

11-1-2016

# On Generalizing Cumulative Ordered Regression Models

Robert W. Walker

*Atkinson Graduate School of Management, Willamette University, [rwalker@willamette.edu](mailto:rwalker@willamette.edu)*



Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

## Recommended Citation

Walker, Robert W. (2016) "On Generalizing Cumulative Ordered Regression Models," *Journal of Modern Applied Statistical Methods*: Vol. 15 : Iss. 2 , Article 28.  
DOI: [10.22237/jmasm/1478003160](https://doi.org/10.22237/jmasm/1478003160)

---

# On Generalizing Cumulative Ordered Regression Models

## **Cover Page Footnote**

The author thanks Alan Agresti, William H. Greene, Andrew Martin and Peter McCullagh, Christopher Winship, and participants in the Statistics Seminar of the Department of Mathematics at Washington University for commentary and suggestions. An equivalent and simpler geometric presentation can be constructed using equiprobability contours, as Peter McCullagh suggested, though the linkages to the coherence/completeness problem are less obvious (though they can also be represented geometrically). This approach seems more direct.

# On Generalizing Cumulative Ordered Regression Models

**Robert W. Walker**  
Willamette University  
Salem, OR

---

We examine models that relax proportionality in cumulative ordered regression models. Something fundamental arising from ordered variables and stochastic ordering implies a partitioning. Efforts to relax proportionality also relax the ability to collapse an inherently multidimensional problem to a partitioning of the (unidimensional) real line. It is surprising and unfortunate to find that deviations from proportionality are sufficient to generate internal contradictions; undecidable propositions must exist by relaxing proportional odds without other relevant and significant changes in the underlying model. We prove a single theorem linking continuous support and partitions of a latent space to show that for these two characteristics to be simultaneously satisfied, the model must be the proportional-odds model. Conditioning on the adjacency that is closely related to the partitioning is fruitful, but at this point we join the class of continuation-ratio models. Alternatively, Anderson's (1984) stereotype model is quite general and nests ordered and unordered choice models, but again we have left the domain of cumulative models. Adopting multidimensional cumulative models or imposing covariate-specific thresholds are the only certain methods for avoiding these troubles in the cumulative framework. It is generically impossible to generalize the cumulative class of ordered regression models in ways consistent with the spirit of generalized cumulative regression models. Monte Carlo studies also demonstrate the general principles.

**Keywords:** Proportional odds models, partial proportional odds models, Monte Carlo simulation

---

## Introduction

Generalizations of common cumulative models for ordered phenomena are considered. The parallelism inherent to cumulative models such as the (proportional odds) ordered logit/probit model (McKelvey & Zavoina, 1975; McCullagh, 1980) is seen as limiting and workers in numerous statistical literatures have worked on generalizations.<sup>1</sup> According to Google Scholar, 407 papers cited

---

*Dr. Walker is an Associate Professor of Quantitative Methods in the Atkinson Graduate School of Management. Email him at: [rwalker@willamette.edu](mailto:rwalker@willamette.edu).*

Peterson and Harrell's (1990) presentation of a partial proportional-odds model and 478 cited the work by Brant (1990) on testing proportional odds as of March 6, 2015; the generalization appears in McCullagh and Nelder (1989) and Agresti (2002). Test statistics have been proposed for testing this specification as null hypothesis against a more general specification (Brant, 1990) and software for Stata (Williams, 2006) and Yee's (2010) library for R (R Development Core Team, 2009) allow these models to be estimated.

Following Peterson and Harrell (1990), Cox (1995), and numerous others in medical statistics, social science has used these models for ordered scales related to social policy and racial attitudes (Branton & Jones, 2005); Fullerton (2009) presents a sociological analysis of income quartiles; Gannon (2009) examines self-reported disability status using this generalization. A spate of articles in the Journal of Modern Applied Statistical Methods (e.g., O'Connell & Liu, 2011; Liu & Koirala, 2012) developed diagnostics for the model and deployed the model for educational outcomes. The model received some attention in the field of health economics (Lindeboom & van Doorslaer, 2004). Generalized threshold models (Maddala, 1983; Terza, 1985) are similar and the general argument applies to the class of location-scale models. To our surprise, there is no obvious way to generalize the model while retaining two basic assumptions that motivate cumulative models. It is cumulative in both probabilistic foundations and in name.

Given the widespread attention to generalizations of the cumulative model, it is surprising to find that the model only sensibly exists in the presence of proportional odds or the parallel equivalent. Partitioning a unidimensional latent space uniquely with functions of covariates is constrained by the requirement that everything match at the boundaries of any two adjacent partitions with cumulative probabilities. Defining a model and finding conditions where optima exist can be quite different from a model with sensible statistical microfoundations; these two ideas diverge when generalizing the cumulative ordered regression model. When models are employed for describing and estimating physical, social, or biological processes, internal contradictions pose significant difficulties because it is not clear how we return the parameters to their substantive context in a way that is consistent with the assumptions that facilitate estimation.

## The Argument

Anderson (1984) distinguished ordered variables that are grouped continuous – ordered groupings of an unobserved continuous outcome – from assessments – judgments or grades somehow combining (possibly) multiple inputs. Ordered

responses can represent a coarsened latent variable, as in income quartiles, Likert scales (in many cases), feeling thermometers, and grades given to things ranging from diamonds to dairy products. Although the continuous variable cannot be observed, the groupings – partitions – may be observed as an ordered outcome. The ordered outcome can be inverted to a partition of the latent scale and differences on this scale are often of substantive interest; many seemingly ordered phenomena do not obviously present for finer measurement but the latent measure remains substantively interesting. Nevertheless, an ordered variable can be an assessed or judged outcome.

A variable of the second type is generated by an *assessor* who possesses an indeterminate amount of information before producing his *judgment* of the grade of the ordered variable. For example, Anderson and Philips (1981) refer to the “extent of pain relief after treatment”: *worse, same, slight improvement, marked improvement or complete relief*. In principle, there is a single, unobservable, continuous variable related to this ordered scale, but in practice, the doctor making the assessment will use several pieces of information in making his judgment on the *observed* category. For example, he might use severity of pain, kind of pain, consistency in time and degree of disability. We will refer to these variables of the second type as “assessed” ordered categorical variables and argue that, in general, a different approach to modelling regression relationships is appropriate for the two types. (Anderson, 1984, p. 2)

Anderson’s argument suggested that the presence of multiple inputs requires a model that need not assume an underlying order but instead allows order to emerge (or not) as a special case of a more general model. The arguments underlying the stereotype model of Anderson (1984) are precisely focused on dimensionality (the number of latent dimensions), ordering (and whether or not it obtains), and distinguishability (do covariates distinguish categories?) with a model that can assess each in a null hypothetical framework. The model derives from category probabilities rather than a cumulative scale. Though, the outcome variable itself,  $y$ , can also be represented by sets of ratios of cumulative probabilities with some assumed distribution, the statistical principles that are engaged require a sensible probability model and the cumulative framework becomes quite limiting. At some point, the cumulative approach requires a well-defined cumulative distribution; this is deeply constraining and leads to an internal contradiction in “generalizations” of cumulative ordered regression models.

Consider a  $J > 2$  category ordinal variable  $y_i$  observed on units  $i \in N$ . The canonical distribution for such outcomes is the multinomial. Frequently, analysts employ the notion of a latent variable,  $y_i^*$ , crossing thresholds defined by the  $J + 1$  vector  $\boldsymbol{\tau}$ , with  $\tau_0 = -\infty$  and  $\tau_J = \infty$ . A key component to cumulative models is a partition linking an observation rule and a latent unidimensional continuous variable. We define such a rule as Assumption 1.

**Assumption 1:** *Mutually exclusive and exhaustive partition:*

$$y_i = j \text{ if and only if } \tau_j > y_i^* \geq \tau_{j-1} \quad (1)$$

for

$$j = 1, 2, \dots, J, \quad i \in N, \quad \sum_{j \in J} \text{pr}(y = j) = 1$$

Equivalently,  $\tau$  could be viewed as a function so that  $\tau: y^* \rightarrow y$ . It is many-to-one, but it is special because it is an ordered partition. The ordering can be inverted to imply a unique set of inequality relations that must apply to  $y^*$ . Though we have yet to define  $y^*$ , we will place some structure on randomness.

**Assumption 2:**  $\epsilon_i$  are independent and identically distributed with probability density function  $f$  and cumulative distribution  $F$  such that  $\text{supp}(\epsilon) = \mathbb{R}$  and  $\mu(\epsilon) = 1$  (measure one).

The errors will ultimately give a distribution to the random variable of interest; the random variable will inherit the distribution of  $\epsilon$  conditional on a true model consisting of covariates. Of a driving force in the statistical logic, the latent variable, as a function of covariates  $X$ , requires structure. This leads to Definition 1.

**Definition 1:** *Proportional Odds: Linearity in latent variables.*

$$y_i^* = \mathbf{X}_i \boldsymbol{\beta} + \epsilon_i \quad (2)$$

where  $\mathbf{X}_i$  is a (row) vector of centered covariates for  $i$ ,  $\boldsymbol{\beta}$  is a (column) vector of parameters of interest with  $(\mathbf{X}^T \mathbf{X}) - 1$  existing.<sup>2</sup> This is the source of our notion of “parallelism”. The latent variable is a linear function of covariates and parameters

and this yields a set of parallel planes. When the covariates are bounded, the randomness in the latent variable is then inherited from  $\epsilon$ .<sup>3</sup> We will assume independence between the covariates and random errors.

The argument will apply equally to commonly used F with continuous support (normal, logistic, cloglog (Gumbel), Cauchy) on  $\mathbb{R}$ . Substituting (2) into (1) yields (for all  $j \in J$ ), for a general cumulative distribution function F determined by the assumed distribution of  $\epsilon_i$ ,

$$\mathcal{L}_F = \sum_{i=1}^N \sum_{j=1}^J \text{Ind}[y_i = j] \times \ln \left[ F(\tau_j - y_j^*) - F(\tau_{j-1} - y_i^*) \right] \quad (3)$$

One appealing feature of this proportional odds model is the ease of use. It is intuitively pleasing to link ordered categories to some underlying continuum that is determined by covariates and to allow the marginal effects to be well defined across all outcomes. In effect, we have a linear regression for ordered outcomes that does not impose an (likely fallacious) interval-scale interpretation. It is also intuitively restrictive.

The “partial-proportional odds” model (Brant, 1990; McCullagh & Nelder, 1989; Peterson & Harrell, 1990) employs subscripted  $\beta$  with the implicit idea that some (or all) regressors have varying impacts that depend on the comparison, as in Definition 2.

**Definition 2:** *Partial/Non Proportional Odds: Conditionally (on j) linear in the latent variable.*

$$y_i^* = \mathbf{X}_i \beta_j + \epsilon_i \quad (4)$$

The proportional-odds assumption is that  $F = \Lambda$  or that  $\epsilon_i$  are i.i.d. logistic and that  $\beta_1 = \beta_2 = \dots = \beta_{J-1}$ ; this model can be estimated by maximizing (3) with (2) under general conditions owing to properties illuminated by Pratt (1981) or by employing generalized linear models with conditional mean function as implied by (2). Insert (4) into (3) to yield

$$\mathcal{L}_F = \sum_{i=1}^N \sum_{j=1}^J \text{Ind}[y_i = j] \times \ln \left[ F(\tau_j - \mathbf{X}_i \beta_j) - F(\tau_{j-1} - \mathbf{X}_i \beta_{j-1}) \right] \quad (5)$$

It is straightforward to identify the parameters as deviations ( $\beta_j = \beta_1 + \delta_j$ ) from a base parameter with a simple view toward whether or not proportional odds obtains in Wald statistics, score tests, or approximating the likelihood ratio.<sup>4</sup> At first glance, (5) is a very useful generalization because the underlying linear structure of the proportional odds type model seems excessively limiting even with creative functional forms for the covariates. Generalizing the model maintains the significant intuitive appeal of the cumulative model for parameters that it is sensible to believe map linearly onto the latent scale with the flexibility of altering relationships in a way that uniquely leverages the adjacency of ordered data. There is obvious gain to the exercise that is quite appealing by retaining the simplicity of unidimensionality; at what cost? The unidimensional cumulative foundation, if the model is an appropriate partition of the latent space, requires that this hold for *each*  $i \in N$ .

In the literature on partial proportional odds models, much has been made of conditions for sensible estimates. Estimating the model is distinct from requiring predicted responses to be nonnegative. Conditions must hold on  $\beta_j$  and  $\tau$  for estimates to exist (the parameters and thresholds are jointly bounded) and these conditions are weaker than those required for nonnegative category probabilities.<sup>5</sup> The latter is the usual criterion for assessing the model. Unfortunately, the set of models we can estimate is itself a proper subset of models that contradict their own probabilistic foundations. Put simply, models may be estimated with nonnegative probabilities for each ordered category that have no well-behaved latent variable satisfying Assumptions 1 and 2 with Definition 2; this is the central demonstration of Theorem 1. Research has remained focused on testing (Brant, 1990), estimating (Peterson & Harrell, 1990), and generalizing (Maddala, 1983; Terza, 1985; Cox, 1995; Williams, 2006) ordered regression models using Definition 2 for which no such generalization exists.

**Theorem 1:** Assumption 1 and Assumption 2 avoid internal contradiction if and only if observational equivalence holds between Definition 1 and Definition 2 ((2) and (4)).

**Proof:**

1. Suppose that Assumption 2 holds and (2)  $\neq$  (4). (4) generically requires  $\exists j: \beta_j \neq \beta_{j-1}$  and, perhaps more importantly that  $y_j^* \neq y_{j-1}^*$ . Assumption 2, recalling parameters to scale, allows us to write,



$\beta_j = \beta_j/\sigma$  which then scales  $y^* = y^*/\sigma$ . With the assumption that  $\mathbf{X}$  is centered, there must be an  $x$  to which one of the following two conditions apply because no column of  $x$  can fail to have support on both sides of the center without contradicting full rank.

**Case (a):** Suppose  $\beta_j > \beta_{j-1}$  in (4).  $\sigma$  is fixed under Assumption 2.<sup>6</sup> Consider  $y_j^* = \tau_j$  which is justified by the support of  $\epsilon$  on  $\mathbb{R}$ . Under the supposition,  $y_j^* = \tau_j > y_{j-1}^*$  (because  $\beta_j > \beta_{j-1} \Rightarrow \beta_j x > \beta_{j-1} x$ ) in (4).  $y$  is undefined; invoking Assumption 1 yields  $y_j^* = j+1$  while  $y_{j-1}^* = j$ .

**Case (b):** Suppose  $\beta_{j-1} > \beta_j$ . Consider  $y_{j-1}^* = \tau_j$  which is justified by the support of  $\epsilon$  on  $\mathbb{R}$ . Under the supposition,  $y_{j-1}^* = \tau_j > y_j^*$  (because  $\beta_{j-1} > \beta_j \Rightarrow \beta_{j-1} x > \beta_j x$ ) in (4).  $y$  is undefined; invoking Assumption 1 yields  $y_{j-1}^* = j+1$  while  $y_j^* = j$ .

2. Suppose that Assumption 1 holds and (2)  $\neq$  (4). Assumption 1 allows us to write the probability that  $y = j$  sums to one, the logic will follow the above. That all of the observations sum to one will contradict continuous support. To show this, generically write

$$\Pr(y = j) = \int_{\{\tau_{\{j-1\}} - y_{\{j\}}^*\}}^{\{\tau_{\{j\}} - y_{\{j\}}^*\}} f(\epsilon) d\epsilon \quad \forall j \in J \quad (6)$$

$$\sum_{j \in J} \Pr(y_i = j) = \int_{-\infty - y_j^*}^{\tau_1 - y_i^*} f(\epsilon) d\epsilon + \int_{\tau_1 - y_i^*}^{\tau_2 - y_i^*} f(\epsilon) d\epsilon + \dots + \int_{\tau_{J-1} - y_i^*}^{\infty - y_i^*} f(\epsilon) d\epsilon \quad (7)$$

$$\int_{-\infty}^{\infty} f(\epsilon) d\epsilon = 1 \quad (8)$$

Under the proportional odds model, all is fine and Assumption 2 is satisfied. We have integrated the real line satisfying the restrictions on  $f$  and  $F$ . Now let us examine (4).

$$\sum_{j \in J} \Pr(y_i = j) = \int_{\tau_0 - y_{i0}^*}^{\tau_1 - y_{i1}^*} f(\epsilon) d\epsilon + \int_{\tau_1 - y_{i2}^*}^{\tau_2 - y_{i2}^*} f(\epsilon) d\epsilon + \dots + \int_{\tau_{J-1} - y_{i,J-1}^*}^{\tau_J - y_{i,J-1}^*} f(\epsilon) d\epsilon$$

The boundaries of the integrals were moved by assuming (4) but leaving  $\tau$  fixed. To view this more cleanly, expand the integral about the fact that  $y_j^* \neq y_{j-1}^*$ . The two cases from before will appear parenthetically.

$$\begin{aligned} \sum_{j \in J} \Pr(y_i = j) &= \int_{-\infty - y_{i1}^*}^{\tau_1 - y_{i1}^*} f(\epsilon) d\epsilon \left( \overbrace{\int_{\tau_1 - y_{i1}^*}^{\tau_2 - y_{i2}^*} f(\epsilon) d\epsilon}^{y_{i2}^* > y_{i1}^*} \right) \left[ \overbrace{-\int_{\tau_1 - y_{i2}^*}^{\tau_1 - y_{i1}^*} f(\epsilon) d\epsilon}^{y_{i2}^* < y_{i1}^*} \right] \\ &\quad + \int_{\tau_1 - y_{i2}^*}^{\tau_2 - y_{i2}^*} f(\epsilon) d\epsilon \left( \overbrace{\int_{\tau_2 - y_{i2}^*}^{\tau_2 - y_{i3}^*} f(\epsilon) d\epsilon}^{y_{i3}^* > y_{i2}^*} \right) \left[ \overbrace{-\int_{\tau_2 - y_{i3}^*}^{\tau_2 - y_{i2}^*} f(\epsilon) d\epsilon}^{y_{i3}^* < y_{i2}^*} \right] \quad (9) \\ &\quad + \dots + \int_{\tau_{J-1} - y_{i,J-1}^*}^{\infty - y_{i,J-1}^*} f(\epsilon) d\epsilon \\ &\neq 1 \end{aligned}$$

Generically, the parenthetical (or bracketed) elements will be non-zero unless (2) = (4). Moreover, these regions, given support on  $\mathbb{R}$ , are not countable and the probability that the two or multiple terms offset is a set of measure zero. Up to this set of measure zero,  $\sum_{j \in J} \Pr(y_j = j) \neq 1$  contradicting Assumption 2.

## Discussion

The proportional-odds/parallel model is the  $J - 1$  dimensional solution that uniquely collapses to a marginal distribution. Efforts to make the model more realistic, such as the structure defined in Definition 2, ultimately make it less realistic in the sense that its properties cannot be studied under its assumptions. The reason is that the assumptions are internally contradictory when combined with Definition 2. The models become internally contradictory of their own probability formulations when they deviate from the proportional odds model. The underlying latent variable is a strict order under the proportional-odds assumption and deviations can violate this ordering. These deviations from this underlying ordering wreak havoc on the probabilistic foundations.

We will replicate the Monte Carlo simulation evidence in Peterson and Harrell (1990, Table 6, Design 4) to provide a context. When the odds-ratio formulation common is considered to the ordered logistic regression model, write

$$\frac{\Pr(y_i \leq j | \mathbf{X}_i)}{\Pr(y_i > j | \mathbf{X}_i)} = \exp(-\mathbf{X}_i \boldsymbol{\beta}_j), \quad j = 1, 2, \dots, J-1 \quad (10)$$

Consider the experiment reported as Design 4 in Peterson and Harrell (1990, p. 216) that defines a four-category  $y$ .  $\mathbf{X}$  is a set of five completely crossed binary predictors ( $2^5$ ) of ten observations each ( $N = 320$ ). Peterson and Harrell (1990) set  $\boldsymbol{\beta} = 0.5$  with the exception of  $\beta_{25} = 1$ .  $\alpha$  are constants (or inverse cutpoints such that  $\tau$  in (1)) are set to  $\alpha = \{0.405, -0.847, -2.2\}$ . The key to their strategy is in independent multinomial sampling.

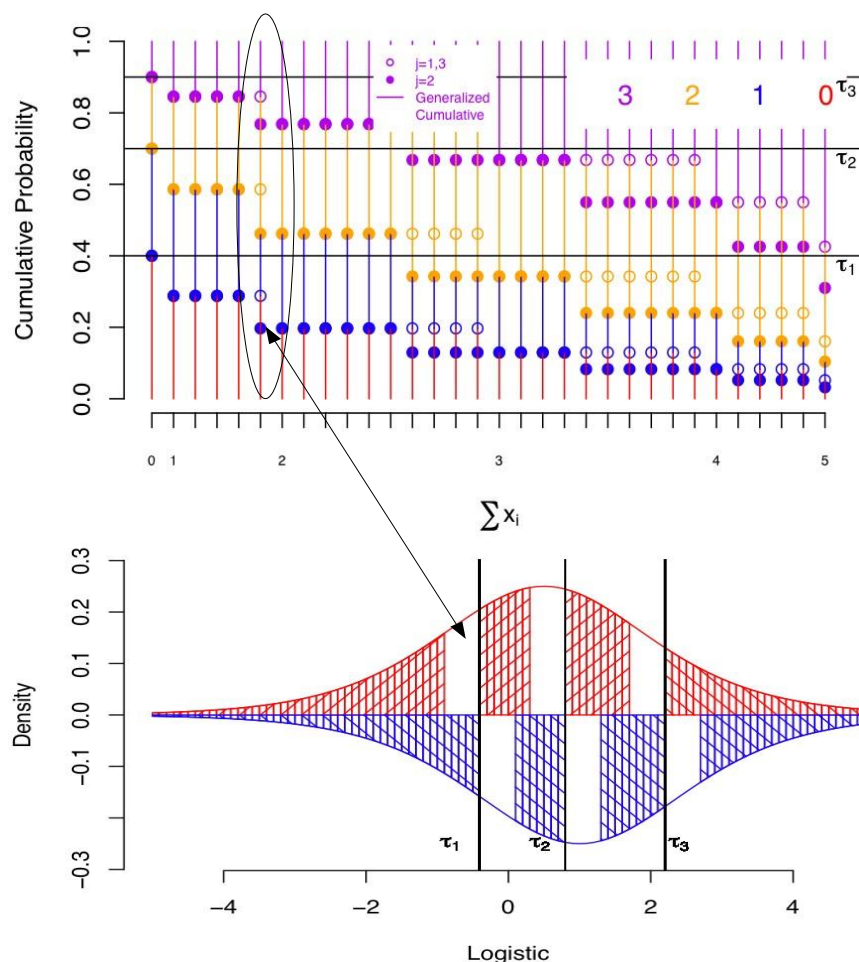
Begin with ratios of categories specified along some cumulative scale but curiously no appeal to a random variable.  $y$  will ultimately result from creating cumulative probabilities and comparing them with model estimates. Peterson and Harrell (1990, p. 208) define

$$C_{ij} = \Pr(y \geq j | \mathbf{X}_i) = \frac{1}{1 + \exp(-\alpha_j - \mathbf{X}_i \boldsymbol{\beta}_j)} \quad (11)$$

Because this defines the cumulative distribution function of a logistically distributed random variable, work backward to examine the distribution(s?) of this logistic random variable. As in Peterson and Harrell (1990, Design 4), suppose  $x_1 = x_2 = x_3 = x_4 = 0$ ,  $x_5 = 1$ ; this implies  $\mathbf{X}_i \boldsymbol{\beta} = \{0.5, 1, 0.5\}$ . Figure 1 illustrates a part of the difficulty.

The top panel of Figure 1 provides the cumulative probability plots obtained from all thirty-two possible combinations of our five binary predictors arranged along the  $x$ -axis ordered lexicographically first by  $\sum_i x_i$  and then by  $i$ . The  $y$ -axis presents the cumulative probabilities (3-purple, 2-orange, 1-blue, 0-red) according the partial proportional-odds model. As the  $x$ -axis increases, the probability of higher categories increases. The unfilled circles represent predictions from the first and third equations (which happen to be equal) and the filled circles represent predictions from the second equation. The unfilled circles define the cumulative probabilities for the lowest (blue) and highest (purple) categories while the (orange) filled circles define outcomes in the interior categories. As expected from the

parameters given before, recalling that these parameters have log odds-ratio interpretations, the highest category becomes quite common.



**Figure 1.** Peterson and Harrell (1990): Table 6, Design 4: The top panel plots cumulative probabilities derived from the partial proportional-odds cumulative logit color coded as in the legend; Filled circles represent probabilities derived from  $j = 2$ ; Open circles represent probabilities derived from  $j = \{1, 3\}$ ; The solid lines capture the cumulative probabilities as they enter the “partial proportional-odds” likelihood; The bottom panel displays logistic densities for  $\mathbf{X}\boldsymbol{\beta}_j = 0.5$  above zero and for  $\mathbf{X}\boldsymbol{\beta}_j = 1$  below zero; The non-hatched areas represent areas such that partitioning fails

That the probability of category two is shrinking is a product of the nonproportional-odds and the oft-noted issue of negative probabilities is a

necessary consequence of this shrinkage. That they do not cross in the observed data is taken as a signal of the underlying validity of the model (and estimates) when the very fact that such lines can cross with a cumulative probability model contradict the foundations of cumulative probabilities.

The scale has been assumed fixed by implication,  $y \sim \lambda(\mathbf{X}_i\boldsymbol{\beta}_j, \sigma = 1)$ , where  $\lambda$  is a logistic distribution characterized by location  $\mathbf{X}_i\boldsymbol{\beta}_j$  and scale equal to one.<sup>7</sup> The density depends on the outcome because  $j$  enters the conditional expectation. The way to resolve this is to set (2) = (4). If it is assumed that the cutpoints between categories for this logistic random variable to be constants and that the model is true, then a logistically distributed random variable with continuous support on  $\mathbb{R}$  arises and Theorem 1 applies. This only works when (2) = (4).

Peterson and Harrell (1990) instead use the sequence of logit cumulative odds. Define a multinomial random variable using  $C_{ij}$  as a partition of the unit interval (0, 1) for input probabilities. This is the equivalent of invoking Assumption 1. The problem is that drawing cumulative probabilities in a uniform fashion over the unit interval and inverting them to the logit-scale, given that the logit is a one-to-one transformation, implies continuous support on  $\mathbb{R}$ . Under Theorem 1, this cannot be valid unless it is done under the proportional variant. On a superficial level, the approach resolves an inconsistency such that simulation succeeds with probability one.

In the process, avoid defining a random variable excepting  $y$  and take a cumulative probability over an undefined logistic random variable. Invoke a logistic random variable to estimate  $C_{ij}$  alongside  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}_j$ . Order only enters to the extent that the multinomial distribution is drawn as a partition of the logistic distribution. But here is where the problem emerges. A uniform random variable, call it  $u$ , gives the hypothetical cumulative probability. Taking  $C_{ij}$  as given can generate  $y$  according to which interval  $u$  happens to fall into for each  $i$ . Theorem 1 dictates a generic problem with this strategy; either the logistic distribution does not have continuous support or it does not generically integrate to one. In either case, sidestepping the specification of the random variable also allows us to sidestep the uncomfortable realization that the random variable we invoke does not and cannot have the properties that we have assumed. This is illustrated in the bottom panel of Figure 1.

The bottom panel of Figure 1 displays an example of the implied logistic densities from Peterson and Harrell (1990), Design 4. The hatched areas represent portions of the density that satisfy partitioning while the blank areas showcase the area of partitioning failure. How does partitioning fail? In general, if one equation

produces some outcome  $\tilde{y}$  and the other equation produces  $\hat{y}$  and  $\tilde{y} \neq \hat{y}$ , we can say the partition failed because the resulting value of  $y$  is not unique (or does not exist).

The problem is very similar to the issues of completeness and coherence in the econometric study of simultaneous equations with limited outcomes (Heckman, 1978; Gourieroux, Laffont, & Monfort, 1980; Dagenais, 1999; Tamer, 2003; Lewbel, 2007).<sup>8</sup> Depending on the sign of differences in adjacent regression coefficients, the offending regions are characterized in (9). This is of consequence because the size of the regions in which the model is internally contradictory is increasing in the differences (the size of deviations from parallelism) and the boundaries depend on  $\mathbf{X}_i$ . Deviations from proportional odds (or an equivalent parallelism of planes in  $\mathbb{R}^p$  where  $p$  is the column rank of  $\mathbf{X}$ ) are sufficient to break the most basic of assumptions about sources of randomness and notions of order, both quite sensible. Larger deviations from parallelism increase the measure of contradictions.

The generic fact that  $y^*$  becomes multidimensional under deviations from parallelism or proportionality causes the problem. Under deviations from proportional odds with a well behaved cumulative distribution, no such constraint exists that is not a jointly identified function of thresholds for each unique  $x$  and the parameters of interest  $\beta_j$ . Thought of as a function,  $\tau$  must match at the boundaries of the cumulative distribution for  $\mathbf{y}^*$  (and  $\epsilon$ ) to have continuous support. For this to work,  $\tau_j$  cannot be invariant to  $i$  unless  $\mathbf{X}$  is also invariant to  $i$ . Of course, if  $\mathbf{X}$  is globally invariant to  $i$ ,  $\mathbf{X}$  is a constant. When  $\mathbf{X}$  is a variable, the trouble reemerges. Before presenting a Monte Carlo simulation, two related issues are mentioned.

There are a variety of ways to test parallelism of the regression slopes. The previous demonstration begs the question of what alternatives such tests embody. If the model does not exist except at the null hypothesis, a rejection of the null seems entirely uninformative because it offers no insights into the nature of the problem. These tests do not obviously lead to some more general class of models in which parallelism is a restriction. With this in mind, turn to an analysis of solutions to the more general problem in models that are not internally contradictory.

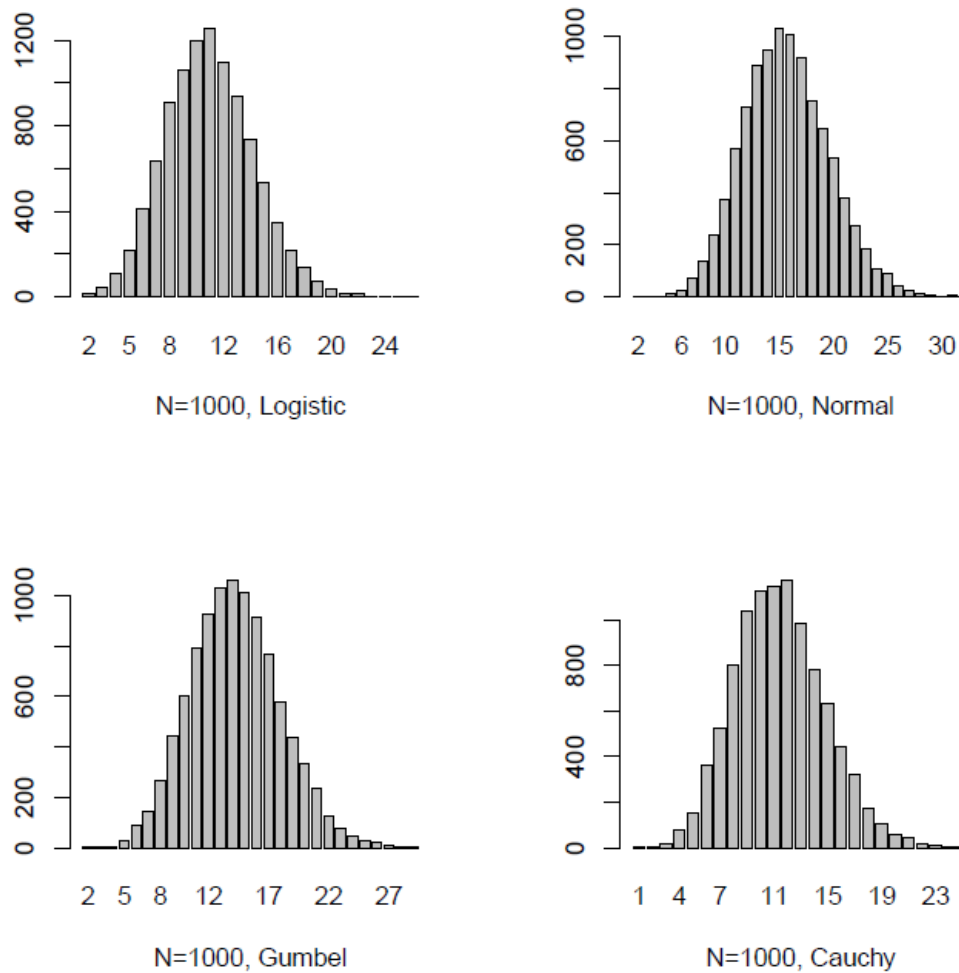
In the nonrecursive simultaneous equations setting, Dagenais (1999) restricts the support of  $\epsilon$ . Although this is technically correct and logically sound, it seems hard to intuitively justify for most substantive applications and impossible to verify in practice. Another obvious solution emerges from the ideas of Maddala (1983) and Terza (1985). The generalized threshold model parameterizes the thresholds

$(\tau_j)$ , instead of allowing regression coefficients to vary on the basis of the comparisons  $(\beta_j \neq \beta_{j-1})$ . To be precise, write

$$\tau_{ij} = \mathbf{X}_i \boldsymbol{\psi}_j \quad (12)$$

and expand the vector  $\mathbf{X}_i$  to include a constant. As it happens, the model is an isomorphism to the previous case (4) and all the same results apply. Suppose instead, construct a model based on the varying thresholds where the variation in the thresholds is specific to the row rank of  $\mathbf{X}$ . In other words, maintain the aforementioned parallelism, but allow the thresholds to be specific to observed covariates. Of course, with continuous covariates, this is not at all helpful, but with discrete regressors and large samples, such a model can be estimated and all of the relevant thresholds can also be estimated so long as each  $J$  is observed for each unique row of  $\mathbf{X}$ . When there are no such observations, the problems of Chamberlain's (1980) fixed effects estimator when outcomes do not vary arises.

A brief R ([R Development Core Team, 2009](#)) simulation example showcases the severity of the problem (Appendix A provides a logistic example).<sup>9</sup> The simulation is constructed with a single binary regressor and a uniform regressor on  $[-1, 1]$ . Set  $\boldsymbol{\tau} = \{-0.5, 0.5\}$ ,  $\beta_{11} = 0.05$ ,  $\beta_{12} = 0.1$ ,  $\beta_2 = 1$  for 1000 observations and perturbed the latent variable with standard logistic, normal, Gumbel (cloglog), and Cauchy errors before applying Assumption 1 to yield results. It is important to note that with 1000 observations and relatively small effect sizes (as these are), under the proportional odds logistic regression, roughly 6.25% of 10,000 iterations reject the hypothesis that  $\beta = 0$  when  $\beta = 0.10$ . The effects are so small there is almost no power. Even under these minute deviations, answers fail to exist. Turning to the evidence reported in [Figure 2](#), the number of failures in invoking partitioning is bounded below by zero and bounded above by just over 3%. The graphic makes clear that a non-zero fraction of outcomes are undefined (in all 10000 Monte Carlo trials) as reported in [Figure 2](#). Given Theorem 1, it comes as no surprise that all are susceptible; the model contains an internal contradiction unless it is the parallel version.



**Figure 2:** Undefined outcomes from partial-proportional odds: 10,000 Monte Carlo trials

## Conclusion

Modeling multidimensionality is a useful endeavor and it is not prohibitive. The difficulties are in conceptualizing the substance of such dimensions in applications and linking them together to obtain a stochastic order. It is important that our efforts remain true to the underlying probability structures that generate the data. Models that cannot be inverted cannot be studied in any meaningful way. To the extent that models are meant to capture the processes that generated them, generalizing



nonparallelism in a cumulative framework under Assumptions 1 and 2 is impossible and the probability of contradiction goes to one as sample sizes become infinite. The parallel version of the model is exceptionally useful for many problems, but generalizations of the model must carefully handle the restrictions imposed by their cumulative foundations.

Although the main demonstration is a negative one, hope is not lost. A well-studied and widely known class of ordered regression models can accommodate non-parallel effects and retain some cumulative foundations. Form odds-ratios for the sequential (Fahrmeir & Tutz, 1994) or continuation-ratio logit (Agresti, 2002) as

$$\frac{\Pr(y_i = j | \mathbf{X}_i, y_i \geq j)}{\Pr(y_i > j | \mathbf{X}_i, y_i \geq j)} = \exp(-\mathbf{X}_i \boldsymbol{\beta}_j), \quad j = 1, 2, \dots, J-1 \quad (13)$$

The solution has two parts. First, condition on the observed data and this resolves the incompleteness of the generalized cumulative regression model. Second, the models are mixtures of category and cumulative probabilities and, more importantly, are inherently multidimensional in the non-parallel case. However, such models conform to the more basic intuition that each unique linear function captured by  $\boldsymbol{\beta}_j$  must yield a unique dimension. The generalized cumulative regression model is a multidimensional model collapsed to a single dimension. The collapsing works if and only if the model is the proportional odds model.

## Acknowledgements

The author thanks Alan Agresti, William H. Greene, Andrew Martin and Peter McCullagh, Christopher Winship, and participants in the Statistics Seminar of the Department of Mathematics at Washington University for commentary and suggestions. An equivalent and simpler geometric presentation can be constructed using equiprobability contours, as Peter McCullagh suggested, though the linkages to the coherence/completeness problem are less obvious (though they can also be represented geometrically). This approach seems more direct.

## References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York, NY: John Wiley and Sons.

- Anderson, J. A. (1984). Regression and ordered categorical variables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 46(1), 1-30. Retrieved from <http://www.jstor.org/stable/2345457>
- Anderson, J. A., & Philips, P. R. (1981). Regression, discrimination, and measurement models for ordered categorical variables. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 30(1), 22-31. doi: 10.2307/2346654
- Boes, S., & Winkelmann, R. (2006). Ordered response models. *Allgemeines Statistisches Archiv*, 90(1), 167-181. doi: 10.1007/s10182-006-0228-y
- Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, 46(4), 1171-1178. doi: 10.2307/2532457
- Branton, R. P., & Jones, B. S. (2005). Reexamining racial attitudes: The conditional relationship between diversity and socioeconomic environment. *American Journal of Political Science*, 49(2), 359-372. doi: 10.1111/j.0092-5853.2005.00128.x
- Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The Review of Economic Studies*, 47(1), 225-238. doi: 10.2307/2297110
- Cox, C. (1995). Location-scale cumulative odds models for ordinal data: A generalized non-linear model approach. *Statistics in Medicine*, 14(11), 1191-1203. doi: 10.1002/sim.4780141105
- Dagenais, M. (1999). A simultaneous probit model. *Cahiers Économiques de Bruxelles*, 163(3), 325-346.
- Fahrmeir, L., & Tutz, G. (1994). *Multivariate statistical modelling based on generalized linear models*. New York, NY: Springer.
- Fullerton, A. S. (2009). A conceptual framework for ordered logistic regression models. *Sociological Methods & Research*, 38(2), 306-347. doi: 10.1177/0049124109346162
- Gannon, B. (2009). The influence of economic incentives on reported disability status. *Health Economics*, 18(7), 743-759. doi: 10.1002/hec.1399
- Gourieroux, C., Laffont, J. J., & Monfort, A. (1980). Coherency conditions in simultaneous linear equation models with endogenous switching regimes. *Econometrica*, 48(3), 675-696.
- Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equations system. *Econometrica*, 46(4), 931-959.

- Lewbel, A. (2007). Coherency and completeness of structural models containing a dummy endogenous variable. *International Economic Review*, 48(4), 1379-1392. doi: [10.1111/j.1468-2354.2007.00466.x](https://doi.org/10.1111/j.1468-2354.2007.00466.x)
- Lindeboom, M., & van Doorslaer, E. (2004). Cut-point shift and index shift in self-reported health. *Journal of Health Economics*, 23(6), 1083-1099. doi: [10.1016/j.jhealeco.2004.01.002](https://doi.org/10.1016/j.jhealeco.2004.01.002)
- Liu, X., & Koirala, H. (2012). Ordinal regression analysis: Using generalized ordinal logistic regression models to estimate educational data. *Journal of Modern Applied Statistical Methods*, 11(1), 242-254. Retrieved from <http://digitalcommons.wayne.edu/jmasm/vol11/iss1/21/>
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Beverly Hills, CA: Sage Publications.
- Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge, UK: Cambridge University Press.
- McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(2), 109-142. Retrieved from <http://www.jstor.org/stable/2984952>
- McCullagh, P. J. (2005). The proportional odds model. In P. Armitage & T. Colton (Eds.), *Encyclopedia of biostatistics* (2nd ed.). New York, NY: Wiley.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.
- McKelvey, R. D., & Zavoina, Z. (1975). A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology*, 4(1), 103-120. doi: [10.1080/0022250X.1975.9989847](https://doi.org/10.1080/0022250X.1975.9989847)
- O'Connell, A. A., & Liu, X. (2011). Model diagnostics for proportional and partial proportional odds models. *Journal of Modern Applied Statistical Methods*, 10(1), 139-175. Retrieved from <http://digitalcommons.wayne.edu/jmasm/vol10/iss1/15/>
- Peterson, B., & Harrell, F. E. (1990). Partial proportional odds models for ordinal response variables. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 39(2), 205-217. doi: [10.2307/2347760](https://doi.org/10.2307/2347760)
- Pratt, J. W. (1981). Concavity of the log likelihood. *Journal of the American Statistical Association*, 76(373), 103-106. doi: [10.1080/01621459.1981.10477613](https://doi.org/10.1080/01621459.1981.10477613)
- R Development Core Team. (2009). R: A language and environment for statistical computing [computer software]. Vienna, Austria: R Foundation for Statistical Computing.

Tamer, E. (2003). Incomplete simultaneous discrete response models with multiple equilibria. *Review of Economic Studies*, 70(1):147-165. doi: [10.1111/1467-937X.00240](https://doi.org/10.1111/1467-937X.00240)

Terza, J. V. (1985). Ordinal probit: A generalization. *Communications in Statistics – Theory and Methods*, 14(1), 1-11. doi: [10.1080/03610928508828893](https://doi.org/10.1080/03610928508828893)

Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *The Stata Journal*, 6(1), 58-82.

Winship, C., & Mare, R. D. (1984). Regression models with ordinal variables. *American Sociological Review*, 49(4), 512-525. Retrieved from <http://www.jstor.org/stable/2095465>

Yee, T. W. (2010). The VGAM package for categorical data analysis. *Journal of Statistical Software*, 32(10), 1-34. doi: [10.18637/jss.v032.i10](https://doi.org/10.18637/jss.v032.i10)

## Endnotes

1. To our knowledge, the first suggestion of such a model is contained in Winship and Mare (1984, p. 519). Long (1997, ch. 6) calls this “parallel regressions”. It is latent parallelism. Parallelism holds in the latent variable representation though not in probabilities under asymmetry.
2. We could only require full column rank and finite  $x$ .
3. In the class of models we consider, parameters are generally estimated to scale. The standard deviation of this error is the most commonly used method of scaling.
4. Boes and Winkelmann (2006) show that such a model is similar, in likelihood terms, to what is known as the generalized threshold model of Maddala (1983) and Terza (1985).
5. These conditions are elaborated by McCullagh (2005).  

$$\tau_j - \tau_{j-1} > \mathbf{X}_i (\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j-1})$$
6. But see the class of location-scale models. Cox (1995) discussed generalizations of the location-scale model that nest, as special cases, the partial-proportional odds model of Peterson and Harrell (1990). These results generalize to that case because the scale parameters cannot collapse to zero and the measure of the set of contradictions, though possibly shrinking, similarly does not collapse to zero.

7. The standard logistic distribution has variance equal to  $\frac{\pi^2}{3}s^2$ , where  $s$  is a scaling parameter. If we set  $s = \frac{\sqrt{3}}{\pi}$ , we can make the variance one.
8. Coherence, in simultaneous equations with limited outcomes, refers to nonexistence of solutions. Completeness refers to multiplicity. These problems often arise in the analysis of simultaneous move, discrete action game theory and are tantamount to lack of existence, in the coherence case, and lack of uniqueness, in the completeness case, of equilibrium.
9. Peterson and Harrell (1990) were able to undertake the Monte Carlo simulations that they report because the parameters, as they set them, do not cross and they rely on probabilities fed to the canonical multinomial distribution rather than simulating latent quantities. Were they to have done the latter, they would have realized this.

## Appendix A. An R Example

```

> x1 <- sample(c(0, 1), size = 1000, replace = TRUE)
> x2 <- runif(1000, -1, 1)
> eps <- rlogis(1000)
> y.star1 <- 0.05 * x1 + x2 + eps
> y.star2 <- 0.1 * x1 + x2 + eps
> y1 <- (y.star1 < -0.5 & y.star2 < -0.5)
> y2 <- (y.star1 > -0.5 & y.star2 > -0.5 & y.star1 < 0.5 & y.star2 <
+ 0.5)
> y3 <- (y.star1 > 0.5 & y.star2 > 0.5)
> y <- y1 + 2 * y2 + 3 * y3
> table(y)
y
 0   1   2   3
14 392 216 378
> bad.result <- data.frame(y.star1, y.star2, y1, y2, y3, y)
> bad.result[y == 0, ]
      y.star1  y.star2  y1  y2  y3 y
40  -0.5114175 -0.4614175 FALSE FALSE FALSE 0
152  0.4553046  0.5053046 FALSE FALSE FALSE 0
163  0.4736140  0.5236140 FALSE FALSE FALSE 0
333 -0.5033633 -0.4533633 FALSE FALSE FALSE 0
417  0.4519173  0.5019173 FALSE FALSE FALSE 0
449  0.4507807  0.5007807 FALSE FALSE FALSE 0
464  0.4668629  0.5168629 FALSE FALSE FALSE 0
468  0.4720030  0.5220030 FALSE FALSE FALSE 0
663  0.4846675  0.5346675 FALSE FALSE FALSE 0
676  0.4669751  0.5169751 FALSE FALSE FALSE 0
677  0.4820676  0.5320676 FALSE FALSE FALSE 0
833 -0.5296321 -0.4796321 FALSE FALSE FALSE 0
834 -0.5144424 -0.4644424 FALSE FALSE FALSE 0
880  0.4776592  0.5276592 FALSE FALSE FALSE 0

```