Longitudinal Modeling

Nathan Gillespie & Dorret Boomsma

\\nathan\2008\Longitudinal

- **neuro_f_chol.mx**
- **neuro_f_simplex.mx**
- **jepq6.dat**

Why conduct longitudinal analyses?

1. Can improve power by using multiple observations from the same individual
    - Cross twin cross trait correlation
2. Can examine and estimate time-dependent genetic and environmental effects
    - Changing magnitude of genetic & environmental influence across time
    - Same versus different genes across development

Methods for Longitudinal Data Analysis

1. Cholesky Models
2. Simplex Models

    Eaves *et al.*, 1986

    Boomsma & Molenaar, 1987
3. Growth Curve Models

Aims

1. Revisit the Mx trivariate Cholesky
2. Take a look at simplex or auto-regression models

   - Explain some of the underlying theory of this form of longitudinal modelling

   - Run through an Mx script
3. Compare the Cholesky and simplex models

Longitudinal modeling of adolescent personality

Introduce longitudinal modeling in the context of personality change

Eaves, Eysenck & Martin (1989)

- Adult personality - High of genetic continuity over time
- Effect stronger in Neuroticism vs Extraversion

Genetic continuity in adolescents?

Personality Data

Twin Mole and Twin MAPS projects

Assess the genetic / environmental etiology of Melanocytic Naevi (common moles) in twins aged 12 & 14 years + Cognition at 16

81 items JEPQ: Psychoticism (P), Extraversion (E), and Neuroticism (N)

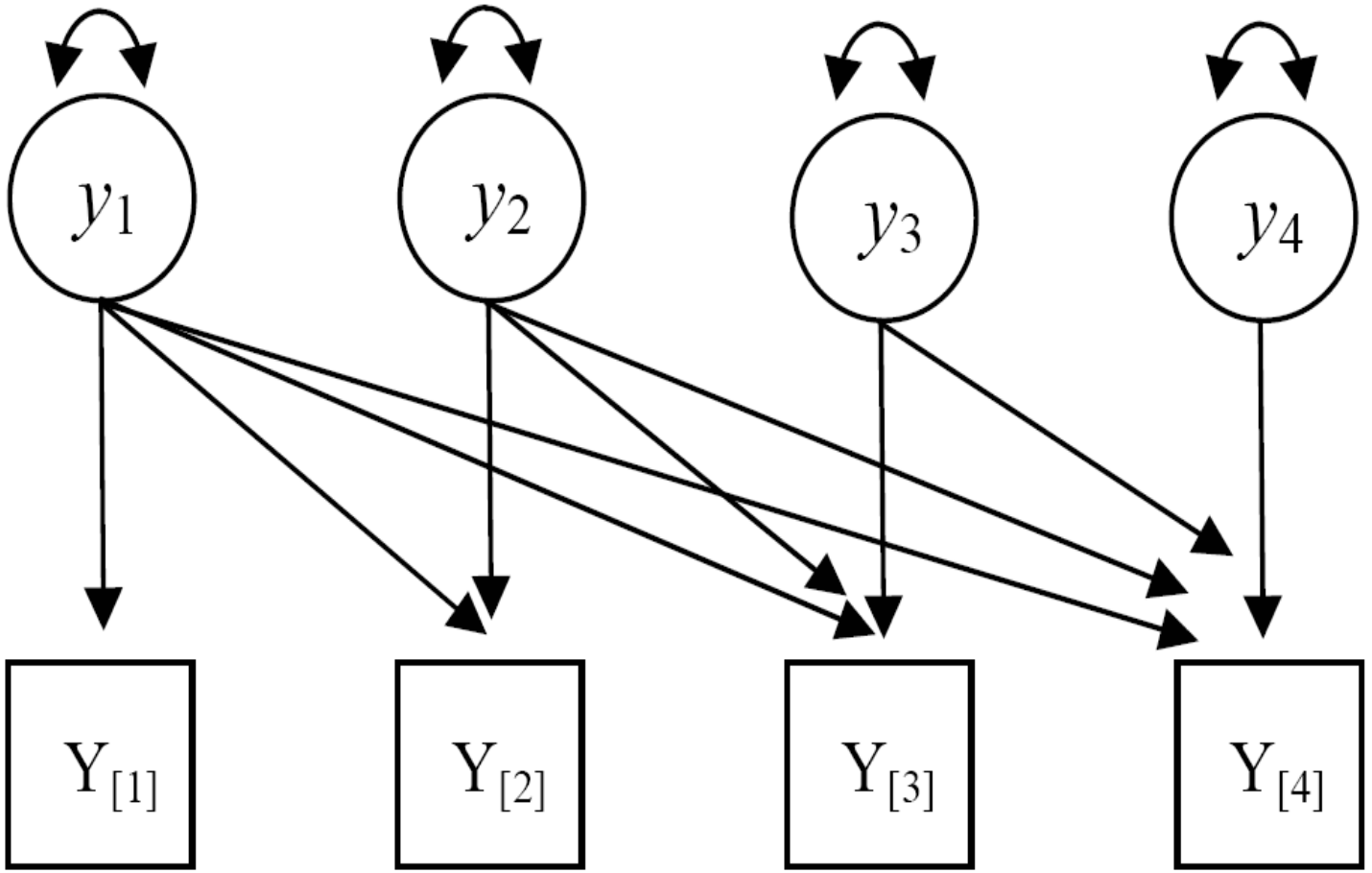
	Twin Mole								Twin MAPS			
	12 yrs				14 yrs				16 yrs			
	P	E	N	L	P	E	N	L	P	E	N	L
Male	603	604	606	605	465	466	466	467	416	412	416	415
Female	605	602	607	609	471	470	473	473	442	438	442	442

Personality Data

Raw continuous data methods

Assumptions of mean and variance
homogeneity by twin order and zygosity
(and necessarily by sex and age)

1. Cholesky Model



Cholesky Model

Advantages:

- Logical: organized such that all factors are constrained to impact later, but not earlier time points
- 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 time points)
- 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

Run ACE Cholesky Model

neuro_f_chol.mx

```
! Female Neuroticism at 12, 14 and 16 years
! Multivariate Cholesky
#Ngroups 5
#define nvar  3 ! variables
#define nsib  2 ! twin-1 & twin-2
#define nitem 6
!
!_____________________________________________________
!

G1: Model paramaters
Calculation
Begin Matrices;
X Lower nvar nvar Free
Y Lower nvar nvar Free
Z Lower nvar nvar Free
H Full 1 1
End Matrices;

Matrix H 0.5
```

Model Fit for Female Neuroticism

Model	LL	df	∆2LL	∆df	p	BIC
ACE Cholesky	10424.88	1803	-	-	-	2352.134

Cholesky Model Fitting Results

Phenotypic correlations

MATRIX Q
This is a computed FULL matrix of order 6 by 6
[=\STND(A+C+E|A+C_A+C|A+C+E)]

	1	2	3	4	5	6
1	1.0000	0.5320	0.4530	0.3099	0.2877	0.1969
2	0.5320	1.0000	0.5867	0.2877	0.4616	0.3246
3	0.4530	0.5867	1.0000	0.1969	0.3246	0.2716
4	0.3099	0.2877	0.1969	1.0000	0.5320	0.4530
5	0.2877	0.4616	0.3246	0.5320	1.0000	0.5867
6	0.1969	0.3246	0.2716	0.4530	0.5867	1.0000

Proportions of variance

MATRIX S
This is a computed FULL matrix of order 3 by 9
[=A%(A+C+E)|C%(A+C+E)|E%(A+C+E)]

	1	2	3	4	5	6	7	8	9
1	0.2884	0.5030	0.3395	0.0215	0.0377	0.0951	0.6901	0.4593	0.5654
2	0.5030	0.4429	0.4848	0.0377	0.0187	0.0685	0.4593	0.5384	0.4467
3	0.3395	0.4848	0.1853	0.0951	0.0685	0.0863	0.5654	0.4467	0.7284

A
C
E

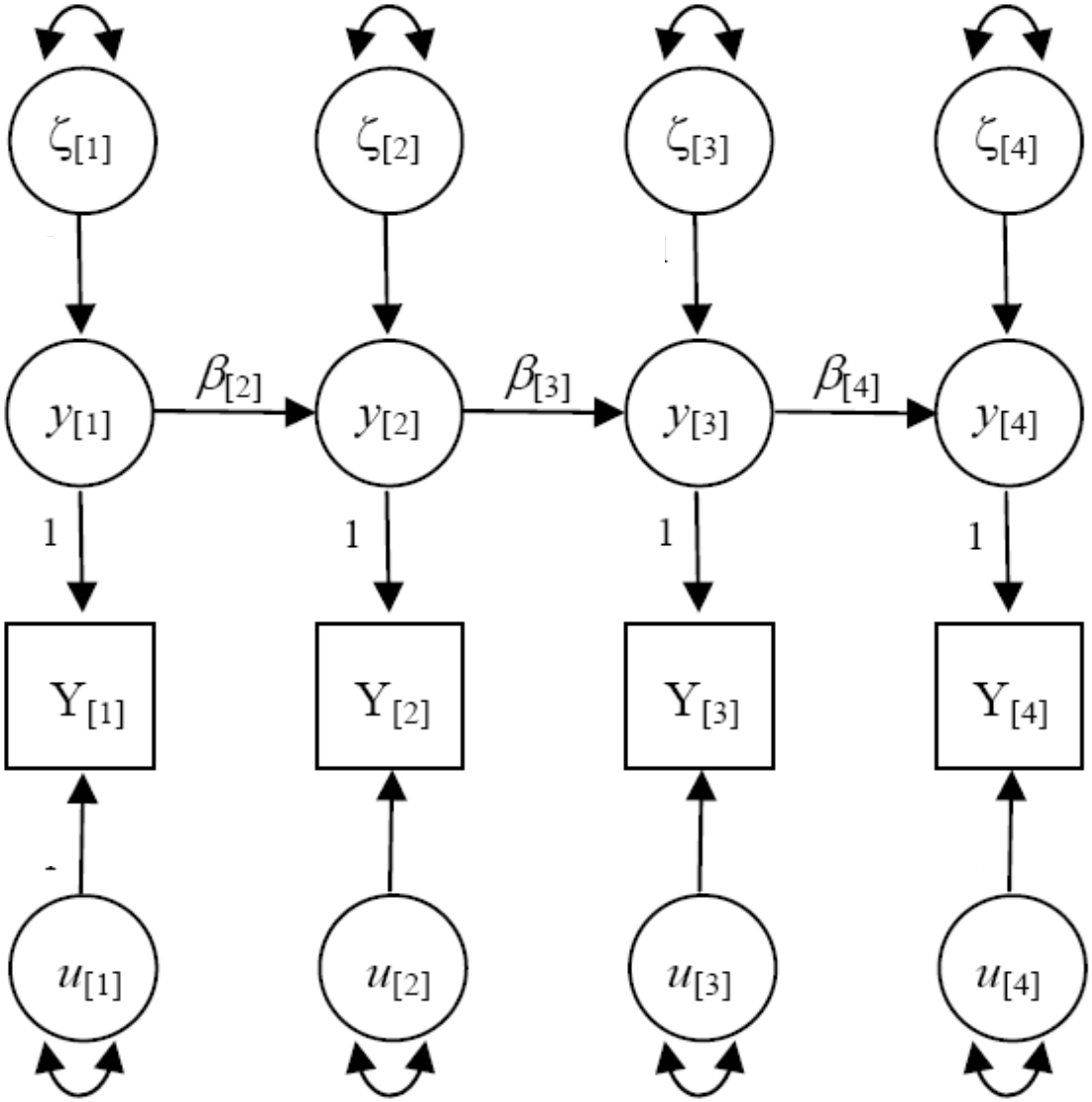
Cholesky Model Fitting Results

```
 MATRIX T
This is a computed FULL matrix of order    3 by    3
 [=\STND(A)]
              1          2          3
1      1.0000     0.7487     0.6652
2      0.7487     1.0000     0.9930
3      0.6652     0.9930     1.0000

 MATRIX U
This is a computed FULL matrix of order    3 by    3
 [=\STND(C)]
              1          2          3
1      1.0000     1.0000     1.0000
2      1.0000     1.0000     1.0000
3      1.0000     1.0000     1.0000

 MATRIX V
This is a computed FULL matrix of order    3 by    3
 [=\STND(E)]
              1          2          3
1      1.0000     0.4009     0.3613
2      0.4009     1.0000     0.4185
3      0.3613     0.4185     1.0000
```

2. Simplex Models



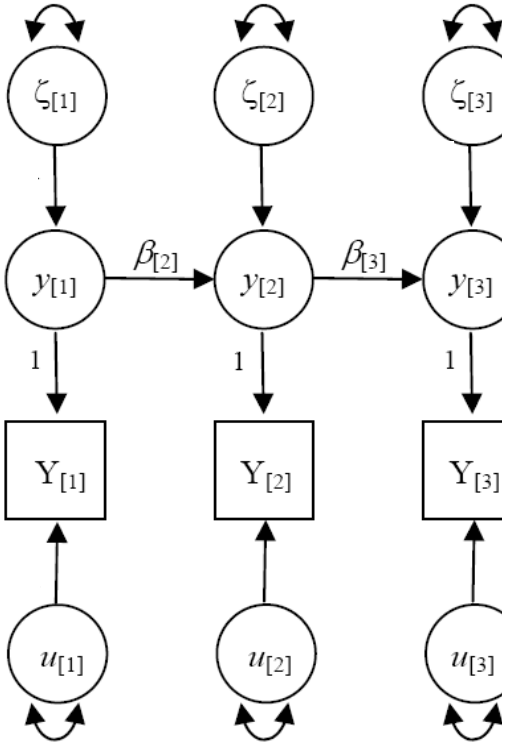
Simplex or autoregressive models

Simplex designs model changes in true scores ($y_{[t]}$) over time by fitting auto-regressive or Markovian chains

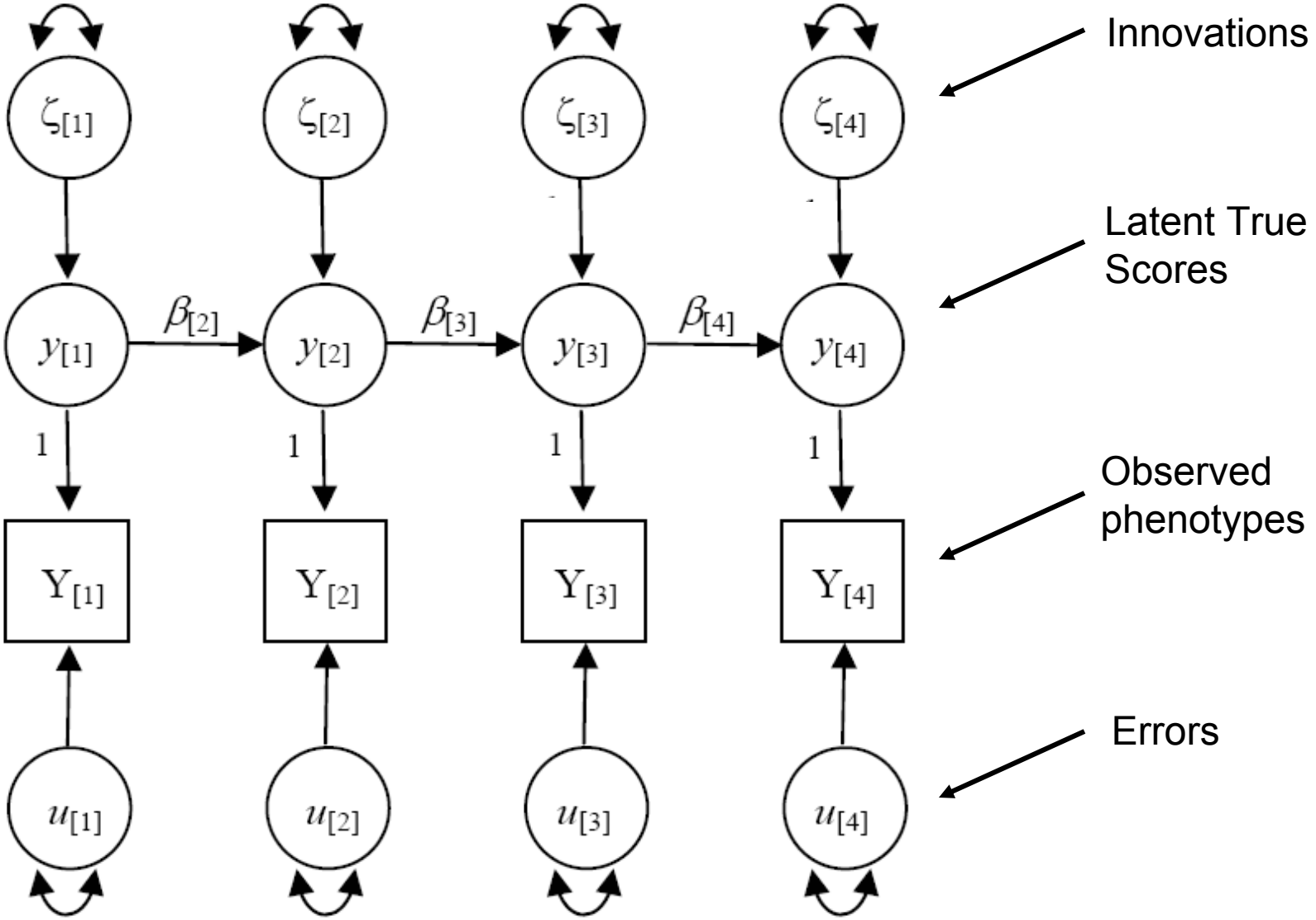
Each true score is predicted to be causally related to the immediately preceding latent true score in a linear fashion

$$y_{[t]n} = \beta_{[t]n} \times y_{[t-1]n} + \zeta_{[t]n}$$

$\beta_{[t]}$ = linear regression of latent factor ($y_{[t]}$) on the previous latent factor ($y_{[t-1]}$), $\zeta_{[t]}$ = new input, change or innovation at time $_{[t]}$, uncorrelated with $y_{[t-1]}$



Simplex Model



ACE Simplex Model script

neuro_f_simplex.mx

```
! Neuroticism - Females
!
#Ngroups 5      !
#define nsib=2  !number of siblings
#define nvar=3  !number of variables

G1: Model paramaters
Calculation
Begin Matrices;
X Diag  nvar nvar Free  ! Genetic innovations
N Lower nvar nvar Free  ! Genetic transmission paths
Y Diag  nvar nvar Free  ! Common environmental innovations
O Lower nvar nvar Free  ! Common environmental transmission paths
Z Diag  nvar nvar Free  ! Specific environmental innovations
P Lower nvar nvar Free  ! Specific environmental transmission paths
U Diag  nvar nvar Free  ! Measurement error variance
I Identity nvar nvar
H Full 1 1
End Matrices;

Matrix H .5

Specify N
0
10 0
0 11 0
```

Model Fitting Results

Model	LL	df	Δ2LL	Δdf	p	BIC
ACE Cholesky	10424.88	1803	-	-	-	2352.13
Full Simplex						
AE Simplex						
CE Simplex						
E Simplex						

Model Fitting Results

Model	LL	df	∆2LL	∆df	p	BIC
ACE Cholesky	10424.88	1803	-	-	-	2352.13
Full Simplex	10425.113	1805	0.23	2	0.89	2349.08
AE Simplex	**10425.339**	**1810**	**0.23**	**5**	**0.99**	**2341.26**
CE Simplex	10432.45	1810	7.34	5	0.20	2344.81
E Simplex	10473.31	1815	48.20	10	<0.001	2357.31

Test for non-significant parameters

Run confidence intervals on all parameters

```
! AE Simplex Structure
Get neuro_f.mxs
Dr 12 13                ! C transmission coefficients
DR 4 5 6                ! C innovations
IN X 1 3 3
ENd
```

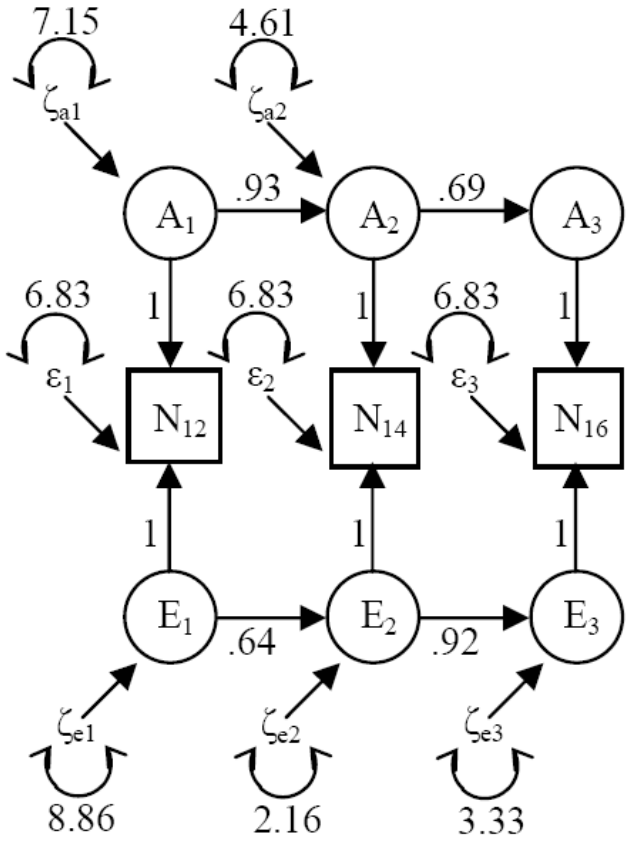
Matrix	Element			Int.	Estimate	Lower	Upper	Lfail	Ufail
X	1	3	3	95.0	0.8752	-1.7949	1.7949	0 2	0 2

```
! AE Simplex Structure, Remove final genetic innovation
Get neuro_f.mxs
Dr 12 13                ! C transmission coefficients
DR 4 5 6                ! C innovations
DR 3
ENd
```

Model Fitting Results

Model	LL	df	Δ2LL	Δdf	p	BIC
ACE Cholesky	10424.88	1803	-	-	-	2352.13
Full Simplex	10425.113	1805	0.23	2	0.89	2349.08
AE Simplex	10425.339	1810	0.23	5	0.99	2341.26
Dropζ_{a3}	**10425.70**	**1811**	**0.58**	**6**	**1.00**	**2339.85**
CE Simplex	10432.45	1810	7.34	5	0.20	2344.81
E Simplex	10473.31	1815	48.20	10	<0.001	2357.31

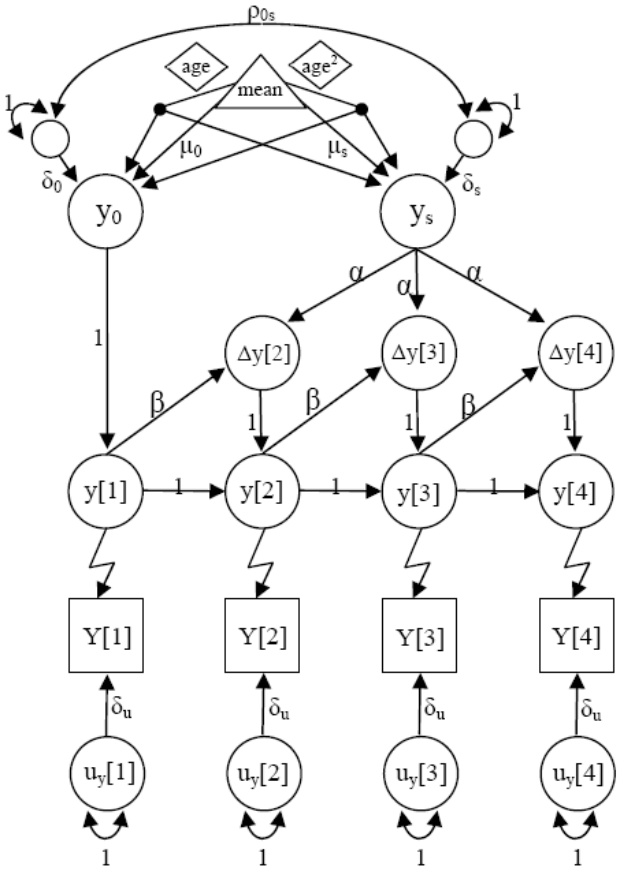
Best Fitting Model for Female Neuroticism



- Degree of genetic continuity
- Age specific genetic effects
- Genetic innovation at 14 years
- Is it related to developmental or hormonal changes during puberty and psychosexual development?

Additional Longitudinal Models

Dual Change Score (DCS) Model for Ordinal Data



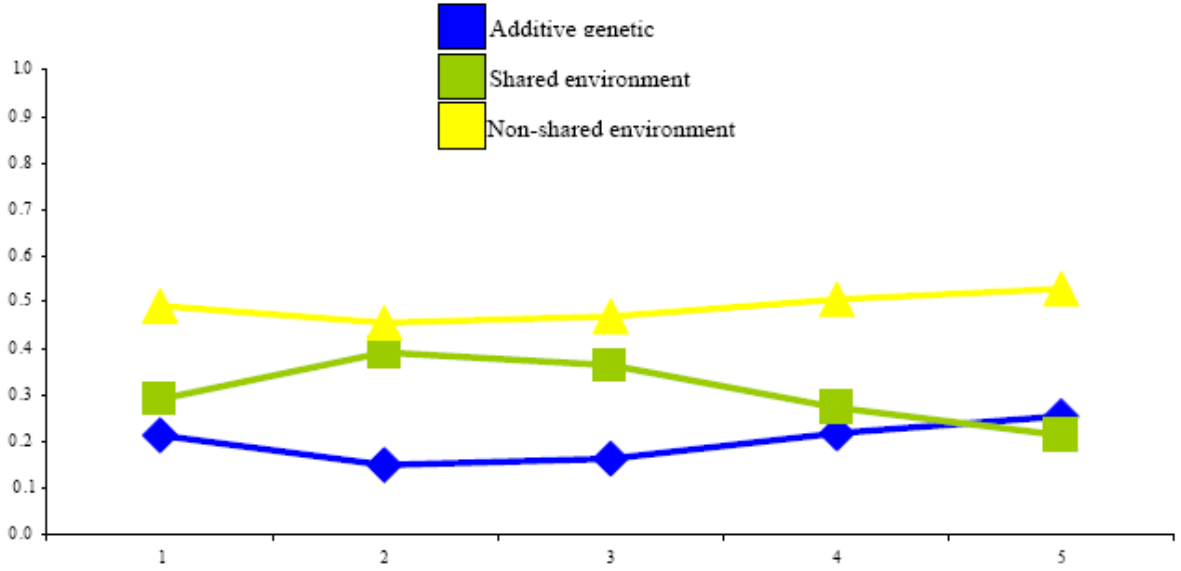
Y_0 = latent intercept, y_s = rate of change, μ_0, μ_s = latent intercept and slope means, δ_0, δ_s = latent intercept and slope variances, ρ_{0s}= correlation between intercept and slope, =$\Delta y[2\text{-}4]$ = latent difference scores, $u_y[1\text{-}4]$ = random error, α = systematic change. B = systematic proportional change. Also included are definition variables (diamonds) to adjust the mean (triangle) intercept and slope for the linear and quadratic effects of age at time of measurement. The sharp S-shaped single headed arrows represent the links between the observed ordinal measures (squares) and their corresponding underlying latent variables (circles).

Squeezing Interval Change From Ordinal Panel Data: Latent Growth Curves With Ordinal Outcomes

Paras D. Mehta
University of Illinois at Chicago

Michael C. Neale
Virginia Commonwealth University

Brian R. Flay
University of Illinois at Chicago

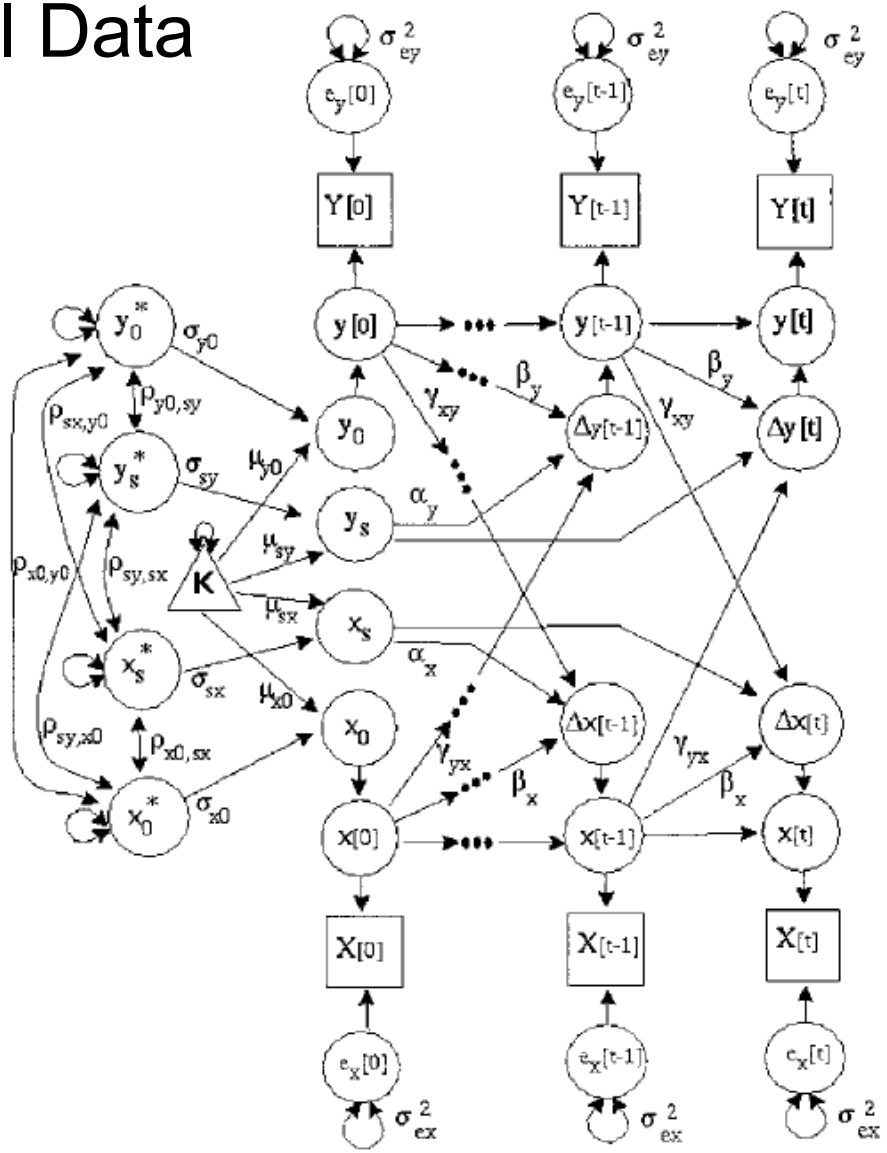


Longitudinal modeling of genetic and environmental influences on self-reported availability of psychoactive substances: alcohol, cigarettes, marijuana, cocaine and stimulants

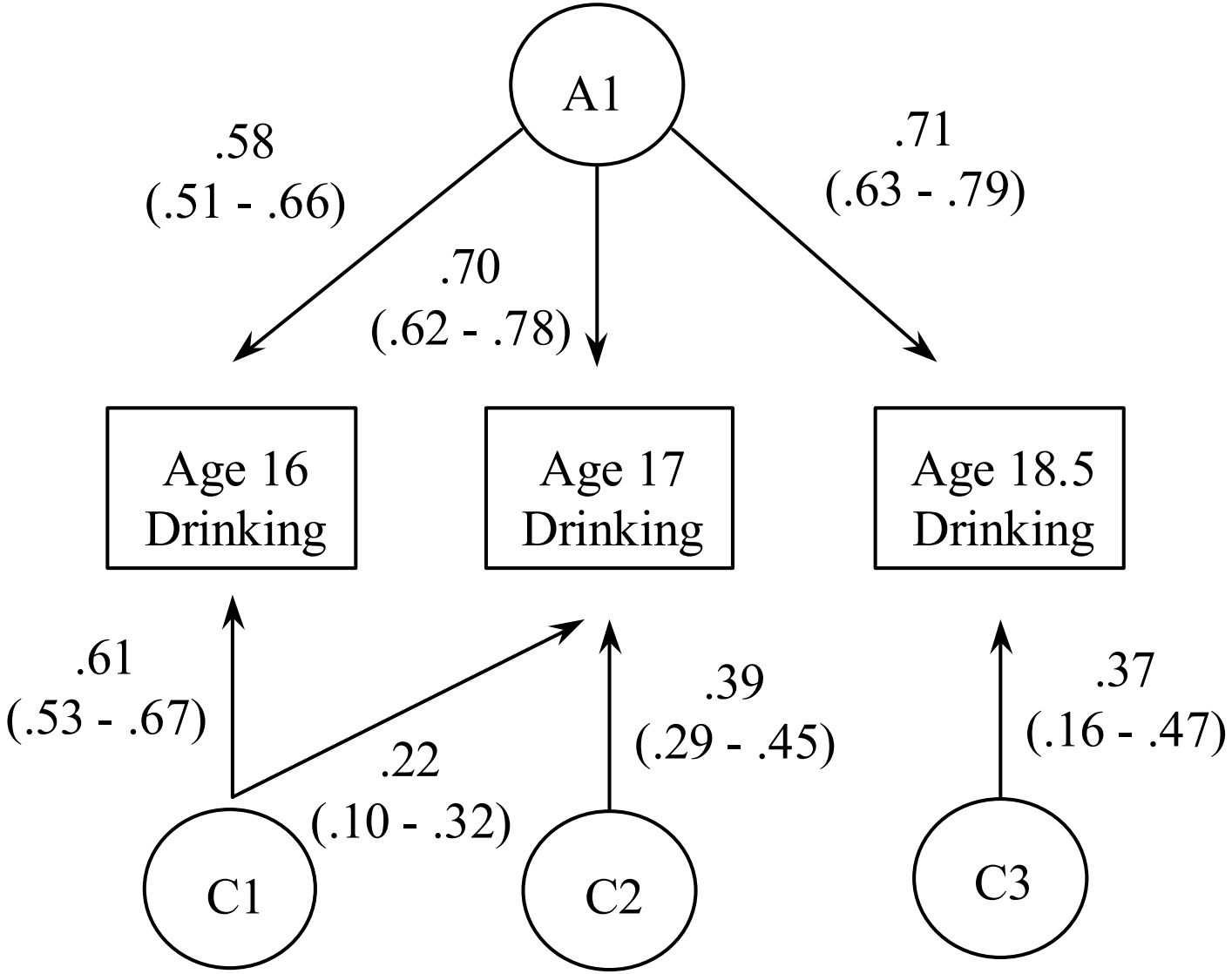
NATHAN A. GILLESPIE[1*], KENNETH S. KENDLER[1,2], CAROL A. PRESCOTT[2], STEVEN H. AGGEN[1], CHARLES O. GARDNER JR[1], KRISTEN JACOBSON[4] AND MICHAEL C. NEALE[1]

Additional Longitudinal Models

Bivariate Dual Change Score (DCS) Model for Ordinal Data



Best-fitting model for drinking

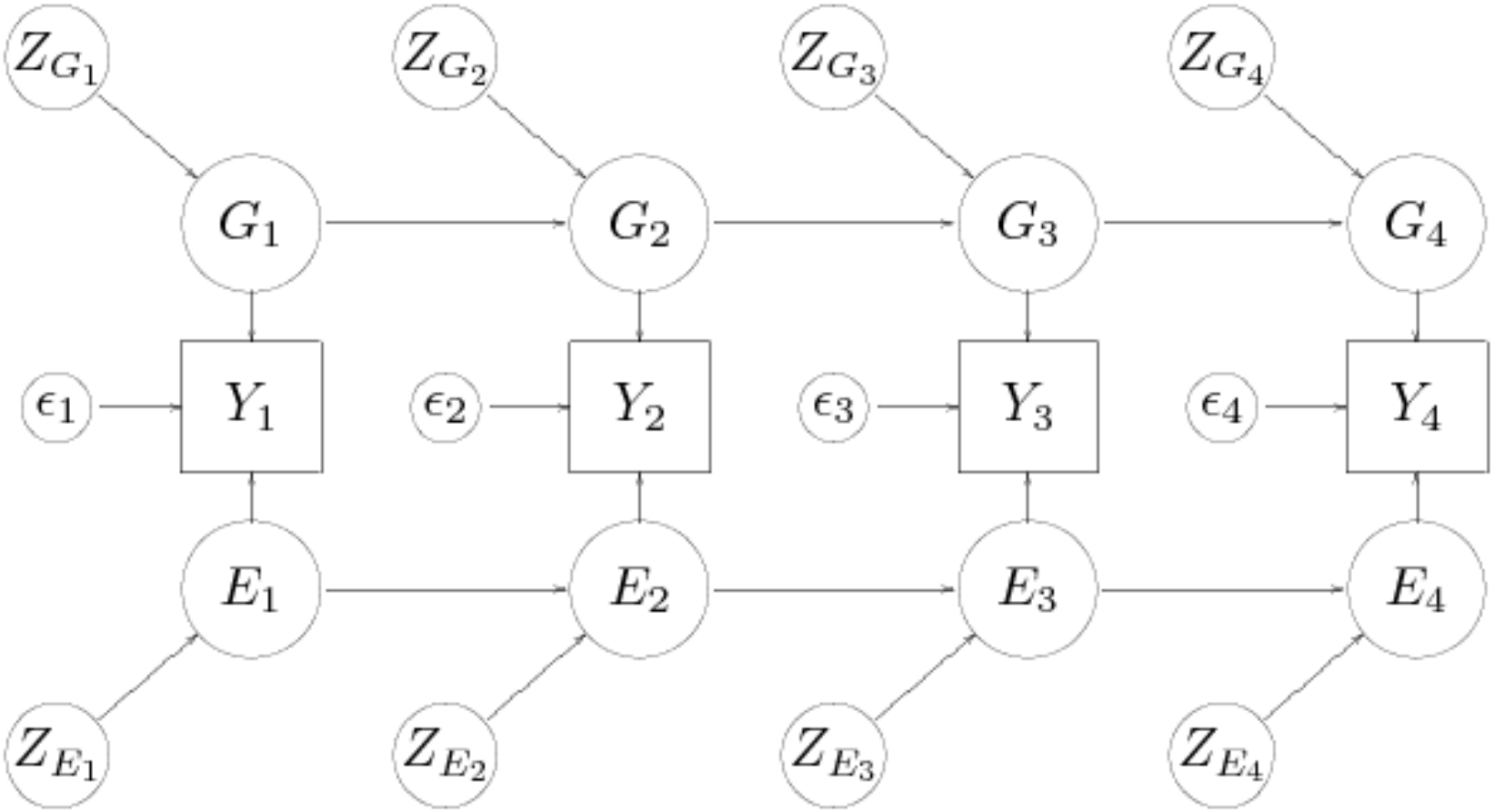


Methods for Longitudinal Analysis

- Cholesky Models
- Simplex Models
- Growth Curve Models

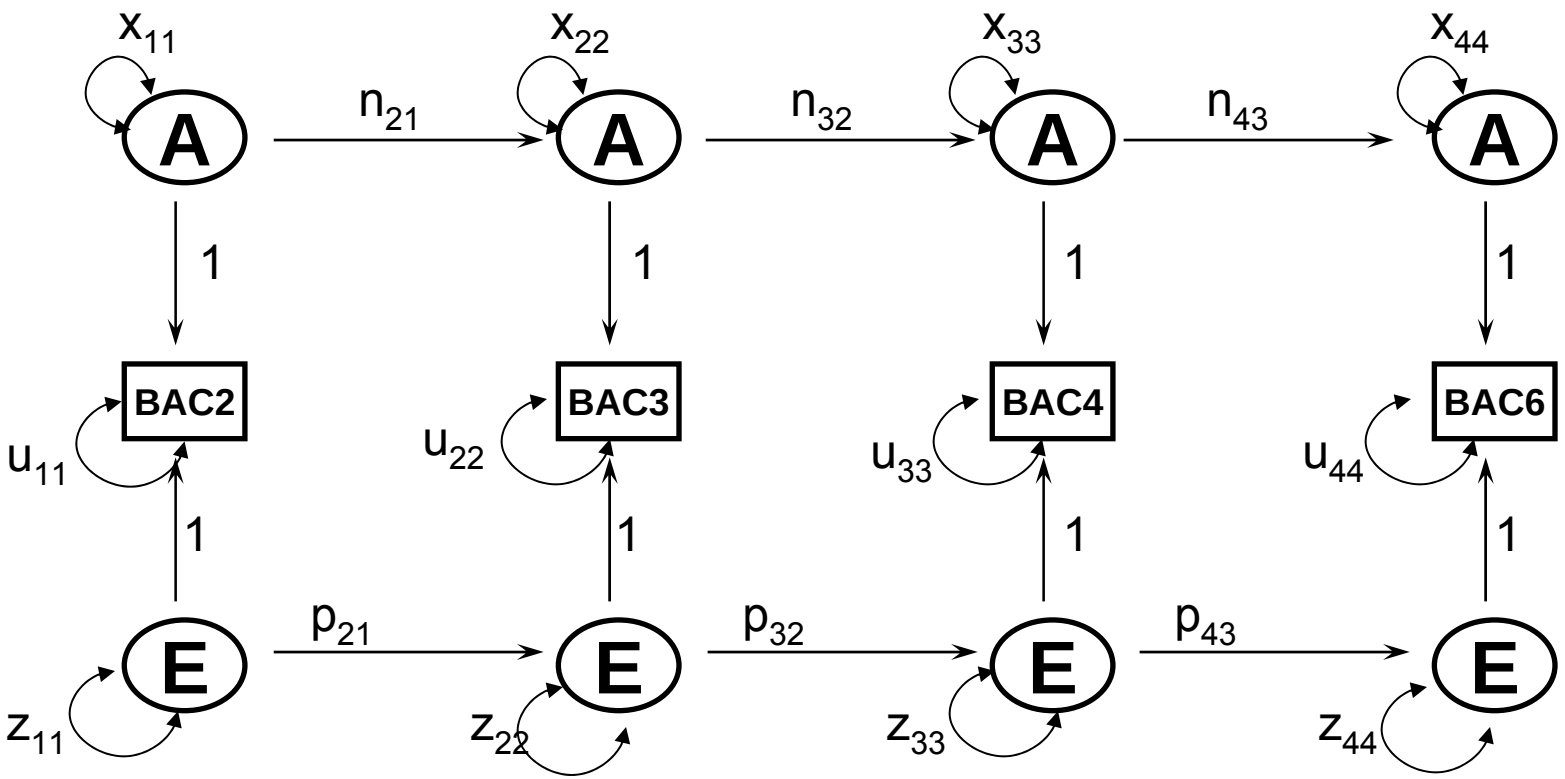
Simplex Model

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

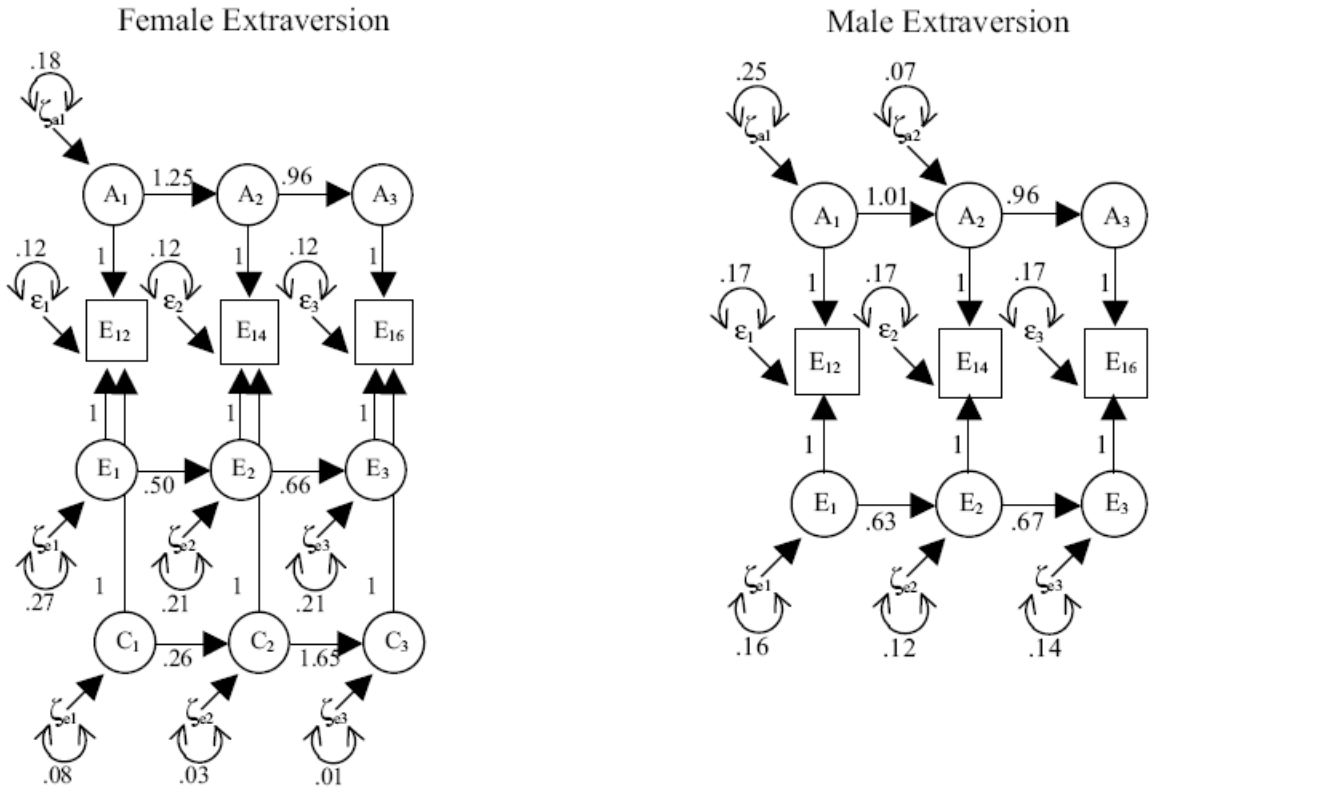


Figure 2

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

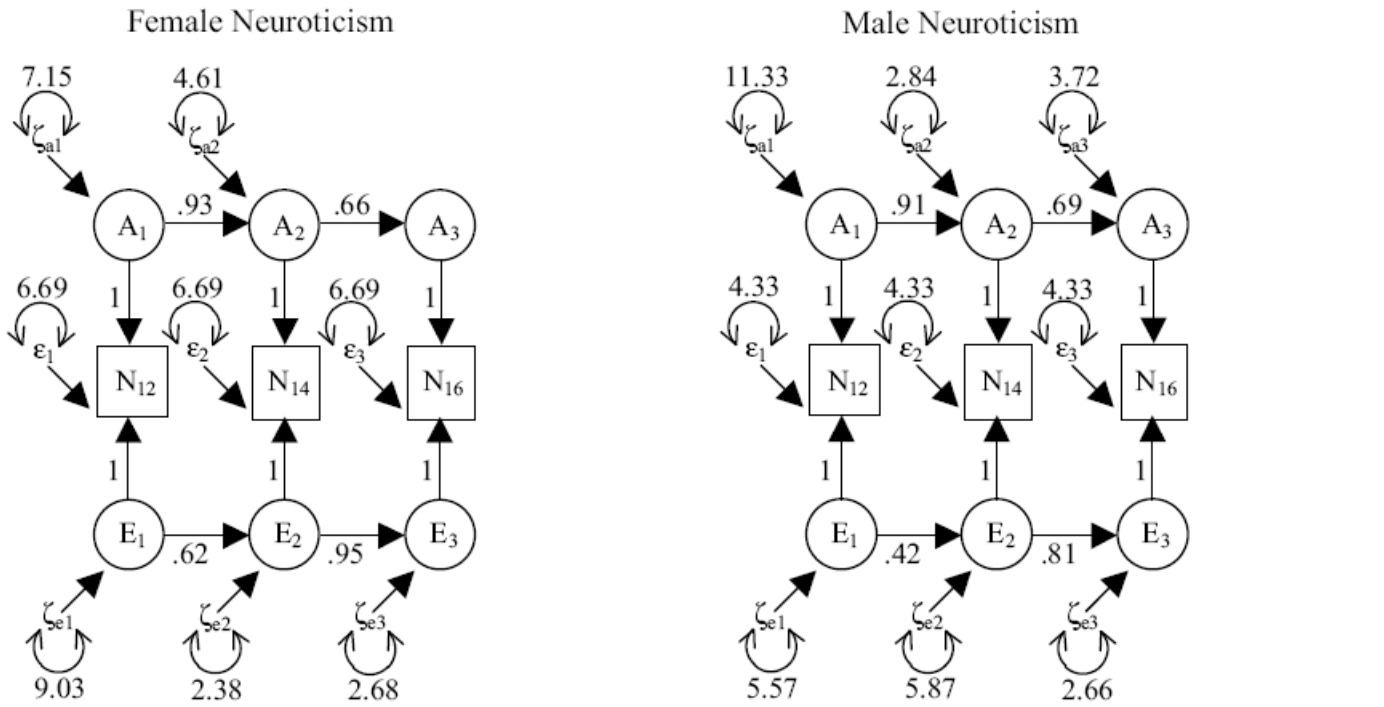


Figure 3

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

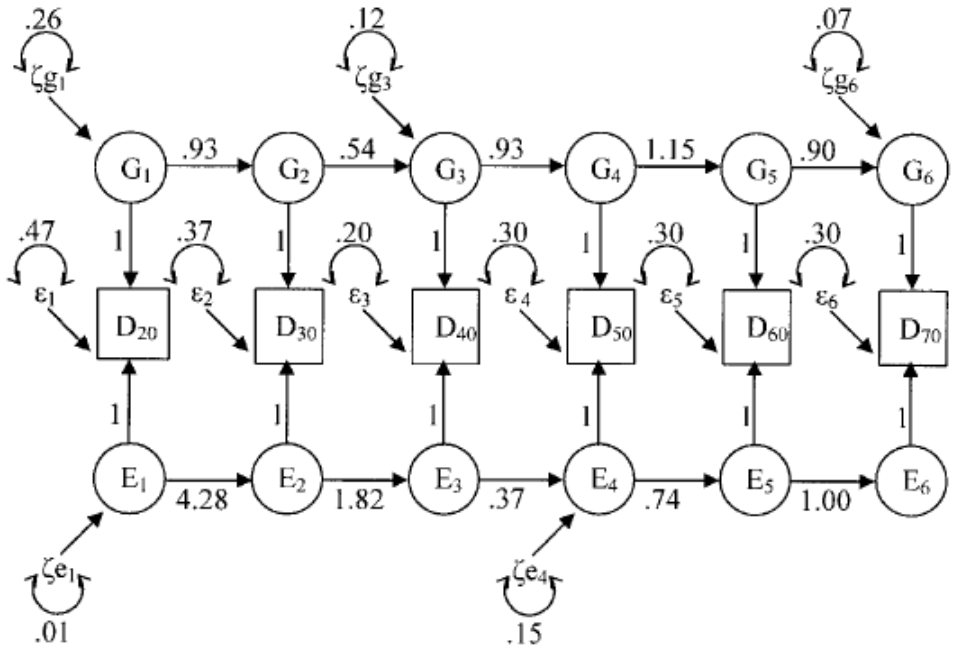


Figure 3

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, ζg = additive genetic innovations, ζe = 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:

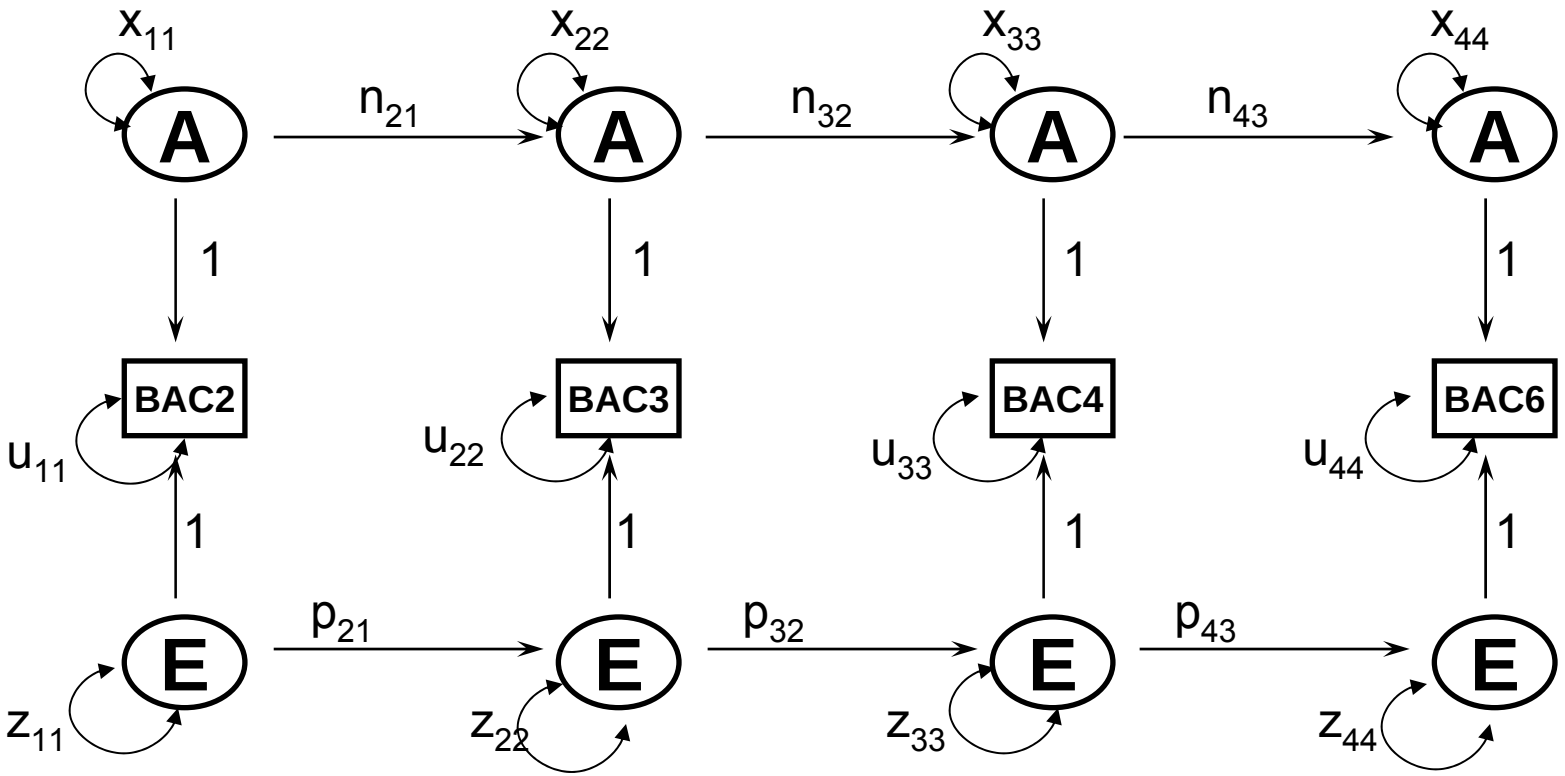
	Minutes post-consump.	Mean BAC	# of Individuals with data	
			MZM (43 prs)	DZM (37 prs)
BAC 1	56	89.0	83	72
BAC 2	**68**	**88.9**	**83**	**74**
BAC 3	**83**	**88.8**	**84**	**71**
BAC 4	**123**	**80.9**	**86**	**74**
BAC 5	140	70.0	52	52

A simplex correlation pattern…

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

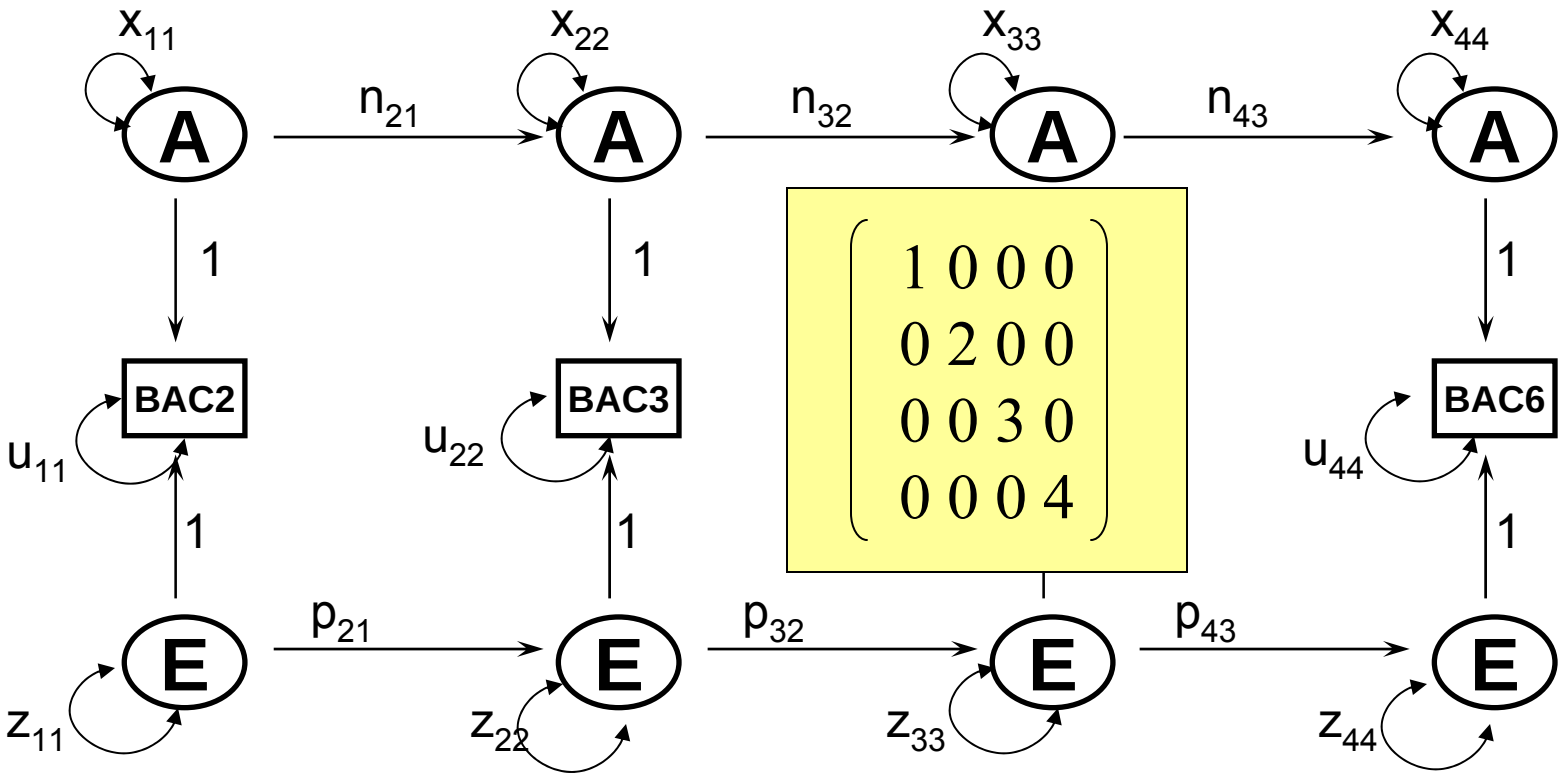
	BAC 2	BAC 3	BAC 4	BAC 6
BAC 2	1.00			
BAC 3	0.90	1.00		
BAC 4	0.69	0.84	1.00	

Practical - Simplex Model



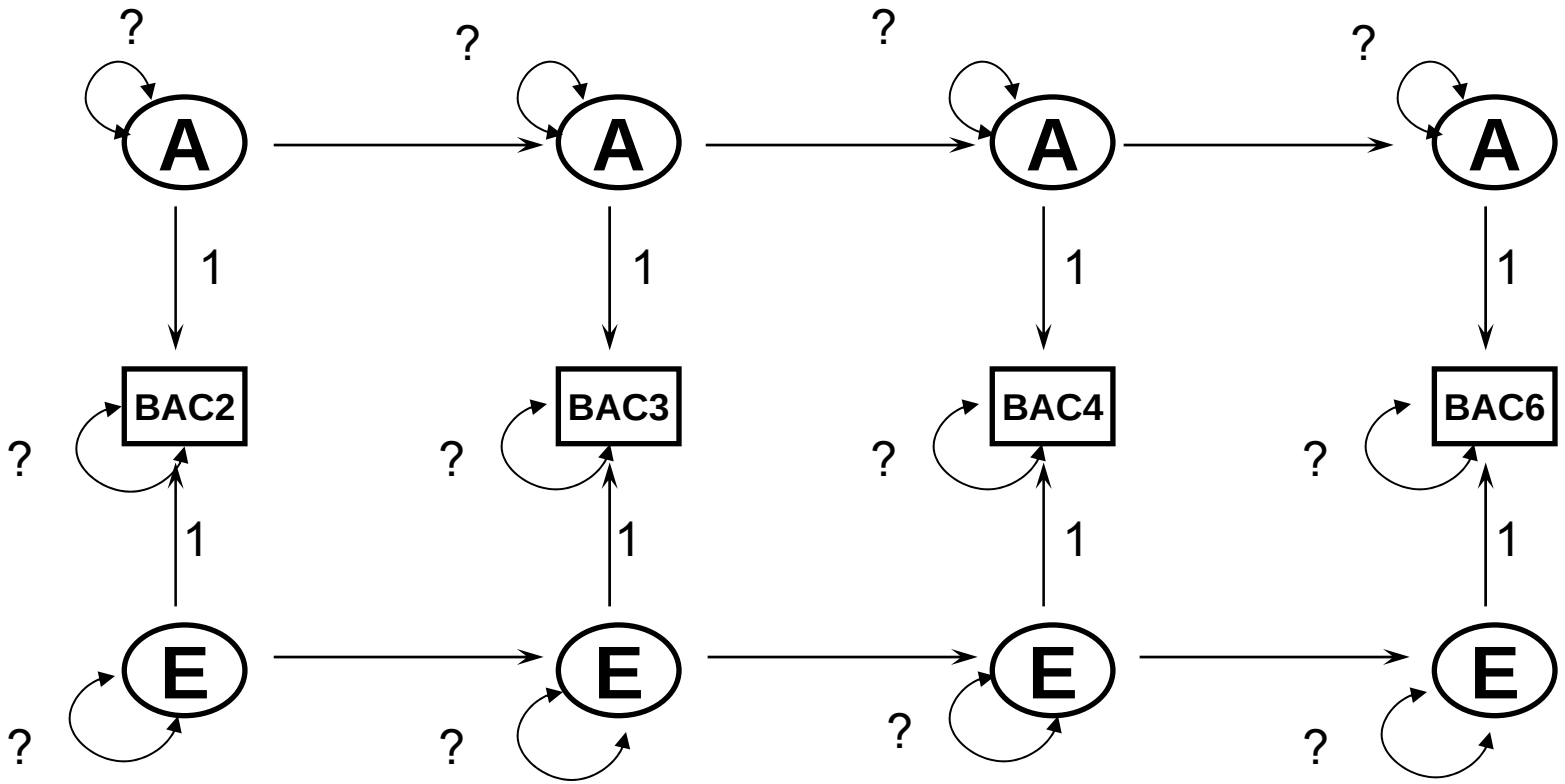
x and z = genetic and nonshared environmental **innovations** respectively
n and p = genetic and nonshared environmental **transmission** respectively
u = **error** variances

Practical - Simplex Model



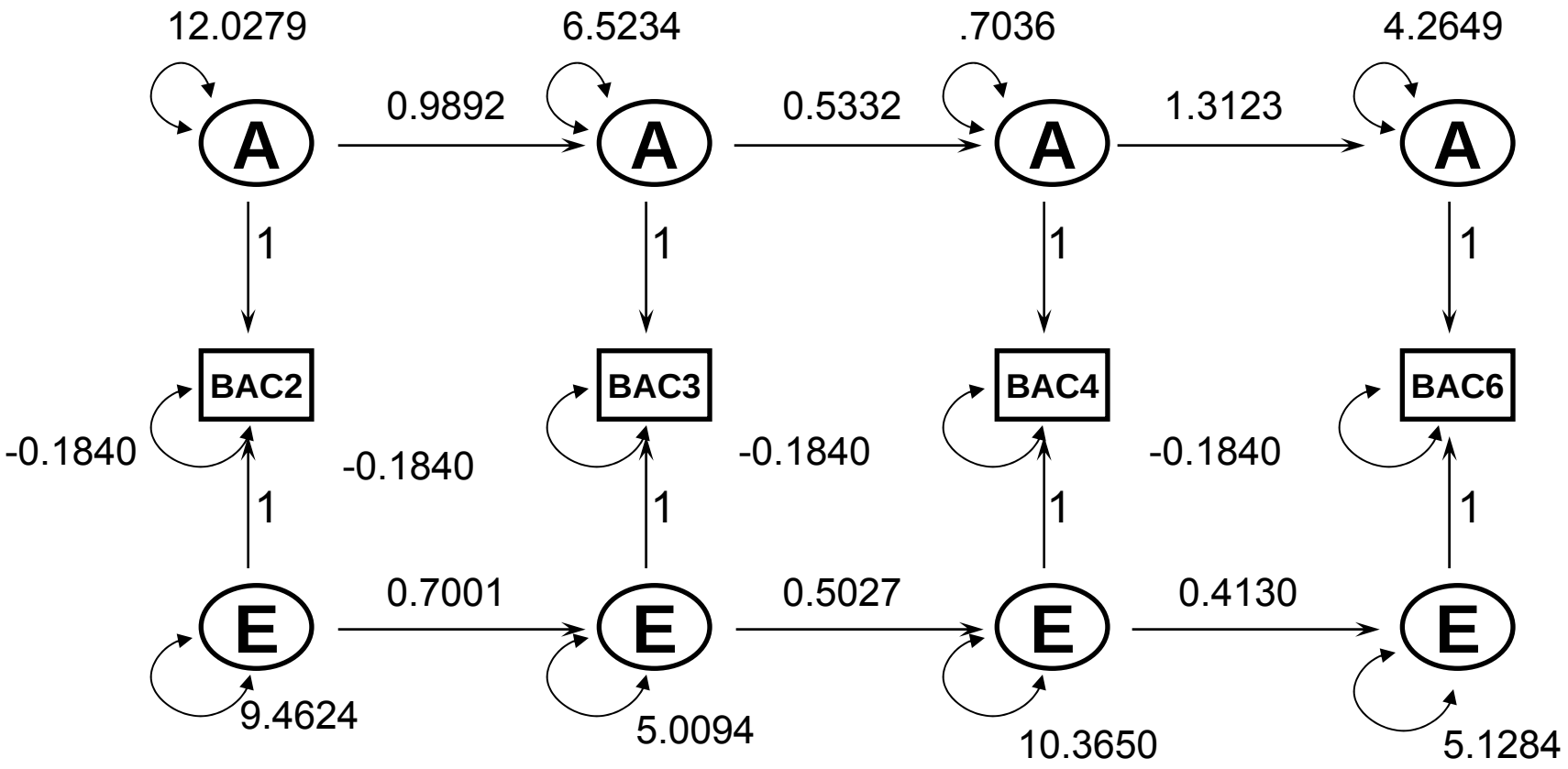
x and z = genetic and nonshared environmental **innovations** respectively
n and p = genetic and nonshared environmental **transmission** respectively
u = **error** variances

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

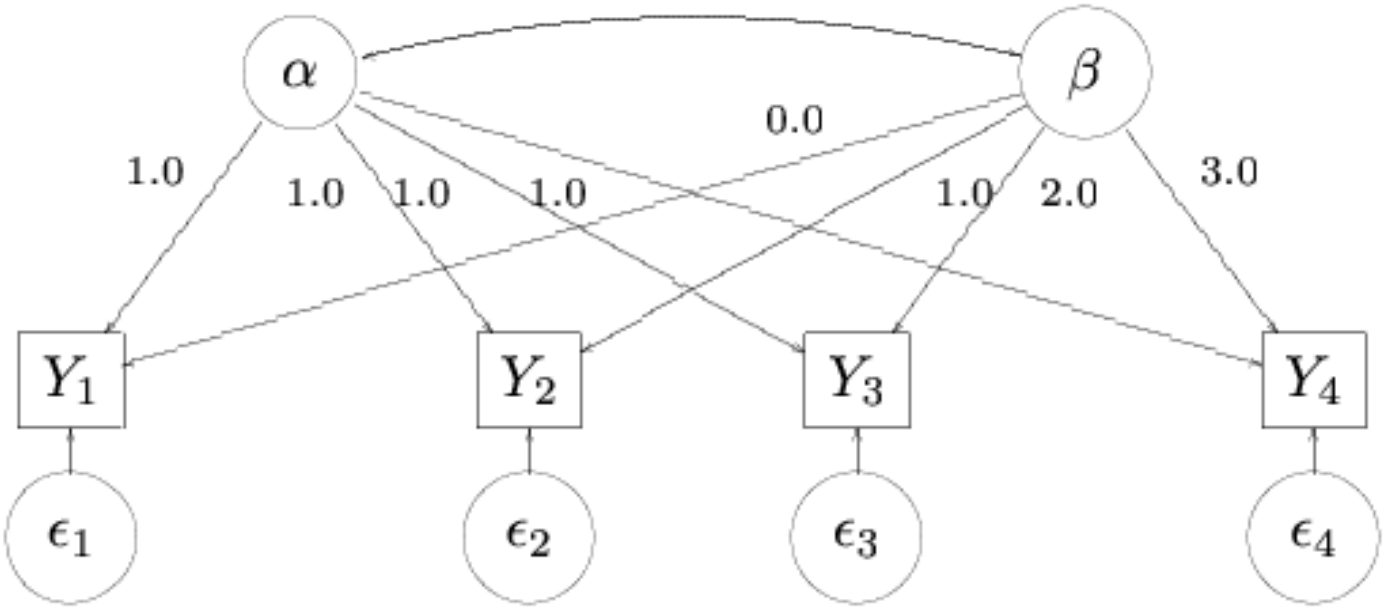
- Advantages:
  - Makes restrictive predictions about covariance pattern
  - Falsifiable
- Disadvantages:
  - Makes restrictive predictions about covariance pattern (future depends on current state only)
  - Number of parameters increases with number of measurements

Methods for Longitudinal Analysis

- Cholesky Models
- Simplex Models
- Growth Curve Models

Latent Growth Curve Model

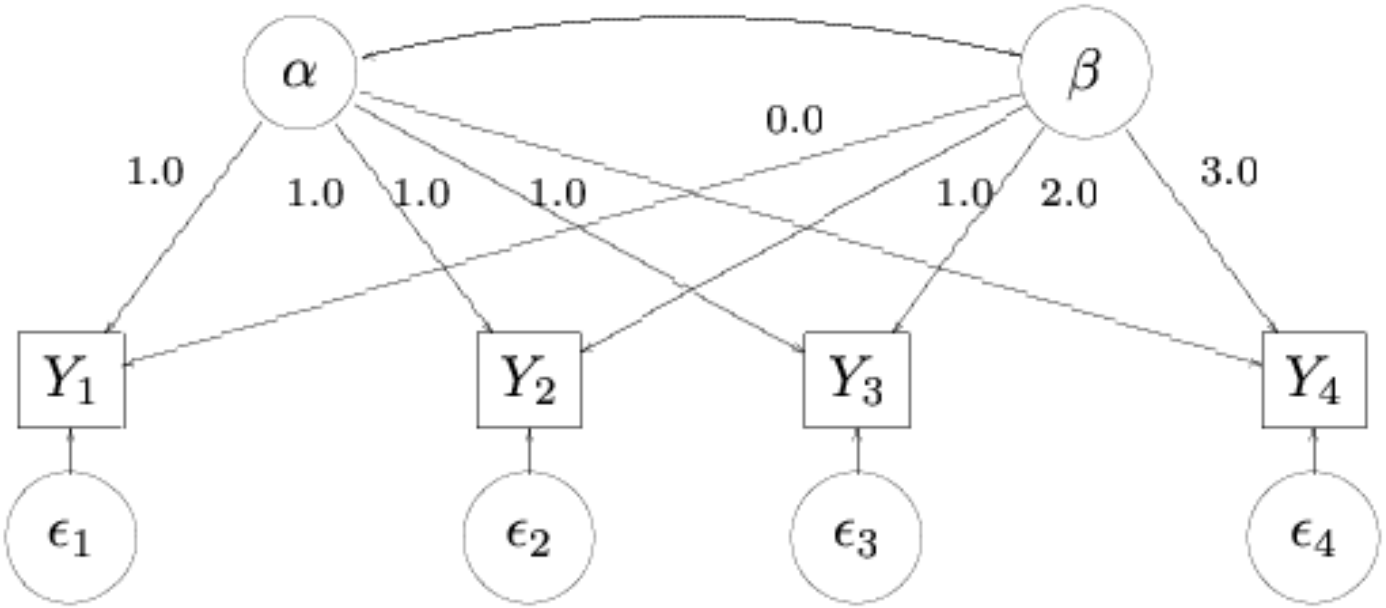
(shown here as linear)



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

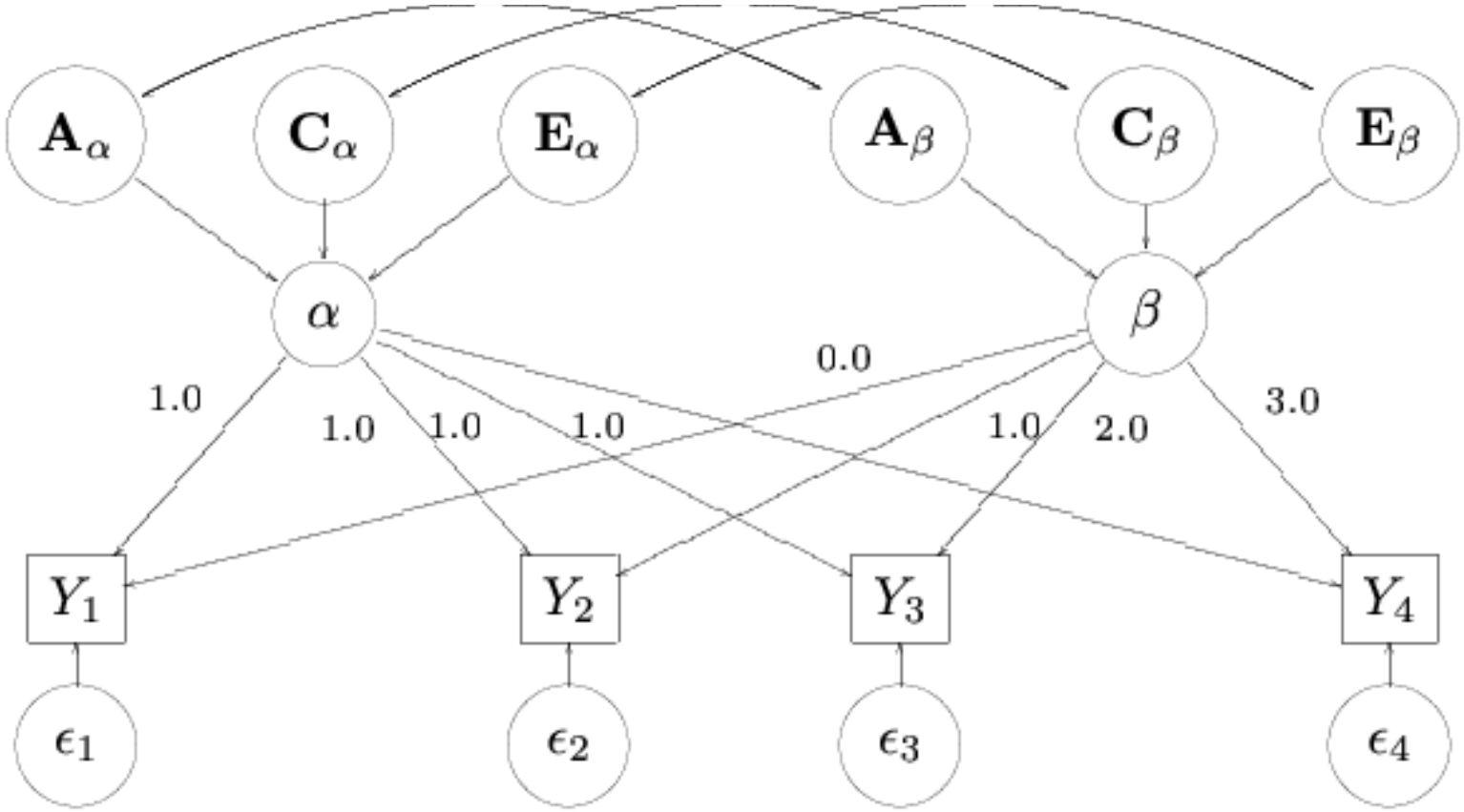
Latent Growth Curve Model

(shown here as linear)

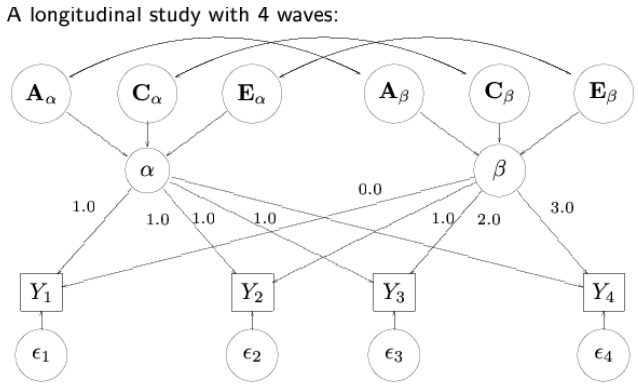


$$Y_{i,t} = \alpha_i + \beta_i(t-1) + \epsilon_{i,t}, \quad i = 1, 2 \quad t = 1, 2, 3, 4$$

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

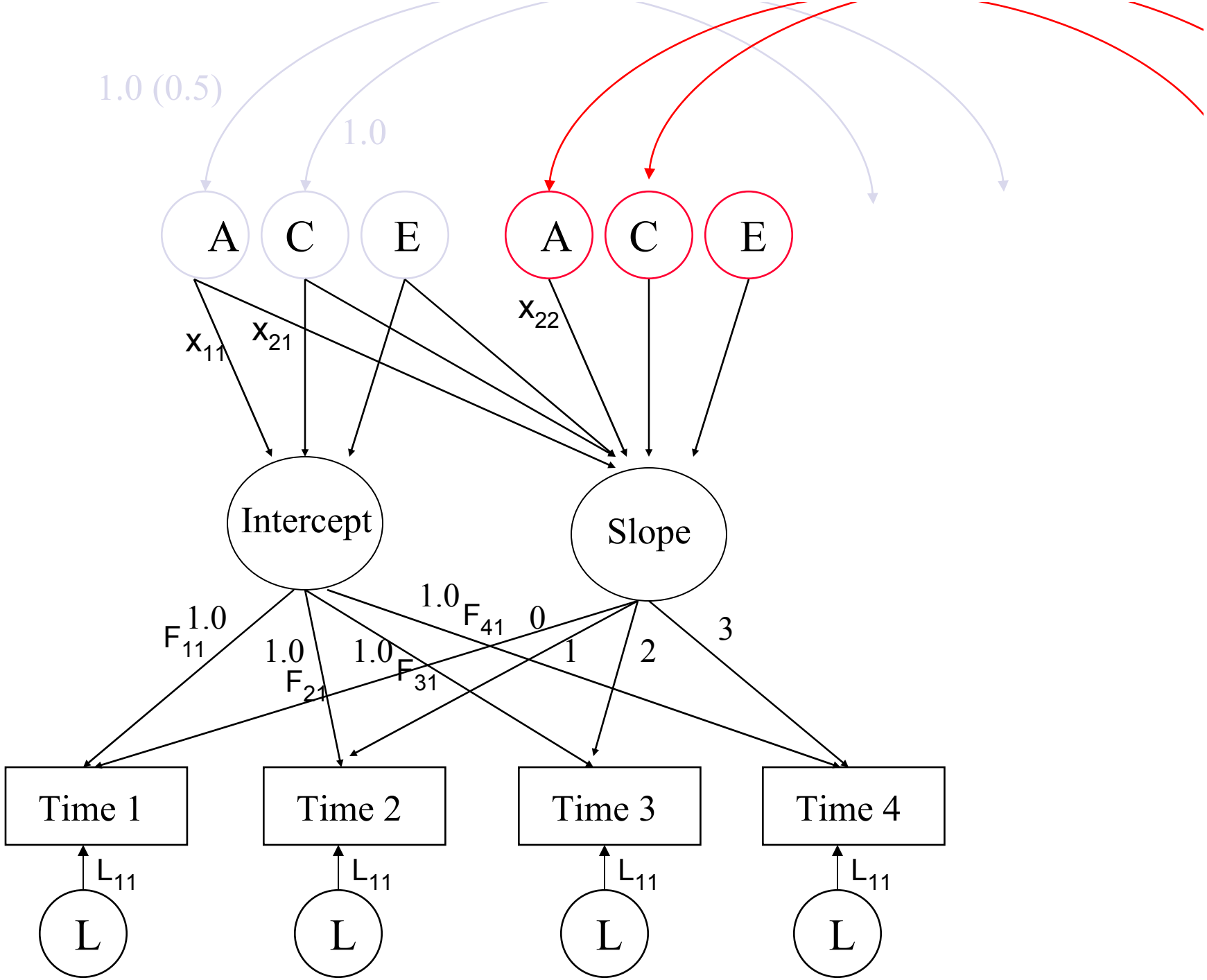
- What is the contribution of genetic/environmental factors to the variation of α (intercept) and β (slope)?
- Same or different genes influencing α (intercept) and β (slope)?
- Same or different environments influencing α (intercept) and β (slope)?

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

Growth Curve Model

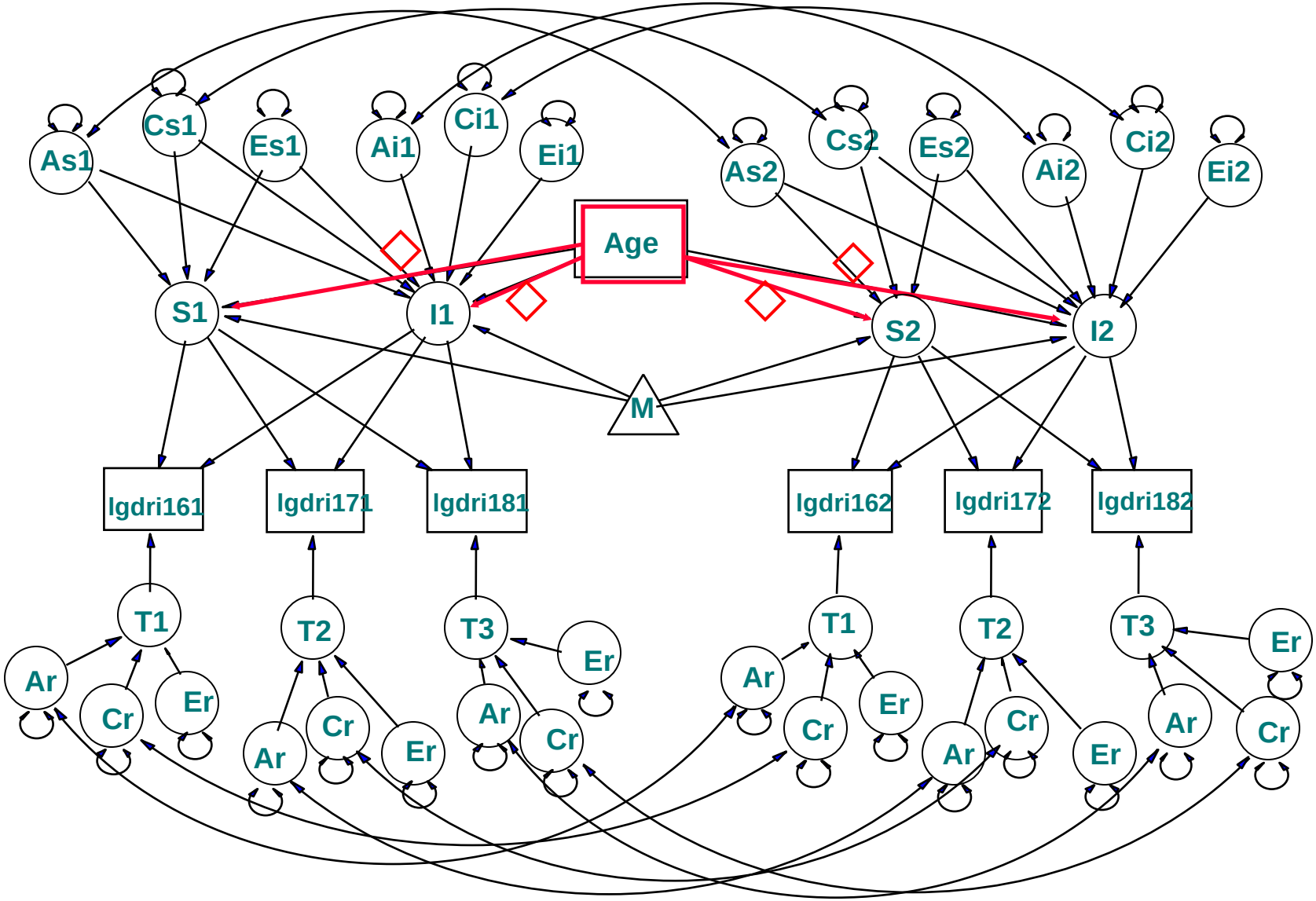
- Advantages:
  - Very efficient: number of parameters does not increase with number of measurements
  - Provides prediction about behavior beyond measured timepoints
- Disadvantages:
  - Note regarding slope parameters
  - Can be computationally intense
  - Assumptions to reduce computational burden
    - 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)

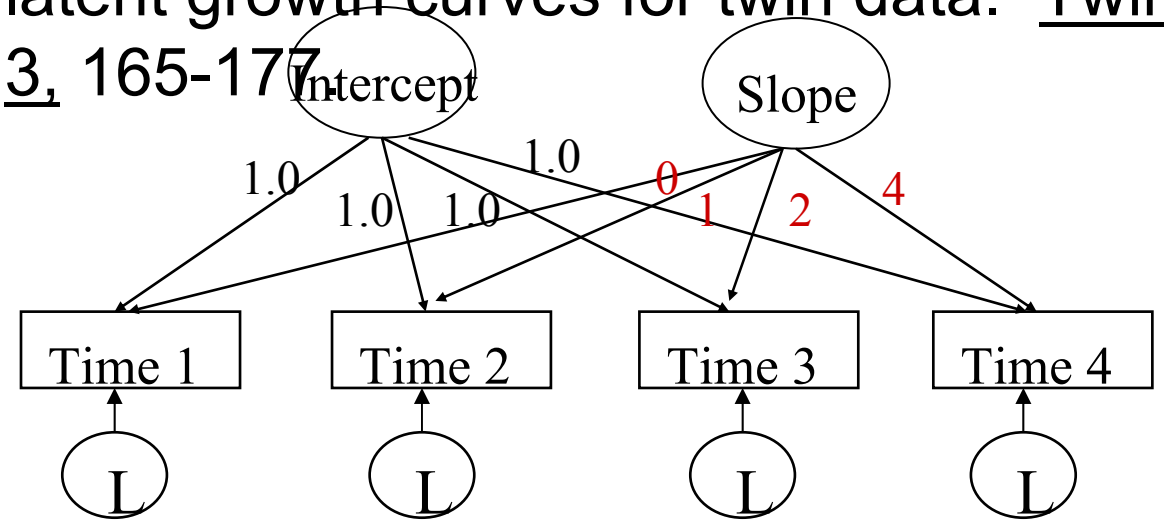
- Age heterogeneity and unequal spacing of measurements can be handled using definition variables
  - Mehta & West, 2000, Psychological Methods

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