

Markov Modeling with Genetically Informative Data

Michael C Neale, PhD

Virginia Institute for Psychiatric and Behavioral Genetics

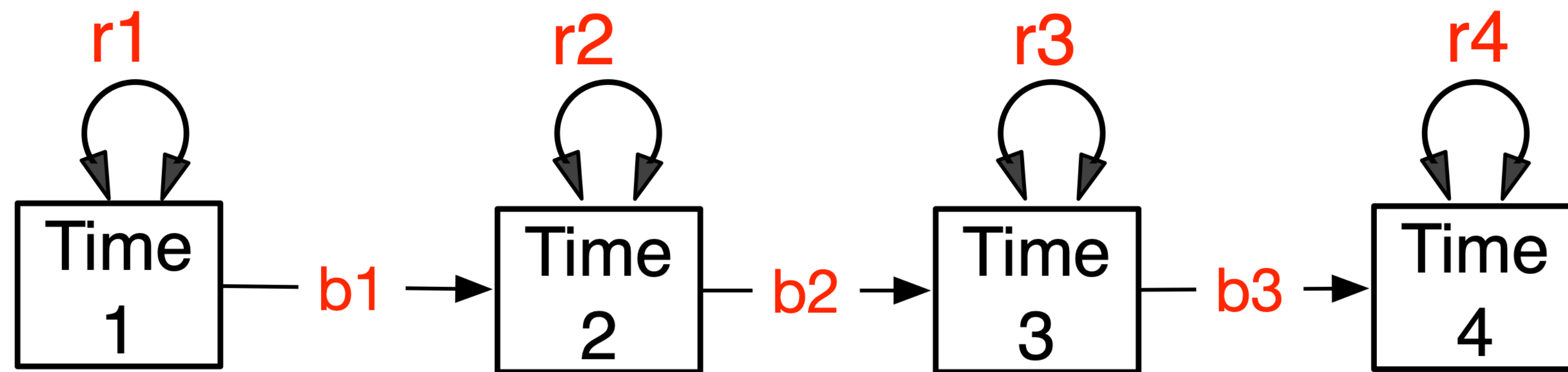
Virginia Commonwealth University

Boulder Colorado NIMH Workshop June 2022

Overview

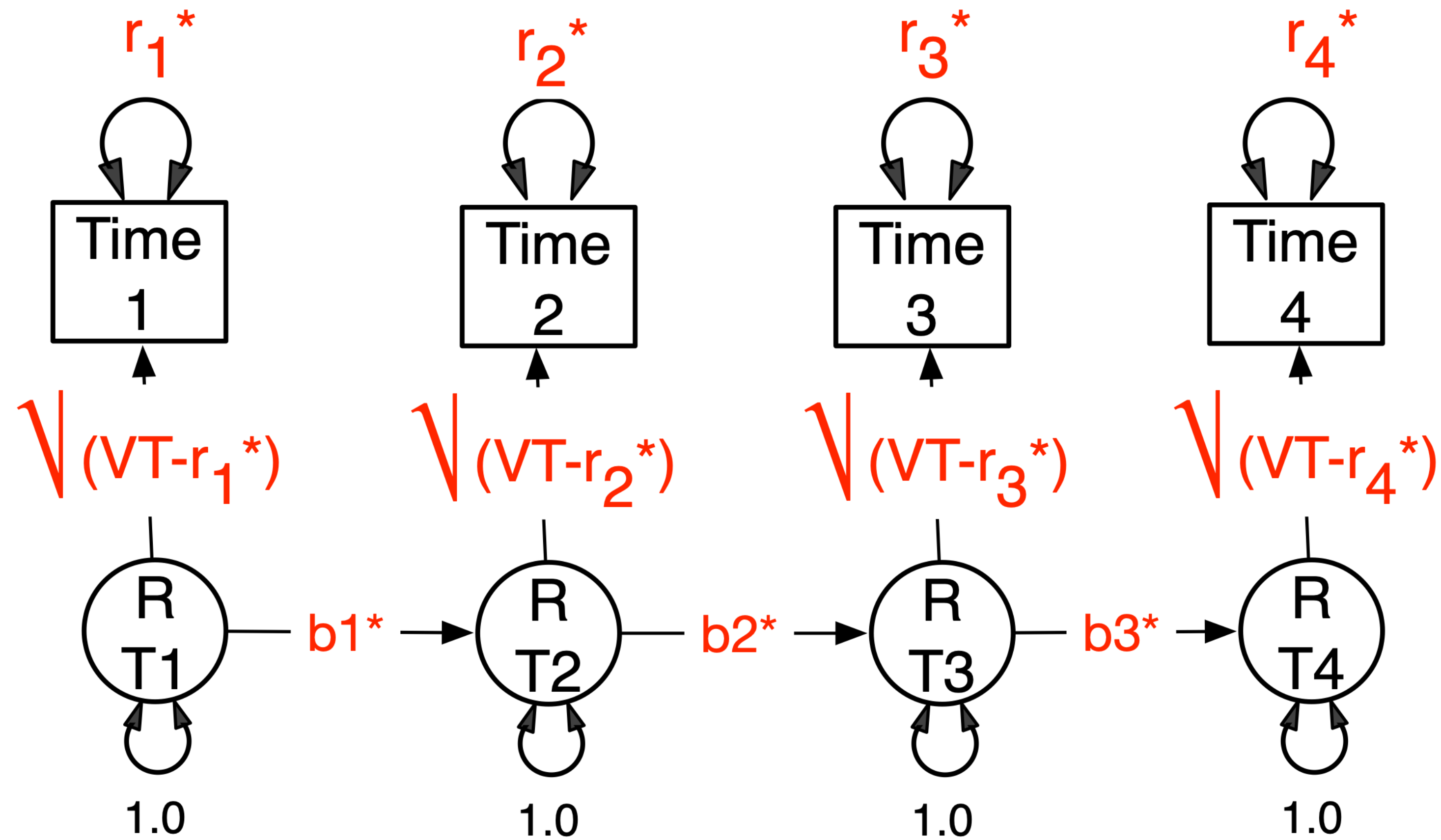
- Markov Models Generally
- Markov Models Genetically
- Markov Model with Niche Selection
- Genomic possibilities

A Markov/Simplex/Autoregressive/AR1 Model



- $r2=r3=r4$ and $b1=b2=b3$ are testable hypotheses
- What if there's measurement error?
- What if intervals between assessments are unequal?

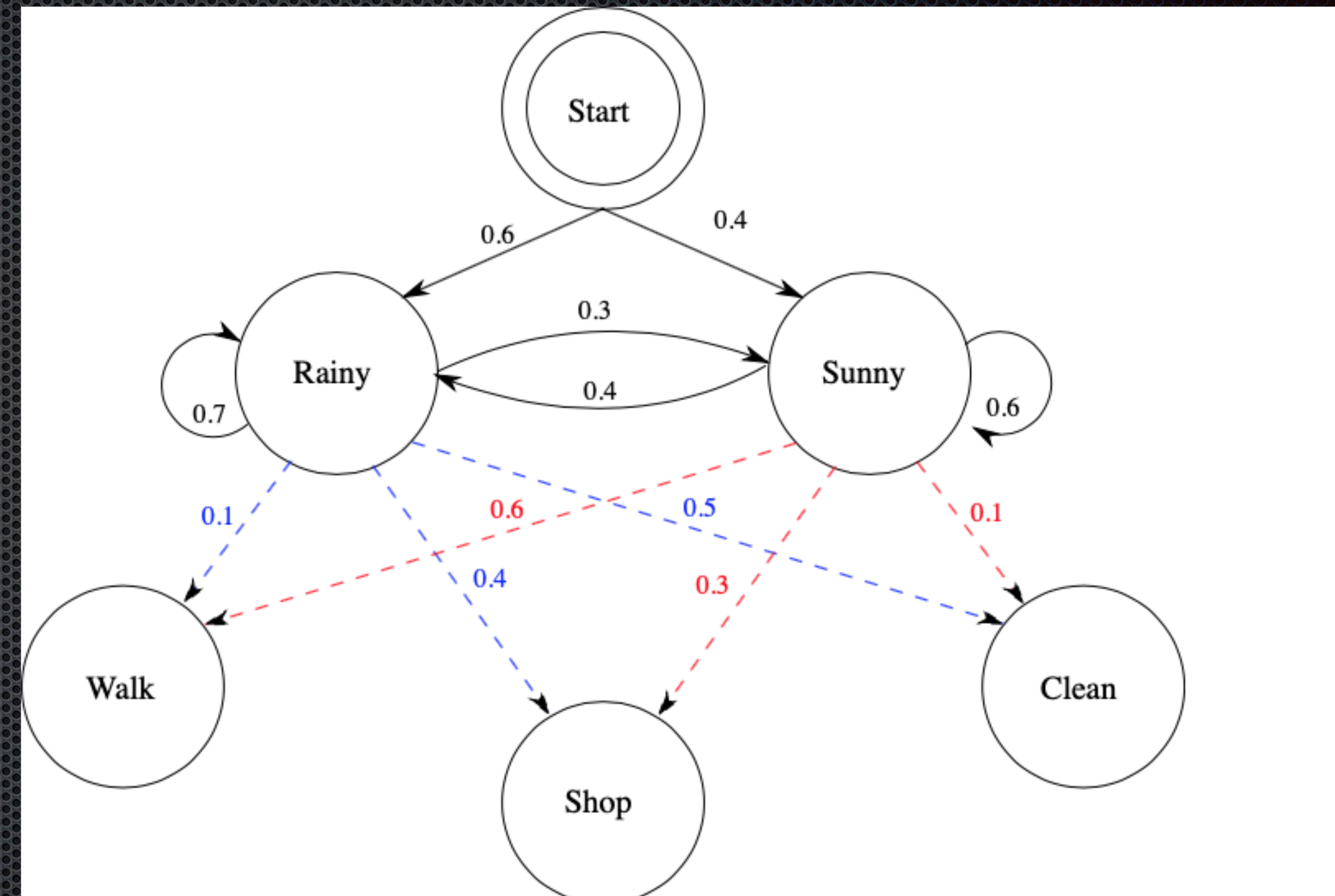
A Markov/Simplex/Autoregressive/AR1 Model with Measurement Error



- $r_1^*:r_4^*$ need external (test-retest) data

Hidden Markov Models

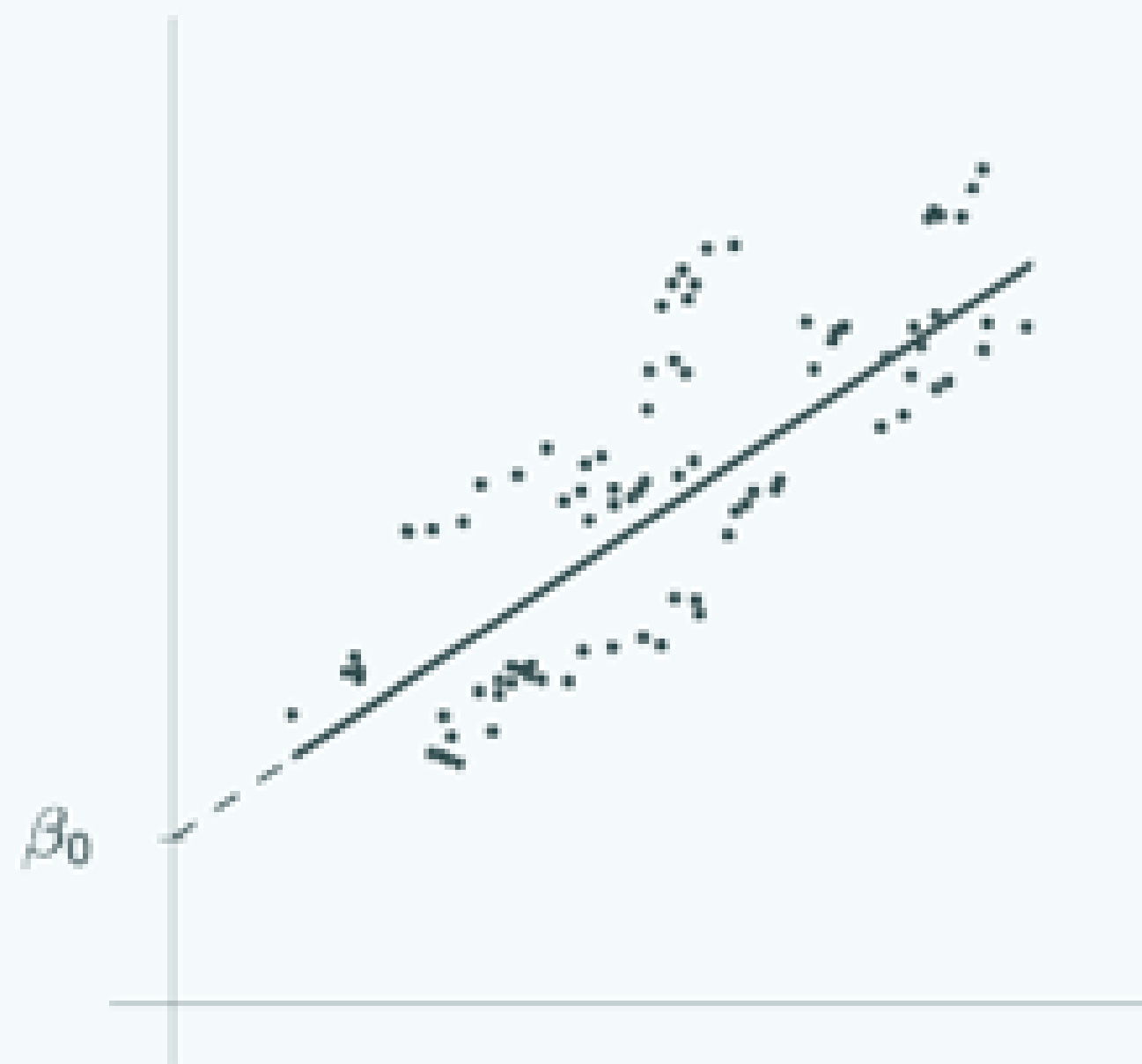
- HMMs have multiple uses, e.g., natural language processing
- Viterbi algorithm to reduce number of calculations
- Kalman Filter for long time series
- OpenMx has fully developed State Space Modeling capabilities



Example HMM: infer rain/sun from sequence of activities

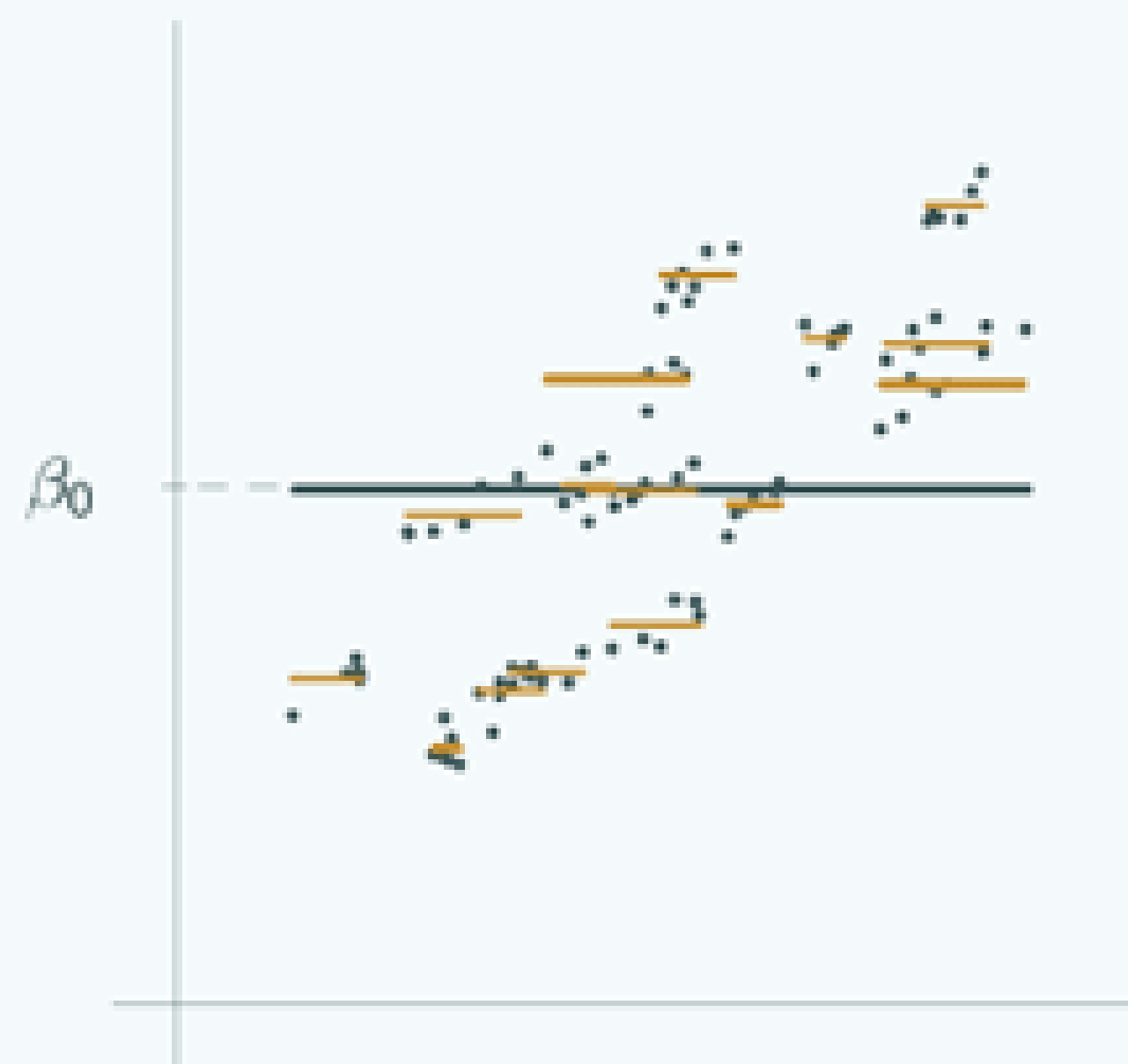
Three Models for Multigroup Data

Single level regression model



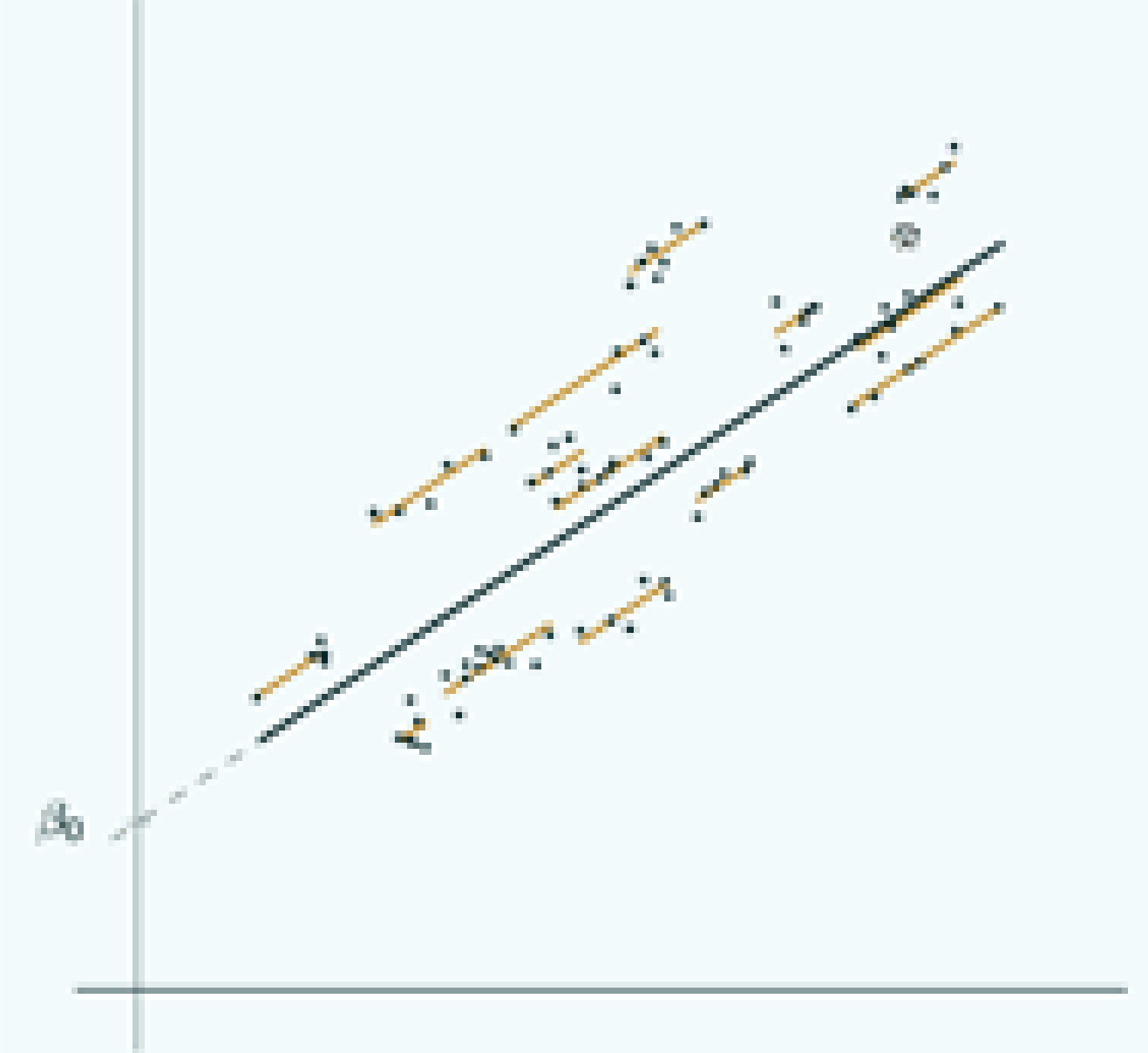
Ignore groups, estimate
One x-y regression

Variance components model



Estimate group means and
variances. No regression.

Random intercept model

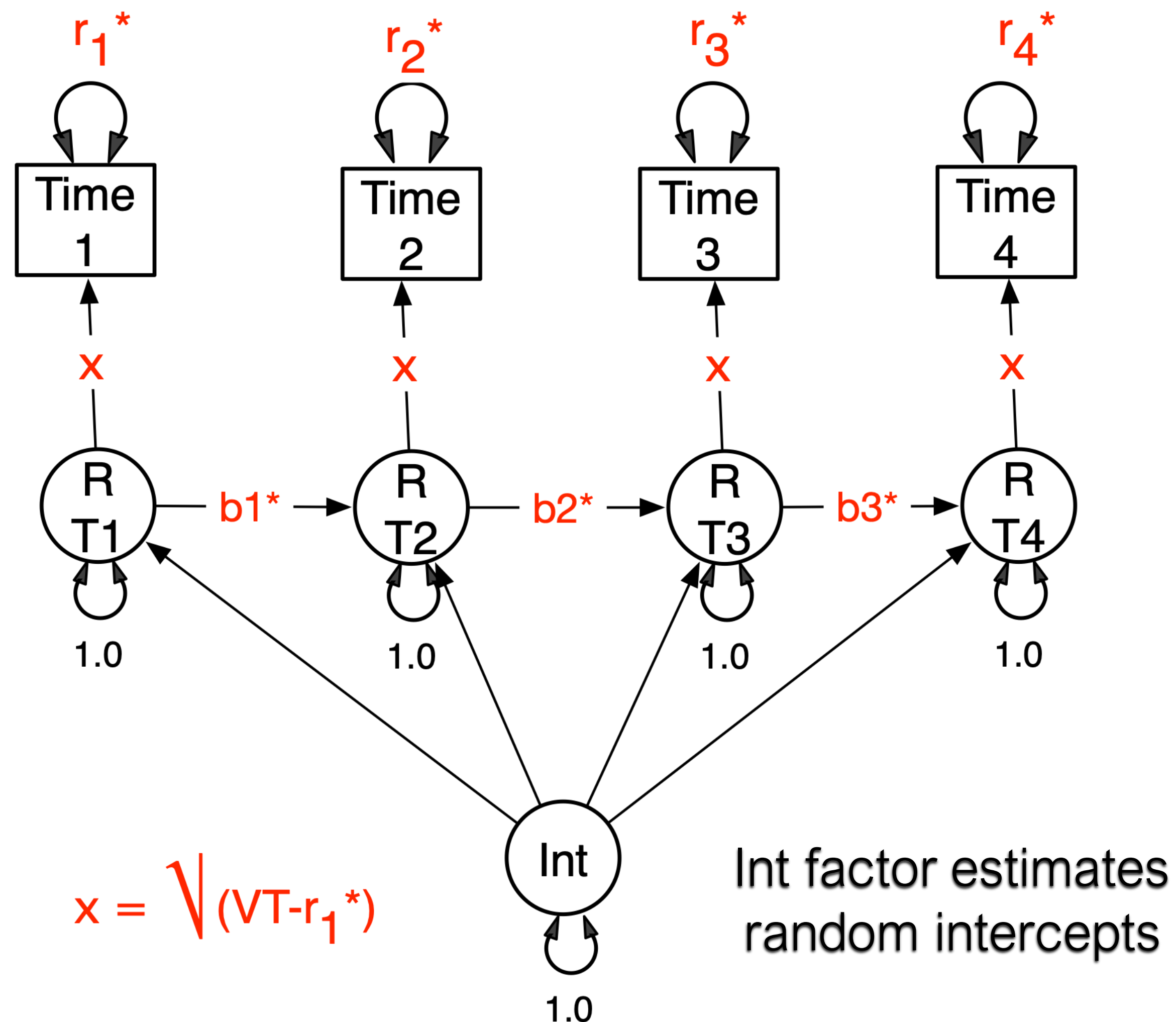


Estimate same regression,
Different intercepts.

Alternative approaches to random intercepts and slopes

- Multigroup approach
 - Estimate separate means, variances, covariances for every group
- Multilevel approach
 - Assume these statistics come from a population with a given (usually normal) distribution & estimate the parameters of this distribution

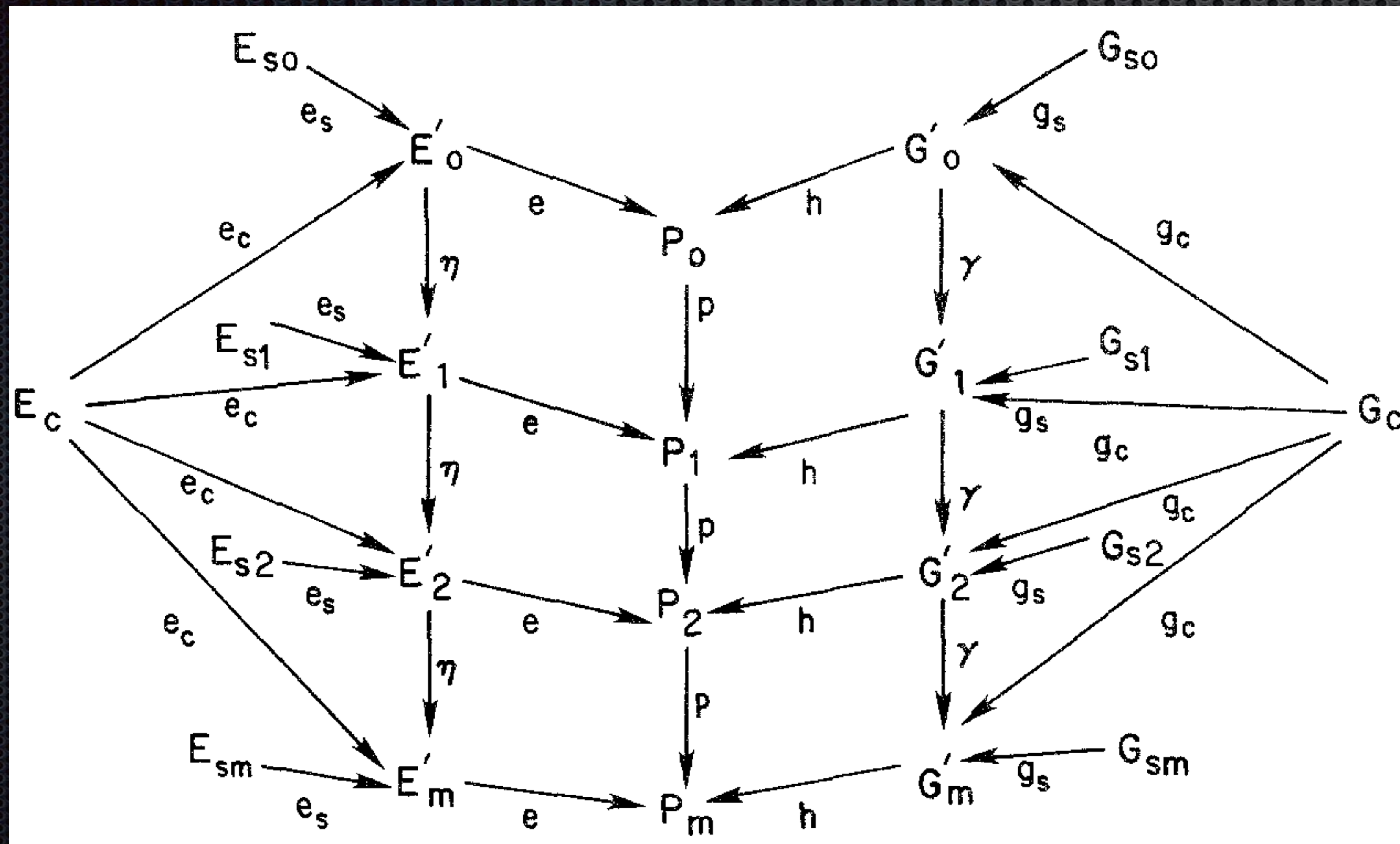
AR1 Model with Random Intercepts



What do random intercepts do?

- Models individual change around the individual's mean, not the group mean
- Adds only 1 parameter to regular AR models
- Accounts for common influences on each timepoint
- Does not automatically add random slopes

Biometrical AR1 Model with Random Intercepts



Markov Process
(time to time)

Factors G_c & E_c
(constant)

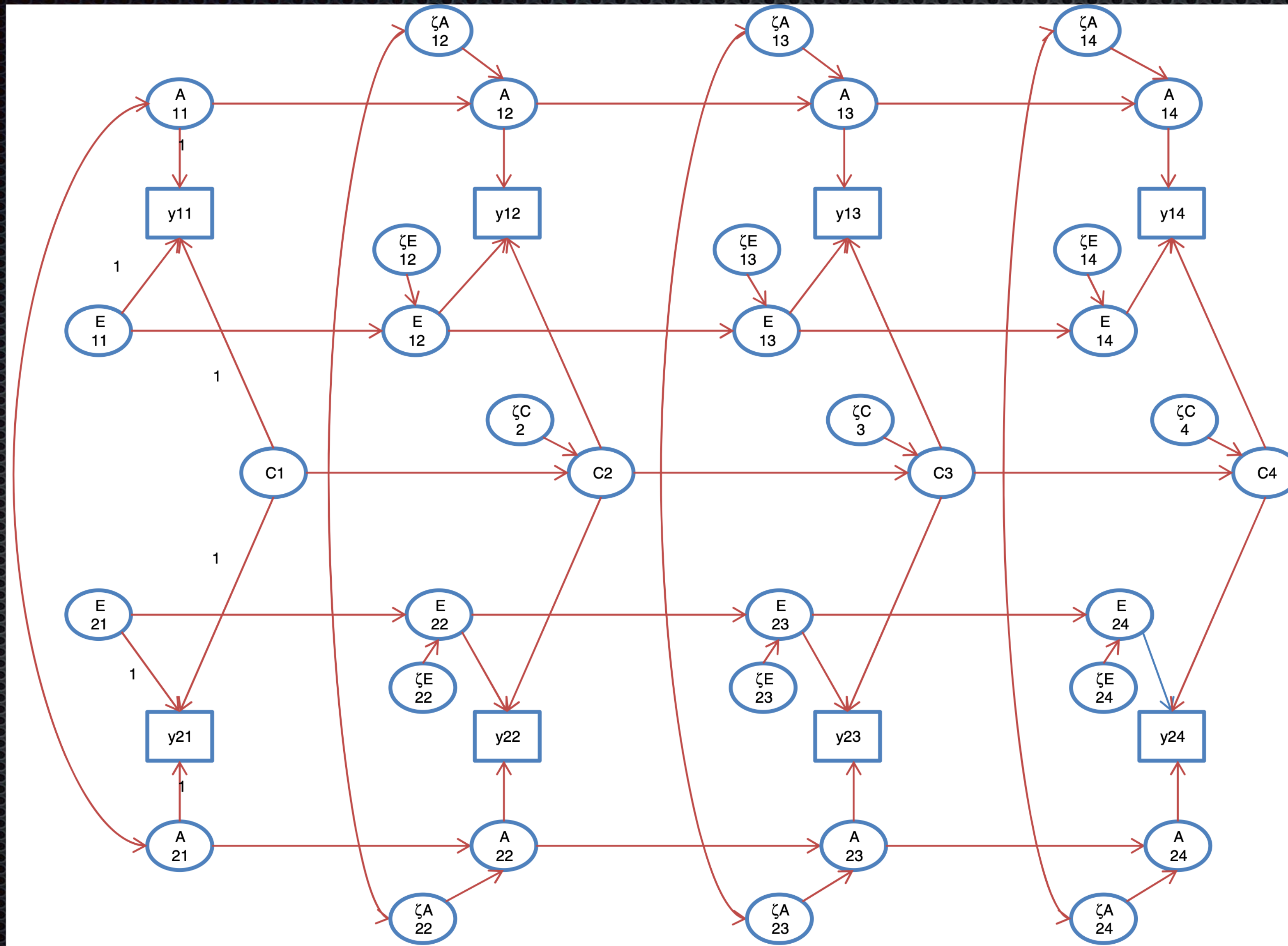
Time-specific Factors
 $G_0 \dots G_m$ etc. (transient)

Eaves et al 1986 A theory of developmental
change in quantitative phenotypes applied to
cognitive development. Behavior Genetics
16(1):143-62

Biometrical AR1 Model with Random Intercepts

- Model not identified as shown
- Must fix $\rho=0$ or $\gamma=\varepsilon=0$
- Can add C or D factors and transmission
- Relative to population or individual mean
- What about cross-variable transmission?

ACE Simplex à la Dolan et al 2014



Transmission Infidelity

- $A_1 \rightarrow A_2, C_1 \rightarrow C_2, E_1 \rightarrow E_2$ or $P_1 \rightarrow P_2$: Natural Markov
- $P_1 \rightarrow E_2$: niche generation
- $E_1 \rightarrow P_2$: a 'pleiotropic' effect of environment not via $E \rightarrow E$

Extended ACE Simplex: P- \rightarrow E

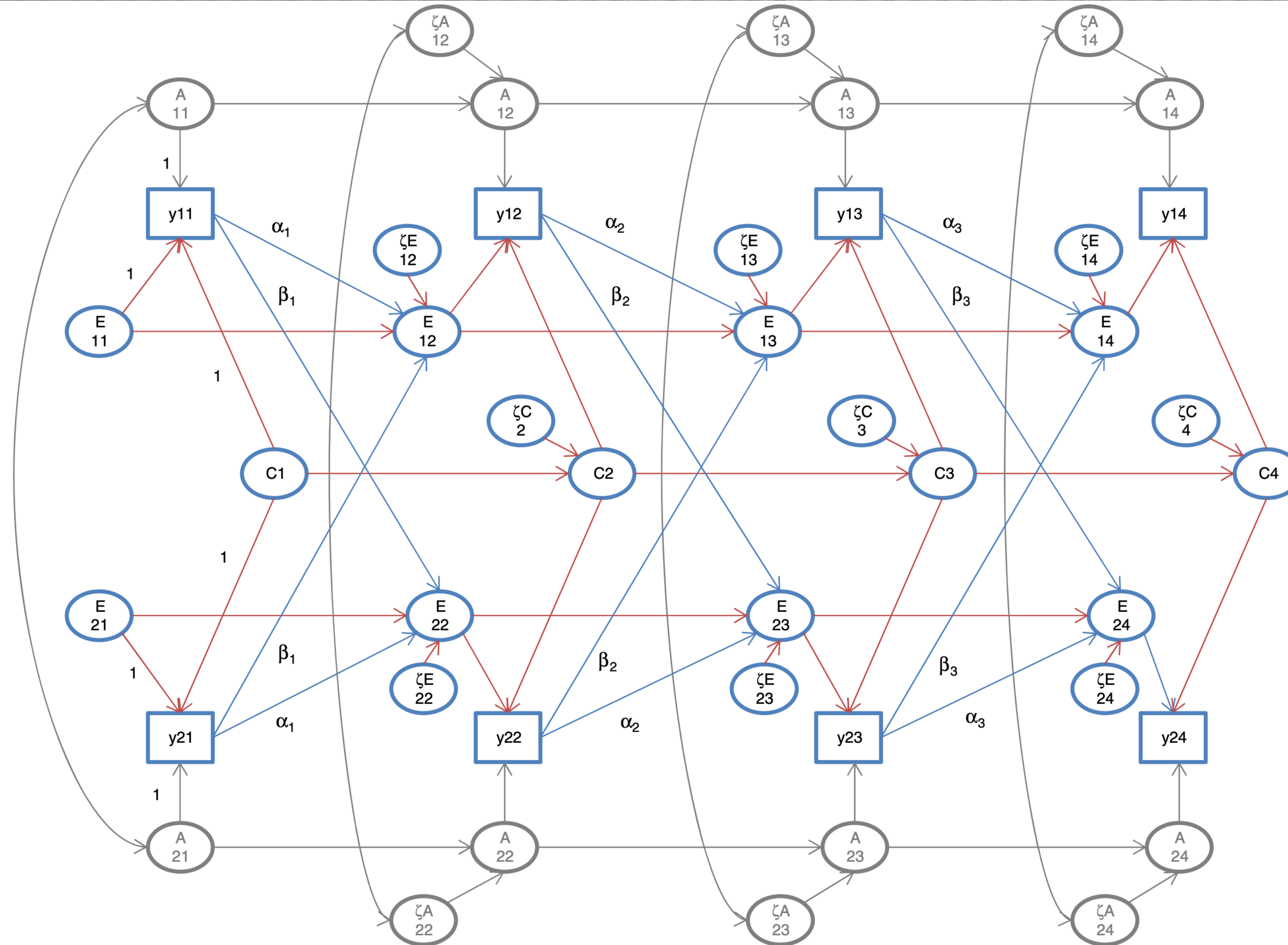


Fig. 2 The extended ACE simplex model. Occasion-specific influences are not shown. The scaling used is shown only at $t = 1$. The extension comprises the arrows from the phenotype y^* at t to the E

variables at $t + 1$ (i.e., parameters α_k and β_k). For the distinction between y (Fig. 1) and y^* in this Figure, see the text

Extended ACE Simplex: P- \rightarrow E

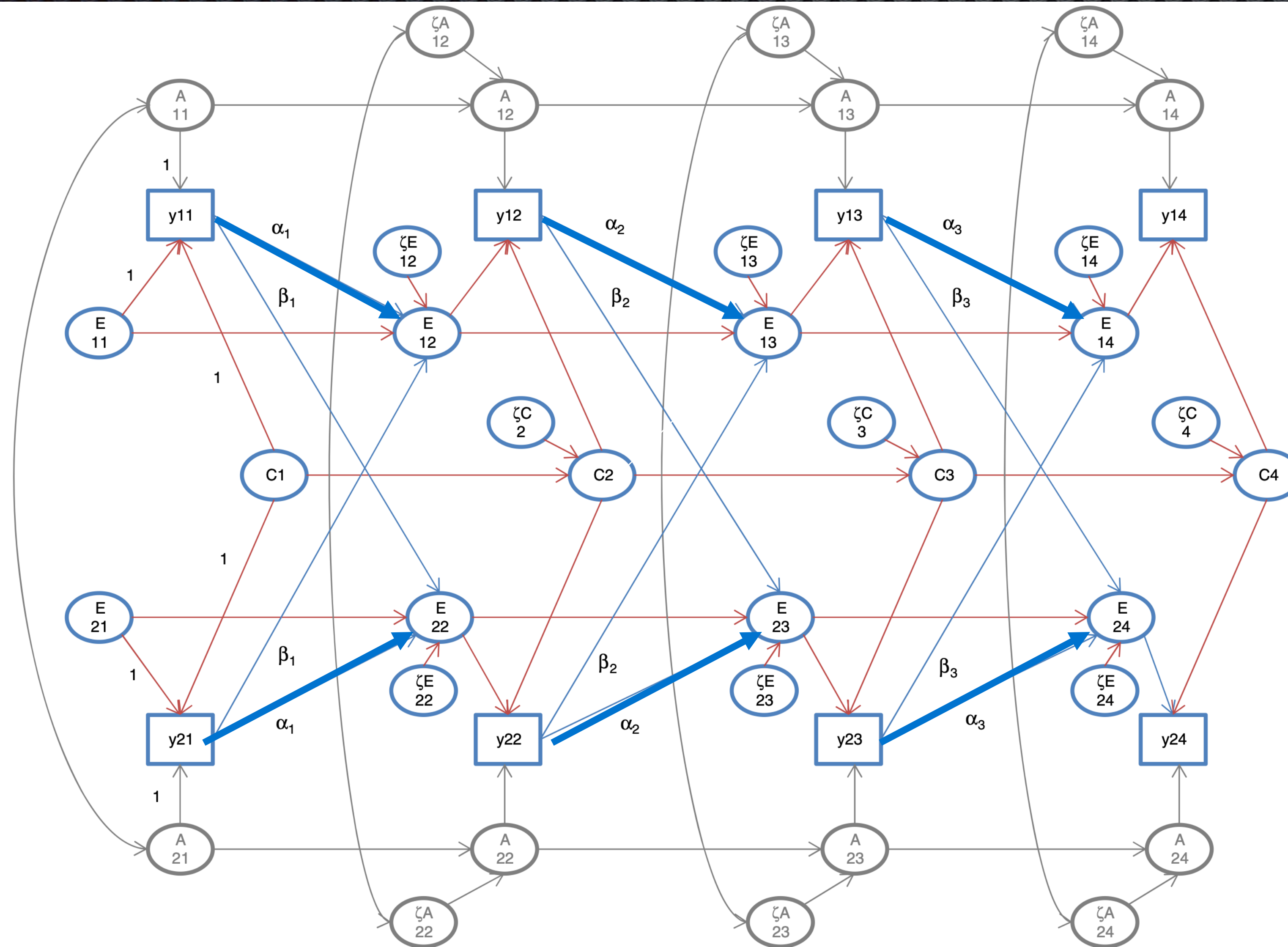


Fig. 2 The extended ACE simplex model. Occasion-specific influences are not shown. The scaling used is shown only at $t = 1$. The extension comprises the arrows from the phenotype y^* at t to the E

variables at $t + 1$ (i.e., parameters α_k and β_k). For the distinction between y (Fig. 1) and y^* in this Figure, see the text

Sibling Effects PT1->ET2 & PT2 -> ET1

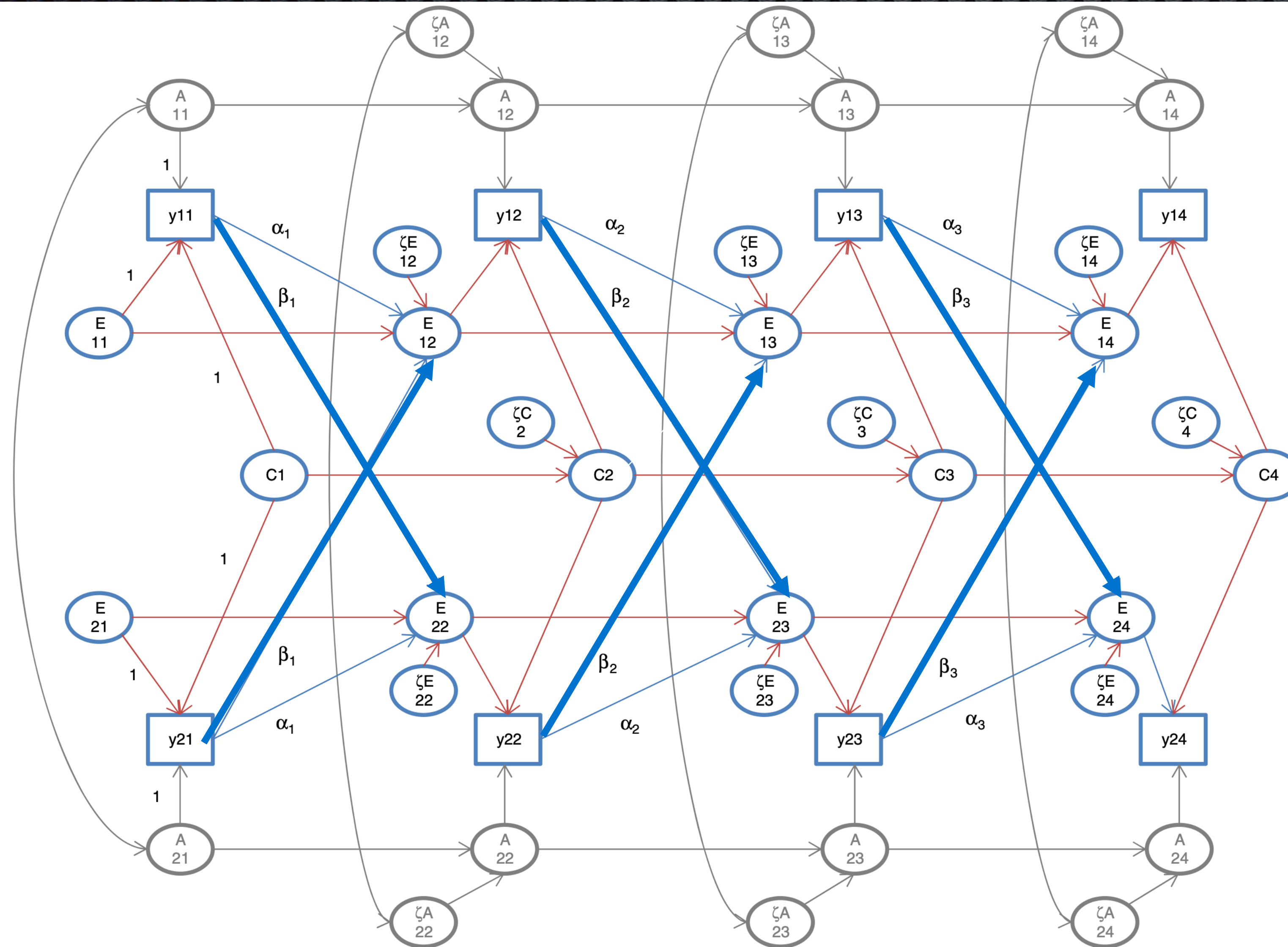


Fig. 2 The extended ACE simplex model. Occasion-specific influences are not shown. The scaling used is shown only at $t = 1$. The extension comprises the arrows from the phenotype y^* at t to the E

variables at $t + 1$ (i.e., parameters α_k and β_k). For the distinction between y (Fig. 1) and y^* in this Figure, see the text

Conclusion

- Simplex/Markov Models very simple in principle
- Usually only first-order (no Time1 \rightarrow Time3 except via Time2)
- Readily extend to latent variable Markov series
- Usefully implemented with twin/sibling data to test niche selection and sibling interaction hypotheses

Future directions

- Including Polygenic Scores in Time Series Models
- Mendelian Randomization
- Multivariate/Network Applications