EFFECT DISPLAYS FOR MULTINOMIAL AND PROPORTIONAL-ODDS LOGIT MODELS

*John Fox**
*Robert Andersen**

An "effect display" is a graphical or tabular summary of a statistical model based on high-order terms in the model. Effect displays have previously been defined by Fox (1987, 2003) for generalized linear models (including linear models). Such displays are especially compelling for complicated models—for example, those including interactions or polynomial terms. This paper extends effect displays to models commonly used for polytomous categorical response variables: the multinomial logit model and the proportional-odds logit model. Determining point estimates of effects for these models is a straightforward extension of results for the generalized linear model. Estimating sampling variation for effects on the probability scale in the multinomial and proportional-odds logit models is more challenging, however, and we use the delta method to derive approximate standard errors. Finally, we provide software for effect displays in the R statistical computing environment.

This is a revised version of a paper read at the ASA Methodology Conference 2004. Please address correspondence to John Fox, Department of Sociology, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4M4; jfox@mcmaster.ca. We are grateful to Georges Monette for checking some of the derivations in this paper, and to Michael Ornstein and two anonymous reviewers for helpful suggestions.

* McMaster University

1. INTRODUCTION

Effect displays, in the sense of Fox (1987, 2003), are tabular or—more often—graphical summaries of statistical models. Fox (1987) introduces effect displays for generalized linear models (including linear models); Fox (2003) refines these methods and provides software for their essentially automatic implementation.

The general idea underlying effect displays—to represent a statistical model by showing carefully selected portions of its response surface—is not limited to generalized linear models, however, nor even to models that incorporate linear predictors. Moreover, the essential idea of effect displays is not wholly original with Fox (1987). For example, adjusted means in analysis of covariance (introduced by Fisher, 1936) are a precursor to more general effect displays. Goodnight and Harvey's (1978) "least-squares means" in analysis of variance and covariance, and Searle, Speed, and Milliken's (1980) "estimated population marginal means" are effect displays restricted to interactions among factors (i.e., categorical predictors) in a linear model.

King, Tomz, and Wittenberg (2000) and Tomz, Wittenberg, and King (2003) have presented similar ideas, but their approach is based on Monte-Carlo simulation of a model. In contrast, the analytical results that we give below can be computed directly. Long (1997) discusses several strategies for presenting statistical models fit to categorical response variables, including displaying estimated probabilities. Hastie, Tibshirani, and Friedman's (2001, sec. 10.13.2) "partial dependence plots" and Weisberg's (2005: 185–90) "marginal model plots" are also related to effect displays.

The primary purpose of this paper is to extend the effect displays in Fox (1987, 2003) to the multinomial logit model and to the proportional-odds logit model, statistical models that find common application in social research. As we will show, this extension is largely straightforward, although the derivation of standard errors is challenging, particularly in the proportional-odds model. We begin by reviewing effect displays for generalized linear models, using as examples a binary logit model and a linear model. We then present results for the multinomial and proportional-odds logit models. In each of these sections, we illustrate the results with examples.

We see the main contribution of this paper as twofold: First, by extending the methods in Fox (1987, 2003) to models commonly used for polytomous data, we provide a means for carefully visualizing

models of this type that have complex structure, such as polynomial or spline regressors and interactions. Second, we derive standard errors for fitted probabilities in multinomial and proportional-odds logit models, permitting us to show statistical uncertainty in effect displays constructed for these models.

2. EFFECT DISPLAYS FOR GENERALIZED LINEAR MODELS: BACKGROUND AND PRELIMINARY EXAMPLES

A general principle of interpretation for statistical models containing terms that are marginal to others (in the sense of Nelder 1977) is that high-order terms should be combined with their lower-order relatives—for example, an interaction between two factors should be combined with the main effects marginal to the interaction. In conformity with this principle, Fox (1987) suggests identifying the high-order terms in a generalized linear model and computing fitted values for each such term. The lower-order "relatives" of a high-order term (e.g., main effects marginal to an interaction, or a linear and quadratic term in a third-order polynomial, which are marginal to the cubic term) are absorbed into the term, allowing the predictors appearing in the high-order term to range over their values. The values of other predictors are fixed at typical values: For example, a covariate could be fixed at its mean or median, a factor at its proportional distribution in the data, or to equal proportions in its several levels.

Some models have high-order terms that "overlap"—that is, that share a lower-order relative (other than the constant). Consider, for example, a generalized linear model that includes interactions AB, AC, and BC among the three factors A, B, and C. Although the three-way interaction ABC is not in the model, it is nevertheless illuminating to combine the three high-order terms and their lower-order relatives (see Fox 2003 and the example developed in Section 2.1).

Let us turn now to the generalized linear model (e.g., McCullagh and Nelder 1989 or Firth 1991) with linear predictor $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ and link function $g(\boldsymbol{\mu}) = \boldsymbol{\eta}$, where $\boldsymbol{\mu}$ is the expectation of the response vector $\mathbf{y}$. Here, everything falls into place very simply: We have an estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, along with the estimated covariance matrix $\widehat{V(\hat{\boldsymbol{\beta}})}$ of $\hat{\boldsymbol{\beta}}$.

Let the rows of $\mathbf{X}^*$ include all combinations of values of predictors appearing in a high-order term, along with typical values of the remaining predictors. The structure of $\mathbf{X}^*$ with respect, for example, to

interactions, is the same as that of the model matrix $\mathbf{X}$. Then the fitted values $\hat{\boldsymbol{\eta}}^* = \mathbf{X}^*\hat{\boldsymbol{\beta}}$ represent the effect in question, and a table or graph of these values—or, alternatively, of the fitted values transformed to the scale of the response, $g^{-1}(\hat{\boldsymbol{\eta}}^*)$—is an effect display. The standard errors of $\hat{\boldsymbol{\eta}}^*$, available as the square-root diagonal entries of $\mathbf{X}^*\widehat{V(\hat{\boldsymbol{\beta}})}\mathbf{X}^{*\prime}$, may be used to compute pointwise confidence intervals for the effects, the endpoints of which may then also be transformed to the scale of the response.

In an application, as we will illustrate presently, we prefer plotting on the scale of the linear predictor (where the structure of the model—for example, with respect to linearity—is preserved) but labeling the response axis on the scale of the response. This approach has the advantage of making the configuration of the display invariant with respect to the values at which the omitted predictors are held constant, in the sense that only the labeling of the response axis changes with a different selection of these values.[1]

2.1. *A Binary Logit Model: Toronto Arrests for Marijuana Possession*

Following Fox (2003), we construct effect displays for a binary logit model fit to data on police treatment of individuals arrested in Toronto for simple possession of small quantities of marijuana. (The data discussed here are part of a larger data set featured in a series of articles in the Toronto Star newspaper.) Under these circumstances, police have the option of releasing an arrestee with a summons to appear in court—similar to a traffic ticket—or bringing the individual to the police station for questioning and possible indictment. The principal question of interest is whether and how the probability of release is influenced by the subject's sex, race, age, employment status, and citizenship, the year in which the arrest took place, and the subject's previous police record. Most of these variables are self-explanatory, with the following exceptions:

- Race appears in the model as "color," and is coded as either "black" or "white." The original data included the additional categories

[1]As David Firth has pointed out to us, however, this invariance does not hold with respect to standard errors, which are affected by the fixed elements of $\mathbf{X}^*$, a fact that follows from considering effects as fitted values. Standard errors will tend to be smaller for components of $\mathbf{x}'$ near the center of the data.

"brown" and "other," but their meaning is ambiguous and their use relatively infrequent. Moreover, the motivation for collecting the data was to determine whether blacks and whites are treated differently by the police.

- The observations span the years 1997 through (part of) 2002. A few arrests in 1996 were eliminated. In the analysis reported below, year is treated as a factor (i.e., as a categorical predictor).
- When suspects are stopped by the police, their names are checked in six databases—of previous arrests, previous convictions, parole status, and so on. The variable "checks" records the number of databases on which an individual's name appeared.

Preliminary analysis of the data suggested a logit model including interactions between color and year and between color and age, and main effects of employment status, citizenship, and checks. The effects of age and checks appear to be reasonably linear on the logit scale and are modeled as such.

Estimated coefficients and their standard errors are shown in Table 1. Where predictors are represented by dummy regressors, the

TABLE 1
Maximum-Likelihood Estimates and Standard Errors for Coefficients in the Logit Model for the Toronto Marijuana-Arrests Data

Coefficient	Estimate	Standard Error
Constant	0.344	0.310
Employed (yes)	0.735	0.085
Citizen (yes)	0.586	0.114
Checks	−0.367	0.026
Color (white)	1.213	0.350
Year (1998)	−0.431	0.260
Year (1999)	−0.094	0.261
Year (2000)	−0.011	0.259
Year (2001)	0.243	0.263
Year (2002)	0.213	0.353
Age	0.029	0.009
Color (white) × Year (1998)	0.652	0.313
Color (white) × Year (1999)	0.156	0.307
Color (white) × Year (2000)	0.296	0.306
Color (white) × Year (2001)	−0.381	0.304
Color (white) × Year (2002)	−0.617	0.419
Color (white) × Age	−0.037	0.010

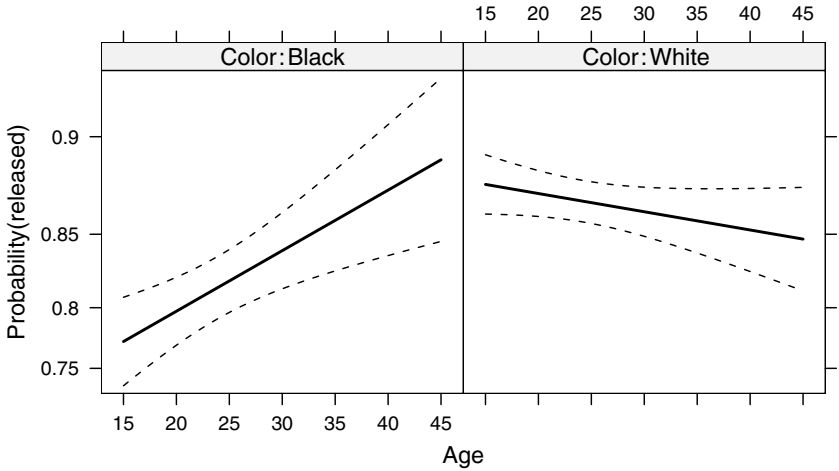

FIGURE 1. Effect display for the interaction of color and age in the logit model fit to the Toronto marijuana-arrests data. The vertical axis is labeled on the probability scale, and a 95-percent point-wise confidence envelope is drawn around the estimated effect. This graph, and those in Figures 2 and 3, are produced by the software described in Fox (2003).

category coded one is given in parentheses; for year, the baseline category is 1997. A fundamental point to be made with respect to this table is that it is difficult to tell from the coefficients alone how the predictors combine to influence the response. This difficulty is primarily a function of the complex structure of the model—that is, the interactions of color with year and age—but partly due to the fact that the coefficients are effects on the logit scale.[2] It is true that with some mental arithmetic we can draw certain conclusions from the table of coefficients. For example, the fitted probability of release declines with age for whites but increases with age for blacks. Grasping the color-by-year interaction is more difficult, however, as is discerning the combined effect of these three predictors.

Two effect displays for the model fit to the Toronto marijuana-arrests data appear in Figures 1 and 2. Figure 1 depicts the interaction between color and age. The lines in this graph are plotted on the logit

[2]A common device, which speaks partly to the second problem but not the first, is to exponentiate the coefficients in the logit model. The exponentiated coefficients are interpretable as multiplicative effects on the odds of the response. Interpreting interactions using exponentiated coefficients becomes even more difficult because it requires mental multiplication rather than addition.

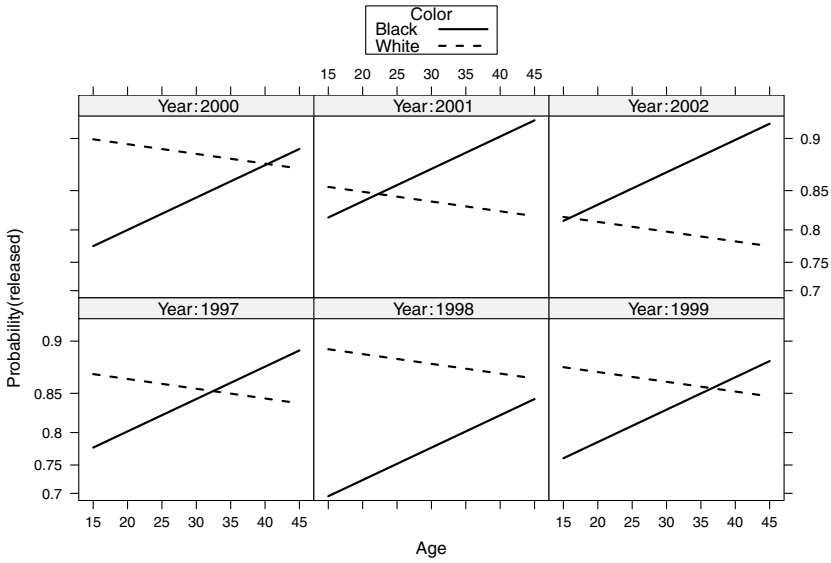

FIGURE 2. An effect display that combines the color-by-year and color-by-age interactions.

scale (i.e., the scale of the linear predictor), but the vertical axis of the graph is labeled on the probability scale (the scale of the response); the broken lines give point-wise 95-percent confidence envelopes around the fitted values. Figure 2 combines the color-by-age interaction with the color-by-year interaction. Because there is no three-way interaction (and no interaction between age and year), the lines for blacks are parallel across the six panels of the graph, as are the lines for whites. A graph such as Figure 2 effectively communicates what the model has to say about how color, age, and year combine to influence the probability of release.[3]

The $\mathbf{X}^*$ matrix for the effect display in Figure 1 has the following form:

[3]This and other examples in this paper include interactions between quantitative and categorical predictors, but effect displays are equally applicable to interactions between and among categorical predictors (and between and among quantitative predictors—e.g., see the example in Fox 1987). Indeed, handling categorical predictors is more straightforward because there is no need to make an arbitrary selection of values to plot, while two-dimensional plots of interactions between quantitative predictors require slicing the response surface. Effect displays for linear models containing only factors (i.e., analysis-of-variance models) are sometimes termed "least-squares means" (Goodnight and Harvey 1978) or "population marginal means" (Searle, Speed, and Milliken 1980).

$$
\begin{bmatrix}
(b_1) & (b_2) & (b_3) & (b_4) & (b_5) & (b_6) & (b_7) & (b_8) & (b_9) & (b_{10}) & (b_{11}) & (b_{12}) & (b_{13}) & (b_{14}) & (b_{15}) & (b_{16}) & (b_{17}) \\
1 & 0.79 & 0.85 & 1.64 & 0 & 0.17 & 0.21 & 0.24 & 0.23 & 0.05 & 15 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0.79 & 0.85 & 1.64 & 1 & 0.17 & 0.21 & 0.24 & 0.23 & 0.05 & 15 & 0.17 & 0.21 & 0.24 & 0.23 & 0.05 & 15 \\
1 & 0.79 & 0.85 & 1.64 & 0 & 0.17 & 0.21 & 0.24 & 0.23 & 0.05 & 16 & 0 & 0. & 0 & 0 & 0 & 0 \\
1 & 0.79 & 0.85 & 1.64 & 1 & 0.17 & 0.21 & 0.24 & 0.23 & 0.05 & 16 & 0.17 & 0.21 & 0.24 & 0.23 & 0.05 & 16 \\
1 & 0.79 & 0.85 & 1.64 & 0 & 0.17 & 0.21 & 0.24 & 0.23 & 0.05 & 17 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0.79 & 0.85 & 1.64 & 1 & 0.17 & 0.21 & 0.24 & 0.23 & 0.05 & 17 & 0.17 & 0.21 & 0.24 & 0.23 & 0.05 & 17 \\
1 & 0.79 & 0.85 & 1.64 & 0 & 0.17 & 0.21 & 0.24 & 0.23 & 0.05 & 18 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0.79 & 0.85 & 1.64 & 1 & 0.17 & 0.21 & 0.24 & 0.23 & 0.05 & 18 & 0.17 & 0.21 & 0.24 & 0.23 & 0.05 & 18 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & 0.79 & 0.85 & 1.64 & 1 & 0.17 & 0.21 & 0.24 & 0.23 & 0.05 & 45 & 0.17 & 0.21 & 0.24 & 0.23 & 0.05 & 45
\end{bmatrix}
$$

The columns in the matrix are labeled with the coefficients to which they pertain and are in the same order as in Table 1:

- The ones in the first column represent the regression constant.
- The second column contains the proportion of arrestees who were employed—that is, the mean of the dummy regressor for employment.
- The third column contains the proportion of arrestees who were Canadian citizens.
- The fourth column contains the average number of checks.
- The fifth column cycles through the two values of the dummy regressor for color.
- Columns six through ten contain the proportions of arrestees in the years 1998 through 2002; recall that 1997 is the baseline level for the dummy regressors for year.
- Column 11 cycles through the values of age, from 15 through 45. Because there are therefore $2 \times 31 = 62$ combinations of values of color and age, the $\mathbf{X}^*$ matrix has 62 rows.
- Columns 12 through 16 represent the interaction of color and year. Because year is "held constant" at its marginal distribution, this term is absorbed in the color main effect.
- Column 17 represents the color by age interaction.

2.2. *A Linear Model: Canadian Occupational Prestige*

The data for our second example, also adapted from Fox (2003), pertain to the rated prestige of 102 Canadian occupations. The prestige of the

TABLE 2
Coefficients for Regression of Occupational Prestige on Income and Education Levels of Occupations and on Percentage of Occupational Incumbents Who are Women

Coefficient	Estimate	Standard Error
Constant	−72.92	15.49
Log income	12.67	1.84
Education (1)	−8.20	7.8
Education (2)	25.66	5.50
Education (3)	30.42	4.59
Women (linear)	11.98	9.38
Women (quadratic)	18.47	6.83

Education is represented in the model by a three degree-of-freedom B-spline, percentage women by a second-order orthogonal polynomial.

occupations is regressed on three predictors, all derived from the 1971 Census of Canada: the average income of occupational incumbents, in dollars (represented in the model as the log of income); the average education of occupational incumbents, in years (represented by a B-spline with three degrees of freedom); and the percentage of occupational incumbents who were women (represented by an orthogonal polynomial of degree two). Estimated coefficients and standard errors for this model are shown in Table 2.

This model does a decent job of summarizing the data, but the meaning of its coefficients is relatively obscure—despite the fact that the model includes no interactions. The coefficient of log income, for example, would be more easily interpreted had we used logs to the base two rather than natural logs. The coefficients corresponding to the different elements of the B-spline basis do not have straightforward individual interpretations. Finally, although we can see from the coefficients for the orthogonal polynomial fit to the percentage of women that the linear trend in this predictor is nonsignificant while the quadratic trend is highly significant, these two coefficients are best interpreted in combination. It is therefore much more straightforward to apprehend these terms graphically as effect displays (Figure 3). We prefer to plot income on the natural scale rather than using a log horizontal axis, making the income effect nonlinear.

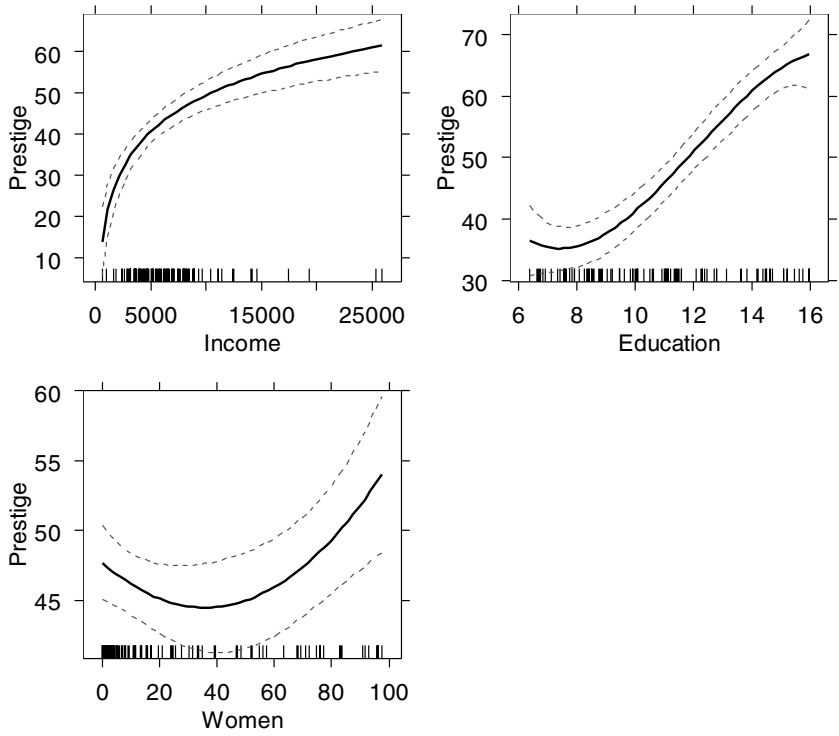

FIGURE 3. Effect plots for the predictors of prestige in the Canadian occupational prestige data. The model includes the log of income, a three-degree-of-freedom B-spline in education, and a quadratic in the percentage of occupational incumbents who are women. The "rug plot" (one-dimensional scatterplot) at the bottom of each graph shows the distribution of the corresponding predictor. The dashed lines give point-wise 95-percent confidence intervals around the fitted values.

3. EFFECT DISPLAYS FOR THE MULTINOMIAL LOGIT MODEL

3.1 *Basic Results*

The multinomial logit model is arguably the most widely used statistical model for polytomous (multicategory) response variables (e.g., McCullagh and Nelder 1989, chap. 5; Fox 1997, chap. 15; Long 1997, chap. 6; Powers and Xie 2000, chap. 7). Letting μ_{ij} denote the probability that observation i belongs to response category j of m categories, the model is given by

$$\mu_{ij} = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta}_j)}{\sum_{k=1}^{m} \exp(\mathbf{x}_{ik}'\boldsymbol{\beta}_{kj})} \quad \text{for} \quad j = 1, \ldots, m, \tag{1}$$

where $\mathbf{x}_i' = (1, x_{i2}, \ldots, x_{ip})$ is the model vector for observation i and $\boldsymbol{\beta}_j = (\beta_{j1}, \beta_{j2}, \ldots, \beta_{jp})'$ is the parameter vector for response category j. Observations may represent individuals, who therefore fall into a particular category of the response, or a vector of category counts for a multinomial observation (as in a contingency table, where both the predictors and the response variables are discrete); the first situation is a special case of the second, setting all of the multinomial total counts (i.e., the "multinomial denominators") n_i to 1.

As it stands, model 1 is overparametrized because of the constraint that the probabilities for each observation sum to one: $\sum_{j=1}^{m} \mu_{ij} = 1$. The resulting indeterminacy can be handled by a normalization, placing a linear constraint on the parameters, $\sum_{j=1}^{m} a_j\boldsymbol{\beta}_j = 0$, where the a_j are constants, not all zero. The choice of constraint is inessential: Fitted probabilities, $\hat{\mu}_{ij}$, and hence the likelihood, under the model are unaffected by the constraint, and consequently the effect displays developed in this paper are invariant with respect to the specific constraint employed; indeed, this invariance is a strength of effect displays. In contrast, the meaning of specific parameters depends upon the constraint, and as we will explain, adds to the difficulty of directly interpreting coefficient estimates for the model. The most common constraint is to set one of the $\boldsymbol{\beta}_j$ to zero (i.e., to set one of the a_j to 1 and the rest to 0); for convenience, we will set $\boldsymbol{\beta}_m = 0$, allowing us to rewrite equation (1) as

$$\mu_{ij} = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta}_j)}{1 + \sum_{k=1}^{m-1} \exp(\mathbf{x}_{ik}'\boldsymbol{\beta}_{kj})} \quad \text{for} \quad j = 1, \ldots, m-1$$

$$\mu_{im} = 1 - \sum_{k=1}^{m-1} \mu_{ik} \quad \text{(for category } m\text{)}. \tag{2}$$

Algebraic manipulation of model 2 suggests an interpretation of the coefficients of the model

$$\log \frac{\mu_{ij}}{\mu_{im}} = \mathbf{x}_i' \boldsymbol{\beta}_j \quad \text{for} \quad j = 1, \ldots, m-1, \tag{3}$$

and thus the coefficient vector $\boldsymbol{\beta}_j$ is for the log-odds of membership in category j versus the "baseline" category m. We can, moreover, express the log-odds of membership for any pair of categories in terms of *differences* in their coefficient vectors:

$$\log \frac{\mu_{ij}}{\mu_{ij'}} = \mathbf{x}_i'(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j'}) \quad \text{for} \quad j, j' \neq m. \tag{4}$$

All this is well and good, but it does not produce intuitively easy-to-grasp coefficients, since pair-wise comparison of the categories of the response is not in itself a natural manner in which to think about a polytomous variable. This difficulty of interpretation pertains even to models in which the structure of the model vector $\mathbf{x}'$ is simple.

Our strategy for building effect displays for the multinomial logit model is essentially the same as for generalized linear models: Find fitted values—in this case, fitted probabilities—under the model for selected combinations of values of the predictors. The fitted values on the probability scale, $\hat{\boldsymbol{\mu}}_{ij}$, are given by model 2, substituting estimates $\hat{\boldsymbol{\beta}}_j$ for the parameter vectors $\boldsymbol{\beta}_j$.

Finding standard errors for fitted values on the probability scale is more of a challenge, however. As is obvious from model 2, the fitted probabilities are nonlinear functions of the model parameters. We did not encounter this difficulty in the binary logit model because we could work on the scale of the linear predictor, translating the endpoints of confidence intervals to the probability scale (or equivalently, relabeling the logit axis). In the multinomial logit model, however, as noted, the linear predictor $\eta_{ij} = \mathbf{x}_i' \boldsymbol{\beta}_j$ is for the logit comparing category j to category m, not for the logit comparing category j to its complement, $\log[\mu_{ij}/(1-\mu_{ij})]$.

Suppose that we compute the fitted value at $\mathbf{x}'_0$ (e.g., a focal point in an effect display). Differentiating μ_{0j} with respect to the model parameters yields

$$\frac{\partial \mu_{0j}}{\partial \boldsymbol{\beta}_j} = \frac{\exp(\mathbf{x}_0' \boldsymbol{\beta}_j)\left[1 + \sum_{k=1, k\neq j}^{m-1} \exp(\mathbf{x}_0' \boldsymbol{\beta}_k)\right]\mathbf{x}_0}{\left[1 + \sum_{k=1}^{m-1} \exp(\mathbf{x}_0' \boldsymbol{\beta}_k)\right]^2}$$

$$\frac{\partial \mu_{0j}}{\partial \boldsymbol{\beta}_{j' \neq j}} = -\frac{\left\{\exp\left[\mathbf{x}_0'\left(\beta_{j'} + \boldsymbol{\beta}_j\right)\right]\right\}\mathbf{x}_0}{\left[1 + \sum_{k=1}^{m-1}\exp(\mathbf{x}_0'\boldsymbol{\beta}_k)\right]^2}$$

$$\frac{\partial \mu_{0m}}{\partial \boldsymbol{\beta}_j} = -\frac{\exp(\mathbf{x}_0'\boldsymbol{\beta}_j)\mathbf{x}_0}{\left[1 + \sum_{k=1}^{m-1}\exp(\mathbf{x}_0'\boldsymbol{\beta}_k)\right]^2}.$$

Let the estimated asymptotic covariance matrix of the (stacked) coefficient vectors be given by

$$\hat{V}(\hat{\boldsymbol{\beta}}) = \hat{V}\begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \\ \vdots \\ \hat{\boldsymbol{\beta}}_{m-1} \end{bmatrix} = [v_{st}], s, t = 1, \ldots, r$$

Here, $r = p(m-1)$ represents the total number of parameters in the combined parameter vectors. $\hat{v}(\hat{\boldsymbol{\beta}})$ is typically computed along with $\hat{\boldsymbol{\beta}}$ when the model is estimated. Then, by the delta method (Rao 1965: 321–27),

$$\hat{V}(\hat{\mu}_{0j}) \simeq \sum_{s=1}^{r}\sum_{t=1}^{r}\hat{v}_{st}\frac{\partial \hat{\mu}_{0j}}{\partial \hat{\boldsymbol{\beta}}_s}\frac{\partial \hat{\mu}_{0j}}{\partial \hat{\boldsymbol{\beta}}_t} \tag{5}$$

(where $\simeq$ denotes approximation).

Because the $\hat{\mu}_{0j}$ are bounded by 0 and 1, confidence intervals on the probability scale are problematic, especially for values near the boundaries. We therefore suggest the following refinement: Re-express the category probabilities μ_{0j} as logits,

$$\lambda_{0j} = \log\frac{\mu_{0j}}{1 - \mu_{0j}}. \tag{6}$$

These are *not* the paired-category (i.e., "baseline") logits (given in equations 3 and 4) to which the parameters of the multinomial logit model directly pertain but rather the log-odds of membership in each category relative to all others. Differentiating equation (6) with respect to μ_{0j} produces

$$\frac{d\lambda_{0j}}{d\mu_{0j}} = \frac{1}{\mu_{0j}(1-\mu_{0j})}$$

and, consequently, by a second application of the delta method,

$$\hat{V}(\hat{\lambda}_{0j}) \simeq \frac{1}{\hat{\mu}_{0j}^2(1-\hat{\mu}_{0j})^2}\hat{V}(\hat{\mu}_{0j}).$$

Using this result, we can form a confidence interval around $\hat{\lambda}_{0j}$, and translate the endpoints back to the probability scale.

This procedure applies regardless of the method used to estimate the parameters of the model and their covariances. For example, especially when the multinomial logit model is fit to aggregated data, overdispersion can be a concern. Following McCullagh and Nelder (1989, chap. 5), we can estimate the overdispersed multinomial-logit model by quasi-likelihood, producing the usual estimates of the regression coefficients, but inflating the coefficient variances (and covariances) by the estimated dispersion parameter, which is implicitly set to one in the traditional multinomial logit model. A similar remark applies to the proportional-odds logit model discussed in Section 4.[4]

3.2. *Example: Political Knowledge and Party Choice in Britain*

The example in this section is adapted from work by Andersen, Heath, and Sinnott (2002) on political knowledge and electoral choices in Britain (see also Andersen, Tilley, and Heath 2005). The data are from the 1997–2001 British Election Panel Study (BEPS). Although the same respondents were questioned at eight points in time, we use information only from the final wave of the study, which was conducted following the 2001 British election. After removing cases with missing data, the sample size is 2206.

We fit a multinomial logit model to describe how attitude toward European integration—an important issue during the 2001 British election—and knowledge of the major political parties' stances on

[4]As it turns out, overdispersion is not a problem for the examples developed in this and the following section: In both instances, the estimated dispersion—based, as suggested by McCullagh and Nelder (1989), on the Pearson statistic for the model—is slightly less than 1.

Europe interact in their effect on party choice. The variables in the model are as follows:

- The response variable is party choice, which has three categories: Labour, Conservative, and Liberal Democrat. Those who voted for other parties are excluded from the analysis. The Conservative platform was decidedly Eurosceptic, while both Labour and the Liberal Democrats took a clear pro-Europe position.
- "Europe" is an 11-point scale that measures respondents' attitudes toward European integration. High scores represent "Eurosceptic" sentiment.
- "Political knowledge" taps knowledge of party platforms on the European integration issue. The scale ranges from 0 (low knowledge) to 3 (high knowledge). An analysis of deviance suggests that a linear specification for knowledge is acceptable.
- The model also includes age, gender, perceptions of economic conditions over the past year (both national and household), and evaluations of the leaders of the three major parties.

Estimated coefficients and their standard errors from a final multinomial logit model fit to the data are shown in Table 3. We have already argued that interpreting coefficients in logit models is not simple, especially in the presence of interactions. Interpretation of the multinomial logit model is further complicated because the coefficients refer to contrasts of categories of the response variable with a baseline category. Nonetheless, we can see even from the coefficients that attitude toward Europe was related to party choice and that this relationship differed according to level of political knowledge. An analysis of deviance confirms that the interaction between attitude toward Europe and political knowledge is statistically significant. As was the case with the binary logit model, however, further interpretation is simplified by plotting this interaction as an effect display.

Figure 4 shows the relationship between attitude toward Europe and the fitted probability of voting for each of the three parties at the several levels of political knowledge (ranging from 0 to 3). An alternative display, with 95 percent confidence intervals around the fitted probabilities, appears in Figure 5. A third display, in Figure 6, shows the response categories in a manner similar to a stacked bar graph.[5]

[5]We are grateful to Michael Ornstein for suggesting this display.

TABLE 3
Coefficients for a Multinomial Logit Model Regressing Party Choice on Attitude Toward European Integration, Political Knowledge, and Other Explanatory Variables

Labour/Liberal Democrat		
Coefficient	Estimate	Standard Error
Constant	−0.155	0.612
Age	−0.005	0.005
Gender (male)	0.021	0.144
Perceptions of economy	0.377	0.091
Perceptions of household economic position	0.171	0.082
Evaluation of Blair (Labour leader)	0.546	0.071
Evaluation of Hague (Conservative leader)	−0.088	0.064
Evaluation of Kennedy (Liberal Democrat leader)	−0.416	0.072
Europe	−0.070	0.040
Political knowledge	−0.502	0.155
Europe × Knowledge	0.024	0.021
Conservative/Liberal Democrat		
Coefficient	Estimate	Standard Error
Constant	0.718	0.734
Age	0.015	0.006
Gender (male)	−0.091	0.178
Perceptions of economy	−0.145	0.110
Perceptions of household economic position	−0.008	0.101
Evaluation of Blair (Labour leader)	−0.278	0.079
Evaluation of Hague (Conservative leader)	0.781	0.079
Evaluation of Kennedy (Liberal Democrat leader)	−0.656	0.086
Europe	−0.068	0.049
Political knowledge	−1.160	0.0219
Europe × Knowledge	0.183	0.028

It is much easier to interpret the interaction between attitude and knowledge in these effect plots than directly from the coefficients: At the lowest level of knowledge, there is apparently no relationship between attitude toward Europe and party choice. In contrast, as knowledge increases, respondents are progressively more likely to match their votes to party platforms—that is, the more Eurosceptic votes are, the more likely they are to support the Conservative Party and the less likely they are to support Labour or the Liberal Democrats. We therefore see much

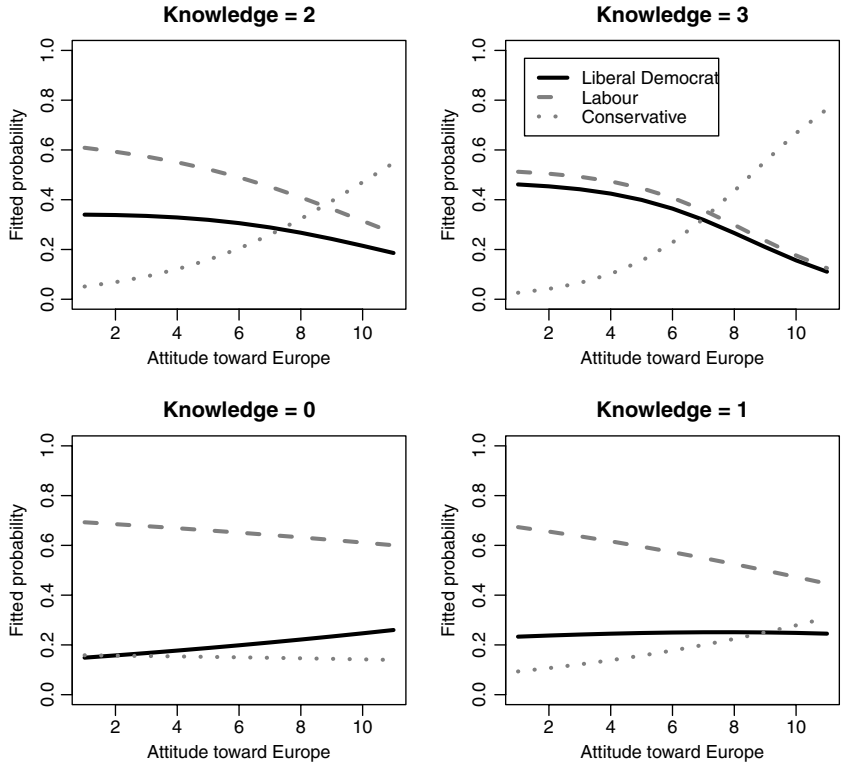

FIGURE 4. Display of the interaction between attitude toward Europe and political knowledge, showing the effects of these variables on the fitted probability of voting for each of the three major British parties in 2001.

more clearly than we could from Table 3 the importance of information to voting behavior—issues do matter in elections, but only to those who have knowledge of party platforms (a point discussed at greater length in Andersen 2003).

4. EFFECT DISPLAYS FOR THE PROPORTIONAL-ODDS LOGIT MODEL

4.1. *Basic Results*

The proportional-odds logit model is a common model for an ordinal response variable (e.g., McCullagh and Nelder 1989, chap. 5; Fox 1997,

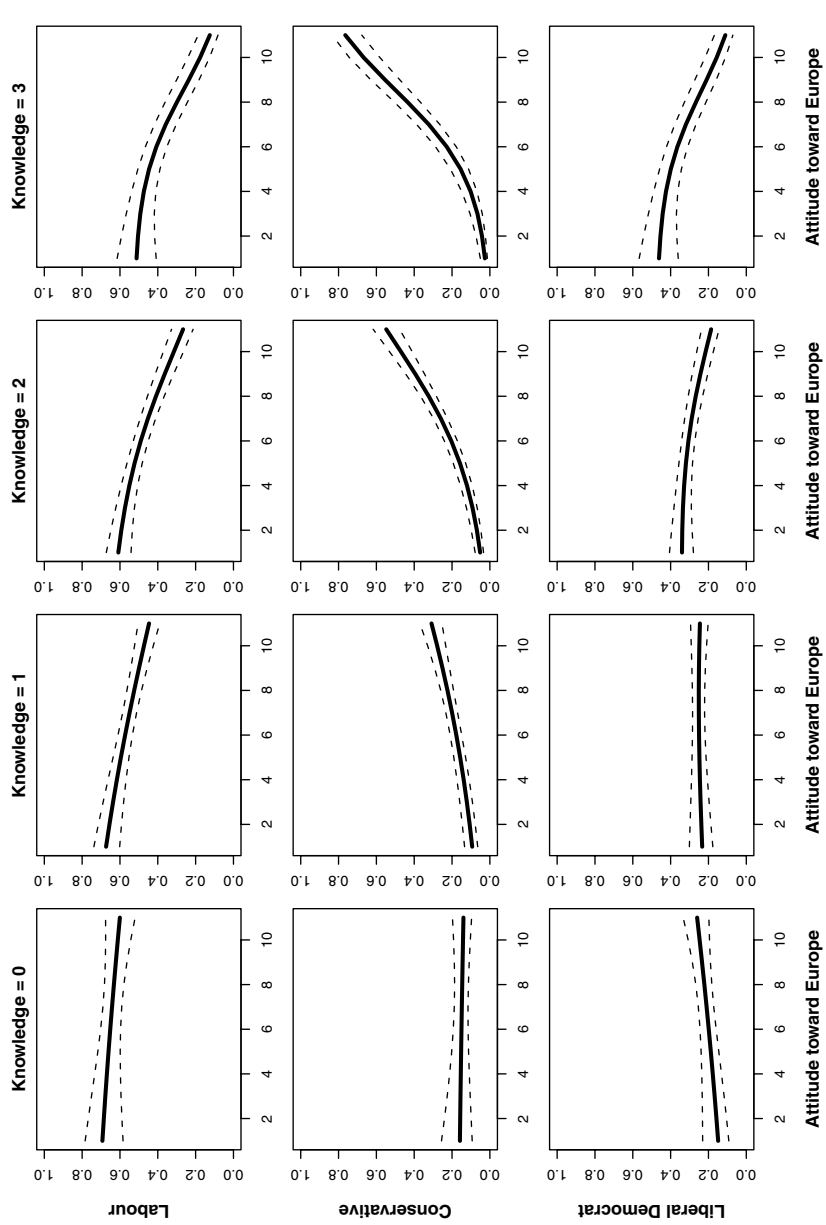

FIGURE 5. Alternative display of the interaction between attitude toward Europe and political knowledge. The dashed lines give point-wise 95-percent confidence intervals around the fitted probabilities.

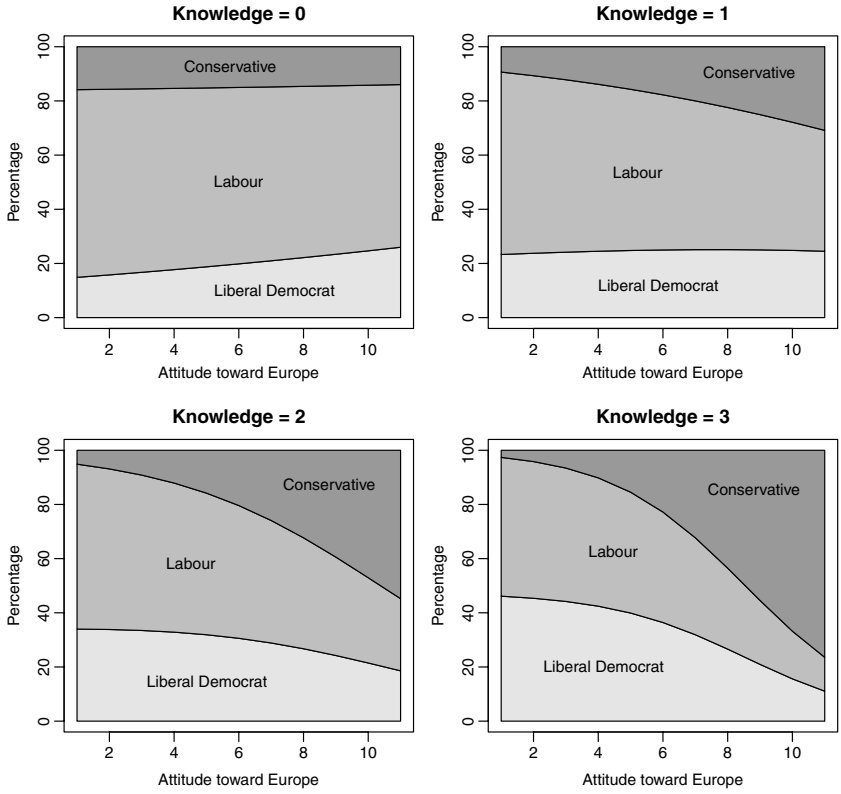

FIGURE 6. A third version of the effect display for the interaction between attitude toward Europe and political knowledge.

chap. 15; Long 1997, chap. 5; Powers and Xie 2000, chap. 6). The model is often motivated as follows: Suppose that there is a continuous, but unobservable, response variable, ξ, which is a linear function of a predictor vector $\mathbf{x}'$ plus a random error:

$$\begin{aligned} \xi_i &= \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i \\ &= \eta_i + \varepsilon_i \end{aligned}$$

We cannot observe ξ directly, but instead implicitly dissect its range into m class intervals at the (unknown) thresholds $\alpha_1 < \alpha_2 < \cdots < \alpha_{m-1}$, producing the observed ordinal response variable y. That is,

$$y_i = \begin{cases} 1 & \text{for} \quad \xi_i \leq \alpha_1 \\ 2 & \text{for} \quad \alpha_1 < \xi_i \leq \alpha_2 \\ \vdots & \\ m-1 & \text{for} \quad \alpha_{m-2} < \xi_i \leq \alpha_{m-1} \\ m & \text{for} \quad \alpha_{m-1} < \xi_i. \end{cases}$$

The cumulative probability distribution of y_i is given by

$$\begin{aligned} \Pr(y_i \leq j) &= \Pr(\xi_i \leq \alpha_j) \\ &= \Pr(\eta_i + \varepsilon_i \leq \alpha_j) \\ &= \Pr(\varepsilon_i \leq \alpha_j - \eta_i) \end{aligned}$$

for $j = 1, 2, \ldots, m-1$. If the errors ε_i are independently distributed according to the standard logistic distribution, with distribution function

$$\Lambda(z) = \frac{1}{1+e^{-z}},$$

then we get the proportional-odds logit model

$$\begin{aligned} \text{logit}[\Pr(y_i > j)] &= \log_e \frac{\Pr(y_i > j)}{\Pr(y_i \leq j)} \\ &= -\alpha_j + \boldsymbol{\beta}' \mathbf{x}_i \end{aligned} \tag{7}$$

for $j = 1,2,\ldots, m-1$. (The similar ordered probit model is produced by assuming that the ε_i are normally distributed.)

Model 7 is overparametrized: Since the $\boldsymbol{\beta}$ vector typically includes a constant, say β_1, we have $m-1$ regression equations, the intercepts of which are expressed in terms of m (i.e., one too many) parameters. A solution is to eliminate the constant from $\boldsymbol{\beta}$. Setting $\beta_1 = 0$ in this manner in effect establishes the origin of the latent continuum ξ; we already implicitly established the scale of ξ by fixing the variance of the error to the variance of the standard logistic distribution ($\pi^2/3$). For convenience, we will absorb the negative sign into the intercept,

rewriting the model as

$$\text{logit}[\Pr(y_i > j)] = \alpha_j + \boldsymbol{\beta}'\mathbf{x}_i, \quad \text{for} \quad j = 1, 2, \ldots, m-1.$$

The thresholds are then the negatives of the intercepts α_j. Because fitted probabilities under the model are unaffected by this reparametrization, effect displays are invariant as well.

The proportional-odds model is more parsimonious than the multinomial logit model (and other models for unordered polytomies): While the proportional-odds model has $m + p - 2$ independent parameters, the multinomial logit model has $p(m-1)$ independent parameters. Of course, the proportional-odds model is also less flexible, and may not adequately represent the data.

We propose two strategies for constructing effect displays for the proportional-odds model. The more straightforward strategy is to plot on the scale of the latent continuum, using the estimated thresholds, $-\hat{\alpha}_j$, to show the division of the continuum into ordered categories. There is not much more to say about this approach, since—other than marking the thresholds (as illustrated in the example in Section 4.2)—one proceeds exactly as for a linear model.

The second approach is to display fitted probabilities of category membership, as we did for the multinomial logit model. Suppose that we need the fitted probabilities at $\mathbf{x}'_0$ (where the constant regressor has been removed from the design vector $\mathbf{x}'$, and the intercept from the parameter vector $\boldsymbol{\beta}$). Let $\eta_0 = \mathbf{x}'_0\boldsymbol{\beta}$, and let $\mu_{0j} = \Pr(Y_0 = j)$. Then

$$\begin{aligned}
\mu_{01} &= \frac{1}{1 + \exp(\alpha_1 + \eta_0)} \\
\mu_{0j} &= \frac{\exp(\eta_0)\left[\exp(\alpha_{j-1}) - \exp(\alpha_j)\right]}{\left[1 + \exp(\alpha_{j-1} + \eta_0)\right]\left[1 + \exp(\alpha_j + \eta_0)\right]}, \; j = 2, \ldots, m-1 \\
\mu_{0m} &= 1 - \sum_{j=1}^{m-1} \mu_{0j}
\end{aligned}$$

As in the case of the multinomial logit model, we derive approximate standard errors by the delta method. The necessary derivatives are messier here, however:

$$\frac{\partial \mu_{01}}{\partial \alpha_1} = -\frac{\exp(\alpha_1 + \eta_0)}{[1 + \exp(\alpha_1 + \eta_0)]^2}$$

$$\frac{\partial \mu_{01}}{\partial \alpha_j} = 0,\ j = 2, \ldots, m-1$$

$$\frac{\partial \mu_{01}}{\partial \boldsymbol{\beta}} = -\frac{\exp(\alpha_1 + \eta_0)\mathbf{x}_0}{[1 + \exp(\alpha_1 + \eta_0)]^2}$$

$$\frac{\partial \mu_{0j}}{\partial \alpha_{j-1}} = \frac{\exp(\alpha_{j-1} + \eta_0)}{\left[1 + \exp(\alpha_{j-1} + \eta_0)\right]^2}$$

$$\frac{\partial \mu_{0j}}{\partial \alpha_j} = -\frac{\exp(\alpha_j + \eta_0)}{\left[1 + \exp(\alpha_j + \eta_0)\right]^2}$$

$$\frac{\partial \mu_{0j}}{\partial \alpha_{j'}} = 0,\ j' \neq j, j-1$$

$$\frac{\partial \mu_{0j}}{\partial \boldsymbol{\beta}} = \frac{\exp(\eta_0)\left[\exp(\alpha_j) - \exp(\alpha_{j-1})\right]\left[\exp(\alpha_{j-1} + \alpha_j + 2\eta_0) - 1\right]\mathbf{x}_0}{\left[1 + \exp(\alpha_{j-1} + \eta_0)\right]^2\left[1 + \exp(\alpha_j + \eta_0)\right]^2}$$

$$\frac{\partial \mu_{0m}}{\partial \alpha_{m-1}} = \frac{\exp(\alpha_{m-1} + \eta_0)}{[1 + \exp(\alpha_{m-1} + \eta_0)]^2}$$

$$\frac{\partial \mu_{0m}}{\partial \alpha_j} = 0,\ j = 1, \ldots, m-2$$

$$\frac{\partial \mu_{0m}}{\partial \boldsymbol{\beta}} = \frac{\exp(\alpha_{m-1} + \eta_0)\mathbf{x}_0}{[1 + \exp(\alpha_{m-1} + \eta_0)]^2}.$$

Let us stack up all of the parameters in the vector $\boldsymbol{\gamma} = (\alpha_1, \ldots, \alpha_{m-1}, \boldsymbol{\beta}')'$, and let

$$\hat{V}(\hat{\boldsymbol{\gamma}}) = [v_{st}],\ s, t = 1, \ldots, r,$$

where $r = m + p - 2$. Then, as for the multinomial logit model,

$$\hat{V}(\hat{\mu}_{0j}) \simeq \sum_{s=1}^{r}\sum_{t=1}^{r} v_{st}\frac{\partial \hat{\mu}_{0j}}{\partial \hat{\gamma}_s}\frac{\partial \hat{\mu}_{0j}}{\partial \hat{\gamma}_t}$$

and

$$\hat{V}(\hat{\lambda}_{0j}) \simeq \frac{1}{\hat{\mu}_{0j}^2(1 - \hat{\mu}_{0j})^2}\hat{V}(\hat{\mu}_{0j}),$$

where

$$\lambda_{0j} = \log \frac{\mu_{0j}}{1 - \mu_{0j}}$$

are the individual-category logits—that is, the log-odds of membership in a particular category versus all others, *not* the cumulative logits modeled directly by the proportional-odds model (given in equation 7).

4.2. *Example: Cross-National Differences in Attitudes Toward Government Efforts to Reduce Poverty*

We now turn to an application of effect displays to a proportional-odds logit model. Data for this example are taken from the World Values Survey of 1995–1997 (Inglehart et al., 2000). We use a subset of the World Values Survey, focusing on four countries (with sample sizes in parentheses): Australia (1874), Norway (1127), Sweden (1003), and the United States (1377). Although the variables that we employ are available for more than 40 countries, we restrict attention to these four nations to simplify the example. The variables in the model are as follows:

- The response variable is produced from answers to the question, "Do you think that what the government is doing for people in poverty in this country is about the right amount, too much, or too little?" We order the responses as too little < about right < too much.
- Explanatory variables include gender, religion (coded 1 if the respondent belonged to a religion, 0 if the respondent did not), education (coded 1 if the respondent had a university degree, 0 if not), and country (dummy coded, with Sweden as the reference category).

Preliminary analysis of the data suggested modeling the effect of age as a cubic polynomial (we use an orthogonal cubic polynomial) and including an interaction between age and country.[6] The coefficients and

[6]As a reviewer has pointed out, because of their nonlocal character, higher-order polynomial fits can be risky. Although we generally prefer more local fits such as regression splines, we use a cubic polynomial here to make a point about interpretation—that is, that multiple-degree-of-freedom effects, particularly involving interactions, are difficult to interpret from the coefficients. This is true of

TABLE 4
Coefficients for a Proportional-Odds Logit Model Regressing Attitude Toward Government Efforts to Help People in Poverty on Gender, Age, Religion, Education, and Country

Coefficient	Estimate	Standard Error
Gender (male)	0.169	0.053
Religion (yes)	−0.168	0.078
University degree (yes)	0.141	0.067
Age (linear)	10.659	5.404
Age (quadratic)	7.535	6.245
Age (cubic)	8.887	6.663
Norway	0.250	0.087
Australia	0.572	0.823
USA	1.176	0.087
Norway × Age (linear)	−7.905	7.091
Australia × Age (linear)	9.264	6.312
USA × Age (linear)	10.868	6.647
Norway × Age (quadratic)	−0.625	8.027
Australia × Age (quadratic)	−17.716	7.034
USA × Age (quadratic)	−7.692	7.352
Norway × Age (cubic)	0.485	8.568
Australia × Age (cubic)	−2.762	7.385
USA × Age (cubic)	−11.163	7.587
Thresholds		
Too little | about right	0.449	0.106
About right | too much	2.262	0.111

[a]Age is represented in the model by a cubic orthogonal polynomial, and interactions between age and country are included in the model.

their standard errors from a final model fit to the data are displayed in Table 4.

The complexity of the nonlinear trend for age, its interaction with country, and coefficients for cumulative logits make it extremely difficult to interpret the parameter estimates associated with age. Instead, we construct effect displays for the interaction of age with country. Figure 7 plots fitted probabilities for each category of the response variable

orthogonal polynomials, as used here, of ordinary polynomials (which provide the same fit to the data), and of regression splines. In fact, the cubic fit that we have employed represents the data well, and provides results similar to a regression spline. See Hastie and Tibshirani (1990, sec. 2.9) for a good discussion of regression splines and their general advantages relative to polynomial regression.

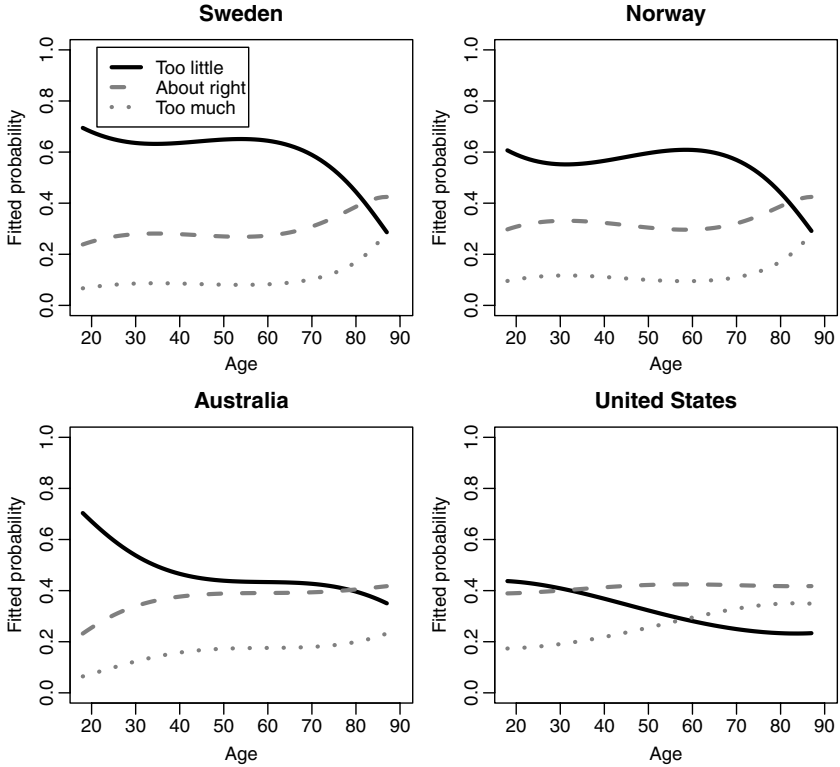

FIGURE 7. Display of the interaction between age and country, showing the effects of these variables on attitude toward government efforts to help people in poverty; the graphs indicate the fitted probability for each of the three categories of the response variable.

in the same manner as for the multinomial logit model of Section 2.2. Because country takes on only four values while age is continuous, we construct a separate plot for each country, placing age on the horizontal axis. There are three fitted lines in each plot—representing the fitted probability of choosing each response category. Figure 8 is generally similar, but with 95 percent point-wise confidence intervals around the fitted probabilities (and separate panels for each response category, so as not to clutter the plots). Figure 9 shows an alternative display with stacked response categories.

Although the graphs in Figures 7–9 are informative—we see, for example, that age differences are relatively muted in the United States

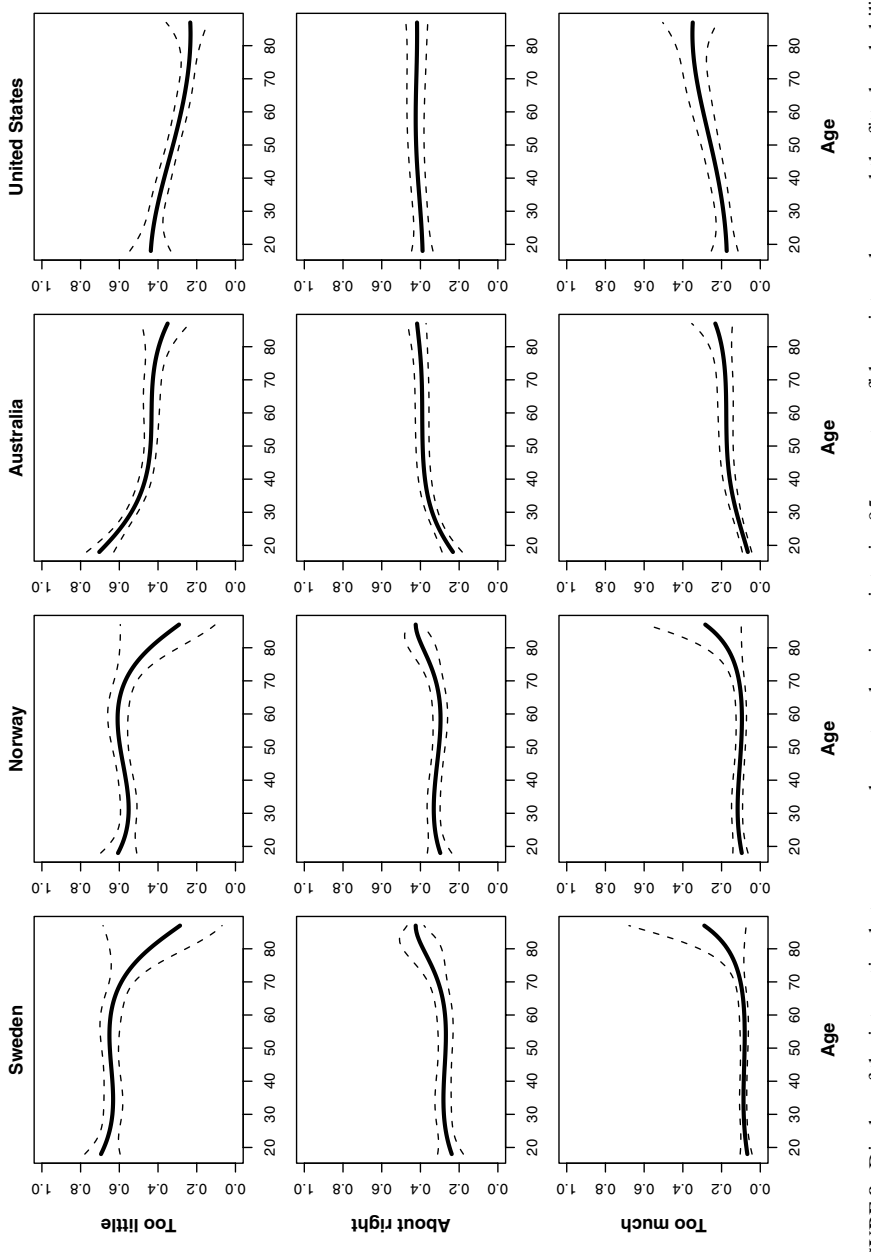

FIGURE 8. Display of the interaction between age and country, showing point-wise 95 percent confidence intervals around the fitted probabilities.

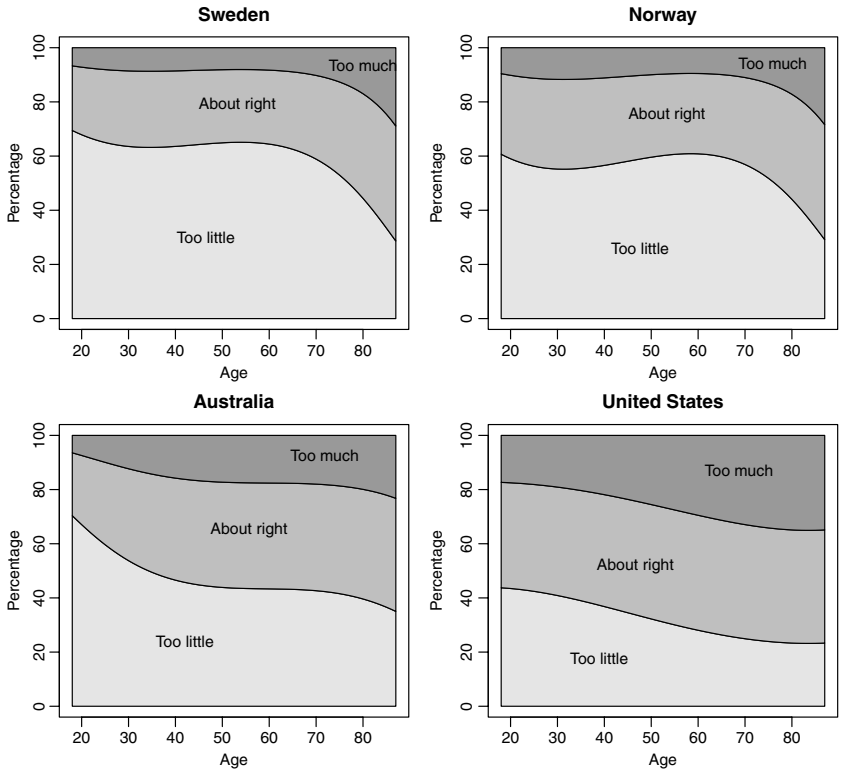

FIGURE 9. Alternative effect display of the interaction between age and country.

and that respondents there are less likely than others to feel that the government is not doing enough for the poor—the displays do not take full advantage of the parsimony of the proportional-odds model. We can capitalize on the structure of the proportional-odds model to plot the fitted response on the scale of the latent attitude continuum. We pursue this strategy in Figure 10, in which there is only one line for each country.[7] The estimated thresholds from the proportional-odds model are displayed as horizontal lines, dividing the latent continuum into three categories. Notice that none of the fitted curves exceeds the

[7]Abstract versions of Figure 10 are often used to explain the proportional-odds model (e.g., see Agresti 1990, fig. 9.2), but not typically to present the results of fitting the model to data and not for the kind of partial-effect plot developed in this paper.

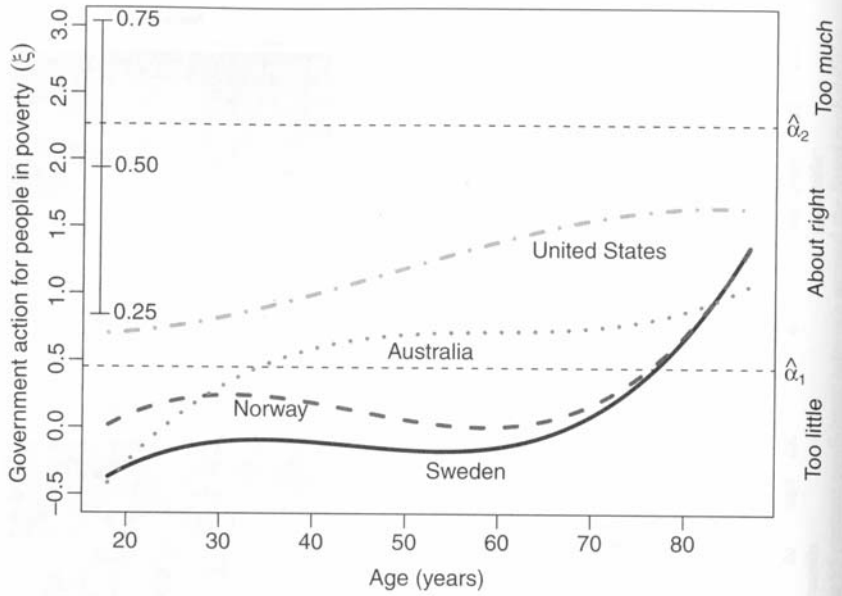

FIGURE 10. Plotting the interaction between age and country on the latent attitude continuum, ξ. The horizontal lines at $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are the thresholds between adjacent categories of the response.

second cut-point, and it is therefore unnecessary to include this cut-point in the graph; we do so to show explicitly that "too much" is never the modal response. The scale at the upper left of the graph shows the range spanned by the middle half of the standardized logistic distribution (i.e., the interquartile range, approximately $2 \times 1.1 = 2.2$ on the scale of the latent response), suggesting variation around the expected response; this is not to be confused with a confidence interval around the fitted response.

The patterns revealed by the effect displays are quite interesting: Even though their countries do more than the others to help those in poverty, people in Norway and Sweden are generally more likely than those in the United States or Australia to feel that the effort is insufficient. Moreover, attitudes are relatively similar among all age groups in the Scandinavian countries, with the exception of those at the highest ages, while in the United States and Australia, there are more general age trends toward decreased sympathy with the poor.

5. DISCUSSION

Statistical models for polytomous response variables are increasingly employed in social research. Too frequently, however, the results of fitting these models are described perfunctorily. Efforts to ensure careful model specification can be largely wasted if the results are not conveyed clearly. Although it is difficult to interpret the coefficients of complex statistical models that transform response probabilities nonlinearly, simply discussing their signs and statistical significance tells us little about the structure of the data. The approach described and illustrated in this paper, in contrast, goes a long way toward clarifying the fit of multinomial logit and proportional-odds models and simplifying their interpretation.

Effect displays allow us to visualize key portions of the response surface of a statistical model, and thus to understand better how explanatory variables combine to influence the response. The computation of effect displays for models of polytomous response variables is fairly straightforward and can be implemented in most statistical software. Computations associated with standard errors and confidence intervals for these effect displays are more difficult, however. We intend to extend the `effects` package for R (described in Fox 2003) to cover multinomial and proportional-odds logit models, making the construction of effect displays for these models essentially automatic. Until that time, a program described in the appendix to this paper may be employed for computing effects, their standard errors, and confidence limits.

APPENDIX: COMPUTING

Fitted values and their standard errors for effect displays may be computed with an R function (program), `polytomousEffects`, available on the web at <http://socserv.socsci.mcmaster.ca/jfox/Papers/polytomous-effect-displays.html>. Also available are code and data for the examples in this paper. R (Ihaka and Gentleman, 1996; R Development Core Team, 2004) is a free, open-source implementation of the S statistical computing environment now in widespread use, particularly among statisticians. The `polytomousEffects` function uses the strategy for safe prediction described in Hastie (1992, sec. 7.3.3) to ensure that fitted values are computed correctly in models with terms

(such as orthogonal polynomials and B-splines) whose basis depends upon the data.

REFERENCES

Agresti, A. 1990. *Categorical Data Analysis*. New York: Wiley.

Andersen, R. 2003. "Do Newspapers Enlighten Preferences? Personal Ideology, Party Choice, and the Electoral Cycle: The United Kingdom, 1992–97." *Canadian Journal of Political Science* 36:601–20.

Andersen, R., A. Heath, and R. Sinnott. 2002. "Political Knowledge and Electoral Choice." *British Elections and Parties Review* 12:11–27.

Andersen, R., J. Tilley, and A. Heath. 2005. "Political Knowledge and Enlightened Preferences." *British Journal of Political Science* 35:285–303.

Firth, D. 1991. "Generalized Linear Models." Pp. 55–82 in *Statistical Theory and Modeling: In Honour of Sir David Cox, FRS* edited by D. V. Hinkley, N. Reid, and E. J. Snell. London: Chapman and Hall.

Fisher, R. A. 1936. *Statistical Methods for Research Workers, 6th ed*. Edinburgh: Oliver and Boyd.

Fox, J. 1987. "Effect Displays for Generalized Linear Models." Pp. 347–61 in *Sociological Methodology*, vol. 17, edited by C. C. Clogg. Washington, DC: American Sociological Association.

—— 1997. *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, CA: Sage.

—— 2003. "Effect Displays in R for Generalised Linear Models." *Journal of Statistical Software* 8(15):1–27.

Goodnight, J. H., and W. R. Harvey. 1978. "Least Squares Means in the Fixed-Effect General Linear Model." Technical Report No. R-103. Cary, NC: SAS Institute.

Hastie, T. J. 1992. "Generalized Additive Models." Pp. 249–307 in *Statistical Models in S* edited by J. M. Chambers and T. J. Hastie. Pacific Grove, CA: Wadsworth.

Hastie, T. J., and R. J. Tibshirani. 1990. *Generalized Additive Models*. London: Chapman and Hall.

Hastie, T., R. Tibshirani, and J. Friedman. 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer.

Ihaka, R., and R. Gentleman. 1996. "R: A Language for Data Analysis and Graphics." *Journal of Computational and Graphical Statistics* 5:299–314.

Inglehart, R. E. A. 2000. *World values surveys and European value surveys, 1981–1984:1990–1993, and 1995–1997* [computer file]. Ann Arbor, MI: Institute for Social Research [producer], Inter–University Consortium for Political and Social Research [distributor].

King, G., M. Tomz, and J. Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44:347–61.

Long, J. S. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage.

McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*. 2d ed. London: Chapman and Hall.

Nelder, J. A. 1977. "A Reformulation of Linear Models" (with commentary). *Journal of the Royal Statistical Society*, Series A, 140:48–76.

Powers, D. A., and Y. Xie. 2000. *Statistical Methods for Categorical Data Analysis*. San Diego: Academic Press.

R Core Development Team. 2004. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rao, C. R. 1965. *Linear Statistical Inference and Its Applications*. New York: Wiley.

Searle, S. R., F. M. Speed, and G. A. Milliken. 1980. "Population Marginal Means in the Linear Model: An Alternative to Least Squares Means." *The American Statistician* 34:216–21.

Tomz, M., J. Wittenberg, and G. King. 2003. "Clarify: Software for Interpreting and Presenting Statistical Results." *Journal of Statistical Software* 8:1–29.

Weisberg, S. 2005. *Applied Linear Regression.* 3d ed. New York: Wiley.