MASTER

ACTUARIAL SCIENCE

MASTER'S FINAL WORK

INTERNSHIP REPORT

ASSESSING THE IMPACT OF EXCESS OF LOSS REINSURANCE ON A WORKERS' COMPENSATION PORTFOLIO

JÚLIA DOS SANTOS NASCIMENTO

JUNE 2025

MASTER

ACTUARIAL SCIENCE

MASTER'S FINAL WORK

INTERNSHIP REPORT

ASSESSING THE IMPACT OF EXCESS OF LOSS REINSURANCE ON A WORKERS' COMPENSATION PORTFOLIO

JÚLIA DOS SANTOS NASCIMENTO

SUPERVISION:
ALEXANDRA BUGALHO DE MOURA
SÓNIA GOMES DE ARAÚJO

JUNE 2025

Acknowledgements

My sincerest gratitude to everyone that supported me in this journey. I could not have done this without you.

First, I truly appreciate the support and mentorship of my supervisors, Professor Alexandra Bugalho de Moura and Sónia Gomes de Araújo, whose advice and encouragement made this project possible. I would also like to thank my colleagues at Lusitania for their constant support throughout this project. They were always ready to listen, offer assistance and share their expertise, which made a world of difference in the development of this MFW.

I am deeply thankful to my friends who supported me from afar, your patience and constant encouragement meant more than words can express. I am equally grateful to the new friends I made along this journey, who made a new country feel like home. Thank you for sharing the ups and downs and for offering comfort, laughter and companionship when it was needed most.

Lastly, I would like to thank my dad, my sister and my grandma for their unwavering support throughout this adventure, even from a distance. Your belief in me has been a great source of strength and I cannot wait to hug you all again. A special and loving thank you to my late mother, whose encouragement and faith in my potential made me believe that I was capable of anything and everything, and I could not be prouder to be her daughter.

ABSTRACT

In Portugal, workers' compensation (WC) insurance is mandatory and must be secured through private insurance companies. The law states that employers must provide insurance coverage to all employees, irrespective of the nature of their employment contracts, to ensure protection against work-related injuries and occupational diseases.

Reinsurance is an essential tool to help insurers manage capital and mitigate exposure to large or unpredictable losses, particularly in long-tail lines such as WC. Given the growing complexity and cost of covering long-duration liabilities like disability pensions, the need for efficient and well-structured reinsurance arrangements has become increasingly important.

This study evaluates the effectiveness of a non-proportional excess of loss (XL) reinsurance treaty in mitigating risk and improving efficiency for a WC insurance portfolio. The analysis is motivated by the high cost of long-term annuity payments, which impose sustained financial obligations on the insurer and must be protected by adequate mathematical reserves. We explore whether the current treaty structure appropriately balances retained risk and ceded premium.

Using actuarial modelling techniques, including Generalized Linear Models (GLMs) to analyse claim frequency and severity, and the Generalized Pareto Distribution (GPD) to capture tail behaviour, we assess the treaty's risk transfer performance. We apply Loss Elimination Ratios (LERs) and capital-based metrics to compare various attachment points and layers of reinsurance under real-world constraints.

The results indicate that the current treaty structure provides minimal reduction in retained losses, largely due to high retention thresholds and steep ceded premium loadings. While lowering the attachment point offers limited benefit in terms of reducing expected losses, given the high cost imposed by reinsurers, it would contribute more significantly to reducing retained volatility. This suggests that although the current structure is not particularly efficient in terms of risk transfer, adjustments could enhance protection against large claim variability.

KEYWORDS: Workers' compensation; Excess of Loss; Reinsurance; Pensions; Generalized Linear Models; Generalized Pareto Distribution.

RESUMO

Em Portugal, o seguro de acidentes de trabalho (AT) é obrigatório e deve ser segurado através de companhias de seguros privadas. As entidades patronais são legalmente obrigadas a assegurar a cobertura de todos os trabalhadores, independentemente do tipo de contrato, para garantir a proteção contra acidentes de trabalho e doenças ocupacionais.

O resseguro desempenha um papel fundamental para ajudar as seguradoras a gerir o capital e a atenuar a exposição a perdas grandes ou imprevisíveis, em especial nos ramos de cauda longa, como AT. Dada a crescente complexidade e o custo da cobertura de responsabilidades de longa duração, como as pensões de invalidez, a necessidade de acordos de resseguro eficientes e bem estruturados tornou-se cada vez mais importante.

Este estudo avalia a eficácia de um tratado de resseguro não proporcional de excesso de perdas (XL) na atenuação do risco e na melhoria da eficiência de uma carteira de seguros de acidentes de trabalho. A análise é motivada pelo elevado custo dos pagamentos de anuidades a longo prazo, que impõem obrigações financeiras sustentadas à seguradora e devem ser protegidas por reservas matemáticas adequadas. Exploramos se a atual estrutura do contrato equilibra adequadamente o risco retido e o prémio cedido.

Utilizando técnicas de modelação atuarial, incluindo Modelos Lineares Generalizados (MLG) para analisar a frequência e a severidade dos sinistros, e a Distribuição de Pareto Generalizada (DPG) para captar o comportamento da cauda, nós avaliamos o desempenho da transferência de risco do tratado. São aplicados rácios de eliminação de perdas e métricas baseadas no capital para comparar vários limites de retenção e camadas de resseguro em condições realísticas.

Os resultados indicam que a atual estrutura do contrato permite uma redução mínima das perdas retidas, em grande parte devido aos elevados limiares de retenção e aos elevados níveis de prémios cedidos. Embora a redução do limite de retenção ofereça um benefício limitado em termos de redução das perdas esperadas, dado o elevado custo imposto pelos resseguradores, contribuiria de forma mais significativa para reduzir a volatilidade retida. Isto sugere que, embora a estrutura atual não seja particularmente eficiente em termos de transferência de riscos, ajustamentos poderiam aumentar a proteção contra a variabilidade dos sinistros de grande dimensão.

PALAVRAS-CHAVE: Acidentes de Trabalho; Excesso de Perdas; Resseguro; Pensões, Modelos Lineares Generalizados; Distribuição de Pareto Generalizada.

Table of Contents

List of Figures

List of Tables

Abbreviations

AIC Akaike Information Criterion

ASF Autoridade de Supervisão de Seguros e Fundos de Pensões (Portuguese Insurance and Pension Funds Supervisory Authority)

AT Acidentes de Trabalho (Workers' Compensation)

BIC Bayesian Information Criterion

CAE Classificação Portuguesa das Actividades Económicas (Portuguese Classification of Economic Activities

CAS Casualty Actuarial Society

CATXL Catastrophe Excess-of-loss

CDF Cumulative Distribution Function

COLA Cost Of Living Adjustments

DF Degrees of Freedom

EDM Exponential Dispersion Model family

FAT Fundo de Acidentes de Trabalho (Workers' Compensation Fund)

FMM Finite Mixture Models

GOF Goodness-of-fit

GLM Generalized Linear Models

GPD Generalized Pareto Distribution

GVIF Generalized Variance Inflation Factor

IBNR Incurred But Not Reported

IPAH Incapacidade Permanente Absoluta para Trabalho Habitual (Permanent Absolute Incapacity for the Usual Occupation)

IPAT Incapacidade Permanente Absoluta para Todo Trabalho (Permanent Absolute Incapacity for Any and All Work)

IPP Incapacidade Permanente Parcial (Partial Permanent Incapacity)

LER Loss Elimination Ratio

LOB Line of Business

LRT Likelihood Ratio Test

LTA Lifetime Assistance

MCAR Missing Completely at Random

MCMC Markov chain Monte Carlo

MFW Master's Final Work

NOR Non-obligatorily redeemable

NP Non-proportional

NSLT Non-Similar to Life Techniques

OB Other Benefits

OR Obligatorily Redeemable

PC Property and Casualty

QS Quota-share

RA Retirement Age

SLT Similar to Life Techniques

SXL Surplus Excess of Loss

TD 88/90 Mortality Table TD 88/90

VIF Variance Inflation Factor

WC Workers' Compensation

XL Excess of Loss

1 INTRODUCTION

Workplace injuries and occupational diseases can be costly for workers and employers and, without workers' compensation (WC) insurance, employees who suffer injuries may face significant healthcare bills and loss of income, and they may struggle to return to work (Guyton, 1999). On the other hand, employers could be liable for the costs associated with these injuries, including legal fees and settlements in the event of a lawsuit. WC offers coverage for employees who sustain injuries or illnesses from occupational causes. This insurance protects employers from liability, while ensuring that workers receive adequate medical care and financial support for their work-related injuries or illnesses (Lewis, 2012). Thus, insurers play a key role in mitigating social risk.

In Portugal, private WC coverage is compulsory (according to Law $n°$ 98, 2009) and insurers must employ safeguards to guarantee that they are able to meet their obligations towards their clients. They often do so by sharing the risk with reinsurance companies.

During my internship with Lusitania, I had the opportunity to observe the reserving practices for WC and to conduct a careful assessment of the excess of loss (XL) reinsurance treaty that provides coverage for this line of business (LOB). The primary objective was to gain a deeper understanding of the treaty's structure, assess its performance in terms of risk mitigation and cost-efficiency, and explore potential areas for optimization. This involved examining the interaction between the reinsurance treaty and the underlying claim distribution, as well as evaluating how well the current reinsurance arrangement aligns with the risk profile of the WC portfolio. To do so, we applied statistical modelling techniques and actuarial principles, such as Generalized Linear Models (GLMs), extreme value theory (EVT) and Loss Elimination Ratios (LER), to quantify the effectiveness of the reinsurance layer in reducing retained losses.

However, some practical limitations shaped the scope and outcomes of the analysis. First, due to restricted access to the full WC portfolio, we used a subset of simple claims as a proxy to estimate the occurrence of long-term liabilities ("severe claims"). The lognormal distribution provided a reasonable fit within the GLM framework, but it did not fully capture the heavy tail of the claim distribution. To address this, we incorporated a generalized Pareto component to better model high-severity losses. Additionally, the sample size was relatively limited for stable parameter estimation, and, for some variables, the available classifications lacked the granularity needed for more detailed segmentation.

Moreover, in assessing the reinsurance treaty, we found that its current structure provides limited reduction in retained losses. While lowering the attachment point under the excess of loss arrangement could theoretically increase risk transfer, the relatively high ceded premium, particularly due to the security loading, reduces the cost-effectiveness of such an adjustment. As a result, the current treaty structure may offer limited financial benefit relative to its cost.

To explore these findings in greater depth, this study is structured into five subsequent chapters. Chapter 2 provides an integrated overview of workers' compensation (WC) and its reinsurance, beginning with the historical and legal context of WC and its implementation in Portugal. It then explores the structure of the WC line of business at Lusitania, outlines key features of reinsurance contracts (with a focus on their application to WC) and concludes with a literature review establishing the theoretical framework for the analysis. Chapter 3 focuses on presenting the available data, as well as its limitations. Chapter 4 outlines the methodological statistical framework used to estimate aggregate claims, focusing on a GLM-GPD combination to model claim frequency and severity. Chapter 5 builds on these results to analyse the structure and performance of the current reinsurance treaty, assessing its effectiveness in mitigating the insurer's risk exposure given the modelled loss distribution. Finally, Chapter 6 synthesizes the main findings and acknowledges the limitations in the research, while also presenting the opportunities for further research.

2 Workers' compensation (WC) insurance and reinsurance

We begin with a review of the relevant background, combining foundational context with theoretical grounding. This chapter is structured into three main sections: first, we explore the historical and legal foundations of workers' compensation, with particular attention to the system currently in place in Portugal. Second, we examine the structure of the WC line of business at Lusitania, including its specific operational and regulatory context. Lastly, we provide an overview of reinsurance concepts, with a focus on non-proportional treaties, and review the academic literature that frames and motivates our approach to modelling and analysis in subsequent chapters.

2.1 History and legal framework

The origins of workers' compensation can be traced back to the Code of Ur-Nammu in ancient Mesopotamia, dated approximately to 2050 B.C., which stipulated financial indemnification for bodily injures sustained by workers (Guyton,1999). While most empires throughout ancient history had detailed schedules for payments depending on the injured (or lost) body part, we see a greater development in the laws surrounding this topic in the period between the Middle Ages and the Industrial Revolution. They had, however, extremely restrictive definitions of the employers' liability, wherein a worker had to prove through extensive legal disputes that their injuries were the result of risks that went beyond the inherent risk of the position and that they had not been negligent in order to receive compensation.

Draper, Ladou, and Tennenhouse (2011) posit that by the early 1900's, every European country had enacted legislation regarding workers' compensation, largely influenced by two models: the British and the German. The British model featured an elective process with administration managed by the courts and insurance provided by private firms. In contrast, the German model was based on the existing centralized social security system.

The modern design of WC in Europe falls largely within two categories, with some countries combining elements of both, all funded by employer contributions, as outlined by Walters (2007). The key distinction lies in the management of the funds: one model involves a state-managed system, whereas the German-inspired schemes rely on employer contributions to private insurers.

Portugal's workers' compensation bears more similarities to the latter. Its history dates back to the First Republic, with the Law $n°$ 83 of July 24 (1913) establishing the foundation for the current structure by introducing the right to clinical assistance, medicine, and compensation for employees who are victims of work accidents. Today, Portugal maintains a comprehensive scheme, based on the employer's strict liability if a causal relationship is established, and regulated by Law $n°$ 98 (2009), in accordance with articles 283 and 284 of the Labour Code

(2009).

As established in Law $n°$ 98 (2009), private insurance is mandatory for all employers, including the self-employed, and covers a multitude of work-related injuries, illnesses, and even death. Disability evaluation follows a standardized process based on the national disability rating scale for workplace injuries and occupational illnesses.

The insurer is responsible for compensation for any incident that occurs at the workplace or during work hours, leading to physical injury, loss of function, or illness that reduces the worker's ability to work or earn, or results in death, including:

i. Accidents occurring during commuting to (or from) work, including travel between multiple work locations, between the worker's home and the workplace, and between work-related locations.

ii. Accidents occurring during voluntary services provided for the employers benefit.

iii. Accidents during activities related to worker representation or participation in training, either at or outside the workplace, when authorized by the employer.

iv. Accidents occurring while seeking employment, when granted hours for this purpose by law, in the case of termination of the employment contract.

Finally, in the case of accidents during commuting, the employer whose workplace the worker is travelling to is held responsible for the accident. Moreover, the law specifies that the employer will not be deemed liable for damages resulting from a work accident if:

i. The accident was deliberately caused by the worker, or resulted from actions or omissions that violated legal or employer-mandated safety rules, without valid justification.

ii. The accident was strictly due to the worker's gross negligence.

iii. The accident occurred due to the worker's loss of mental capacity (whether temporary or permanent) unless this condition was work-induced, beyond the worker's control, or the employer knowingly allowed the worker to continue under unsafe conditions.

In the event of a loss, the employee (or their beneficiary) has the right to compensation either in-kind (medical, surgical, pharmaceutical, hospital, and other necessary services aimed at restoring the individuals health and work capacity, allowing them to return to active life) or in cash (including indemnities, pensions, allowances, and subsidies meant to compensate the financial losses caused by the injury or incapacity to work).

Cash benefits are frequently structured as long-term annuities, particularly in cases of permanent disability or death. These annuities create ongoing, sometimes lifetime, payment obligations that represent long-duration liabilities on the insurer's balance sheet. To guarantee solvency and compliance with regulatory standards, these obligations must be backed by mathematical reserves, reflecting both demographic and financial risks. Given the potential for high-severity, low-frequency events to generate disproportionate losses, reinsurance can be used as a risk mitigation tool (Vaughan and Vaughan, 2013).

2.2 WC line of business in Lusitania

Workers' compensation insurance is comprehensive in nature, encompassing both life and non-life risks. This presents a unique challenge for the actuary in terms of risk assessment, pricing, and reserving, as it requires expertise in both life and non-life practices, as well as knowledge of occupational health and disability. At Lusitania, WC claims are divided into two main categories: simple or severe claims.

Simple claims are sometimes referred to as other benefits (OB) and are not covered by the reinsurance treaty under analysis in this thesis. They encompass all the in-kind compensations (medical, surgical, pharmaceutical, hospital, and other necessary services aimed at restoring the individuals health and work capacity, allowing them to return to an active life) and are modelled using non-life techniques (Non-SLT health, or NSLT), such as the Chain-Ladder and Bornhuetter-Ferguson methods, due to their short-term nature.

Once a worker is discharged from medical care, their physician might determine that they have suffered from lasting effects from the accident that diminishes their ability to return to the workforce (either fully or partially). When that happens, a severe claim is then notified to the insurer.

Despite the terminology, "severe" claims in workers' compensation do not necessarily refer to their immediate monetary value, but rather to their long-term impact on the injured worker's life and capacity to work. They are further divided into two subgroups: pensions and lifetime assistance (LTA).

Lifetime Assistance is provided to victims with injuries requiring lifelong medical care. It covers medical expenses that are not regular payments but offer long-term support and are calculated by using life actuarial techniques.

An annual pension is also granted to the injured worker to compensate for the loss of working or earning capacity, or to their legal beneficiaries in the event of death. Pensions are covered by the Mathematical Reserves, which are established by employing life actuarial techniques (SLT) in accordance with local legislation and the company's technical basis.

As defined by Law 98 (2009), the awarded pension benefit depends on the degree of disability and the beneficiary's gross annual salary plus vacation and Christmas benefits in the 12 months preceding the end of of exposure to the risk (or the date of homologation of the illness that determines incapacity, if applicable). In cases of permanent absolute incapacity for any and all work (IPAT, from the Portuguese *Incapacidade Permanente Absoluta para Todo Trabalho*) or death, the pension is equal to the full remuneration. For permanent absolute incapacity for the usual occupation (commonly referred to as IPAH, from the Portuguese *Incapacidade Permanente Absoluta para Trabalho Habitual*), the pension ranges from 70% and 100% of the remuneration, depending on the extent of the worker's residual functional capacity to perform another compatible job. In cases of partial incapacity (IPP, from the Portuguese *Incapacidade Permanente Parcial*) the pension amount is determined based on the degree of capacity loss resulting from the accident.

The mathematical reserves are thus the product of the aforementioned pension benefit by the corresponding monthly annuity, which is determined for each beneficiary based on several factors, including their relationship to the victim, the type of pension, their age, their medical condition, and the degree of the victim's disability. By legal requirement, they are calculated using common commutation functions (Anderson, 1999), with basic structure similar to what was presented by Rosa (2012), where Y_x represents the present value of unit monthly payments paid in advance for a pensioner aged x with a fixed interest rate:

Victims are subject to a whole life annuity in advance:

$$E[Y_x] = \ddot{a}_x^{(12)} \approx \ a_x + \frac{13}{24} \tag{1}$$

Spouses or Ascendants are also modelled by whole life annuity in advance with an additional payment that starts at the national retirement age (represented by the variable r):

$$E[Y_x] = \ddot{a}^{(12)}_{x:\overline{r-x}|} + \frac{4}{3} \times \ {}_{r-x}E_x \ \times \ \ddot{a}_r^{(12)} \tag{2}$$

Descendants (orphans) follow a temporary life annuity in advance until the age of 25:

$$E[Y_x] = \ddot{a}^{(12)}_{x:\overline{25}|} = \ddot{a}_x^{(12)} - {}_{25-x}E_x \ \times \ \ddot{a}_{25}^{(12)} \tag{3}$$

where,

$$a_x = \sum_{k=1}^{\infty} v^k \ {}_kp_x \text{ is the whole life annuity in arrear} \tag{4}$$

$${}_tP_x = \Pr[T_x > t] = S_x(t) \text{ is the prob. a life aged } x \text{ survives for at least } t \text{ more years} \tag{5}$$

$${}_tq_x = \Pr[T_x \leq t] = 1 - S_x(t) \text{ is the prob. a life aged } x \text{ does not survive beyond age } x+t \tag{6}$$

$$ {}_nE_x = v^n \, {}_nP_x \text{ is the expected present value of the pure endowment} \tag{7} $$

Other considerations, such as 50% aggravation on the degree of disability for pensioners at the age of 50, with a limit of 100% of the loss of ability to work and a provision in the event that the widower remarries, are also incorporated in the company's model, in line with National Table of Disabilities due to Work Accidents and Occupational Diseases (TNI, from the Portuguese *Tabela Nacional de Incapacidades por Acidentes de Trabalho e Doenças Profissionais*), as established in Decree $n°$ 352 (2007).

It is also worth noting that pensions are further split into obligatorily redeemable (OR) and non-obligatorily redeemable (NOR), which impacts their calculations. OR pensions have their technical basis set by law, whereas NOR pensions allow for more flexibility and judgment from the actuary. Article 75 from Law $n°$ 98 (2009) establishes as obligatorily redeemable every annual lifelong pension owed to an injured person with a permanent partial disability of less than 30%, provided that the annual pension amount does not exceed six times the guaranteed minimum monthly wage in effect on the day following the date of discharge or death. Their calculation is subject to a 5.25% interest rate to be applied to the TD 88/90 mortality table.

2.3 Reinsurance for WC

Despite being administered by private insurers, workers' compensation insurance fundamentally addresses a social risk. As such, regulatory mechanisms are in place to ensure that insurance companies can fulfil their obligations to the workers. This system incorporates two notable safeguards.

Fundo de Acidentes de Trabalho (FAT), or Workers' Compensation Fund, was created by Decree no.142 (1999), and has been in force since January 2000. This fund, managed by the Portuguese Insurance and Pension Funds Supervisory Authority (ASF), acts as a guarantor, stepping in to provide benefits in cases where the responsible insurer is unable to meet its financial obligations. FAT thus ensures continuity of protection for workers, even in the event of insurer's insolvency or other financial distress.

Alternatively, insurance companies can also rely on reinsurance to manage their risk.

2.3.1 Reinsurance

Reinsurance is essentially a safety net, an agreement between two companies: the direct insurer (called the "cedent") and another insurance company that is responsible for "insuring the insurer" (the "reinsurer") (Vaughan and Vaughan, 2013). In this setup, the cedent transfers part of its risks to the reinsurer so that the reinsurer is liable for the agreed potential costs associated with those policies, upon the payment of a premium. (Centeno and Simões, 2009)

The core purpose of reinsurance is to spread out risk. Insurance companies that underwrite policies with high risk or high insured amounts are inherently exposed to the potential for significant and unforeseen losses. They can mitigate those risks by transferring a portion of their liabilities' portfolio to another insurer. Additionally, reinsurance can enhance key financial indicators that are closely monitored by regulators, such as solvency ratios, and allow insurers to underwrite policies that would have exceeded their risk tolerance or capital reserves otherwise (Friedland, 2023).

However, reinsurance also comes with limitations and potential drawbacks. While those contracts offer protection against high variance in the expected loss ratio, their costs can reduce underwriting margins and profits, especially in the presence of elevated security loadings in reinsurance premiums. As explored by Cummins, Dionne, Gagné, and Nouira (2021), reinsurers often charge high premiums in order to properly compensate for their administrative expenses and the inherent uncertainty of covering tail risks, generating substantial costs for the insurer. Other researchers, such as Froot (2001), further highlight how premiums can considerably outweigh expected losses in the case of non-proportional reinsurance.

There are two primary forms of reinsurance: obligatory reinsurance and facultative certificates. Facultative certificates are negotiated on a case-by-case basis. The ceding insurer offers specific risks or policies to a reinsurer, who can choose to accept or decline each one individually. However, despite offering more flexibility, this approach tends to be more complex and costly, placing a heavier administrative burden on both sides. On the other hand, obligatory reinsurance is placed as a treaty where the cedent and reinsurer agree on terms that apply to all policies within a defined category or line of business, thus covering a wider range of risks (NASDAQ, 2024).

Within those contracts there is still a distinction on how losses are divided between ceding insurers and reinsurers, by classifying the contracts as proportional or non-proportional, as described by Centeno and Simões (2009). The former consists of the proportional allocation of premiums and losses between the parties according to a set percentage, which can be done in one of two ways: either the reinsurer accepts a fixed portion of all premiums and risks of a determined portfolio (quota-share, or QS); or the cedent can retain a predetermined percentage of the risk and cede only the excess to the reinsurer (surplus share).

In non-proportional reinsurance (NP), the cedent agrees to cover all losses up to a predetermined level (priority or retention) and the reinsurer then assumes responsibility for the portion of claims above that threshold, usually with a coverage limit of their own. Premiums for such arrangements are typically based on the ceding insurer's claims history and risk exposure, with possible adjustments at year-end. NP contracts can be essentially classified into stop-loss and excess of loss (XL).

Stop loss contracts protect insurance companies from high frequency of claims. Coverage is triggered when the aggregate loss over a period exceeds a predetermined threshold, either as a percentage of premium income or a fixed amount, known as the "attachment point", providing stability against adverse volatility in the loss ratio. (MunichRe, 2022)

Excess of loss (XL) reinsurance addresses high-severity claims from individual risks or events. In this arrangement, the reinsurer is liable for losses in excess of a specified retention limit. This protection can be applied per individual risk, per occurrence, or in aggregate (Friedland, 2023). XL contracts are typically divided into two types: Risk XL (working covers) and Catastrophe XL (CATXL).

Given their complexity, reinsurance contracts must clearly define responsibilities and exclusions. Supplementary clauses are often used to enhance clarity, avoid ambiguity, and limit potential litigation.

One example of such provision that is commonly seen in reinsurance contracts is the sunset clause. Sunset clauses establish a deadline for the cedent to report claims or adjustments to their reinsurer, thereby curtailing the duration of emerging loss development and promoting faster notification of claims (Hutter and Graves, 1992). As such, reinsurers implement sunset clauses in order to incentivize prompt reporting practices, as the cedent risks losing coverage if they fail to meet the specified time frame. For further reading, Bear and Nemlick (1990) studied the impact of other loss sharing provisions, such as loss ratio caps, limited reinstatement, and loss corridor clauses, in XL treaties.

2.3.2 Reinsurance for WC in Lusitania

The worker's compensation portfolio at Lusitania is protected by an excess of loss reinsurance treaty. Coverage is limited to the mathematical reserves for pensions related to permanent disability and death, as per Law n° 98 (2009), as well as the mathematical reserves associated with additional pensions related to the total salary. It extends to the entire calendar year of the contract (from January 1 to December 31), on a loss occurring basis.

Under this agreement, the reinsurer covers losses from individual events that exceed a specified retention amount borne by Lusitania. However, the reinsurer's total liability is capped per layer, meaning there is a maximum payout limit for all losses within a single treaty year.

Correspondingly, the contract also stipulates the parameters that shall be used by the cedent to calculate the mathematical reserves (to be calculated with life actuarial techniques as stipulated in the previous section 2.2):

- NOR: a chosen interest rate applicable to the INE 2020-2022 table; and
- OR: interest rate of 5.25% applicable to the TD 88/90 table in accordance with the legal requirement.

Furthermore, per the sunset clause, the liability of the reinsurer is limited to a number of years from the date of the claim. Any subsequent liability arising from claims that were not reported or from changes in the mathematical reserves are considered non-existent for reinsurance purposes.

Finally, it is important to highlight that the treaty only covers the ultimate net loss, meaning that if the responsibility of indemnification is split amongst multiple insurance companies, the reinsurer is only liable if Lusitania's portion of the risk exceeds the priority. This can happen, for example, in the case of coinsurance or when the accident occurs during the insured's commute, which can be covered by both motor and WC policies from different insurers.

2.3.3 Literature review

Specific literature on reinsurance for workers' compensation has developed over time, but it still remains considerably sparse. Earlier works focused on straightforward risk transfer models, but by the 1990's, researchers began to include long-tail liability and impact of inflation of medical costs in their estimations. In recent years, more sophisticated predictive modelling techniques have been studied, examining how factors such as claim development patterns, legislative changes, and catastrophic exposures affect optimal reinsurance structures.

When reviewing the literature on workers' compensation reinsurance, we categorized the relevant works into two primary groups: studies on the estimation of high-severity losses in this line of business and studies on optimal reinsurance strategies.

Estimation of high-severity losses in workers' compensation portfolios

Milicia and Dolsky (2022) review several actuarial approaches used to determine ceded reserves, especially under working-layer excess of loss arrangements. Among the four main techniques discussed, the most commonly applied relies on estimating both gross and net reserves independently. This method is often favoured due to the relative stability of the net data and the more predictable development of gross claims over time. Direct estimation of ceded reserves requires access to detailed reinsurance records. Since these data may not be as complete or developed as gross or net figures, actuaries must exercise greater caution when setting assumptions, particularly when dealing with varying treaty layers and limited historical information. As an alternative, some approaches apply reinsurance terms to either gross or net figures to take advantage of their relative reliability. However, these approximations may not fully capture the complexities or specific conditions of the actual reinsurance treaty.

Several studies have examined the long-term liabilities associated with workers' compensation claims by integrating mortality assumptions into traditional property and casualty (P&C) actuarial methods (which commonly rely on historical loss development patterns without explicitly accounting for survival probabilities) due to the aforementioned unique nature of workers' compensation, where benefits can extend for decades or even the claimant's lifetime.

Sherman and Diss (2005) proposed a stochastic model that integrates a Markov chain framework with incremental claim payment data, allowing for more dynamic tail factor projections. This approach attempts to balance two opposing factors: the escalation of medical costs and the gradual attrition of claimants due to mortality. However, the practical applications of this model have been challenged by Shane and Morelli (2013) due to insufficient data and overly parametrised models. In response, they suggested a more practical reserving method that incorporates life expectancy percentiles into tail estimations, using mortality patterns as a proxy for claim run-off horizons.

Schmid (2012) introduced a Bayesian model using Markov Chain Monte Carlo (MCMC) simulations to better understand how long-tail WC liabilities behave, while Jones, Scukas, Frerman, Holt, and Fendley (2013) extended this work with further empirical validation. Moreover, a recent study by Duncan (2020) focused on the regulatory context in California, comparing standard actuarial mortality tables with termination rates estimated via survival models (using Kaplan Meier and Cox proportional hazard models). The findings suggested that conventional tables often understate actual mortality, which can lead to overestimated liabilities.

The theoretical framework for excess-of-loss reinsurance in WC is primarily shaped by four key papers. Ferguson (1971) proposed separating gross reserves into net and ceded components using a life annuity approach that accounts for mortality and discounting. This work was later updated by Steeneck (1996), who examined the effect of inflation and claim severity on catastrophic exposures. He showed that traditional incurred but not reported (IBNR) methods are not as reliable when dealing with layered reinsurance structures and long-duration claims. The results suggest that annuity-based models estimate both medical and indemnity components more accurately.

Since then, a few studies have explored alternatives to improve the precision of the models by adopting stochastic techniques to estimate ceded loss. Blumsohn (1997) studied the limitations of deterministic reinsurance commutation models by replacing fixed assumptions for variables such as medical costs and inflation, and cost-of-living adjustments (COLAs) with stochastic counterparts. The study demonstrated improvements in reserve accuracy and reduced bias, particularly in the estimation of ceded losses under non-proportional reinsurance agreements.

As one can see, those studies primarily focus on mortality and longevity risks, whereas the effects of longevity are somewhat curtailed in our reinsurance analysis by the presence of the sunset clause in the treaty. More recently, Fu and Liu (2016) performed a comparison of finite mixture models (FMMs) and standard generalized linear models (GLMs) in workers' compensation data. They estimated claim counts by value range and provided a comparative analysis between both frameworks, finding that a combination of multiple gammas improved on various distribution-related challenges, including heavy tails, in the forecasting of excess losses above deductibles.

Ghaddab, Kacem, de Peretti, and Belkacem (2023) were among the first to successfully apply a combined GLM-GPD framework, as proposed by Laudagé, Desmettre, and Wenzel (2019), to real insurance data demonstrating its practical effectiveness. This model consists in applying a Generalized Linear Model (GLM) to the bulk (attritional) portion of the claims distribution and a Generalized Pareto Distribution (GPD) to the tail. While their original application focused on fire insurance, the core modelling structure proved highly adaptable to other lines of business, thus providing the main methodological reference for our estimation of claim amounts, particularly due to the heavy-tailed nature of the workers' compensation dataset. In our case, the very right-skewed distribution of workers' compensation severity data made this two-part approach especially relevant. Although the exposure types and claim drivers differ between fire and workers' compensation, the statistical challenges posed by rare but severe claims are very similar. By adapting their methodology to our dataset, we were able to capture both the body and tail of the distribution more accurately and apply excess-of-loss treaty terms in a consistent manner.

Optimal reinsurance strategies

Optimal reinsurance is a strategy used by insurers to transfer risk to reinsurers while minimising costs and maximising benefits under specific constraints. To the best of our knowledge, despite the extensive research on optimal reinsurance strategies for other lines of business in the past decades, there is a notable lack of studies specifically addressing the issue for WC.

In contrast, several researchers have examined optimal reinsurance for life annuities, often using stochastic simulations to address longevity and financial market risks. Many studies conclude that combining proportional and non-proportional reinsurance improves risk transfer and capital efficiency. For example, De Angelis (2011) and Baione, Angelis, Menzietti, and Tripodi (2017) show that hybrid structures like Surplus-Excess of Loss (SXL) perform well under solvency and profitability constraints, while Veprauskaite and Sherris (2014) and Marcarelli and D'Ortona (2012) emphasise the benefits of multi-layered strategies and the interaction between mortality and market risks.

Although demographic risks are associated with the payment of the claims, they are already incorporated into the calculation of the mathematical reserves and the presence of the sunset clause, which excludes deaths and other adjustments that occur after a predefined number of years from reinsurance considerations, further limits their impact. As a result, it would also be reasonable to adapt certain non-life actuarial techniques in our analysis.

The research on non-life optimal reinsurance spans a wide range of analytical frameworks.

Centeno and Simões (2009) provide a synthesis of key models and their practical implications, examining both proportional and non-proportional structures under various objective functions. This paper consolidates theoretical advances while also highlighting their relevance to real-world decision-making, thus becoming a foundational reference in this area of study.

Early seminal studies, such as Borch (1960), showed that under the expected value principle, proportional reinsurance is often optimal, while variance loading favours stop-loss contracts. More recent work, including Kaluszka and Okolewski (2008), Cai, Tan, and Weng (2007), and Cheung (2010), has explored how different premium principles affect retention levels and risk-sharing preferences based on tail sensitivity and risk appetite.

Optimal excess of loss reinsurance continues to receive attention across various objectives and modelling approaches. For instance, Bazyari (2025) and Sun, Rong, and Zhao (2019) derive optimal XL contracts under ruin probability and utility-based criteria, respectively. Alcaide (2022) and Serrano (2015) focus on practical deductible selection through simulation and empirical analysis, with the latter applying those concepts to real WC insurance data from a Portuguese insurer. These studies emphasise how design choices in XL structures influence profitability, solvency, and technical results.

More broadly, the literature recognizes that combining proportional and non-proportional reinsurance allows insurers to tailor risk transfer to both frequency and severity of claims, enhancing solvency and capital efficiency. These combination strategies are especially relevant for insurers seeking to balance protection against large individual losses and aggregate claim volatility. Centeno and Simões (1991) investigated how insurers can optimally blend quota-share and XL reinsurance for portfolios of independent risks, using the adjustment coefficient (linked to ruin probability) as the optimization criterion. Their results show that a combination of these two forms can yield superior risk management compared to using either alone.

3 DATA

We focus on losses covered by reinsurance, therefore, we need to analyse both the frequency and severity of claims over the period of study. In this chapter, we provide an overview of the variables included in the dataset.

Data quality plays a crucial role in actuarial analysis, particularly when estimating ceded losses under an excess of loss treaty. As Milicia and Dolsky (2022) note, common issues include limited data credibility and irregular loss patterns, with periods of little to no activity punctuated by sudden, significant spikes. These challenges can complicate reliable loss estimation and model performance. Moreover, these issues become even more complex in portfolios like Lusitania's, where factors such as mergers, acquisitions and regulatory changes, including the introduction of Law no. 98 (2009), marked turning points in data consistency.

In this work, we utilise data from 2015-2024 (by accident year) with the purpose of ensuring a comprehensive analysis of the reinsurance treaty's performance under the current retention limit and regulation. In order to maintain confidentiality, the dataset underwent a masking procedure, where the data, some assumptions and supplementary details were transformed.

3.1 Data description

Two primary databases supplied by the company were considered:

1. Base_N contains records of simple claims within the portfolio that may potentially escalate into severe cases over time.

Due to data access limitations, we did not have the full portfolio of WC insurance policies available for analysis. As a result, we rely on simple claims as a proxy for the broader population of WC claims. While this restricts our ability to account for the complete exposure base or premium structure, simple claims provide a consistent and representative subset for studying claim-level characteristics, such as injury types, causes, and costs.

As outlined in Section 2.2, when a worker is formally discharged from medical care, the initial simple claim, which was already recorded by the insurer, has the potential to escalate. Understanding the progression from simple to severe claims is crucial, as it provides insight into how various risk characteristics may influence this transition. Thus, the simple claims database serves as the foundation for estimating the number of severe claims.

2. Base_X offers detailed information regarding the nature and financial impact of severe claims only. This database consists in the claim amounts calculated using life actuarial techniques under the terms of the reinsurance treaty in force at the time of the accident (as described in Section 2.3.2) and provides us with an outlook on the value and characteristics of each severe claim that has developed from the simple claims.

Both datasets are fairly similar in structure and content. The main difference is that Base_N includes an indicator variable (Severe claim) identifying whether a severe claim occurred (0 or 1), whereas Base_X extends this by also including the actual amounts associated with those severe claims. The variables included in the datasets are listed below, with a more detailed description provided in Appendix A.

- Severe claim
- Accident year
- Weekday
- Gender
- Work
- Cause
- Body part
- Bodypart count
- Injury
- Injury count
- Age
- Salary
- Gross cost
- Claim amount

3.2 Data preparation

Our first step was to identify the cases of missing information for each explanatory variable in order to understand their impact, as seen on Tables IX and X. It is known that the presence of incomplete information can introduce bias, reduce model accuracy, and compromise the credibility of the results (Little and Rubin, 2002). In particular, the `glm` command in `R` performs listwise deletion, excluding rows with missing values in any of the variables. In the context of this analysis, missing values in categorical variables appeared either as blank fields or as "Other"/"Unknown" labels. While both indicate incomplete information, the latter may reflect deliberate classifications or limitations in category definitions. Blank entries were treated as missing (NA), while "Other"/"Unknown" were kept as separate, explicit factor levels to preserve potentially meaningful distinctions and avoid conflating structurally different types of missingness in the analysis (especially in variables like Cause, where the frequency of "Other"/"Unknown" was high).

To assess whether the missing values are randomly distributed or systematically related to other variables, Little (1988)'s MCAR (Missing Completely At Random) test was conducted. The results of this test are presented in Table XI and indicate that missing data can reasonably be assumed to be missing at random, which mitigates concerns about bias.

Given these results, imputation was applied. This is a reasonable approach, as simply excluding missing values could lead to biased estimates. Missing numeric variables were replaced with the mean, while missing categorical variables were subject to mode imputation. These techniques were chosen for their interpretability and minimal impact on data distribution.

For numeric variables, we also addressed data quality issues such as invalid values and outliers through the visual inspection of Figure 10. The distribution of salaries is highly skewed to the right, with a large number of outliers, including some values that far exceed the whiskers.

One could opt to treat these extremely high values with techniques such as winsorising or mean imputation, however, given the relationship between Salary and claim amounts (which was described on Section 2.2) and our ultimate goal of better understanding those rare and severe claims, we chose to retain those values in the analysis.

When we look at the age of the claimants at the time of the accident, most values fall within a typical adult working range (roughly 25-70). However, a few clear outliers, such as ages below the minimum working age in Portugal, likely due to data entry errors or anomalies, were identified. These values were not representative and were therefore replaced using mean imputation to maintain data consistency.

Finally, to enhance model interpretability and account for potential non-linear effects, the continuous variables Age, Salary, and the Gross Cost of Simple Claims were transformed into categorical variables using k-means clustering. This unsupervised technique divides the data into a set number of groups in a way that maximizes similarity within each group and differences between groups. It helps reveal underlying structures in the data that would likely be missed if only linear relationships were assumed (James, Witten, Hastie, and Tibshirani, 2021). Discretizing continuous variables in this way can simplify model specification in GLMs, allowing the model to better accommodate non-linearities and improve robustness (Harrell, 2015).

In addition to the numerical features, cluster analysis was applied to reduce the dimensionality of the work categories, which were highly granular and differed in structure between worker types. The dataset comprised both self-employed and employer-funded claims. For self-employed individuals, occupations were listed as one of 492 general job titles, while for employer-funded workers, job classifications followed the *Classificação Portuguesa das Actividades Económicas* (Portuguese Classification of Economic Activities, or CAE), which categorizes work by economic sector, in accordance with Decree $n°$ 9 (2025).

This heterogeneity, along with the large number of distinct categories, posed modelling challenges such as overparameterization and data sparsity, limiting the direct use of these variables in GLM modelling (Fahrmeir, Kneib, Lang, and Marx, 2013; Goldburd, Khare, Tevet, and Guller, 2025). To address this, separate cluster analyses were conducted for each group, using patterns in the response variable to group similar occupations. Six clusters were created for each: labeled "Job_1" through "Job_6" for self-employed workers, and "CAE_1" through "CAE_6" for employer-funded workers. These aggregated categories captured similar risk patterns while preserving interpretability, improving the stability and explanatory power of the GLM.

3.3 Exploratory and descriptive data analysis

This section presents a detailed analysis of the available datasets, structured separately for Base_N and Base_X.

3.3.1 Analysis of claim counts

Base_N is a longitudinal database containing $142{,}133$ observations, with each row representing a unique simple claim. This dataset will be used to analyse claim frequency in the following chapters.

To explore broad trends, we plotted the average rate of severe claims across all categorical variables, including standard deviation whiskers (Figure 11), and highlight a few key findings. Notably, the probability of a severe claim increases with age. Injury type also plays a crucial role: "amputations" and "asphyxiation and gas inhalation" show particularly high rates, while "superficial wounds" and "sprains" rarely lead to severe claims. "Full-body" and "brain" injuries exhibit the highest rates, and cases involving high costs and salaries are also associated with increased severe claim incidence, indicating concentrations of both medical and financial risk. While these visualizations provide valuable initial insights, such as the elevated risk linked to causes like "explosions" or "electrical current", more robust patterns are captured through the GLM analysis discussed in Section 4.1.

Then, to evaluate the relationships between categorical explanatory variables, we calculated Cramér (1946)'s V statistic, which quantifies association strength on a scale from 0 (no association) to 1 (perfect association), and whose values are presented in Table I. Most variables show low association values (< 0.1), indicating minimal collinearity. Notable exceptions include the moderate association between Work and Gender (0.413), and between Cause and Body part (0.235).

TABLE I: Cramér's V Associations Between Categorical Variables

	Accident_year	Weekday	Cause	Bodypart	Bodypart_count	Injury	Injury_count	Work	Gender	Age_band	Salary_band	Gross_cost_band
Accident_year	1.000											
Weekday	0.011	1.000										
Cause	0.089	0.024	1.000									
Bodypart	0.036	0.023	0.235	1.000								
Bodypart_count	0.014	0.007	0.064	0.051	1.000							
Injury	0.068	0.021	0.238	0.249	0.057	1.000						
Injury_count	0.012	0.008	0.031	0.034	0.293	0.037	1.000					
Work	0.037	0.059	0.051	0.038	0.013	0.041	0.018	1.000				
Gender	0.082	0.071	0.202	0.118	0.022	0.099	0.015	0.413	1.000			
Age_band	0.035	0.017	0.073	0.059	0.021	0.044	0.014	0.075	0.040	1.000		
Salary_band	0.185	0.013	0.063	0.036	0.008	0.047	0.007	0.067	0.097	0.067	1.000	
Gross_cost_band	0.013	0.011	0.045	0.057	0.009	0.078	0.008	0.024	0.027	0.019	0.025	1.000

3.3.2 Analysis of claim amounts

Base_X consists of $9{,}408$ unique claims from the years 2015 through 2024. To gain a better understanding of the characteristics and distribution of the claims data, a descriptive statistical and graphical analysis was conducted on the key loss variables.

The empirical loss data, as seen on Figure 1 and Table II, shows heavy concentration of claims below the mean loss of EUR $11{,}541.92$, with 84.13% falling beneath this threshold. This reflects a typical pattern of many low-cost, high-frequency claims. An analysis of the statisti-

Percentiles	Value
25%	2,636.09
50% (Median)	4,629.64
75%	8,831.06
84.1% (Mean)	11,541.92
90%	17,913.10
95%	33,399.60
99%	161,067.49
99.5%	226,844.45
99.9%	410,199.91

TABLE II: Percentiles of the claim amounts

Category	Value
Mean	11,541.92
Standard Deviation	34,279.50
Coefficient of Variation	2.97
Skewness	9.71
Kurtosis	141.99

TABLE III: Statistical moments of the claim amount distribution

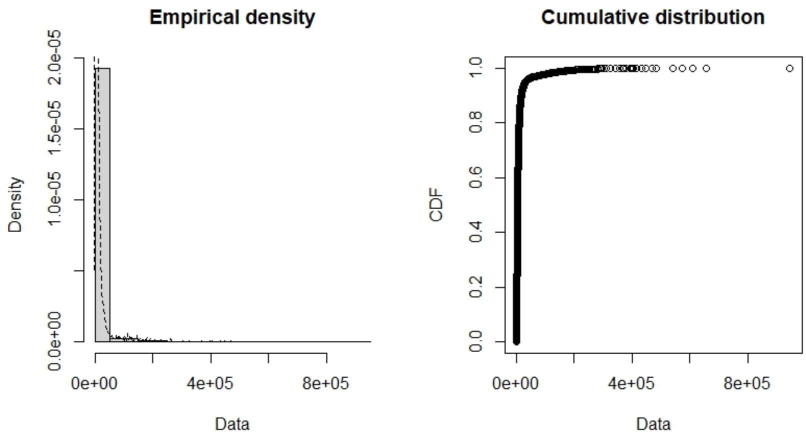


FIGURE 1: Distribution of the claim amounts

cal moments, as seen in Table III, indicates that the distribution is also markedly right-skewed (9.71) and heavy-tailed, with a high coefficient of variation (2.97) and kurtosis (141.99), indicating substantial variability and the presence of a few exceptionally large losses. Moreover, the probability of a claim exceeding the current retention limit (i.e., $P[X > 400,000]$) is approximately 0.1363%. Claims must also be notified to the reinsurer if they surpass 50% of the priority, as they have potential to develop and eventually reach the reinsurance treaty's threshold. In our sample, $P[X > 200,000] = 0.7232$. This suggests that while extreme claims are rare, they are not negligible and must be accounted for in the design of the optimal reinsurance structure.

To further analyse the datas distribution, we plotted the Cullen and Frey (1999) graph (as shown in Figure 2), where the red dot represents the empirical data and orange dots are 2,500 bootstrap replicates showing variability. The empirical data lies far from the normal distribution, clustering closer to lognormal or gamma distributions (based on the position relative to those curves).

Then, similarly to what was done for claim counts, we individually explored each predictor

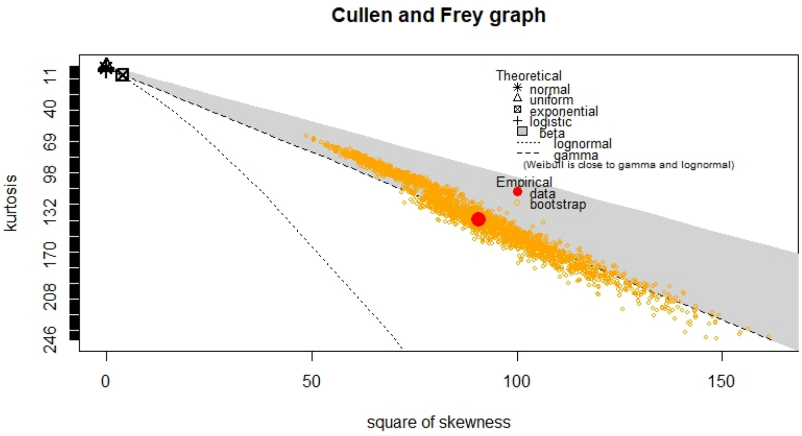


FIGURE 2: Cullen and Frey graph of claim amounts

variable as well as their relationship with our response variable (claim amounts). The distribution of claim amounts across multiple variables was analysed using box plots, with the claim values presented on a logarithmic scale to account for the wide variance and presence of high-value outliers (Figure 12). This approach allowed for clearer identification of differences in central tendency and spread across categories. The main finding was that Weekday showed little impact on claim costs, with similar distributions across groups. In contrast, Injury, Body part, Work, and Cause significantly influenced claim amounts. Serious injuries (like "amputations" and "burns"), self-employed work clusters, and high-impact causes such as "explosions" and "electrical current" were associated with higher and more variable claim costs.

Finally, Table IV shows the results for Cramér (1946)'s V correlation matrix, revealing generally low-to-moderate associations among categorical variables. The strongest relationship appears between Bodypart_count and Injury_count (0.402), suggesting a consistent link between the number of injured body parts and total injuries reported. Similarly, Gender and Work show moderate correlation (0.394), possibly reflecting gendered job roles in the dataset. Cause and Injury (0.266), as well as Bodypart and Injury (0.321), also demonstrate meaningful associations, indicating consistent injury patterns related to certain causes and affected body parts. Most other pairwise associations remain below 0.15, suggesting weak or negligible associations.

TABLE IV: Cramér's V Correlation Matrix for Categorical Variables

	Accident_year	Weekday	Cause	Bodypart	Bodypart_count	Injury	Injury_count	Work	Gender	Age_band	Salary_band	Gross_cost_band
Accident_year	1.000											
Weekday	0.029	1.000										
Cause	0.086	0.041	1.000									
Bodypart	0.068	0.040	0.216	1.000								
Bodypart_count	0.036	0.027	0.077	0.113	1.000							
Injury	0.064	0.041	0.266	0.321	0.074	1.000						
Injury_count	0.044	0.022	0.038	0.100	0.402	0.064	1.000					
Work	0.051	0.054	0.097	0.069	0.018	0.075	0.031	1.000				
Gender	0.070	0.055	0.247	0.143	0.024	0.126	0.035	0.394	1.000			
Age_band	0.056	0.034	0.079	0.077	0.030	0.069	0.020	0.068	0.092	1.000		
Salary_band	0.164	0.029	0.072	0.057	0.022	0.064	0.021	0.104	0.137	0.066	1.000	
Gross_cost_band	0.074	0.027	0.096	0.136	0.049	0.092	0.026	0.053	0.019	0.032	0.132	1.000

4 MODELLING AGGREGATE LOSSES

This chapter outlines the methodology and results used to estimate the expected aggregate loss for the WC portfolio under study. The modelling approach is grounded in statistical methods commonly employed in actuarial science, with a particular emphasis on Generalized Linear Models (GLMs) and extreme value theory (EVT).

Dunn and Smyth (2018) provide a comprehensive overview of GLMs coupled with a practical application using `R`. As the name suggests, GLMs generalize traditional linear regression by including response variables with distributions from the exponential family (EDM). Instead of simply assuming constant variance and normality, they use a link function to connect the expected outcome to explanatory variables, allowing for non-normal responses.

In recent years, GLMs have become an essential tool for actuaries, offering flexibility in modelling the frequency and severity of insurance losses while maintaining statistical rigour (Goldburd et al., 2025). Moreover, GLMs enable actuaries to directly integrate rating variables and policyholder characteristics into the modelling process, making them particularly valuable in non-life insurance tasks such as pricing (tariff setting) and claims reserving (Ohlsson and Johansson, 2010).

Another statistical framework that will be used to model the severity of losses is extreme value theory. EVT pays particular attention to understanding tail behaviour of distributions, offering tools such as the Generalized Pareto Distribution (GPD). Notably, McNeil, Frey, and Embrechts (2015) show that "the GPD is the canonical distribution for modelling excess losses over high thresholds". Multiple papers have explored the application of EVT in insurance, with Henry and Hsieh (2009), for example, concentrating on modelling the tail of the distribution under the assumption of Pareto-type behaviour and by assessing the estimator's performance via simulations and real-world insurance data.

The modelling of the aggregate loss, S, in this thesis is developed through a two-part framework by assuming independence between the severity and frequency of losses, as is common in insurance applications.

The number of claims, N, was modelled using a Binomial GLM, not to estimate the overall probability of any claim occurring, but rather to model the likelihood of a simple (initial) claim evolving into a more severe, long-term liability. Due to limited access to the full workers' compensation portfolio data, we focused on simple claims as a proxy starting point. This approach is supported by the structure of WC claims, where more severe claims are notified after an initial, simple claim (i.e., following a worker's discharge from medical care). The Binomial GLM framework captures the binary outcome (whether a simple claim becomes severe) and enables the inclusion of relevant policyholder or claim-level covariates to understand this progression.

Claim severities, X, were divided into two segments based on a threshold of EUR $200{,}000$, reflecting the distinct behaviour of smaller versus larger claims. This segmentation allows the model to better capture the heterogeneity in claim sizes, following a similar approach to that used in fire insurance by Ghaddab et al. (2023), as discussed in Section 2.3.3. Claims below this threshold (X_1) were modelled using a Lognormal GLM, which is well-suited for right-skewed data typical of moderate insurance losses. Claims equal to or above the threshold (X_2) were modelled using a Generalized Pareto Distribution (GPD), consistent with extreme value theory, to appropriately capture the behaviour of large, rare losses. The choice of the EUR $200{,}000$ threshold is further motivated and justified in Section 4.2.

Finally, it is important to highlight that all analyses were conducted in the statistical software `R` with the help of packages such as `stats`, `actuar`, `MASS`, `fitdistrplus`, and `evir` for model fitting and diagnostics.

4.1 The number of claims

We used a Binomial Generalized Linear Model (GLM) with a logit link function to estimate the probability that a simple claim would trigger a severe one, which we then used to infer expected claim counts. By modelling the binary outcome variable (indicating whether a serious claim had been generated from the simple claim), the GLM model provided estimates of how each predictor influenced the likelihood of a claim. The logit function allowed us to estimate the log-odds of a severe claim occurring.

In order to ensure the model's predictive performance and avoid overfitting, the dataset was randomly split into a training set (80%) and a test set (20%). We first specified a baseline binomial GLM using both categorical and continuous predictors on the training set (train_data), as follows:

```
glm(formula = Severe claim ~ Accident_year + Weekday + Cause + Bodypart +
Bodypart_count + Injury + Injury_count + Work + Gender + Age_band + Salary_band
+ Gross_cost_band, family = binomial(link = "logit"), data = train_data)
```

To identify the most parsimonious model while preserving predictive accuracy, we performed stepwise model selection using the function `step` in `R`. In this procedure, predictors are added or removed one at a time, and at each step the algorithm chooses the model with the lowest Akaike Information Criterion (AIC), thus balancing goodness-of-fit against model complexity. The process continued until no further addition or removal of variables improved the AIC. Ultimately, all candidate predictors remained in the final GLM, indicating that each variable provides meaningful explanatory information despite a few exhibiting p-values above conventional significance threshold.

To evaluate the discriminatory power of the final GLM on the test set, we computed the Gini coefficient, a commonly used metric for assessing the performance of binary classification models. Predicted probabilities for severe claim occurrence were generated using the `predict()` function with `type = "response"` on the held-out test data. The resulting Gini coefficient was 0.5084524, indicating adequate discriminatory ability of the model in predicting the occurrence of severe claims.

Moreover, Variance Inflation Factor (VIF) results (Table XII) do not indicate significant multicollinearity in our chosen model. All generalized VIF values, when adjusted for the degrees of freedom $GVIF^{1/(2Df)}$, are close to 1, suggesting low correlation among the predictor variables. Even the highest adjusted GVIF values fall considerably below common thresholds of 5 or 10, confirming that the model's parameter estimates are stable and not distorted by multicollinearity.

We highlight a few key findings from the GLM (Appendix B):

- **Accident Year:** the coefficient for accident year was negative and statistically significant ($p < 0.001$), indicating a modest but consistent decline in the likelihood of severe claims over time.
- **Weekday of Accident:** accidents on Saturdays were significantly more likely to result in severe claims, possibly due to reduced supervision, less experienced staff, or riskier weekend tasks. Other weekdays showed no significant effect, suggesting comparatively safer or better-controlled conditions during the regular work week.
- **Cause of Injury:** several causes were strongly predictive of severe claims. Injuries resulting from "commuting" (especially on two-wheel vehicles), "electrical currents", "explosives", and "falling" were all positively and significantly associated with the outcome. For example, the coefficient for "explosives" was 3.80 ($p < 0.001$), indicating a substantial increase in the log-odds of a serious claim when such a cause was present.
- **Body Part Affected:** injuries involving the "full body", "multiple body parts", and "lower extremities" were significantly associated with higher odds of severe claims. Conversely, injuries to the "chest", "teeth", and "face" showed negative associations, likely indicating that these injuries tend to be less severe or more recoverable.
- **Injury Type:** most injury types displayed negative associations with severe claims. For instance, "superficial wounds", "dislocations", and "internal injuries" were highly significant predictors of lower claim severity. These results suggest that certain injury types, despite their frequency, may not typically lead to claims classified as serious.
- **Work Classification:** clusters "CAE_4" and "CAE_5", representing sectors like education, healthcare, public administration, and hospitality, were associated with lower odds of severe claims, likely reflecting safer and more regulated work environments. In contrast, self-employed clusters "Job_2" and "Job_3", which include physically demanding

and higher-risk occupations such as welders, drivers, and agricultural workers, showed higher odds of severe claims, likely due to greater exposure to hazardous conditions.

- **Demographic Factors:** being male was significantly associated with higher odds of severe injury claims ($p < 0.001$). Age also showed a strong and consistent pattern: each successive age group exhibited increasing odds of severe claims, with workers aged 59 and older having the highest risk.

- **Salary and Gross Cost Bands:** the effects of salary bands were mixed and mostly not statistically significant, except for the 35k-60k band, which had a negative association with severe claims. Higher gross cost bands tended to be negatively associated with severe claims, which could reflect differences in the nature of the claims or the allocation of medical expenditures.

4.2 The severity of claims: a two-part approach

We initially intended to fit a Generalized Linear Model (GLM) to the entire portfolio of claim severities. While exploratory analysis identified the lognormal distribution as the best-fitting candidate among several considered, it still failed to adequately capture the behaviour of the distribution's tail. In particular, the lognormal fit underestimated the frequency and magnitude of extreme losses, leading to concerns about model reliability in the upper quantiles.

To accurately capture the distinct behaviours present in the distribution of claim severities, we adopt a two-part modelling approach that separately estimates moderate and extreme losses. This decision was motivated by both statistical and operational considerations.

First, exploratory analysis of the claim amount distribution in Section 3.3.2 revealed a strong positive skew. This indicates the presence of a heavy right tail, suggesting that extreme values could exert disproportionate influence on standard statistical models, potentially leading to biased estimates if modelled jointly with lower-valued claims. Secondly, we incorporated practical constraints tied to the reinsurance structure of the portfolio. Specifically, any claim exceeding EUR $200{,}000$ is reported to the reinsurer, establishing a natural and relevant operational threshold. This reporting requirement, aligned with the statistical properties of the data, provides a meaningful cut-off point for differentiating between attritional and extreme losses.

Based on these factors, we define two claim severity components:

- X_1: Claims lower than EUR $200{,}000$, representing the bulk of the portfolio. These are modelled using a Generalized Linear Model (GLM) with a lognormal specification.

- X_2: Claims equal to or greater than EUR $200{,}000$, which constitute the tail of the distribution. These are modelled using a Generalized Pareto Distribution (GPD), in accordance with extreme value theory.

This modelling framework allows us to tailor estimation techniques to the specific characteristics of each subset of data, improving both model accuracy and interpretability.

4.2.1 A lognormal GLM application to attritional claims (X_1)

Prior to fitting a Generalized Linear Model (GLM) to the subset of claims below EUR $200{,}000$, it was necessary to identify an appropriate distributional form for modelling the severity of these attrition claims. Several candidate distributions were evaluated using empirical data from the portfolio, including the lognormal, gamma, Weibull, and Pareto distributions.

Let X_1 denote the random variable representing claim amounts less than EUR $200{,}000$. The following procedure was used to compare potential distributions:

1. The lognormal, gamma, and Weibull distributions were fitted using maximum likelihood estimation.
2. For the Pareto distribution, parameters were initialised with a scale parameter equal to the minimum observed claim, and the shape parameter was estimated via the method of moments.
3. The goodness of fit was evaluated using Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling statistics, as well as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

The results of the fit evaluation are summarised in Table V.

TABLE V: Goodness-of-Fit Statistics for Candidate Distributions for Attritional Claims (X_1)

Statistic	Lognormal	Gamma	Pareto	Weibull
KolmogorovSmirnov	0.061	0.151	0.123	0.117
Cramérvon Mises	12.75	82.76	41.71	62.65
AndersonDarling	82.04	446.21	257.58	377.45
AIC	193069.6	196682.7	194023.8	195937.1
BIC	193083.9	196697.1	194038.1	195951.5

Based on the combination of goodness-of-fit statistics and information criteria, the lognormal distribution clearly outperforms the alternatives. It exhibited the lowest AIC and BIC values, as well as superior performance across all empirical fit tests, reinforcing our findings from the analysis of Figure 2.

Having identified the lognormal distribution as a suitable basis for modelling X_1, we fit a GLM to estimate expected claim values. The GLM employed a Gaussian distribution for the residuals and an identity link applied to the natural logarithm of the claim amount.

An initial full model was constructed using all available covariates, including demographic, contextual, and claim-related features. However, to improve model interpretability and reduce multicollinearity, several simplified variants were tested. Weekday was preemptively excluded, as exploratory plots showed no clear variation in claim amounts by day of the week (Figure 12). Additionally, Bodypart_count, Injury_count, and Gender were considered for removal based on moderate pairwise correlations with other predictors (e.g., Injury_count and Bodypart_count: $r = 0.42$; Gender and Work: $r = 0.38$).

To formally assess the importance of these variables, we conducted likelihood ratio tests comparing the full model to reduced models omitting each of them (Harrell, 2015). Table VI summarizes the deviance comparisons. Likelihood ratio tests (LRTs) were used to evaluate whether the removal of each variable led to a significant loss in model fit. A significant p-value indicates that the excluded variable contributes meaningfully to the model.

TABLE VI: Likelihood Ratio Tests for Candidate Variable Exclusion

Model Comparison	Deviance (χ^2)	p-value
Full vs. –Injury_count	11.45	0.026
Full vs. –Bodypart_count	4.02	0.591
Full vs. –Gender	93.67	$< 2.2 \times 10^{-16}$

The test results indicate that removing Injury_count results in a modest but statistically significant decline in model fit ($p = 0.026$), suggesting it should be retained. In contrast, exclusion of Bodypart_count did not significantly affect fit ($p = 0.591$), supporting its removal. The effect of removing Gender was highly significant ($p < 2.2 \times 10^{-16}$), confirming its importance in the model.

Based on these results, the final model excludes both Weekday and Bodypart_count, preserving variables with demonstrable impact while improving parsimony. This specification was also consistent with the results of a stepwise selection procedure, which independently retained a similar set of covariates.

The final selected model includes the following predictors:

```
lognormal_glm_minusbp <- glm(log(Claim_amount) ~ Accident_year +
Cause + Bodypart + Injury + Injury_count + Work + Gender + Age_band
+ Salary_band + Gross_cost_band, data = claims_under, family =
gaussian(link = "identity"))
```

Additionally, we evaluated multicollinearity among the predictors in the reduced model using Variance Inflation Factors (VIF) (James et al., 2021). All VIF values fell below the standard threshold of 5, with the highest generalized VIF ($\text{GVIF}^{1/(2*\text{Df})}$) observed for Injury

(1.10), thus not indicating multicollinearity concerns in this model specification.

The estimated coefficients suggest that some variables, such as Accident_year (positive effect, $p < 0.001$), significantly influence claim severity. Among the causes of accidents, "explosives" and "road accident or vehicles" showed significant positive effects, while others, such as "commuting" and "falling", were mostly not significant.

Classifications such as "brain", "eyes", "full body", and "multiple parts" had significant positive coefficients, indicating higher claim amounts, while regions like "feet and toes", "hands and fingers", and "lower extremities" were associated with significantly lower claim amounts. Injuries such as "concussions" and "internal injuries", "dislocations, sprains and strains", and "shocks" had significant negative effects, indicating lower expected claim severity compared to the baseline.

The Gender variable showed males tend to have higher claim amounts ($p < 0.001$). Higher salary and gross cost bands corresponded with increasing claim severity for values below EUR $200{,}000$. This trend aligns with the expected economic relationship, as claim payouts are based on a percentage of the claimant's annual income, as detailed in Section 2.2.

Overall, the model identifies meaningful predictors of attritional claim amounts, supporting its use for estimating expected losses below the $200k$ threshold. This model was used to predict the logarithm of claim amounts for the bulk of the portfolio, and the outputs were exponentiated to recover predictions in the original monetary scale.

4.2.2 A GDP application to extreme claims (X_2)

In this section, we focus on modelling the extreme values of claim costs, specifically those exceeding a high threshold. Accurately estimating the tail behaviour of the claim distribution is crucial for risk management and pricing strategies. To achieve this, we employ the Generalized Pareto Distribution (GPD), a standard tool in extreme value theory for modelling excess values over a threshold.

The use of the GPD is theoretically justified by the Pickands-Balkema-de Haan theorem (Pickands (1975), Balkema and de Haan (1974)), which states that for a wide class of underlying distributions, the distribution of excesses over a sufficiently high threshold converges to a GPD as the threshold increases. This result provides a strong foundation for approximating the tail of the claim amount distribution with the GPD.

Following this theoretical basis, we proceed with the empirical fitting of the GPD to our claim data exceeding the chosen threshold of EUR $200{,}000$, using the `R` package `evir`. For a random variable Y representing the excess over a threshold u, the cumulative distribution function (CDF) of the GPD is:

$$F(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\beta}\right)^{-1/\xi}, & \text{if } \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\beta}\right), & \text{if } \xi = 0 \end{cases} \quad \text{for } y > 0,\ 1 + \frac{\xi y}{\beta} > 0 \tag{8}$$

where:

- ξ is the shape parameter (tail index), which determines the heaviness of the tail.
- β is the scale parameter, controlling the dispersion of excesses over the threshold.
- $y = x - u$ is the excess over the threshold u.

The fitted Generalized Pareto Distribution (GPD) has a shape parameter $\hat{\xi} = 0.0952$ and a scale parameter $\hat{\beta} = 106{,}002.17$. These estimated using maximum likelihood estimation based on exceedances over the threshold. The positive shape parameter indicates a heavy tail, though not extremely so, consistent with $\xi > 0$ and finite upper support. The scale parameter $\hat{\beta}$ reflects the typical magnitude of excesses above the threshold, functioning as a spread parameter: higher β values correspond to more widely dispersed extreme outcomes. The results suggest that, on average, claims tend to exceed the 200k threshold by approximately 100k.

We then simulated synthetic excesses based on the fitted GPD parameters, which enables us to compare the observed values with the model predictions. This approach allows us to evaluate the goodness of fit and understand the tail risk in our claims data.

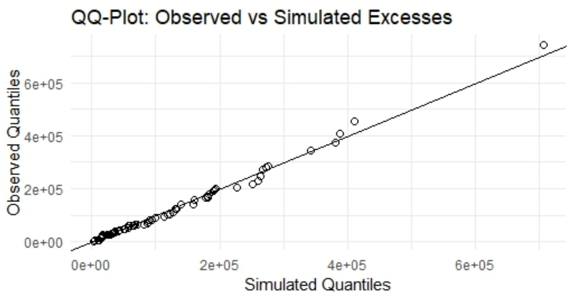


FIGURE 3: QQ-plot comparing observed excesses to simulated GPD excesses.

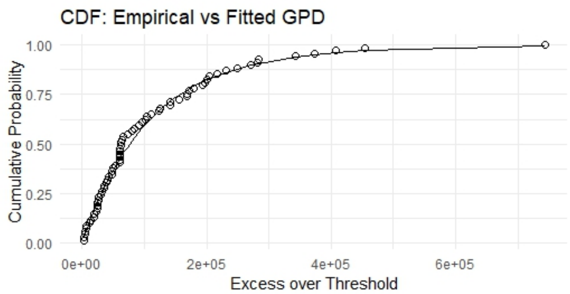


FIGURE 4: Empirical vs fitted GPD CDF.

Figures 3 and 4 present visual diagnostics of the GPD fit to the claim excesses. The QQ-plot shows that the observed exceedances align closely with the simulated GPD quantiles, demonstrating that the tail behaviour of the data is well captured by the model. Similarly, the empirical and fitted cumulative distribution functions closely match, confirming the suitability of the GPD for modelling the extreme claims.

4.2.3 *Combining the two-part model for claim amounts*

After fitting a Generalized Linear Model (GLM) to the claims below the threshold of EUR $200{,}000$ and a Generalized Pareto Distribution (GPD) to the excesses above this threshold, we constructed a unified two-part model to estimate the entire claims distribution.

To evaluate the predictive performance of this combined two-part model, we employed the Gini coefficient, a widely used measure in actuarial science to assess the discriminatory power of prediction models (Frees, Meyers, and Cummings, 2014). A higher Gini coefficient indicates a better ranking of claims according to their magnitude, even if the exact values are not perfectly predicted.

Our result of 0.791 indicates a solid predictive performance of the two-part model, demonstrating that it ranks claim severities effectively and is particularly suitable for datasets with heavy-tailed characteristics.

To represent the full range of potential claim amounts, we construct the expected value of a claim X as a combination of two sub-models: a lognormal GLM for moderate claims (i.e., $X \leq u$), and a Generalized Pareto Distribution (GPD) for extreme claims (i.e., $X > u$), where $u =$ EUR 200,000 is the threshold.

Let p denote the empirical probability that a claim falls below the threshold u, i.e.,

$$p = \mathbb{P}(X \leq u), \quad \text{and} \quad 1 - p = \mathbb{P}(X > u).$$

Then, the combined expected value of a claim X can be written as:

$$\mathbb{E}[X] = p \cdot \mathbb{E}[X \mid X \leq u] + (1 - p) \cdot \mathbb{E}[X \mid X > u]. \tag{9}$$

Here:

- $\mathbb{E}[X \mid X \leq u]$ is the expected claim amount estimated using the lognormal GLM,
- $\mathbb{E}[X \mid X > u] = u + \frac{\beta}{1-\xi}$, assuming $\xi < 1$, is the mean of the GPD excess distribution above the threshold,
- β and ξ are the scale and shape parameters of the fitted GPD.

The probability p is estimated empirically based on historical data as the proportion of claims below the threshold.

4.3 *Aggregate losses, S*

The aggregate loss S over a given period is defined as the total amount of all claims:

$$S = \sum_{i=1}^{N} X_i,$$

where N is the number of claims (frequency), and X_i are the individual claim severities (severity).

Assuming independence between N and X_i, the expected aggregate loss is given by the classical formula (Klugman, Panjer, and Willmot, 2008):

$$\mathbb{E}[S] = \mathbb{E}[N] \times \mathbb{E}[X].$$

In our case, the claim frequency N is modeled using a binomial GLM, which estimates the expected number of claims as

$$\mathbb{E}[N] = m \times \hat{p},$$

where m is the number of exposure units (e.g., claimants with simple claim that might still have a right to receive a pension) and $\hat{p}$ is the estimated claim probability.

The claim severity X follows the two-part model detailed in (4.2.3), with an assumption of $p = 0.993$, based on historical information.

Thus, the expected aggregate loss is

$$\mathbb{E}[S] = m \times \hat{p} \times (p \times \mathbb{E}[X_1] + (1 - p) \times \mathbb{E}[X_2]).$$

For illustrative purposes, the moments of the aggregate loss for 2024 are in Table VII:

TABLE VII: Aggregate Loss Moments for 2024 (Illustrative)

Moment	Value
Expected Aggregate Loss, $\mathbb{E}[S]$	8,667,302
Variance of Aggregate Loss, $\mathrm{Var}(S)$	8.77×10^{11}
Third Central Moment, $\mu_3(S)$	1.35×10^{17}

5 ANALYSIS OF THE REINSURANCE TREATY

Having provided a conceptual overview of excess of loss (XL) reinsurance in Section 2.3, we now turn our attention to its practical implementation and financial implications within the context of our claim amounts data. In this chapter, we quantify the impact of XL reinsurance structures on aggregate losses and assess the effectiveness of various attachment points.

We consider a typical XL treaty covering the layer L xs M, meaning the reinsurer covers losses exceeding a retention level M up to a limit L. For an individual claim X, the amount ceded to the reinsurer is:

$$Z(M,L) = \min(L, (X-M)_+) = \begin{cases} 0, & \text{if } X \leq M, \\ X-M, & \text{if } M < X \leq M+L, \\ L, & \text{if } X > M+L, \end{cases}$$

where $(x)_+ = \max(0, x)$.

This function captures the reinsured loss, which is zero when the claim is below the retention, grows linearly within the layer, and is capped at L for large losses.

The insurer retains the remaining portion of the loss, defined as:

$$Y(M,L) = X - Z(M,L) = \begin{cases} X, & \text{if } X \leq M, \\ M, & \text{if } M < X \leq M+L, \\ X-L, & \text{if } X > M+L. \end{cases}$$

This formula reflects the net retained claim, where the insurer pays the full amount for small claims, a fixed retention M within the reinsured layer, and an increasing share again for extremely large claims above the covered layer.

The distributions of these components can be derived from the original claim distribution $F_X(x)$. The cumulative distribution function (CDF) of the retained loss is:

$$F_{Y(M,L)}(x) = \begin{cases} F_X(x), & \text{if } x < M, \\ F_X(x+L), & \text{if } x \geq M. \end{cases}$$

Similarly, the CDF of the ceded (reinsured) amount is:

$$F_{Z(M,L)}(x) = \begin{cases} F_X(x+M), & \text{if } 0 \leq x < L, \\ 1, & \text{if } x \geq L. \end{cases}$$

These expressions help us evaluate how XL reinsurance reshapes the distribution of losses faced by the insurer and reinsurer, respectively.

5.1 Evaluating treaty effectiveness: loss elimination ratio

A common metric for evaluating the effectiveness of an excess of loss treaty is the Loss Elimination Ratio (LER) (Klugman et al., 2008). The LER quantifies the proportion of gross losses that are eliminated by the reinsurance arrangement, and is defined as:

$$LER = 1 - \frac{Retained\ Loss}{Total\ Loss}$$

A low LER suggests that the reinsurance treaty is triggered infrequently, typically due to a high attachment point. Conversely, a high LER implies significant loss transfer, which may reduce the insurer's retained risk but generally comes at a higher premium cost.

The reinsurance treaty applied in this analysis features an attachment point of $M = 400,000$ and a limit of $L = 39,600,000$. After applying this structure to both observed and model-estimated claims, the Loss Elimination Ratio (LER) was calculated to assess the proportion of losses transferred to the reinsurer.

The results indicate that only 1.44% of total observed losses are eliminated by the treaty. For model-estimated claims, the LER is slightly higher at 1.45%. This marginal difference between the observed and estimated LER further suggests that the model aligns reasonably well with historical experience. Most importantly, the LER values are relatively low, reflecting that the vast majority of claims fall below the attachment point and are therefore retained by the insurer. This implies that the reinsurance treaty, in its current form, has limited impact on overall loss transfer, providing protection for only a small portion of the portfolio.

5.2 Sensitivity analysis of reinsurance attachment points

Now that we have established the mechanics and impact of the current XL reinsurance treaty, we can shift our focus to evaluating whether its structure could be refined to better serve the needs of the insurer. In particular, we revisit the attachment point (the threshold above which claims are ceded to the reinsurer) and consider whether alternative values might yield more favourable outcomes.

The LER results suggest that the current attachment point may be set too high to provide meaningful risk relief. This observation motivated us to conduct a sensitivity analysis, where we vary the attachment point and evaluate the resulting impact on treaty effectiveness.

The aim of this exploration is to determine whether adjusting the threshold could offer:

- Improved efficiency in transferring risk to the reinsurer,

- Greater reduction in the volatility of retained losses, and
- A more optimal balance between the cost of premiums and the acquired protection.

Figure 5 illustrates the Loss Elimination Ratio (LER) calculated at various attachment points for both the observed claim data and the model-based estimated claims. As expected, the graph demonstrates a monotonically-decreasing relationship between LER and the retention limit. As the cedent's priority increases, the LER declines rapidly at first, and then gradually flattens. This behaviour reflects the diminishing marginal impact of increasing deductibles: at low levels, small increases in the attachment point eliminate a substantial portion of losses, whereas at higher levels, further increases result in only marginal loss elimination.

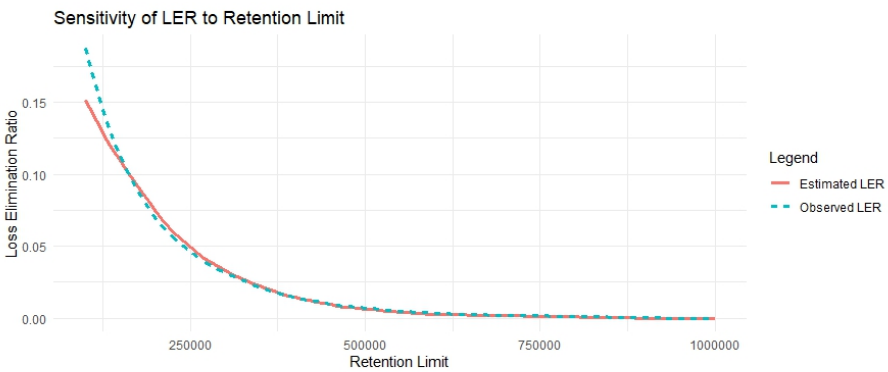


FIGURE 5: Loss Elimination Ratio (LER) as a function of the attachment point for observed and estimated claims.

Moreover, both curves exhibit similar shapes, indicating that the model captures the general trend of the LER well. However, slight deviations between the curves, particularly at lower attachment points, may indicate model limitations in capturing a portion of the loss distribution.

To better understand the insurer's exposure before incorporating reinsurance premium costs, we next examine the statistical properties of retained losses across varying attachment points. In particular, we present plots of:

- The expected value $\mathbb{E}[Y(M, L)]$ of the insurer's retained loss for each candidate attachment point.
- The standard deviation $\mathrm{SD}[Y(M, L)] = \sqrt{\mathrm{Var}[Y(M, L)]}$, reflecting the variability in retained losses across attachment points.

These visualizations provide an understanding of how the insurer's loss exposure changes with treaty design, highlighting the balance between risk retention and transfer before considering the cost of reinsurance.

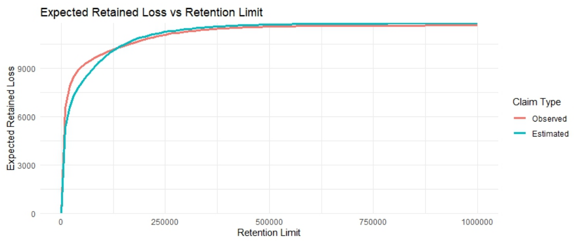


FIGURE 6: Expected Value of the Retained Loss

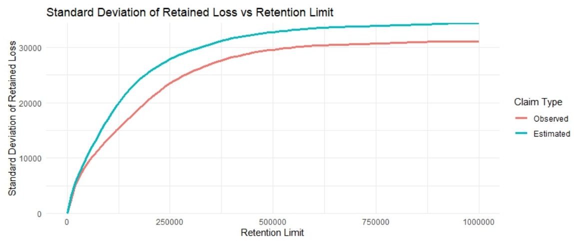


FIGURE 7: Standard Deviation of the Retained Loss

Figures 6 and 7 help illustrate how both the expected value and the variability of retained losses evolve as the retention limit increases. In Figure 6, we see a sharp initial rise in expected retained losses, which gradually levels off at higher retention levels. This pattern makes sense when we consider the distribution of claim amounts discussed earlier in Section 3.3.2, with over 80% of claims falling below the portfolio mean. In other words, most claims are relatively small, so increasing the retention limit in the lower range has a significant effect, as it exposes the insurer to a large share of the total claim volume. Beyond a certain point, however, further increases in retention mostly affect less frequent, high-cost claims, which have a much smaller impact on the overall expected loss.

A similar trend is evident in Figure 7, which shows how the standard deviation of retained losses changes with retention. The initial rise reflects the increased volatility that comes with covering a broader portion of the claim distribution. But just like the expected value, the standard deviation starts to level off as the retention limit grows. This suggests that the insurer's exposure to volatility does not increase much beyond a certain threshold, reinforcing the idea that most of the risk is concentrated in the more frequent, lower-severity claims. However, unlike the expected value, the standard deviation curves for observed and estimated retained losses do not align as closely. In fact, the estimated retained losses tend to exhibit a higher standard deviation across all retention levels. This suggests that while the model captures the average loss well, it may be overestimating the volatility in retained outcomes, potentially due to assumptions made in the frequency-severity modelling or limitations in the underlying data.

Those figures highlight how important it is to choose a retention limit that carefully balances risk exposure and reinsurance cost efficiency in the design of the reinsurance treaty.

To evaluate how the choice of attachment point influences the insurer's financial outcomes, we carried out a sensitivity analysis using two widely adopted premium principles for the reinsurance premium: the Expected Value (EV) Principle and the Standard Deviation (SD) Principle.

For each candidate attachment point within a predefined range, we calculated the total cost to the insurer, defined as the sum of expected retained losses and the reinsurance premium. The process involved the following steps:

1. **Model Retained Loss:** for each claim, we applied an excess-of-loss structure with attachment point M and limit L to determine the retained loss:

$$Y(M, L) = \min(\text{Loss}, M) + \max(0, \text{Loss} - M - L)$$

This was calculated using both observed and simulated claim data.

2. **Calibrate Safety Loadings:** we estimated the safety loading coefficients α_{EV} and α_{SD} using historical reinsurance premium data from the existing excess-of-loss treaty with an attachment point of EUR $400{,}000$. The values were calculated as follows:

$$\alpha_{\text{EV}} = \left(\frac{P_{\text{obs}}}{\mathbb{E}[Z]}\right) - 1, \quad \alpha_{\text{SD}} = \frac{P_{\text{obs}} - \mathbb{E}[Z]}{\text{SD}(Z)}$$

where P_{obs} is the observed (historical) ceded premium actually paid to the reinsurer, and Z is the ceded loss.

3. **Calculate Ceded Premium:** the calibrated values of α_{EV} and α_{SD} are then held constant and applied uniformly across all alternative retention levels in our ceded premium estimation.

   - *Expected Value Principle:*

$$\Pi_{\text{EV}} = (1 + \alpha_{\text{EV}}) \cdot \mathbb{E}[Z]$$

   - *Standard Deviation Principle:*

$$\Pi_{\text{SD}} = \mathbb{E}[Z] + \alpha_{\text{SD}} \cdot \text{SD}(Z)$$

   where Z is the ceded loss.

4. **Calculate Total Expected Retained Cost:** for each candidate retention level, we computed the insurer's total expected cost:

$$\text{Expected Retained Cost} = \mathbb{E}[Y(M, L)] + \Pi$$

   where Π corresponds to either Π_{EV} or Π_{SD} depending on the premium principle in use.

This framework enabled us to explore how expected costs evolve under different retention strategies, guided by empirical loading coefficients derived from real treaty pricing. Although it assumes a uniform loading structure, potentially overlooking how reinsurers adjust loadings by layer, it offers a practical and tractable basis for premium comparison.

By comparing the total retained costs across different retention limits and premium principles, we gained insights into the trade-offs between risk retention and cost efficiency. This analysis enabled us to identify treaty structures that optimise the insurer's overall financial outcome. The results are presented graphically to illustrate how sensitive the insurer's costs are to various treaty design choices.

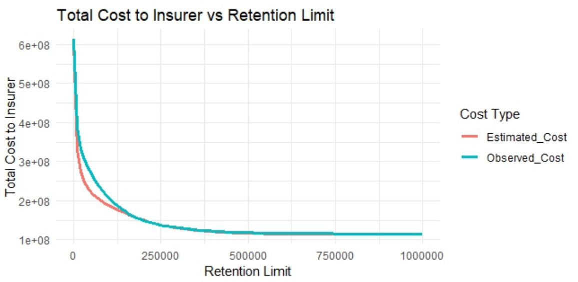


FIGURE 8: Estimated Total Cost for the Insurer vs Retention Limit: Expected Value Premium Principle

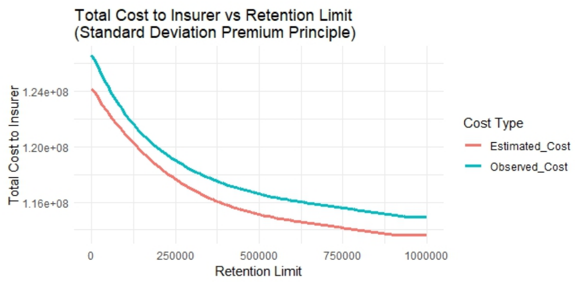


FIGURE 9: Estimated Total Cost for the Insurer vs Retention Limit: Standard Deviation Premium Principle

Figures 8 and 9 show the total cost to the insurer under the Expected Value and Standard Deviation Premium Principles, respectively, as functions of the retention limit. For comparability, we used the same calibrated loading coefficient α_{SD} in both cases, as it reflects realistic pricing observed in the historical excess-of-loss treaty and allows for a consistent evaluation of the impact of the premium principle alone.

In both cases, the total cost decreases with higher retention levels, reflecting reduced reliance on reinsurance. This pattern is robust: we obtained similar qualitative results (i.e., decreasing cost curves) when using not only the calibrated loading coefficients α_{EV} and α_{SD} (shown in Table VIII), but also with arbitrary values such as $\alpha = 0.2$ and $\alpha = 0.5$. This is expected, as the ceded premium always exceeds the expected ceded loss when a positive loading is applied.

TABLE VIII: Estimated loading coefficients from historical premium data

Premium Principle	**Loading α**
Expected Value Principle	$\alpha_{\text{EV}} = 4.27$
Standard Deviation Principle	$\alpha_{\text{SD}} = 3.38$

These high loading values highlight how expensive excess-of-loss (XL) reinsurance can be in practice, which aligns with real-world market behaviour.

In Figure 8, under the Expected Value Principle, the sharp drop in cost at low retention levels reflects this high premium load: even modest risk transfers come at a steep price. As the insurer retains more risk, reinsurance becomes less necessary and premium costs fall, eventually flattening as nearly all risk is retained.

In Figure 9, the Standard Deviation Principle produces a similar shape but with a more gradual decline. This is due to the volatility loading: even as premium falls with higher retention, the total cost remains elevated due to the added risk charge. Particularly, the gap between model estimates and observed data is more pronounced here, suggesting the model may overstate the volatility component, impacting pricing accuracy.

Taken together, the results illustrate the fundamental trade-offs in reinsurance design. Lower retention levels reduce volatility in retained losses but at high premium costs, while higher retentions lower premiums but increase risk exposure. The choice of premium principle is also highly relevant, shaping not just the cost level but also how sharply those costs fall with increased retention. The Expected Value Principle, by ignoring volatility, naturally leans toward more aggressive retention strategies. In contrast, the Standard Deviation Principle incorporates a risk loading and therefore favours more cautious approaches that reduce exposure to large claim fluctuations.

As noted in the literature review (Section 2.3.3), there is no universally optimal retention strategy. The best design depends on what the insurer aims to prioritise: minimising cost, reducing volatility, or balancing the two. These analyses provide a practical framework for evaluating how different structures perform under various assumptions and help insurers align treaty design with their specific risk appetite and strategic goals.

6 CONCLUSION

This analysis of the workers' compensation portfolio provided several useful insights, despite some limitations. A notable portion of the dataset fell into vague categories like "Other" or "Unknown," which made it difficult to extract reliable patterns, particularly when fitting generalized linear models (GLMs) to estimate claim frequency and severity for losses under EUR $200{,}000$. These classification issues reduced the precision of the model and limited the conclusions we could draw for lower-severity claims.

For larger losses (above EUR $200{,}000$) we used a Generalized Pareto Distribution (GPD) to capture tail risk. While the fit was generally reasonable, the model's predictive power was constrained, likely due to both limited data volume and the inherent difficulty of modelling rare, high-impact events.

When evaluating the current excess-of-loss treaty, we found that the Loss Elimination Ratio (LER) was fairly low, at just 1.44% based on observed data and 1.45% in simulated scenarios. This suggests that the treaty does relatively little to reduce retained losses and may not be cost-effective in its current form. We therefore tested alternative retention levels and examined their impact using both expected value and standard deviation reinsurance premium principles.

The analysis highlighted a key trade-off: lower retentions reduce loss volatility but significantly increase reinsurance costs, especially under a high loading coefficient (α) as estimated in our pricing model. On the other hand, higher retention reduces premiums but exposes the insurer to more uncertainty. We also observed that while the expected value estimates for retained losses aligned fairly well with observed outcomes, the model tended to overstate volatility, which in turn affected cost projections under the standard deviation principle.

As discussed in the literature review (Section 2.3.3), there is no single optimal retention level, it ultimately depends on the insurer's priorities. If minimising volatility is a key concern, a more conservative retention strategy may be warranted, despite the higher cost. If cost efficiency is the primary goal, a higher retention may be more suitable. Our analysis provides a practical framework to support this decision-making process, grounded in both observed data and model-based projections.

In summary, while data availability posed some challenges, the project still allowed for a meaningful assessment of the current reinsurance structure and pointed toward more tailored, informed strategies going forward. Looking ahead, future research could focus on improving data granularity, particularly around claim classification, and on refining the modelling of high-severity losses. Exploring alternative risk transfer mechanisms, such as mixing proportional and non-proportional reinsurance arrangements, or the inclusion of other coverages (such as LTA or OB) in the treaty could also offer new angles for optimising reinsurance in this line of business.

REFERENCES

Alcaide, H. d. C. (2022). An optimal deductible evaluation under an excess of loss reinsurance treaty for an automobile insurance portfolio. *Master's Thesis. NOVA Information Management School*.

Anderson, J. F. (1999). Commutation functions. *Education and Examination Committee of the Society of Actuaries*.

Baione, F., Angelis, P. D., Menzietti, M., & Tripodi, A. (2017). A comparison of risk transfer strategies for a portfolio of life annuities based on rorac. *Journal of Applied Statistics*, *44*(10), 1875–1892.

Balkema, A. A., & de Haan, L. (1974). Residual life time at great age. *Annals of Probability*, *2*(5), 792–804.

Bazyari, A. (2025). Optimal excess-of-loss reinsurance contract in a dynamic risk model. *Statistics, Optimization Information Computing*, *13*(4), 1480-1504.

Bear, R. A., & Nemlick, K. J. (1990). Pricing the impact of adjustable features and loss snaring provisions of excess-of-loss treaties. *Proceedings of the Casualty Actuarial Society*, *LXXVII*, 60-123.

Blumsohn, G. (1997). Levels of determinism in workers' compensation reinsurance commutations. *Casualty Actuarial Society E-Forum*, *Spring*, 53-114.

Borch, K. (1960). An attempt to determine the optimum amount of stop loss reinsurance reinsurance. *Transactions of the 16th International Congress of Actuaries*, *2*, 597-610.

Cai, J., Tan, K. S., & Weng, C. (2007). Optimal retention for a stop-loss reinsurance under the var and cte risk measures. *ASTIN Bulletin: The Journal of the IAA*, *37*(1), 93–112.

Centeno, M. d. L., & Simões, O. (1991). Combining quota-share and excess of loss treaties on the reinsurance of n independent risks. *ASTIN Bulletin*, *21*, 41-55.

Centeno, M. d. L., & Simões, O. (2009). Optimal reinsurance. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A: Matemáticas (RACSAM)*, *103*(2), 387-405.

Cheung, K. C. (2010). Optimal reinsurance revisiteda geometric approach. *ASTIN Bulletin: The Journal of the IAA*, *40*(1), 221–239.

Cramér, H. (1946). *Mathematical methods of statistics (volume 9)*. Princeton, NJ: Princeton University Press.

Cullen, A. C., & Frey, H. C. (1999). *Probabilistic techniques in exposure assessment*. Springer.

Cummins, J. D., Dionne, G., Gagné, R., & Nouira, A. (2021). The costs and benefits of reinsurance. *Geneva Papers on Risk and Insurance Issues and Practice*, *46*, 177-199.

Código do trabalho. (2009). *Assembleia da República. Portugal*.

De Angelis, D. N. M. G., P. (2011). Optimal reinsurance programs for a portfolio of life annnuities. *Insurance Market and Companies: Analyses and actuarial computations*, *2*, 31-42.

Decreto-lei no. 142/99, de 30 de abril. (1999). *Assembleia da República. Portugal*.

Decreto-lei no. 352/2007, de 23 de outubro. (2007). *Assembleia da República. Portugal*.

Decreto-lei no. 9/2025, de 12 de fevereiro. (2025). *Assembleia da República. Portugal*.

Draper, E., Ladou, J., & Tennenhouse, D. J. (2011). Occupational health nursing and the quest for professional authority. *Iowa Orthop J.*, *21*(1), 57-88.

Duncan, I. (2020). Using survival analysis to predict workers compensation termination. *Variance*, *13*(1), 31-53.

Dunn, P., & Smyth, G. (2018). *Generalized linear models with examples in r*. Springer.

Fahrmeir, L., Kneib, T., Lang, S., & Marx, B. (2013). *Regression: Models, methods and applications*. Springer.

Ferguson, R. E. (1971). Actuarial note on workmen's compensation loss reserves. *Proceedings of the Casualty Actuarial Society*, *LVIII*, 51-58.

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

Friedland, J. (2023). Exam 7 - reserving for reinsurance. *CAS Study Note*.

Froot, K. A. (2001). The market for catastrophe risk: a clinical examination. *Journal of Financial Economics*, *60*, 529-571.

Fu, L., & Liu, X. (2016). Finite mixture model and workers compensation large loss regression analysis. *in Edward W. Frees, Glenn Meyers, and Richard A. Derrig, eds. Predictive Modeling Applicationsin Actuarial Science*, *II*, 224-260.

Ghaddab, S., Kacem, M., de Peretti, C., & Belkacem, L. (2023). Extreme severity modeling using a glm-gpd combination: application to an excess of loss reinsurance treaty. *Empirical Economics*, *65*, 1105-1127.

Goldburd, M., Khare, A., Tevet, D., & Guller, D. (2025). Generalized linear models for insurance rating. *Casualty Actuarial Society*, *Second Edition (2025 revision)*.

Guyton, G. P. (1999). A brief history of workers' compensation. *Iowa Orthop J.*, *19*, 106-110.

Harrell, F. E. (2015). *Regression modeling strategies*. Springer.

Henry, J. B., & Hsieh, P.-H. (2009). Extreme value analysis for partitioned insurance losses. *Variance*, *3*(2), 214-238.

Hutter, H. E., & Graves, G. T. (1992). Reserving for special provisions in reinsurance contracts. In *Casualty loss reserve seminar.*

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning with applications in r*. Springer.

Jones, B. A., Scukas, C. J., Frerman, K. S., Holt, M. S., & Fendley, V. A. (2013). A mortality-based approach to reserving for lifetime workers compensation claims. *Casualty Actuarial Society E-Forum*, *Fall*(1), 1-26.

Kaluszka, M., & Okolewski, A. (2008). An extension of arrow's result on optimal reinsurance contract. *The Journal of Risk and Insurance*, *75*(2), 275–288.

Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2008). *Loss models: From data to decisions*. Wiley Series in Probability and Statistics.

Laudagé, C., Desmettre, S., & Wenzel, J. (2019). Severity modeling of extreme insurance claims for tariffication. *Insurance: Mathematics and Economics*, *88*, 77-92.

Lei no. 83, de 24 de julho. (1913). *Assembleia da República. Portugal*.

Lei no. 98/2009, de 04 de setembro. (2009). *Assembleia da República. Portugal*.

Lewis, R. (2012). Employers' liability and worker's compensation: England and wales. *Berlin/Boston: De Gruyter*, 137-202.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values.

Journal of the American Statistical Association, *83*(404), 1198-1202.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. John Wiley Sons, Inc.

Marcarelli, G., & D'Ortona, N. (2012, 01). Insurers optimal reinsurance programs. *DYNAMICS OF SOCIO- ECONOMIC SYSTEMS*, *3*, 1-18.

McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools*. Princeton Series in Finance.

Milicia, J., & Dolsky, R. (2022). Ceded reserving - its not as simple as subtraction. In *2022 casualty loss reserving seminar.*

MunichRe. (2022). Types of reinsurance: Non-proportional. Retrieved from `https://www.munichre.com/content/dam/munichre/contentlounge/website-pieces/documents/NL-Non-Proportional_30-03-2023.pdf/_jcr_content/renditions/original./NL-Non-Proportional_30-03-2023.pdf`

NASDAQ. (2024). *Facultative vs. treaty reinsurance: What's the difference?* Retrieved from `https://www.nasdaq.com/articles/facultative-vs-treaty-reinsurance-whats-difference`

Ohlsson, E., & Johansson, B. (2010). *Non-life insurance pricing with generalized linear models*. Springer.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, *3*(1), 119–131.

Rosa, C. E. D. (2012). An internal model for workers compensation. *PhD Thesis. ISEG*.

Schmid, F. A. (2012). The workers compensation tails. *Variance*, *6*(1), 48-77.

Serrano, S. C. (2015). *Analysis of the reinsurance treaty for a workers compensation portfolio* (Unpublished master's thesis). ISEG.

Shane, M., & Morelli, D. (2013). Using life expectancy to inform the estimate of tail factors for workers compensation liabilities. *Casualty Actuarial Society E-Forum*, *XCII*, 579-678.

Sherman, R. E., & Diss, G. F. (2005). Estimating the workers compensation tail. *Proceedings of the Casualty Actuarial Society*, *XCII*, 579-678.

Steeneck, L. R. (1996). Actuarial note on workmen's compensation loss reserves - 25 years later. *Casualty Actuarial Society Forum*, 245-272.

Sun, Q., Rong, X., & Zhao, H. (2019). Optimal excess-of-loss reinsurance and investment problem for insurers with loss aversion. *Theoretical Economics Letters*, *9*(4), 1129-1151.

Vaughan, E., & Vaughan, T. (2013). *Fundamentals of risk and insurance, 11th edition: 11th edition*. John Wiley & Sons, Incorporated.

Veprauskaite, E., & Sherris, M. (2014). Reinsurance decisions in life insurance: An empirical test of the riskreturn criterion. *International Review of Financial Analysis*, *35*, 128-139.

Walters, D. (2007). An international comparison of occupational disease and injury compensation schemes. *A research report prepared for the industrial injuries advisory council (IIAC). Cardif University*.

Appendix

A Data

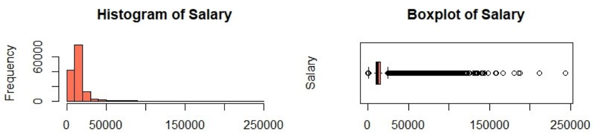
- **Severe claim:** a binary random variable that determines whether a severe claim was generated from the already-reported simple claim.
- **Accident year:** the year in which the event occurred.
- **Weekday:** day of the week in which the accident occurred.
- **Gender:** the gender identity of the worker, categorized as male or female.
- **Work:** occupation of the claimant. For employer-funded policies, they are categorized into professional groups as defined by the Portuguese Classification of Economic Activities (or CAE). For self-employed, they inform the actual profession of the worker themselves.
- **Cause:** the reported cause of the injury or accident (e.g., fall, collision, equipment-related).
- **Body part:** the primary specific body part(s) that were injured in the accident (e.g., arm, leg, head).
- **Body part count:** an integer describing the number of distinct body parts injured in a single incident.
- **Injury:** the nature or type of the primary injury suffered (e.g., fracture, sprain, laceration).
- **Injury count:** an integer describing total number of injuries sustained in a single event.
- **Age:** an integer representing the age of the claimant at the time of the accident, rounded to the nearest.
- **Salary:** a continuous variable indicating the income level of the claimant.
- **Gross cost:** a continuous variable representing the cost of the simple claim.
- **Claim amount:** a continuous variable representing the calculated mathematical reserve.

TABLE IX: Summary of Missing Values by Field for Base_N

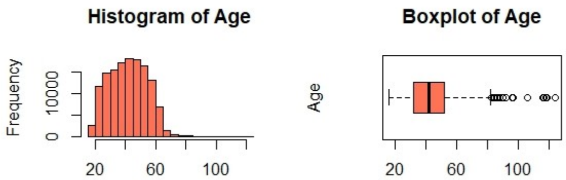
Field	Weekday	Accident_year	Salary	Severe claim	Injury_count	Bodypart_count	Bodypart
Missing	0.00%	0.00%	0.39%	0.00%	0.01%	0.01%	0.01%
	Cause	Injury	Age	Gender	Job/Activity	Gross_cost	Net_cost
Missing	0.03%	0.01%	1.79%	0.00%	0.01%	0.00%	0.00%

TABLE X: Summary of Missing Values by Field for Base_X

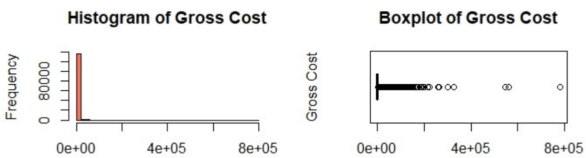
Field	Weekday	Accident_year	Salary	Injury_count	Bodypart_count	Bodypart	Cause
Missing	0.00%	0.00%	0.30%	0.00%	0.00%	0.00%	0.01%
	Injury	Age	Gender	Job/Activity	Gross_cost	Net_cost	Claim_amount
Missing	0.00%	0.03%	0.00%	0.02%	0.00%	0.00%	0.00%



(a) Distribution of Salaries



(b) Distribution of Ages

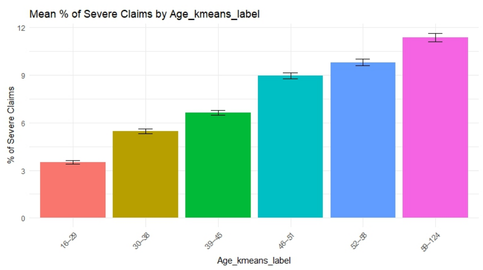


(c) Distribution of Gross Cost of Simple Claims

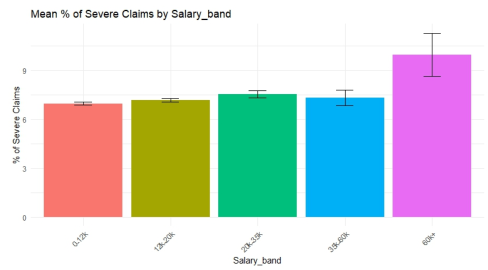
FIGURE 10: Distribution of the primary continuous variables (Age, Salary, and the Gross Cost of Simple Claims)

TABLE XI: Little's MCAR Test for Base_X

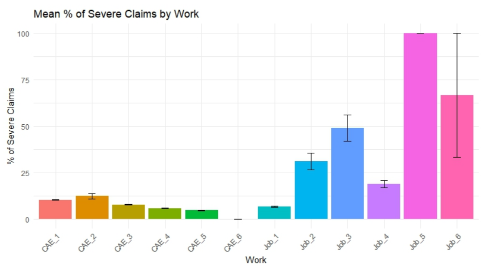
Statistic	Degrees of Freedom	p-value	Missing Patterns
1067.0	115	0	11



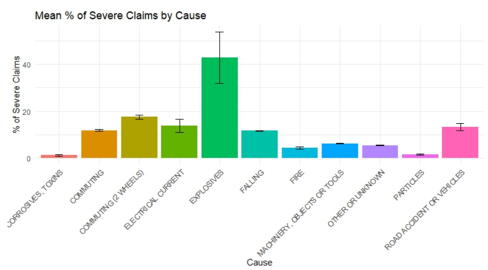
(a) Probability of severe claim by age



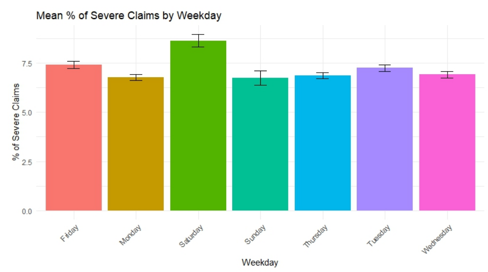
(b) Probability of severe claim by salary



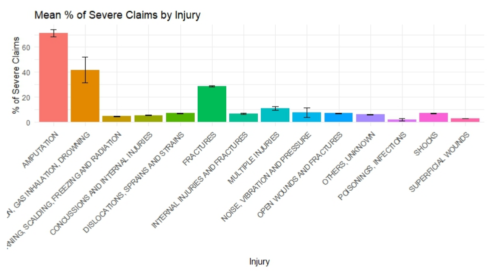
(c) Probability of severe claim by work cluster



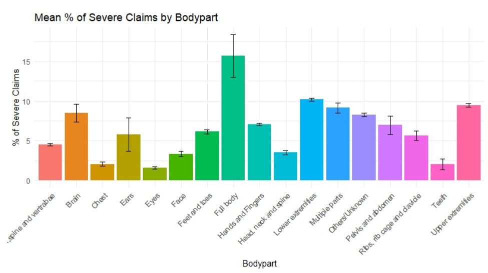
(d) Probability of severe claim by cause



(e) Probability of severe claim by weekday

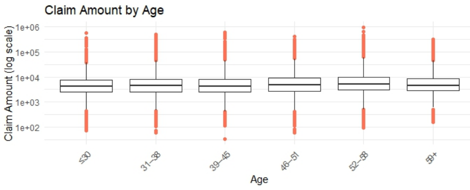


(f) Probability of severe claim by injury

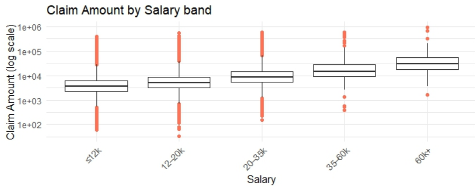


(g) Probability of severe claim by body part

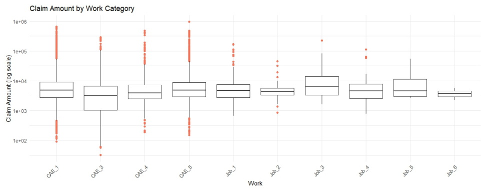
FIGURE 11: Relationship between the explanatory variables and claim amount



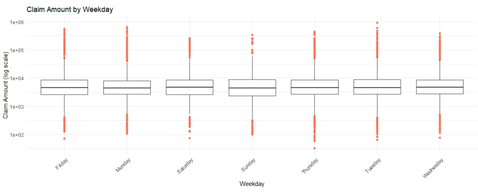
(a) Claim amount by age



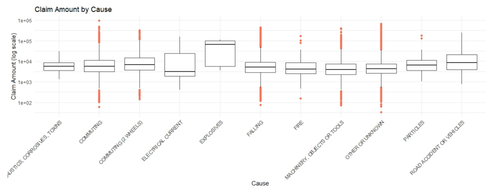
(b) Claim amount by salary



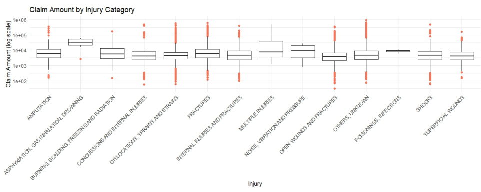
(c) Claim amount by work cluster



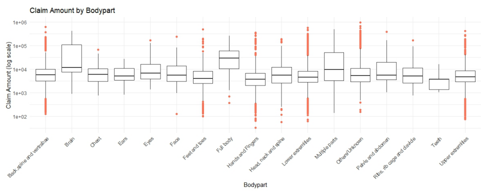
(d) Claim amount by weekday



(e) Claim amount by cause



(f) Claim amount by injury



(g) Claim amount by body part

FIGURE 12: Box-plots for the relationship between the explanatory variables and claim amount

B RESULTS FROM THE GLM FOR N

TABLE XIII: GLM Results for N

Variable	**Estimate (Std. Error)**
Accident_year	-0.024^{***} (0.005)
WeekdayMonday	-0.066^{*} (0.040)
WeekdaySaturday	0.190^{***} (0.058)
WeekdaySunday	0.071 (0.076)

Continued on next page

TABLE XIII – *Continued from previous page*

Variable	**Estimate (Std. Error)**
WeekdayThursday	-0.057 (0.042)
WeekdayTuesday	-0.023 (0.041)
WeekdayWednesday	-0.047 (0.041)
CauseCOMMUTING	1.914*** (0.397)
CauseCOMMUTING (2 WHEELS)	2.208*** (0.402)
CauseELECTRICAL CURRENT	1.907*** (0.484)
CauseEXPLOSIVES	3.856*** (0.633)
CauseFALLING	1.715*** (0.396)
CauseFIRE	1.038** (0.430)
CauseMACHINERY, OBJECTS OR TOOLS	1.221*** (0.396)
CauseOTHER OR UNKNOWN	1.123*** (0.395)
CausePARTICLES	0.557 (0.422)
CauseROAD ACCIDENT OR VEHICLES	1.829*** (0.430)
BodypartBrain	0.530*** (0.180)
BodypartChest	-1.211*** (0.163)
BodypartEars	0.232 (0.478)
BodypartEyes	-0.414***(0.125)
BodypartFace	-0.482*** (0.139)
BodypartFeet and toes	0.028 (0.069)
BodypartFull body	1.340*** (0.244)
BodypartHands and Fingers	0.528*** (0.054)
BodypartHead, neck and spine	-0.267*** (0.095)
BodypartLower extremities	0.744*** (0.048)
BodypartMultiple parts	0.467*** (0.105)
BodypartOthers/Unknown	0.782*** (0.064)
BodypartPelvis and abdomen	0.562*** (0.205)
BodypartRibs, rib cage and clavicle	-0.928*** (0.151)
BodypartTeeth	-1.874*** (0.347)
BodypartUpper extremities	0.598*** (0.051)
Bodypart_count	0.167*** (0.054)
InjuryASPHYXIATION, GAS INHALATION, DROWNING	-0.864 (0.599)
InjuryBURNING, SCALDING, FREEZING AND RADIATION	-3.586*** (0.226)

Continued on next page

TABLE XIII – *Continued from previous page*

Variable	Estimate (Std. Error)
InjuryCONCUSSIONS AND INTERNAL INJURIES	-3.805^{***} (0.169)
InjuryDISLOCATIONS, SPRAINS AND STRAINS	-3.501^{***} (0.163)
InjuryFRACTURES	-1.968^{***} (0.164)
InjuryINTERNAL INJURIES AND FRACTURES	-3.439^{***} (0.175)
InjuryMULTIPLE INJURIES	-3.350^{***} (0.233)
InjuryNOISE, VIBRATION AND PRESSURE	-2.921^{***} (0.641)
InjuryOPEN WOUNDS AND FRACTURES	-3.468^{***} (0.164)
InjuryOTHERS, UNKNOWN	-3.687^{***} (0.166)
InjuryPOISONINGS, INFECTIONS	-3.973^{***} (0.540)
InjurySHOCKS	-3.626^{***} (0.169)
InjurySUPERFICIAL WOUNDS	-4.187^{***} (0.167)
Injury_count	0.535^{***} (0.079)
WorkCAE_2	0.142 (0.157)
WorkCAE_3	-0.234^{***} (0.032)
WorkCAE_4	-0.692^{***} (0.056)
WorkCAE_5	-0.637^{***} (0.038)
WorkCAE_6	-10.083 (161.569)
WorkJob_1	-0.539^{***}(0.067)
WorkJob_2	1.253^{***} (0.247)
WorkJob_3	1.581^{***} (0.346)
WorkJob_4	0.338^{**} (0.160)
WorkJob_5	14.546 (89.508)
WorkJob_6	2.593^{*} (1.361)
GenderMale	0.250^{***} (0.030)
Age_band3038	0.424^{***} (0.050)
Age_band3945	0.569^{***} (0.049)
Age_band4651	0.911^{***} (0.049)
Age_band5258	0.962^{***} (0.048)
Age_band59124	1.100^{***} (0.052)
Salary_band12k-20k	-0.008 (0.029)
Salary_band20k-35k	-0.018 (0.042)
Salary_band35k-60k	-0.286^{***} (0.085)
Salary_band60k+	-0.216 (0.176)
Gross_cost_band12k-20k	-0.017 (0.048)

Continued on next page

TABLE XIII – *Continued from previous page*

Variable	**Estimate (Std. Error)**
Gross_cost_band20k-35k	-0.209^{***} (0.041)
Gross_cost_band35k-60k	-0.115^{***} (0.036)
Gross_cost_band60k+	0.120^{***} (0.036)
Constant	45.503^{***} (9.488)
Observations	109,628
Log Likelihood	−24,625.210
Akaike Inf. Crit.	49,398.420

Note: *p<0.1; **p<0.05; ***p<0.01

C RESULTS FROM THE GLM FOR X_1

TABLE XIV: GLM Results

Variable	**Estimate (Std. Error)**
Accident_year	0.029^{***} (0.003)
CauseCOMMUTING	0.321 (0.279)
CauseCOMMUTING (2 WHEELS)	0.358 (0.282)
CauseELECTRICAL CURRENT	0.132 (0.332)
CauseEXPLOSIVES	1.644^{***} (0.397)
CauseFALLING	0.303 (0.278)
CauseFIRE	0.136 (0.302)
CauseMACHINERY, OBJECTS OR TOOLS	0.274 (0.278)
CauseOTHER OR UNKNOWN	0.221 (0.278)
CausePARTICLES	0.103 (0.299)
CauseROAD ACCIDENT OR VEHICLES	0.673^{**} (0.298)
BodypartBrain	0.526^{***} (0.132)
BodypartChest	−0.137 (0.119)
BodypartEars	0.363 (0.392)
BodypartEyes	0.461^{***} (0.091)
BodypartFace	0.048 (0.099)
BodypartFeet and toes	-0.339^{***} (0.049)
BodypartFull body	0.934^{***} (0.174)
BodypartHands and Fingers	-0.346^{***} (0.040)
BodypartHead, neck and spine	−0.025 (0.069)

Continued on next page

TABLE XIV – *Continued from previous page*

Variable	**Estimate (Std. Error)**
BodypartLower extremities	-0.261^{***} (0.035)
BodypartMultiple parts	0.370^{***} (0.073)
BodypartOthers/Unknown	0.122^{**} (0.048)
BodypartPelvis and abdomen	0.223 (0.154)
BodypartRibs, rib cage and clavicle	-0.366^{***} (0.102)
BodypartTeeth	-0.616^{**} (0.287)
BodypartUpper extremities	-0.208^{***} (0.037)
InjuryASPHYXIATION, GAS INHALATION, DROWNING	0.665^{**} (0.293)
InjuryBURNING, SCALDING, FREEZING AND RADIATION	-0.473^{***} (0.141)
InjuryCONCUSSIONS AND INTERNAL INJURIES	-0.696^{***} (0.077)
InjuryDISLOCATIONS, SPRAINS AND STRAINS	-0.646^{***} (0.070)
InjuryFRACTURES	-0.408^{***} (0.070)
InjuryINTERNAL INJURIES AND FRACTURES	-0.608^{***} (0.082)
InjuryMULTIPLE INJURIES	-0.388^{***} (0.135)
InjuryNOISE, VIBRATION AND PRESSURE	-0.649 (0.523)
InjuryOPEN WOUNDS AND FRACTURES	-0.528^{***} (0.070)
InjuryOTHERS, UNKNOWN	-0.744^{***} (0.073)
InjuryPOISONINGS, INFECTIONS	-0.192 (0.432)
InjurySHOCKS	-0.796^{***} (0.076)
InjurySUPERFICIAL WOUNDS	-0.615^{***} (0.075)
Injury_count	0.143^{***} (0.042)
WorkCAE_3	-0.546^{***} (0.031)
WorkCAE_4	-0.038 (0.035)
WorkCAE_5	0.047^{**} (0.023)
WorkJob_1	0.058 (0.047)
WorkJob_2	0.014 (0.149)
WorkJob_3	0.339^{*} (0.175)
WorkJob_4	0.060 (0.098)
WorkJob_5	0.213 (0.237)
WorkJob_6	-0.220 (0.602)
GenderMale	0.239^{***} (0.021)
Age_band3138	0.012 (0.035)

Continued on next page

TABLE XIV – *Continued from previous page*

Variable	**Estimate (Std. Error)**
Age_band3945	−0.031 (0.034)
Age_band4651	0.035 (0.034)
Age_band5258	0.154*** (0.033)
Age_band59+	0.085** (0.036)
Salary_band1220k	0.131*** (0.020)
Salary_band2035k	0.533*** (0.030)
Salary_band3560k	0.881*** (0.062)
Salary_band60k+	1.096*** (0.126)
Gross_cost_band1220k	0.652*** (0.025)
Gross_cost_band2035k	1.033*** (0.035)
Gross_cost_band3560k	1.602*** (0.061)
Gross_cost_band60k+	2.565*** (0.088)
Constant	−49.795*** (6.479)
Observations	7,526
Log Likelihood	−12,123.460
Akaike Inf. Crit.	24,376.920

Note: *p<0.1; **p<0.05; ***p<0.01

TABLE XII: Variance Inflation Factors (VIF) for Stepwise Binomial GLM Model

Variable	GVIF	Df	$GVIF^{1/(2Df)}$
Accident_year	1.180	1	1.086
Weekday	1.032	6	1.003
Cause	3.326	10	1.062
Bodypart	3.896	16	1.043
Bodypart_count	1.703	5	1.055
Injury	5.163	13	1.065
Injury_count	1.632	3	1.085
Work	1.325	11	1.013
Gender	1.278	1	1.130
Age_band	1.090	5	1.009
Salary_band	1.214	4	1.025
Gross_cost_band	1.033	4	1.004

TABLE XV: Variance Inflation Factors (VIF) for Lognormal GLM (X_1)

Variable	GVIF	Df	$GVIF^{1/(2\times Df)}$
Accident_year	1.173	1	1.083
Cause	4.243	10	1.075
Bodypart	8.132	16	1.068
Injury	11.765	13	1.099
Injury_count	1.029	1	1.014
Work	1.428	9	1.020
Gender	1.281	1	1.132
Age_band	1.142	5	1.013
Salary_band	1.301	4	1.033
Gross_cost_band	1.202	4	1.023

DISCLAIMER

The writing process for this Master's Final Work was performed in accordance with the academic integrity standards set forth by ISEG - Universidade de Lisboa. All research, modelling, and written analysis are the product of my own work, except where specific sources are cited. The ideas presented are the result of my independent interpretation, supported by appropriate academic references.

In the interest of transparency, I disclose that artificial intelligence (AI) tools were used during the preparation of this work to assist with translation, improve language clarity, and support LaTeX formatting. These tools were not used to generate arguments, structure the content or perform data analysis. All final writing, analysis, and critical interpretation are entirely my own.

The use of AI did not compromise the originality or academic responsibility of this thesis.

All sources, whether AI-assisted or traditional, have been properly acknowledged in accordance with academic standards.

Júlia dos Santos Nascimento, 27th June 2025