

# Multivariate Analysis

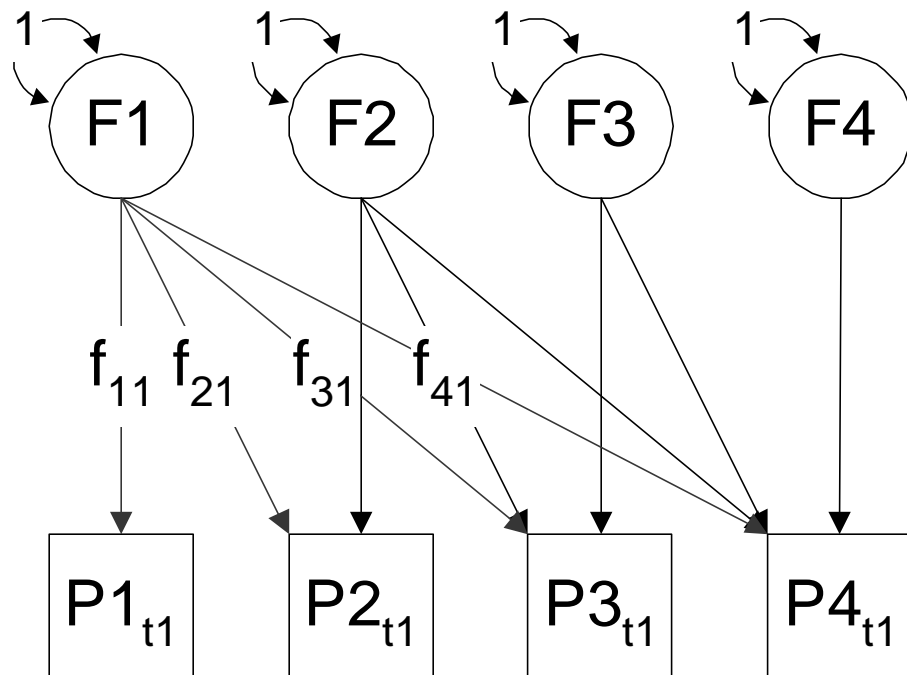
Mx class 2004



# Multivariate Questions I

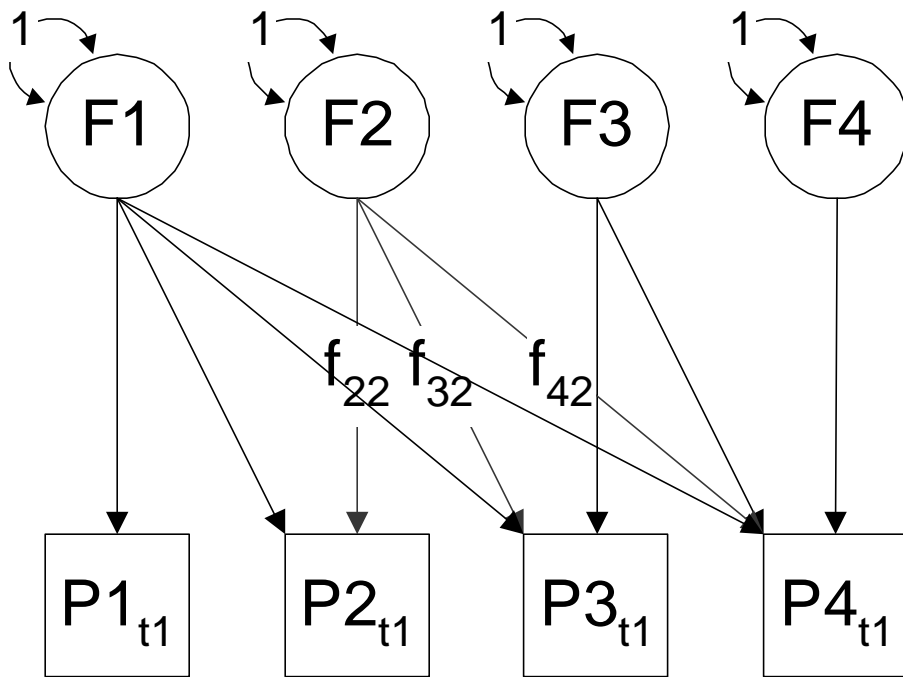
- Bivariate Analysis: What are the contributions of genetic and environmental factors to the covariance between two traits?
- Multivariate Analysis: What are the contributions of genetic and environmental factors to the covariance between more than two traits?

# Phenotypic Cholesky F1



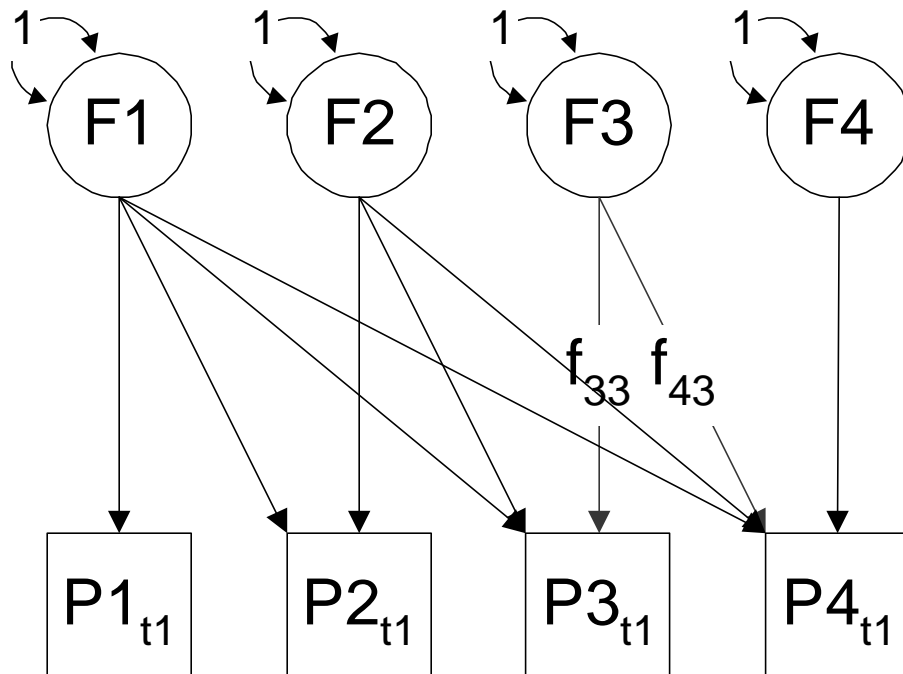
$$\begin{matrix} & F1 & F2 & F3 & F4 \\ P1 & f_{11} & & & \\ P2 & f_{21} & & & \\ P3 & f_{31} & & & \\ P4 & f_{41} & & & \end{matrix}$$

# Phenotypic Cholesky F2



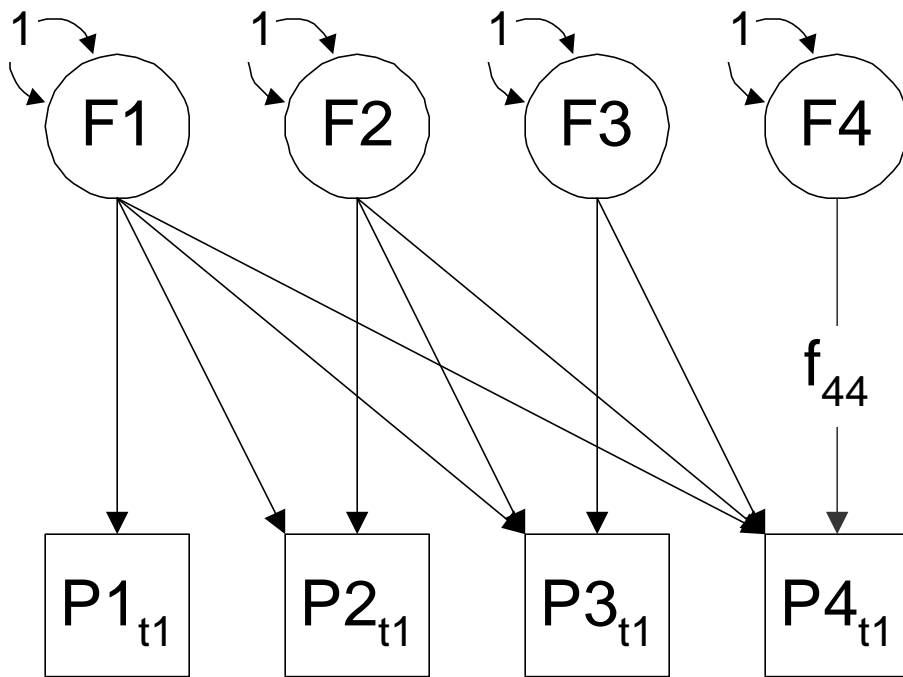
$$\begin{array}{c} \text{P1} \\ \text{P2} \\ \text{P3} \\ \text{P4} \end{array} \begin{bmatrix} \text{F1} & \text{F2} & \text{F3} & \text{F4} \\ f_{11} & 0 & & \\ f_{21} & f_{22} & & \\ f_{31} & f_{32} & & \\ f_{41} & f_{42} & & \end{bmatrix}$$

# Phenotypic Cholesky



$$\begin{array}{c} \text{P1} \\ \text{P2} \\ \text{P3} \\ \text{P4} \end{array} \begin{bmatrix} \text{F1} & \text{F2} & \text{F3} & \text{F4} \\ f_{11} & 0 & 0 & \\ f_{21} & f_{22} & 0 & \\ f_{31} & f_{32} & f_{33} & \\ f_{41} & f_{42} & f_{43} & \end{bmatrix}$$

# Phenotypic Cholesky



$$\begin{array}{c} P1 \\ P2 \\ P3 \\ P4 \end{array} \begin{array}{c} F1 \\ F2 \\ F3 \\ F4 \end{array} \begin{bmatrix} f_{11} & 0 & 0 & 0 \\ f_{21} & f_{22} & 0 & 0 \\ f_{31} & f_{32} & f_{33} & 0 \\ f_{41} & f_{42} & f_{43} & f_{44} \end{bmatrix}$$

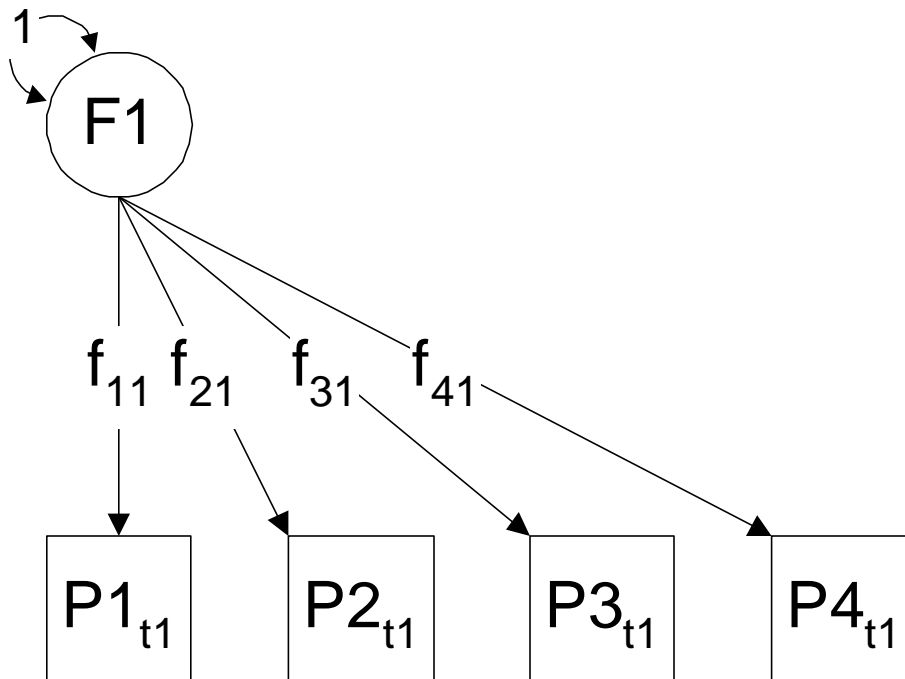
# Cholesky Decomposition

$$\begin{array}{c} \text{P1} \\ \text{P2} \\ \text{P3} \\ \text{P4} \end{array} \begin{array}{c} \text{F1} \text{ F2} \text{ F3} \text{ F4} \\ \left[ \begin{array}{cccc} f_{11} & 0 & 0 & 0 \\ f_{21} & f_{22} & 0 & 0 \\ f_{31} & f_{32} & f_{33} & 0 \\ f_{41} & f_{42} & f_{43} & f_{44} \end{array} \right] \\ \mathbf{F} \end{array} * \begin{array}{c} \left[ \begin{array}{cccc} f_{11} & f_{21} & f_{31} & f_{41} \\ 0 & f_{22} & f_{32} & f_{42} \\ 0 & 0 & f_{33} & f_{43} \\ 0 & 0 & 0 & f_{44} \end{array} \right] \\ \mathbf{F}' \end{array}$$

# Saturated Model

- Use Cholesky decomposition to estimate covariance matrix
- Fully saturated
- Model:  $\text{Cov } P = F^*F'$ 
  - F: Full nvar nvar

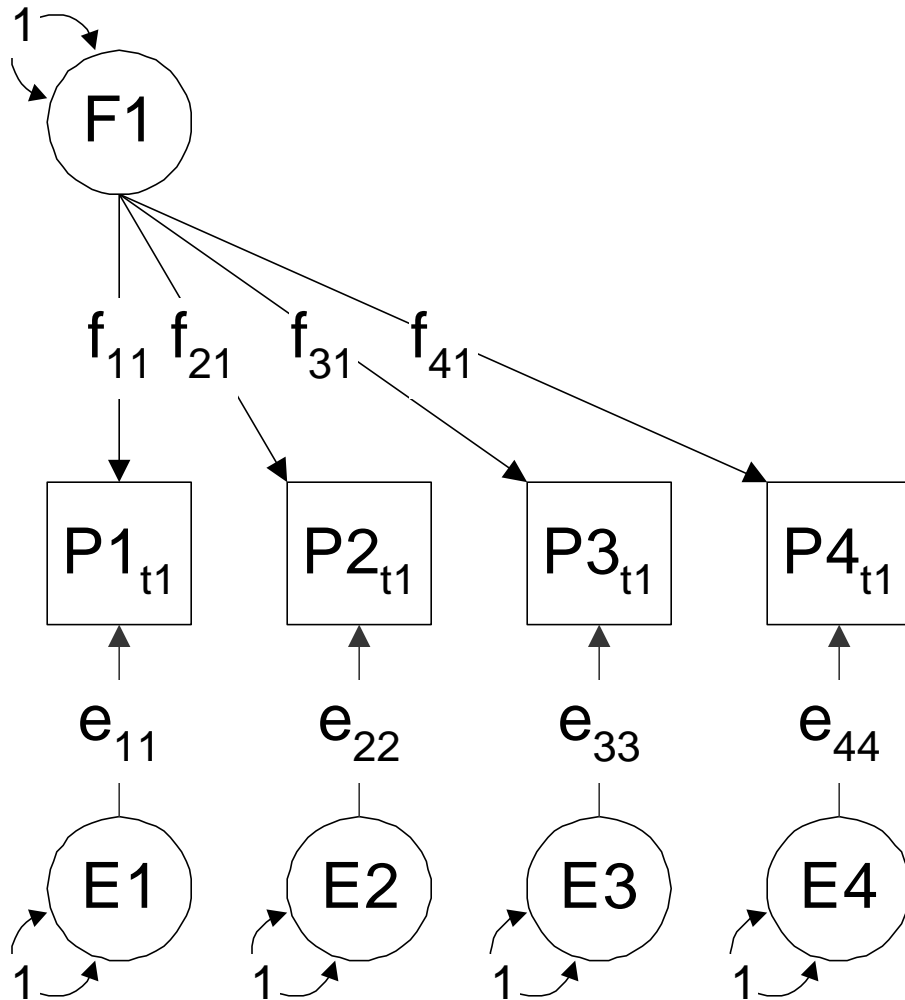
# Phenotypic Single Factor



$$\begin{array}{c} F1 \\ P1 \\ P2 \\ P3 \\ P4 \end{array} \begin{bmatrix} f_{11} \\ f_{21} \\ f_{31} \\ f_{41} \end{bmatrix} * \begin{bmatrix} f_{11} & f_{21} & f_{31} & f_{41} \end{bmatrix}$$

**F** \* **F'**

# Residual Variances

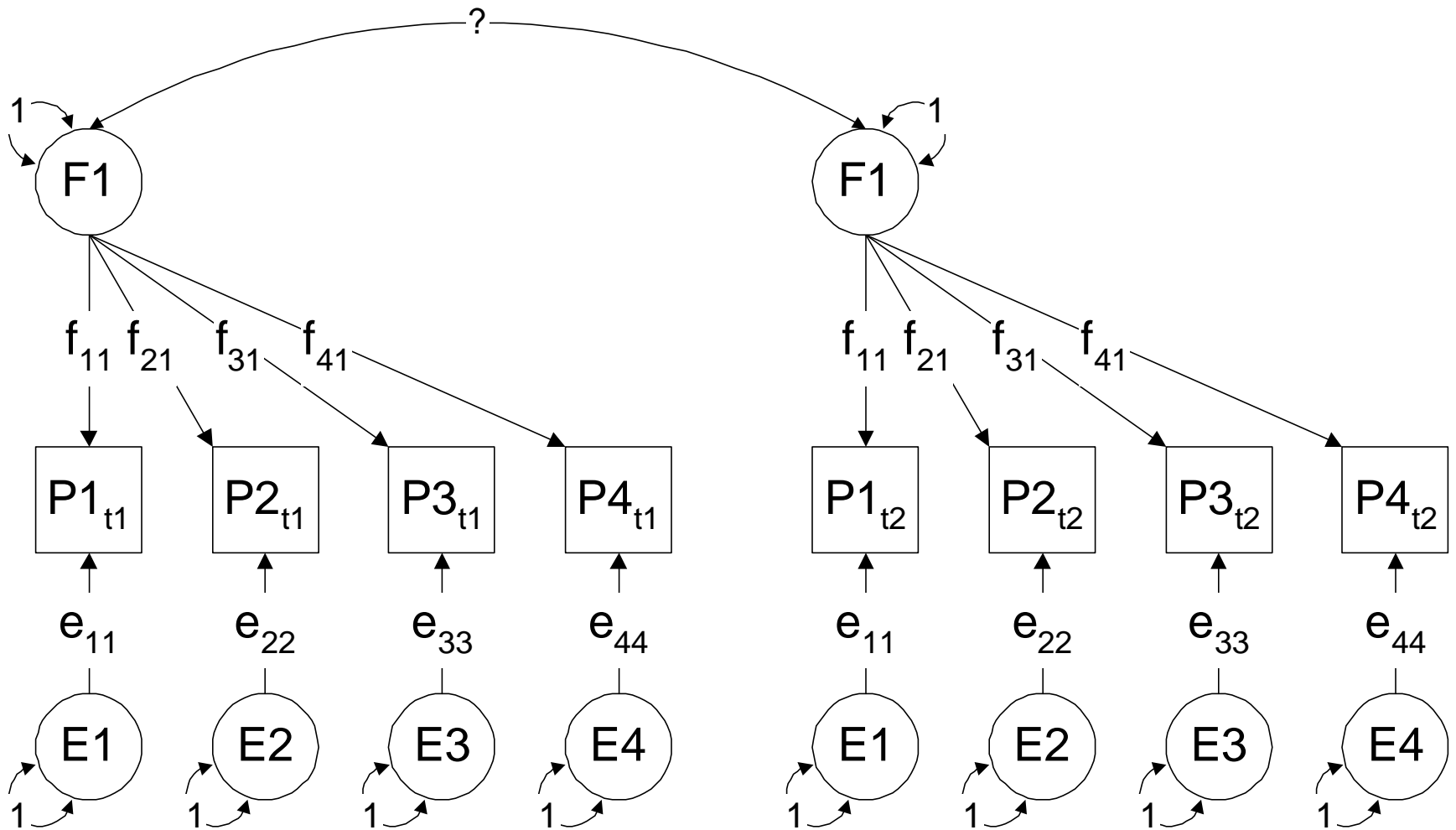


$$\begin{array}{c}
 \text{E1 E2 E3 E4} \\
 \text{P1} \begin{bmatrix} e_{11} & 0 & 0 & 0 \\ 0 & e_{22} & 0 & 0 \\ 0 & 0 & e_{33} & 0 \\ 0 & 0 & 0 & e_{44} \end{bmatrix} * \begin{bmatrix} e_{11} & 0 & 0 & 0 \\ 0 & e_{22} & 0 & 0 \\ 0 & 0 & e_{33} & 0 \\ 0 & 0 & 0 & e_{44} \end{bmatrix} \\
 \text{P2} \\
 \text{P3} \\
 \text{P4} \\
 \mathbf{E} \quad * \quad \mathbf{E}'
 \end{array}$$

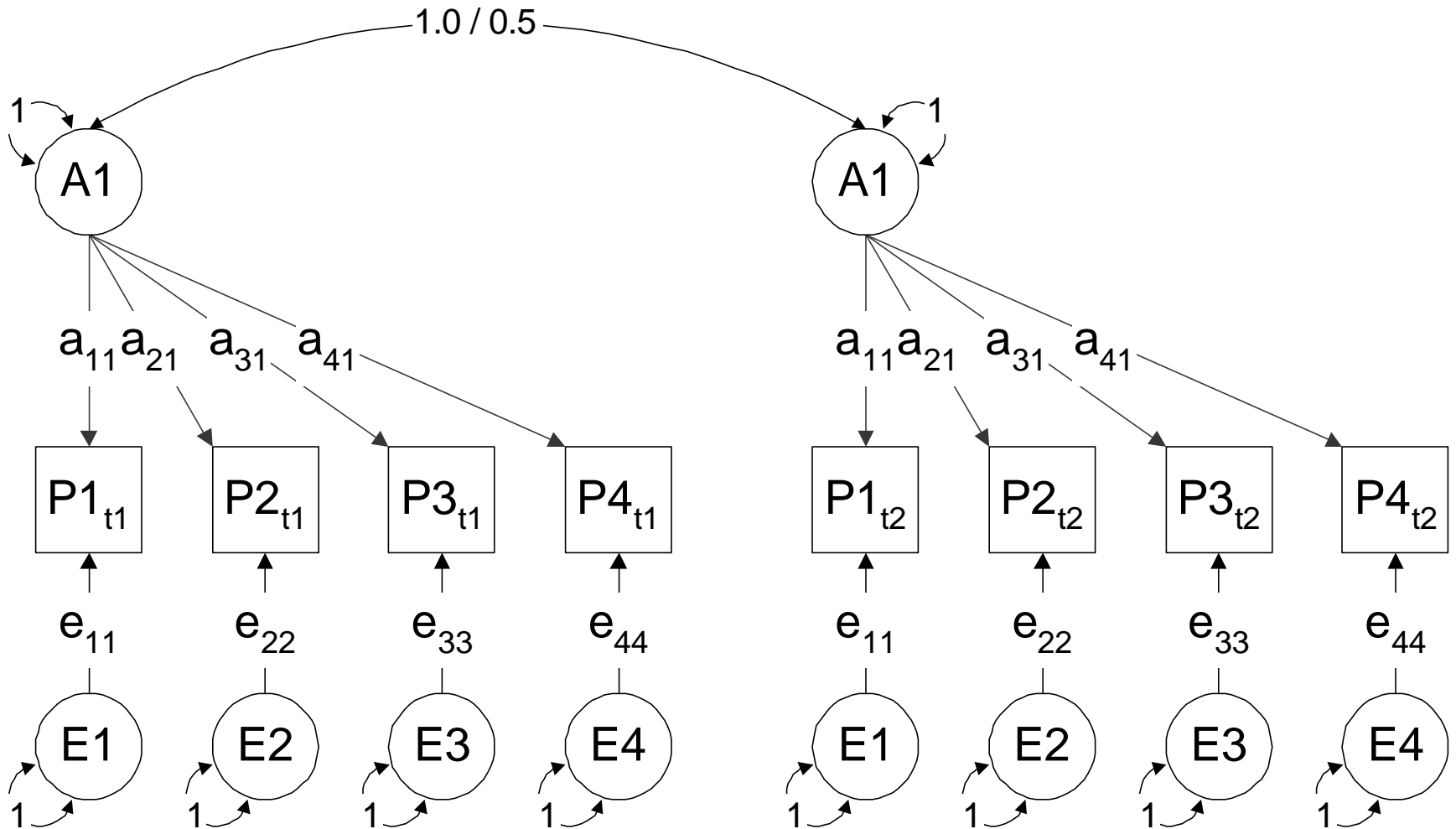
# Factor Analysis

- Explain covariance by limited number of factors
- Exploratory / Confirmatory
- Model:  $\text{Cov } P = F * F' + E * E'$ 
  - F: Full nvar nfac
  - E: Diag nvar nvar

# Twin Data



# Genetic Single Factor



# Single [Common] Factor

- X: genetic

- Full 4 x 1

- Full nvar x nfac

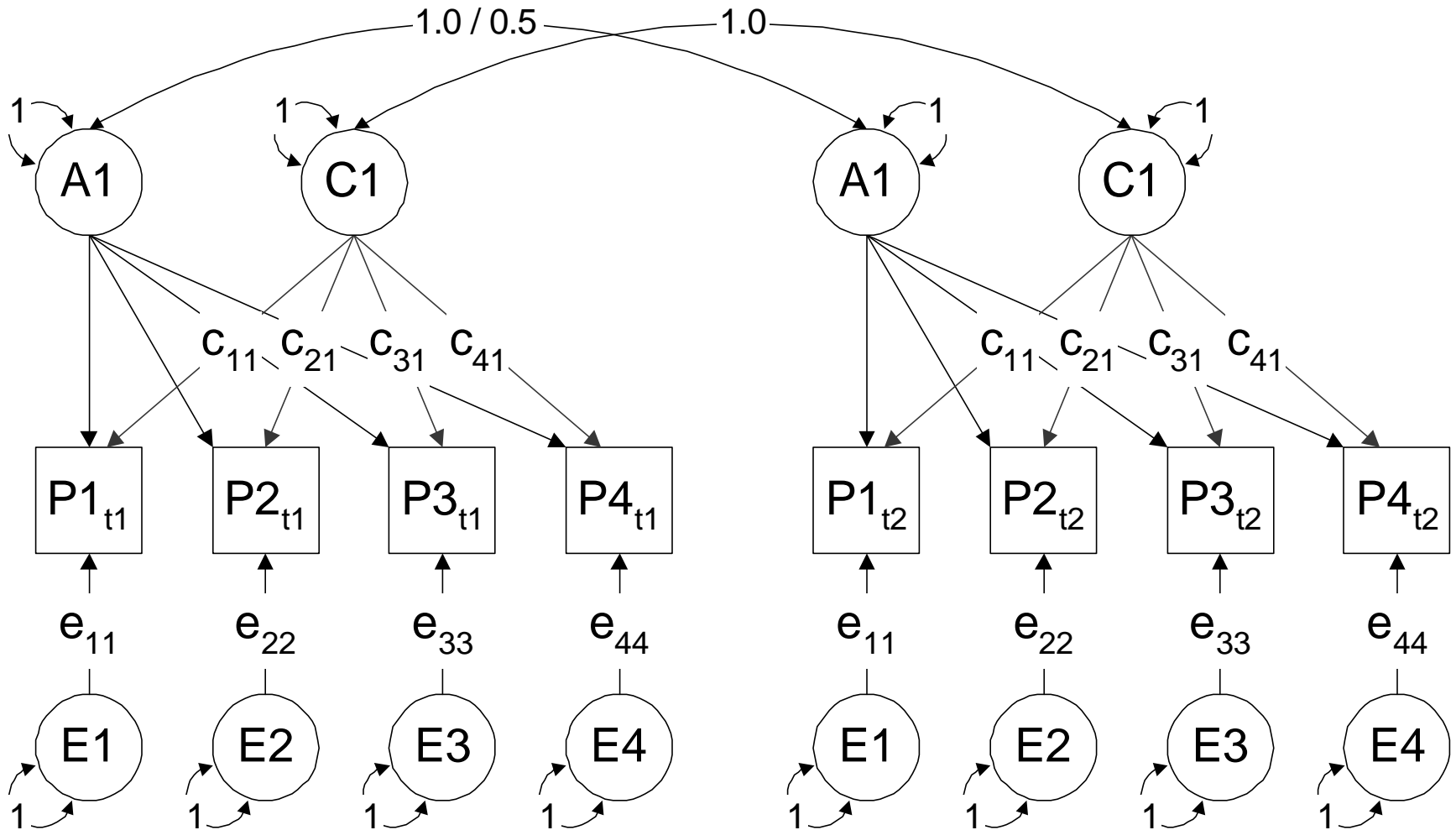
$$\begin{array}{c} \text{P1} \\ \text{P2} \\ \text{P3} \\ \text{P4} \end{array} \begin{array}{c} \text{A1} \\ \left[ \begin{array}{c} a_{11} \\ a_{21} \\ a_{31} \\ a_{41} \end{array} \right] \end{array} * \begin{array}{c} \left[ a_{11} \ a_{21} \ a_{31} \ a_{41} \right] \end{array}$$

**X** \* **X'**

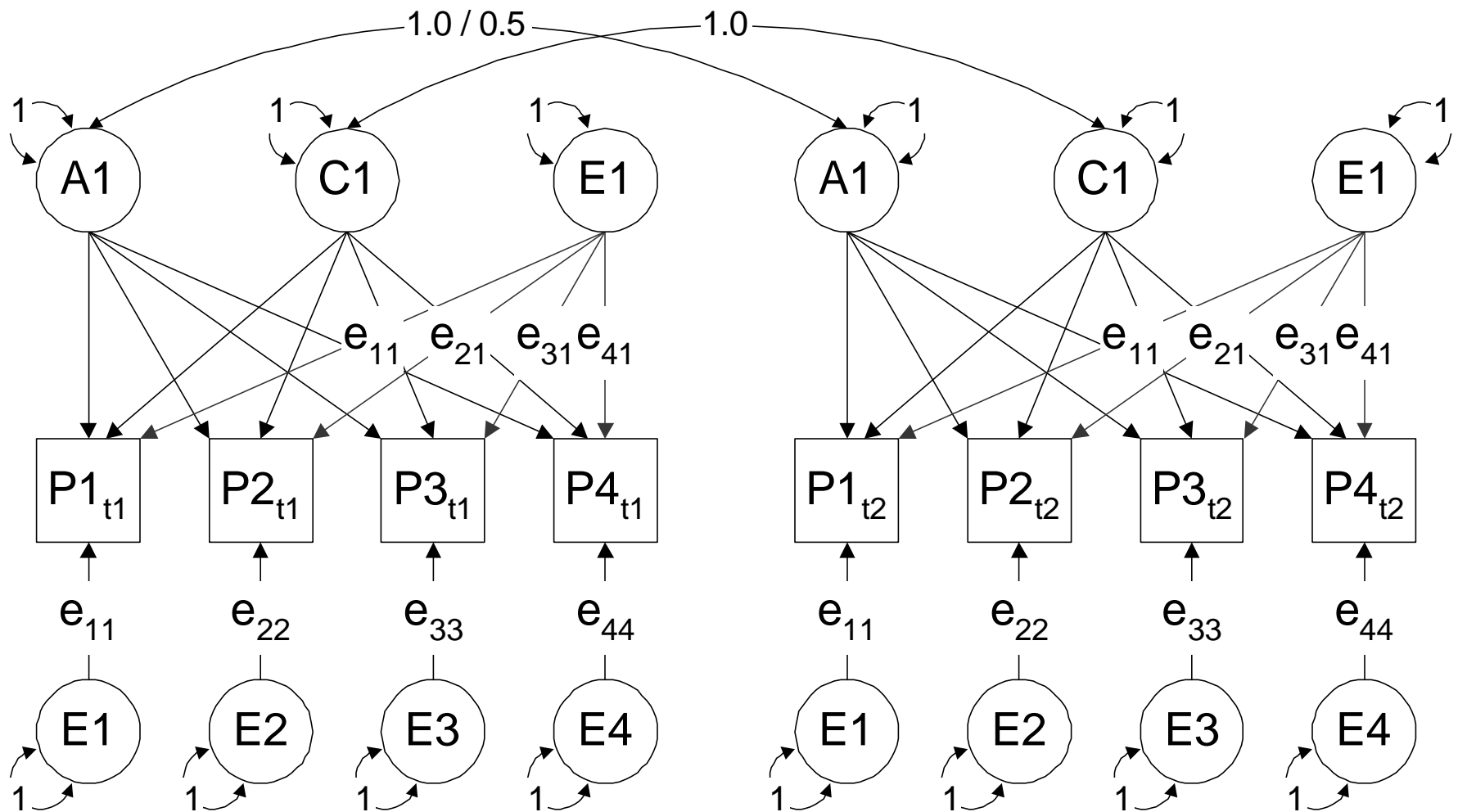
- Y: shared environmental

- Z: specific environmental

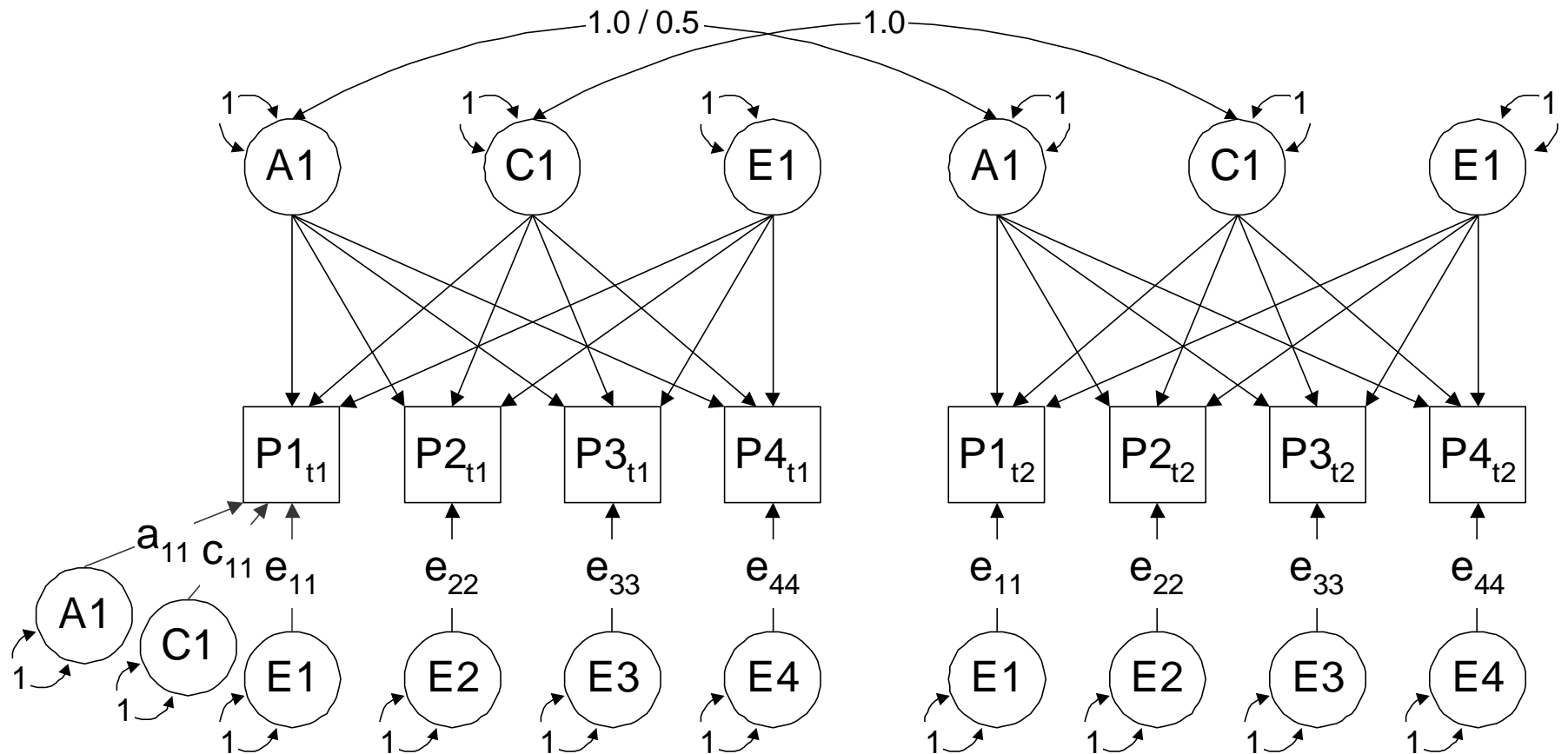
# Common Environmental Single Factor



# Specific Environmental Single Factor



# Residuals partitioned in ACE



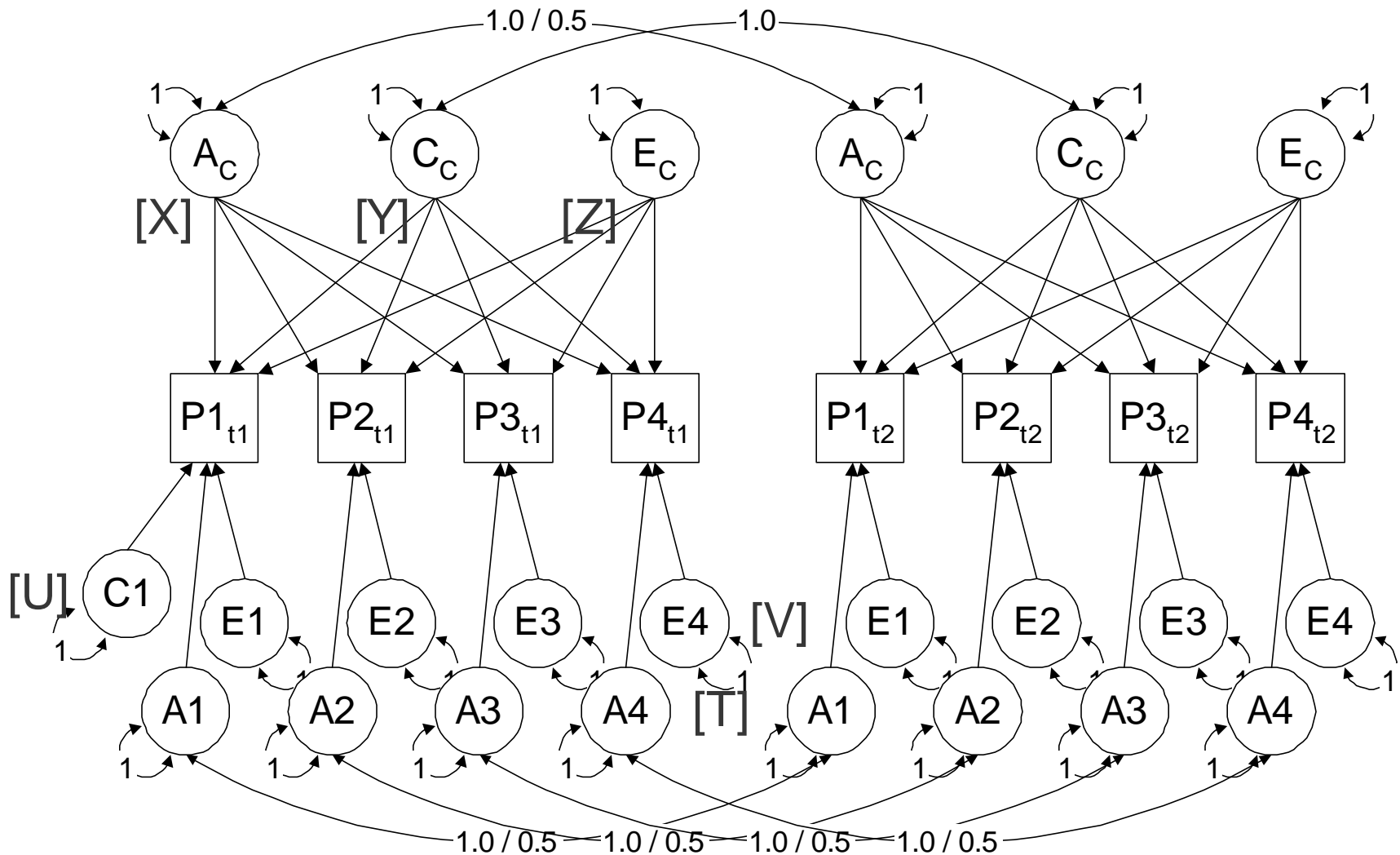
# Residual Factors

- T: genetic
- U: shared environmental
- V: specific environmental
  - Diag 4 x 4
  - Diag nvar x nvar

$$\begin{array}{c} \text{P1} \\ \text{P2} \\ \text{P3} \\ \text{P4} \end{array} \begin{array}{c} \text{E1} \text{ E2} \text{ E3} \text{ E4} \\ \left[ \begin{array}{cccc} e_{11} & 0 & 0 & 0 \\ 0 & e_{22} & 0 & 0 \\ 0 & 0 & e_{33} & 0 \\ 0 & 0 & 0 & e_{44} \end{array} \right] \end{array} * \begin{array}{c} \left[ \begin{array}{cccc} e_{11} & 0 & 0 & 0 \\ 0 & e_{22} & 0 & 0 \\ 0 & 0 & e_{33} & 0 \\ 0 & 0 & 0 & e_{44} \end{array} \right] \end{array}$$

**V** \* **V'**

# Independent Pathway Model



# Path Diagram to Matrices

Variance Component	$a^2$	$c^2$	$e^2$	
Common Factors	[X] 6 x 1	[Y] 6 x 1	[Z] 6 x 1	
Residual Factors	[T] 6 x 6	[U] 6 x 6	[V] 6 x 6	
<code>#define nvar 6</code> <code>#define nfac 1</code>				

# Independent Pathway I

- G1: Define matrices
- Calculation
- Begin Matrices;
- X full nvar nfac Free ! common factor genetic path coefficients
- Y full nvar nfac Free ! common factor shared environment paths
- Z full nvar nfac Free ! common factor unique environment paths
- T diag nvar nvar Free ! variable specific genetic paths
- U diag nvar nvar Free ! variable specific shared env paths
- V diag nvar nvar Free ! variable specific residual paths
- M full 1 nvar Free ! means
- End Matrices;
- Start ...
- Begin Algebra;
- $A = X*X' + T*T'$  ; ! additive genetic variance components
- $C = Y*Y' + U*U'$  ; ! shared environment variance components
- $E = Z*Z' + V*V'$  ; ! nonshared environment variance components
- End Algebra;
- End

indpath.mx

# Independent Pathway II

- G2: MZ twins
- `#include iqnlmz.dat`
- `Begin Matrices = Group 1;`
- `Means M | M ;`
- `Covariance A+C+E | A+C _`
- `A+C | A+C+E ;`
- `Option Residuals`
- `End`
  
- G3: DZ twins
- `#include iqnl dz.dat`
- `Begin Matrices= Group 1;`
- `H full 1 1`
- `End Matrices;`
- `Matrix H .5`
- `Means M | M ;`
- `Covariance A+C+E | H@A+C _`
- `H@A+C | A+C+E ;`
- `Option Residuals`
- `End`

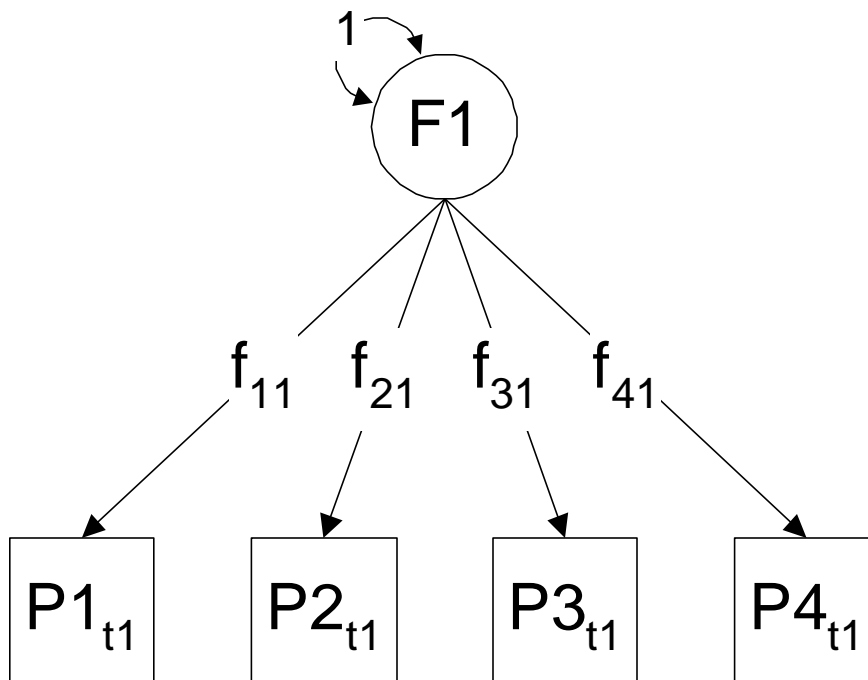
# Independent Pathway III

- G4: Calculate Standardised Solution
- Calculation
- Matrices = Group 1
- I Iden nvar nvar
- End Matrices;
- Begin Algebra;
- R=A+C+E; ! total variance
- S=(\sqrt(I. R))~; ! diagonal matrix of standard deviations
- P=S\*X\_ S\*Y\_ S\*Z; ! standardized estimates for common factors
- Q=S\*T\_ S\*U\_ S\*V; ! standardized estimates for specific factors
- End Algebra;
- Labels Row P a1 a2 a3 a4 a5 a6 c1 c2 c3 c4 c5 c6 e1 e2 e3 e4 e5 e6
- Labels Col P var1 var2 var3 var4 var5 var6
- Labels Row Q as1 as2 as3 as4 as5 as6 cs1 cs2 cs3 cs4 cs5 cs6 es1 es2 es3 es4 e\$
- Labels Col Q var1 var2 var3 var4 var5 var6
- Options NDecimals=4
- End

# IP

- Independent pathways
  - Biometric model
  - Different covariance structure for A, C and E

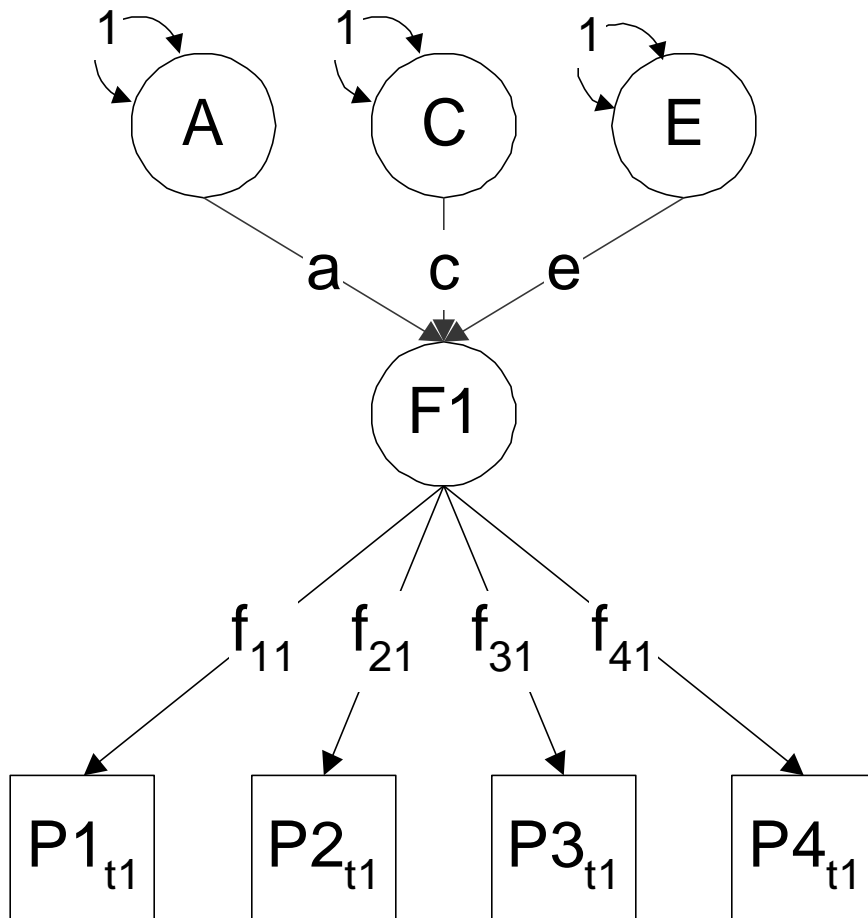
# Phenotypic Single Factor



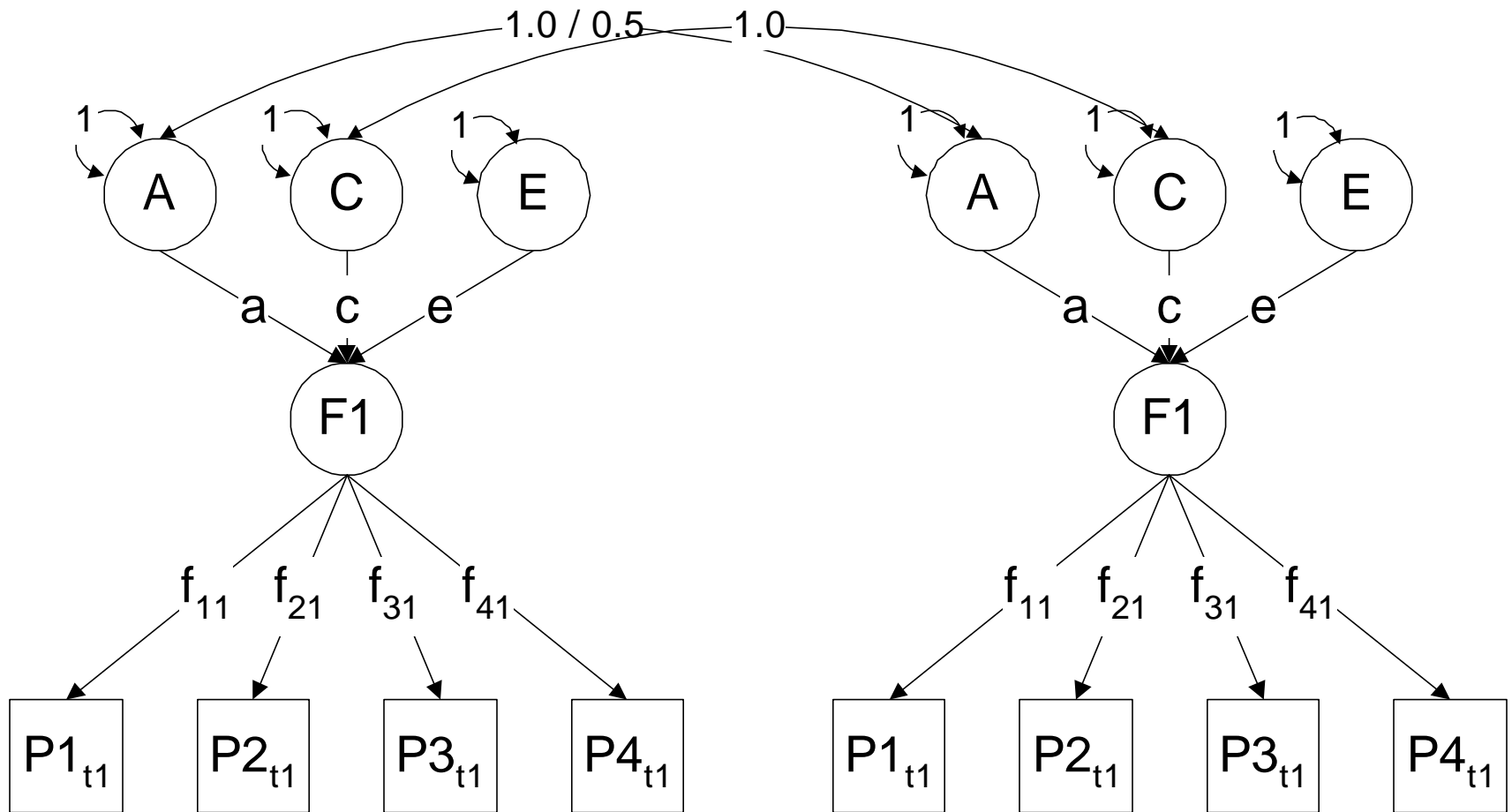
$$\begin{array}{l} F1 \\ P1 \\ P2 \\ P3 \\ P4 \end{array} \begin{bmatrix} f_{11} \\ f_{21} \\ f_{31} \\ f_{41} \end{bmatrix} * \begin{bmatrix} f_{11} & f_{21} & f_{31} & f_{41} \end{bmatrix}$$

**F** \* **F'**

# Latent Phenotype



# Twin Data

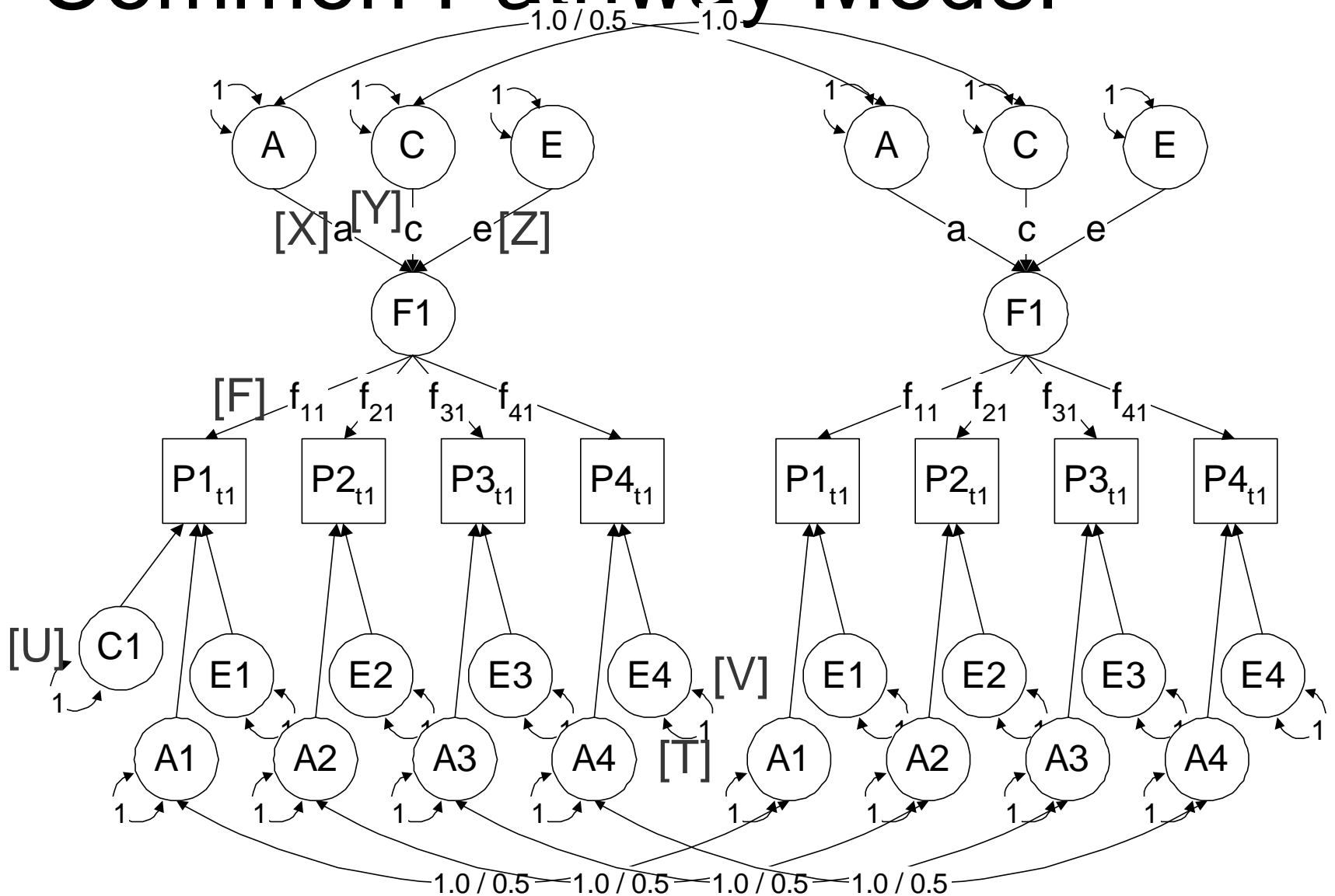


# Factor on Latent Phenotype

$$\begin{array}{c} \text{P1} \\ \text{P2} \\ \text{P3} \\ \text{P4} \end{array} \begin{array}{c} \text{F1} \\ \left[ \begin{array}{c} f_{11} \\ f_{21} \\ f_{31} \\ f_{41} \end{array} \right] \end{array} * \begin{array}{c} \left[ a \right] \\ \left[ a \right] \end{array} * \begin{array}{c} \left[ f_{11} \ f_{21} \ f_{31} \ f_{41} \right] \end{array}$$

$$\begin{aligned} & \mathbf{F} * \mathbf{X} * \mathbf{X}' * \mathbf{F}' \\ = & \mathbf{F} \& (\mathbf{X} * \mathbf{X}') \end{aligned}$$

# Common Pathway Model



# Path Diagram to Matrices

Variance Component	$a^2$	$c^2$	$e^2$	
Common Factor	[X] 1 x 1	[Y] 1 x 1	[Z] 1 x 1	[F] 6 x 1
Residual Factors	[T] 6 x 6	[U] 6 x 6	[V] 6 x 6	
<pre>#define nvar 6 #define nfac 1</pre>				

# Common Pathway Model I

- G1: Define matrices
- Calculation
- Begin Matrices;
- X full nfac nfac Free ! latent factor genetic path coefficient
- Y full nfac nfac Free ! latent factor shared environment path
- Z full nfac nfac Free ! latent factor unique environment path
- T diag nvar nvar Free ! variable specific genetic paths
- U diag nvar nvar Free ! variable specific shared env paths
- V diag nvar nvar Free ! variable specific residual paths
- F full nvar nfac Free ! loadings of variables on latent factor
- I Iden 2 2
- M full 1 nvar Free ! means
- End Matrices;
- Start ..
- Begin Algebra;
- $A = F(X*X') + T*T'$ ; ! genetic variance components
- $C = F(Y*Y') + U*U'$ ; ! shared environment variance components
- $E = F(Z*Z') + V*V'$ ; ! nonshared environment variance components
- $L = X*X' + Y*Y' + Z*Z'$ ; ! variance of latent factor
- End Algebra;
- End

# Common Pathway II

- G4: Constrain variance of latent factor to 1
- Constraint
- Begin Matrices;
- L computed =L1
- I unit 1 1
- End Matrices;
- Constraint L = I ;
- End
  
- G5: Calculate Standardised Solution
- Calculation
- Matrices = Group 1
- D Iden nvar nvar
- End Matrices;
- Begin Algebra;
- R=A+C+E; ! total variance
- S=(\sqrt(D.R))~; ! diagonal matrix of standard deviations
- P=S\*F; ! standardized estimates for loadings on F
- Q=S\*T\_ S\*U\_ S\*V; ! standardized estimates for specific factors
- End Algebra;
- Options NDecimals=4
- End

# CP

- Common pathway
  - Psychometric model
  - Same covariance structure for A, C and E

# Practical Example

- Dataset: NL-IQ Study
- 6 WAIS-III IQ subtests
  - var1 = onvolledige tekeningen / incomplete pict.
  - var2 = woordenschat / vocabulary
  - var3 = paren associeren / digit symbol pairing
  - var4 = incidenteel leren / incidental learning
  - var5 = overeenkomsten / similarities
  - var6 = blokpatronen / block design
- N MZF: 27, DZF: 70



# Summary

- Independent Pathway Model
  - Biometric Factor Model
  - Loadings differ for genetic and environmental common factors
- Common Pathway Model
  - Psychometric Factor Model
  - Loadings equal for genetic and environmental common factor