Phenomics

Michael C Neale
Virginia Commonwealth University
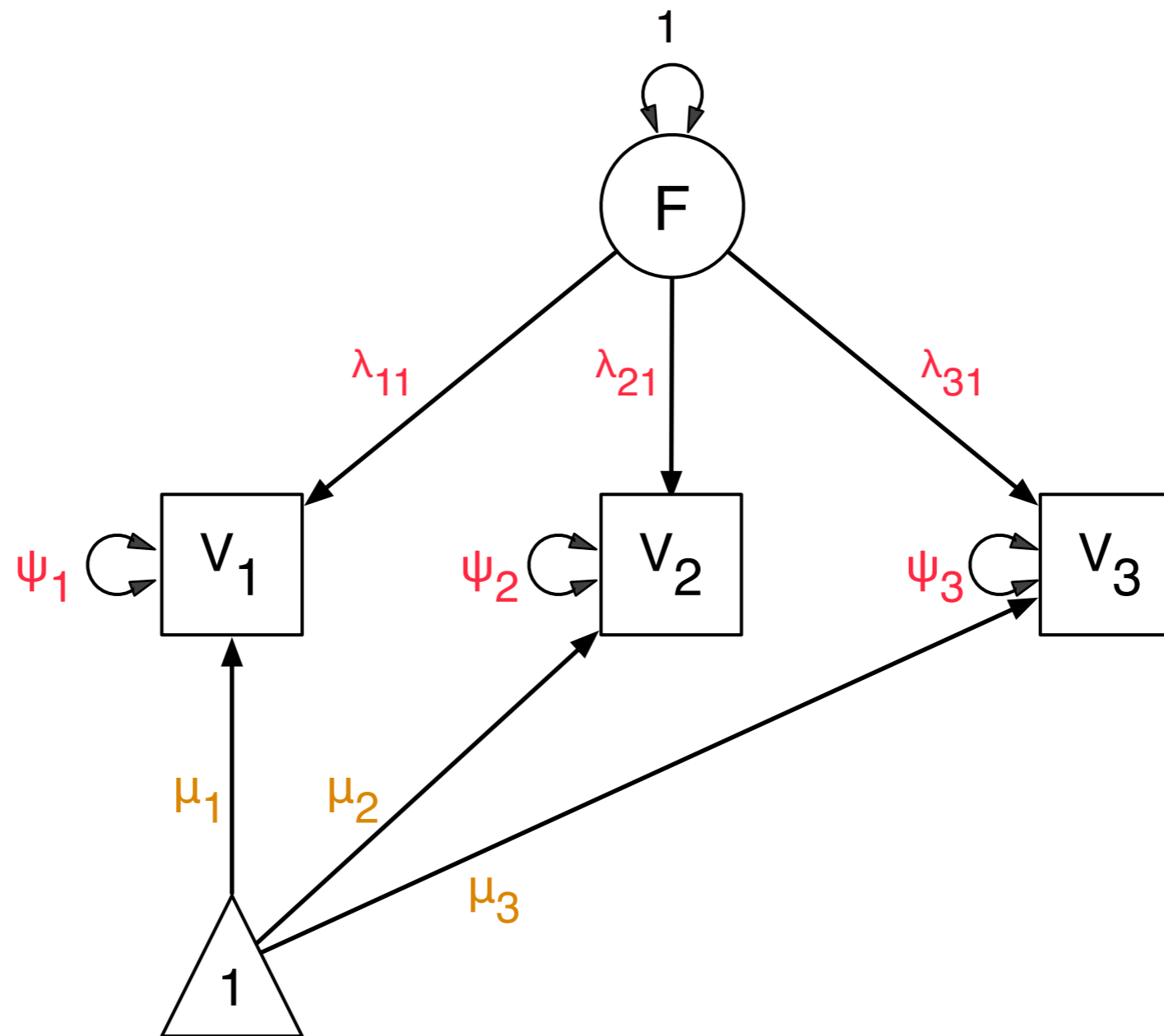
Boulder Workshop March 8 2013
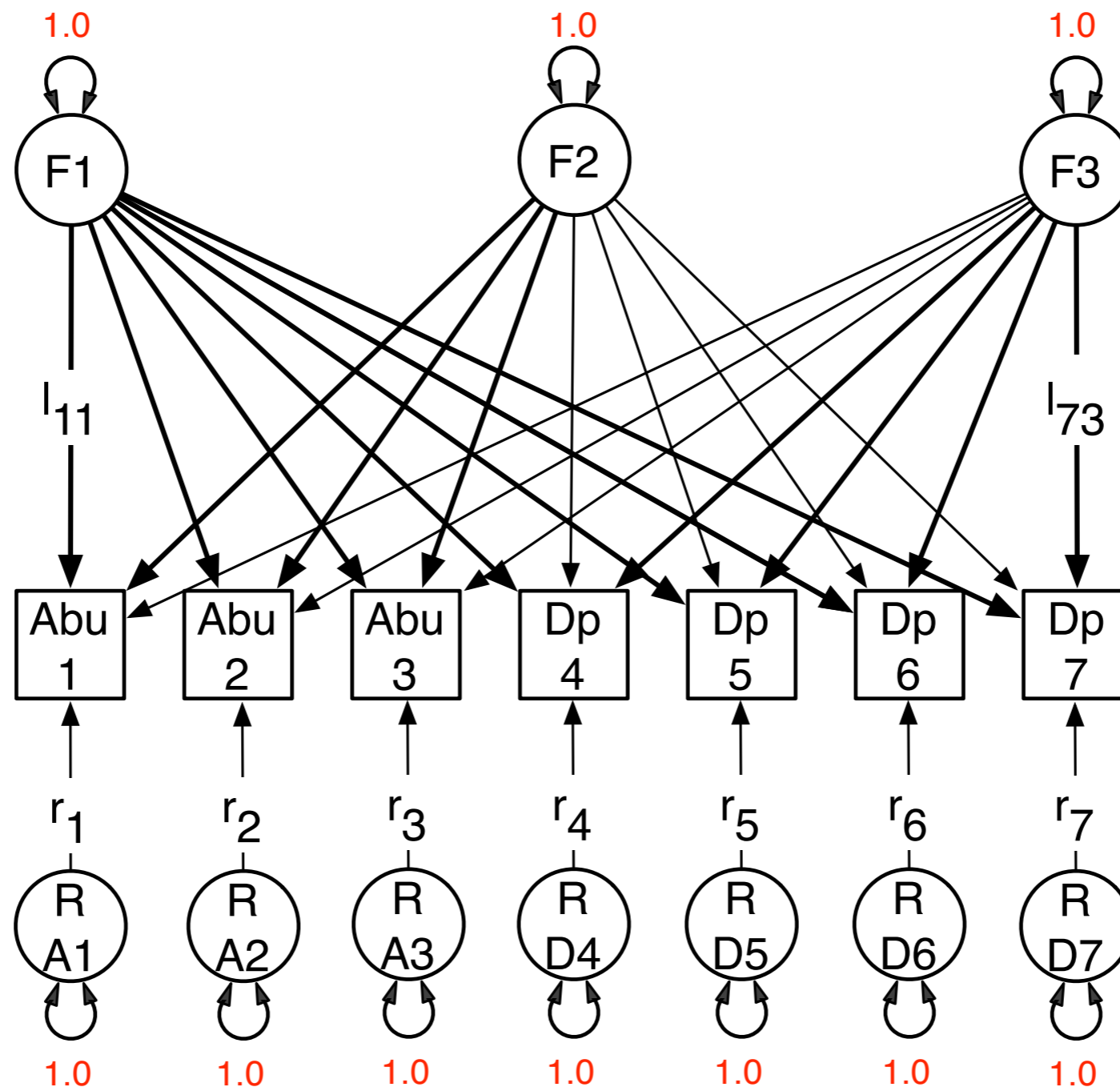
VIPBG

VCU

Measurement

- Latent Trait Models

- Measurement Invariance

- Factor Scores

- Mixture Distributions

- Latent Class Models

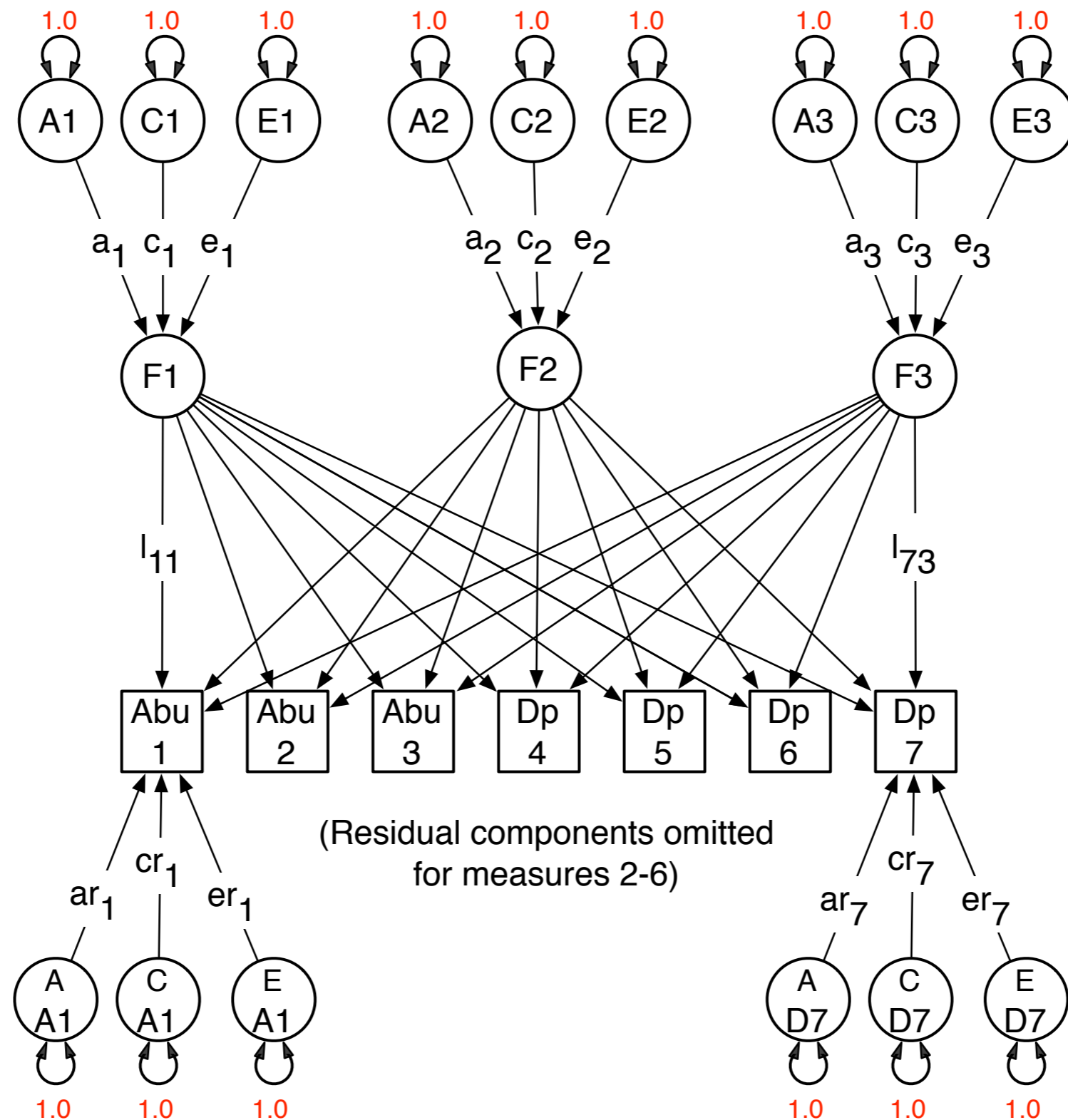- Heterogeneity & Age

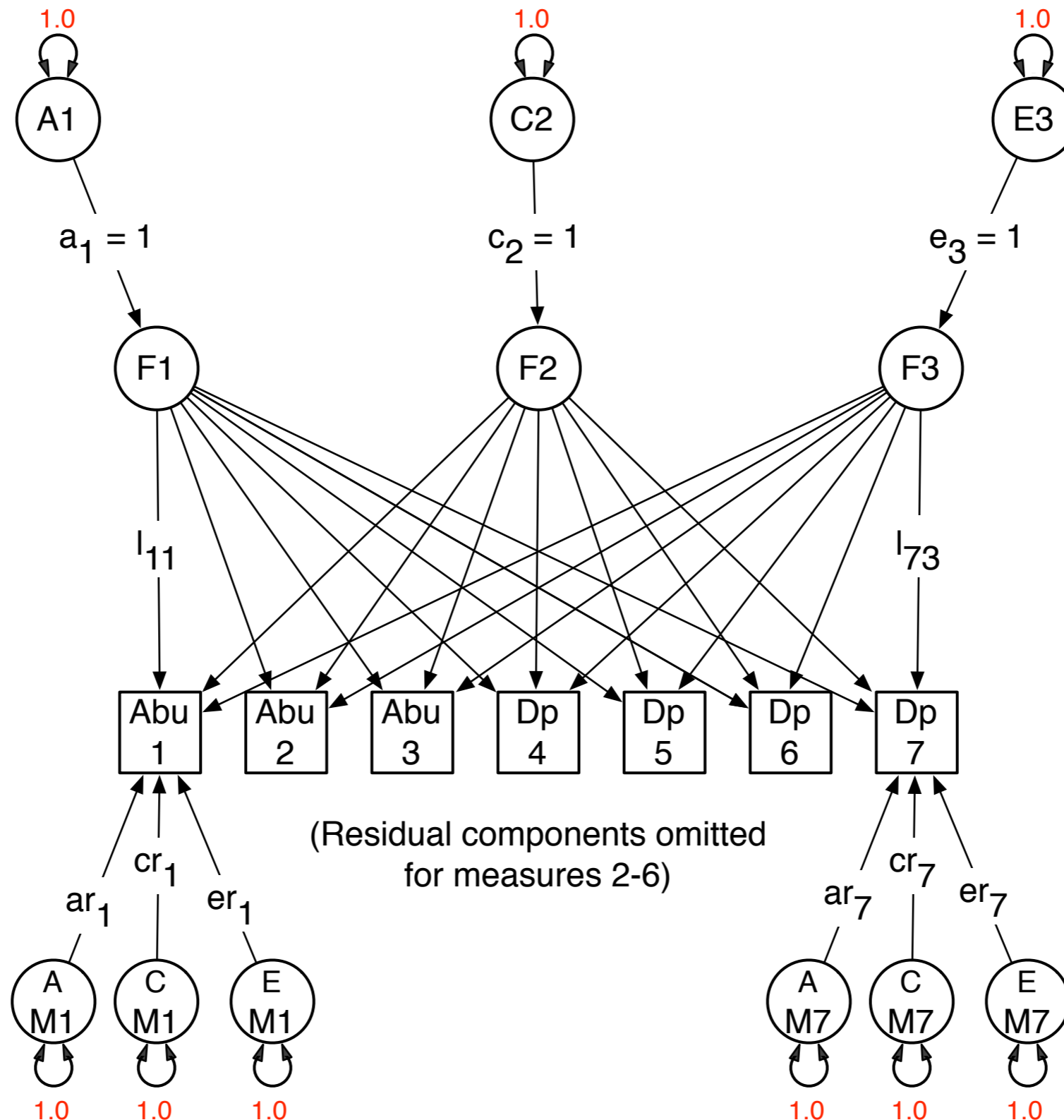# Simple Single Factor Model
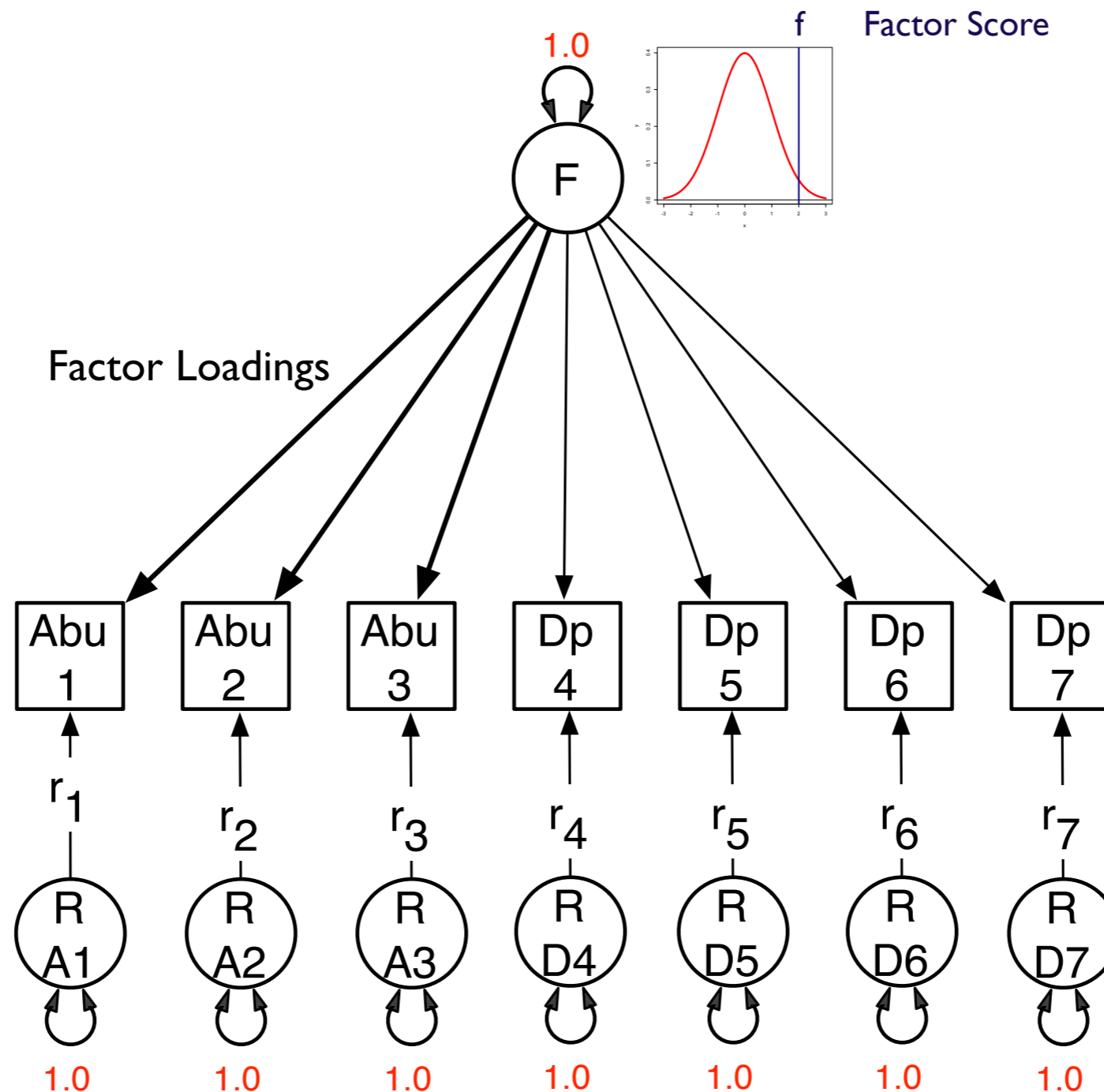
Multiple Factor Model
Beware Rotation

3 Genetic and Environmental Factors: Common Pathway Model
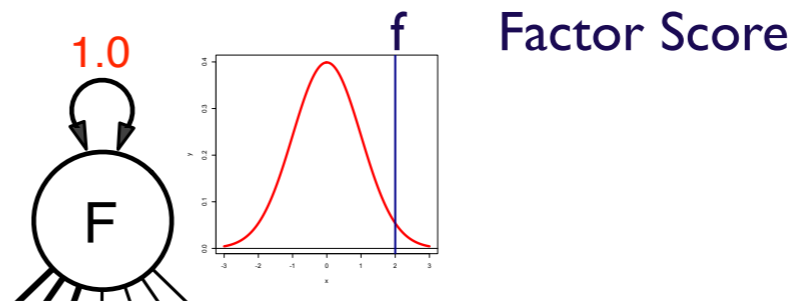

(Residual components omitted for measures 2-6)

3 Genetic and Environmental Factors: Independent Pathway Model
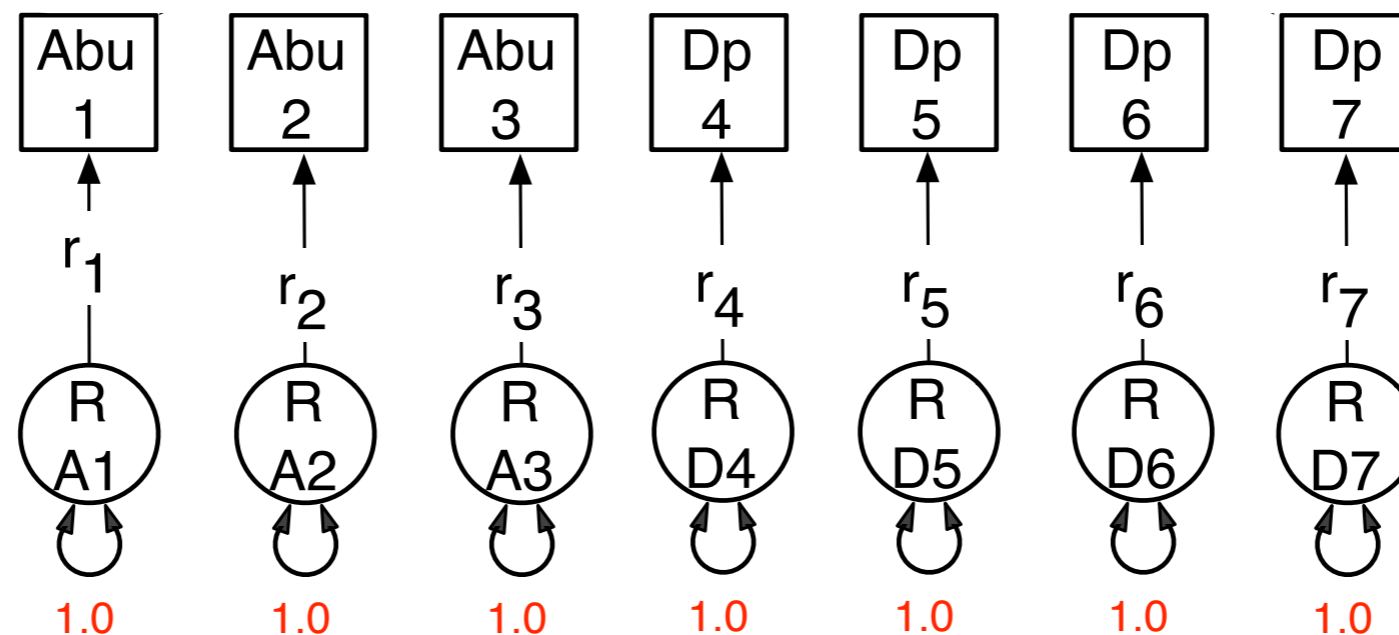
Estimating Factor Scores

ML Estimation of Factor Scores

1.0

f Factor Score

F

Factor Score * Likelihood of items conditional on factor score

Items independent conditional on factor score:
Means and variances change according to size of factor loadings

| Abu 1 | Abu 2 | Abu 3 | Dp 4 | Dp 5 | Dp 6 | Dp 7 |

r_1 r_2 r_3 r_4 r_5 r_6 r_7

R A1   R A2   R A3   R D4   R D5   R D6   R D7

1.0   1.0   1.0   1.0   1.0   1.0   1.0

Application

ORIGINAL ARTICLE

Association between glutamic acid decarboxylase genes and anxiety disorders, major depression, and neuroticism

JM Hettema, SS An, MC Neale, J Bukszar, EJCG van den Oord, KS Kendler and X Chen

Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA

- Used genetic factor scores to select extreme groups

- Found significant association

Notes

- Factor scores do not all have same error variance

- Factor scores of A, C & E components may correlate highly

- Latent trait may be non-normal (Schmitt et al 2006 *Multiv Behav Res*

- Factor loadings may vary across the distribution

- Variation may be discrete not continuous

Measurement Invariance

Equality constraints imposed across groups in steps towards strict factorial invariance

No.	Description	factor loadings	residual variances	intercepts	factor means
1	Configural invariance	free	free	free	fixed at 0
2	Metric/weak invariance	<u>invariant</u>	free	free	fixed at 0
3	Equal residual variances	invariant	<u>invariant</u>	free	fixed at 0
4	Strict factorial invariance	invariant	invariant	<u>invariant</u>	free[1]

Dolan, C. V., Oort, F. J., Stoel, R. D., and Wicherts, J. M. (2009). Testing Measurement Invariance in the Target Rotated Multigroup Exploratory Factor Model. Structural Equation Modeling, 16(2): 295–314.
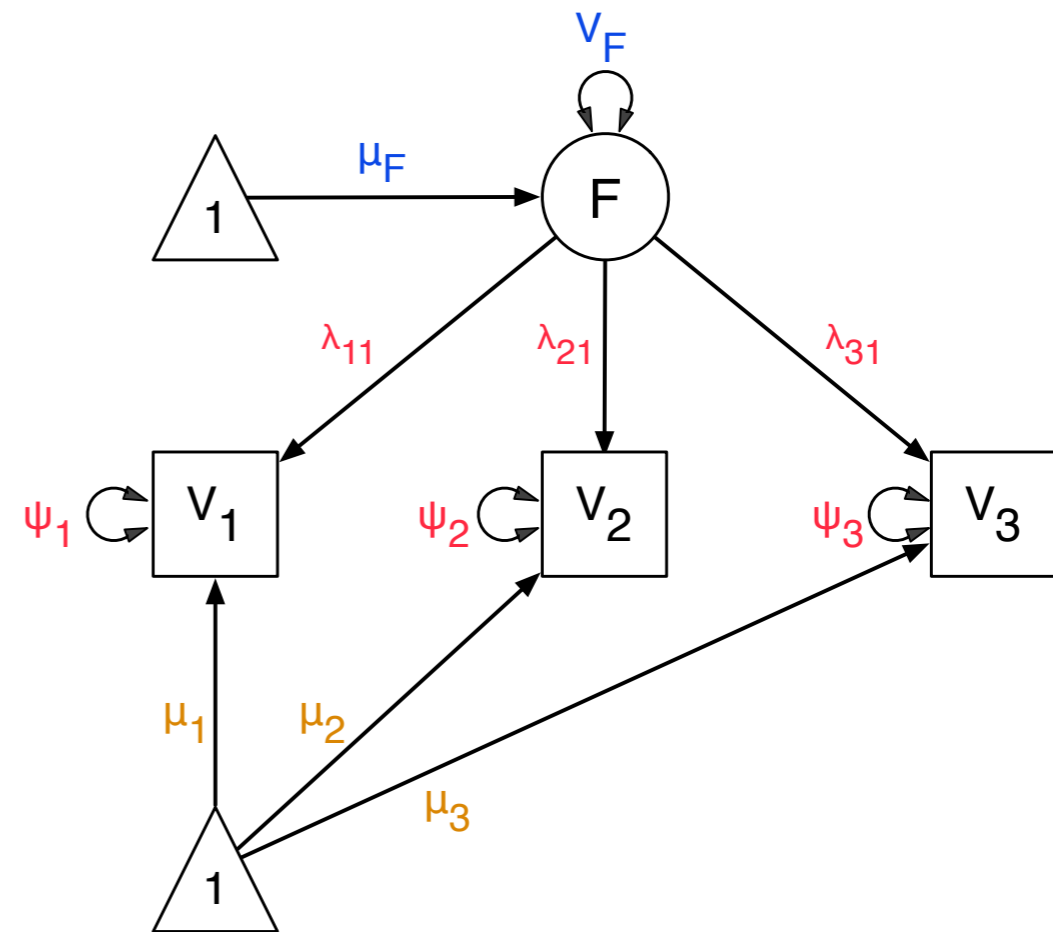
Wicherts J & Dolan CV (In Press) Educational Measurement: Issues and Practice
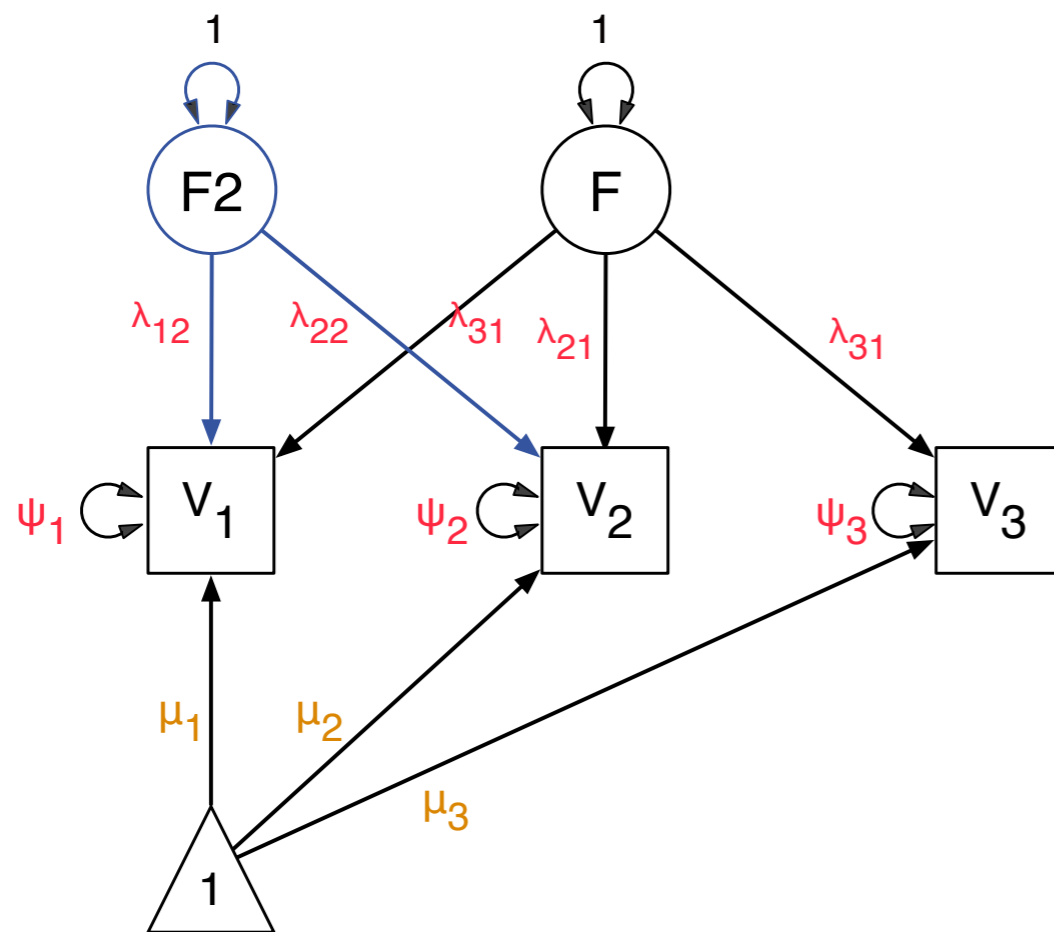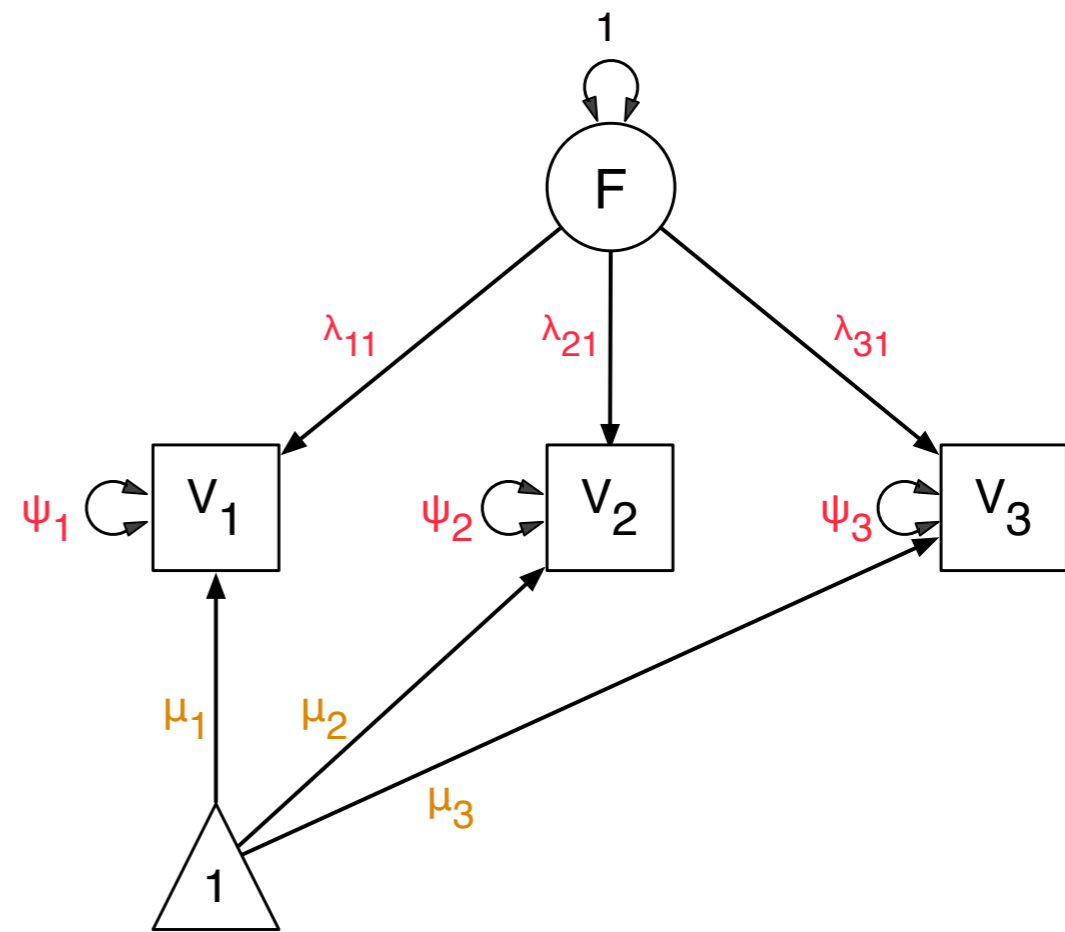
Strict Factorial Invariance

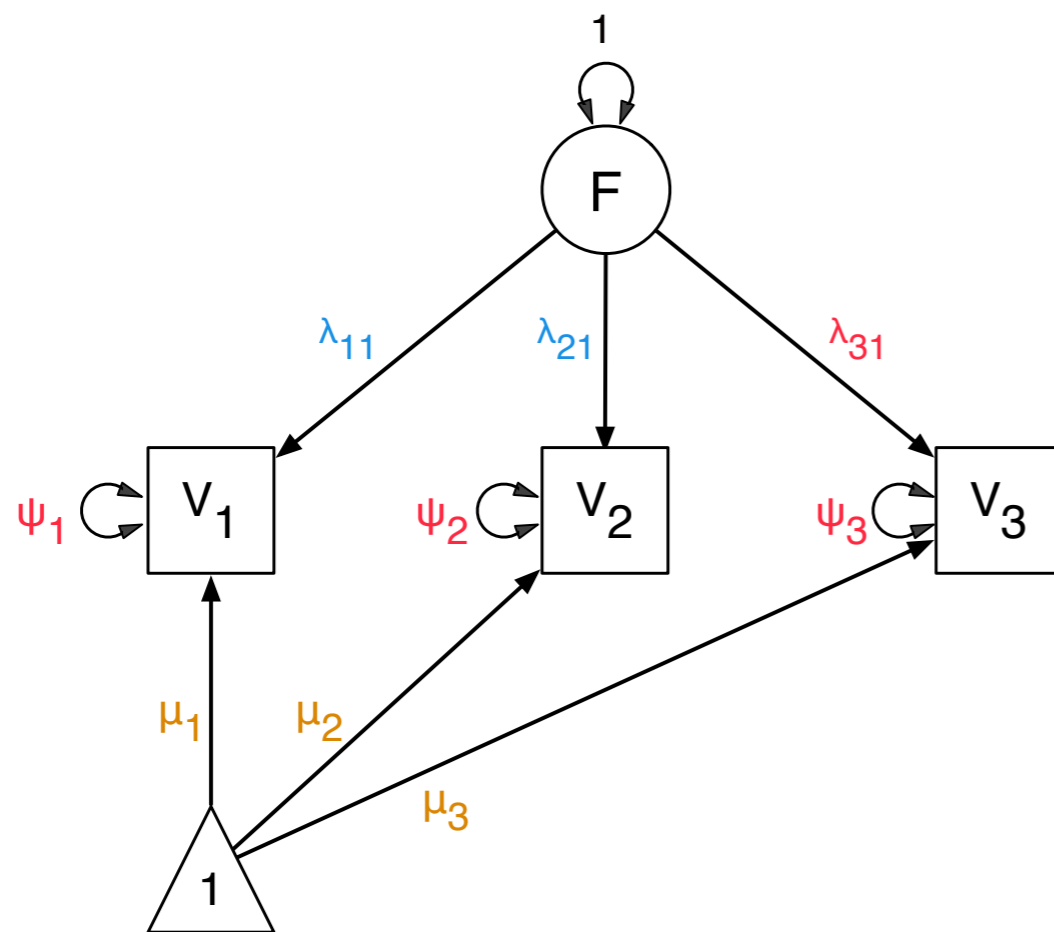
Males
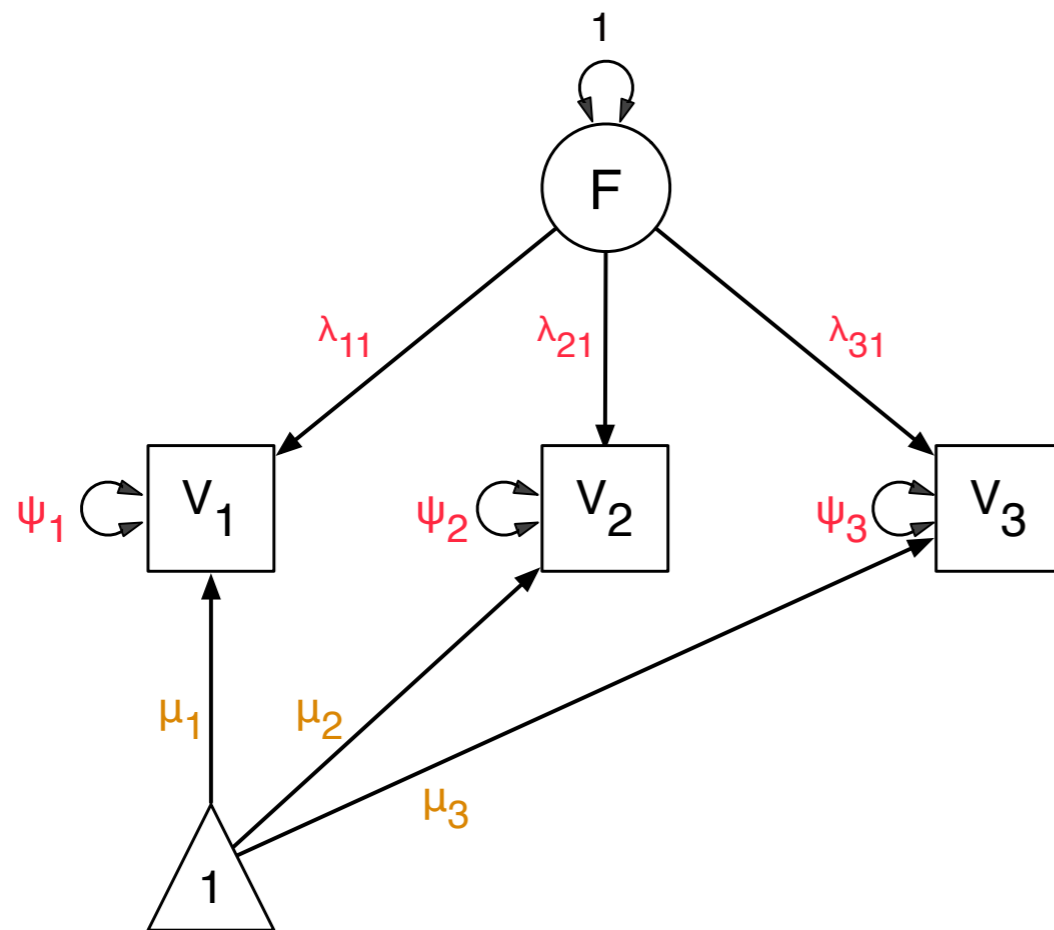
Females

Failure of Configural Invariance
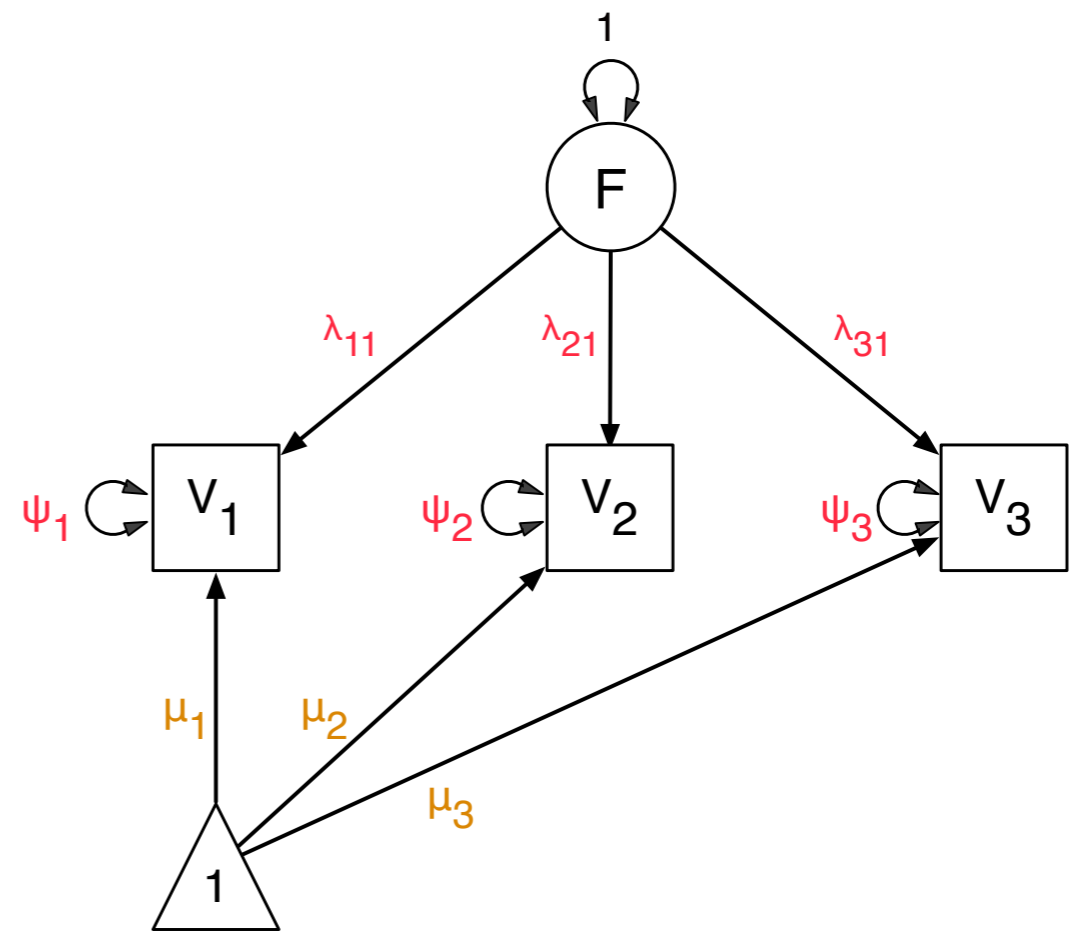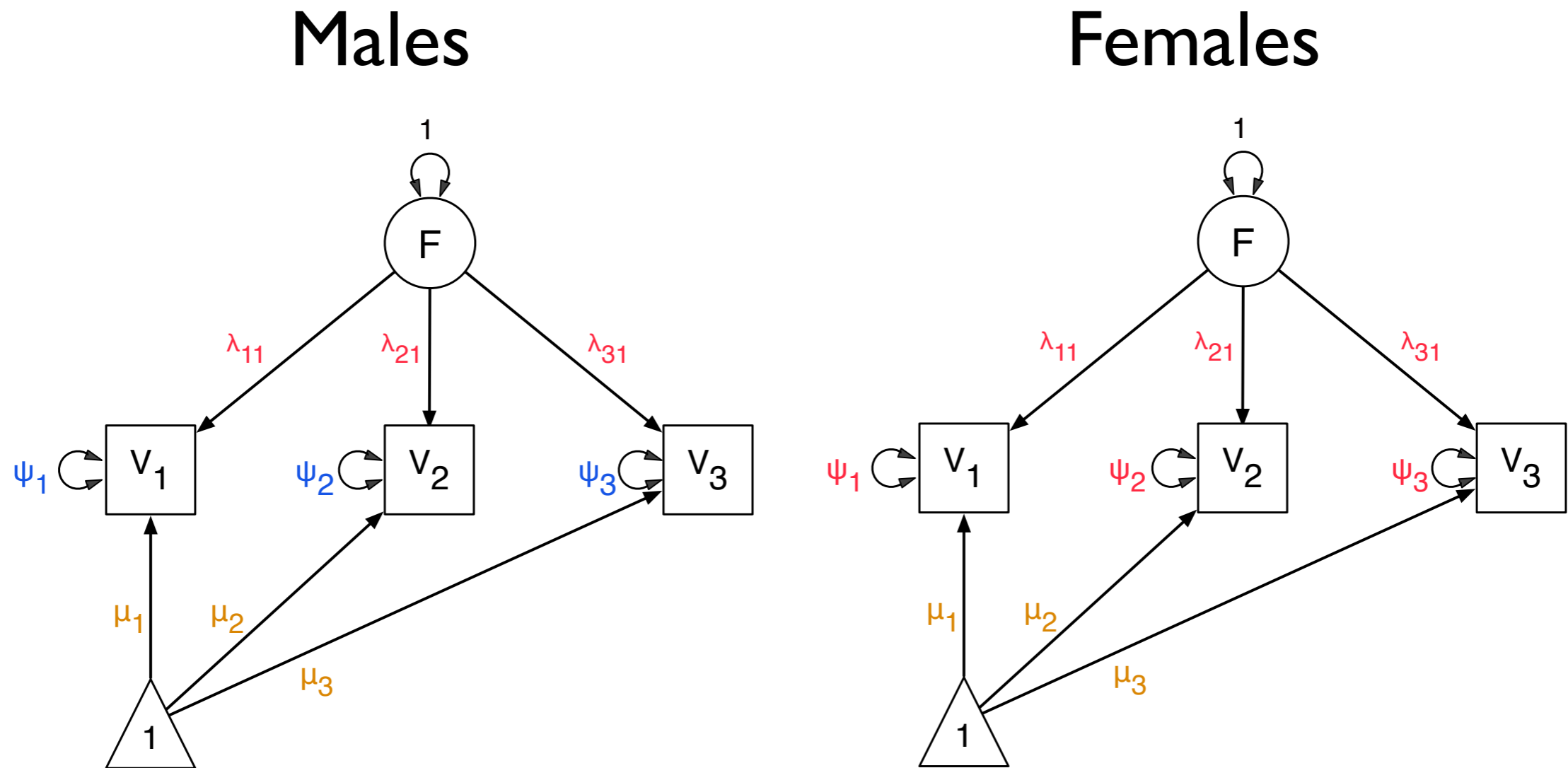
Failure of Metric Invariance

Failure of Residual Invariance

Medland, S. E. and Neale, M. C. (2010). An integrated phenomic approach to multivariate allelic association. Eur J Hum Genet, 18(2):233–9.

Correlations across Substances

	Stimulants	Tranquilizers	Marijuana
Stimulants	1		
Tranquilizers	0.74	1	
Marijuana	0.63	0.66	1
Factor Loadings	0.84	0.87	0.75

DRD2 Association Results

- **Univariate association**

  - Stimulants: $\chi^2=3.88$, $\beta= -.18$, $p < .05$

  - Tranquilizers: $\chi^2=1.65$, $\beta= .13$, NS

  - Marijuana: $\chi^2=2.60$, $\beta= .11$, NS

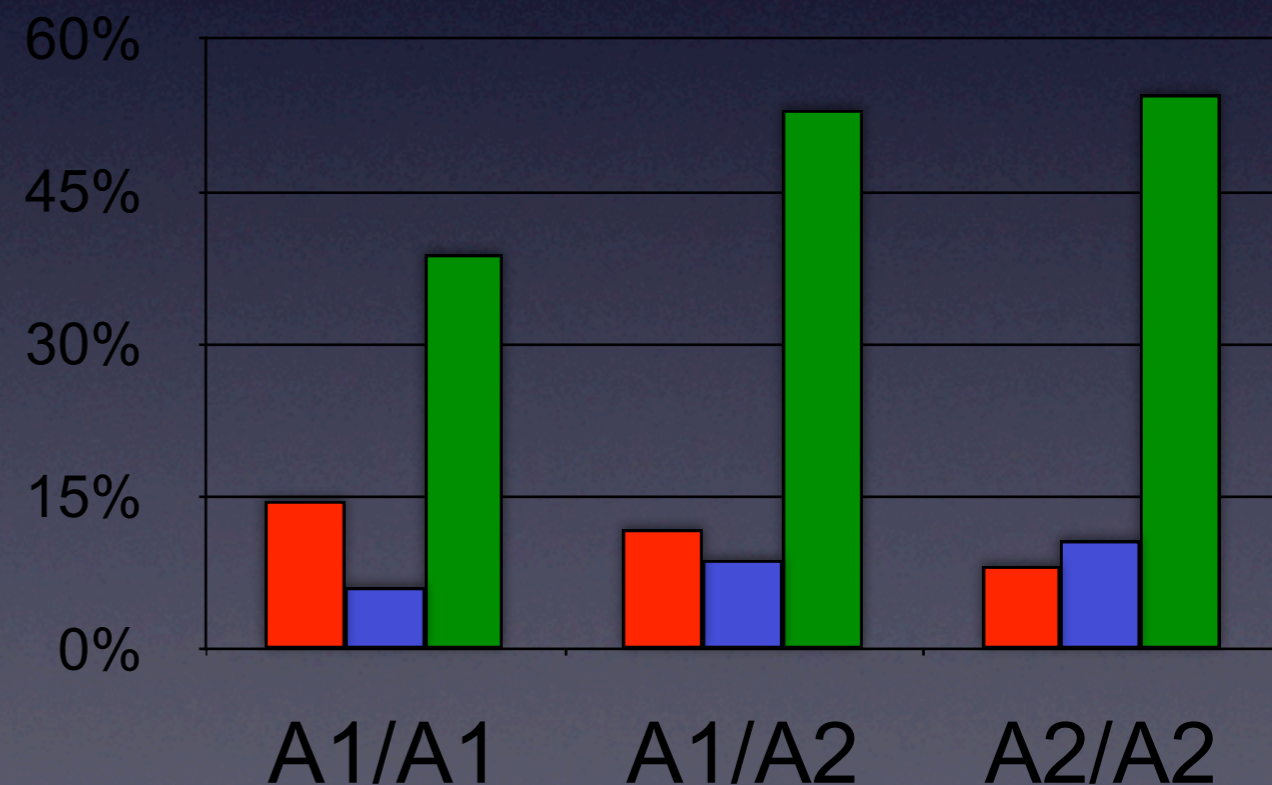- **Factor level association**

  - $\chi^2=0.65$, $kF= .06$, NS

- **Multivariate association**

  - $\chi^2=13.91$ (3df; $p < 0.005$)

 − $k_{Stimulants}$ = -0.19

  − $k_{Tranquilizers}$ = 0.14

  − $\beta_{Marijuana}$ = 0.11

Legend:
- Stimulants (red)
- Tranquilizers (blue)
- Marijuana (green)

Chart axis: 60%, 45%, 30%, 15%, 0% — categories A1/A1, A1/A2, A2/A2

Detecting 1% GV at 80% Power

Phenotypic Complexity, Measurement Bias, and Poor Phenotypic Resolution Contribute to the Missing Heritability Problem in Genetic Association Studies

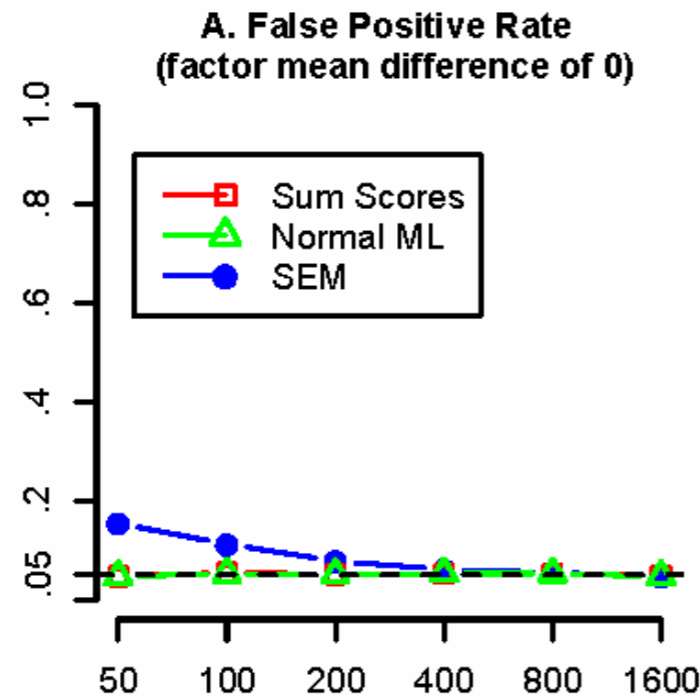Van der Sluis, Verhage, Posthuma & Dolan
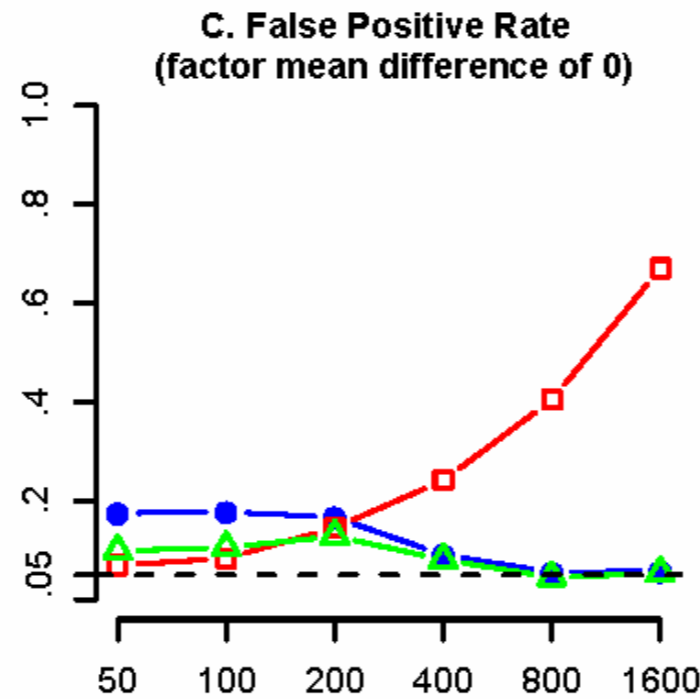Plos One 2010
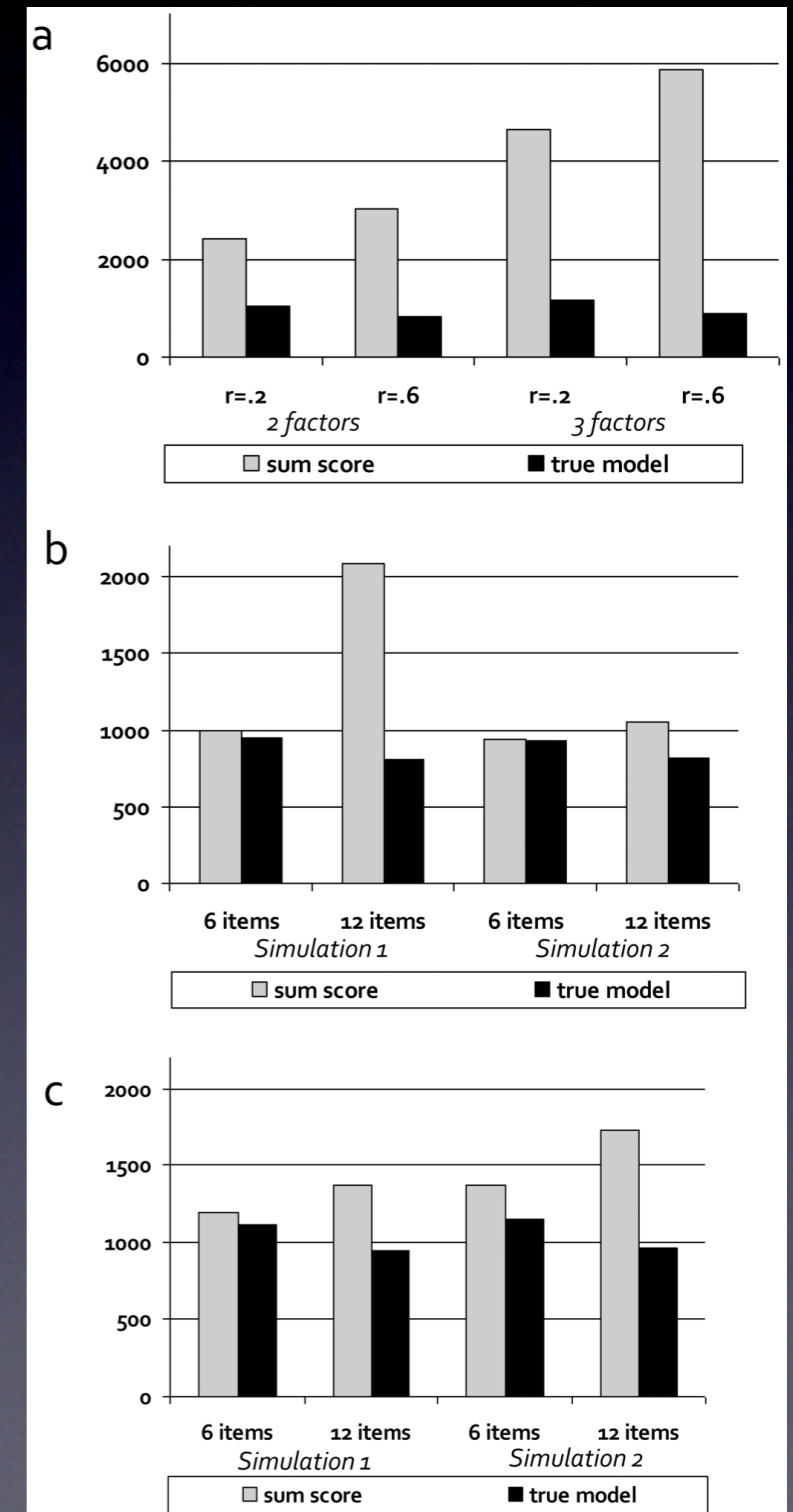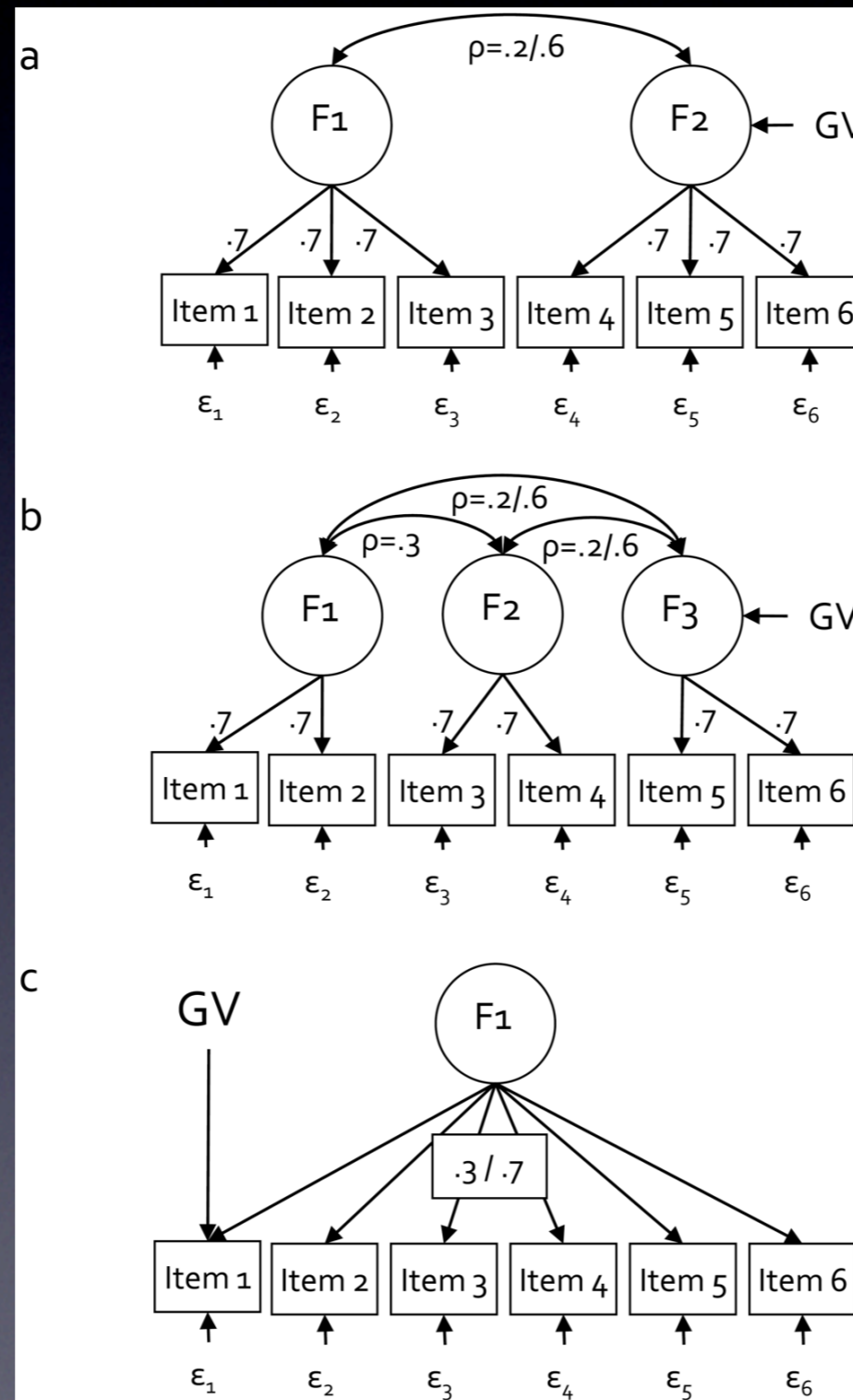
Three methods of scoring

- Sum score
  - Simple & Practical
  - Widely Used

- Maximum likelihood factor score
  - More complex (need computer)
  - Less widely used
  - Can test assumptions

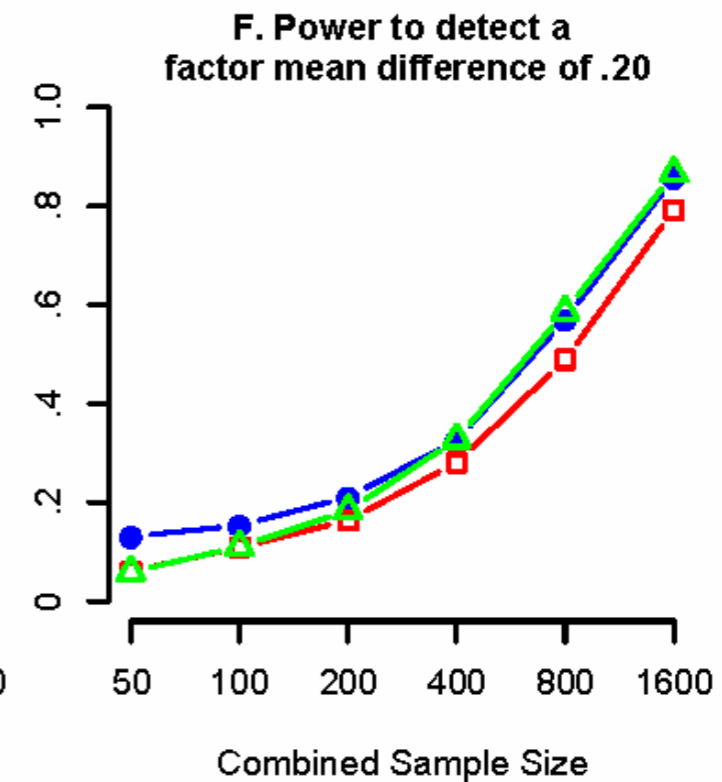- Neither - use SEM framework for testing
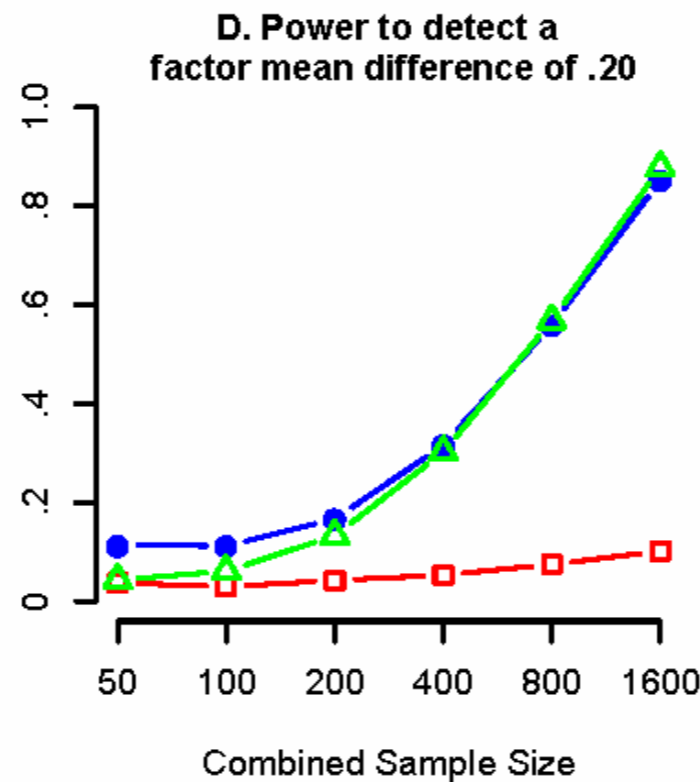
Non-Invariance Effects
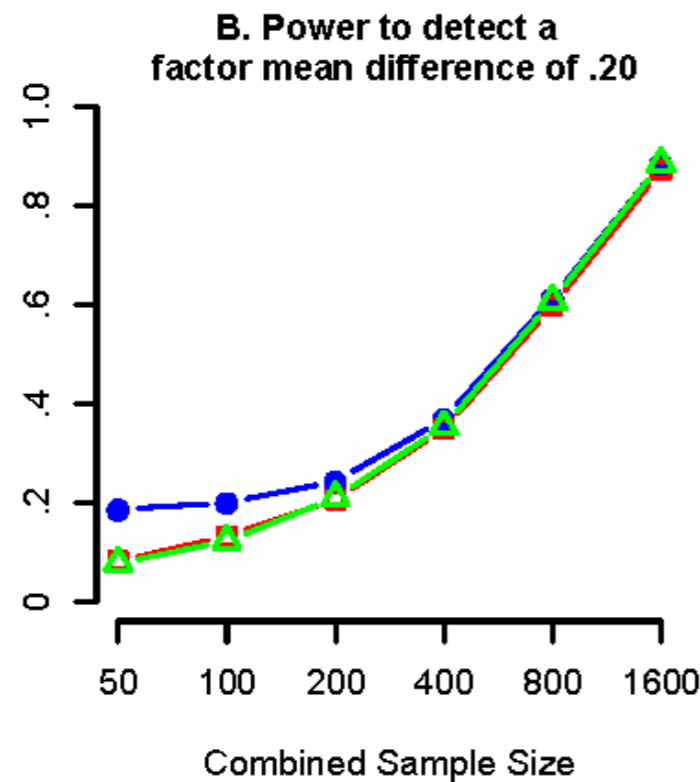

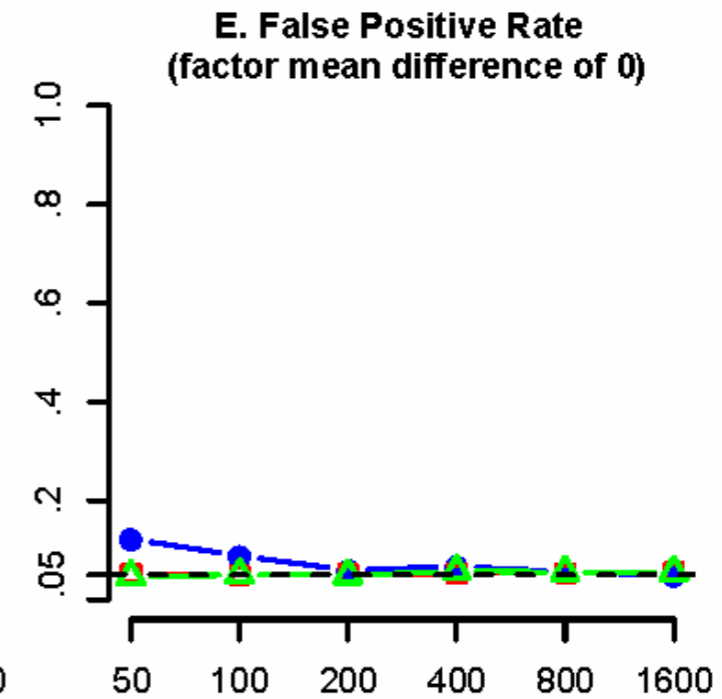
Measurement Invariance Upheld

Non−Invariance of a single Threshold

Non−Invariance of a single Factor Loading

A. False Positive Rate (factor mean difference of 0)

C. False Positive Rate (factor mean difference of 0)

E. False Positive Rate (factor mean difference of 0)

Sum Scores
Normal ML
SEM

B. Power to detect a factor mean difference of .20

D. Power to detect a factor mean difference of .20

F. Power to detect a factor mean difference of .20

Combined Sample Size

Combined Sample Size

Combined Sample Size

What if Variation is Discrete?

- Latent Class and Latent Profile Models

- Factor Mixture Models

- Latent Growth Curve Mixture Models

- Regime Switching

Mixture Distributions

Pearson, K. (1894). Contributions to the mathematical theory of evolution. II. skew variation in homogeneous material. Philosophical Transactions of the Royal Society of London A, 186, 343-414.

Finite Mixture Models

Geoffrey McLachlan

David Peel

WILEY SERIES IN
PROBABILITY AND STATISTICS
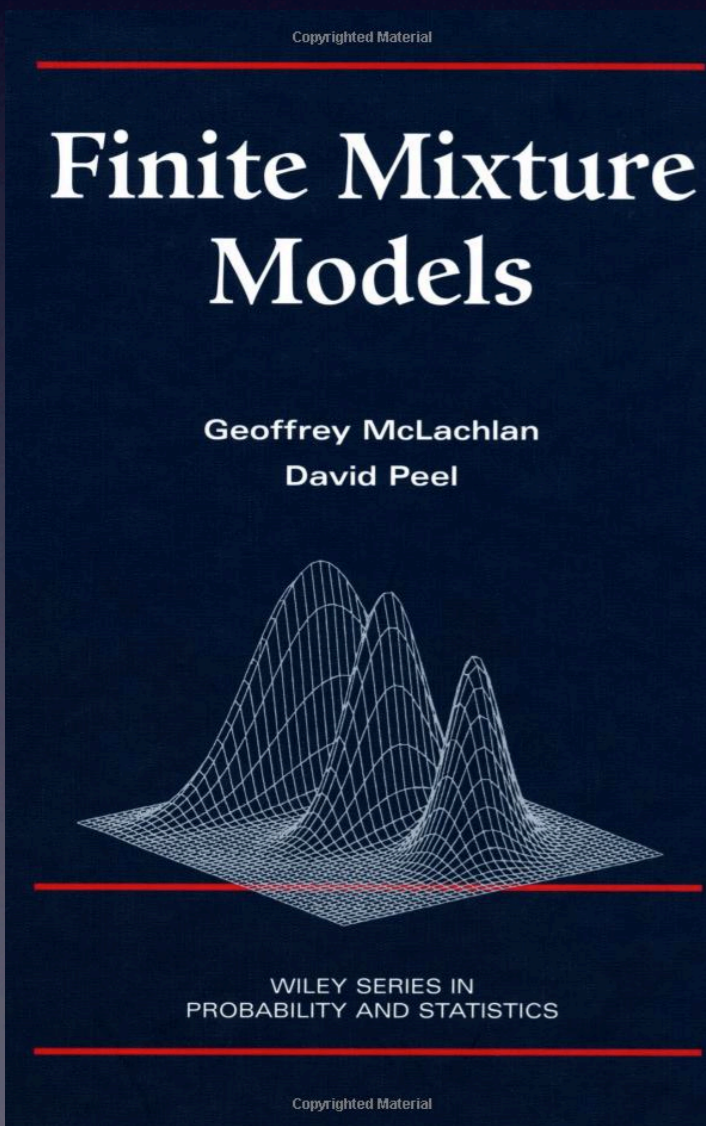
- Skewness in a set of measurements of the ratio of forehead to body length of crabs

- Two species or one?
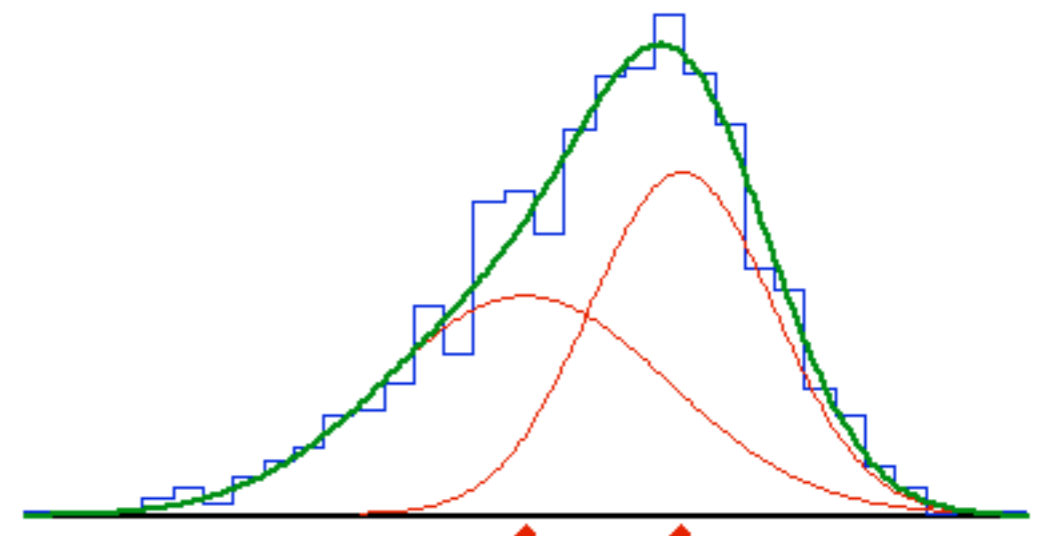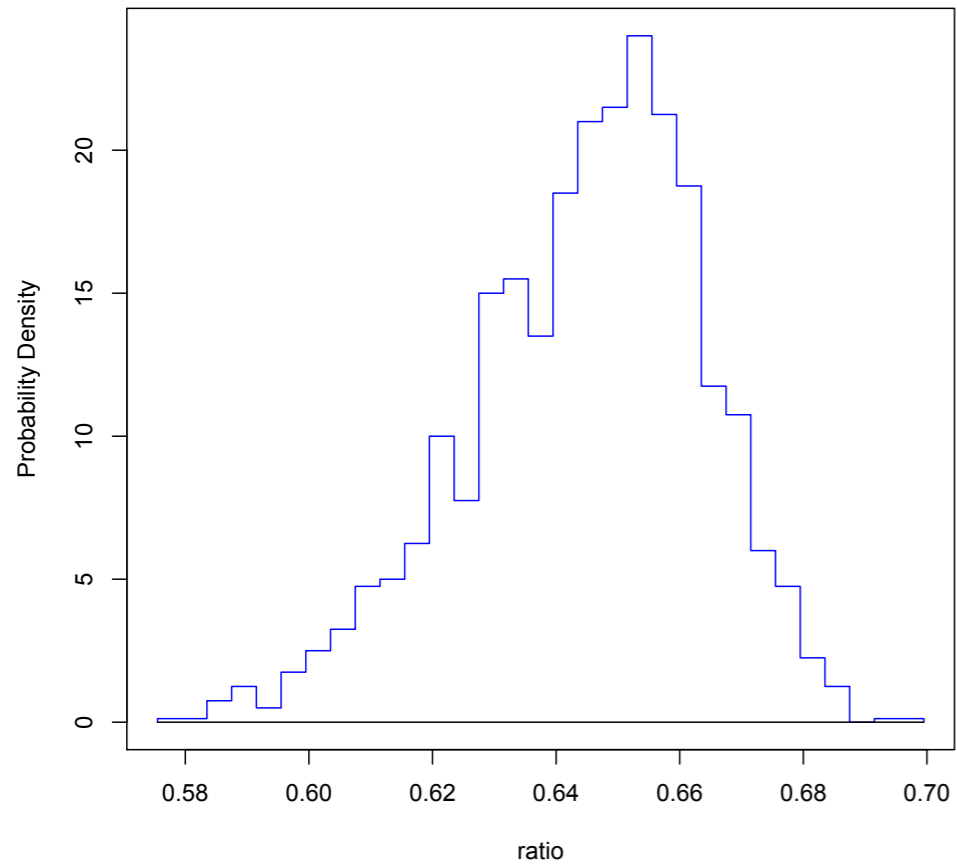
Data & Model

Fit Mixture Distribution

```r
#
# Analysis of Pearson crab ratio data
# Mixture distribution, 2 normal components
#
# Load libraries & data
require(OpenMx)
load("pearson.Rda")
data(pearson)
summary(pearson)
head(pearson)
pearson[pearson$ratio==Inf,1]<-NA

# Set number of variables & classes
nvar <- 1
nClass <-2
selVars<-c("ratio")

g1Model <-  mxModel("group1",
            mxMatrix("Symm", nvar, nvar, values=1, name="expCov", lbound=1.E-10, free=T),
            mxMatrix("Full", 1, nvar, free=T, name="expMean"),
            mxData(pearson, type="raw"),
            mxFIMLObjective("expCov", "expMean", dimnames="ratio", vector=TRUE)
            )
g2Model <-  mxModel(g1Model, name="group2",
            mxMatrix("Full", 1, nvar, values=.5, free=T, name="expMean")
            )
# can repeat above "duplication" step for more classes, in a loop if needed, and stick them in a list[]
```

Fit Mixture Distribution

```r
# Put the two distributions together in one model, add proportion parameter (standardize it to p)
# and set up the objective function as a weighted (freq) mixture distribution

mixtureModel <- mxModel("mixture", g1Model, g2Model,
                mxMatrix(type="Full", nrow=nClass, ncol=1, values=runif(nClass),
                    free=c(rep(T,nClass-1),F), lbound=1.E-1, name="praw"),
                mxAlgebra(praw %x% (1/sum(praw)), name="p"),
                mxMatrix(type="Full", nrow=dim(pearson)[1], ncol=1,
                    values=as.matrix(pearson$freq), name="freq"),
                mxAlgebra( -2*sum (freq * (log(cbind(group1.objective, group2.objective) %*% p ))),
                    name="min2LL"),
                mxAlgebraObjective("min2LL")
            )

summary(mixtureModelFit <- mxRun(mixtureModel, unsafe=T))


#
# Fix proportion parameter to 1 and fix mean & variance parameters of group 2
#

# Copy Model
nonMixtureModel <- mixtureModel

# (could use omxSetParameters(nonMixtureModel,parameters) if they had been labeled)
nonMixtureModel$praw@free[]<-F
nonMixtureModel$praw@values[]<-c(1,0)

nonMixtureModel$group2.expCov@free[]<-F
nonMixtureModel$group2.expMean@free[]<-F

summary(nonMixtureModelFit <- mxRun(nonMixtureModel))

mxCompare(mixtureModelFit,nonMixtureModelFit)
```
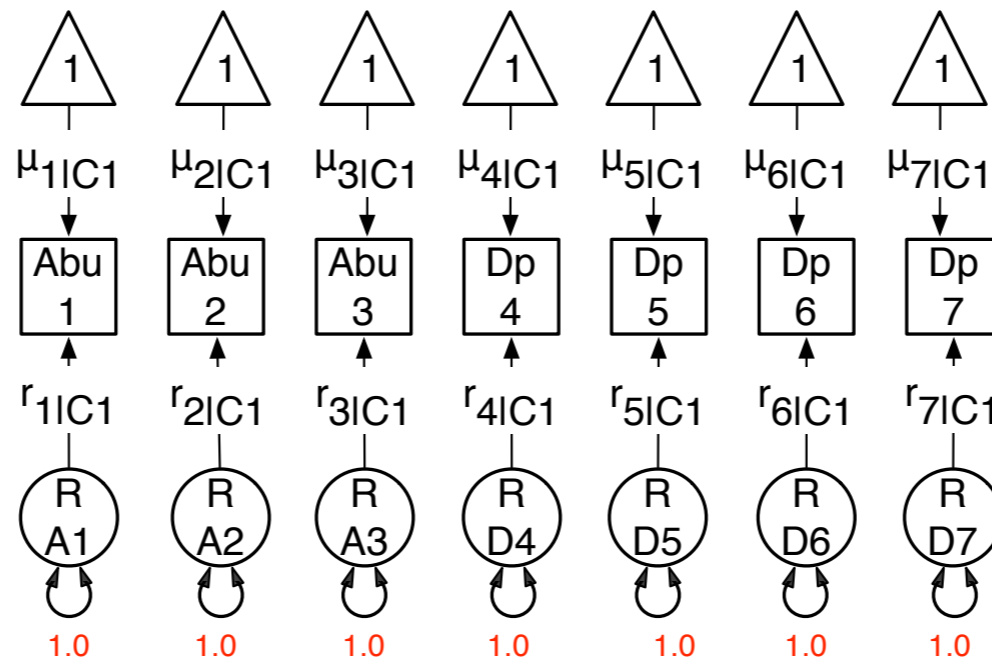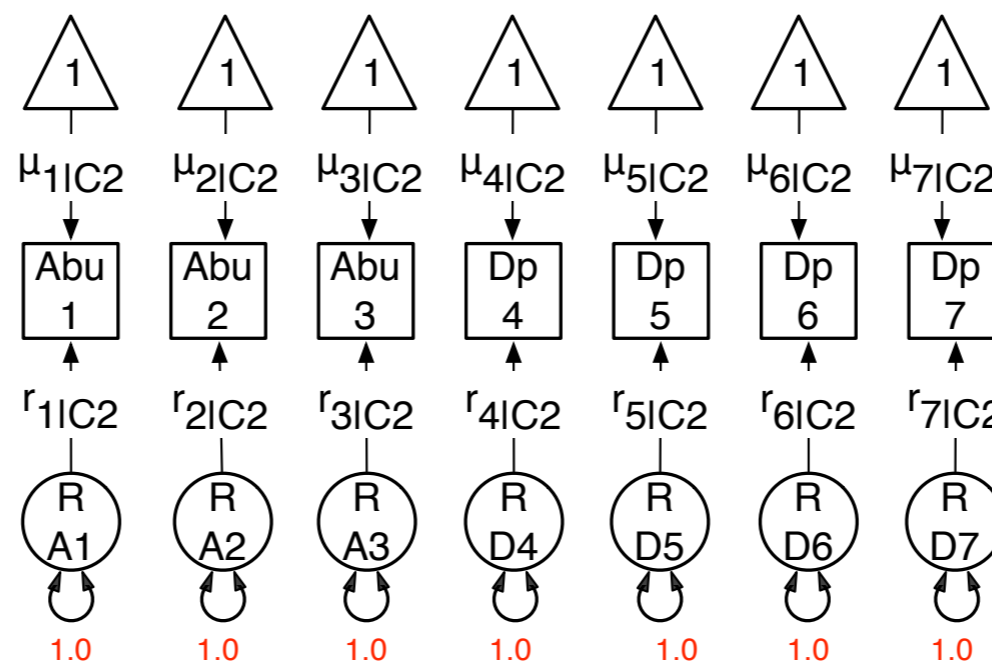
Latent Class (Subgroup) Model

Class 1 probability *p*

Class 2 probability (1-*p*)

Conditionally Independent?!

Expensive!

Factor Mixture Model
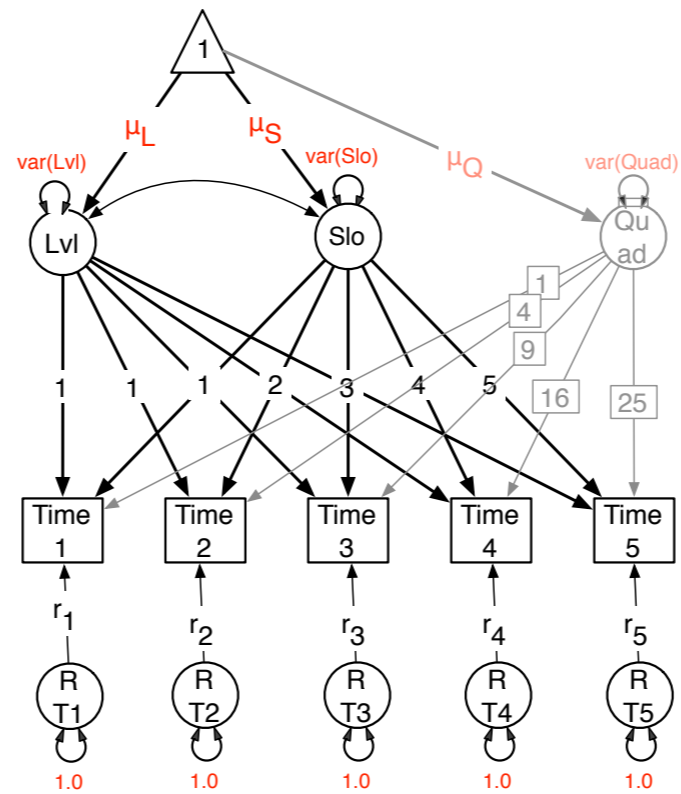
Growth Curve Mixture Model

Class 1
probability
p
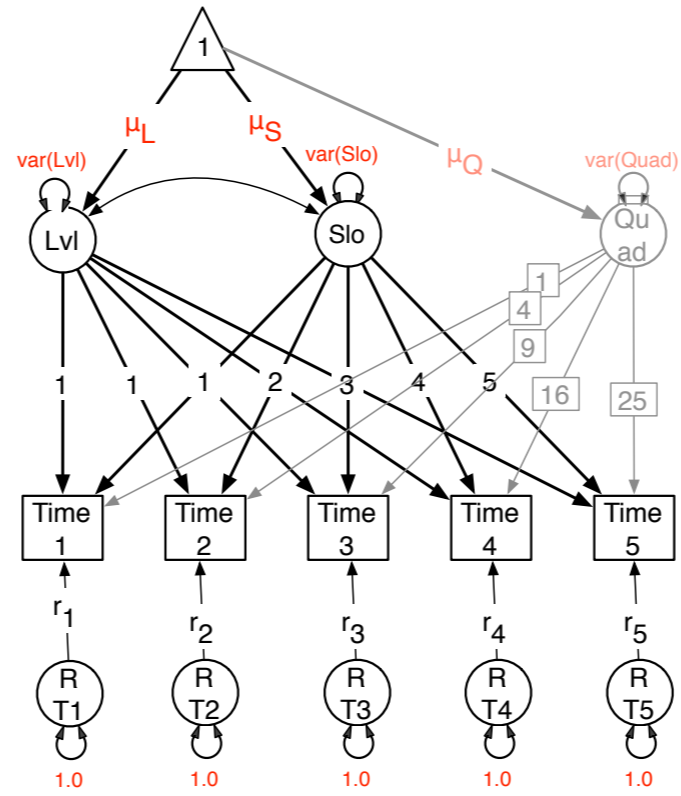
Class 2
probability
$(1-p)$

Regime Switching Model

Posterior Probabilities of Trajectories for Individual 46
−2lnL= 16.108

Regime Switching Model

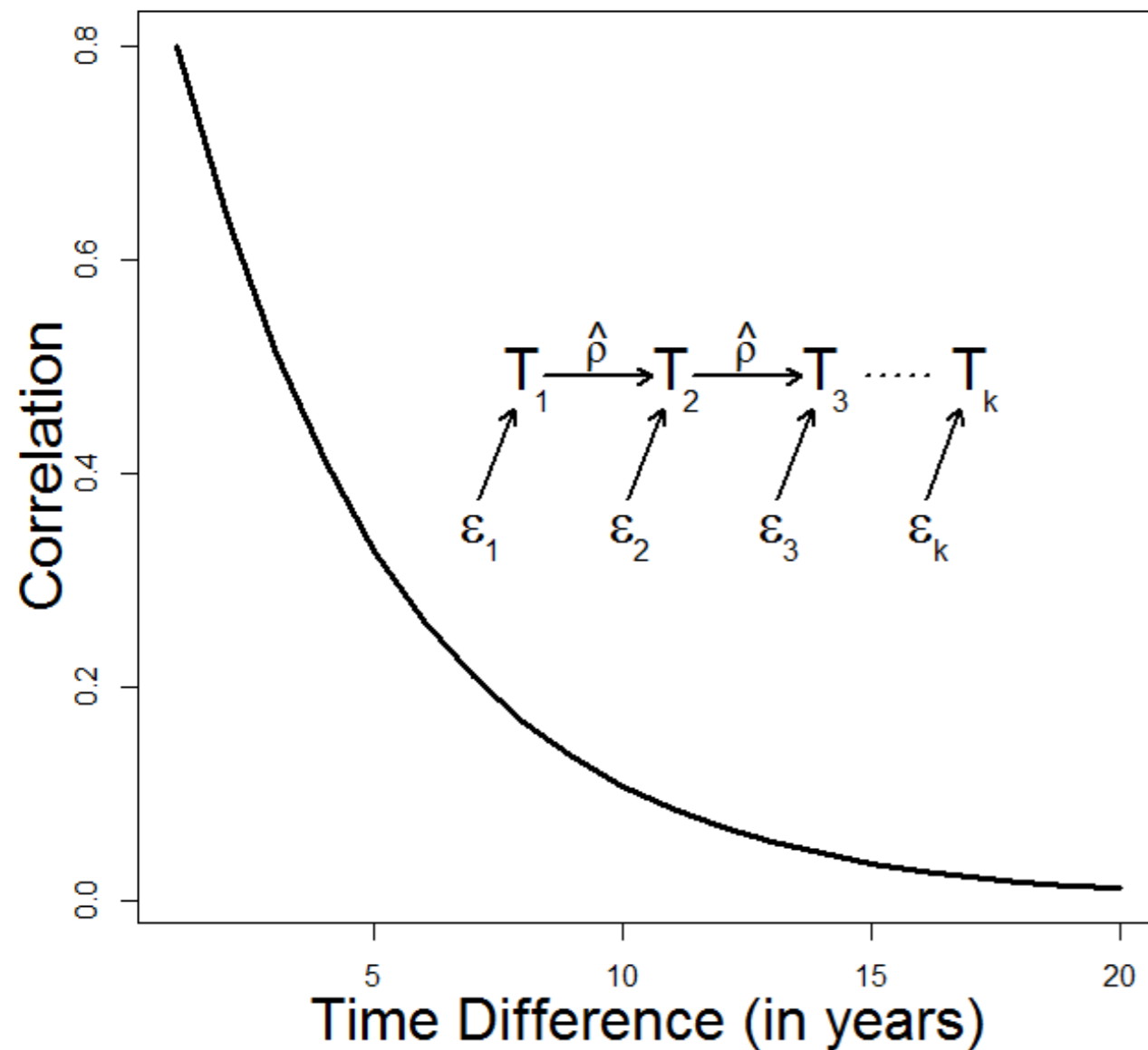Posterior Probabilities of Trajectories for Individual 711
−2lnL= 21.257
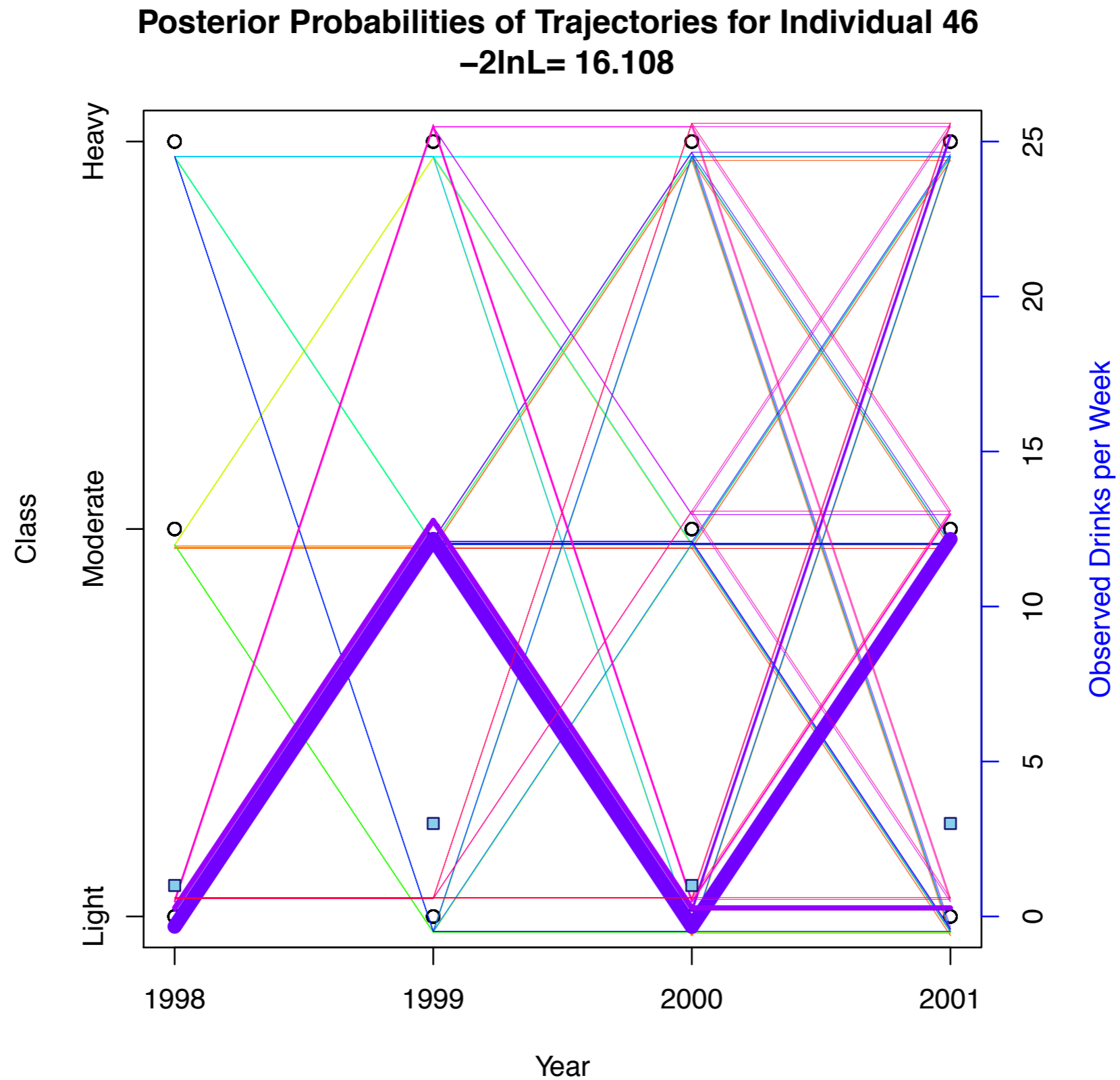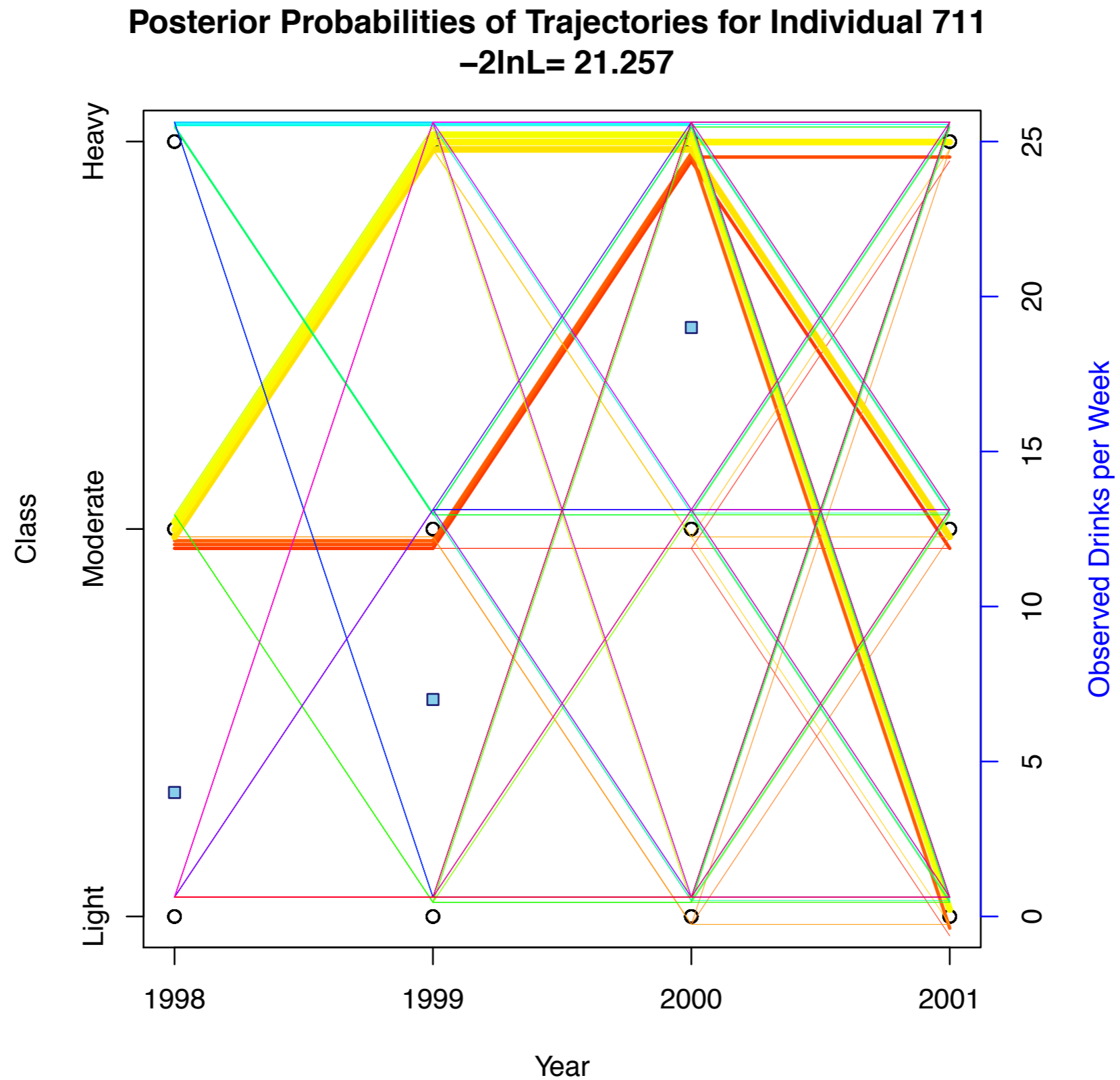
Genetic Heterogeneity

- Genetic factors change during development

- Height

- Neuroticism
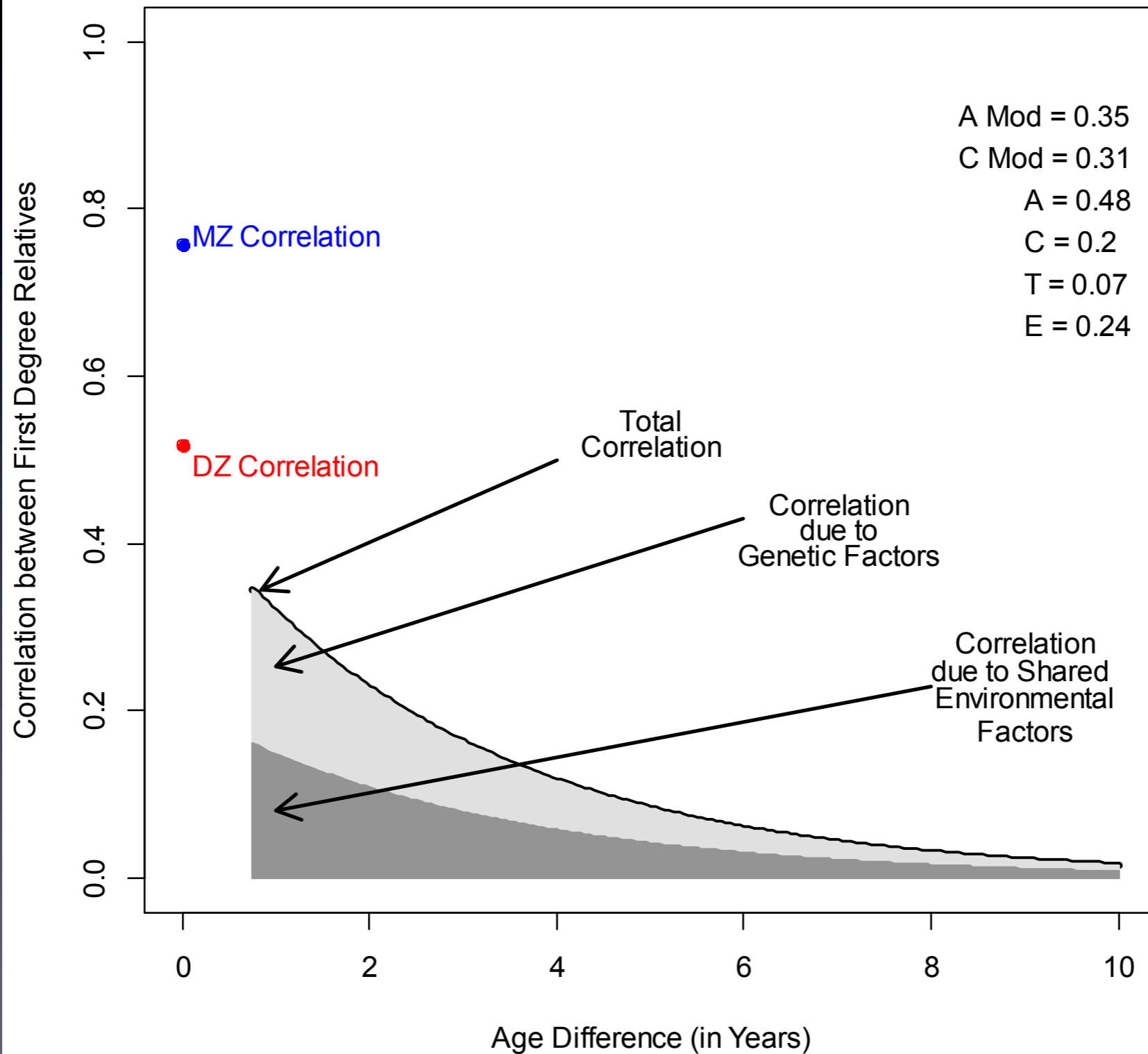
- Detection

Different age, different genes?

Verhulst, Eaves & Neale

$$Cov = Acov \ast e^{-|\Delta age| \ast \alpha_a} + Ccov \ast e^{-|\Delta age| \ast \alpha_c} + Tcov$$

Verhulst, Eaves & Neale

The Decay in the Correlation between First Degree Relatives as a Function of Age Difference

Care with Ascertainment

- Factor Analysis in Cases

- Latent Class Analysis in Cases

- Selection for Case Status

- Selection of Controls

Summary

- Measurement of complex traits is complex

- Measurement invariance desirable

- ML factor scores good start

- Mixture distribution models should be tested

- Choose your study participants carefully

- Analyze what you measure, and measure well what you analyze

Summary

- Measurement of complex traits is complex

- Measurement invariance desirable

- ML factor scores good start

- Mixture distribution models should be tested

- Analyze what you measure, and measure well what you analyze