Preface

Generalized Linear Model (GLM) is a popular statistical tool now used by insurance companies. The model is most often applied in risk assessment for short-term non-life insurance schemes generating mass risk portfolios and in loss reserve prediction. The two groups of problems are now treated as disciplines of actuarial sciences which are referred to in literature as *ratemaking* and *loss (claim) reserving*. Due to the insurance data specificity and considering the progress made in computational techniques as well as the growing amount of information gathered by insurers, different kinds of modifications of GLMs are now in practical use. Zero-inflated General Poisson claim frequency, overdispersion or heavy-tailed empirical claim severity distribution are the examples here.

This book aims to present program **R** and its capabilities in ratemaking and loss reserving. There are a few good titles, which addressed similar topics and the **R** code is partly provided, e.g. [16] or [35]. This book presents the models, complementing the models in these works (hence the title "selected"). Only the first chapter may contain elements that are similar, but it is indispensable for those readers who are not proficient in ratemaking or loss reserving. The characteristic feature of selected models is that they allow to capture different dependencies by using random effects or copulas. Since the GLMs are greatly extended by adding one or more sets of random effects on the same linear scale, it allows to use likelihood-based inferences based on the *h-likelihood*. In this case the estimation method does not require the use of numerical integration as in marginal likelihood, and neither are prior probabilities required as in bayesian approach.

The practical use of more and more complex statistical models creates the main problem of how to estimate their parameters. Usually, such estimation is possible only if numerical algorithms and suitable computational

techniques are applied. Numerous implementations of models dispersed throughout various packages can be found in program **R**. The functions generally have a similar, intuitive structure but they differ from each other when it comes to details. It is quite often the case that the image of what a given function produces and what exactly the estimated quantity is may differ from the actual shape of the implemented model. Therefore, the Authors wish to show the model theoretical fundamentals with respect to a specific function of program **R** on the one hand, and to present the function application in ratemaking and loss reserving on the other. Each issue of the book consists of the following parts:

- theoretical introduction - ratemaking and loss reserving
- theoretical introduction and **R** function - general statistical models
- models and empirical examples in **R**

The first chapter of the book provides a brief overview over the GLMs and their modifications used in insurance. After that Tweedie family of distributions applicable in the modeling of claim severity is considered in detail. Next a variety of models for the count variable, which is the number of claims are presented. The chapter ends considerations about the goodness of fit of the model (Akaik Information Critera (AIC), Bayesian Information Critera (BIC), deviance) and the method of comparing models (cross-validation, bootstrapping). The second chapter presents a class of models which is not popularized in the actuarial literature and which is referred to as Hierarchical Generalized Linear Model (HGLM) that make it possible to take account of dependency occurring between random variables under consideration. The dependency can be both structural and empirical. Except for mere mentions, this issue is not widely discussed in literature. In the third part, copulas, which are well-known and extensively described in the actuarial literature, are applied. However, this book presents their practical use in ratemaking. The last chapter discusses the methods of loss reserving, but only this which are connected with the GLM. As it is popular to use bootstrapping in the estimation of error of prediction of loss reserve this problem is considered as well. The bootstrap estimator of error of prediction is proposed.

The book is intended for BSc, MSc and PhD students. It supplements the knowledge gained in subjects/courses concerning Non-Life Insurance: Statistical Techniques. It may also prove useful in preparation for actuarial examinations. A strong emphasis is laid on practical applications with the use of **R**. Therefore, the second circle of readers that this book is addressed to are practitioners working in different departments of a non-life insurance company.

The readers have to be familiar with the basic course in mathematical statistics including the methodology concerning random variables and their distributions, the maximum likelihood estimation method and linear models. It is also necessary the knowledge of basic theory of insurance issues. Extensive consideration of ratemaking and GLMs models can be found in [46], [55], [14], [19], [79], [35], whereas issues related to loss reserving methods using GLMs models are discussed in detail in [95]. One of the modifications is the change in the assumption concerning the response variable in the GLM model. A wide range of different distributions has so far been presented by many researchers, e.g. [22], [45], [3]. Another modification is taking account of relationships between the risk portfolio individual risks. The relations can be mapped by introducing random effects into GLM models. The use of mixed fixed and random effects models in ratemaking and loss reserving is described for example in [76], [2], [1], [7], [36], [42], [43]. An alternative method of modelling the relationships is the copula function. The application of copula-based models in insurance is the subject of the following works: [37], [18], [60].

Moreover, readers should have basic skills required to use **R**, such as software package installation, data downloading, use of functions/facilities included in packages. The book contains a supplement which is the web-based Git repository [1]

https://github.com/woali/RatemakingLossReserve

with **R** codes for all examples prepared in this book. Most of the code is presented and described in the text, but in case of some complex calcu-

[1]Publication was co-financed from the funds granted to the Faculty of Finance and Law at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential.

lation the code is only in the electronic version. It regards in particular codes generating figures. We believe that the analysis and applications of available **R** codes help readers in understanding the subject. But nothing will replace own solving a problem in **R**.

Bibliography

[1] ANTONIO, K., AND BEIRLANT, J. Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and Economics 40*, 1 (2007), 58–76.

[2] ANTONIO, K., BEIRLANT, J., HOEDEMAKERS, T., AND VERLAAK, R. Lognormal mixed models for reported claims reserves. *North American Actuarial Journal 10*, 1 (2006), 30–48.

[3] ANTONIO, K., AND VALDEZ, E. A. Statistical concepts of a priori and a posteriori risk classification in insurance. *AStA Advances in Statistical Analysis 96*, 2 (2012), 187–224.

[4] AUGUIE, B. *gridExtra: Miscellaneous Functions for "Grid" Graphics*, 2016. R package version 2.2.1.

[5] BATES, D., MÄCHLER, M., BOLKER, B., AND WALKER, S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*, 1 (2015), 1–48.

[6] BJØRNSTAD, J. F. On the generalization of the likelihood function and the likelihood principle. *Journal of the American Statistical Association 91*, 434 (1996), 791–806.

[7] BOUCHER, J., AND DENUIT, M. Fixed versus random effects in poisson regression models for claim counts: A case study with motor insurance. *Astin Bulletin 36*, 1 (2006), 285.

[8] BOUCHER, J.-P., DENUIT, M., AND GUILLÉN, M. Risk classification for claim counts: A comparative analysis of various zero-inflated mixed poisson and hurdle models. *North American Actuarial Journal 11*, 4 (2007), 110–131.

[9] BOUCHER, J.-P., DENUIT, M., AND GUILLÉN, M. Number of accidents or number of claims? an approach with zero-inflated poisson models for panel data. *Journal of Risk and Insurance 76*, 4 (2009), 821–846.

[10] BOUCHER, J.-P., AND GUILLÉN, M. A survey on models for panel count data with applications to insurance. *RACSAM-Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas 103*, 2 (2009), 277–294.

[11] BRESLOW, N. E., AND CLAYTON, D. G. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association 88*, 421 (1993), 9–25.

[12] BROSTRĂ¶M, G. *glmmML: Generalized linear models with clustering*, 2013. R package version 1.0.

[13] BROSTRÖM, G., AND HOLMBERG, H. Generalized linear models with clustered data: Fixed and random effects models. *Computational Statistics & Data Analysis 55*, 12 (2011), 3123–3134.

[14] BÜHLMANN, H., AND GISLER, A. *A course in credibility theory and its applications*. Springer, 2005.

[15] BURNECKI, K., AND NOWICKA-ZAGRAJEK, J. Składka kwantylowa w modelu ryzyka kolektywnego a dane szkodowe z obcieciem dolnym. *Ostasiewicz (red.), Statystyka aktuarialna-stan i perspektywy rozwoju z Polsce, Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, 1108 (2006).

[16] CHARPENTIER, A. *Computational Actuarial Science with R*. CRC Press, 2014.

[17] CONSUL, P., AND FAMOYE, F. Generalized poisson regression model. *Communications in Statistics-Theory and Methods 21*, 1 (1992), 89–109.

[18] CZADO, C., KASTENMEIER, R., BRECHMANN, E. C., AND MIN, A. A mixed copula model for insurance claims and claim sizes. *Scandinavian Actuarial Journal 2012*, 4 (2012), 278–305.

[19] DE JONG, P., AND HELLER, G. Z. *Generalized linear models for insurance data*. Cambridge University Press, 2008.

[20] DELIGNETTE-MULLER, M. L., AND DUTANG, C. fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software 64*, 4 (2015), 1–34.

[21] DENUIT, M., DHAENE, J., GOOVAERTS, M., AND KAAS, R. *Actuarial theory for dependent risks: measures, orders and models*. John Wiley & Sons, 2006.

[22] DENUIT, M., MARÉCHAL, X., PITREBOIS, S., AND WALHIN, J.-F. *Actuarial modelling of claim counts: Risk classification, credibility and bonus-malus systems*. Wiley. com, 2007.

[23] DHAENE, J., DENUIT, M., GOOVAERTS, M., KAAS, R., AND VYNCKE, D. The concept of comonotonicity in actuarial science and finance: applications. *Insurance: Mathematics and Economics 31*, 2 (2002), 133 – 161.

[24] DHAENE, J., DENUIT, M., GOOVAERTS, M., KAAS, R., AND VYNCKE, D. The concept of comonotonicity in actuarial science and finance: theory. *Insurance: Mathematics and Economics 31*, 1 (2002), 3 – 33. Special Issue: Papers presented at the 5th {IME} Conference, Penn State University, University Park, PA, 23-25 July 2001.

[25] DIONNE, G., AND VANASSE, C. A generalization of automobile insurance rating models: the negative binomial distribution with a regression component. *Astin Bulletin 19*, 2 (1989), 199–212.

[26] D'ORAZIO, M. *StatMatch: Statistical Matching*, 2016. R package version 1.2.4.

[27] DUNN, P. K. *tweedie: Tweedie exponential family models*, 2014. R package version 2.2.1.

[28] DUNN, P. K., AND SMYTH, G. K. *dglm: Double Generalized Linear Models*, 2014. R package version 1.8.1.

[29] EFRON, B., AND TIBSHIRANI, R. J. *An introduction to the bootstrap*. CRC press, 1994.

[30] ENGLAND, P., AND VERRALL, R. Analytic and bootstrap estimates of prediction errors in claims reserving. *Insurance: mathematics and economics 25*, 3 (1999), 281–293.

[31] ESTRADA, E. G., AND ALVA, J. A. V. *gPdtest: Bootstrap goodness-of-fit test for the generalized Pareto distribution*, 2012. R package version 0.4.

[32] FAMOYE, F., AND SINGH, K. P. Zero-inflated generalized poisson regression model with an application to domestic violence data. *Journal of Data Science 4*, 1 (2006), 117–130.

[33] FAMOYE, F., WULU, J., AND SINGH, K. P. On the generalized poisson regression model with an application to accident data. *Journal of Data Science 2*, 2004 (2004), 287–295.

[34] FARAWAY, J. *faraway: Functions and datasets for books by Julian Faraway.*, 2014. R package version 1.0.6.

[35] FREES, E. W. *Regression modeling with actuarial and financial applications*. Cambridge University Press, 2009.

[36] FREES, E. W., DERRIG, R. A., AND MEYERS, G. *Predictive modeling applications in actuarial science*. Cambridge University Press, 2014.

[37] FREES, E. W. J., MEYERS, G., AND CUMMINGS, A. D. Dependent multi-peril ratemaking models. *Astin Bulletin 40*, 02 (2010), 699–726.

[38] GENEST, C., HUANG, W., AND DUFOUR, J.-M. A regularized goodness-of-fit test for copulas. *Journal de la Société Française de Statistique 154*, 1 (2013), 64–77.

[39] GENEST, C., RĂ©MILLARD, B., AND BEAUDOIN, D. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics 44*, 2 (2009), 199 – 213.

[40] GESMANN, M., MURPHY, D., ZHANG, Y. W., CARRATO, A., CRUPI, G., WUTHRICH, M., AND CONCINA, F. *ChainLadder: Statistical Methods and Models for Claims Reserving in General Insurance*, 2015. R package version 0.2.2.

[41] GHITANY, M., KARLIS, D., AL-MUTAIRI, D., AND AL-AWADHI, F. An em algorithm for multivariate mixed poisson regression models and its application. *Applied Mathematical Sciences 6*, 137 (2012), 6843–6856.

[42] GIGANTE, P., PICECH, L., AND SIGALOTTI, L. Claims reserving in the hierarchical generalized linear model framework. *Insurance: Mathematics and Economics 52*, 2 (2013), 381–390.

[43] GIGANTE, P., PICECH, L., AND SIGALOTTI, L. Prediction error for credible claims reserves: an h-likelihood approach. *European Actuarial Journal 3*, 2 (2013), 453–470.

[44] GILLELAND, E., AND KATZ, R. W. extRemes 2.0: An extreme value analysis package in R. *Journal of Statistical Software 72*, 8 (2016), 1–39.

[45] GRAY, R. J., AND PITTS, S. M. *Risk Modelling in General Insurance: From Principles to Practice*. Cambridge University Press, 2012.

[46] HABERMAN, S., AND RENSHAW, A. E. Generalized linear models and actuarial science. *The Statistician* (1996), 407–436.

[47] HALLIN, M., AND INGENBLEEK, J.-F. The swedish automobile portfolio in 1977: A statistical study. *Scandinavian actuarial journal 1983*, 1 (1983), 49–64.

[48] HARDY, M. R. An introduction to risk measures for actuarial applications. *SOA Syllabus Study Note* (2006).

[49] HOFERT, M., KOJADINOVIC, I., MAECHLER, M., AND YAN, J. copula: Multivariate dependence with copulas. *R package version 0.999-7* (2013).

[50] IWANIK, J., AND NOWICKA-ZAGRAJEK, J. *Premiums in the Individual and Collective Risk Models*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 407–426.

[51] J, L. Plotrix: a package in the red light district of r. *R-News 6*, 4 (2006), 8–12.

[52] JOE, H., AND ZHU, R. Generalized poisson distribution: the property of mixture of poisson and comparison with negative binomial distribution. *Biometrical Journal 47*, 2 (2005), 219–229.

[53] JØRGENSEN, B. Some properties of exponential dispersion models. *Scandinavian Journal of Statistics* (1986), 187–197.

[54] JORGENSEN, B. Exponential dispersion models. *Journal of the Royal Statistical Society. Series B (Methodological)* (1987), 127–162.

[55] JØRGENSEN, B., AND PAES DE SOUZA, M. C. Fitting tweedie's compound poisson model to insurance claims data. *Scandinavian Actuarial Journal 1994*, 1 (1994), 69–93.

[56] KARLIS, D. A general em approach for maximum likelihood estimation in mixed poisson regression models. *Statistical Modelling 1*, 4 (2001), 305–318.

[57] KARLIS, D. Em algorithm for mixed poisson and other discrete distributions. *Astin bulletin 35*, 1 (2005), 3–24.

[58] KOJADINOVIC, I., YAN, J., ET AL. Modeling multivariate distributions with continuous margins using the copula r package. *Journal of Statistical Software 34*, 9 (2010), 1–20.

[59] KRAEMER, N., BRECHMANN, E., SILVESTRINI, D., AND CZADO, C. Total loss estimation using copula-based regression models. *Insurance: Mathematics and Economics 53* (2013), 829 – 839.

[60] KRÄMER, N., BRECHMANN, E. C., SILVESTRINI, D., AND CZADO, C. Total loss estimation using copula-based regression models. *Insurance: Mathematics and Economics 53*, 3 (2013), 829–839.

[61] KREMER, E. Ibnr-claims and the two-way model of anova. *Scandinavian Actuarial Journal 1982*, 1 (1982), 47–55.

[62] LAIRD, N. M., AND WARE, J. H. Random-effects models for longitudinal data. *Biometrics* (1982), 963–974.

[63] LAMBERT, D. Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics 34*, 1 (1992), 1–14.

[64] LAWLESS, J. F. Negative binomial and mixed poisson regression. *Canadian Journal of Statistics 15*, 3 (1987), 209–225.

[65] LEE, Y., AND NELDER, J. Extended-reml estimators. *Journal of Applied Statistics 30*, 8 (2003), 845–856.

[66] LEE, Y., AND NELDER, J. A. Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), 619–678.

[67] LEE, Y., AND NELDER, J. A. Hierarchical generalised linear models: a synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika 88*, 4 (2001), 987–1006.

[68] LEE, Y., NELDER, J. A., AND PAWITAN, Y. *Generalized linear models with random effects: unified analysis via H-likelihood*. CRC Press, 2006.

[69] LINDSEY, J. K. *Applying generalized linear models*. Springer, 1997.

[70] MACK, T. A simple parametric model for rating automobile insurance or estimating ibnr claims reserves. *Astin Bulletin 21*, 1 (1991), 93–109.

[71] MACK, T. Which stochastic model is underlying the chain ladder method? *Insurance: mathematics and economics 15*, 2 (1994), 133–138.

[72] MACK, T., AND VENTER, G. A comparison of stochastic models that reproduce chain ladder reserve estimates. *Insurance: Mathematics and Economics 26*, 1 (2000), 101–107.

[73] MCCULLAGH, P., AND NELDER, J. Generalized linear models. second edition. *Monographs on statistics and applied probability.* (1989).

[74] MEVIK, B.-H., WEHRENS, R., AND LILAND, K. H. *pls: Partial Least Squares and Principal Component Regression*, 2015. R package version 2.5-0.

[75] MOLENBERGHS, G., AND VERBEKE, G. *Models for discrete longitudinal data*. Springer, 2005.

[76] NELDER, J., AND VERRALL, R. Credibility theory and generalized linear models. *Astin Bulletin 27* (1997), 71–82.

[77] NELDER, J., AND WEDDERBURN, W. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General) 135(3)* (1972), 370–384.

[78] NELSEN, R. B. *An introduction to copulas*. Springer Science & Business Media, 2007.

[79] OHLSSON, E., AND JOHANSSON, B. *Non-life insurance pricing with generalized linear models*. Springer, 2010.

[80] PINHEIRO, J. C., AND BATES, D. M. *Linear mixed-effects models: basic concepts and examples*. Springer, 2000.

[81] PITSELIS, G. Quantile credibility models. *Insurance: Mathematics and Economics 52*, 3 (2013), 477 – 489.

[82] R CORE TEAM. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.

[83] RENSHAW, A. E., AND VERRALL, R. J. A stochastic model underlying the chain-ladder technique. *British Actuarial Journal 4*, 04 (1998), 903–923.

[84] RÖNNEGÅRD, L., SHEN, X., AND ALAM, M. hglm: A package for fitting hierarchical generalized linear models. *The R Journal 2*, 2 (2010), 20–28.

[85] RONNEGARD, L., SHEN, X., AND ALAM, M. hglm: A package for fitting hierarchical generalized linear models. *The R Journal 2*, 2 (2010), 20–28.

[86] RUOYAN, M. *Estimation of dispersion parameters in GLMs with and without random effects*. Citeseer, 2004.

[87] SARKAR, D. *Lattice: Multivariate Data Visualization with R*. Springer, New York, 2008. ISBN 978-0-387-75968-5.

[88] SCHEPSMEIER, U., STOEBER, J., BRECHMANN, E. C., AND GRAELER, B. *VineCopula: Statistical Inference of Vine Copulas*, 2015. R package version 1.4.

[89] SMYTH, G., HU, Y., DUNN, P., PHIPSON, B., AND CHEN, Y. *statmod: Statistical Modeling*, 2015. R package version 1.4.21.

[90] SMYTH, G. K. Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society. Series B (Methodological)* (1989), 47–60.

[91] STASINOPOULOS, M., WITH CONTRIBUTIONS FROM CALLIOPE AKANTZILIOTOU, B. R., HELLER, G., OSPINA, R., MOTPAN, N., MCELDUFF, F., VOUDOURIS, V., DJENNAD, M., ENEA, M., AND GHALANOS., A. *gamlss.dist: Distributions to be Used for GAMLSS Modelling*, 2015. R package version 4.3-5.

[92] TWEEDIE, M. An index which distinguishes between some important exponential families. In *Statistics: Applications and New Directions: Proc. Indian Statistical Institute Golden Jubilee International Conference* (1984), pp. 579–604.

[93] VENABLES, W. N., AND RIPLEY, B. D. *Modern Applied Statistics with S*, fourth ed. Springer, New York, 2002. ISBN 0-387-95457-0.

[94] VERBEKE, G., AND LESAFFRE, E. A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association 91*, 433 (1996), 217–221.

[95] VERRALL, R. J. An investigation into stochastic claims reserving models and the chain-ladder technique. *Insurance: Mathematics and Economics 26*, 1 (2000), 91–99.

[96] WANAT, S. Modele zależności w agregacji ryzyka ubezpieczyciela. *Zeszyty Naukowe/Uniwersytet Ekonomiczny w Krakowie. Seria Specjalna, Monografie*, 211 (2012).

[97] WICKHAM, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.

[98] WINKELMANN, R. *Econometric analysis of count data*. Springer, 2003.

[99] WOLFINGER, R., AND O'CONNELL, M. Generalized linear mixed models a pseudo-likelihood approach. *Journal of statistical Computation and Simulation 48*, 3-4 (1993), 233–243.

[100] WOLNY-DOMINIAK, A., AND TRZESIOK, M. *insuranceData: A Collection of Insurance Datasets Useful in Risk Classification in Non-life Insurance.*, 2014. R package version 1.0.

[101] WOLNY-DOMINIAK, A., AND TRZESIOK, M. *MixedPoisson: Mixed Poisson models*, 2015. R package version 1.0.

[102] YANG, Z., HARDIN, J. W., AND ADDY, C. L. Testing overdispersion in the zero-inflated poisson model. *Journal of Statistical Planning and Inference 139*, 9 (2009), 3340–3353.

[103] YIP, K. C., AND YAU, K. K. On modeling claim frequency data in general insurance with extra zeros. *Insurance: Mathematics and Economics 36*, 2 (2005), 153–163.

[104] ŻĄDŁO, T. On parametric bootstrap and alternatives of mse. In *Proceedings of the 31st International Scientific Conference Mathematical Method in Economics, (Hana Vojáčková ed.), College of Polytechnics Jihlava, Jihlava* (2013), pp. 1081–1086.

[105] ZEILEIS, A., KLEIBER, C., AND JACKMAN, S. Regression models for count data in r. *Journal of statistical software 27*, 8 (2008), 1–25.

[106] ZHANG, Y. Likelihood-based and bayesian methods for tweedie compound poisson linear mixed models. *Statistics and Computing 23*, 6 (2013), 743–757.

Appendix A

The List of R packages in order of appearance

Table A.1: **R** packages used in examples

Library	Citation
library(stats4)	[82]
library(MASS)	[93]
library(insuranceData)	[100]
library(tweedie)	[27]
library(statmod)	[89]
library(dglm)	[28]
library(faraway)	[34]
library(gamlss.dist)	[91]
library(pscl)	[105]
library(cplm)	[106]
library(pls)	[74]
library(fitdistrplus)	[20]
library(gridExtra)	[4]
library(lattice)	[87]
library(ggplot2)	[97]
library(extRemes)	[44]
library(gPdtest)	[31]
library(lme4)	[5]
library(glmmML)	[12]
library(MixedPoisson)	[101]
library(hglm)	[85]
library(StatMatch)	[26]
library(graphics)	[82]
library(gridExtra)	[4]
library(copula)	[49]
library(CopulaRegression)	[59]
library(plotrix)	[51]
library(ChainLadder)	[40]