SEM in R:
Local Estimation Using
piecewiseSEM

Jim Grace

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:


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.

How would evaluate this model using local estimation methods?

Here we consider the same model we did in our brief “Intro to Lavaan”. This model represents the hypothesis that the effect of x1 on y3 occurs because of two processes, one propagated through y1 and the other through y2.
Again, we focus on the mechanics of
- specification
- estimation
- model assessment.

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. \( y_2 \perp y_j \mid x_j \)
2. \( y_3 \perp x_j \mid y_1, y_2 \)

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.
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.

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

# plot residual relationships
```

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 x1, 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 (y2 affected by y1 or y1 affected by y2) 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 x1 should be in the equation for y3. 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.

```r
### 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)
```

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')
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

```r
# 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)
```

<table>
<thead>
<tr>
<th>Independent Claim</th>
<th>Estimate</th>
<th>Std.Error</th>
<th>DF</th>
<th>Crit.Value</th>
<th>P.Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>y2 ~ y1 + ...</td>
<td>0.05628295</td>
<td>0.4836885</td>
<td>87</td>
<td>0.116362</td>
<td>0.9076339</td>
</tr>
<tr>
<td>Fisher.C df P.Value</td>
<td>0.194 2 0.908</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fit is very close (near-perfect).

We could request a full summary of the results with

```r
summary(pw.mod2)
```

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

<table>
<thead>
<tr>
<th>Response</th>
<th>Predictor</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</td>
<td>y1</td>
<td>x1</td>
<td>0.3998</td>
<td>0.0823</td>
<td>88</td>
<td>4.8562</td>
</tr>
<tr>
<td>2</td>
<td>y2</td>
<td>x1</td>
<td>0.8747</td>
<td>0.3715</td>
<td>88</td>
<td>2.3546</td>
</tr>
<tr>
<td>3</td>
<td>y3</td>
<td>y1</td>
<td>0.5925</td>
<td>0.1771</td>
<td>86</td>
<td>3.3455</td>
</tr>
<tr>
<td>4</td>
<td>y3</td>
<td>y2</td>
<td>0.0929</td>
<td>0.0393</td>
<td>86</td>
<td>2.3676</td>
</tr>
<tr>
<td>5</td>
<td>y3</td>
<td>x1</td>
<td>0.6817</td>
<td>0.1576</td>
<td>86</td>
<td>4.3253</td>
</tr>
</tbody>
</table>

P-values indicate support for the individual links.
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).

For additional illustrations of local estimation in SEMs, see
Available at http://www.esajournals.org/doi/pdf/10.1890/ES12-00048.1

Also,
And