Multivariate Insurance Loss Models with Applications in Risk Retention

By

Gee Yul Lee

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Business)

at the

UNIVERSITY OF WISCONSIN-MADISON

2017

Date of final oral examination: 05/03/2017

The dissertation is approved by the following members of the Final Oral Committee:
Edward W. Frees, Professor, Business
Marjorie A. Rosenberg, Professor, Business
Justin Sydnor, Associate Professor, Business
Peng Shi, Associate Professor, Business
Zhengjun Zhang, Professor, Statistics

Acknowledgements

I am thankful to have studied from the best PhD learning environment. The Wisconsin School of Business persistently supported me. For this, I thank Professor Joan Schmit, Department chair of Risk and Insurance, who is a great leader with love and passion.

I thank my PhD advisor Professor Edward (Jed) W. Frees, who supervised me academically throughout my studies. Professor Frees read all my reports with detail and patience and provided advice. After years, the reports matured into papers with sophisticated models, until the memorable moment, when he recommended me to present at a conference for the first time in my life. He gave me motivation when I needed it, sometimes using the stick, sometimes rewards, and sometimes using hundreds of comments to a report. He is an advisor with endless love and passion for students, and about academic research in Actuarial Science. I am thankful for Professor Frees, for guiding me through my studies.

I thank members of my dissertation committee for helpful insights. I thank Professor Margie Rosenberg, a loving mentor of Actuarial Science, who helped me throughout my doctoral studies with helpful comments and suggestions. I thank Professor Justin Sydnor for providing mentorship throughout my studies, and motivating me through his Risk and Insurance seminar. I thank Professor Peng Shi, a motivating mentor and researcher of Actuarial Science, who helped my research. I thank Professor Zhengjun Zhang of Statistics, for agreeing to be on my dissertation committee, and providing suggestions regarding my work. In addition to the committee members, I also thank Risk and Insurance mentors Mark Browne, Martin Halek, Ty Leverty, Anita Mukherjee, Patricia Mullins, Paul Johnson, and Kirk Peter. In addition, I thank my fellow PhD students at the Department of Risk and Insurance.

This dissertation work has been generously funded in part by the Society of Actuaries (SOA)

Research Studies Proposal (RSP) in General Insurance, and in part by the Wisconsin Office of the Commissioner of Insurance (OCI). I thank Brynn Bruijn-Hansen of the OCI for support. I also thank Scott Lennox, Erika Schulty, Dale Hall, Steven Siegel, Barbara Scott of the SOA, and the project oversight group members Robert Winkelman, Michael Ewald, Jean-Marc Fix, Josh Newkirk, Rafi Herzfeld, Andy Ferris, Crispina Caballero, Joe Wurzburger, Julianne Callaway, and Tina Liu, for their valuable insights.

Chapter 2 of this dissertation includes joint work with Professor Jed Frees and Lu Yang. Chapter 4 is based on joint work with Professor Jed Frees. The Appendix (Chapter 7) contains material created during the Local Government Property Insurance Fund ratemaking project. Contributors to the ratemaking project include: Gregory Wanner, Whitney Fose, Jordan Paszek, Francis Lei, James Farley, Henry Chao, Emily Jenkins, Katherine Nondahl, Mike Delucca, Corinna G. Olson, Travis Millikin, Judith-Anne Lesko.

Abstract

This dissertation contributes to the risk and insurance literature by expanding our understanding of insurance claims modeling, deductible ratemaking, and the insurance risk retention problem. In the claims modeling part, a data-driven approach is taken to analyze insurance losses using statistical methods. It is often common for an analyst to be interested in several outcome measures depending on a large set of explanatory variables, with the goal of understanding both the average behavior, and the overall distribution of the outcomes. The use of multivariate analysis has an advantage in a broad context, and the literature on multivariate regression modeling is extended with a focus on dependence among multiple insurance lines. In this process, a deductible is an important feature of an insurance policy to consider, because it may influence the frequency and severity of claims to be censored or truncated. Standard textbooks have approached deductible ratemaking using models for coverage modification, utilizing parametric loss distributions. In practice, regression could be used with explanatory variables including the deductible amount. The various approaches to deductible ratemaking are compared in this dissertation. Ultimately, an insurance manager would be interested in understanding the influence of a retention parameter change to the risk of a portfolio of losses. The retention parameter may be deductible, upper limit, or coinsurance. This dissertation contributes to the statistics and actuarial literature by introducing and applying the 01-inflated negative binomial frequency model (a frequency model for observations with an inflated number of zeros and ones), and illustrating how discrete and continuous copula methods can be empirically applied to insurance claims analysis. In the process, the dissertation provides a comparison among various deductible analysis procedures, and shows that the regression approach has an advantage in problems of moderate size. Finally, the dissertation attempts to broaden our understanding of the risk retention problem within a constrained optimization framework, and demonstrates the quasiconvexity of the objective function in this problem. The dissertation reveals that the loading factor of a reinsurance premium has a risk measure interpretation, and relates to the risk measure relative margins (RMRM). Concepts are illustrated using the Wisconsin Local Government Property Insurance Fund (LGPIF) data.

Contents

Abstract iii

1 Introduction 1

1.1 Multivariate Insurance Claims Modeling 2

1.2 Deductible Ratemaking 4

1.3 Insurance Portfolio Optimization 5

1.4 Endorsements Ratemaking 7

1.5 Conclusion of Introduction 8

2 Multivariate Insurance Claims Modeling 9

2.1 Introduction to Multivariate Claims Modeling 10

2.2 Univariate Foundations 13

2.2.1 Frequency-Severity 14

2.2.2 Modeling Frequency 14

2.2.3 Modeling Severity 16

2.2.4 Tweedie Model 19

2.3 Multivariate Models and Methods 21

2.3.1 Copula Regression 21

2.3.2 Multivariate Severity 22

2.3.3 Multivariate Frequency 23

2.3.4 Multivariate Tweedie 24

2.3.5 Association Structures and Elliptical Copulas 25

2.3.6 Assessing Dependence 27

2.3.7 Frequency-Severity Modeling Strategy . . . 27

2.4 Frequency Severity Dependency Models . . . 30

2.5 Chapter Summary . . . 32

3 Deductible Ratemaking **34**

3.1 Introduction . . . 35

3.2 Literature Review . . . 37

3.3 Theory of Coverage Modification . . . 42

3.4 Approaches to Deductible Ratemaking . . . 47

3.4.1 Maximum Likelihood Approach to Deductible Ratemaking . . . 47

3.4.2 Regression Approach to Deductible Ratemaking . . . 48

3.4.3 Relativity Calculation . . . 50

3.5 Applications . . . 52

3.5.1 Simulation . . . 52

3.5.2 Relativities . . . 54

3.5.3 Aggregate Claims Prediction . . . 59

3.5.4 Comparison of Frequency Models . . . 62

3.6 Summary . . . 64

4 Insurance Portfolio Optimization **66**

4.1 Introduction . . . 67

4.2 Literature Review . . . 68

4.3 Framework . . . 72

4.4 Single Loss Risk Retention . . . 76

4.5 Single Policy with Omit-i Portfolio . . . 81

4.5.1 Non-convexity . . . 81

4.5.2 Quasi-convexity . . . 83

4.5.3 Non-convexity Example . . . 86

4.6 Aggregate Loss Risk Retention . . . 88

4.6.1 Theory . . . 88

4.6.2 Numerical Optimization for the Aggregate Loss Problem . . . 95

4.7 Blocks of Policies . . . 97

4.8 Supplementary Notes . . . 99

4.8.1 Single Policy per Occurrence and per Policy Retention . . . 100

4.8.2 With Portfolio Loss . . . 101

4.8.3 Future Work . . . 102

4.8.4 RM^2 Computation . . . 103

4.8.5 Conditions for Existence and Uniqueness of Global Optima . . . 106

5 LGPIF Case Study **108**

5.1 Insurance Claim Modeling . . . 109

5.1.1 Basic Summary Statistics . . . 109

5.1.2 Marginal Model Fitting—Zero/One Frequency, *GB2* Severity . . . 112

5.1.3 Copula Identification and Fitting . . . 116

5.1.4 Other Lines . . . 122

5.1.5 Tweedie Margins . . . 127

5.1.6 Out-of-sample Validation . . . 128

5.1.7 Gini Index . . . 130

5.1.8 Summary of Insurance Claim Modeling . . . 131

5.2 Deductible Ratemaking . . . 133

5.2.1 LGPIF Claim-Level Data . . . 133

5.2.2 Censored and Truncated Estimation . . . 135

5.2.3 01-Inflated Model Estimation . . . 151

5.2.4 Summary of Deductible Ratemaking . . . 153

5.3 Insurance Portfolio Optimization . . . 154

5.3.1 Framework . . . 154

5.3.2 Madison Metropolitan School District Policy . . . 154

5.3.3 The Omit-i Portfolio . . . 155

5.3.4 Results . . . 156

5.3.5 Summary of Insurance Portfolio Optimization Case Study . . . 160

6 Conclusion and Future Work **161**

6.1 Insurance Claims Modeling 161

6.2 Deductible Ratemaking 162

6.3 Insurance Portfolio Optimization 163

6.4 Summary of Conclusion and Future Work 165

Bibliography **165**

7 Appendix: Rating Engine Summary **173**

7.1 Overview 173

7.2 Rating Engine 174

7.2.1 Building and Contents 174

7.2.2 Motor Vehicle 176

7.2.3 Contractor's Equipment 176

7.2.4 Endorsements 176

7.3 Comparison of Rating Engine Proposals 178

Chapter 1

Introduction

Insurance analytics enables approaches for insurance organizations to make decisions based on the analysis of data, in order to define problems, analyze and devise solutions of the problems using models, and monitor the outcome of the solution based on data. This process of decision making begins by first observing a problem under consideration, and identifying the risks and external factors associated with the problem. Consider the task of developing a rating engine for an insurance organization. The actuary is aware of the external environment in which the organization operates, including the regulatory environment, and the stakeholders surrounding the organization. Based on various factors, the problem could be defined as updating the rating engine based on empirical data. Once a problem is defined, an appropriate solution could be devised, and presented.

This dissertation has been motivated by a ratemaking project for the Local Government Property Insurance Fund (LGPIF), which is introduced further in Chapter 7. The initial request from the LGPIF concerned the *surplus ratio* (a measure of an insurer's financial health) of the property fund approaching the target ratio above which the surplus should be maintained. Historically, the property fund was able to maintain a surplus ratio above the target ratio, however in recent years due to severe weather events in 2010, the surplus ratio was approaching the target rapidly. In this case, the first task of the actuary is to define the problem. In the case of the LGPIF, implementing a more accurate and cost efficient rating engine was mutually agreed to be the primary goal of the project. Hence, a team including myself, worked on the ratemaking of the LGPIF building and contents, contractor's equipment, and motor vehicles coverage groups, in order to present a rating engine to the LGPIF. Details of this rating engine, and the LGPIF can be found in Chapter 7 of this dissertation.

When defining and solving a problem, what is observed may be only part of a larger problem, and an actuary must look below the surface in order to have an idea of the full problem. After completing the LGPIF ratemaking project, a researcher's perspective naturally leads to the question of whether a new and accurate rating engine would suffice for the LGPIF, or could the actuary do more in terms of providing an answer to broader problems concerning the adequacy of rates. What more could an actuary do? This question leads to the primary chapters of this dissertation. This dissertation provides an overview of insurance claims modeling in Chapter 2, an overview of insurance deductible ratemaking in Chapter 3, and an application of insurance claim models in risk retention problems in Chapter 4. Each of the concepts are applied in Chapter 5 in the form of a case study using the LGPIF data set. To provide helpful background, the LGPIF data and the rating algorithm is summarized in detail in Chapter 7.

In the following few pages, a brief summary of the entire dissertation is provided. Section 1.1 summarizes multivariate insurance claims modeling. Section 1.2 summarizes approaches for deductible ratemaking, and Section 1.3 summarizes insurance risk retention applications. Section 1.4 summarizes the endorsement ratemaking approach used in the LGPIF rating algorithm.

1.1 Multivariate Insurance Claims Modeling

The first question is whether the modeling sufficiently accounts for the characteristics of the outcome variables. This includes the dependencies among variables. Correctly understanding the losses of an insurance company involves not only modeling the average behavior of variables, but also understanding the worst possible scenarios of the losses. One approach to understanding the risk of a portfolio of losses is to analyze large potential loss amounts which could happen with very small probabilities. Potentially large losses to a portfolio have significant meanings to the risk imposed to a company. Dependencies among risks either at the individual level or at the line level may influence the worst-case outcome, and hence the overall risk measure of an insurance portfolio. Therefore, correctly understanding insurance risks involves understanding potential dependencies among risks.

Regression models for understanding the distribution of claim outcomes continue to be developed, and sophisticated models are used in the literature. For the purpose of insurance ratemaking, it is recommended to model the frequency of claims, and the severity of claim amounts. In addition to the frequency-severity components, outcomes can be further categorized by entity type and peril types, requiring complex modeling. Hence, an overview of multivariate frequency-severity regression models and dependence models are useful, and the application of these approaches are covered in Chapter 2 of this dissertation.

Section 2.2 starts with univariate foundations, with each subsection introducing the frequency-severity model using GLMs, and the 01-inflated negative binomial model (Section 2.2.2), as well as details of the severity model using the *GB2* model (Section 2.2.3). The chapter puts emphasis particularly on the *GB2* model, because of its flexibility in modeling losses with long-tail distributions. A χ^2-goodness of fit test is used for each of the six coverage groups (building and contents, equipments, and four different motor vehicle coverages), to assess the best model for each case of frequency data. The QQ-plot is utilized to assess the goodness of fit of the *GB2* distribution for the severity model of each coverage group. The Tweedie model, which is a compound Poisson model of a summation of gamma distributed random variables, is also fit as a comparison.

Section 2.3 describes multivariate foundations. The section provides an overview of copula methods, and provides references to various literature related to the usage of copulas in modeling multivariate data. Section 2.3.2 begins by multivariate models for severities, followed by the subsequent section dealing with dependence models for discrete outcomes. Section 2.3.4 discusses the dependence model for the Tweedie distribution, which we determine to be less desirable than the frequency-severity dependence model, based on the jittered plots of the Cox-Snell residuals.

Finally, in Section 2.4, the dependence between frequency and severity responses is considered. The Vuong's test is utilized as a procedure to determine the better model in terms of the log-likelihood, in case there are two competing models of dependence. The main text of the dissertation illustrates the use of Vuong's test result as a supporting evidence that a normal copula suffices for the modeling of dependence between frequencies and severities. Empirical results are shown in the LGPIF case study (Chapter 5), Section 5.1.

1.2 Deductible Ratemaking

When understanding insurance loss portfolios, an important question is whether deductibles are accurately priced. As we will see, deductibles and upper limits have mechanically similar features to insurance contracts because the upper limit of a contract to the insurer is essentially a deductible from a reinsurance company's perspective. Hence, the deductible can be understood as a parameter to determine the amount of insurance purchased by an insurance entity. Deductible choices may influence the frequency and severity of insurance claims, by truncating and censoring the observed data. A policyholder with $100,000 deductible may have a different risk profile, compared to a policyholder with a $500 base deductible. Deductible choices may also depend on psychological characteristics of the policyholder. Both the average behavior (the mean) and the worst-case behavior of the claim distribution is influenced by the deductible for various reasons. In particular, the average behavior determines the rates for the insurance contract. Hence, understanding the relationship between the deductible and the average behavior of an insurance portfolio is a fundamentally important task.

Meanwhile, the upper limit of a policy may be invisible to the policyholder when effected through a reinsurance arrangement. Psychological considerations may apply differently, when the upper limit rating problem is considered. For this reason, different approaches to ratemaking may be needed, depending on whether the deductible or the upper limit is rated. Standard actuarial textbooks have approached this problem by modeling the coverage modification amount using parametric loss distributions. Meanwhile, the regression approach could be used with explanatory variables including the deductible amount. When various approaches for deductible ratemaking are compared, it can be discovered that the regression approach enables reasonable assessment of deductible relativities for problems of moderate size. The regression approach also has the advantage in terms of incorporating endorsements using regularization methods (see Frees and Lee (2017)), which is an approach used for the rating engine of the LGPIF. The maximum likelihood approach may be more appropriate for the upper limit ratemaking problem. Various approaches to deductible ratemaking, and their pros and cons are discussed in Chapter 3 of this dissertation.

Section 3.2 begins with a review of the literature related to insurance deductible ratemaking. In this and subsequent sections, the concept of a relativity curve is introduced, and used as an important tool in assessing the loss elimination behavior of various models. The chapter proceeds to reviewing the theory of coverage modification in Section 3.2. The main text of the theory of coverage modification (Section 3.3) builds up foundational formulas for deductible ratemaking under the assumption that the underlying loss distribution has been estimated. Once a theory is built, the foundation is applied to two different empirical approaches to deductible ratemaking. The first approach is to estimate the underlying loss distribution using maximum likelihood, and to apply the theoretical formulas directly. The second approach is to perform a regression of the response variable on log deductibles, and to obtain the coefficient of the log deductible term, in order to calculate the relativities. The relativity curves for both approaches are compared in Section 3.4.3. A simulation study shows that the relativity curve for the regression approach is an approximation to the theoretically correct relativity curve, using one parameter (See Section 3.5.2).

Finally, an empirical study using the LGPIF data shows that the regression approach and the maximum likelihood approach of deductible ratemaking has pros and cons in their own respect (Section 3.5.3). The result of the chapter is that the regression approach has an advantage in deductible ratemaking in case a large number of covariates are used in the ratemaking model, and the deductible is of moderate size in a relative sense. The maximum likelihood approach turns out to have an advantage in assessing relativities for deductible levels beyond the point where empirical data is available.

Details of the model estimation results used for the empirical studies are shown in Section 5.2.1. Details of the truncated estimation of peril-type models and 01-inflated frequency models are shown in Section 5.2.2.

1.3 Insurance Portfolio Optimization

Once all of the previous problems are solved, the actuary is ready to consider broader problems. Insurance portfolio optimization is a fundamental problem of risk and insurance analytics, requiring modeling of insurance claims, analysis of risk retention parameters, and numerical optimization techniques. During the modeling stage, the insurance analyst builds sophisticated claims models for the purpose of ratemaking and risk assessment.

What do we mean by “insurance portfolio optimization?” Consider a single loss for simplicity.

The loss distribution would have a mean and a variance. If an imaginary insurance company were to cover this loss, what could be ways to optimize the surplus standing of the insurance company, theoretically speaking? Recalling concepts from the Markowitz portfolio optimization, a portfolio is better off when the variability of the outcome is minimal, given some predetermined requirement on the average outcome realized. The Markowitz portfolio optimization problem could be extended to the insurance liability case, and we may use the quantile of the insurance loss as a measure of variability. Now the question for the insurance company is, given a loss, what is the optimal amount of insurance/reinsurance to purchase (how much of the risk to retain). In considering this question, the goal is to minimize the variability of the retained insurance loss, subject to a minimum amount of premium needed to be collected. This relates to the concept illustrated in Markowitz (1952).

In Chapter 4, Section 4.1, this problem is formulated and introduced, with an emphasis on the loading factor $(1 + \lambda^*)$. Basically, this loading factor is found in every literature of reinsurance optimization. The easiest way to understand the loading factor is to think of the case where the loading factor is 1, thus a loading doesn't exist. If the loading doesn't exist, then the price of holding the risky loss, versus paying the safe reinsurance premium, becomes identical and hence the riskless reinsurance premium becomes the optimal choice. Hence full insurance coverage is sought. In the framework used in this dissertation, this corresponds to an upper limit parameter of zero so that all of the loss is covered. Under a certain loading amount, the optimal insurance amount is some partial coverage, and the optimal upper limit becomes an intermediate value between zero and infinity.

The interesting discovery is that since the loading factor is a Lagrangian multiplier, it could be understood as the $RM^2(\theta^*)$ risk measure, introduced in Frees (2016) in the one dimensional problem case. This means imposing a constraint on the premiums is actually analogous to fixing the loading factor to a constant, and vice versa. In the framework introduced in Section 4.3, the optimization problem is formulated and the analogy is illustrated more rigorously. Section 4.4 and 4.5 explores the single loss case in more detail, and demonstrates how constrained optimization could be used to determine the optimal insurance amount depending on the premium constraint imposed. In Section 4.6, the formulas for the aggregate loss case is introduced, and the non-convexity, and quasiconvexity of the optimization problem is discussed. Basically, the quantile function of the optimization problem to be solved is a monotonic function (as is proved in Section

4.6 for each of the d, u, c parameters), and hence is quasiconvex. This means the level sets of the function is convex. Thus, the local optima found from optimization becomes the global optima. Given a non-empty interval of the parameter, there exists a unique solution, according to Section 4.8.5. Non-convexity is illustrated using an example, in order to demonstrate the need for the establishment of quasi-convexity of the quantile function.

Ultimately, an insurance manager would be interested in the influence of a risk retention parameter change to a portfolio of retained risks. Furthermore, the manager may be interested in optimizing the portfolio of risks by adjusting the deductible or upper limit parameter for a block of insurance policies, or for multiple layers of parameters. These more complicated optimization problems are covered in Sections 4.7, and 4.8.

Concepts of the insurance portfolio optimization chapter are illustrated using the Wisconsin Local Government Property Insurance Fund (LGPIF) data in Section 5.3, where the Madison Metropolitan School District policy of the LGPIF is used as an example for optimizing the upper limit parameter u.

1.4 Endorsements Ratemaking

In addition to the material in Sections 1.1, 1.2, and 1.3, the Appendix of this dissertation (Chapter 7, Section 7.2.4) provides results from practical applications of regularization methods for endorsements ratemaking. The endorsement ratemaking approach using shrinkage estimation is illustrated in Frees and Lee (2017), where the endorsement coefficients are "shrunk" towards zero in order to regulate the size. Endorsements are optional coverages, such as zoo animals coverage for a building and contents insurance. Policyholders who elect to purchase the endorsement are subject to additional premium for the optional coverage, and hence an appropriate rate should be calculated. Shrinkage methods are able to moderate the size of regression coefficients, so as to regulate the rates for the endorsements in a disciplined way. Depending on the external environment, the manager is able to control how high or low the endorsement rate should be, without influencing the overall predictive ability of the rating engine.

Chapter 7 begins with a summary of the LGPIF, in order to provide an overview of the problem in hand. Section 7.2 explains the basic structure of the rating engine, where the insurance premium is determined by various multiplicative factors, including endorsement factors. Section 7.2.4 summarizes the endorsements offered by the property fund. The claim frequency and severity is summarized by each endorsement, in order to provide some intuition of the relationship between the endorsement and the claims. Shrinkage estimation is used to calculate the rates for each endorsement, as explained in Frees and Lee (2017). Section 7.3 shows various different proposals for the final rating engine, which depends on the rating variables and tuning parameter used.

1.5 Conclusion of Introduction

This dissertation takes a data-driven approach to insurance claims modeling, and illustrates various applications of insurance claim models. The work contributes to the risk and insurance literature by expanding our understanding of insurance claims modeling, deductible ratemaking, as well as insurance risk retention parameter optimization problems. The material consisting each chapter of this dissertation can be found in the following papers:

1. Frees, Edward W., and Gee Lee (2017). "Rating Endorsements using Generalized Linear Models," *Variance,* Vol. 10(1).
2. Frees, Edward W., and Gee Lee, and Lu Yang (2015). "Multivariate Frequency-Severity Regression Models in Insurance," *Risks,* Vol. 4(1)
3. Lee, Gee Y., "General Insurance Deductible Ratemaking," *Conditionally Accepted by the North American Actuarial Journal.*

Chapter 2

Multivariate Insurance Claims Modeling

Abstract

In insurance analytics, it is common to have several outcome measures that the analyst wishes to understand using explanatory variables. Often times, outcome measures may be related by common hazards and hence have dependencies. For example, in property insurance, a common hazard may impose risk of accident to the insured's building and contents as well as motor vehicles. It is common to be interested in the frequency of accidents, and the severity of the claim amounts, with such dependencies. Outcomes can be categorized by entity types, time, and space, requiring complex dependency modeling. In this chapter of the dissertation, methods for dependence modeling are reviewed. The work contributes to the statistics and actuarial literature by introducing and applying the 01-inflated negative binomial frequency model, and illustrating how discrete and continuous copula methods can be empirically applied to the insurance claims prediction problem.

This chapter is based on Frees, Edward W., and Gee Lee, and Lu Yang (2015). "Multivariate Frequency-Severity Regression Models in Insurance," *Risks,* Vol. 4(1).

Empirical results are shown in Chapter 5, Section 5.1.

2.1 Introduction to Multivariate Claims Modeling

Many insurance data sets feature information about how often claims arise, the frequency, in addition to the claim size, the severity. Random variables can include:

- N, the number of claims (events),
- $y_k,\ k = 1, ..., N,$ the amount of each claim (loss), and
- $S = y_1 + \cdots + y_N$, the aggregate claim amount.

By convention, the set $\{y_j\}$ is empty when $N = 0$.

Importance of Modeling Frequency. The aggregate claim amount S is the key element for an insurer's balance sheet, as it represents the amount of money paid on claims. So, why do insurance companies regularly track the frequency of claims as well as the claim amounts? As in Frees (2014), the reasons can be categorized into four categories: (i) features of contracts; (ii) policyholder behavior and risk mitigation; (iii) databases that insurers maintain; and (iv) regulatory requirements.

1. Contractually, it is common for insurers to impose deductibles and policy limits on a per occurrence and on a per contract basis. Knowing only the aggregate claim amount for each policy limits any insights one can get into the impact of these contract features.
2. Covariates that help explain insurance outcomes can differ dramatically between frequency and severity. For example, in healthcare, the decision to utilize healthcare by individuals (the frequency) is related primarily to personal characteristics whereas the cost and insurance access per user (the severity) may be more related to characteristics of the healthcare provider (such as the physician). Covariates may also be used to represent risk mitigation activities whose impact varies by frequency and severity. For example, in fire insurance, lightning rods help to prevent an accident (frequency) whereas fire extinguishers help to reduce the impact of damage (severity).
3. Many insurers keep data files that suggest developing separate frequency and severity models. For example, insurers maintain a "policyholder" file that is established when a policy is written. A separate file, often known as the "claims" file, records details of the claim against the insurer, including the amount. These separate databases facilitate separate modeling of frequency and severity.

4. Insurance is a closely monitored industry sector. Regulators routinely require the reporting of claims numbers as well as amounts. Moreover, insurers often utilize different administrative systems for handling small, frequently occurring, reimbursable losses, versus rare occurrence, high impact events. Every insurance claim means that the insurer incurs additional expenses suggesting that claims frequency is an important determinant of expenses.

Importance of Including Covariates. In this dissertation, the interest is in the joint modeling of frequency and severity of claims. In actuarial science, there is a long history of studying frequency, severity and the aggregate claim for homogeneous portfolios; that is, identically and independently distributed realizations of random variables. See any introductory actuarial text, such as Klugman et al. (2012), for an introduction to this rich literature.

Meanwhile, a modeler may be interested in incorporating explanatory variables (covariates, predictors) into the analysis. Historically, this additional information has been available from a policyholder's application form, where various characteristics of the policyholder were supplied to the insurer. For example, in motor vehicle insurance, classic rating variables include the age and sex of the driver, type of the vehicle, region in which the vehicle was driven, and so forth. The current industry trend is towards taking advantage of "big data", with attempts being made to capture additional information about policyholders not available from traditional underwriting sources. An important example is the inclusion of personal credit scores, developed and used in the industry to assess the quality of personal loans, that turn out to also be important predictors of motor vehicle claims experience. Moreover, many insurers are now experimenting with global positioning systems combined with wireless communication to yield real-time policyholder usage data and much more. Through such systems, they gather micro data such as the time of day that the car is driven, sudden changes in acceleration, and so forth. This foray into detailed information is known as "telematics". See, for example, Frees (2015) for further discussion.

Importance of Multivariate Modeling. Reasons for examining insurance outcomes on a multivariate basis is well summarized in Frees et al. (2013). Yet, in Frees et al. (2013), the frequencies were restricted to binary outcomes, corresponding to a claim or no claim, known as "two-part" modeling. In contrast, this dissertation chapter describes more general frequency modeling, although the motivation for examining multivariate outcomes are similar. Analysts and managers gain useful insights by studying the joint behavior of insurance risks, i.e., a multivariate approach:

- For some products, insurers must track payments separately by component to meet contractual obligations. For example, in motor vehicle coverage, deductibles and limits depend on the coverage type, e.g., bodily injury, damage to one's own vehicle, or damage to another party. Hence, it is natural for the insurer to track claims by coverage type.
- For other products, there may be no contractual reasons to decompose an obligation by components and yet the insurer does so to help better understand the overall risk. For example, many insurers interested in pricing homeowners insurance are now decomposing the risk by "peril", or cause of loss. Homeowners insurance is typically sold as an all-risk policy, which covers all causes of loss except those specifically excluded. By decomposing losses into homogenous categories of risk, actuaries seek to get a better understanding of the determinants of each component, resulting in a better overall predictor of losses.
- It is natural to follow the experience of a policyholder over time, resulting in a vector of observations for each policyholder. This special case of multivariate analysis is known as "panel data", see, for example, Frees (2004).
- In the same fashion, policy experience can be organized through other hierarchies. For example, it is common to organize experience geographically and analyze spatial relationships.
- Multivariate models in insurance need not be restricted to only insurance losses. For example, a study of term and whole life insurance ownership is in Frees and Sun (2010). As an example in customer retention, both Brockett et al. (2008); Guillén et al. (2008) advocate for putting the customer at the center of the analysis, meaning that one needs to think about several products that a customer owns simultaneously.

An insurer has a collection of multivariate risks and the interest is managing the distribution of outcomes. Typically, insurers have a collection of tools that can then be used for portfolio management including deductibles, coinsurance, policy limits, renewal underwriting, and reinsurance arrangements. Although pricing of risks can often focus on the mean, with allowances for expenses, profit, and "risk loadings", understanding capital requirements and firm solvency requires understanding of the portfolio distribution. For this purpose, it is important to treat risks as multivariate in order to get an accurate picture of their dependencies.

Dependence and Contagion. We have seen in the above discussion that dependencies arise naturally when modeling insurance data. As a first approximation, we typically think about risks in a portfolio as being independent from one another and rely upon risk pooling to diversify portfolio risk. However, in some cases, risks share common elements such as an epidemic in a population, a natural disaster such as a hurricane that affects many policyholders simultaneously, or an interest rate environment shared by policies with investment elements. These common (pandemic) elements, often known as “contagion”, induce dependencies that can affect a portfolio’s distribution significantly.

Thus, one approach is to model risks as univariate outcomes but to incorporate dependencies through unobserved “latent” risk factors that are common to risks within a portfolio. This approach is viable in some applications of interest. Meanwhile, one can also incorporate contagion effects into more general multivariate approaches. This dissertation will consider situations where data are available to identify models and so we will be able to use the data to guide our decisions when formulating dependence models.

Modeling dependencies is important for many reasons. These include:

1. Dependencies may impact the statistical significance of parameter estimates.
2. Dependence modeling allows for an analyst to obtain the distribution of one variable conditional on another.
3. The degree of dependency affects the degree of reliability of our predictions.
4. To understand the distribution of an insurance product with many identifiable components, one strategy is to describe the distribution of each product and a relationship among the distributions. Insurers may be interested in this approach to assess the possibility of extreme variation in the losses of their insurance products.

2.2 Univariate Foundations

This section summarizes modeling approaches for a single outcome. For notation, define N for the random number of claims, S for the aggregate claim amount, and $\bar{S} = S/N$ for the average claim amount (defined to be 0 when $N = 0$). To model these outcomes, we use a collection of covariates $\mathbf{x}$, some of which may be useful for frequency modeling whereas others will be useful for severity modeling. The dependent variables N, S, and $\bar{S}$ as well as covariates $\mathbf{x}$ vary by the risk $i = 1, \ldots, n$. For each risk, we also are interested in multivariate outcomes indexed by $j = 1, \ldots, p$. So, for example, $\mathbf{N}_i = (N_{i1}, \ldots, N_{ip})'$ represents the vector of p claim outcomes from the ith risk. A more detailed review for the single outcome, $p = 1$ case, can be found in Frees (2014).

2.2.1 Frequency-Severity

For modeling the joint outcome (N, S), or equivalently, $(N, \bar{S})$, it is customary to first condition on the frequency and then model the severity. Suppressing the $\{i\}$ subscript, we decompose the distribution of the dependent variables as:

$$\begin{aligned} \mathrm{f}(N,S) &= \mathrm{f}(N) \times \mathrm{f}(S|N) \\ \text{joint} &= \text{frequency} \times \text{conditional severity}, \end{aligned} \tag{2.1}$$

where $\mathrm{f}(N, S)$ denotes the joint distribution of (N, S). Through this decomposition, we do *not* require independence of the frequency and severity components.

There are many ways to model dependence when considering the joint distribution $\mathrm{f}(N, S)$ in Equation (2.1). For example, one may use a latent variable that affects both frequency N and loss amounts S, thus inducing a positive association. Copulas are another tool used regularly by actuaries to model non-linear associations and will be described in subsequent Section 2.4. The conditional probability framework is a natural method of allowing for potential dependencies and provides a good starting platform for empirical work.

2.2.2 Modeling Frequency

It has become routine for actuarial analysts to model the frequency N_i based on covariates $\mathbf{x}_i$ using generalized linear models, GLMs, cf., De Jong and Heller (2008). For binary outcomes, logit and probit forms are most commonly used, cf., Guillén (2014). For count outcomes, one begins with a Poisson or negative binomial distribution. Moreover, to handle the excessive number of zeros relative to that implied by these distributions, analysts routinely examine zero-inflated models, as described in Boucher (2014).

A strength of GLMs relative to other non-linear models is that one can express the mean as a simple function of linear combinations of the covariates. In insurance, it is common to use a "logarithmic link" for this function and so express the mean as $\mu_i = \mathrm{E}[N_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is a vector of parameters associated with the covariates. This function is used because it yields desirable parameter interpretations, seems to fit data reasonably well, and ties well with other approaches traditionally used in actuarial ratemaking applications. See Mildenhall (1999).

It is also common to identify one of the covariates as an "exposure" that is used to calibrate the size of a potential outcome variable. In frequency modeling, the mean is assumed to vary proportionally with E_i, for exposure. To incorporate exposures, we specify one of the explanatory variables to be $\ln E_i$ and restrict the corresponding regression coefficient to be 1; this term is known as an *offset*. With this convention, we have

$$\ln \mu_i = \ln E_i + \mathbf{x}_i'\boldsymbol{\beta} \Leftrightarrow \frac{\mu_i}{E_i} = \exp(\mathbf{x}_i'\boldsymbol{\beta}).$$

When there are excessive numbers of 0 s and 1 s in the outcome, a "zero-one-inflated" model can be used. As an extension of the zero-inflated method, a zero-one-inflated model employs two generating processes. The first process is governed by a multinomial distribution that generates structural zeros and ones. The second process is governed by a Poisson or negative binomial distribution that generates counts, some of which may be zero or one.

Denote the latent variable in the first process as $I_i, i = 1, \ldots, 2$, which follows a multinomial distribution with possible values 0, 1 and 2 with corresponding probabilities $\pi_{0,i}, \pi_{1,i}, \pi_{2,i} = 1 - \pi_{0,i} - \pi_{1,i}$. Here, N_i is frequency. With this, the probability mass function of N_i is

$$f_{N,i}(n) = \pi_{0,i} I_{\{n=0\}} + \pi_{1,i} I_{\{n=1\}} + \pi_{2,i} P_i(n),$$

where $P_i(n)$ may be the density for a Poisson or negative binomial distribution. A logit specification is used to parameterize the probabilities for the latent variable I_i. Denote the covariates associated with I_i as $\mathbf{z}_i$. A logit specification is used to parameterize the probabilities for the latent variable I_i. Using level 2 as the reference category, the specification is

$$log \frac{\pi_{j,i}}{\pi_{2,i}} = \mathbf{z}_i'\boldsymbol{\gamma}_j, j = 0, 1.$$

Correspondingly,

$$\pi_{j,i} = \frac{\exp(\mathbf{z}_i'\boldsymbol{\gamma}_j)}{1+\exp(\mathbf{z}_i'\boldsymbol{\gamma}_j)+\exp(\mathbf{z}_i'\boldsymbol{\gamma}_j)}, j=0,1.$$

$$\pi_{2,i} = 1-\pi_{0,i}-\pi_{1,i}$$

Maximum likelihood estimation is used to fit the parameters.

2.2.3 Modeling Severity

Modeling Severity Using GLMs. For insurance analysts, one strength of the GLM approach is that the same set of routines can be used for continuous as well as discrete outcomes. For severities, it is common to use a gamma or inverse Gaussian distribution, often with a logarithmic link (primarily for parameter interpretability).

The linear exponential family forms the basis of GLMs. One strength of the linear exponential family is that a sample average of outcomes comes from the same distribution as the outcomes. Specifically, suppose that we have m independent variables from the same distribution with location parameter θ and scale parameter ϕ. Then, the sample average comes from the same distributional family with location parameter θ and scale parameter ϕ/m. This result is helpful as insurance analysts regularly face grouped data as well as individual data. For example, Frees (2014) provides a demonstration of this basic property.

To illustrate, in the aggregate claims model, if individual losses have a gamma distribution with mean μ_i and scale parameter ϕ, then, conditional on observing N_i losses, the average aggregate loss $\bar{S}_i$ has a gamma distribution with mean μ_i and scale parameter ϕ/N_i.

Generalized Beta Family. The GLM is the workhorse for industry analysts interested in analyzing the severity of claims. Naturally, because of the importance of claims severity, a number of alternative approaches have been explored, cf., Shi (2014) for an introduction. This section reviews the generalized beta family models for insurance loss severities. Commonly used distributions are exponential, gamma, and Pareto distributions, where the practice is to replace the location parameter with $\boldsymbol{x}'\boldsymbol{\beta}$. For this, it is common to assume a parametric model with a logarithmic link for parameter interpretability. The generalized beta random variable Y has the density

$$f_Y(y; a, b, c, \alpha_1, \alpha_2) = \frac{|a| y^{a \cdot \alpha_1 - 1}(1 - (1-c)(y/b)^a)^{\alpha_2 - 1}}{b^{a \cdot \alpha_1} B(\alpha_1, \alpha_2)(1 + c(y/b)^a)^{\alpha_1 + \alpha_2}},$$

where $0 < c < 1$, and $b, \alpha_1, \alpha_2 > 0$. Here, $B(\alpha_1, \alpha_2)$ is the beta function. The generalized beta family contains many familiar distributions as special cases: *GB1*, *GB2*, gamma, generalized gamma, Weibull, Burr type 3, Burr type 12, Dagum, log-normal, Lomax (Pareto Type II), *F*, Rayleigh, chi-square, half-normal, half-Student-t, exponential and log-logistic. Klugman et al. (2012) provides an introductory overview of the generalized beta family distributions. Each special case is obtained by restricting the parameter of the distribution to a specific value or taking the limiting case of the parameter. Some special cases have been more popular than others in the loss-modeling context. The *GB* family is defined for response values between

$$0 < y^a < \frac{b^a}{1-c}.$$

Yet limiting cases are defined for arbitrarily large y values. The *GB1* distribution is obtained by restricting $c = 0$, while the *GB2* distribution is obtained by restricting $c = 1$ from the *GB* family. The *GB2* density provides a flexible class of distributions for insurance loss modeling, and is described in detail next.

Modeling Severity Using GB2. A random variable with a “generalized beta of the second kind”, or in short *GB2*, distribution can be written as

$$e^{\mu}\left(\frac{G_1}{G_2}\right)^{\sigma} = C_1 e^{\mu} F^{\sigma} = e^{\mu}\left(\frac{Z}{1-Z}\right)^{\sigma},$$

where the constant $C_1 = (\alpha_1/\alpha_2)^{\sigma}$, G_1 and G_2 are independent gamma random variables with scale parameter 1 and shape parameters α_1 and α_2, respectively. Further, the random variable F has an F-distribution with degrees of freedom $2\alpha_1$ and $2\alpha_2$, and the random variable Z has a beta distribution with parameters α_1 and α_2. The density is

$$f_Y(y; a, b, \alpha_1, \alpha_2) = \frac{|a|(y/b)^{a \cdot \alpha_1}}{y B(\alpha_1, \alpha_2)[1 + (y/b)^a]^{\alpha_1 + \alpha_2}},$$

and is defined for

$$0 < y < \infty.$$

For incorporating covariates, it is straightforward to show that the regression function is of the form

$$\mathrm{E}\left(y|\mathbf{x}\right) = C_2 \exp\left(\mu\left(\mathbf{x}\right)\right) = C_2\ e^{\mathbf{x}'\boldsymbol{\beta}},$$

where the constant C_2 can be calculated with other (non-location) model parameters. Under the most commonly used way of parametrization for the GB2, where μ is associated with covariates, if $-\alpha_1 < \sigma < \alpha_2$, then we have

$$C_2 = \frac{B(\alpha_1 + \sigma, \alpha_2 - \sigma)}{B(\alpha_1, \alpha_2)}$$

where $B(\alpha_1, \alpha_2) = \Gamma(\alpha_1)\Gamma(\alpha_2)/\Gamma(\alpha_1 + \alpha_2)$. Thus, one can interpret the regression coefficients in terms of a proportional change. That is, $\partial\left[\ln \mathrm{E}(y)\right]/\partial x_k = \beta_k$. In principle, one could allow for any distribution parameter to be a function of the covariates. However, following this principle would lead to a large number of parameters; this typically yields computational difficulties as well as problems of applications, Sun et al. (2008). An alternative way to choose the location parameter in *GB2* is through log linear model in Prentice (1975). In this dissertation, μ is used to incorporate covariates. The density of $GB2(\sigma, \mu, \alpha_1, \alpha_2)$ is

$$f(y; \mu, \sigma, \alpha_1, \alpha_2) = \frac{[\exp(z)]^{\alpha_1}}{y\sigma B(\alpha_1, \alpha_2)[1 + \exp(z)]^{\alpha_1+\alpha_2}}$$

where $z = \frac{\ln(y)-\mu}{\sigma}$. As pointed out in Yang (2011), if

$$Y \sim GB2(\sigma, \mu, \alpha_1, \alpha_2),$$

$$\log(Y) = \mu + \sigma(\log\ \alpha_1 - \log\ \alpha_2) + \sigma\log\ F(2\alpha_1, 2\alpha_2).$$

This is actually the log linear model used with errors following the log-F distribution. Thus $\mu + \sigma(\log\ \alpha_1 - \log\ \alpha_2)$ can be used as the location parameter associated with covariates.

Thus, the *GB2* family has four parameters (α_1, α_2, μ and σ), where μ is the location parameter. The *GB2* includes limiting distributions such as the generalized gamma, exponential, Weibull, and so forth. It also encompasses the "Burr Type 12" (by allowing $\alpha_1 = 1$), as well as other families of interest, including the Pareto distributions. The *GB2* is a flexible distribution that accommodates positive or negative skewness, as well as heavy tails. See, for example, Klugman et al. (2012) for an introduction to these distributions.

As a special case of *GB2*, the location parameter of *GG* (generalized gamma) can be derived based on *GB2*. Let $\alpha_2 \to \infty$, and the density of $GG(a, b, \alpha_1)$ is

$$GG(y; a, b, \alpha_1) = \frac{a}{\Gamma(\alpha_1) y} (y/b)^{a\alpha_1} e^{-(y/b)^a}.$$

Reparametrizing the $GB2(a, b, \alpha_1, \alpha_2))$ with $a = \frac{1}{\sigma}$, $b = \exp(\mu)$, we have

$$GG(a, b, \alpha_1) = \lim_{\alpha_2 \to \infty} GB2(a, b\alpha_2^{1/a}, \alpha_1, \alpha_2).$$

The location parameter for $GG(a, b, \alpha_1)$ should be $\log(b) + \sigma\log(\alpha_2) + \sigma(\log(\alpha_1) - \log(\alpha_2)) = \log(b) + \sigma\log(\alpha_1)$. This is consistent with the results in Prentice (1974). When $a = 1$, the *GG* distribution becomes the gamma distribution with shape parameter α_1 and scale parameter b. $\log(b) + \log(\alpha_1)$ is the location parameter, which is the log-mean of the gamma distribution, and is hence consistent with the GLM framework.

2.2.4 Tweedie Model

Frequency-severity modeling is widely used in insurance applications. However, for simplicity, it is also common to use the aggregate loss S as a dependent variable in a regression. Because the distribution of S typically contains a positive mass at zero representing no claims, and a continuous component for positive values representing the amount of a claim, a widely used mixture is the Tweedie (1984) distribution. The Tweedie distribution is defined as a Poisson sum of gamma random variables. Specifically, suppose that N has a Poisson distribution with mean λ, representing the number of claims. Let Y_j be an i.i.d. sequence, independent of N, with each Y_j having a gamma distribution with parameters α and β, representing the amount of a claim. Note, β is standard notation for this parameter used in loss-model textbooks, and the reader should understand it is different from the bold-faced $\boldsymbol{\beta}$, as the latter is a symbol we will use for the coefficients corresponding to explanatory variables. Then, $S = y_1 + \ldots + y_N$ is a Poisson sum of random gammas.

To understand the mixture aspect of the Tweedie distribution, first note that it is straightforward to compute the probability of zero claims as $\Pr(S = 0) = \Pr(N = 0) = e^{-\lambda}$. The distribution function can be computed using conditional expectations,

$$\Pr(S \le y) = e^{-\lambda} + \sum_{n=1}^{\infty} \Pr(N = n) \Pr(S_n \le y), \qquad y \ge 0.$$

Because the sum of i.i.d. gammas is a gamma, $S_n = y_1 + \ldots + y_n$ (not S) has a gamma distribution with parameters $n\alpha$ and β. For $y > 0$, the density of the Tweedie distribution is

$$\mathrm{f}_S(y) = \sum_{n=1}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \frac{\beta^{n\alpha}}{\Gamma(n\alpha)} y^{n\alpha-1} e^{-y\beta}.$$

From this, straight-forward calculations show that the Tweedie distribution is a member of the linear exponential family. Now, define a new set of parameters μ, ϕ, P through the relations

$$\lambda = \frac{\mu^{2-P}}{\phi(2-P)}, \qquad \alpha = \frac{2-P}{P-1} \qquad \text{and} \qquad \frac{1}{\beta} = \phi(P-1)\mu^{P-1}.$$

Easy calculations show that

$$\mathrm{E}[S] = \mu \qquad \text{and} \qquad \mathrm{Var}[S] = \phi\mu^P,$$

where $1 < P < 2$. The Tweedie distribution can also be viewed as a choice that is intermediate between the Poisson and the gamma distributions.

In the basic form of the Tweedie regression model, the scale (or dispersion) parameter ϕ is constant. However, if one begins with the frequency-severity structure, calculations show that ϕ depends on the risk characteristics i, cf., Frees (2014). Because of this and the varying dispersion (heteroscedasticity) displayed by many data sets, researchers have devised ways of accommodating and/or estimating this structure. The most common way is the so-called “double GLM” procedure proposed in Smyth (1989) that models the dispersion as a known function of a linear combination of covariates (as well as the mean, hence the name “double GLM”).

2.3 Multivariate Models and Methods

2.3.1 Copula Regression

Copulas have been applied with GLMs in the biomedical literature since the mid-1990s; Meester and Mackay (1994); Lambert (1996); Song (2000). In the actuarial literature, the t-copula and the Gaussian copula with GLMs as marginal distributions were used to develop credibility predictions in Frees and Wang (2005). In more general cases, Song (2007) provides a detailed introduction of copula regression that focuses on the Gaussian copula and Kolev and Paiva (2009) surveys copula regression applications.

Introducing Copulas. Specifically, a *copula* is a multivariate distribution function with uniform marginals. Let $U_1, \ldots, U_p$ be p uniform random variables on $(0, 1)$. Their joint distribution function

$$\mathrm{C}(u_1, \ldots, u_p) = \Pr(U_1 \leq u_1, \ldots, U_p \leq u_p) \tag{2.2}$$

is a copula. We seek to use copulas in applications that are based on more than just uniformly distributed data. Thus, consider arbitrary marginal distribution functions $\mathrm{F}_1(y_1), \ldots, \mathrm{F}_p(y_p)$. Then, we can define a multivariate distribution function using the copula such that

$$\mathrm{F}(y_1, \ldots, y_p) = \mathrm{C}(\mathrm{F}_1(y_1), \ldots, \mathrm{F}_p(y_p)). \tag{2.3}$$

If outcomes are continuous, then we can differentiate the distribution functions and write the density function as

$$\mathrm{f}(y_1, \ldots, y_p) = \mathrm{c}(\mathrm{F}_1(y_1), \ldots, \mathrm{F}_p(y_p)) \prod_{j=1}^{p} \mathrm{f}_j(y_j), \tag{2.4}$$

where f_j is the density of the marginal distribution F_j and c is the copula density function.

It is easy to check from the construction in Equation (2.3) that $\mathrm{F}(\cdot)$ is a multivariate distribution function. Sklar established the converse in Sklar (1959). He showed that any multivariate distribution function $\mathrm{F}(\cdot)$ can be written in the form of Equation (2.3), that is, using a copula representation. Sklar also showed that, if the marginal distributions are continuous, then there is a unique copula representation. See, for example, Nelsen (1999) and Frees and Valdez (1998) for an introduction to copulas, and Joe (2014) for a comprehensive modern treatment.

Regression with Copulas. In a regression context, we assume that there are covariates $\mathbf{x}$ associated with outcomes $\mathbf{y} = (y_1, \ldots, y_p)'$. In a parametric context, we can incorporate covariates by allowing them to be functions of the distributional parameters.

Specifically, we assume that there are n independent risks and p outcomes for each risk i, $i = 1, \ldots, n$. For this section, consider an outcome $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})'$ and $K \times 1$ vector of covariates $\mathbf{x}_i$, where K is the number of covariates. The marginal distribution of y_{ij} is a function of $\mathbf{x}_{ij}$, $\boldsymbol{\beta}_j$ and $\boldsymbol{\theta}_j$. Here, $\mathbf{x}_{ij}$ is a $K_j \times 1$ vector of explanatory variables for risk i and outcome type j, a subset of $\mathbf{x}_i$, and $\boldsymbol{\beta}_j$ is a $K_j \times 1$ vector of marginal parameters to be estimated. The systematic component $\mathbf{x}_{ij}'\boldsymbol{\beta}_j$ determines the location parameter. The vector $\boldsymbol{\theta}_j$ summarizes additional parameters of the marginal distribution that determine the scale and shape. Let $F_{ij} = F(y_{ij}; \mathbf{x}_{ij}, \boldsymbol{\beta}_j, \boldsymbol{\theta}_j)$ denote the marginal distribution function.

This describes a classical approach to regression modeling, treating explanatory variances/covariates as non-stochastic ("fixed") variables. An alternative is to think of the covariates themselves as random and perform statistical inference conditional on them. Some advantages of this alternative approach are that one can model the time-changing behavior of covariates, as in Patton (2006), or investigate non-parametric alternatives, as in Acar et al. (2011). These represent excellent future steps in copula regression modeling that are not addressed further in this article.

2.3.2 Multivariate Severity

For continuous severity outcomes, we may consider the density function $\mathrm{f}_{ij} = \mathrm{f}(y_{ij}; \mathbf{x}_{ij}, \boldsymbol{\beta}_j, \boldsymbol{\theta}_j)$ associated with the distribution function F_{ij} and c the copula density function with parameter vector $\boldsymbol{\alpha}$. With this, using Equation (2.4), the log-likelihood function of the ith risk is written as

$$l_i(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_{j=1}^{p} \ln \mathrm{f}_{ij} + \mathrm{c}(\mathrm{F}_{i1}, \ldots, \mathrm{F}_{ip}; \boldsymbol{\alpha}), \tag{2.5}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p)$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p)$ are collections of parameters over the p outcomes. This is a fully parametric set-up; the usual maximum likelihood techniques enjoy certain optimality properties and are the preferred estimation method.

If we consider only a single outcome, say y_{i1}, then the associated log-likelihood is $\ln \mathrm{f}_{i1}$. Thus, the set of outcomes $y_{11}, \ldots, y_{n1}$ allows for the usual "root-n" consistent estimator of $\boldsymbol{\beta}_1$, and similarly for the other outcomes $y_{ij}, j = 2 \ldots, p$. By considering each outcome in isolation of the others, we can get desirable estimators of the regression coefficients $\boldsymbol{\beta}_j, j = 1, \ldots, p$. These provide excellent starting values to calculate the fully efficient maximum likelihood estimators using the log-likelihood from Equation (2.5). Joe coined the phrase "inference for margins", sometimes known by the acronym *IFM* in Joe (1997), to describe this approach to estimation.

In the same way, one can consider any pair of outcomes, (y_{ij}, y_{ik}) for $j \neq k$. This permits consistent estimation of the marginal regression parameters as well as the association parameters between the jth and kth outcomes. As with the IFM, this technique provides excellent starting values of a fully efficient maximum likelihood estimation recursion. Moreover, they provide the basis for an alternative estimation method known as "composite likelihood", cf., Song et al. (2013) or Nikoloulopoulos (2013a), for a description in a copula regression context.

2.3.3 Multivariate Frequency

If outcomes are discrete, then one can take differences of the distribution function in Equation (2.3) to write the probability mass function as

$$\mathrm{f}(y_1, \ldots, y_p) = \sum_{j_1=1}^{2} \cdots \sum_{j_p=1}^{2} (-1)^{j_1 + \cdots + j_p} \, \mathrm{C}(u_{1,j_1}, \ldots, u_{p,j_p}). \tag{2.6}$$

Here, $u_{j,1} = \mathrm{F}_j(y_j-)$ and $u_{j,2} = \mathrm{F}_j(y_j)$ are the left- and right-hand limits of F_j at y_j, respectively. For example, when $p = 2$, we have

$$\begin{aligned} \mathrm{f}(y_1, y_2) &= \mathrm{C}(\mathrm{F}_1(y_1), \mathrm{F}_2(y_2)) - \mathrm{C}(\mathrm{F}_1(y_1-), \mathrm{F}_2(y_2)) \\ &\quad - \mathrm{C}(\mathrm{F}_1(y_1), \mathrm{F}_2(y_2-)) + \mathrm{C}(\mathrm{F}_1(y_1-), \mathrm{F}_2(y_2-)). \end{aligned}$$

It is straightforward in principle to estimate parameters using Equation (2.6) and standard maximum likelihood theory.

In practice, two caveats should be mentioned. The first is that the result of Sklar (1959) only guarantees that the copula is unique over the range of the outcomes, a point emphasized with several interesting examples in Genest and Nešlehová (2007). In a regression context, this nonidentifiability is less likely to be a concern, as noted in Joe (2014); Song et al. (2013); Nikoloulopoulos and Karlis (2010); Nikoloulopoulos (2013a). Moreover, the last reference emphasizes that the Gaussian copula with binary data has been used for decades by researchers as this is just another form for the commonly used multivariate probit.

The second issue is computational. As can be seen in Equation (2.6), likelihood inference involves the computation of multidimensional rectangle probabilities. The review article Nikoloulopoulos (2013a) describes several variations of maximum likelihood that can be useful as the dimension p increases, see also Nikoloulopoulos (2013b). As in Genest et al. (2013), the composite likelihood method is used for computation. For large values of p, the pair (also known as "vine") copula approach described in Panagiotelis et al. (2012) for discrete outcomes seems to be a promising approach.

2.3.4 Multivariate Tweedie

As emphasized in Joe (2014) (p. 226), in copula regression it is possible to have outcomes that are combinations of continuous, discrete, and mixture distributions. One case of special interest in insurance modeling is the multivariate Tweedie, where each marginal distribution is a Tweedie and the margins are joined by a copula. Specifically, Shi (2016) considers different types of insurance coverages with Tweedie margins.

To illustrate the general principles, consider the bivariate case ($p = 2$). Suppressing the i index and covariate notation, the joint distribution is

$$\mathrm{f}(y_1, y_2) = \begin{cases} C(F_1(0), F_2(0)) & y_1 = 0, y_2 = 0 \\ f(y_1)\partial_1 C(F_1(y_1), F_2(0)) & y_1 > 0, y_2 = 0 \\ f(y_2)\partial_2 C(F_1(0), F_2(y_2)) & y_1 = 0, y_2 > 0 \\ f(y_1)f(y_2)c(F_1(y_1), F_2(y_2)) & y_1 > 0, y_2 > 0 \end{cases},$$

where $\partial_j C$ denotes the partial derivative of copula with respect to the jth component. See Shi (2016) for additional details of this estimation where he also described a double GLM approach to accommodate varying dispersion parameters.

2.3.5 Association Structures and Elliptical Copulas

First consider an outline of the evolution of multivariate regression modeling.

1. The multivariate normal (Gaussian) distribution has provided a foundation for multivariate data analysis, including regression. By permitting a flexible structure for the mean, one can readily incorporate complex mean structures including high order polynomials, categorical variables, interactions, semi-parametric additive structures, and so forth. Moreover, the variance structure readily permits incorporating time series patterns in panel data, variance components in longitudinal data, spatial patterns, and so forth. One way to get a feel for the breadth of variance structures readily accommodated is to examine options in standard statistical software packages such as `PROC Mixed` in SAS (2010) (for example, the `TYPE` switch in the `RANDOM` statement permits the choice of over 40 variance patterns).
2. In many applications, appropriately modeling the mean and second moment structure (variances and covariances) suffices. However, for other applications, it is important to recognize the underlying outcome distribution and this is where copulas come into play. As we have seen, copulas are available for any distribution function and thus readily accommodate binary, count, and long-tail distributions that cannot be adequately approximated with a normal distribution. Moreover, marginal distributions need not be the same, e.g., the first outcome may be a count Poisson distribution and the second may be a long-tail gamma.
3. Pair copulas (cf., Joe (2014)) may well represent the next step in the evolution of regression modeling. A copula imposes the same dependence structure on all p outcomes whereas a pair copula has the flexibility to allow the dependence structure itself to vary in a disciplined way. This is done by focusing on the relationship between pairs of outcomes and examining conditional structures to form the dependence of the entire vector of outcomes. This approach is useful for high dimensional outcomes (where p is large), an important developing area of statistics. This represents an excellent future step in copula regression modeling. See Frees and Lee (2017).

As described in Joe (2014), there is a host of copulas available depending on the interests of the analyst and the scientific purpose of the investigation. Considerations for the choice of a copula may include computational convenience, interpretability of coefficients, a latent structure for interpretability, and a wide range of dependence, allowing both positive and negative associations.

For our applications of regression modeling, we typically begin with the elliptical copula family. This family is based on the family of elliptical distributions that includes the multivariate normal and t-distributions. See more in Frees and Wang (2006).

This family has most of the desirable traits that one would seek in a copula family. From our perspective, the most important feature is that it permits the same family of association matrices found in the multivariate Gaussian distribution. This not only allows the analyst to investigate a wide degree of association patterns, but also allows estimation to be accomplished in a familiar way using the same structure as in the Gaussian family, e.g., SAS (2010).

For example, if the ith risk evolves over time, we might use a familiar time series model to represent associations, e.g.,

$$\mathbf{\Sigma}_{AR1}(\rho) = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$

such as an autoregressive of order 1 (AR1). See Frees and Wang (2005) for an actuarial application.

For a more complex example, suppose that $\mathbf{y}_i = (\mathbf{y}_{i1}, \mathbf{y}_{i2}, \mathbf{y}_{i3})'$ represents three types of expenses for the ith company observed over 4 time periods. Then, we might use the following dependence structure

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{AR1}(\rho_1) & \sigma_{12}\mathbf{\Sigma}_{12} & \sigma_{13}\mathbf{\Sigma}_{13} \\ \sigma_{12}\mathbf{\Sigma}'_{12} & \mathbf{\Sigma}_{AR1}(\rho_2) & \sigma_{23}\mathbf{\Sigma}_{23} \\ \sigma_{13}\mathbf{\Sigma}'_{13} & \sigma_{23}\mathbf{\Sigma}'_{23} & \mathbf{\Sigma}_{AR1}(\rho_3) \end{pmatrix},$$

as in Shi (2012). This is a commonly used specification in models of several time series in econometrics where σ_{jk} represents a cross-sectional association between $\mathbf{y}_{ij}$ and $\mathbf{y}_{ik}$ and $\mathbf{\Sigma}_{jk}$ represents cross-associations with time lags.

2.3.6 Assessing Dependence

Dependence can be assessed at all stages of the model fitting process:

- Copula identification begins after marginal models have been fit. Then, use the "Cox-Snell" residuals from these models to check for association. Create simple correlation statistics (Spearman, polychoric) as well as plots (*pp* and tail dependence plots) to look for dependence structures and identify a parametric copula.
- After a model identification, estimate the model and examine how well the model fits. Examine the residuals to search for additional patterns using, for example, correlation statistics and t-plot (for elliptical copulas). Examine the statistical significance of fitted association parameters to seek a simpler fit that captures the important tendencies of the data.
- Compare the fitted model to alternatives. Use overall goodness of fit statistics for comparisons, including AIC and BIC, as well as cross-validation techniques. For nested models, compare via the likelihood ratio test and use Vuong's procedure for comparing non-nested alternative specifications.
- Compare the models based on a held-out sample. Use statistical measures and economically meaningful alternatives such as the Gini statistic.

In the first and second step of this process, a variety of hypothesis tests and graphical methods can be employed to identify the specific type of copula (e.g., Archmedian, elliptical, extreme value, and so forth) that corresponds to the given data. Researchers have developed a graphical tool called the Kendall plot, or the K-plot for short, to detect dependence. See Genest and Boies (2003). To determine whether a joint distribution corresponds to an Archimedean copula or a specific extreme-value copula, goodness-of-fit tests developed by Genest and Rivest (1993); Kojadinovic et al. (2011); Genest et al. (2011); Bahraoui et al. (2014) can be helpful. The reader may also reference Joe (2014) for a comprehensive coverage of the various assessment methods for dependence.

2.3.7 Frequency-Severity Modeling Strategy

Multivariate frequency-severity modeling strategies are a subset of the usual regression and copula identification and inference strategies. In absence of a compelling theory to suggest the appropriate covariates and predictors (which is generally the case for insurance applications), the modeling strategy consists of model identification, estimation, and inference. Typically, this is done in a recursive fashion where one sets aside a random portion of the data for identification and estimation (the "training" sample), and one proceeds to validate and conduct inference on another portion (the "test" sample). See, for example, Hastie et al. (2009) for a description of this and many procedures for variable selection, mainly in a cross-sectional regression context. One may think about the identification and estimation procedures as three components in a copula regression model:

1. Fit the mean structure. Historically, this is the most important aspect. One can apply robust standard error procedures to get consistent and approximately normally distributed coefficients, assuming a correct mean structure.
2. Fit the variance structure with a selected distribution. In GLMs, the choice of the distribution dictates the variance structure that can be over-ruled with a separately specified variance, e.g., a "double GLM."
3. Fit the dependence structure with a choice of copula.

For frequency-severity modeling, there are two mean and variance structures to work with, one for the frequency and one for the severity.

Identification and Estimation

Although the estimation of parametric copulas is fairly established, the literature on identification of copulas is still in the early stage of development. As described in Section 2.3.2, maximum likelihood is the usual choice with an inference for margins and/or composite likelihood approach for starting values of the iterations. A description of composite likelihood in the context of copula modeling can be found in Joe (2014). As noted here, composite likelihood may be particularly useful for multivariate discrete data when univariate margins have common parameters; page 233, Joe (2014). Another variation of maximum likelihood in copula regression is the "maximization by parts" method, as described in Song (2007) and utilized in Czado et al. (2012). In the context of copula regressions, the idea behind this is to split the likelihood into two pieces, an easier part corresponding to the marginals and a more difficult part corresponding to the copula. The estimation routine takes advantage of these differing levels of difficulty in the calculations.

Identification of copula models used in regression typically starts with residuals from marginal fits. For severities, the idea is to estimate a parametric fit to the marginal distribution, such as normal or gamma regression. Then, one applies the distribution function (that depends on covariates) to the observation. Using notation, we can write this as $F_i(y_i) = \hat{\epsilon}_i$. This is known as the "probability integral transformation." If the model is correctly specified, then the $\hat{\epsilon}_i$ has a uniform (0,1) distribution. This is an idea that dates back to works by Cox and Snell (1968) and so these are often known as "Cox-Snell" residuals. For copula identification, it is recommended in Joe (2014) to take an inverse normal distribution transform (i.e., $\Phi^{-1}(\hat{\epsilon}_i)$, for a standard normal distribution function Φ) to produce "normal scores."

Because of the discreteness with frequencies, these residuals are not uniformly distributed even if the model is correctly specified. In this case, one can "jitter" the residuals. Specifically, define a modified distribution function $F_i(y, \lambda) = \Pr(Y_i < y) + \lambda \Pr(Y_i = y)$ and let V be a uniform random number that is independent of Y_i. Then, we can define the jittered residual to be $F_i(y_i, V) = \tilde{\epsilon}_i$. If the model is correctly specified, then jittered residuals have a uniform (0,1) distribution, cf., Rüschendorf (2009).

Compared to classical residuals, residuals from probability integral transforms have less ability to guide model development—we can only tell if the marginal models are approximately correct. The main advantage of this residual is that it is applicable to all (parametric) distributions. If you are working with a distribution that supports other definitions of residuals, then these are likely to be more useful because they may tell you how to improve your model specification, not whether or not it is approximately correct. If the marginal model fit is adequate, then we can think of the residuals as approximate realizations from a uniform distribution and use standard techniques from copula theory to identify a copula. We refer to Joe (2014) for a summary of this literature.

Model Validation

After identification and estimation, it is customary to compare a number of alternative models based on the training and on the test samples. For the training sample, the "in-sample" comparisons are typically based on the significance of the coefficients, overall goodness of fit measures (including information criteria such as AIC and BIC), cross-validation, as well as likelihood ratio comparisons for nested models.

For comparisons among non-nested parametric models, it is now common in the literature to cite a statistic due to Vuong (1989). For this statistic, one calculates the contribution to the logarithmic likelihood such as in Equation (2.5) for two models, say, $l_i^{(1)}$ and $l_i^{(2)}$. One prefers Model (1) compared to Model (2) if the average difference, $\bar{D} = m^{-1}\sum_{i=1}^{m} D_i$, is positive, where $D_i = l_i^{(1)} - l_i^{(2)}$ and m is the size of the validation sample. To assess the significance of this difference, one can apply approximate normality with approximate standard errors given as $SD_D/\sqrt{m}$ where $SD_D^2 = (m-1)^{-1}\sum_{i=1}^{m}(D_i - \bar{D})^2$. In a copula context, see; page 257, Joe (2014) for a detailed description of this procedure, where sample size adjustments similar to those used in AIC and BIC are also introduced.

Comparison among models using test data, or "out-of-sample" comparisons are also important in insurance because many of these models are used for predictive purposes such as setting rates for new customers. Out-of-sample measures compare held-out observations to those predicted by the model. Traditionally, absolute values and squared differences have been used to summarize differences between these two. However, for many insurance data sets, there are large masses at zero, meaning that these traditional metrics are less helpful. To address this problem, a newer measure is developed in Frees et al. (2011b) that they call the "Gini index." In this context, the Gini index is twice the average covariance between the predicted outcome and the rank of the predictor. In order to compare models, Theorem 5 of Frees et al. (2011b) provides standard errors for the difference of two Gini indices.

2.4 Frequency Severity Dependency Models

In traditional models of insurance data, the claim frequency is assumed to be independent of claim severity. However, the average severity may depend on frequency, even when this classical assumption holds.

One way of modeling the dependence is through the conditioning argument developed in Section 2.2.1. An advantage of this approach is that the frequency can be used as a covariate to model the average severity. See Frees et al. (2011a) for a healthcare application of this approach. For another application, a Bayesian approach for modeling claim frequency and size was proposed in Gschlößl and Czado (2007), with both covariates as well as spatial random effects taken into account. The frequency was incorporated into the severity model as covariate. In addition, they checked both individual and average claim modeling and found the results were similar in their application.

As an alternative approach, copulas are widely used for frequency severity dependence modeling. In Czado et al. (2012), the authors fit Gaussian copula on Poisson frequency and gamma severity and used an optimization by parts method from Song et al. (2005) to do the estimation. They derived the conditional distribution of frequency given severity. In Krämer et al. (2013), the distribution of policy loss is derived without the independence assumption between frequency and severity. They also showed that the ignoring of dependence can lead to underestimation of loss. A Vuong's test was adopted to select the copula.

To see how the copula approach works, recall that $\bar{S}$ represents average severity of claims and N denotes frequency. Using a copula, we can express the likelihood as

$$f_{\bar{S},N}(s,n) = \begin{cases} f_{\bar{S},N}(s,n|N>0)P(N>0) & \text{for } n>0 \\ P(N=0) & \text{for } n=s=0 \end{cases}$$

Denote

$$D_1(u,v) = \frac{\partial}{\partial u}C(u,v) = P(V \le v|U=u).$$

With this,

$$\begin{aligned} P(\bar{S}=s, N \le n|N>0) &= \frac{\partial}{\partial s}P(\bar{S} \le s, N \le n|N>0) \\ &= \frac{\partial}{\partial s}C(F_{\bar{S}}(s), F_N(n|N>0)) \\ &= f_{\bar{S}}(s)D_1(F_{\bar{S}}(s), F_N(n|N>0)). \end{aligned}$$

This yields the following expression for the likelihood

$$f_{\bar{S},N}(s,n) = \begin{cases} f_{\bar{S}}(s)P(N>0)(D_1(F_{\bar{S}}(s), F_N(n|N>0)) & \\ \qquad -D_1(F_{\bar{S}}(s), F_N(n-1|N>0))) & \text{for } s>0, n \ge 1 \\ P(N=0) & \text{for } s=0, N=0. \end{cases}$$

For another approach, Shi et al. (2015) built a dependence model between the frequency and severity. They used an extra indicator variable for occurrence of claim to deal with the zero-inflated part, and built a dependence model between frequency and severity conditional on positive claim. The two approaches described previously were compared; one approach using frequency as a covariate for the severity model, and the other using copulas. They used a zero-truncated negative binomial for positive frequency and the GG model for severity. In Hua (2015), a mixed copula regression based on GGS copula (see Joe (2014) for an explanation of this copula) was applied on a medical expenditure panel survey (MEPS) dataset. In this way, the negative tail dependence between frequency and average severity can be captured.

Brechmann et al. applied the idea of the dependence between frequency and severity to the modeling of losses from operational risks in Brechmann et al. (2014). For each risk class, they considered the dependence between aggregate loss and the presence of loss. Another application of this methodology in operational risk aggregation can be found in Li et al. (2014). Li et al. focused on two dependence models; one for the dependence of frequencies across different business lines, and another for the aggregate losses. They applied the method on Chinese banking data and found significant difference between these two methods.

2.5 Chapter Summary

In this chapter, multivariate insurance claims modeling is reviewed with an emphasis on the usage of copulas. First, marginal models for the frequencies and severities are explained, and then these marginal models are extended into multivariate models using copulas. The chapter is a broad survey of frequency-severity regression models with an emphasis on insurance analytics applications, and the chapter contributes to the literature by recommending specific models for the LGPIF application. The chapter is based on Frees et al. (2016).

Empirical results of the application of the methods in this chapter are shown in Chapter 5 using the LGPIF data. Specifically, Section 5.1 begins with brief summary statistics of the rating variables used in the empirical application, and Section 5.1.2 shows the coefficient estimation results and model fits for the frequency and severity modeling for the building and contents line. Section 5.1.3 shows the dependency modeling results. Section 5.1.4 shows the coefficient estimation results for other lines. Application of the Tweedie model to the LGPIF is illustrated in Section 5.1.5. Out-of-sample validation results are shown in Section 5.1.6, and the usage of the Gini index analysis is shown in Section 5.1.7.

Chapter 3

Deductible Ratemaking

Abstract

Insurance claims have deductibles, which must be considered when pricing for insurance premium. Deductibles may cause censoring and truncation of insurance losses. For these types of claims, the regression approach is often used with deductible amount included as an explanatory variable inside a frequency-severity model. In this way, the resulting regression coefficient can be used to calculate the deductible rates. On the one hand, this approach has the advantage of incorporating the psychological effect of policyholders making deductible choices into the ratemaking. On the other hand, standard actuarial textbooks recommend the maximum likelihood approach for estimating parametric loss models, which can be used for calculating the coverage modification amounts due to the deductibles. In this chapter, a comprehensive overview of deductible ratemaking is provided, and the pros and cons of various approaches under different parametric models are compared. The regression approach proves to have an advantage in predicting aggregate claims. The maximum likelihood approach becomes necessary for calculating theoretically correct relativities for deductible levels beyond those observed, for each policyholder. Models for specific peril types can be combined to improve the ratemaking, and estimation issues for such models under truncation and censoring are discussed.

This chapter is based on Lee, Gee Y., "General Insurance Deductible Ratemaking," *Conditionally Accepted by the North American Actuarial Journal.*

Details of empirical results are shown in Chapter 5, Section 5.2.

3.1 Introduction

A deductible is an important feature of an insurance contract. Deductibles influence the number of times the insured will make a claim and will influence the amount that is reimbursed to the insured in the event of an insured loss. In many cases, deductibles may cause insurance claims to be observed with censoring and truncation. These aspects must be addressed when pricing insurance premiums, and the theoretically correct approach can be discussed from the standpoint of actuarial theory.

To formalize our framework for modeling, let N be the loss frequencies for each policyholder, and random variable Y_j the severities of the losses (for each $j = 1, ..., N$), which are assumed to be independent of N. Suppose a deductible d is applied, so that the risk-sharing function from the perspective of the insurer is defined as

$$g(Y_j; d) = \begin{cases} 0 & Y_j < d \\ Y_j - d & d \leq Y_j < \infty. \end{cases}$$

We can be consistent in notating censored random variables with a subscript g, to remember they have extra zeros below the censoring point. Also, we will denote truncated random variables and their corresponding parameters with a subscript $*$. It is helpful to understand that truncation is basically observing a subset of a full sample, under some truncation mechanism. Hence, the notation $\cdot \mid d \leq Y_j$ means observing a variable conditional on $d \leq Y_j$. Then, the observed, censored and truncated random variable for claim frequencies and severities for each policyholder can be denoted as

$$
\begin{aligned}
N_g(d) &= \sum_{j=1}^{N} \mathrm{I}(d < Y_j) && \text{(number of claims)} \\
Y_{g,j}(d) &= \begin{cases} 0 & Y_j < d \\ Y_j - d & d \le Y_j < \infty \end{cases} && \text{(censored severities)} \\
Y_{*,j}(d) &= Y_j - d \mid d \le Y_j && \text{(truncated severities)} \\
S_g(d) &= \sum_{j=1}^{N} Y_{g,j}(d) && \text{(aggregate claims)}
\end{aligned}
$$

where $\mathrm{I}(\cdot)$ is an indicator function, taking on the value 1 if the input condition is true, and 0 otherwise. Note that $N_g(d)$ is a summation of Bernoulli random variables. These concepts are demonstrated using empirical data in Chapter 5, Section 5.2.1, Table 5.34 and Table 5.33. The concept of censoring and truncation is illustrated graphically in Figure 5.11.

The textbook Klugman et al. (2012) shows in detail how coverage modification affects the claim frequency and severity distributions. The advantage of applying parametric loss models for deductible ratemaking is that accurate, theoretically correct deductible rates can be calculated for insurance losses. When covariates are incorporated into the models, deductibles can be priced in a subject-specific manner, which allows a rating engine to be theoretically correct for all of the policyholders within an insurance company. Empirical work using truncated estimation for insurance claims with data can give practitioners an illustration of the application of loss models for deductible ratemaking.

Although accurate rates can be calculated using such textbook approaches, a practitioner may be interested in the regression approach for deductible ratemaking by treating the deductible level as an explanatory variable in a regression model, as in Frees and Lee (2017). This intuitive solution is to use the coefficient estimates for log deductible to calculate the relativities for various deductible levels. This approach is taken for simplicity of implementation, and practicality in ratemaking applications. The approach becomes particularly useful when a large number of explanatory variables are used for ratemaking. A practitioner may be interested in learning when to apply truncated estimation techniques and when the regression approach suffices. Hence, in this chapter of the dissertation, a detailed analysis of deductible rating approaches is conducted.

For some more motivation for the study, the reader may consider the situation where an analyst would be interested in developing a pricing structure that incorporates deductibles in a disciplined way, and in knowing how to change prices when the deductibles change. Meanwhile, only the reported losses above a certain deductible level may be observed by the analyst. An actuarial analyst may need an assessment of the price of a particular insurance policy or a portfolio of policies under this circumstance. For these considerations, an overview of the available methods for deductible ratemaking, and a comparison of the approaches using empirical applications, would be a meaningful contribution to the literature.

3.2 Literature Review

There is a large literature discussing problems related to deductible ratemaking. Some foundational literature on deductible pricing, exposure rating and coverage modification is summarized in the following subsections. Statistical methods related to censored and truncated estimation have a long history, as does the insurance economics literature, where the deductible choice of policyholders is studied for the assessment of risk preferences. Readers who are interested in the main subject of the paper may skip this section and go directly to Section 3.3.

Deductible Pricing

A standard reference for deductible pricing in actuarial science is in Brown and Lennox (2015), where the *indicated deductible relativity* for a single loss of an insurance policy is given by the relationship

$$\text{Indicated deductible relativity} = \frac{E[Y_g(d)]}{E[Y]} = 1 - \text{LER}(d),$$

where LER is an abbreviation of *loss elimination ratio*. The indicated deductible relativity provides an assessment of how much an insurance loss cost is reduced by a deductible, from a per-loss perspective, while the loss elimination ratio provides an assessment of how much the covered loss is reduced by introducing a deductible d. If the policy has a density function f_Y, then the loss elimination ratio is

$$\text{LER}(d) = \frac{\displaystyle\int_0^d y\ \text{f}_Y(y)dy + d\int_d^\infty \text{f}_Y(y)dy}{\displaystyle\int_0^\infty y\ \text{f}_Y(y)dy}.$$

This principle can be applied to excess-of-loss treaty pricing for per-loss insurance and reinsurance policies, where losses beyond a retention level are covered by a reinsurer. For the frequency-severity framework, it is helpful to use the notations in the following Section 3.3, to define the relativity of an *aggregate loss* as

$$\text{REL}(d_0, d) = \frac{E[S_g(d)]}{E[S_g(d_0)]} = \frac{E[N]\displaystyle\int_d^\infty (1 - \text{F}_Y(y))dy}{E[N]\displaystyle\int_{d_0}^\infty (1 - \text{F}_Y(y)dy} = \frac{\displaystyle\int_d^\infty (1 - \text{F}_Y(y))dy}{\displaystyle\int_{d_0}^\infty (1 - \text{F}_Y(y))dy}, \tag{3.1}$$

where d_0 is a base deductible. In the textbook, Brown and Lennox (2015), the *experience rating* approach, and the *exposure rating* approach for reinsurance pricing are introduced in relation to deductible ratemaking. The former uses a company's historical loss experience, for a best predictor of future experiences. In the latter approach, claim severity distributions are based on industry data. The literature has some work related to excess-of-loss layer rating methodologies.

An article by Bernegger (1997) uses the Maxwell-Boltzmann, Bose-Einstein, Fermi-Dirac distribution to model losses for reinsurance applications. Statistical properties of this distribution is introduced further in Wu and Cai (1999). To summarize their approach, an insurance company may be given an increasing curve in δ, say $H(\delta)$, where $0 < \delta < 1$. In this case, $\delta = d/u$ is a normalized deductible in the $[0, 1]$ interval. This curve is differentiated to obtain an expression for the loss distribution.

Several related studies have been interested in the rating of large insurance losses and the excess-of-loss layer rating. For example, Ludwig (1991) provides an overview of the exposure rating approach. Fasen and Kluppelberg (2014) discusses risk processes for large insurance losses without empirical examples. Several actuarial seminars, such as White and Mrazek (2004) and White (2005), have introduced advanced practical methodologies for exposure rating approaches. A recent article by Chavez-Demoulin et al. (2016) applies extreme value models to operational risk. Some researchers have studied the data misspecification issue under left-truncation, as Gurnecki et al.

(2006). These studies provide good motivation for further studies. This chapter of the dissertation provides an empirical demonstration of how coverage modification effects can be incorporated into deductible ratemaking. The approach in this dissertation chapter is distinct from existing work, in that the interest is more focused on the experience rating approach, using data from the Local Government Property Insurance Fund, introduced further in Chapter 5 and 7.

Coverage Modification

In many cases, the deductible levels correspond to only small values in the lower tail of the claim distribution. Hence, it is often most efficient to use the regression approach with independent explanatory variables, for both the frequencies and the severities of insurance claims. However, for large deductible amounts, there may be motivation to use other approaches.

For a specific class of frequency distributions, called the $(a,\ b,\ 0)$ class distributions, the modification to the frequencies, due to deductibles, has been understood quite well. The $(a,\ b,\ 0)$ class distributions, summarized in Table 3.1, are explained in detail by Klugman et al. (2012). These frequency distributions have the property such that a scale in the parameter θ results in the same scale to the mean of the distribution. If the mean of the distribution with parameter θ is given as $E[Y]$, the mean of the distribution with the scaled parameter θv has mean $E[Y]v$. This property can be easily observed by inspecting the last column of Table 3.1.

Table 3.1: (a, b, 0) Class Distributions

Name	$B(z)$	$B(\theta(z-1))$	Mean
Poisson	e^z	$e^{\theta(z-1)}$	θ
Binomial	$(1+z)^m$	$((1-\theta)+\theta z)^m$	$m\theta$
Geometric	$(1-z)^{-1}$	$(1-\theta(z-1))^{-1}$	θ
Negative Binomial	$(1-z)^{-r}$	$(1-\theta(z-1))^{-r}$	$r\theta$

In Table 3.1, $B(\theta(z-1))$ denotes the probability generating function. To understand the effect of left-truncation of the severities on the frequencies, an effect known as *coverage modifications*, in the language of Klugman et al. (2012), is introduced. Let $v = 1 - \mathrm{F}_Y(d)$ be the probability that a loss results in a claim (a payment). The probability-generating function for the modified claim count random variable can be obtained by modifying the probability-generating function of the underlying (a, b, 0) class distribution $P(z) = B[\theta(z-1)]$ using the probability-generating function $P_I(z) = 1 - v + vz$ of the Bernoulli random variable, which takes on the value 1 when a loss results in a claim:

$$P_g(z) = P(P_I(z)) = P(1 + v(z-1)) = B(\theta(1 + v(z-1) - 1)) = B(\theta v(z-1)) = P(z; \theta v). \quad (3.2)$$

The sum of N Bernoulli random variables has an $(a,\ b,\ 0)$ class primary distribution and a Bernoulli secondary distribution, in which case the probability-generating function for the secondary distribution can be plugged into the probability-generating function of the primary distribution to obtain the resulting compound distribution. This allows expression (3.2) to be so simple and intuitive.

The two most often used frequency distributions for counts in actuarial science are the Poisson distribution and the negative binomial distribution. The Poisson distribution is a popular choice for practical modeling. Let N_g be the observed counts, excluding the unobserved claims due to truncation, for a policyholder. Note that from (3.2), we know the underlying frequencies also follow a Poisson distribution, so that

$$N \sim \text{Poisson}\,(\theta) \iff N_g \sim \text{Poisson}\,(v\theta)\,,$$

where $v = \Pr(Y > d)$ is the amount of coverage modification. In practice, the negative binomial is also often used, in order to accommodate for over-dispersion. In this case, a mean parametrization is used, and the underlying frequencies can be retrieved in a similar way:

$$N \sim \text{NB}\,(r, \theta) \iff N_g \sim \text{NB}\,(r, v\theta)\,,$$

where $v = \Pr(Y > d)$. These properties are valid under the assumption that N and v are independent, meaning the factors determining the severity are independent of the number of claims.

There has been little work on how to use an estimated loss distribution for deductible ratemaking in conjunction with the loss frequencies. When the loss frequencies and severities are empirically analyzed together, the sampling frame also becomes an issue. Cummings (2005) recommends the use of the generalized linear models (GLM) approach to deductible ratemaking. His presentation discusses the limitation of the method of Guiahi (2001), where the relationship between the claim frequencies and the deductibles is not considered. For practical reasons, Cummings (2005) recommends using standard GLM models with deductibles as an independent explanatory variable.

Truncated Estimation

There is a vast literature on censored and truncated data modeling in statistics. However, most of these papers focus on the estimation problem, where the goal is to obtain estimates of parameters for a specified distribution from censored or truncated data. Kaplan and Meier (1958) introduced the product limit estimator for censored and truncated data. There have been a number of follow-up studies, including Woodroofe (1985) and Lai and Ying (1991). Kalbfleisch and Prentice (2002) provides treatment of modeling for censored and truncated data for survival models. Finkelstein and Wolfe (1985) take a semi-parametric approach for interval censored failure time data. There have also been a number of studies focusing on estimation problems for particular distributions: Barr and Sherrill (1999) on the truncated normal, Aban et al. (2006) on the truncated Pareto distribution, and Chapman (1956) on the truncated gamma distribution. Discrete data with zero truncation is discussed in Plackett (1953) and Klugman et al. (2012). Recently, Verbelen and Claeskens (2014) applied multivariate Erlang mixture models to censored and truncated data.

Although the literature is vast, a combined estimation of frequencies and severities under truncation and censoring is rarely found in the statistical estimation literature. The estimation of censored frequencies under fixed coverage modification amounts is treated theoretically in actuarial textbooks, but more empirical treatment seems to be needed.

Insurance Economics

There is another vast but separate literature where the selection effect mentioned in Cummings (2005) is studied in relation to policyholder behavior in the insurance market. The selection effect occurs when specific deductible choices are correlated with the loss profile of a policyholder. The problem of deductible choice is important in insurance economics and risk management, as it is a crucial vehicle for sorting out the adverse-selection and moral-hazard problems in practice. Hence, it is important to study their effects. The pricing of a deductible is an interesting problem, and the precise psychological effect of a deductible choice is a problem under active research.

Hence, the deductible choice problem serves as a framework for understanding economic decisions under uncertainty, whose implications apply more broadly to problems in society. For this reason, traditional economics textbooks cover the deductible choice problem in depth. There is a vast literature in which deductibles are studied in order to understand the risk preference of policyholders and the presence of adverse selection in insurance markets.

In economics and risk management, articles such as Rothschild and Stiglitz (1976) and Halek and Eisenhauer (2001) have been standard references for the need of deductibles for mitigating adverse selection in insurance markets with hidden information. In economics and behavioral economics, deductible choices of policyholders have been used to study the risk preference of decision makers. Treatment of this literature can be found in Mas-Colell et al. (1995), Koszegi and Rabin (2006), and Sydnor (2010). Econometric approaches for measuring the preference of decision makers in a lab setting have been an active topic of research. For example, Holt and Laury (2002) provide standard procedures for measuring risk preferences in a lab environment. Recently, there is interest in extending these studies into real-world problems, through empirical studies such as Cohen and Einav (2007) and Einav et al. (2012). In particular, Sydnor (2010) illustrates how deductible choices are often unexplained by standard economic theory, and more sophisticated models may be necessary.

3.3 Theory of Coverage Modification

This section provides useful theoretical results for deductible ratemaking. Similar results can be found in Klugman et al. (2012), Gray and Pitts (2012), Tse (2009), and Bahnemann (2015); however, here the results are simplified and condensed into more general forms, with an emphasis on small deductible changes. The results apply to any random variable Y, without continuity or the existence of a distribution required. The results for frequencies apply to any count random variable N, not just the $(a, b, 0)$ class distributions.

Let us begin by assuming the censored random variable is observed. From an empirical standpoint, the sample size on which estimation can be performed necessarily gets smaller for the claim severities, as Y is truncated to $Y_*(d)$. See Figure 5.11. The first theorem provides a general expression for the difference in expected aggregate claims, under two different deductibles.

Theorem 3.3.1. *Let N be any count random variable, and Y any random variable, each with finite first moments. If N and Y are independent, then for deductibles $d_1 < d_2$, we have*

$$E[S_g(d_1) - S_g(d_2)] = E[N]E\left[Y_g(d_1) - Y_g(d_2)\right].$$

Proof. We have

$$E[S_g(d)] = E\left[\sum_{i=1}^{N} Y_{g,i}(d)\right] = E\left[E\left[\sum_{i=1}^{N} Y_{g,i}(d)\middle| N\right]\right] = E\left[NE\left[Y_g(d)\right]\right] = E\left[N\right]E\left[Y_g(d)\right],$$

and the theorem follows directly. □

Theorem 3.3.1 provides an expression for the difference between two aggregate claims means, under two different deductibles. When the loss severities have a parametric loss distribution, the following corollaries allow a modeler to calculate the mean average claim and relativities for a given deductible d, relative to a base deductible.

Corollary 3.3.1.1. *If Y has distribution function F_Y, then for $d_2 > d_1$ we have*

$$E[S_g(d_1) - S_g(d_2)] = E[N]\int_{d_1}^{d_2} (1 - F_Y(y))\, dy.$$

Proof. It suffices to provide an expression for $Y_g(d)$. We have

$$\begin{aligned} E[Y_g(d)] &= \int_d^{\infty} (y-d)d\mathrm{F}_Y(y) = (y-d)(1-F_Y(y))\Big|_d^{\infty} + \int_d^{\infty} (1 - \mathrm{F}_Y(y))\, dy \\ &= \int_d^{\infty} (1 - \mathrm{F}_Y(y))\, dy. \end{aligned}$$

In particular, we have

$$E[Y_g(d_1) - Y_g(d_2)] = \int_{d_1}^{d_2} (1 - \mathrm{F}_Y(y))\, dy.$$

Note, $E[Y_g(d)]$ is the partial expectation of Y, given $Y > d$. As a special case, the following provides an expression for the modified aggregate claims under any new per-loss deductible level:

$$E[S_g(d)] = E[S] - E[N] \cdot \int_0^d (1 - \mathrm{F}_Y(y))\, dy.$$

□

Corollary 3.3.1.2. *Using the notation for relativity REL as defined in equation (3.1), we have*

$$E[S_g(d_2)] = E[S_g(d_1)] \times REL(d_1, d_2) = E[S_g(d_1)] \times \frac{\int_{d_2}^{\infty} (1 - F_Y(y))\, dy}{\int_{d_1}^{\infty} (1 - F_Y(y))\, dy},$$

where d_1 may be considered as a base deductible.

This motivates the concept of the *relativity* for aggregate claims, as explained in Section 3.2. Hence, the mean aggregate claim is modified by an amount, depending on the deductible. Given some loss distribution F_Y, the next theorem allows a modeler to recover the underlying loss frequencies, from observed claim frequencies.

Theorem 3.3.2. *Let N be any count random variable, and Y have distribution function F_Y, each with finite first moments. If Y is independent of N, then $N_g(d)$ satisfies*

$$E\left[N_g(d)\right] = E\left[N\right] \cdot (1 - F_Y(d)).$$

Theorem 3.3.2 provides an expression for the mean of the observed, censored frequency distribution, in terms of the underlying loss distribution parameters, under deductible d. When the mean frequency is parametrized using a log-link, for regression purposes, parameters for the underlying loss N can be obtained by a regression, using `offset` $= \ln(1 - \mathrm{F}_Y(d))$. Given the loss distributions, the following formulas provide the theoretical marginal changes in the means, under a small deductible change.

Corollary 3.3.2.1. *If $E[Y_g(d)]$ is differentiable at d, then*

$$\frac{\partial}{\partial d} E[Y_g(d)] = -1 + F_Y(d).$$

If F_Y is differentiable, then

$$\frac{\partial}{\partial d}E[N_g(d)] = -E[N] \cdot f_Y(d).$$

Proof. We have

$$\frac{\partial}{\partial d}E[Y_g(d)] = \lim_{d'\to d} \frac{E\left[Y_g(d') - Y_g(d)\right]}{d' - d} = \lim_{d'\to d} \frac{-1}{d' - d}\int_d^{d'} (1 - \mathrm{F}_Y(y))\, dy = -1 + \mathrm{F}_Y(d).$$

An alternative proof assumes a density for Y. Let f_Y be the density, and we have

$$E[Y_g(d)] = E\left[(Y - d)\cdot \mathrm{I}(Y > d)\right] = \int_d^{\infty} (y - d)\mathrm{f}_Y(y)dy = \int_d^{\infty} y\mathrm{f}_Y(y)dy - d(1 - \mathrm{F}_Y(d)).$$

Differentiation gives

$$\begin{aligned} \frac{\partial}{\partial d}\left[\int_d^{\infty} y\mathrm{f}_Y(y)dy - d\cdot(1 - \mathrm{F}_Y(d))\right] &= -d\cdot \mathrm{f}_Y(d) - (1 - \mathrm{F}_Y(d)) + d\cdot \mathrm{f}_Y(d) \\ &= -1 + \mathrm{F}_Y(d). \end{aligned}$$

For the frequency, we have

$$\begin{aligned} E\left[N_g(d)\right] &= E\left[\sum_{i=1}^{N} \mathrm{I}(y > d)\right] \\ &= E\left[E\left[\sum_{i=1}^{N} \mathrm{I}(y > d)\middle| N\right]\right] \\ &= E\left[N \cdot P(y > d)\right] \\ &= E\left[N\right]\cdot(1 - \mathrm{F}_Y(d)). \end{aligned}$$

If F_Y is differentiable, then the rate of change of the frequencies can be obtained by

$$\frac{\partial E[N_g(d)]}{\partial d} = E[N]\cdot \frac{\partial\left[1 - \mathrm{F}_Y(d)\right]}{\partial d} = -E[N]\cdot \mathrm{f}_Y(d).$$

□

In general, the overall effect of a deductible change on the expected aggregate claim, using any count random variable and any severity distribution, can be obtained by differentiation.

Corollary 3.3.2.2. *Let N be any count random variable, and Y any random variable, each with finite first moments. If N and Y are independent, then $E[S_g(d)]$ satisfies*

$$\frac{\partial}{\partial d}E[S_g(d)] = -E[N]\cdot(1-F_Y(d)).$$

Finally, a formula for the truncated severities is provided.

Theorem 3.3.3 (Truncated Severity Modification)**.** *Let Y be any random variable with density f_Y and distribution F_Y. Then, $Y_*(d)$ satisfies*

$$\frac{\partial}{\partial d}E[Y_*(d)] = \frac{f_Y(d)}{1-F_Y(d)}E[Y_*(d)] - 1.$$

Proof. Using the Libnitz rule, and the notation $\mathrm{TCE}_Y(d) = E[Y|Y>d]$, we have

$$\begin{aligned}
\frac{\partial E[Y_*(d)]}{\partial d} &= \frac{\partial E[Y-d|Y>d]}{\partial d} \\
&= \frac{\partial E[Y|Y>d]}{\partial d} - \frac{\partial}{\partial d}\{d|Y>d\} \\
&= \frac{\partial}{\partial d}\int_d^\infty \frac{y\mathrm{f}_Y(y)}{1-\mathrm{F}_Y(d)}dy - \frac{\partial}{\partial d}\{d|Y>d\} \\
&= \int_d^\infty \frac{\partial}{\partial d}\frac{y\mathrm{f}_Y(y)}{1-\mathrm{F}_Y(d)}dy - \frac{d\mathrm{f}_Y(d)}{1-\mathrm{F}_Y(d)} - \frac{\partial}{\partial d}\{d|Y>d\} \\
&= \frac{\mathrm{f}_Y(d)}{(1-\mathrm{F}_Y(d))^2}\int_d^\infty y\mathrm{f}_Y(y)dy - \frac{d\mathrm{f}_Y(d)}{1-\mathrm{F}_Y(d)} - 1 \\
&= \frac{\mathrm{f}_Y(d)}{1-\mathrm{F}_Y(d)}E[Y|Y>d] - \frac{\mathrm{f}_Y(d)}{1-\mathrm{F}_Y(d)}d - 1 \\
&= \frac{\mathrm{f}_Y(d)}{1-\mathrm{F}_Y(d)}\mathrm{TCE}_Y(d) - \frac{\mathrm{f}_Y(d)}{1-\mathrm{F}_Y(d)}d - 1 \\
&= \frac{\mathrm{f}_Y(d)}{1-\mathrm{F}_Y(d)}(\mathrm{TCE}_Y(d) - d) - \frac{\partial}{\partial d}\{d|Y>d\} \\
&= \frac{\mathrm{f}_Y(d)}{1-\mathrm{F}_Y(d)}E[Y_*] - 1,
\end{aligned}$$

where the last term 1 is obtained by differentiating $\{d|Y>d\}$ with respect to d. Here, TCE is used to denote the tail conditional expectation of a severity distribution. □

3.4 Approaches to Deductible Ratemaking

In this section, we provide an overview of different empirical approaches to deductible ratemaking and how they can be applied in our framework. The two general approaches are the regression approach and the maximum likelihood approach with truncated estimation methods. First, the sampling frame is formalized. For rating purposes, we assume the following variables are observed:

$$\{N_{g,i}(d_i), \boldsymbol{x}_i, d_i\},$$

where $\boldsymbol{x}_i$ is a set of explanatory variables including coverage amounts u_i, which could not be adjusted, while d_i are the deductible choices, which can be adjusted by either the policyholder or the insurance company. The coverage amounts u_i are used as the upper-limit amounts, and these are assumed not adjustable by the policyholder or the insurance company. In other data sets, coinsurance amounts also may be observed. Note that the number of losses N_i are realized prior to the loss amounts. For each loss N_i, the amounts y_{ij} are realized, and

$$y_{*,ij}(d_i) = y_{ij} - d_i | y_{ij} > d_i$$

are observed for each loss $j = 1, \dots N_i$. Hence, the estimation assumes a claims data set and a policyholder data set. If the numbers of observations in these two data sets are considered independent, then standard asymptotic theory could be used for standard error estimates. In many cases, $y_{*,ij}$ may be observed, while y_{ij} is unobserved.

3.4.1 Maximum Likelihood Approach to Deductible Ratemaking

The maximum likelihood approach is a direct application of the theory, outlined in Section 3.3. We provide an overview of how the theory in Section 3.3 and similar results in Gray and Pitts (2012), Tse (2009) and Bahnemann (2015) can be empirically applied to real data. The rating procedure is summarized into four simple steps. The most difficult part is the estimation step, which requires statistical estimation methods for censored and truncated loss distributions. Subsequent steps are simple and straightforward.

Rating Procedure

1. Obtain $\mathrm{F}_Y(y)$ using statistical estimation. This involves censored and truncated estimation methods, (see Section 3.3 for related theorems).

2. Obtain $E[N]$, using Section 3.3 theory. Specifically, use $\ln(1-\mathrm{F}_Y(d))$ as offset and $E[N_g(d)]$ as response in a regression.

3. To calculate the rates for any d_{new}, obtain $E[Y_g(d_{\text{new}})] = \int_0^{d_{\text{new}}} (1-\mathrm{F}_Y(y))\,dy$, using the estimated loss model from Step 1 and numerical integration.

4. Calculate the new $E[S_g(d_{\text{new}})]$ using, $E[S_g(d_{\text{new}})] = E[N]\cdot E[Y_g(d_{\text{new}})]$ (see Corollary 3.3.1.2).

There are two complications to consider. First, for the frequency model parameter estimates, the standard errors become amplified by the modeling error and estimation error from the severity distribution, because of the coverage modification.

Second, if the number of claims is considered random, then the size of the claims data depends on the realization of the claim frequencies and severities for each policyholder, and hence the sampling frame becomes complicated. It is possible to show that for any confidence level, it is possible to find a large enough size for the policyholder sample so that the sample size of the claim data set is ensured to be large enough with any desired confidence. There is a large literature on large-sample theory for the validity of the fixed sample size for the claims data set, given a sequential stopping rule. Anscombe (1952) provides a proof for this result. Siegmund (1985) provides an overview of sequential analysis. In particular, Anscombe (1952) shows that for a sequence of proper random variables taking positive integer values N_r, the sequence of statistics based on N_r observations satisfies convergence and uniform continuity in probability. Assuming the sampling frame described above, application of the rating formulas using the maximum likelihood approach would use the above steps.

3.4.2 Regression Approach to Deductible Ratemaking

The regression approach is to use GLM models with a log deductible covariate. In practice, it is common to assume the after-deductible claims follow a gamma distribution or a Pareto distribution.

Let Y_* be the observed claims. Then common practice is to parameterize the mean of the gamma distribution, using explanatory variables $\boldsymbol{x}_i$ and coefficients $\boldsymbol{\beta}$:

$$E[Y_*(d_i)] = \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right).$$

In this setup, deductibles may be incorporated into the model. If $\ln d_i = \texttt{lnDeduct}_i$ is included as an explanatory variable, then its coefficient, β_d, would satisfy

$$\frac{\partial E[Y_*(d_i)]}{\partial d} = \frac{\partial \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)}{\partial d} = \exp\left(\boldsymbol{x}_i'\boldsymbol{\beta}\right)\beta_d \frac{\partial \ln d_i}{\partial d} = E[Y_*(d_i)]\frac{\beta_d}{d_i}.$$

Hence, for a single policy i, the coefficient β_d can be considered as the deductible *elasticity* of the mean:

$$\beta_d = \frac{\partial E[Y_*(d_i)]/E[Y_*(d_i)]}{\partial d/d_i}.$$

In econometrics, elasticity is a term used to denote the percentage change in a variable, in response to a percentage change in an explanatory variable. However, defining a single quantity β_d for a population of policies can be done in many different ways. When the sample of deductibles is not uniform, or when the deductible choice distribution is correlated with the response variable, the coefficient β_d may reflect this. For this dissertation chapter, assessing how well a calculated β_d summarizes the relativity is best done using graphical approaches explained in Section 3.4.3. Analogously, if used in the regression for $N_g(d_i)$, then the corresponding coefficient γ_d would have the interpretation

$$\gamma_d = \frac{\partial E[N_g(d_i)]/E[N_g(d_i)]}{\partial d/d_i}.$$

In actuarial science, the *pure premium approach* is sometimes used to model the aggregate claims directly, using a compound distribution, such as the *Tweedie* distribution. In this case, a similar approach can be used by including $\ln d_i$ as an explanatory variable in the regression for the aggregate claims. For an overview of the pure premium approach, the reader may refer to Frees (2014) or Shi (2016). In this case, the coefficient ξ_d for $\ln d_i$ would have the interpretation

$$\xi_d = \frac{\partial E[S_g(d_i)]/E[S_g(d_i)]}{\partial d/d_i}.$$

One may compare the preceding derivations with using $\ln u_i = \texttt{LnCoverage}_i$ as an explanatory variable, in order to incorporate the coverage amounts into the regression model, as larger coverage amounts should typically result in higher claims. To ensure the interpretability of the coefficients for log coverage amounts, it is often recommended to use an alternative, *exposure (offset)* approach. According to Frees et al. (2016), in actuarial science *exposures (offsets)* are used to calibrate the size of potential outcome variables. In this case, the mean can be assumed to vary proportionally with an amount E. In this case, the coefficient for $\ln E$ is restricted to be 1 and included in the model as an *offset*. With this convention, we have

$$\mu = E \cdot \exp(\mathbf{x}'\boldsymbol{\beta}) = \exp\left(\mathbf{x}'\boldsymbol{\beta} + \ln E\right).$$

One can consider using $E = u - d$ as the offset amount for the regression, when the deductible amounts are to be incorporated as well. This way, when a policyholder selects a larger coverage amount (upper limit of policy), higher insurance premiums would be charged. Similarly, higher-deductible choices would naturally lead to a discount of the premium. The problem with applying this approach to deductibles is that the precise effects of deductible changes are not considered this way. In most cases, for a unit of deductible change, the scale change of the loss may be different from a unit. Note that in reality, the effect of a deductible change may vary, depending on the loss distributions of the policyholders.

3.4.3 Relativity Calculation

Further insight can be obtained by comparing the relativities for each approach. Recall the expression for REL in equation (3.1). Given a base deductible d_0, for the regression approach, we have

$$\mathrm{REL}_{\mathrm{REG}}(d_0, d) = \frac{\exp\left(\boldsymbol{x}'\boldsymbol{\beta} + \beta_d \ln d\right) \cdot \exp\left(\boldsymbol{x}'\boldsymbol{\gamma} + \gamma_d \ln d\right)}{\exp\left(\boldsymbol{x}'\boldsymbol{\beta} + \beta_d \ln d_0\right) \cdot \exp\left(\boldsymbol{x}'\boldsymbol{\gamma} + \gamma_d \ln d\right)} = \left(\frac{d}{d_0}\right)^{\beta_d + \gamma_d}, \tag{3.3}$$

where we assume $\beta_d + \gamma_d < 0$. When the *pure premium approach* is used, the relativity given base deductible d_0 would be calculated in a similar way:

$$\mathrm{REL}_{\mathrm{REG}}(d_0, d) = \frac{E[S^c(d)]}{E[S^c(d_0)]} = \frac{\exp\left(\boldsymbol{x}'\boldsymbol{\xi} + \xi_d \ln d\right)}{\exp\left(\boldsymbol{x}'\boldsymbol{\xi} + \xi_d \ln d_0\right)} = \left(\frac{d}{d_0}\right)^{\xi_d}, \tag{3.4}$$

where we assume $\xi_d < 0$. For a general link function $\eta : (0, \infty) \to (-\infty, \infty)$, the relativity curve becomes

$$\mathrm{REL}_{\mathrm{REG}}(d_0, d) = \frac{\eta^{-1}\left(\boldsymbol{x}'\boldsymbol{\xi} + \xi_d \cdot \eta(d)\right)}{\eta^{-1}\left(\boldsymbol{x}'\boldsymbol{\xi} + \xi_d \cdot \eta(d_0)\right)},$$

where ξ_d is the coefficient for the covariate $\eta(d_i)$ in the regression. Different link functions may result in different shapes of relativity curves, yet in this paper, let us focus on analyzing the log link. The true relativity is

$$\mathrm{REL}_{\mathrm{CM}}(d_0, d) = \frac{\displaystyle\int_d^u (1 - \mathrm{F}_Y(y))dy}{\displaystyle\int_{d_0}^u (1 - \mathrm{F}_Y(y))dy}, \tag{3.5}$$

where $d > d_0$. For a Pareto model, if $\alpha > 1$, we have

$$\mathrm{REL}_{\mathrm{CM}}(d_0, d) = \frac{\displaystyle\int_d^u \left(\frac{\lambda}{\lambda + y}\right)^{\alpha} dy}{\displaystyle\int_{d_0}^u \left(\frac{\lambda}{\lambda + y}\right)^{\alpha} dy} = \frac{(\lambda + d)^{-\alpha+1} - (\lambda + u)^{-\alpha+1}}{(\lambda + d_0)^{-\alpha+1} - (\lambda + u)^{-\alpha+1}}, \tag{3.6}$$

and taking the upper bound u to infinity gives

$$\mathrm{REL}_{\mathrm{CM}}(d_0, d) \to \left(\frac{\lambda + d}{\lambda + d_0}\right)^{-\alpha+1} \quad \text{as } u \to \infty. \tag{3.7}$$

The reader may compare (3.4) and (3.7), and note the similarity, as both are decreasing functions in d, with the true relativity curve depending on the shape parameter of the distribution. In particular, observe that in equation (3.7), as $\lambda \to 0$ and $u \to \infty$, with $\alpha > 1$, the relativity curve for the Pareto model becomes identical to that of the regression approach, with $\xi_d = -\alpha+1$ in equation (3.4). To compare the performance of the regression approach, one may plot the relativities for deductibles in the range $[d_0, u]$ and compare the fit. Note that in parametric models with covariates, the true relativity curve becomes subject specific through the distribution parameters, whereas the regression approach would allow for subject-specific variation through interaction terms with the deductible covariate. The performance of the regression approach depends on how well the relativity curve approximates the true relativity of a given policyholder.

3.5 Applications

In this section, three applications are provided. Section 3.5.1 performs a simulation study, in order to illustrate the nature of the relativity curve. From the simulation, we learn that the regression approach is an approximation to the true relativity curve using one parameter. Section 3.5.2 calculates the relativities using real data from the LGPIF. The LGPIF data is described in more detail in Chapter 5, and 7. Section 3.5.3 shows that the regression approach performs well in terms of aggregate claims prediction.

3.5.1 Simulation

To assess the performance of the regression approach, claims were generated synthetically, using parameters similar to the Local Government Property Insurance Fund (LGPIF) building and contents claims (explained in more detail in Chapter 5), and the lightning peril type is used for demonstration. The coefficient estimate results for the severity models are shown in Chapter 5, Table 5.35. Claims were synthetically generated using parameters similar to those found from estimation. This way, the deductible level can be adjusted to observe the potential effect on the relativity. Hence, $B = 10,000$ policies were generated using a Pareto distribution, with $E[Y] = 11,087$ and $\alpha = 2.553$. The claim frequency mean $E[N] = 1$ was used. Deductibles were synthetically generated by the following procedure:

- Generate d_i from a multinomial distribution over {500, 1000, 2500, 5000, 10000, 25000}, each with probabilities (0.461, 0.215, 0.118, 0.095, 0.011, 0.100), for $i = 1, \ldots, B$ policyholders.

These numbers were used so that the deductible distribution resembles the LGPIF data. Then the regression approach is used to estimate the elasticities. The results are $\beta_d + \gamma_d = -0.217$ for the deductible elasticity using the Poisson and gamma family. The relativities are then calculated for

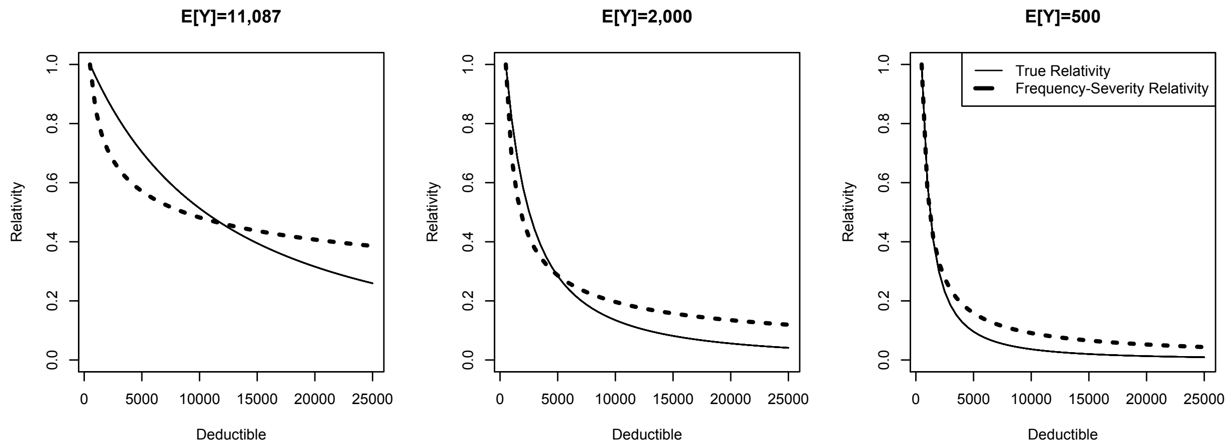


Figure 3.1: Relativities From Regression Using $\ln(d_i)$, for $E[Y] = 11087$, $E[Y] = 2000$, and $E[Y] = 500$

deductible levels ranging between the base 500 and 25,000 using the regression coefficients, and shown in Figure 3.1 along with the true relativity curve for comparison.

From Figure 3.1, observe that the regression approach using $\ln(d_i)$ approximates the true relativity curve in the best possible way, with deviations due to the nature of the log link. In the first panel, notice that for small relativity values, there is a small discrepancy between the dotted and solid lines. If the curve is dilated to the right, eventually the regression approach results in larger and larger deviations from the true relativity, as the curve deviates more from the solid line. For example, if a deductible of 1,000,000 is selected for the reinsurance retention level, the error in the relativity would be substantial when the regression approach is used. The regression approach using $\ln d$ as the explanatory variable results in a curve that is steeper than the true relativity curve for small deductibles, and the curve flattens out eventually due to the nature of the link function. In subsequent panels, the scale parameter λ is increased, showing how the regression approximation becomes closer to the true relativity curve when λ approaches zero. In general, we find that the regression approach would be suitable for moderate-size problems, where the scale parameter λ is moderate in size.

In Figure 3.2, instead of $\ln(d_i)$, we attempt to use $\ln(d_i + \lambda)$ for the explanatory variable. This way, the regression approach relativity curve becomes identical to the true relativity curve, and in all three panels, the curve fits almost exactly. This suggests several valuable insights. First, we learn that the regression is an approximation to the relativity curve using one parameter only,

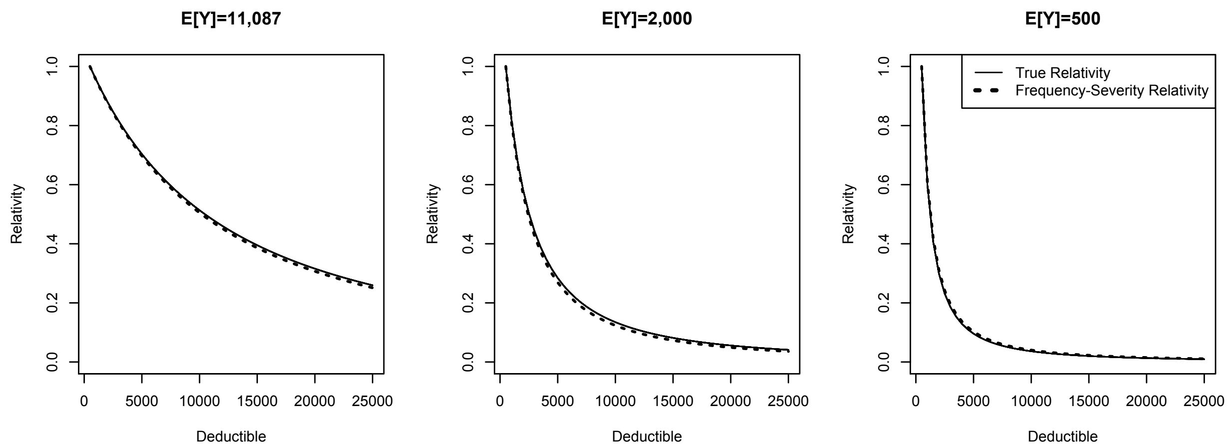


Figure 3.2: Relativities From Regression Using $\ln(d_i + \lambda)$, for $E[Y] = 11087$, $E[Y] = 2000$, and $E[Y] = 500$

or in other words, without a scale parameter. This scale parameter is not known in advance without performing maximum likelihood. Also, we learn that when the problem is of moderate size, regression may perform well.

3.5.2 Relativities

Next, we compare the different approaches using real data. Figure 3.3 shows a histogram of the lightning losses and the fitted distributions using truncated estimation for the exponential, gamma and Pareto models. From the figure, observe that the exponential and gamma model fits look about the same, whereas the Pareto model fit is slightly better. Details of the coefficient estimation results are shown in Chapter 5, Section 5.2, Tables 5.35, 5.36.

The regression approach is implemented using a standard `glm` software package with Poisson and gamma families. Log deductibles are included as covariates in each regression model, by peril type. To compare the single relativity obtained from the regression approach, we calculate single relativities for the gamma and Pareto maximum likelihood approaches by

$$\mathrm{REL}_m(d_0, d) = \frac{E\left[\sum_i S_{g,i,m}(d)\right]}{E\left[\sum_i S_{g,i,m}(d_0)\right]}$$

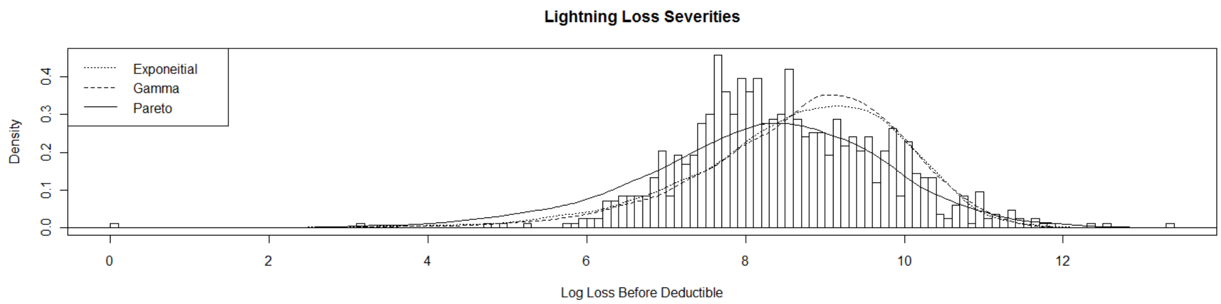


Figure 3.3: Density of Lightning Losses

where $S_{g,i,m}(d)$ indicates the aggregate claims for peril type m, for policyholder i, under deductible d. This single quantity allows for a comparison of the relativity obtained from the maximum likelihood approach and the regression approach.

The regression approach coefficient estimates for the $\ln(d)$ explanatory variable are shown in Table 3.2. These quantities are used to calculate the relativities for the regression approach, by peril type. Relativities for the maximum likelihood approach are calculated using the estimated models in Chapter 5, Section 5.2, Tables 5.35, 5.36.

Table 3.2: Coefficients from Regression Approach

	Poisson (γ_d)	Gamma (β_d)	Sum ($\gamma_d + \beta_d$)	Adjustment	$1-\alpha$**
Fire	-0.407	0.228	-0.179		-0.012
Vandalism	-1.335	0.981	-0.354		-0.357
Lightning	-0.822	0.489	-0.332		-0.874
Wind	-0.567	0.259	-0.308		-0.242
Hail*	-0.202	0.594	0.391	-0.202	-0.543
Vehicle	-1.125	0.429	-0.697		-2.924
Water (Non-weather)*	-0.579	0.714	0.135	-0.579	-0.127
Water (Weather)*	-0.375	0.949	0.574	-0.375	-0.000
Misc.*	-0.734	0.716	-0.019	-0.734	-0.063

*Adjustment made for parameter interpretability.
**Shape parameter of Pareto model is shown for reference.

In Figure 3.4, the peril types are categorized into nine categories, as explained in more detail in Chapter 5, Section 5.2, Table 5.37. In the figure, relativities for the regression approach and those obtained from maximum likelihood are overlayed, allowing for comparisons. The numeric values of the relativities are shown in Table 3.3.

In each panel of Table 3.3, the relativities are shown in the column for each deductible level. The reader may compare the first panel, showing the regression approach, with the second and third panels, showing the gamma model and Pareto model, using truncated estimation and the rating formulas in Section 3.3. The leftmost column is the relativity for the 1,000 deductible level. The relativities for the next deductible level, 2,500, are lower as expected, and so on. The lowest relativity indicates the ratio of the aggregate claims under 50,000 to the aggregate claims under the base deductible, 500. Hence, the single quantities in the tables allow for a comparison of the relativity levels for the single elasticity obtained from the regression approach and the maximum likelihood approach.

The reader may observe that regression provides lower relativities in general. In particular, the relativities in the regression approach are somewhat more uniform over different deductible levels, except for very small deductible levels, when compared with the second panel, which uses the same distributional assumption with a maximum likelihood approach. This is due to the nature of the log link. As we will see in subsequent sections, the performance of the aggregate claim prediction is unaffected by this phenomenon. The Pareto distribution has a heavier tail and in general results in higher relativities. The calculations are shown with holdout sample claims for year 2011, while the models were fit using data for years 2006–2010.

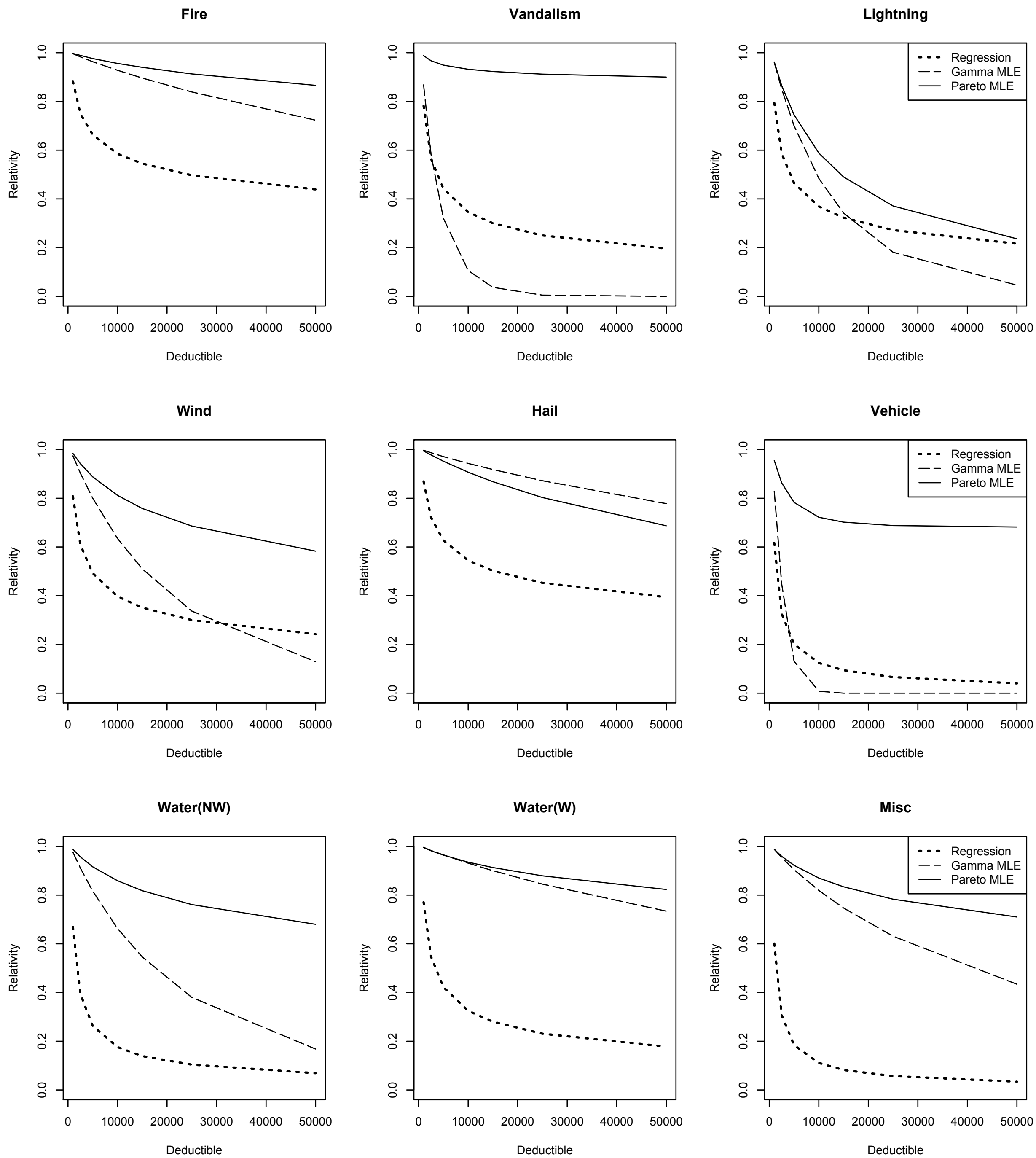


Figure 3.4: Plot of Relativities for Regression Approach and Selected MLE Approaches

We have been able to make a few observations while implementing each approach. First, when the maximum likelihood approach is used with the gamma distribution assumption, some of the policyholders resulted in relativity values of zero. This shows that although the maximum

Table 3.3: Comparison of Relativities for Regression Approach and Selected MLE Approaches

Regression Approach

Deductible	1,000	2,500	5,000	10,000	15,000	25,000	50,000
Fire	0.883	0.750	0.663	0.585	0.545	0.497	0.439
Vandalism	0.782	0.565	0.442	0.346	0.300	0.250	0.196
Lightning	0.794	0.586	0.465	0.369	0.323	0.272	0.216
Wind	0.808	0.609	0.492	0.397	0.351	0.300	0.242
Hail*	0.869	0.722	0.627	0.545	0.502	0.453	0.394
Vehicle	0.617	0.326	0.201	0.124	0.094	0.066	0.040
Water (Non-weather)*	0.669	0.394	0.263	0.176	0.139	0.104	0.069
Water (Weather)*	0.771	0.547	0.422	0.325	0.280	0.231	0.178
Misc.*	0.601	0.307	0.184	0.111	0.082	0.057	0.034

*lnDeduct has been included only in the frequency regression, for interpretability.

Poisson-gamma MLE

Deductible	1,000	2,500	5,000	10,000	15,000	25,000	50,000
Fire	0.996	0.983	0.963	0.928	0.896	0.839	0.723
Vandalism	0.868	0.583	0.320	0.106	0.037	0.005	0.000
Lightning	0.960	0.851	0.700	0.484	0.342	0.181	0.046
Wind	0.974	0.903	0.800	0.635	0.510	0.337	0.129
Hail	0.997	0.987	0.971	0.943	0.918	0.872	0.778
Vehicle	0.829	0.448	0.131	0.008	0.000	0.000	0.000
Water (Non-weather)	0.976	0.911	0.816	0.663	0.546	0.380	0.168
Water (Weather)	0.996	0.983	0.965	0.931	0.900	0.845	0.734
Misc.	0.988	0.954	0.904	0.819	0.747	0.631	0.434

Poisson-Pareto MLE

Deductible	1,000	2,500	5,000	10,000	15,000	25,000	50,000
Fire	0.997	0.989	0.976	0.956	0.940	0.913	0.866
Vandalism	0.988	0.967	0.949	0.932	0.923	0.912	0.900
Lightning	0.962	0.865	0.745	0.588	0.490	0.371	0.236
Wind	0.984	0.942	0.888	0.812	0.758	0.686	0.583
Hail	0.994	0.978	0.952	0.907	0.868	0.803	0.687
Vehicle	0.955	0.863	0.783	0.722	0.702	0.688	0.682
Water (Non-weather)	0.988	0.957	0.916	0.859	0.818	0.761	0.680
Water (Weather)	0.995	0.982	0.964	0.935	0.913	0.879	0.823
Misc.	0.988	0.959	0.922	0.870	0.834	0.783	0.710

likelihood approach has the flexibility of providing policyholder-specific relativities by varying the parametrization, sophisticated distributional assumptions, such as a long-tail, Pareto distribution, would be needed for the relativities to be interpretable. In contrast, the regression approach provides a single relativity value, which allows for easier interpretation. The regression approach, although not perfect, seems to provide reasonable single-value relativities for an analyst to use.

In particular, observe that the maximum likelihood experienced difficulty in assessing the vandalism relativities, because claims in this category are influenced substantially by the deductible, as claim sizes are small. Regression seems to provide more reasonable relativities for this category. The reader may observe this in Table 3.3.

Second, the severity model for the regression approach sometimes provided coefficients that could not be interpreted. This situation is illustrated in Table 3.2. Specifically, hail, water and miscellaneous peril types resulted in β_d values too high compared to γ_d. In these cases, the regression model must be fixed, with β_d omitted from the severity model. In Section 3.5.3, a similar situation was observed for the aggregate claims model. The reader may see the coefficient for `lnDeductBC` in Chapter 5, Section 5.2, Table 5.39.

3.5.3 Aggregate Claims Prediction

How well does each method, including the regression approach, perform in aggregate claims prediction? We compare the regression approach with the maximum likelihood rating approach for total aggregate claims prediction for an entire line. From this study, we demonstrate that, although the regression approach provides smaller-than-reasonable relativities for some policies, it performs quite well in terms of total aggregate claims prediction. We expect the regression approach to provide a practical solution for applications with a large number of explanatory variables, where the aggregate claims prediction is of primary interest. For a demonstration, we compare the results of three different approaches:

(A) Regression approach, using Poisson-gamma, with the `lnDeduct` covariate

(B) Maximum likelihood approach, using Poisson-gamma truncated estimation

(C) Maximum likelihood approach, using Poisson-*GB2* truncated estimation

We did not implement a regression approach for the Poisson-*GB2* model, because comparing the three cases listed would suffice in demonstrating the most commonly encountered assumptions in practice. Advantages and disadvantages of regression and maximum likelihood can be illustrated by comparing these three common modeling assumptions. The estimation results are shown in Chapter 5, where the tables for model A, B and C are each shown in Tables 5.39, 5.41, and 5.42. Using these coefficient estimates, the total aggregate claims are predicted for different deductible levels. For the regression approach, the log deductible amount is multiplied by the coefficient estimate −0.737 in model A. The coefficient estimates from truncated estimation are also used to calculate the aggregate claims, and the result is compared with predictions from the regression approach.

Table 3.4: Out-of-Sample (2011) Performance of Each Approach

	Aggregate Claims	Pearson Correlation with Claims	Spearman Correlation with Claims
(A) Poisson-gamma regression	16,170,966	0.2231	0.3922
(B) Poisson-gamma MLE	11,464,929	0.3358	0.3847
(C) Poisson-*GB2* MLE	20,976,735	0.4157	0.4025
Claims	19,036,189	1.0000	1.0000

According to Table 3.4, the regression approach was effective in identifying the ranking of the policyholders (as could be seen from the Spearman correlation), as well as predicting the amount of aggregate claims, compared with the maximum likelihood approach using the same distribution assumption. The aggregate claim amount for the regression approach, 16,170,966, is in fact closer to the empirical claims than 11,464,929. This shows, for many practical rating problems with small data sets, that the regression approach may provide a good assessment for the actual aggregate claims. That is perhaps because the coefficients β_d and γ_d incorporate the deductible selection effect into the rating.

Theoretically, it is possible to incorporate the deductible selection effect into the Poisson-*GB2* model. For example, one may consider fitting a separate severity model for various deductible levels. A similar approach could be used for the mixture approach as well. If a selection effect exists, then the severity model coefficient estimates would provide different parameters for each deductible level. Another potential way to incorporate the selection effect is to use dependence modeling, considering

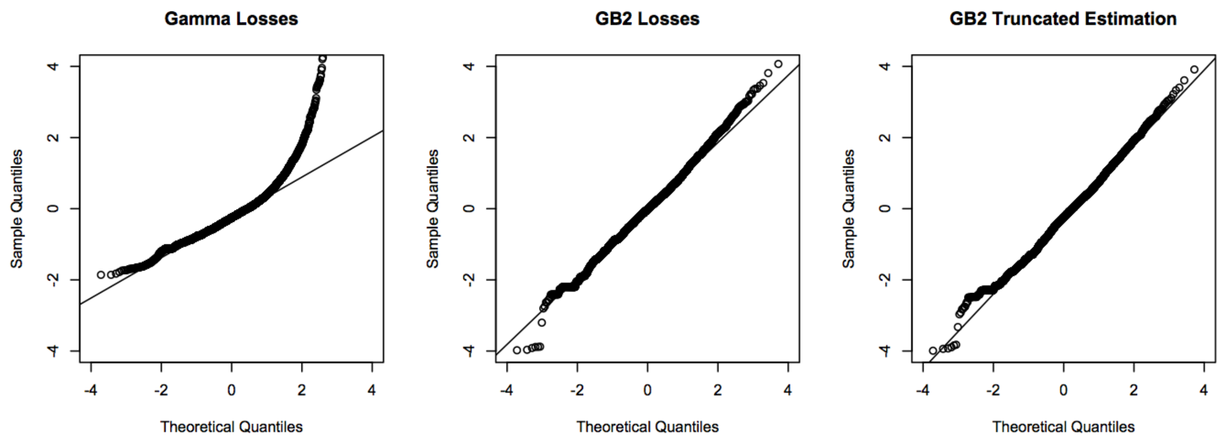


Figure 3.5: Comparison of Severity Distributions. Panel 1: gamma losses fit; Panel 2: *GB2* losses fit; Panel 3: *GB2* truncated estimation (Kolmogorov-Smirnov test statistics (p-values) 0.199 (0.000), 0.025 (0.003), 0.103 (0.000))

the dependence between deductible levels and the loss severity. This approach is left as future work.

According to Table 3.4, long-tail loss models, such as the *GB2* model with truncated statistical estimation, and coverage modification theory can improve the deductible rating, as the third row of the table shows. Figure 3.5 provides some insight into why the prediction improves. The left panel shows the Q-Q plot when the *GB2* model is fit to the underlying loss distribution, assuming it is observed. The LGPIF data records both the underlying loss and the deductible amounts, which allows this figure to be shown as a comparison. The first panel shows that the fit of the gamma distribution suffers for the lower and upper tails of the claims distribution. The middle panel shows that the *GB2* fit is better, and the third panel shows that truncated estimation has recovered the underlying distribution quite well. In general, a better fit of the underlying loss model would result in a better assessment of the coverage modification of the deductibles, for either small values or large values of deductibles.

During our analysis, we performed comparisons of claim scores obtained from different models, for various hypothetical deductible levels. For example, the aggregate claims can be predicted for increasing deductible levels, and applied to all policyholders throughout the property fund. The predicted aggregate claims can then be compared with the hypothetical empirical observed claims, which can be obtained by applying the hypothetical deductible level to the underlying losses. In general, our analyses have shown that, for large deductibles, the *GB2*-01NB model performs best.

Defining how large a deductible level is required, for the maximum likelihood approach to become necessary, would be an interesting research question for future studies. The regression approach may be a good method for predicting aggregate claims, for moderate-size data and small deductibles. We find that when accurate deductible relativities are of interest for large losses, more elaborate methods, such as the maximum likelihood approach, are needed with truncated estimation. In particular, if subject-specific deductible relativities are needed or when an excess of loss layer is to be priced, then the maximum likelihood approach would be necessary. In contrast, when a single relativity value is desirable, the regression approach may turn out to be useful.

3.5.4 Comparison of Frequency Models

The severity distribution influences the truncation of the underlying frequency distribution through the coverage modification amount $v = 1 - F_Y(d)$. This section compares different frequency model assumptions under such deductible influence. We compare the Poisson model and 01-inflated Poisson model by fitting the two distributions to the underlying loss frequencies and then attempt to estimate the same parameters using the censored frequency observations. Details of the estimation issue for 01-inflated count models under deductible influence are covered in Chapter 5, Section 5.2.3. For an introduction to 01-inflated count distributions, the reader may refer to Chapter 2, Section 2.2.2.

Table 3.5 shows that the 01-inflated Poisson model fit is better than that of the Poisson model, when the underlying losses are observed. This situation is shown in the first panel, where the 01-Poisson with underlying frequencies in general fit the empirical counts better, in each count category. For example, the number of zero observations is predicted to be 768, which is close to the empirical 766.

Table 3.5: Comparison of Predicted Counts Using Validation Sample

Count	(1) Poisson Underlying	(2) 01-Poisson Underlying	(3) Poisson Censored Estimation	(4) 01-Poisson Censored Estimation	Empirical Counts (2011)
0	703	768	614	603	766
1	194	195	231	236	187
2	84	54	108	109	57
3	42	29	56	57	28
4	23	16	31	32	17
5	14	10	18	19	14
6	8	6	11	12	4
7	6	4	7	7	8
8	4	3	5	5	2
9	3	2	3	4	2
10	2	2	2	2	2
11	2	1	2	2	1
12	1	1	1	1	0
13	1	1	1	1	1
14	1	0	1	1	0
15	1	0	1	1	1
16	1	0	0	0	1
17	0	0	0	0	0
18	0	0	0	0	0
19	0	0	0	0	0

Note: Each column attempts to predict the underlying loss frequencies. Loss count categories 0–19 are shown for illustration. Notice that (3) and (4) overpredict the 0-losses and 1-loss categories and underpredict the 2-losses category. This is as expected, and the reader may understand the reason from the fit of the severity distribution below the deductible. Compare the Q-Q plots in Figure 3.5.

The situation is different when a deductible causes censoring. In terms of the predicted counts, the 01-Poisson does not perform better than the Poisson. This is because prediction of the observations below the deductible has become difficult, as the censoring causes data below the deductible to be unobserved. This motivates using a basic model for the counts when censoring is in place. Hence, the Poisson model is used in Section 3.5.3, which also allows using the $\ln(1 - v_i)$ offset technique for estimation. The reader may refer to Chapter 5, Section 5.2.3 for details of the estimation procedure when a 01-inflated Poisson model assumption is used, and the coefficient estimate results under this assumption.

3.6 Summary

Loss models are built on positive responses. For this reason, log links or other such link functions $\eta : (0, \infty) \to (-\infty, \infty)$ are required to implement a regression approach. As a result, when the deductible amount is included as a covariate in a regression, the resulting relativities become an approximation of the theoretically true relativities. Ratemaking using the regression approach has practical advantages, because it can be used with the standard generalized linear models framework. As we have seen, the regression approach to the ratemaking of deductibles proves to be valuable when the aggregate claims prediction is of most interest.

This dissertation chapter provides an overview of the rating of deductibles for per-loss insurance deductibles. The chapter demonstrates how textbook methods for coverage modification compares with the regression approach. Theoretical results have been generalized to loss variables without continuous distributions. The chapter also provides a comparison of the regression approach to deductible ratemaking, with the maximum likelihood approach.

To summarize the work in this chapter, a comprehensive overview of deductible ratemaking has been performed. For cases where small deductibles are applied and the aggregate claim amounts are the primary interest, the deductible amounts may be used as covariates in the scale parametrization of a regression model. The reader may compare equations (3.3) and (3.7) to understand how the regression approach provides a reasonable approximation of the true relativity curve, and the meaning of the shape parameter of a distribution in relation to the curvature of the relativity curve. This approach is not suitable when deductibles are large or when the precise relativities are of interest for large losses, as shown in Figure 3.1.

When large deductibles are to be priced, the maximum likelihood approach is recommended. For example, excess of loss layers may be priced better by fitting an underlying loss model. Yet, as Table 3.4 shows, when different deductible-rating approaches were used to predict the aggregate claims, the regression approach performed reasonably well, outperforming the maximum likelihood approach using an identical distributional assumption, in terms of predicted aggregate claim amounts. The explanation is that regression utilized the deductible selection effect within the data. Specifically, the deductible elasticity of the claim frequency may likely have included effects due to deductible selection.

Future Work

For future studies, it may be interesting to apply more elaborated estimation procedures for the case when the observed data contain only the average severities. Further studies on estimation issues for compound distributions also may be of interest. The effect of an aggregate deductible may also be a natural extension, where an additional layer of deductible is applied to the losses. In addition, the claims may be classified using more elaborated techniques, to allow for efficient peril-specific loss models for deductible ratemaking, using regression and maximum likelihood. The influence of classification on deductible rates may also be potentially interesting to study further. With big-data becoming an important issue in analytics, consideration of truncation and censoring issues in classification methods may be a very important extension to consider.

Deductible Selection Effects

In my opinion, the most interesting extension would be the consideration of deductible selection effects. For example, separate models may be imposed for each deductible choice for different sub-populations. The difference between two sub-populations can sometimes be captured by a single parameter, but sometimes not. In the latter case, different parameters, and hence different models can be used for the two sub-populations.

For example, consider a high risk-type group, and a low risk-type group. The risk-type may be indicated by the deductible choice of the policyholders. The fitted loss distribution for the two groups may be identical, or it may be different. The difference may be indicated by the difference in the model parameter estimated from the two populations. If the estimated parameters are different and significant, then it may provide evidence of deductible selection effects from the population.

Chapter 4

Insurance Portfolio Optimization

Abstract

In this chapter of the dissertation, the insurance risk retention problem is explored. Given an underlying claims distribution and a premium constraint, what is the optimal amount of risk to retain, or equivalently which level of risk retention parameters should be chosen by an insurance entity? The risk retention parameter may be deductible (d), coinsurance (c), or upper limit (u). Is it possible to find the optimal parameter, and hence the optimal retention? The problem is studied using numerical optimization techniques, utilizing the estimated claims distributions using loss modeling methods. In our study, the minimum amount of premium collected is used as a constraint to the optimization, and the deductible, upper limit, or coinsurance is optimized for each policyholder. The optimization is performed under consideration of the portfolio of all other policies carried by an insurance company. The chapter discovers that the RM^2 measure is equivalent to the loading factor of a reinsurance premium, and also illustrates that the risk retention problem is not convex, but quasi-convex. This means unique solutions can be found within intervals of the parameter space.

Empirical results using the Local Government Property Insurance Fund (LGPIF) are shown in Chapter 5, Section 5.3 in the form of a case study.

4.1 Introduction

Determining the optimal form of an insurance contract can mutually benefit the insurance company and the policyholder. What does it mean to optimize an insurance portfolio? This question relates to the classic mean-variance trade-off of the Markowitz model of portfolios in finance.

To briefly summarize this chapter of the dissertation, a constrained optimization problem is formulated, similar to the Markowitz framework, so that the VaR minimizing portfolio is found under a certain premium constraint (Note: VaR is the amount of assets needed to cover possible losses). The parameters over which the optimization happens are the deductible, upper limit, and coinsurance. Because these risk retention parameters form a non-convex optimization problem, we resort to numerical techniques to find optima for empirical studies.

Section 4.2 reviews related literatures, and introduces the reader to the concept of a loading factor. Basically speaking, the tradeoff between having reinsurance versus taking one's own risk, happens because of what is known as the loading factor in the optimal reinsurance literature. If there were no loading factor, then an insurer would have incentive to reinsure all of the risk. Under a certain loading factor, now the insurer has a decision to make: How much should be retained, and how much should be reinsured? The decision is influenced by the loading factor. This framework is summarized in more detail in Section 4.3.

In subsequent sections, specific sub-cases of the full framework are explored one at a time. Section 4.4 and 4.5 explores the single risk case, with and without the concept of a portfolio of risks insured. Basically, consider a single policy, and all the rest of the policies, and consider changing the parameter for the single policy. What are the optimal parameters, or in other words, what is the optimal amount of insurance/reinsurance to purchase, under knowledge of the distribution of the losses carried by an insurer. Section 4.6 extends the study to aggregate losses, and demonstrates the quasi-convexity of the problem. An important class of quasi-convex functions is monotonic functions, and the chapter demonstrates that our portfolio optimization problem deals with monotonic functions, and hence has unique solutions under premium constraints. Section 4.7 explores a specific case of the aggregate loss optimization problem, where blocks of policies are optimized. A block of policies is basically a collection of policies that follow the same parameter, such as a collection of school entities, etc. Subsequent chapters provide further insights. For example, Section

4.8.1 explores the case, where two layers of parameters are optimized.

Insurance and reinsurance takes various forms, such as quota-share, stop-loss, or change-loss, to name a few. In many cases, the goal is to determine the price for an insurance scheme, under some type of budget constraint, couched in terms of preferences of the risk taker. Under restrictive assumptions, both the utility theoretic and the risk measure approaches reduce to thinking about a premium as an expected value and measuring the uncertainty through the variance. We begin by reviewing relevant literature, and formalizing the concept of a loading factor $1 + \lambda^*$, where the superscript $*$ is used to recognize the fact that the loading factor is actually identical to what is known as the $RM^2(\theta^*)$ risk measure, introduced in Frees (2016), evaluated at the optimal parameter θ^*. This basically means that the loading factor could be numerically computed, if desired.

4.2 Literature Review

Risk Retention Literature

For an example of a related actuarial textook, Gray and Pitts (2012) compares the risk-sharing arrangements involving deductibles, coinsurance, and upper limits. Traditional approaches in the insurance economics and risk management literature have sought to find an optimal risk retention function. Specifically, using conditions on an insured's preferences as quantified by a utility function or a risk measure, substantial work has gone into determining the optimal form of a contract and, in some cases, the optimal level of contract parameters such as deductible, upper limit, and coinsurance. Also discussed in the traditional insurance economics literature is the issue of function optimality in the presence of "background risks," cf. Schlesinger (2013). For example, we could represent the insurer's total risk as the sum of the insurer's share of the loss, and all other risks. The insurance economics literature focuses on the robustness of the optimality results in the presence of background risks, typically making as few assumptions as possible about the background risk distribution.

Insurance Economics Literature

There is a long history of determining the risk retention parameters in the insurance economics literature. Using this approach, an optimal choice of risk retention parameters can be determined through expected utility. A classic problem in insurance economics is to find the value of c to maximize expected utility, $Eu[w_0 - (1-c)Y - cE[Y](1+\lambda^*)]$, where $u(\cdot)$ denotes an insured's utility function and w_0 represents initial wealth. Under mild conditions, the result of Mossin (1968) indicates that full coverage corresponding to $c = 1$ is optimal when insurance is fair, $\lambda^* = 0$. When a loading is in place, $\lambda^* > 0$, then a partial insurance coverage corresponding to a value of $c < 1$ is optimal, cf., Schlesinger (2013). Schlesinger also notes that this insurance result has a portfolio interpretation. One can define $A = w_0 - E[Y](1+\lambda^*)$ to be a non-risky asset and the weighted average to be a combination of a non-risky and risky asset y. In this sense, the insurance choice problem is equivalent to the portfolio problem in financial economics. As another example, the optimality of deductible policies ($d > 0$) was first established by Arrow (1974). See Gollier (2013), for a description of this and related results.

As with utility theoretic approaches, there is a long although less well-known, literature on the choice of optimal risk retention parameters beginning with the work of Borch (1960). Related to this, Assa (2015) provides a recent overview of this literature. As described in Assa (2015), in insurance one worries about optimal decisions from both the insured and the insurer's viewpoint. In the context of this literature, one may consider the $cE[Y](1+\lambda^*)$ term as a reinsurance premium for a quota share of the risk cY. Thus, $(1+\lambda^*)$ becomes a loading factor for the reinsurance premium. Note, this formulation disregards the credit risk of a reinsurance contract.

Optimal Reinsurance Literature

Notations: The following are common notations in the literature.

$$\text{Insured risk} : I_f = Y - g(Y)$$

$$\text{Quota-share reinsurance} : g(Y) = c \cdot Y$$

$$\text{Stop-loss reinsurance} : g(Y) = \max\{0, Y - d\}$$

$$\text{Change-loss reinsurance} : g(Y) = c \cdot \max\{0, Y - d\}$$

Bowers et al. (1997), Daykin et al. (1994), and Kaas et al. (2001) emphasizes that for a fixed reinsurance premium, the stop-loss contract is the optimal solution among a wide array of reinsurance schemes, in the sense that it gives the smallest variance of the insurer's retained risk. Recently, Cai and Tan (2007) provided a framework for finding the optimal stop-loss insurance contract under VaR and CTE risk measures. In this framework, an analytic form of the optimal contract could be determined, which depends on the safety loading of the reinsurance premium. In their model, the safety loading and assumed loss models are critical factors for determining the optimal retention levels. They consider the optimal reinsurance $P(g(Y)) = (1+\lambda^*)E(g(Y))$, with λ^* as the relative security loading. For example, for the VaR, when it exists, the optimal retention level would be $d^* = VaR_{\lambda'}(Y)$, where $\lambda' = \lambda^*/(1+\lambda^*)$. The authors also illustrate that the results for the individual risk models may be extended to dependent risks, and show the use of multivariate phase type distributions, multivariate Pareto distribution, and multivariate Bernoulli distribution, for the assessment of the effect of dependence on optimal retention levels.

Another related work, Cai et al. (2008) establishes that depending on the risk measure level of confidence and the safety loading for the reinsurance premium, the optimal reinsurance can be in the forms of stop-loss, quota-share, or change-loss. The authors find the optimal reinsurance contract over a class of convex functions, so that the optima is $g^*(X) = (\min(Y, VaR_\alpha(Y)) - VaR_{\lambda'}(Y))_+$, where $\lambda' = \lambda^*/(1+\lambda^*)$. This is a form of stop-loss insurance.

Follow-up studies have furthered this approach using optimization criteria involving both the insurer and the reinsurer. See Cai et al. (2015), Cai and Wei (2012a), Cai and Wei (2012b), Embrechts et al. (2017), Cai et al. (2017). When there are multiple insurance players, Asimit et al. (2013) considers the optimal risk transfer under quantile-based risk measure criteria. The optimality problem of the risk transfer contract between two insurance companies with a one-period setting appears in the literature. Specifically, let ϑ be the risk measure, such as VaR or CTE. Let Y be the loss, $g(Y)$ is the amount retained by the insurer and $Y - g(Y)$ be the amount retained by the policyholder, or the cedent. The function g is assumed to be smooth, specifically, Lipschitz such that $|g(y_1) - g(y_2)| \leq |y_1 - y_2|$ for all y_1, y_2. Then, for a premium functional P, the decision becomes $\min\{\vartheta[Y] - \vartheta[Y - g(Y)] + P[g(Y)]\}$, where $P[g(Y)] = E[g(Y)](1+\lambda^*)$, with λ^* as the relative security loading.

Sung et al. (2011) studies the optimal insurance policy offered by an insurer under the cumulative prospect theory framework of Kahneman and Tversky, with convex probability distortions. They show that under a fixed premium rate, the optimal insurance policy is either an insurance layer or a stop-loss insurance. This work is part of a vast literature, which attempts to better explain the observed insurance buying behavior using non-expected utility models, including subjectively weighted utility. In their work, the proportional premium principle is used, so that $P[g(Y)] = E[g(Y)](1 + \lambda^*)$, with λ^* as the relative security loading. More studies on probability distortions and risk preferences can be found in Mas-Colell et al. (1995), Koszegi and Rabin (2006), Sydnor (2010).

Quantile Sensitivity Literature

Recently, Frees (2016) used results in the management science literature based on quantile sensitivities. A quantile sensitivity measures the change in a risk measure, such as a value at risk, based on changes in an input, such as a deductible or other retention parameter. By assessing the (local) direction and size of a risk measure change per unit change in a retention parameter, valuable advice on which parameters to change can be formed. Sharp, actionable results were provided in Frees (2016) in part due to the knowledge of the background risk distribution but also because the portfolio results were only local, not global, as in much of the literature. Hong (2009) proposes a kernel estimator for estimating quantile sensitivities, which is consistent and asymptotically normal. The authors provide a numerical example, where the method is applied to a portfolio consisting of three assets. The authors also apply the method to a production-inventory example, where the quantile sensitivity of the total cost over n periods is computed, as well as a queueing example, where the quantile sensitivity of the service time is calculated numerically.

In this chapter of my dissertation, I will demonstrate how the risk measures literature ties together with the framework of Markowitz, and how to rethink the security loading factor $(1 + \lambda^*)$ found consistently throughout the literature. I will extend the work of Frees (2016) using aggregate loss models, and demonstrate applications of the risk retention framework.

4.3 Framework

Using general notations, let the total losses for an insurer's portfolio be

$$S = \sum_{\forall i} S_i$$

where S_i is the aggregate loss for policy i. Let the aggregate loss for policy i be

$$S_i = \sum_{k=1}^{p} \sum_{j=1}^{N_{ik}} Y_{ijk},$$

where Y_{ijk} is the jth claim for the ith policyholder in coverage group (line) k. For an important special case of transformations of insured losses, consider the risk-sharing function, where the wedge symbol is defined as $Y \wedge u = \min(Y, u)$.

$$g(Y; d, c, u) = c\,(Y \wedge u - Y \wedge d) = \begin{cases} 0 & Y < d \\ c(Y - d) & d \leq Y < u \\ c(u - d) & Y \geq u \end{cases} . \tag{4.1}$$

Here, the risk retention parameter d is the deductible, c is the coinsurance percentage, and u is the upper limit of coverage. Let the retained claim be $S_{g,i}$. Then we have:

Per-loss level risk retention function: $g_1(Y_{ijk}) = g(Y_{ijk}; c_{i1}, d_{i1}, u_{i1})$

Per-line level risk retention function: $g_2(S_{g_1,ik}) = g\left(\sum_{j=1}^{N_{ik}} g_1(Y_{ijk}); c_{i2}, d_{i2}, u_{i2}\right)$

Per-policy level risk retention function: $g_3(S_{g_1,g_2,i}) = g\left(\sum_{k=1}^{p} g_2(S_{g_1,ik}); c_{i3}, d_{i3}, u_{i3}\right)$

Retained claim: $S_{g,i} = g_3(S_{g_1,g_2,i})$

where Y_{ijk} is the loss for the jth loss of the ith policyholder in coverage group k.

The above framework uses the notation

$$S_{g_1,ik} = \sum_{j=1}^{N_{ik}} g_1(Y_{ijk})$$

which is the aggregate claim amount for policy i in the k coverage group, and

$$S_{g_1,g_2,i} = \sum_{k=1}^{p} g_2(S_{g_1,ik})$$

is the aggregate claims for policy i, which is obtained by summing all the losses in coverage groups $k = 1, ...p$. Also, N_{ik} is the number of losses for the ith risk, which could be a random variable. Define:

$$R_{g,i} = S_i - S_{g,i} \tag{4.2}$$

to denote the **reinsured loss**. In this case, $\sum_k \sum_j g(Y_{ijk})$ denotes the **insured loss**, the amount that the insurer will pay for the ith contract. Let $\boldsymbol{\theta}$ be the collection of risk retention parameters, including deductible, coinsurance, and upper limit. In this dissertation, the primary interest is in

$$S_g = g\left(\sum_{j=1}^{N_i} g(Y_{ij}, \theta_{i1}), \theta_{i2}\right) + S_{(i)}$$

where $S_{(i)}$ represents other risks in an insurer's portfolio (and so the insurer has knowledge of the distribution). The focus may be on a specific coverage group, and hence the subscript k is omitted. Think of the quantile $\xi_{g,i}$ as the amount of assets that the insurer needs to retain for the insured loss at a given parameter $\boldsymbol{\theta}$. Then

$$q_g = \inf\{y : F_g(y) \geq \alpha\} \tag{4.3}$$

where F_g is the distribution of the insured loss S_g. Suppose the insurer wishes to minimize q_g subject to a restriction on the premiums. We can write this as a constrained optimization problem. Specifically, over different choices of $\boldsymbol{\theta}$, we seek to

$$\begin{aligned} &\text{minimize} && q_g \\ &\text{subject to} && E[S_g] \geq P_{min} \end{aligned} \tag{4.4}$$

for the minimal premium P_{min}. From this perspective, the insurance portfolio problem is comparable to the classical portfolio optimization problem introduced by Markowitz in a finance context.

For us, it will be important to check that both the objective function q_g and the constraint function $E[S_g]$ are quasiconvex in $\boldsymbol{\theta}$ in order to ensure a global optima. A function is called *quasiconvex* if its domain and all its sublevel sets are convex. For a function in $\mathbb{R}$, quasiconvexity requires that each sublevel set be an interval. If the function is strictly quasiconvex, then any local optimal point will be a global optimal point (See Section 4.8.5). Further, as is common with portfolio problems, there is a frontier of optimal points that one can visualize using a plot of $E[S_g]$ versus q_g. The corresponding Lagrangian is

$$L = q_g - \{E[S_g] - P_{min}\}(1 + \lambda^*)$$

so that the Lagrange multiplier is

$$(1 + \lambda^*) = \frac{\partial_\theta \ q_g}{\partial_\theta E[S_g]} = RM^2(\theta^*). \tag{4.5}$$

This is the **risk measure relative margin** introduced in (Frees, 2016). Our discovery is that the first order condition from the optimal reinsurance literature can be reinterpreted in terms of the RM^2. Specifically, in the optimal reinsurance literature, the first order condition

$$\frac{\partial q_{S_g}}{\partial \theta} + \frac{\partial E[R_{g,i}]}{\partial \theta}(1 + \lambda^*) = 0$$

is used, assuming a fixed loading factor $(1+\lambda^*)$, which we can now understand as the RM^2 measure at the optimal parameter θ^*. For multiple dimension, we have $\boldsymbol{\theta}^*$ a vector. Taking the gradient of the Lagrangian gives

$$\nabla_{\boldsymbol{\theta}} q_g - \nabla_{\boldsymbol{\theta}} E[S_g](1 + \lambda^*) = 0$$
$$E[S_g] \geq P_{min}$$

Now that we have formalized the framework in terms of the RM^2 measure, our next step is to demonstrate how the the optima for specific examples can be determined using constrained optimization. Focusing on a single coverage group, we will be interested in exploring the following specific cases:

Main Text

- Single loss risk retention $S_g = S_{g,i} = g(Y_i)$ (Section 4.4).
- Single loss with omit-i portfolio, $S_g = S_{g,i} + S_{(i)} = g(Y_i) + S_{(i)}$ (Section 4.5).
- Aggregate losses, $S_g = S_{g,i} + S_{(i)} = g(Y_{i1}) + ... + g(Y_{iN_i}) + S_{(i)}$ (Section 4.6).
- Blocks of policies, $S_g = S_{g,i} + S_{(i)} = g(Y_{i1}) + ... + g(Y_{ip}) + S_{(i)}$ (Section 4.7).

Supplementary Notes

- Two layers of parameters, $S_g = S_{g,i} + S_{(i)} = g_2(g_1(Y_{i1}) + ... + g_1(Y_{iN})) + S_{(i)}$ (Section 4.8.1).

4.4 Single Loss Risk Retention

For a fixed $0 < \alpha < 1$, and a specific policy (so that the subscript i is omitted), let $q = F^{-1}(\alpha)$ be the corresponding quantile. Let $b = c(u - d)$, which represents the maximum payout. The distribution function of $g(Y)$, shown in (4.1), is

$$F_{g(Y;\theta)}(z) = \begin{cases} F(d) & z = 0 \\ F\left(\frac{z}{c} + d\right) & z < b \\ 1 & z \geq b \end{cases}.$$

There is additional discreteness due to the parameters d and u (or b). To compute the mean insured loss, use integration by parts to get (see Chapter 3, Corollary 3.3.1.1)

$$E[g(Y;\theta)] = \quad c\int_d^u (1 - F(y))dy$$

This result assumes that d is finite. The case of $d = -\infty$ can be handled in a straightforward fashion. From Frees (2016), if $F(d) < \alpha < F(b-)$, then one finds q_g has the solution z of the equation $\alpha = F\left(\frac{z}{c} + d\right)$. Straight-forward algebra shows this to be $q_g = c\,(q - d)$, where q is the quantile of the underlying variable Y. If $\alpha \geq F(b-)$, then $q_g = b = c\,(u - d)$. Summarizing, we have

$$q_g = \begin{cases} 0 & \alpha < F(d) \\ c(q-d) & F(d) \leq \alpha < F(b-) \\ c(u-d) & \alpha \geq F(b-) \end{cases}. \tag{4.6}$$

Optimization of Upper Limit

In this section, we demonstrate how the risk retention parameter u can be optimized for a simple example. Assume that the distribution of Y follows a Pareto distribution with distribution function

$$F_Y(y) = 1 - \left(\frac{\gamma}{x + \gamma}\right)^{\eta}$$

so that the mean is

$$\mathrm{E}[Y] = \frac{\gamma}{\eta - 1}.$$

Taking a single Building and Contents loss for the City of Green Bay, from the LGPIF data, assume that $\gamma = 26425.53$, $\eta = 1.846395$, so that the mean is 31221.28 (see Chapter 5, Section 5.1, Table 5.11 for parameter estimation details). If $\alpha = 0.95$, then the quantile is 107436.1 with a zero deductible. For this distribution, the quantile is

$$q = \gamma\left((1-\alpha)^{-1/\eta} - 1\right) = 107436.1$$

and the expected value is

$$\mathrm{E}[g(Y;d,c,u)] = \frac{c\gamma}{\eta - 1}\left[\left(\frac{\gamma}{d+\gamma}\right)^{\eta-1} - \left(\frac{\gamma}{u+\gamma}\right)^{\eta-1}\right] = 31221.28.$$

The expected value calculation is based on limited expected value calculations documented in, for example, (Klugman et al., 2012). Focusing on the upper limit, so that $c = 1$ and $d = 0$,

$$q_g = \begin{cases} q & \alpha < F(u-) \\ u & \alpha \geq F(u-) \end{cases}$$

which is linear in the upper limit parameter. We also have

$$\mathrm{E}[g(Y;u)] = \frac{\gamma}{\eta - 1}\left[1 - \left(\frac{\gamma}{u+\gamma}\right)^{\eta-1}\right]$$

Figure 4.1 is a plot of the quantile and the premium, both as a function of the upper limit u. The quantile function is linear in u and hence quasi-convex. The premium function is concave in u, as well as quasi-convex.

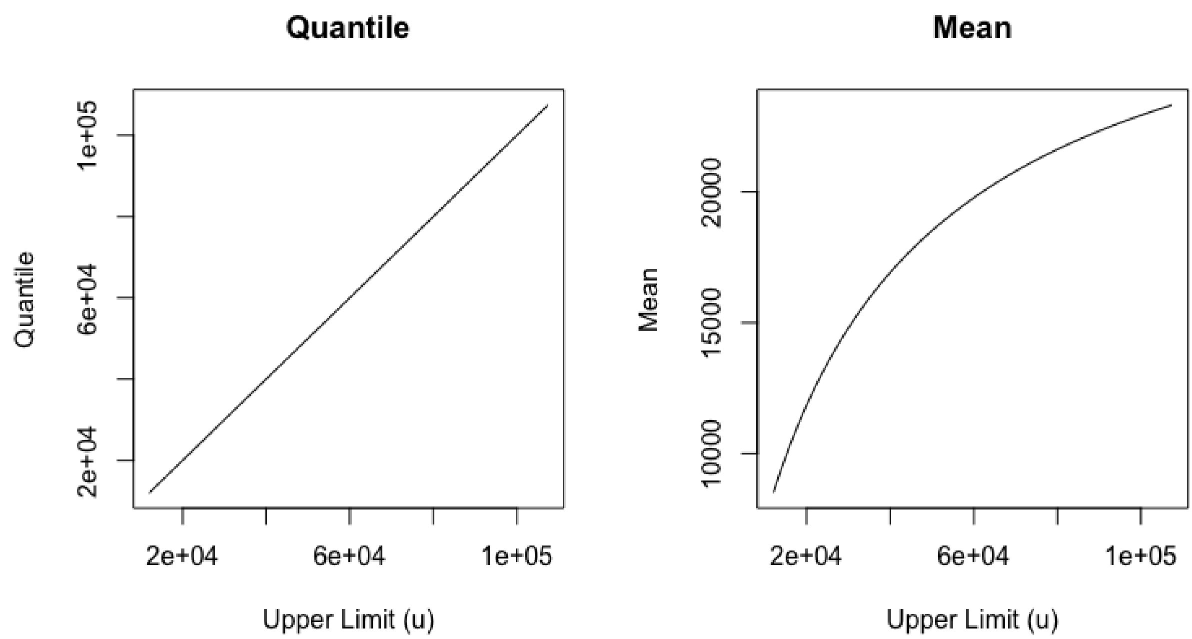


Figure 4.1: Convexity of Single Policy Quantile Function

An insurer would like to make the quantile (left-hand panel) as small as possible by lowering the upper limit. The smallest value occurs when the constraint is binding. Thus, the required premium is $E[Y] = P_{min}$, so that the optimal upper limit is

$$u_{min} = \gamma \left[\left(1 - \frac{(\eta - 1)P_{min}}{\gamma} \right)^{-1/(\eta-1)} - 1 \right].$$

Figure 4.2 is a graph of the optimal upper limit and quantile as a function of the required premium. As the insurer wants more premium, the upper limit increases and the necessary assets (the quantile) increases.

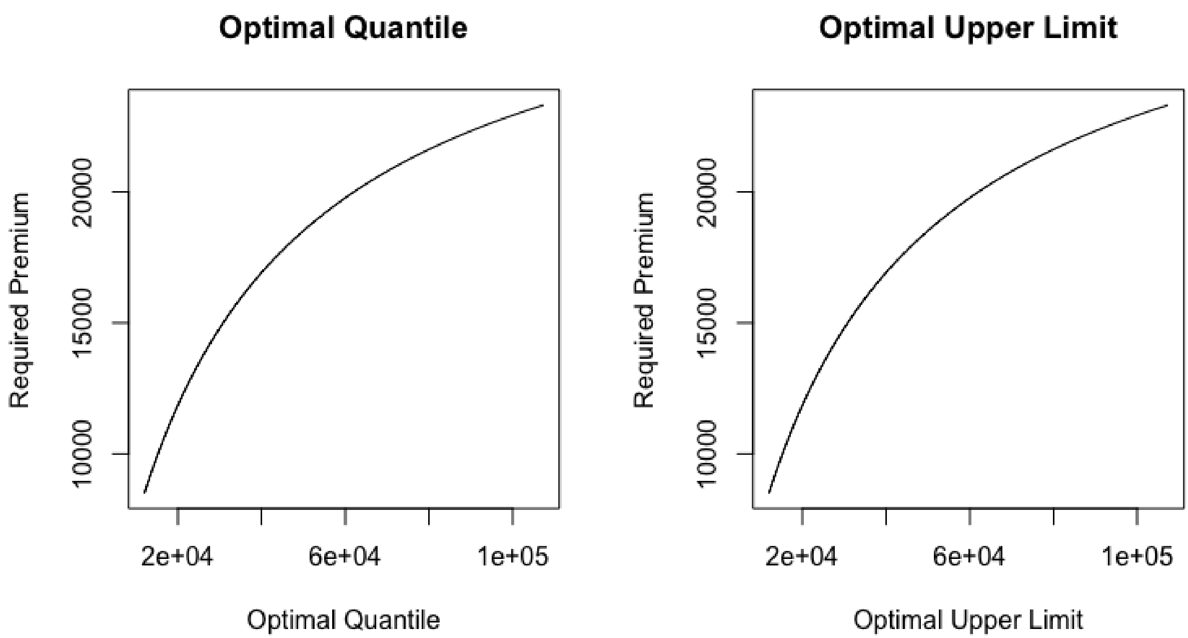


Figure 4.2: Optimal Upper Limit and Quantile

Figure 4.2 shows how to calculate explicit results for a single policy retention with upper limit only, for a Pareto distributed risk.

Numerical Optimization of Upper Limit

Now, we want to use constrained optimization software to do this same calculation with the idea of extending it to more general situations. The simulation solves the contstrained optimization problem for a Pareto distributed risk. The model is exactly same as in Figure 4.1, so not reproduced. When generating Figure 4.1, we obtained closed form expressions using the simplicity of the Pareto distribution. This is not typically possible in an aggregate loss problem. The current exercise can be readily adapted to solve problems with general distributions. The code optimizes $q_g(u)$, which is a function of u, using the R function `auglag`. In order to ensure a convex domain, the following restriction is given.

$$P(u) \geq P_{min}$$

where we call P_{min} the required premium. The optimal upper limit u^*, and the optimal quantile $q_g(u^*)$ is found by optimizing the analytical function for the quantile reconciled with Figure 4.1.

Now, instead of using the analytical quantile and limited expected value functions, we simulate random variables and optimize the simulated approximations of these functions. We first do so in the case of the single risk that follows a Pareto distribution so we can check our procedures. $B = 10,000$ replicates of the Pareto distributed loss is simulated with scale γ and shape η. Then the upper limit is applied, and the 95% quantile of these censored losses is called q_g. The quantile is defined for each upper limit level u, and hence a function $q_g(u)$ can be defined. This function is optimized using the R `auglag` routine, under the constraint

$$P_{sim}(u) > P_{min}$$

where this time $P_{sim}(u)$ is defined as the mean of the 10,000 generated censored losses. The resulting optimal deductible u^* and $q(u^*)$ reconciled with Figure 4.2.

Table 4.1: Upper Limits and Corresponding RM^2

u^*	$E[S_g(u^*)]$	$q_g(u^*)$	$RM^2(u^*)$
12,039	8,498.719	12,039	2.093
39,295	16,781.803	39,295	5.524
66,552	20,456.149	66,552	10.402
93,808	22,561.210	93,808	16.648

This provides motivation, so as to extend our study into more complicated cases, where the objective function is the quantile of an entire portfolio of losses and the premium constraint is obtained from a managerial decision regarding the minimum premium required from an entire portfolio. In the following chapters, we explore specific cases using numerical optimization of objective functions, which involve simulation of losses.

An interesting exercise is to set the loading factor to $(1 + \lambda^*) = RM^2(u^*)$ and optimize the following problem to verify that the same value as u^* is obtained:

$$\min \left\{ q_{S_g} + E[R_g](1 + \lambda^*) \right\}$$
$$\text{where } S_g = \min(Y, u) \text{ and } R_g = (Y - u)_+$$

The optima are shown in Table 4.2. In the simulation, $B = 1,000,000$ replicates of the Pareto distribution has been used to obtain the mean and quantile.

Table 4.2: Loading Factors and Corresponding Optima

$(1+\lambda^*)$	u^*
2.093	13,007
5.524	40,335
10.402	67,583
16.648	95,064

Table 4.2 verifies that in one dimension, the RM^2 risk measure in Frees (2016) is equivalent to the loading factor found in the optimal reinsurance literature.

4.5 Single Policy with Omit-i Portfolio

4.5.1 Non-convexity

In general, the quantile function is not convex. To help understand this problem, we look to the differentiability of the quantile to see if we get conditions for convexity for a single policy problem within a portfolio. To begin, assume that θ is univariate and drop the α subscript to write $q(\theta)$ for the quantile $q_\alpha(\theta)$. Starting with $F(q(\theta);\theta) = \alpha$, differentiate to get

$$\partial_\theta \ F(y;\theta)|_{y=q(\theta)} + \partial_y \ F(y;\theta)|_{y=q(\theta)} \ \partial_\theta q(\theta) = 0.$$

Let's use the shorthand notation

$$F_\theta + F_y \ q_\theta = 0.$$

Thus,

$$q_\theta = -\frac{F_\theta}{F_y},$$

which is known as a "quantile sensitivity." For second derivatives, we have

$$F_{\theta\theta} + F_{y\theta} \ q_\theta + F_y \ q_{\theta\theta} = 0,$$

which can be solved for $q_{\theta\theta}$:

$$q_{\theta\theta} = -\frac{F_{\theta\theta}}{F_y} - \frac{F_{y\theta}}{F_y} q_\theta$$

To understand the expression for $q_{\theta\theta}$ more, we are interested in simulating an insurer's portfolio, including a specific policy i. Thus, let Y_1 simulate the ith policy, and let Y_2 simulate the omit-i portfolio, so that:

$$S_g = g(Y_1) + Y_2$$

To simulate an insurer's portfolio using parametric models, consider a random variable Y_1 and another random variable Y_2 representing the sum of all other losses within a portfolio. We assume

$$Y_1 \sim \text{Gamma}(\alpha_1, \beta)$$
$$Y_2 \sim \text{Gamma}(\alpha_2, \beta)$$

where α_1, α_2 are shape parameters, and β is a scale parameter, which is assumed to be identical for the two random variables for simplicity. Then, recalling that the sum of gamma distributed random variables is also a gamma distribution, we have

$$Y_1 + Y_2 \sim \text{Gamma}(\alpha_1 + \alpha_2, \beta)$$

For the loss random variable Y_1, consider applying an upper limit u. Then, for a given a, we have two cases for the distribution of the censored loss, depending on the value of a:

$$Pr(\min(Y_1, u) \leq a) = \begin{cases} F_{Y_1}(a) & u \geq a \\ 1 & u < a \end{cases}$$

With $\theta = u$, we have:

$$S_g = \min(Y_1, u) + Y_2 \tag{4.7}$$

Thus,

$$
\begin{aligned}
F_{S_g}(y) &= Pr(\min(Y_1, u) + Y_2 \le y) \\
&= \int_0^y Pr(\min(Y_1, u) \le y - z) dF_{Y_2}(z) \\
&= \int_{y-u}^y F_{Y_1}(y-z)\ dF_{Y_2}(z) + \int_0^{y-u} 1\ dF_{Y_2}(z) \\
&= \int_{y-u}^y F_{Y_1}(y-z)\ dF_{Y_2}(z) + F_{Y_2}(y-u) - F_{Y_2}(0) \qquad (4.8)
\end{aligned}
$$

In order to obtain the terms F_{uu}, F_{yu}, and F_y, differentiate (4.8).

$$
\begin{aligned}
F_u &= \int_{y-u}^y f_1(y-z)dF_2(z) - F_1(u)f_2(y-u) - f_2(y-u) \\
F_{uu} &= -f_1(u)f_2(y-u) - f_1(u)f_2(y-u) + F_1(u)f_2'(y-u) + f_2'(y-u) \\
&= f_2'(y-u)\,[1 + F_1(u)] - 2f_1(u)f_2(y-u) \\
F_y &= \int_{y-u}^y f_1(y-z)dF_2(z) + F_1(0)f_2(y) - F_1(u)f_2(y-u) + f_2(y-u) \\
&= F_u + 2f_2(y-u) + F_1(0)f_2(y) \\
F_{yu} &= F_{uu} - 2f_2'(y-u) \\
F_{yy} &= F_{uy} + 2f_2'(y-u) + F_1(0)f_2'(y)
\end{aligned}
$$

Simplifying, we have

$$
q_{uu} = -\frac{F_{uu}}{F_u + 2f_2(y-u) + F_1(0)f_2(y)} - \frac{F_{uu} - 2f_2'(y-u)}{F_u + 2f_2(y-u) + F_1(0)f_2(y)} q_\theta
$$

The quantity q_{uu} gives little intuition, and in general the quantile function cannot be shown to be convex.

4.5.2 Quasi-convexity

In this section, we demonstrate using a specific example, that when a single policy risk is convoluted with a portfolio risk, the resulting quantile of the aggregate loss is still quasi-convex in the optimization parameter. Under the framework of equation (4.7), the following simplification is valid:

$$\begin{aligned}
Pr(\min(Y_1,u)+Y_2 \le y) &= \int_0^y Pr(\min(Y_1,u) \le y-z|Y_2=z) f_{Y_2}(z)dz) \\
&= \int_0^y I(u > y-z) F_{Y_1}(y-z) f_{Y_2}(z)dz + \int_0^y I(u \le y-z) f_{Y_2}(z)dz \\
&= \int_0^y I(z > y-u) F_{Y_1}(y-z) f_{Y_2}(z)dz + \int_0^y I(z \le y-u) f_{Y_2}(z)dz \\
&= \int_{y-u}^y F_{Y_1}(y-z) f_{Y_2}(z)dz + F_{Y_2}(y-u)
\end{aligned}$$

Using this relation, the quantile function can be plotted analytically with respect to different u values. The equation

$$F_{Y_2}(y-u) + \int_{y-u}^y F_{Y_1}(y-z) f_{Y_2}(z)dz = 0.95$$

is solved for y numerically, in order to calculate the quantile corresponding to each u. Let us assume

$$\alpha_1 = 1, \alpha_2 = 3, \beta = 1.$$

The resulting quantiles are plotted with respect to the parameter u. Alternatively, the 95th percentile corresponding to the parameter u may be found from simulation. Specifically, the parameter u can be applied to $B = 1{,}000{,}000$ randomly generated gamma random variables Y_1, and then added to Y_2, so that

$$S_g = \min(Y_1, u) + Y_2$$

The 95th value of the simulated S_g values is obtained. This quantity can be plotted for each u. The result is shown in the following figure. In the left panel, the analytically obtained quantiles are plotted with respect to u, while on the right panel the simulated quantiles are plotted.

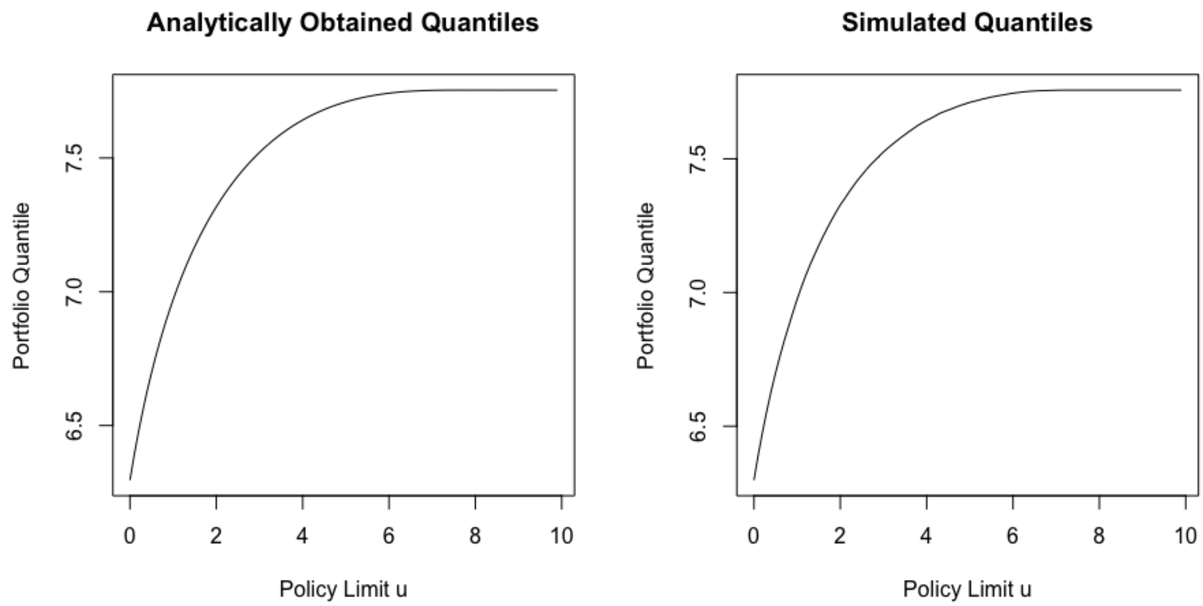


Figure 4.3: Comparison of Analytical and Simulated Quantile Function

The two plots are identical. Figure 4.3 provides several insights. First, it provides motivation to calculate and optimize quantile based risk measures using Monte Carlo simulations. Furthermore, it shows that for this particular example, the portfolio quantile function is quasi-convex in the upper limit parameter u. We will be able to establish this property for general aggregate loss risk retention situations. Specifically, we will be able to establish

$$\frac{\partial q}{\partial u} > 0$$

for the upper limit parameter u, and

$$\frac{\partial q}{\partial d} < 0$$

for the deductible parameter d. Hence, for single parameter optimization of risk retention parameters, such as the deductible or the upper limit, unique global optima could be found under premium constraints. Section 4.5.3 provides an example to demonstrate that in general the quantile function is not convex.

4.5.3 Non-convexity Example

To demonstrate the non-convexity of the quantile function, we plot the quantile function with respect to the parameter u for a specific example. As before, let the losses for policy i be Y_1, and introduce an omit-i portfolio loss for the rest of the portfolio, Y_2. Suppose this time losses are generated from two Pareto distributions with the property, such that $\eta = 1.846395, \gamma = 26425.53$, and

$$Y_1 \sim \text{Pareto}\left(\gamma, \frac{\eta}{2}\right)$$
$$Y_2 \sim \text{Pareto}\left(\gamma, \frac{\eta}{5}\right)$$

The total loss is generated by adding the censored losses and the omit-i portfolio losses, Y_2.

$$\text{Total Loss} = \min(Y_1, u) + Y_2$$

The $\alpha = 0.95$ quantile $q_g(u)$ is a function of u, and hence, can be plotted for values of u in the range $[0, q_g(\infty)]$. Figure 4.4 shows the result.

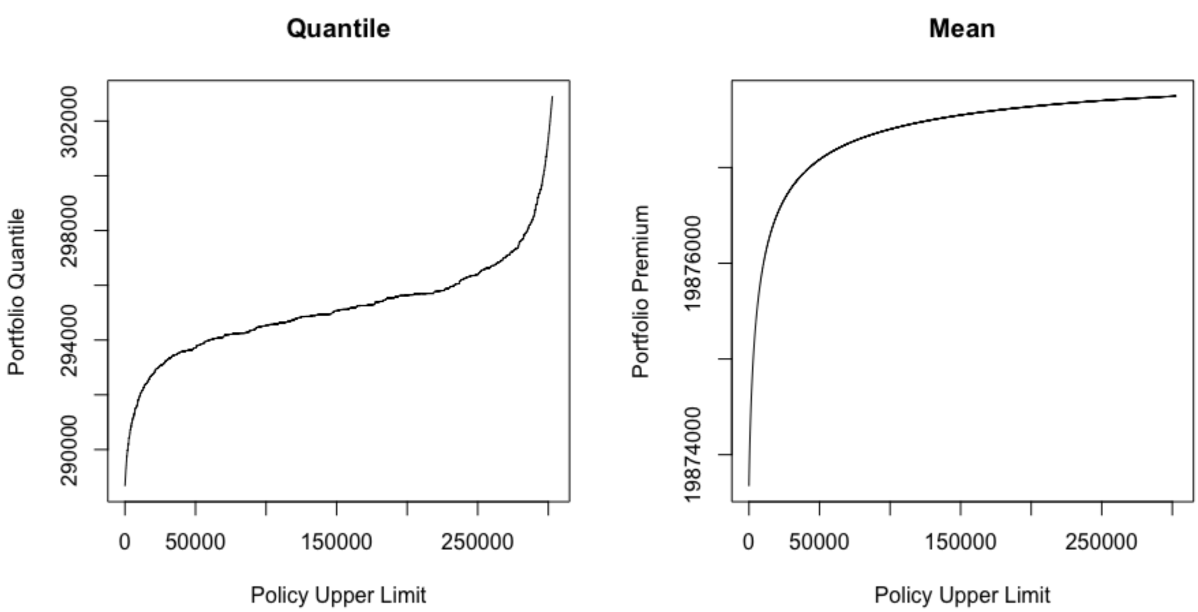


Figure 4.4: Single Policy with Portfolio: Quantile and Premium

The left panel of Figure 4.4 shows that $q_g(u)$ is not convex, but quasi-convex. The quantile can be optimized under the premium constraint

$$P_{sim}(u) \geq P_{min}$$

where $P_{sim}(u)$ is the mean of the simulated losses, and P_{min} defines the minimum requirement for the premium. The following figures illustrate the quantile function and premium function with respect to u, the parameter of interest. $B = 1,000,000$ replicates of the Pareto distribution were used to calculate the quantile and mean function values. From Figure 4.4, we can check that the quantile function $q_g(u)$ is not convex in the parameter u, due to the curvature introduced by the omit-i portfolio. Nevertheless, an optimal parameter u^* can be found under the constrained optimization framework, due to the quasi-convexity of $q_g(u)$.

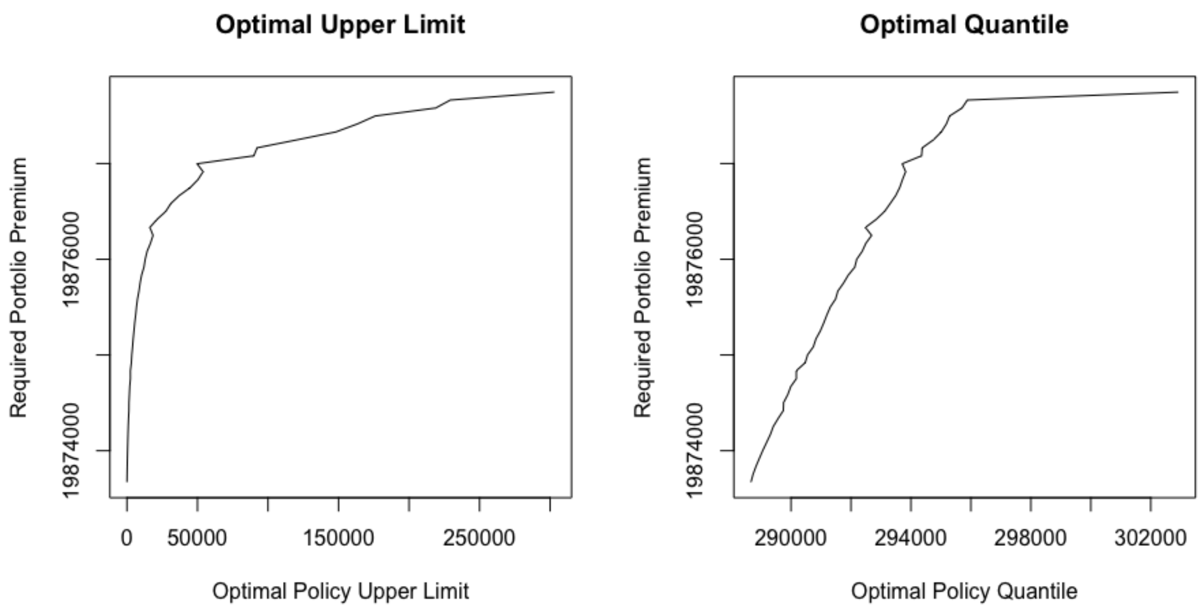


Figure 4.5: Single Policy with Portfolio: Optima

4.6 Aggregate Loss Risk Retention

4.6.1 Theory

We are now ready to see the general result for aggregate losses. We believe that these results are new to the literature. For a single policy with a single deductible, we have losses in the amount $Y_1, Y_2, \ldots, Y_N$ where N is the random number of claims. If each loss is subject to a per occurrence deductible d, then the insured loss (from the insurer's perspective) is

$$\text{Aggregate-Loss}_i(d_i) = \sum_{j=1}^{N_i} (Y_{ij} - d_i)_+ + S_{(i)}$$

for a given policy i. In general, the insurance company or the policyholder may be interested in adjusting retention parameters d, c, or u. In case the parameter change of interest is the coinsurance, we focus on the following aggregate loss:

$$\text{Aggregate-Loss}_i(c_i) = \sum_{j=1}^{N_i} c_i \cdot Y_{ij} + S_{(i)}$$

whereas if the upper limit parameter is of interest, then:

$$\text{Aggregate-Loss}_i(u_i) = \sum_{j=1}^{N_i} \min(Y_{ij}, u_i) + S_{(i)}$$

In each case, the aggregate loss quantile is a function of the risk retention parameter $\theta_i \in \{d_i, c_i, u_i\}$, so that $q_g(\theta_i)$ is to be minimized, under a premium constraint

$$P_{sim}(\theta_i^*) \geq P_{min}.$$

This is a quasiconvex optimization problem, which has a global solution under strict quasi-convexity of the quantile function $q_g(\theta_i)$. For background in quasiconvex optimization problems, see Boyd and Vandenberghe (2004), Section 3.4.5, and 4.2.5.

The following theorems establish quasi-convexity of the quantile functions, with respect to each of the parameters, d_i, c_i, and u_i.

Theorem 4.6.1. *$q_g(d_i)$ is strictly quasi-convex, and satisfies*

$$\frac{\partial q_g}{\partial d_i} = -\frac{c_i}{f_{S_g}(q_g)} \cdot E\left[\left(\sum_{j=1}^{N_i} I(Y_{ij} > d_i)\right) f_{S_{(i)}}\left(q_g - \sum_{j=1}^{N_i} g(Y_{ij})\right)\right] < 0$$

Proof. It suffices to show monotonicity, since monotonic functions are quasi-convex; See Boyd and Vandenberghe (2004), Page 99. For fixed $N_i = n_i$, define

$$\begin{aligned} H^{*n_i}(t) &= E_{Y_{i1},\dots Y_{in_i}}\left[Pr\left(\sum_{j=1}^{n_i} g(Y_{ij}) + S_{(i)} \le t)\right)\right] & (4.9) \\ &= E_{Y_{i1},\dots Y_{in_i}}\left[Pr\left(S_{(i)} \le t - \sum_{j=1}^{n_i} g(Y_{ij}))\right)\right] \\ &= E_{Y_{i1},\dots Y_{in_i}}\left[F_{S_{(i)}}\left(t - \sum_{j=1}^{n_i} g(Y_{i1})\right)\right] & (4.10) \end{aligned}$$

for policy i. If the portfolio consists of only one policy, then let

$$H^{*n_i}(t) = Pr(S_g \le t | N_i = n_i) = F_{g,S_i,n_i}(t) \quad (4.11)$$

Then, assuming differentiation and integration can be interchanged, we have:

$$
\begin{aligned}
\frac{\partial}{\partial d_i} H^{*n_i}(t) &= \frac{\partial}{\partial d_i} Pr(S_{g,n_i} \le t) \\
&= \frac{\partial}{\partial d_i} Pr\left(S_{(i)} + \sum_{j=2}^{n_i} g(Y_{ij}) + g(Y_{i1}) \le t\right) \\
&= \frac{\partial}{\partial d_i} E_{Y_{i1}}\left[H^{*(n_i-1)}(t - g(Y_{i1}))\right] \\
&= \frac{\partial}{\partial d_i}\left\{\int_{d_i}^{u_i} H^{*(n_i-1)}(t-g(y))dF_i(y) + \right. \\
&\qquad \left. H^{*(n_i-1)}(t)F_i(d_i) + H^{*(n_i-1)}(t-g(u_i))(1-F_i(u_i))\right\} \\
&= \int_{d_i}^{u_i} \frac{\partial}{\partial d_i} H^{*(n_i-1)}(t-g(y))dF_i(y) - H^{*(n-1)}(t)f_i(d_i) + H^{*(n_i-1)}(t)f_i(d_i) \\
&\quad + \frac{\partial}{\partial d_i} H^{*(n_i-1)}(t)F_i(d_i) + \frac{\partial}{\partial d_i} H^{*(n_i-1)}(t-g(u_i))(1-F_i(u_i)) \\
&= \int_{d_i}^{u_i} \frac{\partial}{\partial d_i} H^{*(n_i-1)}(t-g(y))dF_i(y) \\
&\quad + \frac{\partial}{\partial d_i} H^{*(n_i-1)}(t)F_i(d_i) + \frac{\partial}{\partial d} H^{*(n_i-1)}(t-g(u_i))(1-F_i(u_i)) \\
&= \int_0^\infty \frac{\partial}{\partial d_i} H^{*(n_i-1)}(t-g(y))dF_i(y), \qquad (4.12)
\end{aligned}
$$

where $H^{*(n_i-1)}(t-g(y)) = 0$ when $g(y) > t$. For the base case, we have

$$
\begin{aligned}
\frac{\partial}{\partial d_i} H^{*1}(t) &= \frac{\partial}{\partial d_i} Pr(S_g \le t | N_i = 1) \\
&= \frac{\partial}{\partial d_i} E_{Y_{i1}}\left[Pr\left(S_{(i)} \le t - g(Y_{i1})\right)\right] \\
&= \frac{\partial}{\partial d_i} E_{Y_i}\left[F_{S_{(i)}}(t - g(Y_i))\right] \\
&= \frac{\partial}{\partial d_i}\left\{\int_{d_i}^{u_i} F_{S_{(i)}}(t-g(y))dF_i(y) + \right. \\
&\qquad \left. F_{S_{(i)}}(t)F_i(d_i) + F_{S_{(i)}}(t-g(u))(1-F_i(u_i))\right\} \\
&= \int_{d_i}^{u_i} \frac{\partial}{\partial d_i} F_{S_{(i)}}(t-g(y))dF_i(y) - F_{S_{(i)}}(t)f_i(d_i) + F_{S_{(i)}}(t)f_i(d_i) \\
&\quad + \frac{\partial}{\partial d_i} F_{S_{(i)}}(t)F_i(d_i) + \frac{\partial}{\partial d_i} F_{S_{(i)}}(t-g(u_i))(1-F_i(u_i)) \\
&= \int_0^\infty \frac{\partial}{\partial d_i} F_{S_{(i)}}(t-g(y))dF_i(y). \qquad (4.13)
\end{aligned}
$$

Plugging in Equation (4.13) into (4.12), we have

$$\begin{aligned}\frac{\partial}{\partial d_i}H^{*n_i}(t) &= \int_0^\infty \ldots \int_0^\infty \frac{\partial}{\partial d_i}F_{S_{(i)}}\left(t-\sum_{j=1}^{n_i} g(y_{ij})\right) dF_i(y_{i1})\ldots dF_i(y_{in_i}) \\ &= -\int_0^\infty \ldots \int_0^\infty f_{S_{(i)}}\left(t-\sum_{j=1}^{n_i} g(y_{ij})\right)\left(\sum_{j=1}^{n_i}\frac{\partial g(y_{ij})}{\partial d_i}\right) dF_i(y_{i1})\ldots dF_i(y_{in}) \\ &= \int_0^\infty \ldots \int_0^\infty f_{S_{(i)}}\left(t-\sum_{j=1}^{n_i} g(y_{ij})\right)\left(c_i\cdot\sum_{j=1}^{n_i} I(y_{ij}>d_i)\right) dF_i(y_{i1})\ldots dF_i(y_{in}). \end{aligned} \tag{4.14}$$

Expression (4.13) is in fact identical to $A_1(t)$ in (Frees, 2016). When the deductible change is considered, we have

$$A_1(t) = \int_{t-b_i}^{t} f_i(g^{-1}(t-s))\frac{\partial g^{-1}}{\partial d_i}(t-s)dF_{S_{(i)}}(s) - [1-F_i(u_i)]f_{S_{(i)}}f_{S_{(i)}}(t-g(u_i))(-c_i) \tag{4.15}$$

where $\frac{\partial g^{-1}}{\partial d}(t-s)=1$. Simplifying this results in

$$A_1(t) = -c_i\left\{f_{S,\theta}(t) - F_i(d_i)f_{S_{(i)}}(t)\right\} \tag{4.16}$$

So that

$$\frac{\partial}{\partial d_i}q_{S,\theta} = -\frac{A_1(q_{S,\theta})}{f_{S,\theta}(q_{S,\theta})} = -c_i\left\{1-F_i(d_i)\cdot\frac{f_{S_{(i)}}(q_{S,\theta})}{f_{S,\theta}(q_{S,\theta})}\right\} \tag{4.17}$$

Let's compare this with expression (4.13), which is the case when a policy i has a single loss. Assuming a continuous distribution for $S_{(i)}$, application of chain rule to the expectation of (4.13) results in

$$
\begin{aligned}
\frac{\partial}{\partial d} q_{S,\theta} &= -\frac{1}{f_{S,\theta}(q_{S,\theta})} E\left[\frac{\partial}{\partial d_i} H^{*1}(q_{S,\theta})\right] \\
&= -\frac{1}{f_{S,\theta}(q_{S,\theta})} \int_0^\infty \frac{\partial}{\partial d_i} F_{S_{(i)}}(q_{S,\theta} - g(y)) dF_i(y) \\
&= -\frac{1}{f_{S,\theta}(q_{S,\theta})} \int_0^\infty f_{S_{(i)}}(q_{S,\theta} - g(y)) \cdot \left(\frac{\partial g(y)}{\partial d_i}\right) dF_i(y) \\
&= -c \int_0^\infty \frac{f_{S_{(i)}}(q_{S,\theta} - g(y))}{f_{S,\theta}(q_{S,\theta})} \cdot I(y > d_i) dF_i(y) \\
&= -cE\left[I(y > d_i) \frac{f_{S_{(i)}}(q_{S,\theta} - g(y))}{f_{S,\theta}(q_{S,\theta})}\right] \\
&= -c\left\{1 - F_i(d) \cdot \frac{f_{S_{(i)}}(q_{S,\theta})}{f_{S,\theta}(q_{S,\theta})}\right\} \qquad (4.18)
\end{aligned}
$$

where the last equality can be established using the identity

$$
E\left[I(y > d_i)\frac{f_{S_{(i)}}(t - g(y))}{f_{S,\theta}(t)} + I(y \leq d_i)\frac{f_{S_{(i)}}(t - g(y))}{f_{S,\theta}(t)}\right] = 1
$$

with $f_{S_{(i)}}(t - g(y)) = f_{S_{(i)}}(t)$ when $y \leq d_i$. □

Theorem 4.6.3 establishes strict quasi-convexity of the quantile function in the parameter d_i. The next results establish quasi-convexity in the parameter u_i and c_i.

Theorem 4.6.2. *$q_g(u_i)$ is strictly quasi-convex, and satisfies*

$$
\frac{\partial q_g}{\partial u_i} = E\left[\left(\sum_{j=1}^{N_i} I(Y_{ij} > u_i)\right) f_{S_{(i)}}\left(q_g - \sum_{j=1}^{N_i} g(Y_{ij})\right)\right] \frac{1}{f_{S_g}(q_g)} \cdot (c_i) > 0
$$

Proof. Assuming differentiation and integration can be interchanged, we have:

$$\begin{aligned}
\frac{\partial}{\partial u_i} H^{*n_i}(t) &= \frac{\partial}{\partial u_i} Pr(S_{g,n_i} \leq t) \\
&= \frac{\partial}{\partial u_i} Pr\left(S_{(i)} + \sum_{j=2}^{n_i} g(Y_{ij}) + g(Y_{i1}) \leq t\right) \\
&= \frac{\partial}{\partial u_i} E_{Y_{i1}}\left[H^{*(n_i-1)}(t - g(Y_{i1}))\right] \\
&= \frac{\partial}{\partial u_i}\left\{\int_{d_i}^{u_i} H^{*(n_i-1)}(t-g(y))dF_i(y) + \right. \\
&\qquad \left. H^{*(n_i-1)}(t)F_i(d_i) + H^{*(n_i-1)}(t-g(u_i))(1-F_i(u_i))\right\} \\
&= \int_{d_i}^{u_i} \frac{\partial}{\partial u_i} H^{*(n_i-1)}(t-g(y))dF_i(y) + \\
&\qquad H^{*(n-1)}(t-g(u_i))f_i(u_i) + \frac{\partial}{\partial u_i} H^{*(n_i-1)}(t)F_i(d_i) + \\
&\qquad \frac{\partial}{\partial u_i} H^{*(n_i-1)}(t-g(u_i))(1-F_i(u_i)) - H^{*(n-1)}(t-g(u_i))f_i(u_i) \\
&= \int_{d_i}^{u_i} \frac{\partial}{\partial u_i} H^{*(n_i-1)}(t-g(y))dF_i(y) + \\
&\qquad \frac{\partial}{\partial u_i} H^{*(n_i-1)}(t)F_i(d_i) + \frac{\partial}{\partial u_i} H^{*(n_i-1)}(t-g(u_i))(1-F_i(u_i)) \\
&= \int_0^{\infty} \frac{\partial}{\partial u_i} H^{*(n_i-1)}(t-g(y))dF_i(y), \qquad (4.19)
\end{aligned}$$

For the base case, a similar proof shows that the base case is given by

$$\frac{\partial}{\partial u_i} H^{*1}(t) = \int_0^{\infty} \frac{\partial}{\partial u_i} F_{S_{(i)}}(t - g(y))dF_i(y). \qquad (4.20)$$

Plugging in (4.20) into (4.19), we have

$$\frac{\partial}{\partial u_i} H^{*n_i}(t) = -\int_0^{\infty} \ldots \int_0^{\infty} f_{S_{(i)}}\left(t - \sum_{j=1}^{n_i} g(y_{ij})\right)\left(\sum_{j=1}^{n_i} \frac{\partial g(y_{ij})}{\partial u_i}\right) dF_i(y_{i1}) \ldots dF_i(y_{in}) \qquad (4.21)$$

where

$$\frac{\partial g(y_{ij})}{\partial u_i} = \begin{cases} c_i & \text{for } y_{ij} > u_i \\ 0 & \text{otherwise} \end{cases}$$

□

Theorem 4.6.3. *$q_g(c_i)$ is strictly quasi-convex, and satisfies*

$$\frac{\partial q_g}{\partial c_i} = E\left[\left(\sum_{j=1}^{N_i}(Y_{ij}-d_i)\cdot I(Y_{ij}>d_i)-(Y_{ij}-u_i)\cdot I(Y_{ij}>u_i)\right)\right.$$
$$\left.\cdot f_{S_{(i)}}\left(q_g-\sum_{j=1}^{N_i}g(Y_{ij})\right)\right]\frac{1}{f_{S_g}(q_g)}>0$$

Proof. Assuming differentiation and integration can be interchanged, we have:

$$\begin{aligned}
\frac{\partial}{\partial c_i}H^{*n_i}(t) &= \frac{\partial}{\partial c_i}Pr(S_{g,n_i}\le t)\\
&= \frac{\partial}{\partial c_i}Pr\left(S_{(i)}+\sum_{j=2}^{n_i}g(Y_{ij})+g(Y_{i1})\le t\right)\\
&= \frac{\partial}{\partial c_i}E_{Y_{i1}}\left[H^{*(n_i-1)}(t-g(Y_{i1}))\right]\\
&= \frac{\partial}{\partial c_i}\left\{\int_{d_i}^{u_i}H^{*(n_i-1)}(t-g(y))dF_i(y)+H^{*(n_i-1)}(t)F_i(d_i)+\right.\\
&\qquad \left.H^{*(n_i-1)}(t-g(u_i))(1-F_i(u_i))\right\}\\
&= \int_{d_i}^{u_i}\frac{\partial}{\partial c_i}H^{*(n_i-1)}(t-g(y))dF_i(y)+\frac{\partial}{\partial c_i}H^{*(n_i-1)}(t)F_i(d_i)+\\
&\qquad \frac{\partial}{\partial c_i}H^{*(n_i-1)}(t-g(u_i))(1-F_i(u_i))\\
&= \int_0^{\infty}\frac{\partial}{\partial c_i}H^{*(n_i-1)}(t-g(y))dF_i(y). \qquad (4.22)
\end{aligned}$$

For the base case, we have:

$$\begin{aligned}
\frac{\partial}{\partial c_i}H^{*1}(t) &= \frac{\partial}{\partial c_i}Pr(S_g\le t|N_i=1)\\
&= \frac{\partial}{\partial c_i}E_{Y_{i1}}\left[Pr\left(S_{(i)}\le t-g(Y_{i1})\right)\right]\\
&= \frac{\partial}{\partial c_i}E_{Y_i}\left[F_{S_{(i)}}(t-g(Y_i))\right]\\
&= \frac{\partial}{\partial c_i}\left\{\int_{d_i}^{u_i}F_{S_{(i)}}(t-g(y))dF_i(y)+F_{S_{(i)}}(t)F_i(d_i)+\right.\\
&\qquad \left.F_{S_{(i)}}(t-g(u))(1-F_i(u_i))\right\}\\
&= \int_0^{\infty}\frac{\partial}{\partial c_i}F_{S_{(i)}}(t-g(y))dF_i(y). \qquad (4.23)
\end{aligned}$$

Hence, assuming $S_{(i)}$ has a continuous and differentiable distribution, we have

$$\frac{\partial}{\partial c_i} H^{*n_i}(t) = -\int_0^\infty \cdots \int_0^\infty f_{S_{(i)}}\left(t - \sum_{j=1}^{n_i} g(y_{ij})\right) \cdot \left(\sum_{j=1}^{n_i} \frac{\partial g(y_{ij})}{\partial c_i}\right) dF_i(y_{i1}) \ldots dF_i(y_{in}) \tag{4.24}$$

where

$$\frac{\partial g(y_{ij})}{\partial c_i} = \begin{cases} y_{ij} - d_i & \text{for } y_{ij} \le d_i \\ u_i - d_i & \text{for } y_{ij} > d_i \end{cases}$$

□

Hence, we can see that monotonicity of the quantile function ensures quasi-convexity. Meanwhile, the reader may realize that quasi-convexity of the quantile function is ensured for the single-loss cases shown in Section 4.5, by restricting $N_i = 1$.

4.6.2 Numerical Optimization for the Aggregate Loss Problem

Let us first take a look at the convexity of the associated premium and quantile functions. Suppressing the subscript i, let the losses be generated from the assumptions $\eta = 1.846395$, $\gamma = 26425.53$, $\lambda = 1$, which represents Green Bay city within the LGPIF data (see Chapter 5, Section 5.1, Table 5.11 for parameter estimation details) using the models:

$$N \sim \text{Poisson}(\lambda)$$
$$Y_j \sim \text{Pareto}(\gamma, \eta), \quad \text{for } j = 1, ..., N$$

Each loss can be transformed, and aggregated, so that the Aggregate-Loss(u) which is a function of u, can be optimized over u with constraint

$$P_{sim}(u^*) \ge P_{min}$$

where $P_{sim}(u^*)$ is the mean of all the replicates of Aggregate-Loss(u^*).

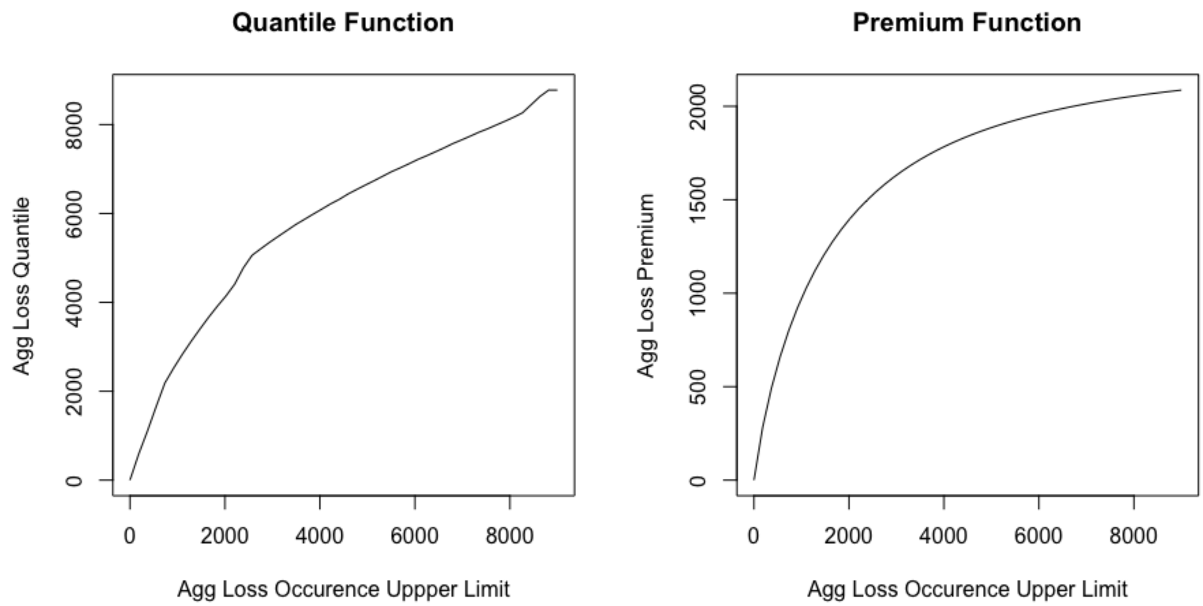


Figure 4.6: Aggregate Risk: Quasi-Convexity

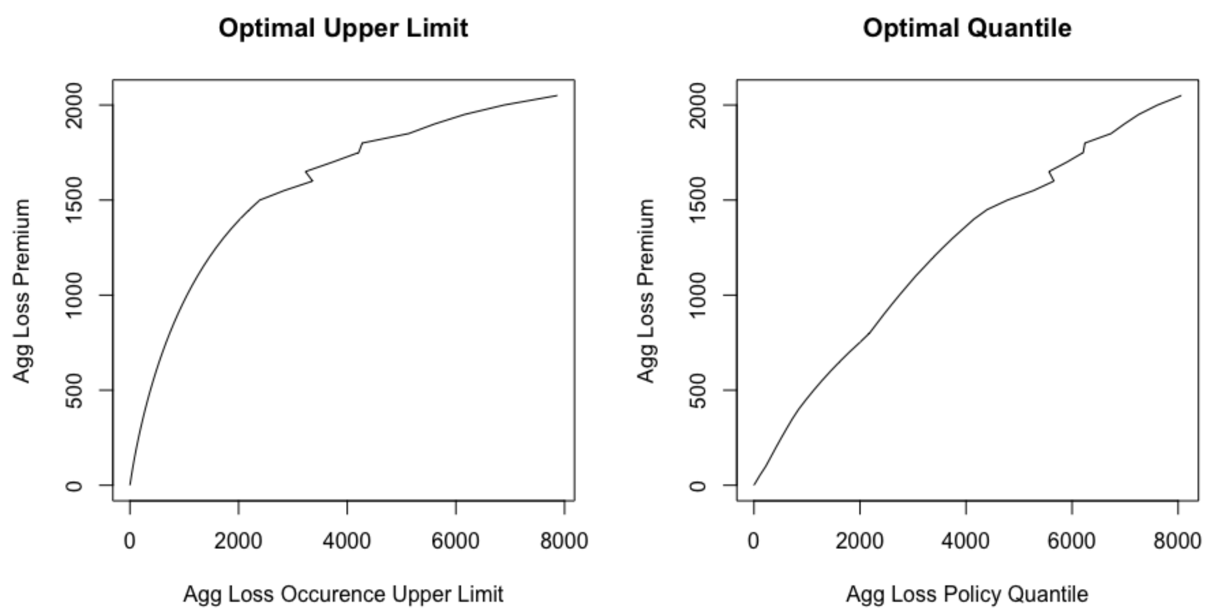


Figure 4.7: Aggregate Risk: Optima

These figures suggest that both the quantile and the premium functions are quasi-convex in the upper limit u. Quasi-convexity allows us to optimize the quantile function using constrained optimization routines such as `auglag` in R. Thus, the optimal upper limit u^* and optimal quantile $q(u^*)$ can be found.

4.7 Blocks of Policies

Now that we are familiar with how the optimization works, let us consider more practical situations. Suppose for a single policy, we have losses in the amount $Y_1, Y_2, \ldots, Y_p$ where p is the number of policies within a block of policies sharing similar features. For example, an insurance company may be providing coverage for different entities. Thus, if there are say, p school entities, then the manager may be interested in adjusting the upper limit for all school entities. Let's focus on the upper limit case, so that each loss is subject to a per occurrence upper limit u, then the insured loss is

$$\text{Block-Loss}(u) = \sum_{j=1}^{p} \min(Y_j, u)$$

Suppose we are interested in applying specific upper limits for each block j. In this case we have

$$\text{Block-Loss}(u) = \sum_{j=1}^{p} \min(Y_j,\ \mu_j \cdot u), \qquad \text{where } \mu_j = E[Y_j]$$

where each of the random variables $Y_1, ..., Y_p$ represents a coverage within a block. For example, p could be the number of school entities, each of which is subject to an upper limit of $\mu_j \cdot u$, for some u. In the following simulation, the optimal u is found by constrained optimization.

Simulation

$p = 10$ policies are randomly generated so that

$$\text{Block-Loss}(u) = \sum_{j=1}^{p} \min(Y_j,\ \mu_j \cdot u)$$

$$Y_j \sim \text{Pareto}(\gamma_j, \eta)$$

$$\gamma_j \sim \text{Uniform}[50, 5000]$$

$$\eta = 3, \quad j = 1, ..., 10$$

where γ_j is the scale parameter of the policies $j = 1, ...10$, and η is the shape parameter of the Type-II Pareto distribution. Initially, the mean and quantile functions are plotted with respect to u, where $u \in [0, 5]$. The first two panels of the plot shows the mean and quantile functions are quasi-convex in the parameter u.

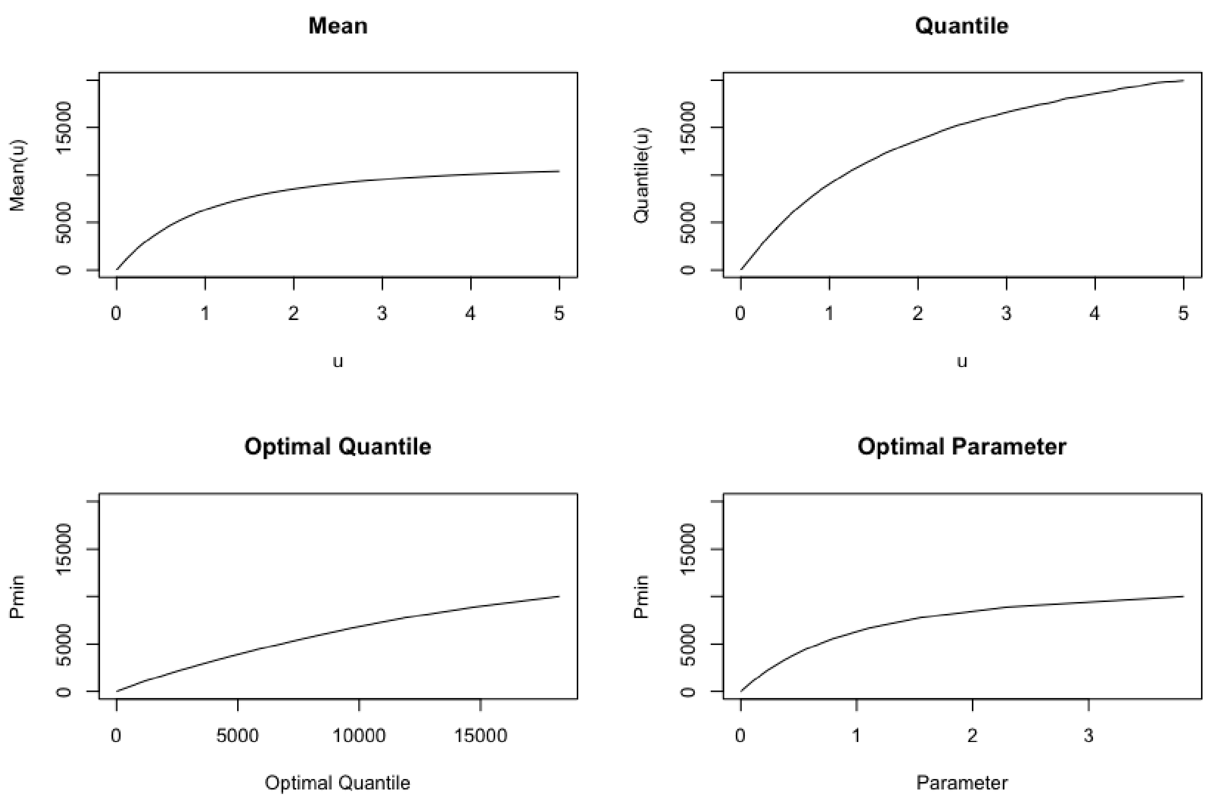


Figure 4.8: Block Policies Example

The first two panels of Figure 4.8 suggest that the premium level may vary within the interval

$[0, 10, 000]$. For this reason, restricting the domain of premium (mean) to $[0, 10, 000]$ results in a compact parameter space, and the optimal quantile can be found. The last two panels show the optimal quantile, and the optimal parameter u^*, as the minimum premium restriction is varied from 0 to 10,000.

4.8 Supplementary Notes

The risk-sharing parameters can potentially be in effect at 4 levels: 1. individual claims; 2. all claims within a line, 3. all claims within a policy, and 4. claims within the portfolio. We also index these parameters by i for the first three levels to indicate that they may vary by policy. For a specific case, We can collect all the risk-sharing parameters into a single vector, so that

$$\boldsymbol{\theta} = \{(c_{il}, d_{il}, u_{il}, c_4, d_4, u_4);\ i = 1, ..., n,\ l = 1, ..., 3\}.$$

Note that if we do not assume that g is smooth in $\boldsymbol{\theta}$, then taking derivatives with respect to $\boldsymbol{\theta}$ is not an option. Consider a policyholder and a typical Z consisting of

$$Z = \big(N, Y_1, ..., Y_N, S_{Y,N}, \bar{S}_{Y,N}\big)$$

for number of claims N, individual claims $\{Y_j\}$, aggregate severity S_{YN} and average severity $\bar{S}_{YN}$. Based on observations of this process, one can develop simple rules to determine prices P, say, for each outcome, by using regression models for each marginal. For a general random variable Z, and a transformation g, we denote $g(Z)$ as the transformed random variable. Our interest is in summarizing this random variable by a pricing functional

$$P(g(Z; \boldsymbol{\theta})) = \int p(g(z)) dF(z)$$

and a risk measure functional

$$RM(g(Z; \boldsymbol{\theta})) = \int F^{-1}_{g(Z)}(t) dK(t)$$

where $F_{g(Z)}$ is the distribution function of $g(Z)$. In general, the pricing functional looks at the center of the retained risks distribution and ignores dependencies. In contrast, the risk measure functional focuses on the tails and incorporates dependencies. We can think of the pricing function as just the identity operator so that prices are expected values. We can think of the risk measure at the 99^{th} percentile of the distribution, known as the VAR.

Take $n = 1, p = 1$ and $N = 1$, so that we have one monoline policy with a single claim. In this special case, the risk-sharing parameters at levels 2, 3, and 4 are redundant and can be ignored. With the identity operator for prices, we have

$$P(g(Z); c, d, u) = \int g(z; c, d, u) dF_Y(z) = c \cdot E_Y(\min(Y, u) - \min(Y, d)),$$

a familiar relationship in actuarial science, e.g., (Klugman et al., 2012), page 188. To date, the substantial work in the literature has focused on risk retention parameters that are based on policies. That is, discovery of an optimal policy is based on preferences concerning a single risk Y. In contrast, very common contract features allow for risk retention parameters to be on a per-occurrence, or per-loss basis. In this dissertation chapter, advice to insurers are provided when parameters are defined on a per-loss basis. To illustrate, consider the situation where an insurer's ith risk has N_i, losses $Y_{i1}, \ldots, Y_{iN_i}$. For a single loss, Y_{ij}, the insurer pays $g(Y_{ij})$. Then, the total risk for the insurer's portfolio can be expressed as

$$S_g = \sum_{j=1}^{N_i} g(Y_{ij}) + S_{(i)}.$$

We know of no other work in the risk and insurance literature that provides advice to risk managers about risk retention parameters that operate on a per loss basis. In Section 4.8.1, advice for two layers of parameters is provided.

4.8.1 Single Policy per Occurrence and per Policy Retention

Focusing on the upper limit, so that $c = 1$ and $d = 0$, we are now interested in exploring the case with two layers of risk retention parameters. We have:

$$\text{Aggregate-Loss}(u_1, u_2) = \min\left(\left[\sum_{j=1}^{N} \min(Y_j, u_1)\right], u_2\right)$$

In the following example, losses are simulated from identical parameters as previous sections, after which u_1 and u_2 are applied. The resulting quantile q of the aggregate loss is a function in the two parameters u_1 and u_2,

$$q_{sim}(u_1, u_2) = \text{Quantile of the aggregate loss with parameters } u_1, u_2$$

which can be optimized under the constraint

$$P_{sim}(u_1^*, u_2^*) \geq P_{min}$$

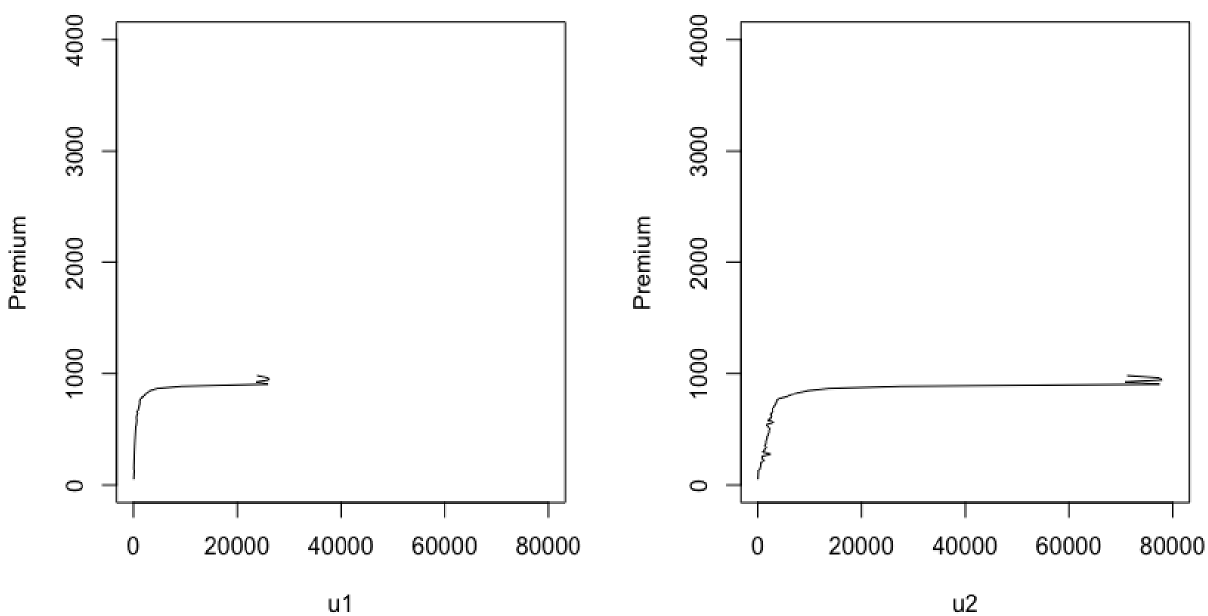


Figure 4.9: Two Upper Limits: Optima

4.8.2 With Portfolio Loss

For the two upper-limit parameter case with portfolio losses, we have

$$\text{Aggregate-Loss}(u_1, u_2) = \min\left(\left[\sum_{j=1}^{N} \min(Y_j, u_1)\right],\ u_2\right) + Y_2$$

The result is shown below:

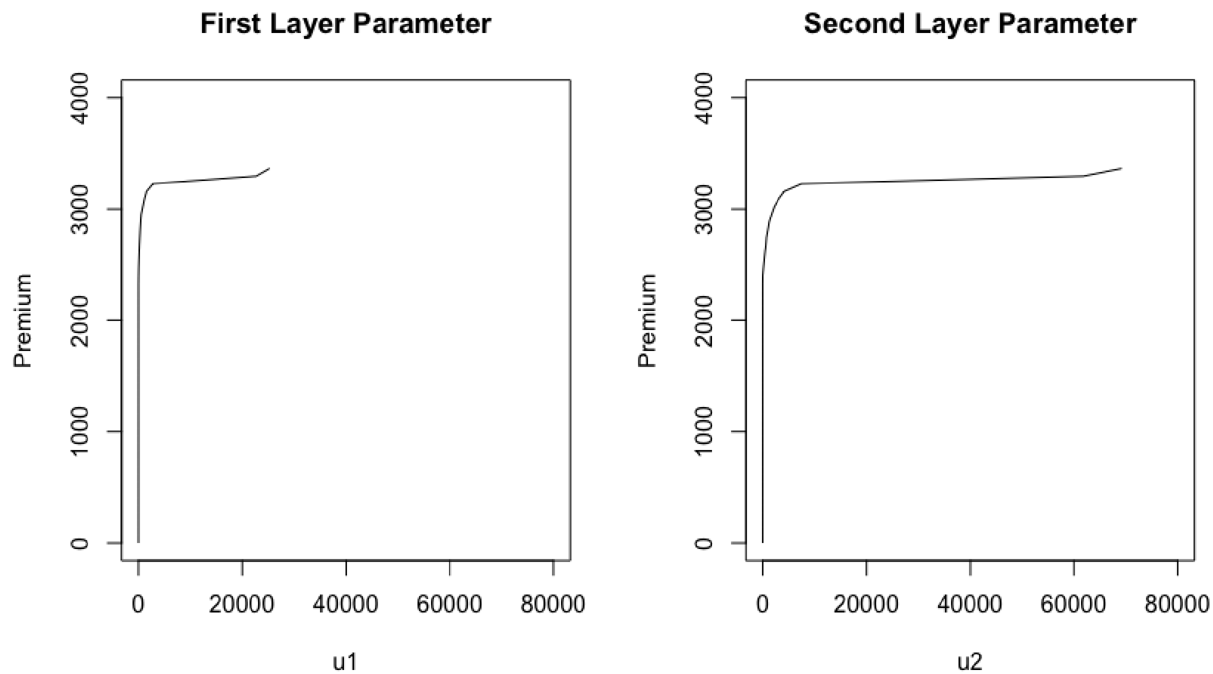


Figure 4.10: Two Upper Limits with Portfolio: Optima

The framework in this section is applied to the LGPIF Madison Metropolitan School District policy optimization case study in Chapter 5, Section 5.3.

4.8.3 Future Work

In insurance risk management, it is helpful to provide advice to risk managers on the selection of risk retention parameters. The retention parameters may include deductibles (d), coinsurance (c), and upper limits (u). In order to provide solutions to this question, statistical models may be helpful. In this chapter, we illustrate how the statistical model in Chapter 2 can help answer the risk retention problem. Subsequently, it may be interesting to investigate altering these initial approaches based on various risk assessment models.

One may think about a general risk retention rule "$f(\cdot)$" that is applied to outcome Z. For this function, we may be interested in (i) applying to individual or aggregate claims by line, (ii) representing deductibles, policy limits, or coinsurance, by line, (iii) Restricting and introducing endorsements (extra coverages), (iv) applying $f(\cdot)$ to number of claims (v) applying $f(\cdot)$ to several lines simultaneously. The special case, $f(\cdot) = 0$, means that the insurer retains 0 claims, as in rejecting a policy. Let

$$S_Z(f) = \sum_{i=1}^{n} f(Z_i)$$

represent the sum of retained claims. We could apply another layer of risk retention rules to the entire portfolio (for reinsurance).

For future work, we would like to understand the distribution of $S_Z(f)$, paying particular attention to (a) affects of contagion, (b) large exposures, large claims, (c) association among lines. Guidelines for risk acceptance and retention may be established using our framework. In our framework, we are especially interested in association over time, such as dividends, credibility. Contagion may be easier to model if we look at causes/perils of risks, e.g. hail for homeowners, earthquakes, floods, etc. Are there interesting risk-sharing rules that occur over time? There are potential applications in dividends, credibility, and the like, where another risk/portfolio, say $S_{Z,2}(f)$ is potentially added to the portfolio. We are interested in providing guidelines for when we might accept additional risks. In conjuction with the determination of risk acceptance guidelines, we might think about alternative pricing structures.

4.8.4 RM^2 Computation

In this section, Monte-Carlo procedures for quantiles and the portfolio densities are discussed. We first review some fundamental concepts of distribution function and quantile sensitivities. Consider a generic random variable Y and evaluate this using the parametric function $g(Y, \theta)$. Use the notation $F_g(y; \theta) = Pr(g(Y, \theta) \leq y)$ for the corresponding distribution function. Further, use

$$\partial_\theta = \frac{\partial}{\partial \theta}$$

for a partial derivative with respect to θ and similarly for ∂_y and so forth. For a fixed α, define the quantile function $q_\alpha = F_g^{-1}(\alpha)$.

Assumption 1. The function $\partial_\theta g(Y, \theta)$ exists with probability 1 for any θ in an open set Θ. There *exists a function* $k(\cdot)$ *such that* $Ek(Y) < \infty$ *and*

$$|g(Y, \theta_2) - g(Y, \theta_1)| \le k(Y)|\theta_2 - \theta_1|$$

for all θ_1, θ_2 *in* Θ.

Assumption 2. For θ *in an open set* Θ, $\partial_y F_g(y;\theta) = f_g(y;\theta)$ *is continuous in a neighborhood of* y, *and* $\partial_y F_h(y;\theta)$ *exists and is continuous with respect to* θ *and* y.

Under assumption 1 and 2, the distribution function sensitivity is

$$\frac{\partial}{\partial\theta} F_g(y;\theta) = -\frac{\partial}{\partial y} E\left\{\frac{\partial}{\partial\theta} g(Y,\theta) I(g(Y,\theta) \le y)\right\} \tag{4.25}$$

The quantile sensitivity is

$$\frac{\partial}{\partial\theta} q_\alpha = -\frac{\partial_\theta F_g(q_\alpha(\theta))}{f_g(q_\alpha(\theta))} = -\frac{\partial_\theta F_g(q_\alpha(\theta))}{\partial_y F_g(q_\alpha(\theta))}\bigg|_{y=q_\alpha(\theta)} \tag{4.26}$$

The proof of the Lemmas is part of Theorem 1 and 2 of (Hong, 2009) (see also the discussion in (Jiang and Fu, 2015)). We now apply the Lemmas to the general framework of risk retention problems. This is known in actuarial mathematics as the "aggregate loss model." It is helpful to review the assumptions underpinning this model.

Let N_i represent the number of claims for the ith risk. Given $N_i = n$, the claims $\{Y_{i1}, \ldots, Y_{in}\}$ are i.i.d. The sum of claims from other risks is $S_{(i)}$. Assume that N_i, $\{Y_{ij}, j = 1, 2, \ldots\}$ and $S_{(i)}$ are mutually independent. The sum of portfolio claims is

$$S_g = \sum_{j=1}^{N_i} g(Y_{ij}) + S_{(i)}.$$

Using the Lemmas, we can express the distribution function sensitivity as

$$\frac{\partial}{\partial\theta} Pr(S_g \le t) = -\frac{\partial}{\partial t} E\left\{\sum_{j=1}^{N_i} \frac{\partial}{\partial\theta} g(Y_{ij}) I(S_g \le t)\right\} \tag{4.27}$$

Following the notation in Frees (2016), we use $q_{S,g,\alpha}$ for the α quantile of S_g. Then, using Equation (4.27), we can express the quantile sensitivity as

$$\partial_\theta q_{S,g,\alpha} = -\frac{\partial_\theta Pr(S_g \le q_{S,g,\alpha})}{f_{S_g}(q_{S,g,\alpha})} \tag{4.28}$$

To evaluate the quantile sensitivity, we now describe the use of Monte-Carlo techniques. With simulated realizations of S_g, it is straightforward to evaluate quantiles $q_{S,g,\alpha}$ and density functions f_{S_g}. More effort is needed to evaluate the distribution function sensitivity (at the quantile $q_{S,g,\alpha}$). To this end, we first briefly outline a procedure to evaluate quantiles and densities and then describe procedures to evaluate distribution function sensitivities.

For each risk $i = 1, \ldots, n$ in the portfolio, generate $\left\{N_i^{(r)}, Y_{i1}^{(r)}, \ldots Y_{iN_i}^{(r)}\right\}$ for $r = 1, \ldots, R$ simulated replications. From this, determine

$$S_{(i)}^{(r)} = \sum_{k \neq i} \sum_{j=1}^{N_k^{(r)}} Y_{kj}^{(r)}$$

and

$$S_g^{(r)} = \sum_{j=1}^{N_i^{(r)}} g\left(Y_{ij}^{(r)}\right) + S_{(i)}^{(r)}$$

From this sequence

$$\left\{S_g^{(r)}\right\}_{r=1}^{R},$$

we determine a Monte-Carlo approximation of $q_{S,g,\alpha}$ and the density evaluated at the quantile, $f_{S_g}(q_{S,g,\alpha})$. Beginning by taking expectations over $S_{(i)}$ and simplifying allows writing

$$\begin{aligned}
\frac{\partial}{\partial\theta} Pr(S_g \le t) &= -\frac{\partial}{\partial t} E\left\{\sum_{j=1}^{N_1} \frac{\partial}{\partial\theta} g(Y_{ij};\theta)\ I\left(\sum_{j=1}^{N} g(Y_{ij};\theta) + S_{(i)} \le t\right)\right\} \\
&= -\frac{\partial}{\partial t} E\left\{\sum_{j=1}^{N_1} \frac{\partial}{\partial\theta} g(Y_{ij};\theta)\ F_{S_{(i)}}\left(t - \sum_{j=1}^{N} g(Y_{ij};\theta)\right)\right\} \\
&= -E\left\{\sum_{j=1}^{N_i} \frac{\partial}{\partial\theta} g(Y_{ij};\theta)\ f_{S_{(i)}}\left(t - \sum_{j=1}^{N_i} g(Y_{ij};\theta)\right)\right\}
\end{aligned}$$

assuming the omit i portfolio distribution is continuous with density $f_{S_{(i)}}(\cdot)$. Define $k_R(\cdot;\cdot)$ to be a kernel density estimator with bandwidth b_R, e.g., $k_R(X;t) = I(t - b_R < X \le t + b_R)/(2b_R)$ for a generic random variable X. For each risk retention parameter change, we can obtain compact notations by conditioning the expression on the frequency of the ith policyholder. The distribution function for the aggregate claims conditional on the frequency is defined by

$$H^{*n_i}(t) = Pr(S_{g,n_i} \leq t) = Pr(S_g \leq t | N_i = n_i).$$

We let $h^{*n_i}(t) = \partial_t H^{*n_i}(t)$ be the corresponding density. Then, for each special case $\theta_i = d_i, c_i, u_i$, it has been shown (in Section 4.6) that

$$\frac{\partial}{\partial \theta_i} H^{*n_i}(t) = \int_0^\infty \dots \int_0^\infty \left(\sum_{j=1}^{n_i} \frac{\partial}{\partial \theta_i} g(y_{ij}) \right) f_{S_{(i)}} \left(t - \sum_{j=1}^{n_i} g(y_{ij}) \right) dF_i(y_{i1}) \dots dF_i(y_{in}), \tag{4.29}$$

which could be evaluated by simulating the claims, and recording the resulting $\partial_{\theta_i} g(y_{ij})$ values and $g(y_{ij})$ values for each replicate. Taking the expectation of Equation (4.29) gives

$$\frac{\partial}{\partial \theta_i} Pr(S_g \leq t) = E_{N_i} \left[\frac{\partial}{\partial \theta_i} H^{*n_i}(t) \right]. \tag{4.30}$$

The gradient to find the aggregate claim quantile minimizing parameter choice for a single parameter optimization over θ_i becomes

$$\partial_{\theta_i} q_{S,g} = \frac{1}{f_{S_g}(q_{S,g,\alpha})} \cdot E_{N_i} \left[\frac{\partial}{\partial \theta_i} H^{*n_i}(t) \right] \tag{4.31}$$

4.8.5 Conditions for Existence and Uniqueness of Global Optima

In this subsection, let us formalize the condition under which a unique global optima for the risk retention parameter could be found. For a review of convex optimization, see: Simon and Blume (1994), Sundaram (1996), Boyd and Vandenberghe (2004). The following theorem provides a basic condition in which a solution exists.

Weierstrass Theorem: Let D be a compact subset of $\mathbb{R}^n$ and $f : D \to \mathbb{R}^n$ be a continuous function on D. Then there exists $\boldsymbol{x}_m$ and $\boldsymbol{x}_M$ in D such that $f(\boldsymbol{x}_m) \leq f(x) \leq f(\boldsymbol{x}_M)$ for all $x \in D$.

See Sundaram (1996), Page 90.

For a special class of functions, known as *quasi-concave* (or *quasi-convex*) functions, the global optima is unique in D. *quasi-concave* functions are defined. Let $f : D \to \mathbb{R}$. The *upper-contour set* of f at $a \in \mathbb{R}$ is defined as

$$U_f(a) = \{x \in D | f(x) \geq a\}$$

while the *lower-contour set* of f at $a \in \mathbb{R}$ is defined as

$$L_f(a) = \{x \in D | f(x) \leq a\}$$

The function f is said to be *quasi-concave* on D if $U_f(a)$ is a convex set for each a. It is called *quasi-convex* on D if $L_f(a)$ is a convex set for each a. A theorem in optimization theory states that f is *quasi-concave* on D if and only if for all $x, y \in D$ and for all $\lambda \in (0, 1)$, it is the case that

$$f[\lambda x + (1 - \lambda)y] \geq \min[f(x), f(y)]$$

The function f is *quasi-convex* on D if and only if for all $x, y \in D$ and for all $\lambda \in (0, 1)$, it is the case that

$$f[\lambda x + (1 - \lambda)y] \leq \max[f(x), f(y)]$$

The following theorem provides a condition under which, a local optima for a quasi-concave (quasi-convex) function is global.

Theorem: Suppose $f : D \to \mathbb{R}$ is strictly quasi-concave (quasi-convex), where $D \in \mathbb{R}^n$ is convex. Then, any local maximum (minimum) of f on D is also a global maximum (minimum) of f on D. Moreover, the set of maximizers (minimizers) of f on D is either empty or a singleton. See Sundaram (1996), Page 213.

Thus, unique global optima may be found, by restricting the search domain to closed intervals. This will be sufficiently useful for many practical applications in insurance risk retention.

Chapter 5

LGPIF Case Study

Abstract

We demonstrate the approaches in this dissertation using a data set from the Wisconsin Local Government Property Insurance Fund (LGPIF). A detailed description of the LGPIF is included in Chapter 7. Here, brief summary statistics and details of the application of the approaches are described. The estimated loss models are applied and focused towards the final section, which applies the loss distributions to a case study on the Madison Metropolitan School District risk retention problem. We find that optimal risk retention parameters can be determined using the loss models developed in earlier chapters of the dissertation.

This chapter is based on the empirical results from:

- Frees, Edward W., and Gee Lee, and Lu Yang (2015). "Multivariate Frequency-Severity Regression Models in Insurance," *Risks,* Vol. 4(1)
- Lee, Gee Y., (forthcoming). "General Insurance Deductible Ratemaking," *Conditionally Accepted by the North American Actuarial Journal.*

Section 5.1, 5.2, and 5.3 each show empirical results using the LGPIF data for chapters 2, 3, 4 respectively.

5.1 Insurance Claim Modeling

Chapter 7 provides a more detailed summary of the LGPIF rating engine. Section 5.1.1 provides basic summary statistics for the LGPIF policyholder level data.

5.1.1 Basic Summary Statistics

The data are split into a training sample, and a validation sample. Table 5.1 shows the sample size for the training sample. Table 5.2 shows the validation sample. The data consist of six coverage groups; building and content (BC), contractor's equipment (IM), comprehensive new (PN), comprehensive old (PO), collision new (CN), collision old (CO) coverage. The data are longitudinal, and Tables 5.1 and 5.2 provide summary statistics for the frequencies and severities of claims within the in-sample years 2006 to 2010, and the validation sample 2011.

Table 5.1: Data Summary by Coverage, 2006–2010 (Training Sample).

	Average Frequency	Average Severity	Annual Claims in Each Year	Average Coverage (Million)	Number of Claims	Number of Observations
BC	0.879	9868	17,143	37.050	4992	5660
IM	0.056	624	766	0.848	318	4622
PN	0.159	197	466	0.158	902	1638
PO	0.103	311	504	0.614	587	2138
CN	0.127	374	744	0.096	720	1529
CO	0.120	538	951	0.305	680	2013

Table 5.2: Data Summary by Coverage, 2011 (Validation Sample).

	Average Frequency	Average Severity	Annual Claims in Year	Average Coverage (Million)	Number of Claims	Number of Observations
BC	0.945	8352	20,334	42.348	1038	1095
IM	0.076	382	645	0.972	83	904
PN	0.224	307	634	0.172	246	287
PO	0.128	220	312	0.690	140	394
CN	0.125	248	473	0.093	137	268
CO	0.081	404	656	0.375	89	375

Table 5.3 describes each coverage group. Automobile coverage is subdivided into four subcategories, which correspond to combinations for collision versus comprehensive and for new versus old cars.

Table 5.3: Description of Coverage Groups

Code	Name of Coverage	Description
BC	Building and Contents	This coverage provides insurance for buildings and the properties within. In case the policyholder has purchased arider, claims in this group may reflect additional amounts covered under endorsements.
IM	Contractor's Equipment	IM, an abbreviation for "inland marine" is used as the coverage code for equipments coverage, which originally belong to contractors.
C	Collision	This provides coverage for impact of a vehicle with an object, impact of vehicle with an attached vehicle, or overturn of a vehicle.
P	Comprehensive	Direct and accidental loss or damage to motor vehicle, including breakage of glass, loss caused by missiles, falling objects, fire, theft, explosion, earthquake, windstorm, hail, water, flood, malicious mischief or vandalism, riot or civil common, or colliding with a bird or animal.
N	New	This code is used as an indication that the coverage is for vehicles of current model year, or 1∼2 years prior to the current model year.
O	Old	This code is used as an indication that the coverage is for vehicles three or more years prior to the current model year.

From Table 5.3, there are collision and comprehensive coverages, each for new and old vehicles of the entity. Hence, an entity can potentially have collision coverage for new vehicles (CN), collision coverage for old vehicles (CO), comprehensive coverage for new vehicles (PN), and comprehensive coverage for old vehicles (PO). Hence, in our analysis, we consider these sub-coverages as individual lines of businesses, and work with six separate lines, including building and contents (BC), and contractor's equipment (IM) as separate lines also.

Preliminary dependence measures for discrete claim frequencies and continuous average severities can be obtained using polychoric and polyserial correlations. These dependence measures both assume latent normal variables, whose values fall within the cut-points of the discrete variables. The polychoric correlation is the inferred latent correlation between two ordered categorical variables; the polyserial correlation is the inferred latent correlation between a continuous variable and an ordered categorical variable, cf. Joe (2014).

Table 5.4 shows the polychoric correlation among the frequencies of the six coverage groups. Note that these dependencies in Table 5.4 are measured before controlling for the effects of explanatory variables on the frequencies. As Table 5.4 shows, there is evidence of correlation across different lines, however these cross-sectional dependencies may be due to correlations in the exposure amounts or, in other words, the sizes of the entities.

Table 5.4: Polychoric Correlation among Frequencies of Claims.

	BC	**IM**	**PN**	**PO**	**CN**
IM	0.506				
PN	0.465	0.584			
PO	0.490	0.590	0.771		
CN	0.492	0.541	0.679	0.566	
CO	0.559	0.601	0.642	0.668	0.646

The dependence between frequencies and average claim severities is often of interest to modelers, to correctly understand the risk involved in the claims. In Section 2.4 we have reviewed methods to assess the dependency between the frequency and average severity of insurance claims. Our data are suitable for applying this approach. The diagonal entries of Table 5.5 show the polyserial correlations between the frequency and severity of each coverage group.

Table 5.5: Polyserial Correlation between Frequencies and Severities.

	BC Freq.	**IM Freq.**	**PN Freq.**	**PO Freq.**	**CN Freq.**	**CO Freq.**
BC Severity	−0.033	0.029	−0.063	−0.069	0.020	−0.050
IM Severity	−0.033	−0.078	0.110	0.249	0.159	0.225
PN Severity	0.074	0.275	−0.146	−0.216	0.119	0.143
PO Severity	0.111	0.171	−0.161	−0.119	0.258	0.137
CN Severity	−0.112	−0.174	−0.003	0.135	0.032	−0.175
CO Severity	−0.099	−0.079	−0.055	−0.083	−0.068	−0.032

According to Table 5.5, the observed correlation between frequency and severity is small. For the CN line, a positive correlation can be observed although very small (0.032, while the other correlations between frequency and severity are negative). Again, these numbers only provide a rough idea of the dependency. Table 5.6 shows the Spearman correlation between the average severities, for those observations with at least one positive claim. The correlation among the severities of new and old car comprehensive coverage is high.

Table 5.6: Correlation among Average Severities.

	BC	**IM**	**PN**	**PO**	**CN**
IM	0.220				
PN	0.098	0.095			
PO	0.229	0.118	0.415		
CN	0.084	0.237	0.166	0.200	
CO	0.132	0.261	0.075	0.140	0.244

In summary, these summary statistics show that there are potentially interesting dependencies among the response variables.

Explanatory Variables

Table 5.7 shows the number of observations available in the data set, for years 2006–2010.

Table 5.7: Number of Observations.

	BC	**IM**	**PN**	**PO**	**CN**	**CO**
Coverage > 0	5660	4622	1638	2138	1529	2013
Average Severity > 0	1684	236	315	263	370	362

Explanatory variables used are summarized in Table 5.8. The marginal analyses for each line are performed on the subset for which the coverage amounts shown in Table 5.7 are positive.

Table 5.8: Summary of Explanatory Variables.

Variable Name	**Description**	**Mean**
`lnCoverageBC`	Log of the BC coverage amount.	37.050
`lnCoverageIM`	Log of the IM coverage amount.	0.848
`lnCoveragePN`	Log of the PN coverage amount.	0.158
`lnCoveragePO`	Log of the PO coverage amount.	0.614
`lnCoverageCN`	Log of the CN coverage amount.	0.096
`lnCoverageCO`	Log of the CO coverage amount.	0.305
`NoClaimCreditBC`	Indicator for no BC claims in prior year.	0.328
`NoClaimCreditIM`	Indicator for no IM claims in prior year.	0.421
`NoClaimCreditPN`	Indicator for no PN claims in prior year.	0.110
`NoClaimCreditPO`	Indicator for no PO claims in prior year.	0.170
`NoClaimCreditCN`	Indicator for no CN claims in prior year.	0.090
`NoClaimCreditCO`	Indicator for no CO claims in prior year.	0.140
`EntityType`	City, County, Misc, School, Town (Categorical)	
`lnDeductBC`	Log of the BC deductible level, chosen by the entity.	7.137
`lnDeductIM`	Log of the IM deductible level, chosen by the entity.	5.340

5.1.2 Marginal Model Fitting—Zero/One Frequency, *GB2* Severity

For each coverage type, a frequency-severity model is fit marginally.

BC (Building and Contents) Frequency Modeling

In the frequency part, we fit several commonly employed count models: Poisson, negative binomial (NB), zero-inflated Poisson, zero-inflated negative binomial. Our data not only exhibit a large mass at 0, as with many other insurance claims data, but also an inflated number of 1 s. For BC, there are 997 policies with 1 claim. This can be compared to the expected number under zero-inflated Poisson, 754, and under the zero-inflated negative binomial, 791. (See Table 5.9 for details). These zero-inflated models underestimate the point mass at 1 due to the shrinkage to 0. Thus, alternative "zero-one-inflated" models are introduced in Section 2.2.2.

Table 5.9 shows the expected count for each frequency value under different models and the empirical values from the data. A Poisson distribution underestimates the zero proportions while zero-inflated and negative binomial models underestimate the proportion of 1 s. The zero-one inflated models do provide the best fits for simultaneously estimating the probability of a zero and a one.

Chi-square goodness of fit statistics can be used to compare different models. Table 5.10 shows the result. It is calculated depending on Table 5.9. The zero-one-inflated negative binomial is significantly better than other methods. Each column represents the result for the following cases:

(1) Empirical

(2) Zero-inflated Poisson

(3) Zero-one-inflated Poisson

(4) Poisson

(5) Negative binomial

(6) Zero-inflated Negative binomial

(7) Zero-one-inflated Negative binomial

Table 5.9: Comparison between Empirical Values and Expected Values for the building and contents (BC) Line.

Category	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
0	3976	4038.125	3975.403	3709.985	4075.368	4093.699	3996.906
1	997	754.384	1024.219	1012.267	809.077	791.424	1003.169
2	333	355.925	276.082	417.334	313.359	314.618	280.600
3	136	187.897	146.962	202.288	155.741	157.282	136.758
4	76	106.780	82.052	106.874	88.866	89.615	75.822
5	31	63.841	48.426	60.160	55.484	55.697	46.021
6	19	39.850	30.212	36.540	36.919	36.845	29.854
7	19	26.082	19.850	24.261	25.765	25.553	20.379
8	16	18.025	13.670	17.440	18.663	18.395	14.482
9	5	13.165	9.808	13.222	13.932	13.652	10.632
10	7	10.087	7.269	10.305	10.664	10.393	8.016
11	2	8.007	5.505	8.124	8.336	8.084	6.180
12	4	6.505	4.219	6.427	6.636	6.406	4.855
13	5	5.357	3.248	5.086	5.367	5.159	3.875
14	5	4.441	2.502	4.024	4.401	4.214	3.136
15	2	3.690	1.925	3.182	3.653	3.485	2.569
16	4	3.062	1.479	2.519	3.066	2.914	2.127
17	3	2.530	1.134	1.999	2.598	2.460	1.777
18	1	2.077	0.867	1.597	2.221	2.095	1.498
≥ 19	19	10.168	5.167	16.366	19.876	18.004	11.343
0 proportion	0.702	0.713	0.702	0.655	0.720	0.723	0.706
1 proportion	0.176	0.133	0.181	0.179	0.143	0.140	0.177

Table 5.10: Goodness of Fit Statistics for BC Line.

(2)	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
154.573	77.064	105.201	88.086	98.400	34.515

BC (Building and Contents) Severity Modeling

In the average severity part, the most commonly used distribution, gamma, is fit and compared with the *GB2* model. To do the goodness of fit test, the quantiles of normal Cox-Snell residuals are compared with normal quantiles.

Figure 5.1 shows the residual plot of severity fitted with gamma and *GB2*. Clearly, the gamma does not fit well especially in the tail part.

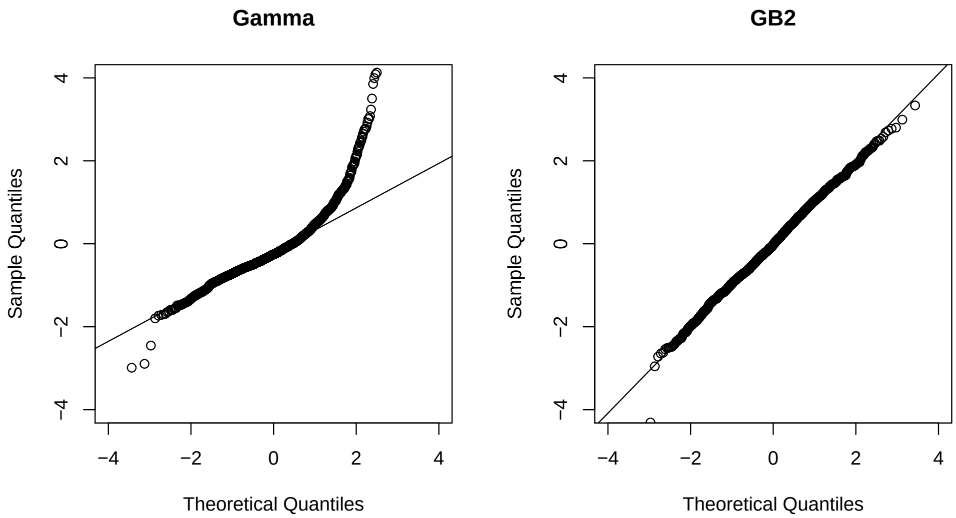


Figure 5.1: QQ Plot for Residuals of Gamma and *GB2* Distribution for BC.

Building and Contents Model Summary

Table 5.11 shows the coefficients for the fitted marginal models. Here, coefficients of *GB2*, NB and the zero-one-inflated parts are provided.

Table 5.11: Coefficients of Marginal Models for BC Line.

	Variable Name	***Coef.***	***Standard Error***	
GB2	(Intercept)	5.620	0.199	***
	`lnCoverageBC`	0.136	0.029	***
	`NoClaimCreditBC`	0.143	0.076	.
	`lnDeductBC`	0.321	0.034	***
	`EntityType`: City	−0.121	0.090	
	`EntityType`: County	−0.059	0.112	
	`EntityType`: Misc	0.052	0.142	
	`EntityType`: School	0.182	0.092	*
	`EntityType`: Town	−0.206	0.141	
	σ	0.343	0.070	
	α_1	0.486	0.119	
	α_2	0.349	0.083	
NB	(Intercept)	−0.798	0.198	***
	`lnCoverageBC`	0.853	0.033	***
	`NoClaimCreditBC`	−0.400	0.132	**
	`lnDeductBC`	−0.232	0.035	***
	`EntityType`: City	−0.074	0.090	
	`EntityType`: County	0.015	0.117	
	`EntityType`: Misc	−0.513	0.188	**
	`EntityType`: School	−1.056	0.094	***
	`EntityType`: Town	−0.016	0.160	
	log(size)	0.370	0.115	
Zero	(Intercept)	−6.928	0.840	***
	`CoverageBC`	−0.408	0.135	**
	`lnDeductBC`	0.880	0.108	***
	`NoClaimCreditBC`	0.954	0.459	*
One	(Intercept)	−5.466	0.965	***
	`CoverageBC`	0.142	0.117	
	`lnDeductBC`	0.323	0.137	*
	`NoClaimCreditBC`	0.669	0.447	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1.

Marginal Models for Other Lines

Section 5.1.4 provides the model selection and marginal model results for lines other than building and contents.

5.1.3 Copula Identification and Fitting

Dependence is fit at two levels. The first is between frequency and average severity within each line. The second is among different lines.

Frequency Severity Dependence

Vuong's test, as described in Section 2.3.7, is used for copula selection. Specifically, we consider two models $M^{(1)}$ and $M^{(2)}$, in our example, $M^{(1)}$ is Gaussian copula while $M^{(2)}$ is t copula. Let Δ_{12} be the difference in divergence from models $M^{(1)}$ and $M^{(2)}$. When the true density is h, this can be written as

$$\Delta_{12} = n^{-1} \sum_{i} \left\{ E_h[logf^{(2)}(Y_i; x_i, \theta^{(2)})] - E_h[logf^{(1)}(Y_i; x_i, \theta^{(1)})] \right\}.$$

A large sample 95% confidence interval for Δ_{12}, $\bar{D} \pm 1.96 \times n^{-1/2} SD_D$, is provided in Table 5.12. Table 5.12 shows the comparison of Gaussian copula against t copula with commonly used degrees of freedom for frequency and severity dependence in BC line. An interval completely below 0 indicates that copula 1 is significantly better than copula 2. Thus, the Gaussian copula is preferred.

Table 5.12: Vuong Test of Copulas for BC Frequency and Severity Dependence.

Copula 1	**Copula 2**	**95% Interval**	
Gaussian	$t(\text{df} = 3)$	-0.0307	-0.0122
Gaussian	$t(\text{df} = 4)$	-0.0202	-0.0065
Gaussian	$t(\text{df} = 5)$	-0.0147	-0.0038
Gaussian	$t(\text{df} = 6)$	-0.0114	-0.0023

Maximum likelihood estimation with the full multivariate likelihood, which estimates parameters in marginal and copula models simultaneously, is fit here. Table 5.13 shows parameters of BC line with the full likelihood method. Here the marginal dispersion parameters are fixed from marginal models. By comparing Tables 5.13 and 5.11, it can be seen that the coefficients are close. As pointed out in Joe (2014), inference functions for margins, with the results in Table 5.11, is efficient and can provide a good starting point for the full likelihood method, as in Table 5.13.

Table 5.13: Coefficients of Total Likelihood for BC Line.

	Variable Name	***Coef.***	***Standard Error***	
GB2	(Intercept)	5.629	0.195	***
	`lnCoverageBC`	0.144	0.029	***
	`NoClaimCreditBC`	0.222	0.076	**
	`lnDeductBC`	0.320	0.031	***
	`EntityType`: City	−0.148	0.090	.
	`EntityType`: County	−0.043	0.111	
	`EntityType`: Misc	0.158	0.143	
	`EntityType`: School	0.225	0.092	*
	`EntityType`: Town	−0.218	0.141	
	σ	0.343	0.070	
	α_1	0.486	0.119	
	α_2	0.349	0.083	
NB	(Intercept)	−0.789	0.083	***
	`lnCoverageBC`	1.003	0.001	***
	`NoClaimCreditBC`	−0.297	0.172	.
	`lnDeductBC`	−0.230	0.001	***
	`EntityType`: City	−0.068	0.097	
	`EntityType`: County	−0.489	0.109	***
	`EntityType`: Misc	−0.468	0.202	*
	`EntityType`: School	−0.645	0.083	***
	`EntityType`: Town	0.267	0.166	
	log(Size)	0.370	0.115	

Table 5.13: *Continued*

	Variable Name	***Coef.***	***Standard Error***	
Zero	(Intercept)	−6.246	0.364	***
	`lnCoverageBC`	−0.338	0.047	***
	`lnDeductBC`	0.910	0.050	***
	`NoClaimCreditBC`	0.888	0.355	*
One	(Intercept)	−5.361	0.022	***
	`lnCoverageBC`	0.345	0.013	***
	`lnDeductBC`	0.335	0.010	***
	`NoClaimCreditBC`	0.556	0.431	
ρ	Dependence	−0.132	0.033	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1.

For other lines, the results of the full likelihood method are summarized in Table 5.14.

Table 5.14: Coefficients of Total Likelihood for Other Lines. As described in Section 5.1.4, the other lines use a negative binomial model for claim frequencies, not the 0–1 inflated model introduced in Section 2.2.2. For the CO line severity, $\frac{1}{\sigma}$ is fitted for the purpose of computation. Model selection and marginal model results can be found in Section 5.1.4.

		IM			PN			PO			CN			CO		
		Coef.	***Std. Error***		***Coef.***	***Std. Error***		***Coef.***	***Std. Error***		***Coef.***	***Std. Error***		***Coef.***	***Std. Error***	
GB2	(Intercept)	8.153	0.823	***	7.918	0.046	***	7.554	0.092	***	6.773	0.059	***	9.334	0.000	***
	`lnCoverage`	0.304	0.065	***	0.078	0.045	.	0.081	0.057		0.137	0.039	***	0.161	0.000	***
	`NoClaimCredit`	0.190	0.202		0.021	0.209		0.695	0.194	***	0.140	0.144		−0.296	0.001	***
	`lnDeduct`	0.028	0.125													
	σ	0.955	0.365		0.047	0.043		0.100	0.130		0.863	0.513		40.193	31.080	
	α_1	1.171	0.630		0.054	0.050		0.102	0.137		4.932	6.441		0.038	0.030	
	α_2	1.337	0.856		0.076	0.068		0.108	0.145		1.279	1.131		0.025	0.019	
NB	(Intercept)	−1.331	0.594	*	−2.160	0.284	***	−2.664	0.297	***	−0.467	0.158	**	−1.746	0.187	***
	`Coverage`	0.796	0.077	***	0.239	0.065	***	0.490	0.067	***	0.487	0.054	***	0.782	0.056	***
	`NoClaimCredit`	−0.371	0.141	**	−0.588	0.194	**	−0.612	0.177	***	−0.668	0.157	***	−0.324	0.139	*
	`lnDeduct`	−0.140	0.085	.												
	`EntityType`: City	−0.306	0.235		0.574	0.330	.	0.411	0.376		0.433	0.186	*	0.680	0.232	**
	`EntityType`: County	0.139	0.274		3.083	0.294	***	2.477	0.329	***	1.131	0.172	***	1.284	0.211	***
	`EntityType`: Misc	−2.195	1.024	*	−0.060	0.642		−0.508	0.709		−0.323	0.456		0.486	0.442	
	`EntityType`: School	−0.032	0.292		0.389	0.297		0.926	0.327	**	−0.192	0.185		1.350	0.208	***
	`EntityType`: Town	−0.405	0.277		−0.579	0.481		−1.022	0.650		−1.529	0.385	***	−0.450	0.355	
	size	0.724			1.004			0.766			1.420			1.302		
ρ	Dependence	−0.109	0.097		−0.154	0.064	*	−0.166	0.073	*	0.171	0.064	**	0.009	0.045	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1.

Table 5.13 shows significantly strong negative association between frequency and average severity for the building and contents (BC) line. In contrast, the results are mixed for other lines. Table 5.14 shows no significant relationships for the CO and IM lines, mild negative relationships for the PN and PO lines, and a strong positive relationship for the CN line. For the BC and CN lines, these results are consistent with the polyserial correlations in Table 5.5, calculated without covariates.

Dependence between Different Lines

The second level of dependence lies between different lines. In this section, the dependence model for frequencies, severities and aggregate loss with Tweedie margins, as in Section 2.3.4, are fit. Here, we use marginal results from the inference functions for margins method. In principle, full likelihood can be used. As mentioned previously in this section, in our case, the results of inference functions for margins are close to full likelihood estimation.

Tables 5.15 and 5.16 show the dependence parameters of copula models for frequencies and severities, respectively. A Gaussian copula is applied and the composite likelihood method is used for computation. Comparing Tables 5.4 and 5.15, it can be seen that frequency dependence parameters decrease substantially. This is due to controlling for the effects of explanatory variables. In contrast, comparing Tables 5.6 and 5.16, there appears to be little change in the dependence parameters. This may be due to the smaller impact that explanatory variables have on the severity modeling when compared to frequency modeling.

Table 5.15: Dependence Parameters for Frequency. PN and PO shows the highest amount of correlation, indicating a common hazard within the comprehensive coverages. New and old car collisions also shows a high correlation, presumably due to a common hazard within the motor vehicles.

	BC	**IM**	**PN**	**PO**	**CN**
IM	0.190				
PN	0.141	0.162			
PO	0.054	0.206	0.379		
CN	0.101	0.149	0.271	0.081	
CO	0.116	0.213	0.151	0.231	0.297

Table 5.16: Dependence Parameters for Severity.

	BC	**IM**	**PN**	**PO**	**CN**
IM	0.145				
PN	0.134	0.051			
PO	0.298	0.099	0.498		
CN	0.062	0.110	0.156	0.168	
CO	0.106	0.215	0.083	0.080	0.210

Table 5.17 shows the result of dependence parameters for different lines with Tweedie margins. The coefficients of marginal models are in Section 5.1.5.

Table 5.17: Dependence Parameters for Tweedies.

	BC	**IM**	**PN**	**PO**	**CN**
IM	0.210				
PN	0.279	0.367			
PO	0.358	0.412	0.559		
CN	0.265	0.266	0.553	0.328	
CO	0.417	0.359	0.496	0.562	0.573

5.1.4 Other Lines

IM (Contractor's Equipment)

The property fund uses IM as a symbol to denote contractor's equipment, and we follow this notation. Figure 5.2 shows the residual plot of severity fitted with gamma and *GB2* in the IM line. Based on the plot, the *GB2* is chosen.

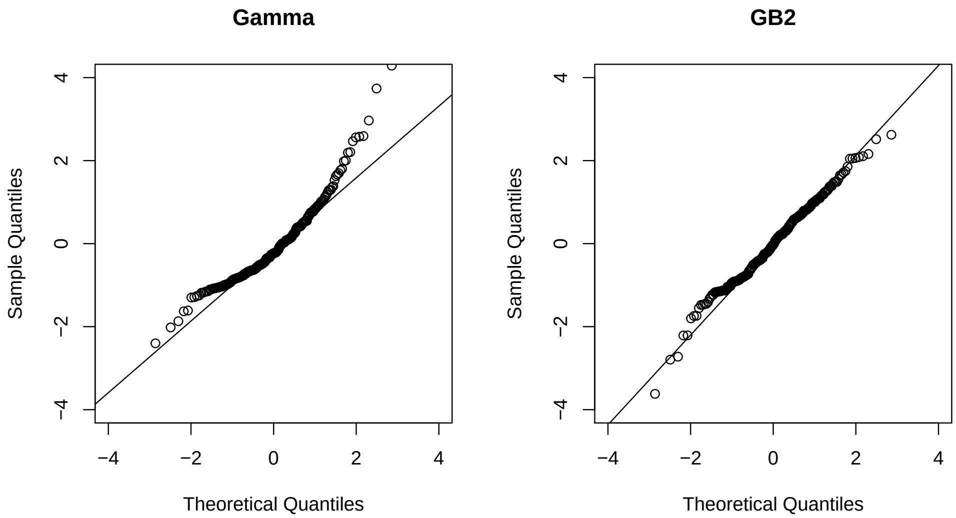


Figure 5.2: QQ Plot for Residuals of Gamma and *GB2* Distribution for contractor's equipment (IM)

Table 5.18 shows the expected count for each frequency value under different models and empirical values from the data. The proportion of 1 s for the IM line is not high, and hence most models are able to capture this.

Table 5.18: Comparison between Empirical Values and Expected Values for IM line.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
0 proportion	0.949	0.948	0.948	0.941	0.948	0.949	0.949
1 proportion	0.039	0.041	0.040	0.050	0.041	0.041	0.041

Table 5.19: Goodness of Fit Statistics for IM Line.

(2)	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
13.046	11.204	74.788	7.335	6.497	6.493

Table 5.19 shows goodness of fit tests result. It has been calculated using the results in Table 5.18. The parsimonious model, negative binomial, is preferred.

PN (Comprehensive New)

The property fund uses P for comprehensive, and N to denote new vehicles. Hence PN would mean comprehensive coverage for new vehicles. Figure 5.3 shows the residual plot of severity, fitted with gamma and *GB2* for PN. Based on the plot, the *GB2* is chosen for the PN line.

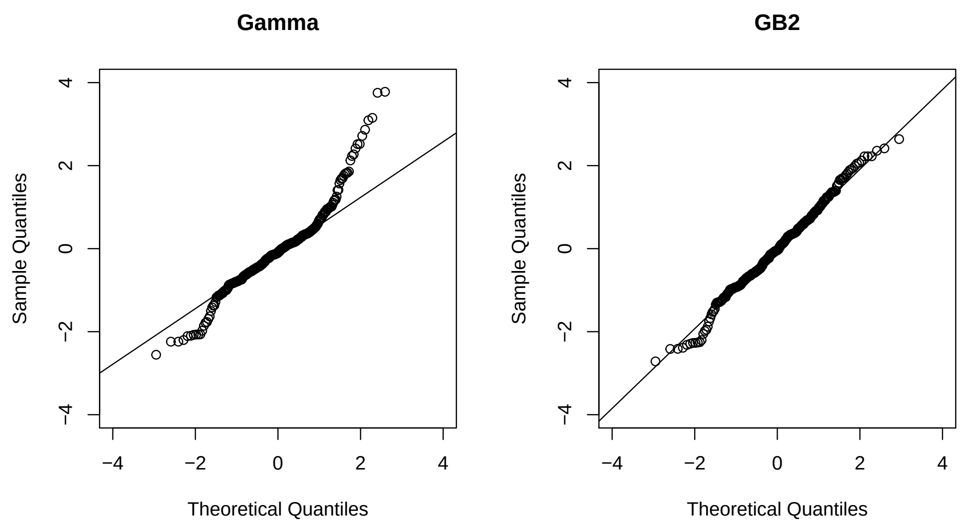


Figure 5.3: QQ Plot for Residuals of Gamma and *GB2* Distribution for comprehensive new (PN).

Table 5.20 shows the expected count for each frequency value under different models, and the empirical values from the data.

Table 5.20: Comparison between Empirical Values and Expected Values for PN Line.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
0 proportion	0.808	0.805	0.800	0.769	0.809	0.811	0.809
1 proportion	0.094	0.071	0.098	0.094	0.090	0.086	0.092

Table 5.21 shows goodness of fit tests result. It has been calculated using the results in Table 5.20. The simpler model, negative binomial, is preferred.

Table 5.21: Goodness of Fit Statistics for PN Line.

(2)	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
31,776.507	2113.085	93,179.199	11.609	14.537	11.853

PO (Comprehensive Old)

The property fund uses symbol O to denote old, hence PO would be comprehensive coverage for old vehicles. Figure 5.4 shows the residual plot of severity fitted with gamma and *GB2*, for the PO line.

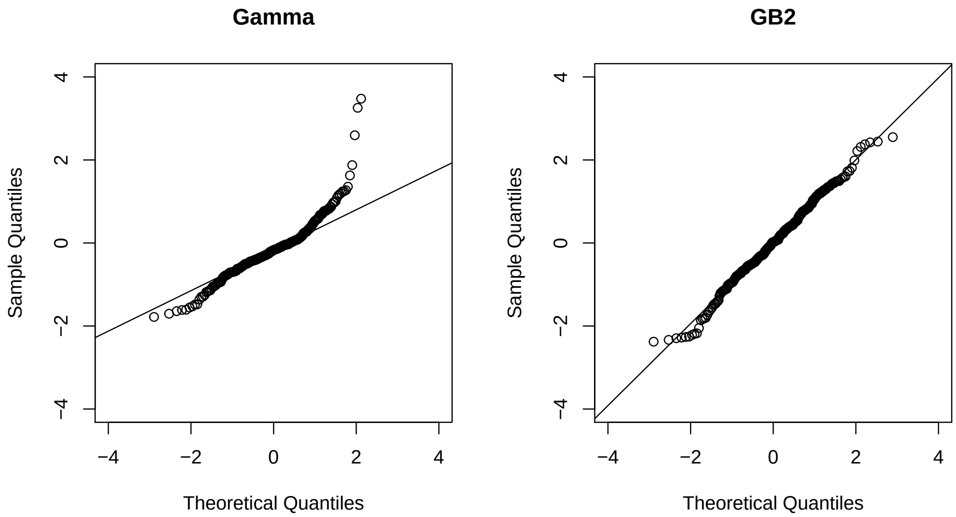


Figure 5.4: QQ Plot for Residuals of Gamma and *GB2* Distribution for comprehensive old (PO).

Table 5.22 shows the expected count for each frequency value under different models and empirical values from the data.

Table 5.22: Comparison between Empirical Values and Expected Values for PO Line.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
0 proportion	0.877	0.876	0.873	0.847	0.879	0.880	0.879
1 proportion	0.072	0.057	0.072	0.084	0.067	0.065	0.068

Table 5.23 shows goodness of fit tests result. It has been calculated using the results in Table 5.22. Negative binomial model is selected, based on the test results.

Table 5.23: Goodness of Fit Statistics for PO Line.

(2)	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
387.671	43.127	5365.758	2.995	3.824	2.512

CN (Collision New)

Figure 5.5 shows the residual plot of severity fitted with gamma and *GB2* for the CN line.

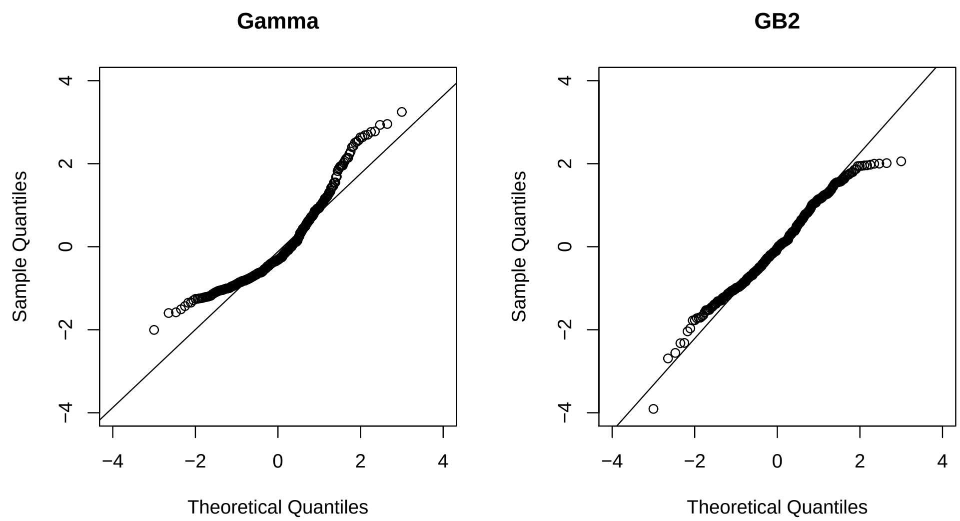


Figure 5.5: QQ Plot for Residuals of Gamma and *GB2* Distribution for collision new (CN).

Table 5.24 shows the expected count for each frequency value under different models, and the empirical values from the data.

Table 5.24: Comparison between Empirical Values and Expected Values for CN Line.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
0 proportion	0.758	0.757	0.754	0.713	0.764	0.765	0.761
1 proportion	0.149	0.132	0.150	0.179	0.138	0.137	0.148

Table 5.25 shows the goodness of fit tests result. It has been calculated using the values in Table 5.24. The parsimonious model, negative binomial, is selected.

Table 5.25: Goodness of Fit Statistics for CN Line.

(2)	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
10,932.035	1791.868	15,221.056	29.911	30.378	22.574

CO (Collision, Old)

Figure 5.6 shows the residual plot of severity fitted with Gamma and *GB2* for the CO line. *GB2* is preferred. Note, here $\frac{1}{\sigma}$ instead of σ is fitted for computational stability.

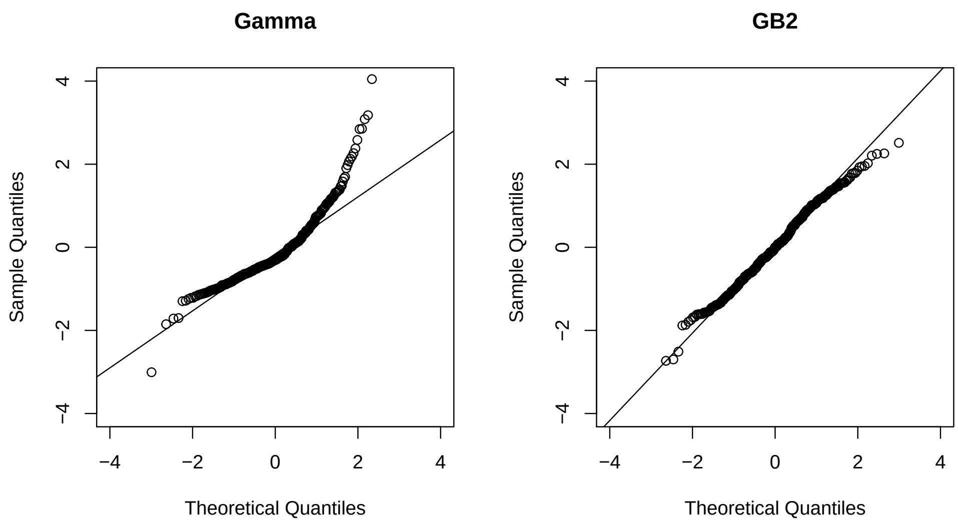


Figure 5.6: QQ Plot for Residuals of Gamma and *GB2* Distribution for collision old (CO).

Table 5.26 shows the expected count for each frequency value under different models, and the empirical values from the data.

Table 5.26: Comparison between Empirical Values and Expected Values for CO Line.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
0 proportion	0.820	0.819	0.818	0.795	0.822	0.823	0.822
1 proportion	0.111	0.098	0.109	0.130	0.109	0.105	0.109

Table 5.27 shows the goodness of fit tests result. It has been calculated using the results in Table 5.26. The parsimonious model, negative binomial, is selected.

Table 5.27: Goodness of Fit Statistics for CO Line.

(2)	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
60.691	25.206	121.987	10.387	11.440	10.370

5.1.5 Tweedie Margins

Table 5.28 shows marginal coefficients for each line.

Table 5.28: Marginal Coefficients of Tweedie Model.

	BC			IM			PN		
Variable Name	***Coef.***	***Std. Error***		***Coef.***	***Std. Error***		***Coef.***	***Std. Error***	
(Intercept)	5.855	0.969	***	8.404	1.081	***	6.284	0.437	***
`lnCoverage`	0.758	0.155	***	1.022	0.134	***	0.395	0.107	***
`lnDeduct`	0.147	0.148		−0.277	0.154	.			
`NoClaimCredit`	−0.272	0.371		−0.330	0.244		−0.570	0.296	.
`EntityType`: City	0.264	0.574		0.223	0.406		0.930	0.497	.
`EntityType`: County	0.204	0.719		0.671	0.501		2.550	0.462	***
`EntityType`: Misc	−0.380	0.729		−1.945	1.098	.	−0.010	0.942	
`EntityType`: School	0.072	0.521		−0.340	0.520		0.036	0.474	
`EntityType`: Town	0.940	0.658		−0.487	0.476		0.185	0.586	
ϕ	165.814			849.530			376.190		
P	1.669			1.461			1.418		
	PO			**CN**			**CO**		
Variable Name	***Coef.***	***Std. Error***		***Coef.***	***Std. Error***		***Coef.***	***Std. Error***	
(Intercept)	5.868	0.489	***	8.263	0.294	***	7.889	0.340	***
`lnCoverage`	0.860	0.119	***	0.474	0.098	***	0.841	0.117	***
`lnDeduct`									
`NoClaimCredit`	0.155	0.319		−0.369	0.253		−1.025	0.331	**
`EntityType`: City	0.747	0.612		0.169	0.347		−0.723	0.540	
`EntityType`: County	1.414	0.577	*	1.112	0.325	***	0.863	0.434	*
`EntityType`: Misc	0.033	0.925		−0.596	0.744		−0.579	0.939	
`EntityType`: School	0.989	0.544	.	−0.631	0.316	*	0.477	0.399	
`EntityType`: Town	−2.482	1.123	*	−1.537	0.499	**	−0.628	0.564	
ϕ	322.662			336.297			302.556		
P	1.508			1.467			1.527		

Notes: ϕ: dispersion parameter, P: power parameter, $1 < P < 2$.
Signif. codes: 0 '***' 0.001'**' 0.01 '*' 0.05 '.' 0.1

Figure 5.7 shows the cdf plot of jittered aggregate losses, as described in Section 2.3.7. For most lines, the plots do not show a uniform trend. This tells us that the Tweedie model may not be ideal for such cases.

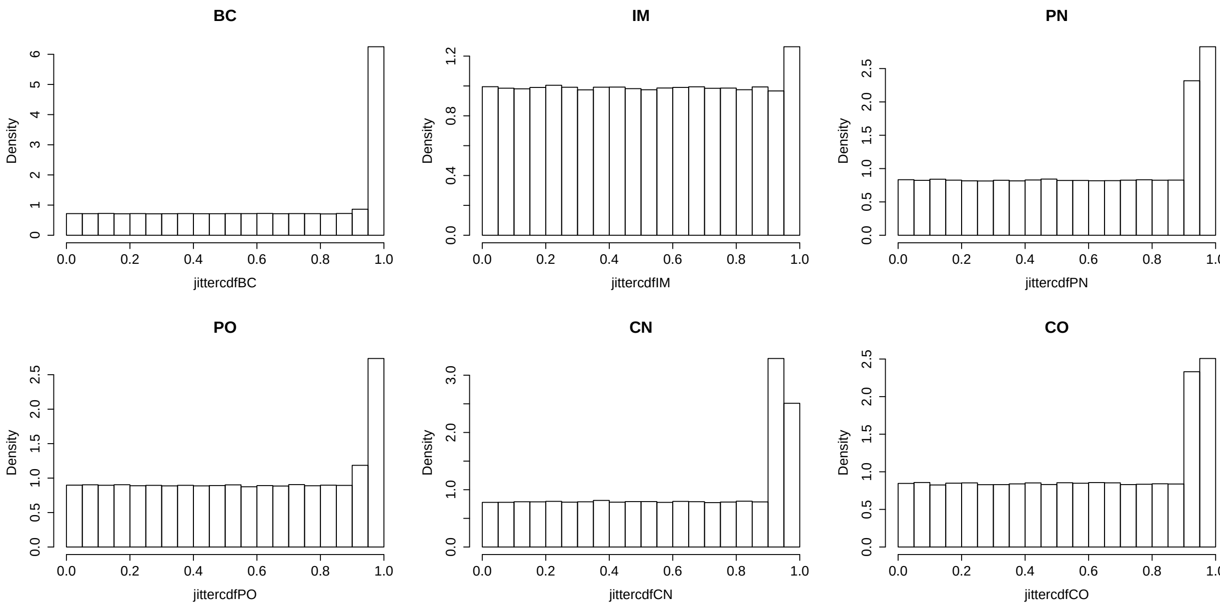


Figure 5.7: Jittering Plot of Tweedie.

5.1.6 Out-of-sample Validation

In this section, the independent case, dependent frequency-severity model, and the dependent pure premium approach are compared. The out of sample validation is performed on the 2011 held-out data, where there are 1098 observations. The claim scores for the pure premium approach are obtained using the conditional mean of the Tweedie distribution for each policyholder. For the frequency-severity approach, the conditional mean for the zero-one-inflated negative binomial distribution is multiplied to the first moment of the *GB2* severity distribution for the policyholder. Claim scores for the dependent pure premium approach and the dependent frequency-severity approach are computed using a Monte Carlo simulation of the normal copula.

We first consider the nonparametric Spearman correlation between model predictions and the held-out claims. Four models are considered: the frequency-severity and pure premium (Tweedie) model, assuming independence among lines, and assuming a Gaussian copula among lines. As can be seen from Table 5.29, the predicted claims are about the same whether dependence is considered or not. The interesting question is how much improvement the zero-one-inflated negative binomial model, and the long-tail distribution (*GB2*) marginals bring. We observe that the long-tail nature of the severity distribution sometimes results in a large predicted claim. We found that prediction of the mean, using the first moment, can be numerically sensitive.

Figure 5.8 shows a plot of the predicted claims against the out-of-sample claims, for the independent pure premium approach. Figure 5.9 shows the dependent frequency-severity approach. These figures illustrate that more sophisticated marginal models improve the prediction for the building and contents coverage group.

Table 5.29: Out-of-Sample Correlation.

	BC	**IM**	**PN**	**PO**	**CN**	**CO**	**Total**
Independent Tweedie	0.410	0.304	0.602	0.461	0.512	0.482	0.500
Dependent Tweedie (Monte Carlo)	0.412	0.305	0.601	0.462	0.511	0.481	0.501
Independent Freq-Sev	0.440	0.308	0.590	0.475	0.525	0.469	0.498
Dependent Freq-Sev (Monte Carlo)	0.435	0.308	0.590	0.477	0.525	0.485	0.521

Figure 5.8: Out-of-Sample Validation for Independent Tweedie. (In these plots, the conditional mean for each policyholder is plotted against the claims.)

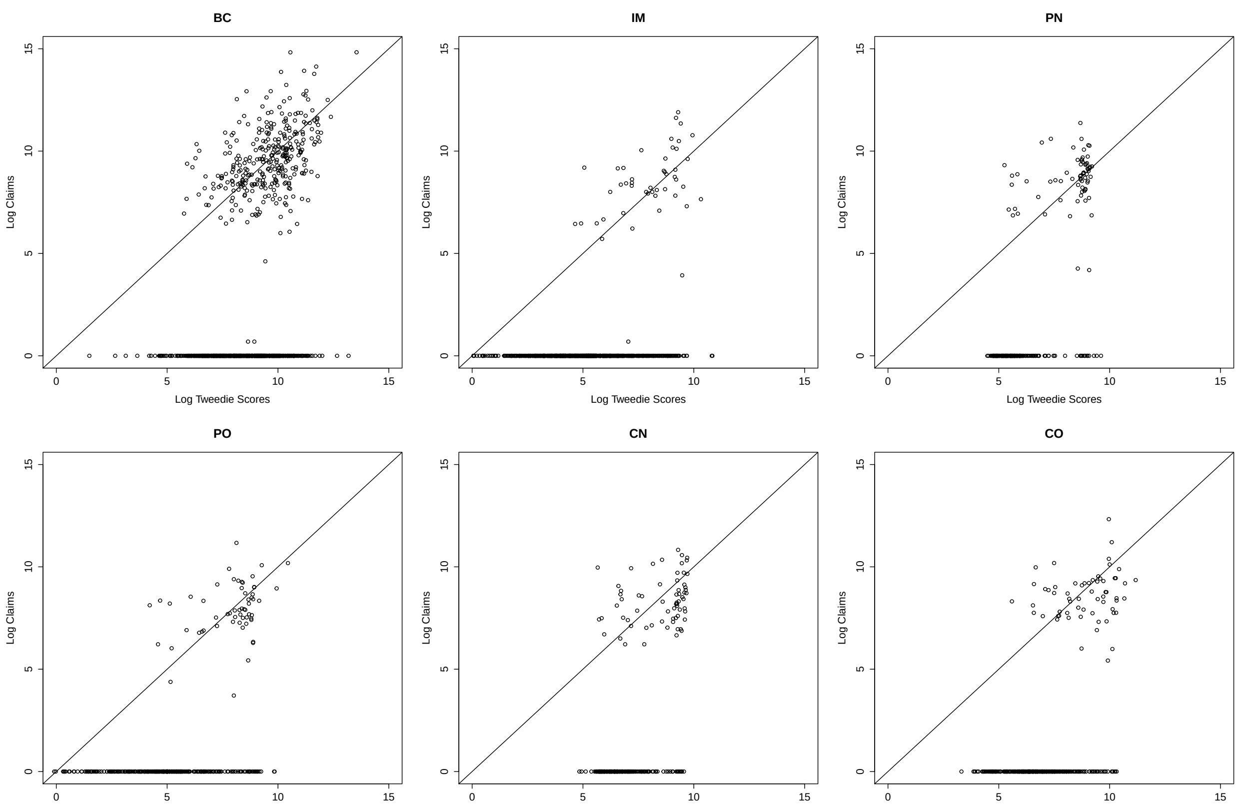
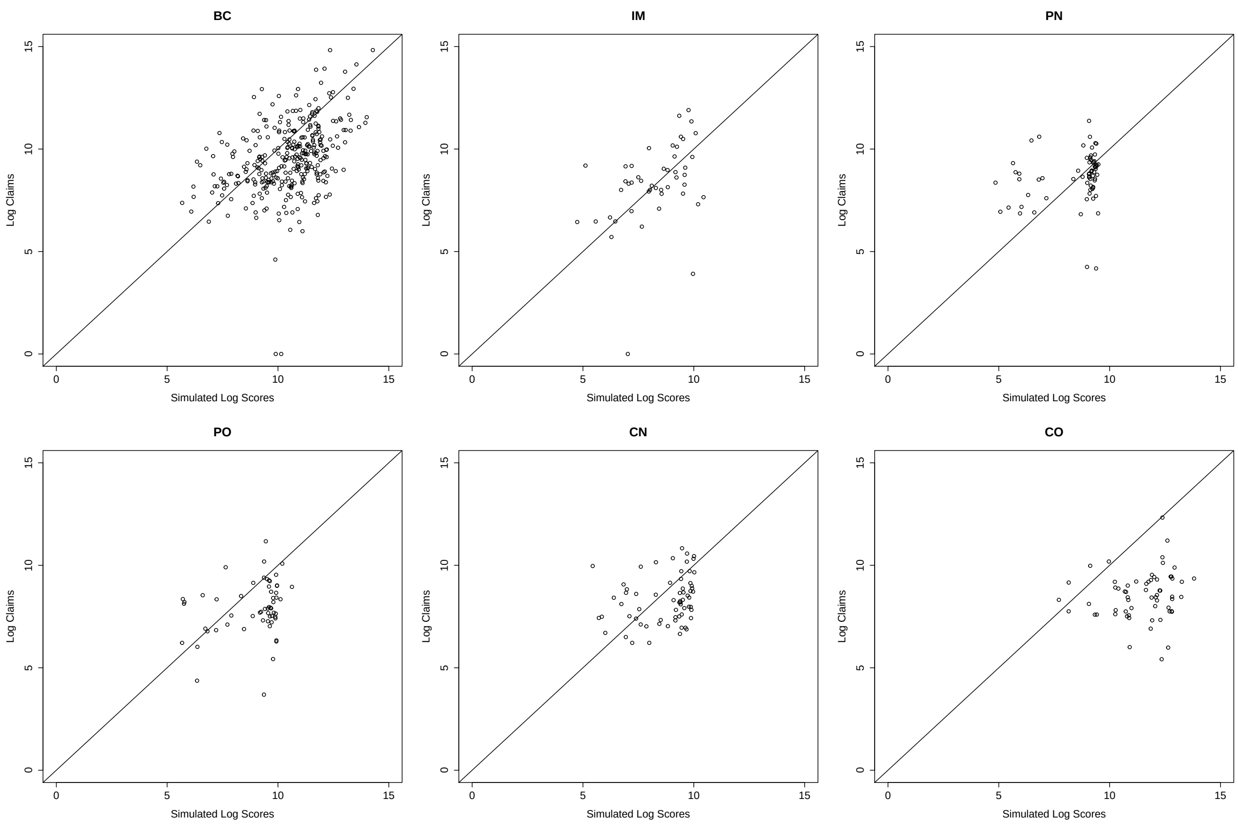


Figure 5.9: Out-of-Sample Validation for Dependent Frequency-Severity. (In these plots, the claim scores for each line is simulated from the frequency-severity model with dependence, using a Monte Carlo approach with B = 50,000 samples from the normal copula. The model with 01-NB and *GB2* marginals show clear improvement for the BC line, in particular for the upper tail prediction. For other lines such as CO, the *GB2* marginal results in miss-scaling).

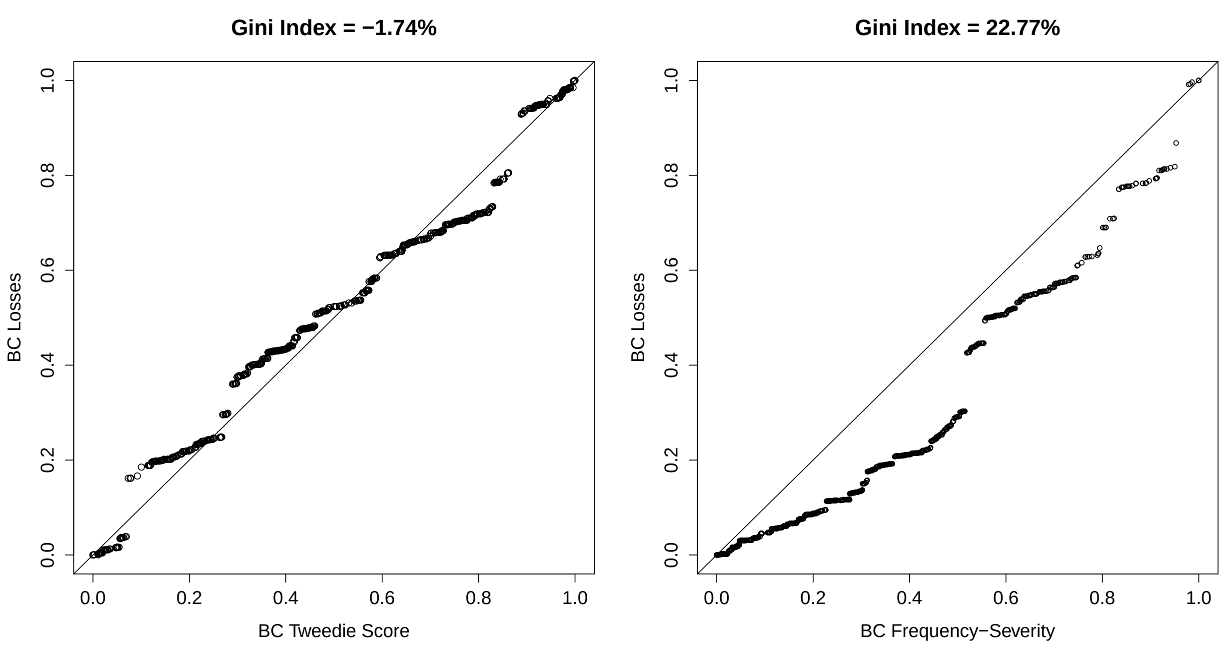


5.1.7 Gini Index

To further validate our results, we use the Gini index to measure the satisfaction of the fund manager with each score. The Gini index is introduced in Frees et al. (2011b). In this section, the Gini index is calculated using relativities computed with the actual premium collected by the LGPIF in 2011 as the denominator, with the scores predicted by each model as numerator. This means we are looking for improvements over the original premium scores used by the LGPIF. We expect the Gini index to be higher with the frequency-severity approach, as the fit for the upper tail is better. Figure 5.10 compares the independent Tweedie approach, and the dependent frequency-severity approach. For the dependent frequency-severity approach, a random 6-dimensional vector is sampled from the normal copula, and the quantiles are converted to frequencies and severities.

For the BC line scores calculated using the Tweedie model, we obtain −1.74% Gini index, meaning this model does not improve the existing premium scores used by the fund. Note that in Frees and Lee (2017), where the interest is more in the regularization problem, a constant premium is used as denominator for assessing the relativity. Here, the denominator used is the original premiums, which means in order for the index to be positive, there must be an improvement over the original premiums. The dependent frequency-severity scores with B = 50,000, normal copula, and zero-one-inflated NB and *GB2* margins results in a Gini index of 22.77%, meaning a clear improvement from the original premium scores. As a side note, the Spearman correlations are: original BC premiums 42.59%, Tweedie model 40.97%, and Frequency-severity model 43.52%, with the out-of-sample claims. Also the reader may observe from Table 5.29 that the improvement is mostly due to the better marginal model fit, instead of the dependence modeling.

Figure 5.10: Ordered Lorenz Curves for BC.



5.1.8 Summary of Insurance Claim Modeling

In Section 5.1, the models introduced in Chapter 2 have been applied to the LGPIF data. The six coverage groups of the LGPIF have been each modeled using the best fitting marginal models. The 01-inflated negative binomial model, and the *GB2* model have been utilized where appropriate.

Copula models have been used to capture the dependence between the frequency and severity, and the dependence among the frequency of different coverage groups, as well as the dependence among the average severity of different coverage groups.

The simulation section used the fitted models to predict the 2011 losses for each of the six coverage groups. According to the results, sophisticated marginal models, such as the 01-inflated negative binomial model, and the *GB2* severity model improves the prediction. The frequency-severity approach performed better according to the Gini index measure shown in Section 5.1.6. Overall, the results verify that dependence modeling is useful for capturing the worst-case behavior of a portfolio of losses, while the average behavior is more influenced by the marginal models.

The models used in Section 5.1 assumed the response variable for the severity part is the average loss severity over a year. The response variable is defined for each policyholder-year for the frequency response. The average severity is defined for the policyholder-years in which at least one claim occurs. In the following section (Section 5.2), model fitting results for claim level responses will be shown. Note that in Section 5.1, the underlying loss variable is assumed to be observed. This is often not the case in reality, and hence Section 5.2 will focus on the case where censoring and truncation is in place.

5.2 Deductible Ratemaking

Chapter 3 compared two different approaches to deductible ratemaking. The regression approach performs statistical estimation of models with the log deductible explanatory variable included as a covariate. The maximum likelihood approach uses the coverage modification formulas. For both approaches, censored and truncated response variables are used for the modeling. In this section, coefficient estimation results are shown for models, which assume censored and truncated response variables are observed.

Section 5.1 assumed the average severities are observed. In Section 5.2, the claims are assumed to be directly observed without an averaging process.

5.2.1 LGPIF Claim-Level Data

This section shows the details of the deductible rating procedure using the LGPIF data, focusing on the building and contents (BC) coverage group. Table 5.30 summarizes the number of building and contents (BC) policies in force.

Table 5.30: Summary Statistics of BC (Primary Coverage) Claims

Year	Average Deductible	Loss Frequency	Claim Frequency	Loss Total	Claim Total	Number of Policyholders
2006	3,048	0.735	0.525	20,313,812	18,161,172	1,158
2007	3,233	0.926	0.611	17,230,457	15,261,868	1,142
2008	3,412	0.747	0.518	11,060,356	9,160,440	1,129
2009	3,517	0.925	0.443	11,047,677	8,774,310	1,113
2010	3,599	1.089	0.633	36,659,296	33,328,603	1,113

The LGPIF data is ideal for this study, because the underlying losses and claims are both recorded in the data server. 4285 losses are observed in the training sample, resulting in 3089 claims exceeding the deductible. In many practical situations, the former may not be available. Hence, our goal in this study is to assume the former variable is unobserved and to test our result using the observed, empirical underlying losses, which various hypothetical deductible levels can be applied to and compared against. Table 5.31 shows a summary of the frequency and severity of claims by deductible choice, for 2006–2010. There are a number of instances with a high deductible level, say $100,000$, which implies that this data set may be studied in relation to risk retention problems in the reinsurance context.

Table 5.31: Summary Statistics of BC Claims by Deductible

Deductible	Avg. Loss Frequency	Avg. Claim Frequency	Average Loss	Average Claim	Number of Observations
500	0.628	0.621	6,197	5,884	2,674
1,000	0.668	0.641	6,808	6,147	1,067
2,500	0.539	0.506	12,923	11,610	686
5,000	0.606	0.362	39,229	36,987	716
10,000	0.378	0.196	13,044	10,692	209
15,000	0.672	0.224	22,426	17,615	67
25,000	5.973	0.202	34,679	21,654	183
50,000	17.290	1.355	530,867	411,317	31
75,000	7.400	0.000	44,897	0	5
100,000	0.294	0.235	486,350	459,880	17

Table 5.32 shows the explanatory variables in the policyholder data, given that an entity has purchased BC coverage. Table 5.34 shows a summary of both the underlying loss data, which usually isn't observable, and the claims data for those losses above the chosen deductible, which is observable in most common practices.

The observed claims are summarized in Table 5.34, allowing for a comparison of `LossBeforeDeductible` and `LossAfterDeductible` in both cases. Because Table 5.33 is conditional on `LossAfterDeductible` > 0, the minimum value for `LossAfterDeductible` is zero in Table 5.34 (losses), whereas it is positive in Table 5.33 (claims).

Table 5.32: Policyholder Data Summary

	Min.	*Median*	*Mean*	*Max.*	*N*
`CoverageBC`	8,937	11,310,000	37,190,000	2,445,000,000	5,655
Log(`CoverageBC`)	-4.718	2.426	2.128	7.802	
`DeductBC`	500	1,000	3,356	100,000	

Table 5.33: Summary of Claims Above the Deductible

	Min.	*Median*	*Mean*	*Max.*	*N*
`LossBeforeDeduct`	504	4094	29,920	12,920,000	3,089
After Deduct	4	2,982	27,420	12,920,000	

Table 5.34: Summary of All Losses

	Min.	*Median*	*Mean*	*Max.*	*N*
`LossBeforeDeduct`	1	2,243	19,300	12,920,000	4,285
After Deduct	0	750	16,970	12,920,000	

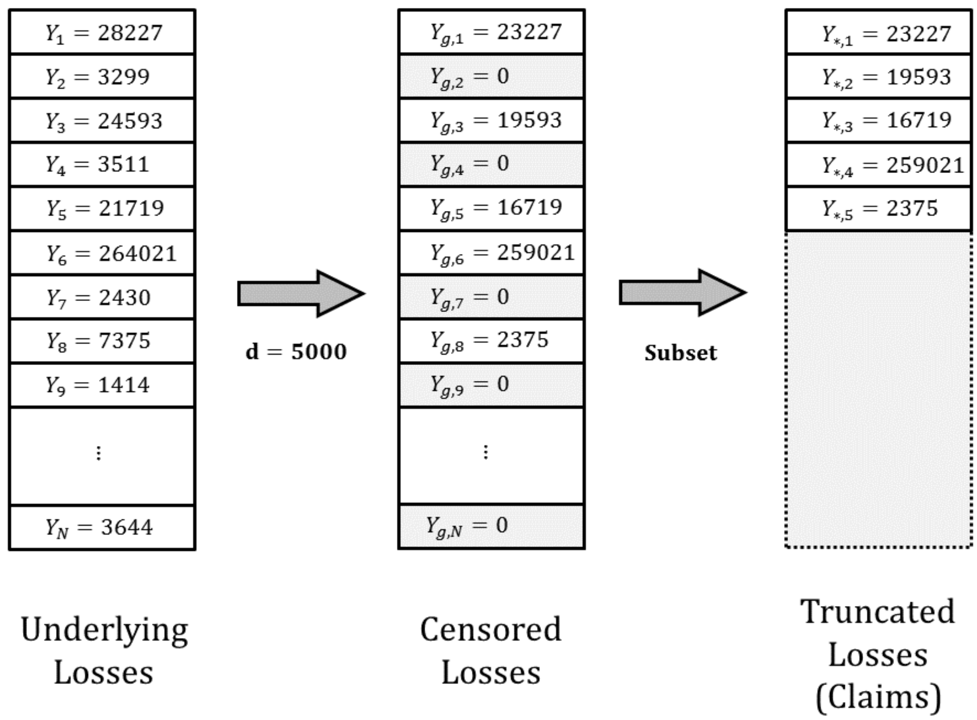


Figure 5.11: Illustration of Censoring and Truncation of Insurance Claims.

In Section 5.1, the average severities are used for the modeling, whereas here the claims are directly used without the averaging. Figure 5.11 illustrates the data generation mechanism for the censored losses and the claims.

5.2.2 Censored and Truncated Estimation

This section shows the statistical estimation methods needed for the deductible ratemaking technique shown in Chapter 3, Section 3.4.1. The general situation is that the jth observed claim for policyholder i is forced to be in the interval $(0, \infty)$, due to left truncation point d_i. We indicate those claim observations above the deductible by using $j(\varsigma)$, where the indices ς (varsigma) take on values $1, \ldots N_{g,i}(d)$. To specify the likelihood, we consider modeling the following observed variables:

$$
\begin{aligned}
N_{g,i}(d) &= \sum_{j=1}^{N_i} I(Y_{ij} > d_i), \\
Y_{*,i,j(\varsigma;i)}(d) &= \sum_{\varsigma=1}^{N_{g,i}} Y_{i,j(\varsigma)} - d_i | Y_{i,j(\varsigma;i)} > d_i, \\
j(\varsigma;i) &= \text{ index of } i\text{th loss above } d_i,\ \varsigma = 1, \ldots, N_{g,i}, \\
i(\iota) &= \text{ index of } j\text{th positive } Y_{*,i,j(\varsigma;i)}(d),
\end{aligned}
$$

Severity

Here, the likelihood for severities is specified. In most practical situations in actuarial science, upper-tail truncation rarely happens, and we are interested in ordinary left truncation only. Then, for a specific peril type $M = m$, the likelihood becomes

$$
\begin{aligned}
L_{Y|M} = &\prod_{M_\iota = m} \prod_{\varsigma=1}^{n_{g,i(\iota)}} \frac{\mathrm{f}_{Y|M}\left(y_{*,i(\iota),j(\varsigma;i)} + d_{i(\iota)}\right)}{1 - \mathrm{F}_{Y|M}\left(d_{i(\iota)}\right)} \cdot \mathrm{I}\left(y_{*,i(\iota),j(\varsigma;i)} < u_{i(\iota)} - d_{i(\iota)}\right) \\
&+ \prod_{M_\iota = m} \prod_{\varsigma=1}^{n_{g,i(\iota)}} \frac{1 - \mathrm{F}_{Y|M}\left(u_{i(\iota)}\right)}{1 - \mathrm{F}_{Y|M}\left(d_{i(\iota)}\right)} \cdot \mathrm{I}\left(y_{*,i(\iota),j(\varsigma;i)} = u_{i(\iota)} - d_{i(\iota)}\right),
\end{aligned}
$$

where y_* is used to denote a realization of $Y_*(d)$, and the second term will be nonzero if there is right-censoring due to a policy limit $u_{i(\iota)}$. This provides the likelihood of the conditional severity distribution. Coefficients have been estimated for the exponential, gamma and Pareto distributions, for each peril type separately. Results are shown in Table 5.35.

Table 5.35: Exponential, Gamma, Pareto Model Coefficient Estimates

		Exponential		**Gamma**		**Pareto**	
		Coef.	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
Fire	(Intercept)	9.484	0.246	9.524	0.385	11.628	1.029
	Coverage	0.455	0.060	0.447	0.096	0.571	0.093
	shape			0.423	0.041	1.012	0.012
	log L	1,723		1,674		1,626	
		Coef.	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
Vandalism	(Intercept)	7.713	0.112	7.782	0.001	8.806	0.305
	Coverage	0.116	0.025	0.016	0.001	-0.110	0.041
	shape			0.454	0.001	1.357	0.131
	log L	4,543		4,435		4,324	
		Coef.	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
Lightning	(Intercept)	8.028	0.105	8.032	0.101	8.419	0.177
	Coverage	0.324	0.028	0.325	0.027	0.226	0.040
	shape			1.079	0.051	1.874	0.201
	log L	7,346		7,345		7,213	
		Coef.	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
Wind	(Intercept)	8.829	0.139	8.860	0.171	9.258	0.607
	Coverage	0.250	0.035	0.235	0.044	0.257	0.066
	shape			0.715	0.058	1.242	0.177
	log L	2,645		2,635		2,552	
		Coef.	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
Hail	(Intercept)	9.658	0.181	9.672	0.239	10.481	0.550
	Coverage	0.595	0.051	0.593	0.068	0.276	0.091
	shape			0.584	0.085	1.543	0.436
	log L	857		849		827	
		Coef.	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
Vehicle	(Intercept)	7.917	0.117	7.921	0.075	8.123	0.149
	Coverage	0.049	0.029	0.087	0.018	-0.021	0.036
	shape			2.217	0.103	3.924	0.622
	log L	5,753		5,645		5,690	
		Coef.	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
Water (NW)	(Intercept)	8.058	0.201	8.072	0.246	10.167	1.460
	Coverage	0.400	0.043	0.386	0.053	0.137	0.088
	shape			0.707	0.070	1.127	0.204
	log L	1,705		1,698		1,653	
		Coef.	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
Water (W)	(Intercept)	9.493	0.180	9.516	0.300	12.953	0.438
	Coverage	0.428	0.042	0.428	0.071	0.270	0.059
	shape			0.375	0.023	1.007	0.002
	log L	4,310		4,141		3,910	
		Coef.	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
Misc.	(Intercept)	8.959	0.148	9.023	0.230	10.561	1.662
	Coverage	0.351	0.036	0.333	0.057	0.174	0.069
	shape			0.436	0.031	1.063	0.112
	log L	3,075		2,990		2,817	

The fit of these models can be assessed by looking at the Q-Q plots. From Figures 5.12, 5.13, 5.14, 5.15, 5.16, 5.17, 5.18, 5.18, 5.19, 5.20, the reader may see that the Pareto model fits best for most of the peril types, demonstrating the long-tail nature of the claim severities.

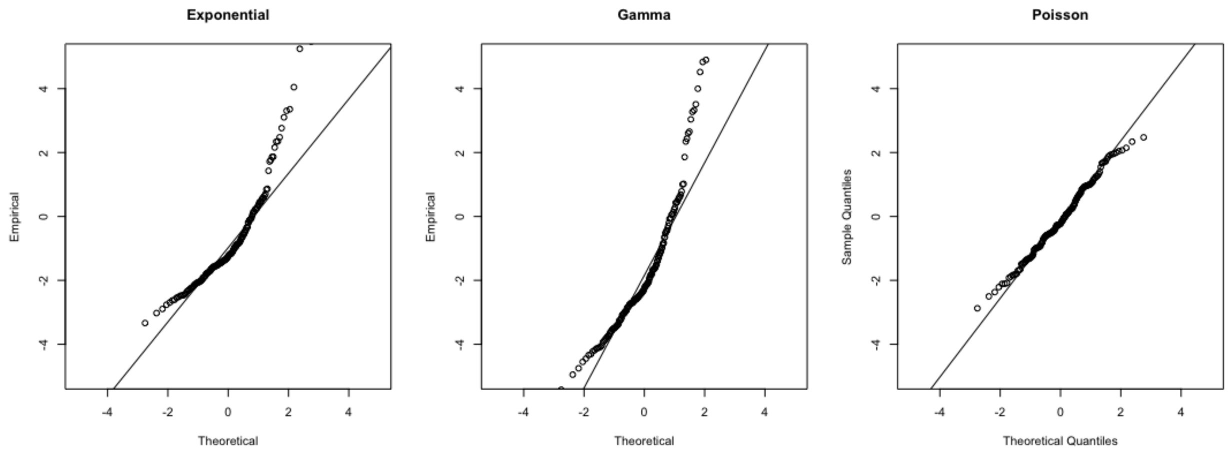


Figure 5.12: Fire Claim Model Q-Q Plots (Kolmogorov-Smirnov test statistics (p-values): 0.442(0.000), 0.594(0.000), 0.094(0.097)

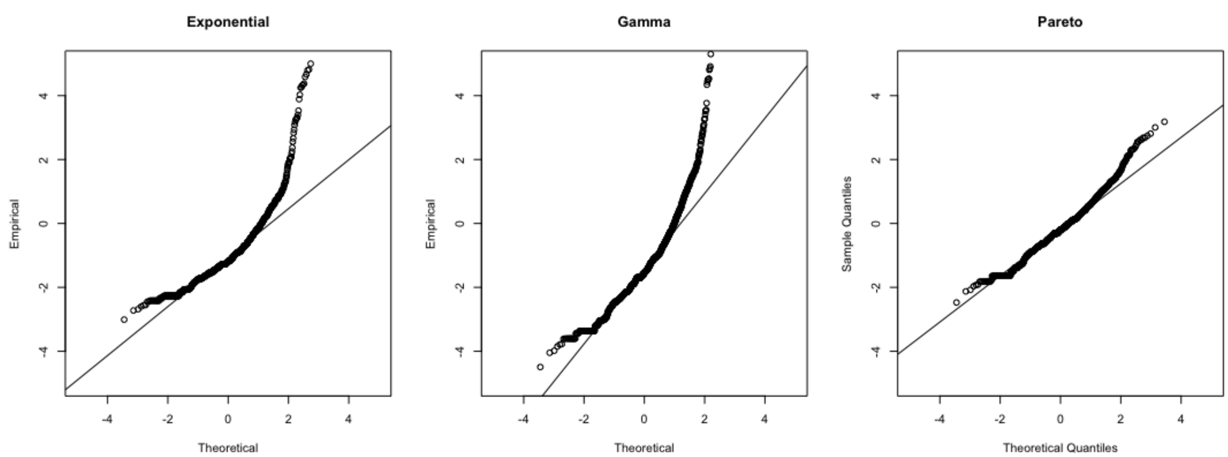


Figure 5.13: Vandalism Claim Model Q-Q Plots (Kolmogorov-Smirnov test statistics (p-values): 0.472(0.000), 0.520(0.000), 0.143(0.000)

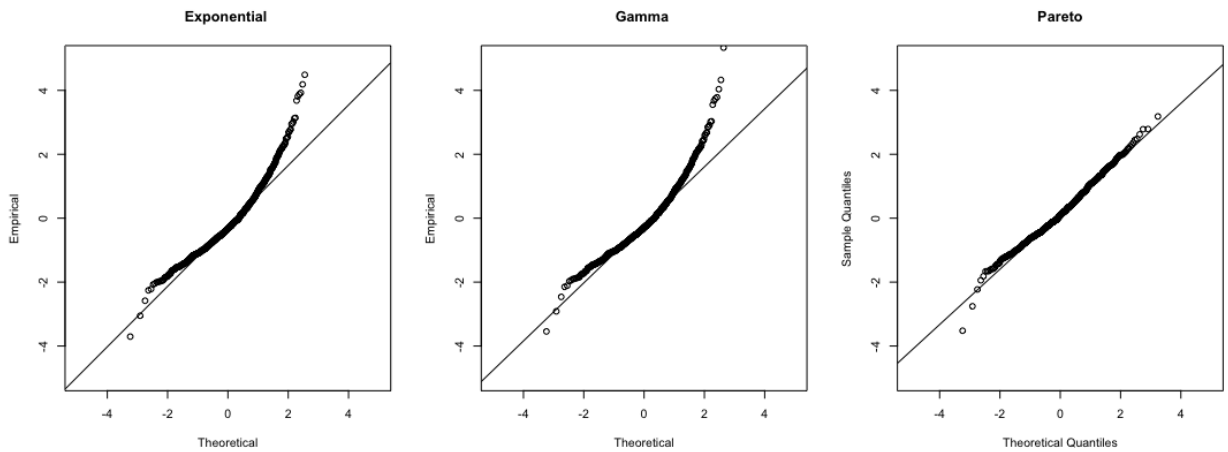


Figure 5.14: Lightning Claim Model Q-Q Plots (Kolmogorov-Smirnov test statistics (p-values): 0.148(0.000), 0.138(0.000), 0.105(0.000)

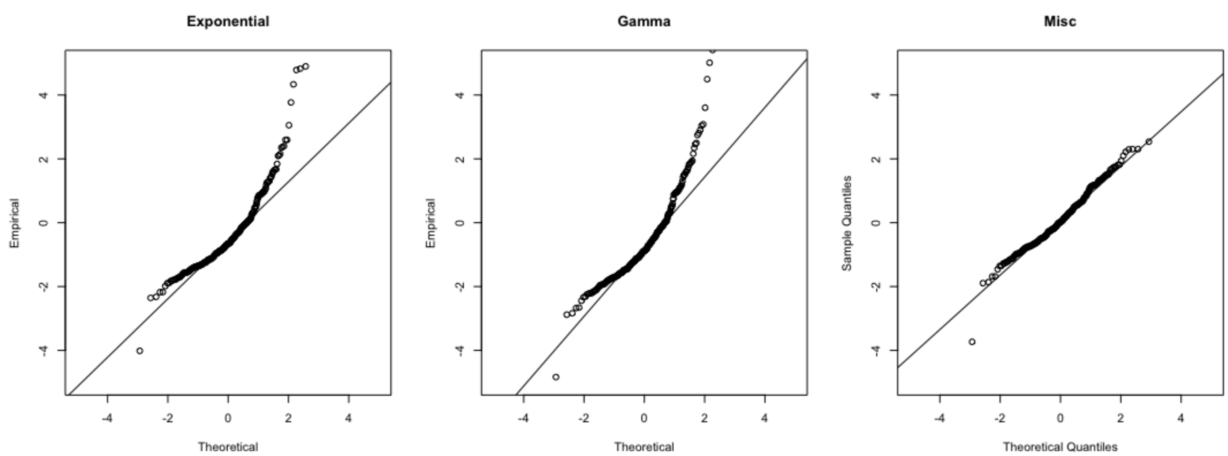


Figure 5.15: Wind Claim Model Q-Q Plots (Kolmogorov-Smirnov test statistics (p-values): 0.262(0.000), 0.328(0.000), 0.095(0.010)

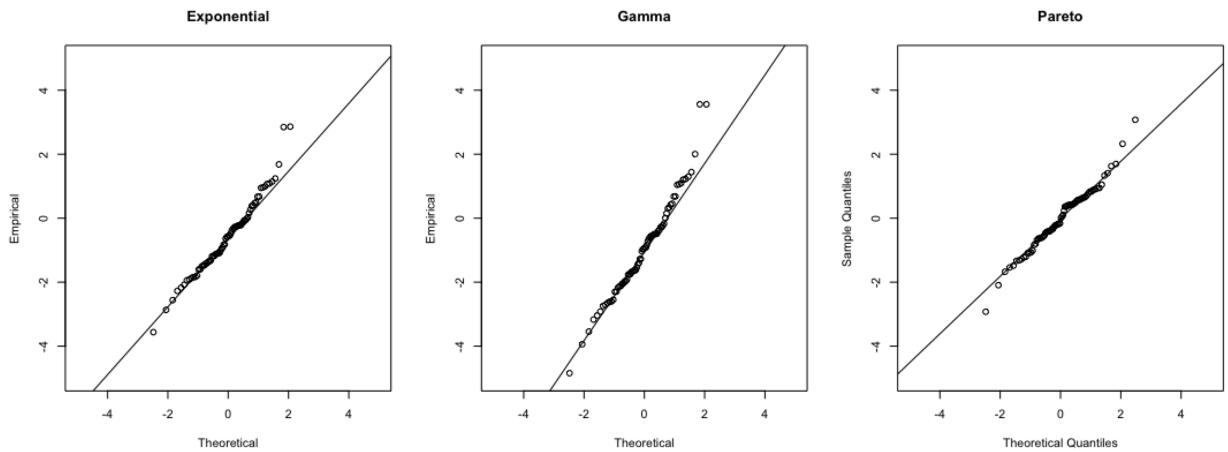


Figure 5.16: Hail Claim Model Q-Q Plots (Kolmogorov-Smirnov test statistics (p-values): 0.257(0.000), 0.359(0.000), 0.088(0.565)

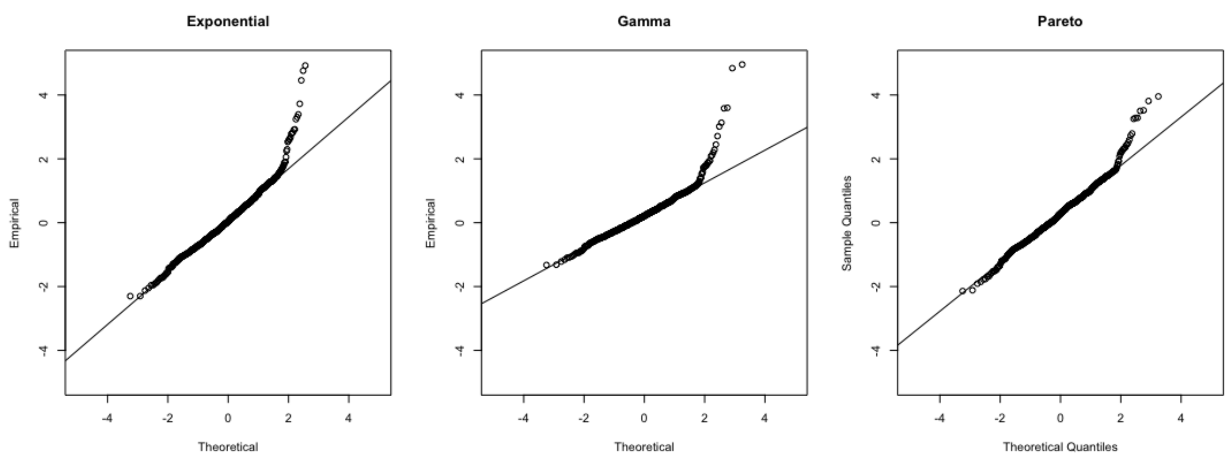


Figure 5.17: Vehicle Claim Model Q-Q Plots (Kolmogorov-Smirnov test statistics (p-values): 0.084(0.000), 0.246(0.000), 0.159(0.000)

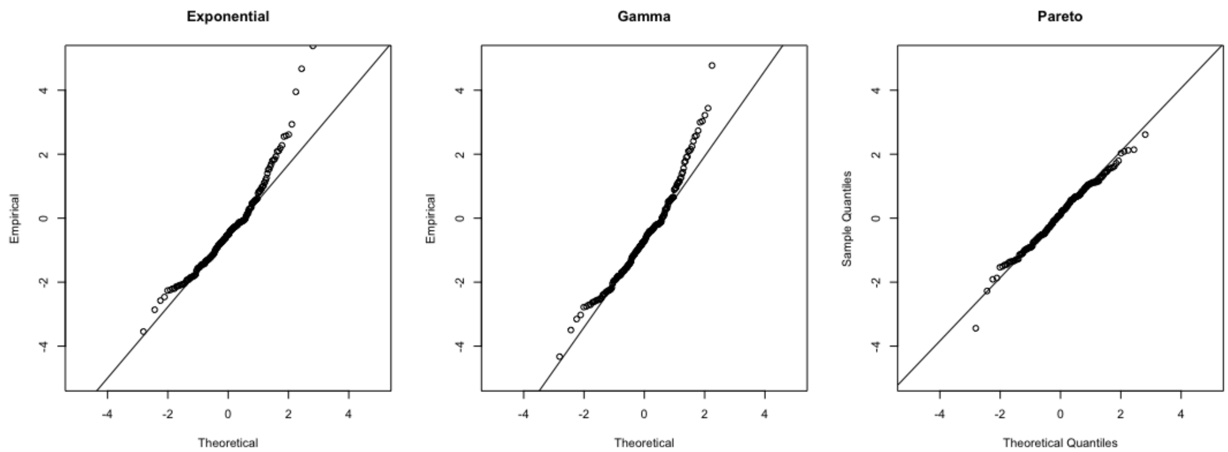


Figure 5.18: Water (Non-Weather) Claim Model Q-Q Plots (Kolmogorov-Smirnov test statistics (p-values): 0.230(0.000), 0.267(0.000), 0.083(0.128)

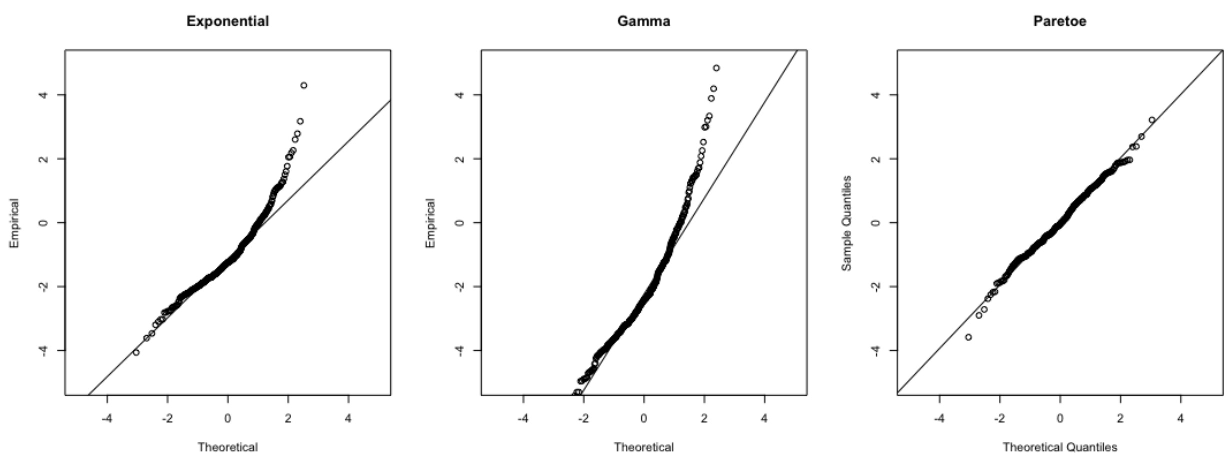


Figure 5.19: Water (Weather) Claim Model Q-Q Plots (Kolmogorov-Smirnov test statistics (p-values): 0.120(0.000), 0.648(0.000), 0.044(0.377)

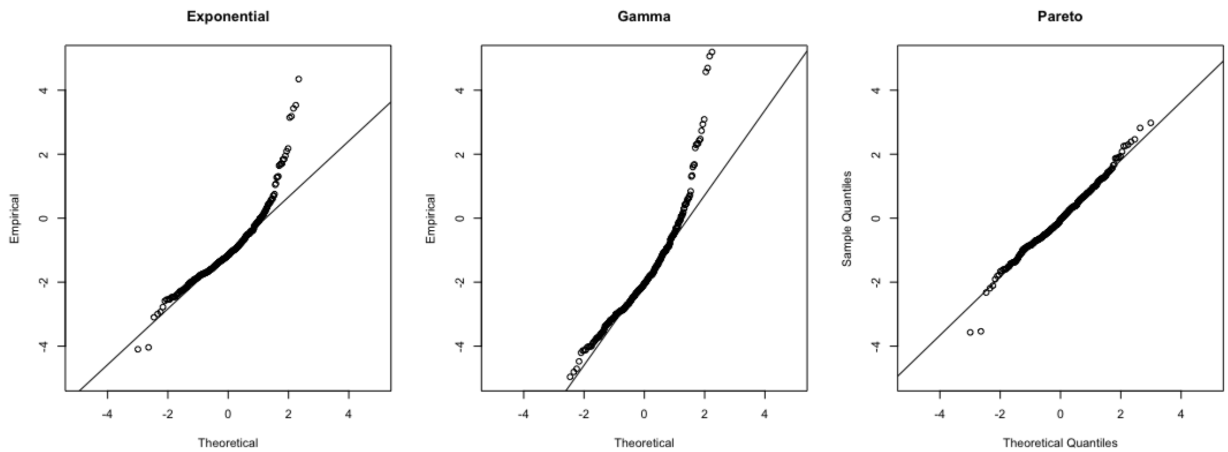


Figure 5.20: Miscellaneous Claim Model Q-Q Plots (Kolmogorov-Smirnov test statistics (p-values): 0.453(0.000), 0.610(0.000), 0.059(0.160)

Frequency

Obtaining the underlying frequency parameters considers the coverage modification by the deductible. For regression models with a log link, the quantity $\ln(1-\mathrm{F}_Y(d_i))$ can be used as an offset in standard regression routines. Specifically, for a mean parametrization

$$E[N_i] = \exp\left(\boldsymbol{x}_i'\boldsymbol{\gamma}\right),$$

consider the observed frequencies, $E[N_{g,i}]$. Then,

$$E[N_{g,i}] = E[N_i](1-\mathrm{F}_Y(d_i)) = \exp\left(\boldsymbol{x}_i'\boldsymbol{\gamma}\right)(1-\mathrm{F}_Y(d_i)) = \exp\left(\boldsymbol{x}_i'\boldsymbol{\gamma} + \texttt{offset}\right).$$

Hence, for most frequency regression models, the offset variable $\texttt{offset} = \ln(1-\mathrm{F}_Y(d_i))$ can be used, in order to recover the underlying loss frequency distribution parameters. For $(a, b, 0)$ class distributions with log link, this approach would work. In particular, the approach would work for the Poisson regression or the negative binomial regression. This can be potentially extended to a general link function η, where the mean is parametrized by

$$E[N_i] = \eta^{-1}\left(\boldsymbol{x}_i'\boldsymbol{\gamma}\right).$$

Estimation becomes more complicated for zero-inflated models, or zero-one-inflated models. For a treatment of zero-one-inflated models, see Section 5.1. In these models, the modification to each component of the primary and secondary probability mass function should be specified in the likelihood function. The estimation of zero-one-inflated Poisson models under censoring and truncation is covered in Section 5.2.3. In this chapter of the dissertation, we will use the Poisson frequency model for most illustrations. The coefficient estimates for the Poisson model are shown, using different coverage modification assumptions from the severity part. In general, the log likelihood tends to improve when the long-tail, Pareto model is assumed.

Table 5.36: Poisson Frequency Model Coefficient Estimates

		Fire		**Vandalism**		**Lightning**	
		Coef.	*Std.Err.*	*Coef.*	*Std.Err.*	*Coef.*	*Std.Err.*
Exponential	(Intercept)	-5.366	0.225	-5.082	0.156	-3.400	0.093
	Coverage	0.551	0.055	1.007	0.036	0.484	0.024
	-log L	604		1,466		2,091	
Gamma	(Intercept)	-5.374	0.225	-5.374	0.155	-3.242	0.096
	Coverage	0.556	0.055	1.099	0.036	0.518	0.025
	-log L	603		1,590		1,985	
Pareto	(Intercept)	-5.124	0.234	-5.118	0.154	-3.245	0.097
	Coverage	0.557	0.057	1.112	0.035	0.560	0.025
	-log L	594		1,396		1,949	

		Wind		**Hail**		**Vehicle**	
		Coef.	*Std.Err.*	*Coef.*	*Std.Err.*	*Coef.*	*Std.Err.*
Exponential	(Intercept)	-4.389	0.161	-4.681	0.210	-4.155	0.122
	Coverage	0.503	0.042	0.137	0.065	0.847	0.030
	-log L	947		375		1,799	
Gamma	(Intercept)	-4.454	0.166	-4.719	0.209	-3.839	0.121
	Coverage	0.599	0.044	0.141	0.065	0.799	0.030
	-log L	933		376		1,837	
Pareto	(Intercept)	-4.250	0.163	-4.693	0.212	-4.179	0.122
	Coverage	0.525	0.043	0.163	0.066	0.883	0.030
	-log L	937		376		1,786	

		Water (NW)		**Water (W)**		**Misc.**	
		Coef.	*Std.Err.*	*Coef.*	*Std.Err.*	*Coef.*	*Std.Err.*
Exponential	(Intercept)	-6.242	0.262	-4.910	0.157	-4.546	0.161
	Coverage	0.867	0.058	0.685	0.036	0.550	0.040
	-log L	601		1,155		998	
Gamma	(Intercept)	-6.652	0.282	-5.245	0.174	-4.710	0.172
	Coverage	1.043	0.063	0.899	0.041	0.697	0.043
	-log L	594		1,142		965	
Pareto	(Intercept)	-6.491	0.277	-4.973	0.166	-4.543	0.169
	Coverage	0.990	0.061	0.776	0.038	0.647	0.042
	-log L	593		1,134		963	

Peril Type Categories

Table 5.37 summarizes the losses and claims with respect to the peril type categories. The building and contents coverage has subcoverages, each of which consists of different peril types, which could be clustered into different categories. The property fund classifies the claims into three categories

Table 5.37: Peril Types of BC Losses

Peril	Average Loss	N	Prob.
Fire	87,168	172	0.034
Vandalism, Theft, Etc.	2,084	1,774	0.355
Lightning	11,087	832	0.167
Wind	18,125	296	0.059
Hail	145,488	76	0.015
Damage by Vehicle	3,905	852	0.171
Water (Weather)	80,432	426	0.085
Water (Non-Weather)	23,974	202	0.040
Misc.	29,150	362	0.073

by default. Here they have been manually recategorized into nine broad categories: fire, vandalism, lightning, wind, hail, vehicle, water damages (weather and non-weather) and other perils. Because the scale of the loss distribution is highly dependent on the peril type, it is worthwhile to consider specific peril type categories and fit loss distributions for each. Hence, the question is, Given a claim, how can we accurately classify it into one of these categories? For example, given a claim, could we classify it into either a vandalism claim or other? Analyses using basic discrete choice models have revealed that more categories result in higher standard errors in the coefficients of the peril type model. Rosenberg et al. (1999) and Yuan and Lin (2006) are some motivating articles for the classification problem. The claim classification should use explanatory variables of the policyholder, instead of any properties of the claim that is unknown at the instance of the claim. Discrete categories may be easier to formulate the classification problem. Classifying claims using continuous mixture models is left as future work.

When the log coverage is plotted with the average severities of claims, usually a positive correlation can be observed. However, when the underlying losses are plotted without the averaging over policy-year observations, the variation in the response variable is larger. Hence, when a single severity model is used with the coverage amount as an explanatory variable, interpretable coefficients may not be obtained without categorizing the claims into peril type categories. For this reason, we are interested in considering the different severity distributions with respect to various peril types. Section 5.2.2 have shown the coefficient estimation results for selected models using the peril type categories specified in Table 5.37. The density plot in Section 3.5.2, Figure 3.3 is for the lightning peril type shown here, and relativity curves in Figure 3.4 follows the nine categories shown here also.

Mixture Models

For modeling insurance losses with different profiles, depending on specific cases, mixture models have been used in the literature. In practice, exponential mixtures and mixed Pareto approaches have been used. For cases with mixture weights in multiple categories, practitioners have used the term *Pareto soup* model; see White (2005). One motivation for mixture models may be the different peril types, under which the loss severities experience different profiles. The modeling of severities, given a specific peril type, can be performed conditional on each peril type to obtain a set of parameters for each category of claim peril. This approach is taken in our aggregate claims prediction. Hence estimation issues for mixture models under censoring and truncation are discussed in detail here.

Suppose the number of claims N and Y are independent. Let M be the peril type categorical variable, so that there are several peril types with respective loss severities, conditional on the peril type category. Then the conditional density for claim severities can be written as

$$\mathrm{f}_{Y|M}(y|m) = \mathrm{f}_Y(y;\theta_m),$$

where we allow the distribution parameters θ_m to vary over different peril types. The unconditional distribution for the severity can be obtained using the mixture

$$\mathrm{F}_Y(y) = \sum_m \mathrm{F}_{Y|M}(y|m)\mathrm{f}_M(m).$$

In this case, the expected loss severity becomes

$$E[Y] = \sum_m E[Y|M=m]\mathrm{f}_M(m).$$

The underlying peril type probabilities, $\mathrm{f}_M(m)$, are required for this. In practice, fixed probabilities $\mathrm{f}_M(m) = p_m$ may be used. Here, in order to determine the peril type probabilities in a data-driven way, we specify the joint density for the frequencies and severities as

$$f_{Y,N}(y,n) = \sum_m f_{Y|M}(y|m) \cdot f_M(m) \cdot f_N(n),$$

for each peril type m. For related work, Frees and Valdez (2008) use a discrete choice model, focusing on the dependency among auto insurance claim types, and Fu and Liu (forthcoming) use the EM algorithm for estimation of a finite mixture model. Other approaches may be to impose parametric models for the peril type probabilities. The approach we take is to model each specific peril type and consider the coverage modification for that specific peril only.

Hence, calculation of the total building and contents density requires the peril type probabilities, since the above results are peril dependent. The peril type probabilities may be modeled either for all of the perils or for the reclassified peril type categories. For the Pareto mixture, the former approach was tried, while for the *GB2* mixture, the reclassified approach has been used with categories shown in Table 5.38. Other data-driven clusterings may be possible. In this case, the clustering has been manually performed into three arbitrarily chosen categories, using the average loss severities. The advantage of clustering the perils into categories is the reduction in the number of parameters to be estimated, especially when covariates are used for the peril type model.

Table 5.38: Average Loss and Claim Severity by Peril

Type	Perils	Average Loss	Average Claim	N
Low	Vandalism, theft, burglary, damage by vehicle	2,675	1,495	2,626
Medium	Lightning, wind, non-weather water damage, misc.	17,721	15,020	1,692
High	Fire, hail, water damage by weather	89,487	83,617	674
Total		19,496	17,167	4,992

In this chapter of the dissertation, a basic discrete choice model is used for the peril type probabilities, whose coefficients are estimated using maximum likelihood:

$$f_M(m) = \begin{cases} \dfrac{\exp\left(\boldsymbol{x}'\boldsymbol{\omega}_m\right)}{1 + \displaystyle\sum_{\varphi \neq m_0} \exp\left(\boldsymbol{x}'\boldsymbol{\omega}_\varphi\right)} & \text{for peril type } m \neq m_0 \\ \dfrac{1}{1 + \displaystyle\sum_{\varphi \neq m_0} \exp\left(\boldsymbol{x}'\boldsymbol{\omega}_\varphi\right)} & \text{for base peril type } m = m_0, \end{cases}$$

Here, $\boldsymbol{\omega}_m$ are the regression coefficients of interest. With deductible d, the observed peril type probabilities are altered. The underlying ω_m can be recovered by first calculating

$$1 - \mathrm{F}_{Y|M}(d|m), \qquad \text{for each } m,$$

which are simply the coverage modification amounts for each peril type. Using the truncated claims, given $Y > d$, the likelihood for peril type is specified using a multinomial model:

$$\begin{aligned} L_{M|Y>d} = \Pr(M = m|Y > d) &= \frac{\Pr(Y > d|M = m)\mathrm{f}_M(m)}{\sum_{\varphi} \Pr(Y > d|M = \varphi)\mathrm{f}_M(\varphi)} \\ &= \frac{\left(1 - \mathrm{F}_{Y|M}(d|m)\right)\mathrm{f}_M(m)}{\sum_{\varphi}\left(1 - \mathrm{F}_{Y|M}(d|\varphi)\right)\mathrm{f}_M(\varphi)}. \end{aligned} \tag{5.1}$$

The left side of (5.1) is an observed quantity from the truncated data. The unknowns are the parameters for the unconditional probabilities $\mathrm{f}_M(m)$ for each m and the conditional severity distribution parameters for each peril type. In a maximum likelihood context, these quantities would need to be included in the likelihood for the severity model. The goal is to estimate the $\boldsymbol{\omega}$ in $\mathrm{f}_M(m; \boldsymbol{\omega})$. For this, each observation of $M|Y > d$ within the truncated claims data is used in the following model for the categorical response M:

$$\begin{aligned} L_{Y,M|Y>d} &= \Pr(Y = y, M = m|Y > d) \\ &= \Pr(Y = y|M = m, Y > d) \cdot \Pr(M = m|Y > d) \\ &= L_{Y|M,Y>d} \cdot L_{M|Y>d}. \end{aligned} \tag{5.2}$$

Taking the log of (5.2) and summing, we have

$$\log L_{Y,M|Y>d} = \log L_{Y|M,Y>d} + \log L_{M|Y>d}, \tag{5.3}$$

where the first term is the log-likelihood for the severity for a specific peril type category, and the second term is the likelihood for the peril type. Note that the first term is the likelihood for the severity for a given observed claim, and the second is a probability weight, which is due to the modification by a deductible d.

Poisson-Gamma Regression (Model A)

This section details the models in the aggregate claims study in Chapter 3, Section 3.5.3. Two different regression approaches are compared with the gamma model with truncated estimation. We also compare a *GB2* mixture model with entity types used as predictors for peril type categories. We begin here by showing the two regression approaches. Table 5.39 shows the coefficient estimates for the first approach. The frequency model contains `lnDeductBC` as an explanatory variable. Here, the peril type model has used a simple classification scheme, where the peril types are categorized as having low, medium and high loss severity, before fitting the model. This was necessary because fitting too many peril type predictors resulted in high standard errors for the peril type model. Different approaches may be used for the reclassification of peril type categories.

Table 5.39: Poisson-Gamma Regression (Model A)

Gamma			Poisson		
Variable	*Coef.*	*Std. Err.*	*Variable*	*Coef.*	*Std. Err.*
(Intercept)	8.442	0.094	(Intercept)	-7.901	0.170
`CoverageBC`	0.353	0.026	`CoverageBC`	0.889	0.018
			`lnDeductBC`	-0.737	0.020
			`NoClaimCreditBC`	-0.409	0.060
`Type`:City	-0.095	0.105	`Type`:City	-0.068	0.063
`Type`:County	-0.093	0.123	`Type`:County	-0.323	0.075
`Type`:Misc	0.358	0.185	`Type`:Misc	-0.335	0.112
`Type`:School	0.764	0.111	`Type`:School	-0.745	0.065
`Type`:Town	0.115	0.194	`Type`:Town	0.021	0.116
ϕ	2.847				
AIC	64,444		*AIC*	9,054	

Table 5.40 shows the estimation results when $\ln E = \ln(u - d)$ is used as an offset variable in both the frequency and severity regressions. Because the deductible amounts and coverages are used as offsets, they are excluded from the set of explanatory variables in the regression.

Table 5.40: Regression Approach With Offset

Gamma			Poisson		
Variable	*Coef.*	*Std. Err.*	*Variable*	*Coef.*	*Std. Err.*
(Intercept)	7.295	0.076	(Intercept)	-2.960	0.046
			`NoClaimCreditBC`	-0.669	0.058
`Type`:City	-1.191	0.093	`Type`:City	-0.852	0.056
`Type`:County	-1.953	0.103	`Type`:County	-1.263	0.062
`Type`:Misc	0.796	0.185	`Type`:Misc	-1.904	0.110
`Type`:School	-0.676	0.095	`Type`:School	-1.599	0.057
`Type`:Town	1.476	0.191	`Type`:Town	0.407	0.114
ϕ	3.196				
AIC	65,037		*AIC*	11,828	

Poisson-Gamma MLE (Model B)

Table 5.41 shows the coefficient estimates when truncated estimation procedures are used with the Poisson frequency and gamma severity distributions. Note that the AIC for the severity model is high, most likely because of the poor model fit in the upper tail. Estimates of the Poisson coefficient are shown in the second panel, where the recovered coefficients are different from Tables 5.39 and 5.40.

Table 5.41: Poisson-Gamma Maximum Likelihood (Model B)

Gamma			Poisson		
Variable	*Coef.*	*Std. Err.*	*Variable*	*Coef.*	*Std. Err.*
(Intercept)	8.157	0.091	(Intercept)	-1.924	0.059
`CoverageBC`	0.373	0.025	`CoverageBC`	0.591	0.016
			`NoClaimCreditBC`	-0.725	0.059
`Type`:City	-0.114	0.103	`Type`:City	0.118	0.064
`Type`:County	-0.109	0.133	`Type`:County	0.087	0.075
`Type`:Misc	0.395	0.185	`Type`:Misc	-0.635	0.112
`Type`:School	0.743	0.110	`Type`:School	-0.774	0.065
`Type`:Town	0.150	0.196	`Type`:Town	-0.195	0.117
ϕ	0.272	0.007			
AIC	70,656		AIC	10,282	

Poisson-*GB2* MLE (Model C)

Table 5.42 shows the coefficient estimates for the *GB2* severity model. For the severity model, notice that only the scale parameters are peril dependent, in order to reduce the number of parameters used in the model. Also, the reader can observe that the AIC for the severity model becomes much lower under the *GB2* model, compared with the gamma model in Table 5.41.

Table 5.42: *GB2* Maximum Likelihood (Model C)

Term	*Variable*	*Coef.*	*Std. Err.*
Low	(Intercept)	7.235	0.138
Medium	(Intercept)	7.619	0.164
	`CoverageBC`	0.198	0.028
High	(Intercept)	8.332	0.216
	`CoverageBC`	0.250	0.040
$\boldsymbol{\omega}_M$	(Intercept)	0.368	0.102
	`Type`:City	-0.705	0.124
	`Type`:County	0.467	0.142
	`Type`:Misc	-0.191	0.248
	`Type`:School	0.056	0.129
	`Type`:Town	-0.314	0.254
$\boldsymbol{\omega}_H$	(Intercept)	-0.669	0.134
	`Type`:City	-0.605	0.166
	`Type`:County	0.444	0.183
	`Type`:Misc	0.081	0.309
	`Type`:School	0.344	0.164
	`Type`:Town	-0.081	0.323
	σ	1.213	0.154
	α_1	1.653	0.337
	α_2	1.677	0.344
AIC		17,737	

Table 5.42 shows only the severity distribution coefficients. This could be paired with the Poisson model or more advanced frequency model assumptions such as zero-one inflated models. Comparison of various frequency model assumptions and coefficient estimates under *GB2* loss severities are shown in Table 5.43. In Table 5.43, the estimated frequency parameters differ, depending on whether the underlying losses are observed or unobserved. Estimation of the coefficients in Table 5.43 is explained in the following Section 5.2.3.

Table 5.43: Comparison of Coefficients for Frequency Models

		(1) Poisson Underlying		(2) 01-Poisson Underlying		(3) Poisson Censored Estimation		(4) 01-Poisson Censored Estimation	
		Coef.	*Std.E.*	*Coef.*	*Std.E.*	*Coef.*	*Std.E.*	*Coef.*	*Std.E.*
Poisson	(Intercept)	-2.874	0.054	-1.841	0.098	-1.955	0.060	-1.913	0.066
	`CoverageBC`	0.993	0.012	0.753	0.019	0.733	0.017	0.734	0.020
	`NoClaimCreditBC`	-0.668	0.047	-0.289	0.122	-0.574	0.059	-0.587	0.065
	`Type`:City	-0.597	0.058	-0.025	0.077	0.055	0.063	0.038	0.064
	`Type`:County	-0.540	0.064	-0.130	0.086	-0.190	0.075	-0.214	0.076
	`Type`:Misc	-1.884	0.113	-0.388	0.168	-0.494	0.112	-0.542	0.121
	`Type`:School	-0.988	0.056	-1.070	0.083	-0.792	0.065	-0.786	0.067
	`Type`:Town	0.360	0.113	-0.117	0.144	-0.045	0.117	-0.130	0.124
Zero	(Intercept)			-0.684	0.325			-13.048	9.776
	`CoverageBC`			0.074	0.071			1.228	1.017
	`NoClaimCreditBC`			1.147	0.243			-5.461	44.880
One	(Intercept)			-3.507	0.436			-6.400	3.860
	`CoverageBC`			0.481	0.089			-0.196	0.276
	`NoClaimCreditBC`			0.900	0.337			2.167	3.894

5.2.3 01-Inflated Model Estimation

To estimate the 01-inflated Poisson model from censored observations, we define $v = 1 - \mathrm{F}_Y(d)$ for notational convenience, since only those losses above the deductible would be observed as a claim. Then the observed zero probabilities would satisfy

$$
\begin{aligned}
&\Pr(N_{\lambda,g}(d) = 0) \\
&= \pi_0 + \pi_2 \mathrm{P}_\lambda(0) + \pi_1(1-v) + \pi_2 \mathrm{P}_\lambda(1)(1-v) + \pi_2 \mathrm{P}_\lambda(2)(1-v)^2 + \pi_2 \mathrm{P}_\lambda(3)(1-v)^3 + \dots \\
&= \pi_0 + \pi_2 \mathrm{P}_\lambda(0) + \pi_1(1-v) + \pi_2 \left[\lambda e^{-\lambda}(1-v) + \frac{e^{-\lambda}\lambda^2}{2!}(1-v)^2 + \frac{e^{-\lambda}\lambda^3}{3!}(1-v)^3 + \dots\right] \\
&= \pi_0 + \pi_2 \mathrm{P}_\lambda(0) + \pi_1(1-v) + \pi_2 \frac{e^{-\lambda}}{e^{-\lambda(1-v)}} \left[\lambda(1-v)e^{-\lambda(1-v)} + \frac{e^{-\lambda(1-v)}(\lambda(1-v))^2}{2!} + \dots\right] \\
&= \pi_0 + \pi_2 \mathrm{P}_\lambda(0) + \pi_1(1-v) + \pi_2 e^{-\lambda v}(1 - \mathrm{P}_{\lambda(1-v)}(0)) \\
&= \pi_0 + \pi_2 \mathrm{P}_\lambda(0) + \pi_1(1-v) + \pi_2 \mathrm{P}_{\lambda v}(0)(1 - \mathrm{P}_{\lambda(1-v)}(0)),
\end{aligned}
$$

where we use the notation $\mathrm{P}_{\lambda(1-v)}(n)$ to denote the probability of a random variable with secondary Poisson distribution with parameter $\lambda(1-v)$ being n.

The probability of one claim being observed is

$$
\begin{aligned}
&\Pr(N_{\lambda,g}(d)=1)\\
&=\pi_1+\pi_2\mathrm{P}_\lambda(1)v\binom{1}{1}+\pi_2\mathrm{P}_\lambda(2)(1-v)v\binom{2}{1}+\pi_2\mathrm{P}_\lambda(3)(1-v)^2v\binom{3}{1}+\ldots\\
&=\pi_1+\pi_2v\left[\mathrm{P}_\lambda(1)+2\mathrm{P}_\lambda(2)(1-v)+3\mathrm{P}_\lambda(3)(1-v)^2+\ldots\right]\\
&=\pi_1+\pi_2v\left[\lambda e^{-\lambda}+2\frac{\lambda^2e^{-\lambda}}{2!}(1-v)+3\frac{\lambda^3e^{-\lambda}}{3!}(1-v)^2+\ldots\right]\\
&=\pi_1+\pi_2v\lambda\frac{e^{-\lambda}}{e^{-\lambda(1-v)}}\left[e^{-\lambda(1-v)}+\frac{(\lambda(1-v))e^{-\lambda(1-v)}}{1!}+\frac{(\lambda(1-v))^2e^{-\lambda(1-v)}}{2!}+\ldots\right]\\
&=\pi_1+\pi_2\lambda ve^{-\lambda v}(1-\mathrm{P}_{\lambda(1-v)}(0))\\
&=\pi_1+\pi_2\cdot\mathrm{P}_{\lambda v}(1)\cdot(1-\mathrm{P}_{\lambda(1-v)}(0)),
\end{aligned}
$$

and the probability of n claims being observed is

$$
\begin{aligned}
&\Pr(N_{\lambda,g}(d)=n)\\
&=\pi_2\left[\mathrm{P}_\lambda(n)v^n\binom{n}{0}+\mathrm{P}_\lambda(n+1)v^n(1-v)\binom{n+1}{1}+\mathrm{P}_\lambda(n+2)v^n(1-v)^2\binom{n+2}{2}+\ldots\right]\\
&=\pi_2v^n\lambda^ne^{-\lambda}\frac{1}{n!}\left[1+\frac{\lambda(1-v)}{1!}+\frac{(\lambda(1-v))^2}{2!}+\ldots\right]\\
&=\pi_2\frac{(\lambda v)^ne^{-\lambda v}}{n!}\\
&=\pi_2\Pr\left(N_{\lambda v}=n\right)\qquad\text{for } n\geq 2.
\end{aligned}
$$

Using these terms, the log-likelihood for policyholder i can be specified as

$$
\begin{aligned}
\log L_i=\log\{&\Pr(N_{\lambda_i,g}(d_i)=0)\cdot\mathrm{I}(N_{\lambda_i,g}(d_i)=0)\\
&+\Pr(N_{\lambda_i,g}(d_i)=1)\cdot\mathrm{I}(N_{\lambda_i,g}(d_i)=1)\\
&+\Pr(N_{\lambda_i,g}(d_i)=n_{\lambda_i,g})\cdot\mathrm{I}(N_{\lambda_i,g}(d_i)=n_{\lambda_i,g}(d_i))\}.
\end{aligned}
$$

Because the estimation of 01-inflated models under deductible influence is an interesting application, the details have been shown here. Section 3.5.4 of the dissertation illustrates the performance of this model in comparison with the basic Poisson model, with and without deductible influence.

5.2.4 Summary of Deductible Ratemaking

Section 5.2 has shown the censored and truncated estimation results, in relation to Chapter 3. The difference between the models in Section 5.1 and 5.2 is the type of observed variable. Section 5.1 assumes the average severity is observed over policyholder-years, while in Section 5.2 the loss severities are assumed to be observed at the claim level without averaging.

The following, Section 5.3 is a case study of insurance portfolio optimization (introduced in Chapter 4) using the models in Section 5.1.

5.3 Insurance Portfolio Optimization

Chapter 4 introduced insurance portfolio optimization methods, where the goal is to determine the optimal risk retention parameter, subject to a premium constraint. Using the models in Section 5.1, the goal in this chapter is to find the optimal risk retention parameter for a single policyholder. Hence, the insurance portfolio optimization approach is applied to the models of the LGPIF, in order to determine the optimal per-loss and annual aggregate upper limit for the school entity Madison Metropolitan School District (MMSD). Specifically, the building and contents claim model introduced in Section 5.1.2, Table 5.11 is used for illustration.

5.3.1 Framework

Using the framework in Section 4.8.1, let

$$S_{g,i}(u_1, u_2) = \text{Aggregate-Loss}_i(u_1, u_2) = \min\left(\sum_{j=1}^{N_i} \min(Y_{ij}, u_1), u_2\right)$$

with

$$q_{g,i} = \alpha\text{-quantile of the aggregate loss for policyholder } i$$

where we assume the insurer wishes to minimize $q_{g,i}$ subject to a restriction on the premiums. As in the framework of Section 4.8.1, our goal is to

$$\begin{aligned} &\text{minimize} && q_{g,i} \\ &\text{subject to} && E[S_{g,i}] \geq P_{min} \end{aligned}$$

5.3.2 Madison Metropolitan School District Policy

A portfolio manager may enter into a reinsurance treaty analogous to the framework in Section 5.3.1 for the largest school district in the portfolio, in order to mitigate the impact of a single large loss causing unexpected amounts in claims. The manager may be interested in a per-loss coverage, as well as a coverage on the aggregate losses at the annual aggregate loss level. The question is, how much of the risk should be retained, and how much should be reinsured?

For this study, the policy i of interest is MMSD, which is a school entity with $585,877,641$

in coverage. MMSD has *GB2* parameters $\sigma = 0.343355$, $\mu = 11.29235$, and $\alpha_1 = 0.4863737$, $\alpha_2 = 0.3488046$, and the probability of zero and one claims each 0.4993782, and 0.2791071, and $E[P_i(n)] = 13.899$ using a 01-inflated negative binomial model as described in Section 5.1.2.

5.3.3 The Omit-i Portfolio

The omit-i portfolio is simulated assuming a hypothetical insurance company, carrying the BC claims only.

Table 5.44: Summary Statistics for Simulated Aggregate Portfolio Losses

Min.	Median	Mean	Max.	B
759,300	6,245,000	31,380,000	4,753,000,000	10,000

Table 5.44 shows summary statistics for the simulated omit-i portfolio, using the 01-inflated negative binomial models, and *GB2* severity models for each of the policies. The long-tail nature of the *GB2* distribution is illustrated in Table 5.44. The average aggregate losses for the omit-i portfolio turned out to be 31.38 million, which is reasonably close to the realized empirical losses in year 2011. Nevertheless, the maximum simulated aggregate portfolio loss is 4,753 million, which is extremely high.

To explain the details of the simulation of the omit-i portfolio, among the 1095 policies in the out-of-sample data with BC coverage, the 1094 policies excluding MMSD are each simulated with $B = 10,000$ replicates of frequency and severity models. The resulting claims are aggregated to form $B = 10,000$ replicates of omit-i portfolio aggregate losses $S_{(i)}$. Thus

$$S_g = \min\left(\left[\sum_{i=1}^{N_i} \min(Y_{ij}, u_1)\right], u_2\right) + S_{(i)}$$

can be obtained, and the quantile and mean-excess function can be used for the optimization. That is, the quantile

$$\begin{aligned} q_g &= \alpha\text{-quantile of the aggregate loss for the entire portfolio} \\ &= \inf\left\{F_{S_g}(y) = \alpha\right\} \end{aligned}$$

where F_{S_g} is the distribution for S_g can be optimized. Because the portfolio losses consist of more policies than a single loss case, the premium constraints and resulting quantile should be larger than the first case. With the omit-i portfolio, the problem becomes

$$\begin{aligned} \text{minimize} \quad & q_g \\ \text{subject to} \quad & E[S_g] \geq P_{min}. \end{aligned}$$

Both cases (without the omit-i portfolio, and with the omit-i portfolio) are shown in Tables 5.45 and 5.46 of the following section.

5.3.4 Results

Without Omit-i Portfolio

First, it is natural to ask, what would be the optimal parameter, if we ignore the omit-i portfolio and simply optimize the single MMSD policy parameters? This section shows the result.

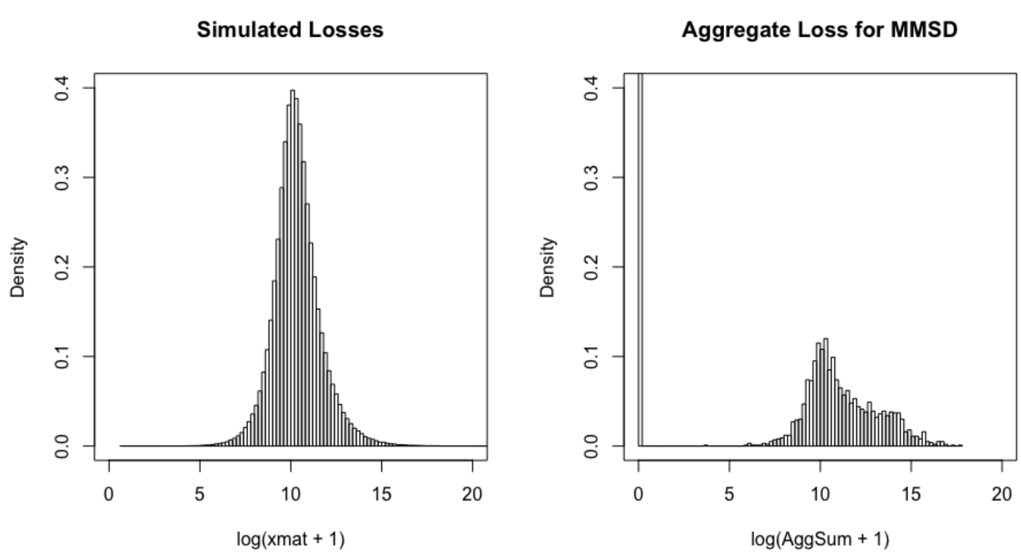


Figure 5.21: Histogram of Simulated Loss (Left Panel) and Aggregate Loss (Right Panel) for MMSD

Figure 5.21 shows the distribution of simulated losses, and the aggregate loss density, on a log-scale. In the left panel of Figure 5.21, each of the simulated Y_{ij} losses for policy MMSD are plotted. Each loss is $GB2$ distributed, and hence all positive. The right panel shows the aggregate losses, $\sum_{j=1}^{N_i} Y_{ij}$. If the realized N_i value is zero, then the aggregate loss is zero, hence the aggregate loss density has a mass point at zero. The distribution is shown before applying the parameter u_i.

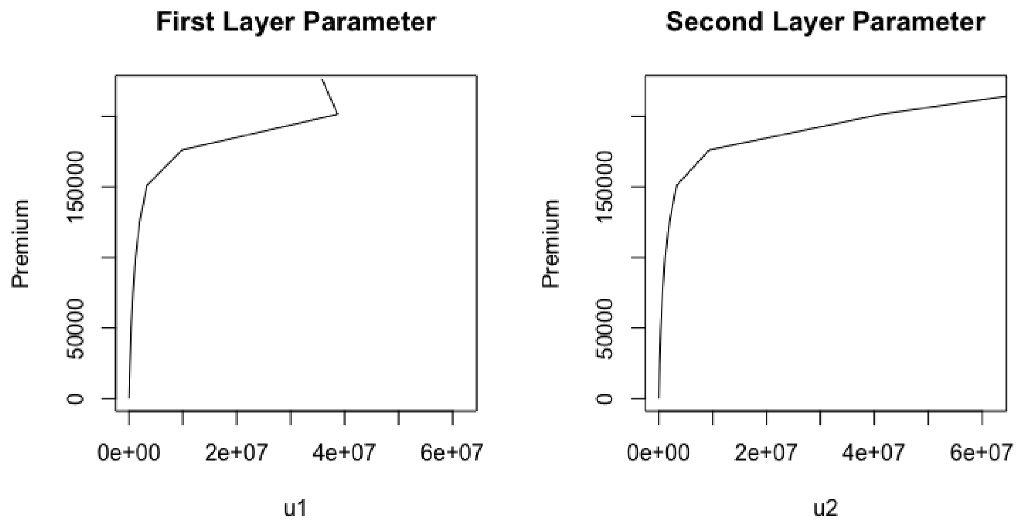


Figure 5.22: Optimization Result Without Omit-i Portfolio

In Figure 5.22, the optimization result is shown. The left panel of 5.22 shows the graph of u_1^* for various $E[S_{g,i}] \geq P_{min}$ constraints , and the right panel shows the graph of u_2^* for various $E[S_{g,i}] \geq P_{min}$ constraints. The parameter pair (u_1, u_2) is optimized simultaneously, and an optimal value (u_1^*, u_2^*) is found for each given P_{min} value. The first parameter u_1 applies to each of the losses, and the second parameter u_2 applies at the aggregate loss level.

In Table 5.45, the actual optimal upper limit values are shown in a table. As P_{min} is varied from 25,181 to 226,626, the optimal parameter pair (u_1^*, u_2^*) varies from small values to high values. Notice that the third column is the constraint, whereas the fourth column is the actual realized empirical expected value under the parameter (u_1^*, u_2^*). It can be observed that the constraint is satisfied. The quantile increases from 127,654 to 4,219,594. The value 4,219,594 is the quantile of the underlying loss, ie. when $(u_1, u_2) = (\infty, \infty)$.

From Table 5.45, the reader may observe that the quantile is fixed to 4,219,594, once the upper limits are large enough. The premium constraint is varied within the interval $[0, 226, 626]$. As the premium constraint is increased, the optimal upper limit increases, until the constraint is binding. The parameter pair with the minimum $q_{g,i}$ is returned.

Another observation is that the second upper limit parameter is higher than the 95th per-

Table 5.45: Optimization Result Without Portfolio

u_1	u_2	P_{min}	$E[S_{g,i}]$	$q_{g,i}$
198,929	127,654	25,181	25,181	127,654
413,984	370,951	50,361	50,361	370,951
743,103	711,706	75,542	75,542	711,706
1,215,740	1,213,764	100,723	100,723	1,213,764
1,998,173	1,998,052	125,903	125,903	1,998,052
3,325,589	3,325,580	151,084	151,084	3,325,580
9,873,521	9,412,247	176,264	191,726	4,219,594
38,706,607	41,235,780	201,445	230,101	4,219,594
35,779,709	86,894,710	226,626	228,892	4,219,594

centile value $q_{g,i}$. The reason why it is possible for the quantile to be fixed at 4,219,594 while the mean $E[S_{g,i}]$ continues to increase while the upper limit parameters increase to extremely high $u_1 = 35,779,709$ and $u_2 = 86,894,710$, is due to the long-tail nature of the *GB2* distribution. Presumably, even after the upper limit is as high as $u_1 = 1,215,740$ and $u_2 = 1,213,764$, the mean is only 100,723, meaning that a lot of mass is towards the tail of the aggregate loss distribution.

With Omit-i Portfolio

When the omit-i portfolio is added, the aggregate loss is much larger. The first panel of Figure 5.23 shows the distribution of $S_{g,i} + S_{(i)}$, without any influence of the upper limit parameter. The plot is generated by simulating $B = 10,000$ replicates of $S_{g,i}$ from Figure 5.21 and adding the result to $B = 10,000$ replicates of $S_{(i)}$. From Figure 5.21, notice that the mass at zero has disappeared. This is because at a portfolio level, mostly all of the simulated portfolio losses ($B = 10,000$ replicates) will have at least one policy with a positive loss.

In Figure 5.24, the optimal upper limit parameter (u_1^*, u_2^*) is plotted by optimizing the 99% quantile for $S_{g,i} + S_{(i)}$ under increasing premium constraints $E[S_g] \geq P_{min}$. Notice that the curve for both u_1^* and u_2^* are not completely smooth depending on how the optimization routine has converged. The upper limit parameters apply to the ith policy (MMSD) only, and not the rest of the portfolio.

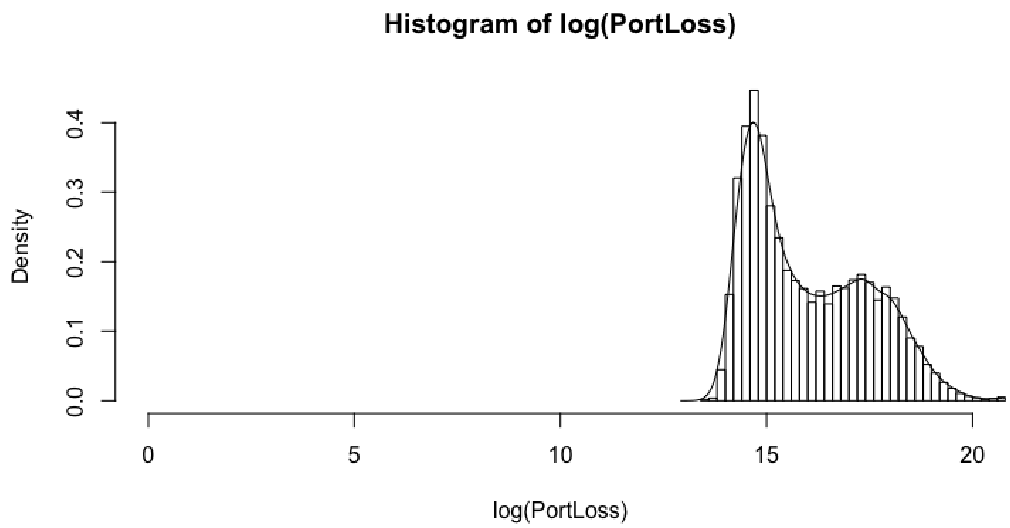


Figure 5.23: Histogram of Simulated Portfolio Aggregate Losses

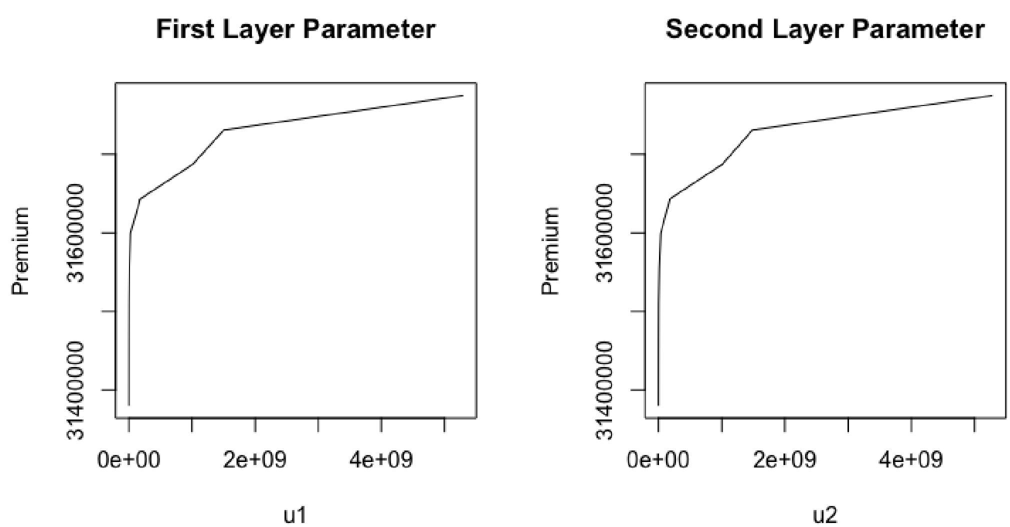


Figure 5.24: Optimization Result With Portfolio

Table 5.46 shows the actual optimal parameters, when the portfolio loss is optimized. The P_{min} constraint is varied from between $[31.4 \text{ mil}, 31.8 \text{ mil}]$, where the lowest possible P_{min} value is the average of the omit-i portfolio loss distribution. The highest possible value is determined by simulating S_g using $u_1 = u_2 = \infty$. It is advised that the manager selects the P_{min} value wisely, and in the process use Table 5.46 as a reference for the decision making.

Table 5.46: Optimization Result With Portfolio

u_1	u_2	P_{min}	$E[S_g]$	q_g
231,289	387,532	31,423,842	31,428,300	288,201,063
729,649	1,222,013	31,467,683	31,476,526	288,201,063
2,012,678	3,370,341	31,511,525	31,524,752	288,201,063
7,594,563	12,716,460	31,555,366	31,570,239	288,201,063
23,065,340	35,309,580	31,599,208	31,599,235	288,201,063
179,376,300	181,474,100	31,643,049	31,643,910	288,389,004
1,009,057,000	1,007,782,000	31,686,891	31,686,892	288,389,004
1,504,771,000	1,486,168,000	31,730,732	31,698,717	288,389,004
5,293,603,000	5,280,993,000	31,774,574	31,774,574	288,389,004

5.3.5 Summary of Insurance Portfolio Optimization Case Study

In this section, the LGPIF data has been applied to the insurance portfolio optimization problem for the Madison Metropolitan School District Policy. Optimal upper limit parameters have been determined, depending on the premium constraint to the policy. If the premium constraint is imposed to the portfolio, then the omit-i portfolio can be added, in order to determine the parameter pair satisfying the constraint at the portfolio level.

Using this methodology, the insurance manager has a framework to determine the optimal upper limit, deductible, and coinsurance for each policy. Improvements are possible, and some potential future work is summarized in the following, Section 6.

Chapter 6

Conclusion and Future Work

To summarize, we have started the dissertation by motivating actuarial problem solving under the framework of first defining the problem, then devising a solution, and finally monitoring the solution. This framework motivated the development of more accurate rating algorithms for the LGPIF. Meanwhile, motivation for the search of more advanced and fundamental solutions to the solvency issue of the LGPIF lead us to study advanced insurance claims models, deductible rating methodologies, and insurance risk retention problems.

Based on the situation in the LGPIF, we have first determined that accurate rating algorithms, and optimal risk retention could help improve the solvency of the LGPIF. Depending on the problem definition, our models and methodologies began to focus on specific chapters of the dissertation.

6.1 Insurance Claims Modeling

In Chapter 2, multivariate approaches to insurance claims modeling were reviewed. For future work, claims may be categorized into further detail using the peril types, and the spatial relationship may be explored. Various categorization methods can be used to cluster and categorize the claims into categories, so as to discover interesting relationships among the categories of insurance claims. The longitudinal nature of the data may be explored further.

In this dissertation, the hierarchical nature of insurance claims and the influence of censoring and truncation has been explored. More case studies using various models, and studying the influence of censoring and truncation could be interesting extensions to the work done in this dissertation.

An example of a hierarchical model in the insurance context is Frees et al. (2009).

In addition, advanced classification and clustering methods can be used for insurance claims modeling. Specifically, textual information could be utilized to classify insurance claims. This is known as *claims triaging* methods. With new machine learning methods emerging, insurance claims modeling using non-traditional data may become an active field of research in the future. For example, in the future it may be possible to utilize bitmap image data to more accurately predict the claim frequency and severity of insurance risks, and classify insurance policies.

6.2 Deductible Ratemaking

In Chapter 3, deductible ratemaking approaches have been reviewed. The standard way of modeling insurance claims treats deductible choices as exogenously given. Deductible levels are either included in a regression model, or used as a constant input to calculating the coverage modification amount. In both cases, relativities for various deductible levels can be obtained for insurance ratemaking purposes.

An new interpretation of deductibles may be suggested in the future. This new interpretation may allow for the flexibility of applying multivariate analysis to deductible choices and other response variables common in insurance analytics, such as claim frequencies, and severities. A survey of various ordinal response models may be a possible next step for flexible modeling of deductible choices. It may be interesting to compare various approaches to deductible choice modeling: One may interpret deductibles as a continuous random variable, while other approaches may treat them as ordinal responses.

Once a model for deductible choice is built, it may be possible to obtain distributions of loss frequencies and severities conditional on the deductible choice, using dependence models. This would allow for underlying loss models to be dependent on the deductible choice made by the insurance policyholder. Several potential models for deductible choice are possible. The proportional odds model for deductible choice may be given by

$$P(Y \leq j|\boldsymbol{x}) = \text{logit}^{-1}\left(\alpha_j + \boldsymbol{\beta}'\boldsymbol{x}\right), \quad j = 1, \ldots, J-1.$$

The cutpoints α_j are increasing in j, and the cumulative probabilities $P(Y \leq j|\boldsymbol{x})$ are increasing in j as well. For example, let $\boldsymbol{x} = \boldsymbol{x}_1 = x_1$ if the entity type is city, and $\boldsymbol{x} = \boldsymbol{x}_2 = x_2$ if the entity type is county, and let $\boldsymbol{\beta} = \beta$. Then, for a fixed j, the odds ratio of cumulative probabilities is proportional to the slope β, providing an expression for the difference in the logit of the cumulative probabilities under the two entity types. Hence, the name proportional odds model. More general models are also possible. Let

$$P(Y \leq j|\boldsymbol{x}) = F_Z(z)$$

for a latent variable Z with any parametric distribution, possibly fat-tailed. The choice of distribution for the latent variable Z would depend on the nature of the underlying data.

6.3 Insurance Portfolio Optimization

In Chapter 4, we have been able to optimize the risk retention parameters for simple example problems of risk retention. The general procedure is to visualize the domain over which the quantile function is quasi-convex, and obtain a global minimizer of the function within a compact domain satisfying the premium constraint. We have shown that the following expressions for the premium (μ_g) and quantile (q_g) sensitivities can be obtained using quantile sensitivity formulas:

$$\begin{aligned}
\frac{\partial q_g}{\partial d_i} &= -\frac{c_i}{f_{S_g}(q_g)} E\left[\left(\sum_{j=1}^{N_i} I(Y_{ij} > d_i)\right) f_{S_{(i)}}\left(q_g - \sum_{j=1}^{N_i} g(Y_{ij})\right)\right] \\
\frac{\partial q_g}{\partial u_i} &= \frac{c_i}{f_{S_g}(q_g)} \cdot E\left[\left(\sum_{j=1}^{N_i} I(Y_{ij} > u_i)\right) f_{S_{(i)}}\left(q_g - \sum_{j=1}^{N_i} g(Y_{ij})\right)\right] \\
\frac{\partial q_g}{\partial c_i} &= \frac{1}{f_{S_g}(q_g)} E\Bigg[\Bigg(\sum_{j=1}^{N_i} (Y_{ij} - d_i) \cdot I(Y_{ij} > d_i) \\
&\qquad - (Y_{ij} - u_i) \cdot I(Y_{ij} > u_i)\Bigg) f_{S_{(i)}}\left(q_g - \sum_{j=1}^{N_i} g(Y_{ij})\right)\Bigg]
\end{aligned}$$

The mean sensitivities are:

$$\frac{\partial}{\partial d_i}\mu = -c_i \cdot E[N_i] \cdot (1 - F_Y(d_i))$$
$$\frac{\partial}{\partial u_i}\mu = c_i \cdot E[N_i] \cdot (1 - F_Y(u_i))$$
$$\frac{\partial}{\partial c_i}\mu = E[N_i] \cdot \int_{d_i}^{u_i} (1 - F_Y(y))dy.$$

For commonly used loss distributions, we have been able to see that

$$\frac{\partial q_g}{\partial d_i} < 0, \quad \text{and} \quad \frac{\partial q_g}{\partial u_i} > 0, \quad \text{and} \quad \frac{\partial q_g}{\partial c_i} > 0$$

and we may also notice that

$$\frac{\partial \mu_g}{\partial d_i} < 0, \quad \text{and} \quad \frac{\partial \mu_g}{\partial u_i} > 0, \quad \text{and} \quad \frac{\partial \mu_g}{\partial c_i} > 0$$

with the curvature (second derivative) of the quantile function depending largely on the curvature of the omit-i portfolio distribution. Thus, for simple one-dimensional optimization problems, we have been able to provide answers to the optimal risk retention parameter, d_i, c_i, and u_i. When multiple parameters are optimized simultaneously, quasi-convexity of the quantile function is not guaranteed. However, for the case where two upper limits are present each at different layers, we were able to obtain reasonable optima for the upper limits u_1 and u_2, for an empirical study using the LGPIF data.

There are many avenues for future work related to the risk retention and portfolio optimization chapter of this dissertation. One possible future work may be to focus on the multiple layers problem. It may be potentially interesting to obtain real reinsurance market data to apply the methodologies in this part of the dissertation. For a multiple parameter problem, how should the RM^2 measure be defined, and how should it relate to the loading factor of the reinsurance premium? These are open questions to explore.

6.4 Summary of Conclusion and Future Work

In conclusion, this dissertation provided an overview of multivariate insurance claims modeling, as well as an overview of insurance ratemaking in the presence of deductibles. The dissertation provided a rigorous overview of insurance claims modeling, and insurance deductible ratemaking. Based on a data-driven way of assessing the underlying loss distribution, models are estimated and utilized for ratemaking and insurance portfolio optimization.

For this reason, this dissertation assumes the deductible parameter d_i, nor other risk retention parameters such as u_i, or c_i alters the underlying distribution of losses. In reality, the parameter changes may influence the underlying distributions, and hence may be a topic for future work.

Bibliography

Aban, Inmaculada B., Mark M. Meerschaert, and Anna K. Panorska (2006). “Parameter Estimation for the Truncated Pareto Distribution,” *Journal of the American Statistical Association*, Vol. 101(473), pp. 270–277.

Acar, Elif F., Radu V. Craiu, and Fang Yao (2011). “Dependence Calibration in Conditional Copulas: A Nonparametric Approach,” *Biometrics*, Vol. 67, pp. 445–453.

Anscombe, F. J. (1952). “Large-sample Theory of Sequential Estimation,” *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 48, pp. 600–607.

Arrow, Kenneth J. (1974). “Optimal Insurance and Generalized Deductibles,” *Scandinavian Actuarial Journal*, Vol. 1, pp. 1–42.

Asimit, Alexandru V., Alexandru M. Badescu, and Tim Verdonck (2013). “Optimal Risk Transfer Under Quantile-Based Risk Measures,” *Insurance: Mathematics and Economics*, Vol. 53, pp. 252–265.

Assa, Hirbod (2015). “On Optimal Reinsurance Policy with Distortion Risk Measures and Premiums,” *Insurance: Mathematics and Economics*, Vol. 61, pp. 70–75.

Bahnemann, David (2015). *Distributions for Actuaries.* CAS Monograph Series, Number 2.

Bahraoui, Zuhair, Catalina Bolance, and Ana M. Perez-Marin (2014). “Testing Extreme Value Copulas to Estimate the Quantile,” *SORT*, Vol. 38, pp. 89–102.

Barr, Donald R. and Todd Sherrill (1999). “Mean and Variance of Truncated Normal Distributions,” *American Statistician*, Vol. 53(4), pp. 357–361.

Bernegger, Stefan (1997). “The Swiss Re Exposure Curves and the MBBEFD Distribution Class,” *ASTIN Bulletin*, Vol. 27, pp. 99–111.

Borch, K. (1960). “An Attempt to Determine the Optimum Amount of Stop Loss Reinsurance,” *Transactions of the 16th International Congress of Actuaries*, Vol. 1(3), pp. 597–610.

Boucher, Jean-Philippe (2014). “Regression with Count Dependent Variables,” in Edward W. Frees, Glenn Meyers, and Richard A. Derrig eds. *Predictive Modeling Applications in Actuarial Science*, Cambridge. Cambridge University Press.

Bowers, N.J., H.U. Gerber, J.C. Hickman, D.A. Jones, and C.J. Nesbitt (1997). “Actuarial Mathematics,” Vol. 2, pp. 201–223.

Boyd, Stephen and Lieven Vandenberghe (2004). *Convex Optimization.* Cambridge University Press.

Brechmann, Eike, Claudia Czado, and Sandra Paterlini (2014). “Flexible Dependence Modeling of Operational Risk Losses and its Impact on Total Capital Requirements,” *Journal of Banking & Finance*, Vol. 40, pp. 271–285.

Brockett, Patrick L., Linda L. Golden, Montserrat Guillén, Jens P. Nielsen, Jan Parner, and Ana M. Pérez-Marín (2008). “Survival analysis of a household portfolio of insurance policies: how much time do you have to stop total customer defection?” *Journal of Risk and Insurance*, Vol. 75, pp. 713–737.

Brown, Robert L. and W. Scott Lennox (2015). *Ratemaking and Loss Reserving for Property and Casualty Insurance*. ACTEX Publications.

Cai, Jun, Christiane Lemieux, and Fangda Liu (2015). "Optimal Reinsurance from the Perspectives of Both an Insurer and a Reinsurer," *Astin Bulletin*, Vol. 46, pp. 815 – 849.

Cai, Jun, Haiyan Liu, and Ruodu Wang (2017). "Pareto-optimal Reinsurance Arrangements Under General Model Settings," *forthcoming*.

Cai, Jun and Ken Seng Tan (2007). "Optimal Retention for a Stop-loss Reinsurance Under the VaR and CTE Risk Measures," *ASTIN Bulletin*, Vol. 37(1), pp. 93–112.

Cai, Jun, Ken Seng Tan, Chengguo Weng, and Yi Zhang (2008). "Optimal Reinsurance under VaR and CTE Risk Measures," *Insurance: Mathematics and Economics*, Vol. 43, pp. 185–196.

Cai, Jun and Wei Wei (2012a). "Optimal Reinsurance with Positively Dependent Risks," *Insurance: Mathematics and Economics*, Vol. 50, pp. 57 – 63.

——— (2012b). "Optimal Reinsurance with Positively Dependent Risks," *Insurance: Mathematics and Economics*, Vol. 50, pp. 57–63.

Chapman, Douglas G. (1956). "Estimating the Parameters of a Truncated Gamma Distribution," *Annals of Mathematical Statistics*, Vol. 27(2), pp. 498–506.

Chavez-Demoulin, Valerie, Paul Embrechts, and Marius Hofert (2016). "An Extreme Value Approach for Modeling Operational Risk Losses Depending on Covariates," *The Journal of Risk and Insurance*, Vol. 83(3), pp. 735–776.

Cohen, Alma and Liran Einav (2007). "Estimating Risk Preferences from Deductible Choice," *American Economic Review*, Vol. 97(3), pp. 745–788.

Cox, David R and E Joyce Snell (1968). "A General Definition of Residuals," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 248–275.

Cummings, David (2005). "Practical GLM Analysis of Homeowners," *Presentation, State Farm Insurance Companies*.

Czado, Claudia, Rainer Kastenmeier, Eike Christian Brechmann, and Aleksey Min (2012). "A Mixed Copula Model for Insurance Claims and Claim Sizes," *Scandinavian Actuarial Journal*, Vol. 2012, pp. 278–305.

Daykin, C.D., T. Pentiakainen, and M. Pesonen (1994). *Practical Risk Theory for Actuaries*. Chapman and Hall, London.

De Jong, Piet and Gillian Z. Heller (2008). *Generalized Linear Models for Insurance Data*. Cambridge University Press Cambridge.

Einav, Liran, Amy Finkelstein, Iuliana Pascu, and Mark R. Cullen (2012). "How General Are Risk Preferences? Choices Under Uncertainty in Different Domains," *American Economic Review*, Vol. 102(6), pp. 2606–2638.

Embrechts, Paul, Haiyan Liu, and Ruodu Wang (2017). "Quantile-based Risk Sharing," *forthcoming*.

Fasen, Vicky and Claudia Kluppelberg (2014). "Large Insurance Losses Distributions," *Wiley StatsRef: Statistics Reference Online*.

Finkelstein, D. M. and R. A. Wolfe (1985). "A Semiparametric Model for Regression Analysis of Interval-Censored Failure Time Data," *Biometrics*, Vol. 41(4), pp. 933–945.

Frees, E. W., J. Gao, and M. A. Rosenberg (2011a). "Predicting the Frequency and Amount of Health Care Expenditures," *North American Actuarial Journal*, Vol. 15(3).

Frees, Edward W. (2004). *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences*. Cambridge University Press.

——— (2014). "Frequency and Severity Models," in Edward W. Frees, Glenn Meyers, and Richard A. Derrig eds. *Predictive Modeling Applications in Actuarial Science*, Cambridge. Cambridge University Press.

——— (2015). "Analytics of Insurance Markets," *Annual Review of Finacial Economics*, Vol. 7, URL: http://www.annualreviews.org/loi/financial, Will be available at: http://www.annualreviews.org/loi/financial.

——— (2016). "Insurance Risk Retention," *forthcoming*.

Frees, Edward W., Xiaoli Jin, and Xiao Lin (2013). "Actuarial applications of multivariate two-part regression models," *Annals of Actuarial Science*, Vol. 7, pp. 258–287.

Frees, Edward W. and Gee Lee (2017). "Rating Endorsements Using Generalized Linear Models," *Variance*, Vol. 10(1), Will be available at: http://www.variancejournal.org/issues/.

Frees, Edward W., Gee Lee, and Lu Yang (2016). "Multivariate Frequency-Severity Regression Models in Insurance," *Risks*, Vol. 4(1).

Frees, Edward W., Glenn Meyers, and A. David Cummings (2011b). "Summarizing Insurance Scores Using a Gini Index," *Journal of the American Statistical Association*, Vol. 106.

Frees, Edward W., Peng Shi, and Emiliano A. Valdez (2009). "Actuarial Applications of a Hierarchical Insurance Claims Model," *Astin Bulletin*, Vol. 39, pp. 165–197.

Frees, Edward W and Yunjie Sun (2010). "Household Life Insurance Demand: A Multivariate Two-part Model," *North American Actuarial Journal*, Vol. 14, pp. 338–354.

Frees, Edward W. and Emiliano A. Valdez (1998). "Understanding Relationships using Copulas," *North American Actuarial Journal*, Vol. 2, pp. 1–25.

——— (2008). "Hierarchical Insurance Claims Modeling," *Journal of the American Statistical Association*, Vol. 103(484), pp. 1457–1469.

Frees, Edward W. and Ping Wang (2005). "Credibility using Copulas," *North American Actuarial Journal*, Vol. 9, pp. 31–48.

——— (2006). "Copula Credibility for Aggregate Loss Models," *Insurance: Mathematics and Economics*, Vol. 38, pp. 360–373.

Fu, Luyang and Xianfang Liu (forthcoming). "Finite Mixture Model and Workers Compensation Large Loss Regression Analysis," in Edward W. Frees, Glenn Meyers, and Richard A. Derrig eds. *Predictive Modeling Applications in Actuarial Science, Vol. II*, Cambridge. Cambridge University Press.

Genest, Christian and Jean-Claude Boies (2003). "Detecting Dependence with Kendall Plots," *Journal of the American Statistical Association*, Vol. 57, pp. 275–284.

Genest, Christian, Ivan Kojadinovic, Johanna Neslehova, and Jun Yan (2011). "A Goodness-of-fit Test for Bivariate Extreme Value Copulas," *Bernoulli*, Vol. 17(1), p. 253275.

Genest, Christian and Johanna Nešlehová (2007). "A Primer on Copulas for Count Data," *Astin Bulletin*, Vol. 37, pp. 475–515.

Genest, Christian, Aristidis K Nikoloulopoulos, Louis-Paul Rivest, and Mathieu Fortin (2013). "Predicting Dependent Binary Outcomes Through Logistic Regressions and Meta-elliptical Copulas," *Brazilian Journal of Probability and Statistics*, Vol. 27, pp. 265–284.

Genest, Christian and Louis-Paul Rivest (1993). "Statistical Inference Procedures for Bivariate Archimedean Copulas," *Journal of the American Statistical Association*, Vol. 88(423), pp. 1034–1043.

Gollier, Christian (2013). "The Economics of Optimal Insurance Design," *Dionne, Georges (ed) Handbook of Insurance, Second Edition*, pp. 107–122.

Gray, Roger J. and Susan M. Pitts (2012). *Risk Modelling in General Insurance: from Principles to Practice*. Cambridge University Press.

Gschlößl, Susanne and Claudia Czado (2007). "Spatial Modelling of Claim Frequency and Claim Size in Non-life Insurance," *Scandinavian Actuarial Journal*, Vol. 2007, pp. 202–225.

Guiahi, Farrokh (2001). "Fitting Loss Distributions with Emphasis on Rating Variables," *CAS Specialty Seminar*.

Guillén, Montserrat (2014). "Regression with Categorical Dependent Variables," in Edward W. Frees, Glenn Meyers, and Richard A. Derrig eds. *Predictive Modeling Applications in Actuarial Science*, Cambridge. Cambridge University Press.

Guillén, Montserrat, Jens P. Nielsen, and Ana M. Pérez-Marín (2008). "The Need to Monitor Customer Loyalty and Business Risk in the European Insurance Industry," *The Geneva Papers on Risk and Insurance-Issues and Practice*, Vol. 33, pp. 207–218.

Gurnecki, Krzysztof, Grzegorz Kukla, and Rafal Weron (2006). "Modelling Catastrophe Claims With Left-Truncated Severity Distributions," *Computational Statistics*, Vol. 21, pp. 537–555.

Halek, Martin and Joseph Eisenhauer (2001). "Demography of Risk Aversion," *Journal of Risk and Insurance*, Vol. 68(1), pp. 1–24.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Science & Business Media.

Holt, Charles A. and Susan K. Laury (2002). "Risk Aversion and Incentive Effects," *American Economic Review*, Vol. 92(5), p. p.1644.

Hong, Jeff L. (2009). "Estimating Quantile Sensitivities," *Operations Research*, Vol. 57(1), pp. 118–130.

Hua, Lei (2015). "Tail Negative Dependence and its Applications for Aggregate Loss Modeling," *Insurance: Mathematics and Economics*, Vol. 61, pp. 135–145.

Jiang, Guangxin and Michael C. Fu (2015). "Technical Note - On Estimating Quantile Sensitivities via Information Perturbation Analysis," *Operations Research*, Vol. 63(2), pp. 435–441.

Joe, Harry (1997). *Multivariate Models and Multivariate Dependence Concepts*. CRC Press.

——— (2014). *Dependence Modeling with Copulas*. CRC Press.

Kaas, R., J. Goovaerts, M. Dhaene, and M. Denuit (2001). "Modern Actuarial Risk Theory."

Kalbfleisch, John D. and Ross L. Prentice (2002). *The Statistical Analysis of Failure Time Data, Second Edition*, New Jersey. John Wiley & Sons, Inc.

Kaplan, E. L. and P. Meier (1958). "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, Vol. 53, pp. 457–481.

Klugman, Stuart A., Harry H. Panjer, and Gordon E. Willmot (2012). *Loss Models: From Data to Decisions, Fourth Edition*, New Jersey. John Wiley & Sons.

Kojadinovic, Ivan, Johan Segers, and Jun Yan (2011). "Large-Sample tests of extreme value dependence for multivariate copulas," *The Canadian Journal of Statistics*, Vol. 39(4), p. 703720.

Kolev, Nikolai and Delhi Paiva (2009). "Copula-based Regression Models: A Survey," *Journal of Statistical Planning and Inference*, Vol. 139, pp. 3847–3856.

Koszegi, Botond and Matthew Rabin (2006). "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, Vol. 121(4), pp. 1133–1165.

Krämer, Nicole, Eike C Brechmann, Daniel Silvestrini, and Claudia Czado (2013). "Total Loss Estimation using Copula-based Regression Models," *Insurance: Mathematics and Economics*, Vol. 53, pp. 829–839.

Lai, Tze Leung and Zhiliang Ying (1991). “Estimating a Distribution Function With Truncated and Censored Data,” *Annals of Statistics*, Vol. 19(1), pp. 417–442.

Lambert, Philippe (1996). “Modelling Irregularly Sampled Profiles of Non-negative Dog Triglyceride Responses under Different Distributional Assumptions,” *Statistics in Medicine*, Vol. 15, pp. 1695–1708.

Li, Jianping, Xiaoqian Zhu, Jianming Chen, Lijun Gao, Jichuang Feng, Dengsheng Wu, and Xiaolei Sun (2014). “Operational Risk Aggregation Across Business Lines Based on Frequency Dependence and Loss Dependence,” *Mathematical Problems in Engineering*, Vol. 2014, Available at: http://www.hindawi.com/journals/mpe/2014/404208/.

Ludwig, Steven J. (1991). “An Exposure Rating Approach to Pricing Property Excess-of-Loss Reinsurance,” *Casualty Actuarial Society*, Vol. 78.

Markowitz, Harry (1952). “Portfolio Selection,” *The Journal of Finance*, Vol. 7, pp. 77 – 91.

Mas-Colell, Andreu, Michael D. Whinston, and Jerry Green (1995). *Microeconomic Theory, First Edition*, Oxford. Oxford University Press.

Meester, Steven G. and Jock Mackay (1994). “A Parametric Model for Cluster Correlated Categorical Data,” *Biometrics*, pp. 954–963.

Mildenhall, Stephen J (1999). “A Systematic Relationship Between Minimum Bias and Generalized Linear Models,” in *Proceedings of the Casualty Actuarial Society*, Vol. 86, pp. 393–487.

Mossin, Jan (1968). “Aspects of Rational Insurance Purchasing,” *Journal of Political Economy*, Vol. 76(4), pp. 553–568.

Nelsen, Roger B (1999). *An Introduction to Copulas*, Vol. 139. Springer Science & Business Media.

Nikoloulopoulos, Aristidis K (2013a). “Copula-based Models for Multivariate Discrete Response Data,” in *Copulae in Mathematical and Quantitative Finance*. Springer, pp. 231–249.

——— (2013b). “On the Estimation of Normal Copula Discrete Regression Models using the Continuous Extension and Simulated Likelihood,” *Journal of Statistical Planning and Inference*, Vol. 143, pp. 1923–1937.

Nikoloulopoulos, Aristidis K and Dimitris Karlis (2010). “Regression in a Copula Model for Bivariate Count Data,” *Journal of Applied Statistics*, Vol. 37, pp. 1555–1568.

Panagiotelis, Anastasios, Claudia Czado, and Harry Joe (2012). “Pair Copula Constructions for Multivariate Discrete Data,” *Journal of the American Statistical Association*, Vol. 107, pp. 1063–1072.

Patton, Andrew J (2006). “Modelling Asymmetric Exchange Rate Dependence*,” *International Economic Review*, Vol. 47, pp. 527–556.

Plackett, R. L. (1953). “The Truncated Poisson Distribution,” *International Biometric Society*, Vol. 9(4), pp. 485–488.

Prentice, Ross L (1974). “A Log Gamma Model and its Maximum Likelihood Estimation,” *Biometrika*, Vol. 61, pp. 539–544.

——— (1975). “Discrimination Among Some Parametric Models,” *Biometrika*, Vol. 62, pp. 607–614.

Rosenberg, Marjorie A., Richard W. Andrews, and Peter Lenk (1999). “A Hierarchical Bayesian Model for Predicting the Rate of Nonacceptable In-Patient Hospital Utilization,” *Journal of Business and Economic Statistics*, Vol. 17(1), pp. 1–8.

Rothschild, Michael and Joseph Stiglitz (1976). “Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information,” *Quarterly Journal of Economics*, Vol. 90(4), pp. 629–649.

Rüschendorf, Ludger (2009). “On the Distributional Transform, Sklar’s Theorem, and the Empirical Copula Process,” *Journal of Statistical Planning and Inference*, Vol. 139, pp. 3921–3927.

SAS (2010). *SAS/STAT 9.2 User's Guide, Second Edition*. SAS Institute Inc. URL: http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#titlepage.htm, Downloaded on 30 September 2015 at: http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#titlepage.htm.

Schlesinger, Harris (2013). "The Theory of Insurance Demand," in *Handbook of Insurance*. Springer New York, pp. 167–184.

Shi, Peng (2012). "Multivariate Longitudinal Modeling of Insurance Company Expenses," *Insurance: Mathematics and Economics*, Vol. 51, pp. 204–215.

——— (2014). "Fat-tailed Regression Models," in Edward W. Frees, Glenn Meyers, and Richard A. Derrig eds. *Predictive Modeling Applications in Actuarial Science*, Cambridge. Cambridge University Press.

——— (2016). "Insurance Ratemaking Using a Copula-Based Multivariate Tweedie Model," *Scandinavian Actuarial Journal*, Vol. 2016(3), pp. 198–215.

Shi, Peng, Xiaoping Feng, and Anastasia Ivantsova (2015). "Dependent Frequency–Severity Modeling of Insurance Claims," *Insurance: Mathematics and Economics*, Vol. 64, pp. 417–428.

Siegmund, David (1985). "The Sequential Probability Ratio Test," in D. Brillinger, S. Fienberg, J. Gani, J. Hartigan, and K. Krickeberg eds. *Sequential Analysis: Tests and Confidence Intervals*, New York. Springer Science and Business Media.

Simon, Carl P. and Lawrence E. Blume (1994). *Mathematics for Economists*. W. W. Norton & Company.

Sklar, M (1959). *Fonctions de Répartition à n Dimensions et Leurs Marges*. Université Paris 8.

Smyth, Gordon K (1989). "Generalized Linear Models with Varying Dispersion," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 47–60.

Song, Peter X-K, Yanqin Fan, and John D Kalbfleisch (2005). "Maximization by Parts in Likelihood Inference," *Journal of the American Statistical Association*, Vol. 100, pp. 1145–1158.

Song, Peter X-K, Mingyao Li, and Peng Zhang (2013). "Vector Generalized Linear Models: A Gaussian Copula Approach," in *Copulae in Mathematical and Quantitative Finance*. Springer, pp. 251–276.

Song, Xue-Kun (2007). *Correlated Data Analysis: Modeling, Analytics, and Applications*. Springer Science & Business Media.

Song, Xue-Kun Peter (2000). "Multivariate Dispersion Models Generated from Gaussian Copula," *Scandinavian Journal of Statistics*, Vol. 27, pp. 305–320.

Sun, Jiafeng, Edward W. Frees, and Marjorie A. Rosenberg (2008). "Heavy-Tailed Longitudinal Data Modeling Using Copulas," *Insurance: Mathematics and Economics*, Vol. 42(2), pp. 817–830.

Sundaram, Rangarajan K. (1996). *A First Course in Optimization Theory*. Cambridge University Press.

Sung, K.C.J., S.C.P. Yam, S.P. Yung, and J.H. Zhou (2011). "Behavioral Optimal Insurance," *Insurance: Mathematics and Economics*, Vol. 49, pp. 418–428.

Sydnor, Justin (2010). "(Over)insuring Modest Risks," *American Economic Journal: Applied Economics*, Vol. 2(4), pp. 177–199.

Tse, Yiu-Kuen (2009). *Nonlife Actuarial Models: Theory, Methods and Evaluation, First Edition*, Cambridge. Cambridge University Press.

Verbelen, R. and G. Claeskens (2014). "Multivariate Mixtures of Erlangs for Density Estimation Under Censoring and Truncation," *Social Science Research Network*.

Vuong, Quang H (1989). "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses," *Econometrica: Journal of the Econometric Society*, pp. 307–333.

White, Steve (2005). "Property Ratemaking — an Advanced Approach: Exposure Rating," *CAS Seminar on Reinsurance.*

White, Steve and Kari Mrazek (2004). "Advanced Exposure Rating: Beyond the Basics," *CAS Seminar on Reinsurance.*

Woodroofe, M. (1985). "Estimating a Distribution Function With Truncated Data," *Annals of Statistics*, Vol. 13(1), pp. 163–177.

Wu, S.Q. and X. Cai (1999). "Maxwell-Boltzmann, Bose-Einstein, Fermi-Dirac Statistical Entropies in a D-Dimensional Stationary Axisymmetry Space-Time," *Cornell University Library.*

Yang, Xipei (2011). "Multivariate Long-tailed Regression with New Copulas," Ph.D. dissertation, University of Wisconsin–Madison.

Yuan, Ming and Yi Lin (2006). "Model Selection and Estimation in Regression With Grouped Variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 68, pp. 49–67.

Chapter 7

Appendix: Rating Engine Summary

.

Abstract

This Appendix provides additional details of the LGPIF, as well as some intuition of how the rating engine for the LGPIF has been implemented.

The rating engine in this chapter is based on Frees, Edward W., and Gee Lee (2017). "Rating Endorsements using Generalized Linear Models," *Variance,* Vol. 10(1).

7.1 Overview

The primary purpose of this Appendix is to document the rating engine, so as to provide some additional details regarding the Local Government Property Insurance Fund (LGPIF). The chapter explains how rates for building & contents, motor vehicle, and contractor's equipment may be calculated. The recommended engine is in Section 7.2. This rating engine is built using modern analytic techniques known as generalized linear models, GLM, and calibrated to existing LGPIF data. A discussion of the techniques may be found in Frees and Lee (2017).

The Wisconsin Office of the Insurance Commissioner administers the LGPIF. The LGPIF was established to provide property insurance for local government entities that include counties, cities, towns, villages, school districts, and library boards. The fund insures local government property such as government buildings, schools, libraries, and motor vehicles. The fund covers all property losses except those resulting from flood, earthquake, wear and tear, extremes in temperature, mold, war, nuclear reactions, and embezzlement or theft by an employee.

In 2011, the fund covered over a thousand local government entities who pay approximately $25 million in premiums each year and receive insurance coverage of about $75 billion. State government buildings are not covered; the LGPIF is for local government entities that have separate budgetary responsibilities and who need insurance to moderate the budget effects of uncertain insurable events. Claims for state government buildings are charged to another state fund that essentially self-insures its properties.

The fund offers three major groups of insurance coverage: building and contents (BC), inland marine (construction equipment), and motor vehicles. In effect, the LGPIF acts as a stand-alone insurance company, charging premiums to each local government entity (policyholder) and paying claims when appropriate. Although the LGPIF is not permitted to deny coverage for local government entities, these entities may go onto the open market to secure coverage. Thus, the LGPIF acts as a "residual" market to a certain extent, meaning that other sources of market data may not reflect its experience.

7.2 Rating Engine

7.2.1 Building and Contents

The rates for the building & contents coverage is calculated using the formula shown below. The endorsement factors should include all endorsement coverage ratios purchased by the policyholder. Here, the endorsement factors are multiplicative. Frees and Lee (2017) explains how multiplicative endorsement factors may be interpreted as additive factors, by using specific covariates for the multiplicative factors. The rating variables are described in detail, in Table 7.1. Details of each term in the formula can be found in Table 7.2 and 7.3.

$$\begin{aligned}\texttt{BC Rate} &= \texttt{Base Rate} \times \texttt{Offset} \times \texttt{Deductible Factor} \times \\ &\quad \texttt{NoClaimCredit Factor} \times \texttt{EntityType Factor} \times \texttt{Endorsement Factors}\end{aligned}$$

Table 7.1: Description of Rating Variables

Rating variable	Formula
lnCoverage	log(building and contents coverage)
lnCoveragePNA	log(comprehensive new ACV coverage)
lnCoveragePNR	log(comprehensive new RC coverage)
lnCoverageCNA	log(collision new ACV coverage)
lnCoverageCNR	log(collision new RC coverage)
lnCoveragePOA	log(comprehensive old ACV coverage)
lnCoveragePOR	log(comprehensive old RC coverage)
lnCoverageCOA	log(collision old ACV coverage)
lnCoverageCOR	log(collision old RC coverage)
lnCoverageIM	log(contractor's equipment coverage)
lnDeduct	log(deductible)
lnDeductIM	log(contractor's equipment deductible)
NoClaimCredit	1 if the insured has no building and contents claim in two years
PWlnCovRat	log(1 + pier and wharf coverages/denom)
SAlnCovRat	log(1 + special use animal coverages/denom)
ZOOlnCovRat	log(1 + zoo animal coverages/denom)
FAlnCovRat	log(1 + fine arts coverages/denom)
GClnCovRat	log(1 + golf course coverages/denom)
BIlnCovRat	log(1 + business interruption coverages/denom)
ARlnCovRat	log(1 + accounts receivable coverages/denom)
MS1lnCovRat	log(1 + money and securities coverages/denom)
MS2lnCovRat	log(1 + money and securities limited term coverages/denom)
VPlnCovRat	log(1 + vacancy permit/denom)
ADDlnCovRat	log(1 + other endorsement coverages/denom)
AC05	1 if policyholder has 5% alarm credit
AC10	1 if policyholder has 10% alarm credit
AC15	1 if policyholder has 15% alarm credit
denom	building coverage/100

Table 7.2: Base Rating Engine

Variable	Value
Base Rate	exp[9.878 − 1.256 + 1.035 × lnCoverage]
Offset	exp[log(0.95) × AC05 + log(0.90) × AC10 + log(0.85) × AC15]
Deductible Factor	exp[−0.332 × lnDeduct]
NoClaimCredit Factor	exp[−0.050 × NoClaimCredit]

Table 7.3: Entity Type Factors

Variable	Value
City	exp[0.332]
County	exp[0.076]
Misc	exp[0.034]
School	exp[−0.032]
Town	exp[0.061]

In Table 7.3, the base entity type is Village, which does not have a multiplicative factor. The following sections describe the rating engine for motor vehicles, and contractor's equipment, as well as more details of the endorsement factors.

7.2.2 Motor Vehicle

Table 7.4: Rating Engine for Motor Vehicle

Coverage	Formula
PNA (Comprehensive, new, actual cash value)	$\exp[0.118 + 7.975 + 0.755 \times \texttt{lnCoveragePNA}]$
PNR (Comprehensive, new, replacement cost	$\exp[0.118 + 7.975 + 0.755 \times \texttt{lnCoveragePNR}] \times \exp[-0.090 + 0.015]$
POA (Comprehensive, old, actual cash value)	$\exp[-1.201 + 8.326 + 0.715 \times \texttt{lnCoveragePOA}]$
POR (Comprehensive, old, replacement cost)	$\exp[-1.201 + 8.326 + 0.715 \times \texttt{lnCoveragePOR}] \times \exp[-0.198 + 0.287]$
CNA (Collision, new, actual cash value)	$\exp[0.238 + 8.508 + 0.893 \times \texttt{lnCoverageCNA}]$
CNR (Collision, new, replacement cost)	$\exp[0.238 + 8.508 + 0.893 \times \texttt{lnCoverageCNA}] \times \exp[-0.009 + 0.276]$
COA (Collision, old, actual cash value)	$\exp[-0.724 + 8.616 + 0.886 \times \texttt{lnCoverageCOA}]$
COR (Collision, old, replacement cost)	$\exp[-0.724 + 8.616 + 0.886 \times \texttt{lnCoverageCOA}] \times \exp[-0.244 + 1.312]$

7.2.3 Contractor's Equipment

Table 7.5: Rating Engine for Contractor's Equipment

Coverage	Formula
IM (Contractor's Equipment)	$\exp[-1.915 + 9.523 + 0.890 \times \texttt{lnCoverageIM} - 0.098 \times \texttt{lnDeductIM}]$

7.2.4 Endorsements

Table 7.6: Endorsement Factors

Variable	**Value**
Pier & Warf	$\exp[0.082 \times \texttt{SAlnCovRat}]$
Special Animal	$\exp[0.045 \times \texttt{SAlnCovRat}]$
Zoo	$\exp[0.041 \times \texttt{ZOOlnCovRat}]$
Fine Arts	$\exp[0.186 \times \texttt{FAlnCovRat}]$
Golf Course	$\exp[0.070 \times \texttt{GClnCovRat}]$
Business Interruption	$\exp[0.093 \times \texttt{BIlnCovRat}]$
Accounts Receivable	$\exp[0.049 \times \texttt{ARlnCovRat}]$
Money & Securities	$\exp[0.005 \times \texttt{MS1lnCovRat}]$
Money & Securities Limited	$\exp[0.002 \times \texttt{MS2lnCovRat}]$
Vacancy Permit	$\exp[0.042 \times \texttt{VPlnCovRat}]$
Add Ins (other endorsements)	$\exp[0.074 \times \texttt{ADDlnCovRat}]$

Details of Endorsements Ratemaking

Table 7.7 describes endorsements, or optional coverages, that are available to LGPIF policyholders. We do not actually observe claims from an endorsement. For example, if a policyholder purchases a Golf Course Grounds endorsement and has a claim that is from this additional coverage, we are not able to observe this connection with our data. We do observe the additional claim, whether the policyholder has the endorsement, and the amount of coverage under the endorsement. In this sense, endorsements can be treated as another rating variable in our algorithms.

Table 7.7: Description of Endorsements

Variable	Description
Business Interruption	Reimburses an insured for business interruption (lost profits and continuing fixed expenses)
Accounts Receivable	Adds coverage for money owed by its debtors during business interruption due to a covered loss.
Pier and Wharf	Loss of watercraft, by the pressure of ice or water on piers and wharves
Fine Arts	Adds coverage (agreed value) on fine arts, either per item or per exhibit
Golf Course Grounds	Adds coverage to golf course type property such as greens, tees, fairways, etc.
Special Use Animal	Adds coverage for police enforcement animals, such as dogs and horses
Zoo Animals	Adds coverage for zoo animals. Animal mortality is specifically excluded.
Vacancy Permit	Allows claims from covered losses arising from vacant property
Monies and Securities	Adds coverage for monies and securities for loss by theft, disappearance, or destruction (A: loss inside premise, B: loss outside premise).
Monies and Securities (limited term)	Adds limited term coverage for monies and securities
Other Endorsements	Other additional endorsements, including ordinance & law, and extra expenses

Table 7.8 summarizes the claims experience by endorsement. Policyholders with the Zoo Animals endorsement experience an average annual claim frequency of 73.9. Presumably, policyholders paying for this extra protection would enjoy higher property claims and so should be charged additional premiums. The most frequently subscribed endorsement is the Monies & Securities, which covers monetary losses by theft, disappearance, or destruction. The average coverage and number of observations are over five years (2006 $\sim$ 2010), the in-sample period. For example, the Zoo Animals coverage consists of 10 observations over five years and these were from the Henry Vilas Zoo in Dane County and the Milwaukee County Zoo in Milwaukee County.

Table 7.8 shows that a policyholder with any type of endorsement has a higher claims frequency compared to the total of all policyholders. Similarly, for most endorsements, policyholders have a higher average severity, with Pier and Wharf, Monies and Securities (limited term), and Other Endorsements being the exceptions. The effect of higher severity seems to be particularly large for certain endorsements, such as Zoo Animals, Golf Course Grounds and Fine Arts.

Table 7.8: Summary of Claim Frequency and Severity by Endorsement

Endorsements	Number of Observations	Claim Frequency	Average Severity	Average Endorsement Coverage	Spearman Correlation of Coverage with Frequency	Spearman Correlation of Coverage with Severity*
Business Interruption	225	6.427	48,612	2,679,595	0.392	0.249
Accounts Receivable	172	5.285	29,743	853,966	0.188	0.097
Pier and Wharf	312	2.510	24,649	245,445	0.067	0.083
Fine Arts	67	13.493	37,896	12,160,956	0.297	0.187
Golf Course Grouds	28	18.000	20,866	237,500	0.749	0.166
Zoo Animals	10	73.900	18,554	1,102,790	0.877	0.462
Special Use Animal	256	5.547	13,127	21,903	0.168	0.073
Vacancy Permit	225	4.902	21,232	1,779,212	0.053	0.316
Monies and Securities (A,B)	2,137	2.000	29,999	58,928	0.255	0.137
Monies and Securities (limited term)	556	1.739	19,811	416,587	0.143	0.091
Other Endorsements	53	4.906	28,245	4,763,019	-0.003	0.334
All Policies						
Total	5,639	1.109	17,287			

Note: *The severity correlations are based on observations with at least one claim using the average severity (amount divided by frequency).

To help establish the relationship between endorsements and claims outcomes, Table 7.8 also shows the average endorsement coverage (the average is over policyholders with some positive coverage). The table summarizes the Spearman correlation of the amount of endorsement coverage, versus the frequency and severity of claims observed. It is not surprising that all of these correlations are positive, indicating that more coverage means both a higher frequency and severity of claims.

7.3 Comparison of Rating Engine Proposals

Several possible rating engines for the LGPIF building and contents, motor vehicles, and contractor's equipment coverages were proposed during the ratemaking project. This section summarizes the projected premiums collected under various proposed engines. The premium amounts collected during the 2011-2013 validation years, and the loss ratios under each proposed rating engine is summarized in Tables 7.9 and 7.10.

For details of the rating engines, the reader may refer to Frees and Lee (2017), where shrinkage estimation is utilized to regulate the regression coefficients for endorsements. More details of each of the rating engine proposals can be found on the LGPIF Project Website. URL: `https://sites.google.com/a/wisc.edu/local-government-property-insurance-fund`. In order for the rating engine to achieve balanced goals, `ProposalG` has been recommended as the final rating engine. This is the version documented in Section 7.2.

Table 7.9: Summary of BC Rating Engine Proposals, for the Validation Years 2011-2013

Building and Contents, Endorsements, 2011-2013							
Rating Agency Premium and Actual Claims		2011	2012	2013	**Total**	**Loss Ratio (2011-13)**	**Loss Ratio (2011-12)**
Claim	Actual BC claims	22,326,420	21,810,101	54,610,875	**98,747,396**		
Premium	Rating Agency Premium	16,779,794	19,216,683	23,466,925	**59,463,402**	**166%**	**123%**
Proposals, with basic severity part		2011	2012	2013	**Total**	**Loss Ratio (2011-13)**	**Loss Ratio (2011-12)**
ProposalA	Entity, Endorsements no-shrink	21,008,944	22,030,612	21,675,099	**64,714,655**	**153%**	**103%**
ProposalB	Endorsements, no-shrink	20,035,598	20,612,146	20,031,720	**60,679,464**	**163%**	**109%**
ProposalC	Entity, County Code, Endorsements, no-shrink	24,190,064	26,061,778	26,177,166	**76,429,008**	**129%**	**88%**
ProposalD	County Code, Endorsements, no-shrink	22,488,344	23,784,246	23,800,179	**70,072,769**	**141%**	**95%**
ProposalE	Shrink Proposal B ($\lambda = 5,090$)	20,623,223	19,674,382	18,880,691	**59,178,296**	**167%**	**110%**
ProposalF	Shrink Proposal B ($\lambda = 840$)	18,192,018	17,801,528	17,091,870	**53,085,416**	**186%**	**123%**
ProposalZ	Coverage, lnDeduct, NoClaimCredit only	19,048,028	19,867,143	19,018,469	**57,933,640**	**170%**	**113%**
ProposalY	Include Fire5 Variable	20,994,111	21,872,717	21,559,197	**64,426,025**	**153%**	**103%**
ProposalW	Include Fire5 and Interaction Terms	20,182,663	21,404,551	21,565,588	**63,152,802**	**156%**	**106%**
ProposalH	Shrink Proposal Y ($\lambda = 1,410$)	20,226,368	20,963,069	20,278,499	**61,467,936**	**161%**	**107%**
ProposalG	Shrink Proposal A ($\lambda = 1,520$)	20,613,830	20,568,182	20,405,217	**61,587,229**	**160%**	**107%**
Proposals, including County Codes in severity part		2011	2012	2013	**Total**	**Loss Ratio (2011-13)**	**Loss Ratio (2011-12)**
ProposalNewA	Entity, Endorsements no-shrink	24,361,156	25,127,193	24,270,236	**73,758,585**	**134%**	**89%**
ProposalNewB	Endorsements, no-shrink	23,167,688	23,654,059	22,409,073	**69,230,820**	**143%**	**94%**
ProposalNewC	Entity, County Code, Endorsements, no-shrink	25,298,499	26,322,859	26,312,040	**77,933,398**	**127%**	**86%**
ProposalNewD	County Code, Endorsements, no-shrink	22,663,382	23,352,261	23,056,013	**69,071,656**	**143%**	**96%**
ProposalNewE	Shrink Proposal B ($\lambda = 5,090$)	19,335,968	18,577,296	17,642,352	**55,555,616**	**178%**	**116%**
ProposalNewF	Shrink Proposal B ($\lambda = 840$)	18,044,740	17,750,364	16,820,564	**52,615,668**	**188%**	**123%**
ProposalNewZ	Coverage, lnDeduct, NoClaimCredit only	22,136,500	22,878,477	20,914,215	**65,929,192**	**150%**	**98%**
ProposalNewY	Include Fire5 Variable	24,385,402	25,073,534	24,248,079	**73,707,015**	**134%**	**89%**
ProposalNewW	Include Fire5 and Interaction Terms	21,410,557	22,115,281	22,341,415	**65,867,253**	**150%**	**101%**
ProposalNewH	Shrink Proposal Y ($\lambda = 1,410$)	23,636,524	24,315,555	22,703,826	**70,655,905**	**140%**	**92%**
ProposalNewG	Shrink Proposal A ($\lambda = 1,520$)	22,286,825	22,328,960	21,888,652	**66,504,437**	**148%**	**99%**
Percentage collected as Endorsement		2011	2012	2013	**Total**	**% of BC Rate**	
Rating Agency		428,582	428,480	459,728	**1,316,790**	**2.28%**	
ProposalG		630,373	612,937	599,784	**1,843,094**	**2.99%**	
ProposalNewG		1,203,646	1,217,098	1,097,528	**3,518,272**	**5.29%**	

Table 7.10: Summary of MV, IM rating Engine Proposals, for the Validation Years 2011-2013

Motor Vehicles, Inland Marine (Equipments), 2011-2013						
Motor Vehicles and Equipment, Actual Claims		2011	2012	2013	**Total**	
PN	Comprehensive New	696,638	450,920	472,960	**1,620,519**	
PO	Comprehensive Old	342,529	769,275	1,530,055	**2,641,859**	
CN	Collision New	519,899	806,637	648,939	**1,975,475**	
CO	Collision Old	720,337	847,595	1,209,702	**2,777,634**	
IM	Inland Marine (equipments)	708,489	1,298,880	798,989	**2,806,358**	
Total (Motor Vehicles and Equipment)		**2,987,891**	**4,173,308**	**4,660,645**	**11,821,844**	
Rating Agency Premiums		2011	2012	2013	**Total**	**Loss Ratio**
PN	Comprehensive New	225,914	216,054	221,501	**663,469**	**244%**
PO	Comprehensive Old	984,117	982,397	1,036,699	**3,003,213**	**88%**
CN	Collision New	514,000	525,083	474,612	**1,513,695**	**131%**
CO	Collision Old	2,269,482	2,292,351	2,338,022	**6,899,855**	**40%**
IM	Inland Marine (equipments)	2,037,347	1,959,159	1,599,688	**5,596,194**	**50%**
Total (Motor Vehicles and Equipment)		**6,030,860**	**5,975,044**	**5,670,522**	**17,676,426**	**67%**
Proposals, basic		2011	2012	2013	**Total**	**Loss Ratio**
PN	Comprehensive New	488,951	487,691	524,768	**1,501,410**	**108%**
PO	Comprehensive Old	572,159	573,495	574,062	**1,719,717**	**154%**
CN	Collision New	731,037	743,717	693,730	**2,168,484**	**91%**
CO	Collision Old	1,167,161	1,194,424	1,219,265	**3,580,850**	**78%**
IM	Inland Marine (equipments)	897,386	913,853	933,368	**2,744,607**	**102%**
Total (Motor Vehicles and Equipment)		**3,856,694**	**3,913,180**	**3,945,194**	**11,715,068**	**101%**
Proposals, using EntityType, CountyCodes, and RC indicator		2011	2012	2013	**Total**	**Loss Ratio**
PN	Comprehensive New	548,385	547,517	534,383	**1,630,285**	**99%**
PO	Comprehensive Old	1,039,040	984,426	892,854	**2,916,320**	**91%**
CN	Collision New	816,755	824,158	765,350	**2,406,263**	**82%**
CO	Collision Old	1,276,916	1,328,641	1,395,729	**4,001,286**	**69%**
IM	Inland Marine (equipments)	906,691	922,785	943,863	**2,773,339**	**101%**
		2011	2012	2013	**Total**	**Loss Ratio**
Total (Motor Vehicles and Equipment)		**4,587,787**	**4,607,528**	**4,532,179**	**13,727,493**	**86%**