Please copy the files for this session

```
mkdir ~/doc
cp  -r  /home/madhur/2024/Day-5/DOC/*  ~/doc/.
```

Twin-based
Direction of Causation (DOC) Modeling

International Statistical Genetics Workshop

IBG, University of Colorado Boulder

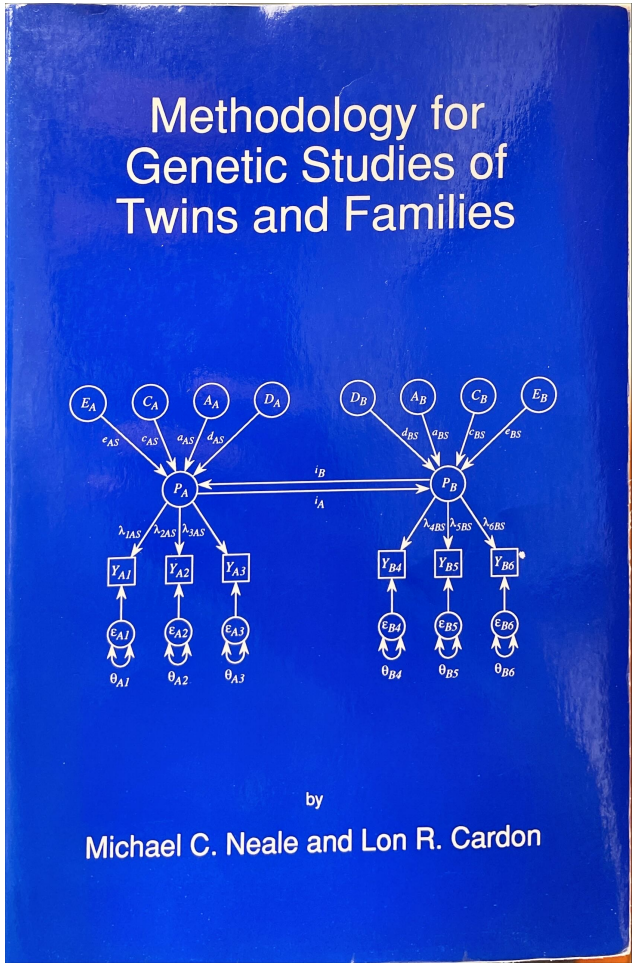
Friday, 8th March 2024

MadhurBain Singh

singhm18@vcu.edu

Virginia Institute for Psychiatric and Behavior Genetics

Virginia Commonwealth University, Richmond, VA



Thanks to Nathan Gillespie, Hermine H. M. Maes, and Michael C. Neale.

Twin-based DOC modeling in SEM

- Causal inference in cross-sectional data from twins
- Extension of bivariate/multivariate twin design

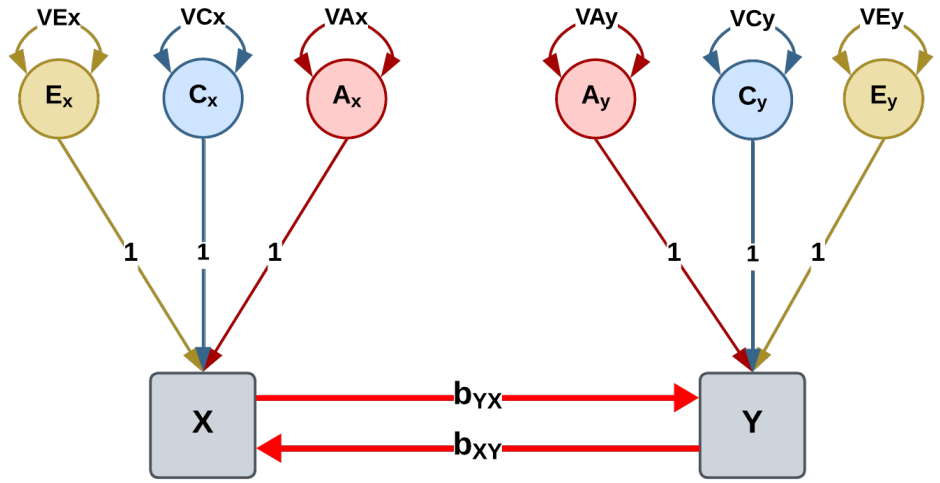
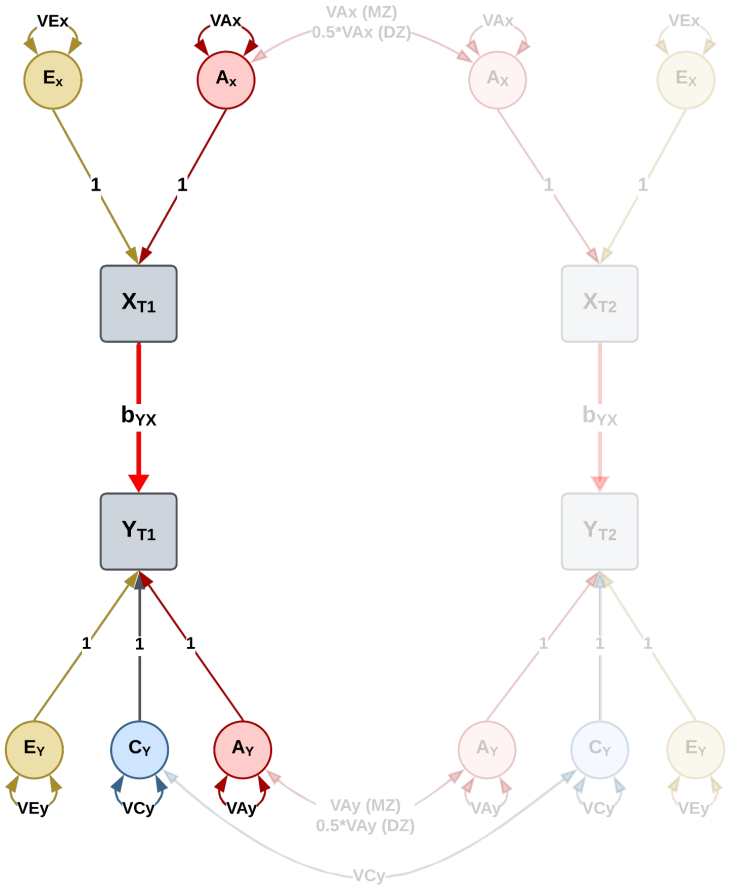


*Diagram showing only the within-individual part of the model.

Power based on the information from the **cross-twin cross-trait covariance**.

Neale and Cardon (1992); Heath et al. (1993); Duffy and Martin (1994).

Causal effect of X on Y within an individual



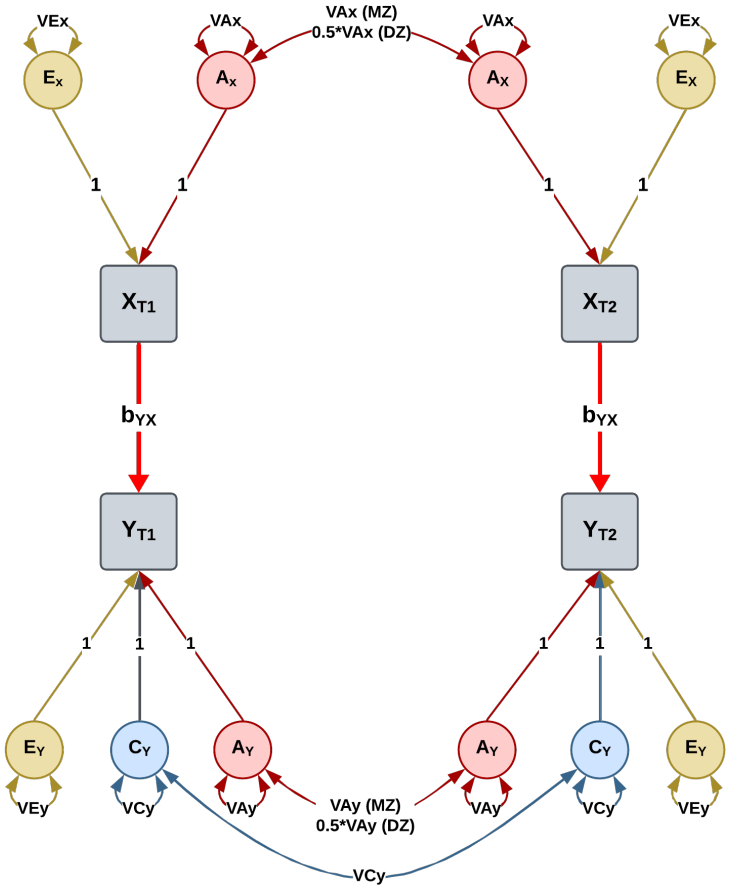
Note

When we model/estimate an effect of *X* on *Y*, the variable *X* need not have a "direct" effect on *Y*.

There may be a chain of unmeasured variables that partly or fully mediate the effect of *X* on *Y*.

This is true for all models of causal inference.

Assumption: The causal process is identical in both siblings.

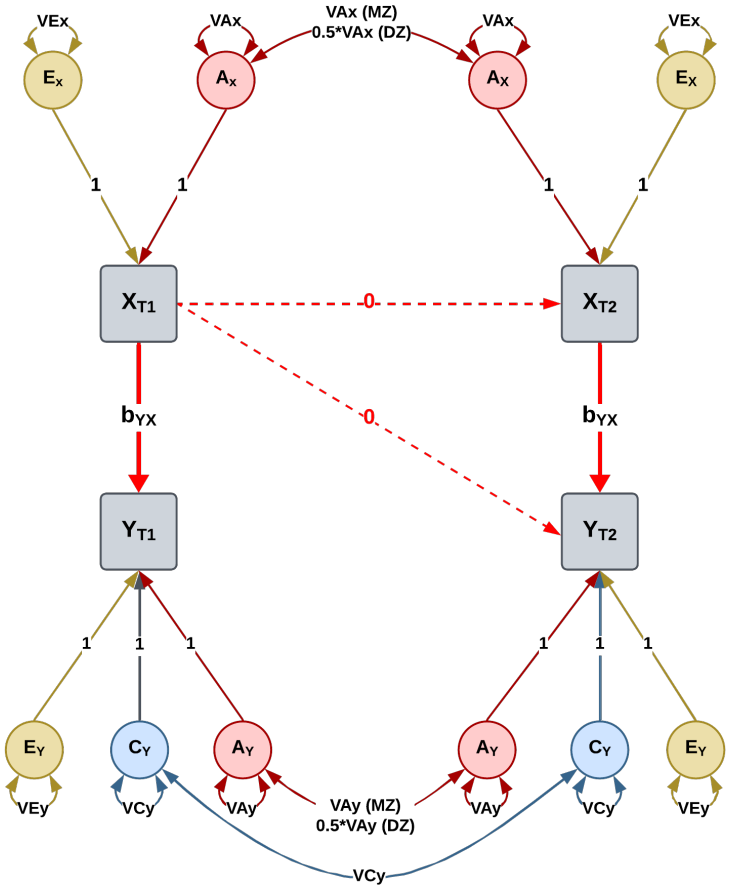


Note

Almost all* causal models assume a homogeneous causal process in the population under study.

*unless explicitly specified.

Assumption: There are no within- or cross-trait sibling interactions.

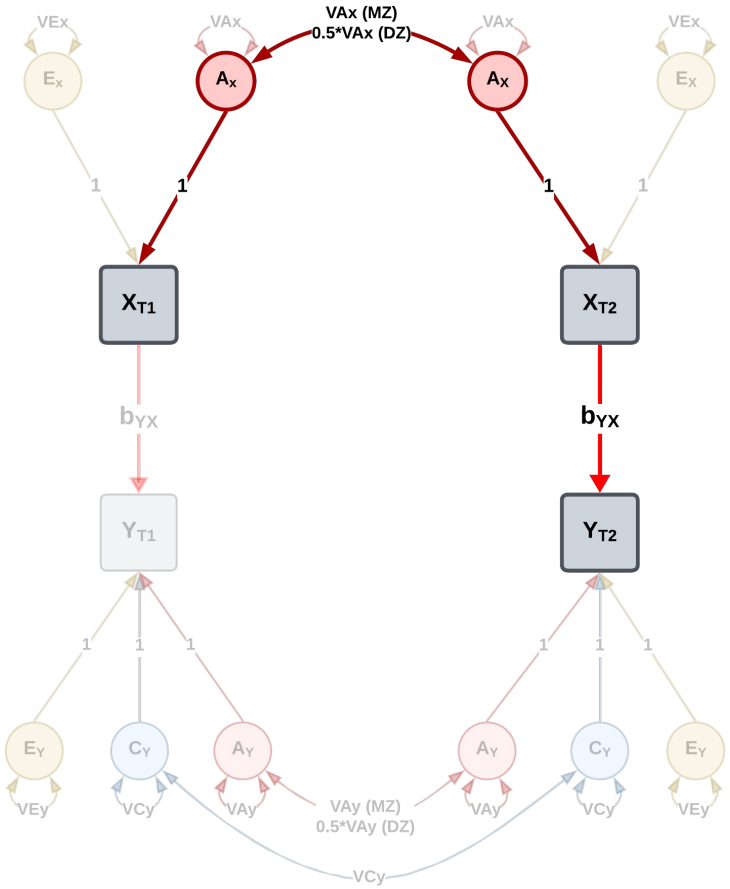


Twin-1's trait *X* has no causal influence on Twin-2's trait *X* or trait *Y*, and *vice versa*.

(e.g., sibling cooperation or rivalry)

Model-implied cross-twin cross-trait covariance

X causes *Y*

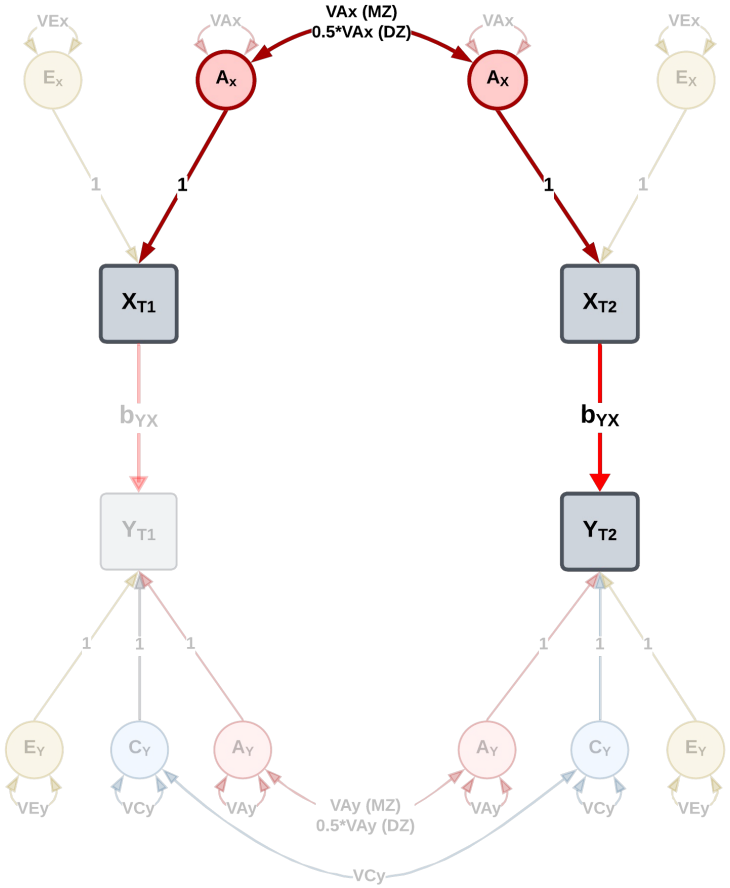


rMZ = VAx * **byx**

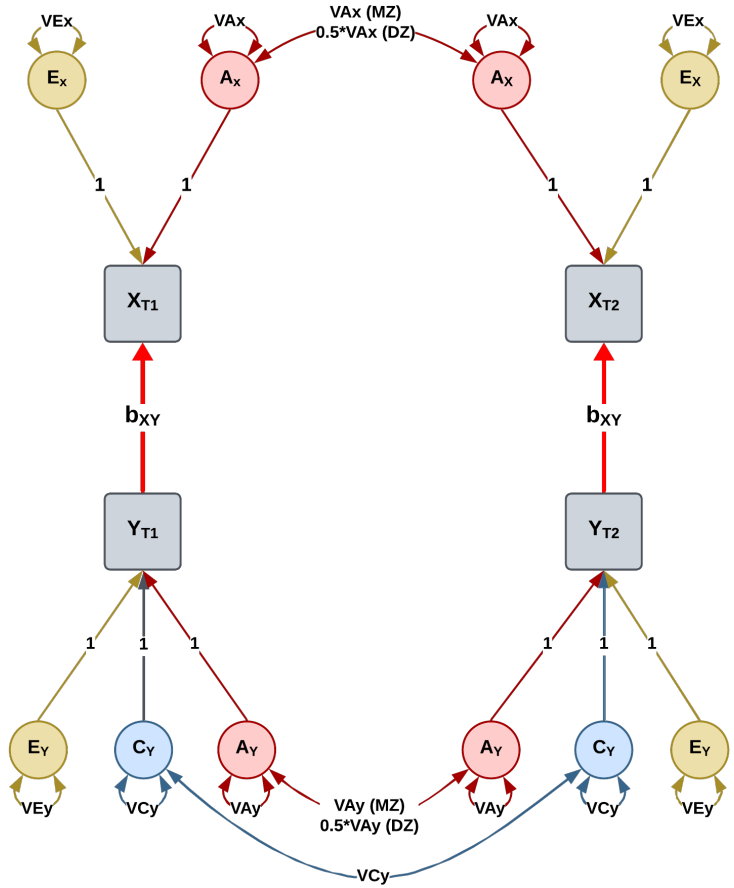
rDZ = ½ VAx * **byx**

Model-implied cross-twin cross-trait covariance

X causes *Y*



Y causes *X*

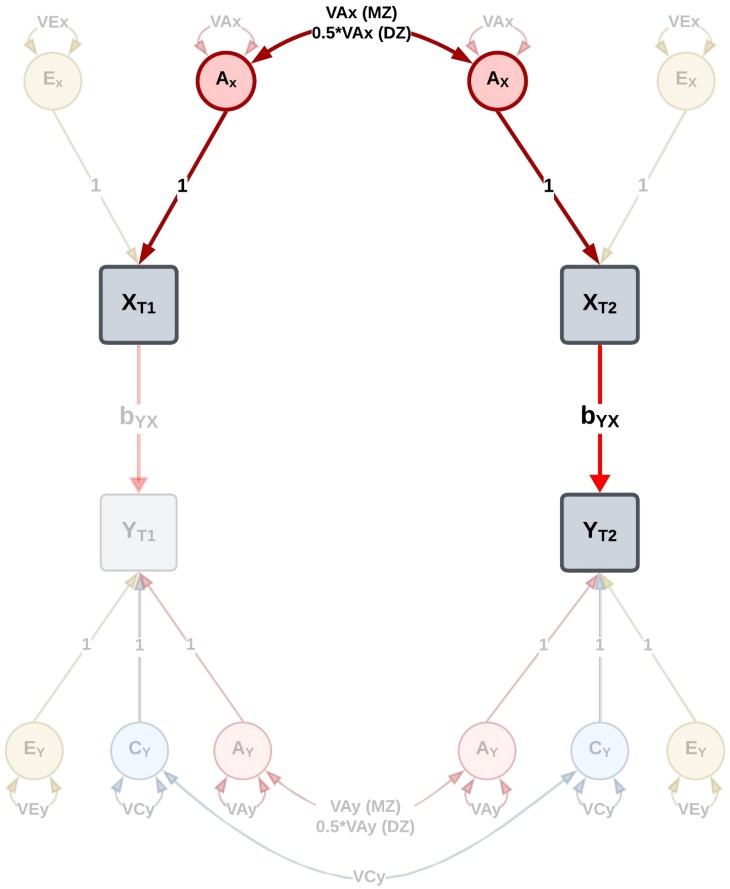


rMZ = VAx * **byx**

rDZ = ½ VAx * **byx**

Model-implied cross-twin cross-trait covariance

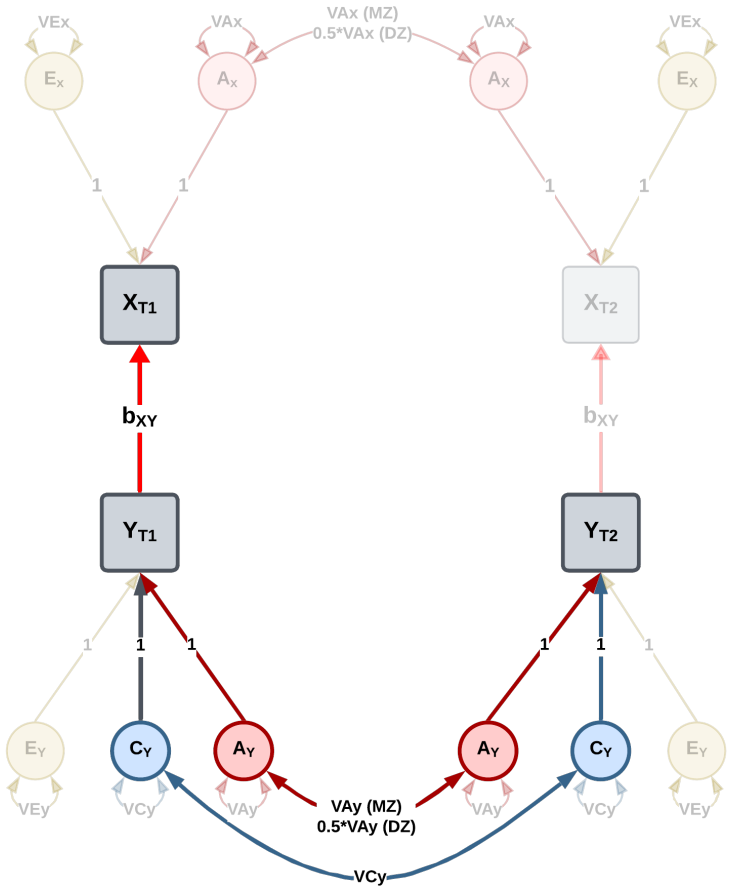
X causes *Y*



rMZ = VAx * **byx**

rDZ = ½ VAx * **byx**

Y causes *X*



rMZ = (VAy + VCy) * **bxy**

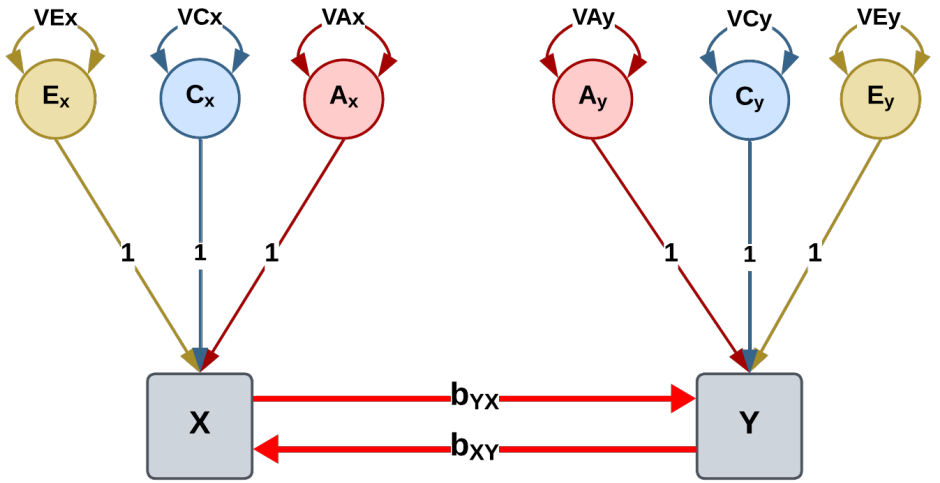
rDZ = (½ VAy + VCy) * **bxy**

DOC depends on the two traits having different twin-pair correlations

- In other words, DOC works if the two traits have different ACE/ADE estimates.
- If the two traits have the same (or very similar*) ACE/ADE estimates, the cross-twin cross-trait covariance is the same (or very similar*) under the two directions of causal effects.
- So, the model has no (or very little*) power to differentiate between the two causal processes.
  - i.e., the model has a similar fit under either direction of causation.

For the DOC model to work,

the ACE estimates of *X* must be different from the ACE estimates of *Y*.

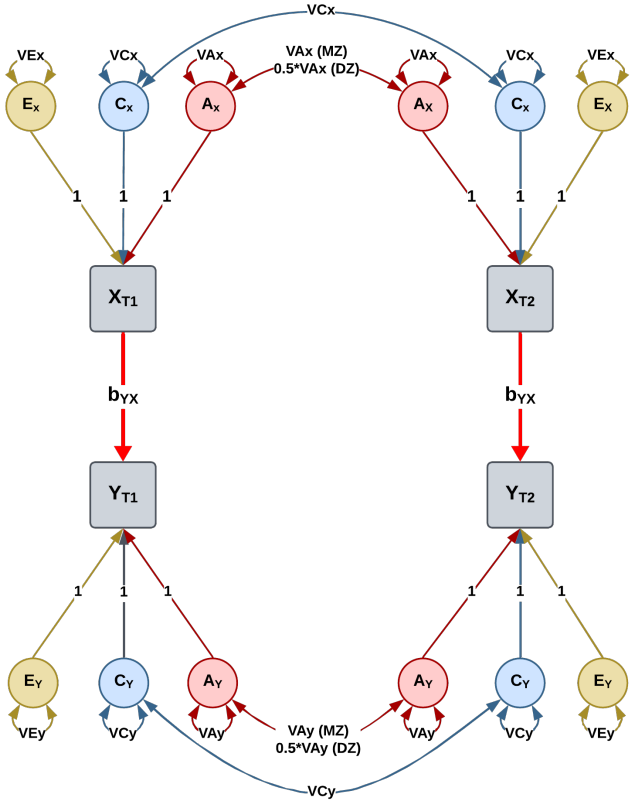


**If the ACE estimates of X and Y are very similar (but not the same), the DOC model would require extremely large sample sizes to differentiate between the two directions of causation.*

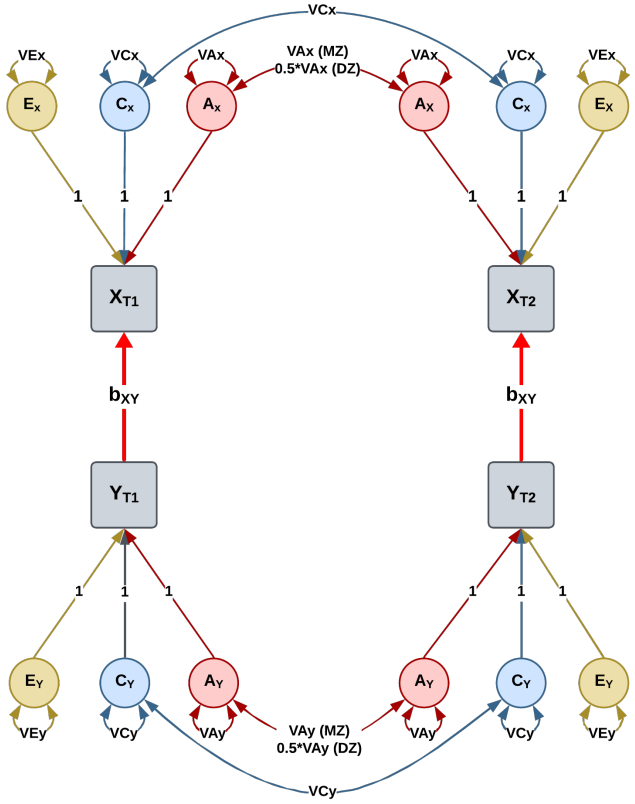
Neale and Cardon (1992); Heath et al. (1993); Duffy and Martin (1994).

Model-implied cross-twin cross-trait covariance

X causes *Y*



Y causes *X*



Both *X* and *Y* have ACE variance decomposition.

Exercise: What is the expected cross-twin cross-trait covariance when

Trait X

VAx = 0.3

VCx = 0.4

VEx = 0.3

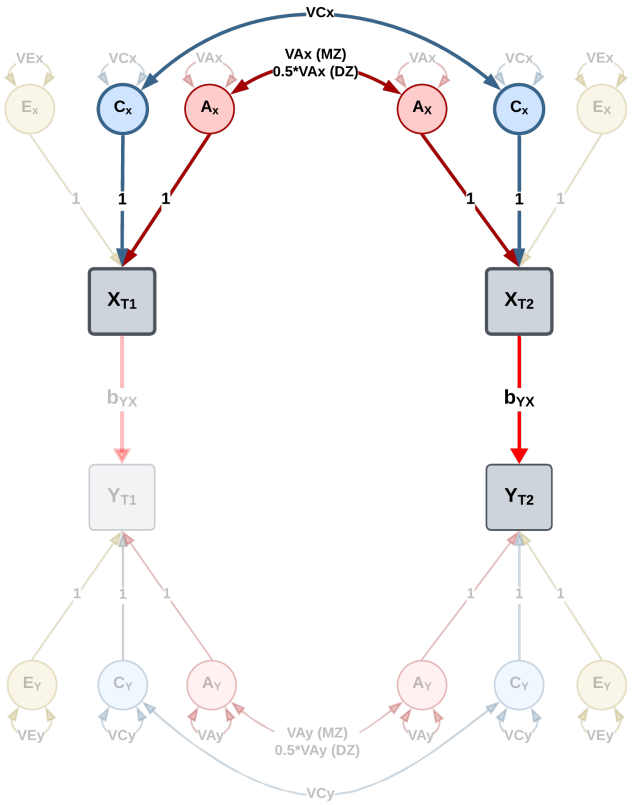
Trait Y

VAy = 0.4

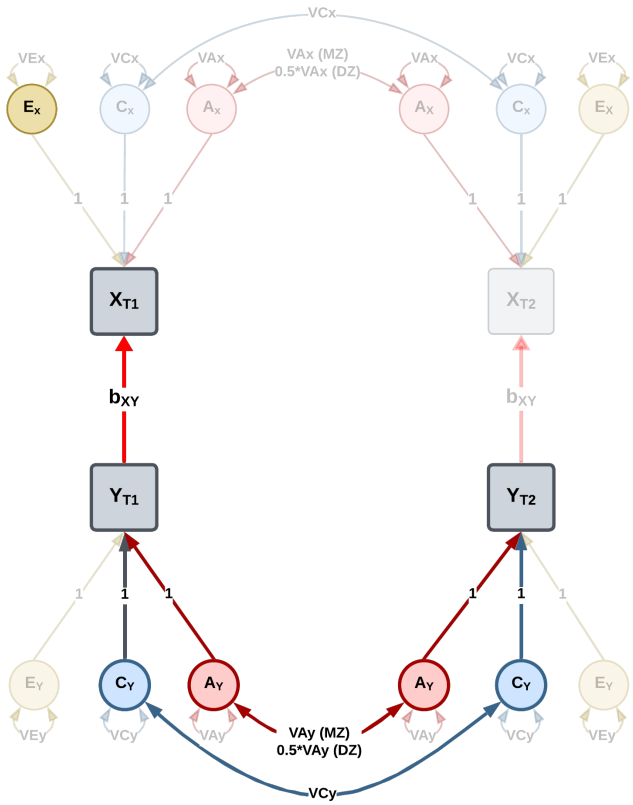
VCy = 0.1

VEy = 0.5

X causes Y



Y causes X



b = 0.5

rMZ = (**VAx** + **VCx**) * **b**

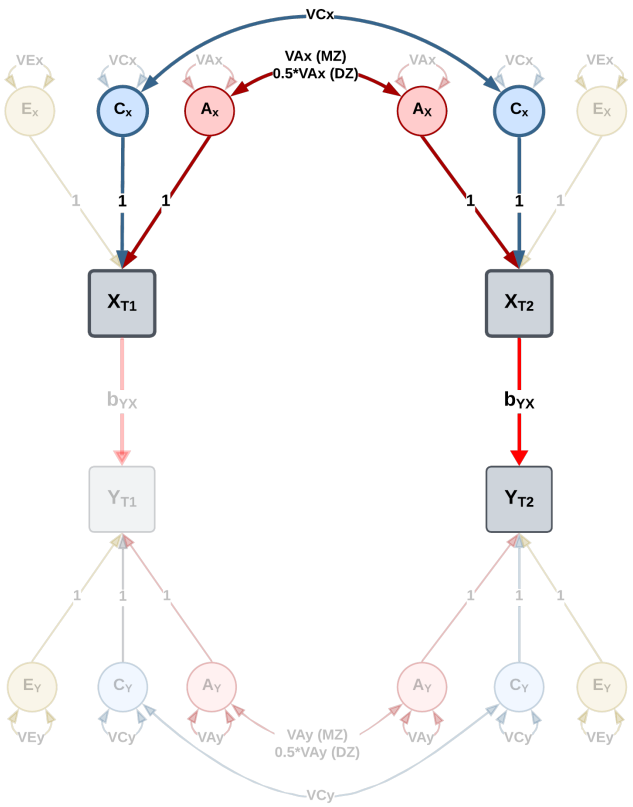
rDZ = (½*VAx + **VCx**) * **b**

rMZ = (**VAy** + **VCy**) * **b**

rDZ = (½*VAy + **VCy**) * **b**

Exercise: What is the expected cross-twin cross-trait covariance when

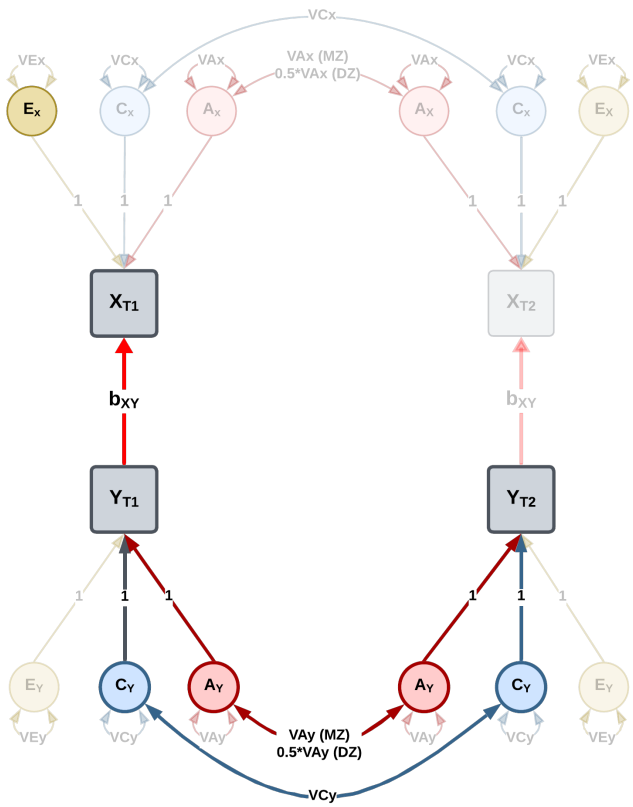
X causes *Y*



rMZ = (**0.3** + **0.4**) * 0.5

rDZ = (½*__0.3__ + **0.4**) * 0.5

Y causes *X*

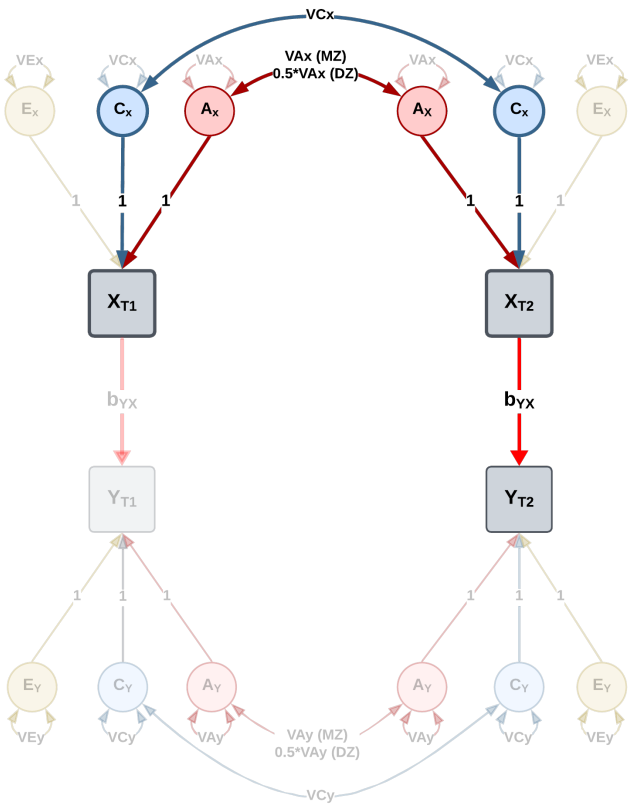


rMZ = (**0.4** + **0.1**) * 0.5

rDZ = (½*__0.4__ + **0.1**) * 0.5

Exercise: What is the expected cross-twin cross-trait covariance when

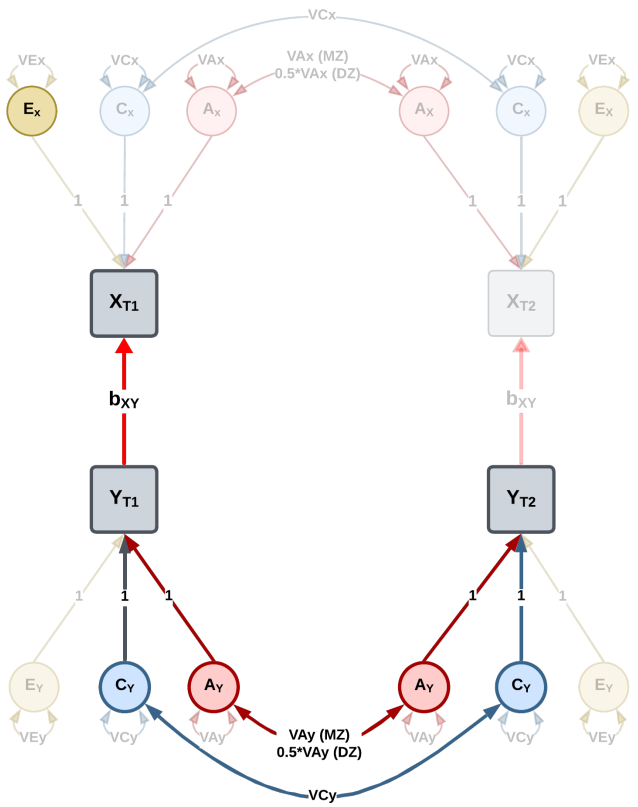
X causes *Y*



rMZ = **0.350**

rDZ = **0.275**

Y causes *X*



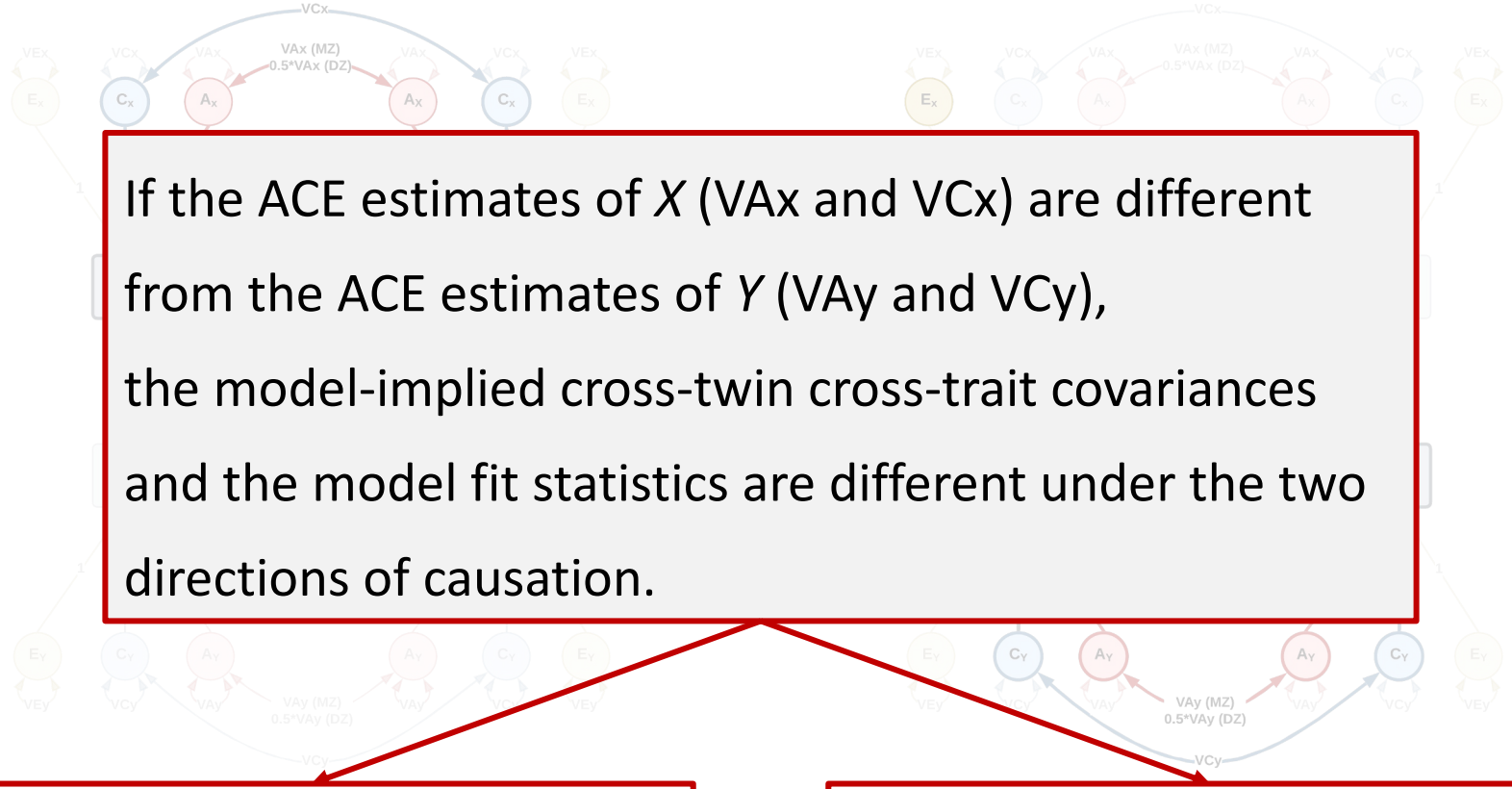
rMZ = **0.250**

rDZ = **0.150**

Compare the model fit statistics with different directions of causation specified

X* causes *Y

Y* causes *X



If the ACE estimates of *X* (VAx and VCx) are different from the ACE estimates of *Y* (VAy and VCy), the model-implied cross-twin cross-trait covariances and the model fit statistics are different under the two directions of causation.

rMZ = (**VAx** + **VCx**) * byx

rDZ = (½ **VAx** + **VCx**) * byx

rMZ = (**VAy** + **VCy**) * bxy

rDZ = (½ **VAy** + **VCy**) * bxy

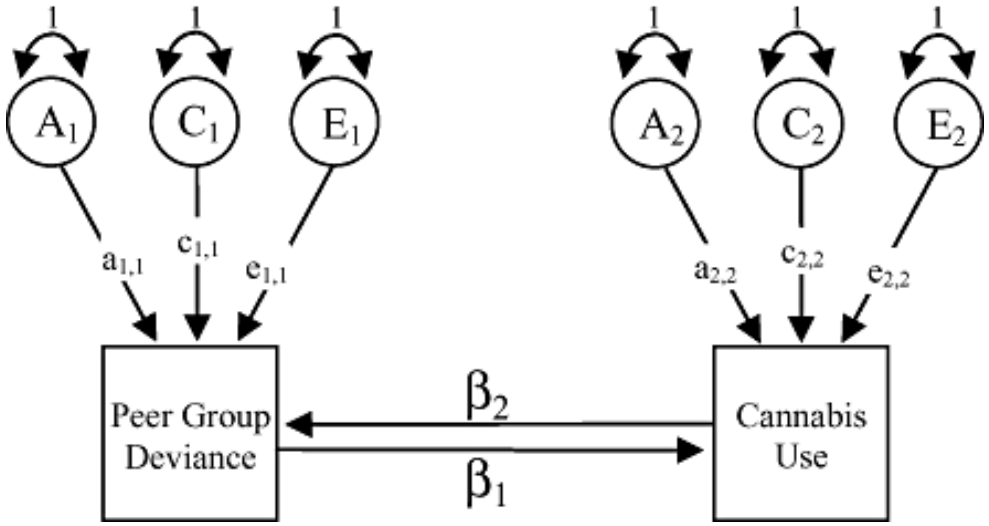
An applied example of the classic DOC model

Addiction
RESEARCH REPORT
doi:10.1111/j.1360-0443.2008.02457.x

Modeling the genetic and environmental association between peer group deviance and cannabis use in male twins

Nathan A. Gillespie[1], Michael C. Neale[1], Kristen Jacobson[2] & Kenneth S. Kendler[1,3]

Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA,[1] Department of Psychiatry, University of Chicago, Chicago, IL, USA[2] and Department of Human Genetics, Virginia Commonwealth University, Richmond, VA, USA[3]



	−2LL	*df*	*Δ-2LL*	*Δdf*	*P*	*BIC*
22–25 years						
Cholesky	7725.38	3542				−2805.25
PGD→CU	7744.32	3544	18.94	2.00	<0.001	−2799.54
CU→PGD	7729.08	3544	3.70	2.00	0.16	−2807.16
PGD↔CU	7728.00	3543	2.62	1.00	0.11	−2805.82

Causal estimates in DOC may be biased by differences in the measurement error of the two traits

- If the two traits have different levels of measurement error (i.e., reliability), the causal estimates in the standard DOC model (applied to observed phenotypes) will be biased.
- **Solutions**
  - *If* the reliability/measurement error of the phenotype is known, SEM allows for its specification in the model.
    - See the path diagrams and scripts in the MR-DOC session this afternoon.
  - More realistically, use latent "true" phenotypes derived from multiple indicators and fit the DOC model to latent phenotypes.

Neale and Cardon (1992); Heath et al. (1993).

DOC model between latent phenotypes with multiple indicators

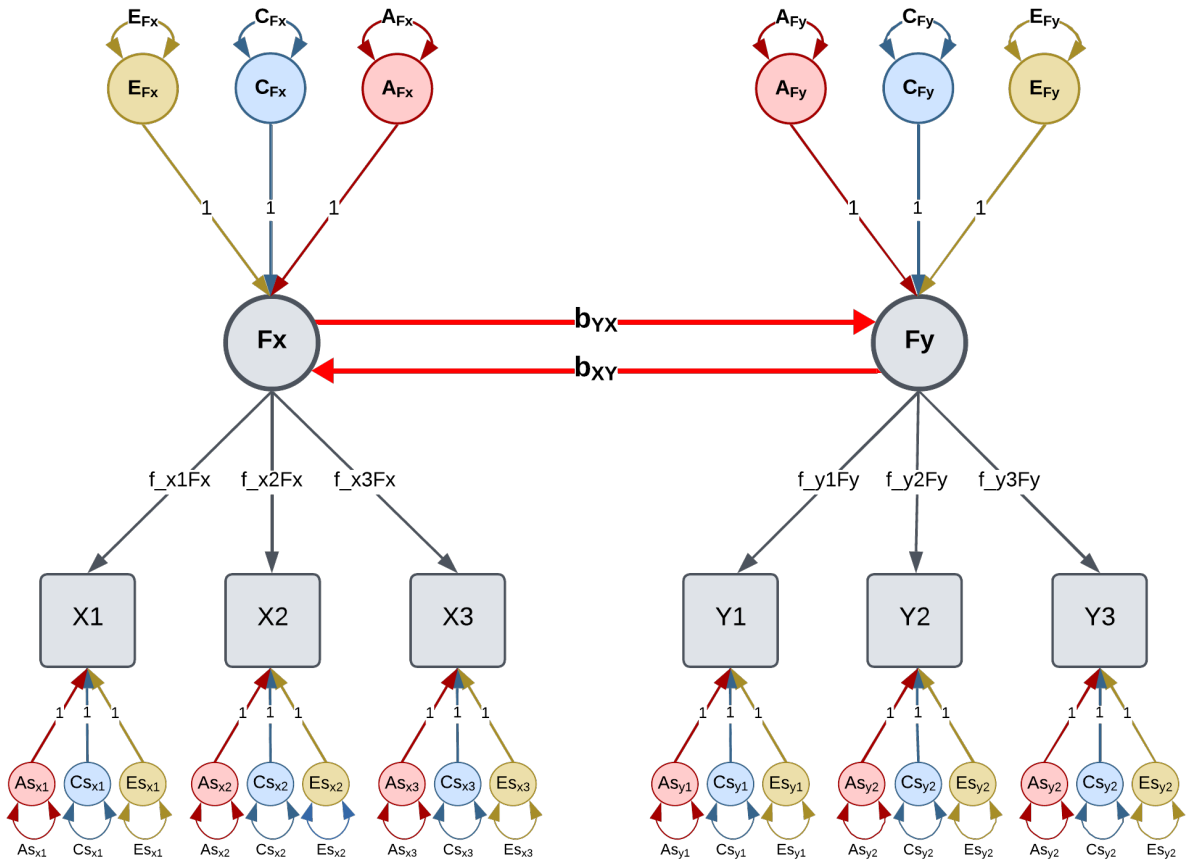
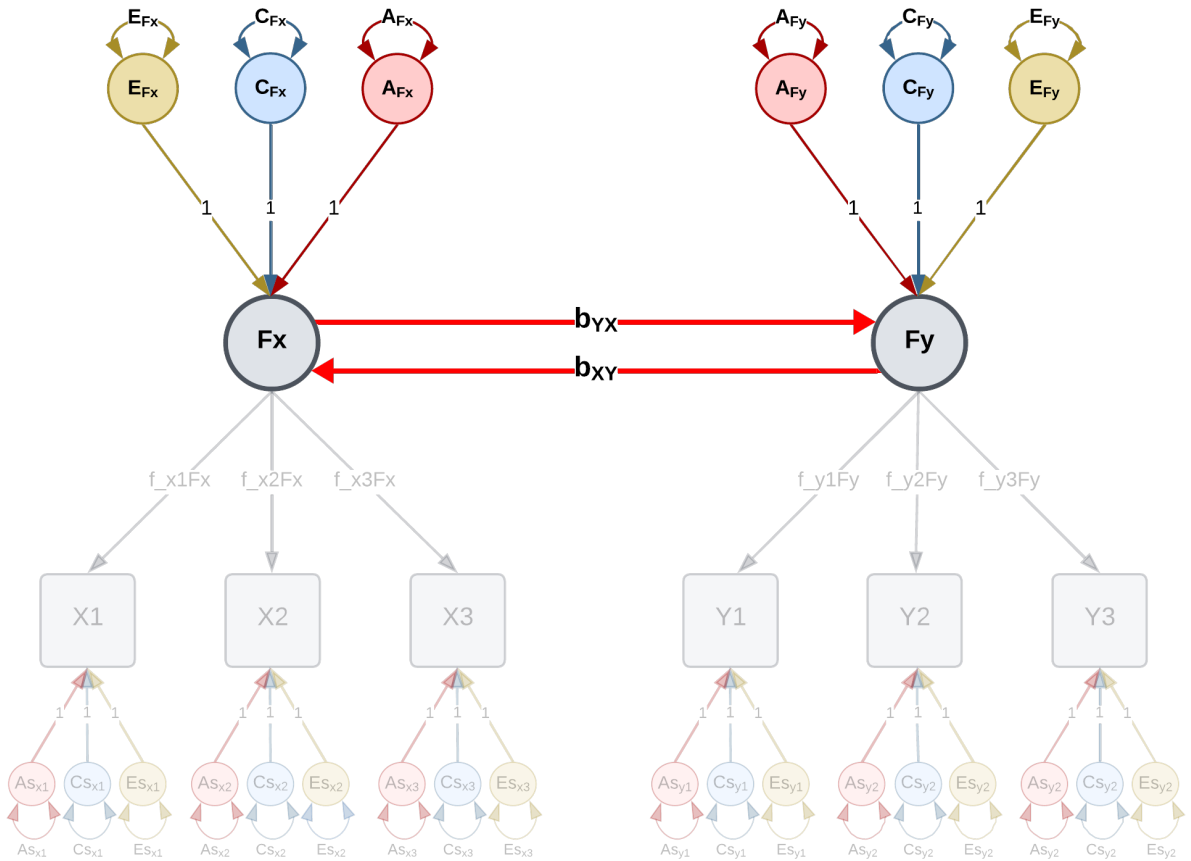


Diagram showing only the within-individual part of the model.

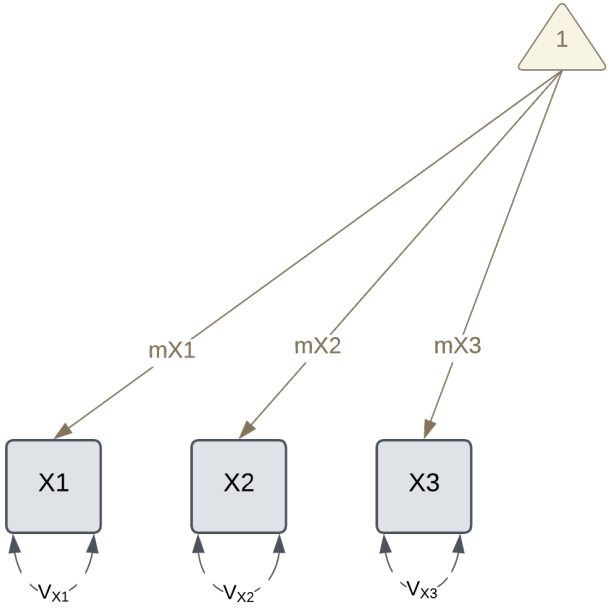
Neale and Cardon (1992); Heath et al. (1993)

DOC model between latent phenotypes with multiple indicators



The DOC model of interest is fitted to latent phenotypes (that have no measurement error).

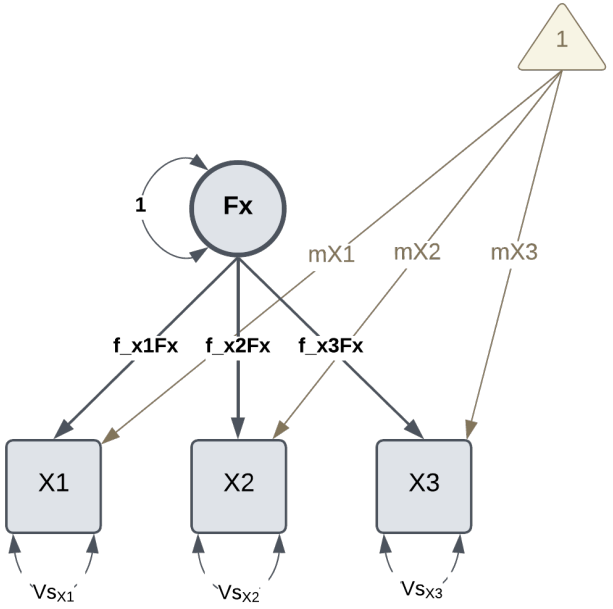
Multiple indicators/items of a phenotype



For example, X_1, X_2, and X_3 may be

- Symptoms of depression (e.g., CIDI-SF)
- Symptoms of alcohol use disorder (e.g., AUDIT)
- Externalizing behaviors in children/adolescents (e.g., the Child Behavior Checklist)
- Measures of cognitive functions in older populations (e.g., the Mini-Mental Status Examination)

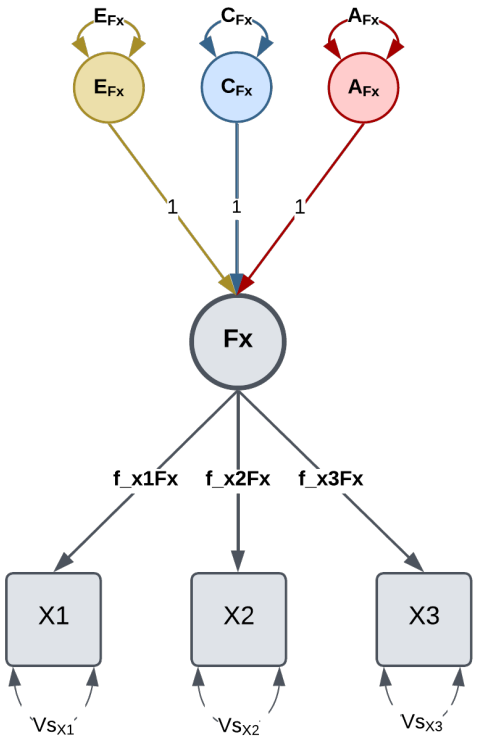
Common latent factor: “True phenotype” from multiple indicators/items



The common latent factor reflects the shared variance between the different items/indicators.

Saturated means model

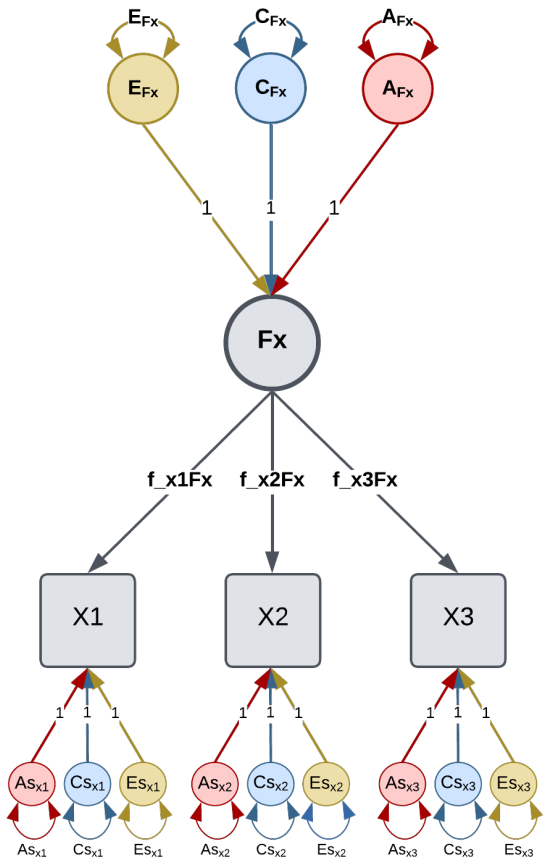
ACE variance decomposition of the latent phenotype (no measurement error)



Measurement error is captured by the indicator-specific variance not explained by the common factor (true phenotype).

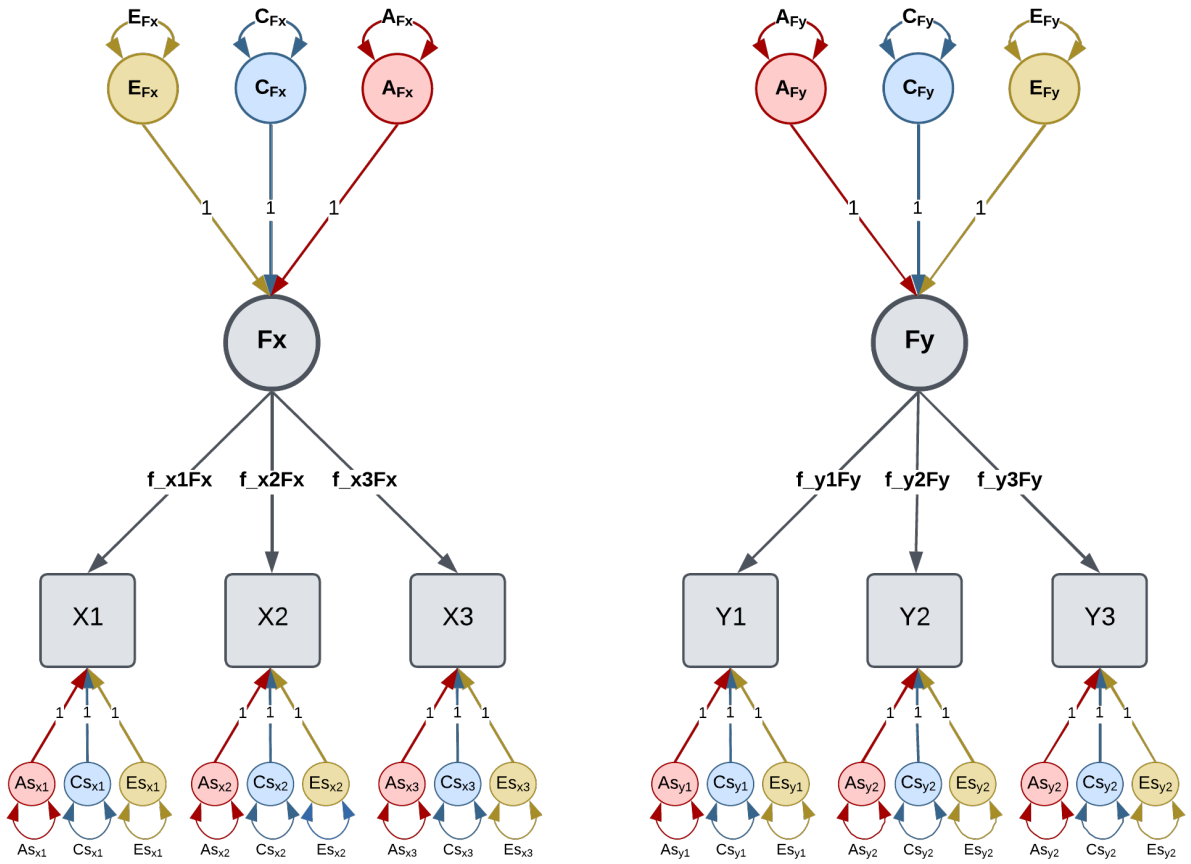
To improve figure readability, the means are not shown.

Plus, ACE decomposition of indicator-specific (residual) variances

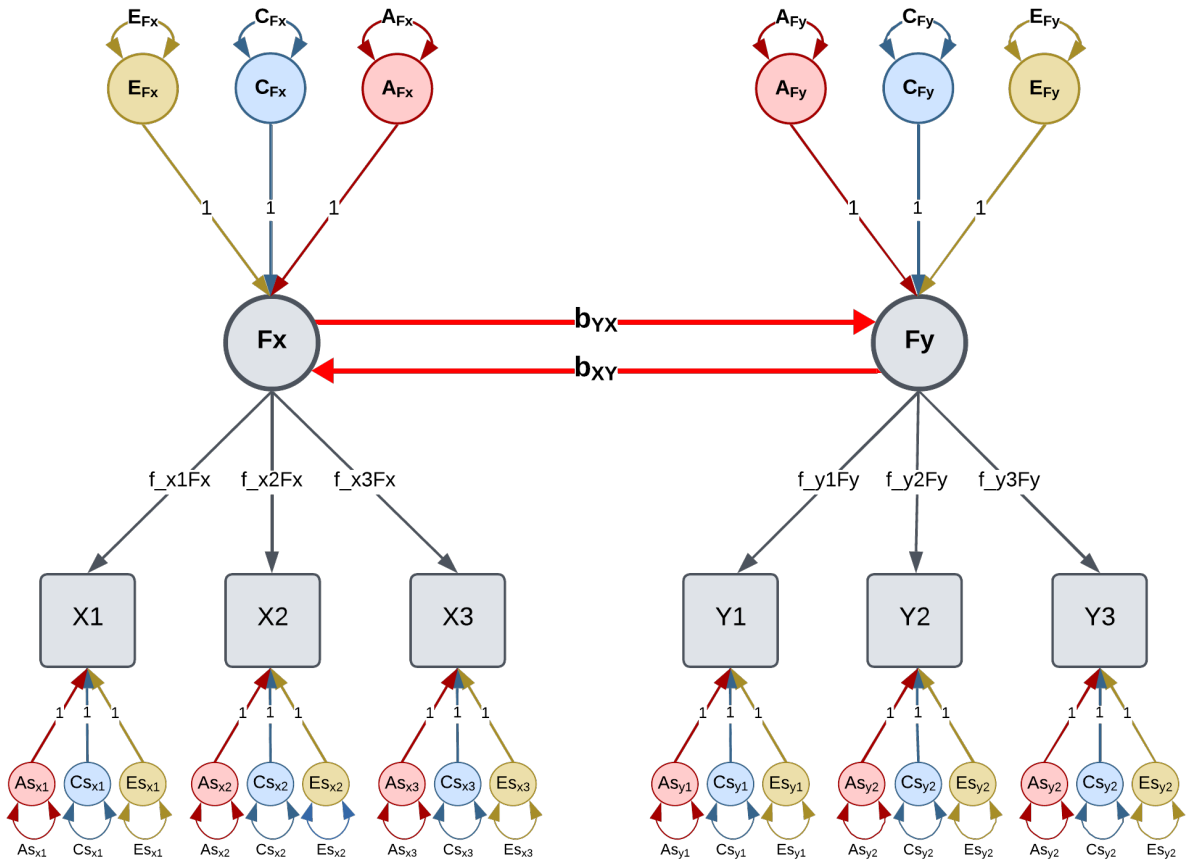


Measurement error is captured by the *Es_{X1}, Es_{X2}, and Es_{X3}* variance components (of the indicator-specific variances).

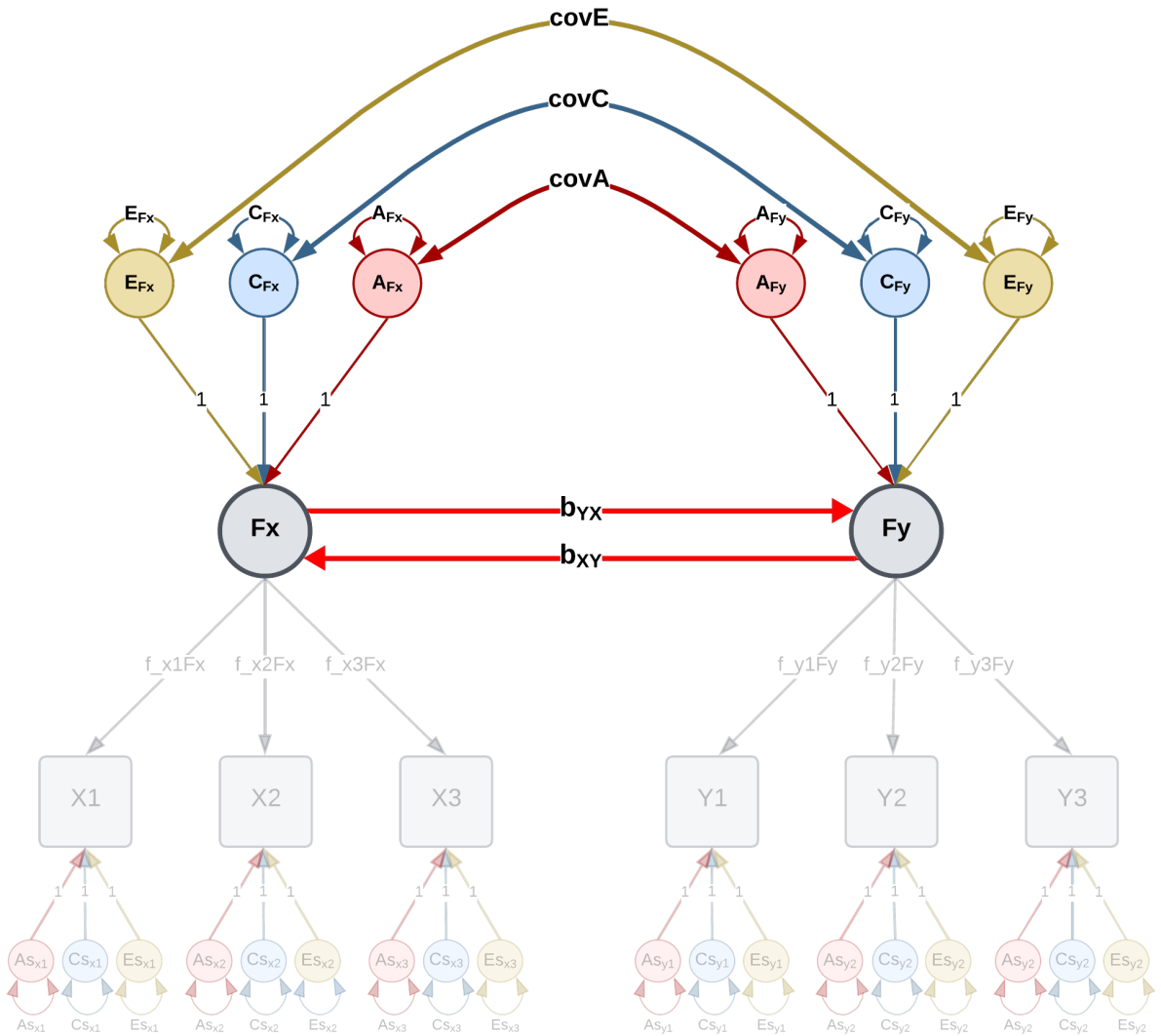
Two latent phenotypes (*X* and *Y*)



Direction-of-causation modeling with two latent phenotypes



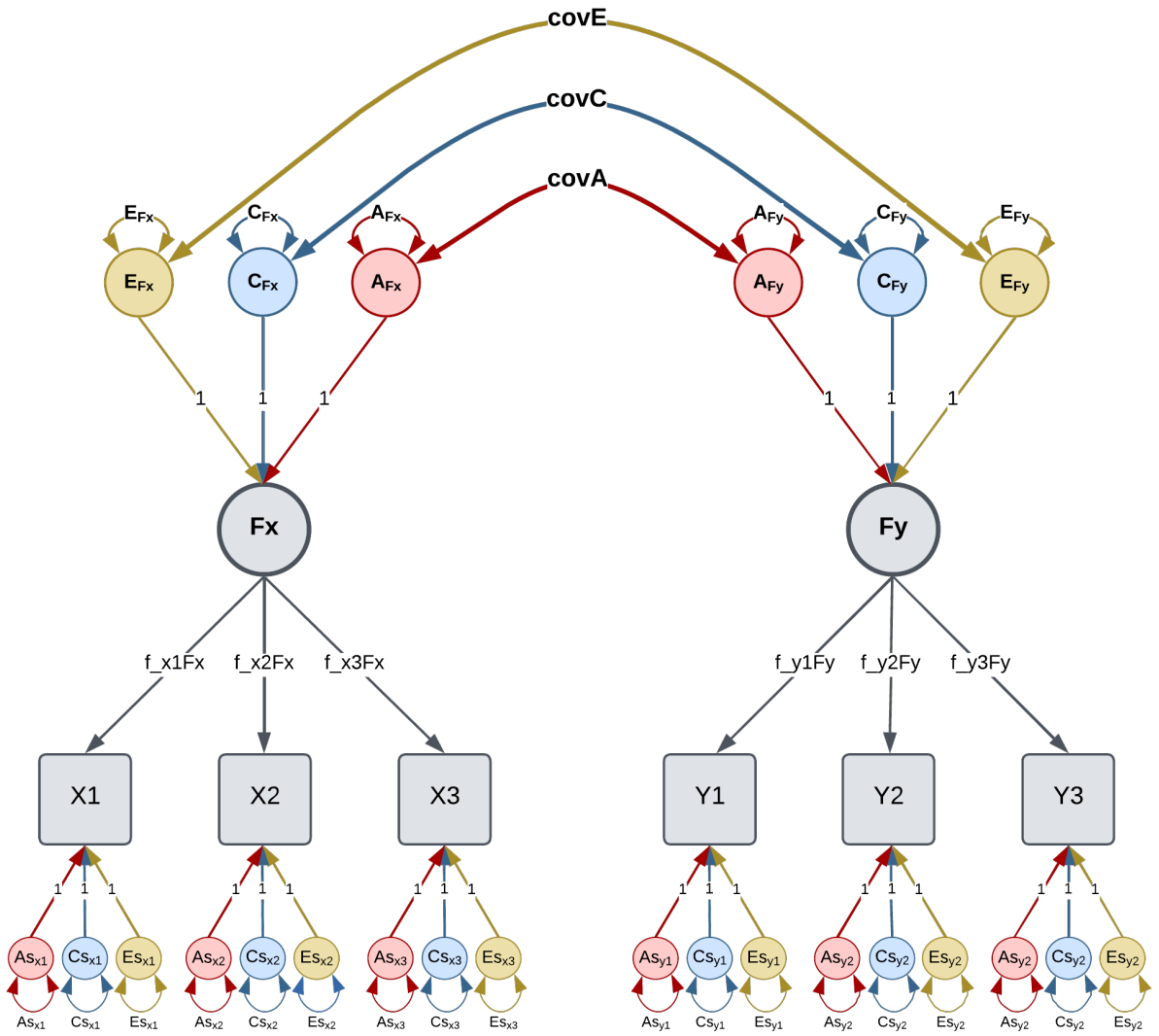
Five possible sources of covariance in the model



The model, as shown, is not identified.

We cannot estimate all five sources of covariance at the same time in this model.

ACE [“confounding-only”] model with two latent phenotypes

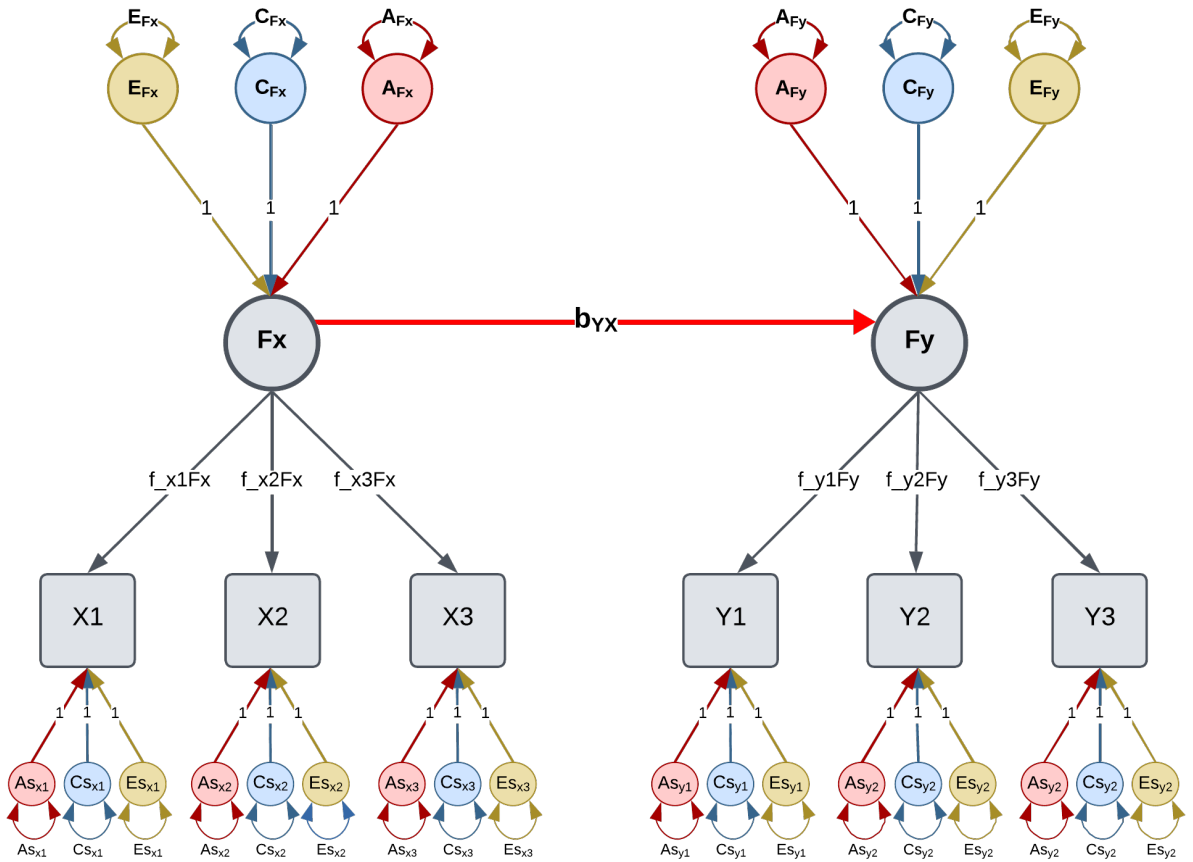


Null hypothesis for DOC

The observed associations between the indicators of X and Y are due to

- *Background confounding due to unmeasured genetic and/or environmental factors*
- *No causation between X and Y*

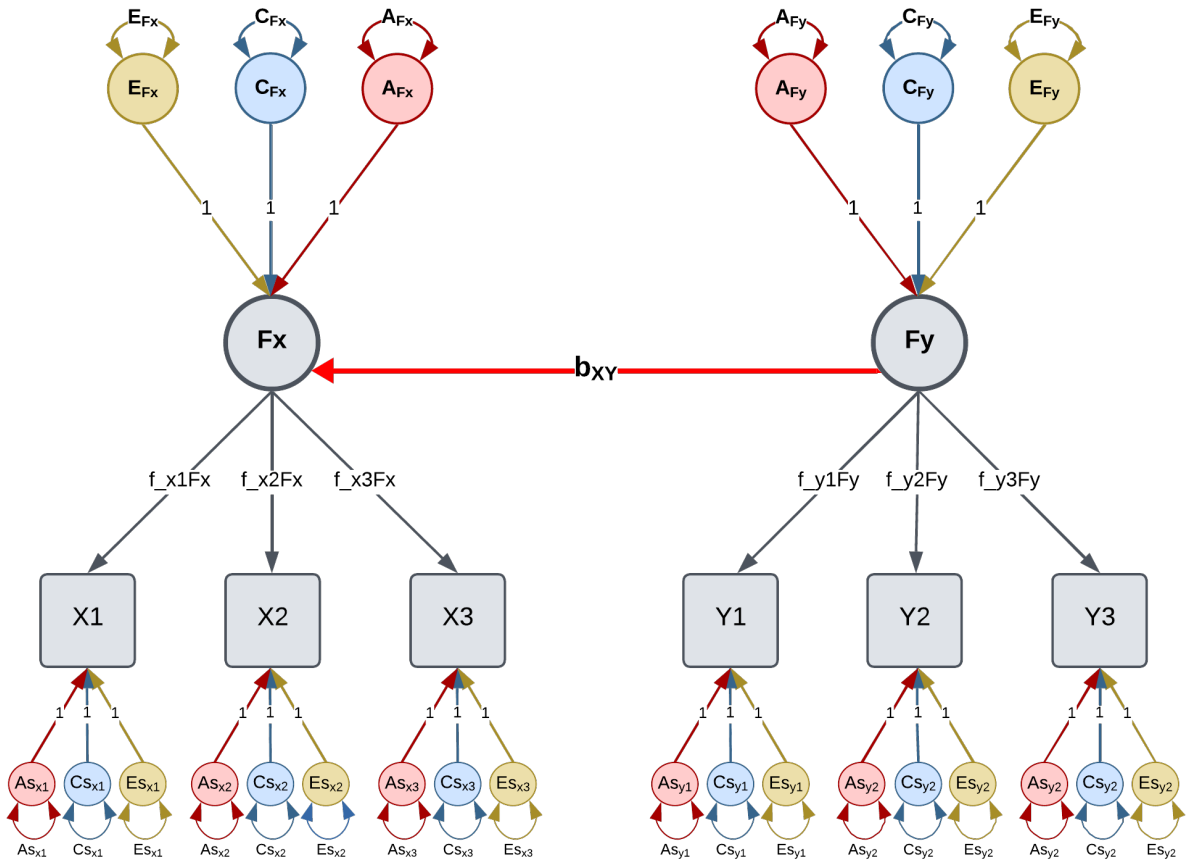
DOC model: *X* causes *Y*



Alternative hypothesis 1

- *X causes Y*
- *No background confounding due to unmeasured variables*

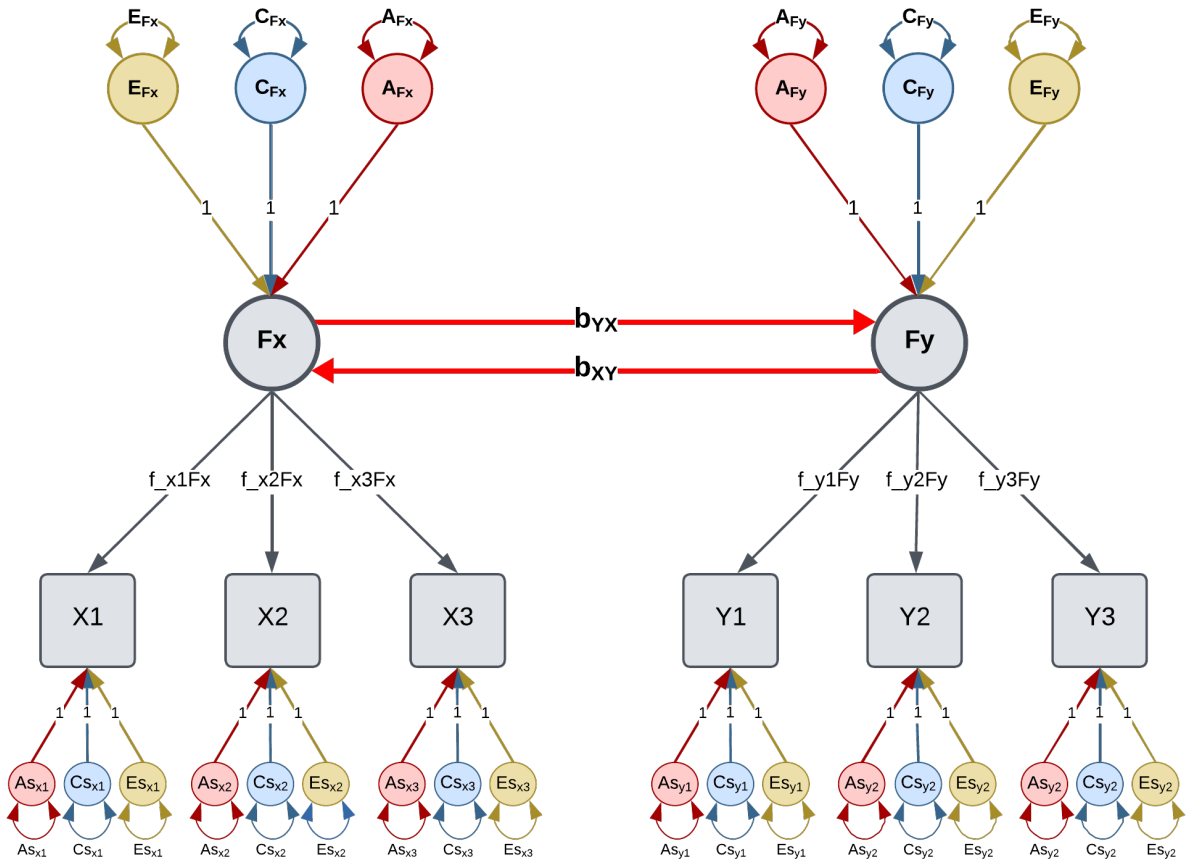
DOC model: *Y* causes *X*



Alternative hypothesis 2

- *Y causes X*
- *No background confounding due to unmeasured variables*

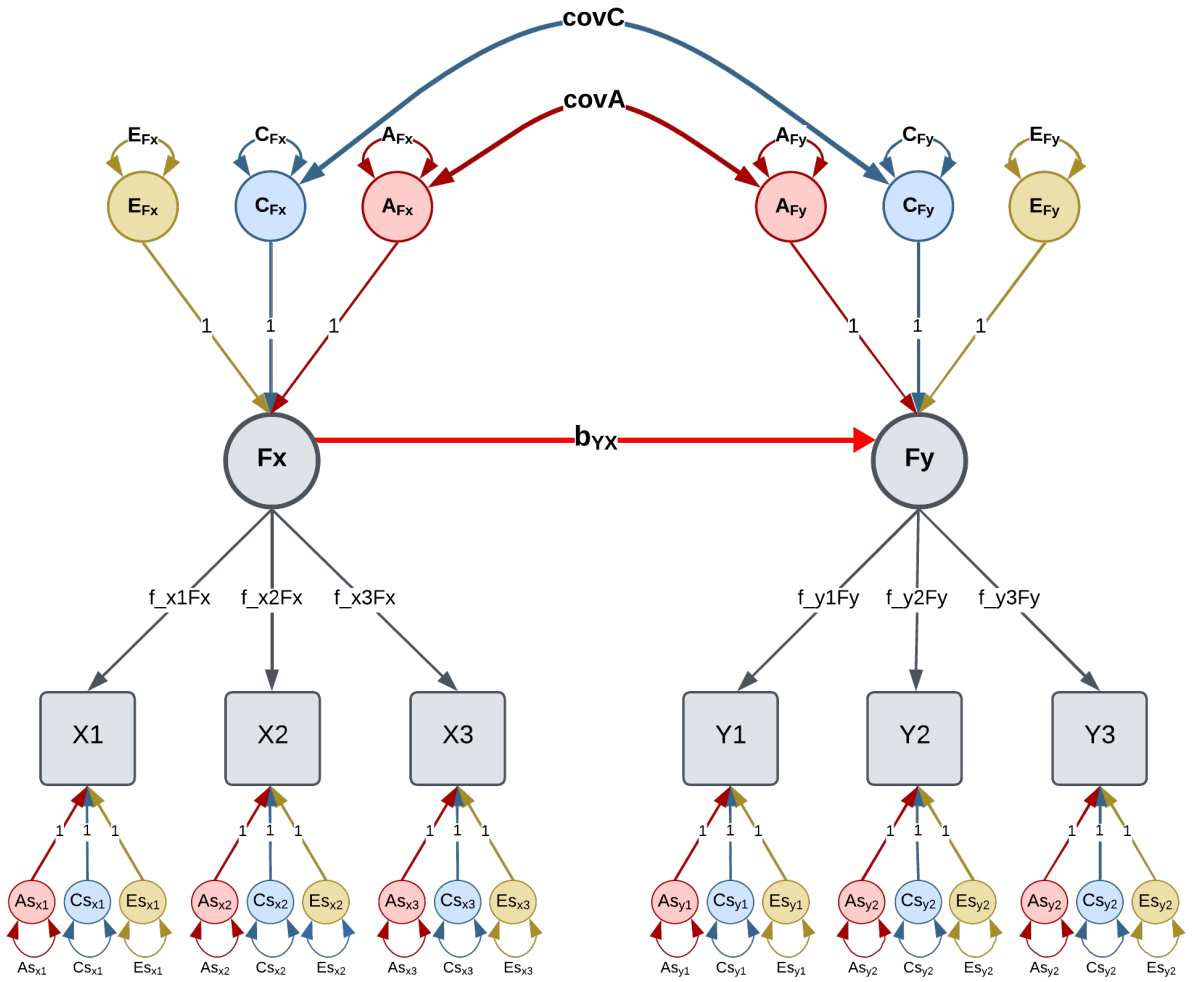
DOC model: Reciprocal causation between *X* and *Y*



Alternative hypothesis 3

- *Reciprocal causation between X and Y*
- *No background confounding due to unmeasured variables*

Hybrid DOC model: Causation + Confounding



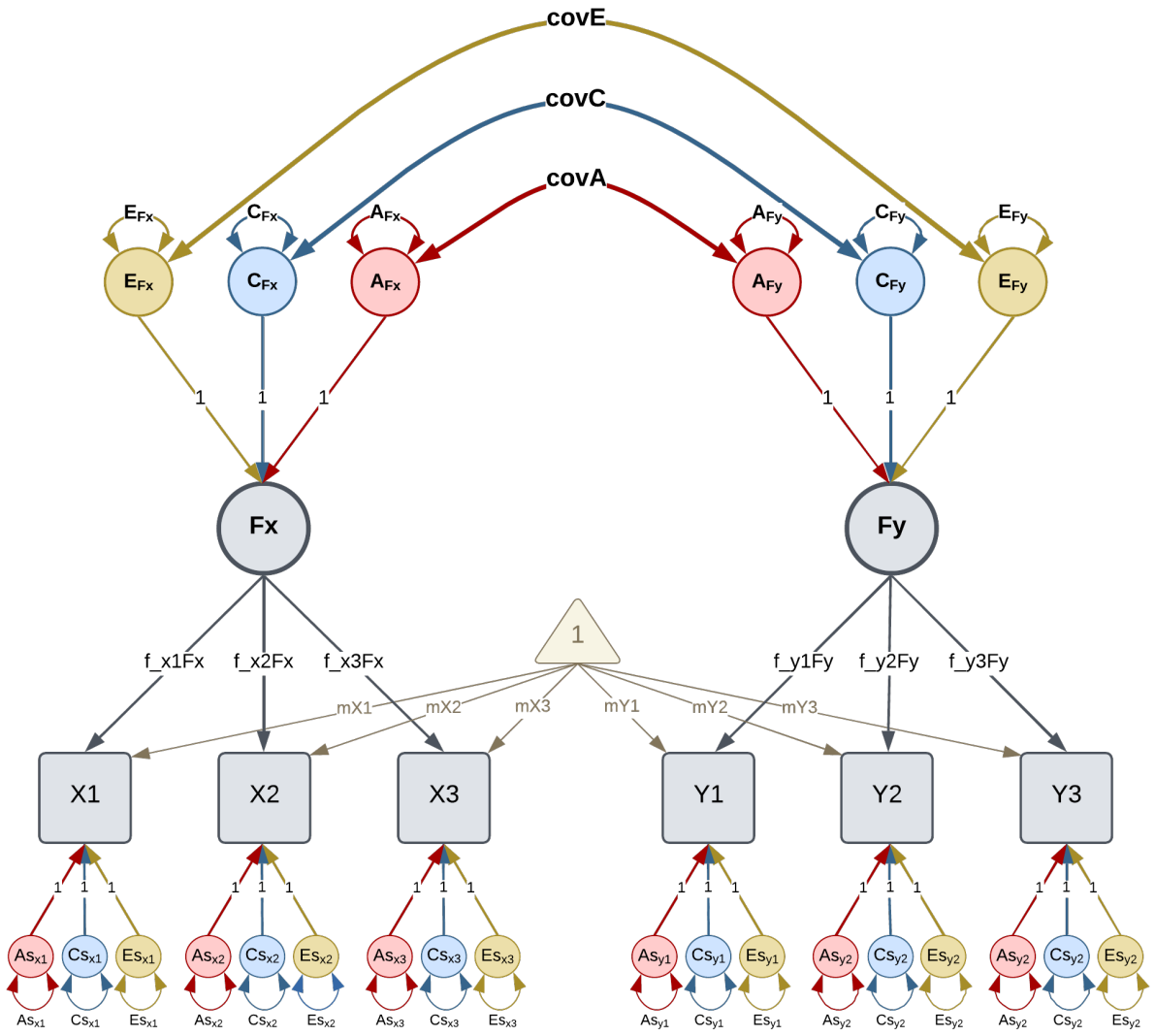
We may estimate any <u>three</u> sources of covariance between X and Y.

For example,

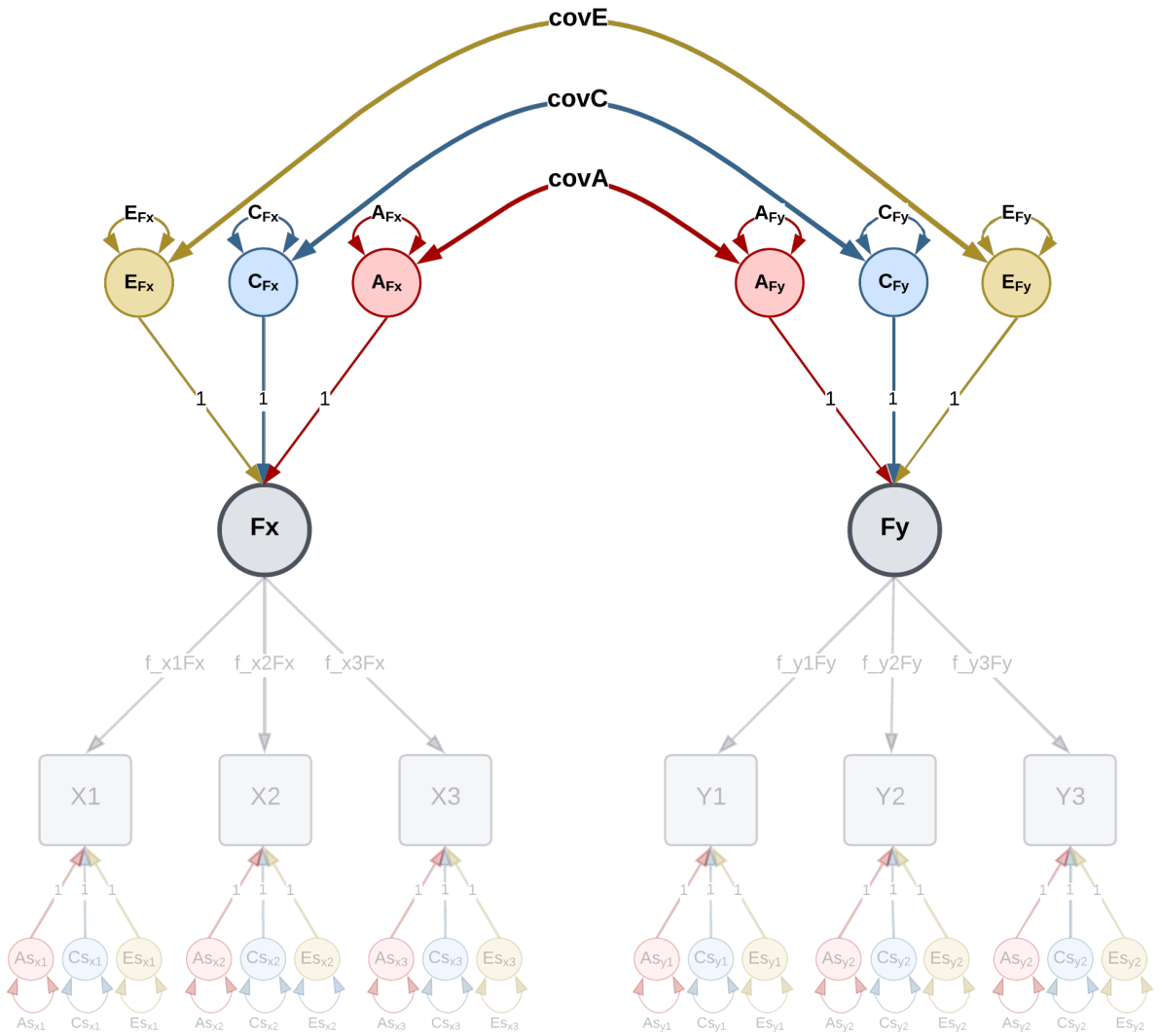
- *X causes Y*
- *Background confounding due to unmeasured additive genetic factors*
- *Background confounding due to unmeasured shared/familial environmental factors*

Script Walk-through

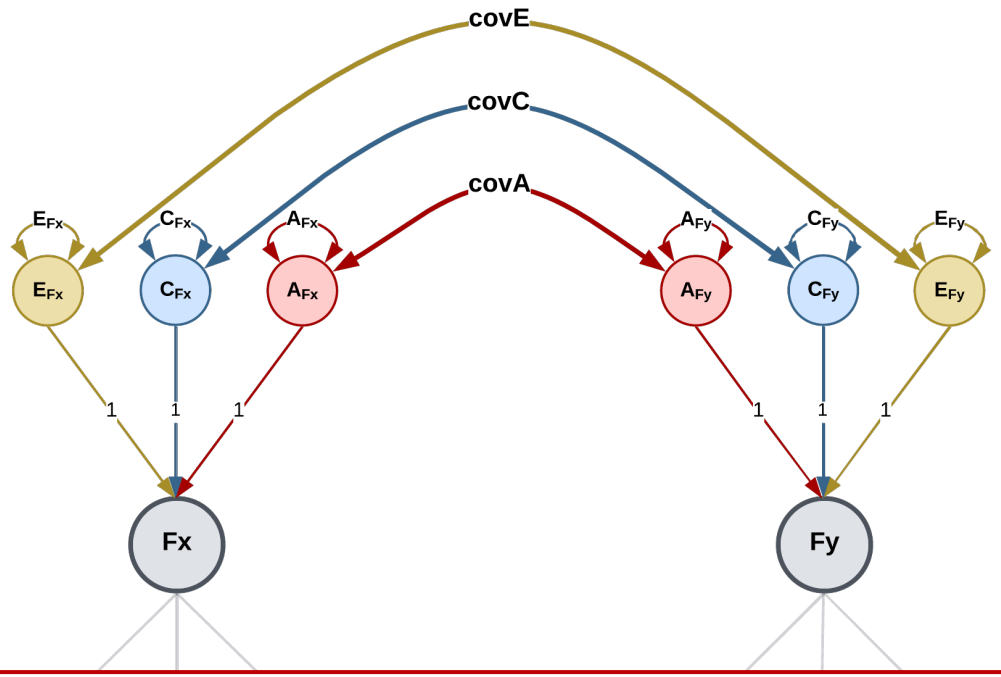
Start with the ACE [“confounding-only”] model



ACE variance-covariance components of the latent factors

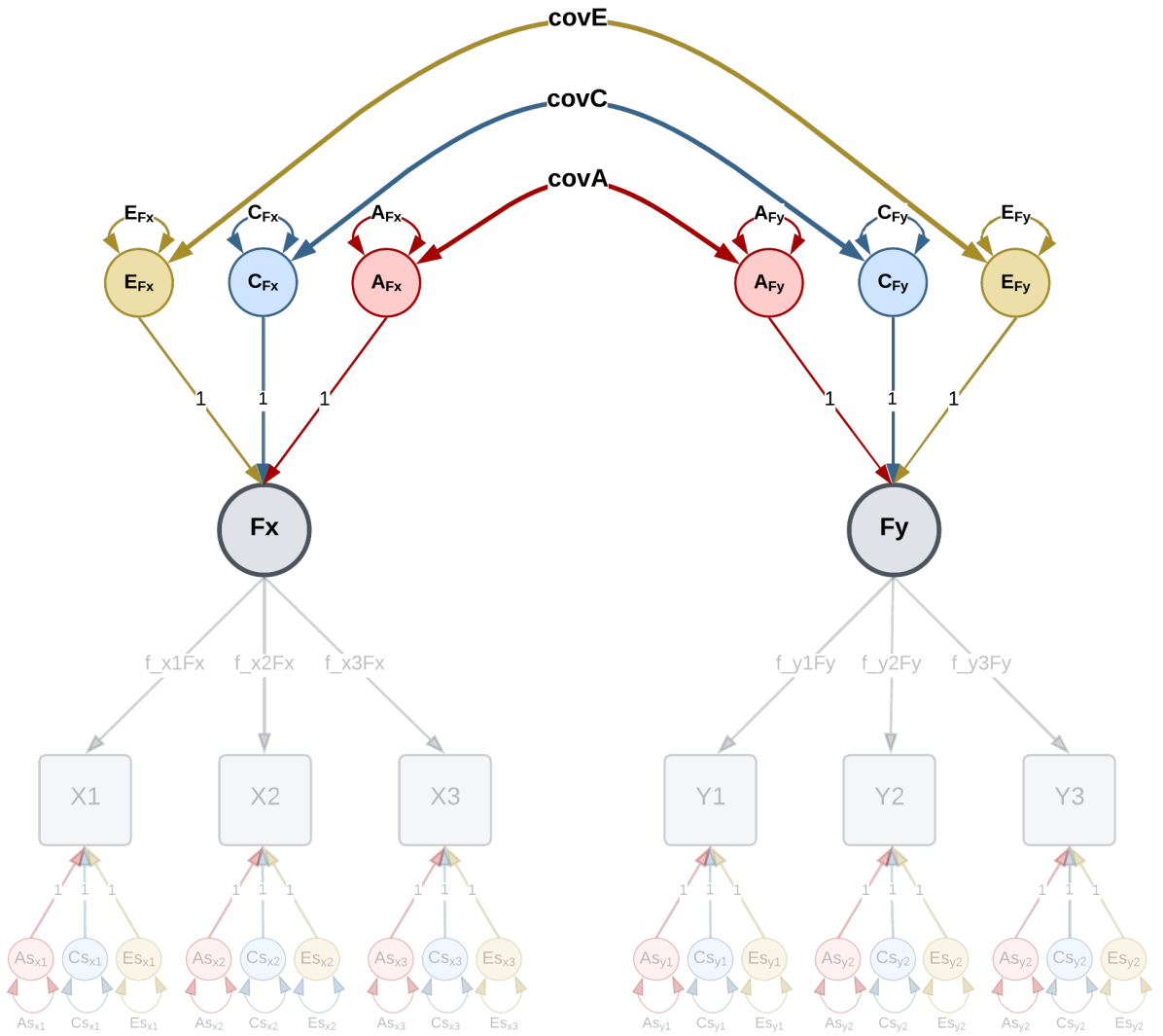


ACE variance-covariance components of the latent factors



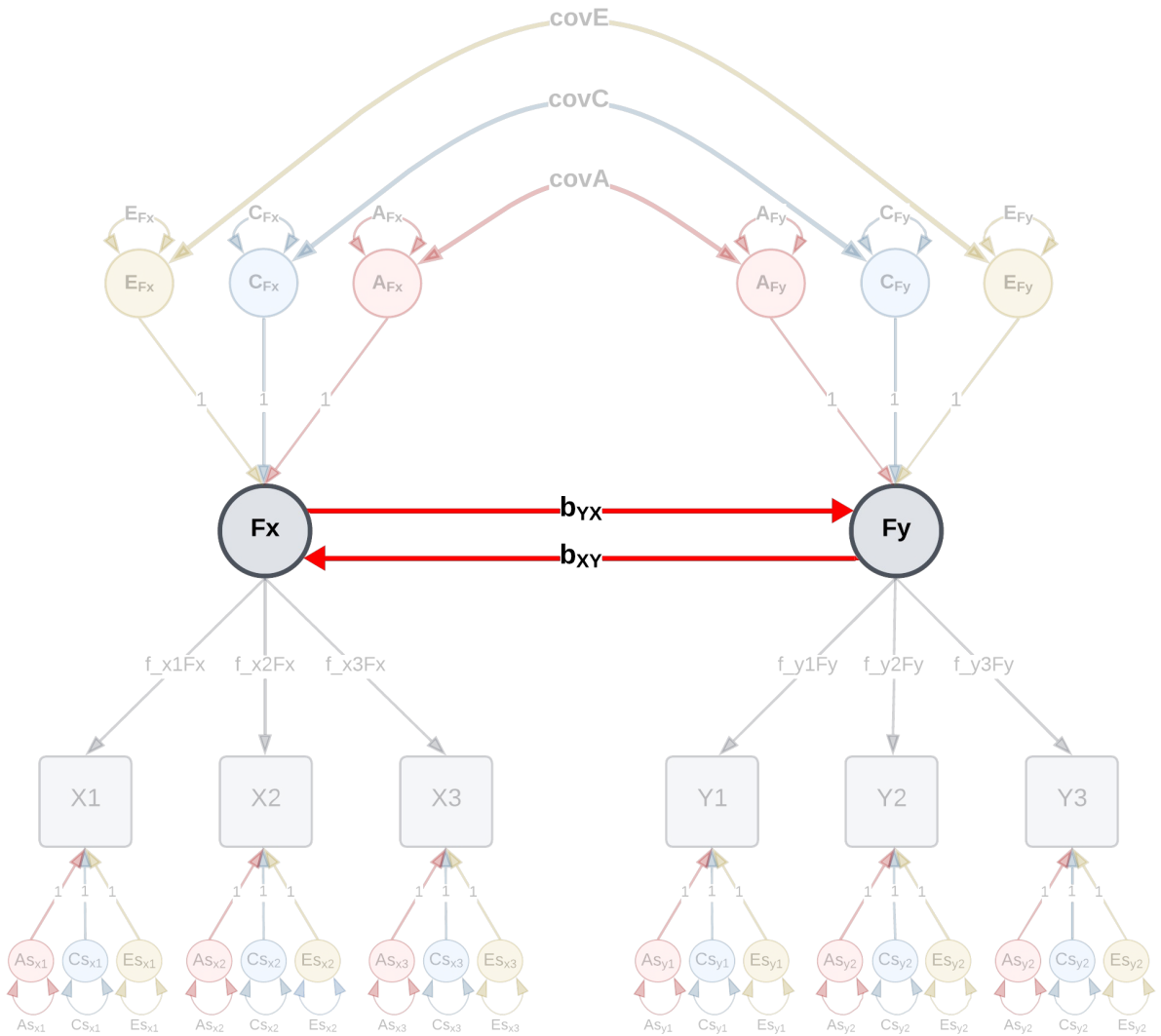

```
Af    <- mxMatrix(type = "Symm", nrow = nFac,ncol=nFac, labels=c("Af_x","covAf","covAf","Af_y"),
                  free=aFree, values=aVals, dimnames = list(latFacs,latFacs),
                  name="Af")
Cf    <- mxMatrix(type = "Symm", nrow = nFac,ncol=nFac, labels=c("Cf_x","covCf","covCf","Cf_y"),
                  free=cFree, values=cVals, dimnames = list(latFacs,latFacs),
                  name="Cf")
Ef    <- mxMatrix(type = "Symm", nrow = nFac,ncol=nFac, labels=c("Ef_x","covEf","covEf","E_Fy"),
                  free=eFree, values=eVals, dimnames = list(latFacs,latFacs),
                  name="Ef")
```

ACE variance-covariance components of the latent factors

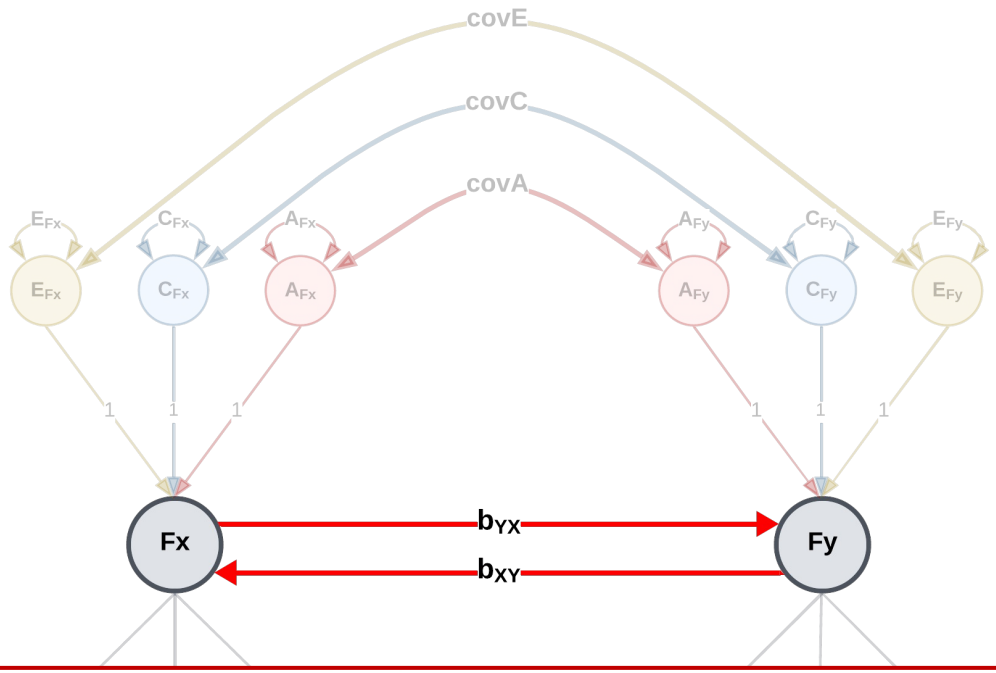



```
> Af
SymmMatrix 'Af'

$labels
   Fx       Fy
Fx "Af_x"   "covAf"
Fy "covAf"  "Af_y"

$values
    Fx  Fy
Fx 0.4 0.2
Fy 0.2 0.4

$free
     Fx   Fy
Fx TRUE TRUE
Fy TRUE TRUE

$lbound: No lower bounds assigned.

$ubound: No upper bounds assigned.
```

Causal Paths – Fixed at zero in the ACE confounding-only model

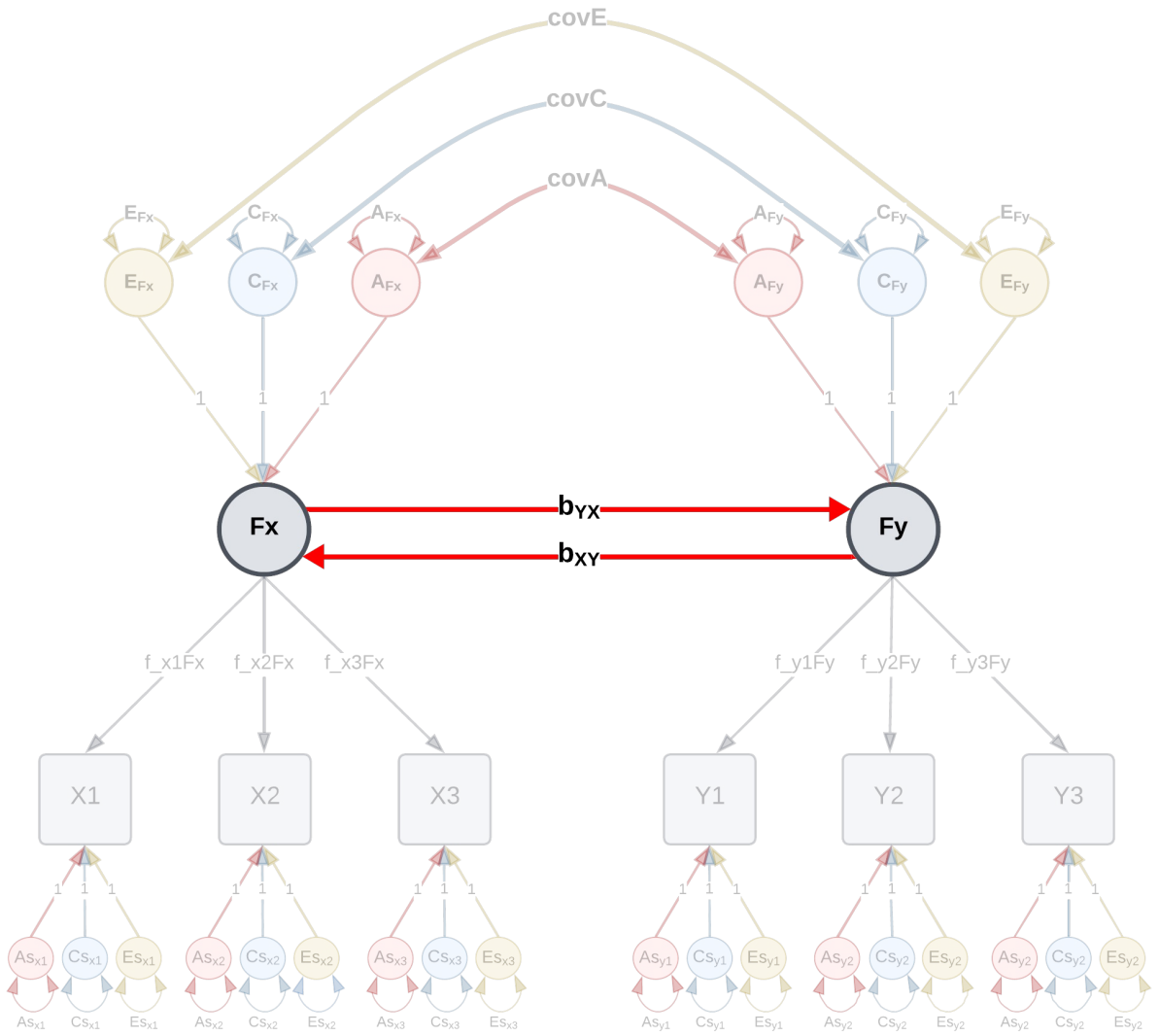


Causal Paths – Fixed at zero in the ACE confounding-only model



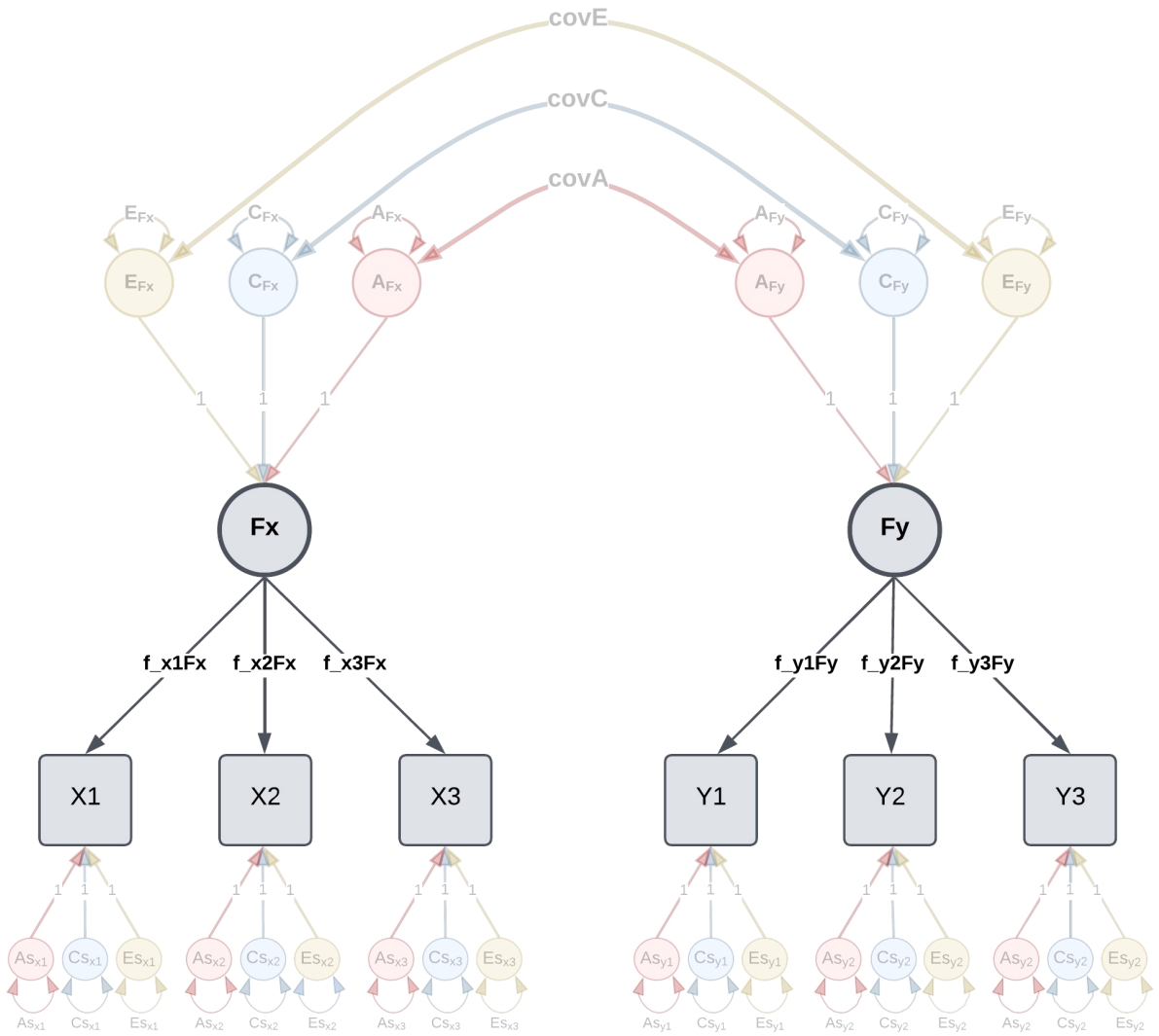

```
Blabs   <- matrix(c(NA,"bxy",
                    "byx",NA),
                  byrow = T,
                  nrow = nFac,
                  ncol = nFac,
                  dimnames = list(latFacs,latFacs))
B       <- mxMatrix(type="Full", nrow=nFac, ncol=nFac,
                    labels=Blabs, free=Bfree, values=Bvals,
                    dimnames = list(latFacs,latFacs), name = "B" )
```

Causal Paths – Fixed at zero in the ACE confounding-only model

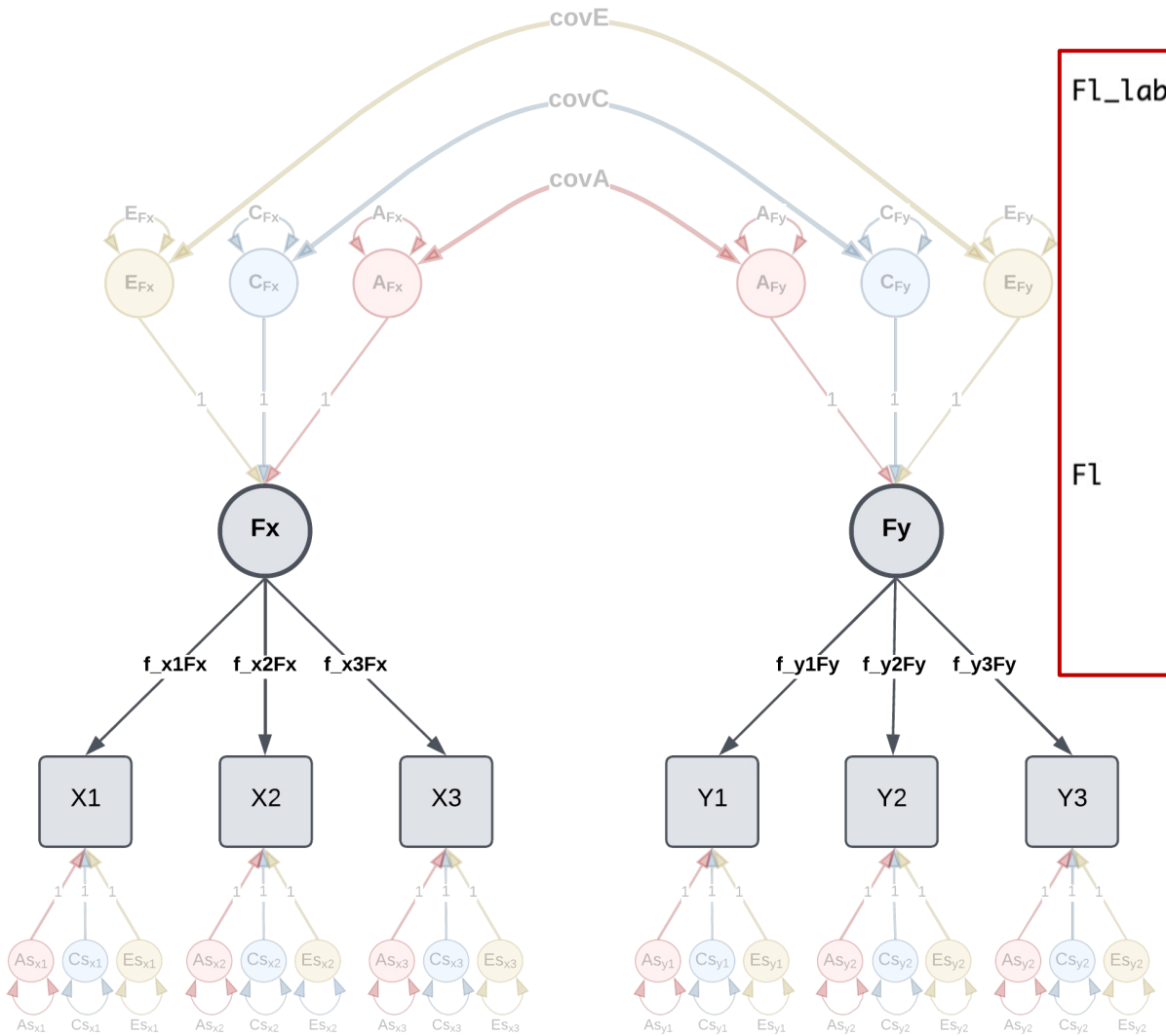



```
> B
FullMatrix 'B'

$labels
   Fx    Fy
Fx NA    "bxy"
Fy "byx" NA

$values
   Fx Fy
Fx  0  0
Fy  0  0

$free: No free parameters.

$lbound: No lower bounds assigned.

$ubound: No upper bounds assigned.
```

Factor loadings

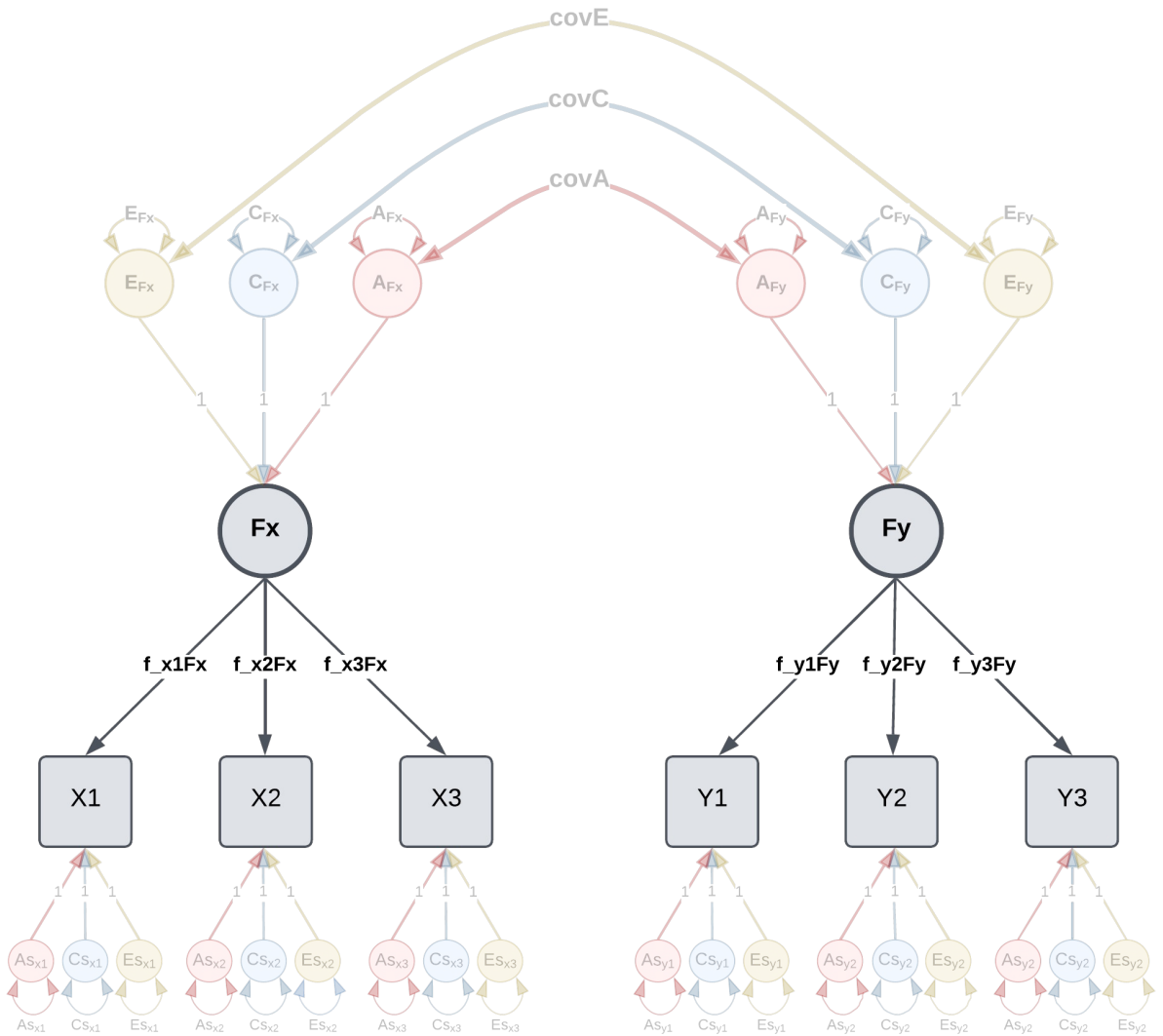


Factor loadings



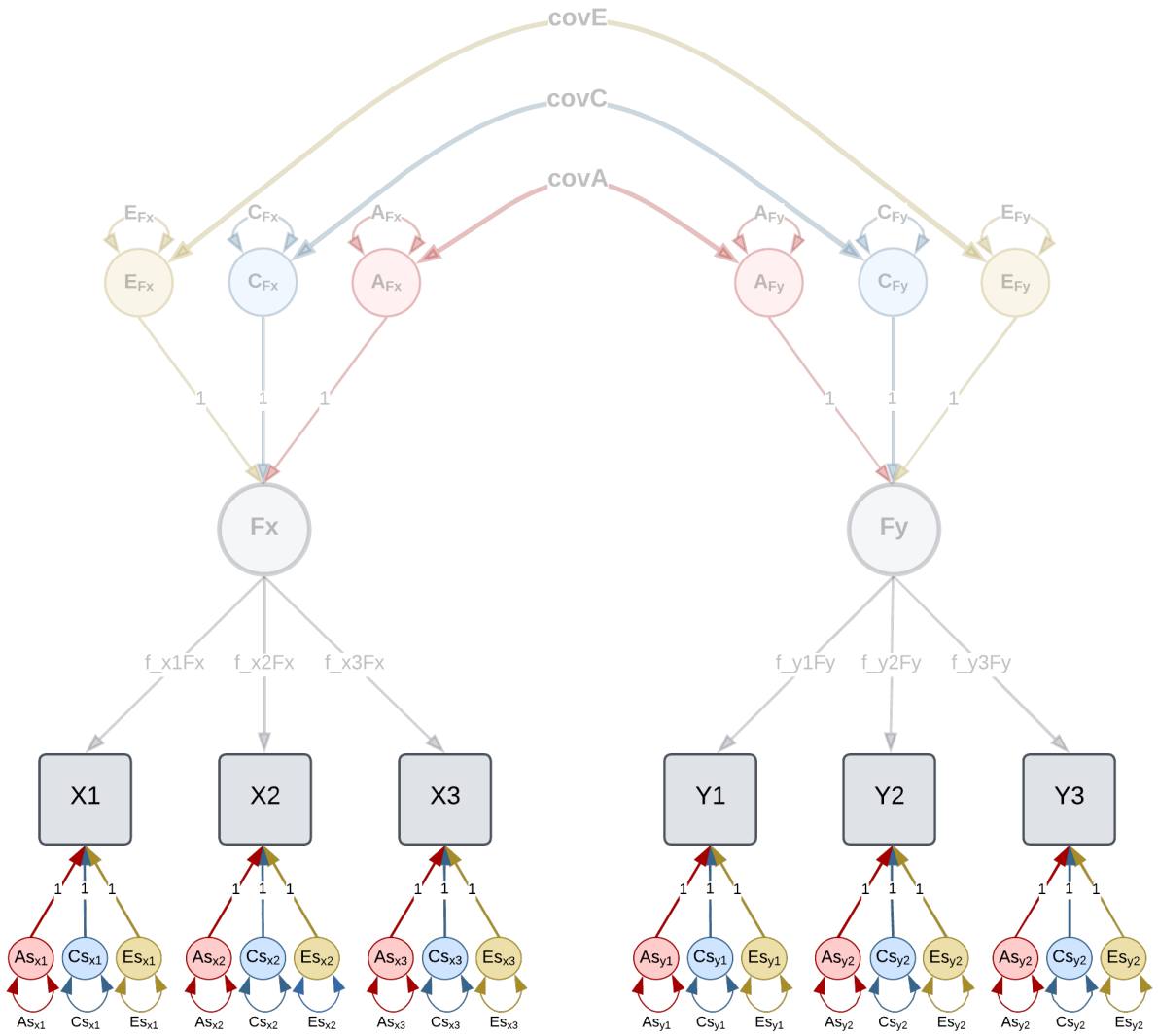

```
Fl_labs      <- matrix(c( "f_x1Fx", NA,
                          "f_x2Fx", NA,
                          "f_x3Fx", NA,
                          NA,       "f_y1Fy",
                          NA,       "f_y2Fy",
                          NA,       "f_y3Fy"),
                       byrow = T,
                       nrow = nv,
                       ncol = nFac,
                       dimnames = list(vars,latFacs))
Fl     <- mxMatrix(type="Full", nrow=nv, ncol=nFac,
                   free = Fl_free,
                   values = Fl_vals, labels=Fl_labs,
                   dimnames = list(vars,latFacs),
                   name = "Fl")
```

Factor loadings



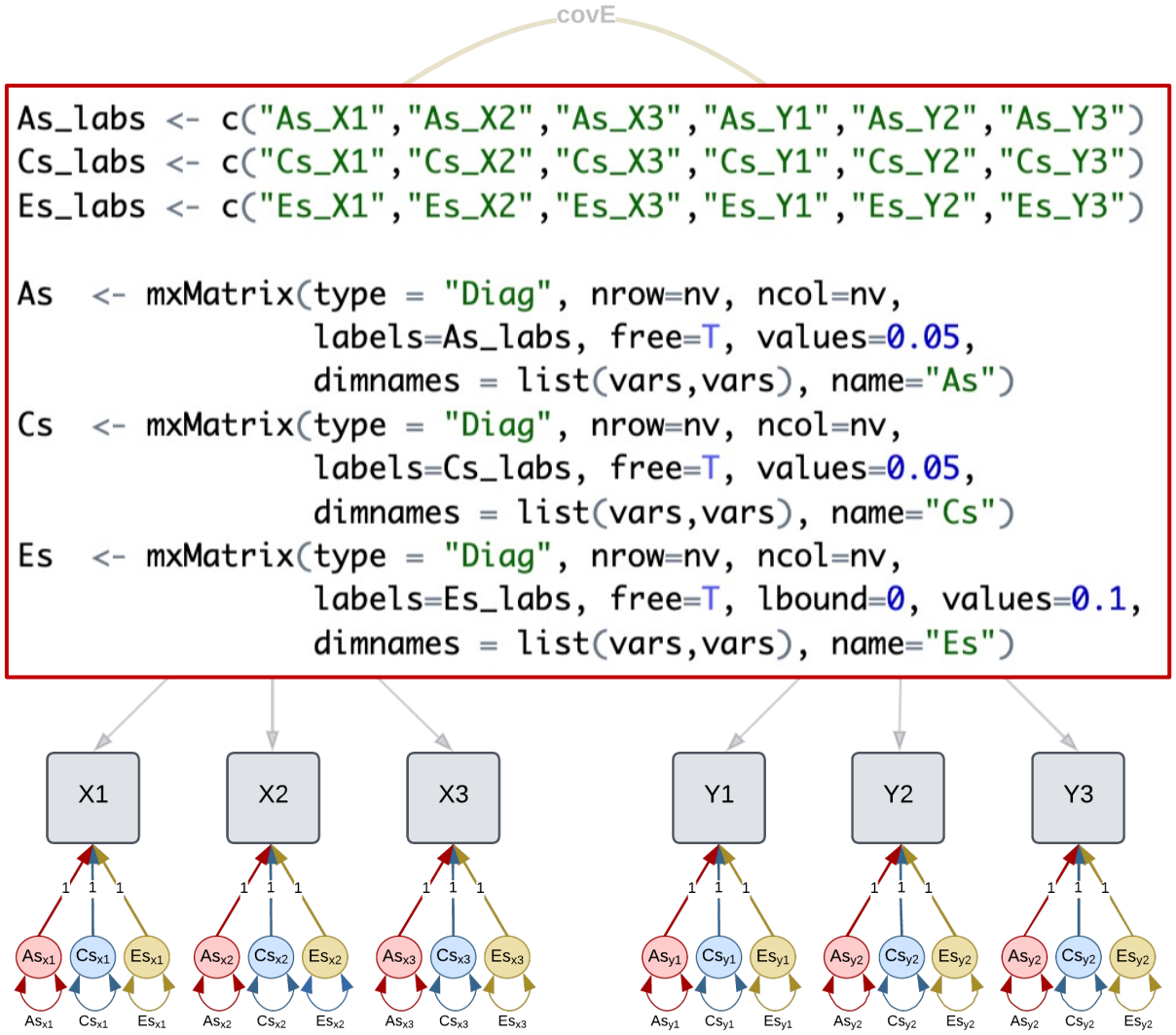

```
> Fl
FullMatrix 'Fl'

$labels
   Fx        Fy
X1 "f_x1Fx" NA
X2 "f_x2Fx" NA
X3 "f_x3Fx" NA
Y1 NA       "f_y1Fy"
Y2 NA       "f_y2Fy"
Y3 NA       "f_y3Fy"

$values
    Fx  Fy
X1 0.8 0.0
X2 0.8 0.0
X3 0.8 0.0
Y1 0.0 0.8
Y2 0.0 0.8
Y3 0.0 0.8

$free
      Fx    Fy
X1  TRUE FALSE
X2  TRUE FALSE
X3  TRUE FALSE
Y1 FALSE  TRUE
Y2 FALSE  TRUE
Y3 FALSE  TRUE

$lbound: No lower bounds assigned.

$ubound: No upper bounds assigned.
```

ACE components of the indicator-specific variances (= residuals)

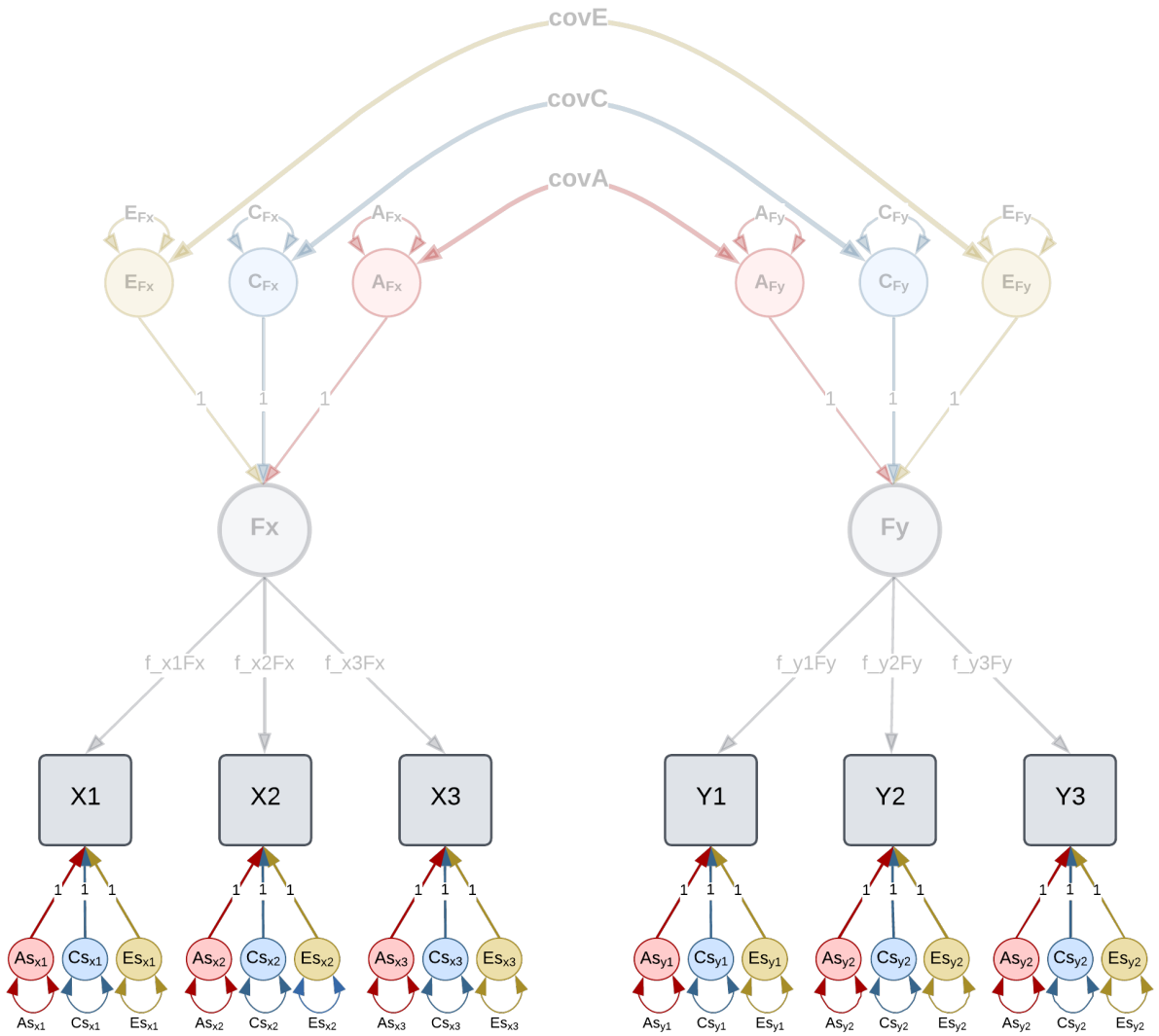


ACE components of the indicator-specific variances (= residuals)

```
As_labs <- c("As_X1","As_X2","As_X3","As_Y1","As_Y2","As_Y3")
Cs_labs <- c("Cs_X1","Cs_X2","Cs_X3","Cs_Y1","Cs_Y2","Cs_Y3")
Es_labs <- c("Es_X1","Es_X2","Es_X3","Es_Y1","Es_Y2","Es_Y3")

As  <- mxMatrix(type = "Diag", nrow=nv, ncol=nv,
                labels=As_labs, free=T, values=0.05,
                dimnames = list(vars,vars), name="As")
Cs  <- mxMatrix(type = "Diag", nrow=nv, ncol=nv,
                labels=Cs_labs, free=T, values=0.05,
                dimnames = list(vars,vars), name="Cs")
Es  <- mxMatrix(type = "Diag", nrow=nv, ncol=nv,
                labels=Es_labs, free=T, lbound=0, values=0.1,
                dimnames = list(vars,vars), name="Es")
```

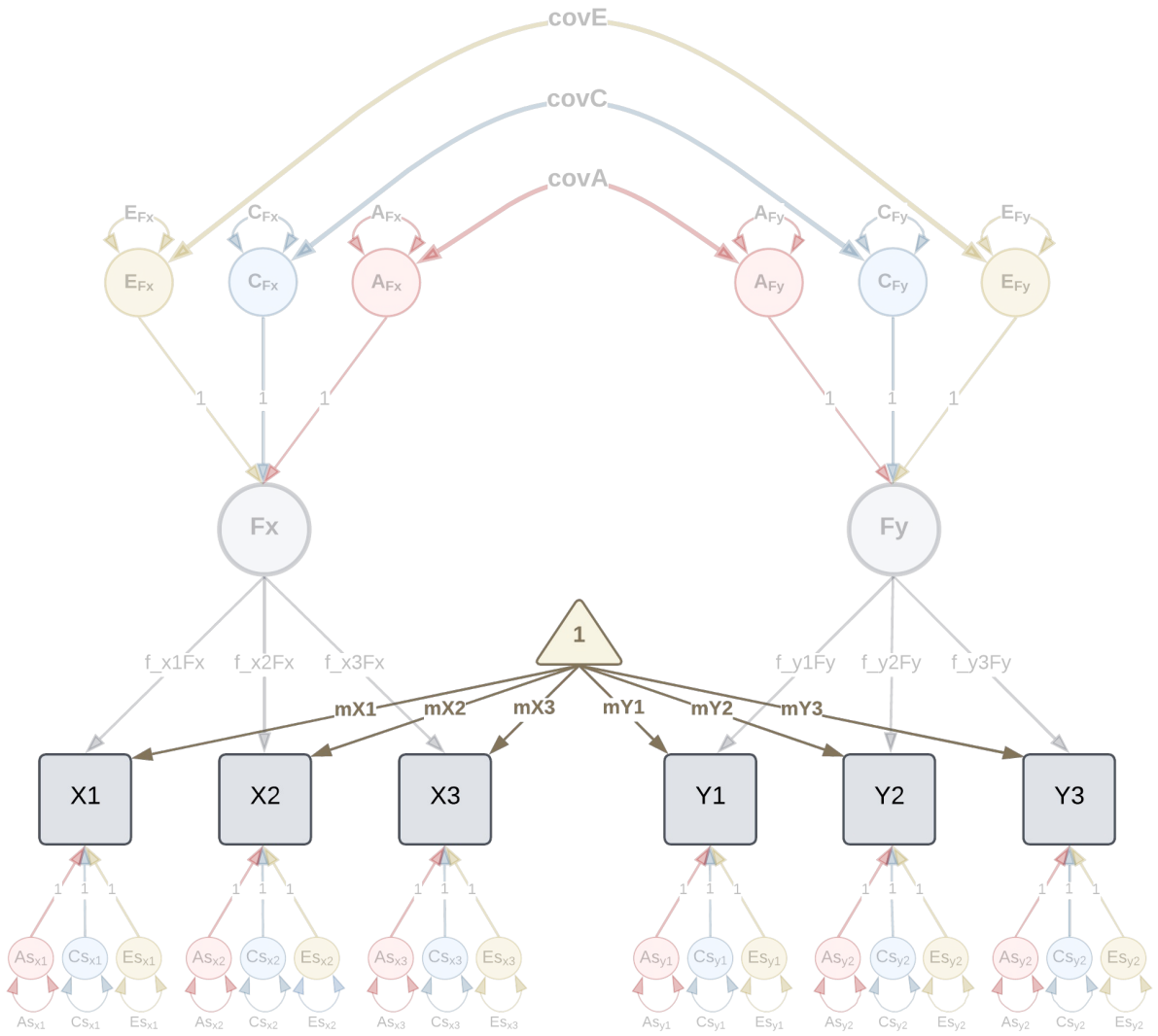


ACE components of the indicator-specific variances (= residuals)

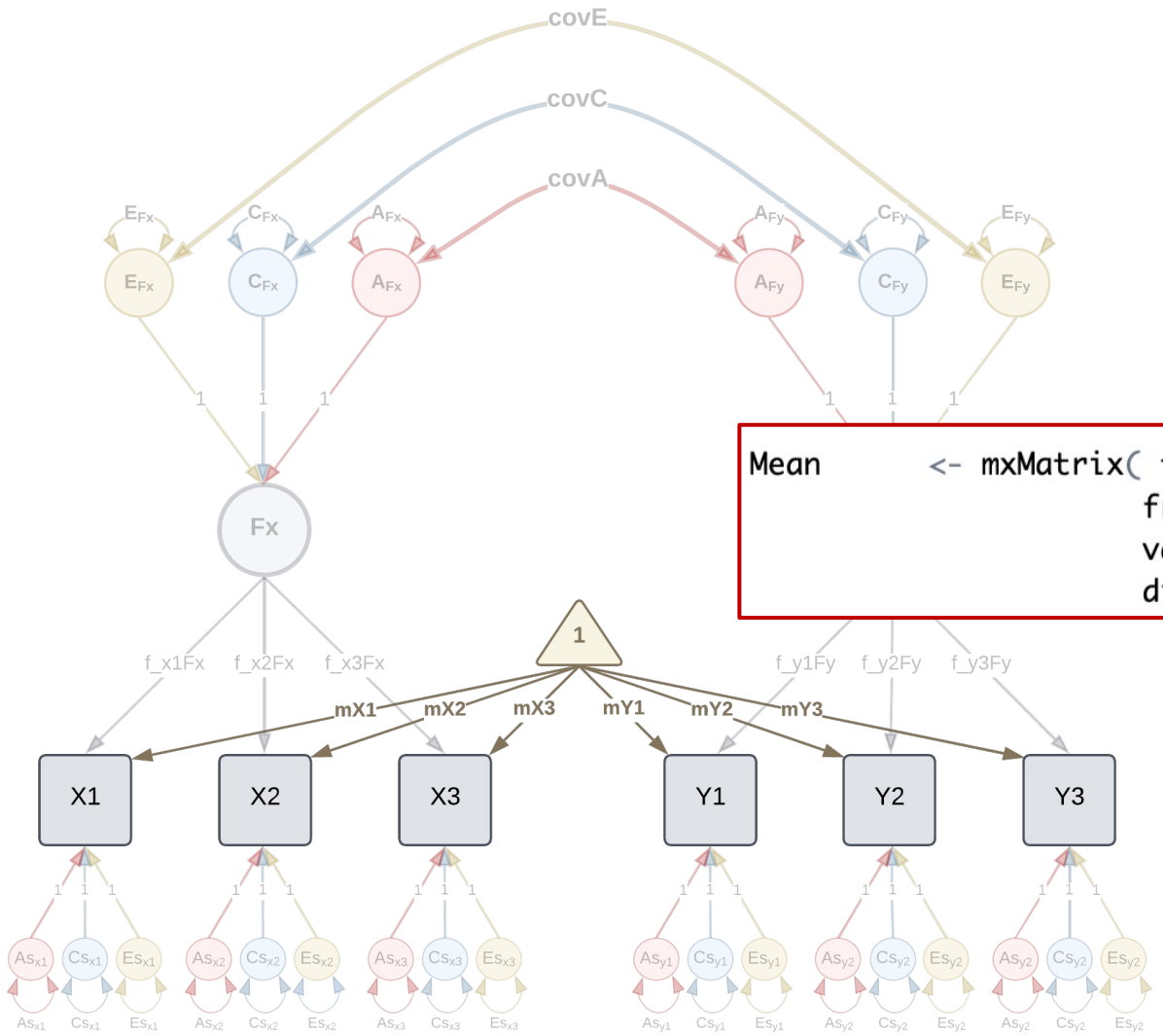



```
> As
DiagMatrix 'As'

$labels
   X1       X2       X3       Y1       Y2       Y3
X1 "As_X1"  NA       NA       NA       NA       NA
X2 NA       "As_X2"  NA       NA       NA       NA
X3 NA       NA       "As_X3"  NA       NA       NA
Y1 NA       NA       NA       "As_Y1"  NA       NA
Y2 NA       NA       NA       NA       "As_Y2"  NA
Y3 NA       NA       NA       NA       NA       "As_Y3"

$values
     X1   X2   X3   Y1   Y2   Y3
X1 0.05 0.00 0.00 0.00 0.00 0.00
X2 0.00 0.05 0.00 0.00 0.00 0.00
X3 0.00 0.00 0.05 0.00 0.00 0.00
Y1 0.00 0.00 0.00 0.05 0.00 0.00
Y2 0.00 0.00 0.00 0.00 0.05 0.00
Y3 0.00 0.00 0.00 0.00 0.00 0.05

$free
      X1    X2    X3    Y1    Y2    Y3
X1  TRUE FALSE FALSE FALSE FALSE FALSE
X2 FALSE  TRUE FALSE FALSE FALSE FALSE
X3 FALSE FALSE  TRUE FALSE FALSE FALSE
Y1 FALSE FALSE FALSE  TRUE FALSE FALSE
Y2 FALSE FALSE FALSE FALSE  TRUE FALSE
Y3 FALSE FALSE FALSE FALSE FALSE  TRUE
```

Specify the means

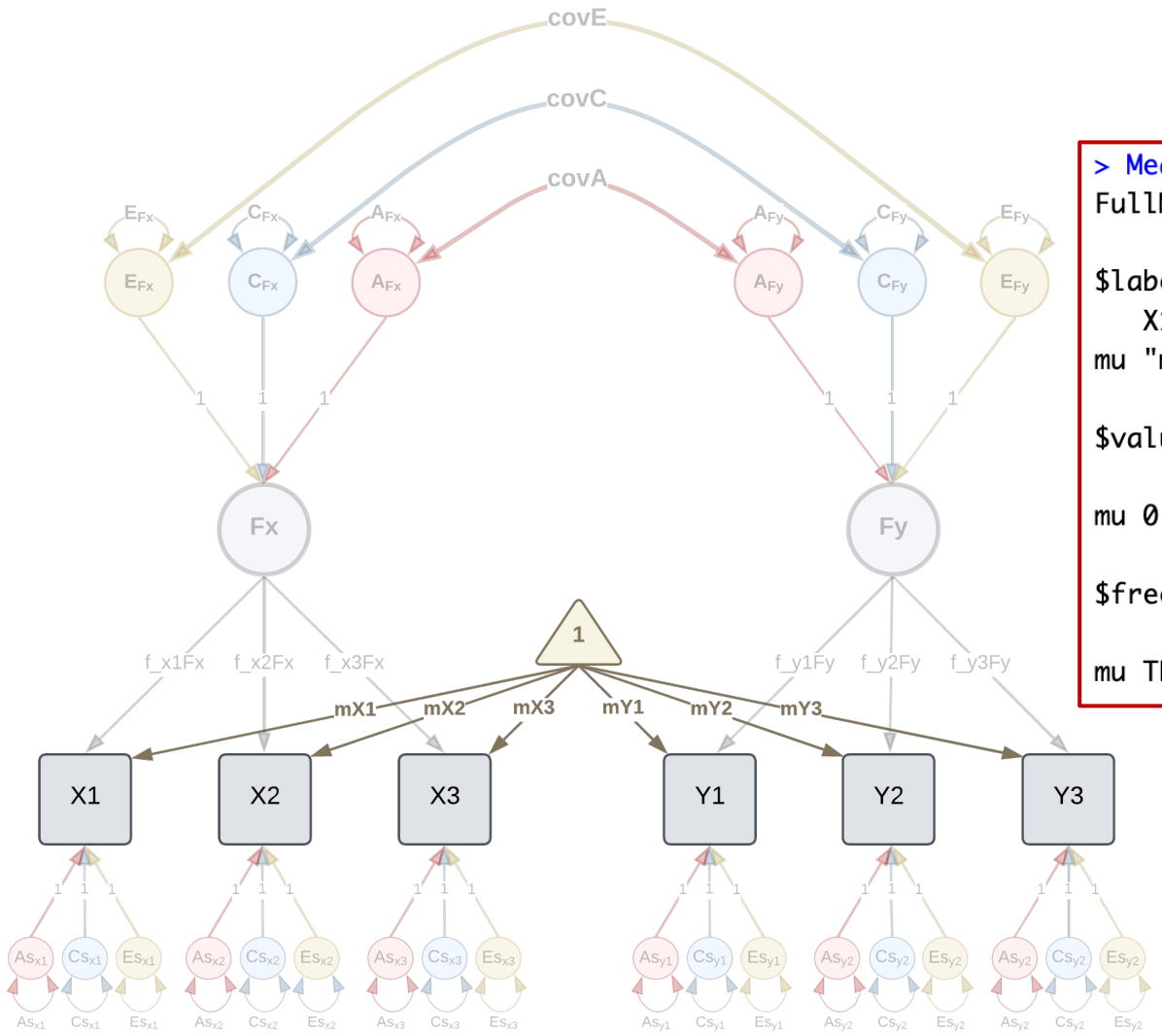


Specify the means




```
Mean      <- mxMatrix( type="Full", nrow=1, ncol=nv,
                       free=TRUE, labels=paste0("mu_",vars),
                       values = c(0.02,0.02,0.02,0.25,0.25,0.25),
                       dimnames = list("mu",vars), name = "Mean" )
```

Specify the means




```
> Mean
FullMatrix 'Mean'

$labels
   X1       X2       X3       Y1       Y2       Y3
mu "mu_X1" "mu_X2" "mu_X3" "mu_Y1" "mu_Y2" "mu_Y3"

$values
     X1   X2   X3   Y1   Y2   Y3
mu 0.02 0.02 0.02 0.25 0.25 0.25

$free
     X1   X2   X3   Y1   Y2   Y3
mu TRUE TRUE TRUE TRUE TRUE TRUE
```

Tutorial

Please copy the scripts for this session

```
mkdir ~/doc
cp  -r  /home/madhur/2024/Day-5/DOC/*  ~/doc/.
```

Qualtrics link

This link is also on top of the R script: ***DOC_multiple_indicator.R***

https://qimr.az1.qualtrics.com/jfe/form/SV_eW1CeUT3ZibFMzQ

Tutorial: DOC modeling with multiple indicators

1. Fit the **ACE "confounding-only" model** = *Null hypothesis for DOC*.

2. Run the **Direction of Causation models** (alternative causal hypotheses):

   a) *X* causes *Y* (no background ACE covariance)

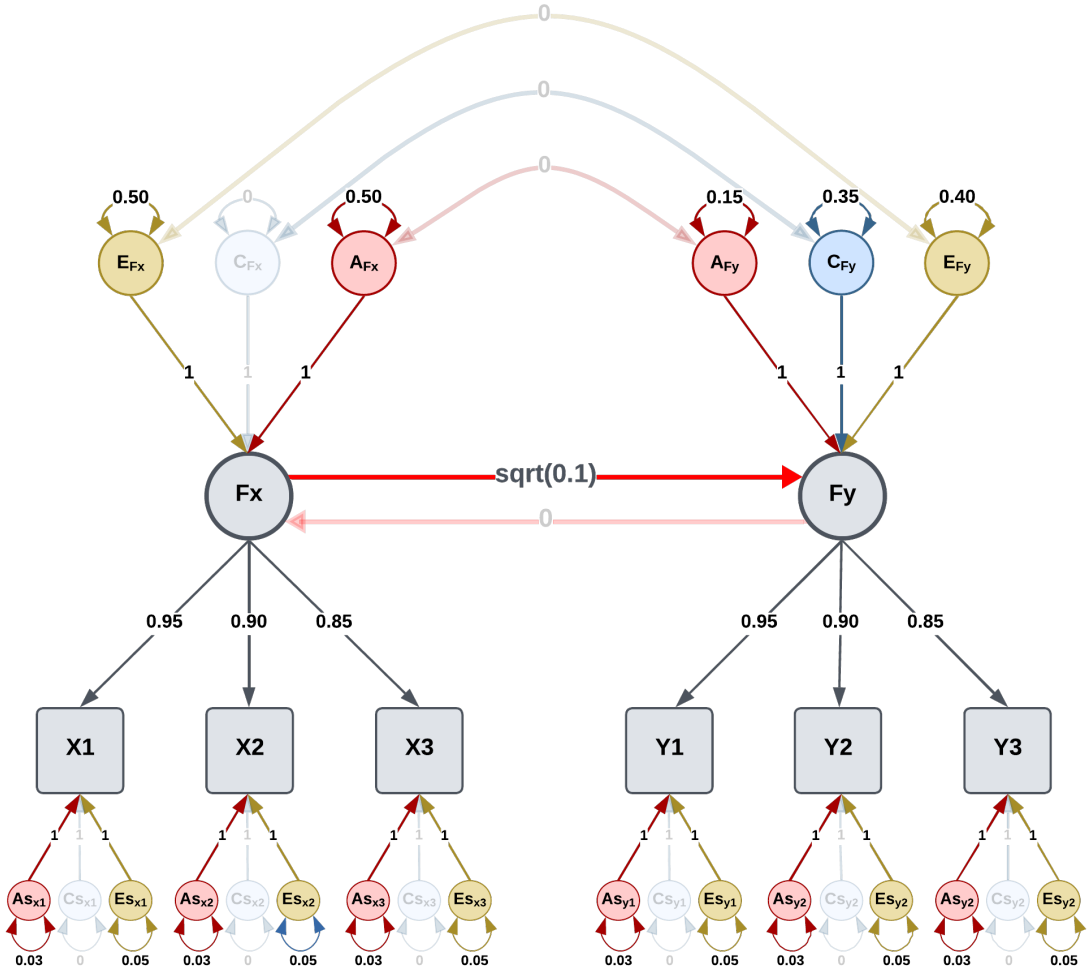
   b) *Y* causes *X* (no background ACE covariance)

   c) Reciprocal causation between *X* and *Y* (no background ACE covariance)

   Infer the mechanism of covariance between *X* and *Y* based on the best-fitting model.

3. Bonus: Fit a **hybrid model with three sources of covariance** (causation + confounding).

The “true” model – as simulated



Data simulated using the script *xtra_Data_Simulation_DoC_Multiple_Indicator.R*

Further Readings

- Maes, H. H. M., Neale, M. C., Kirkpatrick, R. M., & Kendler, K. S. (2021). Using **Multimodel Inference/Model Averaging to Model Causes of Covariation** Between Variables in Twins. *Behavior Genetics*, *51*(1), 82-96. https://doi.org/10.1007/s10519-020-10026-8

  Fit all possible bivariate twin models

  Obtain a weighted average of the parameter estimates across models

  Models are weighted by their goodness of fit

- Neale, M. C., & Kendler, K. S. (1995). **Models of comorbidity for multifactorial disorders**. *American Journal of Human Genetics*, *57*(4), 935–953. https://pubmed.ncbi.nlm.nih.gov/7573055/

  Theoretical and methodological exposition of different causes of covariation between traits

References

- Neale, M. C., & Cardon, L. R. (1992). Direction of Causation. In M. C. Neale & L. R. Cardon (Eds.), *Methodology for Genetic Studies of Twins and Families* (pp. 261-287). Springer Netherlands. https://doi.org/10.1007/978-94-015-8018-2_13
- Heath, A. C., Kessler, R. C., Neale, M. C., Hewitt, J. K., Eaves, L. J., & Kendler, K. S. (1993). Testing hypotheses about direction of causation using cross-sectional family data. *Behavior Genetics*, *23*(1), 29-50. https://doi.org/10.1007/BF01067552
- Duffy, D. L., & Martin, N. G. (1994). Inferring the direction of causation in cross-sectional twin data: Theoretical and empirical considerations. *Genetic Epidemiology*, *11*(6), 483-502. https://doi.org/10.1002/gepi.1370110606