Detecting parameter heterogeneity in psychometric models by means of model-based recursive partitioning with `psychotree`, `stablelearner` & co.

Carolin Strobl

jointly with: Achim Zeileis, Florian Wickelmaier, Julia Kopf, Basil Komboz, Michel Philipp, Thomas Rusch, Kurt Hornik, Lennart Schneider, Rudolf Debelak and Mirka Henninger

University of Zurich[UZH]

CART

Early statistical problems

MOB

for BT models

for Rasch models

for non-Rasch models

Can we trust the results?

Stability

Effect size stopping

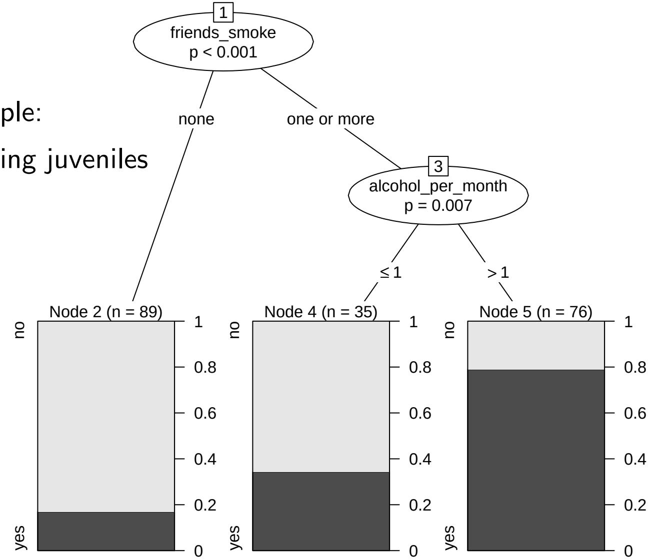
Summary

References

Classification and regression trees (CART)

Example:

Smoking juveniles



R-packages `party` (Hothorn, Hornik, Strobl and Zeileis) and `partykit`

Statistical problems of the early algorithms

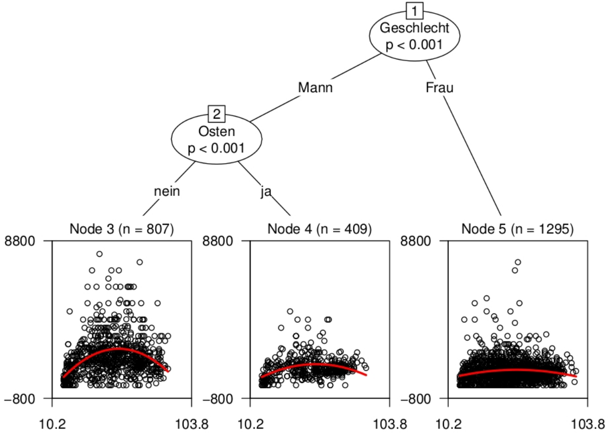
Breiman et al. (1984), Quinlan (1986)

- variable selection bias
- pruning vs. early stopping

e.g.: Hothorn, Hornik and Zeileis (2006, *Journal of Computational and Graphical Statistics*), Strobl, Boulesteix and Augustin (2007, *Computational Statistics & Data Analysis*); tutorial paper: Strobl, Malley and Tutz (2009, *Psychological Methods*)

Model-based recursive partitioning (MOB)

Zeileis, Hothorn and Hornik (2008, *JCGS*)



Example: Income as a function of age (ALLBUS 2008)

Kopf, Augustin and Strobl (2013)

MOB for psychometric models (package `psychotree`)

- **Bradley-Terry models**

  Strobl, Wickelmaier and Zeileis (2011, *Journal of Educational and Behavioral Statistics*); Eugster, Leisch and Strobl (2014, *Computational Statistics & Data Analysis*); Müller, Strobl et al. (2015, *Journal of Applied Ecology*)

- **binary Rasch models**

  Strobl, Kopf and Zeileis (2015, *Psychometrika*)

- **polytomous Rasch models**

  Komboz, Strobl, Zeileis (2016, *Educational and Psychological Measurement*)

- **2PL etc. models**

  Schneider, Strobl, Zeileis and Debelak (2021, Behavior Research Methods); Debelak and Strobl, (2019, Educational and Psychological Measurement)

MOB for psychometric models (package `psychotree`)

- **Bradley-Terry models**

  Strobl, Wickelmaier and Zeileis (2011, *Journal of Educational and Behavioral Statistics*); Eugster, Leisch and Strobl (2014, *Computational Statistics & Data Analysis*); Müller, Strobl et al. (2015, *Journal of Applied Ecology*)

- **binary Rasch models**

  Strobl, Kopf and Zeileis (2015, *Psychometrika*)

- **polytomous Rasch models**

  Komboz, Strobl, Zeileis (2016, *Educational and Psychological Measurement*)

- **2PL etc. models**

  Schneider, Strobl, Zeileis and Debelak (2021, Behavior Research Methods); Debelak and Strobl, (2019, Educational and Psychological Measurement)

Example: MOB for BT models – Topmodels

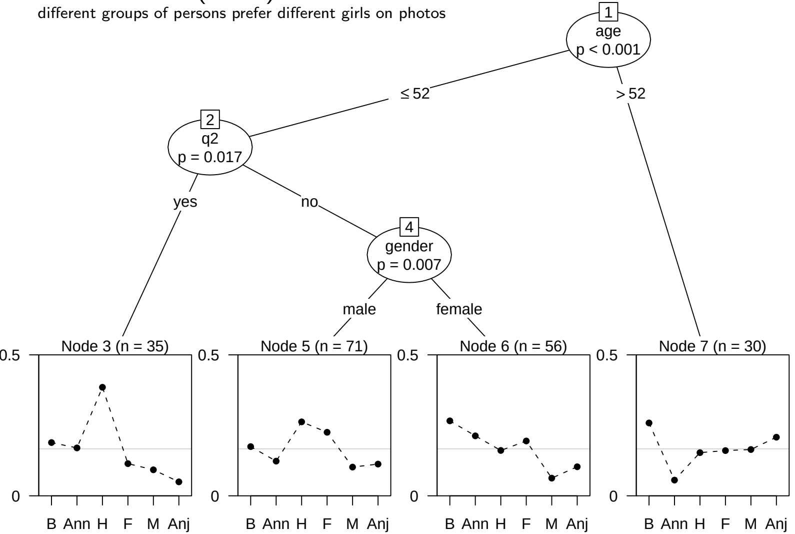
- sample: $n = 192$ (96 female and 96 male) raters between the age of 15 and 77
- covariates: gender, age and
  - (q1) Do you know the women on the photos? Do you know the TV show Germany's Next Topmodel?
  - (q2) Did you watch the latest season of Germany's Next Topmodel regularly? (2007)
  - (q3) Have you seen the final of the latest season of Germany's Next Topmodel? Do you know who won the latest season of Germany's Next Topmodel?

  where "yes" to one or more parts = overall "yes"
- design: forced choice full paired comparison of photos

Example: MOB for BT models – Topmodels

Strobl et al. (2011)

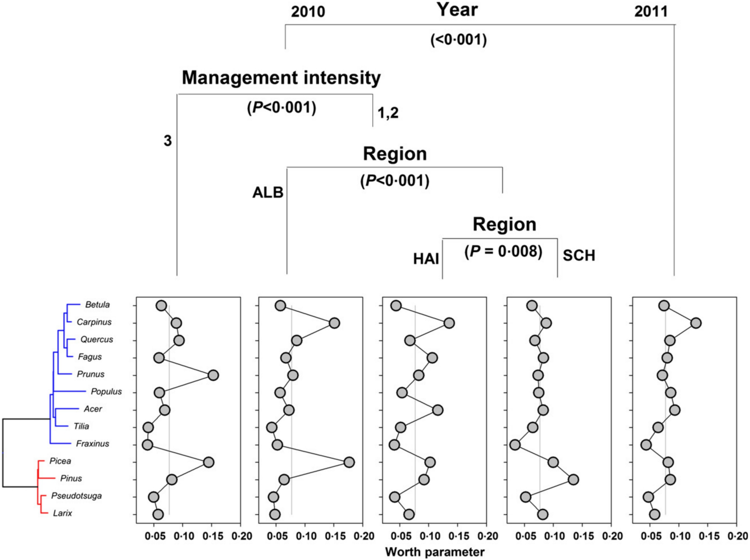
different groups of persons prefer different girls on photos



Example: MOB for BT models – Beetles

Müller et al. (2015)

different types of habitats attract different types of beetles



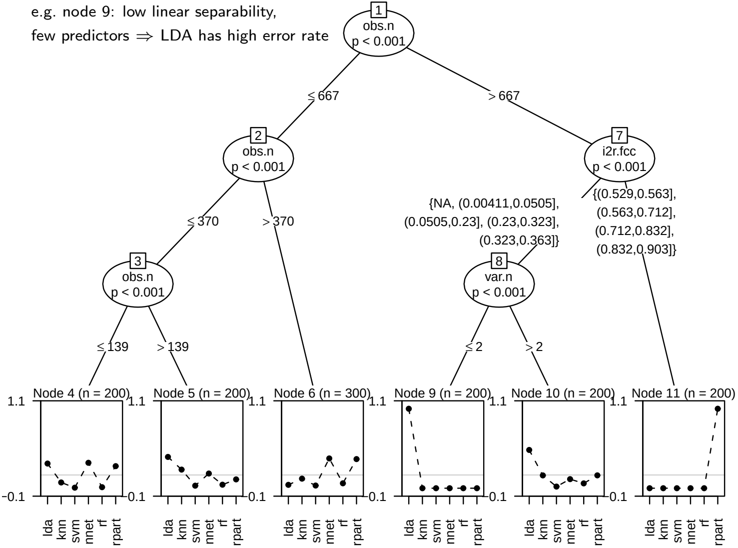
Example: MOB for BT models – Benchmarking

Eugster, Leisch and Strobl (2014)

different types of data sets let different learners perform well

e.g. node 9: low linear separability,

few predictors ⇒ LDA has high error rate



MOB for psychometric models (package `psychotree`)

- **Bradley-Terry models**

  Strobl, Wickelmaier and Zeileis (2011, *Journal of Educational and Behavioral Statistics*); Eugster, Leisch and Strobl (2014, *Computational Statistics & Data Analysis*); Müller, Strobl et al. (2015, *Journal of Applied Ecology*)

- **binary Rasch models**

  Strobl, Kopf and Zeileis (2015, *Psychometrika*)

- **polytomous Rasch models**

  Komboz, Strobl, Zeileis (2016, *Educational and Psychological Measurement*)

- **2PL etc. models**

  Schneider, Strobl, Zeileis and Debelak (2021, Behavior Research Methods); Debelak and Strobl, (2019, Educational and Psychological Measurement)

Detecting parameter heterogeneity

CART

Early statistical problems

MOB

for BT models

for Rasch models

for non-Rasch models

Can we trust the results?

Stability

Effect size stopping

Summary

References

MOB for psychometric models (package `psychotree`)

- **Bradley-Terry models**

  Strobl, Wickelmaier and Zeileis (2011, *Journal of Educational and Behavioral Statistics*); Eugster, Leisch and Strobl (2014, *Computational Statistics & Data Analysis*); Müller, Strobl et al. (2015, *Journal of Applied Ecology*)

- **binary Rasch models**

  Strobl, Kopf and Zeileis (2015, *Psychometrika*)

- **polytomous Rasch models**

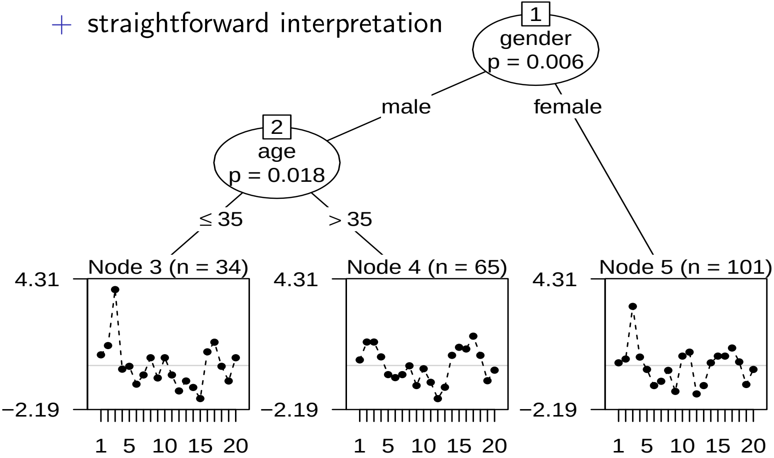
  Komboz, Strobl, Zeileis (2016, *Educational and Psychological Measurement*)

- **2PL etc. models**

  Schneider, Strobl, Zeileis and Debelak (2021, Behavior Research Methods); Debelak and Strobl, (2019, Educational and Psychological Measurement)

MOB for binary Rasch models

+ identifies previously unknown groups with DIF
+ straightforward interpretation



Strobl, Kopf and Zeileis (2015, *Psychometrika*)

cutpoint stopping

CART
Early statistical problems
MOB
for BT models
for Rasch models
for non-Rasch models
Can we trust the results?
Stability
Effect size stopping
Summary
References

Selection of splitting variables

identify groups of people with different parameters

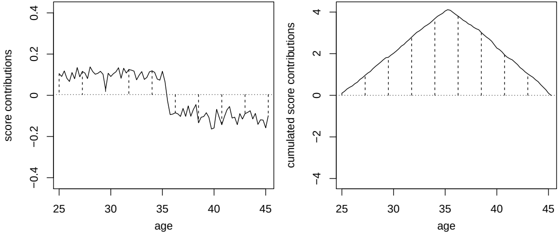
by means of tests for parameter instability (score tests, LM tests):

- individual contributions to the score-funktion

$$\psi(y_i, \theta) = \frac{\partial \Psi(y_i, \theta)}{\partial \theta}$$

- cumulated over all values of covariate ℓ

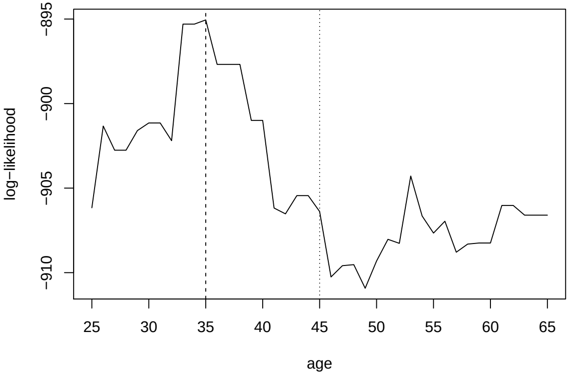
$$W_\ell(t) = \hat{V}^{-1/2} n^{-1/2} \sum_{i=1}^{\lfloor n \cdot t \rfloor} \psi(y_{(i|\ell)}, \hat{\theta})$$



- under H_0 the path fluctuates randomly around zero (→ Brownian bridge, Zeileis & Hornik, 2007)

Selection of cutpoints

maximize partitioned log-likelihood tree



$$\sum_{i \in L(\xi)} \Psi\left(y_i, \hat{\theta}^{(L)}\right) + \sum_{i \in R(\xi)} \Psi\left(y_i, \hat{\theta}^{(R)}\right)$$

CART
Early statistical problems
MOB
for BT models
for Rasch models
for non-Rasch models
Can we trust the results?
Stability
Effect size stopping
Summary
References

Stopping criteria

- p value > 0.05 tree

		Node 1	Node 2	Node 3	...
Age	Statistic	41.237	48.448	28.924	...
	p value	0.171	0.018*	0.593	...
Gender	Statistic	41.479	—	—	...
	p value	0.006*	—	—	...
Motivation	Statistic	112.368	94.680	84.078	...
	p value	0.290	0.740	0.432	...

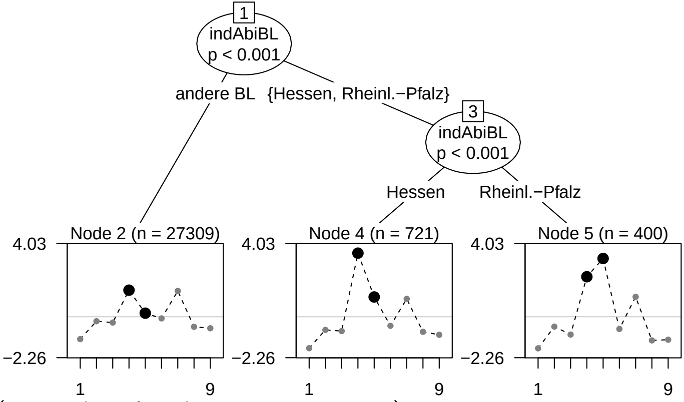
- minimal node-size < 20

Properties

as shown by Strobl et al. (2015):

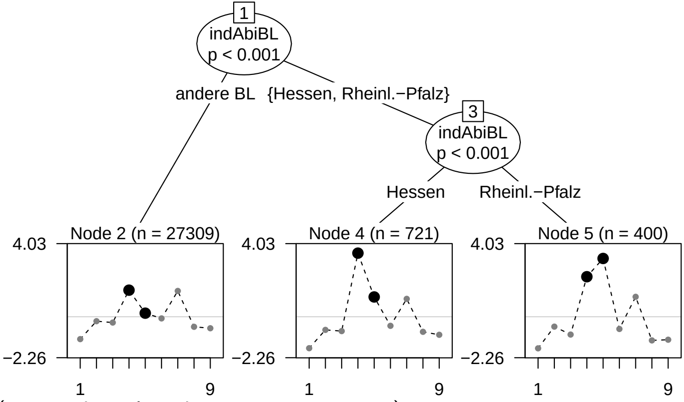
- no alpha inflation
  - correct distributions for optimally selected statistics, closed testing procedure, Bonferroni adjustment
  - even in the presence of ability differences due to CML
- higher power than the LR test to detect DIF
  - in numeric variables (where the true cutpoint is unknown/misspecified)
  - for other non-standard groups (formed, e.g., by interaction or u-shaped patterns)

Example: MOB for RM – Students' PISA



(note: this plot shows item easiness)

Example: MOB for RM – Students' PISA



(note: this plot shows item easiness)

Nr. 4: Where is Hessen? (indicate location on a map)

Nr. 5: What is the capital of Rheinland-Pfalz? (Mainz)

skip PCM

CART
Early statistical problems
MOB
for BT models
for Rasch models
for non-Rasch models
Can we trust the results?
Stability
Effect size stopping
Summary
References

MOB for psychometric models (package `psychotree`)

- **Bradley-Terry models**

  Strobl, Wickelmaier and Zeileis (2011, *Journal of Educational and Behavioral Statistics*); Eugster, Leisch and Strobl (2014, *Computational Statistics & Data Analysis*); Müller, Strobl et al. (2015, *Journal of Applied Ecology*)

- **binary Rasch models**

  Strobl, Kopf and Zeileis (2015, *Psychometrika*)

- **polytomous Rasch models**

  Komboz, Strobl, Zeileis (2016, *Educational and Psychological Measurement*)

- **2PL etc. models**

  Schneider, Strobl, Zeileis and Debelak (2021, Behavior Research Methods); Debelak and Strobl, (2019, Educational and Psychological Measurement)

Detecting parameter heterogeneity

CART

Early statistical problems

MOB

for BT models

for Rasch models

for non-Rasch models

Can we trust the results?

Stability

Effect size stopping

Summary

References

MOB for psychometric models (package psychotree)

- **Bradley-Terry models**

  Strobl, Wickelmaier and Zeileis (2011, *Journal of Educational and Behavioral Statistics*); Eugster, Leisch and Strobl (2014, *Computational Statistics & Data Analysis*); Müller, Strobl et al. (2015, *Journal of Applied Ecology*)

- **binary Rasch models**

  Strobl, Kopf and Zeileis (2015, *Psychometrika*)

- **polytomous Rasch models**

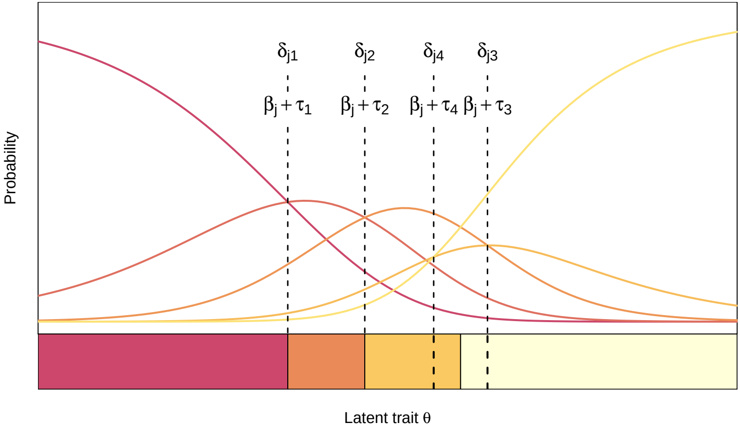
  Komboz, Strobl, Zeileis (2016, *Educational and Psychological Measurement*)

- **2PL etc. models**

  Schneider, Strobl, Zeileis and Debelak (2021, Behavior Research Methods); Debelak and Strobl, (2019, Educational and Psychological Measurement)

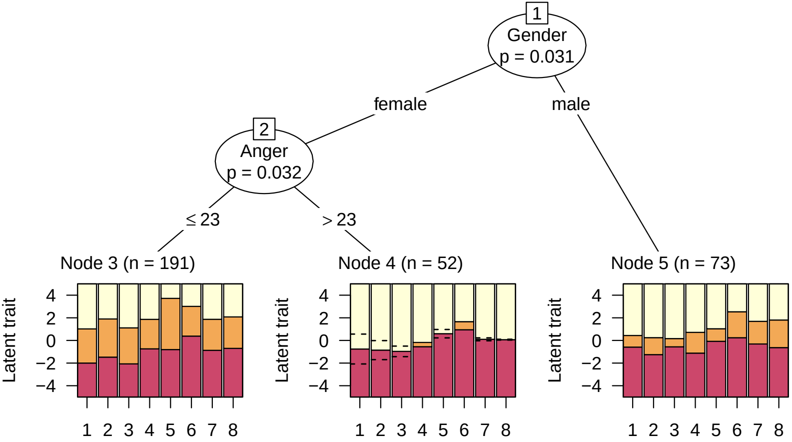
MOB for polytomous Rasch models

"Effect plots", inspired by Van der Linden and Hambleton (1997) and Fox and Hong (2009)



Example: MOB for PCM – Verbal Agression

- Komboz, Strobl and Zeileis (2017, Psychological Measurement)



Differential Step Functioning (DSF)

Detecting parameter heterogeneity

MOB

for BT models

for Rasch models

for non-Rasch models

Can we trust the results?

Stability

Effect size stopping

Summary

References

MOB for psychometric models (package `psychotree`)

- **Bradley-Terry models**

  Strobl, Wickelmaier and Zeileis (2011, *Journal of Educational and Behavioral Statistics*); Eugster, Leisch and Strobl (2014, *Computational Statistics & Data Analysis*); Müller, Strobl et al. (2015, *Journal of Applied Ecology*)

- **binary Rasch models**

  Strobl, Kopf and Zeileis (2015, *Psychometrika*)

- **polytomous Rasch models**

  Komboz, Strobl, Zeileis (2016, *Educational and Psychological Measurement*)

- **2PL etc. models**

  Schneider, Strobl, Zeileis and Debelak (2021, Behavior Research Methods); Debelak and Strobl, (2019, Educational and Psychological Measurement)

MOB for psychometric models (package `psychotree`)

- **Bradley-Terry models**

  Strobl, Wickelmaier and Zeileis (2011, *Journal of Educational and Behavioral Statistics*); Eugster, Leisch and Strobl (2014, *Computational Statistics & Data Analysis*); Müller, Strobl et al. (2015, *Journal of Applied Ecology*)

- **binary Rasch models**

  Strobl, Kopf and Zeileis (2015, *Psychometrika*)

- **polytomous Rasch models**

  Komboz, Strobl, Zeileis (2016, *Educational and Psychological Measurement*)

- **2PL etc. models**

  Schneider, Strobl, Zeileis and Debelak (2021, Behavior Research Methods); Debelak and Strobl, (2019, Educational and Psychological Measurement)

MOB for non-Rasch models

Schneider, Strobl, Zeileis and Debelak (2021, Beh. Res. Meth.)

Model	Response type	Estimation	Model reference	Score-based tests published in
Rasch Model	dichotomous	CML	Rasch (1960)	Strobl et al., (2015)
1PL	dichotomous	MML	Birnbaum (1968)	
2PL	dichotomous	MML	Birnbaum (1968)	Debelak and Strobl (2019)
3PL	dichotomous	MML	Birnbaum (1968)	Debelak and Strobl (2019)
3PLu	dichotomous	MML	Barton and Lord (1981)	
4PL	dichotomous	MML	Barton and Lord (1981)	
ideal point model	dichotomous	MML	Maydeu-Olivares et al., (2006)	
rating scale model	polytomous	CML	Andrich (1978)	Komboz et al., (2018)
partial credit model	polytomous	CML	Masters (1982)	Komboz et al., (2018)
generalized partial credit model	polytomous	MML	Muraki (1992)	
graded response model	polytomous	MML	Samejima (1969)	
nominal response model	polytomous	MML	Bock (1972)	
generalized graded unfolding model	polytomous	MML	Roberts et al., (2000)	
monotonic polynomial model	polytomous	MML	Falk and Cai (2016)	

MOB for non-Rasch models

Schneider, Strobl, Zeileis and Debelak (2021, Beh. Res. Meth.)


Model	Response type	Estimation	Model reference	Score-based tests published in
Rasch Model	dichotomous	CML	Rasch (1960)	Strobl et al., (2015)
1PL	dichotomous	MML	Birnbaum (1968)	
2PL	dichotomous	MML	Birnbaum (1968)	Debelak and Strobl (2019)
3PL	dichotomous	MML	Birnbaum (1968)	Debelak and Strobl (2019)
3PLu	dichotomous	MML	Barton and Lord (1981)	
4PL	dichotomous	MML	Barton and Lord (1981)	
ideal point model	dichotomous	MML	Maydeu-Olivares et al., (2006)	
rating scale model	polytomous	CML	Andrich (1978)	Komboz et al., (2018)
partial credit model	polytomous	CML	Masters (1982)	Komboz et al., (2018)
generalized partial credit model	polytomous	MML	Muraki (1992)	
graded response model	polytomous	MML	Samejima (1969)	
nominal response model	polytomous	MML	Bock (1972)	
generalized graded unfolding model	polytomous	MML	Roberts et al., (2000)	
monotonic polynomial model	polytomous	MML	Falk and Cai (2016)	

MOB for non-Rasch models

Schneider, Strobl, Zeileis and Debelak (2021, Beh. Res. Meth.)


Model	Response type	Estimation	Model reference	Score-based tests published in
Rasch Model	dichotomous	CML	Rasch (1960)	Strobl et al., (2015)
1PL	dichotomous	MML	Birnbaum (1968)	
2PL	dichotomous	MML	Birnbaum (1968)	Debelak and Strobl (2019)
3PL	dichotomous	MML	Birnbaum (1968)	Debelak and Strobl (2019)
3PLu	dichotomous	MML	Barton and Lord (1981)	
4PL	dichotomous	MML	Barton and Lord (1981)	
ideal point model	dichotomous	MML	Maydeu-Olivares et al., (2006)	
rating scale model	polytomous	CML	Andrich (1978)	Komboz et al., (2018)
partial credit model	polytomous	CML	Masters (1982)	Komboz et al., (2018)
generalized partial credit model	polytomous	MML	Muraki (1992)	
graded response model	polytomous	MML	Samejima (1969)	
nominal response model	polytomous	MML	Bock (1972)	
generalized graded unfolding model	polytomous	MML	Roberts et al., (2000)	
monotonic polynomial model	polytomous	MML	Falk and Cai (2016)	

MOB for non-Rasch models

Schneider, Strobl, Zeileis and Debelak (2021, Beh. Res. Meth.)

Model	Response type	Estimation	Model reference	Score-based tests published in
Rasch Model	dichotomous	CML	Rasch (1960)	Strobl et al., (2015)
1PL	dichotomous	MML	Birnbaum (1968)	
2PL	dichotomous	MML	Birnbaum (1968)	Debelak and Strobl (2019)
3PL	dichotomous	MML	Birnbaum (1968)	Debelak and Strobl (2019)
3PLu	dichotomous	MML	Barton and Lord (1981)	
4PL	dichotomous	MML	Barton and Lord (1981)	
ideal point model	dichotomous	MML	Maydeu-Olivares et al., (2006)	
rating scale model	polytomous	CML	Andrich (1978)	Komboz et al., (2018)
partial credit model	polytomous	CML	Masters (1982)	Komboz et al., (2018)
generalized partial credit model	polytomous	MML	Muraki (1992)	
graded response model	polytomous	MML	Samejima (1969)	
nominal response model	polytomous	MML	Bock (1972)	
generalized graded unfolding model	polytomous	MML	Roberts et al., (2000)	
monotonic polynomial model	polytomous	MML	Falk and Cai (2016)	

MML $\Rightarrow$ need to model true group differences in means of person parameter distributions (`impact`)

Can we trust the results?

MOB for psychometric models (package `psychotree`)

- **Bradley-Terry models**

  Strobl, Wickelmaier and Zeileis (2011, *Journal of Educational and Behavioral Statistics*); Eugster, Leisch and Strobl (2014, *Computational Statistics & Data Analysis*); Müller, Strobl et al. (2015, *Journal of Applied Ecology*)

- **binary Rasch models**

  Strobl, Kopf and Zeileis (2015, *Psychometrika*)

- **polytomous Rasch models**

  Komboz, Strobl, Zeileis (2016, *Educational and Psychological Measurement*)

- **2PL etc. models**

  Schneider, Strobl, Zeileis and Debelak (2021, Behavior Research Methods); Debelak and Strobl, (2019, Educational and Psychological Measurement)

MOB for psychometric models (package psychotree)

- **Bradley-Terry models**

  Strobl, Wickelmaier and Zeileis (2011, *Journal of Educational and Behavioral Statistics*); Eugster, Leisch and Strobl (2014, *Computational Statistics & Data Analysis*); Müller, Strobl et al. (2015, *Journal of Applied Ecology*)

- **binary Rasch models**

  Strobl, Kopf and Zeileis (2015, *Psychometrika*)

- **polytomous Rasch models**

  Komboz, Strobl, Zeileis (2016, *Educational and Psychological Measurement*)

- **2PL etc. models**

  Schneider, Strobl, Zeileis and Debelak (2021, Behavior Research Methods); Debelak and Strobl, (2019, Educational and Psychological Measurement)

Detecting parameter heterogeneity

CART

Early statistical problems

MOB

for BT models

for Rasch models

for non-Rasch models

Can we trust the results?

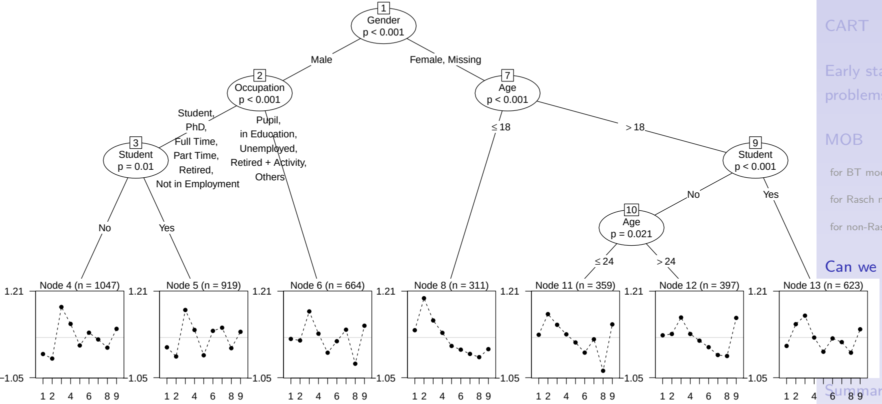
Stability

Effect size stopping

Summary

References

Example: MOB for RM – Students' PISA 2



now: Natural Science items, tree for 5000 students

CART
Early statistical problems
MOB
for BT models
for Rasch models
for non-Rasch models
Can we trust the stopping
Summary
References

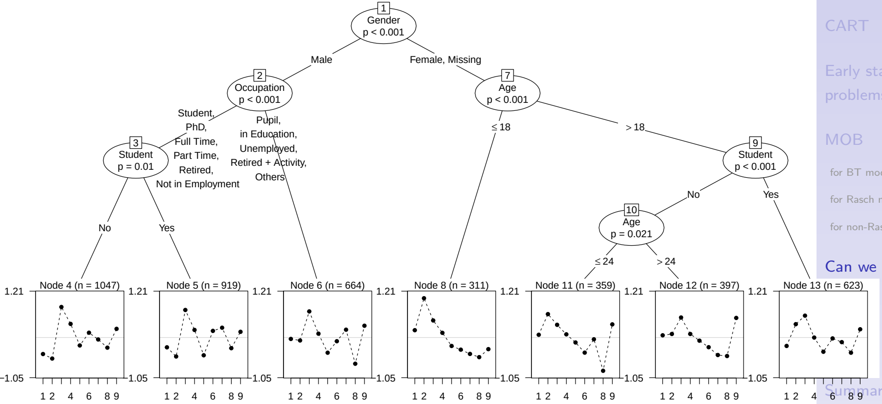
Can we trust the results?

1. stability:

   if we drew another random sample from the data,

   would the tree still look the same?

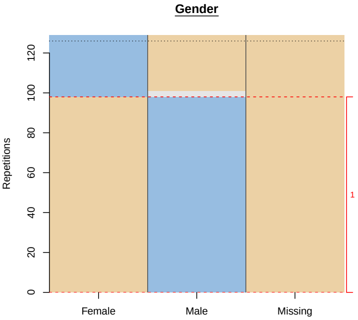
Example: MOB for RM – Students' PISA 2

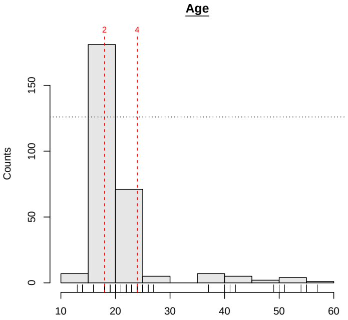


now: Natural Science items, tree for 5000 students

Example: MOB for RM – Students' PISA 2

Which cutpoints were selected in trees for 125 re-samples?



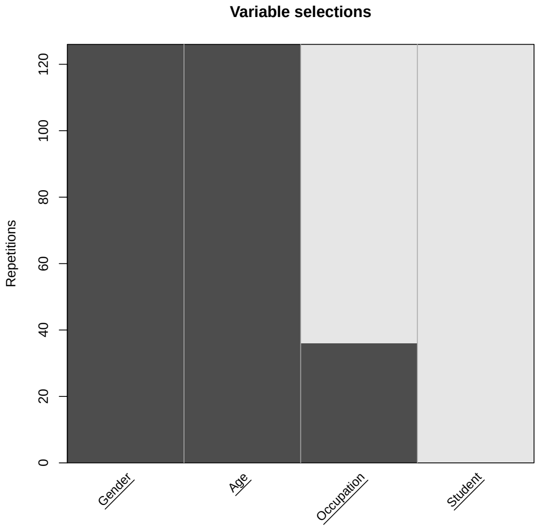


(variable Age appears more than once in many trees ⇒ more than 125 times in total)

Philipp, Zeileis, Strobl (2016, COMPSTAT Proceedings)

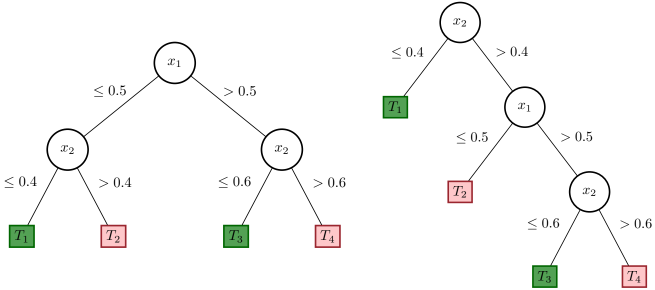
Example: MOB for RM – Students' PISA 2

Which variables were selected in trees for 125 re-samples?



Example: MOB for RM – Students' PISA 2

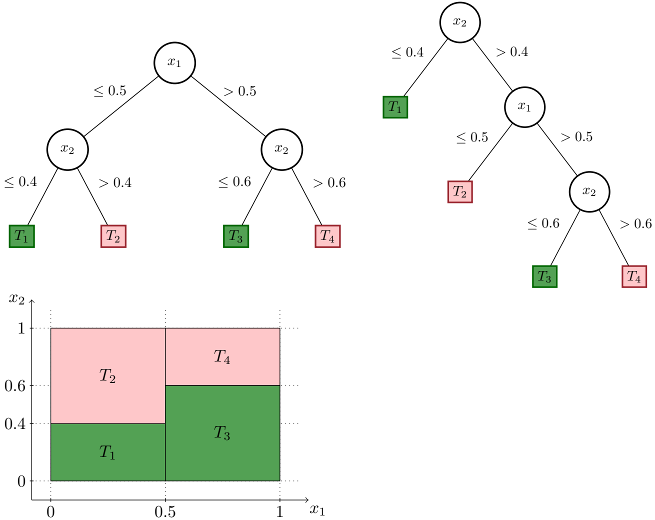
Structural vs. semantic similarity of tree results



Philipp, Rusch, Hornik and Strobl (2018, JCGS)

Example: MOB for RM – Students' PISA 2

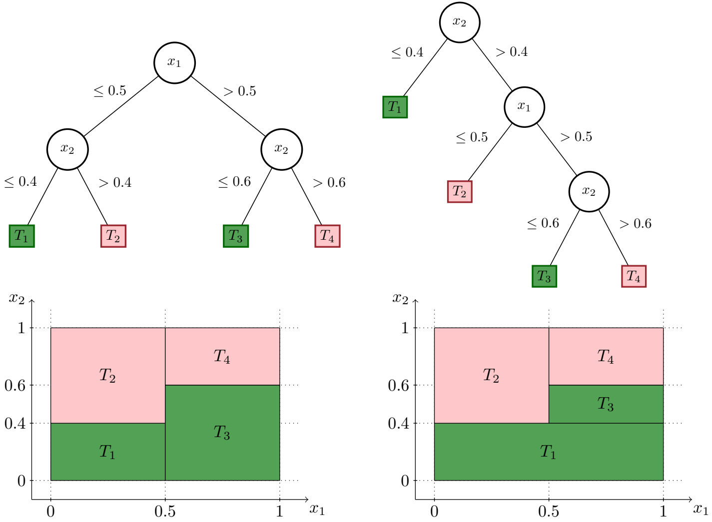
Structural vs. semantic similarity of tree results



Philipp, Rusch, Hornik and Strobl (2018, JCGS)

Example: MOB for RM – Students' PISA 2

Structural vs. semantic similarity of tree results



Philipp, Rusch, Hornik and Strobl (2018, JCGS)

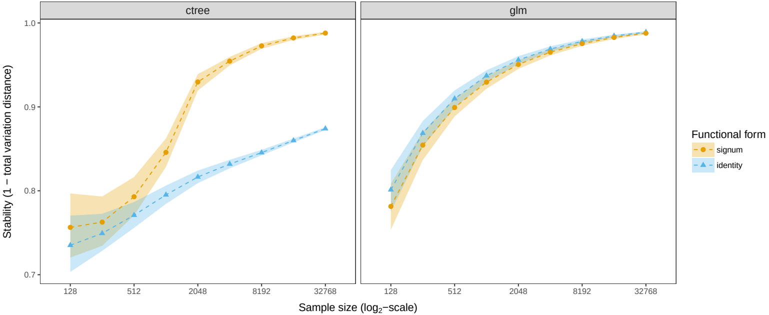
Example: MOB for RM – Students' PISA 2

Are trees always instable?

Example: MOB for RM – Students' PISA 2

CART
Early statistical problems
MOB
for BT models
for Rasch models
for non-Rasch models
Can we trust the results?
Stability
Effect size stopping
Summary
References

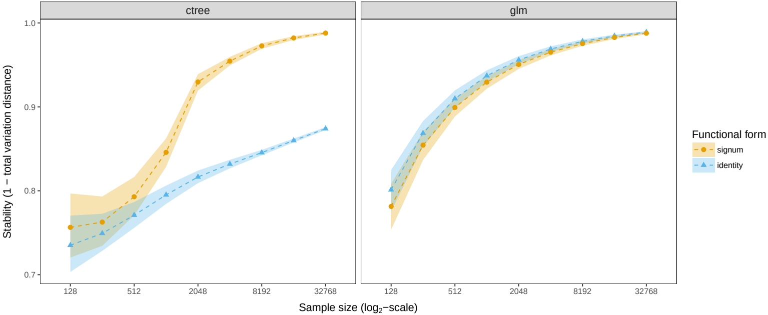
Are trees always instable?



given enough observations, tree shows high stability for step-function (orange results), but not for smooth function (blue results)

Example: MOB for RM – Students' PISA 2

Are trees always instable?



given enough observations, tree shows high stability for step-function (orange results), but not for smooth function (blue results)

$$\text{stability} = f(\text{learner} \times \text{dgp})$$

CART
Early statistical problems
MOB
for BT models
for Rasch models
for non-Rasch models
Can we trust the results?
Stability
Effect size stopping
Summary
References

Can we trust the results?

1. stability:

   if we drew another random sample from the data,

   would the tree still look the same?

Detecting parameter heterogeneity
CART
Early statistical problems
MOB
for BT models
for Rasch models
for non-Rasch models
Can we trust the results?
Stability
Effect size stopping
Summary
References

Can we trust the results?

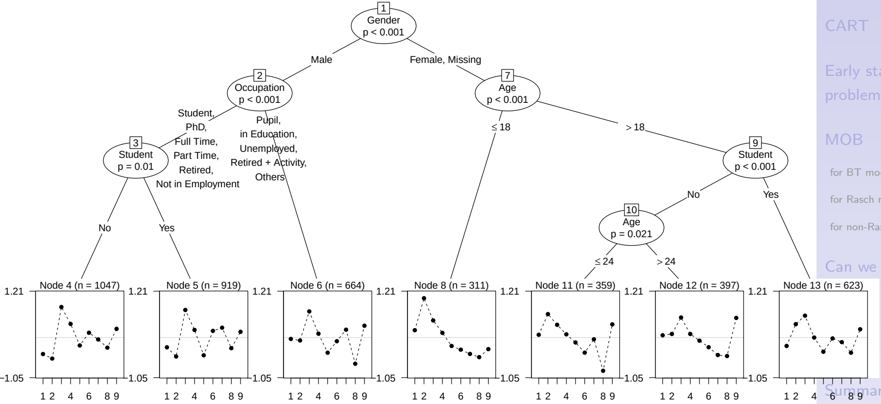
1. stability:

   if we drew another random sample from the data, would the tree still look the same?

2. power vs. effect size:

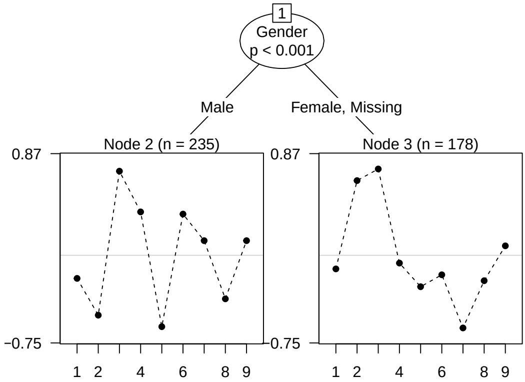
   the group differences are significant – but are they practically relevant?

Example: MOB for RM – Students' PISA 2



Natural Science items, tree for 5000 students

Example: MOB for RM – Students' PISA 2



Natural Science items, tree for 500 students

Example: MOB for RM – Students' PISA 2

Henninger, Debelak and Strobl (2022, Edu. and Psy. Measurement)

additional stopping criterion: Mantel-Haenszel odds ratio = effect size measure for DIF

Educational Testing Service (ETS) classification criteria:

A = negligible DIF

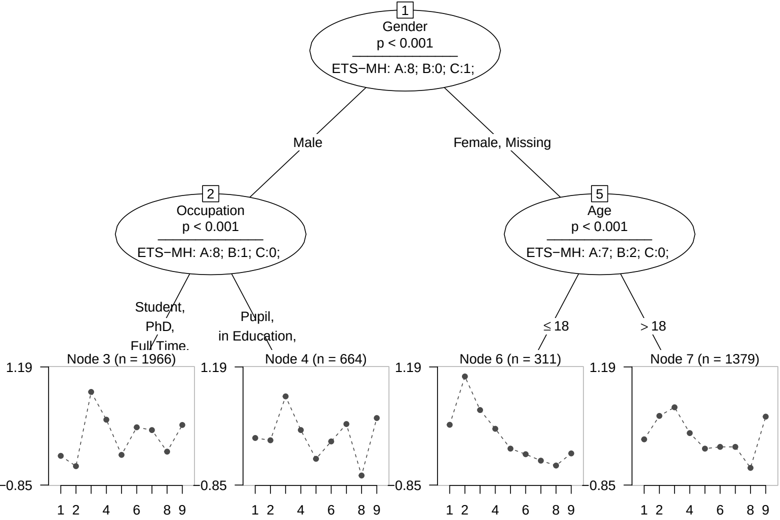
B = medium DIF

C = large DIF

MOB
for BT models
for Rasch models
for non-Rasch models
Can we trust the results?
Stability
Effect size stopping
Summary
References

Example: MOB for RM – Students' PISA 2

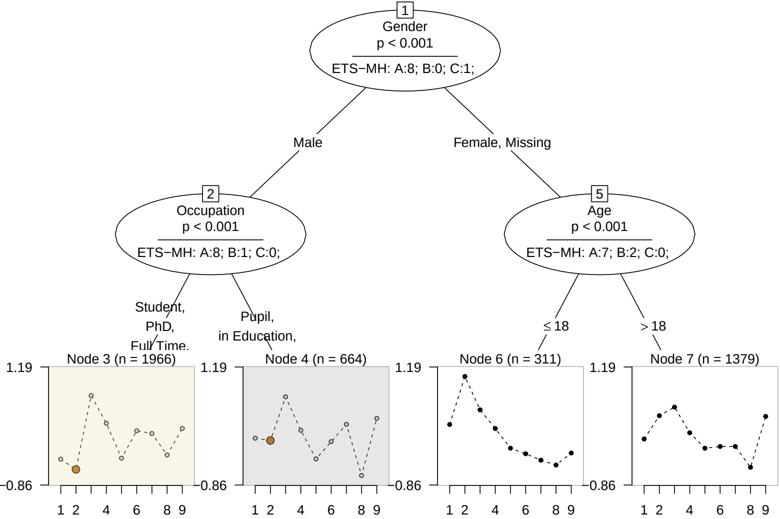
tree stopped with Mantel-Haenszel criterion



summary

Example: MOB for RM – Students' PISA 2

tree stopped with Mantel-Haenszel criterion



summary

Summary

- What I have talked about: (package: `psychotree`) model-based recursive partitioning can
  - detect groups of observations with different parameters
  - need not be specified a priori (data-driven, exploratory)

Summary

- What I have talked about: (package: `psychotree`) model-based recursive partitioning can
  - detect groups of observations with different parameters
  - need not be specified a priori (data-driven, exploratory)
  - can be used to detect DIF and DSF

Summary

- What I have talked about: (package: `psychotree`) model-based recursive partitioning can
  - detect groups of observations with different parameters
  - need not be specified a priori (data-driven, exploratory)
  - can be used to detect DIF and DSF
  - can we trust the results?
    - stability assessment (package: `stablelearner`)
    - effect-size-based stopping (currently on `github.com/mirka-henninger`)

Summary

- What I have talked about: (package: `psychotree`) model-based recursive partitioning can
  - detect groups of observations with different parameters
  - need not be specified a priori (data-driven, exploratory)
  - can be used to detect DIF and DSF
  - can we trust the results?
    - stability assessment (package: `stablelearner`)
    - effect-size-based stopping (currently on `github.com/mirka-henninger`)
- What I haven't talked about:
  - anchoring: making parameters comparable between groups (= end nodes, possibly many) in the presence of DIF and DSF

movie MH anchoring

Key References

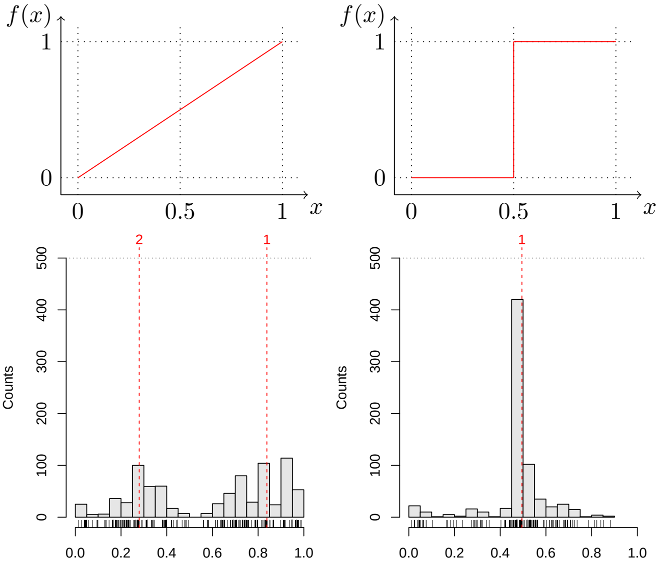
Henninger M, Debelak R, Strobl C (2023). A New Stopping Criterion for Rasch Trees Based on the Mantel–Haenszel Effect Size Measure for Differential Item Functioning. *Educational and Psychological Measurement*, **83**(1), 181–212.

Strobl C, Kopf J, Zeileis A (2015). Rasch Trees: A New Method for Detecting Differential Item Functioning in the Rasch Model. *Psychometrika*, **80**(2), 289–316.

Strobl C, Malley J, Tutz G (2009). An Introduction to Recursive Partitioning: Rationale, Application and Characteristics of Classification and Regression Trees, Bagging and Random Forests. *Psychological Methods*, **14**(4), 323–348.

Philipp M, Rusch T, Hornik K, Strobl C (2018). Measuring the Stability of Results From Supervised Statistical Learning. *Journal of Computational and Graphical Statistics*, **27**(4), 685–700.

Stability of cutpoints

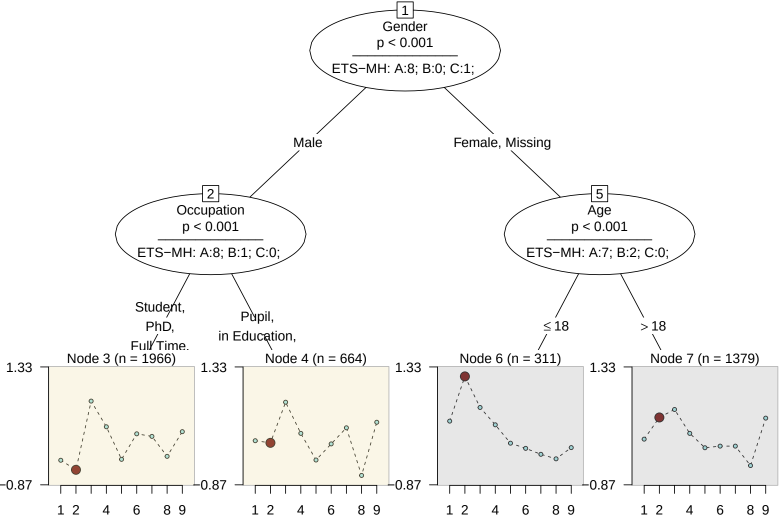


Example: MOB for RM – Students' PISA 2

Which cutpoints were selected in trees for 125 re-samples?

Example: MOB for RM – Students' PISA 2

tree stopped with Mantel-Haenszel criterion



CART
Early statistical problems
MOB
for BT models
for Rasch models
for non-Rasch models
Can we trust the results?
Stability
Effect size stopping
Summary
References