

Tutz, Gerhard; Berger, Moritz

Tree-structured scale effects in binary and ordinal regression. (English) Zbl 1475.62062
Stat. Comput. 31, No. 2, Paper No. 17, 12 p. (2021).

Summary: In binary and ordinal regression one can distinguish between a location component and a scaling component. While the former determines the location within the range of the response categories, the scaling indicates variance heterogeneity. In particular since it has been demonstrated that misleading effects can occur if one ignores the presence of a scaling component, it is important to account for potential scaling effects in the regression model, which is not possible in available recursive partitioning methods. The proposed recursive partitioning method yields two trees: one for the location and one for the scaling. They show in a simple interpretable way how variables interact to determine the binary or ordinal response. The developed algorithm controls for the global significance level and automatically selects the variables that have an impact on the response. The modeling approach is illustrated by several real-world applications.

MSC:

62-08 Computational methods for problems pertaining to statistics

Keywords:

recursive partitioning; tree-structured modeling; location-scale model; heterogeneity of variances; ordinal responses

Full Text: [DOI](#) [arXiv](#)

References:

- [1] Allison, PD, Comparing logit and probit coefficients across groups, *Sociol. Methods Res.*, 28, 2, 186-208 (1999) · [doi:10.1177/0049124199028002003](#)
- [2] Alvarez, RM; Brehm, J., American ambivalence towards abortion policy: development of a heteroskedastic probit model of competing values, *Am. J. Polit. Sci.*, 39, 1055-1079 (1995) · [doi:10.2307/2111669](#)
- [3] Archer, KJ, rpartordinal: an R package for deriving a classification tree for predicting an ordinal response, *J. Stat. Softw.*, 34, 7 (2010) · [doi:10.18637/jss.v034.i07](#)
- [4] Bender, R.; Grouven, U., Using binary logistic regression models for ordinal data with non-proportional odds, *J. Clin. Epidemiol.*, 51, 809-816 (1998) · [doi:10.1016/S0895-4356\(98\)00066-3](#)
- [5] Berger, M.; Tutz, G., Tree-structured clustering in fixed effects models, *J. Comput. Graph. Stat.*, 27, 2, 380-392 (2017) · [doi:10.1080/02664763.2017.1383370](#)
- [6] Breen, R.; Holm, A.; Karlson, KB, Correlations and nonlinear probability models, *Sociol. Methods Res.*, 43, 4, 571-605 (2014) · [doi:10.1177/0049124114544224](#)
- [7] Breiman, L.; Friedman, JH; Olshen, RA; Stone, JC, *Classification and Regression Trees* (1984), Monterey: Wadsworth, Monterey · [Zbl 0541.62042](#)
- [8] Coppersmith, D.; Hong, SJ; Hosking, JR, Partitioning nominal attributes in decision trees, *Data Min. Knowl. Discov.*, 3, 2, 197-217 (1999) · [doi:10.1023/A:1009869804967](#)
- [9] Cox, C., Location-scale cumulative odds models for ordinal data: a generalized non-linear model approach, *Stat. Med.*, 14, 11, 1191-1203 (1995) · [doi:10.1002/sim.4780141105](#)
- [10] Fisher, WD, On grouping for maximum homogeneity, *J. Am. Stat. Assoc.*, 53, 284, 789-798 (1958) · [Zbl 0084.35904](#) · [doi:10.1080/01621459.1958.10501479](#)
- [11] Fullerton, AS; Xu, J., The proportional odds with partial proportionality constraints model for ordinal response variables, *Soc. Sci. Res.*, 41, 1, 182-198 (2012) · [doi:10.1016/j.ssresearch.2011.09.003](#)
- [12] Galimberti, G.; Soffritti, G.; Maso, MD, Classification trees for ordinal responses in R: the rpartscore package, *J. Stat. Softw.*, 47, i10, 1-25 (2012)
- [13] Hauser, RM; Andrew, M., 1. Another look at the stratification of educational transitions: the logistic response model with partial proportionality constraints, *Sociol. Methodol.*, 36, 1, 1-26 (2006) · [doi:10.1111/j.1467-9531.2006.00172.x](#)
- [14] Hedeker, D.; Mermelstein, RJ; Demirtas, H., An application of a mixed-effects location scale model for analysis of ecological momentary assessment (ema) data, *Biometrics*, 64, 2, 627-634 (2008) · [Zbl 1137.62085](#) · [doi:10.1111/j.1541-0420.2007.00924.x](#)
- [15] Hedeker, D.; Demirtas, H.; Mermelstein, RJ, A mixed ordinal location scale model for analysis of ecological momentary

- assessment (ema) data, *Stat. Interface*, 2, 4, 391 (2009) · [Zbl 1245.62171](#) · [doi:10.4310/SII.2009.v2.n4.a1](#)
- [16] Hedeker, D.; Mermelstein, RJ; Demirtas, H., Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models, *Stat. Med.*, 31, 27, 3328-3336 (2012) · [doi:10.1002/sim.5338](#)
- [17] Hornung, R.: Ordinal forests. *J. Classif.* 37, 4-17 (2020) · [Zbl 07223586](#)
- [18] Hothorn, T.; Lausen, B., On the exact distribution of maximally selected rank statistics, *Comput. Stat. Data Anal.*, 43, 121-137 (2003) · [Zbl 1429.62542](#) · [doi:10.1016/S0167-9473\(02\)00225-6](#)
- [19] Hothorn, T.; Hornik, K.; Zeileis, A., Unbiased recursive partitioning: a conditional inference framework, *J. Comput. Graph. Stat.*, 15, 651-674 (2006) · [doi:10.1198/106186006X133933](#)
- [20] Ishwaran, H.; Gatsonis, CA, A general class of hierarchical ordinal regression models with applications to correlated roc analysis, *Can. J. Stat.*, 28, 4, 731-750 (2000) · [Zbl 0966.62046](#) · [doi:10.2307/3315913](#)
- [21] Janitza, S.; Tutz, G.; Boulesteix, AL, Random forests for ordinal responses: prediction and variable selection, *Comput. Stat. Data Anal.*, 96, 57-73 (2016) · [Zbl 1468.62089](#) · [doi:10.1016/j.csda.2015.10.005](#)
- [22] Karlson, KB; Holm, A.; Breen, R., Comparing regression coefficients between same-sample nested models using logit and probit: a new method, *Sociol. Methodol.*, 42, 1, 286-313 (2012) · [doi:10.1177/0081175012444861](#)
- [23] Loh, WY, Fifty years of classification and regression trees, *Int. Stat. Rev.*, 82, 3, 329-348 (2014) · [Zbl 1416.62347](#) · [doi:10.1111/insr.12016](#)
- [24] Long, J.S., Allison, P.D., McGinnis, R.: Rank advancement in academic careers: sex differences and the effects of productivity. *Am. Sociol. Rev.* 58(5), 703-722 (1993)
- [25] McCullagh, P., Regression model for ordinal data (with discussion), *J. R. Stat. Soc. B*, 42, 109-127 (1980) · [Zbl 0483.62056](#)
- [26] Mood, C., Logistic regression: why we cannot do what we think we can do, and what we can do about it, *Eur. Sociol. Rev.*, 26, 1, 67-82 (2010) · [doi:10.1093/esr/jcp006](#)
- [27] Murphy, AH, The ranked probability score and the probability score: a comparison, *Weather*, 81, 82 (1970)
- [28] Piccarreta, R., Classification trees for ordinal variables, *Comput. Stat.*, 23, 3, 407-427 (2008) · [Zbl 1223.62106](#) · [doi:10.1007/s00180-007-0077-5](#)
- [29] Ripley, BD, *Pattern Recognition and Neural Networks* (1996), Cambridge: Cambridge University Press, Cambridge · [doi:10.1017/CBO978051181265](#)
- [30] Rohwer, G., A note on the heterogeneous choice model, *Sociol. Methods Res.*, 44, 1, 145-148 (2015) · [doi:10.1177/0049124114552750](#)
- [31] Shih, YS, A note on split selection bias in classification trees, *Comput. Stat. Data Anal.*, 45, 457-466 (2004) · [Zbl 1429.62264](#) · [doi:10.1016/S0167-9473\(03\)00064-1](#)
- [32] Shih, Y.S., Tsai, H.: Variable selection bias in regression trees with constant fits. *Comput. Stat. Data Anal.* 45, 595-607 (2004) · [Zbl 1429.62725](#)
- [33] Strobl, C.; Boulesteix, AL; Augustin, T., Unbiased split selection for classification trees based on the gini index, *Comput. Stat. Data Anal.*, 52, 483-501 (2007) · [Zbl 1452.62469](#) · [doi:10.1016/j.csda.2006.12.030](#)
- [34] Strobl, C.; Malley, J.; Tutz, G., An introduction to recursive partitioning: rationale, application and characteristics of classification and regression trees, bagging and random forests, *Psychol. Methods*, 14, 323-348 (2009) · [doi:10.1037/a0016973](#)
- [35] Tutz, G., Binary response models with underlying heterogeneity: identification and interpretation of effects, *Eur. Sociol. Rev.*, 34, 211-221 (2018) · [doi:10.1093/esr/jcy001](#)
- [36] Tutz, G.; Berger, M., Separating location and dispersion in ordinal regression models, *Econom. Stat.*, 2, 131-148 (2017)
- [37] Tutz, G.; Berger, M., Tree-structured modelling of categorical predictors in generalized additive regression, *Adv. Data Anal. Classif.*, 12, 737-758 (2018) · [Zbl 1416.62364](#) · [doi:10.1007/s11634-017-0298-6](#)
- [38] Williams, R., Using heterogeneous choice models to compare logit and probit coefficients across groups, *Sociol. Methods Res.*, 37, 4, 531-559 (2009) · [doi:10.1177/0049124109335735](#)
- [39] Williams, R., Fitting heterogeneous choice models with oglm, *Stata J.*, 10, 4, 540 (2010) · [doi:10.1177/1536867X1101000402](#)
- [40] Wright, MN; König, IR, Splitting on categorical predictors in random forests, *PeerJ*, 7, e6339 (2019) · [doi:10.7717/peerj.6339](#)
- [41] Zeileis, A.; Hothorn, T.; Hornik, K., Model-based recursive partitioning, *J. Comput. Graph. Stat.*, 17, 2, 492-514 (2008) · [doi:10.1198/106186008X319331](#)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.