

# Screening, Transforming, and Fitting Predictors for Cumulative Logit Model

Bruce Lund, Independent Consultant

## ABSTRACT

The cumulative logit model is a logistic regression model where the target (or dependent variable) has 2 or more ordered levels. If only 2 levels, then the cumulative logit is the binary logistic model. Predictors for the cumulative logit model might be “NOD” (nominal, ordinal, discrete) where typically the number of levels is under 20. Alternatively, predictors might be “continuous” where the predictor is numeric and has many levels. This paper discusses methods that screen and transform both NOD and continuous predictors before the stage of model fitting. Once a collection of predictors has been screened and transformed, the paper discusses predictor variable selection methods for model fitting. One focus of this paper is determining when a predictor should be allowed to have unequal slopes. If unequal slopes are allowed, then the predictor has  $J-1$  distinct slopes corresponding to the  $J$  values of the target variable. SAS® macros are presented which implement screening and transforming methods. Familiarity with PROC LOGISTIC is assumed.

## INTRODUCTION

In ordinal logistic regression the target (or dependent variable) has 3 or more levels and these levels are ordered.<sup>1</sup> For example, ordinal logistic regression applies to fitting a model where the target is a satisfaction rating (e.g. good, fair, poor). Ordinal logistic regression becomes binary logistic regression if the target has 2 levels. Ordinal logistic regression might be used in customer relationship management or credit risk models as well as models found in social or medical science. In these models it is common to consider nominal, ordinal, or discrete (NOD) predictors.<sup>2</sup> Additionally, continuous predictors are often considered for these models.<sup>3</sup> Generally, there are many potential predictors to consider.

The cumulative logit model is one form of the ordinal logistic model. This paper focuses exclusively on the cumulative logit model. A special case of the cumulative logit model is the proportional odds (PO) model. The general model is called the partial proportional odds (PPO) model. Exact definitions are given in the next section.

The following steps in developing a cumulative logit model are discussed in this paper:

**Screening:** SAS macros are presented to identify predictors that are too weak to be further considered.

**Transforming:** For NOD predictors the transformation involves binning (reducing the number of levels). Then either dummy variable or weight of evidence coding is applied. A SAS macro is discussed that performs binning. For a continuous predictor, a different SAS macro looks for a good transformation (e.g.  $\text{Log}(X)$ ,  $X^2$ , etc.). The use of proportional odds versus partial proportional odds is considered.

**Predictor Selection Methods:** Predictor selection methods for model fitting are discussed. These methods select which binned or transformed predictors to include in the final model. For a PO model there are (at least) three SAS procedures that provide predictor variable selection methods, PROC LOGISTIC, PROC HPLOGISTIC, PROC HPGENSELECT.<sup>4</sup> PROC LOGISTIC also fits a PPO model. If the input data set is suitably modified, then there is an unconventional approach that enables PROC HPLOGISTIC or PROC HPGENSELECT to fit a PPO model. This approach is discussed in Appendix B.

Finally, some of the methods of this paper also add value for the fitting of a binary logistic model. But the discussion of PPO models is not relevant in the binary case.

---

<sup>1</sup> Allison (2012, Chapter 6) gives an introduction to ordinal logistic models. See also Agresti (2010).

<sup>2</sup> Nominal, ordinal, or discrete predictors with less than (perhaps) 20 levels. This cut-off of 20 is subjective.

<sup>3</sup> A continuous predictor is interval-scaled with “many levels”. Distinction between discrete and continuous is subjective.

<sup>4</sup> Methods include FORWARD, BACKWARD, STEPWISE using significance levels. For HPLOGISTIC and HPGENSELECT the predictors can be selected by SBC, AIC, and other methods. HPGENSELECT includes LASSO.

## CUMULATIVE LOGIT MODEL WITH PROPORTIONAL ODDS (PO)

To describe PO, the following example is given: Assume there are 3 levels for ordered target Y: A, B, C and 2 numeric predictors X<sub>1</sub> and X<sub>2</sub>. Let X<sub>k,1</sub> be the value of predictor X<sub>1</sub> for the k<sup>th</sup> observation. Similarly, for X<sub>k,2</sub> be the value for X<sub>2</sub>. Let p<sub>k,j</sub> = probability that k<sup>th</sup> observation has the target value j = A, B or C.

A cumulative logit model with 4 parameters α<sub>A</sub> α<sub>B</sub> β<sub>X1</sub> β<sub>X2</sub> is given via these two response equations:<sup>5</sup>

$$\text{Log} (p_{k,A} / (p_{k,B} + p_{k,C})) = \alpha_A + \beta_{X1} * X_{k,1} + \beta_{X2} * X_{k,2}$$

$$\text{Log} ((p_{k,A} + p_{k,B}) / p_{k,C}) = \alpha_B + \beta_{X1} * X_{k,1} + \beta_{X2} * X_{k,2}$$

Notably, coefficient β<sub>X1</sub> of predictor X1 is the same in both response equations. Similarly, for predictor X2.

Formulas for the probabilities p<sub>k,A</sub>, p<sub>k,B</sub>, p<sub>k,C</sub> can be derived from the two response equations. To simplify the formulas, let T<sub>k</sub> and U<sub>k</sub>, for the k<sup>th</sup> observation be defined by the equations below:

T <sub>k</sub>	= exp (α <sub>A</sub> + β <sub>X1</sub> *X <sub>k,1</sub> + β <sub>X2</sub> *X <sub>k,2</sub> )
U <sub>k</sub>	= exp (α <sub>B</sub> + β <sub>X1</sub> *X <sub>k,1</sub> + β <sub>X2</sub> *X <sub>k,2</sub> )

**Table 1. Supporting Notation**

Then, after algebraic manipulation, the probability equations in Table 2 are derived:

Response	Probability Formula
A	p <sub>k,A</sub> = 1 - 1/(1+T <sub>k</sub> )
B	p <sub>k,B</sub> = 1/(1+T <sub>k</sub> ) - 1/(1+U <sub>k</sub> )
C	p <sub>k,C</sub> = 1/(1+U <sub>k</sub> )

**Table 2. Cumulative Logit Model - Equations for Probabilities**

This cumulative logit model satisfies the following conditions for X<sub>1</sub> (and the analogous conditions for X<sub>2</sub>):

Let “r” and “s” be two values of X<sub>1</sub> and fix the value for X<sub>2</sub> at w. Let k<sub>1</sub> be an observation at (r, w) and k<sub>2</sub> be an observation at (s, w). Using probability formulas from Table 2:

$$\text{Log} \left[ \frac{p_{k1,A}/(p_{k1,B} + p_{k1,C})}{p_{k2,A}/(p_{k2,B} + p_{k2,C})} \right] = \text{Log} (p_{k1,A} / (p_{k1,B} + p_{k1,C})) - \text{Log} (p_{k2,A} / (p_{k2,B} + p_{k2,C})) = (r - s) * \beta_{X1}$$

$$\text{Log} \left[ \frac{(p_{k1,A} + p_{k1,B})/p_{k1,C}}{(p_{k2,A} + p_{k2,B})/p_{k2,C}} \right] = \text{Log} ((p_{k1,A} + p_{k1,B}) / p_{k1,C}) - \text{Log} ((p_{k2,A} + p_{k2,B}) / p_{k2,C}) = (r - s) * \beta_{X1}$$

These equations display the “proportional odds” property. Specifically, the difference of cumulative logits for values of X<sub>1</sub> at r and s is proportional to the difference (r - s). This proportional odds property is a result of the equality of the coefficients of predictor X<sub>1</sub> across the cumulative logit response equations.

The parameters for the cumulative logit model are estimated by maximizing the log-likelihood equation in a manner similar to the binary case (Agresti 2010, p 58).

### EXAMPLE OF FITTING A PO MODEL

Consider a data set called Test. Let Y be the target and suppose there are numeric predictors X<sub>1</sub> and X<sub>2</sub>. If Y has 3 or more levels, the default PROC LOGISTIC model is the proportional odds cumulative logit model. No LINK statement is needed, although, optionally, “/ LINK=CLOGIT” could be added to the MODEL statement. To run the PO cumulative logit model the PROC LOGISTIC statements are simply:

```
PROC LOGISTIC DATA= Test;
MODEL Y= X1 X2;
run;
```

<sup>5</sup> If J levels for Target, there are J-1 response equations. A “cumulative logit” is the log of the ratio of the “cumulative probability up to j” (in the ordering of the target) in the numerator over “one minus the cumulative probability up to j” in the denominator. When the target has levels A, B, C, then j goes from A to B. These cumulative logits are:

Log (p<sub>k,A</sub> / (p<sub>k,B</sub> + p<sub>k,C</sub>)) for the first response equation  
 Log ((p<sub>k,A</sub> + p<sub>k,B</sub>) / p<sub>k,C</sub>) for the second response equation

## CUMULATIVE LOGIT MODEL WITH PARTIAL PROPORTIONAL ODDS (PPO)

To describe PPO, the following example is given: Assume there are 3 levels for ordered target Y: A, B, C and there are 3 numeric predictors R, S and Z. Let  $p_{k,j}$  = probability that  $k^{\text{th}}$  observation has the target value  $j = A, B$  or  $C$

In this example the PPO Model will have 6 parameters  $\alpha_A \alpha_B \beta_R \beta_S \beta_{Z,A} \beta_{Z,B}$  given in 2 equations:

$$\text{Log} (p_{k,A} / (p_{k,B} + p_{k,C})) = \alpha_A + \beta_R * R_k + \beta_S * S_k + \beta_{Z,A} * Z_k$$

$$\text{Log} ((p_{k,A} + p_{k,B}) / p_{k,C}) = \alpha_B + \beta_R * R_k + \beta_S * S_k + \beta_{Z,B} * Z_k$$

Here, Z has different coefficients for the 2 response equations. In general, for PPO some predictors may have coefficients with different values across the response equations. Predictor Z is said to have “unequal slopes”.

Formulas for probabilities  $p_{k,A}, p_{k,B}, p_{k,C}$  continue to be given by Table 2 after modifications to definitions of T and U to reflect the PPO model. In unusual cases it is possible for a PPO probability to be negative.<sup>6</sup>

### EXAMPLE OF FITTING A PPO MODEL

Suppose Y is an ordered target with 3 levels and that X1 and X2 are numeric predictors. Assume the modeler wants to allow X1 to have unequal slopes. The PROC LOGISTIC statements are:

```
PROC LOGISTIC DATA= Test;  
MODEL Y= X1 X2 / UNEQUALSLOPES= (X1);  
run;
```

### A SIMPLIFICATION FOR THIS PAPER

All examples of the cumulative logit model in this paper will assume the target has 3 levels. However, the concepts presented in the paper extend to targets with more than 3 levels.

## GOALS OF THE PAPER

Now that the cumulative logit model (PO and PPO) has been defined, the goals of the paper can be listed.

- This paper provides techniques for screening, binning, and transforming of predictors for the cumulative logit model. Additionally, SAS macros are discussed which implement the screening, binning, and transforming. Issues related to the use of PO versus PPO are discussed.
- After screening, binning, and transforming the predictors the modeler may have many candidate predictors for potential use in a final model. This paper discusses methods for predictor selection to fit a final model. In particular, there is a discussion of SAS procedures which can fit the PO cumulative logit model, and of how these procedures extend to the fitting of a PPO model.

## TEST OF THE PROPORTIONAL ODDS ASSUMPTION

PROC LOGISTIC, when fitting a PO cumulative logit model, performs a test of the proportional odds assumption. A rejection of PO implies that at least one predictor should have unequal slopes across the response equations. If there are K predictors and J target levels, then the PO test statistic is chi-square with  $(J-2)*K$  d.f. A small right-tail probability of the test statistic rejects the proportional odds assumption.<sup>7</sup>

---

<sup>6</sup> See NOTE at bottom of SAS webpage of Example 72.18 for discussion of the possibility of negative probabilities. [http://support.sas.com/documentation/cdl/en/statug/68162/HTML/default/viewer.htm#statug\\_logistic\\_examples22.htm](http://support.sas.com/documentation/cdl/en/statug/68162/HTML/default/viewer.htm#statug_logistic_examples22.htm) Also see Richard Williams (2008, slide 47) at <https://www.stata.com/meeting/germany08/GSUG2008.pdf>

<sup>7</sup> For description of proportional odds assumption test statistic, see SAS/STAT(R) 9.3 User's Guide at: [https://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug\\_logistic\\_sect039.htm](https://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_logistic_sect039.htm) Also see Allison (2012 p. 168) for discussion of the performance of the test statistic.

## EXAMPLE: TEST OF THE PROPORTIONAL ODDS ASSUMPTION

```

DATA Test;
X1=1; X2=3; Y="A"; output;
X1=1; X2=3; Y="B"; output;
X1=1; X2=3; Y="C"; output;
X1=1; X2=3; Y="A"; output;
X1=2; X2=4; Y="A"; output;
X1=2; X2=3; Y="C"; output;
X1=2; X2=3; Y="C"; output;
X1=2; X2=4; Y="C"; output;
X1=2; X2=3; Y="B"; output;
X1=3; X2=3; Y="C"; output;
X1=3; X2=3; Y="A"; output;
X1=3; X2=3; Y="A"; output;
X1=3; X2=4; Y="C"; output;
X1=3; X2=4; Y="B"; output;
run;
PROC LOGISTIC DATA= Test;
MODEL Y= X1 X2;
run;

```

PROC LOGISTIC produces Table 3. Here, the proportional odds (PO) assumption test has p-value of 0.830. This is a strong acceptance of PO.

Score Test for the Proportional Odds Assumption		
Chi-Square	DF	Pr > ChiSq
0.372	2	0.830

**Table 3. PROC LOGISTIC Report of Test of Proportional Odds Assumption**

## SATURATED PPO CUM LOGIT MODEL WITH ONE NOD PREDICTOR

Again consider data set Test. The saturated model with single predictor X1 is the model with X1 in the CLASS statement and X1 in the UNEQUALSLOPES statement.

Since X1 has 3 levels, the CLASS statement produces 2 dummies for the 2 non-reference levels. This leads to 4 coefficients for X1 corresponding to the product of 2 dummies and 2 response equations. The results are shown in Table 4.

```

PROC LOGISTIC DATA= Test;
CLASS X1;
MODEL Y= X1 / UNEQUALSLOPES= (X1);
run;

```

Analysis of Maximum Likelihood Estimates							
Parameter		Y	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept		A	1	-0.597	0.585	1.041	0.308
Intercept		B	1	0.366	0.577	0.402	0.526
X1	1	A	1	0.597	0.822	0.528	0.468
X1	1	B	1	0.732	0.882	0.690	0.406
X1	2	A	1	-0.789	0.871	0.820	0.365
X1	2	B	1	-0.772	0.782	0.974	0.324

**Table 4. PROC LOGISTIC Report for a Saturated PPO Model**

Table 5 gives the X1 x Y frequency table with row percentages. The notation below will be used for the cell counts and marginal row and column counts from Table 5:

$n_{i,j}$  = count for cell (X1=i, Y=j) for  $i = 1, 2, 3$  and  $j = A, B, C$

$n_{i,\cdot}$  or  $n_{\cdot,j}$  give the marginal counts, where the “.” indicates summation across the index

Probabilities for A, B, or C are:  $P(Y = j | X1 = i) = p_{i,j} = n_{i,j} / (n_{i,\cdot})$  for  $i = 1, 2, 3$  and  $j = A, B, C$ . PROC LOGISTIC does not need to be run to find these probabilities. See the row percentages in Table 5.

X1	Y			Tot
	A	B	C	
1	2 .50	1 .25	1 .25	$n_{1,\cdot} = 4$
2	1 .20	1 .20	3 .60	$n_{2,\cdot} = 5$
3	2 .40	1 .20	2 .40	$n_{3,\cdot} = 5$
	$n_{\cdot,A} = 5$	$n_{\cdot,B} = 3$	$n_{\cdot,C} = 6$	$n = 14$

**Table 5. Probabilities for A, B, C for the Saturated PPO Model**

The log-likelihood and likelihood ratio chi-square can be computed from these probabilities using the formulas shown below (if  $n_{i,j} = 0$ , then regard  $n_{i,j} * \log(p_{i,j})$  as 0):

$$LL(\text{full}) = \text{Log-Likelihood (intercept and covariates)} = \sum_{i=1}^3 \sum_{j=A}^C n_{i,j} * \log(p_{i,j}) = -14.185$$

$$LL(\text{restricted}) = \text{Log-Likelihood (intercept Only)} = \sum_{j=A}^C n_{\cdot,j} * \log(n_{\cdot,j} / n) = -14.853$$

$$\text{Likelihood Ratio Chi-Sq} = -2 * \{ LL(\text{restricted}) - LL(\text{full}) \} = 1.337 \text{ (4 d.f.)}$$

The widely used model c (or c-statistic) for the binary logistic model is also defined for the cumulative logit.<sup>8</sup> With additional programming the model c can be computed for the saturated model using the formula for  $P(Y = j | X1 = i)$ . In summary, PROC LOGISTIC is not required in order to compute likelihood ratio chi-square or the model c for the saturated cum logit model with one NOD predictor.

## MEASURES FROM THE SATURATED MODEL AND SCREENING

The saturated model for predictor X has probabilities equaling the X by Y row percentages (e.g. Table 5). But, in the worst case, X could be independent of Y. Then X would add no information about Y versus the intercepts-only model. Otherwise, how much improvement does saturated X give? Two ways to measure improvement are likelihood ratio chi-square and model c. Predictors can be screened using these two measures. Predictors that are strong on both would be good candidates to be kept for further study.

But if X passes the screening, the actual usage of X in a model is unlikely to be as the saturated version. Instead, unequal slopes may not be needed and X may be transformed or binned.

The availability of the simple formulas to calculate likelihood ratio chi-square and model c for the saturated model provide a way to efficiently compute these summary statistics for many NOD predictors with minimal passing of data. These facts are utilized by the macro %CUM\_LOGIT\_SCREEN\_1, which is discussed in the next section.

<sup>8</sup> Definition of Model c. For each observation:

- Let Probabilities for A, B, C be  $p_k$  for  $k = 1, 2, 3$
- Compute  $Mscore = \sum_{k=1}^3 p_k * (k - 1)$ . e.g. If  $p_2 = 0.4, p_3 = 0.1$ , then  $Mscore = 0.4 + 2*0.1 = 0.6$
- IP = “Informative Pairs” of obs (r, s) where Targets  $Y_r \neq Y_s$ 
  - If  $Y_r > Y_s$  and  $Mscore_r > Mscore_s$ , then CONCORDANT
  - If  $Y_r > Y_s$  and  $Mscore_r < Mscore_s$ , then DISCORDANT
  - Else TIE

$$\text{Model c} = \{ \text{CONCORDANT} + 0.5 * \text{TIE} \} / \text{IP}$$

## **%CUM\_LOGIT\_SCREEN\_1 (Dataset, Target, Input, Sort);**

```
DATA Test2;
X1=1; X2="3"; Y="A"; output;
X1=1; X2="3"; Y="B"; output;
X1=1; X2="3"; Y="C"; output;
X1=1; X2="3"; Y="A"; output;
X1=2; X2="2"; Y="A"; output;
X1=2; X2="3"; Y="C"; output;
X1=2; X2="3"; Y="C"; output;
X1=2; X2="2"; Y="C"; output;
X1=2; X2="3"; Y="B"; output;
X1=3; X2="3"; Y="C"; output;
X1=3; X2="3"; Y="A"; output;
X1=3; X2="3"; Y="A"; output;
X1=3; X2="4"; Y="C"; output;
X1=3; X2="4"; Y="B"; output;
run;
%CUM_LOGIT_SCREEN_1(Test2, Y, X1 X2, );
```

**%CUM\_LOGIT\_SCREEN\_1 (DATASET, TARGET, INPUT, SORT)** computes the likelihood ratio chi-square (LRCS), its significance level, and model c for the saturated model for each predictor listed in the INPUT parameter.<sup>9</sup> Predictors can be numeric or character. The right-most parameter (a space in this example) is called SORT, It designates the sort sequence of the output Table. The choices for SORT are “LRCS” (for significance of LRCS), “Model\_c”, or “space”. If space, then the order in INPUT is used.

The macro is illustrated for target Y and predictors X1 (numeric) and X2 (character) from data set Test2. SORT is space. The macro call produces Table 6.

Var_name	Levels	LRCS	Pr > ChiSq	Model c	ASE Model c
X1	3	1.337	0.8551	0.6349	0.1123
X2	3	3.063	0.5474	0.5556	0.0624

**Table 6. Report from %CUM\_LOGIT\_SCREEN\_1**

The significance level of the LRCS can be used to rank the predictors. But it is not possible to specify an absolute cut-off significance level that goes across all applications. The significance level depends on sample size. For example, if a data set is duplicated (two copies), then LRCS is doubled but has the same degrees of freedom. Similarly, model c can rank the predictors. Using the two rankings, along with the modeler’s subject matter expertise, the modeler can screen out obviously weak predictors. (Is there a cut-off value for model c? Perhaps a simulation study may be designed that can provide guidelines.)

A second macro is described below which also bears on the screening question.

## **%CUM\_LOGIT\_SCREEN\_2 (Dataset, Target, N\_Input, C\_Input, IV\_Adj, Miss);**

In the case of binary logistic models, weight of evidence (WOE) is often used to transform a NOD predictor. WOE is also utilized in computing Information Value (IV). For the binary logistic model IV is a measure of the predictive power of a predictor X for the target Y.<sup>10</sup>

WOE and IV can be defined for each “binary split” of the target levels in a cumulative logit model. In the case where the target has levels A, B, C there are two splits:

A vs. B,C and A,B vs. C

---

<sup>9</sup> The macro passes the full data set only once. Then a second pass is made of the output from a PROC SUMMARY.

<sup>10</sup> See Appendix A for an example calculation of IV and WOE

These two splits lead to two WOE predictors and two IV's. The first is for the binary pair A vs. B,C and the second is for the binary pair A,B vs. C. If the target has J levels, then there are J-1 WOE's and J-1 IV's. A numeric X is monotonic for a "split" when the associated WOE is monotonic vs. X. The example in Table 7 shows a predictor X and target Y. WOE1\_X is monotonic while WOE2\_X is not. The information values, IV1 and IV2, for the associated binary splits are given in the right two columns.

X	Frequencies			Binary: A vs. BC WOE1_X	Binary AB vs. C WOE2_X	Binary: A vs. BC IV1	Binary AB vs. C IV2
	Y=A	Y=B	Y=C				
1	4	1	1	0.811	0.734	0.225	0.159
2	3	1	3	-0.170	-0.588	0.012	0.157
3	1	2	1	-0.981	0.223	0.204	0.011
				Monotonic	NOT Monotonic	IV1= 0.441	IV2= 0.327

**Table 7. WOE and IV for Cumulative Logit with Target Y and Predictor X**

Macro %CUM\_LOGIT\_SCREEN\_2 looks at the relationship of a predictor (numeric or character) to the binary splits of the target. The macro parameters are shown:

- Dataset:** Data set containing the Target and Predictors
- Target:** At least 2 levels (observations with Target = missing are ignored)
- N\_Input:** Numeric Predictors (space delimited)
- C\_Input:** Character Predictors (space delimited)
- IV\_Adj:** YES, adds 0.1 to a zero cell to allow IV calculation
- Miss:** If YES, statistics for each INPUT variable are computed for its non-missing values. If not YES, then only "complete cases" are used (all variables must be non-missing for the observation to be used).

```
%CUM_LOGIT_SCREEN_2(Test2, Y, X1, X2, YES , );
```

Var _Name	Spit _Point	Levels	Character	Monotonic	C_Stat	Model c saturated	IV
X1	A - B	3	NO		0.533	0.644	0.312
X1	B - C	3	NO		0.552	0.656	0.347
X2	A - B	3	YES	YES	0.633	0.633	0.559
X2	B - C	3	YES		0.500	0.542	0.034

**Table 8. Report from %CUM\_LOGIT\_SCREEN\_2**

In Table 8 the Split-Point column shows where the target was split to form the cumulative logit.

The macro produces familiar and readily interpretable statistics which apply to the binary splits. The IV column shows if the predictor provides an useful information value with respect to the associated split.

Binary IV guidelines are given below <sup>11</sup>

IV Range	Interpretation
IV < 0.02	"Not Predictive"
IV in [0.02 to 0.1)	"Weak"
IV in [0.1 to 0.3)	"Medium"
IV ≥ 0.3	"Strong"

**Table 9. Practical Guide for Interpreting IV**

<sup>11</sup> Siddiqi (2017, p.179)

The c-stat (always computed, even if X is character) and model c (computed for the saturated binary model) provide additional measures of association for the predictor and the binary splits.<sup>12</sup>

For this very small example, X1 is strongly associated with both binary splits of the cum logit but this association is not close to being monotonic.<sup>13</sup> X2 is strongly associated with the first binary split A vs BC and very weakly associated with the second.

One strategy may be to insist that the predictor provide a “medium” or “strong” IV for at least one split in order to pass the screening.

## EXAMPLE FOR CUM\_LOGIT\_SCREEN 1 and 2

The data set CLS\_Example is used to illustrate the two Cum\_Logit\_Screen macros. The target is Y with 3 levels and there are 6 predictors X1-X5, C5. By construction, X5 has unequal slopes.

```
%LET ERROR = 0.01;
%LET SLOPE1 = 0.01;
%LET SLOPE2 = 0.05;
%LET SLOPE3 = 0.10;
%LET SLOPE4 = 0.20;
%LET SLOPE5 = 0.99;
%LET P_Seed = 5;

DATA CLS_Example;
Do i = 1 to 8000;
  X1 = floor(12*ranuni(2)) - 1.5;
  X2 = floor(2*ranuni(2)) - .5;
  X3 = floor(2*ranuni(2)) - .5;
  X4 = floor(2*ranuni(2)) - .5;
  X5 = floor(2*ranuni(2)) - .5;
  C1 = put(floor(4*ranuni(2)), z2.);
  C1_all = &SLOPE1*(C1='00') + &SLOPE2*(C1='01') + &SLOPE3*(C1='02');
  rannorx = rannor(1);
  /* See Table 1 - note different slopes in T and U for X5 */
  T = exp(0 + C1_all + &SLOPE1*X1 + &SLOPE2*X2 + &SLOPE3*X3 + &SLOPE4*X4 +
    &SLOPE5*X5 + &ERROR*rannorx);
  U = exp(1 + C1_all + &SLOPE1*X1 + &SLOPE2*X2 + &SLOPE3*X3 + &SLOPE4*X4 +
    &SLOPE1*X5 + &ERROR*rannorx);
  /* See Table 2 */
  PA = 1 - 1/(1 + T);
  PB = 1/(1 + T) - 1/(1 + U);
  PC = 1 - (PA + PB);
  /* Assign Target Values to match model probabilities */
  R = ranuni(&P_Seed);
  if R < PA then Y = "A";
  else if R < (PA + PB) then Y = "B";
  else Y = "C";
  OUTPUT;
end;
run;
```

```
%CUM_LOGIT_SCREEN_1(CLS_Example, Y, X1 X2 X3 X4 X5 C1, LRCS);
```

The predictors are sorted by Pr > ChiSq (parameter=LRCS). The best ranked predictor is X5. The Model c for X5 is **0.573** (best among the six). Predictor X5 would be retained for further analysis. A question to consider for X5 would be whether unequal slopes are required.

<sup>12</sup> For binary logistic, see Lund and Brotherton (2013) for a study of model c (here, model c is called x-stat).

<sup>13</sup> X is monotonic for a split if and only if c-stat = model c. Otherwise, c-stat < model c.



Obs	Var_Name	Levels	Log_L_Intercept	Log_L	LRCS	df	Pr>ChiSq	Model c
1	X5	2	-8209.1	-7822.8	772.64	2	0.0000	0.573
2	X4	2	-8209.1	-8198.4	21.42	2	0.0000	0.522
3	X1	12	-8209.1	-8190.8	36.63	22	0.0260	0.523
4	X2	2	-8209.1	-8205.7	6.82	2	0.0331	0.510
5	C1	4	-8209.1	-8205.1	8.00	6	0.2383	0.513
6	X3	2	-8209.1	-8207.7	2.68	2	0.2613	0.508

**Table 10: Cum\_Logit\_Screen\_1 for CLS\_Example**

`%CUM_LOGIT_SCREEN_2(CLS_Example, Y, X1 X2 X3 X4 X5, C1, , );`

Predictor is X5 is very strong for split A v. BC but weak for AB v. C. Strength on one split is a reason to keep X5. The difference in strength of X5 for the 2 binary splits suggests X5 may have unequal slopes.

Obs	Var_Name	Split_Point	Levels	Character	Monotonic	C_Stat	Model c	IV (Info Value)
1	C1	A - B	4	YES		0.503	0.515	0.003
2	C1	B - C	4	YES		0.503	0.516	0.003
3	X1	A - B	12	NO		0.518	0.524	0.011
4	X1	B - C	12	NO		0.517	0.532	0.014
5	X2	A - B	2	NO	YES	0.514	0.514	0.003
6	X2	B - C	2	NO	YES	0.506	0.506	0.001
7	X3	A - B	2	NO	YES	0.509	0.509	0.001
8	X3	B - C	2	NO	YES	0.507	0.507	0.001
9	X4	A - B	2	NO	YES	0.525	0.525	0.010
10	X4	B - C	2	NO	YES	0.523	0.523	0.009
11	X5	A - B	2	NO	YES	0.621	0.621	0.240
12	X5	B - C	2	NO	YES	0.500	0.500	0.000

**Table 11: Cum\_Logit\_Screen\_2 for CLS\_Example (Note: predictors with 2 levels are monotonic)**

## BINNING OF NOD PREDICTORS

Binning is the process of reducing the number of levels of a NOD predictor to achieve parsimony while preserving, as much as possible, the predictive power of the predictor. In addition to parsimony, a purpose of binning is to detect a relationship between the bins and empirical cumulative logits that satisfies business expectations and common sense.

A macro `%CUMLOGIT_BIN` is presented in Lund (2017) for binning for a cumulative logit model. The predictor X may be integer (with values 0 to 99) or character. The target variable is character or numeric with at least 2 levels. Observations with missing values are ignored.

One of the macro parameters is `MODE`. `MODE` defines the pairs of levels of predictor X that are eligible for collapsing together at each step in the binning process. Choices are:

- “A”: Any pairs are eligible.
- “J”: Only pairs of levels that are adjacent in the ordering of X (this includes character valued X)

Whatever value of `MODE`, a criterion is needed to decide which specific levels of X to collapse. The macro parameter `METHOD` is used to specify the criterion. Choices are:

- “IV” for `TOTAL_IV`. `Total_IV` is the sum of IV’s from binary splits of the cum logit model. The eligible pair which gives the greatest “IV” after collapsing vs. all others is the pair which is collapsed.
- “LL” for `-2*LOG(L)` computed for the saturated model. The eligible pair which gives the smallest “-2\*LOG(L)” after collapsing vs. all others is the pair which is collapsed.

Once specified, `MODE` and `METHOD` are fixed throughout the binning.

## %CUMLOGIT\_BIN

The parameters for %CUMLOGIT\_BIN are given below:

**DATASET:** Data set to be processed

**TARGET:** Target variable (numeric or character) with at least 2 levels. Missing values are ignored.

**X:** Predictor variable (numeric or character).

If numeric, then X must have integer values from 0 to 99. If character, then “embedded space”, !, +, \_, # must not occur anywhere in the value. If character values represent an ordered predictor, then care must be taken to assign values that give the intended ordering. X must have at least 2 levels. Missing values are ignored.

**W:** A frequency variable if present in DATASET. Otherwise enter 1. Space is not permitted as entry.

**METHOD:** IV or LL <sup>14</sup>

**MODE:** A or J

**ONE\_ITER:** YES | <other>.

YES | <other> reports statistics for no-binning solution. No binning performed. Priority over MIN\_BIN

**MIN\_BIN:**

INTEGER > 1 | space. Integer value restricts the processing to bin solutions where the number of BINs is greater or equal to the INTEGER. If <space>, then all bin solutions are processed.

**VERBOSE:** If not YES, then only Summary Report is displayed.

YES | <other>. YES significantly increases volume of displayed output. VERBOSE=YES can be run to obtain SAS code statements for weight of evidence and classification level coding as well as step-by-step detail.

### EXAMPLE

Consider data set called BACKACHE in Table 12. It gives the age of women who were pregnant and the severity of backache they experienced. Severity has three levels: 1, 2, and 3 with 1 being least severe.<sup>15</sup>

Age Group	Severity		
	1	2	3
15to19	10	5	2
20to22	20	12	2
23to24	19	8	3
25to26	15	13	4
27to28	8	7	2
29to30	6	7	3
31to32	4	5	3
33to35	5	1	4
36andUP	6	2	4
Total	93	60	27

**Table 12. Data Set BACKACHE in Pregnancy. N = 180**

The macro call is given for BACKACHE data with TARGET= SEVERITY, X= AGE\_GROUP, MODE= A, METHOD= IV, VERBOSE= NO. Since VERBOSE= NO, only the summary report is printed.

```
%CUMLOGIT_BIN (with keyword parameters);
```

<sup>14</sup> Variants of IV could be considered for future versions of CUMLOGIT\_BIN such as MAX(of binary split IV's). The binning solution that is found after the first collapse may be suboptimal with respect to IV (or LL). That is, even though IV (or LL) is stepwise optimized to reach “k” bins, there can be a better “k” bin solution.

<sup>15</sup> AGE\_GROUP of pregnant women with one of 3 levels of SEVERITY (1 = none or very little pain, 2 = troublesome pain, 3 = severe pain) of backache from “BACKACHE IN PREGNANCY” data set in Chatfield (1995, Exercise D.2).

Table 13 shows columns of the summary report. In column “Collapsed” the pair being collapsed at that step has “+” between the levels. When previously collapsed, “\_” appears. Minus2\_LL (-2\*Log Likelihood) and Total\_IV are given, followed by IV\_1 and IV\_2. Corr\_woe\_1\_2 gives correlation of the two WOE’s.

A good solution is the 5 bin solution because there is a large drop-off in Total\_IV when going to 4 bins (0.598 down to 0.561). Guidelines (like those in Table 12 for binary targets) are needed to interpret TOTAL\_IV. Alternatively, the binary guidelines might be applied to individual IVs from the binary splits.

Step	Collapsed	minus2_LL	Total_IV	IV_1	IV_2	Corr_woe_1_2
9	None	339.5	0.614	0.138	0.476	0.581
8	25to26+27to28	339.5	0.613	0.138	0.476	0.581
7	15to19+23to24	339.6	0.609	0.135	0.474	0.578
6	33to35+36UP	339.9	0.605	0.135	0.470	0.579
5	29to30+31to32	340.0	0.598	0.134	0.464	0.578
4	15to19_23to24+20to22	341.0	0.561	0.133	0.429	0.638
3	25to26_27to28+29to30_31to32	342.4	0.493	0.111	0.381	0.607
2	25to26_27to28_29to30_31to32+33to35_36andUP	350.4	0.324	0.103	0.221	1

**Table 13. Step by Step Results from %CUMLOGIT\_BIN with MODE=A**

The 5-bin solution is shown in Table 14. The bins are not monotonic nor required to be since MODE= A.

BIN1	BIN2	BIN3	BIN4	BIN5
15to19_23to24	20to22	25to26_27to28	29to30_31to32	33to35_36andUP

**Table 14. Five-Bin Solution from %CUMLOGIT\_BIN with MODE=A**

If dummy variables for the 5 bins are entered into a model, there is a question of whether unequal slopes should be allowed. Score Test for the proportional odds assumption can be used to test whether there is a proportional odds violation. The assignment statements, below, are provided by %CUMLOGIT\_BIN.

```
DATA BACKACHE_5; SET BACKACHE;
  if Age_group in ( "15to19","23to24" ) then Age_group_bin = 1;
  if Age_group in ( "20to22" ) then Age_group_bin = 2;
  if Age_group in ( "25to26","27to28" ) then Age_group_bin = 3;
  if Age_group in ( "29to30","31to32" ) then Age_group_bin = 4;
  if Age_group in ( "33to35","36andUP" ) then Age_group_bin = 5;
PROC LOGISTIC DATA= BACKACHE_5;
CLASS Age_group_bin;
MODEL SEVERITY= Age_group_bin;
run;
```

The Score Test chi-square is 14.00 with 4 d.f. and p-value of 0.007. Proportional odds is rejected.

If, instead, the 2 WOE transforms enter the model, a similar Score Test for the proportional odds assumption can be used. %CUMLOGIT\_BIN provides the coding for the two WOE transforms for the 5-bin solution (but not shown here).

```
PROC LOGISTIC DATA= BACKACHE_5;
MODEL SEVERITY= Age_group_woe1 Age_group_woe2;
run;
```

The Score Test chi-square is 12.94 with 2 d.f. and p-value of 0.002. Proportional odds is rejected.

Two comments: Sample size (N=180) is small and this could raise a doubt about the PO rejections. For discussion, see Allison (2012 p. 168). For the WOE transform approach, how many d.f. should be assigned to Score Test to reflect the WOE coding? (The answer is at least 2 and may be as high as 4, but there is no final answer.) More discussion of using PO or PPO is given in the last section of the paper.

## TRANSFORMATION OF CONTINUOUS PREDICTORS FOR CUM LOGIT MODEL

A modeler may decide to bin a continuous predictor  $X$ . Some form of fine classing is needed to create preliminary bins. Then a binning algorithm such as %CUMLOGIT\_BIN would be applied. However, binning creates arbitrary cut-points (discontinuities) which could distort identification of a true trend. Binning is especially inappropriate for small sample data sets since reproducible bins are unlikely to be found during binning. The alternative to binning is to find a good transformation for a continuous  $X$ .

Royston and Sauerbrei (2008) (hereafter R-S) in their book *Multivariate Model-building* discuss the Function Selection Procedure (FSP). In R-S the FSP is applied to the binary logistic model as well as to ordinary regression and Cox regression. See R-S for more references.

In this paper FSP is extended to the PO cumulative logit model. The mechanics of FSP extend without change from the binary case. See Lund (2018) for more detail about FSP and extension to the PO cumulative logit model. Later in the paper an extension to the PPO cumulative logit model is discussed.

### FSP HAS TWO PRELIMINARY STEPS

First, the predictor  $X$  must be positive and, if needed, a translation of  $X$  is applied. In fact, for numerical stability, it is better to translate  $X$  so that the minimum of  $X$  is 1 (if minimum is not already greater than 1).

Next a class of transformations of  $X$ , called fractional polynomials (FP), is defined. These fractional polynomials are given by:

$$X^p \text{ where } p \text{ is taken from } S = \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\} \text{ and where "X}^0\text{" denotes } \log(X)$$

FP1 refers to the collection of linear functions formed by the selection of one  $X^p$ . That is,

$$g(X,p) = \beta_0 + \beta_1 X^p$$

FP2 refers to the collection of linear functions formed by selection of two  $X^p$  as shown below:

$$G(X,p_1,p_2) = \beta_0 + \beta_1 X^{p_1} + \beta_2 X^{p_2} \quad p_1 \neq p_2$$

However, there is also the special case where the same fractional polynomial is used twice:

$$G(X,p_1,p_1) = \beta_0 + \beta_1 X^{p_1} + \beta_2 X^{p_1} \log(X) \quad p_1 = p_2$$

FP1 produces only monotonic curves. FP2 produces curves with a variety of non-monotonic shapes. Such an example is given in Figure 1.

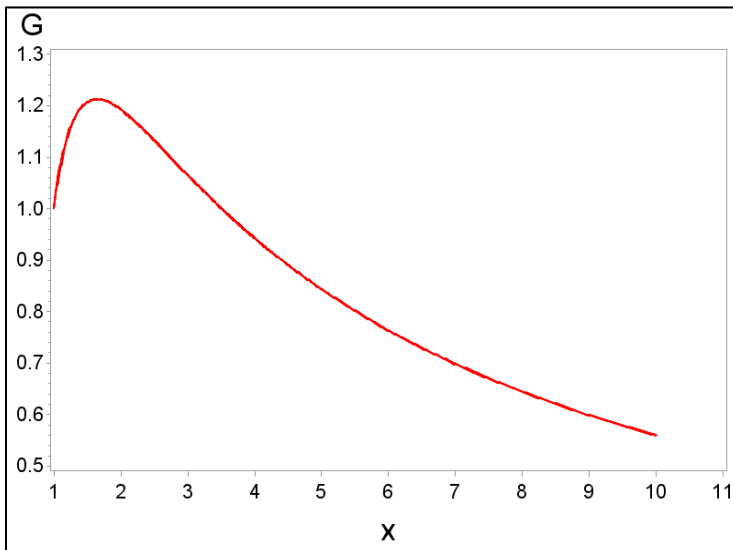


Figure 1. Graph of  $G(X,-1,-1) = X^{-1} + 2 X^{-1} \log(X)$

## FSP HAS TWO MAIN STEPS

**I. Searching for Best Transformations:** There is an exhaustive search of FP1 to find the function with maximum likelihood in a cumulative logit regression. This is called the FSP1 solution. Then, a second exhaustive search of FP2 is conducted to find the function with maximum likelihood within this collection. This is called the FSP2 solution. (This model fitting is for proportional odds models.)

**II. Performing Significance Testing:** The FSP significance testing follows these three steps. The test-statistic is displayed following the description of steps 1, 2, 3.

1. Perform a 4 d.f. test at the  $\alpha$  level of the FSP2 solution against the null model (no predictor). If the test is not significant, drop X from consideration and stop; otherwise continue.
2. Perform a 3 d.f. test at the  $\alpha$  level of the FSP2 solution against X. If the test is not significant, stop, the final model is linear X; otherwise continue.
3. Perform a 2 d.f. test at the  $\alpha$  level of the FSP2 solution against the FSP1 solution. If the test is significant, the final model is the FSP2, otherwise the FSP1 is the final model.<sup>16</sup>

The test-statistics for these three tests are the difference of -2 Log Likelihood's, as shown below:

$$\text{Test-Statistic} = (-2 \text{ Log Likelihood}_{\text{restricted model}}) - (-2 \text{ Log Likelihood}_{\text{FSP2 Solution}})$$

For large samples, the Test-Statistic is approximately a chi-square. The rationale for the degrees of freedom (4, 3, 2) in the 3-step tests of FSP is discussed in R-S (p. 79) for binary logistic models.<sup>17</sup>

This rationale generalizes to the cumulative logit model. See Lund (2018) for discussion.<sup>18</sup>

## IMPLEMENTATION OF FSP BY %FSP\_8LR FOR PO CUMULATIVE LOGIT MODELS

The macro call is %FSP\_8LR (DATASET, TARGET, INPUT, VERBOSE, ORDER);

Parameter definitions:

**DATASET:** The data set containing the target and predictors

**TARGET:** Target variable (character or numeric). At least two non-missing levels

**INPUT:** Numeric predictors (at least one). Predictors are delimited by a space. ("- " convention is implemented ... e.g. X1 - X6).

**VERBOSE:** YES ... "YES" produces more output.

**ORDER:** A | D ... The order for modeling the TARGET (A=ascending, D=descending).

%FSP\_8LR passes the data once to translate all predictors to have a minimum value of at least 1 and then, once again, to form the 8 fractional polynomial transforms for these predictors.

Next, for each predictor in INPUT, there are 8 PROC LOGISTIC's that are run. For each of the 8 PROC LOGISTIC runs, there are 9 predictors listed in the MODEL statement.

The 8 MODEL statements have 8 predictors in common but with different orderings and a ninth predictor that is unique to the statement. For the unique predictor see the right-most column in Table 15. The ordering matters since an INCLUDE is present in the MODEL statement.

---

<sup>16</sup> If FSP2 solution is selected as the *final* FSP transformation, then, normally, the two fractional polynomial transforms are entered separately into the model. However, if the linear combination has a special appeal to the modeler, then the FSP2 solution (with associated coefficients) can be entered.

<sup>17</sup> Loosely, for the first test ("Reject X") there are four parameters to be selection for the FSP2 solution. These are the two exponents (transforms) and the two coefficients. No parameters are required for the null model. The difference is 4 degrees of freedom. Similar logic applies to steps 2 and 3 which have test-statistics of 3 and 2 degrees of freedom.

<sup>18</sup> A simulation is given in Lund (2017) which extends the use of 4, 3, 2 d.f. to the PO cumulative logit model. No simulations have been conducted to justify the d.f. suggested in the later section for the PPO cumulative logit model.

	COMMON								UNIQUE
%LET Var1=	X	X <sup>-2</sup>	X <sup>-1</sup>	X <sup>-5</sup>	X <sup>.5</sup>	X <sup>2</sup>	X <sup>3</sup>	Log(X)	X Log(X)
%LET Var2=	X <sup>-2</sup>	X	X <sup>-1</sup>	X <sup>-5</sup>	X <sup>.5</sup>	X <sup>2</sup>	X <sup>3</sup>	Log(X)	X <sup>-2</sup> Log(X)
%LET Var3=	X <sup>-1</sup>	X	X <sup>-2</sup>	X <sup>-5</sup>	X <sup>.5</sup>	X <sup>2</sup>	X <sup>3</sup>	Log(X)	X <sup>-1</sup> Log(X)
%LET Var4=	X <sup>-5</sup>	X	X <sup>-2</sup>	X <sup>-1</sup>	X <sup>.5</sup>	X <sup>2</sup>	X <sup>3</sup>	Log(X)	X <sup>-5</sup> Log(X)
%LET Var5=	X <sup>.5</sup>	X	X <sup>-2</sup>	X <sup>-1</sup>	X <sup>-5</sup>	X <sup>2</sup>	X <sup>3</sup>	Log(X)	X <sup>.5</sup> Log(X)
%LET Var6=	X <sup>2</sup>	X	X <sup>-2</sup>	X <sup>-1</sup>	X <sup>-5</sup>	X <sup>.5</sup>	X <sup>3</sup>	Log(X)	X <sup>2</sup> Log(X)
%LET Var7=	X <sup>3</sup>	X	X <sup>-2</sup>	X <sup>-1</sup>	X <sup>-5</sup>	X <sup>.5</sup>	X <sup>2</sup>	Log(X)	X <sup>3</sup> Log(X)
%LET Var8=	Log(X)	X	X <sup>-2</sup>	X <sup>-1</sup>	X <sup>-5</sup>	X <sup>.5</sup>	X <sup>2</sup>	X <sup>3</sup>	Log(X) Log(X)

**Table 15. Eight Sets of Variables in MODEL statement for each predictor in INPUT**

PROC LOGISTIC is run 8 times for each INPUT predictor, once for each MODEL statement in Table 15.

```
PROC LOGISTIC; MODEL Y= &Var<K>
/ SELECTION=FORWARD INCLUDE=1 START=1 STOP=2 SLE=1;
```

The INCLUDE=1 forces the selection of the left most variable in the list. Then FORWARD with STOP=2 and SLE=1 will select one more variable. By this method all 36 FP2 pairs of fractional polynomials have a chance to be selected.<sup>19</sup> But the selection of the second variable by FORWARD, to add to the first variable forced in by INCLUDE=1, may be sub-optimal. The reason is that the second variable is selected by the best score chi-square, not by maximizing log likelihood of the model.

E.g. Consider row#1 in Table 15.

- First, X is forced in by INCLUDE=1
- Now perhaps the FORWARD criterion picks X<sup>-2</sup> to enter as the second variable.
- But the best log likelihood might be given by X and X<sup>3</sup>.

Examples where %FSP\_8LR produces a suboptimal FP2 solution do exist.<sup>20</sup> However, of all examples so far examined, the occurrence rate is not high and the severity is not material.

### EXAMPLE OF %FSP\_8LR

The macro code below creates a data set for use by %FSP\_8LR.<sup>21</sup> The target Y has three levels. The predictor X has levels 1 through 8, each with a frequency of 500. An “FP2” function is used to generate the values of Y. This FP2 function is: 0.2\*LOG(X) - 0.5\*(1/X) + 0.01\*error-term. The random error-term is drawn from a standard normal and then multiplied by 0.01.

```
%LET ERROR = 0.01;
%LET SLOPE1 = 0.2;
%LET SLOPE2 = -0.5;
%LET P_Seed = 5;
%MACRO SIM(NUM);
%DO Seed = 1 %TO &NUM;
  DATA SIM_&Seed;
  do i = 1 to 8000;
```

<sup>19</sup> There might appear to be redundancy in Table 15. Specifically, consider row #1 where the variable X is forced-in. Then in the FORWARD step the next variable to be selected is the one with the greatest score chi-square. Suppose FORWARD selects X<sup>-2</sup>. Can row#2, where X<sup>-2</sup> is forced by INCLUDE=1, be simplified by checking only the newly occurring predictor X<sup>-2</sup> Log(X)? The apparent reason would be that X and X<sup>-2</sup> were paired by the logistic model of row#1. But this reasoning is not valid. In row#2 the predictor selected by FORWARD can be any of the other 8.

<sup>20</sup> See R-S (2008) p 266 Whitehall I Data. In this data set there are CIGS and ALL10 (values 0 and 1). Let INPUT consist of X where X = CIGS + 1. Let TARGET = ALL10 (binary). Run %FSP\_8LR. The FP2 solution is X<sup>-1</sup> and X<sup>1</sup>\*Log(X). This compares with the R-S solution of X<sup>-2</sup> and X<sup>-1</sup>. The difference in -2\*Log(L) is 10708.270 vs. 10707.827

<sup>21</sup> This macro program is used in Lund (2018) for a simulation study to support the “d.f. of 4, 3, 2” for FSP significance testing in the case of the PO cumulative logit model.

```

X = mod(i,16) + 1;
rannorx = rannor(&Seed);
/* See Table 1 - Note: equal slopes for LOG(X) and 1/X */
T = exp(0 + &SLOPE1*LOG(X) + &SLOPE2*(1/X) + &ERROR*rannorx);
U = exp(1 + &SLOPE1*LOG(X) + &SLOPE2*(1/X) + &ERROR*rannorx);
/* See Table 2 */
PA = 1 - 1/(1 + T);
PB = 1/(1 + T) - 1/(1 + U);
PC = 1 - (PA + PB);
/* Assign Target Values to match model probabilities */
R = ranuni(&P_Seed);
if R < PA then Y = "A";
else if R < (PA + PB) then Y = "B";
else Y = "C";
output;
end;

run;
%END;
%MEND;
%SIM(1);

```

The macro call is: %FSP\_8LR (SIM\_1, Y, X, NO, A); The summary report is given in Table 16.

Test Step	-2*Log(L)	Test Stat	d.f.	P-Value	Trans1	Trans2	ChiSq _PO	d.f. _PO	Prob _PO
Eliminate X	15824.5	172.05	4	0					
Use Linear	15709.3	56.86	3	0			1.934	1	0.164
<b>Use FP1</b>	15654.3	1.90	2	<b>0.387</b>	<b>p=-0.5</b>		3.112	1	<b>0.078</b>
Use FP2	15652.4				p=-2	log	2.479	2	0.290

**Table 16. Report from %FSP\_8LR for Target Y and Predictor X**

Based on three-step testing the recommended transformation is the FSP1 solution  $f(X) = X^{-0.5}$ . The first two steps (“eliminate X” and “Use Linear”) are both strongly rejected with p-values of essentially 0. The FSP1 solution is accepted with p-value of 0.387.

Since Y has more than 2 levels, the test of proportional odds (PO) is given. In this example, the p-value for the proportional odds test for best FP1 solution is **0.078** (a borderline rejection).

Given the borderline rejection the modeler might choose to add unequal slopes to  $X^{-0.5}$ .

### BETA VERSION: %FSP\_8LR\_PPO

A second macro %FSP\_8LR\_PPO implements FSP where UNEQUALSLOPES is specified for each transformation. Since additional parameters are required for PPO, the new 3-step testing involves degrees of freedom as given by these formulas: (Target\_Levels is the number of levels of Target.)

FSP2 vs Null: d.f. =  $2*(\text{Target\_Levels} - 1) + 2$

FSP2 vs Linear: d.f. =  $2*(\text{Target\_Levels} - 1)$

FSP2 vs FSP1: d.f. =  $(\text{Target\_Levels} - 1) + 1$

However, these degrees of freedoms have not been tested via simulations as was done for the PO cumulative logit model in Lund (2018).

The macro call is: %FSP\_8LR\_PPO (SIM\_1, Y, X, NO, );

The parameters are the same except that 5<sup>th</sup> parameter “ORDER” is now a dummy parameter with no meaning. (Since coefficients are not reported by %FSP\_8LR\_PPO, there is no reason to specify ORDER.)

Again, the FSP1 Solution  $f(X) = X^{-0.5}$  is the final transformation selected by FSP. The degrees of freedom for the three-step testing are 6, 4, 3. There is little to gain by using the PPO solution in this case.

Test Step	-2*Log(L)	Test Stat	d.f.	P-Value	Trans1	Trans2
Eliminate X	15824.5	175.53	6	0	Null	
Use Linear	15707.4	58.40	4	0		
<b>Use FP1</b>	15651.3	2.37	3	<b>0.499</b>	p=-0.5	
Use FP2	15649.0				p=-1	p=2

**Table 17. Report from %FSP\_8LR\_PPO for Target Y and Predictor X**

A PPO cum logit model may have difficulty in converging and some probabilities from a PPO may be negative. This behavior and the greater model complexity of a PPO model argues against the automatic generalization to PPO.

## MODEL FITTING METHODS

After binning and transforming the predictors, modeler will probably need to use a variable selection method in order to select the predictors for the final model.

The PO cumulative logit model is fit by three procedures that provide predictor variable selection methods: LOGISTIC, HPLOGISTIC, HPGENSELECT.

PROC LOGISTIC supports FORWARD, BACKWARD, STEPWISE selection where the predictor variables are selected according to user-specific significance levels (also the SCORE method is supported). Both HPLOGISTIC and HPGENSELECT can fit the PO cumulative logit model, and all advanced selection methods that are available for the binary logistic are also available here. These methods include selection by SBC, AIC, VALIDATE, LASSO (for HPGENSELECT).

With SAS/STAT 12.1 (2012) the UNEQUALSLOPES statement was added to PROC LOGISTIC in the MODEL statement to enable the fitting of PPO models. UNEQUALSLOPES allowed the designated predictors to have coefficients with different values across the response equations. The default for a predictor is equal slopes. If UNEQUALSLOPES appears in the PROC LOGISTIC MODEL statement, then all predictor SELECTION methods can be used except SELECTION = SCORE.

With SAS/STAT 14.1 (2015) more features were added to PROC LOGISTIC. Now the simultaneous use of EQUALSLOPES and UNEQUALSLOPES in MODEL / SELECTION= method<sup>22</sup> determines whether a predictor enters the model and, if so, whether with equal slopes or unequal slopes. Hilliard (2017) provides explanations and examples.<sup>23</sup> See Derr (2013) for an earlier paper on related topics.

The example below follows Hilliard (2017, p. 10).

### EXAMPLE: PREDICTOR SELECTION BY PROC LOGISTIC, EQUAL V. UNEQUAL SLOPES

The predictor selection methods of PROC LOGISTIC can be used to “decide” if designated predictors should have unequal slopes. Consider the data set RANDOM with predictors X1 - X4.

```
DATA RANDOM;
Do ID = 1 to 5000;
  random = ranuni(1);
  If random < .5 then Y = "A";
  else if random < .8 then Y = "B";
  else Y = "C";
  X1 = floor(ranuni(1)*5) * random;
```

<sup>22</sup> Using STEPWISE, FORWARD, or BACKWARD

<sup>23</sup> Also see Hilliard (2017 p. 6) for a short summary and review of other software packages that fit the cumulative logit model and their capabilities to handle selection of predictors with unequal slopes.



```

X2 = rannor(1) * random;
X3 = ranuni(10) * random;
X4 = X3*ranuni(10);
output;
end;
run;

```

FORWARD selection is used in the example below where significance level to enter (SLE) is set at 0.05. The inclusion of both EQUALSLOPES and UNEQUALSLOPES determines whether a predictor from X1 - X4 will enter the model and, if so, whether with equal slopes or unequal slopes.

```

PROC LOGISTIC DATA= RANDOM;
MODEL Y= X1 X2 X3 X4
/ SELECTION= FORWARD SLE= 0.05 EQUALSLOPES UNEQUALSLOPES;
run;

```

As shown by Table 18, X2 and X4 are not selected and X1 and X3 are selected with unequal slopes (as indicated by the prefix “U”).

Analysis of Maximum Likelihood Estimates						
Parameter	Y	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	A	1	2.5651	0.0721	1265.20	<.0001
Intercept	B	1	3.4682	0.0899	1489.38	<.0001
U_X1	A	1	-1.1298	0.0485	541.76	<.0001
U_X1	B	1	-0.6621	0.0373	315.68	<.0001
U_X3	A	1	-7.0304	0.2412	849.54	<.0001
U_X3	B	1	-3.9414	0.1800	479.34	<.0001

**Table 18. Coefficients for PPO Model by PROC LOGISTIC**

The criterion for predictor selection provided by PROC LOGISTIC in this approach is “significance level”. The many criticisms of FORWARD, BACKWARD, STEPWISE selection by “p-values” will therefore apply. An alternative to significance level selection is proposed below.

### PPO MODELS BY HPLOGISTIC OR HPGENSELECT?

HPLOGISTIC and HPGENSELECT do not directly fit the PPO model. But can HPLOGISTIC and HPGENSELECT somehow be “tricked” to fit a PPO cumulative logit Model? The answer is “a qualified yes”, at least for the simple case where only a few predictors are designated for unequal slopes. Such a method is given in Appendix B.

## CONCLUSIONS

This paper describes methods for the process of Screening, Transforming, and Fitting Predictors for Cumulative Logit Model:

1. **Screening:** Two macros were given to screen NOD predictors. FSP and an associated macro screens continuous predictors.
2. **Binning and Transforming:** A binning macro was provided for NOD predictors. FSP and an associated macro selects a transformation for continuous predictors. The use of unequal slopes must be resolved.
3. **Predictor Selection:** The PO model can be fit by LOGISTIC, HPLOGISTIC, and HPGENSELECT with all their selection methods as used for a binary model. PROC LOGISTIC with UNEQUALSLOPES can fit the PPO model using any selection method except SCORE.

As discussed in Appendix B: For PPO models it is possible to use HPLOGISTIC or HPGENSELECT to fit a model. Their advanced predictor selection methods are used to find *candidate* models. Then these *candidate* models are refit by PROC LOGISTIC with UNEQUALSLOPES to produce the gold standard coefficients for a final model.

SGF, Dallas 2019 v05

## REFERENCES

- Agresti, A (2010). *Analysis of Ordinal Categorical Data, 2<sup>nd</sup> Ed.*, Hoboken, NJ, John Wiley & Sons.
- Allison, P.D. (2012), *Logistic Regression Using SAS: Theory and Application 2<sup>nd</sup> Ed.*, Cary, NC, SAS Institute Inc.
- Chatfield, C. (1995). *Problem Solving: A Statistician's Guide, 2<sup>nd</sup> Ed.*, Boca Raton, FL: Chapman & Hall/CRC.
- Derr, B. (2013). Ordinal Response Modeling with the LOGISTIC Procedure, *Proceedings of the SAS Global Forum 2013 Conference*, Cary, NC, SAS Institute Inc.
- Hilliard, P. (2017). Using New SAS 9.4 Features for Cumulative Logit Models with Partial Proportional Odds, *Proceedings of the SAS Global Forum 2017 Conference*, Cary, NC, SAS Institute Inc.
- Lund, B. (2017). Binning of Predictors for the Cumulative Logit Model, SESUG 2017, *Proceedings*, Southeast SAS Users Group, paper SD-69.
- Lund, B. (2018). The Function Selection Procedure, *Proceedings of the SAS Global Forum 2018 Conference*, Cary, NC, SAS Institute Inc., paper 2390.
- Lund B. and Brotherton D. (2013). Information Value Statistic, *MWSUG 2013, Proceedings*, Midwest SAS Users Group, paper AA-14.
- Royston P. and Sauerbrei W. (2008). *Multivariate Model-building*, John Wiley & Sons, West Sussex, UK.
- Siddiqi, N. (2017). *Intelligent Credit Scoring, 2<sup>nd</sup> edition*, Hoboken, NJ, John Wiley & Sons, Inc.

## CONTACT INFORMATION

Your comments, questions, and requests for macro code are valued and encouraged. Contact author at:

Bruce Lund, Independent Consultant  
 blund\_data@mi.rr.com or blund.data@gmail.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

## APPENDIX A

Table 19 shows the calculations for WOE and IV for a binary target.

X	Y = 0 "B <sub>k</sub> "	Y = 1 "G <sub>k</sub> "	Col % Y=0 "b <sub>k</sub> "	Col % Y=1 "g <sub>k</sub> "	Log(g <sub>k</sub> /b <sub>k</sub> ) = <b>X_woe</b>	g <sub>k</sub> - b <sub>k</sub>	IV Terms (g <sub>k</sub> - b <sub>k</sub> ) * Log(g <sub>k</sub> /b <sub>k</sub> )
X1	2	1	25.0%	12.5%	<b>-0.69315</b>	-0.125	0.08664
X2	1	1	12.5%	12.5%	<b>0.00000</b>	0	0.00000
X3	5	6	62.5%	75.0%	<b>0.18232</b>	0.125	0.02279
SUM	8	8	100%	100%		<b>IV =</b>	<b>0.10943</b>

Table 19. Binary Target Example of WOE and IV for Binary-Target Y and Predictor X

## APPENDIX B

### THE DATA RECODING TRICK

The data set "RANDOM" was created earlier in the MODEL FITTING METHODS section of the paper. The target Y has 3 levels (A, B, C) in RANDOM. In data set "RECODE" there are two outputs for every input observation from RANDOM. The split number (0 or 1) corresponds to the first or second output and

is assigned to SPLIT. The first output statement effectively corresponds to the first cumulative logit response equation and the second output statement to the second cumulative logit response equation in a cumulative logit model (either PO or PPO). A new binary variable TARGET is created according to the code below.

```
DATA RECODE; Set RANDOM;
  Do; if Y="A" then TARGET=0; else TARGET=1; SPLIT=0; output;
  end;
  Do; if Y="A" or Y="B" then TARGET=0; else TARGET=1; SPLIT=1; output;
  end;
```

PROC GENMOD can fit a PPO Cumulative Logit model if the input data is recoded as above. But this recoding is not sufficient. The GENMOD REPEATED statement is needed in addition to the recoding.<sup>24</sup>

## PROC GENMOD

If the modeler wants predictors X1 and X3 to have unequal slopes, then this is achieved in GENMOD by:

```
PROC GENMOD DATA= RECODE; CLASS ID SPLIT;
MODEL TARGET= SPLIT X1 X2 X3 X4 X1*SPLIT X3*SPLIT
/ LINK= logit DIST= bin type3;
REPEATED subject= ID / type= unstructured;
run;
```

The unequal slopes for X1 and X3 is accomplished through the inclusion in the MODEL statement of predictors X1, X3, the interaction of X1 with SPLIT, and interaction X3 with SPLIT.<sup>25</sup> The statement "REPEATED subject=ID / type =unstructured" corrects for correlation of TARGET values within the two observations of an ID.<sup>26</sup> The output from GENMOD is not needed below, so it is not displayed.

## PROC LOGISTIC WITH UNEQUALSLOPES, THE GOLD STANDARD

The "gold standard" for fitting the PPO model to Data Set RANDOM with X1 and X3 having unequal slopes is PROC LOGISTIC with UNEQUALSLOPES = (X1 X3).

```
PROC LOGISTIC DATA= RANDOM; /* Model #1 */
MODEL Y= X1 X2 X3 X4 / UNEQUALSLOPES= (X1 X3);
run;
```

## PROC HPLOGISTIC AND PROC HPGENSELECT WITH DATA CODING TRICK

The data coding trick that was employed for PROC GENMOD can also be used to enable HPLOGISTIC and HPGENSELECT to approximate the fitting of a PPO method, thereby making their advanced predictor SELECTION methods available for model fitting. However, neither procedure has a statement like the REPEATED statement of GENMOD.

How different will be the results from those of Model #1 (i.e. PROC LOGISTIC above)?

```
PROC HPLOGISTIC DATA= RECODE; /* Model #2 */
CLASS SPLIT /PARAM= ref;
MODEL TARGET= SPLIT X1 X2 X3 X4 X1*SPLIT X3*SPLIT ;
run;
PROC HPGENSELECT DATA= RECODE; /* Model #3 */
CLASS SPLIT /PARAM= ref;
MODEL TARGET= SPLIT X1 X2 X3 X4 X1*SPLIT X3*SPLIT / DISTRIBUTION= BINARY;
run;
```

<sup>24</sup> Stokes, Davis, Koch (2000) Categorical Data Analysis, 2<sup>nd</sup> ed. P. 533

<sup>25</sup> The coefficient of X1 for the first response equation is the sum of the coefficient of X1 plus the coefficient associated with X1\*SPLIT(0). For the second response function it is coefficient of X1. Similarly for X3.

<sup>26</sup> Discussion of REPEATED:

[http://documentation.sas.com/?docsetId=statug&docsetVersion=14.3&docsetTarget=statug\\_genmod\\_syntax26.htm&locale=en](http://documentation.sas.com/?docsetId=statug&docsetVersion=14.3&docsetTarget=statug_genmod_syntax26.htm&locale=en)

## COEFFICIENTS FROM THE THREE MODELS

Some coefficients from Models #2, #3 must be added to together in order to be comparable to Model #1. After these adjustments Models #2, #3 have very similar coefficients to Model #1 as shown in Table 20.

	LOGISTIC with unequal slopes RANDOM Data MODEL #1	HPLOGISTIC RECODE Data MODEL #2	HPGENSELECT RECODE Data MODEL #3
Intercept A	2.5653	2.6183	2.6183
Intercept B	3.4682	3.4639	3.4639
X1 A	-1.1292	-1.1363	-1.1363
X1 B	-0.6613	-0.6607	-0.6607
X2	0.0340	0.0328	0.0328
X3 A	-6.9985	-7.1992	-7.1992
X3 B	-3.9089	-3.9153	-3.9153
X4	-0.0683	-0.0944	-0.0944

**Table 20. Coefficients from Fitting PPO Model to Data Set Random or Recode**

A comparison of probabilities was made for PROC HPLOGISTIC (Model #2) and PROC LOGISTIC with UNEQUALSLOPES (Model #1). In Table 21a the “Diff A” column is the difference in the probability of A for the same ID’s from Model #1 and Model #2. Similarly for “Diff B” and “Diff C”. The average value of the probability of A, B, and C in Model #1 and in Model #2 is shown in Table 21b.

	Diff A	Diff B	Diff C
Level:	Quantile	Quantile	Quantile
100% Max	0.0122	0.0097	0.0192
99%	0.0105	0.0083	0.0162
95%	0.0073	0.0065	0.0127
90%	0.0056	0.0055	0.0107
75% Q3	0.0025	0.0042	0.0080
50% Median	-0.0009	0.0026	0.0031
25% Q1	-0.0039	-0.0006	-0.0030
10%	-0.0052	-0.0035	-0.0088
5%	-0.0062	-0.0053	-0.0125
1%	-0.0080	-0.0084	-0.0187
0% Min	-0.0095	-0.0092	-0.0211

**Table 21a. Difference in Probabilities for PPO Fit by LOGISTIC vs. HPLOGISTIC with coding trick**

	Prob A	Prob B	Prob C
Mean Prob (Model #1)	0.4871	0.3147	0.1982
Mean Prob (Model #2)	0.4876	0.3130	0.1994

**Table 21b. Mean Probabilities for PPO Fit by LOGISTIC #1 vs. HPLOGISTIC #2 with coding trick**

As shown in Table 21a the differences are fairly small. There are moderately important differences in probabilities for A, B, C for high and low quantiles. Table 21b shows that the mean of probabilities for A, B, C are very similar for Model #1 and Model #2.

The lack of a REPEATED statement in Model #2 (and Model #3) has not greatly degraded the approximation to Model #1.<sup>27</sup>

<sup>27</sup> Contact the author for the SAS code that generated Table 21a,b.

## THIS SUGGESTS NEW TACTICS FOR PREDICTOR SELECTION IN PPO MODELS

The hope is that HPLOGISTIC or HPGENSELECT will select the predictors, using an advanced selection method, with or without unequal slopes, to find the “ideal” PPO model. With this belief, then it is attractive to use advanced predictor selection methods of HPLOGISTIC or HPGENSELECT to find good candidate models for this “ideal” model.

Once a good candidate model or models were identified, then a final fit of “gold standard” coefficients could be found by running PROC LOGISTIC with UNEQUALSLOPES.

An example is given next.

### EXAMPLE OF SELECTION OF PREDICTORS BY HPLOGISTIC WITH FORWARD AND SBC

HPLOGISTIC runs FORWARD with selection of predictors and choosing of the final model by SBC (see CHOOSE=SBC). The SBC criterion is “demanding” and is more likely than AIC to exclude a predictor.

```
PROC HPLOGISTIC DATA= RECODE;
CLASS SPLIT /PARAM = ref;
MODEL TARGET= SPLIT X1 X2 X3 X4 X1*SPLIT X3*SPLIT / INCLUDE=1;
SELECTION METHOD= FORWARD (SELECT=SBC CHOOSE=SBC STOP=NONE) DETAILS=ALL;
run;
```

Here is the selection summary. SPLIT is forced in (INCLUDE=1) to ensure that two intercepts are created. The lowest SBC is achieved at Step 4 and therefore X2 and X4 are not entered. Both X1 and X3 enter with unequal slopes since both X1 and X3 as well as X1\*SPLIT and X3\*SPLIT are selected.

Selection Summary			
Step	Effect Entered	Number Effects In	SBC
0	Intercept	1	
	SPLIT	2	11943
1	X3	3	9482.59
2	X1	4	8357.88
3	X3*SPLIT	5	8248.08
4	X1*SPLIT	6	8196.23*
5	X2	7	8203.75
6	X4	8	8212.85

**Table 22. FORWARD Selections by HPLOGISTIC with SELECT=SBC CHOOSE=SBC**

Now a final model would be fit by PROC LOGISTIC with the predictors identified by HPLOGISTIC. This provides the “gold standard” coefficients.

```
PROC LOGISTIC DATA= RANDOM;
MODEL Y= X1 X3 /* X2 X4 */ / UNEQUALSLOPES= (X1 X3);
run;
```

For this example, all this “extra work” has produced the same PPO model that was shown in Table 18 when PROC LOGISTIC was run with FORWARD and EQUALSLOPES UNEQUALSLOPES.