Analysis of Twin Data: Multivariate Models

J. Hjelmborg and C. Fagnani

Epidemiology, Biostatistics and Biodemography, Institute of Public Health,
University of Southern Denmark, Denmark

May 2015

Prologue

Effect?

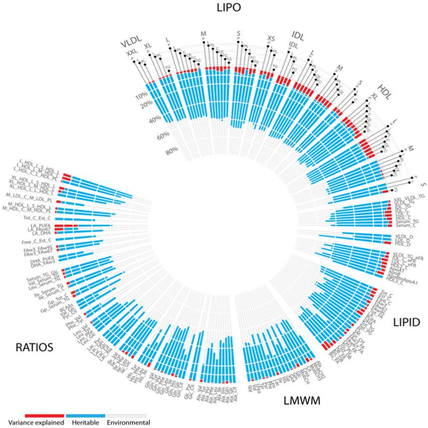
Exposure→Outcome

- Outcome: There are multiple outcomes! (eg. Telomere length, HDL, and BMI).
- What is the contribution of genetic and environmental factors to the variation in outcome?

$$\begin{cases} Y = \text{Genes} + \text{Environment} \\ \Sigma_Y = \Sigma_{\text{Genes}} + \Sigma_{\text{Environment}} \end{cases}$$

- What kind of genetic and environmental influences to expect?
- Are the same or different genes influencing the traits?

Metabonomics with twins



Kaprio et al

Aims of multivariate twin analyses

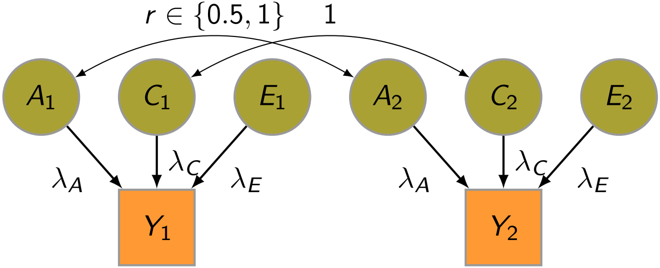
Scope of study

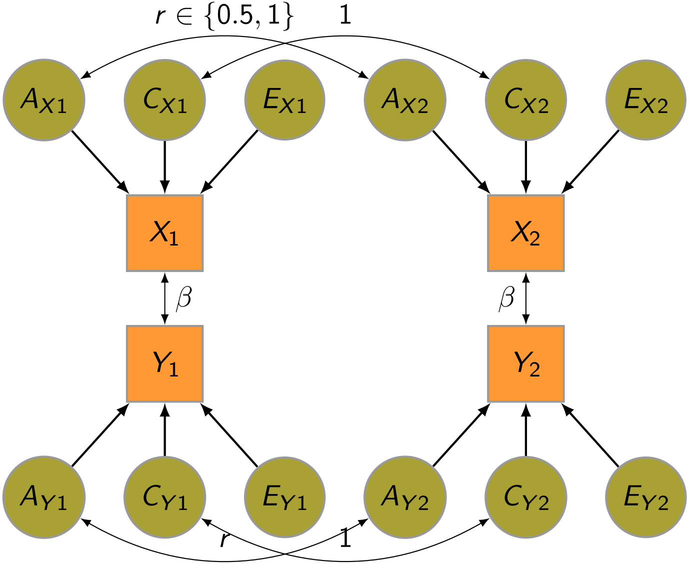
- Co-occurrence or co-morbidity of different diseases.
- Inter-relations, interactions, confounding and moderation effects.
- Genetic or environmental overlap between traits, that is, origin of comorbidity
  - pleiotropic genetic effects
  - environmental overlap: prevention strategies impacting on multiple diseases.
- Developmental changes (longitudinal data).

Outline - Multiple phenotypes measured in twins

Aims of Multivariate Analysis

- Structural Equation Modeling
- Example: BMI in Finnish adult twins
    - the full multivariate ACE model
    - the independent pathway model
    - the common pathway model
    - the growth curve model
    - modeling extremes


$r \in \{0.5, 1\}$
1
A_1
C_1
E_1
A_2
C_2
E_2
λ_A
λ_C
λ_E
λ_A
λ_C
λ_E
Y_1
Y_2


$r \in \{0.5, 1\}$
1
A_{X1}
C_{X1}
E_{X1}
A_{X2}
C_{X2}
E_{X2}
X_1
X_2
β
β
Y_1
Y_2
A_{Y1}
C_{Y1}
E_{Y1}
A_{Y2}
C_{Y2}
E_{Y2}
r
1

- Contributing factors to the variation in outcome:

$$\Sigma_Y = \begin{pmatrix} \Sigma_A & r\Sigma_A \\ r\Sigma_A & \Sigma_A \end{pmatrix} + \begin{pmatrix} \Sigma_C & \Sigma_C \\ \Sigma_C & \Sigma_C \end{pmatrix} + \begin{pmatrix} \Sigma_E & 0 \\ 0 & \Sigma_E \end{pmatrix}$$

where $r = 1$ for MZ pairs and $z = \frac{1}{2}$ for DZ pairs.

In particular, we obtain

- Heritability:

$$h_Y^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_C^2 + \sigma_E^2}$$

- Shared environmental effect:

$$c_Y^2 = \frac{\sigma_C^2}{\sigma_A^2 + \sigma_C^2 + \sigma_E^2}$$

Main assumptions

- Equal environments assumption for MZ and DZ twins.
- No gene-environment interaction and correlation.
- No gene-gene interaction (link: epistasis).
- Equal mean and variance of twin 1 and twin 2, MZ and DZ.
- Estimation and inference by maximum likelihood principle assuming bivariate normality of paired observations (as before).

SEM - General form

SEM

- A general SEM is of form, $\boxed{\eta = B\eta + \Gamma\xi + \zeta}$, where η and ξ denotes endogenous and exogenous variables respectively, B and Γ are matrices of coefficients and ζ denotes errors.
- -this induces the covariance matrix, $\Sigma(\theta)$, to be compared with the observed covariance matrix from the sample.
- -indeed the purpose of programs LISREL, Mx, OpenMx, M-Plus, lava, Stata SEM, ...

SEM - Great many models

- The focus in SEM is the basically the covariance matrix, $\Sigma = \Sigma(\theta)$.
- Great many statistical methods can be formulated via SEM.
- Simple linear regression, $Y = \beta X + \epsilon$, corresponds in SEM to
$$\Sigma_Y = \begin{pmatrix} \beta^2\sigma_X^2 + \sigma_\epsilon^2 & \beta\sigma_X^2 \\ \beta\sigma_X^2 & \sigma_X^2 \end{pmatrix}$$
- -now, find parameters $\beta \in \theta$ minimizing the difference between the sample covariance matrix and the one predicted by the model, the right-side.

OpenMx

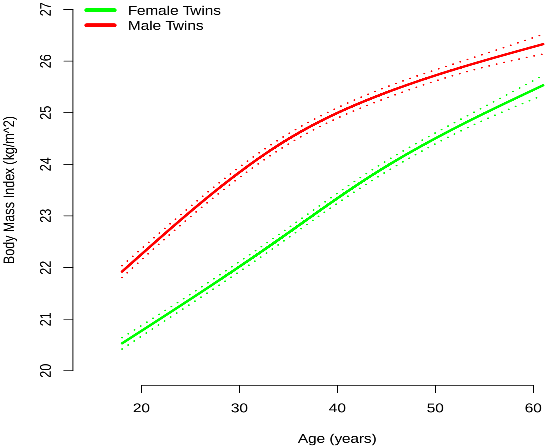
- Statistical package in R for Structural Equation Modeling (SEM)
- The focus in SEM is the basically the covariance matrix, $\Sigma = \Sigma(\theta)$.

The Finnish Twin Cohort: BMI

- There are seven waves of BMI in twins observed i 1975, 1981 and 1990.
- References: **Genetic influences on growth traits of bmi, *Obesity*, 16(4):847–852, 2008.**

  The paper deals with the problem of analysing longitudinal data in twins. A bivariate growth-curve model is suggested for measurements of BMI in Finnish twin data that allows for studying genetic influence of latent traits and their mutual dependence (genetic pleiotropy). The results recovered substantial genetic influence on level and found moderate influence on change in BMI. Software for the analysis was implemented in order to obtain specific constrains on marginals which is desirable for randomly selected twin pairs, MZ and DZ.
- Let's consider analysis using `R`

BMI by age (fitted by a 'gamm model')



Finnish males: BMI over seven waves

The Data - Description

```
ext <- read.delim("data/bmicotwinIntact.out", header=T
out <- lm(bmi ~ age+sex+factor(wave), data=ext)
kable(summary(out)$coef, digits=2)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	21.27	0.06	327.90	0.00
age	0.14	0.00	72.22	0.00
sex	-1.61	0.02	-64.86	0.00
factor(wave)2	-0.07	0.05	-1.41	0.16
factor(wave)3	-0.20	0.05	-3.92	0.00
factor(wave)4	-0.30	0.05	-5.41	0.00
factor(wave)5	-0.31	0.06	-5.26	0.00
factor(wave)6	-0.31	0.06	-4.90	0.00
factor(wave)7	-0.48	0.07	-7.30	0.00

```
library(OpenMx)
```

Male BMI data - descriptives

```
kable(describe(mzData, skew=F, ranges=F))
```

	vars	n	mean	sd	se
bmi_1_1	1	612	22.28269	2.181211	0.0881702
bmi_2_1	2	612	23.35844	2.364580	0.0955825
bmi_3_1	3	612	24.15402	2.627597	0.1062143
bmi_4_1	4	612	24.43920	2.860610	0.1156333
bmi_5_1	5	612	24.87601	3.155865	0.1275683
bmi_6_1	6	612	25.31446	3.305714	0.1336256
bmi_7_1	7	612	25.17276	3.312678	0.1339071
bmi_1_2	8	612	22.41243	2.365447	0.0956175
bmi_2_2	9	612	23.46456	2.559846	0.1034756
bmi_3_2	10	612	24.05678	2.755968	0.1114034
bmi_4_2	11	612	24.39862	2.744412	0.1109363
bmi_5_2	12	612	24.86262	2.982877	0.1205757
bmi_6_2	13	612	25.24010	3.072810	0.1242110
bmi_7_2	14	612	25.13669	2.941579	0.1189063

Male BMI data - descriptives

```r
kable(describe(dzData, skew=F, ranges=F))
```

	vars	n	mean	sd	se
bmi_1_1	1	1228	22.34400	2.166324	0.0618193
bmi_2_1	2	1228	23.41275	2.438479	0.0695857
bmi_3_1	3	1228	24.17729	2.788454	0.0795727
bmi_4_1	4	1228	24.60156	2.883396	0.0822821
bmi_5_1	5	1228	24.99099	3.056260	0.0872150
bmi_6_1	6	1228	25.48215	3.209967	0.0916013
bmi_7_1	7	1228	25.36131	3.178992	0.0907173
bmi_1_2	8	1228	22.38894	2.182843	0.0622907
bmi_2_2	9	1228	23.48180	2.610679	0.0744996
bmi_3_2	10	1228	24.18079	2.696729	0.0769552
bmi_4_2	11	1228	24.58169	2.771909	0.0791006
bmi_5_2	12	1228	24.94828	3.008025	0.0858385
bmi_6_2	13	1228	25.35170	3.099764	0.0884564
bmi_7_2	14	1228	25.22387	3.048204	0.0869851

```r
plot(bmi_1_1 ~ bmi_1_2, data=mzData)
plot(bmi_1_1 ~ bmi_1_2, data=dzData)
```

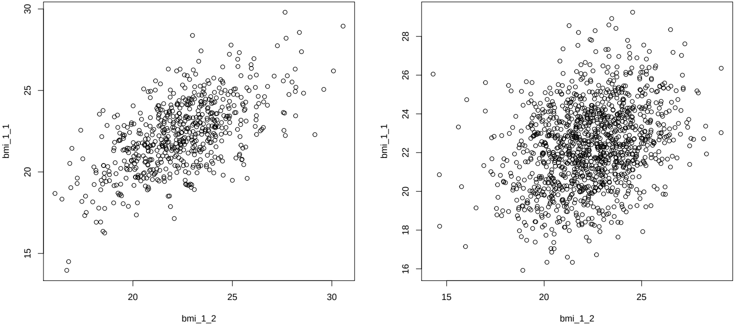


Figure: Male BMI versus co twin BMI for MZ and DZ pairs at first wave

Correlation matrix of seven waves in MZ pairs

```
x<-round(cov2cor(cov(mzData,use="complete")),1)
```

$$\begin{bmatrix}
1 & 0.7 & 0.6 & 0.5 & 0.5 & 0.5 & 0.5 & 0.6 & 0.5 & 0.4 & 0.4 & 0.4 & 0.4 & 0.4 \\
0.7 & 1 & 0.8 & 0.8 & 0.7 & 0.7 & 0.7 & 0.5 & 0.6 & 0.6 & 0.6 & 0.6 & 0.5 & 0.5 \\
0.6 & 0.8 & 1 & 0.8 & 0.8 & 0.8 & 0.8 & 0.4 & 0.6 & 0.7 & 0.7 & 0.6 & 0.6 & 0.6 \\
0.5 & 0.8 & 0.8 & 1 & 0.9 & 0.9 & 0.9 & 0.4 & 0.6 & 0.7 & 0.7 & 0.7 & 0.6 & 0.6 \\
0.5 & 0.7 & 0.8 & 0.9 & 1 & 0.9 & 0.9 & 0.4 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 \\
0.5 & 0.7 & 0.8 & 0.9 & 0.9 & 1 & 1 & 0.4 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 \\
0.5 & 0.7 & 0.8 & 0.9 & 0.9 & 1 & 1 & 0.4 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 \\
0.6 & 0.5 & 0.4 & 0.4 & 0.4 & 0.4 & 0.4 & 1 & 0.7 & 0.6 & 0.5 & 0.5 & 0.4 & 0.4 \\
0.5 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.7 & 1 & 0.8 & 0.8 & 0.8 & 0.7 & 0.7 \\
0.4 & 0.6 & 0.7 & 0.7 & 0.6 & 0.6 & 0.6 & 0.6 & 0.8 & 1 & 0.8 & 0.8 & 0.8 & 0.8 \\
0.4 & 0.6 & 0.7 & 0.7 & 0.6 & 0.6 & 0.6 & 0.5 & 0.8 & 0.8 & 1 & 0.9 & 0.8 & 0.8 \\
0.4 & 0.6 & 0.6 & 0.7 & 0.6 & 0.6 & 0.6 & 0.5 & 0.8 & 0.8 & 0.9 & 1 & 0.9 & 0.9 \\
0.4 & 0.5 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.4 & 0.7 & 0.8 & 0.8 & 0.9 & 1 & 1 \\
0.4 & 0.5 & 0.6 & 0.6 & 0.6 & 0.6 & 0.6 & 0.4 & 0.7 & 0.8 & 0.8 & 0.9 & 1 & 1
\end{bmatrix}$$

Correlation matrix of seven waves in DZ pairs

```r
x<-round(cov2cor(cov(dzData,use="complete")),1)
```

$$\begin{bmatrix}
1 & 0.7 & 0.6 & 0.6 & 0.5 & 0.5 & 0.5 & 0.3 & 0.3 & 0.3 & 0.2 & 0.2 & 0.2 & 0.3 \\
0.7 & 1 & 0.8 & 0.8 & 0.7 & 0.7 & 0.7 & 0.3 & 0.4 & 0.4 & 0.3 & 0.3 & 0.3 & 0.3 \\
0.6 & 0.8 & 1 & 0.9 & 0.9 & 0.8 & 0.8 & 0.3 & 0.4 & 0.4 & 0.4 & 0.4 & 0.3 & 0.4 \\
0.6 & 0.8 & 0.9 & 1 & 0.9 & 0.9 & 0.9 & 0.2 & 0.3 & 0.4 & 0.4 & 0.4 & 0.3 & 0.4 \\
0.5 & 0.7 & 0.9 & 0.9 & 1 & 0.9 & 0.9 & 0.2 & 0.3 & 0.4 & 0.3 & 0.3 & 0.3 & 0.3 \\
0.5 & 0.7 & 0.8 & 0.9 & 0.9 & 1 & 1 & 0.2 & 0.3 & 0.4 & 0.3 & 0.3 & 0.3 & 0.3 \\
0.5 & 0.7 & 0.8 & 0.9 & 0.9 & 1 & 1 & 0.2 & 0.3 & 0.4 & 0.3 & 0.3 & 0.3 & 0.3 \\
0.3 & 0.3 & 0.3 & 0.2 & 0.2 & 0.2 & 0.2 & 1 & 0.7 & 0.6 & 0.6 & 0.5 & 0.5 & 0.5 \\
0.3 & 0.4 & 0.4 & 0.3 & 0.3 & 0.3 & 0.3 & 0.7 & 1 & 0.7 & 0.7 & 0.7 & 0.6 & 0.6 \\
0.3 & 0.4 & 0.4 & 0.4 & 0.4 & 0.4 & 0.4 & 0.6 & 0.7 & 1 & 0.9 & 0.8 & 0.8 & 0.8 \\
0.2 & 0.3 & 0.4 & 0.4 & 0.3 & 0.3 & 0.3 & 0.6 & 0.7 & 0.9 & 1 & 0.9 & 0.9 & 0.8 \\
0.2 & 0.3 & 0.4 & 0.4 & 0.3 & 0.3 & 0.3 & 0.5 & 0.7 & 0.8 & 0.9 & 1 & 0.9 & 0.9 \\
0.2 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.5 & 0.6 & 0.8 & 0.9 & 0.9 & 1 & 1 \\
0.3 & 0.3 & 0.4 & 0.4 & 0.3 & 0.3 & 0.3 & 0.5 & 0.6 & 0.8 & 0.8 & 0.9 & 1 & 1
\end{bmatrix}$$

Multivariate ACE model fitted to seven waves

```
mxCompare(CholAceFit,CholAeFit)  # C seems important.


##      base comparison  ep minus2LL    df      AIC   diffLL diffdf           p
## 1 CholACE       <NA> 119 91335.02 25641 40053.02       NA     NA          NA
## 2 CholACE     CholAE  91 91389.54 25669 40051.54 54.51227     28 0.001945057
```

- Common environmental effects seems important.
- Further model selection and check of assumptions in Practicals to follow.

Multivariate ACE model fitted to seven waves

```
x<-round(CholAceFit$H2$result,2)
```

$$\Sigma_{H^2} = \begin{bmatrix} 0.48 & 0.54 & 0.51 & 0.64 & 0.6 & 0.66 & 0.65 \\ 0.54 & 0.44 & 0.54 & 0.63 & 0.58 & 0.63 & 0.64 \\ 0.51 & 0.54 & 0.48 & 0.62 & 0.6 & 0.62 & 0.63 \\ 0.64 & 0.63 & 0.62 & 0.6 & 0.61 & 0.63 & 0.63 \\ 0.6 & 0.58 & 0.6 & 0.61 & 0.52 & 0.58 & 0.59 \\ 0.66 & 0.63 & 0.62 & 0.63 & 0.58 & 0.52 & 0.55 \\ 0.65 & 0.64 & 0.63 & 0.63 & 0.59 & 0.55 & 0.53 \end{bmatrix}$$

- Heritabilities of seven waves along diagonal.
- Bivariate heritabilities off the diagonal. It is 0.65 between wave 1 and 7, hence 65 percent of phenotypic correlation is mediated by shared genetic influence.

Multivariate ACE model fitted to seven waves

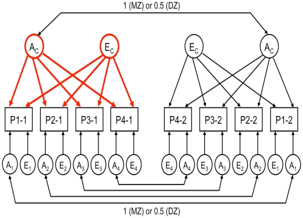
```
x<-round(cov2cor(CholAceFit$A$result),2)
```

$$\Sigma_{\mathrm{Corr}_A} = \begin{bmatrix} 1 & 0.81 & 0.62 & 0.65 & 0.63 & 0.62 & 0.62 \\ 0.81 & 1 & 0.88 & 0.92 & 0.89 & 0.86 & 0.86 \\ 0.62 & 0.88 & 1 & 1 & 1 & 0.97 & 0.97 \\ 0.65 & 0.92 & 1 & 1 & 0.99 & 0.96 & 0.96 \\ 0.63 & 0.89 & 1 & 0.99 & 1 & 0.98 & 0.98 \\ 0.62 & 0.86 & 0.97 & 0.96 & 0.98 & 1 & 1 \\ 0.62 & 0.86 & 0.97 & 0.96 & 0.98 & 1 & 1 \end{bmatrix}$$

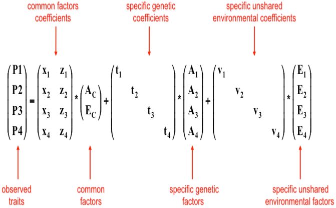
- Genetic correlation of seven waves of BMI, that is, correlation of genetic effects regardless of heritability.
- -the likelihood that a gene found to be associated with one trait is also associated with the other trait.

Independent pathway ACE model

- The covariation between traits is caused by genetic and environmental factors common to traits, each having its own paths to each trait

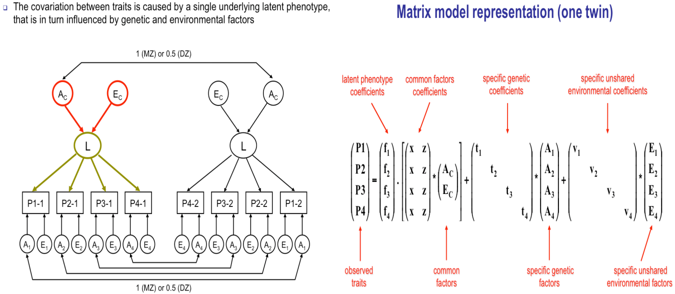


Matrix model representation (one twin)



$$\begin{pmatrix}P1\\P2\\P3\\P4\end{pmatrix} = \begin{pmatrix}x_1 & z_1\\ x_2 & z_2\\ x_3 & z_3\\ x_4 & z_4\end{pmatrix} * \begin{pmatrix}A_C\\ E_C\end{pmatrix} + \begin{pmatrix}t_1 & & & \\ & t_2 & & \\ & & t_3 & \\ & & & t_4\end{pmatrix} * \begin{pmatrix}A_1\\A_2\\A_3\\A_4\end{pmatrix} + \begin{pmatrix}v_1 & & & \\ & v_2 & & \\ & & v_3 & \\ & & & v_4\end{pmatrix} * \begin{pmatrix}E_1\\E_2\\E_3\\E_4\end{pmatrix}$$

Common pathway ACE model



- The covariation between traits is caused by a single underlying latent phenotype, that is in turn influenced by genetic and environmental factors

Matrix model representation (one twin)

$$\begin{pmatrix}P1\\P2\\P3\\P4\end{pmatrix}=\begin{pmatrix}f_1\\f_2\\f_3\\f_4\end{pmatrix}\cdot\left[\begin{pmatrix}x&z\\x&z\\x&z\\x&z\end{pmatrix}*\begin{pmatrix}A_C\\E_C\end{pmatrix}\right]+\begin{pmatrix}t_1&&&\\&t_2&&\\&&t_3&\\&&&t_4\end{pmatrix}*\begin{pmatrix}A_1\\A_2\\A_3\\A_4\end{pmatrix}+\begin{pmatrix}v_1&&&\\&v_2&&\\&&v_3&\\&&&v_4\end{pmatrix}*\begin{pmatrix}E_1\\E_2\\E_3\\E_4\end{pmatrix}$$

latent phenotype coefficients

common factors coefficients

specific genetic coefficients

specific unshared environmental coefficients

observed traits

common factors

specific genetic factors

specific unshared environmental factors

Which model to report?

```
kable(mxCompare(CholAceFit,list(IndAceFit,IndAceFit2,ComAceFit)), digits=2)
```

base	comparison	ep	minus2LL	df	AIC	diffLL	diffdf	p
CholACE	NA	119	91335.02	25641	40053.02	NA	NA	NA
CholACE	IndACE	77	92579.58	25683	41213.58	1244.56	42	0
CholACE	IndACE2	75	92298.62	25685	40928.62	963.59	44	0
CholACE	ComACE	66	95949.10	25695	44559.10	4614.08	54	0

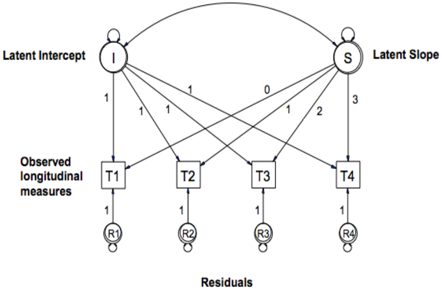
- The multivariate ACE model has the lowest AIC.
- Choosing among non-nested models can be delicate.
- The CP model can be tested as a sub-model of the IP model by likelihood-ratio chi-square test with $\mathrm{nparIP} - \mathrm{nparCP} = (\mathrm{nfac} - 1) * (\mathrm{nvar} - 1)$ degrees of freedom (e.g. $nvar = 4$, $nfac = 2$ (AE model), $\mathrm{degrees\ of\ freedom} = 3$).
- If χ^2-test is not significant, the CP should be preferred over the IP (CP more parsimonious).

The Growth Curve Model for Longitudinal Data

```
#plot(bmi ~ age, data=ext))
```

- The multivariate models above allows for
    - Magnitude of genetic and environmental influences at each time point (wave).
    - Extent to which genetic and environmental influences overlap across time points.
- Focus now on growth variables, eg. initial level ('intercept') and rate of change ('slope') - to predict level at a series of time points.

The Growth Curve model



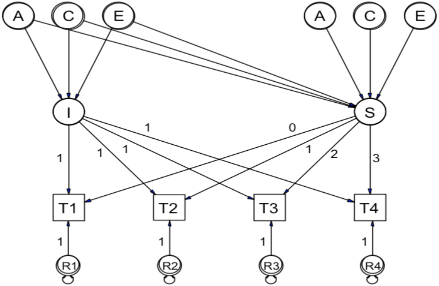
- The growth curve model allows for
  - modeling any number of time points (waves).
  - modeling the individual trajectory.

The Growth Curve model

Aims for the linear growth curve model

- Are there inter-individual differences in initial level and rate of change? - Variance of intercept and slope
- Are initial level and rate of change associated within an individual? - Within-twin correlation between intercept and slope
- Do genetic or environmental factors explain inter-individual differences in initial level and rate of change? - Cross-twin within-trait correlation of intercept and of slope in MZ and DZ twins
- Do genetic or environmental factors explain the within-individual association between initial level and rate of change? - Cross-twin/cross-trait correlation between intercept and slope in MZ and DZ twins
- To what extent are inter-individual differences in each of the longitudinal measures accounted for by initial level and rate of change? - Variance of residuals

The Growth Curve ACE model



- The growth curve ACE model allows for
    - Genetic and environmental influences on initial level and rate of change, and on their mutual interplay modelling any number of time points.
    - Very efficient: number of parameters does not increase with number of measurements.

The Growth Curve ACE model

```
kable(round(LgcAceFit$vars2@result,2)[,7:12])
```

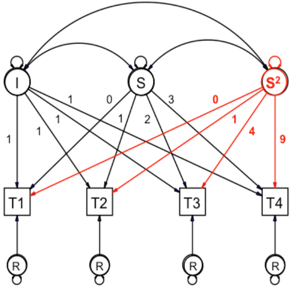
	SAis	SAis	SCis	SCis	SEis	SEis
Int	0.51	13.58	0.20	-1.77	0.28	-10.8
Slope	13.58	0.30	-1.77	0.00	-10.80	0.7

Aims for the linear growth curve ACE model

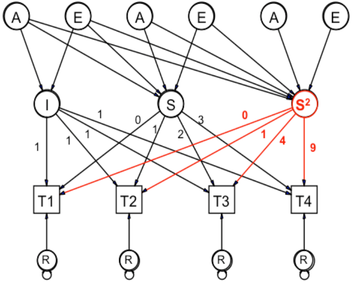
- What is the contribution of genetic factors to inter-individual variation in initial level and rate of change? - Heritability of intercept and slope.
- What is the contribution of environmental factors to inter-individual variation in initial level and rate of change? - Shared and unique environmental proportions of variance of intercept and slope.
- Same or different genes influencing initial level and rate of change? - Genetic correlation between intercept and slope. For BMI: 0.5046
- Same or different environments influencing initial level and rate of change? - Shared and unique environmental correlations between intercept and slope

The Growth Curve ACE model - extensions

General

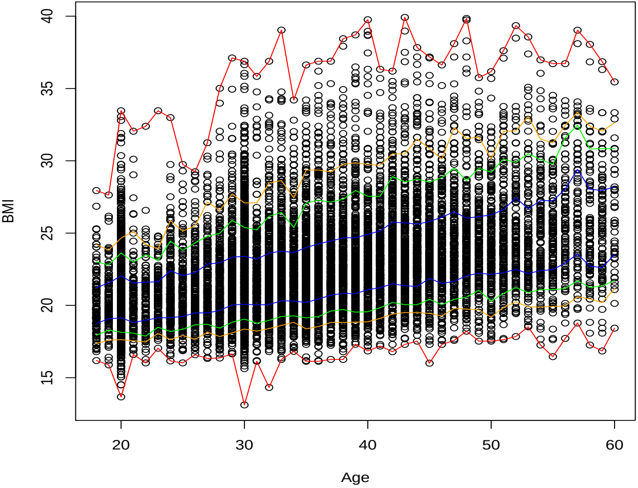


Biometric (AE)



- Easy extension of linear to quadratic model:
  - - loadings of the quadratic factor equal the respective squared loadings of the linear factor.
  - - the quadratic factor covaries with both initial level and linear factor.

How about extremes?



BMI upper tail - Osmann, Kaprio, Goegebeur and Hjelmborg

	N	η	χ	$\overline{\chi}$	$\overline{\nu}(X,Y)$	$\overline{h}_*^2$
MZ Overall	20048	0.80	0.49	0.59	0.23	0.24
Men	8212	0.78	0.45	0.56	0.22	0.13
Men (Age 18-29)	2241	0.83	0.44	0.67	0.29	0.57
Men (Age 30-39)	3356	0.87	0.41	0.74	0.13	0.21
Men (Age 40-60)	2615	0.70	0.39	0.39	0.22	0.14
DZ Overall	37476	0.61	0.26	0.21	0.11	
Men	16354	0.60	0.22	0.20	0.16	
Men (Age 18-29)	4356	0.60	0.22	0.20	0.00	
Men (Age 30-39)	6657	0.55	0.16	0.10	0.03	
Men (Age 40-60)	5341	0.65	0.14	0.30	0.15	

Within pair similarity - Questions

- Within pair similarity is measured by correlations.
- Correlations are further modeled by genetic and environmental variance components via the polygenic ADCE model.
- For instance, the polygenic ACE model relates to correlations via $\rho_{mz} = h^2 + c^2$ and $\rho_{dz} = \frac{1}{2}h^2 + c^2$.
- In table below, give examples corresponding to each row.

Heuristics of MZ and DZ correlations

	Interpretation	
Relation	Genetics	Environment
$\rho_{mz} > 4\rho_{dz}$	Epistasis	
$\rho_{mz} > 2\rho_{dz}$	Genetic dominance D	
$\rho_{mz} = 2\rho_{dz}$	Additive effect A (mono- or polygenic) and small D	Small C
$2\rho_{dz} > \rho_{mz} > \rho_{dz}$	Additive genes A	Shared environment C
$\rho_{mz} = \rho_{dz} > 0$	No genetic effect	C
$\rho_{mz} = \rho_{dz} = 0$	No genetic effect	No familial aggregation

Biometric analyses - polygenic model

In multivariate ACE model below for two traits X, Y, identify

- -genetic correlation in traits X, Y, that is, $\mathrm{corr}(A_X, A_Y)$.

- Contributing factors to the variation in outcome:

$$\Sigma_Y = \begin{pmatrix} \Sigma_A & r\Sigma_A \\ r\Sigma_A & \Sigma_A \end{pmatrix} + \begin{pmatrix} \Sigma_C & \Sigma_C \\ \Sigma_C & \Sigma_C \end{pmatrix} + \begin{pmatrix} \Sigma_E & 0 \\ 0 & \Sigma_E \end{pmatrix}$$

where $r = 1$ for MZ pairs and $z = \frac{1}{2}$ for DZ pairs.

Questions

- Obtain the BMI variance- covariance matrix for
    - MZ pairs.
    - DZ pairs.
- Identify regions for individual variances for twin and co-twin.
- Transform into corresponding within pair correlation matrices for MZ and DZ pairs
- What kind and size of genetic influence do you expect based on correlation matrices?

Variance covariance matrix of seven waves

```
x<-round(cov(mzData,use="complete"),2)
```

$$\Sigma_{\mathrm{MZ}} = \begin{bmatrix} 4.76 & 3.37 & 3.21 & 3.18 & 3.59 & 3.49 & 3.47 \\ 3.37 & 5.59 & 4.72 & 5.1 & 5.39 & 5.33 & 5.29 \\ 3.21 & 4.72 & 6.9 & 6.37 & 6.82 & 6.82 & 6.75 \\ 3.18 & 5.1 & 6.37 & 8.18 & 8.29 & 8.31 & 8.34 \\ 3.59 & 5.39 & 6.82 & 8.29 & 9.96 & 9.48 & 9.46 \\ 3.49 & 5.33 & 6.82 & 8.31 & 9.48 & 10.93 & 10.66 \\ 3.47 & 5.29 & 6.75 & 8.34 & 9.46 & 10.66 & 10.97 \end{bmatrix}$$

Variance covariance matrix of seven waves

```r
x<-round(cov(dzData,use="complete"),2)
```

$$\Sigma_{\mathrm{DZ}} = \begin{bmatrix} 4.69 & 3.61 & 3.6 & 3.55 & 3.49 & 3.2 & 3.13 \\ 3.61 & 5.95 & 5.18 & 5.34 & 5.58 & 5.24 & 5.18 \\ 3.6 & 5.18 & 7.78 & 7.06 & 7.26 & 6.98 & 6.88 \\ 3.55 & 5.34 & 7.06 & 8.31 & 7.99 & 7.92 & 7.81 \\ 3.49 & 5.58 & 7.26 & 7.99 & 9.34 & 8.73 & 8.55 \\ 3.2 & 5.24 & 6.98 & 7.92 & 8.73 & 10.3 & 9.78 \\ 3.13 & 5.18 & 6.88 & 7.81 & 8.55 & 9.78 & 10.11 \end{bmatrix}$$

Correlation matrix of seven waves

```
x<-round(cov2cor(cov(mzData,use="complete")),2)
```

$$\mathrm{Corr}_{\mathrm{MZ}} = \begin{bmatrix} 1 & 0.65 & 0.56 & 0.51 & 0.52 & 0.48 & 0.48 \\ 0.65 & 1 & 0.76 & 0.75 & 0.72 & 0.68 & 0.67 \\ 0.56 & 0.76 & 1 & 0.85 & 0.82 & 0.79 & 0.77 \\ 0.51 & 0.75 & 0.85 & 1 & 0.92 & 0.88 & 0.88 \\ 0.52 & 0.72 & 0.82 & 0.92 & 1 & 0.91 & 0.9 \\ 0.48 & 0.68 & 0.79 & 0.88 & 0.91 & 1 & 0.97 \\ 0.48 & 0.67 & 0.77 & 0.88 & 0.9 & 0.97 & 1 \end{bmatrix}$$

Correlation matrix of seven waves

```
x<-round(cov2cor(cov(dzData,use="complete")),2)
```

$$
\mathrm{Corr}_{\mathrm{DZ}} = \begin{bmatrix}
1 & 0.68 & 0.6 & 0.57 & 0.53 & 0.46 & 0.45 \\
0.68 & 1 & 0.76 & 0.76 & 0.75 & 0.67 & 0.67 \\
0.6 & 0.76 & 1 & 0.88 & 0.85 & 0.78 & 0.78 \\
0.57 & 0.76 & 0.88 & 1 & 0.91 & 0.86 & 0.85 \\
0.53 & 0.75 & 0.85 & 0.91 & 1 & 0.89 & 0.88 \\
0.46 & 0.67 & 0.78 & 0.86 & 0.89 & 1 & 0.96 \\
0.45 & 0.67 & 0.78 & 0.85 & 0.88 & 0.96 & 1
\end{bmatrix}
$$

Variance covariance matrix of seven waves for pairs

```
x<-round(cov(mzData,use="complete"),1)
```

$$\begin{bmatrix}
4.8 & 3.4 & 3.2 & 3.2 & 3.6 & 3.5 & 3.5 & 3.1 & 2.6 & 2.3 & 2.4 & 2.6 & 2.5 & 2.4 \\
3.4 & 5.6 & 4.7 & 5.1 & 5.4 & 5.3 & 5.3 & 2.8 & 3.7 & 3.7 & 3.8 & 3.9 & 3.8 & 3.7 \\
3.2 & 4.7 & 6.9 & 6.4 & 6.8 & 6.8 & 6.7 & 2.7 & 3.9 & 4.8 & 4.7 & 4.9 & 4.9 & 4.8 \\
3.2 & 5.1 & 6.4 & 8.2 & 8.3 & 8.3 & 8.3 & 3 & 4.7 & 5.2 & 5.5 & 5.7 & 5.7 & 5.5 \\
3.6 & 5.4 & 6.8 & 8.3 & 10 & 9.5 & 9.5 & 3 & 4.5 & 5.3 & 5.5 & 5.7 & 5.8 & 5.7 \\
3.5 & 5.3 & 6.8 & 8.3 & 9.5 & 10.9 & 10.7 & 3.1 & 4.7 & 5.6 & 5.7 & 6.1 & 6.4 & 6.2 \\
3.5 & 5.3 & 6.7 & 8.3 & 9.5 & 10.7 & 11 & 3.1 & 4.8 & 5.6 & 5.7 & 6 & 6.4 & 6.2 \\
3.1 & 2.8 & 2.7 & 3 & 3 & 3.1 & 3.1 & 5.6 & 4.2 & 3.7 & 3.5 & 3.8 & 3.2 & 3.1 \\
2.6 & 3.7 & 3.9 & 4.7 & 4.5 & 4.7 & 4.8 & 4.2 & 6.6 & 5.5 & 5.6 & 5.9 & 5.4 & 5 \\
2.3 & 3.7 & 4.8 & 5.2 & 5.3 & 5.6 & 5.6 & 3.7 & 5.5 & 7.6 & 6.3 & 6.7 & 6.7 & 6.3 \\
2.4 & 3.8 & 4.7 & 5.5 & 5.5 & 5.7 & 5.7 & 3.5 & 5.6 & 6.3 & 7.5 & 7.4 & 7 & 6.8 \\
2.6 & 3.9 & 4.9 & 5.7 & 5.7 & 6.1 & 6 & 3.8 & 5.9 & 6.7 & 7.4 & 8.9 & 8.2 & 7.6 \\
2.5 & 3.8 & 4.9 & 5.7 & 5.8 & 6.4 & 6.4 & 3.2 & 5.4 & 6.7 & 7 & 8.2 & 9.4 & 8.7 \\
2.4 & 3.7 & 4.8 & 5.5 & 5.7 & 6.2 & 6.2 & 3.1 & 5 & 6.3 & 6.8 & 7.6 & 8.7 & 8.7
\end{bmatrix}$$

Variance covariance matrix of seven waves for pairs

```
x<-round(cov(dzData,use="complete"),1)
```

$$\left[\begin{array}{cccccccccccccc}
4.7 & 3.6 & 3.6 & 3.6 & 3.5 & 3.2 & 3.1 & 1.5 & 1.5 & 1.5 & 1.4 & 1.5 & 1.6 & 1.7 \\
3.6 & 5.9 & 5.2 & 5.3 & 5.6 & 5.2 & 5.2 & 1.7 & 2.5 & 2.4 & 2.2 & 2.4 & 2.3 & 2.5 \\
3.6 & 5.2 & 7.8 & 7.1 & 7.3 & 7 & 6.9 & 1.8 & 2.8 & 3.3 & 3 & 3.1 & 2.9 & 3.1 \\
3.6 & 5.3 & 7.1 & 8.3 & 8 & 7.9 & 7.8 & 1.5 & 2.5 & 3.1 & 3 & 3.1 & 2.9 & 3.1 \\
3.5 & 5.6 & 7.3 & 8 & 9.3 & 8.7 & 8.6 & 1.4 & 2.7 & 3.1 & 2.9 & 3.2 & 3 & 3.1 \\
3.2 & 5.2 & 7 & 7.9 & 8.7 & 10.3 & 9.8 & 1.2 & 2.5 & 3.2 & 3 & 3.1 & 3 & 3.3 \\
3.1 & 5.2 & 6.9 & 7.8 & 8.6 & 9.8 & 10.1 & 1.2 & 2.5 & 3.2 & 3 & 3.1 & 3 & 3.3 \\
1.5 & 1.7 & 1.8 & 1.5 & 1.4 & 1.2 & 1.2 & 4.8 & 4 & 3.5 & 3.3 & 3.5 & 3.4 & 3.4 \\
1.5 & 2.5 & 2.8 & 2.5 & 2.7 & 2.5 & 2.5 & 4 & 6.8 & 5 & 5.3 & 5.5 & 5 & 4.9 \\
1.5 & 2.4 & 3.3 & 3.1 & 3.1 & 3.2 & 3.2 & 3.5 & 5 & 7.3 & 6.5 & 6.9 & 6.6 & 6.4 \\
1.4 & 2.2 & 3 & 3 & 2.9 & 3 & 3 & 3.3 & 5.3 & 6.5 & 7.7 & 7.6 & 7.3 & 7.1 \\
1.5 & 2.4 & 3.1 & 3.1 & 3.2 & 3.1 & 3.1 & 3.5 & 5.5 & 6.9 & 7.6 & 9 & 8.2 & 7.9 \\
1.6 & 2.3 & 2.9 & 2.9 & 3 & 3 & 3 & 3.4 & 5 & 6.6 & 7.3 & 8.2 & 9.6 & 9 \\
1.7 & 2.5 & 3.1 & 3.1 & 3.1 & 3.3 & 3.3 & 3.4 & 4.9 & 6.4 & 7.1 & 7.9 & 9 & 9.3
\end{array}\right.$$

Multivariate models fitted to seven waves

- Consider assumptions by fitting the saturated model and submodels thereof.
- Common environmental effects seems important.

```
mxCompare(CholAceFit,CholAeFit)  # C seems important.

##      base comparison  ep minus2LL    df      AIC   diffLL diffdf           p
## 1 CholACE       <NA> 119 91335.02 25641 40053.02       NA     NA          NA
## 2 CholACE     CholAE  91 91389.54 25669 40051.54 54.51227     28 0.001945057
```

  consider other genetic submodels of the ACE or ADE models.
- Similarly for the independent and common pathway models.
- Report means, heritabilities, genetic correlations and other quantities of relevance.
- Consider questions on above slide for the linear growth curve ACE model.