

# Standard genetic simplex models in the classical twin design with phenotype to E transmission

Conor Dolan & Janneke de Kort  
Biological Psychology, VU

## Two general approaches to longitudinal modeling

Markov models:

(Vector) autoregressive models for continuous data

(Hidden) Markov transition models discrete data

(Mixtures thereof)

Growth curve models (Brad Verhulst, Lindon Eaves):

Focus on linear and non-linear growth curves

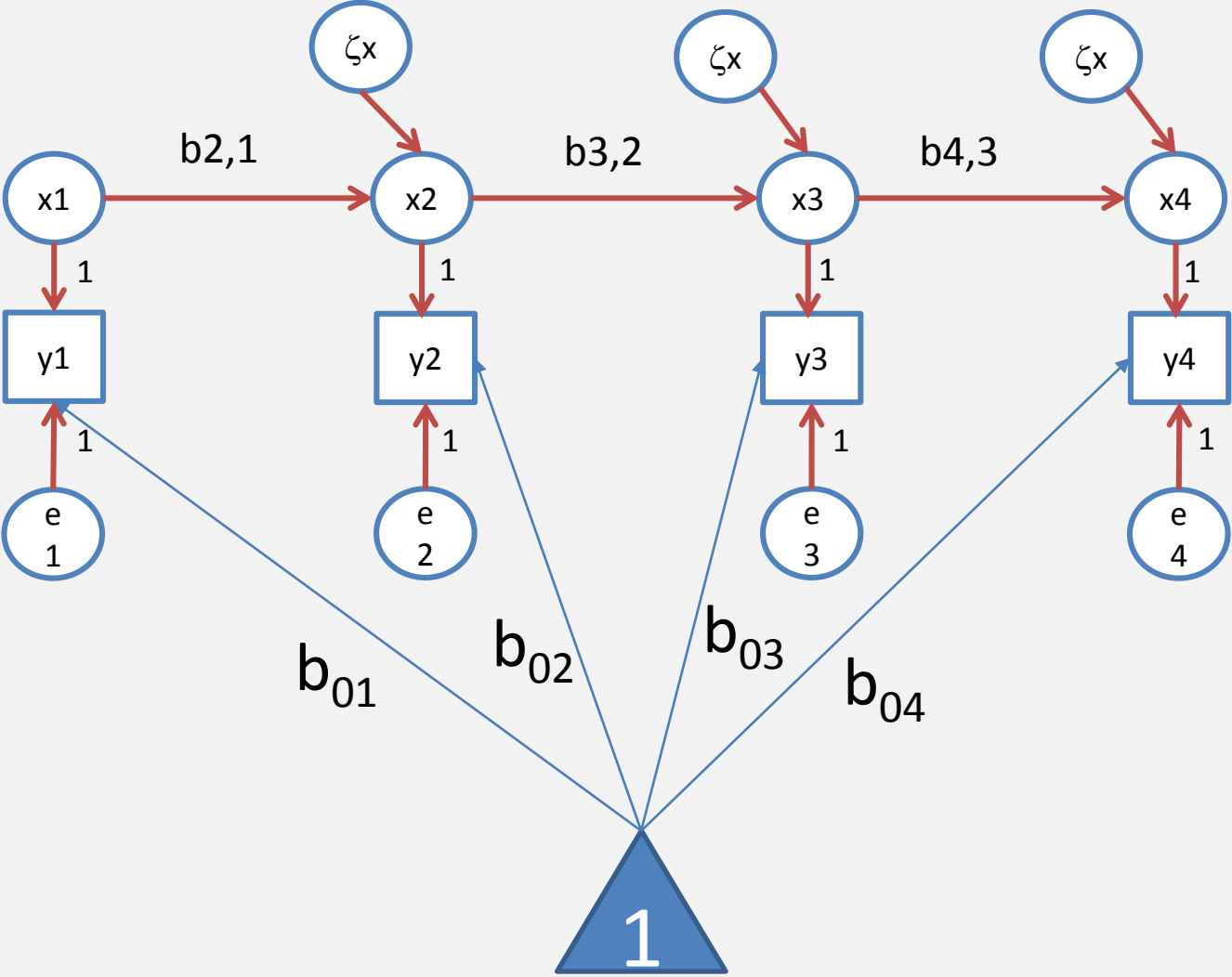
Typically multilevel or random effects model

(Mixtures thereof)

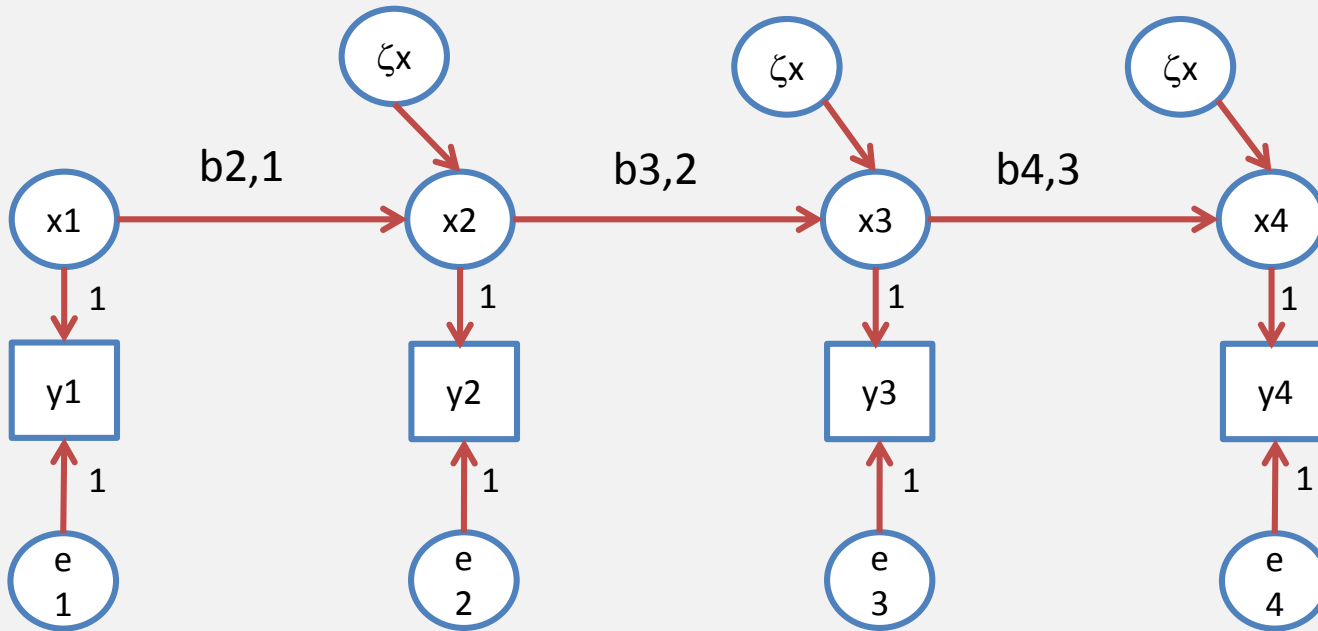
Which to use?

Use the model that fit the theory / data / hypotheses

First order autoregression model. A (quasi) simplex model ( $\text{var}(e) > 0$ ).



First order autoregression model. A quasi simplex model ( $\text{var}(e) > 0$ ).



$$y_{ti} = b_{0t} + x_{ti} + e_{ti}$$

$$x_{1i} = x_{1i} = \zeta x_{1i}$$

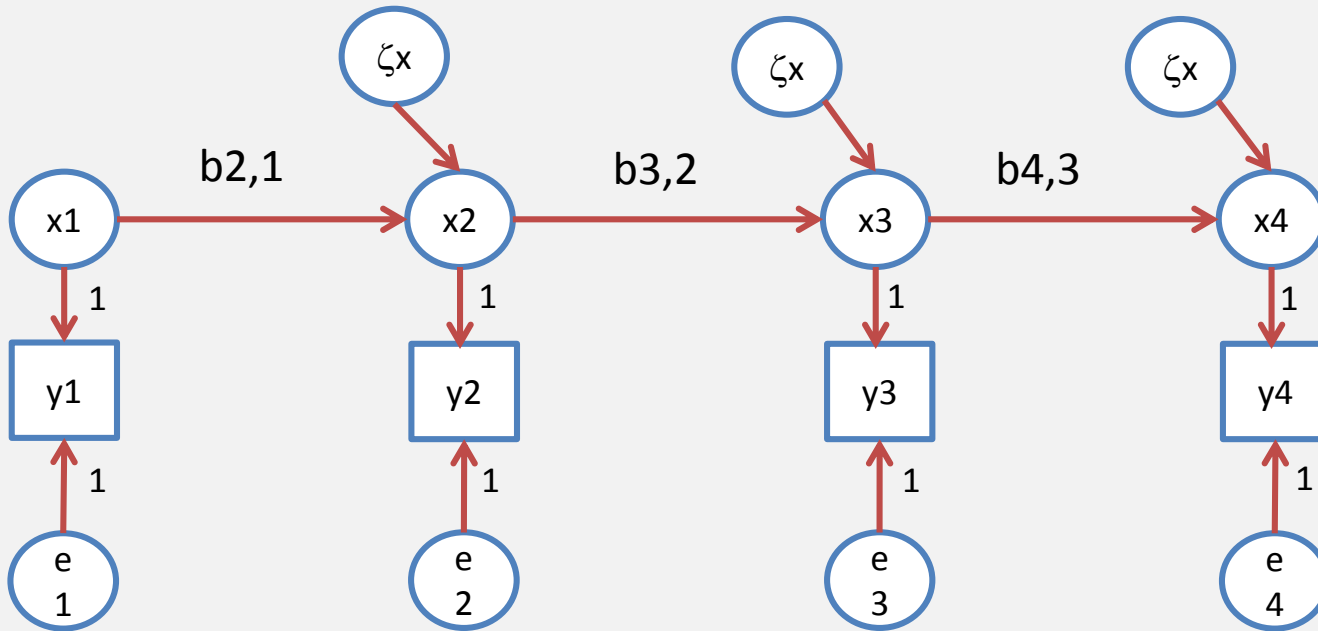
$$x_{ti} = b_{t-1,t} x_{t-1i} + \zeta x_{ti}$$

$$\text{var}(y_1) = \text{var}(x_1) + \text{var}(e_1)$$

$$\text{var}(x_t) = b_{t-1,t}^2 \text{var}(x_{t-1}) + \text{var}(\zeta x_t)$$

$$\text{cov}(x_t, x_{t-1}) = b_{t-1,t} \text{var}(x_{t-1})$$

First order autoregression model. A quasi simplex model ( $\text{var}(e) > 0$ ).



Identification issue:  $\text{var}(e_1)$  and  $\text{var}(e_t)$  are not identified. Solution set to zero, or equate  $\text{var}(e_1) = \text{var}(e_2)$ ,  $\text{var}(e_3) = \text{var}(e_4)$

$$\text{var}(y_1) = \text{var}(x_1) + \text{var}(e_1)$$

$$\text{var}(x_t) = b_{t-1,t}^2 \text{var}(x_{t-1}) + \text{var}(\zeta_t)$$

$$\text{cov}(x_t, x_{t-1}) = b_{t-1,t} \text{var}(x_{t-1})$$

Standardized stats I:

**Reliability at each t, rel(t) :**

$$\text{rel}(x_1) = \text{var}(x_1) / \{\text{var}(x_1) + \text{var}(e_1)\}$$

**Interpretation:**

% of variance in y at t due to latent x at t

$$\text{var}(y_1) = \text{var}(x_1) + \text{var}(e_1)$$

$$\text{var}(x_t) = b_{t-1,t}^2 \text{var}(x_{t-1}) + \text{var}(\zeta x_t)$$

$$\text{cov}(x_t, x_{t-1}) = b_{t-1,t} \text{var}(x_{t-1})$$

Standardized stats II:

**Stability at each t,t-1, stab(t,t-1):**

$$b_{t-1,t}^2 \text{var}(x_{t-1}) / \{b_{t-1,t}^2 \text{var}(x_{t-1}) + \text{var}(\zeta x_t)\}$$

**Interpretation:**

% of the variance in x at t explained by regression on x at t-1 (latent level!)

$$\text{var}(y_1) = \text{var}(x_1) + \text{var}(e_1)$$

$$\text{var}(x_t) = b_{t-1,t}^2 \text{var}(x_t) + \text{var}(\zeta x_t)$$

$$\text{cov}(x_t, x_{t-1}) = b_{t-1,t} \text{var}(x_t)$$

Standardized stats III:

**Correlation t,t-1, cor(t,t-1):**

$$b_{t-1,t}^2 \text{var}(x_t) / \{\text{sd}(y_{t-1}) * \text{sd}(y_t)\}$$

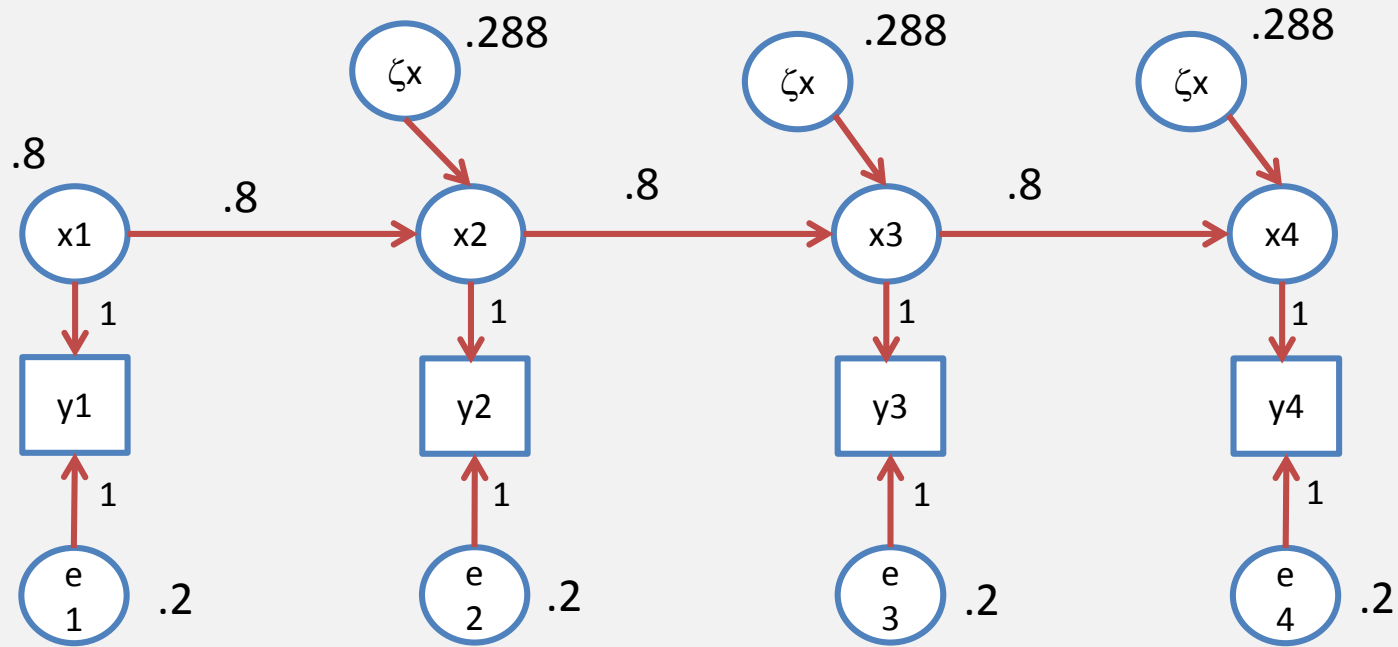
$$\text{sd}(y_t) = \text{sqrt}(\text{var}(x_t) + \text{var}(e_t))$$

$$\text{var}(x_t) = b_{t-1,t}^2 \text{var}(x_t) + \text{var}(\zeta x_t)$$

**Interpretation:**

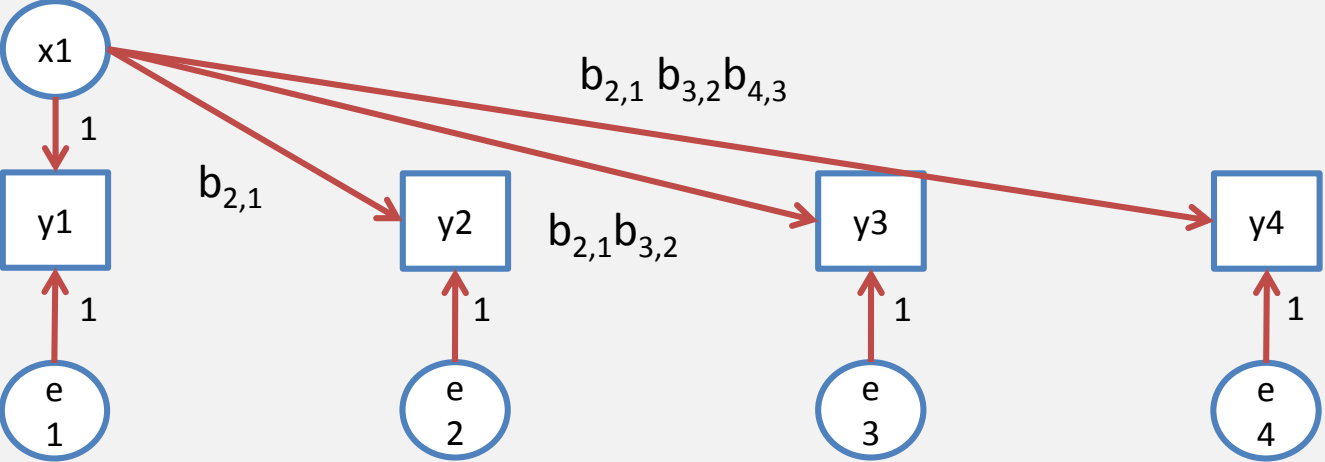
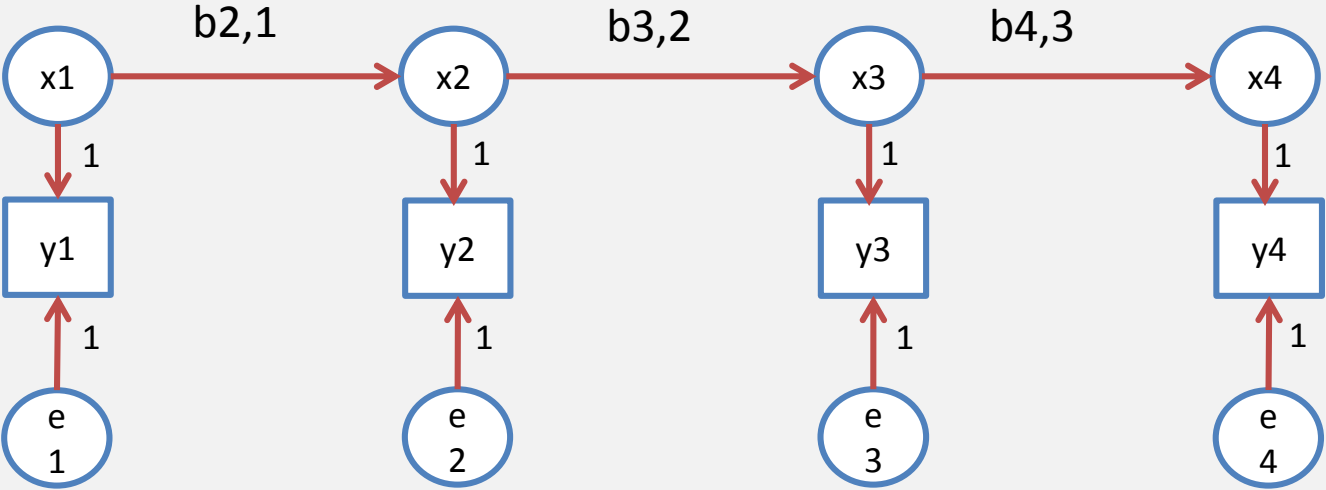
strength of linear relationship





	1.0000			
	0.6400	1.0000		
	0.5120	0.6400	1.0000	
	0.4096	0.5120	0.6400	1.0000

Special case: factor model  $\text{var}(\zeta x_t) \text{ (} t=2,3,4 \text{)} = 0$



1.0000			
0.6400	1.000		
0.5120	0.640	1.000	
0.4096	0.512	0.640	1.0000

$$\text{var}(\zeta x_t) \text{ (t=2,3,4) } = 0$$

1.000			
0.758	1.000		
0.705	0.668	1.000	
0.640	0.607	0.564	1.000



Multivariate decomposition of phenotypic covariance matrix:

$$\Sigma_{\text{ph}} = \Sigma_A + \Sigma_C + \Sigma_E$$

$$\begin{matrix} \Sigma_{\text{ph1}} \\ \Sigma_{\text{ph12}} \end{matrix} \quad \begin{matrix} \Sigma_{\text{ph2}} \\ \end{matrix} =$$

$$\begin{matrix} \Sigma_A + \Sigma_C + \Sigma_E \\ r\Sigma_A + \Sigma_C + \Sigma_E \end{matrix} \quad \Sigma_A + \Sigma_C + \Sigma_E$$

$$\Sigma_{\text{ph}} = \Sigma_A + \Sigma_C + \Sigma_E$$

Estimate  $\Sigma_A$  using a Cholesky-decomp

$$\Sigma_A = \Delta_A \Delta_A^t$$

$$\Delta_A = \begin{matrix} \delta_{11} & 0 & 0 & 0 \\ \delta_{21} & \delta_{22} & 0 & 0 \\ \delta_{31} & \delta_{32} & \delta_{33} & 0 \\ \delta_{41} & \delta_{42} & \delta_{43} & \delta_{44} \end{matrix}$$

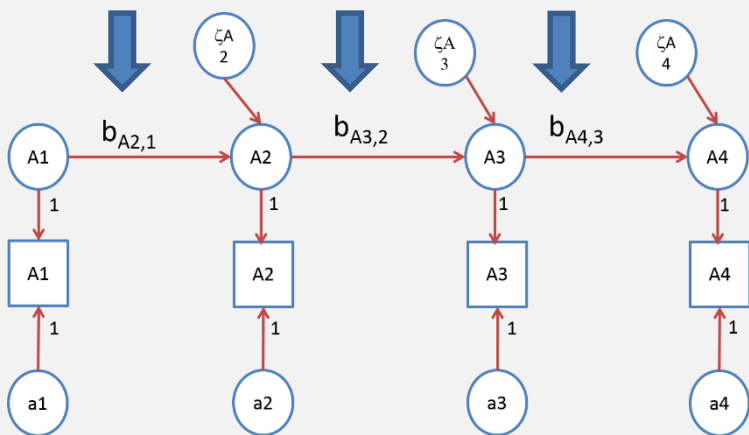
$$\Sigma_{\text{ph}} = \Sigma_A + \Sigma_C + \Sigma_E$$

Model  $\Sigma_A$  a simplex model

$$\Sigma_A = (\mathbf{I}-\mathbf{B})_A \Psi_A (\mathbf{I}-\mathbf{B})_A^t + \Theta_A$$

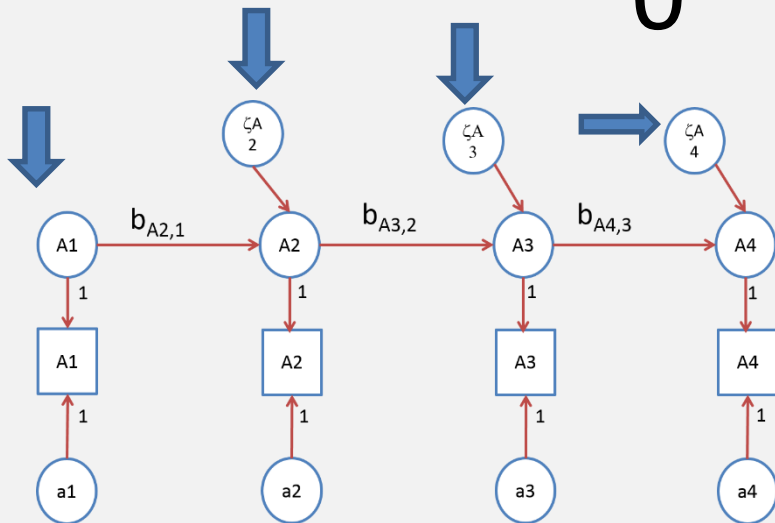
$$\Sigma_A = (I - B_A) \Psi_A (I - B_A)^t + \Theta_A$$

$$B_A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ b_{A21} & 0 & 0 & 0 \\ 0 & b_{A32} & 0 & 0 \\ 0 & 0 & b_{A43} & 0 \end{bmatrix}$$



$$\Sigma_A = (I - B_A) \Psi_A (I - B_A)^t + \Theta_A$$

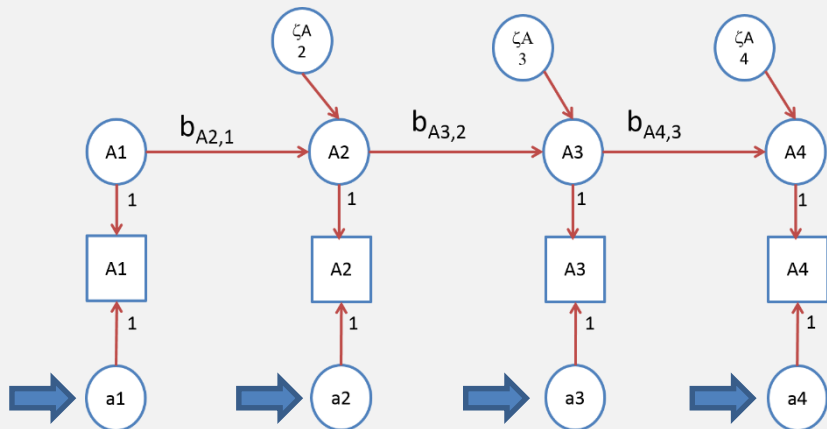
$$\Psi_A = \begin{bmatrix} \text{var}(A_1) & 0 & 0 & 0 \\ 0 & \text{var}(\zeta_{A2}) & 0 & 0 \\ 0 & 0 & \text{var}(\zeta_{A3}) & 0 \\ 0 & 0 & 0 & \text{var}(\zeta_{A4}) \end{bmatrix}$$

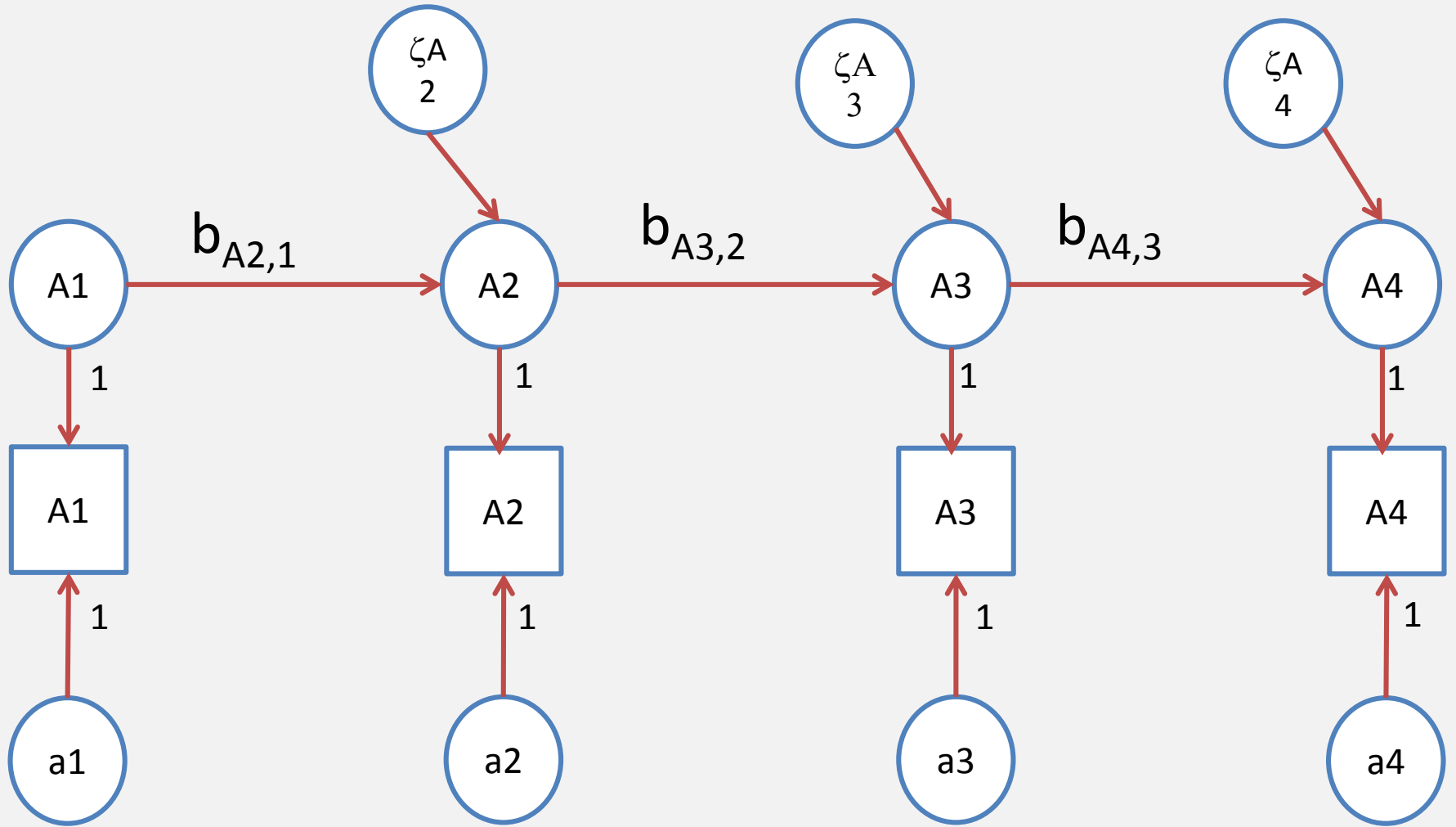




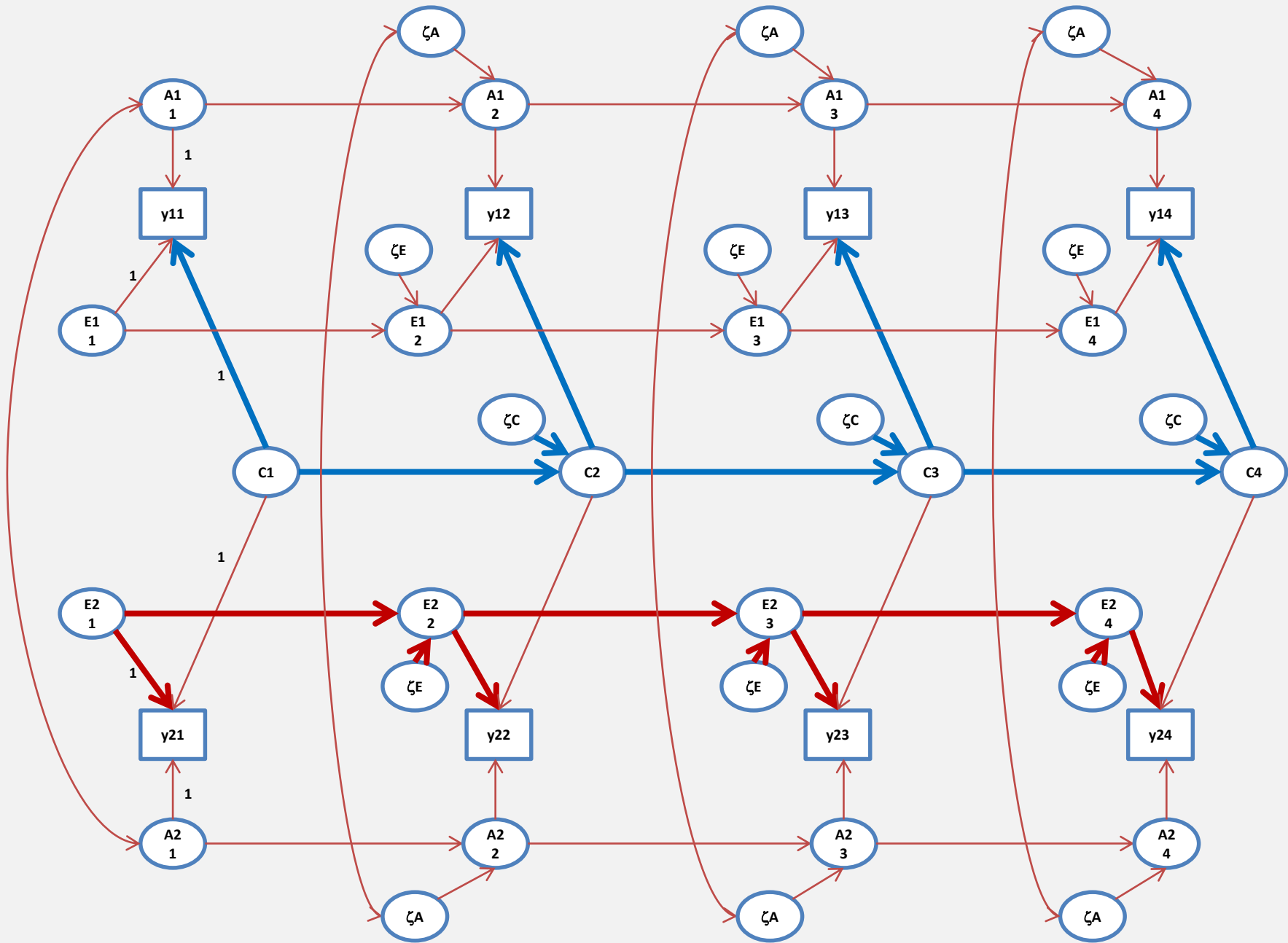
$$\Sigma_A = (I - B_A) \Psi_A (I - B_A)^t + \Theta_A$$

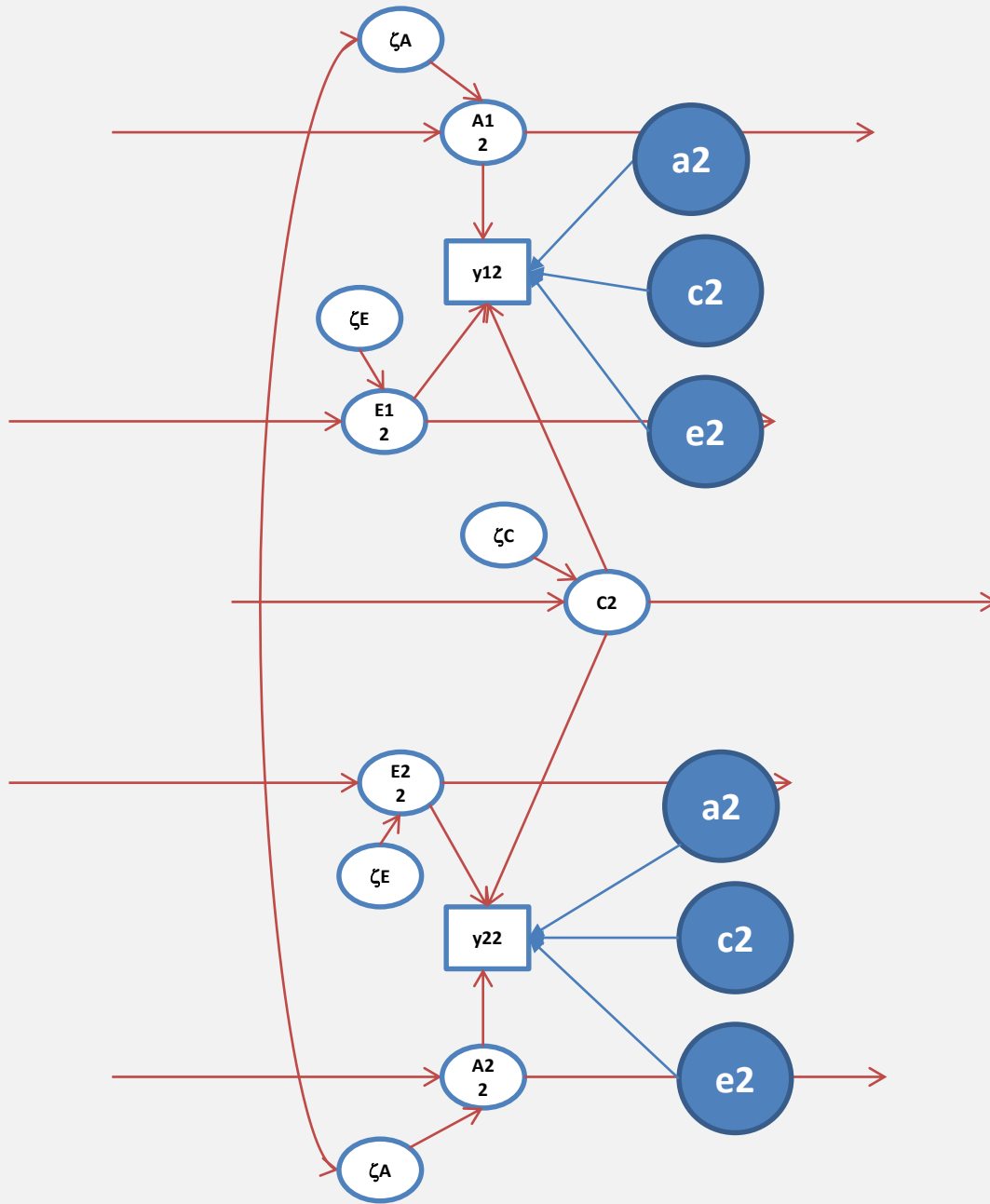
$$\Theta_A = \begin{matrix} \text{var}(a_1) & 0 & 0 & 0 \\ 0 & \text{var}(a_2) & 0 & 0 \\ 0 & 0 & \text{var}(a_3) & 0 \\ 0 & 0 & 0 & \text{var}(a_4) \end{matrix}$$





The genetic simplex (note my scaling)





Occasion  
specific  
effects

Question:  $h^2$ ,  $c^2$ , and  $e^2$  at each time point?

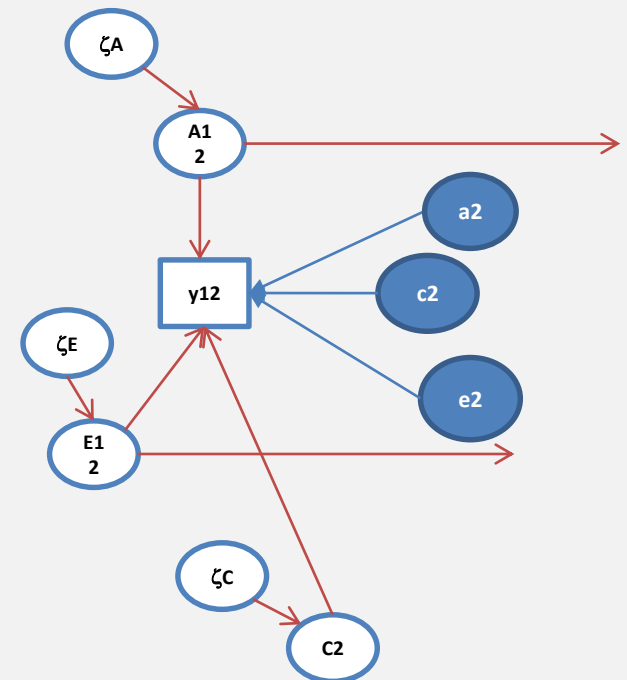
$\text{var}(y_t) =$

$$\{\text{var}(A_t) + \text{var}(a_t)\} + \{\text{var}(C_t) + \text{var}(c_t)\} + \{\text{var}(E_t) + \text{var}(e_t)\}$$

$$h^2 = \{\text{var}(A_t) + \text{var}(a_t)\} / \text{var}(y_t)$$

$$c^2 = \{\text{var}(C_t) + \text{var}(c_t)\} / \text{var}(y_t)$$

$$e^2 = \{\text{var}(E_t) + \text{var}(e_t)\} / \text{var}(y_t)$$

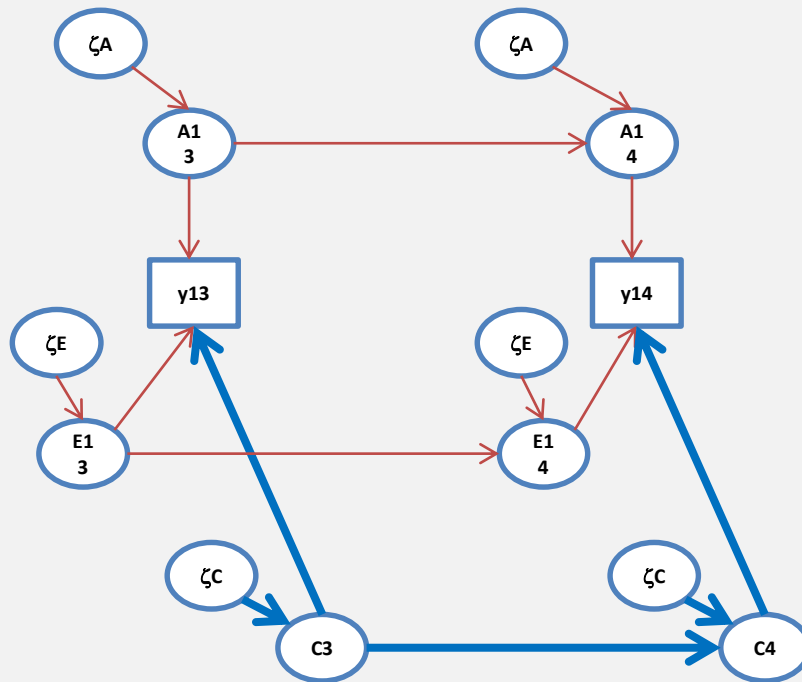


# Question: contributions to stability t-1 to t

$$b_{A_{t-1,t}}^2 \text{var}(A_t) / \{b_{A_{t-1,t}}^2 \text{var}(A_t) + \text{var}(\zeta A_t) + \text{var}(a_t)\}$$

$$b_{C_{t-1,t}}^2 \text{var}(C_t) / \{b_{C_{t-1,t}}^2 \text{var}(C_t) + \text{var}(\zeta C_t) + \text{var}(c_t)\}$$

$$b_{E_{t-1,t}}^2 \text{var}(E_t) / \{b_{E_{t-1,t}}^2 \text{var}(E_t) + \text{var}(\zeta E_t) + \text{var}(e_t)\}$$



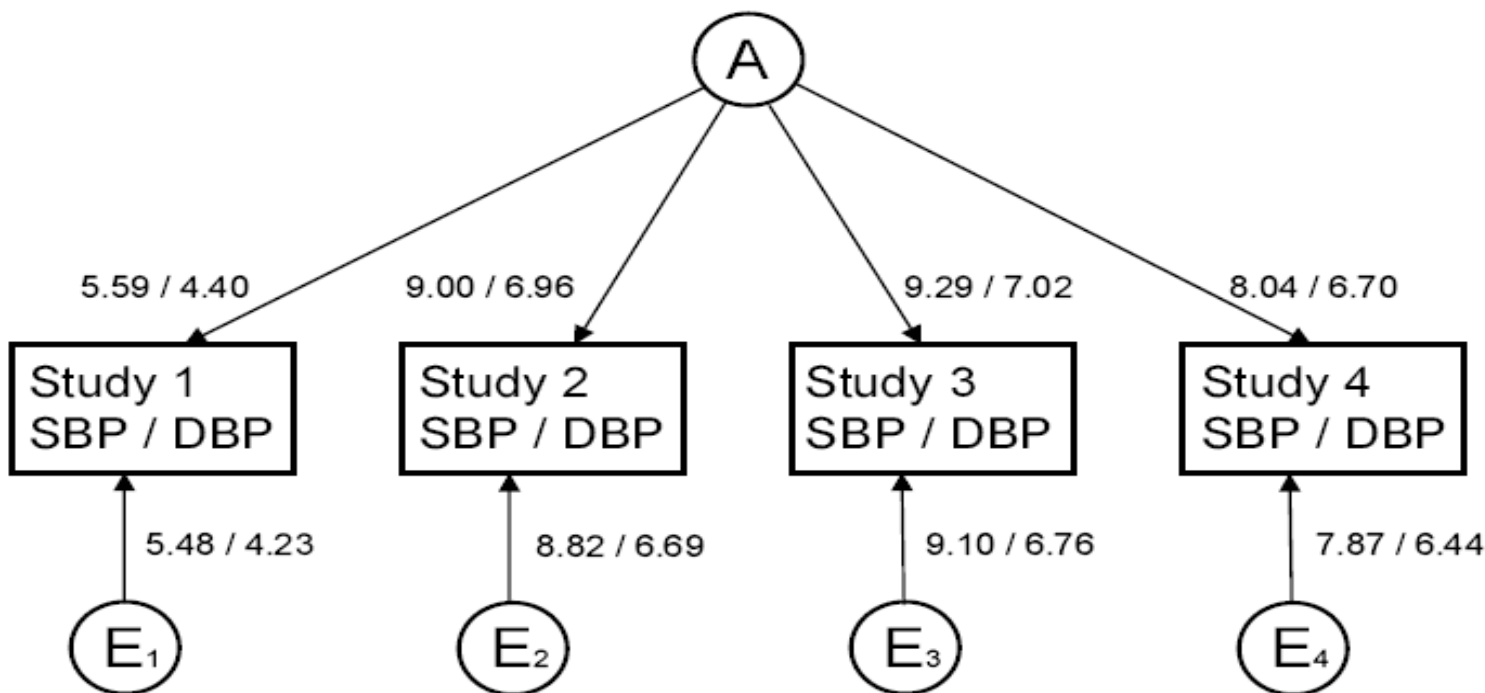
Question: contributions to stability t-1 to t

$$\{b_{A_{t-1,t}}^2 \text{var}(A_t) + b_{C_{t-1,t}}^2 \text{var}(C_t) + b_{E_{t-1,t}}^2 \text{var}(E_t)\}$$

---

$$\begin{aligned} & [\{b_{A_{t-1,t}}^2 \text{var}(A_t) + \text{var}(\zeta A_t) + \text{var}(a_t)\} \\ & + \{b_{C_{t-1,t}}^2 \text{var}(C_t) + \text{var}(\zeta C_t) + \text{var}(c_t)\} \\ & + \{b_{E_{t-1,t}}^2 \text{var}(E_t) + \text{var}(\zeta E_t) + \text{var}(e_t)\}] \end{aligned}$$

Decompose the phenotypic  
covariance into A,C,E components



**Figure 1**

Pathway model showing latent genetic and environmental influences on the measured systolic and diastolic blood pressure corrected for sex and age at measurement.

Note: A represents the additive genetic factors common to the four measurements. E<sub>1</sub>-E<sub>4</sub> shows the unique environmental influences at each time point/study. Path coefficients are shown. The proportions of measured variance for the latent variables are SBP: A = 51%, E = 49%; and DBP: A = 52%, E = 48%.



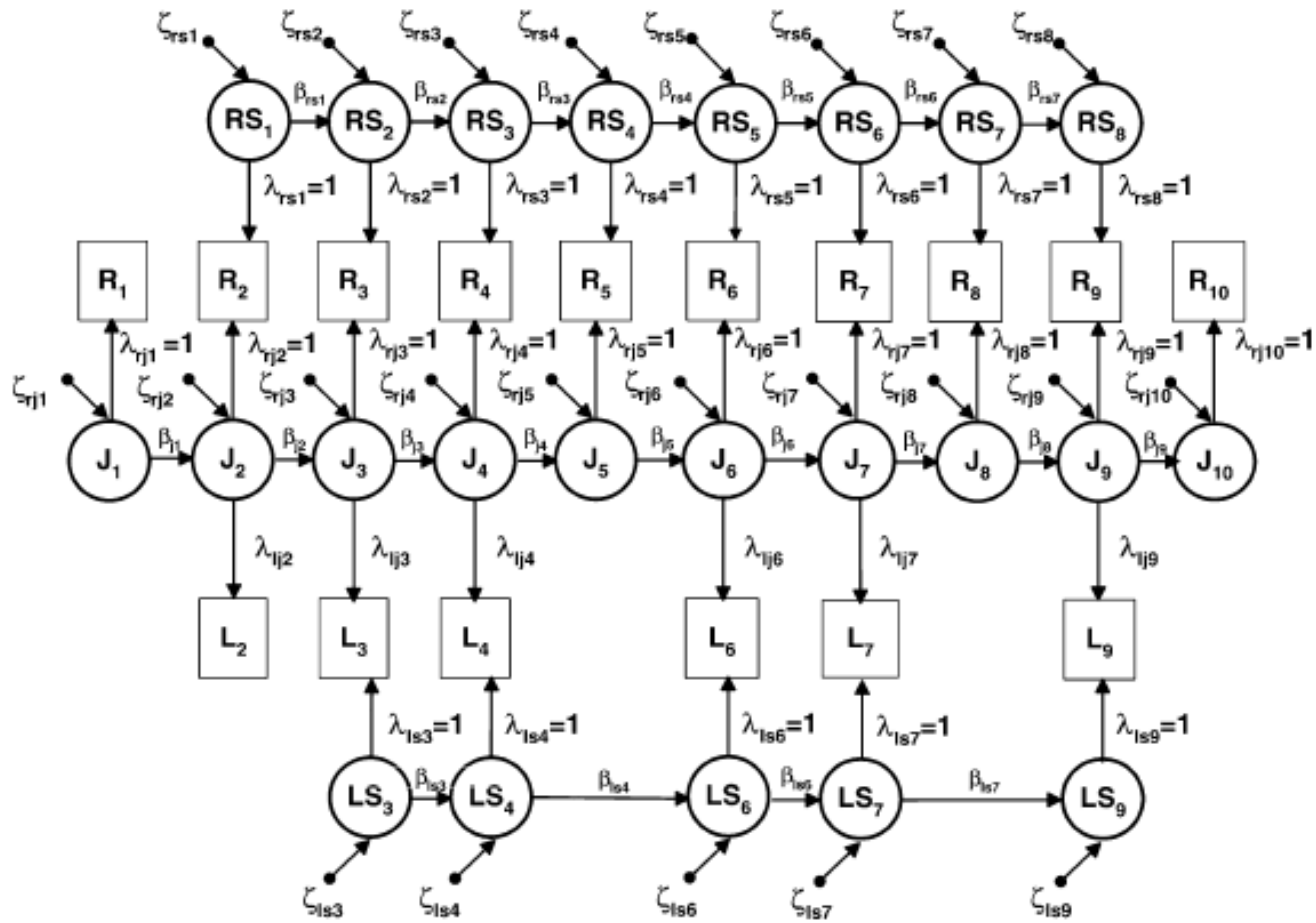
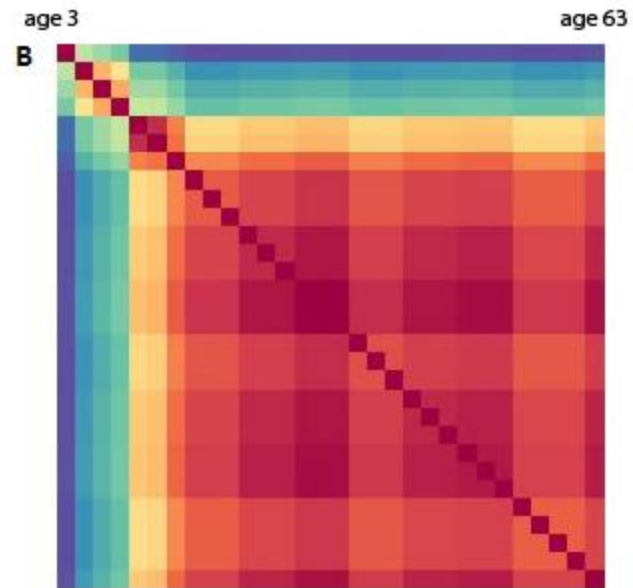
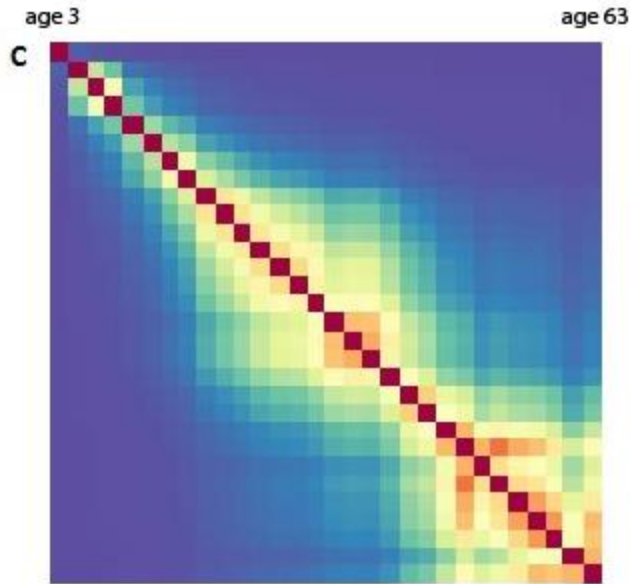
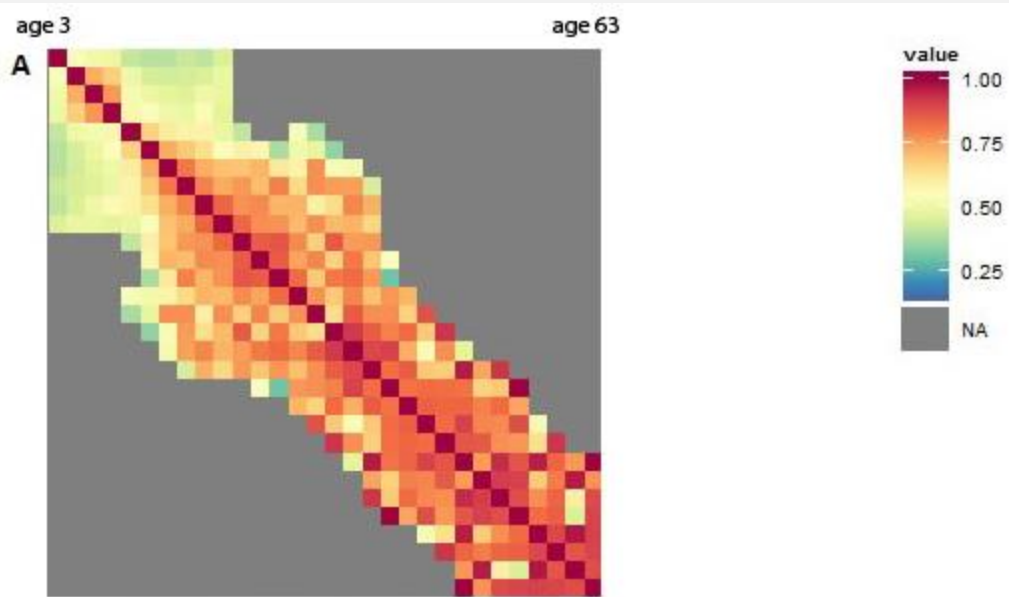
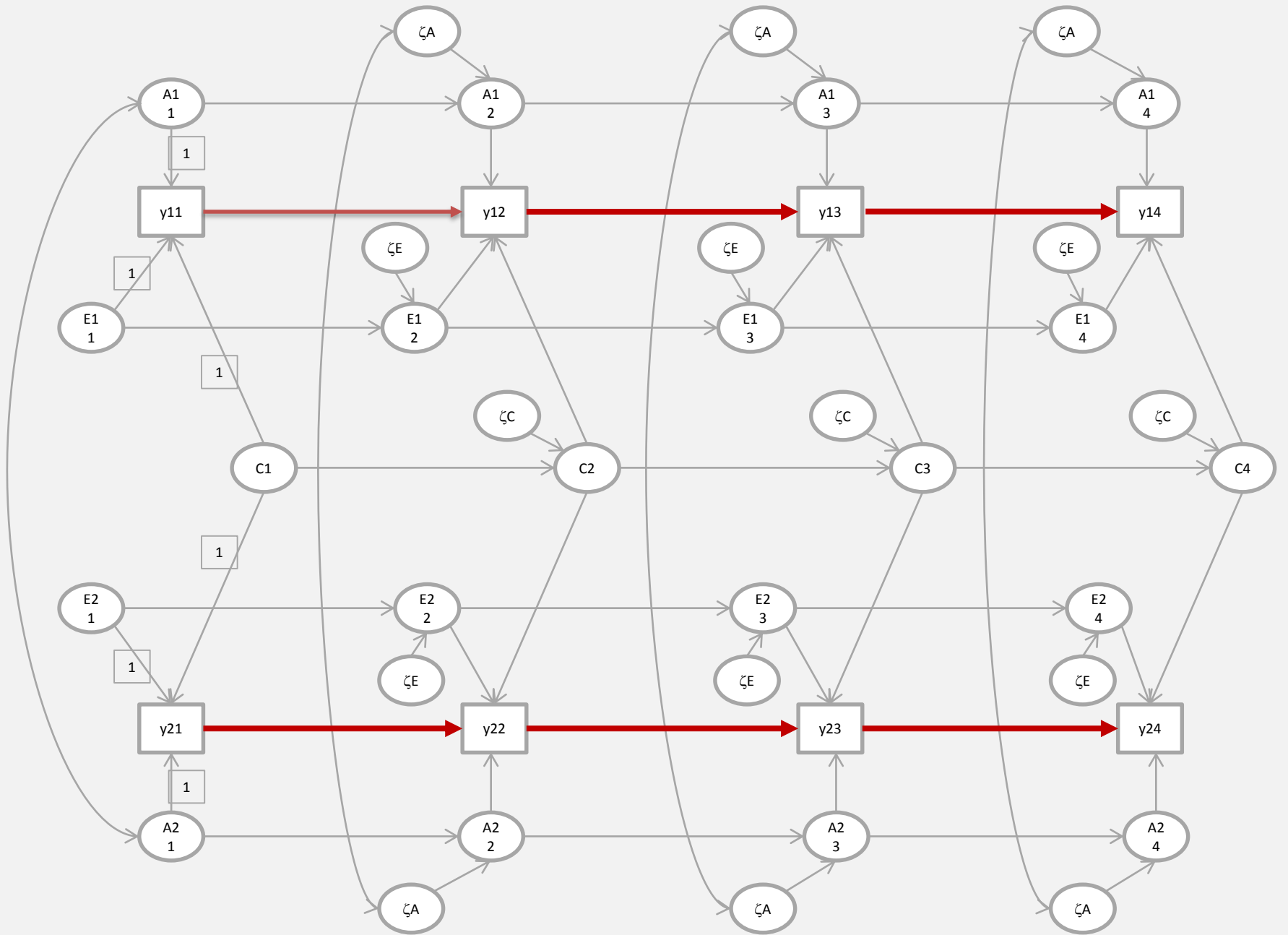


Fig. 3. Specification of the bivariate simplex model for the time course of the observed measurements  $\square$ . There are 10 time points for breath and 6 time points for the blood alcohol concentrations. The joint latent genetic ( $A$ ) (or QTL ( $Q$ )) process as estimated from the covariance between blood and breath alcohol readings is indicated by  $\circ$  as with labels  $J_{1-10}$ . The symbols,  $RS_{2-8}$  or  $LS_{2,3,4,6,7}$  and  $9$ , relate to those simplex processes that are respectively specific to breath or blood measures at the indicated times. Innovations of variance ( $\zeta_{rj1-10}$ ) are shown for the common genetic (or QTL) covariance, the covariance specific to blood and that specific breath, are shown as  $\zeta_{rs}$ . They are depicted in the identified model in relation to the latent variables. Transmission paths ( $\beta$ ) for the specific process for blood, breath and those for the process common to both breath and blood are related to the timing of readings by the respective suffixes, ( $\beta_{ls}$ ,  $\beta_{rs}$ , and  $\beta_{j}$ ). The transmission paths for blood readings take into account missing time points. Latent or  $\lambda$  variables are specified for the common pathway to enable the relative contributions of blood and breath readings in the time series to be compared. The model shown in this diagram is fitted for both  $A$  and  $Q$ . In addition a Cholesky decomposition of  $E$  is fitted.

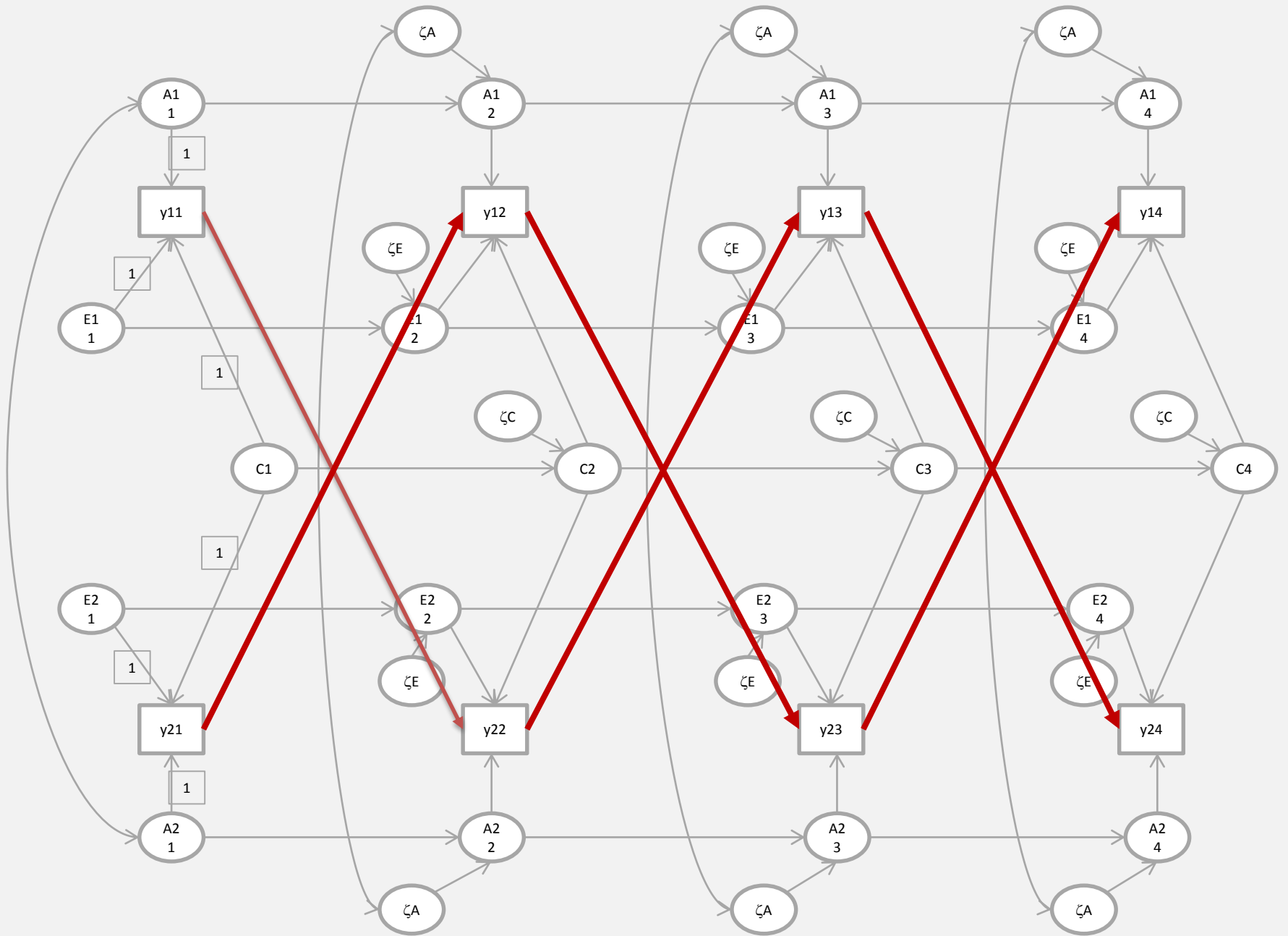


Anx/dep stability due to A and E from 3y to 63 years

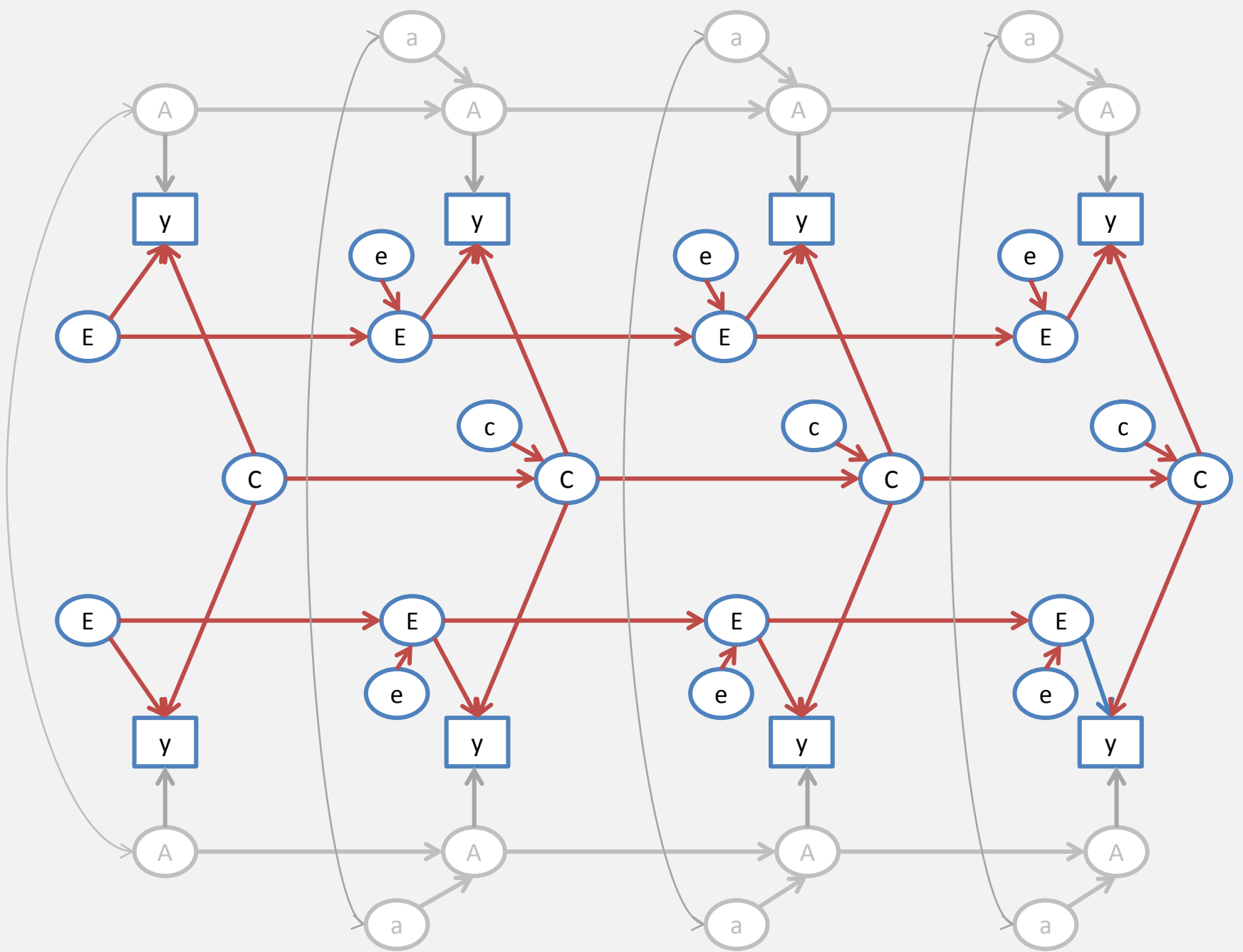
*Variations on the theme*

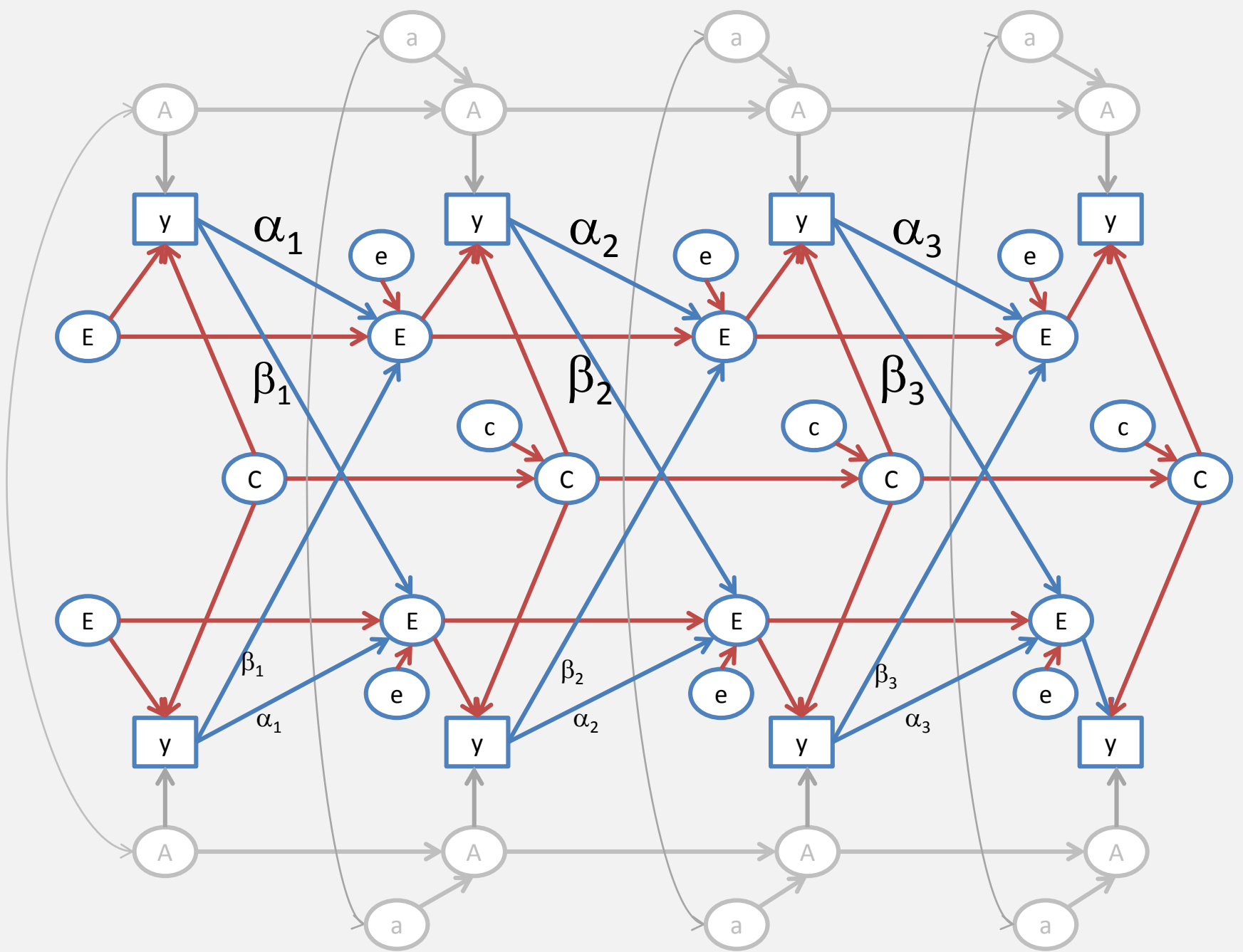


Phenotype-to-phenotype transmission (Eaves et al. 1986)



Sibling "interaction" mutual direct phenotypic influence (Eaves, 1976; Carey, 1986)





## “Niche-picking”

During development children seek out and create and are furnished surrounding (E) that fit their phenotype.

A smart child growing up will pick the niche that fits her/her phenotypic intelligence.

A anxious child growing up may pick out the niche that least aggravates his / her phenotypic anxiety.

Phenotype of twin 1 at time  $t$   $\rightarrow$  environment of twin 1 at time  $t+1$  (parameters  $\alpha_t$ )



## Mutual influences

During development children's behavior may contribute to the environment of their siblings.

A smart child growing up will pick the niche that fits her/her phenotypic intelligence and in so doing may influence (contribute to) the environment of his or her sibling.

A behavior of an anxious child may be a source of stress for his or her siblings.

Phenotype of twin 1 at time  $t$   $\rightarrow$  environment of twin 2 at time  $t+1$  (parameters  $\beta_t$ )

# ACE simplex T=4

Identification in ACE simplex **with no additional constraints.**

**Except:** Occasion specific residual variance decomposition:

$$\text{Var}[y(t) | A(t), E(t), C(t)] = \text{var}[\varepsilon(t)]$$

$$\text{Var}[\varepsilon(t)] = \text{var}[a(t)] + \text{var}[e(t)], t=1, \dots, 4$$

# Identification #1 T=4

$$ph_{ti} \rightarrow E_{(t+1)j} \quad (i=j)$$

$$\alpha_1, \alpha_2, \alpha_3$$

$$\alpha_1, \alpha_2 = \alpha_3$$

$$\alpha_k = b_{0\alpha} + (k-1) * b_{1\alpha} \quad \beta_k = b_{0\beta} + (k-1) * b_{1\beta} \quad (k=1,2,3)$$

but

$$\alpha_1 = \alpha_2, \alpha_3$$

$$ph_{ti} \rightarrow E_{(t+1)j} \quad (i \neq j)$$

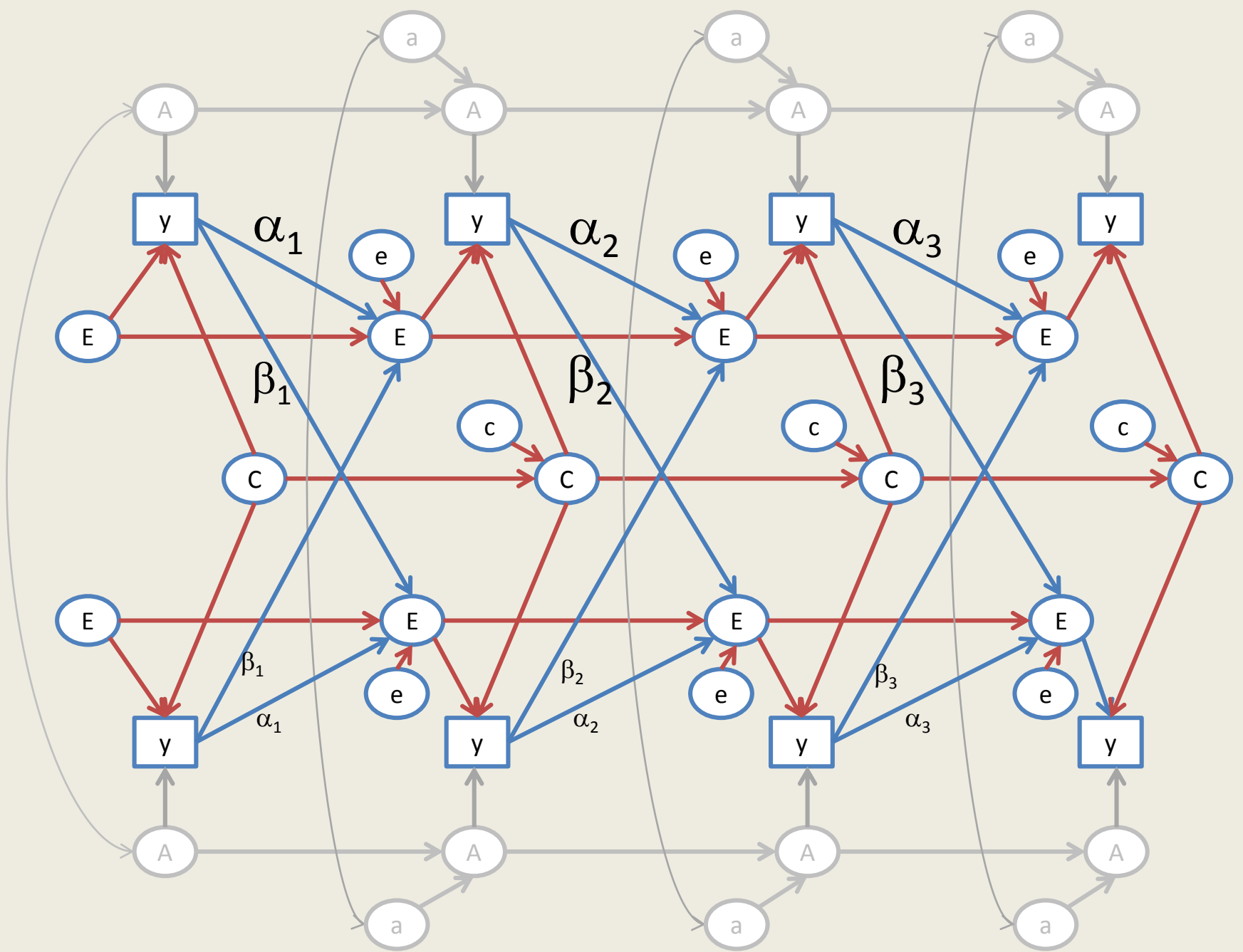
$$\beta_1, \beta_2, \beta_3 \quad (\text{not ID})$$

$$\beta_1, \beta_2 = \beta_3$$

$$\beta_1 = \beta_2, \beta_3 \quad (\text{not ID})$$

Presence of C **not relevant** to this results

Good.....?



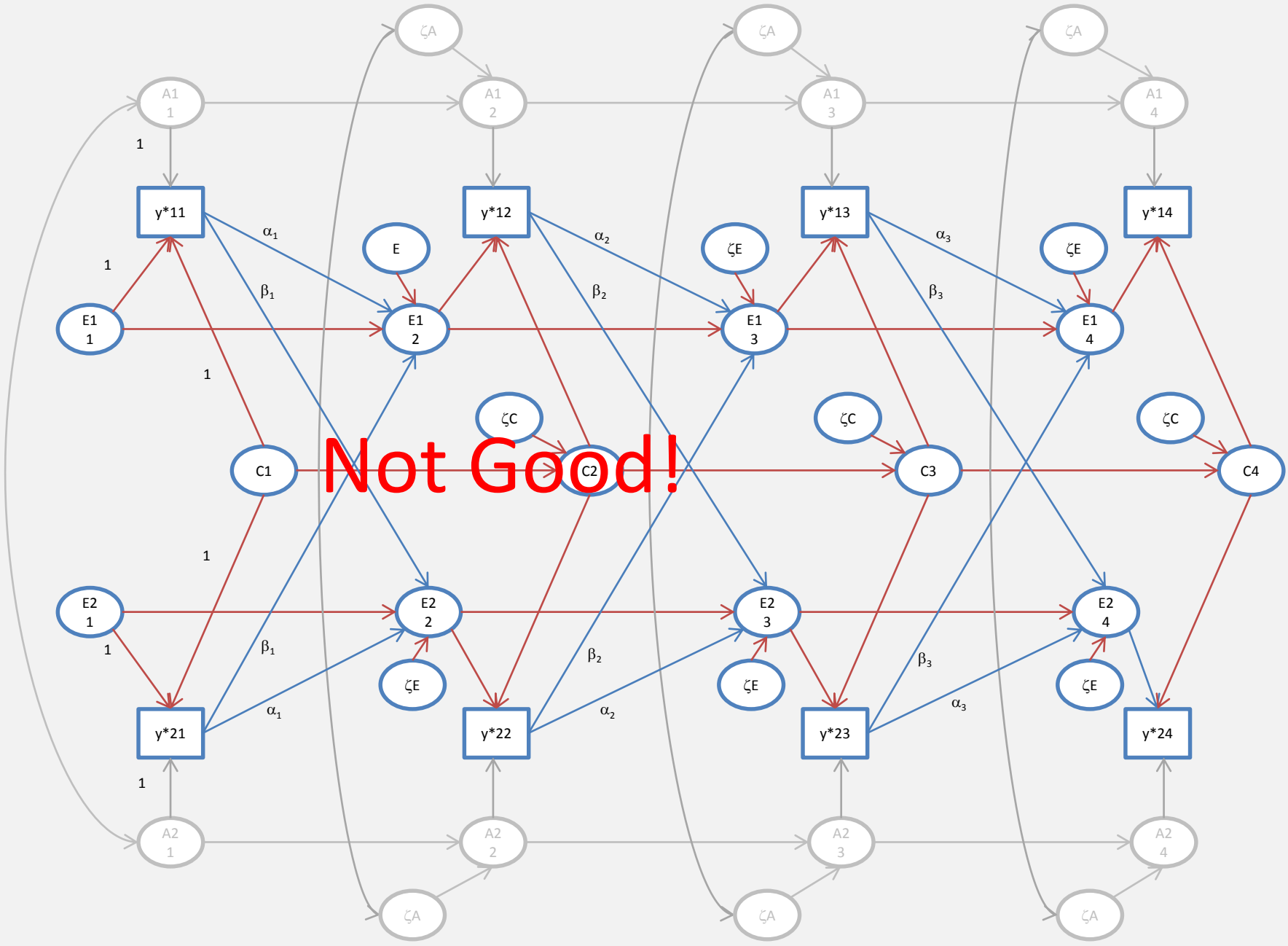
# N required given plausible values ACE

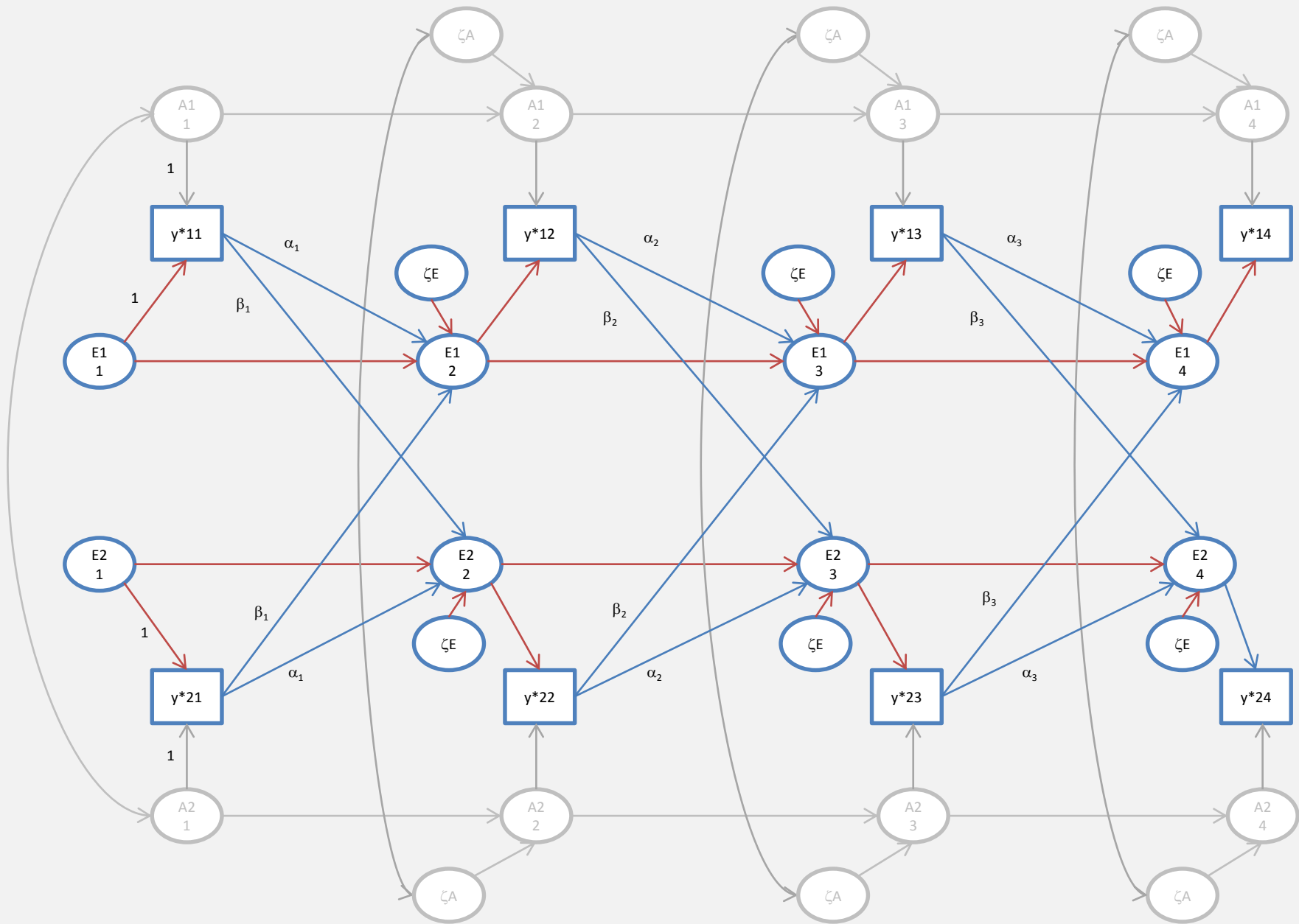
$$\alpha_1, \alpha_2 = \alpha_3 \text{ \& } \beta_1, \beta_2 = \beta_3$$

$\alpha_k$	$\beta_k$	$\sim N$
(power=.80)		
.10	.10	11700
.10	.15	4700
.15	.10	12100
.15	.15	4800

Hypothesis:  $\alpha_1 = \alpha_2 = \alpha_3 = 0$  &  $\beta_1 = \beta_2 = \beta_3 = 0$  (4df)

Good....?





N required given plausible values **AE**

$$\alpha_1, \alpha_2 = \alpha_3 \text{ \& } \beta_1, \beta_2 = \beta_3$$

$\alpha_k$	$\beta_k$	$\sim N$ (power=.80)
.10	.10	620
.10	.15	260
.15	.10	580
.15	.15	240

Good...?

Hypothesis:  $\alpha_1 = \alpha_2 = \alpha_3 = 0$  &  $\beta_1 = \beta_2 = \beta_3 = 0$  (4df)



# True $AE + \alpha_k$ & $\beta_k$ ( $Nmz = Ndz = 1000$ )

$\alpha_k$      $b_k$     ACE simplex  
df=61

**.10    .10    5.44**

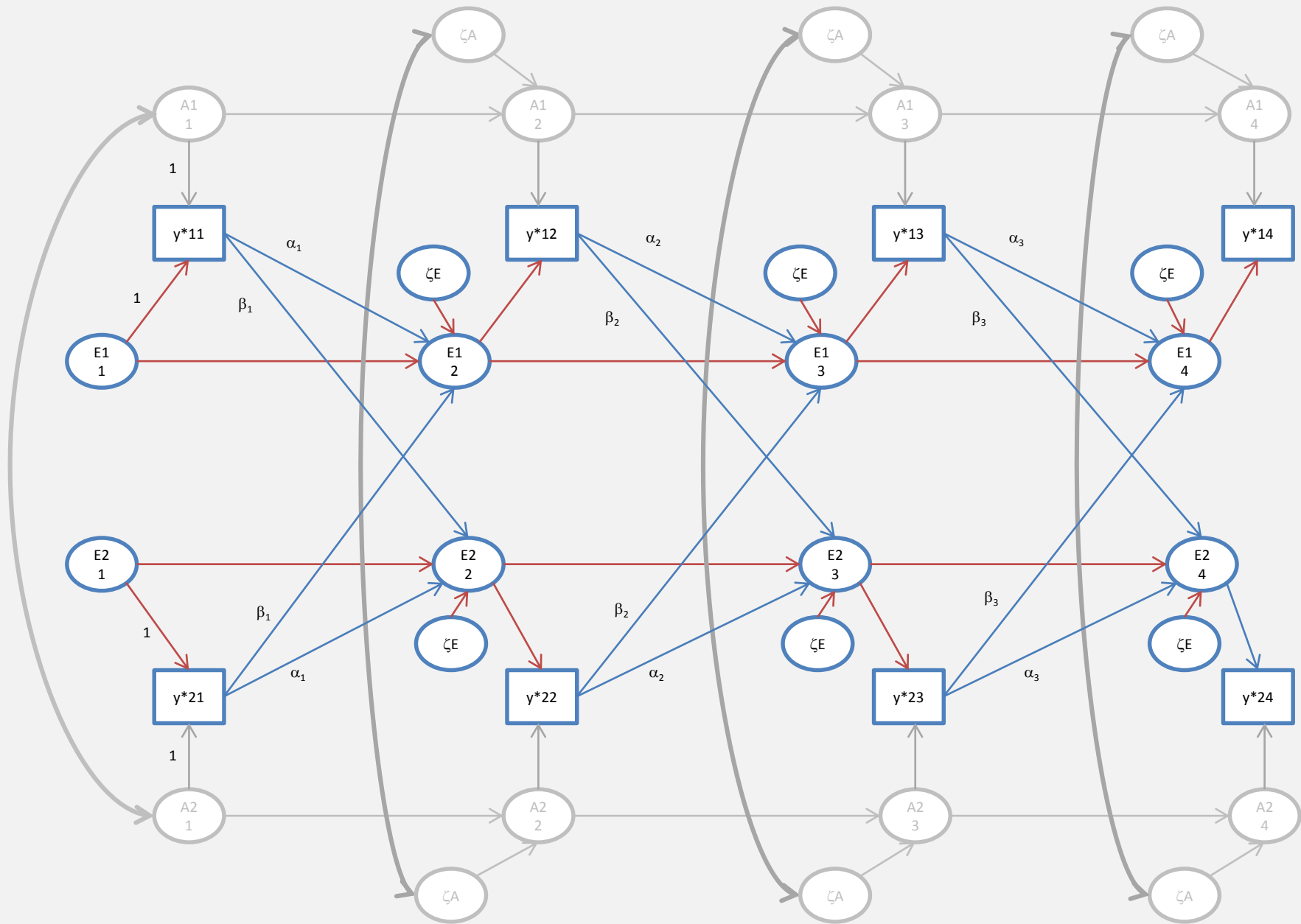
**.10    .15    11.22**

**.15    .10    5.62**

**.10    .15    11.54**

Good...?

(approximate model equivalence)



Full scale IQ.

261 MZ and 301 DZ twin pairs.

mean (std) ages

5.5y (.30), 6.8y (.19), 9.7y (.43), and 12.2y (.24).

The proportions of observed FSIQ data

0.812, 0.295, 0.490, 0.828 (MZ twin 1)

0.812, 0.295, 0.490, 0.828 (MZ twin 2)

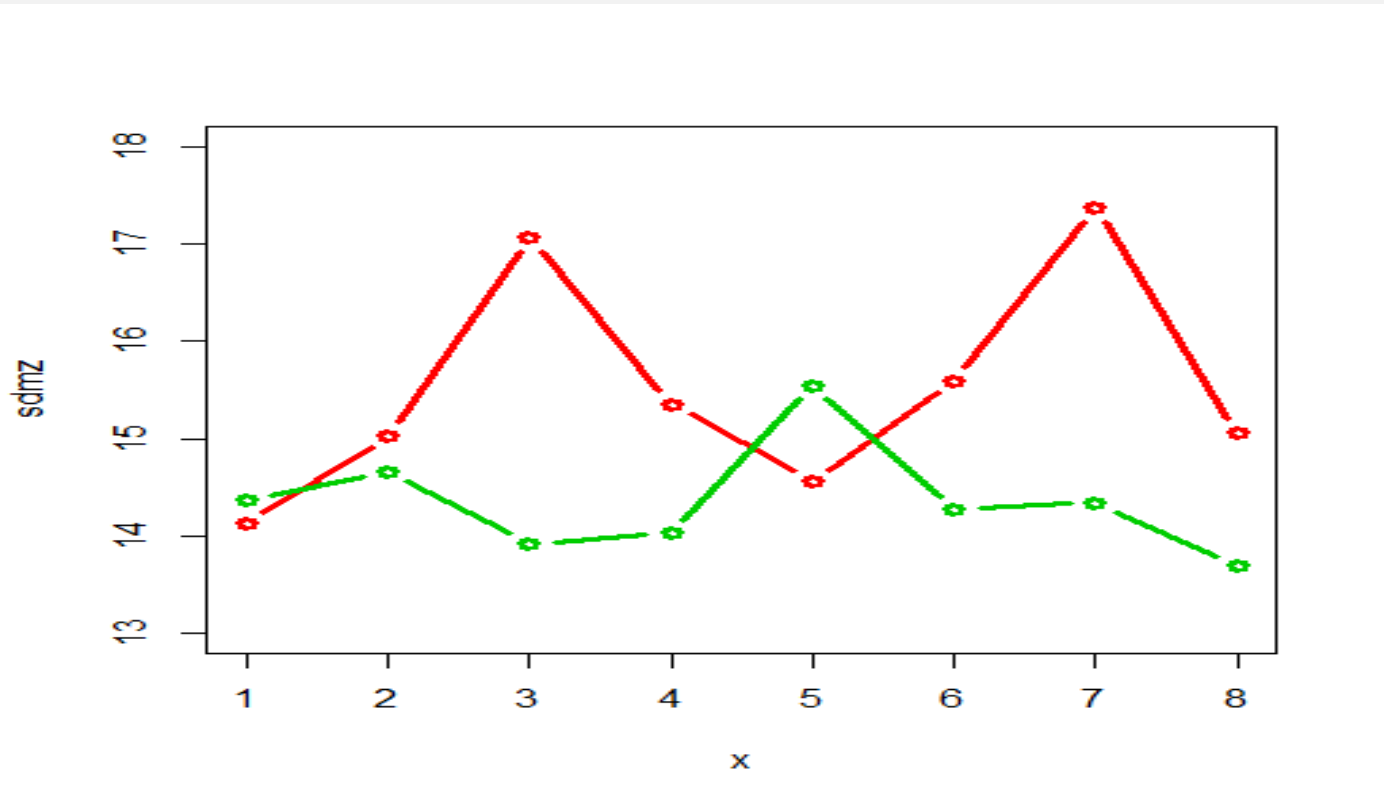
0.774, 0.379, 0.598, 0.797 (DZ twin1)

0.774, 0.379, 0.598, 0.797 (DZ twin 2)

5.5y 6.8y 9.7y 12.2y

0.770 0.674 0.840 0.802 MZ FSIQ correlation

0.641 0.482 0.481 0.500 DZ FSIQ correlation



Red: MZ stdevs

Green: DZ stdevs

## Fitted standard simplex

$\text{Var}(a_t) = \text{zero}$                       time specific A zero

$\text{Var}(\zeta_{A3}) = \text{var}(\zeta_{A4}) = 0$  A inno at t=3,4 zero

$\text{Chi2}(63) = 76.8, p = 0.11$

## FSIQ

5.5y

6.8y

9.7y

12.2y

A

1.000

0.801

0.801

0.801

variance

59.1

$h^2=.27$

1.000

1.000

1.000

104.3

$h^2=.48$

1.000

1.000

140.1

$h^2=.60$

1.000

110.8

$h^2=.54$

# E

1.000			
0.112	1.000		
0.056	0.212	1.000	
0.030	0.115	0.119	1.000
variance			
49.3	62.9	46.2	46.0
$e^2=.23$	$e^2=.29$	$e^2=.20$	$e^2=.22$

# C

1.000			
0.756	1.000		
0.457	0.478	1.000	
0.400	0.418	0.675	1.000
variance			
107.8	50.1	46.0	49.8
$c^2=.50$	$c^2=.23$	$c^2=.20$	$c^2=.24$



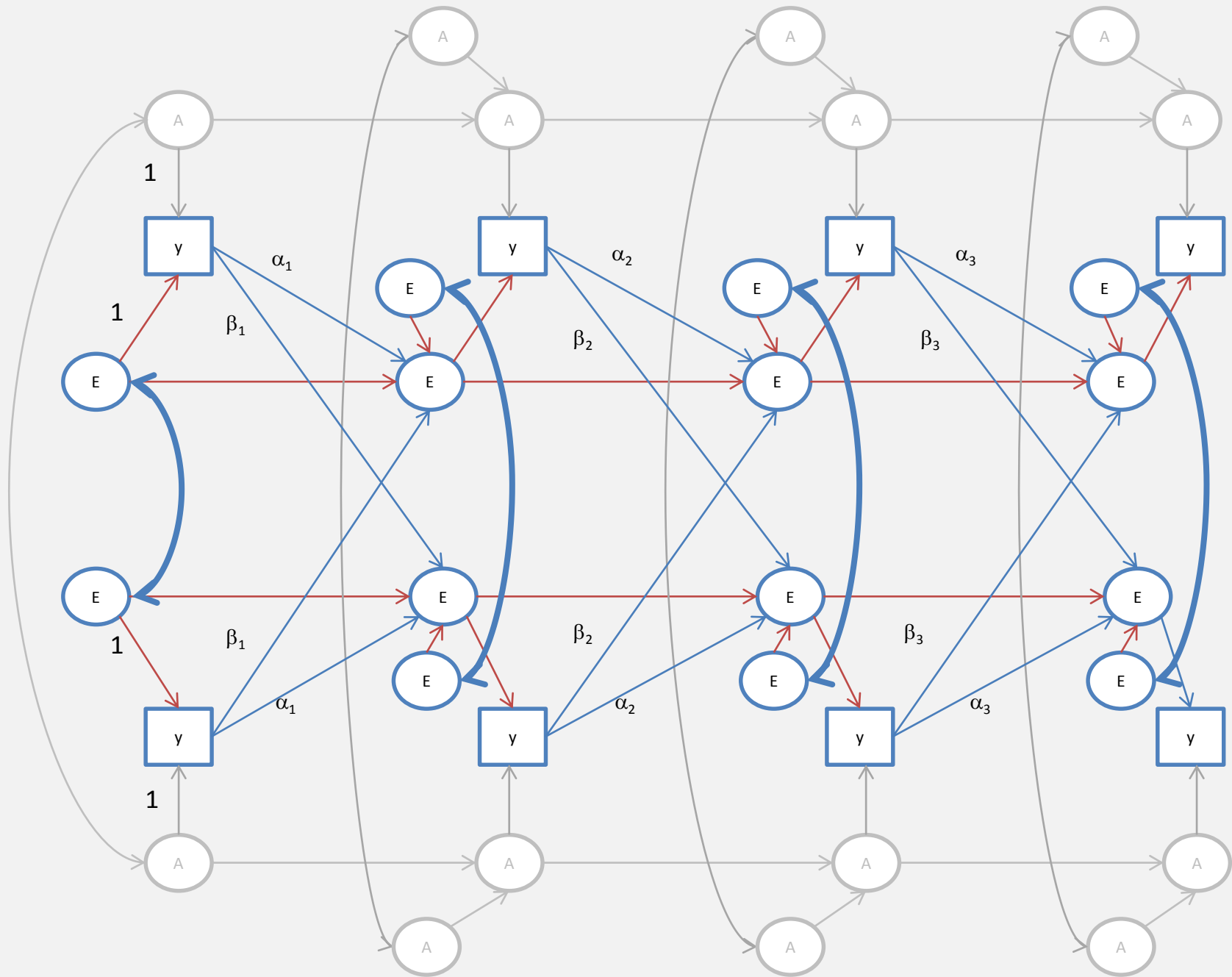
# Accommodation of genotype-environment covariance in a longitudinal twin design

*Where:* Netherlands Journal of Psychology, Volume 67, 81-90

Authors: Johanna M. de Kort\*, Conor V. Dolan\*,\*\* and Dorret I. Boomsma\*\*

Dolan, C.V, Janneke M. de Kort, Kees-Jan Kan, C. E. M. van Beijsterveldt, Meike Bartels and Dorret I. Boomsma (2014). Can GE-covariance originating in phenotype to environment transmission account for the Flynn effect? *Journal of Intelligence. Submitted.*

Dolan, C.V., Johanna M. de Kort, Toos C.E.M. van Beijsterveldt, Meike Bartels, & Dorret I. Boomsma (2014). GE Covariance through phenotype to environment transmission: an assessment in longitudinal twin data and application to childhood anxiety. *Beh. Gen. In Press.*



# Anxious depression

- Twins aged 3,7,10,12 Netherlands Twin Register (NTR), which includes the Young NTR (YNTR; van Beijsterveldt, Groen-Blokhuis, Hottenga, et al., 2013)
- ASEBA CBCL instruments (Achenbach), maternal ratings
- Observed 89%, 54%, 45%, 37% (NMZ=3480)
- Observed 89%, 50%, 39%, 32% (NDZ=3145)

# Phenotypic correlations

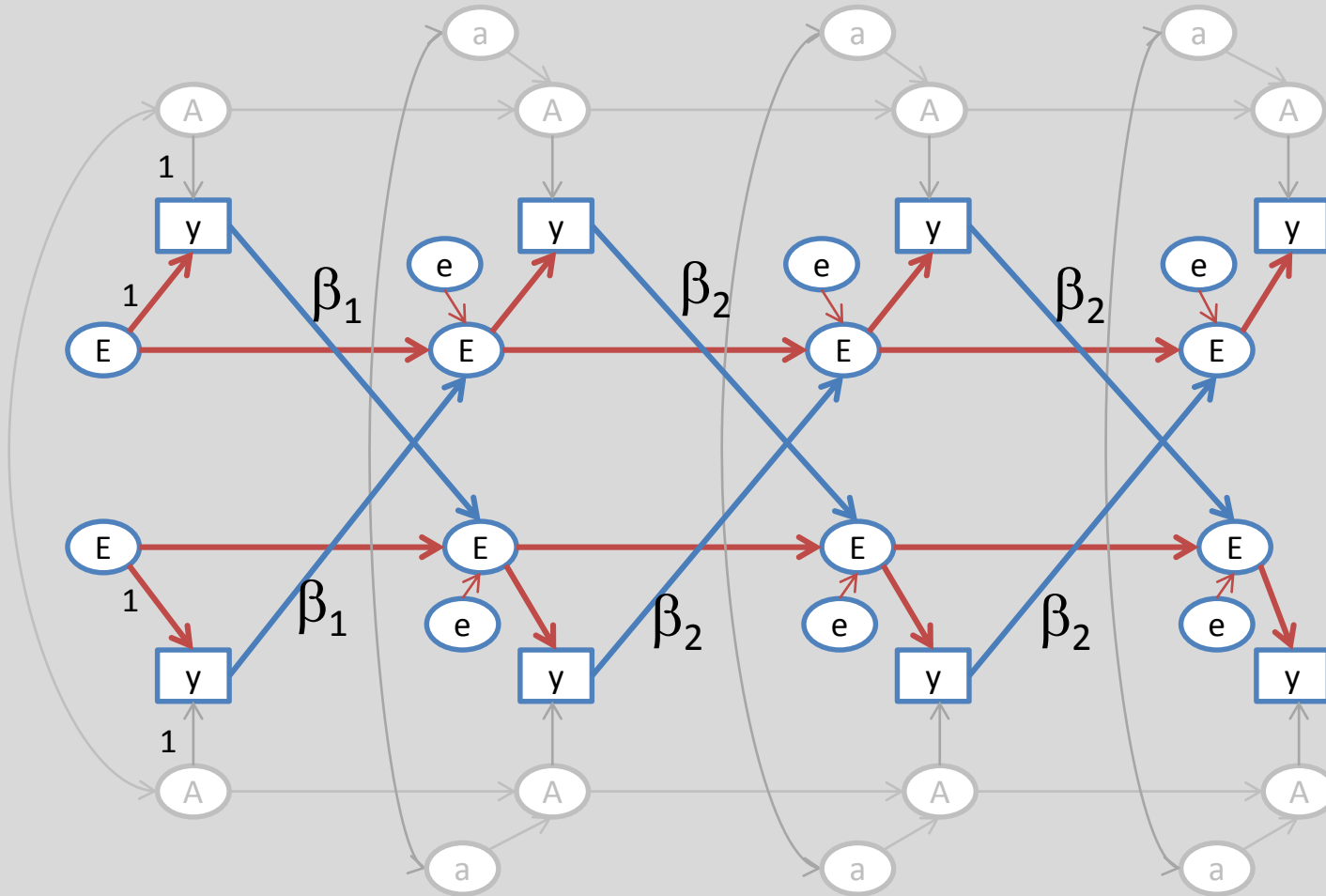
## FIML phenotypic twin correlations

MZ: .71 (3), .58 (7), .58 (10), and .63 (12).

DZ: .31 (3), .36 (7), .35 (10), and .40 (12).

From 7y onwards looks like ACE model

Model	logl	npar	AIC	BIC
1 ACE	-69528.5	28	139113	139303
3 AE	-69537.9	20	139115	139251
4 AE + $\alpha_k, \beta_k$	-69519.3	24	<u>139086</u>	139249
5 AE + $\beta_k$	<b>-69522.1</b>	<b>22</b>	<b>139088</b>	<u><b>139237</b></u>
6 ACE ph1->ph2 (2)	<b>-69522.0</b>	<b>24</b>	<b>139093</b>	<b>139256</b>
7 ACE ph->ph (1)	-69537.1	22	139119	139268



$\beta_1=0.123$  (s.e. .041) and  $\beta_2 = \beta_3 = 0.062$  (s.e. .027)

Parameter estimates in the analysis of Anxiety from 3y to 12y. ML estimates and robust standard errors in parentheses in the AE simplex with parameter  $\beta_k$ ,  $\log l = -69522.1$ ).

	t=1 (3y)	t=2 (7y)	t=3 (10y)	t=4 (12y)
$b_{A_t,t-1}$	-	0.277 (.062)	0.886 (.099)	0.790 (.100)
$\sigma[\zeta_{A_t}]$	2.450 (.098)	1.881 (.152)	1.189 (.302)	1.214 (.277)
$b_{E_t,t-1}$	-	0.414 (.182)	1.017 (.232)	0.799 (.125)
$\sigma[\zeta_{E_t}]$	1.034 (.240)	1.173 (.229)	1.212 (.307)	0.728 (.477)
$\sigma[a_t]$	0.956 (.227)	0.956 (.227)	0.956 (.227)	0.956 (.227)
$\sigma[e_t]$	1.375 (.173)	1.375 (.173)	1.375 (.173)	1.375 (.173)
$\beta_1$	0.123 (.041)			
$\beta_2=\beta_3$	0.062 (.027)			

A1

1.00  
 0.34 1.00  
 0.28 0.83 1.00  
 0.23 0.67 0.81 1.00

**0.00 0.00 0.00 0.00**  
**0.12 0.04 0.03 0.03**  
**0.11 0.07 0.06 0.04**  
**0.12 0.10 0.10 0.08**

0.50 0.17 0.14 0.11

0.17 0.50 0.42 0.34

0.14 0.42 0.50 0.41

0.11 0.34 0.41 0.50

**0.00 0.00 0.00 0.00**

**0.23 0.08 0.07 0.05**

**0.20 0.13 0.11 0.09**

**0.20 0.18 0.18 0.14**

E1

G-E covariance

1.00  
 0.33 1.00  
 0.25 0.74 1.00  
 0.22 0.66 0.89 1.00

A1

E1

A2

Correlated E – But not C!

**0.00 0.10 0.09 0.09**

**0.10 0.09 0.11 0.15**

**0.09 0.11 0.12 0.18**

**0.09 0.15 0.18 0.22**

E2



## Identification #2 T=4

$$b_{A2,1} = b_{A3,2} = b_{A4,3} \text{ OR}$$

$$b_{C2,1} = b_{C3,2} = b_{C4,3} \text{ OR}$$

$$b_{E2,1} = b_{E3,2} = b_{E4,3} \text{ OR (of course)}$$

$$b_{C2,1} = b_{C3,2} = b_{C4,3} = b_{E2,1} = b_{E3,2} = b_{E4,3}$$

$$ph_{ti} \rightarrow E_{(t+1)j} \text{ (} i=j \text{)}$$

$$\alpha_1, \alpha_2, \alpha_3$$

$$ph_{ti} \rightarrow E_{(t+1)j} \text{ (} i \neq j \text{)}$$

$$\beta_1, \beta_2, \beta_3$$