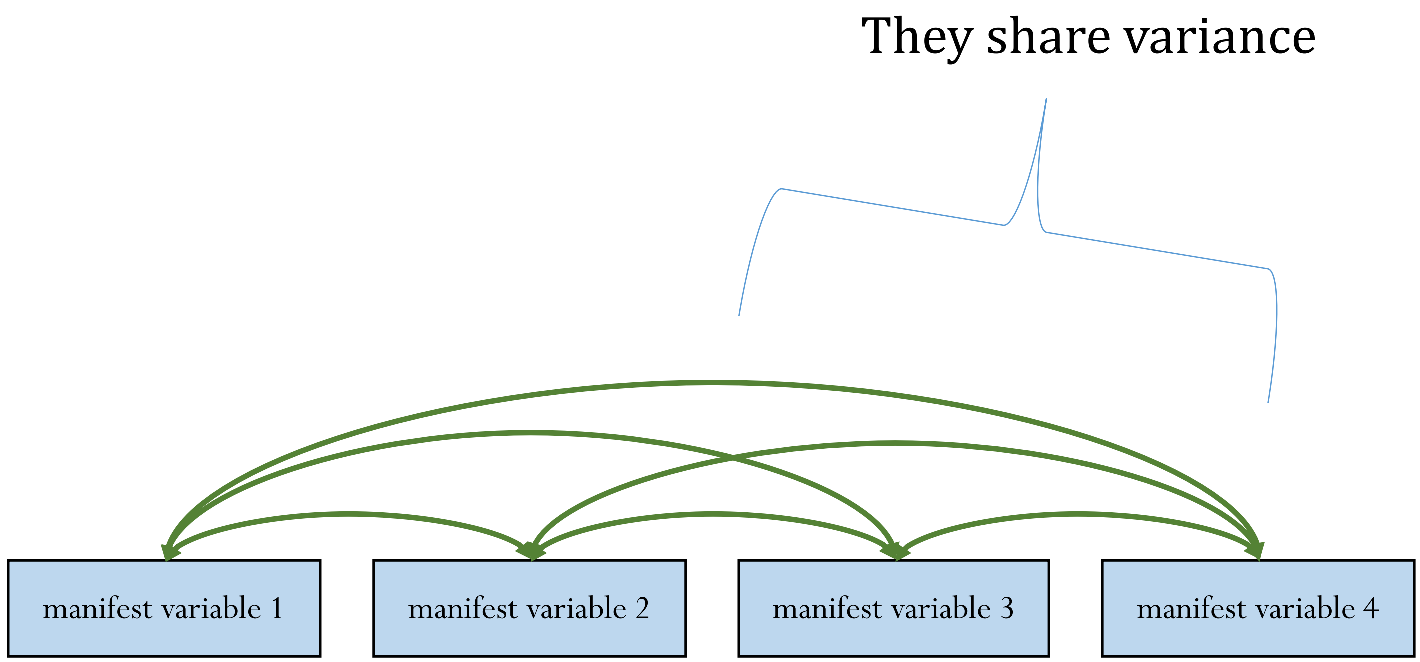
Latent Variable and Network Model Implications for Partial Correlation Structures

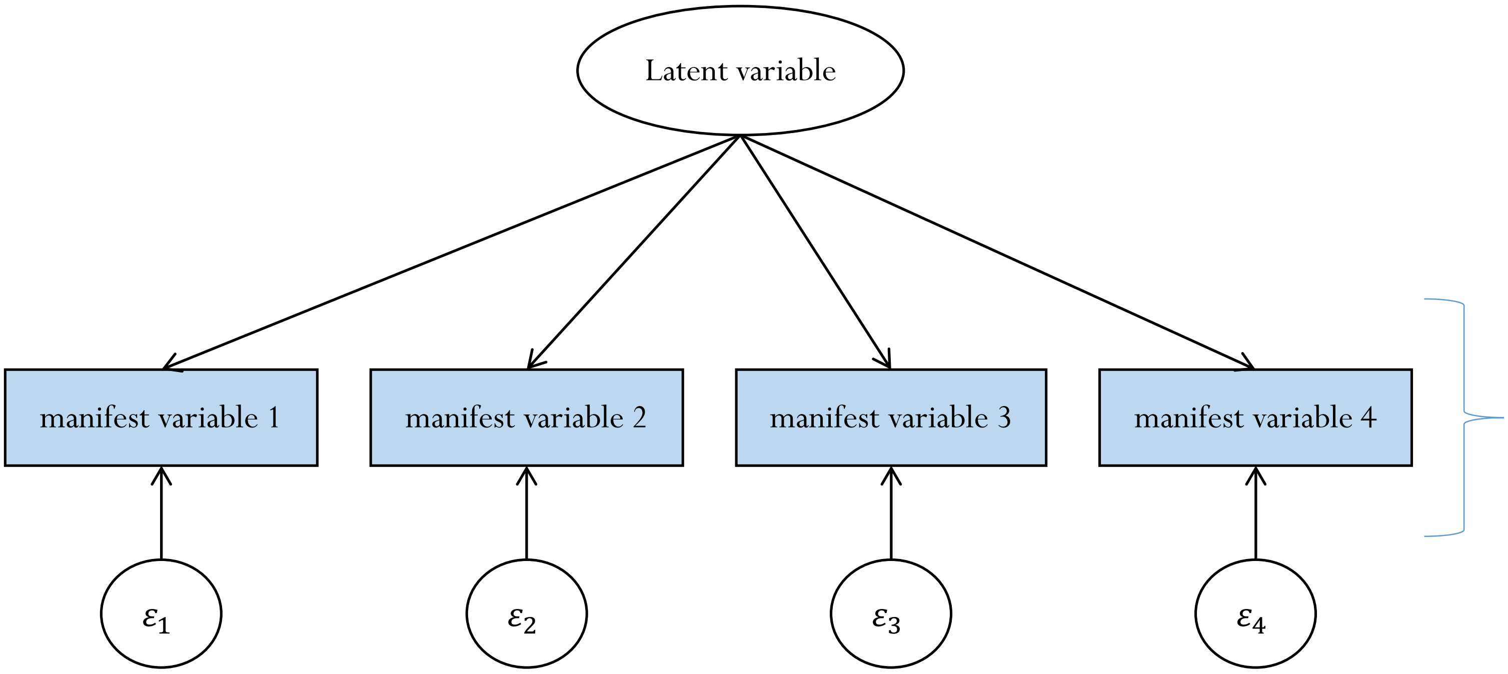
Riet van Bork

University of Amsterdam

The latent variable model

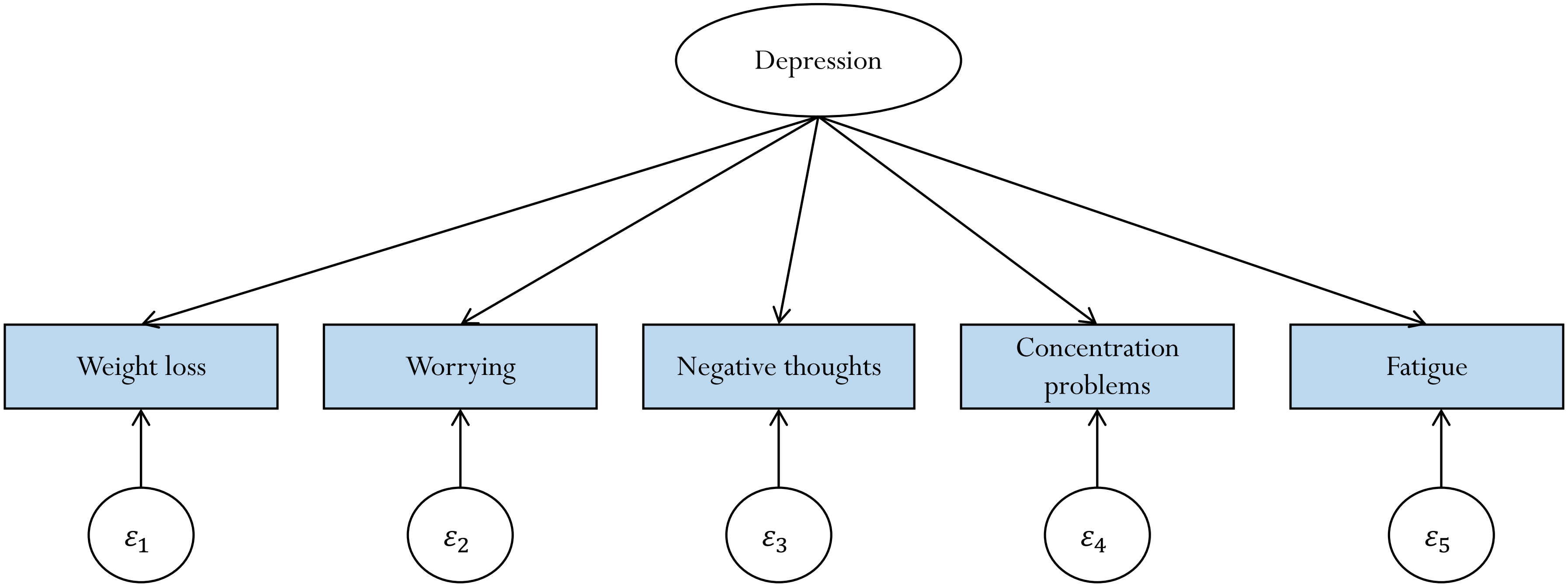


The latent variable model

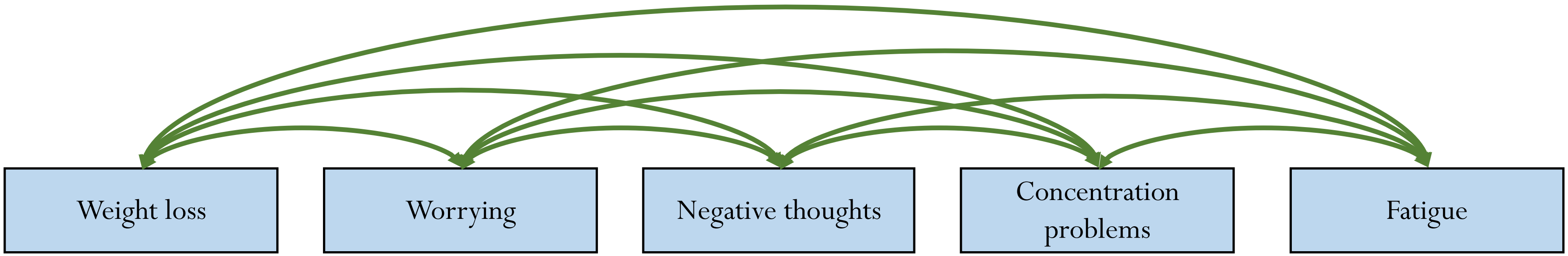


This shared variance is a reflection of their common cause.

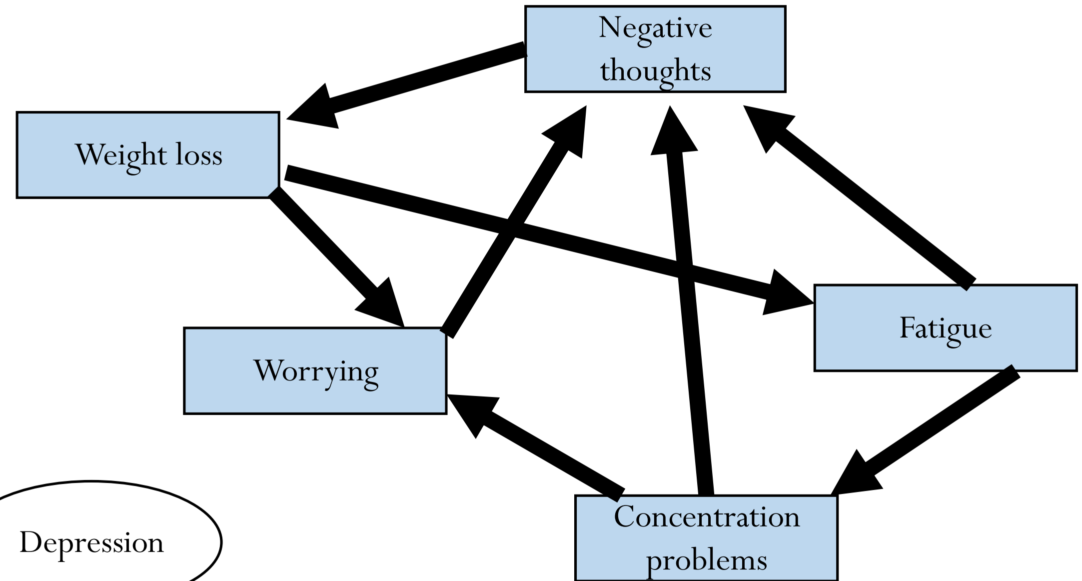
Depression as a latent variable

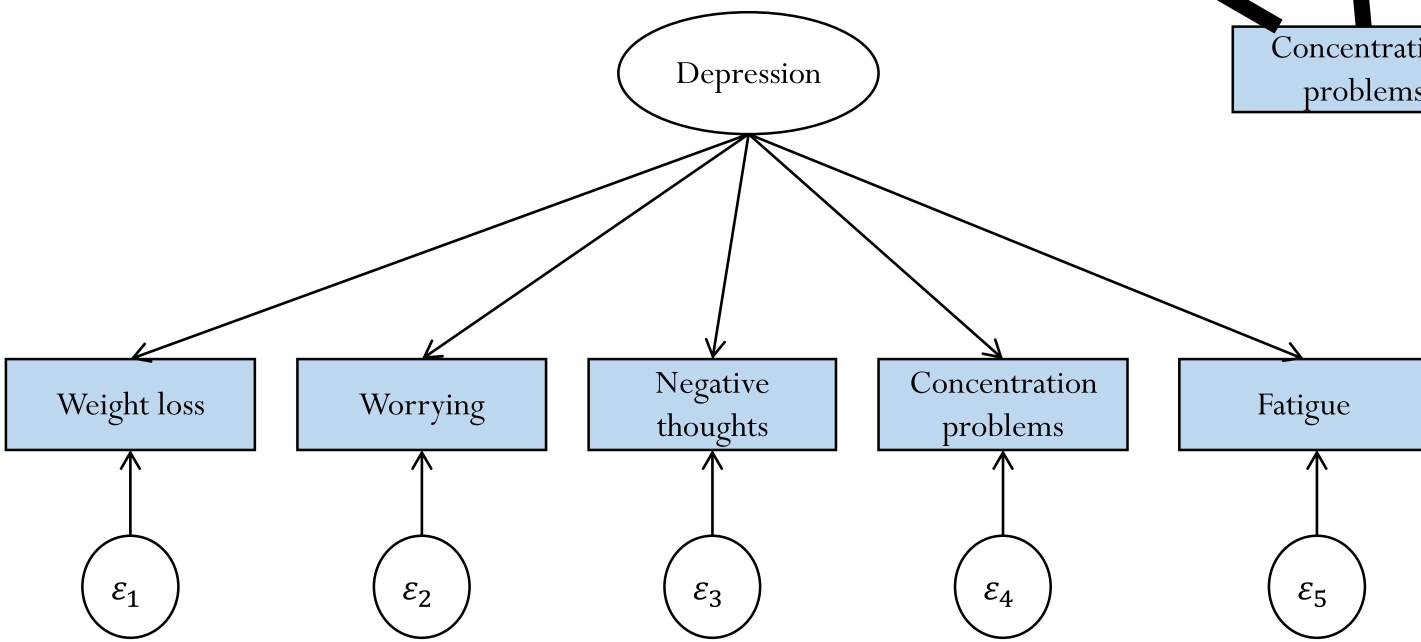


But all we observed were correlations..



LV models vs Networks

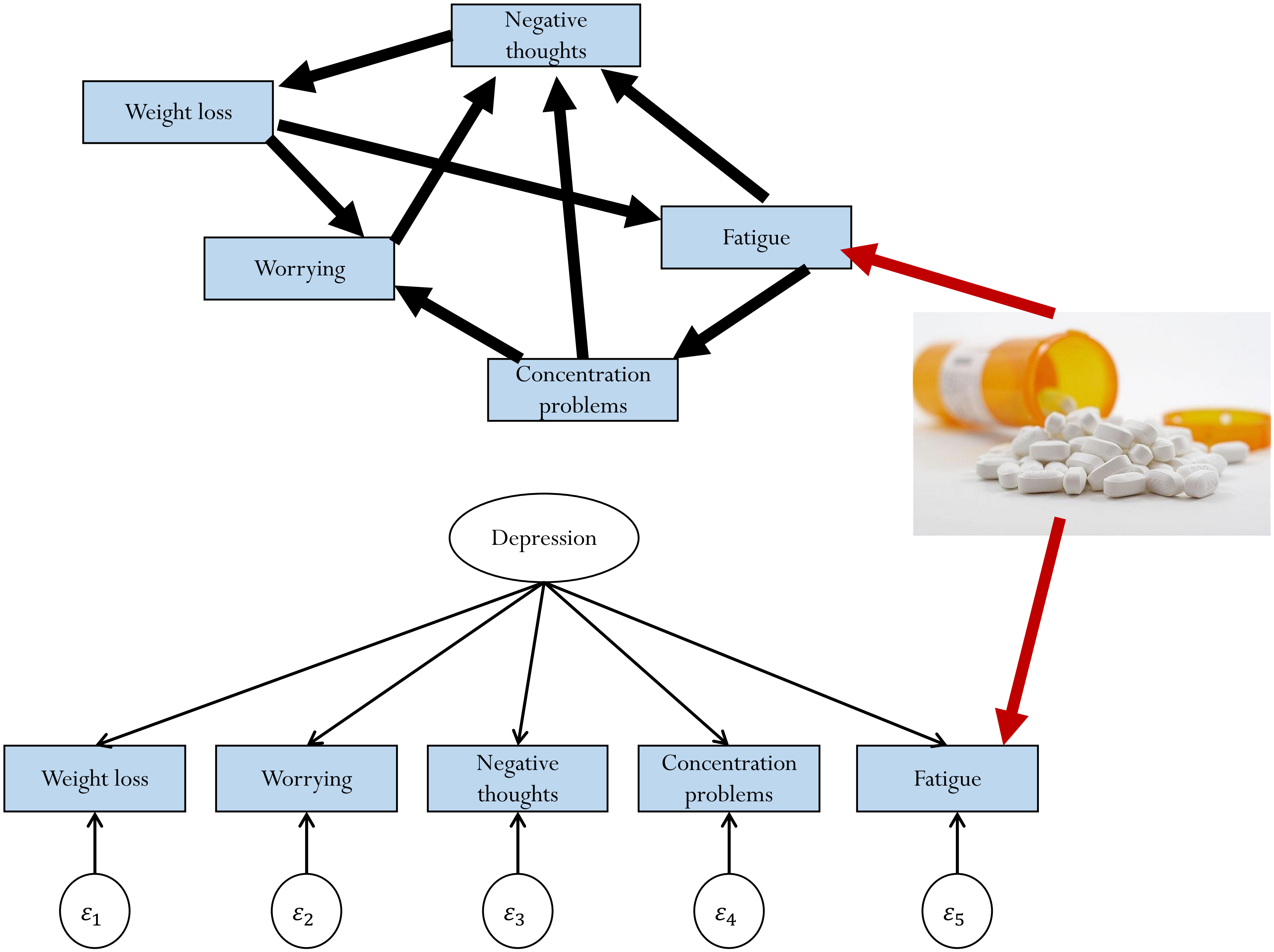


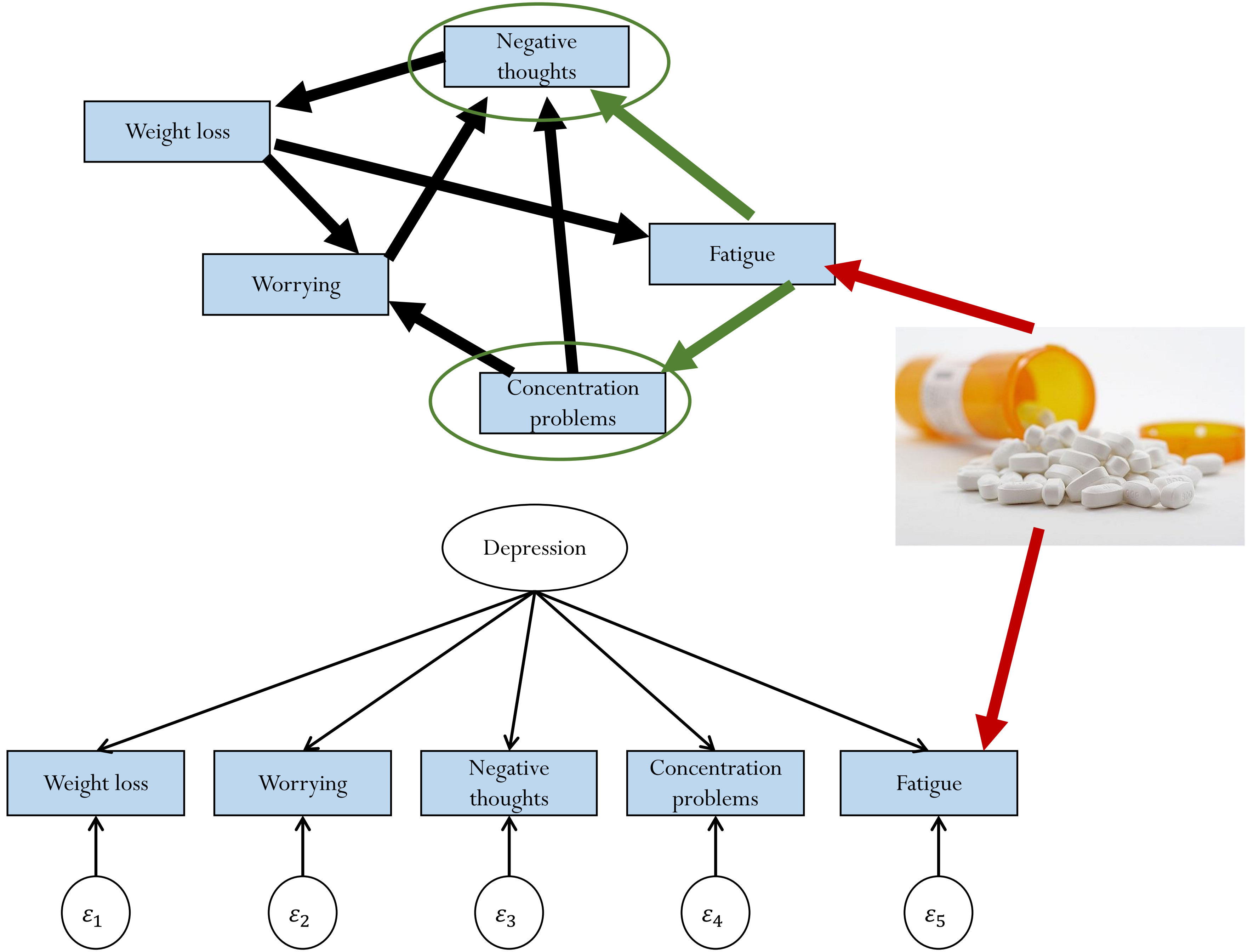


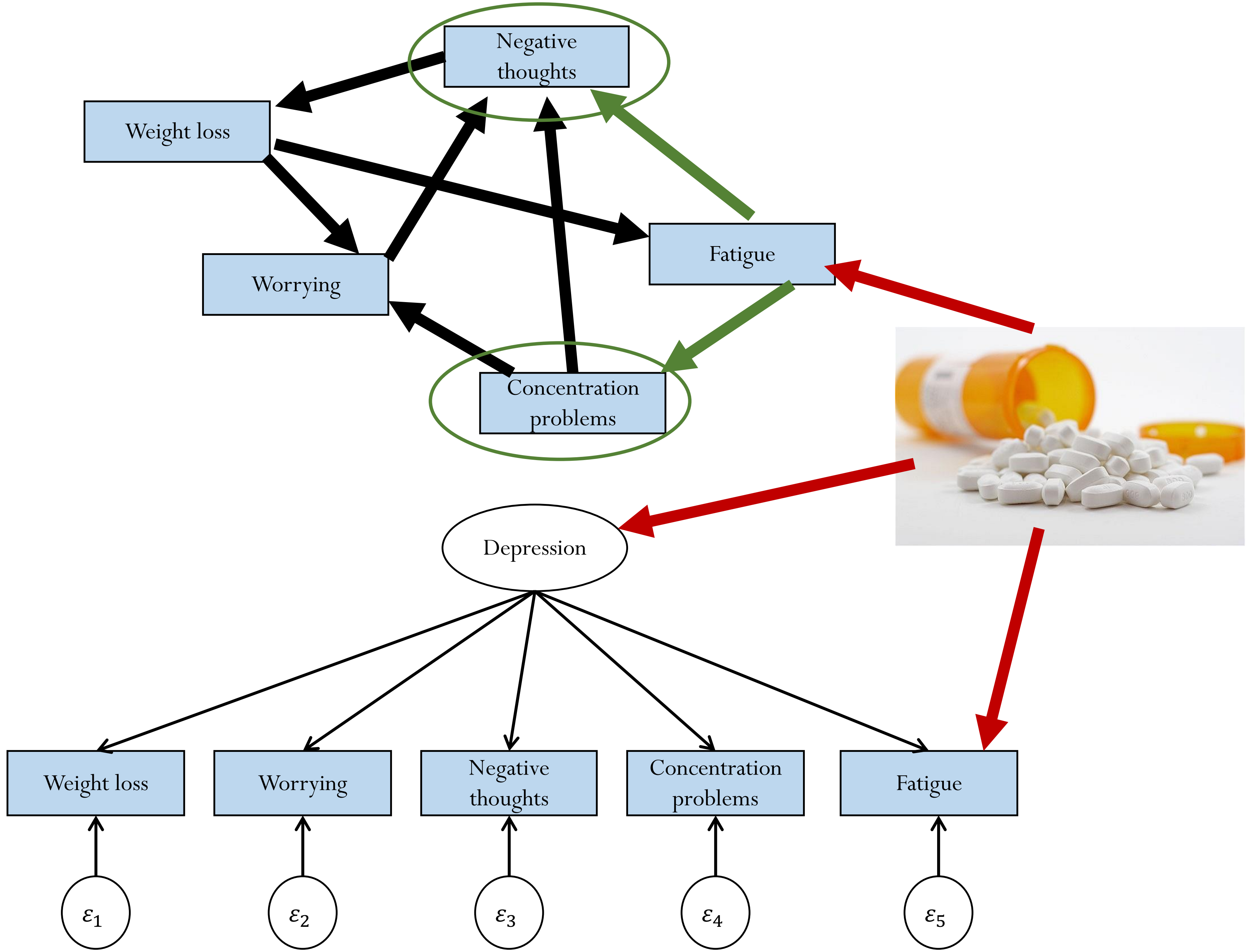
What underlies the correlations we observe?

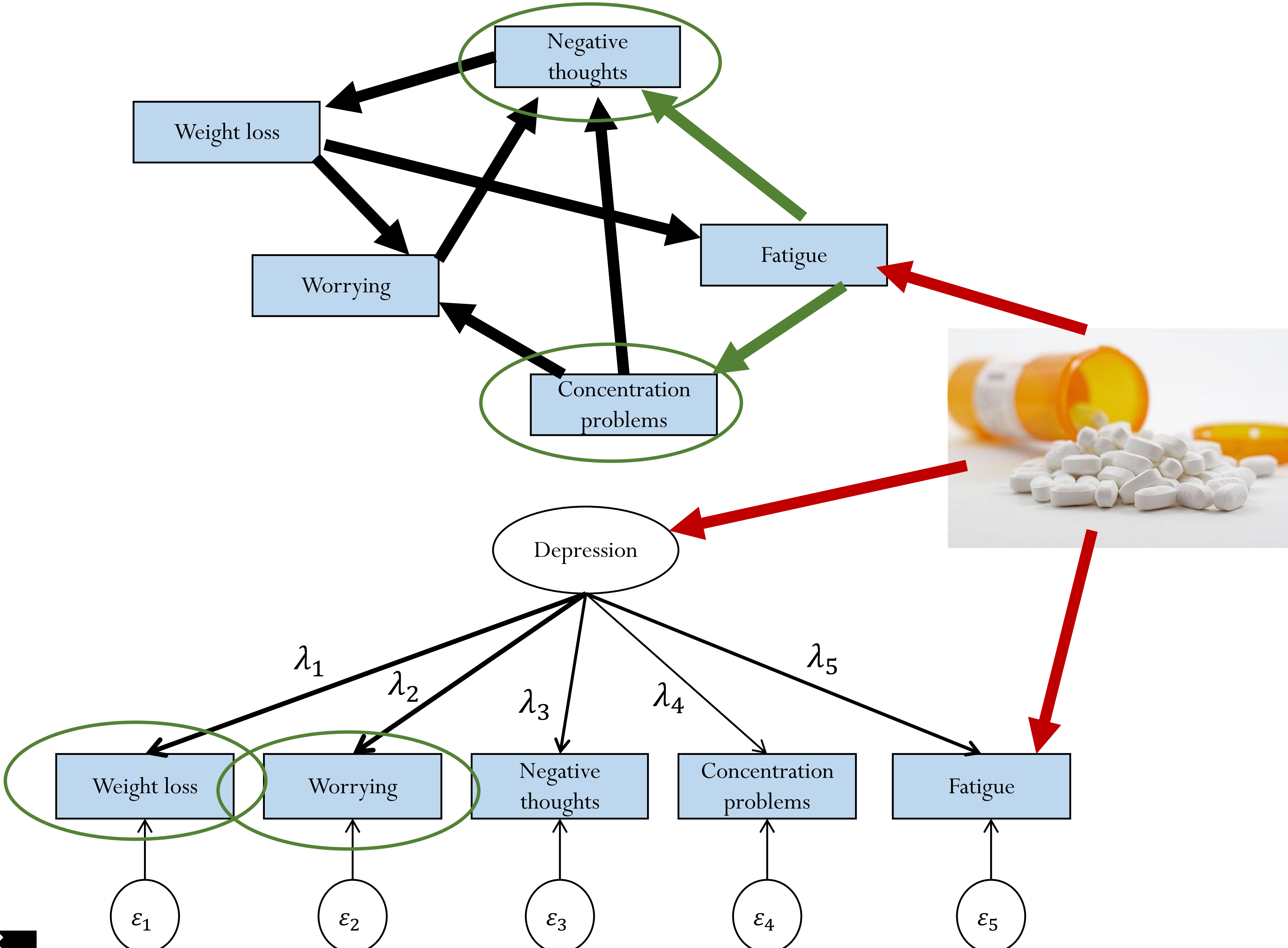
Is it possible to determine whether correlations reflect a common factor or direct relations in a network?

In some cases it is possible to experimentally intervene.


Negative thoughts
Weight loss
Worrying
Fatigue
Concentration problems
Depression
Weight loss
Worrying
Negative thoughts
Concentration problems
Fatigue
ε_1
ε_2
ε_3
ε_4
ε_5

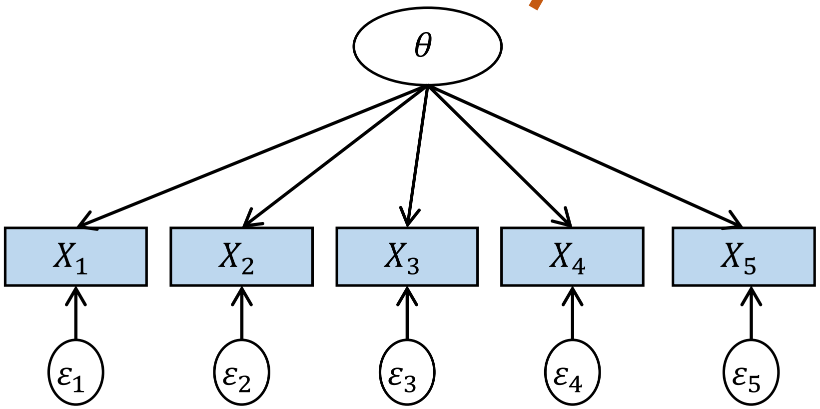

Negative thoughts
Weight loss
Worrying
Fatigue
Concentration problems
Depression
Weight loss
Worrying
Negative thoughts
Concentration problems
Fatigue
ε_1
ε_2
ε_3
ε_4
ε_5

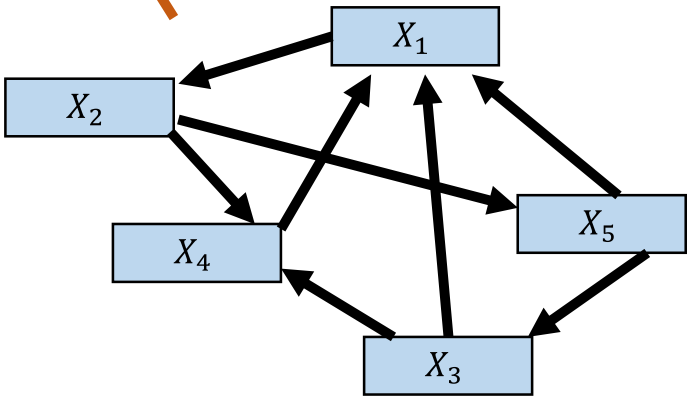

Negative thoughts
Weight loss
Worrying
Fatigue
Concentration problems
Depression
Weight loss
Worrying
Negative thoughts
Concentration problems
Fatigue
ε_1
ε_2
ε_3
ε_4
ε_5


Negative thoughts
Weight loss
Worrying
Fatigue
Concentration problems
Depression
λ_1
λ_2
λ_3
λ_4
λ_5
Weight loss
Worrying
Negative thoughts
Concentration problems
Fatigue
ε_1
ε_2
ε_3
ε_4
ε_5

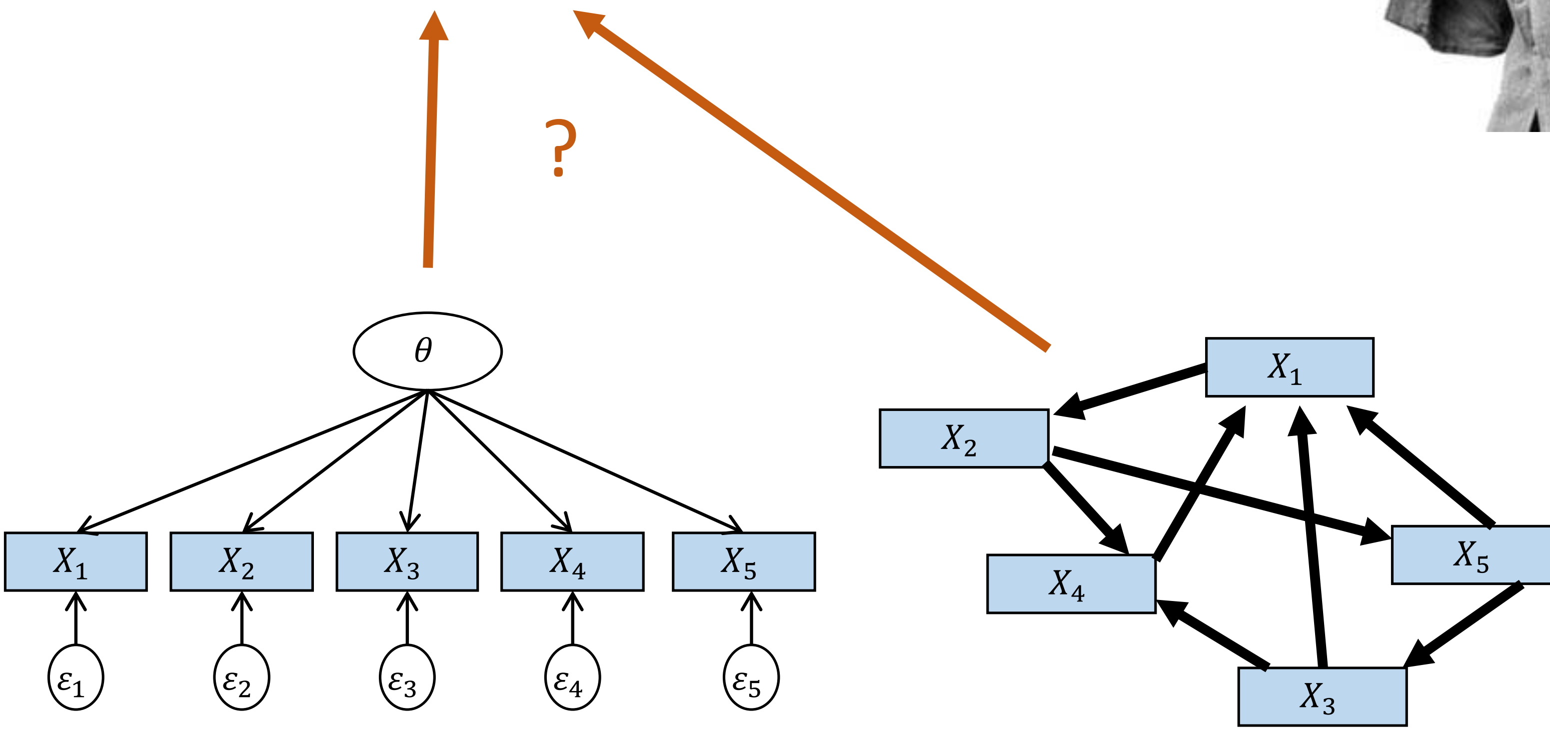
$$\mathbf{S} = \begin{pmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ s_{21} & s_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_p^2 \end{pmatrix}$$

?





$$\mathbf{S} = \begin{pmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ s_{21} & s_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_p^2 \end{pmatrix} \longrightarrow$$ Partial correlations



Factor models and partial correlations

Goal: Determine the underlying data generating model

- 2 implications of factor models for partial correlations
  - 2 tests: pcor zero-test & pcor increase-test
  - Performance of tests
- Conclusion and difference between tests

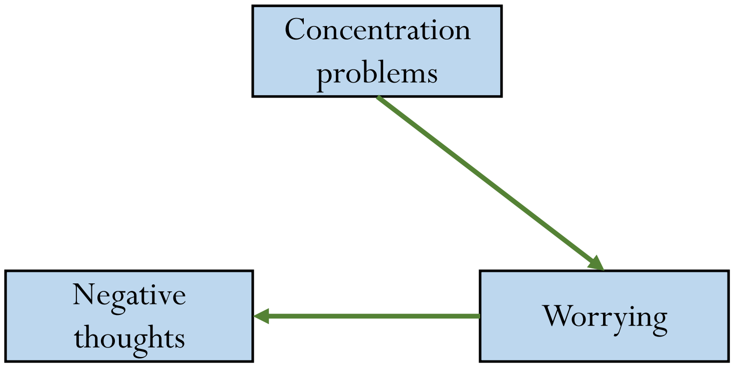
Test 1: pcor zero-test

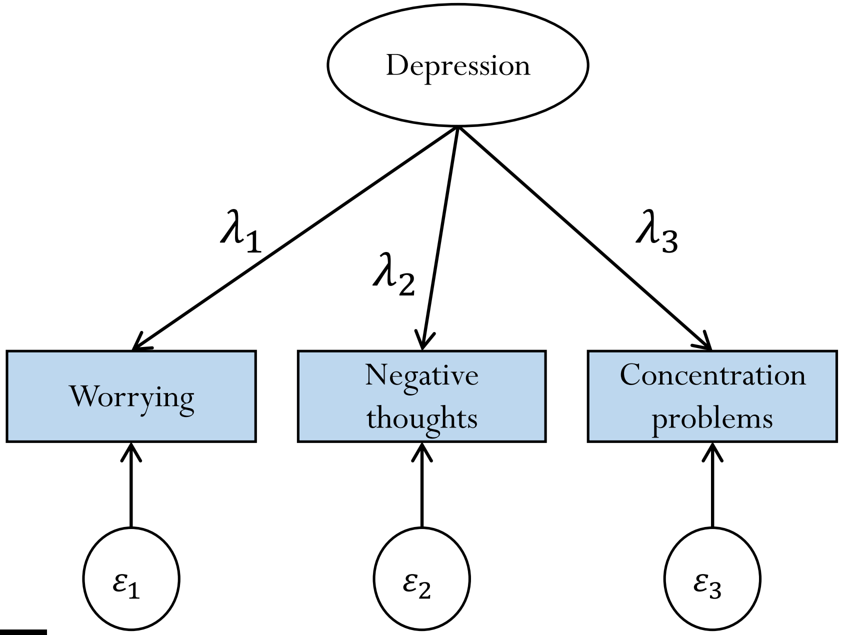
Unidimensional Factor models imply:

1. that the partial correlations between two manifest variables cannot be zero (Holland & Rosenbaum, 1986).

Test 1: pcor zero-test

$$\rho_{xy.z} = \frac{\rho_{xy} - \rho_{yz}\rho_{xz}}{\sqrt{(1-\rho_{yz}^2)(1-\rho_{xz}^2)}}$$





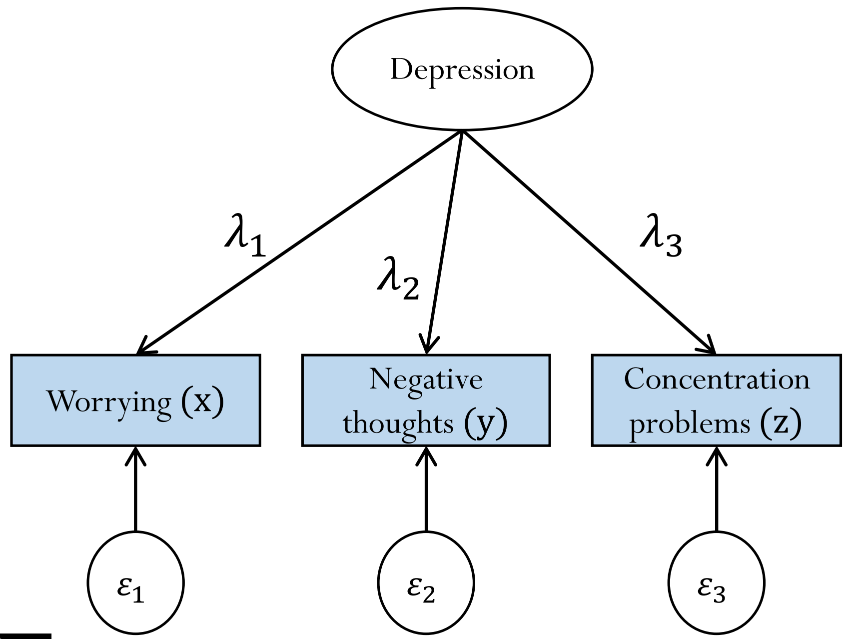
Test 1: pcor zero-test

$$\rho_{xy.z} = \frac{\rho_{xy} - \rho_{yz}\rho_{xz}}{\sqrt{(1-\rho_{yz}^2)(1-\rho_{xz}^2)}}$$

$\rho_{xy} = \lambda_1 * \lambda_2$
$\rho_{xz} = \lambda_1 * \lambda_3$
$\rho_{yz} = \lambda_2 * \lambda_3$

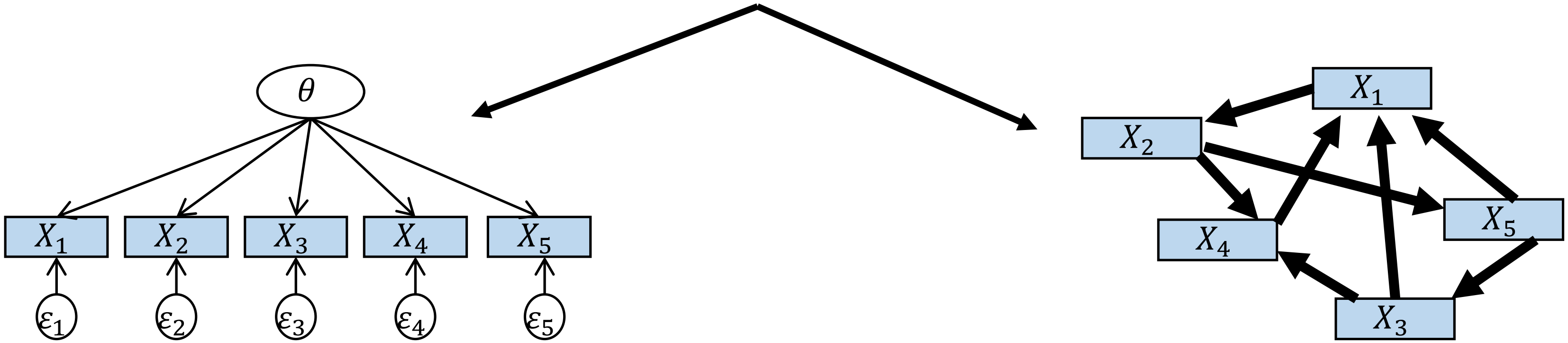
$\rho_{xy} - \rho_{yz}\rho_{xz} = (\lambda_1 * \lambda_2) - (\lambda_1 * \lambda_3 * \lambda_2 * \lambda_3)$
$= (\lambda_1 * \lambda_2) - (\lambda_1 * \lambda_2) * \lambda_3 * \lambda_3$
$\rho_{xy.z} = 0$ iff $\lambda_3 = 1$ or -1

So, for partial correlations to be zero at least one of the variables partialled out should have a factor loading of 1 or -1.



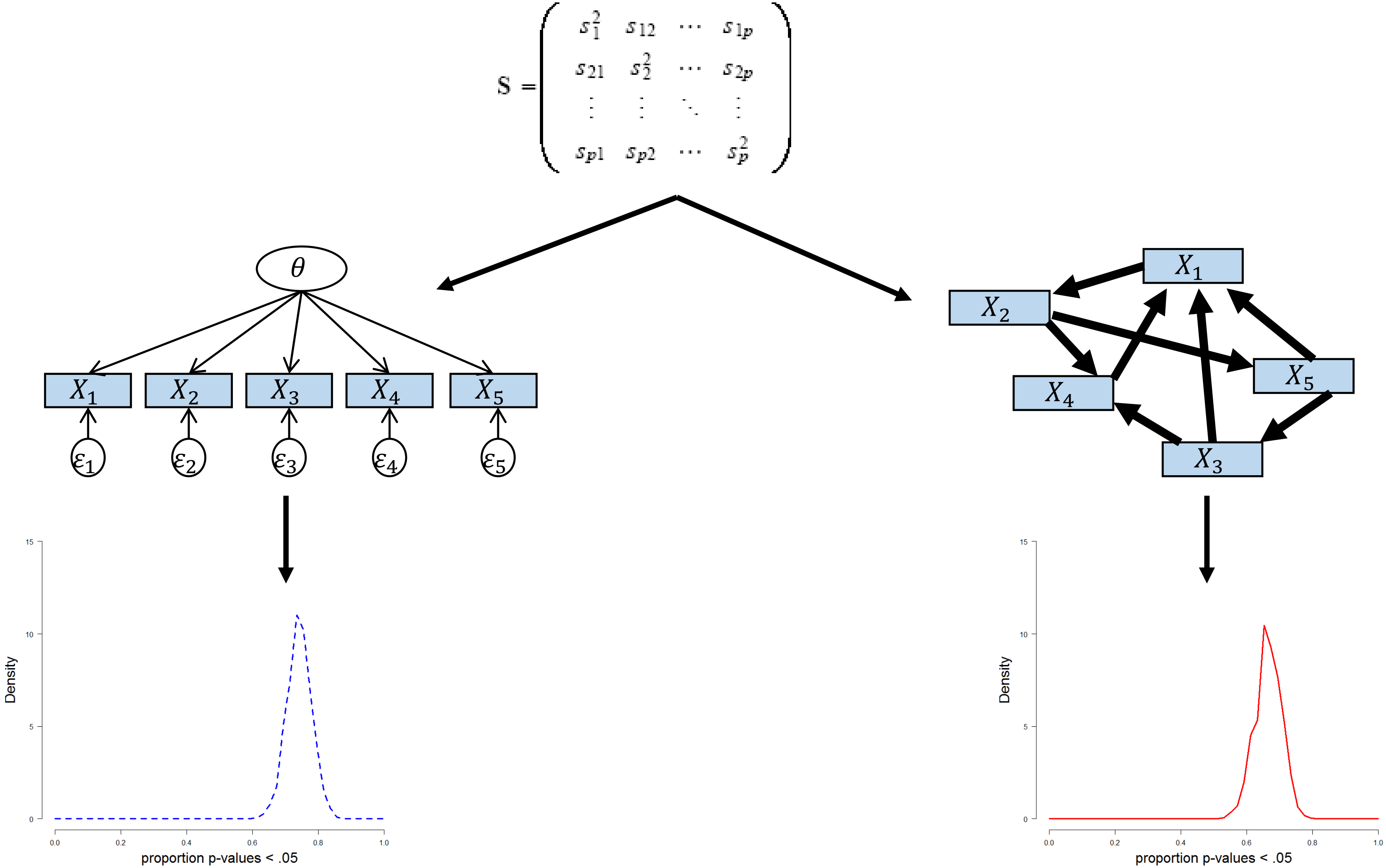
$$\mathbf{S} = \begin{pmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ s_{21} & s_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_p^2 \end{pmatrix}$$

$$\mathbf{S} = \begin{pmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ s_{21} & s_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_p^2 \end{pmatrix}$$

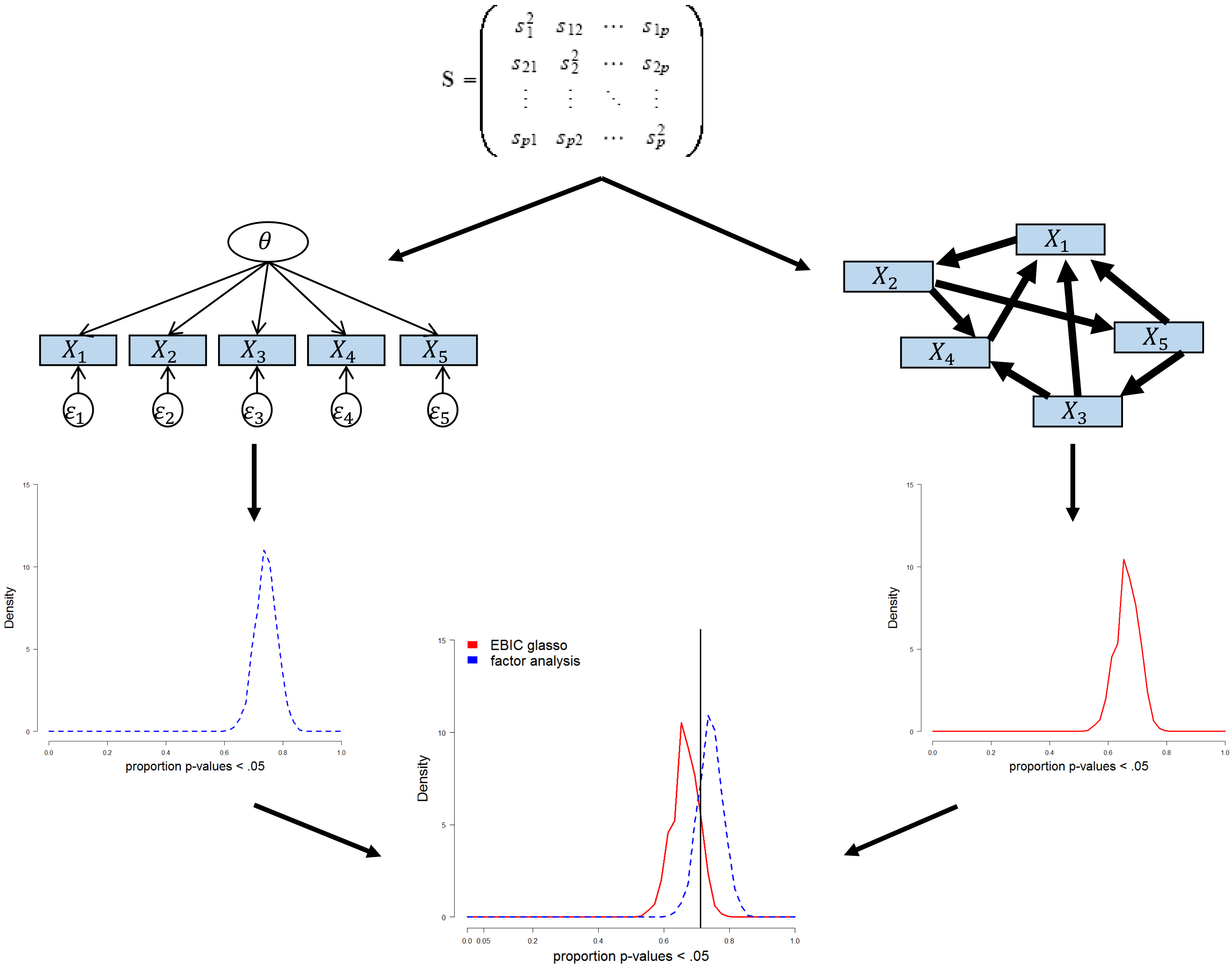


Lavaan

Extended BIC gLasso



S =
θ
X1
X2
X3
X4
X5
ε1
ε2
ε3
ε4
ε5
Density
proportion p-values < .05


S =
EBIC glasso
factor analysis
Density
proportion p-values < .05

Recap in words:

1. Estimate a one-factor model (with Lavaan) and a graphical lasso (gLasso) network based on the Extended BIC (EBIC) (with qgraph).

2. Simulate multiple datasets* according to both estimated models and calculate the proportion significant partial correlations.

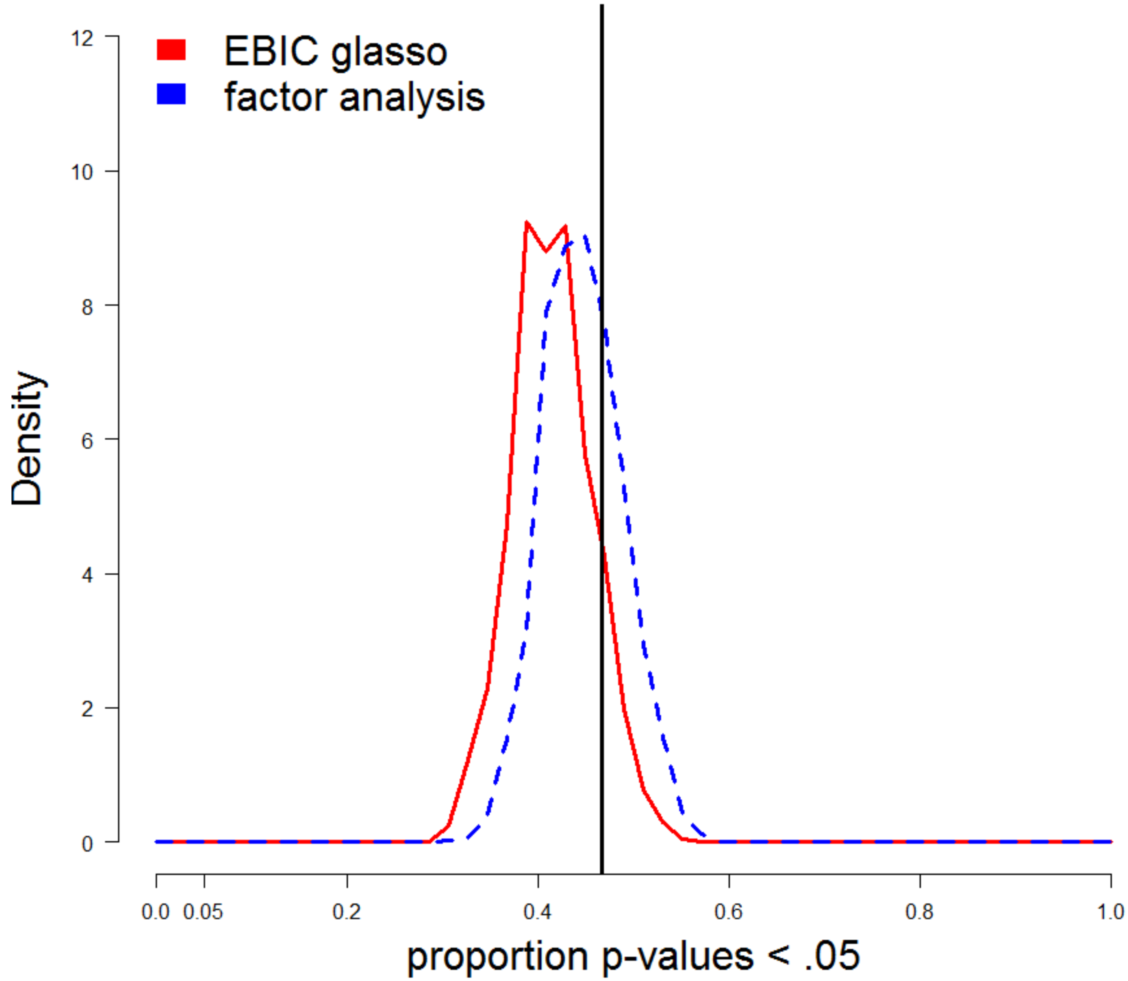
3. Use these proportions to make two density plots; one for the one-factor model and one for the network model.

4. Does the proportion significant partial correlations in the data have a higher density in the PDF of the one-factor model or the network model?

*equal N as in data!

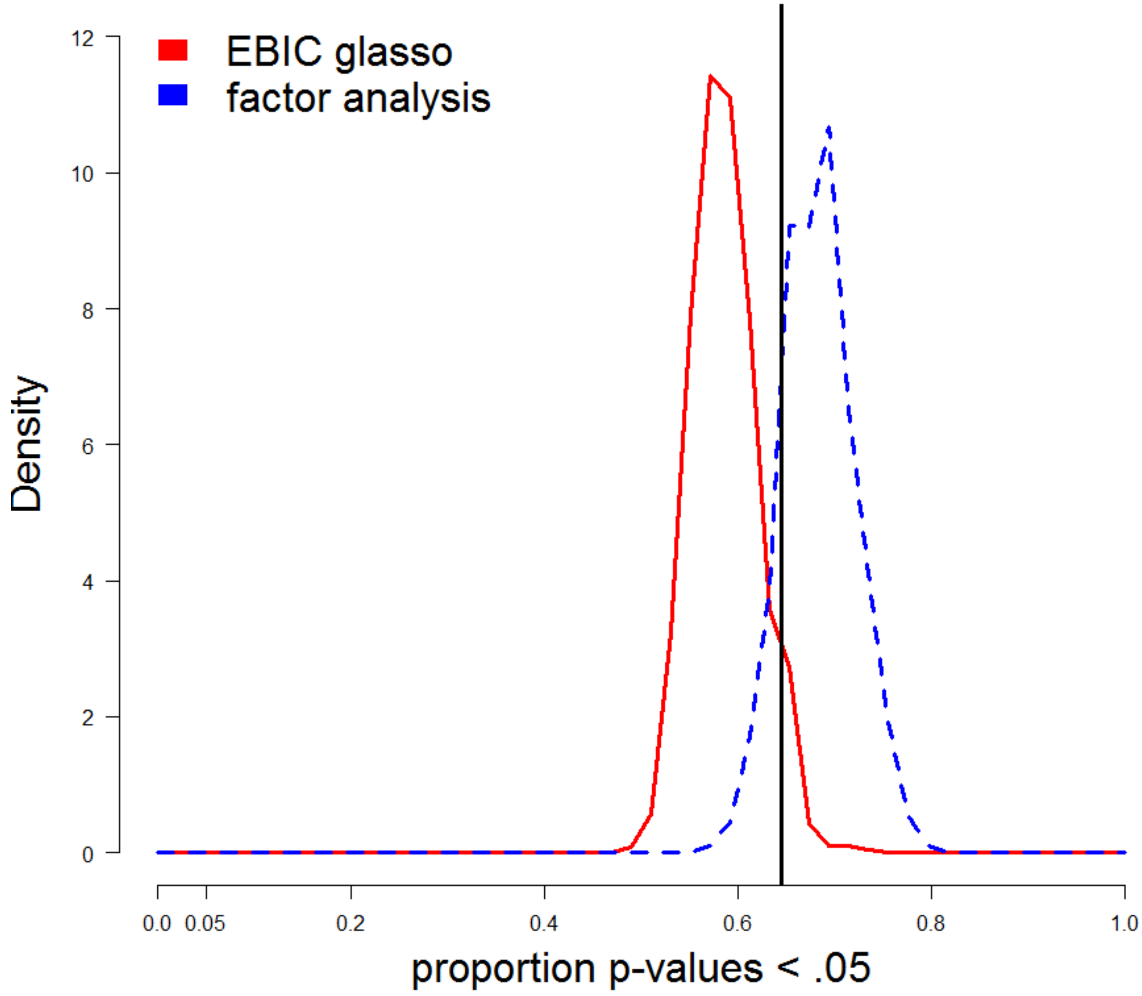
Results 1 : Factor model

- True model = Factor model
- N= **1000**
- $\lambda \sim U(.1, .9)$
- 10 manifest variables



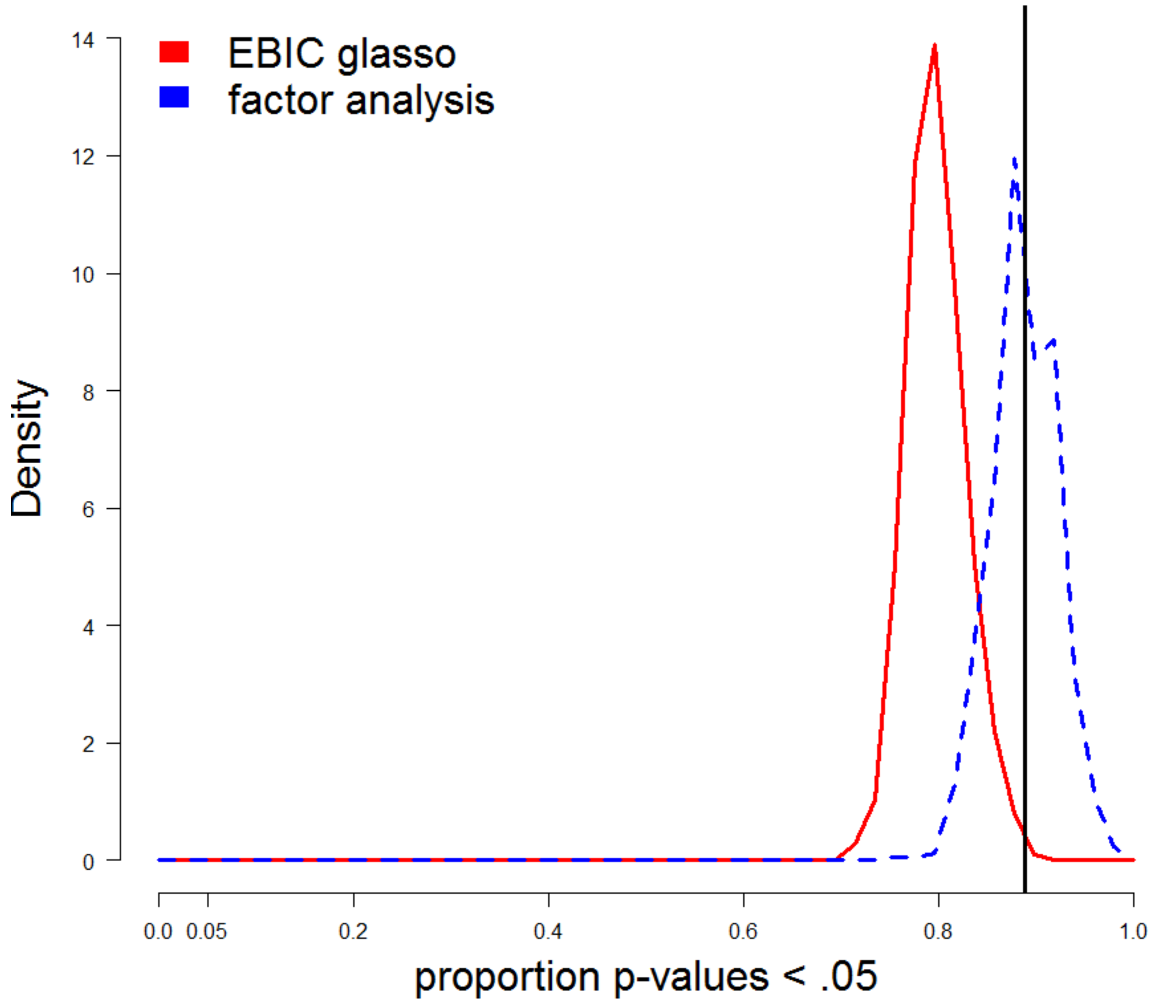
Results 1 : Factor model

- True model = Factor model
- N= **5000**
- $\lambda \sim U(.1\ ,.9)$
- 10 manifest variables



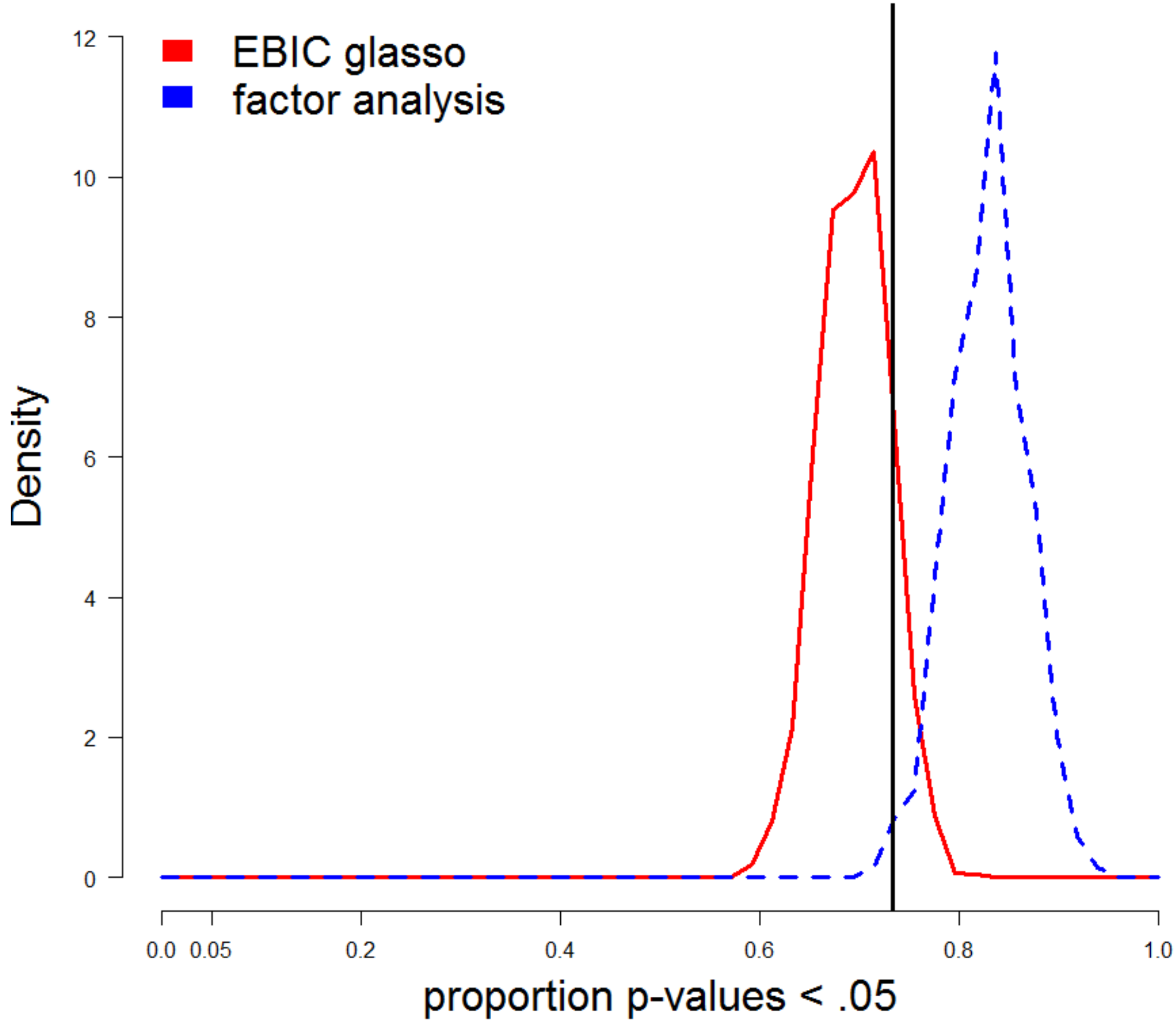
Results 1 : Factor model

- True model = Factor model
- N= **10.000**
- $\lambda \sim U(.1 , .9)$
- 10 manifest variables



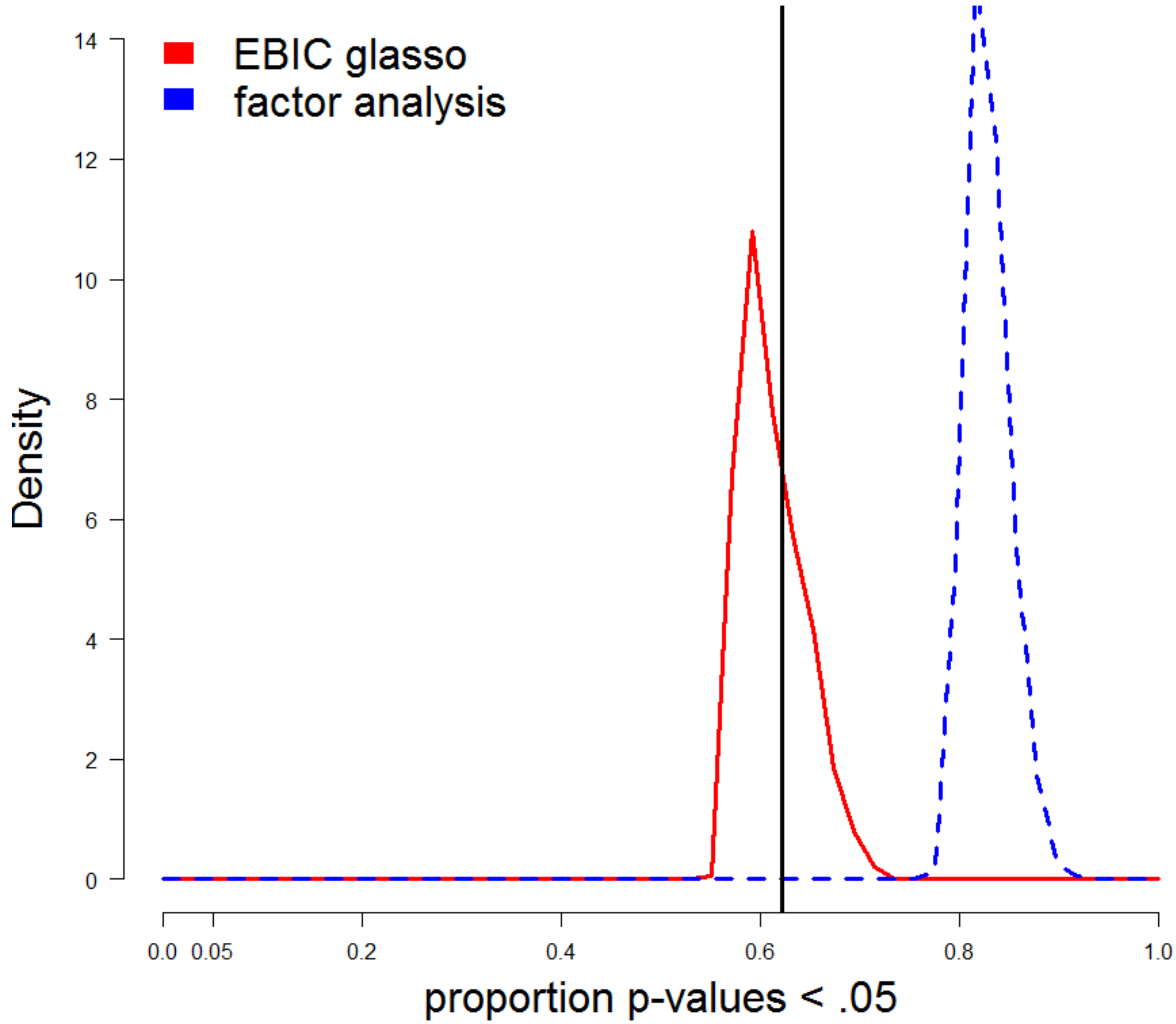
Results 1 : Network model

- True model = network
- N= **1000**
- Proportion partial correlations zero in population = 0.5
- 10 manifest variables



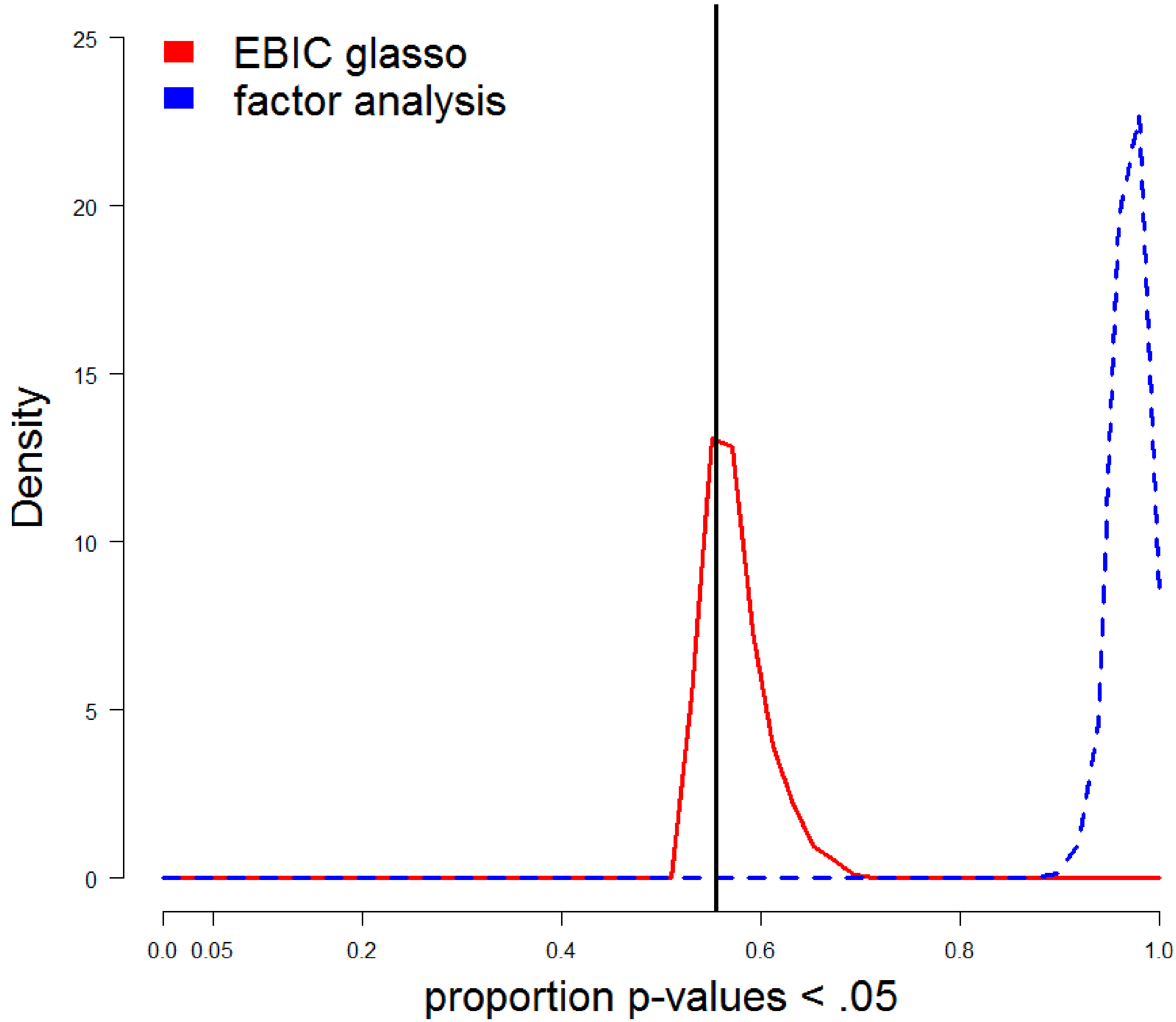
Results 1 : Network model

- True model = network
- N= **5000**
- Proportion partial correlations zero in population = 0.5
- 10 manifest variables



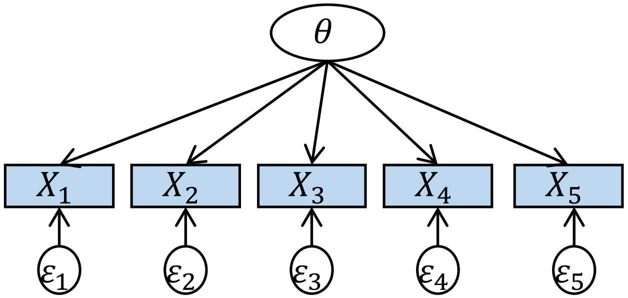
Results 1 : Network model

- True model = network
- N= **10.000**
- Proportion partial correlations zero in population = 0.5
- 10 manifest variables

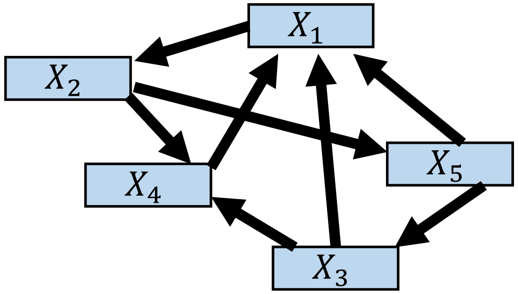


Results 1

How often does this test choose the right model?



65% (N=1000)



91.3% (N=1000)

Test 2: pcor increase-test

Unidimensional Factor models imply:

1. that the partial correlations between two manifest variables cannot be zero (Holland & Rosenbaum, 1986).

Test 2: pcor increase-test

Unidimensional Factor models imply:

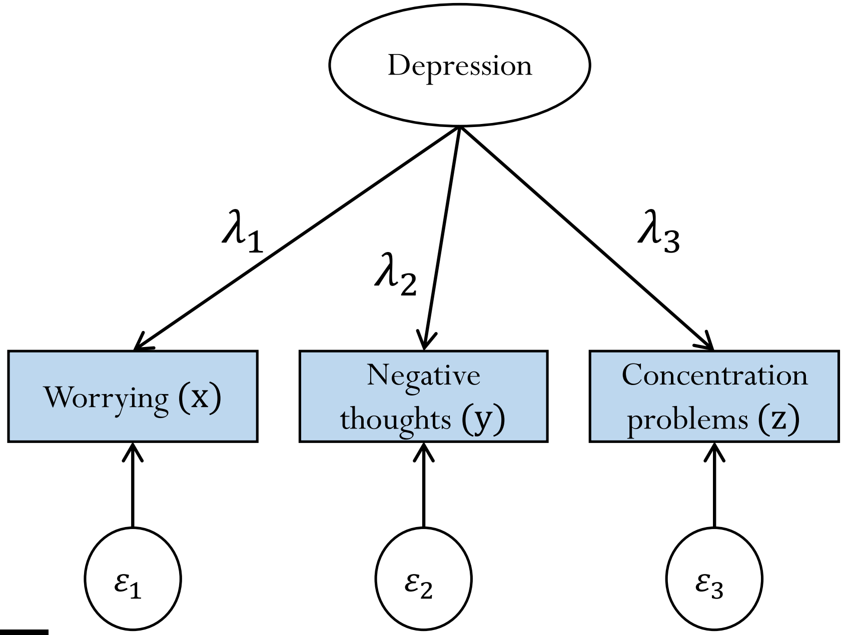
1. that the partial correlations between two manifest variables cannot be zero (Holland & Rosenbaum, 1986).

2. that the partial correlations are always weaker than the corresponding simple correlations.

Test 2: pcor increase-test

$$\rho_{xy.z} = \frac{\rho_{xy} - \rho_{yz}\rho_{xz}}{\sqrt{(1-\rho_{yz}^2)(1-\rho_{xz}^2)}}$$

$\rho_{12} = \lambda_1 \times \lambda_2$
$\rho_{13} = \lambda_1 \times \lambda_3$
$\rho_{23} = \lambda_2 \times \lambda_3$

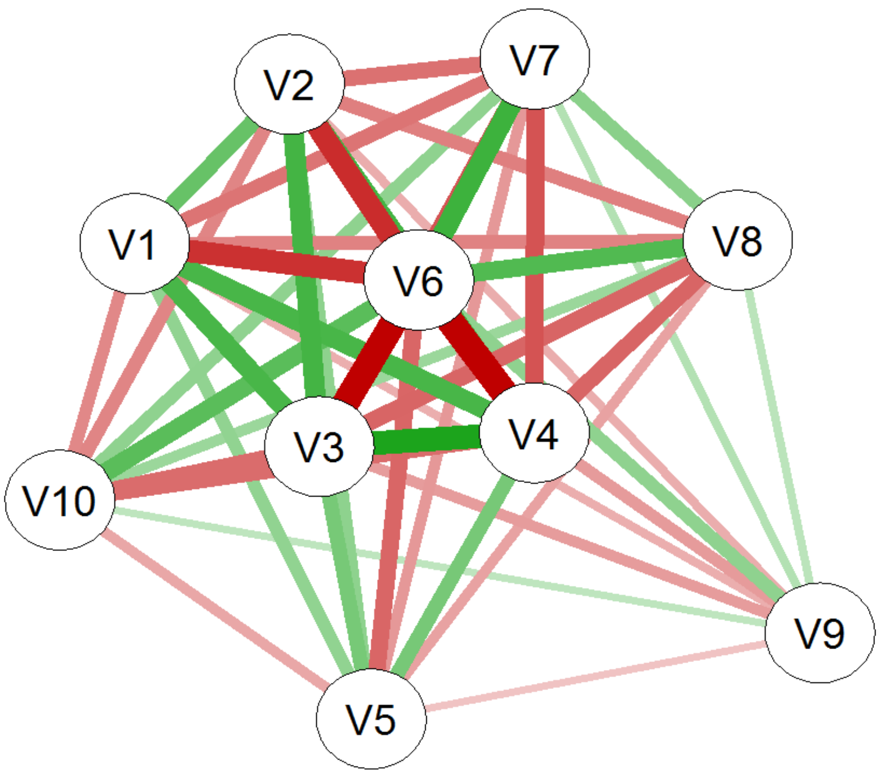


One of the correlations has to be negative to get stronger partial correlation than corresponding simple correlations.

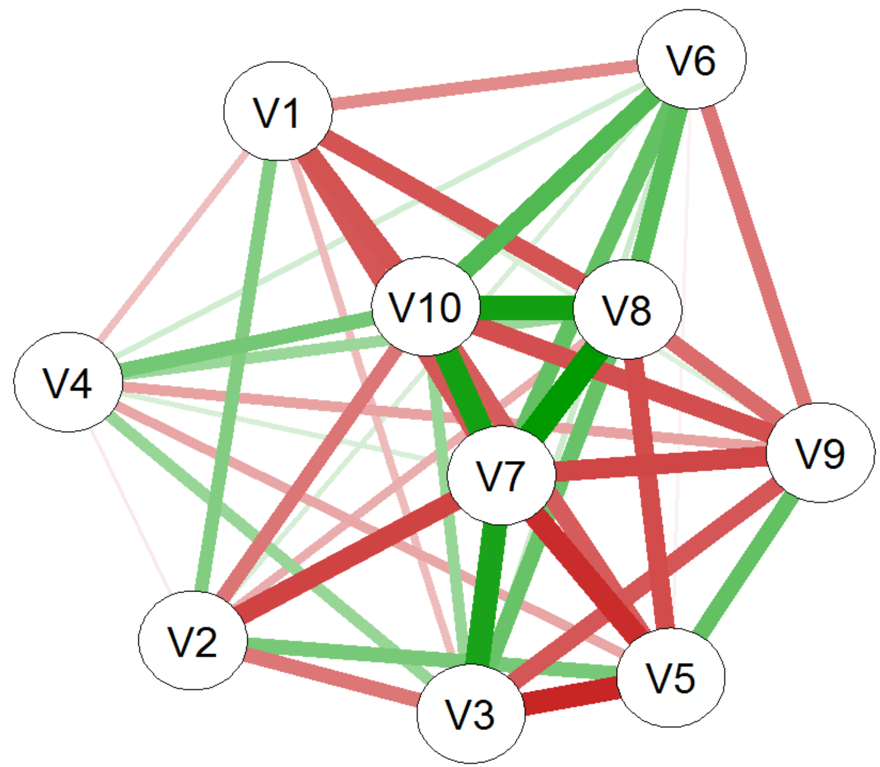
This is not possible for a one factor model.

Test 2: pcor increase-test

One-Factor model



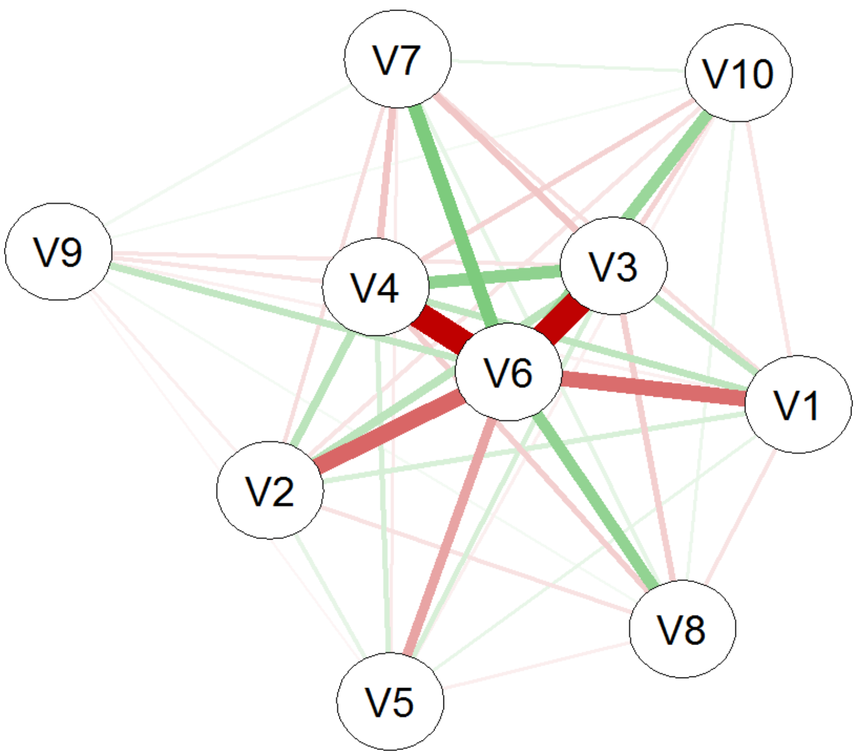
Network model



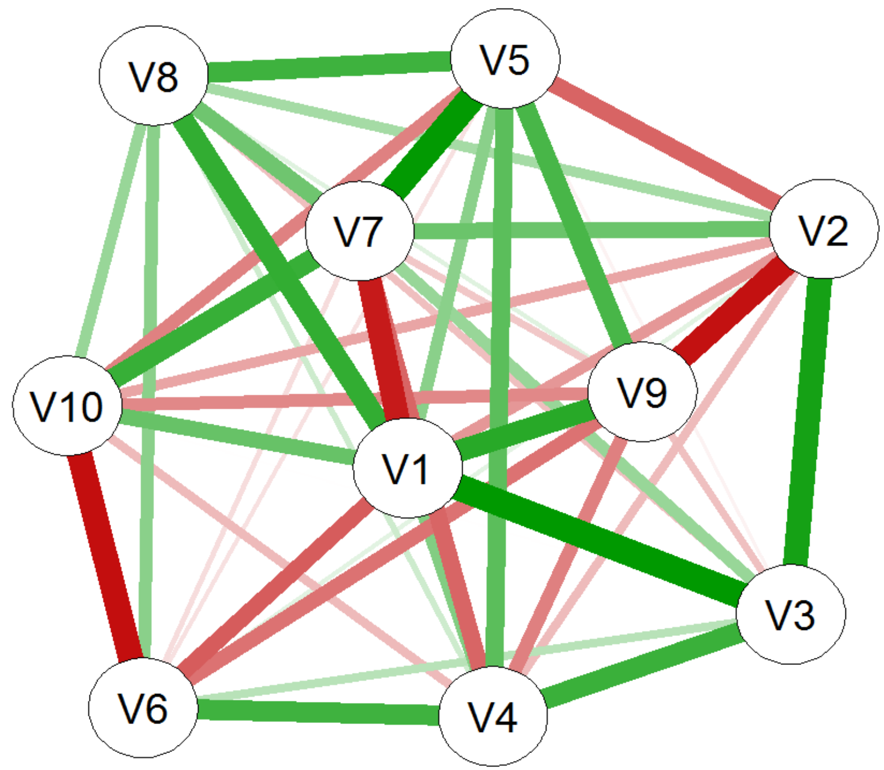
Correlations

Test 2: pcor increase-test

One-Factor model



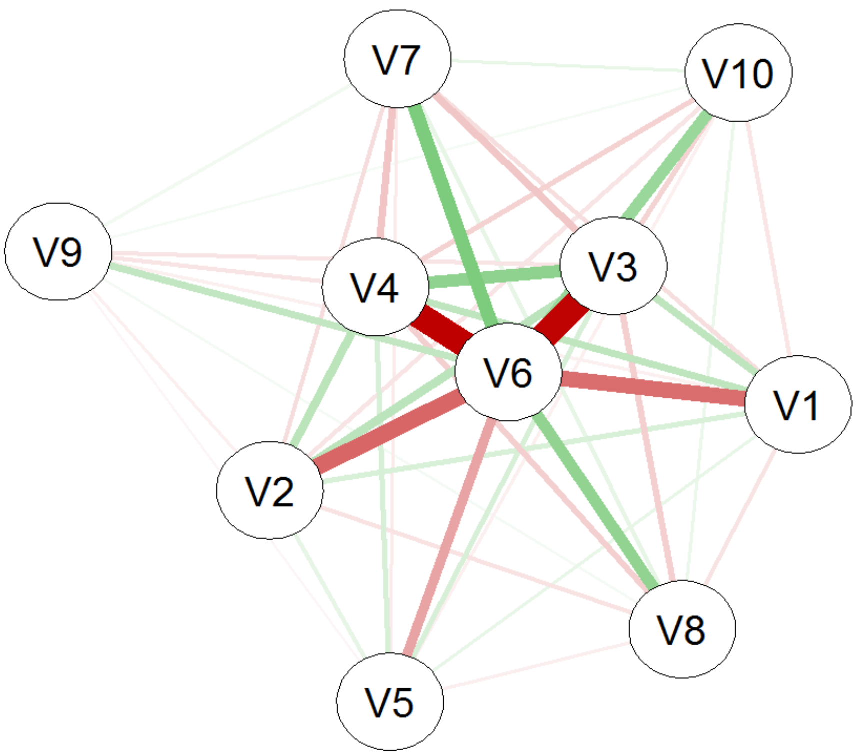
Network model



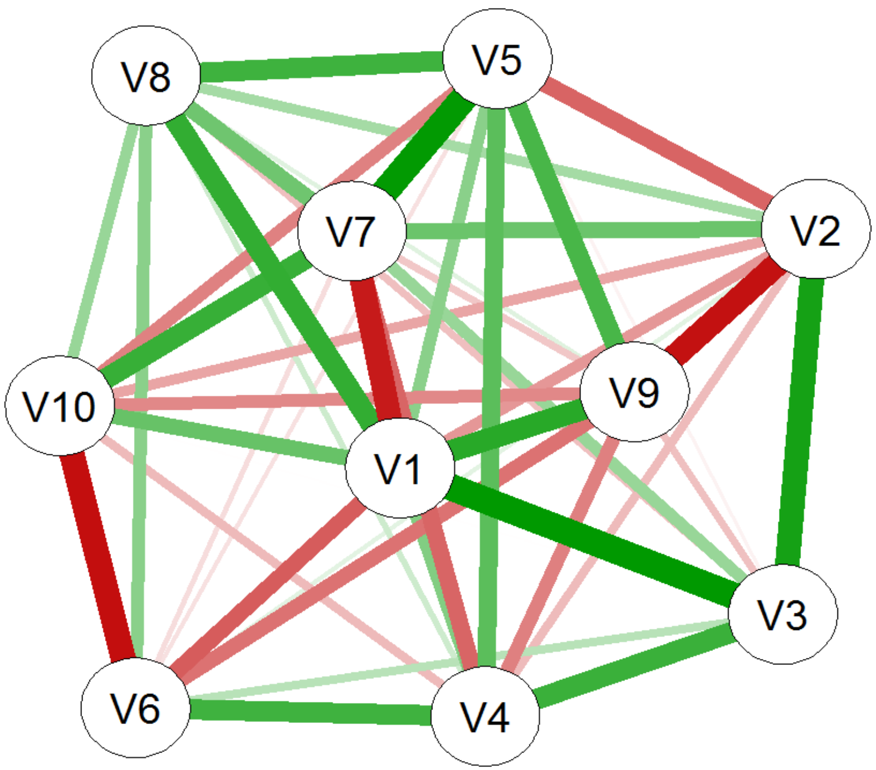
Partial Correlations

Test 2: pcor increase-test

One-Factor model : 0



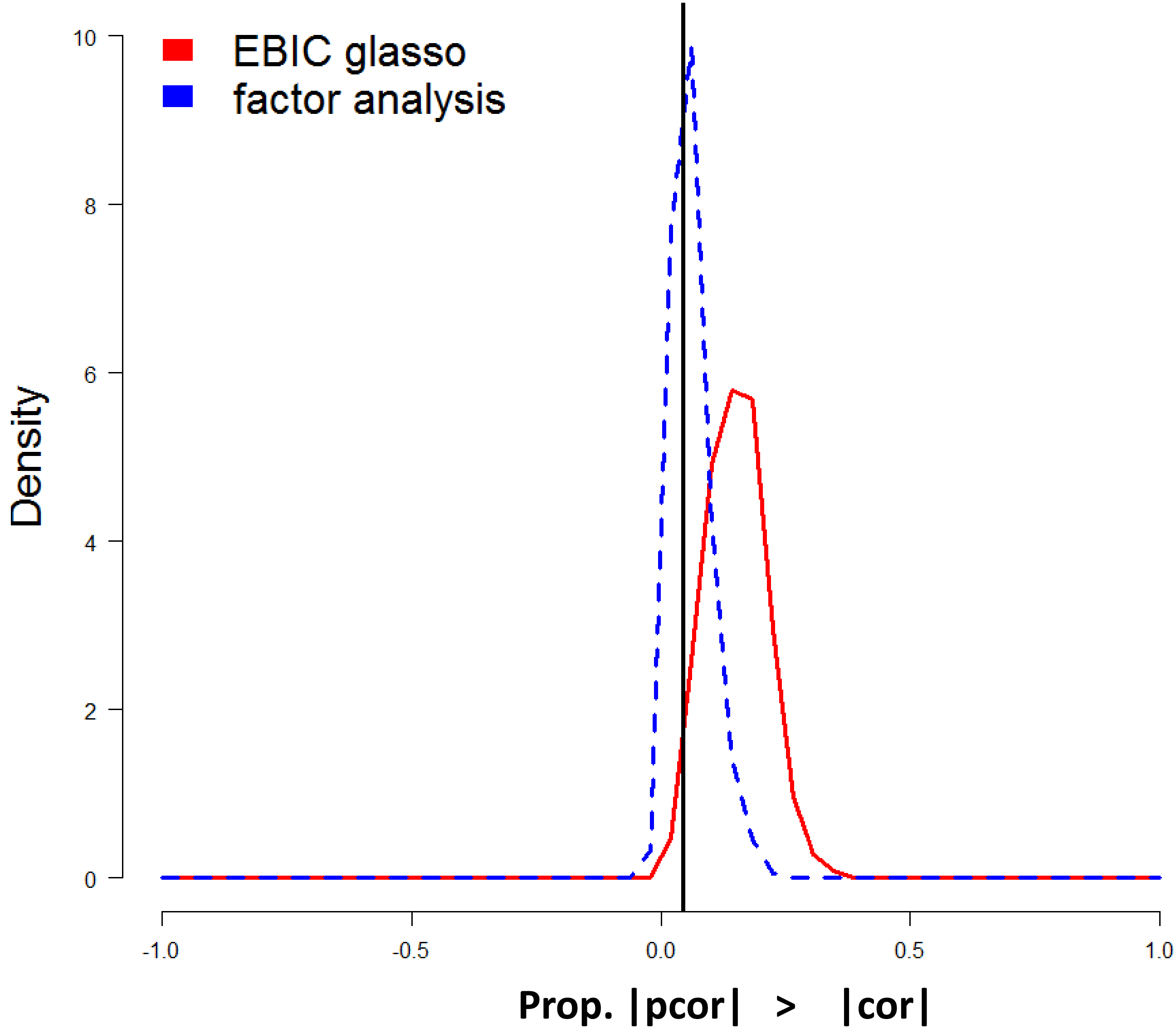
Network model: 17



Partial Correlations

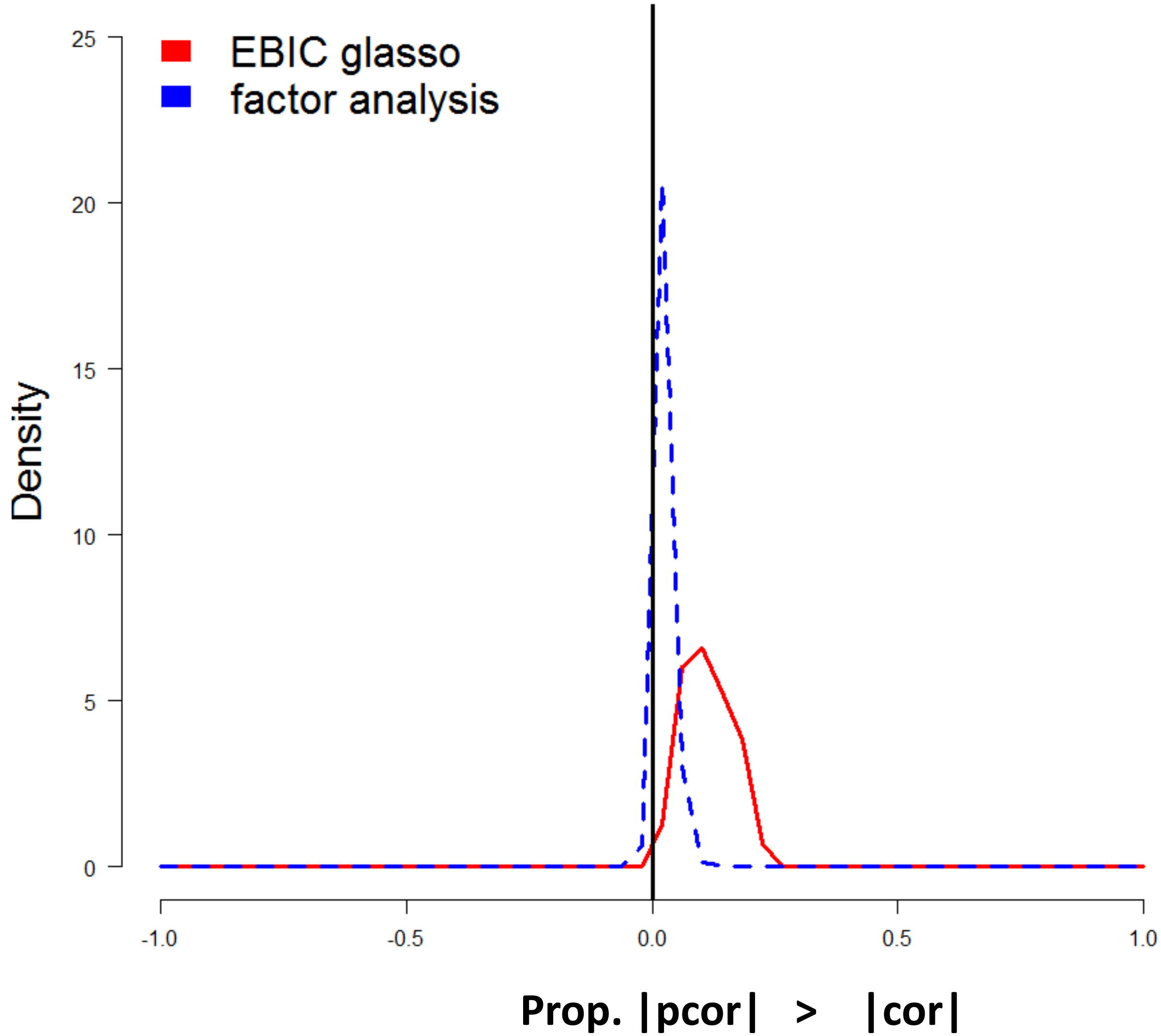
Results 2 : Factor model

- True model = Factor model
- N= **1000**
- $\lambda \sim U(.1\ ,.9)$
- 10 manifest variables



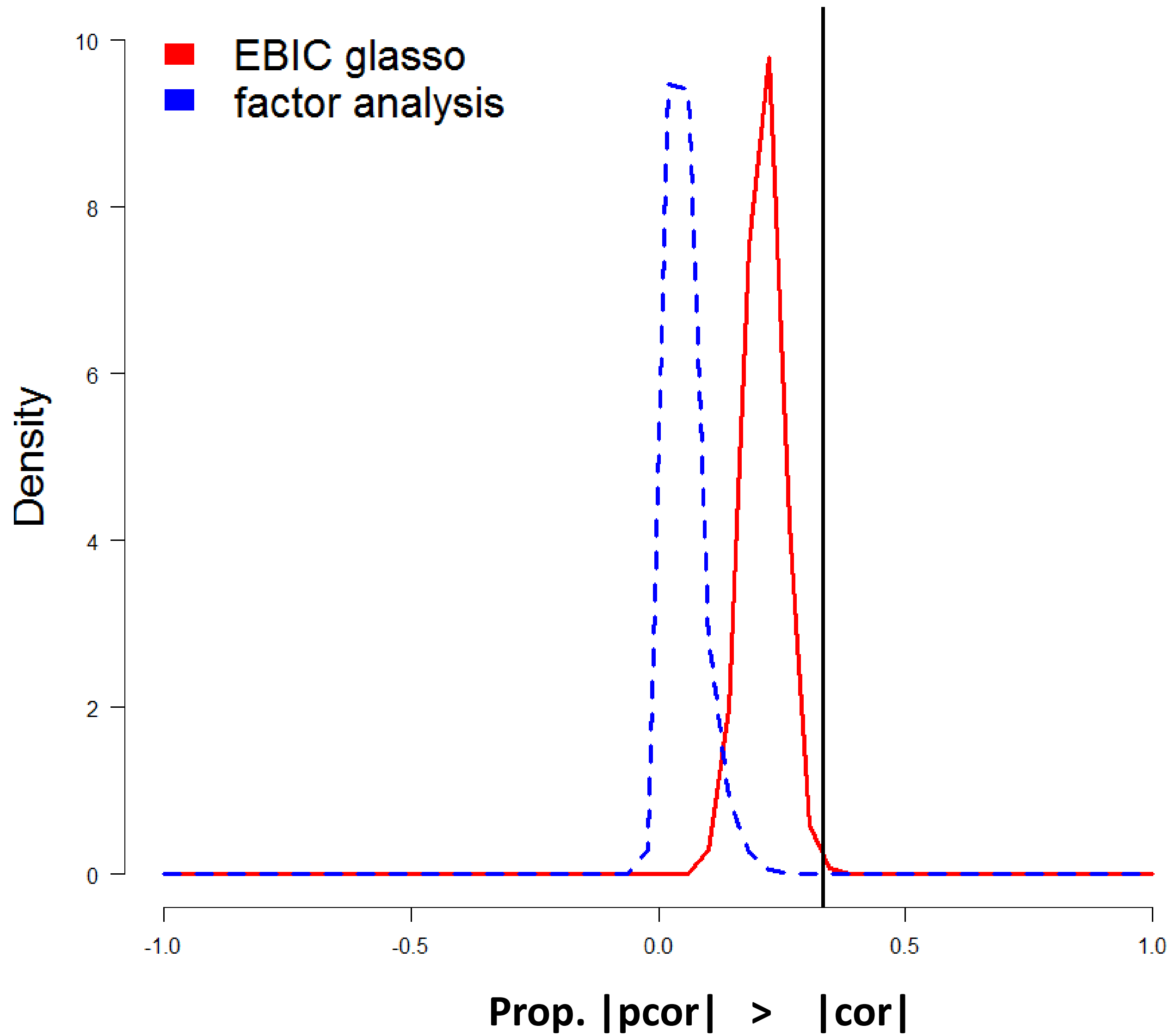
Results 2 : Factor model

- True model = Factor model
- N= **10.000**
- $\lambda \sim U(.1\ ,.9)$
- 10 manifest variables



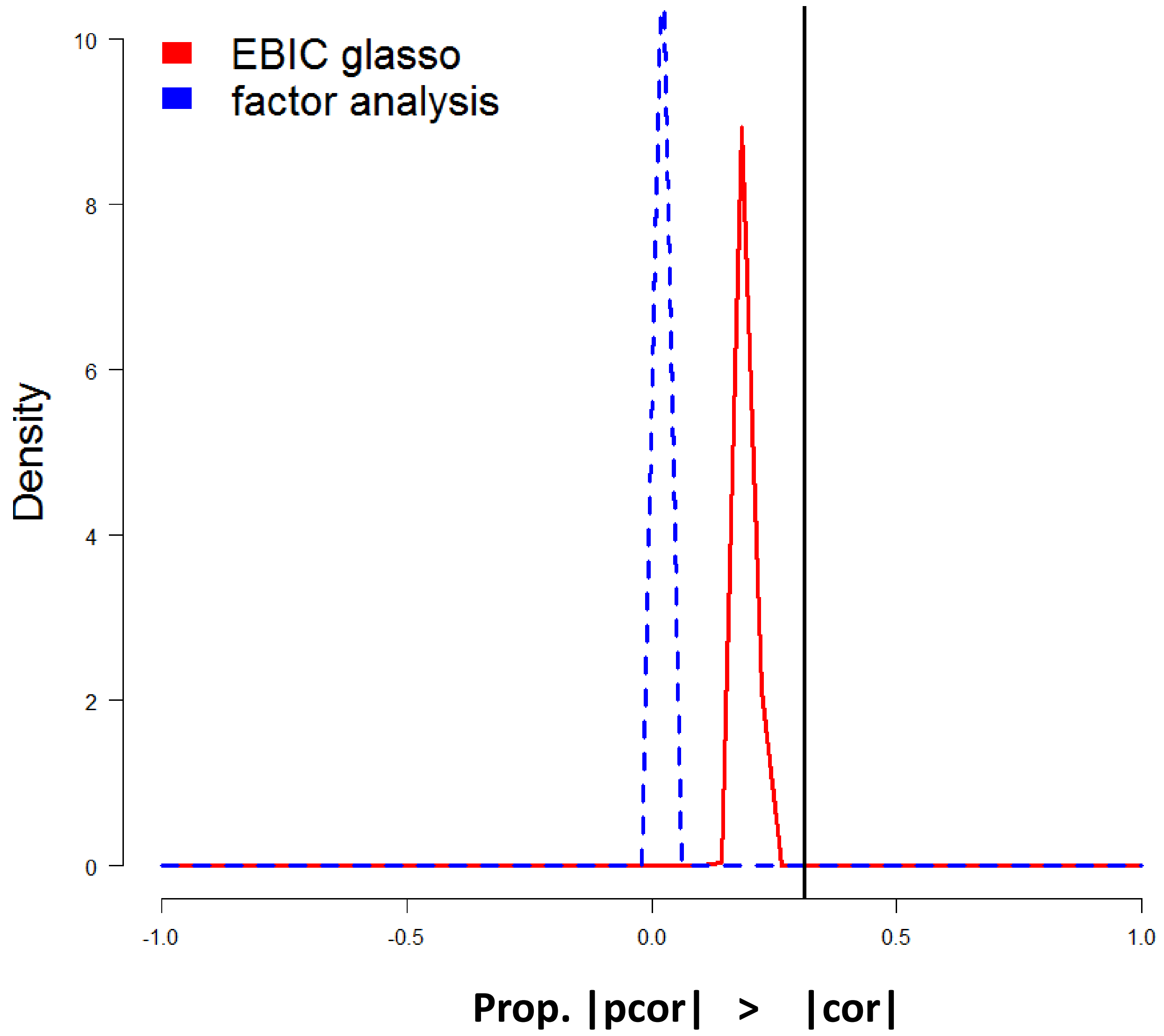
Results 2 : Network model

- True model = Network model
- N= **1000**
- Proportion partial correlations zero in population = 0
- 10 manifest variables



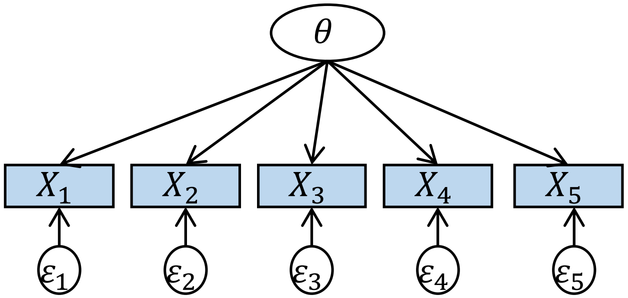
Results 2 : Network model

- True model = Network model
- N= **10.000**
- Proportion partial correlations zero in population = 0
- 10 manifest variables

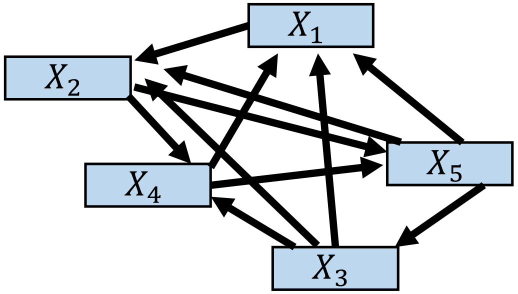


Results 1

How often does this test choose the right model?



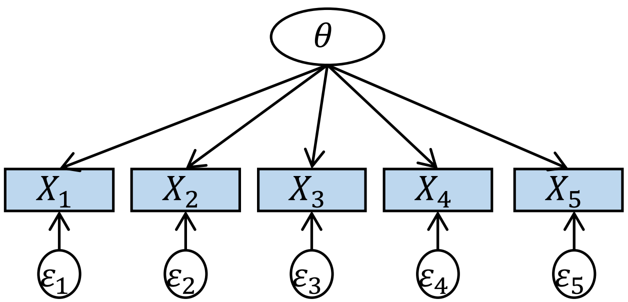
81.3% (N=1000)



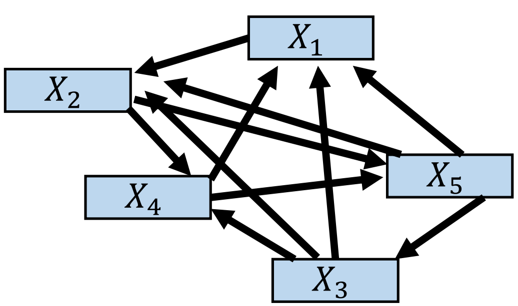
73.2% (N=1000)
12.4% no model

Results 1

How often does this test choose the right model?



81.3% (N=1000)

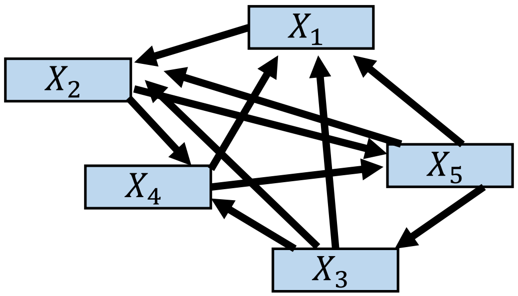


73.2% (N=1000)
12.4% no model

Lasso penalty is too strict for fully connected networks

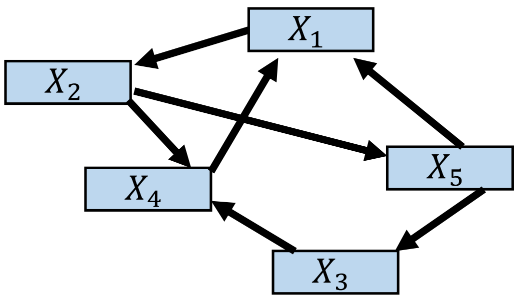
Results 1

How often does this test choose the right model?



73.2% (N=1000)
12.4% no model

Performance increases for sparse networks



88.4% (N=1000)
0.5% no model

Conclusion

Two tests that decide whether a certain property* of the sample partial correlation matrix has a higher probability density under a factor model or under a network model.

*

1. proportion partial correlations significant
2. proportion partial correlations stronger than the corresponding simple correlations

Test 1 distinguishes between *sparse* networks and factor models.
Test 2 distinguishes between networks and factor models that imply some negative correlations.

Questions?

Collaborators:

Mijke Rhemtulla
Denny Borsboom
Lourens J. Waldorp

Disclaimer

To distinguish between these models I assume that these models are not merely statistical models but *causal* models that generate the data and are therefore able to *explain* the correlational structure of the data.