A new implementation of Item Factor Analysis: Accuracy, flexibility, and speed

Joshua N. Pritikin

Department of Psychology
University of Virginia

28 Mar 2013

Why a new implementation?

- ConQuest, $750
- IRTPRO, $495
- flexMIRT, $\approx$ $100 per year

Weaknesses:

- flexibility, customization
- Windows-centric
- non-zero $ barrier to entry

Also, current open-source software is fragmented and uncompetitive.

Goal

Meet or exceed capabilities of all commercial software

Where to start?

Until recently, marginal maximum likelihood (Bock & Aitkin, 1981) with adaptive Gauss-Hermite quadrature (Schilling & Bock, 2005) was the leading algorithm (Wirth & Edwards, 2007).

The IRTPRO manual recommended

- marginal maximum likelihood (Bock & Aitkin, 1981) with analytic dimension reduction (Cai, 2010b)
- Metropolis-Hastings Robbins-Monro (Cai, 2010a)

A starting point

- marginal maximum likelihood (Bock & Aitkin, 1981) with analytic dimension reduction (Cai, 2010b)
- Metropolis-Hastings Robbins-Monro (Cai, 2010a)

Which? *Both*

- Radically different approaches
- Similar performance on a broad class of problems

An open source implementation of MH-RM exists (Chalmers, 2012).

Start with marginal maximum likelihood.

Analytic dimension reduction

Necessary restriction of the factor covariance structure

$$\begin{pmatrix} \Sigma & \\ 0 & diag(\tau) \end{pmatrix}$$

where Σ is any covariance structure. For example,

$$\begin{pmatrix} 1 & & & \\ \rho & 1 & & \\ 0 & 0 & 1 & \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

All factors have unit variance. In a multigroup model, the primary factors can be correlated. The specific factors are mutually uncorrelated and uncorrelated with the primary factors (Cai, 2010b).

Analytic dimension reduction

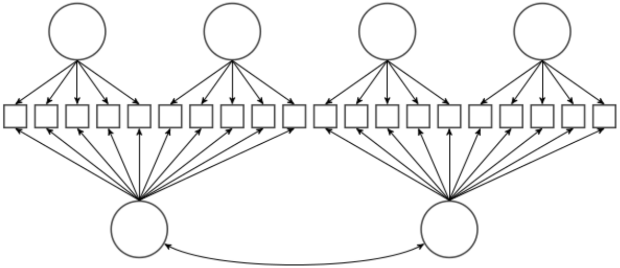


FIGURE 1.
A two-tier model for testlet-based assessments. The items measure 2 correlated primary dimensions and the test is based up of 4 testlets, creating 4 additional specific dimensions.

(Cai, 2010b, p. 583)

Analytic dimension reduction

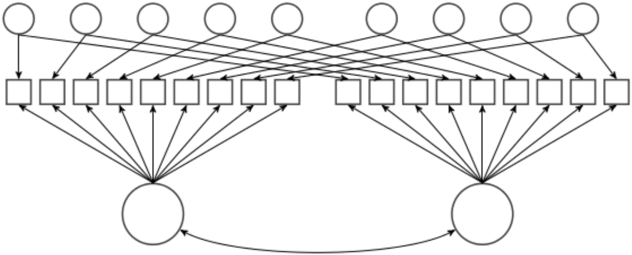


FIGURE 2.
A two-tier model for longitudinal item response data. The 9 items are administered to the same group of individuals at two time points, creating two time-specific primary dimensions and 9 item-specific doublets capturing residual correlations.

(Cai, 2010b, p. 584)

Analytic dimension reduction

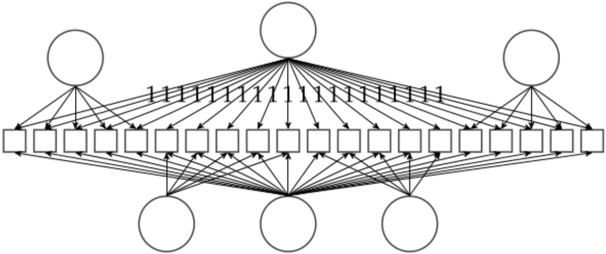


FIGURE 3.
Random intercept item bifactor model as a two-tier model. The inclusion of a random intercept that is orthogonal to the other factors is a distinguishing feature of this model. The general factor, on which all items load, and the random intercept are the primary dimensions.

(Cai, 2010b, p. 585)

Status

- R, $\approx$ 1300 lines
- C, $\approx$ 2350 lines

Implements marginal maximum likelihood (Bock & Aitkin, 1981) with analytic dimension reduction (Cai, 2010b)

OpenMx plugin, not yet merged

EAP person scores (Bock & Mislevy, 1982)

Item models: M3PL, GPCM, or invent your own

$S-\chi^2$ with polytomous extension (Orlando & Thissen, 2000)

Infit & outfit (Wright & Masters, 1982)

Plots: item characteristic curve (1d), data vs model, and information

$S-\chi^2$

(Explain on blackboard)

High-priority wish list

- Merge into OpenMx (Boker et al., 2011)
- Nominal model w/ analytic Newton-Raphson (Baker & Kim, 2004) (open source math exist; Chalmers, 2012)
- Item parameter standard errors (Cai, 2008)
- Multi-group (Cai, Yang, & Hansen, 2011)
- Structural latent trait model
- Hierarchical factor model

Item parameter standard errors

What do they mean?

$$\mathcal{I}(\hat{\theta}|Y_o) = \mathcal{I}_c(\hat{\theta}) - \mathcal{I}_m(\hat{\theta})$$

where the information matrix $\mathcal{I}$ at the item parameter vector estimate $\hat{\theta}$ conditional on the observed data Y_o is the complete information matrix $\mathcal{I}_c$ minus the information matrix of the missing data $\mathcal{I}_m$ (Cai, 2008).

Accuracy, performance

Let's take a look ...

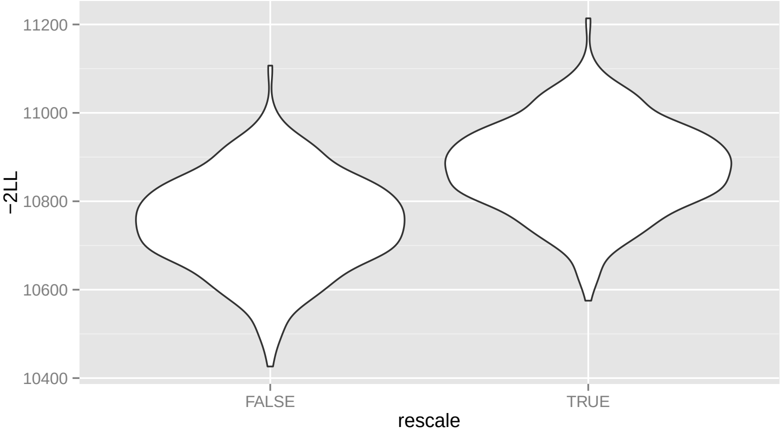
Standardize (rescale) the latent distribution or not?

- 20 2PL items
- 500 persons
- 17 point GH quadrature
- 0% missing
- 500 Monte Carlo replications M
- Rescale or not (Liu, Rubin, & Wu, 1998)

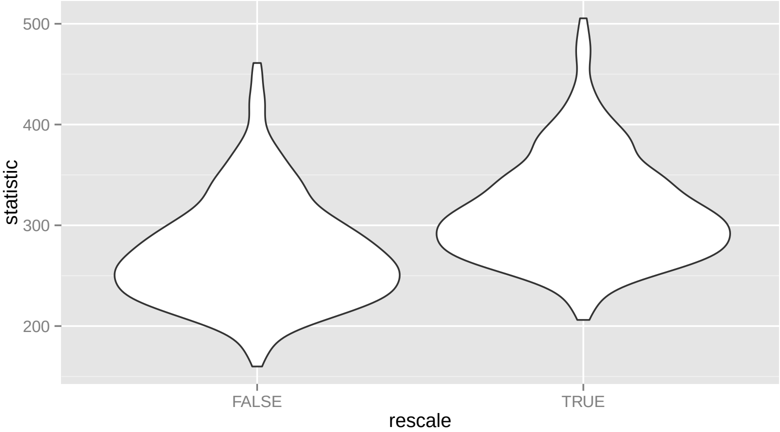
Examine -2LL, $S-\chi^2$, and bias.

$$\text{bias} = \hat{\theta} - \theta_{true} \text{ where } \hat{\theta} = \frac{1}{M}\sum_{m=1}^{M} \theta_m$$

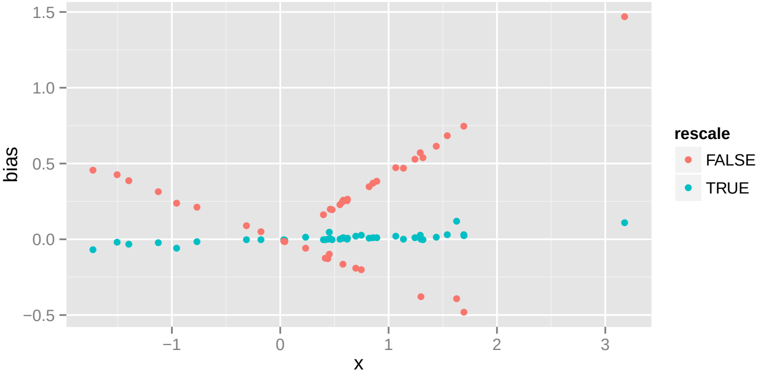
-2LL



$S-\chi^2$



Bias



Bias ranged from -0.0695 to 0.1188 with 50% of the bias between -0.0033 to 0.0174 with the median at 0.0056. For comparison, Winstep obtained bias ranging from 0.01 to 0.13 (Wang & Chen, 2005).

Which model? How many quadrature points?

- 20 2PL items
- 500 persons
- 0% missing
- GH quadrature
- This and subsequent studies are rescaled (Liu et al., 1998)

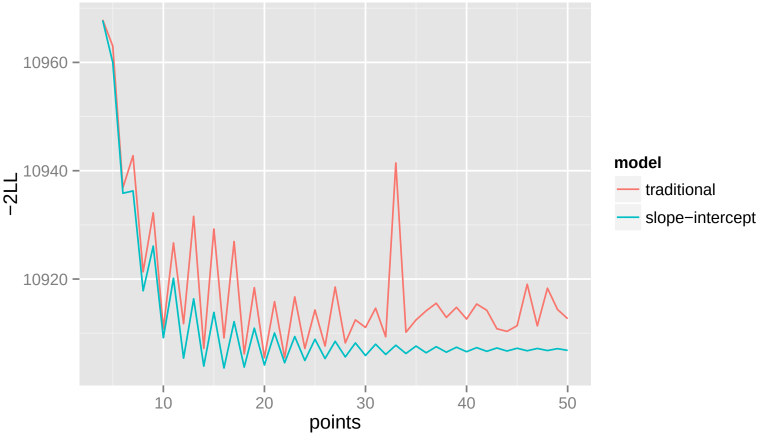
Traditional parameterization

$$\frac{1}{1+\exp(-a(\theta-b))}$$

or slope-intercept form

$$\frac{1}{1+\exp(-(a\theta+c))} \text{ where } b=\frac{c}{-a}$$

Likelihood by item model and quadrature points

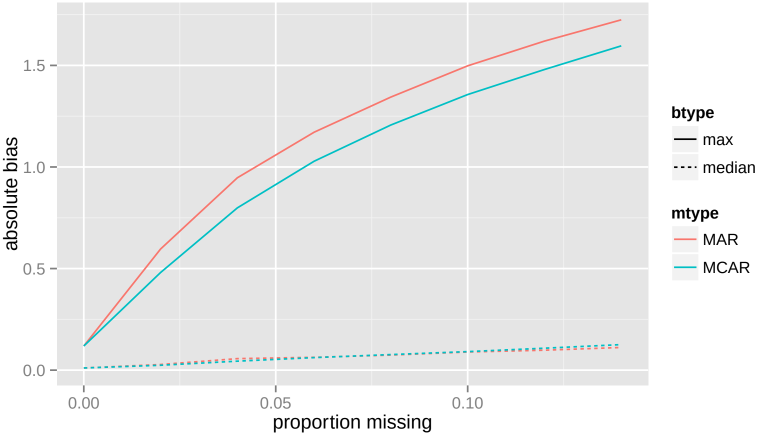


Bias by % missing

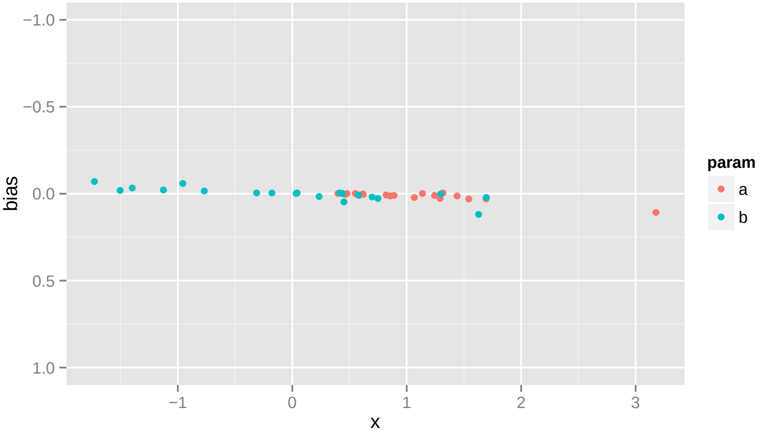
- 20 2PL items
- 500 persons
- 17 point GH quadrature
- 500 Monte Carlo replications

For the missing at random condition, data was replaced by NA depending on the first 5 items.

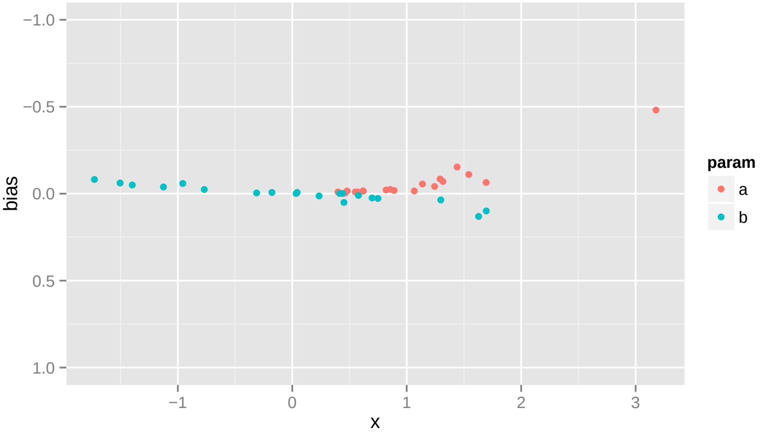
Bias by % missing



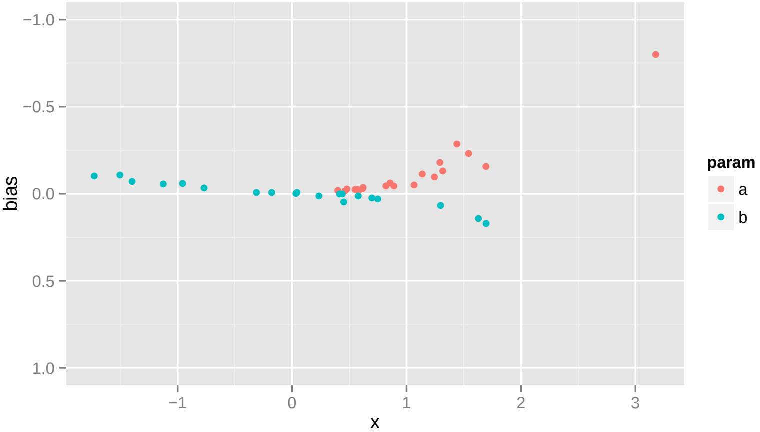
Bias by % missing, MCAR 0%



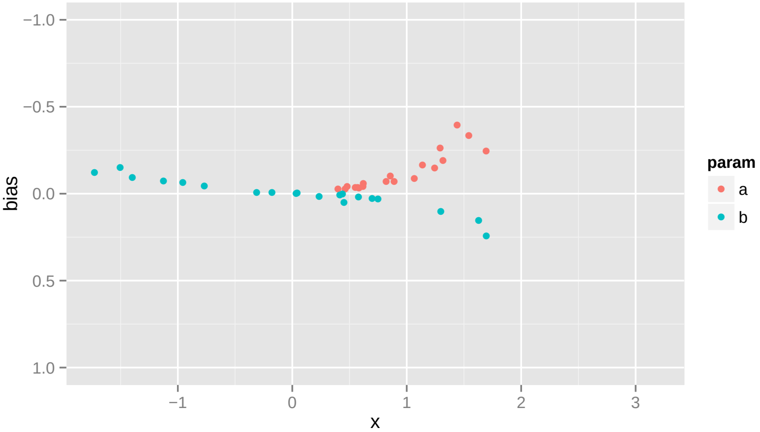
Bias by % missing, MCAR 2%



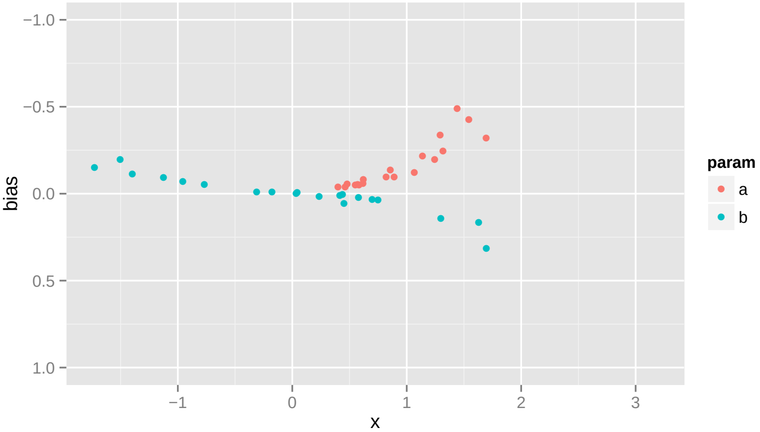
Bias by % missing, MCAR 4%



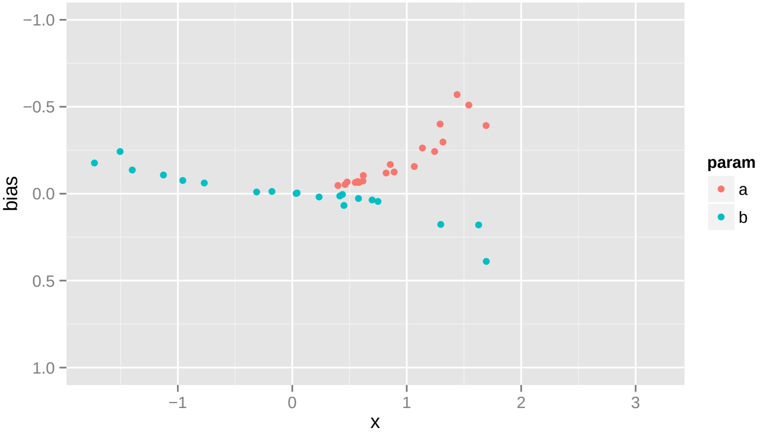
Bias by % missing, MCAR 6%



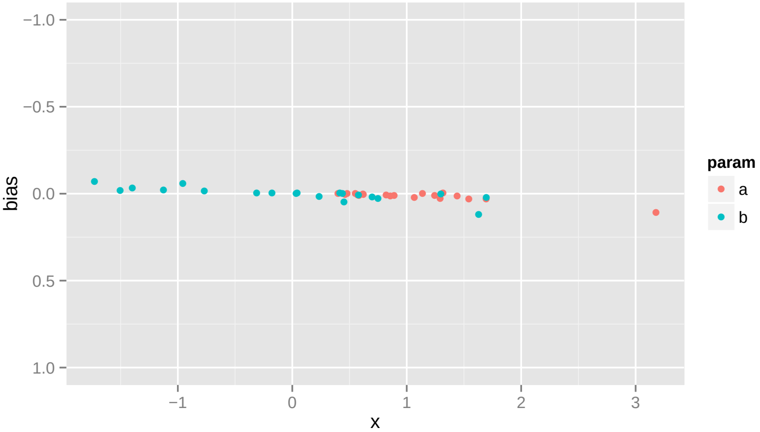
Bias by % missing, MCAR 8%



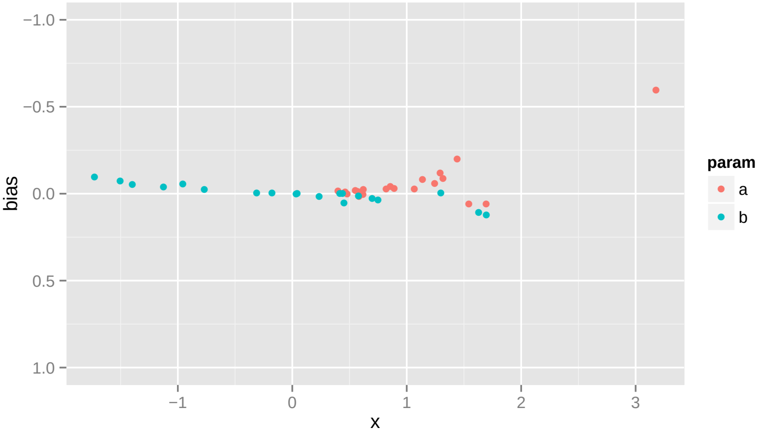
Bias by % missing, MCAR 10%



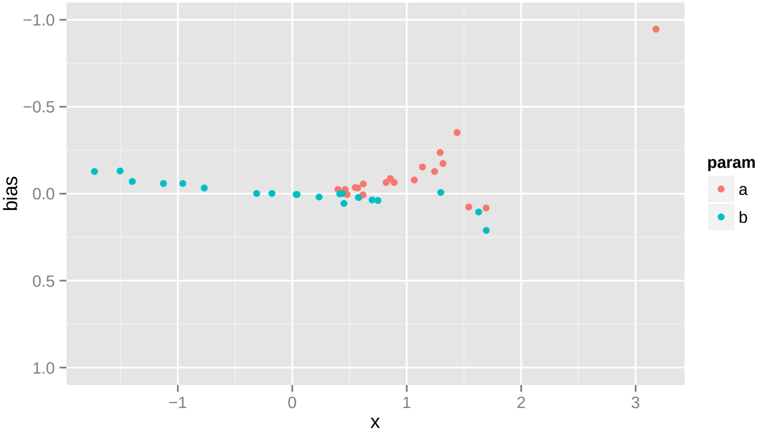
Bias by % missing, MAR 0%



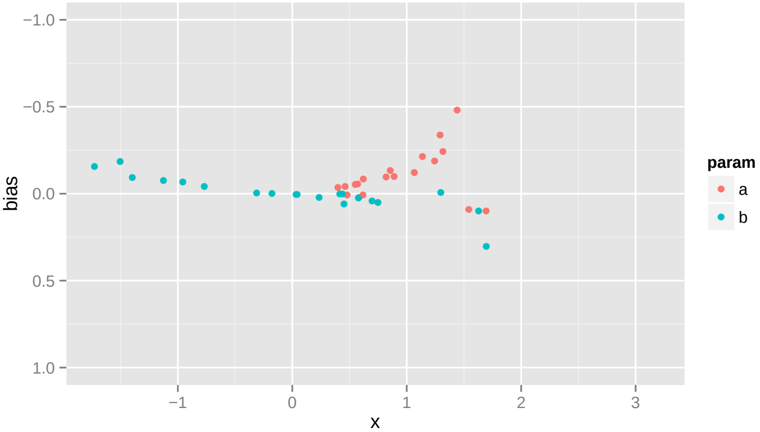
Bias by % missing, MAR 2%



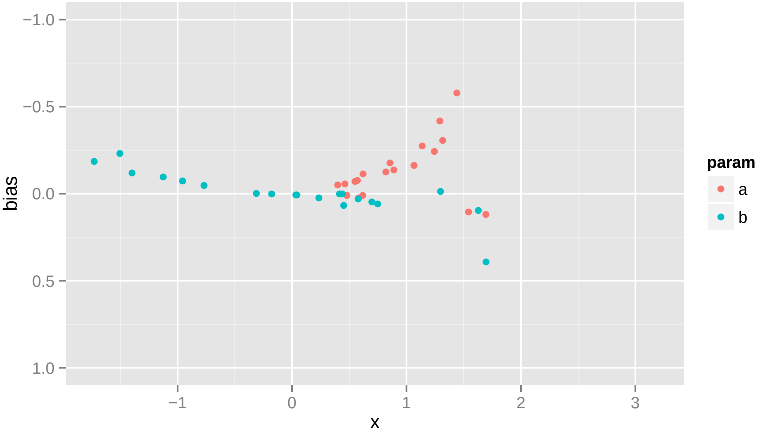
Bias by % missing, MAR 4%



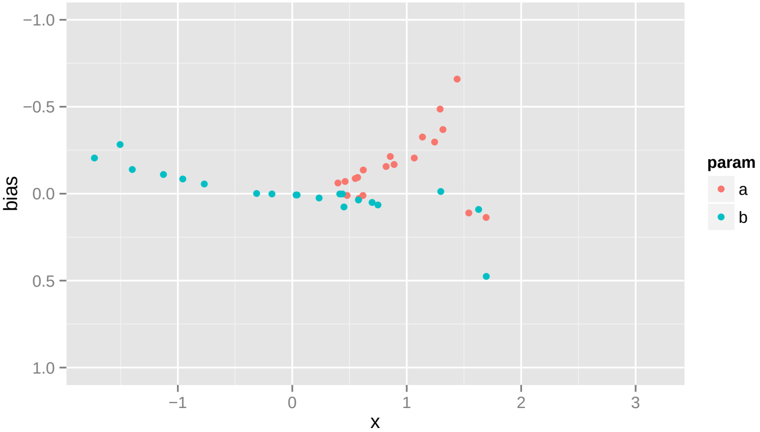
Bias by % missing, MAR 6%



Bias by % missing, MAR 8%



Bias by % missing, MAR 10%

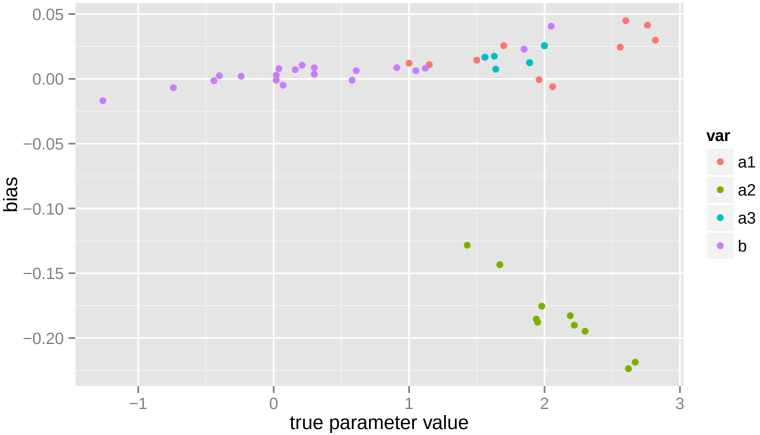


Cai (2010b) parameter recovery simulation

- 20 M2PL items
- 2 primary dimensions
- 4 specific dimensions formed by 4 pairs of item doublets
- $N = 500$ per replication
- 13 point GH quadrature[1]
- 500 Monte Carlo replications

[1] IRTPRO uses a 21 point equally spaced quadrature by default.

Cai (2010b) parameter recovery simulation



Note: slight pos bias; comparable to Cai (2010b) except for green; *very* slow

Acknowledgment

- Karen
- OpenMx development team
- Steve
- Timo
- UVa grad students

& colleagues who I forgot to mention

Questions?

Baker, F. B., & Kim, S. H. (2004). *Item response theory: Parameter estimation techniques* (2nd ed.). CRC Press.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*, 443-459.

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, *6*(4), 431–444.

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., … others (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika*, *76*(2), 306–317.

Cai, L. (2008). SEM of another flavour: Two new applications of the supplemented EM algorithm. *British Journal of Mathematical and Statistical Psychology*, *61*, 309–329.

Cai, L. (2010a). High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. *Psychometrika*, *75*, 33-57.

Cai, L. (2010b). A two-tier full-information item factor analysis model with applications. *Psychometrika*, *75*, 581–612. Retrieved from http://dx.doi.org/10.1007/s11336-010-9178-0 doi:

10.1007/s11336-010-9178-0

Cai, L., Yang, J. S., & Hansen, M. (2011). Generalized Full-Information Item Bifactor Analysis. *Psychological Methods*, *16*(3), 221–248.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29. Retrieved from `http://www.jstatsoft.org/v48/i06/`

Liu, C., Rubin, D. B., & Wu, Y. N. (1998). Parameter expansion to accelerate EM: the PX-EM algorithm. *Biometrika*, *85*(4), 755–770.

Orlando, M., & Thissen, D. (2000). Likelihood-Based Item-Fit Indices for Dichotomous Item Response Theory Models. *Applied Psychological Measurement*, *24*(1), 50–64.

Schilling, S., & Bock, R. D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika*, *70*(3), 533–555.

Wang, W.-C., & Chen, C.-T. (2005). Item parameter recovery, standard error estimates, and fit statistics of the Winsteps program for the family of Rasch models. *Educational and*

Psychological Measurement, *65*(3), 376–404.

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*(1), 58.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.