Behavior Genetics Models
as Multilevel and then . . .

Michael D. Hunter

School of Psychology
Georgia Institute of Technology

International Workshop on Statistical Genetics and
Methodology of Twin and Family Studies
Boulder, CO; March 3, 2020

Georgia
Tech

Outline

- ▶ Code warm-up
- ▶ ACE gets a multilevel makeover
- ▶ Relations among
    - ▶ Multilevel SEM
    - ▶ Polyphenotype (i.e. Multitrait) ACE
    - ▶ GCTA-GREML
- ▶ Small Simulation
- ▶ Conclusions
- ▶ Code Play

Georgia
Tech

Code warm-up

- Copy over to your space: /home/mikeh/...
- rSemMono.R

Georgia
Tech

Previous Work

Monophenotype ACE models have been shown to be estimable as multilevel models (van den Oord, 2001; Guo & Wang, 2002; McArdle & Prescott, 2005; Rabe-Hesketh, Skrondal, & Gjessing, 2007)

- ▶ Difficult to specify
- ▶ Parameters must be transformed for interpretation
- ▶ Do not scale well with increasing relatedness groups
- ▶ Monophenotype only

Anecdotally, multilevel models in M*plus* have been used for single trait analyses.

Georgia Tech

Conceptual Comparison

Multilevel Models

- ▶ Accounts for within-cluster covariance
- ▶ Students nested within classrooms
- ▶ Within-class and between-class variability
- ▶ Covariance across rows
- ▶ (Often) single outcome variable spread over multiple related units

Behavior Genetics Models

- ▶ Accounts for within-cluster covariance
- ▶ Siblings nested with families
- ▶ Within-family and between-family variability
- ▶ Covariance across cols
- ▶ (Often) single outcome variable spread over multiple related units

Georgia Tech

Standard ACE
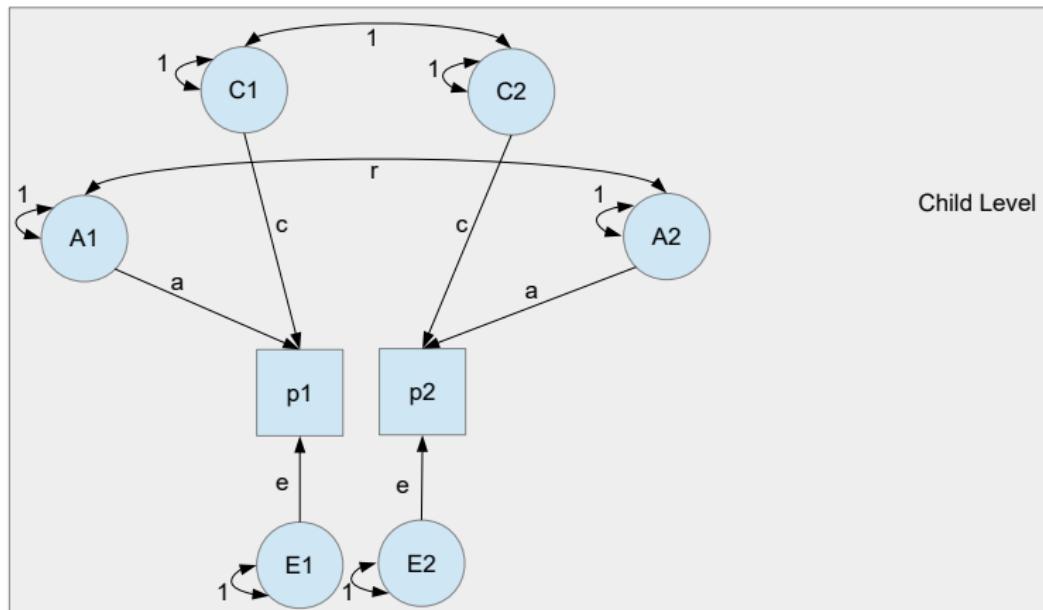
Independent Components ACE
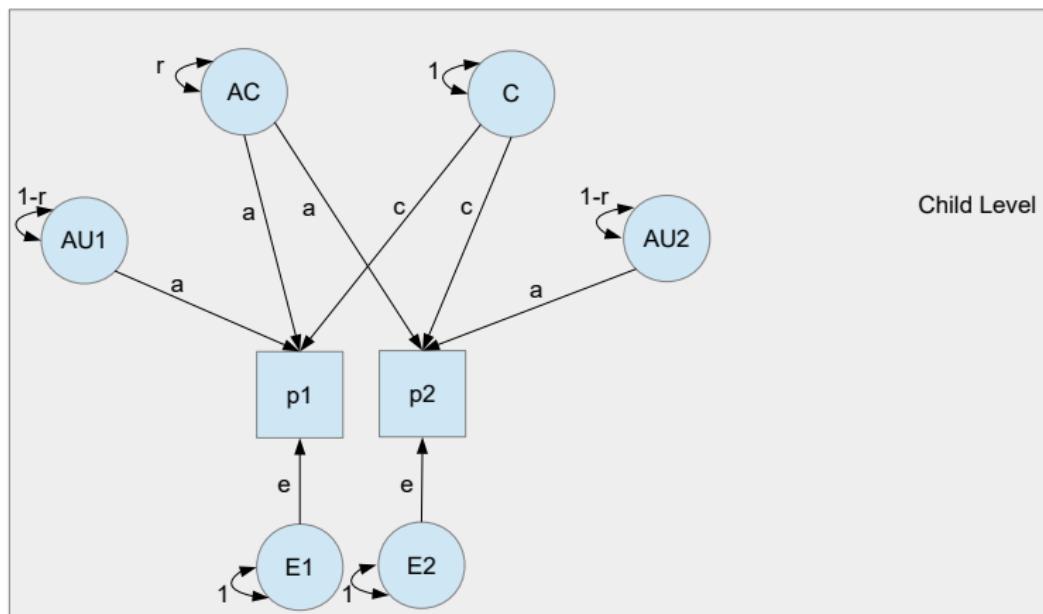
Little Ditty by Jack and Carol
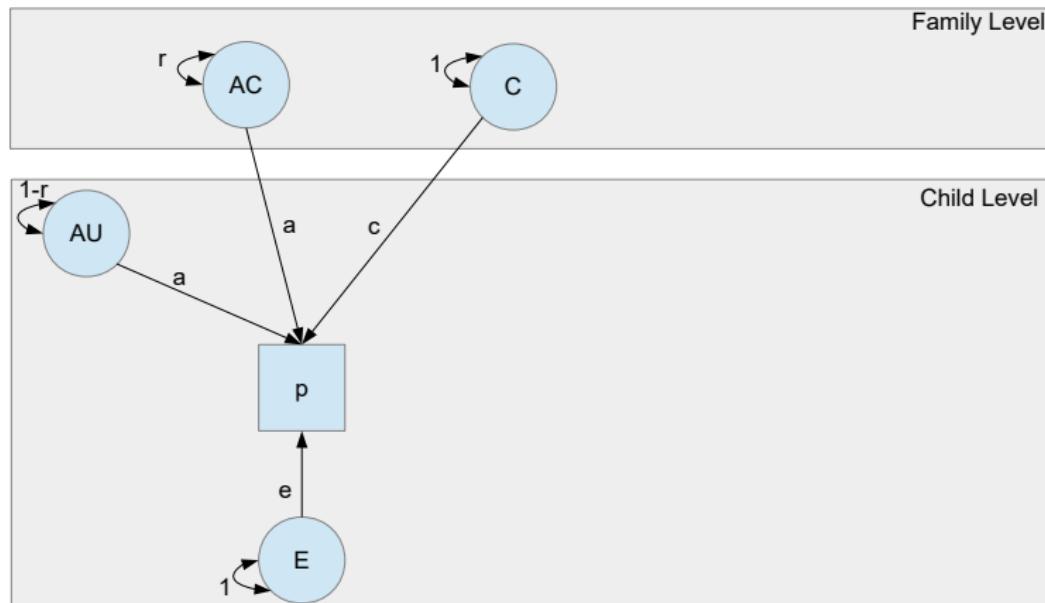
What does the multilevel form of an ACE model get you?

Georgia
Tech

What does the multilevel form of an ACE model get you?

▶ Sensible terminology of multilevel models

▶ Easy extension to multiple phenotypes

▶ No more bivariate or diphenotype confusion

Georgia
Tech

Polyphenotype vs Multivariate

Dr. Kirkpatrick, are you reading this?

Univariate or Bivariate? p-variate or $2p$-variate?
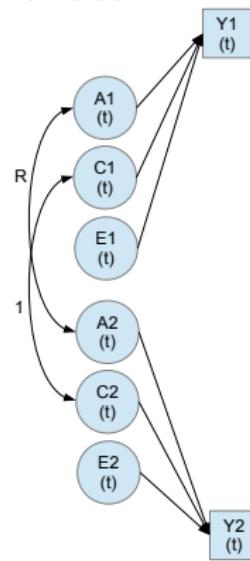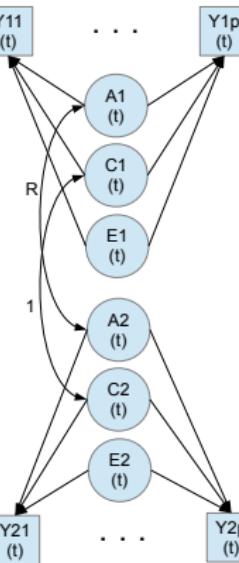
What does the multilevel form of an ACE model get you?

▶ Sensible terminology of multilevel models

▶ Easy extension to multiple phenotypes

▶ No more bivariate or diphenotype confusion

What does the multilevel form of an ACE model get you?

▶ Extension to higher levels of nesting

Georgia
Tech

Multilevel ACE (McArdle & Prescott, 2005)
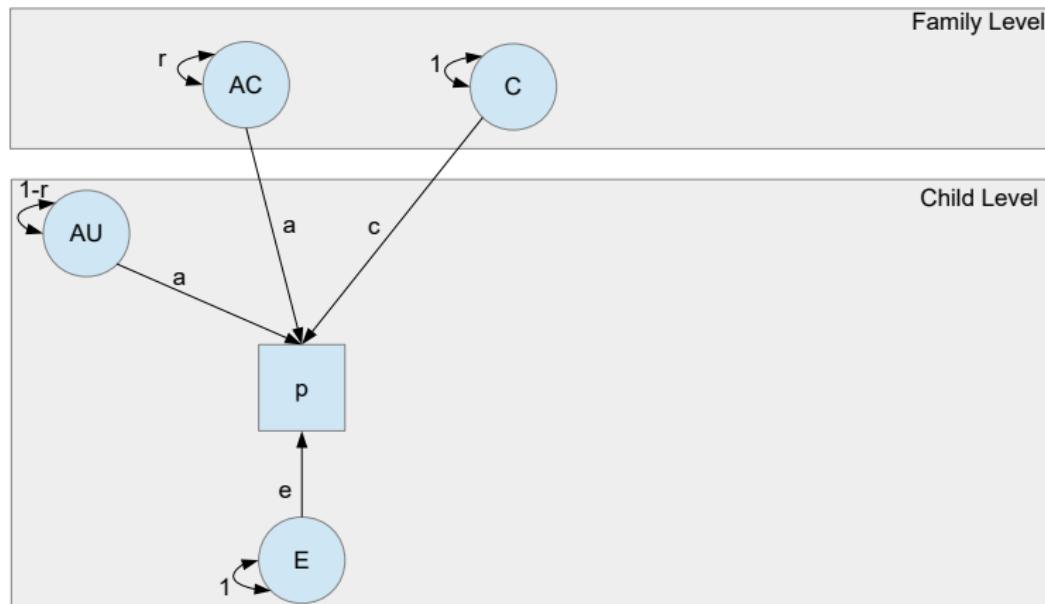
Multilevel ACE (McArdle & Prescott, 2005)
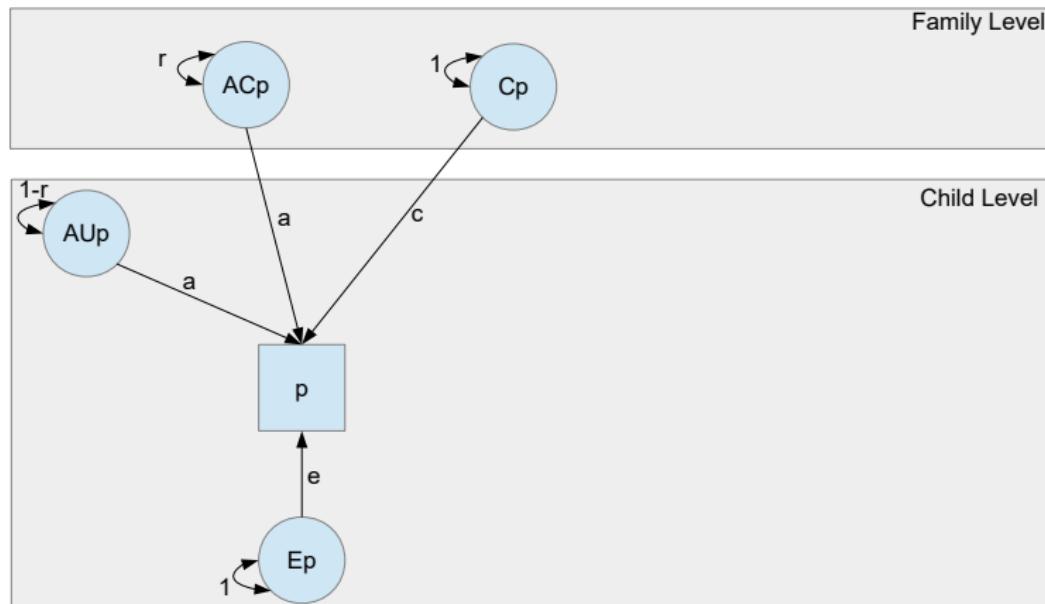
3-Level Multilevel ACE
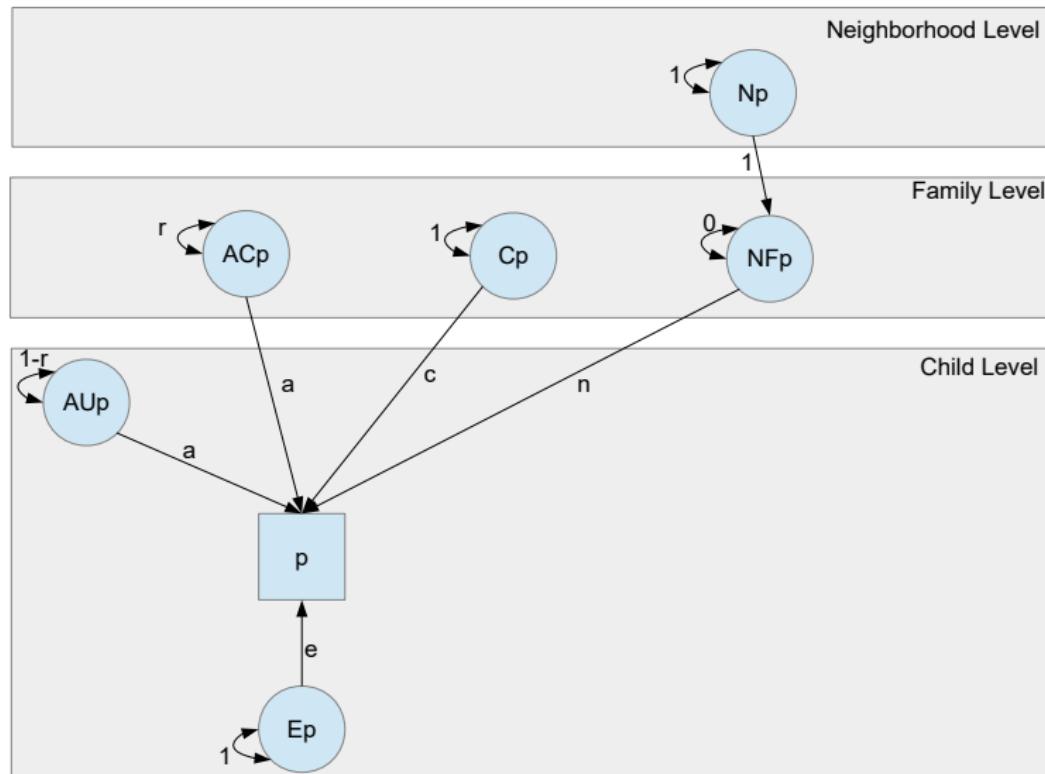
What if you don't have neighborhoods?

▶ Other geo-spatial arrangements (Counties, States, Regions, Nations)

▶ Multiple sites or consortia

▶ Extended families or pedigrees

▶ Drug treatment facilities

▶ Brain scanners within a hospitals

▶ Data collectors that see fixed sets of families

Questions so far?

Blockwise Covariance

$$
\begin{array}{c}
\quad\quad y_{11} \quad\quad\quad\quad y_{12} \quad\quad\quad\quad y_{21} \quad\quad\quad\quad y_{22} \\
\begin{array}{c} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{array}
\left(
\begin{array}{cccc}
A + C + E & & & \\
.5A + C & A + C + E & & \\
\mathbf{0} & \mathbf{0} & A + C + E & \\
\mathbf{0} & \mathbf{0} & A + C & A + C + E
\end{array}
\right)
\end{array}
$$

NLSY79 Gen 1 Height

National Longitudinal Survey of Youth 1979 Height among DZ and MZ Twins
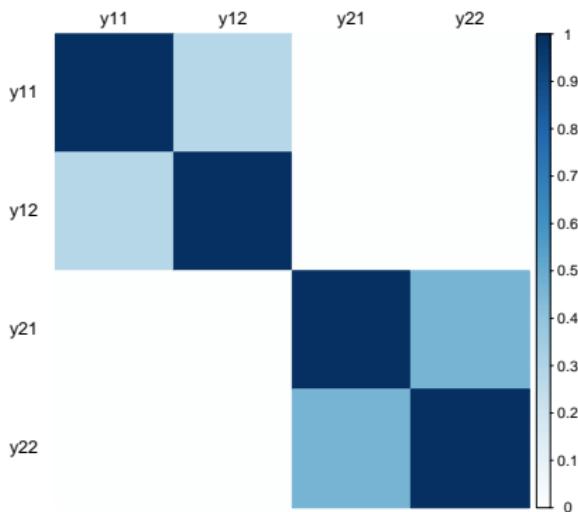
NLSY79 Gen 1 Height

National Longitudinal Survey of Youth 1979 Height among Cousins, Siblings, and MZ Twins
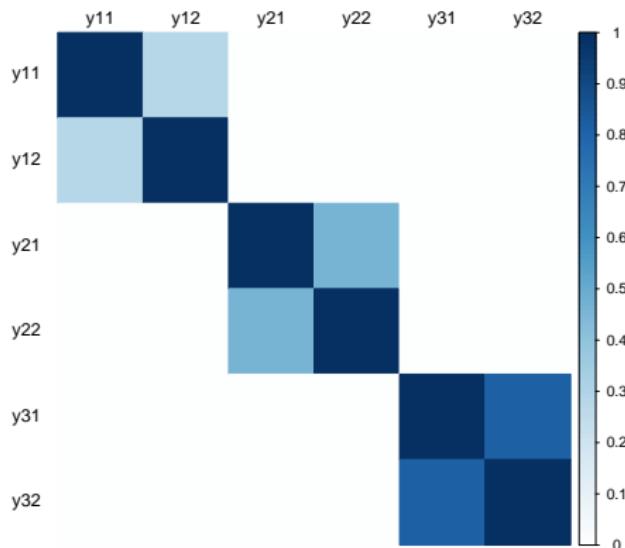
NLSY79 Gen 1 Height

National Longitudinal Survey of Youth 1979 Height 20 Covariance Blocks

Georgia Tech
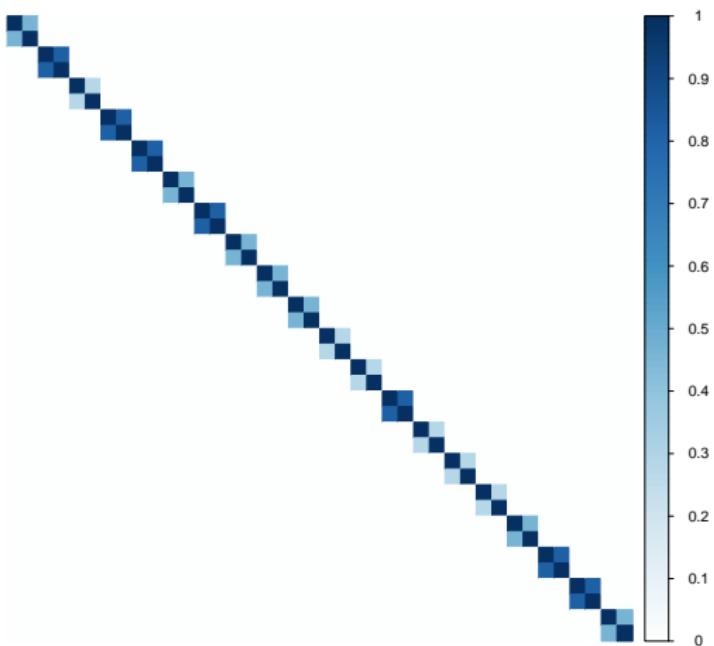
As a Bigger Multilevel SEM

Pedigrees

- ▶ Tall Data

- ▶ One Child per Row

- ▶ Families of Varying Sizes per Cluster

- ▶ Large Family Designs

- ▶ Many Diverse Cluster Sizes and Structures
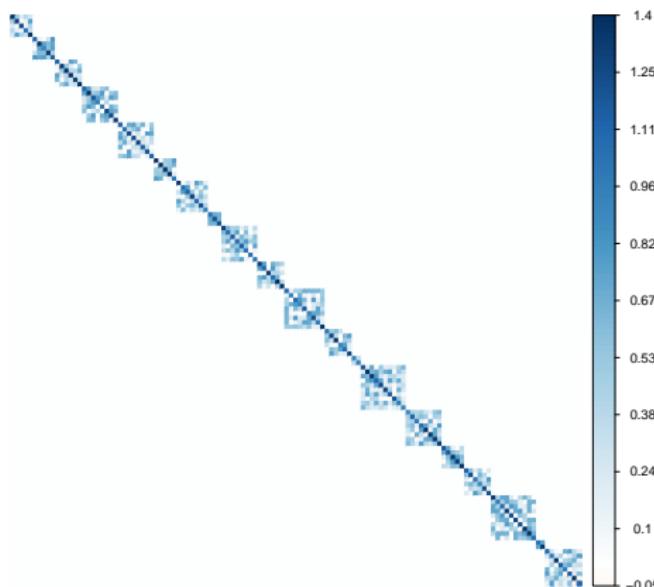
Multilevel BG

As a Very Big Multilevel Regression Model
GCTA

▶ Tall Data

▶ One person per Row

▶ One Family of All
   Related Persons
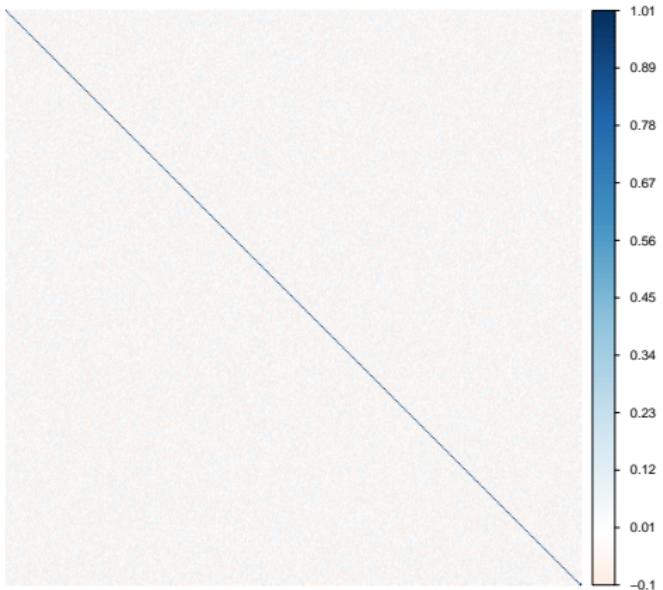
Another way to show ACE models are multilevel.
(For people who prefer equations)

Georgia
Tech

Multitrait Multilevel Models

$$\Sigma = \begin{matrix} & y_{11} & y_{12} & y_{13} & y_{21} & y_{22} \\ \begin{matrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \end{matrix} & \begin{pmatrix} \Sigma_B + \Sigma_W & & & & \\ \Sigma_B & \Sigma_B + \Sigma_W & & & \\ \Sigma_B & \Sigma_B & \Sigma_B + \Sigma_W & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \Sigma_B + \Sigma_W & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \Sigma_B & \Sigma_B + \Sigma_W \end{pmatrix} \end{matrix}$$

Muthén (1991, 1994)

Multitrait ACE Models

MZ Twins

$$\Sigma = \begin{array}{c} \\ y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{array} \begin{pmatrix} \overset{y_{11}}{A+C+E} & \overset{y_{12}}{} & \overset{y_{21}}{} & \overset{y_{22}}{} \\ A+C & A+C+E & & \\ \mathbf{0} & \mathbf{0} & A+C+E & \\ \mathbf{0} & \mathbf{0} & A+C & A+C+E \end{pmatrix}$$

$\Sigma_B = A + C$

$\Sigma_W = E$

Multitrait ACE Models
DZ Twins

$$\Sigma = \begin{array}{c} \\ y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{array} \begin{pmatrix} \overset{y_{11}}{A + C + E} & \overset{y_{12}}{} & \overset{y_{21}}{} & \overset{y_{22}}{} \\ 0.5\ A + C & A + C + E & & \\ \mathbf{0} & \mathbf{0} & A + C + E & \\ \mathbf{0} & \mathbf{0} & 0.5\ A + C & A + C + E \end{pmatrix}$$

$\Sigma_B = 0.5\ A + C$
$\Sigma_W = 0.5\ A + E$

Multitrait ACE Models

$$
\Sigma = \begin{array}{c} \\ y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{array}
\begin{array}{cccc}
y_{11} & y_{12} & y_{21} & y_{22} \\
\left(\begin{array}{cccc}
A + C + E & & & \\
h\ A + C & A + C + E & & \\
\mathbf{0} & \mathbf{0} & A + C + E & \\
\mathbf{0} & \mathbf{0} & h\ A + C & A + C + E
\end{array} \right)
\end{array}
$$

Polyphenotype ACE Models

It's multilevel!

$$\Sigma = \begin{matrix} & \begin{matrix} y_{11} & \quad\quad y_{12} & \quad\quad\quad y_{21} & \quad\quad\quad\quad y_{22} \end{matrix} \\ \begin{matrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{matrix} & \left(\begin{matrix} A+C+E & & & \\ h\,A+C & A+C+E & & \\ \mathbf{0} & \mathbf{0} & A+C+E & \\ \mathbf{0} & \mathbf{0} & h\,A+C & A+C+E \end{matrix} \right) \end{matrix}$$

$\Sigma_B = h\,A + C$ and $\Sigma_W = (1-h)\,A + E$

$\Sigma_B + \Sigma_W = h\,A + C + (1-h)\,A + E =$

$h\,A + (1-h)\,A + C + E = A + C + E$

Polyphenotype ACE Models
Uniform Pedigree

$$\Sigma = \begin{array}{c} \\ y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{22} \end{array} \begin{array}{c} y_{11} \\ \left(\begin{array}{cccc} A+C+E & & & \\ h\ A+C & A+C+E & & \\ h\ A+C & h\ A+C & A+C+E & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & A+C+E \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & h\ A+C \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & h\ A+C \end{array} \right. \end{array} \begin{array}{cc} y_{12} & y_{13} \\ & \\ & \\ & \\ & \\ A & \\ \end{array}$$

$\Sigma_B = h\ A + C$ and $\Sigma_W = (1 - h)\ A + E$

Kronecker Product

Replicator, Activate!

$$
\begin{pmatrix}
h_{11} & h_{12} & h_{13} \\
h_{21} & h_{22} & h_{23} \\
h_{31} & h_{32} & h_{33}
\end{pmatrix} \otimes A =
\begin{pmatrix}
h_{11}A & h_{12}A & h_{13}A \\
h_{21}A & h_{22}A & h_{23}A \\
h_{31}A & h_{32}A & h_{33}A
\end{pmatrix}
\quad (1)
$$

Georgia
Tech

Polyphenotype ACE Models

General Pedigree

$$
\Sigma = \begin{array}{c} \\ y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{22} \end{array}
\begin{array}{cccc}
y_{11} & y_{12} & y_{13} & y_{21} \\
\left(A + C + E \right. & & & \\
h_{12}\, A + C & A + C + E & & \\
h_{13}\, A + C & h_{23}\, A + C & A + C + E & \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & A + C + E \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & h_{12}\, A + C \quad A \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & h_{13}\, A + C \quad h
\end{array}
$$

$$\boldsymbol{\Sigma}_B = \boldsymbol{H} \otimes \boldsymbol{A} + \boldsymbol{U} \otimes \boldsymbol{C} \text{ and } \boldsymbol{\Sigma}_W = \boldsymbol{E}$$

GCTA and GREML

$$\boldsymbol{\Sigma}_B = \boldsymbol{H} \otimes \boldsymbol{A} + \boldsymbol{U} \otimes \boldsymbol{C} \text{ and } \boldsymbol{\Sigma}_W = \boldsymbol{E}$$

$$\text{ACE Model: } \boldsymbol{\Sigma}_{block} = \boldsymbol{H} \otimes \boldsymbol{A} + \boldsymbol{U} \otimes \boldsymbol{C} + \boldsymbol{I} \otimes \boldsymbol{E}$$

$$\text{AE Model: } \boldsymbol{\Sigma}_{block} = \boldsymbol{H} \otimes \boldsymbol{A} + \boldsymbol{I} \otimes \boldsymbol{E}$$

$$\text{GREML (AE Notation): } \Sigma = H\sigma_A^2 + I\sigma_E^2$$

$$\text{GREML (Standard Notation): } \Sigma = \underbrace{\sum_{i=1}^{r} \sigma_i^2 A_i}_{\Sigma_B} + \underbrace{\sigma_\varepsilon^2 I}_{\Sigma_W}$$

Simulation Study

Data Generation Parameters

- ▶ Number of Phenotypes: 5
- ▶ Structure of Variance Components: Cholesky
- ▶ Number of Families: 150, 600, 1200
- ▶ Number of Kids in Each Family: 2, 10, 50, 100
- ▶ Relatedness Patterns: Twins, Siblings, Cousins

Simulation Study

Key Questions

▶ What is the relationship between the multilevel and wide SEM parameters?

▶ What is the relationship between the multilevel and wide SEM standard errors?

▶ What factors influence model estimation time?

Georgia
Tech

Relationship between parameters

numKids=2, numFamilies=150, numPhenotypes=5
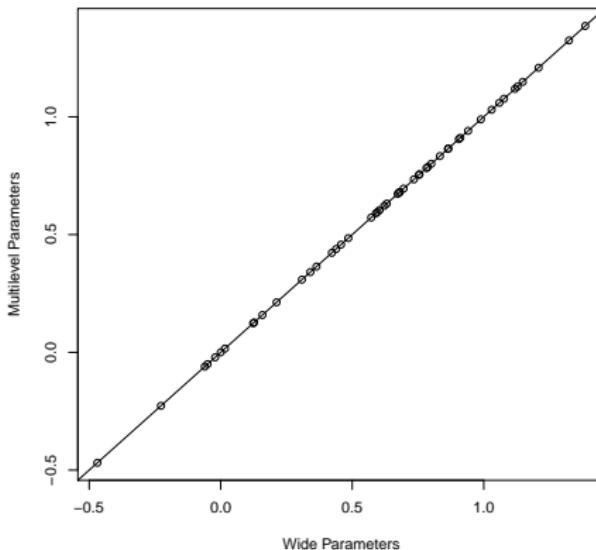
r=0.999999999; rmse $\approx 1 \times 10^{-6}$

Relationship between parameters
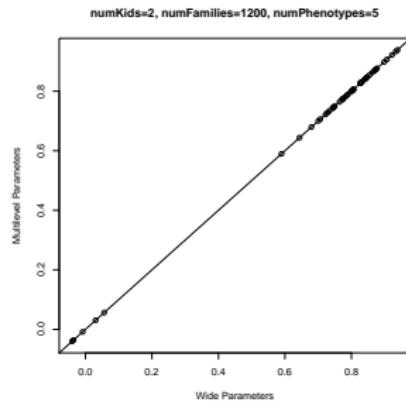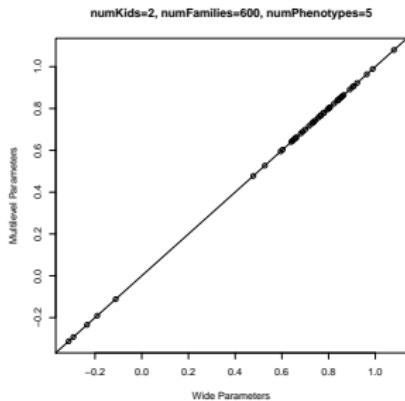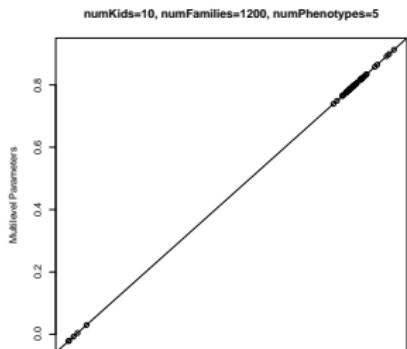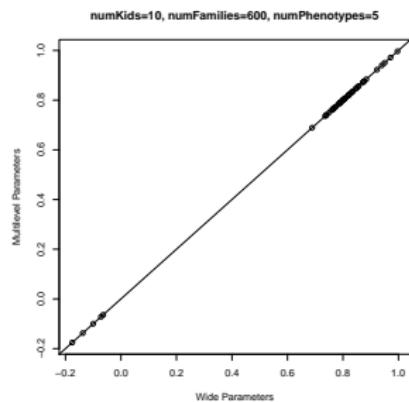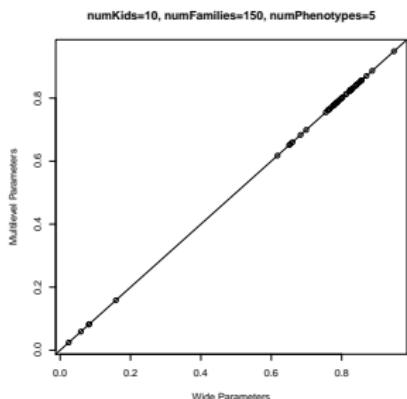
Relationship between parameters

Relationship between parameters
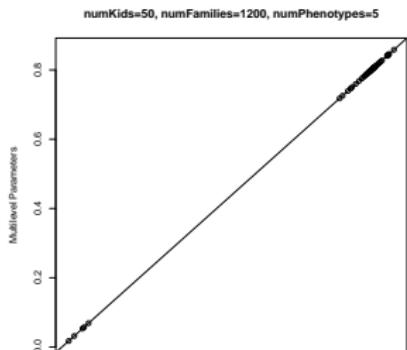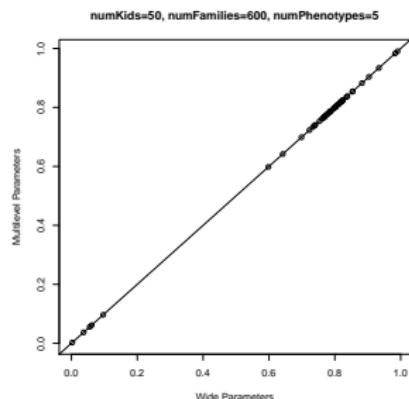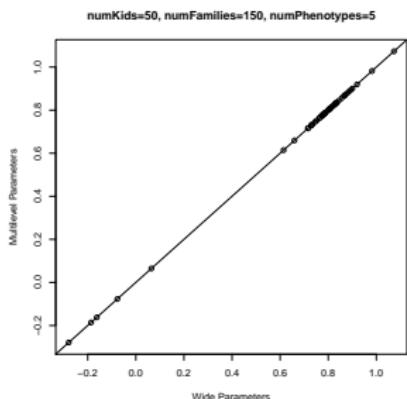
Relationship between parameters
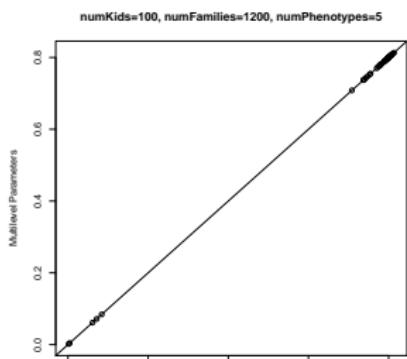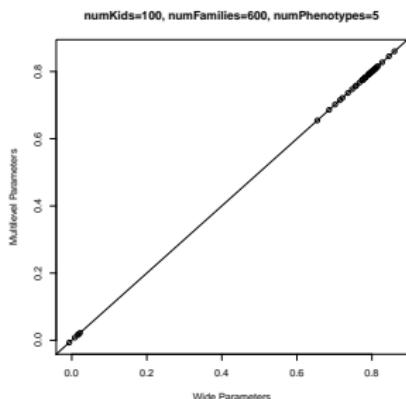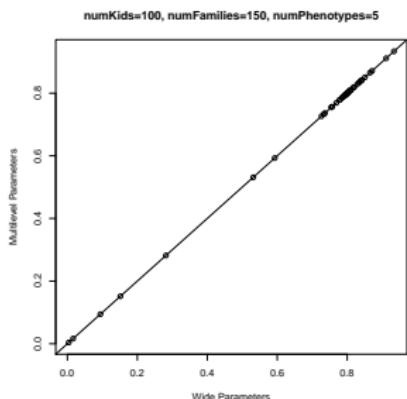
Multilevel BG

Relationship between standard errors

- ▶ They are the same!
- ▶ rmse $\approx 1 \times 10^{-4}$
- ▶ They are not a similar/transformed/approximate models.
- ▶ The multilevel ACE model is identical to the wide SEM ACE model.

Timing
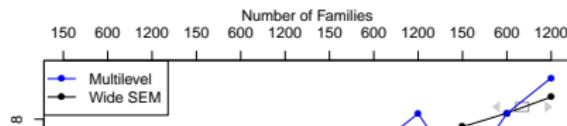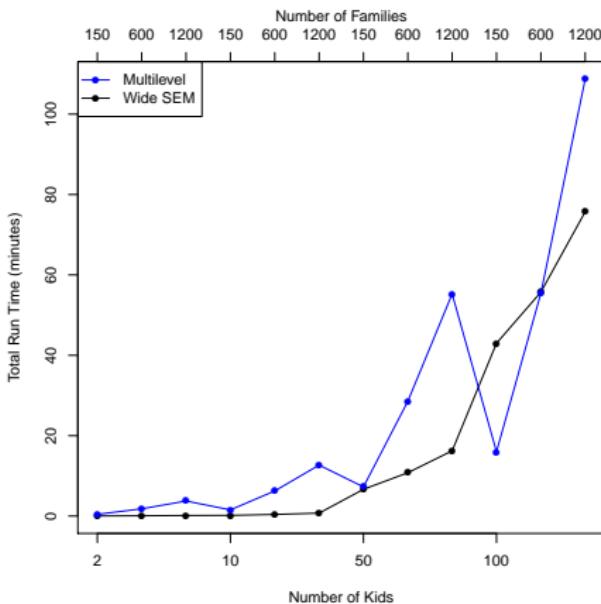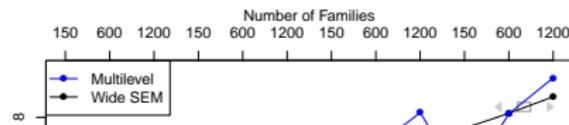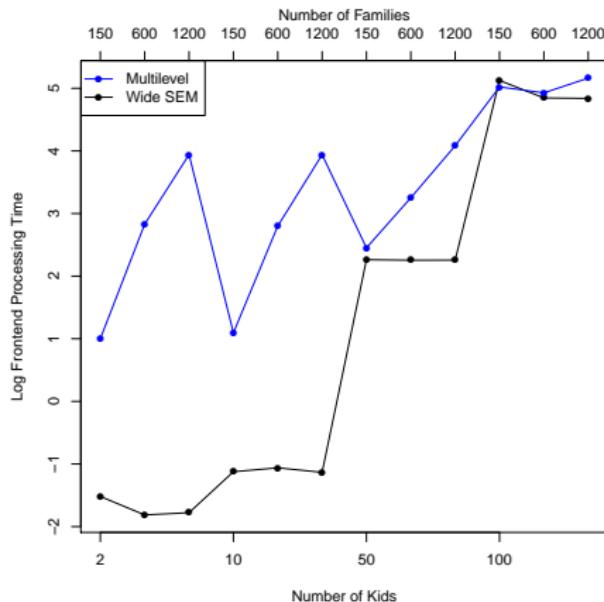
Multilevel BG

Timing

Many Level SEM

Routledge
Taylor & Francis Group

Many-Level Multilevel Structural Equation Modeling:
An Efficient Evaluation Strategy

Joshua N. Pritikin,[1] Michael D. Hunter,[2] Timo von Oertzen,[3] Timothy R. Brick,[4] and Steven M. Boker[5]

[1]*Virginia Commonwealth University*
[2]*University of Oklahoma Health Sciences Center*
[3]*Universität der Bundeswehr, München*
[4]*Pennsylvania State University*
[5]*University of Virginia*

Georgia Tech

Linking Tables

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	Finance	Harriet
Harriet	Sales	George

Georgia
Tech

Example in Child Behavioral Health

(Pritikin, Hunter, von Oertzen, Brick, & Boker, 2017)

OUHSC & OKDHS Partnership
Repeatedly administer the pediatric symptom checklist
(Jellinek et al., 1988)

- ▶ May 2015 to December 2016
  - ▶ 14,436 screeners
  - ▶ 6,076 children aged 4 to 17
  - ▶ 1,280 case workers
  - ▶ 83 county offices
  - ▶ 34 district offices
- ▶ Now
  - ▶ ≈ 35,000 screeners
  - ▶ ≈ 10,000 children

Georgia
Tech

Performance

Intercept only, variance decomposition

TABLE 1

Variance Decomposition of the Pediatric Symptom Checklist in Common Mixed Effects Programs

	3-Level				5-Level	
	lme4	*nlme*	*Mplus*	*OpenMx*	*nlme*	*OpenMx*
District variance	—	—	—	—	0.913	0.874
County variance	—	—	—	—	0.096	0.122
Worker variance	9.235	9.235	1.063	9.231	8.247	8.202
Child variance	32.255	32.255	39.914	32.234	32.257	32.239
Residual variance	17.946	17.946	17.922	17.960	17.939	17.955
Intercept	10.103	10.103	10.003	10.116	10.115	10.125
Estimation time (sec)	11.14	2.21	53.40	4.66	15.68	12.28

Note. Code to take advantage of shared memory parallel processing was disabled, reduce the variance in estimation time.

Georgia Tech

Summary

- ▶ ACE models are multilevel
- ▶ GCTA-GREML models are multilevel
- ▶ The AE model and a GCTA-GREML model differ only in how genetic relatedness is measured
- ▶ Either the multilevel or the wide SEM form of these models may be preferred in different situations
- ▶ The multilevel form of ACE models extends easily to multiple phenotypes and more levels
- ▶ Let's see some code!

Georgia
Tech

OpenMx $\geq$ 2.0

Recent Features

- objective function = $\underbrace{\text{expectation function}}_{\text{data generation}} + \underbrace{\text{fit function}}_{\text{data comparison}}$

- Multiple optimizers: SLSQP, NPSOL, CSOLNP

- New Models: Item factor analysis (Pritikin, Hunter, & Boker, 2015), GREML, state space, LISREL

- New Helpers: `mxFactorScores`, `mxGenerateData`, `mxMI`, `mxRefModels`, `mxTryHard`

OpenMx $\geq$ 2.7
More Recent Features

▶ Handy S3 methods: `logLik`, `confint`, `anova`, `coef`, `simulate`

▶ New Fit Functions for multigroup, mixtures, hidden Markov models, weighted least squares

▶ Utilities: `mxSE`, `mxAutoStart`, nonparametric and parametric `mxBootstrap`, `mxCheckIdentification`, `mxGetExpected`

Code Demos

- ▶ `rSemMono.R` replicates a 2-trait ACE model as a multilevel model
- ▶ `threeLevelCholesky.R` has children nested in families nested in neighborhoods

Thank You

michael-hunter@gatech.edu

Guo, G., & Wang, J. (2002). The mixed or multilevel model for behavior genetic analysis. *Behavior Genetics*, *32*(1), 37-49. doi: 10.1023/a:1014455812027

Jellinek, M. S., Murphy, J. M., Robinson, J., Feins, A., Lamb, S., & Fenton, T. (1988, feb). Pediatric symptom checklist: Screening school-age children for psychosocial dysfunction. *The Journal of Pediatrics*, *112*(2), 201–209. doi: 10.1016/s0022-3476(88)80056-8

McArdle, J. J., & Prescott, C. A. (2005, Sep). Mixed-effects variance components models for biometric family analyses. *Behavior Genetics*, *35*(5), 631-652. doi: 10.1007/s10519-005-2868-1

Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, *28*(4), 338-354. doi: 10.1111/j.1745-3984.1991.tb00363.x

Georgia
Tech

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research*, *22*(3), 376-398. doi: 10.1177/0049124194022003006

Pritikin, J. N., Hunter, M. D., & Boker, S. M. (2015). Modular open-source software for item factor analysis. *Educational and Psychological Measurement*, *75*(3), 458-474. doi: 10.1177/0013164414554615

Pritikin, J. N., Hunter, M. D., von Oertzen, T., Brick, T. R., & Boker, S. M. (2017). Many-level multilevel structural equation modeling: An efficient evaluation strategy. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(5), 684–698. doi: 10.1080/10705511.2017.1293542

Rabe-Hesketh, S., Skrondal, A., & Gjessing, H. K. (2007, May). Biometrical modeling of twin and family data using standard mixed model software. *Biometrics*, *64*(1),

280-288. doi: 10.1111/j.1541-0420.2007.00803.x

van den Oord, E. J. C. G. (2001). Estimating effects of latent and measured genotypes in multilevel models. *Statistical Methods in Medical Research*, *10*, 393-407. doi: 10.1177/096228020101000603

Georgia
Tech