The Korean Communications
in Statistics Vol. 7, No. 3, 2000
pp. 711-720

MARS Modeling for Ordinal Categorical Response Data: A Case Study[1)]

Ji-Hyun Kim[2)]

Abstract

A case study of modeling ordinal categorical response data with the MARS method is done. The study is to analyze the effect of some personal characteristics and socioeconomic status on the teenage marijuana use. The MARS method gave a new insight into the data set.

Keywords : Ordinal response, regression splines, proportional odds, logistic regression.

1. Introduction

This case study is on the analysis of a data set which has an ordinal categorical response variable and mixed (categorical and continuous) explanatory variables. This study was motivated by Du et al. (1998), which analyzes the effect of personal characteristics (gender, school performance) and a socioeconomic status (parent's education) on teenage marijuana use while accounting for unstructured socializing activities.

Logistic regression model with proportional odds assumption, called *proportional odds model*, is widely used for ordinal categorical response data. Du et al. (1998) adopted a completely nonparametric approach mainly because the proportional odds assumption of logitic regression model is violated.

On the other hand, Friedman (1991, 1993) advocated the MARS (Multivariate Adaptive Regression Splines) method for a flexible regression modeling of high dimensional data. The MARS method is best suitable for continuous response data even though Friedman (1991) gives one real example of binary logistic regression.

In this study we apply the MARS method to ordinal categorical response data to find a flexible and effective set of functions of explanatory variables, called a set of basis functions. For inferential purpose we fit the logistic regression model with the set of basis functions found by the MARS method. In return we get a new insight into the data set.

In Section 2, we describe the data and introduce briefly the nonparametric approach and the

---

1) This work was supported by Soongsil University Research Fund, 2000.

2) Associate Professor, Department of Statistics, Soongsil University, Dongjak-Ku Sangdo-Dong, Seoul 156-743, Korea.
E-mail: jhkim@stat.soongsil.ac.kr

MARS method. Data analysis is done in Section 3 with some appropriateness checking of the final model. Then concluding remarks are given in Section 4.

2. Data Description and Possible Methods

Data Description A nationally representative sample of 8,777 high school seniors in the United States was drawn and questionnaires were administered. Among many variables observed, we are concerned on the effect of four explanatory variables (index of unstructured socializing with peers, sex, self-reported average high school grade and parent's education) on the marijuana use in the past 12 months. For a detailed description of the more comprehensive study on the relationship between deviant behavior (marijuana use is one measure of it) and the way people spend their time, see Osgood et al. (1996).

The index of unstructured socializing with peers scored 4 to 21 is a sum of four items. Three items (Ride around in a car just for fun, Get together with friends informally, Go to parties or other social affairs) are on a scale of 1 (Never) to 5 (Almost everyday). And one item (During a typical week, on how many evenings do you go out for fun and recreation?) is on a scale of 1 (Less than one) to 6 (Six or seven). The other three explanatory variables reflect social differentiation. Sex was coded as 1 for males and 2 for females. Respondents' self-reported average high school grades as the indicator of respondents' future socioeconomic prospects were coded 1 for D through 9 for A. Parent's education as the indicator of socioeconomic status for the family is taken as the highest of both parents if both reported and is on a scale of 1 (grade school or less) to 6 (graduate or professional school). The response variable marijuana use has a scale ranging from 1 (for no use in the past 12 months) through 10 (40 or more times in the last 30 days).

Table 1 gives some descriptive statistics.

Table1: Descriptive statistics for variables

	Min.	Max.	Mean	Std. Dev.	Median	Mode
y(marijuana use)	1	10	2.57	2.49	1	1
x_1(socializing)	4	21	14.66	3.04	15	15
x_2(sex)	1	2	1.51	0.50	2	2
x_3(grades)	1	9	5.73	1.93	6	6
x_4(parent's education)	1	6	4.10	1.31	4	3

Analysis Based on Nonparametric Model The most popular model for ordinal categorical response data is the proportional odds model

$$\text{logit}\,[F_{\boldsymbol{x}}(j)] = \alpha_j - \boldsymbol{\beta}'\boldsymbol{x},\ \ j = 1, \ldots, J-1,$$

where $F_{\boldsymbol{x}}(j) = P(Y \le j \mid \boldsymbol{x})$. It assumes the proportional odds, or equal slope parameter $\boldsymbol{\beta}$ for all j. For our data set, however, the proportional odds assumption is violated with or without the interaction terms. (Specifically the SAS procedure LOGISTIC gives p-value 0.0001. For the score test of the proportional odds assumption in SAS, see Stokes et al., 1996, p.221.)

For the analysis of the marijuana use data, Du et al. (1998) proposed nonparametric hypotheses and test statistics, which were developed in Akritas et al. (2000). To briefly introduce their model, assume we have only one nominal categorical factor V (e.g. sex) and a continuous covariate W (e.g. index of unstructured socializing with peers). The nonparametric model decomposes the conditional distribution function $F_{ix}(j) = P(Y \leq j \mid V = i, W = x)$ as

$$F_{ix}(j) = M(j) + A_i(j) + D_x(j) + C_{ix}(j).$$

(In fact, they use a slightly modified definition of the distribution function $F_{ix}(j) = \frac{1}{2}P(Y \leq j \mid V = i, W = x) + \frac{1}{2}P(Y < j \mid V = i, W = x)$ to handle ties.) The function A_i can be regarded as the effect of factor level i averaged over the continuous covariate. For more precise definition of the functions M, A_i, D_x and C_{ix}, see Akritas et al. (2000). They consider the nonparametric hypotheses

$$H_0(A): \ A_i(j) = 0 \text{ for all } i \text{ and all } j$$

for no factor V effect after *adjusting* for the covariate effect and provide the corresponding test statistic. This model can be extended to multi-factor with a continuous covariate case (Akritas et al. 2000).

The nonparametric model focused on hypotheses testing has its advantage. Unlike the logistic regression model with the proportional odds assumption, the nonparametric model does not depend on the proportional odds assumption and is not affected by any monotone transformation of the ordered covariate. In addition to the hypotheses testing on main effects and interaction effects of factors, Du et al. (1998) also gives tests against patterned alternatives for main effects and some descriptive plots. It is desirable to have procedures testing for the covariate effect related to $D_x(j)$ and the interaction effect between factor and covariate, related to $C_{ix}(j)$. Such test procedures are not yet available.

With all the credibility of the nonparametric approach, the semiparametric approach like logistic regression still has its own merits if the underlying assumption holds. In this paper, we try to fit the logistic regression model with the help of the MARS method in finding an effective set of explanatory variables.

MARS Method The MARS model invented by Friedman (1991) is a flexible nonparametric regression model for high dimensional data. For the presumed system that generated the data

$$y = f(x_1, \cdots, x_p) + \varepsilon,$$

MARS model gives as an approximation to f

$$\hat{f}(\mathbf{x}) = \Sigma_{m=1}^{M} a_m B_m(\mathbf{x}), \tag{2.1}$$

where $B_m(\mathbf{x})$ takes the form of product spline basis functions. The number of basis functions M

and the form of basis function $B_m(\mathbf{x})$ (i.e. product degree and knot locations) as well as the parameters a_m are automatically determined by the data. The basic underlying assumption of MARS is that the function f is locally smooth. Friedman (1993) extended the MARS methodology to the model having nominal categorical explanatory variables to which the usual definition of smoothness does not apply. For the case of a simple categorical variable x such that $x \in \{c_1, \cdots, c_K\}$, the function estimate (2.1) can be written as

$$\hat{f}(x) = \Sigma_{m=1}^{M} a_m I(x \in A_m), \quad M \le K$$

where I is the indicator function and $A_1, \cdots, A_M$ are subsets of $\{c_1, \cdots, c_K\}$. The estimate with smaller M is said to be smoother. Friedman (1993) developed the MARS algorithm which accommodates mixed (i.e. categorical and continuous) explanatory variables.

3. Data Analysis

MARS Model for Teenage Marijuana Use Data The MARS modeling being best suited for a continuous response, it can also be applied to ordinal categorical response since the squared distance $[y - \hat{f}(\mathbf{x})]^2$ is still meaningful. The MARS model is fitted to our data set using the publicly available FORTRAN program MARS 3.6. Sex, variable x_2, is treated as nominal categorical and other explanatory variables are taken as ordinal variables.

MARS use the modified form of the generalized cross-validation criterion

$$GCV = \frac{1}{n} \Sigma_{i=1}^{n} [y_i - \hat{f}(\mathbf{x}_i)]^2 / [1 - \frac{\tilde{C}}{n}]^2$$

as a model selection criterion. $\tilde{C}$ is a complexity cost function of the model generating $\hat{f}$. The complexity cost function is a function of smoothing parameter d among others. Larger values of d will lead to smoother function estimates. We use sample reuse technique to automatically estimate the smoothing parameter d from the data. Such option can be activated in MARS 3.6 by simply calling subroutine **xvalid**. We have used 10-fold cross-validation for the estimation of optimal d.

A piecewise-linear approximation is employed, and the fully unconstrained MARS model allowing maximum order of interactions is used. The maximum number of basis functions, M_{max} is set to 40.

The MARS algorithm has selected the model with 9 basis functions. The estimated optimal d is 10.81, and the value of GCV is 5.115. The ANOVA decomposition on these 9 basis functions is given on Table 2 for easy interpretation.

Table 2: ANOVA decomposition on 9 basis functions

set of functions	std. dev.	-gcv	no. of basis functions	variable(s)
1	1.0150	5.787	4 (B_1, B_2, B_3, B_4)	1
2	0.1846	5.125	1 (B_5)	3
3	0.2364	5.125	3 (B_6, B_7, B_8)	2 3 4
4	0.1456	5.119	1 (B_9)	1 2 4

The 9 basis functions are; $B_1(\boldsymbol{x})=(x_1-16)_+$, $B_2(\boldsymbol{x})=(x_1-16)_-$, $B_3(\boldsymbol{x})=(x_1-19)_+$, $B_4(\boldsymbol{x})=(x_1-13)_+$, $B_5(\boldsymbol{x})=(x_3-4)_+$, $B_6(\boldsymbol{x})=I(x_2=1)(x_3-5)_+(x_4-1)_+$, $B_7(\boldsymbol{x})=I(x_2=1)(x_3-5)_-(x_4-1)_+$, $B_8(\boldsymbol{x})=I(x_2=1)(x_3-3)_+(x_4-1)_+$, $B_9(\boldsymbol{x})=I(x_2=2)(x_1-16)_+(x_4-1)_+$, where $(x-a)_+=(x-a)I(x\geq a)$ and $(x-a)_-=-(x-a)I(x<a)$.

The exact equation of the selected model is

$$\begin{aligned}\hat{f}(\boldsymbol{x}) &= 2.171+.405B_1(\boldsymbol{x})-.0881B_2(\boldsymbol{x})+.626B_3(\boldsymbol{x})+.181B_4(\boldsymbol{x}) \\ &\quad -.113B_5(\boldsymbol{x})-.0926B_6(\boldsymbol{x})+.0847B_7(\boldsymbol{x})+.0613B_8(\boldsymbol{x})-.0609B_9(\boldsymbol{x})\end{aligned} \tag{3.1}$$

From the selected model, we determine to omit one basis function representing the three-factor interaction term $B_9(\boldsymbol{x})=I(x_2=2)(x_1-16)_+(x_4-1)_+$, because first, the model without it incurs little increase in GCV value (from 5.115 to 5.119 as in the third column of Table 2), and secondly this term is not retained consistently in the final models with other options of MARS algorithm (e.g. if we allow the maximum order of interaction to be three rather than four, $B_8(\boldsymbol{x})$ is still selected but $B_9(\boldsymbol{x})$ is not).

Without $B_9(\boldsymbol{x})$ term, we get by the least squares method,

$$\begin{aligned}\hat{f}(\boldsymbol{x}) &= 2.127+.332B_1(\boldsymbol{x})-.0864B_2(\boldsymbol{x})+.662B_3(\boldsymbol{x})+.179B_4(\boldsymbol{x}) \\ &\quad -.111B_5(\boldsymbol{x})-.110B_6(\boldsymbol{x})+.0982B_7(\boldsymbol{x})+.0752B_8(\boldsymbol{x}).\end{aligned} \tag{3.2}$$

Figure 1 depicts the model (3.2). (There is little difference in plots between the models (3.1) and (3.2)) The equation looks complicated but the interpretation is quite simple using ANOVA decomposition as in Table 2 and Figure 1. The variable x_1, the index of unstructured socializing with peers, is the dominating predictor since without it the GCV increases most and its impact on $\hat{f}$ is the largest (see Table 2 and Figure 1(a)). The school performance and the socializing index have shown nonlinear effect as in Figure 1 (a) and (b). Figure 1(c) depicting $B_6(\boldsymbol{x})$ to $B_8(\boldsymbol{x})$ tells that male students with poor grades and highly educated parents tend to use more marijuana. It is interesting that only for male students there is an interaction effect between the school performance and the parent's education. Figure 1(d) combining Figure 1(b) and 1(c) should be used for the interpretation of the effect of x_3 for male students since an effect of a factor should be interpreted with other factors together if it is related with them. For female students, Figure 1(b) explains the effect of x_3.

Even though model (3.2) with Table 2 and Figure 1 gives a lot of insight, it lacks inferential measures. Thus we fit the proportional odds model with explanatory variables in (3.2). The justification for this comes from the property of the proportional odds model. If

$$\text{logit}\,[P(Y\leq j\mid \boldsymbol{x})]=\alpha_j-\boldsymbol{\beta}'\boldsymbol{x},\quad j=1,\ldots,J-1,$$

and

$$Y^{*}=\beta' x+\varepsilon$$

where ε has logistic distribution which looks much like normal distribution, then it is known that Y^{*} is an underlying continuous response variable in the sense that

$$Y=j \quad \text{if} \quad \alpha_{j-1} < Y^{*} \leq \alpha_{j,} \tag{3.3}$$

where $-\infty=\alpha_0 < \alpha_1 < \cdots < \alpha_J=\infty$. So, the model for Y^{*} is expected to be equally valid for $\text{logit}\,[P(Y\leq j)]$. However, model (3.2) is on the response Y, not on the unobservable Y^{*}. We expect the regression model on Y would not be much different from the model on Y^{*}. (All we need is the same set of predictors, not the same coefficients.)

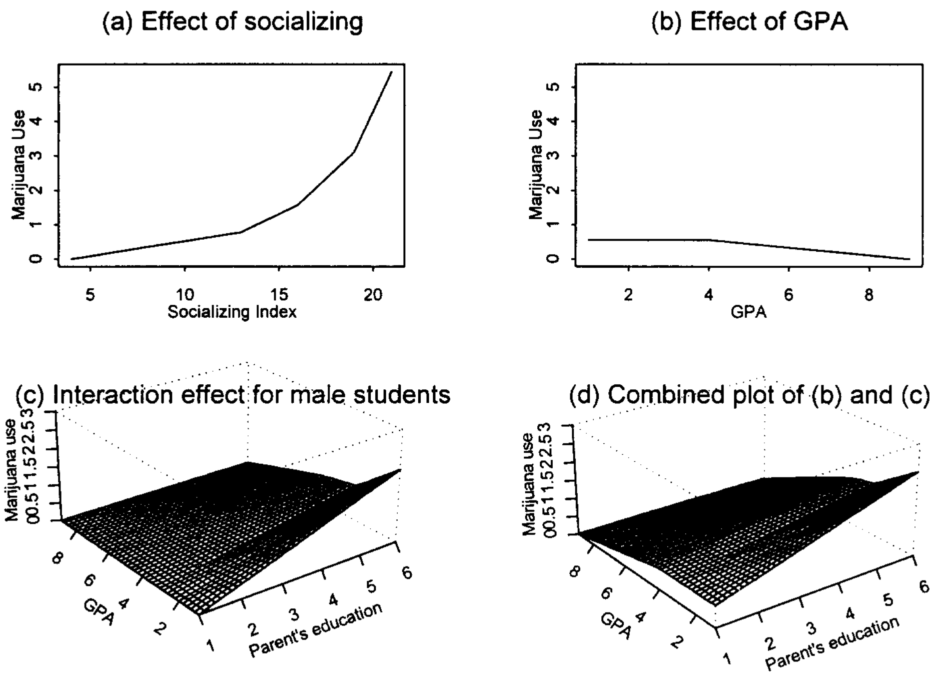

Figure 1. ANOVA functions for the MARS model (3.2)

The SAS procedure PROC LOGISTIC for the logit model

$$\text{logit}\,[P(Y\leq j\,|\,x)]=\alpha_j-\beta_1 B_1(x)-\cdots-\beta_8 B_8(x), \quad j=1,\ldots,J-1 \tag{3.4}$$

produces the coefficients and the corresponding p-values in Table 3.

We also fitted the model with $B_9(x)$ term, in which $\widehat{\beta}_9=.028$ (std.err. = .010, p-value = .0066) with minor changes in other coefficients. Although it is statistically significant, we prefer the model (3.4) without $B_9(x)$ term because; 1) its p-value is relatively large compared with other terms; 2) a statistical significance does not necessarily mean a practical significance in large sample (Agresti, 1996, p.161); 3) the inclusion of $B_9(\boldsymbol{x})$ term unnecessarily complicates the interpretation. On the other hand we retain $B_4(\boldsymbol{x}) = (x_1 - 13)_+$ term since it does not complicate the interpretation and it is an important term for the model (3.2) on Y.

Table 3: Estimates of Coefficients for Model

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$
coeff.	.260	-.219	.444	.052	-.139	-.077	.054	.054
std. err.	.060	.027	.107	.044	.018	.017	.010	.009
p-value	.0001	.0001	.0001	.2419	.0001	.0001	.0001	.0001

Appropriateness of the Proportional Odds Model The proportional odds assumption of the logit model (3.4) is still violated. The score test of SAS procedure PROC LOGISTIC for the assumption gives the p-value .0001 (chi-squared value 230.8 with $(10-2)\times 8 = 64$ degree of freedom). To see how much the assumption is violated, we fit nine binary logistic regression models for $\tilde{Y}_j = I(Y \leq j)$, $j = 1, \cdots, 9$, i.e.

$$\text{logit}\,[P(\tilde{Y}_j = 1 \mid \boldsymbol{x})] = \text{logit}\,[P(Y \leq j \mid \boldsymbol{x})] = \tilde{\alpha}_j - \beta_{1j}B_1(\boldsymbol{x}) - \cdots - \beta_{8j}B_8(\boldsymbol{x}). \quad (3.5)$$

The proportional odds model assumes H_0: $\beta_{i1} = \beta_{i2} = \cdots = \beta_{i9} (= \beta_i)$, $(i = 1, \cdots, 8)$ for our data set. Table 4 helps us to see how different the coefficients are for different categories of Y.

Table 4: Estimates of Coefficients for Model (3.5)

j	$\hat{\beta}_{1j}$	$\hat{\beta}_{2j}$	$\hat{\beta}_{3j}$	$\hat{\beta}_{4j}$	$\hat{\beta}_{5j}$	$\hat{\beta}_{6j}$	$\hat{\beta}_{7j}$	$\hat{\beta}_{8j}$
1	.182	-.216	.222	.064	-.142	-.058	.036	.040
2	.297	-.267	.252	-.003	-.122	-.073	.051	.046
3	.269	-.267	.307	.037	-.164	-.068	.040	.053
4	.236	-.230	.320	.079	-.149	-.090	.060	.067
5	.306	-.225	.225	.077	-.146	-.097	.085	.081
6	.284	-.218	.219	.105	-.145	-.140	.110	.113
7	.336	-.232	.172	.121	-.145	-.144	.095	.110
8	.346	-.167	.126	.149	-.222	-.120	.092	.097
9	.326	-.030	.174	.258	-.258	-.149	.111	.129

There is no reverse sign of coefficients in each column except one case of $\hat{\beta}_{4j}$. (Note that $\hat{\beta}_4$ is insignificantly different from 0 in Table 3.) And most of them are within the boundaries of a half the value of the common coefficient of model (3.4) in Table 4. (All of the exceptions occur in $\tilde{Y}_j$, $j \geq 6$, especially in $\tilde{Y}_9$. Note that only 2.8%, i.e. 244 out of 8,777, have the value Y equal to 10, and less than 15% have the value Y greater than 5.) A large sample as our data set gives small standard deviations of coefficients, so that the score test results are highly significant. But we do not require $\beta_{i1} = \beta_{i2} = \cdots = \beta_{i9} (= \beta_i)$ for all $i = 1, \cdots, 8$ in a strict sense. Practically we might allow as much differences as in Table 4. It is noteworthy that the score test for a naive model

$$\begin{aligned}\text{logit}\,[P(Y\le j\,|\,\boldsymbol{x})] &= \alpha_j-\beta_1x_1-\beta_2x_2^I-\beta_3x_3-\beta_4x_4-\beta_5x_2^Ix_3\\ &\quad -\beta_6x_2^Ix_4-\beta_7x_3x_4-\beta_8x_2^Ix_3x_4,\end{aligned}$$

where $x_2^I=I(x_2=1)$, gives chi-squared value 245.5 with 64 degrees of freedom. Thus MARS helped to reduce the level of violation of the proportional odds assumption.

Another measure for the appropriateness of the model (3.4) is a prediction error. To estimate a prediction error, we calculate

$$y^*=\hat{\beta}_1B_1(\boldsymbol{x})+\cdots+\hat{\beta}_8B_8(\boldsymbol{x}).$$

And we predict $\hat{y}=j$ if $\hat{\alpha}_{j-1}<y^*\le\hat{\alpha}_j,\ j=1,\cdots,10,$ where $\hat{\alpha}_0=-\infty$, $\hat{\alpha}_1,\ldots,\hat{\alpha}_9$ are obtained from PROC LOGISTIC and $\hat{\alpha}_{10}=\infty$. Out of 8,777 observations, model (3.4) gives 4,464 (50.9%) predictions $\hat{y}$ matching with y. If we allow error of prediction as much as $|y-\hat{y}|\le1$, the accuracy of prediction increases to 6,036/8,777(=68.8%). In order to see how good or bad this is, we fit a *saturated model*. Treating all explanatory variables as nominal categorical variables, we have $18\times2\times9\times6=1,944$ categories. Among them we have 1,253 nonempty categories excluding empty cells. (310 cells out of 1,253 have only one observation.) If we are free to choose a prediction $\hat{y}$ in each of 1,253 cells, we would take as our prediction the mode, i.e. the most frequently observed value of y in each cell to maximize the number of correct matches. We call it prediction from a saturated model. The numbers of observations predicted correctly or with some error by a saturated model have been counted by a FORTRAN program and are shown in Table 5. The saturated model gave 888($=5,352-4,464$) more correct predictions than model (3.4). But we got this gain at the cost of 1,236($=1,253-17$) more parameters. This fact confirms how effectively the model (3.4) describes the data.

Table 5: Number of Matched Observations for Various Models

no. of obs. such that	Saturated Model	Model (3.4)	Model (3.4) with $B_9(\mathbf{x})$	Naive Model
$y=\hat{y}$	5352	4464	4442	4186
$\|y-\hat{y}\|\le1$	6753	6036	6036	5968
$\|y-\hat{y}\|\le2$	7524	6826	6824	6837

Table 5 suggests that we cannot do much with the given data set. To see how much we can explain y with the given explanatory variables x_1,x_2,x_3,x_4, we calculated $R^2=[\,\text{corr}(y,\hat{y})]^2$. To maximize R^2 in the saturated model, it is better to take the mean of observations in each cell instead of the mode. $R^2=0.307$ for the saturated model and $R^2=0.189$ for the MARS regression model (3.2). It should be mentioned again that the gain of R^2 of the saturated model was at the cost of 1,236 more parameters. Furthermore, for future data set independent of the given data set, the saturated model perform poorly while the parsimonious models like (3.2) or (3.4) would not be affected much. Small R^2 of the saturated model indicates that we need other explanatory variables such as self control to explain the variation of y more satisfactorily.

Interpretation of the Logit Model If we can practically assume the proportional odds assumption, we can interpret model (3.4) with the odds of response below any given category or the cumulative probabilities. For instance, females ($x_2 = 2$) with poor school performance ($x_3 = 1$) and moderate socializing ($x_1 = 14$) has the probability that they did not use marijuana at all in the past 12 months as

$$\hat{P}(Y=1 \mid x_1 = 14, x_2 = 2, x_3 = 1) = \frac{\exp(\hat{\alpha}_1 - \hat{\beta}_2 \times 2 - \hat{\beta}_4 \times 1)}{1 + \exp(\hat{\alpha}_1 - \hat{\beta}_2 \times 2 - \hat{\beta}_4 \times 1)} = .62,$$

irrespective of their parents' education level. On the other hand, for male students ($x_2 = 1$) with the same school performance and socializing index, the estimated probability is .62 when the parents' education level is low ($x_4 = 1$), while the probability goes down to .35 when the parents' education level is high ($x_4 = 6$). And the odds ratio

$$\frac{P(Y \leq j \mid \mathbf{x} = (14,1,1,1))/P(Y > j \mid \mathbf{x} = (14,1,1,1))}{P(Y \leq j \mid \mathbf{x} = (14,1,1,6))/P(Y > j \mid \mathbf{x} = (14,1,1,6))}$$

is estimated as $e^{\hat{\beta}_7 \times 20} = 3.0$ for any response category $j = 1, \cdots, 9$, which tells that the parent's education is a considerably significant factor on the marijuana use for male students with poor school performance.

In order to see that the two main effects and one three-factor interaction effect in model (3.4) remain valid without the proportional odds assumption, we did a series of significance tests on those 3 effects in model (3.5). (There are $9 \times 3 = 27$ null hypotheses to be tested, such that $H_{01j}: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$; $H_{02j}: \beta_5 = 0$; $H_{03j}: \beta_6 = \beta_7 = \beta_8 = 0$, $j = 1, \ldots, 9$.) They are all highly significant. (With two exceptions, .004 and .009 on the effect of x_3, all p-values are less than .0001. Since the p-values are so small, type I error due to multiple tests does not cause problem.) So, our findings on the two main effects and one interaction effect are valid at least in a qualitative sense. Du et al. (1998) also found that marijuana use tends to increase with parent's education. We can conclude more specifically in this analysis that the effect of parent's education tends to be larger for male students with lower school performance.

4. Conclusion

The MARS modeling on the teenage marijuana use data has revealed that the male students with low school performance tend to use more marijuana when their parents get more highly educated. It has also found some nonlinear effects of the unstructured socializing with peers and the school performance. More quantitative interpretation can be made using either model (3.2) or model (3.4).

For inferential measures, proportional odds model was attempted with the basis functions found by MARS as explanatory variables. Although the assumption of proportional odds failed in the statistical significance test, some findings and arguments were made, asserting that the assumption can be made for a practical purpose. Even without the proportional odds assumption, the qualitative findings by MARS on the effects of unstructured socializing, school performance and the interaction between school performance and parent's education for male students remain valid since they are all highly significant in model (3.5).

This case study suggests that the MARS method is very helpful in the exploratory step and also in the modeling step for the regression with ordinal categorical response as well as with continuous response.

Acknowledgments. Part of this work was done while the author was on a sabbatical leave at the Department of Statistics, the Pennsylvania State University. The author wishes to thank the Department for the hospitality. He is also very grateful to Professors Akritas and Osgood for providing the teenage marijuana use data.

References

[1] Agresti, A., (1996). *An Introduction to Categorical Data Analysis.* New York: Wiley.

[2] Akritas, M. G., Arnold, S. F. and Du, Y., (2000). Nonparametric models and methods for nonlinear analysis of covariance. *Biometrika.* In press.

[3] Du, Y., Akritas, M. G., Arnold, S. F. and Osgood, D. W., (1998). Analysis of teenage deviant behavior data. Submitted.

[4] Friedman, J. H., (1991). Multivariate adaptive regression splines (with discussion). *Annals of Statistics*, 19, 1-141.

[5] Friedman, J. H., (1993). Estimating functions of mixed ordinal and categorical variables using adaptive splines. *New Directions in Statistical Data Analysis and Robustness*, 73-113.

[6] Osgood, D. W., Wilson, J. K., O'Malley, P. M., Bachman, J. G. and Johnston, L. D., (1996). Routine activities and individual deviant behavior. *American Sociological Review*, 61, 635-665.

[7] Stokes, M. E., Davis, C. S. and Koch, G. G., (1996). *Categorical Data Analysis using the SAS System.* SAS Institute, Inc., Cary, NC.