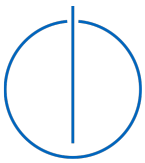
DEPARTMENT OF MATHEMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis

Risk Analysis and Risk Modeling for Natural Catastrophic Events - in the Caribbean

Author:	Abhilakha ("Abby") Das
Supervisors:	Prof. Claudia Czado, Ph.D., Marija Tepegjozova, M.Sc., Francisco Payno y Alegría, M.Sc.
Advisors:	Prof. Claudia Czado, Ph.D.,Marija Tepegjozova, M.Sc.
Submission Date:	26.07.2022



I confirm that this master's thesis is my own work and I have documented all sources and material used.

Munich, 26.07.2022 Abhilakha ("Abby") Das

Acknowledgments

Firstly, I would like to express my sincere gratitude to Professor Claudia Czado for the opportunity, constant support and guidance throughout this project. I am very thankful for the frequent meetings and the substantial time she invested in me during this process.

Secondly, I would like to express my deepest gratitude to Marija Tepegjozova, who gave me this opportunity and constantly encouraged me throughout this project. She has been a constant inspiration; this endeavour would not have been possible without her guidance, advice, support and compassion.

I also want to thank Francisco Payno y Alegría and Munich RE for their generous support and for professionally impacting my life. Without Francisco, I could not have embarked on this journey. Working and collaborating with him and my other colleagues at Munich RE has been my absolute pleasure.

Finally, I am incredibly thankful to my family, friends, and, especially, my partner Christoph Küpper and his family - for the unconditional love and always patiently supporting me throughout these years. Words cannot express my gratitude.

Abstract

Assessing the risks for natural catastrophes in property (re) insurance continues to be notably challenging for underwriters. This is due to the low frequency and high severity nature of catastrophic events. By using statistical models with high precision and strong predictive capabilities - insurers can accurately assess the varying risks found in insurance data. This allows them to measure their potential losses with confidence. Currently, the practical use of available predictive models often does not reflect these conditions and lacks the flexibility required. To address these issues, in this study, we use real-life insurance pricing data and propose models to predict and assess the average rate of loss - in the Caribbean. Specifically, our modelling approach starts by assessing the classical linear models and then extending to the linear mixed and generalized linear models. Here we focus on log-normal and gamma models to accurately capture the high severity of large losses. This study aims to propose the best suitable model given the underlying data, which attains high predictive accuracy. Our results show that the linear mixed model can predict the average loss ratio with high precision. These multilevel models account for varying effects found in different risk class levels (random effects), with fixed and interaction effects of various risk factors. We confirmed these findings by evaluating the model's performance with new unseen (test) data sets.

Contents

List of Figures

List of Tables

1 Introduction

The general idea behind **reinsurance** is often described as the 'insurance of insurance companies' (Hewitt, 2014). In other words, it is the practice of an insurance company contractually transferring portions of underlying insured risks to another insurance company (the reinsurer). In return, the primary insurer of the policy (who seeks to transfer portions of the financial burden) shares the amount paid out by the policyholder (for the insured risk) - known as the policy **premiums** - with the reinsurer (Cremer, 2020). This scenario includes three parties: a **policyholder**, a **primary insurer**, and a **reinsurer** - such that there is no contractual agreement between the insured and the reinsurer (Scherer et. al, 1998). From an insurer's point of view, **a claim** is 'a loss event', as the insurer is required to economically compensate the policyholder based on the reported (insured) event. In other words, the costs paid out by the insurer for the damages (covered by the insurance policy) are known as **losses**.

'Non-life insurance' policies often provide coverage for damages concerning automobiles, business interruptions, or property damages caused by catastrophic events. The burden of large losses motivates primary insurers to seek reinsurance solutions (Ng et al., 2019). Based on Pfeifer and Langen (2021), the insurance company has limited capital to cover all losses caused by catastrophic events; therefore, they need to transfer the risk to the reinsurance, that is, to purchase a catastrophic reinsurance contract. The catastrophic reinsurance contract will protect the insurance company from the immense risk of claims from a group of policyholders that collectively claim their loss caused by a particular event. A reinsurance agreement (referred to as a **'treaty'**) involves only the primary insurer (known as the **'client'** in this study) and the reinsurer - who shares the losses. This implies that reinsurers support their clients with key risk management tasks. Primarily, to withstand the financial burdens of **high-severity** and **low-frequency events**. For this reason, reinsurance portfolios typically consist of a higher number of risks compared to their clients (the primary insurers). This means that not only is it crucial for reinsurers to estimate these varying risks accurately, but it is also significantly challenging (Hewitt, 2019).

Estimating these risks (based on the client) with unpredictable fluctuations in the loss expenditures (also referred to as the **'claims expenditures'**) - are known as the **underwriting risk** (Torre-Enciso and Barros, 2013). Underwriting risk measures the difference between the actual total loss (i.e. 'actual costs of claims) and the expected loss burden incurred in claims expenditures. Hence, given a set of **risk attributes** (properties related to the risk)- an **underwriter** is responsible for estimating and determining each underwriting risk - to ultimately decide on whether or not the risk should be insured (by the insurer or reinsurer). This means that underwriters (who forecast the underwriting risk for each client) are also responsible for pre-determining ***the appropriate premiums and premium rates*** for each risk per client, treaty, market, or policy year. In other words, they must determine the appropriate costs of insurance protection - which are ***unknown at the sale of the treaty*** (Pantelous and Passalidou, 2013). However, given the volatile nature of these *'property catastrophe treaties'* - especially in the Caribbeans (due to the low loss frequencies and high loss severity loss events) - it

is often extremely challenging for an underwriter to accurately estimate the different risks (Pollner, 2001). Specifically, due to the nature of these loss events, it is challenging to forecast, calculate and estimate the overall risks at each **risk class level** (i.e. per client, treaty, market, or policy year). To ultimately ensure that these estimates of premiums (per client) are based on highly accurate statistics. Although the overall premium per client is mainly based on the historical losses and risks, estimates of predicted future insured losses, the generated revenue or capital income of future losses based on treaty coverage ('reserves'), and the cost of capital (expected profit). If the premiums are too low and the risk is underestimated, both the client and reinsurer bear great economic losses. These calculations, historically, in insurance pricing and underwriting risk analysis (especially in reinsurance) did not utilize nor involve much statistical analysis and methods (Ohlsson and Johansson, 2010). Thus, there is a need for statistical methods to ***estimate the expected losses*** - especially in varying risks.

Therefore, this study focuses on underwriting risks in reinsurance arising from the nature of loss events - in the Caribbeans. We specifically look at property catastrophe and large loss treaties designed to protect against large cumulative losses under multiple policies caused by a single natural disaster or other large-scale loss events (Roth Sr and Kunreuther, 1998). The first objective is to provide tools and statistical measures to **compare** the historic observed ('realized') losses given **client and treaty data** (given the expected loss output provided by internal insurance *pricing models*). Then secondly, to ***estimate the predicted loss in property underwriting risk*** at different **risk class levels** (per client, treaty, and country - in the Caribbeans). By proposing a model which describes how the aggregated loss varies and depends on various significant **rating factors** (i.e. covariates based on the properties of clients, policyholders, treaty, or loss events).

To achieve this, we utilize two important key **risk measures** in property insurance treaties - given a certain risk class of losses - per client - (categorized by types of loss events) to answer the following key questions:

1) How often does this event occur? (**Frequency**)
2) What is the aggregated amount of loss - given the type of loss event? (**Severity**)

By modelling the total loss by the **frequency-severity** method - defined as the **Loss Ratio** in this study (often also known as the ***"pure premiums"***) - we can estimate the relationship between the different risk attributes (i.e. rating factors such as properties based on the clients, treaties, countries, and duration of the treaty). To examine this relationship, we also utilize common statistical tools (found in actuarial statistics) to model the given insurance data (provided from the reinsurance's point-of-view). This allows us to estimate the cost of a loss event (Borowicz and Norman, 2006).

There exists a lack of studies focused on assessing internal models implemented by insurance companies (Valeckỳ et al., 2017). Many studies (see Zapart (2013) and Gómez Déniz and Calderín Ojeda (2013)) state that the internal insurance model should be primarily based on either: (1) the constructed probability distribution based on the aggregated losses (derived using classical risk theory) or (2) by **Generalized Linear Models (GLMs)** utilizing individual rating factors. The first types of models are known as *"collective risk models"* (as the losses considered are aggregated with collective risk). Even though, in non-life insurance, typically regression models are often used to describe the total claim amount - or *aggregated losses* - considering the whole portfolio as a collective. The cost of insurance (or annual premiums) may not reflect relevant rating factors (individual characteristics) nor account for the increasingly annual premium differentiation resulting from recent events (Eling et al., 2007). Given our study design of natural catastrophes or large losses in the Caribbean - here, we focus on statistical models which allow us to estimate losses that vary between clients, treaties, countries, and treaty years (in other words, considering the *risk classes*). More specifically, to best ***estimate the***

frequency-severity (the loss ratio) for property insurance in underwriting, concerning individual rating factors, we focus on GLMs.

Since the first application of GLMs by McCullagh and Nelder (1989) to model the claim frequency - GLMs have been proven to be quite efficient and the standard approach for many insurance companies and actuaries around the world (Nelder and Verrall, 1997). To date, the applications of GLMs in insurance can be found in loss or claims reserving, reinsurance, tariff analysis or mortality forecasting or in the reinsurance context (for examples, see Xie et al. (2018)). Since in non-life insurance pricing, the classical **linear models (LMs)** are not entirely suitable or may not allow for appropriate modelling.

This is due to the key assumptions concerning:

i) *the probability distributions or normality* (where the random errors are assumed to be normally distributed) and,

ii) *linearity* (such that the model of the mean or that the expectation of the response variable is assumed to be a linear function of the explanatory variables).

Thus, the popularity of GLMs methods mainly stems from solving these problems. Firstly, it extends to distributions to the exponential family, allowing for regression analysis in non-normal data. It allows for appropriate modelling in non-life insurance pricing data, containing heavily skewed count or binary data (Antonio and Beirlant, 2007). Secondly, it allows for modelling the additive effect of the covariates based on the transformation of the mean (where the linear and multiplicative models are special cases); see Ohlsson and Johansson (2010) for examples.

Additionally, through the applications of GLMs, we can also account for ***heterogeneous risks*** (varying risks between rating factors). A more in-depth discussion on the key assumptions is later provided in this study. Nevertheless, GLM theory generally requires a sample of independent random variables. Though this is a fundamental assumption, due to the nature of insurance data, this assumption is generally not fulfilled in many actuarial and statistical problems (Lee et al., 2018). Mainly, for this study, given the nature of dependency in natural catastrophic events, a large number of treaties, clients, or countries are significantly affected by one natural disaster (such as, for example, a hurricane or a flood). For this reason, we could potentially also look at the applications of **Generalized Linear Mixed Models (GLMMs)**, an extension of GLMs and appropriate for dealing with heterogeneity risks and for data types such as longitudinal, spatial or generally clustered data (where the *independence assumption* is often not fulfilled, see for instance Edwards et al. (2008)).

In non-life insurance, we often deal with **longitudinal data** as our data sets consist of loss observations or risk characteristics that belong to the same policyholder (which may be related). More specifically, in our case, our longitudinal data consists of ***repeated measurements on a group of risk classes*** - which share risk characteristics (for a client, such as for a collection of treaties, countries, treaty years, etc.) and losses observed for the same client over time. Hence, it is noteworthy that correlation structures between a client's loss observations and independent risks cannot be ignored and, thus, should be accounted for or further investigated through various statistical modelling tools (such as linear mixed models).

Linear Mixed Models (LMMs) have comprehensively been utilized for statistical tools and modelling in longitudinal data. According to Antonio and Beirlant (2007), in an insurance context, LMMs extends the classic linear regression model to incorporate client-specific individual effects (***random effects***) alongside rating factors - treated as ***fixed effects*** - in the structure for the mean. For premium rate-making, Frees et al. (1999) and Ohlsson and Johansson (2010) illustrate how LMMs and mixed effect models provide great advantages in actuarial data, precisely for handling categorical variables

with a large number of groups or levels. Therefore, in this study, given a set of ***reinsurance data consisting of large catastrophic losses***, we investigate how the estimates based on the underlying data can be improved by using different specifications and classes of models (i.e. comparing LMs, LMMs, or GLMs). Specifically, we compare how well these models perform or how well our data-sets can be estimated in different given scenarios: such as when we account for random effects or not, with respect to the frequency-severity (or loss ratio) rates. In other words, we need to investigate and compare how the estimates of loss ratios are affected for each model class. We also analyze the estimates using different GLMs and LMMs model specifications - concerning the distribution of ***the loss ratio using the loss severity and frequency***.

1.1 Outline

First, in Chapter 2, we briefly define key concepts in an insurance context and the mathematical and statistical concepts in risk theory related to this study. Specifically, in Section 2.1, we first define the hierarchical structure of the risk class levels found in our data set (i.e. the clients, treaties, and countries) and risk measures used in this study. Based on this, in Section 2.2, we then provide the mathematical framework of each risk measure considered in this study. This includes outlining the frequency-severity method to model the response - the loss ratio per exposure volume (in Section 2.2.6). Chapter 3 provides the theoretical background of the models and performance measures used in this study. We start by introducing the *classical normal linear model* (LM) framework in Section 3.1, which includes the basic model assumptions and general formulation. This also includes the parameters and specifications of the model building. We then repeat this for the model formulation of LMMs (in Section 3.2.2) and GLMs (in Section 3.3.1). Whereas, in Section 3.4 we define the performance evaluation measurements and diagnostics measures used in our study. This includes defining the goodness of fit measures and the statistical tests used per model class.

Furthermore, the analysis of the insurance data and exploration were conducted and outlined in Chapter 4 - not included nor provided in this publication due to confidentiality reasons. This included a description of all variables contained in the data sets for this study. Additionally, we also utilize additional ***Explanatory Data Analysis (EDA)*** tools to provide more insights into our data set and further analyze the response at each risk class level (through univariate and multivariate analysis. Following this, the discussion of the results of all the natural catastrophe models analyzed is provided in Chapter 4. We start by comparing the results of the LMs (in Section 4.1) and then perform residual analysis on the best performing LM (see Section 4.1.1). The same process of assessing then repeated for LMMs, in Section 4.2, and for GLMs in Section 4.3. Then, in Chapter 5, we select the best performing model out of all model classes considered in this study. This includes analyzing each selected model's performance on the testing data and then providing a final interpretation of the selected model (see Section 5.1). Finally, our main findings and study conclusions are discussed in Chapter 6.

2 Basic Concepts in Risk Modelling for Insurance Pricing

In this chapter, we start by presenting basic concepts in non-life insurance, and define the key risk components found in this study. Then, based on these definitions and concepts, we mathematically define our random variables and our variable of interest, i.e. our risk measures.

2.1 Risk Components

The costs of insurance protection (the *'premiums'*) are unknown at the sale of the treaty (Frees, 2018). This means both primary insurers and reinsurers must predetermine the appropriate price for each insured risk. For this reason, "individual insurance characteristics" (policyholder or client level risk attributes) are often analyzed and "pooled" together - to forecast the premium rates (the cost of insurance). Risk attributes at each risk class level - are referred to as the *'risk rating factors* (Ohlsson and Johansson, 2010). This means we aggregate information of treaties (policies) with common properties to group together attributes into classes - with varying risks, losses, or premiums. This information on the risks include, for instance, data regarding the premium, claim counts, exposures, losses, or regions (typically obtained by various databases (Dahen and Dionne, 2010)).

Payments caused by insured random events occurring within an active treaty period is called *'insurance claims'*. From an insurer's point of view, the total cost of paid and outstanding claims arising within the treaty, the period is known as a *loss*. For this study, we look at two specific aggregate claim databases (from the reinsurer's point of view):

- ***Treaty (Policy) Database***: contains data about the risk concerning the insured, and the properties of the policyholder (i.e. line of business, location, insured objects) with the treaty provisions (type of treaty)
- **Claims Database**: contains data and properties of the losses and claims - based on the policy database. This also contains related geographic policy properties (such as countries, markets, or geographic regions, or general properties related such as income per capita, economic statistics, population density, etc.).

To estimate the loss ratio (or loss severity and frequency), we analyze risks given individual rating factors grouped by risk class levels. This allows us to model the statistical relationship between the expected loss ratio and the different rating factors, given the groups of risk class levels.

2.1.1 Risk Class Levels

Typically in *multiplicative models* in actuarial pricing or tariff analysis - it is assumed that the structure of the rating classes reflects the quality of each risk class group appropriately. Therefore, we can estimate the frequency and severity based on the groups of the risk class levels by aggregating or computing the *total losses* based on their '**risk class levels'** which share similar risks. To illustrate this, we outline an example of the simple non-parametric modelling approach provided by Wuthrich (2020).

Assume $\boldsymbol{R}$ denotes the number of rating factors or explanatory variables. Each rating factor $r \in R$ is then divided into risk classes. For a rating factor i the number of risk classes is denoted by r_i. For simplicity, let us consider two covariates, $R = 2$. Then, suppose for the first covariate we choose $\boldsymbol{\mathcal{I}}$ risk levels (labels) such that $i \in \{1, \ldots, \boldsymbol{\mathcal{I}}\}$, while for the second covariate we choose $\mathcal{J}$ risk levels $j \in \{1, \ldots, \boldsymbol{\mathcal{J}}\}$. Then, as shown in tab:riskclasses, we have $R = \boldsymbol{\mathcal{I}} \cdot \boldsymbol{\mathcal{J}}$ risk classes (categorical).

	$1 \cdots j \cdots J$
1	
$\vdots$	
i	risk classes (i, j)
$\vdots$	
I	

Table 2.1. Example of risk class groups and levels. $\boldsymbol{\mathcal{I}}$ and $\boldsymbol{\mathcal{J}}$ for $R = 2$ covariates.

In this example, the risk classes (i, j) represent the risk class belonging to the first and second covariate, respectively. Hence, by "clustering" or grouping similar risk class levels, we can build risk cells (which share similar policy-based characteristics). This allows for non-parametric modelling and, more importantly, introduces a categorical approach for modelling heterogeneous risks.

Hence, in this study, we use the following risk class definitions and levels:

Clients. Consider $\boldsymbol{K}$ **total clients**, such that each index $\boldsymbol{k, k = 1, , ..., K}$ denotes a unique independent **client** $\boldsymbol{k}$.

Treaties. In this study, an insurance contract between client $\boldsymbol{k}$ (the primary insurer) and the reinsurer is referred to as a **'treaty'**. Such that, if there are $\boldsymbol{P^k}$ **total number of treaties of client** $\boldsymbol{k}$ (insurance contracts), then a treaty of client k is indexed by $\boldsymbol{i,\ i = 1, .., P^k}$.

Countries. Due to the nature of this study (with multi-island clients in the Caribbeans), we also analyze the realized incurred losses per market (referred to as 'country') per client $\boldsymbol{k}$. Let $\boldsymbol{\mathcal{M}^{ki}}$ be the **set of all countries** considered in this study (located in the Caribbeans), by client k. Then the country insured by $\boldsymbol{i}$-th treaty, for a client $\boldsymbol{k}$ - is given by $\boldsymbol{j,\ j \in \mathcal{M}^{ki}}$.

Treaty Year. (Underwriting Year). The calendar year to which business is allocated to monitor underwriting statistics - is known as the (underwriting) treaty year is fixed. We consider each incurred loss at the time (given in years) known as the **'treaty year'** Let $\boldsymbol{\mathcal{T}^{kij}}$ be the **set of all treaty years** for country $\boldsymbol{j}$, within treaty $\boldsymbol{i}$ of client $\boldsymbol{k}$. Then, each $\boldsymbol{t,\ t \in \mathcal{T}^{kij}}$ is the treaty year in the $\boldsymbol{i}$-th treaty of client $\boldsymbol{k}$ (in country $\boldsymbol{j}$). For this study, each treaty year $\boldsymbol{t}$ is fixed to one accounting (underwriting) year.

By aggregating the losses at each risk class, we can control some of these risks in a acceptable way

if the number of treaties and risk classes are large. Hence, we first group together all risk attributes (covariates) belonging to an individual client and aggregate them based on common shared properties (i.e. their treaties, countries, and treaty years). This allows us to model the ***'total loss amount'*** with differing risks (heterogeneous risks). However, to model the loss ratio in relation to the risk-related variables individually - known as ***'risk attributes'*** (i.e. properties of risk class levels and attributes, policyholder, etc.) - we must first define the ***'risk measures'*** - required to calculate the loss ratio - at each risk class level.

2.2 Risk Measures

Following the definitions outlined by Ohlsson and Johansson (2010), we introduce the risk measures utilized in this study (including the premium rating factors, commonly found in models for non-life insurance pricing) as:

Loss Amount. The random loss, denoted by $X_{\{\bullet\}}$, is the loss amount reported for the coverage - corresponding to the risk class level (given in monetary units). In this study, this is our variable of interest (in addition to the exposure volume). The average total loss amount given the number of losses per risk class level - is often known as the **loss severity** denoted by $S_{\{\bullet\}}$.

Loss Frequency. The random number of loss events per risk class level, denote as $N_{\{\bullet\}}$. In this study, we consider the average number of losses per exposure volume unit known as the **'standardized loss frequency'**, $N^{v}_{\{\bullet\}}$.

Exposure Volume. In this study, we consider a collection of different *heterogeneous risks*. Thus, to conduct comparable analysis based on the different individual risks for a client k based on the risk class levels (i.e. for all P^{k} treaties, $\boldsymbol{\mathcal{M}}^{ki}$ countries, and, $\boldsymbol{\mathcal{T}}^{kij}$ treaty years) - we standardized our random variables with a volume measure - referred to as the 'risk exposure', $v_{\{\bullet\}}$, (a deterministic variable).

Loss Ratio (Key Loss Ratios - Response). As the random loss amount, $X_{\{\bullet\}}$, depends on the exposure volume the response variable, given by $v_{\{\bullet\}}$ - rather than considering only the response variable - analysis is conducted on the ratios of the random variable $X_{\{\bullet\}}$ with different exposure volume measures, i.e. $\frac{X_{\{\bullet\}}}{v_{\{\bullet\}}}$ (at different risk class levels). In other words, if X is our random variable of interest and v is our volume of exposure. Hence, here the general ratio, given by LR, is defined as $LR = X/v$ (also called the **key loss ratio**).

Note, all *key loss ratios* often used in similar studies are of the same type (Ohlsson and Johansson 2010). In other words, the loss frequency, severity, and loss ratios defined above are all types of 'key loss ratios'. These ratios differ based on the different exposure measures $v_{\{\bullet\}}$ (at different risk class levels) and the random variable of interest X (for example, see Table 2.2).

Due to the scope of this study, we focus on key loss ratios concerning only the random loss amount $X_{\{\bullet\}}$ - as our response variable - while utilizing various exposure volume units for different loss events. Specifically, we use the amount of insurance coverage as our exposure measure for **natural catastrophic** loss events (in property insurance). This is often used in commercial business properties as the amount of insurance increases as property values grow with inflation (i.e. estimates are less sensitive to inflation). Whereas for **large losses** we choose the amount of premiums (cost of insurance coverage paid by the policyholder) as our exposure volume measure.

We now mathematically define each risk measure considered in this study - in order to estimate the

Variable of Interest X	**Exposure** V	**Key Loss Ratios:** $LR = X/V$
Counts of Random losses	Length of Treaty Coverage (years)	Loss Frequency
Amount of Random Losses (cost)	Counts of Random Losses	Loss Severity
Amount of Random Losses (cost)	Length of Treaty Coverage (years)	Known as 'Pure Premium' (referred to as 'Loss Cost')
Amount of Random Losses (cost)	Earned Premium (currency)	Total Ratio of Claims / Losses
Count of Large Losses	Counts of (all) Random Losses	Proportion of Large Losses
Total Premium (currency)	Total Amount of Sum Insured (currency)	Rate

Table 2.2. Description of typical key loss ratios commonly found in actuarial studies.

loss ratio, $LR_{\{\bullet\}}$, at each risk class level.

2.2.1 Random Loss Amount

The classical individual risk model (in risk theory) often focuses on each random loss arising from each risk class level, aggregated over different risk attributes or rating factors, such as treaty i for a client k. In this section, we first define the ***random loss amount*** (given in monetary values) based on the k-th client (required to later calculate the ***'severity'***, i.e. the average loss per event, in Section Section 2.2.5). Specifically, we define the individual amount of random individual losses (or simply 'random losses') - given in monetary values (i.e. a positive continuous random variable) - and aggregate the individual losses at each risk class level (i.e. based on each k-th client per i-th treaty and j-th country, in t treaty years).

Definition 2.1: Random Individual Loss Amount

Assume the random variable $\boldsymbol{X_{kijtl}}$ denotes **an individual random loss amount** for a loss $\boldsymbol{l, l \in \mathcal{L}^{kijt}}$ observed for a client $k,\ k = 1, ..., K$, in treaty $i,\ i = 1, ..., \boldsymbol{P}^k$ at country $j \in \boldsymbol{\mathcal{M}}^{ki}$ - which occurred in treaty year $t \in \boldsymbol{\mathcal{T}}^{kij}$. Then the sum of all l losses, given by,

$$X_{kijt} := \sum_{t \in \boldsymbol{\mathcal{T}}^{kij}} X_{kijtl}\ , \quad \text{for } k = 1, .., K,\ i = 1, .., P^k,\ j \in \boldsymbol{\mathcal{M}}^{ki}, \tag{2.1}$$

is the random loss amount which belongs to the risk class level group (k, i, j, t).

This means that ***the aggregated individual random losses***, $\boldsymbol{X_{kij}}$, over all $\boldsymbol{\mathcal{T}}^{kij}$ ***treaty years*** can be defined as

$$X_{kij} = \sum_{t \in \boldsymbol{\mathcal{T}}^{kij}} X_{kijt}\ , \quad \text{for } k = 1, .., K,\ i = 1, .., P^k,\ j \in \boldsymbol{\mathcal{M}}^{ki}, \tag{2.2}$$

based on a client k, within treaty i and country j.

Whereas,

$$X_{ki} = \sum_{j \in \mathcal{M}^{ki}} X_{kij} = \sum_{j \in \mathcal{M}^{ki}} \sum_{t \in \mathcal{T}^{kij}} X_{kijt} \ , \quad \text{for } k = 1, .., K, \ i = 1, .., P^k, j \in \mathcal{M}^{ki}. \tag{2.3}$$

is the ***the aggregated of random loss events per treaty over all $\mathcal{M}^{ki}$ countries.*** Based on this, we can now calculate **total random loss for each client** k based on all above risk classes.

Definition 2.2: Total Random Loss

The **the total random loss of client k** is denoted by X_k, calculated as

$$X_k = \sum_{i=1}^{P^k} X_{ki} \quad = \sum_{i=1}^{P^k} \sum_{j \in \mathcal{M}^{ki}} X_{kij} \quad = \sum_{i=1}^{P^k} \sum_{j \in \mathcal{M}^{ki}} \sum_{t \in \mathcal{T}^{kij}} X_{kijt} \ , \quad \text{for } k = 1, .., K, \tag{2.4}$$

which is the sum of random losses of client k over all P^k treaties and $\mathcal{M}^{ki}$ countries.

This means that (from a reinsurer's point of view) **the 'ultimate' total amount of random loss based on all K clients** can now be defined as

$$X = \sum_{k=1}^{K} X_k. \tag{2.5}$$

For an example of the breakdown of each total random loss based on client k (such that $k = 1$) see Figure 2.1. Based on a client ($k = 1$), it illustrates a flowchart for two treaties i with countries $j \in \mathcal{M}^{ki}$, given the $t \in \mathcal{T}^{kij}$ treaty years. The total random loss amount for a client $k = 1$ is given by each node (risk class level). Here, the client k has two treaties $i = 1, 2$ that provides coverage provided in countries j. Treaty $i = 1$ is active for two countries j - such that, in the first country is active for 2 treaty years t and the second country is active for one treaty year t. Whereas, treaty $i = 2$ is only active in one country j but provides coverage for three treaty years t.

2.2.2 Random Loss Counts

By analyzing the random counts or ***the number of loss events*** (also known as the number of risks or 'claim counts'), enables us to account for the risk attributes that influence the occurrence of a loss event. Also, by doing this, we can further investigate high or low occurring loss events - based on the k-th client who has a treaty i in-country j, for the treaty year t.

Similar to the set-up for the random losses above, we can now also define the frequency $N_{\{\bullet\}}$ per risk class level.

Definition 2.3: Random Counts of Loss

Let the random count (discrete) variables $\boldsymbol{N_{kijt}}$ denote **the number of random loss events (random counts)**, for each random loss in $t \in \mathcal{T}^{kij}$ treaty year - in country j of treaty i for

each client $k,\ k = 1, ..., K$. Such that, we set $l, l \in \mathcal{L}^{kijt}$ as an index for a loss occurring in treaty year t (in country j of treaty i for each client $k,\ k = 1, ..., K$). Then, the number of random loss events for a random loss t are given by

$$N_{kijt} = \sum_{l \in \mathcal{L}^{kijt}} I_{kijtl} \ , \quad \text{for } k = 1, .., K,\ i = 1, .., P^k,\ j \in \boldsymbol{\mathcal{M}}^{ki},\ l \in \mathcal{L}^{kijt}, \tag{2.6}$$

where $\boldsymbol{I_{kijtl}}$ are indicator (count) random variables defined as

$$I_{kijtl} := \begin{cases} 1, & \text{if the } l\text{-th loss occurred in year t, at country j for treaty i of client k} \\ 0, & \text{otherwise,} \end{cases} \tag{2.7}$$

based on a client $k, k = 1, .., K$, with treaty $i, i = 1, .., P^k$, valid in country $j, j \in \boldsymbol{\mathcal{M}}^{ki}$ - for a loss occurring within the period of the treaty year $t, t \in \boldsymbol{\mathcal{T}}^{kij}$.

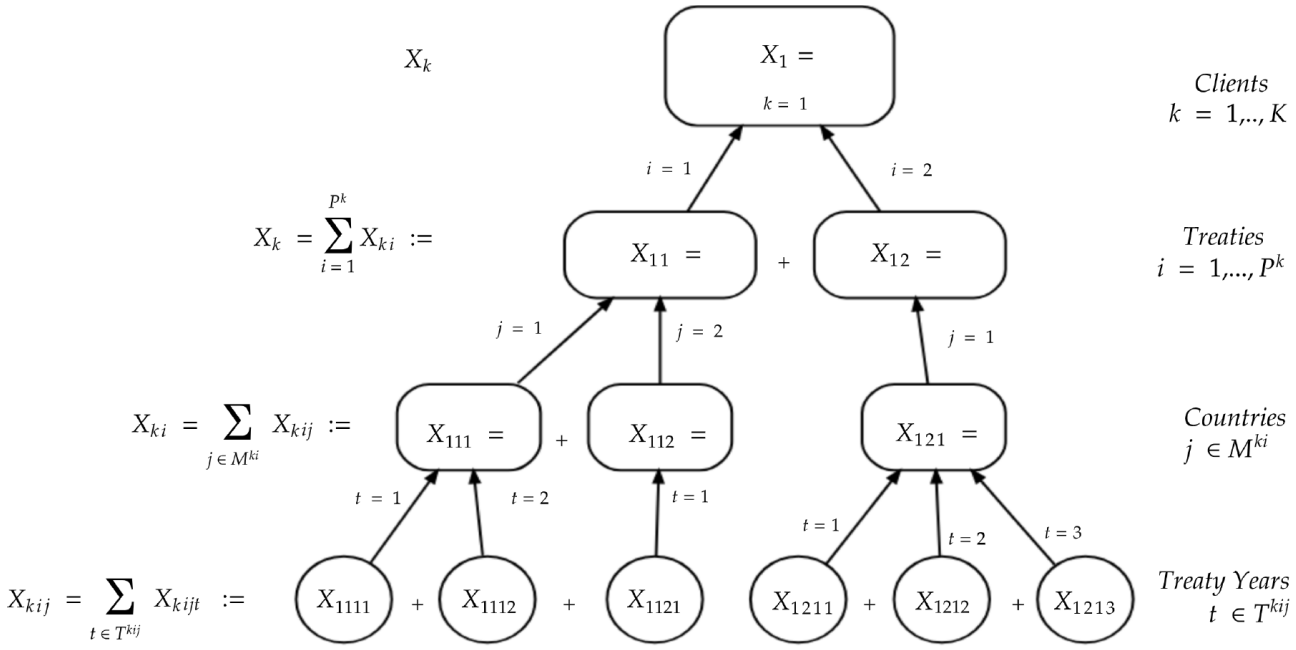


Figure 2.1. Breakdown of Random Losses per Risk Class Level. Based on a client ($k = 1$), a flowchart is illustrated for two treaties i with (overall) two countries j, given the respective treaty years t.

Thus, if the total number of random losses which occur within a treaty year $t, t \in \boldsymbol{\mathcal{T}}^{kij}$ (based on country j active in treaty i for a client $k,\ k = 1, ..., K$) - is given by N_{kijt}. Then the corresponding **frequency of random loss events over all $\boldsymbol{\mathcal{T}}^{kij}$ treaty years**, denoted by $\boldsymbol{N_{kij}}$, is calculcated as:

$$N_{kij} = \sum_{t \in \boldsymbol{\mathcal{T}}^{kij}} N_{kijt} \ , \quad \text{for } k = 1, .., K,\ i = 1, .., P^k,\ j \in \boldsymbol{\mathcal{M}}^{ki} \ . \tag{2.8}$$

for the k-th client's treaty i in country j.

Whereas, **the frequency of random loss events over all $\boldsymbol{\mathcal{M}}^{ki}$ countries** (in treaty i for client k)

is given by N_{ki}, where

$$N_{ki} = \sum_{j \in \boldsymbol{\mathcal{M}}^{ki}} N_{kij} = \sum_{j \in \boldsymbol{\mathcal{M}}^{ki}} \sum_{t \in \boldsymbol{\mathcal{T}}^{kij}} N_{kijt} \ , \quad \text{for } k = 1, .., K, \ i = 1, .., P^k. \tag{2.9}$$

This allows us to now calculate **the frequency of the total of random loss events** of a client k.

Definition 2.4: Total Random Loss Counts

The total random loss counts for the k-th client is given by

$$\begin{aligned} N_k &= \sum_{i=1}^{P^k} N_{ki}, \quad \text{for } k = 1, .., K \\ &= \sum_{i=1}^{P^k} \sum_{j \in \boldsymbol{\mathcal{M}}^{ki}} N_{kij} \\ &= \sum_{i=1}^{P^k} \sum_{j \in \boldsymbol{\mathcal{M}}^{ki}} \sum_{t \in \boldsymbol{\mathcal{T}}^{kij}} N_{kijt} \ , \end{aligned} \tag{2.10}$$

which are aggregated over all, based on the client k $\boldsymbol{\mathcal{T}}^{kij}$ treaty years in $\boldsymbol{\mathcal{M}}^{ki}$ countries and P^k treaties.

This implies that **the frequency of the "ultimate" random loss for the reinsurer**, for all K clients, is

$$N = \sum_{k=1}^{K} N_k. \tag{2.11}$$

Figure 2.2 provides an example of the flowchart of how each random loss count $N_{\{\bullet\}}$ (frequency) is aggregated and the respective breakdown of a client k (indexed as $k = 1$). In this example, the client $k = 1$ has two treaties (i.e. $P^k = 2$)- indexed by (treaty ID's) $i = 1$ and $i = 2$. Such that the first treaty is active in two countries - labelled as $j = \{1, 2\}$ (for simplicity) for a country $j \in \boldsymbol{\mathcal{M}}^{ki}$. The first country is valid for two treaty years t while the first treaty $i = 1$ provides one year of coverage for the second country. Whereas for the second treaty, $i = 2$, is active for only one country in $\boldsymbol{\mathcal{M}}^{ki}$ but provide coverage for three treaty years.

As each i-th treaty for the k-th client is not the same - we must consider a collection of different heterogeneous risks (within the individual risks). In other words, there exist variations of risks for the client k, based on different attributes (risk characteristics) within all P^k treaties, for $\boldsymbol{\mathcal{M}}^{ki}$ countries, and, during $\boldsymbol{\mathcal{T}}^{kij}$ treaty years. Hence, we introduce 'risk exposure measures' (deterministic) to make these risks for all K clients more comparable - by standardizing or adjusting our losses by the exposure (deterministic) units.

2.2.3 Exposure Volume

Estimating and utilizing the appropriate risk exposure for the upcoming treaty year, t, is a key task in underwriting (Blakley et al., 2001). As it is crucial we take into account the client k's historic losses,

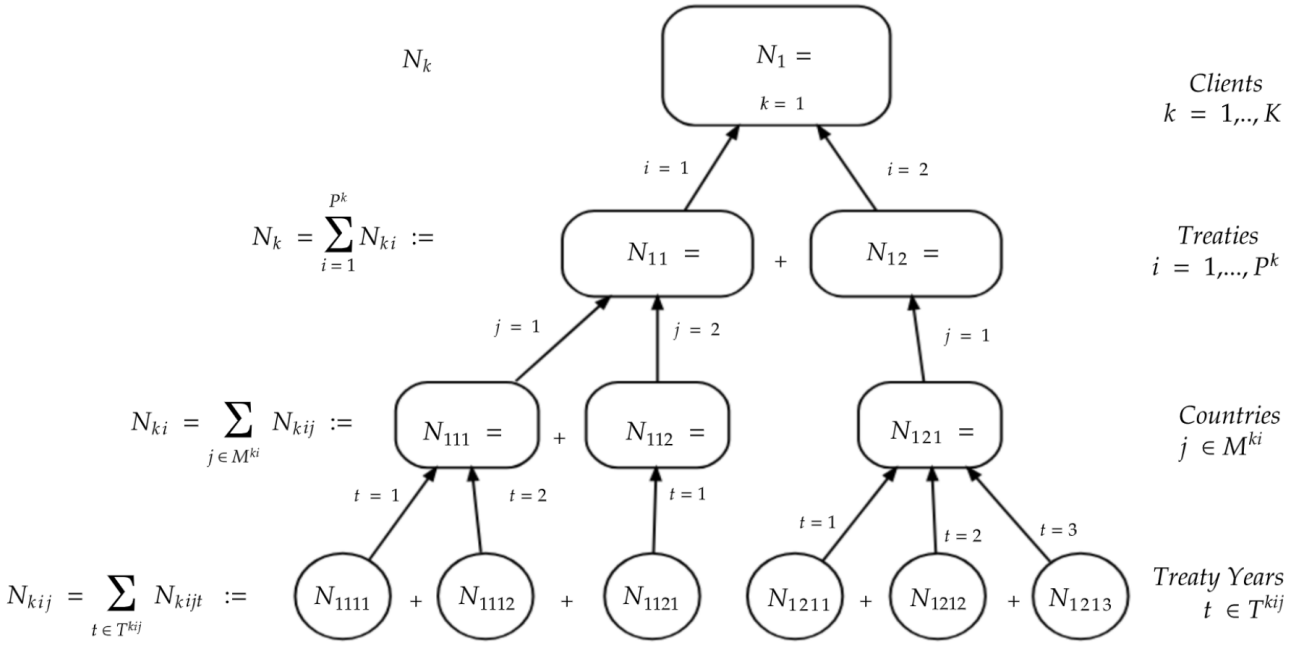


Figure 2.2. Flow chart of Random Counts of Losses, $N_{\{\bullet\}}$. Here, each node represents aggregated random counts for a client k , indexed as $k = 1$.

their strategy - while also adjusting for current economic factors and situations. Though there exist different exposure measures, for this study, we consider and compare two different exposure measures (premiums and the sum insured) - suitable for property insurance - referred to as the ***volume of exposure***

More specifically, we use the following measures for the exposure volumes:

(1) Premiums. Fixed amount paid (to the primary insurer) by the policyholder to cover the risk. In other words, the cost of insurance coverage provided by a policy. Based on the various factors of the policyholder, such as historic losses, type of business, estimated loss size, and frequency. Specifically, in this study, we look at 'written premiums' - which is the cost of coverage of the i-th treaty (per j country) issued ('underwritten') during the treaty year t in question (given monetary values).

(2) Sum Insured. Total amount of coverage payable for a single or multiple loss during a treaty period t. Often referred to as ***'Aggregates'***, it is the volume or portion of the total loss to be covered by the annual risk premium.

Note, for this study; we only consider the total or entire amount of insurance losses per k client - before any treaty deductibles, policy limits, or any application of any retention or reinsurance (known as the 'Ground Up' Loss).

Hence, here we introduce ***'risk exposure measures'***, v, (deterministic) to make these risks for all K clients more comparable - by standardizing or adjusting our random loss counts by the exposure (deterministic) units. Similar to the set-up of the total random loss counts and loss severity, we now define the exposure volume for each client k, treaty i, j treaty and treaty years t.

Definition 2.5: Volume of Risk Exposure

Let $\boldsymbol{v_{kijt}}$ denote **the individual volume of the exposure** for client $\boldsymbol{k}$, $\boldsymbol{(k = 1, .., K)}$ in treaty $\boldsymbol{i}$ $\boldsymbol{(i = 1, .., P^k)}$ and country $\boldsymbol{j \in \mathcal{M}^{ki}}$ per treaty year $\boldsymbol{t \in \mathcal{T}^{kij}}$. Then, for client $\boldsymbol{k}$,

$$v_{kij} = \sum_{t \in \mathcal{T}^{kij}} v_{kijt} \ , \quad \text{for } k = 1, .., K, \ i = 1, .., P^k, \ j \in \mathcal{M}^{ki} \ , \tag{2.12}$$

is **the exposure volume over all $\mathcal{T}^{kij}$ treaty years**, in the i-th treaty coverage for a country j.

This means that,

$$v_{ki} = \sum_{j \in \mathcal{M}^{ki}} v_{kij} = \sum_{j \in \mathcal{M}^{ki}} \sum_{t \in \mathcal{T}^{kij}} v_{kijt} \ , \quad \text{for } k = 1, .., K, \ i = 1, .., P^k, \ j \in \mathcal{M}^{ki}. \tag{2.13}$$

denotes **the k-th client's volume of exposure in all $\mathcal{M}^{ki}$ countries and $\mathcal{T}^{kij}$ treaty years**, based on client k with treaty i.

Similarly, *the total volume of exposure* can also be measured for each k-th client.

Definition 2.6: Total Volume of Exposure

For all P^k treaties belonging to the k-th client, with insurance coverage in $\boldsymbol{\mathcal{M}^{ki}}$ countries, and for $\boldsymbol{\mathcal{T}^{kij}}$ treaty years - the total volume of exposure for a client $\boldsymbol{k}$ is given by

$$\begin{aligned} v_k &= \sum_{i=1}^{P^k} v_{ki}, \quad \text{for } k = 1, .., K \\ &= \sum_{i=1}^{P^k} \sum_{j \in \mathcal{M}^{ki}} v_{kij} \\ &= \sum_{i=1}^{P^k} \sum_{j \in \mathcal{M}^{ki}} \sum_{t \in \mathcal{T}^{kij}} v_{kijt} \ . \end{aligned} \tag{2.14}$$

Then, the reinsurer can calculate the **the cumulative volume of exposure over all clients,** K as

$$v = \sum_{k=1}^{K} v_{ki}. \tag{2.15}$$

It is important to note that, in this study, we require two different exposure measures - since the appropriate exposure volume is dependent on the loss event type of interest (based on the client k). For instance, assume we only considered premiums as the exposure for catastrophic or large loss events (entire loss or ground up). Then, not all loss or pricing developments attached to the risk coverage will be reflected in the premiums for client k (which may produce inaccuracies in key measures, such as loss severity). Since the coverage for previous treaty years t have different development lengths - not all premiums for all P^k treaties are written until the first evaluation date of treaty t (i.e. incomplete or missing data for exposure volume).

2.2.4 Loss Frequency (Standardized with Exposure Volume)

It is generally assumed that the number of losses is proportional to the exposure. This is especially important in non-life insurance pricing, which means that we must standardize and scale our frequency (random counts of losses $N_{\{\bullet\}}$) with the underlying volume of exposure. In this section, we now define the standardized frequency of random loss counts to account for the exposure volume - given in Equation (2.12) - often referred to as the claims frequency or loss frequency (Fowler, 1960). For simplicity, because we are consistently using exposure to describe our random loss count variable N in this study - we refer to the standardized loss frequency as the **'random loss frequency'**, while the random loss frequencies without adjusting for exposure volume is referred to as the **'random loss counts'**.

Definition 2.7: Random Loss Frequency - with Exposure

Let N_{kijt} be the number of random losses and v_{kijt} be the volume of exposure units - in treaty $i,\ i = 1, ..., P^k$, country $j \in \boldsymbol{\mathcal{M}}^{ki}$, per treaty year $t \in \boldsymbol{\mathcal{T}}^{kij}$ (for client k). Then, the loss frequency based on all $\boldsymbol{\mathcal{T}}^{kij}$ treaty years is denoted by $\boldsymbol{N^v_{kij}}$ and calculated as

$$N^v_{kij} = \frac{\sum\limits_{t\in\boldsymbol{\mathcal{T}}^{kij}} N_{kijt}}{\sum\limits_{t\in\boldsymbol{\mathcal{T}}^{kij}} v_{kijt}}, \quad \text{for } k = 1, .., K,\ i = 1, .., P^k,\ j \in \boldsymbol{\mathcal{M}}^{ki}. \tag{2.16}$$

Then the loss frequency over all countries $\boldsymbol{\mathcal{M}}^{ki}$ and $\boldsymbol{\mathcal{T}}^{kij}$, for a client $\boldsymbol{k}$ in treaty $\boldsymbol{i}$ is given by

$$N^v_{ki} = \frac{\sum\limits_{j\in\boldsymbol{\mathcal{M}}^{ki}} N_{kij}}{\sum\limits_{j\in\boldsymbol{\mathcal{M}}^{ki}} v_{kij}}, \quad \text{for } k = 1, .., K,\ i = 1, .., P^k. \tag{2.17}$$

Similarly, for the k-th client having N_k total counts of random losses, while the v_k denotes the corresponding volume of exposure for all P^k treaties. Then we can define the total loss frequency.

Definition 2.8: Total Random Loss Frequency - with Exposure

For the k-th client with P^k treaties and $\boldsymbol{\mathcal{M}}^{ki}$ countries, during the period of $\boldsymbol{\mathcal{T}}^{kij}$ treaty years, the total loss frequency of random counts given by $\boldsymbol{N^v_k}$ can be calculated as,

$$N^v_k = \frac{\sum\limits_{i=1}^{P^k} N_{ki}}{\sum\limits_{i=1}^{P^k} v_{ki}}, \quad \text{for } k = 1, .., K, \tag{2.18}$$

where N_{ki} represents the random loss counts for the risk class (k, i), with corresponding exposure volume v_{ki}.

Finally, **the ultimate loss based on all K clients with P^k treaties active in $\mathcal{M}^{ki}$ countries** can be calculated as

$$N^v = \frac{\sum_{k=1}^{K} N_k}{\sum_{k=1}^{K} v_k}, \tag{2.19}$$

Based on this, if the (cumulative) exposure volume $v > 0$ represents the number of insured risks, then the *expected number of loss counts* is given by

$$\mathrm{E}(N) = \lambda v, \tag{2.20}$$

where $\lambda > 0$ represents the *total expected loss frequency* given by Equation (2.19). Note, at every risk class level (k, i, j, t) we can express the expected loss counts or loss frequency similarly. Using this we can later describe probability distributions for modelling loss counts N given the risk class levels appropriately.

Additionally, given the above breakdown of the random losses and the random counts of losses (in Definition 2.3), we can now also calculate the *'Loss Severity'* - which is the average loss (amount) per loss count or event for the k-th client.

2.2.5 Loss Severity

Often, the **severity** is also referred to as the 'claim severity', the *'average cost per loss ('claim')* or the *'average size of claim or loss'*. This means, for example - given the risk classes (k, i) - the loss severity is the average loss X_{ki} (of the i-th treaty for a client k) per loss count N_{ki}. In this study, we denote the severity as $S_{\{\bullet\}}$.

Definition 2.9: Individual Loss Severity

Assume the following,

- X_{kijt} is the random loss amount for a client k, with treaty $i, i = 1, .., P^k$ active in country $j, j \in \mathcal{M}^{ki}$ - which is valid for the treaty year $t, t \in \mathcal{T}^{kij}$,
- The total number of random loss events based on client k, treaty i, country j occurring in treaty year t is given by N_{kijt} (frequency),

Then, if it holds that $N_{kijt} > 0$, then *the (random) loss severity based on client k, treaty i, country j - for treaty year t* - is the amount of random losses, X_{kijt} divided by the number of random losses N_{kijt} (for the respective client k, treaty i, country j, in treaty year t).

This is denoted as $\boldsymbol{S_{kij}}$ and given as

$$S_{kij} = \frac{\sum_{t \in \mathcal{T}^{kij}} X_{kijt}}{\sum_{t \in \mathcal{T}^{kij}} N_{kijt}}, \quad \text{for } k = 1, .., K, \ i = 1, .., P^k, \ j \in \mathcal{M}^{ki}. \tag{2.21}$$

This means that if the k-th client's loss severity per $t \in \mathcal{T}^{kij}$ treaty year, i-th treaty, and country j is given by $\boldsymbol{S_{kijt}}$ - then the loss severity for each all $\mathcal{T}^{kij}$ treaty years is given by $\boldsymbol{S_{kij}}$.

Hence, now can calculate the **loss severity per treaty**, $\boldsymbol{S_{ki}}$, for all $\boldsymbol{\mathcal{M}^{ki}}$ countries for a treaty $\boldsymbol{i}$ as

$$S_{ki} \;=\; \frac{X_{ki}}{N_{ki}} \;=\; \frac{\sum\limits_{j\in\boldsymbol{\mathcal{M}}^{ki}} X_{kij}}{\sum\limits_{j\in\boldsymbol{\mathcal{M}}^{ki}} N_{kij}}, \qquad \text{for } k=1,..,K,\;\; i=1,..,P^k. \tag{2.22}$$

Hence, we can now define the total severity of random losses, using Equation (2.3) and Equation (2.9).

Definition 2.10: Total Severity of Random Losses

Over all treaties $\boldsymbol{P^k}$, the **total severity of random losses** for the k-th client (for all $\boldsymbol{i=1,..,P^k}$ treaties, $\boldsymbol{j\in\mathcal{M}^{ki}}$ countries, and $\boldsymbol{t\in\mathcal{T}^{kij}}$ treaty years) is given by

$$S_k \;=\; \frac{X_k}{N_k} = \frac{\sum\limits_{i=1}^{P^k} X_{ki}}{\sum\limits_{i=1}^{P^k} N_{ki}}, \quad \text{for } k=1,..,K,\;\; i=1,..,P^k. \tag{2.23}$$

Similarly, from the reinsurer's point of view, we can now also calculate **the cumulative amount of random loss** for all $\boldsymbol{K}$ clients by

$$S = \frac{X}{N} = \frac{\sum\limits_{k=1}^{P^k} X_k}{\sum\limits_{k=1}^{P^k} N_k}, \tag{2.24}$$

aggregated over all P^k total treaties active in all $\boldsymbol{\mathcal{M}^{ki}}$ countries, based on all $\boldsymbol{\mathcal{T}^{kij}}$ treaty years (for K clients).

Based on Definition 2.23, we can now express the mean of the random losses at each risk class (k,i,j,t) in terms of the loss severity. For instance, consider the aggregated cumulative loss amount X. If we make certain assumptions (based on the "standard" collective or aggregated risk model).

Assumptions 2.1: Collective Risk Model

1. The ***random loss counts*** N (at every risk class level) is a discrete random variable, with values in N_0,
2. The (positive) ***random losses*** are identically independently distributed (iid), i.e. for the k-th client the losses, $X_1, X_2, \ldots X_k \overset{iid}{\sim} F$ (where F is the distribution function with $F(0)=0$), and
3. The ***loss counts and loss amount*** (at each risk class) are independent. If we further assume the ***expected cumulative loss severity***, considering all K clients, is given by μ.

Then given these assumptions are fulfilled - the *expected cumulative loss* (aggregated amount for the reinsurer) can be expressed as

$$\mathrm{E}[X] = \mathrm{E}[N]\mathrm{E}[S] = \lambda\mu, \tag{2.25}$$

for exposure volume $v = 1$, whereas the λ is the expected cumulative loss counts as given by Equation (2.20).

2.2.6 Loss Ratio

Finally, in this section, we can now define our response variable in terms of the risk measures defined above by utilizing the frequency-severity approach to calculate the total **'Loss Ratio'** (commonly referred to as 'pure premiums' or lost cost in pricing principals). It is the loss per exposure volume v. Our aim here is to estimate the average loss ratio and to cumulatively predict the risks based on the risk classes (given by the expected loss frequency times the expected average loss severity). In particular, we do this by (for instance) combining the client k's random loss frequency, N_k^v, with the average severity of random loss, S_k, to determine the loss ratio, LR_k.

This provides us with accurate measurements to estimate the overall loss of the risk or the size of the loss ratio (based on the underwriting loss, see Viscusi et al. (1991) for more details). As both the (scaled) frequency and average severity significant impacts the sum of all aggregated observed losses - based on the combined risks - for all K clients with P^k total treaties in $\boldsymbol{\mathcal{M}}^{ki}$ countries, during $\boldsymbol{\mathcal{T}}^{kij}$ treaty years (Zanjani et al., 2010).

Definition 2.11: Individual Loss Ratio

Based on Definition 2.7 and Definition 2.9, we can similarly decompose and calculate our response variable: the loss ratio for client k for all treaty years t - per i-th treaty, j-th country. Thus, the loss ratio $\boldsymbol{LR_{kij}}$ over all treaty years for a country $j \in \boldsymbol{\mathcal{M}}^{ki}$ is given by

$$\begin{aligned} LR_{kij} &= \underbrace{N_{kij}^v}_{\text{(scaled) loss frequency}} \times \underbrace{S_{kij}}_{\text{loss severity}}, \quad \text{for } k = 1,..,K,\ i = 1,..,P^k, j \in \boldsymbol{\mathcal{M}}^{ki} \\ &= \frac{\sum\limits_{t \in \boldsymbol{\mathcal{T}}^{kij}} N_{kijt}}{\sum\limits_{t \in \boldsymbol{\mathcal{T}}^{kij}} v_{kijt}} \times \frac{\sum\limits_{t \in \boldsymbol{\mathcal{T}}^{kij}} X_{kijt}}{\sum\limits_{t \in \boldsymbol{\mathcal{T}}^{kij}} N_{kijt}} \quad \Rightarrow \quad \frac{\sum\limits_{t \in \boldsymbol{\mathcal{T}}^{kij}} X_{kijt}}{\sum\limits_{t \in \boldsymbol{\mathcal{T}}^{kij}} v_{kijt}} = \frac{X_{kij}}{v_{kij}} \end{aligned} \tag{2.26}$$

for $k = 1,..,K$ clients, $i = 1,..,P^k$ treaties, and $j \in \boldsymbol{\mathcal{M}}^{ki}$ countries.

Consequentially, this means that for **a client k's loss ratio based on a treaty i - over all $\boldsymbol{\mathcal{M}}^{ki}$ countries**, $\boldsymbol{LR_{ki}}$, is given by

$$LR_{ki} = \frac{X_{ki}}{v_{ki}} = \frac{\sum\limits_{j \in \boldsymbol{\mathcal{M}}^{ki}} X_{kij}}{\sum\limits_{j \in \boldsymbol{\mathcal{M}}^{ki}} v_{kij}}, \quad \text{for } k = 1,..,K,\ i = 1,..,P^k. \tag{2.27}$$

From this, we can now calculate the total loss ratio for the k-th client.

Definition 2.12: Total Loss Ratio

It holds that for all P^k treaties and $\boldsymbol{\mathcal{M}}^{ki}$ countries - the total loss ratio, denoted by $\boldsymbol{LR_k}$, is given

by

$$LR_k = \frac{X_k}{v_k} = \frac{\sum_{i=1}^{P^k} X_{ki}}{\sum_{i=1}^{P^k} v_{ki}}, \quad \text{for } k = 1, .., K. \tag{2.28}$$

Finally, this implies, **the cumulative loss ratio**, denoted by $\boldsymbol{LR}$, for all $\boldsymbol{K}$ clients is aggregated as

$$LR = \frac{X}{v} = \frac{\sum_{k=1}^{K} X_k}{\sum_{k=1}^{K} v_k}. \tag{2.29}$$

This implies, for instance, based on the expectation of the cumulative loss counts, λ, in Equation (2.20), and the expected cumulative loss amount - from Equation (2.25) - we can express ***the expected cumulative loss ratio*** for all K clients as

$$\mathrm{E}[LR] = \mathrm{E}\Big(\frac{X}{v}\Big) = \frac{\mathrm{E}(X)}{v} = \frac{\lambda\mu}{v}, \tag{2.30}$$

for fixed $v > 0$ and assuming frequency-severity independence - given the number of total loss events N and the total loss amount X for client k are independent.

Table 2.3. Summary of all defined risk measures and variables - with their respective notation and description per risk class level, (k, i, j, t).

Variable	Description	Formula
$\boldsymbol{k}$	Denotes the k-th client. Client in this study refers to a primary insurer.	$k = 1, \ldots, K$
$\boldsymbol{i}$	Index for i-th treaty belonging to the k-th client. Treaty in this study refers to a policy between the client and the reinsurer.	$i = 1, ..., P^k$
$\boldsymbol{j}$	Country index for j within i-th treaty based on the k-th client. This study only focuses on countries located in the Caribbean.	$j \in \boldsymbol{\mathcal{M}}^{ki}$
$\boldsymbol{t}$	Index for treaty year t risk class. Treaty year is fixed to one year and is based on the underwriting year.	$t \in \boldsymbol{\mathcal{T}}^{kij}$
$\boldsymbol{X_{kijt}}$	Random loss amounts for l observed losses - belonging to k-th client per i treaty, j country, and treaty year t.	$X_{kijt} = \sum_{l \in \boldsymbol{\mathcal{L}}^{kijt}} X_{kijtl}$
$\boldsymbol{X_{kij}}$	The aggregated amount of random loss incurred for all treaty years $\boldsymbol{\mathcal{T}}^{kij}$ within i-th treaty and a country j for the k-th client.	$X_{kij} = \sum_{t \in \boldsymbol{\mathcal{T}}^{kij}} X_{kijt}$
$\boldsymbol{X_{ki}}$	Aggregated Random Loss - for events that occurred in all $\boldsymbol{\mathcal{T}}^{kij}$,treaty years, over countries $\boldsymbol{\mathcal{M}}^{ki}$ in i-th treaty for k client.	$X_{ki} = \sum_{j \in \boldsymbol{\mathcal{M}}^{ki}} X_{kij}$

Table 2.3. Summary of all defined risk measures and variables - with their respective notation and description per risk class level, (k, i, j, t).

Variable	Description	Formula
$\boldsymbol{X_k}$	Total Random Loss - that occurred in all P^k treaties and countries, during treaty years $\boldsymbol{\mathcal{T}}^{kij}$, for a client k.	$X_k = \sum_{i=1}^{P^k} X_{ki}$
$\boldsymbol{X}$	Cumulative Random Losses - aggregated over K clients. From the reinsurer's point of view.	$X = \sum_{k=1}^{K} X_k$
$\boldsymbol{N_{kijt}}$	Random counts of loss events that occurred in all $t \in \boldsymbol{\mathcal{T}}^{kij}$ for a client k, in j country based on i-th treaty.	$N_{kijt} = \sum_{l \in \mathcal{L}^{kijt}} I_{kijtl}$
$\boldsymbol{N_{kij}}$	Random counts of loss events that occurred in all $\boldsymbol{\mathcal{T}}^{kij}$, $\mathcal{M}^{ki}$ in treaty i for client k.	$N_{kij} = \sum_{t \in \boldsymbol{\mathcal{T}}^{kij}} N_{kijt}$
$\boldsymbol{N_{ki}}$	Random counts of loss events that occurred in all $\boldsymbol{\mathcal{T}}^{kij}$ and $\mathcal{M}^{ki}$ for a treaty i, based on client k.	$N_{ki} = \sum_{j \in \boldsymbol{\mathcal{M}}^{ki}} N_{kij}$
$\boldsymbol{N_k}$	Total random loss counts for K clients, P^k treaties and $\mathcal{M}^{ki}$ countries.	$N_k = \sum_{i=1}^{P^k} N_{ki}$
$\boldsymbol{N}$	Cumulative Random Losses for aggregated over K clients, for the reinsurer.	$N = \sum_{k=1}^{K} N_k$
$\boldsymbol{S_{kijt}}$	Loss Severity or the average loss amount for client k, per i treaty, j country, and treaty year t.	$S_{kijt} = \frac{\sum_{l \in \boldsymbol{\mathcal{L}}^{kijtl}} X_{kijtl}}{\sum_{l \in \boldsymbol{\mathcal{L}}^{kijt}} I_{kijtl}}$
$\boldsymbol{S_{kij}}$	Loss Severity incurred for all treaty years $\mathcal{T}^{kij}$ within treaty i and a j country.	$S_{kij} = \frac{X_{kij}}{N_{kij}}$
$\boldsymbol{S_{ki}}$	Loss Severity of the over all $i = 1, \ldots, P_k$ average random loss for a client k.	$S_{ki} = \frac{X_{ki}}{N_{ki}}$
$\boldsymbol{S_k}$	Total Loss Severity of the total random loss for a client k, over all P^k treaties.	$S_k = \frac{X_k}{N_k}$
$\boldsymbol{S}$	Cumulative Loss Severity for all clients K (reinsurance view)	$S = \frac{X}{N}$
$\boldsymbol{v_{kijt}}$	Exposure volume (deterministic) for client k, per i treaty, j country, and treaty year t. Exposure volume units can be premiums or sum insured.	$v_{kijt} = \sum_{l \in \boldsymbol{\mathcal{L}}^{kij}} v_{kijtl}$
$\boldsymbol{v_{kij}}$	Sum of exposure volume units for all treaty years $\mathcal{T}^{kij}$ within treaty i and a j country.	$v_{kij} = \sum_{t \in \boldsymbol{\mathcal{T}}^{kij}} v_{kijt}$
$\boldsymbol{v_{ki}}$	Exposure Volume for all $\mathcal{M}^{ki}$ countries and $\mathcal{T}^{kij}$ treaty years - given a client k, treaty i.	$v_{ki} = \sum_{j \in \boldsymbol{\mathcal{M}}^{ki}} v_{kij}$
$\boldsymbol{v_k}$	The Total Volume of Exposure for a client k. Aggregated over all P^k treaties	$v_k = \sum_{i=1}^{P^k} v_{ki}$,

Table 2.3. Summary of all defined risk measures and variables - with their respective notation and description per risk class level, (k, i, j, t).

Variable	Description	Formula
$\boldsymbol{v}$	The Cumulative Volume of Exposure - aggregated over all K clients.	$v = \sum_{k=1} v_k$
$\boldsymbol{N^v_{kijt}}$	Standardize Loss Frequency that occurred in all $t \in \mathcal{T}^{kij}$ for a client k, in country j of treaty i, given exposure volume v_{kijt}	$N^v_{kijt} = \sum_{l \in \mathcal{L}^{kijt}} \frac{I_{kijtl}}{v_{kijtl}}$
$\boldsymbol{N^v_{kij}}$	Standardized Random counts of loss events (Loss Frequency) that occurred in all $\mathcal{T}^{kij}, \mathcal{M}^{ki}$ in treaty i for client k, given exposure v_{kij}.	$N^v_{kij} = \frac{N_{kij}}{v_{kij}}$
$\boldsymbol{N^v_{ki}}$	Standardized Loss Frequency that occurred in all countries $\mathcal{M}^{ki}$ for a treaty i, based on client k, given exposure v_{ki}.	$N^v_{ki} = \frac{N_{ki}}{v_{ki}}$
$\boldsymbol{N^v_{k}}$	Total Standardized Loss Frequency, aggregated over all P^k treaties, given exposure v_k.	$N^v_{k} = \frac{N_k}{v_k}$
$\boldsymbol{N^v}$	Cumulative Standardized Loss Frequency, aggregated over all K clients, given total exposure v.	$N^v = \frac{N}{v}$
$\boldsymbol{LR_{kijt}}$	Loss Ratio for client k, per i treaty, j country, and treaty year t. The loss ratio is given by the loss amount divided by the exposure unit v_{kijt} per risk class.	$LR_{kijt} = \frac{X_{kijt}}{v_{kijt}}$
$\boldsymbol{LR_{kij}}$	Loss Ratio for based on loss in all treaty years within treaty i for country j per client k.	$LR_{kij} = \frac{X_{kij}}{v_{kij}}$
$\boldsymbol{LR_{ki}}$	Loss Ratio based on all treaty years $\mathcal{T}^{kij}$, and countries $\mathcal{M}^{ki}$ in i-th treaty for client k	$LR_{ki} = \frac{X_{ki}}{v_{ki}}$
$\boldsymbol{LR_{k}}$	Total Loss Ratio per k-th client - over all P^k treaties.	$LR_{k} = \frac{X_k}{v_k}$
$\boldsymbol{LR}$	Loss Ratio which occurred for all K clients	$LR_k = \frac{X}{v}$

3 Statistical Models

This chapter outlines the theory behind the regression models selected for this study. We start at simple regression methods by introducing the simplest and classical **Linear Models (LMs)**. Then, we first extend the LM framework to the **Linear Mixed Models (LMMs)**, and later to **Generalized Linear Models (GLMs)** framework. Thus, before we cover the framework of GLMs - we now provide a brief overview of the LMs and LMMs frameworks.

3.1 Linear Models (LMs)

First, we briefly review some of the main principles of the classical linear models (prior to discussing LMMs and GLMs as an extension of the LMs). This includes the outlying the general formulation of LMs and defining the approaches used in this to estimate the model regression parameters.

3.1.1 General Formulation of LMs

Here, let i $(i = 1, \ldots, n)$ denote the i-th observation and p denote the number of parameters included in the model. Then for the i-th observation, the LM is of the form

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{3.1}$$

where the LM describes the response Y_i as a linear combination of the covariates $x_{i1} \ldots x_{ip}$. The model parameters include the intercept β_0 with p parameters $\beta_1, \ldots \beta_p$, and the random error term given by ε_i. This means, including the intercept, in total we have $p + 1$ regression parameters denoted by $m = p + 1$. Recall that the standard assumptions of the classical LM include the following four regression assumptions: *linearity, independence, variance homogeneity and normality.*

Assumptions 3.1: Linear Model

(A1) Linearity. The response variable Y_i is expressed by a linear combination of the covariates $x_{i1} \ldots x_{ip}$, and includes the error random variable ε_i with mean 0.

(A2) Independence. The errors ε_i are independent random variables.

(A3) Variance Homogeneity. The error terms ε_i are random variables with constant variance,

$$Var(Y_i) = Var(\varepsilon_i) = \sigma^2.$$

(A3) Normality. The errors are normally distributed random variables.

For the error terms, from Assumptions (A2), (A3), and (A4), the random variables $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically normally distributed with $\mathrm{E}(\varepsilon_i) = 0$ and $\mathrm{Var}(\varepsilon_i) = \sigma^2$, i.e.

$$\varepsilon_i \sim N(0, \sigma^2).$$

Then based on the linearity assumption (A1) of Y_i, we can now express ***the expectation of the response variable*** Y_i as a linear function of unknown regression parameter,

$$\mathrm{E}(Y_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip},$$

for n observations. Here we let $\boldsymbol{x_i} \in \mathbb{R}^m$ denote the (row) vector of covariates, $\boldsymbol{x_i} = (1, xi1, \ldots, x_{ip})$ for the i-th observation. Whereas, the (column) vector of the regression coefficients is given by $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)^\top \in \mathbb{R}^m$. Then, based on the assumptions for the error terms ε_i, we have that for n observations the random variables $Y_1, \ldots, Y_n$ are also independent and normally distributed, i.e. $Y \sim (\boldsymbol{x_i}\boldsymbol{\beta}, \sigma)$ - with the conditional expectation given the covariate vector $\boldsymbol{x_i}$,

$$\mathrm{E}(Y_i|\boldsymbol{x_i}) = \mathrm{E}(\boldsymbol{x_i}\boldsymbol{\beta} + \varepsilon_i|\boldsymbol{x_i}) = \boldsymbol{x_i}\boldsymbol{\beta} + \mathrm{E}(\varepsilon_i) = \boldsymbol{x_i}\boldsymbol{\beta}, \tag{3.2}$$

while the variance is given by

$$\mathrm{Var}(Y_i|\boldsymbol{x_i}) = Var(\boldsymbol{x_i}\boldsymbol{\beta} + \varepsilon_i|\boldsymbol{x_i}) = \mathrm{Var}(\varepsilon_i) = \sigma^2. \tag{3.3}$$

Now, we rewrite the LM from Equation (3.1) *in matrix form* - to later compare the linear mixed model framework in the next section.

Definition 3.1: Linear Model in Matrix Form

For n observations, the classical LM in vector notation is given by

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \\ \boldsymbol{\varepsilon} &\sim \mathcal{N}_n(\boldsymbol{0}, \sigma^2 I_n), \end{aligned} \tag{3.4}$$

where

- $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top \in \mathbb{R}^n$ is the response vector, given n observations,
- $\boldsymbol{X} \in \mathbb{R}^{n \times m}$ is the $n \times m$ model design matrix ,
- $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)^\top \in \mathbb{R}^m$ is the vector of regression coefficients, and
- $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top \in \mathbb{R}^n$ is the vector of random error terms, such that $\boldsymbol{\varepsilon} \sim N_n(\boldsymbol{0}, \sigma^2 I_n)$.

Note, here N_n represents the the n-variable multivariate normal distribution with I_n, the $n \times n$ identity matrix, and the $n \times 1$ vector of 0s given by $\boldsymbol{0}$. Hence, using this notation, we can now outline the parameter estimation for LMs.

3.1.2 Parameter Estimation - LMs

The standard approaches for estimating the parameters in LMs include: the **least squares estimation (LSE)** method, and the **maximum likelihood estimation (MLE)** method. Due to the scope of this study, we briefly only discuss the ideas behind these methods for multiple models (to later compare and better understand the techniques used to obtain these estimates - for other models discussed in this study).

Definition 3.2: Least Squares Estimator of the Regression Parameters - in LMs

Suppose the *observations of the response vector* $\boldsymbol{Y}$ from n observations is denoted by the vector $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$. Then, it can be shown that the estimator

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y}, \tag{3.5}$$

is both ***the least square estimator and the maximum likelihood estimator*** of the regression parameters $\boldsymbol{\beta}$.

To find $\hat{\boldsymbol{\beta}}$ using the least squares framework, we do not require any distribution assumptions for the response $\boldsymbol{Y}$. The aim is to find an estimate of $\boldsymbol{\beta}$ - which minimizes the sum of the squared residuals, expressed as

$$\begin{aligned} \boldsymbol{Q}(\boldsymbol{\beta}|\boldsymbol{y}) &:= \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 \\ &= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\ &= \boldsymbol{y}^\top \boldsymbol{y} - 2\boldsymbol{y}^\top \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}, \end{aligned} \tag{3.6}$$

for multiple models with $m = p + 1$ parameters (i.e. we solve for a system with $m = p + 1$ equations, assuming X is a matrix of full rank m). By setting the first derivative of the objective function $\boldsymbol{Q}(\boldsymbol{\beta}|\boldsymbol{y})$ (with respect to $\boldsymbol{\beta}$) to zero and, with rearranging, we find the *normal equations* given by

$$\boldsymbol{X}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}^\top \boldsymbol{y},$$

which means that $\boldsymbol{Q}(\boldsymbol{\beta}|\boldsymbol{y})$ is minimized by

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y}. \tag{3.7}$$

We can now also define the vector of the fitted values as

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y} = \boldsymbol{X}\hat{\boldsymbol{\beta}}, \tag{3.8}$$

where we can define the projection of $\boldsymbol{y}$ onto the column space of $\boldsymbol{X}$, denoted by the matrix $\boldsymbol{H} = \boldsymbol{X}\left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top$ (known as the *"hat matrix"*). Whereas, the ***residual vector***, $\boldsymbol{r}$, based on the LSE, can now be defined as

$$\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}. \tag{3.9}$$

Typically, for GLMs and GLMMs, the regression parameters are estimated using the maximum likelihood. This means that obtaining an analytic form of the MLE typically requires iterative numerical techniques (see Dunn and Smyth (2018) and Klein and Moosbrugger (2000) for detailed descriptions and further discussion). Thus, here we briefly summarize the MLE method for LMs to illustrate the general idea of the MLE method.

Definition 3.3: Log-likelihood (for LMs)

If the normality assumptions of the LM model (in Assumptions 1.1) are satisfied, then the LSE coincides with the MLE. This can be shown by the log-likelihood of

$$L(\boldsymbol{\beta}, \sigma \mid \boldsymbol{y}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2\right\}, \tag{3.10}$$

which is given by

$$\ell(\boldsymbol{\beta} \mid \boldsymbol{y}) = -\frac{n}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2, \tag{3.11}$$

and can also be defined by the objective function given by Equation (3.6).

In other words, if we take the log-likelihood of $(\boldsymbol{\beta}, \sigma)$ for the normal distribution given observations $\boldsymbol{y}$, then by setting its first derivative to zero, we get

$$\begin{aligned} & -2\frac{1}{2\sigma^2}\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = 0, \\ \implies & \boldsymbol{X}^\top\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}^\top\boldsymbol{y}, \\ \implies & \hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\boldsymbol{y}, \end{aligned} \tag{3.12}$$

and

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}.$$

Clearly, these are equivalent to the solutions derived by the normal equation (in Equation (3.6) and Equation (3.8), respectively). This means maximizing the log-likelihood under normality assumption is equivalent to minimizing the sum of the squared residuals.

Additionally, we can also show through MLE, ***the maximum likelihood of the variance***, in terms of the residual vector, is given by

$$\hat{\sigma}^2 := \frac{1}{n}\|\boldsymbol{r}^2\|, \tag{3.13}$$

where, the ***sample variance***, given by

$$s^2 := \frac{1}{n-p}\sum_{i=1}^{n}\left(Y_i - \widehat{Y_i}\right)^2 = \frac{n}{n-p}\hat{\sigma}^2 = \frac{1}{n-p}\|\boldsymbol{r}\|^2, \tag{3.14}$$

is the unbiased estimator of σ^2. Similarly, based on this framework - in addition to using transformation rules (see Gałecki and Burzykowski (2013)) - we can also compute *the expectation and variance of the respective parameter estimators for the LM.*

Remark 3.1: Expectation and Variance of Parameter Estimates (for LMs)

The expectation of the parameter estimators in the LM Model (3.1) can be derived as

$$\begin{aligned} \mathrm{E}(\widehat{\boldsymbol{\beta}}) &= \boldsymbol{\beta}, \\ \mathrm{E}(\widehat{\boldsymbol{Y}}) &= X\boldsymbol{\beta} \\ \mathrm{E}(\boldsymbol{r}) &= \mathbf{0}, \\ \mathrm{E}(\widehat{\sigma}^2) &= \frac{n-p}{n}\sigma^2, \end{aligned} \tag{3.15}$$

while the variances of the corresponding estimators can be calculated by

$$\begin{aligned} \mathrm{Var}(\widehat{\boldsymbol{\beta}}) &= \sigma^2 \left(X^\top X\right)^{-1}, \\ \mathrm{Var}(\widehat{\boldsymbol{Y}}) &= \sigma^2 H, \\ \mathrm{Var}(\boldsymbol{r}) &= \sigma^2 \left(I_n - H\right). \end{aligned} \tag{3.16}$$

From this, for the j-th regression coefficient, we have the following equation.

$$\mathrm{Var}\left(\widehat{\beta}_j\right) = \sigma^2 \left(\left(X^\top X\right)^{-1}\right)_{jj}.$$

and, thus, the estimated standard error of $\hat{\beta}_j$ is given by

$$\widehat{se}\left(\widehat{\beta}_j\right) := s\sqrt{\left(\left(X^\top X\right)^{-1}\right)_{jj}} \tag{3.17}$$

For the distribution of the parameter estimators, under the normality assumptions (given by the LM Assumptions 1.1) we have:

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &\sim N_p\left(\boldsymbol{\beta}, \sigma^2 \left(X^\top X\right)^{-1}\right), \\ \widehat{\boldsymbol{Y}} &\sim N_n\left(X\boldsymbol{\beta}, \sigma^2 H\right), \\ \boldsymbol{r} &\sim N_n\left(0, \sigma^2 \left(I_n - H\right)\right), \end{aligned} \tag{3.18}$$

and

$$\frac{(n-p)s^2}{\sigma^2} = \frac{\sum_{i=1}^n \left(Y_i - \widehat{Y}_i\right)^2}{\sigma^2} \sim \chi^2_{n-p}. \tag{3.19}$$

For more details and further derivations for the LM, see Searle (2006). Whereas, in Section 3.4, we further discuss the goodness of fit and performance measures for model selection used to compare LMs in the study.

It is essential to highlight that the LM is often not suitable in non-life insurance pricing, especially given our case of natural catastrophic events. Mainly since, in linear regression, we assume that the random errors are normally distributed. In our case, the number of losses or loss counts are non-negative integers and assumed to follow a discrete probability distribution. Such that the random loss amounts are non-negative and heavily rightly skewed. Secondly, the assumption of linearity for LMs, implies that the expectation of Y_i is a linear function (see Equation (3.2)). However, given the multiplicative and the hierarchical (or multilevel) structures of the data and risk factors (in our case, grouped by risk class groups) - multiplicative models and multilevel models may be more reasonable (Ohlsson and Johansson, 2010).

3.2 Linear Mixed Models (LMMs)

In this section, we now extend the classical LMs formulation to linear mixed models (LMMs). LMMs allows for **multilevel modelling** with (1) varying intercepts and slopes (with the possibility of using group level predictors). Multilevel models (also known as hierarchical models, random effects or mixed models) can handle data structures with nested or non-nested **clusters or levels** (i.e. hierarchical structure). As longitudinal data is an example of 'clustered data', this is especially useful and applicable in our case - due to the given hierarchy of the risk classes (or levels) - allowing for hierarchy in our model specifications (Antonio and Zhang, 2013).

3.2.1 General formulation of LMMs

Here to illustrate the LMM framework, we consider a two-level hierarchy model, as introduced by Fox and Weisberg (2018) and Antonio and Beirlant (2007). This model framework can also be later extended to account for more hierarchy levels (i.e. for all risk classes included in this study).

Here we consider measurements of the a subject (risk class). Specifically, a ***subject*** is denoted by $\boldsymbol{i}$ for ***total subjects*** $\boldsymbol{N}$, with ***measurements*** (or '*observations*') $\boldsymbol{j = 1, \ldots, n_i}$ belonging to each i-th subject.

Definition 3.1: 2-level Linear Mixed Model (LMM)

For a two-level hierarchy mixed model, suppose we have $i = 1, \ldots, N$ subjects (or groups) for n_i total measurements (or observations), the LMM can be expressed as,

$$\begin{aligned} Y_{ij} =& \beta_0 + \beta_1 x_{1ij} + \cdots + \beta_p x_{pij} + b_{i1} z_{1ij} + \cdots + b_{iq} z_{qij} + \varepsilon_{ij}, \\ b_{ir} \sim& N\left(0, \psi_r^2\right), \quad Cov\left(b_{ir}, b_{ir'}\right) = \psi_{rr'}, \quad r', r = 1, \ldots, q \\ \varepsilon_{ij} \sim& N\left(0, \sigma^2 \lambda_{ijj}\right), \quad Cov\left(\varepsilon_{ij}, \varepsilon_{ij'}\right) = \sigma^2 \lambda_{ijj'}, \end{aligned} \tag{3.20}$$

where

- y_{ij} is the response variable for the values in the i-th subject of N, based on the j-th measurement of n_i,
- $x_{1ij}, \ldots, x_{pij}$ are the known ***(fixed effects)*** covariates for j measurement for subject i,
- $\beta_0, \beta_1, \ldots, \beta_p$ are the unknown ***(fixed effects)*** parameters for all subjects in N (with intercept β_0),
- $b_{i1}, \ldots, b_{iq}$ are the unknown ***(random effects)*** parameters based on the i-th subject and assumed to follow a multivariate normal distribution, with ψ_r^2 variance and mean 0. Here, $\psi_{rr'}$ denotes the *covariance* between the random effects for each i-th subject,
- $z_{1ij}, \ldots, z_{qij}$ are the known ***(random effects)*** covariates for measurement j of subject i, and
- ε_{ij} are the ***random error terms*** for the j-th measurement if subject i, assumed to follow a multivariate distribution, where $\sigma^2 \lambda_{ijj'}$ is the covariance of subject i between the errors ε_{ij} and $\varepsilon_{ij'}$.

Hence, a mixed model includes both ***"fixed"*** and ***"random"*** effects. In this study, *fixed effects* - accounting for qualitative (or "factors") or quantitative variables - represent all levels of interest or the whole population considered, respectively. Meanwhile, the *"random effects"* represent either the qualitative variables with randomly sampled levels from a set of levels of interest or the quantitative variables which measure an individual's deviation from the population (of the fixed effect). For more further discussion on mixed effects modelling for nested data see Zuur et al. (2009).

Next, similar to the LM model formulation in vector notation, we now define the matrix form of the LMM model.

Definition 3.2: 2-level Hierarchical LMMs in Matrix Form (for Longitudinal Data)

$$\begin{aligned} &\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i, \\ &\boldsymbol{\alpha}_i \sim N_q(\mathbf{0}, \boldsymbol{D}), \qquad\quad i.i.d, \quad \forall i = 1, \ldots, N \\ &\boldsymbol{\varepsilon}_i \sim N_{n_i}(\mathbf{0}, \boldsymbol{\Sigma}_i), \qquad i.i.d, \quad \forall i = 1, \ldots, N, \quad j = 1, \ldots, n_i, \end{aligned} \tag{3.21}$$

such that,

- $\boldsymbol{Y}_i \in \mathbb{R}^{n_i \times 1}$ is the vector of response, for measurements (or '*observations*') $j = 1, \ldots, n_i$ of the subject i,
- $\boldsymbol{X}_i \in \mathbb{R}^{n_i \times m}$ is the known **(fixed effects)** design matrix for measurements j of subject i,
- $\boldsymbol{\beta} \in \mathbb{R}^{m \times 1}$ is the unknown **(fixed effects)** vector of coefficients,
- $\boldsymbol{Z}_i \in \mathbb{R}^{n_i \times q}$ is the known **(random effects)** design matrix of coefficients, with based on the measurements in the i-th subject,
- $\boldsymbol{\alpha}_i \in \mathbb{R}^{q \times 1}$ with $\boldsymbol{\alpha}_i \sim N_q(\mathbf{0}, \boldsymbol{D})$ is the unknown **(random effects)** vector per subject i. It is assumed to be normally distributed with mean vector $\mathbf{0}$ and covariance matrix of the random effects $\boldsymbol{D} \in \mathbb{R}^{q \times q}$,
- $\boldsymbol{\varepsilon}_i \in \mathbb{R}^{n_i \times 1}$, with $\boldsymbol{\varepsilon}_i \sim N_{n_i}(\mathbf{0}, \boldsymbol{\Sigma}_i)$, is the vector of residuals based on the number of measurements of i-th subject. Errors are assumed to be normally distributed with covariance matrix $\boldsymbol{\Sigma}_i \in \mathbb{R}^{n_i \times n_i}$ in the i-th subject.

Note, here we assume that the random effects are uncorrelated and independent between the i-th subjects, while $\boldsymbol{\alpha}_i$ and $\boldsymbol{\varepsilon}_i$ are also assumed to be independent. Such that, their covariance matrices $\boldsymbol{D}$ and $\boldsymbol{\Sigma}_i$, respectively, are traditionally assumed to follow a multivariate normal distribution (where different structures of the covariance suitable for longitudinal data are also possible, see Galecki (1994) for further details). In other words, we have

$$\begin{pmatrix} \boldsymbol{\alpha}_i \\ \boldsymbol{\varepsilon}_i \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{D} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_i, \end{pmatrix} \right]. \tag{3.22}$$

Thus, now we can further define the distributional assumptions based on this hierarchical LMM. Since, firstly, this implies $\boldsymbol{Y}_i$ has a combined marginal (unconditional) multivariate normal distribution with mean $\boldsymbol{X}_i\boldsymbol{\beta}$ and (known) covariance matrix $\boldsymbol{V}_i = Var(\boldsymbol{Y}_i)$, given by

$$\boldsymbol{V}_i = \boldsymbol{Z_i D Z_i^\top} + \boldsymbol{\Sigma}_i, \tag{3.23}$$

for all subjects $i = 1, \ldots, N$. Hence, it follows that the marginal distribution of $\boldsymbol{Y_i}$ is

$$\boldsymbol{Y}_i \sim N_{n_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{V}_i), \tag{3.24}$$

given the random effects $\boldsymbol{\alpha}_i$ included in the model, for each subject i. This means we can also calculate the conditional mean and variance of the response as

$$\begin{aligned} \mathrm{E}(\boldsymbol{Y_i}|\boldsymbol{\alpha_i}) &= \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\alpha}_i, \\ Var(\boldsymbol{Y_i}|\boldsymbol{\alpha_i}) &= \boldsymbol{\Sigma}_i, \end{aligned} \tag{3.25}$$

respectively, for a subject i. Together this means the two level hierarchical LMM (for longitudinal data) can be rewritten as

$$\begin{aligned} \boldsymbol{Y}_i|\boldsymbol{\beta}_i &\sim N_{n_i}(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\alpha}_i, \boldsymbol{\Sigma}_i) \\ \boldsymbol{\alpha}_i &\sim \boldsymbol{N}_q(\boldsymbol{0}, \boldsymbol{D}), \end{aligned} \tag{3.26}$$

for all subjects $i = 1, \ldots N$ with corresponding measurements $j = 1, \ldots, n_i$.

This clearly implies that even though the random effects (i.e. the i-th subject-specific effects) describes the covariance structure between the measurements (denoted by $j = 1, \ldots, n_i$ belonging to the i-th subject) - the expectation of the responses $\mathrm{E}(Y_{ij})$ only includes the fixed effects (Fahrmeir et al., 2007). Based on this setup, we can now briefly describe the two estimation methods of LMMs with unknown covariance structures (following the specifications outlined in Antonio and Beirlant (2007)).

3.2.2 Parameter Estimation - LMMs

The two most standard approaches for estimating LMMs with unknown covariance structure include the: *maximum likelihood* **(ML)** and the *restricted maximum likelihood* **(REML)**. Using these methods, our goal is to ultimately find the *the maximum likelihood estimators* **(MLE)** - also known as the minimum variance unbiased estimators - of the fixed effects, $\hat{\boldsymbol{\beta}}$ (based on the marginal distribution of $\boldsymbol{Y}_i$, given in (3.24)), and the random effects $\boldsymbol{\alpha}_i$. This leads to defining the *empirical best linear unbiased predictors* (**EBLUPs**). Here we refer to as the 'best' in terms of the minimal mean squared error (by maximizing the MLE or REML).

First, suppose the variance parameters in $\boldsymbol{V}_i$ (based on Equation (3.30)) are known. Then it can be shown that maximum likelihood estimator (**MLE**) of the fixed effects is

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i^\top \boldsymbol{V}_i^{-1} \boldsymbol{Y}_i. \tag{3.27}$$

for 2-level hierarchy LMM defined in (2.2). Note, under the normality assumptions shown in (3.22), this coincides with the generalized LSE. Whereas, the estimator of the random effects, $\boldsymbol{\alpha}_i$, is given by

$$\hat{\boldsymbol{\alpha}}_i = \boldsymbol{D}\boldsymbol{Z}_i^\top \boldsymbol{V}_i^{-1} \left(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}\right), \tag{3.28}$$

which is often estimated using iterative numerical techniques - such as the iterative generalized least squares methods. For a more detailed description on these methods, we refer to Fahrmeir et al. (2007) and Ng et al. (2019). Here, it can also be proven that the $\hat{\boldsymbol{\alpha}}_i$ is the best linear unbiased predictor (**BLUP**). For instance, if the normality assumptions for Model (3.26) hold then $\boldsymbol{Y}_i$ is based on the

i-th subject - such that, we have

$$\begin{aligned}\text{Cov}\left(\boldsymbol{Y}_i, \boldsymbol{\alpha}_i'\right) &= \text{Cov}\left(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i, \boldsymbol{\alpha}_i'\right) \\ &= \text{Cov}\left(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\alpha}_i'\right) + \boldsymbol{Z}_i\,\text{Var}\left(\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_i'\right) + \text{Cov}\left(\boldsymbol{\varepsilon}_i, \boldsymbol{\alpha}_i'\right) \\ &= \boldsymbol{Z}_i\boldsymbol{D}_i,\end{aligned} \tag{3.29}$$

which leads to the BLUP given in Equation (3.28). It is clear here that the covariance depends on both the unknown fixed effects $\boldsymbol{\beta}$ and on the variance components (accounted for in the covariance structure of $\boldsymbol{V}_i$). However, in practice, it is often the case that the covariance structure is unknown, where instead, it is estimated from the data. In other words, this means that we use the MLE and REML methods to estimate the unknown components in $\boldsymbol{V}_i$ (known as the *the variance components*) from the data, and replace them with their estimates. For this reason, we refer to BLUPs as EBLUP (the estimated or empirical BLUP, respectively). More detailed derivations and further discussion on the parameter estimation for LMMs can be found in Eager and Roy (2017), Zhu and Zou (2014), or Searle (2006).

To describe the unknown covariance structure more clearly, suppose for the covariance matrix $\boldsymbol{V}_i$ the unknown parameters are denoted by $\boldsymbol{\alpha}$ (*the variance components vector*). Then we re-express the individual covariance matrices given in (3.30) as: $\boldsymbol{D}(\boldsymbol{\alpha})$ and $\boldsymbol{\Sigma}_i(\boldsymbol{\alpha})$, for subjects $i = 1, \ldots, N$. This means that, $\boldsymbol{V}_i$ is now expressed as

$$\boldsymbol{V}_i(\boldsymbol{\alpha}) = \boldsymbol{Z}_i\boldsymbol{D}(\boldsymbol{\alpha})\boldsymbol{Z}_i^\top + \boldsymbol{\Sigma}_i(\boldsymbol{\alpha}), \tag{3.30}$$

Under the normality assumptions in Model (3.21), the joint log-likelihood of $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ given the data $\boldsymbol{y}_i$ based on the subject i, is

$$\begin{aligned}\ell(\boldsymbol{\beta}, \boldsymbol{\alpha} \mid \boldsymbol{y}_i) &:= \ln f(\boldsymbol{y}_i \mid \boldsymbol{\beta}, \boldsymbol{V}_i(\boldsymbol{\alpha})) \\ &= -\frac{1}{2}\left\{\ln|\sum_{i=1}^{N}\boldsymbol{V}_i(\boldsymbol{\alpha})| + \sum_{i=1}^{N}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^\top\boldsymbol{V}_i(\boldsymbol{\alpha})^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right\} + c,\end{aligned} \tag{3.31}$$

where c denotes the appropriate constants. Hence, if we maximize the above given joint log-likelihood - with respect to $\boldsymbol{\beta}$ for fixed $\boldsymbol{\alpha}$ - then it can be shown that the MLE is given by $\hat{\boldsymbol{\beta}}$ in Equation (3.27) is now expressed as:

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\sum_{i=1}^{N}\boldsymbol{X}_i^\top\boldsymbol{V}_i(\boldsymbol{\alpha})^{-1}\boldsymbol{X}_i\right)^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i^\top\boldsymbol{V}(\boldsymbol{\alpha})_i^{-1}\boldsymbol{Y}_i. \tag{3.32}$$

Such that, if we replace $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$ in (3.31) then we get the so called *"profile log-likelihood"* (Murphy and Van der Vaart, 2000).

Definition 3.3: Profile Log-Likelihood (for LMMs)

The **profile log-likelihood** for the covariance components vector, denoted by $\boldsymbol{V}_i(\boldsymbol{\alpha})$ is given by

$$\begin{aligned}\ell_P(\boldsymbol{\alpha} \mid \boldsymbol{y}_i) &:= \ell(\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}), \boldsymbol{\alpha} \mid \boldsymbol{y}_i) \\ &\propto -\frac{1}{2}\left\{\sum_{i=1}^{N}\ln|\boldsymbol{V}_i(\boldsymbol{\alpha})| + \sum_{i=1}^{N}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}))^\top\boldsymbol{V}_i(\boldsymbol{\alpha})^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}))\right\},\end{aligned} \tag{3.33}$$

given the observed data $\boldsymbol{y}_i$ for subject $i, i = 1, \ldots, N$.

In practice, LMMs often contain many fixed effects. In these cases, it is important to estimate the (unknown) variance components - involved in estimating the fixed effects $\boldsymbol{\alpha}_i$. Since in this case, an unbiased estimator for the vector of variance components, $\boldsymbol{\alpha}$, cannot be obtained using the profile log-likelihood , ℓ_P (see Verbeke and Molenberghs (2000) for example). Specifically, since we maximize the profile log-likelihood with respect to the variance components $\boldsymbol{\alpha}$ to find its MLE (say, $\boldsymbol{\alpha}_{MLE}$). Hence, for this reason, we use the the *restricted maximum likelihood* **(REML)** - which accounts for the degrees of freedom when estimating the fixed effects $\boldsymbol{\alpha}_i$. Under normality assumptions, the REML is maximized to estimate the variance components vector $\boldsymbol{\alpha}$ using the marginal log-likelihood for the variance components (to find $\boldsymbol{\alpha}_{REML}$). Thus, the log-likelihood REML is given by

$$\ell_{REML}(\boldsymbol{\alpha} \mid \boldsymbol{y}_i) = \ell_P(\boldsymbol{\alpha} \mid \boldsymbol{y}_i) - \frac{1}{2}\sum_{i=1}^{N} \ln \left| \boldsymbol{X}^\top \boldsymbol{V}_i^{-1} \boldsymbol{X}_i \right|, \tag{3.34}$$

where $\ell_P(\boldsymbol{\alpha} \mid \boldsymbol{y}_i)$ is the profile log-likelihood, given by Equation (3.33). Additionally, based on this, the vector of residuals $\boldsymbol{r}_i$ can also be defined as

$$\boldsymbol{r}_i = \boldsymbol{Y}_i - \boldsymbol{X}_i \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{V}_i \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{V}_i^{-1} \boldsymbol{Y}_i \right).$$

The log-likelihood of the REML and the profile log-likelihood are typically maximized by utilizing iterative numerical techniques such as the Newton-Raphson and Fisher scoring (Antonio and Beirlant, 2007). Additionally, here the unknown parameters $\boldsymbol{\alpha}$ are then replaced by $\boldsymbol{\alpha}_{ML}$ and $\boldsymbol{\alpha}_{REML}$ in Equation (3.33) and Equation (3.34), respectively. This leads to a closed-form expression of the MLE such that the EBLUP for the random effects $\boldsymbol{\beta}$ is given by Equation (3.27). Here we predict the random effects using the mean of the posterior distribution of the random effects given the data (i.e. $\boldsymbol{\alpha}_i|\boldsymbol{y_i}$). For more information regarding using the ML and REML methods to estimate the EBLUP and the variance components - see Frees et al. (2014).

Essentially, in our study, these random effects will represent hidden (unseen) characteristics at each risk class level. By incorporating random effects in the structure for the mean, LMMs allows us to incorporate the hierarchical structure used for this study (discussed in Chapter 2). Specifically, since *mixed-effects models* (or simply *'mixed models'*) incorporates additional random effect terms associated with the components of the variance and covariance. For this reason, LMMs can adequately represent the groups of subjects with repeated measurements collected over time in longitudinal data. Specifically helpful in our case since our data contains risk measurements (or simply measures) or factors collected over time hierarchically for the same group of risk class levels, k, i, j, t (same subjects).

3.3 Generalized Linear Models (GLMs)

Now, we outline and provide a brief summary of the framework of the GLM, including its main characteristics. This allows us to compare the GLM framework with the classical LM model, and then later to understand how GLMM framework extends these model types and properties.

3.3.1 General Formulation of GLMs

As previously stated, Generalized Linear Models (GLMs) - introduced by McCullagh and Nelder (1989) - extends the model framework of the LMs, shown in Model (1.1). Specifically, it extends the class of

normal distributions to the class of other distributions belonging to the *exponential family*; such as the Poisson, Binomial or the Gamma distribution. This allows us to model a variety of possible response or outcome measures - such as counts, binary or skewed data (especially crucial in risk modelling). In other words, GLMs assumes the distribution belongs to *the exponential (dispersion) family.*

Definition 3.1: Exponential Family

Let the responses $Y_1, \ldots, Y_n$ be independent random variables (the **random component**) with a probability density function from the **exponential family**, and of the form

$$f(y \mid \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\}, \tag{3.35}$$

where

- $\phi > 0$ is the **dispersion parameter**,
- $\theta \in$ is the (unknown) parameter called the **canonical parameter** of the distribution,
- $a(\cdot)$ is a (known) function a which allows for incorporating weights in the distribution - often known as the dispersion function $a(\phi)$,
- $b(\cdot)$ and $c(\cdot)$ are (known) specific functions for the given distribution.

Typically, for $a(\phi)$ we have $a(\phi) = \phi$ or $a(\phi) = \phi/w_i$ for a known or apriori weight w_i. Proofs regarding the derivations of the mean and variance expressions - based on the log-likelihood (with respect to an exponential family distribution) can be found in Nelder and Verrall (1997).

Given these properties of the exponential family - we can also determine the moments of an exponential family response (random variance) Y_i, such as the estimated mean and variance of Y_i.

Lemma 3.1: Moments of an Exponential Family Distribution

Let $Y_1, \ldots, Y_n$ be the independent random variables with an exponential family distribution (of the form (3.1)). Then, it can be shown that the first moment, i.e. **the expectation of response** Y_i is given by

$$\mathrm{E}(Y_i) = \mu_i = b'(\theta_i),$$

and **the variance of the response** Y_i is given by,

$$\mathrm{Var}(Y_i) = b''(\theta_i)a(\phi) = V(\mu_i)a(\phi),$$

where $V(\cdot)$ is called the **variance function** for an exponential family distribution, such that $V(\mu_i) := b''(\theta_i)$.

Based on this we can further define the three components of the GLM: the *random, systematic* and *link component.*

Here we will briefly outline the assumptions corresponding to this three-part GLM specification, introduced by McCullagh and Nelder (1983).

Assumptions 3.1: Components of the GLM

1. ***The Random Component.*** The observation y_i is assumed to be a realization of the random variable Y_i (where we let i index the observations) - with n components - which are independently distributed (by an exponential family distribution) with mean μ_i.
2. ***The Systematic Component.*** The corresponding covariates $\boldsymbol{x}_i = (1, x_{i1}, \dots, x_{ip})^\top$ produce a *linear predictor* η_i given by:

$$\eta_i(\boldsymbol{\beta}) = \boldsymbol{x}_i\boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}, \tag{3.36}$$

   where $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)$ is the vector of the $m = p + 1$ unknown regression parameters .
3. ***The Parametric Link Component.***] The *link* between the random and systematic components - is described by a monotonic differentiable *link function* $g(\cdot)$ such that

$$g(\mu_i) = \eta_i(\boldsymbol{\beta}) = \boldsymbol{x}_i\boldsymbol{\beta}. \tag{3.37}$$

   In other words, it defines the relationship between the mean μ_i of the random variable Y_i and the linear predictor η_i.

Essentially, this implies that the link function $g(\cdot)$ transforms the expectation of the response variable $E(Y_i) = \mu_i$ to the linear predictor η_i. Hence, based on the chosen monotonic link function $g(\cdot)$, this allows for interpretation - i.e. the relationship between the response and covariates in the GLM can be linearly described - dependent on the linear predictor η_i. This is clear and can be shown - given that the link function is invertible - we have

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(\boldsymbol{x}_i\beta), \tag{3.38}$$

which is known as the *inverse mean (link) function*. Thus, the GLM can be thought as a classical LM, given by Model (3.1), for a transformation of the expected response (or as a nonlinear regression model for the response) (Fox, 2015). In fact, based on this formulation, the classical LM is a special case of a GLM - where the random component (first component of the GLM) has a normal distribution with a link function $g(\cdot)$ (third component) known as the *identity link* function, $g(\mu_i) = \mu_i$. Here the inverse mean function returns the linear predictor (unaltered), such that $\eta_i = g(\mu_i) = \mu_i$ where $\mu_i = g^{-1}(\eta_i) = (\eta_i)$.

Additionally, in the case that $g(\mu_i) = \theta_i = \boldsymbol{x}_i\beta$ - where θ_i is the canonical parameter in Equation (3.1) - is known as the *canonical link*. Essentially, though the canonical link (or "natural" link) function simplifies the GLM, it may not provide an appropriate fit for the given data. As it poses a restriction on the link function, and, hence, on the range of the expected response. For this reason, we consider other link functions - as an appropriate or suitable choice of link will remove these restrictions. Other choices of common link functions with their inverse mean functions are given in Table 3.1 (see Fox (2015) for further discussions).

Examples of the GLM components and characteristics of common univariate distributions - in the exponential family - is outlined in Table 3.2. See McCullagh and Nelder (1983) for more details on their full derivations.

Link Name	**Link Function:** $\eta_i = g(\mu_i)$	**Inverse mean function:** $\mu_i = g^{-1}(\eta_i)$
Identity	μ_i	η_i
Log	$\ln(\mu_i)$	$\exp(\eta_i)$
Inverse	μ_i^{-1}	η_i^{-1}
Inverse-square	μ_i^{-2}	$\eta_i^{-1/2}$
Square-root	$\sqrt{\mu_i}$	η_i^2
Logit	$\ln(\mu_i/(1-\mu_i))$	$\exp(\eta_i)/(1+\exp(\eta_i))$
Probit	$\Phi^{-1}(\mu_i)$	$\Phi(\eta_i)$
Log-log	$-\ln(-\ln(\mu_i))$	$\exp(-\exp(-\eta_i))$
Complementary log-log	$-\ln(-\ln(1-\mu_i))$	$1-\exp(-\exp(\eta_i))$

Table 3.1. Examples of common link functions and their inverses - used in generalized linear models.

	Normal	**Poisson**	**Binomial**	**Gamma**	**Inverse Gaussian**
Notation (Distribution)	$N(\mu, \sigma^2)$	$P(\mu)$	$B(m,\pi)/m$	$G(\mu,\nu)$	$IG(\mu,\sigma^2)$
Range of y	$(-\infty,\infty)$	$0,1,\ldots,\infty$	$\frac{\{0,1,\ldots,m\}}{m}$	$(0,\infty)$	$(0,\infty)$
Dispersion parameter: ϕ	$\phi=\sigma^2$	1	$1/m$	$\phi=\nu^{-1}$	$\phi=\sigma^2$
Cumulant function: $b(\theta)$	$\theta^2/2$	$\exp(\theta)$	$\ln\left(1+e^\theta\right)$	$-\ln(-\theta)$	$-(-2\theta)^{1/2}$
$c(y,\phi)$	$-\frac{1}{2}\left(\frac{y^2}{\phi}+\ln(2\pi\phi)\right)$	$-\ln(y)!$	$\ln\binom{m}{my}$	$\nu\ln(\nu y)-\ln(y)$ - $\ln\Gamma(\nu)$	$-\frac{1}{2}\left\{\ln\left(2\pi\phi y^3\right)+\frac{1}{\phi y}\right\}$
$\mu(\theta)=E(Y;\theta)$	θ	$\exp(\theta)$	$e^\theta/\left(1+e^\theta\right)$	$-1/\theta$	$(-2\theta)^{-1/2}$
Canonical link: $g(\mu)=\theta$	identity	log	logit	reciprocal	$1/\mu^2$
Variance function: $V(\mu)$	1	μ	$\mu(1-\mu)$	μ^2	μ^3

Table 3.2. Example of GLM Characteristics of exponential family for common distribution.

3.3.2 Parameter Estimation - GLMs

Similar to the LMMs and LMs, we estimate the unknown regression parameters β in GLMs through the maximum likelihood (ML) method as well. By maximizing the log-likelihood $\ell(\boldsymbol{\beta}, \phi \mid \boldsymbol{y})$ function to find an estimate for $\boldsymbol{\beta}$ based on the observed data, with respect to the exponential family.

Definition 3.2: Log-Likelihood Functions (for GLMs)

Suppose the observed data of the response $\boldsymbol{Y}$ is denoted by $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ for $i = 1, \ldots, n$ independent observations. Then the log-likelihood in a GLM directly results from Equation (3.1), and defined as

$$\begin{aligned}
\ell(\boldsymbol{\beta}) := \ell(\boldsymbol{\beta}, \phi \mid \boldsymbol{y}) &= \sum_{i=1}^{n} \ln\left(f\left(y_i \mid \theta_i, \phi\right)\right) \\
&= \ln\left\{\exp\left(\frac{y_i\theta_i - b\left(\theta_i\right)}{a(\phi)} + c\left(y_i, \phi\right)\right)\right\} \\
&= \sum_{i=1}^{n} \frac{y_i\theta_i - b\left(\theta_i\right)}{a(\phi)} + \sum_{i=1}^{n} c\left(y_i, \phi\right) \\
&= \sum_{i=1}^{n} \ell_i(\mu_i, \phi \mid y_i) := \ell(\boldsymbol{\mu}, \sigma|\boldsymbol{y}).
\end{aligned} \tag{3.39}$$

For simplicity, we also denote the log likelihood function for an individual observation y_i by: $\ell_i := \ell_i(\mu_i, \phi \mid y_i)$. To find the maximum of the log-likelihood in Equation (3.39) with respect to $\boldsymbol{\beta}$ - we use gradient descent methods to differentiate $\ell(\boldsymbol{\beta})$ with respect to all β_j and solve for

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_j} = 0, \quad \text{for } j = 0, \ldots, p. \tag{3.40}$$

Hence, the partial derivatives (by the chain rule for differentiation) is given by

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}.$$

Then based on the properties of the exponential families (see Definition 3.1) and using it's relations given in Lemma 3.1, we can formulate each expression on the right side as,

$$\begin{aligned}
&\frac{\partial \ell_i}{\partial \theta_i} = \frac{y_i - b'\left(\theta_i\right)}{a(\phi)} = \frac{y_i - \mu_i}{a(\phi)}, \quad \text{since } b'(\theta_i) = \mu_i \\
&\frac{\partial \mu_i}{\partial \theta_i} = \frac{b''(\theta_i)a(\phi)}{a(\phi)} = b''\left(\theta_i\right), \quad \text{thus} \quad \frac{\partial \theta_i}{\partial \mu_i} = \frac{a(\phi)}{b''(\theta)a(\phi)} = \frac{1}{b''(\theta)}, \\
&\frac{\partial \eta_i}{\partial \beta_j} = x_{ij}.
\end{aligned}$$

Thus, from this, it can be shown that we can now re-express Equation (3.40), it can be shown

$$\frac{\partial \ell(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} \frac{y_i - \mu_i}{a(\phi)} \frac{1}{b''(\theta_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij} = \sum_{i=1}^{n} \frac{y_i - \mu_i}{a(\phi)b''(\theta_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij},$$

As previously stated, often the dispersion parameter is specified as $a(\phi) = \phi$ or $a(\phi) = \phi/w_i$ for known

weight w_i. In this study, we focus on the latter - specifically for grouped data we consider $a(\phi) = \phi/w_i$ where w_i is the exposure. Whereas, for loss counts we often use offsets instead (discussed further in the next section). Formally, the **weights in GLMs** can be defined by

$$w_i := w_i(\boldsymbol{\beta}) := \left(\frac{d\mu_i}{d\eta_i}\right)^2 / b''(\theta_i)\,. \tag{3.41}$$

Whereas, for this study, we can incorporate the weights in the log-likelihood of the GLM, and re-define it as:

$$\frac{\partial \ell(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} w_i \frac{y_i - \mu_i}{\phi V(\mu_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij}, \tag{3.42}$$

where, $b''(\theta_i) = V(\mu_i)$ (from Equation (3.1)) with (weighted) dispersion parameter $a(\phi) = \phi/w_i$.

Since there is no general close form solution for the MLE of $\boldsymbol{\beta}$, in Equation (3.42) - iterative numerical techniques are used to find the estimate $\boldsymbol{\beta}$ (such as the Newtorn-Raphson method). Due to the scope of this study, we briefly state the steps of the *iteratively re-weighted least squares* **(IRLS)** estimation algorithm - see Algorithm 1. For more explanation and examples see McCullagh and Nelder (1983), Nelder and Verrall (1997) or Ng et al. (2019).

Algorithm 1 Iterative weighted least squares algorithm for GLMs to estimate $\boldsymbol{\beta}$

Input: data set $\boldsymbol{x}_i$ for observations $i = 1, \ldots, n$
Initialize: start values of $\boldsymbol{\beta}_0$ and $\varepsilon > 0$
while $\left\|\boldsymbol{\beta}^r - \boldsymbol{\beta}^{r+1}\right\| > \varepsilon$ **do**
for $i = 1, \ldots, n$, and $r \geq 0$

1. $\eta_i^r := \boldsymbol{x}_i^\top \beta^r \quad i = 1, \ldots, n$ ▷ (current linear predictors)
2 : $\mu_i^r := g^{-1}(\eta_i^r)$ ▷ (current fitted means)
3: $\theta_i^r := h(\mu_i^r)$ ▷ (current canonical parameters)
4: $Z_i^r := \eta_i^r + (y_i - \mu_i^r)\frac{\partial \mu_i^r}{\partial \eta_i^r}$ ▷ (adjusted dependent variables)
5: $w_i^r := \left(\frac{d\mu_i^r}{d\eta_i^r}\right)^2 / b''(\theta_i)$ ▷ (current weights)
6. regress Z_i^r on $\boldsymbol{x}_i$ with weights $1/w_i^r$

return: $\boldsymbol{\beta}^{r+1}$
Output: maximum likelihood estimate of $\boldsymbol{\beta}$, denoted $\hat{\boldsymbol{\beta}}$

3.4 Model Selection and Comparison

In this section, we discuss the different approaches and the combinations of the metrics we use to compare the models and assess accuracy of the models. To compare or assess the three different classes of models: LMs, LMMs and GLMs, we use standard statistical approaches widely used in actuarial and insurance-related studies, suitable for each model class.

Our goal here is to essentially address three *major questions*:

- How can we assess ***the goodness of fit*** of the models (per model class) and ***assess the performance*** in terms of the overall model fit?
- How can we ensure we select models with ***high predictive power while avoiding overfitting*** (i.e. to achieve optimal model complexity)?
- Given a subset of possible models, how can we *compare* these models (across all model classes) to select the model that ***best represents the underlying data distribution - with high prediction accuracy?***

To answer these questions, we categorized our model assessment and comparison methods into the three following categories: (1) measures of fit and variance, (2) model and variable selection and (3) comparison by predictive accuracy.

3.4.1 Measures of Fit for LMs

In this section, we start by first briefly introduce fundamental analysis of variance formulas, based on the empirical variance of the responses, required to formulate the goodness of fit measure (for LMs). In this study, we use the multiple coefficient of determination (R^2) to assess the goodness of fit and measure the overall fit of LMs.

Goodness-of-Fit Measures for LMs

The estimated R^2 of the fitted LM represents the proportion of variability in the response explained by the linear regression model.

Sums of Squares. If the assumptions of the linear regression are fulfilled, we can define the different sum of squares - required to quantify the amount of the variation explained by the regression.

By first considering the following additive decomposition formula:

$$\sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2 = \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2 + \sum_{i=1}^{n} \hat{\varepsilon}_i^2, \tag{3.43}$$

where

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i.$$

Following this, we can now define the following three sum of squares measures relevant for this study.

Remark 3.1: Analysis of Variance Formulas (for LMs)

The analysis of variance formulas, for LMs, in this study can be formulated as:

SST $:= \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2$, ***(total sum of squares)***
SSR $:= \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2$, ***(regression sum of squares)***
SSE $:= \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n} \hat{\varepsilon}_i^2$, ***(error sum of squares)***

From these definitions and given the decomposition in Equation (3.43) follows from

$$\sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)\left(Y_i - \hat{Y}_i\right) = 0,$$

where we have:

$$\mathbf{SST = SSR + SSE}. \tag{3.44}$$

Coefficient of Determination - LMs

As previously stated, the R^2 statistic is used to assess how well the LM predicts the response $\boldsymbol{Y}$ by measuring how much the models account for variability in the response. This study looks at two types of R^2 quantities as a goodness of fit measure for LMs. Since we aim to compare several multiple linear regression models, and given that including more covariates to an LM always increases R^2 - we also consider the adjusted coefficient of determination $R^2_{adj.}$ (which takes into account the number of regression parameters in the model).

Hence, both goodness of measure can be defined by in terms of the sum of squares and following Equation (3.43).

Definition 3.1: Multiple Coefficient of Determination

The **multiple coefficient of determination** of a LM is defined as,

$$\mathrm{R}^2 := \frac{\mathrm{SSR}}{\mathrm{SST}} = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}}. \tag{3.45}$$

The **adjusted multiple coefficient of determination** $R^2_{adj.}$ is given by

$$\mathrm{R}^2_{adj} := 1 - \frac{n-1}{n-p}\left(1 - R^2\right) = 1 - \frac{\mathrm{SSE}/(n-p)}{\mathrm{SST}/(n-1)}, \tag{3.46}$$

such that both measures lies within the range of zero to one.

Hence, based on the definitions, the closer the estimated multiple coefficient of determination of a model is to one - the better the overall fit and the model's ability to account for variability in the response.

Hypothesis Testing and Inference for LMs

In this section we briefly outline the statistical inference we use to draw conclusions about our parameters in our LM. To test any linear hypothesis about the (unknown) regression coefficients $\boldsymbol{\beta}$, this can be defined generally by the following *general linear hypothesis.*

Definition 3.2: General Linear Hypothesis

The testing problem for ***general linear hypothesis testing*** is formulated as

$$H_0 : C\boldsymbol{\beta} = \boldsymbol{d} \qquad \text{versus} \qquad H_0 : C\boldsymbol{\beta} \neq \boldsymbol{d},$$

such that

- $C \in \mathbb{R}^{r \times m}$ is a $r \times m$ matrix of known elements, where r is the number of linear restrictions to be tested ($r \leq m$), for $m = p + 1$ regression coefficients,
- $\boldsymbol{d} \in \mathbb{R}^{r \times 1}$ is a vector of known elements, such that $rank(C) = r$

Hence, for LMs, the test statistic, using the distributional assumptions of the estimated regression coefficients (in 3.19). Under the null hypothesis, H_0 we have

$$\widehat{\boldsymbol{\alpha}} = C\widehat{\boldsymbol{\beta}} - \boldsymbol{d} \overset{H_0}{\sim} N_r\left(0, \sigma^2 C\left(X^\top X\right)^{-1} C^\top\right),$$

such that

$$\frac{1}{\sigma^2}\widehat{\boldsymbol{\alpha}}^\top \left(C\left(X^\top X\right)^{-1} C^\top\right)^{-1} \widehat{\boldsymbol{\alpha}} \overset{H_0}{\sim} \chi^2_{r\cdot},$$

based on the definition of the χ^2-distribution (see Gourieroux et al. (1982) for more details). Hence from this, under the restrictions imposed under the null hypothesis, $C\beta = \boldsymbol{d}$, we denote the LSE as $\widehat{\boldsymbol{\beta}}_{H_0}$, where the SSE under the H_0 is given by

$$\text{SSE}_{H_0} := \left\|\boldsymbol{Y} - X\widehat{\boldsymbol{\beta}}_{H_0}\right\|^2.$$

where SSE/σ^2 is χ^2-distributed with $n-p$ degrees of freedom (based on our distributional assumptions given in (3.19)). From this, now we can define the F-test, used in the study.

Definition 3.3: F-test for LMs

Under the null hypothesis H_0, for a general linear hypothesis (based on Definition 4.2) we use the following F-**test** statistic

$$F = \frac{\text{SSE}_{H_0}/r}{\text{SSE}\,/(n-p)} \overset{H_0}{\sim} F_{r,n-p^*}, \tag{3.47}$$

for r numerator degrees of freedom (*Num. Df*) and $n - p$ denominator degrees of freedom (*Den. Df*).

In this study, we use the statistical output to obtain our overall F-test - summarized in a tabular form - known as the ***Analysis of Variance (ANOVA)*** Table. The One-Way ANOVA Table (often to test how a independent variable affect the dependent variable) is shown in Table 3.3. Here the mean

square is the mean of the corresponding analysis of variance formulas given in Remark 4.1 (for example SSE/df, where df is the corresponding degrees of freedom) and the F-value is given by mean sqaures (MSR) of regression divided by the mean squares of error (MSE). Additionally, we can use ANOVA tables to compare between reduced and full models, shown in Table 3.4 (where the first model is the null model M_0).

Source	Sum of Squares (SS)	df	Mean Square (MS)	F-value
Regression	SSR	$m-1$	$MSR = SSR/(m-1)$	MSR/MSE
Error	SSE	$N-m$	$MSE = SSE/(N-m)$	
Total	SST	$N-1$		

Table 3.3. One-Way Analysis of Variance Table (ANOVA) for LMs, for F-tests, based on the analysis of variance formulas.

Name	Model	SSE	df	SSE difference	df difference
null M_0	$\boldsymbol{Y}=\beta_0\mathbf{1}_n+\boldsymbol{\varepsilon}$	$\text{SSE}_0=\sum_{i=1}^{n}(Y_i-Y)^2$	1		
reduced M_R	$\boldsymbol{Y}=\text{X}_1\boldsymbol{\beta}_1+\boldsymbol{\varepsilon}$	$\text{SSE}(X_1)=\left\|\boldsymbol{Y}-X_1\widehat{\boldsymbol{\beta}_1}\right\|^2$	p_1	$SSE_0-SSE(X_1)$	p_1-1
full M_F	$\boldsymbol{Y}=\text{X}\ \boldsymbol{\beta}+\boldsymbol{\varepsilon}$	$\text{SSE}(X)=\\|\boldsymbol{Y}-X\widehat{\boldsymbol{\beta}}\\|^2$	p	$SSE(X_1)-SSE(X)$	$p-p_1$

Table 3.4. ANOVA table (for comparing nested models), corresponding to the general hypothesis and based on the SSE (sum of squares error) and degrees of freedom df.

This means, using this ANOVA table, we can test if the reduced model, M_R contains a subset of parameters, p_1, in the full model M_F with p parameters, improves the overall model fit of M_F.

3.4.2 Measures of Fit for LMMs

Unlike the goodness of fit measures for LMs, determining the proportion of variance explained by an LMM is not as straightforward and can be challenging. Since the R^2 given by Equation (3.46) can only be estimated for linear models, it is unable to account for the variability in mixed-effects models with complex random effects structures. For this reason, we now define other goodness-of-fit measures extended and appropriate for LMMs.

Goodness-of-Fit Measures for LMMs

For the reasons discussed, we look at two extensions of the R^2 and other relevant statistics - known as the intraclass correlation coefficient (ICC) introduced by Edwards et al. (2008). All measures are related as they are ratios of the variance components in the model.

Coefficient of Determination - LMMs

To measure multivariate relationships between the repeated outcomes for individuals and the fixed effects in LMMs - we compute by extending the R^2 statistic. Here we follow the example outlined by Nakagawa et al. (2017).

Consider the 2-level hierarchical model given in (3.48) - which satisfy the the assumptions of the LMM - with a (2-level) grouping factor (say for the group effects of individuals i) and $p = k + 1$ fixed effects. This model then can be formulated as

$$\begin{aligned} Y_{ij} &= \beta_0 + \sum_{h=1}^{k} \beta_h x_{h_{ij}} + \alpha_i + \epsilon_{ij}, \\ \alpha_i &\sim \mathcal{N}(0, \sigma_\alpha^2), \qquad i.i.d \;\; \forall i = 1, \dots N, \\ \varepsilon_{ij} &\sim \mathcal{N}(0, \sigma_\varepsilon^2) \qquad i.i.d \;\; \forall i = 1, \dots N, j = 1, \dots, n_i, \end{aligned} \tag{3.48}$$

such that Y_{ij} is the response based on the i-th group of individuals with observation j, where x_{hij} is the h-th of the fixed effects of k covariates in the model. Here β_0 and $\beta_1 \dots \beta_k$ are the regression parameters of the fixed effects. While $\boldsymbol{\alpha_i}$ is the ***random individual (intercept) specific effect*** - normally distributed with zero mean and variance $\boldsymbol{\sigma_\alpha^2}$, and independent of the normally distributed random error terms $\boldsymbol{\varepsilon_{ij}}$ (or the observation specific residual) with zero mean and variance $\boldsymbol{\sigma_\varepsilon^2}$.

Additionally, suppose the ***variance is explained by the fixed effects*** in the LMM is given by

$$\sigma_\beta{}^2 = \text{var}\left(\sum_{h}^{k} \beta_h x_{h_{ij}}\right). \tag{3.49}$$

Then following this LMM framework, we can now define the two types and extensions of the R^2 statistic for mixed models.

Definition 3.4: Marginal and Conditional R^2 (for LMMs)

The marginal estimated multiple coefficient of determination for LMMs is given by

$$\boldsymbol{R^2_{mar.}} = \frac{\sigma_\beta^2}{\sigma_\beta^2 + \sigma_\alpha^2 + \sigma_\varepsilon^2}, \tag{3.50}$$

whereas, the ***conditional multiple coefficient of determination*** is calculated as

$$\boldsymbol{R^2_{con.}} = \frac{\sigma_\beta^2 + \sigma_\alpha^2}{\sigma_\beta^2 + \sigma_\alpha^2 + \sigma_\varepsilon^2}, \tag{3.51}$$

such that, with respect to the level 2 hierarchical model in Equation (3.48), we have

- σ_β^2 is the variance explained by the k fixed effects regression parameters $\beta_0, \dots, \beta_k$,
- σ_α^2 is the variance explained by the random i-th individual group-specific effect α_i,
- σ_ε is the variance of the random error terms (or residual variance), $Var(\varepsilon_{ij}) = \sigma_\varepsilon^2$ for the j-th observation of the i-th individual group.

The marginal $R^2_{mar.}$ is estimates the proportion of the total variance explained by the fixed effects only, given all other effects in the LMM. Whereas, the conditional $R^2_{con.}$ estimates the total proportion of variance explained by both the fixed and random effects in the model. Since for LMMs, in this study, we are interested in assessing both the structure of fixed and random effects - we consider both measures.

The Intraclass Correlation Coefficient (*ICC*)

Once again, consider the 2-level hierarchical LMM given by Equation (3.48). The intraclass correlation coefficient (ICC) essentially measures the proportion of the variance explained by the individual i's grouping or clustering structure in the data set. Often utilized to measure the ***"reliability"*** by comparing the variability (or correlation) within each individual i of the same group or cluster. For this reason the ICC can also account for all sources of uncertainty in the mixed model by calculating the ***"adjusted ICC"*** (random and fixed effects) or the ***"conditional ICC"*** (fixed effects only). (Byrne, 2013). The two types of ICC can be calculated in terms of the same variance components defined in Definition 4.5.

Definition 3.5: The Intraclass Correlation Coefficient (ICC)

The ***adjusted intraclass correlation coefficient (ICC)*** is given by

$$\boldsymbol{ICC_{adj.}} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}, \tag{3.52}$$

whereas is the ***conditional intraclass correlation coefficient*** is calculated by

$$\boldsymbol{ICC_{con.}} = \frac{\sigma_\alpha^2}{\sigma_\beta^2 + \sigma_\alpha^2 + \sigma_\varepsilon^2}, \tag{3.53}$$

Note, if no fixed effects are fitted (other than the intercept) then adjusted and conditional ICC are equivalent. Thus, based on the definition, the ICC is a measurement of the reliability based on the variations of the random effects and lies between zero and one - with one being high reliability. Similar to the R^2 for linear models - with only fixed effects - in the sense that the ICC provides information on the explained variance. Often interpreted as "the proportion of the variance explained by the grouping structure in the population" (Nakagawa et al., 2017). Thus, it is a valuable tool for assessing LMMs as it measures the correlations - within a risk class of data rather than the correlations between two different classes of data.

Hypothesis Testing and Inference for LMMs

Since, we have both fixed and mixed random effects in the model, we require two different statistical hypothesis testing framework. Firstly, based on the LMM framework given in Section 3.2.2, we estimate the standard errors for both the estimator for the fixed effects parameter $\boldsymbol{\beta}$ and for the random components (the BLUP), $\boldsymbol{\alpha}$ (following the framework outlined by Frees et al. (2014)). Since, assessing how much estimated variability is accounted for by the variance components within the model, helps us analyze the overall fit.

Remark 3.2: Estimation of Standard Errors (for LMMs)

Consider the marginal model $\boldsymbol{Y} \sim N(\boldsymbol{X\beta}, \boldsymbol{V}(\boldsymbol{\theta}))$ in (3.21), such that $\boldsymbol{\theta}$ denotes the vector of unknown parameters - used in $\boldsymbol{V}(\boldsymbol{\theta}) = \boldsymbol{ZD}(\boldsymbol{\theta})\boldsymbol{Z'} + \boldsymbol{D}(\boldsymbol{\theta})$ (i.e. the variance components).

Then the covariance of estimator $\hat{\boldsymbol{\beta}}$ (of the *fixed* regression parameters) is given by

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = \left(\boldsymbol{X}'\boldsymbol{V}^{-1}(\boldsymbol{\theta})\boldsymbol{X}\right)^{-1} \tag{3.54}$$

such that, here we use $\text{Cov}(\boldsymbol{Y}) = \boldsymbol{V}(\boldsymbol{\theta})$ is used. By replacing the vector of unknown parameters, $\boldsymbol{\theta}$ with its estimate based on the ML or REML $\hat{\boldsymbol{\theta}}$, and using $\hat{\boldsymbol{V}} := \boldsymbol{V}(\hat{\boldsymbol{\theta}})$, a natural estimate for $\text{Cov}(\hat{\boldsymbol{\beta}})$ is given by $\left(\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)^{-1}$.

Whereas, for the empirical BLUP, the covariance is derived by

$$\begin{aligned}\text{Cov}(\hat{\boldsymbol{\alpha}}) &= \text{Cov}\left(\boldsymbol{D}\boldsymbol{Z}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})\right) \\ &= \boldsymbol{D}\boldsymbol{Z}'\,\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\;\boldsymbol{Z}\;\boldsymbol{D},\end{aligned} \tag{3.55}$$

for the estimator of the random components, $\hat{\boldsymbol{\alpha}}$.

This allows us to estimate the precision of the estimated predictors involving both $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$. Moreover, we use the maximum likelihood estimation (MLE) method, we use the ***likelihood ratio test*** (LRT) to compare and test the fixed effects in nested (or hierarchical) LMMs.

Definition 3.6: Likelihood Ratio Test (LRT) - for Fixed Effects (using MLE)

Reject the null hypothesis H_0 (at significance level $\alpha\prime$)

$$H_0 : C\boldsymbol{\beta} = \boldsymbol{d} \quad \text{versus} \qquad H_1 : C\boldsymbol{\beta} \neq \boldsymbol{d}$$

where the corresponding **LRT statistic** is given by

$$-2\left[\ell(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}} \mid \boldsymbol{y}) - \ell\left(\widehat{\boldsymbol{\beta}}_R, \widehat{\boldsymbol{\alpha}}_R \mid \boldsymbol{y}\right)\right] > \chi^2_{1-\alpha\prime,r}$$

based on the log-likelihood for the fixed effects as given in Section 3.29, while

- the parameter estimates included in the restricted model ($C\boldsymbol{\beta} = \boldsymbol{d}$) are given by $\widehat{\boldsymbol{\beta}}_R$ and $\widehat{\alpha}_R$,
- the parameter estimates included in the unrestricted model is given by $\widehat{\boldsymbol{\beta}}$, and $\widehat{\boldsymbol{\alpha}}$, where
- χ^2 is approximately $r := rank(C)$ distributed (at $1 - \alpha\prime$ significance level).

Note, for clarification, we denote the significance level here as $\alpha\prime$, whereas $\boldsymbol{\alpha}$ denotes the random components. Whereas, if we fit the LMM using the REML method, as discussed in Section 3.34, we used the approximate hypothesis tests for the fixed effects (Wald's test).

Definition 3.7: Approximate Hypothesis Tests for Fixed Effects (using REML)

1. Wald Test

Reject H_0 (at significance level $\alpha\prime$):

$$H_0 : C\boldsymbol{\beta} = \boldsymbol{d} \qquad \text{versus} \qquad H_1 : C\boldsymbol{\beta} \neq \boldsymbol{d}, \tag{3.56}$$

where the Wald test statistic is,

$$W := (C\widehat{\boldsymbol{\beta}} - \boldsymbol{d})^\top \left(C^\top \operatorname{Cov}(\widehat{\boldsymbol{\beta}}))C\right)^{-1} (C\widehat{\boldsymbol{\beta}} - \boldsymbol{d}) > \chi^2_{1-\alpha,\operatorname{rank}(C)}, \tag{3.57}$$

where here we assume $\operatorname{Cov}(\widehat{\boldsymbol{\beta}})$ is fixed and does not depend on $\boldsymbol{Y}$.

2. Approximate F-test

The degrees of freedom, df, of the approximate F-test is given by ν_F (which may be differ from $\operatorname{rank}(C)$). The corresponding hypothesis testing is

Reject H_0:

$$H_0 : C\boldsymbol{\beta} = \boldsymbol{d} \qquad \text{versus} \qquad H_1 : C\boldsymbol{\beta} \neq \boldsymbol{d}, \tag{3.58}$$

at level $\alpha\prime$, when the approximate F-test for the fixed effects is given by

$$F := \frac{(C\widehat{\boldsymbol{\beta}} - \boldsymbol{d})^\top \left(C^\top \operatorname{Cov}(\widehat{\boldsymbol{\beta}}C\right)^{-1} (C\widehat{\boldsymbol{\beta}} - \boldsymbol{d})}{\operatorname{rank}(C)} > F_{1-\alpha,\operatorname{rank}(C),\nu_F}. \tag{3.59}$$

To statistically test if the random effects of different subjects are significantly different, we use the LRT. For instance, to investigate if the intercepts of the different risk class levels of the K clients are significantly different, and, thus, should be included in the model. For this reason, ***investigating the necessity of random effects*** requires a hypothesis test involving the variance components. Hence, we use the following LRT test for random components.

Definition 3.8: Likelihood Ratio Test (LRT) - for Random Effects

Reject the null hypothesis, H_0 (at significance level $\alpha\prime$):

$$H_0 : \sigma^2_\alpha = 0 \qquad \text{versus} \qquad H_0 : \sigma^2_\alpha > 0,$$

using the log-likelihood ratio test (LRT). Here the LRT test for nested models is approximately χ^2_α distributed, where σ^2_α is the estimated variance for the random components α.

Consider a model includes only one variance component. Then hypothesis testing (using LRT) for the fixed effects parameters involving r the reference distribution is $\chi^2_r + \frac{1}{2}\chi^2_{r+1}$. This is because zero is on the boundary of the parameter space allowed for σ^2_2. For this reason, the LRT statistic is not comparable with the distribution of χ^2_1 and should be instead compared with the mixture distribution of $\chi^2_r + \frac{1}{2}\chi^2_{r+1}$ (Frees et al., 1999). However, with more variance components involved, the complexity of the general testing problem and problem increases. See Frees (2018) and Ruppert et al. (2003) for further explanations and examples.

3.4.3 Measures of Fit for GLMs

While we use the residual sums of squares to assess the fit of LMs and different variance components for LMMs, for GLMMs we use the deviance and the generalized Pearson statistic. In this section, we briefly formulate both measures.

Goodness-of-Fit Measures for GLMs

Recall from Equation 3.39 the log-likelihood of the GLM can be formulated as

$$\begin{aligned}\ell(\boldsymbol{\beta}, \phi \mid \boldsymbol{y}) &= \sum_{i=1}^{n}\left[\frac{y_i\theta_i - b\left(\theta_i\right)}{a(\phi)} + c\left(y_i, \phi\right)\right] \\ &= \sum_{i=1}^{n}\left[\frac{y_i h\left(\mu_i\right) - b\left(h\left(\mu_i\right)\right)}{a(\phi)} + c\left(y_i, \phi\right)\right] \\ &=: \ell(\boldsymbol{\mu}, \phi \mid \boldsymbol{y}),\end{aligned} \tag{3.60}$$

such that, $\ell(\boldsymbol{\mu}, \phi \mid \boldsymbol{y})$ is the mean parameterization of the GLM log-likelihood. Following this, we can now define the *scaled and unscaled deviance* - which measures the discrepancy between the observations y_i and the fitted means $\hat{\mu}_i$.

Definition 3.9: Deviance in GLMs

Following the log-likelihood functions given in (3.39), the ***scaled deviance*** is defined as

$$\mathrm{D}_s(\widehat{\boldsymbol{\mu}}, \boldsymbol{y}, \phi) := -2[\ell(\widehat{\boldsymbol{\mu}}, \phi \mid \boldsymbol{y}) - \ell(\boldsymbol{y}, \phi \mid \boldsymbol{y})] = 2\sum_{i=1}^{n}\frac{y_i\left(\tilde{\theta}_i - \widehat{\theta}_i\right) - b\left(\tilde{\theta}_i\right) + b\left(\widehat{\theta}_i\right)}{a(\phi)}, \tag{3.61}$$

such that the estimators here are given by $\tilde{\theta}_i := h(\hat{\mu}_i)$ and $\tilde{\theta}_i := h(y_i)$. While the ***unscaled deviance*** is formulated as

$$\mathrm{D}(\widehat{\boldsymbol{\mu}}, \boldsymbol{y}) := \phi \mathrm{D}_s(\widehat{\boldsymbol{\mu}}, \boldsymbol{y}, \phi), \tag{3.62}$$

assuming the dispersion function satisfies $a(\phi) = \phi/w$.

Note, the unscaled deviance $\mathrm{D}(\widehat{\boldsymbol{\mu}}, \boldsymbol{y})$ here eliminates the influence of the dispersion parameter ϕ. We can also see that for the linear model in Equation (3.1), the scaled deviance can be measured by

$$\mathrm{D}_s(\widehat{\boldsymbol{\mu}}, \boldsymbol{y}, \phi) = -2[\ell(\widehat{\boldsymbol{\mu}}, \phi \mid \boldsymbol{y}) - \ell(\boldsymbol{y}, \phi \mid \boldsymbol{y})] = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - \widehat{\mu}_i\right)^2, \tag{3.63}$$

meanwhile the unscaled deviance is derived by

$$\mathrm{D}(\widehat{\boldsymbol{\mu}}, \boldsymbol{y}) = \sum_{i=1}^{n}\left(y_i - \widehat{\mu}_i\right)^2. \tag{3.64}$$

Recall, this is clearly similar to the residual sums of squares for the LM in (4.1). McCullagh and Nelder (1989) provides more theoretical background and examples on how the deviance of GLMs are derived.

Hypothesis Testing and Inference for GLMs

We construct statistical hypothesis tests for GLMs using the asymptotic distribution of the deviance. The two following statistical tests we utilize are: the *residual deviance test* and *the partial deviance test*. Specifically, to test the goodness of fit of the specifications of a GLM we use the residual deviance

test. Meanwhile, to compare the fit of two nested generalised linear models we use the partial deviance test.

Definition 3.10: Residual Deviance Test

Reject the null hypothesis H_0 (at signifance level α):

- the assumptions of the specified generalized linear model (GLM) are satisfied.

versus the alternative hypothesis H_1: not H_0.

If and only if

$$\frac{D(\hat{\boldsymbol{\mu}}, \boldsymbol{y})}{\hat{\phi}} > \chi^2_{n-q,1-\alpha}, \tag{3.65}$$

given:

- $D(\hat{\boldsymbol{\mu}}, \boldsymbol{y})$ is the observed deviance,
- $\hat{\phi}$ is an estimate of the dispersion parameter ϕ,
- $\chi^2_{n-1,1-\alpha}$ is the $100(1-\alpha)\%$ is the quantile of the χ^2 distribution with $n-q$ degrees of freedom (df)

Here, we test if the model assumptions of the specified GLM are satisfied. This includes the correct specification of the response distribution, the link function and the linear predictors, given the GLM.

Definition 3.11: Partial Deviance Test

Reject the null hypothesis H_0 (at level α):

$$H_0 : \boldsymbol{\beta_2} = \mathbf{0},$$

versus the alternative hypothesis H_1 (not H_0):

$$H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0},$$

if and only if

$$\frac{D_R - D_F}{\hat{\phi}_F} > \chi^2_{p_2,1-\alpha},$$

where we define:

- $\beta_1 \in \mathbb{R}^{p_1}$ and $\beta_2 \in \mathbb{R}^{p_2}$ with $p_1 + p_2 = p$
- D_F is the deviance of the full model (F)
- D_R i the deviance of the reduced model (R)
- $\hat{\phi}_F$ is an estimate for the dispersion parameter ϕ based on the full model (F).

3.4.4 Model and Variable Selection

To ensure we avoid overfitting and achieve optimal model complexity, we use information criterion we look at three types of information criteria. As comparing and evaluating all models based on

the goodness of fit measures is often not sufficient. This means that it is essential we select models balancing model quality of fit against the complexity. This is especially crucial for this study where we look at different classes of models with different numbers of combination of explanatory variables.

Information Criterion

In this section, we briefly provide the formulation of the information criterion used for model performance and comparison analysis in this study. Recall, the information criterion are likelihood based performance measures which includes a penalty for model complexity (specifically, based on the number of parameters). Since different information criterion vary by how much it penalizes the model (in proportion to the number of parameters), in this study we choose to focus on three different information criteria: the ***Akaike information criterion*** (AIC), the ***Bayesian information criterion*** (BIC), and, additionally, the ***conditional Akaike information criterion*** (AICc).
Our goal is to choose the best model - across all model classes - that attains the minimum value across all considered different information criterion (as a small value across all measures indicates the optimal balance of goodness of fit and optimal complexity).

Definition 3.12: AIC, AICc and BIC

Let k denote the number of regression parameters included in the model (including the intercept). Then for n total observations,

- The ***Akaike information criterion (AIC)*** is defined as

$$\text{AIC} : = -2 loglikelihood \pm 2 \cdot k. \tag{3.66}$$

- The ***Bayesian information criterion (BIC)*** is calculated as

$$\text{BIC} := -2 loglikelihood \pm \log(n) \cdot k. \tag{3.67}$$

- The ***conditional Akaike information criterion (AICc)*** is defined as

$$\text{AICc} := -2 loglikelihood \pm 2 \log(n) \cdot k \times \frac{n}{n-k-1}, \tag{3.68}$$

The AICc is especially appropriate for LMMs - focusing on clusters and small data sets. Since it is essentially modified to obtain a bias-corrected version of the AIC for small sample sizes (with 'extra' penalty). This is achieved by increasing the relative penalty for model complexity based on the sample size and the number of fitted parameters.

Remark 3.3: AIC and BIC for LMMs

Let k denote the number of fixed effects and covariance parameters of a LMM. Further, assume the estimated parameters of the mixed models are given by $\widehat{\boldsymbol{\beta}}$ (fixed effects) and $\widehat{\boldsymbol{\alpha}}$ (random effects). Such that by considering the defined information criteria measures, we can incorporate model selection uncertainty into the considered parameter estimates and evaluate precision of LMMs (Thagard et al., 2011).

Hence, the ***AIC*** for **LMMs** is calculated as

$$\text{AIC} := -2\ell(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}} \mid \boldsymbol{y}) + 2k, \tag{3.69}$$

and the ***BIC*** for **LMMs** is given by

$$\text{BIC} := -2\ell(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}} \mid \boldsymbol{y}) \pm \log(n)k.$$

As stated, in general, we aim to select models that attain the lowest values for most of the information criterion defined (or across all measures if possible and applicable).

Prediction Accuracy

For this study, as well as in the reinsurance or insurance context, it is crucial that we also investigate the model's predictive power and its accuracy when predicting the rate of loss (per exposure volume unit per k client) given a new set of data ("test data"). So far, we have only considered metrics that allow us to assess the balance between the goodness of fit and model complexity - with respect only to training data (or "in-sample" data). Since, even if we observe that a model can predict the loss ratios with high accuracy on the training data set ("in-sample" data) - but performs exceptionally poorly on "unseen" or new data, then it is not a valuable model nor the "best" suitable model.

For this reason, now we look at prediction errors as a measurement of predictive performance in models - given a new set of data, the test data - and how we can use these metrics to compare across different model classes. Recall that the test data in our study contains 148 total loss observations (aggregated per risk class level, based on 230 raw observations). In comparison, the training data set contained a total of 290 (with 442 raw observations).

Regarding assessing prediction accuracy for model comparison and selection - with respect to estimating the loss ratio LR - we consider various predictive accuracy measurements to compare and assess models - mainly based on the prediction error. The ***prediction error*** is the difference between its actual value and predicted value. As our test data contains observed ("*actual*") loss ratios per risk class level, this helps us compare and highlight each model's ability to predict the response `ln.lr` properly, given a new set of data.

For this reason, we first look at how we can assess the prediction errors in terms of the "***bias***".

Definition 3.13: Bias

Consider a parameter with true value θ, estimated by $\hat{\theta}$. Then, the mean prediction error - known as the ***bias*** - which is the difference between the expected value of the estimator, $\hat{\theta}$, and the the true value θ

$$\boldsymbol{bias}(\hat{\theta}) := E\left[(\hat{\theta}\right] - \theta = E\left[(\hat{\theta} - \theta)\right]. \tag{3.70}$$

where the the expectation of the estimator $\hat{\theta}$ is the average over all possible observations.

In other words, the *bias* is the expected average prediction error and is essentially used to measure on average how close, for example in our case, the predicted (log) loss ratios are to the observed values. When $\boldsymbol{bias}(\hat{\theta}) = 0$ we say the estimator of θ is unbiased. Whereas, a negative bias indicates underestimation of true or actual values while the a positive bias indicates overestimation. This means that only considering the bias as an measurement to assess the variability (or the spread) of estimates is not sufficient.

For this reason, we also consider the *Mean Squared Error* ***(MSE)*** and *Root Mean Squared Error* ***(RMSE)***. These measurements are not only applicable to compare both nested and non-nested models, but it is an an absolute test which works for all model classes (considered in this study). The RMSE is the absolute the root of the ***MSE***, which measures the average of the squares of the errors. It allows us to evaluate their quality of predictions by measuring the average magnitude of the model error.

Definition 3.14: Mean squared error (MSE) and Root Mean Squared Error (RMSE)

Consider any parameter with true value θ and estimator $\hat{\theta}$, the ***mean squared error***(MSE) of the estimator is given by

$$\boldsymbol{MSE}(\hat{\theta}) := E\left[(\hat{\theta} - \theta)^2\right], \tag{3.71}$$

which is the estimated average squared difference between the estimated values, $\hat{\theta}$ and what is being estimated, θ.

Following this, the ***root mean square error (RMSE)*** is then calculated as

$$\boldsymbol{RMSE}(\hat{\theta}) = \sqrt{E\left[(\hat{\theta} - \theta)^2\right]}. \tag{3.72}$$

By calculating the MSE on an out-of-sample data set for each model, we can compare their degree of performance accuracy based on "unseen" data. While then considering the RMSE allows us to compare quickly and efficiently calculate the predictive power - across all model classes, regardless of the model specification, configuration, choice of regression parameters or variance structure. We aim to minimize the RMSE and MSE values since lower values indicate a better fit. Additionally, in this study we also measure the "spread" or variability in the model estimates based on the ***Mean Absolute Error*** (***"MAE"***).

Definition 3.15: Mean Absolute Error (MAE)

For $i.i.d$ loss observations l in all risk classes levels (k, i, j, t), the ***Mean Absolute Error (MAE)*** is calculated as

$$\boldsymbol{MAE}(\hat{\theta}) := E\left[|\hat{\theta} - \theta|\right], \tag{3.73}$$

for any parameter with true value θ and estimator $\hat{\theta}$.

Note, for both RMSE and MAE - the smaller value, the better the model fit and performance. It can be shown that the RMSE penalizes the variance for larger error (absolute) values (in comparison to errors with smaller values), while the MAE gives the same weight to all errors. This means that, by definition, the RMSE values will never be smaller than the MAE. However, this also implies that RMSE is more sensitive to outliers than the MAE(as it measures the error differences prior to taking the average). Therefore, we use a combination of these metrics to evaluate the performances of all models considered adequately.

4 Natural Catastrophe Modeling

The study aims to estimate the average loss ratio for non-life insurance for total losses incurring from natural catastrophes. The loss ratio, also known as the pure premium or the loss of rate, accounts for the monetary risk an insurer or reinsurer bears. It estimates the cumulative insurance rate required to cover the considered risks. However, the premium rate is only the insurance rate paid by the original policyholder. In contrast, the loss ratio (i.e. the gross rate) incorporates all additional costs of the insurance coverage costs with the premium rates.

Given that it is the total insurance rate and is used to estimate the total cost for the insured loss - it is crucial for the insurer that the actual incurred loss should not excessively differ from its expected value. For this reason, the loss ratio is often the standard measure to assess the cumulative (total) risk transferred - from the policyholder to the insurer (shared by the reinsurer). Hence, in this chapter, we directly model the loss ratios using the following model classes: *Linear Models (LMs), Linear Mixed Models (LMMs), and Generalized Linear Models (GLMs).*

Specifically, we first look at the simplest class of models (as previously outlined in Chapter 2), the normal linear model with main effects. Then we investigate - given the grouping of data - the significance of the group risk class levels; of each risk class (k, i^G, j^G, t) for each client $k = 1, \ldots, K$. We aim to select the simplest and the best-performing model. For this reason, we also analyze the importance of interaction effects in our models. This is repeated for the other two model classes, LMMs and GLMs, where we compare the best-performing models from each class. We aim to find the best suitable model between all model classes for our longitudinal natural catastrophic data.

Note, for all models - based on our EDA findings (not included in this publication for confidentiality reasons) - we only consider the grouped risk class effects of treaty $i^G \in \mathcal{P}^k$ and countries $j^G \in \boldsymbol{\mathcal{M}}^{ki}$, for each client k (whereas we consider the ungrouped effects of the risk classes k, t). For simplicity, in this chapter we denote the grouped risk classes (i^G, j^G) as (i, j). Afterwards, we briefly look at EDA for each model class considered to check and specify the appropriate probability distribution visually. To specify and fit different models of the same model class (including main and interaction effects) - to explore the covariate effects on the loss ratio. This enables us to compare them and find the best fit among the subset of models. Lastly, using various model diagnostics tools, we assess the goodness of fit of the selected model followed by residual analysis (also to detect any outliers).

Recall the ultimate aggregated loss ratio (from the reinsurer's perspective) was previously defined as:

$$LR = \frac{L}{v} = \frac{\sum_{k=1}^{K} L_k}{\sum_{k=1}^{K} v_k}, \qquad \text{for } k = 1, \ldots K, \tag{4.1}$$

where L denotes the cumulative loss amount given the fixed exposure volume v. Hence, this section aims to estimate the loss ratio concerning natural catastrophic loss events in the Caribbeans - occurring between the years 2001 to 2020 (with 290 observations aggregated by the corresponding risk classes in (k, i, j, t)).

4.1 Linear Models for Natural Catastrophic Data

Since the log-normal distribution has the following property- for a random variable $Y \sim LN(\mu, \sigma^2)$ - it follows that $\ln(Y) \sim \mathcal{N}(\mu, \sigma^2)$ where the probability distribution function (pdf) is given by,

$$f_Y(y) = \frac{1}{y\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\ln(y) - \mu}{\sigma}\right)^2\right\}, \qquad y > 0. \tag{4.2}$$

We consider the log-normal distribution in this study since if we look at the distribution of the untransformed observed loss ratios LR_{kijt} (shown in Figure 4.1) - clearly, the first histogram (Column 1) shows a highly right-skewed shape. For this reason, we logramatically transform our response variable in Figure 4.1 (Column 2).

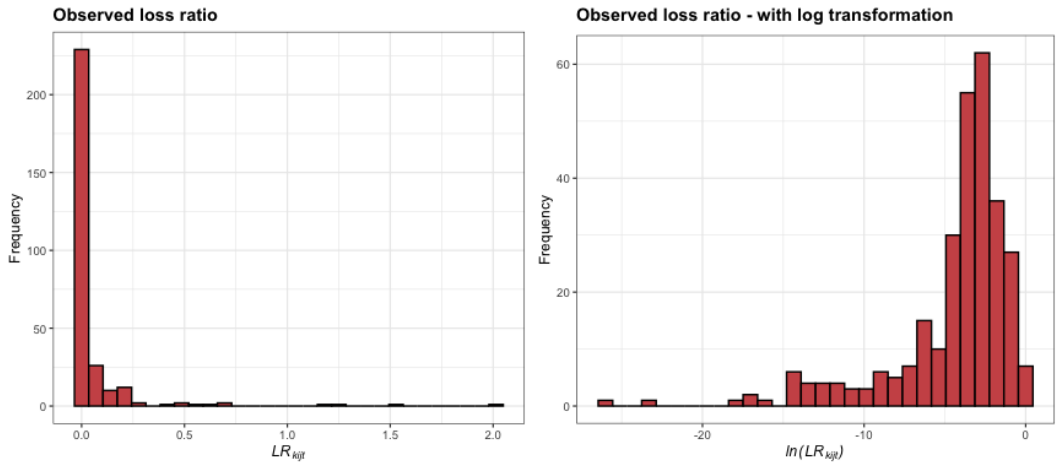


Figure 4.1. Histogram of aggregated observed loss ratios $LR_{k,i,j,t}$. Based on the risk classes clients $k = 1, \ldots, 35$, treaty (grouped) $i = \{1, 2\}$, country (grouped) $j = \{H, L\}$, and years $t = 2001, \ldots, 2020$. Column 1: Untransformed observed loss ratios (raw data). Column 2: Log-transformed Observed loss ratio based on risk classes (k, i, j, t).

Though the average observed loss ratios on the log scale still appear to be skewed towards the left, it appears to satisfy the assumptions of a linear model more. Since it appears it is not as heavily skewed as the observed loss ratios with no transformations. For this reason, moving forward, we model the loss ratios on the log scale for both LMs and LMMs considered in this study.

When investigating the observed loss ratios aggregated by client k with risks in treaty i for country j during treaty t - it is evident that there is a higher portion of small losses per exposure volume - in comparison to more considerable losses (with a maximum loss rate at 201%). Based on these histograms, it is also evident we have very few extreme values (only 5 counts of losses out of the total 290 losses are observed with a loss ratio over 101%). In other words, there exists a very low frequency of observed loss counts with a high loss amount per exposure volume). This aligns with our previous

findings (for observed loss amounts, severity, and exposure volumes). Due to the nature of catastrophic events, we observe a high frequency of low loss amounts and a low frequency of large loss amounts. Hence, in these cases, we expect the estimated mean of our loss ratios to be higher than our estimated median - which may require heavy tail distributions. However, truncating the data may be more reasonable due to the small number of extreme values (while allowing for simpler or straightforward interpretations). Our findings also indicate that a log-normal or Gamma model may also be a more suitable choice, discussed in the next following sections.

Recall, the linear models considered in this study - for natural catastrophic data - are categorized based on risk class level (k, i, j, t). This means we model the response, log loss ratio LR_{kijt}, estimated for each client k with insurance coverage in i-th treaty for country j active in year t. We assume LR_{kijt} follows a log-normal distribution, namely $LR_{kijt} \sim \text{LN}(\boldsymbol{x}_{kijt}^t \boldsymbol{\beta}, \sigma^2)$.

Model Specifications (LMs)

To investigate if the grouped risk classes are required - we first compare the base model (with no risk group effects) with models which analyze the groups of risk class level effects separately.

In summary, the following LMs analyzed in this section, respectively:

Model `lm.M` *(main effects of risk factors only):* 'traditional' linear regression model (or the ***base model***) which includes only the following main effects of risk factors as covariates: the type of the i-th treaty (`t.type`), the class of business insured (`cob`), the peril of the natural disaster (`peril`), the insurance rate (`rate`), the consumer pricing index of the j-th country in year t (`cpi`), and the number of (historic) unique natural disasters (`oci`).This means the model ignores the clustering of the data or the grouping effect of risk class levels in (k, i, j, t), and rather pools together the losses observed for all risk groups (i.e. ***'complete pooling'*** of risks (Frees, 2018)).

Thus, this model can be formulated as

Model `lm.M` *("completely pooled risks" with main risk factors only)*

$$\begin{aligned} \texttt{ln.lr}_{kijt} = \beta_0 + \sum_{q=2}^{4} \beta_1^q \texttt{t.type}_{ki}^q + \sum_{c=2}^{3} \beta_2^c \texttt{cob}_{ki}^c + \sum_{d=2}^{3} \beta_3^d \texttt{peril}_{kijt}^d + \beta_4 \texttt{oci}_{kijt} \\ + \beta_5 \texttt{rate}_{kijt}^d + \beta_6 \texttt{cpi}_{kijt} + \ \varepsilon_{kijt}, \end{aligned} \tag{4.3}$$

where $\texttt{ln.lr}_{kijt}$ is based on loss observations "pooled" together (ignoring clustering effects (k, i, j)). Hence, the intercept β_0 denotes the overall intercept for all risks for clients $k = 1, \ldots, 35$, with (grouped) treaties $\boldsymbol{\mathcal{P}}_G^k$ in all (grouped) countries $\boldsymbol{\mathcal{M}}_G^{ki}$ during all treaty years $\boldsymbol{\mathcal{T}}^{kij}$. Which provides the k-th client coverage for: commercial properties (`'C'`) with a combined reinsurance treaty type (`CB`) - for losses incurring from Earthquakes (`'E'`) based on only one natural catastrophic event (`oci`).

Model `lm.G` *(fixed effects of risk class groups)*: analyzes all main effects given the clustering of the data in the individual risk class levels - of each group of risk class levels in (k, i, j, t) separately (i.e. ***"no pooling"*** of risks). This means, for every client $k, k = 1, \ldots, 35$ each individual risk class group in (k, i, j, t) is included in the model as covariate and analyzed separately.

Hence, we can formulate this model as

Model `lm.G` *(risk class group effects as main fixed effects - with risk factors)*

$$\begin{aligned}\texttt{ln.lr}_{kijt} = \beta_0 + \sum_{k=1}^{35} \beta_1^k \texttt{client}_k + \beta_2 I_k^{\texttt{g.treaty},i} + \beta_3 I_{ki}^{\texttt{g.country},j} + \sum_{t=2}^{18} \beta_4^t \texttt{years}_{kij}^t \\ + \sum_{q=2}^{4} \beta_5^q \texttt{t.type}_{ki}^q + \sum_{c=2}^{3} \beta_6^c \texttt{cob}_{ki}^c + \sum_{d=2}^{3} \beta_7^d \texttt{peril}_{kijt}^d + \beta_8 \texttt{oci}_{kijt} \\ + \beta_9 \texttt{rate}_{kijt}^d + \beta_{10} \texttt{cpi}_{kijt} + \varepsilon_{kijt},\end{aligned} \tag{4.4}$$

where the indicator variables are defined as

$$I_k^{\texttt{g.treaty},i} = \begin{cases} 1, & \text{if loss observation } n_{kijt} \text{ belongs to a client } k \text{ who has more than 1 treaty } i \text{ for} \\ & \text{countries } j \in \mathcal{M}_G^{ki} \text{ during years } t \in \mathcal{T}^{kij}, \\ 0, & \text{otherwise,} \end{cases} \tag{4.5}$$

and

$$I_{ki}^{\texttt{g.country},j} = \begin{cases} 1, & \text{if loss observation } n_{kijt} \text{ belongs to country } j\text{-th group "Low" ("L") for a} \\ & \text{client } k \text{ with insurance coverage in the } i\text{-th treaty group active for the treaty} \\ & \text{years } t \in \mathcal{T}^{kij} \\ 0, & \text{otherwise.} \end{cases} \tag{4.6}$$

Model `lm.I` *(interaction effects)*: based on the results of the previous models, we then investigate if we should include the effects of interaction terms. Given the grouped risks and effects of risk class levels (i, j, t) for client $k, k = 1, \ldots, 35$. The potential interaction effects considered are based on the findings of the EDA and stepwise regression analysis, where we then select the best subset of suitable interaction effects.

Model `lm.M`: Main Effects of Risk Factors Only

As stated, we start by fitting the "base model" given by model `lm.M` in (4.3). It includes only the following 6 risk factors as covariates as the main (fixed) effects; without the grouping or client k specific effect of any risk class levels in (k, i, j, t). With 3 qualitative risk factors: `treaty.type`, `COB`, and `peril`, and three quantitative variables (which enter the model linearly, based on our EDA findings): `CPI`, `OCI` and the premium `rate`.

Model Results: `lm.M`

Here we fit the model using the aggregated data, (`"cat_data"`) with 290 observations (where the loss data is aggregated according to each client k and corresponding risk classes i, j, t). We utilize the `lm` function in `R` - followed by the model results and summary (given by the `summary` function).

```
lm.M <- lm(ln.lr ~  t.type + cob + peril + oci + rate + cpi, data = cat_data)
summary(lm.M)
```

```
Call:
lm(formula = ln.lr ~ t.type + cob + peril + oci + rate + cpi,
    data = cat_data)

Residuals:
   Min     1Q Median     3Q    Max
-21.64  -2.36   0.27   2.38  18.67

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    7.467      2.852     2.6    0.009 **
t.typeCXL     -1.703      1.328    -1.3    0.201
t.typeQS      -2.661      0.894    -3.0    0.003 **
t.typeSP      -2.477      1.181    -2.1    0.037 *
cobM          -0.399      1.328    -0.3    0.764
cobR          -0.028      1.479     0.0    0.985
perilF         2.184      1.355     1.6    0.108
perilH         2.722      1.043     2.6    0.010 **
oci           -5.774      0.268   -21.6   <2e-16 ***
rate          -0.784      0.187    -4.2    4e-05 ***
cpi           -0.033      0.017    -1.9    0.058 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 5.1 on 279 degrees of freedom
Multiple R-squared:  0.68,  Adjusted R-squared:  0.67
F-statistic:  60 on 10 and 279 DF,  p-value: <2e-16
```

The model considers a total of 290 observations with a total of 11 regression parameters and estimates the 'overall' intercept $\hat{\beta}_0 = 7.47$ (with corresponding *standard error*, 's.e.' of 2.85). If we set the significance level to 5%, it appears that all risk factors are significant except for `cob` and `cpi`.

Model Comparison: Full model (`lm.M`) and Reduced Model (`lm.M.r`)

To check if we can drop these terms from our model, we statistical test if the model can be further reduced model (`lm.M.r` without `cob` and `cpi`); by performing a partial F-test in `R` by fitting both the reduced and full model separately (`lm.M`). Such that, the LMs are compared using analysis of variance (ANOVA) techniques (using the `R` function `anova`).

Based on the results, with $p-$value $= 0.3$, at a 5% level of significance (with a F test statistic of 1.22 with 3 *degrees of freedom*, "df") we fail to reject the null hypothesis $H_0 : \beta_2^{(c\ =\ \texttt{M})} = \ldots = \beta_5 = 0$ and conclude that the reduced model does not improve the fit (without the risk factors `cob` and `cpi`).

```
Analysis of Variance Table

Model 1: ln.lr ~ t.type + peril + oci + rate.       # Reduced model: lm.M.r (without cob and cpi)
Model 2: ln.lr ~ t.type + cob + peril + oci + rate + cpi        # Full model: lm.M
  Res.Df  RSS Df Sum of Sq   F   Pr(>F)
1    282 7278
2    279 7184  3        94 1.22    0.3
```

Hence, the preferred reduced model of the risk factors, labelled by `lm.M.r`, can be formulated as:

Model `lm.M.r` *(reduced model with fixed main effects of risk factors only)*

$$\texttt{ln.lr}_{kijt} = \beta_0 + \sum_{q=2}^{4} \beta_1^q \texttt{t.type}_{ki}^q + \sum_{d=2}^{3} \beta_2^d \texttt{peril}_{kijt}^d + \beta_3 \texttt{oci}_{kijt} + \beta_4 \texttt{rate}_{kijt}^d + \varepsilon_{kijt}, \tag{4.7}$$

which models the average log loss ratio for each client k, treaty i, country j and year t. The estimates of regression parameters of the refitted and reduced model, `lm.M.r`, are shown in Table 4.1. Which also shows certain goodness of fit measures, such as the *multiple coefficient of determination* R^2 and adjusted R_a^2 (which, recall, measures the proportion of variability in the response, $\texttt{ln.lr}_{kijt}$). According to the `R` output, the estimated coefficient of determination is $R^2 = 68\%$ (with Adjusted R-squared, $R_a^2 = 67\%$), with regards to model `lm.M.r`. This indicates the regression model accounts for somewhat an adequate portion of variability in the response when accounting for only 4 of risk factors (main effects).

	Estimate	Std. Error	t-value	Pr($>\|t\|$)
(Intercept)	3.35	1.57	2.13	0.03
`t.typeCXL`	-1.72	1.3-	-1.32	0.19
`t.typeQS`	-2.73	0.85	-3.19	0.00
`t.typeSP`	-2.56	1.14	-2.24	0.03
`perilF`	2.12	1.35	1.56	0.12
`perilH`	3.22	0.98	3.27	0.00
`oci`	-5.73	0.26	-21.8	1.6e-6
`rate`	-0.80	0.18	-4.34	1.9e-5
Observations	290			
R^2	0.68			
R^2_{adj}	0.67			
Residual Std. Error	5.10	(df = 282)		
F Statistic	85.0	(df = 7; 282)		

Table 4.1. Summary of reduced model, `lm.M.r`. Table shows the estimated regression parameters (*"Estimate"*), the corresponding standard error (*"Std. Error"*), t-values and p-values. The bottom row provides additional model information: the number of observations (*"Observations"*), the multiple coefficient of determination, R^2 (with adjusted R^2_{adj}), the residual Std. Error and F test statistic with the corresponding degrees of freedom ("df") - based on aggregated data ("cat_data")

However, it is important to note that: firstly, the "traditional" model does not account for the heterogeneity of the risks between each client k for all (k, i, j, t) - where the model ignores the clustering of data in the risk classes (fits an overall intercept β_0 and an overall slope, say β_1, for all main effects and risk classes). Secondly, the model is not suitable for our natural catastrophic data in this case (longitudinal data with repeated measures for the client k or risk class level). For this reason, we investigate if our model fit improves by including the grouping effects of the risk classes (and, later, with LMMs).

Model `lm.G`: Main Effects of Risk Factors with Risk Class Group Effects

So far, we have only explored the main effects on the log loss ratio without individual grouping (or clustering) effects of risk classes. By specifying the regression models at different levels, the effects of our covariates may vary (given the varying effects of each cluster data of the corresponding risk class group). Thus, here we investigate the effect of *"no pooling"* (`lm.G`) based on all groups of the risk class levels (k, i, j, t), given by model `lm.G` (4.4).

Recall, this model was formulated as:

$$\begin{aligned}\textbf{Model } \texttt{lm.G}: \ \texttt{ln.lr}_{kijt} = \beta_0 + \sum_{k=1}^{35} \beta_1^k \texttt{client}_k + \beta_2 I_k^{\texttt{g.treaty},i} + \beta_3 I_{ki}^{\texttt{g.country},j} + \sum_{t=2}^{18} \beta_4^t \texttt{years}_{kij}^t \\ + \sum_{q=2}^{4} \beta_5^q \texttt{t.type}_{ki}^q + \sum_{c=2}^{3} \beta_6^c \texttt{cob}_{kijt} + \beta_7^d \texttt{peril}_{kijt} + \beta_8 \texttt{oci}_{kijt} \\ + \beta_9 \texttt{rate}_{kijt} + \beta_{10} \texttt{cpi}_{kijt} + \varepsilon_{kijt}.\end{aligned}$$

Model Results: `lm.G`

We investigate if the individual effects of every risk class group in (k, i, j, t) on the log loss ratio $\texttt{ln.lr}_{kijt}$ are required. Each risk class group (k, i, j, t) here is included as factor variables - in addition to all the main fixed effects given in Model (4.3) (as the significance of the factors in `lm.M.r` may vary when the group effects are accounted for). This means that, for each given observation belonging to the k-th client, for $k, k = 1, \ldots, 35$, the model estimates a parameter for each group in (k, i, j, t). A total of 58 risk class groups is considered separately (35 `client` + 2 `g.treaty` + 2 `g.country` + 19 `years`).

Such that the intercept β_0 denotes the individual intercept estimated for each k-th client; while also (separately) capturing the effects at each (grouped) treaty i, for the j-th (grouped) country within each treaty year t. We fit this model also using the aggregated data (`cat_data`) with **290 observations**; grouped by each k-th client for $k = 1, ..., 35$ with (grouped) treaty $i \in \{1^k, 2^k\}$ for (grouped) country $j \in \{L^{ki}, H^{ki}\}$ per treaty year $t \in \{2001^{kij}, \ldots, 2021^{kij}\}$. The summary of the estimated coefficients and results of the model fit (in `R`) are shown in Table 4.2. This shows that $\hat{\beta}_0 = 14.83$ (with s.e. or "Std.Error" 5.68).

We notice here that compared to `lm.M.r`, we see a increase in both R^2 (81%) and R^2_{adj} (75%). This means that a higher proportion of variance in `ln.lr` is explained when we introduce the (individual) risk class group effects (k, i, j, t) into the model and, thus, indicates a better fit. Whereas, the coefficient of determination estimated at $R^2 = 68\%$ (with Adjusted R-squared, R^2_{adj} of 67%) - measures a reasonable proportion of variability in loss ratio (on the log scale) - accounted for by the model `lm.G.1`.

Interestingly, it also appears that at a 5% significance level - when accounting for the risk class effects - `cpi` is now a significant influence (p-value = 0.02). While `cob` remains to be not be significant (at $\alpha = 0.05$). However, `peril` is no longer significant in this model either. We once again exclude both covariates and refit the reduced model, `lm.G.r`, to statistically test (using the partial F-test) if by excluding `peril` and `cob` improves the overall model fit.

	Estimate	Std. Error	t-value	Pr(>|t|)
(Intercept)	14.83	5.68	2.61	0.01
client6	-1.04	2.30	-0.45	0.65
client7	0.37	2.38	0.15	0.88
client8	-8.49	2.62	-3.23	0.00
client9	-4.70	2.52	-1.86	0.06
...	...	...	...	...
g.treaty2	1.92	5.32	0.36	0.72
g.countryL	1.02	1.24	0.82	0.41
years2003	-0.10	3.87	-0.03	0.98
years2004	0.96	3.41	0.28	0.78
years2005	1.16	3.46	0.34	0.74
years2007	-1.58	3.50	-0.45	0.65
...	...	...	...	...
t.typeCXL	-0.71	1.67	-0.42	0.67
t.typeQS	-3.19	1.33	-2.39	0.02
t.typeSP	-3.77	1.70	-2.22	0.03
cobM	0.68	1.81	0.38	0.71
cobR	-0.95	1.59	-0.60	0.55
perilF	1.20	1.40	0.86	0.39
perilH	2.08	1.28	1.62	0.11
oci	-5.48	0.32	-17.10	< 2e-16
rate	-1.02	0.26	-3.95	1e-04
cpi	-0.08	0.03	-2.28	0.02
Observations	290			
R^2	0.81			
R^2_{adj}	0.75			
Residual Std.Error	4.4	(df = 226)		
F Statistic	15.0	(df = 63; 226)		

Table 4.2. Summary of estimated regression coefficients of model `lm.G`. LM includes the individual risk class level effects, based on each client k - using aggregated data.

Model Comparison

Once again, we utilize the ANOVA table, shown in Table 4.3, obtained by `anova` - to compare the full model `lm.G` and the reduced model `lm.G.r`. Based on the results, we conclude that at $\alpha = 0.05$ with p-value = 036 the effects of `cob` and `peril` do not improve the model fit. Thus, the reduced model `lm.G.r` is preferred model.

Model	lm Formula (in R)	Res. Df	RSS	Df	Sum of Sq.	t-value	Pr(> |t|)
lm.G.r	lm(ln.lr ~ client + g.treaty + g.country + years+ t.type + oci + rate + cpi, data = cat_data)	230	4453.20				
lm.G	lm(ln.lr ~ client + g.treaty + g.country + years+ t.type + cob + peril + oci + rate + cpi, data = cat_data)	226	4368.88	4	84.32	1.09	0.36

Table 4.3. Analysis of variances (ANOVA) table for LMs, to compare the models with risk class level effects: reduced model (without `cob` and `peril`) and full model `lm.G`. Table shows the degrees of freedom for the residuals ("Res. Df"), the residual sum of squares ("RSS"), df based on the removed from the full model "DF", difference between the RSS of reduced and full model ("Sum of Sq") and the F test-statistic with corresponding p-value.

Table 4.4 compares the performances of the two (reduced) models; `lm.M.r` with no risk class level effects, and `lm.G.r` with individual effect risk class levels. Clearly it is evident that `lm.G.1` is a significantly better fit versus `lm.M` based on the given respective performance measures. For instance, `lm.G.1` accounts for a much higher variability in the log loss ratio with respect to the main fixed effects of the risk factors (with $R^2 = 80\%$ and $R^2 = 75\%$).

This is also evident in Figure 4.2 when we compare the fitted values versus the observed values of the response `ln.lr` obtained by both LMs. Additionally, the model `lm.G.r` also attains the lowest AIC, AICc and BIC value and the lowest prediction error (with RMSE = 3.92, and residual s.e. $\hat{\sigma} = 4.4$), in comparison to model `lm.M.r`.

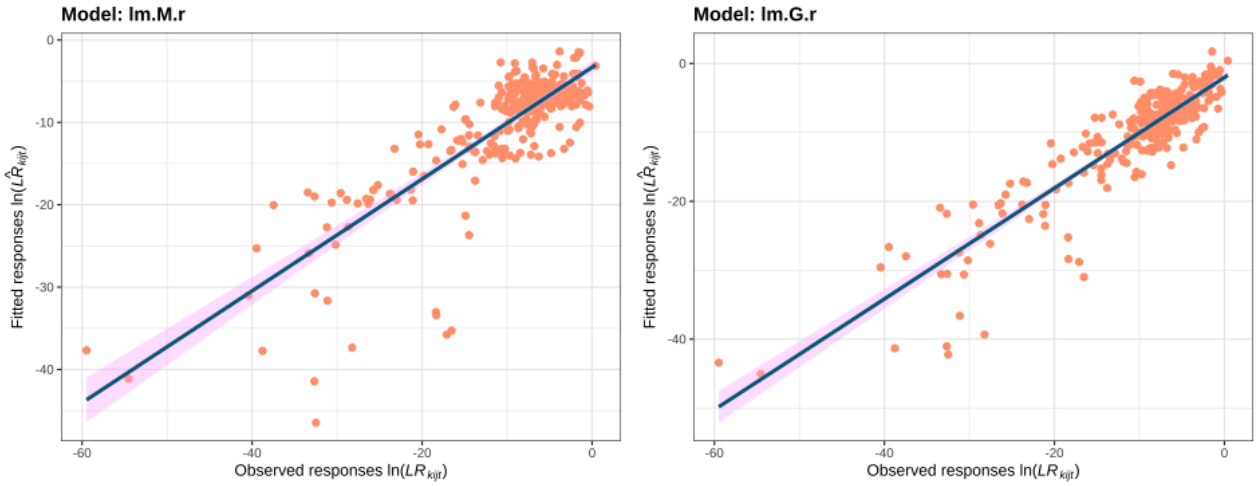


Figure 4.2. Plot of fitted response versus observed values (with corresponding 95% confidence interval bands) of the response: log loss ratio LR_{kijt}. Observed values (in orange) are based on losses aggregated at risk class level (k, i, j, t). **Left plot:**fitted values obtained with reduced LM model (4.3); linear *"base model"* fit with main fixed effects of the risk factors only (`"lm.M.r"`). **Right plot:** obtained by reduced model (4.4) fit, i.e. *"no pooling of risks"*, includes group effects of risk class (k, i, j, t) for each k-th client (`"lm.G.r"`).

Model	p	logLik	AIC	AICc	BIC	R^2	$R^2_{adj.}$	RMSE	$\hat{\sigma}$
`lm.M.r`	9	-878.78	1775.56	1776.20	1808.59	0.68	0.67	5.01	5.08
`lm.G.r`	61	-807.56	1737.12	1770.29	1960.98	0.80	0.75	3.92	4.40

Table 4.4. Comparison of linear model performance. Based on the measures: logLik, AIC, AICc, BIC, R^2, R^2_{adj}, and RMSE and estimated residual standard error $\hat{\sigma}$. The reduced LM with main fixed effects of the risk factors only, `lm.M.r`("base model") - is compared with the reduced model which includes the individual effects of the risk class levels (k, i, j, t). Highlighted cells (green) are the best-performing models with respect to the corresponding performance measure. The number of regression parameters included in each model is denoted by p.

Based on these results, we conclude the model (`lm.G.r`) is our preferred model, given by

Model `lm.G.r` *(reduced model with fixed effects and grouped risk classes)*

$$\begin{aligned}\texttt{ln.lr}_{kijt} = \beta_0 + \sum_{k=1}^{35} \beta_1^k \texttt{client}_k + \beta_2 I_k^{\texttt{g.treaty}_i} + \beta_3 I_{ki}^{\texttt{g.country}_j} + \sum_{t=2}^{18} \beta_4^t \texttt{years}_{kij}^t \\ + \sum_{q=2}^{4} \beta_5^q \texttt{t.type}_{ki}^q \quad + \beta_6 \texttt{oci}_{kijt} + \beta_7 \texttt{rate}_{kijt}^d + \beta_8 \texttt{cpi}_{kijt} + \varepsilon_{kijt}.\end{aligned} \tag{4.8}$$

Though this selected model provides a reasonably good fit (in terms of the $R^2 = 0.80$), the reduced model contains a large number of covariates. This is because each risk class group in model `lm.G.r` is analyzed separately for the k-th client in $k = 1, \ldots, 35$, where even very small clusters in our data "will get a regression parameter." For this reason, this "no pooling" model structure may result in over-fitting (Antonio and Zhang, 2013) while ignoring the nested or hierarchical structure present in the clustered data for the risk classes.

Hence, these findings motivate us to explore linear mixed models, which allow for *"partial pooling"* (allowing for appropriate cluster structuring between the risks). However, beforehand we must first investigate if there exist any strong indications of interaction effects present in our data set, i.e. check if we should allow for significant interaction effects to improve the selected model given by Model (4.8).

Model `lm.I`: Interaction Effects (With Group Effects of Risk Class)

Since we are fitting and specifying regression models at different hierarchical risk class levels, interactions between explanatory variables at different levels (*"cross-level"* effects) may be present. For this reason, here we analyze the model "`lm.I`" with pairwise interaction effects and test if we should include certain interaction effects to essentially improve the model fit of `lm.G.r` (given by Model (4.8)).

We consider two-way interaction effects by first investigating all pairwise interactions (based on the main effects in `lm.G.r`) - using explorative data analysis (EDA) (for example, as shown in Figure 4.3). Given the estimated average log loss ratios (i.e. empirical loss average per exposure volume unit), the plots give us initial ideas about the interaction effects we should further investigate. In the sense that interactions between the covariates may be present - if the polylines of the levels of two covariates are relatively not parallel. For this, in `R`, for two categorical variables we utilize the function `cat_plot` (from the `R` package `interactions`) to test the conditional effect of the "focal predictor" (on the x-axis) at the factor levels of the "moderating" (second) variable (e.g., high, medium, and low). Here the estimated conditional effects (polylines) are often known as the "simple slopes" (see Bauer and Curran (2005) for further examples).

Whereas, we use `interact_plot` (available in the same `R` package `interactions`) for interactions between continuous focal risk predictors and categorical moderating predictors. Note, we only explore the effects without the k-th client, the individual group effects - to investigate if the interaction effects that may be present at the i-th group level and the multi-level data structure.

Model Results: `lm.I`

Our analysis of the corresponding interaction plots (for example, as shown in Figure 4.3) with initial (visual) inspections of all the two-way pairwise interaction effects are summarized in Table 4.5. Here pairwise interaction terms are specified by the operator `":"` for any two variables (i.e. `"variable1:variable2"`, using R's operator syntax). From these results, for example, we observe that there exists strong interaction effects between `oci` and the following variables: `years`, `t.type`, `cpi` and `rate`. Thus, based on these findings we further investigate only the terms with strong or possible (mild) interaction effects by including them into our model `lm.G.r`.

However, due to the high number of potential interaction effects, we first perform a *stepwise regression* analysis to reduce the number of interaction effects - while also supporting our findings from the interaction plots. Our strategy here is to consider a large subset of pairwise interaction - based on the interaction plots using the aggregated `cat_data`). Then we remove and add interaction terms (individually) until we find the best performing reduced interaction model `lm.I` (based on the AIC and R^2_{adj}).

For this reason, we perform this forward and backward stepwise selection using the function `drop1` in `R` (where we add or drop interaction effects using both p-values and the AIC criteria). Table 4.6 shows a summary of our F-tests performed to compare the models - without and with regard to the final selected subset of two-way pairwise interaction effects (based on our stepwise regression analysis). For more details and examples on how step-wise regression methods are implemented in `R`, see Gareth et al. (2013).

	Df	Sum Sq	Mean Sq	F-value	Pr(>F)
`g.country:years`	11	124.11	11.28	0.98	0.46
`g.treaty:oci`	1	15.06	15.06	1.31	0.25
`g.country:oci`	1	59.64	59.64	5.20	0.02*
`years:oci`	15	1497.75	99.85	8.70	1.7e-14***
`rate:years`	17	178.04	10.47	0.91	0.56
`years:cpi`	15	411.88	27.46	2.39	0.004**
`t.type:oci`	3	164.06	54.69	4.76	0.003**
`t.type:cpi`	3	66.53	22.18	1.93	0.13
`oci:cpi`	1	64.77	64.77	5.64	0.02*
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Table 4.6. Summary table of F-tests to find the final subset of two-way pairwise interaction terms selected by the step-wise regression (using the `R` function `step`) based on backward and forward selection. The step-wise selection consists of iteratively adding and removing predictors (one at a time) from the predicted model, `lm.I`.

Here we observe that at a 5% significance level the following pairwise effects may potentially improve our (selected) LM fit: `g.country:oci`, `years:oci`, `years:cpi`, `t.type:oci`, `t.type:cpi` and `oci:cpi`.

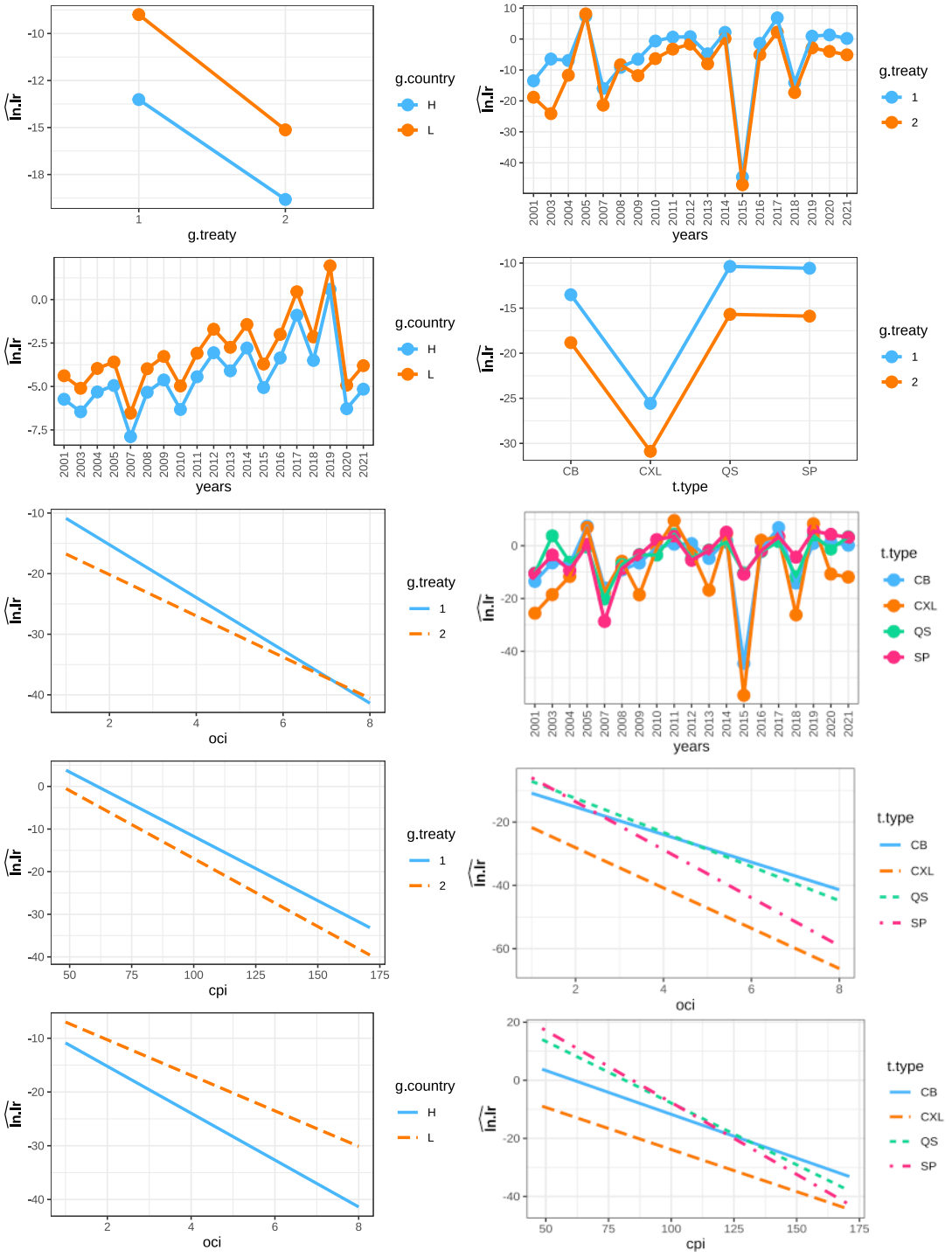


Figure 4.3. Example of *'Pairwise interaction Effects'* plots - for LMs with individual clustering effects in (k, i, j, t), based on aggregated data with 290 observations. Note, for confidentiality reasons not all interaction plots for all variables are shown (within this publication).

		Explorative Data Analysis (Analysis of Interaction Plots)		
Interaction Effects (Two-Way)	**Corresponding Plot (Figure , Row , Column)**	**Strong Indications of Interaction Effects**	**Possible (Mild) Indications of Interaction effects**	**No Indications of Interaction Effects**
g.treaty:g.country	Figure 4.3, Row 1, Column 1			✓
years:g.treaty	Figure 4.3, Row 1, Column 2		✓	
years:g.country	Figure 4.3, Row 2, Column 1		✓	
g.treaty:t.type	Figure 4.3, Row 2, Column 2			✓
oci:g.treaty	Figure 4.3, Row 3, Column 1		✓	
rate:g.treaty	Figure 4.3, Row 3, Column 2		✓	
cpi:g.treaty	Figure 4.3, Row 4, Column 1			✓
g.country:t.type	Figure 4.3, Row 4, Column 2		✓	
oci:g.treaty	Figure Not Shown.			✓
rate:g.country	Figure Not Shown.			✓
oci:g.country	Figure 4.3 Row 5, Column 1		✓	
years:t.type	Figure Not Shown.	✓		
oci:years	Figure Not Shown.	✓		
rate:years	Figure Not Shown.	✓		
cpi:years	Figure Not Shown.	✓		
oci:t.type	Figure 4.3, Row 4, Column 2	✓		
rate:t.type	Figure Not Shown.	✓		
cpi:t.type	Figure 4.3, Row 5, Column 2	✓		
cpi:oci	Figure Not Shown.	✓		
rate:oci	Figure Not Shown.			✓

Table 4.5. EDA analysis and inspection of interaction plots given in Figure 4.3: based on the fitted values of the response, $\widehat{\texttt{ln.lr}}$, and covariates in model `lm.G.r`. Here *"Figure Not Shown"* states the interaction plots not shown in this publication (for simplicity and data confidentiality purposes). Each corresponding interaction term analyzed is given in Column 1, where `":"` denotes the interaction operator in `R`. The green check marks if there any strong, possible (mild) or no indications of interaction effects present between the listed variables, based on their respective plots (Column 2).

Accordingly, we now fit this (full interaction) model `lm.I` - based on our analysis - formulated as,

Model `lm.I` *(full interaction model with fixed effects)*

$$\begin{aligned}\texttt{ln.lr}_{kijt} = \beta_0 &+ \sum_{k=1}^{35}\beta_1^k\texttt{client}_k + \beta_2 I_k^{\texttt{g.treaty}_i} + \beta_3 I_{ki}^{\texttt{g.country}_j} + \sum_{t=2}^{18}\beta_4^t\texttt{years}_{kij}^t + \sum_{q=2}^{4}\beta_5^q\texttt{t.type}_{ki}^q \\ &+ \beta_6\texttt{oci}_{kijt} + \beta_7\texttt{rate}_{kijt}^d + \beta_8\texttt{cpi}_{kijt} + \beta_9 I_{ki}^{\texttt{g.country}_j}\times\texttt{oci}_{kijt} + \sum_{t=2}^{9}\beta_{10}^t\texttt{years}_{kij}^t\times\texttt{oci}_{kijt} \\ &+ \sum_{t=2}^{9}\beta_{11}^t\texttt{years}_{kij}^t\times\texttt{cpi}_{kijt} + \sum_{q=2}^{4}\beta_{12}^q\texttt{t.type}_{ki}^q\times\texttt{oci}_{kijt} + \beta_{13}\texttt{oci}_{kijt}\times\texttt{cpi}_{kijt} + \varepsilon_{kijt},\end{aligned} \tag{4.9}$$

for all client $k = 1, \dots 35$, (grouped) treaty $i \in \{1^k, 2^k\}$, (grouped) country $j \in \{H^{ki}, L^{ki}\}$, and per treaty year $t \in \{2001^{kij}, \dots, 2021^{kij}\}$. Note here the notation `"×"` represents the interaction operator (relating to `":"` in `R`).

By running the following `R` command, we can confirm if we should include all interaction effects into the model or if we can further reduce the number of interaction effects (based on expected loss ratio for the k-th client given the risk class levels).

```
lm.I <- lm(formula = ln.lr ~ client + g.treaty + g.country + years + t.type  + oci + rate + cpi + g.country:oci +
    years:oci + years:cpi + t.type:oci + oci:cpi, data = cat_data, na.action = na.omit)
summary(lm.I.1)
```

The full `R` output is given in Appendix A.1. In this output, $\hat{\beta}_0 = -32.33$ with s.e 18.65. By allowing for interaction effects we see an increase in both the R^2 and R^2_{adj} to 89% and 84% respectively (from 80%, based on the main effects model `lm.G.r`). The results show that the following interactions between the terms are statistically at a 5% significance level: `year` and `cpi`, `t.type` and `oci`, and, between the `oci` and `cpi`. Meanwhile the interaction effects between `g.country` and `oci`, `years` and `oci` may not be significant at the 5% significant level.

Model Comparison: Full Interaction Model (`lm.I`) and Reduced Models (`lm.I.r`)

To statistically test if we can drop these interaction terms (single term deletions), once again we conduct a partial F-test to compare the full and reduced models - referred to as `"lm.I.r1"` without the term `g.country:oci`). Then we investigate if we can reduce the model even further by excluding the effects of `years:oci` (`"lm.I.r2"`). The corresponding ANOVA table is given in Table 4.7 obtained by `anova` in `R`.

From the results, at 5% significance level we conclude that the pairwise effects `g.country:oci` should be dropped from the model (p-value = 0.29). While there is significant evidence that the interaction effects between `years` and `oci` have a significant influence on the average log loss ratio (p-value <0.05), so the model cannot be reduced any further, and we conclude that the model `"lm.I.r1` is our preferred model.

This "final" reduced model (now labelled `lm.I.r` for simplicity) is formulated as

Model	lmFormula (in R)	Res. Df	RSS	Df	Sum of Sq.	t-value	Pr($>\|t\|$)
lm.I.r1	lm(ln.lr ∼ client + g.treaty + g.country + years+ t.type + oci + rate + cpi + years : oci + years : cpi+ t.type : oci + oci : cpi	196	2396				
lm.I	lm(ln.lr ∼ client + g.treaty + g.country + years+ t.type + oci + rate + cpi + g.country : oci + years : oci+ years : cpi + t.type : oci + oci : cpi	195	2382	1	14	1.14	0.29
lm.I.r2	lm(ln.lr ∼ client + g.treaty + g.country + years+ t.type + oci + rate + cpi + years : cpi+ + t.type : oci + oci : cpi	211	3785				
lm.I.r1	lm(ln.lr ∼ client + g.treaty + g.country + years+ t.type + oci + rate + cpi + years : oci+ years : cpi + t.type : oci + oci : cpi	196	2396	15	1390	7.58	3.3e-13

Table 4.7. ANOVA table for model comparison based on LMs with interaction models, using t-tests. First, the full model lm.I is compared with reduced model (by single term deletions, in red) lm.I.r1 (without interaction g.country:oci). Then we compare the preferred model (lm.I.r1) with lm.I.r2 (without years:oci).

Model lm.I.r *(reduced interaction model with fixed effects)*

$$\begin{aligned}\texttt{ln.lr}_{kijt} &= \beta_0 + \sum_{k=1}^{35}\beta_1^k\texttt{client}_k + \beta_2 I_k^{\texttt{g.treaty}_i} + \beta_3 I_{ki}^{\texttt{g.country}_j} + \sum_{t=2}^{18}\beta_4^t\texttt{years}_{kij}^t \\ &+ \sum_{q=2}^{4}\beta_5^q\texttt{t.type}_{ki}^q + \beta_6\texttt{oci}_{kijt} + \beta_7\texttt{rate}_{kijt}^d + \beta_8\texttt{cpi}_{kijt} + \sum_{t=2}^{9}\beta_9^t\texttt{years}_{kij}^t \times \texttt{oci}_{kijt} \\ &+ \sum_{t=2}^{9}\beta_{10}^t\texttt{years}_{kij}^t \times \texttt{cpi}_{kijt} \sum_{q=2}^{4}\beta_9^q\texttt{t.type}_{ki}^q \times \texttt{oci}_{kijt} + \beta_{11}\texttt{oci}_{kijt} \times \texttt{cpi}_{kijt} + \varepsilon_{kijt},\end{aligned} \tag{4.10}$$

given client $k = 1, \dots 35$, (grouped) treaty $i \in \{1^k, 2^k\}$, (grouped) countries $j \in \{H^{ki}, L^{ki}\}$, within treaty years $t \in \{2001^{kij}, \dots, 2021^{kij}\}$.

Overall Comparison of all Selected LMs

Finally, now we compare all (reduced or selected) LMs so far: lm.M.r, lm.G.r and lm.I.r1 (see Table 4.8). From here the estimated coefficient of determination of this reduced model accounts for higher variability in the log rate of loss (with $R^2 = 89\%$ and $R^2_{adj} = 84\%$) - compared to both lm.G.r and lm.M.r. The interaction model performs better in general based on the lowest AIC and AICc criteria - in addition to a lower RMSE and estimated residual standard error $\hat{\sigma}$ (in comparison to all other LM models investigated in this section). Thus, we conclude that lm.I.2 overall performs the best and, thus, is our preferred linear model (based on the respective performance measures; AIC, AICc, R^2, R^2_{adj}, RMSE and residual standard error). However, according to the BIC criteria, lm.I.r does not attain the lowest BIC values. This may be due to the following main reasons:

1) the high number of interactions effects included in the model ($p = 81$),

2) the "no pooling" approach fits each group separately, such that there exists a regression coefficient

Model	p	logLik	AIC	AICc	BIC	R^2	$R^2_{adj.}$	RMSE	$\hat{\sigma}$
`lm.M.r`	9	-878.78	1775.56	1776.20	1808.59	0.68	0.67	5.01	5.08
`lm.G.r`	61	-807.56	1737.12	1770.29	1960.98	0.80	0.75	3.92	4.40
`lm.I.r`	81	-729.76	1625.31	1685.38	1918.77	0.89	0.84	2.98	3.52

Table 4.8. Comparison of all reduced linear model performances: main risk effects model only `lm.M.r`, main effects model with group effects `lm.G.r`, and the (selected reduced) interaction effects model `lm.I.r`. The number of regression parameters is given by p. The following performance measurements are compared: logLik, AIC, AICc, BIC, R^2, R^2_{adj}, RMSE (Root-mean-square error) and estimated residual standard error $\hat{\sigma}$ (also referred to as "*RSE*"). Highlighted cells (green) are the best-performing models with respect to the corresponding performance measure.

even for small clusters.

Often this may indicate over-fitting, especially due to the individual clustering effects and number of interaction factor levels. However, as stated - apart from the BIC criteria - since in this study we take into account all performance measures (in terms of goodness of fit and prediction accuracy); `lm.I.r` we deem the interaction model fit as our preferred and selected "final" linear model. We now further assess the preferred model via model diagnostics and residual analysis.

4.1.1 Model Diagnostics and Residual Analysis

Next, we assess the goodness of fit by first checking the distribution assumptions of the selected normal linear model (`lm.Ir1`). Recall the LM assumptions (summarized below).

(A1) Linearity. The response variable LR_{kij} is expressed by a linear combination of the covariates $x_{i1} \dots x_{ip}$, and includes the error random variable ε_i with mean zero,

(A2) Independence. The errors ε_i are independent random variables,

(A3) Variance Homogeneity. The error terms ε_i are random variables with constant variance,

$$Var(Y_i) = Var(\varepsilon_i) = \sigma^2.$$

(A3) Normality. The errors are normally distributed random variables.

Based on these assumptions, we can assess the validity of the assumptions through residual plots and the presence of outliers. From the model diagnostics plots given in Figure 4.4, we check if there are any violations against our linear model assumptions.

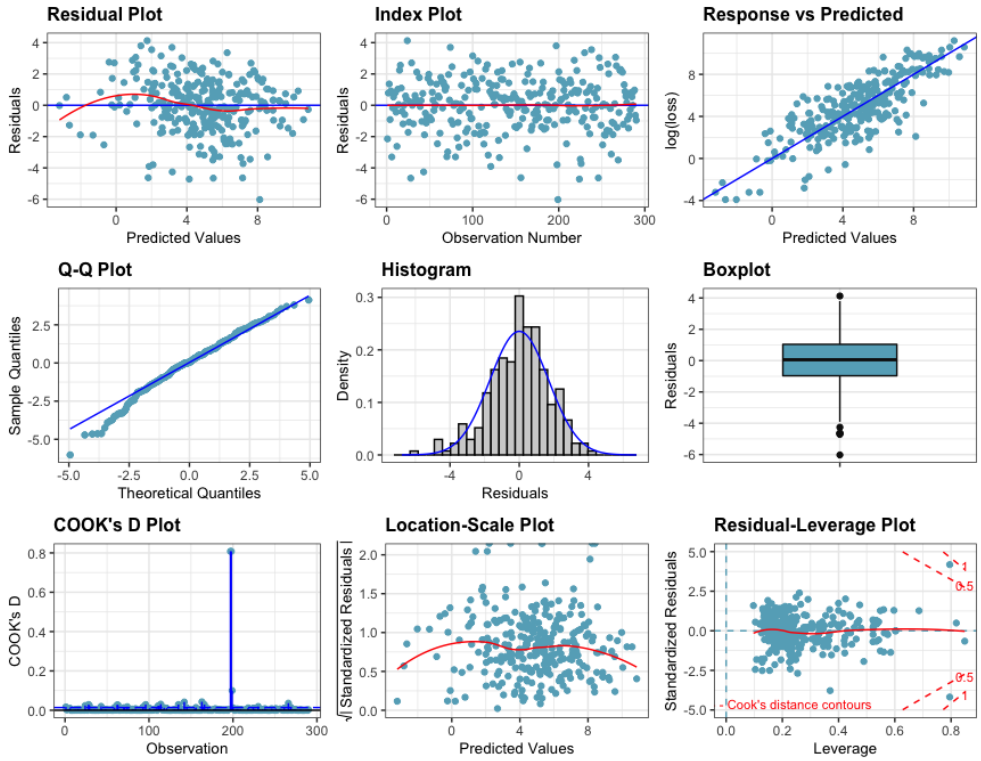


Figure 4.4. Model Diagnostics of selected linear model, `lm.I.r` (Interaction Effects), based on the response `ln.lr`. The residual plots are obtained using package `lme4`. Red lines represent the model fit, while blue lines represent the benchmarks of our LM regression assumptions. Note, lower panels shows outlier detection plots.

Specifically, based on the results, we check if there are any strong violations against the assumptions of the normal (log) linear model - with respect to the selected linear model, `lm.I.r` (with interaction effects). Here, the red lines in these residuals plots are based on the model fit, while the blue lines

represent the benchmarks of our normal regression assumptions (i.e. an appropriate fit given the above assumptions).

Hence, the corresponding examination and analysis of the residuals are summarized as follows:

- To check linearity assumptions, we look at the residuals versus the predicted values (or fitted) of `lm.Ir1`. The results show that there are no strong indications of a non-linear relationship. Similarly, the predicted versus the residual plot shows no indications against our variance homogeneity assumptions, as the predicated values are randomly scattered below and above zero.
- This is also evident when we look at the standardized residuals versus predicted values plot (Row 3, Column 2). Though, this is expected due to the study's design. Specifically, since the study focuses on non-life insurance longitudinal data, we have repeated measures for the given risk classes (k, i, j, t). This may also suggest correlations between the observations of the same group of risk classes and further motivates investing in the inclusion of random effects in the model.
- From the residuals versus predicted values, there are no strong indications against the assumptions of variance homogeneity (residuals are distributed equally above and below the origin, and there are no evident structural relationships).
- Additionally, the residuals' distribution also shows significant large values of around the mean 0. As previously discussed, this may arise due to the low frequency of losses per individual client, and given that not all clients have incurred losses for the risk measure (i.e. given the interaction effects) included in this model.
- However, we also observe that from the histogram (of the residuals), the distribution of the residuals appears to mainly follow a normal distribution (i.e. no strong assumptions against our log-normal distributional assumptions).
- The above is also evident when we look at the corresponding Q-Q plots. Since there are no strong violations of the normality assumption, though there exist several deviations away from the theoretical fit (in the lower tail). For this reason, we reassess the fit by removing the outliers in the data.
- In addition, based on the outliers plots (Cook's D plots and residual-leverage plots), a small number of high leverage loss observations exist. Notably, in the raw data, a small number of observations are over the threshold of $2p/n$.
- We can identify high leverage points (exceeding the threshold) and refit the model without the observations. Figure 4.5 and Figure 4.6 shows the distribution of the model residuals of the response variable (fitted log loss ratios `ln.lr`), when the detected outliers were removed from the data and model `lm.I.1` was refitted.

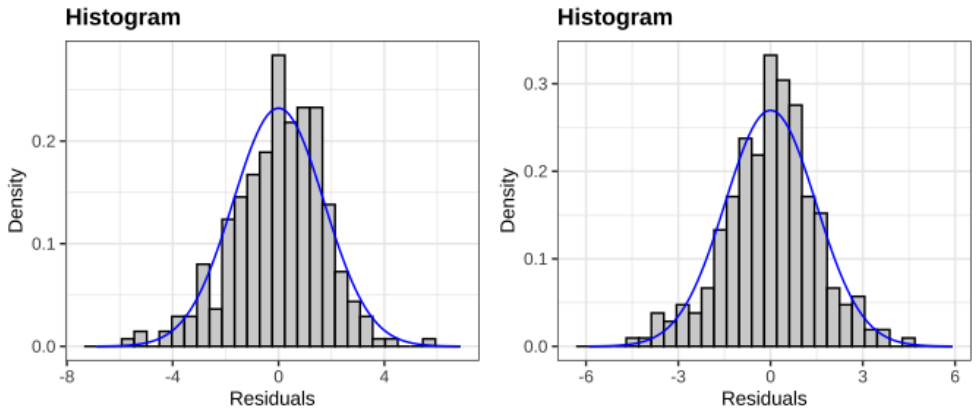


Figure 4.5. Histograms of residual plots based on model `lm.I.r`. Left plot shows the empirical histogram of the model residuals without the outliers. Right plot shows the empirical histograms of the residuals ratios after all 16 outliers were removed from the data set. The blue line represents the normal distribution fit.

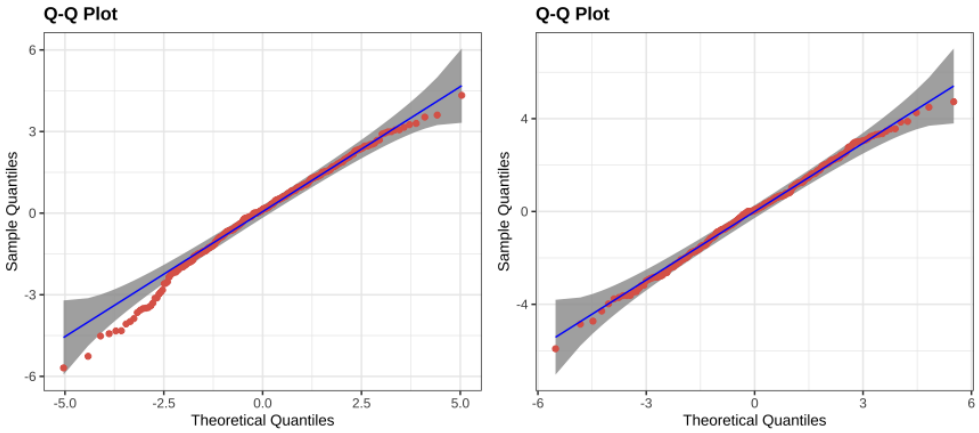


Figure 4.6. Q-Q plots, theoretical versus sample quantiles - comparison of LM Model residuals (with and without outliers). Left Column shows the residuals of model `lmm.I.1` fit without outliers removed. Right Column residuals of model `lmm.I.1` fit with outliers removed.

In total, 16 outliers were detected and removed from the data. From the histograms, we can see that a normal distribution is more evident compared to the fit of the raw data with outliers. Similarly, based on the Q-Q plots, there is no strong evidence against the normality assumptions (with no deviations away from the theoretical fit in the lower tail) - when the outliers were removed. For this reason, we consider either removing or imputing these outliers in our next section (with LMMs).

4.2 Linear Mixed Models for Natural Catastrophic Data

This section investigates and fits the (aggregated) natural catastrophic data using linear mixed models (LMMs) - suitable for longitudinal data. This type of model is an alternative approach and a balance between the "complete pooling" method (`lm.M`) -which did not take into account the clustering in the risk class groups - and "no pooling" (`lm.G`) - which tend to overfits or results in unreasonable parameter coefficients (Frees et al., 2014). Since generally a LMM combines or "mixes" the *"fixed"* and

"random" effects as covariates to model the response. LMMs are a more suitable approach given the longitudinal or (hierarchical structured) data to include more features helpful in this context (than the no or complete pooling LMs). Especially because, in our case, client-specific intercepts are a more meaningful alternative to the no pooling model (`lm.G`) while still allowing for heterogeneity between all the risk classes (k, i, j, t) in this study.

In this section, we follow the typical *"top-down"* approach for analyzing the LMMs - to ultimately find the best LMM that estimates log loss ratio with high precision. This means we include our findings from the LMs and follow the respective steps:

1. **Mean Structure.** Include the significant fixed effects on the average loss ratio (at 5% significance level) - with all significant interactions effects included in our selected LM (i.e., in Model (5.1)).
2. **Random effects.** Using the likelihood ratio tests (LRT), we statistically test if the inclusion of significant random effects improves the model fit. Based on this, we select the preferred model with random effects.
3. **Residual Covariance Structure.** By testing the LMM models with different residual covariance structures (compound symmetry versus unstructured residual covariance structure) - we aim to capture more variability unexplained in our model (by the fixed and random effects).
4. **Model Reduction and Selection.** Then, we try to reduce the model, i.e. reduce the number of fixed effects included in the model (main and interaction terms).

Here we are interested in the including k-th client specific effects: observed for a client $k, k = 1, \dots K$, (level 1), with (grouped) treaty i (level 2) insured in (grouped) country j (level 3) during the treaty years t (level 4). Note (similar to LMs) we only also considered ungrouped treaty years $t \in \mathcal{T}_{kijt}$ in order to investigate the effect on the loss ratio based on the underwriting or treaty year here. Thus, similar to our analysis for LMs, in this section, we first fit the different LMMs considered and then analyze the fit with model diagnostics on the preferred selected model.

Therefore based on the aggregated data, we fit the following three types of log-normal linear mixed models in this chapter, respectively:

Model `lmm.RI.0` *("random client intercept only")*: random intercepts model only; such that there is no random slope or fixed effects, and only includes the random client intercepts for each client $k, k = 1, \dots, 35$ (i.e. ***"null model"*** with one random intercept per k-th client). This means here we investigate if random client effects are required in this study.

Generally, this basic model can be formulated as

Model `lmm.RI.0` *(random client intercept effects only, "null model")*

$$\begin{aligned} \texttt{ln.lr}_{kijt} &= \beta_0 + \alpha_k + \varepsilon_{kijt}, \\ \varepsilon_{kijt} &\sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad i.i.d., \quad \forall\, k, i, j, t \\ \alpha_k &\sim \mathcal{N}(0, \sigma_\alpha^2) \quad i.i.d., \quad \forall\, k = 1, \dots, 35, \end{aligned} \tag{4.11}$$

where the ***random terms*** are given in blue and the ***fixed effects*** are given in red, for all $k = 1, \dots, 35$ with risk class levels (i, j, t).

Each component here in the model is described by the following:

- $\texttt{ln.lr}_{kijt}$ is the ***log loss ratio*** - which is given by ***the average loss per unit of exposure volume (sum insured)*** - for each k-th client in $k = 1, \dots 35$. With a corresponding (grouped) treaty i, $i \in \{1^k, 2^k\}$ providing coverage in (grouped) country $j, j \in \{H^{ki}, L^{ki}\}$, for treaty years $t \in \mathcal{T}^{kij}$. Recall, this includes all losses observed in the years 2001^{kij} to 2020^{kij} for all (k, i, j) groups (with no losses observed in the t-th year 2002^{kij} for all 35 clients).
- β_0 denotes the (unknown) intercept regression parameter for the ***fixed effects***.
- α_k denotes the ***random client intercept*** (known random effects parameters), which is a random variable with mean zero and variance σ^2_α. Here σ^2_α represents the variation between the clients $k = 1 \dots, K$.
- ε_{kijt} denotes the random error terms which are assumed to follow a normal distribution with mean 0 and variance σ^2_ε. Where σ^2_ε represents the variability structure within a client k.

In addition, we also explore the inclusion of weights in the model . Based on these findings, we investigate if the inclusion of weights improves our model (labelled as `lmm.R.0.w`.

Model `lmm.RI.M` *("random client intercepts with main fixed effects")*: fixed effects with random intercepts for random client k effects. Based on our findings from the random intercept model `lmm.RI.0`, we first investigate if by including the random client intercepts improve the model fit of the main fixed effects in model `lm.G.r`. Thus, this extended model here which combines the ***main fixed effects and random client effects*** - referred to as "`lmm.RI.M`" can now be formulated as,

Model `lm.RI.M` *(random client intercept effects with fixed main effects)*

$$\begin{aligned}\texttt{ln.lr}_{kijt} = \beta_0 + \beta_1 I_k^{\texttt{g.treaty}_i} + \beta_2 I_{ki}^{\texttt{g.country}_j} + \sum_{t=2}^{18} \beta_3^t \texttt{years}_{kij}^t + \sum_{q=2}^{4} \beta_4^q \texttt{t.type}_{ki}^q \\ + \ \beta_5 \texttt{oci}_{kijt} + \beta_6 \texttt{rate}_{kijt}^d + \beta_7 \texttt{cpi}_{kijt} + \alpha_k + \varepsilon_{kijt},\end{aligned} \tag{4.12}$$

such that

$$\begin{aligned}\varepsilon_{kijt} &\sim \mathcal{N}(0, \sigma^2_\varepsilon) \quad i.i.d., \quad \forall\ k, i, j, t \\ \alpha_k &\sim \mathcal{N}(0, \sigma^2_\alpha) \quad i.i.d., \quad \forall\ k = 1, \dots, 35,\end{aligned} \tag{4.13}$$

for the random components in the model (in blue) given the fixed main effects (in red). Note in this model the risk class effects `g.treaty`, `g.country` and `years` are still introduced as fixed main effects. While α_k still represents the random client k specific intercepts (with varying σ^2_α).

Model `lmm.RI.I` *("random client intercepts with main and interaction fixed effects")*: In the case that we find that by accounting for random intercept client effects improves the overall model fit of `lm.G.r`); we then allow for interaction effects. Specifically, we test if the LMM fit given in `lmm.RI.M` improves with interaction effects (given the random client intercepts for $k = 1 \dots, 35$).

Note, we only allow for the interaction effects included in the reduced simple linear Model (4.22), selected in Section 4.1, and labelled as `lm.I.r` (with only fixed effects, significant at a 5% level).

Hence the regression equation is

Model `lmm.RI.I` *(random intercept and slope effects with fixed main effects)*

$$\begin{aligned}\texttt{ln.lr}_{kijt} = \beta_0 + \beta_1 I_k^{\texttt{g.treaty}_i} + \beta_2 I_{ki}^{\texttt{g.country}_j} + \sum_{t=2}^{18} \beta_3^t \texttt{years}_{kij}^t + \sum_{q=2}^{4} \beta_4^q \texttt{t.type}_{ki}^q \\ + \beta_5 \texttt{oci}_{kijt} + \beta_6 \texttt{rate}_{kijt}^d + \beta_7 \texttt{cpi}_{kijt} + \sum_{t=2}^{8} \beta_9^t \texttt{years}_{kij}^t \times \texttt{cpi}_{kijt} \\ + \sum_{q=2}^{4} \beta_{10}^q \texttt{t.type}_{ki}^q \times \texttt{oci}_{kijt} + \beta_{11} \texttt{rate}_{kijt} \times \texttt{cpi}_{kijt} + \alpha_k + \varepsilon_{kijt},\end{aligned} \tag{4.14}$$

where,

$$\begin{aligned}\varepsilon_{kijt} &\sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad i.i.d., \quad \forall\, k, i, j, t \\ \alpha_k &\sim \mathcal{N}(0, \sigma_\alpha^2) \quad i.i.d., \quad \forall\, k = 1, \ldots, 35,\end{aligned} \tag{4.15}$$

where similarly, the random client intercepts α_k are independent across all clients $k, k = 1, \ldots, 35$, and independent of the error terms ε_{kijt}, for all risk classes (k, i, j, t).

Model `lmm.RS`: here, we consider different structures of the random effects: multiple random effects per level, nested random effects and crossed random effects. This essentially means we allow for both random intercepts and random slopes into the model - in addition to fixed main or interaction effects. Specifically, we consider random intercepts and slope effects which may vary by the appropriate risk class levels: k, i, j, t, based on the k-th client level.

For the fixed effects, once again, we only include significant main or interaction effects (at the significance level of 5%) with respect to our findings in the previous step (or model `lmm.RI.I`). For example, suppose we allow for the random risk class effects: (grouped) treaties i as random slopes and random client k effects as random intercepts (i.e. multiple random effects per level).

If we consider only the fixed main and interaction effects included in the model `lmm.RI.I`, then this model equation is given by

Model `lmm.RI.S` *(random intercept and slope effects and interaction fixed effects)*

$$\begin{aligned}\texttt{ln.lr}_{kijt} = \beta_0 + \beta_1 I_k^{\texttt{g.treaty}_i} + \beta_2 I_{ki}^{\texttt{g.country}_j} + \sum_{t=2}^{18} \beta_3^t \texttt{years}_{kij}^t + \sum_{q=2}^{4} \beta_4^q \texttt{t.type}_{ki}^q \\ + \beta_5 \texttt{oci}_{kijt} + \beta_6 \texttt{rate}_{kijt}^d + \beta_7 \texttt{cpi}_{kijt} + \sum_{t=2}^{8} \beta_9^t \texttt{years}_{kij}^t \times \texttt{cpi}_{kijt} \\ + \sum_{q=2}^{4} \beta_{10}^q \texttt{t.type}_{ki}^q \times \texttt{oci}_{kijt} + \beta_{11} \texttt{rate}_{kijt} \times \texttt{cpi}_{kijt} \\ + \alpha_{0k} + \alpha_{1k} I_k^{\texttt{g.treaty}_i} + \varepsilon_{kijt},\end{aligned} \tag{4.16}$$

such that here, α_{0k} and α_{1k} are the random intercept and random client slope effects for the the k-th client, while β_0 and $\beta_1 \ldots, \beta_{11}$ are the overall fixed intercept and regression parameters

with random error ε_{kijt} for each k-th client at risk class index (i, j, t). Note in this model our distribution assumptions for the random effects differ from the Model (4.14). Since in this model our random intercept and slope vary on different risk class levels - this means that the variance are given on different levels. Here we assume that for the same k-th client, the random client intercept and random grouped treaty slope effects are correlated. While the random effects are still independent across risk classes, for all clients $k, k = 1 \ldots, 35$, and independent of the random error terms ε_{kijt}. Hence, together, for this model we have the following distribution assumptions for the random effects

$$\begin{bmatrix} \alpha_{0k} \\ \alpha_{1k} \end{bmatrix} \sim \mathcal{N}_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix} \right) \quad i.i.d., \quad \forall\ k = 1, \ldots, 35, \tag{4.17}$$

and for the random error terms we have

$$\varepsilon_{kijt} \ \sim \ \mathcal{N}(0, \sigma_\varepsilon^2) \qquad i.i.d., \quad \forall\ k, i, j, t, \tag{4.18}$$

such that for this model, we also assume an unstructured covariance matrix of the random effects - given by D for simplicity,

$$D = \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}, \tag{4.19}$$

where the vector of the random client effects, denoted by $\boldsymbol{\gamma}_k = (\alpha_{0k}, \alpha_{1k})\prime$, is now bivariate with $\text{Var}(\alpha_{0k}) = \sigma_0^2$, $\text{Var}(\alpha_{1k}) = \sigma_1^2$, and $\text{Cov}(\alpha_{0k}, \alpha_{1k}) = \sigma_{01}$. Note, here for the grouped treaty i, given by

$$I_k^{\texttt{g.treaty}_\texttt{i}} = \begin{cases} 1, & \text{if loss observation } n_{kijt} \text{ belongs to a client } k\text{-th who has more than 1} \\ & \text{treaty } i \text{ for countries } j \in \boldsymbol{\mathcal{M}}_G^{ki} \text{ during years } t \in \boldsymbol{\mathcal{T}}^{kij}, \\ 0, & \text{otherwise.} \end{cases} \tag{4.20}$$

now enters our model as both a random and fixed effect for the k-th client.

Then, based on the results from the models above and after, we then find the best performing model given both the random and fixed effects (with main and interaction effects) of the risk measures - over all risk classes (k, i, j, t) - we investigate for different residual covariance structures to improve the fit.

Model `lmm.RI.0`: Random Client Intercepts Model Only

In this section, we compare the unconditional null or **"basic model"** with only random intercepts (at the k-th client subject level) and the response (loss ratios) without weights, given in model Equation 4.11. These models consider only the first risk class level (i.e. at the k-th client level).

Model Results: Random Client Intercepts Only - With No Weights (`lmm.RI.0`)

Analyzing the basic model allows us to determine if a multilevel model is appropriate given our data regarding the log-transformed ratio of losses. In addition to comparing the influence of the weights.

To fit this LMM, in `R` we use the package `lme4`. By default, the package uses the REML, such that, alternatively, we can set `REML = false` to obtain the maximum likelihood (ML) results.

```r
lmm.RI.0 <- lmer(ln.lr ~ 1 + (1|client), data = cat_data, REML = FALSE)
summary(lmm.RI.0)
```

```
Linear mixed model fit by maximum likelihood . t-tests use  ['lmerModLmerTest']
Formula: ln.lr ~ 1 + (1 | client)
   Data: cat_data

     AIC      BIC   logLik deviance df.resid
    2047     2058    -1020     2041      287

Scaled residuals:
   Min     1Q Median     3Q    Max
-5.005 -0.244  0.134  0.561  2.162

Random effects:
 Groups   Name        Variance Std.Dev.
 client   (Intercept) 14.0     3.74
 Residual             59.6     7.72
Number of obs: 290, groups:  client, 35

Fixed effects:
            Estimate Std. Error     df t value Pr(>|t|)
(Intercept)   -9.223      0.854 39.766   -10.8  2.2e-13 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1
```

From the results of "basic model" (with no fixed effects and only random client intercepts) we can estimate the following parameters: $\hat{\beta}_0$, $\sigma^2_\varepsilon, \sigma^2_\alpha$. In this output $\hat{\beta}_0 = -9.22$ (fixed intercept, with s.e. 0.85), $\hat{\sigma}^2_\alpha = 14.0$ (variance for the random client effects) and $\hat{\sigma}^2_\varepsilon = 59.6$. Note here for the fixed effects for this LMM, in `R`; we utilize Satterthwaite's method for approximating degrees of freedom for the t and F tests (instead of the LRT, using the package `lmerTest`). For more details on this approximation method, see Kuznetsova et al. (2017).

We can also obtain single components of the summary output for the random components such as the estimated variance using the function `VarCorr()` in `R`:

```r
VarCorr(lmm.RI.0)
```

```
# Estimated variances, standard deviations, and correlations between the random-effects.
 Groups   Name        Std.Dev.
 client   (Intercept) 3.74
 Residual             7.72
```

Here, the random effect standard deviation in the first model ($\hat{\sigma}_\alpha = 3.74$) indicates that there exists some variability around the predicted mean, even after accounting for the k-th client. This result suggests that there may be additional (random) effects or grouping structure between the risk class levels (k, i, j, t) - we should potentially introduce into the model.

We can further confirm these findings by calculating the ***"intraclass coefficient"*** (known as the "**ICC**", see Definition 4.5). Recall that the ICC (similar to the R^2 in LMs) measures the proportion of the variance explained by the client k-th grouping structure in the data-set. We can estimate all sources of uncertainty in the mixed model `lmm.RI.0` by the two types of ICC: (Nakagawa et al., 2017):

Adjusted ICC (proportion of the total variance explained by the random client effects $\hat{\alpha}_0$):

$$ICC_{adj.} = \frac{\hat{\sigma}^2_\alpha}{\hat{\sigma}^2_\alpha + \hat{\sigma}^2_\varepsilon} = \frac{14}{59.6 + 14} = 0.19.$$

Conditional ICC (proportion of the total variance explained by both the random client effects $\hat{\alpha}_0$ and mixed-effects $\hat{\beta}_0$):

$$ICC_{con.} = \frac{\hat{\sigma}^2_\alpha}{\hat{\sigma}^2_\alpha + \hat{\sigma}^2_\varepsilon + \hat{\sigma}^2_\beta} = \frac{14}{59.6 + 14 + 0} = 0.19,$$

where $\hat{\sigma}^2_\beta$ denotes the estimated variance of the fixed effects β_0. In this case because no fixed effects were fitted (except for the fixed intercept) we set $\hat{\sigma}^2_\beta = 0$ for the conditional ICC. This implies for `lm.R.0` we have $ICC_{adj.} = ICC_{con.}$.

Alternatively we can obtain both `ICC` from `R` using the `performance::icc()` function,

```
performance::icc(lmm.RI.0)
```

```
# Intraclass Correlation Coefficient
  Adjusted ICC: 0.190
  Conditional ICC: 0.190
```

Thus, based on the ICC of the null model (with intercept only), 19% of the variance in the estimated mean log loss ratio is accounted for by the k-th client-specific effects (without any fixed effects). Therefore, this suggests that the multilevel model may be appropriate.

Recall, to statistically test the significance of the random client effects α_k, we use the likelihood ratio test with respect to the variance components (see Definition 4.8). Here we investigate whether the different intercepts per client k are significantly different or vary and should be included in the model (at a 5% significance level), by testing:

$$H_0 : \sigma^2_\alpha = 0 \qquad \text{versus} \qquad H_0 : \sigma^2_\alpha > 0,$$

using the log-likelihood ratio test (LRT). Note, the LRT test for nested models is approximately χ^2_α distributed (as previously described in Chapter 2). Through the use of the function `ranova` in `R`, we can perform the LRT to test if the structure of the random effects should be included or not (via using the ML or REML if set to `REML = TRUE`).

```
ranova(lmm.RI.0)
```

```
ANOVA-like table for random-effects: Single term deletions

Model:
ln.lr ~ (1 | client)
             npar logLik  AIC  LRT Df Pr(>Chisq)
<none>          3  -1020 2047
(1 | client)    2  -1044 2091 46.2  1    1.1e-11 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Based on the results, we reject the null hypothesis (p-value < 0.05) and conclude that our model's random client effects α_k are necessary (random intercepts), given our data and to estimate the average estimated log loss ratio for risk classes (k, i, j, t).

Random Client Intercepts Only - With weights (`lmm.RI.0.w`)

As previously discussed, the study models the average loss given per unit of exposure volume (per sum of insured amount in thousands). The average loss here is constructed as the total loss - per client

k, treaty i, country k, per t year - divided by the corresponding number of losses (adjusting for the exposure). Therefore, the average loss is more precise for a client when more loss counts (per risk class level) have been observed.

For this reason, here we introduce the use of weights as the number of insured (unique and historic) natural disasters per client k. We first test if the inclusion of weights - since the historic number of natural disasters events insured per k client - improves our model to estimate the average loss ratio. The benefits of including weights in actuarial models have been previously discussed; see Frees et al. (2014).

The basic or null LMM (`lmm.RI.0.w`) with weights can now be reformulated as,

Model `lmm.RI.0` (random client intercept only model with weights)

$$\begin{aligned} \texttt{ln.lr}_{kijt} &= \beta_0 + \alpha_k + \varepsilon_{kijt}, \\ \varepsilon_{kijt} &\sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad i.i.d., \quad \forall\, k, i, j, t \\ \alpha_k &\sim \mathcal{N}(0, \sigma_\alpha^2 / w_k) \quad i.i.d., \quad \forall\, k = 1, \ldots, 35. \end{aligned} \tag{4.21}$$

Here w_k denotes the weights where w_k is given by the number of natural loss events observed per k (`oci`).

Model Results: `lmm.RI.0.w`

Using the function `weights` in `R`, we now investigate if the basic model fit is improved by including the weights.

```
lmm.RI.0.w <- lme4::lmer(ln.lr ~  (1|client), weights = oci, data = cat_data, REML = FALSE)
summary(lmm.RI.0.w)
```

```
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: ln.lr ~ 1 + (1 | client)
   Data: cat_data
Weights: oci

     AIC      BIC   logLik deviance df.resid
    2221     2232    -1108     2215      287

Scaled residuals:
   Min     1Q Median     3Q    Max
-5.870 -0.071  0.198  0.602  2.382

Random effects:
 Groups   Name        Variance Std.Dev.
 client   (Intercept)  38.7     6.22
 Residual             143.7    11.99
Number of obs: 290, groups:  client, 35

Fixed effects:
            Estimate Std. Error t value
(Intercept)   -11.21       1.34   -8.36
```

In this output $\hat{\sigma}_\alpha^2 = 38.7$ and $\hat{\sigma}_\varepsilon^2 = 143.7$. Based on the output, we can also calculate the adjusted $ICC_{adj.}$ (or alternatively (by the `icc` function from the `performance` package in `R`). Recall, this is calculated by dividing the variance of the random effects, for the k-th client, $\hat{\sigma}_\alpha^2$ by the total variance

estimated by `lmm.RI.0.w`,

$$ICC_{adj.} = \frac{\hat{\sigma}_\alpha^2}{\hat{\sigma}_\alpha^2 + \hat{\sigma}_\varepsilon^2} = \frac{38.7}{38.7 + 143.7} = 0.22.$$

Since, here, we still don't allow for any fixed effects, we have $ICC_{adj.} = ICC_{con.}$ with $\hat{\sigma}_\beta^2 = 0$.

Now we compare this model with `lmm.RI.0` (no weights) in Table 4.9. From here we can see that after including for weights `oci` - the log-likelihood, the ICC, the AIC, AICc and BIC - significantly increases (from AIC 2046.85 with log likelihood -1020, without weights, to AIC = 2221.15 and log likelihood -1108). Table 4.9 also provides the ***conditional*** R^2 (denoted as $\boldsymbol{R^2_{con.}}$) which is proportion of variance explained by the "complete" model (considering both fixed and random effects) and the corresponding marginal R^2 which only indicates how much of the variability in the model is explained by the fixed effects (denoted by $\boldsymbol{R^2_{mar.}}$).

From these results, we see that even though there is a slight increase in the variance explained by the random and fixed effects when the weights are accounted for ($R^2_{con.}$), in general - based on the AIC, BIC, RMSE and residual standard error (SE) - `lmm.RI.0` outperforms `lmm.RI.0.w`.

Model	p	logLik	AIC	AICc	BIC	$R^2_{con.}$	$R^2_{mar.}$	$ICC_{adj.}$	$ICC_{con.}$	RMSE	RSE
`lmm.RI.0`	3	-1020.42	2046.85	2046.93	2057.86	0.19	0.00	0.19	0.19	7.46	7.72
`lmm.RI.0.w`	3	-1107.57	2221.15	2221.23	2232.16	0.22	0.00	0.22	0.21	8.18	11.99

Table 4.9. Model performance comparison of random intercept models fit with random k client effects, $k = 1, \ldots, 35$; comparing the results of the performance `lmm.RI.0` (without weights) versus `lmm.RI.0.w` (with weights). Based on the log-likelihood ("logLik") statistic,AIC, AICCc, BIC, the conditional R^2 (based on both fixed and random effects), marginal R^2 (fixed effects only), the ICC (adjusted and conditional), RMSE and the residual standard error("RSE"). Here p denotes the number of parameters in each model.

Additionally, we can also compare the conditional modes of the random effects (the best linear unbiased predictions, BLUPs, in the case of LMMs) estimated under both mixed models. From our results, we note that the 95% prediction intervals are wider in general for `lmm.RI.0.w` (expression more model uncertainty for each client k) in comparison to `lmm.RI.0`. For this reason, moving forward, we should compare the fitted models without weights (see Table 4.9 for comparison) and with the varying effects of clients (as random intercepts).

Model `lmm.RI.M`: Random Client Intercept with Fixed Main Effects

In this section, we briefly investigate if introducing the risk class client effects as a random effect improves the fit of Model (4.8). In other words, if the "classical" linear model's fit, `lm.M.r` (which includes the risk group effects as individual fixed effects and the main fixed effects) is significantly improved. We start by building the LMM with only the main fixed effects included in the reduced LM `lm.M.r` (significant on the estimated log loss ratios, at $\alpha = 0.05$) for each k-th client. Then if the LMM outperforms the LM, we continue further with our analysis with LMMs and investigate if we can further reduce the LMM or if the interaction effects given in Model (5.1) improve our LMM fit.

Here we still consider each risk class group in: treaty i (grouped), country j (grouped), and treaty year t as explanatory variables (fixed effects). Whereas the effect of each k-th client is now included as a

random effect (random intercept, $\boldsymbol{\alpha}_k$). Recall this LMM was formulated as:

$$\begin{aligned}\textbf{Model lmm.RI.M:}\quad \texttt{ln.lr}_{kijt} &= \beta_0 + \beta_1 I_k^{\texttt{g.treaty}_i} + \beta_2 I_{ki}^{\texttt{g.country}_j} + \sum_{t=2}^{18} \beta_3^t \texttt{years}_{kij}^t + \sum_{q=2}^{4} \beta_4^q \texttt{t.type}_{ki}^q \\ &\quad + \beta_5 \texttt{oci}_{kijt} + \beta_6 \texttt{rate}_{kijt}^d + \beta_7 \texttt{cpi}_{kijt} + \alpha_k + \varepsilon_{kijt}, \\ \varepsilon_{kijt} &\sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad i.i.d., \quad \forall\, k, i, j, t, \\ \alpha_k &\sim \mathcal{N}(0, \sigma_\alpha^2) \quad i.i.d., \quad \forall\, k = 1, \ldots, 35,\end{aligned} \tag{4.22}$$

such that we assume the random client effects are independent for each client k with independent error terms, for each $k = 1, \ldots, 35$.

Model Results: `lmm.RI.M`

Once again, we fit this model `lmm.RI.M` in R using ML (instead of REML - by setting `REML = FALSE`). To fit this LMM by using the package `lme4` in R, we use the following command (where blue denotes the random client effects)

```
1 lmm.RI.M <- lmer(ln.lr ~  g.treaty + g.country + years + t.type + oci + rate + cpi + (1|client), na.action=
      na.omit,data = cat_data, REML = F)
```

```
      Linear mixed model fit by maximum likelihood . t-tests use Satterthwaites method ['lmerModLmerTest']
Formula: ln.lr ~ g.treaty + g.country + years + t.type + oci + rate +      cpi + (1 | client)
   Data: cat_data

     AIC      BIC   logLik deviance df.resid
    1741     1847     -841     1683      261

Scaled residuals:
   Min     1Q Median     3Q    Max
  -4.0   -0.5    0.1    0.6    3.6

Random effects:
 Groups   Name        Variance Std.Dev.
 client   (Intercept)  4       2
 Residual             17       4
Number of obs: 290, groups:  client, 35

Fixed effects:
            Estimate Std. Error     df t value Pr(>|t|)
(Intercept)    14.84       4.21 288.90     3.5    5e-04 ***
g.treaty2      -0.46       0.98  40.05    -0.5    0.637
g.countryL      0.47       0.85 212.17     0.6    0.582
years2003      -0.52       3.57 278.59    -0.1    0.883
years2004       0.66       3.19 270.32     0.2    0.837
years2005       0.47       3.26 266.45     0.1    0.884
....
t.typeCXL      -0.87       1.28 236.86    -0.7    0.499
t.typeQS       -2.71       1.00 189.41    -2.7    0.008 **
t.typeSP       -2.74       1.26 187.80    -2.2    0.032 *
oci            -5.69       0.28 278.49   -20.7   <2e-16 ***
rate           -0.84       0.19 256.74    -4.5    9e-06 ***
cpi            -0.09       0.03 288.96    -3.3    0.001 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Here, $\hat{\beta}_0 = 14.84$ (s.e. 4.21), with $\hat{\sigma}^2_\alpha = 4.21$, and $\hat{\sigma}^2_\varepsilon = 0.70$ (a significant decrease from the first "null" LMM, `lmm.RI.0`, and selected LM model, `lm.I.r`).

First of all, if we set the significance level to 5%, the results show that fixed main effects of all the risk factors, `t.type`, `oci`, `rate` and `cpi` remain significant. However, when introducing the random client intercepts into the model, we observe the following risk class groups: `g.treaty`, `g.country` and `years` are no longer significant at a 5% significance level. To simplify the fixed-effects structure, we construct an ANOVA table for the model `lmm.RI.M`, calculate F statistics and corresponding p-values for each fixed-effects dropped from the model (using the F test using the Satterthwaite approximation method, which is compatible for our unbalanced data set where not all clients have the counts of loss). The results are provided in Table 4.10. Alternatively, this can be obtained by the function in `drop1` in `R` package `lmer`.

	Sum Sq	Mean Sq	Num. Df	Den. Df	F-value	Pr($>F$)
`g.treaty`	4	4	1	40	0.23	0.64
`g.country`	5.3	5.3	1	212	0.3	0.58
`years`	913	51	18	279	2.9	1e-04
`t.type`	140	47	3	215	2.7	0.048
`oci`	7449	7449	1	278	427	<2e-16
`rate`	358	358	1	257	21	9e-06
`cpi`	193	193	1	289	11	0.001

Table 4.10. ANOVA table (Type III) for fixed effects. Based on model `lmm.RI.M`, with F-tests for the main fixed-effects (using the Kenward-Roger F-tests with Satterthwaite degrees of freedom) and corresponding p-value and their order of elimination. The results are provided in here are obtained by the package `lmerTest` based on the model fit with `lme4`. Here *"Num. Df"* is the numerator degrees of freedom (number of groups minus one), and *Den. DF* is the denominator degrees of freedom (difference between the number of groups and observations).

The results show after removing `g.treaty` and `g.country` sequentially that `years` is now at a 5% significance level (p-value <0.05), when accounting for the random client k intercept effects. Hence, we conclude that the fixed effects of the group risk class levels `g.treaty` and `g.country` can be dropped from the model. From the results we also see that the following fixed main effects of the risk factors remain significant on the average log loss ratios (at p-value < 0.05): `t.type`, `oci`, `rate` and `oci` - when including the random client effects k into the model.

Model Comparison: Full Model (`lmm.RI.M`) and Reduced Model (`lmm.RI.M.r`)

Since both random clients k intercept mixed models were fitted using maximum likelihood estimation, we can use the likelihood ratio test (LRT) in `R` to test the significance of the fixed effects. In other words, by analyzing the ANOVA output, we can once again check if the model can be reduced by excluding `g.treaty` and `g.country` - by the following `R` command to perform the LRT.

```
1 anova(lmm.RI.M,lmm.RI.M.r1, test = "LRT")
```

```
Data: cat_data
Models:
lmm.RI.M.r1: ln.lr ~ years + t.type + oci + rate + cpi + (1 | client)
lmm.RI.M: ln.lr ~ g.treaty + g.country + years + t.type + oci + rate + cpi + (1 | client)
           npar  AIC  BIC logLik deviance Chisq Df Pr(>Chisq)
```

```
lmm.RI.M.r1   27 1737 1836   -842     1683
lmm.RI.M      29 1741 1847   -841     1683  0.51  2       0.77
```

The results shows that the null hypothesis that both `g.treaty` and `g.country` are not significant, $H_0 : \beta_1 = \beta_2 = 0$, cannot be rejected with, p-value $= 0.77$. Hence, we conclude that the model can be reduced by excluding both covariates.

The performance and accuracy of the reduced model (`lmm.RI.M.r`) is also compared with the full LMM main fixed effects model `lmm.RI.M` in Table 4.11.

Model	p	logLik	AIC	AICc	BIC	$R^2_{con.}$	$R^2_{mar.}$	$ICC_{adj.}$	$ICC_{con.}$	RMSE	RSE
`lmm.RI.M`	29	-841.40	1740.80	1747.49	1847.22	0.77	0.72	0.18	0.05	4.04	4.18
`lmm.RI.M.r`	28	-841.51	1739.02	1745.24	1836.40	0.77	0.73	0.17	0.05	4.05	4.18

Table 4.11. Comparison of performance of full LMM `lmm.RI.M` and the reduced LMM `lmm.RI.M.r`. Both models fit the random client intercept effects and fixed main fixed effects. Performance is compared by the information criteria (AIC,AICc,BIC), the log-likehood (logLik), the proportion of variability explained by each model (i.e. R^2 and ICC) and the residual standard errors (green cells denotes the preferred model per corresponding measure).

When examining the information criteria for the reduced and full model, we see that the reduced model `lmm.RI.M.r` attains the lowest AIC, AICc and BIC values. Though the reduced model results in a slight decrease in terms of the logLik, marginal $R^2_{mar.}$, $ICC_{adj.}$ and RMSE - we choose the simpler and less complex reduced model based on the lowest AIC, AICc and BIC values.

Whereas, if we compare the BIC values attained by `lmm.RI.M.r` to the best performing reduced LM `lm.G.r` (with fixed main effects, BIC = 1960.98) we also attain a better performing model fit with `lmm.RI.M.r` by accounting for the varying random k-th client effect. For this reason, we proceed further with our analysis with LMMs.

Thus, the reduced model (and selected model in this section) can be formulated as,

Model `lmm.RI.M.r` *(reduced model with random intercept and fixed main effects)*

$$\begin{aligned} \texttt{ln.lr}_{kijt} = \beta_0 + \sum_{t=2}^{18} \beta_1^t \texttt{years}_{kij}^t + \sum_{q=2}^{4} \beta_2^q \texttt{t.type}_{ki}^q \quad + \quad \beta_3 \texttt{oci}_{kijt} + \beta_4 \texttt{rate}_{kijt}^d \\ + \beta_5 \texttt{cpi}_{kijt} + \alpha_k + \varepsilon_{kijt}, \qquad (4.23) \\ \varepsilon_{kijt} \quad \sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad i.i.d., \quad \forall\, k, i, j, t \\ \alpha_k \quad \sim \mathcal{N}(0, \sigma_\alpha^2) \quad i.i.d., \quad \forall\, k = 1, \ldots, 35, \end{aligned}$$

Although, an adjusted $ICC_{adj.}$ value of 17% (with conditional $ICC_{con.}$) indicates a low reliable fit, such that unwanted variations may be sufficiently large between the k-th clients. For this reason, we now investigate if the model fit is improved by introducing interaction effects (before we start exploring different residual structures).

Model `lmm.RI.I`: Random Client Intercepts with Fixed Main and Interaction Effects

We now investigate if we can improve the overall fit of the model `lmm.RI.M.r` by introducing certain pairwise interactions. Recall that we only allow for the interaction effects included in the LM interaction model `lm.I.r` (the reduced normal linear model with interaction effects).

Recall `lmm.RI.I` with fixed interactions - with respect to the random k-th component ($\boldsymbol{\alpha}_k$) - was formulated as

$$\begin{aligned}\underline{\textbf{Model lm.RI.I:}}\ \ \texttt{ln.lr}_{kijt} = \beta_0 &+ \beta_1 I_k^{\texttt{g.treaty}_i} + \beta_2 I_{ki}^{\texttt{g.country}_j} + \sum_{t=2}^{18} \beta_3^t \texttt{years}_{kij}^t + \sum_{q=2}^{4} \beta_4^q \texttt{t.type}_{ki}^q \\ &+ \beta_5 \texttt{oci}_{kijt} + \beta_6 \texttt{rate}_{kijt}^d + \beta_7 \texttt{cpi}_{kijt} + \sum_{t=2}^{8} \beta_8^t \texttt{years}_{kij}^t \times \texttt{oci}_{kijt} \\ &+ \sum_{t=2}^{18} \beta_9^t \texttt{years}_{kij}^t \times \texttt{cpi}_{kijt} + \sum_{q=2}^{4} \beta_{10}^q \texttt{t.type}_{ki}^q \times \texttt{oci}_{kijt} + \beta_{11} \texttt{oci}_{kijt} \times \texttt{cpi}_{kijt} \\ &+ \alpha_k + \varepsilon_{kijt},\end{aligned} \tag{4.24}$$

with

$$\begin{aligned}\varepsilon_{kijt} &\sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad i.i.d., \quad \forall\, k, i, j, t, \\ \alpha_k &\sim \mathcal{N}(0, \sigma_\alpha^2) \quad i.i.d., \quad \forall\, k = 1, \ldots, 35,\end{aligned}$$

where the log loss ratio is given for the $k = 1, \ldots, 35$. The random effects are independent across the clients, $k = 1, \ldots, 35$, as well as independent of the error terrors ε_{kijt} for all risk class levels in (k, i, j, t).

Model Results: `lmm.RI.I`

Similarly, we use the package `lme4` with the function `lmer` to fit the above model. Where the `R` operator `":"` denotes pairwise effects between the fixed covariates.

```
lmm.RI.I <- lmer(formula = ln.lr ~ years + t.type  + oci + rate + cpi + years:oci + years:cpi + t.type:oci + oci:cpi +
    (1|client), data = cat_data, na.action = na.omit, REML = F)
summary(lmm.RI.M.int)
```

```
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaites method ['lmerModLmerTest']
Formula: ln.lr ~ years + t.type + oci + rate + cpi + years:oci + years:cpi +
    t.type:oci + oci:cpi + (1 | client)
   Data: cat_data

     AIC      BIC   logLik deviance df.resid
    1637     1860     -757     1515      229

Scaled residuals:
   Min     1Q Median     3Q    Max
-4.701 -0.550  0.027  0.543  2.415

Random effects:
 Groups   Name        Variance Std.Dev.
 client   (Intercept) 2.80     1.67
 Residual             9.57     3.09
Number of obs: 290, groups:  client, 35
```

```
Fixed effects:
              Estimate Std. Error     df t value Pr(>|t|)
(Intercept)      24.64      15.98 274.96     1.5    0.124
years2003        65.47      76.28 283.06     0.9    0.392
years2004       -32.83      16.44 282.45    -2.0    0.047 *
years2005       -51.43      17.68 276.57    -2.9    0.004 **
years2007       -18.09      16.93 280.83    -1.1    0.286
.....
t.typeCXL         5.74       1.97 276.07     2.9    0.004 **
t.typeQS         -1.36       1.27 181.66    -1.1    0.288
t.typeSP          3.11       2.04 267.17     1.5    0.129
oci               7.53       4.07 277.44     1.9    0.065 .
rate             -0.54       0.15 269.84    -3.7    3e-04 ***
cpi              -0.30       0.20 274.02    -1.5    0.144
years2003:oci    -1.43       5.64 287.06    -0.3    0.800
years2004:oci    -4.01       2.73 269.12    -1.5    0.142
years2005:oci    -3.29       2.93 270.42    -1.1    0.263
years2007:oci    -5.57       2.75 263.68    -2.0    0.044 *
....
years2003:cpi    -0.77       0.98 281.74    -0.8    0.434
years2004:cpi     0.47       0.21 281.40     2.2    0.030 *
years2005:cpi     0.67       0.22 275.57     3.0    0.003 **
....
t.typeCXL:oci    -4.43       1.23 282.70    -3.6    4e-04 ***
t.typeQS:oci     -0.46       0.56 289.91    -0.8    0.405
t.typeSP:oci     -4.12       1.39 280.63    -3.0    0.003 **
oci:cpi          -0.11       0.03 284.33    -4.2    3e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Here, for the estimates of the fixed effects we have $\hat{\beta}_0 = 24.64$ (s.e. $= 15.98$), with estimated random error $\hat{\sigma}^2_\varepsilon = 9.6$ and, for the random effects, $\hat{\sigma}^2_\alpha = 2.8$. Alternatively, we can extract the estimated variance explained by both the fixed effects, $\hat{\sigma}^2_\beta$, and random effects or errors - using R function `get_variance` (available in the package `insights`). This allows us to calculate both the adjusted $ICC_{adj.}$ and conditional $ICC_{con.}$ (see Table 4.12).

Model	$\hat{\sigma}^2_\beta$	$\hat{\sigma}^2_\alpha$	$\hat{\sigma}^2_\varepsilon$	$ICC_{adj.}$	$ICC_{con.}$	$R^2_{con.}$	$R^2_{mar.}$
`lmm.RI.I`	66.58	2.80	9.57	0.23	0.04	0.88	0.84

Table 4.12. Estimated Variance of LMM `lmm.RI.I`, the estimated adjusted $ICC_{con.}$ and conditional $ICC_{con.}$, and the estimated coefficients $R^2_{con.}$ and $R^2_{mar.}$. $\hat{\sigma}^2_\beta$ and $\hat{\sigma}^2_\alpha$ is the estimated variance for the fixed and random effects (random client intercepts, and $\hat{\sigma}^2_\varepsilon$ is the estimated variance of the model's random error.

Such that here $ICC_{adj.}$ is calculated by,

$$ICC_{adj.} = \frac{\hat{\sigma}^2_\alpha}{\hat{\sigma}^2_\alpha + \hat{\sigma}^2_\varepsilon} = \frac{2.8}{9.57 + 2.8} = 0.23,$$

and,

$$ICC_{con.} = \frac{\hat{\sigma}^2_\alpha}{\hat{\sigma}^2_\beta + \hat{\sigma}^2_\alpha + \hat{\sigma}^2_\varepsilon} = \frac{2.8}{66.58 + 2.8 + 9.57} = 0.04.$$

Whereas the estimated proportion of the total variance is explained by both the fixed and random effects, $R^2_{con.}$ or by only the fixed effects, $R^2_{mar.}$ - by the model `lmm.RI.I` - are calculated respectively

by

$$R^2_{con.} = \frac{\hat{\sigma}^2_\beta + \hat{\sigma}^2_\alpha}{\hat{\sigma}^2_\beta + \sigma^2_\alpha + \hat{\sigma}^2_\varepsilon} = \frac{66.58 + 2.8}{66.58 + 2.8 + 9.57} = 0.88,$$

$$R^2_{mar.} = \frac{\hat{\sigma}^2_\beta}{\hat{\sigma}^2_\beta + \hat{\sigma}^2_\alpha + \hat{\sigma}^2_\varepsilon} = \frac{66.58}{66.58 + 2.8 + 9.57} = 0.84.$$

Based on this and on the 95% confident interval of both ICC estimates, since the values are less than 0.5 - this is still an indication of poor reliability (based on both the adjusted and conditional ICC). This suggests the random effects structure in this model reflects very little of the remaining variance in the aggregated data. Motivating us to explore other suitable random effects. While based on the estimated proportion of variance explained by both the fixed and mixed effects ($R^2_{con.} = 88\%$) indicates an overall adequate fit.

Additionally, the summary of the model output suggests that at the 5% significance level all interaction effects included in the model are significant ($p < 0.05$): `years:oci`, `years:cpi`, `t.type:oci` and `oci:cpi`. We can confirm these findings by calculating the Type III ANOVA table using Wald χ^2 tests (available in the package `car`) for LMMs, obtained by the following `R` command

```
Anova(lmm.RI.I, type = 3)
```

```
Analysis of Deviance Table (Type III Wald chisquare tests)

Response: ln.lr
            Chisq Df Pr(>Chisq)
(Intercept)   2.4  1       0.12
years        31.4 18       0.03 *
t.type       21.5  3      8e-05 ***
oci           3.4  1       0.06 .
rate         13.7  1      2e-04 ***
cpi           2.2  1       0.14
years:oci   150.7 15     <2e-16 ***
years:cpi    29.1 15       0.02 *
t.type:oci   20.6  3      1e-04 ***
oci:cpi      17.8  1      2e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

The results show confirms our findings, and we conclude all included interaction effects remain significant and cannot be further excluded from the model. While their corresponding main effects (with significant marginal effects, at p-value <0.05) are also necessary to be included in the model, and thus, we conclude that our model cannot be further reduced.

Model Comparison: Main Fixed Effects Model `lmm.RI.M` and Interaction Fixed Effects Model `lmm.RI.I`

Firstly, it is evident from the goodness of fit measurements in Table 4.13, there is an increase in all types of R^2 and $ICC_{adj.}$ (with a very slight decrease in $ICC_{adj.}$) - when accounting for interaction effects (in comparison to both the full and reduced fixed main effects models, `lmm.RI.M` and `lmm.RI.M.r`). This indicates that among the clients k there is more reliability when accounting for fixed interaction effects in the model (though there is a very slight increase in the estimated conditional ICC, i.e. the fixed effects variances). Clearly, the model also outperforms both fixed main effects only models in terms of the performance measures: AIC, and AICc, with higher predictive power based on the RMSE.

Model	p	logLik	AIC	AICc	BIC	$R^2_{con.}$	$R^2_{mar.}$	$ICC_{adj.}$	$ICC_{con.}$	RMSE	RSE
`lmm.RI.M`	29	-841.40	1740.80	1747.49	1847.22	0.77	0.72	0.18	0.05	4.04	4.18
`lmm.RI.M.r`	27	-841.65	1737.31	1743.08	1836.40	0.77	0.73	0.17	0.05	4.06	4.19
`lmm.RI.I`	61	-757.34	1636.68	1669.86	1860.55	0.88	0.84	0.23	0.04	2.98	3.09

Table 4.13. Comparison of performance of full fixed main effects LMM `lmm.RI.M`, the reduced LMM `lmm.RI.M.r`, and the interaction fixed effects LMM `lmm.RI.I` - with random client intercept effects. Performance measures include the log-likelihood, AIC, AICc, BIC, the goodness of fit statistics (all types of R^2 and ICC) with estimated the residual standard errors (RMSE and RSE).

However, we note an increase in BIC compared to the main effects models (as expected since with the inclusion of interaction effects - the complexity of the model increases). This implies that the high number of parameters included in this model - significant on the effect on the estimated log loss ratio - may lead to over-fitting. For this reason, later use the *"step-down strategy"* simplification of the fixed-effects structure. However, since the interaction effects model generally outperforms the fixed effects only models (when considering all performance measurements) - the interaction model is our preferred model.

This model was formulated as:

$$
\begin{aligned}
\textbf{Model } \texttt{lm.RI.I:} \quad \texttt{ln.lr}_{kijt} = {} & \beta_0 + \beta_1 I_k^{\texttt{g.treaty}_i} + \beta_2 I_{ki}^{\texttt{g.country}_j} + \sum_{t=2}^{18} \beta_3^t \texttt{years}_{kij}^t + \sum_{q=2}^{4} \beta_4^q \texttt{t.type}_{ki}^q \\
& + \beta_5 \texttt{oci}_{kijt} + \beta_6 \texttt{rate}_{kijt}^d + \beta_7 \texttt{cpi}_{kijt} + \sum_{t=2}^{8} \beta_8^t \texttt{years}_{kij}^t \times \texttt{oci}_{kijt} \\
& + \sum_{t=2}^{18} \beta_9^t \texttt{years}_{kij}^t \times \texttt{cpi}_{kijt} + \sum_{q=2}^{4} \beta_{10}^q \texttt{t.type}_{ki}^q \times \texttt{oci}_{kijt} + \beta_{11} \texttt{oci}_{kijt} \times \texttt{cpi}_{kijt} \\
& + \alpha_k + \varepsilon_{kijt}, \\
\text{with} \quad & \\
\varepsilon_{kijt} \quad & \sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad i.i.d., \quad \forall\, k, i, j, t, \\
\alpha_k \quad & \sim \mathcal{N}(0, \sigma_\alpha^2) \quad i.i.d., \quad \forall\, k = 1, \ldots, 35,
\end{aligned}
\tag{4.25}
$$

We also note that given the low within-cluster variance ($ICC_{adj.} = 0.23$) suggests that additional grouping structures (per (k, i, j, t)) may be required. To essentially capture a higher portion of variability explained by the random effects structure. Later, after further exploring other potential and more complex structures for random effects and the appropriate covariance structures of our model - we will investigate if the number of fixed effects can be reduced (using statistical hypothesis tests, i.e. LRT, to prevent over-fitting).

Models for Random Intercepts and Slopes, with Fixed Main and Interaction Effects

So far, in the previous sections, we have only analyzed the simplest and most type of LMMs; *"simple random effect per level"*, such that the random effect corresponded to only one specific grouping risk level (i.e. the k-th client). Due to this study's design of risk class groups and structure of longitudinal data, we may have to account for the effects at different hierarchical levels and, thus, allow for nested,

clustered, or even crossed random effects.

For this reason, in the following sections, we explore multiple and more complex structures in the random effects and then select the best suitable model (using statistical hypothesis tests and with respect to the performance measures).

1. Specifying the Random Effect Structure

Here, we first consider other common types of LMMs, including random intercepts and random slopes that correlate or vary by the risk class groupings (k, i, j, t) Frees et al. (1999). Afterwards, we will investigate more sophisticated residual covariance structures (later discussed in Section 4.7) to capture more variability unexplained by the selected mixed model.

The random-effects structures analyzed in this study, respectively, are described by the following:

Multiple random effects per level: this model includes both ***random intercepts and random slopes*** effects, which vary by a group of risk class levels (or factors). Here, the correlation between the multiple random effects is generally assumed, implying that the covariance matrix, say $\boldsymbol{D}$, is not diagonal and is assumed to be unstructured.

Hence, for example, with random `client` intercepts and random `g.treaty` slopes, this model (also given in Equation (4.17)) may be formulated as

Model `lmm.RS.M1`: *(LMM with multiple random effects and fixed effects)*

$$\begin{aligned}\texttt{ln.lr}_{kijt} = \beta_0 + \beta_1 I_{ki}^{\texttt{g.country}_j} + \sum_{t=2}^{18} \beta_2^t \texttt{years}_{kij}^t + \sum_{q=2}^{4} \beta_3^q \texttt{t.type}_{ki}^q + \beta_4 \texttt{oci}_{kijt} \\ + \beta_5 \texttt{rate}_{kijt}^d + \beta_6 \texttt{cpi}_{kijt} + \sum_{t=2}^{8} \beta_7^t \texttt{years}_{kij}^t \times \texttt{oci}_{kijt} + \sum_{t=2}^{18} \beta_8^t \texttt{years}_{kij}^t \times \texttt{cpi}_{kijt} \\ + \sum_{q=2}^{4} \beta_9^q \texttt{t.type}_{ki}^q \times \texttt{oci}_{kijt} + \beta_{10} \texttt{oci}_{kijt} \times \texttt{cpi}_{kijt} \\ + \alpha_{0k} + \alpha_{1k} I_k^{\texttt{g.treaty}_i} + \varepsilon_{kijt},\end{aligned} \tag{4.26}$$

whereas, for the random components, we have

$$\begin{aligned}\varepsilon_{kijt} &\sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad i.i.d., \quad \forall\, k, i, j, t, \\ \begin{bmatrix}\alpha_{0k} \\ \alpha_{1k}\end{bmatrix} &\sim \mathcal{N}_2\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}\sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2\end{bmatrix}\right) \quad i.i.d., \quad \forall\, k = 1, \ldots, 35,\end{aligned} \tag{4.27}$$

which implies that the random effects are independent for each k-th client, and i-th treaty - across all the risk classes (k, i, j, t), and independent of the error terms ε_{kijt}. Recall, the random effects is now a bivariate vector with residual covariance matrix (as a diagonal) with homogeneous variance σ_ε^2.

Nested random effects: in this classification, the levels of one risk class factor can occur only within certain levels of the first factor (hierarchical or multilevel). This type of random effects structure

allows us to form a ***hierarchical structure with multiple grouping factors***. The nested structure of random effects accounts for model variation in the random intercepts, where intercepts vary among two groups within a (nested) group. See Bates et al. (2014) for more examples.

For instance, consider the observations grouped by the i-th risk class level (level 2), `g.treaty`. Since each i-th treaty group is a subset for $k = 1, \ldots, 35$ (i.e. "nested"), we can then include random intercepts varying among each client k and i-th treaty group - within each k-th client risk class group - through this mixed model:

Model `lmm.RS.N1`: ***(LMM with nested random effects and fixed effects)***

$$\begin{aligned}\texttt{ln.lr}_{kijt} = \beta_0 + \beta_1 I_{ki}^{\texttt{g.country}_j} + \sum_{t=2}^{18} \beta_2^t \texttt{years}_{kij}^t + \sum_{q=2}^{4} \beta_3^q \texttt{t.type}_{ki}^q + \beta_4 \texttt{oci}_{kijt} + \beta_5 \texttt{rate}_{kijt}^d \quad (4.28)\\ + \beta_6 \texttt{cpi}_{kijt} + \sum_{t=2}^{8} \beta_7^t \texttt{years}_{kij}^t \times \texttt{oci}_{kijt} + + \sum_{t=2}^{18} \beta_8^t \texttt{years}_{kij}^t \times \texttt{cpi}_{kijt}\\ + \sum_{q=2}^{4} \beta_9^q \texttt{t.type}_{ki}^q \times \texttt{oci}_{kijt} + \beta_{10} \texttt{oci}_{kijt} \times \texttt{cpi}_{kijt} + \alpha_k^{(1)} + \alpha_{ki}^{(11)} + \varepsilon_{kijt},\end{aligned}$$

such that,

$$\begin{aligned}\varepsilon_{kijt} &\sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad i.i.d., \quad \forall\, k, i, j, t,\\ \alpha_k^{(1)} &\sim \mathcal{N}(0, \sigma_1^2) \quad i.i.d., \quad \forall\, k = 1, \ldots, 35,\\ \alpha_{ki}^{(12)} &\sim \mathcal{N}(0, \sigma_{12}^2) \quad i.i.d., \quad \forall\, k = 1, \ldots, 35,\ i \in \{1^k, 2^k\},\end{aligned} \quad (4.29)$$

where,

- the random errors ε_{kijt}, with variance σ_ε^2, structure the variability within all the risk classes levels or groups (k, i, j, t),
- $\alpha_k^{(1)}$ is the random client intercept for risk class level k, given by the grouping level (1) (i.e. `client`) with zero mean and variance σ_1^2,
- $\alpha_{ki}^{(12)}$ is the random intercept effect varying among client k and treaty i (given at level (2)) - within client k (grouping level (1)), with zero mean and variance σ_{12}^2.

It is important to note that for this specification, we allow for two random intercepts - which vary at different risk class grouping levels.

Crossed random effects: a model specification with within-subject random effects, where ***multiple (non-nested)*** observations of a risk class group are associated with multiple risk class levels (i.e. grouping variables). This means that each group of a risk class can occur at each risk class level (often expressed as a "special kind of interaction" between the groupings of risk class levels).

For example, consider a LMM with crossed random effects between the client k and treaty i. Then, this model can be formulated as:

Model `lmm.RS.C1`**:** ***(LMM with crossed random effects and fixed effects)***

$$\begin{aligned}\texttt{ln.lr}_{kijt} = \beta_0 + \beta_1 I_{ki}^{\texttt{g.country}_j} + \sum_{t=2}^{18} \beta_2^t \texttt{years}_{kij}^t + \sum_{q=2}^{4} \beta_3^q \texttt{t.type}_{ki}^q \quad + \quad \beta_4 \texttt{oci}_{kijt} + \beta_5 \texttt{rate}_{kijt}^d \\ + \beta_6 \texttt{cpi}_{kijt} + \sum_{t=2}^{8} \beta_7^t \texttt{years}_{kij}^t \times \texttt{oci}_{kijt} \sum_{t=2}^{18} \beta_8^t \texttt{years}_{kij}^t \times \texttt{cpi}_{kijt} + \sum_{q=2}^{4} \beta_9^q \texttt{t.type}_{ki}^q \times \texttt{oci}_{kijt} \\ + \beta_{10} \texttt{oci}_{kijt} \times \texttt{cpi}_{kijt} + \alpha_{ki} + \varepsilon_{kijt},\end{aligned} \tag{4.30}$$

where,

$$\begin{aligned}\varepsilon_{kijt} &\sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad i.i.d., \quad \forall\, k, i, j, t, \\ \alpha_{ki} &\sim \mathcal{N}(0, \sigma_\alpha^2) \quad i.i.d., \quad \forall\, k = 1, \ldots, 35, \quad i \in \{1^k, 2^k\}.\end{aligned}$$

Therefore, in comparison to the nested effects random effects, we allow for only one random intercept here - which varies among client k and treaty i (i.e. random effect for the interaction of risk class level k and i) - with variance σ_α^2. The random error terms are based on each observation at risk class level t - corresponding to the combined risks at group levels k, i and j.

The differences between the crossed and nested random effects structures are shown in Equation 4.2. It is important to note that while that the first specification (defined as the "Multiple random effects per level") - allows us to include additional crossed and nested random effects - by specifying in the random client intercept effects). For the remaining two specifications, in this study, we specify the random intercept components either as nested random effects or as crossed random effects (i.e. in this study, we choose between the random effects structures).

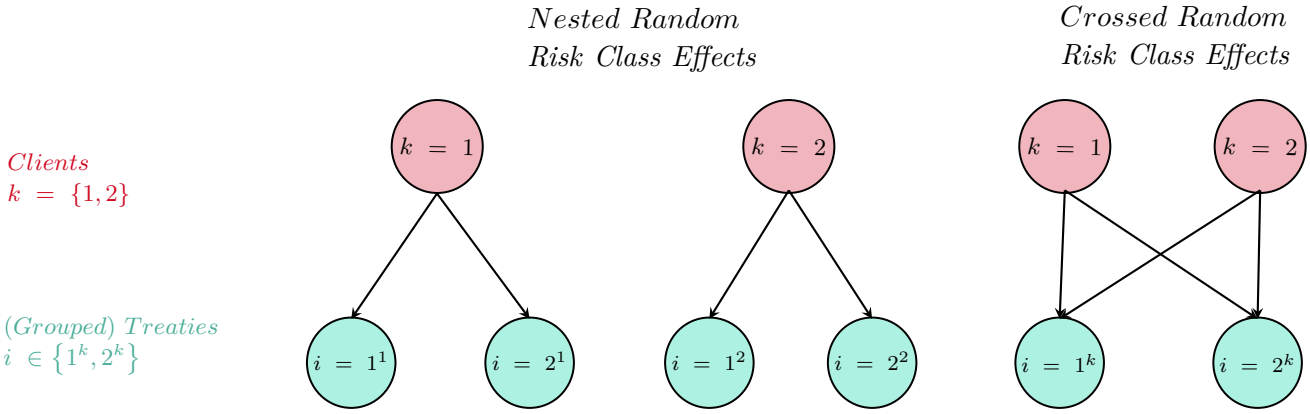


Figure 4.7. Example of random risk class effects structure, based on clients k and treaties i, which may be crossed or nested. **Left Diagram:** Nested random effects of clients $k = 1, 2$ (first level) with (grouped) treaty i effects. This structure introduces nested levels of a random factor - unique to each client k risk class level which are also nested. **Right Diagram:** examines every combination of risk class level groups (k, i) for all $k = 1, 2$ and $i \in \{1^k, 2^k\}$.

In this section, our ***model building strategy*** for LMMs is as follows:

Step 1. Considering the fixed main and interaction effects included in model `lmm.RI`, compare and fit the three structures of random effects: multiple random effects per level, nested and crossed random effects, based on the grouping structure of the risk classes (k, i) only.

Step 2. Using statistical hypothesis tests for random effects to determine the preferred specification of random effects structure. If comparing non-nested models, we will instead compare the models based on the performance measures (previously outlined in Section 3.4).

Step 3. Continue model building (with the selected random effects specification) by allowing for more significant random effects (or reduce if possible).

Step 4. Select and assess the best possible fit based on the model precision and complexity.

Table 4.14 shows the corresponding model formulas in `R` (based on the function `lmer` - available in the `R` package `lme4`) - to model the respective random effects structures (as outlined above): `lmm.RS.M1` (multiple random effects per level), `lmm.RS.N1` (nested random effects), and `lmm.RS.C1` (crossed random effects).

Model Name	**Formula** (as given by the `R` package `lme4`)	**Random Effects Structure and Description**
`lmm.RS.M1`	`ln.lr ~ g.country + years + t.type + oci + rate + cpi + years:oci + years:cpi + t.type:oci + oci:cpi + (1 + g.treaty|client)`	**Multiple random effects per risk class level:** random `client` intercepts for $k = 1, \ldots, 35$ and random `g.treaty` slopes for $i \in \{1^k, 2^k\}$ that vary over all risk classes (k, i, j, t).
`lmm.RS.N1`	`ln.lr ~ g.country + years + t.type + oci + rate + cpi + years:oci + years:cpi + t.type:oci + oci:cpi + (1|client/g.treaty)`	**Nested random effects:** random intercepts varying among `client` k and `g.treaty` i groups - within the risk class level of `client` $k = 1, \ldots, 35$.
`lmm.RS.C1`	`ln.lr ~ g.country + years + t.type + oci + rate + cpi +years:oci + years:cpi + t.type:oci + oci:cpi + (1|client:g.treaty)`	**Crossed random effects:** random intercept effects for the "interaction" (or groupings) of the risk class levels `client` k and `g.treaty` i.

Table 4.14. Model specifications with different structures of the random effects, based on the risk class levels `client`, $k = 1, \ldots, 35$ and `g.treaty`, $i \in \{1^k, 2^k\}$. Corresponding model formulas in `R` are given (using the LMM syntax for `R` function `lmer` - provided by the package `lme4`), where the model `lmm.RS.M1` allows for multiple random effects per level (`"M"`), model `lmm.RS.N1` allows for nested random effects (`"N"`), and `lmm.RS.N1` allows for crossed random effects (`"C"`). Note, ***fixed effects*** are given in red, and ***random effects*** are given in blue.

Step 1: Results of Random Effects Structures

Recall that our goal here is select the best performing random effects structure that best represents the underlying data. Once again, here, we consider both interaction and main fixed effects and, then use the R function `lmer` to fit all three models (setting `REML = FALSE`, package `lme4`).

Model Results: `lmm.RS.M1`, `lmm.RS.N1`, and `lmm.RS.C1`

The summary of the model output of the three models (previously defined in Table 4.14) with respect to the fixed effects and estimated variances of random components are shown in Table 4.15. Specifically, the table first shows the estimates of the fixed regression parameters - with their corresponding standard error, the 95% confidence intervals (*"C.I"*) and p-values (found under "*"fixed effects"*). Whereas, for the models' random components, the estimated covariance ("*Cov.*") and variances (*"Var."*) are shown under the corresponding *"random Effects"*. Note, the estimated variances for the random components can also obtained using the R function `VarCorr` for all models (also provided by R package `lme4`).

From this output, we observe the following:

- At a 5% significance level, all three models indicate that the fixed (main and interaction) effects of the following covariates are significant (p-value <0.05): `years`, `t.type`, `rate`, `cpi`, `years:oci`, `years:cpi`, `t.type:oci`, `oci:cpi`. While `g.country` is not significant (across all three models).
- While `oci` is no longer significance when allowing for nested random effects(`lmm.RS.N1` in the model (with p-value = 0.09) or when allowing for crossed random `client:g.treaty` intercepts (with p-value = 0.09).
- The models with only random intercepts (`lmm.RS.N1` and `lmm.RS.C1`) estimate the regression parameters (for fixed effects) and random error are equivalent (with $\hat{\beta}_0 = 24.67$ and $\hat{\sigma}^2_\varepsilon = 9.53$).
- However, for model `lmm.RS.N1`, the estimated variances of the nested intercept of the `client` and `g.treaty` is $\hat{\sigma}^2_{12} = 1.51$, and the estimated variance client intercept only is $\hat{\sigma}^2_1 = 1.37$. Meanwhile, the estimated variance of the crossed random `client` and `g.treaty` intercept effects (model `lmm.RS.C1`) is $\sigma^2_\alpha = 2.88$.
- In comparison, the model `lmm.RS.M1` estimates the fixed intercept effects as $\hat{\beta}_0 = 23.54$ (and $\hat{\sigma}^2_\varepsilon = 9.59$). Recall, for this model, the random effects structure is now bivariate.
- Hence, the output also shows the estimated variances for the following random components: the random `client` intercept effects $\text{Var}(\hat{\alpha}_{0k}) = \hat{\sigma}^2_0 = 0.31$, the random `g.treaty` slope effects $\text{Var}(\hat{\alpha}_{1k}) = \hat{\sigma}^2_1 = 6.94$, and the estimated covariance of the random `client` intercept, α_{0k}, and random `g.treaty` slope, α_{1k}, is given by $\text{Cov}(\hat{\alpha}_{0k}, \hat{\alpha}_{1k}) = \hat{\sigma}_{01} = -1.13$.

Based on these findings and the results of the fitted models, we now compare their performances to find the most suitable structure for the random components.

Model `lmm.RS.M1` *(Multiple)*					Model `lmm.RS.N1` *(Nested)*				Model `lmm.RS.C1` *(Crossed)*			
Fixed Effects					***Fixed Effects***				***Fixed Effects***			
Predictors	***Estimates***	***std. Error***	***95% C.I.***	*p-value*	***Estimates***	***std. Error***	***95% C.I.***	*p-value*	***Estimates***	***std. Error***	***95% C.I.***	*p-value*
(Intercept)	23.54	16.07	-8.12 – 55.20	0.14	24.67	15.95	-6.76 – 56.10	0.12	24.67	15.95	-6.75 – 56.10	0.12
`g country [L]`	0.46	0.72	-0.96 – 1.88	0.53	0.57	0.73	-0.86 – 2.00	0.43	0.57	0.73	-0.86 – 2.00	0.43
`years [2003]`	65.89	76.15	-84.17 – 215.95	0.39	66.30	76.24	-83.92 – 216.52	0.39	66.30	76.24	-83.92 – 216.52	0.39
`years [2004]`	-32.90	16.52	-65.44 – -0.35	**0.05**	-32.49	16.42	-64.84 – -0.14	**0.05**	-32.49	16.42	-64.84 – -0.14	**0.05**
`years [2005]`	-49.52	17.78	-84.55 – -14.48	**0.01**	-51.26	17.65	-86.03 – -16.48	**4e-03**	-51.26	17.65	-86.03 – -16.48	**2e-03**
.....		...			.....		...		.....		...	
`t type [CXL]`	5.62	2.01	1.65 – 9.58	**0.01**	5.73	1.97	1.85 – 9.61	**4e-03**	5.73	1.97	1.85 – 9.61	**4e-03**
`t type [QS]`	-1.66	1.34	-4.30 – 0.98	0.22	-1.25	1.28	-3.78 – 1.28	0.33	-1.25	1.28	-3.78 – 1.28	0.33
`t type [SP]`	2.49	2.08	-1.61 – 6.58	0.23	3.25	2.05	-0.80 – 7.30	0.12	3.25	2.05	-0.80 – 7.30	0.12
`oci`	8.61	4.02	0.68 – 16.53	**0.03**	7.06	4.10	-1.02 – 15.13	0.09	7.06	4.10	-1.02 – 15.13	0.09
`rate`	-0.47	0.15	-0.76 – -0.18	**2e-03**	-0.52	0.15	-0.81 – -0.22	**2e-03**	-0.52	0.15	-0.81 – -0.22	**2e-03**
`cpi`	-0.29	0.21	-0.70 – 0.11	0.15	-0.30	0.20	-0.70 – 0.10	0.14	-0.30	0.20	-0.70 – 0.10	0.14
`years [2003] :oci`	-1.23	5.75	-12.57 – 10.11	0.83	-1.26	5.64	-12.37 – 9.85	0.82	-1.26	5.64	-12.37 – 9.85	0.82
`years [2004] :oci`	-4.58	2.71	-9.92 – 0.77	0.09	-3.75	2.74	-9.15 – 1.64	0.17	-3.75	2.74	-9.15 – 1.64	0.17
`years [2005] :oci`	-5.53	2.46	-10.37 – -0.69	**0.03**	-5.08	2.46	-9.94 – -0.23	**0.04**	-5.08	2.46	-9.94 – -0.23	**0.04**
.....		...			.....		...		.....		...	
`years [2003]:cpi`	-0.78	0.98	-2.72 – 1.16	0.43	-0.78	0.98	-2.72 – 1.15	0.43	-0.78	0.98	-2.72 – 1.15	0.43
`years [2004]:cpi`	0.47	0.22	0.05 – 0.90	**0.03**	0.46	0.21	0.04 – 0.88	**0.03**	0.46	0.21	0.04 – 0.88	**0.03**
`years [2005]:cpi`	0.66	0.22	0.22 – 1.10	**3e-04**	0.66	0.22	0.23 – 1.10	**3e-04**	0.66	0.22	0.23 – 1.10	**3e-04**
.....		...			.....		...		.....		...	
`t type [CXL] :oci`	-4.21	1.24	-6.65 – -1.77	**8e-04**	-4.45	1.23	-6.87 – -2.03	**<0.001**	-4.45	1.23	-6.87 – -2.03	**<0.001**
`t type [QS] :oci`	-0.41	0.57	-1.53 – 0.70	0.47	-0.48	0.56	-1.58 – 0.62	0.39	-0.48	0.56	-1.58 – 0.62	0.39
`t type [SP] :oci`	-3.88	1.38	-6.60 – -1.15	**0.01**	-4.18	1.39	-6.92 – -1.44	**3e-04**	-4.18	1.39	-6.92 – -1.44	**3e-04**
`oci:cpi`	-0.12	0.03	-0.17 – -0.07	**<0.001**	-0.11	0.03	-0.16 – -0.06	**<0.001**	-0.11	0.03	-0.16 – -0.06	**<0.001**
Random Effects					***Random Effects***				***Random Effects***			
Variance Components					***Variance Components***				***Variance Components***			
$\hat{\sigma}_0^2$ (Var: `client`, Intercept)				0.31	$\hat{\sigma}_{12}^2$ (Var:`client/g.treaty`, Intercept)			1.51	$\hat{\sigma}_{\alpha}^2$ (Var:`client:g.treaty`, Intercept)			2.88
$\hat{\sigma}_1^2$ (Var.: `client, g.treaty2`)				6.94	$\hat{\sigma}_1^2$ (Var.:`client`, Intercept)			1.37	$\hat{\sigma}_{\varepsilon}^2$ (Var.: Residual)			9.53
$\hat{\sigma}_{\varepsilon}^2$ (Var: Residual)				9.59	$\hat{\sigma}_{\varepsilon}^2$ (Var: Residual)			9.53				
$\hat{\sigma}_{01}$ (Cov.: `client`, Intercept, `g.treaty2`, Slope)				-1.13								

Table 4.15. Summary of Model Results, with fixed and random effects (estimated variances). Based on the following models: `lmm.RS.M1` (Multiple random effects per level), `lmm.RS.N1` (Nested random effects), `lmm.RS.C1` (Crossed random effects). Summary output shows: the parameter estimates (*"Estimates"*), the standard error ("*std. Error*"), the 95% confidence interval ("*C.I.*"), and the corresponding *p-value*. p-values < 0.05 are given in bold.

Step 2: Comparison of Random Effects Structures

Since we are comparing models with non-nested random effects; we evaluate the models in terms of complexity, the goodness of fit, and predictive performances, given in Table 4.16 (based on the performance measures discussed in Section 3.4).

Model	p	logLik	AIC	AICc	BIC	$R^2_{con.}$	$R^2_{mar.}$	$ICC_{adj.}$	$ICC_{con.}$	RMSE	RSE
`lmm.RS.M1`	64	-754.90	1637.79	1674.77	1872.66	0.88	0.88	0.03	0.00	3.01	3.10
`lmm.RS.N1`	63	-757.03	1640.07	1675.75	1871.27	0.88	0.84	0.23	0.04	2.97	3.09
`lmm.RS.C1`	62	-757.03	1638.07	1672.48	1865.60	0.88	0.84	0.23	0.04	2.97	3.09

Table 4.16. Comparison of random effects models, with different random effects structures based on the risk class levels `client`, $k = 1, \ldots, K$, and `g.treaty`, $i = 1, \ldots, P^k$.

Once again, here we compare the information criteria for all models, AIC, BIC and AICc (applicable due to the sample size and the large number of parameters included in the models). In addition, to assess the goodness-of-fit of the models - we look at the log-likelihood estimates, ICC, and LMM coefficient of determination $R^2_{con.}$ (or marginal $R^2_{mar.}$). While we evaluate the models' predictive performances based on the rooted mean squared error (RMSE) or the residual error standard deviation (referred to as "RSE"). Additionally, we can also compare the corresponding prediction 95% intervals for each model (not included in this paper). From these results, we note that the selected best-performing model (in terms of the predictive power) with high accuracy - differs across the measures. For instance, based on the three information criteria considered, `lmm.RS.M1` attains the lowest AIC - but in terms of the BIC and AICc values `lmm.RS.C1` is the best fit. Meanwhile, `lmm.RS.M1` accounts for highest proportion of variability in the estimated log loss ratios - explained by the fixed effects ($R^2_{mar.} = 0.88$). However, here we are more interested in the variability also captured by the random effects structure. Hence we are more motivated to consider the models that attain higher values which account for both the fixed and random effects (for instance, such as the $R^2_{con.}$ and $ICC_{adj.}$). From our output, we can also see that model `lmm.RS.M1` results in the widest prediction intervals (indicating the most estimated instability in comparison to the other two models). For this reason, the crossed effects model `lmm.RS.C1` attains higher AICc and BIC values, we statistical test (using LRT) if the model `lmm.RS.M1` is preferred or improves our fit. Likewise, we can also test if the model with random `g.treaty` slope, `lmm.RS.M1`, improves the random `client` intercept model too (given by `lmm.RI.I`). Both results are obtained by the ANOVA table given in Table 4.17.

Model	npar	AIC	BIC	logLik	deviance	Chisq	*df*	Pr(>Chisq)
`lmm.RS.C1`	62.00	1638.07	1865.60	-757.03	1514.07			
`lmm.RS.M1`	64.00	1637.79	1872.66	-754.90	1509.79	4.28	2	0.1179
`lmm.RI.I`	63.00	1640.00	1871.20	-757.00	1514.00			
`lmm.RS.M1`	64.00	1637.79	1872.66	-754.90	1509.79	4.21	1	0.0403

Table 4.17. ANOVA tables, for LRTs (Likelihood ratio tests), comparing models `lmm.RS.C1` and `lmm.RS.M1` (top rows) - test to check if `g.treaty` should be random intercept or random slope.

Clearly based on these results, we conclude that while considering `g.treaty` as a random effect improves the model fit of model `lmm.RI.I` (with client random intercept effects only, p-value = 0.04) - it does not improve the overall fit of the crossed random effects model (p-value = 0.11). For this reason, we

conclude that the crossed random effects structure is our preferred model specification (and hence proceed further with these types of models), given by

Model `lm.RS.C5`: *(crossed random effects model with all risk class levels)*

$$\begin{aligned}\texttt{ln.lr}_{kijt} &= \beta_0 + \beta_1 I_{ki}^{\texttt{g.country}_j} + \sum_{t=2}^{18} \beta_2^t \texttt{years}_{kij}^t + \sum_{q=2}^{4} \beta_3^q \texttt{t.type}_{ki}^q + \beta_4 \texttt{oci}_{kijt} + \beta_5 \texttt{rate}_{kijt}^d + \beta_6 \texttt{cpi}_{kijt} \\ &\quad + \sum_{t=2}^{8} \beta_7^t \texttt{years}_{kij}^t \times \texttt{oci}_{kijt} + \sum_{t=2}^{18} \beta_8^t \texttt{years}_{kij}^t \times \texttt{cpi}_{kijt} + \sum_{q=2}^{4} \beta_9^q \texttt{t.type}_{ki}^q \times \texttt{oci}_{kijt} \\ &= \beta_{10} \texttt{oci}_{kijt} \times \texttt{cpi}_{kijt} + \alpha_{ki} + \varepsilon_{kijt},\end{aligned} \tag{4.31}$$

where for the random components, we have

$$\begin{aligned}\varepsilon_{kijt} &\sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad i.i.d., \quad \forall\, k, i, j, t, \\ \alpha_{ki} &\sim \mathcal{N}(0, \sigma_\alpha^2) \quad i.i.d., \quad \forall\, k = 1, \ldots, 35, \quad i \in \{1^k, 2^k\}.\end{aligned} \tag{4.32}$$

Step 3: Investing More Complex Crossed Random Effects Structures

In this step, we allow for more complex specifications with respect to the crossed random effects structures in our model. Specifically, we start by including the risk classes (k, i, j, t) respectively, based on the hierarchical structure of data. The models investigated in this section are outlined in Table 4.18. Here we denote the random effects in red and fixed effects in blue (for the model formulas).

Model Name	Formula	Random Effects Structure and Description
`lmm.RS.C2`	`ln.lr ~ years + t.type + oci + rate + cpi + years:oci + years:cpi + t.type:oci + oci:cpi + (1|client:g.treaty:g.country)`	**Crossed random effects:** random intercept effects for the "interaction" (or groupings) of the risk class levels `client` k `g.treaty`, i and `g.country` j .
`lmm.RS.C3`	`ln.lr ~ years + t.type + oci + rate + cpi + years:oci + years:cpi + t.type:oci + oci:cpi + (1|client:g.treaty:g.country:years)`	**Crossed random effects:** random intercept effects for the "interaction" (or groupings) of the risk class levels `client` k `g.treaty`, i `g.country` j and `years`, t.
`lmm.RS.C4`	`ln.lr ~ years + t.type + oci + rate + cpi + years:oci + years:cpi + t.type:oci + oci:cpi + (1 + oci|client:g.treaty:g.country)`	**Crossed and Multiple random effects:** random intercept effects for the "interaction" (or groupings) of the risk class levels `client` k and `g.treaty` i. Also includes (multiple) random `oci` slope effects varying for risk class levels (k, i, j, t).
`lmm.RS.C5`	`ln.lr ~ years + t.type + oci + rate + cpi + years:oci + years:cpi + t.type:oci + oci:cpi + (1 + oci|client:g.treaty:g.country:years)`	**Crossed and Multiple random effects:** random intercept effects for the "interaction" (or groupings) of the risk class levels `client` k `g.treaty`, i `g.country` j and `years`, t. Including (multiple) random `oci` slope effects varying for risk class levels (k, i, j, t).

Table 4.18. Model formulas of the different specifications of the random effects, based on the crossed random effect risk class levels `client`, $k = 1, \ldots, 35$, `g.treaty`, $i \in \{1^k, 2^k\}$, `g.country`, $j \in \{H^{ki}, L^{ki}\}$ and `years`, $t \in \{2001^{kij}, \ldots, 2021^{kij}\}$. Corresponding model formulas in `R` are given (using the LMM syntax in package `lme4`), extending the structure of selected model `lmm.RS.C1`. Recall, ***fixed effects*** are given in red, and ***random effects*** are given in blue.

Model Results: `lmm.RS.C2`, `lmm.RS.C3`, `lmm.RS.C4`, and `lmm.RS.C5`.

Here we focus on the estimate variance components of the random effects. The estimated variance components of each model can be obtained (from the model results) using the `R` command `VarCorr` (given in a list here, for all models). Note each model here is fitted using the ML method.

```
list(VarCorr(lmm.RS.C2),VarCorr(lmm.RS.C3),VarCorr(lmm.RS.C4),VarCorr(lmm.RS.C5))
```

```
[[1]] # VarCorr(lmm.RS.C2)
 Groups                    Name        Std.Dev.
 client:g.treaty:g.country (Intercept) 1.76
 Residual                              3.49

[[2]] # VarCorr(lmm.RS.C3)
 Groups                           Name        Std.Dev.
 client:g.treaty:g.country:years (Intercept) 2.83
 Residual                                    2.86

[[3]] # VarCorr(lmm.RS.C4)
 Groups                    Name        Std.Dev. Corr
 client:g.treaty:g.country (Intercept) 2.99
                           oci         3.12     -0.96
 Residual                              2.88

[[4]] # VarCorr(lmm.RS.C5)
 Groups                           Name        Std.Dev. Corr
 client:g.treaty:g.country:years (Intercept) 4.78
                                 oci         3.95     -0.96
 Residual                                    2.22
```

From here we observe that model `lmm.RS.C5` yields the smallest estimated residual variance, $\hat{\sigma}^2_{\varepsilon} = 2.22$, for the random terms ε_{kijt} given the risk class levels (k, i, j, t) (as crossed random intercept effects), in comparison to all other models. This model also includes the `oci` as a random slope effect, varying for all risk classes.

Step 4: Comparison of Crossed Random Effects and Fixed Effects Models

We now compare the results of fitted models, then further analyze the results of the selected best model fit. Similar to previous sections, we once again assess based on the performance measures given in Table 4.19. It is evident from the results that model `lmm.RS.C5` is the best fit. We conclude this

Model	p	logLik	AIC	AICc	BIC	$R^2_{con.}$	$R^2_{mar.}$	$ICC_{adj.}$	$ICC_{con.}$	RMSE	RSE
`lmm.RS.C1`	62	-757.03	1638.07	1672.48	1865.60	0.88	0.84	0.23	0.04	2.97	3.09
`lmm.RS.C2`	61	-737.16	1596.33	1629.51	1820.19	0.85	0.81	0.20	0.04	2.99	3.49
`lmm.RS.C3`	62	-737.11	1598.22	1632.64	1825.75	0.90	0.80	0.49	0.10	2.05	2.86
`lmm.RS.C4`	49	-767.25	1632.50	1652.91	1812.32	0.90	0.68	0.70	0.22	2.50	2.88
`lmm.RS.C5`	63	-696.81	1519.62	1555.30	1750.82	0.95	0.66	0.85	0.29	1.61	2.22

Table 4.19. Model performance summary, for fitted models with crossed random effects structures: (1) `lmm.RS.C1` (random intercept effect for the interaction of level k, i.e. `clients`, and i, i.e.`g.treaty`), (2) `lmm.RS.C2` (random intercept effect for the interaction of levels: k, i and j, i.e.`g.country`), (3) `lmm.RS.C3` (random intercept effect for the interaction of levels: k, i, j and t, i.e. `years`), (4) `lmm.RS.C4` (random intercept effect for the interaction of levels k, i, j, with random `oci` slope effects), (5) `lmm.RS.C5` (random intercept effect for the interaction of levels k, i, j, t, with random `oci` slope effects).

because it outperforms all other models based on all performance measures - while accounting for both the random and fixed effects. This means that, for example, even though `lmm.RS.C1` attains the highest estimated coefficient of $R^2_{mar.}$ - since it only accounts for the fixed effects - we state `lmm.RS.C5` is our preferred model.

Especially since, compared to all other considered models, it also captures the highest portion of variability explained by the whole model, i.e. accounted for by both the fixed and random effects (at conditional $R^2_{con.} = 0.95$). While the model's random effects structure also explains the highest portion of the variation ($ICC_{adj.} = 0.85$) in the grouping of the risk class structure (k, i, j, t) - with crossed levels in the random intercept (for all risk class levels) and random `oci` slope effects. The models' results also suggest `lmm.RS.C5` is the most suitable fit and preferable given the data, based on the lowest RMSE (estimated as 1.81).

This model is given by:

$$\begin{aligned}\textbf{Model } \texttt{lm.RS.C5}\text{:}\quad \texttt{ln.lr}_{kijt} = \beta_0 &+ \sum_{t=2}^{18} \beta_1^t \texttt{years}_{kij}^t + \sum_{q=2}^{4} \beta_2^q \texttt{t.type}_{ki}^q \quad + \beta_3 \texttt{oci}_{kijt} + \beta_4 \texttt{rate}_{kijt}^d \\ &+ \beta_5 \texttt{cpi}_{kijt} + \sum_{t=2}^{8} \beta_6^t \texttt{years}_{kij}^t \times \texttt{oci}_{kijt} + \sum_{t=2}^{18} \beta_7^t \texttt{years}_{kij}^t \times \texttt{cpi}_{kijt} \\ &+ \sum_{q=2}^{4} \beta_8^q \texttt{t.type}_{ki}^q \times \texttt{oci}_{kijt} + \beta_9 \texttt{oci}_{kijt} \times \texttt{cpi}_{kijt} \\ &+ \alpha_{0kijt} + \alpha_{1kijt} oci_{kijt} + \varepsilon_{kijt},\end{aligned} \tag{4.33}$$

with random components,

$$\begin{aligned}\varepsilon_{kijt} &\sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad i.i.d., \quad \forall\, k, i, j, t, \\ \begin{bmatrix}\alpha_{0kijt} \\ \alpha_{1kijt}\end{bmatrix} &\sim \mathcal{N}_2\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}\sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2\end{bmatrix}\right) \quad i.i.d., \quad \forall\, k = 1, \ldots, 35.\end{aligned}$$

We fit this model using the following `R` command (using ML method, via the function `lmer`) to obtain the following model summary (using the ML estimation).

```
lmm.RS.C4 <- lmer(ln.lr ~   years + t.type + oci + rate  + cpi + years:oci + years:cpi + t.type:oci + oci:cpi  +
    (1+oci|client:g.treaty:g.country), na.action = na.omit, data = cat_data, REML = F)
```

```
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: ln.lr ~ years + t.type + oci + rate + cpi + years:oci + years:cpi +
    t.type:oci + oci:cpi + (1 + oci | client:g.treaty:g.country)
   Data: cat_data
Control: lmerControl(check.nobs.vs.nRE = "ignore", check.nobs.vs.nlev = "ignore")

     AIC      BIC   logLik deviance df.resid
    1553     1784     -713     1427      227

Scaled residuals:
   Min     1Q Median     3Q    Max
 -2.93  -0.52   0.06   0.61   2.61

@Random effects:@
 Groups                   Name        Variance Std.Dev. Corr
 client:g.treaty:g.country (Intercept) 7        3
                          oci         8        3        -0.97
 Residual                             6        3
Number of obs: 290, groups:  client:g.treaty:g.country, 43

@@Fixed effects:@@
             Estimate Std. Error     df t value Pr(>|t|)
```

```
(Intercept)      38.33      13.14 243.80     2.9    0.004 **
years2003        20.74      58.47 169.12     0.4    0.723
years2004       -32.85      13.16 265.91    -2.5    0.013 *
years2005       -43.36      14.20 255.37    -3.1    0.002 **
......
t.typeCXL         3.20       1.67 189.46     1.9    0.057 .
t.typeQS         -1.21       0.92  45.29    -1.3    0.194
t.typeSP          1.48       1.70 174.02     0.9    0.386
oci             -18.39       4.86 226.57    -3.8    2e-04 ***
rate             -0.38       0.12  83.47    -3.2    0.002 **
cpi              -0.33       0.16 247.55    -2.0    0.047 *
years2003:oci    11.76       4.70 232.89     2.5    0.013 *
years2004:oci     8.56       2.79 258.73     3.1    0.002 **
years2005:oci     9.29       2.95 268.43     3.2    0.002 **
......
years2003:cpi    -0.38       0.75 160.08    -0.5    0.613
years2004:cpi     0.32       0.17 266.24     1.9    0.062 .
years2005:cpi     0.42       0.18 262.13     2.3    0.020 *
......
t.typeCXL:oci    -2.18       1.16 237.50    -1.9    0.061 .
t.typeQS:oci      0.23       0.47 270.18     0.5    0.630
t.typeSP:oci     -2.03       1.30 237.74    -1.6    0.121
oci:cpi           0.05       0.03 232.90     1.5    0.127
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

For the fixed effects, the output states that the main and pairwise interaction effects of `t.type` are no longer significant (at a 5% level). This means the interaction effects `t.type:oci` are no longer significant, in addition to the pairwise effects of the term `oci:cpi` (at p-value $= 0.13$). However, we will test if we can exclude any fixed effects from our model - after we analyze different structures of the residence covariance structure (in the next section).

While for the random effects, we can utilize the function `ranova` once again to test if the random slope effects of `oci` (i.e. the number of historic of natural catastrophic unique events) - for all risk class levels (k, i, j, t) - can be dropped from the model (to reduce the model). These results are shown in Table 4.20.

Random components, ***Model:*** `lmm.RS.C5`	p	logLik	AIC	LRT	*df*	Pr(>Chisq)
`client:g.treaty:g.country:years`	63.00	-720.87	1567.74			
`oci` in (1 + `oci` | `client:g.treaty:g.country:years`)	61.00	-765.89	1653.77	90.03	2	<2e-16 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

Table 4.20. ANOVA-like table with tests of random-effect terms in the model, for crossed random effects model - with model `lmm.RS.C5` (against the model with no random components "<none>")). Specifically, we test if the random slope effects of `oci` is significant in our model.

Based on the output, we reject the null hypothesis (p-value < 0.05), namely:

$$H_0 : \sigma_1^2 = 0 \qquad \text{versus} \qquad H_0 : \sigma_1^2 > 0,$$

with estimated LRT static of 90.03), and conclude that the random `oci` slop effects cannot be dropped from our model. Hence, the model specification given in Equation (4.33) with varying crossed intercepts and varying slope is our preferred model. Specifically, these models allow for the risk class levels (k, i, j, t) (grouping risk class factors) as varying random intercepts - with random slopes influenced by the number of unique natural disasters (i.e. `oci`) they insure (for example, Hurricane Maria, Hurricane

Katrina, etc.) - taking into account the hierarchical structure with respect to the estimated log loss ratio.

2. The Residual Covariance Structure

In this section, we aim to briefly explore other appropriate structures of the variance-covariance matrix of the residuals. Our goal here is to capture more variability that cannot be explained by fixed effects or the multiple random varying effects (for instance, between the risk class levels (k, i, j)). Here, we extend the selected LMM (`lmm.RS.C5`) to explore more sophisticated residual covariance structures. Since, based on our previous findings, there still exists some heterogeneity between the risk classes (k, i, j, t).

For this reason, in this study, we explore the following residual covariance structures: the *compound symmetry*, defined as

Compound Symmetry: a covariance matrix where all variances and correlations are assumed to be equal. Useful for within-subjects study designs observed under the same conditions (Bentler and Bonett, 1980). The residual covariance matrix can be formulated as

$$\mathbf{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

Unstructured covariance matrix: unlike compound symmetry, no assumptions are made regarding the variances' values or covariance. Here we can allow for heterogeneity between the variances for repeated measures. For example, for a five-parameter model, the general form (with symmetry) can be given by:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

Further details on the derivations regarding more complex covariance structures are discussed in Frees et al. (2014). Since the package `lme4` has no option for dealing with heteroscedasticity (previously utilized for LMM modelling). For this reason we utilize the package `nlme`, shown in Table 4.21. See Gałecki and Burzykowski (2013) for a detailed description of the differences between the packages concerning fitting linear mixed models in `R`.

Name	nlme function
Compound symmetry	corCompSymm
AR1	corAR1
CAR1	corCAR1
Unstructured	corSymm

Table 4.21. Example of functions available in `nlme` package - for defining different residual covariance structures, in `R`.

Model Fitting and Results (with Different Covariance Structures)

To model the compound symmetric covariance structure, allowing for heterogeneous and homogeneous in the variance (for OCI), we use the following R functions: `gls` to model the LMM from the package `nlme`, `corCompSymm` to define the compound symmetric covariance structure in `correlation`, and `weights` is used to define if the model should allow for homogeneous in the variance (for the random components).

We use the following R commands to fit the model `lmm.RS.C5` - with compound symmetric covariance structure with homogeneity.

```
## Model lmm.RS.C5.Comp1: compound symmetry structure - with heterogeneous variance
lmm.RS.C5.Comp1 <- gls(model= ln.lr ~    years+ t.type + oci + rate  + cpi + years:oci + years:cpi + t.type:oci +
    oci:cpi, na.action = na.omit, dat =  cat_data,
    ## Covariance structure
        correlation = corCompSymm(form = ~ 1 + as.numeric(oci) |client/g.treaty/oci),
    ## Variance structure: heterogeneous
      weights  = varIdent(form = ~ 1 |oci)
```

The estimated marginal variance covariance of the model structure can be derived using inbuilt functions in `nlme`; using the function `getVarCov`.

```
getVarCov(lmm.RS.C5.Comp1)
```

```
Marginal variance covariance matrix
     [,1]
[1,]   16
  Standard Deviations: 4
```

While estimated variance components (with respect to the correlation structure, and standard deviations or standardized residuals) with respect to the random effects can be extracted by the model summary.

```
summary(lmm.RS.C5.Comp1)
```

```
Generalized least squares fit by REML
  Model: ln.lr ~ years + t.type + oci + rate + cpi + years:oci + years:cpi +      t.type:oci + oci:cpi
  Data: cat_data

Correlation Structure: Compound symmetry
 Formula: ~as.numeric(oci) | client:g.treaty:g.country:years
 Parameter estimate(s):
Rho
  0
Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | oci
Parameter estimates:
1 2 3 4 5 6 7 8
1 1 1 1 1 1 1 1

Standardized residuals:
  Min    Q1   Med    Q3   Max
-3.67 -0.48 -0.01  0.46  2.62

Residual standard error: 4
Degrees of freedom: 290 total; 225 residual
```

We note here that this model's correlation coefficient, denoted by ρ, is estimated at $\hat{\rho} = 0$. Similarly, the model with this same ***compound symmetric*** covariance structure but without allowing for

heterogeneous effects in the covariance - can be fitted by the following R commands - without weights, labelled as model `lmm.RS.C5.Comp2` (where the results of the estimated marginal compound covariance matrix and the summary is also obtained).

```
## Model lmm.RS.C5.Comp2: compound symmetry structure - with heterogeneous variance
lmm.RS.C5.Comp2<- gls(model= ln.lr ~ years+ t.type + oci + rate  + cpi + years:oci + years:cpi + t.type:oci + oci:cpi,
    na.action = na.exclude, data = cat_data,
    ## Covariance structure: symmetry is the only restriction
    correlation = corCompSymm(form = ~ as.numeric(oci)|client/g.treaty/g.country/years))
    ## Variance structure: homoegenous
summary(lmm.RS.C5.Comp2)
getVarCov(lmm.RS.C5.Comp2)
```

```
Generalized least squares fit by REML
  Model: ln.lr ~ years + t.type + oci + rate + cpi + years:oci + years:cpi +      t.type:oci + oci:cpi
  Data: cat_data

Correlation Structure: Compound symmetry
 Formula: ~as.numeric(oci) | client/g.treaty/g.country/years
 Parameter estimate(s):
Rho
  0

Standardized residuals:
  Min    Q1   Med    Q3   Max
-3.67 -0.48 -0.01  0.46  2.62

Residual standard error: 4
Degrees of freedom: 290 total; 225 residual

### Estimated Marginal Variance Covariance
Marginal variance covariance matrix
     [,1]
[1,]   16
  Standard Deviations: 4
```

From this output, we see that the estimated marginal variance-covariance matrix of `lmm.RS.C5.Comp1` is the same as `lmm.RS.C5.Comp2`. Both results show that the estimated correlation parameter $\hat{\rho}$ is 0 within the residuals, with an estimated marginal variance-covariance matrix with respect to the random intercept (crossed effects of risk classes) of $\hat{\sigma}^2_\alpha = 16$.

Additionally, we also fit and specific an ***unstructured*** covariance matrix into our model (using both functions `gls` and `corSymm`) - where here, the symmetry in the covariance structure is the only restriction. We fit these models using the REML method, with a heterogeneous variance structure for the risk classes (k, i, j, t).

```
## Model lmm.RS.C5.Unstr1: Unstructured covariance matrix
lmm.RS.C5.Unstr1<- gls(model= ln.lr ~    years+ t.type + oci + rate  + cpi + years:oci + years:cpi + t.type:oci +
    oci:cpi,
    na.action = na.exclude,
        data        = cat_data,
          correlation = corSymm(form = ~ 1|client/g.treaty/g.country/years),
        ## Variance structure: heterogeneous
        weights     = varIdent(form = ~ 1|oci),  method = "REML")
summary(lmm.RS.C5.Unstr1)
getVarCov(lmm.RS.C5.Unstr1)
```

```
Generalized least squares fit by REML
  Model: ln.lr ~ years + t.type + oci + rate + cpi + years:oci + years:cpi +      t.type:oci + oci:cpi
  Data: cat_data

Correlation Structure: General
 Formula: ~1 | client/g.treaty/g.country/years
```

```
 Parameter estimate(s):
 Correlation:
  1 2 3 4 5
2 0
3 0 0
4 0 0 0
5 0 0 0 0
6 0 0 0 0 0

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | oci
 Parameter estimates:
1 2 3 4 5 6 7 8
1 1 1 1 1 1 1 1

Standardized residuals:
  Min    Q1   Med    Q3   Max
-3.67 -0.48 -0.01  0.46  2.62

Residual standard error: 4
Degrees of freedom: 290 total; 225 residual

### Estimated Marginal Variance Covariance
Marginal variance covariance matrix
     [,1]
[1,]   16
 Standard Deviations: 4
```

We observe from these results that all models estimate that $\hat{\rho} = 0$, with a marginal variance covariance of $\hat{\sigma}_{\alpha} = 16$, and residual s.e. $\hat{\sigma}_{\varepsilon} = 4$. We now compare these models based on the information criteria - in addition to the model selected in the previous section, model `lmm.RS.C5` (given by Equation 4.33).

Model Results: `lmm.RS.C5`,`lmm.RS.C5.Comp1`,`lmm.RS.C5.Comp2` and `lmm.RS.C5.Unstr1`

First, using the statistical LRTs available from package `lmerTest` (for random effects variance structures); we can test if each the model with the compound symmetric covariance structures with homogeneous or heterogeneous variance or the unstructured, heterogeneous structure - at a 5% significance level - statistically improves the model fits, sequentially. The ANOVA-like table for the results is given in Table 4.22. From here, we observe that, at p-value $= 1$, we reject the null hypothesis and conclude that both models `lmm.RS.C5.Comp2` and `lmm.RS.C5.Unstr1` does not improve the model fit of `lmm.RS.C5.Comp1`. We can also compare the model performances - with the selected model in the previous section. The performances of the above-fitted models, with different residual covariance structures, are shown in Table 4.23.

Model	Formula	df	AIC	BIC	logLik	Test	p-value
(1) lmm.RS.C5.Comp1	gls(model = ln.lr∼ years + t.type + oci + rate + cpi + years:oci + years:cpi + t.type:oci + oci:cpi, correlation = corCompSymm(form = ∼as.numeric(oci) | client/g.treaty/g.country/years), weights = varIdent(form = ∼1 |oci),	67	1603.45	1832.33	-734.73		
(2) lmm.RS.C5.Comp2	gls(model = ln.lr∼ years + t.type + oci + rate + cpi + years:oci + years:cpi + t.type:oci + oci:cpi, correlation = corCompSymm(form = ∼as.numeric(oci) | client/g.treaty/g.country/years)),	74	1617.45	1870.25	-734.73	1 vs 2	1.00
(3) lmm.RS.C5.Unstr1	gls(model = ln.lr∼ years + t.type + oci + rate + cpi + years:oci + years:cpi + t.type:oci + oci:cpi, correlation = corSymm(form = ∼as.numeric(oci) | client/g.treaty/g.country/years), weights = varIdent(form = ∼1 |oci),	88	1645.45	1946.07	-734.73	2 vs 3	1.00

Table 4.22. ANOVA table for testing between different residual covariance structures - based on the regression results of the fitted models, using the generalized least squares (gls) in R.

These results suggest that a more complex covariance structure is not required. This is based on the AIC, AICc and BIC values, where we see an increase across all criteria when we allow for more complex covariance structures. As the previously selected model (lmm.RS.C5) attains the lowest values (at AIC = 1568, AICc = 1603, BIC = 1799) and still results in the lowest RMSE and RSE (at 1.81 and 2.15 respectively). In addition, if we compare the standardized empirical residuals (scaled raw residuals) per client k - in comparison to all models, model lmm.RS.C5 shows the least amount of variability (or, is more homogeneous) estimated between each client k.

Hence, we proceed with our analysis with respect to the selected model - with the simplest covariance structure - in Equation 4.33).

	Name	Model Function (in R)	AIC	AICc	BIC	RMSE	RSE
1	lmm.RS.C5	lmer	1567.74	1603	1798.94	1.81	2.15
2	lmm.RS.C5.Comp1	gls	1617.45	1669	1870.25	3.48	3.95
3	lmm.RS.C5.Comp2	gls	1603.45	1644	1832.33	3.48	3.95
4	lmm.RS.C5.Unstr1	gls	1645.45	1723	1946.07	3.48	3.95

Table 4.23. Model performance summary of LMMs with different residual covariance structures. Performance is compared based on the AIC, AICc, BIC, RMSE and RSE. The first model lmm.RS.C5 is LMM with crossed random effects, with unstructured residual covariance structure. Model (2) lmm.RS.C5.Comp1: crossed random effects, with compound symmetric heterogeneous covariance structure. Model (3) lmm.RS.C5.Comp2: crossed random effects, with compound symmetric homogeneous covariance structure. Model (4) lmm.RS.C5.Unstr1: crossed random effects, unstructured heterogeneous covariance structure with symmetric restrictions.

Model Reduction: for Full Selected Model `lmm.RS.C5`

Before we analyze the residuals and model diagnostics, due to the high number of interaction effects in our model, we now attempt to reduce the number of regression parameters.

Model Results: `lmm.RS.C5`

We can use the function `lmerTest::anova(lmm.RS.C5, type = 3)` - to use Satterthwaite's methods for approximating degrees of freedom to conduct the t and F-tests. To get an idea of which main or interaction terms can be possibly dropped from the full model, `lmm.RS.C5`. In other words, based on the significance level of 5%, we perform ***single term deletions*** using the appropriate F-tests from the full model (utilizing the `R` function `drop1()`, available in the package `lmerTest`). The summary of the results are shown in a ANOVA-like table (Table 4.24). The corresponding p-values given in bold are the significant fixed effects associated with the log loss ratio, given the risk classes (k, i, j, t). Alternatively, we can also use `Anova()` to perform Type III Wald χ^2 tests (suitable for fixed effects,

Fixed Covariate Effects (Main and Interaction Effects)	Sum Sq	Mean Sq	Num. *df*	Den. *df*	F-value	Pr($>F$)
`years`	98.45	5.47	18	180.02	1.19	0.28
`t.type`	45.78	15.26	3	196.36	3.31	**0.02**
`oci`	47.82	47.82	1	58.16	10.37	**2.1e-3**
`rate`	52.60	52.60	1	223.85	11.41	**9.2e-3**
`cpi`	4.17	4.17	1	96.72	0.90	0.34
`years:oci`	93.28	6.22	15	72.67	1.35	0.19
`years:cpi`	118.46	7.90	15	176.33	1.71	**0.04**
`t.type:oci`	37.01	12.34	3	137.36	2.68	**0.03**
`oci:cpi`	1.26	1.26	1	72.65	0.27	0.60

Table 4.24. Single term deletions, using approximate F-tests - based on the Satterthwaite's method for the fixed-effects, for LMMs. The ANOVA-like table is constructed, for Type III hypothesis tests. Recall *"Num. df"* is the numerator degrees of freedom (number of groups minus one), and *Den. df* is the denominator degrees of freedom (difference between the number of groups and observations).

with interactions). These results are shown in the Appendix, Table 1. For more details on hypothesis testing fixed effects in LMMs, specifically for `lmer` model fits - see (Kuznetsova et al., 2017). We see from both ANOVA tables that the interaction fixed effects: `years:oci` (p-value = 0.19) and `oci:cpi` (p-value = 0.6) - is no longer significant (at a 5% significance level) on the estimated log loss ratios when accounting for the crossed random effects of (k, i, j, t).

Hence, we now refit our model by removing these interaction fixed effects and compare this reduced model (labelled as `lmm.RS.C5.r`) to our mixed-effects "full" model, to see if removing these pairwise effects improves the model fit (using ML estimation and χ^2 tests). The results are shown in Table 4.25. The reduced model is the better model with lower AIC =1554.13 and BIC = 1554.13 (where we fail to reject the null hypothesis that the pairwise effects are associated with our response, with p-value = 0.3 and χ^2_{16}).

This means, that we will continue our analysis with the reduced model `lmm.RS.C5.r` without having the interaction fixed effects `years:oci` and `oci:cpi`. Thus, this is our final selected LMM model - with crossed random effects and formulated as:

	p	AIC	BIC	logLik	deviance	χ^2	Df	$\Pr(>\chi^2)$
`lmm.RS.C5.r`	47.00	1554.13	1726.61	-730.06	1460.13			
`lmm.RS.C5`	63.00	1567.74	1798.94	-720.87	1441.74	18.39	16	0.3

Table 4.25. ANOVA table, using χ^2 tests to compare the refitted reduced model, `lmm.RS.C5.r`, and the full model, `lmm.RS.C5`, with ML estimation (instead of REML). p here denotes the number of fixed effects parameters in the model.

Model `lm.RS.C5.r` *(final reduced LMM selected - with crossed random effects)*

$$\begin{aligned}\texttt{ln.lr}_{kijt} = \beta_0 + \sum_{t=2}^{18}\beta_1^t \texttt{years}_{kij}^t + \sum_{q=2}^{4}\beta_2^q \texttt{t.type}_{ki}^q \quad + \ \beta_3 \texttt{oci}_{kijt} + \beta_4 \texttt{rate}_{kijt}^d \\ + \beta_5 \texttt{cpi}_{kijt} + \sum_{t=2}^{18}\beta_6^t \texttt{years}_{kij}^t \times \texttt{cpi}_{kijt} \quad + \sum_{q=2}^{4}\beta_7^q \texttt{t.type}_{ki}^q \times \texttt{oci}_{kijt} \\ + \ \alpha_{0kijt} + \alpha_{1kijt}\texttt{oci}_{kijt} + \varepsilon_{kijt},\end{aligned} \tag{4.34}$$

such that for the random components, we have

$$\begin{aligned}\varepsilon_{kijt} \quad &\sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad i.i.d., \quad \forall\ k, i, j, t, \\ \begin{bmatrix}\alpha_{0kijt} \\ \alpha_{1kijt}\end{bmatrix} \quad &\sim \boldsymbol{\mathcal{N}}\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}\sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2\end{bmatrix}\right) \quad i.i.d., \quad \forall\ k = 1, \ldots, 35.\end{aligned} \tag{4.35}$$

The results of the corresponding final LMM selected are given in Table 5.2.

Model `lmm.RS.C5.r` *(Crossed Random Effects, reduced model)*			
Fixed Effects			
Predictors	*Estimates*	*CI*	*p*
(Intercept)	26.38	3.50 – 49.25	**0.024**
`years [2003]`	53.38	-39.41 – 146.18	0.258
`years [2004]`	-23.01	-47.18 – 1.17	**0.062**
`years [2005]`	-37.59	-67.43 – -7.75	**0.014**
...	...	...	...
`t type [CXL]`	2.95	0.08 – 5.83	**0.044**
`t type [QS]`	-0.83	-2.37 – 0.72	0.293
`t type [SP]`	1.07	-1.87 – 4.01	0.473
`oci`	-6.25	-7.45 – -5.05	**<0.001**
`rate`	-0.35	-0.55 – -0.14	**0.001**
`cpi`	-0.29	-0.56 – -0.01	**0.042**
`years [2003]:cpi`	-0.64	-1.77 – 0.49	**0.264**
`years [2004]:cpi`	0.31	0.02 – 0.61	**0.038**
`years [2005]:cpi`	0.46	0.10 – 0.81	**0.012**
...	...	...	...
`t.type [CXL]:oci`	-2.35	-4.37 – -0.33	**0.023**
`t.type [QS]:oci`	0.07	-0.71 – 0.86	0.854
`t.type [SP]:oci`	-1.48	-3.75 – 0.79	**0.200**
Random Effects			
$\hat{\sigma}_{\varepsilon}^2$			4.54
$\hat{\sigma}_0$ (Intercept, `client:g.treaty:g.country:years`)			18.59
$\hat{\sigma}_1$ (Slope, `client:g.treaty:g.country:years:oci`)			13.85
$\hat{\rho}_{01}$ (Correlation, of `client:g.treaty:g.country:years`)			-0.98
No. of risk class levels `client` k			35
No. of risk class levels `g.treaty` i			2
No. of risk class levels `g.country` j			2
No. of risk class levels `years` t			19
Total Observations			290
$ICC_{adj.}$ / Conditional $ICC_{con.}$			0.84 / 0.26
Marginal $R^2_{mar.}$ / Conditional $R^2_{con.}$			0.690 / 0.951

Table 4.26. Model results of `lmm.RS.C5.r`, reduced final selected LMM to estimate the log loss ratio, $\texttt{ln.lr}_{kijt}$ - with crossed random effects of risk class levels (k, i, j, t), and random `oci` slope effects. Results also show the estimated correlation among the crossed random risk class levels, in the intercept $\hat{\rho}_{01}$, with the corresponding number of risk class groups for all levels considered in $(k, i, j, t,)$, and the estimated coefficients of the types of R^2, and the intraclass correlation coefficient, ICC.

4.2.1 Model Diagnostics and Residual Analysis

We now create residual diagnostic plots (using the `lmer` package in `R`, by the function `resid_panel`, and the package `redres`). Recall, by analyzing these plots, we can evaluate the goodness of fit of the reduced model to visually check if our reduced LMM, `lmm.RS.C5.r`, meets our statistical requirements. Specifically, we check if "strong" indications exist against our normality or variance homogeneity assumptions and potentially identify the outliers (similar to LMs).

In other words, we check the following assumptions of the LMM (by using the corresponding plots):

1. ***Linear relationship*** between the response and the explanatory variables
   - *plots:* response versus predicted plots (see Figure 4.8).
2. ***Constant variance*****:** variance of the model's Pearson residuals are approximately constant (i.e. variance homogeneity)
   - *plots*: residual plot, Pearson residuals versus predicted values (shown in Figure 4.8).
3. ***Normality*****:** normal distributional assumptions of the random effects and error term
   - for the random errors only, we look at histograms for all types of residuals considered (see Figure 4.9), while
   - for both *random errors and components* - we look the normal Q-Q plots (for all residual types, shown in Figure 4.10 and Figure 4.11).

In this study, we compute both the ***marginal*** and ***conditional*** residuals. Analyzing the conditional residuals allows us to check the effects of the random components - while marginal residuals that do not account for the random effects (i.e. considers only the fixed effects of the covariates - without the effects of the risk class levels k, i, j for each client k).

Additionally, first consider the general formulation of the (k-th level) LMM given in Equation 3.21,

$$\begin{aligned} \boldsymbol{Y}_k &= \boldsymbol{X}_k\boldsymbol{\beta} + \boldsymbol{Z}_k\boldsymbol{\alpha}_k + \varepsilon_k, \\ \boldsymbol{\alpha}_k &\sim N(\mathbf{0}, \boldsymbol{D}), \\ \varepsilon_k &\sim N\left(\mathbf{0}, \boldsymbol{\Sigma}_k\right), \end{aligned} \tag{4.36}$$

where generally we assume

$$\begin{bmatrix} \alpha \\ \varepsilon_k \end{bmatrix} \sim \boldsymbol{\mathcal{N}}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{D} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma_k}, \end{bmatrix} \right),$$

for the response vector $\boldsymbol{Y}$, fixed components vector: $\boldsymbol{X}_k\boldsymbol{\beta}$, and random components $\boldsymbol{Z}_k\boldsymbol{\alpha}_k + \varepsilon_k$ (where $\boldsymbol{Z}_k$ denotes the random covariates, with random slopes) - for all client $k = 1, \ldots, K$.

Then, for this LMM (using the `R` package `redres`) we can compute the three types of residuals (for both marginal and conditional) - to investigate our distributional assumptions:

Raw Residuals: measured by the difference between the values of the observed response, Y_k, and the predicted response, $\hat{\boldsymbol{Y}}_k$, for each client k.

Lemma 4.1: Raw Residuals for LMMs

- ***marginal raw residuals*** (based on the fixed effects only)

$$r_k^m = Y_k - \mathbf{x}_k'\widehat{\boldsymbol{\beta}}, \tag{4.37}$$

- ***conditional raw residuals*** (based on the mixed effects, i.e. accounts for both fixed and random effects)

$$r_k^c = Y_k - \boldsymbol{X}_k\boldsymbol{\beta} - \mathbf{z}_k^T\hat{\boldsymbol{\alpha}}, \tag{4.38}$$

Here, the conditional raw residual accounts for the random effects, while the marginal version does not.

Studentized Residuals: is computed by dividing the raw residuals (given by the above equations) square root of the estimated variance of the raw residuals.

Lemma 4.2: Studentized Residuals for LMMs

- ***marginal studentized residuals*** (based on the fixed effects only)

$$r_k^{m,std.} = \frac{r_k^m}{\sqrt{\widehat{Var}\left[r_k^m\right]}}, \tag{4.39}$$

- ***conditional studentized residuals*** (based on both the random and fixed effects)

$$r_k^{c,std.} = \frac{r_k^c}{\sqrt{\widehat{Var}\left[r_k^c\right]}}, \tag{4.40}$$

where $\widehat{Var}\left[r_k^m\right]$ and $\widehat{Var}\left[r_k^c\right]$ is the estimated variance of the marginal and conditional raw residuals respectively. For more details on how this is derived, see Grégoire et al. (1995) (also discussed in Section 3.4.1).

Pearson Residuals: calculated by the type of raw residuals divided by the square root of the estimated variance of the response values, $\boldsymbol{Y_k}$.

Lemma 4.3: Pearson Residuals for LMMs

- ***marginal Pearson residuals*** (estimated for fixed effects only)

$$r_k^{m,per.} = \frac{r_k^m}{\sqrt{\widehat{\text{Var}}\left[Y_k\right]}},$$

- ***conditional Pearson residuals*** (estimated for random and fixed components)

$$r_k^{c,per.} = \frac{r_k^c}{\sqrt{\widehat{\text{Var}}\left[Y_k \mid \boldsymbol{\alpha}_k\right]}},$$

where the estimated variance is based on random and fixed components, i.e. $\text{Var}\left[\boldsymbol{Y}_k\right] = \boldsymbol{Z}_k \boldsymbol{D} \boldsymbol{Z}_k' + \boldsymbol{\Sigma}_k$. (see Section 3.4.1).

Based on these definitions, we now screen the corresponding residual plots and observe the following:

- From Figure 4.8, which shows the Pearson residuals versus the predicted values, first panel (conditional, i.e. accounting for both the fixed and random components). We cannot visually detect any structural dependency. Hence there are no indications that the assumption regarding variance homogeneity is not fulfilled.
- Figure 4.8 (third panel, first row, third column) shows no indications against our linearity assumptions between the observed response and predicted values (`ln.lr`).
- From the histograms in Figure 4.9 and Normal Q-Q plots in Figure 4.10, we check if our assumptions regarding the random errors are normally distributed. From these results, we detect deviations away from normal theoretical fit. Observations with high absolute studentized residuals here can indicate violations against our model distributional assumptions or that the corresponding observations are outliers.
- We detect outliers using the interquartile range (IQR) (by 1.5) to remove them from our data set (16 outliers in total were removed) and then refit the model. The results of the normal Q-Q plots with outliers removed are shown in Figure 4.10 (second Row). From here, the results show after the outliers are removed, there are no longer any strong indications against our normal distribution assumptions on the random errors.
- To check the distributional assumptions for the random effects, we look at the Normal Q-Q plots shown in Figure 4.11. The plots show the studentized residuals of the empirical best linear unbiased predictors (EBLUPs, see Section 3.2.2 with respect to the derivation). From here, we detect a few more outliers towards the lower tail, but there exist no signs of any strong violations against our normality assumptions on the random components.

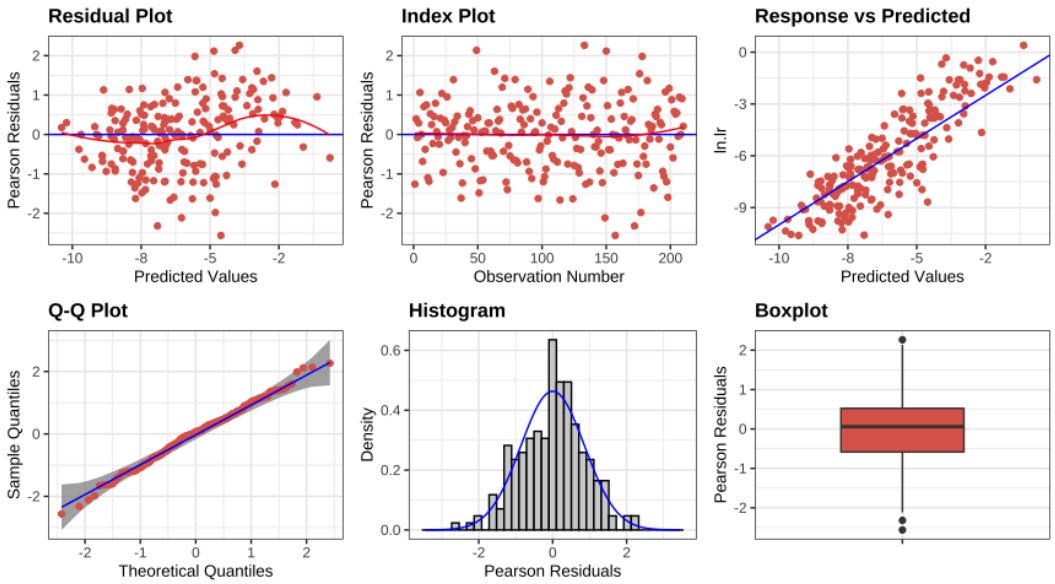


Figure 4.8. Residual plots based on model `lmm.RS.C5.r` (reduced model of crossed risk class levels (k, i, j, t) with main and interaction effects). Plots was obtained by `R resid_pane` available in package `lmer`. Plots include: (first plot, Row 1, Column 1) the Pearson residual plots for the estimated residuals versus the predicted values of the response, $\widehat{\texttt{ln.lr}}_{kijt}$, the Pearson residuals versus the observations (index plot), the response versus the predicted values, the Normal Q-Q plot of the Pearson residuals, and the corresponding histogram and boxplot of the calculated Pearson residuals of the model.

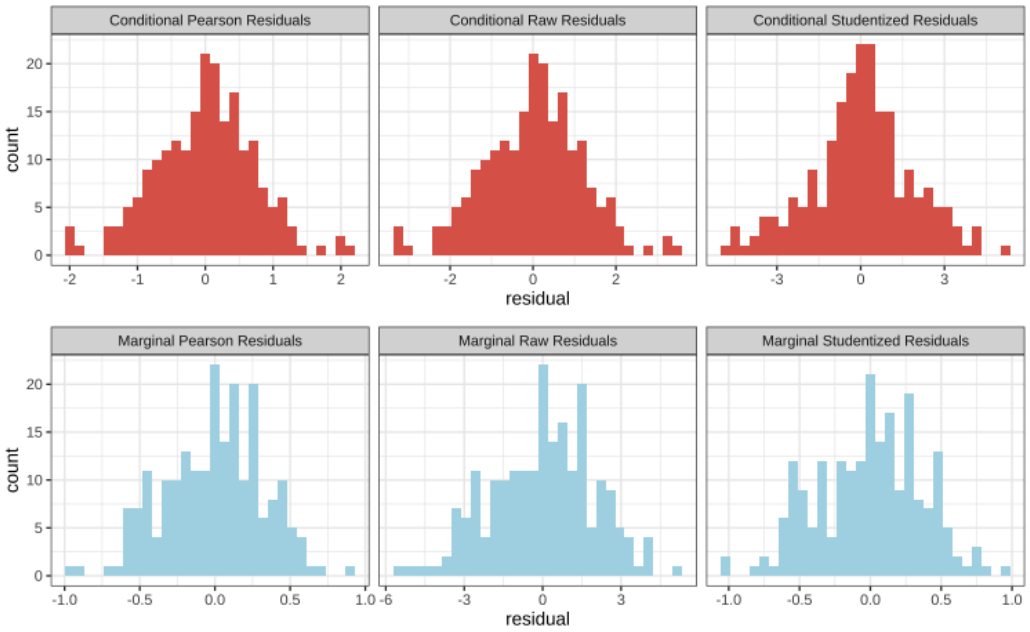


Figure 4.9. Histogram of estimated residuals (random errors), based on model `lmm.RS.C5.r`. Top Row: shows the conditional residuals (which account for both the fixed and random effects in the model). Bottom Row: shows the marginal distribution of the marginal residuals - which only accounts for the fixed effects in the model. Histograms of the following residuals are shown: Pearson, raw, and studentized residuals.

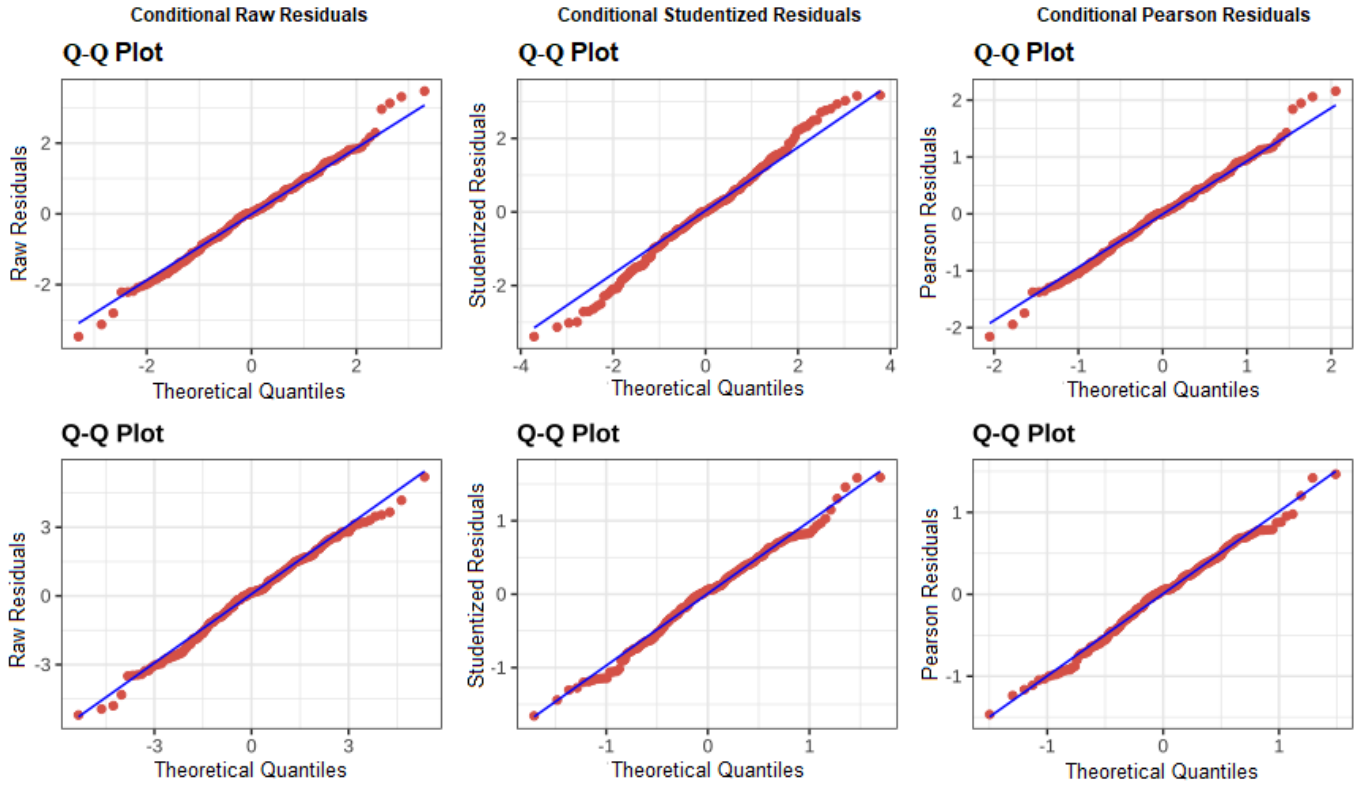
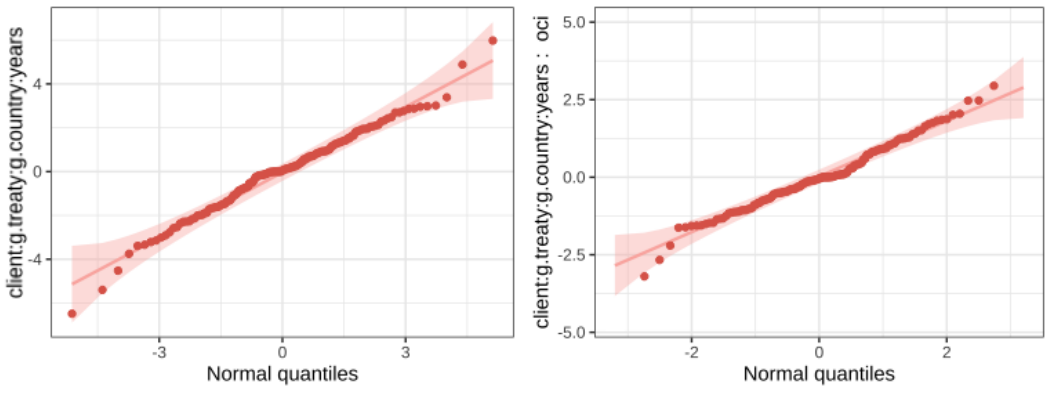


Figure 4.10. Normal Q-Q plots, for conditional residuals (based on the random errors, accounting for both fixed and random components in the model, `lmm.RS.C5.r`). Top Row shows the raw, studentized and Pearson residuals, based on the raw (aggregated) data (labelled as `cat_data`). Bottom row: shows the model's estimated raw, studentized and Pearson residuals - with outliers removed (using interquartile range (IQR) by 1.5).



copy.png

Figure 4.11. Normal Q-Q plots of the standardized (i.e. studentized) residuals of the empirical best linear unbiased predictors (EBLUPs) of model `lmm.RS.C5.r` (based on the data with outliers removed). Left panel: shows the Q-Q plot computed for the crossed random intercept effects, of the risk classes (k, i, j, t). Right panel: shows the Q-Q plot of the standardized residuals based on the EBLUPs, accounting for both the random intercepts and random slope `oci` - for the k-th client, i-th treaty, j-th country during t years.

4.3 Generalized Linear Models for Natural Catastrophic Data

This section models the natural catastrophic data using generalized linear models (GLMs). GLMs have many advantages over LMs and LMMs - since (in our case) the linear models are only applicable after log transformations of the loss ratio LR. This means that interpretations of estimates and predictions are not straightforward because back-transformation to the original scale is required for linear models. Modelling on the original scale is no longer an issue for GLMs and GLMMs.

For this study, we consider the gamma GLM with the log link (often used in various insurance applications, see Frees et al. (2014) for examples). A positive continuous response variable Y follows a gamma probability density function of the form

$$f(y;\mu,k) = \frac{(k/\mu)^k}{\Gamma(k)} e^{\frac{-ky}{\mu}} y^{k-1}, \quad y > 0.$$

with mean parameter μ and shape parameters ν such that $Y \sim \Gamma(\mu,\nu)$. See Ng et al. (2019) for more examples of different parameterizations. Since the gamma model with the log-link generates a multiplicative model, the gamma GLM - with covariates $\boldsymbol{x}_{kijt}$ and regression parameters $\boldsymbol{\beta}$ - for the loss ratio LR_{kijt} for a client k with a treaty i providing coverage in country j, during the year t becomes

$$\mu_{kijt} = E[LR_{kijt}|\boldsymbol{x}_{kijt}] = \exp(\boldsymbol{x}_{kijt}\boldsymbol{\beta}), \tag{4.41}$$

where $LR_{kijt} \sim \Gamma(\mu_{kijt},\nu)$ and μ_{kijt} is the estimated yearly loss ratio for the k-th client, with $i \in P_g^k$, $j \in \mathcal{M}_g^{ki}$ given the years $t \in \mathcal{T}^{kij}$.

Our strategy here is once again to sequentially build a bigger and bigger model (similar to the model building and comparison for LMs and LMMs) starting with the initial subset of risk factors (and removing or adding variables accordingly). Specifically, here we analyze and compare the following models in this section:

Model `glm.M` *(main effects model)* models the the main (fixed) risk factor effects on the loss ratio, `lr`, which are given at different risk class levels, i.e. at the k-th client level with treaty i for j-th country in the t-th year. While not accounting for the nested or clustering effects between the risk class levels (k,i,j,t) individually (i.e. complete pooling of risks). This gamma model (with the log-link) can be formulated as

Model `glm.M` *(main effects model with risk factors only)*

$$\begin{aligned} E[\texttt{lr}_{kijt}] = \ & \beta_0 + \sum_{q=2}^{4} \beta_1^q \texttt{t.type}_{ki}^q + \sum_{c=2}^{3} \beta_2^c \texttt{cob}_{kij}^c + \sum_{d=2}^{3} \beta_3^d \texttt{peril}_{kij}^d \\ & + \beta_4 \texttt{rate}_{kijt} + \beta_5 \texttt{cpi}_{kijt} + \beta_6 \texttt{oci}_{kijt} + \ \varepsilon_{kijt}, \end{aligned} \tag{4.42}$$

where the categorical predictor variables take on the values given in Equation 4.11. Recall, the loss ratio is measured by the loss amount X_{kijt} (`loss`) per exposure volume v_{kijt} (`volume`), for each k-th client (i.e. $LR_{kijt} = \frac{X_{kijt}}{v_{kijt}}$).

For this reason we introduce the log of exposure as an offset as a part of the linear predictor and model the losses of the risk classes (i, j, t) for the k-th client directly as

$$\begin{aligned}\ln(E[\texttt{loss}_{kijt}]) = \; & \beta_{0,} + \sum_{q=2}^{4} \beta_1^q \texttt{t.type}_{ki}^q + \sum_{c=2}^{3} \beta_2^c \texttt{cob}_{kij}^c + \sum_{d=2}^{3} \beta_3^d \texttt{peril}_{kij}^d \\ & + \beta_4 \texttt{rate}_{kijt} + \beta_5 \texttt{cpi}_{kijt} + \beta_6 \texttt{oci}_{kijt} + \ln(\texttt{volume}_{kijt}) + \varepsilon_{kijt},\end{aligned} \tag{4.43}$$

where $X_{kijt} \sim \Gamma(\mu_{kijt}, \mu kijt)$ is measured for each k-th client for risk class levels (k, i, j, t) i.i.d, and the risk class levels take on the values given in Equation 4.4.

With expectation 1 and variance ϕ, for the response random variable, this also implies that the estimate of the expectation μ_{kijt} (in Equation 4.41) is given by $\hat{\mu}_{kijt} = \exp\big(\boldsymbol{x}_{kijt}\hat{\boldsymbol{\beta}}\big)$.

Whereas, in this gamma regression model, for example, for all $k = 1, \ldots, K$ clients, the "total" estimate of the dispersion parameter ϕ is given by

$$\widehat{\phi} = \frac{1}{k-p} \sum_{k=1}^{K} \left(\frac{\texttt{loss}_k - \widehat{\mu}_k}{\widehat{\mu}_k} \right)^2, \tag{4.44}$$

at the k-th client level and p regression parameters (including the intercept). Note, this holds for all other gamma regression models considered in this section.

Model `glm.G`: models the `ln.lr` (or the loss amount X_{kijt} directly with the exposure volume as an offset v_{kijt} (`volume`) - for each cluster group: k-th client with treaty i, and j-th country per t-th year. Including the main (fixed) risk factor effects with clustering effects of (k, i, j, t) as variables individually (i.e. no pooling of risks). This model can be formulated as

Model `glm.G` *(main effects model with risk factors and risk class groups)*

$$\begin{aligned}\ln(E[\texttt{loss}_{kijt}]) = \; & \beta_0 + \sum_{k=1}^{35} \beta_1^k \texttt{client}_k + \beta_2 I_k^{\texttt{g.treaty}_i} + \beta_3 I_{ki}^{\texttt{g.country}_j} + \sum_{t=2}^{18} \beta_4^t \texttt{years}_{kij}^t \\ & + \sum_{q=2}^{4} \beta_5^q \texttt{t.type}_{ki}^q + \sum_{c=2}^{3} \beta_6^c \texttt{cob}_{kij}^c + \sum_{d=2}^{3} \beta_7^d \texttt{peril}_{kij}^d + \beta_8 \texttt{rate}_{kijt} \\ & + \beta_9 \texttt{cpi}_{kijt} + \beta_{10} \texttt{oci}_{kijt} + \ln(\texttt{volume}_{kijt}) + \varepsilon_{kijt},\end{aligned} \tag{4.45}$$

where the is the loss amount for the k-th client for the i-th treaty of country j-th in year t. This models allows for the heterogeneity (individually) between the risk classes for each client k, where we treat the risk class levels as explanatory variables into our model.

Model `glm.I` based on the findings of main effects, from `glm.M`, we then explore if including interaction effects improves our model - to estimate the loss ratio (i.e. the loss amount given the exposures) for the $k = 1, \ldots, K, i \in \mathcal{P}_G^k, j \in \mathcal{M}_G^{ki}$ given the years $t \in \mathcal{T}^{kij}$ (i.e. accounting for individual cluster effects).

Model glm.M: Main Fixed Effects - Without Clustering Effects

We use the glm function in R, to fit the following gamma model (with main risk factors only) - with a log link and the offset of the log exposure (volume). Recall, volume is the total sum insured for the respective loss amount, based on the risk given for each k-th client.

```
glm.M = glm(formula = loss ~   t.type + cob + peril + oci +  rate + cpi, offset = log(volume),family = Gamma(link =
    "log"),data = cat_data, na.action = na.omit)
```

```
Call:
glm(formula = loss ~ t.type + cob + peril + oci + cpi + rate,
    family = Gamma(link = "log"), data = cat_data, na.action = na.omit,
    offset = log(volume))

Deviance Residuals:
   Min      1Q  Median      3Q     Max
 -4.62   -2.53   -1.67   -0.24    4.99

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.1379     1.2618    -1.7    0.091 .
t.typeCXL     2.7597     0.5876     4.7   4e-06 ***
t.typeQS      0.1775     0.3956     0.4    0.654
t.typeSP      0.1469     0.5225     0.3    0.779
cobM         -0.5244     0.5874    -0.9    0.373
cobR         -0.8516     0.6543    -1.3    0.194
perilF        1.8606     0.5996     3.1    0.002 **
perilH        3.7045     0.4615     8.0   3e-14 ***
oci           0.0044     0.1184     0.0    0.970
cpi          -0.0302     0.0076    -4.0   1e-04 ***
rate         -0.2541     0.0830    -3.1    0.002 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for Gamma family taken to be 5)

    Null deviance: 1751.7  on 289  degrees of freedom
Residual deviance: 1333.1  on 279  degrees of freedom
AIC: 4282

Number of Fisher Scoring iterations: 25
```

According to this R output $\hat{\beta}_0 = e^{-2.13} = 0.12$ (with s.e. 1.26). The residual deviance is 1333.1 on 279 degrees of freedom. Recall, the goodness of fit for this gamma GLM can be investigated using the residual deviance statistical test (as defined in Definition 4.10). Note, for the gamma regression model, it can be shown that the (scaled) deviance, for all losses at the k-th client level, can be defined as

$$\mathrm{D}(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) = -2 \sum_{k=1}^{K} \left[\ln\left(\frac{\mathtt{loss}_k}{\widehat{\mu}_k} \right) - \frac{\mathtt{loss}_k - \widehat{\mu}_k}{\widehat{\mu}_k} \right]$$

Model assumptions are satisfied as confirmed by the residual deviance test as

$$\frac{D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}})}{\widehat{\phi}} = \frac{1333.1}{5} = 267 \not> \chi^2_{279,0.95} = 319. \tag{4.46}$$

In R, to efficiently perform the residual test, we defined and constructed the R function residual.deviance.test - which also provides us with the p-value of the corresponding test.

```
residual.deviance.test <- function(model){
{
 # dispersion
 phi.main<- summary(model)$dispersion
 # residual deviance
 dev.main <-model$deviance
 p.main <-length(model$coefficients)

 ## Residual deviance test
 resid.dev.test <- dev.main/phi.main > qchisq(1-0.05,nrow(cat_data)-p.main)
 print(c(message('Residual Deviance Test:'),resid.dev.test))

 # p-value
 p.value <- 1 - pchisq(dev.main/phi.main,nrow(cat_data) - p.main) }

return(c(message('p-value:'),p.value))
}
```

Such that, using this function, for `glm.M` the p-value is given as

```
> residual.deviance.test(glm.M)
Residual Deviance Test:
[1] FALSE
p-value:
[1] 0.72
```

However, based on the model results it appears that both `cob` and `oci` may be possibly non-significant. For this reason we perform a partial deviance test, where we refit the model without each risk factor (reduced models) to test between the reduced and full models (`glm.M`). Table 4.27 compares the reduced models `glm.M.r1` (without `cob`) and `glm.M.r2` (without `cob` and `peril`) with the full model `glm.M` using a analysis of deviance table between the models (implemented via the `R` function `anova` with the appropriate χ^2 test).

	Resid. Df	Resid. Dev	*df*	Deviance	Pr(>Chi)
`glm.M.r1` (without `cob`)	280	1333.62			
`glm.M` (full model)	279	1333.13	1	0.49	0.76
`glm.M.r2` (without `cob` and `oci`)	282	1330.00			
`glm.M.r1` (without `cob`)	281	1329.37	1	0.63	0.74

Table 4.27. `glm.M`: Analysis of deviance table to compare the reduced models and the full gamma glm (with log link) with respect to the main risk factors only. Model `glm.M.r1` compares the full model without `cob`, while Model `glm.M.r2` compares the reduced model with `oci`)

Alternatively, we can also calculate the partial deviance test statistics based on the reduced model (without `cob` and `peril`) with the full model.

```
> 1-pchisq((1330.0/5.8)-(1333/5),3)
[1] 1
```

With residual deviance 1330 (with estimated dispersion of 5.8) for the reduced model (without both covariates) and residual deviance 1333 (with estimated dispersion of 5), we get a corresponding large p-value of approximately 1. This means that both risk factors `cob` and `oci`, at the 5% level is not required in the model.

Thus, we proceed our analysis with the our reduced model (denoted `glm.M.r` for simplicity) - without the following risk factors included in the model - given by

Model `glm.M.r` *(reduced model main effects model with risk factors only)*

$$\ln(E[\texttt{loss}_{kijt}]) = \beta_0 + \sum_{q=2}^{4} \beta_1^q \texttt{t.type}_{ki}^q + \sum_{d=2}^{3} \beta_2^d \texttt{peril}_{kij}^d + \beta_3 \texttt{rate}_{kijt} + \beta_4 \texttt{cpi}_{kijt} + \ln(\texttt{volume}_{kijt}) + \varepsilon_{kijt},$$

where once again we have $\mu_{kijt} = E[LR_{kijt}|\boldsymbol{x}_{kijt}] = \exp(\boldsymbol{x}_{kijt}\boldsymbol{\beta})$ for the loss ratio $\texttt{ln.lr}_{kijt} = \frac{\texttt{loss}_{kijt}}{\texttt{volume}_{kijt}}$, following a gamma distribution and $i.i.d$ for all client k with corresponding risk class groups (i, j, t).

Note, the reduced GLM only includes 4 risk factors as covariates: `t.type`, `peril`, `rate` and `cpi`. This means in this model we ignore the risk class group effects in (k, i, j, t) - which may leave out essentially information when modeling the `ln.lr` per k-th client and may also underfit the underlying data. Additionally, the overall fit of this model is not very good since the residual deviance of 1329 is much larger than residual degrees of freedom of 281. Additionally, when comparing the goodness of fits of base models `lm.M.r` (log normal linear model with main fixed risk effects only) and `lmm.M.r` (LMM with main effects only) with the selected model `glm.M.r`; this model yields a much higher AIC of 4275 (compare to `lm.CP` with AIC = 1783). With a higher BIC value of 4308, and thus a log-likelihood value in comparison to model `lmm.RS.C5.r`. For this reason, we conclude that both the log normal model and LMM is a better fit compared to the gamma GLM with the log link, with respect to the main effects of the risk factors only.

However, it is important to note that when comparing goodness of fits of the linear models on the log scale with the gamma regression models (with respect to the variability explained) - the bias in the intercept of the log normal models (selected in previous sections) has to be corrected for. This is due to the gamma model's assumptions of constant coefficient of variation.

Suppose that Y_k denotes the response random variable, such that

$$\frac{\sqrt{\text{Var}\left[Y_k\right]}}{\text{E}\left[Y_k\right]} \equiv \sigma \quad \forall k = 1, \ldots, K,$$

and

$$\text{Var}\left[Y_k\right] = \sigma^2 \mu_k^2 \quad \forall k = 1, \ldots, K,$$

which implies a quadratic effect of the mean on the variance. If we stabilize the variance of the response it's log transformation, it can be shown that,

$$\text{Var}\left[\ln(Y_k)\right] \approx \ln\left(\mu_k\right)^2 + \sigma^2 - \ln\left(\mu_k\right)^2 = \sigma^2.$$

This means that, for a small σ, here we have $\text{E}\left[\ln(Y_k)\right] \approx \ln\left(\mu_k\right) - \sigma^2/2$ and $\text{Var}\left[Y_k\right] \approx \sigma^2$ for all $k = 1, \ldots, K$.

Hence, for the log normal model, if we assume that $\ln(Y_k)$ is normally distributed with mean $\boldsymbol{x}_i^\top \boldsymbol{\beta}$ and variance σ^2, then Y_k is log normally distributed with mean $\text{E}\left[Y_k\right] = \exp\left\{\boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma^2/2\right\}$. Since, it can be shown that the least squares estimates of the regression parameters is *biased* for the intercept, β_0 with an approximate bias of $-\sigma^2/2$. For this reason, the bias in the intercept of the log normal models must be adjusted for, accordingly. Alternatively, we can also use the standard linear model on the log scale - with a multiplicative error structure, which would allow us to deal with this variance structure. More details and for a full derivation on this, see (McCullagh and Nelder, 1983).

Model glm.G: Main Fixed Effects - With Risk Group Effects

Here we investigate the clustering effects of the risk classes in our model. Especially since: (1) glm.M does not account for any heterogeneity between the risk class levels (k, i, j, t) and (2) ignores the clustering effects at the k-th level, which is crucial in our case given the hierarchical structure of data.

We start by fitting the glm.G given by Model (4.46) in R and include the risk factors as covariates. Specifically, we introduce client, g.treaty, g.country and years as covariates into the model.

```
glm.G = glm(formula = loss ~   client + g.treaty + g.country + years + t.type + cob + peril + oci + cpi + rate, offset
    = log(volume),family = Gamma(link = "log"),data = cat_data, na.action = na.omit)
```

```
Call:
glm(formula = loss ~ client + g.treaty + g.country + years +
    t.type + cob + peril + oci + cpi + rate, family = Gamma(link = "log"),
    data = cat_data, na.action = na.omit, offset = log(volume))

Deviance Residuals:
   Min      1Q  Median      3Q     Max
 -3.92   -1.51   -0.46    0.38    2.83

Coefficients: (1 not defined because of singularities)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -3.912      1.703    -2.3    0.023 *
client6       -0.730      0.690    -1.1    0.292
client7       -2.062      0.713    -2.9    0.004 **
client8       -2.328      0.787    -3.0    0.003 **
client9       -1.623      0.757    -2.1    0.033 *
...
g.treaty2      2.675      1.595     1.7    0.095 .
g.countryL     0.426      0.372     1.1    0.253
years2003      3.533      1.161     3.0    0.003 **
years2004      2.705      1.024     2.6    0.009 **
years2005      1.568      1.037     1.5    0.132
years2007      1.405      1.050     1.3    0.182
...
t.typeCXL     -0.335      0.500    -0.7    0.503
t.typeQS      -0.434      0.400    -1.1    0.279
t.typeSP      -1.170      0.511    -2.3    0.023 *
cobM           0.432      0.543     0.8    0.427
cobR          -0.508      0.476    -1.1    0.287
perilF         1.815      0.420     4.3    2e-05 ***
perilH         2.957      0.385     7.7    5e-13 ***
oci           -0.116      0.096    -1.2    0.228
cpi           -0.025      0.010    -2.5    0.014 *
rate          -0.218      0.078    -2.8    0.005 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for Gamma family taken to be 1.7)

    Null deviance: 1751.73  on 289  degrees of freedom
Residual deviance:  635.92  on 226  degrees of freedom
AIC: 4094

Number of Fisher Scoring iterations: 25
```

The results show that the risk class levels of clients k and years t are statistically significant effects at a 5% level, on the rate of the loss. Whereas, the effects of cob and oci can be dropped from the model. However, even though the model attains a lower AIC value of 4094 compared to glm.M, the model assumptions are not satisfied since based on the residual deviance test with deviance 635.92 with corresponding degrees of freedom 226 and p-value < 0.05.

```
residual.deviance.test(glm.G)
```

```
Residual Deviance Test:
[1] TRUE
p-value:
[1] 8.7e-09
```

Since the model (`glm.G`) assumptions are not met when including the clustering effects of the risk class levels (k, i, j, t) - we conclude that the first reduced gamma model `glm.M.r` is the preferred model.

However, as previously stated since the overall fit of this model is not very good (with residual deviance of 1333.13 and degrees of freedom of 279), we now investigate if including interaction effects improves the model accuracy and performance.

Model `glm.I`: Main and Interaction Fixed Effects of Risk Factors

We investigate the following interaction (fixed) effects based on our findings from the previous sections. Once again, through EDA using the function `cat_plot` from the `interactions` package in `R`, we can select a subset of interaction effects to further investigate. The results were given in interaction plots for GLMs (not shown in this publication, for confidentiality purposes). Note, for comparison and analysis purposes, we also include risk class levels in our interaction plots.

From these results we conclude that there are indications of strong interaction effects between the following risk factors: `t.type` and `peril`, the `cob` and `t.type`. For now we only analyze these two strong interaction effects (to ensure model simplicity and add more variables using a bottom-up approach if we see a significant increase in model performance). As stated, we do not include the effects of any risk class levels, such as `years` in this model (based on our findings from Model 4.3).

This model is fitted using the following equation in `R` followed by the corresponding model results.

```
glm.I <- glm(formula = loss ~ t.type + cob + peril + oci + cpi + rate +  t.type:peril + t.type:cob, family = Gamma(link
    = "log"),
    data = cat_data, na.action = na.omit, offset = log(volume))
```

```
Call:
glm(formula = loss ~ t.type + cob + peril + cpi + rate + t.type:peril +
    t.type:peril + peril:cob + offset(log(volume)), family = Gamma(link = "log"),
    data = cat_data, na.action = na.omit)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-4.162  -2.469  -1.610   0.086   5.328

Coefficients: (1 not defined because of singularities)
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     -2.74483    1.36109   -2.02  0.04472 *
t.typeCXL        2.92901    1.82383    1.61  0.10945
t.typeQS         0.39024    1.00464    0.39  0.69799
t.typeSP         0.30850    1.81891    0.17  0.86545
cobM            -1.49543    1.27203   -1.18  0.24078
cobR            -1.40209    1.14955   -1.22  0.22364
perilF          -0.80528    1.80199   -0.45  0.65532
perilH           3.17113    1.07798    2.94  0.00355 **
cpi             -0.01290    0.00713   -1.81  0.07167 .
Deviance Residuals:
   Min      1Q  Median      3Q     Max
```

```
-3.846  -2.414  -1.481   0.109   4.448

Coefficients: (2 not defined because of singularities)
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      -0.39435    1.62865   -0.24  0.80893
t.typeCXL         6.72496    2.87511    2.34  0.02036 *
t.typeQS         -2.29692    1.49860   -1.53  0.12699
t.typeSP         -2.91153    2.86327   -1.02  0.31050
cobM             -2.43796    1.57381   -1.55  0.12301
cobR             -2.80104    1.32526   -2.11  0.03584 *
perilF           -0.81445    1.92981   -0.42  0.67347
perilH            1.30721    1.31546    0.99  0.32160
oci              -0.00229    0.11167   -0.02  0.98364
cpi              -0.02079    0.00838   -2.48  0.01401 *
rate              1.40277    0.38673    3.63  0.00037 ***
.typeQS:perilF   1.28063    1.45429    0.88   0.3793
t.typeSP:perilF   5.30078    2.50420    2.12   0.0352 *
....
t.typeCXL:cobM   -4.10982    2.27202   -1.81  0.07203 .
t.typeQS:cobM     2.58025    1.43892    1.79  0.07452 .
t.typeQS:cobR     3.22116    1.52575    2.11  0.03605 *
cobM:perilF       2.42991    2.27438    1.07  0.28669
cobR:perilF       2.47829    2.12917    1.16  0.24588
cobM:perilH       2.83010    1.65637    1.71  0.08914 .
....
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for Gamma family taken to be 4.47)

    Null deviance: 1751.7  on 289  degrees of freedom
Residual deviance: 1209.6  on 269  degrees of freedom
AIC: 4261

Number of Fisher Scoring iterations: 25
```

Note, here we reintroduce the main effects of `cob` based on our EDA of possible pairwise effects. Based on the results, the interaction effects of both `t.type` and `cob`, and `t.type` and `peril` are significant at a 5%. Meanwhile, the pairwise interactions of the `cob` and `peril` is not significant on the average `loss` per unit of `volume` (based on clients $k = 1, \ldots, 35$). The model comparison of the (reduced) main effects only model `glm.M.r` and interaction model is shown in Table 4.28.

Model	Resid. Df	Resid. Dev	AIC	BIC
`glm.M.r`	282	1330	4275	4308
`glm.I`	269	1210	4261	4342

Table 4.28. Model comparison of main reduced model `glm.M.r` and interaction effects model `glm.I`, with risk factors only, based on the residual deviance with corresponding degrees of freedom, AIC and BIC.

It is evident that there is a significance decrease in the AIC values and a slight increase in BIC - primarily due to the increase in model complexity. We also note that there is a large decrease in the residual deviance, indicating a better fit in comparison to the main effects model (with residual deviance 836 and residual $df = 197$).

From this analysis, we once again conclude that the pairwise effects of our main (fixed) effects of risk factors provides us with a significantly better fit - to model the rate of loss (per unit of exposure). Therefore, the preferred gamma GLM can be formulated as

Model `glm.I.r` *(reduced gamma model with main and interaction risk effects)*

$$\ln(E[\texttt{loss}_{kijt}]) = \beta_0 + \sum_{q=2}^{4} \beta_1^q \texttt{t.type}_{ki}^q + \beta_2^c \texttt{cob}_{kij}^c + \sum_{d=2}^{3} \beta_3^d \texttt{peril}_{kij}^d \tag{4.47}$$

$$+ \beta_4 \texttt{rate}_{kijt} + \beta_5 \texttt{cpi}_{kijt} + \beta_6 \texttt{t.type}_{ki}^q \times \texttt{cob}_{kij}^c \tag{4.48}$$

$$+ \beta_7 \texttt{t.type}_{ki}^q \times \texttt{peril}_{kij}^d + \ln(\texttt{volume}_{kijt}) + \varepsilon_{kijt},$$

for $i.i.d$ loss observations, such that the response here follows a Based the results, we find that all main effects of the risk model remain significant at the 5% significance level (i.e. should not be reduced any further). While including the interaction effects of the i-th treaty type per k-th client and class of business insured by the i-th treaty and the type of peril of the natural catastrophic resulting in the corresponding loss for the k-th client under i-th treaty, in country j.

4.3.1 Model Diagnostics

Prior to comparing the performance of all model classes, we briefly analyze model diagnostic plots to confirm our gamma model assumptions. Similar to our residual analysis of LMs and LMMs, classical model diagnostic plots for model `glm.I` are shown in Figure 4.12. However, for GLMs, we assess the deviance residual plots. Additionally, we can also look at the corresponding model's Pearson residuals. For example, for the random variable, say Y_k, is formulated as

$$\chi^2(\widehat{\boldsymbol{\mu}}, \mathtt{Y}) = \sum_{K=1}^{n} \frac{(Y_k - \widehat{\mu}_k)^2}{V(\widehat{\mu}_k)} = \sum_{k=1}^{K} \left(\frac{Y_k - \widehat{\mu}_k}{\widehat{\mu}_k}\right)^2,$$

at the k-th client level (for all $k = 1, \ldots, K$). Corresponding Pearson residual plots are also shown in Appendix C Figure 1.

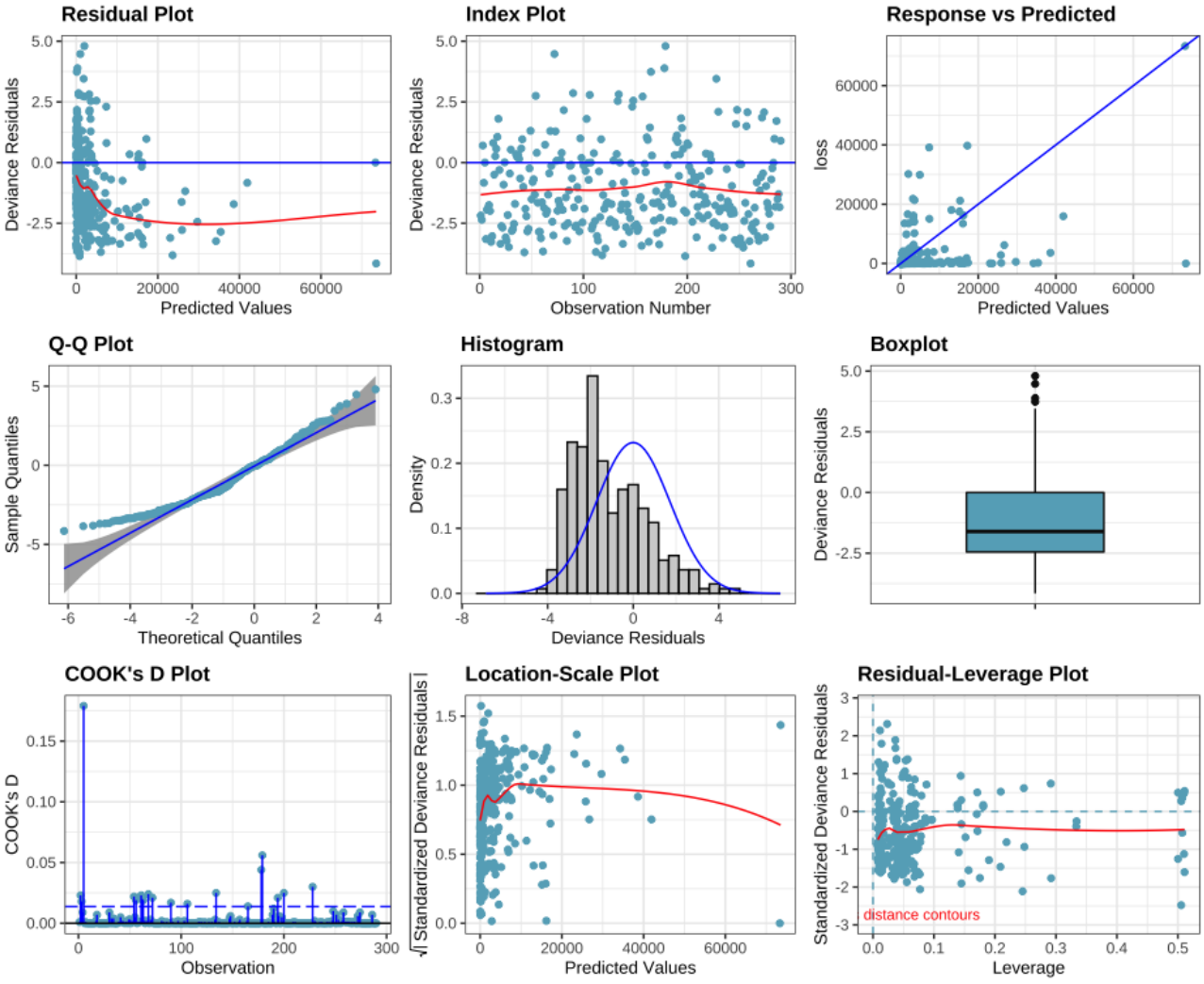


Figure 4.12. Model Diagnostic plots of the selected gamma model `glm.I`, based on deviance (raw) residuals. Note the corresponding plots have been adjusted for the gamma response (using values in the scale of the fitted values, defined by using the option `type` available in the `R` function `resid_panel`, in the library `ggResidPanel`).

Based on the deviance residual plot against the fitted values (Row 1, Column 1), we first observe

random scatters around zero without any particular tendency in the scatter plots. Though, there are generally more residuals lower in magnitude. This plot also shows few instances of very large residuals for lower values (for very low fitted values).

This is also evident when we look at the location-scale residual plot which shows the square root of the absolute value of adjusted deviance residuals (Row 3, Column 2) and the Pearson residual plots (in Appendix C Figure 1). Additionally, the histogram of the (adjusted) deviance residuals also clearly show that the standard normal is not appropriate in this study, as expected (with a highly skewed histogram).

Additionally, the bottom panels provide us additional information regarding the outliers in our data (which we are well aware of from our previous findings). Specifically, based on the Cook's distance plots, we see various instances of influential observations within our data-set. Whereas, though the residual-leverage plots of standardized deviance residuals against leverage values show no high leverage points. From the standardized Pearson residual plots versus leverage values - we observe three high leverage points.

Overall, in comparison to the residual fit of the LMM selected model, our analysis of both model residuals suggest that overall the log linear mixed model (`lmm.RS.C5.r`) provides us with an better residual fit with respect to our natural catastrophic (training) data. Especially since the performance measurements (shown in Table 4.29) also supports our findings. Such that, the LMM clearly outperforms both the GLM and the LM with respect to the AIC, AICc and BIC values (while the gamma model performs the worst, in terms of the information criteria values). Similarly, the results state that that, on average, the LMM fit results in the highest prediction accuracy (with the lowest RMSE at 1.62). Note, all results (up to this point) are based on the training data set. To further confirm these findings, we now proceed with our model comparison and performance analysis of all models with respect to the testing data.

Model Name	**Model**	**AIC**	**AICc**	**BIC**	R^2	**RMSE** (in-sample)
`lm.I.r` (LM)	`ln.lr ~ client + g.treaty + g.country + years+ t.type + oci + rate + cpi + years:oci + years:cpi + t.type:oci + oci:cpi`	1625	1685	1919	0.84	2.98
`lmm.RS.C5.r` (LMM)	`ln.lr ~ years+ t.type + oci + rate + cpi + years:cpi + t.type:oci + (1+oci|client:g.treaty:g.country:years)`	1556	1575	1729	0.95	1.62
`glm.I.r` (GLM)	`ln.lr ~ t.type + peril + cob + cpi + rate + t.type:peril + t.type:cob + offset(log(volume)`	4254	4256	4330	0.85	3.30

Table 4.29. Model performance summary of all selected models - per model class (LM,LMM, and GLM) based on the natural catastrophic data (training or "in-sample" data). Note, here the corresponding R^2 shown for LMMs is based on the conditional $R^2_{con.}$.

5 Out-of-Sample Testing and Performance

In this section, we first assess the performance of the selected fitted models on the test data from each respective model class: LMs, LMMs and GLMs. Then, we can conclude our findings by selecting the overall best model - out of all classes - found in this study based on the underlying natural catastrophic data. Recall that our goal is to find the best suitable model to predict the rate of loss or loss ratio - given the unit of volume exposure (the total sum insured in this study).

5.1 Test Data

Recall, our test data contains incurred losses from $k = 1, \dots, 24$ clients, with 74^k corresponding (ungrouped) treaties - grouped by $i \in \{1^k, 2^k\}$ that provide coverage for $j \in 21^{ki}$ countries over the years 2001-2021 (similar to our training data). Using this testing data set, we assess the fit of the following selected best fitting models within each model class:

(LM) Model `lm.I.r` ***(Reduced log-normal model with fixed main & interaction effects)***

$$\begin{aligned}\texttt{ln.lr}_{kijt} = \beta_0 &+ \sum_{k=1}^{35} \beta_1^k \texttt{client}_k + \beta_2 I_k^{\texttt{g.treaty}_i} + \beta_3 I_{ki}^{\texttt{g.country}_j} + \sum_{t=2}^{18} \beta_4^t \texttt{years}_{kij}^t \\ &+ \sum_{q=2}^{4} \beta_5^q \texttt{t.type}_{ki}^q + \beta_6 \texttt{oci}_{kijt} + \beta_7 \texttt{rate}_{kijt}^d + \beta_8 \texttt{cpi}_{kijt} + \sum_{t=2}^{9} \beta_9^t \texttt{years}_{kij}^t \times \texttt{oci}_{kijt} \\ &+ \sum_{t=2}^{9} \beta_{10}^t \texttt{years}_{kij}^t \times \texttt{cpi}_{kijt} \sum_{q=2}^{4} \beta_9^q \texttt{t.type}_{ki}^q \times \texttt{oci}_{kijt} + \beta_{11} \texttt{oci}_{kijt} \times \texttt{cpi}_{kijt} + \varepsilon_{kijt},\end{aligned} \tag{5.1}$$

given independent losses observed for each client $k = 1, \dots 35$, (grouped) treaty $i \in \{1^k, 2^k\}$, (grouped) countries $j \in \{H^{ki}, L^{ki}\}$, within treaty years $t \in \{2001^{kij}, \dots, 2021^{kij}\}$.

(LMM) Model `lmm.RS.C5.r` *(Mixed log-normal model with crossed random & fixed effects)*

$$\begin{aligned}\texttt{ln.lr}_{kijt} = \beta_0 + \sum_{t=2}^{18}\beta_1^t\texttt{years}_{kij}^t + \sum_{q=2}^{4}\beta_2^q\texttt{t.type}_{ki}^q \quad + \beta_3\texttt{oci}_{kijt} + \beta_4\texttt{rate}_{kijt}^d \\ +\beta_5\texttt{cpi}_{kijt} + \sum_{t=2}^{18}\beta_6^t\texttt{years}_{kij}^t \times \texttt{cpi}_{kijt} \quad + \sum_{q=2}^{4}\beta_7^q\texttt{t.type}_{ki}^q \times \texttt{oci}_{kijt} \\ + \alpha_{0kijt} + \alpha_{1kijt}\texttt{oci}_{kijt} + \varepsilon_{kijt},\end{aligned} \tag{5.2}$$

for clients $k = 1\ldots, K$, with (grouped) treaties $i \in \{1^k, 2^k\}$ for countries in $j \in \{L^{ki}, H^{ki}\}$ over all t years. Recall, here the fixed effects, with regression parameters α_0, α_1 are given in red and the fixed effects parameters $\beta_0, \ldots, \beta_q$ are given in blue. Such that for the random components, we have

$$\begin{aligned}\varepsilon_{kijt} \quad &\sim \mathcal{N}(0, \sigma_\varepsilon^2) \quad i.i.d., \quad \forall\ k, i, j, t, \\ \begin{bmatrix}\alpha_{0kijt} \\ \alpha_{1kijt}\end{bmatrix} \quad &\sim \boldsymbol{\mathcal{N}}\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}\sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2\end{bmatrix}\right) \quad i.i.d., \quad \forall\ k = 1, \ldots, 35.\end{aligned} \tag{5.3}$$

(GLM) Model `glm.I` *(Gamma model with fixed main & interaction effects of risk factors only)*

$$\begin{aligned}\ln(E[\texttt{loss}_{kijt}]) = \ \beta_0 + \sum_{q=2}^{4}\beta_1^q * \texttt{t.type}_{ki}^q + \beta_2^c * \texttt{cob}_{kij}^c + \sum_{d=2}^{3}\beta_3^d * \texttt{peril}_{kij}^d \\ + \beta_4\texttt{rate}_{kijt} + \beta_5\texttt{cpi}_{kijt} + \beta_6\texttt{t.type}_{ki}^q \times \texttt{cob}_{kij}^c \\ + \beta_7\texttt{t.type}_{ki}^q \times \texttt{peril}_{kij}^d + \ln(\texttt{volume}_{kijt}) + \varepsilon_{kijt},\end{aligned} \tag{5.4}$$

for all loss observed for $k = 1, \ldots, 35$ over all risk classes (i, j, t). Note this model does not take into account the varying risks between each risk class level group (k, i, j, t).

In this section, we also look at the *bias* as a metric to assess how close our estimates are on average to the true value of the response (based on Definition 4.13). This is calculated as

$$bias_{kijt} = \frac{1}{n_{kijt}}\sum_{n_{kijt}=1}^{N_{kijtl}} LR_{kijtl} - \hat{LR}_{kijtl}, \tag{5.5}$$

where $n_{kijt} = 1, \ldots, N_{kijtl}$ denotes the independent loss l incurring for $k = 1, \ldots, 24$ per risk class level group (i, j, t). Recall, the negative bias implies that the true value is on average underestimated (while positive values indicate overestimation). Since the positive and negative values will cancel out, we may not get an accurate estimation of the prediction error. For this reason to compare the predictive performances of these models - assessing the spread or variation in the estimates - using both the RMSE (as defined in Definition 4.14) and the mean absolute error (**"MAE"**, see Definition 4.15), given at the most granular risk class level (k, i, j, t). Hence, in our case the MAE is given as

$$MAE_{kijt} = \frac{1}{n_{kijt}}\sum_{n_{kijt}=1}^{N_{kijtl}} |LR_{kijtl} - \hat{LR}_{kijtl}|, \tag{5.6}$$

for $i.i.d$ loss observations l per risk classes levels (k, i, j, t). Recall, for both RMSE and MAE - the smaller value the better the model because they measure errors.

The corresponding results based on these model predictive performance measures are shown in Table 5.1. First, it is evident from the results that the gamma GLM `glm.I` differs the most from the true model

Model Class	Name	Model Formula (as given in R)	BIAS	RMSE	MAE
LM	`lm.I.r`	`ln.lr ~ client + g.treaty + g.country + years+ t.type + oci + rate + cpi + years:oci + years:cpi + t.type:oci + oci:cpi`	1.12	3.59	2.83
LMM	`lmm.RS.C5.r`	`ln.lr ~ years+ t.type + oci + rate + cpi + years:cpi + t.type:oci + (1+oci|client:g.treaty:g.country:years)`	-0.05	1.91	1.28
GLM	`glm.I.r`	`ln.lr ~ t.type + peril + cob + cpi + rate + t.type:peril + t.type:cob + offset(log(volume)`	-4.30	5.12	4.31

Table 5.1. Results of model performance based on the test data - for selected models per model class. Measures include the *bias*, the root square error *RMSE*, and the mean absolute error (*MAE*).

and also provides the worst results. In the sense that compared to the LM and LMM, it attains the highest RMSE and MAE - indicating poor predictive abilities and the most variation in the estimates. It is also evident that, on average, the gamma model `glm.I` clearly underestimates the rate of loss given the `volume` units (with a negative bias of -4.3). The predicted values of our response versus the observed values are shown in Figure 5.1. It is also clear from these results that (for the gamma GLM) there is an obvious lack of fit.

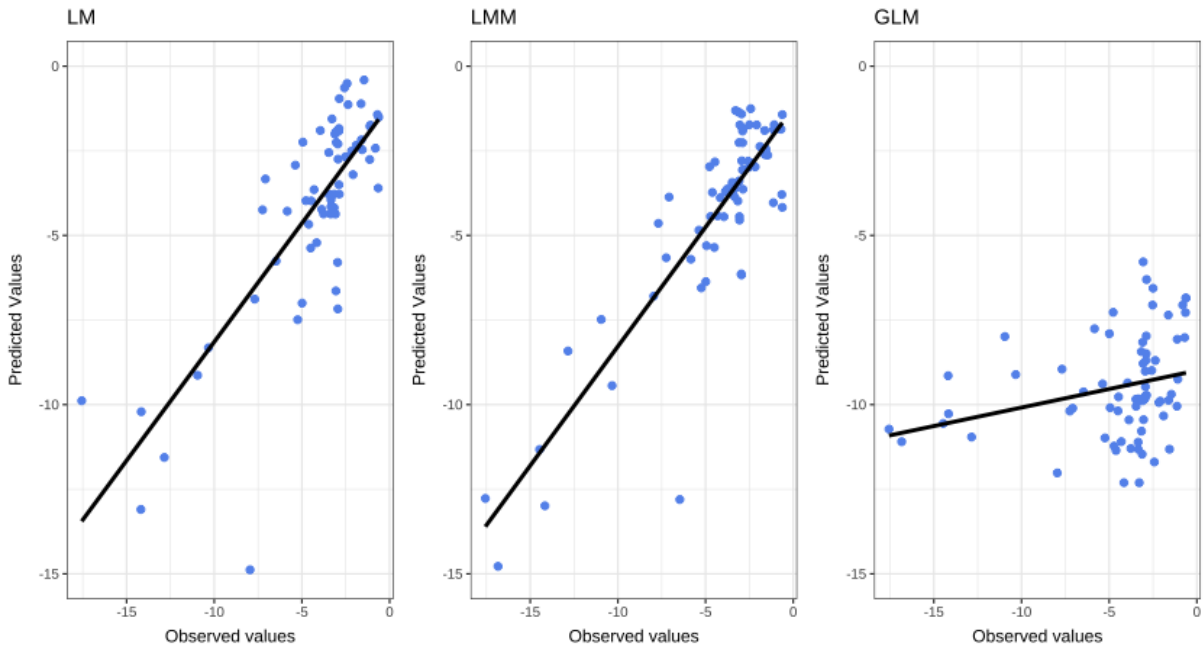


Figure 5.1. Observed values of the response (`ln.lr`) versus predicted values, based on the test data set ("out-sample" data) - per selected model in each model class: (1) LM with (fixed) interaction effects, (2) LMM with crossed random effects (of all risk classes (k, i, j, t) and fixed effects (including pairwise effects), (3) Gamma GLM with only main fixed effects of risk factors (no effects of rick classes).

However, this is somewhat expected and supports our previous findings (when we compared the AIC,

AICc or BIC based on the training data). The large difference in bias and variation (for the GLM) - compared to the selected LM and LMM - may be due to the fact that the model ignores the grouping effects of all risk class levels (k, i, j, t). Specifically, as a result of the fact it does not reflect the heterogeneous risk within any of the groups in the risk class levels (i.e. the losses incurred per client k for treaty i in the country j within treaty year t) - crucial for underlying insurance data and in our study. Since the clustering in groups within this data is ignored - this gamma GLM is not suitable for the longitudinal nature of the data set.

Next, we compare the selected "standard" linear model and the linear mixed model. Clearly, on average, the LMM outperforms the selected LM; in terms of the average prediction error (RMSE = 1.91 and MSE = 1.28), lower bias (bias = -0.05) and the proportion explained by the model in the response based on the test data (with $R^2 = 0.75$). We note here that the in-sample RMSE (1.62) is very close to the out-sample RMSE (1.91). Though the selected "standard" LM accounts for the effects of the risk class levels (k, i, j, t) - each risk class group in this model is analyzed separately. The "classical" log-normal model - attributing to the higher bias = 0.29, compared to the LMM (because here, even small clusters of (k, i, j, t) will have a corresponding regression parameter).

Hence, following our previous findings in Section 4.7, we once again conclude that the LMM (`lmm.RS.C5.r`) with crossed random effects (i.e. with random intercepts with interactions effects of all the risk class levels, and random `oci` slope effects) is our final and optimal performing model - with the lowest variance and bias attained compared to all models considered in this study - given the natural catastrophic data. In other words, it achieves the highest predictive accuracy for the rate of loss per exposure volume unit.

This mixed or multi-level model is not only a compromise between the two extremes (of completing ignoring the effects of the risk class levels, such as the GLM, or analyzing each class separately, such as the LM) - but it also enables us to account for the heterogeneity risks - crucial in this study. Specifically, since the model allows for the varying risks and effects at every risk class level: k, i, j and t, it is suitable for the underlying longitudinal data with repeated measurements per k-th client.

5.2 Final Model Interpretation

Once again, we provide the full model summary of our selected LMM model, in Table 5.2 from Section 4.7 (for easy readability purposes).

Model `lmm.RS.C5.r` *(Crossed Random Effects, reduced model)*			
Fixed Effects			
Predictors	*Estimates*	*CI*	*p*
(Intercept)	26.38	3.50 – 49.25	**0.024**
`years [2003]`	53.38	-39.41 – 146.18	0.258
`years [2004]`	-23.01	-47.18 – 1.17	**0.062**
`years [2005]`	-37.59	-67.43 – -7.75	**0.014**
...	...	...	...
`t type [CXL]`	2.95	0.08 – 5.83	**0.044**
`t type [QS]`	-0.83	-2.37 – 0.72	0.293
`t type [SP]`	1.07	-1.87 – 4.01	0.473
`oci`	-6.25	-7.45 – -5.05	**<0.001**
`rate`	-0.35	-0.55 – -0.14	**0.001**
`cpi`	-0.29	-0.56 – -0.01	**0.042**
`years [2003]:cpi`	-0.64	-1.77 – 0.49	**0.264**
`years [2004]:cpi`	0.31	0.02 – 0.61	**0.038**
`years [2005]:cpi`	0.46	0.10 – 0.81	**0.012**
...	...	...	...
`t.type [CXL]:oci`	-2.35	-4.37 – -0.33	**0.023**
`t.type [QS]:oci`	0.07	-0.71 – 0.86	0.854
`t.type [SP]:oci`	-1.48	-3.75 – 0.79	**0.200**
Random Effects			
$\hat{\sigma}^2_{\varepsilon}$			4.54
$\hat{\sigma}_0$ (Intercept, `client:g.treaty:g.country:years`)			18.59
$\hat{\sigma}_1$ (Slope, `client:g.treaty:g.country:years:oci`)			13.85
$\hat{\rho}_{01}$ (Correlation, of `client:g.treaty:g.country:years`)			-0.98
No. of risk class levels `client` k			35
No. of risk class levels `g.treaty` i			2
No. of risk class levels `g.country` j			2
No. of risk class levels `years` t			19
Total Observations			290
$ICC_{adj.}$ / Conditional $ICC_{con.}$			0.84 / 0.26
Marginal $R^2_{mar.}$ / Conditional $R^2_{con.}$			0.690 / 0.951

Table 5.2. Model summary and estimated regression parameters selected of the final and best performing model, `lmm.RS.C5.r` (LMM), across all model classes - to estimate the average (log) loss ratio, $\texttt{ln.lr}_{kijt}$.

The following recaps our key findings based on the selected model results:

- The average rate of loss per unit of exposure volume significantly varies per individual risk. Specifically, with respect to the interaction or grouping effects of the risk class groups (k, i, j, t) (i.e. the random intercepts, with estimated variance $\hat{\sigma}^2_{\alpha} = 18.59$).
- In this model, the regression coefficient for the number of unique natural catastrophic events (`oci`) varies across the group of risk class levels (k, i, j, t). Here we assumed the varying regression slopes for `oci` followed a normal distribution (with a estimated variance of $\hat{\sigma}^2_{\alpha} = 13.85$).
- In our model, we found the (fixed) main effects of the i-th treaty type per k-th client (for countries j within the t-th year) to be significant (at a 5% significance level) on the average loss ratio. Such that, for instance, the rate of loss ratio on average, per exposure volume unit, is

exponentially higher for a i-th treaty with type `CXL` in comparison to all other types of treaties (i.e. increases the average loss ratio by a factor of $e^{2.95} = 19.10$). This is somewhat expected as these types of treaties are designed only to cover excess losses (over a defined loss limit) arising from catastrophic events, significantly increasing the expected rate of loss per k-th client for that treaty i.

- The fixed main effects of premium rates (`rate`) per k-th client (with groups in (i, j, t)) was also significant on the average loss ratio, where for every one percent increase in the premium rates decreases the average loss ratio by 29.5%. This intuitively makes sense because higher premium rates indicate a higher amount of total coverage or sum insured (exposure volume units). Similarly, we found that for every one-point increase in the change of the consumer price index (`cpi`) - for the j-th country for time t per k-th client - on average decreases the estimated loss ratio.
- Note, the multi-level model includes the marginal effects of the t-th treaty year both as random and fixed effects due to its cross-level interaction effects with the (fixed) effects of the `cpi`. Specifically, the results show how the t-th treaty year (per (k, i, t) facilitates the effects of the `cpi` on the average loss ratio.
- The cross-level pairwise effects of the type of i-th treaty and the number of natural catastrophes were also found to be significant (p-value < 100). The model indicates that the effects of `oci` on the average loss ratio are moderated by the type of i-th treaty. For example, the effects of the number of natural catastrophic events on the average loss ratio - per risk class groups (k, i, j, t) - is significantly higher for a client k with quota-shares treaties (in comparison to all other treaties).

6 Conclusion

Given the nature of natural catastrophic data -with low frequency and high severity - assessing an individual risk in property underwriting can be extremely challenging. In actuarial statistics, a wide variety of models are considered to measure the effects of risk factors - considering a set of corresponding "tariffs" for different risk class groups - on the average rate of loss (per exposure volume). This may include considering predictive models from several model classes, such as LMs, LMMs, GLMs, and GLMMs. The approaches to model the expected average loss ratio may also vastly differ. Our modelling approach in this study was to start with simple and standard familiar models (LMs) and then gradually extend or add complexity demanded by the situations (LMMs and GLMs).

We first constructed a framework to directly model the average loss ratio (instead of the frequency-severity approach) at every hierarchical risk class level (k, i, j, t). The grouping effects of all risk class levels and risk factors were analyzed based on the natural catastrophic data (the training data or "in-sample" data set) - for 24 countries or islands located in the Caribbeans from 2001-2021 (based on the policy and treaty insurance data). The wide range of t years considered also meant that our data set contained a large set of individual risk class levels per k-th client. For this reason, before modelling, we also performed exploratory data analyses to screen the varying effects per risk class level and aggregate our data based on suitable risk groups.

Then based on the aggregated data, our goal was to extensively study the different possible models and find the best suitable or fitting model for the underlying data. This included investigating various model specifications per model class and discussing the appropriate measures to assess all the models. As previously stated, due to the low frequency and high severity loss events, the distributions of the loss ratios exhibited heavy skewness and long tails (as typically expected in the insurance context). This meant that the large losses are more likely to occur than insinuated by the normal distribution (Frees et al., 2014). For this reason, we focused on the two heavy-tailed regression techniques: the log-normal and gamma models.

In addition, given our data insurance loss data (from the reinsurance perspective) had a longitudinal data structure, we also needed to account for the effects of the risk factors on our response - while still being able to account for the varying dynamics in the risk class groups appropriately. For this reason, we first assessed the ordinary LM with and without the grouping effects of risk class levels (k, i, j, t) individually (as covariates). Our LM results showed that ignoring the clustering effects of (k, i, j, t) - provides a poor fit to the data. Meanwhile, accounting for specific pairwise effects improves our overall model fit - in addition to the risk classes. For this reason, next, we extended our findings from the selected LMs to the LMMs. Extending the LM class to LMMs allowed us to compare our findings (from LMs models) on the effect of loss ratios - when we account for the mixed effects of the client-related risk attributes.

Based on our findings from the LMs, we compared a variety of different model specifications. Here, we found the LMM with fixed effects (including main and interaction effects) and random crossed effects specification provided the best LMM fit. Allowing for random varying risk class intercept effects and random slope effects of the number of natural catastrophic events (historical) - provided the best suitable model fit. Following the same model-building process, lastly, we looked at GLMs. We fitted the gamma regression model with a log-link function (instead of the canonical inverse link function) so that the model regression parameters were comparable (to the LMs and LMMs) and, hence, more interpretable. We compared and assessed each model's performance during each model-building process, based on the "in-sample" dataset, concerning the AIC, AICc, BIC and the appropriate estimates for the coefficient of determination (R^2). We used the RMSE for both the training and testing data to analyze the predictive performance across all models and looked at the MSE and bias concerning the testing data ("out-sample" data).

In conclusion, we found that the LMM outperformed the standard LM and the GLM. The excellent performance of this hierarchical linear mixed-effects model was confirmed by our model evaluation results - based on the test data. As the respective LMM yielded the lowest bias and variance, indicating that it is the most suitable model given the underlying data. By correcting and specifying an appropriate variance structure corresponding to the hierarchical data structure, we were able to include the clustering effects of the risk classes based on the k-th client into our model (crucial for insurance rate making). This mixed-effects model enabled us to account for nested effects and allowed for flexibility per risk class group, which allowed each group to have its unique regression relationship. For this reason, we could capture the heterogeneity per k-th client treaty i for all countries j and years t considered in this study. However, it will be interesting also to consider additional and more complex multi-level random effects structures for future studies - with a larger set of loss data.

In contrast, our results (based on the testing data) showed that the standard LM, on average, overestimated the actual values of the loss ratio. This is may be due to the model's high degree of complexity, where each risk class group effect was also analyzed separately (entering the model as covariates). At the same time, the gamma model underestimated the actual values on average. As this model did not account for any risk class level (due to the convergence issues) - underfitting, in this case, may have resulted from an excessively simple model specification. It is essential to note that this study was specifically based on the reinsurance data available for only countries located in the Caribbean. For this reason, the risk factors considered in this study and loss data were limited to the study design. Including additional risk factors or loss observations for the risk class groups with different underlying assumptions and model structures may lead to different empirical results and conclusions. However, in our case, the underestimation may be primarily because the GLM does not account for the heterogeneity risk between the risk classes (per client).

For this reason, future studies should focus on one of GLMs' popular extensions applicable to this study - known as the *Generalized Linear Mixed Models (GLMMs)*. These models are especially useful for *longitudinal data* (with repeated measurements over time on a group of individuals) - while allowing for more complex multi-level structures. Extending the GLM will allow us to include random effects in the linear predictor - which also determines the correlation structure between the loss observations for a client over the treaty years. In other words, it will allow us to account for unobserved heterogeneity risk characteristics among our clients. Another key advantage of GLMMs is that it ensures our regression models are not restricted to normal data - by considering other distributions from the exponential family. However, since the modelling building processes of GLMMs are still in the statistical frontier - available statistical software is often limited, unstable or too complex for practical solutions. Alternatively, based on the recommendation of many previous studies (for example, see Bermúdez et al. (2014)),

using copulas to model the dependencies over time or for treaty dependencies between the lines of businesses (based on the insurance coverage in property insurance) may be interesting. Sun et al. (2008) demonstrates the advantages of copulas to accommodate longitudinal data. They show how copulas may be especially useful for modelling dependencies over time and for representing the marginal distributions, where extreme values are highly likely to occur in longitudinal data. However, this may be challenging given the low occurrence of high severity loss events. For this reason, it might also be worth looking at other heavy-tailed distributions (such as Tweedie or Pareto).

The scope of this study and the methods employed here primarily focuses on the statistical modelling with the available insurance-related measurements. This provides practical and valuable solutions in insurance pricing. The findings from this study not only help researchers to assess different predictive models in the (re) insurance context - but also provide property underwriters and reinsurers with practical and ready-to-use solutions. This allows them to precisely and directly estimate the average loss ratio for insurance pricing with heterogeneity risks - in the natural catastrophic loss data.

Appendix

A Model Results: LMMs

A.1 Model lm.I: Interaction Effects (With Group Effects of Risk Class)

```
Call:
lm(formula = ln.lr ~ client + g.treaty + g.country + years + t.type + oci + rate + cpi + g.country:oci + years:oci +
    years:cpi + t.type:oci + oci:cpi, data = cat_data, na.action = na.omit)

Residuals:
   Min     1Q Median     3Q    Max
   -13     -2      0      2      9

Coefficients: (7 not defined because of singularities)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      32.33      18.65     1.7    0.085 .
client6          -3.31       2.20    -1.5    0.134
client7           0.46       2.55     0.2    0.858
client8          -7.67       2.77    -2.8    0.006 **
....
g.treaty2         0.85       4.37     0.2    0.847
g.countryL        0.01       1.54     0.0    0.992
years2003       134.00      95.76     1.4    0.163
years2004       -30.74      19.39    -1.6    0.115
years2005       -49.69      20.55    -2.4    0.017 *
t.typeCXL         4.54       2.60     1.7    0.083 .
t.typeQS         -3.09       1.86    -1.7    0.099 .
t.typeSP          1.32       2.74     0.5    0.631
oci               3.89       4.91     0.8    0.430
rate             -0.48       0.21    -2.3    0.021 *
cpi              -0.32       0.24    -1.3    0.183
g.countryL:oci    0.58       0.54     1.1    0.288
years2003:oci     4.70       7.03     0.7    0.505
years2004:oci    -2.23       3.20    -0.7    0.486
years2005:oci    -1.81       3.44    -0.5    0.599
....
years2003:cpi    -1.69       1.24    -1.4    0.174
years2004:cpi     0.41       0.25     1.6    0.108
years2005:cpi     0.63       0.26     2.5    0.015 *
....
t.typeCXL:oci    -3.72       1.46    -2.5    0.012 *
t.typeQS:oci     -0.01       0.69     0.0    0.989
t.typeSP:oci     -3.44       1.65    -2.1    0.038 *
oci:cpi          -0.09       0.03    -2.7    0.008 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 3.5 on 195 degrees of freedom
Multiple R-squared:  0.89,  Adjusted R-squared:  0.84
F-statistic:  18 on 94 and 195 DF,  p-value: <2e-16
```

A.2 Model `lmm.RS.C5`: ANOVA table for Fixed and Interaction Effects (With Crossed Effects of Risk Class) - using the Wald Chi-sq Tests (Type III)

Fixed Covariate Effects (Main and Interaction Effects)	χ^2	Df	Pr(>Chisq)
`years`	76.45	18.00	**3.6e-09**
`t.type`	4.72	3.00	**0.02**
`oci`	201.31	1.00	**<2e-16**
`rate`	11.41	1.00	**7.3e-04**
`cpi`	16.94	1.00	**3.9e-05**
`years:oci`	20.23	15.00	0.16
`years:cpi`	25.69	15.00	**0.04**
`t.type:oci`	8.03	3.00	**0.05**
`oci:cpi`	0.27	1.00	0.60

Table 1. Analysis of Deviance Table (Type III Wald chi-square tests).

B Model Results: GLMs

B.1 Model Diagnostic Plots (Pearson): for Gamma Interaction Model `glm.I`

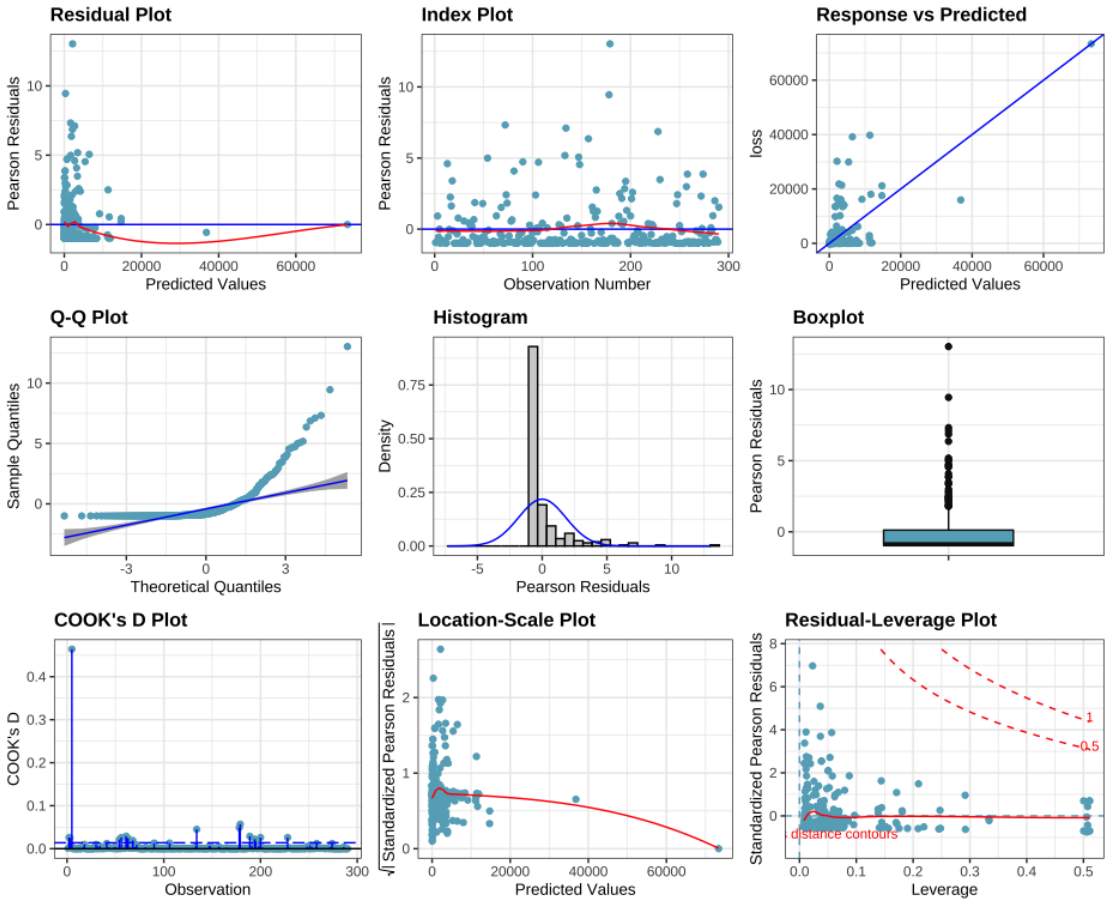


Figure 1. Model Diagnostic plots of model `glm.I.r`, based on pearson residuals.

Bibliography

K. Antonio and J. Beirlant. Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and Economics*, 40(1):58–76, 2007.

K. Antonio and Y. Zhang. Linear mixed models for predictive modelling in actuarial science. *Chapter*, 8:266–312, 2013.

D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*, 2014.

D. J. Bauer and P. J. Curran. Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate behavioral research*, 40(3):373–400, 2005.

P. M. Bentler and D. G. Bonett. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, 88(3):588, 1980.

L. Bermúdez, A. Ferri, and M. Guillén. On the use of risk measures in solvency capital estimation. *International Journal of Business Continuity and Risk Management 21*, 5(1):4–13, 2014.

B. Blakley, E. McDermott, and D. Geer. Information security is information risk management. In *Proceedings of the 2001 workshop on New security paradigms*, pages 97–104, 2001.

J. M. Borowicz and J. P. Norman. The effects of parameter uncertainty in the extreme event frequency-severity model. In *28th International Congress of Actuaries, Paris*, volume 28. Citeseer, 2006.

B. M. Byrne. *Structural equation modeling with EQS: Basic concepts, applications, and programming*. Routledge, 2013.

F. Cremer. En_proceedings2020_20210109_finale version. 2020. URL https://www.th-koeln.de/mam/downloads/deutsch/hochschule/fakultaeten/wirtschafts_und_rechtswissenschaften/en_proceedings_2020.pdf.

H. Dahen and G. Dionne. Scaling models for the severity and frequency of external operational loss data. *Journal of Banking Finance*, 34(7):1484–1496, 2010. ISSN 0378-4266. doi: https://doi.org/10.1016/j.jbankfin.2009.08.017. URL https://www.sciencedirect.com/science/article/pii/S0378426609002210. Performance Measurement in the Financial Services Sector.

P. K. Dunn and G. K. Smyth. Generalized linear models: Estimation. In *Generalized Linear Models With Examples in R*, pages 243–263. Springer, 2018.

C. Eager and J. Roy. Mixed effects models are sometimes terrible. *arXiv preprint arXiv:1701.04858*, 2017.

L. J. Edwards, K. E. Muller, R. D. Wolfinger, B. F. Qaqish, and O. Schabenberger. An r2 statistic for fixed effects in the linear mixed model. *Statistics in medicine*, 27(29):6137–6157, 2008.

M. Eling, H. Schmeiser, and J. T. Schmit. The solvency ii process: Overview and critical analysis. *Risk management and insurance review*, 10(1):69–85, 2007.

L. Fahrmeir, T. Kneib, and S. Lang. Regressionsmodelle. *Regression: Modelle, Methoden und Anwendungen*, pages 19–58, 2007.

R. K. Fowler. Loss frequency. *Ins. LJ*, page 135, 1960.

J. Fox. *Applied regression analysis and generalized linear models*. Sage Publications, 2015.

J. Fox and S. Weisberg. *An R companion to applied regression*. Sage publications, 2018.

E. Frees. Loss data analytics. *arXiv preprint arXiv:1808.06718*, 2018.

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

E. W. Frees, R. A. Derrig, and G. Meyers. *Predictive modeling applications in actuarial science*, volume 1. Cambridge University Press, 2014.

A. Gałecki and T. Burzykowski. Linear mixed-effects model. In *Linear mixed-effects models using R*, pages 245–273. Springer, 2013.

A. T. Galecki. General class of covariance structures for two or more repeated factors in longitudinal data analysis. *Communications in Statistics-Theory and Methods*, 23(11):3105–3119, 1994.

J. Gareth, W. Daniela, H. Trevor, and T. Robert. *An introduction to statistical learning: with applications in R*. Spinger, 2013.

E. Gómez Déniz and E. Calderín Ojeda. The compound dgl/erlang distribution in the collective risk model. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 16:121–142, 2013.

C. Gourieroux, A. Holly, and A. Monfort. Likelihood ratio test, wald test, and kuhn-tucker test in linear models with inequality constraints on the regression parameters. *Econometrica: journal of the Econometric Society*, pages 63–80, 1982.

T. G. Grégoire, O. Schabenberger, and J. P. Barrett. Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. *Canadian Journal of Forest Research*, 25(1): 137–156, 1995.

B. Hewitt. *The Breadth and Scope of the Global Reinsurance Market and the Critical Role Such Market Plays in Supporting Insurance in the United States*. 2014.

B. Hewitt. fio -reinsurance reportpdf. 2019. URL https://www.treasury.gov/initiatives/fio/reports-and-notices/documents/fio%20-reinsurance%20report.pdf.

A. Klein and H. Moosbrugger. Maximum likelihood estimation of latent interaction effects with the lms method. *Psychometrika*, 65(4):457–474, 2000.

A. Kuznetsova, P. B. Brockhoff, and R. H. Christensen. lmertest package: tests in linear mixed effects models. *Journal of statistical software*, 82:1–26, 2017.

Y. Lee, J. A. Nelder, and Y. Pawitan. *Generalized linear models with random effects: unified analysis via H-likelihood*, volume 153. CRC Press, 2018.

P. McCullagh and J. Nelder. Generalized linear models ii. 1989.

P. McCullagh and J. A. Nelder. 1989. *Generalized linear models*, 37, 1983.

S. A. Murphy and A. W. Van der Vaart. On profile likelihood. *Journal of the American Statistical Association*, 95(450):449–465, 2000.

S. Nakagawa, P. C. Johnson, and H. Schielzeth. The coefficient of determination r 2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, 14(134):20170213, 2017.

J. A. Nelder and R. J. Verrall. Credibility theory and generalized linear models. *ASTIN Bulletin: The Journal of the IAA*, 27(1):71–82, 1997.

S. Ng, D. Lestari, and S. Devila. Generalized linear model for deductible pricing in non-life insurance. *AIP Conference Proceedings*, 2168(1):020038, 2019. doi: 10.1063/1.5132465. URL https://aip.scitation.org/doi/abs/10.1063/1.5132465.

E. Ohlsson and B. Johansson. *Non-life insurance pricing with generalized linear models*, volume 174. Springer, 2010.

A. A. Pantelous and E. Passalidou. Optimal premium pricing policy in a competitive insurance market environment. *Annals of Actuarial Science*, 7(2):175–191, 2013.

D. Pfeifer and V. Langen. Insurance business and sustainable development. *arXiv preprint arXiv:2102.02612*, 2021.

J. D. Pollner. *Managing catastrophic disaster risks using alternative risk financing and pooled insurance structures*, volume 495. World Bank Publications, 2001.

R. J. Roth Sr and H. Kunreuther. *Paying the price: The status and role of insurance against natural disasters in the United States*. Joseph Henry Press, 1998.

D. Ruppert, M. P. Wand, and R. J. Carroll. *Semiparametric regression*. Number 12. Cambridge university press, 2003.

Scherer et. al. Insurance mathematics. *Insurance: Mathematics and Economics*, 22(3):305–306, 1998. ISSN 01676687. doi: 10.1016/S0167-6687(98)80115-7.

S. R. Searle. *Linear models for unbalanced data*, volume 639. John Wiley & Sons, 2006.

J. Sun, E. W. Frees, and M. A. Rosenberg. Heavy-tailed longitudinal data modeling using copulas. *Insurance: Mathematics and Economics*, 42(2):817–830, 2008.

P. Thagard, M. R. Forster, J. Woods, and P. S. Bandyopadhyay. Philosophy of statistics. 2011.

M. I. M. Torre-Enciso and R. H. Barros. Operational risk management for insurers. *International Business Research*, 6(1):1, 2013.

J. Valeckỳ et al. Calculation of solvency capital requirements for non-life underwriting risk using generalized linear models. *Prague Economic Papers*, 26(4):450–466, 2017.

G. Verbeke and G. Molenberghs. Estimation of the marginal model. *Linear mixed models for longitudinal data*, pages 41–54, 2000.

W. K. Viscusi et al. *Reforming products liability*. Harvard University Press, 1991.

M. V. Wuthrich. Non-life insurance: mathematics & statistics. *Available at SSRN 2319328*, 2020.

Y.-t. Xie, Z.-x. Li, and R. A. Parsa. Extension and application of credibility models in predicting claim frequency. *Mathematical Problems in Engineering*, 2018, 2018.

G. Zanjani, M. Suher, and J. D. Cummins. 4. federal financial exposure to natural catastrophe risk. In *Measuring and managing federal financial risk*, pages 61–96. University of Chicago Press, 2010.

T. Zapart. The role and importance of an actuary in the insurance. 2013.

R. Zhu and G. H. Zou. Blup estimation of linear mixed-effects models with measurement errors and its applications to the estimation of small areas. *Acta Mathematica Sinica, English Series*, 30(12): 2027–2044, 2014.

A. F. Zuur, E. N. Ieno, N. J. Walker, A. A. Saveliev, G. M. Smith, et al. *Mixed effects models and extensions in ecology with R*, volume 574. Springer, 2009.