



SEM.3.Ex-1 Model Evaluation: Practice Exercise

Jim Grace

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

1

This module provides some practice with model evaluation. This module builds on the teaching module for Model Evaluation (SEM.3).

An appropriate general citation for this material is

Burnham, K.P. and Anderson, D.R. 2002. Model Selection and Multimodel Inference. Springer Verlag

Notes: IP-056512; Support provided by the USGS Climate & Land Use R&D and Ecosystems Programs. I would like to acknowledge formal review of this material by Jesse Miller and Phil Hahn, University of Wisconsin. Many helpful informal comments have contributed to the final version of this presentation. The use of trade names is for descriptive purposes only and does not imply endorsement by the U.S. Government,

Last revised 22.07.06.

Source: <https://www.usgs.gov/centers/wetland-and-aquatic-research-center/science/quantitative-analysis-using-structural-equation>

Example Model for Evaluation: Cardinale et al. 2009.

Ecology, 90(5), 2009, pp. 1227–1241
© 2009 by the Ecological Society of America

Does productivity drive diversity or vice versa? A test of the multivariate productivity–diversity hypothesis in streams

BRADLEY J. CARDINALE,^{1,3} DANUTA M. BENNETT,¹ CRAIG E. NELSON,¹ AND KEVIN GROSS²

¹*Department of Ecology, Evolution, and Marine Biology, University of California–Santa Barbara, Santa Barbara, California 93106 USA*

²*Biomathematics Graduate Program, North Carolina State University, Raleigh, North Carolina 27695 USA*



2

For this example, we will use a paper involving stream ecosystems.

Published Model

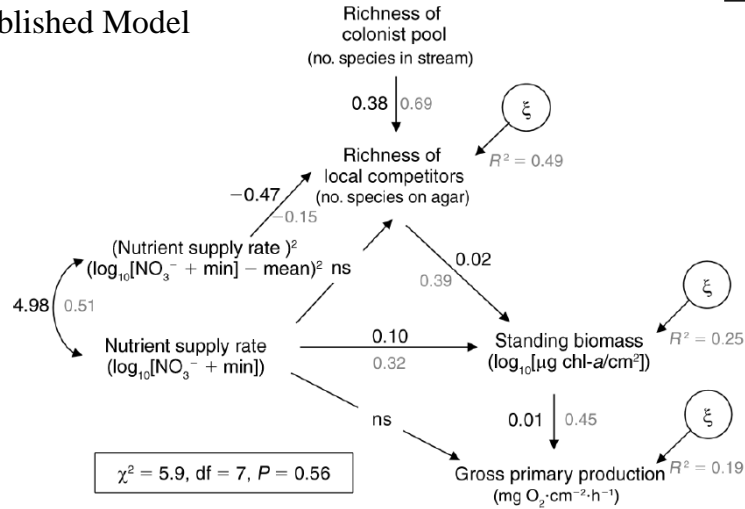
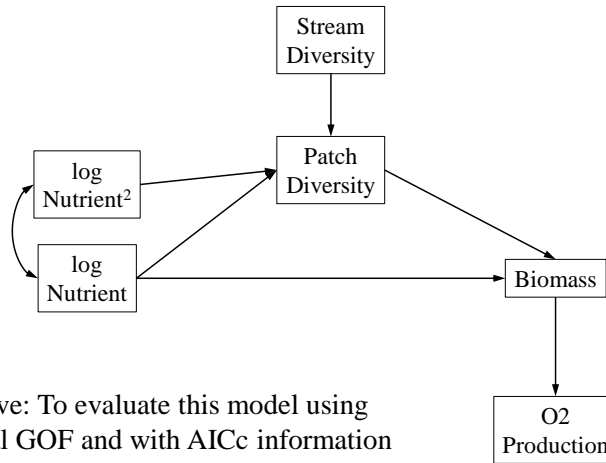


FIG. 5. Results of a structural-equations model used to test whether covariance among variables measured in the stream experiment conform to predictions of the multivariate productivity–diversity hypothesis in Fig. 1. Two values are reported for each path. Unstandardized regression coefficients are shown as the larger-font black numbers presented at the left or top of each arrow. These coefficients give the amount of change in a downstream variable Y per unit change in upstream variable X . The standardized regression coefficients are shown as the smaller-font gray numbers presented at the right or bottom of each arrow. These give the standard deviation change in Y given a standard deviation change in X . Epsilons represent the error term for each “downstream” variable, with the amount of variation explained by the model given by R^2 values at right. Metrics of overall model fit are given in the box at the lower left. These metrics indicate that the model cannot be rejected as a potential explanation of covariance in the data set.

For background, I provide the final model from their paper.

Initial Model for Testing



Objective: To evaluate this model using classical GOF and with AICc information theory approaches.



4

Try starting with this initial model. Before doing any statistical evaluations using the data, provide a logical evaluation considering reasonable alternatives.

Shouldn't O₂ production be influenced by both the quantity of biomass and the patch diversity?

Is it reasonable to expect that productivity will be a function of biomass, patch diversity, and nutrients?

The Data



Cardinale.data.csv - Excel Jim Grace

	A	B	C	D	E	F	G
	PatchDiversity	Biomass	O2Production	logNutrient	logNutrient2	StreamDiversity	Stream
1							
2	26	-2.0256	0.0111	-6	6.06	105	Adobe Creek
3	24	-2	0.0132	-5.7	4.67	105	Adobe Creek
4	20	-1.2127	0.012	-4	0.21	105	Adobe Creek
5	23	-0.9987	0.0184	-2	2.37	105	Adobe Creek
6	20	-0.699	0.0196	0	12.52	105	Adobe Creek
7	21	-1.4236	0.0131	-6	6.06	105	Adobe Creek
8	37	-2.0256	0.01	-5.7	4.67	105	Adobe Creek
9	29	-1.4816	0.0146	-4	0.21	105	Adobe Creek
10	24	-0.2161	0.0154	-2	2.37	105	Adobe Creek
11	18	-0.0389	0.0192	0	12.52	105	Adobe Creek
12	51	-0.7139	0.0343	-6	6.06	139	Buckeye Creek
13	61	-0.3979	0.0487	-5.7	4.67	139	Buckeye Creek
14	60	-0.1118	0.0599	-4	0.21	139	Buckeye Creek
15	63	0.8191	0.0688	-2	2.37	139	Buckeye Creek
16	66	1.0349	0.0549	0	12.52	139	Buckeye Creek
17	37	-2.699	0.0129	-6	6.06	68	Convict Creek
18	14	-2.3267	0.0156	-5.7	4.67	68	Convict Creek
19	15	-1.7246	0.0153	-4	0.21	68	Convict Creek
20	33	-1.2475	0.0227	-2	2.37	68	Convict Creek
21	5	-0.8539	0.0246	0	12.52	68	Convict Creek
22	31	-3	0.0167	-6	6.06	68	Convict Creek
23	19	-1.301	0.0168	-5.7	4.67	68	Convict Creek
24	13	-1.0256	0.02	-4	0.21	68	Convict Creek
25	25	-0.4816	0.024	-2	2.37	68	Convict Creek
26	9	-0.6198	0.0211	0	12.52	68	Convict Creek
27	12	-0.2332	0.0531	-6	6.06	47	Cottonwood Creek
28	11	-0.1757	0.0543	-5.7	4.67	47	Cottonwood Creek
29	24	-0.1253	0.0556	-4	0.21	47	Cottonwood Creek

Cardinale.data

Ready

Data file is named "SEM_3_Ex1_Rcode.txt".

Original data file "Cardinale.data.csv" was extracted from the lavaan.survey package where its analysis is discussed further.

The Exercise

To complete this exercise:

- (1) Use lavaan and the data provided to fit the model shown.
- (2) Check model fit to first see if any important links appear to be missing.
- (3) Add additional links and reevaluate model if and as needed.
- (4) Request summary of model and examine p-values to see if some links appear weak.
- (5) Remove weakest link and create simpler models as you wish.
- (6) Take a set of alternative models and compare using AICc.



6

- (1) You may wish to review the tutorial/module on Model Evaluation.

Preliminary steps in R.

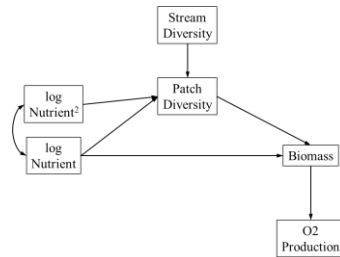
```
### Model Evaluation Exercise
### exercise adapted from Cardinale et al. 2009.

setwd("")
dat <- read.csv("SEM_3_Model Evaluation exercise data.csv")
attach(dat)

# load libraries and functions
library(lavaan)
library(AICcmodavg)
```



Lavaan code for initial model.



```
# Specify initial model
mod1 <-
'PatchDiversity ~ logNutrient +logNutrient2 +StreamDiversity
Biomass ~ PatchDiversity +logNutrient
O2Production ~ Biomass
logNutrient ~~ logNutrient2'
```



8

Here is the code for the initial model.



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

Analysis Code for Initial Model

```
# Fit (estimate) model
mod1.fit <- sem(mod1, data = dat, fixed.x = FALSE,
               estimator = "MLM") #estimator=MLM not required
```

Warning message:

```
In lav_data_full(data = data, group = group, group.label =
group.label, :
lavaan WARNING: some observed variances are (at least) a
factor 1000 times larger than others; use varTable(fit) to
investigate
>
```

```
> varTable(mod1.fit) #used in response to error statement
```

	name	idx	nobs	type	exo	user	mean	var	nlev	lnam
1	PatchDiversity	1	127	numeric	0	0	25.724	189.281	0	
2	Biomass	2	127	numeric	0	0	-0.995	0.492	0	
3	O2Production	3	127	numeric	0	0	0.026	0.000	0	
4	logNutrient	4	127	numeric	0	0	-3.545	5.372	0	
5	logNutrient2	5	127	numeric	0	0	5.329	18.078	0	
6	StreamDiversity	6	127	numeric	1	0	72.992	610.436	0	



10

Here is the code to estimate the initial model, plus a warning and some diagnostics. Note I present code in black, error statements in red, and output in blue. Note that the included option estimator = “MLM” is discussed in the “lavaan options” module.

Recode and Rerun

```
# Recode data
dat$PatchDiversity <- dat$PatchDiversity/100
dat$StreamDiversity <- dat$StreamDiversity/100
dat$O2Production <- dat$O2Production*100

# Rerun mod1 with coded data
mod1.fit <- sem(mod1, data = dat, fixed.x = FALSE,
               estimator = "MLM")

# Extract results
mod1.fit #prints fit stats
```

```
> mod1.fit #prints fit stats
lavaan (0.5-20) converged normally after 52 iterations
  Number of observations              127
  Estimator                          ML      Robust
  Minimum Function Test Statistic    6.130    6.423
  Degrees of freedom                  8        8
  P-value (Chi-square)               0.633    0.600
  Scaling correction factor          0.954
    for the Satorra-Bentler correction
```



11

I recoded some variables and reran model 1. Then extract results by requesting model chi-square statistics.

Model fit is good in general terms

The Search for Missing Links.

```
# Request modification indices greater than 3.0
mil <- modindices(mod1.fit); print(mil[mil$mi > 3.0,])
```

	lhs	op	rhs	mi	mi.scaled
24	Biomass	~~	O2Production	3.410	3.573
31	Biomass	~	O2Production	3.410	3.573
34	O2Production	~	PatchDiversity	3.304	3.461
37	O2Production	~	StreamDiversity	3.752	3.931

>

```
### Model2: add PatchDiversity -> O2Production
### I am adding PatchDiversity to O2Production submodel
### because of logical priority, not largest MI
mod2 <-
'PatchDiversity ~ logNutrient +logNutrient2 +StreamDiversity
Biomass ~ PatchDiversity +logNutrient
O2Production ~ Biomass +PatchDiversity
logNutrient ~~ logNutrient2'
```



12

I have trimmed the verbose output to show only the essential stuff.

Codes are:

“lhs” = left-hand side

“op” = operator

“rhs” = right-hand side

“mi” = modification index (in expected chi-square reduction units)

“mi.scaled” = based on Scaling correction factor for model

Inspection of mod indices led to decision to add PatchDiversity -> O2Production. This was based on my initial suspicion that this link could be expected to be important. Did not hypothesize stream diversity would play a direct role on O2Production.

The Search for Missing Links (cont.).

```
# Request modification indices for Model2
mi2 <- modindices(mod2.fit); print(mi2[mi2$mi > 3.0,])
```

```
> mi2 <- modindices(mod2.fit); print(mi2[mi2$mi > 3.0,])

[1] lhs      op      rhs      mi      mi.scaled
<0 rows> (or 0-length row.names)
```

```
## Compare mod1 and mod2 using chi-square criterion
anova(mod1.fit, mod2.fit)
```

```
Scaled Chi Square Difference Test (method = "satorra.bentler.2001")

      Df      AIC      BIC  Chisq Chisq diff Df diff Pr(>Chisq)
mod2.fit  7 1794.8 1851.7  2.7757
mod1.fit  8 1796.2 1850.2  6.1298      3.2933      1    0.06956 .
```

Classic conclusion would be that mod2 not significantly better than mod1;
thus, addition of adding PatchDiversity -> O2Production not supported.



13

The addition of adding PatchDiversity -> O2Production not supported.

The Search for Missing Links (cont.).

```
### Model3: add StreamDiversity -> O2Production
mod3 <-
'PatchDiversity ~ logNutrient +logNutrient2 +StreamDiversity
Biomass ~ PatchDiversity +logNutrient
O2Production ~ Biomass +StreamDiversity
logNutrient ~~ logNutrient2'
```

```
> mi3 <- modindices(mod3.fit); print(mi3[mi3$mi > 3.0,])
```

```
[1] lhs      op      rhs      mi      mi.scaled
<0 rows> (or 0-length row.names)
```

No indication of additional links to add to this model. How does Model3 compare to Model1?

```
> anova(mod1.fit, mod3.fit)
Scaled Chi Square Difference Test (method = "satorra.bentler.2001")
```

	Df	AIC	BIC	Chisq	Chisq diff	Df diff	Pr(>Chisq)
mod3.fit	7	1794.3	1851.2	2.2296			
mod1.fit	8	1796.2	1850.2	6.1298	4.0524	1	0.04411 *

14

Addition of this link IS supported.

Here are more details from the results.

The Question of Simplification

```
> summary(mod3.fit)
```

Regressions:

	Estimate	Std.Err	Z-value	P(> z)
PatchDiversity ~				
logNutrient	0.004	0.004	0.835	0.403
logNutrient2	-0.005	0.002	-1.950	0.051
StreamDiversity	0.384	0.038	9.980	0.000
Biomass ~				
PatchDiversity	2.003	0.449	4.459	0.000
logNutrient	0.097	0.025	3.823	0.000
O2Production ~				
Biomass	1.287	0.219	5.870	0.000
StreamDiversity	-1.233	0.612	-2.015	0.044

Link from logNutrient -> PatchDiversity appears to not be supported.

15

Parameter p-values are not considered to be definitive, but model Chi-square tests are proposed to be.

The Question of Simplification (cont.)

```
### Model 4: omit logNutrient -> PatchDiversity
mod4 <-
'PatchDiversity ~ logNutrient2 +StreamDiversity
Biomass ~ PatchDiversity +logNutrient
O2Production ~ Biomass +StreamDiversity
logNutrient ~~ logNutrient2'
```

```
> mi4 <- modindices(mod4.fit); print(mi4[mi4$mi > 3.0,])

[1] lhs      op      rhs      mi      mi.scaled
<0 rows> (or 0-length row.names)
```

No indication of additional links to add to this model. How does Model3 compare to Model4?

```
> anova(mod3.fit, mod4.fit)
Scaled Chi Square Difference Test (method = "satorra.bentler.2001")

  Df    AIC    BIC  Chisq Chisq diff Df diff Pr(>Chisq)
mod3.fit  7 1794.3 1851.2  2.2296
mod4.fit  8 1793.0 1847.0  2.9114    0.74711    1    0.3874
```

This simplification does not hurt model fit, so Model4 chosen.

16

Always keep in mind that we seek models that are scientifically solid. This means we are not slaves to the concept of parsimony; rather, we aim for a balance between strength of theory and weight of evidence. As you can tell working through this exercise, retaining a weak effect in our model has little influence on the other parts of the model. As we proceed through the possible simplifications we accumulate a set of models we will use for model selection (slide 20).

The Question of Simplification (cont.)

```
> summary(mod4.fit)
```

Regressions:

	Estimate	Std.Err	Z-value	P(> z)
PatchDiversity ~				
logNutrient2	-0.004	0.002	-1.594	0.111
StreamDiversty	0.383	0.039	9.900	0.000
Biomass ~				
PatchDiversity	2.003	0.454	4.412	0.000
logNutrient	0.097	0.026	3.782	0.000
O2Production ~				
Biomass	1.287	0.222	5.797	0.000
StreamDiversty	-1.233	0.612	-2.014	0.044

Link from logNutrient2 -> PatchDiversity questionable.

17

I always make 1 change at a time and remove least significant element each time.

The Question of Simplification (cont.)

```
### Model 5: omit logNutrient2 -> PatchDiversity
mod5 <-
  'PatchDiversity ~ StreamDiversity
  Biomass ~ PatchDiversity +logNutrient
  O2Production ~ Biomass +StreamDiversity
  logNutrient ~~ logNutrient2'
```

```
> mi5 <- modindices(mod5.fit); print(mi5[mi5$mi > 3.0,])
      lhs op      rhs      mi mi.scaled
22 PatchDiversity ~~ logNutrient2 3.832      3.904
```

Removal of this link is protested by the modification indices.

```
> anova(mod4.fit, mod5.fit)
Scaled Chi Square Difference Test (method = "satorra.bentler.2001")

      Df      AIC      BIC Chisq Chisq diff Df diff Pr(>Chisq)
mod4.fit  8 1793.0 1847.0 2.9114
mod5.fit  9 1794.2 1845.4 6.1675      2.6115      1      0.1061
```

However, the chi-square test says “let it go”, so we accept Model5,
the simpler model.

18

Note first that the satorra bentler method may not be working properly with anova function. Chisq diff does not equal difference between chi-squares! Also, note that chi-square decision is same as BIC would be, but you can see the lower AIC would favor mod4.

The Question of Simplification (cont.)

```
> summary(mod5.fit)
```

Regressions:

	Estimate	Std.Err	Z-value	P(> z)
PatchDiversity ~				
StreamDiversity	0.383	0.039	9.844	0.000
Biomass ~				
PatchDiversity	2.003	0.450	4.454	0.000
logNutrient	0.097	0.025	3.830	0.000
O2Production ~				
Biomass	1.287	0.218	5.891	0.000
StreamDiversity	-1.233	0.612	-2.016	0.044

No more indications of need to prune model.

19

Recall that link from StreamDiversity -> O2Production already brought in as significant improvement.

Comparing the models examined using AICc.

```
### Compare all mods using AICc criterion
aictab(list(mod1.fit, mod2.fit, mod3.fit, mod4.fit,
  mod5.fit), c("Model1", "Model2", "Model3", "Model4",
  "Model5"))
```

Model selection based on AICc:

	K	AICc	Delta_AICc	AICcWt	Cum.Wt	LL
Model4	19	1793.91	0.00	0.39	0.39	-877.49
Model5	18	1794.93	1.01	0.23	0.62	-879.11
Model3	20	1795.51	1.60	0.17	0.79	-877.15
Model2	20	1796.06	2.14	0.13	0.92	-877.42
Model1	19	1797.13	3.22	0.08	1.00	-879.10

This approach selects the more complex model, Model 4. I chose this model for interpretation.

Results from the selected model – part 1

```
> summary(mod4.fit)
```

```
lavaan (0.5-20) converged normally after 54 iterations
```

Number of observations	127	
Estimator	ML	Robust
Minimum Function Test Statistic	2.911	3.070
Degrees of freedom	8	8
P-value (Chi-square)	0.940	0.930
Scaling correction factor		0.948
for the Satorra-Bentler correction		

Results from the selected model – part 2

Regressions:

	Estimate	Std.Err	Z-value	P(> z)	Std.all
PatchDiversity ~					
logNutrient2	-0.004	0.002	-1.594	0.111	-0.115
StreamDiversty	0.383	0.039	9.900	0.000	0.688
Biomass ~					
PatchDiversity	2.003	0.454	4.412	0.000	0.395
logNutrient	0.097	0.026	3.782	0.000	0.323
O2Production ~					
Biomass	1.287	0.222	5.797	0.000	0.484
StreamDiversty	-1.233	0.612	-2.014	0.044	-0.164

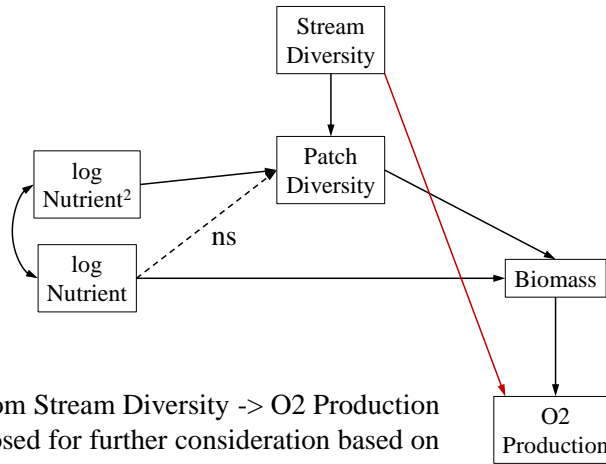
R-Square:

	Estimate
PatchDiversity	0.487
Biomass	0.246
O2Production	0.218

22

There are additional results you will find when you run the summary statement.

Final Model Selected



Link from Stream Diversity -> O₂ Production is proposed for further consideration based on this analysis.



23

One lesson here is that different approaches to model evaluation can lead to the selection of alternative models. We have more confidence in the parts of a model that are consistently chosen using different methods.

Want to take the example further?

For an additional exercise:

- (1) Use your own data to construct a small-sized model. If you choose only 4 or 5 variables, you will still find there are a number of alternative models possible.



24

You can be as creative here as you like. Note that generally, the AICc criterion is liberal and includes more predictors in the submodels. This is consistent with the basis for AICc theory, which presumes true models are complex.