Twin Strutural Equation Models in R & Lavaan

Michel G Nivard

6/21/2021

Contents

1 Introduction

1.1 Resemblance between relatives

There is a long history of questioning the nature of the resemblance between family members, specifically sibling and even more specifically twins. Reasons that are commonly put forward for these similarities are the obvious genetic similarity, similarity in socio-economic position and similarity in upbringing shared between siblings. Twin and family models leverage variation in genetic and environmental relatedness between family members to estimate the relative contributions of genetics, the environment, their interacting, their correlation and other process to the similarities between relatives.

Twins offer an excellent “natural experiment” where some twins are genetically identical (identical, or monozygotic twins or MZ twins or MZs) and some twins share halve their segregating DNA (fraternal, or dizygotic twins, DZ twins or DZs). Like any natural experiment, twin studies aren’t true experiments, and so the estimated quantities come with various assumptions, some of which are specific to twin models, and I’ll make sure to catalog the assumptions we are making along the way.

If certain assumptions are satisfied the causes of resemblance between twins can be generalized to be causes on individual differences in the general population, provided the twins we sample are representative of the populations, and the population in question is homogeneous, as the relative contribution of genes, the environment their correlation and interaction arent fixed quantities or inescapable facts of life, they are but temporal phenomena that reflect the economic/political and societal status quo in the sample of population one studies.

1.2 SEM Models for twin data

This document describes various twin structural equation models (SEM) to estimate the nature of the relationship between family members in order to learn about the contribution of genes and the social or rearing environment to complex (behavioral) outcomes. The goal is to provide a basic understands of these models with the means to fit the models in lavaan (Rosseel 2012), lavaan is an R package that allow the use to define a structural equation model in terms of regression, variances and covariances and will be familiar to users of M-Plus. The package is (IMO) more accessible to beginners then another excellent SEM R package: OpenMx (Neale et al. 2015), but the accessibility comes at the cost of less flexibility, In terms of flexibility OpenMx is truly unrivaled. Its worth pointing out that the developers behind OpenMx have an academic interest in twin models, which means scripts and support for users in those models is often especially excellent. The best (to my knowledge) repository of twin models in OpenMx is maintained by Hermine Maes and there is an R package that wraps OpenMx twin models in simpler R commends called umx (Bates, Maes, and Neale 2019) which is build and maintained by Tim Bates. Some models considered fundamental to our understanding of gene-environment interaction cannot (to my knowledge) be fitted in lavaan, despite these limitation there are many advantages to learning lavaan. The syntax used to specify lavaan models also features in blavaan (Merkle and Rosseel 2018), its Bayesian cousin providing an easy entry into Bayeslian twin modeling. Similarly the package RegSEM (Jacobucci, Grimm, and McArdle 2016) and lslx (Huang, Chen, and Weng 2017) allows for *regularized* structural equation modeling and use lavaan, or very lavaan like, syntax and could be a point of departure for various novel and innovative twin models.

All things considered it makes sense to collect and document lavaan syntax and examples for various twin models.

2 Quickly set up my R environment

We will need the following packages, some may be optional some will be required for any lavaan analysis (like lavaan).

```
library(lavaan)
library(MASS)
library(tidySEM)
library(ggplot2)
library(knitr)
```

3 Lavaan

3.1 Syntax

Lavaan requires the user to specify a SEM model in terms of regression, which are directional relations between observed variables, factor loadings which are directional relations between a latent factor and an observed variable, variances and variances which are undirected relationships between variables. The model is defined in a text string within R.

Table 1: Table 1: Lavaan Syntax Elements and their meaning

Syntax	Description	Meaning
`f =~ In1 + In2 + In3`	Factor loading	The factor “f” is measure by 3 indicators “In1,” “In2” and “In3,”
`y ~ x`	Regression	The variable “y” is regressed on the variable “x”
`x1 ~~ x2`	Covariance	the variable “x1” and “x2” covary
`x1 ~~ x1`	Variance	the variable x1 has a freely estimated variance

You can use addition to add multiple elements to a regression in lavaan: `y ~ x1 + x2` and to covariance: `x1 ~~ x1 + x2 + x3`. Lavaan will map the variable names in the syntax to variable names in your dataset. Lavaan expects all the relationships in a model to be contained in a single string variable in R let me show you an example model fit to simulated data:

```
# make the data
f <- rnorm(100)
i1 <- 1*f + rnorm(100)
i2 <- 0.6*f + rnorm(100)
i3 <- 1.4*f + rnorm(100)

x <- rnorm(100)
y <- f + x

dataset <- cbind.data.frame(i1,i2,i3,x,y)

# write the lavaan model:
example.model <- "
f1 =~ i1 + i2 + i3
y ~ x + f1
x ~~ f1
```

```
"
```

Okay, now notice how `f` does not exist in the dataframe: `dataset` (its going to be latent), lets fit the model to the data and have a look at the results.

```
model.fit <- sem(model = example.model,data=dataset)
```

```
## Warning in lav_object_post_check(object): lavaan WARNING: some estimated ov
## variances are negative
```

```
model.fit
```

```
## lavaan 0.6-7 ended normally after 27 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                         11
##
##   Number of observations                           100
##
## Model Test User Model:
##
##   Test statistic                                 4.271
##   Degrees of freedom                                 4
##   P-value (Chi-square)                           0.371
```

So the model ran and converged, the fit (according to the chi2) is adequate. Thats not entirely unexpected as we are fitting the true model to the data. Lets look at the estimates:

```
parameterEstimates(model.fit)
```

```
##    lhs op rhs    est    se      z pvalue ci.lower ci.upper
## 1   f1 =~  i1  1.000 0.000     NA     NA    1.000    1.000
## 2   f1 =~  i2  0.813 0.142  5.725  0.000    0.535    1.091
## 3   f1 =~  i3  1.808 0.213  8.486  0.000    1.391    2.226
## 4    y  ~   x  1.020 0.061 16.837  0.000    0.901    1.139
## 5    y  ~  f1  1.177 0.133  8.870  0.000    0.917    1.437
## 6   f1 ~~   x -0.150 0.089 -1.690  0.091   -0.323    0.024
## 7   i1 ~~  i1  0.819 0.121  6.742  0.000    0.581    1.057
## 8   i2 ~~  i2  1.072 0.153  6.990  0.000    0.771    1.372
## 9   i3 ~~  i3  0.954 0.181  5.258  0.000    0.598    1.309
## 10   y ~~   y -0.006 0.051 -0.111  0.912   -0.107    0.095
## 11   x ~~   x  0.754 0.107  7.071  0.000    0.545    0.964
## 12  f1 ~~  f1  0.796 0.200  3.975  0.000    0.403    1.188
```

As you can see the factor loading are close to their true simulated values of 1, 0.6 and 1.4 as are all of the other parameters.

3.2 Estimators

Int he examples below is simulate data and have no missingness or non-normal data to deal with, in reality you will so it makes sense to discuss the various estimators you could use in lavaan.

4 Single trait twin model in lavaan

4.1 ACE model

The "ACE" model operationalized a phenotype as a function of additive genetic variance (A), common environmental influences(C) (can be rearing, can be societal influences can be governmental policies), and environment unique to each of the individual twins (E) (can be private friends, being in separate classrooms, but also measurement error). The correlation between the genetic influences and common environmental is 1 for MZ twins while the correlation between the genetic influences is 0.5 for DZ twins.

the ACE model assumes among other things:

1. The absence of non-additive genetic effects (gene gene interaction and dominance)
2. The absence of sibling influences or rater contrast effect
3. The absence of gene by environment correlation
4. The absence of gene by environment interaction
5. The environments MZ twin grow up are as similaras the environments DZ twins grow up in.

Lets simulate same data where the resemblance between twins is a function of equal parts (33.3%/33.3%/33.4%)

```
A <- matrix(1,2,2) # genetic correlation for MZ's = 1
C <- matrix(1,2,2)
E <- diag(2)
Adz <- matrix(c(1,.5,.5,1),2,2) # genetic correlation for DZ's = 0.5

# make 1000 pairs of MZ twins
MZ <- mvrnorm(1000,mu=c(0,0),Sigma = A+C+E)

# Add a column to label as MZ:
MZ<- cbind.data.frame("MZ",MZ)
colnames(MZ) <- c("zyg","P1", "P2")


# make 1500 DZ twin pairs
DZ <- mvrnorm(1500,mu=c(0,0),Sigma = Adz+C+E)

# add variable too label as DZ:
DZ <- cbind.data.frame("DZ",DZ)
colnames(DZ) <- c("zyg","P1", "P2")

# Combine MZ and DZ twins
dataset <- rbind(MZ,DZ)
```

We then define the lavaan model that can express the variance in the trait P explained by latent variables A, C and E:

```
ace.model<-"
A1=~ NA*P1 + c(a,a)*P1
A2=~ NA*P2 + c(a,a)*P2
C1 =~ NA*P1 + c(c,c)*P1
C2 =~ NA*P2 + c(c,c)*P2
# variances
A1 ~~ 1*A1
```

```
A2 ~~ 1*A2
C1 ~~ 1*C1
C2 ~~ 1*C2
P1~~c(e2,e2)*P1
P2~~c(e2,e2)*P2
# covariances
A1 ~~ c(1,.5)*A2
A1 ~~ 0*C1 + 0*C2
A2 ~~ 0*C1 + 0*C2
C1 ~~ c(1,1)*C2"
```

Lets look at some of the critical lines of code in the model:

`A1=~ NA*P1 + c(a,a)*P1` Here we create the latent variable A1, the phenotype P for twin 1 (P1) loads ion this variable, and in both groups (groups being MZ and DZ twins) the influence of this latent variable on the outcome is the same (contained using `c(a,a)`). Similar code is used to defien the latent variables C1 and C2. Now the effect of genes on an outcome is assumed the same for everyone regardless of whether they are twins, or not, the resemblance between twin 1 and twin 2 difference for MZ and DZ twins. We define/fix the resemblance later in the model here: `A1 ~~ c(1,.5)*A2`, because A1 and A2 are variance 1: `A1 ~~ 1*A1` and `A2 ~~ 1*A2` the constrained implies a correlation of 1 for the MZ twins and a correlation of 0.5 for the DZ twins. The common environment is correlated 1 regardless of twin status: `C1 ~~ c(1,1)*C2`, while the unshared environment E is conceptualized as a residual variance of the trait P (P1 or P2 respectively): `P1~~c(e2,e2)*P1`

We assume the latent variables A(1/2) and C(1/2) are uncorrelated, and fix their covariance to 0:

```
A1 ~~ 0*C1 + 0*C2
A2 ~~ 0*C1 + 0*C2
```

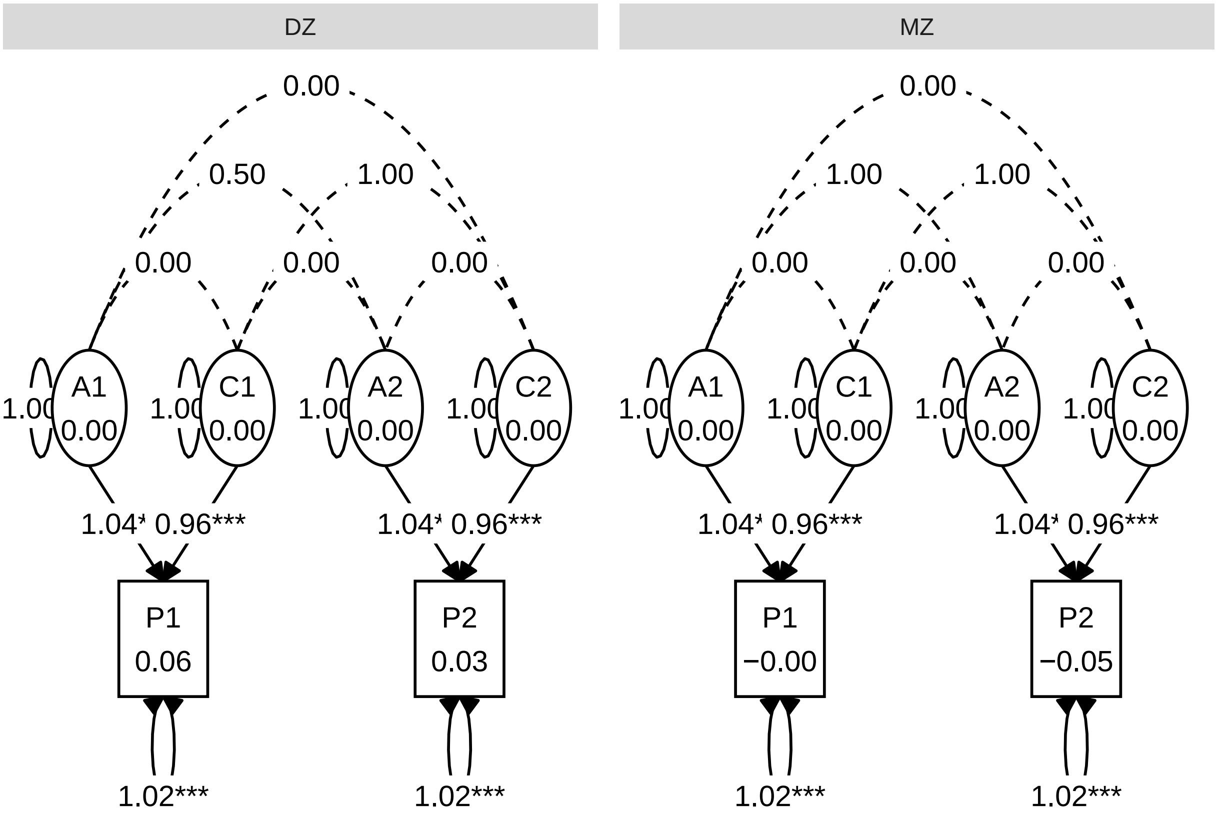
We also assume the residual variance (E) is uncorrelated to A and C, but fortunately for us this is a lavaan default. We proceed to fit the model to the simulated data:

```
# Standard ace model:
ace.fit<-cfa(ace.model, data = dataset,group = "zyg")
parameterEstimates(ace.fit)
```

```
##    lhs op rhs block group label    est    se      z pvalue ci.lower ci.upper
## 1   A1 =~  P1     1     1     a  1.037 0.068 15.223  0.000    0.903    1.170
## 2   A2 =~  P2     1     1     a  1.037 0.068 15.223  0.000    0.903    1.170
## 3   C1 =~  P1     1     1     c  0.959 0.065 14.747  0.000    0.831    1.086
## 4   C2 =~  P2     1     1     c  0.959 0.065 14.747  0.000    0.831    1.086
## 5   A1 ~~  A1     1     1        1.000 0.000     NA     NA    1.000    1.000
## 6   A2 ~~  A2     1     1        1.000 0.000     NA     NA    1.000    1.000
## 7   C1 ~~  C1     1     1        1.000 0.000     NA     NA    1.000    1.000
## 8   C2 ~~  C2     1     1        1.000 0.000     NA     NA    1.000    1.000
## 9   P1 ~~  P1     1     1    e2  1.022 0.045 22.650  0.000    0.933    1.110
## 10  P2 ~~  P2     1     1    e2  1.022 0.045 22.650  0.000    0.933    1.110
## 11  A1 ~~  A2     1     1        1.000 0.000     NA     NA    1.000    1.000
## 12  A1 ~~  C1     1     1        0.000 0.000     NA     NA    0.000    0.000
## 13  A1 ~~  C2     1     1        0.000 0.000     NA     NA    0.000    0.000
## 14  A2 ~~  C1     1     1        0.000 0.000     NA     NA    0.000    0.000
## 15  A2 ~~  C2     1     1        0.000 0.000     NA     NA    0.000    0.000
## 16  C1 ~~  C2     1     1        1.000 0.000     NA     NA    1.000    1.000
## 17  P1 ~1         1     1       -0.001 0.055 -0.012  0.990   -0.108    0.107
## 18  P2 ~1         1     1       -0.045 0.055 -0.821  0.412   -0.153    0.063
## 19  A1 ~1         1     1        0.000 0.000     NA     NA    0.000    0.000
## 20  A2 ~1         1     1        0.000 0.000     NA     NA    0.000    0.000
## 21  C1 ~1         1     1        0.000 0.000     NA     NA    0.000    0.000
## 22  C2 ~1         1     1        0.000 0.000     NA     NA    0.000    0.000
## 23  A1 =~  P1     2     2     a  1.037 0.068 15.223  0.000    0.903    1.170
## 24  A2 =~  P2     2     2     a  1.037 0.068 15.223  0.000    0.903    1.170
## 25  C1 =~  P1     2     2     c  0.959 0.065 14.747  0.000    0.831    1.086
## 26  C2 =~  P2     2     2     c  0.959 0.065 14.747  0.000    0.831    1.086
## 27  A1 ~~  A1     2     2        1.000 0.000     NA     NA    1.000    1.000
## 28  A2 ~~  A2     2     2        1.000 0.000     NA     NA    1.000    1.000
## 29  C1 ~~  C1     2     2        1.000 0.000     NA     NA    1.000    1.000
## 30  C2 ~~  C2     2     2        1.000 0.000     NA     NA    1.000    1.000
## 31  P1 ~~  P1     2     2    e2  1.022 0.045 22.650  0.000    0.933    1.110
## 32  P2 ~~  P2     2     2    e2  1.022 0.045 22.650  0.000    0.933    1.110
## 33  A1 ~~  A2     2     2        0.500 0.000     NA     NA    0.500    0.500
## 34  A1 ~~  C1     2     2        0.000 0.000     NA     NA    0.000    0.000
## 35  A1 ~~  C2     2     2        0.000 0.000     NA     NA    0.000    0.000
## 36  A2 ~~  C1     2     2        0.000 0.000     NA     NA    0.000    0.000
## 37  A2 ~~  C2     2     2        0.000 0.000     NA     NA    0.000    0.000
## 38  C1 ~~  C2     2     2        1.000 0.000     NA     NA    1.000    1.000
## 39  P1 ~1         2     2        0.061 0.045  1.355  0.175   -0.027    0.149
## 40  P2 ~1         2     2        0.026 0.045  0.588  0.557   -0.062    0.114
## 41  A1 ~1         2     2        0.000 0.000     NA     NA    0.000    0.000
## 42  A2 ~1         2     2        0.000 0.000     NA     NA    0.000    0.000
## 43  C1 ~1         2     2        0.000 0.000     NA     NA    0.000    0.000
## 44  C2 ~1         2     2        0.000 0.000     NA     NA    0.000    0.000
```

And finally lets have a look at the model in a path diagram with tidySEM:

```
lay <- get_layout("","","","",    "","","",
                  "A1","","C1","",    "A2","","C2",
                  ""  ,"P1","","",    ""  ,"P2","", rows = 3)

graph_sem(model = ace.fit,layout=lay,variance_diameter=.3,angle = 180,rect_height=.35,ellipses_height=.
```



4.2 ADE model

The "ADE" model operationalized a phenotype as a function of dditive genetic variance (A), non additive genetic effects (D) (can be geneXgene interacting, can be dominant inheritance where a single allele is enough to express the trait regardless of the state of the other allele), and environment unique to each of the individual twins (E) (can private friends, being in separate classrooms, but also measurement error). The correlation between the genetic influences is 1 and non-additive genetic influences is 1 for MZ twins while the correlation between the additive genetic influences is 0.5 for DZ twins and the correlation for the non-additive genetc infleunces is 0.25.

the ACE model assumes among other things:

1. The absence of influences of the shared environment in which twins grow up on the outcome
2. the absence of sibling influences or rater contrast effect
3. The absence of gene by environment correlation
4. The absence of gene by environment interaction
5. The environments MZ twin grow up are as similar the environments DZ twins grow up in.

Lets simulate same data where the resemblance between twins is a function of equal parts (33.3%/33.3%/33.4%) a

```
A <- matrix(1,2,2) # genetic correlation for MZ's = 1
D <- matrix(1,2,2)
E <- diag(2)
Adz <- matrix(c(1,.5,.5,1),2,2) # additive genetic correlation for DZ's = 0.5
```

```
Ddz <- matrix(c(1,.25,.25,1),2,2)  # non-additive genetic correlation for DZ's = 0.5

# make 1000 pairs of MZ twins
MZ <- mvrnorm(1000,mu=c(0,0),Sigma = A+D+E)

# Add a column to label as MZ:
MZ<- cbind.data.frame("MZ",MZ)
colnames(MZ) <- c("zyg","P1", "P2")


# make 1500 DZ twin pairs
DZ <- mvrnorm(1500,mu=c(0,0),Sigma = Adz+Ddz+E)

# add variable too label as DZ:
DZ <- cbind.data.frame("DZ",DZ)
colnames(DZ) <- c("zyg","P1", "P2")

# Combine MZ and DZ twins
dataset <- rbind(MZ,DZ)
```

We then define the lavaan model that can express the variance in the trait P explained by latent variables A, C and E:

```
ade.model<-"
A1=~ NA*P1 + c(a,a)*P1
A2=~ NA*P2 + c(a,a)*P2
D1 =~ NA*P1 + c(d,d)*P1
D2 =~ NA*P2 + c(d,d)*P2
# variances
A1 ~~ 1*A1
A2 ~~ 1*A2
D1 ~~ 1*D1
D2 ~~ 1*D2
P1~~c(e2,e2)*P1
P2~~c(e2,e2)*P2
# covariances
A1 ~~ c(1,.5)*A2
A1 ~~ 0*D1 + 0*D2
A2 ~~ 0*D1 + 0*D2
D1 ~~ c(1,.25)*D2"
```

Lets look at some of the critical lines of code in the model:

`A1=~ NA*P1 + c(a,a)*P1` Here we create the latent variable A1, the phenotype P for twin 1 (P1) loads on this variable, and in both groups (groups being MZ and DZ twins) the influence of this latent variable on the outcome is the same (contained using `c(a,a)`). Similar code is used to define the latent variables D1 and D2. Now the effect of genes on an outcome is assumed the same for everyone regardless of whether they are twins, or not, the resemblance between twin 1 and twin 2 difference for MZ and DZ twins. We define/fix the resemblance later in the model here: `A1 ~~ c(1,.5)*A2`, because A1 and A2 are variance 1: `A1 ~~ 1*A1` and `A2 ~~ 1*A2` the constrained implies a correlation of 1 for the MZ twins and a correlation of 0.5 for the DZ twins. The non-additive genetic effects are correlated 1 for MZ twins and .25 for DZ twins `C1 ~~ c(1,.25)*C2`, while the unshared environment E is conceptualized as a residual variance of the trait P (P1 or P2 respectively): `P1~~c(e2,e2)*P1`

We assume the latent variables A(1/2) and D(1/2) are uncorrelated, and fix their covariance to 0:

```
A1 ~~ 0*D1 + 0*D2
A2 ~~ 0*D1 + 0*D2
```

We also assume the residual variance (E) is uncorrelated to A and D, but fortunately for us this is a lavaan default.We proceed to fit the model to the simulated data:

```
# Standard ace model:
ade.fit<-cfa(ade.model, data = dataset,group = "zyg")
parameterEstimates(ade.fit)
```

```
##    lhs op rhs block group label    est    se      z pvalue ci.lower ci.upper
## 1   A1 =~  P1     1     1     a  0.933 0.156  5.978  0.000    0.627    1.238
## 2   A2 =~  P2     1     1     a  0.933 0.156  5.978  0.000    0.627    1.238
## 3   D1 =~  P1     1     1     d  1.077 0.137  7.885  0.000    0.809    1.344
## 4   D2 =~  P2     1     1     d  1.077 0.137  7.885  0.000    0.809    1.344
## 5   A1 ~~  A1     1     1        1.000 0.000     NA     NA    1.000    1.000
## 6   A2 ~~  A2     1     1        1.000 0.000     NA     NA    1.000    1.000
## 7   D1 ~~  D1     1     1        1.000 0.000     NA     NA    1.000    1.000
## 8   D2 ~~  D2     1     1        1.000 0.000     NA     NA    1.000    1.000
## 9   P1 ~~  P1     1     1    e2  0.955 0.042 22.630  0.000    0.873    1.038
## 10  P2 ~~  P2     1     1    e2  0.955 0.042 22.630  0.000    0.873    1.038
## 11  A1 ~~  A2     1     1        1.000 0.000     NA     NA    1.000    1.000
## 12  A1 ~~  D1     1     1        0.000 0.000     NA     NA    0.000    0.000
## 13  A1 ~~  D2     1     1        0.000 0.000     NA     NA    0.000    0.000
## 14  A2 ~~  D1     1     1        0.000 0.000     NA     NA    0.000    0.000
## 15  A2 ~~  D2     1     1        0.000 0.000     NA     NA    0.000    0.000
## 16  D1 ~~  D2     1     1        1.000 0.000     NA     NA    1.000    1.000
## 17  P1 ~1         1     1       -0.057 0.055 -1.038  0.299   -0.164    0.050
## 18  P2 ~1         1     1       -0.013 0.055 -0.235  0.814   -0.120    0.094
## 19  A1 ~1         1     1        0.000 0.000     NA     NA    0.000    0.000
## 20  A2 ~1         1     1        0.000 0.000     NA     NA    0.000    0.000
## 21  D1 ~1         1     1        0.000 0.000     NA     NA    0.000    0.000
## 22  D2 ~1         1     1        0.000 0.000     NA     NA    0.000    0.000
## 23  A1 =~  P1     2     2     a  0.933 0.156  5.978  0.000    0.627    1.238
## 24  A2 =~  P2     2     2     a  0.933 0.156  5.978  0.000    0.627    1.238
## 25  D1 =~  P1     2     2     d  1.077 0.137  7.885  0.000    0.809    1.344
## 26  D2 =~  P2     2     2     d  1.077 0.137  7.885  0.000    0.809    1.344
## 27  A1 ~~  A1     2     2        1.000 0.000     NA     NA    1.000    1.000
## 28  A2 ~~  A2     2     2        1.000 0.000     NA     NA    1.000    1.000
## 29  D1 ~~  D1     2     2        1.000 0.000     NA     NA    1.000    1.000
## 30  D2 ~~  D2     2     2        1.000 0.000     NA     NA    1.000    1.000
## 31  P1 ~~  P1     2     2    e2  0.955 0.042 22.630  0.000    0.873    1.038
## 32  P2 ~~  P2     2     2    e2  0.955 0.042 22.630  0.000    0.873    1.038
## 33  A1 ~~  A2     2     2        0.500 0.000     NA     NA    0.500    0.500
## 34  A1 ~~  D1     2     2        0.000 0.000     NA     NA    0.000    0.000
## 35  A1 ~~  D2     2     2        0.000 0.000     NA     NA    0.000    0.000
## 36  A2 ~~  D1     2     2        0.000 0.000     NA     NA    0.000    0.000
## 37  A2 ~~  D2     2     2        0.000 0.000     NA     NA    0.000    0.000
## 38  D1 ~~  D2     2     2        0.250 0.000     NA     NA    0.250    0.250
## 39  P1 ~1         2     2        0.008 0.045  0.186  0.852   -0.079    0.096
## 40  P2 ~1         2     2        0.050 0.045  1.121  0.262   -0.037    0.137
## 41  A1 ~1         2     2        0.000 0.000     NA     NA    0.000    0.000
## 42  A2 ~1         2     2        0.000 0.000     NA     NA    0.000    0.000
## 43  D1 ~1          2     2        0.000 0.000     NA     NA    0.000    0.000
## 44  D2 ~1          2     2        0.000 0.000     NA     NA    0.000    0.000
```

4.3 Sibling interactions

Sibling interactions, or one sibling influencing the others outcome, are an additional mechanism by which twins and siblings can become more alike, or if it concerns a negative sibling interaction become less alike. Sibling interaction cannot be distinguished from rater contrast effects where one child's trait changes to norm or view the rater (usually parent) has of the other child. If one of my children is very quite, the other might seem louder especially in contrast to the other. It is reasonable to assume that sibling interacting effects that persist across self, parental and teacher ratings (especially in the case of twins rated by different teachers) and external or formal measurements are more likely to reflect actual sibling interaction then rater contrast effects.

The "ACE" sibling interaction model (ACEx) operationalized a phenotype as a function of additive genetic variance (A), common environmental influences(C) (can be rearing, can be societal influences can be governmental policies), and environment unique to each of the individual twins (E) (can be private friends, being in separate classrooms, but also measurement error). The correlation between the genetic influences and common environmental is 1 for MZ twins while the correlation between the genetic influences is 0.5 for DZ twins in addition to these influences the sibling phenotypes are regressed on each other concurrently and the regression in each direction and across MZ and DZ twins is set to be equal.

The ACE sibling interaction model assumes among other things:

1. The absence of non-additive genetic effects (gene gene interaction and dominance)
2. The absence of gene by environment correlation
3. The absence of gene by environment interaction
4. The environments MZ twin grow up are as similaras the environments DZ twins grow up in.

There is also a ADEx model which is a model one can consider when there is evidence for D in the basic twin model, it is pasted below the example for the ACEx. The variable "beta" represents the magnitude of the sib interaction, its set high (0.4) and we cranked up the simulated sample size as an ACEx is very power hungry. In practice you could consider comparing the AEx (settign C to 0) with an ACE model without sibling interaction (setting the interaction to 0) or, if you suspect negative interactions compare the the respective AEx and ADE models.

```r
A <- matrix(1,2,2)
C <- matrix(1,2,2)
E <- diag(2)
Adz <- matrix(c(1,.5,.5,1),2,2)

#sibling interactie effect:
beta <- .4

# MZ twins

MZ <- mvrnorm(4000,mu=c(0,0),Sigma = A+C+E)

# regress the sibs on eachother
MZ <- t(matrix(c(1,beta,beta,1),2,2) %*% t(MZ))

# MZ data frame mwithcolumns zygositiy ("MZ"), twin1 data, twin2 data
MZ <- cbind.data.frame("MZ",MZ)
```

```
colnames(MZ) <- c("zyg","P1", "P2")


# DZ twins
DZ <- mvrnorm(6000,mu=c(0,0),Sigma = Adz+C+E)

# add sib interaction
DZ <- t(matrix(c(1,beta ,beta ,1),2,2) %*% t(DZ))

DZ <- cbind.data.frame("DZ",DZ)

colnames(DZ) <- c("zyg","P1", "P2")

# cmbineer MZ en DZ in een dataset:
dataset <- rbind(MZ,DZ)
```

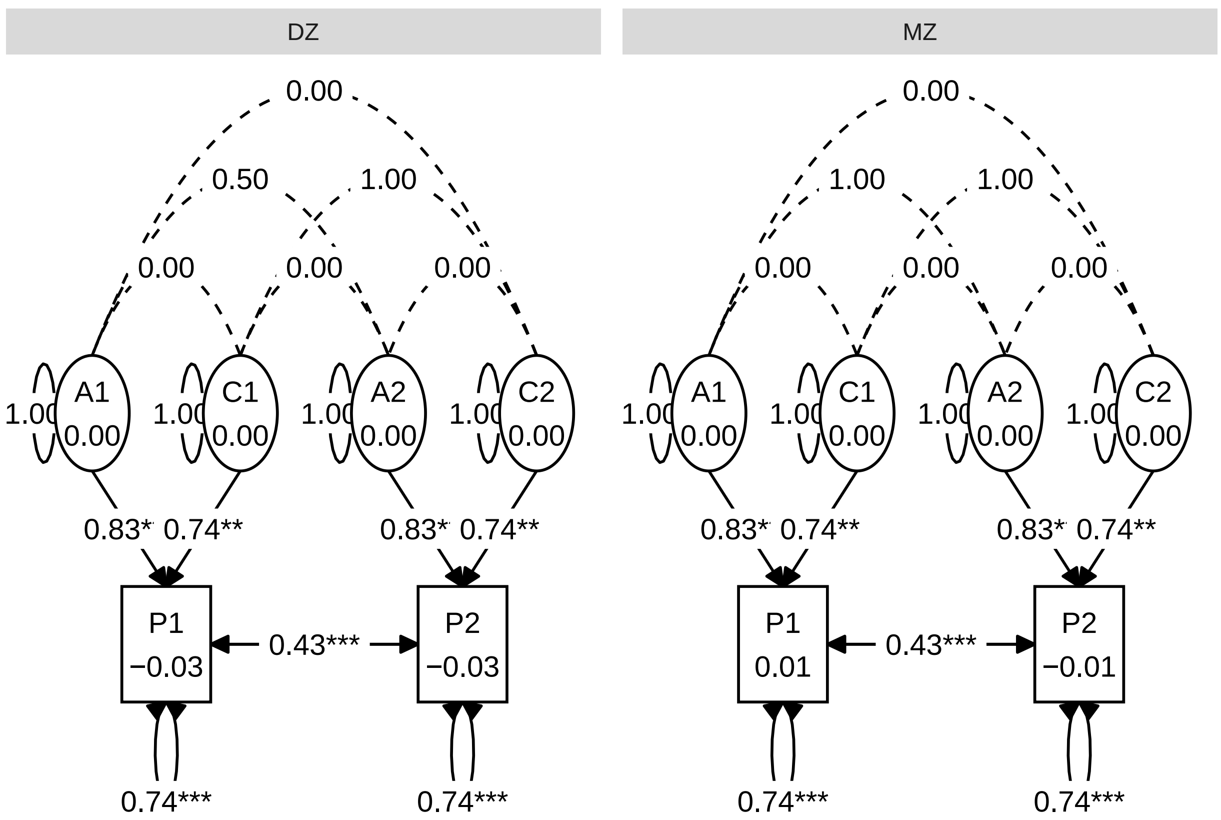
Having generated data with a sibling interaction we can go ahead and fit the sibling interaction model below to retrieve the simulate parameters.

```
# Models

# Sibling interaction model:
ace.model.sib.int<-"
A1=~ NA*P1 + c(a,a)*P1
A2=~ NA*P2 + c(a,a)*P2
C1 =~ NA*P1 + c(c,c)*P1
C2 =~ NA*P2 + c(c,c)*P2
# variances
A1 ~~ 1*A1
A2 ~~ 1*A2
C1 ~~ 1*C1
C2 ~~ 1*C2
P1~~c(e2,e2)*P1
P2~~c(e2,e2)*P2
# covariances
A1 ~~ c(1,.5)*A2
A1 ~~ 0*C1 + 0*C2
A2 ~~ 0*C1 + 0*C2
C1 ~~ c(1,1)*C2

# regs
P1 ~ c(beta,beta)*P2
P2 ~ c(beta,beta)*P1
"

# Sibling interaction:
ace.fitsib <-cfa(ace.model.sib.int, data = dataset,group = "zyg")
parameterEstimates(ace.fitsib)
```

```
##    lhs op rhs block group label   est    se      z pvalue ci.lower ci.upper
## 1   A1 =~  P1     1     1     a 0.828 0.038 21.919  0.000    0.754    0.902
## 2   A2 =~  P2     1     1     a 0.828 0.038 21.919  0.000    0.754    0.902
## 3   C1 =~  P1     1     1     c 0.739 0.238  3.101  0.002    0.272    1.206
## 4   C2 =~  P2      1     1     c  0.739 0.238  3.101  0.002    0.272    1.206
## 5   A1 ~~  A1      1     1        1.000 0.000     NA     NA    1.000    1.000
## 6   A2 ~~  A2      1     1        1.000 0.000     NA     NA    1.000    1.000
## 7   C1 ~~  C1      1     1        1.000 0.000     NA     NA    1.000    1.000
## 8   C2 ~~  C2      1     1        1.000 0.000     NA     NA    1.000    1.000
## 9   P1 ~~  P1      1     1    e2  0.743 0.062 11.972  0.000    0.621    0.864
## 10  P2 ~~  P2      1     1    e2  0.743 0.062 11.972  0.000    0.621    0.864
## 11  A1 ~~  A2      1     1        1.000 0.000     NA     NA    1.000    1.000
## 12  A1 ~~  C1      1     1        0.000 0.000     NA     NA    0.000    0.000
## 13  A1 ~~  C2      1     1        0.000 0.000     NA     NA    0.000    0.000
## 14  A2 ~~  C1      1     1        0.000 0.000     NA     NA    0.000    0.000
## 15  A2 ~~  C2      1     1        0.000 0.000     NA     NA    0.000    0.000
## 16  C1 ~~  C2      1     1        1.000 0.000     NA     NA    1.000    1.000
## 17  P1  ~  P2      1     1  beta  0.432 0.055  7.887  0.000    0.325    0.540
## 18  P2  ~  P1      1     1  beta  0.432 0.055  7.887  0.000    0.325    0.540
## 19  P1 ~1          1     1        0.006 0.022  0.283  0.777   -0.037    0.050
## 20  P2 ~1          1     1       -0.007 0.022 -0.324  0.746   -0.051    0.036
## 21  A1 ~1          1     1        0.000 0.000     NA     NA    0.000    0.000
## 22  A2 ~1          1     1        0.000 0.000     NA     NA    0.000    0.000
## 23  C1 ~1          1     1        0.000 0.000     NA     NA    0.000    0.000
## 24  C2 ~1          1     1        0.000 0.000     NA     NA    0.000    0.000
## 25  A1 =~  P1      2     2     a  0.828 0.038 21.919  0.000    0.754    0.902
## 26  A2 =~  P2      2     2     a  0.828 0.038 21.919  0.000    0.754    0.902
## 27  C1 =~  P1      2     2     c  0.739 0.238  3.101  0.002    0.272    1.206
## 28  C2 =~  P2      2     2     c  0.739 0.238  3.101  0.002    0.272    1.206
## 29  A1 ~~  A1      2     2        1.000 0.000     NA     NA    1.000    1.000
## 30  A2 ~~  A2      2     2        1.000 0.000     NA     NA    1.000    1.000
## 31  C1 ~~  C1      2     2        1.000 0.000     NA     NA    1.000    1.000
## 32  C2 ~~  C2      2     2        1.000 0.000     NA     NA    1.000    1.000
## 33  P1 ~~  P1      2     2    e2  0.743 0.062 11.972  0.000    0.621    0.864
## 34  P2 ~~  P2      2     2    e2  0.743 0.062 11.972  0.000    0.621    0.864
## 35  A1 ~~  A2      2     2        0.500 0.000     NA     NA    0.500    0.500
## 36  A1 ~~  C1      2     2        0.000 0.000     NA     NA    0.000    0.000
## 37  A1 ~~  C2      2     2        0.000 0.000     NA     NA    0.000    0.000
## 38  A2 ~~  C1      2     2        0.000 0.000     NA     NA    0.000    0.000
## 39  A2 ~~  C2      2     2        0.000 0.000     NA     NA    0.000    0.000
## 40  C1 ~~  C2      2     2        1.000 0.000     NA     NA    1.000    1.000
## 41  P1  ~  P2      2     2  beta  0.432 0.055  7.887  0.000    0.325    0.540
## 42  P2  ~  P1      2     2  beta  0.432 0.055  7.887  0.000    0.325    0.540
## 43  P1 ~1          2     2       -0.029 0.018 -1.556  0.120   -0.064    0.007
## 44  P2 ~1          2     2       -0.028 0.018 -1.528  0.126   -0.064    0.008
## 45  A1 ~1          2     2        0.000 0.000     NA     NA    0.000    0.000
## 46  A2 ~1          2     2        0.000 0.000     NA     NA    0.000    0.000
## 47  C1 ~1          2     2        0.000 0.000     NA     NA    0.000    0.000
## 48  C2 ~1          2     2        0.000 0.000     NA     NA    0.000    0.000
```

Lets proceed to visualize the path diagram

```
lay <- get_layout("A1","","C1","",    "A2","","C2",
                  ""  ,"P1","","",    ""  ,"P2","", rows = 2)

graph_sem(model = ace.fitsib,layout=lay,variance_diameter=.3,angle = 180,rect_height=.35,ellipses_heigh
```



And for completeness here is the ADE sibling interaction model:

```
# Sibling interaction model:
ade.model.sib.int<-"
A1=~ NA*P1 + c(a,a)*P1
A2=~ NA*P2 + c(a,a)*P2
D1 =~ NA*P1 + c(c,c)*P1
D2 =~ NA*P2 + c(c,c)*P2
# variances
A1 ~~ 1*A1
A2 ~~ 1*A2
D1 ~~ 1*D1
D2 ~~ 1*D2
P1~~c(e2,e2)*P1
P2~~c(e2,e2)*P2
# covariances
A1 ~~ c(1,.5)*A2
A1 ~~ 0*D1 + 0*D2
A2 ~~ 0*D1 + 0*D2
D1 ~~ c(1,.25)*D2

# regs
P1 ~ c(beta,beta)*P2
P2 ~ c(beta,beta)*P1
"
```

4.4 Gene-environment interaction

The most intuitive models for gene-environment interaction with a continuous environment(Purcell 2002) cannot be fitted in lavaan, please consider OpenMx or Mplus.

4.5 Gene-environment correlation

4.6 Sex specific effects

4.7 Binary/Ordinal data

4.8 Rater bias models

5 Multitrait twin models in lavaan

5.1 Cholesky decomposition

5.2 Direction of Causation models

6 Longitudinal twin models

6.1 Auto regressive models

6.1.1 Phenotype to environment effects

6.2 Growth curve models

References

Bates, Timothy C., Hermine Maes, and Michael C. Neale. 2019. “Umx: Twin and Path-Based Structural Equation Modeling in R.” *Twin Research and Human Genetics* 22 (1): 27–41. https://doi.org/10.1017/thg.2019.2.

Huang, Po-Hsien, Hung Chen, and Li-Jen Weng. 2017. “A Penalized Likelihood Method for Structural Equation Modeling.” *Psychometrika* 82 (2): 329–54. https://doi.org/10.1007/s11336-017-9566-9.

Jacobucci, Ross, Kevin J. Grimm, and John J. McArdle. 2016. “Regularized Structural Equation Modeling.” *Structural Equation Modeling: A Multidisciplinary Journal* 23 (4): 555–66. https://doi.org/10.1080/10705511.2016.1154793.

Merkle, Edgar C., and Yves Rosseel. 2018. “Blavaan: Bayesian Structural Equation Models via Parameter Expansion.” *Journal of Statistical Software* 85 (4). https://doi.org/10.18637/jss.v085.i04.

Neale, Michael C., Michael D. Hunter, Joshua N. Pritikin, Mahsa Zahery, Timothy R. Brick, Robert M. Kirkpatrick, Ryne Estabrook, Timothy C. Bates, Hermine H. Maes, and Steven M. Boker. 2015. “OpenMx 2.0: Extended Structural Equation and Statistical Modeling.” *Psychometrika* 81 (2): 535–49. https://doi.org/10.1007/s11336-014-9435-8.

Purcell, Shaun. 2002. “Variance Components Models for GeneEnvironment Interaction in Twin Analysis.” *Twin Research* 5 (6): 554–71. https://doi.org/10.1375/136905202762342026.

Rosseel, Yves. 2012. “Lavaan: AnRPackage for Structural Equation Modeling.” *Journal of Statistical Software* 48 (2). https://doi.org/10.18637/jss.v048.i02.