



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUT FÜR STATISTIK



Gerhard Tutz & Gunther Schauberger

Visualization of Categorical Response Models – from Data Glyphs to Parameter Glyphs

Technical Report Number 117, 2012
Department of Statistics
University of Munich

<http://www.stat.uni-muenchen.de>



Visualization of Categorical Response Models - from Data Glyphs to Parameter Glyphs

Gerhard Tutz & Gunther Schauberger

Ludwig-Maximilians-Universität München

Akademiestraße 1, 80799 München

March 5, 2012

Abstract

The multinomial logit model is the most widely used model for nominal multi-category responses. One problem with the model is that many parameters are involved, another that interpretation of parameters is much harder than for linear models because the model is non-linear. Both problems can profit from graphical representations. We propose to visualize the effect strengths by star plots, where one star collects all the parameters connected to one term in the linear predictor. In simple models one star refers to one explanatory variable. In contrast to conventional star plots, which are used to represent data, the plots represent parameters and are considered as parameter glyphs. The set of stars for a fitted model makes the main features of the effects of explanatory variables on the response variable easily accessible. The method is extended to ordinal models and illustrated by several data sets.

Keywords: Glyphs, star plots, visualization, multinomial logit model, ordered response models.

1 Introduction

Multinomial response models are a common tool in categorical data analysis with well-established theory. But in applications, in particular in the case of many response categories, it is often tedious to keep track and interpret all of the parameters. Therefore tools for visualization of the effects of explanatory variables will be helpful for practitioners.

In multivariate data analysis visualization techniques have a long tradition. Skillfully devised graphical methods allow to look into data and uncover features of the underlying data generating process. They are used to explore data and also

to present results. Various books and articles are devoted to graphical representations of data, see, in particular, Cleveland (1985), Kastelec and Leoni (2007), and the Handbook of Data Visualization (Chen et al., 2008).

In the following the focus is not on visualization of data but on the visualization of fitted models to help in the interpretation of parameters. The aim of closer linkage of statistical modelling with graphics is investigated in the case of categorical response models. Categorical response models like the multinomial logit models represent a challenge if the number of response categories and/or the number of explanatory variables is large. Even for moderate numbers of explanatory variables one obtains a large number of parameters and the impact of the predictors on the response variable is hard to investigate because of the transformation to logits. While the increase or decrease of the mean response is easily seen in linear models, the effect on logits is much harder to explain to practitioners.

There has been some work in the visualization of categorical data. In particular graphical methods for the analysis of multiway contingency tables in the form of mosaic plots (Friendly, 1994; Theus and Lauer, 1999; Hofmann, 2001; Meyer et al., 2008) are widely used. But categorical response models that also contain continuous predictors cannot be reduced to contingency tables without loss of information. Therefore, for the general case of categorical responses mosaic plots are not very helpful. More recently, in Fox and Andersen (2006) and Fox and Hong (2009) the work on effect displays for generalized linear models (Fox, 2003) was extended to multinomial and proportional-odds logit models, available in the effects package (Fox and Andersen, 2006; Fox et al., 2011). The proposed effect displays depict fitted category probabilities including pointwise confidence envelopes and are typically used for visualization of high-order terms. The package provides several kinds of displays for polytomous logit models.

The objective of the present paper is to develop alternative graphical methods for the general case of categorical response models with all types of regressors. In Section 2 we briefly sketch the multinomial logit model and the interpretation of parameters. In Section 3 more traditional tools for the graphical representation of the effect of explanatory variables in the form of probability plots are considered. The main tool, graphical tools for the visualization of parameters, is given in Section 4. We conclude with an extension to ordinal response models.

2 The Multinomial Logit Model

In the following we shortly summarize the essential properties of the multinomial logit model, which is the most frequently used model in regression analysis for unordered categorical responses and is extensively treated, for example, in Agresti (2002). For response $Y \in \{1, \dots, k\}$ and the vector of explanatory variables \mathbf{x} it has the form

$$P(Y = r|\mathbf{x}) = \frac{\exp(\beta_{r0} + \mathbf{x}^T \boldsymbol{\beta}_r)}{\sum_{s=1}^k \exp(\beta_{s0} + \mathbf{x}^T \boldsymbol{\beta}_s)}, \quad (1)$$

where $\boldsymbol{\beta}_r^T = (\beta_{r1}, \dots, \beta_{rp})$. Since parameters $\beta_{10}, \dots, \beta_{k0}$, $\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_k^T$ are not identifiable additional constraints are needed. One option is to chose one of the response categories as reference category. For example, by setting $\beta_{k0} = 0$, $\boldsymbol{\beta}_k = \mathbf{0}$, category k is chosen as the reference category and interpretation of all parameters refers to this category. Alternatively one can use the symmetric side constraints $\sum_{s=1}^k \beta_{s0} = 0$, $\sum_{s=1}^k \boldsymbol{\beta}_s^T = (0, \dots, 0)$. In both cases one has $k - 1$ intercepts and $p(k - 1)$ effects of predictors, where p denotes the length of \mathbf{x} . Even for moderate number of predictors, say 10, and 5 response categories, one obtains 40 parameters that represent effects of predictors. The result is a lengthy list of parameter estimates that contains the relevant information but it takes some skill and time to evaluate the effects.

The large number of parameters is due to the multi-dimensionality of the model. The response variable $Y \in \{1, \dots, k\}$ hides that the response is actually multivariate. This becomes obvious by considering the distribution of the response. By defining dummy variables y_1, \dots, y_{k-1} with $Y = r \Leftrightarrow y_r = 1$ the possible outcome vectors of length $k - 1$ are given by $(1, 0, \dots), (0, 1, 0, \dots) \dots (0, 0, \dots, 0)$. With probabilities given by $\pi_r(\mathbf{x}) = P(Y = r|\mathbf{x}) = P(y_r = 1|\mathbf{x})$ the vector $\mathbf{y}^T = (y_1, \dots, y_{k-1})$ follows a multinomial distribution $\mathbf{y} \sim M(1, \boldsymbol{\pi}(\mathbf{x}))$, where $\boldsymbol{\pi}^T(\mathbf{x}) = (\pi_1(\mathbf{x}), \dots, \pi_{k-1}(\mathbf{x}))$ represents the vector of response probabilities. A closed representation of the $(k - 1)$ -dimensional model as a multivariate generalized linear model (GLM) uses the form $g(\boldsymbol{\pi}(\mathbf{x})) = \mathbf{X}\boldsymbol{\beta}$ with $(k - 1)$ -dimensional link function g , design matrix \mathbf{X} and all the parameters collected in $\boldsymbol{\beta}^T = (\beta_{10}, \dots, \beta_{k-1,0}, \boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_{k-1}^T)$. Maximum likelihood estimation and parameter tests can be derived within the framework of multivariate GLMs (see, for example, Tutz, 2012).

For the interpretation of the parameters it is essential to specify the identifiability constraint that is used. If k is chosen as the reference category one obtains

$$\log \left(\frac{P(Y = r|\mathbf{x})}{P(Y = k|\mathbf{x})} \right) = \beta_{r0} + \mathbf{x}^T \boldsymbol{\beta}_r, \quad r = 1, \dots, k - 1, \quad (2)$$

where the log-odds compare $P(Y = r|\mathbf{x})$ to the probability $P(Y = k|\mathbf{x})$. Then the parameters reflect the effect of predictors on the relation between category r and the reference category k . Symmetric side constraints are less often used although there is a nice interpretation of parameters. For symmetric side constraints the interpretation refers to the "mean" response defined by the geometric mean

$$GM(\mathbf{x}) = \sqrt[k]{\prod_{s=1}^k P(Y = s|\mathbf{x})} = \left(\prod_{s=1}^k P(Y = s|\mathbf{x}) \right)^{1/k}.$$

It is easily derived that

$$\log \left(\frac{P(Y = r|\mathbf{x})}{GM(\mathbf{x})} \right) = \beta_{r0} + \mathbf{x}^T \boldsymbol{\beta}_r, \quad r = 1, \dots, k,$$

holds. Therefore, $\boldsymbol{\beta}_r$ reflects the effects of \mathbf{x} on the logits when $P(Y = r|\mathbf{x})$ is compared to the geometric mean response $GM(\mathbf{x})$.

When visualizing effects we will focus on symmetric side constraints because effects do not refer to the assigned reference category but to all of the categories. Also results of testing of hypothesis and corresponding p -values are easier to interpret. If $H_0 : \beta_{rj} = 0$ is rejected for the model with reference category k the j th variable distinguishes significantly between response $Y = r$ and $Y = k$. If $H_0 : \beta_{rj} = 0$ is rejected for the model with symmetric side constraint the j th variable distinguishes between response $Y = r$ and $Y \neq r$.

3 Traditional Methods of Visualization: Probability Plots

When visualizing the effects of predictors the main problem with the multinomial logit model is that the link function is not linear. Although odds are an intuitive concept the log-odds in equ. (2) are not appropriate to obtain some feeling for the impact of predictors. Therefore, a traditional way to visualize the effect of explanatory variables is the plotting of response probabilities against the values of specific covariates, see, for example, Agresti (2002).

For illustration we will consider the modelling of party choice with data from the German Longitudinal Election Study. The response categories refer to the dominant parties in Germany, in particular, the Christian Democratic Union (CDU: 1), the Social Democratic Party (SPD: 2), the Liberal Party (FDP: 3), the Green Party (4) and the Left Party (Die Linke: 5). With the five response categories nine predictors were collected, age, political interest (1: less interested 0: very interested), religion (1: evangelical, 2: catholic, 3: otherwise), regional provenance (west; 1: former West Germany, 0: otherwise), gender (1: male, 0: female), union (1: member of a union 0: otherwise), satisfaction with the functioning of democracy (democracy; 1: not satisfied 0: satisfied), unemployment (1: currently unemployed, 0: otherwise), and high school degree (1: yes, 0: no).

Table 1 shows the estimated parameters together with standard errors. It is seen that even in this simple example with moderate number of predictors and response categories many parameters have to be investigated. A simple way to illustrate the effect of a metric covariate like age is to plot the response probabilities against age. But, of course, in a non-linear model as the logit model, the form of the function strongly depends on the values of the other parameters. In Figure 1 the probabilities are given for two sets of values, one where all other predictors have value 0, one where all other predictors have value 1. It is seen that not only the level but also the slope of the curves can vary with the chosen

TABLE 1: *Estimates of multinomial logit model for party preference data, symmetric side constraints.*

	Intercept	Age	Religion (2)	Religion (3)	Democracy (2)	Pol.Interest (1)
CDU	1.397	0.308	0.404	-0.358	-0.766	0.202
SPD	0.469	0.148	-0.196	-0.428	-0.360	0.337
FDP	-0.345	-0.111	0.090	0.326	0.002	-0.264
Greens	-1.096	-0.398	-0.127	0.286	0.008	0.214
Left Party	-0.425	0.053	-0.171	0.174	1.116	-0.488

	Unemployment (2)	Highschool (1)	Union (2)	West (1)	Gender (1)
CDU	-0.514	0.156	-0.408	-0.330	-0.262
SPD	0.127	-0.221	0.400	0.389	-0.191
FDP	-0.560	0.051	-0.509	0.025	0.254
Greens	-0.071	0.563	-0.391	0.639	-0.019
Left Party	1.018	-0.549	0.907	-0.723	0.218

Standard Errors

	Intercept	Age	Religion (2)	Religion (3)	Democracy (2)	Pol.Interest (1)
CDU	0.224	0.069	0.163	0.168	0.139	0.147
SPD	0.245	0.072	0.166	0.172	0.148	0.160
FDP	0.312	0.094	0.239	0.218	0.194	0.191
Greens	0.327	0.097	0.234	0.213	0.193	0.203
Left Party	0.313	0.094	0.233	0.205	0.229	0.185

	Unemployment (2)	Highschool (1)	Union (2)	West (1)	Gender (1)
CDU	0.366	0.169	0.212	0.156	0.135
SPD	0.314	0.189	0.184	0.172	0.142
FDP	0.498	0.218	0.289	0.207	0.187
Greens	0.421	0.208	0.273	0.222	0.184
Left Party	0.301	0.243	0.216	0.194	0.181

value for the other variables. For example, the curve for the Social Democratic Party (SPD) is rather flat in the upper panel, but increasing in the lower panel.

When explanatory variables are categorical bar plots with the probabilities corresponding to the height of the bars can be used. Figure 2 shows the effect of unemployment on the choice probabilities. It shows, for example, that unemployed persons have a stronger preference for the left party, preference for CDU decreases. The tendency is the same if different values are chosen for the other variables (Figure 3), but effect strength is quite different. If the other variables have value 1, the probability for CDU is among the lowest if voters are unemployed.

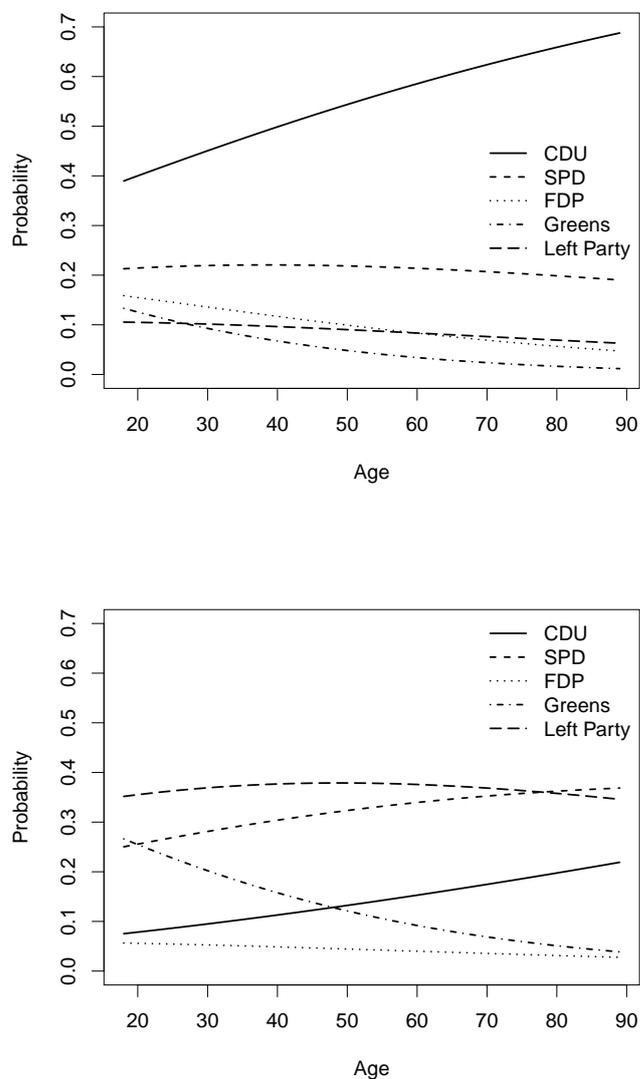


FIGURE 1: *Estimated probabilities for party preference against age, all other variables fixed at value 0 (upper panel), all other variables fixed at value 1 (lower panel).*

4 Glyphs for the Visualization of Parameters

The disadvantage of bar plots as well as curves is that they show effects under the constraint that the other predictors have fixed chosen values. The plots vary with the chosen values. An alternative approach that is propagated here is to visualize the effect strength that is contained in the parameters rather than the probabilities.

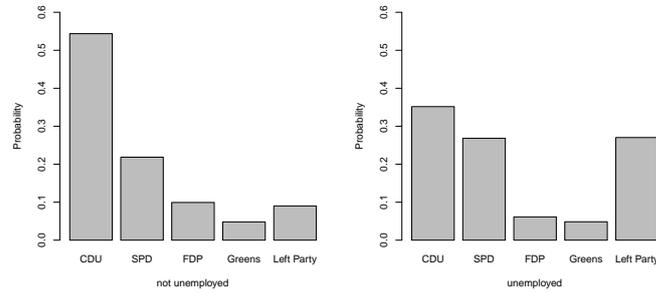


FIGURE 2: Bar plot of estimated probabilities for party preference for unemployment=0 (left) and unemployment=1 (right), all other categorical variables fixed at value 0, age at 50.

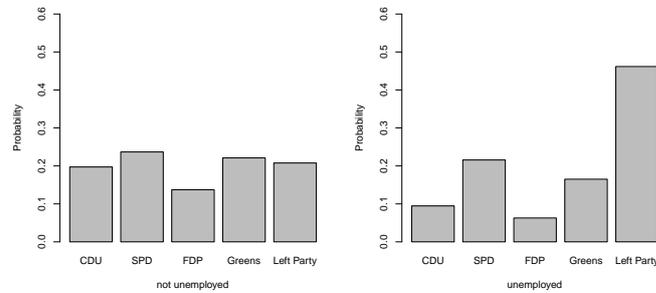


FIGURE 3: Bar plot of estimated probabilities for party preference for unemployment=0 (left) and unemployment=1 (right), all other categorical variables fixed at value 1, age at 50.

We will use glyphs that have traditionally been used to visualize data. Various glyphs have been proposed in the literature, among them profile glyphs (Du Toit et al., 1986), Chernoff faces (Chernoff, 1973) and stars (Anderson, 1957, Siegel et al., 1972, Gnanadesikan, 1977). We will focus on star plots, but instead of using them to visualize data, they are used to visualize parameters. The parameters of the logit model themselves are less appropriate since they contain the effect on logits, which do not carry much intuition. A much better way is to focus on the odds that stand behind the log-odds (or logits).

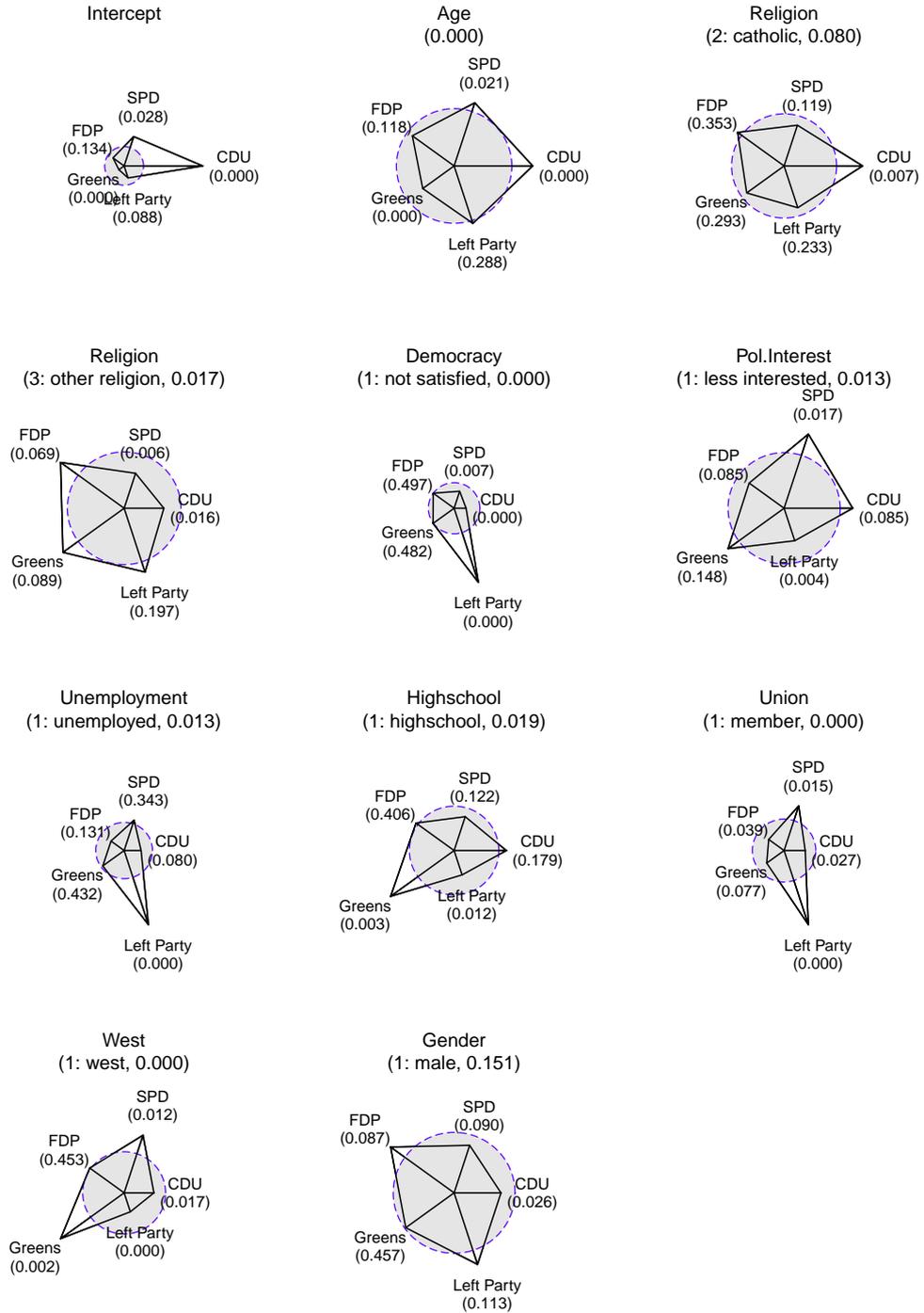


FIGURE 4: *Effect stars showing the exponentials of parameters, p-values at the rays refer to hypothesis $H_0 : \beta_{r,j} = 0$, p-values given with the variable description refer to hypothesis $H_0 : \beta_{1,j} = \dots = \beta_{k,j} = 0$.*

4.1 Star Plots for Parameters

The main tool is the representation of the odds of a model with symmetric side constraints as

$$\frac{P(Y = r|\mathbf{x})}{GM(\mathbf{x})} = \exp(\beta_{r0} + \mathbf{x}^T \boldsymbol{\beta}_r) = e^{\beta_{r0}} e^{x_1 \beta_{r1}} \dots e^{x_p \beta_{rp}} = e^{\beta_{r0}} (e^{\beta_{r1}})^{x_1} \dots (e^{\beta_{rp}})^{x_p}.$$

From

$$\frac{P(Y = r|x_1, \dots, x_j + 1, \dots, x_p)/GM(x_1, \dots, x_j + 1, \dots, x_p)}{P(Y = r|x_1, \dots, x_j, \dots, x_p)/GM(x_1, \dots, x_j, \dots, x_p)} = e^{\beta_{rp}}$$

it is seen that $e^{\beta_{rj}}$ represents the multiplicative effect of variable j on the odds $P(Y = r|\mathbf{x})/GM(\mathbf{x})$ if x_j increases by one unit.

In "effect stars", which are proposed here, the lengths of the rays emanating from the center of the plot represent the exponentials of the parameters. Thus one obtains a star plot for each variable that shows how strong the impact of the predictor on the response is and what form it takes. In addition, we include a (shaded) unit circle around the center that corresponds to the no-effects case, where $\beta_{1j} = \dots = \beta_{kj} = 0$ or, equivalently, $e^{\beta_{1j}} = \dots = e^{\beta_{kj}} = 1$ holds. Therefore, the deviation from the circle shows the strength of the preference for one category as the deviation from the circle. If the ray is outside the circle the increase in the predictor increases the probability of the corresponding category, if it is inside the circle the increase in the predictor decreases the response probability. By default stars are standardized such that the maximal ray lengths of the stars have the same value. This value also scales the radius of the unit circle.

Figure 4 shows the effect stars for the main effect model fitted to the party choice data, where the quantitative variable age has been standardized. Let us consider the effect of age. It is immediately seen that with increasing age the Christian-democratic party (CDU) is more strongly favored while, in particular, the response probability for the Greens decreases. An additional feature that is included is the significance of the deviation. The value in brackets given at each ray is the p -value of the hypothesis $H_0 : \beta_{rj} = 0$ for the model with symmetric side constraint. The effects of age on responses CDU, SPD and Greens turned out to be significant at the level 0.05, the former two with positive (outside the circle), the latter with negative effect (within the circle). In addition, the overall p -value for the hypothesis that one variable can be neglected, that is, $H_0 : \beta_{1j} = \dots = \beta_{kj} = 0$, is given with the description of the variable. For example, age turned out to be highly significant (0.000), whereas the effect of gender was weak (p -value of 0.151).

The advantage of the effect star plots is that all the effects of the variables are shown simultaneously. Metric as well as continuous variables are given in the same representation. In addition to the direction of the effect seen from the shape of the star, information on the significance of specific effects is included, as well as information about the whole variable.

Relevant features are easily seen from the form of the stars. For example, very strong deviations from the circle are found for the variables democracy, unemployment and union. All these variables have a strong effect in favor of the left party. Deviations from the star in favour of the Greens are seen for the variables high school and west. Supporters of the Greens are found among more educated persons from the former west.

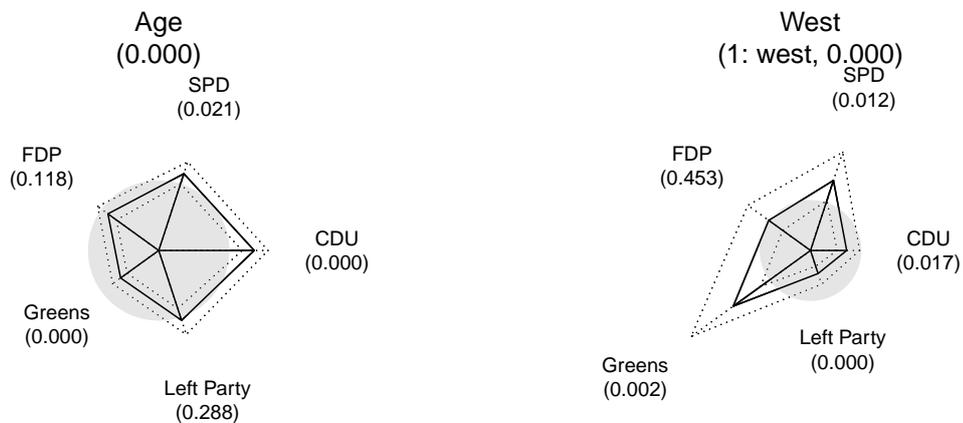


FIGURE 5: *Effect stars with reliability intervals for two variables (party preference data)*

4.2 Extensions and Alternatives

The presentation can be extended to include standard errors. Let σ_{rj} denote the standard error for estimation of β_{rj} . Then, a reliability interval for the exponential is given by $[\exp(\hat{\beta}_{rj} - 1.96\sigma_{rj}), \exp(\hat{\beta}_{rj} + 1.96\sigma_{rj})]$. By plotting the lower and the upper limit one obtains an inner and an outer star.

Figure 5 shows the plots of two predictors for the party preference data. If p -values are large, for example, for FDP and the left party for variable age, and FDP for variable west, the circle is covered by the corresponding intervals whereas for highly significant predictors, for example, CDU for variable age, the corresponding intervals are outside or within the circle. Inclusion of standard errors is certainly helpful but with many stars information content can be high. One strategy is to look first at all the stars without reliability intervals and then pick out the interesting ones and look at them more closely.

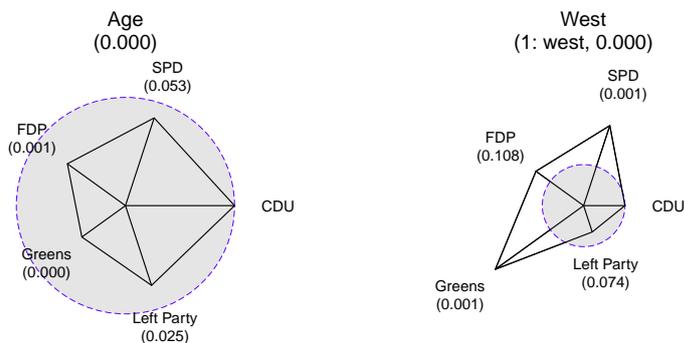


FIGURE 6: *Effect stars with reference category (party preference data)*

It should be noted that star plots for the exponentials of the parameters have the same form if a reference category is chosen. But then a more appropriate circle is the circle with radius defined by the reference category. The radius is fixed by the length of the ray for the reference category. Figure 6 shows effects of two variables with reference category CDU. Now rays inside the circle show that the predictor decreases the preference for the corresponding category when compared to the reference category. Rays outside the circle represent the opposite effect. But in both cases interpretation is in relation to the specified reference category (CDU). Consequently the p -values given now refer to the null hypothesis $H_0 : \beta_{rj} = 0$ for parameters constrained by fixing the reference category.

In Figure 4 the main effect model was represented by star plots. But, of course, also interaction terms can be represented as stars. For simplicity we consider one interaction term that turned out to be significant, namely the interaction between age and democracy. Figure 7 shows the stars for the marginal terms and the interaction. The stars for the other variables do hardly change when the interaction is included and therefore are not shown. When comparing to Figure 4 one sees that the main effect of democracy hardly changes while the main effect of age is quite different. The interaction effect shows that the effect of age is modified by the binary variable democracy (1: not satisfied, 0: satisfied). In particular, the preference for the big parties, SPD and CDU, increases stronger with age than is contained in the marginal effect of age if voters are not satisfied with democracy. For the green and the left party the slope of age is smaller when compared to the marginal effect if voters are not satisfied with democracy.

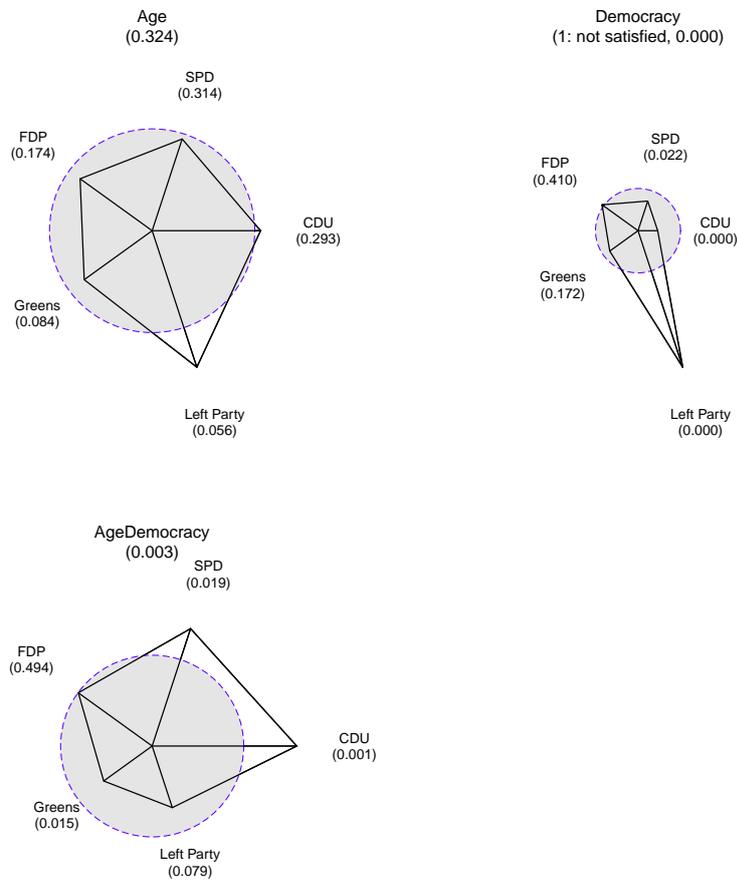


FIGURE 7: Marginal and interaction term stars for age and democracy (party preference data).

4.3 Alternative Displays

Star plots visualize parameters of fitted models. The plots are especially simple for main effect models when predictors are binary or are measured on a metric scale level. Then one star collects all the parameters connected to one explanatory variable. For categorical predictors with more than two categories several stars are linked to one predictor. The same holds when interactions are included. Then one has at least three stars that are linked to two variables. Although the interaction star as a visualization of the underlying effects is interpretable, the effect of a variable is not easily seen since it has to be seen in combination with the variable with which it interacts. The effect displays proposed by Fox and Andersen (2006) are able to visualize the effects of interaction terms quite nicely by allowing other predictors marginal to a given term to be set at average or

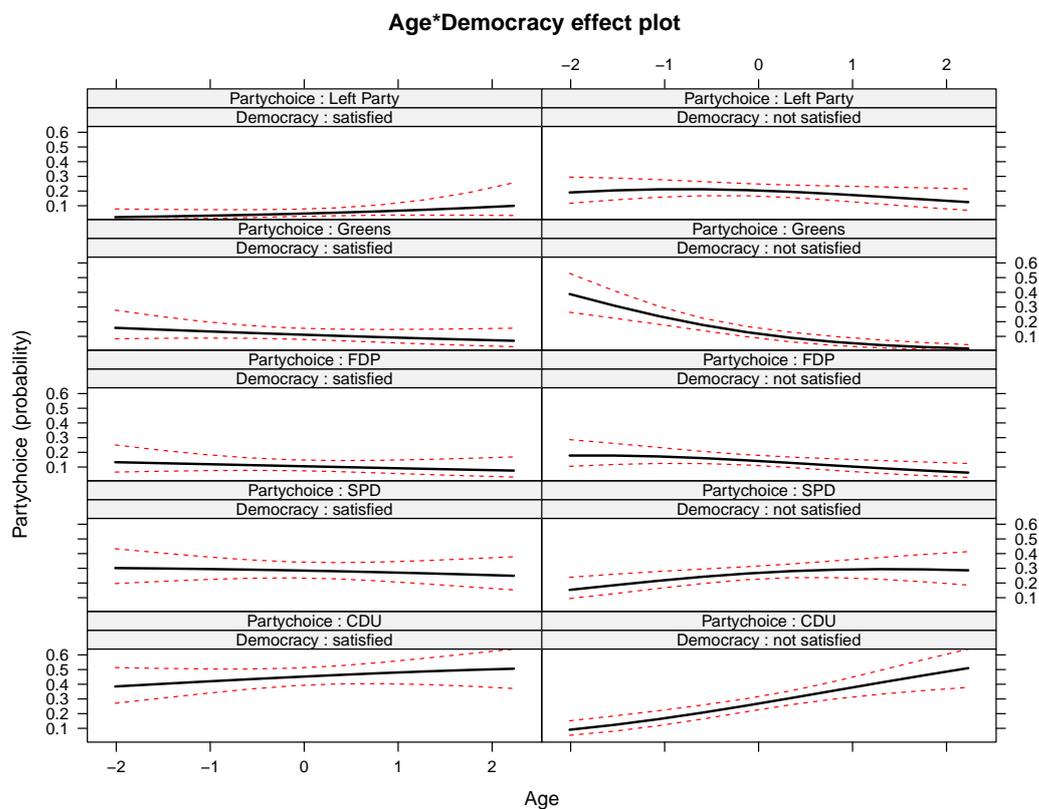


FIGURE 8: *Effect plot for the interaction of age and democracy*

other values.

For illustration we consider the interaction effect between the variables age and democracy, for which stars are given in Figures 8 and 9. Figure 8 shows the typical effect plots as curves and Figure 9 shows the "stacked area" displays also offered by the effects package (Fox et al., 2011). They visualize what the interaction star in Figure 7 shows only qualitatively, that the preference for the big parties, SPD and CDU, increases stronger with age if voters are not satisfied, for the green and the left party the effect slope decreases if voters are not satisfied. In particular the stacked area display visualizes nicely the effect of age and democracy on the response. Nevertheless, it should be noted that the effects on the probabilities are shown for fixed values of the other variables, in our case they have been chosen by mean values. If other values are chosen the effects on probabilities might change.

One can also plot the linear predictor itself, which means the effect on the logits. This plots would essentially show the same form of effects (but shifted) for other values of the rest of the variables, but it has the disadvantage that it is much harder to think in logits than in probabilities. For binary responses the

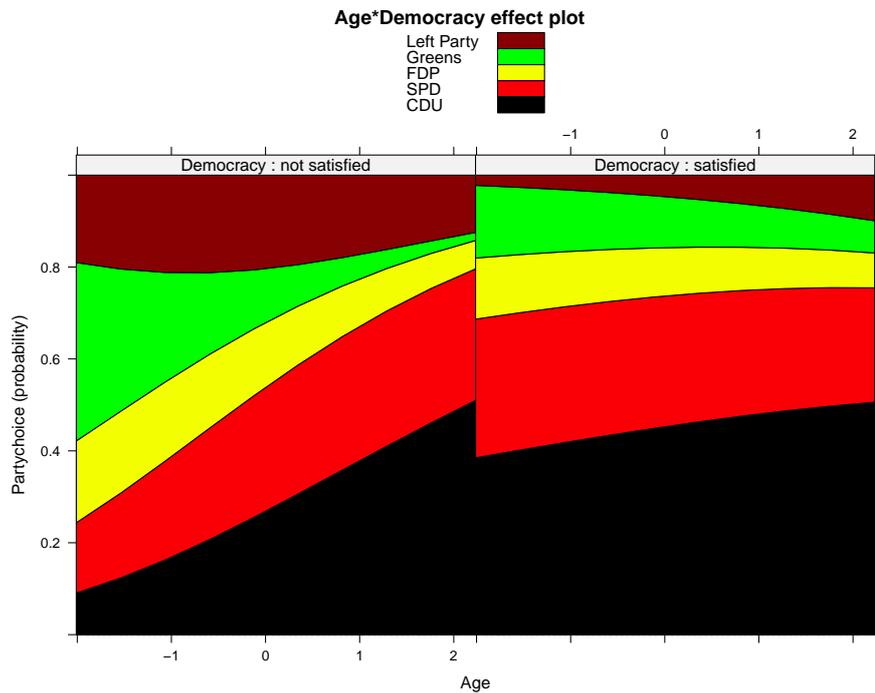


FIGURE 9: *Stacked area display for the interaction of age and democracy.*

effects package offers the option to label the response axis nonlinearly on the probability scale. Then one can see the effect on probabilities from the scaling. For multinomial the scaling is not so straightforward because it depends on the logits that were chosen, that is, the reference category that has been fixed. In their application Fox and Hong (2009) also rely on probability plots to visualize the effects in multinomial models. Star plots avoid the dependence on the reference category by using symmetric side constraints. By using odds rather than logits the effect strength is somewhat more intuitive.

The essential difference between stars and effect displays provided by the effects package is that stars visualize parameters with effect strength referring to specific odds and effect displays visualize the effects on probabilities or logits as curves. Effects displays are strong tools especially for interaction effects because they include the marginal effects. After screening the effects by star plots it is certainly a good idea to look at the plots provided the effects package, which, in particular for metric predictors, show the continuous dependence on the predictor. One other advantage of the effects package is that smooth effects of continuous predictors can be included. Although one might construct stars that visualize smooth effects it would destroy the simplicity of the visualization by stars (see concluding remarks).

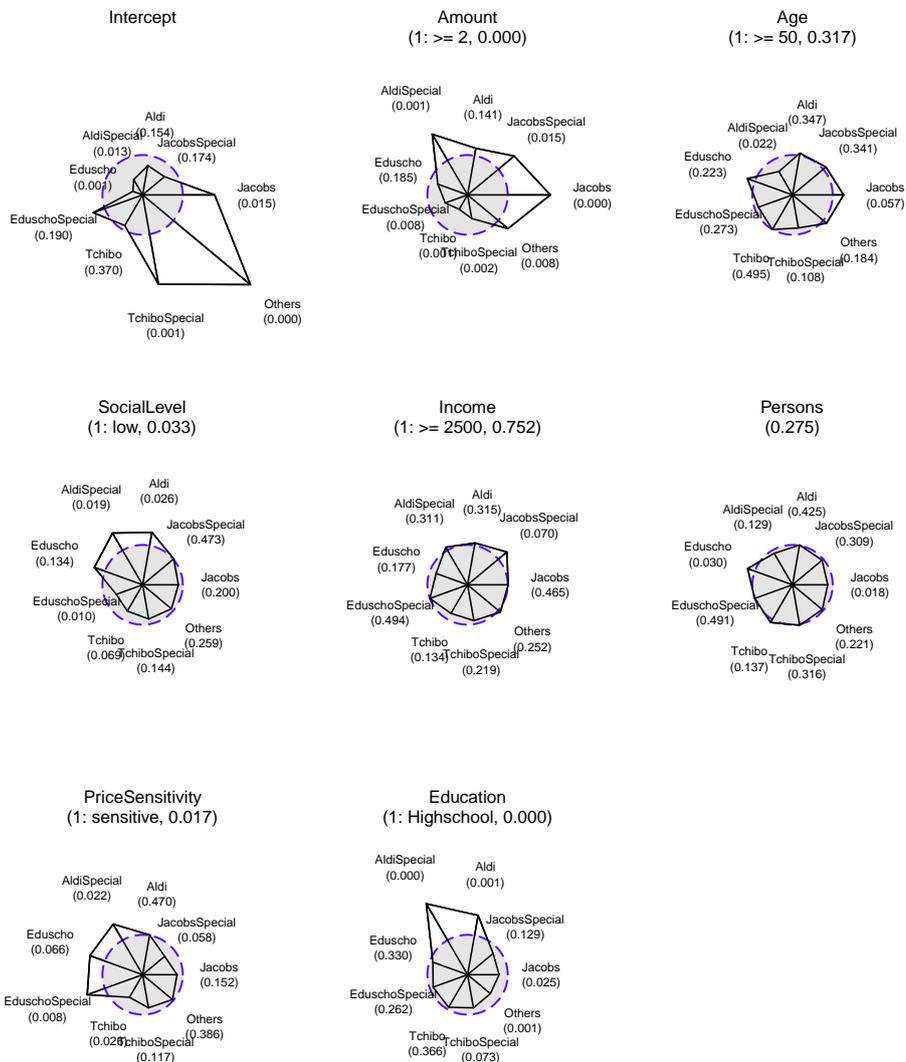


FIGURE 10: Graphs for brand choice data with seven predictors and fixed radius. It is seen that stars for the significant predictors, mount, social level, price sensitivity and education, deviate strongly from the circle.

4.4 Further Examples

For further illustration we consider brand choice data. The data refer to different brands of coffee. The purchases of coffees of 2111 households were collected by the Gesellschaft für Konsumforschung (Society for Consumer Research) and are available at <http://www.stat.uni-muenchen.de/service/datenarchiv/kaffee/>. The brands were named after the shops, which offer a regular brand and a special brand, Aldi, AldiSpecial, Jacobs, JacobsSpecial, Eduscho, EduschoSpecial, Tchibo, TchiboSpecial. The binary covariates were the number of packages

bought (amount; 1: ≥ 2), age (1: ≥ 50), social level (1: low), monthly income (1: ≥ 2500), persons in household, price sensitivity (1: sensitive), education (1: high school). Figure 10 shows the corresponding glyphs. Three of the predictors are not significant, namely age, income, and persons in household. It is seen that the corresponding stars are very close to the circle. For the significant predictors the stars deviate strongly from the no-effects circle. Naturally, the interpretation of the single effects refers to the brands considered. One sees, for example, that the brands, offered by the cheap discounter Aldi, are preferred if the social level is low. Following the suggestion of an Associate Editor the stars are scaled in a different way, namely by fixing the radius of the unit circle. What works well in this example can be less advantageous for other data (see next example).

An often used example with a categorical predictor is the alligator food choice considered in Agresti (2002). In the study by the Florida Game and Fresh Water Commission the response is the primary food type in categories fish, invertebrate, reptile, bird, and other. The explanatory variables are size, dichotomized into $\leq 2.3, > 2.3$, gender (1: male, 0: female), and the lake where the reptiles lived (four categories, 1: George, 2: Hancock, 3: Oklawaha, 4: Trafford), see Agresti (2002). Figure 11 shows the resulting glyphs where lake George is used as the reference category of the predictor lake. Therefore, the intercepts represent the food preference in lake George for small female alligators. It is seen that size of the alligator changes the food preference; larger alligators have a stronger preference of bird and reptiles. Also the lake makes a difference. The effects have to be interpreted with respect to the reference lake George. It is seen that in the other lakes in particular reptiles seem to be strongly preferred. In this example the fixed radius yields stars of strongly varying size. Thus, one preferable scaling of stars depends on the data - therefore, the option to fix the radius is included as an option in the EffectStars package.

5 Ordinal Response Models

The graphical tool of parameter glyphs can also be used to uncover structures in ordinal response models as the cumulative type models or the sequential type models (for example, Agresti, 2009). For simplicity, we restrict consideration to logit models. Let the response Y take values from ordered categories $\{1, \dots, k\}$. The cumulative logit model has the general form

$$\log \left(\frac{P(Y \leq r | \mathbf{x})}{P(Y > r | \mathbf{x})} \right) = \gamma_{0r} + \mathbf{x}^T \boldsymbol{\gamma}_r, \quad r = 1, \dots, k-1,$$

or

$$P(Y \leq r | \mathbf{x}) = \frac{\exp(\gamma_{r0} + \mathbf{x}^T \boldsymbol{\gamma}_r)}{1 + \exp(\gamma_{0r} + \mathbf{x}^T \boldsymbol{\gamma}_r)}, \quad r = 1, \dots, k-1,$$

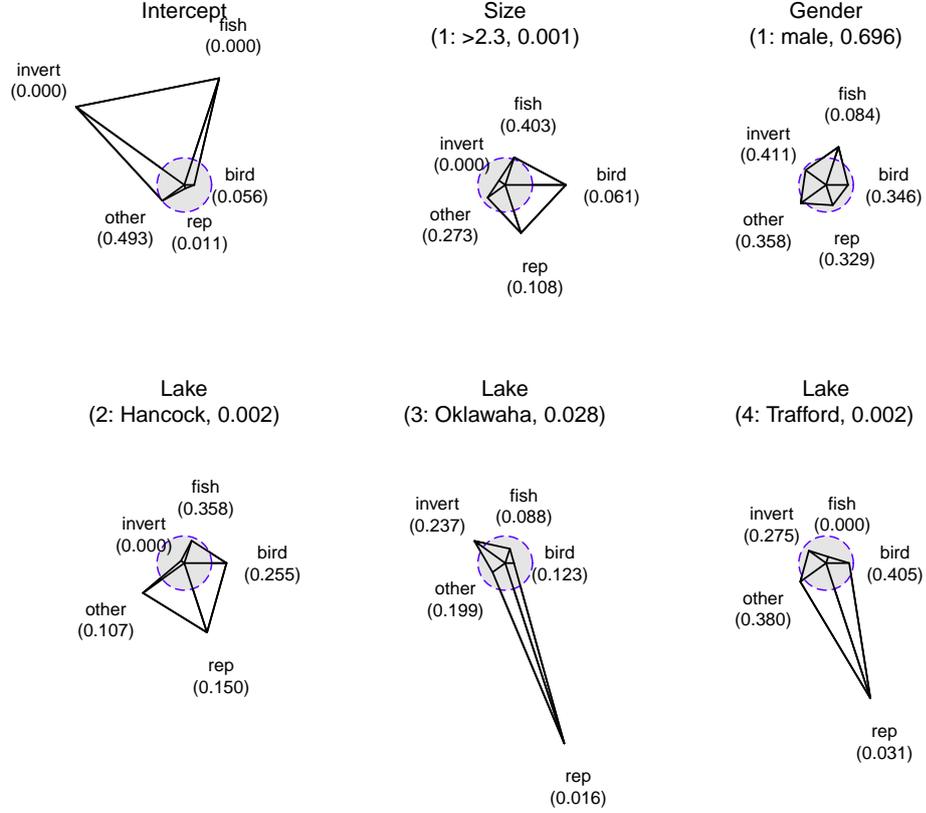


FIGURE 11: Food choice for alligator data depending on size, gender, and lake with fixed radius. Food preference strongly depends on the lake. Fixed radius produces stars of strongly varying size.

The sequential logit model (also called continuation ratio model) has the form

$$\log \left(\frac{P(Y = r | \mathbf{x})}{P(Y > r | \mathbf{x})} \right) = \gamma_{0r} + \mathbf{x}^T \boldsymbol{\gamma}_r, \quad r = 1, \dots, k-1,$$

or

$$P(Y = r | Y \geq r, \mathbf{x}) = \frac{\exp(\gamma_{r0} + \mathbf{x}^T \boldsymbol{\gamma}_r)}{1 + \exp(\gamma_{0r} + \mathbf{x}^T \boldsymbol{\gamma}_r)}, \quad r = 1, \dots, k-1.$$

The model is strongly related to discrete hazard models if the response refers to categorical survival. Then the probability $P(Y = r | Y \geq r, \mathbf{x})$ represents the probability of failure in (time) category r given category r is reached, which is a discrete hazard. For details see, for example, Tutz (2012).

In both models the predictor has the form $\eta_r = \gamma_{0r} + \mathbf{x}^T \boldsymbol{\gamma}_r$. By allowing for *category-specific* effects $\boldsymbol{\gamma}_r^T = (\gamma_{r1}, \dots, \gamma_{rp})$ the model has as many parameters as the multinomial logit model. In its simpler version, where $\boldsymbol{\gamma}_r = \dots = \boldsymbol{\gamma}_{k-1} = \boldsymbol{\gamma}$

holds, the cumulative type model is also called the proportional odds model. Only in this form it fully uses that response categories are ordered. An intermediate case, where only some of the parameters are category-specific is the partial proportional odds model (for example, Cox, 1995, Brant, 1990, Peterson and Harrell, 1990). With many predictors it is a demanding problem to find out which parameters can be specified as *global*, that is, not varying over categories, and which ones as category-specific. In the exploration of the general model star plots can be helpful.

For the representation of effects it is useful to represent the models in a slightly different form. The cumulative logit model can be written as

$$\frac{P(Y \leq r|\mathbf{x})}{P(Y > r|\mathbf{x})} = e^{\gamma_{r0}} e^{x_1 \gamma_{r1}} \dots e^{x_p \gamma_{rp}} = e^{\gamma_{r0}} (e^{\gamma_{r1}})^{x_1} \dots (e^{\gamma_{rp}})^{x_p}.$$

Therefore, the exponential $e^{\gamma_{rj}}$ represents the multiplicative effect of variable j on the cumulative odds $P(Y \leq r|\mathbf{x})/P(Y > r|\mathbf{x})$ if x_j increases by one unit. It is the effect on the dichotomization into response categories $\{1, \dots, r\}$ and $\{r + 1, \dots, k\}$. For the sequential logit model one obtains

$$\frac{P(Y = r|\mathbf{x})}{P(Y > r|\mathbf{x})} = e^{\gamma_{r0}} e^{x_1 \gamma_{r1}} \dots e^{x_p \gamma_{rp}} = e^{\gamma_{r0}} (e^{\gamma_{r1}})^{x_1} \dots (e^{\gamma_{rp}})^{x_p}.$$

Therefore, the exponential $e^{\gamma_{rj}}$ represents the multiplicative effect of variable j on the continuation ratio odds $P(Y = r|\mathbf{x})/P(Y > r|\mathbf{x})$ if x_j increases by one unit.

In a star plot for the effects of variable x_j the length of the rays is given by $e^{\gamma_{1j}}, \dots, e^{\gamma_{k-1,j}}$. As in the multinomial logit model the (dashed) unit circle refers to the case where the j th variable can be neglected, that is, $\gamma_{1j} = \dots = \gamma_{k-1,j} = 0$. The p -value of the likelihood ratio test for the corresponding hypothesis $H_0 : \gamma_{1j} = \dots = \gamma_{k-1,j} = 0$ is denoted by p -rel since the relevance of the j th predictor is tested. When compared to the circle the stars show if the effects are larger than 1 (outside the circle) or smaller than 1 (inside the circle). In the sequential model that means that a variable that has values within the circle decreases the odds $P(Y = r|\mathbf{x})/P(Y > r|\mathbf{x})$, rays outside the circle represent variables that increase the odds. The interpretation of stars is the same as for the multinomial model, that is, closeness to the unit circle means that the variable is not influential.

For illustration we consider the data from the German Munich founder study. Data were collected on business founders who registered their new companies at the local chambers of commerce in Munich and surrounding administrative districts. The focus was on survival of firms measured in 7 categories, the first six represent failure in intervals of six months, the last category represents survival beyond 36 months. Various covariates are available, economic sector (1: industry, manufacturing companies and building sector, 2: commerce, 3: service industry), legal form (1: small trade without entry in the register of companies, 2: one man

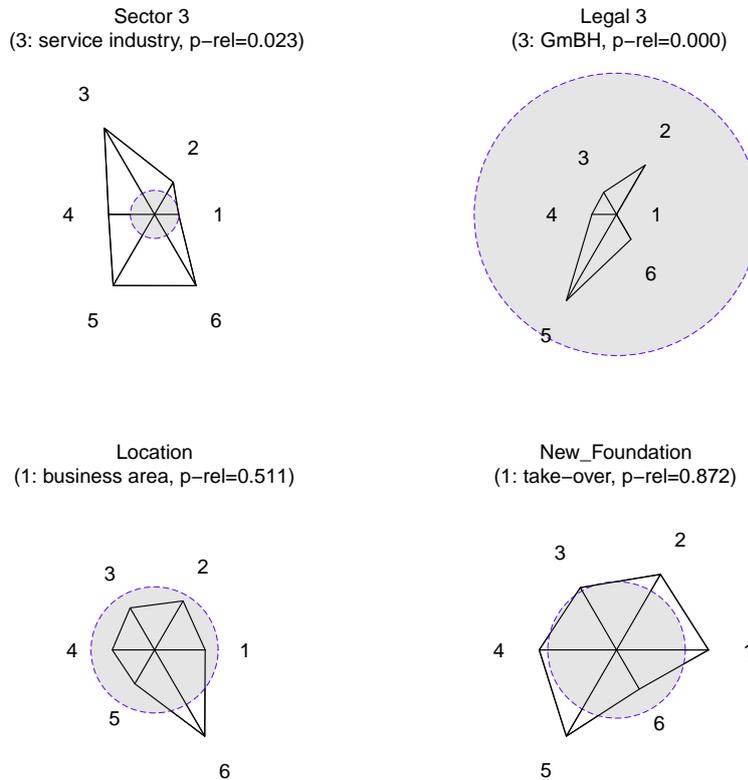


FIGURE 12: Stars for four predictors of the founder study with circles referring to the no-effects case. For the significant variables, sector3 and legal3 the stars are far away from the no-effects circle, for the significant variables, location and new foundation, they are quite close.

business merchant, 3: GmbH, GmbH & CoKG, 4: GbR, KG, OHG), location (0: residential area, 1: business area, industrial area or mixed), new (0: new foundation, 1: partial take-over, take-over, miscellaneous), pecuniary reward (0: main occupation, 1: additional occupation), seed capital (1: > 25000 , 0: ≤ 25000), equity capital (1: yes, 0: no), debt capital (1: yes, 0: no), market (0: local market, 1: national market), clientele (0: wide spread, 1: small amount of important customers), education of founder (1: A-levels, 0: minor), gender of founder (1: male, 0: female), experience (1: > 10 years, 0: ≤ 10), number of employees (1: > 2 , 0: ≤ 2), age of founder. The data of the Munich founder study have also been used by Brüderl et al. (1992) and Kauermann et al. (2005) and are available from the Central Archive for Empirical Social Research, University of Cologne, Germany (<http://www.gesis.org/ZA/>). We restrict our analysis to those firms that were founded completely new, which leaves us with 1224 cases. We fitted the full sequential logit model with all 18 predictors but show only four of the stars that resulted. Figure 12 shows the stars for sector3, legal3, location,

and new foundation. It is seen that the first two variables are highly significant. The variable sector3 has all values outside the circle, meaning that the odds increase if the firm is in the service industry as compared to reference category 1 (industry). For variable legal3 the star is distinctly inside the circle meaning that legal form 3 decreases the odds when compared to reference category 1 (small trade). For the other variables, location and new foundation, the stars are very close to the circle. Consequently both predictors are not significant (see value in brackets).

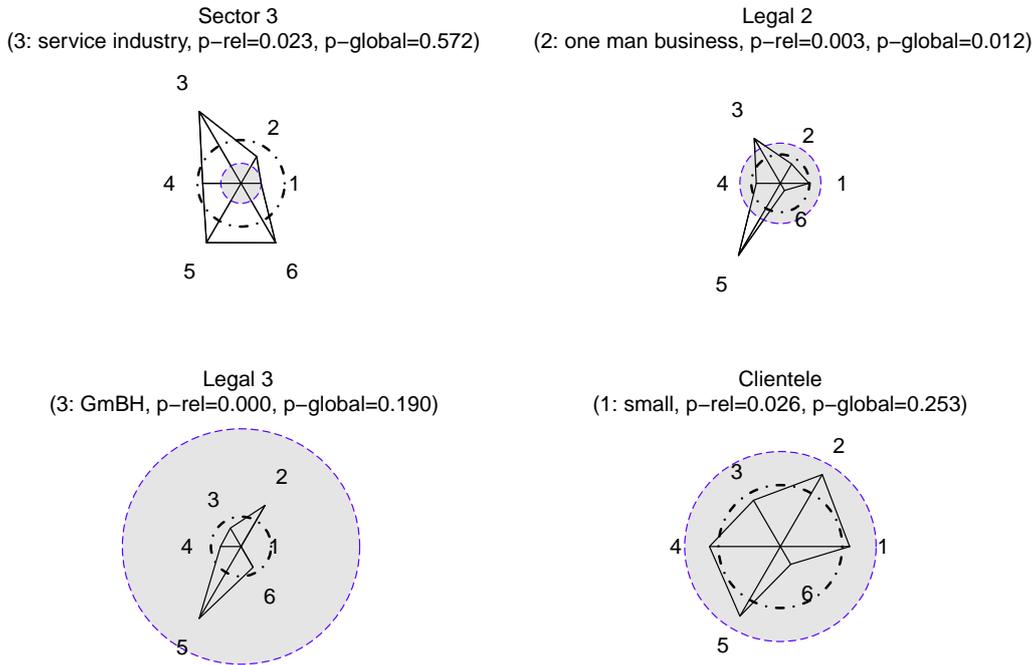


FIGURE 13: Stars for five predictors of the founder study. Deviation from the dashed circle implies relevance of the predictor, deviation from the dashed-dotted circle implies that predictor is category-specific.

In ordinal models a second effect is interesting, namely if the effects are category-specific or global. Therefore, a second (dashed-dotted) circle refers to the model with global effects only, that is, $\gamma_{1j} = \dots = \gamma_{k-1,j} = \gamma_j$. We fit the model that contains all predictors with category-specific effects with the exception of predictor j , which has global effect and include the circle with radius $\exp(\gamma_j)$. The interpretation of stars with respect to the dashed-dotted circle is different. Closeness to this circle means that the variable is global, strong deviation signals that it is category-specific. The hypothesis $H_0 : \gamma_{1j} = \dots = \gamma_{k-1,j} = \gamma_j$ investigates if the proportional odds assumption holds for the j th predictor. The

corresponding p -value of the test is denoted by p -global since the test investigates if the predictor has global effect. Figure 13 again shows the stars for only some predictors although the full model has been fitted. All of the predictors that are shown have significant effects (p -rel < 0.05). For predictors sector3, legal3 and clientele the hypothesis that effects are global is not rejected (p -global > 0.05). The corresponding stars are close to the dashed-dotted circle, although not close to the dashed circle, which represents relevance. For the variable legal2 the star is far away from the dashed-dotted circle and the hypothesis that the effects are global is rejected.

In the illustrations we used the sequential model. There are two reasons. First, the category-specific effects for the sequential model have a simple interpretation. Second, the cumulative model often raises problems when a model with category-specific effects is fitted. Maximum likelihood (ML) estimates may not exist because the parameter space is restricted in a complicated way, one has to postulate that $\gamma_{10} + \mathbf{x}^T \boldsymbol{\gamma}_1 \leq \dots \leq \gamma_{k-1,0} + \mathbf{x}^T \boldsymbol{\gamma}_{k-1}$ holds for all possible predictor values. If the maximum likelihood estimate does not exist an alternative is to use in the star plot for variable x_j values from the fitting of the global model, which gives the circle, and values from the fitting of the model

$$\log \left(\frac{P(Y \leq r | \mathbf{x})}{P(Y > r | \mathbf{x})} \right) = \gamma_{r0} + x_1 \gamma_1 + \dots + x_j \gamma_{rj} + \dots + x_p \gamma_p,$$

where only variable x_j has category-specific effects. But even then ML estimates often deteriorate.

6 Concluding Remarks

We proposed a method to visualize the fitted effects of a categorical response model. The method allows to identify the direction as well as the strength of the effects. For ordinal models it is distinguished between the relevance of a predictor and how strong the effects vary across the categories. Both aspects can be seen from the corresponding stars. The full strength of the visualization method is seen if one looks at the stars for all the covariates. In particular in the ordinal response case we showed only selected stars although much more predictors were used.

Star plots visualize parametric effects and therefore are useful for parametric models. In principle the method can be extended to visualize the effect strength in nonparametric models like the vector generalized additive model (Yee and Wild, 1996). If the predictor has the form $\eta_r = \beta_{r0} + \sum_j f_{rj}(x_j)$, where $f_{rj}(\cdot)$ are unspecified functions, the effect of variable j on the preference of category r is given by the function $f_{rj}(\cdot)$, which is typically centered around zero. The strength of the effect can be measured, for example, by $e_{rj} = \int |f_{rj}(x_j)| dx_j$. Since the function is centered one obtains $e_{rj} = 0$ if the predictor has no effect

and large values if the effect is strong. Stars with length of rays corresponding to $\exp(e_{rj})$ can be used to visualize the effect strengths for all categories. The disadvantage over stars for parametric models is that the direction of the effect gets lost: all strong effects will be found outside the circle, the no-effects case results yields rays close to the circle. The method can also be used for factorial predictors, if the integral is replaced by a sum over the corresponding absolute values of the effects. Then one obtains for one categorical predictor just one star instead of several stars. But, again, direction of effects gets lost. Therefore, in this situation we prefer to look at several stars since information content is higher. For nonparametric models with smooth effects effect displays provided by the effects package (Fox and Hong, 2009) seem to be preferable.

All the computations were done by use of the free software R (R Development Core Team (2010)). The program package Effect-Stars that generates and plots effect stars is available at CRAN and <http://www.statistik.lmu.de/institut/lehrstuhl/semsto/software/index.html>. It contains many options to modify the resulting stars.

Acknowledgement

The presentation of the method and the form of the stars has strongly benefitted from suggestions by an Associate Editor and a reviewer.

References

- Agresti, A. (2002). *Categorical Data Analysis*. New York: Wiley.
- Agresti, A. (2009). *Analysis of Ordinal Categorical Data, 2nd Edition*. New York: Wiley.
- Anderson, E. (1957). A semigraphical method for the analysis of complex problems. *Proceedings of the National Academy of Sciences of the United States of America* 43(10), 923.
- Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics* 46, 1171–1178.
- Brüderl, J., P. Preisendörfer, and R. Ziegler (1992). Survival chances of newly founded business organizations. *American Sociological Review* 57, 227–242.
- Chen, C., W. Härdle, and A. Unwin (2008). *Handbook of data visualization*. Springer Handbooks of Computational Statistics. Springer.
- Chernoff, H. (1973). The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association* 68(342), pp. 361–368.

- Cleveland, W. S. (1985). *The elements of graphing data*. Belmont, CA, USA: Wadsworth Publ. Co.
- Cox, C. (1995). Location-scale cumulative odds models for ordinal data: A generalized non-linear model approach. *Statistics in Medicine* 14, 1191–1203.
- Du Toit, S. H. C., A. G. W. Steyn, and R. H. Stumpf (1986). *Graphical exploratory data analysis*. New York, NY, USA: Springer-Verlag New York, Inc.
- Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software* 8(15), 1–27.
- Fox, J. and R. Andersen (2006). Effect displays for multinomial and proportional-odds logit models. *Sociological Methodology* 36(1), 225–255.
- Fox, J. and J. Hong (2009). Effect displays in R for multinomial and proportional-odds logit models: Extensions to the effects package. *Journal of Statistical Software* 32(1), 1–24.
- Fox, J., S. Weisberg, M. Friendly, and M. Fox (2011). Package 'effects'.
- Friendly, M. (1994). Mosaic displays for multi-way contingency tables. *Journal of the American Statistical Association*, 190–200.
- Gnanadesikan, R. (1977). *Methods for statistical data analysis of multivariate observations*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. Wiley.
- Hofmann, H. (2001). *Graphical Tools for the Exploration of Multivariate Categorical Data*. Books on Demand GmbH.
- Kastellec, J. and E. Leoni (2007). Using graphs instead of tables in political science. *Perspectives on Politics* 5(4), 755.
- Kauermann, G., G. Tutz, and J. Brüderl (2005). The survival of newly founded firms: a case-study into varying-coefficient models. *Journal of the Royal Statistical Society A* 168, 145–158.
- Meyer, D., A. Zeileis, and K. Hornik (2008). Visualizing contingency tables. *Handbook of Data Visualization*, 589–616.
- Peterson, B. and F. E. Harrell (1990). Partial proportional odds models for ordinal response variables. *Applied Statistics* 39, 205–217.
- R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

- Siegel, J., E. Farrell, R. Goldwyn, and H. Friedman (1972). The surgical implications of physiologic patterns in myocardial infarction shock. *Surgery* 72(1), 126.
- Theus, M. and S. Lauer (1999). Visualizing loglinear models. *Journal of Computational and Graphical Statistics* 8(3), 396–412.
- Tutz, G. (2012). *Regression for Categorical Data*. Cambridge University Press.
- Yee, T. W. and C. J. Wild (1996). Vector generalized additive models. *Journal of the Royal Statistical Society B*, 481–493.