Boosting Actuarial Regression Models

Sinan Acemoglu* **Christian Kleiber§**

Preliminary version. Please do not cite or circulate.

Abstract

Machine learning tools are increasingly adopted in the actuarial community in recent years. We explore the usefulness of boosting algorithms for model selection and model fitting in classical problems of non-life insurance. A major advantage of boosting algorithms is that they lead to interpretable models, unlike many other machine learning tools. Also, in the setting of parametric regression models, boosting may be viewed as a regularized version of model fitting via maximum likelihood, with the number of iterations serving as a regularization parameter. Regularization then leads to shrunken parameter estimates, not unlike methods such as LASSO. More specifically, we consider some common regression models for claim incidence, claim frequency and claim severity via gradient boosting of models that are generalized linear models (GLMs) or closely related extensions thereof. Count data models with zero-adjustments receive particular attention, notably hurdle-type models. Computationally, we make use of recent extensions to the `mboost` package for R, among them a new R package named `countreg` that is currently available from the R-Forge development server. An example involving car insurance is used for empirical illustrations.

*University of Basel, Faculty of Business and Economics, Switzerland; E-Mail: `sinan.acemoglu@unibas.ch`.
§University of Basel, Faculty of Business and Economics, Switzerland; E-Mail: `christian.kleiber@unibas.ch`.