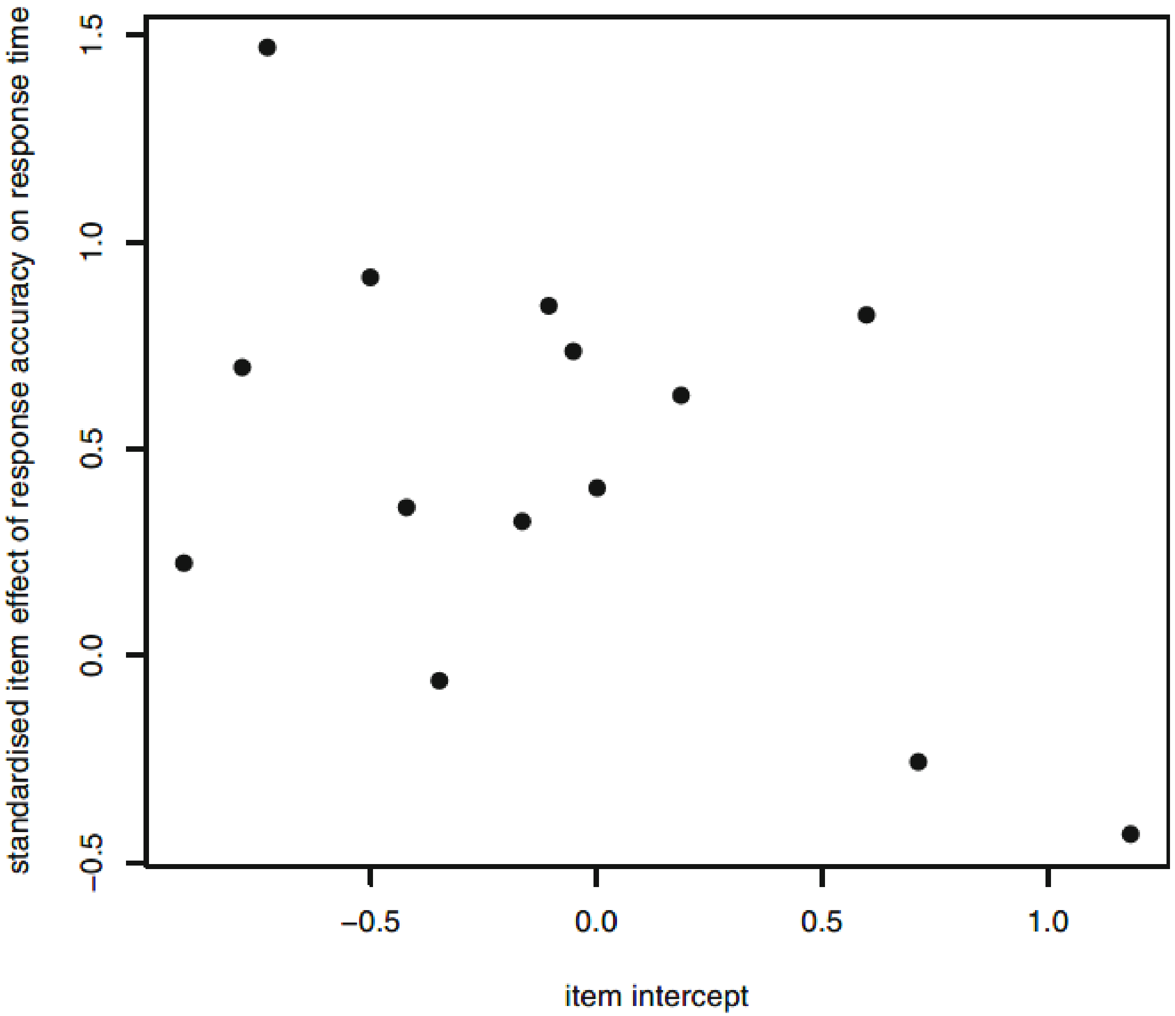
variance explained by the speed

total variance

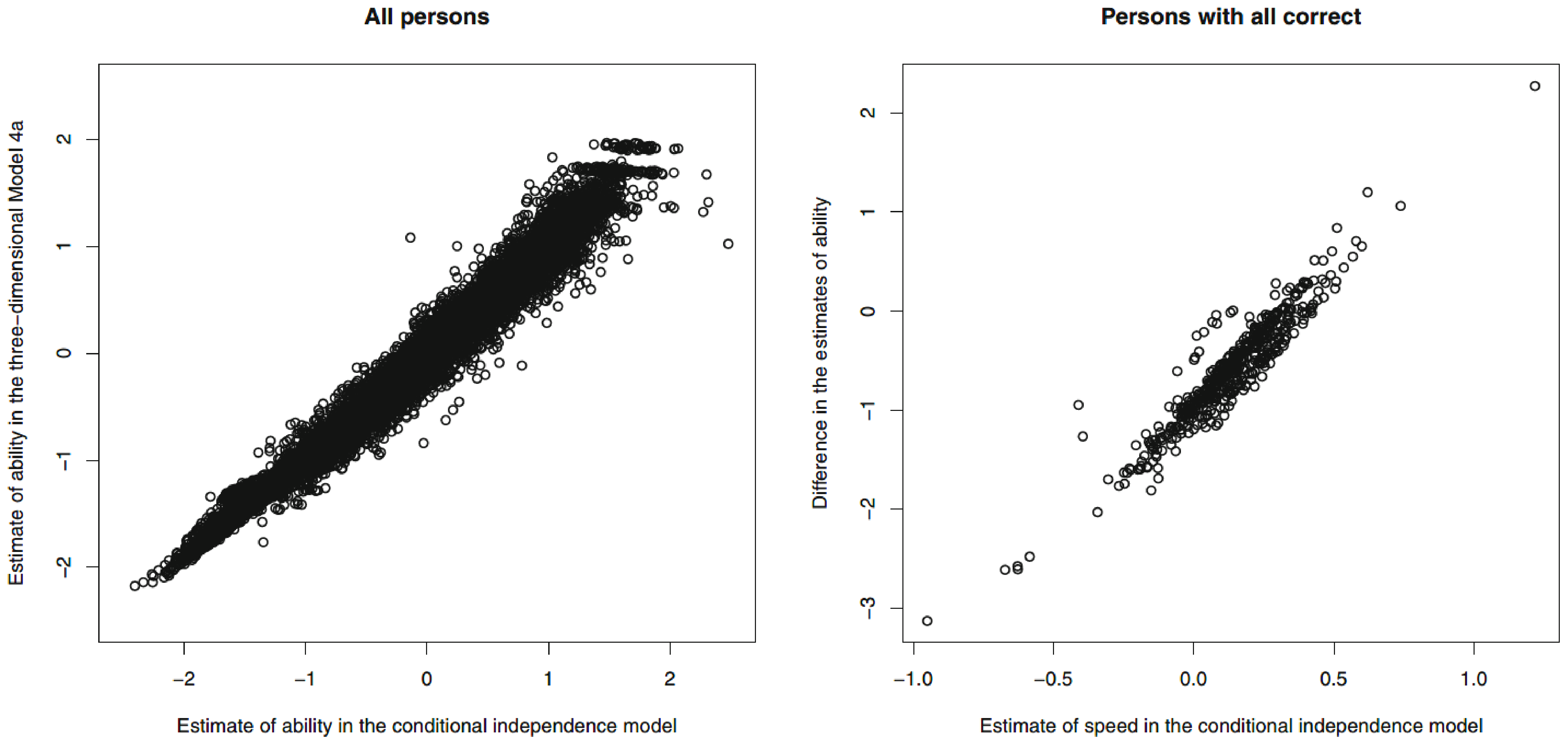
Empirical Example

- quantify the strength and direction of the CD
  - item-specific standardized effect of RA on log-RT

$$\frac{\xi_{i1}-\xi_{i0}}{\sqrt{\frac{(N_{i0}-1)(\sigma_{i0}^2+\lambda_{i0}^2)+(N_{i1}-1)(\sigma_{i1}^2+\lambda_{i1}^2)}{N_{i0}+N_{i1}-2}}}$$



Empirical Example



Simulation Study Based on the Empirical Example

- sample sizes:
  - N = 500, N = 1000, and N = 2000
  - a condition with the same sample size (N = 10,245) and the pattern of missingness

- data generation:
  - The RA: the 2PNO model
  - The RT: the three-dimensional $\mathcal{M}_{4a}$

- estimation:
  - Gibbs Sampler
    (6000 iterations, including 1000 burn-in, and a thinning of 2)

Results

Condition	α	β	ξ	λ	σ^2	Σ_{12}	Σ_{13}	Σ_{23}
	Bias							
$N = 500$	0.033	0.011	0.004	0.001	0.002	0.016	0.019	0.024
$N = 1000$	0.020	0.006	0.001	0.001	0.001	0.011	0.011	0.016
$N = 2000$	0.009	0.003	0.002	0.001	0.001	0.011	0.004	0.009
$N = 10{,}245$	0.004	0.001	0.001	0.001	0.000	0.005	0.005	0.006
	Variance							
$N = 500$	0.019	0.010	0.002	0.002	0.001	0.001	0.004	0.002
$N = 1000$	0.009	0.005	0.001	0.001	0.001	0.001	0.002	0.001
$N = 2000$	0.004	0.002	0.000	0.000	0.000	0.000	0.001	0.000
$N = 10{,}245$	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	Mean squared error							
$N = 500$	0.021	0.010	0.002	0.002	0.001	0.001	0.004	0.002
$N = 1000$	0.010	0.005	0.001	0.001	0.001	0.001	0.002	0.001
$N = 2000$	0.004	0.002	0.000	0.000	0.000	0.000	0.001	0.001
$N = 10{,}245$	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000

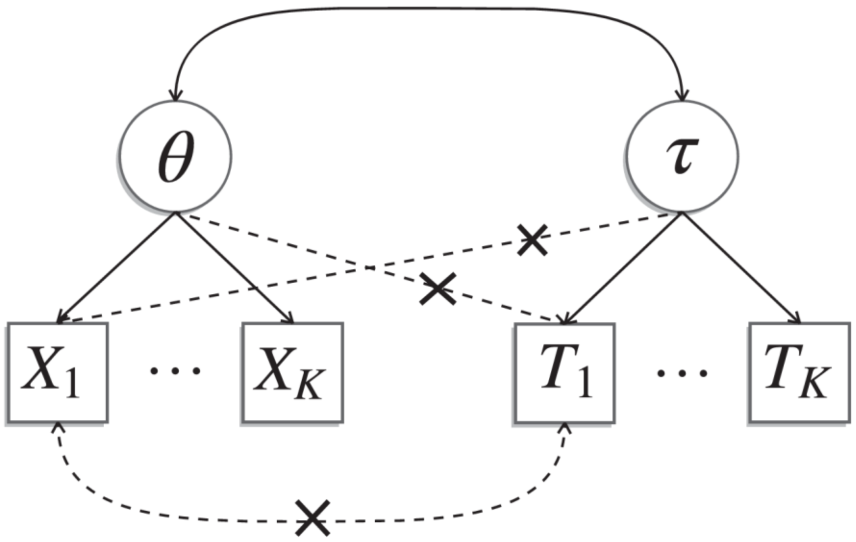
Note: The first three conditions have a complete design, and the last condition has an incomplete design with the missingness patterns matching those in the empirical example.

Discussion

- proposed a framework to directly investigate the differences of RTs between correct and incorrect responses
- all model parameters can generally be **recovered well** if the model is correctly specified
- **the mAIC** with a posterior predictive check is well-suited for selecting the correct model
- there may in practice be **notable relevant differences** between the models for the RTs of correct and incorrect responses
- **two speed** latent variables were needed to best model the empirical data

Future directions

- **other parametric forms** for the RT model for correct and incorrect responses could be explored



ability latent variable(s) is correctly specified,

onfounding

make it possible to deal with **polytomously**

- it is still assumed that RTs only provide collateral information for the estimation of ability through the speed latent variable(s) in the model

THANKS FOR LISTENING!

REPORTER

YINGSHI HUANG