Universität Innsbruck

To Split or to Mix? Tree vs. Mixture Models for Detecting Subgroups

Hannah Frick, Carolin Strobl, Achim Zeileis

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

Outline

- Introduction
- Tools to detect subgroups
  - Mixture models
  - Model-based recursive partitioning
  - Unifying framework
- Simulation study
  - Design
  - Results
- Summary

Introduction

- Basic assumption: One model with one set of parameters fits for all observations.
- However, subgroups might exist for which different sets of parameters hold, e.g., the relationship between some response and regressors might be different for younger and older individuals.
- If breakpoint between “younger” and “older” were known, one can directly compare parameter estimates for both groups.
- If the breakpoint is unknown or there is a smooth transition between “young” and “old”, the subgroups can be still be detected in a data-driven way.
- Mixture model with age as a concomitant variable.
- Model-based recursive partitioning with splits in age.
- Question here: How do those two approaches compare for detecting parameter instability in a linear model?

Mixture Models

- Assumption: Data stem from K subgroups with different regression parameters $\beta_{(k)}$ and error variances $\sigma^2_{(k)}$ ($k = 1, \dots, K$).

$$f(y_i; x_i, z_i, \beta_{(1)}, \sigma_{(1)}, \dots, \beta_{(K)}, \sigma_{(K)}) = \sum_{k=1}^{K} \pi_k(z_i) \cdot \phi(y_i; x_i^\top \beta_{(k)}, \sigma^2_{(k)}).$$

- The component weights may depend on additional covariates z_i through a concomitant variable model, typically a multinomial logit model

$$\pi_k(z_i) = \frac{\exp(z_i^\top \alpha_{(k)})}{\sum_{g=1}^{K} \exp(z_i^\top \alpha_{(g)})}.$$

- For a given K, the EM algorithm is used for ML estimation. Typically, the model is fitted for $K = 1, 2, \dots$ and the best-fitting model, and thus $\hat{K}$, is selected via an information criterion, here the BIC.

Model-Based Recursive Partitioning

Algorithm:

1. Estimate the model parameters in the current subgroup.
2. Test parameter stability along each partitioning variable z_{ij}.
3. If any instability is found, split the sample along the variable z_{ij*} with the highest instability. Choose the breakpoint with the highest improvement in model fit.
4. Repeat 2–4 on the resulting subsamples until no further instability is found.

Unifying Framework

- Since each split can be expressed through an indicator function $I(\cdot)$ (for going left or right), each branch of the tree can be represented as a product of such indicator functions.
- Thus, model-based tree can also be written as a weighted sum over component models, albeit with rather different weights:

$$\pi_k(z_i) = \prod_{j=1}^{J_k} I(s_{(j|k)} \cdot z_{i(j|k)} > b_{(j|k)})$$

where
$z_{(j|k)}$ denotes the j-th partitioning variable for terminal node k,
$b_{(j|k)}$ is the associated breakpoint,
$s_{(j|k)} \in \{-1, 1\}$ the sign (signaling splitting to the left or right), and
J_k the number of splits leading up to node k.

Comparison

- Selection of $\hat{K}$: Based on information criterion for mixture model and on significance tests for trees.
- Covariates: Optional for mixtures which can thus also detect latent classes but required for trees which can thus only detect manifest classes.
- Multinomial logit model for mixtures models a smooth transition while the sample splits of trees model abrupt shifts. Multiple splits can represent a non-monotonic transition. Variable selection is included in trees but requires an additional step for mixture models.
- Clustering: Trees yield a hard clustering, mixture models a probabilistic clustering.

Simulation: Questions

Two basic questions:

1. Is any instability found at all?
2. If so, are the correct subgroups recovered?

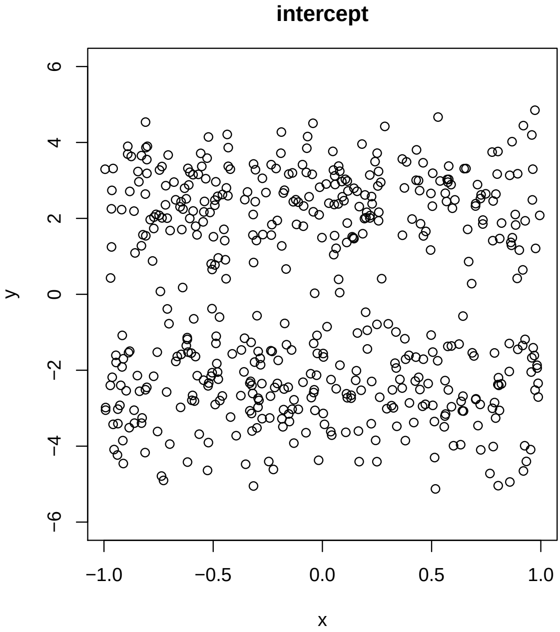
Potentially influential factors:

- *How* does the relationship between the response y and the regressors x differ between the subgroups and *how strongly* does it differ?
- If there are any additional covariates z available, *how* and *how strongly* are those covariates connected to the subgroups?

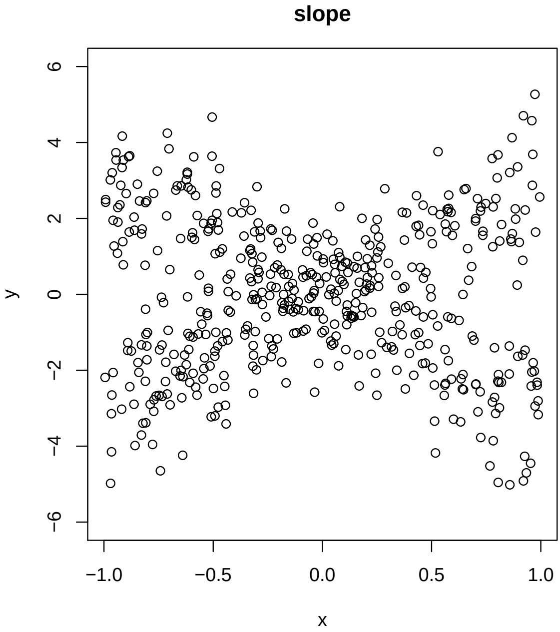
Expectations

- Trees are able to detect smaller differences in $\beta_{(k)}$ than mixtures, given the covariates z are associated strongly enough with the subgroups. In contrast, mixtures are more suitable to detect subgroups if they are only loosely associated with the covariates z, as long as the differences in $\beta_{(k)}$ are strong enough.
- Mixtures are more suitable if the association between covariates and subgroups is smooth and monotonic. Trees are more suitable if the association is characterized by abrupt shifts and possibly non-monotonic.
- If several covariates determine the subgroups simultaneously, mixtures are more suitable, whereas trees are more suitable if z includes several noise variables unconnected to the subgroups.

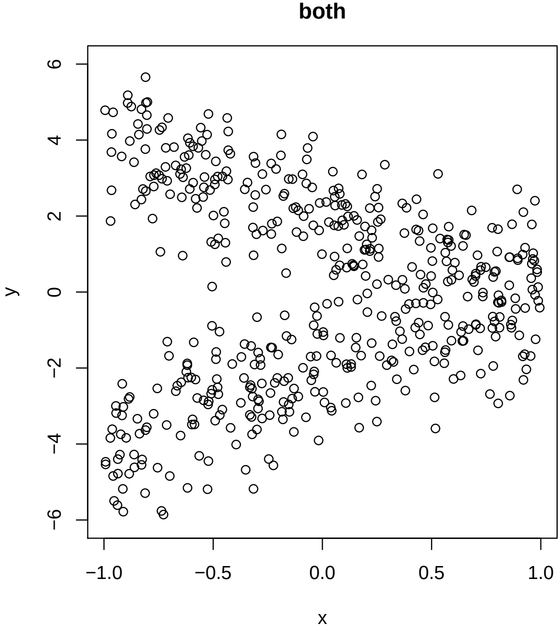
Simulation Design: Coefficients Scenarios



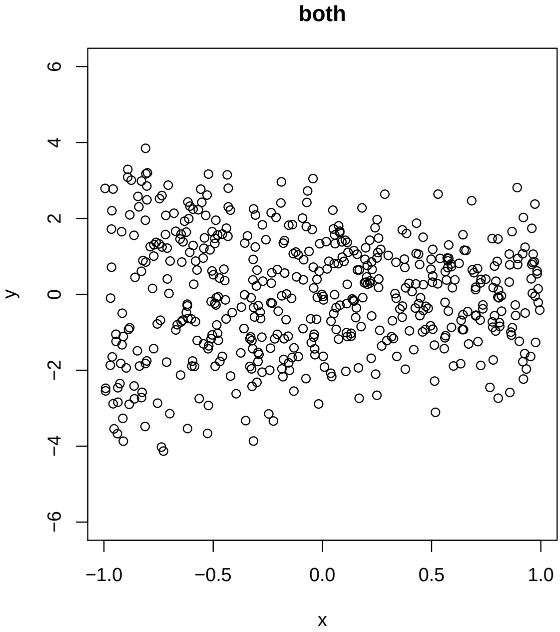
Simulation Design: Coefficients Scenarios



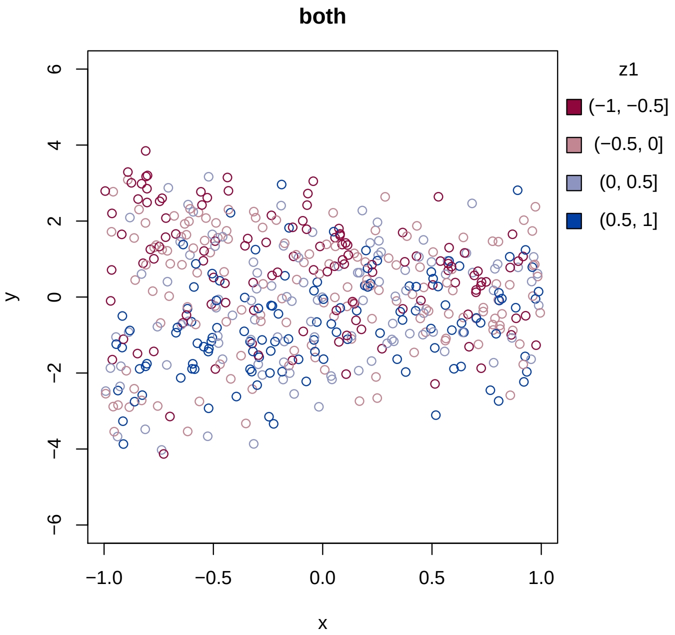
Simulation Design: Coefficients Scenarios



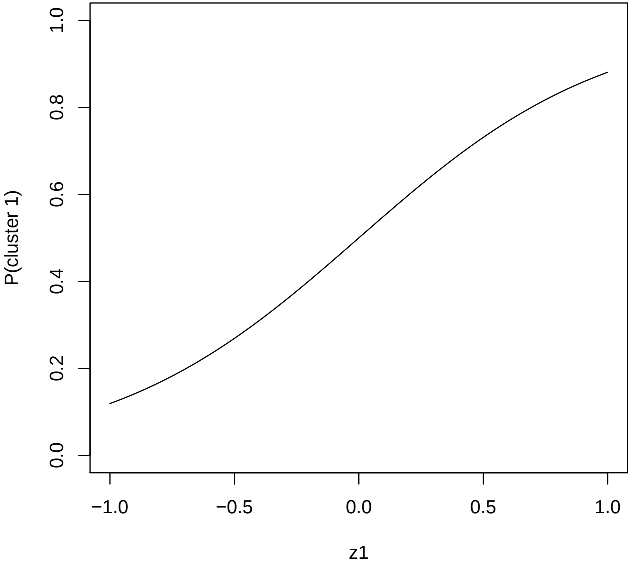
Simulation Design: Coefficients Scenarios



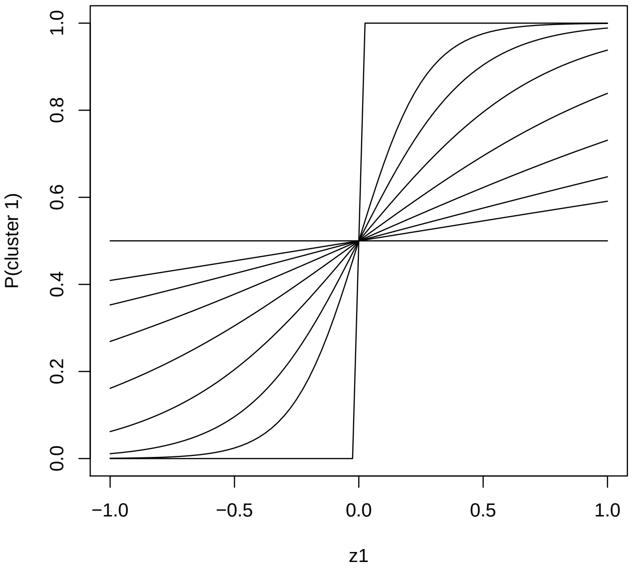
Simulation Design: Coefficients Scenarios



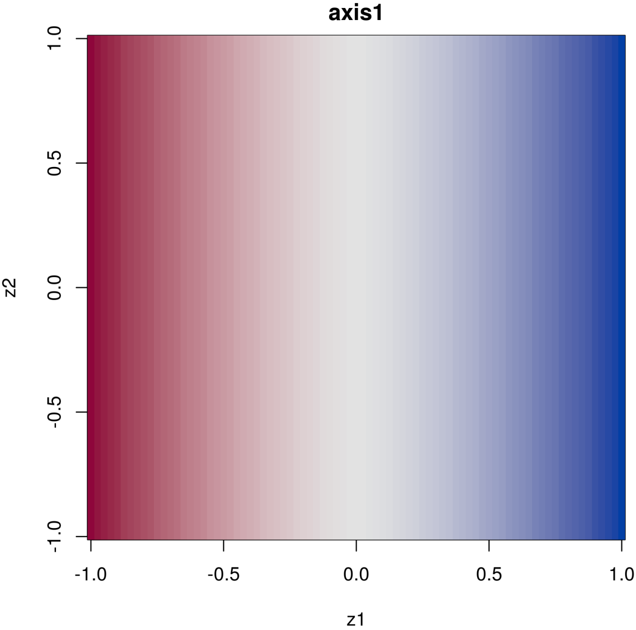
Link between Single Covariate and Clusters



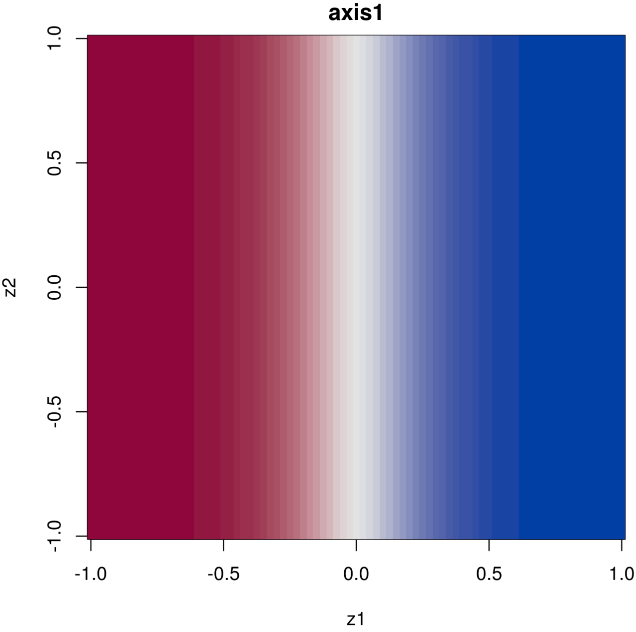
Link between Single Covariate and Clusters



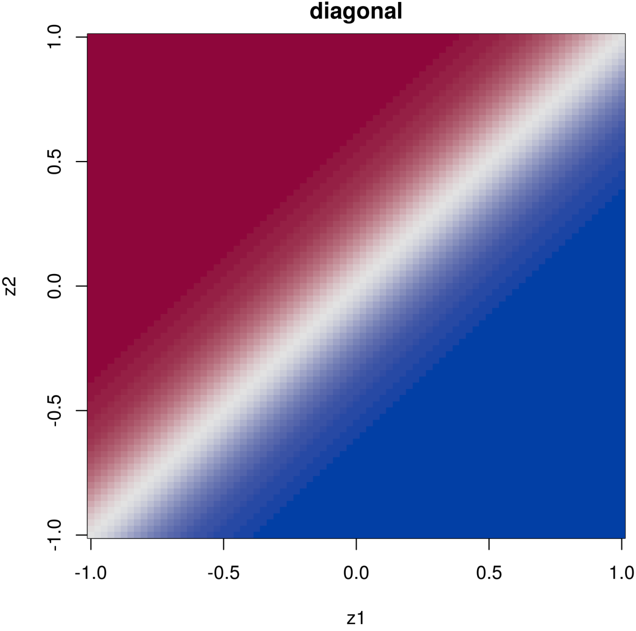
Simulation Design: Covariates Scenario



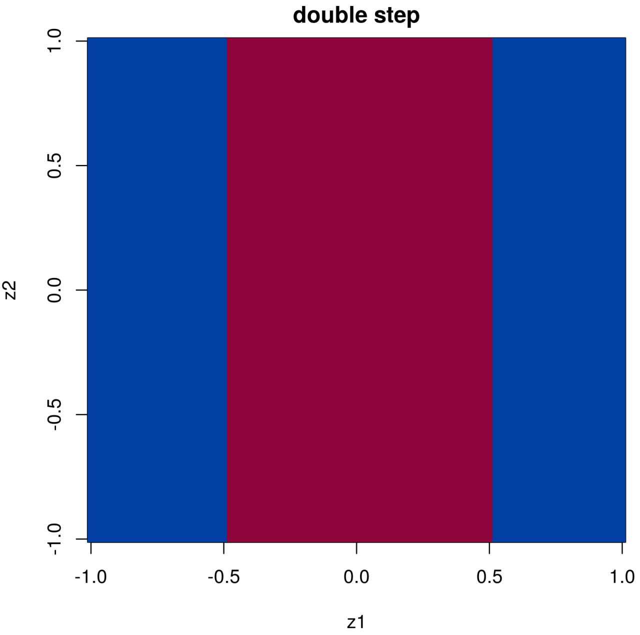
Simulation Design: Covariates Scenario



Simulation Design: Covariates Scenario



Simulation Design: Covariates Scenario



Simulation Design

Relationship between response and regressor:

- *How:* 3 coefficient scenarios `intercept`, `slope`,and `both`
- *How strong:* difference intensity $\kappa \in \{0, 0.05, \ldots, 1\}$.

Covariates connected to subgroups:

- *How:* 2 logistic scenarios `axis1`, `diagonal`, and 1 step scenario with a `double step`.
- *How strong:* separation intensity $\nu \in \{-1, -0.5, \ldots, 2\}$ (only for logistic scenarios).

Simulation Design

Further variations:

- Sample size $n \in \{200, 500, 1000\}$.
- Include 2 additional noise covariates z_3 and z_4, or not.

Technical details:

- The 2 subgroups are of equal size.
- 500 datasets per condition.
- Fitted models: mixture with and without concomitants, and a tree.
- Selection of $\hat{K}$ via BIC from $K = \{1, \dots, 4\}$.

Simulation Design

Outcome assessment:

1. *Hit rate*: Rate of selecting more than one subgroup, i.e., splitting more than once or selecting $\hat{K} > 1$.
2. Compare estimated clustering to true clustering via *Cramér's coefficient*.

Software:

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

Simulation Results

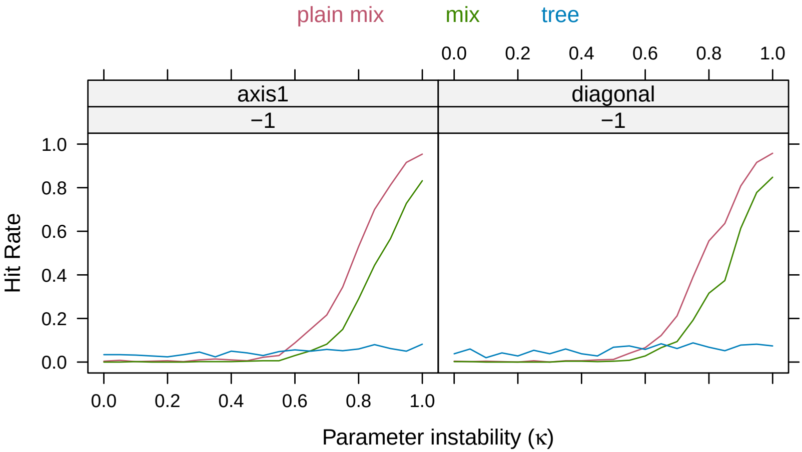
Here, exemplary results for

- Coefficients scenario: `both`.
- Sample size: $n = 200$.
- Without noise variables z_3 and z_4.
- Covariates scenarios: `axis1` and `diagonal` with three levels $\nu = \{-1, 0, 1\}$ of separation between subgroups, and `double step`.

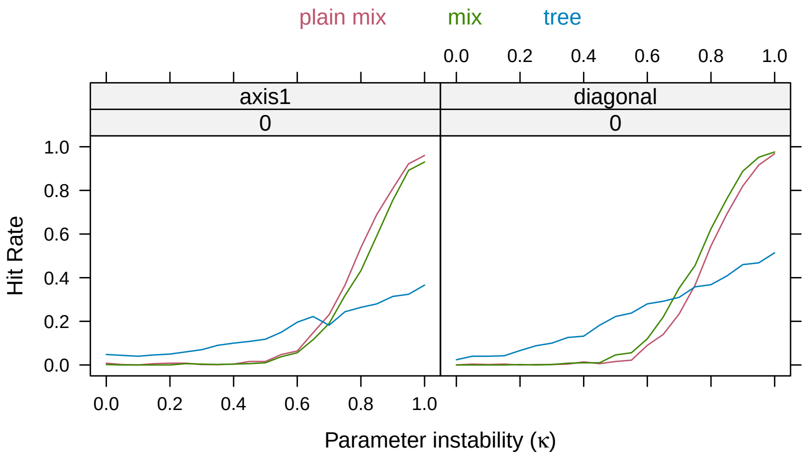
Not shown:

- Results more pronounced for larger numbers of observations ($n = 500$ or 1000) and the other two coefficients scenarios (`intercept` and `slope`).
- Results less pronounced if the two additional noise covariates are included, with hit rates dropping slightly stronger for the mixture than the tree.

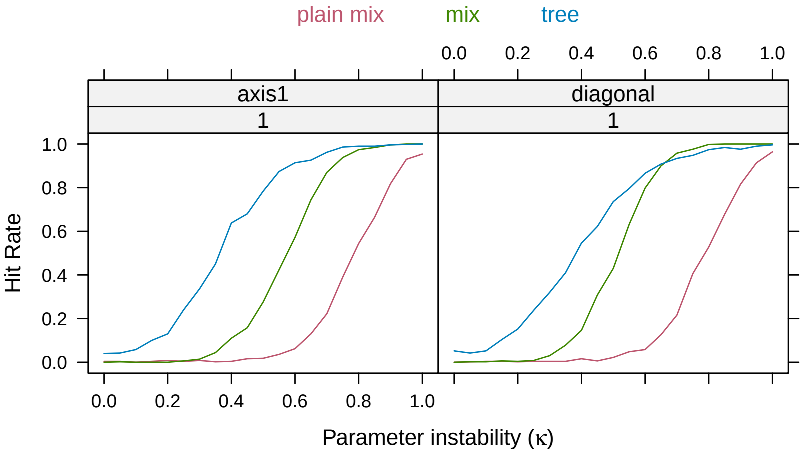
Simulation Results: Hit Rate



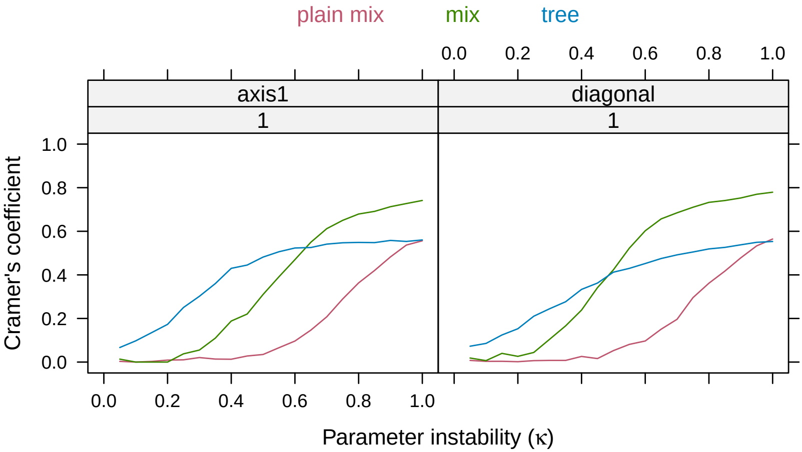
Simulation Results: Hit Rate



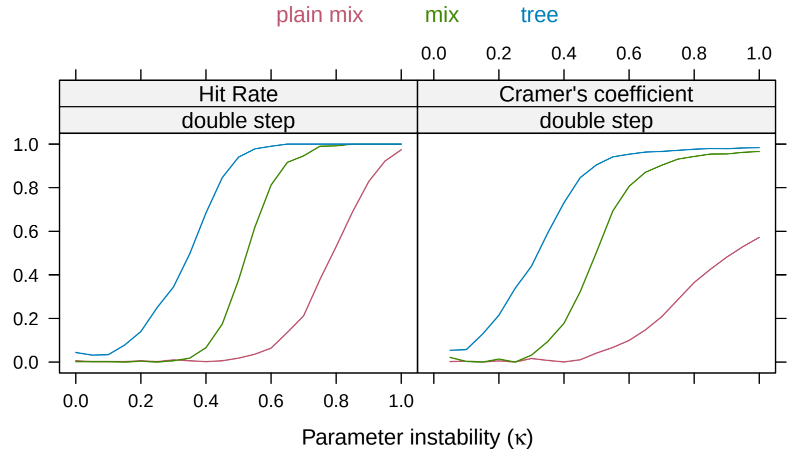
Simulation Results: Hit Rate



Simulation Results: Cramér's Coefficient



Simulation Results: Double Step Scenario



Summary

- Both methods are suitable to detect parameter instability (or lack thereof) and recover the subgroups (if any) but no method universally outperforms the other.
- Strong association between subgroups and covariates: The tree is able to detect smaller differences in the parameters than the mixtures.
- Weak association but reasonably strong difference between parameters: Mixtures outperform the tree.
- Mixture models can also detect latent subgroups without any association to covariates.
- The approximation of a smooth transition between classes through sample splits works rather well.
- Our suggestion: Keep both methods in your toolbox.

References

Frick H, Strobl C, Zeileis A (2014). “To Split or to Mix? Tree vs. Mixture Models for Detecting Subgroups.” *COMPSTAT 2014 – Proceedings in Computational Statistics.*

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

Hothorn T, Zeileis A (2014). “partykit: A Modular Toolkit for Recursive Partytioning in R.” *Working Paper 2014-10, Working Papers in Economics and Statistics, Research Platform eeecon, Universität Innsbruck.*

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

Grün B, Leisch F (2008) “FlexMix version 2: Finite mixtures with concomitant variables and varying and constant parameters.” *Journal of Statistical Software*, 28(4), 1–35.