The Journal of Risk and Insurance

DOI: 10.1111/jori.12185

Book Review

Computational Actuarial Science With R by Arthur Charpentier (editor), 2015, Boca Raton, FL: CRC Press, 618 pages, ISBN: 978-1-4665-9259-9.

Reviewer: Tatjana Miljkovic, Ph.D., Department of Statistics, Miami University; miljkot@miamioh.edu

The popularity of R software in data science, statistical analysis, and predictive analytics jobs has grown tremendously in the past decade. Based on the study by Muenchen (2016) on the number of scholarly articles found in 2015 for each commercial software, R software is reported in second place following SPSS, surpassing SAS. A number of books using R have already been written in statistics, economics, engineering, psychology, and other disciplines.

Most books written in the actuarial science area focus exclusively on theory while lacking practical applications, especially related to a particular use of computational methods and software. To my knowledge, the first attempt to integrate R with actuarial science applications was made in the book *Modern Actuarial Theory With R*, written by Kaas et al. (2008), focusing mostly on nonlife insurance topics. The book *Computational Actuarial Science With R* provides a much broader and comprehensive review of actuarial topics related not only to nonlife insurance but also to life insurance and finance areas of actuarial practice.

As the actuarial science field has changed in the past two decades with advances in predictive modeling, modern financial economics, and statistical computing methods, there has been a great need for developing modern actuarial methods that focus on the computational aspects of actuarial science. Implementation of these methods in R software not only allows the actuarial field to keep up with computational science (e.g., computational statistics) but also to remain competitive in the marketplace. *Computational Actuarial Science With R* elegantly covers a great deal of useful material and applications of R in actuarial science and leaves out much of the actuarial theory that is commonly found in other actuarial books. Numerous real data sets that accompany the book come from 14 countries, bundled up in an R package, “CASdatasets,” and allow researchers, industry practitioners, and students to get a hands-on, efficient implementation of actuarial concepts and data analysis.

A beginning user of R can use the introduction of the book to get up to speed on basic terminology and expressions of the R language. Certain sections of the book can also serve as supplemental material to regular textbooks used with actuarial courses taught at the university level.

Computational Actuarial Science With R is divided into four main parts: Introduction to R Language, Statistical Models With R, Methodology, and Life Insurance, Finance, and Non-Life Insurance. As the editor of this scholarly book, Arthur Charpentier, a professor of actuarial science at the University of Québec at Montréal, has put together a fine collection of articles prepared by 26 contributors (including himself) from the industry and universities around the world.

The first section of the book includes several methodological concepts such as statistical inference and learning, Bayesian philosophy, spatial analysis, reinsurance, and extreme events. Drawing heavily on the theory presented in Klugman, Panjer, and Willmot (2012), Chapter 2 presents the most common discrete, continuous, and mixed distributions used in actuarial science and their implementation in R. Here, a reader should be aware when using "mixtools" and "norm1mix" packages to model mixtures based on normal distributions because they are not suitable for modeling loss data that are typically defined on a positive domain. In the same chapter, the definitions of linear regression model, aggregate loss distribution, copulas, and multivariate distributions are explained, followed by R code illustrating the implementation of these statistical methods. The Bayesian approach to solving actuarial problems with a rich set of R tools is presented in Chapter 3. Chapter 4 on statistical learning summarizes how well-known statistical models can be built based on the real data. Spatial analysis with GIS is discussed as part of Chapter 5, with R files available in an R repository at https://github.com/CASwithR/stats/. Developing maps and embracing stochastic models for spatial variation is an integral part of actuarial jobs in the P&C industry. Readers will learn the capabilities of the "sp" and "maptools" packages in reading and manipulating geographic data. Chapter 6 provides an R framework for basic reinsurance problems. Implementation of basic extreme value theory with statistical inference should be very helpful to those researchers and practitioners dealing with extreme events.

In the second section of the book, the authors review life insurance topics that rely on knowledge of life contingencies, life tables, and survival analysis. The implementation of actuarial mathematics calculations can be easily done in R. I found Chapter 7 to be good supplemental material in teaching life contingencies courses that are traditionally taught with more emphasis on theory and minimal use of software. Chapter 8 illustrates the application of several popular models for forecasting mortality, such as Lee–Carter, Lee–Miller, Booth–Maindonald–Smith, and Hyndman–Ullah, which are also implemented in the "demography" package. A convenient feature of this chapter is that the R code presented in the book is also available in the repository at https://github.com/CASwithR/life/. Prospective mortality tables and portfolio experience are discussed in Chapter 9. Chapter 10 is dedicated to survival analysis, covering the topics of incomplete data, the Kaplan–Meier estimator, the Cox model, and additive models.

The third section of the book deals with finance, covering the topics of stock prices and time series analysis, yield curves and interest rates models, and portfolio allocation. GARCH and GARCH-related models for modeling univariate time series process are introduced for arbitrary distribution of the conditional volatility of financial returns. Estimation of risk measures (e.g., VaR) is considered for the Peak-Over-Threshold (POT) method and GARCH(1,1) models. Parallel R computation with a C/C++ routine in Windows or Unix-like operating systems can speed up R code in these applications.

The fourth section of the book covers nonlife insurance topics with special emphasis on pricing experience rating and loss reserving methods. Numerous predictive models of claim frequency and severity are discussed in Chapter 14. This chapter requires knowledge of generalized linear models (GLMs). Many computational aspects with R used in this chapter can also be found in the book by Faraway (2014).

Pricing with longitudinal data is discussed as part of Chapter 15. The R outputs for fixed-effect and random-effects models are easily reproducible from "AutoClaim" data. Three approaches are discussed to address unobserved heterogeneity and serial correlation in the data. The theory behind the GLM model for longitudinal data such as Poisson GLMM is presented. The reader is also presented with a case study related to Bonus-Malus scales in R that is commonly used as a special type of experience rating system in motor insurance.

Chapter 16 covers some basic popular methods used in claim reserving, such as the Mack chain-ladder and bootstrapping chain-ladder methods that have been in practice for a while. The reader should use the R file "Claims_Reserving_and_IBNR.R," stored at repository: https://github.com/CASwithR/nonlife, when reviewing the R output along with the theory, remarks, and discussion provided in the book. Having this R file conveniently available, this chapter of the book can be used as supplemental teaching material in a classroom setting.

For users' convenience, it would be of great help if the R library "CASdatasets" included the naming conventions of data files and their variables in line with the book. For example, the name of the data file "CONTRACTS" used in writing Chapter 14 corresponds to the name "freMTPLfreq" in the R library "CASdatasets." Additionally, the names of variables included in "CONTRACTS" differ from those variable names associated with "freMTPLfreq." A user can intuitively translate between file names "CLAIMS" and "freMTPLsev" and names of the variables included in these files; however, it takes an advanced user to follow the R code in the book and debug any syntax errors that occur because of the differences in naming. A similar mismatch between file names can also be observed in Chapters 1, 9–11, and 15.

The authors state that the book is written for someone who has some knowledge of R. While this is true for some chapters, I believe that only an advanced user of R can take full advantage of the book. It would be of great help to readers if the contributors to this book would consider adding the R files used in writing the book to the CASwithR repository for each chapter. This would also attract many beginning users of R.

In conclusion, I believe that not only contributors to the *Journal of Risk and Insurance*, but the actuarial community in general, will find this book to be an excellent reference for R code related to actuarial science applications, and that *Computational Actuarial Science With R* will quickly find its place on many bookshelves.

References

Faraway, J. J., 2014, *Linear Models With R* (Boca Raton, FL: CRC Press).

Kaas, R., M. Goovaerts, J. Dhaene, and M. Denuit, 2008, *Modern Actuarial Risk Theory Using R*, Vol. 128 (New York: Springer Science & Business Media).

Klugman, S. A., H. H. Panjer, and G. E. Willmot, 2012, *Loss Models: From Data to Decisions*, Vol. 715 (Hoboken, NJ: John Wiley & Sons).

Muenchen, R.A., 2016. The Popularity of Data Analysis Software. Available at http://r4stats.com/articles/popularity/ (accessed June 15, 2016).