Developmental Models/ Longitudinal Data Analysis

Danielle Dick & Nathan Gillespie

Boulder, March 2006

Washington
WASHINGTON UNIVERSITY IN ST.LOUIS
School of Medicine

Why conduct longitudinal analyses?

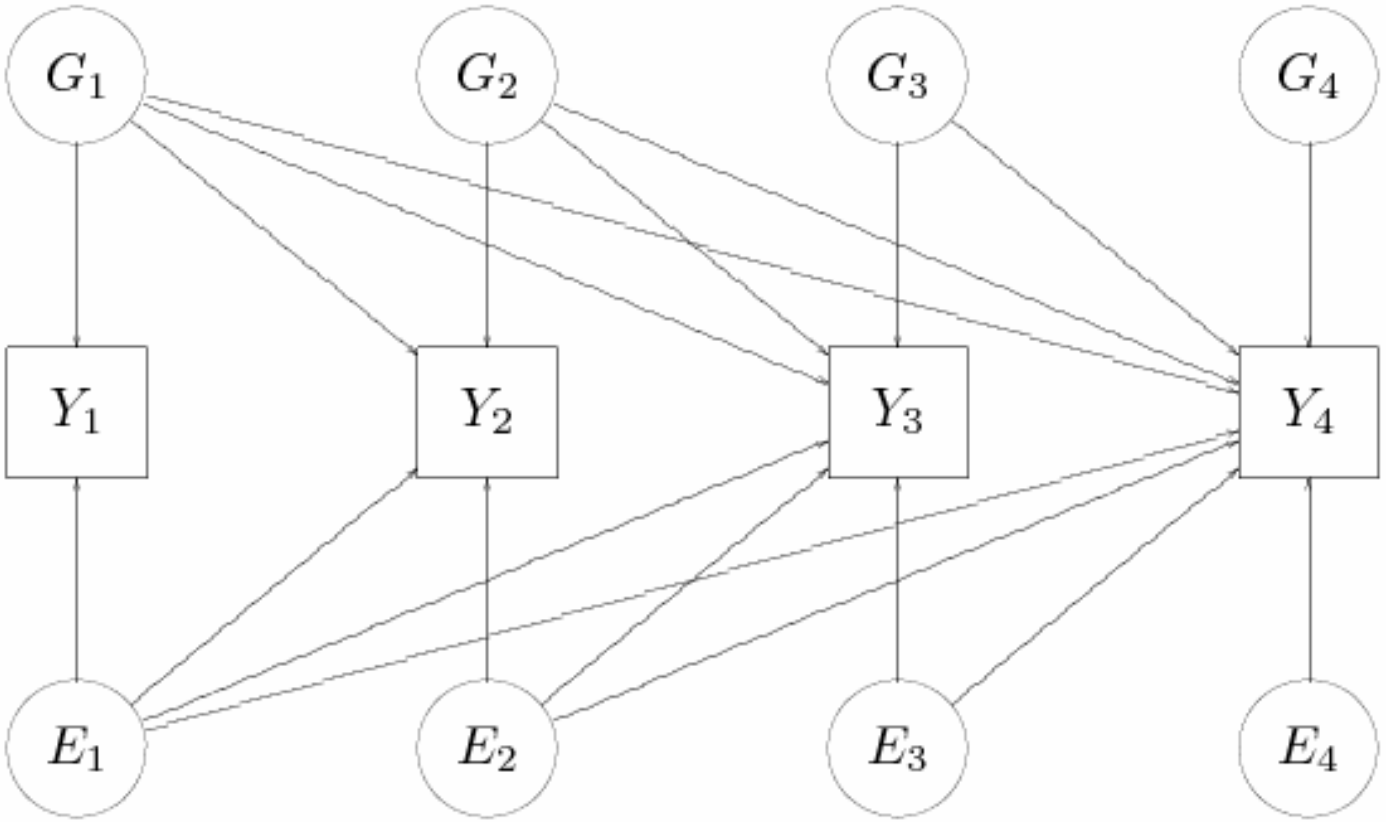
- Can improve power by using multiple observations from the same individual
- Can examine time-dependent genetic and environmental effects
  - Changing magnitude of genetic/environmental influence across development
  - Same versus different genes across development

Methods for Longitudinal Analysis

- Cholesky Models
- Simplex Models
- Growth Curve Models

Cholesky Model

A longitudinal study with 4 waves (single individual):



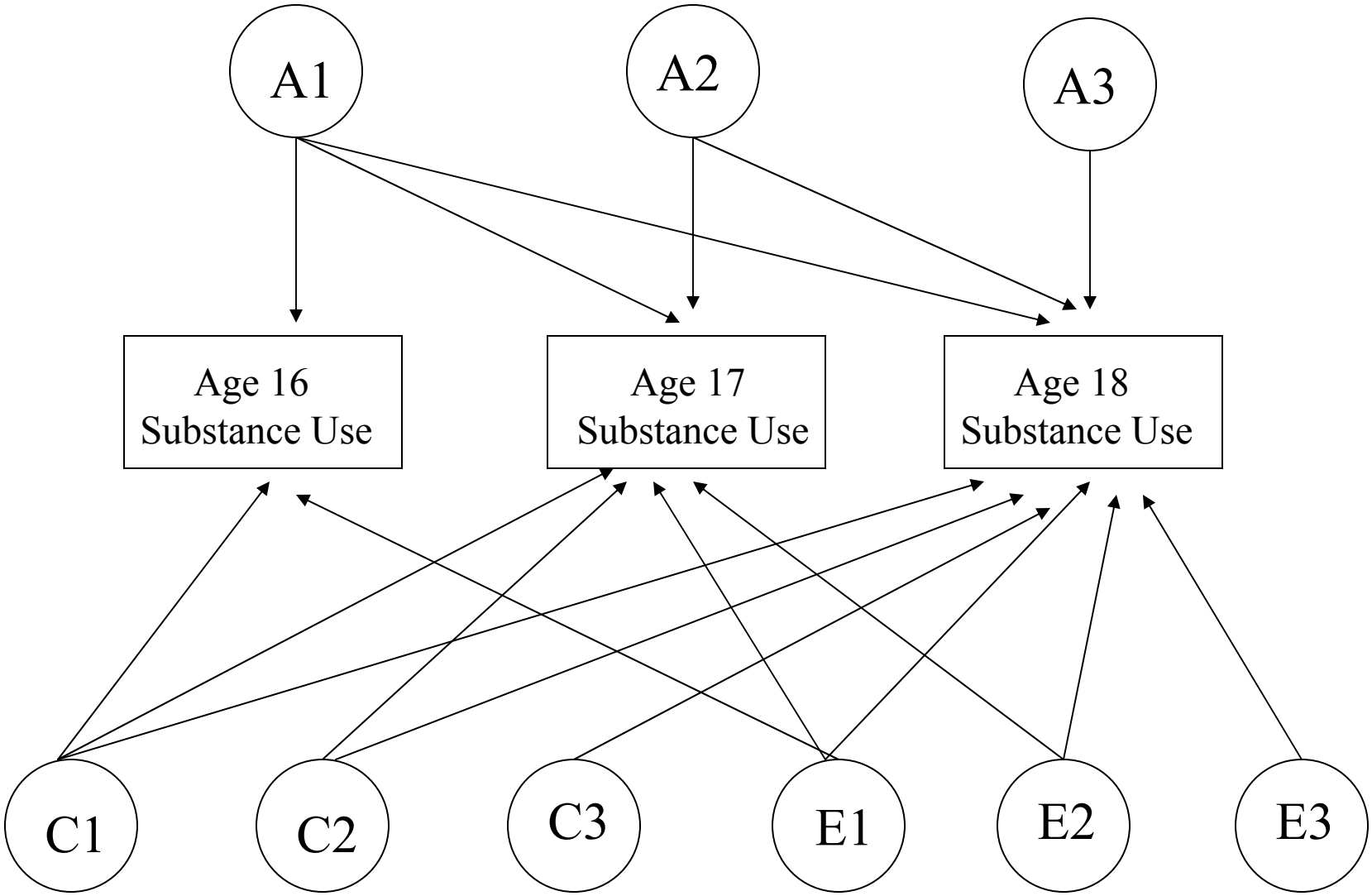
Cholesky Model

- Advantages:
  - Logical: organized such that all factors are constrained to impact later, but not earlier timepoints
  - Requires few assumptions, can predict any pattern of change
- Disadvantages:
  - Not falsifiable
  - Does not make predictions about what will happen in the future (as yet unmeasured timepoints)
  - Only feasible for limited number of measurements

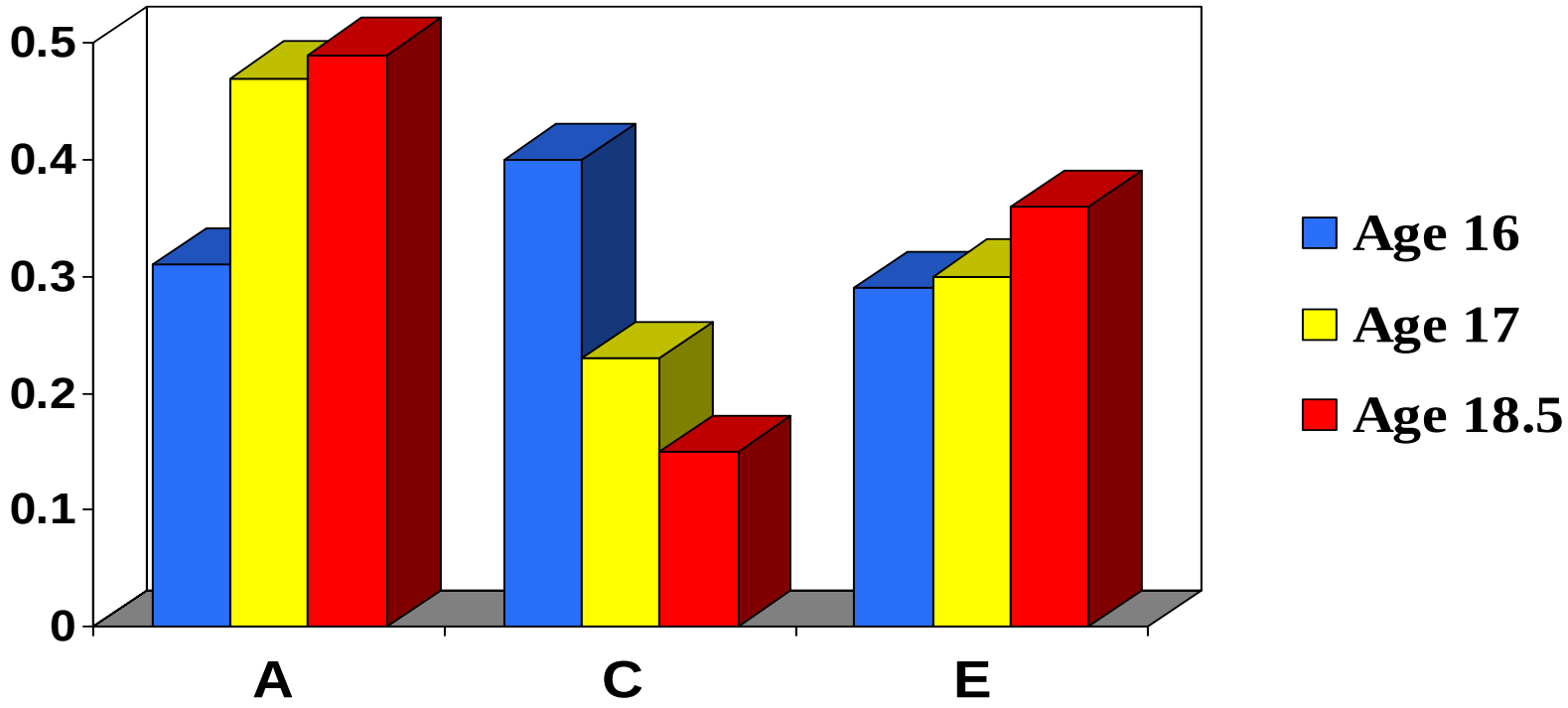
Cholesky Model

- Questions you can address:
  - Magnitude of genetic/environmental influence at each time
  - Extent to which genetic/environmental influences overlap across time

Trivariate Cholesky Decomposition

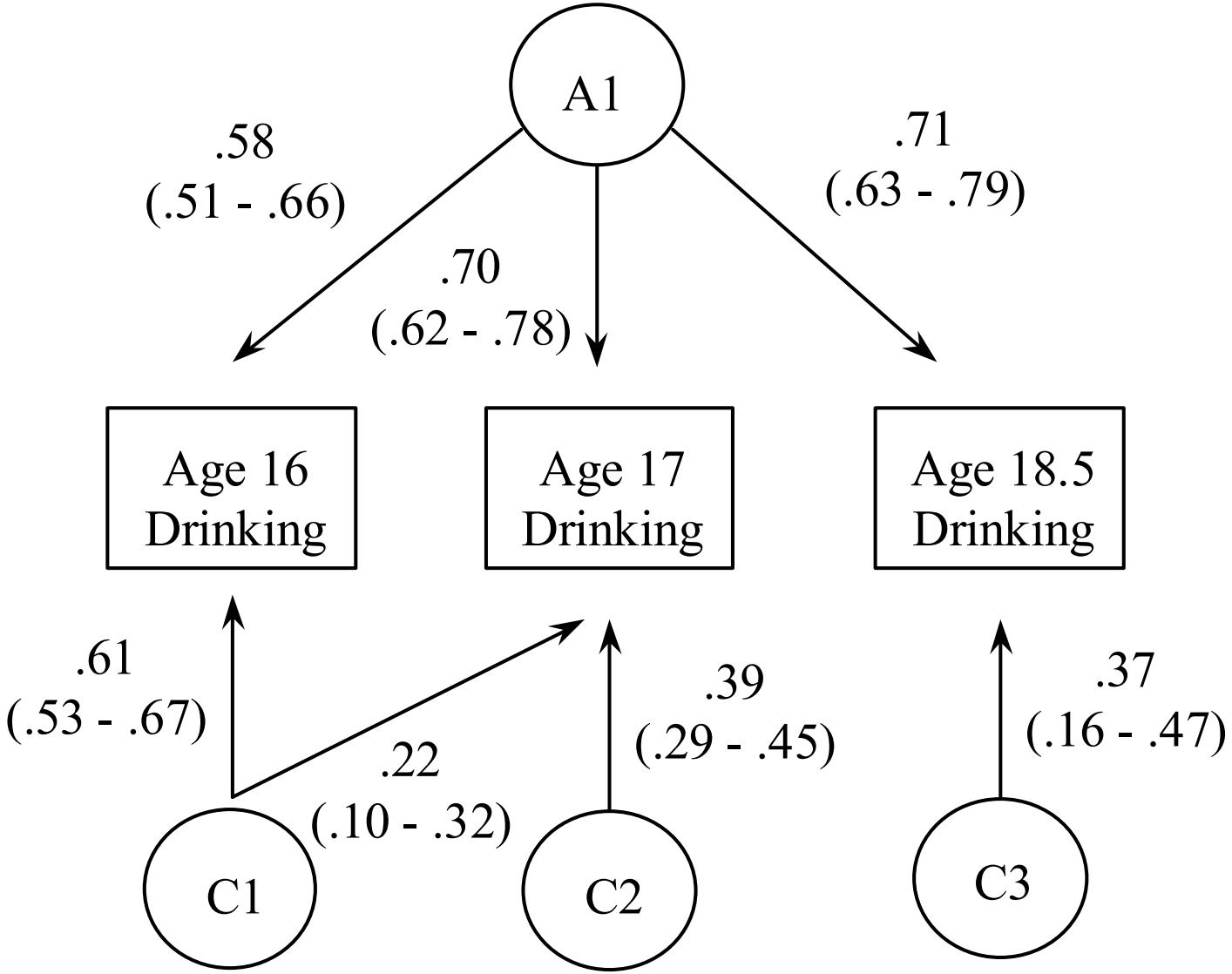


Partitioning of Variance from Longitudinal Drinking Data



Rose, Dick 2001, ACER

Best-fitting model for drinking



Cholesky Model

- Questions you can address:
  - Magnitude of genetic/environmental influence at each time
  - Extent to which genetic/environmental influences overlap across time
  - Other standard multivariate extensions (e.g., multiple group models)

Methods for Longitudinal Analysis

- Cholesky Models
- Simplex Models
- Growth Curve Models

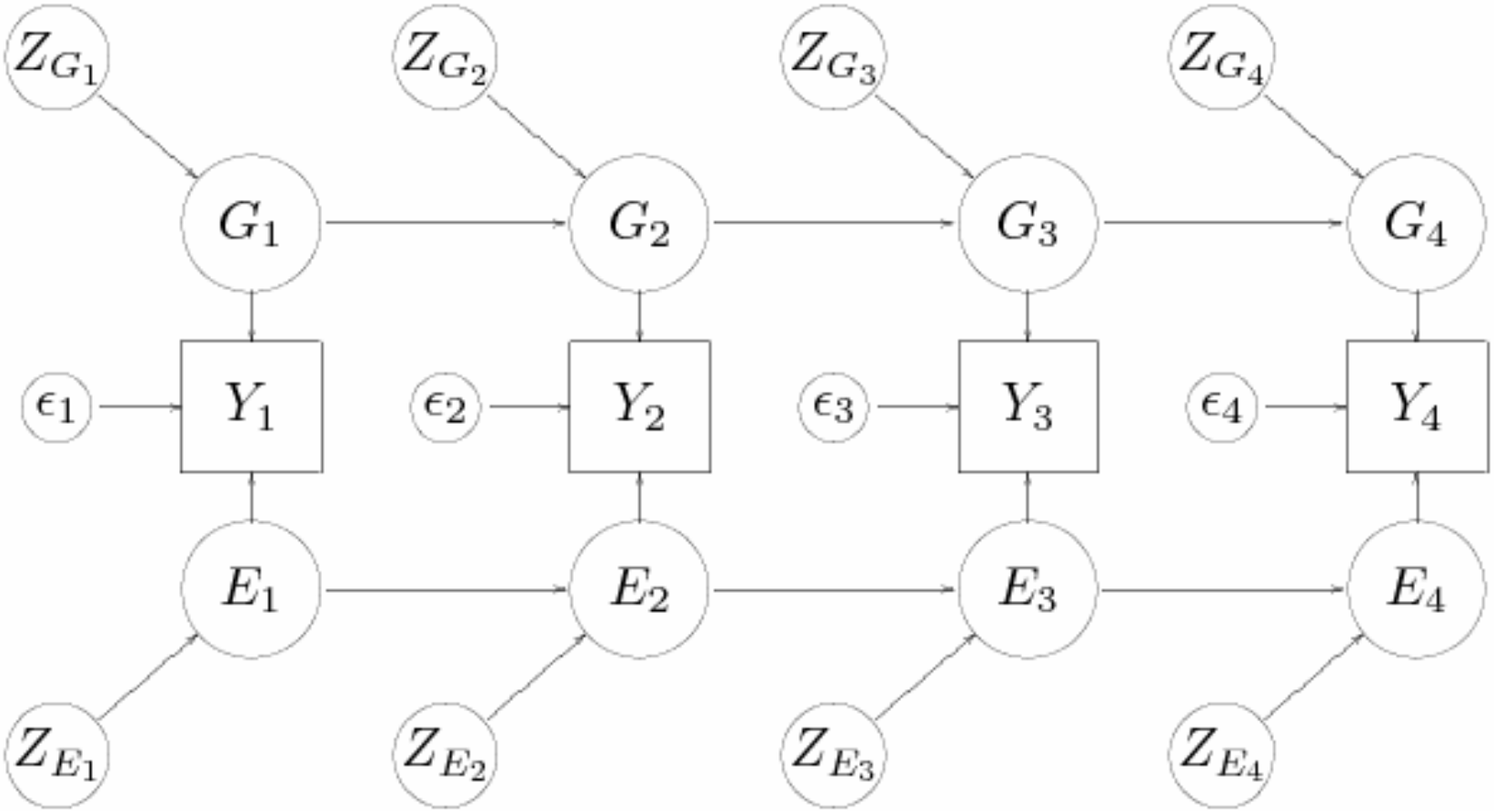
What are simplex/autoregressive models?

- Models with a specific structure of association whereby correlations are largest between adjoining measures, and drop off systematically as distance between variables increases

- This structure is frequently observed with longitudinal data:
  - Classic example: weight measurements
    - 2 year intervals from age 5 to 15
    - you would expect that weight at age 5 would be most highly correlated with weight at 7, somewhat less correlated with weight at 9, less correlated with weight at 11, etc.

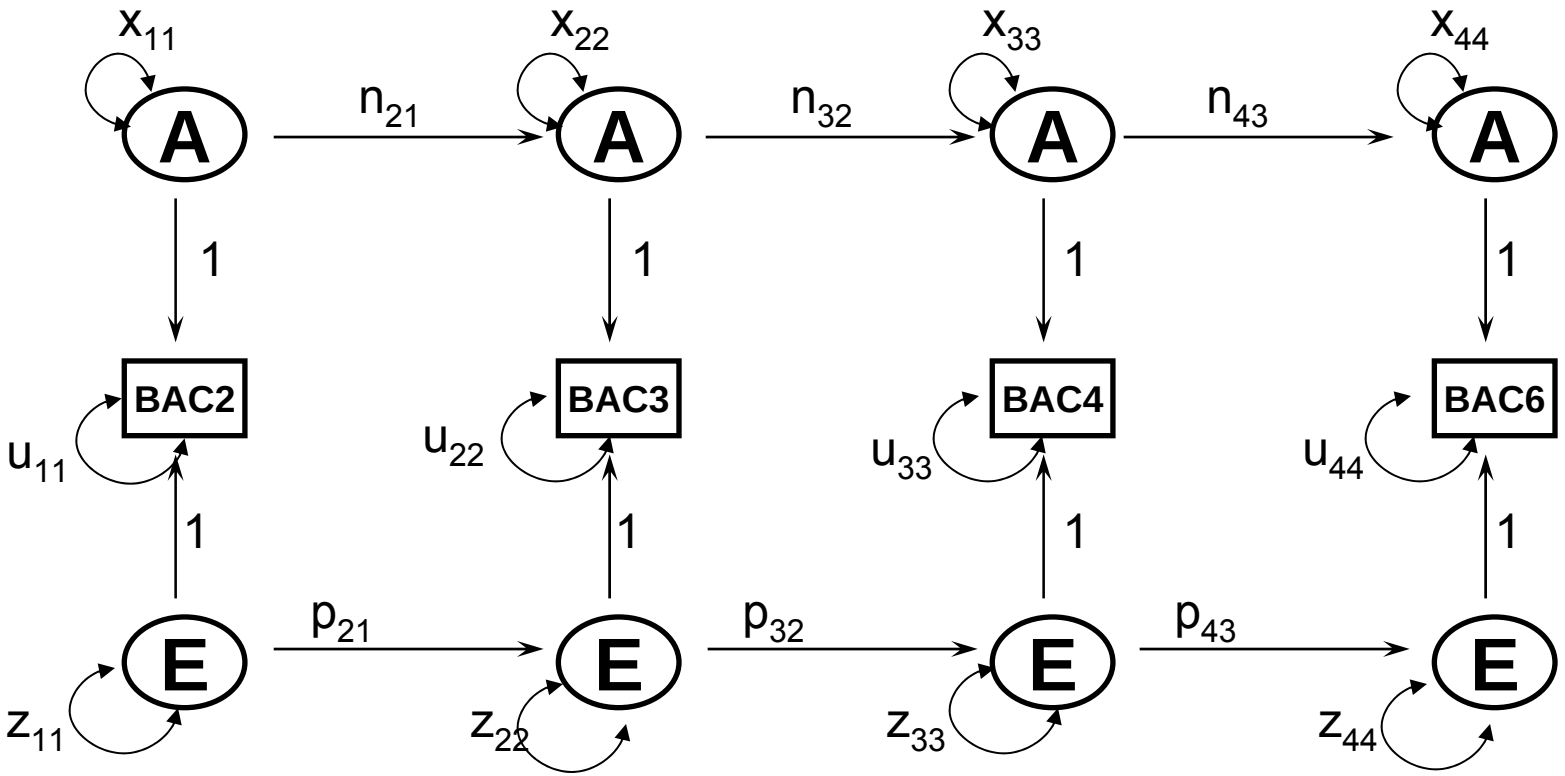
Simplex Model

A longitudinal study with 4 waves (single individual):



(Boomsma & Molenaar, 1987)

Genetic Simplex Model



x and z = genetic and nonshared environmental **innovations** respectively
n and p = genetic and nonshared environmental **transmission** respectively
u = **error** variances

Simplex Model

- Advantages:
  - Makes restrictive predictions about covariance pattern
  - Falsifiable

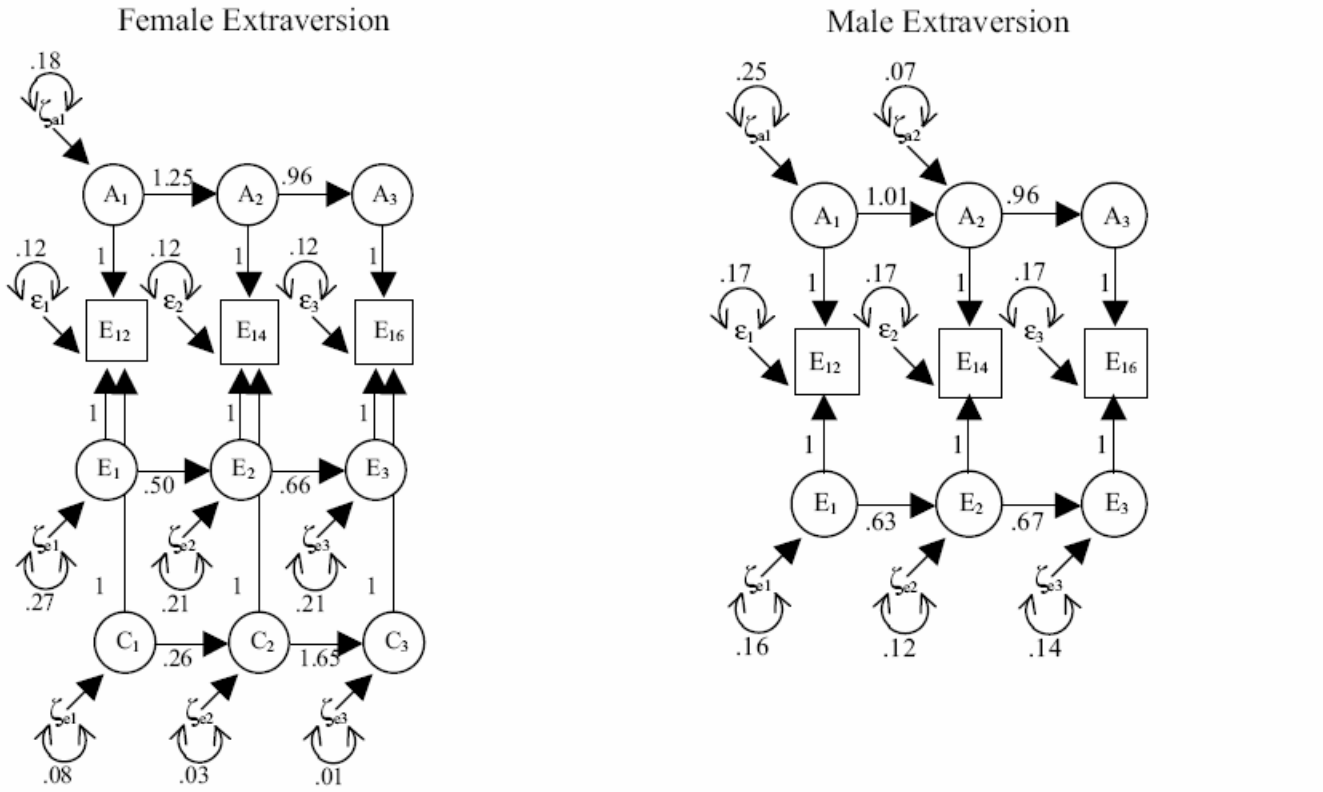


Figure 2

Best fitting genetic simplex model for female and male extraversion.

E_{12-16} = extraversion 12–16 yrs

A_{1-3}, E_{1-3}, C_{1-3} = additive genetic and nonshared and shared environmental effects

ζ_{a1-3}, ζ_{e1-3}, ζ_{c1-3} = additive genetic innovations, nonshared and shared environmental innovations

ε_{1-3} = error parameters 12–16 yrs

double/single headed arrows = variance components/path coefficients

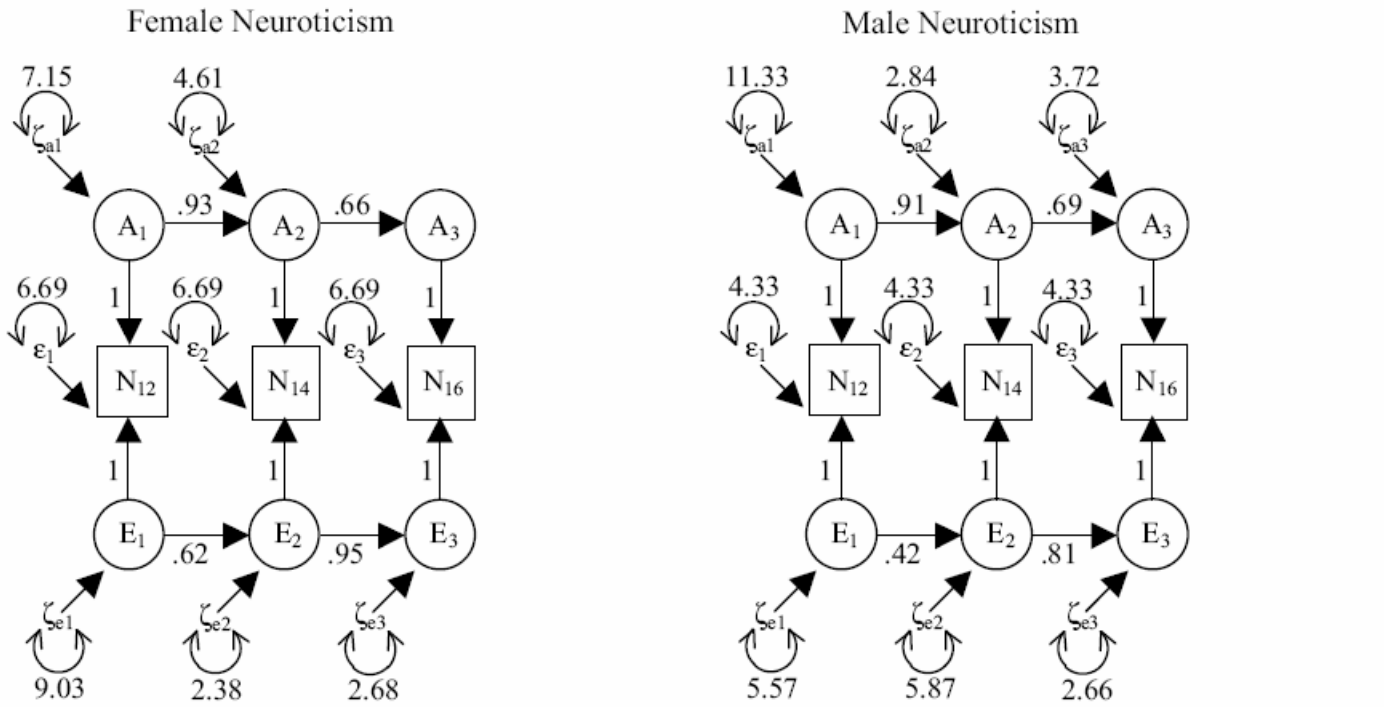


Figure 3

Best fitting genetic simplex model for female and male neuroticism.

$N_{12–16}$ = neuroticism 12–16 yrs

$A_{1–3}$, $E_{1–3}$ = additive genetic and nonshared environmental effects

$\zeta_{a1–3}$, $\zeta_{e1–3}$ = additive genetic innovations and nonshared environmental innovations

$\varepsilon_{1–3}$ = error parameters 12–16 yrs

double/single headed arrows = variance components/path coefficients

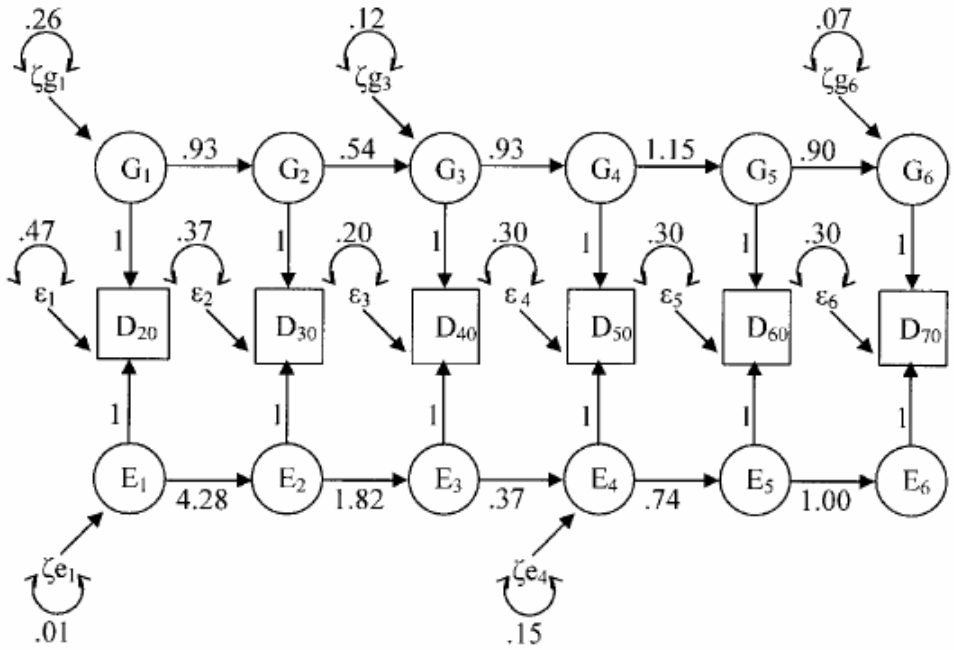


Figure 3

Best fitting simplex model for female depression with unstandardized variance components and path coefficients.

Note: $G_1 - G_6$ = additive genetic effects, $E_1 - E_6$ = non-shared environmental effects, ζ_g = additive genetic innovations, ζ_e = non-shared environmental innovations, ε = error terms (where $\varepsilon_3 = \varepsilon_4$)

Today's example

- Grant et al., 1999, Behavior Genetics, 29, 463-472.
- Australian alcohol challenge data, collected between 1979 and 1981
  - Mean age = 23.5 years
- Subjects drank 0.75 g/kg alcohol at a steady rate over a 20-minute period. Blood Alcohol Concentration (BAC) was assessed at 6 points after consumption:

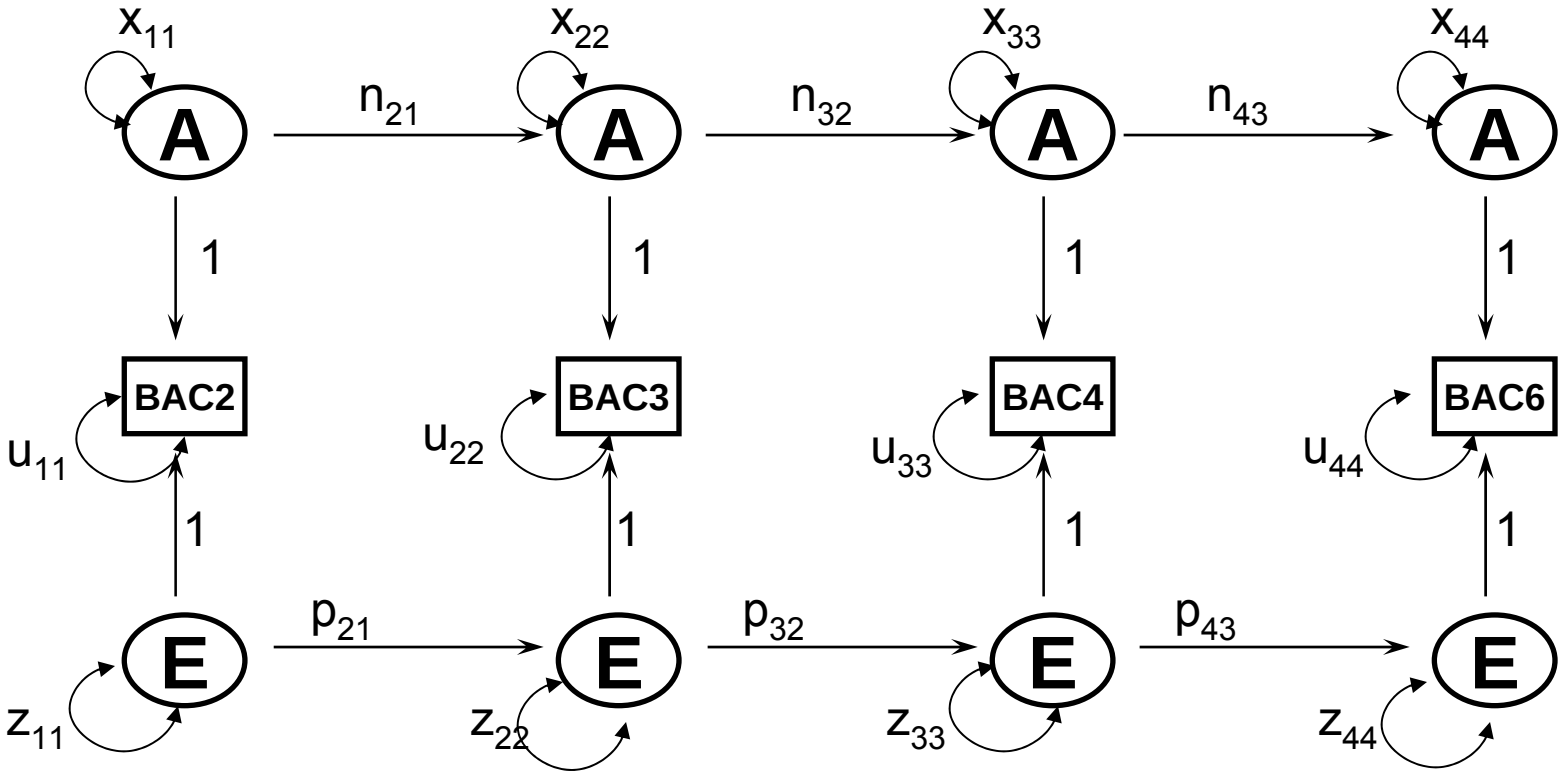
	Minutes post-consump.	Mean BAC	# of individuals with data	
			MZM (43 prs)	DZM (37 prs)
BAC 1	56	89.0	83	72
BAC 2	**68**	**88.9**	**83**	**74**
BAC 3	**83**	**88.8**	**84**	**71**
BAC 4	**123**	**80.9**	**86**	**74**
BAC 5	143	76.6	52	56
BAC 6	**182**	**67.7**	**83**	**74**

A simplex correlation pattern…

- Sample correlations (the DZM twin A quadrant of an intraclass correlation matrix)

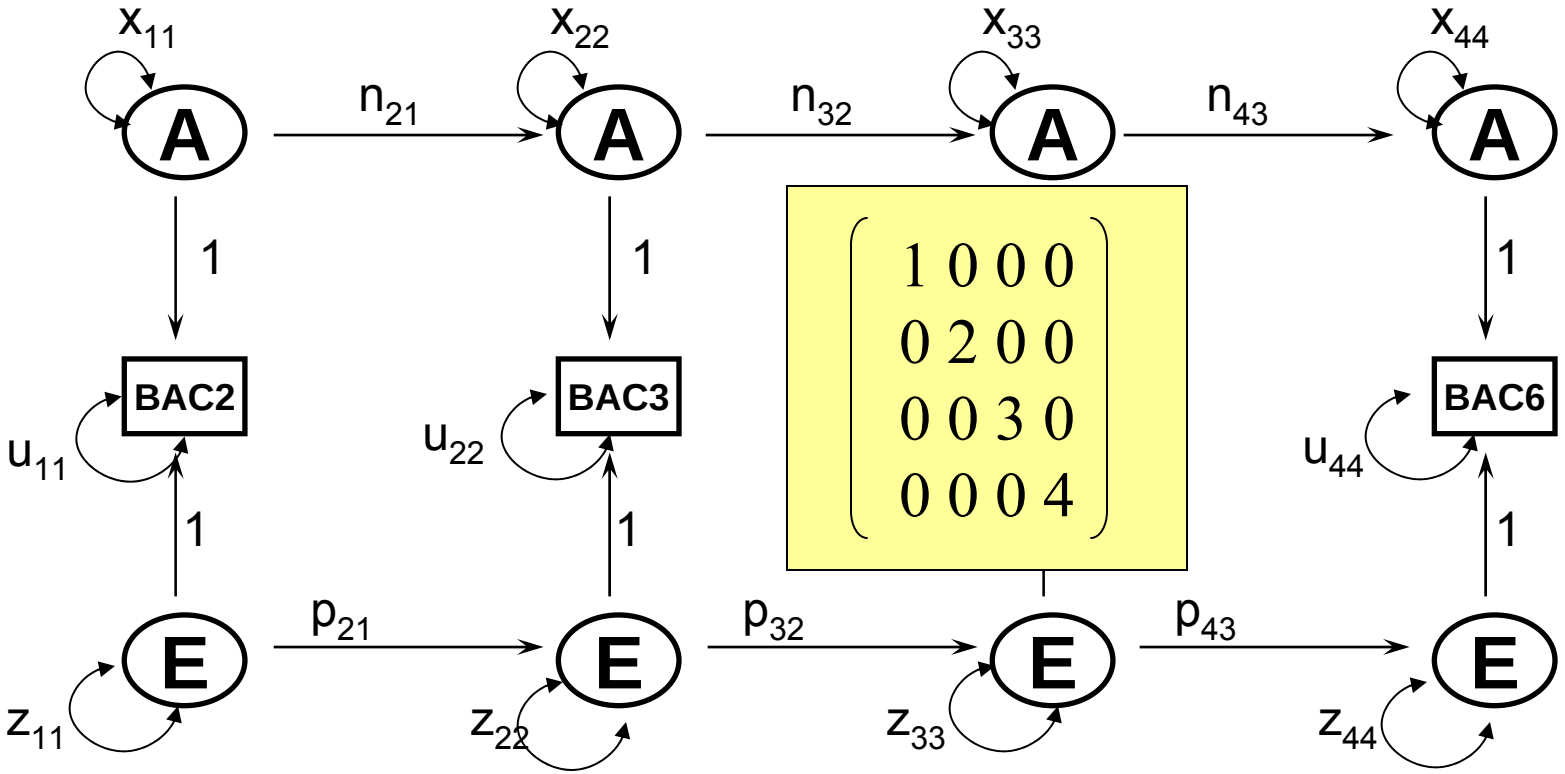
	BAC 2	BAC 3	BAC 4	BAC 6
BAC 2	1.00			
BAC 3	0.90	1.00		
BAC 4	0.69	0.84	1.00	
BAC 6	0.58	0.77	0.93	1.00

Practical - Simplex Model



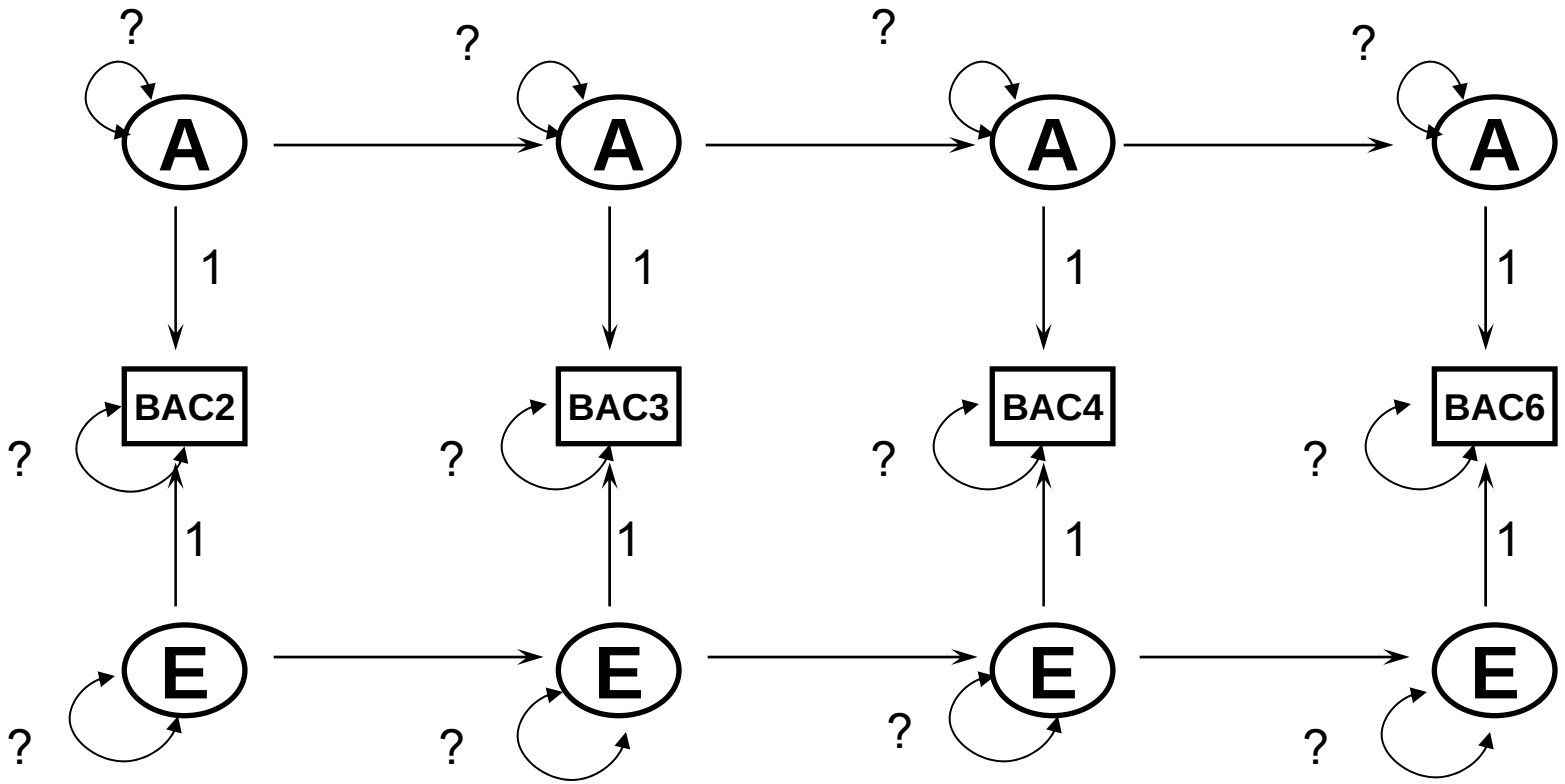
x and z = genetic and nonshared environmental **innovations** respectively
n and p = genetic and nonshared environmental **transmission** respectively
u = **error** variances

Practical - Simplex Model



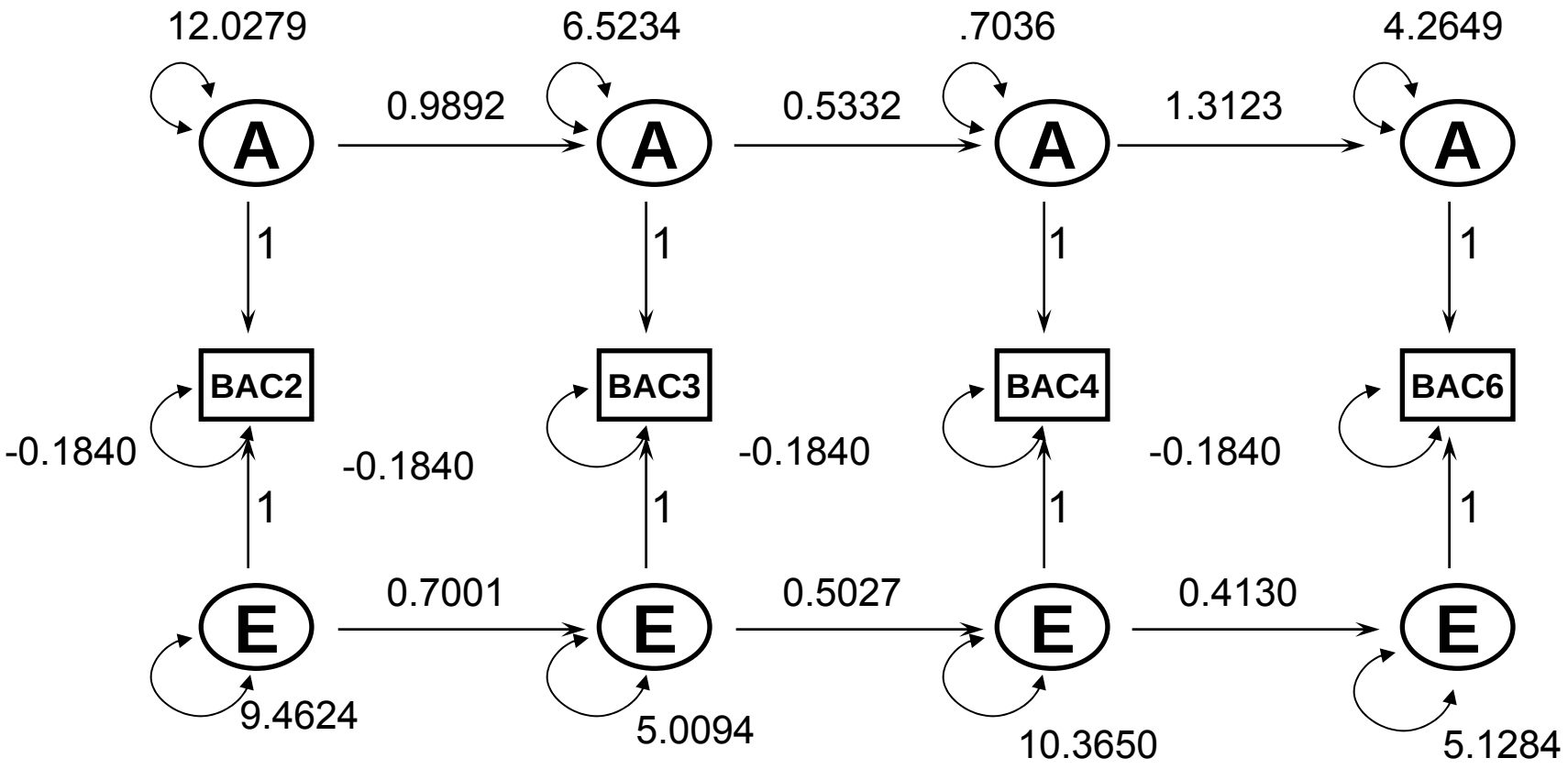
x and z = genetic and nonshared environmental **innovations** respectively
n and p = genetic and nonshared environmental **transmission** respectively
u = **error** variances

Practical - Simplex Model



x and z = genetic and nonshared environmental **innovations** respectively
n and p = genetic and nonshared environmental **transmission** respectively
u = **error** variances

Full Genetic Simplex Model



Basic_simplex.mxo -2*LL=4620.028, 23 est. parameters, 606 df

Sub-Models

1) Is the error variance on individual variable assessments significant?

2) Is the genetic innovations on BAC6 significant? BAC4? BAC2?

Sub-Models

1) Is the error variance on individual variable assessments significant?

    - drop 200

2) Is the genetic innovations on BAC6 significant? BAC4? BAC2?

    - drop 4, 3, 2

Simplex Model

- Advantages:
  - Makes restrictive predictions about covariance pattern
  - Falsifiable

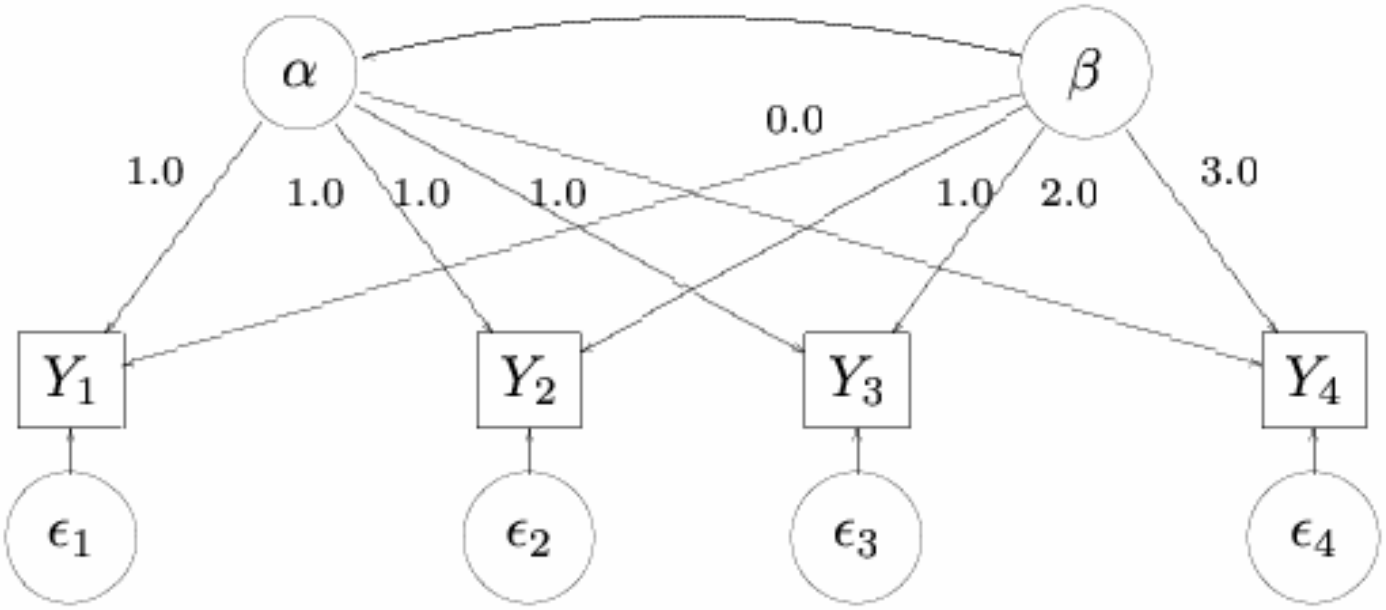
- Disadvantages:
  - Makes restrictive predictions about covariance pattern (future depends on current state only)
  - Number of parameters increases with number of measurements

Methods for Longitudinal Analysis

- Cholesky Models
- Simplex Models
- Growth Curve Models

Latent Growth Curve Model

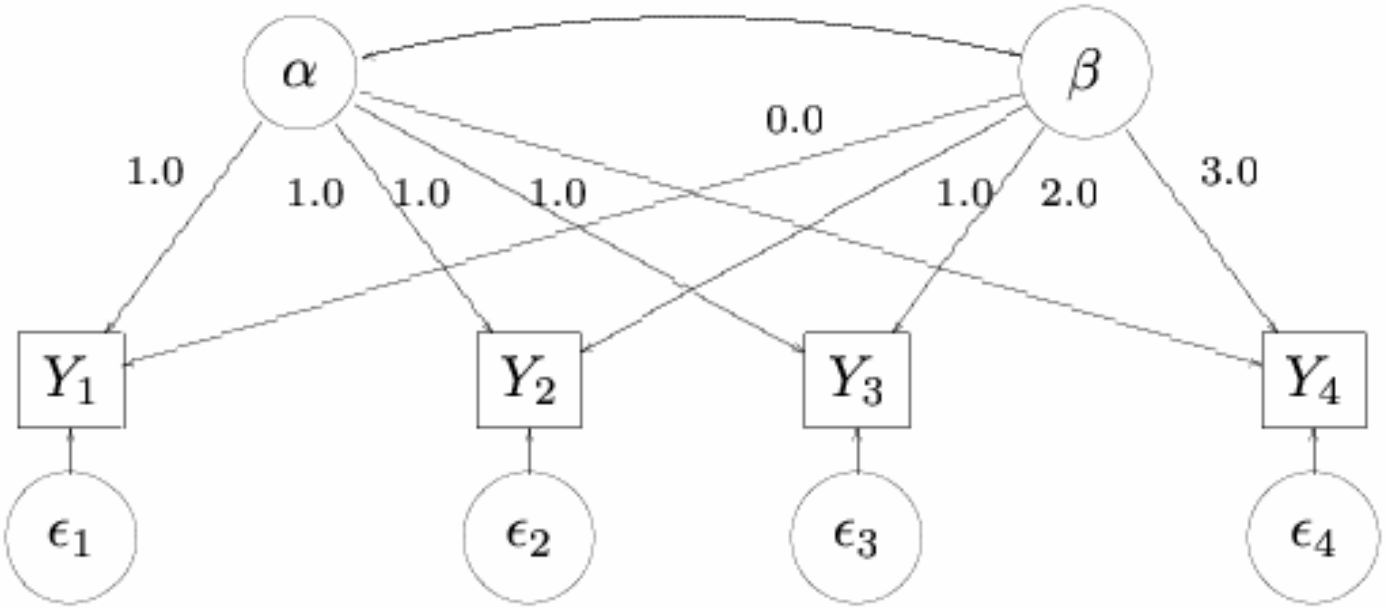
(shown here as linear)



- Mean Level of the Trait (Intercept)
- Rate of Change In Trait (Slope)

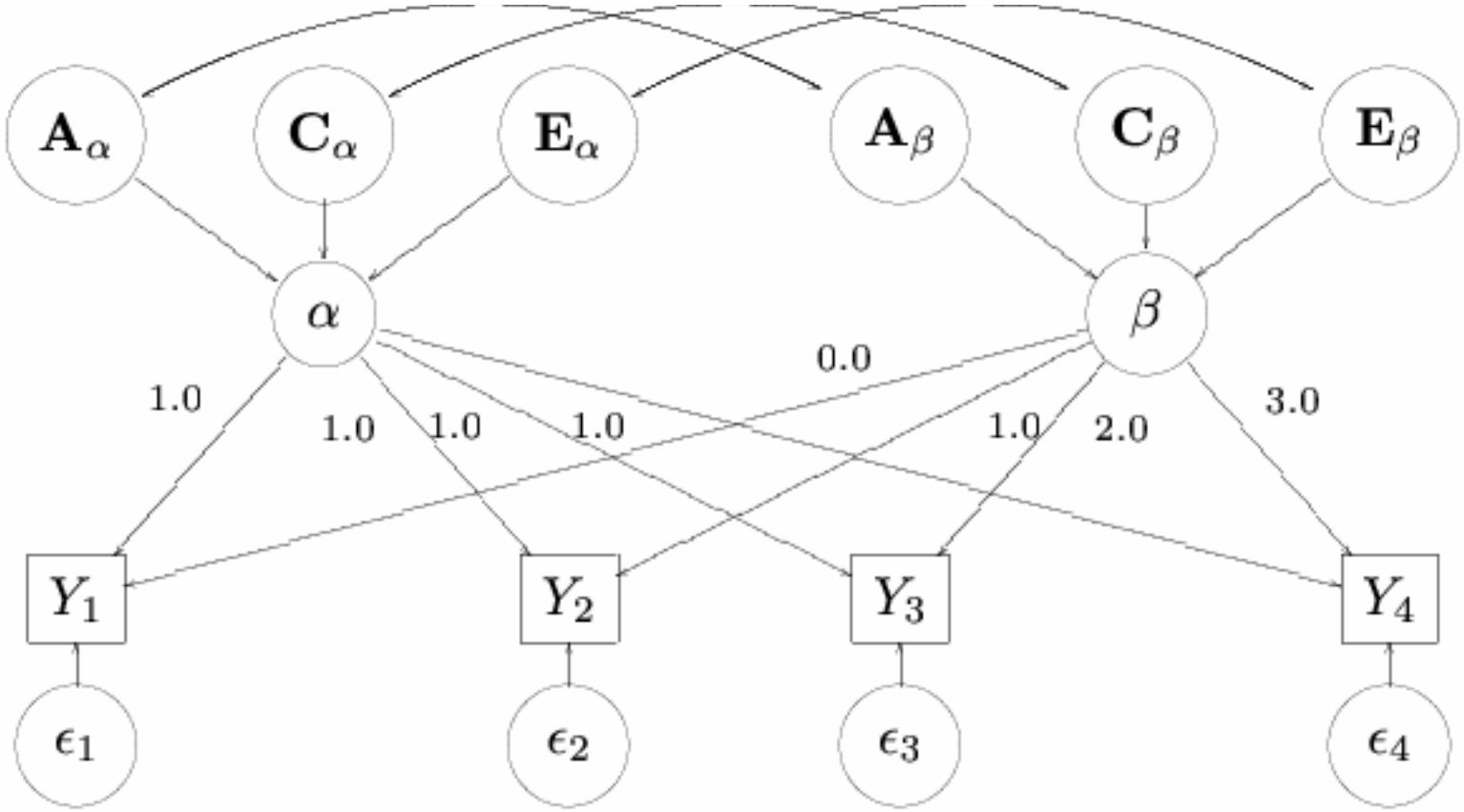
Latent Growth Curve Model

(shown here as linear)

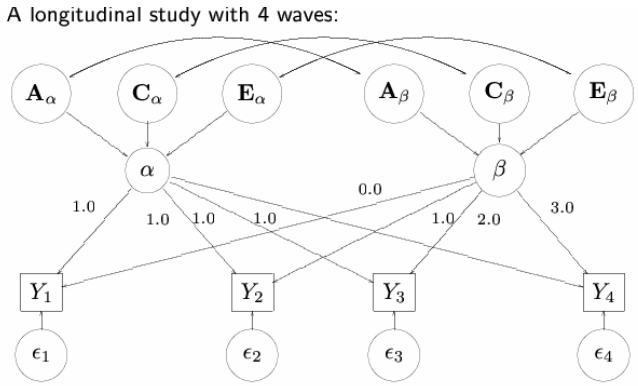


$$Y_{i,t} = \alpha_i + \beta_i(t-1) + \epsilon_{i,t}, \quad i = 1, 2 \quad t = 1, 2, 3, 4$$

Genetically Informative
Latent Growth Curve Model



Genetically Informative Latent Growth Curve Model



Methods of quantitative genetics:

$$\begin{cases} \alpha_i = \mathbf{G}_{\alpha,i} + \mathbf{E}_{\alpha,i} = A_{\alpha,i} + D_{\alpha,i} + C_{\alpha,i} + E_{\alpha,i} \\ \beta_i = \mathbf{G}_{\beta,i} + \mathbf{E}_{\beta,i} = A_{\beta,i} + D_{\beta,i} + C_{\beta,i} + E_{\beta,i} \end{cases}$$

for twin $i = 1, 2$.

Where:

- $(A_{\alpha,i}, A_{\beta,i})$ are the additive genetic components of intercept and slope
- $(D_{\alpha,i}, D_{\beta,i})$ are the respective dominant genetic components
- $(C_{\alpha,i}, C_{\beta,i})$ are the respective shared environmental components
- $(E_{\alpha,i}, E_{\beta,i})$ are the respective non-shared environmental components

→Like a bivariate model

Growth Model Questions

- What is the contribution of genetic/environmental factors to the variation of α (intercept) and β (slope)?
- Same or different genes influencing α (intercept) and β (slope)?
- Same or different environments influencing α (intercept) and β (slope)?

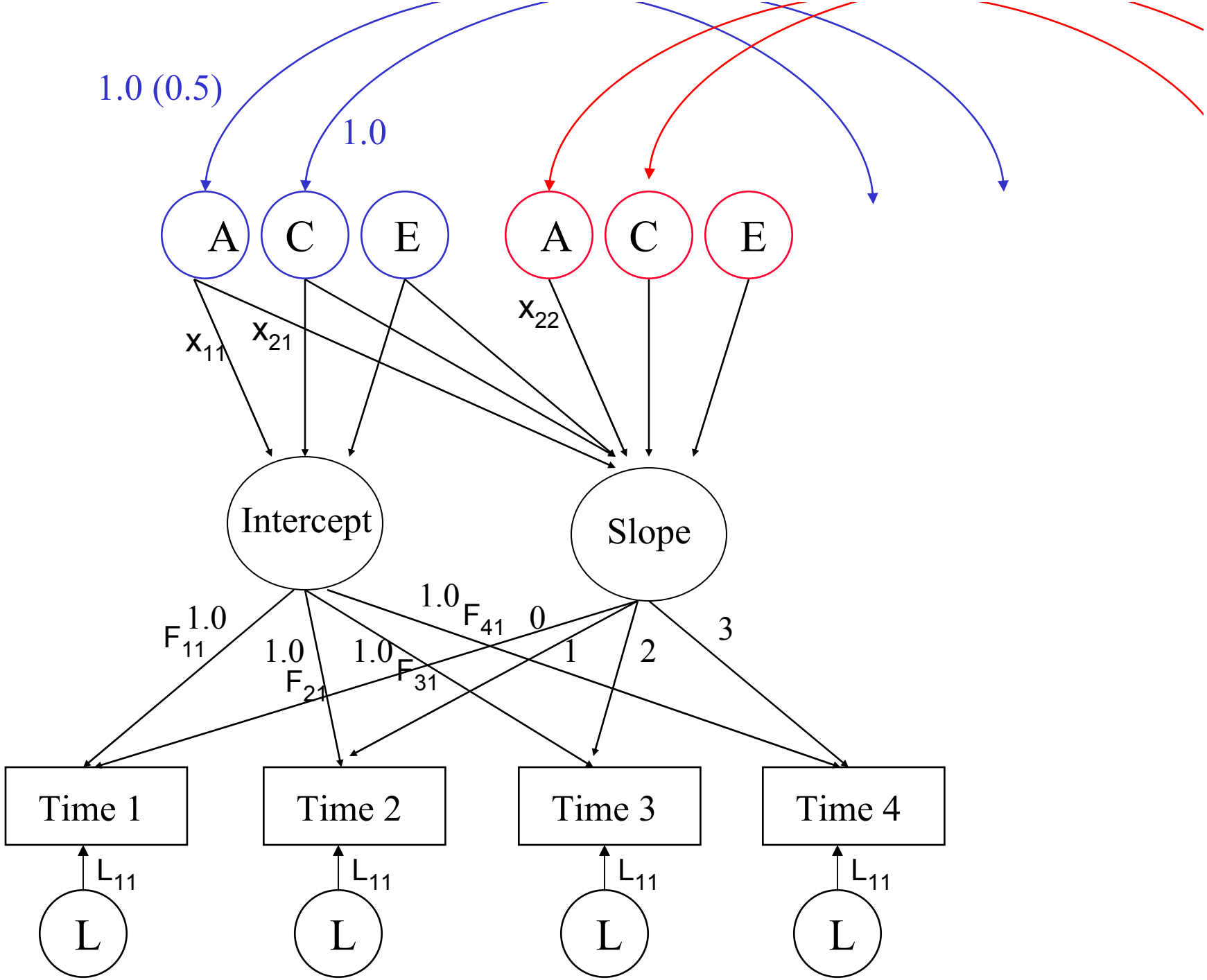
Practical

- Mx latent growth curve example

  (script from http://www.psy.vu.nl/mxbib/)

Submodels to test:

1. No covariance between slope and intercept
2. No genetic effect on intercept
3. No genetic effect on slope
4. No common environmental effect on intercept
5. No common environmental effect on slope
6. Best fitting model? (i.e., ACE, AE, CE, E?)


1.0 (0.5)
1.0
A
C
E
A
C
E
x11
x21
x22
Intercept
Slope
1.0
F11
1.0
F21
1.0
F31
1.0
F41
0
1
2
3
Time 1
Time 2
Time 3
Time 4
L11
L11
L11
L11
L
L
L
L

Practical

- Mx latent growth curve example

  (script from http://www.psy.vu.nl/mxbib/)

Submodels to test:

1. No covariance between slope and intercept – signif decrease in fit
2. No genetic effect on intercept – signif decrease in fit
3. No genetic effect on slope – signif decrease in fit
4. No common environmental effect on intercept -- ns
5. No common environmental effect on slope -- ns
6. Best fitting model? (i.e., ACE, AE, CE, E?) -- AE

Growth Curve Model

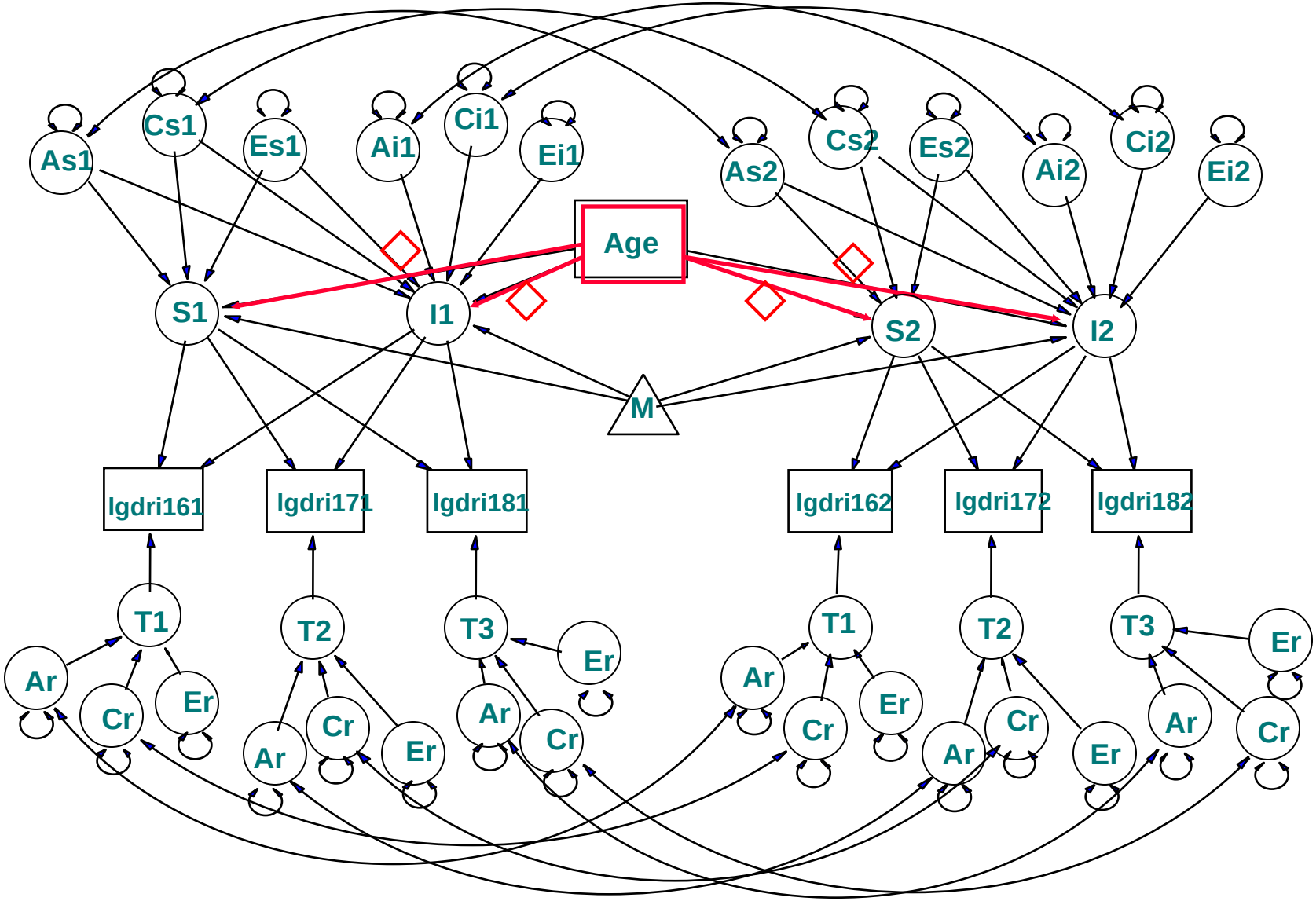
- Advantages:
  - Very efficient: number of parameters does not increase with number of measurements
  - Provides prediction about behavior beyond measured timepoints
- Disadvantages:
  - Note regarding slope parameters
  - Can be computationally intense
  - Assumptions to reduce computational burden
    - Linearity, no genetic effects on residuals, equal variance among residuals at differing timepoints

Latent Growth Curve Modeling Additional Considerations

- Standard approach assumes data are collected at identical set of fixed ages for all individuals (e.g., start at age 12, yearly assessments)

- Age heterogeneity and unequal spacing of measurements can be handled using definition variables
  - Mehta & West, 2000, Psychological Methods

Latent Growth Curve Model with Measured Variable

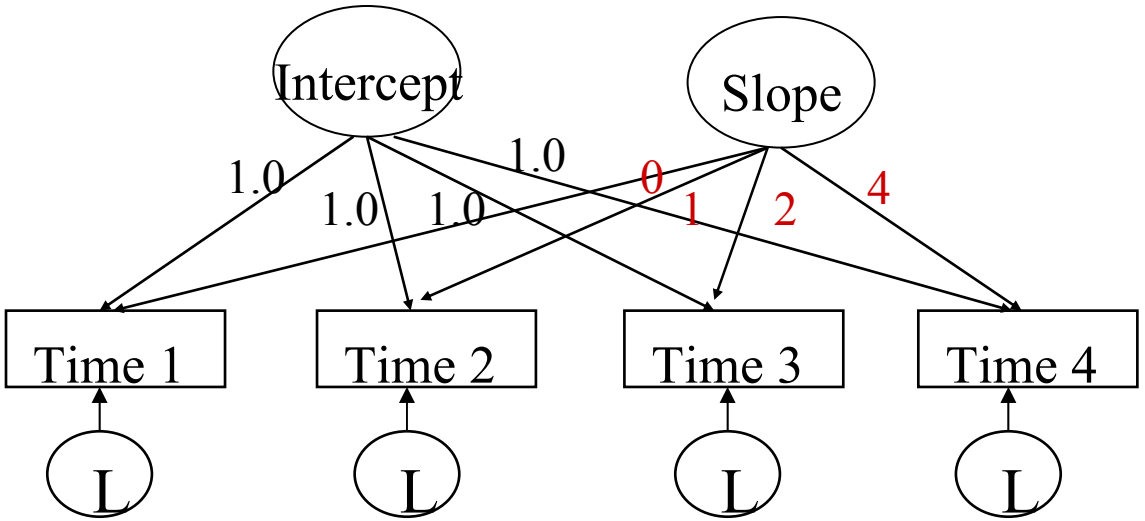


Extensions of Growth Curve Models

- Incorporation of measured variables (genotype, environment)

- Nonlinear growth

  - Neale, MC & McArdle, JJ (2000). A structured latent growth curves for twin data. Twin Research, 3, 165-177.



Latent Growth Curve Modeling

- McArdle, JJ (1986). Latent variable growth within behavior genetic models. Behavior Genetics, 16, 163-200.
- Baker, LA et al. (1992). Biometrical analysis of individual growth curves. Behavior Genetics, 22, 253-264.
- McArdle, JJ et al. (1998). A contemporary method for developmental -genetic analyses of age changes in intellectual abilities. Developmental Neuropsychology, 14, 69-114.

Summary of Longitudinal Models

- Cholesky Model
  - Few assumptions, predict any pattern of correlations
  - Not falsifiable
  - Limited measurements
- Simplex Model
  - Falsifiable
  - Limited measurements
- Growth Curve Model
  - G, E influences on initial level, rate of change
  - Unlimited measurements
  - Computationally intensive, assumptions