To Split or to Mix?
Uncovering Group Structures with Trees and Finite Mixture Models

Hannah Frick, Carolin Strobl, Achim Zeileis

http://eeecon.uibk.ac.at/~frick/

Outline

- Introduction
- Tools to uncover group structures
  - Model-based recursive partitioning
  - Mixture models
- Simulation study
  - Design
  - Results
- Empirical example

Introduction

- Basic assumption: One model with one set of parameters fits for all observations.
- However, subgroups of observations might differ in, e.g., their relationship between explanatory variables and response, preference scales, or item difficulties.
- In the context of psychometric models: Violation of measurement invariance.
- Various tests available to check for such violations. However, they usually require an a priori division of the subjects in reference and focal group.
- Two options to establish such subgroups in a data-driven way: model-based recursive partitioning and mixture models.
- Question here: How do they compare?

General Questions

Is one model / set of parameters enough? Or are there clusters?

- If so, how many?
- How do they differ?
- Who belongs to which cluster?

If there are covariates available:

- Are they connected to the clustering?
- How do they separate the clusters? Abrupt shifts or smooth transitions?
- How well do they separate the clusters?
- Multiple covariates: In which pattern are they connected to the clustering?

Model-Based Recursive Partitioning

Underlying idea:

- Fit model of interest.
- Test parameter stability along the covariates z.
- Split sample along greatest instability.
- Repeat on resulting subsamples until no further instability is found.
- Yields tree structure.

Thus:

- Sample splits lead to a hard clustering, no smooth transitions.
- Tree structure represents the connection to the covariates.
- Comparison of clusters via models in each terminal node.

Mixture Models

- Basic structure: weighted sum over several components which represent the (latent) classes

$$f(y_i|\alpha, \xi) = \sum_{k=1}^{K} \pi(k|\alpha, z_i) f(y_i|\xi^{(k)}) \quad \text{with}$$

  y_i denoting the response of observation i, and
  $f(y_i|\xi^{(k)})$ the model of interest in class k with parameters $\xi^{(k)}$.
- Weights can depend on covariates z through a concomitant variable model $\pi(k|\alpha, z_i)$ with parameters α, e.g., a multinomial logit model.
- If no covariates are available, non-parametric estimates $\pi^{(k)}$ can be used as weights.

Mixture Models

- Estimation via EM algorithm: Alternate estimating components and class membership.
- Number of components is not a model parameter but needs to be established separately. Likelihood ratio test is not applicable due to violation of the regularity conditions. Information criteria are often used instead, here the BIC.

Differences to trees:

- Number of groups established via heuristic rather than significance test.
- Smooth transition between classes rather than hard sample splits.
- Can be used with and without covariates.

Bradley-Terry Model: Estimating Preferences

- Aim: Estimate a preference scale.
- Binary decisions are easier than rankings, thus data in form of paired comparisons.
- Bradley-Terry(-Luce) model based on *worth* parameters representing the popularity or ability of each object.
- Probability of choosing object a over object b:

$$P(a > b) = \frac{\text{worth of object a}}{\text{worth of object a} + \text{worth of object b}}$$

Simulation Study

Goals:

- “Test”: Detection of parameter instability / groups of observations.
- Clustering of individuals.
- Parameter estimation of the Bradley-Terry models.

Influential factors:

- How different are the groups with regard to the model of interest? Here: How strongly do the groups differ in their preferences?
- How strongly do the covariates differentiate between the groups?
- In which pattern do they differentiate?

Simulation Study

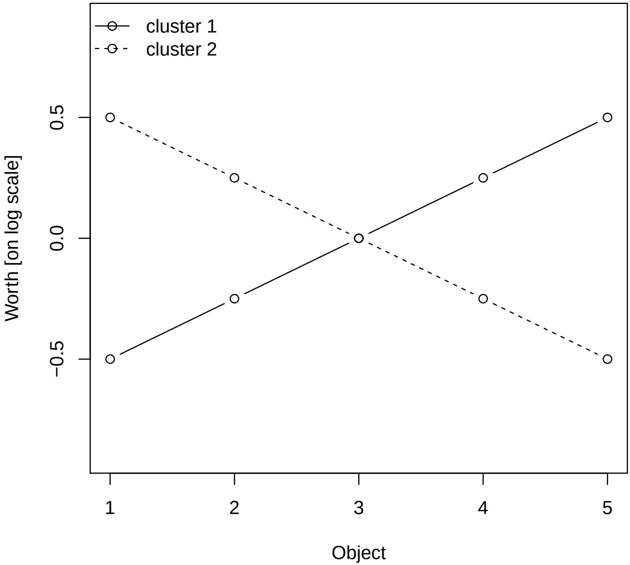
Data generating process:

- Paired comparisons on 5 objects by 100 subjects.
- Two equally-sized clusters.
- Preferences differ between clusters up to γ.
- Two covariates, `z1` and `z2`, logistically linked to the true clustering via parameter θ.
- Patterns in covariates: `axis1`, `diagonal`, `corner`, `xor`.

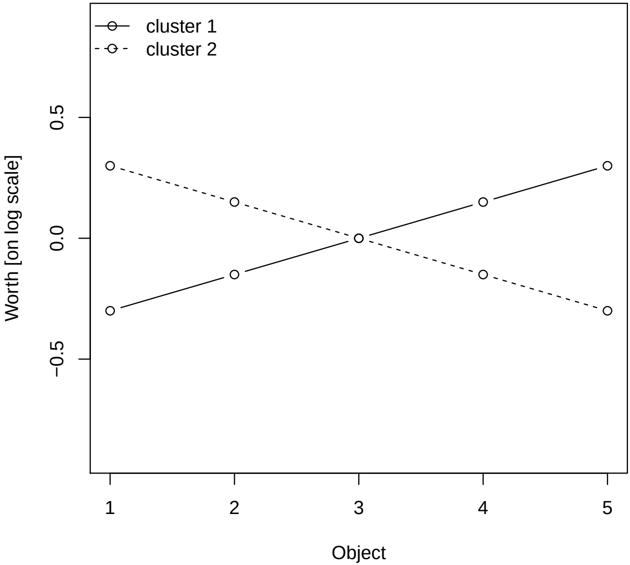
Models:

- BT tree.
- Plain mixture of BT models (without concomitant variables).
- Mixture of BT models with concomitant variables (as main effects).

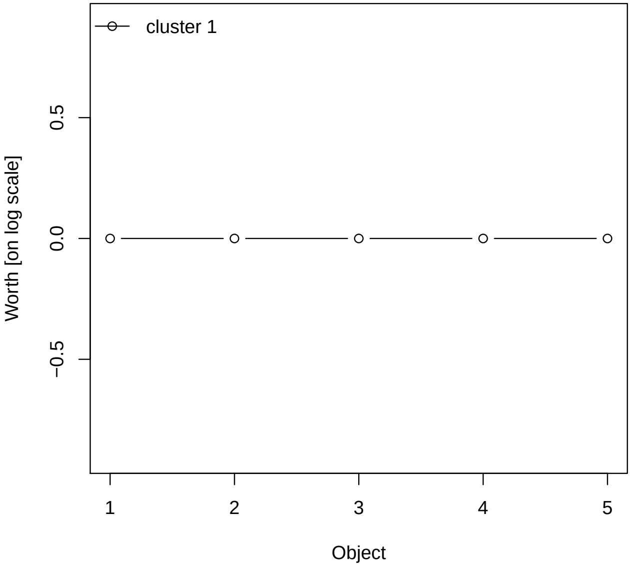
Preferences with Maximum Difference $\gamma = 1$



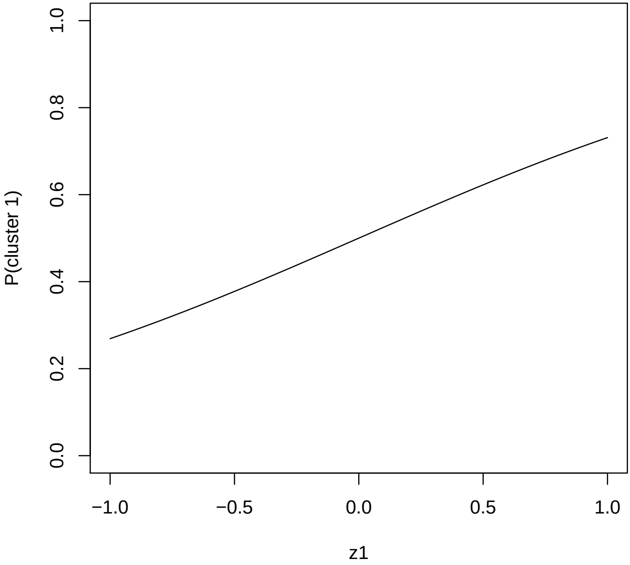
Preferences with Maximum Difference $\gamma = 0.6$



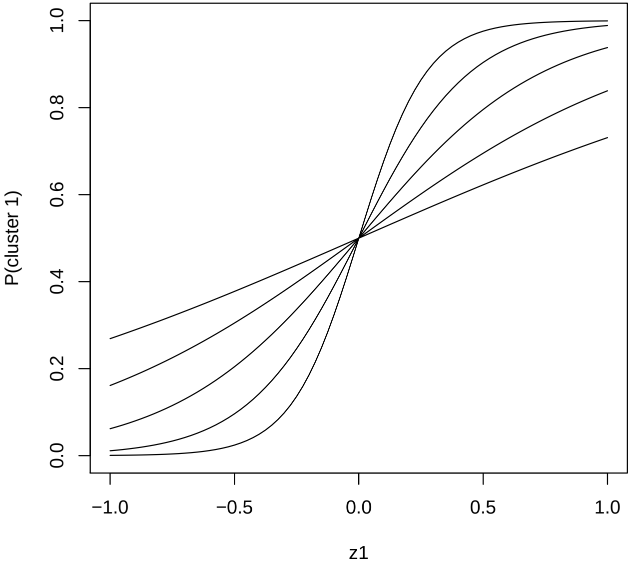
Preferences without Differences $(\gamma = 0)$



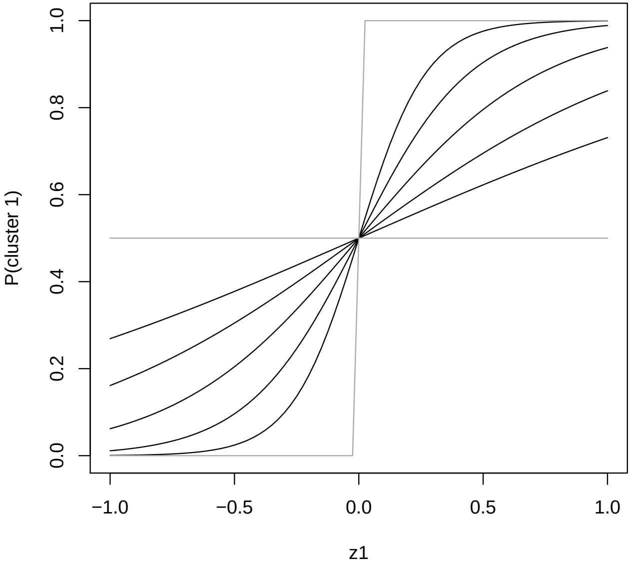
Link between Covariate and Clusters



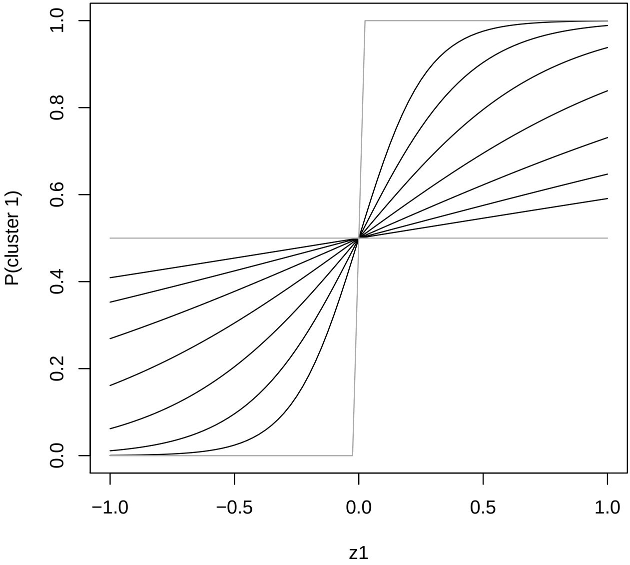
Link between Covariate and Clusters



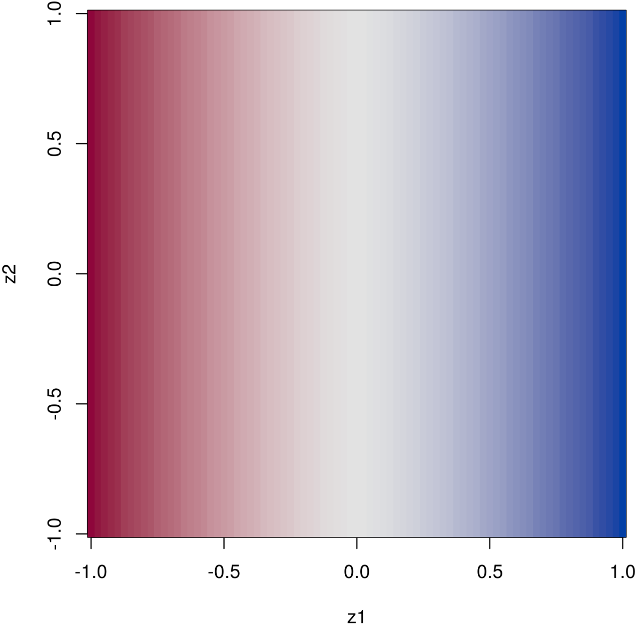
Link between Covariate and Clusters



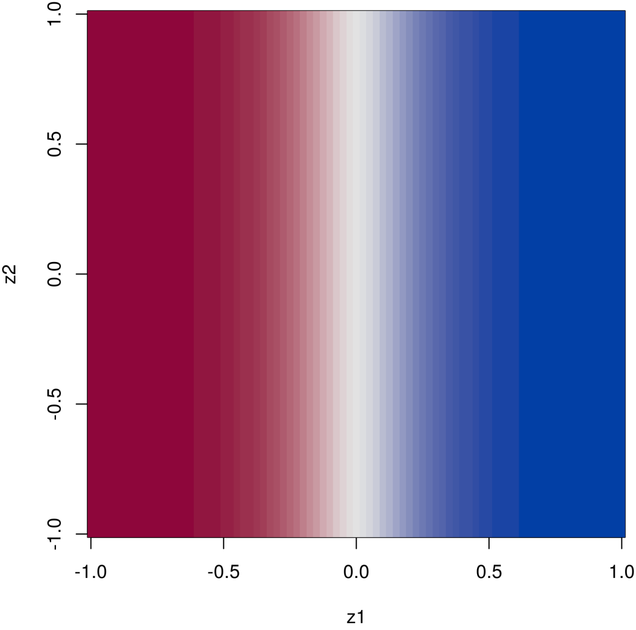
Link between Covariate and Clusters



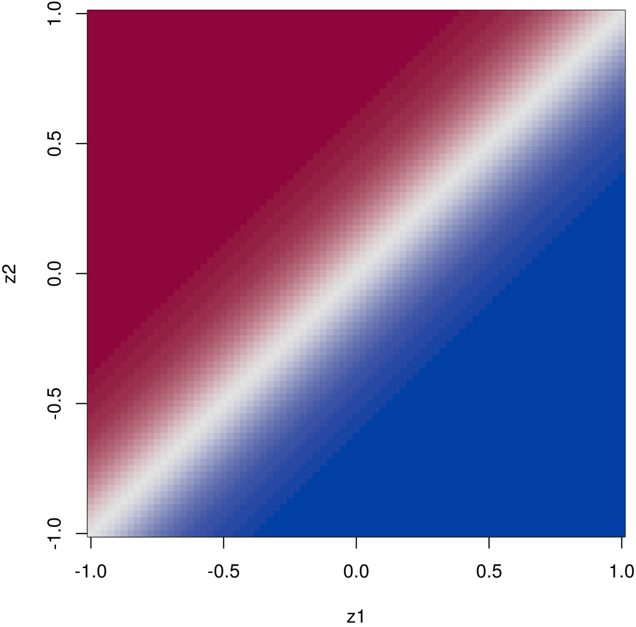
Pattern in Covariates: One axis only $(\theta = -1)$



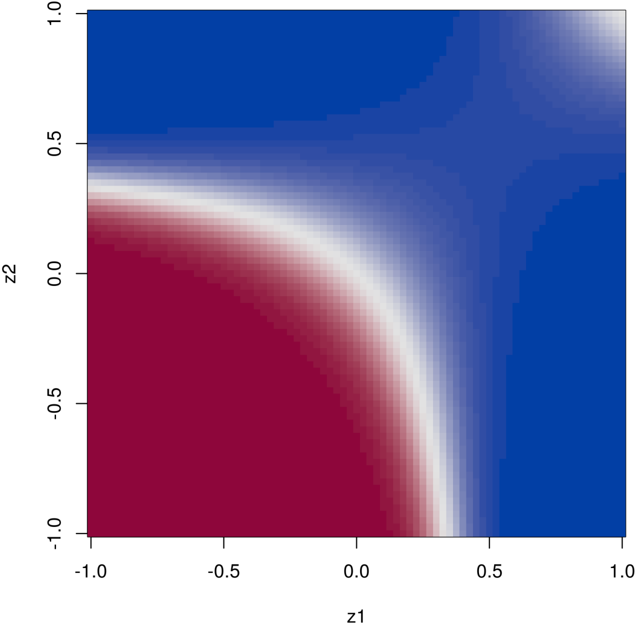
Pattern in Covariates: One axis only $(\theta = 2)$



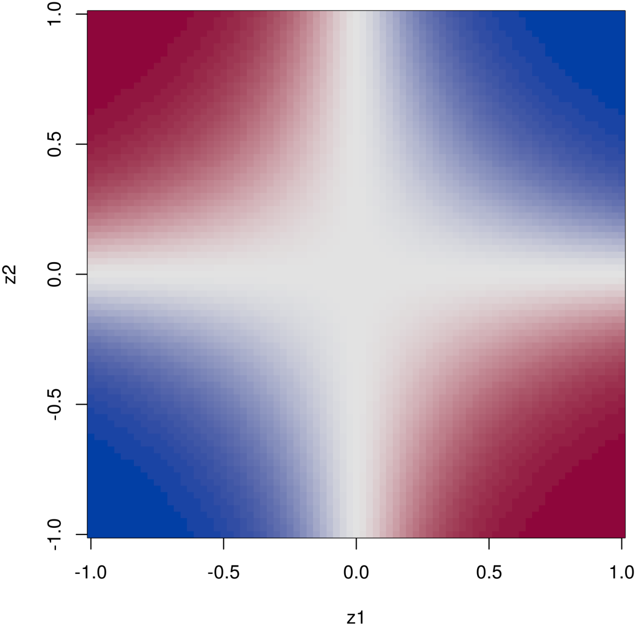
Pattern in Covariates: Diagonal $(\theta = 2)$



Pattern in Covariates: Corner $(\theta = 2)$



Pattern in Covariates: XOR $(\theta = 2)$



Simulation Study

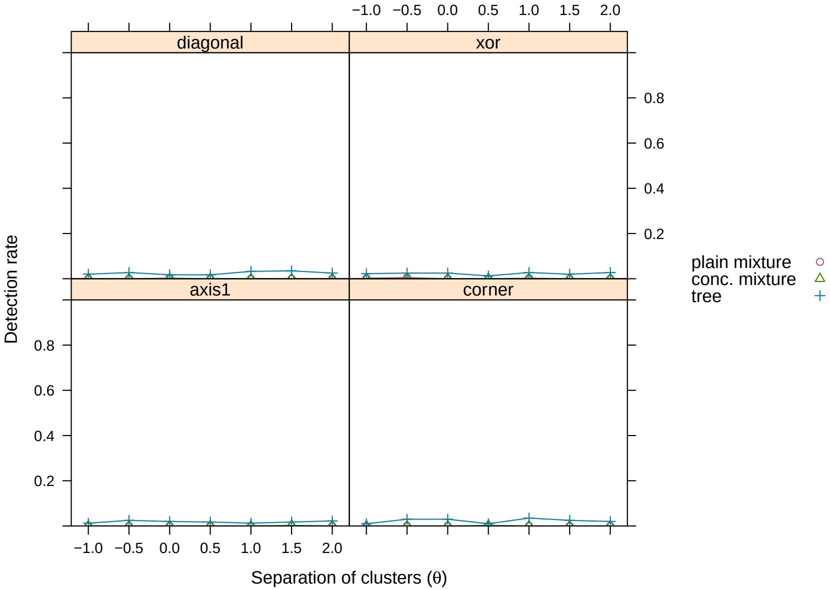
Outcome measures based on 400 replications:

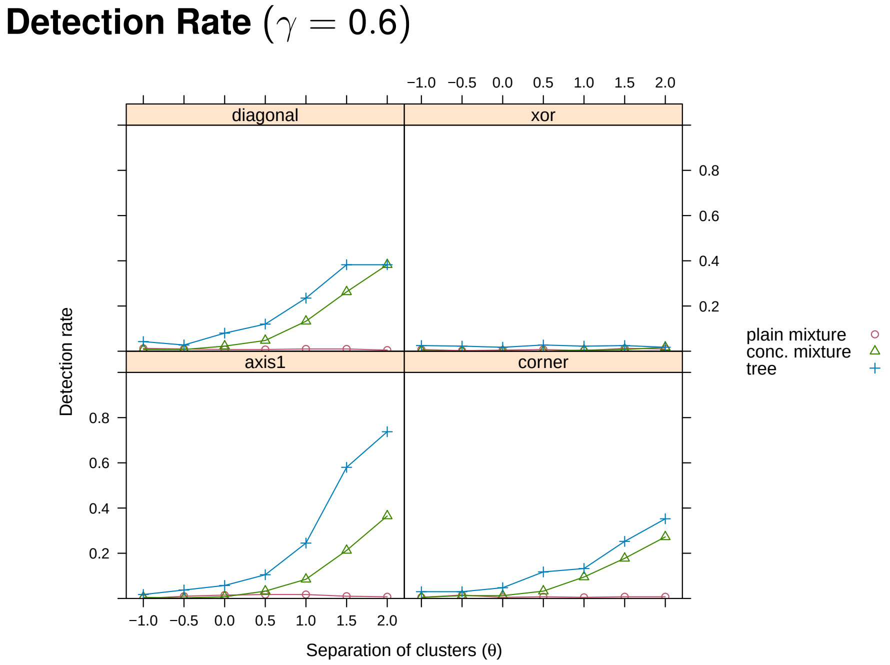
- Detection rate, comparable to power of a test.
- Recovery of clustering: Cramér's coefficient.
- Parameter recovery: mean squared error.

Software:

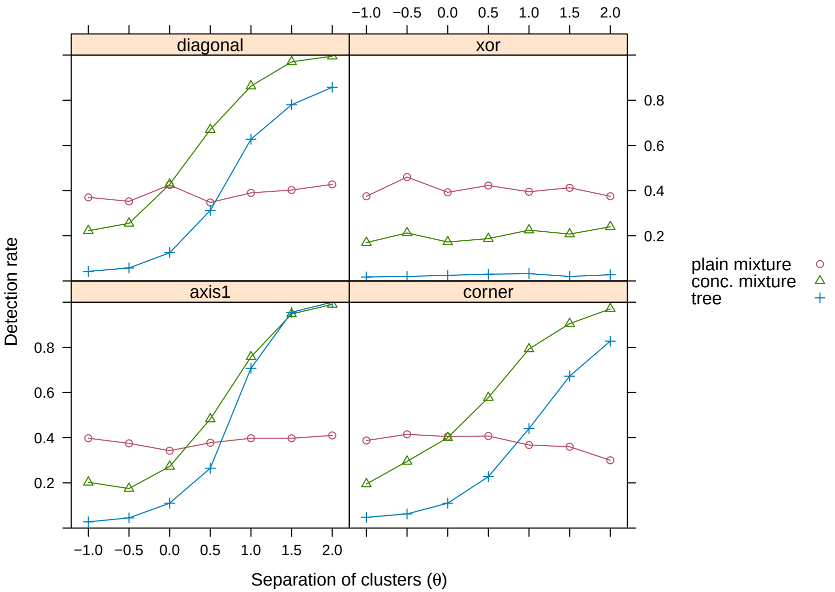
- R package **psychomix**, available at
  `http://CRAN.R-project.org/package=psychomix`
- R package **psychotree**, available at
  `http://CRAN.R-project.org/package=psychotree`

Detection Rate $(\gamma = 0)$

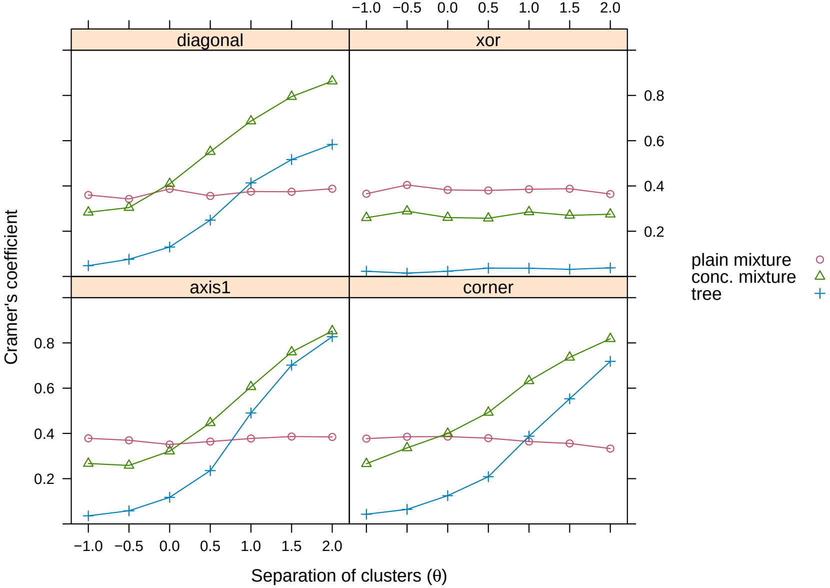



Detection Rate $(\gamma = 0.6)$
diagonal
xor
axis1
corner
Detection rate
Separation of clusters (θ)
plain mixture
conc. mixture
tree
−1.0 −0.5 0.0 0.5 1.0 1.5 2.0
0.2 0.4 0.6 0.8

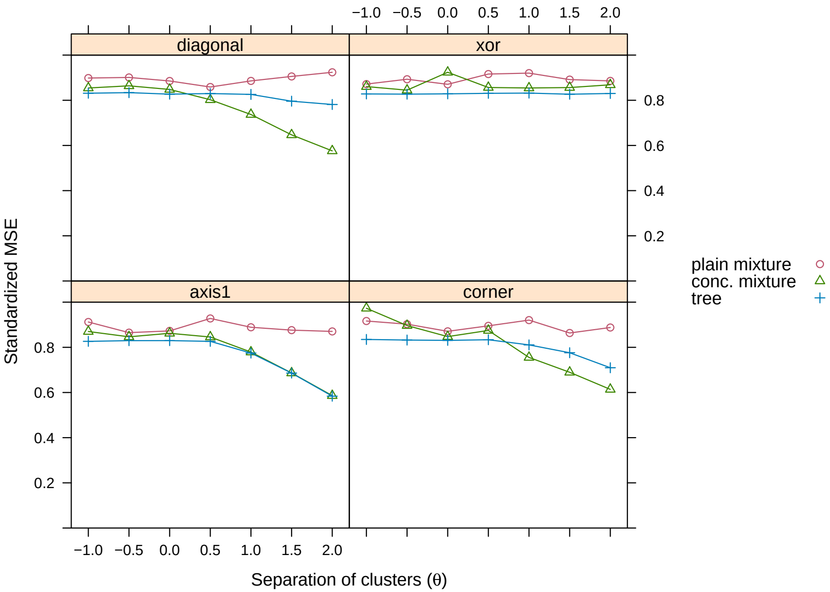
Detection Rate $(\gamma = 1)$



Cramér's Coefficient $(\gamma = 1)$



Standardized MSE $(\gamma = 1)$



Empirical Example

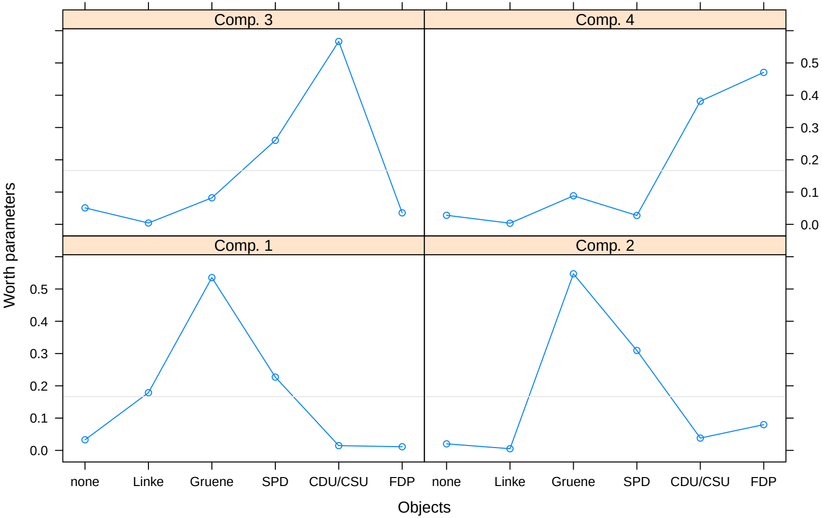
German Parties 2009

Three months before the German general elections of 2009, 192 respondents were asked about their political preferences. They were presented with pairs of five different parties: Die Linke (leftist party), Die Grünen (green party), SPD (social democrats), CDU/CSU (conservatives), and FDP (liberals). Also included was the option to refrain from voting.

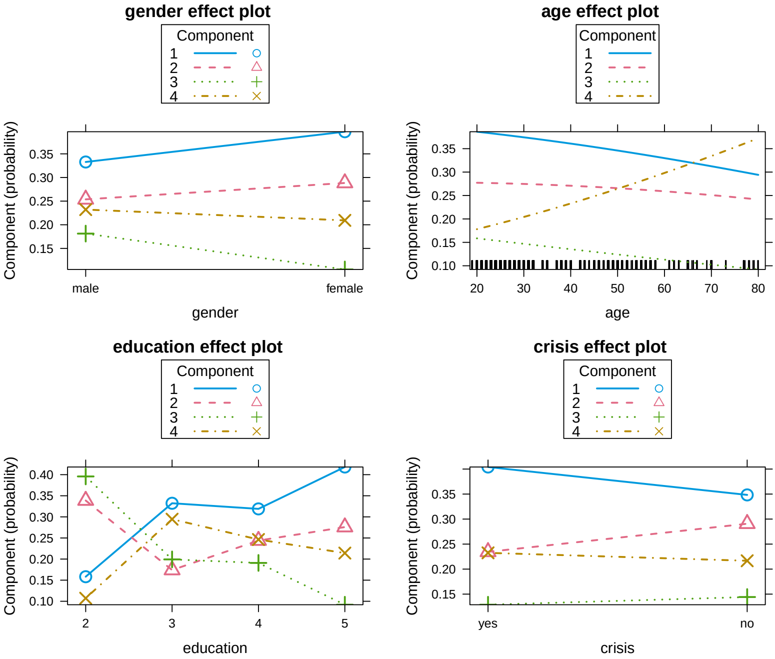
In addition, respondents were asked about the covariates gender, age, education, and whether they felt affected by the economic crisis.

Interviews were conducted by psychology students from Tübingen, mainly asking people they knew. Therefore, this it not a representative sample.

German Parties 2009



German Parties 2009



Summary

Detection of parameter instability:

- Tree:
  Detection depends on covariates via both degree of separation and type of pattern.
  Trees can detect smaller differences in model parameters than mixtures if covariates structure can be approximated through sample splits sufficiently well.
- Plain mixture:
  Per construction only sensitive to differences in model parameters. Thus, detection of clusters works equally well (or bad) across patterns.
- Mixture with concomitants:
  Capable of dealing with all covariate patters. Additional information in covariates improves detection rates in a fashion similar to trees.

Summary

Clustering of individuals:

- Given similar detection rates, all methods do equally well in separating individuals.

Parameter estimation:

- All methods recover the preferences scales similarly well, given similar detection rates.

Suggestion:
Having both methods in your toolbox enables you to uncover potential parameter instability in various settings, e.g., even when differences between groups are small, covariates are not available or only loosely connected to the group structure.

References

Zeileis A, Hothorn T, Hornik K (2008). "Model-Based Recursive Partitioning." *Journal of Computational and Graphical Statistics*, 17(2), 492–514.

McLachlan G, Peel D (2000). *Finite Mixture Models*. John Wiley & Sons, New York.

Bradley RA, Terry ME (1952). "Rank Analysis of Incomplete Block Designs. I. The Method of Paired Comparisons." *Biometrika*, 39(3/4), 324–345.

Strobl C, Wickelmaier F, Zeileis A (2011). "Accounting for Individual Differences in Bradley-Terry Models by Means of Recursive Partitioning." *Journal of Educational and Behavioral Statistics*, 36(2), 135–153.

Frick H, Strobl C, Leisch F, Zeileis A (2012). "Flexible Rasch Mixture Models with Package psychomix." *Journal of Statistical Software*, 48(7), 1–25. `http://www.jstatsoft.org/v48/i07/`