Fixed effect regression for clustered data with tree-structured modeling of unit-specific slopes

Modeling longitudinal or cross-sectional data that come in clusters (i.e. data with more than one observation from the same unit) requires accounting for the heterogeneity between the units to avoid biased estimates and misleading standard errors. In regression analysis, the most widely applied approach to account for this unobserved heterogeneity is mixed effects modeling. The classical linear mixed effects model comprises fixed effects of a set of covariates and random effects for the units that are commonly assumed to follow a normal distribution. To avoid the normality assumption of the random effects and to detect clusters of units that share the same effect, Berger and Tutz (2018) proposed a fixed effect model that parsimoniously accounts for between-unit heterogeneity using tree-based splits. Their model allows for a unspecified, discrete distribution of the unit-specific intercepts and was recently extended to enable tree-structured modeling of covariate effects by Spuck et al. (2026). However, these works focus on differences in the intercept only and do not consider potential differences in slopes, which may be too restrictive in real-world applications.

The proposed project aims to extend the modeling approaches in Berger and Tutz (2018) and Spuck et al. (2026) by developing a fixed effect model, where, in addition to the tree-structured modeling of unit-specific intercepts, the linear effects of the covariates may vary between units and are also determined by a tree. The proposed model can be fitted within the framework of tree-structured varying coefficient models and allows for a flexible modeling of clustered data that is able to capture heterogeneity between different units in a data-driven way (Berger et al., 2019). This extension conceptually reflects the shift from random intercept to random slope models in mixed effects regression.

References

M. Berger and G. Tutz. Tree-structured clustering in fixed effects models. *J Comput Graph Stat*, 27:380–392, 2018. doi: 10.1080/10618600.2017.1371030.

M. Berger, G. Tutz, and M. Schmid. Tree-structured modelling of varying coefficients. *Stat Comput*, 29:217–229, 2019. doi: 10.1007/s11222-018-9804-8.

N. Spuck, M. Schmid, and M. Berger. Flexible tree-structured regression for clustered data with an application to quality of life in older adults. *Advances in Data Analysis and Classification*, 2026.