Piecewise SEM Model Intro

Rosalee Elting

2023-11-08

SEM Overview

An Sem is a model that interprets the regression between a series of variables and their relative impacts on a downstream response. These responses could be behavioral (decision making, phenotype-performance regression), ecological (succession of plants, community impacts of one community species), or many other types of responses. Structural Equation Models (SEM) are useful for this type of work. These models utilize **a priori** knowledge of proposed relationships between data collected in an experiment to organize data into a “structured”equation. This type of model is often represented in a path model, with regressions between variables serving as parameters in the model (Shipley, 2016). An extension upon the standard SEM is the piecewise SEM, where each relationship between variables is independently evaluated (Lefcheck, 2021).

What SEM can and cannot do:

Yes!	No...
Falsify a causal hypothesis	Prove causal relationships
Determine if causal inferences are consistent with data	—–
Tell us if the assumptions we make may be valid	—–

Variables

Endogenous and Exogenous In a piecewise SEM model, data can be separated into endogenous and exogenous variables. **Exogenous variables** are those that are solely predictors and are only the beginnings of paths in a SEM. No other data predicts these data. On the other hand, **endogenous variables** are those that are predicted. These variables themselves can be predictors of tertiary or quaternary variables, but by definition are predicted by exogenous variables. The connection between these two types of variables are regression or correlation coefficients.

An example from my work Exogenous variables for this experiment are the following: body mass, wing aspect ratio (Equation 1) , burst muscle capacity (or load-lifting), and whether an individual was in a flight arena by themselves or with a competitor.

Composite Variables Sometimes there won’t be a single variable that can be used to assess an outcome. This is especially true anytime we use words like “performance” or “quality”. That doesn’t mean that these things cannot be quantified! It just means that these are often many variables summarized into a single values. If you want to get into composite variables, Lefcheck’s online book is really helpful!

An example from my work I assess these fifteen flight kinematic measurements as a composite variable, hereafter referred to as “flight performance” . Flight performance is a result of fifteen manifested flight kinematics and is the result of their combined impacts. In this model, flight performance is an endogenous variable that does not have any resulting variables. It is the final variable in the path, though continued work may use this variable as a predictor of fitness. As a result, it is unknown the relative importance of each individual flight kinematic on the overall observed flight performances. Therefore a fixed composite was used for this model, where each flight kinematic's impact was assessed as equally impacting the observed flight performance (Grace, 2010).

Variance and Covariance

- want to define the relatedness between variables
- *Variance:* degree of spread in the data, deviation of each point from the mean, always positive because it is a squared value
- *Covariance:* measure of dependency between the two variables. Higher value means more relatedness between the variables and they co-vary.

```r
x <- c(1,2,3,4)
y <- c(2,3,4,5)
var(x)
```

```
## [1] 1.666667
```

```r
cov(x,y)
```

```
## [1] 1.666667
```

Transforming Data

But what if my variance and covariance are wildly different? This could be explained by large differences in the units of the two variables. To control for this, we can standardize the variables using a Z-transformation. The formula for Z-transfomation is as follows:

$Z_x = \frac{X_i - \bar{X}}{sd(X)}$ where: X_i is the observed X value $\bar{X}$ is the mean value of the X variable

This can be done in R with the following base R code:

```r
scale(x)
```

```
##            [,1]
## [1,] -1.1618950
## [2,] -0.3872983
## [3,]  0.3872983
## [4,]  1.1618950
## attr(,"scaled:center")
## [1] 2.5
## attr(,"scaled:scale")
## [1] 1.290994
```

Regression Coefficients

To understand the partial effects of each predictor on an outcome, we need to remove the other predictor's influence on the first predictor, as well as the outcome. This is a lot of math, but is done in the psem() function in R.

An example from my work This is complicated and again, probably worth mentioning that you won't actually have to do it. But for calrity sake, here is an example of what this would look like. Here, I want to look at the *impact of body mass on flight performance, removing the effect of the other variables.* The variables: Bm = Body mass Mc = Muscle capacity Ar = Wing aspect ratio C = Competition Fp = Flight performance

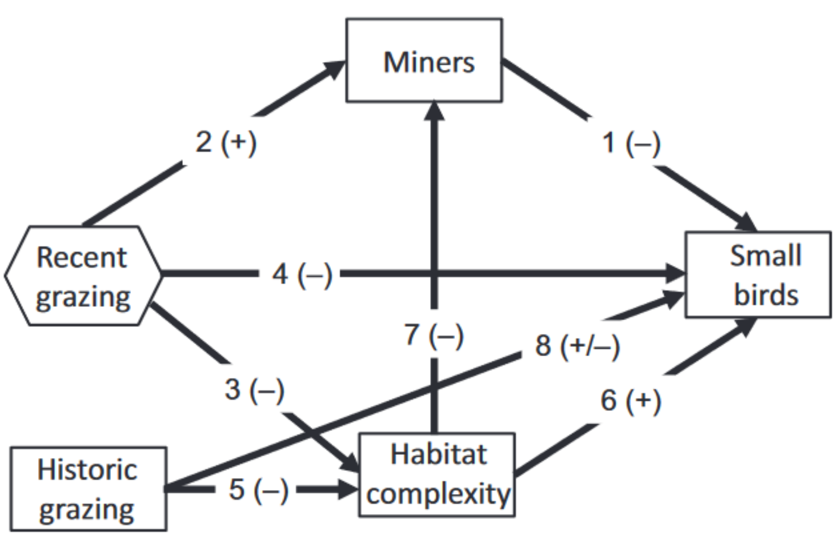
To accomplish this removal we use this formula: $b_{BMFP} = \frac{r_{BmFp} - (r_{BmMc} * r_{McFp} * r_{BmAr} * r_{ArFp} * r_{BMc} * r_{CFp})}{1 - r^2_{BmMc} - r^2_{BmAr} - r^2_{BmC}}$ WheW! A lot, right?

Regression vs. Correlation Coefficients Typically when we have relationships between two variables, we get regression coefficients (think the slope of a line). But, when we have standardized data, these regression coefficeints are also correlation coefficients. These will create the paths between variables in the SEM model.

A Tour through SEM

Val et al. 2018, Livestock grazing reinforces the competitive exclusion of small-bodied birds by large aggressive birds This is a paper looking at bird communities and the effect of previous grazing on the community assemblage of (large and small) birds.

```
val_apriori <- "sem_img/val_apriori.png"
knitr::include_graphics(val_apriori)
```



This is based on *previous literature* to establish relationships between each variable. And here is one of the results, with paths that are solid representing positive relationships (correlation coefficients), and negative ones represented with dashed, red lines. Note that not every relationship from the a priori model is included in this path diagram.

Example Building an SEM

Sometimes it helps to just do it, so here is some code to play around with building an SEM. This was written by Daniel Laughlin, 2022, obtained from Bob Hall.

Load Libraries

```
library(piecewiseSEM)
```

```
## Warning: package 'piecewiseSEM' was built under R version 4.3.2
```

```
##
##   This is piecewiseSEM version 2.3.0.
##
##
##   Questions or bugs can be addressed to <LefcheckJ@si.edu>.
```

```
library(semPlot)
```

```
## Warning: package 'semPlot' was built under R version 4.3.2
```

Next, we'll simulate a multiple regression relationship, where multiple variables affect an outcome. Here, we create data under the assumption that pine and litter are independent, and they both negatively affect grass biomass.

Create data set

```
pine <- rnorm(100)
litter <- rnorm(100)
grass <- -litter - pine + rnorm(100)

simdat <- data.frame(pine=pine, litter=litter, grass=grass)
```

Next we'll create a model (from a priori data) to inform R, specifically the psem() function, what we expect the relationship between these variables to be.

Specify, summarize, and plot model

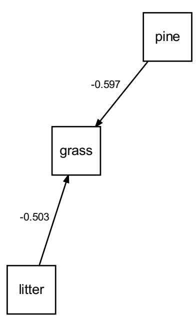
```
sem1<- psem(
    lm(grass ~ pine + litter, data=simdat)
)

summary(sem1, data=simdat)
```

```
##
## Structural Equation Model of sem1
##
## Call:
##   grass ~ pine + litter
##
##     AIC
##  283.690
```

```
##
## ---
## Tests of directed separation:
##
##  No independence claims present. Tests of directed separation not possible.
##
## --
## Global goodness-of-fit:
##
## Chi-Squared = 0 with P-value = 1 and on 0 degrees of freedom
## Fisher's C = NA with P-value = NA and on 0 degrees of freedom
##
## ---
## Coefficients:
##
##   Response Predictor Estimate Std.Error DF Crit.Value P.Value Std.Estimate
##      grass      pine  -0.9237    0.0897 97   -10.2944       0      -0.5968 ***
##      grass    litter  -0.8928    0.1028 97    -8.6830       0      -0.5033 ***
##
##   Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
##
## ---
## Individual R-squared:
##
##   Response method R.squared
##      grass   none      0.68
```

```r
plot(sem1)
```



Next, we can simulate a second model, where it is possible to have paths between varaibles beyond the output variable. This is far more realistic, and can be again informed by an a priori model. This model assumes all our variables are not aligned linearly, but rather have a "web" format. Here, we create data that assumes pine is independent, litter is positively impacted by pine, and grass is negatively impacted by both pine and litter.

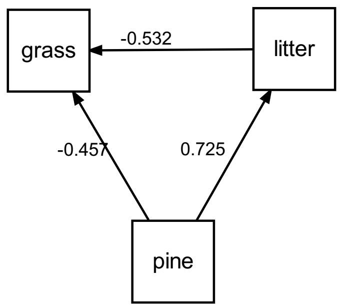
```r
pine <- rnorm(100)
litter <- pine +rnorm(100)
grass <- -litter -pine + rnorm(100)
simdat2 <- data.frame(pine=pine, litter=litter, grass=grass)

sem2 <- psem(
  lm(grass ~ pine + litter, data = simdat2),
  lm( litter~ pine, data= simdat2)
)

summary(sem2, data=simdat2)
```

```
##
## Structural Equation Model of sem2
##
## Call:
##   grass ~ pine + litter
##   litter ~ pine
##
##     AIC
##  578.029
##
## ---
## Tests of directed separation:
##
##  No independence claims present. Tests of directed separation not possible.
##
## --
## Global goodness-of-fit:
##
## Chi-Squared = 0 with P-value = 1 and on 0 degrees of freedom
## Fisher's C = NA with P-value = NA and on 0 degrees of freedom
##
## ---
## Coefficients:
##
##   Response Predictor Estimate Std.Error DF Crit.Value P.Value Std.Estimate
##      grass      pine  -1.1249    0.1427 97    -7.8838       0      -0.4571 ***
##      grass    litter  -0.9759    0.1063 97    -9.1813       0      -0.5324 ***
##     litter      pine   0.9738    0.0933 98    10.4334       0       0.7254 ***
##
##   Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
##
## ---
## Individual R-squared:
##
##   Response method R.squared
##      grass   none      0.85
##     litter   none      0.53
```

```r
plot(sem2)
```



Helpful Papers

Books

Lefcheck's Online Book: https://jslefche.github.io/sem_book/global-estimation.html

Shipley, B. 2016. Cause and Correlation in Biology. Cambridge University Press.

Papers about SEM

Grace, J.B., 2010. Structural equation modeling and natural systems. Cambridge University Press. Lefcheck JS (2016). "piecewiseSEM: Piecewise structural equation modeling in R for ecology, evolution,and systematics." *Methods in Ecology and Evolution*, *7*(5), 573-579. doi:10.1111/2041-210X.12512https://doi.org/10.1111/2041-210X.12512.

Papers that use SEM

Val et al. 2018. Livestock grazing reinforces the competitive exclusion of small-bodied birds by large aggressive birds. *Journal of Applied Ecology*, *55*(4), 1919-1929. doi: 10.1111/1365-2664.13078