



SEM in R: Local Estimation Using `piecewiseSEM`

Jim Grace

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

Here I provide just a very brief introduction to the concept of local estimation and demonstrate the use of the `piecewiseSEM` package

Citation for the package is:

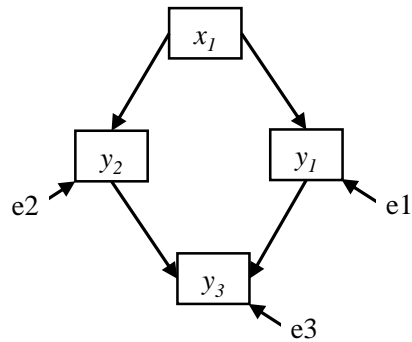
Lefcheck, J.S., 2016. `piecewiseSEM`: Piecewise structural equation modelling in r for ecology, evolution, and systematics. *Methods in Ecology and Evolution*, 7(5), pp.573-579.

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 18.07.11.

Source: <https://bit.ly/graceSEM>

How would we evaluate this model using local estimation methods?

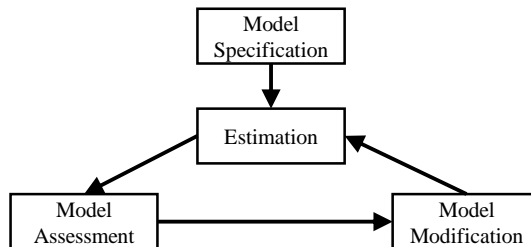


2

Here we consider the same model we did in our brief “Intro to Lavaan”. This model represents the hypothesis that the effect of x_1 on y_3 occurs because of two processes, one propagated through y_1 and the other through y_2 .

Again, we focus on the mechanics of

- specification
- estimation
- model assessment.

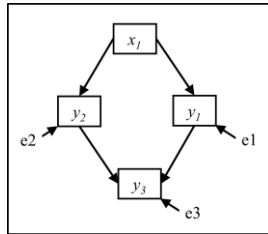


3

As before, we will want to

1. specify our model,
2. estimate the parameter values
3. assess how well our data correspond to our model.

Identify conditional independences as a way of thinking about specification alternatives.



What are the conditional independence claims?

- (1) $y_2 \perp y_1 \mid x_1$
- (2) $y_3 \perp x_1 \mid y_1, y_2$

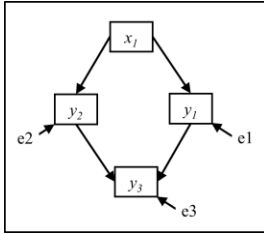


4

It is handy to consider the possible alternative models up front when one is doing local estimation, since we will have to check our models by hand rather than having software do that for us. As discussed in the module on Model Evaluation, the first order of business is to determine if there are any important omitted links. The reason this is the first order of business is because when links that are important in the data generating process are omitted from the model, the estimates for other links can be way off. In contrast, including unimportant links in models has a comparatively smaller effect on the estimates for other links.

Regarding how we evaluate missing links, I introduce the concept of “conditional independence” in the module “SEM Essentials – Basics of Estimation”. It is also covered in greater depth in “SEM Essentials – Path Rules”. In this example, there are two implied independences in our model. There is no link directly from x_1 to y_3 and none connecting y_1 and y_2 . We need to know if those pairs are indeed conditionally independent.

Specification and estimation of equations/submodels in R:



```
# Specify model as series of regressions
y1.mod <- lm(y1 ~ x1, data=t.dat)
y2.mod <- lm(y2 ~ x1, data=t.dat)
y3.mod <- lm(y3 ~ y1 + y2, data=t.dat)
```



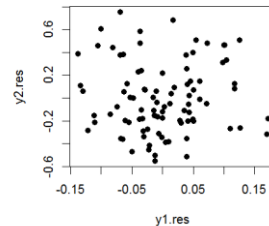
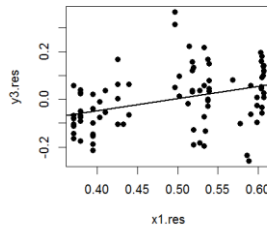
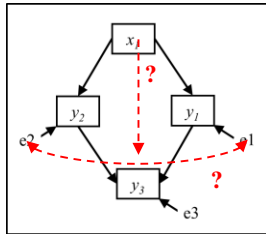
5

Here I show a simple “by-hand” approach. Recall that a network model can be represented by an equation for each endogenous variable. We can use the base function “lm” to model each of our endogenous variables as a function of its parent predictors, creating three model objects that collectively summarize the network hypothesis. This is an example of local or piecewise approach to estimation.

Model assessment by hand.

```
# Capture residuals
x1.res <- x1
y1.res <- resid(y1.mod)
y2.res <- resid(y2.mod)
y3.res <- resid(y3.mod)

# plot residual relationships
```



6

In linear Gaussian models, indications of missing links can be detected through the observation of residual relationships. We can look at this by obtaining residuals for our endogenous variables with the function “resid”. Note that for x_1 , the residuals are the raw values, as this exogenous variable has no predicted scores. Note that the ultimate test will be to include a directed path in one of the endogenous nodes associated with the residual association. The choice among the possibilities (y_2 affected by y_1 or y_1 affected by y_2) is a decision to be based on theory and appropriate causal assumptions.

Examining residual relationships simply involves looking to see if the residuals for two unconnected variables are significantly related.

When we test for significance in this way, we find evidence that perhaps x_1 should be in the equation for y_3 . This evidence should not be considered the final “test” for inclusion, but only a diagnostic. Also, this scatterplot approach permits us to find nonlinear residual relationships, which would require special modeling techniques to include in equations.

All this (and more) is now automated in the “piecewiseSEM” package developed by Jon Lefcheck.

```
### Install Lefcheck's piecewiseSEM package
library(piecewiseSEM)

# Model 1 - SE model is a list of the local models
pw.mod1 = psem(
  lm(y1 ~ x1, t.dat),
  lm(y2 ~ x1, t.dat),
  lm(y3 ~ y1 + y2, t.dat))

# request full summary
summary(pw.mod1, t.dat)
```



7

A package developed by Jon Lefcheck (and now authored by Jon Lefcheck, Jarrett Byrnes, and Jim Grace) is called “piecewiseSEM”. This package handles the bundling of individual local models and d-sep testing for us.

The SEM is a list of R models. Here I only show the simplest sort, “lm” models. The real utility of this package, however, is that it can bundle many different sorts of R models, such as glms, lmers, and more.

For more information on syntax in the package can be found at:

<https://cran.r-project.org/web/packages/piecewiseSEM/piecewiseSEM.pdf>

Once you load piecewiseSEM’ you can request to see

```
> vignette('piecewiseSEM')
```

Summary Output (partial)

```
> summary(pw.mod1, t.dat)
Structural Equation Model of pw.mod1

Call:
lm(y1 ~ x1, data = t.dat)
lm(y2 ~ x1, data = t.dat)
lm(y3 ~ y1 + y2, data = t.dat)

      AIC      BIC
40.405   65.403

---
Tests of directed separation:

      Independ.Claim Estimate Std.Error DF Crit.Value P.Value
y3 ~ x1 + ...    0.6817    0.1576 86     4.3253 0.0000 ***
y2 ~ y1 + ...    0.0563    0.4837 87     0.1164 0.9076

Global goodness-of-fit:

Fisher's C = 20.405 with P-value = 0 and on 4 degrees of freedom
```



8

Here we see the summary results for the same Model 1 as in the lavaan tutorial.

Results indicate we need to add a link from x1 to y3 (shown in next slide).

Revised model and results

```
# Revised Model based on dSep test
pw.mod2 = psem(
  lm(y1 ~ x1, t.dat),
  lm(y2 ~ x1, t.dat),
  lm(y3 ~ y1 + y2 + x1, t.dat))

# request the dSep tests and overall fit (fisherC)
dSep(pw.mod2)
fisherC(pw.mod2)
```

```
Independ.Claim  Estimate Std.Error DF Crit.Value  P.Value
1 y2 ~ y1 + ... 0.05628295 0.4836885 87  0.116362 0.9076339
>
Fisher.C df P.Value
1 0.194 2 0.908
```

Fit is very close (near-perfect).



9

We could request a full summary of the results with
`summary(pw.mod2)`

Here I request the fit statistics for the local fits using “dSep”
and global fit using “fisherC”.

Revised model coefficients

```
> coefs(pw.mod2, standardize="none")
```

	Response	Predictor	Estimate	Std.Error	DF	Crit.Value	P.Value	
1	y1	x1	0.3998	0.0823	88	4.8562	0.0000	***
2	y2	x1	0.8747	0.3715	88	2.3546	0.0208	*
3	y3	y1	0.5925	0.1771	86	3.3455	0.0012	**
4	y3	y2	0.0929	0.0393	86	2.3676	0.0201	*
5	y3	x1	0.6817	0.1576	86	4.3253	0.0000	***



10

P-values indicate support for the individual links.

Compare models

```
# compare models  
AIC(pw.mod1, pw.mod2, aicc=T)
```

```
> AIC(pw.mod1, pw.mod2, aicc=T)  
  df    AIC  
x 10 46.031  
y 11 25.608
```



11

We can compare models using AICc also.

Piecewise (local) estimation opens up many possibilities for modeling more complex specifications. We are working on additional features (and illustrations).

12

For additional illustrations of local estimation in SEMs, see
Grace, J.B., Schoolmaster, D.R. Jr., Guntenspergen, G.R., Little, A.M.,
Mitchell, B.R., Miller, K.M., and Schweiger, E.W. 2012. Guidelines for
a graph-theoretic implementation of structural equation modeling.
Ecosphere 3(8): article 73

Available at <http://www.esajournals.org/doi/pdf/10.1890/ES12-00048.1>

Also,

Shipley, B. 2009. Confirmatory path analysis in a generalized
multilevel context. *Ecology* 90: 363-368.

Shipley, B. (2013) The AIC model selection method applied to path
analytic models compared using a d-separation tests. *Ecology*, 94, 560–
564.

And

Lefcheck, J. 2015. PIECEWISESEM: Piecewise structural equation
modelling in R for ecology, evolution, and systematics. *Methods in
Ecology and Evolution* 7:573-5789.