Modelling Assumed Metric Paired Comparison Data – Application to Learning Related Emotions

Alexandra Grand

WU - Vienna University of
Economics and Business

Regina Dittrich

WU - Vienna University of
Economics and Business

Abstract

In this article we suggest a beta regression model that accounts for the degree of preference in paired comparisons measured on a bounded metric paired comparison scale. The beta distribution for bounded continuous random variables assumes values in the open unit interval $(0, 1)$. However, in practice we will observe paired comparison responses that lie within a fixed or arbitrary fixed interval $[-a, a]$ with known value of a. We therefore transform the observed responses into the interval $(0, 1)$ and assume that these transformed responses are each a realization of a random variable which follows a beta distribution. We propose a simple paired comparison regression model for beta distributed variables which allows us to model the mean of the transformed response using a linear predictor and a logit link function – where the linear predictor is defined by the parameters of the logit-linear Bradley-Terry model. For illustration we applied the presented model to a data set obtained from a student survey of learning related emotions in mathematics.

Keywords: beta regression, logistic Bradley-Terry model, metric paired comparisons.

1. Introduction

The method of paired comparisons is a well established method for analysing preferences in many sciences. In a paired comparison study individuals are asked to repeatedly judge which of the pairwise presented objects (of a set of J objects) they prefer according to an attribute. The aim is to obtain an ordering of the objects on a preference continuum. In each of the $\binom{J}{2}$ paired comparisons (denoted by (jk)) there are two possible decisions: preference for object j, $jk(j)$, or preference for object k, $jk(k)$. For the analysis of such paired comparison data we refer to the well known Bradley-Terry (BT) model (Bradley and Terry 1952), see Equation (4).

Paired comparisons with only two response categories are easy to perform but in a way they forces the judge to decide for one of the two presented objects, which could be questionable if the judge has no preference. Several authors (e.g. Rao and Kupper 1967; Kousgaard 1976; Dittrich, Hatzinger, and Katzenbeisser 1998) have extended the BT model to incorporate a third category (ties) for indifference, i.e. no preference. To obtain more information the BT model has further been extended by allowing the judge to specify the degree of preference on an ordered response scale – for example on a seven point scale with labelled ordered response

categories: *strong preference for object j*, *moderate preference for j*, *mild preference for j*, *no preference*, *mild preference for k*, *moderate preference for k*, *strong preference for object k* (see e.g. Agresti 1992; Böckenholt and Dillon 1997; Dittrich, Hatzinger, and Katzenbeisser 2004; Dittrich, Francis, Hatzinger, and Katzenbeisser 2007). However, judges may differ in their interpretation of labelled response categories or in their response scale usage. Little literature was found about metric paired comparisons where the judge is asked to report the degree of preference for a particular object in a given comparison (jk) on a bounded metric scale. Although there still might exist the problem of possible response style bias, more (detailed) information may be obtained and individuals often felt to be more consistent when using continuous response formats (McKelvie, 1978 cited in Ferrando 2002).

De Ruiz (1990), for example, proposed a logistic model for a paired comparison experiment on a continuum of response where the parameters are estimated with an integrated minimum mean squared error approach. Stern (2011) suggested a model that accounts for continuous paired comparison data. In this study cricket teams (the objects) were pairwise compared in the matches according to a predefined rule to receive the *margin* (i.e. magnitude) of victory. In each cricket match (paired comparison) Stern (2011) obtained the margin of victory by calculating the proportion of available resources which were unused by the winning team. The sign of the margin of victory indicates which team won the match. Stern (2011) treated this paired comparison outcome which was transformed to take values in the unit interval (0, 1), as the response variable in a beta regression model. The parameters of interest (the relative strengths of the cricket teams) were estimated using a penalized log-likelihood method. This penalty function allows specification of the trade-off between the relative importance of the information that a team wins or loses (i.e. the dichotomous outcome) and the continuous margin of victory (the degree).

Following the approach of Stern (2011), the purpose of this article is to suggest a simple beta regression model for (assumed) metric paired comparison data without using a penalty function, where only the degree of preference should be indicative of the preference parameters. In each paired comparison the decision has to be communicated respectively specified on a given response scale with (arbitrary) fixed bounds. We assume that the more one object dominates the other according to a particular attribute the more likely a judge states a high degree of preference for this object in a given paired comparison. The responses made are bounded and (assumed to be) interval scaled so that we found the assumption of a beta distributed response variable appropriate. Note that the response variable still has to be transformed to lie within the open unit interval (0, 1), see Section 2.1. The proposed paired comparison approach for (assumed) metric responses is suitable in situations where a *relative* preference ordering is of interest.

A practical example for the usage of a bounded (assumed) metric response scale in a paired comparison study would be the measurement of relative preferences for J different services (the objects) in tourism where the judge is asked to state on an assumed metric response scale how much more she/he would prefer one of the pairwise presented services. We assume that in the process of coming up with a mark on a bounded response scale in each paired comparison, a person first imagines the worth of each of the two services and then compare these. Finally the judge has to communicate respectively specify her/his degree of preference for one of the objects being compared on a given (arbitrary) bounded response scale. Let us assume that person A would appreciate the possibility of a rent-a-car service and do not require wireless LAN. Confronting person A in a paired comparison with these two services, person A might quantify her/his preference for the rent-a-car service somewhere at the end of the response scale indicating a very high preference for this service compared to the other one. As a result, over all paired comparisons and persons, we obtain estimates of worth parameters (see Section 2.1) for each of the J services which can be located on a continuum. A high value of the estimated worth parameter indicates high preference for a certain service in relation to the other services. In many fields in tourism it is unrealistic to serve all possible services tourists would wish to receive, so a specialization in a few of them is required (due

to economical and other reasons).

Therefore it might be interesting which of the services of a set of J services to pursue and specialize in. In such a case an ordering of the services on a continuum would be a guidance, where also distances between the J services can be reasonably interpreted. Using another approach like a rating scale of preference, we might observe the possible phenomenon that when asking judges they rate lots of services equally high which is less informative. Choosing a paired comparison approach the persons are forced to make several relative, pairwise judgements about the objects and therefore to think more about their responses.

Another example could be a *derived* paired comparison study where the time persons need for solving J mathematical tasks (the objects) is collected. And where in a second step the tasks are pairwise compared by computing for each individual the time difference, i.e. the degree of solving one maths task more rapid than the other. This relative time degree (the derived response) can then be located on a bounded metric paired comparison scale. As these paired comparisons will not originate from real paired comparisons, we term this kind of data *derived* paired comparison data. On basis of these derived responses worth parameters for each of the J maths tasks can be estimated. They can be located on a time continuum where high values indicate that a long time is needed to solve a certain task (implying that this is a relative difficult task) relative to the other tasks and vice versa. As a result, we could obtain information about the rank order (e.g.: the most and least time intense maths task of a set of J maths tasks, respectively) and interpret the distances between the parameters of the maths tasks (e.g.: the distance between the first and second ranked maths task is larger than between the second and third ranked maths task). Moreover we could check if the ordering of the maths task parameters changes for various groups of students (e.g. female students might solve a certain task significantly quicker than male students). The maths tasks can further be classified according to one or more attribute(s) maths tasks can have and incorporate object-specific covariates into the model. Such information might also be useful for psychological assessments or for longitudinal studies, for example.

In this article we applied the beta regression model to a data set of a study concerning learning related emotions in maths. The application (see Section 3) covers paired comparisons where students were asked to mark how much more often one emotion is typically experienced while learning compared to another on a visual scale with arbitrary fixed bounds. Here the emotions are the objects of interest. We still have to bear in mind that we do not know exactly if the same marks of different judges on a visual horizontal line imply that they have experienced the same degree of frequency (see Aitken 1969). However, we assume that the responses are ordered for all judges and on an interval scale level. We term such a scale an *assumed* metric paired comparison response scale.

Emotions are very subjective, complex interrelated constructs for which it is very difficult to come up with suitable measures. Using a traditional IRT approach we would require item sets each measuring one (latent) emotion of interest and the quality of the items has to be checked according to the assumptions made by the IRT models. The responses in these models are sort of *absolute* in the sense that a subject might agree (or not agree) or an item can be solved (or not solved). One model assumption in the IRT approach is that the probability for a response depends on item as well as subject parameters. If the model holds we would get *unidimensional* interval scaled measures for each of the latent constructs of interest separately. The assumption of unidimensionality requires that the item parameters are the same (up to a constant) for all subgroups of respondents. Differences between subgroups would indicate IRT model violations in this context.

Furthermore, in the framework of Item Response Theory (IRT) only a few authors have analysed (bounded) continuous responses (see e.g. Noel and Dauvier 2007 or Müller 1987). In these approaches the respondents were asked for *absolute continuous* judgements for each item.

However, we are not interested in scaling emotions but in directly modelling *relative* responses by comparing emotions to obtain an ordering of a set of J emotions on an interval scaled continuum where we could also interpret the distances between the estimated parameters representing the emotions (i.e. the objects) of interest. In the paired comparison approach the probability for a response only depends on the strength (or worth) of the involved emotions and the effect of the subject characteristics can be modelled explicitly by interaction effects between emotions and subject characteristics. Therefore the proposed metric paired comparison model allows the incorporation of subject covariates, where possible effects of the subject covariates on the ordering of the emotion parameters can be assessed. This can be seen as an advantage of the paired comparison approach compared to the IRT models where item differences between subgroups are interpreted as model violations. As in the IRT models it is also possible to incorporate object-specific covariates. Model selection can easily be done through a likelihood ratio test of nested models.

2. Beta regression model

The beta distribution is restricted to the standard unit interval $(0, 1)$. Let Y be a beta distributed random variable with realization y; the density of Y is then given by:

$$f(y; p, q) = \frac{1}{B(p, q)} y^{p-1}(1-y)^{q-1} = \frac{\Gamma(p+q)}{\Gamma(p) \cdot \Gamma(q)} y^{p-1}(1-y)^{q-1} \ , \ 0 < y < 1 \ . \tag{1}$$

Here, the shape parameters of the beta density are symbolized with p and q, where $p > 0, q > 0$. The beta function $B(\cdot, \cdot)$ is related to the gamma function $\Gamma(\cdot)$ by $B(p, q) = \frac{\Gamma(p) \cdot \Gamma(q)}{\Gamma(p+q)}$ and the term $\frac{1}{B(p,q)}$ is a normalizing constant so that (1) integrates to unity. Several authors substituted one (e.g. Kieschnick and McCullough 2003) or both (e.g. Paolino 2001; Ferrari and Cribari-Neto 2004; Smithson and Verkuilen 2006) shape parameters of the beta distribution (1) and imposed a regression structure for the mean response or for both the mean and the precision parameter to simplify interpretation.

The beta regression model, as introduced by Ferrari and Cribari-Neto (2004), is an alternative parameterization of the beta density for modelling continuous response variables restricted to $(0, 1)$. To be able to model the mean of the response variable as a function of explanatory variables (via a linear predictor) along with a precision parameter, Ferrari and Cribari-Neto (2004) proposed the different parameterization of the beta density:

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi) \cdot \Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} \ , \ 0 < y < 1 \ ,$$

where $\mu = \frac{p}{p+q}$ and $\phi = p+q$ (i.e. $p = \mu\phi$ and $q = (1-\mu)\phi$), $0 < \mu < 1$, $\phi > 0$. The parameter μ is a location parameter (the mean of the response variable), ϕ is a precision parameter and ϕ^{-1} a dispersion parameter; the expectation of Y, $\text{E}(Y) = \mu$ and $\text{VAR}(Y) = \frac{\mu(1-\mu)}{1+\phi}$.

In this article we refer to a beta regression model that includes a linear predictor and a link function for both the location parameter μ and the precision parameter ϕ. This model is similar to a generalized linear model (GLM). For a random sample we write $Y_i \sim B(\mu_i, \phi_i)$, $i = 1, \ldots, N$, where N is the sample size. The beta regression model is defined as (cf. Smithson and Verkuilen 2006; Cribari-Neto and Zeileis 2010; Simas, Barreto-Souza, and Rocha 2010):

$$\begin{aligned} g_1(\mu_i) &= \boldsymbol{x}_i^T \boldsymbol{\lambda} = \eta_{1i} \ , \\ g_2(\phi_i) &= \boldsymbol{z}_i^T \boldsymbol{\gamma} = \eta_{2i} \ , \end{aligned} \tag{2}$$

where $g(\cdot)$ is a link function. $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)^T$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_h)^T$, $k + h < N$, are $k \times 1$ and $h \times 1$ column vectors of unknown regression parameters, respectively. $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ are covariate vectors: $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ik})^T$, $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{ih})^T$ and η_{1i} and η_{2i} are the linear predictors.

The notation of a beta regression model as defined in (2) is suitable for modelling paired comparison data, where we want to model the mean of the response made on a bounded metric scale (i.e. the dependent variable) as a function of a set of covariates (independent variables) and parameters of the objects (λ) via a linear predictor. In the following section we present a beta regression model for paired comparisons where we assume that the precision parameter ϕ is constant for all observations, i.e. $g_2(\phi) = \gamma$.

2.1. Beta regression model for paired comparison data

In most of the paired comparison studies the (assumed) metric responses, the $y_{jk,i}$'s, of the ith individual, $i = 1, \ldots, N$ in the $\binom{J}{2}$ comparisons, will lie in an interval $[-a, a]$. The bounds $-a$ and a are predefined with known value of a which refer to theoretical or arbitrary values. Following the proposed transformation method of Smithson and Verkuilen (2006) we can transform each random variable $Y_{jk,i}$ of a given paired comparison (jk) of person i into the standard unit interval $(0, 1)$ by two steps:

First we squeeze $Y_{jk,i}$ with realization $y_{jk,i}$ in $[-a, a]$ into the interval $[0, 1]$ and obtain the transformed random variable $Y_{jk,i}^*$ by means of the transformation:

$$Y_{jk,i}^* = \frac{Y_{jk,i} + a}{2a}, \quad \text{with} \quad 0 \le Y_{jk,i}^* \le 1 \ .$$

In a second step we make sure that the two times transformed random variable $Y_{jk,i}^{**}$ cannot take on the values zero and one, i.e. takes on values in the interval $(0, 1)$:

$$Y_{jk,i}^{**} = \frac{Y_{jk,i}^* \cdot (N - 1) + 0.5}{N}, \quad \text{with} \quad 0 < Y_{jk,i}^{**} < 1 \ ,$$

where N is the sample size. In our study the $Y_{jk,i}$'s are always (i.e. in each paired comparison (jk)) associated with the first object (j) being compared. For example: In the interval $[-50, 50]$, an observed response $y_{jk,i} = -40$ denotes that object j is 40 units preferred compared to object k by judge i. And $y_{jk,i} = 40$ indicates that object k is 40 units preferred compared to object j. For further discussion about the transformation see e.g. the supplementary material of Smithson and Verkuilen (2006). As a result of the transformation process we obtain values in the interval $(0, 1)$, where values < 0.5 indicate preference for the first object (j) in a paired comparison (jk) and values > 0.5 preference for the second object (k). The closer the value to zero the greater the degree of preference for object j (and vice versa) where the most and the least favourable response for object j is about 0 and about 1. A tied response (i.e. no preference; $y_{jk,i} = -a + \frac{a+a}{2} = 0$, the middle of the response scale) is indicated by the two times transformed value of $y_{jk,i}^{**} = 0.5$.

In the case of a response variable $Y_{jk,i}$ with observation $y_{jk,i}$ on a bounded $[-a, a]$ metric paired comparison scale, let the transformed variable $Y_{jk,i}^{**}$ be a beta distributed random variable, $Y_{jk,i}^{**} \sim B(\mu_{jk,i}, \phi)$. We further assume independence between the decisions of the N judges and between the paired comparisons.

The log-likelihood function over all paired comparisons (jk) (where $j < k; j = 1, 2, \ldots, J - 1; k = 2, 3, \ldots, J$) and over all N judges is defined by:

$$\ell(\mu, \phi) = \sum_{i=1}^{N} \sum_{j=1}^{J-1} \sum_{k=j+1}^{J} \ell_{jk,i}(\mu_{jk,i}, \phi) \ , \tag{3}$$

where the log-likelihood, $\ell_{jk,i}(\mu_{jk,i}, \phi)$ associated with the transformed response $Y_{jk,i}^{**}$ of judge i in a given paired comparison (jk), is given by:

$$\ell_{jk,i}(\mu_{jk,i}, \phi) = \ln \Gamma(\phi) - \ln \Gamma(\mu_{jk,i}\phi) - \ln \Gamma((1 - \mu_{jk,i})\phi) + (\mu_{jk,i}\phi - 1) \ln y_{jk,i}^{**} +$$
$$\{(1 - \mu_{jk,i})\phi - 1\} \ln(1 - y_{jk,i}^{**}), \quad y_{jk,i}^{**} \in (0, 1) \ .$$

As link function $g(\cdot)$ for $\mu_{jk,i}$, we choose the logit link and for ϕ the identity link:

$$g_1(\mu_{jk,i}) = \text{logit}\,(\mu_{jk,i}) = \ln\left(\frac{\mu_{jk,i}}{1 - \mu_{jk,i}}\right) = \eta_{1jk,i}\ ,$$

$$g_2(\phi) = \gamma\ .$$

In general, the expectation $\mu_{jk,i}$ of $Y_{jk,i}^{**}$ can be obtained by inverting the logit link function:

$$\mu_{jk,i} = g_1^{-1}(\eta_{1jk,i}) = \frac{\exp(\eta_{1jk,i})}{1 + \exp(\eta_{1jk,i})}\ .$$

We apply the logit-linear Bradley-Terry model to the logistic mean structure. The Bradley-Terry model (Bradley and Terry 1952) defines the probability that object j is preferred in the comparison (jk), $P_{jk(j)}$, as follows:

$$P_{jk(j)} = \frac{\pi_j}{\pi_j + \pi_k}\ , \tag{4}$$

where the π_j's are positive worth parameters specified with the requirement that $\sum_{j=1}^{J} \pi_j = 1$. The worth parameters can be interpreted as locations of the objects on a preference continuum that ranges from zero to one. The probability of preferring object k over object j, $P_{jk(k)}$ is: $1 - P_{jk(j)}$ or $P_{jk(k)} = \frac{\pi_k}{\pi_j + \pi_k}$. Hence, the log-odds of preferring object j compared to object k is:

$$\ln\left(\frac{P_{jk(j)}}{P_{jk(k)}}\right) = \ln\left(\frac{\pi_j}{\pi_k}\right) = \lambda_j - \lambda_k = \eta_{1jk,i}\ ,$$

where $\lambda_j = \ln \pi_j$ or $\pi_j = \exp(\lambda_j)$. The parameter λ_j characterizes object j. For identifiability we set the object parameter λ_J to be zero. The λs are related to the worth parameter π by $\pi_j = \frac{\exp(\lambda_j)}{\sum_{j=1}^{J} \exp(\lambda_j)}$.

We parameterized the logistic mean structure of the beta regression model for a given paired comparison (jk) for judge i so that $\text{logit}\,(\mu_{jk,i}) = \lambda_j - \lambda_k$, where $\mu_{jk,i} = g_1^{-1}(\lambda_j - \lambda_k)$ is the expected degree of preference for object j compared to object k (cf. Stern 2011). The beta Bradley-Terry regression (BBTR) model for judge i for the paired comparison (jk) is defined as:

$$\text{logit}\,(\mu_{jk,i}) = \lambda_j - \lambda_k\ ,$$

$$\phi = \gamma\ .$$

Note that a tied response $(y_{jk,i}^{**} = 0.5)$ means that the judge has no preference so that $\text{logit}\,(\mu_{jk,i}) = 0$.

In general, for all comparisons (jk) and for all judges i the BBTR model is given by:

$$\text{logit}\,(\mu_{jk,i}) = \boldsymbol{x}_{jki}^T \boldsymbol{\lambda} = \eta_{1jk,i}\ ,$$

$$\phi = \gamma\ . \tag{5}$$

The vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)^T$ is a $J \times 1$ column vector of unknown object parameters, and $\boldsymbol{x}_{jki}$ is the corresponding vector of covariates for a particular paired comparison (jk) for judge i, $\boldsymbol{x}_{jki} = (x_{jki,1}, \ldots, x_{jki,J})^T$ with $x_{jki,j} \in (-1, 0, 1)$ (see Table 1). The expectation of $Y_{jk,i}^{**}$ is: $\text{E}(Y_{jk,i}^{**}) = \mu_{jk,i}$.

The parameters of the BBTR model (5) for paired comparisons can be estimated by maximizing the log-likelihood (3) using a quasi Newton method. The underlying design structure of the beta regression model (5) was constructed in R (R Development Core Team 2013) by using elements of the package `prefmod` (Hatzinger 2012). For model estimation we used the R-package `betareg` (Zeileis, Cribari-Neto, Grün, Kosmidis, Simas, and Rocha 2013; see also Cribari-Neto and Zeileis 2010).

Table 1: Design structure of a BBTR model for $J = 5$.

paired comparisons	response	λ_1	λ_2	λ_3	λ_4	λ_5
$(jk)i$	$y^{**}_{jk,i}$	$x_{jki,1}$	$x_{jki,2}$	$x_{jki,3}$	$x_{jki,4}$	$x_{jki,5}$
$(12)1$	$y^{**}_{12,1}$	1	−1	0	0	0
$(13)1$	$y^{**}_{13,1}$	1	0	−1	0	0
$(23)1$	$y^{**}_{23,1}$	0	1	−1	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
$(35)1$	$y^{**}_{35,1}$	0	0	1	0	−1
$(45)1$	0.5	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮
$(35)111$	$y^{**}_{35,111}$	0	0	1	0	−1
$(45)111$	$y^{**}_{45,111}$	0	0	0	1	−1

The design structure consists of a column with entries of the values of the transformed responses $y^{**}_{jk,i}$ and a $\binom{J}{2} \cdot N \times J$ design matrix with J columns for the objects of the paired comparison study (here labelled with $1, 2, \ldots, J$). Each row represents one of the $\binom{J}{2}$ paired comparisons (ordered according to a predefined rule: $(12), (13), (23), (14), (24), (34), (15), \ldots, (J-1, J)$) which were stacked over all judges N. Note that we set λ_J equal to be zero for parameter estimation. Example: For $J = 5$ we obtain 10 paired comparisons and for N=111 the dimension of the design matrix is 1110×5.

2.2. Extensions of the BBTR model for paired comparisons

The BBTR model (5) can be extended for modelling possible effects of categorical and numerical subject covariates on the preference ordering of objects or incorporating object-specific covariates into the model. In the following we give some examples for BBTR models, each for the paired comparison (jk) for judge i.

Examples:

One categorical subject covariate S with s levels, $s = 1, 2$ (cf. Dittrich *et al.* 1998):

$$\text{logit}\,(\mu_{jk,i|s}) = \lambda_j - \lambda_k + \lambda^{OS}_{j2} - \lambda^{OS}_{k2} + \lambda^S_1 + \lambda^S_2$$

where λ_j is the object parameter for the subject reference group, λ^{OS}_{js} is an interaction parameter between object j and subject level s and λ^S_s is the main effect of covariate S on level s (i.e. a nuisance parameter).

One numerical subject covariate (cf. Francis, Dittrich, Hatzinger, and Penn 2002):

$$\text{logit}\,(\mu_{jk,i}) = \lambda_j - \lambda_k + x_i(\beta_j - \beta_k) + \lambda^M x_i$$

where x_i corresponds to the numerical covariate for individual i, β_j describes the effect of this covariate on object j and λ^M is the main effect of the numeric subject covariate M (i.e. a nuisance parameter).

One object-specific covariate (cf. Dittrich *et al.* 1998):

$$\text{logit}\,(\mu_{jk,i}) = (x_j - x_k)\beta \;.$$

Here, the object parameters λ are replaced by $\lambda_j = x_j\beta$ where x_j is a covariate which describes a particular property of object j and β is an unknown object-specific parameter.

In all examples given we assumed that the precision parameter ϕ is constant for all observations, $g_2(\phi) = \gamma$. However, it is straightforward to specify a BBTR model for paired comparisons that allow the precision parameter to vary.

3. Application – Learning related emotions in mathematics

In the following we refer to achievement emotions (i.e. emotions that occur in achievement contexts) students typically experience when learning mathematics. Emotions are highly subjective experiences, sets of psychological processes for which it is difficult to come up with a reasonable measure (see also Pekrun, Goetz, Frenzel, Barchfeld, and Perry 2011).

In our study the emotions are the *objects* of interest for which we wanted to obtain an ordering on a continuum of frequency. In winter 2013 we carried out an online survey of students studying at the WU (Vienna university of economics and business) concerning learning related emotions in mathematics. By eliminating the response vector of a person with a missing value we received a sample size of N=111 (male=44, female=67).

In this study we considered five emotions taken from the study of Götz, Keller, and Martiny (2012): enjoyment, pride, anger, anxiety and boredom. We were interested if the ordering of learning related emotions on a frequency scale depends on particular subject variables. Several authors like Goetz, Pekrun, Zirngibl, Jullien, Kleine, vomHofe, and Blum (2004) and Frenzel, Pekrun, and Goetz (2007), for example, showed that the subject covariate gender, the individual achievement and the average class achievement have an influence on emotions experienced in mathematics. In this study we were interested if the gender and/or the students' *comparative ability* in mathematics (labelled `cab`), i.e. the self concept of ability compared to the perceived average of students' ability in mathematics, have an effect on learning related emotions in maths. The self concept of ability and the perceived students' ability were measured on a scale ranging from 0 to 100, where 0 indicated low and 100 high ability in mathematics. Comparing for each individual these two covariates we derived the subject covariate comparative ability (`cab`), where a negative sign indicates that student i has rated her/his ability in maths below the averaged perceived ability of other students (the social reference frame) and vice versa.

From this five learning related emotions we obtained ten emotion pairs. In each comparison students had to place a mark (with the cursor) on a bounded horizontal response scale (see Figure 1).
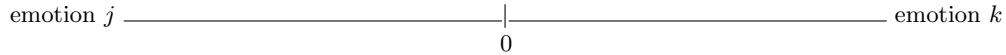
emotion j ——————————————|—————————————— emotion k

0

Figure 1: Assumed metric response scale with fixed bounds for the comparison (jk).

The given response corresponds to the degree of which an emotion is typically experienced more often than the other when learning maths. The response scale of each paired comparison had an arbitrary fixed length of 100 millimeters, i.e. 101 units including the response option in the middle of the scale (labelled with 0) which indicated a tie. The responses of the judges were scored by measuring the length or the units from the zero point (the middle of the scale) to the marked location. The random variables $Y_{jk,i}$ that are all associated with the first object in each of the $\binom{J}{2}$ comparisons can take on values in the interval $[-50, 50]$. An assigned value of $y_{jk,i} = -50$ indicates the most favourable response for the first emotion j in a given paired comparison (jk) and a value of $y_{jk,i} = 50$ the most favourable for the second emotion k. To avoid misunderstandings students were instructed how to use the response scale.

For the analysis we first squeezed the response variables $Y_{jk,i}$ with values in the interval $[-50, 50]$ into the interval $[0, 1]$ and then compressed the transformed variable $Y_{jk,i}^*$ so that the two times transformed variable $Y_{jk,i}^{**}$ assumes values in the standard interval $(0, 1)$ (see Section 2.1).

We started our model selection process by fitting a model with two main effects. One of them is the categorical subject covariate gender (`sex`) and the other the numerical subject covariate (`cab`). This main effects model gives a log likelihood of 141.4 (on 16 estimated parameters). For model comparison of nested BBTR models we used a likelihood ratio test. A reduction of the main effects model to either a model with the covariate gender (`sex`) or the covariate

comparative ability (cab) significantly worsened the model fit so that we preferred the main effects model (sex+cab). As models with constant precision parameter ϕ (as defined in Equation (5)) might be questionable, we also fitted a BBTR model that allows the precision parameter to vary, i.e. to depend on the subject covariates gender (sex) and comparative ability (cab). A general notation of the BBTR model with a variable precision parameter ϕ_i is given in Equation (2). We compared these two BBTR main effects models – one with a constant precision parameter (ϕ) for all observations i and the other one with sex and cab as additional regressors for the variable precision parameter (ϕ_i) in Equation (2) – by specifying for both models the log link function for the precision parameter. A likelihood ratio test gave evidence that the BBTR main effects model with a variable precision parameter significantly (p<0.001) improves the model fit and was therefore preferred for further analysis (see Table 2, main effects BBTR model).
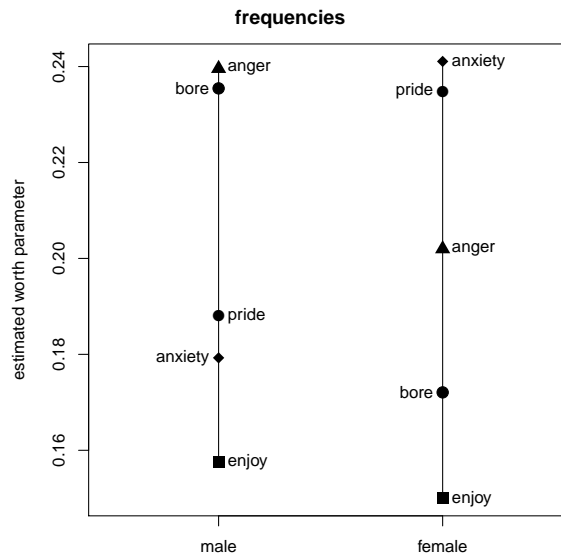

Figure 2: Plot of worth parameters for male and female students.

In Figure 2 the worth parameters of the BBTR model with the significant (p-value<0.001) covariate sex for both male and female students are shown. From this worth plot we can see that both groups differ in their ordering for the five emotions. Male students experience anger most often directly followed by boredom. For female students anxiety on the first place and pride on the second are the emotions most often experienced during learning mathematics. For both groups enjoyment is with distance the least experienced emotion. From the object parameter estimates (see Table 2, BBTR model covariate sex) of the BBTR model with covariate gender (sex) we noticed that female students experience the emotions pride and anxiety significantly (p-value=0.004 and p-value<0.001) more often than male students. From the fitted BBTR model with the significant (p-value<0.001) covariate comparative ability (cab) we can see in the worth plot (Figure 3) a tendency that the lower the comparative ability the more often anxiety is experienced. For students with relative low comparative ability (e.g. cab = −60) the distance between the first (anxiety) and second (anger) often experienced emotion is very large. This means that anxiety is with great distance the most often experienced emotion while learning maths. In Figure 3 we can see that there is a notable gap between the first three ranked emotions, i.e. anxiety, anger and pride. Whereas the worths of the emotions pride, boredom and enjoyment (the last three ranked emotions) are relative close together, which means that they are similar often experienced without big differences. We can see that the worth of all five emotions become more similar the closer the self concept of maths ability to the averaged perceived ability of other students. For students with high comparative ability enjoyment is ranked on the second place whereas for students who rate

their ability lower than those of their colleagues enjoyment is the least experienced emotion. We can further see in Figure 3 that for the group of students with relative high comparative ability (e.g. `cab` = 40) the considerable distance between the three top ranked emotions, noticed from students with relative low comparative ability, diminished. For students with relative high comparative ability there is not one predominant emotion (compared to students with low comparative ability) but instead three emotions (i.e. pride, enjoyment and boredom) compete on the top of the most often experienced emotions. However, the gaps between the worths of the three top ranked emotions and the fourth ranked emotion (anger) and between anger and the least often experienced emotion pride widen for the group of students with high comparative ability. Let us now have a more detailed look at the estimated object parameters for students with `cab` = −60. The distance between the estimated parameters of anxiety and anger i.e. $\Delta_{anxiety,anger} = |(\hat{\lambda}_{o4} + \hat{\beta}_{o4} * -60) - (\hat{\lambda}_{o3} + \hat{\beta}_{o3} * -60)| = 0.660$ (see Table 2, BBTR model covariate `cab`). The distance between anger and pride $\Delta_{anger,pride} = 0.628$. For students with `cab` = 40 the distances between the first three ranked emotions are each much smaller than for students with low comparative ability (`cab` = −60), i.e. $\Delta_{pride,enjoyment} = 0.166$ and $\Delta_{enjoyment,boredom} = 0.100$.
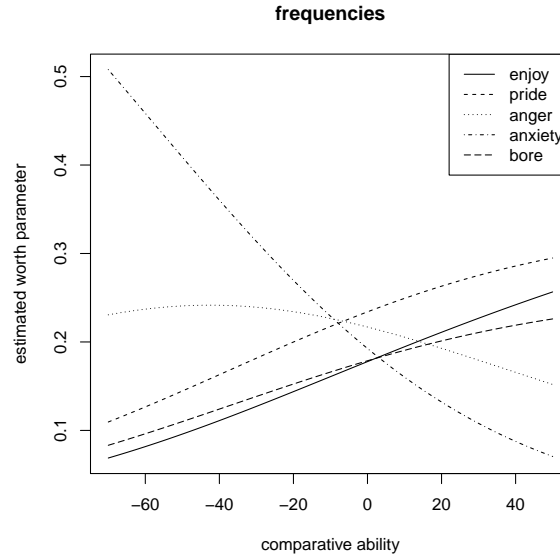

Figure 3: Plot of worth parameters for the subject covariate comparative ability in maths.

4. Discussion

In this article we suggested a simple beta Bradley-Terry regression (BBTR) model for modelling (assumed) metric paired comparison data. The structure of the BBTR model allows extensions for various variables, e.g. subject covariates or object-specific covariates. Model fitting can easily be done in a similar manner as for generalized linear models (GLMs) with the R-package `betareg` (Zeileis *et al.* 2013), provided that a corresponding design structure has been built up. Model selection of nested BBTR models can be done through likelihood ratio tests.

In our study we conducted a subjective self-rated survey which required both the ability and the willingness to deal with or to actually think about the experienced emotions and make them mentally present.

However, we selected and fitted a BBTR model with the two main effects gender (`sex`) and comparative ability (`cab`). A higher order model, i.e. an interaction model (`sex*cab`) would not significantly improve the model fit. In small samples the ML-parameter estimates may be biased, therefore we also fitted a BBTR model with bias correction (denoted by BC) and bias

Table 2: Estimates of nested BBTR models.

estimates	main effects BBTR model (s.e.)		BBTR model covariate sex (s.e.)		BBTR model covariate cab (s.e.)		BBTR 0-model (s.e.)	
o1	0.434	(0.176)	0.402	(0.179)	0.005	(0.060)	0.005	(0.062)
o2	0.241	(0.148)	0.225	(0.151)	−0.270	(0.060)	−0.268	(0.062)
o3	−0.027	(0.126)	−0.017	(0.128)	−0.193	(0.060)	−0.218	(0.062)
o4	0.268	(0.112)	0.273	(0.114)	−0.075	(0.062)	−0.144	(0.063)
o5	0	(NA)	0	(NA)	0	(NA)	0	(NA)
o1:sex2	−0.318	(0.218)	−0.264	(0.223)	-		-	
o2:sex2	−0.573 *	(0.183)	−0.536 *	(0.187)	-		-	
o3:sex2	−0.083	(0.154)	−0.143	(0.157)	-		-	
o4:sex2	−0.492 *	(0.136)	−0.610 *	(0.139)	-		-	
o5:sex2	0	(NA)	0	(NA)	-		-	
o1:cab	−0.005	(0.005)	-		−0.003	(0.005)	-	
o2:cab	−0.003	(0.004)	-		0.000	(0.005)	-	
o3:cab	0.010 *	(0.004)	-		0.012 *	(0.004)	-	
o4:cab	0.022 *	(0.003)	-		0.025 *	(0.003)	-	
o5:cab	0	(NA)	-		0	(NA)	-	
sex1 (male)	−0.096	(0.089)	−0.102	(0.091)	-		-	
sex2 (female)	−0.203	(0.065)	−0.188	(0.068)	-		-	
cab	−0.002	(0.003)	-		−0.001	(0.003)	-	
precision submodel:								
intercept	0.960	(0.059)	0.876	(0.058)	1.182	(0.038)	1.070	(0.037)
sex2	0.511	(0.078)	0.438	(0.076)	-		-	
cab	0.004	(0.002)	-		0.001	(0.002)	-	
log-likelihood	163.5		96		116.2		56.39	
number of estimated parameters	18		12		11		5	

The five emotions are shortnamed as follows: o1=enjoyment, o2=pride, o3=anger, o4=anxiety, o5=boredom and the numeric subject covariate comparative ability is labelled cab. Note that the worth parameters illustrated in Figure 2 and 3 were obtained from reversed object parameter estimates so that higher values indicate a more frequent experience.

reduction (denoted by BR) of the maximum likelihood (ML) parameter estimates using the betareg-package (see Grün, Kosmidis, and Zeileis 2012). We obtained very similar parameter estimates i.e. the differences in the parameter estimates were not worth mentioning.

The results of the BBTR model with the subject covariate gender (sex) indicate that male and female students differ in their ordering of learning related emotions. Male students typically experience most often anger while learning maths. Female students place anxiety on the first place of frequent experienced emotions. This outcome could be due to the fact that female students may fear (failure in) solving difficult maths tasks as they might have a tendency to underestimate their abilities in maths for various reasons (e.g. stereotype thinking).

One aim in the future is to define a latent class BBTR model for being able to model possible effects of latent (unobservable) subgroups.

Acknowledgement

References

Agresti A (1992). "Analysis of Ordinal Paired Comparison Data." *Applied Statistics*, **41**(2), 287–297.

Aitken R (1969). "Measurement of Feelings Using Visual Analogue Scales." *Proceedings of the Royal Society of Medicine*, **62**, 989–993.

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

Böckenholt U, Dillon W (1997). "Modeling Within-Subject Dependencies in Ordinal Paired Comparison Data." *Psychometrika*, **62**(3), 411–434.

Cribari-Neto F, Zeileis A (2010). "Beta Regression in R." *Journal of Statistical Software*, **34**(2), 1–24.

De Ruiz K (1990). *A Mathematical Model for a Paired Comparison Experiment on a Continuum of Response.* PhD dissertation, University of California-Riverside.

Dittrich R, Francis B, Hatzinger R, 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, Katzenbeisser W (1998). "Modelling the Effect of Subject-Specific Covariates in Paired Comparison Studies with an Application to University Rankings." *Applied Statistics*, **47**(4), 511–525.

Dittrich R, Hatzinger R, Katzenbeisser W (2004). "A Log-Linear Approach for Modelling Ordinal Paired Comparison Data on Motives to Start a PhD Programme." *Statistical Modelling*, **4**, 181–193.

Ferrando P (2002). "Theoretical and Empirical Comparisons between Two Models for Continuous Item Response." *Multivariate Behavioral Research*, **37**(4), 521–542.

Ferrari S, Cribari-Neto F (2004). "Beta Regression for Modelling Rates and Proportions." *Journal of Applied Statistics*, **31**(7), 799–815.

Francis B, Dittrich R, Hatzinger R, Penn R (2002). "Analysing Partial Ranks by Using Smoothed Paired Comparison Methods: An Investigation of Value Orientation in Europe." *Applied Statistics*, **51**(3), 319–336.

Frenzel A, Pekrun R, Goetz T (2007). "Girls and Mathematics – A "Hopeless" Issue? A Control-Value Approach to Gender Differences in Emotions Towards Mathematics." *European Journal of Psychology of Education*, **22**(4), 497–514.

Goetz T, Pekrun R, Zirngibl A, Jullien S, Kleine M, vomHofe R, Blum W (2004). "Leistung und Emotionales Erleben im Fach Mathematik." *Zeitschrift für Pädagogische Psychologie*, **18**(3/4), 201–212.

Grün B, Kosmidis I, Zeileis A (2012). "Extended Beta Regression in R: Shaken, Stirred, Mixed, and Partitioned." *Journal of Statistical Software*, **48**(11), 1–25.

Götz T, Keller M, Martiny S (2012). "Emotionales Erleben in den MINT-Fächern: Ursachen, Geschlechtsunterschiede und Interventionsmöglichkeiten." In H Stöger, A Ziegler, M Heilemann (eds.), *Mädchen und Frauen in MINT.*, pp. 135–161. Berlin: Lit Verlag.

Hatzinger R (2012). "Prefmod: Utilities to Fit Paired Comparison Models for Preferences." R package Version 0.8-31. URL http://cran.R-project.org/web/packages/prefmod/index.html.

Kieschnick R, McCullough B (2003). "Regression Analysis of Variates Observed on (0,1): Percentages, Proportions and Fractions." *Statistical Modelling*, **3**, 193–213.

Kousgaard N (1976). "Models for Paired Comparisons with Ties." *Scandinavian Journal of Statistics*, **3**(1), 1–14.

Müller H (1987). "A Rasch Model for Continuous Ratings." *Psychometrika*, **352**(2), 165–181.

Noel Y, Dauvier B (2007). "A Beta Item Response Model for Continuous Bounded Responses." *Applied Psychological Measurement*, **31**(1), 47–73.

Paolino P (2001). "Maximum Likelihood Estimation of Models with Beta-Distributed Dependent Variables." *Political Analysis*, **9**, 325–346.

Pekrun R, Goetz T, Frenzel A, Barchfeld P, Perry R (2011). "Measuring Emotions in Students' Learning and Performance: The Achievement Emotions Questionnaire (AEQ)." *Contemporary Educational Psychology*, **36**, 36–48.

R Development Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL http://www.R-project.org.

Rao P, Kupper L (1967). "Ties in Paired-Comparison Experiments: A Generalization of the Bradley-Terry Model." *Journal of the American Statistical Association*, **62**(317), 194–204.

Simas A, Barreto-Souza W, Rocha A (2010). "Improved Estimators for a General Class of Beta Regression Models." *Computational Statistics and Data Analysis*, **54**, 348–366.

Smithson M, Verkuilen J (2006). "A Better Lemon Squeezer? Maximum-Likelihood Regression with Beta-Distributed Dependent Variables." *Psychological Methods*, **11**(1), 54–71.

Stern S (2011). "Moderated Paired Comparisons: A Generalized Bradley-Terry Model for Continuous Data Using a Discontinuous Penalized Likelihood Function." *Applied Statistics*, **60**(3), 397–415.

Zeileis A, Cribari-Neto F, Grün B, Kosmidis I, Simas A, Rocha A (2013). "Betareg." R package Version 3.0-2. URL http://CRAN.R-project.org/package=betareg.

Affiliation:

Alexandra Grand and Regina Dittrich
Institute for Statistics and Mathematics
WU - Vienna University of Economics and Business
Welthandelsplatz 1
A-1020 Vienna, Austria
E-mail: alexandra.grand@gmx.at, Regina.Dittrich@wu.ac.at