

Subgroup Discovery in Structural Equation Models

Christoph Kiefer, Bielefeld University, Germany

joint work with

Florian Lemmerich, University of Passau, Germany

Benedikt G. Langenberg and Axel Mayer, Bielefeld University, Germany

- **SubgroupSEM**
mining for subgroups with exceptional parameter constellations in a structural equation model (SEM)
- **Structural Equation Modeling**
Framework Combining Path and Measurement Error Models
- **R package** subgroupsem
 - SEM R package lavaan with
 - Python module pysubgroup
 - via R package reticulate

Introduction

- **SubgroupSEM**
mining for subgroups with exceptional parameter constellations in a structural equation model (SEM)
- **Structural Equation Modeling**
Framework Combining Path and Measurement Error Models
- **R package** subgroupsem
 - SEM R package lavaan with
 - Python module pysubgroup
 - via R package reticulate

Let us see, how it works!

A First Example

1. The Target Model
2. The Covariate Space
3. The Interestingness Measure
4. The Search Algorithm

A First Example

1. The Target Model

2. The Covariate Space

3. The Interestingness Measure

4. The Search Algorithm

Example Dataset: *HolzingerSwineford1939*

```
> head(HolzingerSwineford1939)
  id sex ageyr age  school grade  x1  x2  x3
1   1   1   13    1 Pasteur    7 3.33 7.75 0.375
2   2   2   13    7 Pasteur    7 5.33 5.25 2.125
3   3   2   13    1 Pasteur    7 4.50 5.25 1.875
4   4   1   13    2 Pasteur    7 5.33 7.75 3.000
5   5   2   12    2 Pasteur    7 4.83 4.75 0.875
6   6   2   14    1 Pasteur    7 5.33 5.00 2.250
```

A First Example

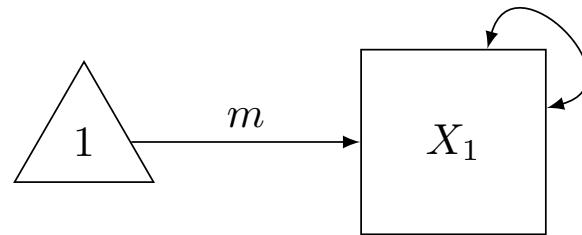
1. The Target Model

2. The Covariate Space

3. The Interestingness Measure

4. The Search Algorithm

- **Path Diagram of Model**



- **lavaan Syntax of Model:**

```
model <- 'x1 ~ m*1'
```

A First Example

1. The Target Model

2. The Covariate Space

3. The Interestingness Measure

4. The Search Algorithm

- **Names of Variables**

Character Vector

```
covariates <- c("school",  
               "grade",  
               "sex")
```

- **Search Space**

generated from possible combinations

- students in Pasteur school
- students in grade 7
- students in grade 7 of Pasteur school
- ...

A First Example

1. The Target Model

2. The Covariate Space

3. The Interestingness Measure

4. The Search Algorithm

- *What is interesting for you?*

- lavaan **Syntax for Interestingness Measure**

```
qf <- 'subsem_qf := m'
```

- *Or something more sophisticated:*

```
qf <- '  
  subsem_qf := abs(4.93577 - m)  
,
```


A First Example

1. The Target Model

2. The Covariate Space

3. The Interestingness Measure

4. The Search Algorithm

- **For the Moment:** *Run the defaults!*

```
fit <- subsem(  
  model = model,  
  data = HolzingerSwineford1939,  
  qf = qf,  
  predictors = covariates  
)
```

A First Example

```
> summary(fit)
```

```
General information:
```

```
Elapsed time: 3.151902 secs
```

```
Summary of subgroup search:
```

	quality	subgroup	size_sg
1	4.958299	grade==7 AND sex==2	83
2	4.953876	grade==7	157
3	4.953876	grade==8	143
4	4.952451	grade==7 AND school=='Pasteur' AND sex==2	40
5	4.949242	grade==8 AND school=='Pasteur' AND sex==1	36
6	4.946136	grade==8 AND sex==1	71
7	4.945550	grade==7 AND school=='Pasteur'	78
8	4.943906	grade==8 AND school=='Grant-White'	65
9	4.943282	school=='Pasteur' AND sex==2	82
10	4.943181	school=='Pasteur' AND sex==1	74

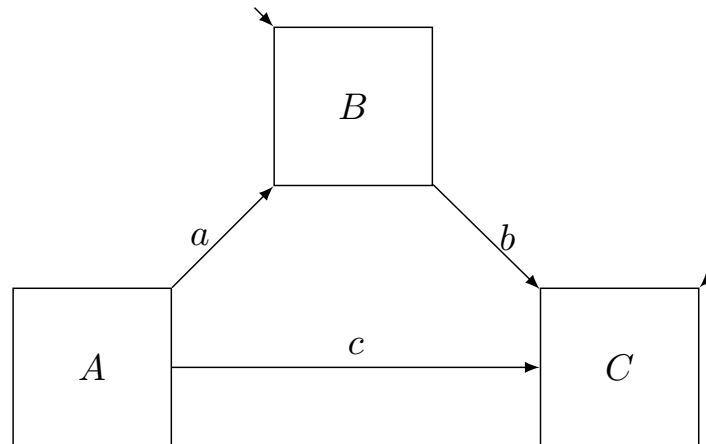
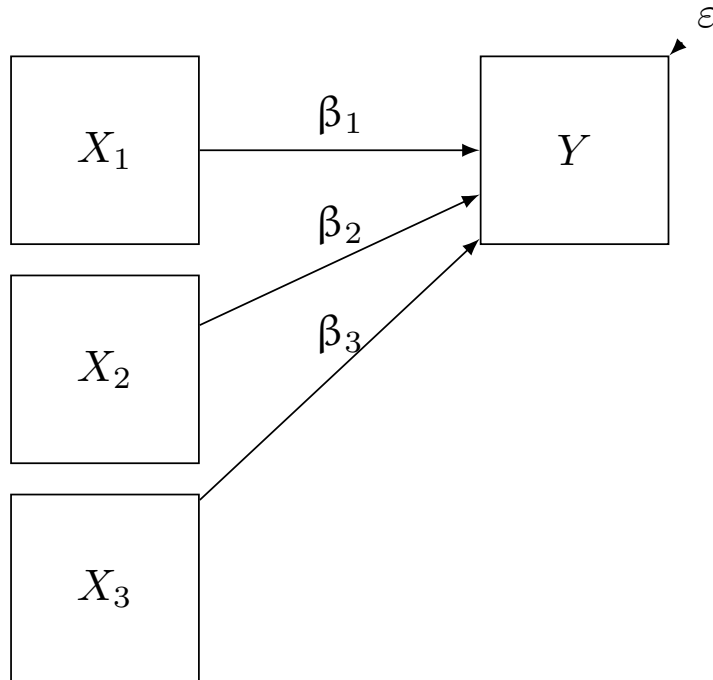
Structural Equation Modeling

Framework Combining Path and Measurement Error Models

Getting Deeper Into SEM

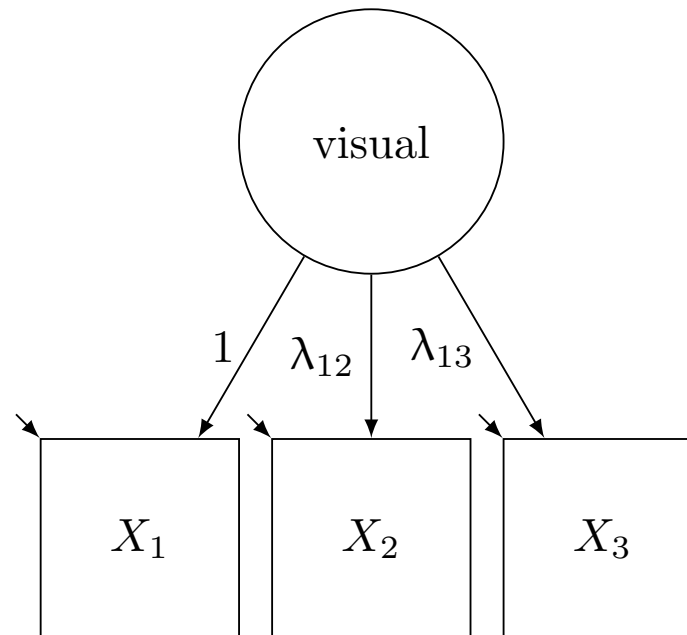
Structural Equation Modeling

Framework Combining Path and Measurement Error Models



Structural Equation Modeling

Framework Combining Path and Measurement Error Models



Getting Deeper Into SEM

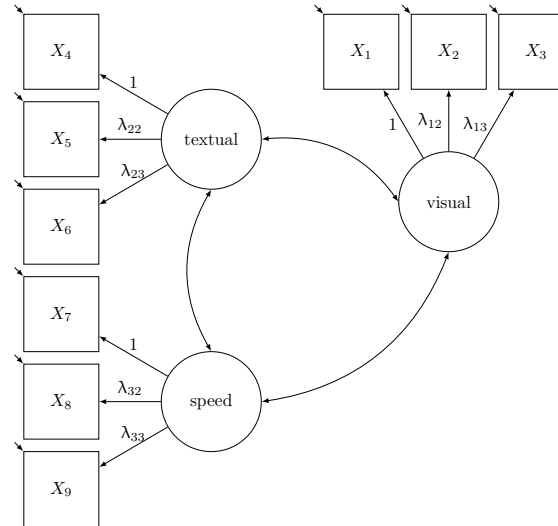
1. The Target Model

2. The Covariate Space

3. The Interestingness Measure

4. The Search Algorithm

• Path Diagram of Model



• lavaan Syntax of Model:

```
model <- "  
eta1 =~ x1 + x2 + x3  
eta2 =~ x4 + x5 + x6  
eta3 =~ x7 + x8 + x9  
"
```

1. The Target Model

2. The Covariate Space

3. The Interestingness Measure

4. The Search Algorithm

- **Exploratory Interestingness Measures**
any function of estimated parameters

- **Test Statistic-based Measures**
 - Likelihood Ratio Test
 - Wald Test
 - Delta Method

- **Example:**
Likelihood Ratio Test (LRT) Statistic

```
fit <- subsem_lrt(  
  model = model,  
  data = HolzingerSwineford1939,  
  predictors = covariates  
)
```

Getting Deeper Into SEM

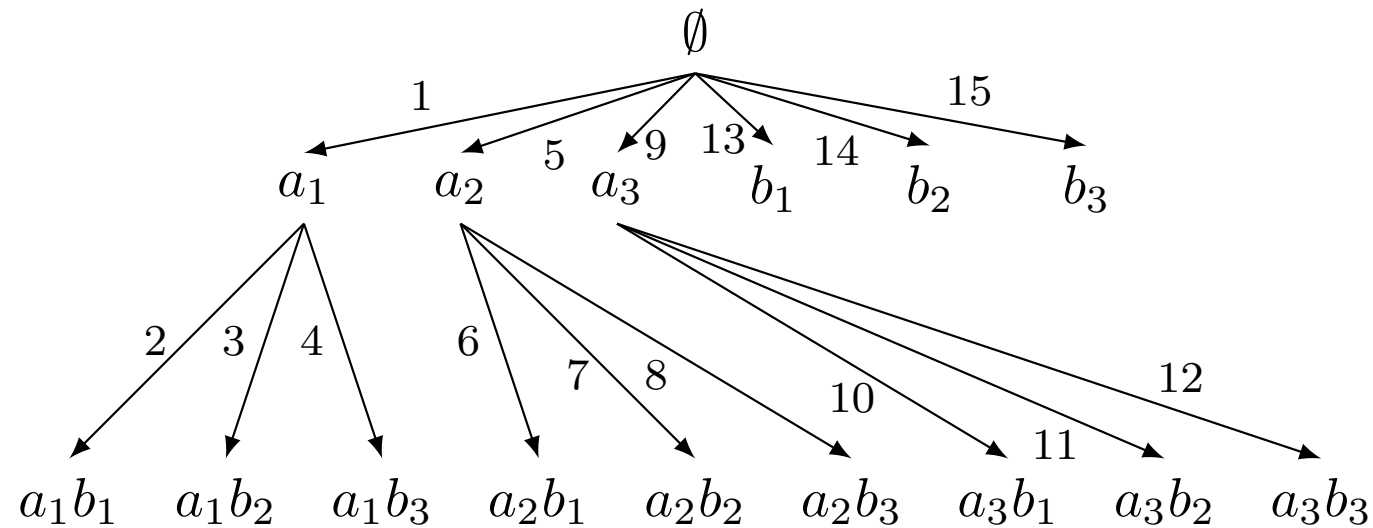
Summary of subgroup search:

	quality	subgroup	size_sg
1	111.09483	school=='Grant-White'	145
2	111.09483	school=='Pasteur'	156
3	107.82037	grade==7 AND school=='Grant-White'	79
4	99.36034	grade==7 AND school=='Grant-White' AND sex==2	43
5	97.03249	grade==7	157
6	97.03249	grade==8	143
7	94.98001	grade==8 AND school=='Grant-White'	65
8	92.99673	grade==8 AND sex==2	72
9	90.87883	grade==8 AND school=='Pasteur'	78
10	89.37181	school=='Grant-White' AND sex==2	73

Subgroup Discovery Algorithms

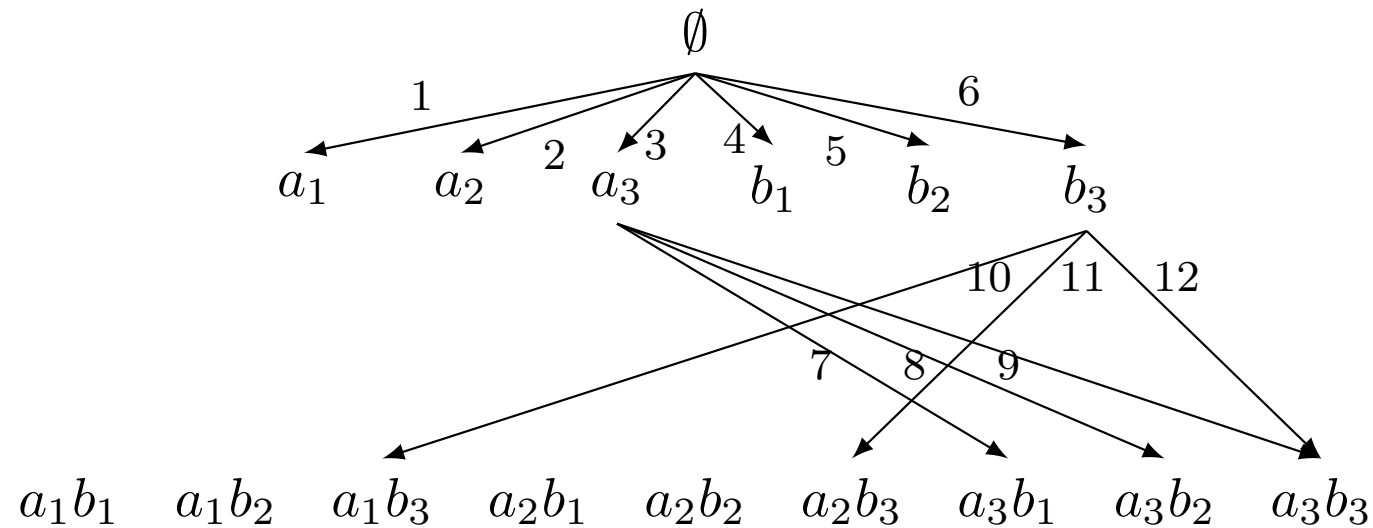
Exhaustive Search

Example: *Simple depth first search*



Heuristic Search

Example: *Beam search*



1. The Target Model

2. The Covariate Space

3. The Interestingness Measure

4. The Search Algorithm

- **Subgroup Discovery Algorithms**

- Depth First Search
- Beam Search
- more to come...

- **Pruning Options**

- Search Depth
- Minimum Subgroup Size

- **R Code**

```
fit <- subsem_lrt(  
  \\ ... as above... \\  
  subsem_options =  
    list(algorithm = "Beam",  
          search_depth = 3L,  
          min_subgroup_size = 50L)  
)
```

Last Slide

Further Information

- Kiefer, C., Lemmerich, F., Langenberg, B. G., & Mayer, A. (in press). **Subgroup Discovery in Structural Equation Models**. *Psychological Methods*. doi: 10.1037/met0000524
 - comparison to finite mixture and decision tree approaches
 - more details on statistical and algorithmic components
 - extended example with large-scale panel data and high-dimensional covariate space
- **Preprint** on <https://psyarxiv.com/c6qv4>
- **R package** subgroupsem installable from <https://github.com/chkiefer/subgroupsem>

Thank you for your attention!