Advanced Topics in SEM for Ecology & Evolutionary Biology

Jarrett E. K. Byrnes
UMass Boston
An Advanced Outline

1. Revisiting Sample Size
2. Revisiting Dsep in lavaan
3. Analysis of Nested Survey Data
4. Multilevel Generalized Piecewise SEM
5. Additional Spatial Techniques
6. Panel Models for Lagged Time Effects
7. Growth Curve Models & Time Series
Revisiting Sample Size

1. The further you are in a model from an exogenous data-generating, the weaker it's influence.

2. Our ability to detect the these tapering effect sizes is proportional to our information (especially sample size) and the number of parameters being estimated.

3. Our sample size sets an upper limit for the complexity of the model we can obtain.

4. Rules of thumb for sample size -- we hope to have at least 5 samples per estimated parameter and would prefer 20 samples per parameter.

5. Path coefficients add to our parameter list, not the variances
There are a total of 12 parameters shown. However, only 6 of these require unique information...

Number of Estimated Parameters

- Chi-sqr = 5.147; df = 3; p = .161
- AIC = 29.147; NPAR = 12
Parameters Needing Unique Information

Variances & covariance of exogenous variables can be obtained from the data. For “pesticide”, “Macroalgae”, and “Grass”, this yields 4 parameters.

Error variances (and R-sqrs) for endogenous variables are calculated from other parameters. This is 2 parameters.

Only 6 parameters require unique information.
Samples/parameters = 40/6 = 6.7.

Chi-sqr = 5.147; df = 3; p = .161
AIC = 29.147; NPAR = 12
For our more complex model, we would want to set non-contributing paths to zero to minimize estimated parameters.

Here estimated parameters = 8, samples/parameters = 5.

If we can combine Caprellids and Gamarids, we could reduce parameters further.
An Advanced Outline

1. Revisiting Sample Size
2. Revisiting Dsep in lavaan
3. Analysis of Nested Survey Data
4. Multilevel Generalized Piecewise SEM
5. Additional Spatial Techniques
6. Panel Models for Lagged Time Effects
7. Growth Curve Models & Time Series
1. $\chi^2$ gives you information regarding the discrepancy between your observed and predicted covariance matrix.

2. The test of D-Separation gives you information regarding whether you have missed key associations between variables.

3. We can test for D-Separation in recursive models without correlated error simply
D-Separation in lavaan

Two options
1) Feed model to DAG in ggm

#Full Mediation

distModel2 <- 'rich ~ abiotic + hetero
  hetero ~ distance
  abiotic ~ distance'
D-Separa8on in lavaan

2) Use script (and this will be in future lavaan versions)

```
> source("./dsepTest.R")
> dsepTest(distFit2)
$ctest
[1] 21.86173

$df
[1] 4

$pvalue
[1] 0.0002135289
```
D-Separation in lavaan

\[
\text{dsepTest}(\text{distFit2, showall}=\text{T}) \\
\text{ctest} \\
[1] 21.86173 \\
\text{df} \\
[1] 4 \\
\text{pvalue} \\
[1] 0.0002135289 \\
\text{dsep} \\
\begin{array}{ccc}
\text{Pair} & \text{Conditioning} & \text{P.t.} \\
\text{distance} & \text{distance,rich hetero,abiotic} & 9.564005e-05 \\
\text{abiotic} & \text{abiotic,hetero} & 1.871306e-01 \\
\end{array}
\]
An Advanced Outline

1. Revisiting Sample Size
2. Revisiting Dsep in lavaan
3. Analysis of Nested Survey Data
4. Multilevel Generalized Piecewise SEM
5. Additional Spatial Techniques
6. Panel Models for Lagged Time Effects
7. Growth Curve Models & Time Series
How might structure in surveys affect SEM results?

• Parameter estimates & variation biased
  – Depends on structure of sampling

• Parameter estimates & variation inconsistent
  – Dependent on groups with largest range of variation and/or sample size

• Covariance matrix may also be biased

• Fit measures likely incorrect
Clusters and Strata in Surveys

• Strata: levels that divide a population into different classes
  – E.g. age classes of an organism, grades in high school, income level

• Clusters
  – Blocks sampled at random
  – Not all blocks may be sampled
  – Can be levels of nesting
Example Survey

Sample of Algal Composition

- Islands randomly sampled – CLUSTER
- West/East side of island - STRATA
Example Survey

Sample of High School Test Scores

• District = CLUSTER
• No strata
• UNLESS: grade = STRATA
How do we Correct our Estimation?

1. Create a weight matrix based on the survey structure and then either

2. Correct error and $\chi^2$ of model fit using a corrected estimator or robust correction (e.g., S-B)

Or

3. Use weight matrix with WLS estimation
The lavaan.survey package

• Blends lavaan fits & survey structure

• Survey structure described using survey package

• Produces models with corrected errors and fit statistics
Survey Package

• Designed to provide robust estimation & SE under clustered, stratified surveys

• [http://r-survey.r-forge.r-project.org/survey/index.html](http://r-survey.r-forge.r-project.org/survey/index.html)


svydesign to Describe Survey

```
svydesign(
  ids =~ district,
  strata=grade,
  probs =~1, data=d)
```

probs means schools from each cluster have an equal probability of being sampled in the survey
A Multi-Stream Experiment

Also wells grouped in blocks within a stream

Cardinale et al 2009
Nonlinear Relationship Between Nutrient Addition and Richness

Cardinale et al 2009
Note that Treatment's Don't Covary with Regional Richness

\[
\text{cardModel} <- \'
\begin{align*}
\text{SA} & \sim \text{logN} + \text{logNcen2} + \text{SR} \\
\text{logChl} & \sim \text{SA} + \text{logN} \\
\text{logN} & \sim \text{logNcen2}'
\end{align*}
\]

cardFit <- sem(cardModel, data=cards, fixed.x=F, estimator="MLM")
But...no Nutrient Effect?

Was the lack of an effect all just due to collinearity?

Did clustering obscure the effect?

Regressions:

| SA ~           | Estimate | Std.err | Z-value | P(> |z|) | Std.lv | Std.all |
|---------------|----------|---------|---------|------|-------|---------|
| logN          | -2.996   | 2.960   | -1.012  | 0.311| -2.996| -0.504  |
| logN2         | -0.475   | 0.456   | -1.041  | 0.298| -0.475| -0.506  |
| SR            | 0.384    | 0.035   | 10.859  | 0.000| 0.384 | 0.688   |
| logChl ~      |          |         |         |      |       |         |
| SA            | 0.020    | 0.005   | 3.768   | 0.000| 0.020 | 0.393   |
| logN          | 0.097    | 0.023   | 4.194   | 0.000| 0.097 | 0.321   |

Cardinale et al 2009
The Survey Design

cardSrvy <- svydesign(
  ids=~Stream+Sub,
  probs=~1,
  data=cards)

Also wells grouped in blocks within a stream
But...no Nutrient Effect?

fit.card.survey <- lavaan.survey(cardFit, cardSrvy, estimator = "MLM")

Colinearity still problematic in separating linear v. nonlinear effect – but, there is an N->SA effect!
Exercise: Shipley’s Nested Data

- Simulated data from a fit model
- 20 sites
- 5 trees measured per site

```r
> Shipley <- read.table("./Shipley.dat")
> head(Shipley)
```

<table>
<thead>
<tr>
<th>site</th>
<th>tree</th>
<th>lat</th>
<th>year</th>
<th>Date</th>
<th>DD</th>
<th>Growth</th>
<th>Survival</th>
<th>Live</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>40.38063</td>
<td>1970</td>
<td>115.4956</td>
<td>160.5703</td>
<td>61.36852</td>
<td>0.9996238</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>40.38063</td>
<td>1970</td>
<td>118.4959</td>
<td>158.9896</td>
<td>43.77182</td>
<td>0.8433521</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>40.38063</td>
<td>1970</td>
<td>115.8836</td>
<td>159.9262</td>
<td>44.74663</td>
<td>0.9441110</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>40.38063</td>
<td>1970</td>
<td>110.9889</td>
<td>161.1282</td>
<td>48.20004</td>
<td>0.9568525</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>40.38063</td>
<td>1970</td>
<td>120.9946</td>
<td>157.3778</td>
<td>50.02237</td>
<td>0.9759584</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>40.38063</td>
<td>1972</td>
<td>114.2315</td>
<td>160.6120</td>
<td>56.29615</td>
<td>0.9983398</td>
<td>1</td>
</tr>
</tbody>
</table>
Nested Structure in the Data
Exercise: Shipley’s Nested Data

- Simulated data from a fit model
- 20 sites
- 5 trees measured per site

```r
> Shipley<-read.table("./Shipley.dat")
> head(Shipley)

     site tree lat year Date DD Growth Survival Live
  1      1    1 40.38063 1970 115.4956 160.5703 61.36852 0.9996238    1
  2      1    2 40.38063 1970 118.4959 158.9896 43.77182 0.8433521    1
  3      1    3 40.38063 1970 115.8836 159.9262 44.74663 0.9441110    1
  4      1    4 40.38063 1970 110.9889 161.1282 48.20004 0.9568525    1
  5      1    5 40.38063 1970 120.9946 157.3778 50.02237 0.9759584    1
  6      1    1 40.38063 1972 114.2315 160.6120 56.29615 0.9983398    1
```
Exercise: Shipley’s Nested Data

shipSurv <- svydesign(ids =~site + tree,
                       probs=~1,
                       data=Shipley)
Exercise: Shipley’s Nested Data

\[
\begin{align*}
\text{Latitude} & \quad \rightarrow \\
\text{Degree Days} & \quad \rightarrow \\
\text{Date of Bud Burst} & \quad \rightarrow \\
\text{Diameter Growth} & \quad \rightarrow \\
\text{Survival} &
\end{align*}
\]

\[
\begin{array}{c}
R^2=0.50 \\
-0.78 \\
R^2=0.43 \\
-0.65 \\
R^2=0.04 \\
0.22 \\
R^2=0.33 \\
0.57
\end{array}
\]

\[
\text{shipFit} \leftarrow \text{sem(shipMod1, data=Shipley, estimator="MLM")}
\]

\[
\text{shipCorrect} \leftarrow \text{lavaan.survey(shipFit, shipSurv)}
\]
An Advanced Outline

1. Revisiting Sample Size
2. Revisiting Dsep in lavaan
3. Analysis of Nested Survey Data
4. Multilevel Generalized Piecewise SEM
5. Additional Spatial Techniques
6. Panel Models for Lagged Time Effects
7. Growth Curve Models & Time Series
D-Separation in Piecewise models beyond linear regression

1. We have models that deal with
   1. Hierarchical/nested data (mixed models)
   2. Nonlinear relationships
   3. Non-normal error distributions (glms)

2. The test of the effect of a variable in one of those models serves the same purpose as a partial correlation test in a linear model

3. These p-values can be used for tests of D-Separation

The Simulated Data
Nested Structure in the Data
# Piecewise Hierarchical Model Fitting

![Diagram showing the flow from lat to DD to Date to Growth to Live.]

#e.g., for DD -> lat
Shipley<-read.table("./Shipley.dat")
library(nlme)

#model with random intercept
#tree nested in site
Date_dd<-lme(Date~DD,data=Shipley,
    random=~1|site/tree,na.action=na.omit)
The Basis Set Needs to Accommodate the Nested Structure

To calculate the partial regression slope, use hierarchical models

<table>
<thead>
<tr>
<th>D-sep claim of independence</th>
<th>Mixed model†</th>
<th>Variable whose partial regression slope should be zero</th>
<th>Null probability (distribution)</th>
</tr>
</thead>
<tbody>
<tr>
<td>((X_1, X_3)</td>
<td>{X_2}</td>
<td></td>
<td>(X_3 \sim X_2 + X_1 + (1</td>
</tr>
<tr>
<td>((X_1, X_4)</td>
<td>{X_3}</td>
<td></td>
<td>(X_4 \sim X_3 + X_1 + (1</td>
</tr>
<tr>
<td>((X_1, X_3))</td>
<td>{X_4}</td>
<td></td>
<td>(X_5 \sim X_4 + X_1 + (1</td>
</tr>
<tr>
<td>((X_2, X_4))</td>
<td>{X_1, X_3}</td>
<td></td>
<td>(X_4 \sim X_3 + X_1 + X_2 + (1</td>
</tr>
<tr>
<td>((X_2, X_3))</td>
<td>{X_1, X_4}</td>
<td></td>
<td>(X_5 \sim X_4 + X_1 + X_2 + (1</td>
</tr>
<tr>
<td>((X_3, X_4))</td>
<td>{X_2, X_4}</td>
<td></td>
<td>(X_5 \sim X_4 + X_2 + X_3 + (1</td>
</tr>
</tbody>
</table>
Evaluate Independence Claims with Hierarchical Models

#Independence claim: (Date,lat)|{DD}

fit1<-lme(Date~DD+lat,data=Shipley,
    random=~1|site/tree,na.action=na.omit)

summary(fit1)$tTable
Evaluate Independence Claims with Hierarchical Models

#Independence claim: (Date, lat) | {DD}

fit1 <- lme(Date ~ DD + lat, data = Shipley,
random = ~ 1 | site/tree, na.action = na.omit)
Evaluate Independence Claims with Hierarchical Models

> summary(fit1)$tTable

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std.Error</th>
<th>DF</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>198.915223483</td>
<td>7.337099813</td>
<td>1330</td>
<td>27.11087876</td>
<td>3.185667e-129</td>
</tr>
<tr>
<td>DD</td>
<td>-0.497660383</td>
<td>0.004936809</td>
<td>1330</td>
<td>-100.80608521</td>
<td>0.000000e+00</td>
</tr>
<tr>
<td>lat</td>
<td>-0.009051378</td>
<td>0.113476607</td>
<td>18</td>
<td>-0.07976426</td>
<td>9.373049e-01</td>
</tr>
</tbody>
</table>
We Have Nonlinear Relationships with Non-Normal Distributions

Use generalized linear models – e.g., logit curve with a binomial error
Evaluate Independence Claims with GLMMs

###need lme4 for the glmmns
library(lme4)

#Independence claim with glmm (Live,lat)|{Growth}
fit4<-lmer(Live~Growth+lat+(1|site)+(1|tree),
data=Shipley, na.action=na.omit,
family=binomial(link="logit"))
Evaluate Independence Claims with GLMMs

> summary(fit4)@coefs

|                | Estimate | Std. Error | z value | Pr(>|z|)   |
|----------------|----------|------------|---------|-----------|
| (Intercept)    | -14.4384 | 2.6539     | -5.4403 | 5.3174e-08|
| Growth         | 0.3553   | 0.0455     | 7.8012  | 6.1304e-15|
| lat            | 0.0305   | 0.0282     | 1.0823  | 2.7911e-01|
Putting it All Together in Shipley's Test

```
#note, since we're logging things
#we can use log(a)+log(b) = log(a*b)

> fisherC <- -2* log(9.373049e-01 * 3.836896e-01 *
                      7.667083e-01 * 2.791099e-01 *
                      3.159286e-01 * 1.519170e-01)

> fisherC
[1] 11.20225

> 1-pchisq(fisherC, 2*6)
[1] 0.5116698
```
AIC and D-Sep

AIC = C + 2K

Why? Shipley has proven that:

\[-2 \ln(L(\text{model} \mid \text{data})) = -2 \sum \ln(p) = \text{Fisher's C}\]

Shipley, B. In Press. The AIC model selection method applied to path analytic models compared using a d-separation tests. Ecology.
#each piece except last has 5 parameters - slope, # intercept, variance, and random variance for #slope & intercept, so, K=5*4-1

> fisherC + 2*(5*4-1)
[1] 49.20225
Using `piecewiseSEM` for GLMMs

```
library(piecewiseSEM)

shipley2009.modlist = list(
  lme(DD~lat, random = ~1|site/tree, na.action = na.omit, 
       data = Shipley),

  lme(Date~DD, random = ~1|site/tree, na.action = na.omit, 
       data = Shipley),

  lme(Growth~Date, random = ~1|site/tree, na.action = na.omit, 
       data = Shipley),

  glmer(Live~Growth+(1|site)+(1|tree), 
       family=binomial(link = "logit"), data = Shipley) )
```
Using piecewiseSEM for GLMMs

```
> sem.fit(shipley2009.modlist, data=Shipley)
```

$missing.paths

<table>
<thead>
<tr>
<th>missing.path</th>
<th>estimate</th>
<th>std.error</th>
<th>DF</th>
<th>crit.value</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Date &lt;- lat</td>
<td>-0.009</td>
<td>0.113</td>
<td>18</td>
<td>-0.080</td>
<td>0.937</td>
</tr>
<tr>
<td>2 Growth &lt;- lat</td>
<td>-0.099</td>
<td>0.111</td>
<td>18</td>
<td>-0.893</td>
<td>0.384</td>
</tr>
<tr>
<td>3 Live &lt;- lat</td>
<td>0.030</td>
<td>0.030</td>
<td>NA</td>
<td>1.028</td>
<td>0.304</td>
</tr>
<tr>
<td>4 Growth &lt;- DD</td>
<td>-0.011</td>
<td>0.036</td>
<td>1329</td>
<td>-0.297</td>
<td>0.767</td>
</tr>
<tr>
<td>5 Live &lt;- DD</td>
<td>0.027</td>
<td>0.027</td>
<td>NA</td>
<td>1.004</td>
<td>0.315</td>
</tr>
<tr>
<td>6 Live &lt;- Date</td>
<td>-0.047</td>
<td>0.030</td>
<td>NA</td>
<td>-1.562</td>
<td>0.118</td>
</tr>
</tbody>
</table>
Using piecewiseSEM for GLMMs

\[ \text{lat} \rightarrow \text{DD} \rightarrow \text{Date} \rightarrow \text{Growth} \rightarrow \text{Live} \]

\[
> \text{sem.fit(\text{shipley2009.modlist}, \text{data=Shipley})} \\
... \\
\]

\[
\text{Fisher.C} \\
\text{Fisher.C} \quad k \quad P \\
11.54 \quad 12.000 \quad 0.484 \\
... \\
\]
AIC from piecewiseSEM

```r
> sem.fit(shipley2009.modlist, data=Shipley)
...
```

$AIC$

<table>
<thead>
<tr>
<th></th>
<th>AIC</th>
<th>AICc</th>
<th>K</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>49.54</td>
<td>50.079</td>
<td>19</td>
<td>1431</td>
</tr>
</tbody>
</table>

...
What do we gain from random effects in piecewiseSEM?

# Want to compare to a lavaan fit?
> sem.lavaan(shipley2009.modlist, data=Shipley)

lavaan (0.5-20) converged normally after 27 iterations

<table>
<thead>
<tr>
<th></th>
<th>Used</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations</td>
<td>1431</td>
<td>1900</td>
</tr>
<tr>
<td>Estimator</td>
<td>ML</td>
<td></td>
</tr>
<tr>
<td>Minimum Function Test Statistic</td>
<td>38.433</td>
<td></td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>P-value (Chi-square)</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

Hierarchical structure & non-normality necessary for fit
## Variation Explained

- Mixed Models have two kinds of $R^2$
  - **Marginal**: $R^2$ due to fixed effects only
  - **Conditional**: $R^2$ due to fixed & random effects

```r
> sem.model.fits(shipley2009.modlist)

<table>
<thead>
<tr>
<th>Class</th>
<th>Family</th>
<th>Link</th>
<th>Marginal</th>
<th>Conditional</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>lme</td>
<td>gaussian</td>
<td>identity</td>
<td>0.4864825</td>
<td>0.6990231</td>
</tr>
<tr>
<td>2</td>
<td>lme</td>
<td>gaussian</td>
<td>identity</td>
<td>0.4095855</td>
<td>0.9838829</td>
</tr>
<tr>
<td>3</td>
<td>lme</td>
<td>gaussian</td>
<td>identity</td>
<td>0.1079098</td>
<td>0.8366353</td>
</tr>
<tr>
<td>4</td>
<td>glmerMod</td>
<td>binomial</td>
<td>logit</td>
<td>0.5589201</td>
<td>0.6291994</td>
</tr>
</tbody>
</table>
```
Exercise: Multilevel Model with Richness Model

- Use `piecewiseSEM` to fit and evaluate this model
- Use the squared term (may need to center)
- Stream and Sub nested within Stream are your random effects
Final Thoughts on Piecewise Fits

• You can use anything: generalizes linear models, mixed models, generalized least squares fits with temporal or spatial autocorrelation built-in

• Bayesian methods also provide flexible frameworks for piecewise models, but cannot calculate omnibus fit tests
An Advanced Outline

1. Revisiting Sample Size
2. Revisiting Dsep in lavaan
3. Analysis of Nested Survey Data
4. Multilevel Generalized Piecewise SEM
5. Additional Spatial Techniques
6. Panel Models for Lagged Time Effects
7. Growth Curve Models & Time Series
Spatial Effects

There are two key issues regarding space:

(1) Are there things to learn about the other factors that could explain variations in the data that vary spatially?

(2) Do we have nonindependence in our residuals?

Recent reference on the subject:
Spatial References

Reference where mechanistic questions have been asked:

(Distance from mouth of river and edge of shore served as proxies for past storm-driven saltwater intrusions.)


(Showed fine-scale matching of plant to abiotic conditions in severe environments. No evidence of mass effects.)
Spatial References

Reference where autocorrelation has been adjusted for in SEM studies:

Degrees of freedom and sample size adjusted using Moran's I.
Adjusting for Spatial Autocorrelation

1. Is there spatial autocorrelation?
   – Calculate Moran’s I on residuals

2. If yes, it may bias the SE of parameters
   – Calculate new effective sample size
   – Recalculate parameter SE from VCOV matrix

3. Correct by recalculating SE and Z-tests
   – By hand...for now
Example: NDVI in a Boreal System
Data Contains Spatial Information

```r
> boreal <- read.table("./Boreality.txt", header=T)

> head(boreal)

       point       x       y Oxalis   boreal nBor nTot   Grn   NDVI    T61
1      1 2109.70 2093.52      0 15.38462    2   13 0.0597027 0.480180 296.367
2      2 2190.18 2105.71      1 19.04762    4   21 0.0514881 0.483990 296.367
3      3 2064.48 2052.77      1 20.00000    6   30 0.0509510 0.489213 296.367
4      4 2277.34 2103.42      0 15.38462    2   13 0.0521183 0.473226 296.367
5      5 2347.91 2074.81      0  0.00000    0   13 0.0422267 0.405898 296.785
6      6 2437.21 2086.95      0 16.66667    1    6 0.0417779 0.424769 296.367

    Wet nBroSq
1  -0.0264378      4
2  -0.0234048     16
3  -0.0189264     36
4  -0.0280431      4
5  -0.0292287      0
6  -0.0229209      1
```
Model of NDVI in a Boreal System

\[ \text{borModel} \leftarrow ' \]
\[ \text{NDVI} \sim \text{nTot} + \text{T61} + \text{Wet} \]
\[ \text{nTot} \sim \text{T61}' \]

\[ \text{borFit} \leftarrow \text{sem(borModel, data=boreal, meanstructure=T)} \]
Residuals Might be Spatially Correlated

Cool colors cluster in Center

borRes <- as.data.frame(residuals(borFit, "casewise"))
Calculating a Distance Matrix

1. Distance matrices tell us how close points are in space
   • ape library calculates matrix and Moran’s I

   ```
   library(ape)
   distMat <- as.matrix(dist(
       cbind(boreal$x, boreal$y)))
   ```

2. We take the inverse, as closer points have greater similarity
   • The diagonal is 0, as there is no correlation within a point

   ```
   distsInv <- 1/distMat
   diag(distsInv) <- 0
   ```
Moran’s I using Residuals

> mi.ndvi <- Moran.I(borRes$NDVI, distsInv)
> mi.ndvi
$observed
 [1] 0.08265236

$expected
 [1] -0.001879699

$sd
 [1] 0.003985846

$p.value
 [1] 0

Data is more spatially correlated than expected – need a correction
Adjusting the Effective Sample Size

• Estimate of effective sample size (Fortin & Dale 2005, p. 223, Equation 5.15):

\[
n' = \frac{n^2}{\sum_{i=1}^{n} \sum_{j=1}^{n} \text{Cor}(x_i, x_j)}
\]

• For first-order autocorrelation \( \rho \) and large \( n \):

\[
n' \approx n \cdot \frac{1 - \rho}{1 + \rho}
\]
Where did that SE Come From?

SE’s calculated as the square root of the variance-covariance matrix of parameters

```r
> vcov(borFit)
```

<table>
<thead>
<tr>
<th></th>
<th>NDVI~nT</th>
<th>NDVI~T6</th>
<th>NDVI~W</th>
<th>nT~T61</th>
<th>NDVI~~</th>
<th>nTt~~T</th>
<th>NDVI~1</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDVI~nTot</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDVI~T61</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDVI~Wet</td>
<td>0.000</td>
<td>0.000</td>
<td>0.017</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nTot~T61</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.298</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDVI~~NDVI</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nTot~~nTot</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>47.113</td>
<td></td>
</tr>
<tr>
<td>NDVI~1</td>
<td>0.000</td>
<td>-0.002</td>
<td>0.027</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.448</td>
</tr>
<tr>
<td>nTot~1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>-88.329</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Given that \( \text{Var}(x) = \frac{\text{sum}[(x-x')^2]}{n} \), all we need to do is change \( n \)!
Calculating a New Standard Error

```r
#New SE
ndvi.var <- diag(vcov(borFit))[1:3]

ndvi.se <- sqrt(ndvi.err*nrow(boreal)/n.ndvi)

ndvi.se

  NDVI~nTot  NDVI~T61  NDVI~Wet
0.0001848868 0.0024493462 0.1436405689

Compare to former SE – new SE is wider

> sqrt(diag(vcov(borFit)))[1:3]
  NDVI~nTot  NDVI~T61  NDVI~Wet
0.0001701878 0.0022546163 0.1322207383
```
Z-Tests Show Not All Paths Still ≠ 0

\[ \text{not all paths are significant} \]

\[ > z <- \text{coef(borFit)}[1:3]/\text{ndvi.se} \]

\[ > 2*\text{pnorm(abs}(z), \text{lower.tail}=F) \]

\[
\begin{align*}
\text{NDVI~nTot} & \quad \text{NDVI~T61} & \quad \text{NDVI~Wet} \\
5.366259e-02 & \quad 1.517587e-47 & \quad 3.404230e-194
\end{align*}
\]
Or...do it the easy way

\[ T61 \rightarrow nTot \]
\[ Wet \rightarrow NDVI \]
\[ T61 \rightarrow Wet \]
\[ nTot \rightarrow NDVI \]

\[ \xi_1 \]
\[ \xi_2 \]

\[
\begin{align*}
-0.04 \\
-0.35 \\
-0.71 \\
0.09 \\
-0.04
\end{align*}
\]

\[
\begin{align*}
\text{> library(semTools)} \\
\text{> spatialCorrect(borFit, boreal$x, boreal$y)}
\end{align*}
\]
Or...do it the easy way

$\text{Morans}_I$
$\text{Morans}_I$\text{NDVI}

<table>
<thead>
<tr>
<th></th>
<th>observed</th>
<th>expected</th>
<th>sd</th>
<th>p.value</th>
<th>n.eff</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.08265236</td>
<td>-0.001879699</td>
<td>0.003985846</td>
<td>0.451.6189</td>
<td></td>
</tr>
</tbody>
</table>

$\text{Morans}_I$\text{nTot}

<table>
<thead>
<tr>
<th></th>
<th>observed</th>
<th>expected</th>
<th>sd</th>
<th>p.value</th>
<th>n.eff</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.03853411</td>
<td>-0.001879699</td>
<td>0.003998414</td>
<td>0.493.4468</td>
<td></td>
</tr>
</tbody>
</table>

...
Or...do it the easy way

\[
\begin{align*}
\text{T61} & \quad 0.09 & \quad \text{nTot} & \quad \zeta_1 \\
\text{Wet} & \quad -0.35 & \quad \text{NDVI} & \quad \zeta_2 & \quad -0.71
\end{align*}
\]

\[
\begin{array}{cccccc}
\text{Parameter} & \text{Estimate} & \text{n.eff} & \text{Std.err} & \text{Z-value} & \text{P(>|z|)} \\
\hline
\text{NDVI-nTot} & -0.0003567484 & 451.6189 & 0.0001848868 & -1.92955 & 5.366259e-02 \\
\text{NDVI-T61} & -0.0354776273 & 451.6189 & 0.0024493462 & -14.48453 & 1.517587e-47 \\
\text{NDVI-Wet} & -4.2700526589 & 451.6189 & 0.1436405689 & -29.72734 & 3.404230e-194 \\
\text{NDVI-NDVI} & 0.0017298286 & 451.6189 & 0.0001151150 & 15.02696 & 4.889505e-51 \\
\text{NDVI-1} & 10.8696158663 & 451.6189 & 0.7268790958 & 14.95382 & 1.470754e-50
\end{array}
\]

\[
\begin{array}{cccccc}
\text{Parameter} & \text{Estimate} & \text{n.eff} & \text{Std.err} & \text{Z-value} & \text{P(>|z|)} \\
\hline
\text{nTot-nTot} & 1.170661 & 493.4468 & 0.5674087 & 2.063171 & 3.909634e-02 \\
\text{nTot-nTot} & 112.051871 & 493.4468 & 7.1336853 & 15.707431 & 1.345204e-55 \\
\text{nTot-1} & -322.936937 & 493.4468 & 168.1495917 & -1.920534 & 5.479054e-02
\end{array}
\]
An Advanced Outline

1. Revisiting Sample Size
2. Revisiting Dsep in lavaan
3. Analysis of Nested Survey Data
4. Multilevel Generalized Piecewise SEM
5. Additional Spatial Techniques
6. Panel Models for Lagged Time Effects
7. Growth Curve Models & Time Series
Problem of Non-Recursive Models

Longitudinal Studies – Time-Step (Panel) Model for Lagged Effects

An Advanced Outline

1. Revisiting Sample Size
2. Revisiting Dsep in lavaan
3. Multilevel Generalized Piecewise SEM
4. Additional Spatial Techniques
5. Panel Models for Lagged Time Effects
6. Growth Curve Models & Time Series
Latent Trajectory Models for Timeseries & Repeated Measures

Latent Trajectory Models for Repeated Measures

random slopes and intercepts

upper-level covariate

lower-level covariate
Means Structures: Acquiring Intercepts from SEM!

```
meanMod<-'Giant.Kelp ~ Purple.Urchins'
meanFit <- sem(meanMod, data=kfm, meanstructure=T)
```
Means Structures: Acquiring Intercepts from SEM!

Regressions:
- Giant.Kelp ~ Purple.Urchin: $-0.366, 0.029, -12.397, 0.000$

Intercepts:
- Giant.Kelp: $1.590, 0.076, 20.791, 0.000$

Variances:
- Giant.Kelp: $0.579, 0.045, 12.961, 0.000$
Latent Variable Growth Model

Initial Density

Kelp in Year 1
Kelp in Year 2
Kelp in Year 3
Kelp in Year 4

intercept
slope

ξ₁ ξ₂ ξ₃ ξ₄
Example: Channel Islands Kelp Dynamics

\[
\text{gMod<-'}
\]

\[
\text{Initial} \sim 1 \cdot \text{KelpT1} + 1 \cdot \text{KelpT2} + 1 \cdot \text{KelpT3} + 1 \cdot \text{KelpT4}
\]

\[
\text{Growth} \sim 0 \cdot \text{KelpT1} + 1 \cdot \text{KelpT2} + 2 \cdot \text{KelpT3} + 3 \cdot \text{KelpT4}
\]

\[
\text{gFit<-growth(gMod, data=kelpTseries)}
\]
Example: Channel Islands Kelp Dynamics

\[ \begin{align*}
R^2 &= 0.5 - 0.67 \\
1 &\quad 0.763
\end{align*} \]

**Conclusions:**

At minimum, no linear trajectory.

At most, kelp densities stay constant with some small variation

|                    | Estimate | Std.err | Z-value | P(>|z|) |
|--------------------|----------|---------|---------|---------|
| KelpT1             | 0.000    |         |         |         |
| KelpT2             | 0.000    |         |         |         |
| KelpT3             | 0.000    |         |         |         |
| KelpT4             | 0.000    |         |         |         |
| Initial            | 0.763    | 0.096   | 7.976   | 0.000   |
| Growth             | 0.027    | 0.032   | 0.837   | 0.403   |

Intercepts:

Initial Density

Kelp in Year 1

Kelp in Year 2

Kelp in Year 3

Kelp in Year 4

\[ \begin{align*}
\xi_1 &= 0 \\
\xi_2 &= 1 \\
\xi_3 &= 1 \\
\xi_4 &= 0 \\
\end{align*} \]
Growth Models and Autoregressive Relationship

Initial Density

Growth

Kelp in Year 1

Kelp in Year 2

Kelp in Year 3

Kelp in Year 4

$\zeta_1$ $\zeta_2$ $\zeta_3$ $\zeta_4$
Growth Models and Autoregressive Relationship

- $a = 0.266$
- Fit not different

\[ \begin{align*}
\zeta_1 & \to 1 & \zeta_2 & \to 1 & \zeta_3 & \to 1 & \zeta_4 & \to 1 \\
1 & \to 0 & 1 & \to 2 & 1 & \to 3 & 0 & \to 1 \\
\end{align*} \]

\[ a = 0.266 \]

\[ \text{Fit not different} \]

\[ 0.704 \]
Other Processes Affect Growth Curves

![Diagram showing the relationship between nutrient delivery, initial density, growth, and kelp in years 1 to 4.](image-url)
Other Processes Affect Growth Curves

Nutrient Delivery

- Initial Density
- Growth

Kelp in Year 1 → Kelp in Year 2 → Kelp in Year 3 → Kelp in Year 4

\( \zeta_1 \) \( \zeta_2 \) \( \zeta_3 \) \( \zeta_4 \)

\( a = 0.304 \)
Final Comments on Advanced Topics

1. Often, our concern for spatial and temporal effects is due to our deep ecological fear of pseudoreplication.

2. If you can account for the drivers that create spatial or temporal blocks, you gain information.

3. Many cases are more easily dealt with in a piecewise approach.

4. But, many special cases have techniques in the literature that YOU can now use!
Prediction in SEM Using lavaan
If I have new Distance values, what is richness?

- Predict in lavaan merely gives factor loadings
- To get predicted values, we need to think about the underlying structural equations!
If I have new Distance values, what is richness?

Richness ~ a*distance + b1*b*distance + c1*c*distance

But what about intercepts?

Intercept = ia + ib*b1 + ic*c1
Estimating Direct and Indirect Effects with Named Coefficients

totDistModel <- ' 
rich ~ a*distance + b*abiotic + c*hetero 
abiotic ~ b1*distance 
hetero ~ c1*distance 

rich ~ ia*1 
abiotic ~ ib*1 
hetero ~ ic*1 

direct:= a 
indirect:= b1*b + c1*c 
total:= b1*b + c1*c + a 
totalIntercept := il + ib*b1 + ic*c1 
'
Estimates of Compounded Effects

Defined Parameters:

| Parameter       | Estimate | Std.Err | Z-value | P(>|z|) |
|-----------------|----------|---------|---------|---------|
| direct          | 0.640    | 0.156   | 4.117   | 0.000   |
| indirect        | 0.360    | 0.103   | 3.488   | 0.000   |
| total           | 1.000    | 0.145   | 6.918   | 0.000   |
| totalIntercept  | -19.070  | 9.683   | -1.969  | 0.049   |
Predictions with those Coefficients!

> #Get new predictions
> newDistance <- 57

> #naive prediction
> newDistance*1 + -19.070

[1] 37.93
You have Compound Parameter Uncertainty

Defined Parameters:

| Parameter            | Estimate | Std.Err | Z-value | P(>|z|) |
|----------------------|----------|---------|---------|---------|
| direct               | 0.640    | 0.156   | 4.117   | 0.000   |
| indirect             | 0.360    | 0.103   | 3.488   | 0.000   |
| total                | 1.000    | 0.145   | 6.918   | 0.000   |
| totalIntercept       | -19.070  | 9.683   | -1.969  | 0.049   |

Diagram: Abiotic factors (abiotic) influence richness, mediated by distance and heterogeneity (hetero).
You have Compound Parameter Uncertainty

```
> predDist <- rnorm(100, 1, 0.145)*newDistance +
    rnorm(100, -19.070, 9.683)
> quantile(predDist)
          0%        25%        50%        75%       100%
10.58009   31.15239   39.55937   46.08567   66.19564
```
Creating Ranges of Predictions

code:

```r
new_data <- data.frame(distance = seq(57,83,.01))
new_data$predicted_richness <- new_data$distance*1 + -19.070

new_data <- cbind(new_data,
                   t(sapply(new_data[,1], function(x)
                   quantile(rnorm(100, 1, 0.145)*x +
                               rnorm(100, -19.070, 9.683),
                               probs=c(0.025, 0.975))))
```