Advanced Topics in SEM for Ecology & Evolutionary Biology

Jarrett E. K. Byrnes

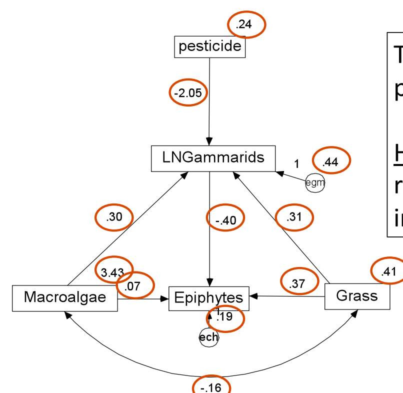
An Advanced Outline

1. Revisiting Sample Size
2. Revisiting Dsep in lavaan
3. Multilevel Generalized Piecewise SEM
4. Additional Spatial Techniques
5. Panel Models for Lagged Time Effects
6. Growth Curve Models & Time Series

Revisiting Sample Size

1. The further you are in a model from an exogenous data-generating, the weaker it's influence.
2. Our ability to detect the these tapering effect sizes is proportional to our <u>information</u> (especially sample size) and the <u>number of parameters being estimated</u>.
3. Our sample size sets an upper limit for the complexity of the model we can obtain.
4. <u>Rules of thumb for sample size</u> -- we hope to have at least 5 samples per estimated parameter and would prefer 20 samples per parameter.
5. Path coefficients add to our parameter list, not the variances

Number of Estimated Parameters

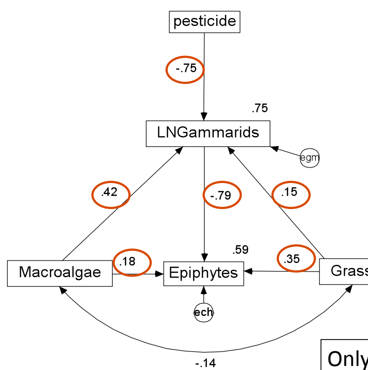


Chi-sqr = 5.147; df = 3; p = .161
AIC = 29.147; NPAR = 12

There are a total of <u>12</u> parameters shown.

<u>However</u>, only <u>6</u> of these require unique information…

Parameters Needing Unique Information



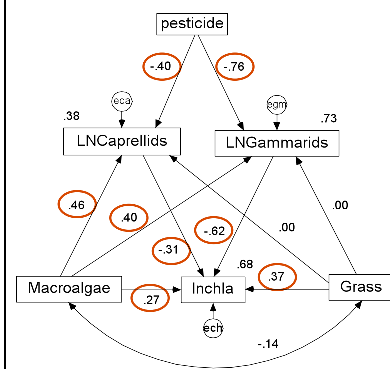
Chi-sqr = 5.147; df = 3; p = .161
AIC = 29.147; NPAR = 12

Variances & covariance of exogenous variables can be obtained from the data. For "pesticide", "Macroalgae", and "Grass", this yields 4 parameters.

Error variances (and R-sqrs) for endogenous variables are calculated from other parameters. This is 2 parameters.

Only 6 parameters require unique information.
Samples/parameters = 40/6 = 6.7.

Removing Unimportant Paths



chi-sqr = 9.545; df = 6; p = .145;
AIC = 39.545; NPAR = 15

For our more complex model, we would want to set non-contributing paths to zero to minimize estimated parameters.

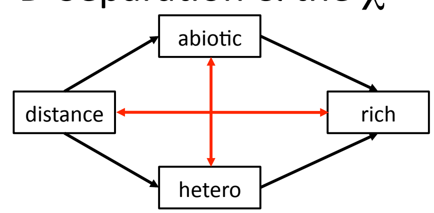
Here estimated parameters = 8, samples/parameters = 5.

If we can combine Caprellids and Gamarids, we could reduce parameters further.

An Advanced Outline

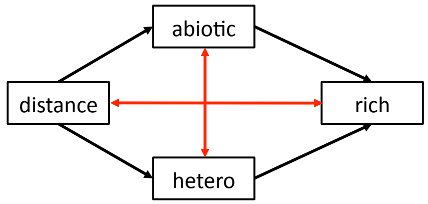
1. Revisiting Sample Size
2. Revisiting Dsep in lavaan
3. Multilevel Generalized Piecewise SEM
4. Additional Spatial Techniques
5. Panel Models for Lagged Time Effects
6. Growth Curve Models & Time Series

D-Separation & the χ^2



1. χ^2 gives you information regarding the discrepancy between your observed and predicted covariance matrix.
2. The test of D-Separation gives you information regarding whether you have missed key associations between variables.
3. We can test for D-Separation in recursive models without correlated error simply

D-Separation in lavaan

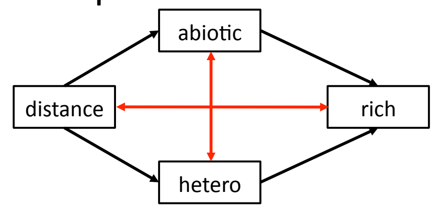


Two options

1) Feed model to DAG in ggm

```
#Full Mediation
distModel2 <- 'rich ~ abiotic  + hetero
              hetero ~ distance
              abiotic ~ distance'
```

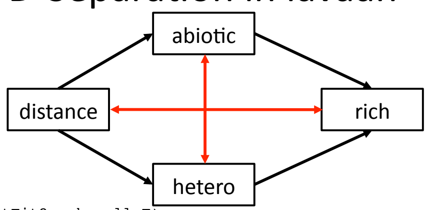
D-Separation in lavaan



2) Use script (and this will be in future lavaan versions)

```
> source("./dsepTest.R")
> dsepTest(distFit2)
$ctest
[1] 21.86173

$df
[1] 4

$pvalue
[1] 0.0002135289
```

D-Separation in lavaan




```
> dsepTest(distFit2, showall=T)
$ctest
[1] 21.86173

$df
[1] 4

$pvalue
[1] 0.0002135289

$dsep
                   Pair   Conditioning         P.t.
distance  distance,rich hetero,abiotic 9.564005e-05
abiotic   abiotic,hetero       distance 1.871306e-01
```

An Advanced Outline

1. Revisiting Sample Size
2. Revisiting Dsep in lavaan
3. Multilevel Generalized Piecewise SEM
4. Additional Spatial Techniques
5. Panel Models for Lagged Time Effects
6. Growth Curve Models & Time Series

D-Separation in Piecewise models beyond linear regression

1. We have models that deal with
   1. Hierarchical/nested data (mixed models)
   2. Nonlinear relationships
   3. Non-normal error distributions (glms)
2. The test of the effect of a variable in one of those models serves the same purpose as a partial correlation test in a linear model
3. These p-values can be used for tests of D-Separation

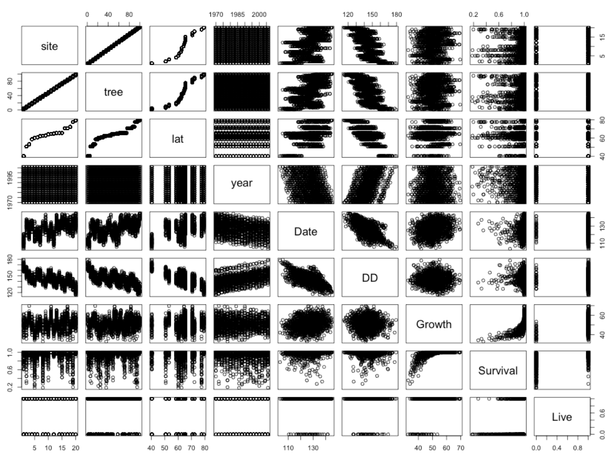
Shipley, B. (2009). Confirmatory path analysis in a generalized multilevel context. *Ecology, 90, 363–368.*

The True Model

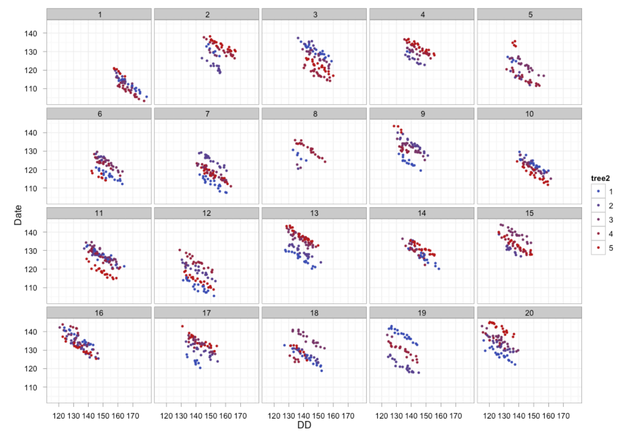
Latitude → Degree Days → Date of Bud Burst → Diameter Growth → Survival

- Simulated data from a fit model
- 20 sites
- 5 trees measured per site
- Replicated measurements biannually from "1970-2006"

The Simulated Data



Nested Structure in the Data



Piecewise Hierarchical Model Fitting

lat → DD → Date → Growth → Live

```
#e.g., for DD -> lat
Shipley<-read.table("./Shipley.dat")
library(nlme)

#model with random intercept
#tree nested in site
Date_dd<-lme(Date~DD,data=Shipley,
  random=~1|site/tree,na.action=na.omit)
```

The Basis Set Needs to Accommodate the Nested Structure

lat → DD → Date → Growth → Live

D-sep claim of independence	Mixed model†	Variable whose partial regression slope should be zero	Null probability (distribution)
$(X_1, X_3) \mid \{X_2\}$	$X_3 \sim X_2 + X_1 + (1 \mid \text{site}) + (1 \mid \text{tree})$	X_1	0.9373 (normal)
$(X_1, X_4) \mid \{X_3\}$	$X_4 \sim X_3 + X_1 + (1 \mid \text{site}) + (1 \mid \text{tree})$	X_1	0.3837 (normal)
$(X_1, X_5) \mid \{X_4\}$	$X_5 \sim X_4 + X_1 + (1 \mid \text{site}) + (1 \mid \text{tree})$	X_1	0.2800 (binomial)
$(X_2, X_4) \mid \{X_1, X_3\}$	$X_4 \sim X_3 + X_1 + X_2 + (1 \mid \text{site}) + (1 \mid \text{tree})$	X_2	0.9839 (normal)
$(X_2, X_5) \mid \{X_1, X_4\}$	$X_5 \sim X_4 + X_1 + X_2 + (1 \mid \text{site}) + (1 \mid \text{tree})$	X_2	0.9839 (binomial)
$(X_3, X_5) \mid \{X_2, X_4\}$	$X_5 \sim X_4 + X_2 + X_3 + (1 \mid \text{site}) + (1 \mid \text{tree})$	X_3	0.1890 (binomial)

To calculate the partial regression slope, use hierarchical models

Evaluate Independence Claims with Hierarchical Models

lat → DD → Date → Growth → Live

```
#Independence claim: (Date,lat)|{DD}

fit1<-lme(Date~DD+lat,data=Shipley,
  random=~1|site/tree,na.action=na.omit)

summary(fit1)$tTable
```

Evaluate Independence Claims with Hierarchical Models

lat → DD → Date → Growth → Live

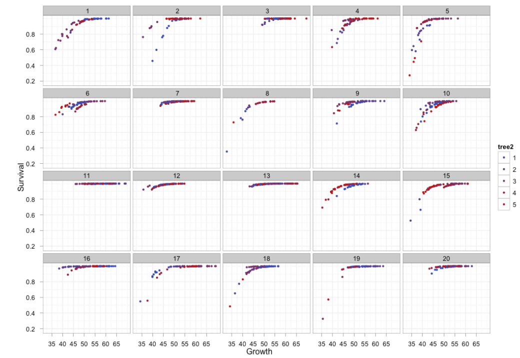
```
#Independence claim: (Date,lat)|{DD}

fit1<-lme(Date~DD+lat,data=Shipley,
  random=~1|site/tree,na.action=na.omit)
```

Evaluate Independence Claims with Hierarchical Models

lat → DD → Date → Growth → Live

```
> summary(fit1)$tTable
                   Value   Std.Error   DF      t-value       p-value
(Intercept) 198.915223483 7.337099813 1330    27.11087876 3.185667e-129
DD           -0.497660383 0.004936809 1330  -100.80608521  0.000000e+00
lat          -0.009051378 0.113476607   18    -0.07976426  9.373049e-01
```

We Have Nonlinear Relationships with Non-Normal Distributions



Use generalized linear models – e.g., logit curve with a binomial error

Evaluate Independence Claims with GLMMs

lat → DD → Date → Growth → Live

```
###need lme4 for the glmms
library(lme4)

#Independence claim with glmm (Live,lat)|{Growth}
fit4<-lmer(Live~Growth+lat+(1|site)+(1|tree),
        data=Shipley, na.action=na.omit,
        family=binomial(link="logit"))
```

Evaluate Independence Claims with GLMMs

lat → DD → Date → Growth → Live

```
> summary(fit4)@coefs
                Estimate Std. Error   z value     Pr(>|z|)
(Intercept) -14.43837636 2.65394004 -5.440355 5.317446e-08
Growth        0.35530576 0.04554481  7.801235 6.130440e-15
lat           0.03051257 0.02819180  1.082321 2.791099e-01
```

Putting it All Together in Shipley's Test

lat → DD → Date → Growth → Live

```
#sorry, you have to do this by hand
#note, since we're logging things
#we can use log(a)+log(b) = log(a*b)

> fisherC <- -2* log(9.373049e-01 * 3.836896e-01 *
                   7.667083e-01 * 2.791099e-01 *
                   3.159286e-01 * 1.519170e-01)

>1-pchisq(fisherC, 2*6)

[1] 0.5116698
```

AIC and D-Sep

lat → DD → Date → Growth → Live

$$\boxed{\textbf{AIC = -2C+ 2K}}$$

Why? Shipley has proven that:

$$-2\ \ln(L(\text{model} \mid \text{data})) = -2\ \Sigma\ \ln(p) = \text{Fisher's C}$$

Shipley, B. In Press. The AIC model selection method applied to path analytic models compared using a d-separation tests. *Ecology*.

AIC and D-Sep

lat → DD → Date → Growth → Live

```
> #each piece has 5 parameters - slope, intercept,
> #variance, and random variance for
> #slope & intercept, so, K=5*4

> fisherC + 2*(5*4)
[1] 51.20225
```

Final Thoughts on Piecewise Fits

- You can use anything: generalizes linear models, mixed models, generalized least squares fits with temporal or spatial autocorrelation built-in
- Currently, it's difficult to code complex models, but that does not mean they should not be attempted!
- Bayesian methods also provide flexible frameworks for piecewise models

An Advanced Outline

1. Revisiting Sample Size
2. Revisiting Dsep in lavaan
3. Multilevel Generalized Piecewise SEM
4. Additional Spatial Techniques
5. Panel Models for Lagged Time Effects
6. Growth Curve Models & Time Series

Spatial Effects

There are two key issues regarding space:

(1) Are their things to learn about the other factors that could explain variations in the data that vary spatially?

(2) Do we have nonindependence in our residuals?

Recent reference on the subject:
Hawkins, BA (2011) Eight (and a half) deadly sins of spatial analysis. Journal of Biogeography. doi:10.1111/j.1365-2699.2011.02637.x

Spatial References

Reference where mechanistic questions have been asked:
Grace JB and Guntenspergen, GR (1999) The effects of landscape position on plant species density: Evidence of past environmental effects in a coastal wetland. Ecoscience Vol. 6, pp. 381-391.

(Distance from mouth of river and edge of shore served as proxies for past storm-driven saltwater intrusions.)

Mancera et al. (2005) Fine-scale spatial variation in plant species richness and its relationship to environmental conditions in coastal marshlands. Plant Ecology 178:39-50.

(Showed fine-scale matching of plant to abiotic conditions in severe environments. No evidence of mass effects.)

Spatial References

Reference where autocorrelation has been adjusted for in SEM studies:
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.

Degrees of freedom and sample size adjusted using Moran's I.

Adjusting for Spatial Autocorrelation

Is there residual spatial autocorrelation and does it bias the estimates of standard errors?

Here we see a file containing residuals for floral resources and pollinators, along with (X,Y) spatial coordinates.

We can use R to examine autocorrelation and compute Moran's I.

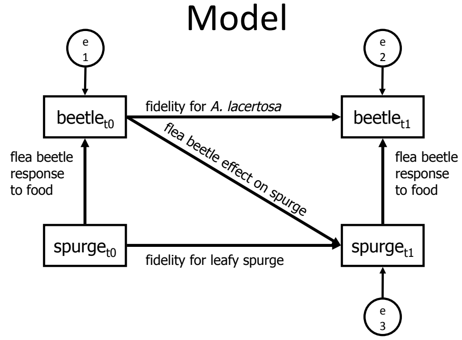
If Moran's I significant, adjust sample size.

row	N.case	N.point.X	N.point.Y	N.floralLV.res	N.pollinators.res
1	1	589216	4516511	0.104	0.286
2	2	588677	4517998	-0.109	-0.081
3	3	589302	4518208	0.377	0.352
4	4	589175	4517933	0.142	0.375
5	5	589558	4517821	0.031	0.309
6	6	589765	4516871	0.167	-0.414
7	7	588114	4514632	-0.049	0.246
8	8	588116	4514434	-0.220	0.147
9	9	588231	4514192	0.241	-0.178
10	10	588234	4513999	0.385	-0.681
11	11	588316	4513682	0.630	-0.382
12	12	588936	4516330	0.115	0.431
13	13	588142	4513542	-0.038	0.071
14	14	587824	4513656	0.388	-0.501
15	15	597328	4509049	-0.032	0.150
16	16	604711	4513071	0.125	-0.162
17	17	603973	4512903	-0.053	-0.283
18	18	603337	4512699	-0.230	-0.465
19	19	597021	4508971	-0.357	0.143
20	20	583437	4506379	-0.180	-0.033
21	21	583230	4506639	-0.353	0.126
22	22	588570	4513731	-0.585	-0.186
23	23	593364	4524027	-0.310	0.131

An Advanced Outline

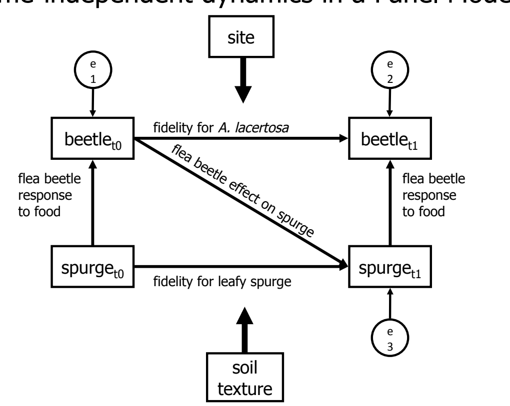
1. Revisiting Sample Size
2. Revisiting Dsep in lavaan
3. Multilevel Generalized Piecewise SEM
4. Additional Spatial Techniques
5. Panel Models for Lagged Time Effects
6. Growth Curve Models & Time Series

Longitudinal Studies – Time-Step (Panel) Model



Larson, DL and Grace, JB (2004) Temporal Dynamics of Leafy Spurge (Euphorbia esula) and Two Species of Flea Beetles (Aphthona spp.) Used as Biological Control Agents. Biological Control 29:207–214.

Time-independent dynamics in a Panel Model

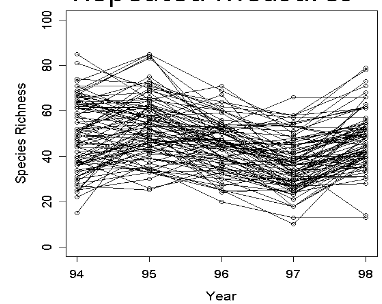


Larson, D.L., Grace, J.B., and Larson, J.L. 2008. Long-term dynamics of leafy spurge (*Euphorbia esula*) and its biocontrol agent, the flea beetle *Aphthona lacertosa*. *Biological Control* 47:250-256.

An Advanced Outline

1. Revisiting Sample Size
2. Revisiting Dsep in lavaan
3. Multilevel Generalized Piecewise SEM
4. Additional Spatial Techniques
5. Panel Models for Lagged Time Effects
6. Growth Curve Models & Time Series

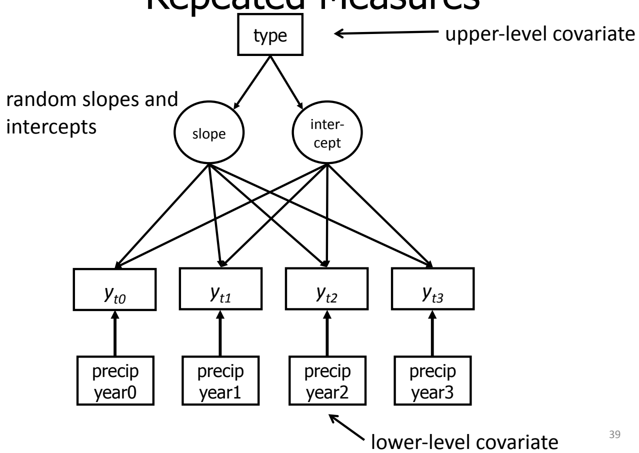
Latent Trajectory Models for Timeseries & Repeated Measures



Grace, J.B., Keeley, J., Johnson, D., and Bollen, K.A. 2012. Structural equation modeling and the analysis of long-term monitoring data. In: Gitzen, R.A., Millspaugh, J.J., Cooper, A.B., and Licht, D.S. Design and Analysis of Long-Term Ecological Monitoring Studies. Cambridge University Press.

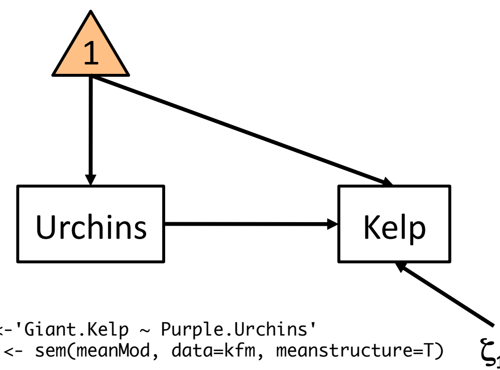
38

Latent Trajectory Models for Repeated Measures



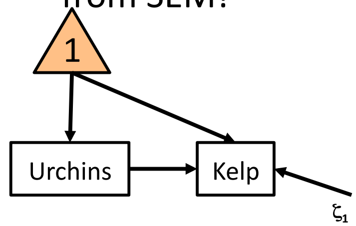
39

Means Structures: Acquiring Intercepts from SEM!



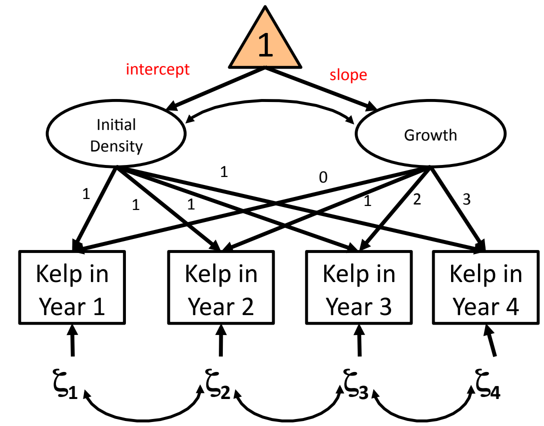

```
meanMod<-'Giant.Kelp ~ Purple.Urchins'
meanFit <- sem(meanMod, data=kfm, meanstructure=T)
```

Means Structures: Acquiring Intercepts from SEM!

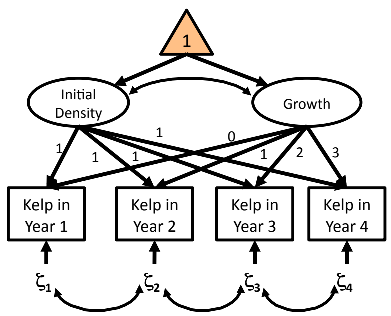



```
            Estimate  Std.err  Z-value  P(>|z|)
Regressions:
  Giant.Kelp ~
    Purple.Urchin    -0.366    0.029  -12.397    0.000

Intercepts:
    Giant.Kelp        1.590    0.076   20.791    0.000

Variances:
    Giant.Kelp        0.579    0.045   12.961    0.000
```

Latent Variable Growth Model



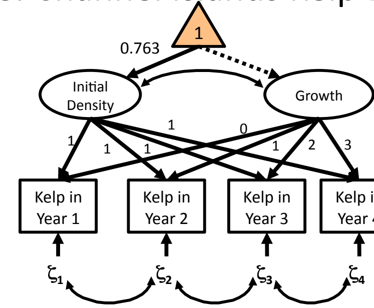
Example: Channel Islands Kelp Dynamics




```
gMod<-'
  Initial =~ 1*KelpT1 + 1*KelpT2 + 1*KelpT3 + 1*KelpT4
  Growth =~ 0*KelpT1 + 1*KelpT2 + 2*KelpT3 + 3*KelpT4
  '

gFit<-growth(gMod, data=kelpTseries)
```

Example: Channel Islands Kelp Dynamics

R^2=0.5-0.67



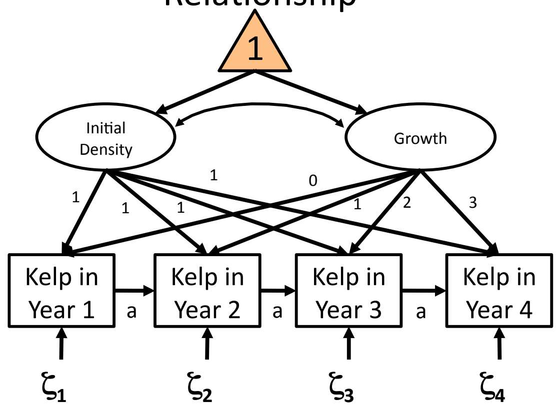
Conclusions:

At minimum, no linear trajectory.

At most, kelp densities stay constant with some small variation

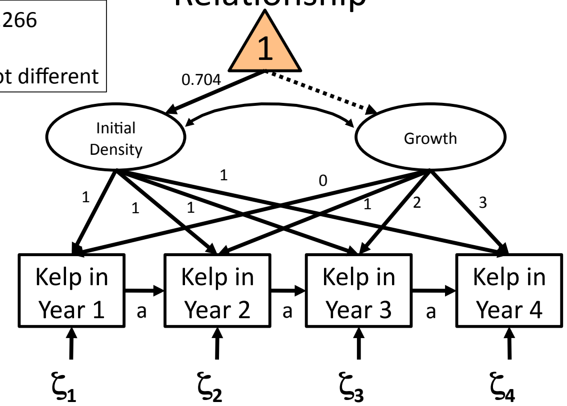
```
                    Estimate  Std.err  Z-value  P(>|z|)
Intercepts:
    KelpT1             0.000
    KelpT2             0.000
    KelpT3             0.000
    KelpT4             0.000
    Initial            0.763    0.096    7.976    0.000
    Growth             0.027    0.032    0.837    0.403
```

Growth Models and Autoregressive Relationship

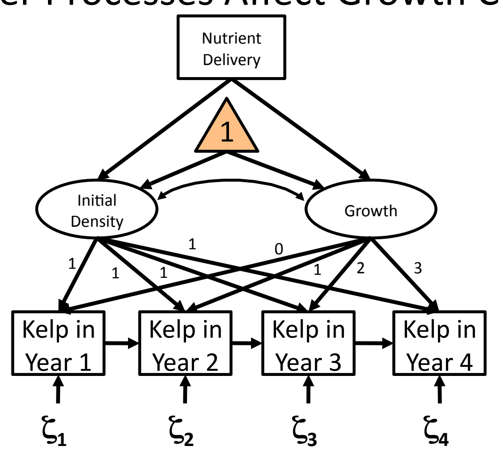


Growth Models and Autoregressive Relationship

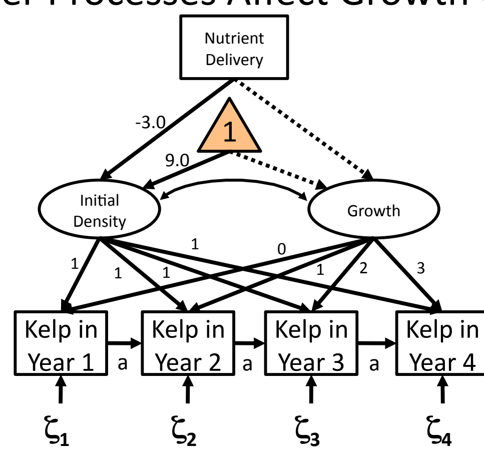
- a = 0.266
- Fit not different



Other Processes Affect Growth Curves



Other Processes Affect Growth Curves



- a = 0.304

Final Comments on Advanced Topics

1. Often, our concern for spatial and temporal effects is due to our deep ecological fear of pseudoreplication.
2. If you can account for the drivers that create spatial or temporal blocks, you gain information.
3. Many cases are more easily dealt with in a peicewise approach – still a developing story.
4. But, many special cases already have techniques in the literature that YOU can now use!