A Two Parameter Asymptote IRT Model

Laine Bradshaw
Assessment & Measurement, James Madison University
Jonathan Templin
Research, Evaluation, Measurement & Statistics, University of Georgia

Introduction

- Item response theory (IRT) models are the most commonly used psychometric models in educational measurement
- Well-known phenomena occurs when a student guesses the correct answer to an item that is more difficult than his or her ability
  - » Yields model-data misfit for the 2-PL
- 3-PL IRT model models guessing
- This study introduces a new IRT model
  - » Accounts for guessing with only 2 item parameters

Three-Parameter IRT Model

- The 3-PL
  - » Estimates three parameters per item: a_i, b_i, and c_i
  - » More accurately portrays item response process than 2-PL if guessing is occurring
    - In theory, the model is preferred for data from multiple choice tests
  - » Will fit better than the 2-PL when guessing is occurring if it has enough data to estimate the additional parameter per item
    - In practice, there are some limitations
    - The c parameter is difficult to estimate
  - » Most common recommendation: use 2-PL—even if guessing is occurring— if the sample is less than 1000 examinees

Guessing Parameter Estimation Strategies

- A variety of strategies can be used to estimate the 3-PL with smaller samples
  - » Constrain all c's to be equal (estimate a common lower-asymptote)
  - » Set difficult-to-estimate c's equal to the mean of the estimable c's
  - » Put a Bayesian prior on c
- This study's strategy was to develop a new model that could provide a lower asymptote for an item response for small-scale assessments without having to estimate any additional item parameters

2-PA IRT Model

3-PL

- Models the conditional probability an examine *e* provides the correct response to item *i* as a function of continuous ability (θ_e) as (omitting the scaling constant 1.7):

$$P(X_{ei} = 1|\theta_e) = c_i + (1 - c_i)\frac{\exp(a_i(\theta_e - b_i))}{1 + \exp(a_i(\theta_e - b_i))}$$

2-PL

- The 3-PL where $\mathrm{c_i} = 0$ for every item:

$$P(X_{ei} = 1|\theta_e) = \frac{\exp(a_i(\theta_e - b_i))}{1 + \exp(a_i(\theta_e - b_i))}$$

2-PA

$$P(X_{ei} = 1|\theta_e) = \frac{\exp(\lambda_i + \exp(\delta_i\theta_e))}{1 + \exp(\lambda_i + \exp(\delta_i\theta_e))}$$

2-PA IRT Model

$$P(X_{ei} = 1|\theta_e) = \frac{\exp(\lambda_i + \exp(\delta_i \theta_e))}{1 + \exp(\lambda_i + \exp(\delta_i \theta_e))}$$

- δ_i is akin to discrimination α_i in the 2-PL and 3-PL models
  - » As δ_i increases, the slope of the item response function (on the logit scale) increases
- λ_i is the akin to the intercept $-a_i b_i$ in the 2-PL model
  - » As the intercept increases, the item becomes more difficult
  - » Also uniquely determines the lower asymptote:

$$\lim_{\theta \to -\infty} (P(X_{ei} = 1|\theta_e)) = \frac{\exp(\lambda_i)}{1 + \exp(\lambda_i)}$$

Empirical Data Analysis

Results Overview

- Data
- Estimation
  - Custom MCMC estimation algorithm written in Fortran
- Model Fit
  - Relative Fit (Deviance Information Criterion)
  - Absolute Fit (Yen's Q_1 (1981))
- Parameter estimates
  - Item Characteristic Curves (ICCs)
  - Ability distribution
  - Standard Error of Ability

Data

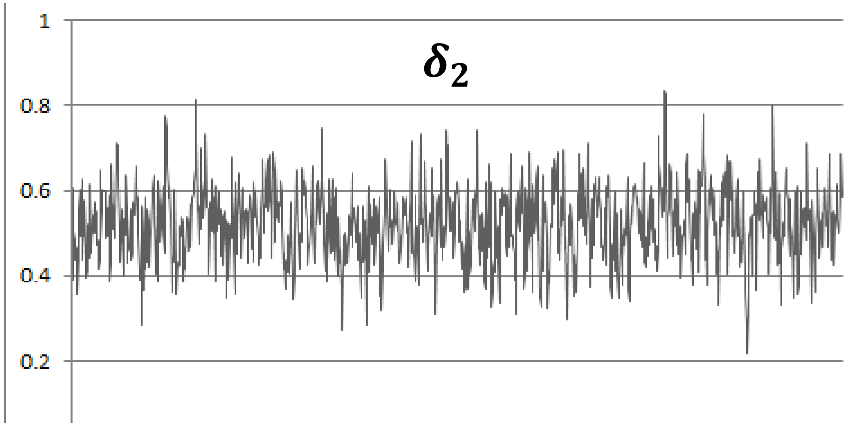
- Test of American History
  - » Administered to a random sample of 670 incoming freshman at a mid-sized Southeastern university
  - » 40 item multiple choice test
  - » 4 alternatives per item

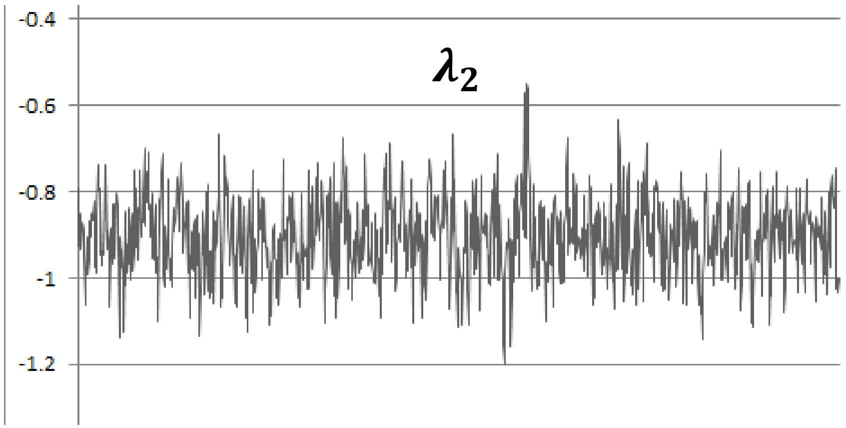
Estimation

- Convergence assessed by
  - » Gelman and Rubin's (1992) $\hat{R}$
    - % of converged parameters, by type

Model	Slope	Intercept	c
2-PA	97.5	95	-
3-PL	80	85	77.5
2-PL	85	90	-

  - » Chain Plots
    - Examples:





Model Fit

- Deviance Information Criterion (DIC) for relative model-data fit
  - » Appropriate criterion when MCMC estimation is used

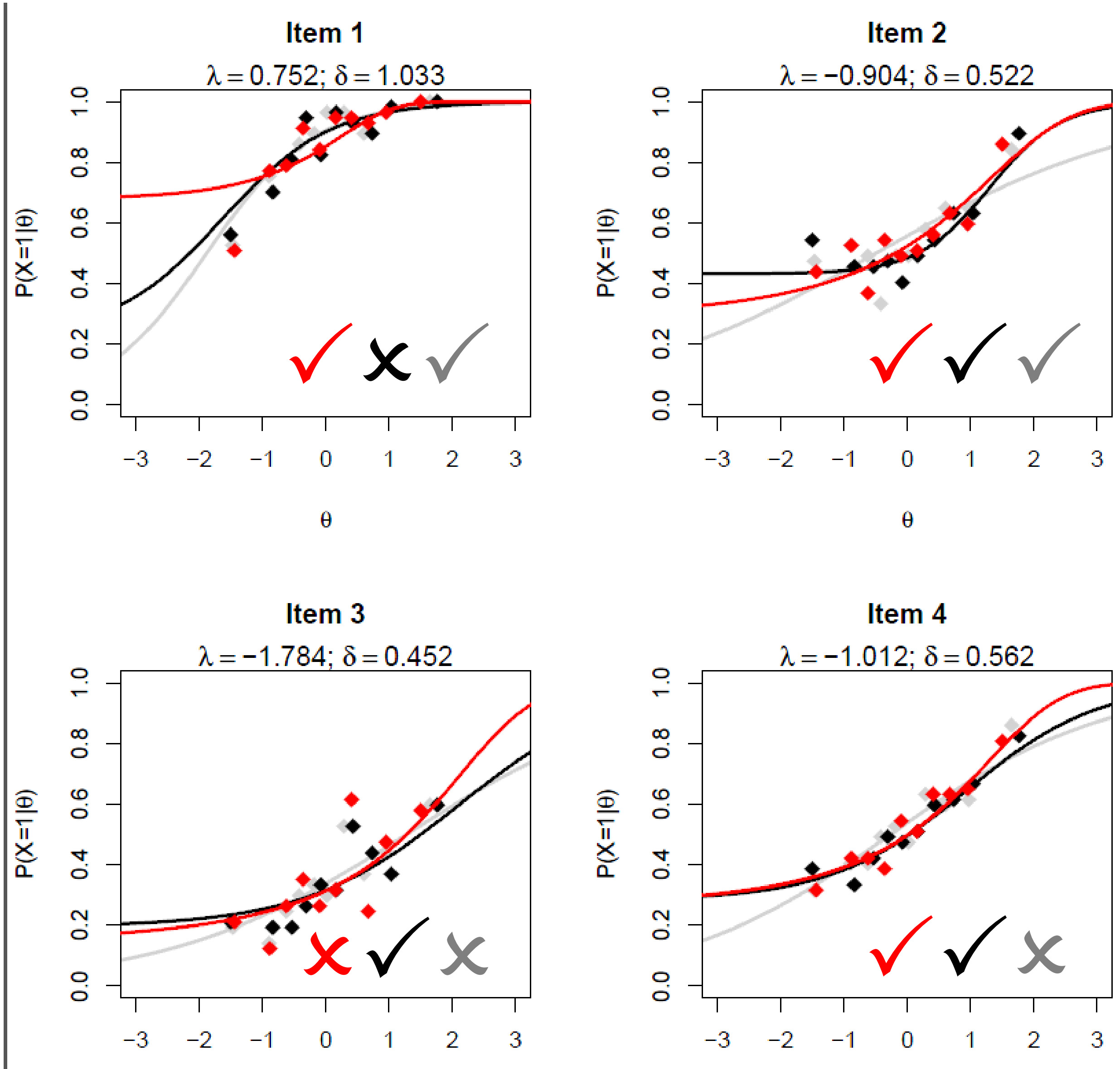
Model	Loglikelihood	Parameters	DIC
2-PA	-12494.4	80	23187.43
3-PL	-12446.3	120	**23024.31**
2-PL	-12465.5	80	23075.65

- Yen's Q_1 fit statistic for item-level absolute fit

Model	Number of Misfitting Items (p<.05)
2-PA	9
3-PL	9
2-PL	3

ICCs: First 4 Items

- 2-PA
- 3-PL
- 2-PL



Q_1 fit?

✓ = yes

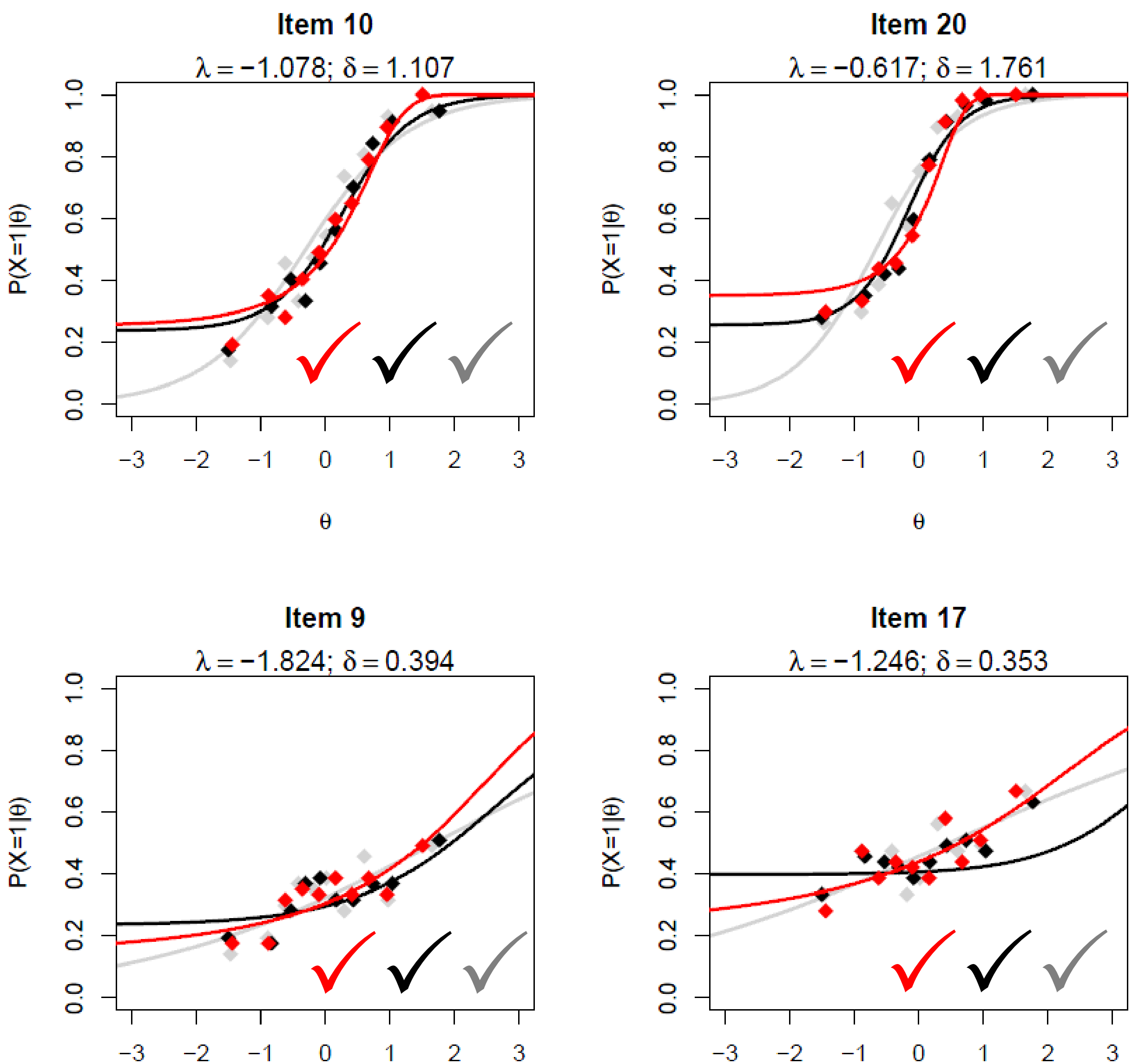
✗ = no

ICCs: High/Low Discrimination (δ)

2-PA

3-PL

2-PL



High δ

Low δ

Q_1 fit?

✓ = yes

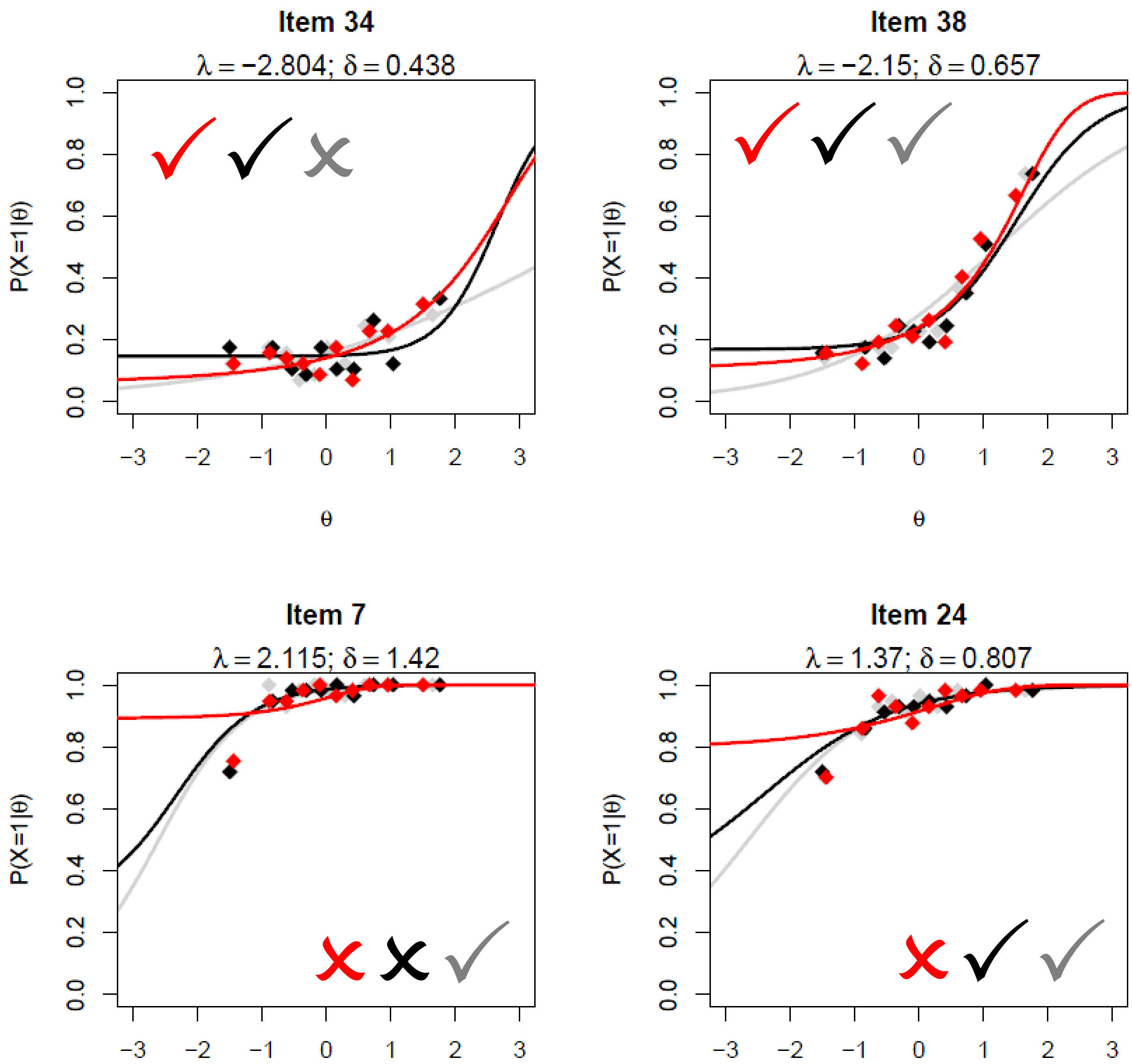
✗ = no

ICCs: High/Low Intercept (λ)

2-PA

3-PL

2-PL



High λ

Low λ

Q_1 fit?

✓ = yes

× = no

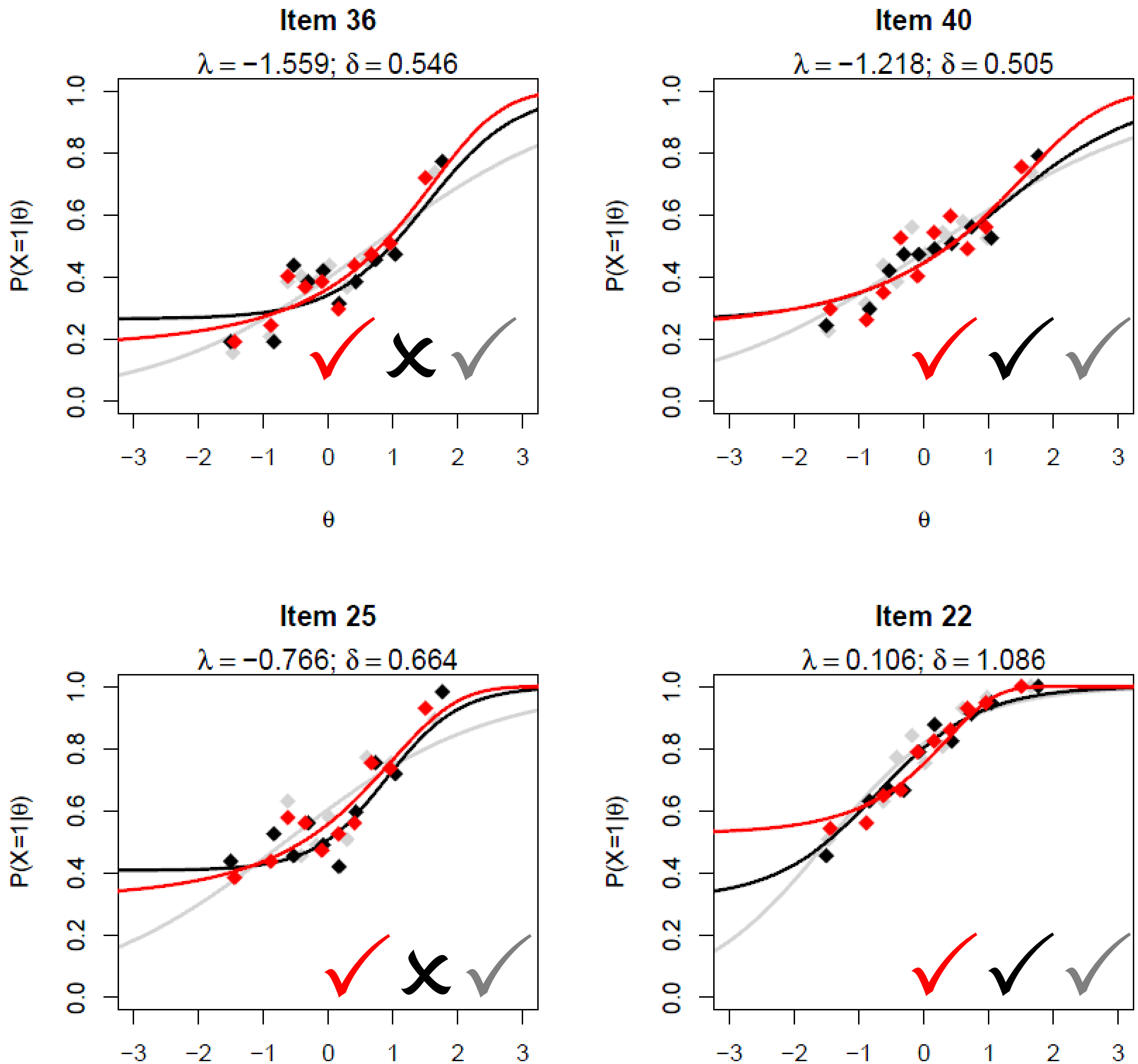
ICCs: Range of Asymptotes

- **2-PA** (red)
- **3-PL** (black)
- **2-PL** (grey)

Q_1 fit?

✓ = yes

✗ = no

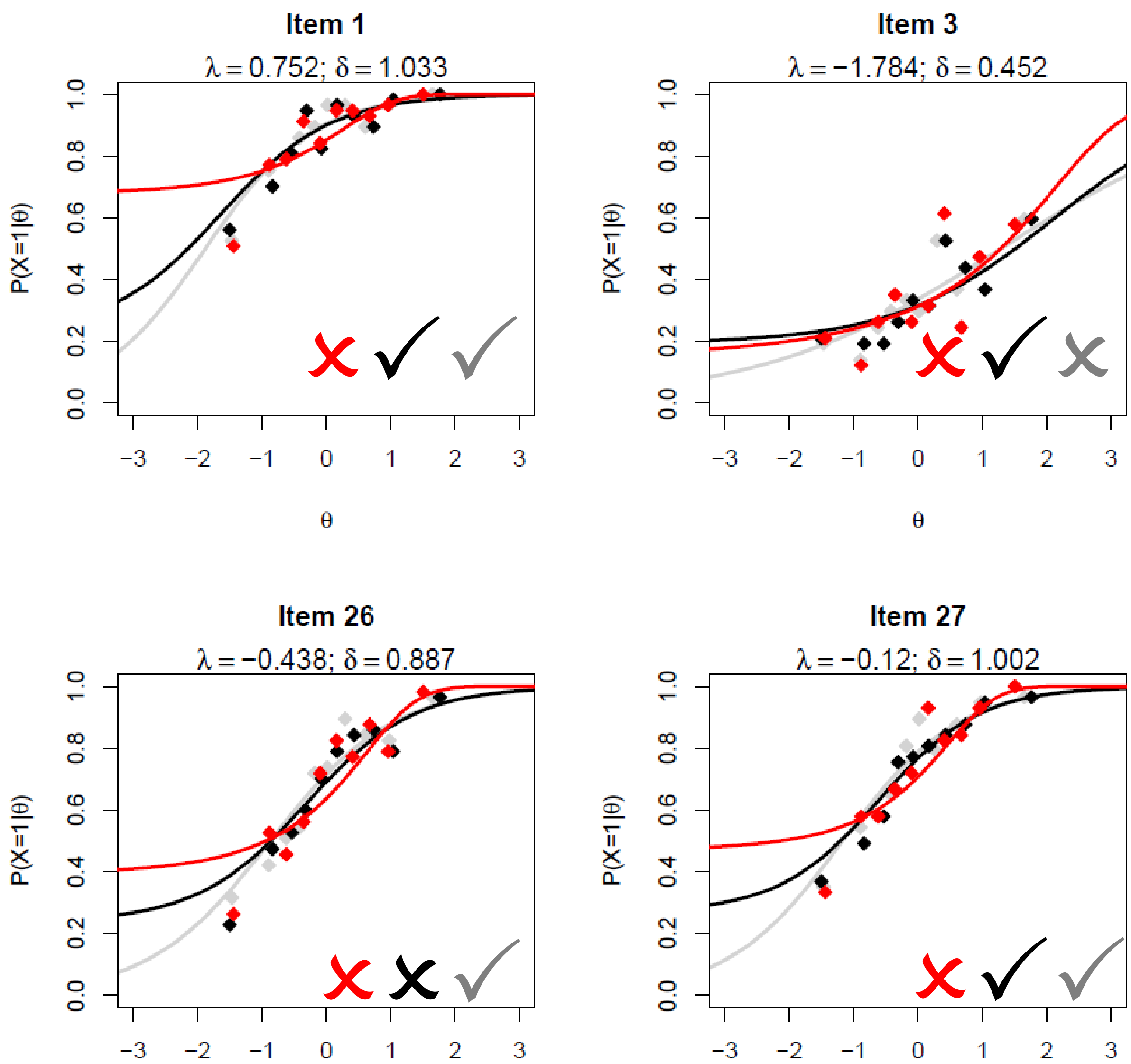


ICCs: Worst Fitting Items

2-PA

3-PL

2-PL

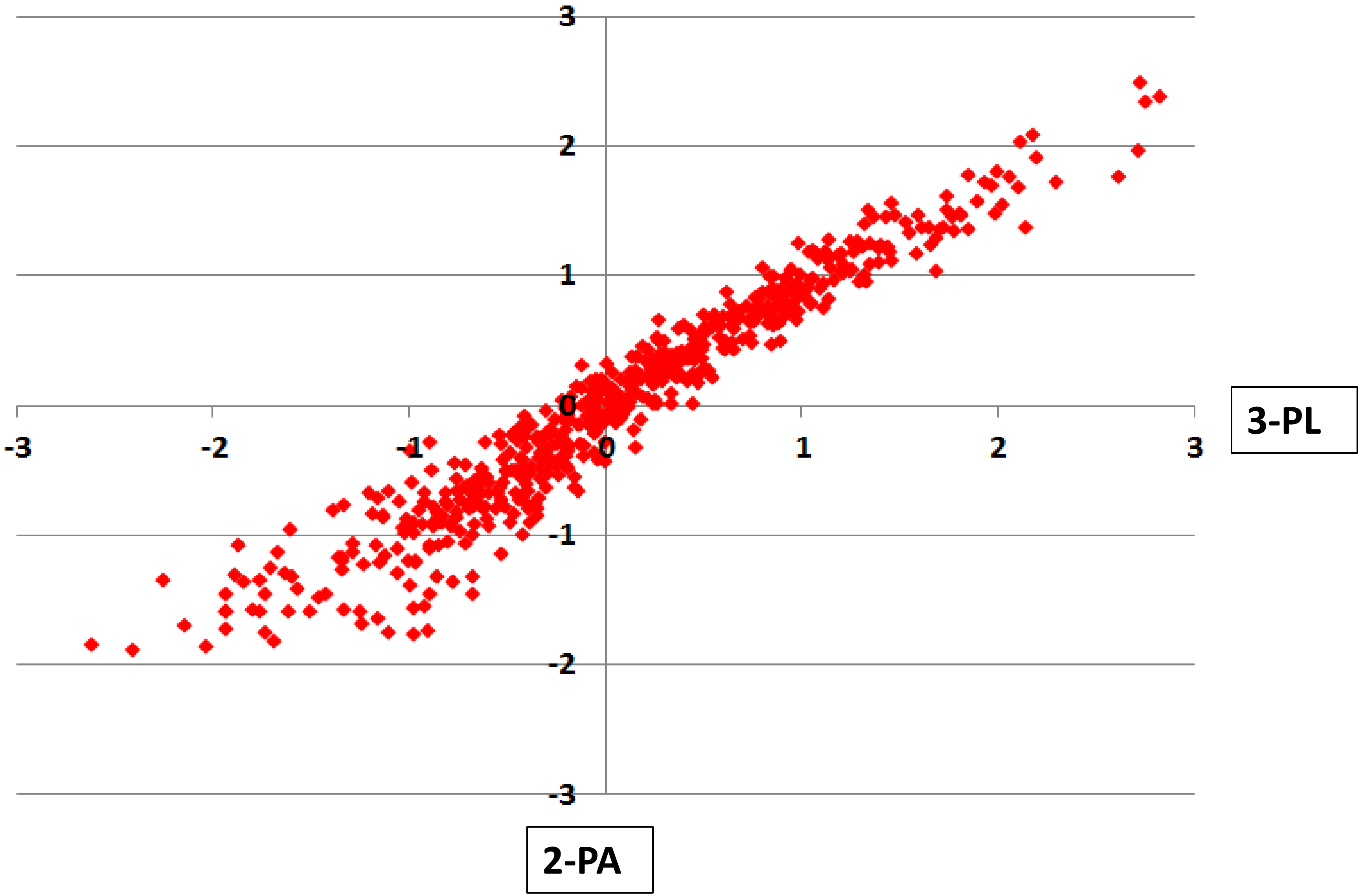


Q_1 fit?

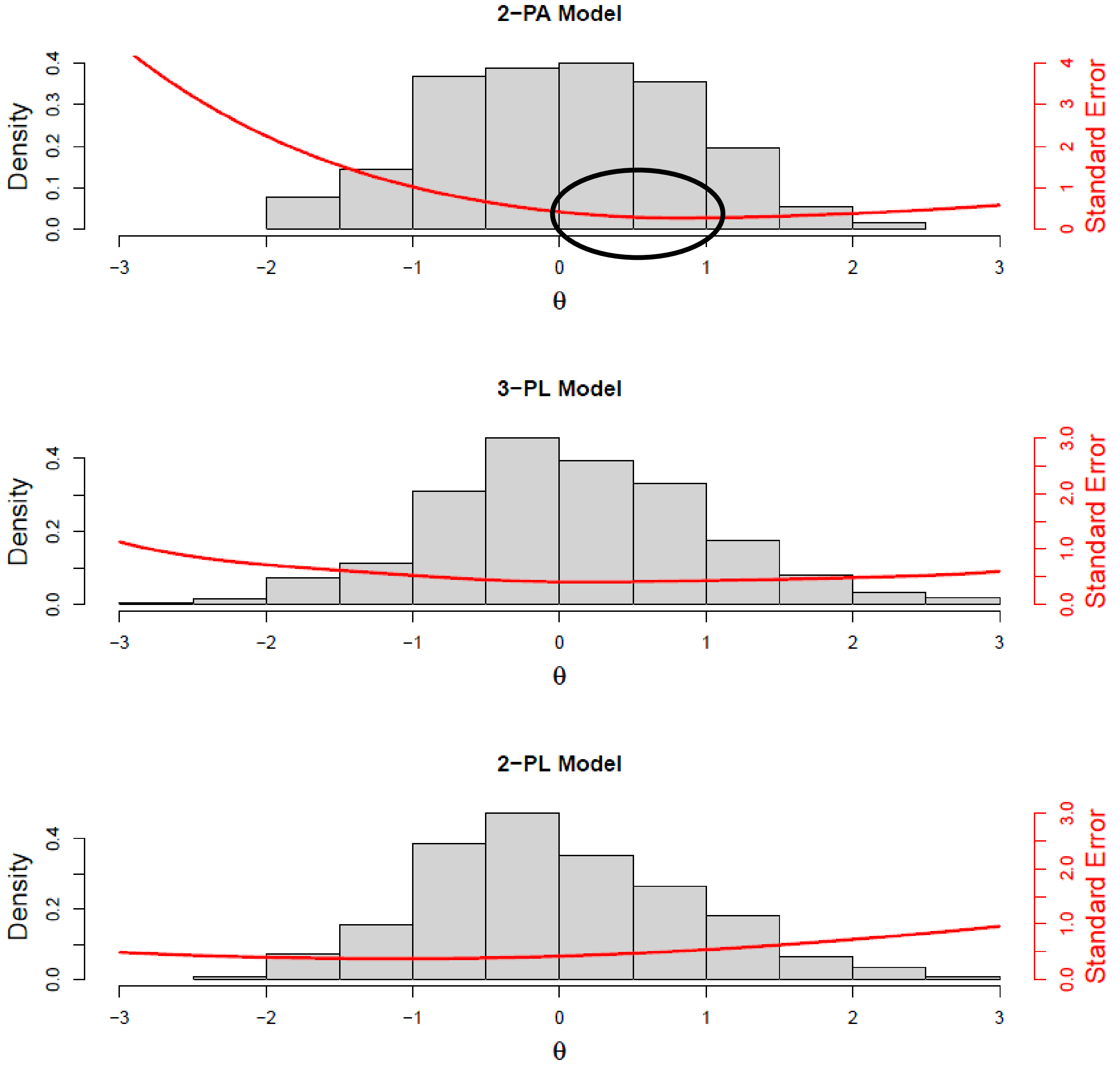
✓ = yes

✗ = no

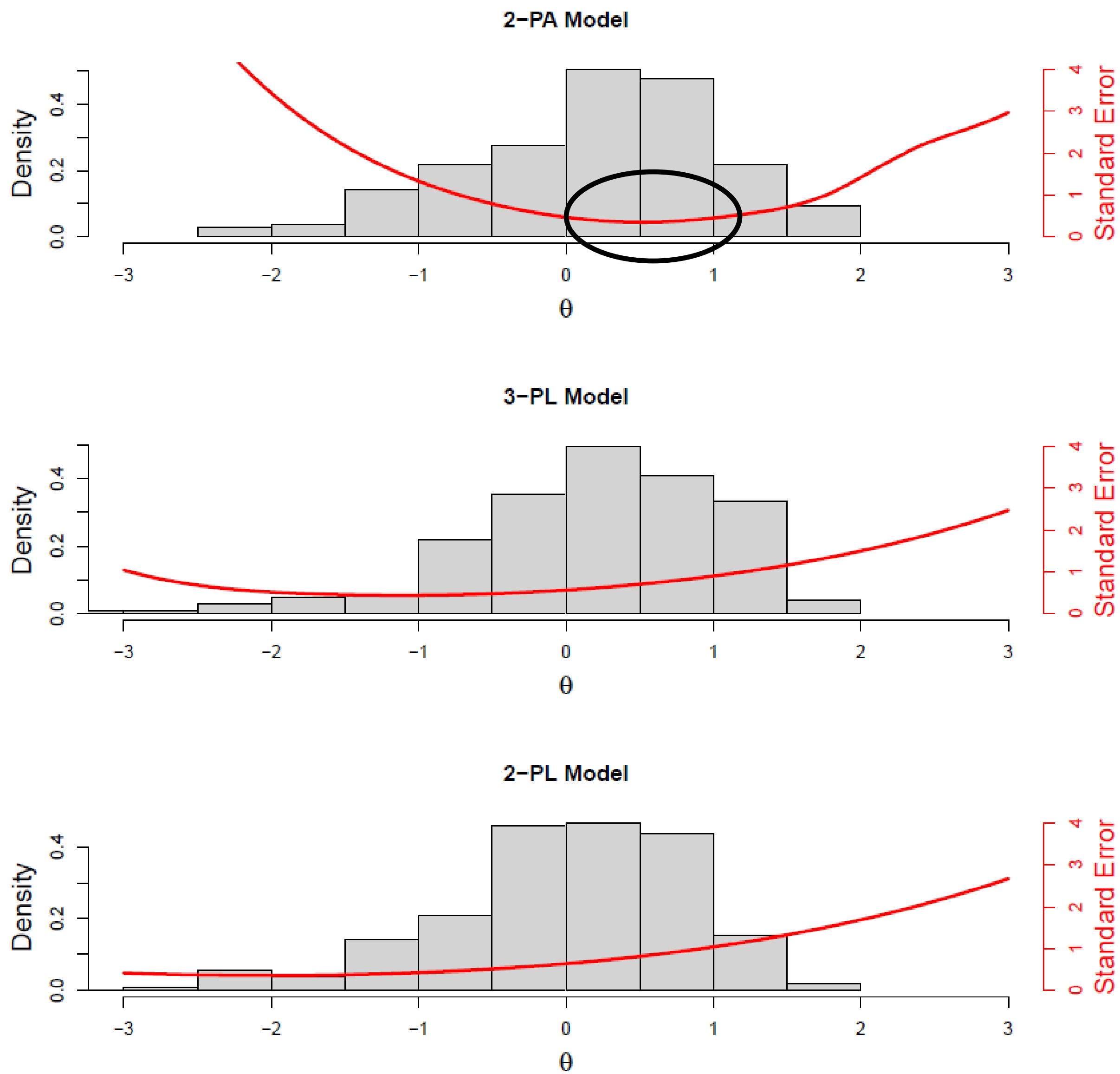
Ability Estimates



Ability and Standard Error of Ability



From a Two-Option Data Set



What's next?

- We are not sure where to go from here
  - » Simulation study may answer some of our concerns
- Remaining questions:
  - » Is ability on the same scale?
  - » Should we use a base other than e?
    - Perhaps 2?
    - We could estimate a base, β:

$$P(X_{ei} = 1|\theta_e) = \frac{\exp(\lambda_i + \beta^{\delta_i \theta_e}))}{1 + \exp(\lambda_i + \beta^{\delta_i \theta_e}))}$$

  - » Or we could estimate the base as as an item parameter, β_i:

$$P(X_{ei} = 1|\theta_e) = \frac{\exp(\lambda_i + {\beta_i}^{\delta_i \theta_e}))}{1 + \exp(\lambda_i + {\beta_i}^{\delta_i \theta_e}))}$$

    - Then we're back to 3 parameters!

Thank you!

If you have questions or comments, please feel free to email me:

bradshlp@jmu.edu