Longitudinal Modeling

Nathan, Lindon & Mike

LongitudinalTwinAnalysis_MatrixRawCon.R

GenEpiHelperFunctions.R

jepq.txt

Why run longitudinal models?

Estimate time-dependent genetic and environmental effects

- changes in the magnitude of genetic & environmental influence across time
- same versus different genes across development
- identify factors driving change versus factors maintaining stability

Improve power by using multiple observations from the same individual and the cross twin cross trait correlations

Common methods for longitudinal data analyses

Cholesky Decomposition

- 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
  - No predictions
  - Feasible for limited number of measurements

Latent Growth Curve Modeling

Simplex Modeling

Presentation layout

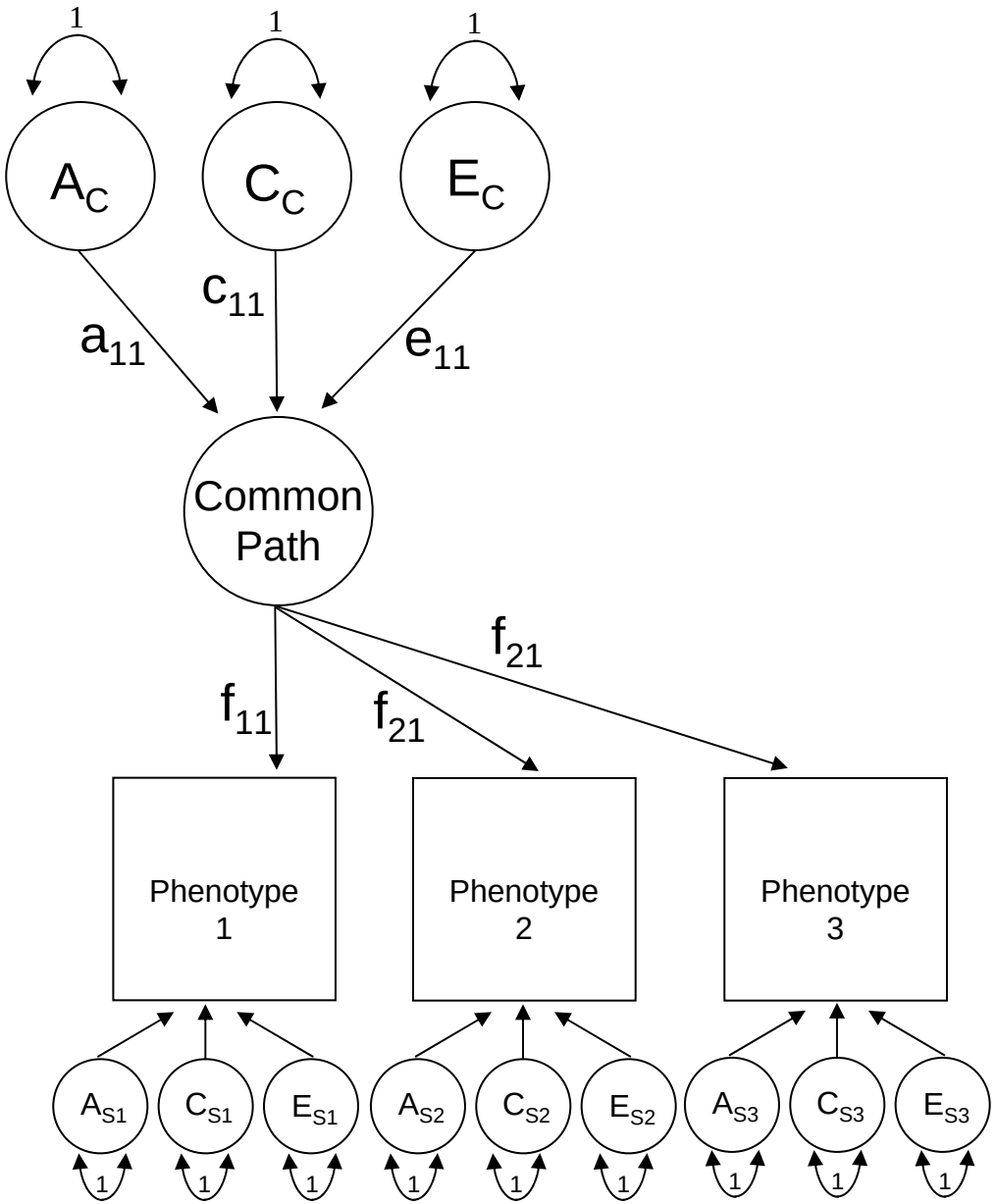
Recap common pathway model

Latent Growth Models

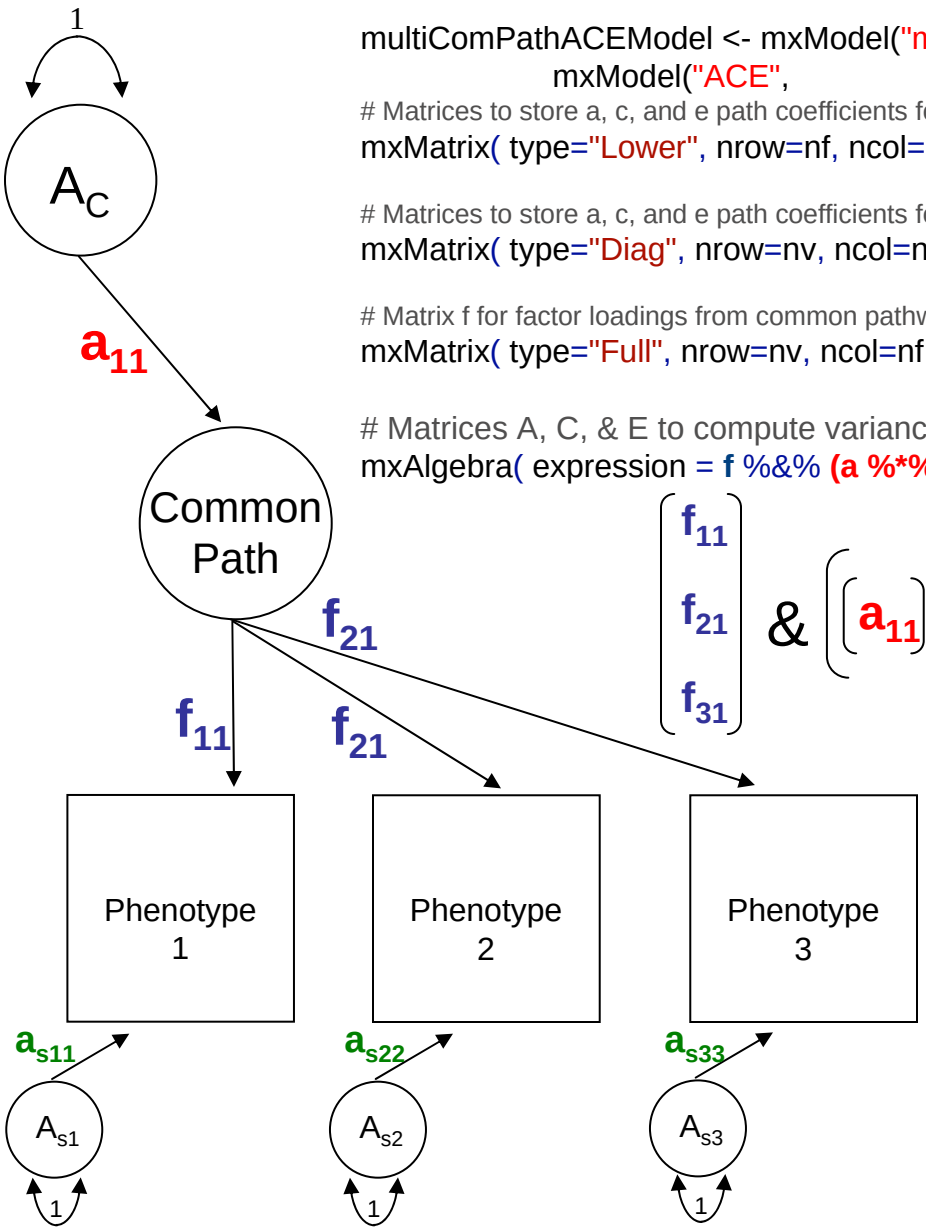
Simplex Models

Lindon’s esoteric input

Common Pathway



Common Pathway: Genetic components of variance




```
multiComPathACEModel <- mxModel("multiComPathACE",
                    mxModel("ACE",
# Matrices to store a, c, and e path coefficients for latent phenotype(s)
mxMatrix( type="Lower", nrow=nf, ncol=nf, free=TRUE, values=.6, name="a" ),

# Matrices to store a, c, and e path coefficients for specific factors
mxMatrix( type="Diag", nrow=nv, ncol=nv, free=TRUE, values=4, name="as" ),

# Matrix f for factor loadings from common pathway to observerd phenotypes
mxMatrix( type="Full", nrow=nv, ncol=nf, free=TRUE, values=15, name="f" ),

# Matrices A, C, & E to compute variance components
mxAlgebra( expression = f %&% (a %*% t(a)) + as %*% t(as), name="A" ),
```

$$\begin{pmatrix} f_{11} \\ f_{21} \\ f_{31} \end{pmatrix} \& \left(\begin{pmatrix} a_{11} \end{pmatrix} X \begin{pmatrix} a_{11} \end{pmatrix}' \right) + \begin{pmatrix} a_{s11} & & \\ & a_{s22} & \\ & & a_{s33} \end{pmatrix} X \begin{pmatrix} a_{s11} & & \\ & a_{s22} & \\ & & a_{s33} \end{pmatrix}' = A$$

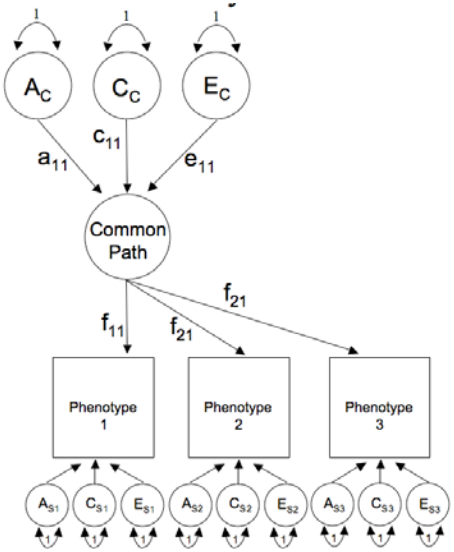
Common Pathway: Matrix algebra + variance components

```
multiComPathACEModel <- mxModel("multiComPathACE",
                mxModel("ACE",
# Matrices to store a, c, and e path coefficients for latent phenotype(s)
mxMatrix( type="Lower", nrow=nf, ncol=nf, free=TRUE, values=.6, name="a" ),
mxMatrix( type="Lower", nrow=nf, ncol=nf, free=TRUE, values=.6, name="c" ),
mxMatrix( type="Lower", nrow=nf, ncol=nf, free=TRUE, values=.6, name="e" ),

# Matrices to store a, c, and e path coefficients for specific factors
mxMatrix( type="Diag", nrow=nv, ncol=nv, free=TRUE, values=4, name="as" ),
mxMatrix( type="Diag", nrow=nv, ncol=nv, free=TRUE, values=4, name="cs" ),
mxMatrix( type="Diag", nrow=nv, ncol=nv, free=TRUE, values=4, name="es" ),

# Matrix f for factor loadings from common pathway to observerd phenotypes
mxMatrix( type="Full", nrow=nv, ncol=nf, free=TRUE, values=15, name="f" ),

# Matrices A, C, & E to compute variance components
mxAlgebra( expression = f %&% (a %*% t(a)) + as %*% t(as), name="A" ),
mxAlgebra( expression = f %&% (c %*% t(c)) + cs %*% t(cs), name="C" ),
mxAlgebra( expression = f %&% (e %*% t(e)) + es %*% t(es), name="E" ),
```

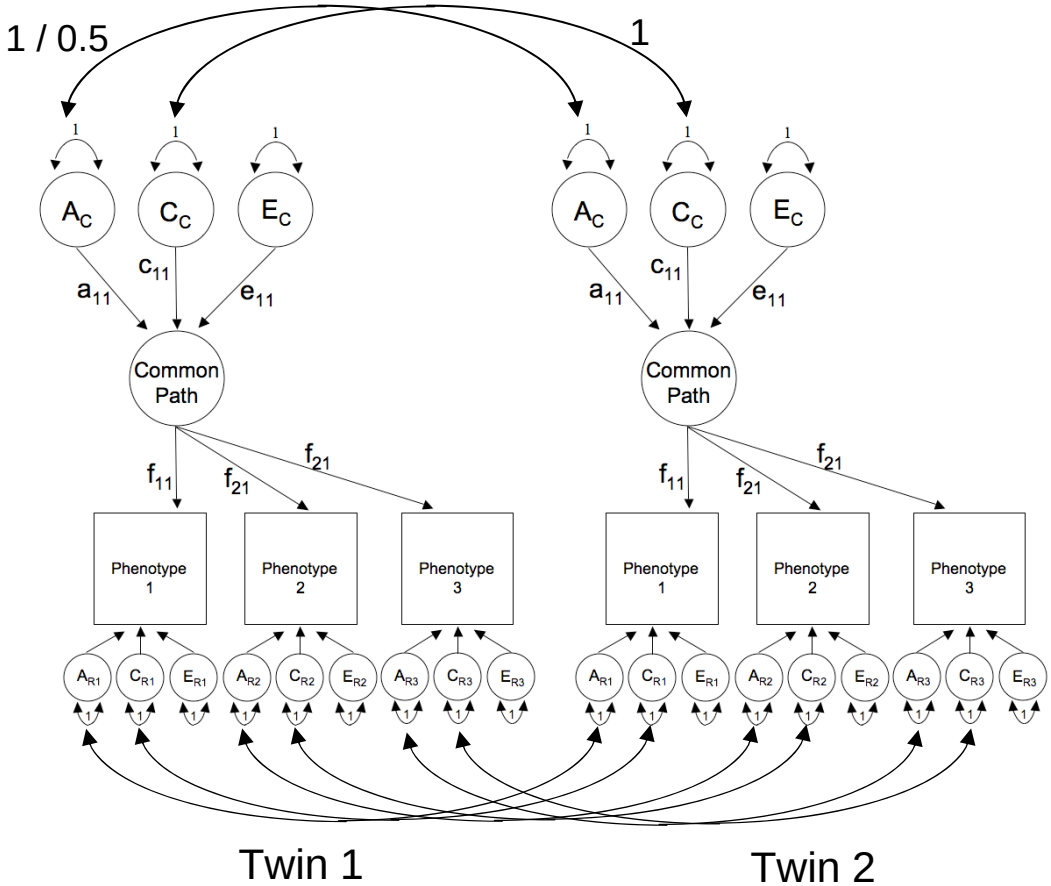


Within twin (co)variance

	T_1	T_2
T_1	A+C+E	
T_2		A+C+E

CP Model: Expected covariance

```
# Algebra for expected variance/covariance matrix in MZ
mxAlgebra( expression= rbind  ( cbind(A+C+E , A+C),
                                cbind(A+C   , A+C+E)), name="expCovMZ" ),
# Algebra for expected variance/covariance matrix in DZ
mxAlgebra( expression= rbind  ( cbind(A+C+E ,   0.5%x%A+C),
                                cbind(0.5%x%A+C ,A+C+E) ),  name="expCovDZ" )
```



MZ	T_1	T_2
T_1	A+C+E	A+C
T_2	A+C	A+C+E

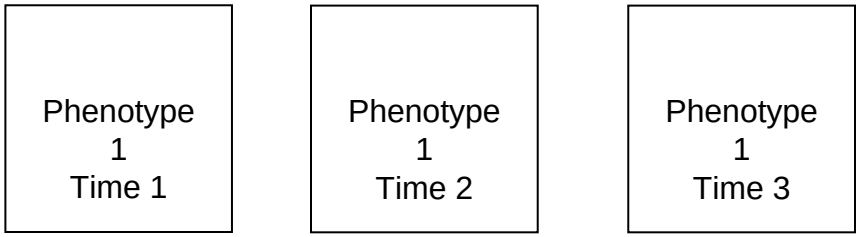
DZ	T_1	T_2
T_1	A+C+E	0.5@A+C
T_2	0.5@A+C	A+C+E

Got longitudinal data?

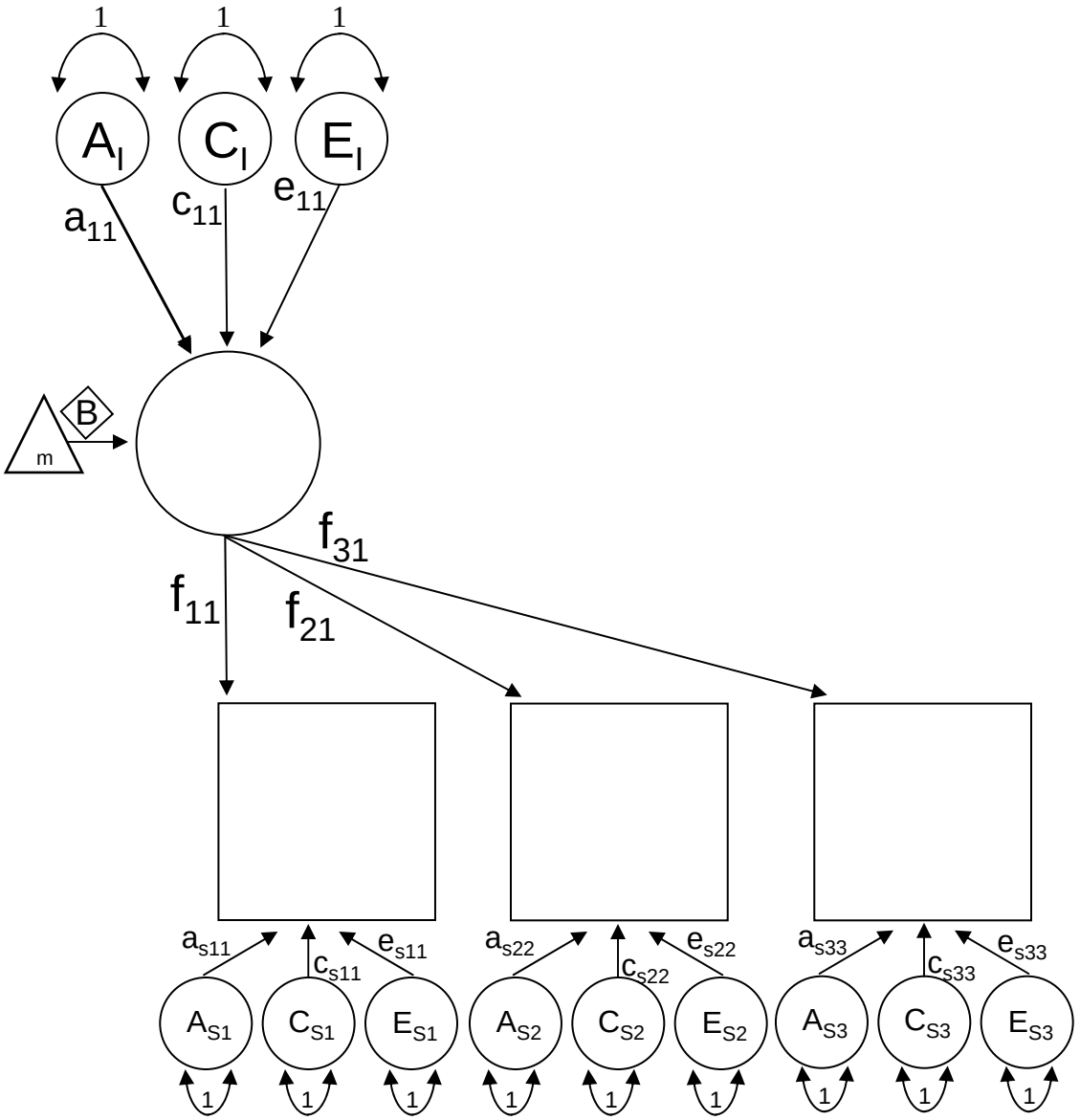
How do variance components change over time?

Are they stable?

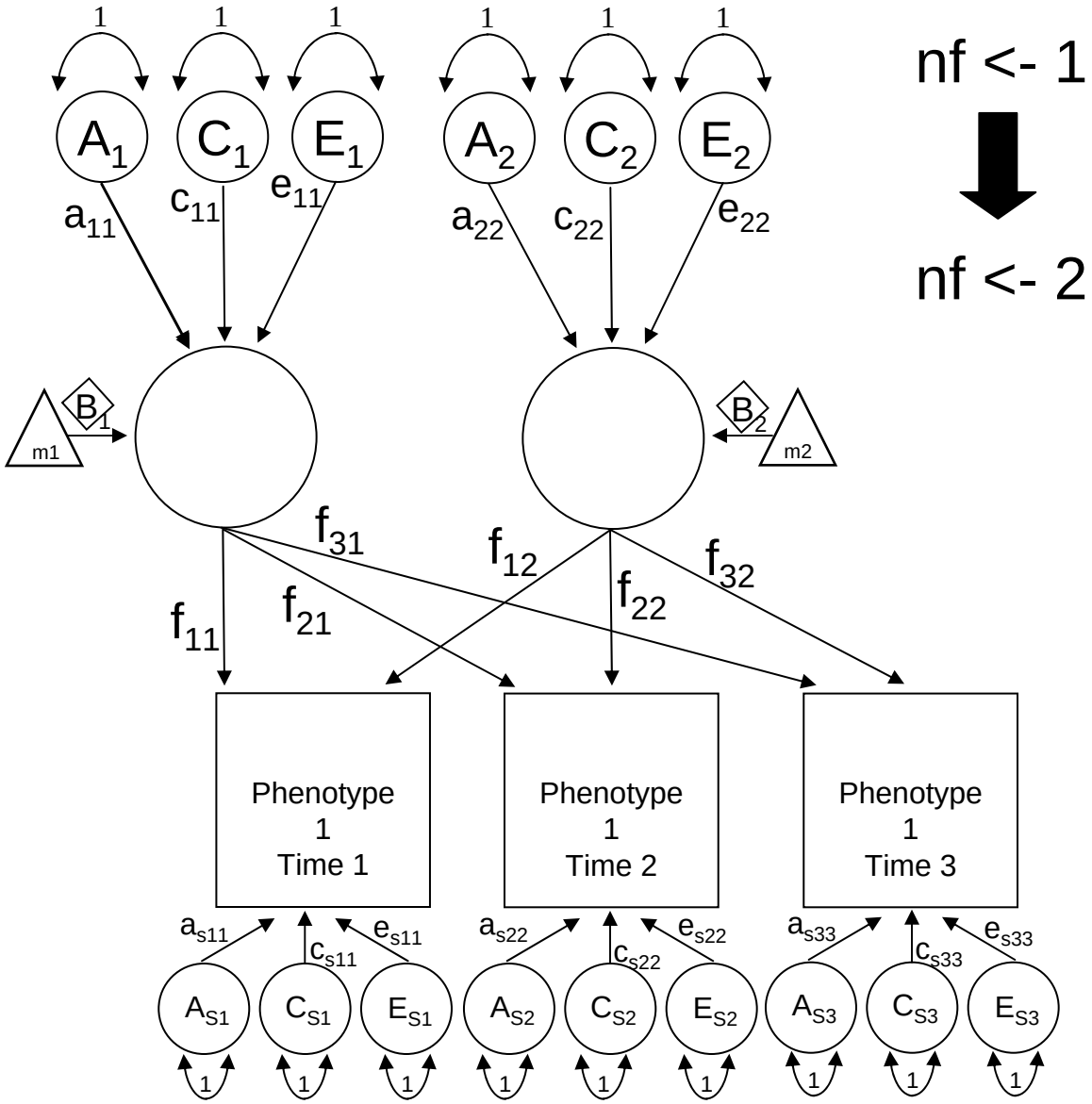
How to best explain change? Linear, non-linear?



Common Pathway Model



CP to Latent Growth Curve Model



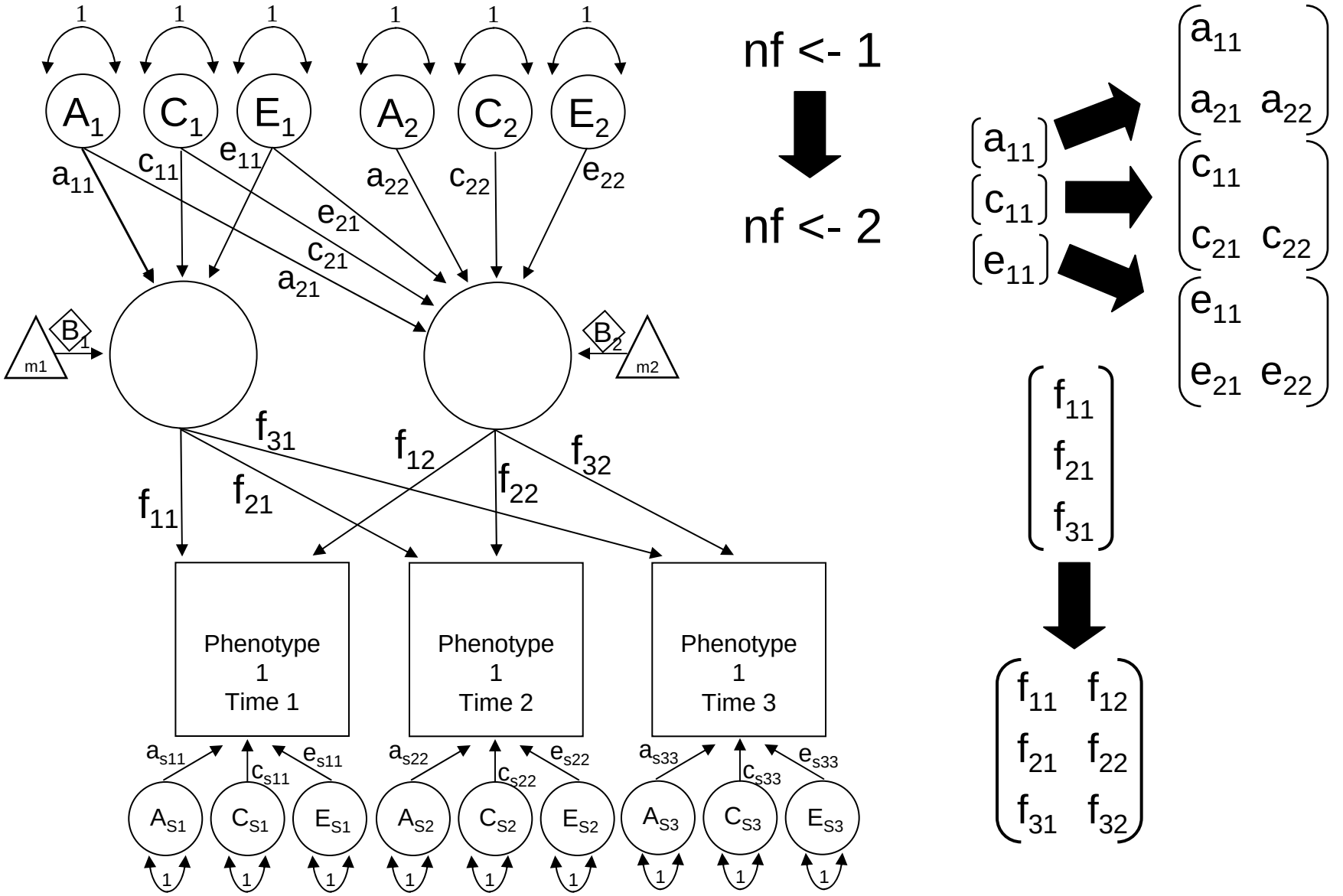
nf <- 1

nf <- 2

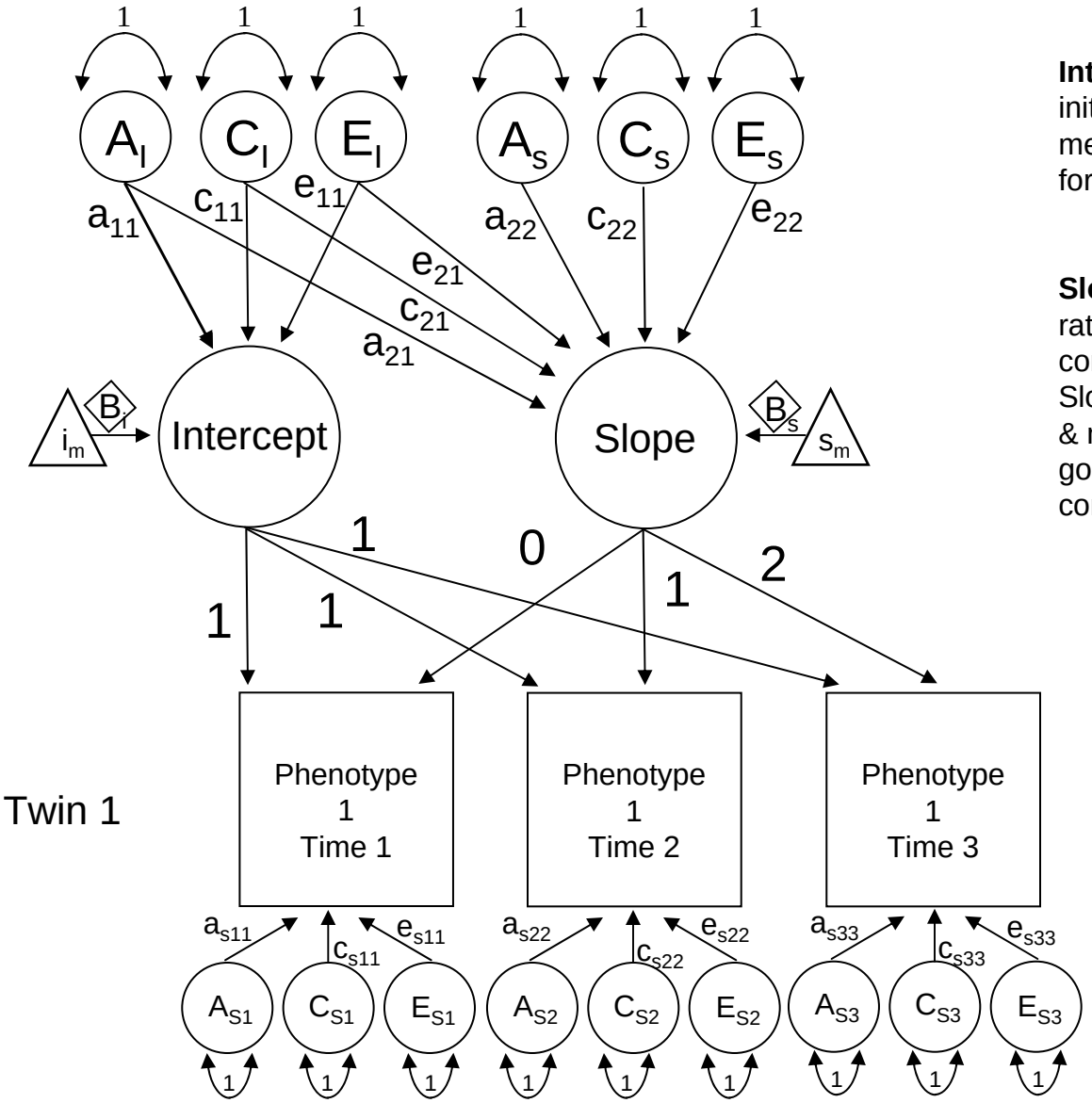
$$\begin{pmatrix} a_{11} \\ c_{11} \\ e_{11} \end{pmatrix} \rightarrow \begin{pmatrix} a_{11} & \\ & a_{22} \\ c_{11} & \\ & c_{22} \\ e_{11} & \\ & e_{22} \end{pmatrix}$$

$$\begin{pmatrix} f_{11} \\ f_{21} \\ f_{31} \end{pmatrix} \rightarrow \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \\ f_{31} & f_{32} \end{pmatrix}$$

CP to Latent Growth Curve Model



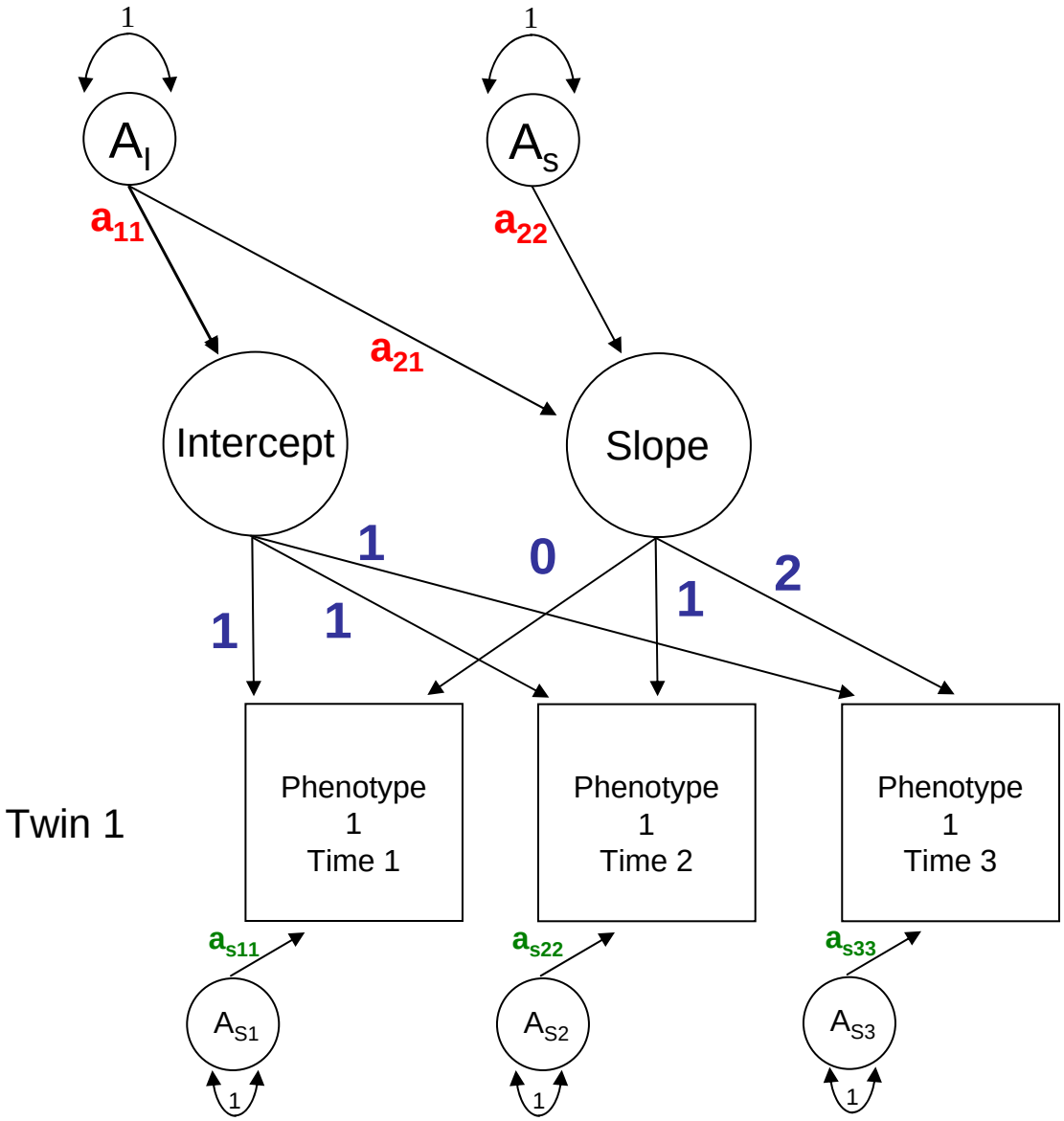
Latent Growth Curve Model



Intercept: Factor which explains initial variance components (and mean) for all measures. Accounts for the stability over time.

Slope: Factor which influences the rate of change in the variance components (and mean) over time. Slope(s) is (are) pre-defined: linear & non linear (quadratic, logistic, gompertz etc) hence factor loading constraints required.

LGC Model: Within twin genetic components of variance



$$\begin{pmatrix} a_{11} & \\ a_{21} & a_{22} \end{pmatrix}$$

Genetic pathway coefficients matrix

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$$

Factor loading matrix

$$\begin{pmatrix} a_{S11} & & \\ & a_{S22} & \\ & & a_{S33} \end{pmatrix}$$

Residual genetic pathway coefficients matrix

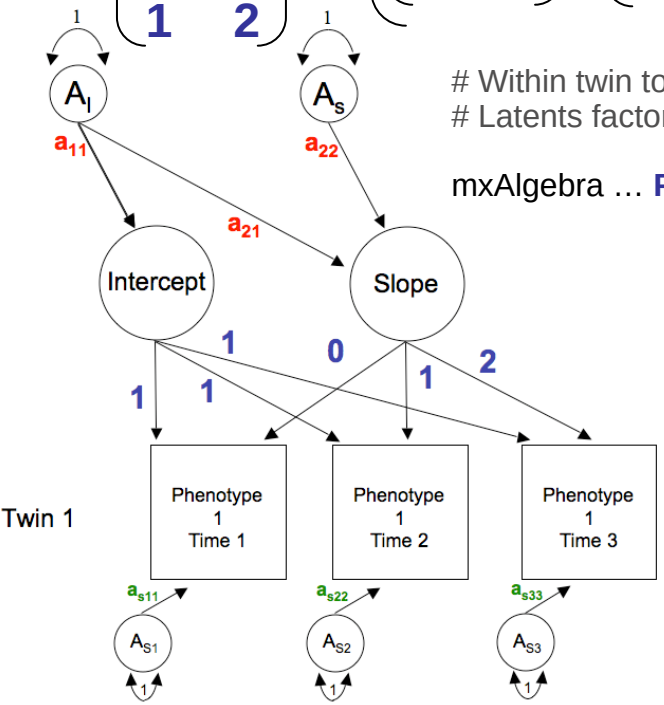
LGC Model: Specifying variance components in R

```
lgcACEModel <- mxModel("lgcACE",
# Matrix for a path coefficients from latent factors to Int' & Slope latent factors
mxMatrix( type="Lower", nrow=nf, ncol=nf, free=T, values=0.2, name="a" ),

# Matrix for a path coefficients from residuals to observed phenotypes
mxMatrix( type="Diag",  nrow=nv, ncol=nv, free=T, values=0.2, name="as"

# Factor loading matrix of Int & Slop on observed phenotypes
mxMatrix( type="Full", nrow=nv, ncol=nf, free=F, values=c(1,1,1,0,1,2), name="F" ),
```

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \& \left(\begin{pmatrix} a_{11} & \\ a_{21} & a_{22} \end{pmatrix} X \begin{pmatrix} a_{11} & a_{21} \\ & a_{22} \end{pmatrix}' + \begin{pmatrix} a_{S11} & & \\ & a_{S22} & \\ & & a_{S33} \end{pmatrix} X \begin{pmatrix} a_{S11} & & \\ & a_{S22} & \\ & & a_{S33} \end{pmatrix}' \right) = A$$




```
# Within twin total genetic  (co) variance components
# Latents factors + residuals

mxAlgebra … F %&% (a %*% t(a)) + as %*% t(as), name="A" ),
```

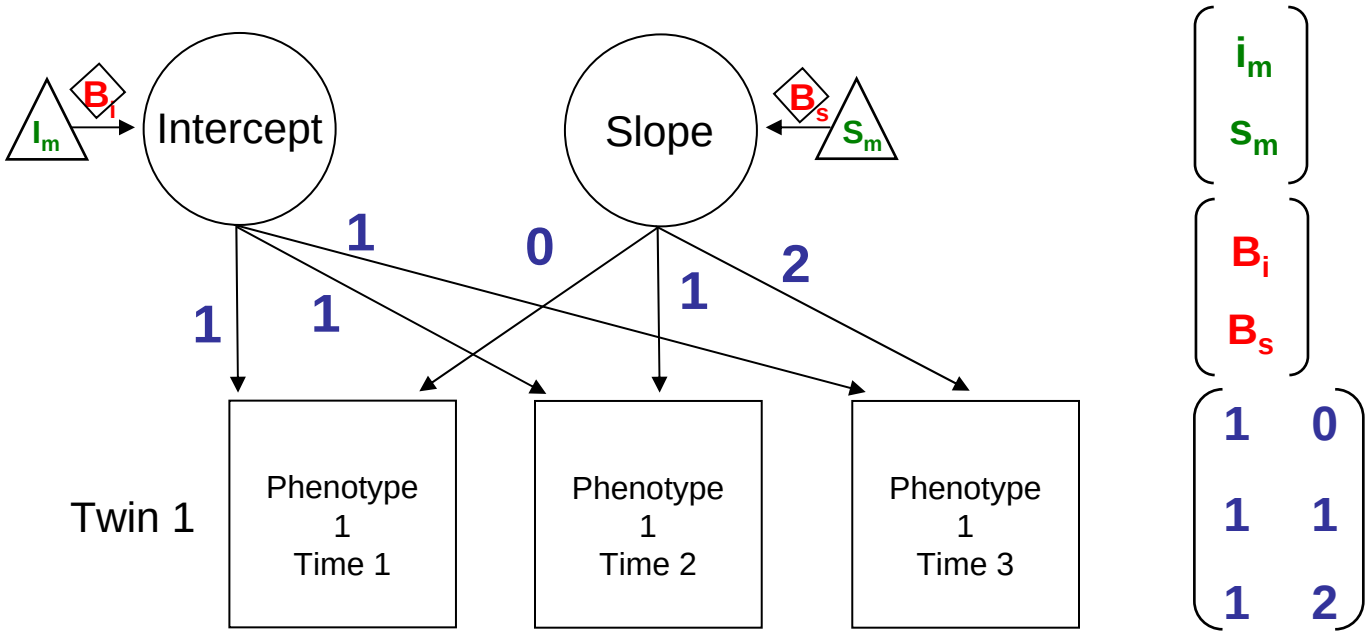
$$a_{11}^2 + a_{S11}^2 = A_{\text{var time 1}}$$

$$a_{11}^2 + a_{21}a_{11} + a_{21}a_{11} + a_{22}^2 + a_{S22}^2 = A_{\text{var time 2}}$$

$$a_{11}^2 + 2a_{21}a_{11} + 2a_{21}a_{11} + 2a_{22}^2 + a_{S33}^2 = A_{\text{var time 3}}$$

LGC Model: Means & sex in R

Means on observed phenotypes versus means on Intercept & Slope?



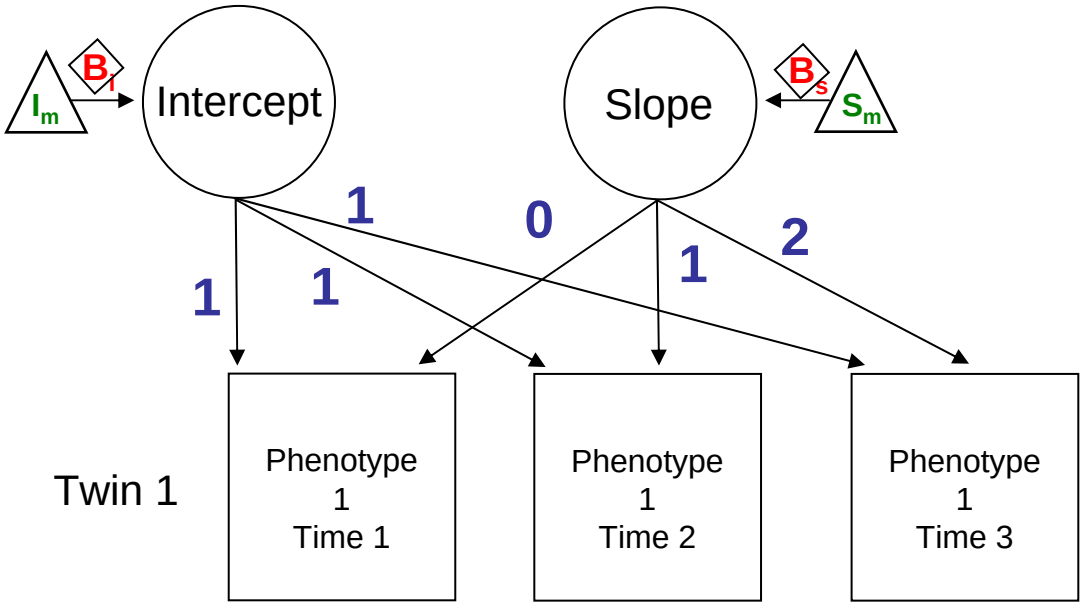

```
# Means for Intercept and Slope
# mxMatrix( type="Full", nrow=1, ncol=nv, free=TRUE, values= 80, name="Mean" ),
mxMatrix(type="Full", nrow=2, ncol=1, free=T, values=c(1,0.1), labels=c("Im","Sm"), name="Mean"),
# Betas / Sex effects on Int & Slope means
mxMatrix(type="Full", nrow=2, ncol=1, free=T, values=0.4,labels=c("Bi","Bs"), name="Beta" ),
# Factor loading matrix of Intercept and Slope on observed phenotypes
mxMatrix( type="Full", nrow=nv, ncol=2, free=F, values=c(1,1,1,0,1,2), name="F" ),
```

LGC Model: Sex on the Means Algebra

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \times \left(\begin{pmatrix} i_m \\ s_m \end{pmatrix} + \left(\text{Sex}_{T1} @ \begin{pmatrix} B_i \\ B_s \end{pmatrix} \right) \right) = \text{Expected means for Twin 1}$$

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \times \begin{pmatrix} i_m + B_i \text{Sex}_{T1} \\ s_m + B_s \text{Sex}_{T1} \end{pmatrix} = \begin{pmatrix} (i_m + B_i \text{Sex}_{T1}) \\ (i_m + B_i \text{Sex}_{T1}) + 1(s_m + B_s \text{Sex}_{T1}) \\ (i_m + B_i \text{Sex}_{T1}) + 2(s_m + B_s \text{Sex}_{T1}) \end{pmatrix} \begin{matrix} \text{Time 1} \\ \text{Time 2} \\ \text{Time 3} \end{matrix}$$

```
mxAlgebra ... (F %*% (Mean + (sex_t1 %x% Beta))) ...  name = "expMean"),
```



LGC Model: LongitudinalTwinAnalysis_MatrixRawCon.R

1. Inspect proportions of variance explained by A, C & E
2. Standardize pathway coefficients

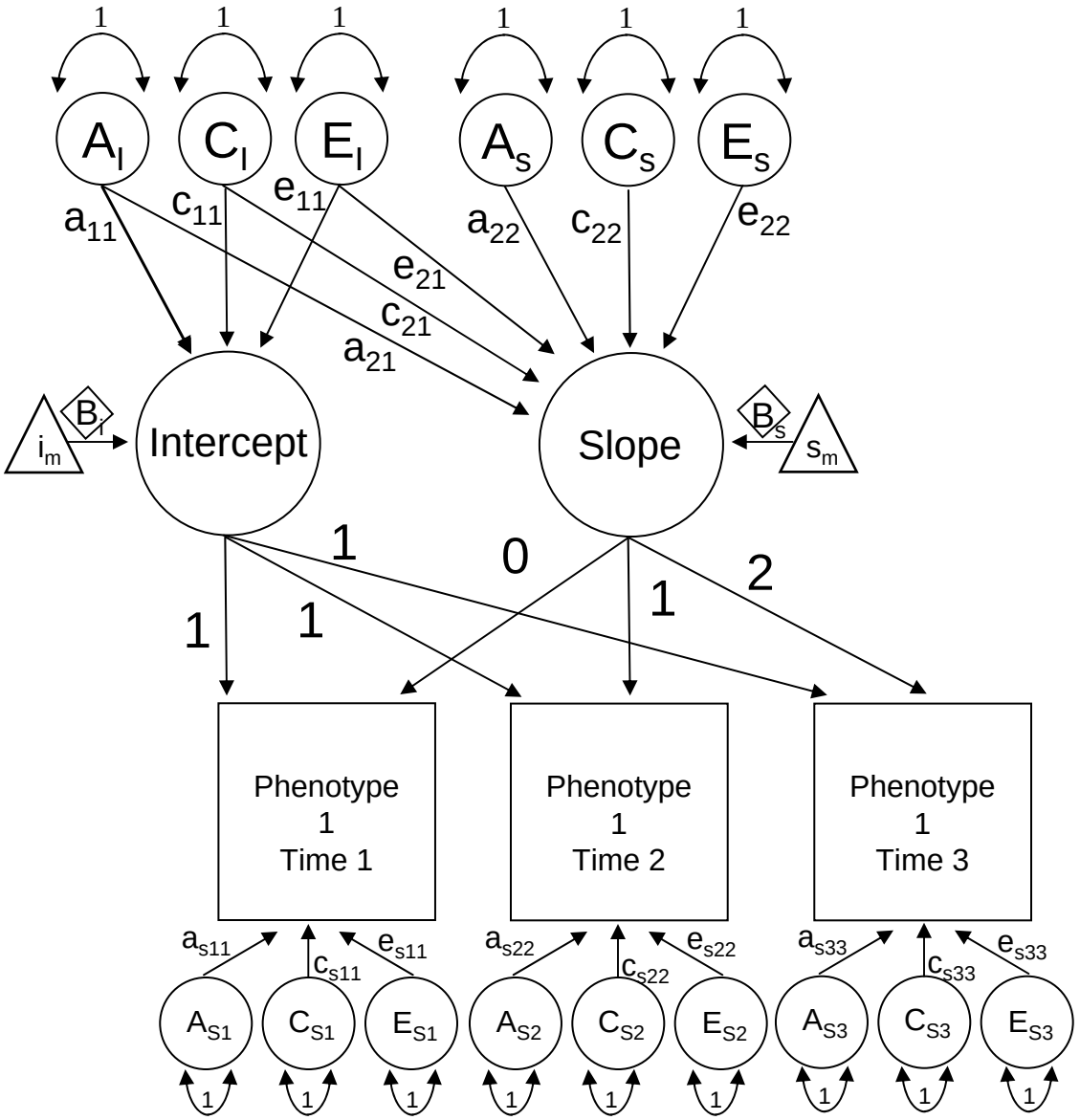
3. **Practical**: Fit nested models (AE, CE & E) & compare to ACE
4. **Practical**: Change factor loadings to -1, 0, & 1 on slope & compare fit to ACE
5. **Practical**: Calculate genetic factor correlations between Intercept & Slope

6. Estimate the genetic factor correlation under the first model i.e. F=c(1,1,1,0,1,2)

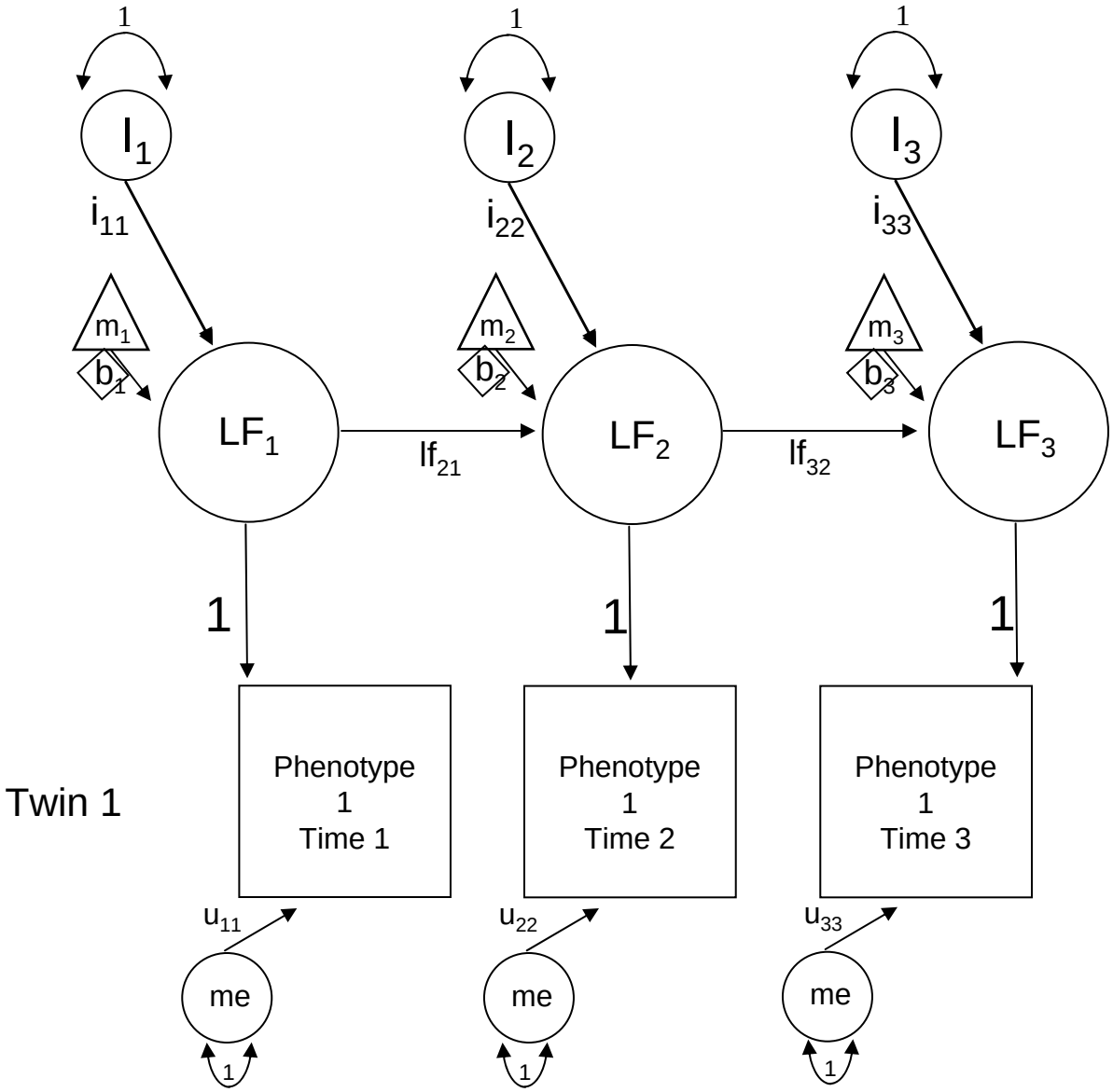
7. **Practical**: Estimate the genetic factor correlation under the 2nd model i.e. F=c(1,1,1,-1,0,1)

8. **Practical** 3: Drop C specific effect on Slope & compare to full model

Latent Growth Curve Model



Simplex Models



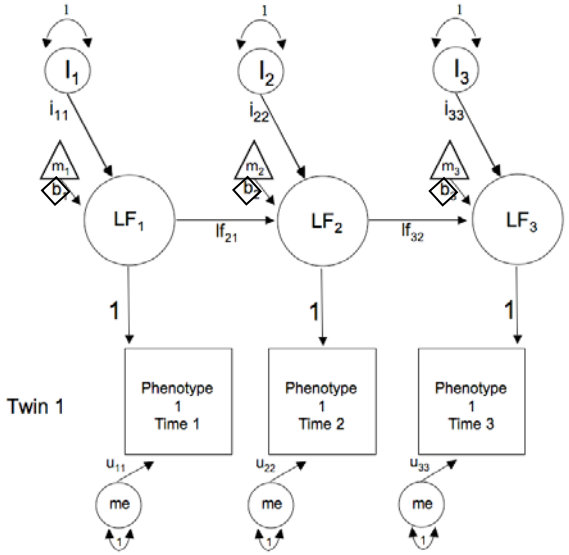
Simplex Models

Simplex designs model changes in the latent factor structure over time by fitting auto-regressive or Markovian chains

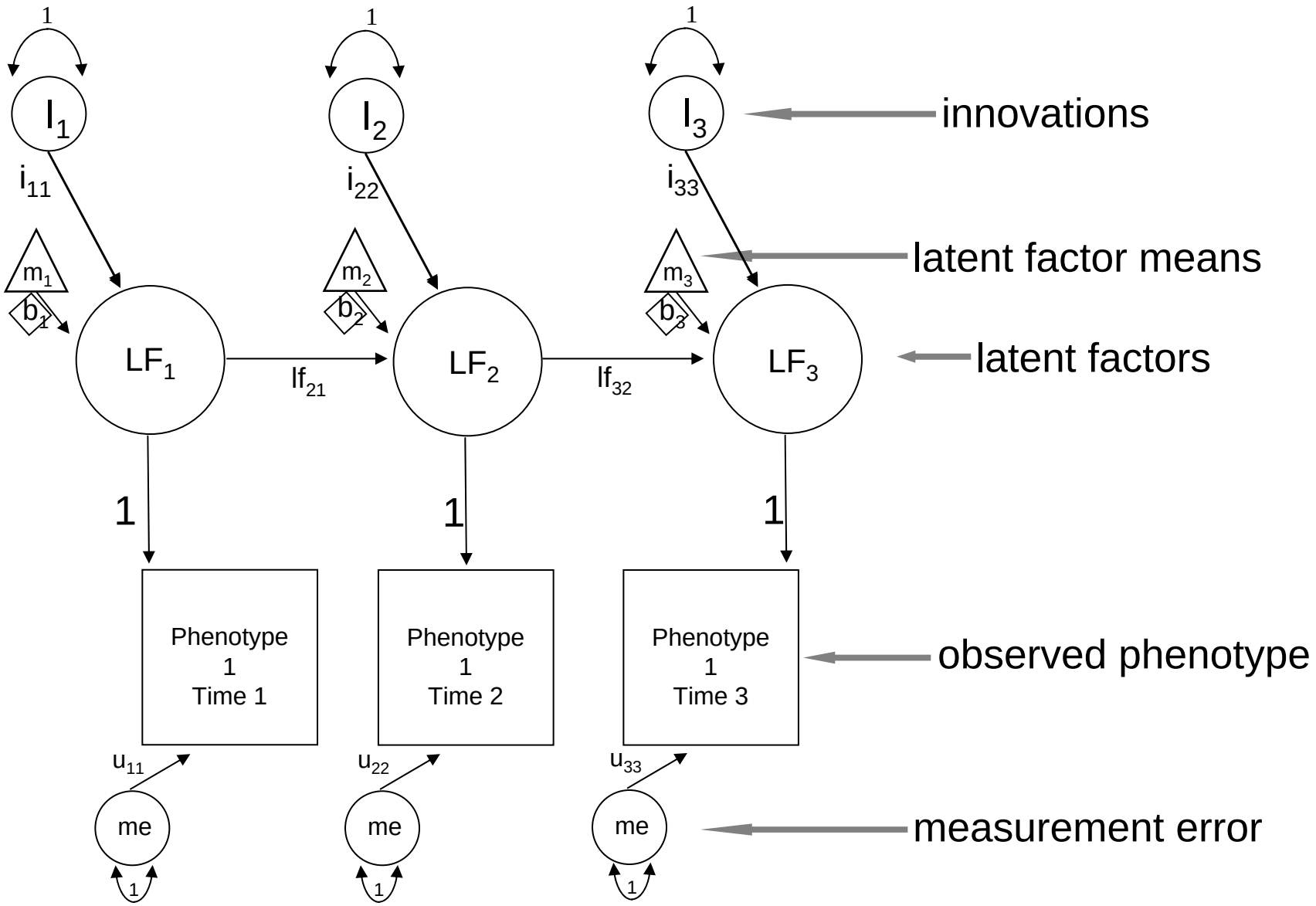
Determine how much variation in a trait is caused by stable & enduring effects versus transient effects unique to each time

The chief advantage of this model is the ability to partition environmental & genetic variation at each time point into:

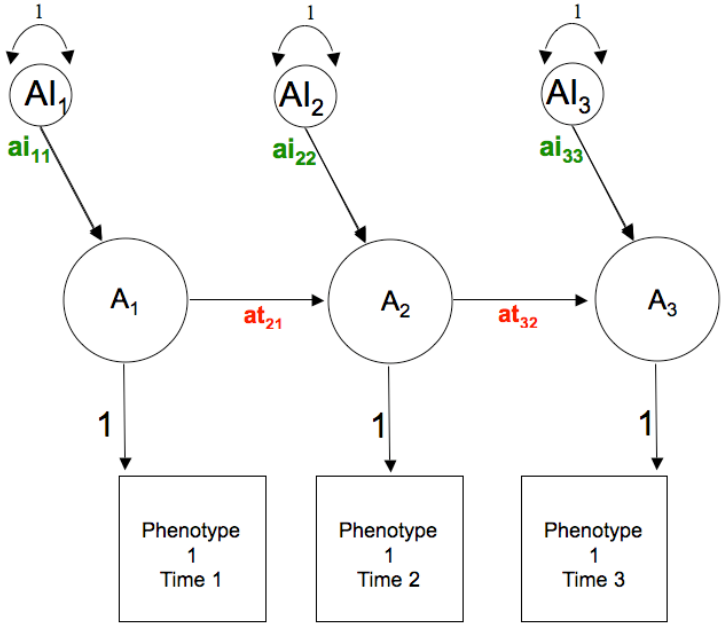
- genetic & environmental effects unique to each occasion
- genetic and environmental effects transmitted from previous time points



Simplex Models



Simplex Models: Within twin genetic (co)variance



$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

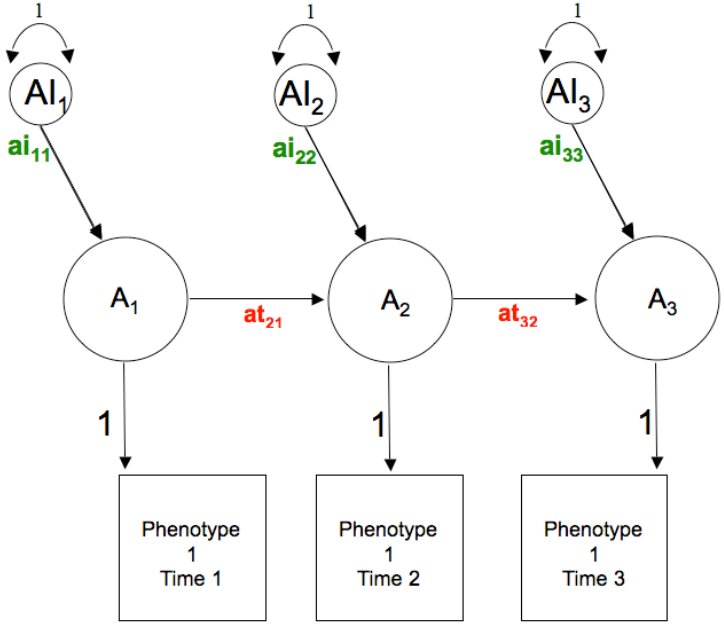
$$\begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \end{array} \begin{array}{c} \begin{array}{ccc} A_1 & A_2 & A_3 \end{array} \\ \begin{pmatrix} ? & & \\ ? & ? & \\ ? & ? & ? \end{pmatrix} \end{array}$$

Transmission pathways

$$\begin{pmatrix} ai_{11} & & \\ & ai_{22} & \\ & & ai_{33} \end{pmatrix}$$

Innovation pathways

Simplex Models: Within twin genetic (co)variance



$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \end{array} \begin{array}{c} \begin{array}{ccc} A_1 & A_2 & A_3 \end{array} \\ \begin{pmatrix} 0 & & \\ at_{21} & 0 & \\ 0 & at_{32} & 0 \end{pmatrix} \end{array}$$

Transmission pathways

$$\begin{pmatrix} ai_{11} & & \\ & ai_{22} & \\ & & ai_{33} \end{pmatrix}$$

Innovation pathways

$$\left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & & \\ at_{21} & 0 & \\ 0 & at_{32} & 0 \end{pmatrix} \right)^{-1} \& \begin{pmatrix} ai_{11} & & \\ & ai_{22} & \\ & & ai_{33} \end{pmatrix} * \begin{pmatrix} ai_{11} & & \\ & ai_{22} & \\ & & ai_{33} \end{pmatrix}' = A$$

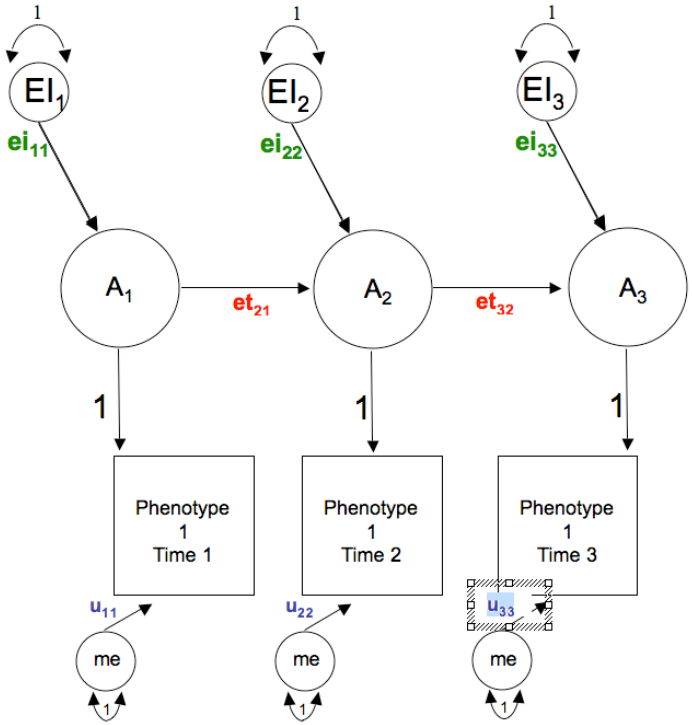
```
SimplexACEModel <- mxModel("SimplexACE",
mxModel("ACE",
mxMatrix( type="Iden", nrow=nv, ncol=nv, name="I"),
mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(F,T,F,F,T,F), name="at" ), # Transmissions
mxMatrix( type="Diag", nrow=nv, ncol=nv, free=TRUE, name="ai" ), # Innovations

mxAlgebra( expression=solve(I-at) %&% (ai %*% t(ai)), name="A" ), # A variance component
```

Simplex Models: E (co)variance + measurement error

```
SimplexACEModel <- mxModel("SimplexACE",
mxModel("ACE",
mxMatrix( type="Iden", nrow=nv, ncol=nv, name="I"),
mxMatrix( type="Lower", nrow=nv, ncol=nv, free=c(F,T,F,F,T,F), name="et" ), # Transmissions
mxMatrix( type="Diag", nrow=nv, ncol=nv, free=TRUE, name="ei" ), # Innovations
mxMatrix( type="Diag", nrow=nv, ncol=nv, free=TRUE, labels=c("u","u","u"), name="me" ),
mxAlgebra( expression=solve(I-et) %&% (ei %*% t(ei)) + (me %*% t(me)), name="E" ),# E var
```

$$\left(\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} - \begin{pmatrix}0&&\\et_{21}&0&\\0&et_{32}&0\end{pmatrix}\right)^{\sim} \& \begin{pmatrix}ei_{11}&&\\&ei_{22}&\\&&ei_{33}\end{pmatrix} * \begin{pmatrix}ei_{11}&&\\&ei_{22}&\\&&ei_{33}\end{pmatrix}' + \begin{pmatrix}u_{11}&&\\&u_{22}&\\&&u_{33}\end{pmatrix} * \begin{pmatrix}u_{11}&&\\&u_{22}&\\&&u_{33}\end{pmatrix}' = E$$

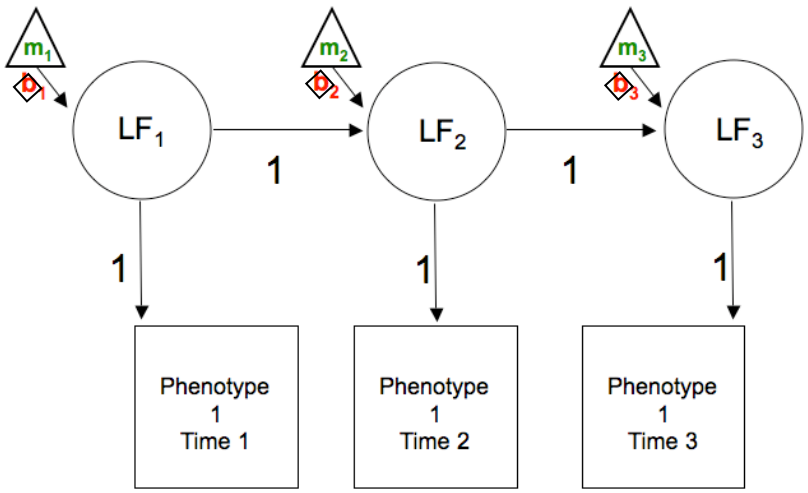


Simplex Models: Means & sex in R

```
# Estimate latent variable means
mxMatrix( type="Lower", nrow=nv, ncol=nv, free=FALSE, values=1, name="T"),
mxMatrix( type="Full", nrow=nv, ncol=1, free=TRUE, labels=c("m1","m2","m3"), name="m" ),
mxAlgebra( expression= t(T %*% m), name="Means" ),
```

$$\left(\begin{bmatrix} 1 & & \\ 1 & 1 & \\ 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix}\right)' = \begin{array}{ccc} \text{Time 1} & \text{Time 2} & \text{Time 3} \\ \begin{pmatrix} m_1 & m_1+m_2 & m_1+m_2+m_3 \end{pmatrix} \end{array}$$

```
mxMatrix( type="Full", nrow=1, ncol=nv, free=TRUE, labels=c("b1","b2","b3"), values=c(1,1,1), name="Beta" ),
mxModel("MZ",
 mxData(data.frame(mzData,mzDefs), type="raw" ),
 mxMatrix( type="Full", nrow=1, ncol=1, free=FALSE, labels=c("data.sex_1"), name="sex_t1"),
 mxMatrix( type="Full", nrow=1, ncol=1, free=FALSE, labels=c("data.sex_2"), name="sex_t2"),
 mxAlgebra(expression=cbind(   (ACE.Means + (sex_t1 %x% ACE.Beta)),
                               (ACE.Means + (sex_t2 %x% ACE.Beta)) ), name="expMean"),
```



$$\begin{pmatrix} Sex_{T1} \end{pmatrix} @ \begin{pmatrix} b_1 & b_2 & b_3 \end{pmatrix} = \begin{pmatrix} b_1 Sex_{T1} & b_2 Sex_{T1} & b_3 Sex_{T1} \end{pmatrix}$$

$$\begin{array}{ccc} \text{Time 1} & \text{Time 2} & \text{Time 3} \\ \begin{pmatrix} m_1 + b_1 Sex_{T1} & m_1 + m_2 + b_2 Sex_{T1} & m_1 + m_2 + m_3 + b_3 Sex_{T1} \end{pmatrix} \end{array}$$

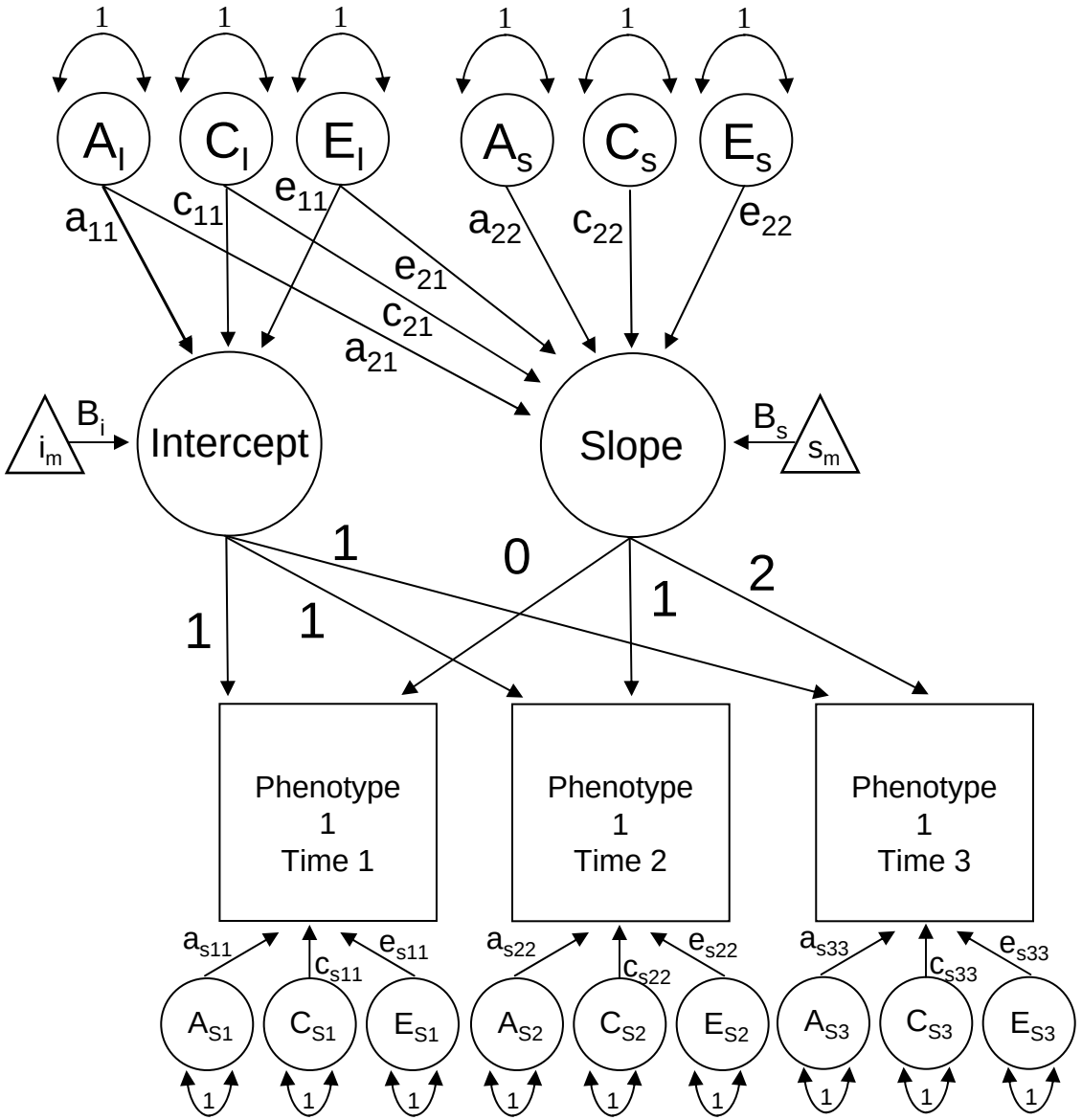
Simplex Model: LongitudinalTwinAnalysis_MatrixRawCon.R

1. Derive standardized pathway transmission coefficients
2. Derive standardized innovation pathway coefficients
3. Illustrate mean changes over time

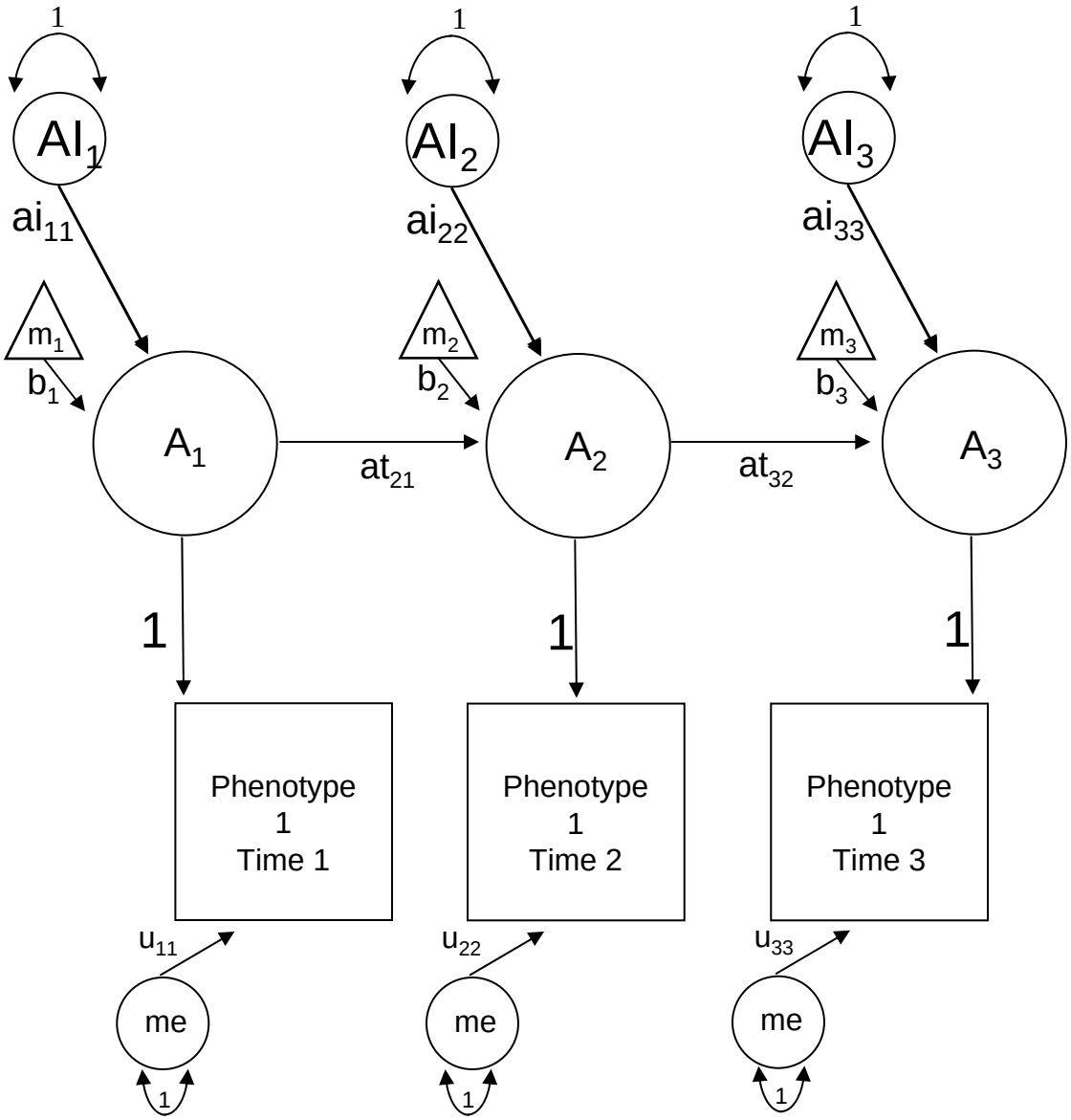
4. **Practical**: Fit nested models (AE, CE & E) & compare to ACE
5. **Practical**: Estimate genetic & environmental correlations between phenotypes

6. Demonstrate how to drop ‘C’ innovations at Time 2 & Time 3 & compare to ACE

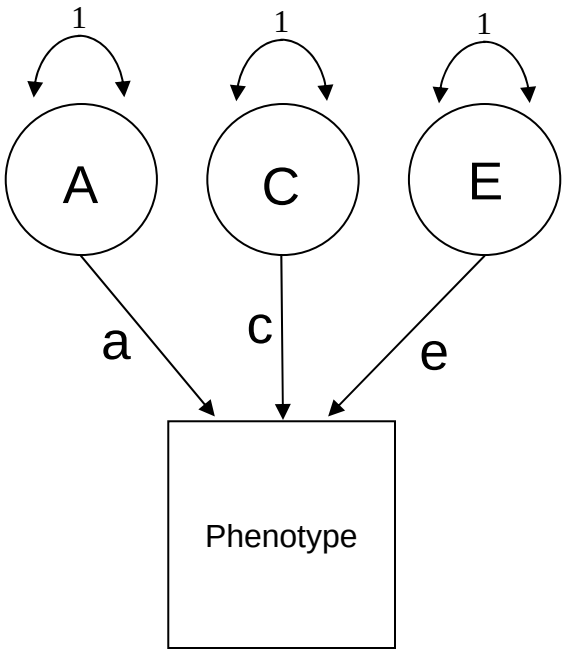
Latent Growth Curve Model

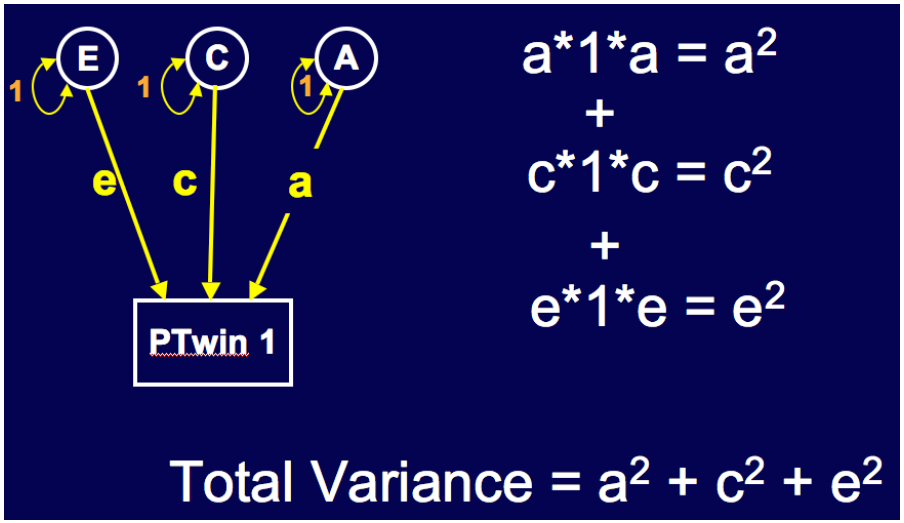


Simplex Models



Phenotypic variance decomposition

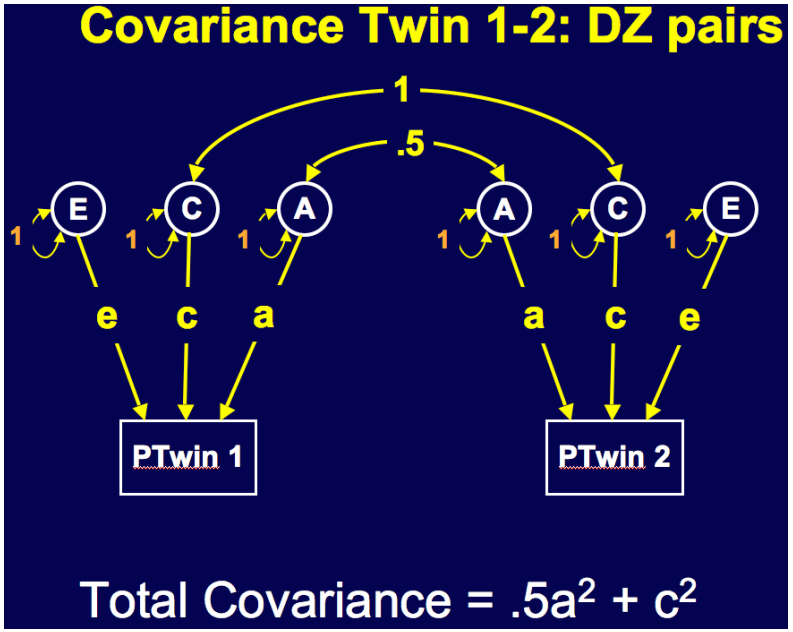




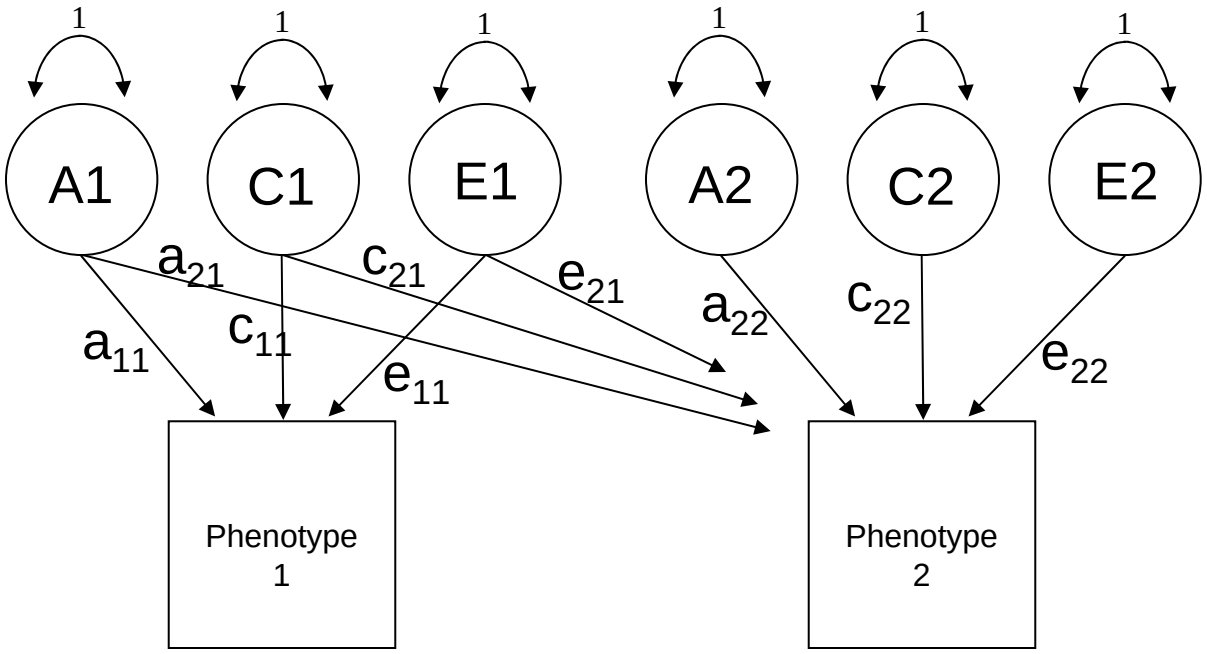
Univariate analysis: Expected within twin & cross-twin (co)variance

MZ	T_1	T_2
T_1	A+C+E	A+C
T_2	A+C	A+C+E

DZ	T_1	T_2
T_1	A+C+E	0.5@A+C
T_2	0.5@A+C	A+C+E



Bivariate & multivariate analyses (Cholesky decompositions)

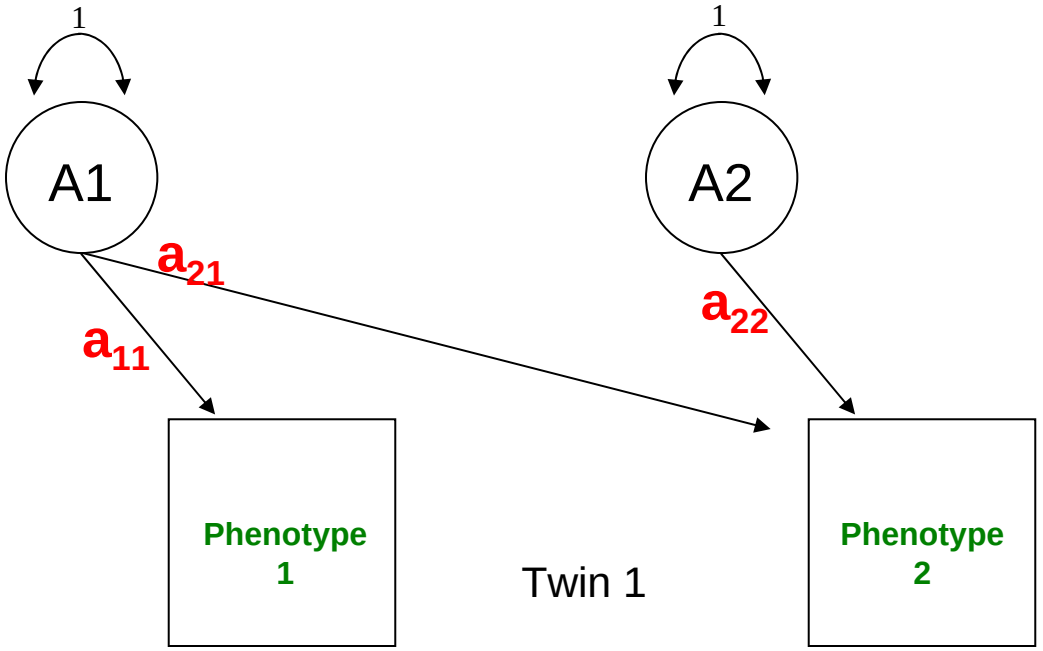


Twin 1

Calculating within twin genetic (co)variance matrix

```
twinACEModel <- mxModel("ACE",
mxMatrix( type="Lower", nrow=nv, ncol=nv, free=TRUE, values=.6, name="a"),
mxAlgebra( expression=a %*% t(a), name="A" ), #Total genetic variance
```

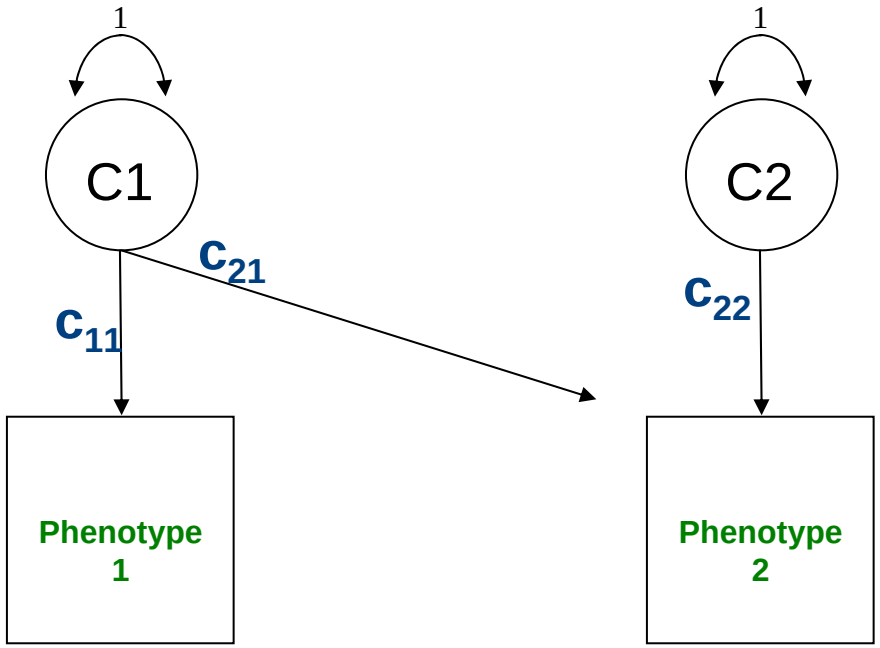
$$\begin{array}{c} \\ \text{Pheno1} \\ \text{Pheno2} \end{array}\begin{array}{c} \begin{array}{cc} \text{A1} & \text{A2} \end{array} \\ \begin{pmatrix} a_{11} & \\ a_{21} & a_{22} \end{pmatrix} \end{array} \times \begin{pmatrix} a_{11} & a_{21} \\ & a_{22} \end{pmatrix} = \begin{array}{c} \\ \text{Pheno1} \\ \text{Pheno2} \end{array}\begin{array}{c} \begin{array}{cc} \text{Pheno1} & \text{Pheno2} \end{array} \\ \begin{pmatrix} a_{11}^2 & a_{11}a_{21} \\ a_{11}a_{21} & a_{22}^2 \end{pmatrix} \end{array} = A$$



Calculating within twin (co)variance matrices

```
twinACEModel <- mxModel("ACE",
mxMatrix( type="Lower", nrow=nv, ncol=nv, free=TRUE, values=.6, name="c"),
mxAlgebra( expression=a %*% t(a), name="C" ), #Total genetic variance
```

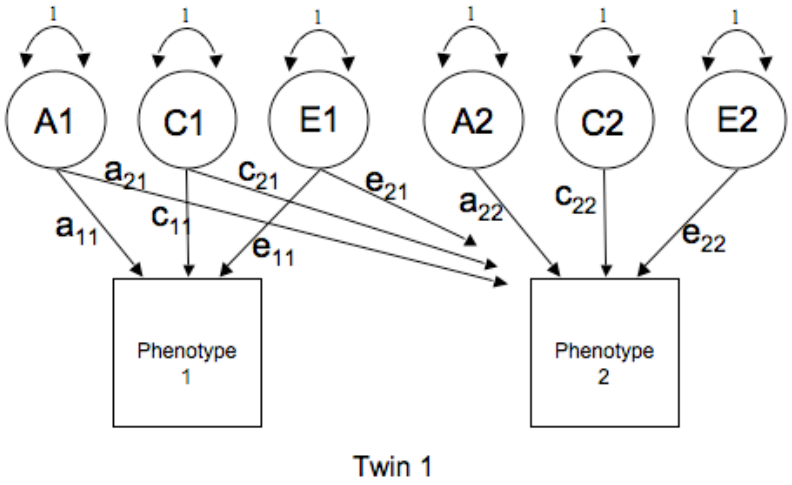
$$\begin{array}{c} \\ \text{Pheno1} \\ \text{Pheno2} \end{array}\begin{array}{c} \begin{array}{cc} \text{C1} & \text{C2} \end{array} \\ \begin{pmatrix} c_{11} & \\ c_{21} & c_{22} \end{pmatrix} \end{array} \times \begin{pmatrix} c_{11} & c_{21} \\ & c_{22} \end{pmatrix} = \begin{array}{c} \\ \text{Pheno1} \\ \text{Pheno2} \end{array}\begin{array}{c} \begin{array}{cc} \text{Pheno1} & \text{Pheno2} \end{array} \\ \begin{pmatrix} c_{11}^2 & c_{11}c_{21} \\ c_{11}c_{21} & c_{22}^2 \end{pmatrix} \end{array} = C$$



Calculating within twin (co)variance matrix

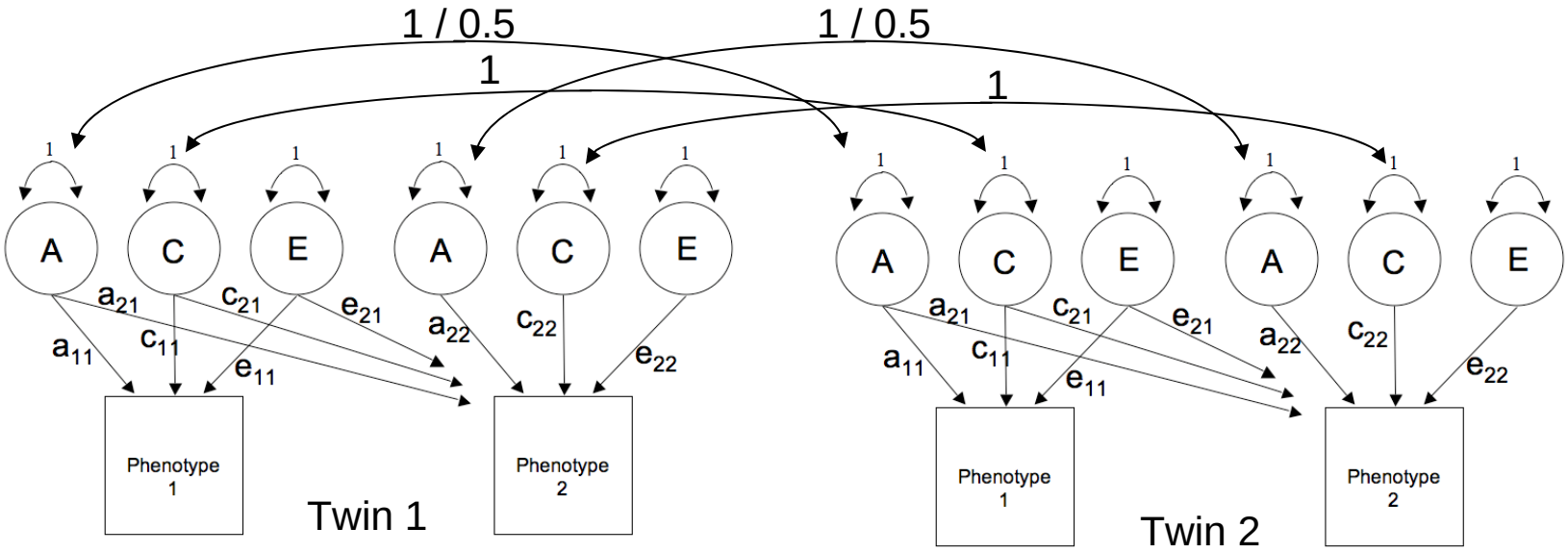
mxAlgebra(expression=A+C+E, name="V"),

	T_1	
T_1	$a_{11}^2 + c_{11}^2 + e_{11}^2$	$a_{11}a_{21} + c_{11}c_{21} + e_{11}e_{21}$
	$a_{11}a_{21} + c_{11}c_{21} + e_{11}e_{21}$	$a_{22}^2 + c_{22}^2 + e_{22}^2$



Bivariate & multivariate analyses

```
# Algebra for expected variance/covariance matrix in MZ
mxAlgebra( expression= rbind ( cbind(A+C+E , A+C),
                               cbind(A+C   , A+C+E)), name="expCovMZ" ),
# Algebra for expected variance/covariance matrix in DZ
mxAlgebra( expression= rbind ( cbind(A+C+E ,  0.5%x%A+C),
                               cbind(0.5%x%A+C , A+C+E) ), name="expCovDZ" )
```



MZ	T_1	T_2
T_1	A+C+E	A+C
T_2	A+C	A+C+E

DZ	T_1	T_2
T_1	A+C+E	$0.5_{@}$A+C
T_2	$0.5_{@}$A+C	A+C+E