

serp: An R package for smoothing in ordinal regression

Ejike R. Ugba¹

¹ Department of Mathematics and Statistics, School of Economics and Social Sciences, Helmut Schmidt University, Hamburg, Germany

DOI: [10.21105/joss.03705](https://doi.org/10.21105/joss.03705)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: Hugo Gruson ↗

Reviewers:

- [@bernardsilenou](#)
- [@wesleyburr](#)

Submitted: 25 August 2021

Published: 26 October 2021

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

The use of specialized methods for the analysis of categorical data has been on the rise in recent years ([Agresti, 2010](#)). For instance, scientists frequently use the different forms of the ordinal model to analyze relationships between an ordinal response variable and covariates of interest. In particular, the cumulative link model (CLM) propagated by [McCullagh \(1980\)](#) finds a wide range of empirical applications in clinical trials, social surveys, market research, etc. However, in high-dimensional settings with a large number of unknown parameters, e.g., if several potential predictors and response categories are modeled, the so-called identification problem could be somewhat unavoidable ([Bartels, 1985](#); [Fisher, 1966](#)). Thankfully, regularization techniques (see, e.g., [Bühlmann & Van de Geer, 2011](#); [Hastie et al., 2009](#)) among other approaches, offer a remedy in such situations.

The `serp` software package ([Ugba, 2021](#)), available in the Comprehensive R Archive Network (CRAN) ([R Core Team, 2021](#)), implements a unique regularization algorithm that enforces smoothness on the adjacent categories of CLMs. The approach provides flexible modeling of CLMs that uses the smooth-effects-on-response penalty (SERP) discussed in [Ugba et al. \(2021\)](#) and [Tutz & Gertheiss \(2016\)](#). For instance, when applied to the general form of the cumulative logit model, also known as the non-proportional odds model (NPOM), one obtains the proportional odds model (POM) under extreme parameter shrinkage. Fitting in `serp` is carried out by a modified Newton's optimization method that facilitates an easy convergence and estimation speed. As an open-source software package, `serp` aims to provide a platform for advanced scientific modeling in empirical research. The core features of `serp`, as well as details about usage, are provided in the software package [documentation](#).

Statement of Need

Although a very useful tool in regression analysis, the development of regularization methods for categorical response models is still in its elementary stage. This explains why so many popular statistical software packages for ordinal regression do not currently implement such methods. The `'vglm'` function in the VGAM R package ([Yee, 2010](#); [Yee & Wild, 1996](#)), for instance, fits the cumulative family of the ordered model, with several link functions, but with no provision for regularization. The same is also true for the `'clm'` and `'polr'` functions in the `ordinal` ([Christensen, 2019b](#)) and MASS ([Venables & Ripley, 2002](#)) R packages, respectively. Other non-R functions for ordinal regression, such as the SAS CATMOD procedure ([SAS Institute Inc, 2018](#)) and the SPSS PLUM procedure ([IBM Corp., 2021](#)) also do not currently support regularization. `serp` fills this gap by providing a very unique and convenient means of regularizing estimates in ordinal models. The software package is intended for diverse forms of scientific applications. It is particularly suited for studies involving ordinal categorical outcomes. Users already conversant with existing R packages for ordinal regression will

by no means find it difficult to use `serp`. The package, moreover, comes with several useful functions that support empirical research while being also equipped with standard model performance and descriptive methods.

State of the Field

As previously noted, although a topic of intensive research in statistics for decades, regularization methods that are specifically designed for categorical data are relatively new (Ugba et al., 2021). So far, existing software packages implementing such techniques are mainly extensions of methods typically found in the high-dimensional linear or generalized linear model framework. Mostly, penalties in the tradition of lasso (Tibshirani, 1996), ridge (Hoerl & Kennard, 1970) and elastic net (Zou & Hastie, 2005) are implemented. The `glmnetcr` (Archer, 2014a) and `glmptcr` (Archer, 2014b) R packages, for instance, fit continuation ratio models with the elastic net penalty, while the `rms` R package (Harrell Jr, 2021) fits the cumulative logit model with quadratic (ridge regression) penalty. Also, the `ordinalgmifs` R package provided by Archer et al. (2014) implements a generalized monotone incremental forward stagewise (GMIFS) algorithm for regularized ordinal regression models, with a solution path akin to the L1 norm (lasso) penalty. Furthermore, the `ordinalNet` R package (Wurm et al., 2021) fits the elementwise link multinomial-ordinal (ELMO) class of models with the elastic net using a coordinate descent algorithm.

A common principle behind the regularization procedures with the highlighted software packages is to make the model's maximum likelihood estimates biased a bit towards zero or fully shrunk to zero, of course, with the so-called bias-variance trade-off in mind (Fortmann-Roe, 2012). However, the regularization method implemented in `serp` makes a paradigm shift from the traditional form of regularization where estimates are shrunk towards zero. `serp` provides a unique form of penalization that shrinks the category effects in the non-proportional cumulative link model towards global effects. Under extreme shrinkage, this form of penalization results in the proportional odds model. A somewhat similar form of smoothing for categorical response data is provided in the `mgcv` R package (Wood et al., 2016) but for un-ordered categories and with smoothing done in the linear predictors. By and large, regularization with `serp` provides a means of reducing model complexity without necessarily chopping off important features in the model.

A Minimal Example

The `wine dataset` from Randall (1989) available in `serp` software package represents the outcome of a factorial experiment on factors determining the bitterness of wine. Two treatment factors (temperature and contact) with two levels each are provided. Indeed, temperature (warm or cold) and contact (yes or no) between juice and skins can be controlled when crushing grapes during wine production. Nine judges each assessed wine from two bottles from each of the four treatment conditions, making a total of 72 observations. Rating of wine bitterness ranges from 1 = 'least bitter' to 5 = 'most bitter.' A typical modeling framework in this instance is an ordinal regression where the cumulative logit model is used to predict the scores of wine bitterness using the stated treatment factors (i.e., a simple case where all observations are aggregated over bottles and judges, compare e.g., the examples given in the `ordinal` R package manual (Christensen, 2019a)). However, the fitted NPOM (see, Table 1) using the `VGAM-vglm` function (Yee, 2010), for instance, is not fully identifiable. As observed, unbounded estimates with large standard errors ($SE > 10^3$) were obtained for covariate temperature with respect to the first and last response category, and also the last threshold (Intercept:4). In the actual sense, the absolute values of the estimates and the estimated standard errors for the unbounded parameters seem to diverge to ∞ as the

stopping criteria of the iterative fitting procedure used become stricter and the number of iterations allowed increases (see, e.g., Kosmidis, 2014). Other software implementations of the same model without regularization run into a similar problem, forcing users to adopt a different modeling approach which may or may not be appropriate for the data. However, by incorporating SERP into the modeling framework, a fully identified model with bounded estimates could be achieved. The realized coefficient paths for the estimated model via `serp`, using increasing values of the tuning parameter $[\lambda : 10^{-3}, 10^5]$, are shown in Figure 1.

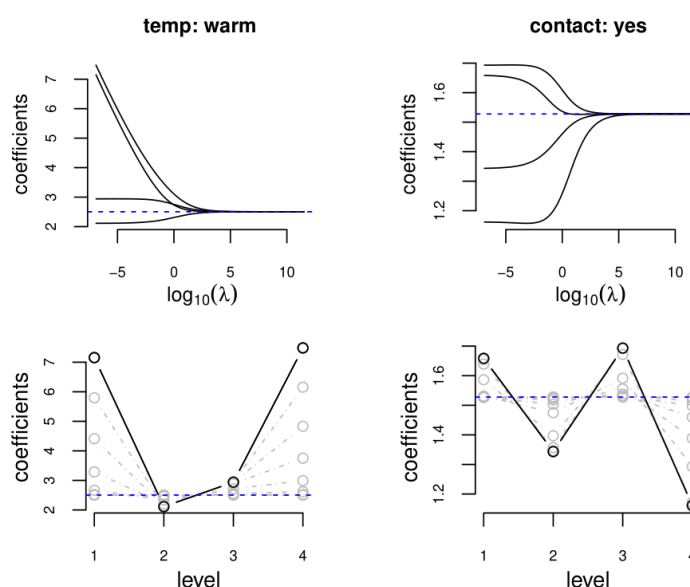


Figure 1: Estimated coefficient paths for the ordinal model of the wine data when using the smooth-effects-on-response penalty (SERP). The thick lines on the top displays are the category-specific coefficients associated with the two predictors, under increasing values of the tuning parameter (λ) on the range $(10^{-3}, 10^5)$. The dashed horizontal (blue) lines denote the parallel estimates. The bottom displays further illustrate SERP's smoothing steps from the category-specific to the parallel estimates, with the solid black, grey and dashed (blue) line strokes respectively the category-specific, the penalized and the parallel estimates.

Table 1: Estimates and standard errors (in parenthesis) of regression coefficients of the non-proportional odds model (NPOM) of the wine dataset, having temperature:warm (TW) and contact:yes (CY) as predictors, with estimates obtained using the `serp` and `vglm` R functions respectively penalized and non-penalized.

Coefficients	vglm	serp
(Intercept):1	1.226 (0.557)	1.344 (0.509)
(Intercept):2	-1.033 (0.481)	-1.251 (0.439)
(Intercept):3	-3.946 (0.902)	-3.467 (0.597)
(Intercept):4	-19.184 ($>10^3$)	-5.006 (0.729)
TW:1	19.243 ($>10^3$)	2.503 (0.532)
TW:2	2.111 (0.601)	2.503 (0.532)
TW:3	2.940 (0.828)	2.503 (0.532)
TW:4	17.064 ($>10^3$)	2.503 (0.532)
CY:1	1.659 (1.179)	1.528 (0.474)
CY:2	1.343 (0.583)	1.528 (0.474)
CY:3	1.693 (0.660)	1.528 (0.474)
CY:4	1.162 (0.905)	1.528 (0.474)

As observed, with extreme shrinkage all subject-specific estimates (including the unbounded estimates) get smoothed out to parallel estimates (see the dashed blue line in each panel). Thus, estimates between the parameter space of NPOM and POM inclusive could be determined via an appropriate tuning procedure. In this instance, a 5-fold cross-validated tuning with SERP resulted in POM (see Table 1) with no boundary estimates. The intercept parameters too, with no penalty on them, were all stabilized by the smoothing penalty on the remaining coefficients. This obviously, demonstrates the amount of flexibility that could be achieved with `serp` in fitting cumulative models in general. In addition to the ordinal logit model, `serp` also fits ordinal models with the probit, loglog, cloglog and cauchit link functions. A penalized partial proportional odds model (Peterson & Harrell, 1990) with the different link functions is also possible with `serp`.

Conclusion

The R add-on package `serp` contains functions for regularization/smoothing across response categories in the non-proportional cumulative ordinal regression model. Beyond the highlighted functionality, `serp` provides a collection of tools that promote stress-free modeling in empirical research. Moreover, standard function names and arguments already known to users familiar with related libraries are also used in `serp`, reducing unnecessary ambiguity. Lastly, details about usage and more elaborate examples are hosted online through a pkgdown (Wickham & Hesselberth, 2020) website on Github Pages.

Acknowledgements

The author would like to thank Jan Gertheiss for his invaluable suggestions and contributions towards the development of `serp`.

References

- Agresti, A. (2010). *Analysis of ordinal categorical data* (2nd ed.). John Wiley and Sons. <https://doi.org/10.1002/9780470594001>
- Archer, K. J. (2014a). *glmnetcr: Fit a penalized constrained continuation ratio model for predicting an ordinal response*. <https://cran.r-project.org/package=glmnetcr>
- Archer, K. J. (2014b). *glmpathcr: Fit a penalized continuation ratio model for predicting an ordinal response*. <https://cran.r-project.org/package=glmpathcr>
- Archer, K. J., Hou, J., Zhou, Q., Ferber, K., Layne, J. G., & Gentry, A. E. (2014). *ordinalgmifs: An R package for ordinal regression in highdimensional data settings*. *Cancer Informatics*, 13, 187–195. <https://doi.org/10.4137/CIN.S20806>
- Bartels, R. (1985). Identification in econometrics. *The American Statistician*, 39, 102–104. <https://doi.org/10.2307/2682804>
- Bühlmann, P., & Van de Geer, S. (2011). *Statistics for high-dimensional data: Methods, theory and applications*. Springer. <https://doi.org/10.1007/978-3-642-20192-9>
- Christensen, R. H. B. (2019a). *Cumulative link models for ordinal regression with the R package ordinal*. https://cran.r-project.org/web/packages/ordinal/vignettes/clm_article.pdf.
- Christensen, R. H. B. (2019b). *ordinal—regression models for ordinal data*. <https://CRAN.R-project.org/package=ordinal>

- Fisher, F. (1966). *The identification problem in econometrics*. McGraw-Hill.
- Fortmann-Roe, S. (2012). *Understanding the bias-variance tradeoff*. <http://scott.fortmann-roe.com/docs/BiasVariance.html>
- Harrell Jr, F. E. (2021). *rms: Regression modeling strategies*. <https://CRAN.R-project.org/package=rms>
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference and prediction* (2nd ed.). Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Hoerl, A., & Kennard, R. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12, 55–67. <https://doi.org/10.1080/00401706.2000.10485983>
- IBM Corp. (2021). *IBM SPSS statistics for windows* (Version 28.0) [Computer software]. <https://www.ibm.com/analytics/spss-statistics-software>
- Kosmidis, I. (2014). Bias in parametric estimation: Reduction and useful side-effects. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6, 185–196. <https://doi.org/10.1002/wics.1296>
- McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42, 109–142. <https://doi.org/10.1111/j.2517-6161.1980.tb01109.x>
- 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, 205–217. <https://doi.org/10.2307/2347760>
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Randall, J. (1989). The analysis of sensory data by generalised linear model. *Biometrics*, 31, 781–793. <https://doi.org/10.1002/bimj.4710310703>
- SAS Institute Inc. (2018). *SAS/STAT user's guide procedures*. SAS Institute Inc. <http://support.sas.com/documentation/onlinedoc/stat/indexproc.html>
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Tutz, G., & Gertheiss, J. (2016). Regularized regression for categorical data (with discussion and rejoinder). *Statistical Modelling*, 16, 161–260. <https://doi.org/10.1177/1471082X16642560>
- Ugba, E. R. (2021). *serp: Smooth effects on response penalty for CLM*. <https://CRAN.R-project.org/package=serp>
- Ugba, E. R., Mörlein, D., & Gertheiss, J. (2021). Smoothing in ordinal regression: An application to sensory data. *Stats*, 4, 616–633. <https://doi.org/10.3390/stats4030037>
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (Fourth). Springer. <https://www.stats.ox.ac.uk/pub/MASS4/>
- Wickham, H., & Hesselberth, J. (2020). *pkgdown: Make static HTML documentation for a package*. <https://CRAN.R-project.org/package=pkgdown>
- Wood, S. N., Pya, N., & Saefken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111, 1548–1575. <https://doi.org/10.1080/01621459.2016.1180986>

- Wurm, M. J., Rathouz, P. J., & Hanlon, B. M. (2021). Regularized ordinal regression and the ordinalNet R package. *Journal of Statistical Software*, 99, 1–42. <https://doi.org/10.18637/jss.v099.i06>
- Yee, T. W. (2010). The VGAM package for categorical data analysis. *Journal of Statistical Software*, 32, 1–34. <https://doi.org/10.18637/jss.v032.i10>
- Yee, T. W., & Wild, C. J. (1996). Vector generalized additive models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 481–493. <https://doi.org/10.1111/j.2517-6161.1996.tb02095.x>
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Methodological)*, 67, 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>