Structural Equation Modeling with R using lavaan

Alexander M. Schoemann

SPSP 2015

East Carolina University

What is SEM?

- AKA:
  - Simultaneous Equations
  - Covariance Structure Analysis
  - Path Analysis
  - Confirmatory Factor Analysis (CFA)

What is a model?

- A model is a set of relations between observed and latent variables that can be represented by a set of equations.
- Or...
- A useful representation of reality

Box: "essentially, all models are wrong, but some are useful"

Latent variables

- A latent variable:
  - is a construct that is hypothesized to exist but is not directly measurable
  - causes behaviors that are directly measurable; these behaviors are called indicators of the latent variable
  - is defined as everything that is in common between its manifest (observed) indicators

SEM in R

- There are (at least) four packages that fit SEM models in R
  - lavaan, OpenMx, sem, and lava

lavaan

a free open-source, but commercial-quality package for latent variable modeling.

– `From lavaan.org`

- Features
  - Full support for mean structures and multiple groups
  - Several estimators available (including ML, GLS, WLS. . .)
  - Standard and robust standard errors and test statistics (bootstrapping too!)
  - Missing data handling through FIML
  - Linear and non-linear inequality constraints
  - Support for categorical data (and mixture of binary, ordered and continuous observed variables)

lavaan syntax

- lavaan specifies relationships based on a path diagram
  - Every arrow in the path diagram is a line of syntax in lavaan
- lavaan syntax uses a few different operators to specify relationships between variables
  - $=\sim$ is used for factor loadings
    - `Positive =~ Great`
  - $\sim\sim$ is used for variance and covariances
    - `Great ~~ Great`
  - $\sim$ is used for regressions
  - * is used to fix a parameter to a specific value
    - `Positive ~~ 1*Positive`

lavaan: Example

- Two factor CFA based on the msq daa (from the psych package)

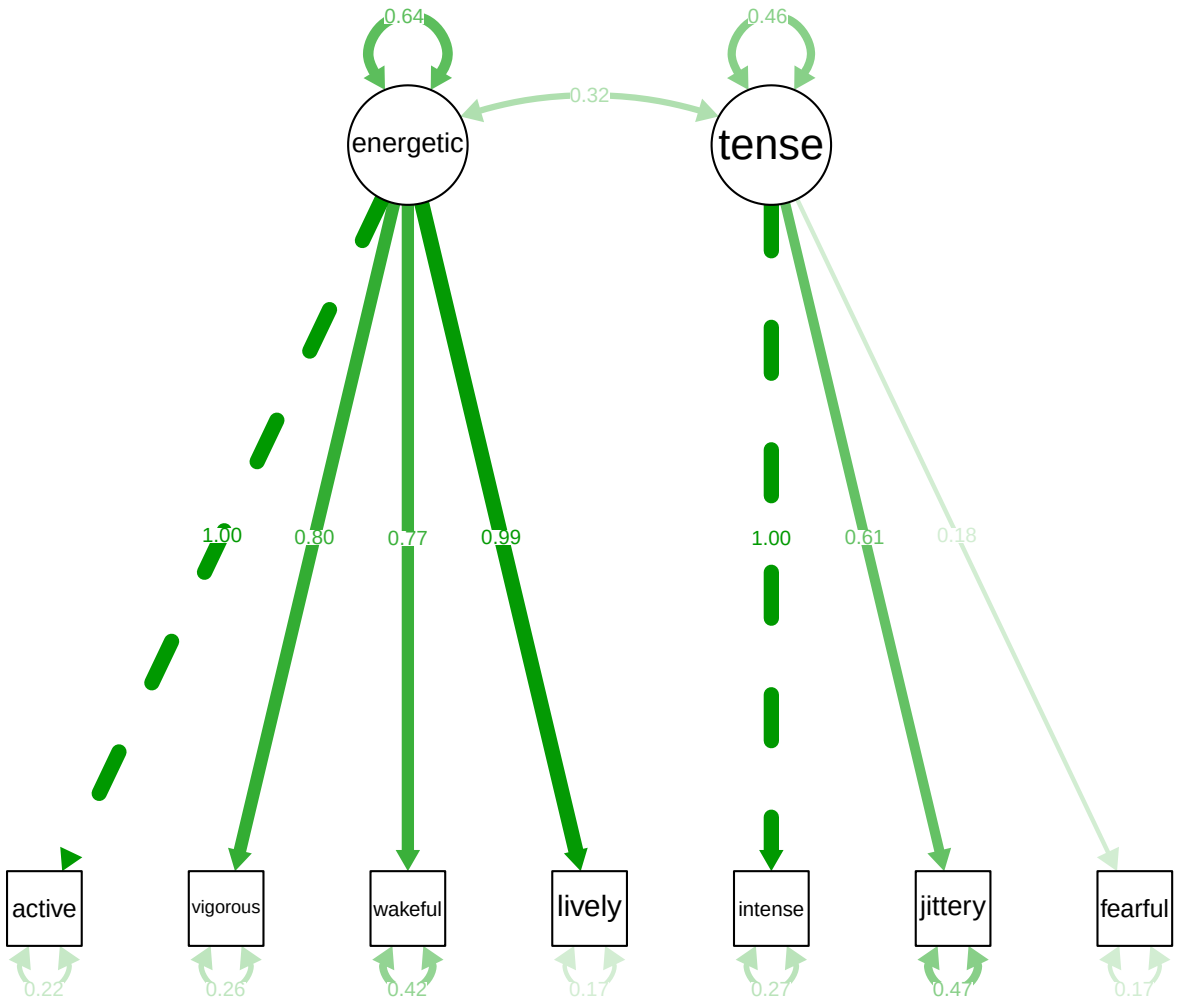
```
mod <- '
energetic =~ active + vigorous + wakeful + lively
tense =~ intense + jittery + fearful
'

fit <- cfa(mod, data = msq)
```

lavaan: Example

By default lavaan uses a marker variable method of scale setting

```
semPaths(fit, "est", nCharNodes = 0)
```

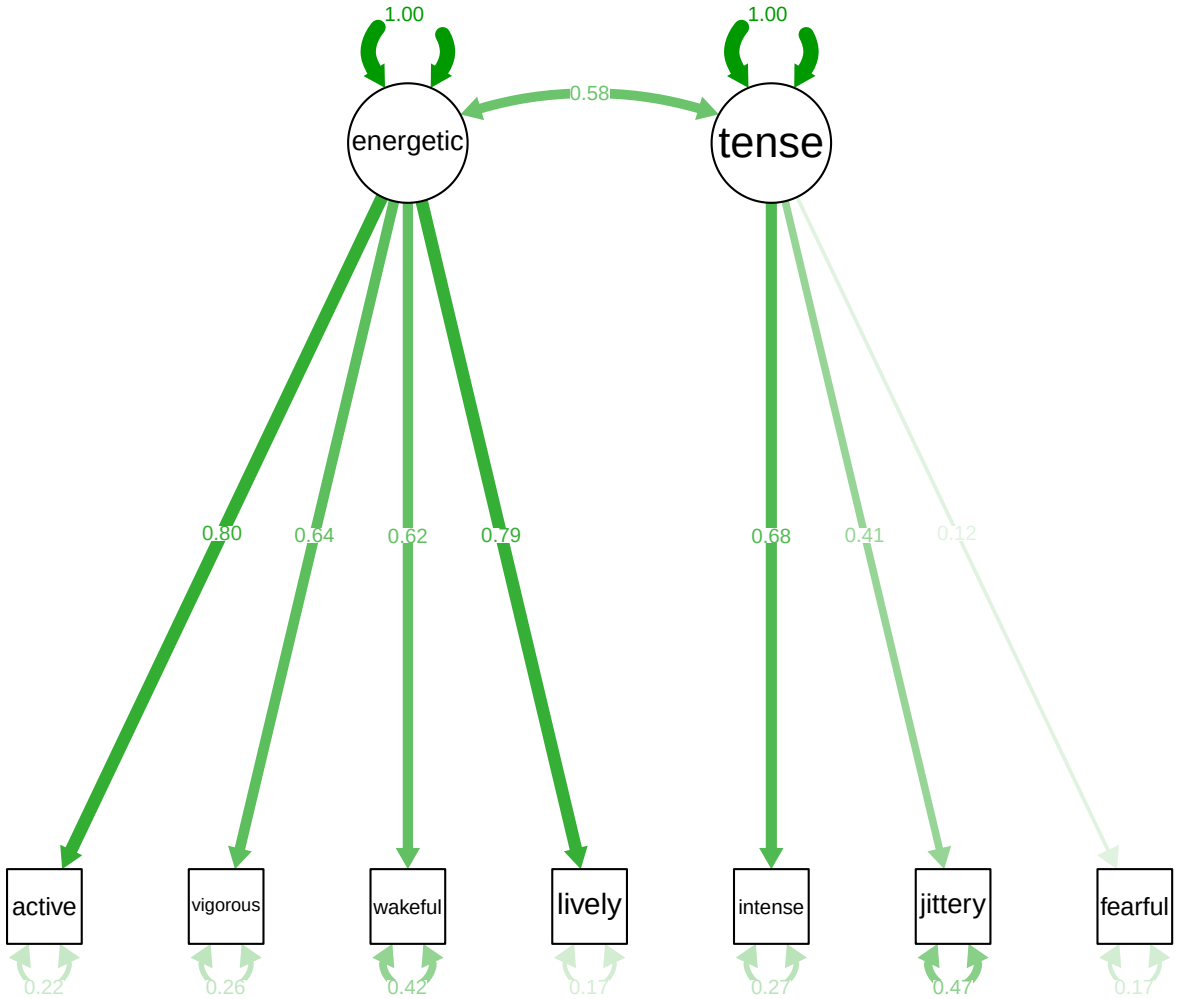


lavaan: Example

```
fit1 <- cfa(mod, data = msq, std.lv = TRUE)
```

The std.lv option will use a fixed factor method of identification

lavaan: Example



lavaan: Example

- Model results can be inspected with summary, nested models can be compared with anova

```
Estimator                                         ML
  Minimum Function Test Statistic           323.554
  Degrees of freedom                             13
  P-value (Chi-square)                        0.000

User model versus baseline model:

  Comparative Fit Index (CFI)                 0.970
  Tucker-Lewis Index (TLI)                    0.952

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)          -26806.206
  Loglikelihood unrestricted model (H1)  -26644.430
```

lavaan: Example

```
Root Mean Square Error of Approximation:

  RMSEA                                          0.078
  90 Percent Confidence Interval          0.071  0.086
  P-value RMSEA <= 0.05                          0.000

Standardized Root Mean Square Residual:

  SRMR                                           0.048
```

lavaan: Example

	Estimate	Std.err	Z-value	P(>|z|)
Latent variables:				
energetic =~				
active	0.798	0.012	64.708	0.000
vigorous	0.639	0.011	56.253	0.000
wakeful	0.614	0.013	46.768	0.000
lively	0.785	0.012	67.324	0.000
tense =~				
intense	0.681	0.019	35.674	0.000
jittery	0.413	0.016	25.531	0.000
fearful	0.121	0.008	14.830	0.000
Covariances:				
energetic ~~				
tense	0.579	0.017	34.129	0.000

lavaan: Example Multiple Groups

- Two factor CFA based on the msq daa (from the psych package)
  - Compare the msq with the revised msq

```
mod <- '
energetic =~ active + vigorous + wakeful + lively
tense =~ intense + jittery + fearful
'

fitg <- cfa(mod, data = msq, group = "scale")
```

lavaan: Example Multiple Groups

- Parameters can constrained with group.equal option

```
fitgW <- cfa(mod, data = msq, group = "scale",
             group.equal = "loadings")

anova(fitg, fitgW)
```

```
## Chi Square Difference Test
##
##       Df   AIC   BIC  Chisq Chisq diff Df diff Pr(>Chisq)
## fitg  26 52979 53255 343.51
## fitgW 31 52978 53222 352.36      8.851       5     0.1152
```

lavaan: Example Multiple Groups

- Or invariance testing can be conducted using the measurementInvariance function
  - In the semTools package

```
measurementInvariance(mod, data = msq, group = "scale")
```

lavaan: Example Multiple Groups

```
Model 1: configural invariance:
    chisq        df    pvalue       cfi     rmsea       bic
  343.506    26.000     0.000     0.970     0.080 53254.619

Model 2: weak invariance (equal loadings):
    chisq        df    pvalue       cfi     rmsea       bic
  352.357    31.000     0.000     0.969     0.073 53222.196

[Model 1 versus model 2]
  delta.chisq      delta.df delta.p.value     delta.cfi
        8.851         5.000         0.115         0.000

Model 3: strong invariance (equal loadings + intercepts):
    chisq        df    pvalue       cfi     rmsea       bic
  376.702    36.000     0.000     0.967     0.070 53205.266
```

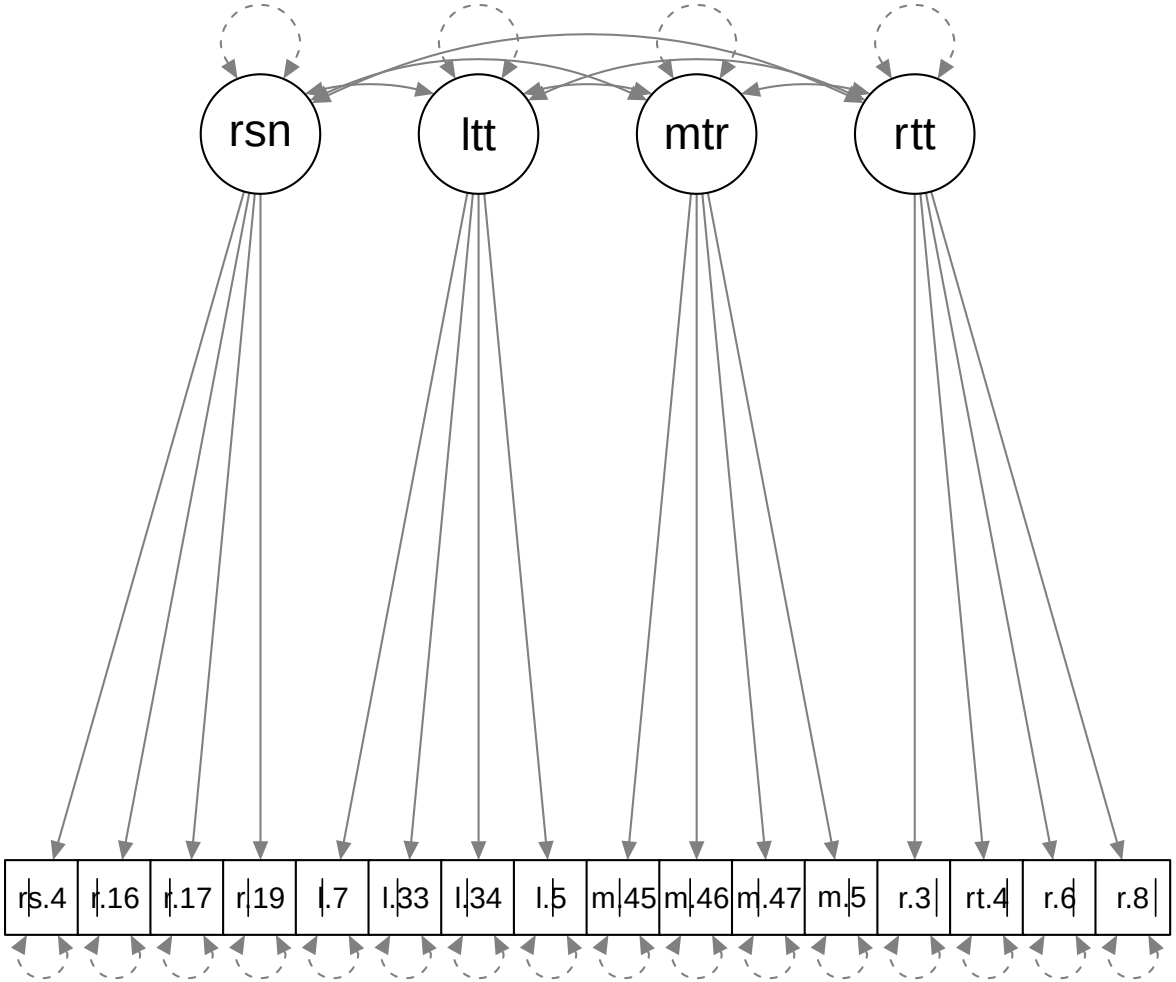
lavaan: Example Categorical Indicators

- Two factor CFA based on the ability data (from the psych package)

```
modCat <- '
reason =~ reason.4 + reason.16 + reason.17 + reason.19
letter =~ letter.7 + letter.33 + letter.34 + letter.58
matrix =~ matrix.45 + matrix.46 + matrix.47 + matrix.55
rotate =~ rotate.3 + rotate.4 + rotate.6 + rotate.8
'

fitCat <- cfa(modCat, data = ability, std.lv = TRUE,
              ordered = names(ability))
```

lavaan: Example Categorical Indicators



lavaan: Extensions

- There are many packages providing additional functionality to lavaan:
  - semPlot - Path diagrams for lavaan (used to make the above plots)
  - semTools - useful functions for SEM in R
    - ★ Includes functions for latent interactions, multiple imputation and more
  - simsem - simulation package for SEM
  - lavaan.survey - apply survey weights, clustering corrections and other corrections for lavaan models
  - Onyx - graphical user interface for SEM (`http://onyx.brandmaier.de/`)

lavaan: Resources

- The lavaan website is extremely helpful (lavaan.org)
  - It include a tutorial in lavaan
  - Links to examples from popular SEM books in lavaan
  - Link to a discussion board about lavaan

Thank you!

- Questions?
- email: schoemanna@ecu.edu
- Slides from today at: