Non-Proportional-Odds-Modelle sind weitgehend verzichtbar: Sparsame Modellierung über Location-Shift-Ansätze

Moritz Berger

Universität Bonn

In Zusammenarbeit mit Gerhard Tutz

Kolloquium "Statistische Methoden in der empirischen Forschung", Berlin
17. Januar 2023

Related Articles

1. Tutz, Gerhard & Berger, Moritz (2017): Separating location and dispersion in ordinal regression models, Econometrics and Statistics 2, 131-148.

2. Tutz, Gerhard & Berger, Moritz (2022): Sparser ordinal regression models based on parametric and additive location-shift approaches, International Statistical Review 90(2), 306-327.

Cumulative Regression Models

General representation

$$P(Y \leq r|\boldsymbol{x}) = F(\eta_r(\boldsymbol{x})) \,, \quad r = 1, \dots, k-1$$

- $Y \in \{1, \dots, k\}$: ordinal response
- $\boldsymbol{x}$: vector of covariates
- $F(\cdot)$: cumulative distribution function
- $\eta_r(\cdot)$: predictor function

Proportional odds model (McCullagh, 1980)

$$P(Y \leq r|\boldsymbol{x}) = \frac{\exp(\eta_r)}{1+\exp(\eta_r)} = \frac{\exp(\beta_{0r} + \boldsymbol{x}^\top \boldsymbol{\beta})}{1+\exp(\beta_{0r} + \boldsymbol{x}^\top \boldsymbol{\beta})} \qquad \text{or equivalent}$$

$$\log\left(\frac{P(Y \leq r|\boldsymbol{x})}{P(Y > r|\boldsymbol{x})}\right) = \beta_{0r} + \boldsymbol{x}^\top \boldsymbol{\beta}$$

- β_{0r} : category-specific intercepts
- $\boldsymbol{\beta}$: regression coefficients

Proportional Odds Model

Interpretation of parameters

Define $\gamma(r|\boldsymbol{x}) = P(Y \leq r|\boldsymbol{x})/P(Y > r|\boldsymbol{x})$. Then the proportion of cumulative odds for two sets of covariates is given by

$$\frac{\gamma(r|\boldsymbol{x})}{\gamma(r|\tilde{\boldsymbol{x}})} = \exp\left((\boldsymbol{x} - \tilde{\boldsymbol{x}})^\top \boldsymbol{\beta}\right) ,$$

which does not depend on the category r.

In particular, $\exp(\beta_j)$ represents the factor by which the cumulative odds change, if x_j increases by one unit

$$\exp(\beta_j) = \frac{\gamma(r|x_1, \ldots, x_j + 1, \ldots, x_p)}{\gamma(r|x_1, \ldots, x_j, \ldots, x_p)} ,$$

which is the same for all odds.

Category-Specific Effects

Non-proportional odds model

$$\log\left(\frac{P(Y \leq r|\boldsymbol{x})}{P(Y > r|\boldsymbol{x})}\right) = \beta_{0r} + \boldsymbol{x}^\top \boldsymbol{\beta}_r$$

This general model usually provides a better fit to the data, but postulates that

$$\beta_{01} + \boldsymbol{x}^\top \boldsymbol{\beta}_1 \leq \ldots \leq \beta_{0,k-1} + \boldsymbol{x}^\top \boldsymbol{\beta}_{k-1}\,,$$

which is a strong restriction and may cause problems when estimating probabilities for future observations (Walker, 2016).

Partial proportional odds model

$$\log\left(\frac{P(Y \leq r|\boldsymbol{w},\boldsymbol{z})}{P(Y > r|\boldsymbol{w},\boldsymbol{z})}\right) = \beta_{0r} + \boldsymbol{w}^\top \boldsymbol{\beta}^w + \boldsymbol{z}^\top \boldsymbol{\beta}_r^z$$

The effects of $\boldsymbol{w}$ are global, while the effects of $\boldsymbol{z}$ are category-specific. In sociology, partial proportional odds models are also referred to as generalised ordered logit models (Williams, 2006).

Location-Scale Model

A cumulative model that accounts for additional dispersion is given by

$$\log\left(\frac{P(Y \leq r|\boldsymbol{x})}{P(Y > r|\boldsymbol{x})}\right) = \frac{\beta_{0r} + \boldsymbol{x}^\top \boldsymbol{\beta}_r}{\tau_x}, \quad r = 1, \ldots, k-1,$$

where τ_x is a variance parameter, which may depend on $\boldsymbol{x}$.

- In social science the model is also known as heterogeneous choice model (Williams, 2010).

Shortcomings:

- model is highly non-linear
- model is not a member of the class of GLMs

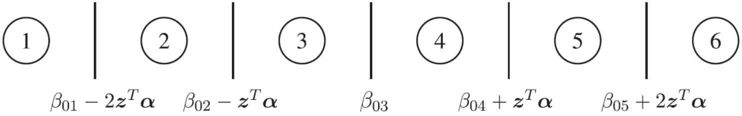
Location-Shift Models

An alternative extension of the proportional odds model assumes that the thresholds β_{0r} may change with an (additional) set of covariates $\boldsymbol{z}$. This yields the so-called location-shift model, which is given in closed form by

$$\log\left(\frac{P(Y \leq r|\boldsymbol{x},\boldsymbol{z})}{P(Y > r|\boldsymbol{x},\boldsymbol{z})}\right) = \beta_{0r} + \boldsymbol{x}^\top\boldsymbol{\beta} + (r - k/2)\boldsymbol{z}^\top\boldsymbol{\alpha}, \quad r = 1,\ldots,k-1\,.$$

- $\boldsymbol{x}^\top\boldsymbol{\beta}$: location term
- $\boldsymbol{z}^\top\boldsymbol{\alpha}$: dispersion term
- $(r - k/2)$: scaling factor

Illustration for $k = 6$



- $\boldsymbol{z}^\top\boldsymbol{\alpha}$ positive: intervals are widened $\Rightarrow$ weaker dispersion
- $\boldsymbol{z}^\top\boldsymbol{\alpha}$ negative: intervals are shrunk $\Rightarrow$ stronger dispersion

Application - Eye Vision

Stuart's (1953) quality of right eye vision data:

	Vision Quality			
	Highest (1)	2	3	Lowest (4)
Men	1053	782	893	514
Women	1976	2256	2456	789

Parameter estimates and standard errors:

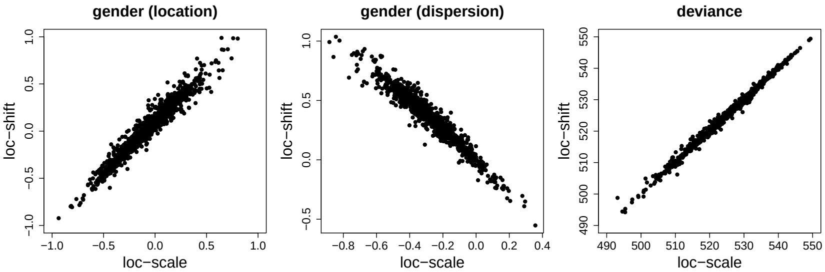
Covariate	**Proportional Odds Model**			**Location-Shift Model**		
	estimate	std error	z value	estimate	std error	z value
Intercept1	-0.905	0.034	-26.613	-0.721	0.037	-19.397
Intercept2	0.293	0.033	8.911	0.236	0.033	7.104
Intercept3	2.005	0.039	50.398	1.710	0.045	37.563
gender (female)						
location	-0.038	0.038	-1.003	0.042	0.038	1.109
dispersion				0.353	0.031	11.348

- LR test ($H_0 : \alpha = 0$): 122.5 on 1 df $\Rightarrow$ dispersion effect present!
- $e^{-\hat{\alpha}} = 0.70$ decreases odd for category 1
- $e^{\ \hat{\alpha}} = 2.01$ increases odd for categories ≤ 3

Comparison - Eye Vision

Location effects, dispersion effects and deviances for 100 sub-samples of size 200 of the eye vision data, when estimating

- the locoation-scale model (x-axis)
- the location-shift model (y-axis)



- relation between the parameters α and γ non-linear!

Hierarchy of Models

Let us consider the case when $\boldsymbol{x} = \boldsymbol{z}$. Then, one has

$$\log\left(\frac{P(Y \leq r|\boldsymbol{x})}{P(Y > r|\boldsymbol{x})}\right) = \beta_{0r} + \boldsymbol{x}^\top(\boldsymbol{\beta} + (r - k/2)\boldsymbol{\alpha}) = \beta_{0r} + \boldsymbol{x}^\top\boldsymbol{\beta}_r \,,$$

where $\boldsymbol{\beta}_r = \boldsymbol{\beta} + (r - k/2)\,\boldsymbol{\alpha}$.

$\Rightarrow$ The location-shift model is equivalent to a category-specific model with constraints.

Because the proportional odds model is a submodel of the location-shift model, the following nested structure holds

proportional odds model $\subset$ location-shift model $\subset$ non proportional odds model

$\Rightarrow$ One can investigate if the models can be simplified by testing the sequence of nested models.

Location-Shift and Partial Proportional Odds

Let $\boldsymbol{x}$ be partitioned into two subvectors $\boldsymbol{x}^\top = (\boldsymbol{w}^\top, \boldsymbol{z}^\top)$, such that $\boldsymbol{x}^\top \boldsymbol{\beta} = \boldsymbol{w}^\top \boldsymbol{\beta}^w + \boldsymbol{z}^\top \boldsymbol{\beta}^z$. Then one obtains

$$\begin{aligned}\eta_r =& \beta_{0r} + \boldsymbol{x}^T \boldsymbol{\beta} + (r - k/2) \boldsymbol{z}^T \boldsymbol{\alpha} \\ =& \beta_{0r} + \boldsymbol{w}^T \boldsymbol{\beta}^w + \boldsymbol{z}^T \boldsymbol{\beta}^z + (r - k/2) \boldsymbol{z}^T \boldsymbol{\alpha} \\ =& \beta_{0r} + \boldsymbol{w}^T \boldsymbol{\beta}^w + \boldsymbol{z}^T \boldsymbol{\beta}^z_r \,,\end{aligned}$$

where $\boldsymbol{\beta}^z_r = \boldsymbol{\beta}^z + (r - k/2)\, \alpha$.

$\Rightarrow$ The location-shift model is a submodel of the partial proportional odds that allows for an easier interpretation of effects.

+ The proposed model is between the most general model and a model with global effects, in which the impact of a single variable is described by just two parameters (instead of one or $k - 1$).

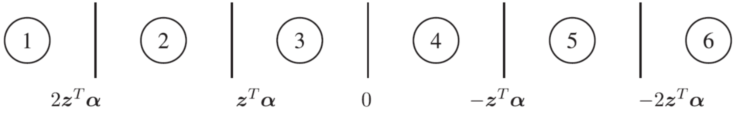
Adjacent Categories Models

Location-shift version of the model

$$\log \left(\frac{P(Y = r+1|\boldsymbol{x})}{P(Y = r|\boldsymbol{x})} \right) = \beta_{0r} + \boldsymbol{x}^T \boldsymbol{\beta} + (k/2 - r)\boldsymbol{z}^T \boldsymbol{\alpha}, \quad r = 1, \ldots, k-1.$$

- no ordering of intercepts has to be postulated!

Illustration for $k = 6$



- $\boldsymbol{z}^\top \boldsymbol{\alpha}$ positive: tendency to middle categories
- $\boldsymbol{z}^\top \boldsymbol{\alpha}$ negative: tendency to extreme categories

Estimation and Implementation

Estimation and evaluation by embedding the model into the framework of multivariate generalized linear models (GLMs)

- Data: $(\mathrm{y}_i, \mathrm{x}_i, \mathrm{z}_i)$, $i = 1, \ldots, n$
- Distributional assumption: $\mathrm{y}_i | \mathrm{x}_i, \mathrm{z}_i \sim M(1, \boldsymbol{\pi}_i)$, with $\boldsymbol{\pi}_i^\top = (\pi_{i,1}, \ldots, \pi_{i,k})$
- General form: $g(\boldsymbol{\pi}_i) = \mathrm{X}_i \boldsymbol{\delta}$, with total parameter vector $\boldsymbol{\delta}$

- Application of ML estimation and inference for multivariate GLMs by representation of the model with specific design matrix X_i
- Implemented in the R add-on package **ordDisp** (Berger, 2020), which internally calls function `vglm()` from R-package **VGAM**
- Easy handling (by argument `xij`) and quite fast computation

Application 1: Retinopathy Status

We consider a 6-year followup study on diabetes and retinopathy status (Bender and Grouven, 1998).

- Ordinal outcome: Retinopathy (1: no retinopathy, 2: nonproliferative retinopathy, 3: advanced retinopathy or blind)
- Risk factors
  - Smoking (SM; 0: no, 1: yes)
  - Diabetes duration (DIAB; in years)
  - glycosylated hemoglobin (GH; in percent)
  - diastolic blood pressure (BP; in mmHg)
- Available from the R-package **catdata** (Schauberger and Tutz, 2014)

Application 1: Retinopathy Status

	Covariate	Proportional Odds Model		
		estimate	std error	z value
location effects	SM	-0.254	0.191	-1.328
	DIAB	-0.139	0.013	-10.368
	GH	-0.459	0.074	-6.175
	BP	-0.072	0.013	-5.357
AIC			916.14	
BIC			942.65	
Deviance			904.14	

Application 1: Retinopathy Status

	Covariate	Location-Shift Model			Location-Scale Model		
		estimate	std error	z value	estimate	std error	z value
location effects	SM	-0.159	0.198	-0.802	-0.348	0.371	-0.938
	DIAB	-0.148	0.014	-10.524	-0.108	0.102	-1.066
	GH	-0.485	0.076	-6.324	-0.311	0.296	-1.053
	BP	-0.071	0.014	-5.204	-0.051	0.047	-1.092
dispersion effects	SM	0.491	0.235	2.087	-0.256	0.150	-1.707
	DIAB	-0.037	0.016	-2.254	0.035	0.010	3.490
	GH	-0.101	0.092	-1.099	0.043	0.053	0.805
	BP	-0.007	0.015	-0.465	-0.012	0.010	-1.186
AIC			912.45			907.17	
BIC			956.63			951.36	
Deviance			892.45			887.17	

Application 1: Retinopathy Status

	Covariate	Non Proportional Odds Model		
		estimate	std error	z value
	SM1	-0.405	0.205	-1.972
	SM2	0.086	0.254	0.340
	DIAB1	-0.129	0.014	-8.889
	DIAB2	-0.166	0.018	-9.264
	GH1	-0.435	0.080	-5.426
	GH2	-0.535	0.097	-5.470
	BP1	-0.068	0.014	-4.627
	BP2	-0.075	0.017	-4.432
AIC			912.45	
BIC			956.63	
Deviance			892.45	

Application 1: Retinopathy Status

Maybe the best solution?

	Covariate	Partial Proportional Odds Model		
		estimate	std error	z value
	SM1	-0.398	0.205	-1.943
	SM2	0.062	0.249	0.249
	DIAB1	-0.129	0.014	-8.895
	DIAB2	-0.165	0.017	-9.467
	GH	-0.467	0.074	-6.271
	BP	-0.071	0.013	-5.228
AIC		909.77		
BIC		945.12		
Deviance		893.77		

Application 2: Safety in Naples

We consider data of a survey conducted in the metropolitan area of Naples, Italy.

- Ordinal outcome: Feeling safe on a 10-point scale
- Large categories refer to high safety
- Data of 2225 participants
- Covariates:
  - Age
  - Gender (0: male, 1: female)
  - Residence (1: City of Naples, 2: District of Naples, 3: Other Campania, 4: Others Italia)
  - Educational degree (1: compulsory school, 2: high school diploma, 3: Graduated-Bachelor degree, 4: Graduated-Master degree, 5: Post graduated)
- Available from the R-package CUB (Iannario et al., 2015)

Application 2: Safety in Naples

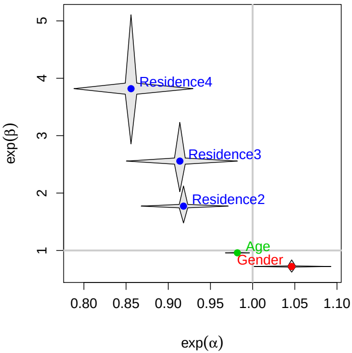
Fits of models with logistic link

	Deviance	df	Difference in deviances	df	p-value
Cumulative models					
Non proportional odds model	9825.78	19935			
Location-shift model	9899.67	19998	73.89	63	0.1640
Proportional odds model	9948.99	20007	49.32	9	0.0000
Adjacent categories models					
Model with category-specific effects	9828.07	19935			
Location-shift model	9902.43	19998	74.36	63	0.1549
Model with global effects	9959.00	20007	56.57	9	0.0000

- The full model with category-specific effects has 90 parameters, which reduces to 27 parameters in the location-shift model.

Application 2: Safety in Naples

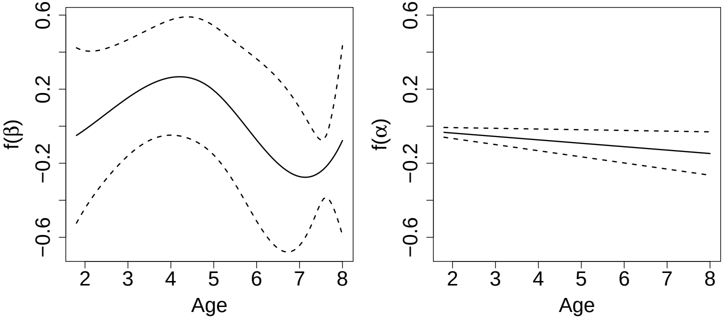
Illustration of Location and Dispersion Effects



- Cumulative model (left) and adjacent categories model (right)

Application 2: Safety in Naples

Cumulative model with smooth effect of age



- A test if the smooth effect of age is needed yields a p-value of 0.046.

Application 3: Nuclear Energy

We consider data from the German Longitudinal Election Study (GLES), a long-term study of the German electoral process (Rattinger et al., 2014).

- Pre-election survey for the German federal election in 2017
- Ordinal outcome: Fear due to the use of nuclear energy on a 7-point scale
- Large categories refer to high fear
- Covariates
    - Age
    - Gender (0: female, 1: male)
    - EastWest (1: East Germany/former GDR, 0: West Germany/former FRG)

Application 3: Nuclear Energy

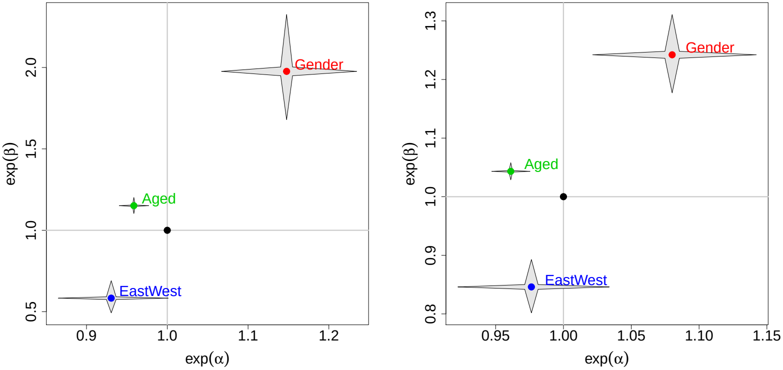
Fits of models with logistic link

	Deviance	df	Difference in deviances	df	*p*-value
Cumulative models					
Model with category-specific effects	7499.61	12192			
Location-shift model	7506.36	12204	6.75	12	0.873
Model with global effects	7544.60	12207	38.24	3	0.000
Adjacent categories models					
Model with category-specific effects	7500.77	12192			
Location-shift model	7508.72	12204	7.95	12	0.997
Model with global effects	7545.41	12207	36.69	3	0.000

- The full model with category-specific effects has 24 parameters, which reduces to 12 parameters in the location-shift model.

Application 3: Nuclear Energy

Illustration of Location and Dispersion Effects



- Cumulative model (left) and adjacent categories model (right)

Conclusion

The proposed location-shift model ...

- ... simultaneously accounts for location effects and dispersion effects (or tendencies to respond).
- ... enables an easy interpretation of effects in terms of log-odds.
- ... typically is sufficiently complex to approximate the underlying probability structure.
- ... is often a parsimonious alternative to the use of category-specific parameters.
- ... can be embedded into the framework of multivariate GLMs, which allows to use inference techniques and asymptotic results.

Main References

[1] Bender, R. , Grouven, U. (1998). Using binary logistic regression models for ordinal data with non proportional odds. J. Clin. Epidemiol. 51 (10), 809–816 .

[2] Berger, M. (2020). orddisp: Separating location and dispersion in ordinal regression models. R package version 2.1.1.

[3] Iannario, M., Piccolo, D. & Simone, R. (2015). CUB: A class of mixture models for ordinal data. R package version 1.1.3.

[4] McCullagh, P. (1980). Regression model for ordinal data (with discussion). J. R. Stat. Soc. B 42, 109–127.

[5] Rattinger, H., Roßteutscher, S., Schmitt-Beck, R., We els, B. & Wolf, C. (2014). Pre-election cross section (GLES 2013). GESIS Data Archive, Cologne ZA5700 Data file Version 2.0.0.

[6] Schauberger, G., Tutz, G. (2014). Catdata: Categorical Data. R package version 1.2.1

[7] Walker, R. W. (2016). On generalizing cumulative ordered regression models. J. Mod. Appl. Stat. Methods 15(2), 28.

[8] Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. Stata J. 6(1), 58.

[9] Williams, R. (2010). Fitting heterogeneous choice models with oglm. Stata J. 10 (4), 540–567 .