Claim incidence models with varying exposure in automobile insurance

Linda E. Walter[1]
[1]Department of Business and Economics, Universität Basel,
4002 Basel, Switzerland, linda.walter@unibas.ch

We study claim incidence in a unique Swiss motor third-party liability insurance data set. We apply extended generalized linear models (GLM) with noncanonical link functions to account for insurance-specific features such as varying exposure times and changes in the contracts over the observation period resulting in multiple observations per contract. An exposure-adjusted logit link function is used to model the relationship between the probability of reporting a claim and several covariates. We employ cluster-robust standard errors as the number of observations per contract can vary greatly. Among other findings, our results demonstrate that leasing a car strongly influences the probability of reporting a claim, whereas the gender of the most frequent driver does not seem to be relevant. We show that including contracts with varying exposure times changes the results substantially, whereas tracking changes in a contract seems less important. We provide evidence that these results are robust to different choices of the link function.

Key Words: Generalized linear models, noncanonical link functions, pseudo-maximum likelihood, cluster-robust standard errors.

1. Introduction

The claim incidence is analysed using a generalized linear model (GLM) with a non-canonical link function. Certain characteristics that are typical features of insurance data do not permit the use of standard link functions to model the probability of reporting a claim. In particular, the time a person is exposed to risk may differ between the contracts. In addition, and in contrast to the existing literature, we allow risk characteristics to change over the year. Thus, there may be more than one observation available per contract. In order to handle these insurance-specific features, we use a logit link function adjusted by the exposure period as proposed by Heller et al. (2006) and De Jong and Heller (2008) and make use of cluster-robust standard errors. We provide evidence that results differ if only contracts with equal exposure time are included in the analysis.

2. Data

The analysis is conducted using a unique representative data set of a Swiss insurance company, which consists of more than 150,000 observations. The considered claim type is a motor third-party liability claim. We concentrate on one calendar year. A large amount of information is available and has been included in the analysis. In particular, characteristics on the most frequent driver, the car and the contract may be of importance when modelling the probability of reporting a claim.

3. Model

A straightforward approach for binary dependent variables is the application of a generalized linear model (GLM) with a logit link function. However, an adjustment has to be made to enable us to account for the different exposure times of the observations. The modified exposure-adjusted link function is given by (Heller et al., 2006)

$$g(p_{ij}^*) = \ln\left[\frac{p_{ij}^*/t}{1-p_{ij}^*/t}\right] = x_{ij}^T\beta,$$

where p_{ij}^* denotes the probability of reporting a claim in a specific time period t, which is expressed as a proportion such that $0 < t \leq 1$. The parameters are estimated by pseudo-maximum likelihood. Following Binder (1983), we employ cluster-robust standard errors to account for unbalanced clusters.

4. Results

The following risk factors are found to be statistically significant at the 1% significance level: the age of the vehicle, the number of kilometres per year, the use of the car (private, private with journey to work or business use), leasing, the age and residence permit of the most frequent driver, his/her canton of residence, and a deductible or an additional insurance coverage if one exists.

Receiver operating characteristic (ROC) analysis shows that the performance of the exposure-adjusted logit model is statistically significantly better than the performance of a common logit model without exposure adjustment (p-value of Delong AUC test $<10^{-5}$). For the latter, only contracts with an exposure of exactly one year are used for the analysis. Thus, we cannot conclude that contracts with an exposure of less than one year are similar to contracts with exposure periods of exactly one year.

However, the cluster correction only slightly improves the performance. Thus, the common approach not to account for changes in the contract and to observe them only on the first day of risk exposure leads to similar results.

These results are robust to different link functions as well as to different forms of inclusion of the exposure term.

5. Conclusions

Accounting for the variation in the exposure time is common in modelling count data, but there is little knowledge of how to include different periods of exposure in binary response models. We provide several approaches and show for a unique insurance portfolio that accounting for varying exposure times in binary models is essential because neglecting them may lead to biased results.

References

Binder, D.A. (1983) “On the variances of asymptotically normal estimators from complex surveys,” *International Statistical Review/Revue Internationale de Statistique,* 51, 279–292.

De Jong, P. and Heller G.Z. (2008) *Generalized linear models for insurance data*, Cambridge University Press.

Heller, G., Stasinopoulos M. and Rigby B. (2006) “The zero-adjusted Inverse Gaussian distribution as a model for insurance claims,” in *Proceedings of the 21th International Workshop on Statistical Modelling*, pp. 226–233.