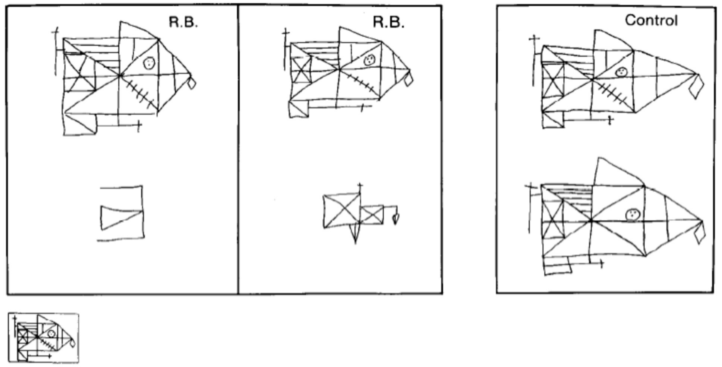
Special topics:
Latent variable models

April 4, 2022

What is a latent variable?



- *Observed variable* versus *latent variable*

- <u>Observed variables</u>: words correctly recalled out of a list of 20, number of features correctly reproduced on drawing of Rey-Osterreith figure from memory, score on Everyday Memory Test
  - anything that can be measured directly

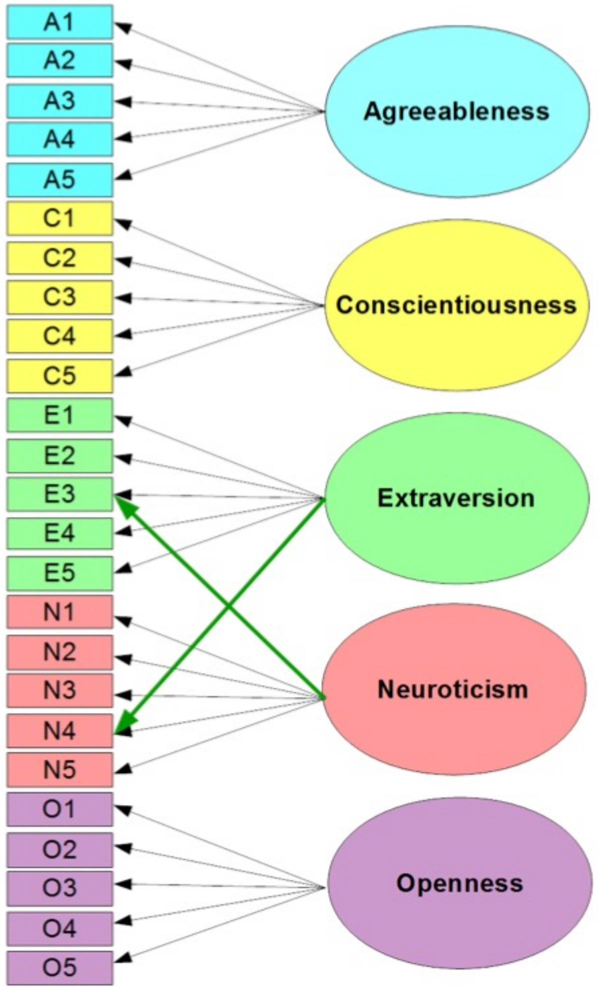
- <u>Latent variables</u>: "episodic memory"
  - anxiety, conscientiousness, creativity, ...
  - Hypothetical construct that cannot be measured directly but whose presence is inferred based on ***patterns of correlation*** between observed variables

What is a latent variable?

- *Observed variable* versus *latent variable*

- Observed variables: words correctly recalled out of a list of 20, number of features correctly reproduced on drawing of Rey-Osterreith figure from memory, score on Everyday Memory Test
  - anything that can be measured directly

- Latent variables: "episodic memory"
  - anxiety, conscientiousness, creativity, ...
  - Hypothetical construct that cannot be measured directly but whose presence is inferred based on ***patterns of correlation*** between observed variables



Latent variable models

- Further extension of general linear model

- Essentially: allows you to do regression analyses relating *latent variables* to one another

- Measurement model: how *indicators* relate to *latent variables*
- Latent variable (structural) model: how *latent variables* relate to one another

- "LISREL" models (**LI**near **S**tructural **REL**ations)

Path diagrams

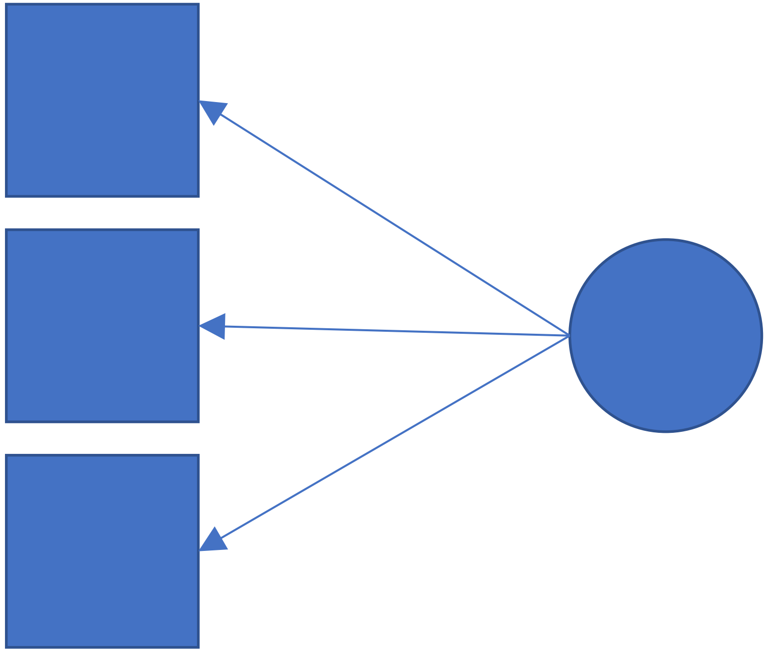
observed/manifest variable
"indicator"

latent variable "factor"

Path diagrams

observed/manifest variable
"indicator"

latent variable "factor"



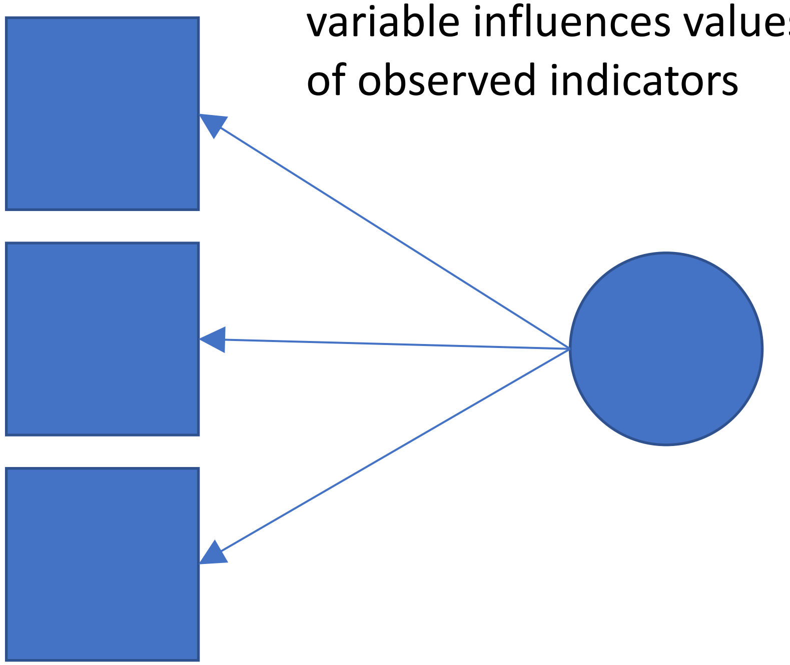
Path diagrams

observed/manifest variable
"indicator"

latent variable "factor"

direction of arrows
implies influence

value of unobserved latent
variable influences values
of observed indicators



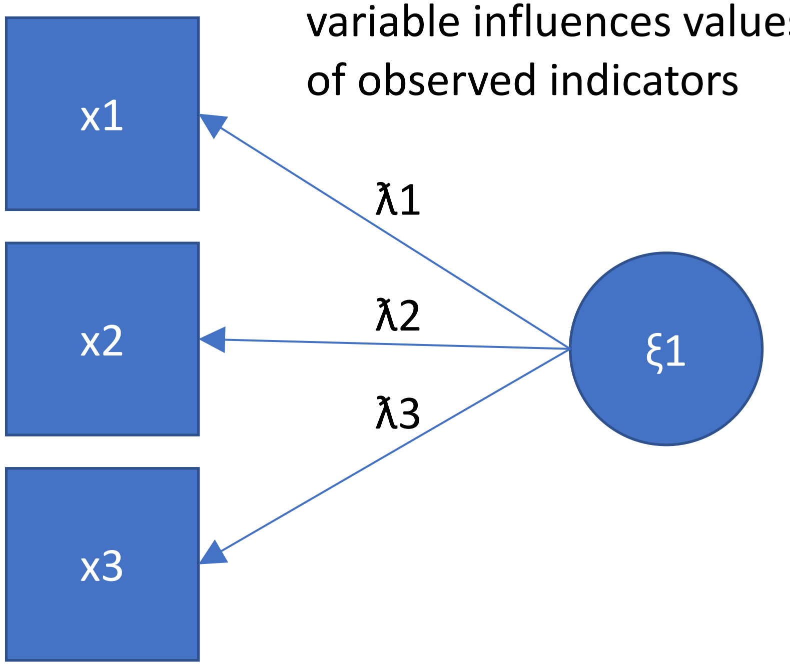
Path diagrams

observed/manifest variable
"indicator"

latent variable "factor"

direction of arrows
implies influence

value of unobserved latent
variable influences values
of observed indicators



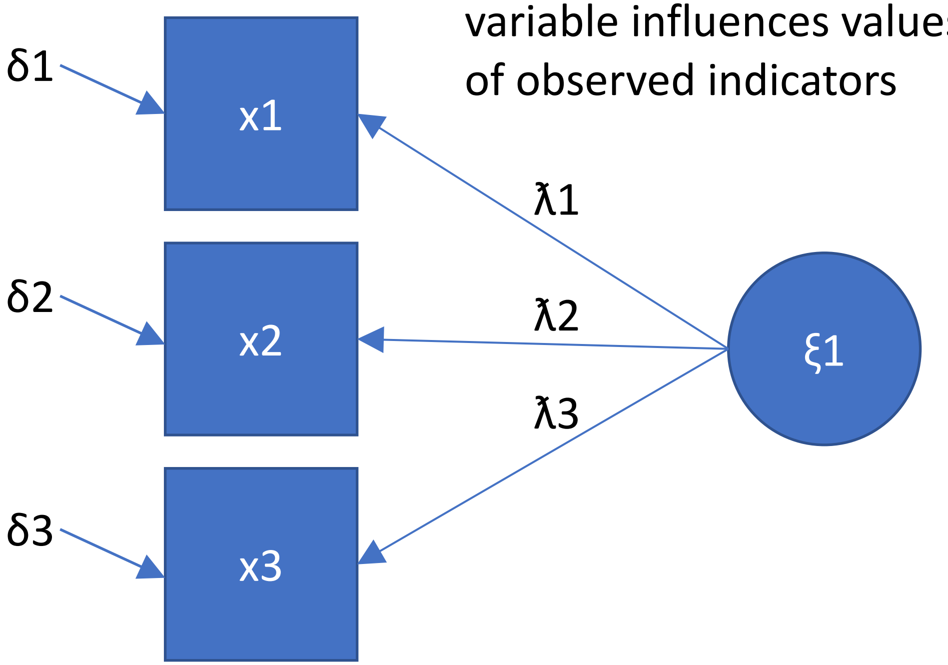
Path diagrams

direction of arrows implies influence

value of unobserved latent variable influences values of observed indicators

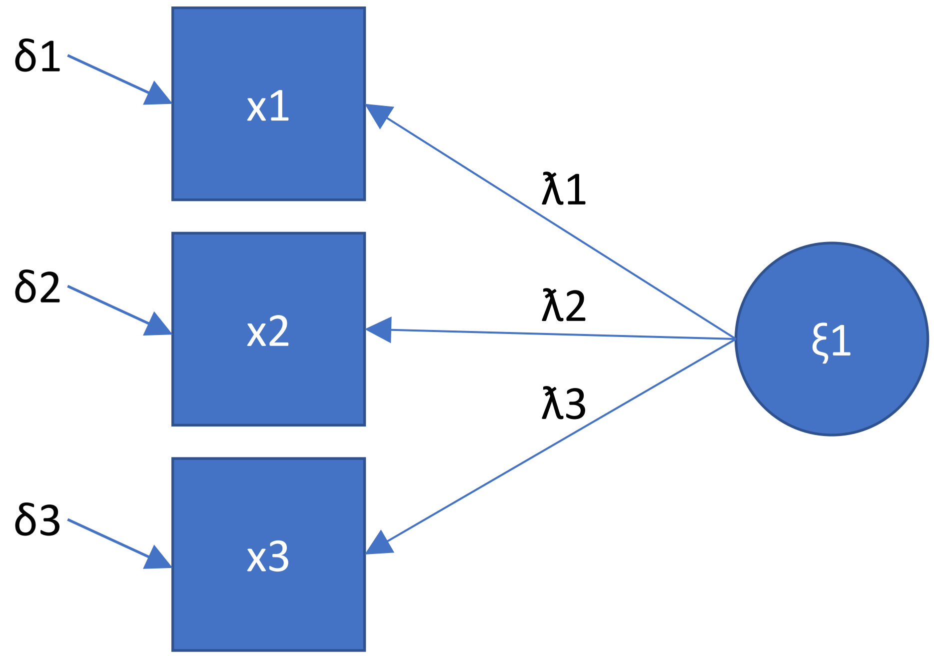
observed/manifest variable "indicator"

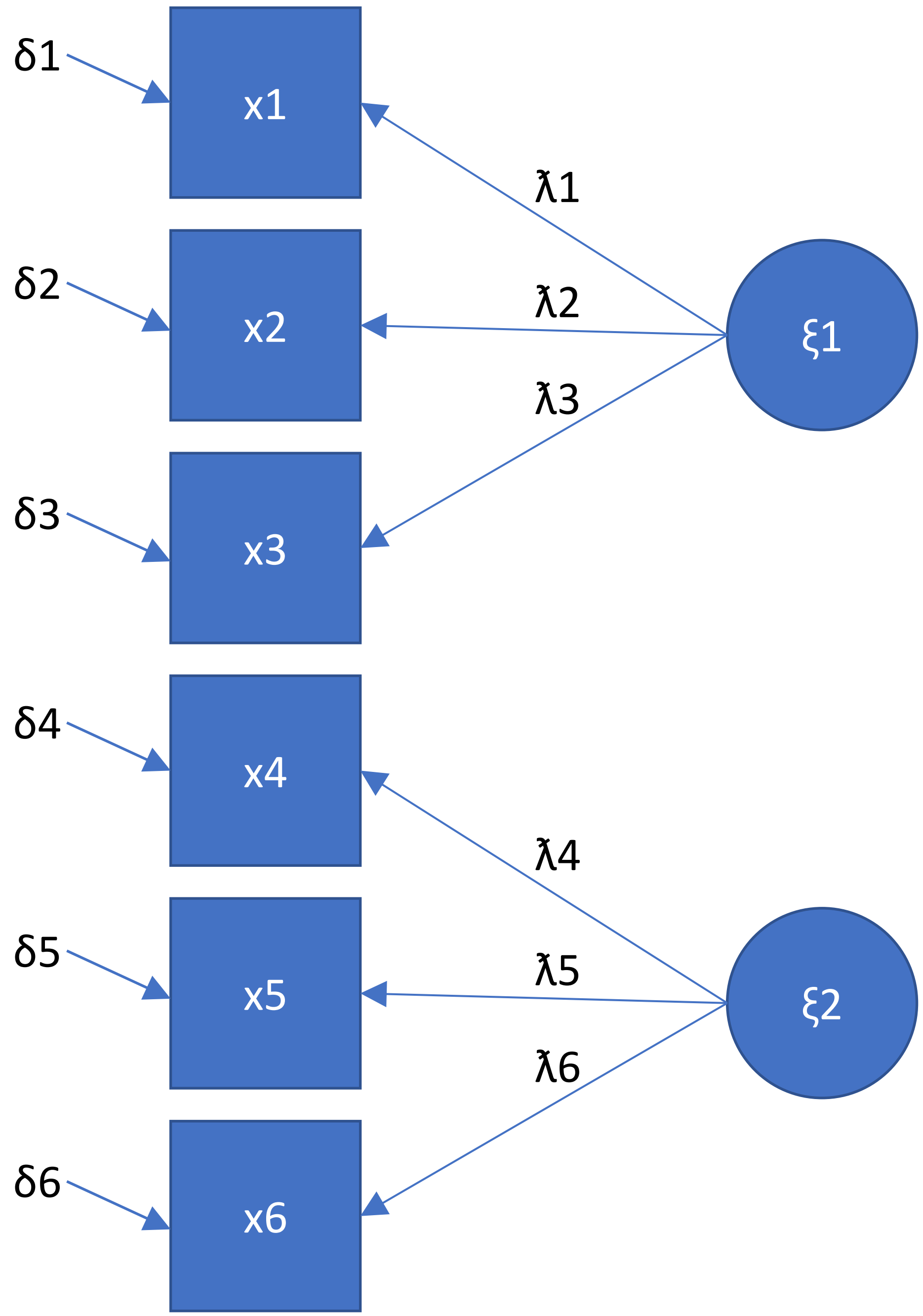
latent variable "factor"



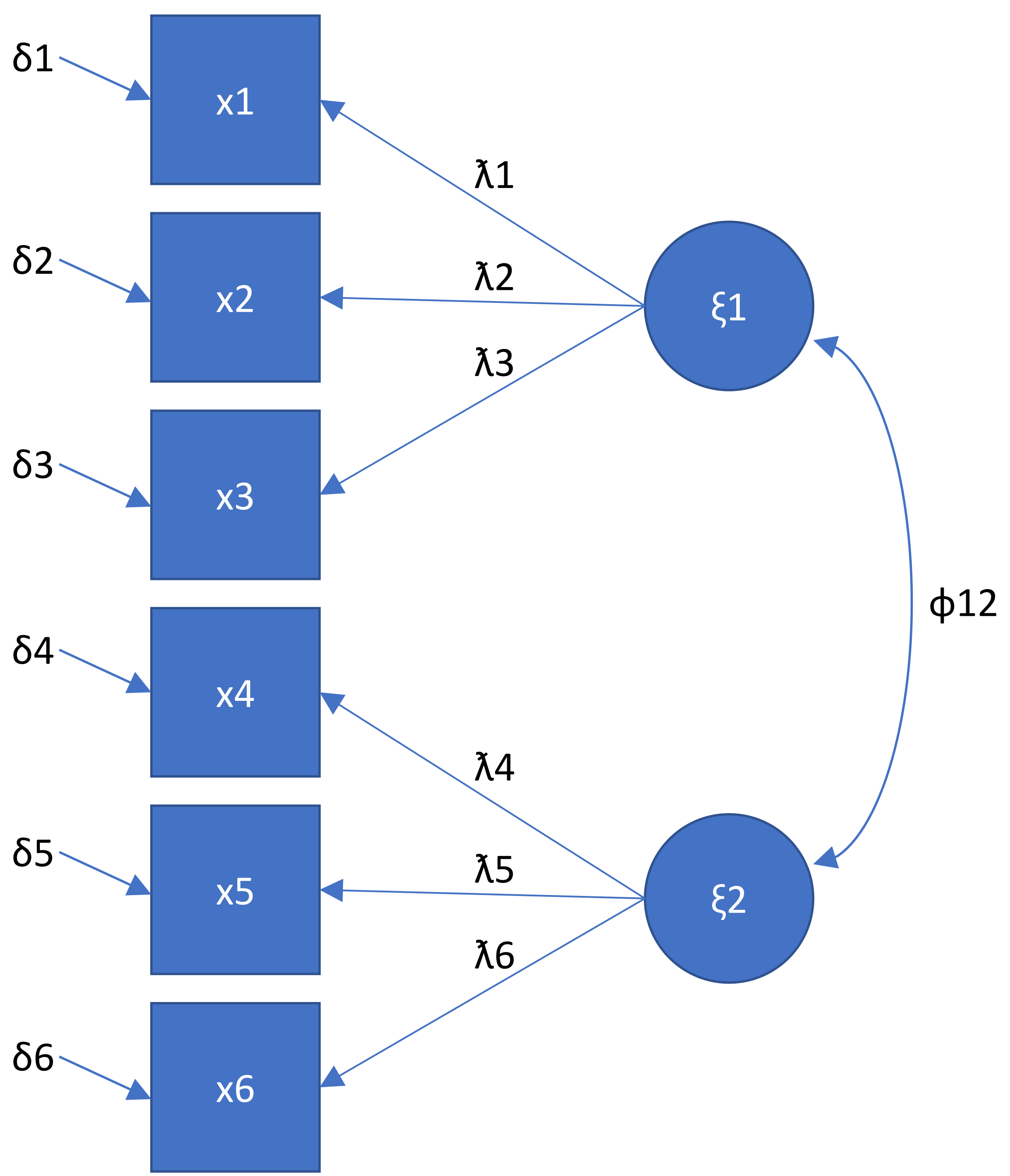
measurement error in indicators: only modeling covariance among indicators that represents relationship to latent variable

(factor analysis vs component analysis)


δ1
δ2
δ3
x1
x2
x3
λ1
λ2
λ3
ξ1

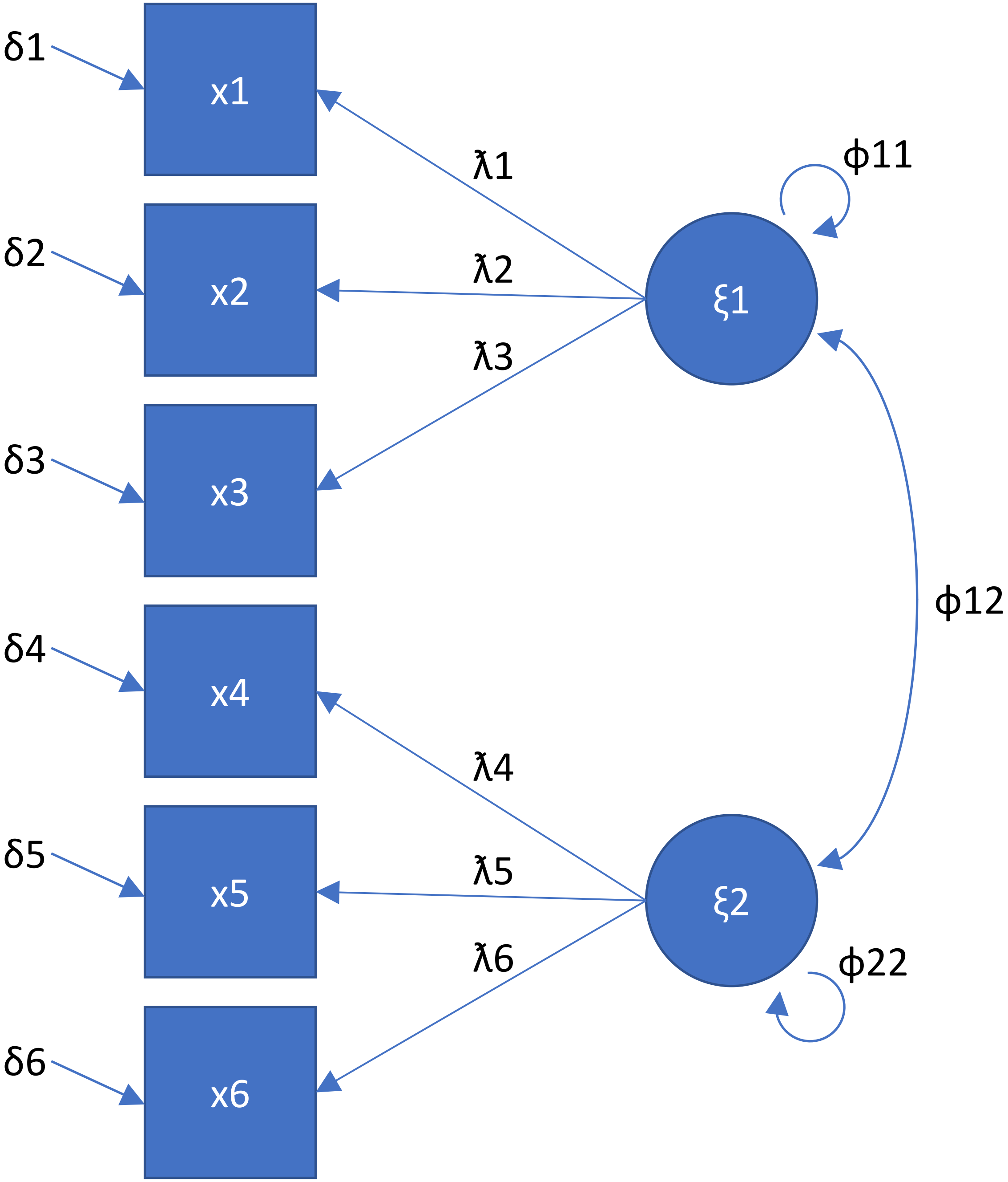


δ = residuals "disturbances"
x = observed variables
ƛ = loadings
ξ = "exogenous" latent variables – not influenced by other latent variables in model / "x" side of structural model



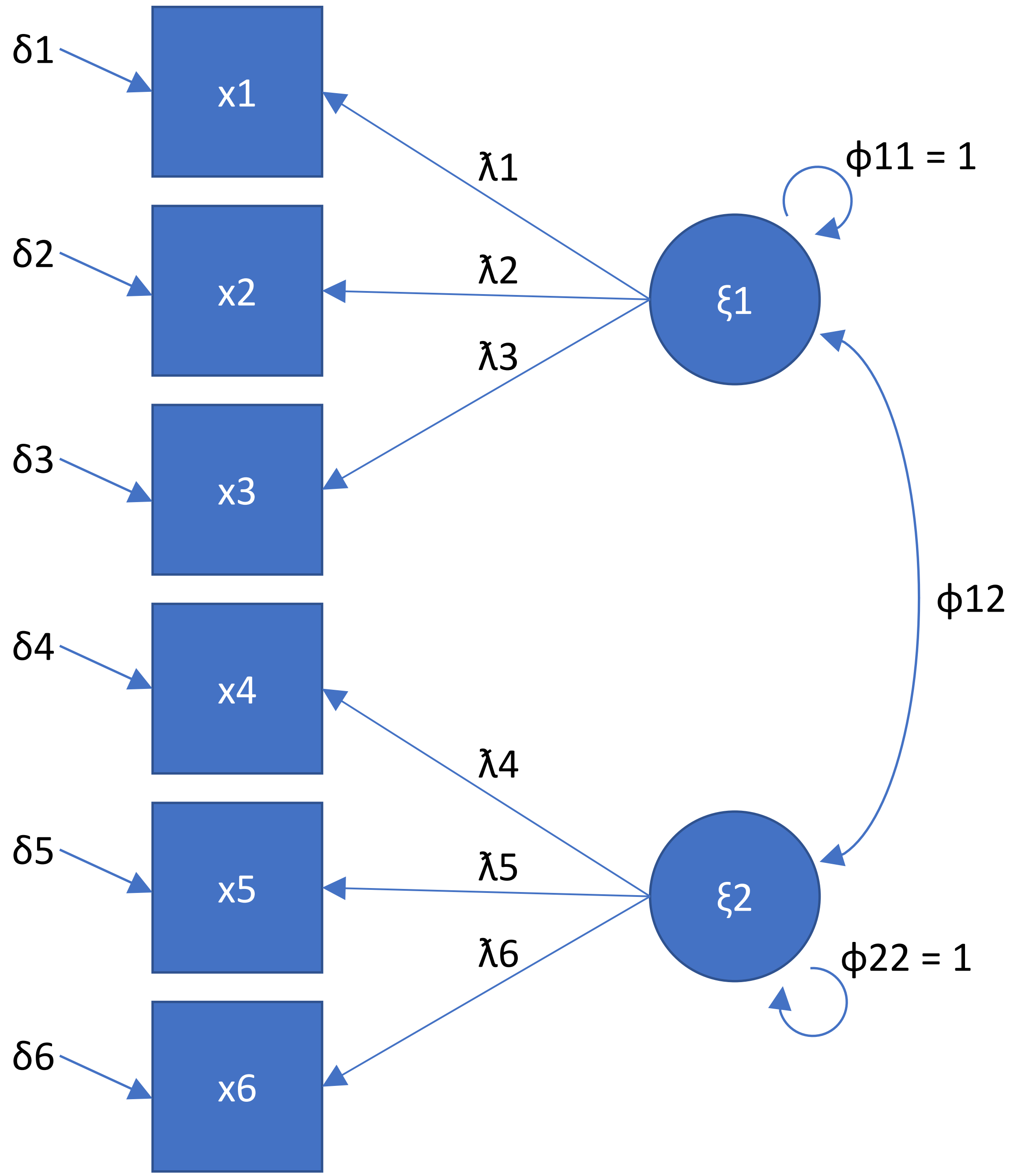
δ = residuals "disturbances"
x = observed variables
λ = loadings
ξ = "exogenous" latent variables – not influenced by other latent variables in model / "x" side of structural model

ϕ = covariance among exogenous latent variables



δ = residuals "disturbances"
x = observed variables
ƛ = loadings
ξ = "exogenous" latent variables – not influenced by other latent variables in model / "x" side of structural model

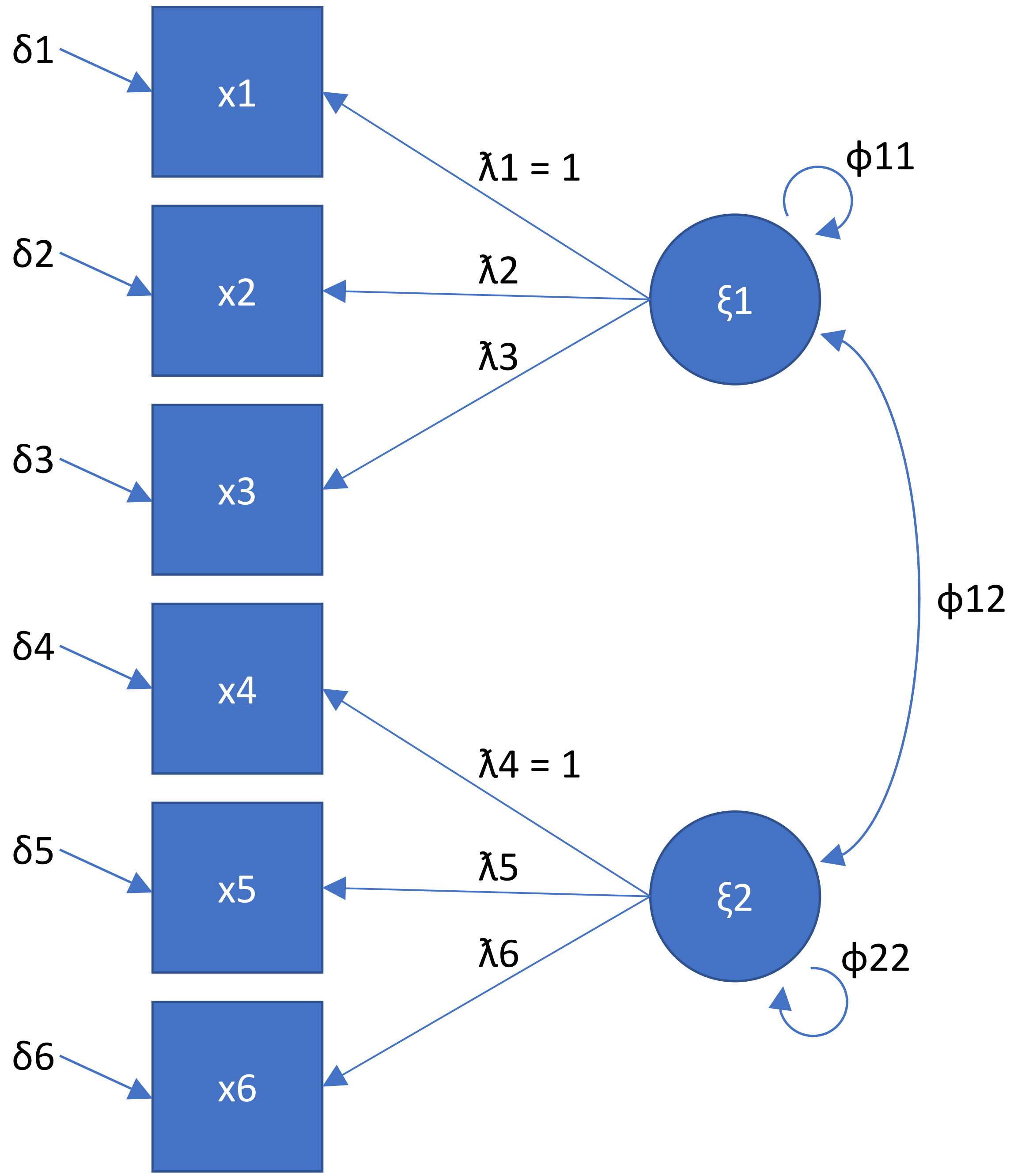
ϕ = covariance among exogenous latent variables



δ = residuals "disturbances"
x = observed variables
ƛ = loadings
ξ = "exogenous" latent variables – not influenced by other latent variables in model / "x" side of structural model

ϕ = covariance among exogenous latent variables

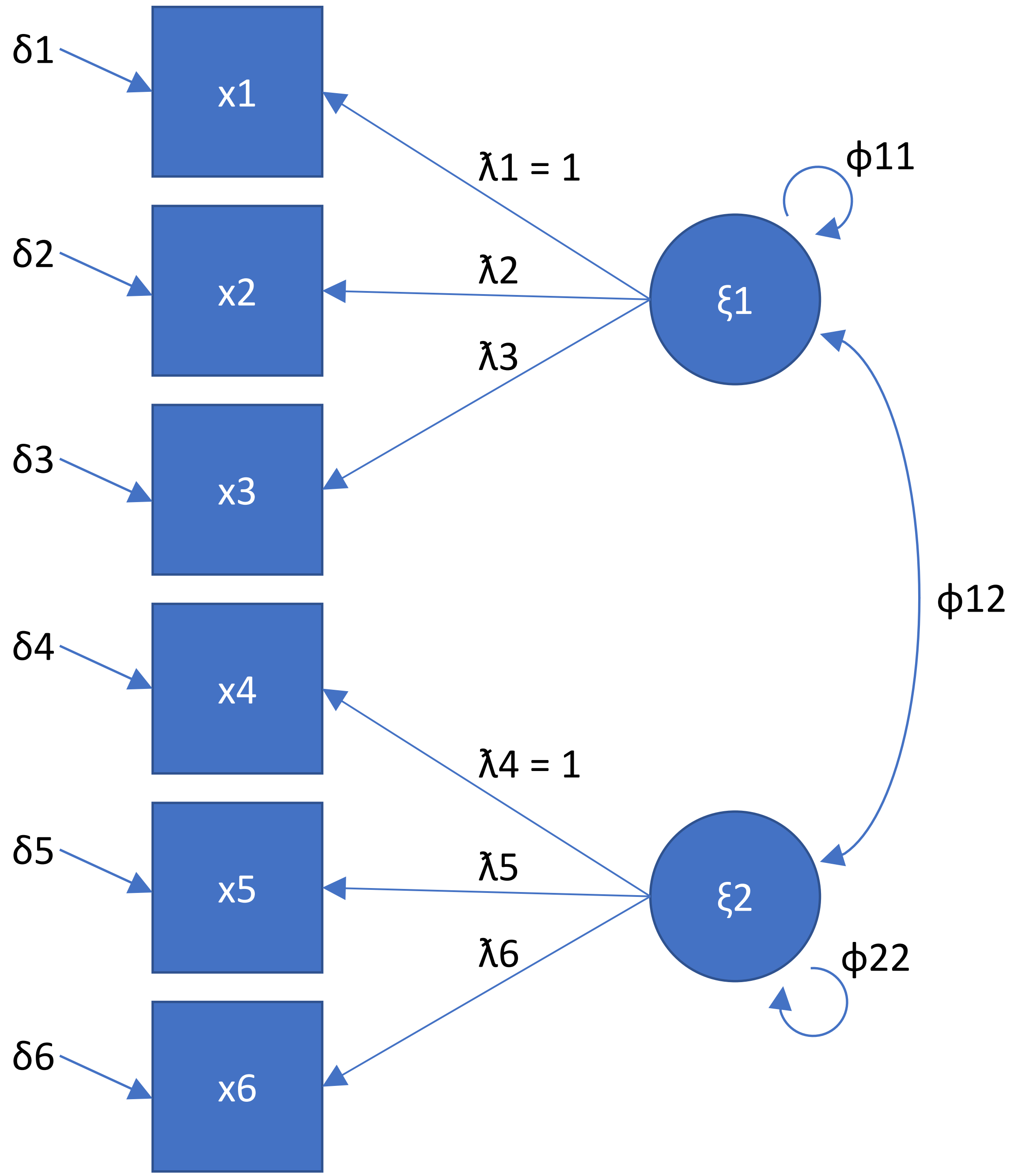
often fix variance of latent variables to 1 to set the scale; alternative is to fix one loading for each latent variable to 1



δ = residuals "disturbances"
x = observed variables
ƛ = loadings
ξ = "exogenous" latent variables – not influenced by other latent variables in model / "x" side of structural model

ϕ = covariance among exogenous latent variables

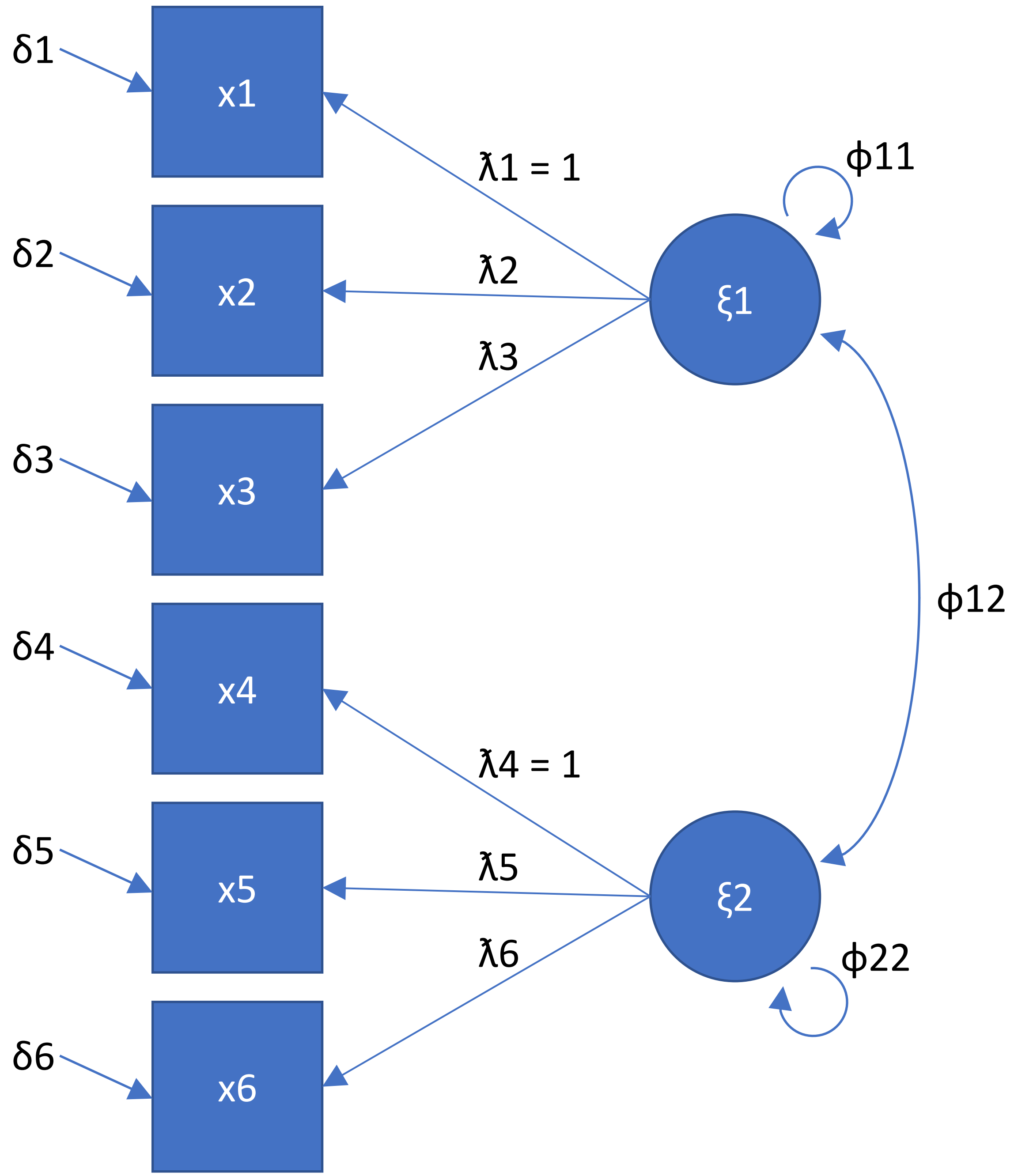
often fix variance of latent variables to 1 to set the scale; alternative is to fix one loading for each latent variable to 1



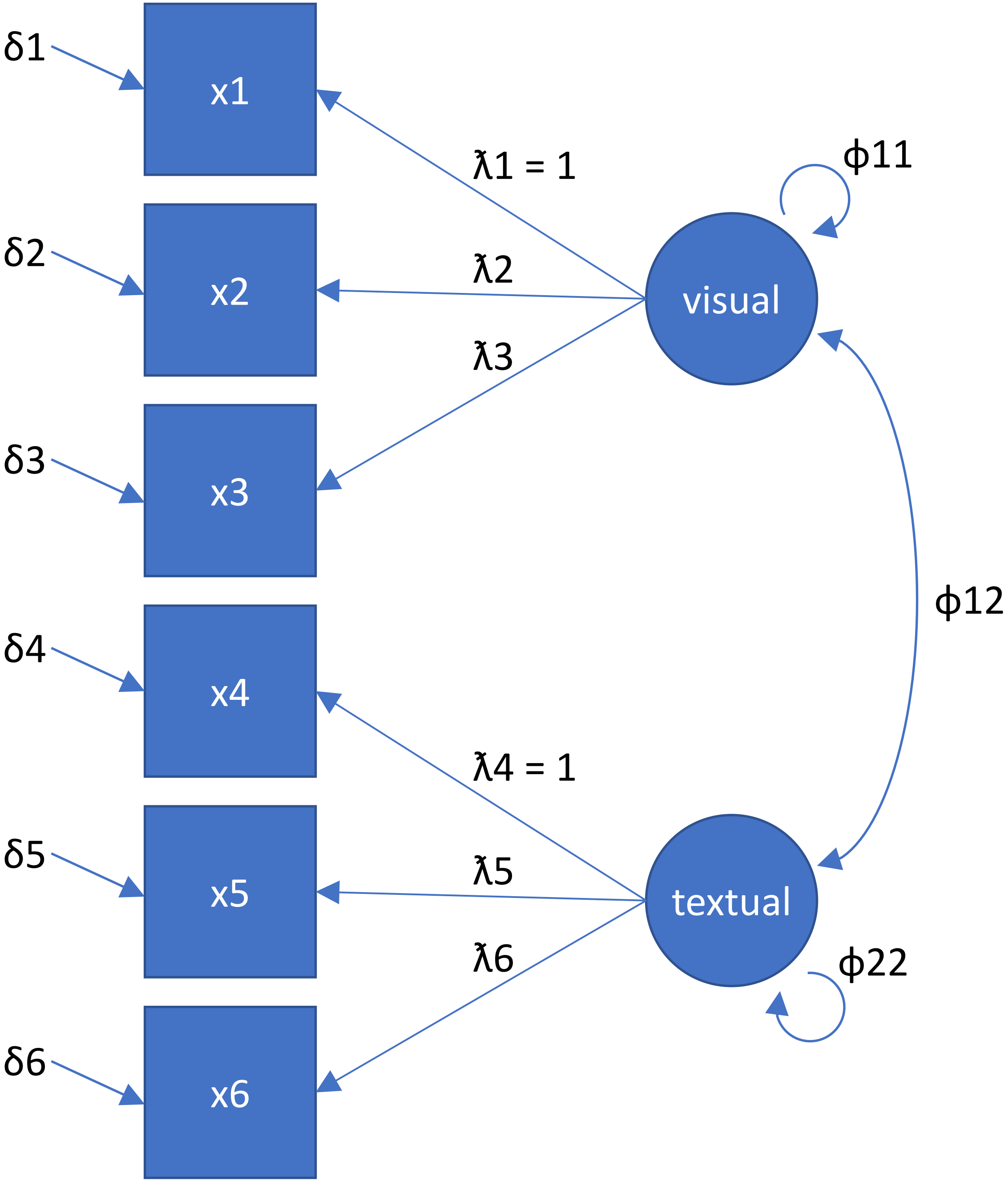
- This type of model is often referred to as a confirmatory factor analysis (CFA) model
- You have decided ahead of time which indicator variables are associated with distinct factors
- This is different from exploratory factor analysis (EFA) where you are trying to detect the factor structure from the data
- Cross-validation: cannot do EFA and CFA on the same data. If you do not know structure, split data set first

Confirmatory factor analysis example

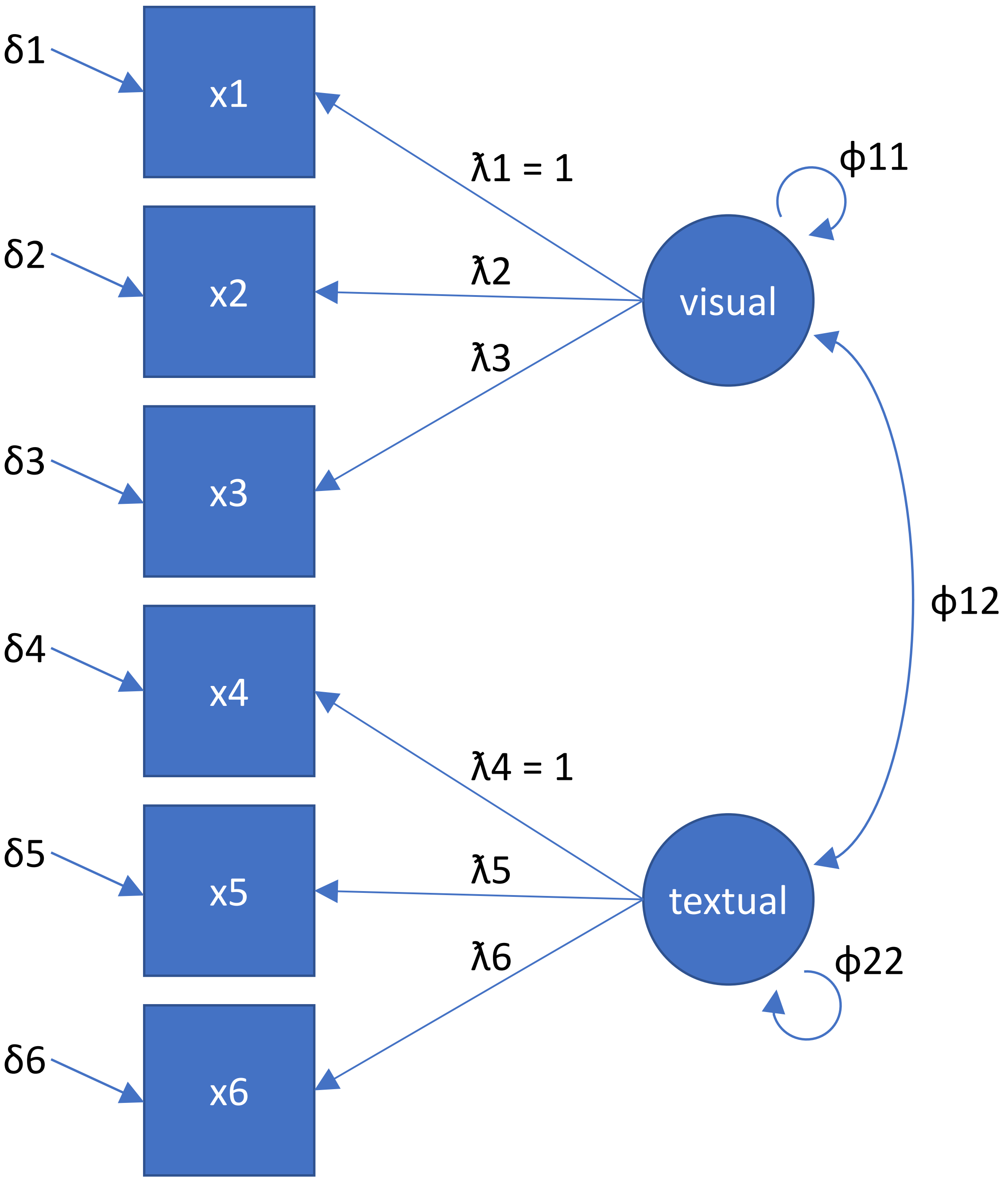
- Holzinger-Swineford (1939) data included with `lavaan` package
- Mental ability test scores of seventh- and eighth-grade children from two different schools
- Example data set includes scores on 9 different tests that should load onto 3 different factors
- **visual**: `x1` ("visual perception"), `x2` ("cubes"), `x3` ("lozenges")
- **textual**: `x4` ("paragraph comprehension"), `x5` ("sentence completion"), x6 ("word meaning")
- **speed**: `x7` (addition), `x8` (counting of dots), `x9` (discrimination of straight and curved capitals)



20220404_SEM_demo.R

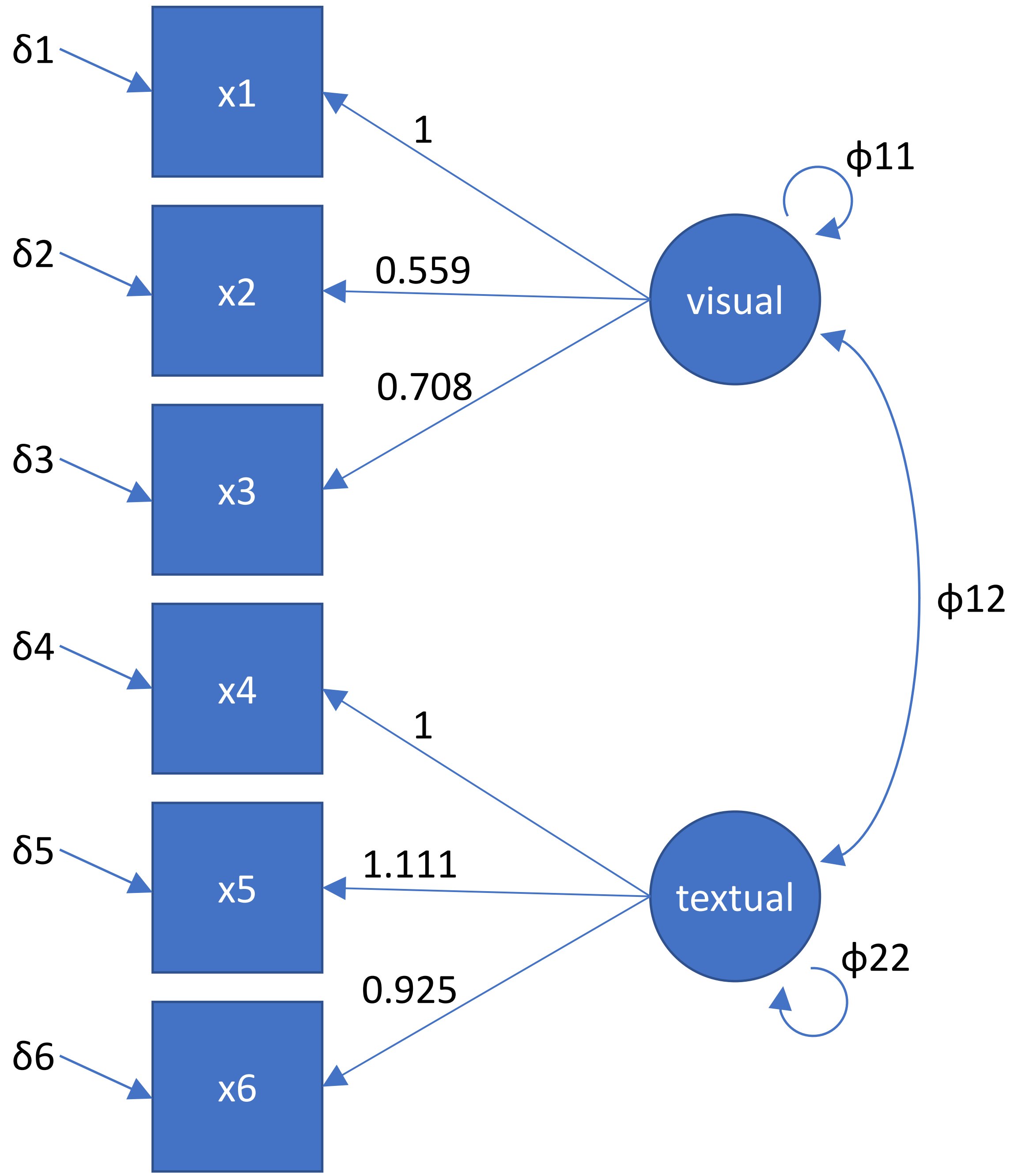



```r
## specify the model
x1to6_model <- ' visual =~ x1 + x2 + x3
                 textual =~ x4 + x5 + x6'
# =~ "is measured by"
# latent variable analog of y ~ x1 + x2 + x3 in linear model

## confirmatory factor analysis: fit the model
fit1 <- cfa(model = x1to6_model, data = hs1939)

## display summary output
summary(fit1, fit.measures = TRUE)
```



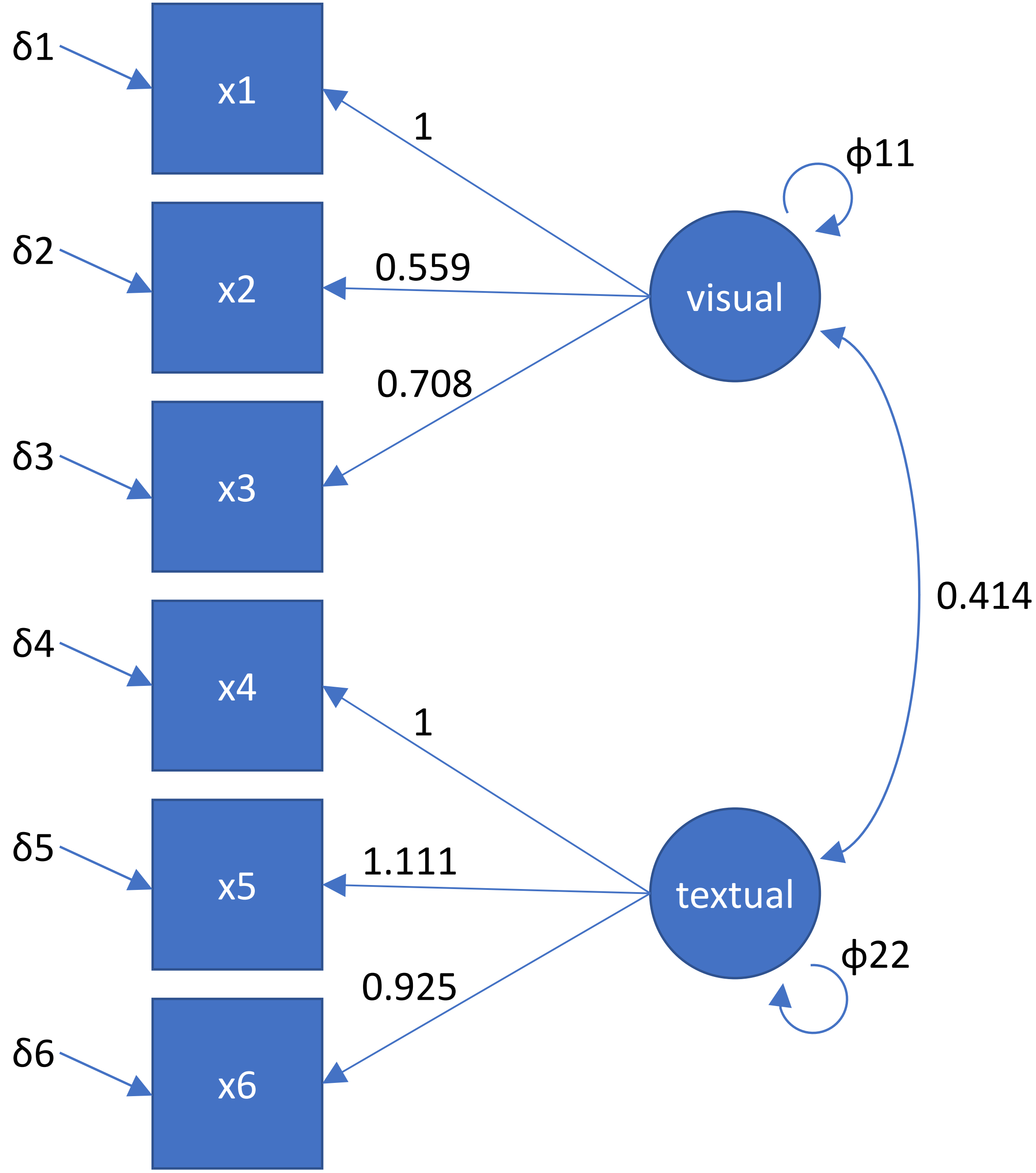

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  visual =~
    x1                1.000
    x2                0.559    0.105    5.301    0.000
    x3                0.708    0.118    6.004    0.000
  textual =~
    x4                1.000
    x5                1.111    0.065   16.996    0.000
    x6                0.925    0.055   16.703    0.000

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  visual ~~
    textual           0.414    0.074    5.562    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .x1                0.536    0.129    4.155    0.000
   .x2                1.125    0.103   10.965    0.000
   .x3                0.863    0.095    9.085    0.000
   .x4                0.369    0.048    7.735    0.000
   .x5                0.449    0.059    7.662    0.000
   .x6                0.356    0.043    8.263    0.000
    visual            0.822    0.158    5.188    0.000
    textual           0.981    0.112    8.745    0.000
```

20220404_SEM_demo.R



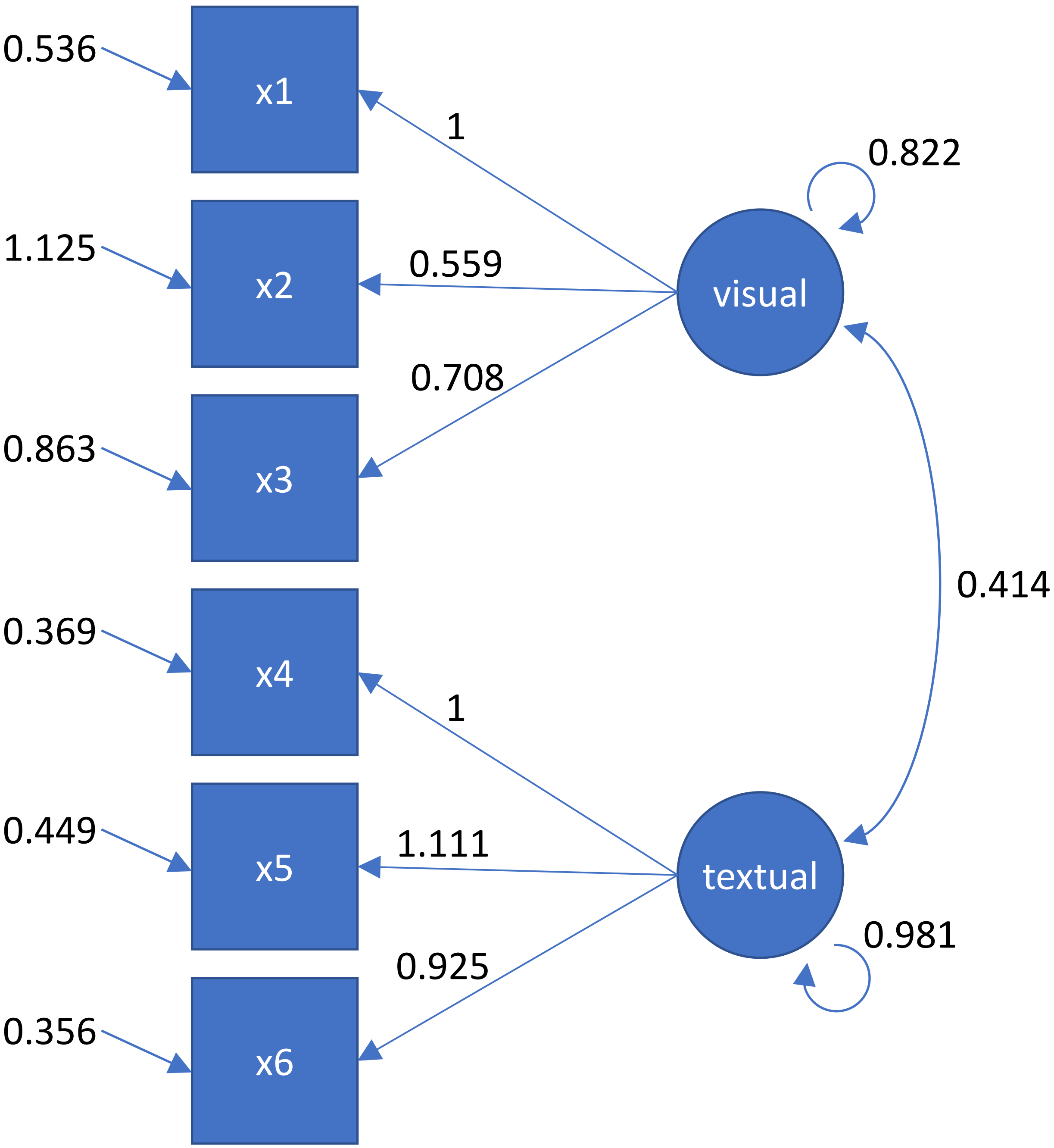

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  visual =~
    x1                1.000
    x2                0.559    0.105    5.301    0.000
    x3                0.708    0.118    6.004    0.000
  textual =~
    x4                1.000
    x5                1.111    0.065   16.996    0.000
    x6                0.925    0.055   16.703    0.000

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  visual ~~
    textual           0.414    0.074    5.562    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .x1                0.536    0.129    4.155    0.000
   .x2                1.125    0.103   10.965    0.000
   .x3                0.863    0.095    9.085    0.000
   .x4                0.369    0.048    7.735    0.000
   .x5                0.449    0.059    7.662    0.000
   .x6                0.356    0.043    8.263    0.000
    visual            0.822    0.158    5.188    0.000
    textual           0.981    0.112    8.745    0.000
```

20220404_SEM_demo.R



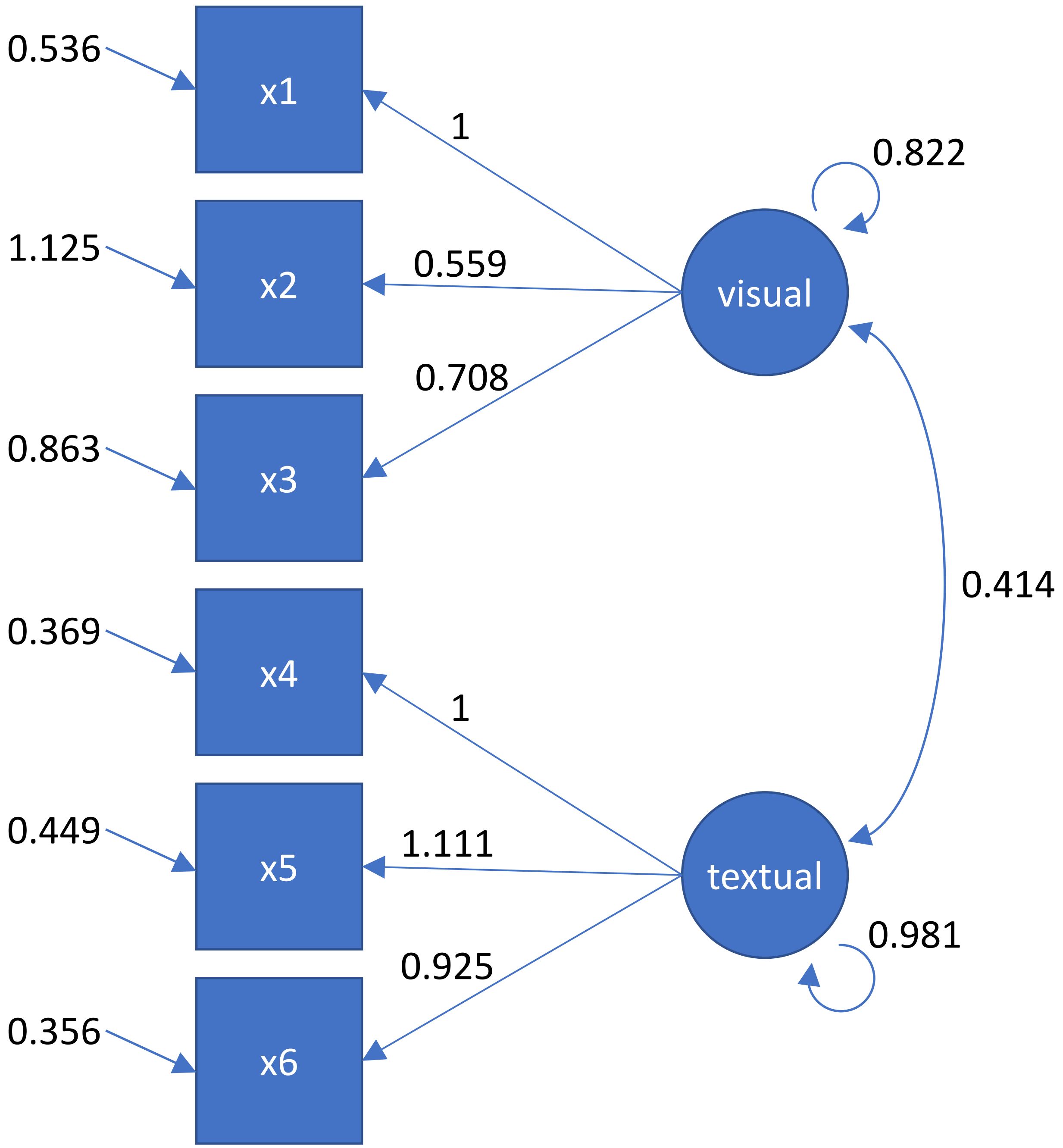

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  visual =~
    x1                1.000
    x2                0.559    0.105    5.301    0.000
    x3                0.708    0.118    6.004    0.000
  textual =~
    x4                1.000
    x5                1.111    0.065   16.996    0.000
    x6                0.925    0.055   16.703    0.000

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  visual ~~
    textual           0.414    0.074    5.562    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .x1                0.536    0.129    4.155    0.000
   .x2                1.125    0.103   10.965    0.000
   .x3                0.863    0.095    9.085    0.000
   .x4                0.369    0.048    7.735    0.000
   .x5                0.449    0.059    7.662    0.000
   .x6                0.356    0.043    8.263    0.000
    visual            0.822    0.158    5.188    0.000
    textual           0.981    0.112    8.745    0.000
```

20220404_SEM_demo.R




```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)
  visual =~
    x1                1.000
    x2                0.559    0.105    5.301    0.000
    x3                0.708    0.118    6.004    0.000
  textual =~
    x4                1.000
    x5                1.111    0.065   16.996    0.000
    x6                0.925    0.055   16.703    0.000

Covariances:
                   Estimate  Std.Err  z-value  P(>|z|)
  visual ~~
    textual           0.414    0.074    5.562    0.000

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)
   .x1                0.536    0.129    4.155    0.000
   .x2                1.125    0.103   10.965    0.000
   .x3                0.863    0.095    9.085    0.000
   .x4                0.369    0.048    7.735    0.000
   .x5                0.449    0.059    7.662    0.000
   .x6                0.356    0.043    8.263    0.000
    visual            0.822    0.158    5.188    0.000
    textual           0.981    0.112    8.745    0.000
```

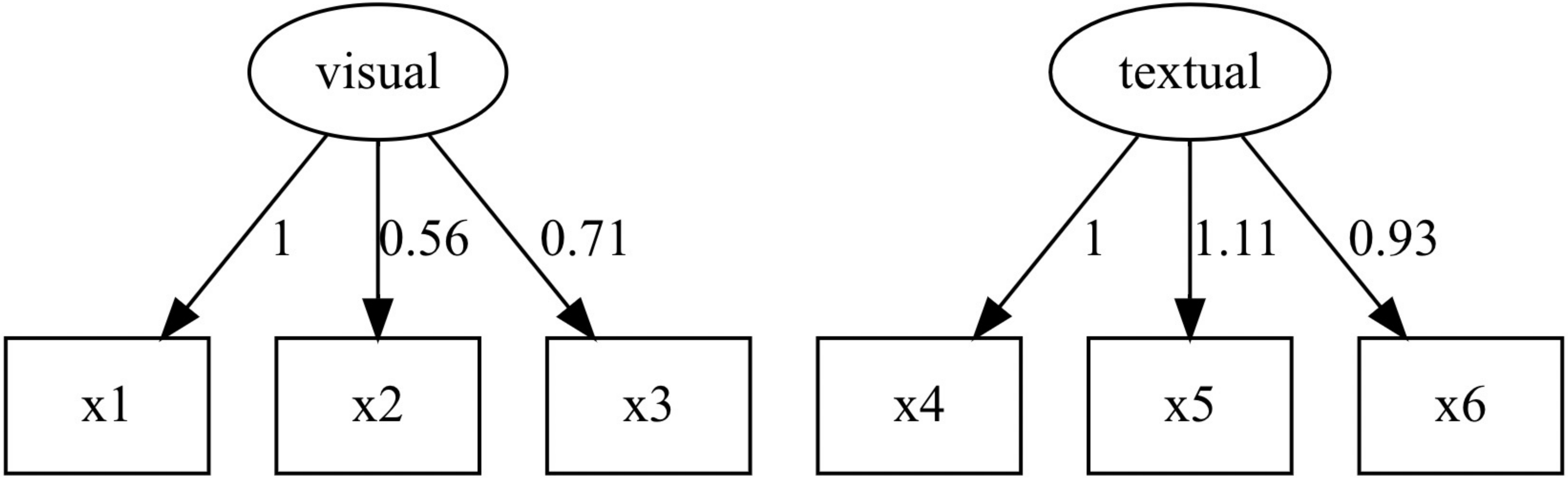
20220404_SEM_demo.R



- visual and textual ability are moderately positively correlated with one another
- x2 is probably more weakly associated with "visual" ability than x1 and x3 (more measurement variance?)
- x4, x5, x6 all strongly associated with "textual" ability

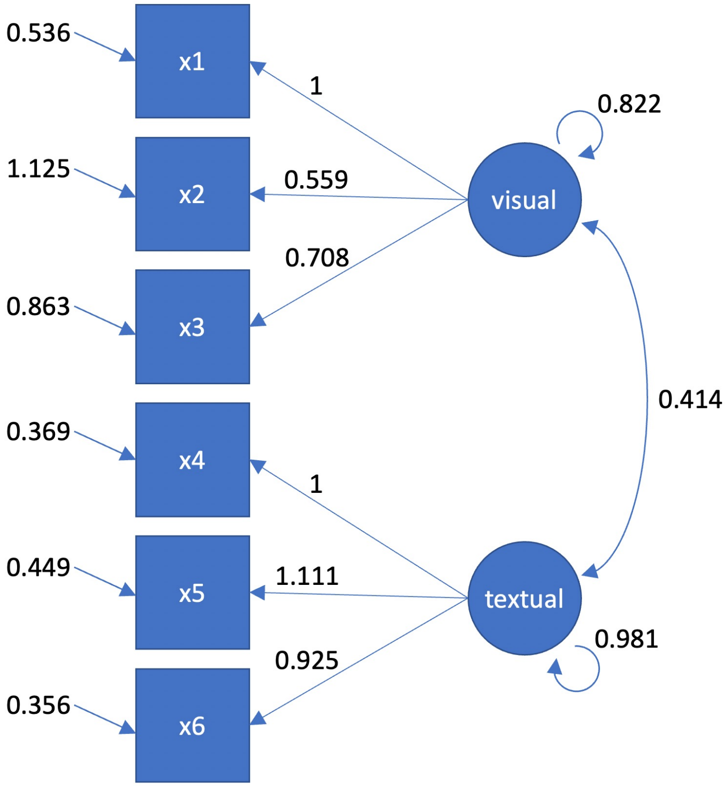
20220404_SEM_demo.R

```r
library(lavaanPlot)
lavaanPlot(model = fit1, coefs = TRUE)
```



Goodness of fit indices

- Oh, so many goodness of fit indices




```
lavaan 0.6-10 ended normally after 28 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        13

  Number of observations                           301

Model Test User Model:

  Test statistic                                24.361
  Degrees of freedom                                 8
  P-value (Chi-square)                           0.002

Model Test Baseline Model:

  Test statistic                               668.643
  Degrees of freedom                                15
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.975
  Tucker-Lewis Index (TLI)                       0.953

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -2520.252
  Loglikelihood unrestricted model (H1)      -2508.071

  Akaike (AIC)                                5066.503
  Bayesian (BIC)                              5114.696
  Sample-size adjusted Bayesian (BIC)         5073.467

Root Mean Square Error of Approximation:

  RMSEA                                          0.082
  90 Percent confidence interval - lower         0.046
  90 Percent confidence interval - upper         0.121
  P-value RMSEA <= 0.05                          0.067

Standardized Root Mean Square Residual:

  SRMR                                           0.047
```

Goodness of fit indices

Chi-squared: is covariance matrix different from 0 (is there anything to model)

baseline model: is there covariance in your data (yes, $p < .0005$)

```
lavaan 0.6-10 ended normally after 28 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        13

  Number of observations                           301

Model Test User Model:

  Test statistic                                24.361
  Degrees of freedom                                 8
  P-value (Chi-square)                           0.002

Model Test Baseline Model:

  Test statistic                               668.643
  Degrees of freedom                                15
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.975
  Tucker-Lewis Index (TLI)                       0.953

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -2520.252
  Loglikelihood unrestricted model (H1)      -2508.071

  Akaike (AIC)                                5066.503
  Bayesian (BIC)                              5114.696
  Sample-size adjusted Bayesian (BIC)         5073.467

Root Mean Square Error of Approximation:

  RMSEA                                          0.082
  90 Percent confidence interval - lower         0.046
  90 Percent confidence interval - upper         0.121
  P-value RMSEA ⩽ 0.05                           0.067

Standardized Root Mean Square Residual:

  SRMR                                           0.047
```

Goodness of fit indices

Chi-squared: is covariance matrix different from 0 (is there anything to model)

baseline model: is there covariance in your data (yes, $p < .0005$)

user model: is there covariance left after you fit your model - estimated 7 parameters: 4 loadings and 3 variance/covariances in latent vars (yes, $p < .0005$, but this is almost always significant especially if you have a half-decent N)

```
lavaan 0.6-10 ended normally after 28 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        13

  Number of observations                           301

Model Test User Model:

  Test statistic                                24.361
  Degrees of freedom                                 8
  P-value (Chi-square)                           0.002

Model Test Baseline Model:

  Test statistic                               668.643
  Degrees of freedom                                15
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.975
  Tucker-Lewis Index (TLI)                       0.953

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -2520.252
  Loglikelihood unrestricted model (H1)      -2508.071

  Akaike (AIC)                                5066.503
  Bayesian (BIC)                              5114.696
  Sample-size adjusted Bayesian (BIC)         5073.467

Root Mean Square Error of Approximation:

  RMSEA                                          0.082
  90 Percent confidence interval - lower         0.046
  90 Percent confidence interval - upper         0.121
  P-value RMSEA <= 0.05                          0.067

Standardized Root Mean Square Residual:

  SRMR                                           0.047
```

Goodness of fit indices

<u>comparison</u> of specified model to baseline (null) model

Comparative Fit Index (CFI): compares fit of model to baseline (null) model and should be above 0.9

Tucker-Lewis Index (TLI): very similar and highly correlated to CFI (only report one)

both of these have penalties for adding parameters

```
lavaan 0.6-10 ended normally after 28 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        13

  Number of observations                           301

Model Test User Model:

  Test statistic                                24.361
  Degrees of freedom                                 8
  P-value (Chi-square)                           0.002

Model Test Baseline Model:

  Test statistic                               668.643
  Degrees of freedom                                15
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.975
  Tucker-Lewis Index (TLI)                       0.953

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -2520.252
  Loglikelihood unrestricted model (H1)      -2508.071

  Akaike (AIC)                                5066.503
  Bayesian (BIC)                              5114.696
  Sample-size adjusted Bayesian (BIC)         5073.467

Root Mean Square Error of Approximation:

  RMSEA                                          0.082
  90 Percent confidence interval - lower         0.046
  90 Percent confidence interval - upper         0.121
  P-value RMSEA <= 0.05                          0.067

Standardized Root Mean Square Residual:

  SRMR                                           0.047
```

Goodness of fit indices

<u>information criteria</u> are only informative when you are comparing two models to one another – for example if you have 2 or more competing CFA models

these differ in how they penalize adding parameters

lower values indicate better fit (better reproduction of covariance structure of actual data by model)

```
lavaan 0.6-10 ended normally after 28 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        13

  Number of observations                           301

Model Test User Model:

  Test statistic                                24.361
  Degrees of freedom                                 8
  P-value (Chi-square)                           0.002

Model Test Baseline Model:

  Test statistic                               668.643
  Degrees of freedom                                15
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.975
  Tucker-Lewis Index (TLI)                       0.953

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -2520.252
  Loglikelihood unrestricted model (H1)      -2508.071

  Akaike (AIC)                                5066.503
  Bayesian (BIC)                              5114.696
  Sample-size adjusted Bayesian (BIC)         5073.467

Root Mean Square Error of Approximation:

  RMSEA                                          0.082
  90 Percent confidence interval - lower         0.046
  90 Percent confidence interval - upper         0.121
  P-value RMSEA <= 0.05                          0.067

Standardized Root Mean Square Residual:

  SRMR                                           0.047
```

Goodness of fit indices

RMSEA and SRMR are absolute measures of goodness of fit: an ideal best fitting model will have value of 0

RMSEA is most common measure (related to chi-squared of fitted model scaled by df and N)
0.01 "excellent"
0.05 "good"
0.08 "mediocre" (McCallum et al., 1996)

SRMR is standardized difference between observed and predicted correlation; values less than .08 are "good fit"

```
lavaan 0.6-10 ended normally after 28 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        13

  Number of observations                           301

Model Test User Model:

  Test statistic                                24.361
  Degrees of freedom                                 8
  P-value (Chi-square)                           0.002

Model Test Baseline Model:

  Test statistic                               668.643
  Degrees of freedom                                15
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.975
  Tucker-Lewis Index (TLI)                       0.953

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -2520.252
  Loglikelihood unrestricted model (H1)      -2508.071

  Akaike (AIC)                                5066.503
  Bayesian (BIC)                              5114.696
  Sample-size adjusted Bayesian (BIC)         5073.467

Root Mean Square Error of Approximation:

  RMSEA                                          0.082
  90 Percent confidence interval - lower         0.046
  90 Percent confidence interval - upper         0.121
  P-value RMSEA <= 0.05                          0.067

Standardized Root Mean Square Residual:

  SRMR                                           0.047
```

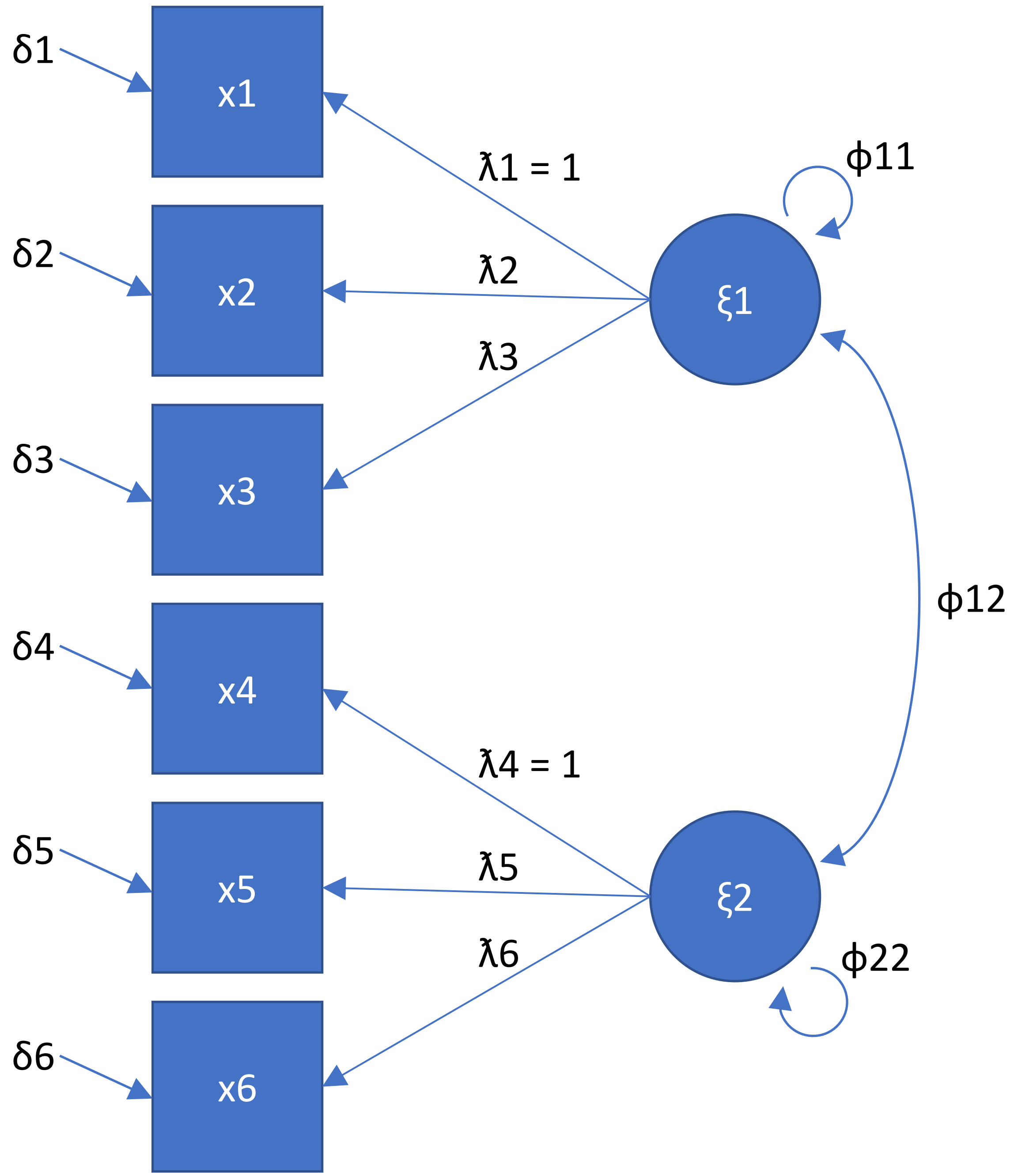
Goodness of fit indices

- There are many more.
- You want to avoid cherry-picking goodness of fit indices.
- Consider the goal of your latent variable modeling approach (most parsimonious model, model comparison, ...)
- If you decide your model needs to be changed, cross-validate!

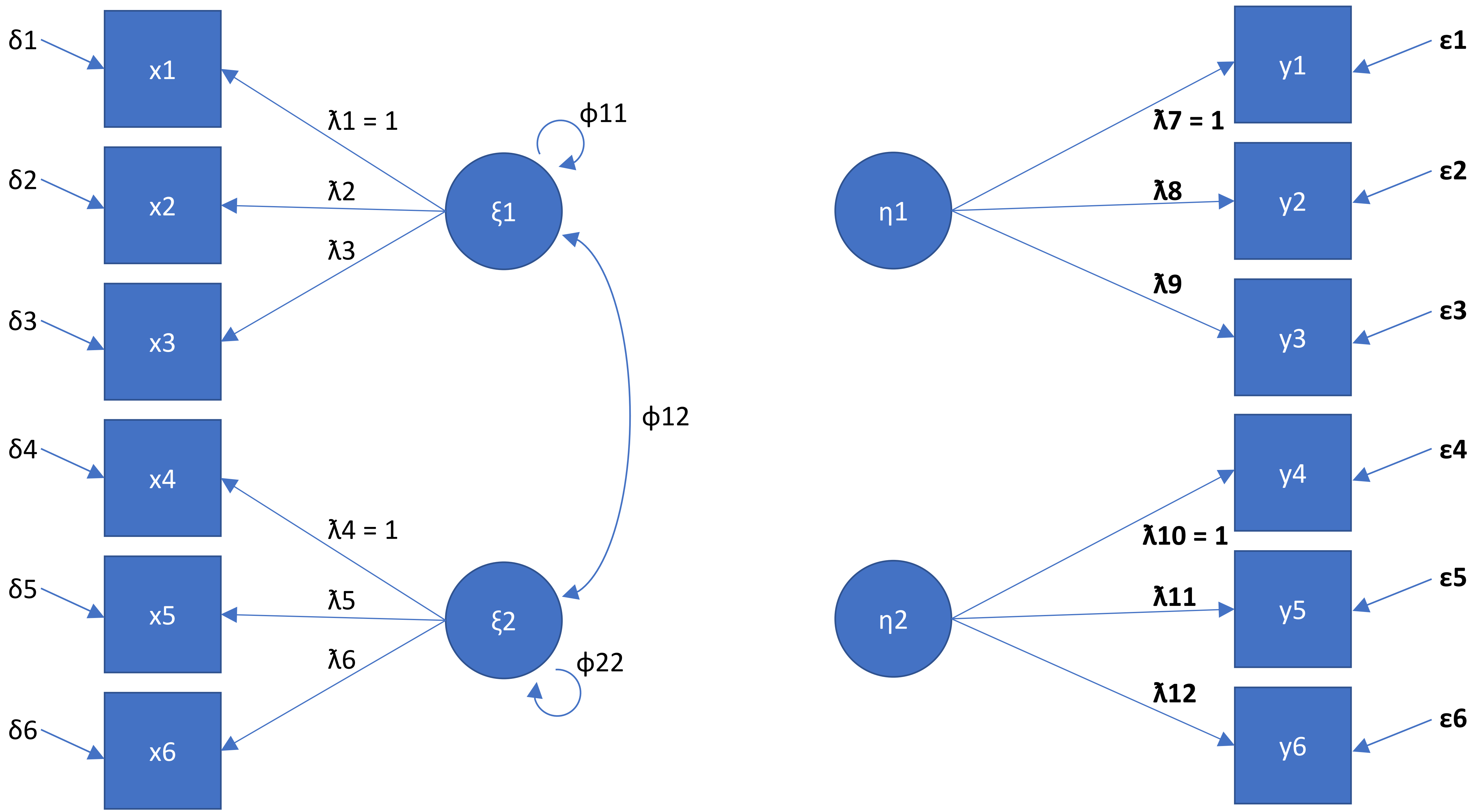
```
lavaan 0.6-10 ended normally after 28 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        13

  Number of observations                           301

Model Test User Model:

  Test statistic                                24.361
  Degrees of freedom                                 8
  P-value (Chi-square)                           0.002

Model Test Baseline Model:

  Test statistic                               668.643
  Degrees of freedom                                15
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.975
  Tucker-Lewis Index (TLI)                       0.953

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)              -2520.252
  Loglikelihood unrestricted model (H1)      -2508.071

  Akaike (AIC)                                5066.503
  Bayesian (BIC)                              5114.696
  Sample-size adjusted Bayesian (BIC)         5073.467

Root Mean Square Error of Approximation:

  RMSEA                                          0.082
  90 Percent confidence interval - lower         0.046
  90 Percent confidence interval - upper         0.121
  P-value RMSEA <= 0.05                          0.067

Standardized Root Mean Square Residual:

  SRMR                                           0.047
```

Full structural equation model example

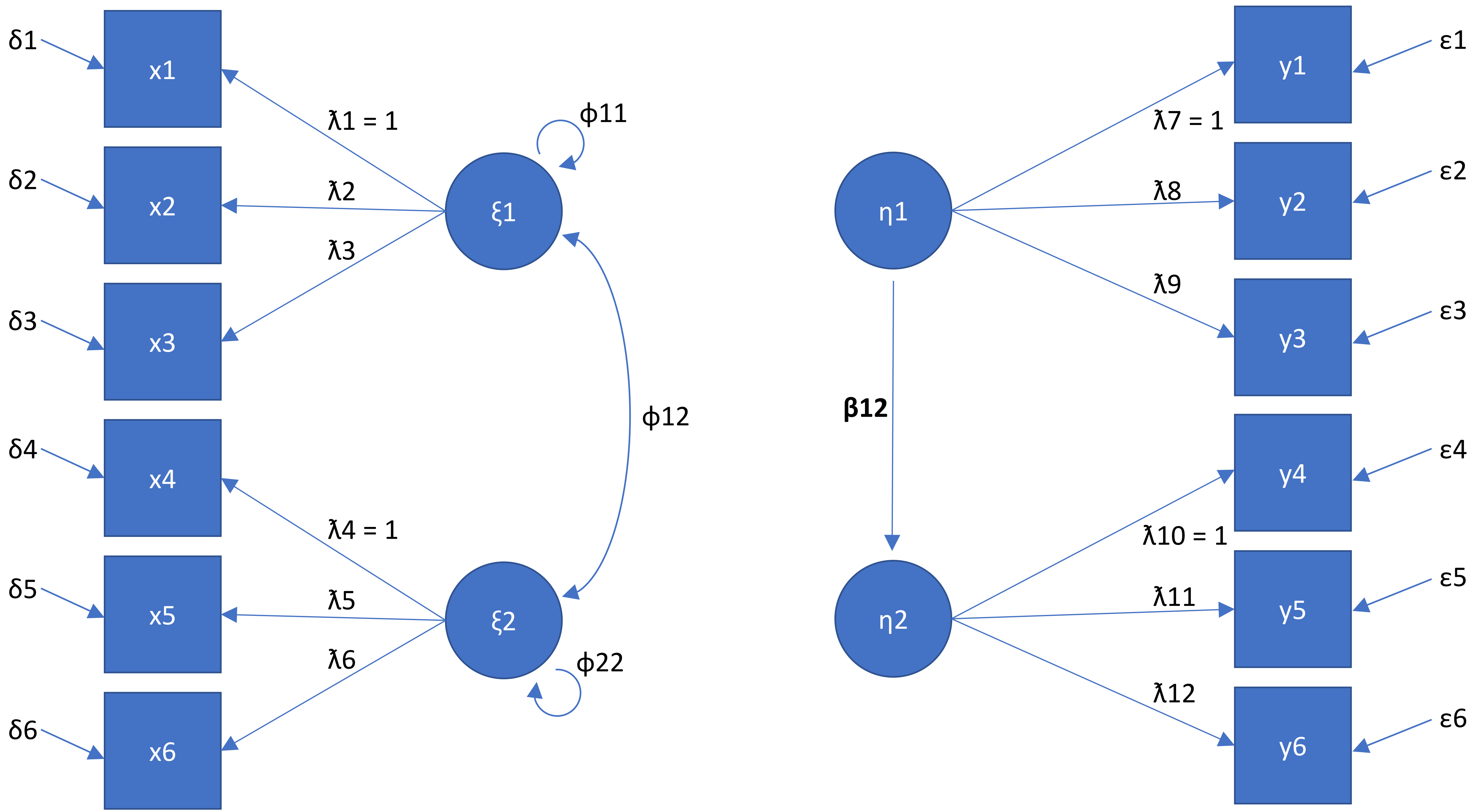
- we have seen the "x" side of a latent variable model (equivalent to a confirmatory factor analysis)
- now we can add a "y" side and see how the "x" latent variables (*exogenous* latent variables) influence the "y" latent variables (***endogenous*** latent variables)
- the influence of exogenous latent variables on endogenous latent variables (and endogenous latent variables on each other) is the **structural model**
- endogenous latent variables incorporate measurement error on the latent variable (may be influences on them beyond the exogenous LVs - like regression where we may not perfectly predict y)



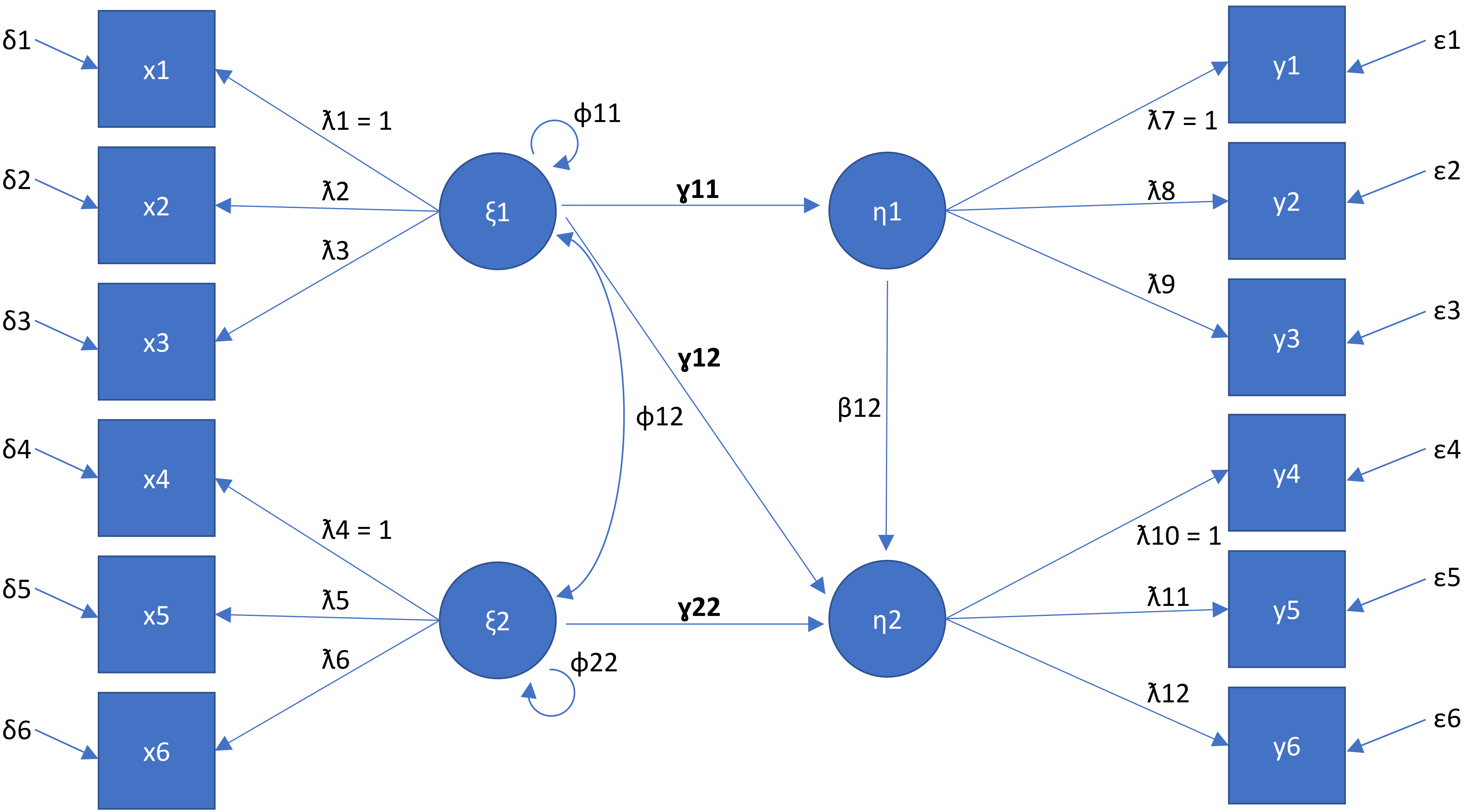
20220404_SEM_demo.R

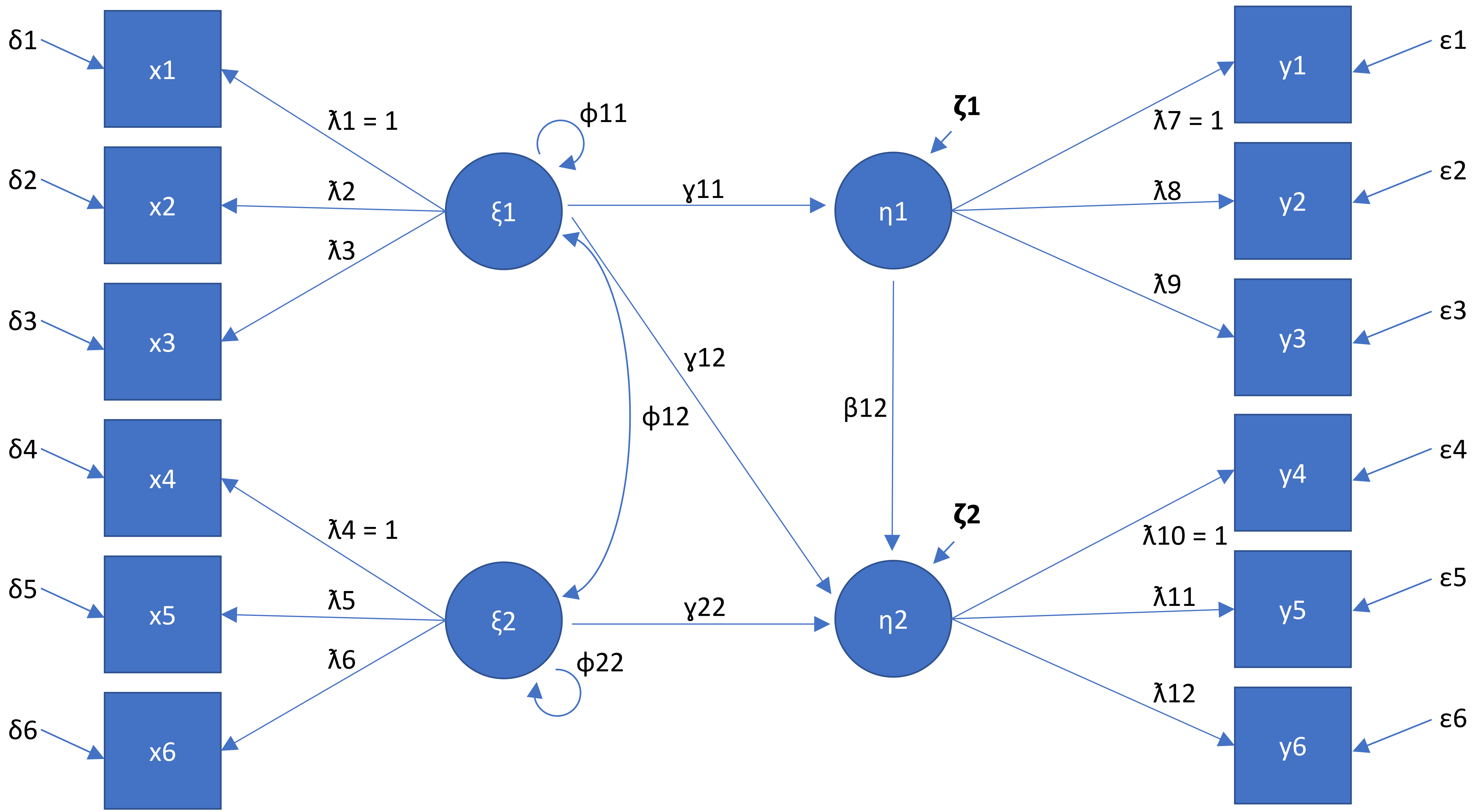


δ1
δ2
δ3
δ4
δ5
δ6
x1
x2
x3
x4
x5
x6
λ1 = 1
λ2
λ3
λ4 = 1
λ5
λ6
ξ1
ξ2
ϕ11
ϕ12
ϕ22
η1
η2
λ7 = 1
λ8
λ9
λ10 = 1
λ11
λ12
y1
y2
y3
y4
y5
y6
ε1
ε2
ε3
ε4
ε5
ε6

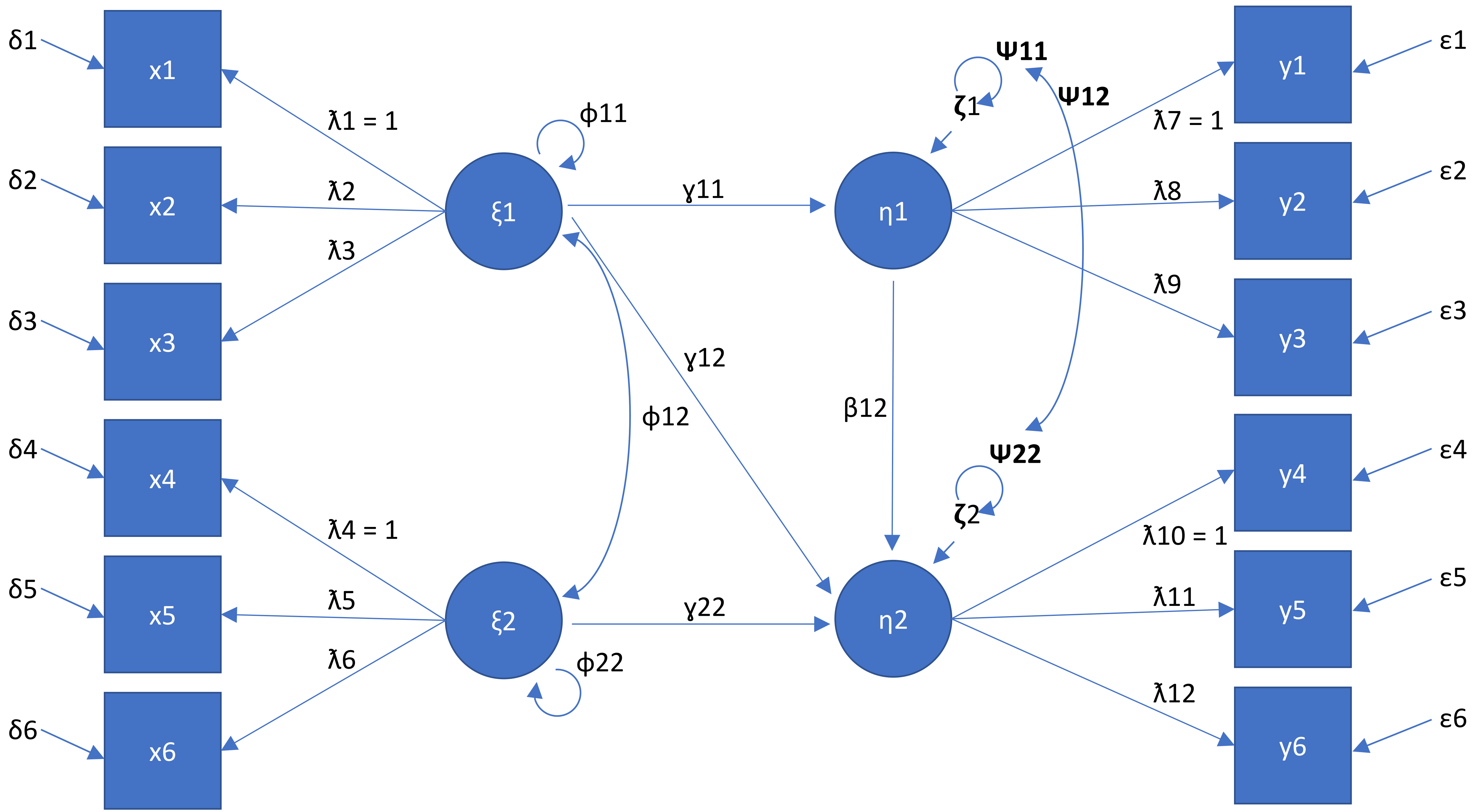


δ1
δ2
δ3
δ4
δ5
δ6
x1
x2
x3
x4
x5
x6
λ1 = 1
λ2
λ3
λ4 = 1
λ5
λ6
ξ1
ξ2
ϕ11
ϕ12
ϕ22
η1
η2
β12
λ7 = 1
λ8
λ9
λ10 = 1
λ11
λ12
y1
y2
y3
y4
y5
y6
ε1
ε2
ε3
ε4
ε5
ε6

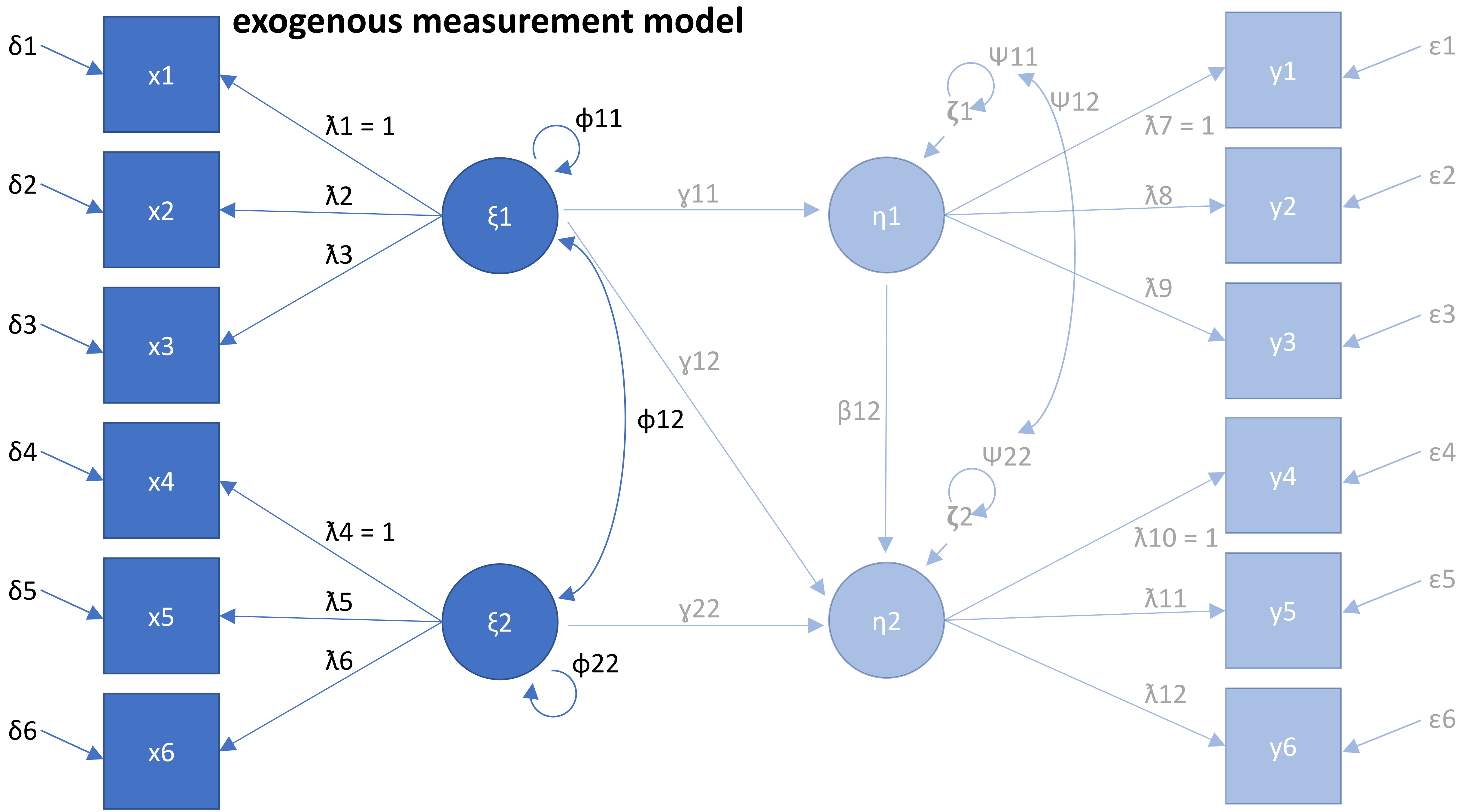

δ1
δ2
δ3
δ4
δ5
δ6
x1
x2
x3
x4
x5
x6
λ1 = 1
λ2
λ3
λ4 = 1
λ5
λ6
ϕ11
ϕ12
ϕ22
ξ1
ξ2
γ11
γ12
γ22
η1
η2
β12
λ7 = 1
λ8
λ9
λ10 = 1
λ11
λ12
y1
y2
y3
y4
y5
y6
ε1
ε2
ε3
ε4
ε5
ε6



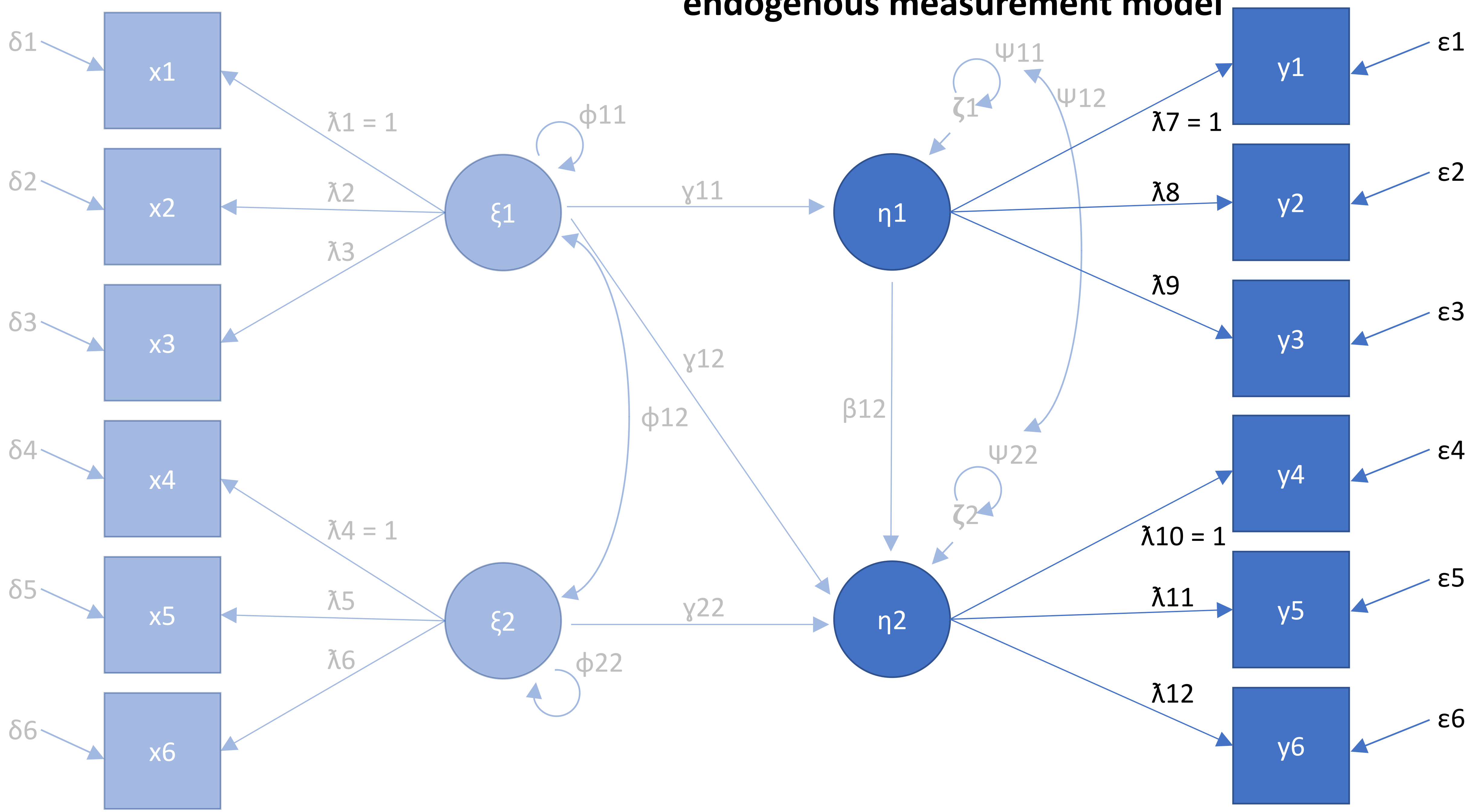
δ1
δ2
δ3
δ4
δ5
δ6
x1
x2
x3
x4
x5
x6
ƛ1 = 1
ƛ2
ƛ3
ƛ4 = 1
ƛ5
ƛ6
ϕ11
ϕ12
ϕ22
ξ1
ξ2
γ11
γ12
γ22
η1
η2
β12
ζ1
ζ2
ƛ7 = 1
ƛ8
ƛ9
ƛ10 = 1
ƛ11
ƛ12
y1
y2
y3
y4
y5
y6
ε1
ε2
ε3
ε4
ε5
ε6

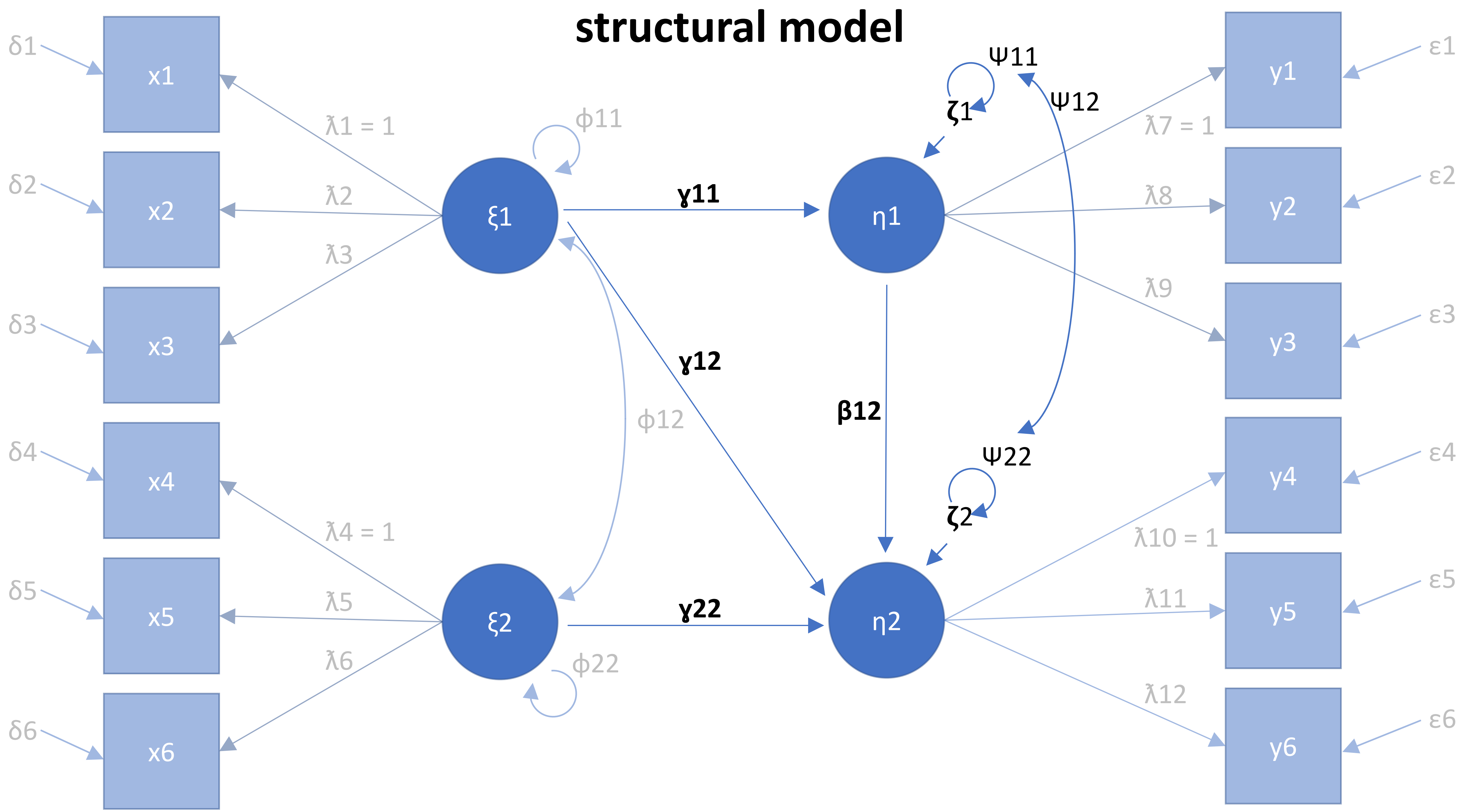


δ1
δ2
δ3
δ4
δ5
δ6
x1
x2
x3
x4
x5
x6
ƛ1 = 1
ƛ2
ƛ3
ƛ4 = 1
ƛ5
ƛ6
ξ1
ξ2
ϕ11
ϕ12
ϕ22
γ11
γ12
γ22
η1
η2
β12
ζ1
ζ2
Ψ11
Ψ12
Ψ22
ƛ7 = 1
ƛ8
ƛ9
ƛ10 = 1
ƛ11
ƛ12
y1
y2
y3
y4
y5
y6
ε1
ε2
ε3
ε4
ε5
ε6


exogenous measurement model
δ1
δ2
δ3
δ4
δ5
δ6
x1
x2
x3
x4
x5
x6
λ1 = 1
λ2
λ3
λ4 = 1
λ5
λ6
ξ1
ξ2
ϕ11
ϕ12
ϕ22
γ11
γ12
γ22
η1
η2
β12
Ψ11
Ψ12
Ψ22
ζ1
ζ2
λ7 = 1
λ8
λ9
λ10 = 1
λ11
λ12
y1
y2
y3
y4
y5
y6
ε1
ε2
ε3
ε4
ε5
ε6

endogenous measurement model




structural model
δ1
δ2
δ3
δ4
δ5
δ6
x1
x2
x3
x4
x5
x6
λ1 = 1
λ2
λ3
λ4 = 1
λ5
λ6
ϕ11
ϕ12
ϕ22
ξ1
ξ2
γ11
γ12
γ22
η1
η2
β12
Ψ11
Ψ12
Ψ22
ζ1
ζ2
λ7 = 1
λ8
λ9
λ10 = 1
λ11
λ12
y1
y2
y3
y4
y5
y6
ε1
ε2
ε3
ε4
ε5
ε6

Components of full SEM model

- exogenous ("x") measurement model
  - observed/manifest variables (x)
  - measurement error on x variables (δ)*
  - exogenous latent variables (ξ)
  - variance/covariance among exogenous LVs (ϕ)
- endogenous ("y") measurement model
  - observed/manifest variables (y)
  - measurement error on y variables (ε)*
  - endogenous latent variables (η)
  - variance/covariance among measurement error (ζ) on endogenous LVs (Ψ)

- **structural** model
  - path coefficients between exogenous LVs and endogenous LVs (ɣ)
  - path coefficients between different endogenous LVs (β)

*variance/covariance; not estimating these directly ("theta-delta", "theta-epsilon")

Full structural equation model example

- "Political Democracy" data included with `lavaan` package
- 3 economic indicators (x1-x3) measured in 1960
- 4 political indicators measured in both 1960 (y1-y4) and 1965 (y5-y8)

Full structural equation model example

- measurement model specifies both exogenous and endogenous LVs

- structural model ("regressions") specifies relationships between LVs

```
pdmodel1 <- '
# measurement model
    ind60 =~ x1 + x2 + x3
    dem60 =~ y1 + y2 + y3 + y4
    dem65 =~ y5 + y6 + y7 + y8
# regressions
    dem60 ~ ind60
    dem65 ~ ind60 + dem60'
```

20220404_SEM_demo.R

Full structural equation model example

- measurement model specifies both exogenous and endogenous LVs

- structural model ("regressions") specifies relationships between LVs

```
pdmodel1 <- '
# measurement model
    ind60 =~ x1 + x2 + x3
    dem60 =~ y1 + y2 + y3 + y4
    dem65 =~ y5 + y6 + y7 + y8
# regressions
    dem60 ~ ind60
    dem65 ~ ind60 + dem60'
```

```
pd_fit1 <- sem(model = pdmodel1, data = poldem)
summary(pd_fit1, standardized = TRUE, fit.measures = TRUE)
```

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  ind60 =~
    x1                1.000                               0.669    0.920
    x2                2.182    0.139   15.714    0.000    1.461    0.973
    x3                1.819    0.152   11.956    0.000    1.218    0.872
  dem60 =~
    y1                1.000                               2.201    0.845
    y2                1.354    0.175    7.755    0.000    2.980    0.760
    y3                1.044    0.150    6.961    0.000    2.298    0.705
    y4                1.300    0.138    9.412    0.000    2.860    0.860
  dem65 =~
    y5                1.000                               2.084    0.803
    y6                1.258    0.164    7.651    0.000    2.623    0.783
    y7                1.282    0.158    8.137    0.000    2.673    0.819
    y8                1.310    0.154    8.529    0.000    2.730    0.847

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  dem60 ~
    ind60             1.474    0.392    3.763    0.000    0.448    0.448
  dem65 ~
    ind60             0.453    0.220    2.064    0.039    0.146    0.146
    dem60             0.864    0.113    7.671    0.000    0.913    0.913

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .x1                0.082    0.020    4.180    0.000    0.082    0.154
   .x2                0.118    0.070    1.689    0.091    0.118    0.053
   .x3                0.467    0.090    5.174    0.000    0.467    0.240
   .y1                1.942    0.395    4.910    0.000    1.942    0.286
   .y2                6.490    1.185    5.479    0.000    6.490    0.422
   .y3                5.340    0.943    5.662    0.000    5.340    0.503
   .y4                2.887    0.610    4.731    0.000    2.887    0.261
   .y5                2.390    0.447    5.351    0.000    2.390    0.355
   .y6                4.343    0.796    5.456    0.000    4.343    0.387
   .y7                3.510    0.668    5.252    0.000    3.510    0.329
   .y8                2.940    0.586    5.019    0.000    2.940    0.283
    ind60             0.448    0.087    5.169    0.000    1.000    1.000
   .dem60             3.872    0.893    4.338    0.000    0.799    0.799
   .dem65             0.115    0.200    0.575    0.565    0.026    0.026
```

loadings of manifest variables on exogenous LVs

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  ind60 =~
    x1                1.000                               0.669    0.920
    x2                2.182    0.139   15.714    0.000    1.461    0.973
    x3                1.819    0.152   11.956    0.000    1.218    0.872
  dem60 =~
    y1                1.000                               2.201    0.845
    y2                1.354    0.175    7.755    0.000    2.980    0.760
    y3                1.044    0.150    6.961    0.000    2.298    0.705
    y4                1.300    0.138    9.412    0.000    2.860    0.860
  dem65 =~
    y5                1.000                               2.084    0.803
    y6                1.258    0.164    7.651    0.000    2.623    0.783
    y7                1.282    0.158    8.137    0.000    2.673    0.819
    y8                1.310    0.154    8.529    0.000    2.730    0.847

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  dem60 ~
    ind60             1.474    0.392    3.763    0.000    0.448    0.448
  dem65 ~
    ind60             0.453    0.220    2.064    0.039    0.146    0.146
    dem60             0.864    0.113    7.671    0.000    0.913    0.913

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .x1                0.082    0.020    4.180    0.000    0.082    0.154
   .x2                0.118    0.070    1.689    0.091    0.118    0.053
   .x3                0.467    0.090    5.174    0.000    0.467    0.240
   .y1                1.942    0.395    4.910    0.000    1.942    0.286
   .y2                6.490    1.185    5.479    0.000    6.490    0.422
   .y3                5.340    0.943    5.662    0.000    5.340    0.503
   .y4                2.887    0.610    4.731    0.000    2.887    0.261
   .y5                2.390    0.447    5.351    0.000    2.390    0.355
   .y6                4.343    0.796    5.456    0.000    4.343    0.387
   .y7                3.510    0.668    5.252    0.000    3.510    0.329
   .y8                2.940    0.586    5.019    0.000    2.940    0.283
    ind60             0.448    0.087    5.169    0.000    1.000    1.000
   .dem60             3.872    0.893    4.338    0.000    0.799    0.799
   .dem65             0.115    0.200    0.575    0.565    0.026    0.026
```

loadings of manifest variables on endogenous LVs

20220404_SEM_demo.R

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  ind60 =~
    x1                1.000                               0.669    0.920
    x2                2.182    0.139   15.714    0.000    1.461    0.973
    x3                1.819    0.152   11.956    0.000    1.218    0.872
  dem60 =~
    y1                1.000                               2.201    0.845
    y2                1.354    0.175    7.755    0.000    2.980    0.760
    y3                1.044    0.150    6.961    0.000    2.298    0.705
    y4                1.300    0.138    9.412    0.000    2.860    0.860
  dem65 =~
    y5                1.000                               2.084    0.803
    y6                1.258    0.164    7.651    0.000    2.623    0.783
    y7                1.282    0.158    8.137    0.000    2.673    0.819
    y8                1.310    0.154    8.529    0.000    2.730    0.847

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  dem60 ~
    ind60             1.474    0.392    3.763    0.000    0.448    0.448
  dem65 ~
    ind60             0.453    0.220    2.064    0.039    0.146    0.146
    dem60             0.864    0.113    7.671    0.000    0.913    0.913

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .x1                0.082    0.020    4.180    0.000    0.082    0.154
   .x2                0.118    0.070    1.689    0.091    0.118    0.053
   .x3                0.467    0.090    5.174    0.000    0.467    0.240
   .y1                1.942    0.395    4.910    0.000    1.942    0.286
   .y2                6.490    1.185    5.479    0.000    6.490    0.422
   .y3                5.340    0.943    5.662    0.000    5.340    0.503
   .y4                2.887    0.610    4.731    0.000    2.887    0.261
   .y5                2.390    0.447    5.351    0.000    2.390    0.355
   .y6                4.343    0.796    5.456    0.000    4.343    0.387
   .y7                3.510    0.668    5.252    0.000    3.510    0.329
   .y8                2.940    0.586    5.019    0.000    2.940    0.283
    ind60             0.448    0.087    5.169    0.000    1.000    1.000
   .dem60             3.872    0.893    4.338    0.000    0.799    0.799
   .dem65             0.115    0.200    0.575    0.565    0.026    0.026
```

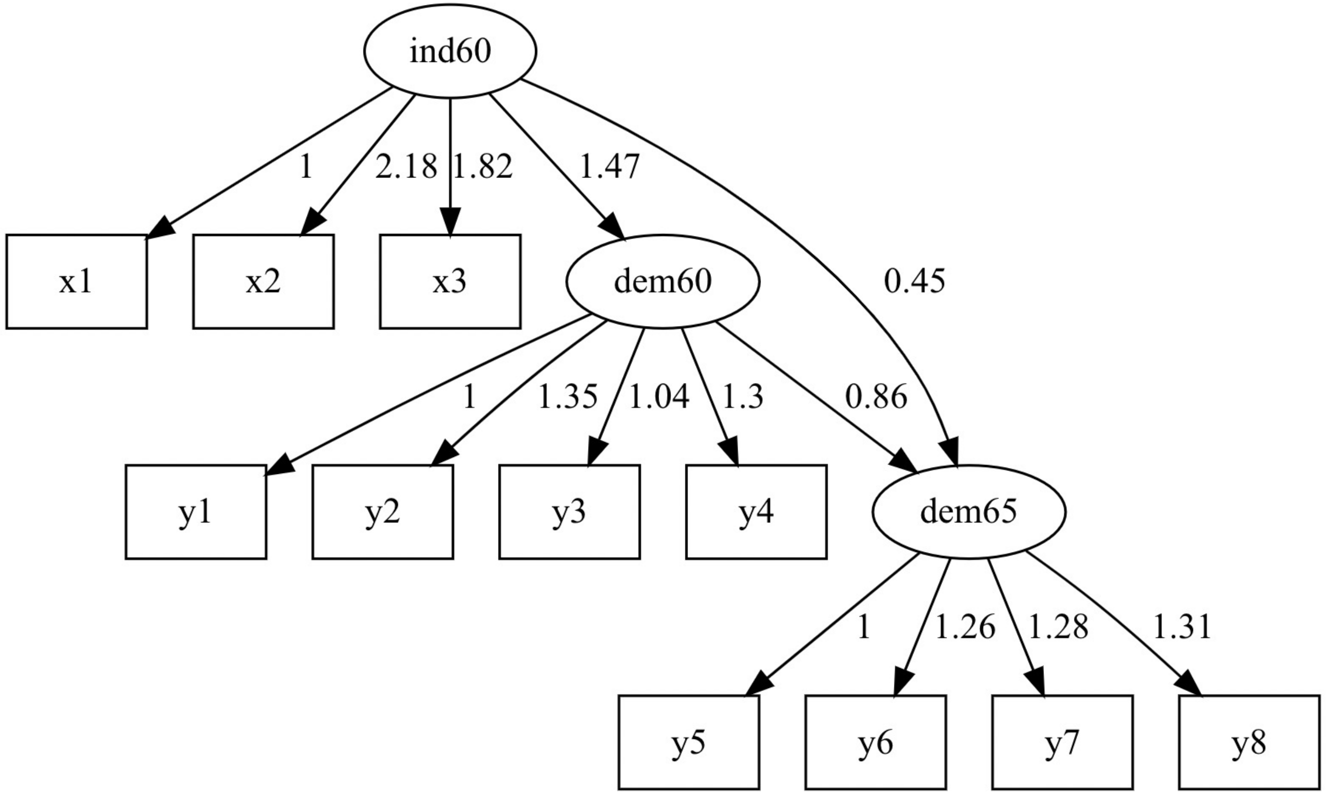
path coefficients of structural model

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  ind60 =~
    x1                1.000                               0.669    0.920
    x2                2.182    0.139   15.714    0.000    1.461    0.973
    x3                1.819    0.152   11.956    0.000    1.218    0.872
  dem60 =~
    y1                1.000                               2.201    0.845
    y2                1.354    0.175    7.755    0.000    2.980    0.760
    y3                1.044    0.150    6.961    0.000    2.298    0.705
    y4                1.300    0.138    9.412    0.000    2.860    0.860
  dem65 =~
    y5                1.000                               2.084    0.803
    y6                1.258    0.164    7.651    0.000    2.623    0.783
    y7                1.282    0.158    8.137    0.000    2.673    0.819
    y8                1.310    0.154    8.529    0.000    2.730    0.847

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  dem60 ~
    ind60             1.474    0.392    3.763    0.000    0.448    0.448
  dem65 ~
    ind60             0.453    0.220    2.064    0.039    0.146    0.146
    dem60             0.864    0.113    7.671    0.000    0.913    0.913

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .x1                0.082    0.020    4.180    0.000    0.082    0.154
   .x2                0.118    0.070    1.689    0.091    0.118    0.053
   .x3                0.467    0.090    5.174    0.000    0.467    0.240
   .y1                1.942    0.395    4.910    0.000    1.942    0.286
   .y2                6.490    1.185    5.479    0.000    6.490    0.422
   .y3                5.340    0.943    5.662    0.000    5.340    0.503
   .y4                2.887    0.610    4.731    0.000    2.887    0.261
   .y5                2.390    0.447    5.351    0.000    2.390    0.355
   .y6                4.343    0.796    5.456    0.000    4.343    0.387
   .y7                3.510    0.668    5.252    0.000    3.510    0.329
   .y8                2.940    0.586    5.019    0.000    2.940    0.283
    ind60             0.448    0.087    5.169    0.000    1.000    1.000
   .dem60             3.872    0.893    4.338    0.000    0.799    0.799
   .dem65             0.115    0.200    0.575    0.565    0.026    0.026
```

variances of residuals and of exogenous LV

20220404_SEM_demo.R

```
Latent Variables:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  ind60 =~
    x1                1.000                               0.669    0.920
    x2                2.182    0.139   15.714    0.000    1.461    0.973
    x3                1.819    0.152   11.956    0.000    1.218    0.872
  dem60 =~
    y1                1.000                               2.201    0.845
    y2                1.354    0.175    7.755    0.000    2.980    0.760
    y3                1.044    0.150    6.961    0.000    2.298    0.705
    y4                1.300    0.138    9.412    0.000    2.860    0.860
  dem65 =~
    y5                1.000                               2.084    0.803
    y6                1.258    0.164    7.651    0.000    2.623    0.783
    y7                1.282    0.158    8.137    0.000    2.673    0.819
    y8                1.310    0.154    8.529    0.000    2.730    0.847

Regressions:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  dem60 ~
    ind60             1.474    0.392    3.763    0.000    0.448    0.448
  dem65 ~
    ind60             0.453    0.220    2.064    0.039    0.146    0.146
    dem60             0.864    0.113    7.671    0.000    0.913    0.913

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .x1                0.082    0.020    4.180    0.000    0.082    0.154
   .x2                0.118    0.070    1.689    0.091    0.118    0.053
   .x3                0.467    0.090    5.174    0.000    0.467    0.240
   .y1                1.942    0.395    4.910    0.000    1.942    0.286
   .y2                6.490    1.185    5.479    0.000    6.490    0.422
   .y3                5.340    0.943    5.662    0.000    5.340    0.503
   .y4                2.887    0.610    4.731    0.000    2.887    0.261
   .y5                2.390    0.447    5.351    0.000    2.390    0.355
   .y6                4.343    0.796    5.456    0.000    4.343    0.387
   .y7                3.510    0.668    5.252    0.000    3.510    0.329
   .y8                2.940    0.586    5.019    0.000    2.940    0.283
    ind60             0.448    0.087    5.169    0.000    1.000    1.000
   .dem60             3.872    0.893    4.338    0.000    0.799    0.799
   .dem65             0.115    0.200    0.575    0.565    0.026    0.026
```



20220404_SEM_demo.R

Correlated residuals

- Because you have 2 sets of variables measured at different time points, might reasonably expect that their measurement errors are correlated.
- Improves fit indices in this case.

```
pdmodel2 <- '
# measurement model
    ind60 =~ x1 + x2 + x3
    dem60 =~ y1 + y2 + y3 + y4
    dem65 =~ y5 + y6 + y7 + y8
# regressions
    dem60 ~ ind60
    dem65 ~ ind60 + dem60
# residual correlations
    y1 ~~ y5
    y2 ~~ y6
    y3 ~~ y7
    y4 ~~ y8'
```

Modification indices (MIs)

- What if your fit is bad?
- `lavaan` will suggest ways to improve it

- <u>do not</u> check MIs and rerun your model unless you can cross-validate modified model on a new sample

```
> modindices(fit2, sort = TRUE, maximum.number = 8)
       lhs op rhs     mi    epc sepc.lv sepc.all sepc.nox
30  visual =~  x9 36.411  0.577   0.519    0.515    0.515
76      x7 ~~  x8 34.145  0.536   0.536    0.859    0.859
28  visual =~  x7 18.631 -0.422  -0.380   -0.349   -0.349
78      x8 ~~  x9 14.946 -0.423  -0.423   -0.805   -0.805
33 textual =~  x3  9.151 -0.272  -0.269   -0.238   -0.238
55      x2 ~~  x7  8.918 -0.183  -0.183   -0.192   -0.192
31 textual =~  x1  8.903  0.350   0.347    0.297    0.297
51      x2 ~~  x3  8.532  0.218   0.218    0.223    0.223
```

20220404_SEM_demo.R

What could go wrong?

- well, quite a bit (model identification can be an issue)
- much of the inference is based on assumptions of multivariate normality (be even more careful about examining indicators, consider transforming them if skewed, check for outliers)
- when you are using these approaches for model construction it is *vital* to cross validate
- many alternative models may fit data just as well and so if you are trying to build models it is important to compare to alternatives
- avoid using causal language even though arrows point in one direction

Breckler (1990) *Psychological Bulletin*