Composite Variables in SEM

Jarrett E.K. Byrnes

To SEM and Beyond!

1. What is a composite variable?

2. Using Composites for nonlinear variables
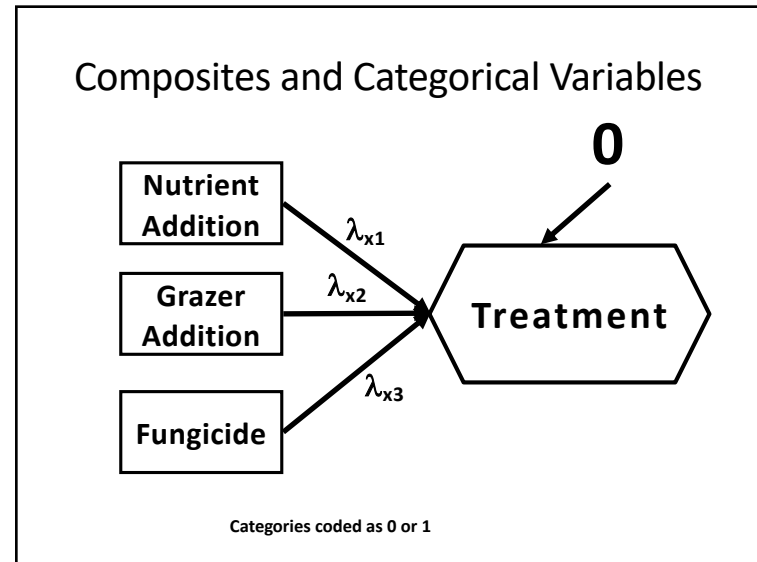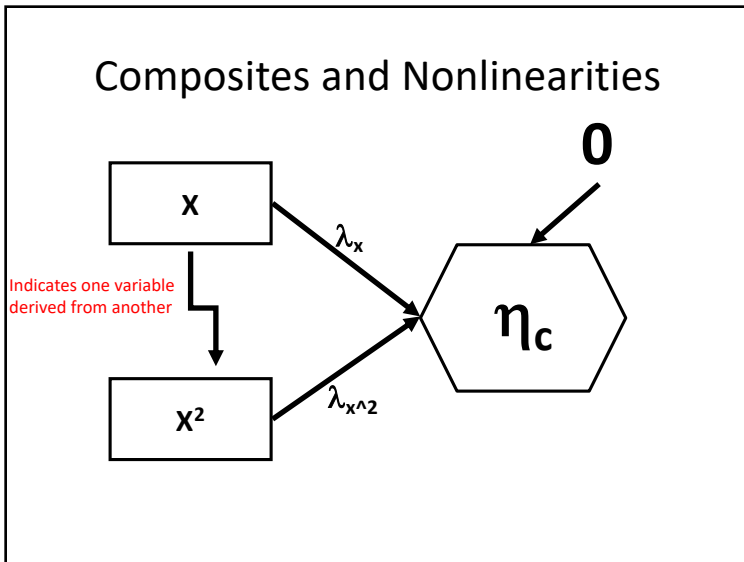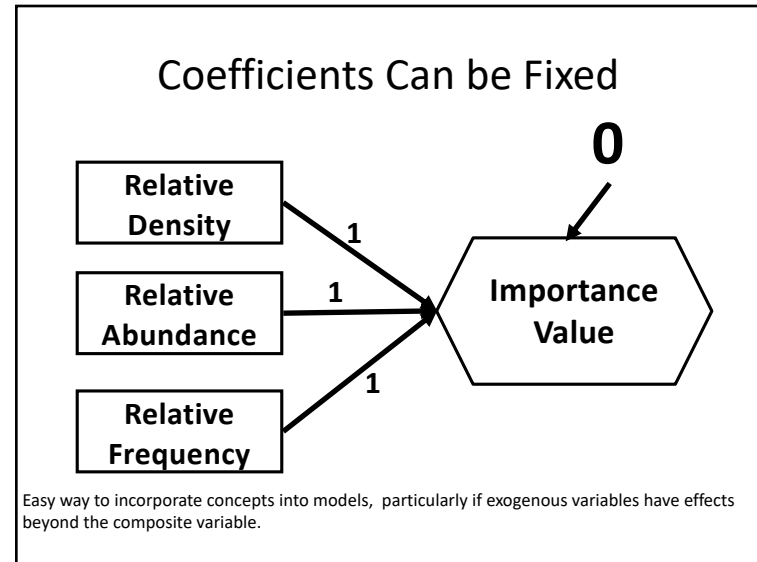
3. Composites v. Latents - when and why?

4. Composites in a `piecewiseSEM` context?

We're Comfortable with Latent Exogenous Variables…

Latent Composite

Composite Variables Reflect Joint Effects

0

X1 $\xrightarrow{\lambda_{x1}}$

X2 $\xrightarrow{\lambda_{x2}}$ η_c

X3 $\xrightarrow{\lambda_{x3}}$

- **Coefficients can be statistically estimated.**
- **Fixing error to 0 aids in identification (otherwise it's a latent composite)**
- **Specifying a scale is often helpful.**

Coefficients Can be Fixed

0

Relative Density $\xrightarrow{1}$

Relative Abundance $\xrightarrow{1}$ **Importance Value**

Relative Frequency $\xrightarrow{1}$

Easy way to incorporate concepts into models, particularly if exogenous variables have effects beyond the composite variable.

Composites and Nonlinearities

0

X $\xrightarrow{\lambda_x}$

Indicates one variable derived from another

X² $\xrightarrow{\lambda_{x^2}}$ η_c

Composites and Categorical Variables

0

Nutrient Addition $\xrightarrow{\lambda_{x1}}$

Grazer Addition $\xrightarrow{\lambda_{x2}}$ **Treatment**

Fungicide $\xrightarrow{\lambda_{x3}}$

Categories coded as 0 or 1

2
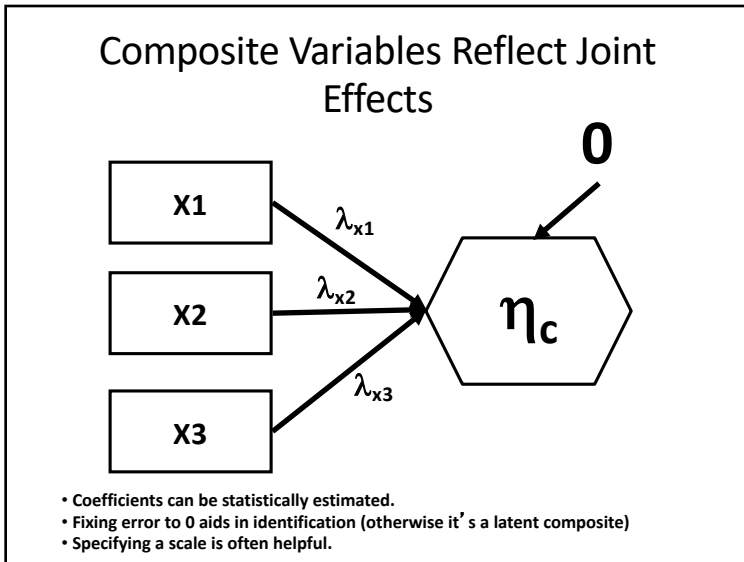
To SEM and Beyond!

1. What is a composite variable?

2. Using Composites for nonlinear variables

3. Composites v. Latents - when and why?

4. Composites in a `piecewiseSEM` context?

Mediation in Analysis of Post-Fire Recovery of Plant Communities in California Shrublands*


*Five year study of wildfires in Southern California in 1993. 90 plots (20 x 50m), (data from Jon Keeley et al.)

Analysis focus: understand post-fire recovery of plant species richness

measured vegetation recovery:
-plant cover
-species richness

Examination of woody remains allowed for estimate of age of stand that burned as well as severity of the fires.
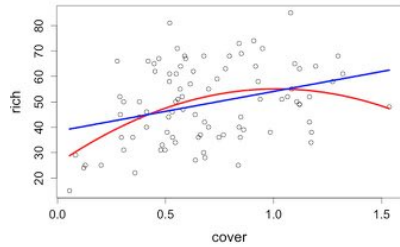
Linear or Nonlinear?
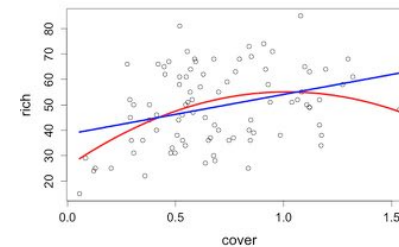

```
linear<-lm(rich ~ cover, data=keeley)

nonlinear<-lm(rich ~ cover+I(cover^2), data=keeley)

aictab(list(linear, nonlinear), c("linear",
   "squared"))
```
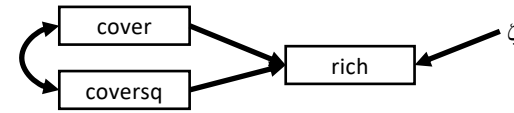
Linear or Nonlinear?

Model selection based on AICc :

```
        K   AICc Delta_AICc AICcWt Cum.Wt      LL
squared 4 735.92      0.00   0.83   0.83 -363.72
linear  3 739.08      3.15   0.17   1.00 -366.40
```
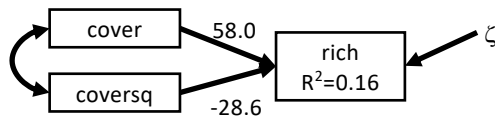
A Simple Nonlinear Model


```
#Create a new nonlinear variable in the data
keeley<-within(keeley, coverSQ<-cover^2)

#Now, for a model
noCompModel <- 'rich ~ cover + coverSQ'

noCompFit <- sem(noCompModel, data=keeley)
```
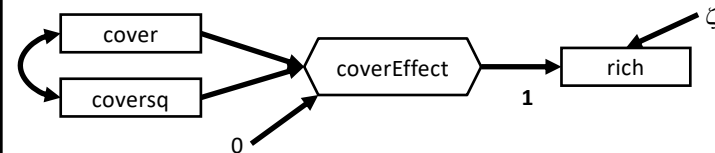
A Simple Nonlinear Model
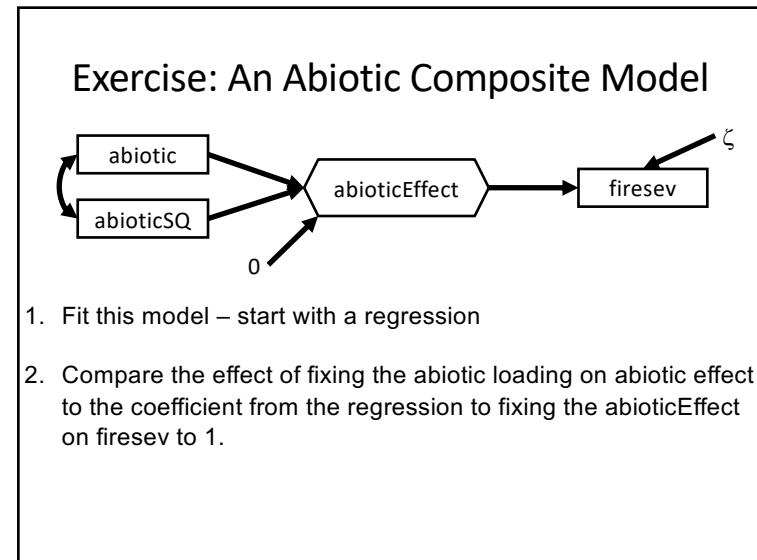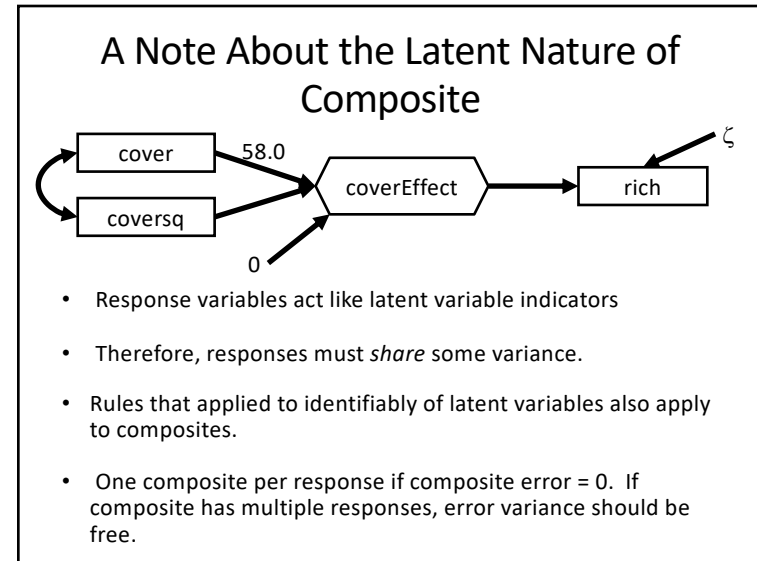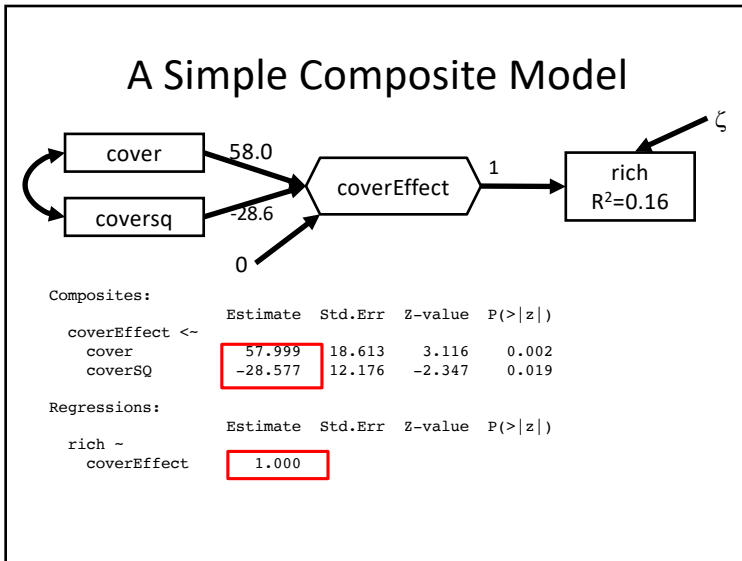


```
> summary(noCompFit)
…
             Estimate  Std.err  Z-value  P(>|z|)
Regressions:
  rich ~
    cover          57.999   18.613    3.116    0.002
    coverSQ       -28.577   12.176   -2.347    0.019
```
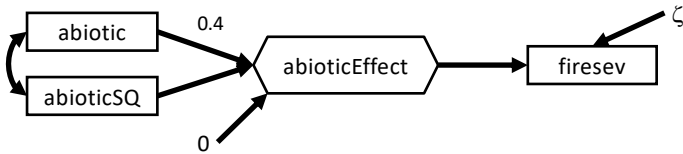
A Simple Composite Model


```
compModel<-'

        coverEffect <~ cover + coverSQ

        rich ~ 1*coverEffect'

compFit <- sem(compModel, data=keeley)
```

4

A Simple Composite Model


```
Composites:
                  Estimate  Std.Err  Z-value  P(>|z|)
  coverEffect <~
    cover           57.999   18.613    3.116    0.002
    coverSQ        -28.577   12.176   -2.347    0.019

Regressions:
                  Estimate  Std.Err  Z-value  P(>|z|)
  rich ~
    coverEffect     1.000
```

A Note About the Latent Nature of Composite

- Response variables act like latent variable indicators

- Therefore, responses must *share* some variance.

- Rules that applied to identifiably of latent variables also apply to composites.

- One composite per response if composite error = 0. If composite has multiple responses, error variance should be free.

Exercise: A Nonlinear Relationship Between abiotic and firesev

Exercise: An Abiotic Composite Model

1. Fit this model – start with a regression

2. Compare the effect of fixing the abiotic loading on abiotic effect to the coefficient from the regression to fixing the abioticEffect on firesev to 1.

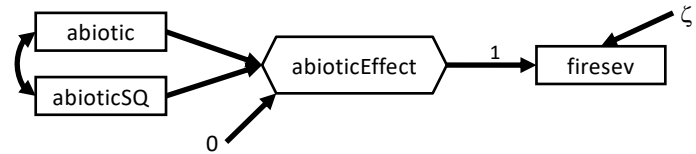## For some reason, this model fails


```
keeley$abioticSQ <- keeley$abiotic^2

abioticCompModelBad<-'
            abioticEffect <~ 0.4 * abiotic +
                                  abioticSQ

            firesev ~ abioticEffect'

abioticCompFitBad <- sem(abioticCompModelBad, data=keeley)
```
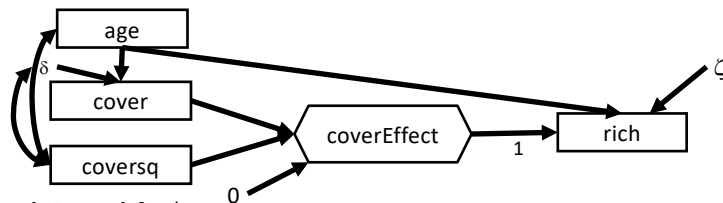
This model does not: try multiple methods with composites!


```
keeley$abioticSQ <- keeley$abiotic^2

abioticCompModel<-'
            abioticEffect <~ abiotic + abioticSQ

            firesev ~ 1*abioticEffect'

abioticCompFit <- sem(abioticCompModel, data=keeley)
```

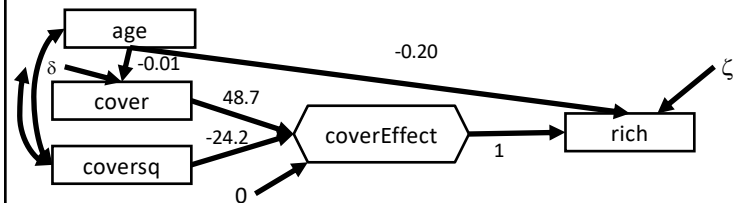## Endogenous Composite Variables


```
endoCompModel<-'
            coverEffect <~ cover + coverSQ

            cover ~~ coverSQ
            age ~~ coverSQ

            cover ~ age
            rich ~ age + 1*coverEffect'

endoCompFit <- sem(endoCompModel, data=keeley, fixed.x=F)
```

Endogenous Composite Variables

Composites:

	Estimate	Std.Err	Z-value	P(>|z|)
coverEffect <~				
cover	48.705	19.246	2.531	0.011
coverSQ	-24.186	12.315	-1.964	0.050

Regressions:

|	Estimate	Std.Err	Z-value	P(>|z|)
cover ~				
age	-0.009	0.002	-3.549	0.000
rich ~				
age	-0.201	0.125	-1.603	0.109
coverEffect	1.000			

To SEM and Beyond!

1. What is a composite variable?

2. Using Composites for nonlinear variables

3. Composites v. Latents - when and why?

4. Composites in a `piecewiseSEM` context?

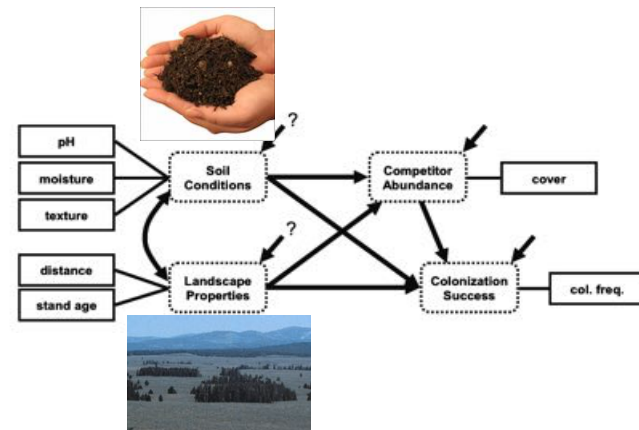Example: Tree Recolonization and Composite Variables

Grace, J.B. & Bollen, K.A. (2008). Representing general theoretical concepts in structural equation models: the role of composite variables. *Environ. Ecol. Stat., 15, 191–213.*

What is the Contribution of Local versus Regional Factors to Recolonization
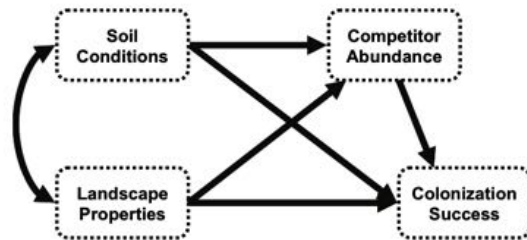
Grace and Bollen 2008
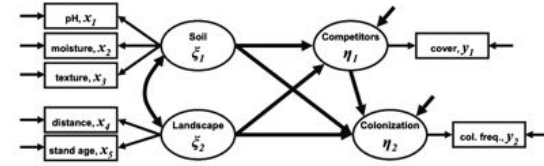

Adding Variables to the Metamodel
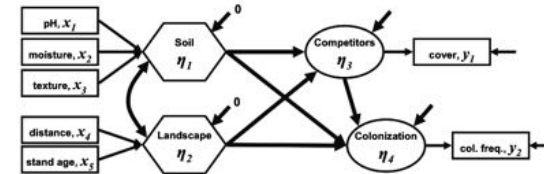

Grace and Bollen 2008

Questions to Ask of Your Latent/Composite Variables

1. What is the direction of causality?

2. Are the indicators in a block interchangeable?

3. Do indicators covary because of joint causes?

4. Do indicators have a consistent set of causal influences?
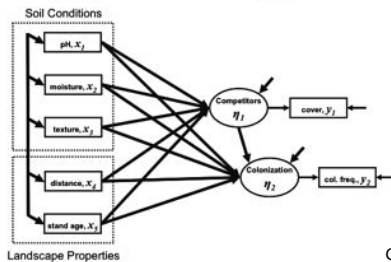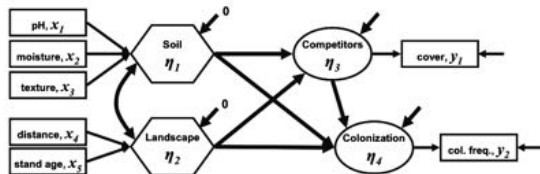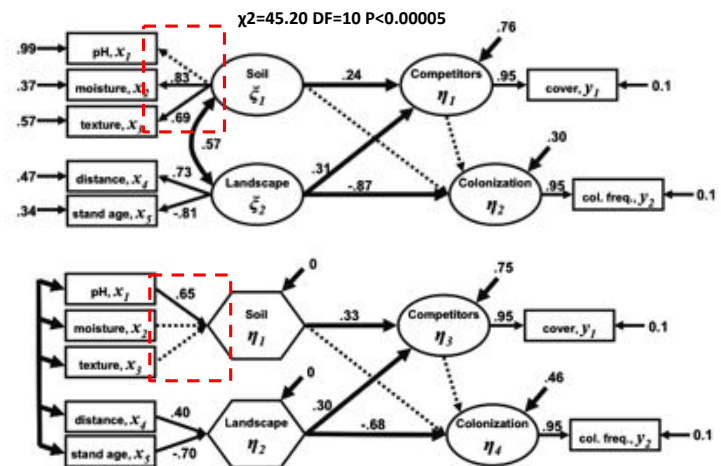
Latent or Composites?

What do you think?
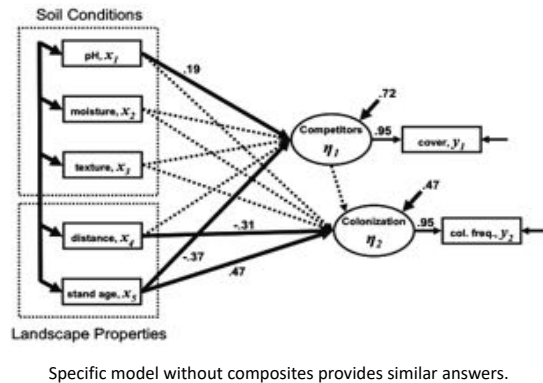
Grace and Bollen 2008

Generality v. Specificity

Grace and Bollen 2008

$\chi 2=45.20$ DF=10 P<0.00005

$\chi 2=6.88$ DF=3 P=0.075

Grace and Bollen 2008

How Confident are We in Composite Loadings and their Conclusions?

Specific model without composites provides similar answers.

Testing our Confidence in Composites
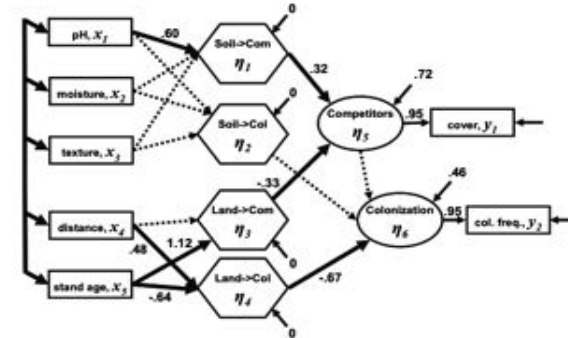
The general composite construct is not obscuring more specific relationships in the data.

Final Notes about Composites

1. The key is causality!

2. Ask what do you gain from a composite versus an observed variable model

To SEM and Beyond!

1. What is a composite variable?

2. Using Composites for nonlinear variables

3. Composites v. Latents - when and why?

4. Composites in a `piecewiseSEM` context?

And Endogenous Nonlinearity Model


```
keeley<-within(keeley, coverSQ<-cover^2)
```
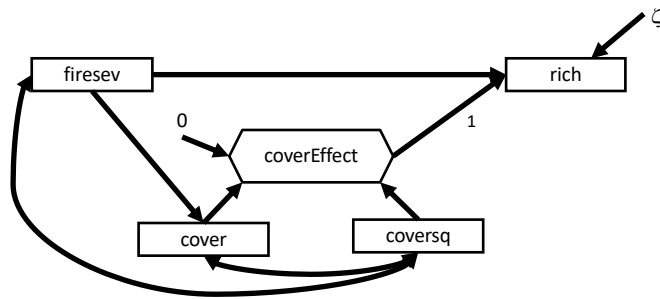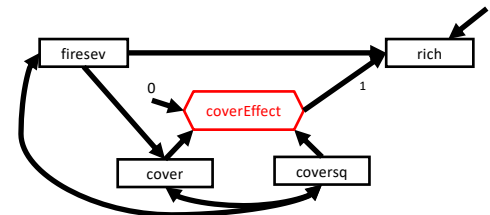
Step 1: Make a Composite via an Observed Variable Model


```
#First, fit the observed only model
rich_obs_mod <- lm(rich ~ cover + coverSQ + firesev,
                data=keeley)

#Now extract a composite
keeley$cover_eff <- with(keeley,
                coef(rich_obs_mod )[2]*cover +
                coef(obs_mod)[3]*cover^2)
```

Step 2: Get the Loadings for the Composite as Part of the Model
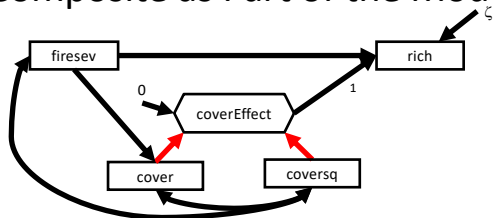

```
#Second, make a loadings relationship
comp_loadings_mod <- lm(cover_eff ~ cover + coverSQ,
                data=keeley)
```

Step 3: Refit with the Composite
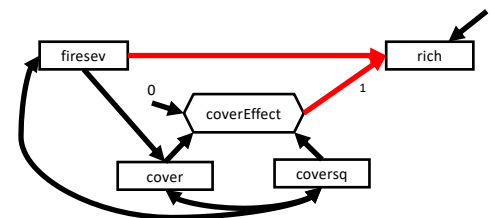

```
#Third, refit the model with the composite
rich_comp_mod <- lm(rich ~ cover_eff + firesev,
                data=keeley)
```

You can use `offset(1*cover_eff)`, *but this causes problems for piecewiseSEM and you don't yet get standardized coefficients for an offset. Fix this, Jon?*

Fit the rest of the model


```
#Now, put it all together
cover_mod <- lm(cover ~ firesev, data=keeley)
```

Make an SEM


```
#Roll it into a pSEM
fire_comp_psem <- psem(
  comp_loadings_mod,
  cover_mod,
  rich_comp_mod,
  coverSQ %~~% firesev,
  coverSQ %~~% cover,
  data = keeley
)
```

The basis set…


```
> basisSet(fire_comp_psem)
$`1`
[1] "coverSQ"    "rich"       "firesev"    "cover_eff"

$`2`
[1] "firesev"    "cover_eff" "coverSQ"    "cover"

$`3`
[1] "cover"      "rich"       "firesev"    "cover_eff"
```

Model Fits…


```
        Independ.Claim       Estimate     Std.Error DF    Crit.Value    P.Value
1      rich ~  coverSQ + ... -5.371133e-15 3.770510e+00 86 -1.424511e-15 1.0000000
2 cover_eff ~  firesev + ... -9.595893e-17 1.104898e-15 86 -8.684868e-02 0.9309937
3   rich ~  cover + ...      -1.327861e-14 7.324056e+00 86 -1.813014e-15 1.0000000


Fisher.C df P.Value
1   0.143  6       1
```

11

Coefficients


```
                Independ.Claim      Estimate      Std.Error DF     Crit.Value   P.Value
1      rich  ~  coverSQ + ...   -5.371133e-15  3.770510e+00 86  -1.424511e-15 1.0000000
2 cover_eff  ~  firesev + ...   -9.595893e-17  1.104898e-15 86  -8.684868e-02 0.9309937
3    rich  ~  cover + ...       -1.327861e-14  7.324056e+00 86  -1.813014e-15 1.0000000


Fisher.C df P.Value
1    0.143  6     1
```

Compare to Traditional Approaches

piecewiseSEM

	Response	Predictor	Estimate
1	cover_eff	cover	45.2741
2	cover_eff	coverSQ	-23.3076
3	cover	firesev	-0.0839
4	rich	cover_eff	1.0000
5	rich	firesev	-2.1610
6	~~coverSQ	~~firesev	-0.3798
7	~~coverSQ	~~cover	-0.1205

lavaan

```
Composites:
                           Estimate
   coverEffect <~
     cover                   45.274
     coverSQ                -23.308

Regressions:
                           Estimate
   cover ~
     firesev                 -0.084
   rich ~
     coverEffect              1.000
     firesev                 -2.161
```

Final Notes on Composites

- You actually do not need to ever fit a composite...

- Can fit an observed variable model and have a 'composite' that is the sum of some of your variables

- Advantage to fitting is to get a summed effect of a suite of influences flowing through one conceptual material
  - This can be done outside of the framework of model fitting and evaluation

Grace et al 2016