USGS
science for a changing world

Lavaan Option for Adjusting for Spatial Autocorrelation Practice Example

Jim Grace

U.S. Department of the Interior
U.S. Geological Survey

The module contains a practice example on adjusting for spatial autocorrelation when modeling using lavaan. It is an accompaniment to the module entitled, "SEM.Sp1-Lavaan Spatial Autocorrelation Procedures."

Notes: IP-056512; Support provided by USGS Climate & Land Use R&D and Ecosystems Programs. I would like to acknowledge the major contribution by Jarrett Byrnes, Univ. Mass. – Boston for the `lavSpatialCorrect` function used in this module. Appreciation also to Darren Johnson for technical advice. Formal review of the material from which this tutorial was derived was provided by Jesse Miller and Phil Hahn, Univ. Wisconsin. Questions about this material can be sent to sem@usgs.gov.

Last revised 15.06.17.

The Example

VOL. 170, SUPPLEMENT THE AMERICAN NATURALIST AUGUST 2007

Biogeographic Affinity Helps Explain Productivity-Richness Relationships at Regional and Local Scales

Susan Harrison[1,*] and James B. Grace[2,†]

ABSTRACT: The unresolved question of what causes the observed positive relationship between large-scale productivity and species richness has long interested ecologists and evolutionists. Here we examine a potential explanation that we call the biogeographic affinity hypothesis, which proposes that the productivity-richness relationship is a function of species' climatic tolerances that in turn are shaped by the earth's climatic history combined with evolutionary niche conservatism. Using botanical data from regions and sites

USGS

2

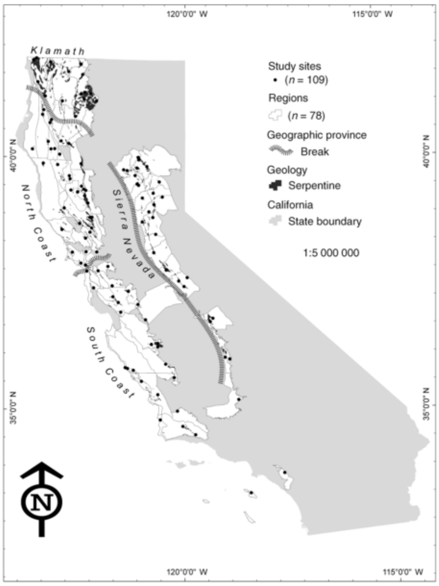
Link for this article:

http://www.jstor.org/stable/10.1086/519010

Cite this example as:

Harrison, S. and Grace, JB. 2007. Biogeographic affinity contributes to our understanding of productivity-richness relationships at regional and local scales. *American Naturalist*. 170:S5-S15.

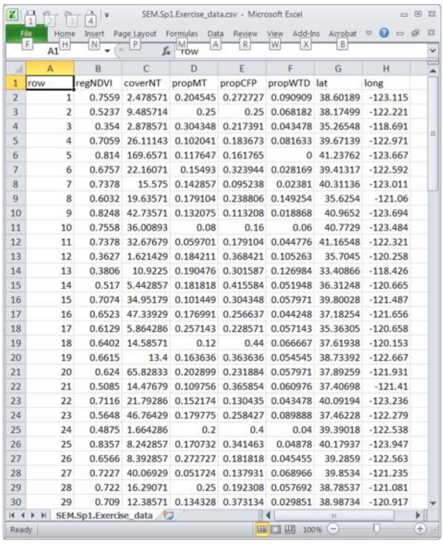
The Sample



The overall objective in this exercise will be to adjust for any spatial autocorrelation in the structural equation model residuals.

3

The Data



SEM.Sp1.Exercise_data.csv - Microsoft Excel

row	regNDVI	coverNT	propMT	propCFP	propWTD	lat	long
1	0.7559	2.478571	0.204545	0.272727	0.090909	38.60189	-123.115
2	0.5237	9.485714	0.25	0.25	0.068182	38.17499	-122.221
3	0.354	2.878571	0.304348	0.217391	0.043478	35.26548	-118.691
4	0.7059	26.11143	0.102041	0.183673	0.081633	39.67139	-122.971
5	0.814	169.6571	0.117647	0.161765	0	41.23762	-123.667
6	0.6757	22.16071	0.15493	0.323944	0.028169	39.41317	-122.592
7	0.7378	15.575	0.142857	0.095238	0.02381	40.31136	-123.011
8	0.6032	19.63571	0.179104	0.238806	0.149254	35.6254	-121.06
9	0.8248	42.73571	0.132075	0.113208	0.018868	40.9652	-123.694
10	0.7558	36.00893	0.08	0.16	0.06	40.7729	-123.484
11	0.7378	32.67679	0.059701	0.179104	0.044776	41.16548	-122.321
12	0.3627	1.621429	0.184211	0.368421	0.105263	35.7045	-120.258
13	0.3806	10.9225	0.190476	0.301587	0.126984	33.40866	-118.426
14	0.517	5.442857	0.181818	0.415584	0.051948	36.31248	-120.665
15	0.7074	34.95179	0.101449	0.304348	0.057971	39.80028	-121.487
16	0.6523	47.33929	0.176991	0.256637	0.044248	37.18254	-121.656
17	0.6129	5.864286	0.257143	0.228571	0.057143	35.36305	-120.658
18	0.6402	14.58571	0.12	0.44	0.066667	37.61938	-120.153
19	0.6615	13.4	0.163636	0.363636	0.054545	38.73392	-122.667
20	0.624	65.82833	0.202899	0.231884	0.057971	37.89259	-121.931
21	0.5085	14.47679	0.109756	0.365854	0.060976	37.40698	-121.41
22	0.7116	21.79286	0.152174	0.130435	0.043478	40.09194	-123.236
23	0.5648	46.76429	0.179775	0.258427	0.089888	37.46228	-122.279
24	0.4875	1.664286	0.2	0.4	0.04	39.39018	-122.538
25	0.8357	8.242857	0.170732	0.341463	0.04878	40.17937	-123.947
26	0.6566	8.392857	0.272727	0.181818	0.045455	39.2859	-122.563
27	0.7227	40.06929	0.051724	0.137931	0.068966	39.8534	-121.235
28	0.722	16.29071	0.25	0.192308	0.057692	38.78537	-121.081
29	0.709	12.38571	0.134328	0.373134	0.029851	38.98734	-120.917

SEM.Sp1.Exercise_data

The data to be used in this exercise can be found in the file "SEM.Sp1.Exercise_data.csv".

To complete this exercise:

(1) Use this data to estimate the model on the next page using lavaan.

(2) Check model fit and respecify if needed.

(3) Use the lavSpatialCorrect function to check for spatial autocorrelation in residuals and to obtain revised stats.

4

The Model

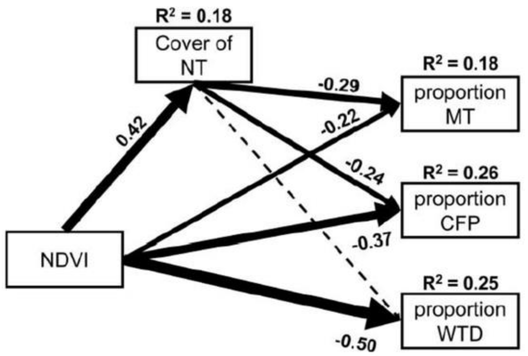


Figure 7: Relationships of the local richness of the three subordinate affinity groups (Madro-Tertiary [MT], California Floristic Province [CFP], warm temperate desert [WTD]) to normalized difference vegetation index (NDVI) and the local cover of north-temperate (NT) species. Standardized path coefficients are shown. The dashed line indicates that proportional representation of WTD species was unrelated to the cover of NT once the effect of NDVI was taken into account.

This model seeks to determine if the local abundance of species with North-temperate affinity (Cover of NT) might be suppressing the richness of other groups of species (ones with different evolutionary origins).

Refer to the paper for more details (but, the punchline is yes, abundance of NT species appears to suppress species from two of the other groups, MT and CFP).

USGS

Give it a try!

Again,

To complete this exercise:

(1) Use the data given to estimate the SE model on the previous page using lavaan.

(2) Check model fit and respecify if needed.

(3) Use the lavSpatialCorrect function to check for spatial autocorrelation in residuals and to obtain revised stats.

USGS [when you have finished with your work, go to the next slides to compare with those anticipated for this exercise]

Lavaan code – part 1: Read data and create data objects.

```
### SEM.Sp1.Exercise-Rcode

### Set directory - example path
setwd("F:/ppt_files/_SEM_educational_materials/Z_SpatialAu
tocorrelation")

### Read and check data
exdat <- read.csv("SEM.Sp1.Exercise_data.csv")
names(exdat)
summary(exdat)
dim(exdat)
attach(exdat)

### Load needed libraries and functions
library(lavaan)
library(ape)
source("lavSpatialCorrect.R") # access the function
```


Here are the preliminary bits of R code needed to get things ready.

Lavaan code – part 2: Specify lavaan model, fit, and correct.

```
### lavaan modeling
# Specify model
ex.mod <- 'coverNT ~ regNDVI
           propMT ~ regNDVI + coverNT
           propCFP ~ regNDVI + coverNT
           propWTD ~ regNDVI + 0*coverNT'

# Fit model
ex.mod.fit <- sem(ex.mod, exdat, meanstructure=T)

# Examine model with uncorrected parameters
summary(ex.mod.fit, rsq=T, standardized=T)

### Correct for spatial autocorrelation
# Execute correction function
lavSpatialCorrect(ex.mod.fit, lat, long)
```

8

≋USGS

And here is the code for specifying and fitting the model. Also shown in the code for feeding in xy coordinates (lat and long) and then correcting for spatial autocorrelation.

Note that I set link “coverNT -> propWTD” to a value of zero. This creates one degree of freedom for model testing. I could have just specified the last line as “propWTD ~ regNDVI” and accomplished the same thing.

Also note: Lavaan automatically estimates error correlations for joint responses. It may be possible to constrain some of these error correlations/covariances between response variables to zero, though I do not work through that here.

Corrected Output: Part I

```
> lavSpatialCorrect(ex.mod.fit, lat, long)
$Morans_I

$Morans_I$coverNT
    observed     expected        sd   p.value n.eff
1 0.01768234 -0.009259259 0.0248945 0.2791499   109

$Morans_I$propMT
    observed     expected         sd      p.value    n.eff
1 0.08604658 -0.009259259 0.02576962 0.0002169805 91.72804

$Morans_I$propCFP
    observed     expected         sd      p.value    n.eff
1 0.07995737 -0.009259259 0.02583072 0.0005525497 92.85982

$Morans_I$propWTD
  observed     expected         sd   p.value n.eff
1 0.029492 -0.009259259 0.02552833 0.1290218   109
```

Modest, but significant amounts of residual autocorrelation for "propMT" and "propCFP", but not for the other variables.

USGS

9

Here is the Moran's I part of the output from the "lavSpatialCorrect" command. Results are given for each endogenous variable.

P-values suggest significant affects of spatial autocorrelation for propMT and propCFP, but not for coverNT or propWTD. Also shown are the effective sample sizes (n.eff) estimated.

Corrected Output (cleaned up a little): Part II

```
$parameters

      Parameter   Estimate n.eff  Std.err   Z-value      P(>|z|)
coverNT~regNDVI   93.57478   109 20.22592  4.626479 3.719346e-06
coverNT~~coverNT 567.15546   109 76.82523  7.382412 1.554478e-13
       coverNT~1 -32.05121   109 13.24998 -2.418963 1.556484e-02

   Parameter        Estimate    n.eff       Std.err    Z-value       P(>|z|)
propMT~regNDVI   -0.1212172170 91.72804 0.0567513469 -2.1359355 3.268467e-02
propMT~coverNT   -0.0006919510 91.72804 0.0002435454 -2.8411583 4.494999e-03
propMT~~propMT    0.0031516740 91.72804 0.0004653773  6.7722979 1.267528e-11
propMT~~propCFP   0.0002531823 91.72804 0.0004605114  0.5497852 5.824667e-01
propMT~~propWTD  -0.0002714040 91.72804 0.0001980094 -1.3706620 1.704804e-01
       propMT~1   0.2434607855 91.72804 0.0349317606  6.9696111 3.178183e-12

   Parameter        Estimate    n.eff       Std.err     Z-value       P(>|z|)
propCFP~regNDVI  -2.743766e-01 92.85982 0.0789390547 -3.47580251 5.093272e-04
propCFP~coverNT  -8.785884e-04 92.85982 0.0003417677 -2.57071790 1.014880e-02
propCFP~~propCFP  6.151883e-03 92.85982 0.0009028366  6.81394963 9.495497e-12
propCFP~~propWTD  1.593965e-05 92.85982 0.0002721264  0.05857441 9.532911e-01
       propCFP~1  4.581109e-01 92.85982 0.0485313122  9.43949218 3.745904e-21
```

Hopefully these results match yours (assuming you ran the same model). If you spot problems, please report to sem@usgs.gov.

USGS

10

And, here are the corrected standard errors and p-values for those relationships affected by spatial autocorrelation.

More information can be found at http://www.nwrc.usgs.gov/SEM

USGS

I hope this overview has been useful. For more information, go to our webpage or search for examples involving your subject of interest. Questions and comments can be sent to sem@usgs.gov. Please note I cannot guarantee responses to individual inquiries, but will definitely incorporate suggestions in future tutorials. – Thanks!