

- ❖ Go to  
Faculty/marleen/Boulder2012/Moderating\_cov
- ❖ Copy all files to your own directory
  
- ❖ Go to  
Faculty/sanja/Boulder2012/Moderating\_covariances  
\_IQ\_SES
- ❖ Copy all files to your own directory

# Moderating covariances

Marleen de Moor & Sanja Franić  
Boulder Twin Workshop  
March 8, 2012

# Heterogeneity – Multigroup models (Tuesday)

- Is the magnitude of genetic influences on ADHD the same in boys and girls?
- Do different genetic factors influence ADHD in boys and girls?
- Multiple Group Models
  - Sex differences: MZM, DZM, MZF, DZF, DOS
  - Cohorts differences: MZyoung, DZyoung, MZold, DZold

# Heterogeneity – Moderation models (Today)

- Some variables have many categories:
  - Socioeconomic status (5 levels)
- Some variables are continuous:
  - Age
  - Parenting
- Grouping these variables into high/low categories loses a lot of information

# Gene-Environment Interaction

GxE:

- Genetic control of **sensitivity** to the environment
- Environmental control of gene **expression**

Examples:

- ✓ Does the heritability of IQ depend on SES?
- ✓ Does the heritability of ADHD depends on age?

# Gene-Environment Correlation

rGE:

- Genetic control of **exposure** to the environment
- Environmental control of gene **frequency**

Examples:

- ✓ Active rGE: Children with high IQ read more books
- ✓ Passive rGE: Parents of children with high IQ take their children more often to the library
- ✓ Reactive rGE: Children with ADHD are treated differently by their parents

# GxE: moderation models

## Variance Components Models for Gene–Environment Interaction in Twin Analysis


---

Shaun Purcell

*Social, Genetic and Developmental Psychiatry Research Centre, Institute of Psychiatry, King's College, London, UK*

Gene–environment interaction is likely to be a common and important source of variation for complex behavioral traits. Often conceptualized as the genetic control of sensitivity to the environment, it can be incorporated in variance components twin analyses by partitioning genetic effects into a mean part, which is independent of the environment, and a part that is a linear function of the environment. The model allows for one or more environmental moderator variables (that possibly interact with each other) that may i) be continuous or binary ii) differ between twins within a pair iii) interact with residual environmental as well as genetic effects iv) have nonlinear moderating properties v) show scalar (different magnitudes) or qualitative (different genes) interactions vi) be correlated with genetic effects acting upon the trait, to allow for a test of gene–environment interaction in the presence of gene–environment correlation. Aspects and applications of a class of models are explored by simulation, in the context of both individual differences twin analysis and, in a companion paper (Purcell & Sham, 2002) sibpair quantitative trait locus linkage analysis. As well as elucidating environmental pathways, consideration of gene–environment interaction in quantitative and molecular studies will potentially direct and enhance gene-mapping efforts.

bivariate twin distribution, where twin pair difference correlates with twin pair sum (Jinks & Fulker, 1970). However, as well as suffering from low power, this test also is sensitive to non-normality in the trait. More importantly, beyond indicating that *some* form of interaction is occurring, it sheds no light on underlying processes. Having both  $G$  and  $E$  as measured variables provides the most power for detecting  $G \times E$ ; the results will potentially be very informative also, beginning to map onto the underlying biology. For example, sex moderates the effect of the *APOE e4* allele on cognitive decline, where women show higher *e4*-associated risk than men (Yaffe et al., 2000). Additionally, the *e4* allele moderates the impact of estrogen in women on cognitive decline, as the estrogen use is associated with less cognitive decline only in women without the *e4* risk allele.

In the present paper we consider the case of latent  $G \times$  measured  $E$ , which is most relevant to the classical twin study. For example, ad  pn

Twin  
Research  
2002

# Application

PSYCHOLOGICAL SCIENCE

## Research Article

### SOCIOECONOMIC STATUS MODIFIES HERITABILITY OF IQ IN YOUNG CHILDREN

Eric Turkheimer, Andreeana Haley, Mary Waldron, Brian D'Onofrio,  
and Irving I. Gottesman

*University of Virginia*

**Abstract**—Scores on the Wechsler Intelligence Scale for Children were analyzed in a sample of 7-year-old twins from the National Collaborative Perinatal Project. A substantial proportion of the twins were raised in families living near or below the poverty level. Biometric analyses were conducted using models allowing for components attributable to the additive effects of genotype, shared environment, and non-shared environment to interact with socioeconomic status (SES) measured as a continuous variable. Results demonstrate that the proportions of IQ variance attributable to genes and environment vary nonlinearly with SES. The models suggest that in impoverished families, 60% of the variance in IQ is accounted for by the shared environment, and the contribution of genes is close to zero; in affluent families, the result is almost exactly the reverse.

uous variable like socioeconomic status (SES) and latent genetic and environmental influences on cognitive ability.

Nevertheless, several previous studies have addressed differential heritability as a function of race, social class, or parental education. Scarr-Salapatek (1971) obtained a sample of twins from the Philadelphia school system and used standardized test scores as a measure of ability. SES was estimated from census-tract information. A major limitation of the study was that it did not include zygosity information about the twin pairs; instead, analyses were based on comparisons of same-sex pairs (that combined monozygotic, or MZ, twins and dizygotic, or DZ, twins) and opposite-sex pairs (all DZ twins), a method with considerably less statistical power than the classical twin design (Eaves & Jinks, 1972). For both Black and White children, estimated heritabilities were lower in children from families in lower-SES census tracts than in those from

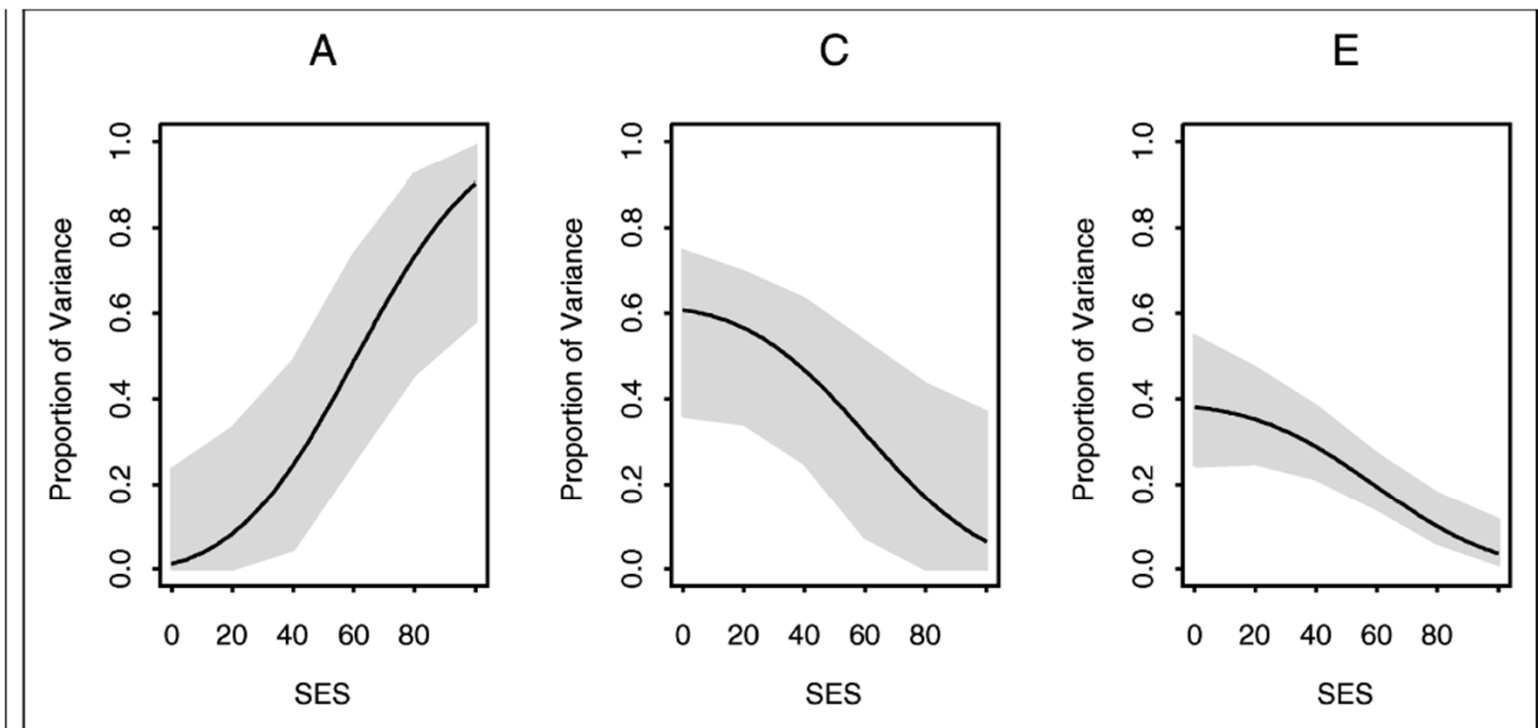
Turkheimer et al. 2003



# Application

PSYCHOLOGICAL SCIENCE

E. Turkheimer et al.



**Fig. 3.** Proportion of total Full-Scale IQ variance accounted for by A, C, and E plotted as a function of observed socioeconomic status (SES). Shading indicates 95% confidence intervals.

# Practical

- Replicate findings of Turkheimer et al.
- Sample of 5-yr old twins from Netherlands Twin Register
- Phenotype: FSIQ
- Environmental/moderator variable:SES

# 'Definition variables' in OpenMx

- General definition: Definition variables are variables that may vary per pair/subject and that are not dependent variables
- In OpenMx: The specific value of the def var for a specific pair/individual is read into an mxMatrix in OpenMx when analyzing the data of that particular pair/individual

# 'Definition variables' in OpenMx

Common uses:

1. As main effects on the means (e.g. age and sex)
2. To model changes in variance components as function of some variable (e.g., age, SES, etc)

# Cautionary note about definition variables

- Def var should not be missing if dependent is not missing
- Def var should not have the same missing value as dependent variable (e.g., use -2.00 for def var, -1.00 for dep var)

# Definition variables as main effects

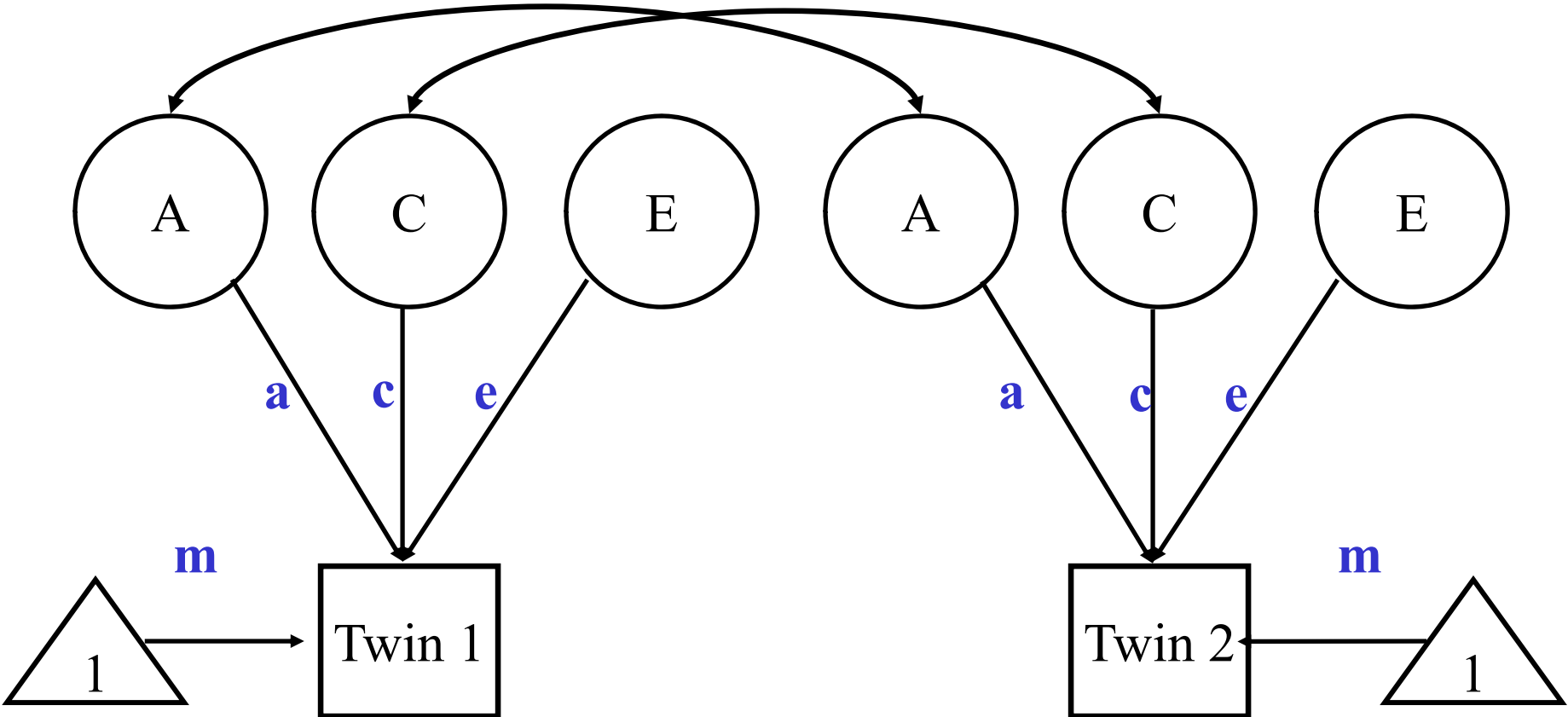
General model with age and sex as main effects:

$$y_i = \alpha + \beta_1(\text{age}_i) + \beta_2(\text{sex}_i) + \varepsilon_i$$

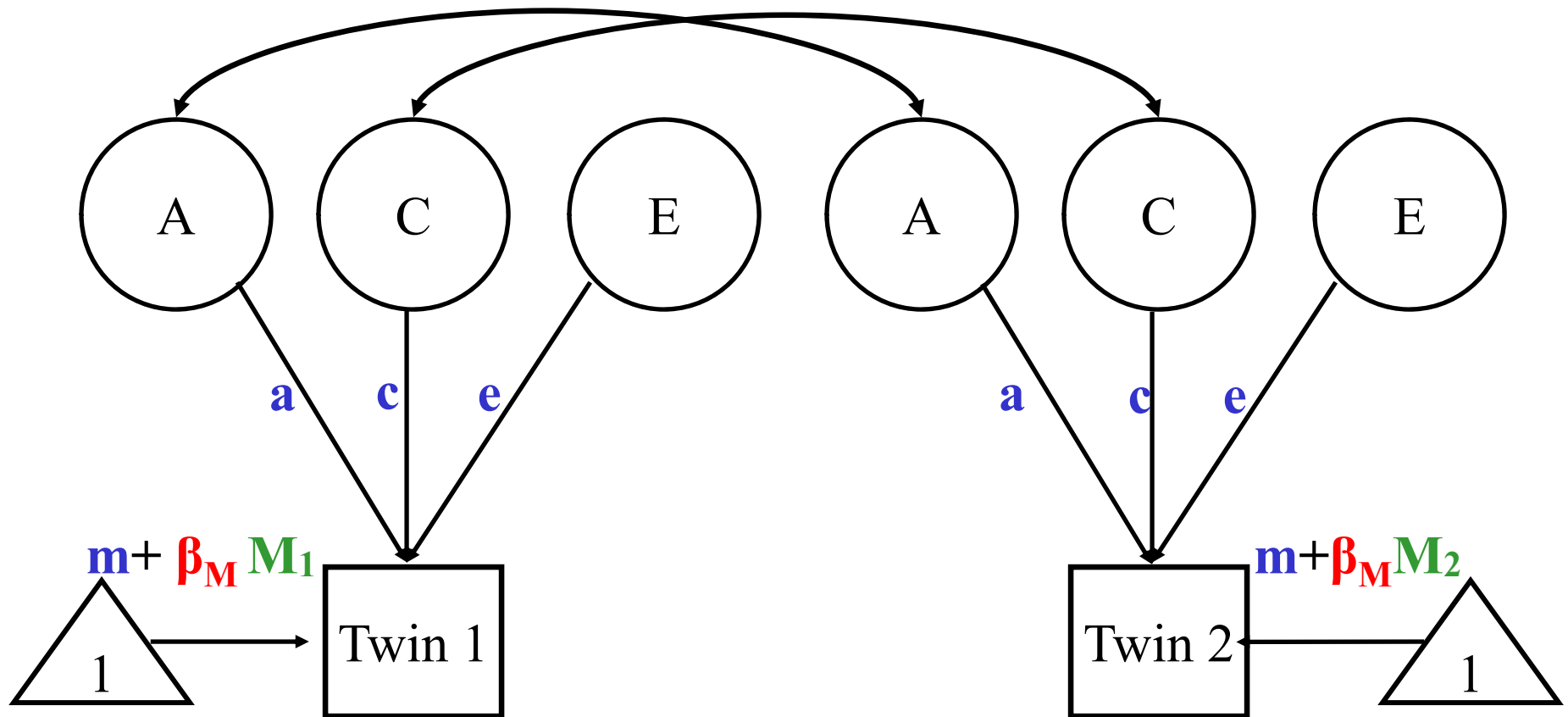
Where:

- ❖  $y_i$  is the observed score of individual  $i$
- ❖  $\alpha$  is the intercept or grand mean
- ❖  $\beta_1$  is the regression weight of age
- ❖  $\text{age}_i$  is the age of individual  $i$
- ❖  $\beta_2$  is the deviation of females (if sex is coded 0= male; 1=female)
- ❖  $\text{sex}_i$  is the sex of individual  $i$
- ❖  $\varepsilon_i$  is the residual that is not explained by the definition variables (and can be decomposed further into ACE etc)

# Standard ACE model



# Standard ACE model + Main effect on Means





# Standard ACE model

- Means vector

$$\begin{pmatrix} m & m \end{pmatrix}$$

- Covariance matrix

$$\begin{pmatrix} a^2 + c^2 + e^2 & \\ Za^2 + c^2 & a^2 + c^2 + e^2 \end{pmatrix}$$

# Allowing for a main effect of $X$

- Means vector

$$\left( m + \beta_M X_{1i} \quad m + \beta_M X_{2i} \right)$$

- Covariance matrix

$$\begin{pmatrix} a^2 + c^2 + e^2 & \\ Za^2 + c^2 & a^2 + c^2 + e^2 \end{pmatrix}$$

# Allowing for a main effect of $X$

$$(m + \beta_M X_{1i} \quad m + \beta_M X_{2i})$$

OpenMx

```
intercept <- mxMatrix( type="Full", nrow=1, ncol=nv, free=TRUE, values=.1, label="interc", name="int" )
PathM <- mxMatrix( type="Full", nrow=1, ncol=1, free=T, values=.6, label=c("m11"), name="m" )
mod <- mxMatrix( type="Full", nrow=1, ncol=1, free=FALSE, labels=c("data.ses"), name="D")

wmod <- mxAlgebra( expression= m %*% D, name="DR")
meanG <- mxAlgebra( expression= cbind((int + DR),(int + DR)), name="expMeanG")
```

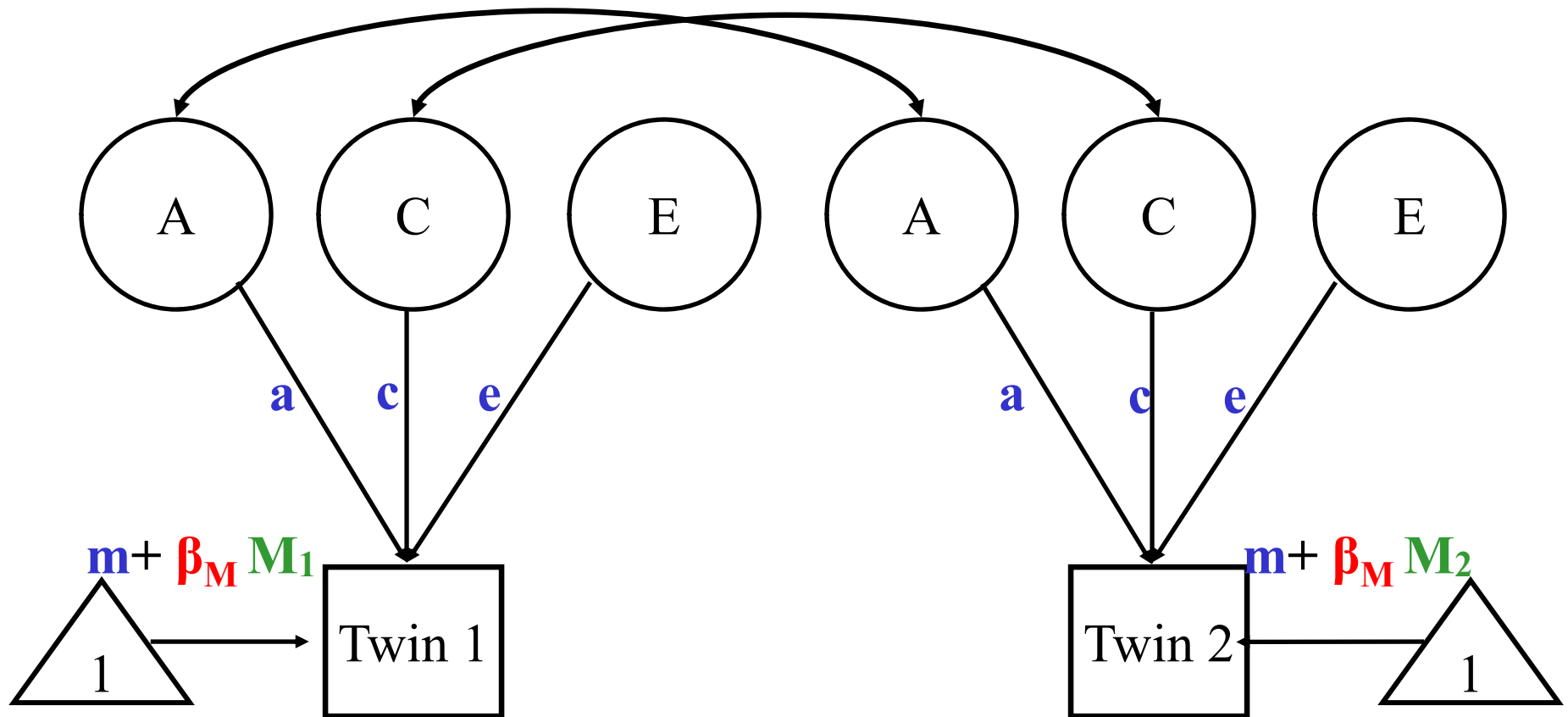
$$\begin{matrix} (\text{int}) \\ (m) \\ (\text{data.ses}) \end{matrix} \quad (\text{int} + m * \text{data.ses} \quad \text{int} + m * \text{data.ses})$$

# 'Definition variables' in OpenMx

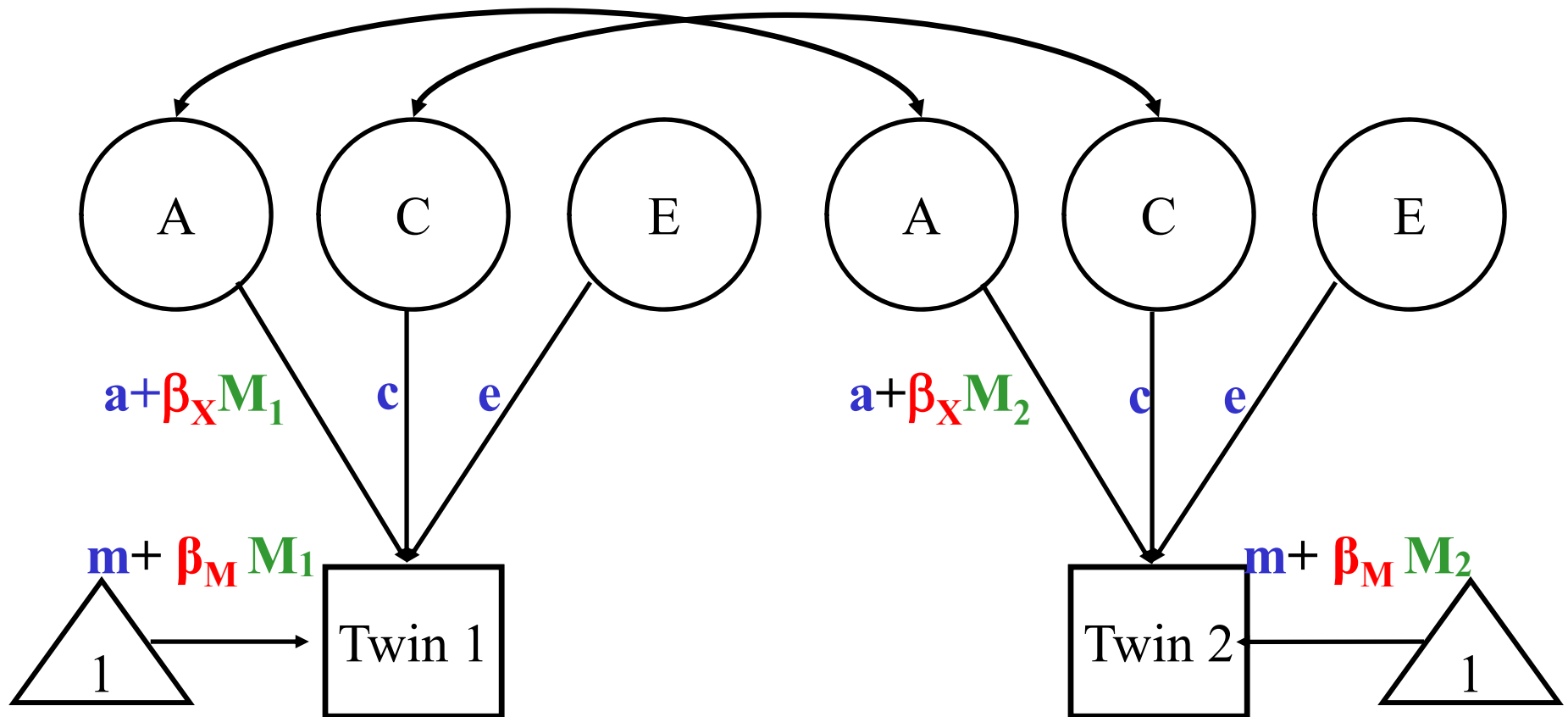
Common uses:

1. As main effects on the means (e.g. age and sex)
2. To model changes in variance components as function of some variable (e.g., age, SES, etc)

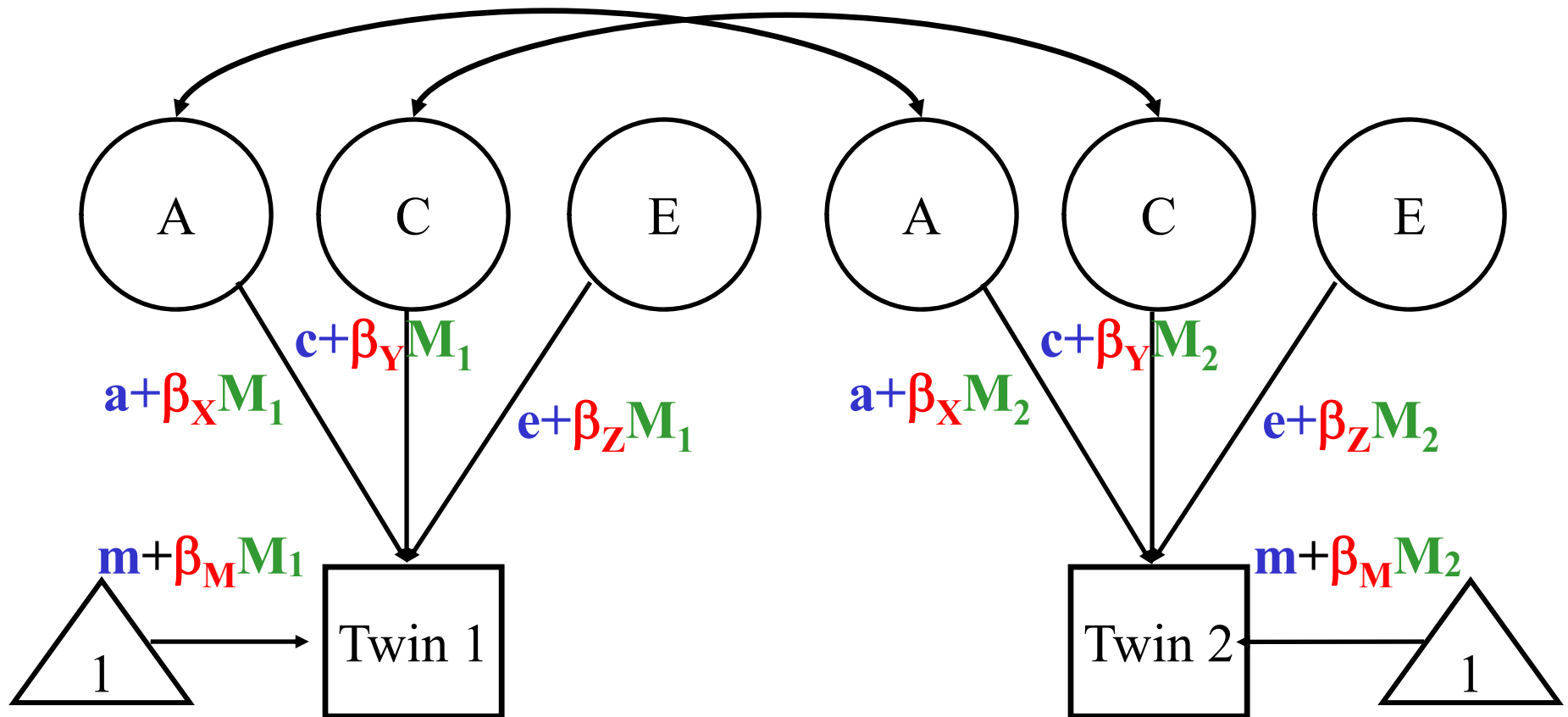
# Standard ACE model + Effect on Means

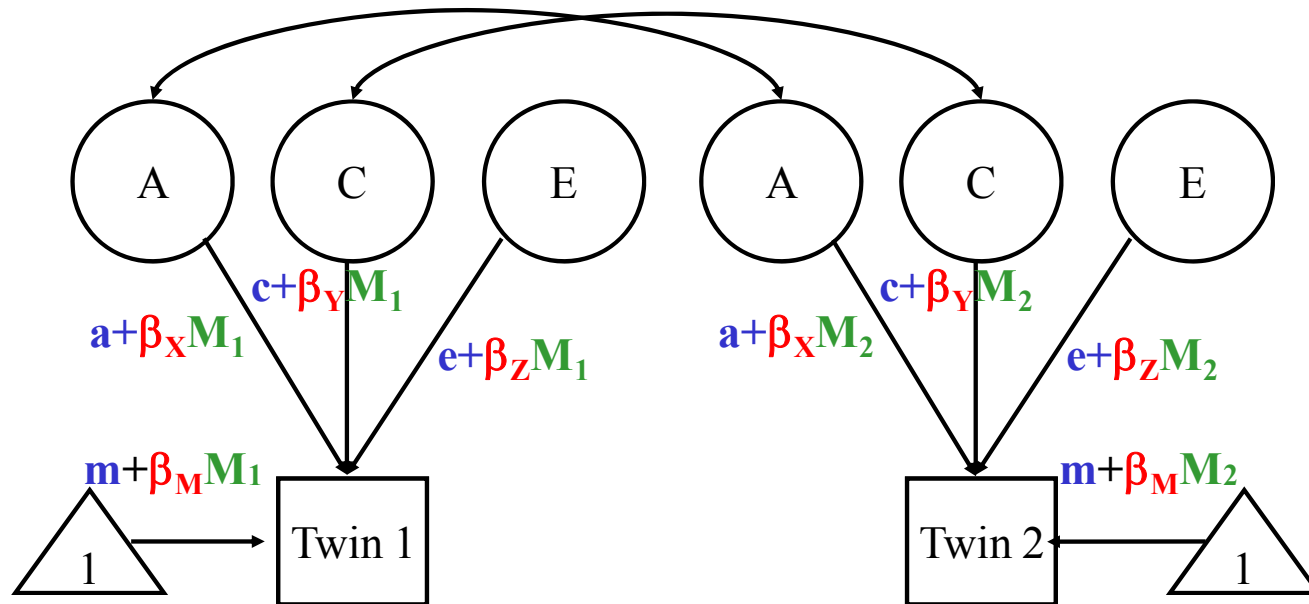


# Standard ACE model + Effect on Means and "a" path



# Standard ACE model + Effect on Means and "a/c/e" paths



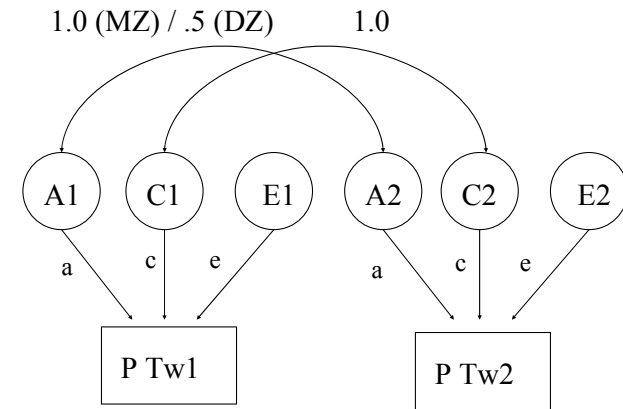


- Effect on means:
  - ❖ Main effects
  - ❖ To account for **gene-environment correlation**
- Effect on a/c/e path loadings:
  - ❖ Moderation effects
  - ❖ To model **gene-environment interaction**  
(and environment-environment interaction)



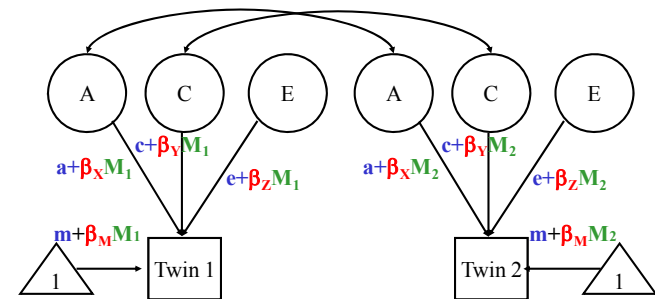
- **Classic Twin Model:**

$$\text{Var}(P) = a^2 + c^2 + e^2$$



- **Moderation Model:**

$$\text{Var}(P) = (a + \beta_X M)^2 + (c + \beta_Y M)^2 + (e + \beta_Z M)^2$$



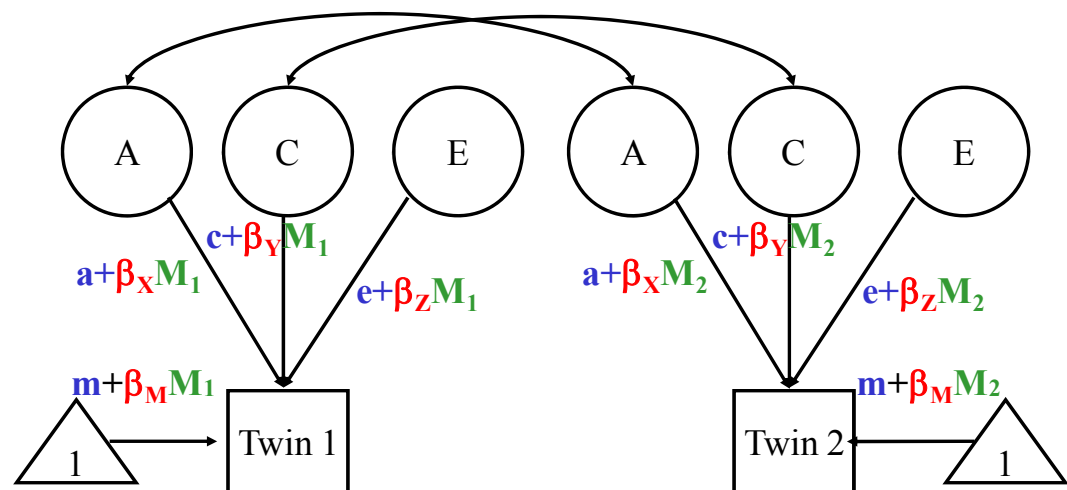
Note: Variances of the latent factors are constrained to 1

# Expected variance

$$\text{Var}(P) = (a + \beta_x M)^2 + (c + \beta_y M)^2 + (e + \beta_z M)^2$$

Where  $M$  is the value of the moderator and

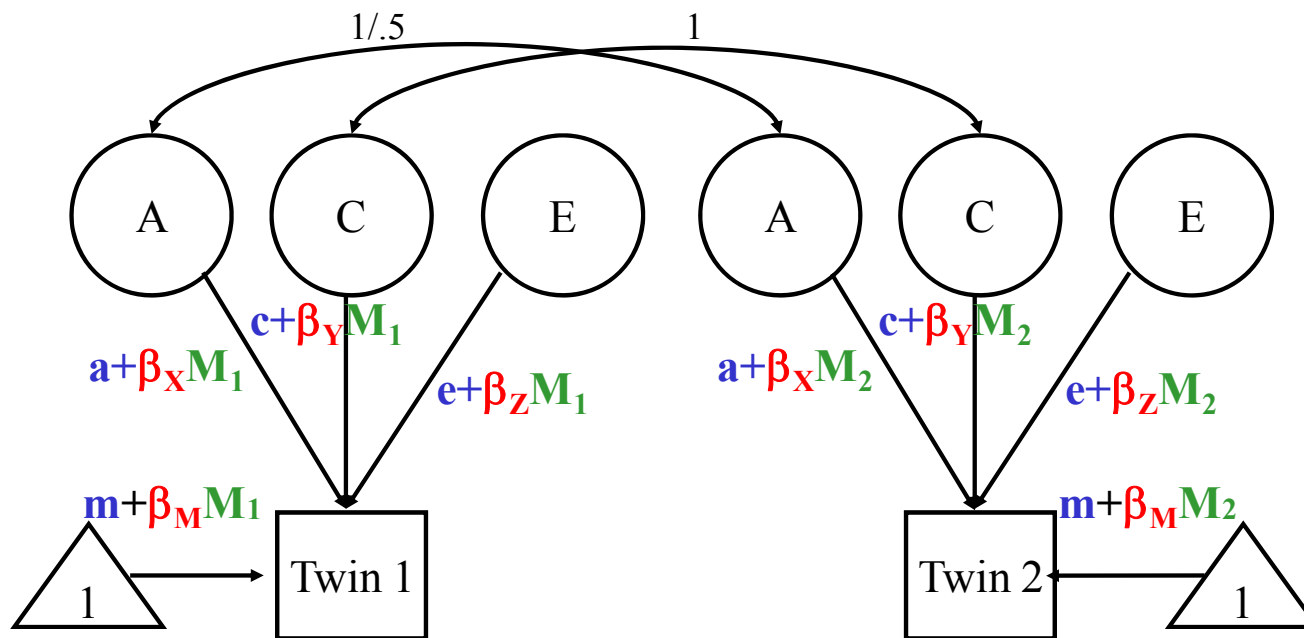
- Significance of  $\beta_x$  indicates genetic moderation
- Significance of  $\beta_y$  indicates common environmental moderation
- Significance of  $\beta_z$  indicates unique environmental moderation



# Expected MZ / DZ covariances

$$\text{Cov}(P_1, P_2)_{\text{MZ}} = (a + \beta_X M)^2 + (c + \beta_Y M)^2$$

$$\text{Cov}(P_1, P_2)_{\text{DZ}} = 0.5 * (a + \beta_X M)^2 + (c + \beta_Y M)^2$$



# Expected **MZ** covariance matrix

$$\begin{pmatrix} (a + \beta_X M)^2 + (c + \beta_Y M)^2 + (e + \beta_Z M)^2 & (a + \beta_X M)^2 + (c + \beta_Y M)^2 \\ (a + \beta_X M)^2 + (c + \beta_Y M)^2 & (a + \beta_X M)^2 + (c + \beta_Y M)^2 + (e + \beta_Z M)^2 \end{pmatrix}$$

```
nv<-1
# Matrices to store a, c, and e Path Coefficients
pathA <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=TRUE, values=.6, label="a11", name="a" )
pathC <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=TRUE, values=.6, label="c11", name="c" )
pathE <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=TRUE, values=.6, label="e11", name="e" )
# Matrices to store the moderated a, c, and e Path Coefficients
modPathA <- mxMatrix( type='Full', nrow=nv, ncol=nv, free=TRUE, values=.6, label="aM11", name="aM" )
modPathC <- mxMatrix( type='Full', nrow=nv, ncol=nv, free=TRUE, values=.6, label="cM11", name="cM" )
modPathE <- mxMatrix( type='Full', nrow=nv, ncol=nv, free=TRUE, values=.6, label="eM11", name="eM" )
# Matrix for the moderator variable
mod <- mxMatrix( type="Full", nrow=1, ncol=1, free=FALSE, labels=c("data.ses"), name="D" )
# Matrices to compute the moderated A, C, and E variance components
covAmod <- mxAlgebra( expression=(a+ D%*%aM) %*% t(a+ D%*%aM), name="A" )
covCmod <- mxAlgebra( expression=(c+ D%*%cM) %*% t(c+ D%*%cM), name="C" )
covEmod <- mxAlgebra( expression=(e+ D%*%eM) %*% t(e+ D%*%eM), name="E" )
# Algebra for the expected mean vector and expected variance/covariance matrices and in MZ and DZ
covMZ <- mxAlgebra( expression= rbind ( cbind(A+C+E , A+C),
                                         cbind(A+C, A+C+E)), name="expCovMZ" )
```

# Expected **MZ** covariance matrix

```
nv<-1
# Matrices to store a, c, and e Path Coefficients
pathA  <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=TRUE, values=.6, label="a11", name="a" )
pathC  <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=TRUE, values=.6, label="c11", name="c" )
pathE  <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=TRUE, values=.6, label="e11", name="e" )
# Matrices to store the moderated a, c, and e Path Coefficients
modPathA <- mxMatrix( type='Full', nrow=nv, ncol=nv, free=TRUE, values=.6, label="aM11", name="aM" )
modPathC <- mxMatrix( type='Full', nrow=nv, ncol=nv, free=TRUE, values=.6, label="cM11", name="cM" )
modPathE <- mxMatrix( type='Full', nrow=nv, ncol=nv, free=TRUE, values=.6, label="eM11", name="eM" )
```

$$a = (a11) \quad aM = (aM11)$$

$$c = (c11) \quad cM = (cM11)$$

$$e = (e11) \quad eM = (eM11)$$

# Expected **MZ** covariance matrix

```
nv<-1
# Matrix for the moderator variable
mod <- mxMatrix( type="Full", nrow=1, ncol=1, free=FALSE, labels=c("data.ses"), name="D")
# Matrices to compute the moderated A, C, and E variance components
covAmod <- mxAlgebra( expression=(a+ D%*%aM) %*% t(a+ D%*%aM), name="A" )
covCmod <- mxAlgebra( expression=(c+ D%*%cM) %*% t(c+ D%*%cM), name="C" )
covEmod <- mxAlgebra( expression=(e+ D%*%eM) %*% t(e+ D%*%eM), name="E" )
```

$$D = (data.ses)$$

$$a = (a11) \quad aM = (aM11) \quad A = (a11 + data.ses * aM11)^2$$

$$c = (c11) \quad cM = (cM11) \quad C = (c11 + data.ses * cM11)^2$$

$$e = (e11) \quad eM = (eM11) \quad E = (e11 + data.ses * eM11)^2$$

# Expected **MZ** covariance matrix

OpenMx

```
nv<-1
```

```
# Algebra for expected variance/covariance matrix and expected mean vector in MZ
```

```
mxAlgebra( expression= rbind ( cbind(A+C+E , A+C),  
                               cbind(A+C , A+C+E)), name="expCovMZ" ),
```

$$A = (a_{11} + data.ses * a_{M11})^2$$

$$C = (c_{11} + data.ses * c_{M11})^2$$

$$E = (e_{11} + data.ses * e_{M11})^2$$

$$\text{expCovMZ} = \begin{bmatrix} (a_{11} + data.ses * a_{M11})^2 + (c_{11} + data.ses * c_{M11})^2 + (e_{11} + data.ses * e_{M11})^2 & (a_{11} + data.ses * a_{M11})^2 + (c_{11} + data.ses * c_{M11})^2 \\ (a_{11} + data.ses * a_{M11})^2 + (c_{11} + data.ses * c_{M11})^2 & (a_{11} + data.ses * a_{M11})^2 + (c_{11} + data.ses * c_{M11})^2 + (e_{11} + data.ses * e_{M11})^2 \end{bmatrix}$$



$$\begin{pmatrix} (a + \beta_X M)^2 + (c + \beta_Y M)^2 + (e + \beta_Z M)^2 & (a + \beta_X M)^2 + (c + \beta_Y M)^2 \\ (a + \beta_X M)^2 + (c + \beta_Y M)^2 & (a + \beta_X M)^2 + (c + \beta_Y M)^2 + (e + \beta_Z M)^2 \end{pmatrix}$$

# Expected **DZ** covariance matrix

$$\begin{pmatrix} (a + \beta_X M)^2 + (c + \beta_Y M)^2 + (e + \beta_Z M)^2 & 0.5(a + \beta_X M)^2 + (c + \beta_Y M)^2 \\ 0.5(a + \beta_X M)^2 + (c + \beta_Y M)^2 & (a + \beta_X M)^2 + (c + \beta_Y M)^2 + (e + \beta_Z M)^2 \end{pmatrix}$$

```

nv<-1
# Matrices to store a, c, and e Path Coefficients
pathA  <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=TRUE, values=.6, label="a11", name="a" )
pathC  <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=TRUE, values=.6, label="c11", name="c" )
pathE  <- mxMatrix( type="Full", nrow=nv, ncol=nv, free=TRUE, values=.6, label="e11", name="e" )
# Matrices to store the moderated a, c, and e Path Coefficients
modPathA <- mxMatrix( type='Full', nrow=nv, ncol=nv, free=TRUE, values=.6, label="aM11", name="aM" )
modPathC <- mxMatrix( type='Full', nrow=nv, ncol=nv, free=TRUE, values=.6, label="cM11", name="cM" )
modPathE <- mxMatrix( type='Full', nrow=nv, ncol=nv, free=TRUE, values=.6, label="eM11", name="eM" )
# Matrix for the moderator variable
mod <- mxMatrix( type="Full", nrow=1, ncol=1, free=FALSE, labels=c("data.ses"), name="D" )
# Matrices to compute the moderated A, C, and E variance components
covAmod <- mxAlgebra( expression=(a+ D%*%aM) %*% t(a+ D%*%aM), name="A" )
covCmod <- mxAlgebra( expression=(c+ D%*%cM) %*% t(c+ D%*%cM), name="C" )
covEmod <- mxAlgebra( expression=(e+ D%*%eM) %*% t(e+ D%*%eM), name="E" )
# Algebra for the expected mean vector and expected variance/covariance matrices and in MZ and DZ
covMZ <- mxAlgebra( expression= rbind ( cbind(A+C+E , 0.5%x%A+C),
                                         cbind(0.5%x%A+C, A+C+E)), name="expCovDZ" )

```



# Expected **DZ** covariance matrix

OpenMx

```
nv<-1
```

```
# Algebra for the expected mean vector and expected variance/covariance matrices and in MZ and DZ
covMZ <- mxAlgebra( expression= rbind ( cbind(A+C+E , 0.5%x%A+C),
                                         cbind(0.5%x%A+C, A+C+E)), name="expCovDZ" )
```

$$A = (a_{11} + data.ses * aM11)^2$$

$$C = (c_{11} + data.ses * cM11)^2$$

$$E = (e_{11} + data.ses * eM11)^2$$

$$\text{expCovDZ} = \begin{bmatrix} (a_{11} + data.ses * aM11)^2 + (c_{11} + data.ses * cM11)^2 + (e_{11} + data.ses * eM11)^2 & 0.5 * (a_{11} + data.ses * aM11)^2 + (c_{11} + data.ses * cM11)^2 \\ 0.5 * (a_{11} + data.ses * aM11)^2 + (c_{11} + data.ses * cM11)^2 & (a_{11} + data.ses * aM11)^2 + (c_{11} + data.ses * cM11)^2 + (e_{11} + data.ses * eM11)^2 \end{bmatrix}$$



$$\begin{pmatrix} (a + \beta_X M)^2 + (c + \beta_Y M)^2 + (e + \beta_Z M)^2 & 0.5(a + \beta_X M)^2 + (c + \beta_Y M)^2 \\ 0.5(a + \beta_X M)^2 + (c + \beta_Y M)^2 & (a + \beta_X M)^2 + (c + \beta_Y M)^2 + (e + \beta_Z M)^2 \end{pmatrix}$$

# Making plots

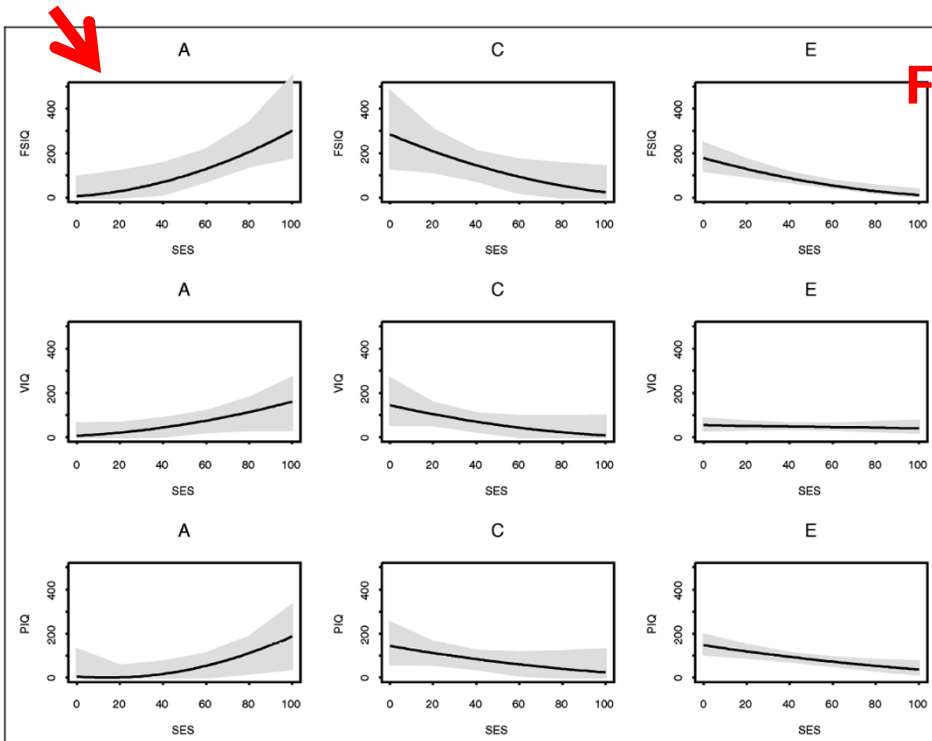
- Linear effect of SES on *path loadings*
- Non-linear effect of SES on *unstandardized variance components*
- Non-linear effect of SES on *standardized variance components*

# Example Turkheimer study

Moderation of **unstandardized** variance components:

Moderation of **standardized** variance components:

**FSIQ**



**FSIQ**

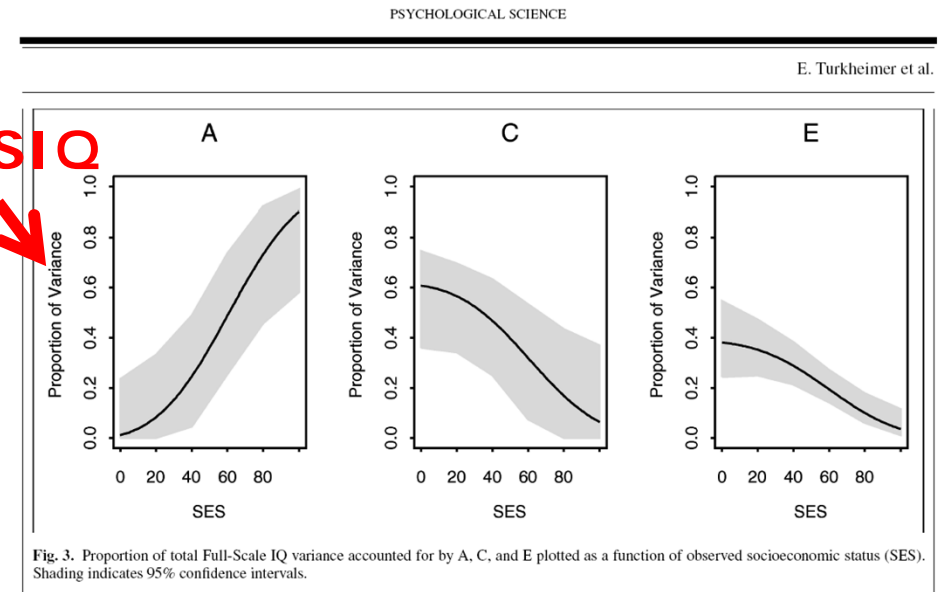
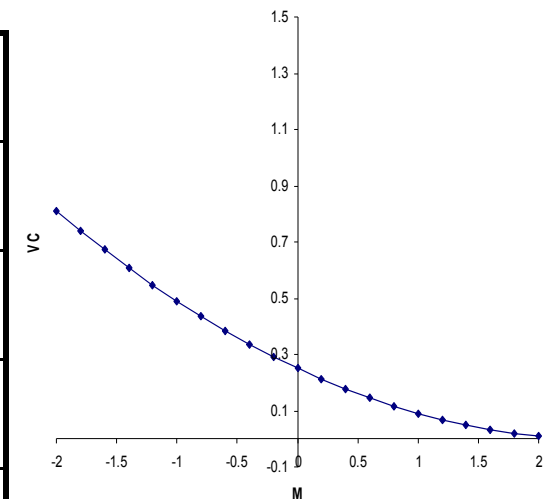


Fig. 3. Proportion of total Full-Scale IQ variance accounted for by A, C, and E plotted as a function of observed socioeconomic status (SES). Shading indicates 95% confidence intervals.

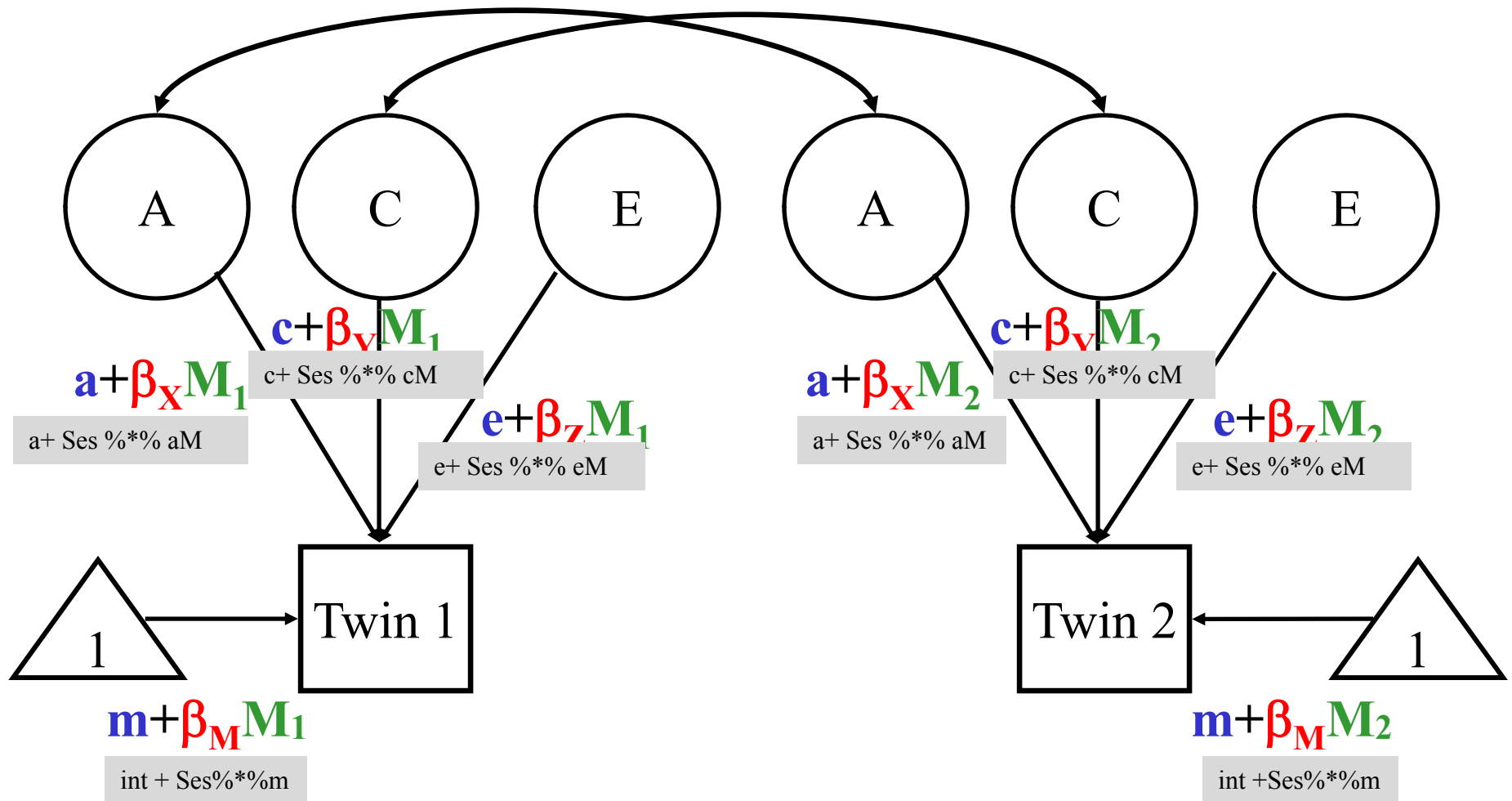
# Calculate it yourself, or plot it in R!

- Moderation of the additive genetic VC:
  - From OpenMx:  $a=0.5$  ;  $aM=-0.2$
  - Range moderator: -2 to 2

<i>SES</i>	$(a+SES*aM)^2$	$(a+SES*aM)^2$
-2	$(0.5+(-2*-0.2))^2$	0.81
-1.5	$(0.5+(-1.5*-0.2))^2$	0.73
...		
+2	$(0.5+(2*-0.2))^2$	0.01



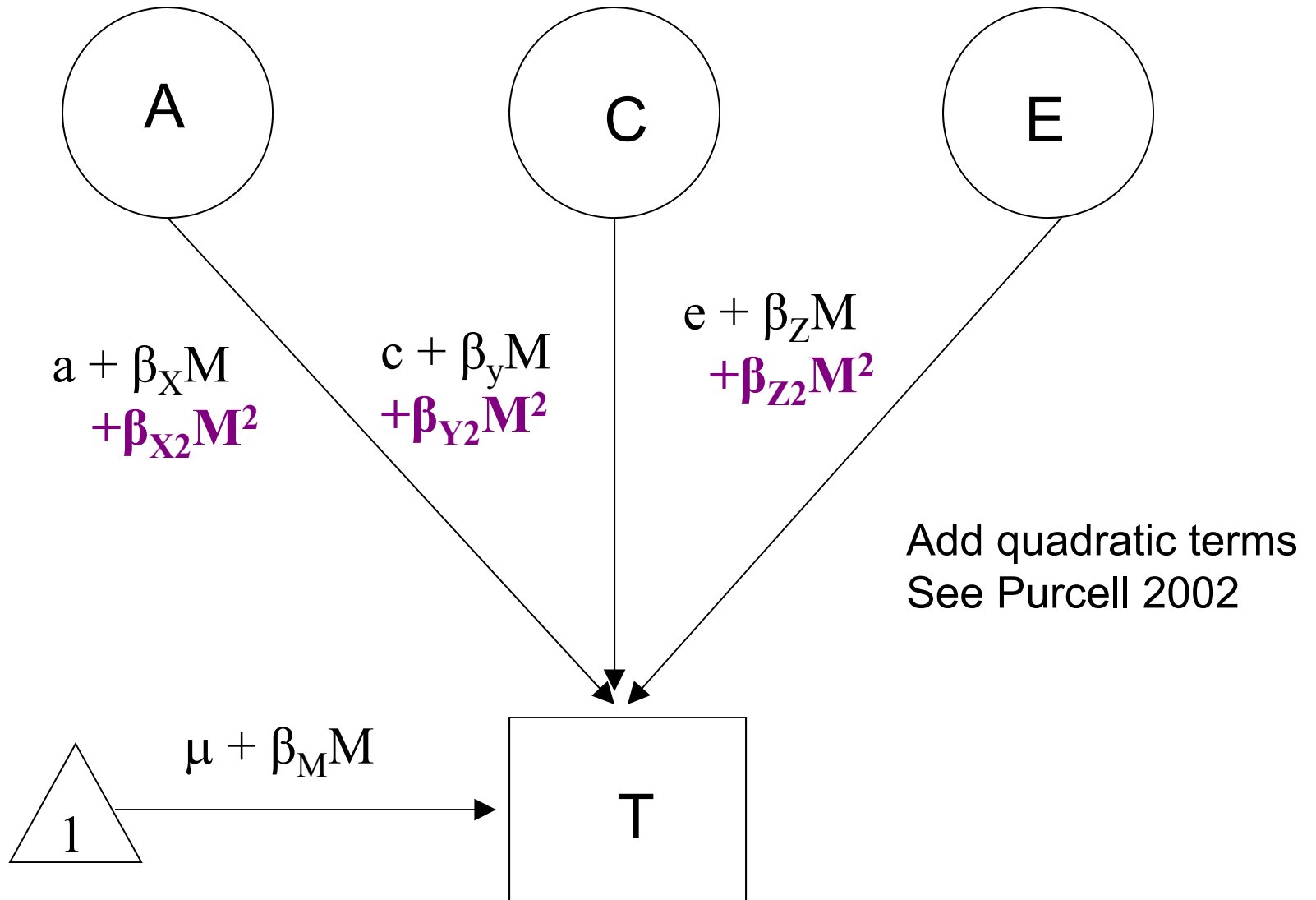
# Path model vs. OpenMx matrices



# More advanced models

- Nonlinear moderation
- GxE for categorical data
- GxE in the context of rGE

# Nonlinear moderation



# GxE for categorical data

- Continuous data
  - Moderation of means and variances
- Ordinal data
  - Moderation of thresholds and variances

Behav Genet (2009) 39:220–229  
DOI 10.1007/s10519-008-9247-7

---

BRIEF COMMUNICATION

## **A Note on the Parameterization of Purcell's $G \times E$ Model for Ordinal and Binary Data**

Sarah E. Medland · Michael C. Neale ·  
Lindon J. Eaves · Benjamin M. Neale



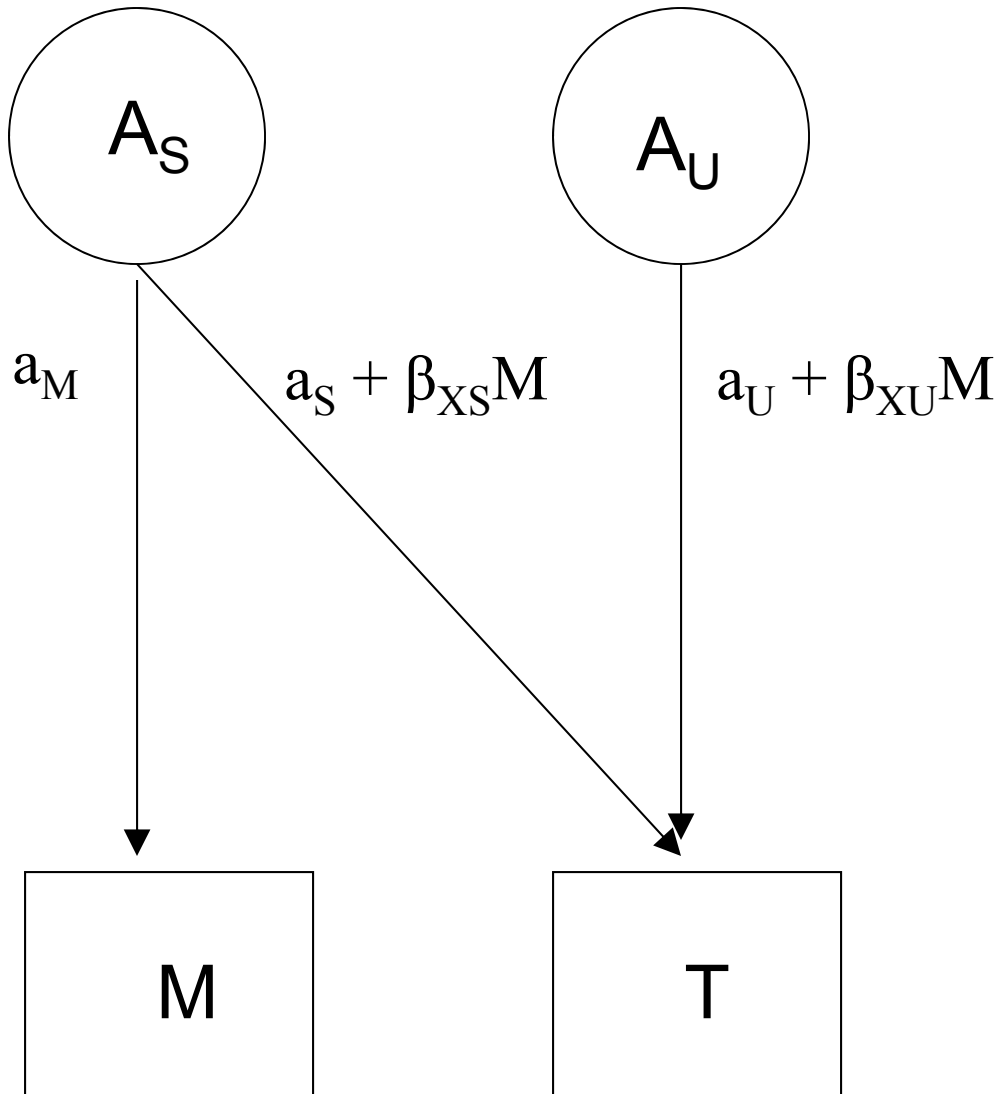
# GxE in the context of rGE

- If there is a correlation between the moderator (environment) of interest and the outcome, and you find a GxE effect, it's not clear if:
  - The environment is moderating the effects of genes
- Or:
  - Trait-influencing genes are simply more likely to be present in that environment

# Ways to deal with rGE

- Limit study to moderators that aren't correlated with outcome
  - Pro: easy
  - Con: not very satisfying
- Moderator in means model will remove from the covariance genetic effects shared by trait and moderator
  - Pro: Any interaction detected will be moderation of the trait specific genetic effects
  - Con: Will fail to detect GxE interaction if the moderated genetic component is shared by the outcome and moderator
- Explicitly model rGE using a bivariate framework

# Bivariate model



$\beta_{XS}$  indicates moderation of shared genetic effects

$\beta_{XU}$  indicates moderation of unique genetic effects on trait of interest

See Johnson, 2007

**NOTE:** this model is not informative for family-level variables (e.g., ses, parenting, etc)

# Practical

- Replicate findings from Turkheimer et al. with twin data from NTR
- Phenotype: FSIQ
- Moderator: SES
- Data: 205 MZ and 225 DZ twin pairs
- 5 years old