Essays on Claims Modelling, Customer Churn Analysis, and Customer Valuation in General Insurance

Yumo Dong

Supervisors: Dr. Fei Huang, Dr. Francis Hui, Prof. Edward (Jed) Frees, and Dr. Honglin Yu

A thesis submitted for the degree of

Doctor of Philosophy

The Australian National University

Research School of Finance, Actuarial Studies & Statistics

Australian National University

February 2024

© Copyright by Yumo Dong, 2024

Spend your life in your own way.

Disclaimer

The work in this thesis is my own except where otherwise stated.

Yumo Dong

Acknowledgements

I have spent five wonderful years as a PhD student at the Australian National University. At the end of this memorable journey, I would like to express my sincere thanks to the people who have supported and helped me.

First and foremost, I would like to thank my supervisory panels, Dr Fei Huang, Dr Francis Hui, Prof Edward (Jed) Frees, and Dr Honglin Yu, for training me to be an independent scholar. I want to thank Fei for taking care of me along the journey and guiding me with great patience. Fei taught me not only extensive knowledge but also to be confident, humble, and detail-oriented. I am grateful for working with Jed, who always gives me forward-looking advice and exposes me to state-of-the-art methodologies. A heartfelt thanks to Francis, a brilliant statistician who taught me to be a "non-significant" perfectionist in doing research. I also want to thank Honglin and Prof Harald van Heerde, who gave invaluable comments on my thesis. These people mentioned above have served as examples to me, and I see in them what I want to be.

It has been a pleasure to work with academic and administrative staff at RSFAS. Special thanks go to Dr Tim Higgins for kindly helping me with all PhD-related matters. I also want to thank my PhD colleagues, Macro Li, Nickson Ning, and Xi Xin, who helped me a lot in doing research. I really enjoy discussing questions with them and wish them all the best with their academic endeavours.

Last but not least, I owe a big thank you to my wife Xiaoxue Zhao and my puppy Blue, for accompanying me through my PhD journey in Canberra. Without their support, it would be very hard for me to finish my thesis. I would like to thank my parents for supporting my academic pursuits emotionally and financially.

Abstract

This thesis presents three essays themed around the topics of claims modelling, customer churn analysis, and customer valuation in general insurance. Chapter 2 (essay) studies multivariate claims modelling with combined deductibles, Chapter 3 explores multi-state customer churn analysis, and Chapter 4 focuses on cost-based customer lifetime value (CLV) models. While each chapter offers a unique and innovative contribution to the existing literature on the aforementioned topics, the research problems tackled in all three chapters are of broad and referential significance to general insurance companies.

In Chapter 2, we study the dependence modelling of multivariate insurance claims with the consideration of a combined deductible. Without combined deductibles, it is straightforward to value bundled insurance contracts as the sum of contracts from each supporting coverage. However, with combined deductibles, assumptions about the relationships among coverages become important. To demonstrate this importance, we evaluate the cost of combined deductibles under dependence via copulas, and show the impact of dependence modelling using three applications (commercial insurance, reinsurance, and personal insurance). This chapter also examines a range of dependence parameters, marginal distributions, and copulas in the simulation study, and supplements the findings from the simulation study with several theoretical propositions.

Chapter 3 explores multi-state customer churn analysis in general insurance. Traditionally, customer churn analyses have employed models that utilise only a binary outcome (churn or not churn) in one period. However, real business relationships are multi-period, and policyholders may reside and transition between a wider range of states beyond that of the simply churn/not churn throughout this relationship. To better understand policyholder behaviours in real business relationships, we propose a multi-state customer churn analysis that aims to model customer behaviour over a larger number of states (defined by different combinations of insurance policies taken) and across multiple periods. Using multinomial logistic regression (MLR) with a second-order Markov assumption, we demonstrate how a policyholder's transition history is associated with their decision-making, whether that be to retain the current set of policies, churn, or to add/drop a policy or coverage.

In Chapter 4, we introduce a CLV modelling framework suitable for the insurance industry. Traditional CLV models focus on modelling revenues and churn but often neglect modelling costs to serve customers. In the insurance industry, a major cost component is insurance claims, which are highly variable and can be extremely large

in size. Existing CLV models are not suitable for the insurance industry because they do not capture costs adequately and ignore the association between costs (claims) and the likelihood to churn. To bridge this gap, we introduce a novel cost-based CLV modelling framework. Utilising a Tweedie regression approach for claims, combined with shared random effects between the claims and churn processes, our model successfully captures unobserved and correlated heterogeneity in claims and churn risks across customers. Applying our proposed model to commercial insurance data, we illustrate how this model can capture latent associations between claims and churn tendencies, enhance predictive CLV performance, and bring novel managerial insights into CLV analyses.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Claims modelling (supply-side) and customer churn analysis (demand-side) are two key aspects of general insurance ratemaking, ensuring the equilibrium of fair pricing for customers and sustainable profitability for insurance businesses. More importantly, claims modelling and customer churn analysis are fundamental techniques that insurers use to value their customers. This thesis presents three essays on several important technical questions that insurers may encounter when building claims models, conducting customer churn analyses, and valuing customers. Specifically, Chapter 2 or essay studies multivariate claims modelling with combined deductibles, Chapter 3 explores multi-state customer churn analysis, and Chapter 4 proposes a joint model of claims and customer churn to estimate customer value. Each chapter makes a unique contribution to the existing literature and offers broad application prospects. Collectively, they form a thesis of both methodological and practical value to general insurance.

In Chapter 2, we study the dependent modelling of multivariate insurance claims with the consideration of a combined deductible. Bundled insurance contracts, providing protection based on several loss coverages, are attractive because they allow insurers to focus on customers' needs (Frees et al. 2016; Yang et al. 2020; Dong, Frees, Huang and Hui 2022). A common contract feature is a deductible set such that the insurer pays the excess over the deductible of the sum of losses from all coverages; we refer to this as a combined deductible. Without combined deductibles, it is straightforward to value a customer's bundled insurance contract as the sum of contracts from each supporting coverage. However, with combined deductibles, assumptions about the relationships among coverages become important, and these dependence relationships may affect the cost to customers through claims modelling (Bäuerle and Müller 1998; Denuit et al. 2006). To demonstrate this importance, this chapter evaluates the cost of combined deductibles under dependence via copulas. In particular, we show the impact of dependence modelling using three applications in the areas of commercial insurance, reinsurance, and personal insurance.

A sensitivity analysis using simulations further provides a sharper perspective on the importance of dependent relationships. This chapter supplements the findings from the empirical study with several theoretical propositions demonstrating how dependence affects the insurance pricing of combined deductibles.

Chapter 3 explores multi-state customer churn analysis in general insurance (Dong, Frees, Huang and Hui 2022). Customer churn (also known as customer attrition, dropout, turnover, or defection) refers to the non-renewal of existing customers for an insurance company. In general insurance, customer churn poses a major challenge to insurers because contracts are renewed periodically (commonly after one year), and moreover with substantial competition and diverse contracts, it is easy for customers to switch insurance providers to satisfy their needs at the time (Brant Carson 2017). Traditionally, customer churn analyses have employed models that utilise only a binary outcome (churn or not churn) in one period (Guillen et al. 2003; Günther et al. 2014; Staudt and Wagner 2018; Frees et al. 2021). However, real business relationships are multi-period, and policyholders may reside and transition between a wider range of states beyond that of the simple churn/not churn throughout this multi-period relationship. To better encapsulate the richness of policyholder behaviours through time, we propose a multi-state customer churn analysis, which aims to model behaviour over a larger number of states (defined by different combinations of insurance coverage taken) and across multiple periods (thereby making use of readily available longitudinal data). Using multinomial logistic regression (MLR) with a second-order Markov assumption in the linear predictor, we demonstrate how multi-state customer churn analysis offers deeper insights into how a policyholder's transition history is associated with their decision-making, whether that be to retain the current set of policies, churn, or add/drop a coverage. Applying this model to commercial insurance data from the Wisconsin Local Government Property Insurance Fund, we illustrate how transition probabilities between states are affected by differing sets of explanatory variables, and how a multi-state analysis can potentially offer stronger predictive performance and more accurate calculations of customer lifetime value compared to the traditional customer churn analysis techniques.

While Chapters 2 and 3 respectively focus on claims modelling and customer churn analysis, Chapter 4 formally introduces an innovative customer lifetime value (CLV) modelling framework suitable for the insurance industry. CLV quantifies the present value of projected profits that a firm expects from a customer over the course of their relationship. It takes projected revenues, costs, relationship duration (retention), and discount rate into consideration. While traditional CLV models focus on mod-

elling revenues and churn, they often neglect modelling costs to serve customers (see Venkatesan and Kumar 2004; Kumar et al. 2008; Ascarza et al. 2018, among others). In many industries, these costs are low and/or steady. But in the insurance industry a major cost component is insurance claims, which are unknown at the point of sale, highly variable, and can be extremely large in size (Donkers et al. 2007). As such, existing CLV models are not suitable for the insurance industry because they do not capture costs adequately. Moreover, existing models do not consider the association between costs (claims) and the likelihood to churn – this is an oversight since a policyholder who has a high intrinsic claim risk may exhibit lower churn propensity. To bridge this gap, we introduce a new cost-based CLV modelling framework. Utilising a Tweedie regression approach for claims along with shared random effects between the claims and churn processes, our model is able to adequately capture unobserved and correlated heterogeneity in claims and churn risks across customers, which is ignored in Chapters 2 and 3. Applying our proposed model to commercial insurance data, we illustrate how this model can capture latent associations between claims and churn tendencies, enhance predictive CLV performance, and bring novel managerial insights into CLV analyses.

Chapter 5 concludes the findings from Chapters 2 to 4, and discusses the scope for future research.

Chapter 2

Deductible Costs for Bundled Insurance Contracts

Yumo Dong, Edward W (Jed) Frees, Fei Huang

This chapter has a working paper version available on SSRN or go to the next url: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4020299`.

2.1 Introduction

A deductible, a specific amount that a policyholder must spend before the insurance company starts to pay the costs, is a common feature of insurance contracts. Deductibles reduce premiums for the policyholders by eliminating a large number of small claims, the costs associated with handling these claims, and the potential moral hazard arising from having insurance; see, for example, Wang et al. (2008). In effect, a deductible aligns insurers' and policyholders' interests so that both parties seek to mitigate the risk of losses.

Traditionally, insurers offered a single contract for a single type of risk, starting with coverages for caravans in ancient China and Mesopotamia. As insurance has evolved, insurers have become more customer-centric and offer to cover multiple risks under a single contract. This includes not only the so-called "all perils" policy under a homeowners insurance policy (covering several causes of loss) but also several risks that have traditionally been treated separately. One example is the "bundling" of auto and homeowners, which has become commonplace in the personal insurance market. For another example, within the U.S. healthcare industry (Barnett and Vornovitsky 2016), of the population with health insurance coverage in 2015, 78.4 per cent had one coverage while 21.6 per cent had multiple coverages. A traditional insurance contract only contains one coverage, whereas, in contrast, a bundled contract contains multiple coverages simultaneously. Insurers could reduce administrative costs, which is reflected in their pricing, by offering bundled coverages. Also, there are marketing advantages, customer retention benefits, and so forth, associated with bundling for insurers. Historically, insurers have been attracting customers to buy multiple coverages through the so-called "bundling" discount. This "bundling" discount is outside of our efforts because we focus on the "technical" costs and prices in this chapter. When considering bundled insurance, it is natural to think about commercial, reinsurance, and personal insurance. Table 2.1 illustrates a few examples from personal insurance. For example, a policyholder can buy a customised travel insurance contract that includes both loss and damage of luggage as well as personal injury coverage. Then, for a claim made by a policyholder, the insurer may simultaneously observe two outcomes, one from each coverage.

Historically, different risk types have been treated separately, leading to two important implications. First, it means that insurers are likely to retain data on each type of claim outcome even when bundled into a single insurance contract. Because the data are retained, it is natural to model their behaviour. Second, when risks are insured under different contracts, it is natural to model them separately, assuming no relationships between them, which essentially means treating them as independent.

Table 2.1: Personal bundled insurance contracts with combined deductibles.

Contract Name	**Coverages**	**Combined Deductibles**
Comprehensive Pet Plan	Accident, Illness	\$200 to \$1000
Health Insurance Hospital	Medical Treatments, Health Services, Overnight Accommodation, Nursing Care, etc	\$250 to \$750
Home Buildings & Contents	Buildings, Contents	\$0 to \$5000
Travel Insurance	Medical Expenses, Personal Baggage	\$500 to \$2000

In contrast, when several outcomes are included within a single contract, it is more natural to be concerned with their dependence. Broadly speaking, it is known that dependencies among interrelated outcomes can be important for quantifying the financial implications of insurance contracts (see, for example, Yang et al. (2020)). In this chapter, we consider the financial impact of deductible costs of a bundled insurance contract. We explicitly investigate the effect of dependencies on deductible costs through the analysis of applications in commercial insurance, reinsurance, and personal insurance.

An interesting feature that often appears in bundled contracts is a deductible that covers multiple coverages. There are two kinds of deductibles in a bundled insurance contract that we call "split" and "combined". A split deductible is applied to an individual coverage or risk type, and a combined deductible is applied to the sum of multiple coverages in a bundled insurance contract. In practice, a combined deductible is widely used to define the potential costs undertaken by different parties involved in a bundled contract. Take the travel insurance with two coverages discussed earlier as an example. We define X_1 as the loss of luggage damage and X_2 as the loss of personal injury. Then, we could use $S_1 = (X_1 - d_1)_+$ to define the reimbursement covering the loss of luggage damage from an insurer. Here, d_1 is a split deductible as it applies only to luggage damage. In the same way, we define $S_2 = (X_2 - d_2)_+$ with d_2 being the split deductible that applies only to personal injury coverage. In this case, the pure premium of an insurance contract paid by a policyholder to an insurer is given by

$$E(S_1 + S_2) = E(X_1 - d_1)_+ + E(X_2 - d_2)_+. \tag{2.1}$$

When there is a combined deductible d applying to the two risks X_1 and X_2, the bundled contract provides reimbursement in the amount of $(X_1 + X_2 - d)_+$ where d is the "combined deductible". The expected value of $(X_1 + X_2)$ is given by

$$E(X_1 + X_2) = E(X_1 + X_2 - d)_+ + E\min(X_1 + X_2, d), \tag{2.2}$$

where $E(X_1 + X_2 - d)_+$ is the expected loss undertaken by the insurer (pure premium), and $E\min(X_1 + X_2, d)$ is the expected loss undertaken by the policyholder, which we call the deductible cost. When two split deductibles and a combined deductible are applied together, the pure premium is $E(S_1 + S_2 - d)_+$, and $\big(E(X_1+X_2)-E(S_1+S_2-d)_+\big)$ is the expected loss undertaken by the policyholder according to equations (2.1) and (2.2).

As noted above, our benchmark model assumes independence among the losses of different coverages when valuing bundled insurance contracts. The assumption of independence works well when only split deductibles are applied in bundled insurance contracts. For the travel insurance example, when two split deductibles are applied to each coverage respectively, the dependent and independent assumptions of X_1 and X_2 have the same pure premiums $E(S_1 + S_2)$; see equation (2.1). Therefore, when only split deductibles are used, the dependence does not matter for the pricing purpose. Hence, we will not consider split deductibles in this chapter. A standard reference (such as Brown and Lennox 2015; Lee 2017) on general insurance deductible pricing can be applied to each coverage in bundled contracts with only split deductibles.

The lack of independence can be important in the presence of combined deductibles. In this chapter, we identify situations in which the independent assumption does not work well and introduce improved pricing models using copulas. For the bundled travel insurance example, when there is a combined deductible d in the contract; see equation (2.2), the dependence between X_1 and X_2 will lead to a different pure premium $E(X_1 + X_2 - d)_+$ compared to using the independent assumption of X_1 and X_2. In a pet insurance example in Section 2.4, the calculated deductible costs are reduced by approximately 10 to 23 per cent when dependence models are used, compared to the benchmark models that assume independence.

The remainder of this chapter is organised as follows. Following a short literature review in Section 2.2 on dependence modelling in insurance pricing, in Section 2.3 we illustrate how to calculate deductible costs for dependence models analytically and get an approximation result through Monte Carlo simulation. In Section 2.4, we examine three practical applications (commercial insurance, reinsurance, and personal insurance) in which combined deductibles are regularly utilised. We find that an adjustment for dependence among coverages leads to materially different prices compared to the benchmark models that assume independence. The contracts studied in Section 2.4 focus on specific situations, so in Section 2.5, we describe sensitivity results using simulations. We find that the choices of dependence parameters, marginal distributions, and copulas all affect the deductible cost. Taking advantage of the

well-known literature on stop-loss premiums, Section 2.6 presents general properties of joint distributions that can lead to materially different prices and shows how the dependence structure affects insurance pricing of combined deductibles. Section 2.7 offers some concluding remarks.

2.2 Literature Review

Extensive literature has been devoted to problems related to split deductible ratemaking. Lee (2017) discussed split deductible ratemaking considering both the regression approach and the maximum likelihood approach under the general insurance context. Ng et al. (2019) found that using the untransformed deductible as a covariate in a generalised linear model could improve accuracy in predicting relativities. However, to the best of our knowledge, there has been limited research in the literature considering the combined deductible ratemaking of bundled insurance contracts. Compared to split deductible ratemaking, combined deductible ratemaking requires the dependence of multiple coverages to be taken into account. To show the importance of dependence for combined deductible ratemaking, we compare the deductible costs of copula models with those of benchmark models that assume independence, comonotonicity, and countermonotonicity. This chapter reduces the gap by investigating the combined deductible ratemaking from the perspectives of empirical analysis, simulations, and theoretical analysis.

Dependence modelling is important in the presence of combined deductibles, and we apply copulas in this chapter to model the dependence of multiple risks. The concept of copulas, a tool for understanding relationships among multivariate outcomes, was first introduced by Sklar (1959) in the context of probabilistic metric spaces. Frees and Valdez (1998) introduced actuaries to copulas by exploring some practical applications and describing basic properties, after which copulas have been widely used for dependence modelling in various actuarial applications. For example, Frees and Wang (2005) and Frees and Wang (2006) developed a link between credibility and loss distributions through the notion of a copula, and showed the importance of dependence in credibility ratemaking. Shi and Frees (2011) proposed a copula regression model for dependent lines of business that can be used to predict unpaid losses.

The stop-loss premium, a concept closely related to deductible costs, has been studied extensively in the insurance literature and can be used in both stop-loss reinsurance contracts (between an insurer and a reinsurer) and insurance contracts with combined deductibles (between a policyholder and an insurer). Dhaene and

Goovaerts (1996) and Dhaene and Goovaerts (1997) investigated how stop-loss premiums are influenced by changing the dependency assumptions between different risks. Müller (1997) showed that supermodular ordering implied the stop-loss order of the portfolios. Denuit et al. (2001) showed that the sum of risks exhibiting a weak form of positive dependence was larger in convex order than the corresponding sum under the theoretical independence assumption. The literature above contributes to the theoretical foundation of the stop-loss premium, which can be extended to the deductible cost. Mathematically, the stop-loss premiums for reinsurance contracts and insurance contracts can be formulated and calculated in the same way. However, in practice, a combined deductible is usually placed at the right tail of a loss distribution for reinsurance contracts while placed at the left tail of a loss distribution for insurance contracts. In this chapter, we will take advantage of the well-known literature on stop-loss premiums and provide several propositions on the impact of dependence on deductible costs.

2.3 Calculating Deductible Costs

This section aims to formalise the modelling framework for calculating deductible costs. We consider an insurance (reinsurance) contract with K coverages and a combined deductible d applying to all the coverages. Let X_k $(k = 1, ..., K)$ denote the loss of each coverage, and S denote the aggregate loss $S = \sum_{k=1}^{K} X_k$. We start with the relationship including the combined deductible d,

$$E(S) = E(S - d)_+ + E\min(S, d). \tag{2.3}$$

Theoretically, equation (2.3) can be applied to both insurance and reinsurance contracts. We explain this equation in the following two scenarios.

In an insurance contract, $E(S - d)_+$ is the pure premium paid by a policyholder to an insurer, and $E\min(S, d)$ is the expected loss undertaken by the policyholder. In this case, d is mostly placed at the left tail of the distribution of the aggregate loss S. When the aggregate loss S is smaller than d, the policyholder is responsible for paying all the loss; otherwise, the policyholder pays the amount of loss d, and the insurer pays the excess loss $S - d$.

In a reinsurance contract, $E(S - d)_+$ is called the stop-loss premium, which is the premium paid by an insurer to a reinsurer, and $E\min(S, d)$ is the expected loss undertaken by the insurer. In this case, d is mostly placed at the right tail of the distribution of the aggregate loss S. For example, an insurer buys stop-loss reinsurance for its portfolio to protect against large claims. When the aggregate

loss S of the insurer's portfolio is smaller than d, the insurer pays all the loss; otherwise, the insurer pays the amount of loss d, and the reinsurer pays the excess loss $S - d$.

In this chapter, we focus on $E\min(S, d)$, which we call the deductible cost,

$$c_d = E\min(S, d) = \int_0^d 1 - F_S(s)\ ds. \tag{2.4}$$

The deductible cost c_d depends on the cumulative distribution function of aggregate loss $F_S(s)$ as shown in equation (2.4), which further depends on the joint distribution of $X_1, X_2, \ldots X_k$, hence the dependence among the losses of K coverages. In this chapter, we use copulas to model the dependence of multiple risks. For simplicity and illustration purposes, we consider a case of $K = 2$ as an example, and illustrate how a copula captures the dependence between the bivariate losses (X_1, X_2). The copula of (X_1, X_2) is defined as the joint cumulative distribution function of (U_1, U_2),

$$C(u_1, u_2) = Pr(U_1 \leq u_1, U_2 \leq u_2),$$

where the random vector $(U_1, U_2) = \big(F_1(x_1), F_2(x_2)\big)$ has marginals that are uniformly distributed on the interval $[0, 1]$. According to Sklar's theorem (Sklar 1959), the joint cumulative distribution function of (X_1, X_2) can be expressed in terms of its marginals $\big(F_1(x_1), F_2(x_2)\big)$ and a copula C,

$$Pr(X_1 \leq x_1, X_2 \leq x_2) = C\big(F_1(x_1), F_2(x_2)\big). \tag{2.5}$$

Differentiating both sides of equation (2.5) with respect to x_1 and x_2, we can obtain the joint density distribution function of (X_1, X_2):

$$f_{1,2}(x_1, x_2) = c\big(F_1(x_1), F_2(x_2)\big) \cdot f_1(x_1) \cdot f_2(x_2),$$

where $c\big(F_1(x_1), F_2(x_2)\big)$ is the density of the copula. Then the joint density distribution function can be used to calculate $F_S(s)$ in equation (2.4) following the convolutional formula:

$$F_S(s) = \textstyle\int_0^\infty \int_0^{s-x_1} f_{1,2}(x_1, x_2) dx_2\ dx_1.$$

However, it quickly becomes difficult to calculate the deductible cost analytically through the convolutional formula when $K > 2$. To approximate deductible costs accurately and consistently, we use Monte Carlo simulation to calculate the deductible costs for dependence and independence models for $K \geq 2$ in this chapter.

Given the copula and the marginal distributions of (X_1, X_2), we can use the Monte Carlo approach to calculate deductible costs in the following steps:

1. We simulate R bivariate samples $(u_{11}, u_{21}),\ldots,(u_{1R}, v_{2R})$ from the bivariate copula $C(u_1, u_2)$, where R is the number of the simulation that is determined using the method described in Appendix A.7. The function that we use to simulate samples is `rCopula` (Hofert et al. 2020) in R.

2. Given the two marginal distributions of the losses, we can calculate their inverse distribution functions $F_1^{-1}(u_1)$ and $F_2^{-1}(u_2)$. Using these two inverse distribution functions and R bivariate samples generated in Step 1, we can generate R samples of the bivariate losses

$$(X_{11}, X_{21}), \ldots, (X_{1R}, X_{2R}) = \left(F_1^{-1}(u_{11}), F_2^{-1}u_{21})\right), \ldots, \left(F_1^{-1}(u_{1R}), F_2^{-1}(u_{2R})\right).$$

3. For any positive combined deductible d, the deductible cost based on a dependence model (c_d^{dep}) is calculated by

$$c_d^{dep} = \frac{\sum_{r=1}^{R} \min(X_{1r} + X_{2r}, d)}{R}.$$

For the deductible cost based on an independence model (c_d^{ind}), we use a Gaussian copula with $\rho_S = 0$ (Spearman's rank correlation coefficient) in Step 1 to simulate samples, then repeat Step 2 to get R samples of the independent bivariate losses, and repeat Step 3 to get c_d^{ind}.

To quantify the impact of dependence modelling on deductible costs, we define the cost relativity (CR) as

$$CR = \frac{c_d^{dep}}{c_d^{ind}} - 1. \tag{2.6}$$

The larger is $|CR|$, the bigger is the difference in deductible costs between the two models. In Section 2.4, we will study CR in three applications (commercial insurance, reinsurance, and personal insurance) to analyse the impact of dependence modelling on deductible costs.

The concepts of comonotonicity and countermonotonicity have been widely studied in economics, financial mathematics, and actuarial science; see Deelstra et al. (2011). Comonotonicity corresponds to the perfect positive dependence structure, while countermonotonicity corresponds to the perfect negative dependence structure. Comonotonicity/countermonotonicity is especially informative and useful to insurers in making conservative estimates about the risks and calculating premi-

ums. Hence in equation (2.6), we can also replace c_d^{ind} by c_d^{co} for the deductible cost based on comonotonic scenario and $c_d^{counter}$ for the deductible cost based on countermonotonic scenario as additional benchmarks to understand the cost relativities. For readers interested in the comonotonic and countermonotonic scenarios, we provide results of three applications (commercial insurance, reinsurance, and personal insurance) in Appendix A.5.

2.4 Applications of Deductible Costs

In this section, we investigate the importance of dependence modelling for deductible costs through three insurance applications: commercial insurance, reinsurance, and personal insurance.

2.4.1 Application 1: Wisconsin Property Fund (Commercial)

This application is based on the application in Frees et al. (2016) using a data set from the Wisconsin Local Government Property Insurance Fund (LGPIF). The bundled contracts in the data set contain at most six coverages. In this chapter, we only consider those contracts with all six coverages, including building and content (BC), contractor's equipment (IM), comprehensive new vehicles (PN), comprehensive old vehicles (PO), collision new vehicles (CN), and collision old vehicles (CO). Table 2.5 shows descriptive statistics for the claims of six coverages.

Table 2.2: Descriptive statistics for the claims of six coverages.

Variables	Min	Median	Mean	Max	SD
BC	0	0	17,195	12,922,218	230,009
IM	0	0	941	485,485	10,842
PN	0	0	1,613	147,534	6,356
PO	0	0	1,337	342,242	11,213
CN	0	0	2,760	129,783	9,890
CO	0	0	2,676	303,289	14,559

The marginal distribution of the loss for each coverage is modelled by a Tweedie distribution, and all marginal distributions are joined by a Gaussian copula. The Tweedie distribution is defined as a Poisson sum of gamma random variables. For information on the estimated marginal coefficients of the Tweedie model for each coverage, we refer to Table A11 in Frees et al. (2016). Table 2.3 shows the estimated dependence parameters of the Gaussian copula for different coverages with Tweedie marginal distributions.

Note that there is no combined deductible in the original bundled contracts, but we can still investigate the impact of dependence modelling on the deductible cost for risk managers interested in introducing this popular feature. The simplest way to introduce the combined deductible is by offering a combined deductible d to all six coverages. We choose to study this combined deductible on all six coverages for illustration purposes in this chapter. However, for the purpose of the application, insurers can flexibly offer one or more combined deductibles to different combinations of coverages in a bundled insurance contract. For example, an insurer can offer a combined deductible d_{auto} to the four auto coverages (PN, PO, CN, CO), and a combined deductible d to the whole contract. In this case, the insurer should cover the loss of $\left[X_{BC}+X_{IM}+(X_{PN}+X_{PO}+X_{CN}+X_{CO}-d_{auto})_+-d\right]_+$. This example of two combined deductibles is studied in Appendix A.1 for interested readers.

We henceforth assume there is a combined deductible d on the aggregate losses of six coverages. Considering the nature of these commercial contracts, we set the upper bound of the combined deductible to be \$10000. A higher upper bound makes it more flexible for commercial policyholders to customise their contracts with insurers (Frees et al. 2022).

Table 2.3: 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

In this case, we write the deductible cost as

$$c_d = E\ \min(X_{BC}+X_{IM}+X_{PN}+X_{PO}+X_{CN}+X_{CO}, d),$$

where X_k is the loss of k-th coverage. It means that the insurer agrees to pay for the excess claim over the combined deductible d.

Figure 2.1 displays CR of twenty-four policyholders, randomly chosen from all 244 policyholders who have six coverages in 2011. When the combined deductible d is no more than \$10000, the CR of each policyholder is negative. We further explain why $\lim_{d\to 0^+} CR$ is negative in Appendix A.2 for interested readers. If we allow the combined deductible d to approach infinity, then the CR will tend to zero as $E\ \min(X_{BC}+X_{IM}+X_{PN}+X_{PO}+X_{CN}+X_{CO}, \infty) = E(X_{BC})+E(X_{IM})+E(X_{PN})+E(X_{PO})+E(X_{CN})+E(X_{CO})$. In this application, the deductible costs are reduced

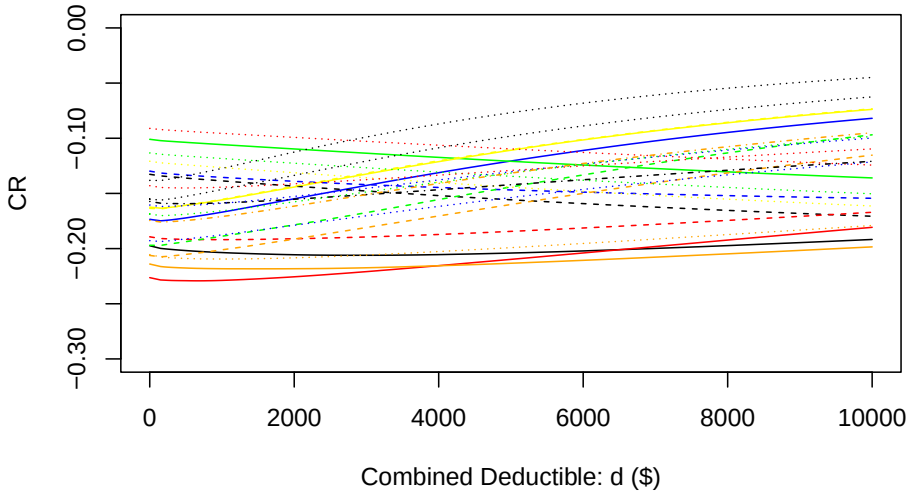

Figure 2.1: Cost Relativities of twenty-four policyholders from LGPIF (each curve represents CR of an individual policyholder).

by approximately 6 to 23 per cent when dependence models are used, compared to the benchmark models that assume independence.

We select four policyholders from the twenty-four policyholders who are most affected by the dependence modelling as examples. Their policy IDs are 140430, 160934, 160730, and 133640 (the bottom four curves in Figure 2.1). Figure 2.2 shows the premiums (left panel) and deductible costs (right panel) of these four policyholders. We can see that deductible costs increase with the increase of combined deductibles. We can also see that the premiums calculated by independence models are smaller than those calculated by dependence models, while the deductible costs calculated by independence models are larger than those calculated by dependence models. This explains the negative CRs we observed in Figure 2.1. We will further explain these observations using propositions in Section 2.6.

A higher combined deductible may be selected in a more expensive insurance contract. We define the d/P ratio to reflect this phenomenon,

$$d/P = \frac{d}{\text{Premium}_{d=0}} \times 100\%,$$

where $\text{Premium}_{d=0}$ is the premium of the policy without a deductible. The main reason we introduce d/P is to measure the relative size of the combined deductible to the premium of a policy with no deductible ($\text{Premium}_{d=0}$). For example, Figure 2.3 shows the differences in pricing between dependence and independence models,

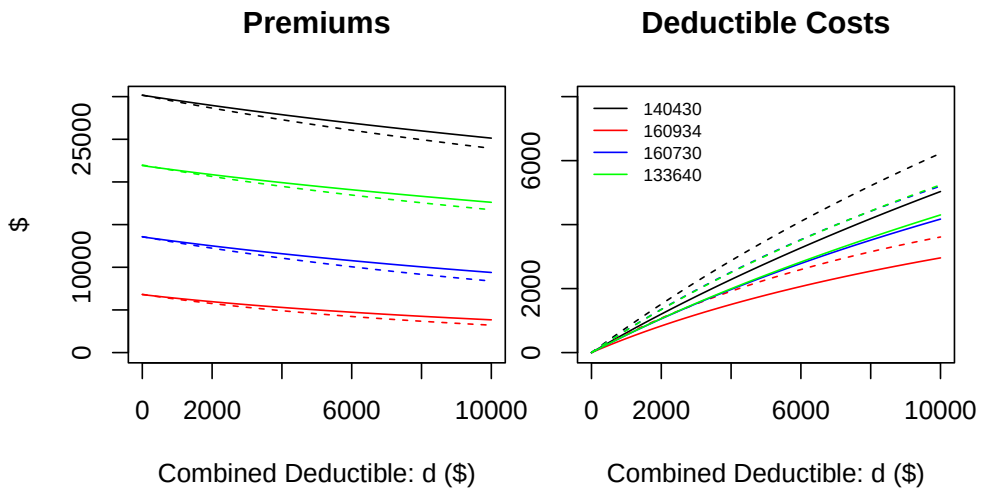

Figure 2.2: Premiums and deductible costs of four policyholders (dependence modelling: solid line, independence modelling: dotted line).

measured by CR; see equation (2.6), with different levels of combined deductibles d and d/P.

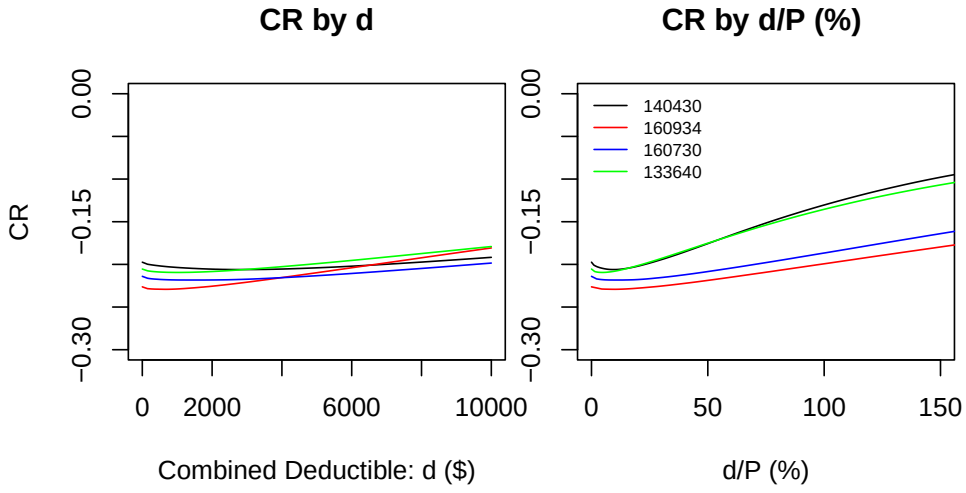

Figure 2.3: Cost Relativities of four policyholders by two forms of combined deductible.

Another way to examine the importance of dependence modelling is to calculate the indicated deductible relativity (IDR) and loss elimination ratio (LER). According to Lee (2017), the relationship between LER and IDR is given as

$$LER(d) = \frac{c_d}{\text{Premium}_{d=0}} = 1 - IDR(d).$$

Actuaries use IDR to assess how much an insurance loss cost is reduced by a deductible from a per-loss perspective, and use LER to assess how much the covered loss is reduced by introducing a deductible. In this case, we examine the difference in $LER(d)$ calculated by dependence and independence models.

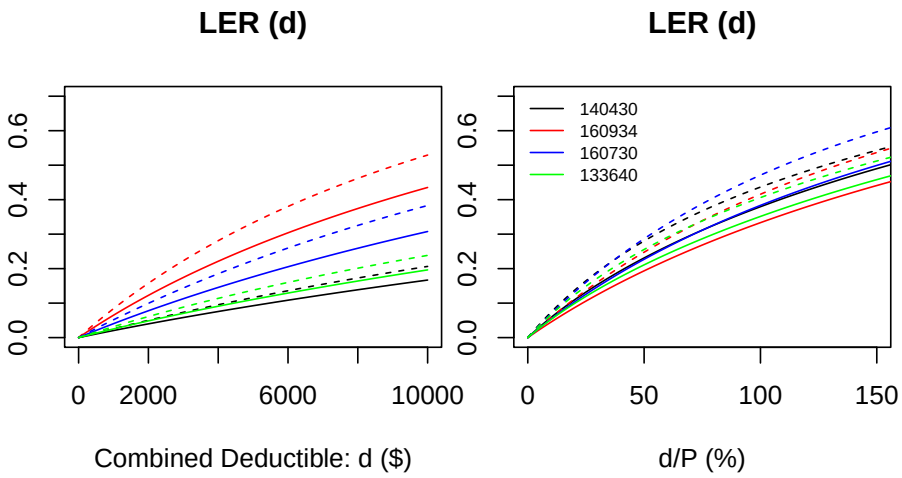

Figure 2.4: LER of four policyholders by two forms of combined deductible (dependence: solid line, independence: dotted line).

From Figure 2.4, we see that a larger combined deductible or d/P leads to a larger difference in LER between dependence models and independence models. We select four d/P ratios (25%, 50%, 75%, 100%) and show their corresponding LERs for the four policyholders in Table 2.4. We find that LERs calculated based on independence (in bracket) are larger than those based on dependence assumptions (out of bracket). This finding will be theoretically explained by Proposition 2.6.1 later. The differences in LERs provide further insights into the importance of dependence modelling.

Table 2.4: Loss Elimination Ratio (%) of four policyholders by d/P (dependence: outside the brackets, independence: in the brackets).

	d/P = 25%	**d/P = 50%**	**d/P = 75%**	**d/P = 100%**
140430	13.21 (16.48)	23.12 (28.04)	31.18 (36.74)	37.98 (43.68)
160934	10.59 (13.64)	19.36 (24.69)	26.84 (33.81)	33.36 (41.47)
160730	12.45 (15.86)	22.62 (28.53)	31.11 (38.71)	38.32 (47.02)
133640	12.00 (14.90)	21.08 (25.48)	28.68 (33.79)	35.17 (40.56)

Adding a combined deductible to bundled contracts can eliminate a large number of small claims for insurers. For readers who are interested in what proportion of claims can be eliminated by certain combined deductibles, we refer to Appendix A.3.

2.4.2 Application 2: Insurance Company Loss and Expense (Reinsurance)

This application is based on the work of Frees and Valdez (1998) who fit copulas (Gumbel and Frank) with Pareto marginal distributions to insurance company indemnity claims. The data used to calibrate a reinsurance contract comprises 1500 general liability claims randomly chosen from late settlement lags. Each claim consists of an indemnity payment (the loss) and an allocated loss adjustment expense (ALAE). Table 2.5 shows descriptive statistics for the loss and ALAE included in this reinsurance contract.

Table 2.5: Descriptive statistics for the loss and ALAE.

Variables	**Min**	**Median**	**Mean**	**Max**	**SD**
Loss	10	12,000	41,208	2,173,595	102,748
ALAE	15	5,471	12,588	501,863	28,146

In this chapter, we consider the excess-of-loss reinsurance and write the deductible cost as

$$c_d = E\ \min(X_{loss} + X_{ALAE}, d),$$

where X_k is the severity of k-th coverage. It means that the reinsurer agrees to pay for each claim the excess over the combined deductible d. For the reinsurers, the combined deductible is naturally small compared to the loss they suffer. From the perspective of insurers, the combined deductible is the maximum loss they undertake, and is mostly applied to the right tail of the aggregate loss distribution. Hence, commonly used risk measures (see Embrechts et al. 2011, for example) for the right tail, such as Value-at-Risk and stop-loss premium, can also be studied in this application. However, this chapter only focuses on the expected cost limited by the combined deductible, which is undertaken by insurers. Also, we only focus on the severity modelling in this case because the frequencies of the loss and ALAE are always the same.

The distribution functions of Gumbel copula, Frank copula, and Pareto distributions are listed in Table 2.6, where $C(u, v)$ is a bivariate copula function and $F_X(x)$ is a cumulative distribution function. The scale parameter and shape parameter of the Pareto distribution are denoted by s and a respectively.

The estimated coefficients of the Gumbel copula model and Frank copula model from Frees and Valdez (1998) are listed in Table 2.7. The parameters of Pareto distributions for the loss and ALAE are (θ_1, a_1) and (θ_2, a_2) respectively.

Although there is no combined deductible in the original contracts, to investigate

Table 2.6: Bivariate copula and marginal distribution functions.

Distributions	**Functions**
Gumbel Copula	$C(u,v) = \exp\left(-\left((-\ln u)^{\theta} + (-\ln v)^{\theta}\right)^{\frac{1}{\theta}}\right)$
Frank Copula	$C(u,v) = -\frac{1}{\theta}\left(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\right)$
Pareto	$F_X(x) = 1 - \left(\frac{s}{s+x}\right)^{a}$

Table 2.7: Parameter estimates using Gumbel copula and Frank copula with Pareto marginals.

Parameters	**Gumbel Model**	**Frank Model**
Loss: (s_1, a_1)	(14036, 1.122)	(14558, 1.115)
ALAE: (s_2, a_2)	(14219, 2.118)	(16678, 2.309)
Dependence Parameter θ	1.453	3.158

the impact of dependence modelling on the deductible cost, we assume a combined deductible d on the loss and ALAE. Considering this is a reinsurance contract, we set the upper bound of combined deductible to be \$50000.

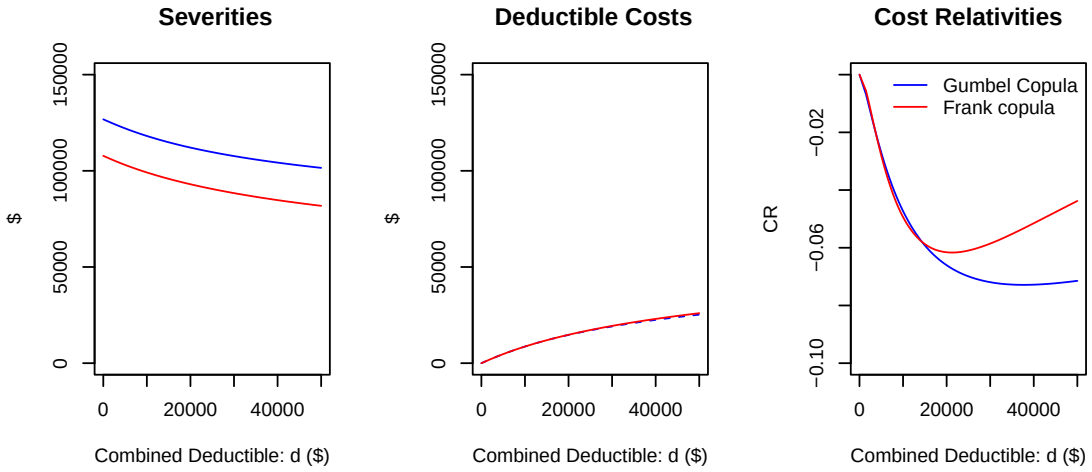

Figure 2.5: Severities, deductible costs, and Cost Relativities of Gumbel model and Frank model.

Figure 2.5 depicts the severity, deductible costs, and cost relativity calculated by two types of copula models. In general, the choice of copulas affects the outcome since different copulas represent different dependence structures. We will investigate the impact of copulas on the deductible costs in Section 2.5.1 later. If we only consider the dependence of severity without the frequency between risks, the CR tends to zero if the combined deductible d tends to zero. This is the main difference between Application 2 and Application 1, as Application 2 focuses on severity modelling

only.

2.4.3 Application 3: Quotes of a Pet Insurance Contract (Personal)

Relatively little literature has been published with respect to data analysis in the pet insurance market. One of the most valuable examples is Pai et al. (2005), which studied the claim transaction data for dogs covered by a single Canadian pet insurer. This data set recorded 15029 individual dogs, 70 different types of accidents, 411 different types of illnesses, and two causes of death (by accident or by illness). Although the data set is not publicly available, we understand that claims are recorded by the types of illnesses and accidents in practice.

Since it is difficult to get claims data of a pet insurance contract, in this section, we calibrate models based on the online quotes of an accident and illness pet insurance provider who provides quotes with choices of four combined limits ($5000, $8000, $10000, $15000), five combined deductibles ($200, $300, $500, $750, $1000) and three reimbursement rates (70%, 80%, 90%). We collect 60 quotes data online to calibrate the underlying parameters and distributions. Table 2.8 shows descriptive statistics for online quotes of the pet insurance contract.

Table 2.8: Descriptive statistics for online quotes of the pet insurance contract.

Variables	Min	Median	Mean	Max	SD
Quote	164	521	533	1,271	259

Figure 2.6 shows the 60 data points collected from the insurer's website. We can see that holding else equal, a higher combined limit, a lower combined deductible, or a higher combined reimbursement rate will lead to a higher premium.

The deductible cost is defined as

$$c_d = E\ \min(X_{illness} + X_{accident}, d),$$

where X_k is the aggregate loss of k-th coverage group (illness and accident). It means that the insurer agrees to pay for each claim the excess over the combined deductible d.

Based on practical pet insurance information (NAIC 2018), we assume the probability of having at least one claim (illness and/or accident) per year is around 30% to 50% and that the frequency models for the illness and accident losses are distributed

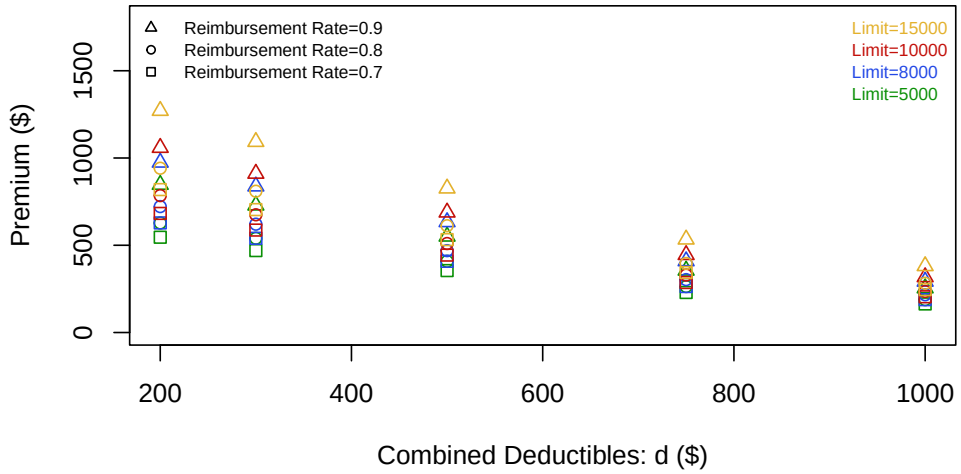

Figure 2.6: Online quotes data of a pet insurance contract.

as Poisson($\lambda = 0.4$) and Poisson($\lambda = 0.25$), respectively. A bivariate Gaussian copula is used to model the dependence of the two risks (illness and accident). We consider the gamma distribution, the Pareto distribution, and the Weibull distribution as three candidates for the two marginal severity distributions. The probability density functions and the corresponding shape parameters and expected values are listed in Table 2.9. In total, five parameters need to be calibrated, including the dependence parameter of the Gaussian copula ρ_S and four parameters of the two marginal severity distributions.

Table 2.9: Pet insurance distribution functions.

Distribution	**Probability Density Function**	**Shape Parameter**	**E(X)**
gamma	$f(x) = \frac{x^{(a-1)}e^{-\frac{x}{s}}}{s^a\Gamma(a)}$	a	as
Weibull	$f(x) = \frac{a}{s}\left(\frac{x}{s}\right)^{a-1}\exp\left(-\left(\frac{x}{s}\right)^a\right)$	a	$s\Gamma\left(1+\frac{1}{a}\right)$
Pareto	$f(x) = \frac{as^a}{(x+s)^{a+1}}$	a	$\frac{s}{a-1}$

We consider four models with different combinations of marginal severity distributions, two Poisson marginal frequency distributions, and a Gaussian copula. The four models are listed as follows:

1. Model 1 : $X_{illness} \sim$ gamma, $X_{accident} \sim$ gamma, Poisson frequency, Gaussian copula.

2. Model 2 : $X_{illness} \sim$ gamma, $X_{accident} \sim$ Pareto, Poisson frequency, Gaussian copula.
3. Model 3 : $X_{illness} \sim$ gamma, $X_{accident} \sim$ Weibull, Poisson frequency, Gaussian copula.
4. Model 4 : $X_{illness} \sim$ Weibull, $X_{accident} \sim$ Weibull, Poisson frequency, Gaussian copula.

We apply the mean absolute percentage error (MAPE) to measure the model performance

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_i - F_i}{A_i}\right|,$$

where A_i is the actual value (quotes) and F_i is the forecast value (estimated premium), and n is the number of samples. We refer to Appendix A.4 for detailed steps of the calibration. The five calibrated parameters of the four models are listed in Table 2.10.

Table 2.10: Five calibrated parameters of four pet insurance models.

Parameters	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Shape - illness	20.00	20.00	15.00	15.00
E($X_{illness}$)	1166.80	1182.00	1177.30	1131.70
Shape - accident	20.00	2.01	10.00	15.00
E($X_{accident}$)	986.90	1049.00	969.00	1070.30
Dependence (Gaussian)	0.95	0.40	0.95	0.95

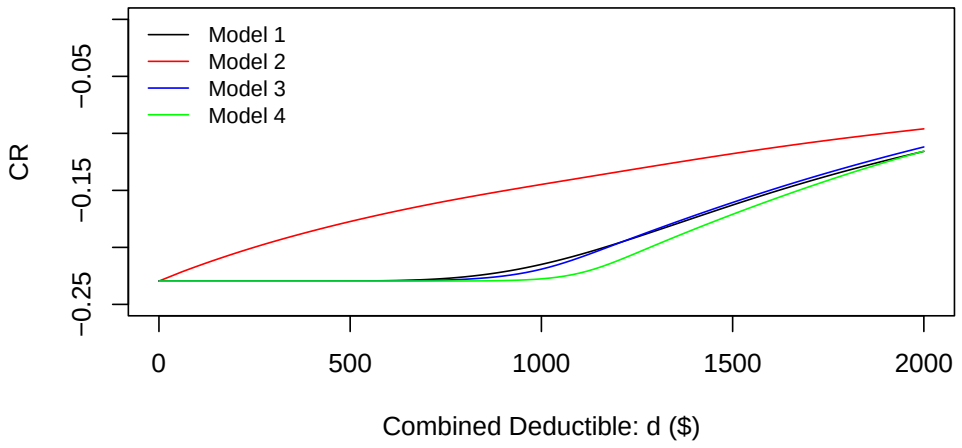

Figure 2.7: Cost Relativities of four pet insurance models.

Figure 2.7 shows the calculated cost relativity using the fitting outcome of the four models. When the marginal distributions are gamma, Weibull, or the combination

of gamma and Weibull distributions, the curves of CR are similar (blue, black, and green curves). A special case is that when we use the combination of gamma and Pareto distributions (Model 2). Here, the red curve in Figure 2.7 is more linear and above the other three curves due to the difference in marginal distributions and a low dependence parameter of the Gaussian copula. In this application, we find that the deductible costs are reduced by approximately 10 to 23 per cent when dependence models are used compared to the benchmark models that assume independence. We will investigate the impact of marginal distributions on the deductible costs in Section 2.5.1 later.

2.4.4 Summary of the Three Applications

In the first application, combined deductibles are applied to commercial insurance contracts with six coverages. Deductible costs are reduced by approximately 6 to 23 per cent, and loss elimination ratios are reduced by 3 to 9 per cent in dependence models compared to the benchmark models that assume independence.

Reinsurance is discussed in the second application, where a combined deductible is applied to an indemnity payment and an allocated loss adjustment expense. We find the choice of copulas could lead to different deductible costs. Although an indemnity payment always comes with an allocated loss adjustment expense (the frequencies are the same for the two coverages), the difference in deductible costs between dependence and independence modelling could reach 8 per cent.

The third application is about personal insurance (i.e. pet insurance), in which deductible costs are reduced by approximately 10 to 23 per cent when dependence models are used compared to the benchmark models that assume independence. Additionally, we find the choice of the marginal distributions of each coverage could impact the model outcome.

Note that in the first and third applications, we work with a combined deductible on the total loss of all coverages, whereas for the second application, we have a combined deductible on the total cost (loss + expense) arising from an individual claim. Therefore, the dependence captured in the two different scenarios has different implications. The dependence we observe in the first and third applications mainly comes from the inherent association between certain claim risks. For example, in private insurance, a pet prone to illness (illness claim) is also likely to die (death claim). For example, in commercial insurance, when a business owner's building is damaged (building damage claim), it is also likely to cause business interruption (business interruption claim). The observed dependence in the second application mainly comes from a single claim that simultaneously affects the amount of the

claim's loss and expenses. While the copula models used in this chapter could flexibly capture the two forms of dependence discussed above as well as other forms of dependence in claims modelling (see Shi et al. 2015; Shi and Lee 2022, for example), we acknowledge that there are other models that can be used to capture different forms of dependence. For example, in Chapter 4, we propose a shared random effects model that could capture the dependence coming from unobserved heterogeneity in claims risk across customers.

A unified conclusion of the three applications is that the ratio CR is non-positive in all cases, and the deductible costs can be affected by the choice of copulas and marginal distributions. The results of these three applications show how dependence affects the insurance pricing of combined deductibles.

2.5 Simulation Study

As described in Section 2.3, we use copulas to model the dependence among insurance losses of different coverages when calculating deductible costs. Intuitively, the choice of copulas will directly determine the dependence structure and further affect the calculation result of deductible costs. Motivated by the results of the three insurance applications in Section 2.4, we further investigate how the dependence structure (dependence parameters and copulas) and marginal distributions could affect the deductible costs and CR using Monte Carlo simulations. The number of simulations is discussed in Appendix A.7.

All settings in this section were purposefully designed to make the illustration relatively simple yet insightful. We consider a bivariate case with two risks X_1 and X_2 for severity modelling and ignore the dependence of frequency between risks. In future simulation studies of deductible costs, one could thoroughly explore the dependence of severity and frequency across multiple risks as well as the dependence between the severity and frequency of each individual risk (Shi et al. 2015). Table 2.11 summarises the four combinations of severity marginal distributions for X_1 and X_2. Marginal 1 (uniform distribution) is only for illustration and comparison purposes. Marginals 2 (gamma distribution), 3 (Pareto distribution), and 4 (Weibull distribution) are widely used in the severity analysis of insurance claims; see Resti et al. (2013), Teodorescu and Vernic (2006), and Ni et al. (2014). The distribution functions of these four distributions are provided in Appendix A.6. In all four cases, we set $E(X_1 + X_2) = 1000$ and the combined deductible ranges from 0 to 1000 for illustration purposes.

Table 2.11: Four combinations of the severity marginal distributions.

Marginal Distributions	X_1	X_2
1	uniform (0, 1000)	uniform (0, 1000)
2	gamma (1, 500)	gamma (2.5, 200)
3	Pareto (2, 400)	Pareto (3, 1200)
4	Weibull (0.5, 300)	Weibull (1.5, 443.1)

2.5.1 Simulation Results

Comparing the Impact of Dependence Parameters

We generate data for the simulation study based on the four combinations of the severity marginal distributions listed in Table 2.11. A bivariate Gaussian copula is used to join a couple of severity marginal distributions (X_1, X_2). The Spearman's rank correlation coefficient of the Gaussian copula, $\rho_S(X_1, X_2)$, takes values of -0.8, -0.4, 0, 0.4, and 0.8 for comparison purposes. When $\rho_S(X_1, X_2) = 0$, the model can be regarded as the independence model.

Figure 2.8 shows the impact of the dependence parameter ρ_S on the deductible cost and CR using the four marginal combinations. When ρ_S is negative, the deductible cost of a dependence model is higher than that of the independence model (positive CR); and when ρ_S is positive, the deductible cost of a dependence model is lower than that of the independence model (negative CR); see Proposition 2.6.1 in Section 2.6 for theoretical explanations. Additionally, when ρ_S is positive (0.4 and 0.8), a stronger positive dependence leads to a lower deductible cost and CR. However, when ρ_S is negative (-0.4 and -0.8), a stronger negative dependence leads to a higher deductible cost and CR; see Proposition 2.6.2 in Section 2.6 for theoretical explanations. Lastly, as d increases, the gap between the deductible costs of a dependence model and that of an independence model increases first and then decreases. For an extreme case, if we allow d to increase to infinity (all costs will be paid by the policyholder), then the gap will reduce to zero ($CR = 0$), as explained by Proposition 2.6.3 in Section 2.6.

One of the primary objectives of risk theory is to model the distribution of total claim costs for portfolios of policies to make business decisions regarding various aspects of insurance contracts. We also investigate the linear approximations of deductible costs for readers interested in the analytical result and calculation. We derive three approximations of $E\ \min(X_1+X_2, d)$ using the Trapezoidal rule (Chiou 1978); see Appendix A.8 for more details.

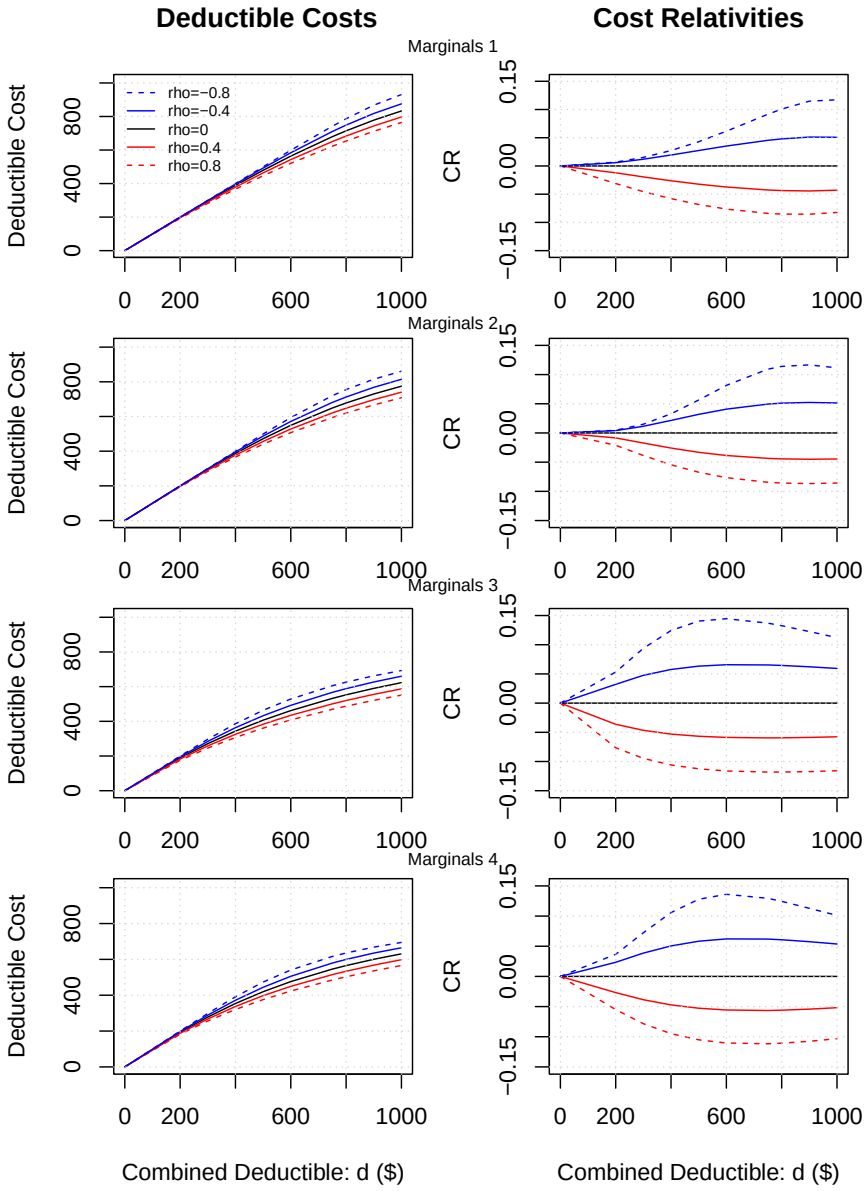

Figure 2.8: Deductible costs and Cost Relativities of four Gaussian copula models with five dependence parameters.

Comparing the Impact of Different Copulas

We examine the impact of different copulas on CR in this section by considering positive dependence parameters ($\rho_S = 0.4$ and $\rho_S = 0.8$) and heavy-tailed Pareto distributions (Marginals 3) and Weibull (Marginals 4) distributions. Note that the Weibull distribution is heavy-tailed for its shape parameter $0 < a < 1$ and light-tailed for $a > 1$. Five widely studied copulas, Gaussian, Student-t, Frank, Gumbel, and Clayton are applied for comparison purposes. The bivariate function and tail dependence of these five copulas are summarized in Table 2.12. The tail dependence

describes the association between the upper or lower tail of marginals (Venter et al. 2002).

Table 2.12: Tail Dependence of Copulas

Copula	$C(u,v) =$	**Lower Tail Dependence**	**Upper Tail Dependence**
Gaussian	$\Phi_\theta\Big(\Phi^{-1}(u), \Phi^{-1}(v)\Big)$, $\theta \in [-1,1]$.	No	No
Student-t	$H_{v,\theta}\Big(F_v^{-1}(u), G_v^{-1}(v)\Big)$, $\theta \in [-1,1],\ v \geq 2$.	Yes	Yes
Frank	$-\frac{1}{\theta}\Big(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\Big)$, $\theta \in \mathbb{R}\setminus 0$.	No	No
Gumbel	$\exp\Big(-\big((-ln(u))^\theta + (-ln(v))^\theta\big)^{\frac{1}{\theta}}\Big)$, $\theta \in [1,\infty)$.	No	Yes
Clayton	$\max(u^{-\theta}+v^{-\theta}-1, 0)^{\frac{-1}{\theta}}$, $\theta \in [-1,\infty)\setminus 0$.	Yes	No

Figure 2.9 shows the impact of different copulas on CR using Marginals 3 and 4. The two plots on the top are based on the copula models with Pareto marginals, while the two bottom plots are based on copula models with Weibull marginals. Compared to the other four copulas, the Clayton copula leads to a sharp decrease of CR and then a slow increase when d increases from 0 to 1000. A reason is that the Clayton copula is an asymmetric Archimedean copula, exhibiting dependence in the lower tail but not in the upper tail. Thus, the CR is very sensitive to small d (in the lower tail) when using the Clayton copula. On the contrary, the Gumbel copula, another asymmetric Archimedean copula, exhibits dependence in the upper tail but not in the lower tail. Hence the CR decreases with d slowly when d is small. For insurance contracts, d is mostly in the lower tail of the aggregate loss distribution, so copulas with lower tail dependence tend to have larger values of $|CR|$ than copulas without lower tail dependence. Similarly, when a combined deductible is placed at the right tail of a loss distribution for reinsurance contracts, we expect that copulas with upper tail dependence tend to have larger values of $|CR|$ than copulas without upper tail dependence.

Comparing the left and right panels in Figure 2.9, we also notice that stronger dependence leads to larger values of $|CR|$, which is explained in Proposition 2.6.2 in Section 2.6. In summary, the level of dependence (ρ_S) is more important than the choice of the copula in calculating the deductible costs.

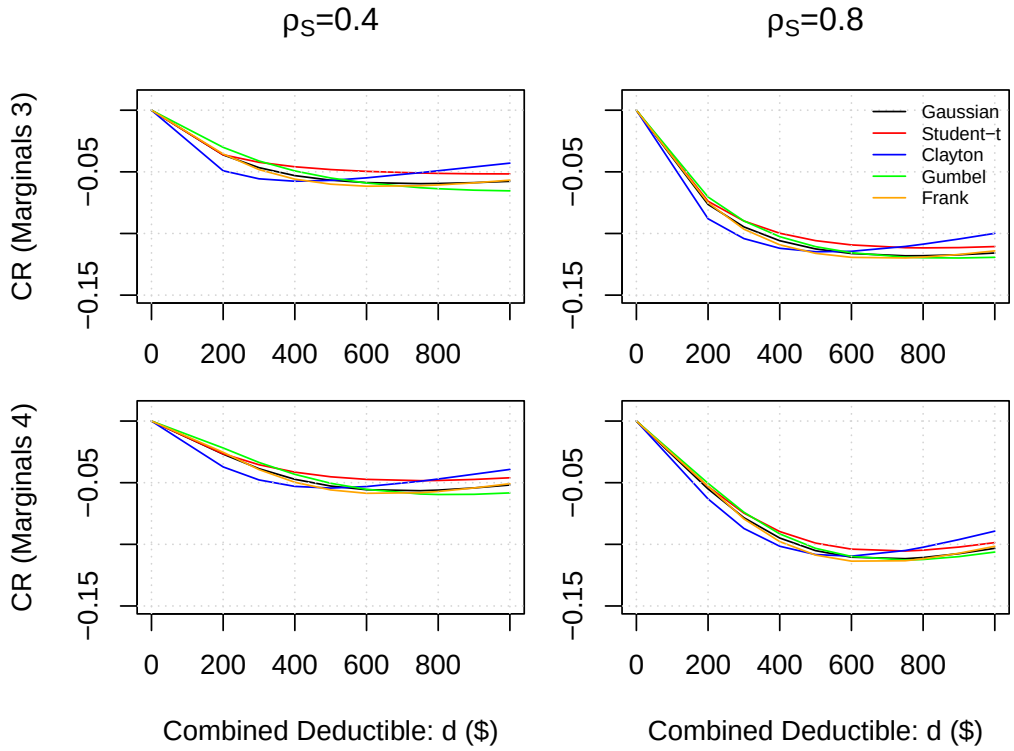

Figure 2.9: Cost Relativities of five copulas.

Comparing the Impact of Marginal Distributions

Data generated in Section 2.5.1 is also used in this analysis. However, we only consider positive dependence parameters $\rho_S = 0.4$ and $\rho_S = 0.8$ to be consistent with real insurance practice; e.g., see the three applications in Section 2.4. In Figure 2.10, we show the impact of different marginal distributions on both the deductible costs (top figures) and CR (bottom figures). We can see that the deductible costs and CR vary by marginal distributions, and the difference in CR becomes larger when the dependence is stronger. It suggests that practitioners should choose marginal distributions carefully, especially when the dependence between risks is strong.

2.6 Propositions of Combined Deductible Costs

In Sections 2.4 and 2.5, we have discussed several findings that shed light on the importance of dependence modelling of combined deductible costs for bundled insurance contracts. For example, in Section 2.4, CR is non-positive in all cases and the deductible costs can be affected by the choice of copulas and marginal distribu-

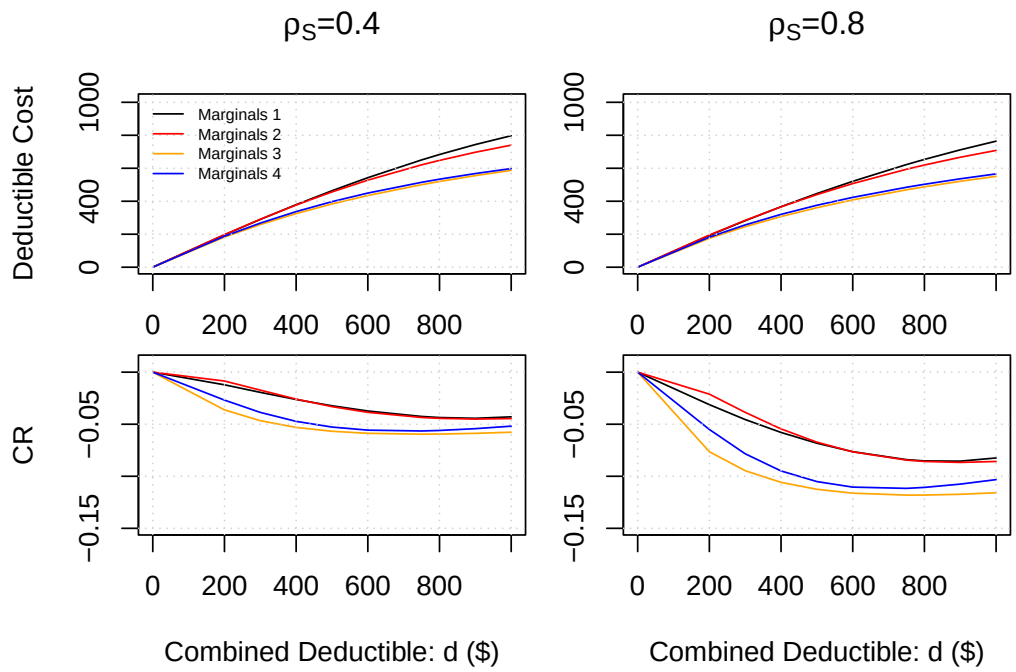

Figure 2.10: Deductible costs and Cost Relativities of four models with different marginals.

tions. Also, in Section 2.5, the level of dependence (ρ_S) is more important than the choice of copulas and marginal distributions to calculate the deductible costs. The current findings lack theoretical support, so they are difficult to understand intuitively. However, actuarial science, essential to the sustainable development of the insurance industry, cannot thrive without practitioners having a solid mathematics foundation. Taking advantage of the well-known theories of stop-loss premiums (Denuit et al. 2006), this section presents general properties of joint distributions that can lead to materially different prices, and shows how the dependence structure affects insurance pricing of combined deductibles.

Propositions introduced in this section help explain the research findings mentioned above, and help practitioners theoretically understand the importance of dependence modelling in the combined deductible ratemaking. Also, for future research on the combined deductible ratemaking, these propositions provide theoretical support to different types of dependence models. For example, the propositions could be applied to study the dependence of different marginals, including severity (see Section 2.4.2), frequency (not covered in this chapter), and aggregate claims (see Sections 2.4.1 and 2.4.3). With these general propositions, researchers can identify problems and solve them in other areas of actuarial science. For example, we could investigate the impact of applying a combined deductible to medical expenditures

with strong associations as studied by Frees et al. (2013). In this section, all propositions are studied in a bivariate case, which constructs a fundamental framework for multivariate studies.

To describe the propositions, we start with four definitions. The first definition is about the concept of positive quadrant dependence introduced by Lehmann (1966). The random couple $\boldsymbol{X} = (X_1, X_2)$ is positively quadrant dependent (PQD) if

$$H_{X_1,X_2}(x_1, x_2) := P(X_1 \leq x_1, X_2 \leq x_2) - P(X_1 \leq x_1)P(X_2 \leq x_2) \geq 0, \tag{2.7}$$

for all $x_1, x_2 \in \mathbb{R}$. Similarly, $\boldsymbol{X}$ is negatively quadrant dependent (NQD) if $H_{X_1,X_2}(x_1, x_2) \leq 0$. According to equation (2.7), $\boldsymbol{X}$ is NQD if, and only if $(X_1, -X_2)$ is PQD. In view of this duality, we shall mainly focus on PQD in the following analysis.

The second definition is about the PQD condition of a copula. Let $C(u_1, u_2)$ be the copula of the random couple $\boldsymbol{X} = (X_1, X_2)$. Then, $C(u_1, u_2)$ is PQD if

$$C(u_1, u_2) - u_1 u_2 \geq 0, \tag{2.8}$$

for all $u_1, u_2 \in [0, 1]$. Based on equations (2.7) and (2.8), $\boldsymbol{X}$ is PQD (resp. NQD) if, and only if its copula $C(u_1, u_2)$ is PQD (resp. NQD); see Corollary 5.3.4 in Denuit et al. (2006). We summarise the PQD and NQD conditions (Salvadori et al. 2007; Nelsen 2007) with respect to ρ_S (Spearman's rank correlation coefficient), τ (Kendall rank correlation coefficient), and θ (dependence parameter of a copula) for some common copulas in Appendix A.10. In many cases, the insurance losses among different coverages are PQD (e.g. the three applications in Section 2.4). If a Gaussian copula or Student-t copula is fitted with $0 \leq \rho_S \leq 1$ (or equivalently, $0 \leq \tau \leq 1$ and $0 \leq \theta \leq 1$), the copula and hence the random couple of losses are PQD.

The third definition is about the correlation order. Consider two random couples $\boldsymbol{X} = (X_1, X_2)$ and $\boldsymbol{Y} = (Y_1, Y_2)$ in the Fréchet space $\mathcal{R}_2(F_1, F_2)$. The Fréchet space gathers together all the probability distributions with fixed univariate marginals (F_1, F_2). Elements in a given Fréchet space only differ in their dependence structures, and not in their marginal behaviours. If two joint cumulative distribution functions $F_{\boldsymbol{X}}(x_1, x_2)$ and $F_{\boldsymbol{Y}}(x_1, x_2)$ satisfy

$$F_{\boldsymbol{X}}(x_1, x_2) \leq F_{\boldsymbol{Y}}(x_1, x_2),$$

for all $x_1, x_2 \in \mathbb{R}$. Then $\boldsymbol{Y}$ is said to be larger than $\boldsymbol{X}$ in correlation order denoted

by $\boldsymbol{X} \preceq_{\text{CORR}} \boldsymbol{Y}$.

The last definition is about the integrated difference in pricing ($IDIP$) of the random couple $\boldsymbol{X} = (X_1, X_2)$, which is defined as

$$IDIP = \int_0^\infty |E\min(X_1^\perp + X_2^\perp, t) - E\min(X_1 + X_2, t)|\ dt.$$

We regard $IDIP$ as the overall difference in pricing of combined deductibles between the dependence model and the independence model. A larger value of $IDIP$ means a larger overall difference in deductible costs between dependence and independence modelling.

Based on the four definitions above and theorems of dependence in Sections 5 and 6 of Denuit et al. (2006), we introduce the following propositions of deductible costs. Detailed proofs are provided in Appendix A.9.

Proposition 2.6.1. *If the random couple $\boldsymbol{X} = (X_1, X_2)$ is PQD, then $c_d^{dep} \leq c_d^{ind}$ (i.e. $CR \leq 0$). If $\boldsymbol{X}$ is NQD, then $c_d^{dep} \geq c_d^{ind}$ (i.e. $CR \geq 0$).*

Proposition 2.6.1 explains why CR is non-positive in Figures 2.1, 2.3, 2.5, and 2.7. It also explains why the LERs based on independent assumptions are larger than those based on dependent assumptions in Table 2.4. This proposition can help practitioners compare deductible costs based on dependence modelling and independence modelling before building any models. In most cases, the losses of different coverages in a bundled insurance contract are positively correlated and satisfy the PQD condition. For example, the loss and ALAE in Frees and Valdez (1998) is PQD, so dependence modelling leads to lower deductible costs and higher premiums compared to independence modelling. If a random couple loss is NQD (rare in practice), then dependence modelling leads to higher deductible costs and lower premiums compared to independence modelling.

Proposition 2.6.2. *If two random couples $\boldsymbol{X} = (X_1, X_2)$ and $\boldsymbol{Y} = (Y_1, Y_2)$ follow $\boldsymbol{X} \preceq_{CORR} \boldsymbol{Y}$, then $CR_{\boldsymbol{X}} \geq CR_{\boldsymbol{Y}}$.*

$CR_{\boldsymbol{X}}$ (resp. $CR_{\boldsymbol{Y}}$) is the cost relativity ratio calculated for $\boldsymbol{X}$ (resp. $\boldsymbol{Y}$). The random couple $\boldsymbol{X}^\perp = (X_1^\perp, X_2^\perp)$ is an independent version of $\boldsymbol{X}$, which means $X_1^\perp$ and $X_2^\perp$ are independent. Positive correlation order means $\boldsymbol{X}^\perp \preceq_{\text{CORR}} \boldsymbol{X}$, which is equivalent to $\boldsymbol{X}$ is PQD; see equation (2.7). Negative correlation order means $\boldsymbol{X} \preceq_{\text{CORR}} \boldsymbol{X}^\perp$, which is equivalent to $\boldsymbol{X}$ is NQD. To compare the correlation order between risks easily, we also have the following relationship using three correlation coefficients, including ρ_P is the Pearson correlation coefficient, ρ_S is the Spearman's rank correlation coefficient, and τ is the Kendall rank correlation coefficient; see

Property 6.2.14 in Denuit et al. (2006). If $\boldsymbol{X} \preceq_{\text{CORR}} \boldsymbol{Y}$, then we have

$$\begin{aligned} \rho_P(X_1, X_2) &\leq \rho_P(Y_1, Y_2), \\ \tau(X_1, X_2) &\leq \tau(Y_1, Y_2), \\ \rho_S(X_1, X_2) &\leq \rho_S(Y_1, Y_2). \end{aligned}$$

Proposition 2.6.2 indicates that a stronger dependence structure leads to a larger pricing difference. It explains why bigger positive ρ_S (PQD) or smaller negative ρ_S (NQD) leads to bigger values of $|CR|$ in Figures 2.8, 2.9, and 2.10. This proposition helps practitioners compare deductible costs based on different forms of dependence by checking the three classic association coefficients. For example, in Section 2.5.1, without doing any calculation, we know that the deductible cost of $\rho_S = 0.8$ is smaller than that of $\rho_S = 0.4$. Hence, this proposition can further guide practitioners in judging the necessity of dependence modelling. Intuitively, it is more necessary to consider dependence modelling when $\rho_S(X_1, X_2)$ is 0.9 than when $\rho_S(X_1, X_2)$ is 0.1, since building a dependence model can be expensive and time-consuming in practice.

Proposition 2.6.3. *Define $A = X_1^{\perp} + X_2^{\perp}$ and $B = X_1 + X_2$. The cumulative distribution functions (CDFs) $F_A(a)$ and $F_B(b)$ at least cross once. If they only cross once, the difference in deductible costs $|E\min(A, d) - E\min(B, d)|$ is maximised at the intersection of two CDFs.*

Proposition 2.6.3 is related to Figure 2.8, in which the difference in deductible costs between a dependence model and the independence model is maximised when two CDFs intersect. Also, we can see that the value of $|CR|$ increases to the maximum point and then decreases as the combined deductible d increases. If d is set to be positive infinity, then the difference in deductible costs between a dependence model and the independence model will be zero (i.e. $CR = 0$). This proposition helps practitioners understand that the difference in deductible costs between dependence and independence modelling is also affected by the size of a combined deductible (Frees et al. 2022, especially for the insurance contracts with a large combined deductible). For example, in Figure 2.8, when the combined deductible tends to zero, the difference in deductible costs between dependence and independence modelling is negligible. However, when the combined deductible is set to be at the intersection

of two CDFs, the difference in deductible costs is maximised, and so dependence modelling becomes necessary.

Proposition 2.6.4. *If* $\boldsymbol{X}$ *is PQD, then* $IDIP = \sigma(X_1, X_2)$. *If* $\boldsymbol{X}$ *is NQD, then* $IDIP = -\sigma(X_1, X_2)$.

Proposition 2.6.4 suggests that $IDIP$ is only based on the covariance (σ) of X_1 and X_2. In other words, a larger positive covariance (resp. smaller negative covariance) between variables leads to a larger overall difference in pricing under the PQD (resp. NQD) assumption. Therefore, Proposition 2.6.2 and Proposition 2.6.4 are mutually supportive. We also find that $IDIP$ equals the integrated stop-loss distance introduced in Section 9.8 of Denuit et al. (2006). This proposition helps practitioners understand that the overall difference in deductible costs between dependence and independence modelling is measured by $|\sigma(X_1, X_2)|$. If $|\sigma(X_1, X_2)|$ is larger, then generally speaking, it will be more important to consider dependence modelling for deductible costs.

2.7 Discussion

In this chapter, we study the cost of combined deductibles under dependence from three perspectives: three applications on real data sets, a simulation study, and several theoretical propositions.

The theoretical propositions present general properties of joint distributions that can lead to materially different prices. The theoretical results explain the observations in both the empirical applications and simulation studies. They also provide insights into the understanding of how the dependence structure affects insurance pricing of combined deductibles. The difference in deductible costs between dependence and independence modelling is affected by the size of combined deductibles and the correlation between risks. With these general propositions, researchers can identify problems and solve them in other areas of actuarial science.

In the simulation study, a variety of dependence parameters, marginal distributions, and copulas are studied. The results are consistent with theoretical propositions and suggest that the strength of dependence is more important than the type of copula dependence function. It is also worth noting that the dependence structure and marginal distributions should be selected carefully in practice, which could lead to different deductible costs.

The results of three applications (commercial insurance, reinsurance, and personal insurance) show the importance of dependence modelling for calculating deductible

costs. According to the theoretical calculations, when the dependence parameter is positive, deductible costs calculated by dependence models are lower than those calculated by benchmark models that assume independence. In the three applications, all dependence parameters are positive. The deductible costs are reduced by approximately 6 to 23 per cent for commercial insurance, up to 8 per cent for reinsurance, and 10 to 23 per cent for pet insurance when dependence models are used compared to the benchmark models that assume independence.

The three perspectives above convey a similar conclusion that deductible costs calculated by dependence models are different from those calculated by the independence models, and the difference can be as large as 23 per cent. The results shed light on the importance of dependence modelling of combined deductible costs for bundled insurance contracts.

For future research, we summarise the following four directions. Firstly, more types of dependence modelling techniques, such as vine copulas, can be applied to study combined deductible ratemaking. Vine copulas are well known for modelling complex dependency patterns through a rich variety of bivariate copulas in a tree structure; see Allen et al. (2017). Moreover, in Chapter 4, we introduce a new joint model for multivariate claims, which could capture the unobserved heterogeneity in claims risks across policyholders. Secondly, the dependence modelling technique introduced in this chapter can also be applied to study the deductible cost of an all-peril insurance policy. For example, a combined deductible (also called an all-peril deductible) under an all-peril homeowners insurance policy covers the aggregate loss of all perils, such as fire and smoke, lightning strikes, windstorms, and hail. Thirdly, Shi and Lee (2022) proposed an endogenous split deductible model and found that the policyholder's split deductible is negatively associated with the claim frequency but positively associated with the claim severity. Thus, future research could examine the association between an endogenous combined deductible and claims of multiple coverages. Lastly, since the deductible cost is undertaken by policyholders, deductible costs based on dependence modelling can be further used to analyse the effect of combined deductibles on consumer behaviour. For example, it is well known that adverse selection may take place if health insurance policies are offered with the option to take a deductible in exchange for a premium reduction; see for example Van de Ven and Van Praag (1981).

Chapter 3

Multi-State Modelling of Customer Churn

Yumo Dong, Edward W (Jed) Frees, Fei Huang, Francis K. C. Hui

This chapter has been published as a paper in ASTIN Bulletin:
Dong, Y., Frees, E. W., Huang, F. and Hui, F. K. C., 'Multi-state modelling of customer churn'. ASTIN Bulletin. 2022;52(3):735-764. doi:10.1017/asb.2022.18

3.1 Introduction

Customer churn (also known as customer attrition, dropout, turnover, or defection) refers to the non-renewal of existing customers for an insurance company. In general insurance, customer churn poses a major challenge to insurers because contracts are renewed periodically (commonly after one year), and with substantial competition and diverse contracts, it is easy for customers to switch insurance providers to satisfy their needs at the time. Customer churn also poses a challenge in life insurance and other industries such as banking and telecommunications; keeping a current customer rather than acquiring a new one is often a more profitable strategy for a company (Simchi-Levi et al. 2008), because long-term consumers typically take up less of the company's resources compared to new customers (Min and Zhou 2002). As such, analyses to determine the drivers and customer behaviours involving churn are critical to a company's customer retention program (Okongwu et al. 2016); identifying those customers with high intention to churn can help insurers secure continued loyalty.

This chapter is motivated by customer churn analysis for commercial insurance, which plays an important role in the general insurance industry. According to the direct premium written reported by Brian Briggs (2021), the commercial market represents 49% of the total direct premium written in the U.S. property & casualty insurance industry. Commercial insurance contracts can include several coverages such as building, equipment, and auto coverage, and it is common for customers to have multiple contracts with the same company, or (alternatively) have multiple coverages under a single bundled contract (Dong, Frees and Huang 2022). Additionally, customers add or drop coverages, or switch providers entirely, on a regular basis to meet their benefits and budgets. Indeed, as revealed in a survey for Australian small commercial insurance consumers (Brant Carson 2017), in the small and medium enterprise insurance market, around three-quarters of small business owners are dissatisfied with their commercial insurance providers. Approximately 19 per cent of small business owners choose to switch insurance providers every year.

Since commercial insurance contracts are typically of short duration (so that customer churn is more about the non-renewal of a contract rather than breaking an existing contract), traditional customer churn analysis has focused on static models applied to data from a single period. This is despite the fact that commercial insurers can trace information about their customers over multiple periods, resulting in readily available longitudinal data. Moreover, with single period data, various methods such as logistic regression, decision trees, and neural networks have been proposed to predict the probability of customer churn for the next contract period (e.g., Guillen

et al. 2003; He et al. 2020; Yeo et al. 2001). However, with multi-period or longitudinal data, customers often traverse a larger number of states beyond simply deciding whether to churn or not. That is, as discussed above, during a multi-period contract term, commercial policyholders may regularly adjust their coverages combination by adding and dropping coverages. For example, an organisation policyholder taking both building and equipment coverages as a bundled contract in one year may decide to drop equipment coverage but add auto coverage in the following year. Put simply, multi-period data more accurately reflects the dynamic changes in the needs of policyholders and the broader market, e.g., the coverage first dropped by a policyholder is usually an important signal to predict subsequent likelihood to churn (Brockett et al. 2008). Thus, there is a strong motivation for commercial insurers to break away from traditional customer churn analysis. In this chapter, we propose and advocate using multi-period data to perform multi-state customer churn analysis, focusing not only on the customer churn *per-se*, but also on the coverage adjustment behaviour of commercial customers over time.

While there are a number of ways by which multi-state customer churn analysis can be performed, in this chapter, we adopt a method that prioritises interpretability over predictive performance, namely multinomial logistic regression (MLR) with a higher-order Markov framework. To our knowledge, this chapter (paper) is the first to use a higher-order Markov model for customer churn research in actuarial science. We illustrate the use of this MLR and multi-state customer churn analysis on a longitudinal commercial insurance dataset from the Wisconsin Local Government Property Insurance Fund (LGPIF). Exploratory analysis of this data reveals that transition probabilities tend to depend strongly on the states of the policyholder in their two previous periods, which motivates considering a second-order MLR model to study policyholders' state transitions. We demonstrate how different sets of predictors tend to drive the various likelihoods of transition, which we visualise through conditional effect and average marginal effect plots. A comparison of multi-state customer churn analysis to traditional (single-state) customer churn analysis reveals that the former can lead to more accurate calculations of customer lifetime value. Furthermore, the second-order MLR model has consistently strong out-of-sample predictive performance relative to several parametric models of different orders, and also compared to more non-parametric techniques such as support vector machines (SVMs) and gradient boosting machines (GBMs).

The remainder of this chapter is organised as follows. In Section 3.2, we review the literature on customer churn and multi-state modelling. In Section 3.3, we illustrate how to perform multi-state customer churn analysis using MLR models under the

first and second-order Markov assumptions. In Section 3.4, we illustrate an application of second-order MLR models to study multi-state customer churn in the LGPIF data, visualising the estimated effects along with demonstrating differences in calculating customer lifetime value between multi-state versus traditional customer churn modelling. In Section 3.5, we compare the out-of-sample prediction performance of the second-order MLR model with several other methods for forecasting transition. Section 3.6 offers some concluding remarks.

3.2 Literature Review

In actuarial science, multi-state models are commonly divided into time-continuous and time-discrete approaches (see Haberman and Pitacco 2018, for an overview), with the primary difference being whether the transitions between states are assumed to occur at any time or at equally spaced time points during a certain research period. In this chapter, motivated by commercial insurance data from the LGPIF and to be consistent with traditional customer churn analysis (e.g., Guillen et al. 2003; Brockett et al. 2008), we adopt the time-discrete approach and assume that the results of policyholders' renewal decisions are realised at the end of each contract period. However, we acknowledge the vast literature on time-continuous multi-state models used in long-term care insurance (e.g., Pritchard 2006; Fong et al. 2015), life insurance (e.g., Norberg 2013; Buchardt 2014; Buchardt et al. 2015), and general insurance (e.g., Orr 2007; Hürlimann 2015).

In terms of traditional customer churn analysis, particularly in general insurance, there exists an array of different techniques for modelling the likelihood of customer churn (a binary response). By far the most popular of these is logistic regression, or some variation thereof, to quantify the factors driving customer churn (see Guillen et al. 2003; Günther et al. 2014; Staudt and Wagner 2018; Frees et al. 2021, among many others). More recently, data mining techniques such as tree-based methods and neural networks have grown in popularity for predicting churn rates (Bolancé et al. 2016; He et al. 2020). Yet another approach uses survival analysis to study how long a customer will stay with a company after their first policy cancellation (see for instance Brockett et al. 2008; Guillen et al. 2009, 2012).

It is important to highlight the general trade-offs among these different techniques for traditional customer churn analysis. Logistic regression is more parsimonious in the sense of having coefficients that explicitly quantify the effects of predictors on the response variable (churn rate), which one can perform statistical inference on. In contrast, machine learning models often exhibit superior predictive capacity due to

their relative flexibility, but the parameters of (most) machine learning models are related more to controlling the learning process, without necessarily having a practical meaning. Such trade-offs also exist in the context of multi-state customer churn analysis. In this chapter, we adopt multinomial logistic regression (MLR), which is a natural extension of logistic regression to the case of categorical responses with more than two levels, as we prioritise interpretability over predictive capacity (although we note that in our application to the LGPIF data, MLR worked reasonably well in predicting transitions compared to some machine learning techniques for multi-state modelling).

MLR has been used in actuarial science to study various multi-class classification problems. For example, Christiansen et al. (2016) employed MLR to study policyholders' decisions to switch insurance coverage in long-term health insurance, Zhu et al. (2015) used MLR to model the claim rate, lapse rate, and in-force rate jointly in life and annuity insurance, while Alai et al. (2015) and Kwon and Nguyen (2019) applied MLR to model cause-specific mortality. A more relevant work is that of Antonio et al. (2016), who used MLR with a first-order Markov assumption to build a multi-state model for the development of a non-life insurance claim process. However, to the best of our knowledge, this chapter is the first to propose its use with a second-order Markov assumption in customer churn analysis.

3.3 Multi-State Customer Churn Modelling

For constructing time-discrete multi-state models of customer churn, there are two strategies we can adopt in terms of fitting MLR models; see Chapter 11.3 of Frees (2004). First, we can split the data into separate subsets based on each state of origin, and build a separate MLR model for each subset. For example, suppose there is a total of three possible states that a policyholder can be in, namely full-coverage, partial-coverage, and churn, and we want to model the probabilities of transitioning into these states based (only) on the current state. Then we could split the data based on the policyholder's current state, such that we have a subset comprising all observations whose current state is full-coverage, and another subset comprising all observations whose current state is partial-coverage (note we ignore the subset comprising all observations whose current state is churn; see also Figure 3.1 later on). Then, we fit a separate MLR model to each subset, thereby modelling the transition probabilities from these two states of origin.

Alternatively, we can fit a single MLR model to the whole data, and use indicator variables to denote the states of origin when calculating the likelihood contribution.

That is, we can construct interaction terms between each of the explanatory variables and an indicator variable indicating the state of origin (in the above example, either full-coverage or partial-coverage), and then include this expanded set of predictors in the MLR model when calculating the likelihood contribution.

Both the above strategies have their advantages, with the former being computationally more efficient (especially using parallel computing), while the latter is more unified and offers a convenient means of (say) testing the equality of parameters, i.e., the effect of a particular explanatory variable is the same irrespective of the state of origin, and thus simplifying the regression structure. Importantly though, both methods of fitting the MLR model will produce the same estimates, statistical inference, and predictions. With this in mind, we will adopt the second strategy in this chapter and use an expanded set of covariates in what follows.

To trace the path of a policyholder's states over time, let $l_{i,t}$ denote the state of i-th policyholder at current time t, and $\{l_{i,1}, \ldots, l_{i,t-1}\}$ denote the history of i-th policyholder's states from time 1 up to the previous time point $(t-1)$. Throughout the chapter, we assume independence among individuals i for $i = (1, 2, \ldots, N)$, which is standard and widely used in the actuarial studies (see for example Frees et al. 2016; Okine et al. 2022, among others for the LGPIF data specifically) as well as more broadly in the longitudinal data analysis literature; see also our discussion in Section 3.6. To build a MLR model (or any general approach for customer churn analysis as a matter of fact), a key decision to be made is how much of the policyholder's state history is associated with the likelihood of being in their current state. That is, what order a Markov assumption should be assumed for transition probabilities. In actuarial science, by far the most common assumption is that of the first-order Markov property (Kwon and Jones 2006, 2008; Bettonville et al. 2021), meaning that the probability of being in the current state depends only on the state at the previous time point, and not on earlier states. That is, $P(l_{i,t}|l_{i,1}, \ldots, l_{i,t-1}) = P(l_{i,t}|l_{i,t-1})$. For multi-state customer churn analysis, this leads to the following first-order MLR model. Assume there are a total of Q states that the policyholder can transition to, i.e., states of destination. Then for the i-th policyholder, the transition probability from state r at time $(t-1)$ to state s at time t is given by

$$P(l_{i,t} = s|l_{i,t-1} = r) = \frac{\exp\left(\boldsymbol{x}_{i,t-1,r}^{\top}\boldsymbol{\beta}_{rs}\right)}{\sum_{s'=1}^{Q}\exp\left(\boldsymbol{x}_{i,t-1,r}^{\top}\boldsymbol{\beta}_{rs'}\right)}, \tag{3.1}$$

where $\boldsymbol{x}_{i,t-1,r}$ is a vector of explanatory variables for the i-th policyholder who was in state r at time $(t-1)$, with its first element set equal to one to reflect an intercept

term, and the quantity $\boldsymbol{\beta}_{rs}$ denotes the vector of regression coefficients quantifying the effect of the explanatory variables on the transition probability from state r to state s. Note the dependence of explanatory variables on the state of origin reflects the idea that one may wish to, whether based on *a-priori* knowledge or exploratory data analysis, include different sets of explanatory variables for the transitions with different states of origin. Indeed, we adopt this approach in the LGPIF application, e.g., see Table 3.3. At the same time, the dependence on r also reflects the construction of an expanded set of covariates, i.e., interaction terms between each explanatory variable and an indicator variable indicating the state of origin. In traditional customer churn analysis, it is assumed that the probability of churn from $(t-1)$ to t depends only on the explanatory variables $\boldsymbol{x}_{t-1}$ at time $(t-1)$; see Guillen et al. (2003); Brockett et al. (2008). In this chapter, we further allow $\boldsymbol{x}_{i,t-1,r}$ in equation (3.1) to include (values of) explanatory variables previous to $(t-1)$, e.g., in the LGPIF application these include time-variant predictors such as the ratio of total premium of this to previous year minus one $(\boldsymbol{x}_{i,t-1,r}/\boldsymbol{x}_{i,t-2,r}-1)$; see Table 3.2.

When building a MLR model as in equation (3.1), we need to decide on a reference state or reference category among all states of destination $s = 1, \ldots, Q$. All the coefficients corresponding to this reference state will then be set to zero for reasons of parameter identifiability. For instance, if we choose $s = 1$ to be the reference state, then we set $\boldsymbol{\beta}_{r1} = \mathbf{0}$ for all states of origin r, and as such equation (3.1) for this particular state of destination simplifies to $P(l_{i,t} = 1|l_{i,t-1} = r) = \left(\sum_{s'=2}^{Q} \exp\left(\boldsymbol{x}_{i,t-1,r}^{\top}\boldsymbol{\beta}_{rs'}\right) + 1\right)^{-1}$. Note this constraint is consistent with logistic regression, and it is also consistent with the interpretation of the coefficients in a MLR model in terms of log-odds. Following on from the above example, we have $\log\left(P(l_{i,t} = s|l_{i,t-1} = r)/P(l_{i,t} = 1|l_{i,t-1} = r)\right) = \boldsymbol{x}_{i,t-1,r}^{\top}\boldsymbol{\beta}_{rs}$, for $s = 2, \ldots, Q$, meaning $\boldsymbol{\beta}_{rs}$ explicitly quantifies changes in the log-odds of transitioning (from state r) to state s relative to transitioning to state 1.

As discussed above, the first-order Markov assumption is a common one made in the context of (traditional) customer churn analysis. In this chapter however, motivated specially by our exploration of the LGPIF data, we consider a second-order MLR model where the probability of being in the current state now depends upon the states in the previous two time points. More generally, a MLR model of order k implies that the probability of being in the current state depends on the k earlier states i.e., $P(l_{i,t}|l_{i,1}, \ldots, l_{i,t-1}) = P(l_{i,t}|l_{i,t-k}, l_{i,t-k+1}, \ldots, l_{i,t-1})$. Mathematically, a second-order MLR model expands on equation (3.1) as follows. For the i-th policyholder, the transition probability from states (q, r) at time points $(t-2, t-1)$ to

state s at time t is given by

$$P(l_{i,t} = s | l_{i,t-2} = q, l_{i,t-1} = r) = \frac{\exp\left(\boldsymbol{x}_{i,t-1,qr}^{\top} \boldsymbol{\beta}_{qrs}\right)}{\sum_{s'=1}^{Q} \exp\left(\boldsymbol{x}_{i,t-1,qr}^{\top} \boldsymbol{\beta}_{qrs'}\right)}. \tag{3.2}$$

The state of origin is now defined by two prior states (q, r) together, while $\boldsymbol{\beta}_{qrs}$ denotes the vector of regression coefficients quantifying the effect of the explanatory variables on the transition from state of origin (q, r) to state of destination s. We use the second-order MLR model given by equation (3.2) as an example to calculate the likelihood. Following the second-order Markov assumption, i.e., $P(l_{i,t}|l_{i,1}, \ldots, l_{i,t-1}) = P(l_{i,t}|l_{i,t-2}, l_{i,t-1})$, we can write the likelihood function as

$$L = \prod_{i=1}^{N} P(l_{i,1}, l_{i,2}) \times P(l_{i,3}|l_{i,1}, l_{i,2}) \times \ldots \times P(l_{i,T_i}|l_{i,T_i-2}, l_{i,T_i-1}), \tag{3.3}$$

where T_i is the total number of observed states for i-th policyholder. Note that if a policyholder churned during the study period, then we treated the churn state as the last observed state (l_{i,T_i}) for this policyholder. For the second-order MLR model, we focus on those policyholders who have entered their contracts for more than one year and ignore the new policyholders; in practice, these new policyholders are treated differently compared to existing policyholders. Hence, we treat the first two states of each policyholder as given, i.e., $P(l_{i,1}, l_{i,2}) = 1$. Then we can replace the second-order conditional probabilities in equation (3.3) with the second-order MLR formula on the right-hand side of equation (3.2). The likelihood of the second-order MLR model can be written as

$$L(\boldsymbol{\beta}|\boldsymbol{l}, \boldsymbol{x}) = \prod_{i=1}^{N} \prod_{t=3}^{T_i} \prod_{s=1}^{Q} \prod_{r=1}^{Q-1} \prod_{q=1}^{Q-1} \left(\frac{\exp\left(\boldsymbol{x}_{i,t-1,qr}^{\top} \boldsymbol{\beta}_{qrs}\right)}{\sum_{s'=1}^{Q} \exp\left(\boldsymbol{x}_{i,t-1,qr}^{\top} \boldsymbol{\beta}_{qrs'}\right)} \right)^{\mathbb{I}(l_{i,t-2}=q)\mathbb{I}(l_{i,t-1}=r)\mathbb{I}(l_{i,t}=s)}, \tag{3.4}$$

where indicator variables $\mathbb{I}(l_{i,t-2} = q)$, $\mathbb{I}(l_{i,t-1} = r)$, and $\mathbb{I}(l_{i,t} = s)$ identify the states of origin and destination separately of the transition from $(t-2, t-1)$ to t for the i-th policyholder. Note the number of states of origin for q and r is $(Q-1)$ because they exclude the churn (absorbing) state. In equation (3.4), three successive observations $(l_{i,t-2}, l_{i,t-1}, l_{i,t})$ of a policyholder can contribute to the likelihood once, which is the reason that fitting a higher-order MLR model requires more data as the order increases. Note that similar to its first-order counterpart, we again need to select a reference state and set the coefficients corresponding to this reference state to zero for reasons of parameter identifiability, e.g., $s = 1$ and $\boldsymbol{\beta}_{qr1} = \mathbf{0}$ for all (q, r).

By comparing equations (3.1) and (3.2), we observe that a second-order MLR model

greatly expands the states of origin, and subsequently accounts for more of policyholders' past history when determining the likelihood of moving to different states or staying in the same state. If such an assumption is appropriate, then equation (3.2) will generally improve model fitting and predictive power compared to a first-order MLR model. On the other hand, fitting a higher-order MLR model requires more data as the order increases, e.g., policyholders who are observed for less than three time points will not be useful for fitting a second-order MLR model, and we require a bigger sample size overall to ensure we have sufficient data within each state of origin to model the transition probabilities effectively.

To better illustrate the difference between using first and second-order MLR models for multi-state customer churn analysis, we consider an insurance company that offers auto insurance (obliged) and homeowners insurance (optional). A policyholder of this company can reside in three possible states: full-coverage (state 1) where a customer is covered by both auto insurance and homeowners insurance, partial-coverage (state 2) where a customer is covered by auto insurance only, and churn (state 3) meaning a customer drops all insurance coverage and leaves the company. If we adopt the first-order MLR model, then we need only consider two states of origin at time $(t-1)$, namely full and partial-coverages (Figure 3.1 left panel). Since churn is an absorbing state, we do not model transition probabilities from this state. Following the notation in equation (3.1), we have $r = \{1, 2\}$, $s = \{1, 2, 3\}$, and $Q = 3$. On the other hand, if we adopt the second-order MLR model, then we will need to consider four potential states of origin, depending on whether the policyholder had full or partial-coverage at time points $(t-2)$ and $(t-1)$ (Figure 3.1 right panel). Following equation (3.2), we have $q = \{1, 2\}$, $r = \{1, 2\}$, and $s = \{1, 2, 3\}$. It is clear from Figure 3.1 that a larger number of parameters need to be involved in the second-order MLR model: after choosing the reference state, under a first-order Markov assumption, we have four sets of coefficients $\boldsymbol{\beta}_{rs}$ to model, while under a second-order Markov assumption we have eight sets of coefficients $\boldsymbol{\beta}_{qrs}$ to model.

3.4 Application to Data from Wisconsin Local Government Property Insurance Fund

We illustrate an application of multi-state customer churn analysis to commercial insurance data, using a data set publicly available from the Wisconsin Local Government Property Insurance Fund (LGPIF); see Frees et al. (2016), Shi and Yang (2018), Dong, Frees and Huang (2022) among others for previous uses of the LGPIF

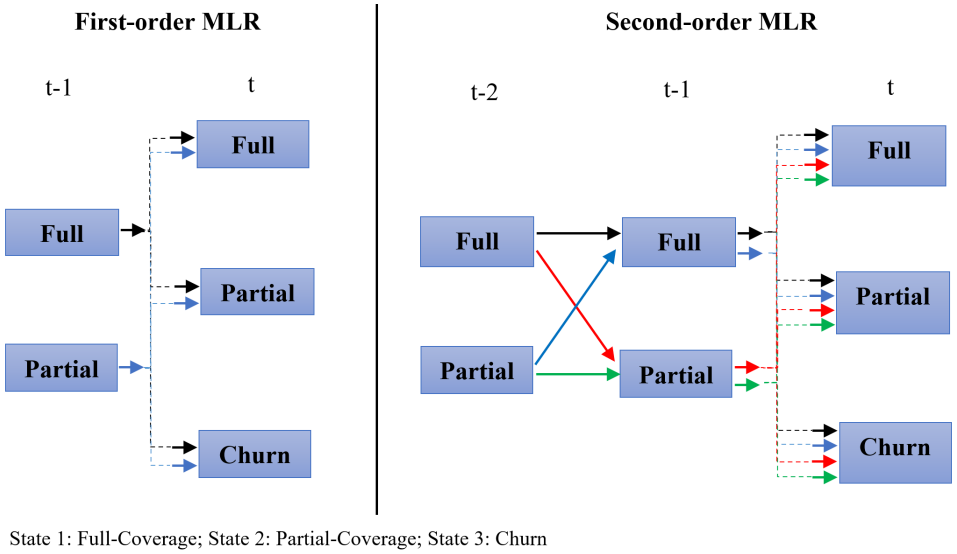

Figure 3.1: Multi-state customer churn analysis: first-order MLR model (left) and second-order MLR model (right).

data in actuarial science. The LGPIF is administered by the Wisconsin Office of the Commissioner of Insurance, with its overarching purpose being to make property insurance available for local government units. The LGPIF offers six coverages for policyholders to choose: building and content (BC), contractor's equipment (IM), comprehensive new vehicles (PN), comprehensive old vehicles (PO), collision new vehicles (CN), and collision old vehicles (CO). Since the last four coverages all represent the vehicle coverage, and also to ensure that there is sufficient data to model the transition probabilities reasonably well, we choose to combine the four vehicle coverages as a single coverage that we refer to simply as "Car" coverage. Only BC coverage is compulsory, while IM and Car coverages are optional for a policyholder.

Based on the above, we set up a three-state transition framework for multi-state customer churn analysis, as exemplified by Figure 3.2. When a policyholder enters a contract, they can take either a full-coverage contract (state 1, which includes BC, IM, and Car coverage) or a partial-coverage contract (state 2, which includes BC plus optional IM or Car coverage). Afterwards, if a policyholder does not adjust their combination of coverages, then the contract will be in the same state in the following period, i.e., they stay in the same state. Alternatively, a policyholder may drop or add certain coverages without churning, which indicates a transition from state 1 to state 2 or vice versa. Finally, if a policyholder decides to cancel all the coverages in a contract, then the state of the contract will move to state 3, i.e., will churn, which is the absorbing state. Note that we did not consider short-term cancellations because such information was unavailable to us. However, we acknowledge

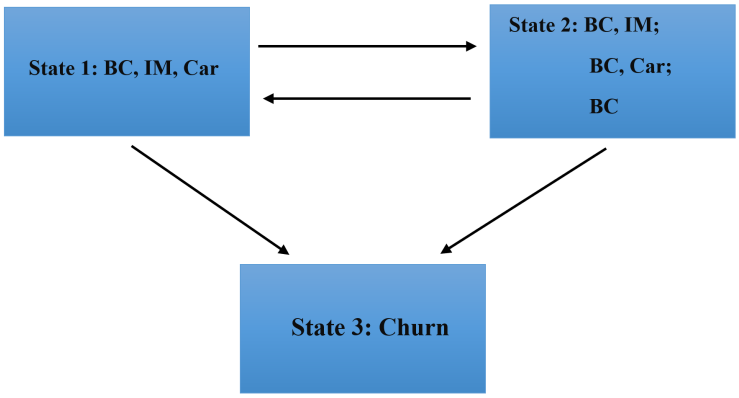

Figure 3.2: A three-state transition diagram of the application to data from LGPIF. BC refers to building and content insurance, IM refers to contractor's equipment insurance, while Car refers to vehicle coverage.

that short-term cancellations could be an important question in some specific situations. For example, in Norway, the insurance regulations changed in 2006, allowing customers to cancel their policies at any time and not only on due date (Günther et al. 2014). For insurers who want to consider short-term cancellations, they could use the multinomial logistic regression introduced in this chapter to model multivariate customer behaviours, including renewal, churn, and short-term cancellations. In addition, because the time of short-term cancellations is random, they could consider building a separate time-continuous model (e.g., Orr 2007; Hürlimann 2015) for short-term cancellations. Doing so helps insurers better understand the reason for short-term cancellations and brings new insights into customer behaviour.

Note state 2, i.e., partial-coverage, actually includes three types of contracts (Figure 3.2); again, the reason we collect these three forms of contracts into one state is to address data imbalance (a very small number of observations for some transitions), and subsequently to ensure that there is sufficient data to model all transition probabilities reasonably well. We refer the reader to Appendix B.1 for a discussion of an expanded five-state transition framework as well as the issue of data imbalance in the LGPIF data.

As part of our exploratory data analysis, Table 3.1 lists the second-order transition data and corresponding empirical transition probabilities from 2006 to 2013. As discussed in Section 3.3, although we will not model new policyholders' behaviour, we list these new policyholders' transitions (referred to as state "0" at time $(t-2)$) in Table 3.1 to offer a complete picture of the data. From time $(t-2)$ to $(t-1)$, there are more observations in the states of origin $(1,1)$ and $(2,2)$, i.e., staying in the same state, compared to changing states. More importantly, we notice that observations staying in the same states over the past *two* years have higher empirical transition

probabilities of staying in the same state next year. For example, one can compare the empirical transition probabilities based on state of origin $(1, 1)$, i.e., having full-coverage in the past two years, with those based on state of origin $(2, 1)$, i.e., having partial-coverage and then full-coverage in the past two years. The empirical transition probability of moving from state of origin $(1, 1)$ to state 3, i.e., churn, is 2.93%, while the empirical transition probability of moving from state of origin $(2, 1)$ to churn is 0%. Furthermore, the empirical transition probability of moving from state of origin $(1, 1)$ to state 2, i.e., partial-coverage, is 2.44%, while the empirical transition probability of moving from $(2, 1)$ to partial-coverage is 11.63%. That is, policyholders who stayed with full-coverage over the past two years are more likely to churn and less likely to move to partial-coverage, relative to policyholders who moved from partial to full-coverage in the past two years. Overall, these initial results suggest that a policyholder's coverage at time $(t-2)$ plays an important role in affecting their behaviour, beyond that seen at time $(t-1)$. With this in mind, we decided to build a second-order MLR model, as exemplified by equation (3.2), to analyse multi-state customer churn in the LGPIF data.

Table 3.1: Second-order transition counts and empirical transition probabilities in per cent (in parentheses) from 2006 to 2013. In each row, the sum of counts represents the total observations in the corresponding state of origin.

	State of destination (t)		
State of origin	State 1	State 2	State 3
$(t-2, t-1)$	(full-coverage)	(partial-coverage)	(churn)
(0, 1)	416 (92.24%)	22 (4.88%)	13 (2.88%)
(1, 1)	2098 (94.63%)	54 (2.44%)	65 (2.93%)
(2, 1)	38 (88.37%)	5 (11.63%)	0 (0.00%)
(0, 2)	14 (1.72%)	780 (95.59%)	22 (2.70%)
(1, 2)	7 (9.33%)	63 (84.00%)	5 (6.67%)
(2, 2)	27 (0.65%)	4015 (96.31%)	127 (3.05%)

Finally, it is important to highlight that we did not consider orders higher than two in our application, primarily because the LGPIF did not contain sufficient information to do so, i.e., the data did not contain enough time points or observations such that we should feasibly fit a higher-order MLR model effectively. For life insurance companies that do have long-term policyholders' data available (say), then an exploration of subsequent higher-order analysis is warranted.

3.4.1 Review of Variables in Customer Churn Analysis

In Table 3.2, we summarise the explanatory variables that we will consider in our multi-state customer churn analysis of the LGPIF data based on a combination of our background knowledge and exploratory data analysis (see also the discussion at the beginning of Section 3.4.2). These explanatory variables are commonly used in customer churn analysis in general insurance, and we refer the reader to Appendix B.2 for a discussion on these variables. Note in particular that we included an indicator variable for the 2008 financial crisis, because it was one of the most important global financial events within the training set and statistically highly significant. As part of our exploratory data analysis for the LGPIF data, we examined several other continuous economic growth variables, such as GDP and federal rate, but found that none of them were statistically significant predictors for the transition probabilities. Also, note that since the LGPIF is a commercial insurance fund targeting government units as policyholders, then variables commonly seen and used in personal insurance, such as age and gender, are not present for inclusion in our analysis.

For the purposes of this chapter, we treat all variables as strictly exogenous, and we do so to be consistent with the existing literature on traditional customer churn analysis (see for instance Guillen et al. 2003; Brockett et al. 2008, among others). We refer the reader to Appendix B.3 for a discussion on exogeneity in the context of general insurance, and leave the issue of treating certain variables as sequentially exogenous/endogenous for future research. For example, in Chapter 4, we treat claims variables as endogenous and build a joint model of claims and customer churn. It is also important to point out that not all of the variables will be used in our final model. Moreover, we will allow different sets of explanatory variables to drive the transition probabilities with different states of origin, i.e., the dependence on the subscripts (q, r) in the covariates $\boldsymbol{x}_{i,t-1,qr}$ in equation (3.2). On the explanatory variables themselves, we calculate and use rates (premium per coverage) instead of using premiums directly for each type of coverage, as we found that the premium of a contract was highly affected by the coverage size of the contract. Since the size of coverage in commercial insurance is relatively large and varies from year to year, then we want to separately study the effect of rate variation and coverage variation on the transition behaviour of policyholders.

3.4.2 Second-Order MLR Model: Estimated Effects

We fitted second-order MLR models to the LGPIF data from years 2006 to 2012, and left year 2013 as a holdout set that we will use later on to assess out-of-sample

Table 3.2: Variable description for our application to the LGPIF data.

Variable name	**Description**
RateBC	Rate of BC, i.e., premium per million BC coverage ($1000 USD)
RateIM	Rate of IM, i.e., premium per million IM coverage ($1000 USD)
RateCar	Rate of Car, i.e., premium per million Car coverage ($1000 USD)
RatioPremium	Ratio of total premium of this to previous year minus one
TotalCoverage	Value of policy coverage (hundred millions USD)
FrequencyBC	Frequency of BC claim
FrequencyIM	Frequency of IM claim
FrequencyCar	Frequency of Car claim
$\mathbb{I}$(ClaimBC)	Indicator of BC claim occurrence (0 = no claim)
$\mathbb{I}$(ClaimIM)	Indicator of IM claim occurrence (0 = no claim)
$\mathbb{I}$(ClaimCar)	Indicator of Car claim occurrence (0 = no claim)
$\mathbb{I}$(Entity: types)	Indicator of types of entity: Ci = city; Co = county; Sc = school; Mi = misc; To = town; Vi = village, e.g., $\mathbb{I}$(Entity: CiSc) = 1 means the entity is either city or school
$\mathbb{I}$(FinaCris)	Indicator of 2008 financial crisis (1 means in 2008)

Note: All the variables take the values at time $(t-1)$, when considering the transition probability from time $(t-1)$ to t.

performance in Section 3.5. We performed exploratory data analysis to assess which explanatory variables to include as part of $\boldsymbol{x}_{i,t-1,qr}$. In particular, we started with all variables listed in Table 3.2, and chose the final set of covariates as the set of explanatory variables which were statistically significant at the 10% level for at least one transition probability (either to partial-coverage or to churn, where full-coverage was set to be the reference state). Note from Table 3.2, there were only a handful of observations with states of origin equal to $(2,1)$ and $(1,2)$. As such, we restricted the number of explanatory variables to two for the transitions with these two states of origin, so as to avoid (excessive) overfitting. We used the `multinom` function from `R` package `nnet` (Venables and Ripley 2002) to fit the second-order MLR model. Note the standard errors calculated are non-robust, as the `multinom` function does not support calculating robust standard errors currently; we refer the reader to Appendix B.4 for the same analysis using Stata 17, which also returns robust standard errors.

Table 3.3 presents the estimated coefficients and corresponding 90% confidence in-

tervals for the explanatory variables included in the fitted second-order MLR model. From the estimated MLR model, we observe that both the set of important explanatory variables and the resulting effect sizes differed greatly depending on the state of origin. This in turn reflected a notion that the second-order states of origin play an important role in distinguishing between different transition probabilities, and we discuss the results for each of the four states of origin separately below.

For policyholders with full-coverage in the previous two time points, i.e., state of origin equal to $(1, 1)$, there was statistically clear evidence of an association between the likelihood of transitioning to partial-coverage and the three coverage rates. In particular, a higher rate of IM was related to a higher transition probability to partial-coverage, while higher rates of BC and Car were associated with a lower transition probability to partial-coverage. Both the occurrence of BC claim and the occurrence of IM claim significantly reduced the likelihood of churn, while the occurrence of BC claim suggested that the customer was likely to drop some coverage (moved to the partial state). That is, we found that different claim types exhibited different effects on transition probabilities, which is an interesting new insight relating to the role of claims in customer churn analysis. The slopes for *TotalCoverage* suggested that big buyers (policyholders with large coverages) were less likely to move to either partial-coverage or churn compared to small buyers. Finally, policyholders who experienced larger total premium increases in the previous two years (i.e., positive values of *RatioPremium*) were significantly more likely to churn the next year, although there was no statistically clear evidence of a similar effect on the likelihood of transitioning to partial-coverage.

For policyholders who went from partial-coverage to full-coverage over the previous two time points, i.e., state of origin equal to $(2, 1)$, the occurrence of Car claim was found to significantly increase the probability of staying in full-coverage. Also, the three coefficients of the transition to churn were all negative, which suggested the predicted probabilities of this type of transition would always be zero. This result was not surprising given such transitions were not found in the training set in the first place; see the transition from $(2, 1)$ to churn in Table 3.1.

For policyholders who dropped from full-coverage to partial-coverage over the previous two time points, i.e., state of origin equal to $(1, 2)$, there was statistically clear evidence of an effect of a higher rate of IM on the transition probability to partial-coverage, and moreover the magnitude of this effect was similar to that of policyholders with state of origin $(1, 1)$.

Finally, for policyholders with partial-coverage over the previous two time points, i.e., state of origin equal to $(2, 2)$, a large number of BC claims was found to be

Table 3.3: The second-order MLR model estimates with 90% confidence interval in parentheses.

	State of destination			
	Partial-coverage (s = 2)		**Churn (s = 3)**	
	Estimate	CI	Estimate	CI
State of origin: $(q = 1, r = 1)$				
Intercept	-23.23	(-25.18, -21.29)*	-21.52	(-23.16, -19.87)*
$\mathbb{I}$(Entity: ScToVi)	19.73	(18.74, 20.72)*	17.01	(16.16, 17.87)*
$\mathbb{I}$(Entity: CiMi)	20.66	(19.63, 21.68)*	16.21	(15.25, 17.18)*
$\mathbb{I}$(ClaimBC)	0.20	(-0.34, 0.74)	-0.64	(-1.26, -0.02)*
$\mathbb{I}$(ClaimIM)	-0.47	(-2.18, 1.24)	-27.55	(-27.55, -27.55)*
RateBC	-2.86	(-4.67, -1.05)*	0.30	(-0.76, 1.35)
RateIM	1.37	(0.16, 2.57)*	0.76	(-0.25, 1.77)
RateCar	-0.39	(-0.61, -0.16)*	-0.04	(-0.25, 0.17)
RatioPremium	0.81	(-0.49, 2.12)	0.93	(0.02, 1.83)*
TotalCoverage	-0.88	(-1.48, -0.27)*	-0.66	(-1.31, -0.00)*
State of origin: $(q = 2, r = 1)$				
Intercept	-15.98	(-16.38, -15.60)*	-16.74	(-16.74, -16.74)*
$\mathbb{I}$(Entity: ScToVi)	14.29	(13.89, 14.69)*	-3.66	(-3.66, -3.66)*
$\mathbb{I}$(ClaimCar)	-11.78	(-11.78, -11.78)*	-2.82	(-2.82, -2.82)*
State of origin: $(q = 1, r = 2)$				
Intercept	20.87	(19.45, 22.30)*	20.84	(19.43, 22.26)*
$\mathbb{I}$(Entity: ScToVi)	-21.24	(-22.68, -19.81)*	-21.01	(-22.43, -19.59)*
RateIM	1.19	(0.07, 2.31)*	-0.07	(-1.23, 1.09)
State of origin: $(q = 2, r = 2)$				
Intercept	23.72	(23.41, 24.03)*	20.26	(19.74, 20.58)*
$\mathbb{I}$(Entity: CiScToVi)	-18.72	(-19.03, -18.40)*	-18.60	(-18.94, -18.27)*
$\mathbb{I}$(FinaCris)	-0.90	(-1.58, -0.20)*	-0.42	(-1.19, 0.35)
FrequencyBC	0.02	(-0.21, 0,25)	-0.71	(-1.16, -0.27)*
RatioPremium	3.25	(0.70, 5.80)*	3.56	(1.00, 6.13)*

Note: * confidence intervals excluding zero

associated with a significantly lower probability of churn (but no effect on the likelihood of transitioning to partial-coverage). Also, policyholders who experienced total premium increases in the previous two years were significantly less likely to obtain full-coverage in the following year. Finally, we found that the 2008 financial crisis only presented a statistically clear negative effect for this particular group of policyholders, as it strongly decreased their likelihood of retaining partial-coverage in the following year.

For all policyholders, we found that entity type played a statistically significant role in driving the transition probabilities, although this differed depending on the precise state of origin. For example, the entity being equal to either school, town, or village, i.e., $\mathbb{I}(\textit{Entity: ScToVi}) = 1$ was found to be associated with a significantly higher probability of churn when the state of origin was $(1, 1)$, but a significantly lower probability of churn when the state of origin was $(2, 1)$, compared to the corresponding reference entity type.

Finally, it should be noted that several of the statistically significant effects estimated from the MLR model were distinctly large (Table 3.3). This potentially reflected a level of overfitting by the second-order MLR model to the LGPIF data. For instance, the estimated pair of effects for *Entity(ScToVi)* when the state of origin was $(1, 2)$ was approximately -21. This large negative coefficient was attributed to the fact that in the LGPIF training data, when the entity type was equal to a school, town, or village, all policyholders with a state of origin $(1, 2)$ retained full-coverage.

The large estimated effects in this application arise more as a feature of the data rather than an actual problem with the MLR model. That is, the sample size of the LGPIF data was not overly large, and its time horizon (2006 to 2012) was not long enough to observe more actual transitions. Importantly, we expect this data problem to be neither common nor serious for most multi-state customer churn analyses, given the majority of insurance companies commonly have (much) more accumulated years of longitudinal data. From a modelling perspective, it is possible to apply regularisation penalties, e.g., the lasso, ridge or bias-correction type penalties, to shrink these coefficients away from large values and hence potentially avoid an excessive degree of overfitting (Madigan et al. 2005), while also ensuring that overall conclusions in terms of statistically significant effects are retained. We leave such extensions of the MLR model for future research, as the main focus of this chapter is on illustrating MLR models as a modelling technique for multi-state customer churn analysis. For readers interested in the goodness-of-fit tests of the fitted MLR model, we provide the results of these in Appendix B.5.

In summary, for the LGPIF data, we have found that premium information, claim

experience, and contract information play statistically significant roles in driving multi-state customer transition probabilities. It is important to highlight that the marketplace we examined in the chapter has not been previously examined in the actuarial customer churn literature, i.e., the LGPIF is a special line of commercial insurance with its overarching purpose being to make property insurance available for local government units. On the one hand, government units differ from personal customers, this is a smaller marketplace with fewer insurance providers, and there may not be readily available alternatives which government customers can choose from (the original motivation for establishing the fund). This lack of availability will certainly affect policyholders' decisions to leave the fund, suggesting that the results from this chapter may not be consistent with other studies that have focused on personal insurance. On the other hand, the complexity of the insurance industry, influenced by factors such as business lines (life/general, personal/commercial), regional laws and regulations, company reputation and service, as well as specific contract terms inherently limit the generalisability of any single dataset.

At the same time, our analysis offers a more complete picture of the different roles that factors play in determining customer behaviour beyond simply whether policyholders churn or not. Indeed, understanding the drivers of customer transition behaviour can assist insurance companies with marketing campaigns, e.g., pushing customers to concentrate all their insurance policies in the same company. For example, the estimation result of *RatioPremium* in Table 3.3 suggests that a small discount (leading to a negative value of *RatioPremium*) may attract policyholders to stay with the insurance company or even purchase additional coverages. An open question for future research is how to use the multi-state customer churn analysis to identify target customers for marketing campaigns and design optimal marketing strategies for insurance companies. Note also that the effects and scope of exploratory variables considered depend on the second-order state of origin, which in turn sheds light on the importance of a higher-order analysis.

3.4.3 Second-Order MLR Model: Visualisation

We examine two types of figures that can be used to better understand the effects of explanatory variables in MLR models, namely conditional effect figures and average marginal effect figures. Briefly, a conditional effect plot quantifies the estimated effect for an individual policyholder (micro-perspective), while an average marginal effect plot quantifies the effect for an entire population of policyholders (macro-perspective).

A conditional effect is based on varying a single predictor (the focal variable) while

fixing all the other variables at certain values, such that it quantifies the effect of the focal variable and ignores how it correlates (and thus might hence) with other variables in practice. As a result, in the context of multi-state customer churn analysis, conditional effect plots are useful for interpreting the model results from the perspective of an individual policyholder. The insurer can draw conditional effect figures by setting the "condition" to reflect the characteristics of a target policyholder, and thus understand the role a focal variable plays in triggering this policyholder's transition behaviour. This in turn can assist the insurer, say, in developing a targeted strategy to maintain a policyholder's loyalty. By default, when varying the focal variable in a conditional effect plot, we can fix all numerical variables at their mean values, and fix all factor variables at their reference levels (say).

To illustrate the use of a conditional effect plot, we consider the second-order MLR model fitted above to the LGPIF data, and target a city entity (policyholder) with partial-coverage over the past two years, i.e., state of origin equal to $(2, 2)$. In this case, there are only four exploratory variables affecting transition probabilities for this policyholder; see Table 3.3. Suppose we want to specifically study how the ratio of total premiums affects this policyholder's transition behaviour. From the resulting conditional effect plot (Figure 3.3 left panel), we observe that, generally speaking, if that target policyholder experiences a larger total premium increase in the previous two years (i.e., positive values of *RatioPremium*), then the likelihood of transitioning to full-coverage and even maintaining partial-coverage decreases, as that targeted policyholder is more likely to churn. This full picture is not immediately obvious when we only examine the corresponding estimated effects in Table 3.3: the estimated coefficients of *RatioPremium* to partial-coverage and to churn are 3.25 and 3.56, respectively. However, as discussed in Section 3.3 these coefficients explicitly quantify the effect of the ratio of premiums on the log-odds relative to full-coverage, and thus only offer an idea of the *relative* changes in transition probabilities. By contrast, from the conditional effect plot we see that, in fact, the probability of transitioning to partial-coverage is concave in shape as a function of *RatioPremium*, while the probability of churn is monotonically increasing.

As an alternative to conditional effects, the average marginal effect calculates each policyholder's effect of transition probabilities from varying a focal variable, and then averages this across the resulting effect estimates among all policyholders. Consequently, this takes into account how variables are naturally correlated with each other as based on the data (Leeper 2017). Average marginal effect figures are useful for interpreting the fitted MLR model at a population level, i.e., macro-perspective

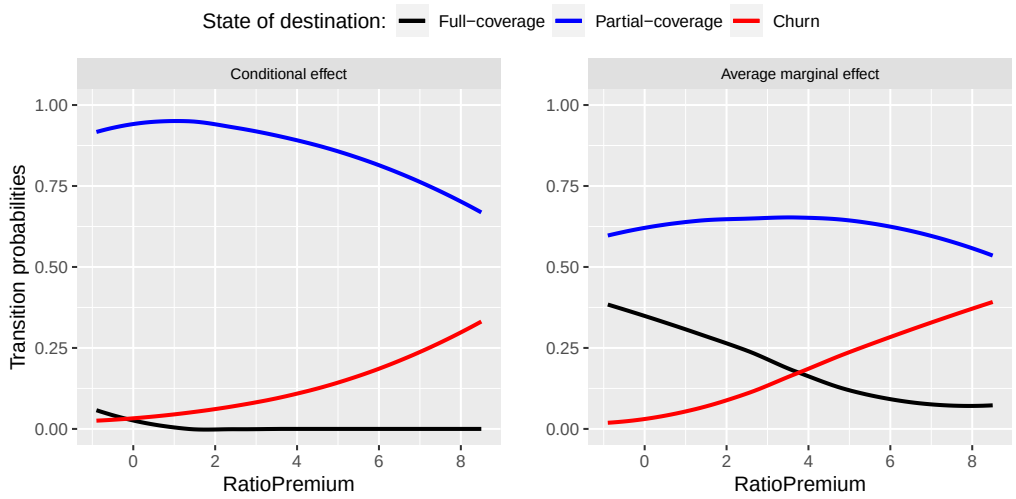

Figure 3.3: The effect of RatioPremium on transition probabilities: conditional effect (left panel) for a city entity with state of origin equal to $(2, 2)$, compared to an average marginal effect (right panel).

for all policyholders. That is, the insurance company can construct average marginal effect plots with different focal variables to understand how the changes affect the transition probabilities across all policyholders currently holding contracts, and in turn, develop general as opposed to personalised strategies to maintain customer loyalty. As an example of this, and to contrast with the conditional effect plot, we again consider the ratio of premiums as a focal variable, but this time study its average marginal effect among all policyholders in the LGPIF data. The resulting plot is shown in the right panel of Figure 3.3, from which we see that the probability of transitioning to partial-coverage is (also) concave in shape as a function of premium ratio, while the probability of churn is (also) monotonically increasing. However, the probability of transitioning to partial-coverage (relative to churn) in the average marginal effect plot is consistently lower compared to that in the conditional effect plot. This difference in probability curves indicates the difference of effects from a focal variable to an individual policyholder versus all policyholders. Notice also that in Table 3.3, the premium ratio was *only* included to model the transition probabilities from states of origin $(1, 1)$ and $(2, 2)$, i.e., policyholders who had the same level of coverage across the two previous years. Interestingly, this alone suffices to induce an effect across the broader population of policyholders with the insurance fund, and the average marginal effect plot of Figure 3.3 reflects this.

3.4.4 Comparing Multi-State and Traditional Customer Churn Analysis

As a comparison between multi-state and traditional customer churn analysis for data from the LGPIF, we use an example based on calculating customer lifetime value (CLV). At the end of each year before the renewal date, an insurer may construct models as part of their customer churn analysis, which are then used to calculate policyholders' CLV, the discounted future income stream derived from acquisition, retention, and expansion projections and their associated costs (Gupta et al. 2004). As a simple illustration of this, we consider an "average" policyholder for a two-year time horizon, defined as a customer whose numerical explanatory variables are set equal to the mean values from the data, and whose factor variables are set equal to their modal categories. All exploratory variables are then fixed for two years. We further assume that the policyholder has had full-coverage over the past two years, i.e., state of origin equal to $(1, 1)$, and we use customer churn models to assess their behaviour over the next two years. For readers interested in the precise values of exploratory variables, we refer to Appendix B.6. We also point out that all settings in this section were purposefully designed to make the illustration relatively simple yet insightful. In future applications of the proposed MLR model for multi-state customer churn modelling, it is perhaps recommended to adopt richer techniques such as generalised linear models or machine learning techniques to predict claims and interest rates for the CLV calculation. For example, in Chapter 4 , we build a joint model of claims and customer churn for a more accurate CLV calculation.

We calculate the CLV of this average policyholder for the next two years via multi-state customer churn modelling and traditional customer churn modelling. The formula for the CLV of a policyholder is given by $\text{CLV} = \sum_{t=1}^{T} \text{P}_t(\text{Revenue}_t - \text{Cost}_t)/(1+\text{Discount Rate}_t)^t$, where Revenue_t and Cost_t denotes the revenue and cost respectively for a customer at time t, P_t denotes the probability that the customer still has a valid contract up to time t, and Discount Rate_t is interest rate used to discount cash flow at time t. For illustrative purposes, we assume Revenue_t is the premium paid in dollar value by the policyholder at time t, and that it is fixed at the state-specific average value in the data set. Moreover, Cost_t is the expected claim cost in dollar value from the policyholder after paying deductibles, and it is also fixed at the state-specific average value in the data set. We fix Discount Rate_t at 1.78%, i.e., the annual average federal funds rate from 2006 to 2012 in the LGPIF data. Table 3.4 lists the values of these quantities according to the state of the policyholder. For a policyholder in full-coverage, the revenue is higher than the expense. However,

the revenue is lower than the expense for a policyholder in partial-coverage.

Table 3.4: CLV calculation: state-specific revenue, state-specific expense, and interest rate.

	Revenue_t	Cost_t	Discount Rate_t
Full-coverage	\$34,556	\$28,283	1.78%
Partial-coverage	\$12,268	\$12,904	1.78%

For traditional customer churn analysis, we assume the insurance company will build a second-order binary logistic regression (BLR) model to predict the probability of churn for this policyholder in the next two years, while for multi-state customer churn analysis, we use the second-order MLR model fitted in Section 3.4.2. Note that one limitation of one-versus-all BLR models is that the transition probabilities from the same state of origin are independently modelled, and thus need not and in general do not sum to one. We refer the reader to Appendix B.7 for the illustration of the second-order one-versus-all BLR models in the context of multi-state customer modelling. Importantly, multi-state customer churn analysis focuses on all types of transition among states (see Figure 3.2), while traditional customer churn analysis only focuses on the transition to churn. Consequently, there are three scenarios to consider if the insurer performs traditional customer churn analysis, but seven scenarios to consider if the insurer performs multi-state customer churn analysis. We refer to Appendix B.8 for the visualisation of all scenarios for traditional and multi-state customer churn analysis.

Table 3.5 presents results of CLV calculations comparing traditional and multi-state customer churn analysis. In the case of the former, the CLV over the three possible scenarios was \$2,917.4. In the latter, we notice that when the (average) policyholder's future path is to partial-coverage for both years or to partial-coverage followed by churn, the expected present value is negative. This is because when the policyholder takes partial-coverage (in state 2), the revenue is less than the expense; see Table 3.4. Across the seven possible scenarios, the CLV of the policyholder calculated by multi-state customer churn analysis was \$2,844.5.

Through Table 3.5, we can see the difference in calculating CLV between traditional versus multi-state customer churn analysis, which in case ends up being \$72.9. In practice, customer lifetime is commonly longer than two years, and so it is likely that an even larger difference in CLV between the two approaches would be observed. Furthermore, while we have considered an "average" policyholder, insurers are typically more interested in those who pay a lot of premiums annually, i.e., big buyers.

Table 3.5: Scenarios of the future path and CLV calculation: traditional customer churn analysis versus multi-state customer churn analysis.

Traditional customer churn analysis (second-order BLR)

Scenario	*Probability (per cent)*	*Expected present value*
(full, full)	92.60%	\$2835.7
(full, churn)	3.63%	\$81.8
(churn)	3.77%	\$0.0
Sum	**100%**	**\$2917.4**

Multi-state customer churn analysis (second-order MLR)

Scenario	*Probability (per cent)*	*Expected present value*
(full, full)	89.15%	\$2730.3
(full, partial)	1.68%	\$36.5
(full, churn)	3.58%	\$80.7
(partial, full)	0.27%	\$1.5
(partial, partial)	1.32%	\$-4.1
(partial, churn)	0.20%	\$-0.5
(churn)	3.80%	\$0.0
Sum	**100%**	**\$2844.5**

The difference in CLV between the two approaches will again likely be larger for this select group, relative to what we saw in this example. To summarise, because multi-state customer churn analysis considers all possible transitions among states when calculating CLV, and since policyholders in different states naturally have different values for insurers, then we believe such an analysis will generally improve the accuracy of CLV calculations. Ultimately, this is relevant to the profitability of insurance companies (Neslin et al. 2006), as it helps them better distinguish valuable customers and trivial customers.

3.5 Assessing Out-of-Sample Performance

Using the LGPIF data, we assessed the out-of-sample prediction performance of the proposed second-order MLR model, and compared it with several other models for predicting customer transition/churn rates. In particular, we also considered a first-

order MLR model as formulated in equation (3.1), along with a first-order and a second-order one-versus-all BLR models (see Appendix B.7). We also included two non-parametric methods that are widely used in traditional customer churn analysis (He et al. 2020), namely gradient boosting machines (GBMs) and support vector machines (SVMs). As both can be applied for multi-class classification problems also, then they can be straightforwardly adapted for use in multi-state customer churn analysis with the first-order assumption. We refer to Appendix B.9 for details of applying GBMs and SVMs to the LGPIF data.

All methods were fitted to LGPIF data from years 2006 to 2012, and out-of-sample validation was performed on data in year 2013. In terms of assessing performance, a widely-used evaluation metric in traditional customer churn analysis is error rate = (1−accuracy), which is defined as the percentage of incorrectly classified observations in the test set (see for example Bolancé et al. 2016; Loisel et al. 2019; Scriney et al. 2020). However, a major disadvantage of error rate is that it is not well suited to imbalanced data with rare events (with imbalanced classes, it is easy to get a low error rate without actually making useful predictions, Morrison 1969), and this is the case here with the LGPIF data, as exemplified in Table 3.1. As such, we require alternatives to error rate, and here we consider the area under the receiver operating characteristic curve (AUC, Cortes and Mohri 2003). The AUC is better suited for assessing classification accuracy in imbalanced data, since it is independent of the choice of classification threshold and avoids choosing the largest estimated probability for the multi-class classification; see Hand and Till (2001) and He and Ma (2013). In this chapter, we calculate both multi-class AUCs and the standard binary-class AUCs for model assessment. Additionally, we calculate the top-decile lift, which is a well-established evaluation metric in traditional customer churn research of marketing, and is noted for its ability to target critical customers (Pendharkar 2009; De Bock and Van den Poel 2011). We demonstrate how to make use of the top-decile lift for multi-state customer churn analysis.

Note we compare different models' predictive power using widely accepted metrics, namely AUCs and the top-decile lift. But we also acknowledge they are not always the best choices depending on the specific setting. For instance, Guelman et al. (2012) suggests that insurers should develop models and metrics to instead identify the target customers who are likely to respond positively to a customer retention campaign. Designing and choosing appropriate metrics in multi-state customer churn analysis for insurers, in order to maximise their utility, is an interesting topic that we leave for future research.

3.5.1 AUC

While AUC is well established for assessing binary classification, here we consider two other types of AUC for multi-class classification problems, namely micro-AUC and macro-AUC (Wu and Zhou 2017). Macro-AUC is derived from the receiver operating characteristic curve based on macro-average true positive and false positive rates (TPR and FPR), where the macro-average is defined by calculating TPR and FPR independently for each class and then taking the average. Because macro-average quantities treat all classes equally, then they better reflect the statistics of the smaller classes, and so are more appropriate when performance on all the classes is equally important. By contrast, micro-AUC is derived from the receiver operating characteristic curve based on micro-average TPR and FPR. The micro-average pools the contributions of all classes before constructing TPRs and FPRs. As such, micro-average quantities are dominated by the more frequent class.

Table 3.6 presents the binary AUCs for each type of transition, along with the micro- and macro-AUCs for all six methods considered. Additionally, we consider applying the methods both to the original training set, and to a balanced training set constructed through oversampling. Since the main aim of this chapter is not about optimising a model's predictive performance, then we used the `upSample` function from `R` package `caret` (Kuhn 2021) to oversample the training set based on the state of destination in order to deal with data imbalance issue. We ensured the numbers of transitions with different states of destination (full-coverage, partial-coverage, and churn) were equal by adding additional samples to the minority transitions with replacement. In contrast to oversampling, undersampling is also widely used in customer churn analysis when a data set includes a large number of records; see Scriney et al. (2020). However, as discussed in Section 3.4, the LGPIF data does not contain enough time points or observations for this to be relevant. Note for each of the BLR and MLR models, we used explanatory analysis to determine the appropriate regression forms based on the original training set (explanatory variables were chosen if they were statistically significant at the 10% level for at least one transition probability). We then used these forms in the models fitted to the balanced training set. For GBMs and SVMs, we tuned the relevant hyperparameters using the original training set and the balanced training set separately.

For methods fitted to the original training set (Table 3.6 top half), there is no single-best method across all AUCs. The second-order BLR model had the highest AUC to churn and macro-AUC, which indicates that it performed well at predicting the likelihood of customer churn specifically. If we compare the two second-order models with the two first-order models, then the former was consistently better

Table 3.6: Out-of-sample validation: AUCs of six models (top: original training set; bottom: balanced training set).

		AUC (to full)	AUC (to partial)	AUC (to churn)	Macro-AUC	Micro-AUC
Original training set	Second-order MLR	0.990	0.974	0.690	0.885	0.987
	First-order MLR	0.985	0.971	0.651	0.869	0.984
	Second-order BLR	0.990	0.979	**0.691**	**0.887**	0.987
	First-order BLR	0.985	0.977	0.651	0.871	0.984
	SVM	0.993	0.976	0.676	0.882	0.988
	GBM	**0.994**	**0.980**	0.654	0.876	**0.988**
Balanced training set	Second-order MLR	0.989	**0.980**	**0.700**	**0.890**	0.980
	First-order MLR	0.985	0.976	0.648	0.870	0.984
	Second-order BLR	0.989	0.979	0.696	0.888	0.978
	First-order BLR	0.985	0.977	0.645	0.869	**0.984**
	SVM	0.972	0.880	0.532	0.795	0.907
	GBM	**0.991**	0.975	0.643	0.870	0.982

Higher AUC indicates better performance.

than the latter in terms of AUC. This further supports the notion that, for MLR models, a second-order Markov assumption leads to more accurate classifications of customer transition. The GBM had the highest AUC to full-coverage, AUC to partial-coverage, and micro-AUC. When we compare methods fitted in the balanced training set (Table 3.6 bottom half), the second-order MLR model had the highest AUC to partial-coverage, AUC to churn, and macro-AUC, while GBM had the highest AUC to full-coverage. Again, if we compare the two second-order models with the two first-order models, then the former had higher AUCs than the latter in all cases except micro-AUC.

Lastly, when comparing methods across the original and balanced training sets, we observe that the second-order models tended to achieve higher AUC to churn in the balanced training set. However, the overall prediction performance of the other four models became worse in the balanced training set compared to the original training set. This suggests that, at least for the LGPIF data, balancing the data set for second-order models had the potential to improve their prediction performance.

3.5.2 Top-Decile Lift

Top-decile lift is commonly used in customer churn analysis to focus on the most critical group of customers and their churn risk (see for example Neslin et al. 2006; Holtrop et al. 2017; Devriendt, Berrevoets and Verbeke 2021). Broadly speaking,

those with the highest 10% predicted churn rates in a test set present the riskiest customers and subsequently an ideal segment for targeting a customer retention program. Top-decile lift is then calculated as the proportion of churners in this riskiest segment, divided by the proportion of churners in the whole test set. It thus acts as a practical measure to compare different methods for predicting customer churn by assessing their relative abilities to identify the set of riskiest customers. Put another way, a predictive method with a high top-decile lift gives confidence to insurers that they can still reach a majority of all potential churners by targeting the top 10% riskiest segment only (Neslin et al. 2006).

We extend the idea of top-decile lift to multi-state customer churn analysis. That is, we calculate top-decile lift and target the corresponding top 10% riskiest segments for transitions with different states of destination. In detail, for a given state of destination s, we first identify the 10% riskiest segment, i.e., the policyholders with the highest 10% of predicted transition probabilities to that state of destination. The top-decile lift is then given by $\text{TDL}_s = \hat{\pi}_s^{10\%}/\hat{\pi}_s$, where $\hat{\pi}_s^{10\%}$ is the proportion of the actual transitions to state s in the corresponding 10% riskiest segment, and $\hat{\pi}_s$ is the proportion of the true transitions to state s across the entire test set. A higher value of top-decile lift reflects a better capacity to predict the riskiest policyholders for a particular type of transition.

Table 3.7 presents the results from calculating top-decile lifts to full-coverage, partial-coverage, and churn in both the original and balanced training sets of the LGPIF data. GBM and SVM had the highest top-decile lift to churn in the original training set, while the second-order MLR model had the highest top-decile lift to churn in the balanced training set. Both first- and second-order MLR models performed equivalently in terms of top-decile lifts to churn in the original training set, but when fitted in the balanced training set, the second-order MLR model performed noticeably better than its first-order counterpart. A similar conclusion can be drawn if we compare first- and second-order BLR models. This again supported the use of the second-order Markov assumption for customer churn analysis in the LGPIF data.

Finally, when we compare Tables 3.6 and 3.7, we can see that balancing the training set had the potential to improve a model's out-of-sample predictive performance, especially for predicting transitions to churn. This reflected the fact that, in the original training set, the transitions to churn were in the minority, and dominated by transitions to full- and partial-coverage (Table 3.1). When we balanced the original training set through oversampling, we oversampled the transitions to churn, and so their corresponding weight was increased when fitting models, thus generally

Table 3.7: Out-of-sample validation: top-decile lifts of six models (top: original training set; bottom: balanced training set).

		TDL (to full)	**TDL (to partial)**	**TDL (to churn)**
Original training set	Second-order MLR	**3.09**	1.50	0.87
	First-order MLR	**3.09**	1.51	0.87
	Second-order BLR	**3.09**	1.51	0.87
	First-order BLR	**3.09**	1.51	0.87
	SVM	3.03	1.51	**1.74**
	GBM	**3.09**	**1.53**	**1.74**
Balanced training set	Second-order MLR	**3.06**	1.51	**2.61**
	First-order MLR	**3.06**	1.51	1.74
	Second-order BLR	**3.06**	1.51	2.17
	First-order BLR	**3.06**	1.51	1.30
	SVM	2.97	1.50	0.00
	GBM	**3.06**	**1.53**	1.74

Higher TDL indicates better performance.

improving out-of-sample predictions for this minority class.

3.6 Discussion

In this chapter, we have proposed the concept of multi-state customer churn analysis for commercial insurance. To study policyholders' behaviour beyond just churn/not churn and across multiple periods, we develop multinomial logistic regression (MLR) models with a second-order Markov assumption. Through the use of such models, which favour interpretability over predictive power, insurance companies can gain a complete picture of how a policyholder's past states affect their decision to move to the next state, as well as the complex relationships between transition probabilities and various explanatory variables related to premium information, claim information, and contract type. To the best of our knowledge, this is the first chapter (paper) to employ a second-order MLR model for customer churn research in actuarial science.

We present an application of multi-state customer churn analysis by fitting second-order MLR models to investigate policyholders' behaviour on data from the Wisconsin LGPIF. Empirical analysis demonstrates that explanatory variables play dif-

fering roles in triggering different types of transitions, and this in turn leads to a more nuanced understanding of how a policyholder's previous history and attributes are associated with their decision to future behaviour. We present conditional and average marginal effect plots to interpret the estimated effects on the probability scale. Insurers can intuitively understand the effect of a focal variable on different types of transitions for a specific policyholder using the former, while the latter allows the study of a broader customer group or population of policyholders. We use the example of an average policyholder to illustrate the difference in calculating customer lifetime value between multi-state and traditional customer churn modelling, the latter based on one-versus-all binary logistic regression (BLR) models. Because policyholders in different states naturally have different values for insurers, we believe that multi-state customer churn analysis will improve the accuracy of customer lifetime value calculation for insurers as a whole.

Although prediction is not the primary aim of this chapter as a whole, the second-order MLR model was shown to predict relatively well. Specifically, compared to a first-order MLR model, BLR models, GBMs and SVMs, the proposed second-order models performed strongly in terms of out-of-sample AUCs and top-decile lift. We also find that balancing the data through oversampling the minority groups tended to improve the predictive performance of most methods, although the improvement for second-order models was more substantial compared to their first-order counterparts in the LGPIF data. Overall, our findings open up a new way of perceiving and understanding customer behaviour in the insurance market, and shed light on the importance of multi-state customer churn modelling.

The multi-state, multi-period modelling used in this chapter presents a new approach to churn modelling, and so naturally, there are several variations that warrant further investigation. First, we have implicitly treated all variables as strictly exogenous, which means that we assume all variables are completely unaffected by the customers' state transitions in the past, present, and future. For future research, variables can be treated as sequentially exogenous, which indicates that current or future customers' state transitions would cause variables to change in the future. For example, premium and claim could be treated as sequentially exogenous variables, an approach that may provide insights into contractual moral hazard and adverse selection. Second, we assume that the effect of each covariate is fixed across all time points. Future research could examine the use of time-varying coefficients instead for multi-state customer churn analysis, perhaps building on the dynamic survival model approach of Guillen et al. (2012). Allowing for time-varying coefficients would enhance the flexibility of the model, and potentially improve prediction performance.

However, it would generally require more data to estimate than fixed coefficients models, noting also that the records in our motivating LGPIF data are in discrete time, in contrast to the time-continuous approach of Guillen et al. (2012).

Third, the assumption that all policyholders are independent is not only questionable in our application, but also in studies of personal homeowners and automobiles where things being insured i.e., houses and cars, are close to one another, sharing common environmental hazards such as hail, windstorm and so forth. Some research has been being done to accommodate these dependencies, such as the spatial autocorrelation approach (Brechmann and Czado 2014) and the more recent marked point process approach (Shi et al. 2022). Both of these, among others, may serve as valuable reference points for relaxing such assumption in future developments of multi-state customer churn modelling. Fourth, we have shown that multi-state customer churn analysis could improve the accuracy of customer lifetime value calculation for insurers. Future research could use multi-state customer churn analysis to improve the accuracy of loss reserving, i.e., predicting the ultimate loss amount at the aggregate level with the consideration of multiple states. Fifth, we only focus on those policyholders who have entered their contracts for more than one year and ignore the new policyholders in this chapter. Consequently, all the selected customers have at least two years of observations (states), making it easier to employ the second-order MLR model. If insurers want to integrate new and existing customers into one model, one could modify the queuing theory approach proposed by Boucher and Couture-Piché (2015, 2016) to support multi-state transitions. For example, instead of using queuing theory to model the number of insured cars per household (Boucher and Couture-Piché 2015), we could use the theory to model the number of coverages per customer.

Sixth, if prediction is the primary goal of a customer churn analysis, then a better-performing classifier can be obtained from a balanced sample with appropriate sampling techniques. Future research could examine the use of bias correction methods such as intercept correction and weighting correction (Lemmens and Croux 2006), along with regularisation and feature selection techniques (e.g., extending the work of Devriendt, Antonio, Reynkens and Verbelen 2021, to support higher-order MLR models). Fifth, for higher-order MLR models, the value of order can be regarded as a hyperparameter that (also) needs to be tuned for optimising the predictive performance. In practice though, building higher-order models (whether they be parametric or more machine-learning based techniques) ideally requires a large data set with many time points, and demands close collaboration between academia and industry. Finally, there are a plethora of advanced multi-class classification models

available in the machine learning literature, such as convolutional neural networks and XGBoost, which have already been used in traditional customer churn analysis (He et al. 2020; Loisel et al. 2019). Applying these to build higher-order models for multi-state customer churn analysis presents an important and challenging avenue for future research.

Chapter 4

Valuing Customers in the Insurance Industry: A Joint Model for Claims and Churn

Yumo Dong, Edward W (Jed) Frees, Fei Huang, Francis K. C. Hui, Harald J. van Heerde

This chapter has a working paper version awaiting submission.

4.1 Introduction

4.1.1 Background: CLV and Costs

Customer lifetime value (CLV) quantifies the present value of projected profits a firm expects from a customer throughout their relationship. It takes projected revenues, costs, relationship duration (retention), and discount rate into consideration. Using CLV in customer relationship management helps firms optimise marketing resources, enhance retention strategies, and ensure long-term profitability (see Venkatesan and Kumar 2004; Kumar et al. 2008; Kumar and Reinartz 2016, 2018; Ascarza et al. 2018, among others). A prevailing view in the literature related to customer management suggests that CLV-based marketing strategies generally outperform those solely based on retention (see Venkatesan and Kumar 2004; Ascarza 2018; Ascarza et al. 2018; Lemmens and Gupta 2020, for example). Indeed, trying to retain a customer with a negative CLV might be a waste of resources, and instead it might be better to let such a customer go.

CLV has been studied and implemented in various industries, such as high-technology services, retailing, financial services, and airlines (see Kumar and Reinartz 2016, for an overview). What is common in these industries is that the marginal costs to serve customers on an ongoing basis are relatively low and/or stable (e.g., the marginal cost for serving broadband internet subscribers or the marketing cost to reach airline customers). However, the nature and magnitude of costs to serve customers is very different in the insurance industry. The major cost component for insurance companies are insurance claims, which are highly variable and can be extremely large in size. While many policyholders may have zero claims in most years, when they do have to submit a claim (e.g., when their car is wrecked or their house is damaged), claims can run into hundreds of thousands or millions of dollars. The challenge for CLV modelling in the insurance industry is thus to model these (claim) costs adequately, as these costs are heterogeneous across customers and are unknown at the point of sale.

Arguably as a result of this challenge, the application of CLV in the insurance industry remains sparse (Donkers et al. 2007), despite the fact this industry accounts for more than 7% of the global economy (Jimenez 2022). As insurance has evolved over time, insurers have become more customer-centric, and their customers (policyholders) are able to buy a variety of insurance products in the market to gain protection against different losses. However, given the stochastic nature of the cost component and their potential magnitude, it remains challenging for insurance providers to value their customers using CLV approaches that can not capture the claims risk

adequately.

Another challenge is that there may be an association between a customer's intrinsic claim risk and churn risk. That is, a policyholder who has a high intrinsic claim risk may be less likely to churn. Such a customer (with private information about their claim risk) may believe it is better to stay insured with the current insurance provider than to switch providers, as other providers may not be willing to insure the customer at the same conditions and premiums as the current provider. Consequently, the CLV modelling framework needs to account for a link between a customer's intrinsic claims risk and churn risk.

4.1.2 Existing CLV Models

Existing CLV modelling frameworks (Table 4.1 in Section 4.2) are not suitable for insurance companies. They focus on revenue and churn predictions, but, for the insurance industry setting, they greatly oversimplify cost assumptions (e.g., assuming a low, steady cost). In doing so, they may potentially produce inaccurate CLV calculations and hence ineffective managerial decisions based on CLV for insurance companies using these models. We contend that the cost assumptions in current CLV studies fall short of capturing individual customer's claims risks in the insurance industry, especially considering the diverse risks in claim frequency (number of claims) and severity (claim amount).

4.1.3 Our CLV Modelling Framework

To bridge this gap, this chapter introduces a new cost-based CLV modelling framework tailored to the insurance industry, and sheds light on the importance of cost modelling in CLV calculation. Our framework brings together actuarial science modelling of insurance claims, and CLV modelling in marketing science. For claims modelling, it uses a multivariate Tweedie regression, and allows for an association between the customers' claims outcomes and churn decisions through shared random effects. In an empirical application to insurance data, we show that this model fits and predicts better than existing CLV models in marketing, and offers insurance managers additional substantive insights on the most and least profitable customers.

To clarify the role of claims in CLV calculations, we express the CLV of the i-th

customer in insurance as

$$\text{CLV}_i = \sum_{t=1}^{T} \frac{\text{Premiums}_{i,t} - \text{Claims}_{i,t}}{(1 + \text{Discount Rate})^t} \times \text{Retention Rate}_{i,t}, \tag{4.1}$$

where premiums and claims respectively represent the expected revenues and costs in the standard time-discrete CLV formula (see Dwyer 1997; Pfeifer and Carraway 2000; Lewis 2004; Gupta and Lehmann 2005; Zhang et al. 2010, among others). Since CLV is a forward-looking metric, we generally need to predict premiums, claims, and retention rates and use their expected values to calculate CLV. For an insurance contract with a deductible, Claim = max(Ground-up Loss − Deductible,0), where the ground-up loss is a customer's actual loss, and the deductible is the amount paid by the customer themself before the insurance company starts to pay. Note there are other cost components such as administrative expense and tax that insurance companies could consider in their CLV calculations. However, those components are unavailable to us, so we decided not to include them in cost computation following the literature (for example, see Guillen et al. 2003; Brockett et al. 2008). Unlike traditional industries where costs are relatively low and under managerial control, claim costs in the insurance industry are stochastic, hard to predict, and not under managerial control. As a result of their potential magnitude, claims costs decisively influence the profitability of the insurance companies, thereby making claims the central concern for insurers.

Claims data, as a valuable asset for insurance companies, are distinct from standard cost data in other industries, and are generally recorded at the individual customer level with detailed summary statistics. Traditionally, historical claims data have been used for "insurance ratemaking" (the process to determine premiums; see Frees, Gao and Rosenberg 2011; Shi et al. 2015; Garrido et al. 2016; Frees 2018, for examples), and for estimating CLV at the individual level (Seyerle 2001). More recently, actuarial studies have investigated the mass at zero and heavy-tailed features of claims data, demonstrating the importance of modelling claims for insurance companies from various perspectives (e.g., Frees et al. 2016; Lee and Shi 2019; Yang and Shi 2019). As many insurance companies insure different risks (e.g., cars and homes), claims made by policyholders of multiple policies are not just univariate but rather multivariate. Leveraging the latest developments in actuarial sciences (e.g., Shi 2016; Frees et al. 2016; Kurz 2017; Frees et al. 2021), we employ multivariate Tweedie regression (defined as a Poisson sum of gamma random variables) to capture these multivariate claims.

Predicting customer churn lies at the heart of any attempt to estimate CLV in cus-

tomer relationship management (Berger and Nasr 1998; Gupta et al. 2004; Venkatesan and Kumar 2004; Fader and Hardie 2016), so it is important to consider whether the customer churn and claims processes are independent or dependent within our proposed cost-based CLV modelling framework. Several studies use past claims as exogenous predictors to predict the customer churn behaviour in insurance (see Guillen et al. 2003; Brockett et al. 2008; Guillen et al. 2009; Haugen and Moger 2016, for example), while emergent research has started simultaneously modelling customer churn and claims (Jeong et al. 2018; Frees et al. 2021). However, a shortcoming of these studies in the actuarial literature is that they only address the (extrinsic) observed heterogeneity, and ignore the (intrinsic) unobserved heterogeneity in claims and churn risk across customers. Our work advances this literature by developing a joint model for claims and churn, utilising so-called shared random effects in the form of random intercepts that are shared between the Tweedie model for claim risk and the logit model for customer churn. This allows for positive or negative associations between a customer's intrinsic claim risk and churn risk.

Applying our proposed shared random effects model to a commercial insurance data set, we illustrate how this model can capture latent tendencies for claims and churn risks simultaneously. The model offers novel managerial implications for CLV studies, and enhances CLV estimation and prediction accuracy compared to benchmark models. The empirical results also present an instructive avenue for other industries aiming to refine their cost modelling techniques. Notably, it is the first, to our knowledge, to apply multivariate Tweedie regression with random effects in the joint modelling of longitudinal and time-to-event outcomes. While the Tweedie regression has not been used in marketing before, it has been applied successfully in modelling health care utilisation cost (Kurz 2017) and youth bullying prevention (Boulton and Williford 2018), among other fields.

To summarise, the two key contributions of this chapter are: 1) it introduces a CLV modelling framework that is suitable for contexts such as the insurance industry, where costs are unknown at the point of sale and are a large and stochastic component of an individual customer's CLV; 2) this modelling framework allows managers to obtain new marketing-relevant insights that are not possible without the framework: what drives claims (and churn), how are the intrinsic claim risk and churn risk associated, which customers are most profitable after taking their cost into account, and how to conduct customer management to realise CLV growth.

The remainder of this chapter is organised as follows. Section 4.2 reviews the relevant literature on CLV and joint modelling frameworks. Section 4.3 formulates a joint model of multivariate claims and customer churn through shared random effects, and

explains the estimation process. Section 4.4 illustrates an insurance application of our proposed model and reports the main empirical findings. Section 4.5 compares our proposed model's predictive performance with several alternative models, and discusses how our proposed model brings new managerial implications into CLV analyses. Section 4.6 provides concluding remarks.

4.2 Literature Review

In CLV research, various models have been developed to estimate CLV in both contractual (lost-for-good) and noncontractual (always-a-share) settings. In contractual settings, firms are aware of customer churn, unlike in noncontractual settings. Table 4.1 provides an overview of selected CLV literature in contractual and noncontractual settings. The CLV studies listed in this table have focused on modelling customer revenue and churn, which are two essential components of CLV as can be seen in equation (1). However, only a handful of CLV studies have modelled cost, and they have only done so in non-contractual settings. Rust et al. (2011), Kumar et al. (2008), and Venkatesan et al. (2007) modelled the frequency distribution of cost, for example, the number of marketing messages sent to customers. In these papers, the cost severity (e.g., cost per marketing message) is assumed constant, which is a reasonable assumption for costs related to specific marketing activities.

By contrast, in the insurance industry, to model CLV appropriately, it is important to account for the cost of insurance claims by their customers. These costs differ in frequency (number of claims per customer per year) and severity (one claim can be a small amount, and another claim can be a huge amount). Unlike other studies, our individual-level CLV calculations capture both cost frequency and severity, along with customer revenue and churn (see the first row in Table 4.1).

As Table 4.1 also shows, most CLV studies account for some form of association between the CLV components. However, what sets our study apart is that we model the association between churn risk and cost, which has not been modelled in CLV studies before. More broadly, a common feature of existing joint models used in marketing is capturing the unobserved heterogeneity in customer behaviour (see Borle et al. 2008; Kumar et al. 2008; Fader and Hardie 2010; Ascarza and Hardie 2013; Papies et al. 2017; Ascarza et al. 2018; Ascarza 2018, among others). In actuarial studies, while unobserved heterogeneity in claims risk has been studied through credibility theory (e.g., Frees et al. 1999; Boucher and Denuit 2006; Cohen and Einav 2007), little attention has been paid to the unobserved heterogeneity in customer churn risk. Our proposed shared random effects model can capture the observed

Table 4.1: Summary of selected customer lifetime value literature.

Representative studies	Individual-specific CLV components: Revenue	Churn rate	Cost frequency	Cost severity	Association between CLV components	Key contribution/insight
Contractual setting						
Current chapter (paper)	✓	✓	✓	✓	✓	Cost modelling is essential for CLV calculation in insurance.
Lemmens and Gupta (2020)	✓	✓			✓	A profit-based loss function to predict the financial impact of a retention intervention.
Ascarza (2018)		✓				Customers having the highest risk of churning are not necessarily the best targets for proactive churn programs.
McCarthy et al. (2017)	Aggregate	Aggregate			✓	A framework for valuing subscription-based firms that focuses on the firm's acquisition and retention processes.
Boucher and Couture-Piché (2016)	✓	✓				A queuing theory based approach to model the number and global value of insured households in an insurance portfolio.
Datta et al. (2015)	✓	✓			✓	Explores how free-trail and regular customers differ in terms of their decision to retain the service.
Ascarza and Hardie (2013)	✓	✓			✓	An integrated model of usage and retention that fits two-clock contractual settings.
Schweidel et al. (2011)	✓	✓			✓	A hidden Markov model to identify latent states governing customers' affinity for available services.
Fader and Hardie (2010)	Aggregate	✓				Failing to recognise cohort-level retention-rate dynamics leads to biased estimates of the residual value of a customer.
Borle et al. (2008)	✓	✓			✓	A hierarchical Bayes approach to estimate CLV by jointly modelling purchase timing, purchase amount, and risk of defection.
Donkers et al. (2007)	✓	✓			✓	Studies the capabilities of a range of models to predict CLV in the insurance industry.
Lewis (2005)	✓	✓			✓	A programming-based approach to creating optimal relationship pricing policies.
Gupta et al. (2004)	Aggregate	Aggregate				Demonstrates how to value firms by CLV.
Rust et al. (2004)	✓	✓			✓	A framework enabling competing marketing strategy options to be traded off based on projected financial returns.
Bolton et al. (2000)	✓	✓			✓	Investigates conditions in which a loyalty program will have a positive effect on customer evaluations, behaviour, and repurchase intentions.
Noncontractual setting						
Rust et al. (2011)	✓	NA	✓		✓	An approach to predict customer profitability in future periods that perform substantially better than naive models.
Kumar et al. (2008)	✓	NA	✓		✓	Shows the implementation of CLV at IBM and its effects on profitability and resource allocation.
Venkatesan et al. (2007)	✓	NA	✓		✓	A customer selection framework that can aid managers in enhancing marketing productivity and estimating return on marketing actions.
Venkatesan and Kumar (2004)	✓	NA			✓	A framework to maintain/improve customer relationships through optimal allocation of marketing resources and maximise CLV.

Aggregate: aggregate level CLV components

heterogeneity in claims and churn risks by covariates, as well as the unobserved and correlated heterogeneity between claims and churn by shared random effects.

Shared random effects models have received considerable attention over the past two decades, particularly for survival analysis in clinical studies (Wulfsohn and Tsiatis 1997; Ibrahim et al. 2010; Ferrer et al. 2016; Hickey et al. 2016). For example, researchers in clinical studies collect a patient's health status (alive/dead/ill) and health indices (also known as biomarkers) repeatedly, similar to tracking customer churn and other behaviours in marketing studies. Researchers who use shared random effects models then generally believe there may be some latent factor underlying a person's characteristics that simultaneously influences both longitudinal (e.g., biomarkers) and time-to-event (e.g., death, recovery) outcomes of the person.

Similarly, marketing researchers suggest latent factors might simultaneously affect customer churn and other behaviours, such as purchase timing, amount, and service usage (Borle et al. 2008; Ascarza and Hardie 2013; Datta et al. 2015). As elaborated on in this chapter, shared random effects models offer a nuanced approach to jointly model customer churn behaviour and insurance claims. In addition, it has been well demonstrated in extensive literature that modelling longitudinal and time-to-event outcomes by shared random effects models can lead to improvements in the efficiency of statistical inferences, reduce bias, and improve predictions (Rizopoulos et al. 2014; Ferrer et al. 2016; Hui and Bondell 2021), and we will see that this is also the case in our empirical study in Sections 4.4 and 4.5.

4.3 Proposed Joint Modelling Framework

In this section, we introduce our joint modelling framework for multivariate claims and customer churn risks. The outputs of the proposed joint model are multi-period predictions of churn probabilities and claim costs, which make up the inputs to subsequent calculations of CLV. As the calculation of customer revenue (premiums) as the third CLV input requires some context about the application, we will defer the discussion of premium revenue prediction to Section 4.5.

The proposed joint model comprises two sub-models: one for the multivariate claims outcome, and the other for the customer churn outcome. We use an association structure based on shared random effects to link the two sub-models. Although the formulation of the shared random effects model can be viewed from a Bayesian hierarchical modelling perspective, in this chapter we focus on likelihood-based estimation and inference.

4.3.1 Shared Random Effects Model

Claims Sub-Model.

Insurers have become more customer-centric and offer to cover multiple risks under a bundled contract, leading to the need for multivariate claims modelling (see Frees et al. 2016, 2021; Jeong et al. 2023). The distribution of claims typically contains a mass at zero representing no claims, along with a continuous component for positive values representing the claim amount. Moreover, the rare but high-claim customers contribute to the heavy-tailed nature of the claim distribution. To model claims adequately, we employ multivariate Tweedie regression, a widely used claims modelling technique in actuarial science (see for example Jørgensen and Paes De Souza 1994; Shi 2016; Frees et al. 2016).

The Tweedie distribution can be defined as a Poisson-sum-of-gamma random variables, so it naturally captures the mass at zero by its Poisson component and the right-skewed feature of cost data by its gamma component (Bonat and Kokonendji 2017). Because the Tweedie distribution belongs to the linear exponential family, then it is straightforward to incorporate covariates via a generalized linear modelling (GLM) framework, e.g., to identify the factors that affect both the cost frequency and severity.

Note instead of a Tweedie regression to model claims, firms could also work with two sub-models, one for frequency and one for severity. This is commonly known as the frequency-severity approach (see Braun et al. 2006; Shi et al. 2015; Alemany et al. 2020; Bolancé and Vernic 2020, among others). However, this approach doubles the number of parameters compared to the Tweedie regression approach. More broadly, the literature is split between the simpler Tweedie approach and the more complex (but arguably more flexible) frequency and severity approach. Since our emphasis is on incorporating features such as unobserved and correlated heterogeneity, then we focus on the simpler Tweedie approach in the below; however in the empirical application section, we also investigate shared random effects models based on the frequency-severity approach for comparison purposes.

Let $y_{i,t,k}$ denote the claim outcome of the k-th insurance category for the i-th customer at time t, for $i = 1, \ldots, N$ individuals, $k = 1, \ldots, K$ insurance categories (e.g., building, car), and $t = 1, \ldots, T_i$ time periods (e.g., years). The Tweedie variable $y_{i,t,k} = \text{Claims}_{i,t,k}$ has the following density function:

$$f(y_{i,t,k}; \mu_{i,t,k}, p_k, \phi_k) = a(y_{i,t,k}, \phi_k) \exp\left[\frac{1}{\phi_k}\left(\frac{y_{i,t,k}\mu_{i,t,k}^{1-p_k}}{1-p_k} - \frac{\mu_{i,t,k}^{2-p_k}}{2-p_k}\right)\right], \tag{4.2}$$

where $1 < p_k < 2$ is the power parameter, $\phi_k > 0$ is the dispersion parameter, and $\mu_{i,t,k}$ is the expected value of $y_{i,t,k}$. We denote the vector of power and dispersion parameters for all K claims outcomes by $\boldsymbol{p} = (p_1, \ldots, p_K)$ and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_K)$, respectively. The normalising constant $a(y_{i,t,k}, \phi_k)$ requires numerical evaluation, and we refer to Dunn and Smyth (2008) along with the `R` package `Tweedie` package (Dunn 2022) for details of its evaluation and estimating the Tweedie density more generally.

In actuarial settings, $\mu_{i,t,k}$ is referred to as the pure premium (Frees et al. 2016), which is the expected claim cost for the k-th insurance category made by the i-th customer at time t. To model this, we employ a generalised linear mixed model (GLMM) approach,

$$\mu_{i,t,k} = \exp\left(\boldsymbol{x}_{i,t,k}^{\top}\boldsymbol{\beta_k} + b_{i,k}\right), \tag{4.3}$$

such that $\boldsymbol{x}_{i,t,k}$ is a vector of covariates capturing observed heterogeneity, $\boldsymbol{\beta_k}$ is the corresponding vector of regression coefficients for the claim risk of the k-th insurance category, and $b_{i,k}$ is an individual-level random intercept capturing unobserved heterogeneity in the claim risk of the k-th insurance category across customers. Note we only include random intercepts in this chapter. That is, we allow each customer to have their own intercept but keep their slopes homogeneous. Capturing unobserved heterogeneity by random intercepts is widely used in marketing research related to customer management (e.g., Borle et al. 2008; Kumar et al. 2008; Datta et al. 2015). However, the inclusion of random slopes further increases the model complexity and generally requires a sufficient number of observations per customer for estimation. The insurance application in this chapter only has five observations (years) per customer for model fitting (training), so we do not consider random slopes and leave it as an avenue for future research.

Let the vector of individual-specific random effects for all K claims outcomes be denoted by $\boldsymbol{b}_i = (b_{i1}, \ldots, b_{iK})$. We assume $\boldsymbol{b}_i$ follows a K-dimensional multivariate normal distribution with zero mean vector and a covariance matrix $\boldsymbol{\Sigma}$. The correlation parameters in $\boldsymbol{\Sigma}$ capture the dependence among multivariate claims outcomes. Note other distribution choices for random effects are possible, but we choose to use the multivariate normal distribution given it is a standard choice in many longitudinal studies and mainstream statistical software packages (Hickey et al. 2016)

Customer Churn Sub-Model.

For the binary customer churn outcome, we utilise logistic regression given its prevalence among many customer churn analyses (e.g., De Caigny et al. 2018; Lu et al. 2012; Frees et al. 2021; Dong, Frees, Huang and Hui 2022), although we also incor-

porate random effects into our model. Specifically, the proposed model with shared random effects is

$$P(l_{i,t}=1|l_{i,t-1}=0)=\frac{\exp\left(\boldsymbol{x}_{i,t,\text{churn}}^{\top}\boldsymbol{\delta}+\sum_{k=1}^{K}\alpha_k b_{i,k}\right)}{1+\exp\left(\boldsymbol{x}_{i,t,\text{churn}}^{\top}\boldsymbol{\delta}+\sum_{k=1}^{K}\alpha_k b_{i,k}\right)}, \tag{4.4}$$

where $l_{i,t}$ represents the customer churn outcome at the end of time t ($l_{i,t}=1$ for churn, and 0 otherwise), $P(l_{i,t}=1|l_{i,t-1}=0)$ denotes the probability of churn for the i-th customer at time t conditional on entering/renewing their contract at time $t-1$, and $\boldsymbol{\delta}$ is a vector of coefficients corresponding to the covariates $\boldsymbol{x}_{i,t,\text{churn}}$. Note that the covariates $\boldsymbol{x}_{i,t,\text{churn}}$ in the churn sub-model can differ from the covariates $\boldsymbol{x}_{i,t,k}$ in the claims sub-model. This is important since the covariates used in the claims sub-models are highly regulated by authorities to avoid discrimination in insurance ratemaking (Avraham 2017; Frees et al. 2021; Xin and Huang 2023), while the covariates used in customer churn analysis is not regulated and can be freely chosen by the firm or researcher.

The term $\sum_{k=1}^{K}\alpha_k b_{i,k}$ is referred to as the association structure, and reflects the latent association between claims (for insurance categories $k=1,...,K$) and churn tendencies across customers as captured by shared random effects. We denote the association parameters in the churn sub-model as $\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_K)$, and the vector of all coefficients in the claim and churn sub-models as $\boldsymbol{\beta}=(\boldsymbol{\delta},\boldsymbol{\beta_1},\ldots,\boldsymbol{\beta_K})$. For the former, the quantity α_k can be interpreted as the association parameter capturing the association of the unobserved heterogeneities in the k-th claim and churn risks: holding all other covariates $\boldsymbol{x}_{i,t,\text{churn}}$ constant, $\exp(\alpha_k)$ is the multiplicative change in the odds of churn for every one unit change increase in $b_{i,k}$. The sign of α_k indicates how the intrinsic k-th claim and churn risks are associated: a negative (positive) sign of α_k indicates that a customer with a higher intrinsic k-th claim risk is associated with a lower (higher) intrinsic churn tendency. We further discuss the sign of association parameters and their practical significance in the empirical application.

Note that the multi-state customer churn analysis proposed in Chapter 3 is not incorporated into Chapter 4. Primarily, this stems from the insufficiency of observations in the LGPIF data to estimate shared random effects for multi-state transitions.

Together, Equations (4.3) and (4.4) lead to a shared random effects model. Equation (4.3) models the evolution of customer claims outcomes over time, while Equation (4.4) models customer churn behaviour. The random effects are shared between

multivariate claims and churn sub-models, capturing unobserved heterogeneity in multivariate claims and churn risks simultaneously. Overall, while this chapter mainly focuses on insurance claims and customer churn risks, the proposed model contributes to the broad research area of longitudinal data analysis. In particular, it provides a flexible joint modelling framework that captures observed heterogeneity by covariates and unobserved and correlated heterogeneity by shared random effects for multivariate longitudinal data. Since our proposed model allows multivariate longitudinal data to accommodate different scales of measurement, including continuous, discrete, and categorical observations, it can be potentially useful for various actuarial studies.

4.3.2 Model Fitting

We use maximum likelihood estimation to fit the proposed shared random effects model. Note to our knowledge, there is no existing software that supports fitting the proposed model. This is not too surprising since this chapter is the first to employ a multivariate Tweedie model in the joint modelling of time-to-event (churn) and multivariate longitudinal outcomes (claims).

Conditional on the shared random effects, the K multivariate claims outcomes are independent of each other. As such, the expression for the likelihood function of the proposed model is given by

$$L(\boldsymbol{\theta}, \boldsymbol{b}) = \prod_{i=1}^{N} \int \left[\prod_{t=1}^{T_i} \left(\prod_{k=1}^{K} f(y_{i,t,k}|b_{i,k}, \boldsymbol{\theta}) \right) P(l_{i,t}|l_{i,t-1}, \boldsymbol{b}_i, \boldsymbol{\theta})^{l_{i,t}} \big(1 - P(l_{i,t}|l_{i,t-1}, \boldsymbol{b}_i, \boldsymbol{\theta})\big)^{1-l_{i,t}} \right] f(\boldsymbol{b}_i|\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{b}_i,$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{p}, \boldsymbol{\phi}, \boldsymbol{\alpha}, \boldsymbol{\Sigma})$ denotes the full parameter vector in our model, and we assume independence among policyholders. Intuitively, the claims outcomes do not affect our ability to observe churn outcomes (conditional on shared random effects). The converse is not true; churn outcomes do affect our ability to observe claims (i.e., when a customer has churned, no claims can be observed). To estimate $\boldsymbol{\theta}$, we maximise $L(\boldsymbol{\theta}, \boldsymbol{b})$, and in this chapter we do so via the Laplace approximation given its relative computing efficiently and scalability in dealing with multiple random effects (Tuerlinckx et al. 2006; Kristensen et al. 2016). Laplace-approximated maximum likelihood estimation remains widely in use across mainstream statistics, as exemplified by popular `R` functions such as the `glmer` function in the `lme4` package (Bates et al. 2015), the `glmmTMB` function in the package of the same name, the `glmm.admb` function in the `glmmADMB` package (Fournier et al. 2012), and the `gam` function in the `mgcv` package (Wood 2011). Even with the Bayesian data analysis, the Laplace approximation remains a popular approach for scalable approximate

estimation of the posterior distribution, as exemplified by the Integrated Nested Laplace Approximation, which is popular for spatial analysis.

In particular, applying the Laplace approximation to the integral in the likelihood function, we obtain

$$\begin{aligned}\log L(\boldsymbol{\theta}, \boldsymbol{b}) &= \sum_{i=1}^{N} \log \left\{ \int \left[\prod_{t=1}^{T_i} \Big(\prod_{k=1}^{K} f(y_{i,t,k}|b_{i,k}, \boldsymbol{\theta}) \Big) \; P(l_{i,t}|l_{i,t-1}, \boldsymbol{b}_i, \boldsymbol{\theta})^{l_{i,t}} \big(1 - P(l_{i,t}|l_{i,t-1}, \boldsymbol{b}_i, \boldsymbol{\theta})\big)^{1-l_{i,t}} \right] f(\boldsymbol{b}_i|\boldsymbol{\theta}) \; \mathrm{d}\boldsymbol{b}_i \right\} \\ &= \sum_{i=1}^{N} \log \left\{ \int \exp\big(M(\boldsymbol{b}_i)\big) \mathrm{d}\boldsymbol{b}_i \right\} \\ &\approx \sum_{i=1}^{N} M(\hat{\boldsymbol{b}}_i) + \frac{K}{2}\log(2\pi) - \frac{1}{2}\log\det\Big(\boldsymbol{H}_i(\hat{\boldsymbol{b}}_i)\Big).\end{aligned}$$

In the above,

$$M(\boldsymbol{b}_i) = \sum_{t=1}^{T_i}\sum_{k=1}^{K} \log(f(y_{i,t,k}|b_{i,k}, \boldsymbol{\theta})) + \sum_{t=1}^{T_i} l_{i,t} \log(P(l_{i,t}|l_{i,t-1}, \boldsymbol{b}_i, \boldsymbol{\theta})) + (1 - l_{i,t}) \log(1 - P(l_{i,t}|l_{i,t-1}, \boldsymbol{b}_i, \boldsymbol{\theta})),$$

$\hat{\boldsymbol{b}}_i$ maximises $M(\boldsymbol{b}_i)$ for a given value of $\boldsymbol{\theta}$, and $\boldsymbol{H}_i(\hat{\boldsymbol{b}}_i)$ denotes the negative Hessian matrix of $M(\boldsymbol{b}_i)$ evaluated at $\hat{\boldsymbol{b}}_i$. To estimate $\boldsymbol{\theta}$ through the above Laplace approximate likelihood function, we iterate between an inner and outer optimisation process, a summary of which is in the provided Algorithm 1; see also (Tuerlinckx et al. 2006; Kristensen et al. 2016).

Algorithm 1 Maximum likelihood estimation via the Laplace approximation.

Initialisation: $\boldsymbol{\theta}^{(0)}$
repeat
  Inner optimisation:
  for i in $1 : N$ **do**
    $\hat{\boldsymbol{b}}_i^{(s+1)} = {}_{\boldsymbol{b}_i} M(\boldsymbol{b}_i)$ given $\boldsymbol{\theta}^{(s)}$
  end for
  Outer optimisation:
  $\boldsymbol{\theta}^{(s+1)} = {}_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta}, \hat{\boldsymbol{b}}^{(s+1)})$ given
  $s = s + 1$
until convergence criterion satisfied e.g., $\|\boldsymbol{\theta}^{(s+1)} - \boldsymbol{\theta}^{(s)}\| < \epsilon$ for some small ϵ.

4.4 Empirical Analysis: Wisconsin Local Government Property Insurance Fund Commercial Insurance Data

4.4.1 Data and Variable Descriptions

The Local Government Property Insurance Fund (LGPIF) was managed by the Wisconsin Office of the Commissioner of Insurance, and offered property insurance to local government entities such as cities, counties, school districts, or fire departments. Within a bundled insurance contract, the LGPIF offered two primary categories: building & content (henceforth abbreviated to "building") and equipment & cars (henceforth "equipment").

It is important to clarify in advance that our primary objective is to show the universality of our proposed model rather than to draw a general conclusion from the LGPIF dataset. On the one hand, the complexity of the insurance industry, influenced by factors such as business lines (life/general, personal/commercial), regional laws and regulations, company reputation and service, as well as specific contract terms, inherently limits the generalisability of any single dataset. On the other hand, the LGPIF data used in this chapter is very unique in actuarial research but common for insurance companies, which simultaneously tracks policyholders' premiums, claims, and contract renewal information over the years (from 2006 to 2012 in this case). We used this dataset as an illustrative example to show how to deploy our proposed shared random effects model. Considering the complexity of the insurance industry, there are reasons to believe that different insurers might get varied modelling outcomes and business insights after deploying our proposed model on their own data.

From exploratory data analysis, we observe a consistent trend where in each year, renewers make higher claims than churners on average (see bottom panel of Table 4.2). This trend hints at a negative correlation between claim risk and churn risk, noting that at this stage, we have not yet taken the influence of covariates and shared random effects into account.

In Table 4.3, we outline the variables considered in the proposed shared random effects model. These variables are commonly used for claims and churn modelling in the actuarial studies related to the LGPIF data (see Frees et al. 2016; Shi and Yang 2018; Quan and Valdez 2018; Shi and Lee 2022, among others). For the claims sub-model, the two responses are *ClaimBuilding* and *ClaimEquipment*, which are the ground-up losses reported by customers. Ground-up losses are any losses a cus-

Table 4.2: Customer churn (top panel) and claims (bottom panel) summary statistics.

	2006	2007	2008	2009	2010	2011	2012
Customer churn summary statistics							
Number of customers	949	938	922	905	900	883	853
Number of churners	22	26	41	14	23	39	19
Churn rate	2.3%	2.8%	4.4%	1.5%	2.6%	4.4%	2.2%
Building and equipment average claims summary statistics ($)							
Renewers' building claim	21,400	18,082	12,688	11,904	26,012	25,363	21,599
Churners' building claim	1,158	5630	5296	336	188	2805	282
Renewers' equipment claim	4,548	4,469	4,607	4,063	3,827	3,513	4,947
Churners' equipment claim	137	1839	276	354	827	307	174

Number of customers is counted at the beginning of the year.

tomer experiences (e.g., due to storm damage), irrespective of whether they exceed the deductibles. As this was a government insurer in a relatively small marketplace, LGPIF administrators enjoyed some freedom not found in other markets. In particular, customers were required to report all losses, not only losses that exceeded a deductible. In this way, the administrators had an extra mechanism to verify claim accuracy and their appropriate adjustments. The expected ground-up losses are transformed into the expected net claim (after applying any deductibles) for CLV calculation using Equation (4.1). For the insurers who do not know about small losses that are below deductibles, they could directly model the (post-deductible) claims with a specific deductible choice.

The independent variables for the claims sub-model are *Tenure*, *log(Cover)*, *log(Deduct)*, *Rate*, *NoClaimCredit*, *ICoverEquipment*, and *TypeEntity*. *TypeEntity* is a covariate indicating the type of a local government entity being insured, and includes cities and counties (City), towns (Town), villages (Village), school districts (School), fire departments, and other miscellaneous entities (Misc). For example, a city entity may need coverage for its snow plowing trucks, in addition to its building and contents. Note that *Cover*, *Deduct*, *Rate*, and *ICoverEquipment* are claim-specific independent variables, only used in the corresponding (building or equipment) claims sub-model.

For the customer churn sub-model, the dependent variable is the indicator for customer churn. Note that we did not consider short-term cancellations because such information was unavailable to us. For the insurers who want to take short-term cancellations into consideration, one could use the multinomial logistic regression (MLR) introduced in Chapter 3 to model multivariate customer behaviours, including renewal, churn, and short-term cancellations. Note that this proposed MLR

Table 4.3: Variable descriptions for the Local Government Property Insurance Fund data.

Variable name	Description	Min	5th percentile	Median	95th percentile	Max	Mean	SD
Dependent Variables								
ClaimBuilding	Building claims (ground-up losses $1,000) .	0.00	0.00	0.00	55.01	6,615.12	19.03	168.41
ClaimEquipment	Equipment claims (ground-up losses $1,000).	0.00	0.00	0.00	22.47	570.42	4.18	22.14
Churn	Indicator for customer churn, and P(Churn=1)=2.9%.							
Independent Variables								
Tenure	Length of time that the policyholder has been insured (years).	0.00	0.00	3.00	6.00	6.00	2.83	2.00
CoverBuilding	Building coverage amount (million $).	0.0004	0.39	15.62	162.36	2,468.84	44.78	117.12
CoverEquipment	Equipment coverage amount (million $).	0.005	0.04	0.50	14.98	136.21	2.59	7.87
DeductBuilding	Building coverage deductible level ($1000).	0.50	0.50	1.00	15.00	100.00	3.77	8.86
DeductEquipment	Equipment deductible level ($1,000).	0.50	0.50	2.50	152.50	160.00	69.06	74.78
RateBuilding	Rate of building coverage ($ per thousand coverage).	0.06	0.28	0.58	0.94	36.72	0.61	0.77
RateEquipment	Rate of equipment coverage ($ per thousand coverage).	0.20	1.75	2.39	4.40	6.40	2.68	0.82
NoClaimCredit	Indicator for no building claims in the prior two years, and P(NoClaimCredit=1)=34.5%.							
ICoverEquipment	Indicator for having equipment coverage, and P(ICoverEquipment=1)=96.7%. It includes two scenarios: equipment only and equipment + car.							
TypeEntity	The type of entity: Misc (6.5%), School (27.4%), Town (16.3%), Village (25.8%), and City (24.0%).							

Number of customers in each year

Year	2006	2007	2008	2009	2010	2011	2012
Observations	949	938	922	905	900	883	853

model also includes shared random effects, which makes the model estimation more challenging. The benefit of building such a MLR model with shared random effects is to gain additional insights into the (unobserved) association between claim risk and short-term cancellation risk across customers.

The independent variables are *Tenure*, *log(Deduct)*, *log(Rate)*, *NoClaimCredit*, and *TypeEntity*. Note that we did not use competitors' premiums as covariates in the churn sub-model because such data was not available to us. There are several examples in the literature where marketplace prices (competition) are available, but this is in personal insurance (see Bikker and Van Leuvensteijn 2008; Frank and Lamiraud 2009, for example), where risks are more homogeneous. We know of no study where this is true in the commercial marketplace (Werner and Modlin 2010). However, if the competitors' premium data was available, we could follow Verschuren (2022)'s work and include a covariate representing the competitiveness of the renewal offer before any rate changes (A) relative to the cheapest competitor (B), i.e., the competitiveness is measured by $(B - A)/A$.

4.4.2 Estimation Results

Tables 4.4 and 4.5 summarise the estimation results for the claims and customer churn sub-models. An immediate and important finding we make is that, even after accounting for the effects of the covariates, there is a notable association between claims and churn risks captured by the shared random effects.

Results for Claims Sub-Model.

For the claims sub-model, we summarise our main findings from Table 4.4 into five aspects. First, a positive association exists between the log of coverage amount and claims risk ($\hat{\beta}_{\text{building}} = 0.920$, $p < 0.01$; $\hat{\beta}_{\text{equipment}} = 1.084$, $p < 0.01$). This is reasonable given a higher coverage amount indicates a greater risk has been insured for buildings, contents, equipment, and cars, thus reflecting a potential for higher claims. Second, the relationship between the deductible level and claims risk is complex: a higher deductible level means customers have to pay more by themselves in case they have a claim before the insurance company starts to pay. By offering customers different deductible levels, insurance companies can sort customers based on their risk tolerance and willingness to bear a portion of the loss, which in turn could mitigate the risk of adverse selection (Van de Ven and Van Praag 1981; Puelz and Snow 1994; Cohen and Einav 2007). Generally speaking, customers with a higher claims risk tend to choose a lower deductible level, so they do not need to pay as much for the claims by themselves (note our government database includes all claims, including those below the deductible, which is uncommon for commercial insurers). This helps explain the negative coefficient of *logDeductBuilding* ($\hat{\beta}_{\text{building}} = -0.171$, $p < 0.01$) in the sub-model for building claims. However, when customers choose a lower deductible level, they have to pay a higher premium to the insurance company in exchange. Hence, the choice of deductible level reflects a game between customers and the insurance company. If a customer thinks the (future) insurance benefit brought by a lower deductible cannot cover the additional premium they have to pay, then choosing a higher deductible is optimal. Therefore, it is also plausible to find a positive coefficient of *logDeductEquipment* ($\hat{\beta}_{\text{equipment}} = 0.266$, $p < 0.01$), indicating that customers with high equipment claim risks tend to choose high deductibles to avoid paying high premiums.

A third interesting finding is that the coefficient of *NoClaimCredit* is marginally statistically significant and positive for building claims ($\hat{\beta}_{\text{building}} = 0.176$, $p < 0.1$), indicating customers are more likely to make building claims after they receive a no-claim credit for their building insurance. A possible reason is that customers might strategically delay reporting building claims until they receive a no-claim credit, which can maximise their utility through a way of cheating (Shi and Lee 2022). We note that the company does not offer a no-claim credit for equipment, but nevertheless we include *NoClaimCredit* (for building) in the equipment claims sub-model to test if the credit has an impact on equipment claims outcome. Unsurprisingly, we find that the credit for building category has no significant impact on equipment claims ($\hat{\beta}_{\text{equipment}} = -0.147$, $p > 0.1$).

A unique aspect of the equipment category is that not all customers are required to insure equipment and cars. For example, 96.7% of customers buy insurance for equipment without covering cars (see also Table 4.3). The negative coefficient of *ICoverEquipment* ($\hat{\beta}_{\text{equipment}} = -0.968$, $p < 0.01$) in the equipment claims sub-model suggests that customers with equipment coverage are less risky than customers with cars coverage only.

Finally, we found a modest positive correlation (correlation between random intercepts= 0.125) between intrinsic building and equipment claims risks, indicating that a customer with a larger building claims risk is also likely to have a larger equipment claims risk. The association between different claims risks under a bundled insurance contract is commonly found to be positive in actuarial studies on claims. A possible reason for this is that the objects being insured (buildings and equipment) are close to one another, and share common environmental hazards such as hail, windstorms, and so forth (Shi and Valdez 2014; Frees et al. 2016; Yang and Shi 2019; Frees et al. 2021).

Table 4.4: Parameter estimates for the claims sub-model in our proposed shared random effects model.

Claims sub-model					
Building claims			**Equipment claims**		
	Estimate	S.E.		Estimate	S.E.
Intercept	7.673	0.502**	Intercept	4.050	0.504***
TypeMisc	-0.937	0.282***	TypeMisc	-1.916	0.524***
TypeSchool	-0.880	0.138***	TypeSchool	-0.272	0.276
TypeTown	-0.777	0.271***	TypeTown	-1.311	0.300***
TypeVillage	-0.337	0.172**	TypeVillage	-0.452	0.247*
Tenure	-0.020	0.018	Tenure	-0.062	0.018***
NoClaimCredit	0.176	0.093*	NoClaimCredit	-0.147	0.118
logCoverBuilding	0.920	0.058***	logCoverEquipment	1.084	0.073***
logDeductBuilding	-0.171	0.054***	logDeductEquipment	0.266	0.037***
RateBuilding	-0.544	0.351	RateEquipment	0.264	0.083***
			ICoverEquipment	-0.968	0.363***
Tweedie-power	1.59			1.44	
Tweedie-dispersion	198.05			368.47	
SD of random intercept	1.198			1.508	
Correlation between random intercepts			0.125		

Signif. codes: ***p< 0.01; **p< 0.05; *p< 0.1.

Results for Customer Churn Sub-Model.

In Table 4.5, we compare the estimation results from our proposed churn sub-model with shared random effects to a logistic regression model without shared random ef-

fects. Results show that, in the former, the association parameters $\alpha_{\text{building}} = -0.531$ and $\alpha_{\text{equipment}} = -0.439$ are statistically significant at 1% and 5%, respectively. This demonstrates the importance of using shared random effects to model claims and churn jointly. The association parameters also suggest that the unobserved heterogeneities in claims and churn risks are negatively associated, indicating a customer with a strong latent tendency to make building or equipment claims is less likely to churn/more likely to renew their contract. This finding is also consistent with our exploratory data analysis (Table 4.2), where we found that the average claims of churners are smaller than those of renewers. While there may be multiple reasons that could potentially explain this negative association, one of the most plausible ones is adverse selection (Pauly 1978) – customers who want to purchase insurance are likely to be those with relatively high claims risks, reflecting an unobserved tendency to renew insurance and make claims. This finding is consistent with the operating feature of the LGPIF, i.e., LGPIF was an insurer of last resort and had little competition pressure. High-claim-risk policyholders of the LGPIF might be charged higher premiums by other insurers in the market, so these policyholders "adversely selected" the LGPIF. Though our empirical conclusion might not be generalised to the broader market, our proposed model could be used by insurers to explore the unobserved association between claims and churn risks for their own data.

The coefficient of *logRateBuilding* ($\hat{\delta} = 1.046$, $p < 0.01$) is positive and significant, indicating that customers paying higher premiums for building coverage are associated with a greater likelihood of churn, in line with higher prices being associated with lower demand. The coefficient for equipment category *logRateEquipment* ($\hat{\delta} = 0.136$, $p > 0.1$) is not statistically significant. One possible reason is that building insurance premiums generally account for a large proportion of the total premium paid by customers, while the equipment insurance premium only accounts for a small proportion. Therefore, a customer's decision to churn is mainly driven by the change in building insurance premiums.

Finally, the positive coefficient of *Tenure* ($\hat{\delta} = 0.111$, $p < 0.05$) indicates that the longer the customer tenure, the higher the likelihood of churn (in line with the results of Guillen et al. 2012). This finding implies that extra attention to loyalty management is required for long-term customers. Moreover, in both claims and churn sub-models, we find that entity type plays a statistically significant role in measuring and differentiating churn and claims risks, indicating different types of local government entities need to be managed differently from the perspective of claims and churn risks.

Table 4.5: Parameter estimates for the churn sub-model in our proposed shared random effects model.

	With shared random effects		**Without shared random effects**	
	Estimate	S.E.	Estimate	S.E.
Intercept	-4.317	0.840**	-3.825	0.750***
TypeMisc	1.021	0.422**	0.894	0.384**
TypeSchool	1.311	0.317***	1.218	0.294***
TypeTown	1.075	0.344***	0.971	0.317***
TypeVillage	0.861	0.318**	0.801	0.299***
Tenure	0.111	0.053**	0.037	0.043
NoClaimCredit	-0.107	0.195	0.172	0.175
logDeductBuilding	-0.086	0.107	-0.073	0.096
logDeductEquipment	0.035	0.037	0.026	0.033
logRateBuilding	1.046	0.277***	0.886	0.223***
logRateEquipment	0.136	0.327	0.036	0.308
α_{building}	-0.531	0.199***		
$\alpha_{\text{equipment}}$	-0.439	0.219**		

Signif. codes: ***p< 0.01; **p< 0.05; *p< 0.1.

4.4.3 Understanding the Role of Shared Random Effects

The association parameters in Table 4.5 show that the shared random effects play a vital role in capturing the unobserved heterogeneity in claims and customer churn risks simultaneously. Below, we illustrate how shared random effects quantitatively translate the unobserved heterogeneity in claims and customer churn risks and their negative latent association, into tangible claims and churn risks.

In Table 4.6, we vary the value of the shared random effects (from -2 to +2), while holding all the covariates at their mean values, to the conditional effect of the shared random effects on the claims and customer churn risks (Hoffmann and Kassouf 2005; de Miguel et al. 2012). Results of this show that the size of a shared random effect is positively associated with the corresponding claims risk, but negatively associated with the churn risk (consistent with the association parameters being negative in Table 4.5). Practically, this demonstrates that customers present substantial differences in claim and churn risks even if they have the same observed characteristics (due to differences in shared random effects). For example, holding all other characteristics fixed, a customer whose building shared random effect is -2 has an expected building claims cost of $350.5 and churn percentage of 4.96%, while a customer whose building shared random effect is +2 has a much larger expected building claims cost of $19,135.1 but a much smaller churn percentage of 0.62%. We refer to Appendix C.1 for the histogram of estimated random effects of all customers.

Table 4.6: The role of shared random effects in influencing claims and churn risks.

Shared random effect	**-2**	**-1**	**0**	**1**	**2**
Varying the shared random effect of building category from -2 to 2.					
At least one building claim probability (%)	12.74	18.57	26.63	37.30	50.53
Expected building claims cost ($)	350.5	952.7	2,589.7	7,039.4	19,135.1
Churn probability (%)	4.96	2.98	1.78	1.05	0.62
Varying the shared random effect of equipment category from -2 to 2.					
At least one equipment claim probability (%)	2.30	4.02	6.95	11.90	19.98
Expected equipment claims cost ($)	16.4	44.6	121.2	329.4	895.3
Churn probability (%)	4.17	2.73	1.78	1.15	0.75

All covariates are fixed at their mean values.

Since we use the Tweedie regression to model claims, we can also calculate the probability of making at least one claim for each customer. For example, holding all other characteristics constant, when we vary the building shared random effect from -2 to +2, the probability of making at least one building claim varies from 12.74% to 50.53% for this synthetic customer. The difference in claims and churn risks across customers due to the difference in their shared random effects, in turn, sheds light on the importance of using shared random effects models to capture the unobserved heterogeneity in claims and customer churn risks.

From a managerial perspective, understanding and capturing the interplay between claims and churn risks through shared random effects is invaluable for insurers. Customers with high shared random effects (noting customer-specific estimates of the random effects are obtained as part of the model fitting in Section 4.3.2) should be closely monitored, and tailored customer management and pricing strategies can be developed for them, as they are relatively more likely to incur long-term costs (high claims risk and low churn risk). This insight aligns with credibility theory in actuarial science (see Frees et al. 1999; Boucher and Denuit 2006; Cohen and Einav 2007, among others), where a customer with higher past claims is associated with a higher random effect in claims risk, meaning they would be predicted to incur higher costs in the future.

4.5 Predictive Performance on Customer Lifetime Value

Customer lifetime value (CLV) is an important metric for insurers to evaluate the sustainability of the insurance pool. To be more specific, insurers know that if all

the "good risks" (high-CLV customers) leave, they will be left with "poor risks" (low-CLV customers), exacerbating the unsustainability issue of the insurance pool. To maintain or more closely reach target levels of return, the insurer is likely to respond by increasing premiums. This, in turn, means that more "good risks" leave, leading to the classic downward spiral of further premium rises (Butler 2002; Siegelman 2003; Suncorp 2013). Therefore, insurers could use CLV to evaluate the sustainability of the insurance pool and manage the dreaded downward spiral.

To evaluate the predictive accuracy of our proposed shared random effects model in estimating CLV, we compared it against a range of other models, as detailed in Table 4.7, based on a two-year hold-out dataset formed from the LGPIF. In detail, the training set spanning 2006 to 2010 was used for model fitting, while the years 2011 and 2012 served as the hold-out or test set for out-of-sample assessment. We designated the beginning of 2011 as the reference "present time" for CLV calculations, such that all data after this time was treated as unobserved and required prediction. That is, we multiply the latest premiums (paid at the beginning of 2011) by the inflation rate to predict the premiums charged at the beginning of 2012, while the inflation rate (also used as the discount rate) is the historical average inflation rate of 2.2% from 2006 to 2010 in the United States. We made this simple assumption on premiums for three main reasons. Firstly, the LGPIF's pricing process is confidential, making it almost impossible to find a model that could perfectly restore the LGPIF's pricing manual. Secondly, insurers generally "cap" premiums to moderate the premium changes from one year to the next (Werner and Modlin 2010; Parodi 2023). This is done to smooth out fluctuations and reduce the chance of bill shock to customers. Adjusting the current premiums by the inflation rate to predict future premiums can be regarded as a feasible capping method. Thirdly, assuming future premiums remain constant is a default method in insurance customer profitability studies (Fang et al. 2016; Abreu 2019; St-Jean 2021; Welikadaarachchi et al. 2023), while we improve this default method by taking the inflation rate into consideration. In summary, this chapter mainly focuses on showing the specific benefits insurers could get by employing our proposed model for their claims and churn risks, given that their premium calculation process is only affected by the inflation rate. Once more and more insurers understand and accept our proposed model, they could use it to improve their technical premiums (expected claims) and demand modelling (customer churn), which might affect total premiums (charged premiums). Consequently, we expect that insurers would be able to keep positive-CLV customers and force negative-CLV customers to churn using our proposed model in practice.

We assume other undetermined covariates, such as coverage, deductible, and no-

claim credit, remain unchanged from their latest values observed at the beginning of 2011. This assumption is made to simplify the prediction process so that we do not have to build models for each undetermined covariate. For example, one needs to use the predicted claims for two consecutive years to predict the expected value of no-claim credit, which further complicates the CLV calculation. However, we acknowledge that building suitable models for undetermined covariates has the potential to improve the overall predictive performance of our proposed model (Shi and Lee 2022), especially when the length of the prediction period is long. Last, new customers who entered their contracts at the beginning of 2012 were treated as completely unobserved, and were excluded from this predictive comparison.

4.5.1 Alternative Models for Customer Lifetime Value

We evaluate the prediction accuracy of our proposed model against several alternative models, as detailed in models A1 to A7 in Table 4.7.

Model A1 operates on the cost assumption that all customers have the same constant claims cost, which can be zero (indicating no cost). This approach is prevalent in the current CLV literature. For example, CLV studies listed in the upper panel of Table 4.1 and the non-contractual setting studied by Venkatesan and Kumar (2004) apply the constant cost assumption in their CLV calculation. The constant claim cost used in Model A1 is estimated by the average claim cost per customer in the training data.

Models A2 and A3 are similar to the marketing contact cost models proposed by Kumar et al. (2008) and Rust et al. (2011), which assume the cost per marketing contact (severity) is constant, while modelling the number of marketing contacts (frequency) at the individual level. Note however that claim frequency is a count variable with a large percentage of values equal to zero, and this can not be modelled appropriately using the linear regression models proposed by Rust et al. (2011) and Kumar et al. (2008). To account for this mass at zero characteristic of claim frequency in models A2 and A3, we use zero-inflated Poisson regression to model claim frequency and treat severity as constant. Briefly, zero-inflated Poisson regression is widely used to model discrete responses with a spike at zero, such as claims frequency in actuarial studies (Yip and Yau 2005; Boucher et al. 2009; Chen et al. 2019) among other fields (e.g., Lambert 1992; Hall 2000; Lee et al. 2006; Datta et al. 2015). The only difference between models A2 and A3 is that the latter includes random intercepts in the Poisson component, to capture potential unobserved heterogeneity in frequency risk.

Models A4 and A5 are commonly adopted in the insurance industry, using zero-

inflated Poisson regression to model claims frequency (A4 without random intercepts, A5 with them) along with gamma regression to model claims severity. Together, this is commonly known as the "frequency-severity approach" (see Danzon 1984; Shi et al. 2015; Frees et al. 2016; Lee and Shi 2019, among others), and is known for its flexibility in applying different distributions for frequency and severity sub-models, and capturing the dependence between frequency and severity models.

Table 4.7: Summary of alternative customer lifetime value models.

Models	**Claims sub-model** $\mu_{i,t,k}$	**Customer churn sub-model** $P(l_{i,t}=1\|l_{i,t-1}=0)=$
Alternative models		
A1: Constant claims	$\mu_{i,t,k}=C_k$.	ilogit $\left(\boldsymbol{x}_{i,t,\text{churn}}^\top\boldsymbol{\delta}\right)$
A2: Zero-inflated Poisson & Constant severity	$\mu_{i,t,k}=\mathbb{E}[\text{Frequency}_{i,t,k}]\times\text{Severity}_k$, where $\text{Frequency}_{i,t,k}\sim\text{ZIP}(\pi_{i,t,k},\lambda_{i,t,k}),\lambda_{i,t,k}=\exp\left(\boldsymbol{x}_{i,t,k}^\top\boldsymbol{\beta}_k^P\right)$, and $\pi_{i,t,k}=\text{ilogit}(\boldsymbol{x}_{i,t,k}^\top\boldsymbol{\beta}_k^{ZI})$; $\text{Severity}_k=S_k$.	ilogit $\left(\boldsymbol{x}_{i,t,\text{churn}}^\top\boldsymbol{\delta}\right)$
A3: Random-intercept zero-inflated Poisson & Constant severity	$\mu_{i,t,k}=\mathbb{E}[\text{Frequency}_{i,t,k}]\times\text{Severity}_k$, where $\text{Frequency}_{i,t,k}\sim\text{ZIP}(\pi_{i,t,k},\lambda_{i,t,k}),\lambda_{i,t,k}=\exp\left(b_{i,k}+\boldsymbol{x}_{i,t,k}^\top\boldsymbol{\beta}_k^P\right)$, and $\pi_{i,t,k}=\text{ilogit}(\boldsymbol{x}_{i,t,k}^\top\boldsymbol{\beta}_k^{ZI})$; $\text{Severity}_k=S_k$.	ilogit $\left(\boldsymbol{x}_{i,t,\text{churn}}^\top\boldsymbol{\delta}\right)$
A4: Zero-inflated Poisson & Gamma severity	$\mu_{i,t,k}=\mathbb{E}[\text{Frequency}_{i,t,k}]\times\mathbb{E}[\text{Severity}_{i,t,k}]$, where $\text{Frequency}_{i,t,k}\sim\text{ZIP}(\pi_{i,t,k},\lambda_{i,t,k}),\lambda_{i,t,k}=\exp\left(\boldsymbol{x}_{i,t,k}^\top\boldsymbol{\beta}_k^P\right)$, and $\pi_{i,t,k}=\text{ilogit}(\boldsymbol{x}_{i,t,k}^\top\boldsymbol{\beta}_k^{ZI})$; $\text{Severity}_{i,t,k}\sim\text{Gamma}(\mu_{i,t,k})$, $\mu_{i,t,k}=\exp\left(\boldsymbol{x}_{i,t,k}^\top\boldsymbol{\beta}_k^G\right)$.	ilogit $\left(\boldsymbol{x}_{i,t,\text{churn}}^\top\boldsymbol{\delta}\right)$
A5: Random-intercept zero-inflated Poisson & Gamma severity	$\mu_{i,t,k}=\mathbb{E}[\text{Frequency}_{i,t,k}]\times\mathbb{E}[\text{Severity}_{i,t,k}]$, where $\text{Frequency}_{i,t,k}\sim\text{ZIP}(\pi_{i,t,k},\lambda_{i,t,k}),\lambda_{i,t,k}=\exp\left(b_{i,k}+\boldsymbol{x}_{i,t,k}^\top\boldsymbol{\beta}_k^P\right)$, and $\pi_{i,t,k}=\text{ilogit}(\boldsymbol{x}_{i,t,k}^\top\boldsymbol{\beta}_k^{ZI})$; $\text{Severity}_{i,t,k}\sim\text{Gamma}(\mu_{i,t,k})$ and $\mu_{i,t,k}=\exp\left(\boldsymbol{x}_{i,t,k}^\top\boldsymbol{\beta}_k^G\right)$.	ilogit $\left(\boldsymbol{x}_{i,t,\text{churn}}^\top\boldsymbol{\delta}\right)$
A6: Tweedie	$\mu_{i,t,k}=\mathbb{E}[\text{Claims}_{i,t,k}]$, where $\text{Claims}_{i,t,k}\sim\text{Tweedie}(\mu_{i,t,k})$, and $\mu_{i,t,k}=\exp\left(\boldsymbol{x}_{\boldsymbol{i,t,k}}^\top\boldsymbol{\beta_k}\right)$.	ilogit $\left(\boldsymbol{x}_{i,t,\text{churn}}^\top\boldsymbol{\delta}\right)$
A7: Type-2 Tobit	$\mu_{i,t,k}=\mathbb{E}[z_{2,i,t,k}]$, where $\log(z_{2,i,t,k})=z_{2,i,t,k}^*$ if $z_{1,i,t,k}^*>0$, otherwise $\log(z_{2,i,t,k})=0$; $z_{1,i,t,k}^*=\boldsymbol{x}_{i,t,k}^\top\boldsymbol{\beta_{1,k}}+\epsilon_{1,i,t,k}$, and $z_{2,i,t,k}^*=\boldsymbol{x}_{i,t,k}^\top\boldsymbol{\beta_{2,k}}+\epsilon_{2,i,t,k}$.	ilogit $\left(\boldsymbol{x}_{i,t,\text{churn}}^\top\boldsymbol{\delta}\right)$
Proposed models		
P1: Random-intercept Tweedie (independent)	$\mu_{i,t,k}=\mathbb{E}[\text{Claims}_{i,t,k}]$, where $\text{Claims}_{i,t,k}\sim\text{Tweedie}(\mu_{i,t,k})$, and $\mu_{i,t,k}=\exp\left(b_{i,k}+\boldsymbol{x}_{\boldsymbol{i,t,k}}^\top\boldsymbol{\beta_k}\right)$.	ilogit $\left(\boldsymbol{x}_{i,t,\text{churn}}^\top\boldsymbol{\delta}\right)$
P2: Random-intercept Tweedie (dependent)	$\mu_{i,t,k}=\mathbb{E}[\text{Claims}_{i,t,k}]$, where $\text{Claims}_{i,t,k}\sim\text{Tweedie}(\mu_{i,t,k})$, and $\mu_{i,t,k}=\exp\left(b_{i,k}+\boldsymbol{x}_{\boldsymbol{i,t,k}}^\top\boldsymbol{\beta_k}\right)$.	ilogit $\left(\boldsymbol{x}_{i,t,\text{churn}}^\top\boldsymbol{\delta}+\sum_{k=1}^K\alpha_k b_{i,k}\right)$

$\text{ZIP}(\pi,\lambda)$ refers to zero-inflated Poisson regression with two parameters π and λ.
$\text{ilogit}(x)=\exp(x)/(1+\exp(x))$ denotes the inverse logit transformation.

Model A6 employs Tweedie regression to model total annual individual-level claims without separating frequency and severity, an approach known in the insurance literature as the "aggregate-loss approach" (e.g., Jørgensen and Paes De Souza 1994; Shi et al. 2015; Frees et al. 2016). Compared to the frequency-severity approach, the aggregate-loss approach is more parsimonious because only one regression model needs to be built. Model A6 does not incorporate any random intercepts for capturing the unobserved heterogeneity in claim risk (in contrast to our proposed models P1 and P2, discussed below).

Model A7 is a Type-2 Tobit regression and uses a (normally distributed) latent variable $z^*_{1,i,t,k}$ to capture the yes/no decision for making a claim (yes if $z^*_{1,i,t,k} > 0$, no else), and a (normally distributed) latent variable $z^*_{2,i,t,k}$ representing the claim amount conditional on a claim being made (Humphreys 2013). Type-2 Tobit models are widely used to model zero-inflated continuous variables in economics and marketing (Bolton et al. 2000; Nizar Al-Malkawi 2007; Randall et al. 2007; van Heerde et al. 2008), though to the best of our knowledge, they have yet to be used for modelling insurance claims data directly.

For the customer churn component of models A1 to A7, we use a logistic regression model that excludes a random intercept (see the right column of Table 4.7). This decision not to include any non-shared (i.e., separate) random intercepts in this component is driven by additional exploration, where we found that churn models that included non-shared random intercepts suffered from persistent model convergence problems and poor prediction performance of customer churn (this is likely a product of the binary response being complete or quasi-separation; see Clark et al. 2023, for instance).

Finally, our proposed shared random effects model is model P2. We also have a simplified version of our approach in the form of model P1, which assumes claims and churn sub-models are independent and the latter uses a logistic regression analogous to those of models A1 to A7. As such, all alternative models along with model P1 are constructed assuming that claims and churn sub-models function independently; note this also means the random intercepts $b_{i,k}$ in the claims sub-models are independent of each other for different insurance categories k. We refer to Appendix C.2 for the estimation results of all alternative models. We conclude this section by remarking that although there are other potentially suitable alternatives for claims and churn sub-models, we focus on the seven alternative models outlined in Table 4.7 to make the illustration relatively comprehensive, yet insightful. We leave the study of comparing a wider range of alternative models for future research.

4.5.2 Claims and Churn Predictions

Before presenting results on CLV predictive performance, we first evaluate each model's performance in predicting individual claims costs (ground-up losses) and churn rates.

To evaluate the customer churn prediction, we use the area under the receiver operating characteristic (ROC) curve (or AUC), balanced accuracy, and precision. These metrics are frequently applied in imbalanced classification tasks such as customer churn, and are derived based on the number of true positives (TP), false positives

(FP), true negatives (TN), and false negatives (FN) in a confusion matrix (see Neslin et al. 2006; Burez and Van den Poel 2009; Bekkar et al. 2013; Holtrop et al. 2017, for examples). The ROC curve is a plot of the TP rate against the FP rate at various threshold settings, while for the other two metrics we have

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}; \quad \text{balanced accuracy} = \frac{1}{2}\Big(\frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FP}}\Big).$$

Turning to evaluate the predictive performance of claims, we use the Gini index, the mean absolute error (MAE), and the Pearson correlation coefficient, three metrics frequently employed in claim modelling (e.g., Frees, Gao and Rosenberg 2011; Frees et al. 2016; Yang and Shi 2019). Note that the Pearson correlation coefficient for Model A1 is not available because Model A1 assumes the claims prediction is constant.

The Gini index (Frees, Meyers and Cummings 2011; Frees et al. 2014; Shi and Zhao 2024) in this context evaluates a claims model's ability to rank customers based on their profitability (measured by relativity). The computation of the Gini index involves defining relativity as

$$R_{i,t,k} = \hat{\mu}_{i,t,k}/\text{Premium}_{i,t,k},$$

where $\hat{\mu}_{i,t,k}$ is the predicted k-th claim by a model for the i-th customer at time t, and $\text{Premium}_{i,t,k}$ is the premium charged to the customer by the insurer (premiums are given in the data). The relativity represents the ratio of the predicted claims by the model, over the premiums charged by the insurer, which measures the profitability of a customer. Here, the lower the predicted claims relative to the premiums, the more profitable the customer is expected to be for the insurer. As such, insurance companies are most interested in customers whose relativity is low (ideally lower than 1). Therefore, the effectiveness of using relativity to identify profitable and non-profitable customers mainly depends on the accuracy of the model for claims prediction. Different claims models would lead to different rankings of customers according to the relativity order determined by the model, so that we could examine which model better identifies a profitable portfolio of customers.

After calculating each customer's relativity, we rank their relativity from low to high, i.e., from the most to the least profitable. We then plot an ordered Lorenz curve with the y-coordinates given by

$$\text{Share of Claims } (r) = \frac{\sum_i \sum_t \text{Claims}_{i,t,k} \mathbb{I}(R_{i,t,k} \leq r)}{\sum_i \sum_t \text{Claims}_{i,t,k}},$$

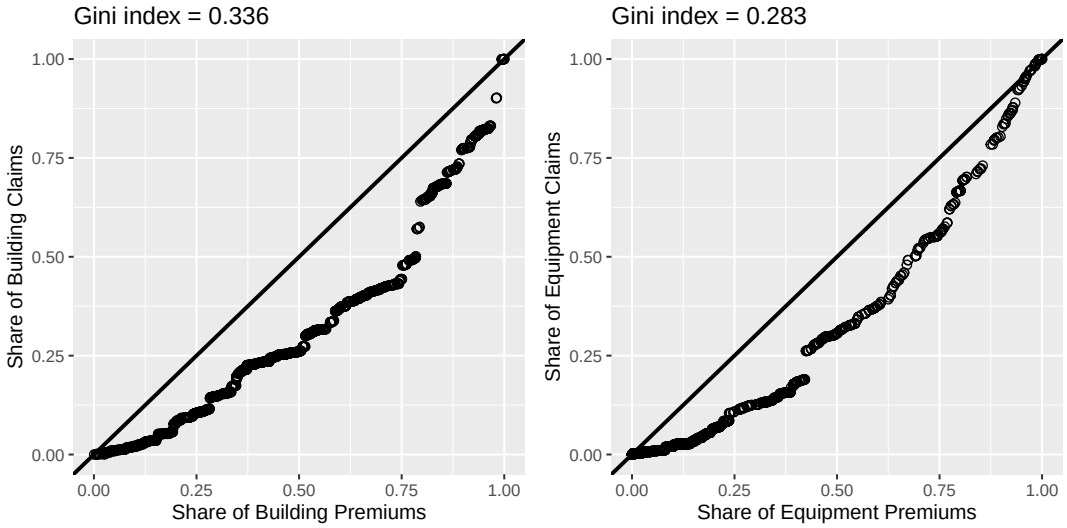

Figure 4.1: Ordered Lorenz curves obtained by model P2 for building (left panel) and equipment (right panel) categories.

and the x-coordinates given by

$$\text{Share of Premiums } (r) = \frac{\sum_i \sum_t \text{Premium}_{i,t,k} \mathbb{I}(R_{i,t,k} \leq r)}{\sum_i \sum_t \text{Premium}_{i,t,k}}.$$

Here, r denotes the ascending thresholds for the ordered relativity.

Figure 4.1 shows the ordered Lorenz curves obtained by our proposed model P2, where the ordered empirical relativities calculated by model P2 are used as the thresholds. Each dot in Figure 4.1 represents a relationship between the share of claims and the share of premiums for a selected portfolio of customers. In other words, each dot represents an underwriting strategy that the insurer uses to select the most profitable customers given a certain relativity upper limit. For example, a dot close to the coordinates $(0, 0)$ represents a portfolio of customers with comparatively low relativities among all customers. As we move in the northeast direction of the plot, customers with comparatively high relativities are added to the corresponding portfolio. From the left panel of Figure 4.1, we see that a 50% share of building premiums corresponds to approximately 25% of building losses, indicating substantial opportunities for risk segmentation for insurers. Indeed, the fact that in both panels the ordered Lorenz curve lies below the 1-1 line of equity indicates that model P2 could accurately identify the order of customers in terms of their profitability.

The Gini index is formally defined to be (twice) the area between the ordered Lorenz curve and the line of equality, and thus can be interpreted as average profit over all underwriting strategies (dots of the ordered Lorenz curves). Thus, a model that

is better at predicting claims would have a higher Gini index, which indicates a higher overall profit. We refer readers to Appendix C.3 for a simple example to help understand the Gini index.

Table 4.8 shows that our proposed model P2 has consistently strong out-of-sample predictive performance relative to all alternative models in predicting both claims and churn. For 7 out of 9 criteria, Model P2 is best, while it is second-best on the other 2 criteria.

Table 4.8: Evaluating the predictive performance of claims and churn in the test set.

Models	**Building claims**			**Equipment claims**			**Customer churn**		
	Gini index ↑	**MAE ($)** ↓	**Pearson** ↑	**Gini index** ↑	**MAE ($)** ↓	**Pearson** ↑	**AUC** ↑	**Balanced accuracy** ↑	**Precision** ↑
Alternative models									
A1: Constant claims	-0.182	33,359.4		-0.112	7,128.1		0.563	0.516	0.037
A2: Zero-inflated Poisson & Constant severity	0.239	30,557.9	0.587	0.186	4,459.4	0.540	0.563	0.516	0.037
A3: Random-intercept zero-inflated Poisson & Constant severity	0.327	29,516.6	0.451	**0.331**	4,604.3	0.516	0.563	0.516	0.037
A4: Zero-inflated Poisson & Gamma severity	0.085	45,364.8	0.605	0.211	5,002.8	0.502	0.563	0.516	0.037
A5: Random-intercept zero-inflated Poisson & Gamma severity	0.260	53,297.4	0.463	0.231	5,479.1	0.457	0.563	0.516	0.037
A6: Tweedie	0.162	28,046.8	0.520	0.183	4,311.4	0.534	0.563	0.516	0.037
A7: Type-2 Tobit	0.115	28,575.7	0.459	0.142	4,667.5	0.492	0.563	0.516	0.037
Proposed models									
P1: Random-intercept Tweedie (independent)	**0.341**	25,719.9	0.621	0.278	3,869.8	0.551	0.563	0.516	0.037
P2: Random-intercept Tweedie (dependent)	0.336	**25,302.3**	**0.625**	0.283	**3,864.9**	**0.552**	**0.595**	**0.569**	**0.047**

↓: the lower the better; ↑: the higher the better. Best values are presented underlined and in bold.

4.5.3 Customer Lifetime Value Predictions

Although it is important to evaluate individual predictions of claims and churn, this alone does not offer a complete insight into a model's effectiveness in estimating CLV. As previously discussed, the projected premiums are the same for all models. Consequently, the predictive performance of CLV depends on how well a model predicts claims and churn rates jointly. To calculate CLV using Equation (4.1), we set the discount rate at 2.2% (reflecting the historical average inflation rate in the U.S. from 2006 to 2010), and note $T = 2$ given the two years in our test data. For simplicity, we assume premiums and claims costs are realised concurrently at each year's end.

Table 4.9 presents various evaluation metrics used for assessing CLV prediction. Five metrics are used to assess how well the models predict CLV among all customers in the test set i.e., overall performance. For illustration purposes, let CLV_i and $\hat{\text{CLV}}_i$

denote the true CLV and predicted CLV for the i-th ($i = 1, \ldots, N_{\text{test}}$) customer in the test set, respectively. Note the true CLV is calculated using the actual payments of claims (after applying deductibles) as customers' costs. When calculating the predicted CLV, we utilise the loss elimination ratio approach, a standard deductible rate-making approach in actuarial practice (Lee 2017), to further adjust the predicted net profits to account for deductibles. We refer readers to Appendix C.4 for the illustration of this approach.

The first metric measures the proportion of customers with positive predicted CLV in the test set,

$$(\text{CLV} > 0)\% = \frac{\sum_{i=1}^{N_{\text{test}}} \mathbb{I}(\hat{\text{CLV}}_i > 0) \times 100\%}{N_{\text{test}}}.$$

The error is then calculated by [(CLV > 0)% $-$ True rate], where the true rate (86.9%) is the proportion of customers with positive true CLV in the test data. For instance, our proposed model P2 predicts (CLV > 0)% $= 89.3\%$, and so its error is $89.3\% - 86.9\% = +2.4\%$. As Table 4.9 shows, model P1 has a predicted (CLV > 0)% closest to the true rate (just 1.9% off), followed closely by model P2 (2.4% off). All alternative models have predicted (CLV > 0)% lower than the true ratio, indicating that they all tend to overestimate the claims risk and underestimate CLV in the test data.

The second metric we consider is rank-based hit rate proposed by Malthouse and Blattberg (2005). To calculate this, we divide customers into quartiles (G_1, G_2, G_3, G_4) according to their rank of CLV. For example, if a customer's CLV is in the top (bottom) 25% among all customers, this customer will be divided into group G_1 (G_4). We then calculate

$$\text{Hit rate} = \frac{\sum_{g=1}^{4} \sum_{i=1}^{N_{\text{test}}} \mathbb{I}(\hat{\text{CLV}}_i \in G_g \ \&\ \text{CLV}_i \in G_g) \times}{N_{\text{test}}} \times 100.$$

From Table 4.9, we observe that our proposed models P1 and P2 outperform all the alternative models in terms of hit rate. Model A1 also achieves relatively high hit rates – this is because the majority of customers did not make any claims in the test, and so premiums mainly determine their CLVs. In this case, model A1 applying the constant claim assumption can reliably predict the rank for the customers who did not make any claim, leading to a relatively high hit rate.

The remaining three metrics are dollar-based metrics. These are the mean absolute error (MAE), the two-sample Kolmogorov-Smirnov (K-S) statistic, and the Pearson correlation coefficient. The two-sample Kolmogorov-Smirnov (K-S) statistic quantifies how close the distributions of predicted and true CLV are (Pettitt

Table 4.9: Evaluating overall predictive performance of customer lifetime value in the test set.

Models	**(CLV>0)%: error** True rate=86.9%; ↓	**Hit rate** ↑	**MAE ($)** ↓	**K-S statistic** ↓	**Pearson correlation** ↑
Alternative models					
A1: Constant claims %	-52.7%	58.7%	65,924.2	0.531	-0.292
A2: Zero-inflated Poisson & Constant severity	-38.7%	28.1%	55,409.9	0.496	0.559
A3: Random-intercept zero-inflated Poisson & Constant severity	-23.6%	45.2%	55,719.1	0.313	0.349
A4: Random-intercept Poisson & Gamma severity	-23.3%	24.5%	87,602.3	0.401	0.587
A5: Random-intercept zero-inflated Poisson & Gamma severity	-12.0%	41.6%	103,936.4	0.250	0.434
A6: Tweedie	-11.1%	47.2%	50,667.1	0.265	0.262
A7: Type-2 Tobit	-19.3%	36.2%	52,799.4	0.379	-0.062
Proposed models					
P1: Random-intercept Tweedie (independent)	**+1.9%**	62.6%	46,025.2	0.094	0.628
P2: Random-intercept Tweedie (dependent)	+2.4%	**63.0%**	**45,925.3**	**0.092**	**0.629**

↓: the lower the better; ↑: the higher the better. Best values are presented underlined and in bold.

and Stephens 1977). That is, suppose $F(v)$ and $W(v)$ are the empirical cumulative distribution functions of true CLV and predicted CLV, respectively. Then K-S statistic $= \max_v |F(v) - W(v)|$, and a lower K-S statistic indicates greater similarity between the distributions of true and predicted CLV. The Pearson correlation coefficient (see Liu and Shih 2005*a*,*b*; Shih and Liu 2008, among others for its usage) evaluates whether the predicted CLV aligns well with the true CLV in a linear sense. For all three dollar-based metrics, model P2 is the winner, followed (closely) by model P1.

Figure 4.2 visualises the true and predicted CLV distributions for our proposed model P2, where the range of true CLV is roughly from -$4,900,000 to $500,000 at the individual customer level (remember that customers are oftentimes large government entities). Our proposed model P2 has the ability to predict the overall shape and some extreme values (such as -$4,500,000) relatively well, although even more extreme true CLV (such as -$4,900,000) caused by very large claims are naturally hard to predict accurately by any model in practice. We also observe that the probability mass below zero in true CLV is larger than in the predicted CLV, which is consistent with the error of **CLV** > 0 (1.9%) in Table 4.9. We refer the reader to Appendix C.5 for the distributions of predicted CLV of other alternative

models.

To summarise, our proposed models P1 and P2 have consistently strong predictive performance in terms of overall performance for CLV prediction. The alternative models exhibit varying shortcomings, potentially leading to sub-optimal customer management in real-world scenarios, which we will explore next.

4.5.4 Profit Analysis based on Customer Lifetime Value Predictions

While our proposed model performs well in the overall prediction of CLV, this does not directly translate to the models' economic significance for insurers. Instead, we can consider an insurer aiming to implement a proactive marketing strategy by targeting customers with a negative predicted CLV for churn. If a model is able to predict CLV accurately at the individual level, then the targeted customers with negative predicted CLV would be likely to have negative true CLV. In this case, proactive marketing management could successfully save money (generate profit) by terminating the contract with customers having a predicted negative CLV, i.e., applying a demarketing strategy for these customers (Lepthien et al. 2017). Note that our demarketing strategy is hypothetical, which ignores the negative word-of-mouth consequences of "firing" customers. We expect to use this hypothetical strategy to generate business insights that could be used as a reference by insurers. In practice, insurers generally have the right to exit from a certain market, which is related to the total CLV of that market. For example, Farmers Insurance ended coverage in Florida (McDaniel 2023), a move that would affect about 100,000 existing policies and represented the latest retreat by insurance companies reluctant to cover Americans living in disaster-prone areas. This example suggests that providing farmers insurance in Florida might not be sustainable, i.e., the total CLV of farmers insurance might be negative in Florida. This type of dilemma for insurers is mainly because the risk of hazards is estimated to be very high by insurers, leading to unaffordable prices for their customers.

With this in mind, for each of the models considered, we implement a hypothetical demarketing strategy that terminates the contract with those customers having a negative predicted CLV. We then aggregate the actual CLVs of these hypothetically terminated customers to understand how much additional profit this strategy can bring. To quantify this, we construct two new metrics called profit increase per customer ($), and profit increase per million coverage ($), to evaluate the potential

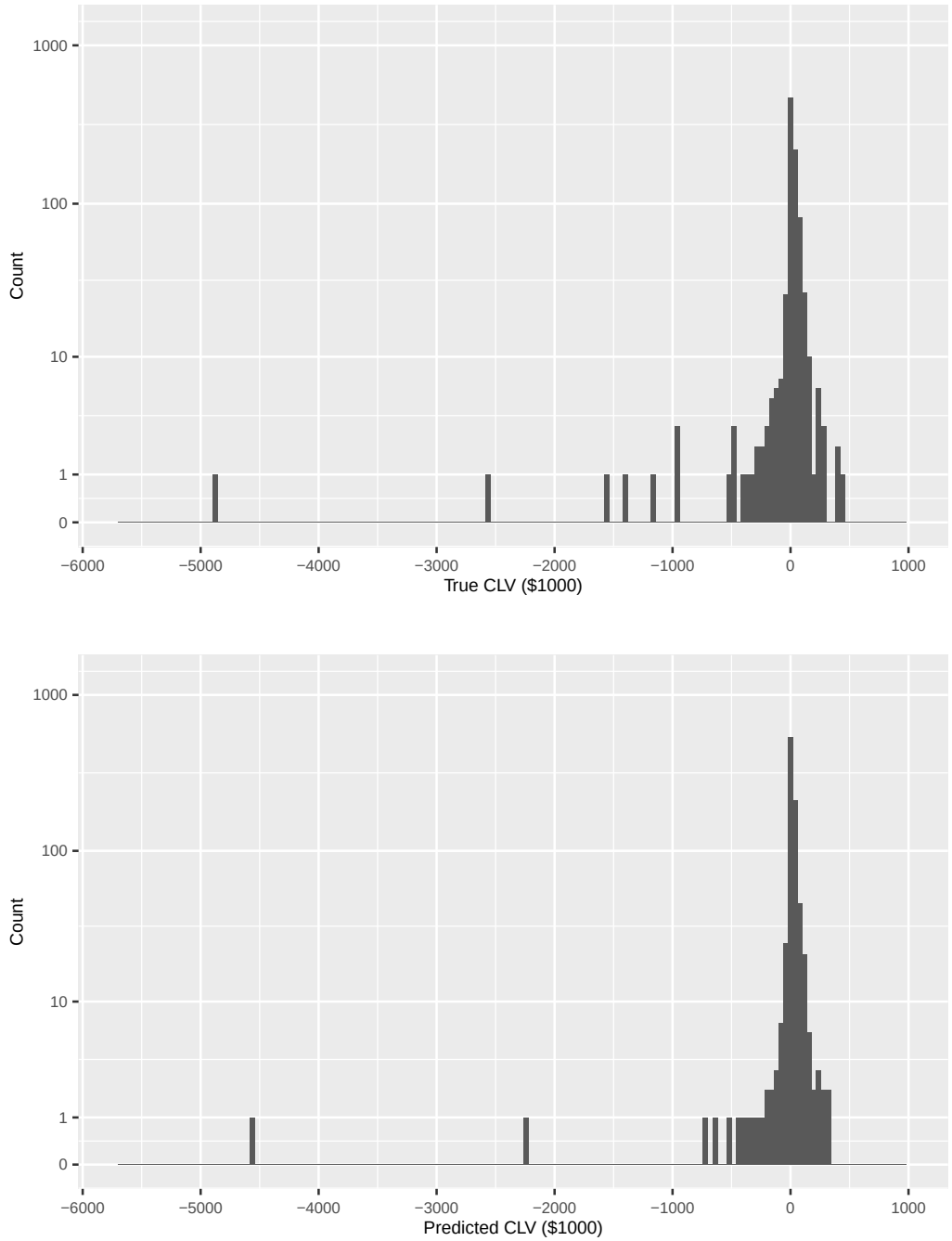

Figure 4.2: The distributions of true CLV (top panel) and predicted CLV of model P2 (bottom panel) in the test set (the height of bars is scaled by $\log_{10}$(count+1)).

economic benefits brought by models to insurers,

$$\text{Profit increase per customer} = \frac{\sum_{i=1}^{N_{\text{test}}} \mathbb{I}(\hat{\text{CLV}}_i < 0) \times (-\text{CLV}_i)}{\sum_{i=1}^{N_{\text{test}}} \mathbb{I}(\hat{\text{CLV}}_i < 0)},$$

$$\text{Profit increase per million coverage} = \frac{\sum_{i=1}^{N_{\text{test}}} \mathbb{I}(\hat{\text{CLV}}_i < 0) \times (-\text{CLV}_i)}{\sum_{i=1}^{N_{\text{test}}} \mathbb{I}(\hat{\text{CLV}}_i < 0) \times \text{Coverage Amount}_i},$$

where Coverage Amount is a customer's total coverage in the test set.

Table 4.10 summarises the profitability of CLV models as measured by these two profit-based metrics. A demarketing strategy using our proposed model P2 could generate a mean profit of $58,374.70 per demarketed customer and $248.0 per million coverage ($5,545,599 profit in total). In contrast, using model A1 could lead to a mean loss of $6,294.40 per demarketed customer and $265.3 per million coverage ($3,657,068 loss in total). We also present the results based on Winsorised and trimmed means, both of which are less sensitive to outliers. Again, our proposed model P2 exhibits strong profitability in these cases compared to other models. The superiority of model P2 over model P1 in particular further lends support to the importance of capturing the association between intrinsic claims and churn risks through shared random effects, from the perspective of models' profitability.

Table 4.10: Evaluating the profitability of CLV models through the proposed marketing strategy.

Models	**Profit increase per customer ($)**			**Profit increase per million coverage ($)**		
	Mean ↑	**Winsorised (1%)** ↑	**Trimmed (1%)** ↑	**Mean** ↑	**Winsorised (1%)** ↑	**Trimmed (1%)** ↑
Alternative models						
A1: Constant claims	-6,294.4	-8,059.7	-8,738.1	-265.3	-290.6	-347.4
A2: Zero-inflated Poisson & Constant severity	100.8	-14,099.7	-17,774.5	0.8	-98.1	-146.9
A3: Random-intercept zero-inflated Poisson & Constant severity	20,167.2	2,385.2	-4,004.3	**277.3**	64.2	-27.5
A4: Random-intercept Poisson & Gamma severity	-5,360.8	-18,198.6	-25,013.7	-26.1	-93.9	-124.6
A5: Random-intercept zero-inflated Poisson & Gamma severity	22,534.9	-1,228.4	-7,877.5	147.9	-10.4	-66.1
A6: Tweedie	10,419.8	-7,712.0	-13,354.7	65.5	-46.2	-114.9
A7: Type-2 Tobit	-12,825.7	-12,886.6	-13,898.9	-130.8	-148.7	-176.6
Proposed models						
P1: Random-intercept Tweedie (independent)	55,635.9	22,801.2	11,046.1	226.6	112.8	59.9
P2: Random-intercept Tweedie (dependent)	**58,374.7**	**24,626.8**	**11,926.0**	248.0	**126.2**	**69.6**

↓: the lower the better; ↑: the higher the better. Best values are presented underlined and in bold.

We conclude this section by noting that the LGPIF was closed at the end of 2017. The Governor's proposed State Budget suggested that property insurance might be

more effectively managed by the private insurance sector (Rick 2015; OCI 2019). The mission of the LGPIF was to make property insurance available for local government units, so it was not allowed to deny coverage, although local government units could secure insurance in the open market. Thus, in reality, the LGPIF did not conduct the demarketing strategy mentioned in the paragraph above, while the local government units with huge negative CLV might have contributed to the closure of the LGPIF. In hindsight, one potential way for the LGPIF to escape the fate of being closed is to adjust the premiums in a timely but appropriate manner; we expand on this point in the next section.

4.5.5 Adjusting Premiums to Increase Customer Lifetime Value

The estimated coefficients of *logRateBuilding* ($\hat{\delta} = 1.046$, $p < 0.01$) and *logRateEquipment* ($\hat{\delta} = 0.136$, $p < 0.01$) from the customer churn sub-model in Table 4.5 suggest that higher rates correlate with an increased likelihood of churn. Consequently, if the LGPIF had increased premiums for customers with negative predicted CLV, then these customers might have sought insurance in the open market rather than renewing their contracts with the LGPIF.

In this section, we assume the LGPIF adopts our proposed model P2 to implement a proactive pricing strategy: increasing the premiums per coverage by 10%, 20%, 30%, and 40% for the customers with negative predicted CLV and keeping the premiums unchanged for the rest. We restrict the maximum variation in the premiums per coverage to 40%, which is closely in line with the empirical rate of change in premium per coverage we observed in the LGPIF data; see Figure 4.3. The underlying rationale for this strategy is to encourage negative-CLV customers to churn with elevated rates, while the LGPIF could increase revenues even if those customers accept the elevated rates and renew their contracts. While these pricing strategies have a simple design for illustrative purposes, it is worth noting that premium adjustments in personal insurance lines could be strictly regulated to prevent discrimination (Avraham 2017; Lindholm et al. 2022; Frees and Huang 2023). However, commercial insurance lines, like the LGPIF, often have more flexibility in premium adjustments.

Table 4.11 illustrates how increasing the premiums for the negative-CLV customers would influence their predicted premiums paid, claims, churn rates, and CLV. As we increase the premiums for the negative-CLV customers from 10% to 40%, their predicted total premiums increase, expected total claims decrease, and the average churn rate among those customers increases. Consequently, these changes jointly

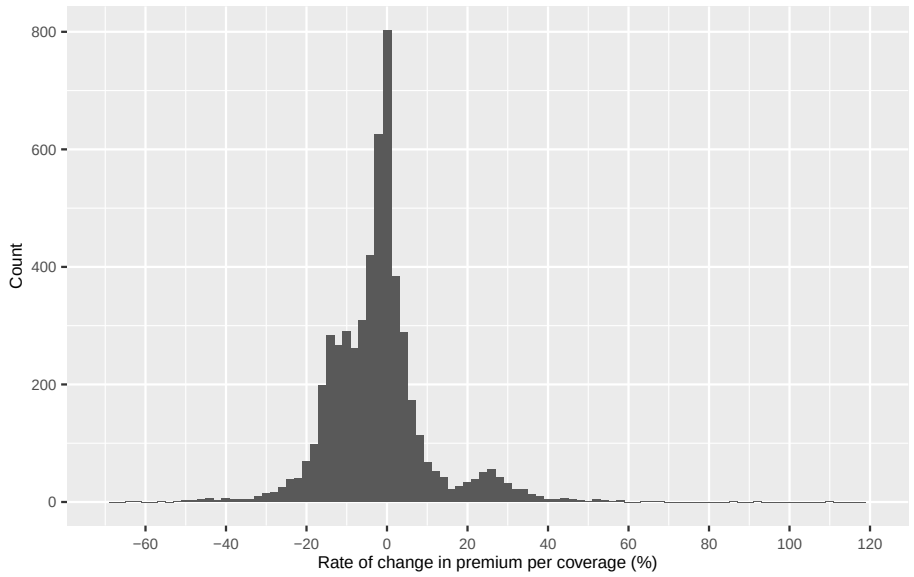

Figure 4.3: The distribution of the rate of change in premium per coverage used by the LGPIF.

lead to an increase in the total predicted CLV of those negative-CLV customers, which is in line with our expectations of the proposed pricing strategy. For example, if we increase premiums by 40% for all the customers with negative predicted CLV, these customers' total predicted CLV would increase by $3.59 million. The analysis in this section thus highlights the importance of strategic pricing in effective cost management. Leveraging our shared random effects model, we show that premium adjustments could enhance an insurance company's total CLV and ensure long-term financial success.

Table 4.11: Evaluating the profitability of model P2 through the pricing strategy for negative-CLV customers.

	Total premiums	**Total claims**	**Average churn rate**	**Total predicted CLV**	**Change in CLV**
Before premium adjustment	$5.21M	$22.38M	0.48%	-$13.50M	
+10% premium adjustment	$5.74M	$22.26M	0.55%	-$12.57M	+$0.93M
+20% premium adjustment	$6.26M	$22.16M	0.61%	-$11.66M	+$1.84M
+30% premium adjustment	$6.78M	$22.09M	0.68%	-$10.77M	+$2.73M
+40% premium adjustment	$7.30M	$22.04M	0.75%	-$9.91M	+$3.59M

Change in CLV represents the change in predicted CLV compared to "Before premium adjustment".

4.6 Discussion

4.6.1 Summary and Takeaways

In this chapter, we have introduced a cost-based customer lifetime value (CLV) modelling framework tailored to the insurance industry, marking a significant advancement in CLV research. By leveraging shared random effects, our model appropriately captures the unobserved heterogeneity in claims and churn risks, and in doing so reveals the latent association between claims and churn tendencies across customers. To the best of our knowledge, this is the first chapter (paper) to employ such a shared random effects model in CLV research.

We applied our shared random effects models to study claims and customer churn behaviours using data from the Wisconsin Local Government Property Insurance Fund (LGPIF). Specifically, we utilise multivariate Tweedie regression models to address the unique challenges posed by claims data, notably its mass at zero and heavy-tailed nature. A logistic regression model serves as the customer churn submodel. Our empirical analysis reveals a negative latent association between claims and churn tendencies across customers of the LGPIF. Furthermore, through a conditional effect analysis, we interpret how shared random effects translate the unobserved heterogeneity and latent association into tangible claims and churn risks, with results showing that customers with differing shared random effects exhibit significant variations in claims and churn risks, even when their observed characteristics are identical.

Given the task of valuing customers in the insurance industry, the proposed model outperforms a diverse array of alternative models in predicting claims, churn, and CLV. Looking beyond the value of these predictions, we develop two evaluation metrics, profit increase per customer and profit increase per million coverage, allowing us to generate managerial insights into model profitability. The strong performance of our proposed model in these two metrics emphasises its potential to bring considerable economic benefits to insurers. Moreover, utilising our proposed shared random effects model, we illustrate that adjusting premiums reasonably could help insurance companies increase total CLV and ensure long-term profitability. In doing so, this chapter presents an instructive avenue for insurance as well as other industries aiming to refine their cost modelling techniques.

4.6.2 Limitations and Future Research

As a new joint modelling framework for CLV research, there are several variations of the shared random effects model that are worthy of further investigation. For instance, beyond shared random intercepts, shared random slopes for the covariates can also be investigated. Doing so, however, not only requires more observations per customer but will also introduce additional covariance structures between random intercepts and slopes that further complicate model fitting. Also, we assume the shared random effects follow a multivariate normal distribution, and future research could examine the sensitivity of this assumption (Hui et al. 2021) as well as the use of non-normal multivariate distributions for the shared random effects, such as a log-gamma distribution (Fabio et al. 2012). Using a suitable multivariate distribution for shared random effects has the potential to improve prediction performance, although again at the expense of more complex and burdensome model estimation.

The coefficients in the claims model and churn models have an associative rather than causal interpretation, in line with the predictive nature of CLV modelling. When the model goal is predictive, there is often no benefit from endogeneity correction; in fact, using 2SLS estimation can make predictions worse (Ebbes et al. 2011). With this in mind, future research could examine the use of machine learning models for both claims and customer churn under the joint modelling framework. However, machine learning models lack interpretability and do not allow us to understand to what extent the improvement in fit is due to allowing for dependencies in the heterogeneities in churn and claim, like in our proposed model. In addition, though we mentioned a proactive marketing strategy in Section 4.5.4 and a pricing strategy in Section 4.5.5 for illustration purposes, we have not discussed the effectiveness of marketing campaigns, interventions, and incentives since we do not have such information in our data. Future research could bridge this gap by linking our proposed models to uplift modelling research (Ascarza 2018; Ascarza et al. 2018; Lemmens and Gupta 2020). Such an extension however requires a suitable dataset tracking customer behaviour before and after a marketing campaign.

One of the limitations of the Tweedie regression approach used in this chapter is that it incorporates covariates to model the mean of claims only. However, there exist practical situations where the variance, as well as the mean of the claims, needs to be modelled. To achieve such flexibility, a standard way is using a double-GLM approach, which models both mean and variance of claims (see Boucher and Davidov 2011; Delong et al. 2021, for examples). It is worth acknowledging that a double-GLM approach also doubles the number of parameters (similar to the frequency-severity approach) compared to the Tweedie regression approach. Consequently,

building a double-GLM generally requires more data in order to estimate it compared to modelling the overall mean using Tweedie regression. Future research could employ a double-GLM approach on Tweedie regression within our joint modelling framework.

Finally, we only study one data set (LGPIF) in this chapter, so our empirical results could be data-specific but not general. If an insurer could price contracts perfectly based on customers' claims risk, then the insurer should be able to retain positive-value customers with attractive prices and force negative-value customers to churn with excessive prices. In this ideal case, the association parameters (alpha in Table 4.5) should be positive instead of negative; i.e., customers with high claims risk are forced to churn via the perfect pricing process. However, the LGPIF studied in this chapter was a special commercial insurance that served government units without profit purposes, so their pricing process was somewhat inefficient. Therefore, future research could test our proposed model on different data sets and explore different association structures. We believe that such exploration could help insurers improve their pricing engine.

Chapter 5

Conclusion

To conclude this thesis, we first offer a summary of the primary findings from Chapters 2 to 4, before providing a discussion of the limitations and possible extensions to this research.

5.1 Summary of Findings

This thesis explores three primary research areas within general insurance: claims modelling, customer churn analysis, and customer valuation. The findings presented in Chapters 2 to 4 provide critical insights for insurers, especially for the actuaries in the general insurance sector.

Chapter 2 studies the cost of combined deductibles under dependence from three perspectives: empirical applications on real data sets, a comprehensive simulation study, and theoretical propositions. The results of the empirical applications, spanning commercial insurance, reinsurance, and personal (pet) insurance, shed light on the importance of dependence modelling for calculating deductible costs. Specifically, the deductible costs are reduced by 6 to 23 per cent for commercial insurance, up to 8 per cent for reinsurance, and 10 to 23 per cent for pet insurance when dependence modelling with copulas is used, compared with benchmark models in the literature that assume independence. We examine a range of dependence parameters, marginal distributions, and copulas in the simulation study, revealing that the strength of dependence emerges as the most influential factor affecting deductible costs. Moreover, the dependence structure and marginal distributions should be selected carefully in practice because they could yield different deductible costs. The theoretical propositions in Chapter 2 present general properties of joint distributions that can lead to materially different prices, thereby explaining the findings in the empirical applications and simulation study. These propositions show the difference in deductible costs between dependence and independence modelling is affected by the size of combined deductibles, and the correlation between risks. With these general

propositions, researchers can effectively address challenges in dependence modelling across diverse contexts. An overarching contribution of the above three perspectives underscores the importance of modelling dependence on combined deductible costs for bundled insurance contracts.

Chapter 3 explores multi-state customer churn analysis in the insurance industry. The primary contribution of this chapter lies in introducing multi-state customer churn analysis, designed to explore customers' behaviours beyond the simple churn/not churn and across multiple periods. We develop multinomial logistic regression (MLR) models with a second-order Markov assumption to conduct multi-state customer churn analysis. Through the use of such models, which prioritise interpretability over predictive power, insurance companies can gain a complete picture of how a policyholder's past states affect their decision to move to the next state, as well as the complex relationships between transition probabilities and various explanatory variables related to premium information, claim information, and contract type. The empirical analysis, based on commercial insurance data from the Wisconsin Local Government Property Insurance Fund (LGPIF), demonstrates that explanatory variables play differing roles in triggering different types of transitions. This nuanced understanding could guide insurers to predict policyholders' future transitions based on their previous states and attributes. To aid in interpretation, we present conditional and average marginal effect plots to interpret the estimated effects on the probability scale. Furthermore, we use the example of an average policyholder to illustrate how multi-state customer churn analysis could improve the accuracy of customer lifetime value calculation for insurers. Although not the primary aim of this chapter as a whole, the second-order MLR model was shown to also predict relatively well compared to a variety of alternative models. Overall, our findings open up a new way of perceiving and understanding customer state transition behaviour in the insurance market, and shed light on the importance of multi-state customer churn modelling.

Chapter 4 introduces a cost-based customer lifetime value (CLV) modelling framework tailored to the insurance industry, marking a significant advancement in CLV research. By leveraging shared random effects, our model appropriately captures the observed heterogeneity in claims and churn risks by covariates, as well as the unobserved and correlated heterogeneity between claims and churn risks by shared random effects. Specifically, we utilise multivariate Tweedie regression models to address the unique challenges posed by claims data, notably its zero-inflated and heavy-tailed nature. A logistic regression model serves as the customer churn sub-model. Our empirical analysis of commercial insurance data reveals a negative latent

association between claims and churn tendencies across customers. Furthermore, through a conditional effect analysis, we interpret how shared random effects translate the unobserved heterogeneity and latent association into tangible claims and churn risks, with results showing that customers with differing shared random effects exhibit substantial variations in claims and churn risks, even when their observed characteristics are identical. Given the task of valuing customers in the insurance industry, the proposed model outperforms a diverse array of alternative models in predicting claims, churn, and CLV. Looking beyond the value of these predictions, we use two hypothetical applications to generate managerial insights into model profitability. The strong performance of our proposed model in these two hypothetical applications emphasises its potential to bring considerable economic benefits to insurers. In doing so, this chapter presents an instructive avenue for insurance as well as other industries aiming to refine their cost modelling techniques.

5.2 Limitations and Future Research

This thesis introduces a variety of innovative approaches to claims modelling, customer churn analysis, and customer valuation. Naturally then, there are several generalisations and extensions that warrant further investigation. First, all the models presented, such as the copula models in Chapter 2, the multinomial logistic regression models in Chapter 3, and the shared random effects models in Chapter 4 are based on association, not causation. For example, the two applications in Sections 4.5.4 and 4.5.5 are based on the association structure drawn from the shared random effects model. Future research could study the claims and customer churn risks from the perspective of causal inference (see Guelman and Guillén 2014; Paredes 2018; Verschuren 2022; Parodi 2023, among others). For example, Paredes (2018) showed that a phone call reminding customers of the benefits of their policy could reduce the auto insurance churn rate by up to 6% from the perspective of causal effects. To our knowledge, there remains a notable gap in understanding the causal effects of insurer interventions on claims and churn risks simultaneously. Addressing this gap could assist insurers in producing more informed decisions, optimising their products and services, improving customer demand modelling, and maximising profits (Ascarza 2018; Ascarza et al. 2018; Lemmens and Gupta 2020). Nevertheless, pursuing this avenue is challenging because it likely requires data collected from carefully planned experimental designs (e.g., randomised control trials), and necessitates close collaboration between academia and industry.

Second, among all the models built in the three chapters, the shared random effects

model proposed in Chapter 4 is the only one that explicitly captures the unobserved and correlated heterogeneity in claims and churn risks across customers. Future research could examine machine-learning-based methods that can enhance the predictive power of claims and churn models through flexibly "learning" the unobserved and correlated heterogeneity in claims and churn risks. One promising avenue involves extending the generalised mixed effects neural network proposed by Avanzi et al. (2023), a hybrid of generalised mixed effects models and neural networks. What calls for special attention though is that modellers need to balance interpretability and prediction accuracy when employing such machine-learning-based models. This is especially important when the models are used for pricing purposes because modellers need to interpret the modelling results to diverse stakeholder audiences.

Third, the multi-state customer churn analysis proposed in Chapter 3 is not incorporated into the joint modelling of claims and churn in Chapter 4. Primarily, this stems from the insufficiency of observations in the LGPIF data to estimate shared random effects for multi-state transitions. With a suitable data set, future research could consider replacing the logistic regression used as the customer churn sub-model in Chapter 4 with a multi-state model, such as multinomial logistic regression. The modified model could help insurers understand how intrinsic multivariate claims risks and multi-state transition tendencies are associated, subsequently leading to a potentially more accurate CLV calculation (considering all paths instead of customer churn only). However, such a model extension also requires a sophisticated algorithm for estimating shared random effects, because the state of a customer will no longer be static.

Fourth, we have not considered geographical variables in this thesis. With insurers and customers having become more concerned about the risk of natural hazards, insurance is increasingly identified as a disaster management technique of choice. A large number of studies has been done to incorporate the natural hazard risk into risk management and insurance pricing (see Hsu et al. 2011; Booth and Harwood 2016; McAneney et al. 2016; Borensztein et al. 2017; Kousky 2019, for examples). Future research could expand the models proposed in this thesis to include geographical variables and explore natural hazard risks. For example, in Chapter 4, we found a modest positive correlation between intrinsic building and equipment claims risks. A possible reason for this positive association is that the insured objects (buildings, equipment, and cars) are close to one another and share common environmental hazards such as hail, windstorms, etc. Future research could incorporate spatial information into shared random effects models for multivariate claims, and examine

whether the correlation between intrinsic claims risks still exists.

Finally, this thesis has consistently treated premiums as exogenous mainly because insurers' pricing process (e.g., models, expenses, and profit margin) is confidential and unavailable to us. For example, we treat future premiums as constant in Section 3.4 and adjust future premiums by inflation rate in Section 4.5, which might introduce bias in CLV calculations. If an insurer is willing to share confidential information about their pricing process, then it could facilitate new research into joint models for premiums, claims, and churn, i.e., three main components of CLV calculation.

Bibliography

Abreu, J. E. C. (2019), Customer lifetime value in insurance, PhD thesis.

Alai, D. H., Arnold, S. and Sherris, M. (2015), 'Modelling cause-of-death mortality and the impact of cause-elimination', *Annals of Actuarial Science* **9**(1), 167–186.

Alemany, R., Bolancé, C., Rodrigo, R. and Vernic, R. (2020), 'Bivariate mixed poisson and normal generalised linear models with sarmanov dependenceâan application to model claim frequency and optimal transformed average severity', *Mathematics* **9**(1), 73.

Allen, D. E., McAleer, M. and Singh, A. K. (2017), 'Risk measurement and risk modelling using applications of vine copulas', *Sustainability* **9**(10), 1762.

Antonio, K., Godecharle, E. and Van Oirbeek, R. (2016), 'A multi-state approach and flexible payment distributions for micro-level reserving in general insurance', *Available at SSRN 2777467* .

Ascarza, E. (2018), 'Retention futility: Targeting high-risk customers might be ineffective', *Journal of Marketing Research* **55**(1), 80–98.

Ascarza, E. and Hardie, B. G. (2013), 'A joint model of usage and churn in contractual settings', *Marketing Science* **32**(4), 570–590.

Ascarza, E., Neslin, S. A., Netzer, O., Anderson, Z., Fader, P. S., Gupta, S., Hardie, B. G., Lemmens, A., Libai, B., Neal, D. et al. (2018), 'In pursuit of enhanced customer retention management: Review, key issues, and future directions', *Customer Needs and Solutions* **5**, 65–81.

Avanzi, B., Taylor, G., Wang, M. and Wong, B. (2023), 'Machine learning with high-cardinality categorical features in actuarial applications', *arXiv preprint arXiv:2301.12710* .

Avraham, R. (2017), 'Discrimination and insurance', *The Routledge handbook of the ethics of discrimination* pp. 335–347.

Barnett, J. C. and Vornovitsky, M. S. (2016), *Health insurance coverage in the United States: 2015*, US Government Printing Office Washington, DC.
URL: *https://www.khi.org/assets/uploads/news/14561/p60-257.pdf*

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2015), 'Fitting linear mixed-effects models using lme4', *Journal of Statistical Software* **67**(1), 1–48.

Bäuerle, N. and Müller, A. (1998), 'Modeling and comparing dependencies in multivariate risk portfolios', *ASTIN Bulletin: The Journal of the IAA* **28**(1), 59–76.

Bekkar, M., Djemaa, H. K. and Alitouche, T. A. (2013), 'Evaluation measures for models assessment over imbalanced data sets', *J Inf Eng Appl* **3**(10).

Berger, P. D. and Nasr, N. I. (1998), 'Customer lifetime value: Marketing models and applications', *Journal of interactive marketing* **12**(1), 17–30.

Bettonville, C., d'Oultremont, L., Denuit, M., Trufin, J. and Van Oirbeek, R. (2021), 'Matrix calculation for ultimate and 1-year risk in the semi-Markov individual loss reserving model', *Scandinavian Actuarial Journal* **2021**(5), 380–407.

Bikker, J. A. and Van Leuvensteijn, M. (2008), 'Competition and efficiency in the dutch life insurance industry', *Applied Economics* **40**(16), 2063–2084.

Bolancé, C., Guillen, M. and Padilla-Barreto, A. E. (2016), Predicting probability of customer churn in insurance, *in* 'International Conference on Modeling and Simulation in Engineering, Economics and Management', Springer, pp. 82–91.

Bolancé, C. and Vernic, R. (2020), 'Frequency and severity dependence in the collective risk model: An approach based on sarmanov distribution', *Mathematics* **8**(9), 1400.

Bolton, R. N., Kannan, P. K. and Bramlett, M. D. (2000), 'Implications of loyalty program membership and service experiences for customer retention and value', *Journal of the academy of marketing science* **28**(1), 95–108.

Bonat, W. H. and Kokonendji, C. C. (2017), 'Flexible tweedie regression models for continuous data', *Journal of Statistical Computation and Simulation* **87**(11), 2138–2152.

Booth, K. and Harwood, A. (2016), 'Insurance as catastrophe: A geography of house and contents insurance in bushfire-prone places', *Geoforum* **69**, 44–52.

Borensztein, E., Cavallo, E. and Jeanne, O. (2017), 'The welfare gains from macro-insurance against natural disasters', *Journal of Development Economics* **124**, 142–156.

Borle, S., Singh, S. S. and Jain, D. C. (2008), 'Customer lifetime value measurement', *Management science* **54**(1), 100–112.

Boucher, J.-P. and Couture-Piché, G. (2015), 'Modeling the number of insureds' cars using queuing theory', *Insurance: Mathematics and Economics* **64**, 67–76.

Boucher, J.-P. and Couture-Piché, G. (2016), 'Modeling the number of insured households in an insurance portfolio using queuing theory', *ASTIN Bulletin: The Journal of the IAA* **46**(2), 401–430.

Boucher, J.-P. and Davidov, D. (2011), 'On the importance of dispersion modeling for claims reserving: An application with the tweedie distribution', *Variance* **5**(2), 158–172.

Boucher, J.-P. and Denuit, M. (2006), 'Fixed versus random effects in poisson regression models for claim counts: A case study with motor insurance', *ASTIN Bulletin: The Journal of the IAA* **36**(1), 285–301.

Boucher, J.-P., Denuit, M. and Guillen, M. (2009), 'Number of accidents or number of claims? an approach with zero-inflated poisson models for panel data', *Journal of Risk and Insurance* **76**(4), 821–846.

Boulton, A. J. and Williford, A. (2018), 'Analyzing skewed continuous outcomes with many zeros: A tutorial for social work and youth prevention science researchers', *Journal of the Society for Social Work and Research* **9**(4), 721–740.

Brant Carson, Christine Korwin-Szymanowska, O. L. (2017), 'Mckinsey Australia small commercial insurance consumer survey'.
URL: *https://www.mckinsey.com/industries/financial-services/our-insights/sme-insurance-in-australia-a-market-ripe-for-change*

Braun, M., Fader, P. S., Bradlow, E. T. and Kunreuther, H. (2006), 'Modeling the âpseudodeductibleâ in insurance claims decisions', *Management Science* **52**(8), 1258–1272.

Brechmann, E. and Czado, C. (2014), 'Spatial modeling', *Predictive Modeling Applications in Actuarial Science: Volume 1, Predictive Modeling Techniques* p. 260.

Brian Briggs, E. C. (2021), U.S. property & casualty and title insurance industries - 2020 full year results, Technical report, National Association of Insurance Commissioners.
URL: *https://content.naic.org/*

Brockett, P. L., Golden, L. L., Guillen, M., Nielsen, J. P., Parner, J. and Perez-Marin, A. M. (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* **75**(3), 713–737.

Brown, R. L. and Lennox, W. S. (2015), 'Ratemaking and loss reserving for property and casualty insurance'.

Buchardt, K. (2014), 'Dependent interest and transition rates in life insurance', *Insurance: Mathematics and Economics* **55**, 167–179.

Buchardt, K., Møller, T. and Schmidt, K. B. (2015), 'Cash flows and policyholder behaviour in the semi-Markov life insurance setup', *Scandinavian Actuarial Journal* **2015**(8), 660–688.

Burez, J. and Van den Poel, D. (2009), 'Handling class imbalance in customer churn prediction', *Expert Systems with Applications* **36**(3), 4626–4636.

Butler, J. R. (2002), 'Policy change and private health insurance: did the cheapest policy do the trick?', *Australian Health Review* **25**(6), 33–41.

Chen, K., Huang, R., Chan, N. H. and Yau, C. Y. (2019), 'Subgroup analysis of zero-inflated poisson regression model with applications to insurance data', *Insurance: Mathematics and Economics* **86**, 8–18.

Chiou, W. L. (1978), 'Critical evaluation of the potential error in pharmacokinetic studies of using the linear trapezoidal rule method for the calculation of the area under the plasma level-time curve', *Journal of pharmacokinetics and biopharmaceutics* **6**(6), 539–546.

Christiansen, M. C., Eling, M., Schmidt, J.-P. and Zirkelbach, L. (2016), 'Who is changing health insurance coverage? empirical evidence on policyholder dynamics', *Journal of Risk and Insurance* **83**(2), 269–300.

Clark, R. G., Blanchard, W., Hui, F. K. C., Tian, R. and Woods, H. (2023), 'Dealing with complete separation and quasi-complete separation in logistic regression for linguistic data', *Research Methods in Applied Linguistics* **2**, 100044.

Cohen, A. and Einav, L. (2007), 'Estimating risk preferences from deductible choice', *American economic review* **97**(3), 745–788.

Cortes, C. and Mohri, M. (2003), 'AUC optimization vs. error rate minimization', *Advances in neural information processing systems* **16**.

Danzon, P. (1984), 'The frequency and severity of medical malpractice claims', *The Journal of Law and Economics* **27**(1), 115–148.

Datta, H., Foubert, B. and Van Heerde, H. J. (2015), 'The challenge of retaining customers acquired with free trials', *Journal of Marketing Research* **52**(2), 217–234.

De Bock, K. W. and Van den Poel, D. (2011), 'An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction', *Expert Systems with Applications* **38**(10), 12293–12301.

De Caigny, A., Coussement, K. and De Bock, K. W. (2018), 'A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees', *European Journal of Operational Research* **269**(2), 760–772.

De la Llave, M. Á., López, F. A. and Angulo, A. (2019), 'The impact of geographical factors on churn prediction: an application to an insurance company in madrid's urban area', *Scandinavian Actuarial Journal* **2019**(3), 188–203.

de Miguel, S., Mehtätalo, L., Shater, Z., Kraid, B. and Pukkala, T. (2012), 'Evaluating marginal and conditional predictions of taper models in the absence of calibration data', *Canadian Journal of Forest Research* **42**(7), 1383–1394.

Deelstra, G., Dhaene, J. and Vanmaele, M. (2011), 'An overview of comonotonicity and its applications in finance and insurance', *Advanced mathematical methods for finance* pp. 155–179.

Delong, Ł., Lindholm, M. and Wüthrich, M. V. (2021), 'Making tweedieâs compound poisson model more accessible', *European Actuarial Journal* **11**(1), 185–226.

Denuit, M., Dhaene, J., Goovaerts, M. and Kaas, R. (2006), *Actuarial theory for dependent risks: measures, orders and models*, John Wiley & Sons.

Denuit, M., Dhaene, J. and Ribas, C. (2001), 'Does positive dependence between individual risks increase stop-loss premiums?', *Insurance: Mathematics and Economics* **28**(3), 305–308.

Devriendt, F., Berrevoets, J. and Verbeke, W. (2021), 'Why you should stop predicting customer churn and start using uplift models', *Information Sciences* **548**, 497–515.

Devriendt, S., Antonio, K., Reynkens, T. and Verbelen, R. (2021), 'Sparse regression with multi-type regularized feature modeling', *Insurance: Mathematics and Economics* **96**, 248–261.

Dhaene, J. and Goovaerts, M. J. (1996), 'Dependency of risks and stop-loss order 1', *ASTIN Bulletin: The Journal of the IAA* **26**(2), 201–212.

Dhaene, J. and Goovaerts, M. J. (1997), 'On the dependency of risks in the individual life model', *Insurance: Mathematics and Economics* **19**(3), 243–253.

Dong, Y., Frees, E. W. and Huang, F. (2022), 'Deductible costs for bundled insurance contracts', *Available at SSRN 4020299* .

Dong, Y., Frees, E. W., Huang, F. and Hui, F. K. C. (2022), 'Multi-state modelling of customer churn', *ASTIN Bulletin: The Journal of the IAA* **52**(3), 735–764.

Donkers, B., Verhoef, P. C. and de Jong, M. G. (2007), 'Modeling clv: A test of competing models in the insurance industry', *Quantitative Marketing and Economics* **5**, 163–190.

Dunn, P. K. (2022), *Tweedie: Evaluation of Tweedie Exponential Family Models*. R package version 2.3.5.

Dunn, P. K. and Smyth, G. K. (2008), 'Evaluation of tweedie exponential dispersion models using fourier inversion', *Statistics and Computing* **18**(1), 73–86.

Dwyer, F. R. (1997), 'Customer lifetime valuation to support marketing decision making', *Journal of Direct Marketing* **11**(4), 6–13.

Ebbes, P., Papies, D. and van Heerde, H. J. (2011), 'The sense and non-sense of holdout sample validation in the presence of endogeneity', *Marketing Science* **30**(6), 1115â1122.

Embrechts, P., Frey, R. and McNeil, A. (2011), 'Quantitative risk management'.

Fabio, L. C., Paula, G. A. and de Castro, M. (2012), 'A poisson mixed model with nonnormal random effect distribution', *Computational Statistics & Data Analysis* **56**(6), 1499–1510.

Fader, P. S. and Hardie, B. G. (2010), 'Customer-base valuation in a contractual setting: The perils of ignoring heterogeneity', *Marketing Science* **29**(1), 85–93.

Fader, P. S. and Hardie, B. G. (2016), 'Reconciling and clarifying clv formulas'.

Fagerland, M. W., Hosmer, D. W. and Bofin, A. M. (2008), 'Multinomial goodness-of-fit tests for logistic regression models', *Statistics in medicine* **27**(21), 4238–4253.

Fang, K., Jiang, Y. and Song, M. (2016), 'Customer profitability forecasting using big data analytics: A case study of the insurance industry', *Computers & Industrial Engineering* **101**, 554–564.

Ferrer, L., Rondeau, V., Dignam, J., Pickles, T., Jacqmin-Gadda, H. and Proust-Lima, C. (2016), 'Joint modelling of longitudinal and multi-state processes: application to clinical progressions in prostate cancer', *Statistics in medicine* **35**(22), 3933–3948.

Fong, J. H., Shao, A. W. and Sherris, M. (2015), 'Multistate actuarial models of functional disability', *North American Actuarial Journal* **19**(1), 41–59.

Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielsen, A. and Sibert, J. (2012), 'Ad model builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models', *Optimization Methods and Software* **27**(2), 233–249.

Frank, R. G. and Lamiraud, K. (2009), 'Choice, price competition and complexity in markets for health insurance', *Journal of Economic Behavior & Organization* **71**(2), 550–562.

Frees, E. (2018), 'Loss data analytics', *arXiv preprint arXiv:1808.06718* .
URL: *https://ewfrees.github.io/Loss-Data-Analytics/*

Frees, E., Butt, A. et al. (2022), 'Anu insurable risks'.

Frees, E. W. (2004), *Longitudinal and panel data: analysis and applications in the social sciences*, Cambridge University Press.

Frees, E. W., Bolancé, C., Guillen, M. and Valdez, E. A. (2021), 'Dependence modeling of multivariate longitudinal hybrid insurance data with dropout', *Expert Systems with Applications* **185**, 115552.

Frees, E. W., Gao, J. and Rosenberg, M. A. (2011), 'Predicting the frequency and amount of health care expenditures', *North American Actuarial Journal* **15**(3), 377–392.

Frees, E. W. and Huang, F. (2023), 'The discriminating (pricing) actuary', *North American Actuarial Journal* **27**(1), 2–24.

Frees, E. W., Jin, X. and Lin, X. (2013), 'Actuarial applications of multivariate two-part regression models', *Annals of Actuarial Science* **7**(2), 258.

Frees, E. W., Lee, G. and Yang, L. (2016), 'Multivariate frequency-severity regression models in insurance', *Risks* **4**(1), 4.

Frees, E. W., Meyers, G. and Cummings, A. D. (2011), 'Summarizing insurance scores using a gini index', *Journal of the American Statistical Association* **106**(495), 1085–1098.

Frees, E. W., Meyers, G. and Cummings, A. D. (2014), 'Insurance ratemaking and a gini index', *Journal of Risk and Insurance* **81**(2), 335–366.

Frees, E. W. and Valdez, E. A. (1998), 'Understanding relationships using copulas', *North American actuarial journal* **2**(1), 1–25.

Frees, E. W. and Wang, P. (2005), 'Credibility using copulas', *North American Actuarial Journal* **9**(2), 31–48.

Frees, E. W. and Wang, P. (2006), 'Copula credibility for aggregate loss models', *Insurance: Mathematics and Economics* **38**(2), 360–373.

Frees, E. W., Young, V. R. and Luo, Y. (1999), 'A longitudinal data analysis interpretation of credibility models', *Insurance: Mathematics and Economics* **24**(3), 229–247.

Friedman, J. H. (2001), 'Greedy function approximation: a gradient boosting machine', *Annals of statistics* pp. 1189–1232.

Garrido, J., Genest, C. and Schulz, J. (2016), 'Generalized linear models for dependent frequency and severity of insurance claims', *Insurance: Mathematics and Economics* **70**, 205–215.

Greenwell, B., Boehmke, B., Cunningham, J. and Developers, G. (2020), *gbm: Generalized Boosted Regression Models*. R package version 2.1.8.
URL: *https://CRAN.R-project.org/package=gbm*

Guelman, L. and Guillén, M. (2014), 'A causal inference approach to measure price elasticity in automobile insurance', *Expert Systems with Applications* **41**(2), 387–396.

Guelman, L., Guillén, M. and Pérez-Marín, A. M. (2012), Random forests for uplift modeling: an insurance customer retention case, *in* 'International Conference on Modeling and Simulation in Engineering, Economics and Management', Springer, pp. 123–133.

Guillen, M., Nielsen, J. P., Scheike, T. H. and Pérez-Marín, A. M. (2012), 'Time-varying effects in the analysis of customer loyalty: A case study in insurance', *Expert systems with Applications* **39**(3), 3551–3558.

Guillen, M., Nielsen, J. and Pérez-Marín, A. (2009), 'Cross buying behaviour and customer loyalty in the insurance sector', *EsicMarket* **132**, 77–105.

Guillen, M., Parner, J., Densgsoe, C. and Perez-Marin, A. M. (2003), Using logistic regression models to predict and understand why customers leave an insurance company, *in* 'Intelligent and other computational techniques in insurance: Theory and applications', World Scientific, pp. 465–490.

Günther, C.-C., Tvete, I. F., Aas, K., Sandnes, G. I. and Borgan, Ø. (2014), 'Modelling and predicting customer churn from an insurance company', *Scandinavian Actuarial Journal* **2014**(1), 58–71.

Gupta, S. and Lehmann, D. (2005), *Managing customers as investments the strategic value of customers in the long run*, Wharton School Publishing.

Gupta, S., Lehmann, D. R. and Stuart, J. A. (2004), 'Valuing customers', *Journal of marketing research* **41**(1), 7–18.

Haberman, S. and Pitacco, E. (2018), *Actuarial models for disability insurance*, Routledge.

Hall, D. B. (2000), 'Zero-inflated poisson and binomial regression with random effects: a case study', *Biometrics* **56**(4), 1030–1039.

Hand, D. J. and Till, R. J. (2001), 'A simple generalisation of the area under the ROC curve for multiple class classification problems', *Machine learning* **45**(2), 171–186.

Haugen, M. and Moger, T. A. (2016), 'Frailty modelling of time-to-lapse of single policies for customers holding multiple car contracts', *Scandinavian Actuarial Journal* **2016**(6), 489–501.

He, H. and Ma, Y. (2013), *Imbalanced learning: foundations, algorithms, and applications*, John Wiley & Sons.

He, Y., Xiong, Y. and Tsai, Y. (2020), Machine learning based approaches to predict customer churn for an insurance company, *in* '2020 Systems and Information Engineering Design Symposium (SIEDS)', IEEE, pp. 1–6.

Hickey, G. L., Philipson, P., Jorgensen, A. and Kolamunnage-Dona, R. (2016), 'Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues', *BMC medical research methodology* **16**(1), 1–15.

Hofert, M., Kojadinovic, I., Maechler, M. and Yan, J. (2020), *copula: Multivariate Dependence with Copulas*. R package version 1.0-1.
URL: *https://CRAN.R-project.org/package=copula*

Hoffmann, R. and Kassouf, A. L. (2005), 'Deriving conditional and unconditional marginal effects in log earnings equations estimated by heckman's procedure', *Applied Economics* **37**(11), 1303–1311.

Holtrop, N., Wieringa, J. E., Gijsenberg, M. J. and Verhoef, P. C. (2017), 'No future

without the past? predicting churn in the face of customer privacy', *International Journal of Research in Marketing* **34**(1), 154–172.

Hosmer, D. W. and Lemesbow, S. (1980), 'Goodness of fit tests for the multiple logistic regression model', *Communications in statistics-Theory and Methods* **9**(10), 1043–1069.

Hsu, W.-K., Huang, P.-C., Chang, C.-C., Chen, C.-W., Hung, D.-M. and Chiang, W.-L. (2011), 'An integrated flood risk assessment model for property insurance industry in taiwan', *Natural Hazards* **58**, 1295–1309.

Hui, F. K. C. and Bondell, H. D. (2021), 'A shared parameter mixture model for longitudinal income data with missing responses and zero rounding', *Australian & New Zealand Journal of Statistics* **63**(2), 221–240.

Hui, F. K. C., Müller, S. and Welsh, A. H. (2021), 'Random effects misspecification can have severe consequences for random effects inference in linear mixed models', *International Statistical Review* **89**(1), 186–206.

Humphreys, B. R. (2013), 'Dealing with zeros in economic data', *University of Alberta, Department of Economics* **1**, 1–27.

Hürlimann, W. (2015), 'A simple multi-state gamma claims reserving model', *International Journal of Contemporary Mathematical Sciences* **10**(2), 65–77.

Ibrahim, J. G., Chu, H. and Chen, L. M. (2010), 'Basic concepts and methods for joint models of longitudinal and survival data', *Journal of Clinical Oncology* **28**(16), 2796.

Jeong, H., Gan, G. and Valdez, E. A. (2018), 'Association rules for understanding policyholder lapses', *Risks* **6**(3), 69.

Jeong, H., Tzougas, G. and Fung, T. C. (2023), 'Multivariate claim count regression model with varying dispersion and dependence parameters', *Journal of the Royal Statistical Society Series A: Statistics in Society* **186**(1), 61–83.

Jimenez, J. H. (2022), 'Insurance accounts for more than 7% of the global economy'. **URL:** *https://www.mapfre.com/en/insights/insurance/insurance-accounts-for-more-than-seven-percent-the-global-economy/*

Jørgensen, B. and Paes De Souza, M. C. (1994), 'Fitting tweedie's compound poisson model to insurance claims data', *Scandinavian Actuarial Journal* **1994**(1), 69–93.

Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. (2008), *Modern actuarial risk theory: using R*, Vol. 128, Springer Science & Business Media.

Kousky, C. (2019), 'The role of natural disaster insurance in recovery and risk reduction', *Annual Review of Resource Economics* **11**, 399–418.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H. and Bell, B. M. (2016), 'TMB: Automatic Differentiation and Laplace Approximation', *Journal of Statistical Software* **70**, 1–21.

Kuhn, M. (2021), *caret: Classification and Regression Training*. R package version 6.0-88.
URL: *https://CRAN.R-project.org/package=caret*

Kumar, V. and Reinartz, W. (2016), 'Creating enduring customer value', *Journal of marketing* **80**(6), 36–68.

Kumar, V. and Reinartz, W. (2018), *Customer relationship management*, Springer.

Kumar, V., Venkatesan, R., Bohling, T. and Beckmann, D. (2008), 'Practice prize reportâthe power of clv: Managing customer lifetime value at ibm', *Marketing science* **27**(4), 585–599.

Kurz, C. F. (2017), 'Tweedie distributions for fitting semicontinuous health care utilization cost data', *BMC Medical Research Methodology* **17**, 1–8.

Kwon, H.-S. and Jones, B. L. (2006), 'The impact of the determinants of mortality on life insurance and annuities', *Insurance: Mathematics and Economics* **38**(2), 271–288.

Kwon, H.-S. and Jones, B. L. (2008), 'Applications of a multi-state risk factor/mortality model in life insurance', *Insurance: Mathematics and Economics* **43**(3), 394–402.

Kwon, H.-S. and Nguyen, V. H. (2019), 'Analysis of cause-of-death mortality and actuarial implications', *Communications for Statistical Applications and Methods* **26**(6), 557–573.

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

Lee, A. H., Wang, K., Scott, J. A., Yau, K. K. and McLachlan, G. J. (2006), 'Multi-level zero-inflated poisson regression modelling of correlated count data with excess zeros', *Statistical methods in medical research* **15**(1), 47–61.

Lee, G. Y. (2017), 'General insurance deductible ratemaking', *North American Actuarial Journal* **21**(4), 620–638.

Lee, G. Y. and Shi, P. (2019), 'A dependent frequency–severity approach to modeling longitudinal insurance claims', *Insurance: Mathematics and Economics* **87**, 115–129.

Leeper, T. J. (2017), 'Interpreting regression results using average marginal effects with r's margins', *Reference manual* **32**.

Lehmann, E. L. (1966), 'Some concepts of dependence', *The Annals of Mathematical Statistics* pp. 1137–1153.

Leiria, M., Rebelo, E. and deMatos, N. (2021), 'Measuring the effectiveness of intermediary loyalty programmes in the motor insurance industry: loyal versus non-loyal customers', *European Journal of Management and Business Economics* .

Lemmens, A. and Croux, C. (2006), 'Bagging and boosting classification trees to predict churn', *Journal of Marketing Research* **43**(2), 276–286.

Lemmens, A. and Gupta, S. (2020), 'Managing churn to maximize profits', *Marketing Science* **39**(5), 956–973.

Lepthien, A., Papies, D., Clement, M. and Melnyk, V. (2017), 'The ugly side of customer management: Consumer reactions to firm-initiated contract terminations', *International Journal of Research in Marketing* **34**(4), 829–850.

Lewis, M. (2004), 'The influence of loyalty programs and short-term promotions on customer retention', *Journal of marketing research* **41**(3), 281–292.

Lewis, M. (2005), 'Research note: A dynamic programming approach to customer relationship pricing', *Management science* **51**(6), 986–994.

Lindholm, M., Richman, R., Tsanakas, A. and Wüthrich, M. V. (2022), 'Discrimination-free insurance pricing', *ASTIN Bulletin: The Journal of the IAA* **52**(1), 55–89.

Liu, D.-R. and Shih, Y.-Y. (2005*a*), 'Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences', *Journal of Systems and Software* **77**(2), 181–191.

Liu, D.-R. and Shih, Y.-Y. (2005*b*), 'Integrating ahp and data mining for product recommendation based on customer lifetime value', *Information & Management* **42**(3), 387–400.

Loisel, S., Piette, P. and Tsai, C.-H. J. (2019), 'Applying economic measures to

lapse risk management with machine learning approaches', *ASTIN Bulletin: The Journal of the IAA* pp. 1–33.

Lu, N., Lin, H., Lu, J. and Zhang, G. (2012), 'A customer churn prediction model in telecom industry using boosting', *IEEE Transactions on Industrial Informatics* **10**(2), 1659–1665.

Madigan, D., Genkin, A., Lewis, D. D. and Fradkin, D. (2005), Bayesian multinomial logistic regression for author identification, *in* 'AIP conference proceedings', Vol. 803, American Institute of Physics, pp. 509–516.

Malthouse, E. C. and Blattberg, R. C. (2005), 'Can we predict customer lifetime value?', *Journal of interactive marketing* **19**(1), 2–16.

Mankiw, N. G. (2003), *Macroeconomics*, eighth edn, New York: Worth Publishers.

McAneney, J., McAneney, D., Musulin, R., Walker, G. and Crompton, R. (2016), 'Government-sponsored natural disaster insurance pools: A view from down-under', *International Journal of Disaster Risk Reduction* **15**, 1–9.

McCarthy, D. M., Fader, P. S. and Hardie, B. G. (2017), 'Valuing subscription-based businesses using publicly disclosed customer data', *Journal of Marketing* **81**(1), 17–35.

McDaniel, J. (2023), *Citing climate change risks, Farmers is latest insurer to exit Florida.*
URL: *https://www.washingtonpost.com/climate-environment/2023/07/12/farmers-insurance-leaves-florida/*

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A. and Leisch, F. (2021), *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien.* R package version 1.7-9.
URL: *https://CRAN.R-project.org/package=e1071*

Min, H. and Zhou, G. (2002), 'Supply chain modeling: past, present and future', *Computers & industrial engineering* **43**(1-2), 231–249.

Morrison, D. G. (1969), 'On the interpretation of discriminant analysis', *Journal of marketing research* **6**(2), 156–163.

Müller, A. (1997), 'Stop-loss order for portfolios of dependent risks', *Insurance: Mathematics and Economics* **21**(3), 219–223.

NAIC (2018), A regulator's guide to pet insurance, Technical report. Draft:

10/31/18.
URL: *https://us.eversheds-sutherland.com/portalresource/cmte$_{ce}$xposure$_{p}$et$_{i}$nsurance$_{w}$hite$_{p}$aper.pdf*

Nelsen, R. B. (2007), *An introduction to copulas*, Springer Science & Business Media.

Neslin, S. A., Gupta, S., Kamakura, W., Lu, J. and Mason, C. H. (2006), 'Defection detection: Measuring and understanding the predictive accuracy of customer churn models', *Journal of marketing research* **43**(2), 204–211.

Ng, S., Lestari, D. and Devila, S. (2019), Generalized linear model for deductible pricing in non-life insurance, *in* 'AIP Conference Proceedings', Vol. 2168, AIP Publishing LLC, p. 020038.

Ni, W., Constantinescu, C., Pantelous, A. A. et al. (2014), 'Bonus-malus systems with weibull distributed claim severities', *Annals of Actuarial Science* **8**(2), 217–233.

Nizar Al-Malkawi, H.-A. (2007), 'Determinants of corporate dividend policy in jordan: an application of the tobit model', *Journal of Economic and Administrative Sciences* **23**(2), 44–70.

Noble, W. S. (2006), 'What is a support vector machine?', *Nature biotechnology* **24**(12), 1565–1567.

Norberg, R. (2013), 'Optimal hedging of demographic risk in life insurance', *Finance and Stochastics* **17**(1), 197–222.

OCI (2019), 'Local government property insurance fund overview'.
URL: *https://oci.wi.gov/Pages/Funds/LGPIFOverview.aspx*

Okine, A. N.-A., Frees, E. W. and Shi, P. (2022), 'Joint model prediction and application to individual-level loss reserving', *ASTIN Bulletin: The Journal of the IAA* **52**(1), 91–116.

Okongwu, U., Lauras, M., François, J. and Deschamps, J.-C. (2016), 'Impact of the integration of tactical supply chain planning determinants on performance', *Journal of Manufacturing Systems* **38**, 181–194.

Orr, J. (2007), 'A simple multi-state reserving model', *ASTIN, Colloqiua Orlando edition* .
URL: *https://www.actuaries.org/ASTIN/Colloquia/Orlando/Papers/Orr.pdf*

Pai, J., Shand, K. and Wang, X. (2005), 'Preliminary analysis of pet insurance data', *SOA ARCH* .

URL: *https://www.soa.org/globalassets/assets/files/static-pages/research/arch/2005/arch05v39n1*$_1$*7.pdf*

Papies, D., Ebbes, P. and Van Heerde, H. J. (2017), 'Addressing endogeneity in marketing models', *Advanced methods for modeling markets* pp. 581–627.

Paredes, M. (2018), A case study on reducing auto insurance attrition with econometrics, machine learning, and a/b testing, *in* '2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)', IEEE, pp. 410–414.

Parodi, P. (2023), *Pricing in general insurance*, Chapman and Hall/CRC.

Pauly, M. V. (1978), Overinsurance and public provision of insurance: The roles of moral hazard and adverse selection, *in* 'Uncertainty in economics', Elsevier, pp. 307–331.

Pendharkar, P. C. (2009), 'Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services', *Expert Systems with Applications* **36**(3), 6714–6720.

Pettitt, A. N. and Stephens, M. A. (1977), 'The kolmogorov-smirnov goodness-of-fit statistic with discrete and grouped data', *Technometrics* **19**(2), 205–210.

Pfeifer, P. E. and Carraway, R. L. (2000), 'Modeling customer relationships as markov chains', *Journal of interactive marketing* **14**(2), 43–55.

Pinquet, J. (2000), Experience rating through heterogeneous models, *in* 'Handbook of Insurance', Springer, pp. 459–500.

Price, K., Storn, R. M. and Lampinen, J. A. (2006), *Differential evolution: a practical approach to global optimization*, Springer Science & Business Media.

Price, K. V. (2013), Differential evolution, *in* 'Handbook of Optimization', Springer, pp. 187–214.

Pritchard, D. (2006), 'Modeling disability in long-term care insurance', *North American Actuarial Journal* **10**(4), 48–75.

Puelz, R. and Snow, A. (1994), 'Evidence on adverse selection: Equilibrium signaling and cross-subsidization in the insurance market', *Journal of Political Economy* **102**(2), 236–257.

Quan, Z. and Valdez, E. A. (2018), 'Predictive analytics of insurance claims using multivariate decision trees', *Dependence Modeling* **6**(1), 377–407.

Randall, T., Terwiesch, C. and Ulrich, K. T. (2007), 'Research noteâuser design of customized products', *Marketing Science* **26**(2), 268–280.

Resti, Y., Ismail, N. and Jamaan, S. H. (2013), 'Estimation of claim cost data using zero adjusted gamma and inverse gaussian regression models', *Journal of Mathematics and Statistics* **9**(3), 186–192.

Rick, K. (2015), 'Qa: Local government property insurance fund'.
URL: *https://www.myknowledgebroker.com/blog*

Rizopoulos, D., Hatfield, L. A., Carlin, B. P. and Takkenberg, J. J. (2014), 'Combining dynamic predictions from joint models for longitudinal and time-to-event data using bayesian model averaging', *Journal of the American Statistical Association* **109**(508), 1385–1397.

Rust, R. T., Kumar, V. and Venkatesan, R. (2011), 'Will the frog change into a prince? predicting future customer profitability', *International Journal of Research in Marketing* **28**(4), 281–294.

Rust, R. T., Lemon, K. N. and Zeithaml, V. A. (2004), 'Return on marketing: Using customer equity to focus marketing strategy', *Journal of marketing* **68**(1), 109–127.

Salvadori, G., De Michele, C., Kottegoda, N. T. and Rosso, R. (2007), *Extremes in nature: an approach using copulas*, Vol. 56, Springer Science & Business Media.

Schweidel, D. A., Bradlow, E. T. and Fader, P. S. (2011), 'Portfolio dynamics for customers of a multiservice provider', *Management Science* **57**(3), 471–486.

Scriney, M., Nie, D. and Roantree, M. (2020), Predicting customer churn for insurance data, *in* 'International Conference on Big Data Analytics and Knowledge Discovery', Springer, pp. 256–265.

Seyerle, M. (2001), *Customer lifetime value and its determination using the SAS Entereprise Miner and the SAS-OROS software.* SAS Institute Germany.
URL: *https://support.sas.com/resources/papers/proceedings-archive/SEUGI2003/SEYERLE_LifetimeValue.pdf*

Shi, P. (2016), 'Insurance ratemaking using a copula-based multivariate tweedie model', *Scandinavian Actuarial Journal* **2016**(3), 198–215.

Shi, P., Feng, X. and Ivantsova, A. (2015), 'Dependent frequency–severity modeling of insurance claims', *Insurance: Mathematics and Economics* **64**, 417–428.

Shi, P. and Frees, E. W. (2011), 'Dependent loss reserving using copulas', *ASTIN Bulletin: The Journal of the IAA* **41**(2), 449–486.

Shi, P., Fung, G. M. and Dickinson, D. (2022), 'Assessing hail risk for property insurers with a dependent marked point process', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* .

Shi, P. and Lee, G. Y. (2022), 'Copula regression for compound distributions with endogenous covariates with applications in insurance deductible pricing', *Journal of the American Statistical Association* pp. 1–16.

Shi, P. and Valdez, E. A. (2014), 'Multivariate negative binomial models for insurance claim counts', *Insurance: Mathematics and Economics* **55**, 18–29.

Shi, P. and Yang, L. (2018), 'Pair copula constructions for insurance experience rating', *Journal of the American Statistical Association* **113**(521), 122–133.

Shi, P. and Zhao, Z. (2024), 'Enhanced pricing and management of bundled insurance risks with dependence-aware prediction using pair copula construction', *Journal of Econometrics* **240**(1), 105676.

Shih, Y.-Y. and Liu, D.-R. (2008), 'Product recommendation approaches: Collaborative filtering via customer lifetime value and customer demands', *Expert Systems with Applications* **35**(1-2), 350–360.

Siegelman, P. (2003), 'Adverse selection in insurance markets: an exaggerated threat', *Yale LJ* **113**, 1223.

Simchi-Levi, D., Kaminsky, P., Simchi-Levi, E. and Shankar, R. (2008), *Designing and managing the supply chain: concepts, strategies and case studies*, Tata McGraw-Hill Education.

Sklar, M. (1959), 'Fonctions de repartition an dimensions et leurs marges', *Publ. inst. statist. univ. Paris* **8**, 229–231.

St-Jean, A. (2021), Customer profitability forecasting using fair boosting: an application to the insurance industry, PhD thesis, Université Laval.

Stata (2022), Estimation and postestimation commands, Technical report.
URL: *https://www.stata.com/manuals13/u20.pdf*

Staudt, Y. and Wagner, J. (2018), 'What policyholder and contract features determine the evolution of non-life insurance customer relationships?', *International Journal of Bank Marketing* .

Suncorp (2013), 'Calculated risk: A look into suncorp's pricing calculations', *Journal of marketing* .

Teodorescu, S. and Vernic, R. (2006), 'A composite exponential-pareto distribution', *An. Stiint. Univ.âOvidiusâ Constanta Ser. Mat* **14**(1), 99–108.

Tuerlinckx, F., Rijmen, F., Verbeke, G. and De Boeck, P. (2006), 'Statistical inference in generalized linear mixed models: A review', *British Journal of Mathematical and Statistical Psychology* **59**(2), 225–255.

Van de Ven, W. P. and Van Praag, B. M. (1981), 'The demand for deductibles in private health insurance: A probit model with sample selection', *Journal of econometrics* **17**(2), 229–252.

van Heerde, H. J., Gijsbrechts, E. and Pauwels, K. (2008), 'Winners and losers in a major price war', *Journal of Marketing Research* **45**(5), 499–518.

Venables, W. N. and Ripley, B. D. (2002), *Modern Applied Statistics with S*, fourth edn, Springer, New York. ISBN 0-387-95457-0.
URL: *https://www.stats.ox.ac.uk/pub/MASS4/*

Venkatesan, R. and Kumar, V. (2004), 'A customer lifetime value framework for customer selection and resource allocation strategy', *Journal of marketing* **68**(4), 106–125.

Venkatesan, R., Kumar, V. and Bohling, T. (2007), 'Optimal customer relationship management using bayesian decision theory: An application for customer selection', *Journal of marketing research* **44**(4), 579–594.

Venter, G. G. et al. (2002), Tails of copulas, *in* 'Proceedings of the Casualty Actuarial Society', Vol. 89, pp. 68–113.

Verschuren, R. M. (2022), 'Customer price sensitivities in competitive insurance markets', *Expert Systems with Applications* **202**, 117133.

Wang, J. L., Chung, C.-F. and Tzeng, L. Y. (2008), 'An empirical analysis of the effects of increasing deductibles on moral hazard', *Journal of Risk and Insurance* **75**(3), 551–566.

Welikadaarachchi, S., Botheju, M., Ariyasena, D., Fernando, V., Gamage, A. and Gunathilake, P. (2023), 'Data driven approach to improve profitability in vehicle insurance sector', *International Research Journal of Innovations in Engineering and Technology* **7**(10), 333.

Werner, G. and Modlin, C. (2010), Basic ratemaking, *in* 'Casualty Actuarial Society', Vol. 4, pp. 1–320.

Wood, S. N. (2011), 'Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **73**(1), 3–36.

Wu, X.-Z. and Zhou, Z.-H. (2017), A unified view of multi-label performance measures, *in* 'International Conference on Machine Learning', PMLR, pp. 3780–3788.

Wulfsohn, M. S. and Tsiatis, A. A. (1997), 'A joint model for survival and longitudinal data measured with error', *Biometrics* pp. 330–339.

Xin, X. and Huang, F. (2023), 'Antidiscrimination insurance pricing: Regulations, fairness criteria, and models', *North American Actuarial Journal* pp. 1–35.

Yang, L., Frees, E. W. and Zhang, Z. (2020), 'Nonparametric estimation of copula regression models with discrete outcomes', *Journal of the American Statistical Association* **115**(530), 707–720.

Yang, L. and Shi, P. (2019), 'Multiperil rate making for property insurance using longitudinal data', *Journal of the Royal Statistical Society Series A: Statistics in Society* **182**(2), 647–668.

Yeo, A. C., Smith, K. A., Willis, R. J. and Brooks, M. (2001), Modeling the effect of premium changes on motor insurance customer retention rates using neural networks, *in* 'International Conference on Computational Science', Springer, pp. 390–399.

Yip, K. C. and Yau, K. K. (2005), 'On modeling claim frequency data in general insurance with extra zeros', *Insurance: Mathematics and Economics* **36**(2), 153–163.

Zhang, J. Q., Dixit, A. and Friedmann, R. (2010), 'Customer loyalty and lifetime value: An empirical investigation of consumer packaged goods', *Journal of marketing theory and practice* **18**(2), 127–140.

Zhu, Z., Li, Z., Wylde, D., Failor, M. and Hrischenko, G. (2015), 'Logistic regression for insured mortality experience studies', *North American Actuarial Journal* **19**(4), 241–255.

Appendix for Chapter 2

A.1 Applying Two Combined Deductibles: One for Four Autos, One for the Whole Contract

In Section 2.4.1 of Chapter 2, we mention that an insurer can offer a combined deductible d_{auto} to the four auto coverages (PN, PO, CN, CO), and then a combined deductible d to the entire contract. This is only one of the examples that insurers may encounter in practice. In this case, the insurer should cover the loss of $\left[X_{BC} + X_{IM} + (X_{PN} + X_{PO} + X_{CN} + X_{CO} - d_{auto})_+ - d\right]_+$ while the remaining loss is undertaken by the policyholder. If we hold d to be 2500 and let d_{auto} change from 0 to 5000, then the CR is shown in Figure A.1. We can see the CR is around 20% for these four policyholders, which is consistent with the finding of the contracts with one combined deductible in Section 2.4.1 of Chapter 2.

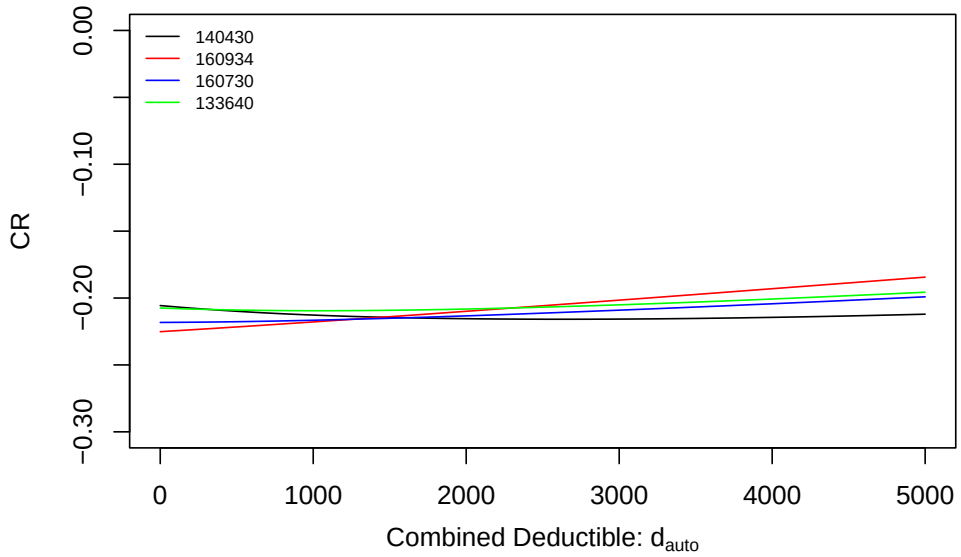

Figure A.1: Cost Relativities of four policyholders with two combined deductibles.

A.2 Discussion on the $\lim_{d\to 0^+} CR$ for Section 2.4.1

In Figure 2.1 and Figure 2.3 of Chapter 2, we see the ratios CR have different starting points when $d \to 0^+$ and starting points are all negative. This phenomenon can be explained through the mathematical derivations below.

$$\begin{aligned}
&\lim_{d\to 0^+} c_d = \lim_{d\to 0^+} E\ \min(X_{BC} + X_{IM} + X_{PN} + X_{PO} + X_{CN} + X_{CO}, d) \\
&= P(\text{0-claim}) \times 0 + \big(1 - P(\text{0-claim})\big) \times d \\
&= d - P(\text{0-claim})d.
\end{aligned} \tag{A.1}$$

$$\begin{aligned}
\lim_{d\to 0^+} CR &= \frac{c_d^{dep}}{c_d^{ind}} - 1 \\
&= \frac{d - P^{dep}(\text{0-claim})d}{d - P^{ind}(\text{0-claim})d} - 1 \\
&= \frac{P^{ind}(\text{0-claim}) - P^{dep}(\text{0-claim})}{1 - P^{ind}(\text{0-claim})}.
\end{aligned} \tag{A.2}$$

In equations (A.1) and (A.2), "0-claim" means that there is no claim for all six

coverages. Equation (A.2) suggests that the starting point of CR is affected by P^{dep}(0-claim) and P^{ind}(0-claim), where P^{dep}(0-claim) (resp. P^{ind}(0-claim)) is the probability of "0-claim" calculated by the dependence model (resp. independence model).

Let $u_{BC}, u_{IM}, u_{PN}, u_{PO}, u_{CN}, and u_{CO}$ denote the probabilities of no claim for the six coverages BC, IM, PN, PO, CN, and CO, respectively. For example, u_{BC} is the probability of no claim for the BC coverage. Then P^{ind}(0-claim) for the independence model and P^{dep}(0-claim) for the dependence model can be calculated by

$$P^{ind}(\text{0-claim}) = u_{BC}u_{IM}u_{PN}u_{PO}u_{CN}u_{CO}, \tag{A.3a}$$

$$P^{dep}(\text{0-claim}) = C(u_{BC}, u_{IM}, u_{PN}, u_{PO}, u_{CN}, u_{CO}). \tag{A.3b}$$

Combining equations (A.2), (A.3a), (A.3b), it is straightforward to have the following conclusions about the starting point of CR.

- Marginal distributions of the losses affect the starting points of CR, as the no-claim probabilities for the six coverages are different.
- The choice of the copula for the dependence model also affects CR.
- The reason that all starting points are negative is due to

$$C(u_{BC}, u_{IM}, u_{PN}, u_{PO}, u_{CN}, u_{CO}) > u_{BC}u_{IM}u_{PN}u_{PO}u_{CN}u_{CO}.$$

  In Table 2.3 of Chapter 2, all dependence parameters are positive, which implies the losses of six coverages modelled by the dependence model tend to be small together compared to the independent case. So, there is a higher chance for the dependence model to have a "0-claim" compared with the independence model.

To visualise the above conclusions, we assume no claim probabilities of six coverages are the same ($u_{BC} = u_{IM} = u_{PN} = u_{PO} = u_{CN} = u_{CO} = u$). Then we can observe how different values of u affect the starting point of CR in Figure A.2. It further explains why the starting points of CR of 24 policyholders are all different in Figure 2.1 of Chapter 2.

In Figure A.3, we consider both a Gaussian copula and Student-t copulas with the same dependence parameters listed in Table 2.3. For the Student-t copulas, we set degrees of freedom to be 2, 4, 6, and 20 for comparison purposes. We find that the Student-t copula with a lower degree of freedom leads to a lower starting point of

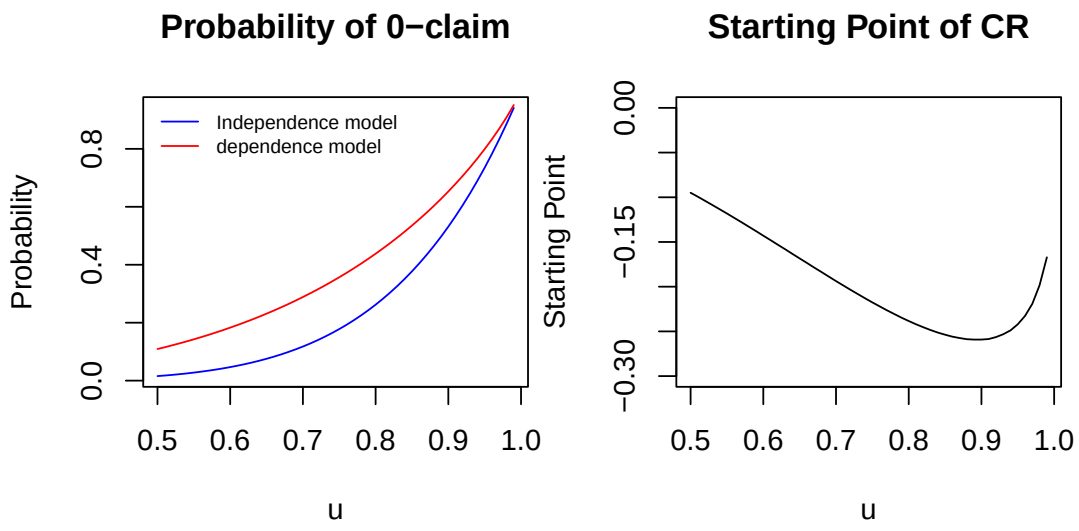

Figure A.2: How u affects starting points of Cost Relativities (dependence and independence model).

CR, holding other conditions the same. The Student-t copula approximates to the Gaussian copula when its degree of freedom approximates to positive infinity, which explains that the green line and black line are close.

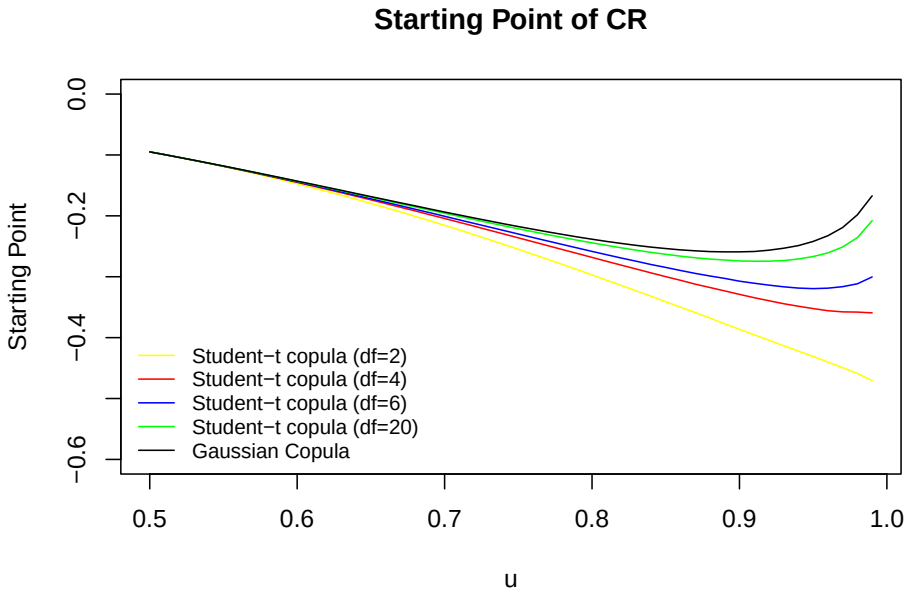

Figure A.3: How u affects the starting points of Cost Relativities (Student-t copula and Gaussian copula).

A.3 Two Probabilities related to Section 2.4.1

Define $S = X_{BC} + X_{IM} + X_{PN} + X_{PO} + X_{CN} + X_{CO}$ to be the total claim amount of six coverages. We introduce two probabilities that practitioners may be considered. These two probabilities are calculated by Monte Carlo simulations and are defined as follows.

$$P(S < d) = \frac{R(0 < S < d) + R(S = 0)}{R},$$

$$P(0 < S < d | S > 0) = \frac{R(0 < S < d)}{R - R(S = 0)},$$

where R is the total number of the simulation, $R(0 < S < d)$ and $R(S = 0)$ are the numbers of simulation that satisfies $0 < S < d$ and $S = 0$ respectively. The relationship between these two probabilities and the combined deductible of the four policyholders discussed in Section 2.4.1 of Chapter 2 is shown in Figure A.4.

We use $P(S \leq d)$ to denote the unconditional probability that the insurer will not pay anything to the policyholder. $P(S \leq d) = P(0 < S \leq d) + P(S = 0)$. As d increases, $P(S \leq d)$ will also increase. If $d = 0$, then $P(S \leq d) = P(S = 0)$, which can be observed in the left figures in Figure A.4. Also, the dependence model leads to higher $P(S \leq d)$ compared with the independence model because of a higher $P(S = 0)$. This is consistent with the "0-claim" probabilities in Figure A.2.

We use $P(0 < S \leq d | S > 0)$ to denote the conditional probability that the insurer will not pay anything to the policyholder, given the policyholder has a loss. It represents the proportional decrement in the probability of paying a claim by setting a combined deductible. For example, when there is no combined deductible, the insurer must pay the policyholder for the loss amount. If there is a combined deductible of 100 and $P(0 < S \leq 100 | S > 0) = 20\%$, then the probability of paying a claim by the insurer will decrease by 20%.

A.4 Models and Calibrated Parameters for Section 2.4.3

We summarise the calibration process in the following two steps.

Step 1. Write a function with the five parameters being inputs and MAPE being the output.

Step 2. Use DE optimisation to get five optimised parameters through minimising

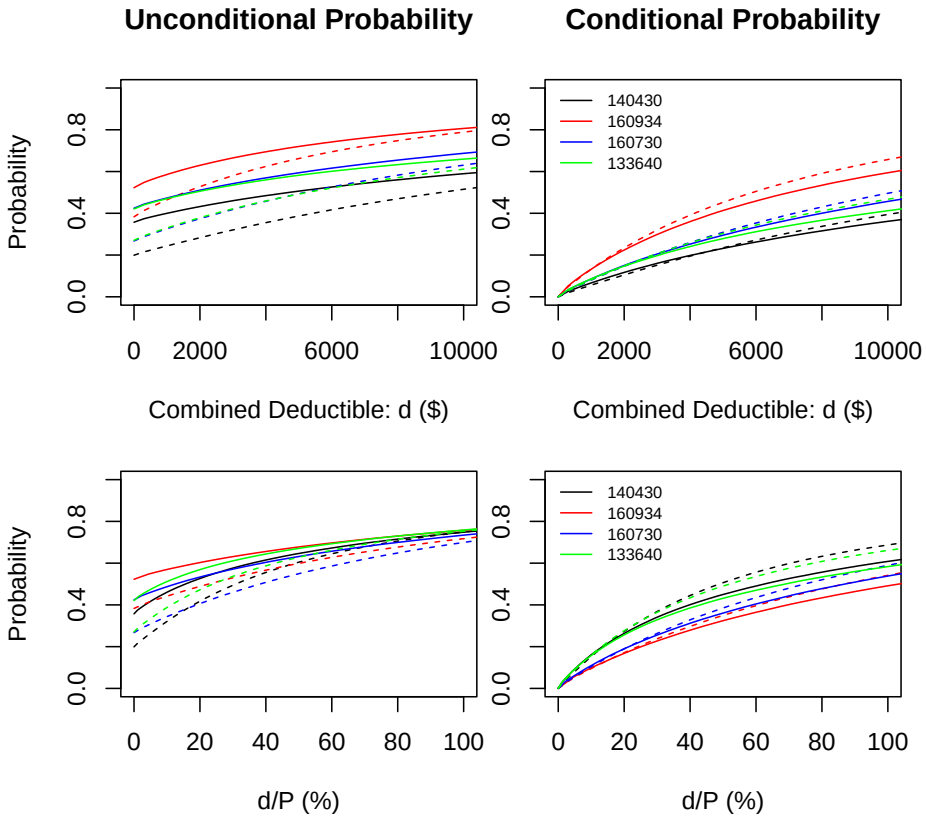

Figure A.4: Two probabilities by two forms of combined deductible (dependence: solid line, independence: dash line).

the MAPE. The optimisation process will stop if the maximum number of iterations is reached or the best parameter vector finds a value. The maximum number of iterations is set to 200 to ensure optimised parameters are stable.

DE optimises a problem by maintaining a population of candidate solutions that will be used to create new candidate solutions. Then, DE optimisation only keeps whichever candidate solution has the best score or fitness on the optimisation problem. In this way, the optimisation problem is treated as a black box that merely provides a measure of quality given a candidate solution. To see the introduction of the DE algorithm and more applications, see Price et al. (2006) and Price (2013).

A.5 Comonotonic and Countermonotonic Results of Three Applications

We use $CR_{co} = \frac{c_d^{dep}}{c_d^{co}} - 1$ and $CR_{counter} = \frac{c_d^{dep}}{c_d^{counter}} - 1$ to denote the cost relativities of the comonotonic scenario and countermonotonic scenario respectively.

A.5.1 Application 1: Wisconsin Property Fund (Commercial)

Since there are six coverages in this bundled contract, we only consider the comonotonic scenario, which is considered as the perfect positive dependence structure. In Figure A.5, the cost relativities based on the comonotonic benchmark are all positive for four specific policies, which means the deductible costs based on the dependence model are larger than the corresponding deductible costs based on the comonotonic model. In general, the deductible costs are increased by approximately 50 to 100 per cent when dependence models are used compared to the benchmark models that assume comonotonicity for four specific policies. This result can be explained by Proposition 2.6.2 in Section 2.6 of Chapter 2. Comonotonicity refers to the perfect positive dependence, so it has the largest correlation order as well as the smallest deductible costs.

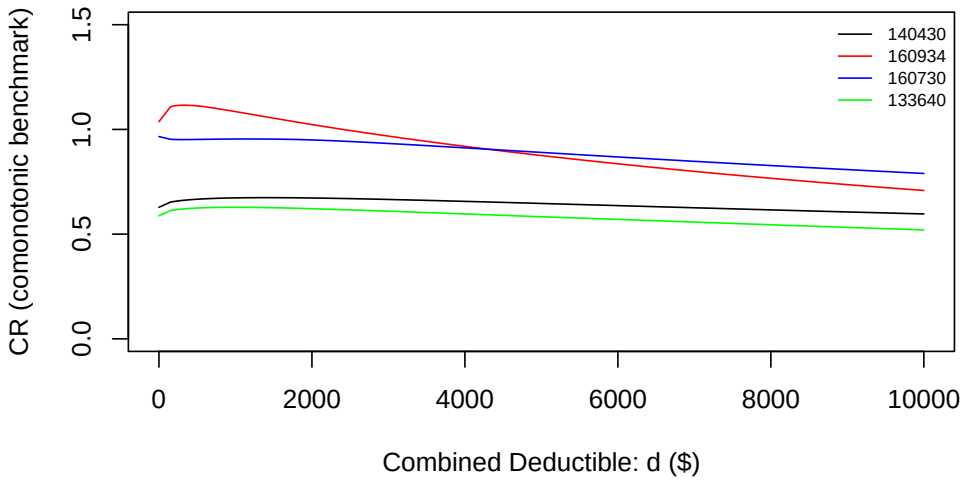

Figure A.5: Cost Relativities of four policyholders with the comonotonic benchmark.

A.5.2 Application 2: Insurance Company Loss and Expense (Reinsurance)

We consider both the comonotonic scenario and countermonotonic scenario for the loss and ALAE. Countermonotonic scenario is often considered by a progressive risk manager who thinks that two risks can mutually hedge, which can be regarded as the least risky scenario. In Figure A.6, the cost relativities are all positive when we assume a comonotonic benchmark and are negative when we assume a countermonotonic benchmark. In general, the deductible costs are increased by approximately 0 to 10 per cent when dependence models are used compared to the benchmark models that assume comonotonicity, and reduced by 0 to 20 per cent compared to the benchmark models that assume comonotonicity. This result can also be explained by Proposition 2.6.2 in Section 2.6 of Chapter 2. Countermonotonicity refers to the perfect negative dependence, so it has the smallest correlation order as well as the highest deductible costs.

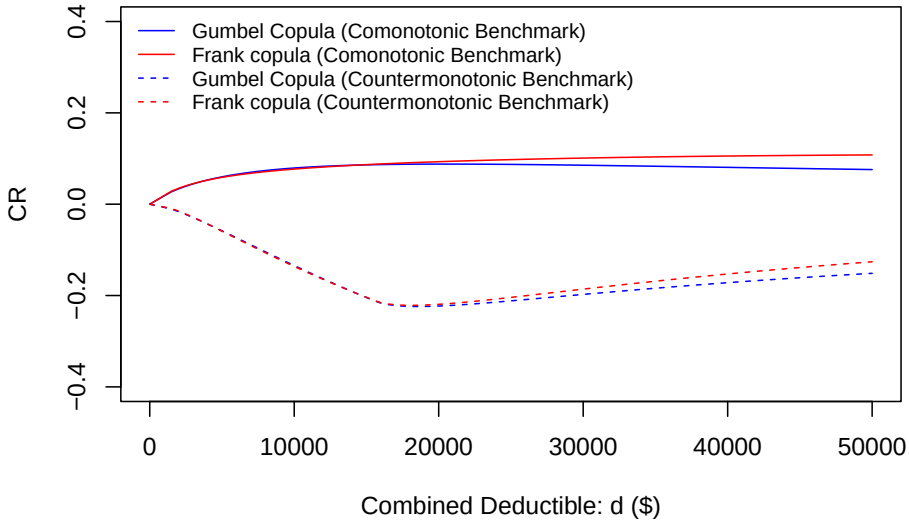

Figure A.6: Cost Relativities of Gumbel model and Frank model with comonotonic and countermonotonic benchmarks.

A.5.3 Application 3: Quotes of a Pet Insurance Contract (Personal)

We consider both the comonotonic scenario and countermonotonic scenario for the illness coverage and accident coverage. We pick Model 1 in the main body of Chap-

ter 2 to illustrate the cost relativity with comonotonic and countermonotonic benchmarks. In Figure A.7, in general, the deductible costs are increased by approximately 4 to 12 per cent when dependence models are used compared to the benchmark models that assume comonotonicity, and reduced by 16 to 33 per cent compared to the benchmark models that assume countermonotonicity. In conclusion, according to Proposition 2.6.2 in Section 2.6 of Chapter 2, countermonotonicity results in the largest deductible costs, and comonotonicity results in the smallest deductible costs.

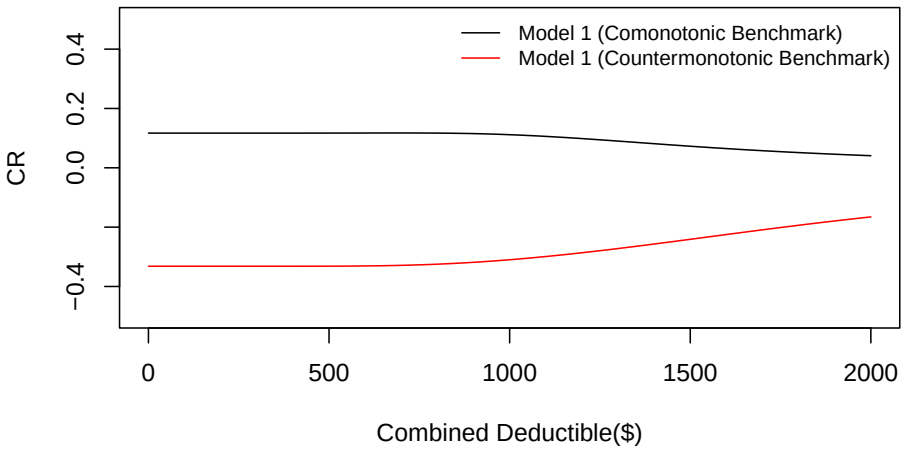

Figure A.7: Cost Relativities of model 1 with comonotonic and countermonotonic benchmarks.

A.6 Distribution Functions in Section 2.5

Table A.1 lists all four marginal distributions we use in Section 2.5 of Chapter 2. We use l and u to denote the lower bound and upper bound respectively for the uniform distribution. We use a and s to denote the shape parameter and scale parameter respectively for the gamma distribution, Pareto distribution, and Weibull distribution.

Table A.1: The distribution functions in Section 2.5.

Distribution	**Probability Density Function**	**E(X)**
uniform (l, u)	$f(x) = \frac{1}{l-u}$	$\frac{l+u}{2}$
gamma (a, s)	$f(x) = \frac{x^{(a-1)}e^{-\frac{x}{s}}}{s^a\Gamma(a)}$	as
Weibull (a, s)	$f(x) = \frac{a}{s}\left(\frac{x}{s}\right)^{a-1}\exp\left(-\left(\frac{x}{s}\right)^a\right)$	$s\Gamma\left(1+\frac{1}{a}\right)$
Pareto (a, s)	$f(x) = \frac{as^a}{(x+s)^{a+1}}$	$\frac{s}{a-1}$

A.7 Determination of the Number of Simulations

We use the central limit theorem to determine the number of simulations (R); see Section 6 in Frees (2018). In Section 2.5 of Chapter 2, we wish to be within 0.4% of the expected total losses with 95% certainty (0.4% of expected total loss is \$4). For $S = X_1 + X_2$, we want $\Pr\left(|\bar{S}_R - \mathrm{E}(S)| \leq 0.004\mathrm{E}(S)\right) \geq 0.95$, where $\bar{S}_R = \frac{1}{R}\sum_{r=1}^{R} S_r$.

According to the central limit theorem, our estimate should be approximately normally distributed, and so we want to have a large enough R to satisfy $0.004\mathrm{E}(S)/\sqrt{\mathrm{Var}(S)/R} \geq 1.96$ (1.96 is the 97.5th percentile from the standard normal distribution). Replacing $\mathrm{E}(S)$ and $\mathrm{Var}(S)$ with the estimates, we have

$$\frac{0.004\bar{S}_R}{s_{S,R}/\sqrt{R}} \geq 1.96,$$

where $s_{S,R}$ is the square root of the simulation variance,

$$s_{S,R} = \sqrt{\frac{1}{R-1}\sum_{r=1}^{R}(S_r - \bar{S}_R)}.$$

This criterion is a direct application of the approximate normality. Note that $\bar{S}_R$ and $s_{S,R}$ (estimates of $\mathrm{E}(S)$ and $\sigma(S)$) are not known in advance, so we will have to come up with the estimates, either by doing a small pilot study in advance or by interrupting our procedure intermittently to see if the criterion is satisfied.

A.8 Linear Approximation in Section 2.5.1.1

One of the primary objectives of risk theory is to model the distribution of total claim costs for portfolios of policies to make business decisions regarding various aspects of insurance contracts. However, the exact probability density distribution (PDF) of the aggregate loss is a weighted sum of convolutions, and the computation can be highly complex. To this end, simulations and approximation methods have been developed. Simulation methods can generate close to accurate results only if the number of simulations is large enough. In this section, we investigate linear approximation methods as an alternative. Data used in this section is generated from Section 2.5.1.1. of Chapter 2, and we focus on Marginals 3 (Pareto) and Marginals 4 (Weibull) with non-negative ρ_S (0, 0.4, 0.8) of the Gaussian copulas.

We derive the following three approximations of $E\ \min(X_1+X_2, d)$ using the Trapezoidal Rule Chiou (1978) based on equation (2.4) of Chapter 2. The approximation becomes more accurate as the resolution of the partition increases.

$$Approx_1 = d - \frac{d}{2}F_S(d),$$

$$Approx_2 = d - \frac{d}{2}F_S\left(\frac{d}{2}\right) - \frac{d}{4}F_S(d),$$

$$Approx_3 = d - \frac{d}{3}F_S\left(\frac{d}{3}\right) - \frac{d}{3}F_S\left(\frac{2d}{3}\right) - \frac{d}{6}F_S(d).$$

To calculate the cumulative distribution function $F_S(s)$ in $Approx_1$, $Approx_2$, $Approx_3$, we again use Monte Carlo simulations. We define an approximation-to-truth ratio

$$AT = \frac{Approx}{c_d}$$

to measure the approximation performance. The upper bound of d is set to be 2000 in order to see the approximation performance when d is relatively large. If an approximation method works well, AT would be close to 1. In the following analysis, we consider the scenarios of both known and unknown dependence structure of the two risks to the insurer. For the case of an unknown dependence structure, we assume the independence of the risks.

Figure A.8 shows the performance of the approximations if the insurer does not know the underlying ρ_S between X_1 and X_2, hence assuming independence of the two risks. The three plots on the top are based on Pareto marginals, and the bottom

plots are based on Weibull marginals. For each marginal, we consider three cases of the true dependence parameters, that is, $\rho_S \in (0, 0.4, 0.8)$. When the combined deductible d is relatively small, all three approximations work similarly and are close to the true values. However, as d increases beyond 1000, $Approx_2$ and $Approx_3$ show better and more stable performance than $Approx_1$.

We also notice that the approximations become worse as the true models move from independent ($\rho_S = 0$) to dependent ($\rho_S = 0.4$ and $\rho_S = 0.8$) cases. Moreover, all three approximations tend to overestimate the true deductible costs when the combined deductible d is bigger than 1000.

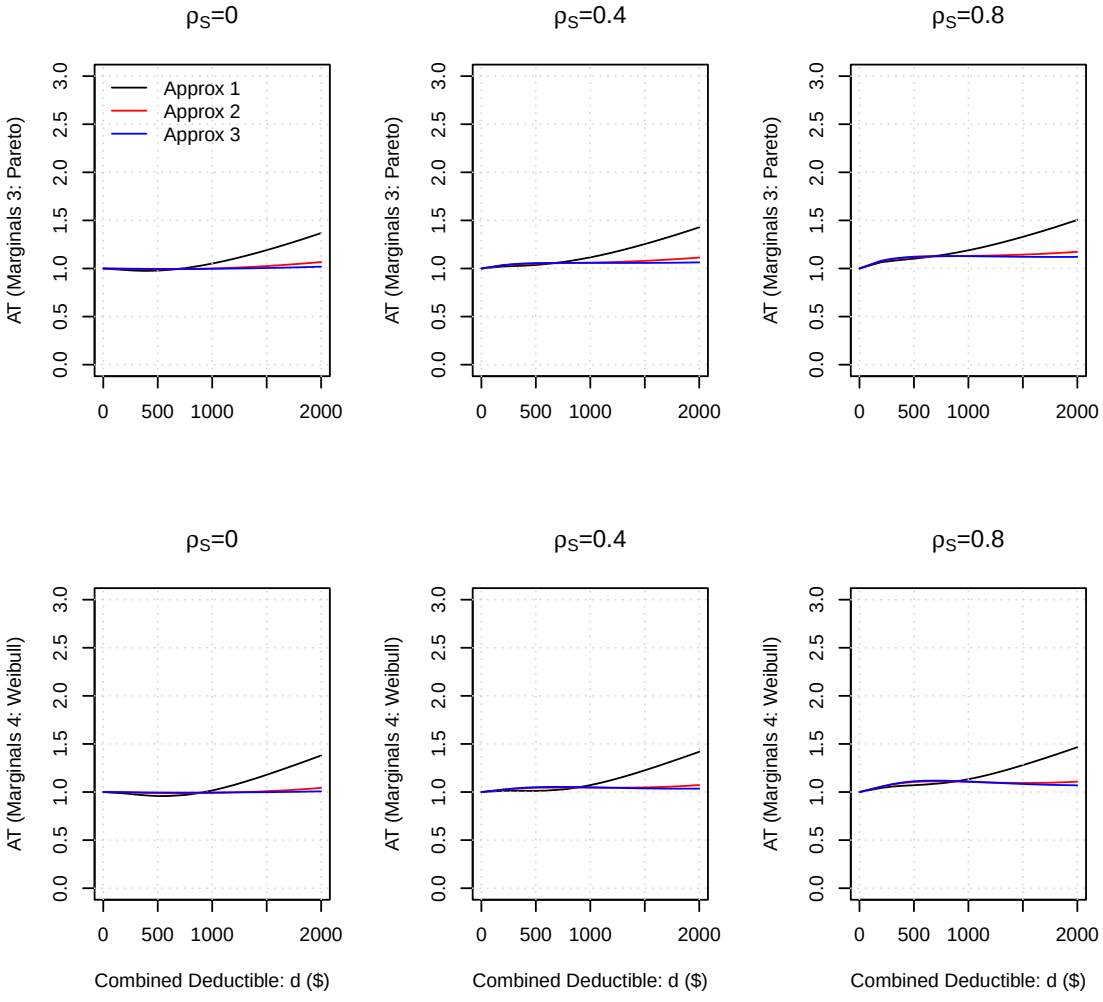

Figure A.8: The approximation-to-truth ratio (unknown dependence structure).

Figure A.9 shows the performance of approximations assuming the true dependence structure of the risks is known to the insurer. The left panel of Figure A.9 is the same as the left panel of Figure A.8, when $\rho_S = 0$. Compared to Figure A.8, the approximations work better when the insurer knows the dependence structure.

When d is large, $Approx_2$ and $Approx_3$ performed better and more stable compared to $Approx_1$.

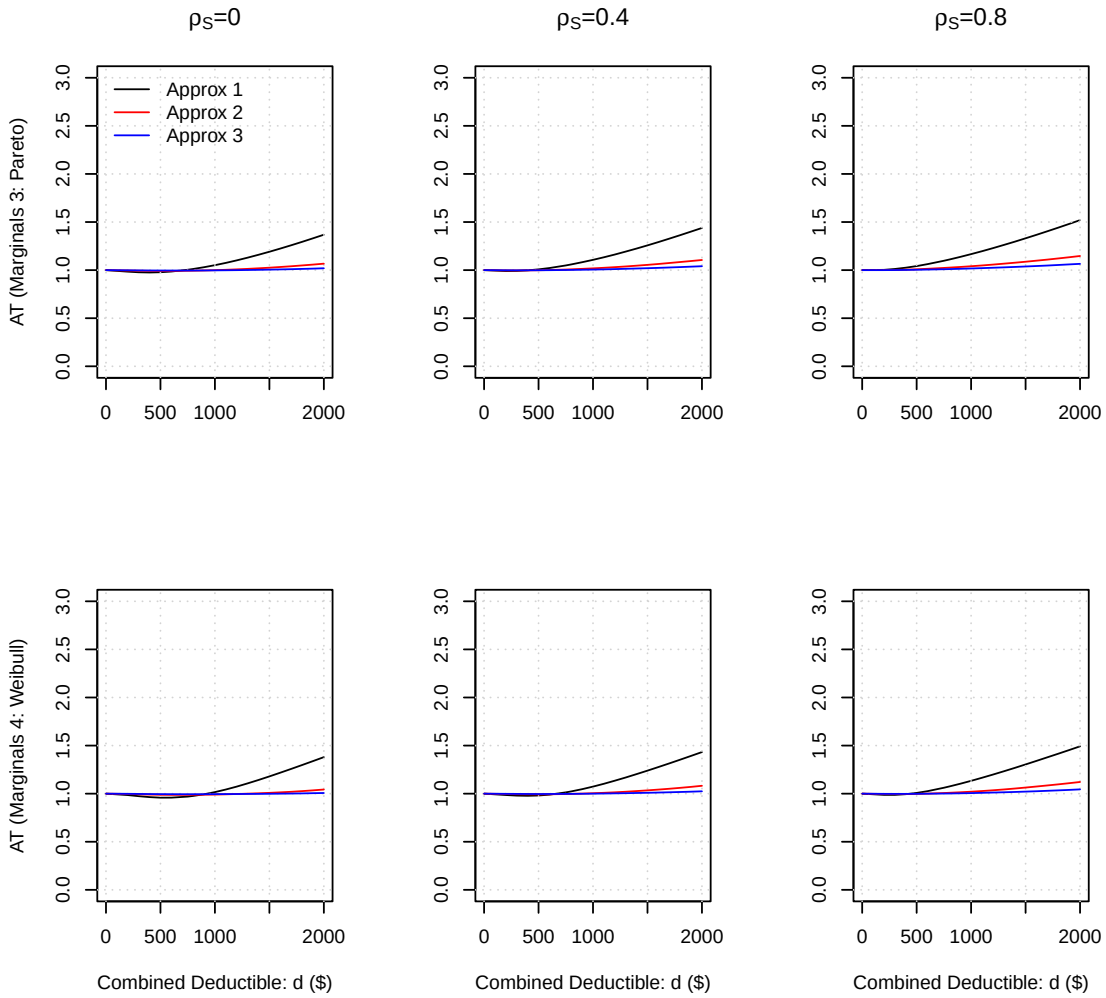

Figure A.9: The approximation-to-truth ratio (known dependence structure).

A.9 Proofs of Propositions in Section 2.6

The proof of Proposition 2.6.1 in Chapter 2 is as follows.

Proof. The random couple $\boldsymbol{X} = (X_1, X_2)$ is PQD. We need to prove $E\min(X_1 + X_2, d) \leq E\min(X_1^{\perp} + X_2^{\perp}, d)$, where $(X_1^{\perp}, X_2^{\perp})$ is an independent version of (X_1, X_2). When $d \geq 0$,

$$E\min(X_1 + X_2, d) = \int_0^d [1 - F_{X_1,X_2}(t, d-t)]\, dt,$$

$$E\min(X_1^{\perp} + X_2^{\perp}, d) = \int_0^d [1 - F_{X_1}(t)F_{X_2}(d-t)]\, dt.$$

Provided $\boldsymbol{X}$ is PQD, which ends the proof by

$$E\min(X_1+X_2,d) \quad -E\min(X_1^{\perp}+X_2^{\perp},d) = \int_0^d [F_{X_1}(t)F_{X_2}(d-t) - F_{X_1,X_2}(t,d-t)]\ dt \leq 0,$$

$$CR \quad = \frac{E\min(X_1+X_2,d)}{E\min(X_1^{\perp}+X_2^{\perp},d)} - 1 = \frac{c_d^{dep}}{c_d^{ind}} - 1 \leq 0.$$

□

The proof of Proposition 2.6.2. in Chapter 2 is as follows.

Proof. Two random couples $\boldsymbol{X} = (X_1, X_2)$ and $\boldsymbol{Y} = (Y_1, Y_2)$ satisfy $\boldsymbol{X} \preceq_{\text{CORR}} \boldsymbol{Y}$. When $d \geq 0$,

$$E\min(X_1+X_2,d) \quad = \int_0^d [1 - F_{\boldsymbol{X}}(t,d-t)]\ dt = \int_0^d \bar{F}_{\boldsymbol{X}}(t,d-t)\ dt,$$

$$E\min(Y_1+Y_2,d) \quad = \int_0^d [1 - F_{\boldsymbol{Y}}(t,d-t)]\ dt = \int_0^d \bar{F}_{\boldsymbol{Y}}(t,d-t)\ dt.$$

Provided $\boldsymbol{X} \preceq_{\text{CORR}} \boldsymbol{Y}$, which ends the proof by

$$E\min(X_1+X_2,d) - E\min(Y_1+Y_2,d) \quad = \int_0^d [F_{\boldsymbol{Y}}(t,d-t) - F_{\boldsymbol{X}}(t,d-t)]\ dt \geq 0,$$

$$CR_{\boldsymbol{Y}} \quad \leq CR_{\boldsymbol{X}}.$$

□

The proof of Proposition 2.6.3 in Chapter 2 is as follows.

Proof. If two CDFs never cross, then $E(A) \neq E(B)$ and it leads to a contradiction [Kaas et al. (2008)]. Assuming two CDFs cross exactly once at d_c ($F_B(d_c) = F_A(d_c)$). Then with Proposition 6.5. in Chapter 2 and (X_1, X_2) is PQD, it suffices to get $F_A(d) \leq F_B(d)$ when $0 \leq d \leq d_c$ and $F_A(d) > F_B(d)$ when $d_c < d$. If (X_1, X_2) is NQD, then $F_A(d) \geq F_B(d)$ when $0 \leq d \leq d_c$ and $F_A(d) < F_B(d)$ when $d_c < d$.

$$\frac{\partial}{\partial d}\big(E\min(A,d) - E\min(B,d)\big) = F_B(d) - F_A(d).$$

Now we summarise the behaviour of the difference in pricing $|E\min(A,d) - E\min(B,d)|$ with respect to the combined deductible d. When d increases from 0 to d_c, $|E\min(A,d) - E\min(B,d)|$ will increase. When two CDFs cross at d_c, $|E\min(A,d) - E\min(B,d)|$ is maximised. When d increases from d_c to infinity, $|E\min(A,d) - E\min(B,d)|$ will decrease. We find that the maximum value

$|E\min(A, d_c) - E\min(B, d_c)|$ equals to the stop-loss distance introduced in Section 9.8 of Denuit et al. (2006). □

The proof of Proposition 2.6.4 in Chapter 2 is as follows.

Proof. If the random couple $\boldsymbol{X} = (X_1, X_2)$ is PQD, $E\min(X_1^{\perp} + X_2^{\perp}, d) \geq E\min(X_1 + X_2, d)$ for all $d \in [0, \infty)$ by Proposition 2.6.3 in Chapter 2. Again assume $A = X_1^{\perp} + X_2^{\perp}, B = X_1 + X_2$. First, we prove that for any random variable X, $\int_0^{\infty}[\min(t, E(X)) - E\min(X, t)]\ dt = \frac{1}{2}\sigma_X^2$.

$$\begin{aligned}
&\int_0^{\infty}[\min(t, E(X)) - E\min(X, t)]\ dt \\
&= \int_0^{E(X)}[t - E\min(X, t)]\ dt + \int_{E(X)}^{\infty}[E(X) - E\min(X, t)]\ dt \\
&= \int_0^{E(X)} E(t - X)_+\ dt + \int_{E(X)}^{\infty} E(X - t)_+\ dt \\
&= \int_0^{E(X)} \int_0^t F_X(x)\ dx\ dt + \int_{E(X)}^{\infty} \int_t^{\infty}[1 - F_X(x)]\ dx\ dt.
\end{aligned}$$

Interchanging the order of the integrations and using partial integration leads to

$$\begin{aligned}
&\int_0^{E(X)} \int_0^t F_X(x)\ dx\ dt = \int_0^{E(X)} \big(E(X) - x\big)F_X(x)\ dx \\
&= \frac{1}{2}E(X)^2 F_X\big(E(X)\big) - \int_0^{E(X)}[E(X)x - \frac{x^2}{2}]\ dF_X(x) \\
&= \frac{1}{2}\int_0^{E(X)} \big(x - E(X)\big)^2\ dF_X(x),
\end{aligned}$$

$$\begin{aligned}
&\int_{E(X)}^{\infty} \int_t^{\infty}[1 - F_X(x)]\ dx\ dt = \int_{E(X)}^{\infty} \big(E(X) - x\big)\big(F_X(x) - 1\big)\ dx \\
&= \frac{1}{2}E(X)^2\big(F_X(E(X)) - 1\big) - \int_{E(X)}^{\infty}[E(X)x - \frac{x^2}{2}]\ d[F_X(x) - 1] \\
&= \frac{1}{2}\int_{E(X)}^{\infty} \big(x - E(X)\big)^2\ dF_X(x),
\end{aligned}$$

$$\begin{aligned}
&\int_0^{\infty}[\min(t, E(X)) - E\min(X, t)]\ dt = \frac{1}{2}\int_0^{\infty} \big(x - E(X)\big)^2\ dF_X(x) \\
&= \frac{1}{2}\sigma_X^2.
\end{aligned}$$

So we now show the following result:

$$\begin{aligned}
IDIP &= \int_0^\infty [E\min(A,t) - E\min(B,t)]\ dt \\
&= \int_0^\infty [\min(t,E(B)) - E\min(B,t) - \min(t,E(A)) + E\min(A,t)]\ dt \\
&= \int_0^\infty [\min(t,E(B)) - E\min(B,t)]\ dt - \int_0^\infty [\min(t,E(A)) - E\min(A,t)]\ dt \\
&= \frac{1}{2}(\sigma_B^2 - \sigma_A^2) \\
&= \frac{1}{2}\big(E(B^2) - E(A^2)\big) \\
&= \frac{1}{2}\big[E(X_1^2 + X_2^2 + 2X_1X_2) - E\big((X_1^\perp)^2 + (X_2^\perp)^2 + 2X_1^\perp X_2^\perp\big)\big] \\
&= E(X_1X_2) - E(X_1)E(X_2) \\
&= \sigma(X_1, X_2).
\end{aligned}$$

If the random couple $\boldsymbol{X} = (X_1, X_2)$ is NQD, it is easy to get $IDIP = -\sigma(X_1, X_2)$.

□

A.10 Copula Table

Table A.2: Copulas PQD/NQD Conditions

Copula	$C(u,v) =$	**PQD Conditions**	**NQD Conditions**
Gaussian	$\Phi_\theta\Big(\Phi^{-1}(u), \Phi^{-1}(v)\Big)$, $\theta \in [-1,1]$.	$0 < \theta \leq 1$, $0 < \tau \leq 1$, $0 < \rho_S \leq 1$.	$-1 \leq \theta < 0$, $-1 \leq \tau < 0$, $-1 \leq \rho_S < 0$.
T-student (v)	$H_{v,\theta}\Big(F_v^{-1}(u), G_v^{-1}(v)\Big)$, $\theta \in [-1,1],\ v \geq 2$.	$0 < \theta \leq 1$, $0 < \tau \leq 1$, $0 < \rho_S \leq 1$.	$-1 \leq \theta < 0$, $-1 \leq \tau < 0$, $-1 \leq \rho_S < 0$.
Frank	$-\frac{1}{\theta}\Big(1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\Big)$, $\theta \in \mathbb{R} \setminus 0$.	$0 < \theta$, $0 < \tau \leq 1$, $0 < \rho_S \leq 1$.	$\theta < 0$, $-1 \leq \tau < 0$, $-1 \leq \rho_S < 0$.
Gumbel	$\exp\Big(-\big((-ln(u))^\theta + (-ln(v))^\theta\big)^{\frac{1}{\theta}}\Big)$, $\theta \in [1,\infty)$.	$1 < \theta$, $0 < \tau \leq 1$.	$\varnothing$.
Clayton	$\max(u^{-\theta} + v^{-\theta} - 1, 0)^{\frac{-1}{\theta}}$, $\theta \in [-1,\infty) \setminus 0$.	$0 < \theta$, $0 < \tau \leq 1$.	$-1 \leq \theta < 0$, $-1 \leq \tau < 0$.
Ali-Mikhail-Haq	$\frac{uv}{1-\theta(1-u)(1-v)}$, $\theta \in [-1,1]$.	$0 < \theta \leq 1$, $0 < \tau \leq 1/3$.	$-1 \leq \theta < 0$, $\frac{5-8ln(2)}{3} \leq \tau < 0$.
Joe	$1-\big((1-v)^\theta + (1-u)^\theta - (1-v)^\theta(1-u)^\theta\big)^{\frac{1}{\theta}}$, $\theta \in [1,\infty)$.	$1 < \theta$, $0 < \tau \leq 1$.	$\varnothing$.
Farlie-Gumbel-Morgenstern	$uv + \theta uv(1-u)(1-v)$, $\theta \in [-1,1]$.	$0 < \theta \leq 1$, $0 < \tau \leq 2/9$, $0 < \rho_S \leq 1/3$.	$-1 \leq \theta < 0$, $-9/2 \leq \tau < 0$, $-1/3 \leq \rho_S < 0$.

Appendix for Chapter 3

B.1 Five-State Transition Framework and Data Imbalance

Figure B.1 shows the original five-state transition diagram of the application to data from LGPIF.

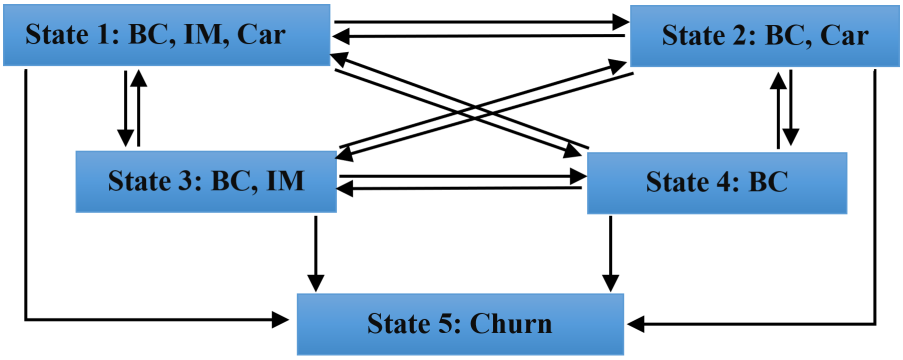

Figure B.1: A five-state transition diagram of the application to data from LGPIF. BC refers to building and content insurance, IM refers to contractor's equipment insurance, while Car refers to vehicle coverage.

Table B.1 lists the five-state transition counts and empirical transition probabilities from 2006 to 2013. We can see the numbers of several types of transition are very small. For example, the number of transitions from state 3 to state 2 is 0, which means that this type of transition has not occurred from 2006 to 2013 among all policyholders. The transitions in the minority are statistically rare events. Consequently, when a multi-state model is estimated on a random sample of the customer population, the majority transitions will dominate the statistical analysis, which may decrease the predictive accuracy of the minority transitions. This is a typical issue of data imbalance, and has been addressed through "balanced sampling" in the literature on traditional customer churn probability prediction; see Lemmens and Croux (2006). However, in Chapter 3, we focus on using the second-order MLR

model to study multi-state customer churn analysis rather than achieving a high degree of predictive accuracy, so we merged states 2, 3, and 4 in Figure B.1 into one state in Chapter 3 to ensure that there is sufficient data to model all transition probabilities reasonably well.

Table B.1: Five-state transition counts and empirical transition probabilities in percent (in parentheses) from 2006 to 2013.

	State of destination				
State of origin	State 1	State 2	State 3	State 4	Churn
State 1	2552 (94.14%)	4 (0.15%)	74 (2.73%)	3 (0.11%)	78 (2.88%)
State 2	11 (4.40%)	217 (87.80%)	1 (0.40%)	9 (3.60%)	12 (4.80%)
State 3	34 (0.94%)	0 (0.00%)	3479 (95.87%)	5 (0.14%)	111 (3.06%)
State 4	3 (0.25%)	7 (0.59%)	29 (2.45%)	1112 (94.08%)	31 (2.62%)

B.2 Explanatory Variables in Customer Churn Analysis in General Insurance

Premium information is found to be the key driver of customer churn, with the majority of literature finding that policyholders are more likely to churn after they experience the premium increase (Brockett et al. 2008; Haugen and Moger 2016; Jeong et al. 2018; De la Llave et al. 2019; Leiria et al. 2021). Premium information is also a frequently used predictor when applying machine learning techniques to predict customer churn rates (see for instance Bolancé et al. 2016; Paredes 2018; Scriney et al. 2020).

The claim experience is a special feature in general insurance. In life insurance, when a policyholder makes a claim, it often means that the policyholder has died and the life insurance contract is terminated. However, in general insurance, it is common to see a policyholder make several claims in a single period, in which case customers will normally face a higher premium for the next period after they had made one or more claims in the previous period. Conversely, customers may receive a bonus that reduces the premium for the next period if they did not make any claim in the previous period. Jeong et al. (2018) and Frees et al. (2021) provide evidence of a strong association between claim occurrence and customer churn. Furthermore, numerous studies have shown that making a claim increases the probability of customer churn and reduces the overall lifetime of a contract (see for example Guillen et al. 2003; Brockett et al. 2008; Guillen et al. 2009; Haugen and Moger 2016).

Finally, contract information refers to a broad range of information that can be

found in signed contracts between policyholders and insurers, including the geographic location of the insured object (e.g., Haugen and Moger 2016; Paredes 2018; Staudt and Wagner 2018; De la Llave et al. 2019; Frees et al. 2021), policyholder characteristics (e.g., Paredes 2018; Staudt and Wagner 2018; Frees et al. 2021), and other useful information such as the date of renewal.

B.3 Endogeneity and Exogeneity

In economic models, variables are commonly divided into two main categories: endogenous variables, i.e., variables that a model tries to explain, and exogenous variables, i.e., variables that a model takes as given (Mankiw 2003). Identifying whether a variable is exogenous versus endogenous poses a challenge for insurers. There are two main forms of exogeneity, depending on the level of independence shown by the variable. A strictly exogenous variable is completely unaffected by the output of a model in the past, present, and future. A sequentially exogenous (also called predetermined) variable is not affected by past instances of the model's output, but future instances may be affected by current or future instances of the model's output. If a variable is neither strictly exogenous nor sequentially exogenous, it is called endogenous. In the context of multi-state customer churn analysis, the endogenous variable is the transition among different states of a contract held by a policyholder, i.e., the values of $l_{i,t}$ in Chapter 3. In order to decide if a new variable is strictly exogenous or sequentially exogenous, we have to decide if the current or future customers' transition decision would cause the new variable to change in the future.

As we have seen in the existing literature, premium information is known to be statistically related to customer churn, e.g., the higher the premium, the higher the churn rate, and there is a causal relationship among the claim, the premium, and the probability of customer churn. To determine whether claims occurrence and frequency and premiums are strictly exogenous or sequentially exogenous in a multi-state transition model, we need to consider the moral hazard and adverse selection under the insurance context. Adverse selection is the tendency of policyholders in high-risk positions to purchase and renew insurance contracts. Moral hazard occurs if policyholders act in a more risky way after they enter insurance contracts. In the absence of moral hazard and adverse selection, a premium can be simply regarded as the insurance company actuary's summary measure of several risk factors with realistic considerations, and a claim is only affected by strictly exogenous factors such as climate and economic change. In this case, a customer's current or future

purchasing decision will not affect his or her future premiums and claims, so both claims and premiums are strictly exogenous. On the other hand, if we take into account moral hazard and adverse selection, then policyholders who keep purchasing insurance contracts are more likely to make claims in the future, and high insurance claims in one year lead to high premiums in subsequent years. From this perspective, both premiums and claims evolve over time, and so both of them should be regarded as sequentially exogenous. For the purposes of Chapter 3, we treat all variables as strictly exogenous, and we do so to be consistent with the existing literature on traditional customer churn analysis (see for instance Guillen et al. 2003; Brockett et al. 2008, among others). We also refer the reader to Pinquet (2000) for a discussion of exogeneity in an insurance rating context, and Chapter 6 of Frees (2004) for an overview of exogeneity in longitudinal models.

B.4 Robust Standard Errors of the Second-Order MLR Model

We provide robust standard errors using Stata 17; see Table B.2. Generally, the formula for the robust estimator of variance is given in Stata (2022) as

$$\hat{V} = \hat{\mathrm{V}} \left(\sum_{j=1}^{M} \mathrm{u}_{\boldsymbol{j}}' \mathrm{u}_{\boldsymbol{j}} \right) \hat{\mathrm{V}}$$

where M is the total number of observations, $\hat{\mathrm{V}} = (-\partial^2 \ln L / \partial \boldsymbol{\beta}^2)^{-1}$ (the conventional estimator of variance), and $\mathrm{u}_{\boldsymbol{j}}$ (a row vector) is the contribution from the j-th observation to $\partial \ln L / \partial \boldsymbol{\beta}$.

For the second-order MLR model, we need to select a reference state and set the coefficients corresponding to this reference state to zero for reasons of parameter identifiability e.g., in Chapter 3, $s = 1$ and $\beta_{qr1} = 0$ for all (q, r). For the transition with state of origin (q, r) and state of destination s, we define

$$\eta_{qrs} = \begin{cases} \frac{\exp\left(\boldsymbol{x}_{qr}^{\top} \boldsymbol{\beta}_{qrs}\right)}{1 + \sum_{s'=2}^{Q} \exp\left(\boldsymbol{x}_{qr}^{\top} \boldsymbol{\beta}_{qrs'}\right)}, & 1 < s \leq Q, \\ \frac{1}{1 + \sum_{s'=2}^{Q} \exp\left(\boldsymbol{x}_{qr}^{\top} \boldsymbol{\beta}_{qrs'}\right)}, & s = 1. \end{cases}$$

To calculate the score vector and Hessian matrix later, we will use the fact that for

all $1 < l, s \leq Q$

$$\frac{\partial}{\partial \beta_{qrs}} \eta_{qrl} = \eta_{qrl}[\mathbf{1}(s = l) - \eta_{qrs}]\boldsymbol{x}_{qr}.$$

We write the score $= (\mathrm{sc}_{1,1,2}, \ldots, \mathrm{sc}_{1,1,Q}, \mathrm{sc}_{1,2,2}, \ldots, \mathrm{sc}_{1,2,Q}, \ldots, \mathrm{sc}_{Q-1,Q-1,2}, \ldots, \mathrm{sc}_{Q-1,Q-1,Q})$. For example, in the application to data from LGPIF in Section 3.4 of Chapter 3, score $= (\mathrm{sc}_{1,1,2}, \mathrm{sc}_{1,1,3}, \ldots, \mathrm{sc}_{2,2,3})$. Any $\mathrm{sc}_{q,r,s}$ can be calculated as

$$\begin{aligned} \mathrm{sc}_{q,r,s} &= \frac{\partial}{\partial \boldsymbol{\beta}_{\boldsymbol{qrs}}} \ln L \\ &= \sum_{j=1}^{M_{qr}} [\mathbf{1}(l_j = s) - \eta_{j,qrs}]\boldsymbol{x}_{j,qr}, \end{aligned}$$

where M_{qr} is the observed number of transitions with states of origin (q, r), l_j is the state of destination for the j-th transition among M_{qr} transitions, and $\boldsymbol{x}_{j,qr}$ is the vector of corresponding covariates used in the second-order MLR model.

The Hessian matrix is a block diagonal matrix, with (r, s)-th block given by

$$\begin{aligned} \mathrm{bl}_{rs} &= \frac{\partial^2}{\partial \boldsymbol{\beta}_{qrs} \partial \boldsymbol{\beta}_{qrl}} \ln L \\ &= -\sum_{j=1}^{M_{qr}} \eta_{j,qrs}[\mathbf{1}(l_j = l) - \eta_{j,qrl}]\boldsymbol{x}_{j,qr}\boldsymbol{x}_{j,qr}^{\top}. \end{aligned}$$

If we compare the coefficients in Table B.2 below with those in Table 3.3 of Chapter 3, we can see the coefficients are slightly different in several cases, which is due to the use of different softwares (R in Chapter 3 and Stata in the supplementary material).

B.5 Goodness-of-Fit Tests

Due to the fact that multinomial models have more than one group of coefficients, assessing Goodness-of-Fit (GoF) is more challenging, and is still an area of intense research. The most approachable method to assess model confidence is the Hosmer-Lemeshow (HL) test (Hosmer and Lemesbow 1980), which is extended in Fagerland et al. (2008) for MLR models. In a MLR model setting, we have data of the form $(x_1, l_1), (x_2, l_2), \ldots, (x_n, l_n)$, where l_i is a k-vector that indicates which class the i-th observation belongs to (exactly one entry contains a one and the rest are zero), x_i

Table B.2: Summary of the second-order MLR model with robust standard errors (in parentheses.)

State of destination	to partial-coverage ($s = 2$)	to churn ($s = 3$)
State of origin: (1,1)		
Intercept	-19.76^{***} (1.88)	-19.88^{***} (1.36)
Entity(ScToVi)	16.25^{***} (0.34)	15.36^{***} (0.25)
Entity(CiMi)	17.18^{***} (0.39)	16.25^{***} (0.53)
RateBC	-2.86^{***} (1.03)	0.30 (0.60)
RateIM	1.37^{*} (0.79)	0.76 (0.61)
RateCar	-0.39^{***} (0.13)	-0.04 (0.13)
I(ClaimBC)	0.20 (0.34)	-0.64^{*} (0.38)
I(ClaimIM)	-0.47 (1.02)	-14.66^{***}(0.31)
TotalCoverage	-0.88^{**} (0.31)	-0.66^{*}(0.51)
RatioPremium	0.81 (1.07)	0.93^{*}(0.51)
State of origin: (2,1)		
Intercept	-18.02^{***} (0.58)	
Entity(ScToVi)	16.34^{***} (0.76)	
I(ClaimCar)	-16.63^{***} (0.67)	
State of origin: (1,2)		
Intercept	16.68^{***} (0.90)	16.65^{***} (1.12)
Entity(ScToVi)	-17.06^{***} (1.01)	-16.82^{***} (0.95)
RateIM	1.19^{*} (0.77)	-0.07 (0.66)
State of origin: (2,2)		
Intercept	19.75^{***} (0.18)	16.29^{***} (0.33)
Entity(CiScToVi)	-14.75^{***} (0.28)	-14.63^{***} (0.37)
FrequencyBC	0.02 (0.09)	-0.72^{***} (0.22)
RatioPremium	3.25^{**} (1.39)	3.56^{**} (1.40)
I(FinaCris)	-0.90^{**} (0.41)	-0.42 (0.46)
Note:		*p<0.1; **p<0.05; ***p<0.01

is the vector of explanatory variables for the i-th observation, and n is the number of observations. After we fit a MLR model that estimates a vector of predicted probabilities for k classes $\hat{p}(x_i) = \big(\hat{p}_1(x_i), \hat{p}_2(x_i), \ldots, \hat{p}_k(x_i)\big)$, the predicted values ($1 - \hat{p}_1(x_i)$ assuming class 1 is the reference level) are arrayed from the lowest to the highest, and then separated into several groups of approximately equal size. Let O_{hj} and E_{hj} denote the sum of the observed frequencies and estimated frequencies for j-th class in h-th group. The HL statistic is calculated as

$$C_g = \sum_{h=1}^{g} \sum_{j=1}^{k} \frac{(O_{hj} - E_{hj})^2}{E_{hj}},$$

where g is the number of groups (normally is set to be 10). Under the null hypothesis that the fitted model is the correct model and the sample is sufficiently large, we expect C_g to have an approximate χ^2 distribution with $(g-2) \times (k-1)$ degrees of freedom.

Besides the HL test, the deviance and Pearson chi-square tests for the binary logistic regression can also be extended to MLR models. The deviance can be written

as

$$D = -2\sum_{i=1}^{n}\sum_{j=1}^{k} l_{ij}\log\left(\frac{\hat{p}_j(x_i)}{l_{ij}}\right),$$

where l_{ij} is an indicator for j-th class in vector l_i, and we treat zero times the log of anything as zero. Under the null hypothesis that the fitted model is the correct model, the distribution of D is chi-squared with $(n-p)\times(k-1)$ degrees of freedom, where p is the number of explanatory variables including intercept.

The Pearson chi-square is calculated as

$$\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{k} \frac{\left(l_{ij} - \hat{p}_j(x_i)\right)^2}{\hat{p}_j(x_i)}.$$

Under the null hypothesis that the fitted model is the correct model, the distribution of χ^2 is chi-squared with $(n-p)\times(k-1)$ degrees of freedom.

In Table B.3, we show the results of the three tests for the second-order MLR model in the application to data from LGPIF. In all cases, p-values are high, so there is no evidence showing that the second-order MLR model is a poor fit.

Table B.3: Goodness-of-fitting tests for the second-order MLR model.

	Pearson Chi-Square			Deviance			HL (10 groups)		
	Statistic	df	p-value	Statistic	df	p-value	Statistic	df	p-value
Second-order MLR	9258.0	10846	1	2191.6	10846	1	15.5	16	0.49

B.6 Exploratory Variables of an "Average" Policyholder

The values of exploratory variables of an average policyholder are listed in Table B.4.

B.7 One-versus-All Binary Logistic Regression

It is worth comparing MLR models to another, an obvious approach that could be used for multi-state customer churn analysis, namely fitting multiple sets of binary logistic regression (BLR) models based on either a one-versus-all (also known as one-versus-rest) or one-versus-one approach. Recall that in traditional customer churn

Table B.4: The state-specific values of exploratory variables of an "average" policyholder.

Exploratory variable	**Full-coverage**	**Partial-coverage**
Entity(ScToVi)	1	1
Entity(CiMi)	0	0
RateBC	0.537	0.652
RateIM	2.214	1.660
RateCar	3.600	0.169
I(ClaimBC)	0	0
I(ClaimIM)	0	0
I(ClaimCar)	0	0
TotalCoverage	0.666	0.285
RatioPremium	0.034	0.018

analysis, BLR models are fitted to a binary response for identifying the explanatory variables driving the probability of customer churn. With multi-state data, one can extend this approach by fitting separate BLR models for each transition, including the transition to churn.

We illustrate this with a one-versus-all, second-order BLR model approach as follows. Consider the example discussed in the right panel of Figure 3.2 of Chapter 3, where there are three possible states a policyholder can transition to at time t (full-coverage or state 1, partial-coverage or state 2, churn or state 3). Then for each state of origin (q, r) where $q = \{1, 2\}$, $r = \{1, 2\}$, a one-vs-all, second-order BLR model would construct the following three separate logistic regression models:

$$\begin{aligned} P(l_{i,t} = 1 | l_{i,t-2} = q, l_{i,t-1} = r) &= \text{logit}^{-1}\left(\boldsymbol{x}_{i,t-1,qr1}^{\top}\boldsymbol{\beta}_{qr1}\right), \\ P(l_{i,t} = 2 | l_{i,t-2} = q, l_{i,t-1} = r) &= \text{logit}^{-1}\left(\boldsymbol{x}_{i,t-1,qr2}^{\top}\boldsymbol{\beta}_{qr2}\right), \\ P(l_{i,t} = 3 | l_{i,t-2} = q, l_{i,t-1} = r) &= \text{logit}^{-1}\left(\boldsymbol{x}_{i,t-1,qr3}^{\top}\boldsymbol{\beta}_{qr3}\right), \end{aligned} \tag{B.1}$$

where $\text{logit}^{-1}(x) = \exp(x)/(1 + \exp(x))$ denotes the inverse link function. It is important to emphasise that in equation (B.1), three sets of BLR models are fitted for each state of origin. This contrasts with the MLR model in equation (3.2) of Chapter 3, i.e., for each state of origin, only a single model needs to be fitted for all transition probabilities. Indeed, this reflects a fundamental difference between a MLR modelling approach and a one-versus-all BLR modelling approach for multi-state analysis: in the latter, transition probabilities from the same state of origin are independently modelled, and thus need not and in general do not sum to one. That is, in equation (B.1) there is no guarantee that $P(l_{i,t} = 1|l_{i,t-2} = q, l_{i,t-1} = r) + P(l_{i,t} = 2|l_{i,t-2} = q, l_{i,t-1} = r) + P(l_{i,t} = 3|l_{i,t-2} = q, l_{i,t-1} = r) = 1$. One consequence of this is that the transition probabilities from BLR models do not properly reflect

the conditional or marginal effects of covariates in multi-state customer analysis, and we discuss this point further in Section 3.4.3 of Chapter 3. Also, it is clear that, conditional on a state of origin (q, r), the coefficients in a one-versus-all BLR model only affect that particular corresponding transition probability. In contrast, for the MLR modelling approach in equation (3.2) of Chapter 3, the coefficients $\boldsymbol{\beta}_{qrs}$ affect all transition probabilities, as they occur in both the numerator and the denominator of the regression function. This is again nothing more than a direct consequence of the sum to one constraint.

In Chapter 3, we choose to use MLR models instead of a set of separate BLR models for the purposes of multi-state customer churn analysis, as the former naturally considers the dependence among the transition probabilities. Indeed, MLR models precisely quantify the effect of covariates on transitions to a multitude of states, and thus offer a more complete picture of dynamic changes in the needs of policyholders over time. That being said, it is important to acknowledge that if the primary question of interest is related to studying *specific* transitions instead of all transitions jointly, then the one-versus-all BLR modelling approach is suited for this. For example, if the primary focus was studying the state of destination as churn, then building a BLR model based on the third line of equation (B.1) would be a more direct way of answering this question since it explicitly and only models the probability of customer churn versus not churn. However, it is also possible to use the MLR model to answer the same question, albeit with more mathematical calculations.

B.8 Scenarios by Conducting Traditional Customer Churn Analysis and Multi-state Customer Churn Analysis

We visualise the scenarios of future paths for traditional and multi-state customer churn analysis in Figure B.2.

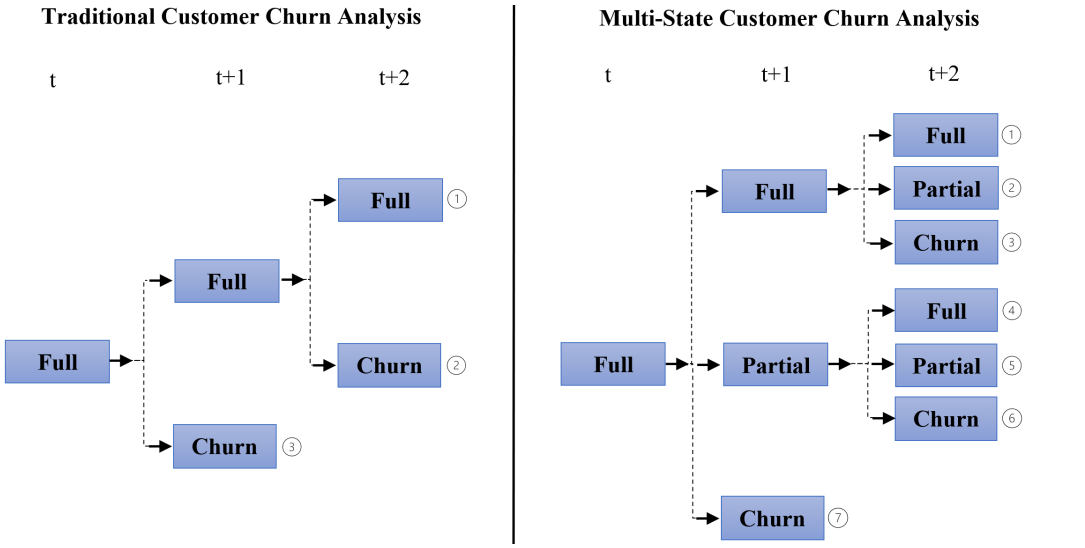

Figure B.2: Scenarios of the future path for CLV calculation: three scenarios for traditional customer churn analysis (left panel) versus seven scenarios for multi-state customer churn analysis (right panel).

B.9 Models in Out-of-Sample Validation

B.9.1 Support Vector Machine

Support vector machine (SVM) is one of the most popular supervised learning algorithms, which is used for classification as well as regression problems (Noble 2006). However, primarily, it is used for classification problems in machine learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that people can easily put the new data points in the correct category in the future. However, one disadvantage of SVM is that it does not output probabilities natively, but probability calibration methods can be used to convert the output to class probabilities. Various methods exist, including Platt scaling (particularly suitable for SVMs) and isotonic regression. In Chapter 3, we use the default probabilities provided in `svm` function in `e1071` package (Meyer et al. 2021). The default probability model for SVM classification fits a logistic distribution using maximum likelihood to the decision values of all binary classifiers, and computes the a-posteriori class probabilities for the multi-class problem using quadratic optimisation.

The SVM is an approximate implementation of a theoretical bound on the generalisation performance that is independent of the dimensionality of the feature space. To build a non-linear SVM classifier, we can use either the polynomial kernel or the radial kernel function. In Chapter 3, we use the radial kernel function. There are

two hyperparameters to be tuned when we use the radial kernel function. `C`, the "cost" of the radial kernel, controls the complexity of the boundary between support vectors. The radial kernel also requires setting a smoothing parameter, `sigma`. The `caret` (Kuhn 2021) package can be used to tune the radial hyperparameters of SVM radial kernel function models by setting `method = ‘‘svmRadial’’`. We use 10-fold cross-validation to find the optimal hyperparameters that can maximise accuracy in the original training set and balanced training set. Accuracy is the ratio of the number of correctly classified instances to the total number of instances. The result of optimal hyperparameters is summarised in Table B.5.

Table B.5: Tuning results of SVM classifiers.

	Original training set	**Balanced training set**
C	1	4096
sigma	0.0326	0.0337

B.9.2 Gradient Boosting Machine

Gradient boosting machine (GBM) combines the predictions from multiple decision trees (weak learners) to generate the final predictions (Friedman 2001). Additionally, each new tree takes into account the errors or mistakes made by the previous trees. So, every successive decision tree is built on the errors of the previous trees. This is how the trees in a gradient-boosting machine algorithm are built sequentially.

The hyperparameters of a GBM model to be tuned are `interaction.depth`, `n.trees`, `shrinkage`, and `n.minobsinnode`. The definition of these hyperparameters can be found in gbm (Greenwell et al. 2020) package. `interaction.depth` specifies the maximum depth of each tree (i.e., the highest level of variable interactions allowed). `n.trees` specifies the total number of trees to fit. `shrinkage` is a shrinkage parameter applied to each tree in the expansion (also known as the learning rate or step-size reduction). `n.minobsinnode` specifies the minimum number of observations in the terminal nodes of the trees. GBM model is tuned by 5-fold CV to minimise logLoss using `caret` (Kuhn 2021) package by setting `distribution="multinomial"`, `method="gbm"`, and `metric="logLoss"`. `logLoss` computes the minus log-likelihood of the multinomial distribution (without the constant term):

$$-\text{logLoss} = \frac{-1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} l_{ij}\log(p_{ij}),$$

where the y values are binary indicators for the classes and p are the predicted class probabilities, n is number of observations, and k is number of classes.

The result of optimal hyperparameters in the original training set and balanced training set is summarised in Table B.6.

Table B.6: Tuning results of GBMs.

	Original training set	**Balanced training set**
interaction.depth	3	2
n.trees	400	5000
shrinkage	0.01	0.01
n.minobsinnode	10	10

Appendix for Chapter 4

C.1 Estimation Results for Shared Random Effects

Figure C.1 shows the histogram of estimated random effects of all customers. Note that we assume the shared random effects follow a multivariate normal distribution, but the estimated shared random effects do not strictly satisfy the normality assumption. Future research could examine the sensitivity of this assumption as well as the use of non-normal multivariate distributions for the shared random effects, such as a log-gamma distribution.

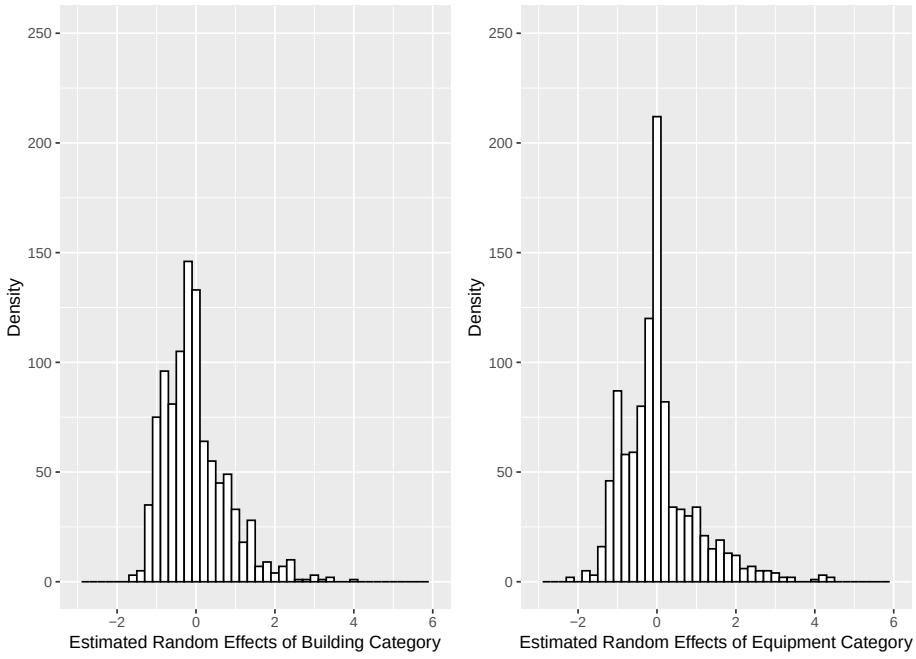

Figure C.1: Estimated shared random effects of all customers.

C.2 Estimation Results for Alternative Models

Since the estimation results of the customer churn sub-model are included in the main body of the paper, this section only includes the estimation results of the claims sub-model for all alternative models. Note feature selection is applied to certain models to avoid the model convergence problem.

Table C.1: Parameter estimates for the claims severity sub-model in models A2, A3, A4 and A5.

Building claims severity			**Equipment claims severity**		
Models A2 and A3 severity: constant severity					
\$18767.2			\$5873.4		
	Estimate	S.E.		Estimate	S.E.
Models A4 and A5 severity: gamma severity					
Intercept	6.843	0.433***	Intercept	8.995	0.227***
TypeMisc	-0.359	0.185*	TypeMisc	0.785	0.304***
TypeSchool	0.252	0.075***	TypeSchool	-0.239	0.118**
TypeTown	0.179	0.200	TypeTown	0.552	0.167***
TypeVillage	-0.269	0.095***	TypeVillage	0.074	0.111
Length	-0.021	0.017	Length	-0.039	0.017**
NoClaimCredit	0.441	0.086***	NoClaimCredit	-0.176	0.095*
logCoverBuilding	0.117	0.039***	logCoverEquipment	0.084	0.030***
logDeductBuilding	0.369	0.037***	logDeductEquipment	0.148	0.045***
RateBuilding	-0.087	0.299	RateEquipment	-0.022	0.041
			ICoverEquipment	-1.077	0.344***

Signif. codes: ***p< 0.01; **p< 0.05; *p< 0.1.

Table C.2: Parameter estimates for the claims frequency sub-model in models A2, A3, A4 and A5.

Building claims frequency			**Equipment claims frequency**		
	Estimate	S.E.		Estimate	S.E.
Models A2 and A4 conditional frequency: Poisson(λ)					
Intercept	-5.035	0.207***	Intercept	-0.670	0.200***
TypeMisc	-0.883	0.111***	TypeMisc	-1.267	0.299***
TypeSchool	-0.020	0.030	TypeSchool	-0.097	0.117
TypeTown	1.295	0.145***	TypeTown	-2.032	0.206***
TypeVillage	0.612	0.065***	TypeVillage	-1.406	0.104***
Length	-0.035	0.007***	Length	-0.035	0.008***
NoClaimCredit	-0.352	0.053***	NoClaimCredit	-0.395	0.078***
logCoverBuilding	1.007	0.015***	logCoverEquipment	0.684	0.023***
logDeductBuilding	0.139	0.016***	logDeductEquipment	0.112	0.019***
RateBuilding	1.526	0.164***	RateEquipment	0.160	0.021***
			ICoverEquipment	-0.407	0.201**
Models A2 and A4 zero-inflation frequency: logit					
Intercept	-8.011	0.715***	Intercept	-2.115	4.918
TypeMisc	0.354	0.301	TypeMisc	0.543	0.527
TypeSchool	1.531	0.125***	TypeSchool	-0.992	0.269***
TypeTown	1.719	0.324***	TypeTown	-2.953	0.890***
TypeVillage	0.848	0.213***	TypeVillage	-2.857	0.527***
Length	-0.094	0.028***	Length	-0.040	0.032
NoClaimCredit	0.504	0.133***	NoClaimCredit	0.115	0.209
logCoverBuilding	-0.069	0.063	logCoverEquipment	-1.031	0.068***
logDeductBuilding	0.763	0.059***	logDeductEquipment	0.056	0.087
RateBuilding	1.978	0.461***	RateEquipment	-0.696	0.094***
			ICoverEquipment	5.087	5.044
Models A3 and A5 conditional frequency: Poisson(λ)					
Intercept	1.149	0.407***	Intercept	-1.777	0.355***
TypeMisc	-0.583	0.263**	TypeMisc	-1.011	0.405**
TypeSchool	-1.008	0.099***	TypeSchool	0.693	0.189***
TypeTown	-0.072	0.234	TypeTown	-0.951	0.223***
TypeVillage	-0.061	0.125	TypeVillage	-0.160	0.183
Length	-0.058	0.008***	Length	-0.032	0.009***
NoClaimCredit	0.648	0.074***	NoClaimCredit	-0.005	0.088
logCoverBuilding	0.740	0.044***	logCoverEquipment	0.991	0.059***
logDeductBuilding	-0.401	0.037***	logDeductEquipment	-0.037	0.038
RateBuilding	-1.350	0.271***	RateEquipment	0.055	0.058
SD of random intercept	0.798			1.088	
Models A3 and A5 zero-inflation frequency: logit					
Intercept	-2.308	1.890	Intercept	2.939	3.713
TypeMisc	1.226	0.673*	TypeMisc		
TypeSchool	0.913	0.321***	TypeSchool		
TypeTown	1.321	0.670**	TypeTown		
TypeVillage	0.648	0.450	TypeVillage		
Length	-0.238	0.086***	Length	-0.225	0.231
NoClaimCredit	2.483	0.344***	NoClaimCredit	1.040	0.616*
logCoverBuilding	-0.141	0.143	logCoverEquipment	-0.933	0.202***
logDeductBuilding	0.151	0.149	logDeductEquipment	-0.271	0.392
RateBuilding	-2.132	1.458	RateEquipment	-2.669	1.002***
			ICoverEquipment	5.594	3.099*

Signif. codes: ***p< 0.01; **p< 0.05; *p< 0.1.

Table C.3: Parameter estimates for the claims sub-model in models A6 and P1.

Building claims (Tweedie)			**Equipment claims (Tweedie)**		
	Estimate	S.E.		Estimate	S.E.
Model A6: Tweedie					
Intercept	6.944	0.388***	Intercept	6.827	0.273***
TypeMisc	-1.042	0.190***	TypeMisc	-1.148	0.269***
TypeSchool	-0.638	0.084***	TypeSchool	0.102	0.136
TypeTown	0.449	0.183**	TypeTown	-0.311	0.155**
TypeVillage	-0.419	0.115***	TypeVillage	-0.358	0.129***
Length	-0.029	0.019	Length	-0.035	0.018*
NoClaimCredit	0.009	0.089	NoClaimCredit	-0.543	0.107*
logCoverBuilding	0.823	0.038***	logCoverEquipment	1.145	0.031***
logDeductBuilding	0.031	0.036	logDeductEquipment	0.123	0.048***
RateBuilding	-0.054	0.253	RateEquipment	0.458	0.045***
			ICoverEquipment	-1.617	0.373***
Tweedie-power	1.66			1.47	
Tweedie-dispersion	164			403	
Model P1: Random-intercept Tweedie (independent)					
Intercept	7.578	0.613***	Intercept	5.649	0.474***
TypeMisc	-0.901	0.282***	TypeMisc	-1.528	0.519***
TypeSchool	-0.857	0.138***	TypeSchool	0.516	0.259**
TypeTown	-0.699	0.275**	TypeTown	-0.911	0.292***
TypeVillage	-0.301	0.173*	TypeVillage	-0.169	0.242
Length	-0.011	0.018	Length	-0.073	0.018***
NoClaimCredit	0.185	0.093**	NoClaimCredit	-0.124	0.117
logCoverBuilding	0.931	0.061***	logCoverEquipment	1.351	0.064***
logDeductBuilding	-0.171	0.058***	logDeductEquipment	0.174	0.085**
RateBuilding	-0.451	0.402	RateEquipment	0.529	0.084***
			ICoverEquipment	-2.121	0.606***
Tweedie-power	1.59			1.44	
Tweedie-dispersion	198			371	
SD of random intercept	1.181			1.497	

Signif. codes: ***p< 0.01; **p< 0.05; *p< 0.1.

Table C.4: Parameter estimates for the claims sub-model in model A7.

Building claims			**Equipment claims**		
	Estimate	S.E.		Estimate	S.E.
Tobit type 2: outcome z_2^*					
Intercept	5.814	0.456***	Intercept	7.831	0.375***
TypeMisc	-0.033	0.208	TypeMisc	0.072	0.372
TypeSchool	-0.355	0.104	TypeSchool	-0.012	0.150
TypeTown	0.144	0.195	TypeTown	-0.098	0.214
TypeVillage	-0.042	0.106	TypeVillage	-0.275	0.139**
Length	-0.004	0.018	Length	-0.038	0.021*
NoClaimCredit	0.072	0.101	NoClaimCredit	-0.187	0.117
logCoverBuilding	0.700	0.060***	logCoverEquipment	0.712	0.073***
logDeductBuilding	0.047	0.049	logDeductEquipment	0.079	0.055
RateBuilding	0.490	0.319	RateEquipment	0.161	0.055***
			ICoverEquipment	-1.000	0.410**
Tobit type 2: Probit selection z_1^*					
Intercept	0.743	0.233***	Intercept	-1.265	0.149***
TypeMisc	-0.426	0.093***	TypeMisc	-0.574	0.163***
TypeSchool	-0.657	0.048***	TypeSchool	0.314	0.076***
TypeTown	-0.144	0.089	TypeTown	-0.371	0.090***
TypeVillage	-0.119	0.057**	TypeVillage	0.016	0.072
Length	0.023	0.010**	Length	-0.006	0.012
NoClaimCredit	-0.336	0.044***	NoClaimCredit	-0.157	0.058***
logCoverBuilding	0.467	0.023***	logCoverEquipment	0.626	0.021***
logDeductBuilding	-0.287	0.022***	logDeductEquipment	-0.062	0.017***
RateBuilding	-0.109	0.154	RateEquipment	0.261	0.028***

Signif. codes: ***p< 0.01; **p< 0.05; *p< 0.1.

C.3 A Simple Example for the Gini Index

This section aims to provide readers with a simple example to help understand the Gini index. Assume there are four customers, A, B, C, and D, in the test set. The upper panel of Table C.5 summarises the premiums paid and claims made by these four customers. For illustration purposes, we compare two models, Model 1 and Model 2, from the perspective of their Gini index. The middle panel of Table C.5 summarises the predicted claims and other information of Model 1 that assumes claims are constant, while the bottom panel summarises the information for Model 2 which is good at predicting claims risk.

Note underlying assumptions for this simple example are the premium charged cannot be perfectly aligned with claims, and that relativity needs to be different for different customers. In practice, these two assumptions are met in most cases.

Table C.5: A simple example for the Gini index with four customers.

Customer	**A**	**B**	**C**	**D**	**Sum**
Premium	1	2	3	4	10
Claim	0	1	4	5	10
Model 1: constant claims					
Predicted claim ($\hat{\mu}$)	2	2	2	2	
Relativity ($R = \frac{\hat{\mu}}{\text{Premium}}$)	$\frac{2}{1}$	$\frac{2}{2}$	$\frac{2}{3}$	$\frac{2}{4}$	
Ascending order of customers based on relativity	D	C	B	A	
Share of Premiums (X-axis)	$\frac{4}{10}$	$\frac{7}{10}$	$\frac{9}{10}$	$\frac{10}{10}$	
Share of Claims (Y-axis)	$\frac{5}{10}$	$\frac{9}{10}$	$\frac{10}{10}$	$\frac{10}{10}$	
Model 2: a powerful model					
Predicted claim ($\hat{\mu}$)	0.1	1	4	5	
Relativity ($R = \frac{\hat{\mu}}{\text{Premium}}$)	$\frac{0.1}{1}$	$\frac{1}{2}$	$\frac{4}{3}$	$\frac{5}{4}$	
Ascending order of customers based on relativity	A	B	D	C	
Share of Premiums (X-axis)	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{7}{10}$	$\frac{10}{10}$	
Share of Claims (Y-axis)	$\frac{0}{10}$	$\frac{1}{10}$	$\frac{6}{10}$	$\frac{10}{10}$	

Here, the lower the relativity, the more profitable the customer is expected to be for the insurer. Thus, according to Model 1, the most profitable customer is D, while the least profitable customer is A. However, this conclusion based on Model 1 is clearly wrong because we observe that customer D takes 40% of the premiums but accounts for 50% of the claims. We can visualise the observation based on Model 1 through the ordered Lorenz curve in Figure C.2. The Gini index of Model 1 is -0.2,

showing Model 1 can not accurately identify the order of customers in terms of their profitability.

Similarly, according to Model 2, the most profitable customer is A, while the least profitable customer is C. This conclusion based on Model 2 is correct because we observe that customer A takes 10% of the premiums while accounting for 0% of the claims. We can visualise the observation based on Model 2 through the ordered Lorenz curve in the right panel of Figure C.2. The Gini index of Model 2 is +0.22, showing Model 2 can accurately identify the order of customers in terms of their profitability.

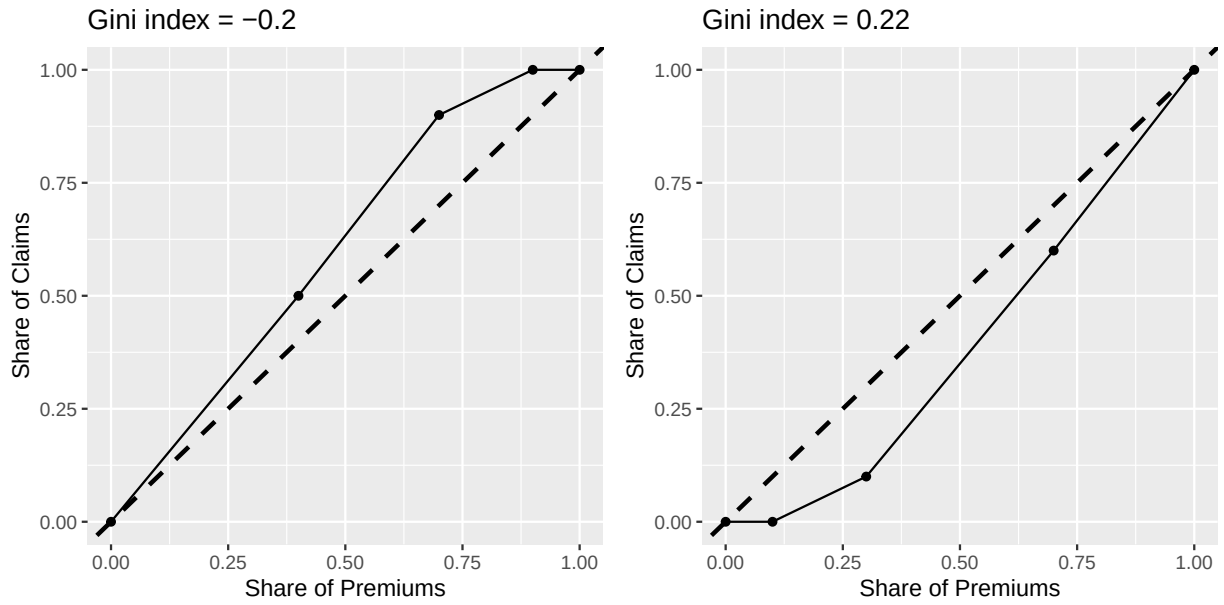

Figure C.2: The ordered Lorenz curve obtained by model 1 (left) and model 2 (right).

In summary, one can intuitively see that a model with a higher Gini index is better at predicting claims risk, leading to a correct order of customers' profitability. For readers who are interested in more applications of the Gini index in the insurance industry, please see Frees et al. (2014).

C.4 The Approach to Adjust the Predicted Net Profits based on Deductibles

To illustrate the approach used in the paper, we start with the predicted profit, which is the value of a customer before applying any deductibles. The predicted profit before applying the deductible for the i-th customer by k-th claim category

at time t is

$$\hat{\text{Profit}}_{i,t,k} = \hat{\text{Premium}}_{i,t,k} - \hat{\mu}_{i,t,k},$$

where $\hat{\text{Premium}}_{i,t,k}$ is the predicted premium and $\hat{\mu}_{i,t,k}$ is the predicted ground-up loss (claims before applying deductibles). To account for the deductible (the amount paid by the customer themself before the insurance company starts to pay), we scale up the $\hat{\text{Profit}}_{i,t,k}$ by scaling up the $\hat{\text{Premium}}_{i,t,k}$. The main reason that we do not adjust the predicted ground-up loss, $\hat{\mu}_{i,t,k}$, is that $\hat{\mu}_{i,t,k}$ is affected by the choice of the claims sub-model, while $\hat{\text{Premium}}_{i,t,k}$ is not. Thus, adjusting $\hat{\text{Premium}}_{i,t,k}$ is consistent with the logic of this article; assuming that predictions of premiums remain unaffected by our models. Consequently, the predictive performance of CLV depends on how well a model predicts ground-up losses and churn rates jointly.

The predicted net profit after applying the deductible for the i-th customer by k-th claim category at time t is

$$\text{Net } \hat{\text{Profit}}_{i,t,k} = \hat{\text{Premium}}_{i,t,k} \times \big(1 + LER_k(d)\big) - \hat{\mu}_{i,t,k},$$

where $LER(d)_k$ is the loss elimination ratio used to scale up the net profit given the deductible level d. In LGPIF, customers have to choose a deductible from a set of deductible levels (can be 0) offered by LGPIF. The following three steps summarise how to estimate $LER_k(d)$ using the training data.

1. Select the observations (customer per year) who choose d as their deductible for their k-th claim category, and the total number of these observations is denoted by $N_k(d)$.

2. Let $y_{i,t,k}$ denote the selected customers' reported loss before applying the deductible d, and $c_{i,t,k}$ denote their actual reimbursement after applying the deductible d. Then, let

   $$\text{Adjustment}_{i,t,k}(d) = y_{i,t,k} - c_{i,t,k},$$

   where $\text{Adjustment}_{i,t,k}(d)$ represents the difference between a customer's reported loss and actual reimbursement due to the deductible level d. Note when $d = 0$, then $y_{i,t,k} = c_{i,t,k}$, and $\text{Adjustment}_{i,t,k}(0) = 0$.

3. Estimate the loss elimination ratio by

   $$LER_k(d) = \sum_{i,t} \big(\frac{\text{Adjustment}_{i,t,k}}{\text{Premium}_{i,t,k}}\big)/N_k(d).$$

   Note it is recommended to use Winsorised or trimmed means to calculate

$LER_k(d)$ in practice, both of which are less sensitive to outliers.

C.5 The Distributions of Predicted CLV of Alternative Models (A1 to A7) and Model P1

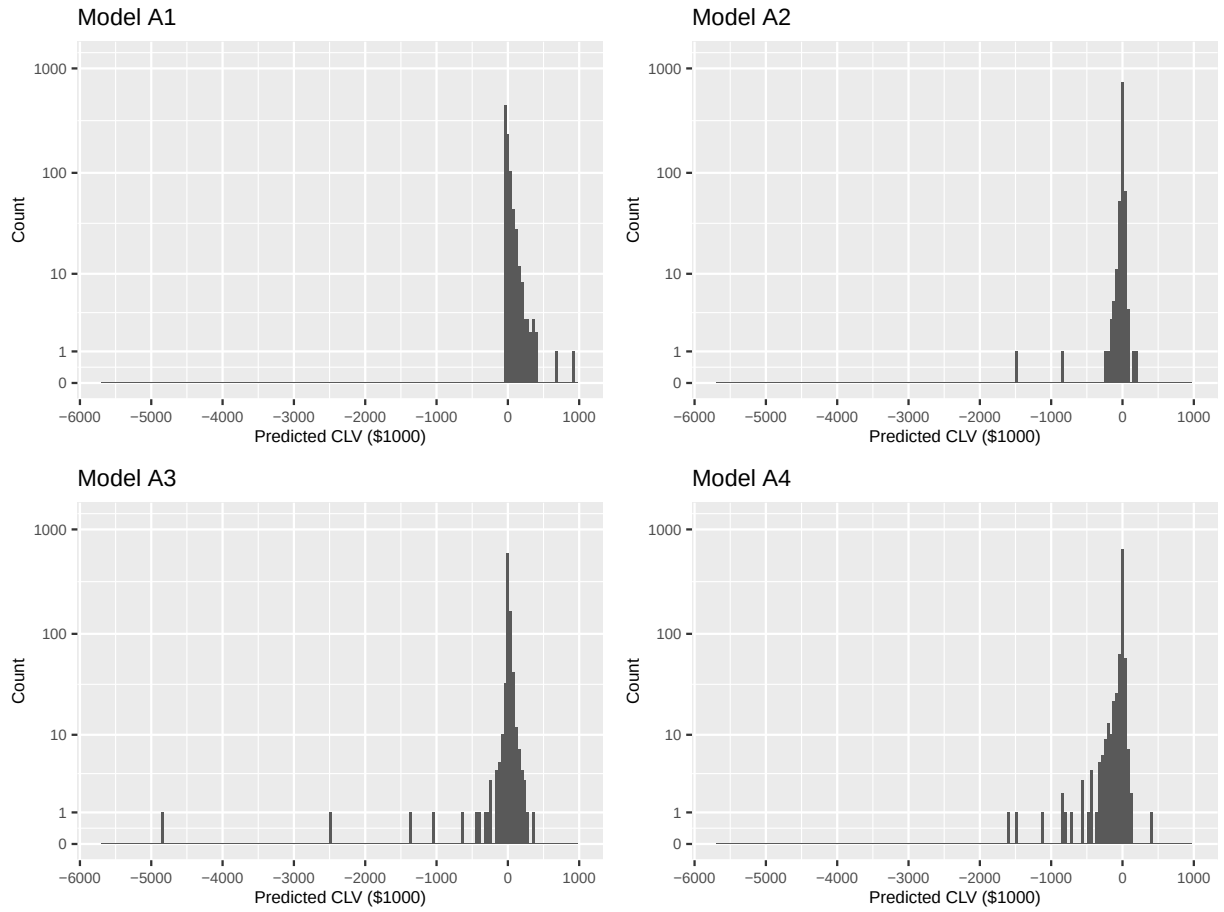

Figure C.3: The distributions of predicted CLV of alternative model A1 to A4. The height of bars is scaled by $\log_{10}$(Count+1).

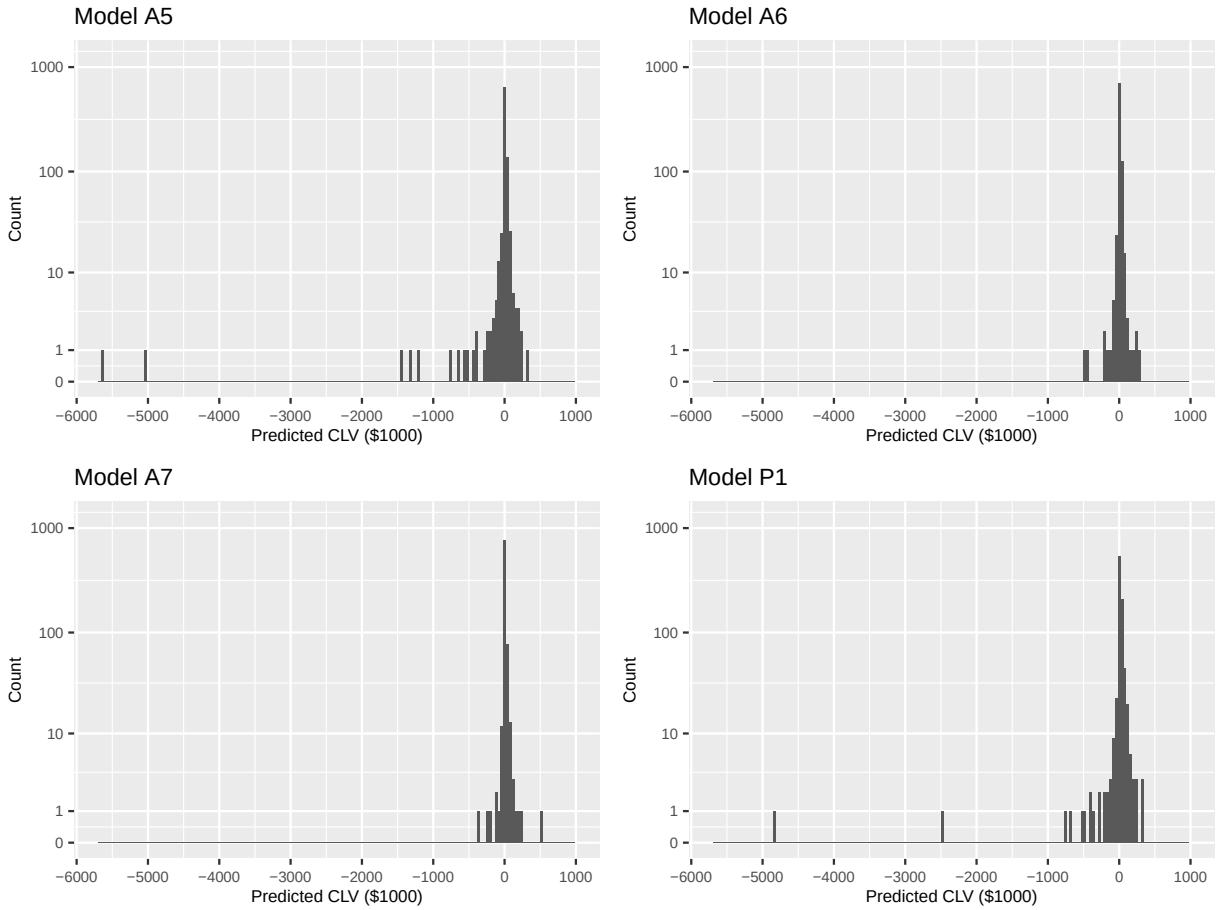

Figure C.4: The distributions of predicted CLV of alternative models A5, A6, A7 and model P1. The height of bars is scaled by $\log_{10}$(Count+1).