Predictive Modeling Applications in Actuarial Science

Volume II: Case Studies in Insurance

Predictive modeling uses data to forecast future events. It exploits relationships between explanatory variables and the predicted variables from past occurrences to predict future outcomes. Forecasting financial events is a core skill that actuaries routinely apply in insurance and other risk-management applications. *Predictive Modeling Applications in Actuarial Science* emphasizes lifelong learning by developing tools in an insurance context, providing the relevant actuarial applications, and introducing advanced statistical techniques that can be used to gain a competitive advantage in situations with complex data.

Volume II examines applications of predictive modeling. Where Volume I developed the foundations of predictive modeling, Volume II explores practical uses for techniques, focusing especially on property and casualty insurance. Readers are exposed to a variety of techniques in concrete, real-life contexts that demonstrate their value, and the overall value of predictive modeling, for seasoned practicing analysts as well as those just starting out.

Edward W. (Jed) Frees is the Hickman-Larson Chair of Actuarial Science at the University of Wisconsin–Madison. He received his PhD in mathematical statistics in 1983 from the University of North Carolina at Chapel Hill and is a Fellow of both the Society of Actuaries (SoA) and the American Statistical Association (the only Fellow of both organizations). Regarding his research, Professor Frees has won several awards for the quality of his work, including the Halmstad Prize for best paper published in the actuarial literature (four times).

Glenn Meyers, PhD, FCAS, MAAA, CERA, retired from ISO at the end of 2011 after a 37-year career as an actuary. He holds a BS in mathematics and physics from Alma College, an MA in mathematics from Oakland University, and a PhD in mathematics from SUNY at Albany. A frequent speaker at Casualty Actuarial Society (CAS) meetings, he has served, and continues to serve, the CAS and the International Actuarial Association on various research and education committees. He has also served on the CAS Board of Directors. He has several published articles in the *Proceedings of the Casualty Actuarial Society*, *Variance*, and *Actuarial Review*. His research contributions have been recognized by the CAS through his being a three-time winner of the Woodward-Fondiller Prize; a two-time winner of the Dorweiler Prize; and a winner of the DFA Prize, the Reserves Prize, the Matthew Rodermund Service Award, and the Michelbacher Significant Achievement Award. In retirement, he still spends some of his time on his continuing passion for actuarial research.

Richard A. Derrig is founder and principal of OPAL Consulting LLC, which is a firm that provides research and regulatory support to property casualty insurance clients. Primary areas of expertise include financial pricing models, database and data-mining design, fraud detection planning and implementation, and expert testimony for regulation and litigation purposes.

INTERNATIONAL SERIES ON ACTUARIAL SCIENCE

The *International Series on Actuarial Science*, published by Cambridge University Press in conjunction with the Institute and Faculty of Actuaries, contains textbooks for students taking courses in or related to actuarial science, as well as more advanced works designed for continuing professional development or for describing and synthesizing research. The series is a vehicle for publishing books that reflect changes and developments in the curriculum, that encourage the introduction of courses on actuarial science in universities, and that show how actuarial science can be used in all areas where there is long-term financial risk.

A complete list of books in the series can be found at www.cambridge.org/statistics. Recent titles include the following:

Computation and Modelling in Insurance and Finance
Erik Bølviken

Solutions Manual for Actuarial Mathematics for Life Contingent Risks (2nd Edition)
David C.M. Dickson, Mary R. Hardy & Howard R. Waters

Actuarial Mathematics for Life Contingent Risks (2nd Edition)
David C.M. Dickson, Mary R. Hardy & Howard R. Waters

Risk Modelling in General Insurance
Roger J. Gray & Susan M. Pitts

Financial Enterprise Risk Management
Paul Sweeting

Regression Modeling with Actuarial and Financial Applications
Edward W. Frees

Predictive Modeling Applications in Actuarial Science, Volume I: Predictive Modeling Techniques
Edited by Edward W. Frees, Richard A. Derrig & Glenn Meyers

Nonlife Actuarial Models
Yiu-Kuen Tse

Generalized Linear Models for Insurance Data
Piet De Jong & Gillian Z. Heller

PREDICTIVE MODELING APPLICATIONS IN ACTUARIAL SCIENCE

Volume II: Case Studies in Insurance

EDWARD W. FREES
University of Wisconsin–Madison

GLENN MEYERS

RICHARD A. DERRIG
OPAL Consulting LLC

CAMBRIDGE UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment,
a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of
education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107029880

First published 2016

A catalogue record for this publication is available from the British Library

ISBN 978-1-107-02988-0 Hardback

Contents

Contributors

Ernesto Schirmacher is Latin America's regional actuary for Liberty International, part of Liberty Mutual Insurance. He is a Fellow of the Society of Actuaries, an affiliate member of the Casualty Actuarial Society, and a member of the American Academy of Actuaries. He has worked in both direct and reinsurance companies in a variety of projects, such as asbestos reserving, economic capital modeling, and ratemaking using generalized linear models and other statistical techniques.

Dan Tevet is a director of data science at Liberty Mutual Insurance in Boston, where he leads a team of actuarial analysts and data scientists on the construction and implementation of predictive modeling and engineering solutions for the Distribution organization. Previously, he worked on product analytics for the personal property department at Liberty Mutual, and before that, he was at ISO in Jersey City, New Jersey, in the predictive modeling and specialty commercial lines divisions. He is a Fellow of the Casualty Actuarial Society and an active volunteer for the CAS.

Greg Taylor holds an honorary professorial position in risk and actuarial studies at the University of New South Wales. He previously spent 44 years in commercial actuarial practice and 8 years as an actuarial academic. He has published two books on loss reserving and numerous other articles in mathematics, statistics, and actuarial science. He is an officer of the Order of Australia and holds a Gold Medal from the Australian Actuaries Institute and a Silver Medal from the United Kingdom Institute and Faculty of Actuaries.

James Sullivan is a commercial pricing actuary specialising in long tailed lines. He is a Fellow of the Institute of Actuaries Australia and was a co-op scholar at the University of New South Wales. His research interests are focused on using generalized linear modeling and gradient boosting algorithms for pricing alternative investments opportunities.

Peng Shi is an assistant professor of actuarial science, risk management, and insurance at the University of Wisconsin–Madison. He is an associate of the Society of Actuaries. His research interests are predictive modeling, multivariate regression and dependence models, longitudinal data, and asymmetric information in insurance.

James Guszcza is the U.S. predictive analytics lead for Deloitte Consulting's Actuarial, Risk, and Advanced Analytics practice. He is a Fellow of the Casualty Actuarial Society and a past faculty member in the Department of Actuarial Science, Risk Management, and Insurance at the University of Wisconsin–Madison.

Mona Salah Ahmed Hammad, ASA, PhD, is assistant professor of the Mathematics and Insurance department, Faculty of Commerce, Cairo University, Egypt. Her current research includes ratemaking, health insurance, and multilevel linear modeling. She has acted as an actuarial advisor to the Egyptian Financial Supervisory Authority (EFSA) and was recently awarded second place for best paper in the nonlife track in the 30th International Congress of Actuaries (ICA2014).

Galal Abdel Haleim Harby, PhD, is currently vice-president of Al Ahram Canadian University in Egypt, where he has also acted as dean of the School of Business Administration since 2008. His permanent affiliation is as a professor of insurance in the Math and Insurance department, Faculty of Commerce, Cairo University, Egypt. He has published several books and papers, has served as an insurance advisor to the Egyptian Financial Supervisory Authority (EFSA), and is a winner of Cairo University's encouraging prize for scientific research in property insurance.

Ji Yao, FIA, CERA, is a lecturer in actuarial science at the University of Kent and an actuarial consultant. He has more than 10 years of experience in general insurance market on pricing, reserving, and capital modeling. His research interests include predictive modeling, stochastic claim reserving, data mining, and, more recently, Big Data.

Louise Francis, FCAS, MAAA, is the consulting principal and founder of Francis Analytics and Actuarial Data Mining Inc., where she leads reserving, pricing, predictive modeling, simulation, and related actuarial projects and engagements. Ms. Francis is a former VP research for the Casualty Actuarial Society (CAS) and has been involved in a number of CAS initiatives, including estimating reserve variability, improving data quality, and reviewing papers for the CAS journal *Variance*. She presents frequently on data mining-related topics and is a five-time winner of the CAS's Data Management and Information call paper program.

Glenn Meyers, FCAS, MAAA, CERA, PhD, recently retired after a 37-year actuarial career that spanned both industry and academic employment. For his last 23 years of working, he was employed by ISO as a research actuary. He has received numerous

awards for his publications from the Casualty Actuarial Society that include being the first recipient of the Michaelbacher Significant Achievement Award, which "recognizes a person or persons who have significantly and fundamentally advanced casualty actuarial science."

Luyang Fu, PhD, FCAS, is the department head of Predictive Analytics at Cincinnati Insurance Companies, where he leads the development of predictive models to support all business units, including personal line, commercial line, claims, and marketing. Prior to joining Cincinnati Insurance, he led the development of personal line pricing models, commercial line underwriting models, and enterprise risk models at State Auto Insurance Companies. He holds a master's degree in finance and a doctorate in agricultural and consumer economics from the University of Illinois at Urbana-Champaign.

Xianfang Liu, PhD, FCAS, is an actuary in the Predictive Analytics department at Cincinnati Insurance Companies, leading the development of commercial property and workers' compensation underwriting and pricing models. He has a PhD in mathematics from East China Normal University and a master's degree in computer science from Johns Hopkins University.

Claudine Modlin is a Fellow of the Casualty Actuarial Society and the leader of Willis Towers Watson's property/casualty pricing and product management team in the Americas. Her primary areas of expertise are predictive analytics and insurance pricing. Claudine has a bachelor's degree in mathematics from Bradley University, and prior to joining Willis Towers Watson in 2002, she was employed as an actuary at Allstate Insurance Company and AIG.

Mohamad A. Hindawi is currently a vice president of quantitative research and analytics at Allstate Insurance Company. Prior to that, he was a senior consultant with Towers Watson. His areas of expertise include predictive modeling and product management. He is a Fellow of the Casualty Actuarial Society and received his PhD in mathematics from the University of Pennsylvania.

Jim Weiss is the director of analytic solutions at ISO, a Verisk Analytics company. He is coauthor of "Beginner's Roadmap to Working with Driving Behavior Data," which was awarded the 2012 Casualty Actuarial Society (CAS) Management Data and Information Prize. He is a Fellow of CAS, a member of the American Academy of Actuaries, and a Chartered Property Casualty Underwriter.

Udi Makov is the head of actuarial research for Telematics at Verisk Insurance Solutions. He's also the director of the actuarial research center at the University of Haifa, where he specializes in statistical methods in insurance, Bayesian statistics, and mixture models. Dr. Makov holds a PhD in mathematical statistics from the University of London.

Preface

In January 1983, the North American actuarial education societies (the Society of Actuaries and the Casualty Actuarial Society) announced that a course based on regression and time series would be part of their basic educational requirements. Since that announcement, a generation of actuaries has been trained in these fundamental applied statistical tools. This two-set volume builds on this training by developing the fundamentals of predictive modeling and providing corresponding applications in actuarial science, risk management, and insurance.

The series is written for practicing actuaries who wish to get a refresher on modern-day data-mining techniques and predictive modeling. Almost all of the international actuarial organizations now require continuing education of their members. Thus, in addition to responding to competitive pressures, actuaries will need materials like these books for their own continuing education. Moreover, it is anticipated that these books could be used for seminars that are held for practicing actuaries who wish to get professional accreditation (known as VEE, or validated by educational experience).

Volume I lays out the foundations of predictive modeling. Beginning with reviews of regression and time series methods, this book provides step-by-step introductions to advanced predictive modeling techniques that are particularly useful in actuarial practice. Readers will gain expertise in several statistical topics, including generalized linear modeling and the analysis of longitudinal, two-part (frequency/severity) and fat-tailed data. Thus, although the audience is primarily professional actuaries, the book exhibits a "textbook" approach, and so this volume will also be useful for continuing professional development.

An international author team (seven countries, three continents) developed Volume I, published in 2014. You can more learn more about Volume I at

http://research.bus.wisc.edu/PredModelActuaries

Volume II examines applications of predictive models, focusing on property and casualty insurance, primarily through the use of case studies. Case studies provide a learning experience that is closer to real-world actuarial work than can be provided

by traditional self-study or lecture/work settings. They can integrate several analysis techniques or, alternatively, can demonstrate that a technique normally used in one practice area could have value in another area. Readers can learn that there is no unique correct answer. Practicing actuaries can be exposed to a variety of techniques in contexts that demonstrates their value. Academic actuaries and students see that there are valid applications for the theoretical material presented in Volume I. As with Volume I, we have extensive sample data and statistical code on the series website so that readers can learn by doing.

The first three chapters of Volume II focus on applications of the generalized linear model (GLM), arguably the workhorse of predictive modeling in actuarial applications. Chapter 1, by Ernesto Schirmacher, gives an overview of the use of GLMs in pricing strategy, focusing on private passenger automobile. Dan Tevet's Chapter 2 reinforces this discussion by examining insurance for motorcycles, emphasizing the comparison between frequency-severity and pure premium models. In Chapter 3, Greg Taylor and James Sullivan demonstrate how to use GLM techniques in loss reserving. Although the two books in the series are written independently, readers with access to the first book will appreciate these three chapters more deeply after reviewing the foundations in Chapters 2–6 of Volume I.

Chapters 4 and 5 provide extensions of the generalized linear model. Like Chapter 1, in Chapter 4, Peng Shi and James Guszcza also examine pricing strategies for personal automobile insurance. However, they show how to price insurance when more than one type of coverage, such as third-party liability and personal injury protection, is available in the database; by taking advantage of the multivariate nature of claims, they are able to incorporate dependencies among coverages in their pricing structure. For another approach to incorporating dependencies, in Chapter 5, Mona S. A. Hammad and Galal A. H. Harby use multilevel models. They provide a unique and interesting case study of group health insurance in the Egyptian market. Chapters 8 and 16 of Volume I provide an introduction to mixed and multilevel modeling.

Chapters 6 and 7 describe applications of unsupervised predictive modeling methods. Most predictive modeling tools require that one or more variables be identified as "dependent variables" or the "outcome of interest," and other variables are used to explain or predict them; this is known as a supervised predictive model. In contrast, unsupervised models treat all variables alike and do not require this identification. Chapter 12 of Volume I, by Louise Frances, introduced unsupervised learning with a focus on common methods of dimension reduction, principal components/factor analysis, and clustering. Chapter 6 of this volume, by Ji Yao, builds on this introduction with an application in insurance ratemaking. Louise A. Frances, the author of Chapter 7 of this volume, also follows up with two advanced unsupervised learning techniques, a variation of principal components known as PRIDIT, and a

(unsupervised) variation of random forests, a tree-based data-mining method. She applies these techniques to help identify predictors of claims that are fraudulent or questionable.

Chapters 8 through 11 show how to use predictive modeling techniques in problems that are currently receiving substantial attention in actuarial science and insurance risk modeling. In Chapter 8, Glenn Meyers shows how to take the output of a Bayesian Monte Carlo Markov chain (MCMC) stochastic loss reserve model and calculate the predictive distribution of the estimates of the expected loss over a finite time horizon. Luyang Fu and Xianfang (Frank) Liu, in Chapter 9, compare GLM modeling to finite mixture models to study claims triaging and high-deductible pricing using workers, compensation data.

Chapter 10, by Mohamad A. Hindawi and Claudine H. Modlin, provides a framework for managing claim escalation. This chapter also discusses claims triaging and text mining, using penalized regression techniques, such as elastic net, to help with variable selection. In Chapter 11, Udi Makov and Jim Weiss describe how to analyze data collected from policyholders' vehicles via telematics to help determine motor vehicle premium rates. Data collected via telematics are volatile and voluminous, and actuaries and data scientists must take particular care when applying predictive modeling techniques.

Acknowledgments

Funding for this project was provided by the Casualty Actuarial Society and the Canadian Institute of Actuaries. The authors also thank the Wisconsin School of Business for hosting the book's website.