University of Innsbruck

Using recursive partitioning to account for parameter heterogeneity in multinomial processing tree models

Florian Wickelmaier, Achim Zeileis

Working Papers in Economics and Statistics

2016-26

Research Platform eeecon
Empirical and Experimental Economics
University of Innsbruck
http://eeecon.uibk.ac.at/

University of Innsbruck
Working Papers in Economics and Statistics

The series is jointly edited and published by

- Department of Banking and Finance
- Department of Economics
- Department of Public Finance
- Department of Statistics

Contact address of the editor:
Research platform “Empirical and Experimental Economics”
University of Innsbruck
Universitaetsstrasse 15
A-6020 Innsbruck
Austria
Tel: + 43 512 507 7171
Fax: + 43 512 507 2970
E-mail: eeecon@uibk.ac.at

The most recent version of all working papers can be downloaded at
http://eeecon.uibk.ac.at/wopec/

For a list of recent papers see the backpages of this paper.

Using Recursive Partitioning to Account for Parameter Heterogeneity in Multinomial Processing Tree Models

Florian Wickelmaier
University of Tübingen

Achim Zeileis
Universität Innsbruck

Abstract

In multinomial processing tree (MPT) models, individual differences between the participants in a study lead to heterogeneity of the model parameters. While subject covariates may explain these differences, it is often unknown in advance how the parameters depend on the available covariates, that is, which variables play a role at all, interact, or have a nonlinear influence, etc. Therefore, a new approach for capturing parameter heterogeneity in MPT models is proposed based on the machine learning method MOB for model-based recursive partitioning. This recursively partitions the covariate space, leading to an *MPT tree* with subgroups that are directly interpretable in terms of effects and interactions of the covariates. The pros and cons of MPT trees as a means of analyzing the effects of covariates in MPT model parameters are discussed based on a simulation experiment as well as on two empirical applications from memory research. Software that implements MPT trees is provided via the `mpttree` function in the *psychotree* package in R.

Keywords: multinomial processing tree, model-based recursive partitioning, parameter heterogeneity.

1. Introduction

Multinomial processing tree (MPT) models are a class of statistical models for categorical data. These models are associated with a graph resembling a probability tree, the links being the parameters, the leaves being the response categories. The path from the root to one of the leaves represents the latent cognitive processing steps a participant executes to arrive at a given response. Since they were introduced in a seminal article (Riefer and Batchelder 1988), MPT models have been applied in numerous ways in cognitive psychology and in related fields (Batchelder and Riefer 1999; Erdfelder, Auer, Hilbig, Aßfalg, Moshagen, and Nadarevic 2009).

As an example, consider an experimental paradigm prevalent in memory research for investigating recognition memory. A recognition-memory experiment consists of two phases: In the learning phase, participants are presented with a list of items to be memorized. In the test phase, old items are presented intermixed with new distractor items, and participants have to classify them as either old or new. Figure 1 displays the structure of the one-high-threshold (1HT) model of recognition (Blackwell 1963; Swets 1961), possibly one of the simplest MPT

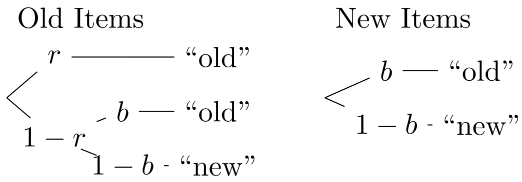

Figure 1: Graph of the one-high-threshold model for recognition memory (Blackwell 1963; Swets 1961). Latent cognitive processes are recognition of an old item (r) and guessing that a not recognized item is old (b).

models. According to this model, an old item is recognized as old with probability r, or, if not recognized, it is guessed that it is old with probability b. Therefore, on the left-hand side of the figure, there are two paths leading from the root of the tree to an old response. Alternatively, displayed on the right-hand side of the figure, a new item can only be guessed as being old with probability b since, according to the model assumptions, such an item never exceeds the recognition threshold.

Frequently, it is the goal of a study to investigate the effects of explanatory variables on the parameters of an MPT model. In order to do so, it is common practice to apply the model to multiple groups defined by these variables and to test for effects (see, e. g., Riefer and Batchelder 1991, who study age effects on memory processes). When the influence of the covariates is linear, it is more powerful to model them directly via specific link functions (Coolin, Erdfelder, Bernstein, Thornton, and Thornton 2015; Michalkiewicz, Coolin, and Erdfelder 2013; Oravecz, Anders, and Batchelder 2015). More generally, covariate effects represent a form of parameter heterogeneity: different settings of covariates may lead to a change in models parameters. Therefore, additional approaches to account for parameter heterogeneity may be employed to study covariate effects; these include latent-class (Klauer 2006) and latent-trait MPT models (Klauer 2010; Smith and Batchelder 2010; Matzke, Dolan, Batchelder, and Wagenmakers 2015). We will discuss these methods in more detail later and compare them to our approach.

In this paper, we introduce MPT trees, a novel approach to incorporating covariates into MPT models. The core of this approach is model-based recursive partitioning (Zeileis, Hothorn, and Hornik 2008), a tree-based computational method from machine learning for detecting parameter heterogeneity across covariates in a data-driven way. The result is a tree-based classification of all individuals into groups where the MPT model parameters are homogeneous within each group. Thus, not only do MPT trees test for the presence of parameter heterogeneity, but they also capture it (if any) in interpretable groups without the need for pre-specification of the relevant covariates or their interactions.

For illustration, Figure 2 depicts an artificial data set following such a tree. In this data set, the responses of all participants are represented by the 1HT model from Figure 1, but the model parameters vary between three groups that are defined in terms of two covariates x_1 and x_2. A conceivable situation would be a recognition experiment where x_1 could be an IQ test score (e. g., Fagan 1984) and x_2 could be the amount of training with the task. The interpretation would then be: The recognition probability r is lowest for participants with lowest IQ scores as measured in x_1 (below some threshold or cutoff ν_1), whereas those with higher IQ scores have a higher recognition probability r, which even increases further

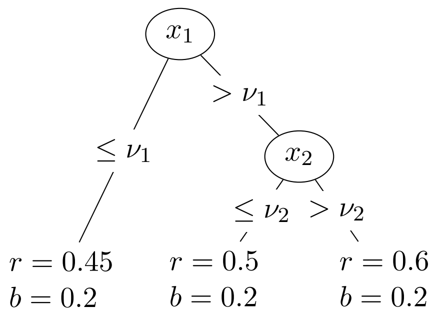

Figure 2: Tree structure for the artificial data. Two covariates (x_1, x_2) along with their binary cutoffs (ν_1, ν_2) define three groups with specific r parameters of the one-high-threshold model.

with sufficient training x_2 above some threshold ν_2. In this artificial data set, the guessing probability b is the same across all groups.

Note that this MPT tree combines two levels of trees. The first level is the tree of the MPT model (Figure 1). Its tree structure has to be specified in advance and is assumed to be constant in the entire population; the parameters (r and b) associated with its links, however, are allowed to vary and need to be estimated. The second level is the recursive partitioning based on the subject covariates (Figure 2). It does *not* have to be specified in advance but is "learned" based on the available data. Specifically, neither the correct order of the variables x_1 and x_2 nor their cutoffs ν_1 and ν_2 have to be pre-specified but are estimated from the data by model-based recursive partitioning.

The remainder of this paper is organized as follows: First, the steps of the model-based recursive partitioning algorithm for MPT models are outlined. Next, the performance of the method is investigated in a simulation study based on the artificial scenario from Figure 2. Then, the use of recursive partitioning for investigating effects of covariates on cognitive processes is illustrated with two examples from memory research. Finally, our approach is discussed in the context of other methods for incorporating covariates or for detecting parameter heterogeneity in MPT models.

2. Recursive partitioning based on MPT models

Model-based recursive partitioning (MOB; Zeileis *et al.* 2008) is a general approach to account for heterogeneity in parametric models. The basic idea of MOB is that the fit of a model may be improved by splitting the sample and fitting the model to subgroups. These subgroups are formed automatically: the algorithm "learns" the optimal partitions using the covariates available. Thus, by recursively partitioning the sample, MOB seeks to explain parameter heterogeneity, which is also called parameter instability in the machine-learning context, by means of main effects and interactions of subject covariates.

There already exist adaptations of the MOB algorithm to (multivariate) linear and generalized linear models (Zeileis *et al.* 2008), to the Bradley-Terry-Luce choice model (Strobl, Wickelmaier, and Zeileis 2011), and to the Rasch model and other psychometric models from item response theory (Komboz, Strobl, and Zeileis 2016; Strobl, Kopf, and Zeileis 2015). Common to these adaptations are the general steps of the MOB algorithm, which are, in summary, as follows:

1. Fit a parametric model to the current (sub-)sample, starting with the full sample, by estimating its parameters via maximum likelihood.
2. Assess the stability of the model parameters with respect to each available covariate. This is done using a parameter instability test based on the maximum likelihood scores.
3. If there is significant instability, select the covariate associated with the strongest instability. Compute the cutpoint that leads to the greatest improvement in the model's likelihood. Split the sample.
4. Repeat steps 1 to 3 until there is no more significant parameter instability or until the minimum sample size is reached.

Thus, all steps are based on the model's likelihood, and the size of the resulting tree is controlled by significance tests.

In this paper, we will introduce MPT trees, an adaptation of model-based recursive partitioning to MPT models. In the following, the steps of the algorithm specific to MPT models are explained. For the general procedure of model-based recursive partitioning we refer to Zeileis *et al.* (2008).

2.1. Likelihood of MPT models

The data consist of the response frequencies for each of $i = 1, \dots, n$ participants in each of $j = 1, \dots, J$ response categories. Let $y_i = (y_{ij})$ be the vector of observed frequencies for participant i in the response categories. Let $\Theta = (\vartheta_k)$, $k = 1, \dots, K$, $\Theta \in [0, 1]^K$, be the vector of MPT model parameters. The MPT model defines the probability of a response in each category, $p_j = p_j(\Theta)$, as a function of the parameters. Assuming independence of the responses, the data follow a multinomial distribution. The joint likelihood becomes

$$L(\Theta; y_1, \dots, y_n) = \prod_{i=1}^{n} \left(y_{i+}! \prod_{j=1}^{J} \frac{p_j(\Theta)^{y_{ij}}}{y_{ij}!} \right), \tag{1}$$

where $y_{i+} = \sum_{j=1}^{J} y_{ij}$, and it only depends on the MPT model parameters Θ. The kernel of the log-likelihood is proportional to

$$\log L(\Theta; y_1, \dots, y_n) \propto \sum_{i=1}^{n} \sum_{j=1}^{J} y_{ij} \log p_j(\Theta) = \sum_{i=1}^{n} \ell(\Theta; y_i), \tag{2}$$

where $\ell(\Theta; y_i)$ denotes the log-likelihood contribution of the i-th person.

For example, in the recognition-memory experiment introduced above, items are either old or new, and participants have to classify them as old or new in a recognition test. Therefore, the responses of an individual fall into one of $J = 4$ categories, resulting in a two-by-two table of response frequencies:

		Response	
		old	new
Item	old	y_{i1}	y_{i2}
	new	y_{i3}	y_{i4}

The 1HT model (Figure 1) has two parameters, $\Theta = (r, b)$, and the predicted probabilities for each response category are

$$\begin{aligned} p_1(\Theta) &= r + (1 - r)b & p_2(\Theta) &= (1 - r)(1 - b) \\ p_3(\Theta) &= b & p_4(\Theta) &= 1 - b. \end{aligned} \tag{3}$$

Many prevalent MPT models consist of multiple category systems, or subtrees. For example, the 1HT model has two response categories for old items and two for new items. Thus, technically, the corresponding likelihood is product (or joint) multinomial. For parameter estimation and for the instability tests presented below, however, this distinction is irrelevant, so we keep the simplified notation of J categories in total.

2.2. Maximum likelihood estimation

Maximum likelihood estimates of MPT model parameters are obtained by maximizing Equation 2 with respect to Θ,

$$\hat{\Theta} = \arg\max_{\Theta} \sum_{i=1}^{n} \ell(\Theta; y_i). \tag{4}$$

One way of solving Equation 4 is by means of the expectation-maximization (EM) algorithm described in Hu and Batchelder (1994). The idea is that parameter estimation would be simplified if not only the category frequencies were known, but also the frequencies of every single branch from the root to the leaves. The latter are missing, of course, but their expected value can be computed given initial parameter values (E step). With the expected branch frequencies at hand, the parameter values are updated (M step). These two steps are iterated until the likelihood converges to a local maximum.

A prerequisite for the application of the EM algorithm is that the link probabilities in a branch take the form

$$\gamma \prod_{k=1}^{K} \vartheta_k^{\alpha} (1 - \vartheta_k)^{\beta}, \tag{5}$$

where $\alpha, \beta \in \{0, 1\}$ indicate the occurrence of either ϑ_k or $1 - \vartheta_k$, and γ is a nonnegative real number. Equation 5 is the structural restriction of the class of MPT models that can be represented by binary trees. Other model types have to be suitably reparameterized for the algorithm to apply.

An alternative way of solving Equation 4 is by directly maximizing the log-likelihood using analytical gradients (Riefer and Batchelder 1988). When doing so, it is advantageous to transform the parameters to the logit scale in order to remove the $[0, 1]$ boundaries.

Interval estimation is straightforward since both parameter estimation methods lead to analytical expressions for the observed Fisher information or negative Hessian matrix (Hu and Batchelder 1994, Equation 16; Riefer and Batchelder 1988, Equation 21). When working on the logit scale, the information matrix may be obtained by the multivariate delta method (Agresti 2002; Bishop, Fienberg, and Holland 1975; Grizzle, Starmer, and Koch 1969). The approximate covariance matrix is available via the inverse information matrix.

Once the model is fit to the full sample, we want to test for parameter heterogeneity that can be attributed to the covariates; this is described next.

2.3. Detection of parameter instability

In the framework of model-based recursive partitioning, a test of parameter instability checks if the model fit can be improved by splitting the sample according to some covariate X and fitting the model to the subsamples. Under the null hypothesis of parameter homogeneity (or stability), it is assumed that Equation 1 holds and thus the parameter vector is equal for all participants,

$$H_0: \ \Theta_i = \Theta_0 \quad (i = 1, \dots, n), \tag{6}$$

where Θ_i is the parameter vector of individual i. The alternative hypothesis is that the parameter vector varies as a function of X with observations $x_1, \dots, x_n$,

$$H_1: \ \Theta_i = \Theta(x_i) \quad (i = 1, \dots, n). \tag{7}$$

The exact pattern of variation is usually unknown. For unordered categorical X, it is common to test for differences in the parameter vector for all categories of X. For continuous and ordinal X, one frequent pattern of interest is an abrupt change in the parameter vector at an unknown cutpoint ν,

$$H_1^*: \ \Theta_i = \begin{cases} \Theta^{(A)} & \text{if } x_i \leq \nu, \\ \Theta^{(B)} & \text{if } x_i > \nu, \end{cases} \tag{8}$$

where $\Theta^{(A)} \neq \Theta^{(B)}$ (Merkle and Zeileis 2013; Merkle, Fan, and Zeileis 2014). Possible examples of such a pattern include effects of age, expertise, intelligence, etc.

To test the above hypotheses, the parameter instability statistics employed here make use of the individual contributions to the score function or subject-wise estimating function, $s(\Theta; y_i)$, and assess the deviations from its mean zero. For MPT models, due to the multinomial form of the likelihood, the contribution of individual i to the score function is given by

$$s(\Theta; y_i) = \frac{\partial \ell(\Theta; y_i)}{\partial \Theta} = \sum_{j=1}^{J} y_{ij} \frac{\partial \log p_j}{\partial \Theta} = \sum_{j=1}^{J} \frac{y_{ij}}{p_j(\Theta)} \frac{\partial p_j}{\partial \Theta}. \tag{9}$$

For example, in the 1HT model, the individual score contributions are determined by first partially differentiating the probabilities in Equation 3 with respect to Θ; this yields

$$\begin{aligned} \frac{\partial p_1}{\partial \Theta} &= \begin{pmatrix} 1-b \\ 1-r \end{pmatrix} & \frac{\partial p_2}{\partial \Theta} &= \begin{pmatrix} b-1 \\ r-1 \end{pmatrix} \\ \frac{\partial p_3}{\partial \Theta} &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} & \frac{\partial p_4}{\partial \Theta} &= \begin{pmatrix} 0 \\ -1 \end{pmatrix}. \end{aligned} \tag{10}$$

Second, substituting these terms into Equation 9 gives

$$s(\Theta; y_i) = \begin{pmatrix} \dfrac{y_{i1}(1-b)}{r+(1-r)b} + \dfrac{y_{i2}(b-1)}{(1-r)(1-b)} + y_{i3} \cdot 0 + y_{i4} \cdot 0 \\ \dfrac{y_{i1}(1-r)}{r+(1-r)b} + \dfrac{y_{i2}(r-1)}{(1-r)(1-b)} + \dfrac{y_{i3}}{b} - \dfrac{y_{i4}}{1-b} \end{pmatrix}. \tag{11}$$

The score contributions behave like residuals and are diagnostic of the model fit. Evaluation of the score function for each individual at the joint maximum likelihood estimate $\hat{\Theta}$ measures the extent to which the model maximizes each individual's likelihood: Scores further from zero indicate that the model provides a poorer description of such individuals. The general idea of the tests applied here is that under the null hypothesis of parameter homogeneity (6), the individual score contributions, when ordered by any covariate X, fluctuate randomly around zero. When parameters are not homogeneous across the entire sample, however, the scores systematically depart from zero. To capture these deviations, the cumulative score process

$$B(t; \hat{\Theta}) = \hat{I}^{-1/2} n^{-1/2} \sum_{i=1}^{\lfloor n \cdot t \rfloor} s(\hat{\Theta}; y_{(i)}) \quad (0 \leq t \leq 1), \tag{12}$$

is employed, where $\lfloor n \cdot t \rfloor$ is the integer part of $n \cdot t$, $\hat{I}$ is an estimate of the covariance matrix of the scores, and $y_{(i)}$ denotes that y_i has been ordered by X. Since the sampling distribution of this process under the null hypothesis is known, critical values and p-values can be derived either analytically or by simulation. The exact from of the test statistic depends on whether the covariate is continuous, categorical, or ordinal.

The tests employed to detect parameter heterogeneity are generalized Lagrange multiplier (LM) tests, also known as score tests. More background information on these tests than provided here is included in several recent articles: Details of the parameter instability tests are discussed by Zeileis and Hornik (2007), who show that they are not restricted to maximum likelihood scores but also apply to other maximum-likelihood-type estimators (M-estimators), like ordinary least squares. Details of the recursive application of these tests and of the model-based recursive partitioning algorithm in general are given by Zeileis *et al.* (2008). Merkle and Zeileis (2013) discuss the tests in the context of measurement invariance with respect to structural equation models. Merkle *et al.* (2014) extend the results to ordered categorical covariates.

2.4. Cutpoint location and recursive partitioning

When all available covariates have been tested for parameter instability using the procedure outlined above and at least one test is significant, the MOB algorithm selects the variable that induces the strongest instability (with the smallest p-value) in order to locate the cutpoint for splitting the sample. The idea behind the estimation of the optimal cutpoint ν is to find the value of the selected covariate with $x_m \leq \nu$ and $x_{m+1} > \nu$ that splits the current sample such that the likelihood in the two subsamples

$$\ell(\hat{\Theta}^{(A)}; y_i, \dots, y_m) + \ell(\hat{\Theta}^{(B)}; y_{m+1}, \dots, y_n) \tag{13}$$

is maximized. For unordered categorical covariates, all possible binary partitions are computed and the one with the maximum segmented likelihood is chosen.

Once the optimal cutpoint is located and the sample is split, the instability tests are recursively conducted in the two subsamples until there is no further significant instability. Within model-based recursive partitioning, there are two built-in mechanisms that prevent inflation of the type I error rate and, consequently, that a tree grows unwarrantedly large: (1) When testing for instability in a subsample, Bonferroni correction is applied. Thus, instability tests become increasingly strict with an increasing number of covariates. (2) Testing proceeds in a nested fashion, that is, only if a test is significant in a subsample will testing continue in nested subsamples. As a consequence of (1) and (2), a tree does not exceed the nominal significance level α (Zeileis *et al.* 2008). We will address the statistical performance of the proposed procedure in a simulation study presented next.

3. Simulation study

This section describes a simulation study to investigate power, type I error rate, and classification accuracy of MPT trees. The focus of this simulation is restricted to one specific MPT model that is observed under realistic magnitudes of parameter instability and moderate sample sizes. More general simulation results have been reported elsewhere and include power and type I error of score tests for measurement invariance in the context of structural equation modeling (Merkle and Zeileis 2013; Merkle *et al.* 2014), performance of recursive partitioning and comparison to mixture models for linear regression (Frick, Strobl, and Zeileis 2014), performance of Rasch, partial credit, and rating scale trees for detecting differential item functioning (Komboz *et al.* 2016; Strobl *et al.* 2015).

3.1. Simulation design and experimental settings

In order to simulate responses, the 1HT model (see Figure 1) is employed as the data-generating process with group-specific r parameters and a constant b parameter, $\Theta = (r_{group}, b = 0.2)$ for $group \in \{1, 2, 3\}$, see Figure 2. Each virtual subject contributes 40 simulated responses (to 20 old and 20 new items). Three subject-specific covariates (x_1, x_2, x_3) are included that are independently uniformly distributed in the interval $[-1, 1]$. The interaction between x_1 and x_2 along with the corresponding binary cutoff values ν_1 and ν_2 defines three groups: $x_1 \leq \nu_1$ versus $x_1 > \nu_1 \wedge x_2 \leq \nu_2$ versus $x_1 > \nu_1 \wedge x_2 > \nu_2$. The noise variable x_3 is unrelated to the groups.

The magnitude of parameter instability is controlled by the deviation $\delta \in \{0, 0.01, 0.02, \ldots, 0.20\}$ from the average recognition probability $r = 0.5$. The group-specific recognition probabilities are $r_1 = 0.5 - \delta/2$, $r_2 = 0.5$, and $r_3 = 0.5 + \delta$. Thus, $\delta = 0$ corresponds to parameter homogeneity across the three groups with $r_1 = r_2 = r_3 = 0.5$. The setup with $\delta = 0.1$ is shown in Figure 2. Moreover, three small to moderate sample sizes $n \in \{80, 120, 200\}$ are considered. We expect that increasing both the magnitude δ and the number n of participants will lead to improved detection performance of the MPT trees.

Two scenarios are considered for the cutoffs ν_1 and ν_2: First, the median value of the distributions of x_1 and x_2 is used, that is, $\nu_1 = \nu_2 = 0$, so that on average the group sizes are 1/2, 1/4, and 1/4, respectively, of the total sample. Second, $\nu_1 = -0.5$ and $\nu_2 = 0.5$ are used as the cutoffs resulting in group sizes of about 1/4, 9/16, and 3/16, respectively. Thus, in the latter scenario, the parameter differences are harder to detect because the middle group (with $r_2 = 0.5$) is the largest and the deviating groups are smaller.

For benchmarking the power and the accuracy of MPT trees (see below for details on the outcome measures), the frequently used likelihood ratio test (LRT) is employed as a reference method. Because the LRT requires a pre-specified split into groups, we consider the common strategy of splitting at the median of a relevant covariate. Here, we consider splitting either x_1 or x_2 at their corresponding medians. Note that this gives the LRT a somewhat unfair advantage, especially in the first scenario where the true cutoffs are at the median of zero. Also, the irrelevant covariate x_3 is not considered at all and no Bonferroni correction is applied for aggregating multiple LRTs.

In summary, for each of the two cutoff scenarios and each combination of magnitude of parameter instability and sample size, 2000 data sets are generated to compute the outcome measures below for the MPT tree method, the LRT with splitting at the median of x_1, and the LRT with splitting at the median of x_2, respectively. All simulations were run in R using software described in the "Computational details" section.

3.2. Outcome measures

Two kinds of outcome measures are considered: (1) the power with which the MPT tree and the two LRTs reject the null hypothesis of parameter stability; (2) the accuracy with which the true groups were recovered. For the MPT tree, the power is the proportion of experiments in which the score test in the root node is significant for x_1 or x_2, that is, in which the sample is split at least once. For comparison, the power of the two LRTs is the proportion of experiments in which the null hypothesis of $r_{x_1 \leq 0} = r_{x_1 > 0}$ or $r_{x_2 \leq 0} = r_{x_2 > 0}$, respectively, is rejected. Note that the hypothesized cutoff value of zero, the median of x_1 and x_2, either coincides with the true cutoff (first cutoff scenario) or differs (second cutoff scenario).

The classification accuracy for MPT trees is assessed using the Cramér coefficient of agreement defined as the normalized χ^2 statistic of the crosstabulated true and predicted group membership (Mirkin 2001). This takes a value of zero if the true and predicted groups are uncorrelated, and a value of one if true and predicted groups essentially match. However, unlike many other cluster indices (e. g., the Rand index), it does not penalize if some of the groups are split up further. This property is particularly useful when assessing recursive partitions that might need several splits to form a certain group. Note that for the LRTs, we do not simulate the Cramér coefficient but simply determine its theoretical value assuming a given cutoff of zero in either one of x_1, x_2, or x_3 alone.

3.3. Results

Figure 3 displays the simulated power of the MPT tree in comparison to LRTs based on x_1 or x_2 as a function of the magnitude of parameter instability (δ) and sample size (n). In the first row, the results for the scenario are shown where the true cutoffs coincide with the medians of x_1 and x_2, respectively. Thus, the LRT that splits at the median of x_1 performs best for all magnitudes and sample sizes as it tests for the correct split of the root node. The MPT tree performs second best (except for very small magnitudes δ), although it neither knows which variable (x_1, x_2 or x_3) nor which cutoff point is correct. Furthermore, under the null hypothesis of homogeneous parameters ($\delta = 0$), the MPT tree holds its nominal significance level of 5%, although it is somewhat conservative, especially for small sample sizes n, due to the asymptotic nature of the tests employed. Finally, the LRT that splits at the median of x_2 performs worst among the three methods despite using the correct split in one of the relevant variables. In the second row, where the true cutoffs do not coincide with the medians, the power of all methods goes down because the groups are more unbalanced (see above) and, more importantly, the search for the correct variables and cutoffs in the MPT tree pays off. This advantage of the MPT tree over the LRTs becomes more pronounced for larger magnitudes of parameter instability and larger sample size.

In summary, because the MPT tree always determines the cutoffs in a data-driven way, it cannot profit from "knowing" the true cutoffs in contrast to the LRTs. Therefore, the latter tests will have a power advantage over MPT trees if the true cutoff and the relevant variables are used. Conversely, when the cutoffs are unknown, the MPT tree has an advantage over LRTs, which depend on an often arbitrary choice of the cutoff (here, the median).

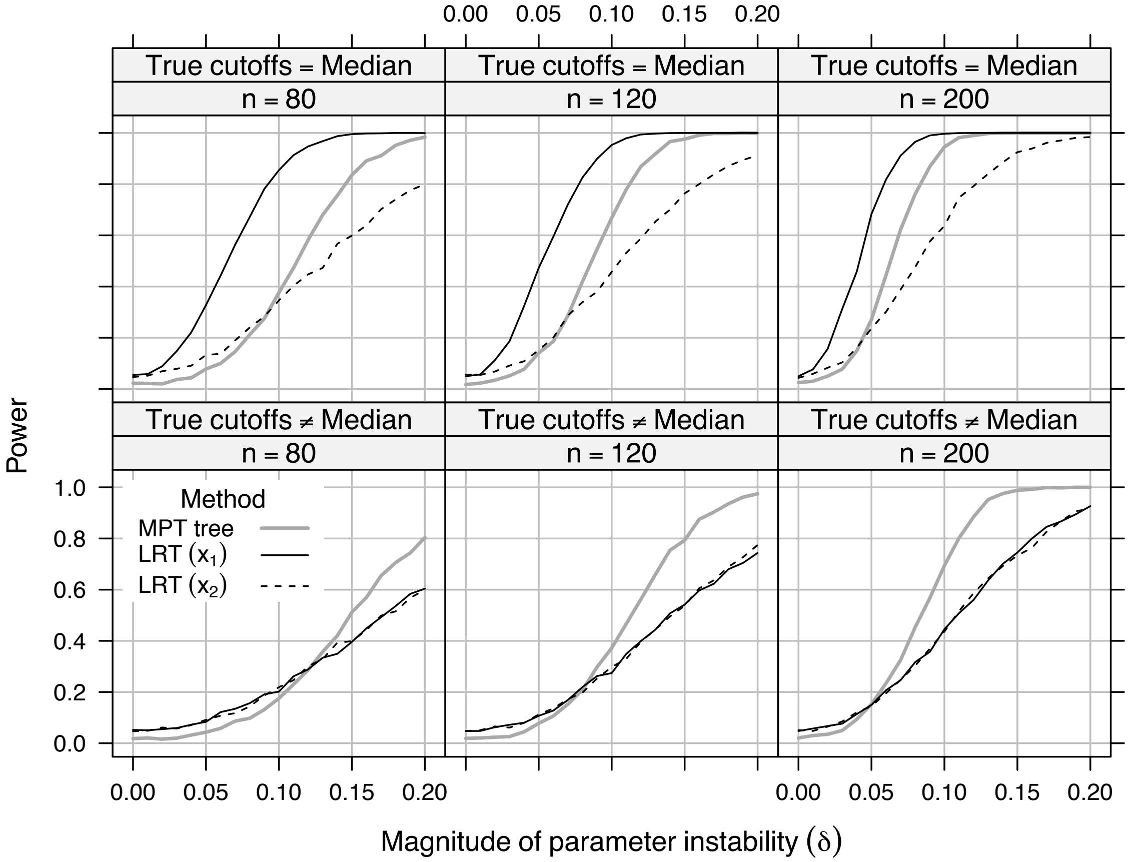

Figure 3: Simulated power as a function of the magnitude of parameter instability (δ), sample size (n), and the method used to test for instability. For the likelihood ratio tests (LRT), the median of x_1 or x_2, respectively, is used that either coincides with the true cutoff (upper panel) or not (lower panel).

The second part of the results shows the accuracy of the MPT tree in recovering the true partitions. Figure 4 displays the average Cramér coefficient of agreement between true and predicted group membership as a function of the magnitude of parameter instability (δ) and sample size (n). In both cutoff scenarios, the Cramér coefficient of the MPT tree increases with increasing parameter instability and sample size; however, it is generally somewhat lower in the second scenario in the right panel. This is due to the fact that the groups 1 and 3, which differ from the middle group 2, are smaller and hence harder to detect. As a reference, both panels show the theoretical Cramér coefficient of the deterministic splits using the medians of x_1, x_2, and x_3, respectively. For the split in x_3, this is generally 0 because this split is completely unrelated to the true groups in either scenario. For a split at the median of x_1 in the first scenario, the Cramér coefficient is 1 because this exactly catches the first split of the tree (and ignoring the second split is not penalized by the Cramér coefficient). However, if the true cutoff in x_1 differs from the median, the theoretical Cramér coefficient drops to 1/3. Similarly, the Cramér coefficient for the deterministic split at the median of x_2 yields 1/2 if this coincides with the true cutoff, and 1/4 otherwise. Thus, in both scenarios, the Cramér coefficient of the MPT tree approaches the best possible value of 1 only for large δ and/or n;

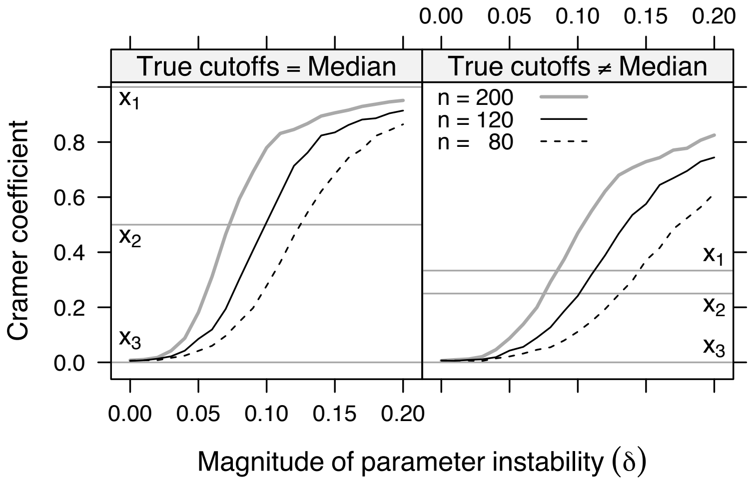

Figure 4: Average Cramér coefficient of agreement between true and MPT-tree-predicted group membership as a function of the magnitude of parameter instability (δ) and sample size (n). Horizontal lines indicate the Cramér coefficient when splitting the sample along the median of x_1, x_2, or x_3, which may either be the true cutoff (left panel) or not (right panel). As x_3 is unrelated to the groups, its Cramér coefficient is zero.

however, in the second scenario, this outperforms the deterministic splits already for values of δ above around 0.1 (depending on the sample size).

In conclusion, these results show that subgroups previously defined on the covariates are satisfactorily recovered by recursive partitioning based on an MPT model. In contrast to the likelihood ratio test, neither the relevant covariates nor the cutpoints have to be known in advance. A limitation of the results presented here is that they were obtained for a single MPT model (the 1HT model) and two similar tree structures (cutoff scenarios). Nevertheless, similar results can be obtained in other setups (see references cited above). Hence, we believe that these insights contribute evidence that MPT trees constitute a useful tool for detecting parameter heterogeneity in realistic settings.

4. Two applications

This section covers two applications of recursive partitioning based on MPT models. The first analyzes a new data set for which the potential partitions were unknown a priori (as in most applications) but were the primary research interest. The second is a reanalysis of a published data set (Riefer, Knapp, Batchelder, Bamber, and Manifold 2002), where the focus is on how well the MPT tree succeeds in uncovering the a priori hypothesized partitions.

4.1. Source monitoring

The first application considers a typical source monitoring experiment: Participants study two lists of items as presented by either Source A or Source B. Afterwards, in a memory test, participants are shown old and new items intermixed and asked to classify them as either A, B, or new (N).

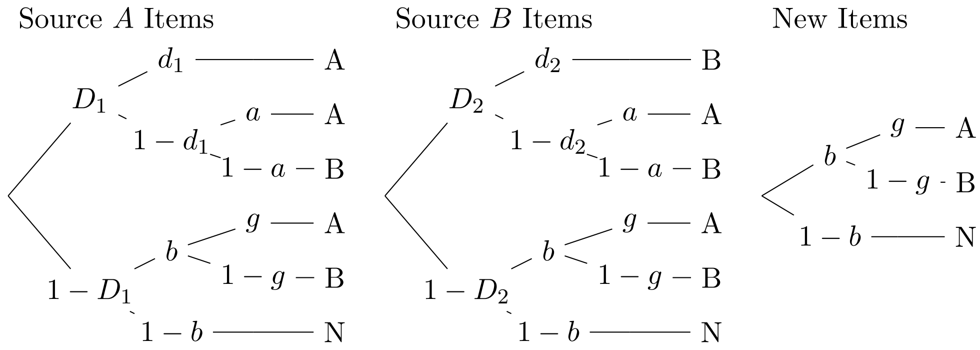

Figure 5: Graph of the MPT model for the source monitoring paradigm (Batchelder and Riefer 1990). Latent cognitive processes are detectability of Source A items (D_1), detectability of Source B items (D_2), source discriminabilities for Source A (d_1) and Source B items (d_2), bias for responding "old" to a nondetected item (b), guessing that a detected but nondiscriminated item belongs to Source A (a), and guessing that the item is a Source A item (g).

Figure 5 displays the MPT model for the source monitoring paradigm by Batchelder and Riefer (1990). To illustrate, consider the paths from the root to an A response for a Source A item (left tree in the figure). With probability D_1, a respondent detects an item as old. If, in a second step, he or she is able to discriminate the item from a Source B item (d_1), then the response will correctly be A; else, if discrimination fails ($1-d_1$), a correct A response can only be guessed with probability a. If the item was not detected as old in the first place ($1 - D_1$), the response will be A only if there are both a response bias for "old" (b) and a guess for the item being Source A (g). The remaining paths in the left tree lead to classification errors (B, N). The middle and right trees in Figure 5 represent processing of Source B or new items, respectively.

Such a source monitoring experiment was conducted at the Department of Psychology, University of Tübingen. The sample consisted of 128 participants (64 female) aged between 16 and 67 years. Two source conditions were used in the study phase: In the first one, respondents had to read the presented items either quietly (think) or aloud (say). In the second one, they wrote them down (write) or read them aloud (say). Items were presented on a computer screen at a self-paced rate. In the final memory test, studied items were mixed with new distractor items, and respondents had to classify them as either A, B, or new by pressing a button on the screen.

The response frequencies are analyzed using the above MPT model for source monitoring (Figure 5; Batchelder and Riefer 1990), where $a = g$ is assumed for identifiability. In addition, discriminability is assumed to be equal for both sources ($d_1 = d_2$). As a research question, we investigate whether there are any effects of source condition, gender, or age on the model parameters. The MPT tree uses a Bonferroni-corrected significance level of $\alpha = 0.05$ and a minimum number of five participants per node.

Figure 6 shows the tree resulting from recursive partitioning of the source monitoring MPT model. The node numbers are labels assigned from left to right, starting from the top, used to identify a given node. Table 1 displays the results of the parameter instability tests for every node. In Node 1, the full sample, only source type is significant, $S = 28.48$, $p < 0.001$,

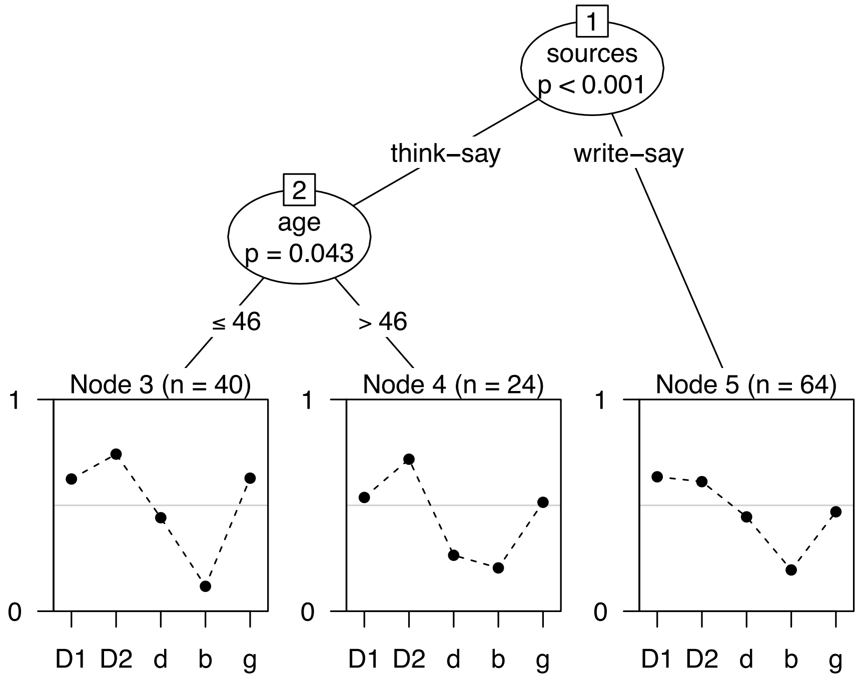

Figure 6: Partitioned MPT model for source monitoring data indicating that parameters vary with combinations of the covariates source type and age.

Node	Sources		Age		Gender	
	S	p	S	p	S	p
1	**28.48**	**0.000**	16.93	0.249	9.00	0.292
2	–	–	**20.77**	**0.043**	2.84	0.924
3	–	–	10.25	0.763	4.28	0.760
4	–	–	8.59	0.822	5.46	0.593
5	–	–	8.06	0.965	7.41	0.347

Table 1: Parameter instability test statistic (S) and Bonferroni-adjusted p-value for each covariate per node (see Figure 6). *Note:* Significant test results are in bold face.

so it is selected for splitting; since it is a binary variable, no cutpoint has to be computed. For the think–say subgroup in Node 2, age is selected for splitting, $S = 20.77$, $p = 0.043$, and the optimal cutpoint is found at age 46. No further parameter instability is detected in the subgroups, so the procedure stops. The fact that gender is never selected as the splitting variable suggests that there is no significant parameter heterogeneity with respect to gender.

The resulting three sets of parameter estimates reflect the combined influence of the covariates. For the think–say sources (Nodes 3 and 4 in Figure 6), D_2 exceeds D_1 indicating an advantage of say items over think items with respect to detectability. For the write–say sources (Node 5), D_2 and D_1 are about the same indicating that for these sources no such advantage exists. The think–say subgroup is further split by age with the older participants having lower values on D_1 and d, which suggests lower detectability of think items and lower discriminability as compared to the younger participants. This age effect seems to depend on the type of sources

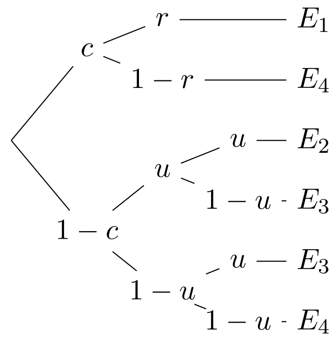

Figure 7: Graph of the storage-retrieval model for pair clustering (Batchelder and Riefer 1986). Latent cognitive processes are clustering of a pair (c), retrieval of a pair (r), and storage/retrieval of a single item (u).

as there is no such effect for the write–say sources. In addition, there are only small effects for the bias parameters b and g, which are psychologically less interesting in this application.

4.2. Storage-retrieval model for pair-clustering data

Riefer *et al.* (2002) report a study on memory deficits in schizophrenic ($n = 29$) and organic alcoholic ($n = 21$) patients, who were compared to two matched control groups ($n = 25$, $n = 21$). Participants were presented with 20 pairs of semantically related words. In a subsequent memory test, they freely recalled the presented words. This procedure was repeated for a total of six study and test trials. Responses were classified into four categories: each pair is recalled adjacently ($E1$), each pair is recalled non-adjacently ($E2$), one word in a pair is recalled ($E3$), neither word in a pair is recalled ($E4$). Riefer *et al.* (2002) analyzed the data using the storage-retrieval model for pair clustering (Batchelder and Riefer 1986) displayed in Figure 7. This model aims at separately measuring storage and retrieval capacities of episodic memory by its parameters c and r.

Figure 8 shows the results of the recursive partitioning based on the storage-retrieval model. Table 2 contains the parameter estimates associated with the end nodes of the MPT tree. The first split separates the two patient and control groups. In the control groups, the parameters improve with repeated presentation of the items: In Node 5, trial is selected as splitting variable, and the optimal cutpoint is $\leq 2, > 2$. Within the ≤ 2 partition, there is again a split into $\leq 1, > 1$. All three parameters constantly increase for one, two, and more than two presentations; the increase is particularly pronounced for the r parameter. The patient groups, on the other hand, do not improve to the same extent. Indeed, their improvement over trials is so weak that it does not attain significance. Neither storage (c) nor retrieval (r) parameters for these groups on average reach the level of the control groups. Marginally significant (Node 2) is the difference between schizophrenic and organic alcoholic patients: While these groups are comparably weak at storing new information, the retrieval is even more impaired in the organic alcoholic patients. The results of our MPT tree analysis of the data are consistent with the findings in Riefer *et al.* (2002). One of the main conclusions is that alcoholic patients with organic brain damage exhibit essentially no improvement in retrieval over trials. Schizophrenic patients improve, albeit less than the control patients, in both storage and retrieval capacities.

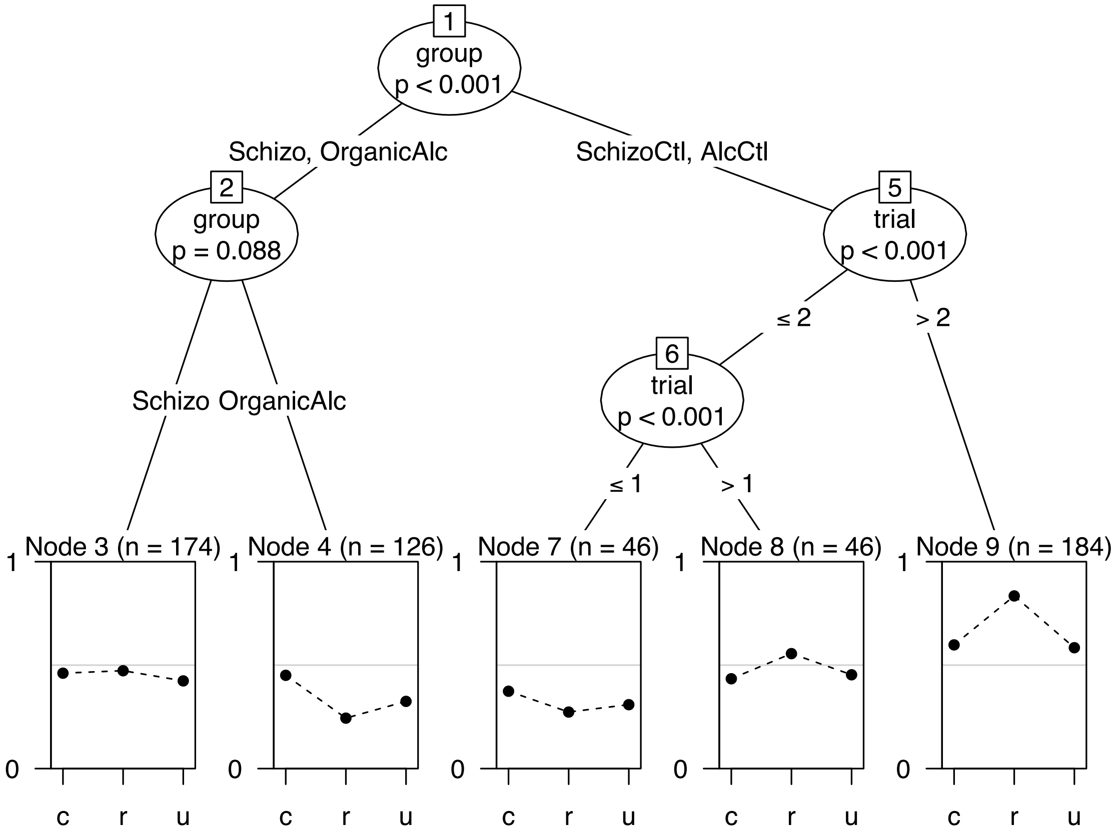

Figure 8: Partitioned storage-retrieval model for pair-clustering data indicating that parameters vary with combinations of the covariates patient group and trial number.

Node	c	r	u
3	0.46	0.47	0.42
4	0.45	0.24	0.32
7	0.37	0.27	0.31
8	0.43	0.56	0.45
9	0.60	0.83	0.58

Table 2: Maximum likelihood estimates of storage-retrieval model parameters associated with the end nodes of the MPT tree in Figure 8.

Other than in the first application, partitioning is done here between observations, not between participants. Each participant contributes six response vectors, one for each trial, to the data set. Consequently, responses from the same participant may appear in more than a single end node. In order to account for the clustering of the responses contributed by the same person, a clustered covariance matrix estimate $\hat{I}$ for the maximum likelihood scores in Equation 12 is employed in the instability tests. Generally, in situations with clustered data, the parameter instability tests within the tree should be considered with care. In the present application, the resulting tree structure is well in line with the hypothesized effects and the results of previous analyses (Riefer *et al.* 2002).

5. Discussion

We introduce MPT trees as a tool for investigating the effects of covariates on MPT model parameters. The core of MPT trees is model-based recursive partitioning (MOB), which recursively searches for covariates that induce parameter heterogeneity. When such a variable is found, an optimal cutpoint is located and the sample is split. As a result, groups of participants are established with (approximately) the same model parameters. As has been illustrated by simulation and in the application examples, the groups do not have to be known beforehand, combinations of relevant covariates are identified, and interactions between covariates are incorporated automatically if the data demand them. The general idea of MOB is not restricted to MPT models but has proved useful in other areas of psychological modeling (Merkle and Zeileis 2013; Strobl, Kopf, and Zeileis 2015; Strobl, Wickelmaier, and Zeileis 2011). Therefore, it seems promising to further extend it to models where individual differences in parameters due to covariate effects need to be accounted for.

There are a number of approaches that partly share the same goals with MPT trees, that is, accounting for individual differences in model parameters by covariate effects or explaining parameter heterogeneity in general. Most notably, such approaches include latent-class MPT models, latent-trait MPT models with random subject effects, and fully parameterized MPT models with covariates as fixed effects. In the remainder, similarities and differences of these methods to MPT trees will be discussed.

MPT trees share similarities with latent-class MPT models (Klauer 2006). As with latent-class models, the sample is partitioned into a discrete number of groups within each of which parameter homogeneity holds, while between groups parameters differ. The difference between these two approaches to parameter heterogeneity is that latent-class models identify a previously specified number of groups on the basis of the response variables only. MPT trees, on the other hand, identify an unknown number of groups based on splits in the available covariates. In doing so, the groups become immediately interpretable in terms of covariate effects and interactions. A caveat is that in MPT trees the parameter heterogeneity is entirely attributed to covariate effects. Thus, without predictive covariates, heterogeneity might go unnoticed. As latent-class MPT models, MPT trees assume homogeneity across items. This is sometimes considered a less problematic assumption than the assumption of subject homogeneity (Klauer 2006); not least because the item material can be experimentally controlled, whereas differences between participants in cognitive processes are often the main focus of the study.

In contrast to models with a discrete number of classes, random effects models represent heterogeneity in a continuous way. The beta MPT model (Smith and Batchelder 2010) uses independent beta hyperdistributions for the MPT parameters to account for individual differences. Similarly, the latent-trait MPT model (Klauer 2010) uses probit-transformed multivariate normal hyperdistributions to represent parameter heterogeneity induced by persons and accounts for correlation between parameters. Both models assume homogeneity of items but can be extended to deal with heterogeneity of persons and items. The crossed random effects extension of the latent-trait MPT model (Matzke *et al.* 2015) accounts for both sources of parameter heterogeneity simultaneously. For these random effects models, parameter estimation and hypothesis testing is carried out in a Bayesian framework using Markov chain Monte Carlo sampling. Whereas random effects models treat parameter heterogeneity by introducing nuisance variables and assumptions about their distributions, MPT trees seek to explain heterogeneity by covariate effects and interactions.

Alternatively to MPT trees, the effects of covariates can be directly incorporated as fixed effects into a parametric model using a specific link function that relates a linear predictor to model parameters. Examples of such an approach include models with probit link function in cultural consensus theory (Oravecz *et al.* 2015), logit-link MPT models (Coolin *et al.* 2015; Coolin, Erdfelder, Bernstein, Thornton, and Thornton 2016; Michalkiewicz *et al.* 2013), and their hierarchical extensions (Arnold, Bayen, and Böhm 2014; Arnold, Bayen, and Smith 2015; Michalkiewicz, Arden, and Erdfelder 2016a; Michalkiewicz, Minich, and Erdfelder 2016b). Such models will have high power for detecting covariate effects if the model specification matches the true data-generating process. The main advantage of MPT trees over direct modeling becomes apparent when such a functional form of the covariate effects cannot be justified or is unknown a priori: Because of its semi-parametric nature, an MPT tree is able to detect even nonlinear effects and interactions between covariates without the need of a fully parameterized model. This flexibility with respect to the functional form and its straightforward graphical representation make MPT trees a useful tool for analyzing the effects of covariates in MPT models.

To summarize, recent methodological, statistical, and computational advances have produced a diversity of methods that account for parameter heterogeneity in MPT models. These methods can be broadly distinguished by whether (1) the heterogeneity-inducing variables are observed and (2) the form of the influence of these variables on the parameters are known. If the relevant variables are not observed, latent class and latent trait MPT models are promising candidates for capturing unobserved heterogeneity. If the variables are observed and the form of their influence is known, fully parameterized MPT models are applicable. If, however, the relevant variables are observed (plus potentially many irrelevant variables) and the form of their influence is unknown, MPT trees provide an elegant approach to detecting and explaining heterogeneity by means of subject covariates.

Computational details

Our results were obtained using R 3.3.1 (R Core Team 2016) and the package *psychotree* 0.15-1 (Zeileis, Strobl, Wickelmaier, Komboz, and Kopf 2016a), which implements MPT trees as introduced in this manuscript. It relies on packages *partykit* 1.1-1 (Hothorn and Zeileis 2015) for recursive partitioning and *psychotools* 0.4-2 (Zeileis, Strobl, Wickelmaier, Komboz, and Kopf 2016), which also contains the data for the source monitoring and the memory-deficits examples. In addition, for the simulation study, *mpt* 0.5-3 (Wickelmaier and Zeileis 2011) was used. R itself and all packages used are freely available under the terms of the General Public License from the Comprehensive R Archive Network (`https://CRAN.R-project.org/`). Code for replicating our analyses is available in the *psychotree* package via `example("mpttree", package = "psychotree")`.

Acknowledgments

We would like to thank William H. Batchelder for making available the data on memory deficits in clinical subpopulations.

References

Agresti A (2002). *Categorical Data Analysis*. John Wiley & Sons, New York.

Arnold NR, Bayen UJ, Böhm FB (2014). "Is Prospective Memory Related to Depression and Anxiety? A Hierarchical MPT Modelling Approach." *Memory*, **23**, 1215–1228. doi:10.1080/09658211.2014.969276.

Arnold NR, Bayen UJ, Smith RE (2015). "Hierarchical Multinomial Modeling Approaches: An Application to Prospective Memory and Working Memory." *Experimental Psychology*, **62**, 143–152. doi:doi:10.1027/1618-3169/a000287.

Batchelder WH, Riefer DM (1986). "The Statistical Analysis of a Model for Storage and Retrieval Processes in Human Memory." *British Journal of Mathematical and Statistical Psychology*, **39**, 129–149. doi:10.1111/j.2044-8317.1986.tb00852.x.

Batchelder WH, Riefer DM (1990). "Multinomial Processing Models of Source Monitoring." *Psychological Review*, **97**, 548–564. doi:10.1037/0033-295x.97.4.548.

Batchelder WH, Riefer DM (1999). "Theoretical and Empirical Review of Multinomial Process Tree Modeling." *Psychonomic Bulletin & Review*, **6**, 57–86. doi:10.3758/bf03210812.

Bishop YMM, Fienberg SE, Holland PW (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge.

Blackwell HR (1963). "Neural Theories of Simple Visual Discriminations." *Journal of the Optical Society of America*, **53**, 129–160. doi:10.1364/josa.53.000129.

Coolin A, Erdfelder E, Bernstein DM, Thornton AE, Thornton WL (2015). "Explaining Individual Differences in Cognitive Processes Underlying Hindsight Bias." *Psychonomic Bulletin & Review*, **22**, 328–348. doi:10.3758/s13423-014-0691-5.

Coolin A, Erdfelder E, Bernstein DM, Thornton AE, Thornton WL (2016). "Inhibitory Control Underlies Individual Differences in Older Adults' Hindsight Bias." *Psychology and Aging*, **31**, 224–238. doi:10.1037/pag0000088.

Erdfelder E, Auer T, Hilbig BE, Aßfalg A, Moshagen M, Nadarevic L (2009). "Multinomial Processing Tree Models: A Review of the Literature." *Zeitschrift für Psychologie*, **217**, 108–124. doi:10.1027/0044-3409.217.3.108.

Fagan JF (1984). "Recognition Memory and Intelligence." *Intelligence*, **8**, 31–36. doi:10.1016/0160-2896(84)90004-7.

Frick H, Strobl C, Zeileis A (2014). "To Split or to Mix? Tree vs. Mixture Models for Detecting Subgroups." In M Gilli, G González-Rodríguez, A Nieto-Reyes (eds.), *COMPSTAT 2014 – Proceedings in Computational Statistics*, pp. 379–386. The International Statistical Institute/International Association for Statistical Computing, Geneva, Switzerland.

Grizzle JE, Starmer CF, Koch GG (1969). "Analysis of Categorical Data by Linear Models." *Biometrics*, **25**, 489–504. doi:10.2307/2528901.

Hothorn T, Zeileis A (2015). "partykit: A Modular Toolkit for Recursive Partytioning in R." *Journal of Machine Learning Research*, **16**, 3905–3909.

Hu X, Batchelder WH (1994). "The Statistical Analysis of General Processing Tree Models with the EM Algorithm." *Psychometrika*, **59**, 21–47. doi:10.1007/bf02294263.

Klauer KC (2006). "Hierarchical Multinomial Processing Tree Models: A Latent-Class Approach." *Psychometrika*, **71**, 7–31. doi:10.1007/s11336-004-1188-3.

Klauer KC (2010). "Hierarchical Multinomial Processing Tree Models: A Latent-Trait Approach." *Psychometrika*, **75**, 70–98. doi:10.1007/s11336-009-9141-0.

Komboz B, Strobl C, Zeileis A (2016). "Tree-Based Global Model Tests for Polytomous Rasch Models." *Educational and Psychological Measurement*. Forthcoming.

Matzke D, Dolan CV, Batchelder WH, Wagenmakers EJ (2015). "Bayesian Estimation of Multinomial Processing Tree Models with Heterogeneity in Participants and Items." *Psychometrika*, **80**, 205–235. doi:10.1007/s11336-013-9374-9.

Merkle EC, Fan J, Zeileis A (2014). "Testing for Measurement Invariance with Respect to an Ordinal Variable." *Psychometrika*, **79**, 569–584. doi:10.1007/s11336-013-9376-7.

Merkle EC, Zeileis A (2013). "Tests of Measurement Invariance without Subgroups: A Generalization of Classical Methods." *Psychometrika*, **78**, 59–82. doi:10.1007/s11336-012-9302-4.

Michalkiewicz M, Arden K, Erdfelder E (2016a). "Do Smarter People Make Better Decisions? The Influence of Intelligence on Adaptive Use of the Recognition Heuristic." Manuscript under review.

Michalkiewicz M, Coolin A, Erdfelder E (2013). "Individual Differences in Use of the Recognition Heuristic." In *MathPsych 2013 Program*. Society of Mathematical Psychology, Potsdam, Germany, August 4–7.

Michalkiewicz M, Minich B, Erdfelder E (2016b). "Explaining Individual Differences in Fast-and-Frugal Decision Making: The Impact of Need for Cognition and Faith in Intuition on Use of the Recognition Heuristic." Manuscript under review.

Mirkin B (2001). "Eleven Ways to Look at Chi-Squared Coefficients for Contingency Tables." *The American Statistician*, **55**, 111–120. doi:10.1198/000313001750358428.

Oravecz Z, Anders R, Batchelder WH (2015). "Hierarchical Bayesian Modeling for Test Theory without an Answer Key." *Psychometrika*, **80**, 341–364. doi:10.1007/s11336-013-9379-4.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Riefer DM, Batchelder WH (1988). "Multinomial Modeling and the Measurement of Cognitive Processes." *Psychological Review*, **95**, 318–339. doi:10.1037/0033-295x.95.3.318.

Riefer DM, Batchelder WH (1991). "Age Differences in Storage and Retrieval: A Multinomial Modeling Analysis." *Bulletin of the Psychonomic Society*, **29**, 415–418. doi:10.3758/bf03333957.

Riefer DM, Knapp BR, Batchelder WH, Bamber D, Manifold V (2002). "Cognitive Psychometrics: Assessing Storage and Retrieval Deficits in Special Populations with Multinomial Processing Tree Models." *Psychological Assessment*, **14**, 184–201. doi:10.1037/1040-3590.14.2.184.

Smith JB, Batchelder WH (2010). "Beta-MPT: Multinomial Processing Tree Models for Addressing Individual Differences." *Journal of Mathematical Psychology*, **54**, 167–183. doi:10.1016/j.jmp.2009.06.007.

Strobl C, Kopf J, Zeileis A (2015). "Rasch Trees: A New Method for Detecting Differential Item Functioning in the Rasch Model." *Psychometrika*, **80**, 289–316. doi:10.1007/s11336-013-9388-3.

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**, 135–153. doi:10.3102/1076998609359791.

Swets JA (1961). "Is There a Sensory Threshold?" *Science*, **134**, 168–177. doi:10.1126/science.134.3473.168.

Wickelmaier F, Zeileis A (2011). "Multinomial Processing Tree Models in R." Presented at useR! – The R User Conference, August 16–18, Coventry, UK, URL https://www.R-project.org/conferences/useR-2011/TalkSlides/Contributed/17Aug_1705_FocusV_3-Psychometrics_1-Wickelmaier.pdf.

Zeileis A, Hornik K (2007). "Generalized M-Fluctuation Tests for Parameter Instability." *Statistica Neerlandica*, **61**, 488–508. doi:10.1111/j.1467-9574.2007.00371.x.

Zeileis A, Hothorn T, Hornik K (2008). "Model-Based Recursive Partitioning." *Journal of Computational and Graphical Statistics*, **17**, 492–514. doi:10.1198/106186008x319331.

Zeileis A, Strobl C, Wickelmaier F, Komboz B, Kopf J (2016). *psychotools: Infrastructure for Psychometric Modeling*. R package version 0.4-2, URL https://CRAN.R-project.org/package=psychotools.

Zeileis A, Strobl C, Wickelmaier F, Komboz B, Kopf J (2016a). *psychotree: Recursive Partitioning Based on Psychometric Models*. R package version 0.15-1, URL https://CRAN.R-project.org/package=psychotree.

Affiliation:

Florian Wickelmaier
Department of Psychology
Universität Tübingen
Schleichstr. 4
72076 Tübingen, Germany
E-mail: Florian.Wickelmaier@uni-tuebingen.de
URL: http://homepages.uni-tuebingen.de/florian.wickelmaier/

Achim Zeileis
Department of Statistics
Faculty of Economics and Statistics
Universität Innsbruck
Universitätsstr. 15
6020 Innsbruck, Austria
E-mail: Achim.Zeileis@R-project.org
URL: http://eeecon.uibk.ac.at/~zeileis/

University of Innsbruck - Working Papers in Economics and Statistics
Recent Papers can be accessed on the following webpage:

http://eeecon.uibk.ac.at/wopec/

2016-26 **Florian Wickelmaier, Achim Zeileis:** Using recursive partitioning to account for parameter heterogeneity in multinomial processing tree models

2016-25 **Michel Philipp, Carolin Strobl, Jimmy de la Torre, Achim Zeileis:** On the estimation of standard errors in cognitive diagnosis models

2016-24 **Florian Lindner, Julia Rose:** No need for more time: Intertemporal allocation decisions under time pressure

2016-23 **Christoph Eder, Martin Halla:** The long-lasting shadow of the allied occupation of Austria on its spatial equilibrium

2016-22 **Christoph Eder:** Missing men: World War II casualties and structural change

2016-21 **Reto Stauffer, Jakob Messner, Georg J. Mayr, Nikolaus Umlauf, Achim Zeileis:** Ensemble post-processing of daily precipitation sums over complex terrain using censored high-resolution standardized anomalies

2016-20 **Christina Bannier, Eberhard Feess, Natalie Packham, Markus Walzl:** Incentive schemes, private information and the double-edged role of competition for agents

2016-19 **Martin Geiger, Richard Hule:** Correlation and coordination risk

2016-18 **Yola Engler, Rudolf Kerschbamer, Lionel Page:** Why did he do that? Using counterfactuals to study the effect of intentions in extensive form games

2016-17 **Yola Engler, Rudolf Kerschbamer, Lionel Page:** Guilt-averse or reciprocal? Looking at behavioural motivations in the trust game

2016-16 **Esther Blanco, Tobias Haller, James M. Walker:** Provision of public goods: Unconditional and conditional donations from outsiders

2016-15 **Achim Zeileis, Christoph Leitner, Kurt Hornik:** Predictive bookmaker consensus model for the UEFA Euro 2016

2016-14 **Martin Halla, Harald Mayr, Gerald J. Pruckner, Pilar García-Gómez:** Cutting fertility? The effect of Cesarean deliveries on subsequent fertility and maternal labor supply

2016-13 **Wolfgang Frimmel, Martin Halla, Rudolf Winter-Ebmer:** How does parental divorce affect children's long-term outcomes?

2016-12 **Michael Kirchler, Stefan Palan:** Immaterial and monetary gifts in economic transactions. Evidence from the field

2016-11 **Michel Philipp, Achim Zeileis, Carolin Strobl:** A toolkit for stability assessment of tree-based learners

2016-10 **Loukas Balafoutas, Brent J. Davis, Matthias Sutter:** Affirmative action or just discrimination? A study on the endogenous emergence of quotas *forthcoming in Journal of Economic Behavior and Organization*

2016-09 **Loukas Balafoutas, Helena Fornwagner:** The limits of guilt

2016-08 **Markus Dabernig, Georg J. Mayr, Jakob W. Messner, Achim Zeileis:** Spatial ensemble post-processing with standardized anomalies

2016-07 **Reto Stauffer, Jakob W. Messner, Georg J. Mayr, Nikolaus Umlauf, Achim Zeileis:** Spatio-temporal precipitation climatology over complex terrain using a censored additive regression model

2016-06 **Michael Razen, Jürgen Huber, Michael Kirchler:** Cash inflow and trading horizon in asset markets

2016-05 **Ting Wang, Carolin Strobl, Achim Zeileis, Edgar C. Merkle:** Score-based tests of differential item functioning in the two-parameter model

2016-04 **Jakob W. Messner, Georg J. Mayr, Achim Zeileis:** Non-homogeneous boosting for predictor selection in ensemble post-processing

2016-03 **Dietmar Fehr, Matthias Sutter:** Gossip and the efficiency of interactions

2016-02 **Michael Kirchler, Florian Lindner, Utz Weitzel:** Rankings and risk-taking in the finance industry

2016-01 **Sibylle Puntscher, Janette Walde, Gottfried Tappeiner:** Do methodical traps lead to wrong development strategies for welfare? A multilevel approach considering heterogeneity across industrialized and developing countries

2015-16 **Niall Flynn, Christopher Kah, Rudolf Kerschbamer:** Vickrey Auction vs BDM: Difference in bidding behaviour and the impact of other-regarding motives

2015-15 **Christopher Kah, Markus Walzl:** Stochastic stability in a learning dynamic with best response to noisy play

2015-14 **Matthias Siller, Christoph Hauser, Janette Walde, Gottfried Tappeiner:** Measuring regional innovation in one dimension: More lost than gained?

2015-13 **Christoph Hauser, Gottfried Tappeiner, Janette Walde:** The roots of regional trust

2015-12 **Christoph Hauser:** Effects of employee social capital on wage satisfaction, job satisfaction and organizational commitment

2015-11 **Thomas Stöckl:** Dishonest or professional behavior? Can we tell? A comment on: Cohn et al. 2014, Nature 516, 86-89, "Business culture and dishonesty in the banking industry"

2015-10 **Marjolein Fokkema, Niels Smits, Achim Zeileis, Torsten Hothorn, Henk Kelderman:** Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees

2015-09 **Martin Halla, Gerald Pruckner, Thomas Schober:** The cost-effectiveness of developmental screenings: Evidence from a nationwide programme *forthcoming in Journal of Health Economics*

2015-08 **Lorenz B. Fischer, Michael Pfaffermayr:** The more the merrier? Migration and convergence among European regions

2015-07 **Silvia Angerer, Daniela Glätzle-Rützler, Philipp Lergetporer, Matthias Sutter:** Cooperation and discrimination within and across language borders: Evidence from children in a bilingual city *forthcoming in European Economic Review*

2015-06 **Martin Geiger, Wolfgang Luhan, Johann Scharler:** When do Fiscal Consolidations Lead to Consumption Booms? Lessons from a Laboratory Experiment *forthcoming in Journal of Economic Dynamics and Control*

2015-05 **Alice Sanwald, Engelbert Theurl:** Out-of-pocket payments in the Austrian healthcare system - a distributional analysis

2015-04 **Rudolf Kerschbamer, Matthias Sutter, Uwe Dulleck:** How social preferences shape incentives in (experimental) markets for credence goods *forthcoming in Economic Journal*

2015-03 **Kenneth Harttgen, Stefan Lang, Judith Santer:** Multilevel modelling of child mortality in Africa

2015-02 **Helene Roth, Stefan Lang, Helga Wagner:** Random intercept selection in structured additive regression models

2015-01 **Alice Sanwald, Engelbert Theurl:** Out-of-pocket expenditures for pharmaceuticals: Lessons from the Austrian household budget survey

University of Innsbruck

Working Papers in Economics and Statistics

2016-26

Florian Wickelmaier, Achim Zeileis

Using recursive partitioning to account for parameter heterogeneity in multinomial processing tree models

Abstract
In multinomial processing tree (MPT) models, individual differences between the participants in a study lead to heterogeneity of the model parameters. While subject covariates may explain these differences, it is often unknown in advance how the parameters depend on the available covariates, that is, which variables play a role at all, interact, or have a nonlinear influence, etc. Therefore, a new approach for capturing parameter heterogeneity in MPT models is proposed based on the machine learning method MOB for model-based recursive partitioning. This recursively partitions the covariate space, leading to an MPT tree with subgroups that are directly interpretable in terms of effects and interactions of the covariates. The pros and cons of MPT trees as a means of analyzing the effects of covariates in MPT model parameters are discussed based on a simulation experiment as well as on two empirical applications from memory research. Software that implements MPT trees is provided via the mpttree function in the psychotree package in R.

ISSN 1993-4378 (Print)
ISSN 1993-6885 (Online)