

# Assessing Practical Significance of the Proportional Odds Assumption

Ji-Hyun Kim <sup>1</sup>

Department of Statistics, Soongsil University, Korea.

## Abstract

The proportional odds model is a regression model widely used for ordinal categorical response, which has a rather strong underlying assumption, the proportional odds assumption. The rejection of the null assumption, however, is not very informative since a statistical significance does not necessarily imply a practical significance. This paper discusses the limitation of the statistical significance test of the proportional odds assumption, and proposes graphical methods that can be used to help assess the practical significance of the assumption.

*Keywords:* Ordinal response, logistic regression, graphical methods.

## 1. Introduction

Categorical response variables measured on an ordinal scale (for example, none, mild, severe) arise in many fields of study. The proportional odds model (McCullagh, 1980) is a popular regression model for ordinal categorical response. The model utilizing the ordinal nature of the response gives a simple interpretation since it can be considered as a regression model for an underlying continuous response variable (Agresti 1990). But the assumption of the model, the proportional odds assumption, is rather strong and needs to be checked. Some standard statistical package produces the test result for the assumption. This test result, however, is of only limited use for a practical purpose. When the null assumption of proportional odds is rejected, we need to know how much it is violated in a practical sense.

This paper discusses the limitation of the null hypothesis test of the proportional odds assumption, and proposes graphical methods that can be used to help assess the practical

---

<sup>1</sup>Corresponding address: Department of Statistics, Soongsil University, Dongjak-Ku Sangdo-Dong, Seoul 156-743, KOREA (E-mail: jhkim@stat.ssu.ac.kr)

significance of the assumption. Section 2 presents the formulation of the proportional odds model, and reports a simulation study which demonstrates the limitation of the null hypothesis significance tests. Section 3 proposes graphical methods helping to measure the degree of violation of the proportional odds assumption. The graphical methods are applied to a real example in Section 4.

## 2. Statistical Significance vs. Practical Significance of Test for Proportional Odds Assumption

We observe data of the form  $\{(y_i, x_{i1}, x_{i2}, \dots, x_{ip}); i = 1, \dots, n\}$ . The  $i$ th response  $y_i$  takes a value in a set of ordered categories  $\{1, 2, \dots, J\}$ , and  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$  is a vector of covariates for subject  $i$ . When response categories are ordered, the most popular model is the proportional odds model

$$\text{logit}[F_{\mathbf{x}}(j)] = \log \frac{F_{\mathbf{x}}(j)}{1 - F_{\mathbf{x}}(j)} = \alpha_j - \mathbf{x}'\boldsymbol{\beta}, \quad j = 1, \dots, J - 1, \quad (2.1)$$

where  $F_{\mathbf{x}}(j) = P(Y \leq j | \mathbf{x})$ . It assumes proportional odds, i.e. equal slope parameter  $\boldsymbol{\beta}$  for all  $j$ .

SAS 6.12 (SAS Institute Inc., Cary, NC), specifically the SAS procedure PROC LOGISTIC, performs a test for the appropriateness of the proportional odds assumption. For this test the alternative model considered is

$$\text{logit}[F_{\mathbf{x}}(j)] = \alpha_j - \mathbf{x}'\boldsymbol{\beta}_j, \quad (2.2)$$

for  $j = 1, \dots, J - 1$ . We assume in this paper that this model fits the data properly. The null hypothesis tested for further reduction is that

$$\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \dots = \boldsymbol{\beta}_{J-1} = \boldsymbol{\beta} \quad (2.3)$$

or equivalently,

$$\beta_{m1} = \beta_{m2} = \dots = \beta_{m,J-1},$$

for all  $m = 1, \dots, p$ , where  $\beta_{mj}$  is the  $m$ th element of  $\boldsymbol{\beta}_j$ . Since the test is comparing  $J - 1$  parameters across  $p$  covariates, and its chi-squared score test statistic is evaluated at  $\hat{\boldsymbol{\beta}}$  which

is the maximum likelihood (ML) estimate under the null hypothesis, it has  $p \times (J - 2)$  degrees of freedom. (cf. SAS/STAT User's guide, version 6, fourth edition, p.1090)

The null hypothesis (2.3) is a strong assumption since it assumes on  $p$  sets of equal  $J - 1$  parameters. We cannot expect this assumption to be true, nor do we need it to be in a strict sense for a practical purpose. A simulation study is done to see how the test behaves for the practically ignorable deviation of the slope parameter values from the null hypothesis.

The model considered for simulation is

$$\text{logit}[F_{\mathbf{x}}(j)] = \alpha_j + \beta_{1j}x_1 - \beta_{2j}x_2 + \cdots + (-1)^{p+1}\beta_{pj}x_p, \quad j = 1, \dots, J - 1, \quad (2.4)$$

where

$$\alpha_j = \log \frac{j/J}{1 - j/J}, \quad (2.5)$$

$$\beta_{ij} = 2\left(1 + \frac{j - J/2}{(J - 1) - J/2} \cdot \delta\right). \quad (2.6)$$

The value of  $\alpha_j$ 's in (2.5) gives  $F_{\mathbf{x}}(j) = j/J$  when  $\mathbf{x} = \mathbf{0}$ . The null hypothesis (2.3) is true when  $\delta = 0$  in (2.6), in which case the common slope is 2. The value of  $\delta$  represents the amount of deviation from the equal slope assumption since  $\beta_{i1}, \dots, \beta_{i,J-1}$  are within the width of  $2 \times \delta$  from the common slope 2. All covariates are independently generated from a uniform distribution over the interval  $(0, 1)$ . Then  $y$  can be generated from a multinomial distribution with  $P(Y = j|\mathbf{x}) = F_{\mathbf{x}}(j) - F_{\mathbf{x}}(j - 1)$ . We generate  $n$  set of values  $(y_i, x_{i1}, x_{i2}, \dots, x_{ip})$ , and obtain the  $p$ -value of the test for the proportional odds assumption from the SAS procedure PROC LOGISTIC.

We simulate each configuration 100 times, and calculate the mean of 100  $p$ -values to summarize 100 test results. PROC LOGISTIC reports  $p$ -value as .0001 (or  $< .0001$  in version 7.0 and higher of SAS) for any value less than .0001, so that we report up to three decimal points of the mean  $p$ -value. We also report the estimated probability of rejecting the null hypothesis under the significance level  $\alpha = .05$  in Table 1.

We consider that  $\delta = .05$  or  $\delta = .1$  gives practically equal slopes. This deviation from the null hypothesis would not matter in practice, so that we might use the proportional odds model for its simple interpretation. The test, however, detects such a negligible deviation

given a large sample, and does not differentiate it from the practically significant one such as  $\delta = 1.0$ . Table 1 shows that as sample size gets larger the test gives a highly significant  $p$ -value less than .010 even though the null hypothesis can be practically assumed to hold. This simulation study demonstrates that we cannot depend solely on the test result. The next section presents graphical methods to assess the practical significance of the assumption.

<< Table 1 around here >>

### 3. Graphical Methods to Assess the Proportional Odds Assumption

We need to check the adequacy of the equal slope assumption more carefully when the test gives a significant result. One natural method to assess the assumption (2.1) is to draw confidence intervals of  $\beta_{i1}, \beta_{i2}, \dots, \beta_{i,J-1}$  simultaneously on one graph. One such graph is illustrated in Figure 3 in the next section. Overlapping confidence intervals do not give strong evidence against the assumption of a common slope. Even for the nonoverlapping confidence intervals, we cannot exclude the possibility of a *practically* common slope since the widths of intervals become narrower as the sample size gets larger, and the confidence coefficient is for each interval, not for a family of intervals.

To tell whether the difference, if there is any, between  $\beta_{i1}, \dots, \beta_{i,J-1}$  is practically important, the probabilities  $P(Y = j|\mathbf{x}), j = 1, \dots, J - 1$  are useful. If estimates of these probabilities under (2.1) and (2.2) differ substantially, the violation of the null assumption would be considered to be practically significant. On the other hand, if there is little difference between the estimates of the probabilities, the possible difference between  $\beta_{i1}, \dots, \beta_{i,J-1}$  can be ignored for a practical purpose.

There is some difficulty in visualizing the probabilities, though. For each observation  $(y_i, \mathbf{x}_i)$  two vectors of probability estimates  $(\hat{P}(Y_i = 1|\mathbf{x}), \hat{P}(Y_i = 2|\mathbf{x}), \dots, \hat{P}(Y_i = J - 1|\mathbf{x}))$  are defined under the model (2.1) and (2.2) respectively, and it is hard to visualize  $n$  pairs of vectors. We circumvent this difficulty by plotting pairs of scalars  $\hat{P}(Y_i = y_i|\mathbf{x})$  under the

model (2.1) and (2.2), denoted by  $\hat{p}_{1i}$  and  $\hat{p}_{2i}$ , respectively. The  $n$  points  $\{(\hat{p}_{1i}, \hat{p}_{2i}); i = 1, \dots, n\}$  are expected to be near the line of slope 1 if the null assumption holds. Figures 1 and 2 illustrate how the plots change according to the degree of the violation of the null assumption. As (a) and (c) in Figures 1 and 2 show, the range of probabilities and the shape of the plots depend on the number of the levels of response, the number and the range of predictor variables, and the strength of the relationship between  $\mathbf{x}$  and  $y$ . Hence, in order to help judge the linearity of the plots, a reference plot is necessary.

The reference plot is proposed as follows:

1. Fit a proportional odds model (2.1) to a given set of data  $\{(y_i, \mathbf{x}_i); i = 1, \dots, n\}$ , then obtain the maximum likelihood estimates  $\hat{\alpha}_j$  ( $j = 1, \dots, J - 1$ ) and  $\hat{\boldsymbol{\beta}}$ .
2. Draw a new set of data  $\{(y_i^*, \mathbf{x}_i); i = 1, \dots, n\}$  based on the estimated proportional odds model in the previous step. In other words, we simulate  $y_i^*$  from a multinomial distribution with the probability belonging to category  $j$  equal to  $\hat{F}_{\mathbf{x}}(j) - \hat{F}_{\mathbf{x}}(j - 1) = [1 + \exp\{-(\hat{\alpha}_j - \mathbf{x}'\hat{\boldsymbol{\beta}})\}]^{-1} - [1 + \exp\{-(\hat{\alpha}_{j-1} - \mathbf{x}'\hat{\boldsymbol{\beta}})\}]^{-1}$  for  $j = 1, \dots, J - 1$ , and equal to  $1 - \hat{F}_{\mathbf{x}}(j)$  for  $j = J$ .
3. Then fit a proportional odds model to the new set of data, and obtain the estimated probability  $\hat{p}_{1i}^* = \hat{P}(Y_i^* = y_i^* | \mathbf{x})$ .
4. Fit model (2.2) to the new set of data, and obtain the estimated probability  $\hat{p}_{2i}^* = \hat{P}(Y_i^* = y_i^* | \mathbf{x})$ .
5. Plot  $\hat{p}_{2i}^*$  against  $\hat{p}_{1i}^*$ ,  $i = 1, \dots, n$ .

Whether or not the raw data set  $\{(y_i, \mathbf{x}_i); i = 1, \dots, n\}$  satisfies the proportional odds assumption, the points  $(\hat{p}_{1i}^*, \hat{p}_{2i}^*)$  in the plot are expected to be near the line of slope 1 since the new set of data  $\{(y_i^*, \mathbf{x}_i); i = 1, \dots, n\}$  satisfies the assumption. But its specific shape depends on the data set as shown in (b) and (d) in Figure 1 and 2.

We expect  $\delta = 0.05$  in (2.6) to give practically equal slopes. The linearity in the plot (a) in Figure 1 tells that the proportional odds model provides a decent fit for the data set generated from the model with  $\delta = 0.05$  and  $J = 3$ . It is confirmed by practically the same reference plot in (b). But, the null assumption of the proportional odds is rejected with the  $p$ -value less than .0001 because of the large sample size ( $n = 8,000$ ). Without these plots it would be hard to tell the practical significance from the statistical significance. On the other

hand, the plots (c) and (d) in Figure 1 shows the case when the slopes are different enough with  $\delta = 0.5$ . The striking difference between the two plots tells us that we cannot assume the null hypothesis in either a practical sense or a statistical sense. Figure 2 shows similar results for the case of  $J = 5$ . In the next section we present a real data example which has motivated this study.

<< Figures 1 and 2 around here >>

## 4. A Real Example

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 (denoted by  $x_1$ ), sex ( $x_2$ ), self-reported average high school grade ( $x_3$ ) and parent's education ( $x_4$ )) on the marijuana use in the past 12 months ( $y$ ). 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 measures the amount of time spent for informal socializing with friends. The other three explanatory variables reflect social differentiation. Except for the sex variable, the index of unstructured socializing with peers (scored 4 to 21), the average high school grade (1 to 9), and the parent's education (1 to 6) are treated as interval variables, i.e. the actual scores of the three ordinal variables are assumed to be equally spaced. The selection of scores of ordinal variables does not matter much because of the characteristic of the MARS (Multivariate Adaptive Regression Splines) method to be adopted for the analysis in the sequel, where local smoothing of each variable is made. The response variable of marijuana use has a scale ranging from 1 (for no use) through 10 (40 or more times in the last 30 days).

For the analysis of this data set, Du et al. (2002) used nonparametric hypotheses and test statistics, which were developed in Akritas et al. (2000). The main reason for applying

the nonparametric approach was that the proportional odds assumption in the logit model is significantly violated.

Kim (2000) used the MARS method for the analysis. The MARS model invented by Friedman (1991) is a flexible nonparametric regression model for high dimensional data with a continuous response  $Y^*$ . For the presumed system that generated the data

$$Y^* = f(x_1, \dots, x_p) + \epsilon,$$

MARS model gives as an approximation to  $f$

$$\hat{f}(\mathbf{x}) = \sum_{m=1}^M a_m B_m(\mathbf{x}), \quad (4.1)$$

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. Friedman (1993) extended the MARS methodology to the model with mixed (i.e. categorical and/or continuous) explanatory variables.

Kim (2000) used the MARS method to find an effective set of explanatory variables, i.e. a set of basis functions, and then applied the proportional odds model (2.1) with the new set of explanatory variables. The justification for applying the MARS method on  $Y$  comes from the characteristic of the proportional odds model, in which we can think of an underlying continuous response variable  $Y^*$  which is a continuous version of the observed ordinal response  $Y$  (Agresti 1990, p.323). 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 explanatory variables, not the same coefficients.)

The 8 basis functions found by the MARS model on  $Y$  are as follows;  $B_1(\mathbf{x}) = (x_1 - 16)_+$ ,  $B_2(\mathbf{x}) = (x_1 - 16)_-$ ,  $B_3(\mathbf{x}) = (x_1 - 19)_+$ ,  $B_4(\mathbf{x}) = (x_1 - 13)_+$ ,  $B_5(\mathbf{x}) = (x_3 - 4)_+$ ,  $B_6(\mathbf{x}) = I(x_2 = 1)(x_3 - 5)_+(x_4 - 1)_+$ ,  $B_7(\mathbf{x}) = I(x_2 = 1)(x_3 - 5)_-(x_4 - 1)_+$ ,  $B_8(\mathbf{x}) = I(x_2 = 1)(x_3 - 3)_+(x_4 - 1)_+$ , where  $(x - a)_+ = (x - a)I(x \geq a)$ ,  $(x - a)_- = -(x - a)I(x < a)$ , and  $I(A)$  is the indicator function of a set  $A$ . The equation (4.1) with these basis functions looks complicated, but the interpretation is quite simple using the ANOVA decomposition, where we collect together all basis functions that involve the same explanatory variables (Friedman

1991). For example, the combined term  $a_1B_1(\mathbf{x}) + \dots + a_4B_4(\mathbf{x})$  represents the nonlinear main effect of  $x_1$ , and the term  $a_5B_5(\mathbf{x}) + \dots + a_8B_8(\mathbf{x})$  reveals the three-variable interaction of  $x_2, x_3$  and  $x_4$ . Kim (2000) presents figures to ease the interpretation. It is interesting that only for male students is there an interaction effect between the school performance and parent's education. More specifically, the male students with low school performance tend to use more marijuana if their parents are more highly educated.

The SAS procedure PROC LOGISTIC for the proportional odds model

$$\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j - \beta_1B_1(\mathbf{x}) - \dots - \beta_8B_8(\mathbf{x}), \quad j = 1, \dots, J - 1 \quad (4.2)$$

produces the coefficients and the corresponding  $p$ -values in Table 2. We retain  $B_4(\mathbf{x}) = (x_1 - 13)_+$  term in spite of its nonsignificance since it does not complicate the interpretation and it is an important term for the MARS model on  $Y$ .

<< Table 2 around here >>

The violation of the proportional odds assumption of model (4.2) is still highly significant with the  $p$ -value less than .0001. Let us examine the practical significance of the test using the graphical methods proposed in the previous section.

To see how much the assumption is violated, we fit nine binary logistic regression models for  $\tilde{Y}_j = I(Y \leq j)$ ,  $j = 1, \dots, 9$ , i.e.

$$\text{logit}[P(\tilde{Y}_j = 1|\mathbf{x})] = \text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j - \beta_{1j}B_1(\mathbf{x}) - \dots - \beta_{8j}B_8(\mathbf{x}). \quad (4.3)$$

The proportional odds model assumes  $H_0 : \beta_{i1} = \beta_{i2} = \dots = \beta_{i9} (= \beta_i)$ , ( $i = 1, \dots, 8$ ) for our data set. Table 3 gives the estimates of coefficients for model (4.3). Note that there is no reverse sign of coefficients in each column. (There is one exception in  $\hat{\beta}_{4j}$ , which is acceptable since  $\beta_4$  is insignificantly different from 0 for model (4.2).)

<< Table 3 around here >>

To assess the amount of violation of the proportional odds assumption, we first draw the confidence intervals for  $\beta_{i1}, \beta_{i2}, \dots, \beta_{i9}$  ( $i = 1, \dots, 8$ ) as in Figure 3. The last three



plots show some evidence of violation, which are related to the interaction term between the school performance and the parent's education for male students. The discrepancy between the confidence intervals for  $\beta_{i1}, \beta_{i2}, \dots, \beta_{i9}$ , however, depends on the sample size  $n$  as well as the scale of the predictors. So the plots in Figure 3 are not very useful to assess the practical significance of the null assumption. For example, the confidence interval plot does not tell whether we can allow for a practical purpose the difference in  $\hat{\beta}_{6j}$  such as  $\hat{\beta}_{61} = -0.058$  and  $\hat{\beta}_{69} = -0.149$ .

As in Figure 4(a) the plot of two estimates of probability  $P(Y_i = y_i|\mathbf{x})$  under the model (4.2) and (4.3) is more helpful to judge the practical significance of the null assumption. Compared with the reference plot in Figure 4(b), the plot in Figure 4(a) indicates that there is a practically significant violation of the proportional odds assumption. Most points of discrepancy between these two plots were identified as with the cases  $y = 10$  as shown by Figure 4(c). We can conclude that except for the cases with  $y = 10$  the proportional odds assumption can be assumed for the practical purpose. (We have 244 cases of  $y = 10$  out of 8,777.) The estimates of  $\beta_{i9}$  ( $i = 1, \dots, 8$ ) for model (4.3) are quite different from the corresponding estimate of  $\beta_i$  for model (4.2), so that the two estimates of  $P(Y_i = 10|\mathbf{x}) = 1 - P(Y_i \leq 9|\mathbf{x})$  show a large discrepancy. But it should be mentioned that as in Table 3 the signs of coefficients are equal so that the qualitative interpretations such as the interaction effect between the school performance and the parent's education for male students are the same for two models (4.2) and (4.3).

<< Figures 3 and 4 around here >>

## 5. Summary and Conclusion

In common with other null hypothesis significance tests, the test of the proportional odds assumption has a limitation. The rejection of the null hypothesis of the proportional odds model is not very informative, especially for a large sample study. The simulation study in Kim (2001) and a real data example in this paper have demonstrated that we should

not depend solely on the significance test since a statistical significance does not necessarily mean a practical significance.

In spite of the subjectiveness of the judgment, the graphical methods proposed in this study can be a useful supplement assessing the practical significance.

**Acknowledgments.** The author wishes to thank Professors Osgood and Akritas at the Pennsylvania State University for providing the teenage deviant behavior data.

## References

- Agresti, A. (1990), *Categorical Data Analysis* (Wiley, New York).
- Akritas, M. G., Arnold, S. F. and Du, Y. (2000), Nonparametric models and methods for nonlinear analysis of covariance, *Biometrika* **87**, 507-526.
- Du, Y., Akritas, M. G., Arnold, S. F. and Osgood, D. W. (2002), Analysis of teenage deviant behavior data, *Sociological Methods and Research*. In press.
- Friedman, J. H. (1991), Multivariate adaptive regression splines, *Annals of Statistics* **19**, 1-141.
- 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.
- Kim, J. H. (2000), MARS modeling for ordinal categorical response data: a case study, *The Korean Communications in Statistics* **7**, 711-720.
- McCullagh, P. (1980), Regression models for ordinal data. *Journal of the Royal Statistical Society Series B* **42**, 109-142.
- 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.



Table 2: Estimates of Coefficients for Model (4.2)

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$
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

Table 3: Estimates of Coefficients for Model (4.3)

$j$	$\hat{\beta}_{1j}$	$\hat{\beta}_{2j}$	$\hat{\beta}_{3j}$	$\hat{\beta}_{4j}$	$\hat{\beta}_{5j}$	$\hat{\beta}_{6j}$	$\hat{\beta}_{7j}$	$\hat{\beta}_{8j}$
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