DEPARTMENT OF STATISTICS

LUDWIG–MAXIMILIANS–UNIVERSITY MUNICH

MASTER THESIS

November 3, 2013

Modelling Comparison Data with Ordinal Response

Author:	Giuseppe Casalicchio
Supervisor:	Prof. Dr. Gerhard Tutz
Advisor:	Gunther Schauberger

Abstract

In this thesis, we present several extensions of the Bradley-Terry-Luce model, which is known as a pair comparison model. The aim of pair comparisons is to elicit an overall ranking for a set of objects that are compared pairwise by judges, or, as we will call them, 'subjects'. The presented extensions allow for ordinal responses and the inclusion of subject-specific covariates as well as order effects (or object-specific order effects). If an order effect is present, there is an advantage or disadvantage in a pair comparison for the object that is presented first. In the context of sport competitions, this effect is equivalent to having a home advantage (or a team-specific home advantage). The inclusion of subject-specific covariates in a Bradley-Terry-Luce model may lead to an over-parameterized model. This is because subject-specific covariates are included in the model as subject-object interactions, and for each additional subject-specific covariate, we need to estimate as many subject-object parameters as there are object parameters. To tackle this problem, we suggest a component-wise boosting algorithm that is able to select the most important subject-object interactions. The performance of this algorithm is investigated through a simulation study. This thesis comes along with an R-Package called `ordBTL` that can be used to fit the proposed models and extensions. Lastly, we will illustrate some extensions by applying them to different datasets.

Contents

List of Notations

numbers and indices

$q = \lfloor \frac{K-1}{2} \rfloor$	number of threshold parameters that have to be estimated
$m = 1, \ldots, M$	objects
$i = 1, \ldots, I$	subjects
$k = 1, \ldots, K$	response categories
$p = 1, \ldots, P$	subject-specific covariates
$c = 1, \ldots, C = P \cdot (M-1)$	candidate covariates (also referred to as subject-object interactions)
$b = 1, \ldots, b_{\text{stop}}$	boosting iterations
r, s	indices referring to object r and object s, respectively

parameters

$\boldsymbol{\theta}$	threshold parameter
$\boldsymbol{\gamma}$	object parameter
$\boldsymbol{\lambda}$	parameter for candidate covariates (also referred to as subject-object interaction parameter)
$\boldsymbol{\beta}^\top = (\boldsymbol{\theta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\lambda}^\top)$	parameter vector containing all parameters
$\boldsymbol{\alpha}$	parameter for home advantages or order effects
$\boldsymbol{\eta}$	linear predictor
π	worth parameter
τ	Kendall's τ rank correlation coefficient
Φ	dispersion parameter

vectors and matrices

$\mathbf{x}_i^\top = (x_{i1}, \ldots, x_{iP})$ vector of subject-specific covariates for subject i

$\left(\mathbf{x}^{(r,s)}\right)^\top = \left(x_1^{(r,s)}, \ldots, x_{M-1}^{(r,s)}\right)$ vector of object indicators when the pair (r,s) is compared

$\left(\mathbf{x}_{iC}^{(r,s)}\right)^\top = \left(x_{i1}^{(r,s)}, \ldots, x_{iC}^{(r,s)}\right)$ vector of candidate covariates for subject i when the pair (r,s) is compared $(x_{ic}^{(r,s)} = x_{ip} x_m^{(r,s)})$

$\left(\mathbf{h}^{(r,s)}\right)^\top = \left(h_1^{(r,s)}, \ldots, h_M^{(r,s)}\right)$ vector of dummy variables indicating object-specific order effects (or team-specific home advantages)

$\mathbf{Q}_{(K-1)\times q}$ constraint matrix (for the thresholds)

$\mathbf{X}_O^{(r,s)}$ comparison-specific design matrix when comparing the pair (r,s) once

$\mathbf{X}_{iO}$ design matrix containing all $\mathbf{X}_O^{(r,s)}$ for all pairs (r,s) that are compared by subject i

$\mathbf{X}_O$ design matrix containing $\mathbf{X}_{iO}$, $\forall i$ (contains all comparisons made by all subjects)

$\mathbf{X}_{iC}^{(r,s)} = \left[\mathbf{1}_{K-1} \otimes \left(\mathbf{x}_{iC}^{(r,s)}\right)^\top\right]$ matrix of candidate covariates for subject i when comparing the pair (r,s)

$\mathbf{X}_C = [\mathbf{x}_{\cdot 1}, \ldots, \mathbf{x}_{\cdot C}]$ matrix containing candidate covariates (contains all $\mathbf{X}_{iC}^{(r,s)}$ $\forall r \neq s, i$)

$\mathbf{X}_i^{(r,s)} = \left[\mathbf{X}_O^{(r,s)}, \mathbf{X}_{iC}^{(r,s)}\right]$ design matrix containing object indicators and candidate covariates for a comparison of the pair (r,s) made by subject i

$\mathbf{X} = [\mathbf{X}_O, \mathbf{X}_C]$ complete design matrix with object indicators and candidate covariates for all comparisons and all subjects

miscellaneous

X_m latent variable referring to object m

$\epsilon_m \sim iid$ random variable referring to object m

$\lfloor x \rfloor = \max\{z \in \mathbb{Z} \mid z \leq x\}$ floor function

$F(\cdot), F^{-1}(\cdot)$ cumulative distribution function and corresponding link function

1. Introduction

The present thesis concentrates on modelling pair comparison data with ordinal response. In general, pair comparison data is produced when a certain number of objects are compared pairwise in order to obtain an overall preference ranking of the objects. We will consider *complete pair comparisons* in which each object is compared to all other objects at least once. In many application fields, such as in marketing research or in psychometric experiments, the objects are presented in a pairwise manner to judges, or, as we call them in this thesis, 'subjects'. Their task for each comparison is to make a preference decision for one of the two presented objects according to specific (subjective) criteria (e.g. the fragrance of perfumes when two perfumes are the objects being compared). This preference decision is called the 'response' and can be binary, meaning that the subjects could either prefer the first object or the second object. Alternatively, the subjects can make their preference decisions based on a (symmetric) Likert scale with more than two categories, such as preferring the first object strongly or weakly, preferring neither of the objects, or preferring the second object strongly or weakly. This yields an ordinal response that allows a more precise preference ranking of the objects because we use more information about how strongly an object is preferred (Tutz, 1986). A slightly different application field can be found in (sport) competitions. Here, the two competing teams or players correspond to the objects that are being compared and for which a preference ranking is of interest. In this context, the 'response' corresponds to the outcome of the competition (e.g. win, tie, or lose). Thus, there is a vast range of application fields for pair comparisons.

The main aim of this thesis is first to describe existing pair comparison models that allow for ordinal responses and then to develop an algorithm that allows for variable selection in a model that contains subject-specific covariates. These covariates are, for example, personal characteristics of the subject, which are assumed to play a role in the decision of a subject. Thus, selecting relevant subject-specific covariates might reveal important personal characteristics that are specific for preferring certain objects.

This thesis is structured as follows: Chapter 2 introduces the basic Bradley-Terry-Luce model, which is a pair comparison model for binary responses. In Chapter 3, we present extensions of the Bradley-Terry-Luce model that allow for ordinal responses. The subsequent Chapter 4 explains how subject-specific covariates and order effects can be included in the presented models. For a better understanding of the data structure, we introduce a matrix notation in Chapter 5, which is then used in Chapter 6 within the proposed boosting algorithm that implicitly allows variable selection. To investigate the performance of this algorithm, a simulation study was conducted in Chapter 7. Finally, some of the proposed models and the boosting algorithm are applied on real data sets in Chapter 8. For the models fitted in this thesis, we used the statistical programming language R (R Core Team, 2013) and the additional ordBTL package (Casalicchio, 2013) that was developed during this thesis. It is available on http://cran.r-project.org/web/packages/ordBTL/.

2. The Bradley-Terry-Luce Model

2.1. The Model

One of the most widely used models for comparisons of M different objects in a pairwise manner is the model suggested by Bradley and Terry (1952). It is closely related to the choice axiom of Luce (1959) with the restriction to choices being between two objects. Thus, the model is named the *Bradley-Terry-Luce (BTL) model*.

We consider a comparison of the pair (r, s), which represents the comparison between object r and object s. The response of this comparison is denoted by Y_{rs}, where $Y_{rs} = 1$ indicates the preference for object r and $Y_{rs} = 2$ indicates the preference for object s. An example for pair comparisons are sport competitions in which a match of team r (or player r) and team s (or player s) corresponds to the comparison of the pair (r, s). In this case, the preference for team r (denoted by $Y_{rs} = 1$) can be interpreted as *team r wins*, whereas the preference for team s (denoted by $Y_{rs} = 2$) implies that team r is not preferred and thus can be interpreted as *team r loses*. The BTL model for such a binary response specifies the probability for preferring object r over object s by

$$\mathbb{P}(Y_{rs} = 1 | (r, s)) = \frac{\pi_r}{\pi_r + \pi_s}, \qquad r, s \in \{1, \dots, M\}, \ r \neq s. \tag{2.1}$$

Here, the π_m's are positive-valued *worth parameters* and represent the *worth* or *ability* of the corresponding object m on a continuous preference scale (Hatzinger and Dittrich, 2012).

Alternatively, model (2.1) can be written as a logistic model. For this purpose, we first apply the logit function $\text{logit}(x) = \log\left(\frac{x}{1-x}\right)$ on both sides of equation (2.1), which yields

$$\text{logit}(\mathbb{P}(Y_{rs} = 1 | (r, s))) = \log\left(\frac{\pi_r/(\pi_r + \pi_s)}{1 - \pi_r/(\pi_r + \pi_s)}\right) = \log\left(\frac{\pi_r}{\pi_s}\right) = \log(\pi_r) - \log(\pi_s).$$

In the next step, we use the re-parameterization $\gamma_m = \log(\pi_m)$, $\forall m \in \{1, \ldots, M\}$ with the restriction $\gamma_M = 0$ (which is necessary for identifiability), such that object M is considered as reference object. For a given pair comparison of (r, s), the logistic model can then be written as

$$
\begin{aligned}
\text{logit}(\mathbb{P}(Y_{rs} = 1|(r,s))) \quad &= \gamma_r - \gamma_s \\
&= x_1^{(r,s)} \gamma_1 + \ldots + x_{M-1}^{(r,s)} \gamma_{M-1} \\
&= \left(\mathbf{x}^{(r,s)} \right)^\top \boldsymbol{\gamma},
\end{aligned}
\tag{2.2}
$$

where

$$
x_m^{(r,s)} = \begin{cases} +1 & \text{if } m = r \\ -1 & \text{if } m = s \\ 0 & \text{otherwise} \end{cases}
\tag{2.3}
$$

is considered as an independent variable that indicates if object m is involved in a given pair comparison. Thus, we call $x_m^{(r,s)}$ the object indicator for object m when the pair (r, s) is compared. The object parameter γ_m reflects, in this case, the logarithm of the worth parameter for object m. Since the logarithm is a monotonic transformation, the object parameter γ_m also reflects the worth or ability for object m.

If independence for all comparisons is assumed, ordinary methods for logistic models (such as maximum likelihood estimation) can be used to fit the BTL model (Turner and Firth, 2010; Agresti, 2002, p. 436).

2.2. Generalisation

A logistic model, such as the one in equation (2.2), is a generalized linear model, where the logit function is used as the link function. In general, the logit link can be replaced with an arbitrary link function F^{-1}, where F is the corresponding distribution function. According to David (1988), this yields the following linear model

$$
\begin{aligned}
\mathbb{P}(Y_{rs} = 1|(r,s)) &= F(\gamma_r - \gamma_s) \\
\Leftrightarrow \quad F^{-1}(\mathbb{P}(Y_{rs} = 1|(r,s))) &= \gamma_r - \gamma_s,
\end{aligned}
\tag{2.4}
$$

which can be derived by assuming an underlying latent variable for the objects. For the derivation, we introduce a latent variable X_m for each object $m \in \{1, \ldots, M\}$ that has the form

$$
X_m = \gamma_m + \epsilon_m,
\tag{2.5}
$$

where $\gamma_m \in \mathbb{R}$ is a constant and $\epsilon_m \sim iid$ is a random variable. The differences $\epsilon_s - \epsilon_r$ are assumed to be symmetrically distributed around zero with the cumulative distribution function F. If $X_r > X_s$, then object r is preferred over object s and the probability of this event is given by

$$
\begin{aligned}
\mathbb{P}(Y_{rs} = 1|(r,s)) = \mathbb{P}(X_r > X_s) \ &= \mathbb{P}(\gamma_r + \epsilon_r > \gamma_s + \epsilon_s) \\
&= \mathbb{P}(\epsilon_s - \epsilon_r < \gamma_r - \gamma_s) \qquad (2.6) \\
&= F(\gamma_r - \gamma_s).
\end{aligned}
$$

Since this is equivalent to $F^{-1}(\mathbb{P}(Y_{rs} = 1|(r,s))) = \gamma_r - \gamma_s$, it is obvious that if we use the logistic distribution function with the corresponding link function $F^{-1} = \text{logit}$, we will return to the BTL model (2.2). It should be mentioned briefly that the Thurstone-Mosteller model (Thurstone, 1927; Mosteller, 1951), which is also a pair comparison model, uses the standard normal distribution function $F = \Phi$ (with the corresponding link function $F^{-1} = \Phi^{-1}$) instead of the logistic distribution function. Thus, the BTL model and the Thurstone-Mosteller model can be seen as special cases of the model (2.4).

3. Extensions for Ordinal Responses

3.1. Motivation

In the previous chapter, we mentioned the example of sport competitions with a response consisting of two categories. Now we will consider sport competitions that can also end in a draw, such as football matches. Thus, we have to take into account a third category that allows for a draw in a competition. This yields a response that consists of three categories and measures the three possible outcomes for a match between team r and team s: $Y_{rs} = 1$ if team r wins, $Y_{rs} = 2$ if the match ends in a draw and $Y_{rs} = 3$ if team r loses. In general, ordinal responses with more than three categories are also conceivable. A response consisting of five categories, for example, could be: *strong preference for object r*, *weak preference for object r*, *no preference*, *weak preference for object s* and *strong preference for object s*. Without further assumptions, the method described in Chapter 2 cannot be applied to such an ordinal response.

In the following sections, we tackle this problem using the *cumulative logit model* and the *adjacent categories logit model*, which are known as multivariate regression models for ordinal responses and allow parameter estimation via maximum likelihood estimation (see Agresti, 1992; Fahrmeir et al., 1994, p. 73 - 99). These two models can be seen as a generalisation of the BTL model. Thus, many other extensions of the BTL model that allow for responses with three categories can be assigned to the multivariate regression models mentioned above (see Section 3.5).

3.2. Symmetry Condition

The ordinal response with $K \geq 2$ categories is denoted by $Y_{rs} \in \{1, \ldots, K\}$ and is assumed to be symmetric, meaning that $Y_{rs} = k \Leftrightarrow Y_{sr} = K - k + 1$ and thus

$$\mathbb{P}(Y_{rs} = k) = \mathbb{P}(Y_{sr} = K - k + 1), \qquad \forall k = 1, \ldots, K. \tag{3.1}$$

To fix the notation, we let lower response categories indicate the preference for object r. Thus, $Y_{rs} = 1$ is the most favourable response category for object r and $Y_{rs} = K$ is the most favourable response category for object s, or, equivalently, the least favourable response category for object r. The symmetry condition (3.1) can then be interpreted as *the probability that object r is preferred over object s with strength k is equal to the probability that object s is preferred over object r with the corresponding strength $K - k + 1$* (Tutz, 1986; Agresti, 1992).

3.3. Cumulative Logit Model

3.3.1. The Model

The cumulative logit model for an ordinal response denoted by $Y_{rs} \in \{1, \ldots, K\}$ is based on the consideration of cumulative probabilities $\mathbb{P}(Y_{rs} \leq k) = \mathbb{P}(Y_{rs} = 1) + \ldots + \mathbb{P}(Y_{rs} = k)$. The logarithm of cumulative odds is called cumulative logit and can be written as

$$\mathrm{logit}(\mathbb{P}(Y_{rs} \leq k)) = \log\left(\frac{\mathbb{P}(Y_{rs} \leq k)}{\mathbb{P}(Y_{rs} > k)}\right), \qquad k = 1, \ldots, K - 1. \tag{3.2}$$

It measures how likely the response Y_{rs} is in category k or below, compared to a category higher than k (Fahrmeir et al., 1994, p. 76-78).

In a comparison of the pair (r, s), the BTL model for ordinal responses proposed by Tutz (1986) can be written as *cumulative logit model*

$$\begin{aligned} \mathrm{logit}(\mathbb{P}(Y_{rs} \leq k | (r, s))) \quad &= \theta_k + (\gamma_r - \gamma_s) \\ &= \theta_k + (x_1^{(r,s)}\gamma_1 + \ldots + x_{m-1}^{(r,s)}\gamma_{m-1}) \\ &= \theta_k + \left(\mathbf{x}^{(r,s)}\right)^\top \boldsymbol{\gamma}, \end{aligned} \tag{3.3}$$

where $-\infty = \theta_0 < \theta_1 < \ldots < \theta_{K-1} < \theta_K = \infty$ are the thresholds (Tutz, 1989, p. 110-114). Due to the symmetry condition, we need the following additional restrictions for the thresholds (for proof, see Tutz (1986)):

(1) If K is odd

$$\theta_k = -\theta_{K-k} \quad \text{for} \quad k = 1, \ldots, \tfrac{K-1}{2} = \left\lfloor \tfrac{K-1}{2} \right\rfloor \tag{3.4}$$

(2) If K is even

$$\theta_{K/2} = 0 \quad \text{and} \quad \theta_{K/2-k} = -\theta_{K/2+k} \quad \text{for} \quad k = 1, \ldots, \tfrac{K}{2} - 1 = \left\lfloor \tfrac{K-1}{2} \right\rfloor \quad (3.5)$$

From these restrictions, we can see that if $K > 2$, then only $q = \left\lfloor \frac{K-1}{2} \right\rfloor$ threshold parameters need to be estimated, although the cumulative logit model requires $K-1$ thresholds (apart from $\theta_0 = -\infty$ and $\theta_K = \infty$). If we consider a response with $K = 2$ categories, we have $q = \left\lfloor \frac{2-1}{2} \right\rfloor = \left\lfloor \frac{1}{2} \right\rfloor = 0$ and $\theta_1 = 0$ due to the constraint (3.5). Thus, we do not have to estimate any threshold parameters. In this case, model (3.3) simplifies to

$$\text{logit}(\mathbb{P}(Y_{rs} \leq 1 | (r, s))) = \text{logit}(\mathbb{P}(Y_{rs} = 1 | (r, s))) = \left(\mathbf{x}^{(r,s)} \right)^{\top} \boldsymbol{\gamma}$$

and we return to the BTL model (2.2), which is a special case of the model described in this section.

3.3.2. Generalisation

If we consider the latent variable $X_m = \gamma_m + \epsilon_m$ from equation (2.5) that underlies each object m, we can use the category boundary approach (also known as the threshold approach) to derive the cumulative logit model (3.3) (Tutz, 1989, p. 66-68; Agresti, 1992). Here, we assume that object r is preferred over object s with strength k if $X_s - X_r$ is between θ_{k-1} and θ_k. This statement can be written as

$$Y_{rs} = k \iff \theta_{k-1} < X_s - X_r \leq \theta_k, \qquad \forall k = 1, \ldots, K \qquad (3.6)$$

and means that Y_{rs} is a categorized version of $X_s - X_r$ (Fahrmeir et al., 1994, p. 74). The lower the response category Y_{rs}, the smaller $X_s - X_r$ and the smaller $X_s - X_r$, the more one prefers object r. Thus, lower response categories indicate the preference for object r. As in Chapter 2, we assume that the differences $\epsilon_s - \epsilon_r$ are symmetrically distributed around zero with the cumulative distribution function F. Therefore, it follows that

$$
\begin{aligned}
\mathbb{P}(Y_{rs} \leq k | (r, s)) \overset{(3.6)}{\iff} \mathbb{P}(X_s - X_r \leq \theta_k) &= \mathbb{P}(\gamma_s - \gamma_r + \epsilon_s - \epsilon_r \leq \theta_k) \\
&= \mathbb{P}(\epsilon_s - \epsilon_r \leq \theta_k + (\gamma_r - \gamma_s)) \qquad (3.7) \\
&= F(\theta_k + (\gamma_r - \gamma_s)),
\end{aligned}
$$

where the thresholds are restricted in the same way as defined in the previous section. Model (3.7) corresponds to the *cumulative link model* supposed by Agresti (1992) and is a generalisation of the cumulative logit model (3.3). This is because we are allowed to use an arbitrary distribution function F (with $F(x) = 1 - F(-x)$) instead of the logistic distribution function used in model (3.3) (Tutz, 1989, p. 71).

3.4. Adjacent Categories Logit Model

The adjacent categories logit model is defined as

$$
\begin{aligned}
\log\left(\frac{\mathbb{P}(Y_{rs} = k|(r,s))}{\mathbb{P}(Y_{rs} = k+1|(r,s))}\right) &= \theta_k + (\gamma_r - \gamma_s) \\
&= \theta_k + (x_1^{(r,s)}\gamma_1 + \ldots + x_{M-1}^{(r,s)}\gamma_{M-1}) \\
&= \theta_k + \left(\mathbf{x}^{(r,s)}\right)^{\top}\boldsymbol{\gamma},
\end{aligned}
\tag{3.8}
$$

again with the constraints (3.4) and (3.5) for the symmetry of the thresholds. It uses adjacent response categories $\mathbb{P}(Y_{rs} = k)$ and $\mathbb{P}(Y_{rs} = k+1)$ instead of cumulative probabilities $\mathbb{P}(Y_{rs} \leq k|(r,s))$ and $\mathbb{P}(Y_{rs} > k|(r,s))$ as in the cumulative link model (3.3). Thus, the interpretation is directly related to the log-odds for two adjacent response categories (namely $Y_{rs} = k$ and $Y_{rs} = k+1$) instead of the log-odds for two groupings of response categories (which, in the cumulative link model, are determined by $Y_{rs} \leq k$ and $Y_{rs} > k$) (Agresti, 2002, p. 286 - 287).

The adjacent categories logit model reduces to the BTL model (2.2) if the response consists of $K = 2$ categories. In this case, the cumulative logit model and the adjacent categories logit model are equivalent. Since the adjacent categories logit model can also be represented as a log-linear model, one can find many extensions for the BTL model based on its log-linear representation (see Agresti, 1992; Dittrich et al., 1998, 2004, 2007).

3.5. Overview

Various approaches have been proposed to extend the BTL model for allowing ordinal responses. According to Agresti (1992), most of these extensions can be assigned

into two general models: the *cumulative link model* and the *adjacent categories logit model* (see also Figure 3.1).
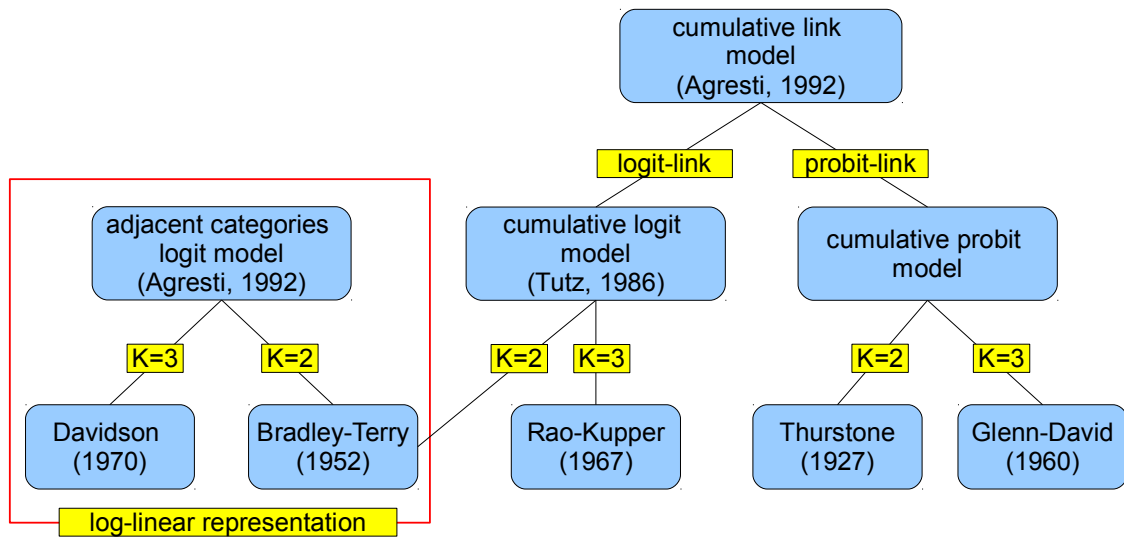

Figure 3.1.: Overview of existing models and their connections.

The *cumulative link model* uses cumulative probabilities in order to model ordinal responses. If the logit link is used as the link function, we obtain the cumulative logit model suggested by Tutz (1986). This again corresponds to the Rao-Kupper model (Rao and Kupper, 1967) if the response consists of $K = 3$ categories, and it is equivalent to the BTL model (Bradley and Terry, 1952) if the response consists of $K = 2$ categories. Using the probit link instead of the logit link yields the Thurstone model if the response consists of $K = 2$ categories, or the model proposed by Glenn and David (1960) if the response consists of $K = 3$ categories.

The *adjacent categories logit model* is based on adjacent response probabilities instead of cumulative probabilities. It corresponds to the BTL model if the response consists of $K = 2$ categories and equates to the model proposed by Davidson (1970) if the response consists of $K = 3$ categories. Dittrich et al. (2004) use a log-linear representation for the BTL model based on the adjacent categories logit model. They also suggest several other extensions for the log-linear BTL model, such as including (subject-specific) covariates or considering order effects. In the next section, we will show how to include these extensions without using a log-linear representation.

4. Extensions for Covariates

The linear predictor for any model presented in the previous chapter is

$$\eta_k^{(r,s)} = \theta_k + \left(\mathbf{x}^{(r,s)}\right)^\top \boldsymbol{\gamma}.$$

The following extensions can be applied on all of these models through an extension or modification of the linear predictor. For illustration purposes, however, we will apply them only on the cumulative logit model (3.3).

4.1. Subject-specific Covariates

According to Dittrich et al. (1998), the preference for a certain object in a pair comparison can depend on personal characteristics of the subject. To include this assumption in the model, we introduce the vector of subject-specific covariates $\mathbf{x}_i^\top = (x_{i1}, \ldots, x_{iP})$ that contains P personal characteristics for subject i. The object parameter γ_m (that reflects the *worth* of object m) is then assumed to depend on the personal characteristics of the subject. To incorporate (binary or metric) subject-specific covariates into the model, we can follow Francis et al. (2002) and express the object parameters as a linear combination of P subject-specific covariates. The object parameter for subject i that depends on P subject-specific covariates is then

$$\gamma_{i(m)} = \gamma_{(m)} + \mathbf{x}_i^\top \boldsymbol{\lambda}_{(m)} = \gamma_{(m)} + \sum_{p=1}^P x_{ip} \lambda_{(m)p}, \tag{4.1}$$

where $\gamma_{(m)}$ is a parameter for object m that is independent of the personal characteristics of the subject, and $\lambda_{(m)p}$ is a modifying effect for object m depending on the p-th subject-specific covariate. According to Francis et al. (2002), we have to constrain $\lambda_{(M)p} = 0$, $\forall p$ so that the model below remains identifiable.

The corresponding model for subject i that considers P subject-specific covariates is then

$$
\begin{aligned}
\text{logit}(\mathbb{P}(Y_{rs} \leq k | (r,s), i)) &= \theta_k + (\gamma_{i(r)} - \gamma_{i(s)}) \\
&= \theta_k + (\gamma_{(r)} - \gamma_{(s)}) + \sum_{p=1}^{P} x_{ip}(\lambda_{(r)p} - \lambda_{(s)p}) \\
&= \theta_k + \sum_{m=1}^{M-1} x_m^{(r,s)} \gamma_{(m)} + \sum_{p=1}^{P} x_{ip} \sum_{m=1}^{M-1} x_m^{(r,s)} \lambda_{(m)p},
\end{aligned}
\tag{4.2}
$$

where the object indicators $x_m^{(r,s)}$ are defined as in equation (2.3) and are used to represent the differences of $\gamma_{(r)} - \gamma_{(s)}$ and $\lambda_{(r)p} - \lambda_{(s)p}$ as a linear combination of all γ's and λ's, respectively. The parameter $\lambda_{(m)p}$ is regarded as a subject-object (interaction) parameter that refers to the m-th object and the p-th subject-specific covariate (Tutz, 1989, p.110-114; Cattelan, 2012).

In the absence of subject-specific covariates, we have $x_{ip} = 0$, $\forall i, p$ and we return to the cumulative logit model without subject-specific covariates as described in Section 3.3. Since subject-specific covariates enter the model as subject-object interactions, we have to estimate $M - 1$ subject-object parameters for each additional subject-specific covariate. This can result in an over-parameterized model. To overcome this problem, we introduce a variable selection procedure in Chapter 6.

4.2. Order Effect and Home Advantage

When comparing two objects, sometimes there can be an advantage (or disadvantage) for the object that is presented first. This advantage refers to the order of the presentation of the objects and is thus called *order effect*. In the context of sport competitions, the order effect is equivalent to a *home advantage*, where the team that plays at its home location has an advantage over the away team (Fahrmeir and Tutz, 1994).

4.2.1. Global Effect

If we consider the pair (r, s), we fix the notation such that object r corresponds to the object that is presented first (or, in the context of sport competitions, the team

that plays at its home location). Thus, the order effect (or home advantage) is only included within the first presented object r, such that the object parameter γ_r can be re-parameterized by $\gamma_r \to \gamma_r + \alpha_0$. This yields the model

$$
\begin{aligned}
\text{logit}(\mathbb{P}(Y_{rs} \leq k | (r,s))) &= \theta_k + \alpha_0 + \gamma_r - \gamma_s \\
&= \theta_k + \alpha_0 + (x_1^{(r,s)}\gamma_1 + \ldots + x_{M-1}^{(r,s)}\gamma_{M-1}) \qquad (4.3) \\
&= \theta_k + \alpha_0 + \left(\mathbf{x}^{(r,s)}\right)^{\top}\boldsymbol{\gamma},
\end{aligned}
$$

where α_0 represents the order effect (or home advantage) that is present in all pair comparisons (Goos and Großmann, 2011).

In some sport competitions, there is the possibility that the teams will play on a neutral field, where neither team r nor team s has a home advantage. This is not considered in the model above because the re-parameterization $\gamma_r \to \gamma_r + \alpha_0$ forces team r to have a home advantage even if the match would be played on a neutral field. In such a case, we can use the re-parameterization $\gamma_r \to \gamma_r + h_0^{(r,s)}\alpha_0$ instead, where $h_0^{(r,s)}$ is a dummy variable that indicates if team r has a home advantage ($h_0^{(r,s)} = 1$) or if it plays on a neutral field without having a home advantage ($h_0^{(r,s)} = 0$). This brings us to the model

$$
\begin{aligned}
\text{logit}(\mathbb{P}(Y_{rs} \leq k | (r,s))) &= \theta_k + h_0^{(r,s)}\alpha_0 + \gamma_r - \gamma_s \\
&= \theta_k + h_0^{(r,s)}\alpha_0 + (x_1^{(r,s)}\gamma_1 + \ldots + x_{M-1}^{(r,s)}\gamma_{M-1}) \qquad (4.4) \\
&= \theta_k + h_0^{(r,s)}\alpha_0 + \left(\mathbf{x}^{(r,s)}\right)^{\top}\boldsymbol{\gamma}.
\end{aligned}
$$

Thus, in general, order effects or home advantages can be seen as comparison-specific or contest-specific covariates, respectively (Harville and Smith, 1994; Turner and Firth, 2010; Masarotto and Varin, 2012).

Note that the models (4.3) and (4.4) are still identifiable due to the symmetry constraints for the thresholds. However, for the model (4.4) care must be taken when having a single round-robin structure for which the pairs (r,s), $\forall r < s$ are compared only once. For example, if we assume that all matches are played on the home field of team 1, only team 1 will have a home advantage and all other matches will be played on a neutral field. Then, because of $r < s$, it follows that object $r \in \{1, \ldots, M-1\}$ and object $s \in \{2, \ldots, M\}$ and thus $s \neq 1$ yielding

$$
x_1^{(r,s)} = \begin{cases} 1 & \text{if } r = 1 \\ 0 & \text{otherwise.} \end{cases} \qquad (4.5)
$$

In such a case, we would have $h_0^{(r,s)} = x_1^{(r,s)}$ for all comparisons $r < s$ and thus the model (4.4) would not be identifiable.

4.2.2. Object-specific or Team-specific Effects

It is also possible to consider object-specific order effects, or equivalently, team-specific home advantages (Harville and Smith, 1994). These effects vary from object to object (or team to team) and are incorporated into model (4.3) using a covariate vector $\left(\mathbf{h}^{(r,s)}\right)^\top = \left(h_1^{(r,s)}, \ldots, h_M^{(r,s)}\right)$, where

$$h_m^{(r,s)} = \begin{cases} 1 & \text{if } m = r \\ 0 & \text{otherwise} \end{cases} \tag{4.6}$$

indicates (when considering the pair (r, s)) if object r or team r has an order effect or home advantage, respectively. The model can then be written as

$$\begin{aligned} \operatorname{logit}(\mathbb{P}(Y_{rs} \leq k | (r, s))) &= \theta_k + (h_1^{(r,s)}\alpha_1 + \ldots + h_M^{(r,s)}\alpha_M) \\ &\quad + (x_1^{(r,s)}\gamma_1 + \ldots + x_{M-1}^{(r,s)}\gamma_{M-1}) \\ &= \theta_k + \left(\mathbf{h}^{(r,s)}\right)^\top \boldsymbol{\alpha} + \left(\mathbf{x}^{(r,s)}\right)^\top \boldsymbol{\gamma}, \end{aligned} \tag{4.7}$$

where the parameter vector $\boldsymbol{\alpha}^\top = (\alpha_1, \ldots, \alpha_M)$ contains the object-specific order effects, or equivalently, the team-specific home advantages. Note that the model described in this section is only identifiable if at least a double round robin structure is present and each object of a fixed pair comparison is allowed to have the object-specific order effect (or equivalently, the team-specific home advantage) once. This is only possible if every pair (r, s) is compared at least twice (e.g. comparing (r, s) and (s, r) giving the order effect once object r and once object s, respectively).

5. Matrix Notation

This chapter is intended to illustrate the structure of the design matrix for pair comparisons with and without considering subject-specific covariates. Having the design matrix at hand, we can use the Fisher scoring algorithm within the framework of multivariate maximum likelihood estimation as described in Fahrmeir et al. (1994, p. 97 - 99) to obtain the parameter estimates. The matrix notation introduced in this chapter will also facilitate the description of the proposed boosting algorithm in the next chapter.

5.1. Constraint Matrix

As mentioned in Section 3.3, we do not have to estimate any threshold parameters for $K = 2$ response categories, whereas for $K > 2$ we have to estimate $q = \left\lfloor \frac{K-1}{2} \right\rfloor$ threshold parameters. Below we consider the case where $K > 2$.

Let $\boldsymbol{\theta}_q^\top = (\theta_1, \ldots, \theta_q)$ denote a vector containing all threshold parameters that have to be estimated and $\boldsymbol{\theta}_{K-1}^\top = (\theta_1, \ldots, \theta_{K-1})$ denote a vector containing all threshold parameters in the model (including those that are determined by the symmetry constraints from equation (3.4) and (3.5)). We then define a constraint matrix $\mathbf{Q}_{(K-1)\times q}$, such that $\mathbf{Q}_{(K-1)\times q} \cdot \boldsymbol{\theta}_q = \boldsymbol{\theta}_{K-1}$ satisfies the symmetry constraints.

For illustration, we will consider a response that consists of $K = 6$ categories. According to constraint (3.5), we have $\boldsymbol{\theta}_{K-1}^\top = \boldsymbol{\theta}_5^\top = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5) = (\theta_1, \theta_2, 0, -\theta_2, -\theta_1)$. We are looking for a constraint matrix $\mathbf{Q}_{5\times 2}$ that satisfies $\mathbf{Q}_{5\times 2} \cdot \boldsymbol{\theta}_2 = \boldsymbol{\theta}_5$, where $\boldsymbol{\theta}_q^\top = \boldsymbol{\theta}_2^\top = (\theta_1, \theta_2)$. This is achieved with the following equation

$$\mathbf{Q}_{5\times 2} \cdot \boldsymbol{\theta}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & -1 \\ -1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ 0 \\ -\theta_2 \\ -\theta_1 \end{pmatrix} = \boldsymbol{\theta}_5, \tag{5.1}$$

which satisfies the constraints from equation (3.5).

Another example with an odd number of response categories, for instance $K = 5$, is presented. According to constraint (3.4), we have $\boldsymbol{\theta}_{K-1}^\top = \boldsymbol{\theta}_4^\top = (\theta_1, \theta_2, \theta_3, \theta_4) = (\theta_1, \theta_2, -\theta_2, -\theta_1)$. Thus, the constraint matrix $\mathbf{Q}_{4\times 2}$ is chosen as follows

$$\mathbf{Q}_{4\times 2} \cdot \boldsymbol{\theta}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \\ -1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ -\theta_2 \\ -\theta_1 \end{pmatrix} = \boldsymbol{\theta}_4. \tag{5.2}$$

To obtain the general structure of $\mathbf{Q}_{(K-1)\times q}$, we use the null vector $\mathbf{0}_q^\top = (0, \dots, 0)$ and the matrices

$$\mathbf{I}_{q\times q} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \text{ and } \mathbf{J}_{q\times q} = \begin{pmatrix} 0 & \cdots & 0 & -1 \\ \vdots & \cdot^{\cdot^{\cdot}} & -1 & 0 \\ 0 & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \vdots \\ -1 & 0 & \cdots & 0 \end{pmatrix}.$$

The constraint matrix is a block matrix and has the form

$$\mathbf{Q}_{(K-1)\times q} = \begin{bmatrix} \mathbf{I}_{q\times q} \\ \mathbf{0}_q^\top \\ \mathbf{J}_{q\times q} \end{bmatrix} \text{ if } K \text{ is even and } \mathbf{Q}_{(K-1)\times q} = \begin{bmatrix} \mathbf{I}_{q\times q} \\ \mathbf{J}_{q\times q} \end{bmatrix} \text{ if } K \text{ is odd.}$$

Using this general structure and considering the case with $K = 6$ response categories, it follows that

$$\mathbf{Q}_{5\times 2} = \begin{bmatrix} \mathbf{I}_{2\times 2} \\ \mathbf{0}_2^\top \\ \mathbf{J}_{2\times 2} \end{bmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & -1 \\ -1 & 0 \end{pmatrix},$$

which corresponds to the constraint matrix from equation (5.1). Consequently, a response consisting of $K = 5$ categories yields the constraint matrix from equation (5.1), namely

$$\mathbf{Q}_{4\times 2} = \begin{bmatrix} \mathbf{I}_{2\times 2} \\ \mathbf{J}_{2\times 2} \end{bmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \\ -1 & 0 \end{pmatrix}.$$

5.2. Design Matrix without Subject-specific Covariates

5.2.1. Single Pair Comparison

For a single comparison of the pair (r, s), we can represent the model without subject-specific covariates in equation (3.2) as

$$\text{logit}(\mathbb{P}(Y_{rs} \leq k|(r,s))) = \mathbf{X}_O^{(r,s)}\boldsymbol{\beta}. \tag{5.3}$$

With the parameter vector

$$\boldsymbol{\beta}^\top = (\boldsymbol{\theta}_q^\top, \boldsymbol{\gamma}^\top) = (\theta_1, \ldots, \theta_q, \gamma_1, \ldots, \gamma_{M-1})$$

and the comparison-specific design matrix

$$\mathbf{X}_O^{(r,s)} = \left[\mathbf{Q}_{(K-1)\times q}, \mathbf{1}_{K-1} \otimes \left(\mathbf{x}^{(r,s)} \right)^\top \right], \tag{5.4}$$

in which the constraint matrix $\mathbf{Q}_{(K-1)\times q}$ and the vector of object indicators $\left(\mathbf{x}^{(r,s)} \right)^\top = \left(x_1^{(r,s)}, \ldots, x_{M-1}^{(r,s)} \right)$ are included. The Kronecker product (denoted by $\otimes$) with the vector $\mathbf{1}_{K-1} = (1, \ldots, 1)^\top$ yields a matrix, in which the object indicators $\left(\mathbf{x}^{(r,s)} \right)^\top$ are repeated $K-1$ times row-wise (see example below). The design matrix is considered as block matrix consisting of two blocks, where the constraint matrix $\mathbf{Q}_{(K-1)\times q}$ forms the first block and the second block contains the object indicators (where each column refers to an object). Thus, the design matrix has $(K-1)$ rows and $(q + (M-1))$ columns.

To illustrate the structure of the comparison-specific design matrix, we consider the case where the response consists of $K = 5$ categories. If only the pair $(1, 2)$ is compared, the design matrix has the form

$$\mathbf{X}_O^{(1,2)} = \left[\mathbf{Q}_{4\times 2}, \mathbf{1}_4 \otimes \left(\mathbf{x}^{(1,2)} \right)^\top \right] = \left[\begin{array}{cc|cccccc} 1 & 0 & 1 & -1 & 0 & \ldots & 0 \\ 0 & 1 & 1 & -1 & 0 & \ldots & 0 \\ 0 & -1 & 1 & -1 & 0 & \ldots & 0 \\ -1 & 0 & 1 & -1 & 0 & \ldots & 0 \end{array} \right],$$

where the first block contains the constraint matrix $\mathbf{Q}_{4\times2}$ and the second block contains a matrix of object indicators $\mathbf{1}_4 \otimes \left(\mathbf{x}^{(1,2)}\right)^\top$, which is repeated for all $K-1 = 4$ rows.

The extension for comparing an arbitrary pair (r,s) is straightforward: in the second block of $\mathbf{X}_O^{(r,s)}$, the 1's are positioned in the r-th column and the -1's in the s-th column. Note that for the pair $(M-1, M)$, the second block is slightly different. This is because, in general, object M is set as the reference object such that the second block has only $M-1$ columns and thus

$$\mathbf{X}_O^{(M-1,M)} = \left[\mathbf{Q}_{4\times2}, \mathbf{1}_4 \otimes \left(\mathbf{x}^{(M-1,M)}\right)^\top\right] = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 1 \\ 0 & 1 & 0 & \dots & 0 & 1 \\ 0 & -1 & 0 & \dots & 0 & 1 \\ -1 & 0 & 0 & \dots & 0 & 1 \end{bmatrix},$$

where in the second block of $\mathbf{X}_O^{(M-1,M)}$ the 1's are positioned in the $(M-1)$-th column.

5.2.2. Complete Pair Comparison

In a complete pair comparison experiment, I subjects and M objects are involved, and each subject has the task of making preference decisions for all $\binom{M}{2}$ comparisons. Here, $\binom{M}{2} = \frac{M!}{2!(M-2)!}$ is read as 'M choose 2' and denotes the binomial coefficient representing the number of all distinct pairs in a set of M objects, namely

$$(1,2),(1,3),\dots,(r,s),\dots,(M-1,M), \forall r < s.$$

The subject-specific design matrix that contains all of these pair comparisons, which were made by a single subject (subject i), has the form

$$\mathbf{X}_{iO} = \begin{bmatrix} \mathbf{X}_O^{(1,2)} \\ \mathbf{X}_O^{(1,3)} \\ \vdots \\ \mathbf{X}_O^{(r,s)} \\ \vdots \\ \mathbf{X}_O^{(M-1,M)} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{(K-1)\times q} & \mathbf{1}_{K-1} \otimes \left(\mathbf{x}^{(1,2)}\right)^\top \\ \mathbf{Q}_{(K-1)\times q} & \mathbf{1}_{K-1} \otimes \left(\mathbf{x}^{(1,3)}\right)^\top \\ \vdots & \vdots \\ \mathbf{Q}_{(K-1)\times q} & \mathbf{1}_{K-1} \otimes \left(\mathbf{x}^{(r,s)}\right)^\top \\ \vdots & \vdots \\ \mathbf{Q}_{(K-1)\times q} & \mathbf{1}_{K-1} \otimes \left(\mathbf{x}^{(M-1,M)}\right)^\top \end{bmatrix}, \forall r < s. \qquad (5.5)$$

The single blocks $\mathbf{X}_O^{(r,s)}$, $\forall r < s$ do not depend on subject i, meaning that all of the subjects have the same design matrix and thus $\mathbf{X}_{1O} = \mathbf{X}_{2O} = \ldots = \mathbf{X}_{IO}$. Thus, the *complete design matrix* containing all of the pair comparisons that were made by any subject is then defined as

$$\mathbf{X}_O = \begin{bmatrix} \mathbf{X}_{1O} \\ \mathbf{X}_{2O} \\ \vdots \\ \mathbf{X}_{IO} \end{bmatrix}. \tag{5.6}$$

5.3. Design Matrix with Subject-specific Covariates

5.3.1. Single Pair Comparison

For estimation purposes and later use in Chapter 6, we define *candidate covariates* of the form $x_{ic}^{(r,s)} = x_{ip} x_m^{(r,s)}$ $\forall p, m$ with $c = 1, \ldots, C = P \cdot (M - 1)$. They serve as a proxy for the subject-object interactions, where each candidate covariate refers to an interaction of the m-th object and the p-th subject-covariate. The model from equation (4.2) can then be expressed as

$$\begin{aligned} \mathrm{logit}(\mathbb{P}(Y_{rs} \leq k | (r,s), i)) &= \theta_k + \sum_{m=1}^{M-1} x_m^{(r,s)} \gamma_{(m)} + \sum_{p=1}^{P} \sum_{m=1}^{M-1} x_{ip} x_m^{(r,s)} \lambda_{(m)p} \\ &= \theta_k + \sum_{m=1}^{M-1} x_m^{(r,s)} \gamma_{(m)} + \sum_{c=1}^{C} x_{ic}^{(r,s)} \lambda_c, \end{aligned} \tag{5.7}$$

where $\lambda_c = \lambda_{(m)p}$ is the associated parameter for the c-th candidate covariate. We introduce the vector $\left(\mathbf{x}_{iC}^{(r,s)} \right)^\top = \left(x_{i1}^{(r,s)}, \ldots, x_{iC}^{(r,s)} \right)$ that contains all of the candidate covariates for subject i when the pair (r, s) is compared and rewrite the model (5.7) as

$$\mathrm{logit}(\mathbb{P}(Y_{rs} \leq k | (r,s), i)) = \theta_k + \left(\mathbf{x}^{(r,s)} \right)^\top \boldsymbol{\gamma} + \left(\mathbf{x}_{iC}^{(r,s)} \right)^\top \boldsymbol{\lambda}. \tag{5.8}$$

The model for a single comparison of the pair (r, s) made by subject i is given in matrix notation as follows

$$\mathrm{logit}(\mathbb{P}(Y_{rs} \leq k | (r,s), i)) = \mathbf{X}_i^{(r,s)} \boldsymbol{\beta}, \tag{5.9}$$

with the parameter vector

$$\boldsymbol{\beta}^\top = (\boldsymbol{\theta}_q^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\lambda}^\top) = (\theta_1, \ldots, \theta_q, \gamma_1, \ldots, \gamma_{M-1}, \lambda_1, \ldots, \lambda_C)$$

and the subject-specific and comparison-specific design matrix

$$\mathbf{X}_i^{(r,s)} = \Big[\underbrace{\mathbf{Q}_{(K-1)\times q}, \mathbf{1}_{K-1} \otimes \left(\mathbf{x}^{(r,s)}\right)^\top}_{\mathbf{X}_O^{(r,s)}}, \underbrace{\mathbf{1}_{K-1} \otimes \left(\mathbf{x}_{iC}^{(r,s)}\right)^\top}_{\mathbf{X}_{iC}^{(r,s)}} \Big].$$

The block matrix $\mathbf{X}_O^{(r,s)}$ from above contains the object indicators and has the same form as in equation (5.4), whereas the matrix $\mathbf{X}_{iC}^{(r,s)}$ contains only the candidate covariates. Thus, the dimension of this matrix is $(K-1) \times (q + (M-1) + C)$.

5.3.2. Complete Pair Comparison

For a complete pair comparison experiment, we have $\binom{M}{2}$ comparisons for each subject. Here, the complete design matrix is a block matrix of the form

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1^{(1,2)} \\ \vdots \\ \mathbf{X}_1^{(M-1,M)} \\ \vdots \\ \mathbf{X}_I^{(1,2)} \\ \vdots \\ \mathbf{X}_I^{(M-1,M)} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_O^{(1,2)} & \mathbf{X}_{1C}^{(1,2)} \\ \vdots & \vdots \\ \mathbf{X}_O^{(M-1,M)} & \mathbf{X}_{1C}^{(M-1,M)} \\ \vdots & \\ \mathbf{X}_O^{(1,2)} & \mathbf{X}_{IC}^{(1,2)} \\ \vdots & \vdots \\ \underbrace{\mathbf{X}_O^{(M-1,M)}}_{\mathbf{X}_O} & \underbrace{\mathbf{X}_{IC}^{(M-1,M)}}_{\mathbf{X}_C} \end{bmatrix}, \tag{5.10}$$

where the block matrix $\mathbf{X}_O$ is defined as in equation (5.6) and the block matrix $\mathbf{X}_C$ contains only the candidate covariates for all of the possible comparisons $r < s$ made by any subject $i = 1, \ldots, I$.

6. Boosting with Implicit Variable Selection

6.1. Introduction

Boosting was developed in the machine learning community to improve classification by combining a set of (iteratively learned) weak learners to create a single, strong learner (Schapire, 1990; Freund, 1990). In a joint work, Freund and Schapire (1997) introduced the well-known AdaBoost algorithm, which can be seen as a weighted refitting of observations based on a binary classificator (that represents the weak learner). In each boosting iteration of the AdaBoost algorithm, a weak learner is fitted to the weighted data, wherein the weights for misclassified observations increase from iteration to iteration. Thus, one concentrates more on misclassified observations. Finally, a linear combination of the weak learners from each iteration is used, which leads to a strong learner.

The idea to use weak learners and 'boost' them to gain a strong learner has also received a great deal of attention in statistics. The breakthrough, from a statistical point of view, came with the reinterpretation of AdaBoost as an optimization algorithm, namely as 'functional gradient descent' that minimizes an exponential loss within the framework of 'forward stage-wise additive modelling' (Friedman et al., 2000; Friedman, 2001). When using a different loss function, such as the squared error loss, this can be seen as iteratively 'fitting of residuals' and allows for an extension of boosting to regression problems. Based on this ground breaking discovery, Bühlmann and Yu (2003) proposed a 'component-wise boosting' procedure that implies variable selection and does a fitting of the residuals using only one (or a subset of) covariates in each iteration. Tutz et al. (2006) use a similar approach, but instead of minimizing a specific loss function with functional gradient descent, they directly maximize the log-likelihood using one-step Fisher scoring to obtain a weak learner (this is known as likelihood-based boosting). Further development of the component-wise boosting was done by Tutz and Binder (2007), who proposed a procedure that allows for distinguishing between obligatory covariates (that have to be included in the model) and optional covariates (that might be of relevance).

6.2. Algorithm

The following boosting algorithm uses the terminology introduced in Chapter 5 and is a (likelihood-based) component-wise boosting procedure based on the `pomBoost` algorithm (Zahid and Tutz, 2011), which is able to handle both obligatory and optional covariates, and allows for ordinal responses (since it is based on the cumulative logit model). The object indicators with the associated parameter vector $\boldsymbol{\gamma}$ are used as obligatory covariates, and the subject-object interactions (referred to as *candidate covariates*) with the associated parameter vector $\boldsymbol{\lambda}$ are used as optional covariates. For simplicity, we assume that the subject-specific covariates, and therefore also the candidate covariates, are either metric or binary. Categorical covariates with $K > 2$ categories can be included by creating $K - 1$ dummy variables, where the K-th category level is set as the reference category.

After the initialization in **Step 1**, the algorithm splits up into two main parts within an iteration. The first part (**Step 2.1**) considers the design matrix $\mathbf{X}_O$, in which the constraint matrix and the object indicators are included. Thus, the parameter estimates for the object indicators are updated along with the estimates for the threshold parameters. The second part (**Step 2.2**) considers the matrix of candidate covariates $\mathbf{X}_C = [\mathbf{x}_{\cdot 1}, \ldots, \mathbf{x}_{\cdot C}]$, where $\mathbf{x}_{\cdot c}$ is the c-th column of $\mathbf{X}_C$ and reflects the vector for a single candidate covariate. In each iteration, only the parameter of a single candidate covariate is updated, namely the one that maximally improves the fit according to a certain criterion, such as the AIC or BIC that are introduced in Section 6.3. This is considered as 'component-wise linear base procedure' because the weak learner, which is also called the base learner, is used to describe a linear effect of the single candidate covariate (see Hofner, 2011, chap 2.3). The algorithm stops if the maximum number of iterations b_{stop} is reached, and it can be described as follows:

Algorithm: `BTLboost`

Step 1: Initialization

Fit the intercept model $\boldsymbol{\mu}^{(0)} = h(\boldsymbol{\eta}^{(0)})$ (for some response function h) by maximizing the likelihood in order to obtain

$$\hat{\boldsymbol{\eta}}^{(0)} = \mathbf{Q}_{(K-1) \times q} \cdot \hat{\boldsymbol{\theta}}_q.$$

This yields only estimates for the threshold parameters, such that the estimated parameter vector has the form

$$\hat{\boldsymbol{\beta}}^{(0)} = (\hat{\boldsymbol{\theta}}_q^\top, \hat{\boldsymbol{\gamma}}^\top, \hat{\boldsymbol{\lambda}}^\top) = (\hat{\theta}_1, \ldots, \hat{\theta}_q, 0, \ldots, 0)^\top.$$

Step 2: For each $b = 1, 2, \ldots, b_{\text{stop}}$

Step 2.1: Consider object indicators and thresholds

– *Estimation:*
Fit the model $\boldsymbol{\mu} = h(\hat{\boldsymbol{\eta}}^{(b-1)} + \mathbf{X}_O \boldsymbol{\beta}_O)$, where $\hat{\boldsymbol{\eta}}^{(b-1)}$ is treated as offset and $\mathbf{X}_O$ is a design matrix based on the object indicators and the constraint matrix for updating the object parameters and threshold parameters, respectively. The estimated parameter vector

$$\hat{\boldsymbol{\beta}}_O^\top = (\hat{\boldsymbol{\theta}}_q^\top, \hat{\boldsymbol{\gamma}}^\top, \hat{\boldsymbol{\lambda}}^\top) = (\hat{\theta}_1, \ldots, \hat{\theta}_q, \hat{\gamma}_{(1)}, \ldots, \hat{\gamma}_{(M-1)}, 0, \ldots, 0)$$

is obtained by one-step Fisher scoring (see Tutz et al., 2006).

– *Update:*
$$\hat{\boldsymbol{\eta}}^{(b)} = \hat{\boldsymbol{\eta}}^{(b-1)} + \mathbf{X}_O \hat{\boldsymbol{\beta}}_O$$
$$\hat{\boldsymbol{\beta}}^{(b)} = \hat{\boldsymbol{\beta}}^{(b-1)} + \hat{\boldsymbol{\beta}}_O$$

Step 2.2: Consider candidate covariates

– *Estimation (for each candidate covariate $c = 1, \ldots, C$):*
Fit the model $\boldsymbol{\mu} = h(\hat{\boldsymbol{\eta}}^{(b)} + \mathbf{x}_{\cdot c} \lambda_c)$, $\forall c = 1, \ldots, C$, where $\hat{\boldsymbol{\eta}}^{(b)}$ is treated as offset and $\mathbf{x}_{\cdot c}$ reflects the c-th candidate covariate. The estimated parameter $\hat{\lambda}_c$ is obtained by one-step Fisher scoring (see Tutz et al., 2006).

– *Selection:*
Choose the candidate covariate $\mathbf{x}_{\text{best}}$ over all $\mathbf{x}_{\cdot 1}, \ldots, \mathbf{x}_{\cdot C}$ covariates that maximally improves the fit with respect to some criteria (e.g. AIC, BIC). The associated estimated parameter $\hat{\lambda}_{\text{best}}$ is included within the parameter vector $\hat{\boldsymbol{\lambda}}$, such that

$$\hat{\boldsymbol{\beta}}_{\text{best}}^\top = (\hat{\boldsymbol{\theta}}_q^\top, \hat{\boldsymbol{\gamma}}^\top, \hat{\boldsymbol{\lambda}}^\top) = (0, \ldots, 0, \hat{\lambda}_{\text{best}}, 0, \ldots, 0).$$

– *Update:*
$$\hat{\boldsymbol{\eta}}^{(b)} = \hat{\boldsymbol{\eta}}^{(b)} + \mathbf{X}_{\text{best}} \hat{\boldsymbol{\beta}}_{\text{best}}, \text{ where } \mathbf{X}_{\text{best}} = [\mathbf{0}, \ldots, \mathbf{0}, \mathbf{x}_{\text{best}}, \mathbf{0}, \ldots, \mathbf{0}].$$
$$\hat{\boldsymbol{\beta}}^{(b)} = \hat{\boldsymbol{\beta}}^{(b)} + \hat{\boldsymbol{\beta}}_{\text{best}}$$

6.3. Stopping Criteria

The stopping iteration b_{stop} can be seen as a tuning parameter in boosting. Common choices for determining the optimal stopping iteration are either information criteria, such as the AIC or BIC, or cross-validation. In this thesis, we consider only the two information criteria mentioned above, which are known measures for the trade-off between goodness-of-fit and model complexity. The computation of both criteria is based on the deviance $\text{Dev}(\hat{\boldsymbol{\eta}}^{(b)})$ and the degrees of freedom $\text{df}(b)$ after b boosting iterations. Since the boosting algorithm is an iterative procedure, the true degrees of freedom after b boosting iterations are unknown and need to be approximated.

In the literature, using the trace of the hat matrix after b boosting iterations as an approximation for the degrees of freedom is often suggested (Tutz et al., 2006; Bühlmann and Hothorn, 2007). Thus, we have $\text{df}(b) \approx \text{df}_{\text{trace}}(b) = \text{trace}(\mathbf{H}_b)$, for which the approximate hat matrix $\mathbf{H}_b$ needs to be computed in each iteration. The estimation formula for the approximate hat matrix and its trace can be found in Zahid and Tutz (2011).

Another possibility proposed by Bühlmann and Hothorn (2008) is to use the number of parameters that have been selected until the b-th boosting iteration (denoted by $\text{df}_{\text{actset}}(b)$) as an approximation for the degrees of freedom. The results in their paper suggest that $\text{df}_{\text{actset}}(b)$ *'is a better approximation for degrees of freedom for boosting with the component-wise linear base procedure.'* Since we use a *component-wise linear base procedure* in the `BTLboost` algorithm, the use of $\text{df}_{\text{actset}}(b)$ seems reasonable and allows for faster and simpler computation because we do not need to estimate the hat matrix in each iteration.

The definition of the AIC and BIC after b iterations is given by

$$\text{AIC}(b) = \text{Dev}(\hat{\boldsymbol{\eta}}^{(b)}) + 2 \cdot \text{df}(b) \tag{6.1}$$

and

$$\text{BIC}(b) = \text{Dev}(\hat{\boldsymbol{\eta}}^{(b)}) + \log(n) \cdot \text{df}(b), \tag{6.2}$$

where n is usually the number of observations, or, in our case, the number of comparisons that are made by all subjects. Thus, in a complete pair comparison experiment $n = \binom{M}{2} \cdot I$. For the degrees of freedom $\text{df}(b)$, we can use the approximation $\text{df}(b) \approx \text{df}_{\text{trace}}(b)$ or $\text{df}(b) \approx \text{df}_{\text{actset}}(b)$ as mentioned above.

The optimal stopping iteration b^*_{stop} is chosen among the iterations $b = 1, \ldots, b_{\text{stop}}$ as the one yielding the best (=lowest) AIC or BIC, respectively. Thus, one has to compute the AIC or BIC for all iterations first. The optimal stopping iteration is then determined afterwards using

$$b^*_{\text{stop}} = b \quad \Leftrightarrow \quad \text{AIC}(b^*_{\text{stop}}) = \min_{b=1,\ldots,b_{\text{stop}}} \text{AIC}(b)$$

when the AIC is chosen as stopping criterion and

$$b^*_{\text{stop}} = b \quad \Leftrightarrow \quad \text{BIC}(b^*_{\text{stop}}) = \min_{b=1,\ldots,b_{\text{stop}}} \text{BIC}(b)$$

when the BIC is chosen as stopping criterion, respectively.

7. Simulation

7.1. Basic Considerations

In this chapter, we investigate the performance of the boosting algorithm for three simulation scenarios with 100 simulations for each scenario. In the first two scenarios, we use $M = 4$ objects, and $M = 8$ objects in the third scenario. All objects are compared pairwise by $I = 150$ subjects on $K = 3$ response categories, meaning that each simulated subject has the possibility to prefer the first object, neither of the objects, or the second object. In all scenarios, we use $P = 6$ subject-specific covariates, which are drawn from the following distributions:

- $X_1 \sim B(1, 0.3)$ and $X_2, X_3 \sim B(1, 0.5)$
- $X_4, X_5, X_6 \sim N(0, 1)$

Each scenario is performed twice, once with the AIC as the stopping criterion and once with the BIC. For the sake of simplicity, the AIC and BIC are computed using the number of parameters that have been selected until the b-th boosting iteration as an approximation for the degrees of freedom. This was also done to investigate whether the results will be meaningful or useful when using $\mathrm{df}(b) \approx \mathrm{df}_{\mathrm{actset}}(b)$. Furthermore, we set the maximum number of iterations to $b_{\mathrm{stop}} = 1000$ and allow the algorithm to stop earlier if the chosen stopping criterion deteriorated in the last 10 consecutive iterations. We then use the estimates of the iteration b_{stop}^* that yielded the best (=lowest) AIC or BIC, respectively (see Section 6.3).

7.2. Simulation Scenarios

For each subject $i = 1, \ldots, 150$, we have the realizations of subject-specific covariates $\mathbf{x}_i^\top = (x_{i1}, \ldots, x_{i6})$ that are used to simulate the cumulative logit model

$$\mathrm{logit}(\mathbb{P}(Y_{rs} \leq k | (r,s), i)) = \theta_k + \sum_{m=1}^{M-1} x_m^{(r,s)} \gamma_{(m)} + \sum_{p=1}^{P} \sum_{m=1}^{M-1} x_{ip} x_m^{(r,s)} \lambda_{(m)p}, \quad k = 1, 2.$$

In the first two scenarios, we use $M = 4$ objects yielding $C = P \cdot (M - 1) = 18$ subject-object interactions that can potentially be included in the model. In scenario 3, we use $M = 8$ objects yielding $C = P \cdot (M - 1) = 42$ possible subject-object interactions. Each component used in the model above is standardized, such that they have a variance of 1.

Hereinafter, we denote $\boldsymbol{\lambda}^\top_{(\cdot)p} = (\lambda_{(1)p}, \ldots, \lambda_{(M-1)p})$ as a vector that contains all $M - 1$ subject-object interactions referring to a single subject-specific covariate p. The first two scenarios split up into 'strong effects' and 'small effects' settings. In the 'small effects' setting, the relevant (= non-zero) subject-object parameters are reduced by a factor of 0.5 compared to those from the 'strong effects' setting. This was done to investigate whether smaller effects could also be identified. The scenarios are organized as follows:

- **Scenario 1** (4 relevant and 14 irrelevant subject-object interactions):
  The corresponding relevant parameters, for which $\lambda_{(\cdot)p} \neq 0$, are displayed in boldface in Table 7.1. Additionally, we use the following irrelevant parameters: $\boldsymbol{\lambda}^\top_{(\cdot)2} = \boldsymbol{\lambda}^\top_{(\cdot)3} = \boldsymbol{\lambda}^\top_{(\cdot)5} = \boldsymbol{\lambda}^\top_{(\cdot)6} = (0.0, 0.0, 0.0)$ that are set to zero.

Parameter	Strong Effects	Small Effects
$\boldsymbol{\theta}^\top$	(-0.8, 0.8)	(-0.8, 0.8)
$\boldsymbol{\gamma}^\top$	(2.0, 1.5, 1.0)	(2.0, 1.5, 1.0)
$\boldsymbol{\lambda}^\top_{(\cdot)1}$	(**-0.8**, **-0.7**, 0.0)	(**-0.4**, **-0.35**, 0.0)
$\boldsymbol{\lambda}^\top_{(\cdot)4}$	(**0.8**, **0.7**, 0.0)	(**0.4**, **0.35**, 0.0)

Table 7.1.: Parameters used in scenario 1

- **Scenario 2** (5 relevant and 13 irrelevant subject-object interactions):
  The corresponding relevant parameters are displayed in boldface in Table 7.2. Additionally, we use the following irrelevant parameters $\boldsymbol{\lambda}^\top_{(\cdot)5} = \boldsymbol{\lambda}^\top_{(\cdot)6} = (0.0, 0.0, 0.0)$.

Parameter	Strong Effects	Small Effects
$\boldsymbol{\theta}^\top$	(-0.8, 0.8)	(-0.8, 0.8)
$\boldsymbol{\gamma}^\top$	(2.0, 1.5, 1.0)	(2.0, 1.5, 1.0)
$\boldsymbol{\lambda}^\top_{(\cdot)1}$	(**-0.8**, 0.0, 0.0)	(**-0.4**, 0.0, 0.0)
$\boldsymbol{\lambda}^\top_{(\cdot)2}$	(0.0, **-0.7**, 0.0)	(0.0, **-0.35**, 0.0)
$\boldsymbol{\lambda}^\top_{(\cdot)3}$	(0.0, **0.6**, 0.0)	(0.0, **0.3**, 0.0)
$\boldsymbol{\lambda}^\top_{(\cdot)4}$	(**0.8**, **0.7**, **0.6**)	(**0.4**, **0.35**, **0.3**)

Table 7.2.: Parameters used in scenario 2

- **Scenario 3** (10 relevant and 32 irrelevant subject-object interactions):
  The corresponding relevant parameters are displayed in boldface in Table 7.3. Additionally, we use the following irrelevant parameters $\boldsymbol{\lambda}_{(\cdot)2}^{\top} = \boldsymbol{\lambda}_{(\cdot)3}^{\top} = \boldsymbol{\lambda}_{(\cdot)5}^{\top} = \boldsymbol{\lambda}_{(\cdot)6}^{\top} = (0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0)$.

Parameter	Effects
$\boldsymbol{\theta}^{\top}$	(-0.8, 0.8)
$\boldsymbol{\gamma}^{\top}$	(2.0, 1.8, 1.6, 1.4, 1.2, 1.0, 0.8)
$\boldsymbol{\lambda}_{(\cdot)1}^{\top}$	(**-0.7**, **-0.6**, 0.0, 0.0, **-0.3**, **-0.2**, **-0.1**)
$\boldsymbol{\lambda}_{(\cdot)4}^{\top}$	(0.0, 0.0, **0.5**, **0.4**, **0.3**, **0.2**, **0.1**)

Table 7.3.: Parameters used in scenario 3

7.3. Results

As in Zahid and Tutz (2011), we introduce two measurements to investigate the performance of the algorithm:

- The hit rate (HR) that represents the percentage of correctly identified relevant subject-object interactions is given by

$$
\text{HR} = \frac{\sum\limits_{p=1}^{P} \sum\limits_{m=1}^{M-1} \mathbf{I}(\lambda_{(m)p} \neq 0) \cdot \mathbf{I}(\hat{\lambda}_{(m)p} \neq 0)}{\sum\limits_{p=1}^{P} \sum\limits_{m=1}^{M-1} \mathbf{I}(\lambda_{(m)p} \neq 0)} \tag{7.1}
$$

- The false alarm rate (FAR) that represents the percentage of irrelevant candidate covariates that are mistakenly identified as relevant candidate covariates is given by

$$
\text{FAR} = \frac{\sum\limits_{p=1}^{P} \sum\limits_{m=1}^{M-1} \mathbf{I}(\lambda_{(m)p} = 0) \cdot \mathbf{I}(\hat{\lambda}_{(m)p} \neq 0)}{\sum\limits_{p=1}^{P} \sum\limits_{m=1}^{M-1} \mathbf{I}(\lambda_{(m)p} = 0)} \tag{7.2}
$$

The $\hat{\lambda}_{(m)p}$'s are the parameters that were identified and estimated by the `BTLboost` algorithm, whereas the $\lambda_{(m)p}$'s are the predefined (true) subject-object parameters from Tables 7.1 – 7.3.

An overview of the hit rate and false alarm rate for all three scenarios is given in Table 7.4. Hereby, we denote HR_{AIC} and HR_{BIC} as the hit rate when the AIC or

BIC is used as stopping criterion, and FAR_{AIC} and FAR_{BIC} are, accordingly, the false alarm rate when the AIC or BIC is used.

In the 'strong effects' setting of scenario 1, both the AIC and the BIC identified all relevant subject-object parameters ($\text{HR}_{\text{AIC}} = \text{HR}_{\text{BIC}} = 1$). In this setting, the $\text{FAR}_{\text{BIC}} = 0.01571$ is lower than the $\text{FAR}_{\text{AIC}} = 0.10929$. This suggests that the BIC performs better than the AIC because the AIC is likely to select irrelevant subject-object interactions more frequently (see also Figure 7.1, which shows that in all scenarios, the relative frequency for including irrelevant subject-object interactions is always higher for the AIC than for the BIC). However, in the 'small effects' setting, the $\text{HR}_{\text{BIC}} = 0.73750$ is much lower than the $\text{HR}_{\text{AIC}} = 0.98500$, whereas both the FAR_{AIC} and the FAR_{BIC} increased by about 0.04 compared to the 'strong effects' setting. This suggests that, in this case, the BIC might be too restrictive. The results in scenario 2 are similar.

In scenario 3, we use strong and small effects at the same time. The $\text{FAR}_{\text{AIC}} = 0.17781$ is much higher compared to the $\text{FAR}_{\text{BIC}} = 0.03500$. This could be explained from the following fact: Within our predefined (true) parameters of $\boldsymbol{\lambda}_{(\cdot)1}^{\top}$ and $\boldsymbol{\lambda}_{(\cdot)4}^{\top}$ from Table 7.3, two subject-object parameters were set to zero (those that are not boldface), and the AIC tends to select these subject-object interactions more frequently than the BIC (see also Figure 7.1, where, for example, in scenario 3 the relative frequency for $\hat{\lambda}_{(4)1}$ is about 90%, although we use the predefined parameter $\lambda_{(4)1} = 0.0$).

Scenario	Setting	HR_{AIC}	HR_{BIC}	FAR_{AIC}	FAR_{BIC}
1	strong effects	1.00000	1.00000	0.10929	0.01571
	small effects	0.98500	0.73750	0.14929	0.06000
2	strong effects	1.00000	1.00000	0.10500	0.01083
	small effects	0.98333	0.62833	0.13583	0.05333
3	strong and small effects	0.88800	0.77900	0.17781	0.03500

Table 7.4.: Averaged hit rate (HR) and false alarm rate (FAR) over all 100 simulations for each setting.

Figures 7.2 - 7.6 display the estimates of 100 simulated data sets for all settings that are mentioned in Section 7.2. The boxplots give an overview of the range of the estimates, where the red squares denote the underlying predefined (true) parameter values that are defined in Tables 7.1 - 7.3. Note that the width of the boxplots is directly related to the number of observations within the boxplot: the thinner the boxplot, the fewer times its associated parameter was selected.

All in all it is difficult to say whether the AIC or the BIC should be chosen as the stopping criterion. If one is mainly interested in identifying 'strong effects', the BIC seems more appropriate due to the lower FAR_{BIC} compared to the FAR_{AIC} from our 'strong effects' settings. If one wants to identify 'small effects' as well, using the BIC might be too restrictive and may lead to a lower HR_{BIC} compared to the HR_{AIC}. Thus, one has to weigh the options between the AIC, which may lead to a higher HR but also a higher FAR, or the BIC, which may lead to a lower FAR but, in some cases (especially when 'small effects' are relevant), also to a much lower HR.
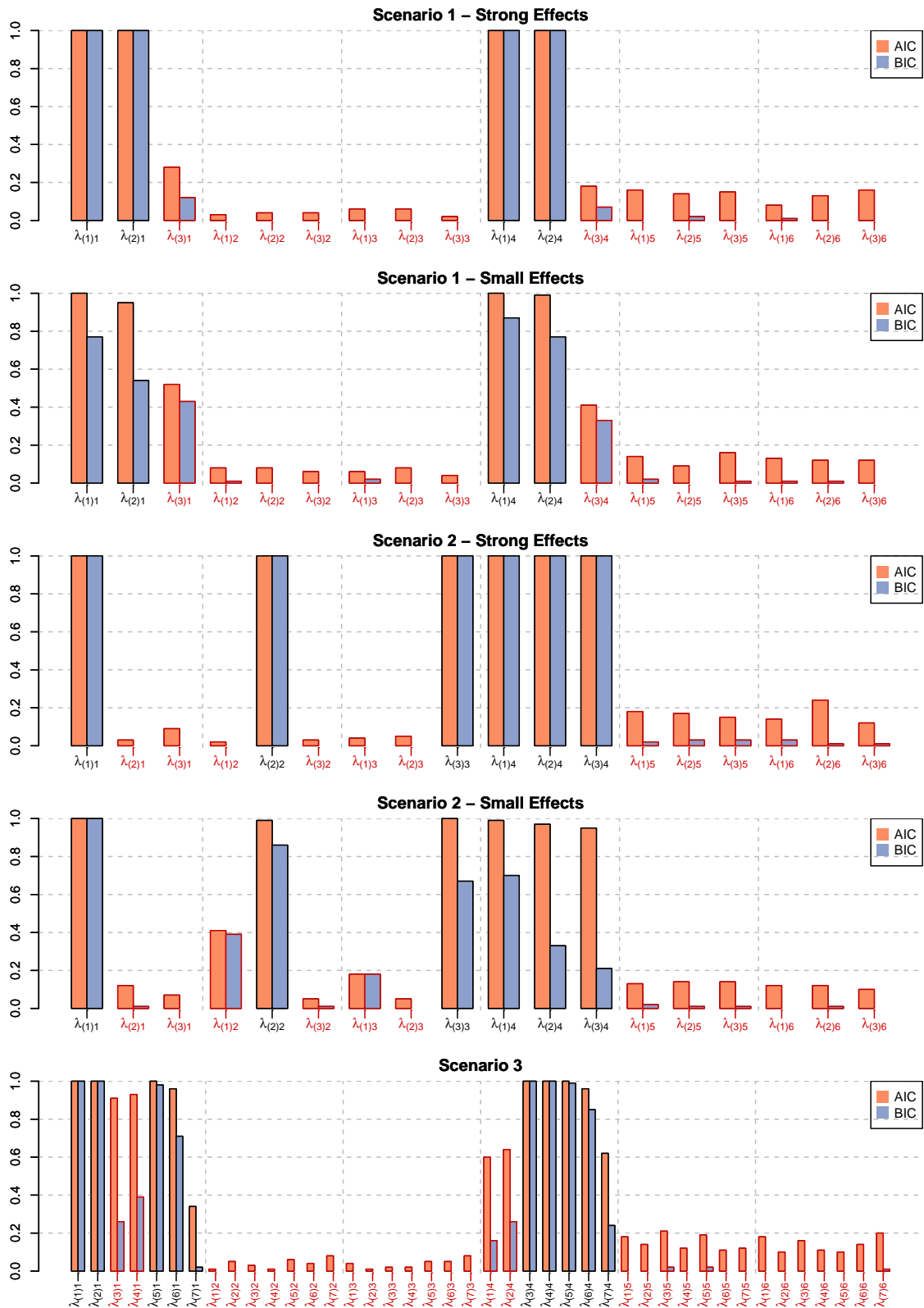
Figure 7.1.: Relative frequency of selected subject-object interactions averaged over the 100 simulations for each scenario. Irrelevant subject-object interactions are labelled in red, whereas the relevant ones are labelled in black.
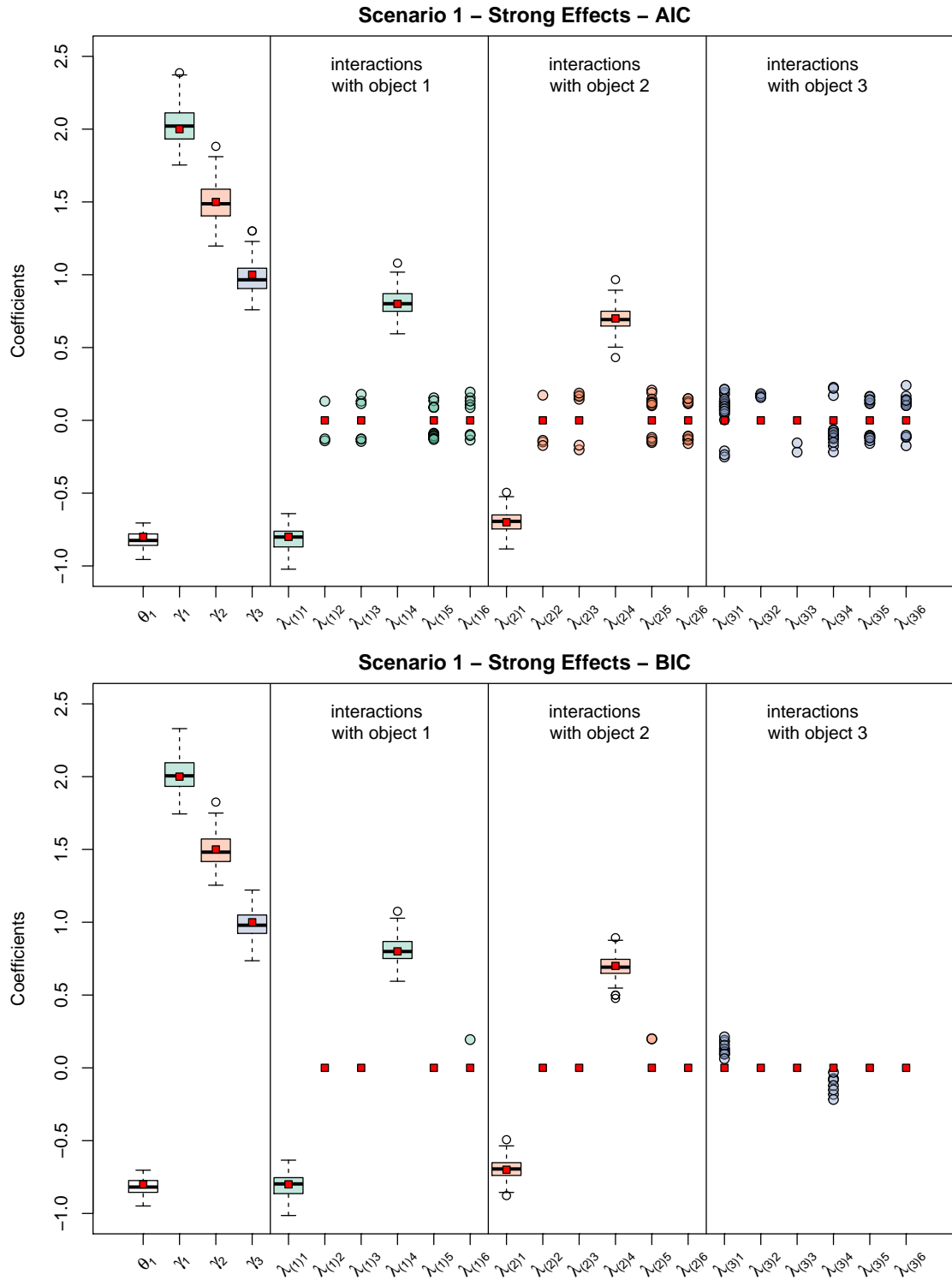
Figure 7.2.: Estimated coefficients after running the `BTLboost` algorithm with AIC and BIC as stopping criteria for scenario 1 (strong effects)
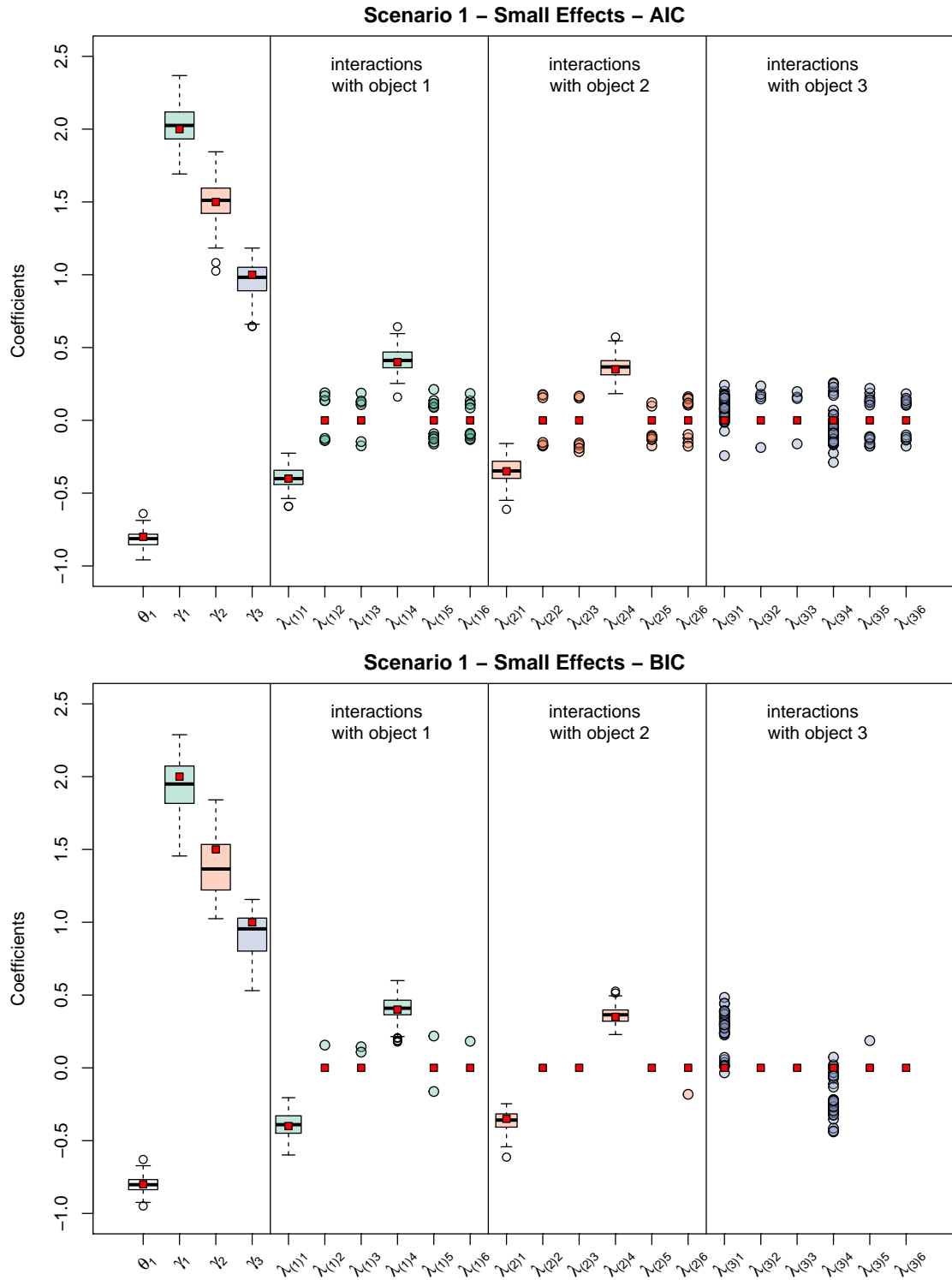
Figure 7.3.: Estimated coefficients after running the `BTLboost` algorithm with AIC and BIC as stopping criteria for scenario 1 (small effects)

Figure 7.4.: Estimated coefficients after running the `BTLboost` algorithm with AIC and BIC as stopping criteria for scenario 2 (strong effects)

Figure 7.5.: Estimated coefficients after running the `BTLboost` algorithm with AIC and BIC as stopping criteria for scenario 1 (small effects)

Figure 7.6.: Estimated coefficients after running the `BTLboost` algorithm with AIC and BIC as stopping criteria for scenario 3

8. Application
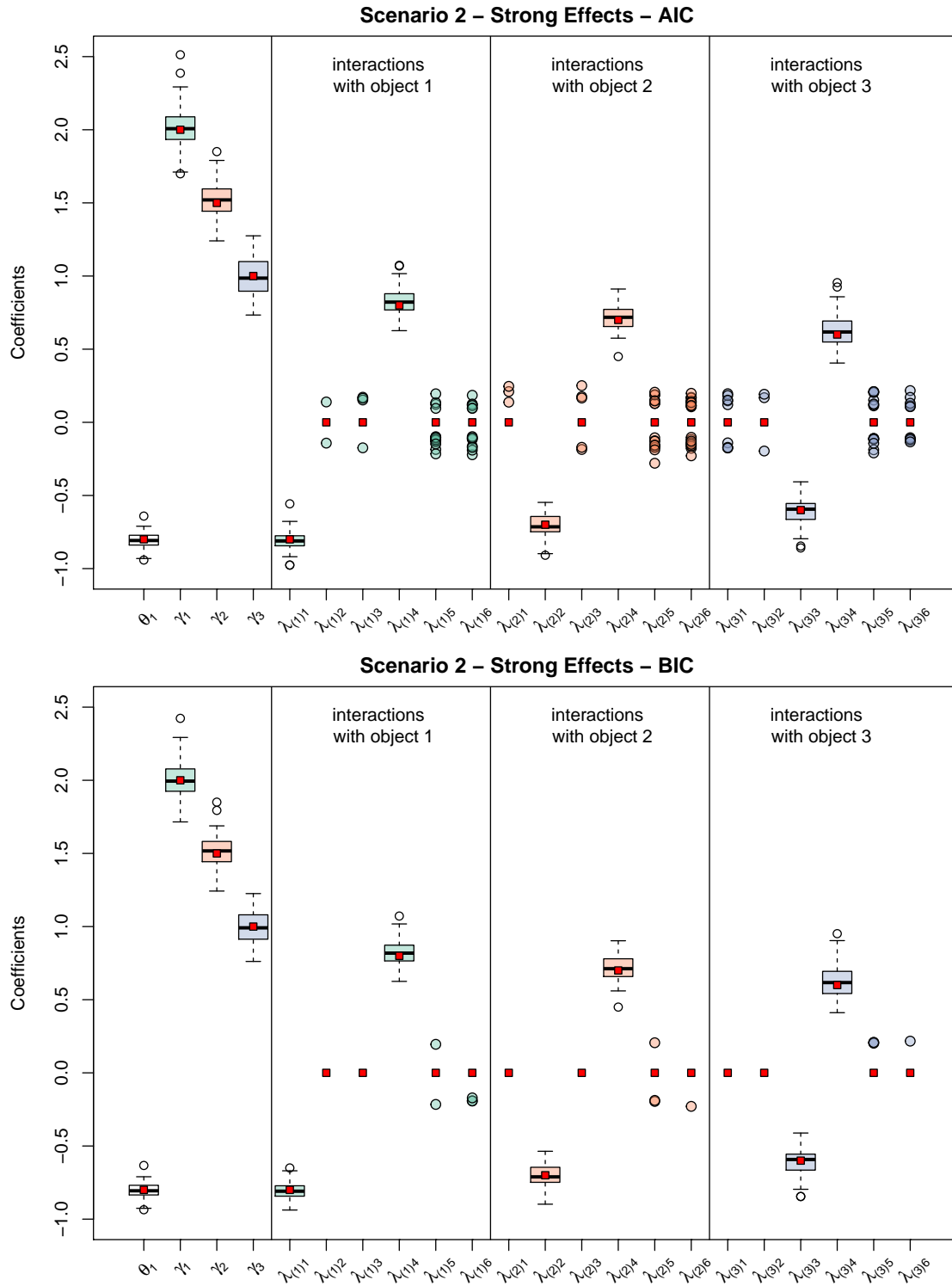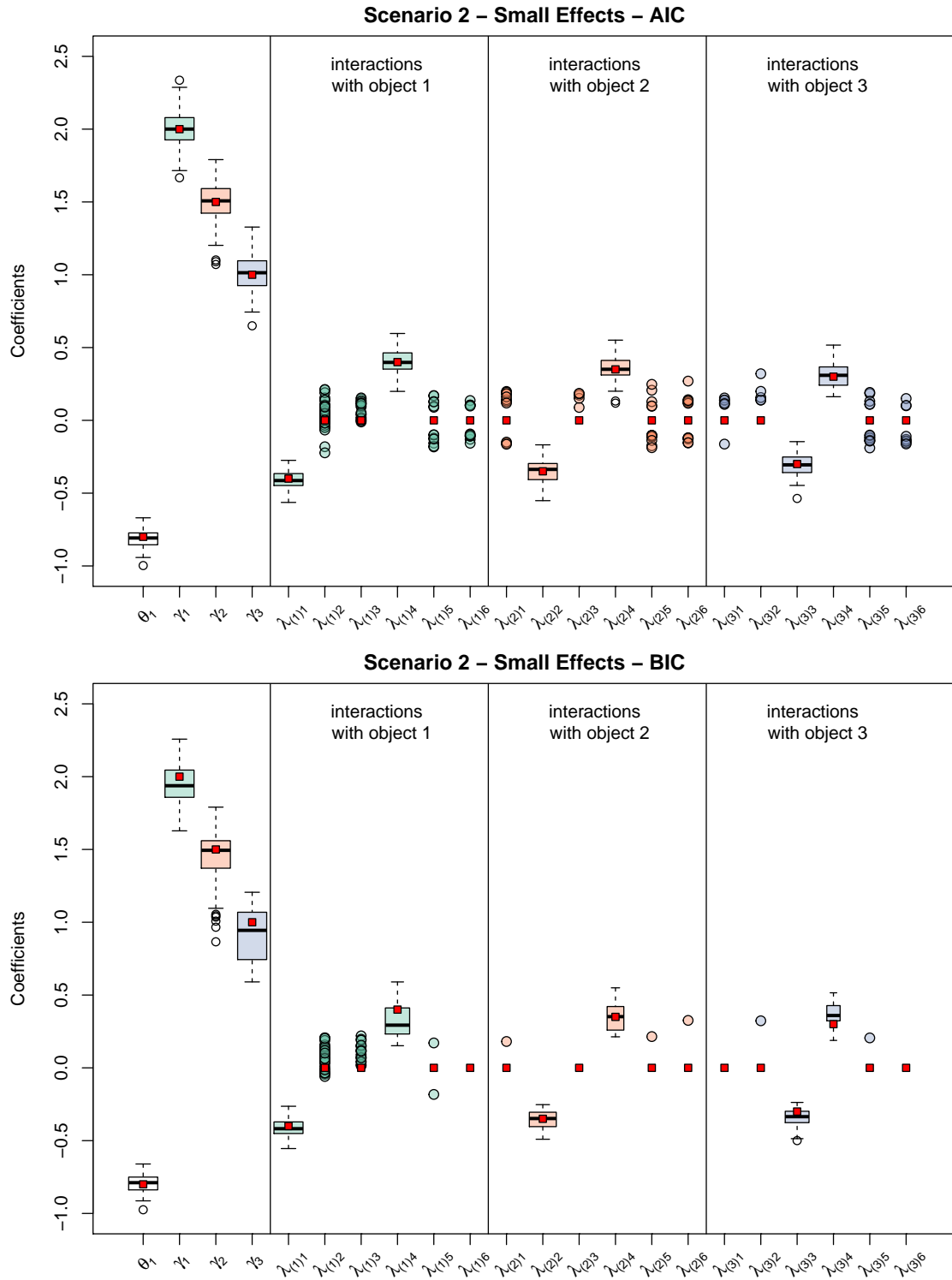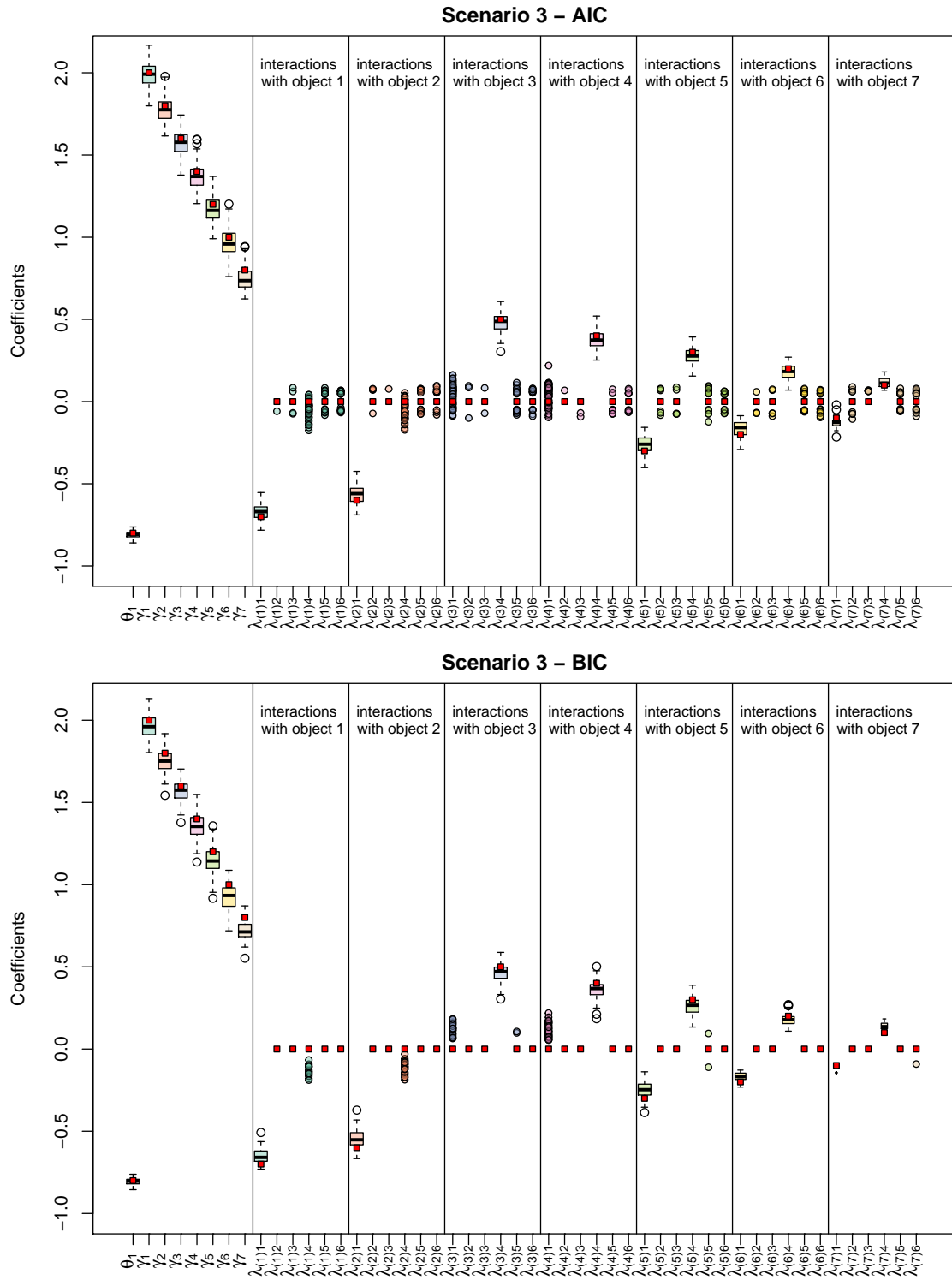
All of the models presented in this chapter are fitted with the `ordBTL` package, which was developed during this master thesis. The `R`-code to reproduce the results from the following applications can be found in Appendix A.2. The first application considers an ordinal response with 7 categories. The second application concentrates on fitting a pair comparison model considering subject-specific covariates. Furthermore, the `BTLboost` algorithm introduced in Chapter 6 is applied. The last application concerns the inclusion of home advantages.

8.1. Typewriter Ribbon Data

Agresti (1992) used a dataset where 5 different typewriter ribbon brands were compared pairwise by 30 secretaries yielding $\binom{5}{2} = 10$ pair comparisons for each secretary. In a comparison of brand r and brand s, the secretaries had the choice to prefer brand r strong, moderate, or mild, prefer neither brand, or prefer brand s mild, moderate, or strong. Thus, the response consists of seven categories, such that $Y_{rs} \in \{1, \ldots, 7\}$.

In Figure 8.1, we see stacked bar plots showing the percentage of how strong a brand was preferred in a pair comparison. The brands displayed on the left green bars have a higher percentage of being preferred than the brands displayed on the yellowish bars on the right side of the plot. The first bar plot, for example, shows that in a comparison between brand 3 and brand 4, the latter was preferred in 3% of the cases and brand 3 was preferred in 87% of the cases, whereas in 10% of the cases none of either brand were preferred. Furthermore, we can see that each brand was involved in four comparisons and brand 3 only shows up on the left green bars. This means that in any comparison in which brand 3 was involved, it was the most preferred brand. In contrast to this, brand 4 was never the most preferred brand since it only shows up on the yellow bars on the right. From this simple visual analysis, we can

assume that brand 3 may be the most preferred brand, whereas brand 4 may be the least preferred brand.
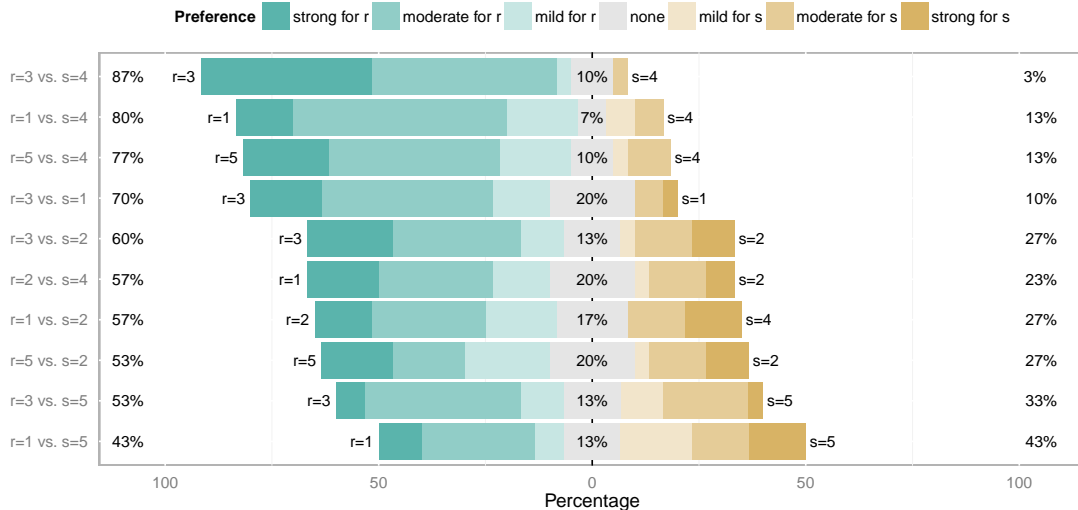
Figure 8.1.: Percentages of how strong a brand was preferred in a comparison between brand r and brand s

For illustration, we fit the cumulative logit model from Agresti (1992):

$$\text{logit}\left(\mathbb{P}(Y_{rs} \leq k | (r, s))\right) = \theta_k + (x_1^{(r,s)}\gamma_1 + \ldots + x_5^{(r,s)}\gamma_5), \quad k = 1, \ldots, 6, \qquad (8.1)$$

which has the constraint that all object parameters sum up to zero: $\sum_{i=1}^5 \gamma_i = 0$. The `ordBTL` package is only able to fit the model with the constraint $\gamma_5 = 0$. Therefore, we have to re-parametrize model (8.1) to get the same results as Agresti. To do this, we use the constraint that all parameters sum up to zero and re-express the parameter γ_5 as a linear combination of all other parameters, which yields

$$\gamma_5 = 0 - \sum_{i=1}^{4} \gamma_i \quad \Leftrightarrow \quad \sum_{i=1}^{5} \gamma_i = 0.$$

This is then plugged in model (8.1), yielding the model

$$\text{logit}\left(\mathbb{P}(Y_{rs} \leq k | (r, s))\right) = \theta_k + (z_1^{(r,s)}\gamma_1 + \ldots + z_4^{(r,s)}\gamma_4), \quad k = 1, \ldots, 6, \qquad (8.2)$$

where $z_i^{(r,s)} = x_i^{(r,s)} - x_5^{(r,s)}$, $\forall i = 1, \ldots, 4$ are used as independent variables to obtain estimates for $\gamma_1, \ldots, \gamma_4$ and the estimate for γ_5 is computed by the linear combination $\hat{\gamma}_5 = 0 - \sum_{i=1}^4 \hat{\gamma}_i$.

The results can be seen in Table 8.1 and are the same as in Agresti (1992, Table 2). Note that $\hat{\theta}_4 = -\hat{\theta}_3$, $\hat{\theta}_5 = -\hat{\theta}_2$ and $\hat{\theta}_6 = -\hat{\theta}_1$ due to the symmetry constraints for the thresholds. The parameter estimates for the adjacent categories logit model and the cumulative probit model are carried out analogously. In each of the three models, the ranking based on the object parameters is the same: brand 3 ranks best, followed by brand 5, brand 1, brand 2 and brand 4, which is ranked worst.

Parameter	Adjacent categories logit	Cumulative logit	Cumulative probit	Rank
$\hat{\theta}_1$	-0.852	-1.379	-2.400	–
$\hat{\theta}_2$	0.833	-0.490	-0.830	–
$\hat{\theta}_3$	-0.544	-0.219	-0.371	–
$\hat{\gamma}_1$	0.042	0.058	0.117	3
$\hat{\gamma}_2$	-0.050	-0.088	-0.196	4
$\hat{\gamma}_3$	0.270	0.494	0.887	1
$\hat{\gamma}_4$	-0.340	-0.607	-1.048	5
$\hat{\gamma}_5$	0.078	0.143	0.239	2

Table 8.1.: Estimated thresholds and object parameters; The last column displays the ranking of the five brands, which is the same for all models.

8.2. CEMS Data

8.2.1. Description

We use the *Community of European management schools (CEMS)* data from Dittrich et al. (1998) that was collected in a survey of 303 students of the Vienna University of Economics. The aim of this study was to investigate the preferences of students for studying at least one semester abroad in one of 6 different universities (London, Paris, Barcelona, St.Gallen, Milan and Stockholm) and to establish an overall ranking of these universities. For each of the $\binom{M}{2} = \binom{6}{2} = 15$ pairs of universities the students could either prefer the first university, none of both universities, or the second university. Additionally, the data contains different personal characteristics (subject-specific covariates) of the students. An overview of the $P = 8$ subject-specific (binary) covariates is given in Table 8.2.

In Figure 8.2, we illustrate the results of the study using stacked bar plots for each possible pair comparison. The green bars on the left always have a higher percentage

of being preferred than the yellow bars on the right. The grey bars reflect the percentage in which the surveyed students had no preference in a given comparison of university r and university s. London appears only on the left green bars, meaning that in any comparison where London was involved, it was the most preferred university. This suggests that, without considering subject-specific covariates, London may be the most preferred university. Consequently, Stockholm may be the least preferred university, since it appears only on the right yellow bars. The results in the following section will confirm these conclusions that were drawn purely from Figure 8.2.

Covariate	Description	Coding
STUD	Main discipline of study	0 = other, 1 = commerce
ENG	Knowledge of English	0 = good, 1 = poor
FRA	Knowledge of French	0 = good, 1 = poor
SPA	Knowledge of Spanish	0 = good, 1 = poor
ITA	Knowledge of Italian	0 = good, 1 = poor
WOR	Full-time employment while studying	0 = no, 1 = yes
DEG	Intention to take an international degree	0 = no, 1 = yes
SEX	Gender	0 = female, 1 = male

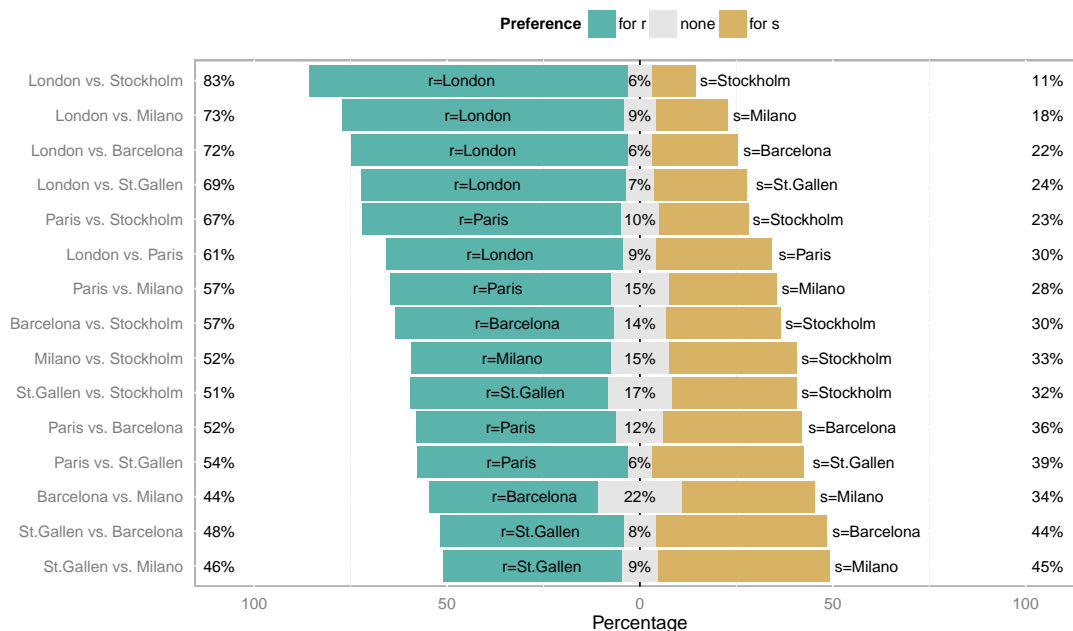Table 8.2.: Description of all $P = 8$ subject-specific covariates.

Figure 8.2.: Percentages of the preferred universities in the comparison of university r and university s.

8.2.2. Model without Subject-specific Covariates

The adjacent categories logit model (3.8) can be represented as a log-linear BTL model (see Agresti, 1992). Thus, we can reproduce the results of the log-linear BTL model from Dittrich et al. (2001) by fitting the adjacent categories logit model using the `ordBTL` package.

In this application, the adjacent categories logit model without considering the subject-specific covariates is

$$
\begin{aligned}
\log\left(\frac{\mathbb{P}(Y_{rs} = k|(r,s))}{\mathbb{P}(Y_{rs} = k+1|(r,s))}\right) &= \theta_k + (\gamma_r - \gamma_s) \\
&= \theta_k + (x_1^{(r,s)}\gamma_1 + \ldots + x_5^{(r,s)}\gamma_5), \quad k = 1,2,
\end{aligned}
\tag{8.3}
$$

where

γ_1 is the parameter for London,
γ_2 is the parameter for Paris,
γ_3 is the parameter for Barcelona,
γ_4 is the parameter for St.Gallen,
γ_5 is the parameter for Milan and
γ_6 is the parameter for Stockholm, which is used as reference and thus $\gamma_6 = 0$.

The resulting estimates are given in Table 8.3, which are the same estimates as in the 'initial model' from Dittrich et al. (2001, Table 3). Here, London is the most preferred university, followed by Paris, Barcelona, St.Gallen, Milan and Stockholm, which is the least preferred university. The interpretation of the threshold θ_1 can be derived when two universities with the same 'worth' and hence the same object parameters $\gamma_r = \gamma_s \Leftrightarrow (\gamma_r - \gamma_s) = 0$ are considered. In this case, the resulting log-odds for the adjacent categories logit model with $K = 3$ response categories are defined as

$$
\log\left(\frac{\mathbb{P}(Y_{rs} = 1|(r,s))}{\mathbb{P}(Y_{rs} = 2|(r,s))}\right) = \log\left(\frac{\mathbb{P}(\text{'}r\text{ is preferred'})}{\mathbb{P}(\text{'no preference'})}\right) = \theta_1
\tag{8.4}
$$

and

$$
\log\left(\frac{\mathbb{P}(Y_{rs} = 2|(r,s))}{\mathbb{P}(Y_{rs} = 3|(r,s))}\right) = \log\left(\frac{\mathbb{P}(\text{'no preference'})}{\mathbb{P}(\text{'}s\text{ is preferred'})}\right) = \theta_2.
\tag{8.5}
$$

Since $\theta_2 = -\theta_1$, we can re-express equation (8.5) as

$$
\log\left(\frac{\mathbb{P}(Y_{rs} = 3|(r,s))}{\mathbb{P}(Y_{rs} = 2|(r,s))}\right) = \log\left(\frac{\mathbb{P}(\text{'}s\text{ is preferred'})}{\mathbb{P}(\text{'no preference'})}\right) = \theta_1.
\tag{8.6}
$$

From equations (8.4) and (8.6), it follows that

$$\log \left(\frac{\mathbb{P}(\text{'}s \text{ is preferred'})}{\mathbb{P}(\text{'no preference'})} \right) = \log \left(\frac{\mathbb{P}(\text{'}r \text{ is preferred'})}{\mathbb{P}(\text{'no preference'})} \right) = \theta_1. \qquad (8.7)$$

Thus, θ_1 can be interpreted as a 'no preference' parameter, where $\theta_1 > 0$ indicates that there is a tendency in favour of a decision (see also Dittrich et al. (1998), in which their $-\delta$ corresponds to our θ_1). As seen in Table 8.3, the estimate for the threshold has a positive value ($\hat{\theta}_1 = 1.317$). Thus, there is a tendency in favour of a decision.

Parameter	University	Estimates
$\hat{\theta}_1$	–	1.317
$\hat{\gamma}_1$	London	0.911
$\hat{\gamma}_2$	Paris	0.516
$\hat{\gamma}_3$	Barcelona	0.306
$\hat{\gamma}_4$	St.Gallen	0.299
$\hat{\gamma}_5$	Milan	0.221
$\hat{\gamma}_6$	Stockholm	(reference)

Table 8.3.: Parameter estimates for the adjacent categories logit model without subject-specific covariates from equation (8.3).

8.2.3. Model with Subject-specific Covariates

Dittrich et al. (1998) considered all $C = P \cdot (M - 1) = 8 \cdot 5 = 40$ subject-object interactions and all $P = 8$ two-way interactions between the subject-specific covariates and the no-preference term (which they denoted by δ). Thus, a total number of 48 covariates could potentially be included in their model. To select only the most important covariates, they used a forward and backward elimination procedure for their log-linear BTL model (see Christensen, 1990, p. 128). In their final model, 13 subject-object interactions and 2 interactions between subject-specific covariates and the no-preference term were selected with this procedure (see Dittrich et al., 2001, Table 6). Below, we apply the `BTLboost` algorithm on the adjacent categories logit model and get similar results. Within this model, the two-way interactions between subject-specific covariates and the no-preference term can be understood as threshold covariates. In this case, the threshold θ_1 is assumed to depend on subject-specific covariates and can be re-parametrized using a linear combination of subject-specific covariates, namely $\theta_1 + \sum_{p=1}^{P} x_p \theta_{1p}$ (see Fahrmeir et al., 1994, p. 79 - 80). Thus, these

two-way interactions can be included in the adjacent categories logit model using the subject-specific covariates and restrict their estimated parameters in the same way as the thresholds (using the symmetry constraints).

Table 8.4 displays the results of the `BTLboost` algorithm, where we the Wilkinson and Rogers notation (Wilkinson and Rogers, 1973) is used as a label for the parameters. Furthermore, we used the AIC as the stopping criterion because we also wanted to identify small effects. Here, we used $b_{\text{stop}} = 1000$ and allowed the algorithm to stop earlier if the AIC deteriorated in the last 10 consecutive iterations. As seen, the `BTLboost` algorithm selected almost the same covariates as the forward and backward elimination procedure that was used by Dittrich et al. (2001). Compared to their 'final model', the `BTLboost` algorithm additionally identifies the subject-object interactions `Barcelona.ENG` and `Milan.FRA`, whereas the interaction `London.ITA` was not selected.

Parameter	Estimate	Parameter	Estimate
London	1.046	Barcelona.WOR	0.548
Paris	0.660	StGallen.DEG	0.277
Milan	1.114	Paris.STUD	0.412
StGallen	0.266	StGallen.STUD	-0.210
Barcelona	1.060	StGallen.ENG	0.106
Stockholm	(reference)	London.FRA	-0.193
1	1.415	Paris.FRA	-0.764
SEX.1	-0.348	Barcelona.SPA	-0.811
WOR.1	0.633	Milan.ITA	-0.911
Paris.WOR	0.706	Milan.SEX	-0.147
Milan.WOR	0.540	**Barcelona.ENG**	**-0.167**
		Milan.FRA	**-0.164**

Table 8.4.: Selected covariates from `BTLboost` with AIC as the stopping criterion; The parameters in boldface do not appear in the final model from Dittrich et al. (2001).

The selected interactions between subject-specific covariates and the no-preference term (denoted by `1`) are `SEX.1` and `WOR.1`. It is found that $\hat{\theta}_{\text{WOR.1}} = 0.633 > 0$, which means that students who are employed full-time have a higher tendency in favour of a decision (compared to students who are not employed full-time) and because $\hat{\theta}_{\text{SEX.1}} = -0.348 < 0$, male students have a lower tendency in favour of a decision (compared to female students). Furthermore, $\hat{\lambda}_{\text{Paris.FRA}}$, $\hat{\lambda}_{\text{Barcelona.SPA}}$, $\hat{\lambda}_{\text{Milan.ITA}}$ have a negative value, which means that students with poor knowledge of French, Spanish and Italian have a lower tendency to prefer the universities in Paris (France),

Barcelona (Spain), and Milan (Italy), respectively. All interactions with the covariate `WOR` are positive valued, meaning that students who are employed full-time have a higher tendency to prefer the universities in Paris (France), Barcelona (Spain), and Milan (Italy). The interpretation for all the other subject-object interactions is analogous: when the subject-specific covariate has the value 1, then all subject-object interaction parameters with a negative value reflect a lower tendency for the corresponding object (here, university) and a positive value of the subject-object interaction parameters means a higher tendency for the corresponding object.

8.3. German Football League 2005/2006

8.3.1. Home Advantage

In this application, we consider the German football league (Bundesliga) of 2005/2006, in which $M = 18$ teams competed pairwise. There were a total of $\binom{M}{2} \cdot 2 = \binom{18}{2} \cdot 2 = 306$ matches. This is because in the Bundesliga, the teams meet twice (double round-robin tournament), where each team is allowed to play at its home location once. Therefore, we have the pairs (r, s), $\forall r \neq s$. The final standings are listed in Table 8.5 and are based on the points that are earned in each match: The winning team earns 3 points, whereas the losing team earns nothing. If the match ends in a draw, both teams get 1 point.

Since the final standings are based on the number of points and the points are given based on the three possible outcomes of a match (win, tie, or lose), we fit a cumulative logit model with $K = 3$ response categories, such that model M_0 (without considering a home advantage) corresponds to

$$\text{logit}(\mathbb{P}(Y_{rs} \leq k|(r, s))) = \theta_k + (x_1^{(r,s)}\gamma_1 + \ldots + x_{17}^{(r,s)}\gamma_{17}),\ k = 1, 2 \qquad (8.8)$$

and model M_1 (that considers a home advantage α_0) corresponds to

$$\text{logit}(\mathbb{P}(Y_{rs} \leq k|(r, s))) = \theta_k + \alpha_0 + (x_1^{(r,s)}\gamma_1 + \ldots + x_{17}^{(r,s)}\gamma_{17}),\ k = 1, 2, \qquad (8.9)$$

where $\gamma_1, \ldots, \gamma_{17}$ are the parameters that represent the 'ability' of a team, and $\gamma_{18} = 0$ (here, MSV Duisburg) is the reference team.

In football competitions there is often the question, whether the team that plays at its home location has an advantage (home advantage). This can be tested using the likelihood ratio test, in which we test $H_0 : \alpha_0 = 0$ against $H_1 : \alpha_0 \neq 0$ to see if the home advantage parameter α_0 can be neglected (Fahrmeir et al., 1994, p. 99-101). The likelihood ratio test statistic is then computed by

$$\lambda = \frac{Dev(M_0) - Dev(M_1)}{\Phi},$$

where, in this case, the dispersion parameter $\Phi = 1$. Since we test only one parameter to be equal to zero, H_0 can be rejected on 5% significance level when $\lambda > \chi^2_{0.95}(1)$.

The estimated parameters for model M_0 and model M_1 are shown in Table 8.5. For the test above, we computed the likelihood ratio test statistic by $\lambda = 18.045 > \chi^2_{0.95}(1) = 3.841$. Thus, $H_0 : \alpha_0 = 0$ can be rejected, meaning that the home advantage parameter could not be neglected.

We can follow Cattelan et al. (2013) and use the Kendall τ rank correlation to see if the ranking based on the estimated team parameters is similar to the (real) final ranking based on the points. The rank correlation of the restricted model without considering a home advantage is $\tau_{M_0} = 0.922$ and $\tau_{M_1} = 0.961$ for the unrestricted model that considers the home advantage. Thus, both models seem appropriate.

Final standings		Model M_0		Model M_1	
Team	Points	Rank	Estimate	Rank	Estimate
FC Bayern Muenchen	75	1	2.152	1	2.324
SV Werder Bremen	70	2	1.971	2	2.100
Hamburger SV	68	3	1.851	3	2.005
FC Schalke 04	61	4	1.537	4	1.648
Bayer Leverkusen	52	5	1.112	5	1.176
Hertha BSC Berlin	48	6	0.969	6	1.063
Borussia Dortmund	46	7	0.887	7	0.991
1. FC Nuernberg	44	10	0.696	9	0.801
VfB Stuttgart	43	8	0.843	8	0.946
Borussia Moenchengladbach	42	9	0.736	10	0.790
1. FSV Mainz 05	38	12	0.465	12	0.514
Hannover 96	38	11	0.679	11	0.748
Arminia Bielefeld	37	15	0.295	13	0.393
Eintracht Frankfurt	36	14	0.334	15	0.373
VfL Wolfsburg	34	13	0.377	14	0.381
1. FC Kaiserslautern	33	16	0.172	16	0.228
1. FC Koeln	30	17	0.064	17	0.115
MSV Duisburg (reference)	27	18	0.000	18	0.000
Estimate for θ_1	–	–	-0.767	–	-0.805
Estimate for α_0	–	–	–	–	0.478

Table 8.5.: The table shows the final standings, the ranking and estimates based on the restricted model M_0 (without considering a home advantage) and the ranking and estimates based on the unrestricted model M_1 (that considers the home advantage).

8.3.2. Team-specific Home Advantage

The consideration of team-specific home advantages enables the reproduction of a ranking

- for matches where the teams playing at their home location have a home advantage (which corresponds to the ranking of the home table) and
- for matches where the teams have no home advantage (which corresponds to the ranking of the away table).

We specify the model M_2 that includes team-specific home advantages as

$$\text{logit}(\mathbb{P}(Y_{rs} \leq k|(r,s))) = \theta_k + \alpha_r + (x_1^{(r,s)}\gamma_1 + \ldots + x_{17}^{(r,s)}\gamma_{17}), \ k = 1, 2, \ \forall r \neq s. \quad (8.10)$$

Using the likelihood ratio test, we can test if including team-specific home advantages α_r, $\forall r = 1, \ldots, 18$ is significantly better than including a single home advantage parameter α_0 (as in the model M_1). The test can be formulated as

$$H_0 : \alpha_1 - \alpha_0 = \ldots = \alpha_{18} - \alpha_0 = 0$$

against $H_1 : \alpha_r - \alpha_0 \neq 0$ for at least one $r \in \{1, \ldots, 18\}$.

Here, the test statistic is $\lambda = \frac{Dev(M_1) - Dev(M_2)}{1} = 15.917 < 27.587 = \chi^2_{0.95}(17)$. Thus, H_0 can not be rejected and the model M_2 with team-specific home advantages is not significantly better than the model M_1 with a single home advantage α_0 for all teams.

Table 8.6 reflects the (real) home table and contains estimates of the team's 'ability' (denoted by γ_m) and the home advantage α_m for all m teams. Since the estimates from the home table represent the 'ability' of a team when it plays at its home location, the estimates are computed by the sum of $\hat{\gamma}_m + \hat{\alpha}_m$ and the ranking is based on this sum. Accordingly, the estimates from the away table, which is given in Table 8.7, represent the 'ability' of a team when it did not play at its home location (thus, the ranking is based on $\hat{\gamma}_m$).

Kendall's $\tau_{\text{home}} = 0.935$ for the home table and $\tau_{\text{away}} = 0.908$ for the away table, therefore, the rankings based on the estimated parameters are similar to the (real) rankings based on the points of the home table and the away table, respectively.

Team	Points	$\hat{\gamma}_m$	$\hat{\alpha}_m$	$\hat{\gamma}_m + \hat{\alpha}_m$	Rank
FC Bayern Muenchen	44	2.391	1.192	3.582	1
SV Werder Bremen	39	2.337	0.693	3.029	2
FC Schalke 04	36	1.949	0.532	2.482	3
Hamburger SV	32	3.110	-1.175	1.935	5
Borussia Moenchengladbach	31	0.415	1.777	2.191	4
1. FC Nuernberg	30	0.737	1.196	1.932	6
Hertha BSC Berlin	29	1.356	0.426	1.782	7
Borussia Dortmund	28	1.400	0.286	1.686	8
Bayer Leverkusen	27	1.848	-0.194	1.654	9
Arminia Bielefeld	26	0.417	1.007	1.423	11
1. FSV Mainz 05	25	0.523	0.974	1.497	10
VfL Wolfsburg	22	0.347	1.011	1.358	12
VfB Stuttgart	22	1.653	-0.408	1.245	13
Hannover 96	21	1.319	-0.126	1.193	14
Eintracht Frankfurt	20	0.933	-0.061	0.871	17
1. FC Kaiserslautern	20	0.505	0.509	1.014	15
1. FC Koeln	19	0.491	0.347	0.838	18
MSV Duisburg (reference)	18	0.000	0.932	0.932	16

Table 8.6.: (real) home table and ranking based on $\hat{\gamma}_m + \hat{\alpha}_m$

Team	Points	$\hat{\gamma}_m$	Rank
Hamburger SV	36	3.110	1
SV Werder Bremen	31	2.337	3
FC Bayern Muenchen	31	2.391	2
Bayer Leverkusen	25	1.848	5
FC Schalke 04	25	1.949	4
VfB Stuttgart	21	1.653	6
Hertha BSC Berlin	19	1.356	8
Borussia Dortmund	18	1.400	7
Hannover 96	17	1.319	9
Eintracht Frankfurt	16	0.933	10
1. FC Nuernberg	14	0.737	11
1. FSV Mainz 05	13	0.523	12
1. FC Kaiserslautern	13	0.505	13
VfL Wolfsburg	12	0.347	17
1. FC Koeln	11	0.491	14
Borussia Moenchengladbach	11	0.415	16
Arminia Bielefeld	11	0.417	15
MSV Duisburg (reference)	9	0.000	18

Table 8.7.: (real) away table and ranking based on $\hat{\gamma}_m$

9. Discussion and Outlook

In this thesis, we described methods that enable making preference rankings for objects that are compared pairwise. The estimated object parameters reflect the 'ability' or 'worth' of an object and are assumed to be fixed values. It is also possible to assume that the object parameters vary over time, which is an approach that was considered by Fahrmeir and Tutz (1994) and more recently by Cattelan et al. (2013). This can make sense in terms of sport competitions if the 'ability' of a team (or player) is assumed to be time-dependent, e.g. if the 'ability' of a team (or player) improves over time.

We also handled the case when subject-specific covariates and order effects (or, equivalently, home advantages) are included in the model. However, for the subject-specific covariates we used only linear effects, which might be too restrictive in some application fields, but theoretically, non-linear effects could also be considered (see Francis et al., 2002). The model considered by Cattelan et al. (2013) is being described as a *'convenient way to induce home abilities which vary between teams and in time depending on past performances of teams'.* In their paper, the authors state that the inclusion of team-specific home advantages is controversial in the literature since different researchers make different conclusions about whether team-specific home advantages should be considered or whether supposing the same home advantage for all teams is appropriate. The same statement can also be found in Cattelan et al. (2010), who also criticised the method of including team-specific home advantages as we have done in Section 4.2.2 due to the additional number of parameters that have to be estimated. Since team-specific home advantages are included in the model as dummy variables (see Section 4.2.2), one can apply the boosting algorithm for variable selection (proposed in Chapter 6) to overcome the criticised issue, such that only important team-specific home advantages might be included.

The inclusion of object-specific covariates was not considered in this thesis. The easiest way to include object-specific covariates is to replace the object parameters γ_m by a linear combination of object-specific covariates $\gamma_m = \sum_{o=1}^{O} \delta_o x_{(m)o}$, where

$x_{(m)1}, \ldots, x_{(m)O}$ are the object-specific covariates and $\delta_1, \ldots, \delta_O$ are the corresponding parameters (Hatzinger and Dittrich, 2012). However, this will only allow for making conclusions about the effect of object-specific covariates (without considering an effect for the objects themselves) and makes sense if the object-specific covariates are assumed to fully represent the 'worth' or 'ability' of the objects.

Threshold covariates, where the thresholds are assumed to be dependent from explanatory variables, could also be considered in the model. This was implicitly done for the adjacent categories logit model with 3 response categories in Section 8.2.3. However, the interpretation of threshold covariates for the cumulative link model or the adjacent categories logit model with more than 3 response categories is more complicated, since more than a single threshold parameter has to be estimated.

The results of the simulation study in Chapter 7 and the application of the boosting algorithm on a real data set in Section 8.2.3 suggest that the algorithm seems to work. However, we used only linear effects for the covariates (linear base learners). Above, we mentioned that non-linear effects could also be included. Thus, further work will be required to extend the `BTLboost` algorithm to also include non-linear base learners.

Bibliography

Agresti, A. (1992). Analysis of ordinal paired comparison data. *Applied Statistics*, pages 287–297.

Agresti, A. (2002). *Categorical data analysis*, volume 359. Wiley-interscience.

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345.

Bühlmann, P. and Hothorn, T. (2007). Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science*, 22(4):477–505.

Bühlmann, P. and Hothorn, T. (2008). Rejoinder: Boosting Algorithms: Regularization, Prediction and Model Fitting. *ArXiv e-prints*.

Bühlmann, P. and Yu, B. (2003). Boosting with the l_2 loss: regression and classification. *Journal of the American Statistical Association*, 98(462):324–339.

Casalicchio, G. (2013). *ordBTL: Modelling comparison data with ordinal response.* R package version 0.7.

Cattelan, M. (2012). Models for Paired Comparison Data: A Review with Emphasis on Dependent Data. *Statistical Science*, 27(3):412–433.

Cattelan, M., **Varin, C.**, **and Firth, D. (2010)**. Stochastic dynamic thurstone-mosteller models for sports tournaments.

Cattelan, M., **Varin, C.**, **and Firth, D. (2013)**. Dynamic bradley–terry modelling of sports tournaments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(1):135–150.

Christensen, R. (1990). Log-linear models.

David, H. A. (1988). *The Method of Paired Comparisons.* Oxford University Press, New York, 2 edition.

Davidson, R. R. (1970). On extending the bradley-terry model to accommodate ties in paired comparison experiments. *Journal of the American Statistical Association*, 65(329):317–328.

Dittrich, R., **Francis, B.**, **Hatzinger, R.**, **and Katzenbeisser, W. (2007)**. A paired comparison approach for the analysis of sets of likert-scale responses. *Statistical Modelling*, 7(1):3–28.

Dittrich, R., **Hatzinger, R.**, **and Katzenbeisser, W. (1998)**. Modelling the effect of subject-specific covariates in paired comparison studies with an application to university rankings. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(4):511–525.

Dittrich, R., **Hatzinger, R.**, **and Katzenbeisser, W. (2001)**. Corrigendum: Modelling the effect of subject-specific covariates in paired comparison studies with an application to university rankings. *Journal of the Royal Statistical Society Series C*, 50(2):247–249.

Dittrich, R., **Hatzinger, R.**, **and Katzenbeisser, W. (2004)**. A log-linear approach for modelling ordinal paired comparison data on motives to start a phd programme. *Statistical Modelling*, 4(3):181–193.

Fahrmeir, L. and Tutz, G. (1994). Dynamic stochastic models for time-dependent ordered paired comparison systems. *Journal of the American Statistical Association*, 89(428):1438–1449.

Fahrmeir, L., **Tutz, G.**, **and Hennevogl, W. (1994)**. *Multivariate statistical modelling based on generalized linear models*, volume 2. Springer New York.

Francis, B., **Dittrich, R.**, **Hatzinger, R.**, **and Penn, R. (2002)**. Analysing partial ranks by using smoothed paired comparison methods: an investigation of value orientation in Europe. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 51(3):319–336.

Freund, Y. (1990). Boosting a weak learning algorithm by majority. In *COLT*, volume 90, pages 202–216.

Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119 – 139.

Friedman, J., **Hastie, T.**, **and Tibshirani, R. (2000)**. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics*, 28(2):337–407.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, pages 1189–1232.

Glenn, W. and David, H. (1960). Ties in paired-comparison experiments using a modified thurstone-mosteller model. *Biometrics*, 16(1):86–109.

Goos, P. and Großmann, H. (2011). Optimal design of factorial paired comparison experiments in the presence of within-pair order effects. *Food Quality and Preference*, 22(2):198–204.

Harville, D. A. and Smith, M. H. (1994). The home-court advantage: How large is it, and does it vary from team to team? *The American Statistician*, 48(1):pp. 22–28.

Hatzinger, R. and Dittrich, R. (2012). prefmod: An r package for modeling preferences based on paired comparisons, rankings, or ratings. *Journal of Statistical Software*, 48(10):1–31.

Hofner, B. (2011). *Boosting in structured additive models*. PhD thesis, München, Ludwig-Maximilians-Universität, Diss., 2011.

Luce, R. D. (1959). *Individual Choice Behavior a Theoretical Analysis*. John Wiley and sons.

Masarotto, G. and Varin, C. (2012). The ranking lasso and its application to sport tournaments. *The Annals of Applied Statistics*, 6(4):1949–1970.

Mosteller, F. (1951). Remarks on the method of paired comparisons: I. the least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, 16(1):3–9.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available from: `http://www.R-project.org/`.

Rao, P. V. and Kupper, L. L. (1967). Ties in paired-comparison experiments: A generalization of the bradley-terry model. *Journal of the American Statistical Association*, 62(317):194–204.

Schapire, R. E. (1990). The strength of weak learnability. *Machine learning*, 5(2):197–227.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological review*, 34(4):273.

Turner, H. and Firth, D. (2010). Bradley-terry models in r: The bradleyterry2 package. *Biometrika*, 714(730):498.

Tutz, G. (1986). Bradley-terry-luce models with an ordered response. *Journal of mathematical psychology*, 30(3):306–316.

Tutz, G. (1989). *Latent Trait-Modelle für ordinale Beobachtungen: die statistische und messtheoretische Analyse von Paarvergleichsdaten*. Lehr- und Forschungs-texte Psychologie. Springer.

Tutz, G. and Binder, H. (2007). Boosting ridge regression. *Computational Statistics & Data Analysis*, 51(12):6044–6059.

Tutz, G., **Binder, H.**, **Discussion, A.**, **and Evans, R. (2006)**. Generalized additive modeling with implicit variable selection by likelihood-based boosting. *Biometrics*, 62(2006):961–971.

Wilkinson, G. and Rogers, C. (1973). Symbolic description of factorial models for analysis of variance. *Applied Statistics*, pages 392–399.

Zahid, F. M. and Tutz, G. (2011). Proportional odds models with high-dimensional data structure.

A. R-Code

A.1. `BTLboost` Implementation

Listing A.1: R function for `BTLboost` algorithm

```r
BTLboost <- function(formula, data, mstop=1000, selection=c("AIC","BIC"),
                     objects, stopif=3, parallel=TRUE, verbose=TRUE, ...){

  require(gtools)
  require(VGAM)
  selection <- match.arg(selection)
  mc <- match.call()
  mc[[1]] <- as.name("ordBTL")
  rm <- c("mstop", "selection","stopif","objects",
          "parallel","verbose")
  mc[rm] <- NULL
  mc$noWarning <- TRUE

  # model formulas
  allVars <- attr(terms(formula),"term.labels")
  pool <- allVars[!allVars%in%objects]
  response <- as.character(formula)[2]
  formobjects <- paste(response, paste(objects, collapse="+"), sep="~")
  formCandidate <- paste(response, paste(pool, "1", sep="-"), sep="~")

  # preallocation
  base <- vector("list", length=length(formCandidate))
  metric <- nparam <- numeric(length(formCandidate))
  crit <- selected <- rep(NA, length(mstop))

  ### Step 1: get initial values (intercept model)
  mc$formula <- as.formula(paste(response,"~", "1"))
```

```r
28  intercept <- eval(mc)
29  f <- intercept@predictors
30  param <- coefficients(intercept)
31  paramlist <- rbind(NA, param)
32
33  # Start loop for m=1,..., mstop
34  for(j in 1:mstop){
35    ### Step 2A: object predictors
36    mc$data <- data
37    mc$formula <- as.formula(formobjects)
38    mc$offset <- mc$etastart <- f
39    mc$maxit <- 1
40    obligatory <- eval(mc)
41    # update step
42    f <- f + obligatory@predictors
43    coefs <- smartbind(param, coefficients(obligatory))
44    param <- colSums(coefs, na.rm=TRUE)
45
46    ### Step 2B: candidate predictors
47    if(parallel){ # parallel computation
48      base <- foreach(i=1:length(formCandidate),
49                      .packages="ordBTL") %dopar% {
50        mc$formula <- as.formula(formCandidate[i])
51        mc$offset <- mc$etastart <- f
52        mc$maxit <- 1
53        mc$data <- data
54        base <- eval(mc)
55      }
56    }else{ # sequential computation (slower)
57      for(i in 1:length(formCandidate)) {
58        mc$formula <- as.formula(formCandidate[i])
59        mc$offset <- mc$etastart <- f
60        base[[i]] <- eval(mc)
61      }
62    }
63
64    # stopping and selection criterion
65    nparam <- sapply(pool, function(X){length(unique(c(names(param), X)))
         })
```

```
66    if(selection=="AIC"){
67      metric <- sapply(base, function(X){deviance(X)}) + 2*nparam
68    } else{
69      metric <- sapply(base, function(X){deviance(X)}) + log(nobs(
             intercept))*nparam
70    }
71
72    # update step
73    best <- which.min(metric)
74    coefs <- smartbind(param, coefficients(base[[best]]))
75    param <- colSums(coefs, na.rm=TRUE)
76    paramlist <- smartbind(paramlist, param)
77    f <- f + base[[best]]@predictors
78    crit[j] <- metric[best]
79    selected[j] <- pool[best]
80
81    # print iteration
82    if(verbose==TRUE){
83      cat(sprintf("% 5.0f", j),
84          paste(selection, ":", sep=""), sprintf("%.3f", crit[j]),
85          "| vars:", sprintf("% 3.0f", length(param)),
86          "| updated:", pool[best], "\n")
87    }
88
89    # stop earlier
90    if(length(na.omit(crit))>=stopif &&
91        !is.unsorted(window(crit, j-(stopif-1),j), na.rm=TRUE)){
92      cat("Finished:", "No improvement of", selection, "in the last",
             stopif, "steps \n")
93      return(list(BEST=paramlist[which.min(crit)+2,], CRITERION=crit,
94                 PATH=paramlist, UPDATED=selected))
95    }
96  }
97  return(list(BEST=paramlist[which.min(crit)+2,], CRITERION=crit,
98            PATH=paramlist, UPDATED=selected))
99 }
```

A.2. Applications

Listing A.2: R code for models from Section 8.1

```
1  library(ordBTL)
2  ################################################################
3  # Reproduce model from Table 8.1
4  ################################################################
5  # Data from Table 1 of Agresti (1992)
6  data(ribbon)
7
8  # design matrix
9  des3 <- design(ribbon, var1="obj1", var2="obj2", use.vars="ALL")
10
11 # Note that Agresti (1992) used the constraint that the object
12 # parameters sum up to 1. To get the same results, we use the model
13 form3 <- cbind(V1,V2,V3,V4,V5,V6,V7)~I(GAMMA.1-GAMMA.5)+
14   I(GAMMA.2-GAMMA.5)+I(GAMMA.3-GAMMA.5)+I(GAMMA.4-GAMMA.5)
15
16 # Fit the adjacent categories logit model
17 ac <- ordBTL(form3, data=des3, family="acat",
18             family.control=list(reverse=TRUE))
19
20 # Fit the cumulative logit model
21 clm.logit <- ordBTL(form3, data=des3)
22
23 # Fit the cumulative probit model
24 clm.probit <- ordBTL(form3, data=des3,
25                   family.control=list(link="probit"))
26
27 # Parameter estimates
28 coefs <- t(rbind("Adjacent categories logit"=coefficients(ac),
29             "Cumulative probit"=coefficients(clm.probit),
30             "Cumulative logit"=coefficients(clm.logit)))
31 coefs <- rbind(coefs, "GAMMA.5"=0-colSums(coefs[4:7,]))
32 coefs
```

Listing A.3: R code for models from Section 8.2

```
library(ordBTL)
###############################################################
# Reproduce model from Table 8.3
###############################################################
# Get the CEMS data and generate design matrix
example(wide2long, package="ordBTL", echo=FALSE)
des1 <- design(CEMSlong, var1="object1", var2="object2",
               use.vars="Y", reference="Stockholm")
# Fit the adjacent categories model
mod1 <- ordBTL(Y~., data=des1, family="acat",
               family.control=list(reverse=TRUE))
# We get the same results from Table 3 of Dittrich et al (2001).
getRank(mod1)


###############################################################
# Apply BTLboost algorithm on CEMS data (reproduce Table 8.4)
###############################################################
# Generate design matrix and specify model formula
des2 <- design(CEMSlong[-which(is.na(CEMSlong$Y)),],
               var1="object1", var2="object2",
               use.vars="ALL", reference="Stockholm")
form.all <- Y ~ (GAMMA.London + GAMMA.Paris + GAMMA.Milano +
                 GAMMA.StGallen + GAMMA.Barcelona)*
                (WOR +SEX + DEG + STUD + ENG + FRA + SPA + ITA)
objects <- c("GAMMA.London", "GAMMA.Paris","GAMMA.Milano",
             "GAMMA.StGallen", "GAMMA.Barcelona")
subject.covs <- c("ENG","FRA","SPA","ITA","WOR","DEG","SEX","STUD")

# convert subject-specific covariates to 0-1 dummy variables
for(i in subject.covs) des2[,i] <- as.numeric(as.character(des2[,i]))-1

# You have to load the BTLboost function from Appendix A.1 first
AICcems <- BTLboost(form.all, data=des2, family="acat", parallel=FALSE,
                    family.control=list(reverse=TRUE), objects=objects,
                    restrict=subject.covs, selection="AIC")
AICcems$BEST
```

Listing A.4: R code for models from Section 8.3

```
 1 library(ordBTL)
 2 #############################################################
 3 # Reproduce models from Table 8.5
 4 #############################################################
 5 # load data
 6 example(design, package="ordBTL", echo=FALSE)
 7
 8 # Model without home advantage
 9 des.nohome <- design(buli0506, var1="Heim", var2="Gast",
10                      use.vars="Y3", home.advantage="no",
11                      reference="GAMMA.MSV.Duisburg")
12 mod.nohome <- ordBTL(Y3~., data=des.nohome)
13 # team 'abilities' (should be the ranking of the final standings)
14 getRank(mod.nohome, prefix="GAMMA", reference="GAMMA.MSV.Duisburg")
15
16 # Model with home advantage
17 des.onehome <- design(buli0506, var1="Heim", var2="Gast",
18                      use.vars="Y3", home.advantage="yes",
19                      reference="GAMMA.MSV.Duisburg")
20 mod.onehome <- ordBTL(Y3~., data=des.onehome)
21 # team 'abilities'
22 getRank(mod.onehome, prefix="GAMMA", reference="GAMMA.MSV.Duisburg")
23 # home advantage
24 getRank(mod.onehome, prefix="ALPHA")
25
26 #############################################################
27 # Reproduce model from Table 8.6 and Table 8.7
28 #############################################################
29 # Model with team-specific home advantage
30 des.teamhome <- design(buli0506, var1="Heim", var2="Gast",
31                       use.vars="Y3", home.advantage="specific",
32                       reference="GAMMA.MSV.Duisburg")
33 mod.teamhome <- ordBTL(Y3~., data=des.teamhome)
34 # team 'abilities' (should be the ranking for the away table)
35 getRank(mod.teamhome, prefix="GAMMA", reference="GAMMA.MSV.Duisburg")
36 # team-specific home advantages
37 getRank(mod.teamhome, prefix="ALPHA")
```

B. Structure of the CD-ROM

The attached CD-ROM contains the following files and file folders:

- `Figures\`: Folder containing all created graphics (as `.pdf` file) used in this thesis.
- `Simulation\`: Folder containing the following files:
  - `SimData\`: Subfolder containing all `.RData` files that are created from the files listed below. They include the simulated data sets and the results.
  - `simFun.R`: Functions used for simulations.
  - `AICsim.R`: R code for Scenario 1 and 2 with AIC as stopping criterion.
  - `BICsim.R`: R code for Scenario 1 and 2 with BIC as stopping criterion.
  - `Scenario3AIC.R`: R code for Scenario 3 with AIC as stopping criterion.
  - `Scenario3BIC.R`: R code for Scenario 3 with BIC as stopping criterion.
  - `simulation_results.R`: R code for tables and figures from Chapter 7.
- `Applications\`: Contains the following files:
  - `AppFun.R`: Functions used for applications.
  - `ribbon.R`: Results for Section 8.1.
  - `cems.R`: Results for Section 8.2.
  - `AICcemsBoost.RData`: Saved results for Table 8.4 generated by `cems.R`.
  - `football.R`: Results for Section 8.3.
  - `home_away_tab.RData`: Saved home table and away table.
- `MA_Casalicchio.pdf`: The present thesis as `.pdf` file.

Affidavit

I hereby declare that the present master thesis has been written only by the undersigned and without any assistance from third parties. Furthermore, I confirm that no sources have been used in the preparation of this thesis other than those indicated in the thesis itself.

Munich, November 3, 2013

....
(Giuseppe Casalicchio)

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Masterarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

München, den 3. November 2013

....
(Giuseppe Casalicchio)