Unbelievably fast estimation of multilevel structural equation models

Joshua N. Pritikin

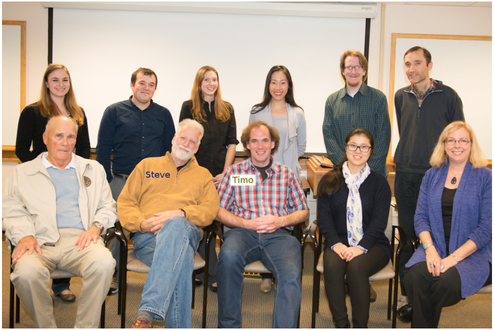
Department of Psychology
University of Virginia

Spring 2016

Abstract

The challenge of quick optimization of multilevel structural equation models (SEM) will be introduced. To provide context, multilevel SEM will be compared with the mixed model. Rampart, a novel method to simplify the multilevel SEM likelihood will be introduced, inspired by the fact that the multivariate normal density is transparent to orthogonal rotation. Assumptions and limitation of Rampart will be discussed.

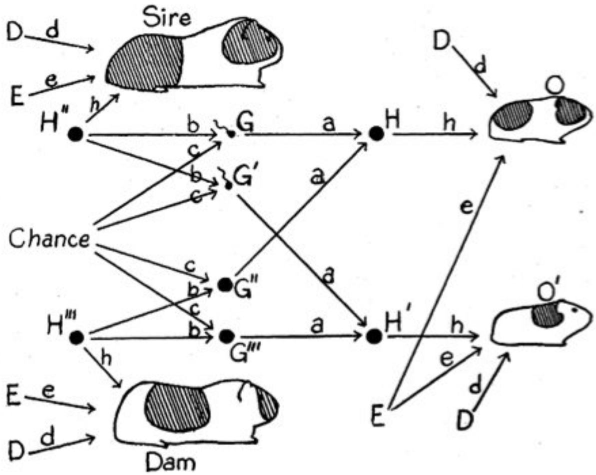
Acknowledgment



This research was aided by

- Timo von Oertzen & Steve Boker (University of Virginia)
- Tim Brick (Pennsylvania State University)
- OpenMx development team

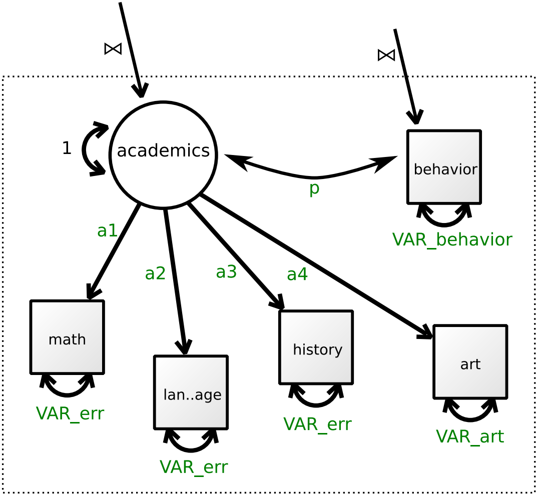
Structural equation models



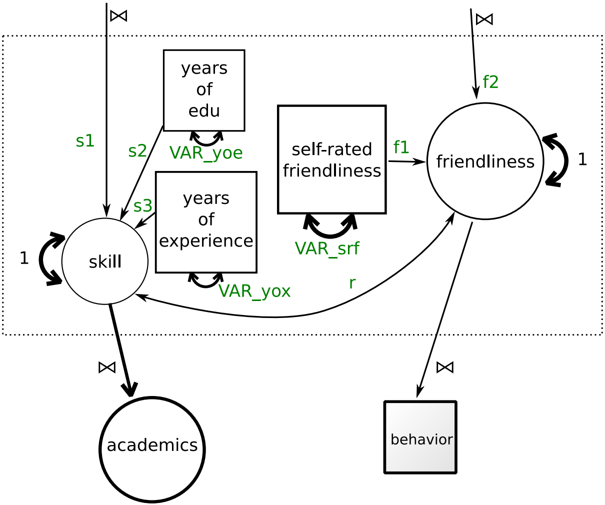
Multilevel structural equation models

A hypothetical example

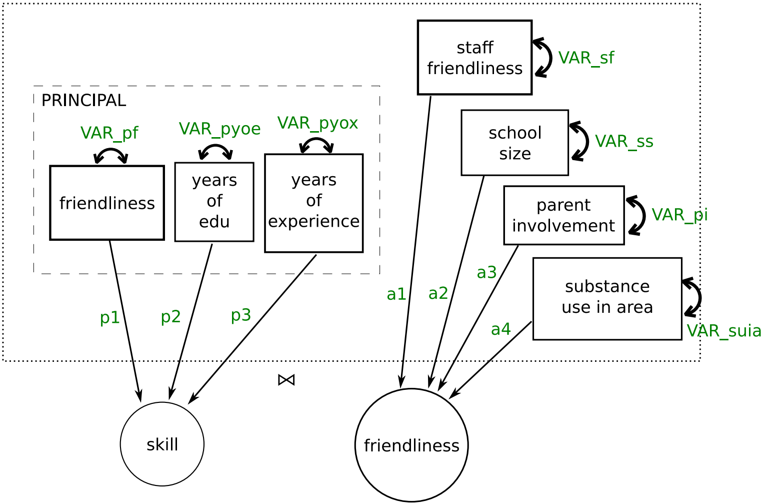
Student


⋈
⋈
1
academics
behavior
p
VAR_behavior
a1
a2
a3
a4
math
lan..age
history
art
VAR_err
VAR_err
VAR_err
VAR_art

Teacher


⋈
⋈
years
of
edu
f2
s1
s2
VAR_yoe
self-rated
friendliness
f1
friendliness
1
s3
years
of
experience
1
skill
VAR_srf
VAR_yox
r
⋈
⋈
academics
behavior

School



PRINCIPAL
VAR_pf
VAR_pyoe
VAR_pyox
friendliness
years of edu
years of experience
p1
p2
p3
staff friendliness
VAR_sf
school size
VAR_ss
parent involvement
VAR_pi
substance use in area
VAR_suia
a1
a2
a3
a4
⋈
skill
friendliness

Estimation time?

How long will this model take to estimate?

Roadmap

- What is hard about multilevel?
- mixed model + SEM = Relational SEM
- Rampart (a novel method that favorably transforms the problem)

Direct sum

$$\boldsymbol{B}_1 \bigoplus \boldsymbol{B}_2 = \begin{pmatrix} \boldsymbol{B}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B}_2 \end{pmatrix}$$

$$\bigoplus_{i=1}^{k} \boldsymbol{B}_i = \begin{pmatrix} \boldsymbol{B}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B}_2 & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{B}_k \end{pmatrix}.$$

- $\bigoplus$ stacks matrices in a block-diagonal arrangement
- Here we assume *nested multilevel* structure

Covariance

Suppose we build a covariance model $\boldsymbol{S}$ for a particular student. A classroom of s students will have covariance matrix

$$\boldsymbol{T} = \begin{pmatrix} \boldsymbol{T}_{1,1} & \boldsymbol{T}_{1,2} \\ \boldsymbol{T}_{2,1} & \bigoplus_{i=1}^{s} \boldsymbol{S}_i \end{pmatrix}.$$

Sparseness pattern

Bottleneck in model evaluation

- Covariance matrix can become very large
- Inverting the covariance is expensive, $O(N^3)$
- Sparse matrix operations help, but not enough
- Conclusion: impractical without cleverness

Structural equation model

What is a SEM?

It's similar to regression. But how?

Structural equation model, 1st moment

$$\text{Regression is } \boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{e}$$
$$\text{SEM is } \boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{e}$$

For $\boldsymbol{y}$ observations, X covariates/predictors, $\boldsymbol{\beta}$ constant coefficients, $\boldsymbol{e}$ residuals

Structural equation model, 2nd moment

Regression

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{e}$$
$$\boldsymbol{e} \sim \mathcal{N}(., \sigma^2 I)$$

SEM

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{e}$$
$$(X, \boldsymbol{e}) \sim \mathcal{N}(., \Sigma(\boldsymbol{\theta}))$$

For $\boldsymbol{y}$ observations, X covariates/predictors, $\boldsymbol{\beta}$ constant coefficients, $\boldsymbol{e}$ residuals, variance σ^2, parameters $\boldsymbol{\theta}$, covariance Σ

Mixed model

lme4 (Bates, Mächler, Bolker, & Walker, 2015)

Multilevel or crossed regression

Fast

How does it work?

Mixed model, 1st moment

$$\boldsymbol{Y} = \underbrace{\boldsymbol{X\beta}}_{\text{constant}} + \underbrace{\boldsymbol{Zu} + \boldsymbol{e}}_{\text{varying}}.$$

- column vector of observations $\boldsymbol{Y}$
- covariates $\boldsymbol{X}$ associated with constant coefficients $\boldsymbol{\beta}$
- covariates $\boldsymbol{Z}$ associated with varying coefficients $\boldsymbol{u}$
- column vector of residuals $\boldsymbol{e}$

$$\mathrm{E}\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{e}\end{pmatrix} = \begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}$$

Mixed model, 2nd moment

$$\boldsymbol{Y} = \underbrace{\boldsymbol{X}\boldsymbol{\beta}}_{\text{constant}} + \underbrace{\boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}}_{\text{varying}}.$$

$$\mathrm{Cov}\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{e}\end{pmatrix} = \begin{pmatrix}\boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{R}\end{pmatrix}$$

Mixed model, unconditional distribution

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$$

where

$$\boldsymbol{e} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^T + \boldsymbol{R})$$

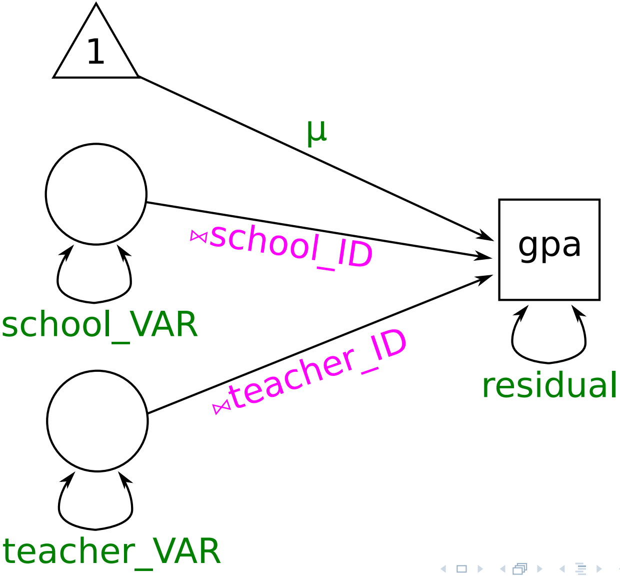
The formulation I showed earlier is actually conditional on a particular realization of the varying coefficients u.

Mixed model, example

```
lmer(gpa ~ 1 + (1 | school) + (1 | school:teacher), ...)
```

- Intercept-only model with 3 levels
- The expression after the vertical bar indicates the partitioning design (like conditional probability)
- 4 free parameters are estimated: the grand intercept; 2 variances, one for each varying coefficient; and the residual variance
- Efficient

How does this look in SEM?



What is $\bowtie$?

Let R and S be tables (or data frames) that contain rows.

$$R \bowtie (F)\ S \equiv \{r \cup s \wedge r \in R \wedge s \in S \wedge F(r \cup s)\}$$

where F is a boolean valued function.

Without loss of generality, here F is whether primary and foreign keys match. We will omit F and write $\bowtie (k)$ where k is the name of the key.

What is ⋈? Example

Employee	Dept
Harry	Sales
Sally	Finance
George	Finance
Harriet	Sales

Dept	Manager
Sales	George
Finance	Harriet
Production	Charles

Employee ⋈(Dept) Manager

Employee	Dept	Manager
Harry	Sales	George
Sally	Finance	Harriet
George	Sales	George
Harriet	Finance	Harriet

Which is easier to understand?

Time point t, individual i, cluster j.

y_{tij} : individual-level, outcome variable
a_{1tij} : individual-level, time-related variable (age, grade)
a_{2tij} : individual-level, time-varying covariate
x_{ij} : individual-level, time-invariant covariate
w_j : cluster-level covariate

Three-level analysis (Mplus considers Within and Between)

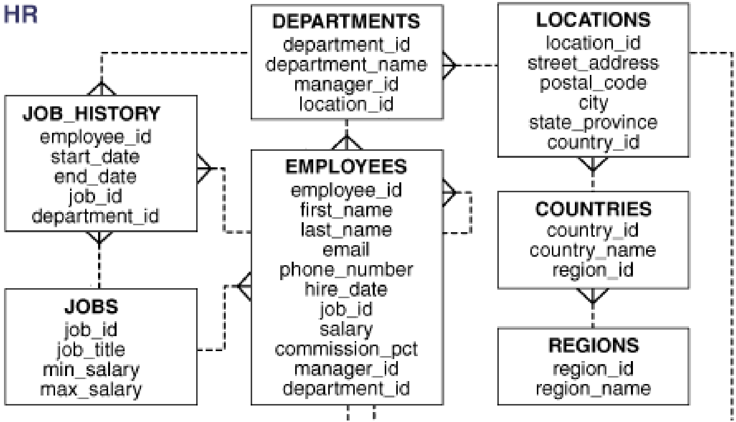
$$\textit{Level 1 (Within)} : y_{tij} = \pi_{0ij} + \pi_{1ij}\, a_{1tij} + \pi_{2tij}\, a_{2tij} + e_{tij}, \tag{1}$$

$$\textit{Level 2 (Within)} : \begin{cases} \pi_{0ij} = \beta_{00j} + \beta_{01j}\, x_{ij} + r_{0ij}, \\ \pi_{1ij} = \beta_{10j} + \beta_{11j}\, x_{ij} + r_{1ij}, \\ \pi_{2tij} = \beta_{20tj} + \beta_{21tj}\, x_{ij} + r_{2tij}. \end{cases} \tag{2}$$

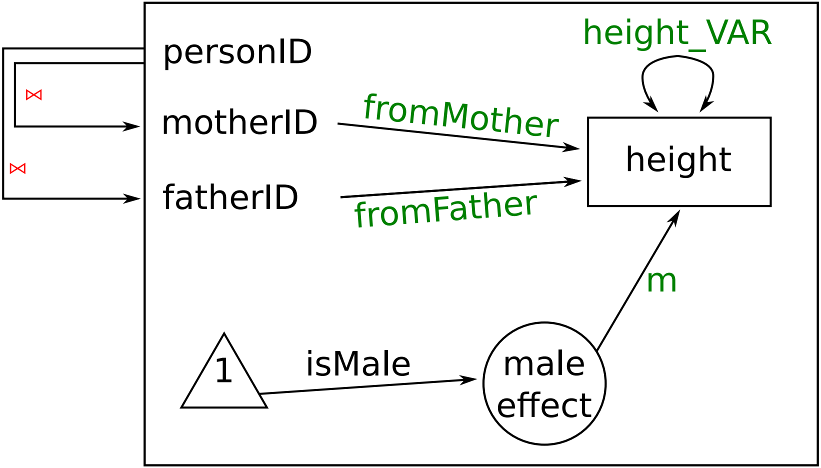
$$\textit{Level 3 (Between)} : \begin{cases} \beta_{00j} = \gamma_{000} + \gamma_{001}\, w_j + u_{00j}, \\ \beta_{10j} = \gamma_{100} + \gamma_{101}\, w_j + u_{10j}, \\ \beta_{20tj} = \gamma_{200t} + \gamma_{201t}\, w_j + u_{20tj}, \\ \beta_{01j} = \gamma_{010} + \gamma_{011}\, w_j + u_{01j}, \\ \beta_{11j} = \gamma_{110} + \gamma_{111}\, w_j + u_{11j}, \\ \beta_{21tj} = \gamma_{2t0} + \gamma_{2t1}\, w_j + u_{2tj}. \end{cases} \tag{3}$$

69

A relational schema



Autoregressive tree (pedigree)



Summary: $\bowtie$ vs conditional probability

Data is joined ($\bowtie$)

Conditional probability is an ingredient in multilevel models, not data

Join commutes, conditional probability doesn't

Mixed model

Limitations:

- missing data $\longrightarrow$ row-wise deletion
- only the lowest level unit has observations
- multivariate (more than one outcome) is very awkward

Sufficient statistic approach

Suppose we have data of N independent observations of K-variate units. Let μ and Σ be the model expected mean vector and covariance matrix, respectively. Let m and S be the mean vector and covariance matrix of the data, respectively.

$$-2 \log L(\text{data}|\theta) = N(K \log(2\pi) + \log(|\Sigma|) + \text{tr}(\Sigma^{-1} S) + \mu^T \Sigma^{-1}(\mu - 2m))$$

Maximum covariance dimension is K.

Sparseness pattern

Uncompressed likelihood

Suppose we have N observations consisting of data vector x. Let μ and Σ be the model expected mean vector and covariance matrix, respectively.

$$-2 \log L(\text{data}|\theta) = N \log(2\pi) + \log(|\Sigma|) + (\mu - x)^T \Sigma^{-1} (\mu - x)$$

Maximum covariance dimension is N.

Rampart

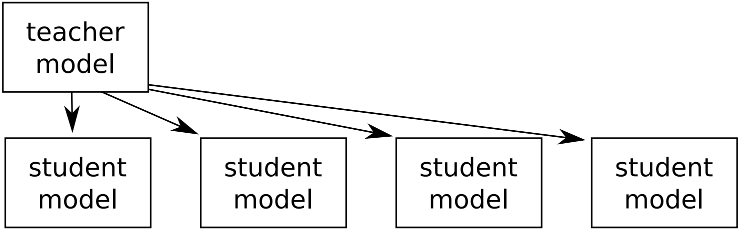
Covariance becomes very large. What to do?

SEMs are specified using the RAM parameterization:

$$\boldsymbol{\mu} = \boldsymbol{F}(\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{M}$$
$$\boldsymbol{\Sigma} = \boldsymbol{F}(\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{S}(\boldsymbol{I} - \boldsymbol{A})^{-T}\boldsymbol{F}^T$$

$\boldsymbol{A}$, $\boldsymbol{S}$, $\boldsymbol{F}$, and $\boldsymbol{M}$ are used for what?

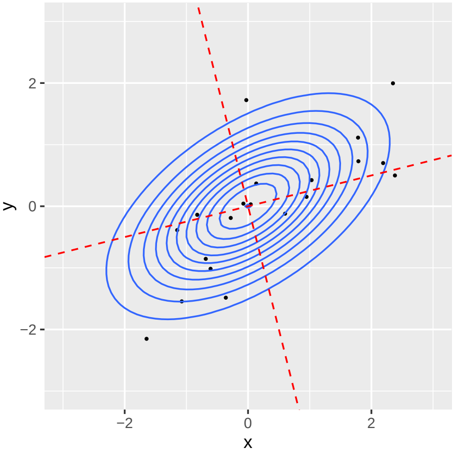
RAM's $\boldsymbol{A}$ matrix



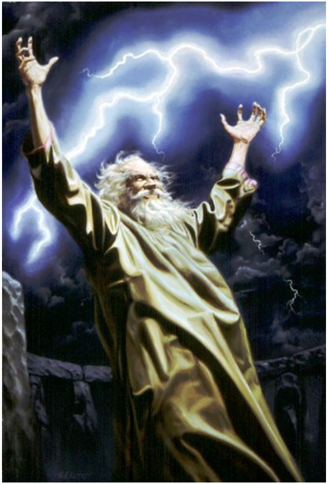
RAM's $\boldsymbol{A}$ matrix

	T	S1	S2	S3	S4
T	0	0	0	0	0
S1	1	0	0	0	0
S2	1	0	0	0	0
S3	1	0	0	0	0
S4	1	0	0	0	0

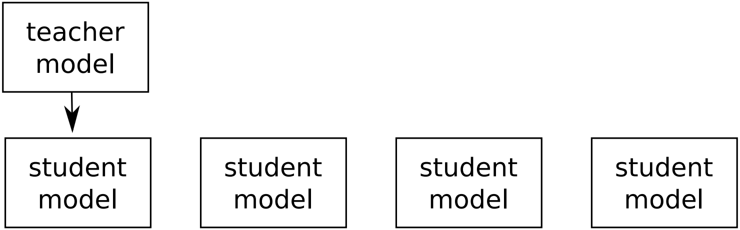
Orthogonal or axis rotation



Call forth the sublime orthogonal rotation



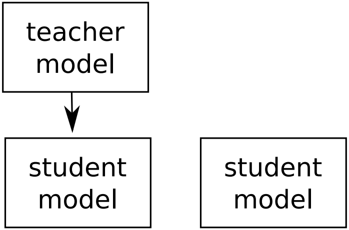
Rampart transformed



Rampart transformed

	T	S1	S2	S3	S4
T	0	0	0	0	0
S1	$\sqrt{4}$	0	0	0	0
S2	0	0	0	0	0
S3	0	0	0	0	0
S4	0	0	0	0	0

Rampart transformed



Rampart algebratically

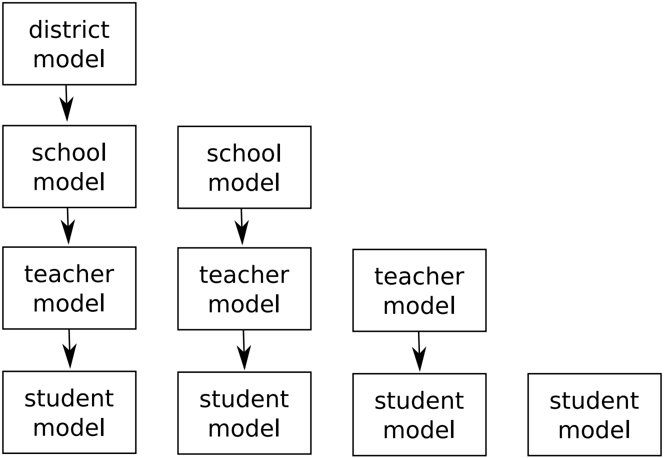
- Upper level has K outgoing regressions
- Lower level has M incoming regressions with data D
- Find an orthogonal matrix $Q \in \mathbb{R}^{M \times M}$ such that the lower $M - K$ rows of QA are zero.
- Define new model A' as the first K rows of $Q^T A$
- Define new lower level dataset $D' = Q^T D$
- Proceed with optimization as usual

Rampart geometry

$$\begin{bmatrix} 1.00 & 6.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 1.00 & -1.00 & 5.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 1.00 & -1.00 & -1.00 & 4.00 & 0.00 & 0.00 & 0.00 \\ 1.00 & -1.00 & -1.00 & -1.00 & 3.00 & 0.00 & 0.00 \\ 1.00 & -1.00 & -1.00 & -1.00 & -1.00 & 2.00 & 0.00 \\ 1.00 & -1.00 & -1.00 & -1.00 & -1.00 & -1.00 & 1.00 \\ 1.00 & -1.00 & -1.00 & -1.00 & -1.00 & -1.00 & -1.00 \end{bmatrix}$$

Use QR decomposition to scale to an orthogonal rotation

Rampart, apply recursively

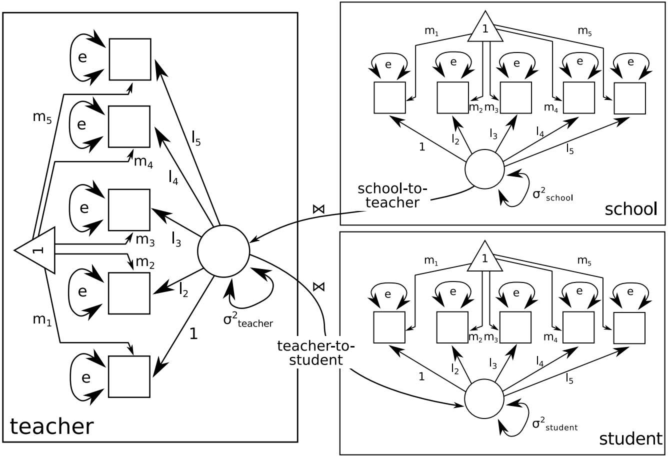


Rampart historical lineage

Who done it?

- summer of 2012: idea conceived by Timo von Oertzen, Steven M. Boker, and Timothy R. Brick, inspired by von Oertzen and Hackett (submitted)
- spring 2013: prototyped in OpenMx (by me)
- *completed predissertation on a different topic*
- spring 2016: optimized implementation in OpenMx (again by me)

Does it work?


e
e
e
e
e
m5
m4
m3
m2
m1
l5
l4
l3
l2
1
1
σ2 teacher
teacher
⋈
⋈
school-to-
teacher
teacher-to-
student
1
m1
m5
e
e
e
e
e
m2
m3
m4
l2
l3
l4
l5
1
σ2 school
school
1
m1
m5
e
e
e
e
e
m2
m3
m4
l2
l3
l4
l5
1
σ2 student
student

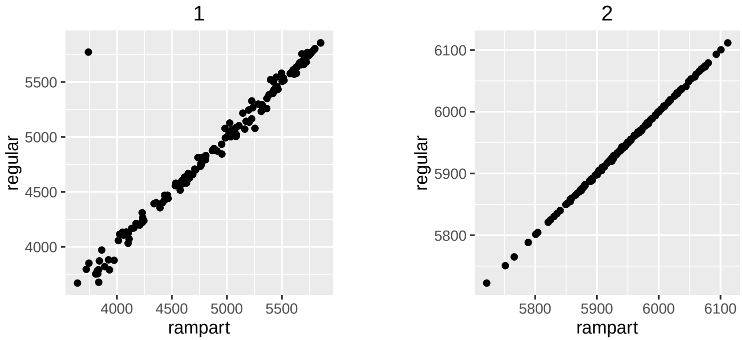
Conditions

- 11 parameters per level, 2 between level regressions (total 35)
- 1st student indicator was set to missing with 20% probability
- 2 sets of true parameters ($\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$)
- Parameter $\boldsymbol{\theta}_1$: 7 schools, 38 teachers, and 293 students
- Parameter $\boldsymbol{\theta}_2$: 7 schools, 37 teachers, and 296 students
- 200 Monte Carlo replications for each condition (Algorithm $\times$ θ)

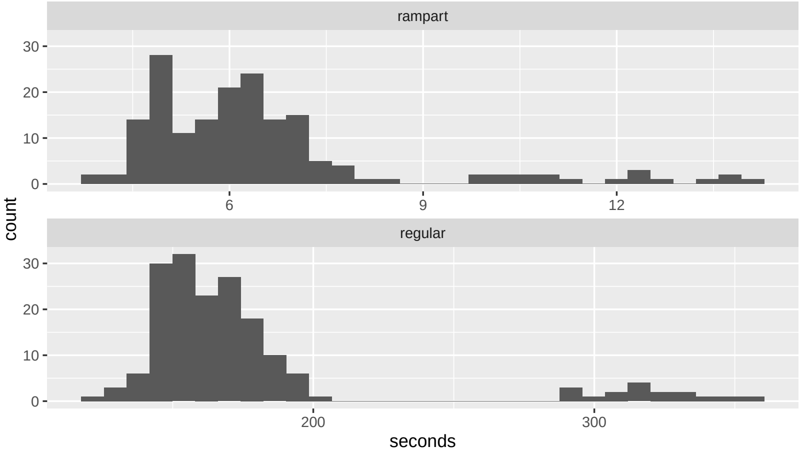
Monte Carlo bias and variance

$\boldsymbol{\theta}$	replications	method	$\|\|$bias$\|\|$	$\|\|\sigma^2\|\|$
1	174	rampart	1.686	0.769
		regular	1.702	0.780
2	171	rampart	2.336	0.557
		regular	2.335	0.560

Scatterplot of deviance at the maximum likelihood



Seconds required per replication



OpenMx is a free and open source extension to the R statistical environment.

Software and support available at
`http://openmx.psyc.virginia.edu/`

Questions?

Appendix

Some extra slides follow

Terminology

Historically, coefficients that help predict all observations are called *fixed effects* whereas the other type of coefficient has been called a *random effect*. These are unfortunate terminology. In the statistical literature, there are at least five definitions of these phrases, all of which differ from each other (Gelman, 2005). Moreover, in computer science, the term *random* is usually associated with draws from a uniform random number generator, not synonymous with *stochastic* that does not suppose a particular distribution. Here we follow Gelman (2005) and use the terms *constant* and *varying*.

Example, lme4

```
lmer(Reaction ~ Days + (Days | Subject), sleepstudy)
```

Example, OpenMx (part 1)

```
bySubj <- mxModel(
    model="bySubj", type="RAM",
    latentVars=c("slope", "intercept"),
    mxData(data.frame(Subject=unique(sleepstudy$Subject)),
           type="raw", primaryKey = "Subject"),
    mxPath(c("intercept", "slope"), arrows=2, values=1),
    mxPath("intercept", "slope", arrows=2,
           values=.25, labels="cov1"))
```

Example, OpenMx (part 2)

```
ss <- mxModel(
    model="sleep", type="RAM", bySubj,
    manifestVars="Reaction", latentVars = "Days",
    mxData(sleepstudy, type="raw", sort=FALSE),
    mxPath("one", "Reaction", arrows=1, free=TRUE),
    mxPath("one", "Days", arrows=1, free=FALSE,
           labels="data.Days"),
    mxPath("Days", "Reaction", arrows=1, free=TRUE),
    mxPath("Reaction", arrows=2, values=1),
    mxPath(paste0('bySubj.', c('intercept','slope')),
           'Reaction', arrows=1, free=FALSE, values=c(1,NA),
           labels=c(NA, "data.Days"), joinKey="Subject"))
```

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Boker, S. M., Brick, T. R., Pritikin, J. N., Wang, Y., von Oertzen, T., Brown, D., . . . Neale, M. C. (2015). Maintained individual data distributed likelihood estimation. *Multivariate Behavioral Research*, *50*(6), 706–720.

Gelman, A. (2005). Analysis of variance–why it is more important than ever. *The Annals of Statistics*, *33*(1), 1–53.

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R., . . . Boker, S. M. (in press). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*. doi: 10.1007/s11336-014-9435-8

Pritikin, J. N. (2016). A computational note on the application of the Supplemented EM algorithm to item response models.

Pritikin, J. N., & Schmidt, K. M. (in press). Model builder for Item Factor Analysis with OpenMx. *R Journal*.

von Oertzen, T., & Hackett, D. C. (submitted). *Pre-processing for efficient maximum likelihood estimation in structural equation models with fixed loadings.* submitted.