Introduction to structural equation modelling - advanced modelling techniques

Frank Pennekamp

05.11.2021

Department of Evolutionary Biology and Environmental Sciences

University of Zurich

Overview

- Advanced SEM topics covered today:

Overview

- Advanced SEM topics covered today:
    - Latent variables

Overview

- Advanced SEM topics covered today:
    - Latent variables
    - Composite variables

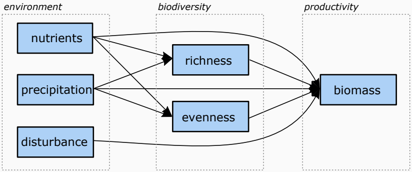
Overview

- Advanced SEM topics covered today:
    - Latent variables
    - Composite variables
    - Interactions

Overview

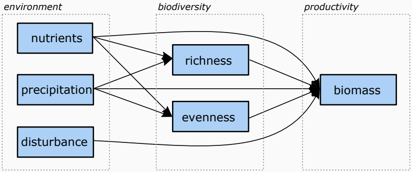
- Advanced SEM topics covered today:
    - Latent variables
    - Composite variables
    - Interactions
    - Complex survey designs

Revisiting the meta-model



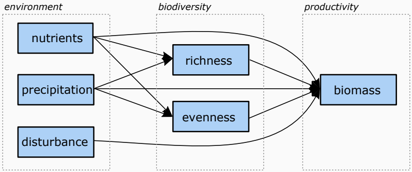
- Meta-model are conceptual models that allow to link data with theory.

Revisiting the meta-model



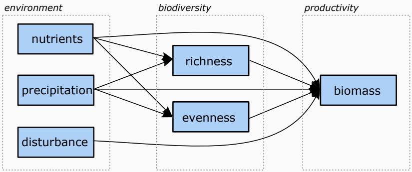
- Meta-model are conceptual models that allow to link data with theory.
- So far, we have only worked with manifest (measured) variables.

Revisiting the meta-model



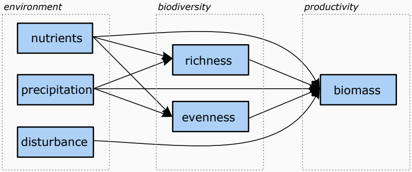
- Meta-model are conceptual models that allow to link data with theory.
- So far, we have only worked with manifest (measured) variables.
- However, some aspects of a meta-model may be difficult to quantify and measure directly.

Revisiting the meta-model



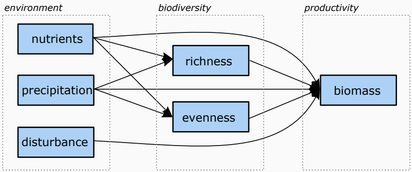
- Meta-model are conceptual models that allow to link data with theory.
- So far, we have only worked with manifest (measured) variables.
- However, some aspects of a meta-model may be difficult to quantify and measure directly.
- This is because they represent an abstract, multifaceted concept (e.g., general intelligence).

Revisiting the meta-model



- Meta-model are conceptual models that allow to link data with theory.
- So far, we have only worked with manifest (measured) variables.
- However, some aspects of a meta-model may be difficult to quantify and measure directly.
- This is because they represent an abstract, multifaceted concept (e.g., general intelligence).
- Or, because we measure variables with error.

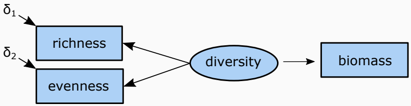
Revisiting the meta-model



- Meta-model are conceptual models that allow to link data with theory.
- So far, we have only worked with manifest (measured) variables.
- However, some aspects of a meta-model may be difficult to quantify and measure directly.
- This is because they represent an abstract, multifaceted concept (e.g., general intelligence).
- Or, because we measure variables with error.
- This is where latent and composite variables are needed.

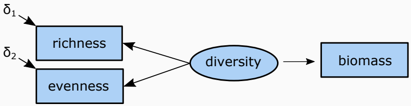
Latent variables

Latent variables



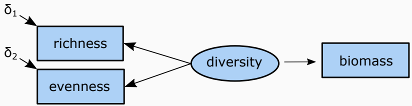
- Latent variables are unobserved, but their influence is captured by indicator variable(s).

Latent variables



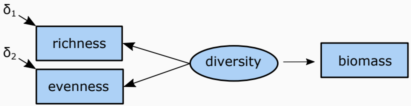
- Latent variables are unobserved, but their influence is captured by indicator variable(s).
- Latents can capture complex or conceptual properties that are difficult to quantify or measure directly.

Latent variables



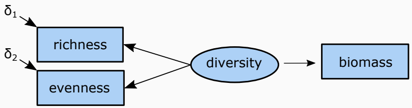
- Latent variables are unobserved, but their influence is captured by indicator variable(s).
- Latents can capture complex or conceptual properties that are difficult to quantify or measure directly.
- Graphically, latent variables are often represented by an oval node shapes (ellipses).

Latent variables



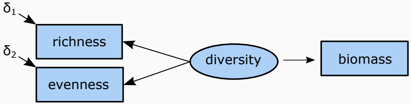
- Direction of causality reversed: from latent variables to the observed variables (reflective indicators).

Latent variables



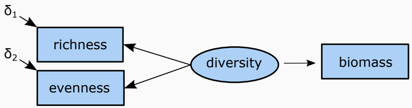
- Direction of causality reversed: from latent variables to the observed variables (reflective indicators).
- Indicator variables are emergent manifestation of the underlying phenomenon represented by the latent variable.

Latent variables



- Direction of causality reversed: from latent variables to the observed variables (reflective indicators).
- Indicator variables are emergent manifestation of the underlying phenomenon represented by the latent variable.
- All indicators should be positively correlated to the latent variable (i.e., driver).

Latent variables



- Direction of causality reversed: from latent variables to the observed variables (reflective indicators).
- Indicator variables are emergent manifestation of the underlying phenomenon represented by the latent variable.
- All indicators should be positively correlated to the latent variable (i.e., driver).
- In contrast to indicators, latent variables are free of random or systematic measurement errors.

Latent variables

- Latents can be both endogenous and exogenous.

Latent variables

- Latents can be both endogenous and exogenous.
- *Measurement model* focuses solely on relating indicators to latent variables.

Latent variables

- Latents can be both endogenous and exogenous.
- *Measurement model* focuses solely on relating indicators to latent variables.
- *Structural model* is one with directed paths between latent variables.

Latent variables

Latents often fitted in two steps:

1) Confirmatory factor analysis

Latents often fitted in two steps:

1) Confirmatory factor analysis

- Multi-indicator latent variables test hypothesis that multiple indicator variables are generated by the same underlying process.

Latent variables

Latents often fitted in two steps:

1) Confirmatory factor analysis

- Multi-indicator latent variables test hypothesis that multiple indicator variables are generated by the same underlying process.
- Precursor to evaluation of any structural models in which the latent variables appear.

Latent variables

Latents often fitted in two steps:

1) Confirmatory factor analysis

   - Multi-indicator latent variables test hypothesis that multiple indicator variables are generated by the same underlying process.
   - Precursor to evaluation of any structural models in which the latent variables appear.

2) Full model including latent variable

Latent variables

Latents often fitted in two steps:

1) Confirmatory factor analysis

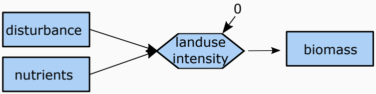
- Multi-indicator latent variables test hypothesis that multiple indicator variables are generated by the same underlying process.
- Precursor to evaluation of any structural models in which the latent variables appear.

2) Full model including latent variable

- To estimate effect of latent variable on other variables.

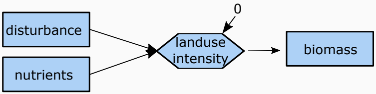
Composite variables

Composite variables



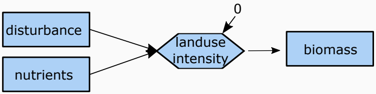
- Composite variables specify the influence of collections of other variables (e.g., landuse).

Composite variables



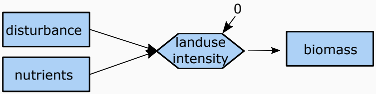
- Composite variables specify the influence of collections of other variables (e.g., landuse).
- In comparison to latent variables, they arise from the indicators.

Composite variables



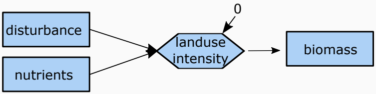
- Composite variables specify the influence of collections of other variables (e.g., landuse).
- In comparison to latent variables, they arise from the indicators.
- Visualized by arrows pointing from the indicators to the composite (formative or causal indicators).

Composite variables



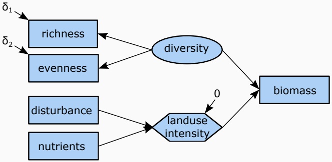
- Composite variables specify the influence of collections of other variables (e.g., landuse).
- In comparison to latent variables, they arise from the indicators.
- Visualized by arrows pointing from the indicators to the composite (formative or causal indicators).
- Values are determined by its causes (indicators), thus error variance is set to 0.

Composite variables



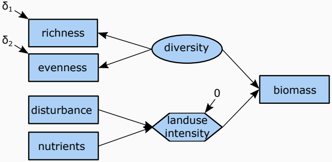
- Composite variables specify the influence of collections of other variables (e.g., landuse).
- In comparison to latent variables, they arise from the indicators.
- Visualized by arrows pointing from the indicators to the composite (formative or causal indicators).
- Values are determined by its causes (indicators), thus error variance is set to 0.
- Composites are often shown as hexangular shapes or ellipses (latent composites).

Key differences between latent and composite variables



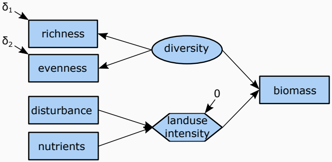
- Flow of causation in a latent variables is from the construct to its indicators, i.e., indicators are driven by an underlying, unmeasured process.

Key differences between latent and composite variables



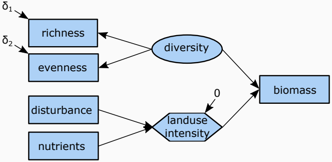
- Flow of causation in a latent variables is from the construct to its indicators, i.e., indicators are driven by an underlying, unmeasured process.
- As a latent variable represents an underlying, data-generating process, its indicators need to be correlated to each other.

Key differences between latent and composite variables



- Flow of causation in a latent variables is from the construct to its indicators, i.e., indicators are driven by an underlying, unmeasured process.
- As a latent variable represents an underlying, data-generating process, its indicators need to be correlated to each other.
- For composite variables, the flow of causation is reversed and the indicators are independent entities.

Key differences between latent and composite variables



- Flow of causation in a latent variables is from the construct to its indicators, i.e., indicators are driven by an underlying, unmeasured process.
- As a latent variable represents an underlying, data-generating process, its indicators need to be correlated to each other.
- For composite variables, the flow of causation is reversed and the indicators are independent entities.
- In the case of a composite variable, there is no assumption about the relation between the indicators.

Key differences between latent and composite variables

How distinguish latent and composites:

- Rule of thumb 1: if the indicators are redundant, they likely belong to a latent variable.

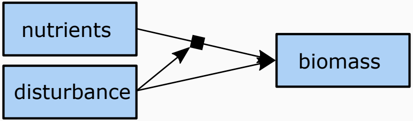
Key differences between latent and composite variables

How distinguish latent and composites:

- Rule of thumb 1: if the indicators are redundant, they likely belong to a latent variable.
- Rule of thumb 2: If the meaning of the construct changes after dropping one of the indicators, the construct likely is a composite.

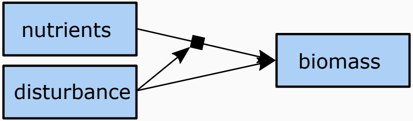
Interactions

Interactions



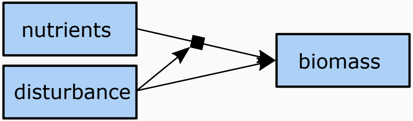
- In nature, things often are contingent on each other.

Interactions



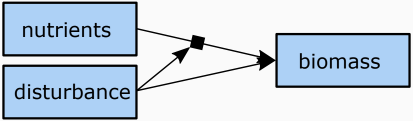
- In nature, things often are contingent on each other.
- For instance, the effect of nutrients on plant growth, may depend on how disturbed the environment is.

Interactions



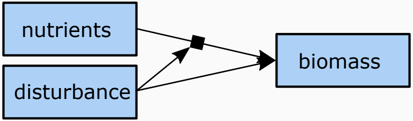
- In nature, things often are contingent on each other.
- For instance, the effect of nutrients on plant growth, may depend on how disturbed the environment is.
- Such a behaviour is called an interaction, which indicates that the effect of the two main effects are different when combined.

Interactions



- In nature, things often are contingent on each other.
- For instance, the effect of nutrients on plant growth, may depend on how disturbed the environment is.
- Such a behaviour is called an interaction, which indicates that the effect of the two main effects are different when combined.
- Both positive and negative interactions are possible.

Interactions



- In nature, things often are contingent on each other.
- For instance, the effect of nutrients on plant growth, may depend on how disturbed the environment is.
- Such a behaviour is called an interaction, which indicates that the effect of the two main effects are different when combined.
- Both positive and negative interactions are possible.
- In regression, the interaction is represented by a coefficient that estimates the effect of the product of the two predictors.

Interactions can be modelled in different ways in lavaan:

1) Multiple groups

Interactions can be modelled in different ways in lavaan:

1) Multiple groups
2) Composites

Interactions (multi group)

```
mod <- sem(model, group = "age_class", data = dd)
```

- Multigroup fitting allows coefficients to vary among groups.

Interactions (multi group)

```
mod <- sem(model, group = "age_class", data = dd)
```

- Multigroup fitting allows coefficients to vary among groups.
- Lavaan offers the “group” argument to specify for which groups coefficients should be estimated.

Interactions (multi group)

```
mod <- sem(model, group = "age_class", data = dd)
```

- Multigroup fitting allows coefficients to vary among groups.
- Lavaan offers the “group” argument to specify for which groups coefficients should be estimated.
- Importantly, groups have to be categorical (e.g., sex, age class).

Interactions (multi group)

Lavaan allows to introduce equality constraints on various aspects via the `group.equal` argument:

```
mod <- sem(model, group = "age_class", group.equal =
c("regressions"), data = dd)
```

Additional constraints could be:

```
group.equal=c(
"intercepts",
"means",
"regressions",
"residuals",
"residual.covariances")
```

- Even more control by having the same name for different parameters:

```
model <- '
y ~ c("b1", "b1") * x1 + c("b2", "b2") * x3
x2 ~ c("b3", "b4") * x1
x3 ~ c("b5", "b5") * x2
'
```

Same coefficients for all but the effect of x1 on x2.

Heywood cases: when things go wrong

- Improper solutions:

Heywood cases : when things go wrong

- Improper solutions:
    - Negative variances

Heywood cases : when things go wrong

- Improper solutions:
    - Negative variances
    - Correlations larger than 1

Heywood cases : when things go wrong

- Improper solutions:
    - Negative variances
    - Correlations larger than 1
- Reasons for the warnings are:

Heywood cases : when things go wrong

- Improper solutions:
  - Negative variances
  - Correlations larger than 1
- Reasons for the warnings are:
  1) slightly negative error estimates (not really a problem)

Heywood cases : when things go wrong

- Improper solutions:
    - Negative variances
    - Correlations larger than 1
- Reasons for the warnings are:
    1) slightly negative error estimates (not really a problem)
    2) indicators for latent variables need to be positively correlated

Heywood cases : when things go wrong

- Improper solutions:
    - Negative variances
    - Correlations larger than 1
- Reasons for the warnings are:
    1) slightly negative error estimates (not really a problem)
    2) indicators for latent variables need to be positively correlated
    3) local non-identification

Heywood cases : when things go wrong

- Improper solutions:
    - Negative variances
    - Correlations larger than 1
- Reasons for the warnings are:
    1) slightly negative error estimates (not really a problem)
    2) indicators for latent variables need to be positively correlated
    3) local non-identification
    4) general misspecification

Complex sampling structure

Complex sampling structure

- Needed when data is nested:

Complex sampling structure

- Needed when data is nested:
    - within sites

Complex sampling structure

- Needed when data is nested:
    - within sites
    - within groups such as families, nests

Complex sampling structure

- Needed when data is nested:
  - within sites
  - within groups such as families, nests
- Nesting violates the principle of being independent and identically distributed.

Complex sampling structure

- Needed when data is nested:
    - within sites
    - within groups such as families, nests
- Nesting violates the principle of being independent and identically distributed.
- Necessary to account for this structure in the data in the model.

Complex sampling structure

```
library("lavaan.survey")
```

- The add-on package lavaan.survey allows the analysis of stratified, clustered or weighted data.

Complex sampling structure

```
library("lavaan.survey")
```

- The add-on package lavaan.survey allows the analysis of stratified, clustered or weighted data.
- lavaan objects can be processed further with a specific data structure:

Complex sampling structure

```
library("lavaan.survey")
```

- The add-on package lavaan.survey allows the analysis of stratified, clustered or weighted data.
- lavaan objects can be processed further with a specific data structure:
    1) initialize the design

Complex sampling structure

```
library("lavaan.survey")
```

- The add-on package lavaan.survey allows the analysis of stratified, clustered or weighted data.
- lavaan objects can be processed further with a specific data structure:
    1) initialize the design
    2) post-process the lavaan object and compute the adjusted results.

```
library("lavaan.survey")
```

- The add-on package lavaan.survey allows the analysis of stratified, clustered or weighted data.
- lavaan objects can be processed further with a specific data structure:
    1) initialize the design
    2) post-process the lavaan object and compute the adjusted results.
    3) Result is a corrected lavaan object.

Complex sampling structure

```
design <- svydesign(ids = ~ plot, strata = ~ field, nest = TRUE,
summary(design)

fit.nested <- lavaan.survey(lavaan.fit = model, survey.design =
```

- Needed to specify the study design

Complex sampling structure

```
design <- svydesign(ids = ~ plot, strata = ~ field, nest = TRUE,
summary(design)

fit.nested <- lavaan.survey(lavaan.fit = model, survey.design =
```

- Needed to specify the study design
- Here we have plots (`ids`) nested in fields (`strata`)

Complex sampling structure

```
design <- svydesign(ids = ~ plot, strata = ~ field, nest = TRUE,
summary(design)

fit.nested <- lavaan.survey(lavaan.fit = model, survey.design =
```

- Needed to specify the study design
- Here we have plots (`ids`) nested in fields (`strata`)
- Next, we can refit the `simple` model from before with `lavaan.survey` using the specified study design as an argument.

Questions?

Live coding session

Your turn: working with the Seabloom dataset

Exercise 1

Start with the following model:

```
library("lavaan")

simple <-
"mass.above ~ nadd + disk + rich + even + precip.mm
rich ~ nadd + precip.mm
even ~ nadd + precip.mm

rich ~~ even"

fit.simple <- sem(simple, data = seabloom, estimator = "MLM")
summary(fit.simple)
```

Exercise 2

- Construct latent variable diversity based on richness, evenness and ens.pie

Exercise 2

- Construct latent variable diversity based on richness, evenness and ens.pie
    - Run confirmatory factor analysis

Exercise 2

- Construct latent variable diversity based on richness, evenness and ens.pie
    - Run confirmatory factor analysis
    - What do you conclude?

Exercise 3

- Construct latent variable diversity only based on richness and evenness

Exercise 3

- Construct latent variable diversity only based on richness and evenness
    - Run confirmatory factor analysis

Exercise 3

- Construct latent variable diversity only based on richness and evenness
  - Run confirmatory factor analysis
  - Include latent variable into full model

Exercise 3

- Construct latent variable diversity only based on richness and evenness
    - Run confirmatory factor analysis
    - Include latent variable into full model
    - What do you conclude?

Exercise 4

- Construct composite variable landuse based on nutrients and disturbance (disk):

Exercise 4

- Construct composite variable landuse based on nutrients and disturbance (disk):
    - What do you conclude?

Exercise 4

- Construct composite variable landuse based on nutrients and disturbance (disk):
    - What do you conclude?
    - Build composite manually.

Exercise 5

- Investigate possible interaction between disturbance and nutrient addition on AGB:

Exercise 5

- Investigate possible interaction between disturbance and nutrient addition on AGB:
    - Construct composite variable to capture a possible interaction:

Exercise 5

- Investigate possible interaction between disturbance and nutrient addition on AGB:
    - Construct composite variable to capture a possible interaction:
    - What do you conclude?

Exercise 5

- Investigate possible interaction between disturbance and nutrient addition on AGB:
    - Construct composite variable to capture a possible interaction:
    - What do you conclude?
    - Incorporate composite into full model

Exercise 6

- Investigate possible interaction between disturbance and nutrient addition on AGB:

Exercise 6

- Investigate possible interaction between disturbance and nutrient addition on AGB:
    - Use multi-group fitting to explore interaction.

Exercise 6

- Investigate possible interaction between disturbance and nutrient addition on AGB:
    - Use multi-group fitting to explore interaction.
    - What do you conclude?

Exercise 7

- Account for nested experimental design with the `lavaan.survey` package:

Exercise 7

- Account for nested experimental design with the `lavaan.survey` package:
    - Add individual plots nested in fields.