Development and application of statistical learning methods in insurance and finance

Despoina Makariou

Department of Statistics
The London School of Economics and Political Science

A thesis submitted for the degree of Doctor of Philosophy

February 2022

This thesis is dedicated to my dear husband Laurynas Gadeikis and academic advisor Prof. Pauline Barrieu for making every day of my PhD journey a happy one.

Declaration

I certify that the thesis I have presented for examination for the Degree of Doctor of Philosophy at the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it). The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party. I declare that my thesis consists of less than 100,000 words.

Statement of co-authored work

I confirm that part of Chapter 3 has been adapted into a manuscript entitled **"A random forest based approach for predicting spreads in the primary catastrophe bond market"** jointly authored with Prof. Pauline Barrieu and Dr. Yining Chen, and is published on an **open access** basis in the peer-reviewed journal namely Insurance: Mathematics and Economics, see Makariou, Barrieu & Chen (2021).

I confirm that part of Chapter 5 has been adapted into a manuscript entitled **"A Finite Mixture Modelling Perspective for Combining Experts' Opinions with an Application to Quantile-Based Risk Measures"** jointly authored with Prof. Pauline Barrieu and Dr. George Tzougas, and is published on an **open access** basis in the peer-reviewed journal namely Risks, see Makariou, Barrieu & Tzougas (2021).

I confirm that part of Chapter 7 has been adapted into a manuscript entitled **"The Multivariate Poisson-Generalized Inverse Gaussian Claim Count Regression Model with Varying Dispersion and Shape"** jointly authored with Dr. George Tzougas, and is submitted, and is currently under minor revision, in a peer-reviewed journal.

Acknowledgements

I would like to express my sincere gratitude and appreciation to my first academic advisor Prof. Pauline Barrieu. I first got to know her through her writings on insurance-linked securities and my admiration for her work has been a huge inspiration for me to pursue this PhD. After meeting her in person, this admiration has been well extended to her exceptional personality and later on supervision. Her dedication to my development has been holistic, at times well beyond the PhD, and her continuous encouragement has been magnificent giving me purpose in every step of the PhD process. Certainly Prof. Pauline Barrieu has been a role model and indeed a great mentor to me in all respects. She is a very dear and important person in my life and I owe her much more than the great achievement of finishing my PhD thesis.

I also want to sincerely thank my second academic advisor Dr. Yining Chen who gave me a great deal of encouragement in research. I learned a lot from the discussions with him. Dr. Yining Chen showed me a very unique way of thinking when addressing theoretical and computational research problems including the importance of attention to every single detail in a manuscript. His commitment as my academic advisor has been shining through throughout my PhD and I am so grateful for his support.

Special thanks also go to Dr. George Tzougas for the exceptional guidance, encouragement, and kindness. His enthusiasm for research and positive energy has always been inspiring. Dr. George Tzougas has been a great influence in me taking further steps for advancing my computing skills. Certainly, he has also taught me very good time management skills and efficiency.

I am deeply grateful in advance to my examiners Prof. Erik Baurdoux and Prof. Stéphane Loisel for taking the time to read and assess my thesis.

My PhD studies are fully funded by the Department of Statistics at LSE to which I am extremely grateful for giving me this fantastic opportunity to begin my journey in research. This studentship has been generous and enabled me to focus on my research as well as being financially able to publish my articles on an open access basis, and present my research in international conferences. With regards to the latter special thanks also go to my first academic advisor Prof. Pauline Barrieu. Moreover, I thank the Department of Statistics at LSE for offering me great teaching opportunities throughout the PhD. It has been a great honour to teach the talented students at LSE. Furthermore, being involved in various committees in the Department of Statistics has increased my awareness on various departmental matters.

I express my sincere gratitude to industry practitioners Luca Albertini, Morton Lane, and Benjamin Jacquet who helped me a great deal to build an understanding of complicated aspects of the insurance-linked securities markets. I am also grateful to my previous employers in the London Insurance Market and to the Chartered Insurance Institute as they have both taught me so much about the fascinating world of insurance.

Last but not least, I thank my beloved husband Laurynas Gadeikis for being a bright source of motivation and the person who has inspired me to follow my PhD dream in the first place. He has provided me with splendid support in my transition from the professional to academic space and celebrated with me every single step of my PhD journey. I am grateful for his love and remarkable encouragement. Special thanks go to my PhD colleague and dear friend Tianlin Xu for sharing desks and being such a great companion at Columbia House, Research Room 5.02. I also thank my parents Theodoros Makarios and Olympia Konstantinidou for promoting curiosity towards learning since the early stages of my life.

Abstract

This thesis deals with the development and application of statistical learning methods in insurance and finance. Firstly, we focus on an insurance-linked financial instrument type namely catastrophe bond. Given the intricacies, and the over-the-counter nature of the market where these instruments are traded, we introduce a flexible statistical learning model called random forest. We use real data in order to predict the spread of a new catastrophe bond at issuance and identify the importance of various variables in their ability to predict the spread in a purely predictive framework. Finally, we develop and implement a series of robustness checks to ensure repeatability of prediction performance and predictors' importance results.

Secondly, we explore a decision-making problem which is faced in an abundance of interdisciplinary settings referring to the combination of different experts' opinions on a given topic. Focusing on the case where opinions are expressed in a probabilistic manner, we suggest employing a finite mixture modelling methodology to capture various sources of heterogeneity in experts' opinions, and assist the decision maker to test their very own judgement on opinions weights allocation too. An application in an actuarial context is presented where different actuaries report their opinions about a quantile-based risk measure to decide on the level of reserves they need to hold for regulatory purposes.

Finally, we focus on the problem of regression analysis for multivariate count data in order to capture the dependence structures between multiple count response variables based on explanatory variables, which is encountered across several disciplines. In particular, we introduce a multivariate Poisson-Generalized Inverse Gaussian regression model with varying dispersion and shape for modelling different types of insurance claims and their associated counts and we provide a real-data application in non-life insurance.

Introduction

Statistical learning, as a set of methods aiming to address the problem of discovering a function from data, can provide a useful framework for managing risk in insurance and finance. This thesis showcases how this can be achieved via the development and application of various statistical learning methods for addressing some problems faced in non-life insurance and its intersection with finance.

Some words on contributions Our contributions to research via means of published or under minor revision articles are three. Firstly, we focus on a spread prediction problem in the "opaque" primary market for catastrophe bonds for which a random forest methodology is deployed. Secondly, we explore model risk in reserve calculation expressing it as an opinion combination problem which can be addressed via a finite mixture modelling methodology. Thirdly, we investigate the modelling of different types of claims and their associated counts in non-life insurance for which purposes we develop a multivariate Poisson-generalised inverse Gaussian regression model with varying dispersion and shape for capturing overdispersion and positive correlation structures in highly dimensional claim count data.

These three topics may seem quite different with respect to their subject, however they are all motivated from daily business problems evidenced throughout the working experience of the author of this thesis within catastrophe insurance. That said, some of the contributions in this thesis are more general and therefore they are not limited to catastrophe insurance as such even though they are inspired by it. In the first contribution, the motivation has been a business need to develop an internal tool to assess quickly whether the price guidance of new catastrophe bond issuance is fair at real time when price negotiations between catastrophe bond sponsors and investors still take place. The second contribution has been motivated from catastrophe model risk as sometimes different licensed vendors models provide different views of risk for the same insurance account often causing confusion to stakeholders which use these results to make business decisions. The third contribution has been motivated from the wealth of multivariate data available to insurers' hands which are not yet extensively utilised for rate making and claims modelling purposes.

Our broader contribution to research, apart from the aforementioned articles, is that we bring together insights from insurance industry practice, as evidenced through the author's personal working experience and discussions with leading market participants, and actuarial academic literature in a variety of topics ranging from London Insurance

Market operations to premium calculation principles, and catastrophe risk modelling process and uncertainties only to name a few. This wealth of information surrounds our article contributions allowing the reader to have a bigger picture of the problems we study and not only. That said, even if each chapter of this thesis could be read individually as a selection of topics, we recommend that there are benefits to be extracted by reading it cumulatively. A chapter by chapter breakdown of the thesis follow.

Thesis structure In Chapter 1, we present a preamble consisting of two parts out of which the first presents the insurance and alternative risk transfer markets whilst the second one focuses on statistical learning especially relevant to research workflow, methods, and tools that are used in the research contributions presented later on. At the end of the chapter, we provide a brief description of the extend at which statistical learning is deployed in the insurance industry practice. We believe that this preamble is important because it allows the reader to form a rounded view about the general context of our research both from a market segment positioning and methodological perspectives. For example, some parts of our study focus on pricing products falling under the traditional insurance framework whilst others within an alternative risk transfer one, therefore the understanding of both is crucial. Moreover, we believe that it is beneficial to present some methods relevant to our research through the lenses of the type of the statistical learning task that we perform.

In Chapter 2, we illustrate various concepts related to non-life catastrophe bond pricing in the primary catastrophe bond market. We consider topics relevant to structuring, notion of catastrophe bond pricing, and some crucial attributes that could affect the issuance price of catastrophe bond. Given the fact that catastrophe bonds payout is linked to an underlying insurable risk, we deem helpful to end with an Appendix illustrating a general pricing formation methodology for insurable risks. It is important to mention that a big part of this chapter is based on the author's working experience in the London Insurance Market and discussions with key catastrophe bond market practitioners. Finally, the material presented here is a building block for Chapter 3 where our first research paper on catastrophe bond pricing using statistical learning methods is presented.

In Chapter 3, we firstly present our article entitled "A random forest based approach for predicting spreads in the primary catastrophe bond market" published at Insurance: Mathematics and Economics. We deploy a random forest methodology to facilitate spread prediction in the primary catastrophe bond market and we assess the importance of each covariate in the prediction of spread. The analysis is based on very rich data set of non-life catastrophe bonds issued in a time period where the market has entered a slightly more established stage after the financial crisis. It is worth mentioning that we perform multiple robustness tests both for random forest generalisation ability and variable importance measures always in comparison with a very competitive benchmark model. Since this work, to the best of our knowledge, is out of the first that suggests a purely predictive framework in a catastrophe bond market setting we also discuss the synergies that can be achieved when examined in line with previous contribution in the explanatory framework. Additional to the main research paper, which is presented as published in terms of format, we consider important to include some supplementing material. The latter is mainly comprised of some modelling considerations at the initial stage of our research, and extra robustness checks which, whilst they are not included in the published version of the article, we believe that they are of material importance.

In Chapter 4, we introduce the concept of capital requirement which is the amount of capital that a financial institution has to put aside for the risk taking activities that it undertakes. The importance of an adequate reserve for the financial health and regulatory compliance of a financial company is highlighted. Moreover, great emphasis is given in presenting monetary risk measures as way to identify the minimum amount of capital to add to a position to make it acceptable from a financial strength and regulatory viewpoint. That said, there is variety of monetary risk measures that one could use and thus we demonstrate some popular options along with their financial interpretation, merits, and limitations. The chapter concludes by emphasising that no matter which risk measure is chosen by a financial institution, there is an inherent model risk in its computation. The latter point is a stepping-stone to Chapter 5 where our second research paper relevant to model risk in reserve calculation.

In Chapter 5, we start from presenting our article entitled "A Finite Mixture Modelling Perspective for Combining Experts' Opinions with an Application to Quantile-Based Risk Measures" published at Risks. There we give a different perspective on expert opinions' combination using finite mixture models with the components of the mixture not necessarily coming from the same parametric family. We flexibly account for multiple sources of heterogeneity involved in the opinions expressed by the experts in terms of the parametric family, the parameters of each component density, and also the mixing weights. The flexibility of our approach in capturing multiple sources of heterogeneity involved in the opinions expressed by the experts in terms of the parametric family, the parameters of each component density, and also the mixing weights can assist a decision maker to make informed decisions. Our proposed models are then used in an actuarial application for numerically computing quantile-based risk measures taking into account model risk. Except for the aforementioned article, we include some supplementary material where we share some ideas regarding potential extensions of our study.

In Chapter 6, we touch upon the phenomenon of multivariate count data in non-life insurance through the angle of current trends in the market related to insurers underwriting several lines of policies, whose claim counts and size may showcase some dependence. We then illustrate some situations stipulated by recent actuarial research to signify the importance of multivariate data modelling with dependence for insurance premium and claims prediction. In particular, we classify previous research studies per reason why such joint modelling methodologies are important in the current non-life insurance practice. Then, we naturally proceed to Chapter 7, where we provide our contribution to the literature of multivariate count data in non-life insurance with dependence.

In Chapter 7, we present our article entitled "Multivariate Poisson-Generalized Inverse Gaussian regression model with varying dispersion and shape" which is under submission and minor revision and it is incorporated in the thesis in the exact format in which it has been submitted. In particular, we introduce and build a multivariate Poisson-Generalised Inverse Gaussian (MVPGIG) regression model with varying dispersion and shape in order to model positively correlated and overdispersed claim counts from different types of non-life insurance coverage flexibly. We implement the model by utilising bodily injury and property damage claim count data from a European motor insurer. Additionally, some supplementary material with research extensions which have not been accompanying the submitted version of the article is included at the end.

It is worth mentioning that the references related to the articles which are, published on an open access basis or under minor revision, and presented in Chapters 3, 5, and 7 are included within the articles. The references listed at the end of this thesis refer to any other citations in this thesis other than those included within the aforementioned articles. Generally, the text of this thesis is written using British English spelling except from the name of some methods which are extensively used in the literature with American English spelling.

On the experience of PhD Whilst this written work forms a representative piece of my PhD studies, it cannot fully reflect this exciting personal development journey which equals more than the sum of this thesis parts. The PhD has been a wonderful experience improving my analysis, problem-solving, project management, collaboration, verbal, and written communication skills among others. Whilst towards this direction, individual effort played a role, I strongly believe that these improvements have been fostered by the mentoring of my academic advisors with whom I have developed a very meaningful connection.

Moreover, my prior professional experience in the London Insurance Market in modelling natural catastrophes, terrorism, and political violence risks across various non-life insurance lines has brought a business perspective when formulating applied research problems and made data search for conducting research easier. Furthermore, I am delighted that, during my PhD, I have advanced my computing skills, by learning how to code, which has in turn allowed me to implement and experiment on various research ideas. Also, attending and presenting research in conferences and seminars introduced me to the making of creative academic dialogues. With regards to my experience in research publication process, it has equipped me with attention to detail, persistence, and most of all appreciation of the merits of constructive feedback.

Having had the honour to be a member of the Department of Statistics at LSE has provided me with a unique exposure to reading groups in several areas of statistics which have ignited my interest in experimenting with a diverse set of research methods. Teaching opportunities allowed for personal interaction with students and learning how to break down complicated material into smaller pieces and then connecting it to real-life applications. Last but not least, as an invited member in various departmental committees, I have increased my awareness in various aspects of academic life.

Contents

Chapter 1

Preamble

In this preamble, we present some preliminaries regarding the methods and tools that we use in addition to the area in which the latter are applied. In particular, this introductory material is divided into two sections. The first part deals broadly with insurance considering also reinsurance, and other alternative risk transfer mechanisms. Except from discussing the aforementioned concepts, we put emphasis on the London Market explaining the reasons behind its unique status globally, as well as the way in which the business flows there, and its regulation. Distinctions between insurance and other forms of risk transfer are drawn. In the second part, we present some preliminary material on statistical learning relevant to the methods and tools that are used in the research contributions presented later on. We also briefly refer to the current level of statistical learning deployment in the current insurance practice.

1.1 Insurance essentials

In the first part of the Preamble, we will provide some background information in the areas of insurance, reinsurance, and insurance-linked securitisation. We will place particular focus on the London Market, and even more so on Lloyd's market as a distinctive insurance market place within the London Market, whilst the insurance sector regulation in the UK will also be discussed.

1.1.1 Risks and risk transfer

The word "risk" is being used in real life in an abundance of different settings. Whilst there is no single accepted definition, see for example Punter (2003), by using a probability theory terminology, risks can be seen as random variables which map uncertain states of the world into values which may represent profits and losses. When a risk is associated only with a possibility of loss to the party that is exposed to it, we speak of a pure risk. When there is possibility for both making a loss or profit out of a situation then, we speak of a speculative risk.

Some notions which are closely related to the concept of risk are these of peril, i.e. the cause of a loss, and hazard, i.e. the event which originates or increases the peril. Both peril and hazard have a bearing on risk but risk is a broader context although in practice there are occasions where these terms may be used interchangeably. Focusing on pure risks, individuals or businesses being exposed to them may decide to contractually shift the latter to another party if they feel that they are not risk tolerant enough to face the possible loss consequences, see Banks (2014). This is a risk management and control strategy called risk transfer. As we will see throughout this thesis, there are many risk transfer mechanisms from traditional insurance to alternative risk transfer options involving the capital markets.

At this point, we refrain from providing a formal definition of insurance, as we deal with this matter in Chapter 1, Section 1.1.2. However, we consider essential to discuss a main concept behind insurance which is known as risk pooling. In particular, as Smith & Kane (1994) nicely explain, based on the law of large numbers, the average of a substantial number of independent identically distributed random variables tends to approach the expected value. The latter means that when an insurer includes extra risks in an insured pool can achieve reduction in the variation of the average loss per insured around the expected value. Now, if every insured contributes more in terms of premium payment into the pool than the expected loss payment, it means that by adding more insureds into the pool, the probability that the pool will not have enough money to pay for any claimed losses decreases.

However, one should bear in mind that whilst risk pooling works well for homogeneous risks, it is not always applicable, see Braun et al. (2020). For instance, in the case of catastrophe risks, which are described by a low frequency of occurrence combined with very high loss severity potential, the insurer cannot easily achieve to diversify their losses even at a global level. That said, we direct the interested reader to Chapter 2, Section 2.3.1 for more details on some limitations of risk pooling.

1.1.2 Towards a legal definition of insurance

Whilst it may be tempting to provide a clear definition of insurance such a formulation is not straightforward for several reasons. From a judicial perspective, any legal statues regarding insurance regulatory matters refrain from providing a clear definition in fear of running the risk to exclude contracts which should have been falling within the scope. Keeping things relatively loose in terms of definitions provide regulators with enough flexibility to judge on a case by case basis. Other reasons potentially why an official strict definition is not formed is that insurance undertakings are very strictly regulated anyway, and also from an insurance law viewpoint, there are some special principles, such as the doctrine of utmost good faith, which before identifying whether the latter applies, one would need to prove first that the contract in question is one of insurance.

From a purely regulatory purposes, we employ a rather loose explanation of what insurance is as described in Birds (2010) according to which insurance is any contract whereby a party agrees to take on a risk associated with the future occurrence of a random event in which the other party has an interest and under which contract the first party is obliged to remunerate the second party via means of money or its equivalent subject to such an event occurring. It follows that any party entering frequently in such contracts as risk bearer is considered an insurer for the purposes of the statue regulating insurance business. In order for a contract to qualify as insurance the cedant must generally demonstrate an insurable interest - it must prove that it has suffered an economic loss once the defined event occurs. Insurable interest exists to reduce gambling or moral hazard. There are other requirements for a valid insurance contract and we direct the interested reader to Warr (2016*a*) and Birds (2010).

1.1.3 Some classifications of insurance

The risks that insurance business covers can be categorised in various ways depending on the perspective that one would employ. For example, Jerry & Richmond (2012) have provided a thorough discussion on the topic of examining several viewpoints including the nature of the risk, nature of the insurer, and nature of marketing. Here, we make the following three distinctions.

First versus third party insurance In Birds (2010), a simpler classification system of two types is suggested on the grounds of legal importance. In particular, the first distinction is between the so called first and third party insurance depending respectively on whether someone purchases insurance to cover risks such as their life, or personal property or the motive behind buying insurance is to protect themselves against the possibility of being held liable in law to pay damages to another party. Obviously, this is not to say that there are no insurance policies that would cover both types of risks at the same time but such a differentiation is important as insurance law signifies this difference by requiring that some types of third party insurance have a compulsory character as well as the fact that both the insured and a third party are involved.

Life versus non-life insurance The second classification distinguishes between life and non-life insurance. Whilst for all types of insurance, an element of uncertainty is required, there is some difference on how uncertainty is expressed between life and non-life insurance. In life insurance policies, for example, it is certain that the person who takes on a life insurance product will die at some point and thus a pay-out will be needed, so the real question is when this will happen.

In contrast, in the context of non-life insurance, the occurrence of an event or suffering of loss based on which a payout may be claimed is not certain. Think for instance, a property policy the insured buys coverage against losses that may because to their house because of fire but the house may eventually never burn. Generally, non-life insurance contracts are contracts of indemnity meaning that they aim to indemnify the policyholder, only with reference to the loss that the latter has suffered, assuming that such loss has really been suffered and only to the size of the loss suffered. Of course, there are exceptions in the above indemnity rule as insurer and insured could agree to waive this out by agreeing that a particular amount of money is payable in case a certain event occurs. Such policies are known as valued policies.

Catastrophe versus non-catastrophe insurance Another classification is to further differentiate between insurance policies which provide coverage against losses arising out of the occurrence of catastrophes or those which not. By catastrophe[1] here, we mean a low frequency event which can have a potentially big financial impact. The cause of such events can be either of natural or man-made origin. Examples of the former case are earthquakes, hurricanes, and wildfires whilst instances of the latter case include terrorism attacks, explosions, and political violence risks among others.

Catastrophe insurance policies are usually specialised and provide coverage for very specific disasters subject to policy terms. Although catastrophes events are of low frequency, the global economic losses arising out of them have increased significantly over

[1]Throughout the thesis, we use the terms catastrophe, catastrophe risk and disaster interchangeably

the last decades for various reasons including climate change and the increase in the population living in catastrophe prone areas to name a few, see Swiss Re (2019). In particular, according to Swiss Re Institute (2021) this number reached 202 billion dollars in 2020, an increase from 150 billion dollars compared to the previous year. Out of these costs, the insurance industry has absorbed only 89 billion US dollars.

With regards to the non-catastrophe policies they cover for perils with relatively higher frequency and lower severity potential, such as automobile, fire, or theft risks. Non-catastrophe risks, from the other hand, are more well understood and can be pooled fairly easily across the insurance sector. For more details on the distinction between catastrophe and non catastrophe risks see Chapter 2, Section 2.3.1. The focus of the current thesis lies on research topics relevant to non-life insurance policies with and without catastrophe coverage.

1.1.4 A short history of insurance

The foundations of modern insurance were set by Italian merchants in the 14th century in an effort to protect themselves against various maritime risks such as losing their cargo or ships whilst in the sea. By the 16th century, the marine insurance tradition has been passed on to London where such transactions have been taking place in a coffee house owned by an individual named Edward Lloyd. There, whoever merchant wanted to buy insurance would write in a slip of paper the information about the object that they would like to insure and the size of coverage needed. Then the insurance buyer would pass this slip around the clients of the coffee house in search for people that would accept to provide insurance. The proportion of coverage which a given client would provide was underwritten on the slip and this process continued up until the desired total amount of insurance has been arranged.

With years, this place has been transformed into a corporation with statutory authority under the same name, i.e. Lloyd's, and whilst it seized from being a coffee shop the way in which the insurance business is conducted up until today have remained the same at its core. Lloyd's have played a great role, not only at the development of insurance but also the very foundations of insurance law. For instance, the Marine Insurance Act 1906 included the standard Lloyd's maritime policy as a statutory form. Based on all these principles other insurance types were born when there was a historic demand for it starting from fire insurance after the Great Fire of London in 1666, and life and personal accident insurance following the industrial revolution in the 19th century to present when one can insure almost anything against the risk of loss or damage, see Warr (2016*a*).

1.1.5 The London Market

The London Market is plausibly one of the most notable segments of the UK insurance and reinsurance sector as this is where the placing of risks characterised by increased complexity, rareness, and size from all over the world happens.

Global appeal There are several reasons behind the global appeal of the London Market, see Warr (2016*a*) for a thorough discussion but some important points are summarised as follows. Firstly, the London Market has enough financial capacity to take on large risks and, as we will see in Chapter 1, Section 1.1.5.1, this is even more evident in the case of Lloyd's Market due to its subscription-based structure. Secondly, the spirit of innovation prevailing in the London Market brings new perspectives in the area or risk transfer and this is appreciated by clients all over the world demanding bespoke solutions. Thirdly, the claims service, which is what attracts an insurer the most in the eyes of the insured, has a strong reputation. Finally, throughout the 300-year history of the London Market, the underwriting expertise has created a competitive advantage in underwriting even emerging risks.

Business lines The classification of business lines across the London Insurance Market is not strictly set at a theoretical level, but here we employ the approach of Warr (2016*b*) employing the following categories, i.e. marine, non-marine, and aviation. Firstly, under marine business, an insurer typically insure vessels, cargoes, and their subsequent liabilities as well as offshore energy, and marine related construction risks. Secondly, non-marine business provides coverage for property risks and associated liabilities including construction for buildings, onshore energy including power generation and alternative energy risks. Within the non-marine business class, one can also find coverage against professional liability and personal accident risks. Finally, the aviation business covers physical damage to air-crafts and associated liabilities.

Participants There are many participants operating in the London market including insurers, reinsurers, protection and indemnity clubs, and Lloyd's syndicates among others. Briefly, insurers provide protection against losses arising from the materialisation of pre-agreed risks for a fee, reinsurers provide insurance to insurers, and the protection and indemnity (P&I) clubs are mutual marine insurance associations providing coverage for its members. Lloyd's syndicates are groups of private individuals or corporate investors who actually carry the risks underwritten in Lloyd's marketplace. Whilst all participants are important in the ecosystem, here, we will focus on the Lloyd's market since it is considered to be the world's oldest and largest insurance marketplace, see Lloyd's (2019), and highlight its differences from the rest of what is called Company Market within the London Market.

1.1.5.1 Lloyd's market

When discussing about Lloyd's, the first thing to clarify is that it is not an insurer but a marketplace which is headquartered at London Lime Street. It is interesting to see that Lloyd's is considered to be the strongest insurance brand in the world without being the insurance provider itself, see Warr (2016*a*).

Transactions In Lloyd's, the insurance transactions are conducted mostly in person in the so called "Underwriting Room". The risks are placed by brokers, who represent the clients willing to buy insurance, and Lloyd's specialist underwriters whose role is the evaluation, pricing and acceptance of risks. To understand the size of the market, we shall mention that according to Lloyd's (2019), on an average working day there are 5,000 people with physical presence in Lloyd's whilst the average daily premiums entering the market exceed 100 million British pounds and the average amount of claims remitted daily is more than 50 million British pounds. Furthermore, there are no limitations in terms of size, complexity, geographical region, or industry segment for the risks that are placed there.

Subscription market One reason behind strong Lloyd's capacity and flexibility to make unusual placements is that risks are written on a subscription basis meaning that more than one carrier accepts a proportion of the same risk. By no party assuming the whole financial liability, larger risks can be accommodated whilst having multiple participants embraces diverse underwriting skills making it possible to create speciality insurance solutions. It is important to mention that the Lloyd's market is regulated by two different regulatory bodies that is the Financial Conduct Authority (FCA) and the Prudential Regulatory Authority (PRA) for which we will speak in detail in Chapter 1, Section 1.1.5.4.

Structure They way in which Lloyd's is structured for management purposes is defined in several Acts of Parliament between 1872 and 1982 which are named Lloyd's Acts, see Warr (2016*a*). Out of these, the Lloyd's Act 1982 showcases that Lloyd's market is managed and supervised by the Council of Lloyd's. The primary goal of the Council is to govern and oversee the market putting the insured at the centre of focus. The Council of Lloyd's is comprised of three working[2]., three external[3] and nine nominated members which are elected by Lloyd's members. The members of the Council of Lloyd's elect the Chairman and the Deputy Chairmen on an annual basis. The Council of Lloyd's can dismiss some of its functions by making decisions and issuing resolutions, requirements, rules, and byelaws Warr (2016*a*). The market strategy of Lloyd's market is set by the Franchise board which is also accountable for meeting risk management and profitability goals at a market level. In particular, it sets a set of rules of how managing agents should operate which it then monitors to guarantee that high standards of underwriting and risk management are met so to ensure that the market as a whole is profitable and financially strong.

Syndicates The risks that are underwritten in Lloyd's are carried and backed financially by groups of individual or corporate investors named syndicates. The terms Names or underwriting members are used interchangeably in reference to the aforementioned investor categories, i.e. individual or corporate, that may be comprising a syndicate. It is important to clarify that the size of a given syndicate is not determined by the count of its Names but from the financial capacity of each syndicate to cover risks. Therefore, we see that even if the number of syndicates have dropped from 400 in 1990s to only 71 currently, the market has actually increased in size. Legally wise, syndicates have a very unique status in that each of them does not exist as a legal entity separately but they simply constitute a sum of their parts i.e. underwriting members.

Another interesting feature of syndicates is that they have an expiry date and in particular, they need to be renewed on an annual basis. Syndicates are identified by a pseudonym, a unique number, and the year of the account. As an example, for a given year of account, HIS 0033 is a syndicate name standing for HISCOX insurer with a unique reference number 0333. Finally, since a syndicate is nothing else than a sum of its parts, the day to day operations need to be made through another entity which is known as a managing agent. The latter is a company founded for the purpose of underwriting management of one or several syndicates. The managing agent as such employs the underwriters and this is the entity that it is subject to regulation for prudential and conduct of business matters[4].

[2]Working members are classified as those who have an active engagement in the market's day to day operations

[3]External members are those who do provide capital in the market but they cannot be classified as working members

[4]More on the UK regulation will be found in Chapter 1, Section 1.1.5.4.

Capital providers The biggest part of capital being invested in Lloyd's is coming from Names who are corporate investors rather than individual ones. However, no matter the Names type, making an investment into this market requires meeting certain financial adequacy criteria. In other words, Names cannot have unlimited liability in covering losses. This is to ensure that the size of claims that would arise as a result of the risks being undertaken in Lloyd's can be capped. Such restrictions were imposed in 1990s straight after a bitter chapter in the history of Lloyd's when in the late 1980s, unprecedented large claims, both in terms of size and count, arose due to the occurrence of various catastrophic events such as hurricane Hugo. Back then, Names had unlimited personal liability and consequently there was no way to limit these extreme claims demand. Due to this incident, Lloyd's had to rebuild itself by creating a run-off agent called Equitas Reinsurance Limited (ERL) to clear out the balance sheet of Lloyd's for the 1992 year of account and prior and let the market start again from a fresh page. For any given year, the capacity of Lloyd's is the result of aggregation of the syndicates capacity.

Members' agents Another Lloyd's market participant is called members' agents. The main responsibility of the latter is to consult Names on which syndicate(s) they should direct their investment to and what amount they shall invest. The advice is provided considering the risk profile of a given Name. For instance, if the latter is risk averse, a member agent may recommend providing capital to a syndicate that focuses on motor insurance because the nature of the risk, mostly high frequency and low severity, would easily ensure a steady and moderate rate of return. On the other hand, if a Name's risk appetite is more opportunistic, the member's agent may recommend classes of business which indicate higher volatility such as offshore energy. Such an approach would certainly have the potential for larger returns but also large losses. Members' agents also keep in touch with Names within the year to communicate any foreseeable deviation in the spread of their investment for the next year of accounts.

1.1.5.2 Company market and differences with Lloyd's

Lloyd's represents only a portion of the London Market. An important other part is what is known as London Company Market and here we will present some of their differences based on the following perspectives, i.e. insurer's structure, market participants, marketplace, regulation, and international liaison, see Warr (2016*a*).

Starting from insurer's structure, London Company Market insurers have the freedom to choose among various structural formats such as limited liability firms, mutual indemnity associations, mutual companies, and captive insurers whilst such flexibility does not exist for Lloyd's insurers. In addition, participants' wise, the London Company Market includes firms which are very diverse with respect to their origin being normally UK branches of large international firms from abroad or UK firms with large presence in the UK and around the world. With regards to the marketplace, unlike Lloyd's, the Company Market does not provide any physical place to transact insurance business except for the underwriting centre in minister court in which few insurers have voluntarily decided to use it as basis.

Moreover, regulation wise, as we will see in Chapter 1, Section 1.1.5.4 in more detail, all insurance providers in the UK are regulated by the Prudential Risk Authority (PRA) for prudential issues such as solvency and capital requirements, and the Financial Conduct Authority (FCA) for business conduct topics. Companies of EU origin also need to satisfy the demands of their home regulator. That said, Lloyd's Market is unique in that, except for the aforementioned rules, whoever operates there, is subject to an additional layer of regulation implied by Lloyd's itself. Moving forward, from an international liaison viewpoint, Lloyd's as a whole liaises with regulators abroad whilst each participant in the London Company Market needs to make separate arrangements with each foreign regulatory body. In what follows, we will discuss some ways in which the business flows in the London Market.

1.1.5.3 Flow of business in the London Market

The placement of risks and any associated processing of claims in the London market follow a specific end to end business process workflow. Here, for expository purposes, we differentiate between the following three stages namely risk quotation, risk placement, and claims handling.

Risk quotation Starting with the placement of a risk, a broker, i.e. a party who intermediates between the insurer and the insured, creates a summary of the risk in question and the recommended terms and conditions in a file called a Market Reform Contract (MRC) or simply "slip" – a traditional notion linked to historical facts described in Chapter 1, Section 1.1.4. The MRC is standardised with respect to its content and format to promote transparency for all stakeholders involved so that there is no ambiguity regarding terms and coverage at the time that parties enter the insurance contract. This is something that in the market is known as contract certainty indicating that prior the inception of a policy all terms between the insurance company and the policyholder need to be agreed upon and any supporting documents proving what has been agreed need to be given by the policy inception or straight after – normally within 30 days. There is an element of flexibility regarding what documents can be considered as proof including the official insurance policy contract, or simply a copy of the slip among others.

Risk placement When an insurance purchase agreement happens with the assistance of a Lloyd's broker, the latter first approaches for a quotation the underwriter who is recognised by the market as being the expert, the so-called "leader", in the relevant class of business. The leader will recommend a certain percentage of risk that they are willing to share and communicate the terms and conditions for such coverage to be extended. Since the Lloyd's broker represents the client, they will try to identify as many leaders as possible, if they exist[5], and come back to the client presenting the quotes acquired and recommending which of them would be most appropriate for the client given their needs.

When the client makes a decision about which leader's quotation they want to proceed with, the Loyd's broker contacts other underwriters who can belong in any market, i.e. Lloyd's, London companies or the international markets, until 100 percent of the coverage required is achieved. The risk placement and its associated documents are reported on a central market database which is known as Xchanging Ins-sure Services (XIS). Then, the broker needs to receive the premium from the client and pass it on to the insurers minus a brokerage fee which is agreed in advance and it is a percentage of the total premium. Once this process finishes, both the broker and the underwriters receive the risk data from XIS.

Claims handling When an insured suffers a loss which they believe that it could be covered by the insurance policy that they have in place, they need to notify the broker as soon as possible. This is because the appropriate combination of insurers has to be identified. After this happens, the broker submits the claim on paper or electronically via the so-called Electronic Claims File (ECF). The broker sits in between the insurer, and the client in all stages of the claims processing acting as they key contact for the insurer, third parties such as loss adjusters and the client. If a claim is qualified and a payment needs to be paid, the funds move from insurers' bank account to the broker's bank account or an agreed third-party such as a lawyer's office and from there directly to the insured. All this happens in a very transparent way with electronic messages being sent to both insurance companies, and the broker via the XIS where all information is recorded regarding the movements of capital, and the claims to which these relate. In what follows, we briefly discuss the regulatory framework in the London Insurance Market.

[5] This highly depends on how unusual is the risk being transferred.

1.1.5.4 The UK regulatory framework

Companies performing insurance undertaking business in the UK are regulated by two bodies, both sitting within the Bank of England, namely Financial Conduct Authority (FCA), and Prudential Regulation Authority (PRA).

Financial Conduct Authority The FCA views the market as a whole from the top or in other words it is responsible for macro-prudential regulation. In particular, the FCA has responsibility for matters relevant to the conduct of business and general market matters for all insurers so that to secure a certain level of protection for the policyholders. In doing so, FCA performs analysis for the whole market, from insurance product design to distribution, in order to identify practices that have the potential to harm policyholders. In case where such threats are identified, the FCA has power to intervene and ban products or services.

Prudential Regulation Authority Moving to PRA, it has a more detailed approach being interested in every single individual part of the market and in this sense it is accountable for the micro-prudential regulation, see Bank of England (2021). Its role involves encouraging the safety and financial fitness of the insurers it regulates and ensure that policyholders can achieve a proper level of protection. Towards this direction, PRA looks at the external environment, business risk, management and governance, risk management and controls, and capital and liquidity, see Warr (2016*b*), to assess and prevent potential harm that insurers could cause to the stability of the UK financial system. It implies rules to ensure that insurers are financially sound and have enough reserves to fulfil the remuneration promise that they give to their policyholders if such a need arises. Towards this direction, the PRA, under a EU Directive called Solvency II, see European Insurance and Occupational Pensions Authority (2015) for more details, sets the requirements for firms to calculate their solvency capital. Nevertheless, no specific method that is implied for its computation, a point that we also touch upon in Chapters 4 and 5 in the current thesis.

Lloyd's regulation status With regards to the regulation of Lloyd's market participants, it is worth mentioning that Lloyd's managing agents are regulated by both the FCA and the PRA whilst members' agents and Lloyd's brokers are regulated by the FCA. For the calculation of the solvency requirements, Lloyd's is treated as a distinct entity which implies that any requirements are applied on a market level. Moreover, it is worth mentioning that the regulators permit Lloyd's to maintain a level of internal regulatory control, see Warr (2016*b*). Because there is a degree of overlapping objectives between the FCA, PRA, and Lloyd's, there are some arrangements for collaboration between the aforementioned parties on topics regarding supervision, and its execution to limit the amount of duplicated work. For more on these arrangements, we direct the interested reader to Prudential Regulation Authority and Society of Lloyd's (2013) and Financial Conduct Authority (2013). Next, we focus on discussing the ways in which insurers can insure themselves against risks undertaken within their normal course of business.

1.1.6 Reinsurance

Insurers may decide that they want to transfer some of the risks they face by buying insurance for themselves. This practice is known as reinsurance and it has a longstanding history as discussed in the classic essay of Kopf (1871). In particular, it may be plausible that in the early years of marine insurance practice, as we also described in Chapter 1, Section 1.1.4, the back then underwriters would not base their risk taking decisions on a solid statistical analysis but they would write risks based on their perception instead. Sometimes, the insurers whilst having assumed the risk, they may have been feeling worried about the faith of the ship or cargo they had agreed to insure and they were seeking to re-sell the risk for the more "dangerous" parts of the voyage at a higher price.

If we apply the definition provided for the term insurance in Chapter 1, Section 1.1.2 to a reinsurance context, reinsurance is a practice where the original insurer enters into a contract with another insurer or reinsurers (in the latter case we speak of retrocession) to take on a partial or the full risk assumed by the primary insurer for an agreed premium. In the London Market Insurance context, by insurers we mean all legal entities which are allowed by the regulators to make contracts of insurance.

In the UK, the purchase of reinsurance does not have a compulsory character as it happens for some classes of primary insurance but it is viewed by the regulatory bodies as a good business practice. It is worth mentioning that reinsurance purchase does not take away from the primary insurer the responsibility to remunerate the originally insured thus in any case the insurer needs to be cautious for the risks it takes so that to avoid insolvency in case claims arise, see Warr (2016*b*). Nevertheless, reinsurance entails multiple benefits both for the primary insurer, and the reinsurer and some of them are presented below.

1.1.6.1 Benefits of reinsurance

We discuss few merits of reinsurance both from the side of the buyer and the seller based on Warr (2016*b*) and Swiss Re (2021). We differentiate between advantages that an insurer may enjoy from buying and selling reinsurance protection.

Buy-side Starting from the buy side, when a primary insurer purchases reinsurance reduces their gross direct exposure meaning that their underwriting capacity for any one risk increases and in doing so they can achieve a bigger market share. That said, if this holds for every insurer, it means that the risk sharing capacity of the market as a whole increases. Moreover, because reinsurance acts as an alternative form of capital, there may be benefits from a regulatory point of view when having to compute the solvency capital requirement - a concept explained in Chapters 4 and 5 in this thesis. Furthermore, reinsurance helps the primary insurer to manage their risk portfolio better as it enables them to spread the cost of potentially very big losses over a longer period of time. An additional benefit of reinsurance is that an insurer can use it as a indicator of good practice in the eyes of the regulator when wishing to enter into an uncharted class of business.

Sell-side When an insurer sells reinsurance, they can enjoy various advantages such as exploring new business opportunities both in terms of geographic location, and risk type. Starting from location, sometimes insurers face barriers for writing direct business in certain places. However, such obstacles may seize to exist when it comes for reinsurance provision enabling a insurer to entry into a new market geography. Moreover, reinsurance is a relatively safe way to explore the feasibility of writing risks outside an insurer's current expertise. In particular, writing a novel risk perhaps requires an amount of investment in underwriting headcount yet with uncertain financial outcomes. Since reinsurance is renewed normally on an annual basis, the reinsurer has some room to experiment with new risk categories. That said, if it appears that expected financial gain is not as high after all, there is a clear exit strategy for the reinsurer which is simply not to renew the risk. In what follows, we present some basic types of reinsurance contracts.

1.1.6.2 Some key types of reinsurance

There are plenty of reinsurance contract types, but most of them can be classified into two forms namely facultative, and obligatory reinsurance. Their main difference lies on whether individual risks are transferred on a standalone basis to the reinsurer in which case we speak of facultative reinsurance or multiple risks are transferred all at once namely obligatory reinsurance, see Swiss Re (2021).

Facultative In a facultative reinsurance contract the original insurer has the freedom to select the risks to be reinsured, but at the same time the reinsurer also maintains the option to accept or reject any given risk being offered by the primary insurer. Because of its flexibility, facultative reinsurance is often used as an addition to obligatory reinsurance in order to provide protection from risks over and above what is covered by an obligatory reinsurance scheme.

Obligatory In cases when an insurer wants to buy reinsurance for the whole portfolio within a given class of business then obligatory reinsurance may be the first choice. The term obligatory signifies that both the insurer and reinsurer are legally obliged to transfer or reinsure alike the batch of risks meaning that there is no room for not accepting specific risks. Because there is an element of automation in the process, obligatory reinsurance is associated with lower administrative costs.

Finally, both types of reinsurance can be written either on a proportional or a non-proportional basis. In the former case, as a result of an in advance agreement, the reinsurer is accepting a certain percentage share of the premium income and claim liabilities that a primary insurer underwrites. On the contrary, in non-proportional reinsurance the reinsurer is only liable for losses exceeding a pre agreed level. More information on specific types of facultative and obligatory reinsurance can be found in Swiss Re (2021),Warr (2016*b*), and Bugmann (1997). Finally, the profitability of the reinsurance industry presents cyclical elements and in what follows we describe this phenomenon.

1.1.6.3 Reinsurance cycle

Both the primary insurance industry, and the reinsurance market are subject to cycles of low and high profitability alike. In the former case, we speak of a soft market whilst in the later case of a hard market. Various works such as these of Cummins et al. (1991), Cummins & Outreville (1992), Meier & Outreville (2003), and Cummins et al. (2021) have showed interest in this phenomenon.

Soft market In a soft market, there is plenty of supply for reinsurance and purchasing it is less costly. Insurers can take this advantage and increase their own underwriting capacity. Generally, a soft market is experienced when it has been long since the occurrence of devastating events that would result in a very high amount of claims.

Hard market On the contrary, normally straight after a big disaster, many industry players may face financial scrutiny or even bankruptcy and as a result a hard market appears having the opposite characteristics; high reinsurance premiums, and tight policy terms reflecting the unwillingness of the market to accept new risks or in other words the power of the few reinsurers that have been left that can now dictate prices.

It is worth mentioning that reinsurance is not the only means by which the primary insurers can transfer risk and some of the alternatives are discussed straight after in Chapter 1, Section 1.1.7.

1.1.7 Insurance-linked securitisation

Apart from the traditional reinsurance, there are various risk mitigation strategies allowing insurers, and reinsurers to transfer their risks directly in the capital markets. Techniques which are used to securitise cashflows based on insurance risk fall under the term insurance-linked securitisation (ILS).

Background details As Nowak et al. (2014) describes, ILS was developed in 1990s, a time when the property and casualty insurance has been struggling post the occurrences of various catastrophic events in the United States, such as Hurricane Andrew and Northridge Earthquake in California. The associated losses for the insurers and reinsurers have been so large that concerns were raised with respect to the capacity of the reinsurance market to absorb them all alone. Then the idea of using the capital markets as a separate source of risk bearing capital was born as the amount of funds existing there is multiple times bigger than in the reinsurance markets. From an investor's perspective, a major benefit is that investment in ILS provides high returns and the fact that insurance-linked events exhibit low correlation with other traditional investments in the financial markets.

Market state ILS market currently accounts for more than 75 billion dollars of the nearly 600 billion dollars of global reinsurer capital and since many transaction happen on an over the counter basis the actual number can well exceed this number, see PwC (2021). Traditionally, the strongest jurisdictions for ILS have been Bermuda, Ireland and Gibraltar. Nevertheless, ILS has a growing importance in the UK market and both

PRA and FCA are in close collaboration with the London Market Group[6] to construct a set comprehensive regulations for the implementation of a strong ILS regime in the UK based on Solvency II provisions for Insurance Special Purpose Vehicles (ISPVs) and some additional features that can increase the appeal of the UK for such transactions, see HM Treasury (2016). Below, we briefly present some popular ILS products.

1.1.7.1 Catastrophe bonds

Hurricane Andrew and other major natural disasters happened in 1990s triggered the development of a new, for that time, risk transfer security called catastrophe bond or simply cat bond. The main reason for this development was to provide issuers with coverage for risks of extremely low frequency (less or equal to 0.01) and very large potential severity that could not otherwise be covered by traditional reinsurance contracts, see Guy Carpenter (2005). Catastrophe bonds are kind of instruments whose payoff is dependent on the occurrence of a predetermined event which may be related either to natural catastrophes or not – such as extreme mortality or longevity in life insurance sector, see Cummins (2008).

What differentiates catastrophe bonds from other bonds is a clause which allows the sponsor to take back the interest payments if the trigger is activated. Just because of this provision investors are compensated by receiving a higher interest rate which as we will see below it is funded via the premium that a SPV receives from the sponsor, see Finken & Laux (2009). In terms of lifespan, both one-year and multi-year bonds can be found. However, the feasibility of being protected for more than one year is one of the reasons of increasing cat bonds attractiveness compared to traditional reinsurance the last years, see Michel-Kerjan & Morlaye (2008).

With regards to the covered perils can be again either single or multiple. The latter case is very convenient for sponsors since there is no need for them to generate special deals for every single peril from which they seek protection. Finally, it should be mentioned that cat bonds are structured in various tranches reflecting in this way different levels of risk and return that satisfy investors' risk appetite. Catastrophe bonds is a very big area of interest in the current thesis, therefore more details on the structuring, characteristics, and pricing of this product are provided in Chapters 2 and 3.

[6] London Market Group is a market-wide body accountable for the representation of the insurance and re-insurance market in London.

1.1.7.2 Other ILS types

Except for catastrophe bonds, there are many other types of insurance-linked securities that could be used to provide additional capacity such as side cars, contingent capital, insurance loss warranties, and catastrophe risk swaps only to name a few. The idea here is not to exhaust the list of products that could be created for alternative risk transfer purposes but to give a short overview of some famous options.

Sidecars Another substitute to traditional reinsurance is sidecars. The latter are special reinsurance structures with limited life span, created by other reinsurance companies in order to transfer risks entailed in their book of business and gain a return on them. Investors should put capital in the vehicle so that in case a particular catastrophic event occurs, claims can get paid.

One of the key benefits of sidecars is that except for giving access to extra capital directly, the time horizon is fixed and short (usually no more than 2 years) whilst enter and exit is an easy task. The latter gives the opportunity to use the sidecar when the reinsurance market is hard in order to access capital directly and in lower prices than otherwise and to leave when the premiums start to fall again, see Greenwald (2006) and NAIC (2021).

Moreover, if the reinsurers who transfer the risk are large with a sound balance sheet, a sidecar transaction may be an appealing option since more investors can be attracted by the high quality of cedants' books. Of course, the opposite also holds, i.e. lack of information regarding the risks in which investors will be exposed to would investment in sidecars less attractive. Another drawback is that, unlike other insurance - linked securities such as cat bonds, sidecars do not provide a standard remuneration to investors and of course that is unknown at the inception of the contract, see Lane (2007).

Contingent capital Another ILS instrument is called contingent capital defined as an option converting debt into equity in case that a predefined trigger event occurs. This may happen if, for instance, the value of the buyer's equity falls in the market below a certain level helping the purchaser to survive under extreme financial distress conditions such as those arising after a severe natural catastrophe, see see Vermaelen & Wolff (2010).

Contingent capital facilities are beneficial for (re)insurers since they provide a cushion in case that their financial results are influenced badly due to the occurrence of particular events, see The Economist (2009). Another benefit of the product is that it can assist in meeting the capital requirements that regulators and sometimes the market itself imposes.

In particular, sometimes it may be preferable to keep reserves to a satisfactory level by having exchangeable debt rather than facing the opportunity cost of retaining capital in order to face the worse-case scenario. Contingent capital is also a good way to recapitalise without forcing governments to pay much for “too big to fail” institutions. Moreover, from a supervision perspective, the fact that the exercise of the options may create a diluting effect for the shareholders, may act as a disincentive for (re)insurers to do the conversion without this being really necessary, see Maes & Schoutens (2012).

Nevertheless, in other circumstances, converting debt into equity can have the opposite effect as it can increase moral hazard. This is because (re)insurance companies, knowing that their debt positions are covered can become more prone to extensive borrowing, see The Economist (2009). Furthermore, systemic risk is another issue of concern as regulators and financial institutions have not agreed yet whether contingent capital should be considered as debt or equity.

Industry loss warranties Industry loss warranties (ILW) is another popular ILS solution. An ILW is a reinsurance or derivative contract that pays out when the financial losses experienced by an industry exceed a specified threshold. That said, the trigger event that activates the payment is predominantly related to the amount of losses suffered by the insurance industry as a whole and not the one suffered by the (re)insurer who buys the protection, see Ishaq (2005).

A significant benefit of ILW is low transaction costs, short implementation times, no credit risk given that these products are fully collaterilised, and that they are normally subject to the same regulation as reinsurance, see Gatzert & Schmeiser (2012). Another advantage of ILW is that moral hazard, i.e. a change observed in the behaviour of the insurance buyer after purchasing protection because they have no longer motives to take all the necessary measures in order to avoid losses, see Parsons (2003), is minimal. The reason why this happens is that an ILW is typically based on a parametric index and thus it cannot be affected in a great degree by the losses of a single (re)insurer, see Gatzert & Kellner (2011). However, one of the most important drawbacks of ILW is that they involve the so called basis risk. The latter is because there is always a probability that the losses suffered by the individual reinsurer can be larger than those recorded in the market for the pre-agreed industry loss threshold to be reached.

Catastrophe risk swaps An alternative mechanism that insurers and reinsurers use in order to manage natural disaster exposures is risk swaps. The latter is a reciprocal agreement in which different kinds of catastrophic risk can be exchanged between counterparties. One of the most important advantages of cat swaps is that enables reinsurers to do business without having to keep as much equity capital as it would be required otherwise, whilst it helps the reinsurer to diversify their portfolio as the swap concerns risks which are not correlated, see Cummins (2008). In addition, the costs associated with the transaction are insignificant especially in the case when many deals are made with the same counterparty. Also, since there is no money transfer until the trigger event happens, the present level of expenditure decreases considerably.

On the other hand, there are also disadvantages regarding cat swaps use. First and foremost, parity is required and in order to accomplish that modelling of the risks in questions should be absolutely precise. In addition, the basis risk involved in cat swap transactions compared to other insurance linked securities is higher especially in case of agreements that an index is used as basis for the trigger and finally the probability that the counterparty may go bankrupt is significant since catastrophe risk swaps are a non-collateralised deal type. After having presented some popular ILS products, we shift our focus to what make ILS distinct from reinsurance.

1.1.8 Some differences between reinsurance and securitisation

ILS differ from traditional reinsurance contracts in various ways and here we provide a brief summary based on some viewpoints shared in Gorge (2009).

In particular, starting from structuring, in reinsurance the underlying insurance risk is seen as liability as opposed to ILS where it is seen as an asset. Moreover, the cost associated with structuring a reinsurance agreement is most of the times much lower than those associated with ILS issuance. Furthermore, in reinsurance it is feasible for the reinsurer to fail to pay claims when due or at all because of several reasons such as inadequate reserves, the effect of very large and unexpected claims after a big loss event etc. Such a risk is not usually present in ILS as they are often fully collateralised.

Moreover, for some ILS products, there is secondary market meaning that there is a clear exit strategy if the investor decides to do so. Finally, if one compares the two markets from a regulation angle, they will realise a big difference. In particular, once a risk is written by an insurer, the latter remains liable to their policyholder even if the risk in question has already been transferred. Nevertheless, when it comes to other institution types within the financial sector this restriction does not apply as banks, for instance, by selling debt sign away any liability associated with the risk sold. As a result, the primary insurer under a reinsurance contract knows exactly who holds the ceded risk. On the contrary, in ILS transactions is more difficult to do so as the underlying philosophy is to resell the risk transferred and not to hold it.

1.2 Statistical learning essentials

The second part of the Preamble deals with statistical learning which is a set of algorithmic methods aiming to address the problem of discovering, otherwise called "learning" a function from data, see James et al. (2013). Here, we provide some key categorisation of the research tasks that it entails. Furthermore, we describe the representative statistical learning workflow that is normally followed by a researcher before presenting some broad information regarding statistical learning models that are relevant for the main research studies of this thesis in Chapters 3, 5, and 7. Finally, we provide some information regarding the application of statistical learning in actuarial science, something that Wuthrich & Merz (2021) calls actuarial learning, in order to showcase the importance of statistical learning in solving insurance problems.

1.2.1 Statistical learning and typical tasks

As evidenced in multiple statistical learning textbooks, see for instance Friedman et al. (2001), Anzai (2012), Murphy (2012), and Zhou (2019) there is a general consensus that one can distinguish among three types of statistical learning tasks namely supervised learning, unsupervised learning, and reinforcing learning. A brief description for each of them follows.

1.2.1.1 Supervised learning

The aim of supervised learning is to learn a function from a data set $\{(\mathbf{x}_n, y_n), n = 1, 2, \dots, N\}$ which is comprised by $n = 1, \dots, N$ distinct input-output pairs, so that this function is able to predict the response y' given a new input $\mathbf{x}'$ as accurately as possible. Because for each input, the output is known, such a data set is called a labelled. By input, we mean a $p = 1, 2, \dots, P$-dimensional, random vector of features, often called attributes, predictors, covariates, or independent variables, denoted by $\mathbf{x}_n = (x_{1n}, x_{2n}, \dots, x_{Pn})$ being an element of $\mathbb{R}^p$. The input vectors can be anything from numbers, to an image, a time series, etc, see Murphy (2012). The output, called response or dependent variable, denoted by y_n is indexed by example number $n = 1, \dots, N$ and once again can take any form, however most of the times it is either a real-valued scalar in which case the supervised learning task is called regression, or it is a categorical or nominal variable from some finite set $y_n \in 1, ..., C$ in which case we speak of a classification task. All of the research problems addressed in this research, see Chapters 3, 5, and 7 involve supervised learning.

1.2.1.2 Unsupervised learning

In unsupervised learning, the approach is different based on the fact that our data set consists only from inputs therefore being of the form $\{\mathbf{x}_n, n = 1, 2, \ldots, N\}$ and the aim is to discover knowledge, or otherwise called patterns, without knowing for what kind of the latter we are searching for. Examples of unsupervised learning tasks include clustering where the goal to discover groups of similar examples within the data, and density may be to discover groups of similar examples within the data, where it is called clustering, or to determine the distribution of data within the input space, known as density estimation, or to project the data from a high-dimensional space down to two or three dimensions for the purpose of visualisation.

1.2.1.3 Reinforcement learning

Reinforcement learning aim is to learn what course of action needs to be taken in a given circumstance in an effort to maximise a numerical reward signal, see Sutton & Barto (2018) and Russell & Norvig (2010) for a comprehensive review. As opposed to supervised learning, the model is not provided with instances of optimal outputs but it has to find them via means of trial and error and observe which action leads to a higher reward. To capture the idea of reinforcement learning one can think of a baby who develops knowledge by interacting with their environment trying things out without having instructions on how to do it - something that is referred to as a closed-loop in the reinforcement learning jargon. Some further characteristics of reinforced learning include that the outcome of actions, as well as reward signals, occur for longer duration in a sense that a present action has an effect not only the reward on the current step but also on the reward of all of the following steps.

1.2.2 Statistical learning workflow

The traditional statistical modelling process involves three steps namely model identification also seen as model selection, estimation, and prediction, see Box et al. (2015) and McCullagh & Nelder (1983). In statistical learning though the aforementioned steps are altered to some degree to include the following processes namely data collection, data cleaning and data pre-processing, model class selection, choice of objective function, providing solution to a a convex or non-convex optimisation problem, and model validation, see Wuthrich & Merz (2021) based on which we provide the following summary.

Data Given that the goal is to "learn from data", the biggest part of the statistical learning process relates to it. Starting from collection, this can range from being an easy to challenging task depending on the public availability of information. As we will see later in Chapter 3, data scarcity can bring difficulties in the modelling procedure. After acquiring data, it is common for the researcher to perform an exploratory data analysis, visualise, assess data quality, and pre-process the data. This data related preparatory step can be very costly with respect to the time that the researcher may need to contribute as a proportion of the overall statistical learning process. However, it is of paramount importance as such exercise builds the researcher's understanding regarding the data in hand and assists in the formulation of both the research question and model.

Model class Given a set of data which is ready to use for modelling purposes, the research shall choose the model class which is deemed to be the most appropriate based on the ability to answer their research question. There is descent flexibility in what is meant by model class here ranging from a stochastic data model to a total algorithmic model, see Breiman et al. (2001) and Shmueli (2010) for a distinction between these two modelling philosophies.

Briefly, in the case of a stochastic data model, the researcher assumes that the data generative mechanism is described by this model. On the other hand, in an algorithmic model, there is no such assumption, and it is a viable option for both big and smaller data sets if the interest is to extract as much information as possible from data. Moreover, in the algorithmic culture, predictive performance key with the model as such being much less important provided accuracy in prediction. Despite these differences, for the researcher whose goal is to use data in order to give solutions to problems, it may be beneficial to create synergies by utilising both modelling philosophies to the degree that they are relevant. Indeed, recently there is an increasing number of works which advocate towards narrowing the perceived gap between the these two alternative modelling perspectives, see Neufeld & Witten (2021) and Imbens & Athey (2021). In Chapter 3 of this thesis, we also address and show some benefits that can be enjoyed by exploring an algorithmic viewpoint on the top of a traditional data model.

Objective function After specifying a model class, a definition of a rule based on which a model class candidate will be picked for the data in-hand is required. There is a tendency for this rule to be an objective function for example a loss function whose goal is to quantify the degree at which our model is misspecified. That said, it is worth mentioning that the choice of the objective function can be easier or less easy to find depending on the learning tasks that one chooses to use, see Zhou (2019). For instance, in supervised learning, where the goal is to predict sufficiently well on unseen data, the choice of an objective function may be relatively more straightforward as the true value of the response variable is known and thus one could assess how good a predictive model is by seeing how close the predicted value for the response is compared to the true value. In the case of unsupervised learning for example the choice of a performance metric is not that obvious as the researcher does not have an in advance knowledge of the patterns that are sought.

Optimisation problem Post choosing the objective function the need arise to solve a optimisation problem which aims to discover the best model within the selected model class with respect to the chosen objective function and data. This optimisation problem may be convenient at times to assume that is convex because there is a good level of comprehension of the way in which convex sets are structured and as a result numerical methods can be utilised.

Nevertheless, with the vast increase of the amount of data available now-days, there are many applications where formulating a research problem in a non-convex manner may be naturally much more suitable in capturing the structure of the research problem. Of course, the researcher shall keep in mind that working in a non-convex environment is inevitably associated with a higher level of computational complexity than in the convex case. Some typical domains when a non-convex formulation may be advantageous is high dimensional problems in the area of signal processing or bioinformatics. In the current thesis, the main focus lies on convex optimisation however the interested reader can find an extensive review on the topic of non-convex optimisation in Jain & Kar (2017).

Model validation The last stage in the statistical learning process is the validation of the model derived in the previous steps. The idea here is to assess the goodness of model fit for the data in hand, the ability of the model to predict unseen data, and overall explore the possibility that there may be a more appropriate model than the one in-hand to work with. In case that in this stage the results are not adequate, the researcher may need to repeat the process from the first step, i.e. data and contemplate whether the data-pre processing for example was not done in a meaningful way, or need to include additional variables etc.

With respect to the performance of any given model, it is worth mentioning that there is not clear consensus of when a model is good or not, i.e. it really depends on the needs of the researcher for the problem in question. That said, different researchers may differ in their expectations regarding learning performance. A typical move is for the researcher to evaluate and estimate the performance of various models then reach a decision or possibly select "the best" out of the personal criteria of the researcher available out of candidate models.

1.2.3 Discussion on some relevant models

We believe that it is important to discuss the general concepts of some key methods that we use in Chapters 3, 5, and 7 and share some ideas regarding how they have been applied in multidisciplinary settings based on literature other than those indicated in this thesis. In most cases, these models are already defined in the aforementioned chapters, thus to avoid repetition, we redirect you there for notations and formal definitions.

1.2.3.1 Random forest

Random forest, see Breiman et al. (2001), is a supervised statistical learning model deployed to solve prediction problems of classification or regression types. The model is comprised of a number of de-correlated decision trees, each grown from a different sample which is taken out of the original data with replacement. The prediction is derived either by taking the the majority vote (in the case of classification) or by taking the average (in the case of regression) among all trees. The rationale behind the model, as well as the model itself, is explained in detail in Chapter 3. However, we consider beneficial to showcase the increasing research interest in random forests for several applications in insurance. That said, the literature presented below is certainly not exhaustive but it gives a flavour of different problems that have been attempted to be solved using random forest across the insurance value chain.

One of the insurance areas where random forest has been applied is claim fraud detection, see Li et al. (2018). Except from direct financial losses, fraud can pose significant reputational damages to an insurer thus early prevention is of utmost importance. One of the insurance classes that faces a significant number of fraudulent claims is automobile insurance which includes a variety of wrongdoings such as misrepresentation when answering questions in an insurance application, overstating the insurance claims size, or faking accidents, just to name a few, see Insurance Information Institute (2021).

That said, Li et al. (2018) has performed a comparative analysis using various statistical learning models for classification tasks and found that random forest achieved better prediction performance and lower level of variance than the competing models. Furthermore, from an insurance offering side, random forest has been deployed for the prediction of future insured's behaviour when they consider various options for buying a new insurance product, see Guo et al. (2019). When compared to other popular algorithmic methods, random forest was shown to have an edge meaning that it could potentially used successfully in practice for insurance products recommendation.

Moreover, random forest has been used in insurance applications relevant to flood risk assessment, Mobley et al. (2021). In particular, in an effort to estimate flood related hazards across a big flood-prone areas of the Gulf of Mexico, it was shown that random forest has been successful in predicting flooding probabilities even when the prediction results have been compared to those derived from the current Federal Emergency Management Agency regulatory floodplain. Finally, one the actuarial side, random forest can be used for predicting the insolvency of insurance companies Rustam & Saragih (2021) and for calculating the claim severity element of non-life insurance premiums, see Staudt & Wagner (2021), where it was shown that random forest can perhaps be a powerful tool for predicting right-skewed claims sizes.

1.2.3.2 Finite mixture models

Finite mixture models is a famous statistical modelling method. Their popularity is due to their flexibility and extensibility for the approximation of general distribution functions semi-parametrically whilst taking into account unobserved heterogeneity, see Tzougas et al. (2014). In particular, Peel & MacLahlan (2000) provide a very thorough review of the history, formulation and interpretation of finite mixture models thus we direct the interested reader there. In brief, traditionally, mixture models have been applied to data having inherently a group-structure with the main modelling goal being to explore it - we call this clustering, see Bishop (2016). However, we see that with time mixture models have been significantly used for inference purposes, apart from clustering, and in particular for modelling unknown distributional shapes on a semi-parametric basis.

As we will see in Chapter 5, a standard way to represent a finite mixture model is to consider that $\mathbf{X} \triangleq \{X_i\}_{i=1}^{\nu}$ is a sample of independent and identically distributed (i.i.d.) random variables from an n-component finite mixture distribution with density function

$$f(x|\mathbf{\Xi}) = \sum_{z=1}^{n} \pi_z f_z(x|\boldsymbol{\theta}_{\boldsymbol{z}}),$$

where $\Xi = (\boldsymbol{\theta}, \boldsymbol{\pi})$, with $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_n)$, where $\boldsymbol{\theta_z}$ denotes the parameters of the z^{th} density function $f_z(\cdot)$, and where $\pi^T = (\pi_1, \pi_2..., \pi_n)$ is the vector of component weights, with π_z the prior (or mixing) probability of the component z, where $0 < \pi_z \leq 1$ $\forall z \in \{1, 2, ..., n\}$ and $\sum_{z=1}^{n} \pi_z = 1$ holds. This is also the angle we take in this thesis.

Apart from this simple finite mixture model representation, there are certainly a variety of other finite mixture models. A helpful extension is to consider that $f_z(\cdot)$ does not only depend on parameters θ_z but also on a set of covariates via a multinomial logistic regression such the one described in Rigby & Stasinopoulos (2010). In this case, the simple finite mixture model form presented above will become

$$f(x|\Xi, \boldsymbol{g}) = \sum_{z=1}^{n} \pi_z f_z(x|\boldsymbol{\theta_z}, \boldsymbol{g_z}),$$

where $\boldsymbol{g} = (\boldsymbol{g}_1, \boldsymbol{g}_2, ..., \boldsymbol{g}_n)$ are the explanatory variables.

Apart from applications in insurance, which we examine in the current thesis, there are various other disciplines in which finite mixture models are used, and having said that, Peel & MacLahlan (2000) groups them in the fields of agriculture, astronomy, bioinformatics, biology, economics, engineering, genetics, imaging, marketing, medicine, neuroscience, psychiatry, and psychology, to name a few. For the interested reader, in Chapter 5, we provide a handful of references that have used various forms of finite mixture models across some of the aforementioned disciplines.

1.2.3.3 Multivariate count regression models

Regression models for multivariate count data are well regarded in the literature for allowing to make inferences regarding any statistical relationship that may exist between multiple response variables given a set of explanatory variables. There are several means by which the generalisation of univariate count regression models to their multivariate peers can be achieved and this is demonstrated by the range of relevant literature across multiple disciplines, see Winkelmann (2008), and Cameron & Trivedi (2013) for a detailed presentation of various advancements in multivariate count data modelling and Jeong et al. (2021) for a thorough and up-to-date literature review which we summarise below.

Over the last decades, the literature in the area of multivariate count data modelling can be classified into three distinct categories namely multivariate Poisson models, multivariate mixed Poisson models, and copula-based models. With regards to multivariate Poisson models, part of the literature has been aiming at developing multivariate

Poisson models which could capture positive correlation between variables offering the additional merit of closed-form densities, see Jung & Winkelmann (1993), Ho & Singer (2001), and Kocherlakota & Kocherlakota (2001). This has been mostly achieved by using the trivariate reduction method, see Kocherlakota (1992) for further details. Except for the trivariate reduction scheme method, Lakshminarayana et al. (1999) built a bivariate Poisson model which accounts for negative, zero or positive dependencies. Later, Karlis & Meligkotsidou (2005) extended the multivariate Poisson model to allow for broader covariance structures among the explanatory variables.

With respect to the multivariate mixed Poisson models, they constitute a substantial model class where dependence structures and overdispersion can be flexibly accounted for via adding variables which capture unobserved heterogeneity in the independent Poisson marginal probabilities which are following few mixing distributions, see Cameron & Trivedi (2013). As it is described in Jeong et al. (2021), there are two general categories of multivariate mixed Poisson models categories seen in the literature. The first one allows for positive dependencies between variables and normally involves a shared random effect which follows a univariate continuous mixing distribution, see Stein & Juritz (1987), Stein et al. (1987), Kocherlakota (1988), Munkin & Trivedi (1999), Gurmu & Elder (2000), Ghitany et al. (2012). As we will see in Chapter 7, this is the part of literature towards which we make a contribution. The second group of multivariate mixed Poisson models account for both positive and negative correlation structures with several random effects following a multivariate continuous mixing distribution, see Aitchison & Ho (1989), Cameron & Trivedi (2013), Munkin & Trivedi (1999), Chib & Winkelmann (2001), Park & Lord (2007), Ma et al. (2008), El-Basyouny & Sayed (2009), Aguero-Valverde & Jovanis (2009), Zhan et al. (2015), Silva et al. (2019), and Chiquet et al. (2020).

Lastly, the copula-based multivariate count data modelling methodology is more recent compared to those described in the two previous paragraphs. The main idea is that the multivariate count distribution can be seen through the lenses of a continuous copula distribution coupled with discrete marginals, see Jeong et al. (2021). Some representative works on regression models using copulas at their core are these of Lee (1999), Cameron et al. (2004), Nikoloulopoulos & Karlis (2010), Nikoloulopoulos (2013), and Nikoloulopoulos (2016). One of the key benefits behind using copula-based regression models is that they can capture both positive and negative dependence structures and in this context we direct the interesting reader to Chen & Hanson (2017) where a comprehensive review is provided regarding the employment of copulas for specifying correlation structures. It is worth mentioning, the combination of copula-based regression models with discrete marginals has the drawback of increasing the computational cost considerably especially when the dimensionality of the model surges, see Zimmer & Trivedi (2006), and Genest & Nešlehová (2007) for a discussion.

1.2.4 Deployment of statistical learning in the insurance industry practice

The current and future role of statistical learning in insurance is important given the breadth of interest in business applications across the whole value chain including pricing, underwriting, claims management, fraud prevention, and customer service just to name a few, see Accenture (2018). A recent report from McKinsey & Company (2021) shows that the main trends shaping the use of statistical learning in insurance are the data wealth arising from connected devices, the rise of physical robotics, the open-source ecosystems, and rapid advances in cognitive technologies. That said, the pace and extend of the adaptation differs across various insurance firms functions, see Swiss Re Institute (2020). For indicative purposes, in what follows, we focus on insurance pricing which is the main interest of this Thesis.

In an insurance context, statistical learning deployment has been relatively fast for the more common supervised learning models such as generalised linear models fitted to historical claims and premiums data, see McCullagh & Nelder (1983) and Wuthrich & Merz (2021) for more on this model class and its use in insurance respectively. However, in the daily practice insurers are relatively hesitant, at least until now, to embrace a more flexible and therefore less interpretable models, for actuarial work, see Gareth et al. (2013) for more details on the trade-off between flexibility and interpretability using several statistical learning models.

Such reservations are mainly due to the difficulty associated with explaining models with rather opaque operations to regulators and non-technical internal stakeholders. That said, regulators do recognise the efficiencies that statistical learning can bring in the insurance and finance sectors in general Financial Conduct Authority (2019), however they would rather prefer the move to be more gradual so that new governance and controls can be put in place. Nevertheless, a challenge here is that the pace that any regulatory changes happen much slower than the technological progress and data availability. Some insurers may be using more advanced statistical learning models as an internal tool on the side to supplement with insights the traditional approaches they have in place. This is something that we also discuss in Chapter 5.

Chapter 2

Some remarks on catastrophe bond pricing

Chapter 2 introduces concepts relevant to pricing formation in the primary non-life catastrophe bond market and it is the stepping stone to Chapter 3. As a starting point, we describe how a catastrophe bond deal is structured - a topic which is not addressed in Chapter 1, Section 1.1. Next, we explain the notion of catastrophe bond pricing and later discuss some important factors which are taken into account when determining the issuance price of catastrophe bond. Since catastrophe bonds involve an underlying insurable risk, and for the sake of completeness, Chapter 2 ends with an Appendix presenting a general pricing formation methodology for insurable risks.

2.1 Catastrophe bond structure

As we have seen in Chapter 1, Section 1.1, catastrophe bonds are insurance-linked financial instruments, first developed in 1990s, in an effort to provide additional capacity to the reinsurance industry post mega-disasters. For illustrative purposes, in Chapter 2, Figure 2.1, a simplified version of the structuring process is presented below based on Risk Management Solutions (2012).

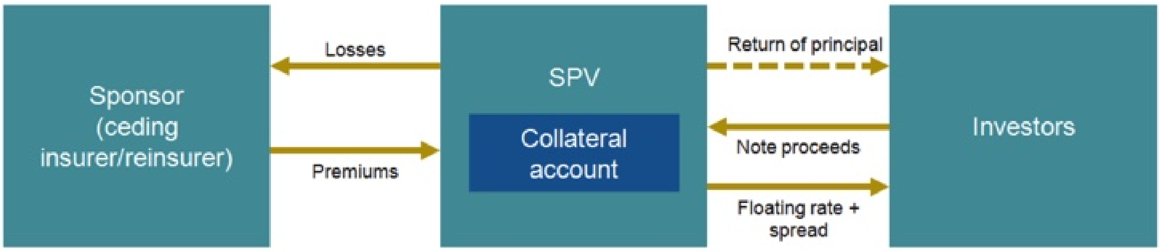


Figure 2.1: Structure of a typical catastrophe bond transaction, see Risk Management Solutions (2012).

The securitisation of insurance risk in a form of a bond is initiated by a sponsor or cedent in reinsurance terminology who can be an insurer, reinsurer or even a government with a purpose of transferring some of their catastrophe risk exposure such as wind or earthquake to name a few, to the capital markets.

For this to be achieved, a Special Purpose Entity (SPE) or otherwise called a Special Purpose Vehicle (SPV) is established which then enters into a reinsurance contract with the cedent. The SPE receives periodical payments called premiums in exchange for future protection provided to the cedent via means of the insurance-linked bonds which SPE issues. After the issuance, the SPE sells these securities to investors and receives principal amounts in return which are kept as collateral into a separate account and invested in low risk funds.

As a remuneration for the risk that the investors assume, the latter receive periodical payments called coupons which include the premiums that the SPE previously earned and the interest made on the principal amounts. If a qualifying event occurs, SPE liquidates the collateral and pays back the cedent for their incurred loss in a way that contract terms indicate. In case of no qualifying event, then the liquidation of the collateral happens at the expiry of the catastrophe bond when is returned to the investors.

2.2 The notion of catastrophe bond price

The notion of catastrophe bond price is not as straightforward as in a traditional zero-coupon bond setting where the nominal value of the bond equals its price. In particular, the price of a catastrophe bond is subject to multiple meanings depending on the context in which it is used. Based on discussions with industry participants, we shall distinguish between three empirical catastrophe bond price categories namely launch price, transaction price, and secondary market price, see Nowak et al. (2014). That said, a short description for each of them follows.

Launch price Following the structuring of the transaction, determination of specific design characteristics and risk modelling analysis, see Chapter 2, Section 2.3 for the latter two, the cedent presents a precise risk-return deal profile, including a price guidance range, to investors for reviewing. Based on this information, their current portfolio consistency and factors affecting their perception of the true risk involved in the transaction, investors give a quote of what price they are willing to pay for a specific size of insurance coverage liability. At this stage, every investor can come up with a different price for certain level of participation in the transaction and this price is usually called as initial offering price, or launch price.

Transaction price The final transaction price usually differs from the initial offering one and it is the same for all investors. For instance, assume that a reinsurer is in need to cover potential future cat losses of 100 million dollars size. Every investor can buy the portion of risk they want until they reach this 100 million dollars threshold. However, at this point in time, the price for each and every investor is the same - obviously denominated for the deal "slice" they buy.

Secondary market price If a catastrophe bond is already issued and traded, its price in the secondary market will not be the same to the one in the primary market, i.e. when the catastrophe bond was first issued. Actually, this price difference and can be large over time due to various forces including news.

In this thesis, the interest lies on the initial offering price stage in the primary market for catastrophe bonds and from now on this is what we mean when using the notion of price in a catastrophe bond setting. As we will see in Chapter 3, Section 3.1, the market convention is to perceive the price as the amount of interest earned by an investor on the top of the risk free rate namely spread. Consequently, in what follows the words spread and price may be used interchangeably. Next, we discuss some elements which we consider relevant in the process of determining a catastrophe bond price at issuance.

2.3 The importance of expected loss

Given that the payout of a catastrophe bond is linked to an insurable risk, an important factor when considering catastrophe bond pricing, at least at the initial offering stage, is the expected loss arising out of the potential insured disaster occurrence - a typical concept when calculating premium in general insurance, see Chapter 2, Appendix for further details. That said, expected loss quantification for catastrophe events is certainly non-trivial.

In what follows, we discuss some expected loss quantification considerations, the catastrophe risk modelling process based on which expected losses are calculated in the industry, and few sources of uncertainty involved in this process. It is worth mentioning that the computation of expected loss for non-life insurance-linked securitisation purposes is the same as the one followed in general (re)insurance. Therefore, here for the purposes of simplicity, we assume the simple case of a (re)insurer wanting to quantify the expected loss for a new (re)insurance policy. The ideas presented are mainly based on Lloyd's Market Association (2013), Lloyd's Market Association (2017), and this thesis author's working experience.

2.3.1 Quantification considerations

Prior to the acceptance of a new risk by a (re)insurer, an estimate of the potential loss arising after the realisation of an event due to an insured peril must be made. This estimation involves calculating the annual expected loss, otherwise called pure premium, which reflects the mean loss per year averaged over multiple years. The average annual loss is usually seen as the product of multiplication between two elements namely expected loss frequency, and expected loss severity. In particular, the annual expected loss in (re)insurance is calculated by utilising historical loss data, various modelling methods or a combination of both. Below, in providing some expected loss quantification considerations, we deem useful to differentiate the case of catastrophe risks versus non-catastrophe ones starting from the latter.

Non-catastrophe risks Risks which have high occurrence probability during one insurance period, i.e. typically a year, but low potential of adverse consequences can be priced based on past insurance claims performance without a problem. When risks of such non-catastrophic nature are included in an insurance pool, then the occurrence of an individual insured event has no effect on the probability of other events included in the pool happening from the same source and in this sense the risks are statistically independent. Empirically, when there are sufficiently many observations of random claims, fluctuations around the average claim become less significant. Then one can reduce the aggregate claims size by simply adding more risks into the insurance pool releasing that way capital needed to be held aside compared to the units of insurance coverage provided Nowak et al. (2014) and Chapter 1, Section 1.1.1 in this thesis for more details on risk pooling.

Catastrophe risks Past claims experience is not reliable source of information when it comes to risks of catastrophic nature, and thus the risk pooling argument cannot be extended to this risk category. There are many reasons why this is the case. First of all, insured events of low frequency and high severity are not statistically independent. One mega-disaster can trigger losses in multiple insured exposures at the same time, even across different lines of business, and thus contravening with the diversification-by-pooling concept. Secondly, historical catastrophe losses recorded are not enough for the development of a reliable events database because the return periods can vary from few years to centuries. Also, even if full records existed, their reliability would still be questionable. This is because the rare nature, and unpredictability of disasters does not change, no matter how many of them have occurred. That said, there is always a chance that the next coming disastrous event will be the worst that has ever happened in history.

Finally, efforts to create an index by linking loss related factors to past disaster claims cannot come into fruition easily because the former is not static. For example, when examining catastrophe risks, the quality of materials with which insured properties are built of can determine their vulnerability to a natural disaster, of course among other factors. Nevertheless, materials used in construction change from period to period. Similarly, the number of inhabitants in areas prone to catastrophes is a factor affecting insured losses but again in such a long return periods population cannot be stable. Under these conditions, it is clear that any projections regarding catastrophe insured losses by analysing past claims history and the employment of deterministic models in the process may not reliable. Probabilistic modelling is a requisite for the analysis of catastrophe risks Cummins & Mahul (2009). Next, we aim to present the stochastic catastrophe risk modelling process as used in the industry.

2.3.2 Catastrophe risk modelling

A catastrophe model is a computerised system whose purpose is to produce catastrophe loss estimates using a simulation methodology. The period in which these simulations expand reach thousand years in order to build the gap that the lack of historical data create. This leads to the development of large simulated loss event catalogue which renders the measurability of the simulated events occurrence frequency. Policy terms are also taken into account and the loss per event can be reduced to an estimate that reflects the insured loss. Based on the above, catastrophe modelling is the process of using a catastrophe model for achieving the aforementioned aims. In the same context, catastrophe risk modeller is a person who manipulates raw data and transforms them into an input that the model comprehends whilst being able to analyse the model output. In the following pages, we present some information regarding the users of catastrophe risk models, a typical catastrophe risk model components, and we identify some sources of uncertainty in the catastrophe risk modelling process.

2.3.2.1 Users

There are many parties across different sectors who use catastrophe models to enhance the precision of their business decisions. Nevertheless, among all players being interested in quantifying catastrophe risks, insurers are definitely at the top of the list. A catastrophe model is capable of evaluating catastrophic risk and assist the insurer to make informed risk management decisions related to risk pricing, mitigation, purchase of reinsurance, or even the issuance of alternative risk transfer products such as catastrophe bonds among others.

Moreover, (re)insurance brokers use catastrophe risk models extensively, given the fact that they are the first stakeholder to gain access to clients' data and by "running" these models, they can provide to (re)insurers a first insight on the risks to be transferred. Finally, at a governmental level, catastrophe models can assist regulatory authorities and emergency management agencies to develop strategies with regards to emergency response post a disaster. In order to understand how the expected loss is derived through the use of catastrophe risk modelling, it is important to comprehend the elements of a typical catastrophe risk model first.

2.3.2.2 Model components

A catastrophe model consists of various components namely hazard module, inventory/exposure data module, vulnerability module, and financial loss module which are presented below.

Exposure data module For a catastrophe model to operate, there is a need to "feed" it with data against which the risk evaluation will be made. Consequently, inseparable part of a catastrophe risk model is the exposure data module. The modeller receives from the underwriting team a Schedule of Values (SoV) which contains information regarding the location, values, and special building characteristics of the insured property. This document is used to feed the location file that will be inserted into the model. The location file normally includes codified information taken originally from the SoV. The role of the catastrophe risk modeller here is to transform this information into a meaningful input that the model will be able to "understand".

In particular, there are certain fields of information that go into the model which need to be filled in a spreadsheet by the modeller. That said, a list of all properties at risk along with geographic coordinates, postcode/zipcode, address, city, county, and country needs to be included. With regards to the location of the property to be insured, a process called geocoding is performed. In particular, given the address of the property at risk or another characteristic which describes its location, such as ZIP code in the US for instance, city, county, and country, it is feasible to assign geographic coordinates to the relevant property. Ideally, a SoV should include all of the aforementioned location information.

In case that the data provided to the modeller do not contain all the required details, the level of geocoding depends on the region where the property at risk is located. This is because different regions across the world have different minimum standards for the geocoding process to be considered accurate. For example, for an insured property in the UK, the postcode is very relevant piece of data as it points out the exact building. In the US, though, zip code is not enough and extra information needs to be included. Moreover, in the case of energy or power generation insurance business classes, the risks may only be possible to be located through given coordinates which of course does not allow the model to geode but there is no other way as these locations are outside the mapping of a city or even pipelines in the middle of the ocean. Now every location also needs to be numbered as depending on the policy terms some locations may need duplication to capture various catastrophe deductibles or sub limits that cannot be applied simultaneously.

Furthermore, each location at risk needs to be associated with an insured value which is normally the replacement cost, i.e. the amount that the (re)insurer would have to pay for the replacement of an insured asset (in case of property insurance) based on its present worth. This value should not generally come as a total but it needs to be split into the following categories in the original SoV namely buildings, contents, others such as improvements, and business interruption values (in case of a commercial property) for a specific time frame normally indicated in number of months. The split of values is very important for the faultless operation of other model components and especially for the vulnerability module as we will later see. For instance, in the event of an earthquake, history shows that contents are more susceptible to damage than the building itself.

Except from the spatial characteristics of the location, there are other details to be included in the exposure module. These include primary characteristics of the property at risk like construction, occupancy, and year built, or secondary characteristics such as roof geometry, shear walls, and custom elevation only to name a few. Such details are important for the vulnerability assessment of the property in question later on. Finally, the modeller needs to read the insurance contract and include detailed policy terms such as deductibles, and limits per location.

Hazard module In the hazard module, fundamental elements of physical hazard are linked to each simulated event. For example, a hurricane can be described by its landfall location, direction, and top wind speed whilst in the case of an earthquake, core components would be its epicentre, magnitude, and the frequency of certain magnitudes occurring. The details of the hazard module are combined with the exposure data, and any other details that the model has on notable features, such as soil type in the case of earthquake, for every single location. Finally, the impact of hazard on every location at risk is determined.

Vulnerability module Thanks to the hazard and exposure data catastrophe risk model components, it is feasible to calculate how vulnerable or susceptible to damage is a building at risk. In particular, vulnerability module consists of many vulnerability curves, and given the primary risk characteristics the most appropriate curve is automatically chosen. Vulnerability curves are normally attained by engineering reports but sometimes can be based on insurers' experience whilst they portray the response of a risk under various conditions. Let's take earthquake as an example; the level of damage that a building will suffer due to shaking is very much dependent on peak ground acceleration, i.e. the maximum ground acceleration that is observed during an earthquake at a certain location. Generally speaking, the higher the peak ground acceleration, the higher the value of the expected damage to the building. This positive aforementioned relationship is represented in a vulnerability curve.

It is also worth noting that the way in which quantification of vulnerability takes place is not standard for all models. There are times where the state of the property post the event is classified in a more descriptive way ranging from slight to complete damage. Nevertheless, for most of the models damage to the property at risk is related to a severity parameter such as peak gust wind speed. What is standard for all model vendors though is that the construction of damage curves happens separately for the following loss categories; buildings, contents, and business interruption. The first two categories represent direct losses, i.e. the cost that the insurer would have to bear for replacing or repairing the damaged property whilst business interruption is an indirect loss. The output of these computations is a damage ratio, which is then applied on the structure at risk.

Financial module Post allocating a damage ratio for each location that needs to be assessed, the catastrophe model takes into account any financial and insurance policy terms populated by the modeller in the exposure module stage. The information about policy terms is summarised in the contract file that a risk modeller prepares. The raw data are coming from the (re)insurance slip, see Chapter 1, Section 1.1.5.3, which is provided by the underwriter through the broker. This document lists all the policy terms; layer, policy level limits and deductibles etc.

Firstly, the ground up loss to each location is calculated and then the model takes into account policy terms and programme level conditions for every structure such as limits, sub-limits and deductibles producing the gross loss. The result of this process is the generation of an event loss table with which an insurer can evaluate the financial risk exposure to individual events. If one combines an event loss table with an exceedance probability curve, they can derive additional measures for the whole risk. Further information about the exceedance probability curve is provided below.

2.3.2.3 Model output

The most famous catastrophe model output is the Exceedance Probability (EP) curve. For a certain portfolio of locations at risk, the EP curve represents graphically the probability that a certain loss level will be exceeded within a particular time framework. For illustrative purposes, we present a graphical example in Chapter 2, Figure 2.2.

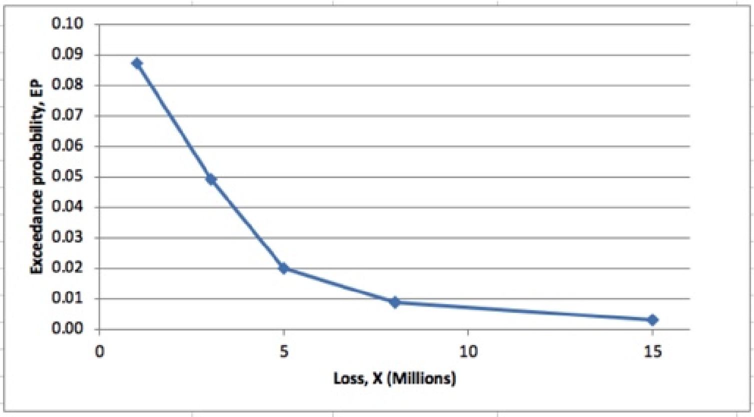


Figure 2.2: Example of an Exceedance Probability (EP) curve, see Casualty Actuarial Society (2020).

The importance of an EP curve lies in determining the size and distribution of (re)insurers' portfolio potential losses with extra attention paid in its right tail where the top losses are located. By looking at the EP curve, insurers are able to identify buildings categories and geographies where they believe it will be profitable to insure, whilst getting an idea of what the policy should cover and at what price.

It is important to keep in mind that every insurance company wants to ensure its solvency, see Chapter 4 for more details on reserving. Consequently except from pricing, the EP curve can also be used for determining risk transfer proportions either to the reinsurance or to the capital markets using alternative risk transfer solutions, see Chapter 1, Section 1.1.7, as a way to maintain the probability of insurer's default at an acceptable level. For indicative purposes, we consider an example assuming that an insurer provides commercial windstorm coverage in Palm Beach, Florida. The acceptable level of loss specified by the insurer for this area is 50 million USD dollars at 1 in 100 exceedance probability. Nevertheless, the EP curve for the Palm Beach, Florida portfolio is 65 million USD dollars at 1 in 100 exceedance probability meaning that the insurer must find ways to reduce their exposure. This can happen in various fashions such as by transferring part of the risk through the purchase of reinsurance or a catastrophe bond or simply by writing the risk under stricter policy terms through the introduction of limits and deductibles.

2.3.2.4 Sources of uncertainty

Catastrophe risk models are certainly important to insurers and other users in quantifying the risk associated to disasters. However, decision makers should keep in mind that there are many elements of uncertainties in the catastrophe risk modelling process meaning that an insurer cannot rely fully on cat-modelling output when making business decisions. The following list of uncertainties is not exhaustive and heavily relies on the working experience of the author of this Thesis. Nevertheless, recently there has been an ongoing academic interest in decision making under uncertainty when using catastrophe risk models in insurance, see for instance Roussos et al. (2021) for an insight of uncertainties in hurricane modelling.

Static events catalogue The development of a catastrophe risk model involves the generation of a fixed set of catastrophic events. The latter obviously represents only few states of the universe even if the scenarios number is satisfactory enough to give a representation of the underlying hazard, and practical enough to facilitate fast decisions from insurer's viewpoint. A crucial observation in the licensed models of famous risk modelling companies, i.e. RMS or AIR for instance, is that the events set used is not random. The latter means that the same events run every single time that the model functions. This simplification has the advantage of allowing for comparisons among risks and other firms which licence the same model vendor. Nevertheless, given the fact that these events reflect just a subset of all potential catastrophic scenarios that can happen, and in the eye of climate change, extra caution should be taken especially for events located at the tail of the distribution.

Non-modelled losses Non-modelled losses is an important source of uncertainty in the insurance industry overall and typically it occurs because of limitations in the model itself. For example, this may be the case when geographical areas or perils are excluded from the modelling assessment, or the insurance account of interest faces risks that the model is not able yet to consider. It is vital for key members in the business to be at least aware of any material missing information as this can impair their decision making process if this is relied on the model output.

That said, there are instances in the past where insurers faced considerable financial pressure just because they did not take into account potential discrepancies between real versus modelled losses. For instance, in Hurricane Katrina storm surge losses peaked in a very unexpected way and proved that the level of water surge can be much above the storm's landfall intensity compared to what the model had predicted. Managers sometimes seem confident about their ability to judge the materiality of missing exposures but overconfidence could lead to serious business mistakes that can jeopardise the solvency of the firm.

Quality of exposure data It is crucial for the model to receive data that represent the location for risk assessment in a best way possible. In case that there is absence of information, assumptions are made about missing key characteristics. This can be done either by the modeller or by the catastrophe risk model itself. In the case of very big data sets such as in binding authorities, mistakes are naturally more in number compared to those in an energy and power generation account where the sites to be insured are limited to one or two normally. Nevertheless, in the latter case characteristics such custom elevation and exact geocoding are super important especially for windstorm hazard investigation and not always easy to map as they are normally located in isolated or offshore locations.

Furthermore, sometimes, in SoV's there is no split between different type of values and it depends on the modeller's discretion how they are going to prorate these values in every value category. In the case of earthquake, contents value is very important as contents are normally more vulnerable than buildings. Another common exposure data pitfall is business interruption values to be provided for different (or even unknown) time frame to the one that the model accepts. In this case, it is again in modellers' discretion how to deal with this, nevertheless pro-rating is again the most common approach.

Another problem is that the quality of data provided by the broker can be incomplete. Usually, no modelling request can be accepted without SoV and insurance slip. Nevertheless, for renewals and in the absence of updated data, it is not uncommon to use last year data until the new schedule/slip is ready. This is very risky strategy as if an unanticipated big loss was to occur before the receipt of the updated SoV, this could be devastating for the insurer.

Finally, geocoding process entails lots of risks as there are again multiple approaches that a modeller can use to develop the location input and each of these could lead to a different modelled outcome. For example, running a UK account with postcode as location information could give reasonably acceptable results. If one though would use ZIP code in the US as main location information though, the resulting model output would be inaccurate and very risky for the financial health of the insurer in case that this would happen for a high hazard territory. For instance, a ZIP code level geocoding in a coastal area of Florida could swift the location at the postcode centroid producing as a result considerably smaller expected loss. Finally, there are countries such as Chile that do not offer lat/long resolution through online platforms making geocoding impossible. In these cases, a lat/long must be manually inserted decreasing the reliability of the model outcome. Added on this, is the difficulty to find these locations manually on the map, as the schedules are normally not in English which makes the task very time consuming and prone to human error.

Special risk characteristics One of the most serious factors of uncertainty in cat modelling is this related to how the characteristics of the hazard are being translated into damage for a certain structure at risk. Whilst special risk characteristics such as roof geometry are taken into account, they do not cause significant changes in a base vulnerability curve. Consequently, the selection of the most appropriate vulnerability curve by the model turns out to be highly generic. No matter the level of model sophistication, generalisations as to the assignment of a certain risk to one of the specified base vulnerability curves are inevitable. In order to moderate this uncertainty, it is important to pursue as minimum the sensitivity of the portfolio to the assumptions made by the vendors and perform testing to comprehend how material every aspect is in changing the curve. Nevertheless, in a fast paced insurance business environment modellers have time to do this only in periods when work-flow is very low and unfortunately it is perceived as a low urgency task.

Financial calculations When assessing the risk of an insurance contract, the modellers need to include financial terms into the model input such as deductibles, sub-limits, covered perils per location, and sometimes even reinsurance terms. This requires the modeller to have experience in understanding of the policy wording included in the slip which is certainly not straightforward and not always possible. Consequently, misunderstanding of policy terms can inject uncertainty in pricing process and when trying to predict future claims.

Moreover, when risk analysis happens at a portfolio level, all underlying exposure from both direct and reinsurance side of business need to be accounted accurately. Modellers must select the appropriate settings and options before they run the model. The majority of the models run in timetables that are practical from time perspective using various approximations whilst computing. This forces the output of individual stages in the modelling process into pre-determined distributions or deployment of sampling techniques making it very complex to make direct calculations.

For instance, consider an insurance policy that is set to provide coverage against wind damage only but the insured structure has faced damage from both wind and storm surge. With the current sophistication level of models, it is not possible to identify post the event occurrence whether the actual loss was caused by wind only or storm surge only or by a mixture of both perils. Also, demand surge, i.e. the increase in the cost of materials and labour post a disaster is another aspect that causes difficulties in translating the pure loss into a claim accurately. Even if most vendors include demand surge in their settings, none can know for sure how much the demand will actually surge, as politics may play a role in restricting the influx of labour and materials in the country hit by the disaster.

Moral hazard Another element of concern is related to the financial incentive of the underwriters to bring revenue in the business. For very long time, insurance business have been highly dependent on business relationship building and pricing decisions were mostly based on intuition and experience. Nowadays, risk modelling has by far more weight in making underwriting decisions as opposed to the past and sometimes this change in business approach comes with conflicts.

For example, in cases when the catastrophe risk model output disagrees with the underwriter's intuition to accept the risk, pressure may be applied on modellers to re-model the insurance account, by offering different interpretation to grey zones of data. This may happen in an effort of the underwriter to align the modelling results with the business network pressures.

Communication of Uncertainty Communication of uncertainty in catastrophe risk model output is vital for making wise insurance business decisions. Being over-confident or under-confident regarding catastrophe risk modelling results can be financially unhealthy for the insurer in the long term. Due to the complexity of the catastrophe risk modelling process, stakeholders who use the modelled output for decision making purposes may not be in position to judge the level of trust that they need to show in it. That said, sometimes, there is too much of reliance on the model, and the modeller being "right". With regards to the latter, it is worth mentioning that the way in which the modeller present the risk analysis results to major stakeholders in the company can influence the level of confidence of the latter in the reliability of the assessment. That said, psychological factors may influence the way in which a decision maker perceives the level of uncertainty in the catastrophe risk modelling process.

For instance, based on observation from personal working experience, when stakeholders are presented with loss numbers with many decimal points, they somehow tend to feel more secure and overconfident about the reliability of the results. From the other hand, when modellers state the assumptions, and uncertainties in the reliability of the modelled output in a very detailed way, underwriters may become more conservative or they may reach the point of ignoring the model altogether and price the risk based on their personal judgement.

2.4 Other important characteristics

Except from the expected loss of the underlying insurance risk in a catastrophe bond, there are other factors which could potentially influence the issuance price. As mentioned in Chapter 2, Appendix, the expected loss can be seen as the pure insurance premium on the top of which an additional loading is added to account for some variability. Thus, in what follows, on the top of the expected loss discussed in Chapter 2, Section 2.3, it is worth presenting some attributes that could potentially have some influence on the risk loading of a catastrophe bond price at issuance.

In particular, we deal with attributes influencing mainly investors' risk perception, and consequently, the catastrophe bond risk load and initial offering price. These characteristics are either linked to the design and contract provisions of the catastrophe bond deal or general conditions in the market which alternate investors' risk perception. That said, the direction in which the risk perception of investors is influenced by these factors may differ from investor to investor depending on their attitude towards risk and the current state of the portfolio. The ideas below are based on Nowak et al. (2014), Boyd (2016), and author's working experience and they are not a result of a statistical analysis. In Chapter 3, though we use some of these factors in a their ability to predict the price, not solely the risk load, of a catastrophe bond at issuance.

2.4.1 Timing

There are at least two ways in which the timing of a catastrophe bond issuance may influence its risk load. Firstly, a catastrophe bond which is issued in a time period of low disaster activity may have a lower risk load compared to the one of an issuance with similar characteristics straight after the occurrence of a mega-catastrophe. Obviously, such a distinction does not usually reflect a true change in the risk involved but perhaps a change of investors' risk perception driven by psychological factors.

Secondly, every issuance relates to a well specified deal size reflecting the amount of coverage that the sponsor of a catastrophe bond needs. Investors buy "slices" of the issuance up to the point that this pre-defined amount is reached. At a time when the participation potential in a given issuance is still high, the associated risk load shall be relatively small. This is because, from a cedent's perspective, there is more confidence that there will sufficient investors' demand to achieve the level of coverage needed. Also, from an investor's perspective there is more flexibility into buying the "slice" size that meets better their portfolio needs, and risk-return desired structure. However, moving closer to the exhaustion of the deal size, uncertainty increases and the associated safety load shall also move upwards.

2.4.2 Peril in conjunction with territory

The rarity of the peril from which a sponsor seeks for coverage, combined with the coverage territory may affect investors' perception regarding the level of a catastrophe bond riskiness. That said, one view is that risk loads of issuances covering perils in peak zones, i.e. areas with high disasters activity, tend to be higher compared to bonds with perils in non-peak zones. This is usually attributed to the diversification effect that rarer risks can offer to investors. In particular, by purchasing a catastrophe bond with an unusual underlying peril, investors can smooth the risk-return structure of their current portfolio. However, a different perspective is that catastrophe bonds with peak zone perils could offer more "certainty" as they are better understood, and "reliable" modelling methods are more likely to exist.

For instance, a catastrophe bond covering losses against US/North American hurricane is considered more of a "mainstream" product compared to a pure flood or wildfire peril in a less developed part of the world. Since US hurricanes happen every year, the availability of loss data, the quality of modelling techniques, and of course the experience of the cedent in insuring these risks in the first place, makes the issuance process more straightforward and increases investors' level of confidence. Even though losses can occur, investors can have a more informed indication of how much they can lose.

Another important aspect regarding the catastrophe bond underlying peril is whether the latter is of long or short tail nature. For example, securitising liability related risks, i.e. casualty lines, may result in a higher risk load because in this class of business claims may take longer to arise, sometimes even well after the expiration date of the coverage, whilst the estimation of loss magnitude is relatively difficult. From the other hand, risks belonging to property business class are more certain with regards to the time span between coverage start date and arrival of claims and the calculation of the expected loss can be perhaps considered more reliable.

2.4.3 Reinsurance cycle

Both catastrophe bond, and reinsurance risk loads alike are related through the forces that reinsurance cycle, see Chapter 1, Section 1.1.6.3, implies on the pricing of catastrophe coverage. In particular, after a period of severe catastrophe losses, the cost of reinsurance normally increases (hard market) and consequently cedents look into alternative ways to transfer risk, i.e. such as a catastrophe bond, in an effort to release capacity. This may increase demand for catastrophe bonds nevertheless the risk load rises in line to reflect the general market conditions. It is important to mention that the market cycle can affect either the whole market or a specific segment of it, depending on the region and peril that caused the loss. For example, a very active US hurricane season would increase the risk load for US perils. However, if no other serious disasters occurred in other parts of the world, the market for non-US perils may still be considered soft.

2.4.4 Trigger

One of the most intricate parts of catastrophe bond issuance is determining what can trigger principal and/or interest impairment to investors. The trigger always include an attachment which needs to be breached for investors to be liable to cover losses. Some popular trigger types are called indemnity, industry loss, parametric, and model ones respectively. In general, the difference among them lies on the level of correlation to the real losses of the ceding reinsurer.

In particular, in the case of a catastrophe bond with an indemnity trigger, the principal/interest payments will be suspended in line with the actual loss which cedent suffered after the occurrence of the trigger event. With an industry loss trigger, the suspension decision is linked to the level of losses sustained by the whole insurance industry as it is reflected through an industry wide loss index. In a catastrophe bond with parametric trigger what matters are not the losses of the individual cedent or industry but the physical parameters of the insured disaster such as specific wind speed or earthquake magnitude in Richter scale. Finally, catastrophe bonds with modelled loss trigger rely on cat claim amounts estimated by cat models.

Given the prolonged period of time associated with claims handling post a catastrophe, it is natural to argue that indemnity triggers would originate more uncertainty to investors as for what percentage of the principal will be lost in claims payments. Also, since indemnity triggers guarantee the coverage of actual losses incurred to the cedent, the process inevitably involve a higher degree of moral hazard which puts investors in a disadvantageous position. As a result, a higher safety loading is necessary and investors would require from the cedent to charge a more competitive price for taking this uncertainty. Obviously, the more irrelevant the actual losses are, the more attention should be paid in the structure and calibration of the cat bond. However, investors could be ready to pay a premium for avoidance of delays in the claims process.

The number of triggers is another factor which can affect risk load. A catastrophe bond can be constructed to have either a single or multiple triggers. In the instance of a single trigger, the first insured loss event exceeding the attachment point triggers the provision of coverage by the investors. For loss events that do not manage to reach the pre-agreed level, the cedent cannot suspend payments to investors as this attachment point acts as deductible.

In catastrophe bonds with multiple triggers, various loss events must exceed various pre-determined attachment points before investors held liable to cover any insured losses. This can be done in various ways. For example, sometimes the occurrence of a very specific first event leads to the trigger activation of a very particular second trigger event which will make investors liable in case that a pre-defined loss threshold will be exceeded. Alternatively, the trigger event can be single but the requirements which need to be met multiple for the activation of the coverage to occur. Based on the above, investors are slightly less prone to be held liable to pay claims when they buy multi-trigger catastrophe bonds as more conditions need to be met before they start losing money compared to coverage activation single trigger instruments. This means that multi-year catastrophe bonds may have smaller risk load.

2.4.5 Resets provision

Normally, catastrophe bonds are designed to offer coverage against disasters over the course of multiple years rather than for a single year as it happens for most of (re)insurance contracts. This multi-year coverage provision can cause complications in the risk quantification process, especially this of non parametric cat-bonds. The reason behind this is the possible size alteration of cedent's exposure base. In particular, the expected loss in these instruments is calculated based on the cedent's inventory at risk at the time of issuance. Nevertheless, the portfolio of the cedent, and thus their exposure, is not stable over the duration of the bond. This randomness of cedent's wealth can overexpose investors to risks they are unaware of and thus not compensated for.

Consequently, the presence of basis risk is apparent and grows in line with the time difference between the risk assessment and the change in the exposure. This problem can be mitigated through the so called "resets". This process entails regular re-modelling of the risk based on the updated exposure followed by adjustment of the trigger to preserve cat bond's loss probability at the level agreed when the latter was issued. In this sense, in an absence of number of resets provision, the risk load that needs to be charged needs to be higher to compensate for this basis risk and equivalently resets happening on an annual basis should lead to a lower safety loading.

2.4.6 Bonds liquidity

Catastrophe bond liquidity can also affect the risk load. Catastrophe bonds that have the potential to be sold fast in the secondary market shall require a smaller load as they are considered less risky. Investors benefit from holding liquid investments because of the versatility they offer for portfolio re-balancing, hedging, or simply obedience to rules regarding assets credit rating minimum acceptable level introduced by some investment funds.

Generally, one could say that a high credit rating, a relatively short coverage period, a well known covered territory, and a quite common peril could make a cat bond qualify as a liquid investment. However, whilst the first two factors are easy to determine, things are not as straightforward when it comes to peril and territory as depending on the current circumstances investor's perception of riskiness can change. For example, after a very active hurricane season with serious landfalls investors may feel that holding a Turkey earthquake based cat bond provides them with more opportunities to cash out without much loss in value. The latter risk though may be proved a very illiquid one, after years of low loss hurricane seasons.

Another factor that it relates to the liquidity of the market as a whole and thus the risk loads is regulation. In the early stages of cat bond market development, the purchase base of these instruments were specialist hedge funds and asset management companies. Nowadays though, there are jurisdictions where regulation allows for pension funds and other institutional investors to buy cat bonds. This means that issuances in areas with looser regulation with respect to this matter may enjoy a lower risk load.

2.4.7 Reputability of issuance participants

A reinsurer with a high credit rating and solid experience in insurance-linked securitisation could fairly result in a lower risk load. Reinsurers such as Zurich Re and Swiss Re are pioneers in securitisation of insurance risks and they have a proven record of successful deals placements. This creates a contract of trust between reputable reinsurance players and investors that affects the perception of catastrophe bond riskiness and risk load, and eventually the catastrophe bond initial offering price.

Similarly, a higher credit rating illustrates the cedent's ability to assess risk and pay valid claims without compromising the soundness of their financial performance. Once again this boosts investors' confidence around the quality of the cedent's risk selection practices minimising the necessity for additional loading. Cedent's trustworthiness is not the only one that matters though. In particular, the risk modelling company used for the risk assessment of the deal may also be important especially when the securitisation comes to periods of model upgrades. This is something that we will also see in Chapter 3, Section 3.1.

2.5 Appendix; On premium calculation principles

The price formation for an insurable risk follows some basic actuarial principles. This methodological framework is not very far away from the one employed in a catastrophe bond pricing setting thus its study becomes relevant. That said, we look into some main traditional premium principles used in non-life insurance along with their properties. For the purposes of this written work, the focus will be mostly on classical and economic premium principles based on the views of Gerber (1979), Deprez & Gerber (1985), and Bühlmann (2005).

Premium calculation in insurance is a topic of tremendous importance. Nevertheless, the fact that there is no standard way to perform premium computation makes the decision of appropriateness of the existing methods intriguing. The discussion starts by establishing the main categories of these actuarial principles. There are mainly three methodologies that actuaries utilise for the development of premium principles namely the classical, the characterisation, and the economic methods. Some brief description for each follows.

In the classical method, the actuary chooses a premium calculation method and then examines whether the chosen principle satisfies any good properties. On the other hand, the characterisation method requires the actuary to produce a list of desired properties first and based on them then to develop the premium principle that satisfies them. The economic method, which is often considered the most meticulous one of all, stems from adopting a specific economic theory and then quantifying the emerging premium principle. However, it is highlighted that in reality it is often difficult to distinguish among these methods, as there are circumstances where a premium principle originates from more than a single method. What is really important is to understand what constitutes such a principle.

2.5.1 Defining a principle of premium calculation

The basis of premium calculation in (re)insurance lies on assuming that fixed payments called premiums can recompense for fortuitous claims. A reinsurer who wants to price an accidental loss operates in a uncertain universe. The latter is described through a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which models a real-world insurance pricing process containing states of random loss occurrence. This random loss, say X is a bounded random variable. Then a principle of premium calculation is a rule, suppose R, whose role is the assignment of a real number, say P, to any bounded random variable, X. The latter mapping can be illustrated as follows.

$$P = R(X)$$

where X represents an insurable risk and P the premium attributed to it.

It is obvious that the aforementioned definition is quite abstract in a sense that no emphasis is given to the fact that the functional $R(X)$ is formed by distributions. Nevertheless, this generality serves a deeper purpose which is to reveal that both the payment that the insurer has to make post the occurrence of the insured event and consequently her gains are random. In particular, this principle says that for any insurable risk the insurer can provide a premium quotation P meaning that they are ready to accept P and as an exchange pay a random loss X. Since X is random, then the insurer's earning from such a policy is $P - X$ also random.

2.5.2 Properties of premium principles

Properties that are usually accepted as true in most of the premium principles, without necessarily being so, are the following.

Independence The premium to be charged depends only on the distribution of the risk X. In other words, the same premium is assigned to insurable risks that happen to be identically distributed. The property of independence presented in the previous section is a very important one in the general insurance industry. Nevertheless, when the underlying peril of a (re)insurance or cat bond transaction is a rare event the assumption of independence needs loosening. The reason behind this will be thoroughly explained later on.

Non-negative risk Loading For any X, $R(X) \geq \mathbb{E}(X)$ meaning that the premium charged should be at least equal to but most of the times larger than the expected loss. The inequality is quite intuitive as the expected loss is just an estimate whose reliability depend on multiple factors. Nevertheless, where there is equality it means that for all constant risks, a safety load is indefensible.

Invariance Take any insurable risk X and any constant, say c. Then a premium augmentation by an additive constant c is equivalent to the original premium plus this constant. This can be expressed as

$$R(X + c) = R(X) + c$$

Additivity Consider that X_1, X_2 are two risks which are independent then it holds that the premium of their sum equals to the sum of their premiums.

$$R(X_1 + X_2) = R(X_1) + R(X_2)$$

Thus, by combining two uncorrelated risks, the total premium should not be affected.

Iterativity Except from risk X consider now randomly another risk Y. Then according to this principle the premium can be calculated in two steps; first apply the rule $R(.)$ to the conditional distribution of X which is a function and thus a random variable itself. Then the application of the rule $R(.)$ to the distribution $R(X|Y)$ follows to derive $R(X) = R(R(X|Y))$. In other words, when risks X and Y are of arbitrary nature,

$$R(X) = R(R(X|Y))$$

holds.

2.5.3 Examples of premium calculation principles

This paragraph introduces some representative examples of premium calculation principles. These are presented below.

Net Premium Principle The net premium principle or otherwise called principle of equivalence says that the premium charged for an insurable risk equals the expected value of the random payment.

$$R(X) = \mathbb{E}(X)$$

The absence of loading is based on the assumption that if an insurer issues a large enough number of identically distributed and independent policies there is effectively no risk. This principle could be appropriate for life insurance but it is out of context in the case of property and casualty risks especially those of low frequency and high severity risks as they are not characterised by homogeneity.

The Standard Deviation Principle The standard deviation principle says that the loading is proportional to standard deviation. Specifically, for a parameter $\beta > 0$

$$R(X) = \mathbb{E}(X) + \beta\sqrt{Var(X)}$$

This principle is perhaps the most famous across property and casualty insurance lines. The reason behind its popularity is the its linearity with reference to a proportional change in (re)insurer's claims experience. Moreover, it is claimed that if the probability distribution of risks included in the portfolio is normal, then all premiums have the same probability of being surpassed by the related claims. Both arguments though are not appropriate for all types of risks and especially those of catastrophic nature as their distributions are highly skewed.

The Variance Principle The variance principle incorporates a security loading which is proportionate to the variance of the expected random payment. In particular, for a parameter $a > 0$ the following equation holds.

$$R(X) = \mathbb{E}(X) + aVar(X)$$

The variance principle is not as extensively used. The theoretical advantage lies on its alleged linearity with reference to the addition of independent risks even though there is no longer linearity with reference to a proportional change in (re)insurer's claims pattern.

Similarly with the standard deviation principle, it is improper for low frequency and high severity risks.

The principle of zero utility The concept behind this principle is that the insurer wants to charge a premium which is fair in terms of monetary units of utility. Consider that $u(.)$ is a function which positions options according to their utility to an insurer. This utility function is increasing and satisfies the property of concavity. The element w reflects the wealth that the insurer has under his procession before the issuance of the new policy. In this context, the premium principle $P = R(X)$ is derived from the following rule showing the fairness of the premium, i.e. the utility of the insurer is unchanged when provides the cover. The latter reflects more the situation where the insurer examines the provision of coverage to the insured but it can be easily adjusted for cases where the insurer wants to transfer the risk X out from his portfolio to a reinsurer or a financial market by simply changing the sign of P and X.

$$\mathbb{E}(u(w + P - X)) = u(w)$$

An important aspect here is that when the insurer's utility is exponential, the pricing equation can be solved explicitly. The result is then called the exponential principle, as shown below, where the level of premium charged increases as the size of the parameter a increases.

$$u(x) = \frac{1}{a}(1 - e^{-aX})$$

$$P = \frac{1}{a} \ln \mathbb{E}(e^{aX})$$

The generalised principle of zero utility The difference between this principle and the previous one lies on the nature of the parameter of wealth. Whilst previously, the fortune of the insurer has been taken as constant, here it is stochastic, W, and unlike earlier, it appears in the explicit solution of the exponential utility principle. Once again, with a change of sign in P and X the principle falls in the context of risk transfer.

$$\mathbb{E}(u(W + P - X)) = \mathbb{E}((W))$$

$$P = \frac{1}{a} \ln \frac{\mathbb{E}[e^{a(X-W)}]}{E[e^{-aW}]}$$

It shall be mentioned that when parameter a is small enough, the below approximation can be derived.

$$P \approx \mathbb{E}(X) + \frac{a}{2} Var(X) - aCov(X, W)$$

This finding demonstrates that the premium to be charged is dependent on the joint distribution of the risk X itself and the "wealth" W of risks that the insurer held already prior considering to cover risk X.

As mentioned before the principle of (generalised) zero utility fits perfectly the concept of reinsurance or risk securitisation and offers the advantage of offering a pricing solution which takes into account the individual preferences and attitudes towards risk of stakeholders involved in the transaction.

To summarise, in a non-life insurance context the price for an insurable risk exceeds the expected loss value by a safety load which compensates the (re)insurer for the uncertainty around real losses. Thus, one aspect of pricing involves the estimation of the expected loss and another to understand the dynamics that force the risk load upwards or downwards. This is relevant not only in the case of reinsurance pricing but also in a cat bond transaction initial valuation. These aspects have been discussed in Chapter 2.

Chapter 3

A random forest based approach for predicting spreads in the primary catastrophe bond market

This chapter is mainly dedicated to our article published on an open access basis at Insurance: Mathematics and Economics entitled "A random forest based approach for predicting spreads in the primary catastrophe bond market"[1], see Makariou, Barrieu & Chen (2021). In particular, the chapter is comprised of two sections out of which the first is the article itself presented as it appears in the journal, whilst the second one includes some supplementing material which is not discussed in the published version of the article but we think that it may be of interest for the readers of the thesis.

3.1 A random forest based approach for predicting spreads in the primary catastrophe bond market

[1]The article can be accessed via the following link: https://doi.org/10.1016/j.insmatheco.2021.07.003

Insurance: Mathematics and Economics 101 (2021) 140–162

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

www.elsevier.com/locate/ime

A random forest based approach for predicting spreads in the primary catastrophe bond market

Despoina Makariou *, Pauline Barrieu, Yining Chen

Department of Statistics, London School of Economics and Political Science, United Kingdom of Great Britain and Northern Ireland

A R T I C L E I N F O

Article history:
Received February 2020
Received in revised form June 2021
Accepted 17 July 2021
Available online 30 July 2021

JEL classification:
G1
G220

Keywords:
Catastrophe bond pricing
Interactions
Machine learning in insurance
Minimal depth importance
Permutation importance
Primary market spread prediction
Random forest
Stability

A B S T R A C T

We introduce a random forest approach to enable spreads' prediction in the primary catastrophe bond market. In a purely predictive framework, we assess the importance of catastrophe spread predictors using permutation and minimal depth methods. The whole population of non-life catastrophe bonds issued from December 2009 to May 2018 is used. We find that random forest has at least as good prediction performance as our benchmark-linear regression in the temporal context, and better prediction performance in the non-temporal one. Random forest also performs better than the benchmark when multiple predictors are excluded in accordance with the importance rankings or at random, which indicates that random forest extracts information from existing predictors more effectively and captures interactions better without the need to specify them. The results of random forest, in terms of prediction accuracy and the minimal depth importance are stable. There is only a small divergence between the drivers of catastrophe bond spread in the predictive versus explanatory framework. We believe that the usage of random forest can speed up investment decisions in the catastrophe bond industry both for would-be issuers and investors.

1. Introduction

Catastrophe bonds are **I**nsurance-**L**inked **S**ecurities (ILS), first developed in 1990s, in an effort to provide additional capacity to the reinsurance industry post mega-disasters. The pricing of these instruments is particularly challenging as most of these securities are traded over the counter. Over the last years, there have been several empirical papers trying to address this difficulty by studying the price of catastrophe bonds using real-market data, mainly in the explanatory framework, see Lane (2000), Lane and Mahul (2008), Lei et al. (2008), Bodoff and Gan (2009), Gatumel and Guegan (2008), Dieckmann (2010), Jaeger et al. (2010), Papachristou (2011), Galeotti et al. (2013), Braun (2016), Gürtler et al. (2016), Götze and Gürtler (2018), Trottier et al. (2018), and only very recently in the context of comparative studies for machine learning algorithms, Götze et al. (2020).

The main orientation of the explanatory based approach was to explain catastrophe bond price via means of identification of variables having a theoretically material and statistically significant link to it. This was mostly achieved through the use of explanatory statistical models. Certainly, the aforementioned works have shed light on the drivers of catastrophe bond prices in the presence of causal theory. However, there are certain limitations, namely, selection bias, predictor interactions, non-linearities, and a non-purely-predictive study goal. Starting from selection bias, the data samples used previously often excluded bonds of certain characteristics, unusual issuances were eliminated as outliers, and observations with missing entries were excluded from data sets, leading to a potential significant loss of information. See Bodoff and Gan (2009), Götze and Gürtler (2018), Galeotti et al. (2013), Braun (2016) and Lane and Mahul (2008). Meanwhile, in Papachristou (2011), concerns about interactions between independent variables were expressed but not investigated. Another limitation is the extensive use of linear regression without justification of its suitability in a catastrophe bond market setting. This was recognised in some cases, see Lane and Mahul (2008) and Papachristou (2011). Finally, as Major (2019) mentioned, in terms of study goal, past works did not aim directly at spread prediction, although there is a business need for it.

In this manuscript, we suggest a supervised machine learning method called random forest (Breiman 2001) to predict spreads in the primary catastrophe bond market. Some reasons about the

* Corresponding author.
E-mail addresses: D.Makariou@lse.ac.uk (D. Makariou), P.M.Barrieu@lse.ac.uk (P. Barrieu), Y.Chen101@lse.ac.uk (Y. Chen).

https://doi.org/10.1016/j.insmatheco.2021.07.003
0167-6687/

model specification are discussed below. The model choice is partially based on the fact that random forest is widely considered as one of the most successful machine learning methods to date, see Berk (2008), and Biau and Scornet (2016) among others. It should be noted that random forest success in providing highly accurate predictions is mainly achieved by resolving the trade-off between over-fitting and prediction accuracy, as discussed in various works such as these of Breiman (2001), Díaz-Uriarte and De Andres (2006), Oh et al. (2003). Moreover, the recent novel research of Götze et al. (2020) compared different machine learning methods in a catastrophe bond market setting, which provides evidence that random forest outperforms neural networks, and linear regression which is combined with variable selection via Lasso and Ridge penalisations. In addition, we believe that the random forest method has a number of particular aspects which could help overcome some of the limitations presented previously in the explanatory framework in the literature. Firstly, random forest is a flexible method in a sense that makes no assumptions about the underlying data generative process. This is an important advantage that could help us effectively tackle the issue of non-linearities in the catastrophe bond market. Secondly, because the building blocks of the method are regression trees, random forest is reasonably robust to outliers. This is very useful given that catastrophe bonds can be extremely heterogeneous and losing information is particularly “costly” in this opaque market segment. Thirdly, once again, due to the tree structure of the method, variables are considered in such a way that allows to capture interactions without the need to specify them (Breiman et al. 1984). Fourthly, internal measures of variables' importance can be derived solely in a prediction context, and selection of the most important variables is feasible. Finally, there is only a small number of hyperparameters to tune, and the need for data pre-processing is minimal because many steps are integrated in the method itself, ensuring time efficiency from a business perspective.

Here, we apply the random forest method to predict spreads in the full spectrum of primary non-life catastrophe bond market. We aim to generate accurate spreads' predictions of new catastrophe bond observations on both temporal and non-temporal bases. Comparisons are made with highly competitive benchmark models. In absence of causal theory, we assess how spread predictors rank in terms of importance using two different methods, namely, permutation importance and minimal depth, where the latter is random forest specific. To our best knowledge, this work is among the first to apply the minimal depth method as described in Ishwaran et al. (2010) in a financial application. In addition, we explore whether the variables found by now to be good at explaining catastrophe bond spreads in the explanatory framework are similar to those good at prediction in absence of causal theory. As mentioned in Shmueli (2010), one should not expect these two to be exactly the same and indeed we find some small level of divergence. From an empirical perspective, we aim at random forest prediction accuracy and variables' importance results to be stable thus this aspect is also evaluated subject to multiple iterations of random subsampling. Besides, we assess the degree at which the prediction accuracy of random forest versus benchmark model is sensitive to simultaneous missingness of more than one predictor. By doing so we also check the degree to which the random forest captures predictors' interactions without specifying them, as well as its ability to extract information from existing variables to recover the loss of predictive power in the absence of other important predictors.

With regards to the benchmark model, first we reproduce and then improve the model of Braun (2016) to account for non-rated catastrophe bond issuances. Braun (2016) is chosen as it indicates the best out of sample performance to date in the relevant explanatory literature. Next, we build a new simple linear regression model based on the same set of variables we use for the random forest generation. For the first time, we include the risk modelling company and coverage type in the analysis as potential catastrophe bond spread drivers making a contribution in the explanatory framework. A potential reason for lack of prior works taking into account these variables is most probably due to the difficulty of finding information about them as they pin-point to very detailed aspects of a transaction - a view that Braun (2012) already expressed regarding the risk modelling company. The newly built regression model outperforms the improved version of Braun (2016) and thus is used as our benchmark in this manuscript.

The rest of the paper is organised as follows. In Section 2, we briefly introduce machine learning concepts. We explain our research methodology in Section 3 and present our catastrophe bond data set details in Section 4. Benchmark models are discussed in Section 5. We then demonstrate the random forest generation based on our catastrophe bond data in Section 6, followed by the evaluation of the random forest's performance in Section 7, and the importance analysis of catastrophe bond spread predictors in Section 8. Furthermore, in Section 9, we provide an example of how the random forest could be used in practice to assist issuers' and investors' decision making when they examine a new catastrophe bond issuance. Finally, concluding remarks follow in Section 10.

2. Machine learning preliminaries

In this section, we introduce some machine learning concepts that will be useful for the comprehension of methods used later on in our study. The explanations to be given are limited to a regression problem because catastrophe bond spread is a quantitative response variable.[1]

2.1. Supervised learning

Machine learning includes a set of approaches dealing with the problem of finding or otherwise learning a function from data (James et al. 2013). Supervised learning is a machine learning task where a function, otherwise called a hypothesis, is learned from a data set - often referred to as training set. The latter consists of a number of input-output pairs where for every single input in the training set the correct output is known. An algorithm is going through all data points in the training set identifying patterns and finding how to map an input to an output. Because the desired answer for the output is known, the algorithm modifies this mapping based on how different algorithm generated outputs are compared to the original ones in the training set (Friedman et al. 2001). Ultimately, the aim is that by the time the learning process finishes, this difference will be small enough for the algorithm to be able to map any set of new inputs the algorithm will come across in the future in a reasonable manner.

2.2. Ensemble learning

Sometimes instead of learning one mapping, it is useful to have a collection of mappings which merge their predictions to create an ensemble (Russell and Norvig 2016). Individual approximation functions in the ensemble are usually called base learners and predictions combination can happen in various ways with most usual ones being voting or averaging. Such techniques have

[1] We clarify that in machine learning literature, the term “regression problem” often refers to prediction using a continuous response variable, see James et al. (2013). We distinguish this from the term linear regression that is used throughout this work to describe either the linear regression models in the literature or our benchmark model.

been investigated quite early on, see for example Breiman (1996c), Clemen (1989), Perrone (1993) and Wolpert (1992). The main benefit of ensembles is that if each single hypothesis is characterised by high degree of accuracy and diversity then the ensemble is going to produce more accurate predictions than any of the individual hypotheses on its own, see Zhou (2012). Here, accuracy means that a hypothesis results in a lower error rate as opposed to one that would be derived from random guessing on new input values, while diversity means that each hypothesis in the ensemble makes different errors on new data points (Dietterich 2000a). Ensembles are usually built by utilising methods to derive various data sets out of the original data set for each base learner. One of the most famous methods to construct an ensemble is briefly discussed below.

2.3. Bagging

Bagging, an acronym for **b**ootstrap **agg**regat**ing** presented by Breiman (1996a), is a powerful ensemble learning method. As the name indicates, the ensemble uses the bootstrap, see Efron (1992), as resampling technique to take multiple data samples from which multiple base learners will be then generated. At the same time, aggregation, which is simple averaging for regression, is the way to combine the predictions of these individual base learners. There are various merits in using bagging for building ensembles. First, using a bootstrap sample to build each base learner means that a part of the original data (normally two third by default) are not used in its construction. Then, these unseen data points can constitute an unbiased test data to quantify how well each base learner generalises (Breiman 2001). Secondly, the method is useful when data is noisy (Opitz and Maclin 1999). Thirdly, and probably most importantly, by aggregating base learners which individually suffer from high variance, e.g. decision trees (Breiman et al. 1984), the ensemble as a whole achieves a variance reduction; see Breiman (1996a), Bauer and Kohavi (1999), Breiman (1996c), Breiman (1996b) and Dietterich (2000b). A pitfall of the method though is that whilst bagging reduces the ensemble variance, there are diminishing returns in variance reductions as the computational cost increases. This is because all bootstrap samples are drawn from the same original data set, meaning that base learners will inevitably be correlated. This latter point is where the idea of random forest is based on and it will be further discussed in Section 3.

3. Research methodology

Having provided necessary background information about certain machine learning concepts, the purpose of this section is twofold. We start by stating our catastrophe bond spread prediction problem introducing notations that will be used later in our study. We then continue by presenting our research methodology.

3.1. Problem statement with notations

Broadly, we use an ensemble algorithmic method to perform a supervised learning task for the primary catastrophe bond market. For now, let $\mathbf{x}$ generally denote[2] the input which reflects characteristics of catastrophe bonds available in the offering circular at the time of issuance and ILS market conditions. At the same time, let symbol y denote catastrophe bond spreads at the time of issuance. A function f of the form $y = f(\mathbf{x})$ relates catastrophe bond characteristics, conditions in the economic environment and possibly random effects to their spreads, however f is unknown. Based on past primary catastrophe bond data including information both for $\mathbf{x} = (x_1, x_2, \dots, x_P)$ where $p = 1, 2, \dots, P$ and y, we first want to find a function that approximates f so that we can predict spreads given new catastrophe bond input.

In particular, experience about past catastrophe bond issuances is captured by collecting $n = 1, \dots, N$ distinct input-output pairs. The input is a vector of predictors, also called features, covariates or independent variables, $\mathbf{x}_n = (x_{1n}, x_{2n}, \dots, x_{Pn})$ indexed by dimension $p = 1, 2, \dots, P$ and it is a element of $\mathbb{R}^P$. The output, also called response or dependent variable, is a real-valued scalar denoted by y_n indexed by example number $n = 1, \dots, N$. By assembling these N pairs, we collectively form a catastrophe bond data set $D = \{(\mathbf{x}_n, y_n), n = 1, 2, \dots, N\}$ based on which the ensemble algorithmic method will search the space H of all feasible functions, in a process called learning, and find a function, denoted by h_{en}, that is able to predict the response y' given a new input $\mathbf{x}'$ as accurately as possible. Because, we use an ensemble method, h_{en} is in reality a collection of functions approximating f. We are also interested in assessing the importance of each input of $\mathbf{x}$ in predicting the spread. Finally, all results will be evaluated on the grounds of them being stable subject to random subsampling of the whole data set.

3.2. Random forest

The ensemble method that we use is called random forest. It is developed by Breiman (2001) and is used to solve prediction problems. Below we present the rationale behind the method, random forest construction process, main hyperparameters, and lastly how random forest is used to make predictions.

3.2.1. Underlying logic

As James et al. (2013) mentioned, the underlying logic of random forest is to “divide and conquer”: split the predictor space into multiple samples, then construct a randomised tree hypothesis on each subspace and end with averaging these hypotheses together. Generally, random forest can be seen as a successor of bagging when the base learners are decision trees. This is because random forest addresses the main pitfall of bagging; the issue of diminishing variance reductions discussed earlier in Section 2.3. This is achieved by injecting an additional element of randomness during decision trees construction for them to be less correlated to one another. At the same time, since the base learners are decision trees there are not many assumptions about the form of the target function resulting in low bias.

3.2.2. Random forest construction process

The process of constructing a random forest involves various steps which are summarised in Fig. 1 and discussed straight after.

The first step in the random forest generation process is bootstrap sampling. In particular, from a data set, like D, we take $1, \dots, K$ samples with replacement each of them having the same size as the original data set. The second stage is regression trees development. From each bootstrap sample, K regression trees are grown using recursive partitioning as done in **C**lassification **A**nd **R**egression **T**rees (CART) (Breiman et al. 1984) but with a smart twist which further randomises the procedure. At each level of the recursive partitioning process, the best predictor to conduct the splitting is considered based on a fresh, each time, random subsample of the full set of predictors denoted as $\mathrm{m_{try}}$. The best split is chosen by examining all possible predictors in this sub-sample and all possible cut-points as of their ability to minimise the residual sum of squares for the resulting tree. A tree stops growing when a minimum number of observations in a given node is reached but generally speaking trees comprising the random forest are fully grown and not pruned. By constructing these K trees we

[2] Our convention is that bold lowercase letters reflect random vectors.

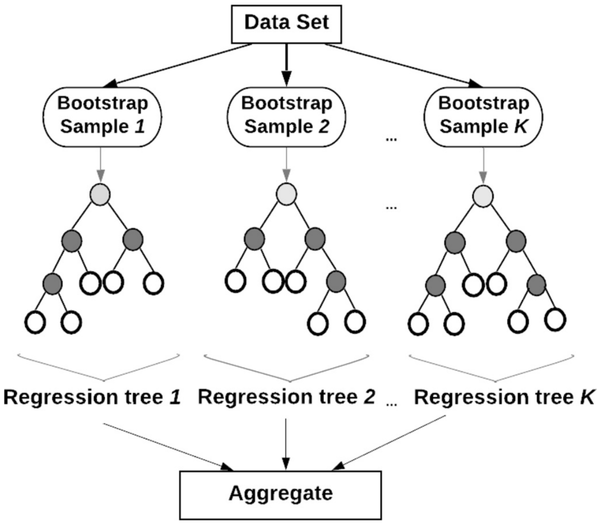


Fig. 1. Random forest construction scheme. For each regression tree, light grey circles indicate the root node, dark grey circles intermediate nodes and white colour circles terminal nodes.

effectively get K estimators of function f namely $h_1, h_2, \ldots, h_K$. The average of these individual estimators $h_{en} = \frac{1}{K}\sum_{k=1}^{K} h_k(\mathbf{x}_n)$ is the random forest.

3.2.3. Hyperparameters

From the above process description, it is evident that there are three parameters whose value needs to be fixed prior to random forest development; namely the number of trees grown, node size, and number of variables randomly selected at each split. Each of them respectively control the size of the forest, the individual tree size and an aspect of the within tree randomness. There are certain default values that have been suggested following empirical experiments on various data sets but one can use an optimising tuning strategy with respect to prediction performance to select the most suitable values specifically for the data set under study (Probst et al. 2018).

3.2.4. Making predictions

After the random forest is built, it can be used to provide predictions of the response variable. To make predictions though, it is necessary to feed the method inputs that have never been seen before during the construction process. As we have briefly mentioned in Section 2.3, due to bootstrap sampling, we can refrain from keeping aside in advance a portion of the original data set for testing purposes. The reason for this is that each tree uses more or less two thirds of the observations, from now on called in-bag observations, whilst the remaining one third of the observations are never used to build a specific tree, from now on called out of bag (OOB) observations. For each tree, the out of bag observations act as a separate test set. To predict the response variable value for the n^{th} observation, one should drop its corresponding input down every single tree in which this observation was out of bag. This means that by doing so one will end up having in hand on average $K/3$ predictions for any $n = 1, \ldots, N$ observation. Then, in order to derive a single response prediction for the n^{th} observation, the average of these predictions is taken. The same procedure is repeated for all other observations. Whether these predictions are good enough or not needs to be evaluated based on certain metrics as shown next.

3.3. Performance evaluation criteria for random forest

To assess the performance of any machine learning algorithm, one needs to set in advance the criterion upon which judgement will be made. In this paper, we employ two criteria for the performance evaluation of our random forest; prediction accuracy and stability. They are discussed below.

3.3.1. Prediction accuracy

Prediction accuracy is one of the most used performance indicators in machine learning algorithms aiming at prediction. This is no different for random forest algorithm as originally presented in Breiman (2001). In the current study, prediction accuracy is assessed based on two different perspectives: a temporal and a non-temporal one. We believe that such a distinction highlights different prediction needs and could add value in a practical context.

By employing a non-temporal approach, one can assess random forest predictions robustness when at the time of the spread prediction, the general catastrophe bond market conditions have been relatively stable over a time period prior the prediction, thus the time element could potentially be ignored. A non-temporal perspective would also be meaningful when simply the character of the prediction is not time relevant. With regards to the latter, an instance would be when there is ambiguity around the accuracy of spread information a company holds for a transaction or in cases that the spread information for a given transaction is unavailable resulting in a company having to face the issue of an incomplete data base. In both cases, the element of time may appear to be less important compared to the need of having a bigger and more diverse training set in the analysis. On the other hand, a temporal approach considers the robustness of the random forest in accurately predicting spreads over time. Effectively, such a point of view allows us to account for regime shifts and examine the degree at which an industry participant would be able to predict a new catastrophe bond spread no matter its features and risk profile. For this purpose, we need to split the data into separate train and test sets for various time periods and then assess the random forest and benchmark model prediction accuracy performance.

In the non-temporal context, prediction accuracy is primarily measured by means of the proportion of the total variability explained by the random forest, here denoted as $\mathrm{R}^2_{\mathrm{OOB}}$. Following Grömping (2009), the latter metric is defined as $\mathrm{R}^2_{\mathrm{OOB}} = 1 - \frac{\mathrm{SE}_{\mathrm{OOB}}}{\mathrm{TSS}}$ where $\mathrm{SE}_{\mathrm{OOB}}$ stands for the total out of bag squared errors and TSS for the total sum of squares. In addition, we denote the out of bag mean squared error as $\mathrm{MSE}_{\mathrm{OOB}} = \mathrm{SE}_{\mathrm{OOB}}/N$. With respect to $\mathrm{MSE}_{\mathrm{OOB}}$, it shows the variability in the response variable that is not forecasted by the random forest. It is calculated as $\mathrm{MSE}_{\mathrm{OOB}} = \{\sum_{n=1}^{N}(y_n - \overline{\hat{y}}_{n_{\mathrm{OOB}}})^2\}/N$ where $\overline{\hat{y}}_{n_{\mathrm{OOB}}}$ is the mean prediction for the n^{th} observation where $n = 1, \ldots, N$ for all trees for which the n^{th} data point was out of bag. In effect, $\mathrm{MSE}_{\mathrm{OOB}}$ is a sound approximation of the test error for the random forest because every single data point is predicted based solely on the trees that were not constructed using this observation. Actually, when the number of trees K is very large then the $\mathrm{MSE}_{\mathrm{OOB}}$ is roughly equivalent to leave one out cross validation James et al. (2013). With regards to TSS, as in linear regression, it reflects the degree at which the response variable, here the catastrophe bond spread, deviates from its mean value. It is defined as $\mathrm{TSS} = \sum_{n=1}^{N}(y_n - \overline{y})^2$ where y_n is the response variable value for the n^{th} observation where $n = 1, \ldots, N$ and $\overline{y}$ the mean value of the response variable. In this study, $\mathrm{R}^2_{\mathrm{OOB}}$ is going to be expressed in percentage terms. The higher the $\mathrm{R}^2_{\mathrm{OOB}}$, the better the prediction accuracy of the random forest is. Whilst for random forest, it is somehow natural to use the $\mathrm{R}^2_{\mathrm{OOB}}$, see Breiman (2001), we deem useful to also present the prediction accuracy results derived by two "more standard" statistical approaches, i.e. 10 fold cross validation and leave one out cross validation even though we expect that results may be fairly similar. In the temporal context, prediction performance will be assessed on the basis of out of sample R^2 denoted as R^2_{OOS}. The metrics presented here both in the non-temporal, and temporal context will be also used for the benchmark model to allow for a fair comparison.

3.3.2. Stability

The term stability here refers to how repeatable random forest results are when different samples taken from the same data generative process are used for its construction, see Turney (1995) and Philipp et al. (2018) for the rationale behind this approach. The rationale for investigating stability is rooted from the fact that consistent results are deemed more reliable, see Stodden (2015), Turney (1995), Yu (2013) and Philipp et al. (2018) for a discussion.

Various ways of measuring the stability of algorithmic results have been presented in Turney (1995), Lange et al. (2004), Ntoutsi et al. (2008), Lim and Yu (2016) and Philipp et al. (2018). In this study, we are inspired by the works of Turney (1995) and Philipp et al. (2018) with regards to stability and its empirical measurement. In particular, the idea is that by obtaining two sets of data from the same phenomenon sampled from the same underlying distribution the algorithm needs to produce fairly similar results from both data sets for it to be considered stable. One way to achieve this is to randomly partition the whole data set into two separate data sets multiple times. An important decision though is how to take the samples. Here, we propose taking the samples using the split-half technique as described in Philipp et al. (2018) meaning that the whole catastrophe data set will be split into two disjoint data sets of roughly equal size. This sampling method ensures that a similarity between the results is not attributed to the same observations being in both samples as this could result in similar results without meaning that the algorithm is actually stable. By choosing a small learning overlap it is possible to examine the degree of a result generalisation for independent draws from the catastrophe bond data generative process.

In particular, following Turney (1995) and Philipp et al. (2018), we obtain two sets of data from the same phenomenon and same underlying distribution with as little learning overlap as possible, then construct two random forests from each one and check whether prediction accuracy is fairly similar. To be more specific, we take a random 50% of the observations without replacement from the initial catastrophe bond data set, namely Sample A. The rest of the original data set observations, not included in Sample A, forms Sample B. Then, two separate random forests are grown out of Sample A and Sample B to assess the stability of random forest prediction accuracy to changes in the initial data set. We repeat this process 100 times. Optimal values for the number of variables randomly selected to be considered at each split are sought in both cases.

3.4. Evaluation of predictors' importance

The random forest algorithm allows for assessing how important each predictor is with respect to its ability to predict the response, a concept that is briefly called as variables importance. Its assessment is executed empirically (Grömping 2009) and see Chen and Ishwaran (2012) for a comprehensive review of various methods that can be used to achieve this. Here, the focus lies on two widely used approaches namely permutation importance, and minimal depth importance.

3.4.1. Permutation importance

The central idea of permutation importance, also known as "Breiman-Cutler importance" (Breiman 2001), is to measure the decrease in the prediction accuracy of the random forest resulting from randomly permuting the values of a predictor. The method provides a ranking for predictors' importance as end result and it is tied to a prediction performance measure. In particular, the permutation importance for x_p predictor is derived as follows. For each of the K trees: firstly, record the prediction error $\text{MSE}_{\text{OOB}_k}$; secondly, noise up, i.e. permute, the predictor x_p in the out of bag sample for the k^{th} tree; thirdly, drop this permuted out of bag sample down the k^{th} tree to get a new $\text{MSE}^{x_p\, perm}_{\text{OOB}_k}$ after the permutation and calculate the difference between these two prediction errors (before and after the permutation). In the end, average these differences over all trees. The mathematical expression of the above description is $I_{x_p} = \sum_{k=1}^{K} [\frac{1}{K}(\text{MSE}^{x_p\, perm}_{\text{OOB}_k} - \text{MSE}_{\text{OOB}_k})]$ where I_{x_p} is the importance of variable x_p, K the number of trees in the forest, $\text{MSE}^{x_p\, perm}_{\text{OOB}_k}$ the estimation error with predictor x_p being permuted for the k^{th} tree, and $\text{MSE}_{\text{OOB}_k}$ the forecasting error with none of the predictors being permuted for the k^{th} tree. The larger the I_{x_p} the stronger the ability of x_p to predict the response. Generally speaking a positive permutation importance is associated with decrease in prediction accuracy after permutation whilst negative permutation importance is interpreted as no decline in accuracy.

3.4.2. Importance based on minimal depth

The other approach for measuring predictors importance is based on measure named minimal depth, presented in Ishwaran et al. (2010) with the latter being motivated by earlier works of Strobl et al. (2007) and Ishwaran (2007). The minimal depth shows how remote a node split with a specific predictor is with respect to the root node of a tree. Thus, here the position of a predictor in the k^{th} tree determines its importance for this tree. The latter means that unlike permutation importance, the importance of each predictor is not tied on a prediction performance measure. Also, in addition to ranking variables, the method also performs variable selection - a very useful feature for elimination of less important predictors.

Specifically, Ishwaran et al. (2010) have formulated the concept of minimal depth based on the notion of maximal sub-tree for feature x_p. The latter is defined as the largest sub-tree whose root node is split using x_p. In particular, the minimal depth of a predictor x_p, a non-negative random variable, is the distance between the k^{th} tree root node and the most proximate maximal sub-tree for x_p, i.e. the first order statistic of the maximal subtree. It takes on values $\{0, \ldots, Q(k)\}$ where $Q(k)$ the depth of the k^{th} tree reflects how distant is the root from the furthermost leaf node, i.e. the maximal depth (Ishwaran et al. 2011). A small minimal depth value for predictor x_p means that x_p has high predictive power whilst a large minimal depth value the opposite. The root node is assigned with minimal depth 0 and the successive nodes are sequenced based on how close they are to the root. The minimal depth for each predictor is averaged over all trees in the forest. Ishwaran et al. (2010) showed that the distribution of the minimal depth can be derived in a closed form and a threshold for picking meaningful variables can be computed, i.e. the mean of the minimal depth distribution. In particular, variables whose forest aggregated minimal depth surpasses the mean minimal depth ceiling are considered irrelevant and thus could be excluded from the model. However, since Ishwaran et al. (2010) suggests that variable selection using the minimal depth threshold is more meaningful for problems with high dimensionality, this aspect is not considered relevant in the current study.

3.4.3. Other evaluation factors

After calculating the importance of predictors using the methods described above, we consider useful to examine the results based on two additional criteria. Firstly, we want to ensure that primarily the importance rankings and secondarily the selected variables are repeatable. Because both permutation and minimal depth importance are linked to the random forest constructed, the stability of predictors' importance results is evaluated in line with the random forest stability evaluation for the catastrophe bond data set, as mentioned in Section 3.3.2. Secondly, we check whether the predictors' importance results reflect investors' knowledge from an empirical perspective. In a business context, it would be uncomfortable for an investor to see good catastrophe bond predictions but with importance rankings of the predictors outside their empirical knowledge, even though this type of agreement is not necessary from a statistical viewpoint.

4. Catastrophe bond data

In this section, we present how the catastrophe bond data used in this study have been collected and processed whilst details are given with respect to the choice of variables and their role in our study.

4.1. Collection

The core of catastrophe bond pricing cross sectional data has been collected from a leading market participant enabling us to work with a data set that is substantially larger than those used in the literature. The websites of ARTEMIS, Lane Financial LLC and Swiss Re Sigma Research have been also extensively used to cross validate data entries that were unclear or non-available in the main data body. Historical values of the Synthetic Rate on Line index have been given by Lane Financial LLC. To the best of our knowledge, our data set refers to all non-life catastrophe bonds issued in the primary market from December 2009 to May 2018, a total of 934 transactions. This time period is particularly interesting since it coincides with the restart of the catastrophe bond sector after almost two years of low activity following the collapse of Lehman Brothers, which played a counterparty role in several bonds and therefore ignited concerns and reflection around the structuring of transaction as to ensure security of collateral, see Hills (2009). The information gathered was related to investors' return, loss potential of the securitized risk, i.e. expected loss and attachment probability, various design characteristics of the risk transfer, i.e. issuance size, coverage period, coverage type, trigger, region, peril, credit score, risk modelling company, price cyclicality in ILS market, and BB corporate bond spreads level.

4.2. Preparation

Since we consolidated data from various sources there were pieces of information referring to the same concept but measured in different units across different data providers. Such scaling issues have been appropriately addressed to maintain consistency. With regards to the spread at issuance, it was derived from the coupon by subtracting the element of the money market rate. In the case of zero coupon catastrophe bonds the spread was derived from the implied coupon by subtracting the element of the money market rate.

Through validating the data across various sources, we ensured that there are no missing values in the study, a pitfall in many previous works. On this note, it needs to be acknowledged that an exception in the above non-missing values claim is very few catastrophe bonds for which there was no information regarding the risk modelling firm because these transactions were privately placed even though our data set contains other private placement deals for which we did not have missing values. For these few deals for which vendor information was missing, we created a separate category level to capture this specific reason for missingness, i.e. private placement. Including this level is considered important via means that the developed algorithmic method will be able to predict spreads for these circumstances also. Further information on this category level can be found in Appendix A.

4.3. Discussion about the choice of variables

The variables included in the data set can be seen in Table 1, presented along with the definition, type, and their role in this study. In Appendix A, one can find basic statistical information and histograms for all variables along with a discussion to enhance the understanding of catastrophe bond data intricacies. With regards to the role of each variable in our research, the spread was chosen as dependent variable as it is an industry wide accepted lens through which one can see catastrophe bond pricing. The spread is of utmost interest to the investors as it indicates how much they could earn on the top of the risk free rate if they decided to employ their capital in this alternative risk transfer segment.

Since the goal of this study lies on the prediction of spread, a major consideration is that the independent variables need to be available at the time of the prediction. This is indeed the case here, as the predictors constitute information included in the placement material offered to investors prior to a new catastrophe bond issuance. Also, in the case of predictor RoL, investors are also aware of the general ILS market conditions and possibly we could assume that the Financial Lane LLC Synthetic Rate on Line index values are readily available at an investment company level. Similar rationale applies for the BB spread regarding its availability at the point of the prediction. The reason why we have incorporated RoL and BB spread in the study is because the prior literature shows that such macroeconomic variables have a relevant influence on catastrophe bond spreads, see Braun (2016) and Gürtler et al. (2016) for example.

We note that there are previous works (see Galeotti et al. (2013), Braun (2016), Gürtler et al. (2016), and Trottier et al. 2018, among others) refraining from using the attachment probability (AP) as a predictor for the spread forecast, even though the reason was not mentioned explicitly. A potential explanation for why the EL was preferred over the AP in these works is because the EL is a coherent risk measure, meaning that if we were to examine catastrophe bonds in a portfolio context, then the EL contributes proportionately to the portfolio EL. This is not the case when a risk measure such as Value at Risk (VaR) is developed using AP as basis because there are instances where the subadditivity condition of coherent risk measures, i.e. $VaR_{AP}(X) + VaR_{AP}(Y) \geq VaR_{AP}(X+Y)$, is not satisfied; see Galeotti et al. (2013). Having said that, whether or not AP might be appropriate for assessing catastrophe bonds at the portfolio level is not examined in the current study. It should also be mentioned that for our purpose, it may be helpful to include the variable AP, thanks to the fact that the correlation between EL and AP appears heterogeneous in this dataset. For example, whilst it is somehow expected that EL and AP are highly correlated, if we were to focus on transactions with large spreads, then the correlation between EL and AP is around 70%, indicating that AP contains information which is not captured by EL for these cases. Moreover, since our study aims at prediction, the addition of an extra variable is not an issue for the random forest. In addition, including AP does not materially affect the performance of LR model in this example either.

With regards to the variable loc_peril, we use a location - peril code categorisation closely in line with the data provided to reflect industry practice. In Appendix A, we provide details regarding all

Table 1
Catastrophe bond data set glossary.

Variable	Description	Type	Role
spread	The amount of interest earned on the top of the risk free rate.	continuous	response
AP	(**A**ttachment **P**robability). The probability of incurred losses surpassing the attachment point. For catastrophe bonds with parametric triggers, AP is translated as the probability that measured parameters will surpass the agreed trigger point.	continuous	predictor
BB spread	U.S. High Yield BB Option-Adjusted Spread for the examined time period computed as the difference between a yield index for the BB rating category and the Treasury spot curve, as in Braun (2016). It reflects the BB rated corporate bond spread, with the BB rating being chosen because, because out of the rated catastrophe bonds, the vast majority of them exhibit this rating. It can be considered as a macroeconomic variable.	continuous	predictor
coverage	Contract term indicating whether protection is offered for a string of loss events or a single loss event.	categorical	predictor
EL	(**E**xpected **L**oss). The annual expected loss within the layer in question divided by the layer size.	continuous	predictor
rating	A dummy variable indicating the credit rating quality of the bond (granular rating), or whether it has not been rated at all.	categorical	predictor
iss_year	The year of issuance of a given catastrophe bond to capture cyclical effects.	continuous	predictor
loc_peril	A location-peril combination.	categorical	predictor
RoL	(**R**ate **o**n **L**ine). Quarterly values of Lane Financial LLC Synthetic Rate on Line Index for the examined time period capturing the level of rates in the ILS, and ILW markets. It can be considered as a macroeconomic variable, see Fig. 8.	continuous	predictor
size	Catastrophe bond nominal amount.	continuous	predictor
term	Years passed from issuance to maturity date.	continuous	predictor
trigger	Mechanism through which a loss payment is activated.	categorical	predictor
vendor	Catastrophe risk modelling software firm.	categorical	predictor

location peril combinations we have considered. Finally, the reason why we have incorporated the issuance year in the predictors set is to account for any other unknown drivers of spread related to a particular issuance year. As an example, one possible instance of such a driver would be the release of an updated model by a risk model vendor which would significantly influence underwriting as it happened in 2011 when RMS released its software Version 11.

To the best of our knowledge, one of the novelties in our study is that we explore the association between coverage type and catastrophe bond spreads. This is in line with current sector discussions as expressed in ILS[3] speciality articles, such as Risk (2019) and Muir-Wood (2017). There, the need to incorporate the coverage type in catastrophe bond pricing was highlighted following the extensive capital freezes investors experienced after California wildfires in 2018. Briefly touching upon this topic, wildfires, a not well understood peril, has been mostly transferred to investors with a provision that losses are covered on an aggregate basis. By design, aggregate deals tend to obtain losses easier, even from small events, compared to their per occurrence counterparts, as a string of loss events triggers the bond. The incapacity of the models to account for this to date led to big losses from aggregate deals and pressure for spreads to incorporate this transaction aspect. This signifies the importance of considering this variable. A further addition into the variables kit for studying the spread is the incorporation of information regarding the modelling company employed to calculate the frequency and severity of the securitised catastrophe risks. The software used for this purpose is firm specific thus it is interesting to explore whether by knowing this information part of the spread can be predicted.

A final note for the variables of this study regards credit ratings. Following Braun (2016), we initially thought to consider whether an issuance was allocated an investment grade by an independent credit rating agency and add an additional categorical value to account for transactions which were not rated as in our data set the majority of catastrophe bonds were issued without a credit rating attached to them. However, as we explain in Section 5, our benchmark model performed better when used granular rating for the rated transactions with the extra categorical value for the non-rated deals - thus our analysis follows this set up. It is worth noting that the absence of credit rating in new issuances is not solely an observation in the current data set. In ILS professional circles, the popularity of non-rated catastrophe bonds is justified from a catastrophe bond market evolution perspective; investors feel more comfortable and trust the risk modelling companies for the calculation of loss and the analysis of the risk return profile more. As a result, credit ratings are somehow no longer seen as essential as they used to be in the past and this is reflected in the increasing issuance pace of non-rated bonds, see ARTEMIS (2019). In the following sections, we choose our benchmark model out of two alternative ones and then apply the research methodology of Section 3 to the catastrophe bond data set that we have just discussed here.

5. Benchmark models

Before we report the random forest generation and prediction accuracy results, we discuss the benchmark models we considered. Even if random forest is not a new approach, it would be helpful to use a benchmark model for its performance assessment and evaluation as its rationale somehow differs from the methods used in most of the previous studies. In search for a benchmark, we looked into the models of Galeotti et al. (2013), Gürtler et al. (2016), Braun (2016), and Trottier et al. (2018), as they are non-fragment[4] and exhibit high out of sample performance. Given

[3] ILS is an abbreviation for Insurance Linked Securities or Insurance Linked Securitisation depending on the context in which it is used.

[4] By non-fragment, we mean that multiple peril - territory coding has been considered.

that the majority of catastrophe bond transactions in our data set are non-rated transactions, we decided to slightly alter the model of Braun (2016) to account for non-rated transactions in addition to those having attached an investment or non-investment grade credit quality tag. Such an alteration allows us to use all 934 observations in our data set. However, the results of the original Braun (2016) model can be found in Appendix C.[5]

As an alternative benchmark model, we also built a new linear regression model (from now on denoted as LR) using the set of variables we consider for the random forest generation, as presented in the previous section.[6] The improved Braun (2016) model is then compared to our LR model by means of in sample overall R^2 and out of sample R^2 resulting from 10 fold cross validation, leave one out cross validation, and bootstrap.[7] The results are presented in Tables 2 and 3.

It appears that LR model outperforms the improved model of Braun (2016) both in terms of in sample and out of sample performance. The overall in sample R^2 for LR is around 85% compared to 80% when using the improved Braun (2016) model. LR also gives consistently better R^2_{OOB}, R^2_{10CV}, and R^2_{LOOCV} results compared to the improved model of Braun (2016). Consequently, the random forest model will be compared to the more competitive LR regression model when we examine its prediction accuracy.

6. Random forest generation

In order to build[8] the random forest using our catastrophe bond data set, we first needed to decide the hyperparameters' values that we will use, i.e. number of trees, number of variables randomly selected at each split and node size. Breiman (2001) has suggested certain default values that seem to work well after multiple empirical experiments; still we have incorporated certain tuning strategies for the most important hyperparameters. Our approach in choosing these values is explained below.

6.1. Number of trees

The number of trees in the random forest controls its size. Generally, it is good to have a large number of trees as their resulting decisions will be complementing each other more, having a positive impact on random forest prediction accuracy. At the same time, a large number of trees is a safe option in case the optimal value of hyperparameter m_{try} is small so that each variable has enough of a chance to be included in the forest prediction process. However, except for the computational cost which is associated with growing large random forests, it was found by Breiman (2001) that there are diminishing returns in the prediction accuracy increase by adding a bigger number of trees. Taking these reflections into account, we start the random forest development process by growing 2000 trees and in Fig. 2 one can see how the MSE_{OOB} converges for various values of random forest size up to this level (the plot is produced on a logarithmic scale for the ease of readability). From a first sight, it does not take a large number of trees for MSE_{OOB} to stabilise. Before even reaching 100 trees, MSE_{OOB} drops from around 35000 to less than 5500. By the time we reach to 200 trees, it seems that the MSE_{OOB} is almost stabilised. Finally, we find that 500 trees, i.e. the default value that Breiman (2001) suggests, is adequate for our problem as it corresponds to virtually the same R^2_{OOB} as when using 2000 trees and has a much smaller computational cost. Therefore, we choose 500 as the number of trees in the random forest.

6.2. Node size

The hyperparameter node size, i.e. the minimum number of data points in the terminal nodes of each tree, controls the size of the tree in the random forest and effectively determines when the recursive partitioning should stop. A large node size results in shallower trees because the splitting process stops earlier. This has the advantage of lower computation times, but it effectively means that the tree will not learn some patterns resulting in lower prediction accuracy. A small node size translates to a higher computational cost but more thorough learning of patterns and consequently a more accurate base learner. The recommended value for node size given by Breiman (2001) is 5 for regression problems. This default value was also suggested and used by many other authors, as Wang et al. (2018), Grömping (2009), and Berk (2008) and therefore we also employ it as node size value here. The random forest needs to consist of trees which are fully or almost fully grown, see Breiman (2001), thus there is not much added value in exploring this aspect further as 5 meets this requirement and there is a general consensus for its appropriateness.

6.3. Number of variables selected at each split

The number of candidate predictors getting randomly considered at each split, m_{try}, is the most important hyperparameter. This is because it mostly affects the performance of the random forest and the predictors' importance measures, see Berk (2008). The significance of m_{try} lies on the fact that it influences at the same time both the prediction accuracy of each individual tree but also the diversity of the trees in the forest. To get the most out of the random forest, one wants each tree to have good prediction performance but at the same time trees not to be correlated to one another. However, these two goals are conflicting. An individual tree will be the most accurate when m_{try} has a high value but this would result in high correlation for the ensemble. In particular, an extreme case of $m_{try} = P$ would force the process to account to simple bagging (James et al. 2013). Generally, a small m_{try} is preferable as, for a sufficiently large number of trees, each predictor will have higher chance to get selected and thus contribute to the forest construction. All in all, the trade-off between individual learner accuracy and diversity needs to be managed by finding an optimal value which secures balance for the data set we study.

In Breiman (2001), the default value of $m_{try} = P/3$ (rounded down) is suggested for regression problems. This means that in our problem where $P = 12$, the algorithm would consider 3 predictors at each potential split. We have investigated the relevance of this empirical rule using a tuning strategy called grid search followed by 5-fold cross validation. The goal was to ensure that the most appropriate m_{try} is chosen. The process started by specifying the range of all possible values that m_{try} can take, namely the grid. In the current study, this is between 1 and 12, i.e. as many as the number of predictors. Then, 12 different versions of the random forest algorithm were built one for each possible value of m_{try}.

[5] Both the in sample, and out of sample results of the improved Braun (2016) model are very similar to those of the original Braun (2016) model, even though the improved model performs slightly better.

[6] For an alternative, yet worse performing, linear regression model where, instead of the variable rating as described in Table 1, we include the variable Investment grade (IG) as per the improved model of Braun (2016), see Appendix B.

[7] The reason why we present the bootstrap results here is because it is used as measure of prediction performance in the following sections.

[8] The statistical software used is R, version 3.5.1. The statistical packages employed to perform computations are the following. `randomForest` (Liaw and Wiener, 2002) for developing the random forest as well as calculating permutation importance values, `randomForestSCR` (Ishwaran and Kogalur, 2019) for calculating minimal depth importance measures, and `caret` (Kuhn, 2008) for tuning the main hyperparameter using grid search methodology. It should be mentioned that whenever packages `randomForestSCR` and Kuhn (2008) were used, algorithm arguments used agreed to those used in package `randomForest` to avoid inconsistencies.

Table 2
In sample fit of the improved linear regression model of Braun (2016) versus the new linear regression model LR. Further information on the category levels of the LR variables can be found in Appendix A.

Improved LR model of Braun (2016)	Estimate	Std. error	t value	Pr(> |t|)
(Intercept)	−801.78	41.53	−19.30	0.000 ***
Swiss Re	16.56	13.31	1.24	0.210
RoL index	6.90	0.40	17.52	0.000 ***
BB spread	68.55	8.35	8.21	0.000 ***
Investment grade no (*baseline*)				
Investment grade yes	−180.01	92.74	−1.94	0.050 *
Investment grade nr	62.68	14.74	4.25	0.000 ***
Peak territory	196.42	17.36	11.31	0.000 ***
Expected Loss	1.13	0.02	44.20	0.000 ***
R^2	80.04%			
Adjusted R^2	79.89%			
Res. Std. Error	183.70 (df = 926)			
F Statistic	530.60 (df = 7; 926)			
LR	**Estimate**	**Std. error**	**t value**	**Pr(> |t|)**
(Intercept)	55940.00	9396.00	5.95	0.000 ***
RoL	6.25	0.41	15.10	0.000 ***
BB spread	46.24	8.39	5.51	0.000 ***
term	−28.83	8.57	−3.36	0.001 ***
size	0.00	0.00	2.521	0.012 *
trigger industry loss index (*baseline*)				
trigger indemnity	−21.10	14.49	−1.46	0.146
trigger model	−69.30	38.66	−1.79	0.073
trigger multiple	−44.55	42.70	−1.04	0.297
trigger parametric index	−28.88	40.60	−0.71	0.477
trigger parametric	−167.60	36.55	−4.59	0.000 ***
coverage aggregate (*baseline*)				
coverage both	52.27	82.10	0.64	0.525
coverage occurrence	−49.45	13.56	−3.65	0.000 ***
vendor AIR (*baseline*)				
vendor AON	98.50	87.51	1.13	0.261
vendor EQECAT	−0.82	32.80	−0.03	0.980
vendor pp	9.08	71.18	0.13	0.899
vendor RMS	27.97	18.73	1.49	0.136
AP	−13.99	5.90	2.37	0.018 *
EL	1.26	0.09	14.91	0.000 ***
iss_year	−27.95	4.65	−6.01	0.000 ***
APAC_Quake (*baseline*)				
loc_peril APAC_Typh	−63.60	43.11	−1.48	0.141
loc_peril Europe_APAC_Multi_Peril	−105.00	125.50	−0.84	0.403
loc_peril Europe_Quake	8.94	55.94	0.16	0.873
loc_peril Europe_Wind	−148.20	39.36	−3.77	0.000 ***
loc_peril NA_APAC_Multi_Peril	55.08	47.07	1.17	0.242
loc_peril NA_Europe_APAC_Multi_Peril	139.00	41.14	3.38	0.001 ***
loc_peril NA_Europe_Multi_Peril	118.10	39.79	2.97	0.003 **
loc_peril NA_Multi_Peril	158.90	27.33	5.82	0.000 ***
loc_peril NA_Quake	−23.64	34.40	−0.69	0.492
loc_peril NA_Wind	86.64	29.74	2.91	0.004 **
loc_peril SA_Quake	139.90	103.80	1.35	0.178
rating B (*baseline*)				
rating BB	−146.80	18.05	−8.13	0.000 ***
rating BBB	−346.70	83.25	−4.16	0.000 ***
rating CCC	−45.11	94.25	−0.48	0.632
rating nr	−2.09	18.90	−0.11	0.912
R^2	85.07%			
Adjusted R^2	84.52%			
Res. Std. Error	161.10 (df = 900)			
F Statistic	155.40 (df = 33; 900)			
Note for signif. codes:	*p < 0.1; **p < 0.05; ***p < 0.01			
Observations number:	934			

Table 3
Out of sample performance measured in terms of R^2_{OOB}, R^2_{10CV}, and R^2_{LOOCV} for the improved linear model of Braun (2016) versus the new linear regression model LR.

Model	R^2_{OOB}	R^2_{10CV}	R^2_{LOOCV}
Improved Braun (2016)	79.71%	80.81%	79.40%
LR	83.30%	84.42%	83.84%

The prediction accuracy of each random forest version, measured by means of R^2_{OOB}, was evaluated through a 5-fold cross validation.

The results, shown in Fig. 3, reveal that there is considerable improvement in random forest performance when the m_{try} value is increased from 1 to 2, and then 3 to 4. No real advantage in terms of prediction accuracy seems to be yielded from further increasing the m_{try} value above 4 which also happens to be the default value $m_{try} = 12/3$ as per the suggestion of Breiman (2001). Moreover, since variable importance measures are to be calculated later on, we deem preferable to choose the smaller value of $m_{try} = 4$, by discipline, as this would lead to less correlated trees giving the

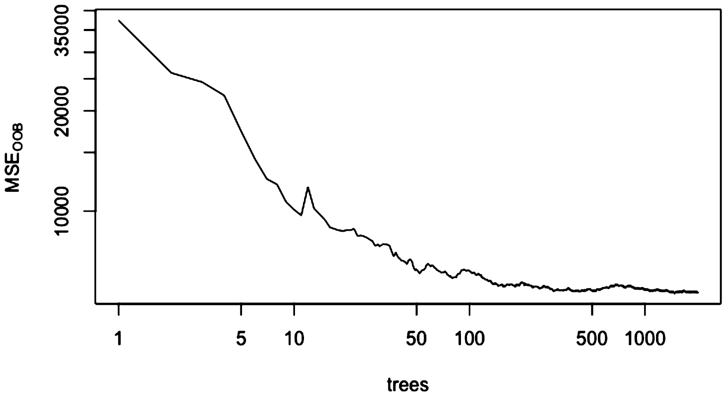


Fig. 2. Out of bag mean squared error convergence with respect to random forest size. The line corresponds to the mean squared error based on out of bag samples (MSE_{OOB}) versus the number of trees in random forest. The plot is produced on a logarithmic scale for the ease of readability.

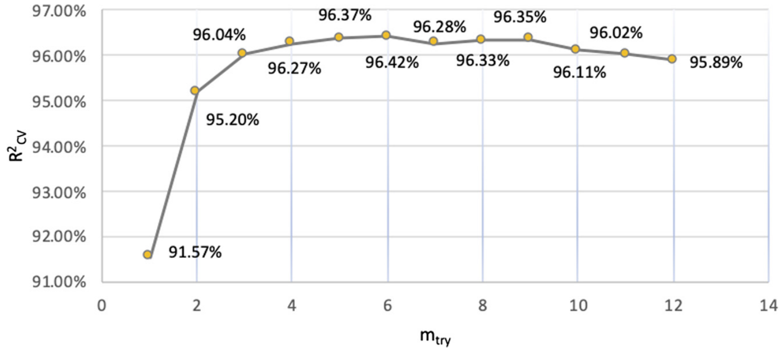


Fig. 3. Tuning of main random forest hyperparameter through grid search followed by 5-fold cross validation. Out of bag based R^2 (R^2_{OOB}) for random forest versus number of candidate predictors getting randomly considered at each split (m_{try}) during forest generation.

Table 4
Description of final random forest in terms of sample size, predictors number, and hyperparameters values.

Final random forest description	
sample size	934
number of predictors	12
random forest type	regression
number of trees	500
no. of variables tried at each split (m_{try})	4
node size	5

opportunity to see the influence of weaker predictors to catastrophe bond spreads prediction. Also, a smaller m_{try} value would lead to a simpler model which would be less costly in terms of computational time. Having decided on the hyperparameter values, the final random forest was generated and a summary description is provided in Table 4. The next section investigates how well the random forest performed in our catastrophe bond setting.

7. Random forest performance evaluation

In this section, we evaluate how well our random forest performs with regards to its prediction accuracy and stability.

7.1. Random forest prediction accuracy

As mentioned in Section 3.3.1, the ability of the random forest to predict catastrophe bond spreads on new inputs is investigated from both a non-temporal and a temporal point of view. In the former case, the prediction accuracy metrics we consider are R^2_{OOB}, R^2_{10CV}, and R^2_{LOOCV} whilst in the latter case R^2_{OOS} is used to assess the out of sample performance. The prediction accuracy results of the random forest versus the benchmark model are presented and discussed below for each of the two perspectives.

7.1.1. Non-temporal prediction accuracy

We start by clarifying what we regarded as new inputs followed by how the catastrophe bond spread predictions were made for the computation of R^2_{OOB} as it may appear to be a less standard approach (especially for the benchmark model) compared to 10 fold, and leave one out cross validation.

Starting from the random forest, as new inputs for a given tree, we have accounted its out of bag observations. Due to the property of sampling with replacement, only around two thirds of $N = 934$ data points were used to build each of the 500 unpruned and almost fully grown (node size = 5) regression trees. For a given tree, the remaining one third of $N = 934$ data points were never used during the building process and as a result they formed a reliable test set for it. Secondly, a prediction for the spread at issuance for the $n = 1$ observation, $\hat{y}_1$, was produced by dropping its corresponding input down every single tree in which the $n = 1$ observation was out of bag. This resulted on average to around one third of 500 catastrophe bond spread predictions for the $n = 1$ observation. Then, a single spread prediction for the $n = 1$ observation was made by taking the average value of these predictions. After having predicted the catastrophe bond spread value for the observation $n = 1$, the same process has been repeated for the $n = 933$ observations left. Finally, in order to evaluate the prediction accu-

Table 5
Prediction accuracy performance measured in terms of R^2_{OOB}, R^2_{10CV}, and R^2_{LOOCV} versus in sample performance measured in terms of R^2 for random forest (RF), and linear regression (LR).

Model	R^2_{OOB}	R^2_{10CV}	R^2_{LOOCV}	R^2
RF	96.57%	96.49%	96.59%	99.25%
LR	83.30%	84.43%	83.84%	84.52%

racy of our random forest, the metrics discussed in Section 3 were calculated. In particular, we have computed the mean squared error based on the out of bag data as $\text{SE}_{\text{OOB}} = \sum_{n=1}^{934}(y_n - \overline{y}_{n_{\text{OOB}}})^2$, the total sum of squares as $\text{TSS} = \sum_{n=1}^{934}(y_n - \overline{y})^2$ and, the variability explained by our random forest as $\text{R}^2_{\text{OOB}} = 1 - \frac{\text{SE}_{\text{OOB}}}{\text{TSS}}$.

With respect to the benchmark model, the calculation of R^2_{OOB} was done as it was described in the case of random forest; we used 500 bootstrap samples to refit the model and for each observation, we only considered predictions from bootstrap samples not including that observation. The results for the prediction accuracy metrics both for the random forest,[9] and the benchmark model are presented in Table 5. Note that in Table 5, we have also included the in sample R^2 for both models as reference.

It stands out that our random forest explains more than 96% of the total variability in the non-temporal context no matter whether using the bootstrap or one of the other two cross validation methods. At the same time, the non-temporal predictive performance of our benchmark model, i.e. linear regression, is lower - the highest total variability it explains across all metrics is 84.43%. Once again the prediction accuracy results for the benchmark model are very similar across different resampling methods. As a result, from now on we will be focusing on the out of bag related metrics in the non-temporal context.

With regards to the in sample R^2, it seems that random forest may lead to some degree of overfitting which is expected as the individual regression trees are fully grown. The latter signifies the fact that measuring performance in terms of R^2 for random forest in this instance might not be as appropriate as in the case of linear regression.[10]

An important aspect is to evaluate whether the 96.57% random forest non-temporal prediction accuracy is high enough given the nature of the problem under study. On a broader perspective, making predictions in a financial market setting is not an easy task. Inefficiencies, multiple market participants and, the influence of psychology on their behaviour are only few of the factors making the prediction task complex. Consequently, one might claim that achieving an R^2_{OOB} of more than 96% here corresponds to a very satisfactory level of prediction accuracy. Of course this also holds true for the around 84% benchmark model prediction performance but since there is a considerable difference in the reported R^2_{OOB}, we would conclude that using random forest in a non-temporal context may be preferable.

7.1.2. Temporal prediction accuracy

In a temporal context, we focus on the forecasting ability of the random forest versus the linear regression model over time. In this case, the training data is not picked randomly thus we are able to assess robustness towards potential regime shifts. Regime shifts

[9] In a robustness check, see Appendix D, we provide the prediction accuracy of the random forest when the categorical variables in the catastrophe bond data set are pre-processed with categorical dummies as in the case of linear regression. As we see the two random forest versions, i.e. with and without dummies, lead to very similar results.

[10] It should be noted that this problem is not specific to random forest but more general and lies in the use of in-sample performance measures in case of over-fitted models and therefore can also be seen in a linear regression context.

Table 6
Prediction accuracy of random forest (RF) versus linear regression (LR) measured in terms of R^2_{OOS} for various train-test sets by issuance year.

Train set	Test set	RF R^2_{OOS}	LR R^2_{OOS}
2009-2010	2011	64.42%	< 0.00%
2009-2011	2012	58.04%	71.36%
2009-2012	2013	45.64%	16.84%
2009-2013	2014	88.74%	72.90%
2009-2014	2015	55.12%	70.47%
2009-2015	2016	84.23%	89.55%
2009-2016	2017	91.24%	91.06%
2009-2017	2018	88.59%	91.69%
Average R^2_{OOS} across all train sets		72.00%	62.98%

in the catastrophe bond market can include regulatory changes, issuances with unusual features, demand forces etc but their identification for the time period we study is beyond the goals of this study. Our aim here is simply to examine the degree by which the trained models (random forest, and linear regression) can accurately predict catastrophe bond spreads even in presence of such changes. The temporal prediction performance challenge between random forest and linear regression is designed by using the train-test data set split approach in eight cycles of operation so that we can have a more complete picture of how models performance compare as the catastrophe bond market evolves. We start by using as train data set, the data from December 2009 to December 2010. We fit the random forest and linear regression models to this train data set and we make spread predictions using data from 2011. The second cycle of operation includes adding bonds from 2011 into the train data set and using bonds from 2012 as test set. The aforementioned process is repeated up until the train data set reflects the period up to 2017 and the test data set includes the catastrophe bond issuances in 2018. The prediction accuracy results measured in terms of out of sample R^2 (R^2_{OOS}) for each cycle of operation are presented in Table 6. It should be mentioned that assessing the prediction performance on a temporal context has a particular limitation. That is, new observations in a given test set cannot (directly) include categorical variable levels which did not appear in the respective train set. Thus, in order to avoid deleting deals having new levels in some of the categorical predictors in any given test year, we have imputed these values based on the most commonly observed categories in the corresponding training set accordingly.

We note that there seem to be some noticeable regime shifts especially in the first few cycles of operation. However, we cannot be definite about which model, the random forest or linear regression, handles regime changes best as in some years random forest does better than linear regression and vice versa. By looking at the variability of the R^2_{OOS} across all cycles for both models, it appears that the R^2_{OOS} range for random forest is between 45% and 91% whilst the respective range for the linear regression model is around between below 0% and 91%. It should be noted that in the first year, LR exhibits a very low temporal predictive power but we believe that this may be the result of the imputation in a small data set sample and perhaps the fact that some categorical variables contain quite granular information; see Table 13 in Appendix A. It is worth mentioning that the worst performance for both models is observed for the 2013 test set. By looking into how the regression model is parameterized for the test sample in 2013, it appears that the poor performance of the regression model on this test sample is largely due to the fact that catastrophe bonds in 2013 indicated record high EL values, i.e. the largest of them almost doubled the maximum EL value that was observed prior to 2013, for which models based on earlier observations might not be entirely suitable for the purpose of prediction. A potential reason

Table 7
A typical realisation regarding random forest prediction accuracy stability results.

Random forest summary	Sample A	Sample B
sample size	467	467
number of predictors	12	12
random forest type	regression	regression
number of trees	500	500
no. of variables tried at each split	5	6
node size	5	5
MSE_{OOB}	13855.34	14744.59
R^2_{OOB}	92.14%	90.71%

Table 8
Random forest stability measured in terms of minimum, mean, and maximum absolute difference of R^2_{OOB} between Sample A and Sample B across 100 iterations.

R^2_{OOB} Min Abs. Dif.	R^2_{OOB} Mean Abs. Dif.	R^2_{OOB} Max Abs. Dif.
0.01%	2.19%	6.92%

is the largest version change in the history of RMS model, implemented towards the end of 2012, which affected all 2013 renewals in having the potential to increase insured loss results even above 100% in some cases. Overall, it appears that the random forest is relatively more robust than linear regression in this respect. More comparisons between RF and LR when taking into account missingness of more than one predictor at a time follow in Section 8.5.

7.2. Random forest stability

We now examine the stability of random forest prediction accuracy results over the entire time period of interest. This is measured empirically from a practitioner's point of view as presented in Section 3.3.2 and in Table 7, we present a typical realisation of 1 out of 100 iterations with respect to the repeatability of prediction accuracy results.

As we observe in Table 8, across all 100 iterations, the recorded mean absolute difference of R^2_{OOB} between Sample A and Sample B for the catastrophe bond data set is 2.19% with the minimum and maximum absolute differences being 0.01% and 6.92% respectively.[11] Given that our problem sits in the intersection of financial and insurance market spheres where many behavioural aspects can affect prices, we consider the reported difference for the catastrophe bond data set being small. In essence, it is unlikely that an ILS fund would reject the use of the method solely for such a level of dissimilarity. In fact, the repeatability of prediction results here means that we can fairly safely say that our initial random forest prediction accuracy result, i.e. of an R^2_{OOB} of 96.57% presented in Table 5, is reliable, in the non-temporal context.

This finding is beneficial for the usage of the method in the industry. With new catastrophe bonds being issued, the random forest would need to be validated at some point in time as any other model in an insurance related firm. Surely, in a business context, there is no point in investing time and capital to introduce a new model if the latter provides accurate predictions strictly for one particular data set. Having gone through the examination of prediction accuracy results stability, we proceed with determining the importance of each independent variable in the study.

[11] As a robustness check, we have also repeated the random forest stability evaluation using two popular Open Source data sets, namely Boston Housing and Abalone, which are also used in the original paper of Breiman (2001) for the empirical assessment of the random forest method and are available at the UCI repository. The results are close with those derived for the catastrophe bond data set.

8. Predictor importance analysis

The importance of predictors is assessed using the methodologies of permutation and minimal depth importance presented in Section 3. It should be highlighted that the goal here is to find how powerful each independent variable is in predicting catastrophe bond spreads at issuance. No kind of relationship between spread at issuance and the predictors is to be established - the focus lies solely on their prediction ability. We then compare the stability of predictors' importance results for both methods. Then based on the ranking of the most stable predictors importance method, we examine the sensitivity of the random forest versus the benchmark to simultaneous missingness of multiple predictors in an effort to reveal and understand variables interactions. Next, by considering once again the most stable importance method, we examine the degree of similarity in predictors importance results in the predictive versus explanatory modelling frameworks. Finally, we discuss whether the rankings make empirical sense from investors' viewpoint.

8.1. Permutation importance

The importance of each independent variable in predicting catastrophe bond spreads has been here assessed on the basis of a percentage increase in MSE_{OOB} when a predictor is randomly permuted from the out of bag data whilst others remain untouched. First, the MSE_{OOB} for each of the 500 trees comprising the random forest, was recorded. The same process was repeated after randomly shuffling the values of a particular x_p across all observations. Then, the change between these two mean squared errors, before and after x_p permutation, has been calculated and averaged across the 500 trees after being normalised by the standard deviations of the differences. In this way, the importance score for x_p has been derived. Finally, based on these scores, an importance ranking has been produced. The ranking of catastrophe bond predictors based on their permutation importance score is shown in Fig. 4. Variables higher on the vertical axis are more important in predicting catastrophe spread at issuance with respect to this measurement.

One of the first observations is that all scores have positive value, indicating that each of the independent variables presented here does contribute towards prediction of catastrophe bond spreads. The predictors EL and RoL followed closely by term appear as the most important predictors of spread at issuance. In particular, when EL is shuffled, the out of bag mean squared error increases by around 41% whilst the respective percentages for RoL and term are slightly lower between 33% and 34%. Next, had any of the predictors; loc_peril and AP been randomly permuted, the prediction performance of the random forest would have been deteriorated between 31% and 32%. By shuffling the predictor iss_year, we see an almost 28% decrease in random forest prediction accuracy whilst the respective percentages for BB spread and size are in the range between 27% and 28%. Rating contributes to the reduction in the prediction accuracy of the random forest by around 19% and the least important predictors are coverage and vendor resulting in an approximately 16% and 13% prediction accuracy decrease respectively.

8.2. Minimal depth importance

The focus is now shifted from using a specific prediction performance measure to assess variables importance to a criterion based on the way that the forest was constructed, namely, the minimal depth. A tour over the constructed random forest was made to find

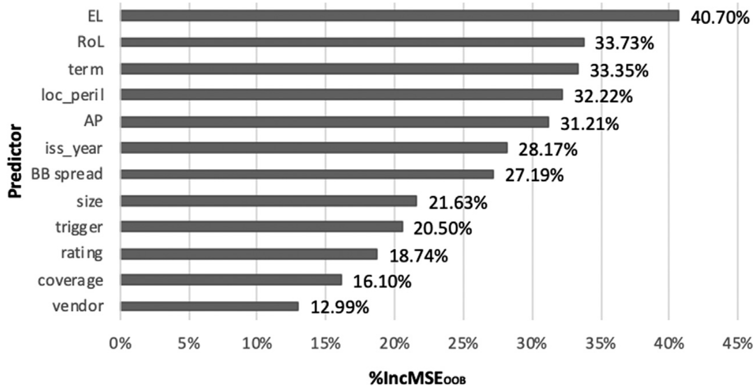


Fig. 4. Permutation importance based ranking of predictors. Predictors being permuted versus percentage increase in MSE_{OOB} as a result of the permutation.

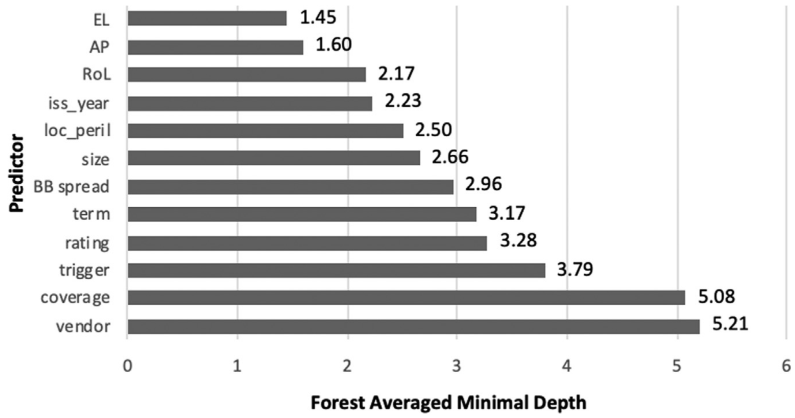


Fig. 5. Minimal depth importance based ranking of predictors. Predictors and their forest averaged minimal depth.

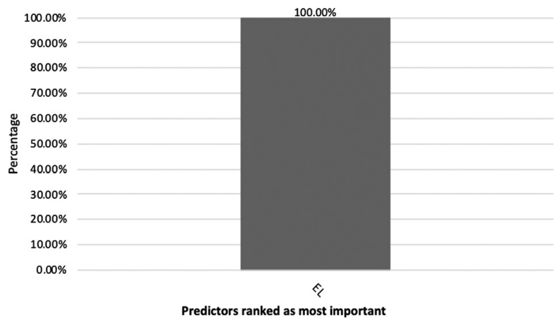
the maximal subtree[12] within each of the $K = 500$ trees for a particular x_p predictor. From there, the minimal depth for x_p within each tree was identified following the rationale explained in Section 3. Then, the forest level minimal depth for x_p was derived by averaging the minimal depth for x_p within each tree among all 500 trees. Fig. 5 illustrates the ranking of the covariates with respect to their average minimal depth; higher values of minimal depth correspond to less predictive variables.

Predictors EL and AP, with random forest average minimal depths of 1.45 and 1.60 respectively, have the largest impact in predicting catastrophe bond spreads. In particular, such small values of minimal depth demonstrate that these two variables were mostly used to split either the root node or any of its child nodes at least in most of the trees in the forest. Straight after in rankings comes the variable RoL followed closely by iss_year which on average were chosen to split a node for the very first time at a depth equal to 2.17 and 2.23 respectively. At similar level of importance stand the loc_peril and size with a level of depth still closer to 2.00 rather than 3.00 implying that they also have a considerable forecasting power. It appears that predictors BB spread, term, rating, and trigger were on average chosen to split the third node in the regression trees comprising the random forest. The aforementioned predictors appear as not being as powerful because they split nodes which naturally have less data points due to their proximity to the terminal nodes. Then the remaining variables, coverage and vendor, have minimal depth measurements of 5.08 and 5.21 respectively. These values are the highest among all predictors, revealing that coverage and vendor have the most limited forecasting ability out of all predictors.

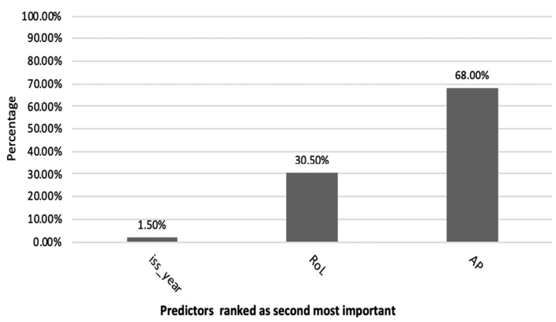
8.3. Divergence between permutation and minimal depth importance results

Permutation and minimal depth importance procedures presented for ranking or selecting catastrophe bond spread predictors above are not directly comparable. This is because, as it has been seen, each of them follows a different approach in defining and quantifying the importance in prediction. However, empirically we would expect that there should be some consensus between the two methods. What we see is that whilst there is indeed a degree of agreement for the very top and bottom of the rankings, there is some divergence at the upper middle ranks. This realisation makes us think which of the two variable importance approaches leads to the most trustworthy results for our catastrophe bond spread prediction problem. Indeed, empirically, an answer to this question would be to examine which ranking makes more sense from a practitioner's perspective. However, we believe that it is also preferable to bring our attention back to the concept of stability, but this time for the catastrophe bonds features importance. If one of the two methods is unstable, then we can shift our focus to the other one that is more robust and then discuss whether the ranking it provides makes sense from an investor's perspective. In the following, we present the stability checks for the importance results derived by both permutation and minimal depth importance.

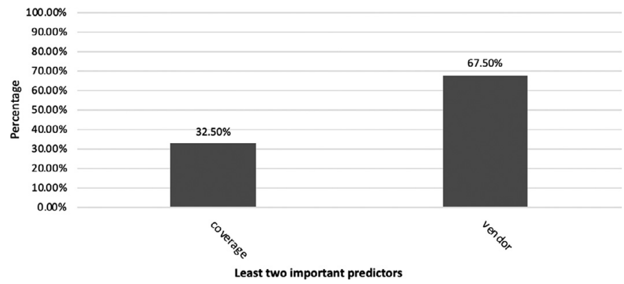
[12] See Section 3.4.2 for an explanation of what constitutes a maximal subtree.



(a) Predictor ranked as most important



(b) Predictors ranked as second most important



(c) Least two important predictors

Fig. 6. Bar plots showing the percentage frequency where a given predictor was ranked as top, second from top and in the last two positions for the most stable variable importance method in terms of ranking, i.e. minimal depth.

Table 9
Stability of ranking of predictors by different importance ranking method.

Ranking position	Agreement % (Permutation)	Agreement % (Mimimal depth)
Top	98%	100%
Second from top	36%	46%
Second from bottom	27%	69%
Bottom	22%	69%
Last two	10%	100%

8.4. Stability checks for predictors importance results

Here a predictor ranking method will be considered reliable if its importance ranking for catastrophe bond spread predictors is fairly robust to certain type of changes in the data set, such as random splitting. If a change in the catastrophe bond data set from which the random forest is constructed lead to a big change at the top and at the bottom of predictors importance rankings, then that particular importance ranking method will be considered unstable and thus probably unreliable.

Towards this direction, since both permutation importance and minimal depth importance are procedures derived internally after the construction of the random forest, the stability of permutation and minimal depth importance has been mainly examined based on the 100 random forests pairs grown out of 100 Sample A and Sample B pairs which have been previously used when the stability of the random forest was investigated in Section 7.2. In Table 9, we report by variable importance method, the percentage of times where there was an agreement between Sample A and Sample B in the predictor chosen at the top, second, and bottom positions of the ranking for all data sets. As bottom positions of the rankings we consider the last two positions jointly. This is because we understand that the further we go down the ranking, the more susceptible variables may jump from the position to its neighbours across different iterations. It is evident that minimal depth importance method provides more stable ranking results for both top, and bottom positions compared to permutation importance method. The biggest differences between the two methods are recorded for the second from bottom, bottom, and last two ranking positions combined where the discrepancy in the agreement percentage reaches 42%, 47% and 90% respectively. As previously highlighted in Chen and Ishwaran (2012), the complex randomisation element of permutation importance procedure makes it difficult to assess the underlying cause for it being relatively more unstable. However, it should be mentioned that this is not the first work when this measure showed an irregular conduct. As an example from bioinformatics, Calle and Urrea (2010) showed that permutation importance rankings were unstable to small perturbations of a gene data set related to the prognosis bladder cancer. All in all, it should be acknowledged that the appropriateness of a feature importance method is mostly data set specific and at least for the catastrophe bond set in hand it seems that permutation importance is not as reliable.[13] Based on the above, any discussion from now on which is relevant to predictors importance will be based on results of minimal depth importance as presented in Section 8.2.

Moving forward, it is interesting to examine stability within the minimal depth importance output with respect to which variable is chosen at a given position of the minimal depth importance rankings. In order to do so, we considered the number of counts out of 200 sub-samples taken in 100 iterations (or 400 samples taken in 100 iterations when we consider the last two ranking positions

[13] We have also examined the robustness of the predictors importance stability results using the Boston Housing and Abalone Open Source data sets, as we did in the case of random forest stability evaluation. The results align with those derived for the catastrophe bond data set, i.e. the minimal depth method, at least for the top ranking positions, appears to be more reliable compared to the permutation importance one.

Table 10
Sensitivity analysis for random forest (RF) versus linear regression (LR) to missing predictors using R^2_{OOB} as performance measure. The performance is examined by removing predictors sequentially based on the minimal depth importance ranking. Here we also report the R^2_{OOB} of RF and LR without any missing predictors to facilitate comparison.

Missing predictors	RF R^2_{OOB}	LR R^2_{OOB}
no missing predictors	96.57%	83.30%
EL	95.74%	79.98%
EL, AP	87.69%	50.68%
EL, AP, RoL	87.86%	47.30%
EL, AP, RoL, iss_year	84.83%	43.65%
EL, AP, RoL, iss_year, loc_peril	84.56%	33.33%
EL, AP, RoL, iss_year, loc_peril, size	68.81%	30.76%
EL, AP, RoL, iss_year, loc_peril, size, BB spread	30.86%	21.18%
EL, AP, RoL, iss_year, loc_peril, size, BB spread, term	18.89%	17.69%
EL, AP, RoL, iss_year, loc_peril, size, BB spread, term, rating	12.23%	10.40%
EL, AP, RoL, iss_year, loc_peril, size, BB spread, term, rating, trigger	7.60%	7.69%
EL, AP, RoL, iss_year, loc_peril, size, BB spread, term, rating, trigger, coverage	5.67%	6.03%

Table 11
Sensitivity analysis for random forest (RF) versus linear regression (LR) to missing predictors using R^2_{OOB} as performance measure. The performance is examined by randomly removing M predictors from the original data set for $M = 1, \dots, 11$. For each M, this experiment is repeated 100 times, and the average R^2_{OOB} is reported. Here we also report the R^2_{OOB} of RF and LR without any missing predictors to facilitate comparison.

Number of missing predictors at random - M	RF R^2_{OOB}	LR R^2_{OOB}
no missing predictors	96.57%	83.30%
1	96.21%	82.47%
2	95.87%	80.58%
3	95.19%	78.19%
4	92.84%	73.56%
5	91.01%	70.78%
6	89.28%	67.34%
7	73.76%	59.74%
8	67.49%	51.74%
9	60.08%	43.45%
10	45.02%	29.22%
11	28.09%	16.97%

jointly), where a given predictor was ranked as top, second from top, or in last two positions in terms of importance by variable importance method. The results are shown in Fig. 6 in terms of percentage frequency. We see that minimal depth method is also fairly stable with regards to its predictors' choices for the examined ranking positions. That said, in the top position the predictor EL was chosen 100% of the times and only a small variation is visible for the second from top and last two ranking positions. In the next section, we provide some further analysis on how well the random forest handles missingness of important predictors as opposed to LR model.

8.5. Further analysis - on handling missingness of important variables

We now assess the sensitivity of prediction accuracy of random forest in the absence of important predictors, and contrast the outcomes with those from the benchmark model. Doing so also allows us to understand and characterise interactions between predictors. Here we consider removing more than one predictor each time and then report the resulting prediction accuracy of both random forest and the benchmark model. The removal of predictors is made firstly, sequentially based on the minimal depth ranking presented in Section 8.2, from the most important one to the least important one, and secondly, by (uniformly) randomly dropping M predictors from the original data set for $M = 1, \dots, 11$. For each M, the second experiment is repeated 100 times, with the average R^2_{OOB} computed for both RF and LR. The sensitivity results are presented in Table 10 and Table 11.

When predictors are removed sequentially according to the minimal depth ranking, it appears that random forest prediction accuracy results seem to be considerably more robust compared to the ones derived from LR when the most important predictors, as identified in the minimal depth analysis, such as EL and AP, are jointly missing. For example, when the most important predictors EL and AP are excluded from the analysis, the RF prediction accuracy drops by around 8% as opposed to 29% in the case of LR compared to the respective prediction performances when only EL, i.e. the most important predictor, is missing. This may be an indication that there are potentially interactions, as well as non-linearities, between the predictors, which random forest appears to be capturing whereas the linear regression model struggles. Another observation is that we see a significant drop in random forest prediction accuracy when size and even more so BB spread are included in the missing predictors set. In particular, when size is removed, RF prediction accuracy deteriorates by 16%, i.e. the biggest drop up to this point since the beginning of the minimal depth based sequential removal of predictors. When BB spread is excluded, RF prediction accuracy declines by an additional 38% which is the highest drop in RF prediction accuracy across the whole experiment. A potential interpretation is that there is a certain degree of information redundancy among all the predictors. Here the predictors size and BB spread contain a large amount of useful information of all its predecessors found to be of higher importance in catastrophe bond spread prediction, which can be effectively extracted by random forest.

Similar observations are made when randomly removing M predictors from the original data set for $M = 1, \dots, 11$ repeated 100 times and taking the average of R^2_{OOB} for the RF and LR respectively. RF still shows a better predictive performance than LR having an average R^2_{OOB} of around 90% even by randomly dropping half of the variables, again forcing the impression that RF is more flexible than LR and is likely better at capturing interactions and dealing with possible missingness of the predictors. It should be noted that thanks to the random dropping mechanism, the results in Table 11 appear smoother than these in Table 10 where we exclude the most important variables first - a strategy which acts more like assessing the worst case scenario. In summary, random forest is better at borrowing strength from existing predictors to (partially) recover the predictive power lost due to the absence of important predictors.

8.6. Predictive versus explanatory importance

Now we discuss whether the importance results in our predictive framework agree with those presented in explanatory models of past works but also the LR model in the current study.

As mentioned in Shmueli (2010), variables which are considered important in explaining the response are tied to theoretical hypotheses which are set at the beginning of the study, and on the notion of statistical significance. These aspects are immaterial in a purely predictive modelling framework as the one we present by using random forests. Exploring the level of this divergence is meaningful, as it can add value in understanding the full spectrum of catastrophe bond spread drivers for both prediction, and explanation. It should be mentioned that this is an exercise that shall be made with extra caution as, to our best knowledge, every study in the explanatory catastrophe bond pricing literature to date and our predictive study has utilized different data sets and made different assumptions (apart from the LR model). However, given the fact that satisfactory level of agreement has been recorded in the past for certain variables in the explanatory framework, even under these constraints, it merits a short discussion.

The starting point is independent variables where harmony with respect to predictive and explanatory importance between this and previous studies has been observed. In particular, in Section 8.2, it is seen that EL is the most major contributor in predicting spreads in the primary catastrophe bond market. This result comes in agreement with our LR model presented in Table 3, and the majority of the previous explanatory oriented literature, see Lane (2000), Lane and Mahul (2008), Bodoff and Gan (2009), Dieckmann (2010), Braun (2016), Galeotti et al. (2013), and Jaeger et al. (2010). In Lei et al. (2008), the conditional expected loss is considered instead of expected loss, despite the fact that the former is not found to be statistically significant, while other variables related to the loss distribution are.

At the same time, in this study we observe that the probability of losses outstripping the attachment point has almost equal forecasting power as the expected loss. Moreover, we see that the predictor AP is statistically significant in LR too. Lane (2000) also supports that the catastrophe bond premium is derived through an interplay between frequency and severity of catastrophe bond expected losses. On the top of this, Lei et al. (2008) and Jaeger et al. (2010) agree with the view that the attachment probability is of high significance in explaining catastrophe bond spreads. Moving forward, the importance of variables reflecting the cyclicality of the market is high both in a predictive, and explanatory context, see LR, Lane and Mahul (2008), and Braun (2016). At the same time, peril-territory combination which is found particular importance for its ability to forecast spreads here and in the explanatory framework. In particular, alike results are obtained by LR, Gatumel and Guegan (2008), Jaeger et al. (2010), and Götze and Gürtler (2018). Similarly, trigger is predictive in the current research whilst Dieckmann (2010), Götze and Gürtler (2018) and Papachristou (2011) also commented about the explanatory significance of this variable in their models. Finally, the predictor rating which is found to be predictive in our study (although not of top importance), is seen as major determinant of spread in our LR model, and also in Lei et al. (2008), and Götze and Gürtler (2018); even though Götze and Gürtler (2018) have examined rating from a different perspective to the one we employ, i.e. the variable related to rating does not refer to the credit quality of the bond but to that of the cedent instead.

With respect to the predictor term, no general consensus on its statistical significance has been reached in the literature up until now, although here it appears to be relevant for both prediction and explanatory purposes as LR reveals. For example, Papachristou (2011) and Braun (2016) exclude the variable term from their analysis whilst on the other hand Dieckmann (2010), Galeotti et al. (2013), and Gürtler et al. (2016) highlight its importance. At the same time, the predictor size is minded as less influential or not significant at all by the models of Papachristou (2011), Lei et al. (2008), Braun (2016), and LR (zero coefficient even if the variable is significant) but it is considered sufficiently important for prediction purposes in our study. This divergence may once again stem from the way weak predictors are treated in a typical linear regression model versus random forests. As it is mentioned by Berk (2008), in a traditional regression framework a variable having a very small association with the response is most often excluded from the model being regarded as noise. Nevertheless, a big number of small associations when considered not on an individual basis but on an aggregate level can have a substantial impact on fitted values. That is not to say that linear regression is not capable of capturing interactions, however to do so any interactions need to be explicitly specified - a complicated task when the number of predictors in the study starts increasing. On the contrary, random forests, as a tree based method, is naturally able to capture associations between predictors without the need to specify them. Indeed, Papachristou (2011) also acknowledges that in the context of his study, the fact that the term is not considered as important enough to be included in the suggested model may be due to the challenge of capturing complex effects between covariates. Coming back to the discrepancy between explanatory and predictive power for predictor size, we recall that in Section 8.5 the interacting behaviour of this variable is also observed in the predictive framework.

Finally, our study indicates that the variables vendor and coverage are predictive despite of their appearance at the bottom of the ranking. Since this is the first time that these variables are studied, we can only compare them with LR in the explanatory framework. In particular, vendor does not appear as a statistically significant variable whilst coverage is. Overall, we can conclude that explanatory (based on LR and past literature) and predictive power appears to coexist for all catastrophe bond spread drivers considered in our study apart from size and vendor.

8.7. Discussion of predictors' importance results from an industry perspective

Looking broadly at the minimal depth ranking presented in Fig. 5, we observe that the predictors may fall into three groups: those of utmost (the top two), medium (the next seven) and low prediction strength (the last two). We acknowledge that the bounds of where medium and lowest importance variables groups start may be subjective. The distinction here is made looking at the ranking from the perspective of a practitioner. The reason why we want to avoid focusing on individual importance scores is that explaining results in such a detailed way would neither be appropriate nor meaningful for a prediction oriented study. This section is not about interpreting results but seeing whether the results capture somehow investors' perception and knowledge of the market.

Having explained our rationale, the group of top importance predictors comprises from the two fundamental ingredients in any risk quantification process, that is the product of severity and frequency of losses, i.e. EL, and AP. This is something that would most probably not surprise insurance professionals, risk managers or even investors if the variable importance results were to be presented to them. Especially with respect to investors, it is well comprehended that the return to be earned by investing into a catastrophe bond deal needs to surpass the expected value of catastrophe bond payouts. Thus, from an empirical viewpoint, investors would expect that by knowing the expected loss and probability of them losing the first dollar, at least a part of the spread value can be predicted.

The second group refers to some cyclical market elements and catastrophe bond features which could influence investors' interest in a deal. The high importance of cyclical aspects in the prediction of a new issuance spread is somehow natural since a hard or soft market directly sets some bounds on the top of which a deal's specific loss profile and characteristics would be assessed. One reason why certain catastrophe bond features could influence an investor's appetite considering a deal, is the effect that these features could have on investors' portfolio returns. In particular, investors would most probably agree with the predictor loc_peril having a high position in the ranks, as this type of information acts as the window shop for them entering the transaction. The rarity of the peril combined with the coverage territory indirectly informs investors about the diversification effect that the particular security can bring into their portfolio; a significant incentive for them to invest in this asset class. We acknowledge that this may not be true for new or rare perils, for which the existing catastrophe models are not yet trusted, however even in this case the peril-territory combination is informative in this sense. Another reason why the predictors of the second group could trigger investment interest is because some of these features are typical in traditional bond types traded in the financial markets and investors are already accustomed to this type of information such as issuance size, BB spread level, time between issuance and maturity date, credit rating related information, and trigger of payment. Consequently, one can say that the location of these variables in the ranking supports the way an average investor would think even for a typical non-insurance linked investment.

Finally, the last group of predictors in the importance ranking comprises from variables having strong technical weight in the securitization process and being insurance sector specific. The first predictor in this group, i.e. coverage type, refers to a contract term found in insurance contract whilst the second one, i.e. vendor, to the software company used to calculate the expected loss and various loss probabilities. Whilst this may not be immaterial information, there is not direct equivalent of such features in the financial markets. Thus, the average investor not specialising in insurance linked securities would not really dig deep into analysing vendor model updates, and historical loss catalogues, or even the wording of the transaction when thinking of returns prediction. Especially for vendor, it is a matter of fact that there is a global oligopoly in firms offering catastrophe risk modelling solutions in the insurance industry. Although the software developed by each of these companies is based on different assumptions, their scientific grounds are not disputed in the marketplace. This can be mostly attributed to the fact that these companies have been founded years before the birth of the first catastrophe bond and also that they have a long track record of being used in the traditional insurance and reinsurance markets. Thus, there is a contract of trust between them and the market participants as all vendors are perceived to be of equivalent reputational standing. Having said that, it does not mean that investors are sure about the reliability of the expected loss computation. It is just that most likely they would not believe that one vendor will have a much more valid estimate of loss compared to another. Similarly, coverage type really matters from an investor's perspective when seen in conjunction with the trigger or the combination of peril and geography. For example, catastrophe bonds with indemnity triggers or not well understood risks when combined with aggregate coverage terms can be risky in trapping investors' capital, as it was seen after 2018 Californian wildfires (Risk 2019). Taking into account all the above, the minimal depth predictors' importance ranking seem to reasonably reflect investors' current understanding of the market.

9. Example of random forest application in the industry

In this section, we present some possible examples of how the random forest could add value to ILS industry participants' daily operations. In particular, we discuss how the random forest could assist a would-be catastrophe bond issuer or investor in making faster and more informed decisions. In other words, we attempt to showcase examples of random forest applicability both from the "buy" and "sell" sides of the catastrophe bond market.

Starting from the sell side, a would-be catastrophe bond issuer along with their investment advisors, prior to finalising the terms of a new catastrophe bond issuance, would use the random forest to predict the likely spread at which investors would accept the offering. Getting to know this information is important as it allows for exploration of terms which would make the deal appear more attractive to an investor. In case this would not be feasible, the would-be issuer would realise faster that it may be preferable to explore alternative risk financing options.

From the buy side point of view, the random forest could also be beneficial to investors. In particular, just before a new catastrophe bond is issued, potential investors are provided with an offering circular. This document includes information about the deal which is to be launched and an invite for them to attend a road show, post which the issuance pricing will be settled. The information disclosed in this package refers to risk details, various design characteristics of the issuance and a price guidance. Investors want to make sure that the suggested spread compensates them enough for the true element of risk that they would undertake had they entered the transaction. However, a detailed analysis of this aspect can be time consuming as various departments and sometimes even external risk modelling firms get involved in the process. Whilst this process is undoubtedly important, investors would like to have a first flavour for a new deal's potential faster. Then, let's imagine how useful a straightforward prediction tool like random forest would be, where investors could plug in details provided in the circular of the new issuance the moment they receive it to get a quick spread prediction for the new transaction they investigate on the spot. This prediction would then be compared with the spread guidance offered and give investors an initial idea on whether the bond is overpriced, under-priced or "fairly" priced based on past catastrophe bond experience. This would direct investors to identify bargains faster and ask more relevant questions about the deal whilst on the road show. Then if the deal would be of interest, they could send all information needed to their modelling teams to perform the usual tasks of remodelling the underlying risk exposure and calculate the marginal impact that this new investment would bring into their portfolio. Overall, random forest is a solution that can speed up the investment decisions and help ILS investment firms not to use their valuable human resources for irrelevant catastrophe bond deals. As mentioned in Section 3.3.1, random forest could also be used to populate incomplete catastrophe bond deals databases when there is uncertainty or missingness of spread values for past transactions. We believe that its suitability for this purpose is very likely given the fact that we have some evidence about its high non-temporal prediction accuracy (see Section 7.1.1), and its "robustness" when information for more than one predictor is missing simultaneously (see Section 8.5) - a relatively usual phenomenon in an opaque market setting.

Besides, one note that needs to be made is that when assessing the discrepancy between the predicted spread value provided by the random forest (which for the buy side is the price guidance, and for the sell side it is the price for which the issuer would think that investors would accept the deal), one might first want to look back at what happened in the past, i.e. the historical discrepancy between the predicted and actual values recorded in the prediction phase post the random forest training. This may shed some light on the level at which a mispriced deal according to random forest is due to the portion of variability that the random forest could not explain or merely due to the fact that the new catastrophe bond has characteristics that have never been recorded in the past. The latter problem, could be mitigated if the random forest would be re-trained at frequent intervals, as part of the model validations taking place at least annually in a business context, enriching the training data set with more deals.

Finally, although many other parameters could be taken into account for random forest to be incorporated into internal business processes, here we give an idea of how the prediction power of random forest can liaise with issuers and investors' personal judgement to make faster and more informed decisions. It should be highlighted that recent developments in the catastrophe risk market also support the use of machine learning techniques. Prime examples are the new cyber risk model of AIR vendor, see AIR (2018), and a new platform for analysing deals and facilitating transparency in the catastrophe bond market, see Jones (2019).

10. Concluding remarks and future research

Until recently, the data-driven catastrophe bond pricing literature was mainly focused on building statistical models with an aim to test causal theory. The centre of interest lied on identification of variables which have a theoretically material and statistically significant link to catastrophe bond price, i.e. hypotheses of relationship between price and each independent variable were made. Then a statistical model, mostly linear regression, was applied to observed data to compute the size of this effect and the statistical significance of each independent variable in relation to the causal hypotheses set at the beginning. For model evaluation, in sample R^2 has been the classical way to assess model success, even though few more recent studies, such as Galeotti et al. (2013), Gürtler et al. (2016), and Braun (2016) have also considered out of sample model performance, and in some cases robustness checks for stability over different time periods. Model selection happened on the basis of keeping statistically significant factors and sometimes those non-significant ones having large coefficients to match the function connecting catastrophe bond spread and factors to the true underlying catastrophe bond data generation process.

The approach presented in the current research study was fundamentally different. A machine learning method called random forest was applied to a rich primary market catastrophe bond data set with a goal to predict catastrophe bond spreads at issuance given information in the offering circular and knowledge about current market conditions available at the time of prediction. Here, we did not focus on the underlying data generation process instead we learned the association between catastrophe bond spreads and predictors from the data directly using the random forest. The performance of our method was assessed on how accurately it predicted spreads based on unseen catastrophe bond observations on both temporal and non-temporal bases as well as the sensitivity of this prediction accuracy when possibly interacting predictors are missing. Variable importance measures referred to predictive ability and not the power to explain how the spreads are generated in this universe. There was also interest in securing repeatable prediction accuracy and predictors' importance results because of the multiple levels of randomness incorporated in random forests thus relevant checks were performed. The degree of divergence between predictive and explanatory importance was also of interest.

It was found that random forest has at least as good prediction performance as linear regression in the temporal context, and better prediction performance in the non-temporal one. Random forest performed better than linear regression when multiple predictors were missing from the model, as it has the ability to capture and extract interactions between existing variables. By assessing variables' importance on a non-explanatory basis, we found that all examined predictors have a say in the prediction of spread even if this is in varying degrees. The prediction accuracy, and predictors' importance results of random forest were stable. Taking prior explanatory literature and LR model into account, it appeared that predictive and explanatory power coexist for all catastrophe bond spread drivers considered in our study apart from size and vendor. There is potential for random forest to be used in the catastrophe bond industry to fast track investment decisions from both the buying and selling sides.

Based on the above findings there are certain aspects that would be interesting to research in the future. Although by using random forest as presented here, an investor, for instance, can see whether a new issuance of any type has a competitive price guidance or not, they do not get informed about the suitability of a new deal given their current portfolio composition. Addressing this need is a significant and important topic for future research. Another subject for future study is to extend our data set prior to 2009 to focus on the years of the financial crisis, and also additionally examine whether the drivers of private placements differ compared to those of non-private catastrophe bond deals. Finally, for the explanatory framework, another direction is for the variables size and BB spread to be further investigated as they stand out due to their potential interactions with other variables when other important variables are missing in the context of random forest.

In conclusion, our research provides some evidence that utilising both predictive and explanatory modelling can enhance the understanding of catastrophe bond market segment, increase its transparency and contribute to its development.

Declaration of competing interest

The authors declare that they have no conflict of interest.

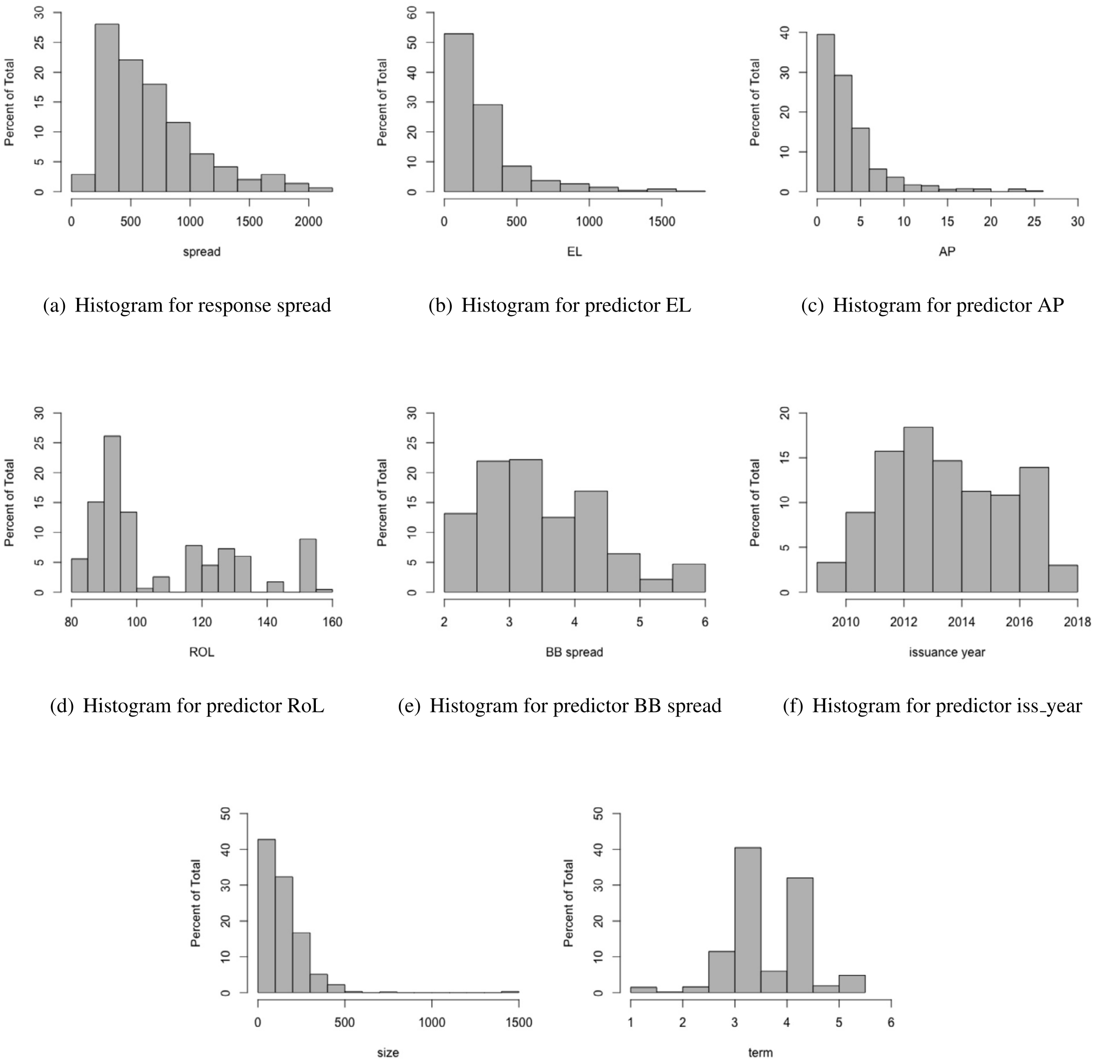
Acknowledgements

We thank the anonymous reviewers for their careful reading, whose insightful comments and suggestions helped significantly improve the content and presentation of this manuscript.

Appendix A. Summary statistics for the catastrophe bond data set

We now provide further information about the catastrophe bond data set used in this research paper. Summary statistics are presented for all variables, both continuous and categorical ones. Starting from the continuous variables, we present histograms in Fig. 7 and measures of central tendency and spread of the observations in our data set in Table 12. In Fig. 7, we see that all continuous variables have a right skewed distribution except variables RoL, BB spread, iss_year, and term. In particular, we see that the majority of catastrophe bond issuances in our data set corresponded to a RoL value of less than 100 indicating a soft market. Moreover, most of catastrophe bonds were issued in the year 2012-2013. It appears that term distribution has two peaks reflecting that most catastrophe bond issuances have a 3 to 5 year time horizon. Looking at Table 12, we notice that the range between minimum and maximum values for all continuous variables as well as the interquartile range are rather broad indicating that data points are well spread out. Such a data structure is anticipated in a catastrophe bond market setting. In essence, each issuance is a bespoke product developed to meet a very specific risk transfer need and consequently the population of catastrophe bond deals is heterogeneous.

Moving forward to categorical variables in Table 13, we present for each of them the number of level and number of observations under each level, with the latter quantity also being expressed as a percentage of the total number of observations. All variables levels are those used by the industry unless otherwise stated. Some comments regarding each categorical variable follow. With regards to coverage type, we find that the majority of catastrophe bonds during the studying period were issued to provide compensation in situations where a single large-scale loss event would activate the



(a) Histogram for response spread (b) Histogram for predictor EL (c) Histogram for predictor AP

(d) Histogram for predictor RoL (e) Histogram for predictor BB spread (f) Histogram for predictor iss_year

(g) Histogram for predictor size (h) Histogram for predictor term

Fig. 7. Histograms for the continuous variables. Percentage of total observations versus different ranges of a given numerical variable.

Table 12
Continuous variables summary statistics. The unit in which each continuous variable is measured is provided in brackets.

Continuous variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
spread (in basis points)	50.00	375.00	590.0	687.70	871.50	2200.00
EL (in basis points)	1.00	111.00	188.50	274.60	333.80	1735.00
AP (%)	0.02	1.36	2.51	3.72	4.68	25.04
RoL (in basis points)	83.57	91.73	96.06	106.97	124.57	159.73
BB spread (%)	2.12	2.61	3.41	3.46	4.14	5.98
iss_year (as numeric value)	2009.00	2012.00	2014.00	2014.00	2016.00	2018.00
size (in million US dollars)	3.00	75.00	130.00	164.70	200.00	1500.00
term (in years)	1.00	3.02	3.18	3.49	4.02	5.12

trigger, i.e. per occurrence coverage, as opposed to this happening due to a collection of insured loss events i.e. aggregate coverage. In very few instances in the data set, such as tranches A and B of Riverfront Re Ltd Series 2017-1 for example, per occurrence and annual aggregate coverage co-existed.

With respect to loc_peril, we shall start by providing some explanations in terms of abbreviations. The first part in each loc_peril level name indicates (a) geographical region(s). In particular, APAC stands for perils specific to Asia Pacific region, NA for perils relevant to North America, SA for prominent perils in South America,

Table 13
Summary statistics for all the categorical variables. Levels of each categorical variable are presented by number of observations and percentage of total observations. Abbreviations are explained in the text.

Categorical variable	Levels	No. of observations	Percentage (%)
coverage	aggregate	303	32.4
	occurrence	627	67.1
	both	4	0.5
loc_peril	APAC_Quake	51	5.46
	APAC_Typh	22	2.36
	Europe_APAC_Multi_Peril	2	0.21
	Europe_Quake	12	1.28
	Europe_Wind	54	5.78
	NA_APAC_Multi_Peril	26	2.78
	NA_Europe_APAC_Multi_Peril	36	3.85
	NA_Europe_Multi_Peril	39	4.18
	NA_Multi_Peril	425	45.50
	NA_Quake	80	8.57
	NA_Wind	184	19.70
	SA_Quake	3	0.32
rating	B	141	15.09
	BB	286	30.62
	BBB	4	0.43
	CCC	4	0.43
	nr (**n**ot **r**ated)	499	53.43
trigger	indemnity	511	54.7
	parametric	29	3.1
	industry loss index	325	34.8
	parametric index	23	2.5
	model	22	2.4
	multiple	24	2.6
vendor	AIR	741	79.3
	AON	4	0.4
	EQECAT	42	4.5
	RMS	141	15.1
	pp	6	0.6

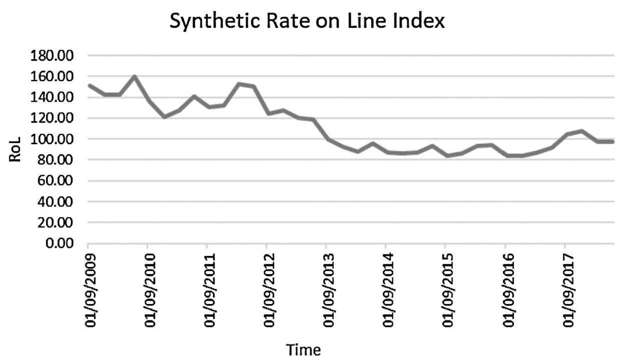


Fig. 8. Historical development of the Lane Financial LLC Synthetic Rate on Line Index (measured in percentage terms). Values above 100 indicate a hard market.

and Europe for perils in the aforementioned region. What follows the geographical region code, for instance APAC, is the peril type covered in the aforementioned location. There, except for those that are self-explanatory, Typh stands for typhoon and Multi_Peril includes various individual perils in the earlier indicated regions. For instance, one out of the NA_Europe_APAC_Multi_Peril tagged transaction provide cover against US named storms, Canadian earthquake, European earthquake, Australian wind and Australian earthquake. We see that almost half of the catastrophe bond deals in the data set had a mixture of perils in NA geographical territory which is quite expected since the perils in the area are generally considered to be more well understood and there is a longer heritage of issuances there. For example, bonds covering wind in North America are very popular even if the assumption of losses in the area is more likely due to the effect of hurricane seasons. Nevertheless, the high frequency of events had allowed risk modelling companies to understand the risk better, and build more trustworthy models with investors feeling more secure to buy exposures in this region. Looking into the credit quality allocation of the bonds issued, it is evident that more than 99% of catastrophe bonds in the data set either were characterised as non-investment grade securities or they did not receive a rating by any independent credit quality agency - the latter point has already been discussed more thoroughly in Section 4.3.

With regards to triggers, indemnity ones were the most popular among the bonds included in the study followed by industry indices. This clearly shows a preference from cedents' perspective to get compensated for the exact level of losses that they anticipate to experience or at least to be compensated in line to industry losses. Deals which are triggered when pre-determined event parameters are satisfied or surpassed accounted only for 5.6% of the total market in the period under study. Examples of parametric index deals in the current data set is Atlas VI Capital Ltd. Series 2010-1 and Bosphorus Ltd. Series 2015-1 whilst IBRD CAR 118-119 is an example of pure parametric trigger deal issued by the International Bank for Reconstruction and Development for Mexico's natural disaster fund named FONDEN. The least used triggers were those combining different trigger types such as Fortius Re II Ltd. Series 2017-1 and those based on the modelled losses of the cedent's exposure portfolio calculated based on event parameters gathered from specified agencies, such as Akibare II Ltd. single tranche.

With respect to the risk modelling company used to calculate the expected loss of investors' exposure to underlying peril, we see that AIR Worldwide is the most widely used followed by RMS. Together, they account for the 94.4% of all non-life securitisations in the data sample followed by EQECAT, AON and pp accounting for the rest 5.6%. It is worth to note that pp abbreviation is not a risk modelling firm but it stands for **p**rivate **p**lacement. Examples are the single tranches of Merna Re Ltd. Series 2016-1, 2017-1, 2018-1 which were privately purchased by specialized ILS funds. Finally, the internal model of AON was used for very few deals where the aforementioned company had acted as the structuring and placement agent, such as in the case of Windmill I Re series 2013-1.

Appendix B. In sample and out of sample performance of LR model using the variable Investment Grade (IG) instead of the variable (granular) rating

See Tables 14 and 15.

Table 15
Out of sample performance measured in terms of R^2_{OOB}, R^2_{10CV}, and R^2_{LOOCV} for the improved linear model of Braun (2016) versus the linear regression model with **I**nvestment **G**rade (IG) variable to indicate credit quality (LR with IG), and the benchmark linear regression model (LR) which includes the variable rating presented in Table 1 and Table 13 in Appendix A.

Model	R^2_{OOB}	R^2_{10CV}	R^2_{LOOCV}
Improved Braun (2016)	79.71%	80.81%	79.40%
LR with IG	82.22%	82.35%	82.72%
LR (with granular rating)	83.30%	84.42%	83.84%

Table 14
In sample fit of the linear regression model with **I**nvestment **G**rade (IG) variable to indicate credit quality (LR with IG) as opposed to variable rating presented in Table 1 and Table 13 in Appendix A.

	Estimate	Std. error	t value	Pr(> |t|)
(Intercept)	61540	9677	6.36	0.000 ***
RoL	6.05	0.43	14.15	0.000 ***
BB spread	50.34	8.66	5.81	0.000 ***
IG 0 (*baseline*)				
IG 1	−253	85.34	−2.96	0.003 **
IG nr (**n**ot **r**ated)	87.99	15.71	5.6	0.000 ***
term	−27.25	8.86	−3.07	0.002 **
size	0.00	0.00	3.06	0.002 **
trigger industry loss index (*baseline*)				
trigger indemnity	−0.58	14.77	−0.04	0.969
trigger model	−92.47	39.91	−2.32	0.021 *
trigger multiple	−44.9	40.13	−1.12	0.264
trigger parametric index	−23.9	42.02	−0.57	0.570
trigger parametric	−122.6	37.4	−3.28	0.001 **
coverage aggregate (*baseline*)				
coverage both	55.80	85.00	0.66	0.512
coverage occurrence	−59.95	13.95	−4.3	0.000 ***
vendor AIR (*baseline*)				
vendor AON	101.1	90.59	1.11	0.265
vendor EQECAT	−0.68	33.96	−0.02	0.98
vendor pp	21.33	73.65	0.29	0.772
vendor RMS	15.4	19.25	0.8	0.424
AP	−16.32	6.09	−2.68	0.008 **
EL	1.33	0.09	15.22	0.000 ***
iss_year	−30.79	4.79	−6.42	0.000 ***
APAC_Quake (*baseline*)				
loc_peril APAC_Typh	−77.74	44.6	−1.74	0.082
loc_peril Europe_APAC_Multi_Peril	−9.11	129.2	−0.07	0.944
loc_peril Europe_Quake	−13.43	57.83	−0.23	0.816
loc_peril Europe_Wind	−138.1	40.7	−3.4	0.001 ***
loc_peril NA_APAC_Multi_Peril	100.1	47.48	2.1	0.035 *
loc_peril NA_Europe_APAC_Multi_Peril	152.7	42.51	3.6	0.000 ***
loc_peril NA_Europe_Multi_Peril	149.9	40.96	3.66	0.000 ***
loc_peril NA_Multi_Peril	166.8	28.25	5.9	0.000 ***
loc_peril NA_Quake	−19.79	35.59	−0.55	0.57
loc_peril NA_Wind	97.83	30.72	3.18	0.002 **
loc_peril SA_Quake	133.1	107.4	1.24	0.216
R^2	83.96%			
Adjusted R^2	83.41%			
Res. Std. Error	166.8 (df = 902)			
F Statistic	152.3 (df = 31; 902)			
Note for signif. codes:	*p < 0.1; **p < 0.05; ***p < 0.01			
Observations number:	934			

Appendix C. In sample and out of sample performance of Braun (2016) model using a subset of our catastrophe bond data

See Tables 16 and 17.

Table 16
In sample fit of candidate benchmark models specification. Here, the linear regression model of Braun (2016) was applied on our catastrophe bond data set. We notice that only 434 data points are considered as the binary Investment grade variable does not take into account non-rated transactions.

	Estimate	Std. error	t value	Pr(> |t|)
(Intercept)	−665.97	40.33	−16.51	0.000***
Swiss Re	−20.12	14.57	−1.38	0.168
RoL index	5.24	0.44	12.02	0.000***
BB spread	57.18	11.66	4.91	0.000***
Investment grade	−39.17	73.32	−0.53	0.593
Peak territory	224.95	19.22	11.70	0.000***
Expected Loss	1.64	0.07	23.85	0.000***
R^2	79.97%			
Adjusted R^2	79.69%			
Res. Std. Error	143 (df = 428)			
F Statistic	284.8 (df = 6; 428)			
Note for signif. codes:	*p < 0.1; **p < 0.05; ***p < 0.01			
Observations number:	434			

Table 17
Out of sample performance measured in terms of R^2_{OOB}, R^2_{10CV}, and R^2_{LOOCV} for the linear model of Braun (2016) versus the linear regression (LR) in this study.

Model	R^2_{OOB}	R^2_{10CV}	R^2_{LOOCV}
Braun (2016)	79.20%	80.40%	79.12%
LR	83.30%	84.43%	83.84%

Appendix D. Prediction accuracy performance of RF with categorical dummy variables

See Table 18.

Table 18
Prediction accuracy performance measured in terms of R^2_{OOB}, R^2_{10CV}, and R^2_{LOOCV} for random forest (RF) when converting all the categorical variables into dummies in the catastrophe bond data set.

Model	R^2_{OOB}	R^2_{10CV}	R^2_{LOOCV}	R^2
RF	96.57%	96.49%	96.59%	99.25%
RF_dummies	96.48%	96.16%	96.63%	99.18%

References

AIR, 2018. AIR develops advanced probabilistic model for global cyber risks. https://www.air-worldwide.com/Press-Releases/AIR-Develops-Advanced-Probabilistic-Model-for-Global-Cyber-Risks/. (Accessed 19 May 2019).
ARTEMIS, 2019. Decline in ILS ratings shows the asset class isn't so alternative: Kbra. http://www.artemis.bm/news/decline-in-ils-ratings-shows-the-asset-class-isnt-so-alternative-kbra/. (Accessed 19 March 2019).
Bauer, E., Kohavi, R., 1999. An empirical comparison of voting classification algorithms: bagging, boosting, and variants. Machine Learning 36 (1–2), 105–139.
Berk, R.A., 2008. Statistical Learning from a Regression Perspective, 2 edn. Springer.
Biau, G., Scornet, E., 2016. A random forest guided tour. Test 25 (2), 197–227.
Bodoff, N.M., Gan, Y., 2009. An analysis of the market price of cat bonds. In: Casualty Actuarial Society E-Forum, Spring 2009.
Braun, A., 2012. Determinants of the cat bond spread at issuance. Zeitschrift für die gesamte Versicherungswissenschaft 101 (5), 721–736.
Braun, A., 2016. Pricing in the primary market for cat bonds: new empirical evidence. The Journal of Risk and Insurance 83 (4), 811–847.
Breiman, L., 1996a. Bagging predictors. Machine Learning 24 (2), 123–140.
Breiman, L., 1996b. Heuristics of instability and stabilization in model selection. The Annals of Statistics 24 (6), 2350–2383.
Breiman, L., 1996c. Stacked regressions. Machine Learning 24 (1), 49–64.
Breiman, L., 2001. Random forests. Machine Learning 45 (1), 5–32.
Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth & Brooks/Cole Advanced Books & Software.
Calle, M.L., Urrea, V., 2010. Letter to the editor: stability of random forest importance measures. Briefings in Bioinformatics 12 (1), 86–89.
Chen, X., Ishwaran, H., 2012. Random forests for genomic data analysis. Genomics 99 (6), 323–329.
Clemen, R.T., 1989. Combining forecasts: a review and annotated bibliography. International Journal of Forecasting 5 (4), 559–583.
Díaz-Uriarte, R., De Andres, S.A., 2006. Gene selection and classification of microarray data using random forest. BMC Bioinformatics 7 (1), 3.
Dieckmann, S., 2010. By force of nature: explaining the yield spread on catastrophe bonds. http://dx.doi.org/10.2139/ssrn.1082879.
Dietterich, T.G., 2000a. Ensemble methods in machine learning. In: International Workshop on Multiple Classifier Systems. Springer, pp. 1–15.
Dietterich, T.G., 2000b. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. Machine Learning 40 (2), 139–157.
Efron, B., 1992. Bootstrap methods: another look at the jackknife. In: Breakthroughs in Statistics. Springer, pp. 569–593.
Friedman, J., Hastie, T., Tibshirani, R., 2001. The Elements of Statistical Learning. Springer Series in Statistics. New York, NY, USA.
Galeotti, M., Gürtler, M., Winkelvos, C., 2013. Accuracy of premium calculation models for cat bonds—an empirical analysis. The Journal of Risk and Insurance 80 (2), 401–421.
Gatumel, M., Guegan, D., 2008. Towards an Understanding Approach of the Insurance Linked Securities Market, Documents de travail du centre d'economie de la sorbonne. Université Panthéon-Sorbonne (Paris 1), Centre d'Economie de la Sorbonne.
Götze, T., Gürtler, M., 2018. Sponsor-and trigger-specific determinants of cat bond premia: a summary. Zeitschrift für die gesamte Versicherungswissenschaft, 1–16.
Götze, T., Gürtler, M., Witowski, E., 2020. Improving cat bond pricing models via machine learning. Journal of Asset Management 21 (5), 428–446.
Grömping, U., 2009. Variable importance assessment in regression: linear regression versus random forest. American Statistician 63 (4), 308–319.
Gürtler, M., Hibbeln, M., Winkelvos, C., 2016. The impact of the financial crisis and natural catastrophes on cat bonds. The Journal of Risk and Insurance 83 (3), 579–612.
Hills, S., 2009. Catastrophe bond market emerges from crisis. https://www.reuters.com/article/us-catbonds-revival-analysis-idUSTRE59F1MK20091016. (Accessed 10 July 2020).
Ishwaran, H., 2007. Variable importance in binary regression trees and forests. Electronic Journal of Statistics 1, 519–537.
Ishwaran, H., Kogalur, U., 2019. Random Forests for Survival, Regression, and Classification (RF-SRC). R package version 2.8.0.
Ishwaran, H., Kogalur, U.B., Chen, X., Minn, A.J., 2011. Random survival forests for high-dimensional data. Statistical Analysis and Data Mining: The ASA Data Science Journal 4 (1), 115–132.
Ishwaran, H., Kogalur, U.B., Gorodeski, E.Z., Minn, A.J., Lauer, M.S., 2010. High-dimensional variable selection for survival data. Journal of the American Statistical Association 105 (489), 205–217.
Jaeger, L., Müller, S., Scherling, S., 2010. Insurance-linked securities: what drives their returns? The Journal of Alternative Investments 13 (2), 9–34.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning. Springer.
Jones, L., 2019. Shah: data-driven cat bond can be replicated. Trading Risk 119. 9–9.
Kuhn, M., 2008. Building predictive models in R using the caret package. Journal of Statistical Software 28 (5), 1–26.
Lane, M., Mahul, O., 2008. Catastrophe Risk Pricing: An Empirical Analysis. The World Bank.
Lane, M.N., 2000. Pricing risk transfer transactions 1. STIN Bulletin: The Journal of the IAA 30 (2), 259–293.
Lange, T., Roth, V., Braun, M.L., Buhmann, J.M., 2004. Stability-based validation of clustering solutions. Neural Computation 16 (6), 1299–1323.
Lei, D.T., Wang, J.-H., Tzeng, L.Y., 2008. Explaining the spread premiums on catastrophe bonds. In: NTU International Conference on Finance. Taiwan.
Liaw, A., Wiener, M., 2002. Classification and regression by randomforest. R News 2 (3), 18–22.
Lim, C., Yu, B., 2016. Estimation stability with cross-validation (escv). Journal of Computational and Graphical Statistics 25 (2), 464–492.
Major, J.A., 2019. Methodological considerations in the statistical modeling of catastrophe bond prices. Risk Management and Insurance Review 22 (1), 39–56.
Muir-Wood, R., 2017. The case of the trapped collateral. https://www.rms.com/blog/2017/11/16/the-case-of-the-trapped-collateral/. (Accessed 19 March 2019).
Ntoutsi, I., Kalousis, A., Theodoridis, Y., 2008. A general framework for estimating similarity of datasets and decision trees: exploring semantic similarity of decision trees. In: Proceedings of the 2008 SIAM International Conference on Data Mining. SIAM, pp. 810–821.
Oh, J., Laubach, M., Luczak, A., 2003. Estimating neuronal variable importance with random forest. In: 2003 IEEE 29th Annual Proceedings of Bioengineering Conference. IEEE, pp. 33–34.

Opitz, D., Maclin, R., 1999. Popular ensemble methods: an empirical study. Journal of Artificial Intelligence Research 11, 169–198.
Papachristou, D., 2011. Statistical analysis of the spreads of catastrophe bonds at the time of issue. ASTIN Bulletin: The Journal of the IAA 41 (1), 251–277.
Perrone, M.P., 1993. Improving regression estimation: Averaging methods for variance reduction with extensions to general convex measure optimization. PhD thesis. Physics Department, Brown University, Providence, RI.
Philipp, M., Rusch, T., Hornik, K., Strobl, C., 2018. Measuring the stability of results from supervised statistical learning. Journal of Computational and Graphical Statistics 27 (4), 685–700.
Probst, P., Bischl, B., Boulesteix, A.-L., 2018. Tunability: importance of hyperparameters of machine learning algorithms. arXiv preprint. arXiv:1802.09596.
Risk, T., 2019. Controlling a blazing risk. Trading Risk ILS Investor Guide, 14–15.
Russell, S.J., Norvig, P., 2016. Artificial Intelligence: A Modern Approach. Pearson Education, Limited.
Shmueli, G., 2010. To explain or to predict? Statistical Science 25 (3), 289–310.
Stodden, V., 2015. Reproducing statistical results. Annual Review of Statistics and Its Application 2, 1–19.
Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics 8 (1), 25.
Trottier, D.-A., Charest, A.-S., et al., 2018. Cat bond spreads via hara utility and nonparametric tests. The Journal of Fixed Income 28 (1), 75–99.
Turney, P., 1995. Bias and the quantification of stability. Machine Learning 20 (1–2), 23–33.
Wang, Z., Wang, Y., Zeng, R., Srinivasan, R.S., Ahrentzen, S., 2018. Random forest based hourly building energy prediction. Energy and Buildings 171, 11–25.
Wolpert, D.H., 1992. Stacked generalization. Neural Networks 5 (2), 241–259.
Yu, B., 2013. Stability. Bernoulli 19 (4), 1484–1500.
Zhou, Z.-H., 2012. Ensemble Methods: Foundations and Algorithms. Chapman and Hall/CRC.

3.2 Supplementing material

There has been a considerable amount of work behind the scene which is meaningful yet not directly visible in the published version of it. Consequently, we have decided to include a supplementing material section where we can share this information with the readers of this thesis. We will talk about issues relevant to the data preparation, choice of variables, development of a random forest model that can be fairly compared to benchmark models, and a comparative analysis of robustness results when using our data set versus some open source data sets from different disciplines.

3.2.1 Data collection and preparation

The collection and preparation of data is important in statistical research. To fulfil our study objectives, it was necessary to gather secondary data, however due to the over the counter nature of the financial products we analyse, and our desire to have the full population of these instruments for the chosen period of study, we had to seek for data from multiple sources. This information came in different formats, some were in pdf files, others in excel files requiring a lot of manual work to bring all information together in a single file. Some examples of such endeavours follow.

Starting from units, there were pieces of information referring to the same concept but measured in different units across different data providers. For example, some sources expressed expected loss as a percentage of issuance size whilst others in basis points terms. Since those measured on percentage terms were the majority, the appropriate transformation was made to change the unit from basis points[2] into percentage terms to maintain consistency within the same data columns.

In addition, there were categorical variables whose levels have not been following a rational flow in the original data set even though from discussion with people in the industry we understand that such categorisations reflected their needs at a business level. To give an example, in the context of the categorisation of the location-peril variable, it was not very reasonable to categorise some variables (NA Quake, NA Wind, SA Quake) at the peril-location level, others at the location level (APAC, Europe), and to neither consider peril nor location for multi peril bonds, which compose the majority of the sample. We do not doubt that such a categorisation serves some business need but from a statistical perspective such option was not as suitable. Thus, we made all categorical levels of this variable consistently reflect a location-peril combination.

[2]The equivalent of 1 basis point is 0.01 percent.

There were also instances where the effect that a wrong categorisation even in numerical variables would distort the prediction accuracy of the benchmark model. In particular, we have seen that there was a small number of observations corresponding to zero coupon deals which in the original data set have been recorded in the industry as if they carried zero spread. This was wrong to assume as the implied spread should have been derived from the implied coupon (which was available in the original data set) by subtracting the element of the money market rate. By modifying zero spread to an implied spread, the prediction performance of our benchmark linear regression when using only the variables describing catastrophe bond characteristics had increased from $R^2_{\text{OOB}} = 47\%$ and $R^2_{\text{10CV}} = 51\%$ to $R^2_{\text{OOB}} = 65\%$ and $R^2_{\text{10CV}} = 67\%$. Next, we discuss the choice of variables.

3.2.2 The choice of variables

At the beginning of our research design process, we thought to focus only on the variables that describe the characteristics of a catastrophe bond issuance without taking into consideration the market conditions at the time of the issuance. The reason for this was partially due to the fact that we wanted our model to enable a fairly accurate prediction accuracy with information that has been readily available in the offering circular of a new catastrophe bond.

Such a strategy did not significantly alter the performance of the random forest perhaps due to its flexibility to extract information by the rest of the predictors. In particular, by adding the Financial Lane Synthetic Rate variable into the study the R^2_{OOB} for random forest increased only by around 3%. However, the improvement in the linear regression results has been been approximately 30% and by incorporating the variable issuance year (iss_year) into the set of predictors, the linear regression prediction accuracy increased by further 2%. The aforementioned insights, as well as the fact that the macroeconomic variables have been included in previous studies which employed linear regression model, we decided to add them in the study.

Another decision that we had to make was how to represent the rating status of each catastrophe bond in our data set. We first tried out a relatively rough subdivision of our issuances into those that have not been rated at all, those that have been allocated an investment grade meaning having received a rating of 'BBB' or higher by Standard and Poor's or Moody's and those who did not qualify for investment grade. This categorisation was our first option because it has been used in the literature. Then, we tried out a finer specification using more granular rating categories but not individual rating notches and keeping of course the category for the non-rated deals. Finally, our analysis continued with the latter specification as it indicated a better prediction performance.

3.2.3 Developing a random forest model that can be comparable to the linear regression benchmark

In the context of the specification of random forest's hyperparameters, it may not be intuitive to the reader that P = 9 is the number of variables considered in the random forest, because in a traditional regression framework some of the categorical characteristics (diversifier, vendor) would enter the model as a series of dummy variables for each category, thus leading to a higher overall number of variables.

In regression models, categorical variables can only be considered through such category dummies. Therefore, one may think that dummy variables for each category of a categorical variable should also be included in the random forest model, to compare it to other models based on an identical set of variables, even though random forest is able to directly process categorical variable. For categorical variables (with more than two levels), one would perhaps agree that in the traditional linear regression that whether using dummy variables or not makes virtually no difference. However, for random forest, despite the fact that we have performed this robustness check, and as seen in Appendix D of the research paper, the prediction accuracy performance of the random forest with categorical dummy variables is not considerable for our catastrophe bond data set, there are a few reasons why we would refrain from following the dummy variables route.

Firstly, we believe that by considering categorical variables through category dummies would make each split in the random forest less flexible because it would limit the cut off point options for the categorical variable chosen to make a split. Maintaining plurality in the levels of categorical predictors provides more cut off point options and this can increase the diversity of each tree in the random forest which, as mentioned across Chapter 3, Section 3.1, is important. In particular, for expository purposes, suppose that we have a categorical variable with 4 levels, A, B, C, and D. Then, with the original implementation, if this particular variable is picked for splitting, then the tree would consider its options over 7 possibilities, i.e. {(A), (B, C, D)}, {(B), (A, C, D)}, {(C), (A, B, D)}, {(D), (A, B, C)}, {(A, B), (C, D)}, {(A, C), (B, D)}, and {(A, D), (B, C)}. On the other hand, using dummy variables, the number of available options for splitting would be effectively reduced to 4, i.e., {(A), (B,C,D)}, {(B), (A, C, D)}, {(C), (A, B, D)}, {(D), (A, B, C)}. This difference would be more dramatic for variables with more levels (such as the loc_peril variable with 12 levels in our catastrophe bond data set). In addition, with many levels, the resulting dummy variables would tend to be unbalanced (i.e. a large proportion of zeros, and a small proportion of ones), which would in turn make the trees less balanced.

Secondly, using dummy variables also make the evaluation of variable importance measures, such as those based on minimum depth, less straightforward. To give more details, for a particular categorical variable with K levels, (depending on the way in which the dummy variables are coded), one would now get K (or $K-1$) different average minimum depth scores $D_1, \ldots, D_K$, one for each dummy variable; but how to aggregate them into one final score that represents the importance of this categorical variable as a whole remains unclear, with many open options such as using the average, or taking the smallest of all, etc. Still, these options are not with their problems. For instance, if we were to use the average of $D_1, \ldots, D_K$, then the entire variable would tend to be assigned a lower importance rankings than what it really deserve even when only one or two levels are not too relevant (because the corresponding D_k tend to be very large). On the other hand, if we were to use the smallest of $D_1, \ldots, D_K$, then one might argue that variables with a larger K (i.e. number of levels) tend to have higher importance rankings than what they really deserve (because for these predictors, we are taking the minimum over a larger set).

Thirdly, to our best knowledge, most of the existing software implementing random forests (in R and Python) does not convert categorical variables into dummies by default, unless the number of levels is very large (because otherwise, going through all possible splitting options would be very time-consuming). As such, for most practitioners, if they were to apply random forest to their data, chances are that categorical variables will not be automatically converted into dummies by their software. Consequently, from a practitioner's viewpoint, we believe that it makes more sense for us to report results where categorical variables in random forest are handled in the default manner.

3.2.4 Assessing the stability of random forest model using as benchmark other random forest models developed with open source data sets

We appreciate that our data set is not publicly available and the reader of this thesis would like to have a benchmark when evaluating the stability of the results our proposed methods. A clear indication of absolute stability would require the establishment of a threshold. However, we recognise that, with our approach, it is perhaps more informative to look at these measurements in relative instead of absolute terms, i.e. by comparing stability of different methods on the same dataset, etc.

In particular, even through there is a temptation for the establishment of clear absolute cut-off point for characterising stability, we believe that may be potentially problematic or misleading because repeatability of the results in practice is often data (or research area) dependent. For instance, what is regarded as stable in financial applications (based on somewhat noisy data) might well be deemed as an unacceptable cut-off point in image processing or speech recognition (where noise level is lower).

In addition, the stability of any sensible methods in these measures would almost always improve in the simulations if we increase the number of observations from the same data generating process, since there would be less uncertainty associated with the data. Consequently, in our research study, the stability results for the random forest are discussed through the prism of the financial nature of our problem, allowing the reader to make their own judgement. Moreover, in the stability analysis of the predictors importance, stability is presented in relative terms - we examine two methods, permutation importance and minimal depth, and results are discussed on this comparative basis.

However as an extra robustness check, we have repeated the whole analysis using two other open source data sets as it is done in other machine learning applications. We have chosen to work with two data sets that were used by Breiman in his original random forest paper. As we observe in Table 3.2, across all 100 iterations, the recorded mean absolute difference of $\mathrm{R}^2_{\mathrm{OOB}}$ between Sample A and Sample B for the catastrophe bond data set is 2.19% with the minimum and maximum absolute differences being 0.01% and 6.92% respectively. We see that these results are in line with the random forest stability performance of the Boston Housing and Abalone data sets, which act as benchmarks for the stability evaluation, as explained in Section 3.3.2 of the published article.

Given that our problem sits in the intersection of financial and insurance market spheres where many behavioural aspects can affect prices, we consider the reported difference for the catastrophe bond data set being small. In essence, it is unlikely that an ILS fund would reject the use of the method solely for such a level of dissimilarity. In fact, the repeatability of prediction results here means that our initial random forest prediction accuracy result, i.e. of an $\mathrm{R}^2_{\mathrm{OOB}}$ of 96.57% is reliable, in the non-temporal context. Similarly, by looking Table 3.3, it is evident that the stability results of variables importance when using the minimal depth method and the catastrophe bond data set are more encouraging compared to when using the Boston using and Abalone data sets. Also, for the top of the variable importance ranks, it is still evident that minimal depth leads to more stable results than permutation importance for the two open source data sets just like it happens in the case of the catastrophe bond data set.

Data set	R^2_{OOB} Min Abs. Dif.	R^2_{OOB} Mean Abs. Dif.	R^2_{OOB} Max Abs. Dif.
Cat bond	0.01%	2.19%	6.92%
Boston Housing	0.14%	2.90%	9.07%
Abalone	0.04%	2.05%	6.04%

Table 3.2: Random forest stability of random forest using the cat bond data set versus these of Boston housing and Abalone data sets respectively.

Data set	Ranking position	Agreement % (Permutation)	Agreement % (Mimimal depth)
Cat bond	Top	98%	100%
	Second from top	36%	46%
	Second from bottom	27%	69%
	Bottom	22%	69%
	Last two	10%	100%
Boston Housing	Top	60%	72%
	Second from top	60%	72%
	Second from bottom	47%	17%
	Bottom	55%	18%
	Last two	44%	1%
Abalone	Top	70%	100%
	Second from top	39%	93%
	Second from bottom	20%	27%
	Bottom	19%	27%
	Last two	4%	0%

Table 3.3: Ranking stability by different importance ranking method for three different datasets: Cat bond, Boston housing, and Abalone.

Chapter 4

Model risk in reserve calculation

Risk taking is inherent to the business nature of all financial institutions, including banks, insurance companies, investment companies just to name a few. As a result, regulatory supervision is needed to ensure the integrity of the global financial system. The main requirement is that prior any financial position is taken, the risk associated with it needs to be quantified. Based on the latter, an equitable amount to the size of the risk has to be put aside as a reserve to ensure the soundness of the financial entity.

Nevertheless, unused capital has an opportunity cost, i.e. this of not realising new business opportunities. Consequently, it is natural for financial institutions to want to know what is the minimum amount of capital to add to a position to make it acceptable from regulatory viewpoint. In this chapter, we will see that this can be achieved by using a monetary measure of risk given its ability to be interpreted as a capital requirement. That said, there is variety of monetary risk measures that once could use. The goal of this chapter is to present some popular options and highlight that no matter which risk measure will be chosen there is an inherent model risk in their computation.

4.1 Monetary risk measures

A risk measure is a mapping from a set of random variables to the real numbers. This set reflects the uncertain terminal worth of a position at the end of a given trading period. That said, the net monetary outcome can be of any sign meaning that it can reflect both profit and loss. Moreover, risk measures need to satisfy certain properties namely monotonicity, translation invariance, and normalisation even though the latter property can be relaxed when it is convenient to do so. There are various risk measures and there is a need to formulate a set of properties which a good risk measure must fulfil. Later on, we explore how these measures are being structured. Examples of various monetary risk measures are also presented.

It is important to mention that whilst literature on the topic of risk measures is massive, there are two classical financial mathematics seminar papers which open the field namely Artzner et al. (1999) and Föllmer & Schied (2002). Since then, there has been a a plethora of great works focusing on the understanding of key properties of these risk measures, see for example Boyd & Vandenberghe (2004), Barrieu & El Karoui (2007), Henderson & Hobson (2009), Föllmer & Schied (2010), Rheinlander & Sexton (2011), Laeven & Stadje (2012), Delong (2013), and Föllmer & Schied (2016). That said, whilst the content of Chapter 4, Section 4.1 in this thesis follows strictly the aforementioned works, this literature list is certainly not exhaustive. Formal mathematical definitions, assumptions, and results are written in italics font for clarity purposes.

4.1.1 Definition and properties

An agent considers to take a financial position which intuitively carries an element of risk. Our aim is to quantify this risk by finding out the minimal amount which if added to the position, it would make it acceptable.

Definition (Financial position)**.** *We denote Ω as an non-empty set representing a fixed set of possible scenarios. A financial position taken by an agent is a mapping $X : \Omega \longrightarrow \mathbb{R} \cup \{+\infty\}$. Let ω be a scenario which is part of Ω. Then $X(\omega)$ reflects the terminal value of the position (profit or loss) at the end of the trading period if the scenario $\omega \in \Omega$ is observed.*

Let $\mathscr{X}$ be a set of financial positions and let the financial position X belong to it. Whilst from an economic perspective $\mathscr{X}$ would had to be of a very large size, preferably the space of all $X : \Omega \longrightarrow \mathbb{R} \cup \{+\infty\}$, it is mathematically convenient to introduce the restriction of boundedness. Furthermore, $\mathscr{X}$ is a linear space containing the constants. At this point we do not fix a probability measure in Ω. Below, we will find some number $\rho(X)$ which quantifies the risk of taking the financial position X.

Definition (Monetary Risk Measure)**.** *A monetary risk measure is a mapping $\rho : \mathscr{X} \longrightarrow \mathbb{R}$ which satisfies the conditions of monotonicity and cash/translation invariance for all X and $Y \in \mathscr{X}$ as follows.*

Monotonicity:

$$\textit{If } X \leq Y, \textit{ then } \rho(X) \geq \rho(Y)$$

Cash Invariance:

$$\rho(X+m) = \rho(X) - m, \textit{ for all } m \in \mathbb{R}$$

The condition of monotonicity simply reflects the fact that a position yielding a higher payoff in all scenarios, i.e. in the whole Ω, carries less risk. The cash invariance property, which can also be seen as translation invariance, demonstrates that risk is measured in monetary units meaning when an amount m is added to a risky position, its risk will decrease by the same amount m. Particularly, the cash invariance property indicates that:

$$\rho(X + \rho(X)) = \rho(X) - \rho(X) = 0$$

and

$$\rho(m) = \rho(0) - m = -m$$

for all m $\in \mathbb{R}$ and given that ρ is normalised, see Assumption straight below.

Thus, $\rho(X)$ accounts for the amount which if added to the position X, it would make the position acceptable from a regulatory perspective. For an agent to be protected against the risk of incurring a loss of size m, she would have to put aside as reserve the same amount m. Consequently, we see that the notions of risk measure and capital requirement are interchangeable.

Assumption. *A monetary measure of risk, generally, satisfies the condition of normalisation, as follows.*

Normalisation:

$$\rho(0) = 0$$

Normalisation implies that if one has nothing, there is no need to put aside any reserve. Given the above mentioned assumption, there is an equivalence between the properties of cash invariance and cash additivity in a way that:

$$\rho(X+m) = \rho(X) + \rho(m)$$

It should be noted though, that there are times when it is not convenient to make the assumption of a normalised monetary risk measure. Nevertheless, in most cases it is considered safe to do so.

Lemma. *Any monetary risk measure ρ is Lipschitz continuous with regards to the supremum norm $||.||$, as follows.*

$$\rho(X) - \rho(Y) \leq ||X - Y||$$

Definition (Convex risk measure)**.** *A convex risk measure is a monetary risk measure $\rho : \mathscr{X} \longrightarrow \mathbb{R}$ which fulfills the property of convexity, as follows.*

Convexity:

$$\rho(\lambda(X) + (1 - \lambda)Y) \leq \lambda\rho(X) + (1 - \lambda)\rho(Y), \textit{ for } 0 \leq \lambda \leq 1$$

Let's think about an agent who wants to make investment but she naturally comes across with the problem of scarce resources. The agent needs to consider the set of probable future outcomes that can be produced given the limited resources available. Let A and B be two investment strategies which lead to X and Y respectively. If the agent decides to build a diversified portfolio, she will invest only a fraction λ of the resources on the first possibility and spend whatever is left for the second option. Then the portfolio will look like λX+$(1 - \lambda)Y$. Coming back now to the property of convexity, it implies that diversification has a risk diminishing effect as it states that the risk of the diversified portfolio λX+$(1 - \lambda)Y$ is less or equal to the weighted average of the individual risks, see Föllmer & Schied (2002). The latter is even more evident if one takes into account that for a monetary measure of risk there is equivalence between the requirements of convexity and quasi-convexity even though the latter one is weaker, see Cerreia-Vioglio et al. (2011) for more on this topic. In particular, the quasi-convexity is given by the following equation.

$$\rho(\lambda X + (1 - \lambda)Y) \leq \max(\rho(X), \rho(Y)), \text{ for } 0 \leq \lambda \leq 1$$

The interpretation here is that the risk associated with the diversified portfolio is not higher than the maximum risk of both positions signifying that diversification should not increase risk.

Definition (Coherent risk measure)**.** *A coherent risk measure ρ is a convex risk measure which satisfies the property of positive homogeneity.*

Positive homogeneity:

$$\rho(\lambda X) = \lambda\rho(X), \textit{ for } \lambda \geq 0$$

The latter property reflects the fact that the risk associated with a liquid position shall be proportionate to its size implying that the absence of liquidity risk in the market. It is worth mentioning that when positive homogeneity axiom holds for a monetary risk measure ρ, then it is also normalized, meaning that $\rho(0) = 0$. Given that positive homogeneity applies, there is equivalence between the properties of convexity and sub-additivity which is described straight below. The difference lies in the fact that in the convexity property risks have weights.

$$\rho(X + Y) \leq \rho(X) + \rho(Y)$$

The aforementioned axiom is particularly useful as it provides scope for risk decentralization. If every department in a financial institution has been provided with different risk limits, then the sub-additivity property ensures that the risk of the aggregate position has bounds; this of the sum of the individual risk limits. Nevertheless, there is a pitfall here as by relying on this property, we acknowledge that the risk rises linearly as the size of the position grows. This is a big assumption which does not always reflect the reality, so it would be convenient to focus on convex risk measures instead.

4.1.2 Acceptance sets and their relation to monetary risk measures

We introduce the concept of acceptance sets which describe the requirements that a financial position needs to meet for it to pass the demands of the regulator, see for instance Föllmer & Schied (2016) and Föllmer & Schied (2010).

Definition (Acceptance set)**.** *A monetary risk measure ρ gives rise to the class of positions which do not require additional capital and are thus acceptable.*

$$\mathscr{A}_\rho := \{X \in \mathscr{X} | \rho(X) \leq 0\}$$

The class $\mathscr{A}_\rho$ will be named as the acceptance set of ρ.

Below we will present the relation between a monetary risk measure and its acceptance set. In this context, one can easily prove:

a) ρ is a convex risk measure if and only if $\mathscr{A}_\rho$ is a convex set.

b) ρ satisfies the property of positive homogeneity if and only if $\mathscr{A}_\rho$ is a cone. Specifically, ρ is coherent if and only if $\mathscr{A}_\rho$ is a convex cone.

c) the acceptance set $\mathscr{A}_\rho$ completely determines ρ, as:

$$\rho(X) := \inf\{m \in \mathbb{R} | m + X \in \mathscr{A}_\rho\} \tag{4.1.2.1}$$

Now, let ρ be a monetary risk measure with an acceptance set $\mathscr{A}$. For $\mathscr{A} := \mathscr{A}_\rho$ to hold the following properties must be satisfied:

1) $\mathscr{A} \cap \mathbb{R} \neq \emptyset$ meaning that there exists a monetary amount m such that having m is for sure acceptable.

2) $\inf\{m \in \mathbb{R} | m \in \mathscr{A}\} > -\infty$

3) $X \in \mathscr{A}$, $Y \in \mathscr{X}$, $Y \geq X$ then $Y \in \mathscr{A}$, implying that if one position is accepted then all better positions will also be accepted.

Conversely, given a subset $\mathscr{A}$ of $\mathscr{X}$ (acceptance set) and a position $X \in \mathscr{X}$, the risk of the position X can be defined as:

$$\rho_{\mathscr{A}}(X) := \inf\{m \in \mathbb{R} | m + X \in \mathscr{A}\} \tag{4.1.2.2}$$

If the properties 1-3 hold, then $\rho_{\mathscr{A}}$ is a monetary risk measure. Then:

a) If $\mathscr{A}$ is convex, then $\rho_{\mathscr{A}}$ is also convex.

b) If $\mathscr{A}$ is a cone, then $\rho_{\mathscr{A}}$ satisfies positive homogeneity.

c) $\mathscr{A}_{\rho\mathscr{A}}$ is equal to the closure of $\mathscr{A}$ with respect to the supremum norm $||.||$. In particular, $\mathscr{A}_{\rho\mathscr{A}} = \mathscr{A}$ holds if and only if $\mathscr{A}$ is $||.||$-closed.

d) Given Equation 4.1.2.2, we see that Equation 4.1.2.1 appears as $\rho_{\mathscr{A}}(X)$=ρ.

4.1.3 Dual Representation

A commonplace in mathematical optimisation theory is the so called duality principle according to which an optimisation problem can be seen from two different viewpoints, i.e, the primal (minimisation) or dual (maximisation) problem, see Laeven & Stadje (2012), Barrieu & El Karoui (2007), Föllmer & Schied (2016), and Föllmer & Schied (2010). Finding solution for the dual problem means that a lower bound to the solution of the primal problem is identified. Nevertheless, equality between the optimal values of of the primal and dual problems is not necessary. In particular, the notation duality gap is exactly used to describe the difference between these two optimal values. However, when considering convex optimisation problems something interesting happens; the duality gap equals to zero under certain conditions.

Coming back to the previous section, it is understandable that the quantification of $\rho(X)$ signifies the need to solve an optimisation problem. Equation 4.1.2.1 reflects the primal approach interpreting the risk measure as a capital requirement. However, under this representation, finding the value of $\rho(X)$ is a quite complicated task. Also, it would be interesting to see whether we can derive a different interpretation of $\rho(X)$ other than the one mentioned above. Indeed, we will see that the dual approach of our optimisation problem is a robust analogue equation to Equation 4.1.2.1 which is easier to calculate plus it shows that a risk measure is also the lens through which model uncertainty can be seen.

Result (Dual representation of convex risk measures)**.** *Let's assume that $\mathscr{X}$ is comprised of measurable functions on $(\Omega, \mathscr{F})$. Then the dual form of a convex risk measure ρ is the following.*

$$\rho(X) = \sup_{\mathbb{Q} \in \mathscr{M}} (\mathbb{E}_{\mathbb{Q}}[-X] - a(Q))$$

Where $\mathscr{M}$ is a set of probability measures on $(\Omega, \mathscr{F})$ and $\mathbb{Q}$ is a subset of probability measures in $\mathscr{M}$. $\mathscr{M}$ is such that $\mathbb{E}_{\mathbb{Q}}[X]$ is well defined for all $\mathbb{Q} \in \mathscr{M}$ and $X \in \mathscr{X}$. The functional $a : \mathscr{M} \longrightarrow \mathbb{R} \cup \{+\infty\}$ is named penalty function.

The probability measures contained in $\mathscr{M}$ can be seen as probabilistic models whose reliability depends on how big or small the penalty $a(Q)$ is. Therefore, the value of $\rho(X)$ is calculated as the worst case expectation taken over all models $Q \in \mathscr{M}$ and then penalized by $a(Q)$.

The main goal in the dual representation theory of convex risk measures is to derive a representation in a systematic way by applying convex duality. In this context, we admit the following.

Result (Minimal penalty function). *For every $Q \in \mathscr{M}$ c, the minimal penalty function ρ is defined by:*

$$a_\rho(Q) := \sup_{X \in \mathscr{X}} \left(\mathbb{E}_Q[-X] - \rho(X)\right) = \sup_{X \in \mathscr{A}_\rho} \left(E_Q[-X]\right)$$

By adding more assumptions on the structure of X and as well as on the continuity properties of ρ, it is feasible for one to represent $\rho(X)$ through Fenchel-Legendre duality as follows:

$$\rho(X) = \sup_{Q \in \mathscr{M}} \left(\mathbb{E}_Q[-X] - a_\rho(Q)\right) \tag{4.1.3.1}$$

In the special case, where a_ρ takes only the values 0 and $+\infty$, ρ is a coherent measure of risk which can be represented as follows:

$$\rho(X) = \sup_{Q \in Q_\rho} \left(\mathbb{E}_Q[-X]\right)$$

where Q_ρ contains all $Q \in \mathscr{M}$ for which $a_\rho(Q) = 0$.

Coming back to Equation 4.1.3.1, we explain the circumstances under which such a representation can be derived. In this setting, it is requisite to extend the set $\mathscr{M}$ to include finitely additive set functions and discuss about risk measures on $L_\infty(\mathbb{P})$.

4.1.4 Risk Measures on $L_\infty(\mathbb{P})$

When a probability measure $\mathbb{P}$ is fixed, we can naturally define risk measures ρ on $L_\infty(\mathbb{P})$ and not on $\mathscr{X}$, see for example Föllmer & Schied (2016), given that the following compatibility condition is satisfied:

$$\text{If } X = Y, \text{ then } \rho(X) = \rho(Y) \quad \mathbb{P}\text{-almost surely.}$$

At this point, we introduce two new notations. In particular, let $M_{1,ac}(\mathbb{P})$ be the set of all finitely additive measures absolutely continuous with respect to $\mathbb{P}$, and $\mathscr{M}_{1,ac}(\mathbb{P})$ be the set of probability measures absolutely continuous with respect to $\mathbb{P}$. We also need to provide the definition of the natural extension of continuity from below and above in the space $L_\infty(\mathbb{P})$ as follows:

Definition (Continuity from below in the space $L_\infty(\mathbb{P})$).

$$X_n \searrow X \quad \mathbb{P}\text{-almost surely} \Longrightarrow \rho(X_n) \nearrow \rho(X)$$

Definition (Continuity from above, Fatou property, in the space $L_\infty(\mathbb{P})$).

$$Y_n \nearrow Y \quad \mathbb{P}\text{-almost surely} \Longrightarrow \rho(Y_n) \searrow \rho(Y)$$

Theorem. *Let $\mathbb{P}$ be a probability measure.*
1. Any convex risk measure ρ on X satisfying the 5.01 may be considered as a risk measure on $\mathbb{L}_\infty(\mathbb{P})$. A dual representation holds true in terms of absolutely continuous additive measures $Q \in M_{1,ac}(\mathbb{P})$.
2. ρ admits a dual representation on $\mathscr{M}_{1,ac}(\mathbb{P})$:

$$\alpha(\mathbb{Q}) = \sup_{X \in L_\infty(\mathbb{P})} \{\mathbb{E}_{\mathbb{Q}}[-X] - \rho(X)\}$$

,

$$\rho(X) = \sup_{\mathbb{Q} \in \mathscr{M}_{1,ac}(\mathbb{P})} \{\mathbb{E}_{\mathbb{Q}}[-X] - \alpha(\mathbb{Q})\}$$

if and only if one of the following equivalent properties is true:
a) ρ is continuous from above (Fatou property);
b) ρ is closed for the weak topology $\sigma(L_\infty, L_1)$;
c) the acceptance set $\{\rho \leq 0$ is a weak-closed in $L_\infty(P)$.
3. Assume that ρ is a coherent (homogeneous) risk measure, satisfying the Fatou property. Then:

$$\rho(X) = \sup_{\mathbb{Q} \in \mathscr{M}_{1,ac}(\mathbb{P})} \{\mathbb{E}_{\mathbb{Q}}[-X] | \alpha(\mathbb{Q}) = 0\} \tag{4.1.4.1}$$

The supremum in Equation 4.1.4.1 is a maximum if and only if one of the following equivalent properties hold:
a) ρ is continuous from below.
b) The convex set $\mathscr{Q} = \{\mathbb{Q} \in \mathscr{M}_{1,ac} | \alpha(\mathbb{Q}) = 0\}$ is weakly compact in $L^1(\mathbb{P})$.

With regards to the latter point, according to the Dunfford-Pettis theorem, the weakly relatively compact sets of $L^1(\mathbb{P})$ are sets of uniformly integrable variables and La Vallée -Poussin gives a criterion to check this property: the subset of $\mathscr{A}$ of $L^1(\mathbb{P})$ is weakly relatively compact iff it is closed and uniformly integrable in the sense that there exists an increasing convex continuous function $\Psi : \mathbb{R}_+ \longrightarrow \mathbb{R}$ (the so called Young's function), such that:

$$\lim_{x \longrightarrow \infty} \frac{\Psi(x)}{x} = +\infty$$

and

$$\sup_{\mathbb{Q} \in \mathscr{A}} \mathbb{E}_{\mathbb{P}} \left[\Psi\Big(\frac{d\mathbb{Q}}{d\mathbb{P}}\Big) \right] < +\infty$$

Having made a formal presentation of the concept of monetary risk measures, we proceed by providing few monetary risk measures examples.

4.2 Some examples

There is a variety of risk measures, see for instance Barrieu & El Karoui (2005), Föllmer & Schied (2010), and Föllmer & Schied (2016). Here, we present some examples of monetary measures of risk, based on the works of Barrieu & El Karoui (2007), Boyd & Vandenberghe (2004), Laeven & Stadje (2012), Delong (2013), Föllmer & Schied (2016), Föllmer & Schied (2010), Henderson & Hobson (2009), and Rheinlander & Sexton (2011), which have quantiles as a key ingredient namely Value at Risk, Average Value at Risk, and also a non-quantile based monetary risk measure called Entropic Risk Measure. The choice of these instances is based on the popularity or the potential value that they can bring in the financial industry practice given their properties. Moreover, some reviews regarding potential drawbacks for some of these methods are also mentioned. Formal mathematical definitions, and results are written in italics font for clarity.

4.2.1 Quantile-based risk measures

Since, we provide examples of quantile-based risk measures, it makes sense to start with an introduction on quantiles. In particular, one of the most common practices in the topic of measuring the risk of a financial position X is to determine a quantile of the distribution of X under a given probability measure P whose definition follows.

Definition (Quantile)**.** *Let $\lambda \in (0,1)$. Then a $\lambda -$ quantile of a random variable on $(\Omega, \mathscr{F}, P)$ is any real number q which satisfies the following property.*

$$P[X \leq q] \geq \lambda$$

and

$$P[X < q] \leq \lambda$$

Now, the set of all $\lambda -$ *quantiles* of X is an interval $[q_X^-(\lambda), q_X^+(\lambda)]$ where the lower quantile function of X is:

$$q_X^-(t) = \sup\{x|P[X < x] < t\} = \inf\{x|P[X \leq x] \geq t\}$$

and the upper quantile function of X is:

$$q_X^+(t) = \inf\{x|P[X \leq x] > t\} = \sup\{x|P[X < x] \leq t\}$$

Having explained the notion of quantile and the bounds of the set of all quantile functions of X, we next introduce Value at Risk and Average Value at Risk.

4.2.1.1 Value at risk

The financial and insurance industry uses extensively a monetary risk measure called Value at risk (V@R) whose mathematical definition is provided below.

Definition (Value at risk)**.** *Fix some level $\lambda \in (0,1)$. The Value at Risk of a financial position X at level λ is defined as follows.*

$$V@R_\lambda(X) := -q_X^+(\lambda) = \inf\{m|P[X+m<0] \leq \lambda\}$$

The financial interpretation of $V@R_\lambda(X)$ is the smallest amount of capital which, if added to the position X and invested in a risk free manner, ensures that the probability of a negative outcome is below the level λ. Thus, we see that by using V@R one can maintain control of the probability of a loss. Nevertheless, V@R does not provide any information about the size of loss in case it occurs meaning that if an agent is looking only at this measure there is no way to know the maximum possible loss that can incur.

It is important to mention that $V@R_\lambda(X)$ is coherent measure of risk as it satisfies the property of positive homogeneity. However, the acceptance set of $V@R_\lambda(X)$ does not fulfil the convexity property and consequently $V@R_\lambda(X)$ is not a convex measure of risk. As we have mentioned in Chapter 4, Section 4.1.1, convexity reflects the diversification effect. Thus, this second limitation of V@R not being convex implies that this measure may penalise diversification instead of encouraging it. Indeed, V@R may actually create an incentive to concentrate risk on a scenario of small probability.

4.2.1.2 Average value at risk

As a result of the drawback of V@R to determine the size of losses, a new risk measure was developed to account for this, i.e., the Average Value at Risk (AV@R). The idea behind it is to calculate the V@R at few levels of the distribution and then take the average. A formal definition of AV@R follows.

Definition (Average value at risk)**.** *The Average Value at Risk at level $\lambda \in (0,1]$ of a position $X \in \mathscr{X}$ is given by*

$$AV@R_\lambda(X) = \frac{1}{\lambda}\int_0^\lambda V@R_\gamma(X)d\gamma$$

For $X \in L^\infty$, we have

$$\lim_{\lambda \downarrow 0} V@R_\lambda(X) = -\operatorname{ess\,inf} X = \inf\{m | P[X + m < 0] \leq 0\}$$

Therefore, we can define

$$AV@R_0(X) = V@R_0(X) := -\operatorname{ess\,inf} X$$

representing the most conservative (worst case) risk measure in L^∞ which is continuous from above but in general not from below, and

$$AV@R_1(X) = -\int_0^1 q_X^+(t)dt = -\mathbb{E}[X]$$

It is worth mentioning that despite the aforementioned correction of the original V@R, the Average Value at Risk is still not convex and the problem of penalising diversification is not solved. Consequently, we thought that it would be interesting present a risk measure example, which happens to be non-quantile based this time, which satisfies the convexity property. In this context, next we introduce the entropic risk measure.

4.2.2 Entropic risk measure

Another alternative to V@R and AV@R is the entropic risk measure. We could say that it enjoys the highest popularity among convex risk measures on L^∞ and it is a prime example of a convex risk measure which is not coherent, i.e. the property of positive homogeneity is not satisfied. We will define this measure as e_γ functional which is continuous from below. Entropic risk measure derives its name from the notion of relative entropy which appears in its dual representation. That said, we start by defining the relative entropy and we continue with presenting results with respect to the dual and primary formulation of the entropic risk measure.

Definition (Relative entropy)**.** *The relative entropy of $H(\mathbb{Q}, \mathbb{P})$ of a probability measure $\mathbb{Q}$ with respect to a probability measure $\mathbb{P}$ is given as:*

$$H(\mathbb{Q}, \mathbb{P}) = \begin{cases} \mathbb{E}[\frac{d\mathbb{Q}}{d\mathbb{P}} \log \frac{d\mathbb{Q}}{d\mathbb{P}}], \textit{if} \quad \mathbb{Q} \ll \mathbb{P} \\ +\infty, \textit{otherwise} \end{cases}$$

Result (Dual formulation of entropic risk measure)**.** *We now regard a penalty function $\alpha : \mathscr{M}(\mathbb{P}) \longrightarrow (0, +\infty]$ which is defined by:*

$$\alpha(Q) := \frac{1}{\gamma} H(\mathbb{Q}, \mathbb{P})$$

Here γ *is a given constant and* $H(\mathbb{Q}|\mathbb{P}) = \mathbb{E}_Q[\log \frac{dQ}{dP}]$ *is the relative entropy of* $\mathbb{Q} \in \mathscr{M}(\mathbb{P})$ *with respect to* $\mathbb{P}$.

Then, the dual representation of the corresponding entropic risk measure $e_\gamma(X)$ *is given by*

$$e_\gamma(X) = \sup_{\mathbb{Q} \in \mathscr{M}(\mathbb{P})} \Big(\mathbb{E}_{\mathbb{Q}}[-X] - \frac{1}{\gamma} H(\mathbb{Q}|\mathbb{P}) \Big)$$

It can be shown that α is in reality the minimal penalty function representing $e_\gamma(X)$, as follows.

$$\alpha^{min}(Q) = sup_{X \in L^{+\infty}} \Big(\mathbb{E}_{\mathbb{Q}}[-X] - \frac{1}{\gamma} H(\mathbb{Q}|\mathbb{P}) \Big) = \frac{1}{\gamma} H(\mathbb{Q}|\mathbb{P})$$

Result (Primal Formulation of entropic risk measure)**.** *Above, we presented the dual form of the entropic risk measure. One, though, can see the optimization problem from the primal point of view which is described below.*

$$e_\gamma(X) = \gamma \ln \mathbb{E}_{\mathbb{P}} \Big[exp\Big(-\frac{1}{\gamma} x \Big) \Big]$$

Let's now consider the Theorem 1. Since e_γ is continuous from below in $\mathbb{L}_\infty$, then $\forall X \in \mathbb{L}_\infty(\mathbb{P})$ the following holds true.

$$e_\gamma(X) = \gamma \ln \mathbb{E}_{\mathbb{P}} \Big[exp\Big(-\frac{1}{\gamma} x \Big) \Big] = \max_{\mathbb{Q} \in \mathscr{M}_{1,ac}} \{ \mathbb{E}_{\mathbb{Q}}[-X] - \gamma H(\mathbb{Q}|\mathbb{P}) \}$$

.

This is very important as it indicates that we can find an explicit solution to our optimisation problem. Moreover, the entropic risk measure is closely related to the exponential utility function and to the associated indifference price, more details on both of these two aspects can be found in Chapter 4, Appendix.

As we have done already for V@R and AV@R in Chapter 4, Section 4.2.1.1 and 4.2.1.2 respectively, it is interesting to see how we can interpret the relative entropy from a financial viewpoint. In particular, the quantification of risk with the relative entropy has a somehow natural interpretation in a financial setting. In particular, a financial agent has a reference model $\mathbb{Q}$. Obviously, the measure Q is not the true measure but a mere approximation to the probabilistic model of the payoff X. As a result, an issue of "trust" arises.

In particular, the agent disputes the model $\mathbb{Q}$ and takes into consideration many models $\mathbb{P}$ whose worth of being accepted as true or reasonable decreases proportionally to their distance from the reference model $\mathbb{Q}$. It is apparent that for every given X, the mapping $\gamma \longrightarrow e_\gamma(X)$ is increasing. As a result, the parameter γ may be seen as a tool which quantifies the degree of trust the agent puts in the reference measure $\mathbb{Q}$. If $\gamma = 0$ then $e_0(X) = -\operatorname{ess\,inf} X$ which corresponds to maximal level of mistrust. In the latter case only the zero sets of the measure $\mathbb{Q}$ are considered trustworthy. On the other hand, if $\gamma = 0$ then the $e_\infty(X) = -\mathbb{E}[X]$ showing a maximal level of reliability in the measure P.

Having explained what risk measures are and shown some alternative risk measures that can be used in effort to compute a financial institution's reserve, we next focus on some model risk considerations.

4.3 Model risk in reserve computation

Until now we have seen that there are various risk measures although some of them lack important properties which can in turn lead to poor risk management strategies eventually. For example, Value at Risk is a risk measure that the industry uses massively as it is implied by regulators. Nevertheless, it is well known that it can lead to the accumulation of shortfall risk in some scenarios, see Embrechts & Wuttrich (2022).

Here, we ignore what makes a financial institution choose one risk measure over another, for instance AV@R instead of V@R, and we focus on uncertainties regarding the computation of any risk measure. In particular, this is important given that the regulatory rules for financial institutions, such as these of Basel II and Solvency II directives, grant total flexibility in the choice of internal models for the computation of a monetary risk measure. Similarly, there are no specific requirements regarding the assumptions deployed in the stochastic modelling process apart from the fact that resulting reserve shall respond to a certain quantile of the aggregated loss data over the period of a year among some other data related details. Moreover, there are not explicit rules with regards to how the correlation between different risks that financial institutions take should be captured, except from a general guideline that any assumptions need to be justifiable, see Embrechts et al. (2013).

Consequently, post choosing a monetary risk measure to work with, and no matter how the decision to work with a particular risk measure over another is made, an actuary faces the hurdle of deciding which is an appropriate model to use for its computation, and straight after this which are the suitable parameters for the chosen model. In other words the actuary faces model risk, Alexander & Sarabia (2012). In particular, model choice is inherent in the decision making of an actuary when calculating the reserve as assumptions need to be made for the random variable of interest, the data generative mechanism, and for the statistical model itself as a stochastic model inevitable involves an element of randomness in its output, see Cairns (2000). With respect to the model parameters, they are subject to estimation but given the fact that there is only a finite amount of data available at any given point in time, one can never be rest assured that the parameters estimation is truly reflective of the underlying phenomenon of interest. Also, parameter estimation can be achieved by using a variety of numerical methods each of which could yield dissimilar results for a given set of data and model.

It is worth mentioning that the aforementioned decomposition of model risk into model choice and parameter ambiguity is not ultimate as there is no clear agreement in the literature regarding the origins of model risk. We direct the interested reader to Cont (2006), Kerkhof et al. (2010), and Cairns (2000) on some views on this matter. Also, some selective, and certainly not exhaustive, literature aiming to address model risk relevant to the the computation of capital reserves include the works of Kerkhof et al. (2010), Boucher et al. (2014), Barrieu & Scandolo (2015), and Bertram et al. (2015).

Finally, in the subsequent Chapter 5, we see model risk for the computation of the reserve through the lenses of an opinion combination problem.

4.4 Appendix; Entropic risk measure and indifference price

The entropic risk measure is closely related to the exponential utility function and to the associated indifference price, where the latter refers to the price at which an agent is indifferent between entering a financial transaction or not based on an expected utility level argument. We deem interesting to talk about this topic, as literature, see Barrieu & El Karoui (2005), suggests that this mathematical model can be used for the pricing of financial products in incomplete markets. That said, in Chapter 3, we have seen an example of such a market for catastrophe bonds.

4.4.1 Indifference pricing

Indifference pricing is an alternative approach to hedge a contingent claim. Under this framework, the agent is able to maximise her expected utility of wealth and lower the risk associated with the payoff uncertainty even in an incomplete market setting. At the same time, when the market is complete this method advances to the standard risk neutral hedge and as a result it is a powerful pricing tool.

In any case, the agent's risk preferences with regards to risk exposure are quantified through the use of utility functions. The choice of a representative utility function has even greater importance when an incomplete market is present. Ultimate goal is to find an amount which confirms that the agent's utility is unchanged by purchasing/selling the contingent claim. This price is called utility indifference or reservation price.

In practice, assuming that the agent has a utility function U, we can say that from the buyer's viewpoint, the reservation price shows the maximum amount π that the agent is willing to pay for a claim X, thus:

$$\mathbb{E}[U(X-\pi)] = U(0)$$

Similarly, from a seller's aspect this price reflects the minimum amount π the agent is ready to accept in order to sell claim c.

$$\mathbb{E}[U(\pi-X)] = u(0)$$

Thus, depending from which side we are looking into this (”buy” or ”sell”) we can determine an upper or lower bound respectively so that the transaction will happen. Consequently, the utility indifference price can just provide us with a decision rule and it is not the actual transaction price.

The indifference price satisfies certain properties:

i) it is a monotonically increasing function of the claim X such that

$$\pi(X) \leq \pi(Y), \text{ if } X \leq Y.$$

ii) it is also a convex function of the claim X, thus allowing one to capture the effect of diversification

$$\pi(\lambda X + (1-\lambda)Y) \leq \lambda\pi(X) + (1-\lambda)\pi(Y), \forall\lambda \in [0,1].$$

iii) it is cash translation invariant meaning that if the payoff of the claim X can be translated by a riskless amount m then the price π can be also translated by the same amount

$$\pi(X+m) = \pi(X) + m, \forall m \in \mathbb{R}$$

.

Next, we discuss the way in which the exponential utility function is linked to the entropic risk measure presented in Chapter 4, Section 4.2.2.

4.4.2 Relation between exponential utility and entropic risk measure

We start describing the relation between the exponential utility function, and entropic risk measure by considering an agent who operates in an uncertain universe which is modelled by a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let T be the time horizon of interest. It is assumed that the agent's risk preferences are best described by an exponential utility function for the computational convenience it offers and takes the form:

$$U(X) = -\gamma \exp(-\frac{1}{\gamma}X)$$

where γ is the risk tolerance coefficient reflecting how tolerant is the agent to the part of risk that cannot be hedged. As a result, an agent with a risk averse profile would require a higher price for a non-perfectly replicated claim as remuneration for accepting this residual risk. When this risk aversion goes to the limit, the agent leans towards demanding the super-replication price for the contingent claim. The agent has wealth W reflecting the state of their portfolio at time T. At this point, we assume that the wealth W is random in our model. The random variables X and W are bounded and given the current low interest rate environment, we can safely ignore interest rate fluctuations.

For the agent to decide whether or not to buy a contingent claim[1] with a payoff X at time T. Then, it is enough to find the pricing rule which makes the agent indifferent between keeping his wealth stable and entering the transaction. As it was described

[1]That is a financial instrument whose payout depends on the occurrence of a random future event.

earlier, there are two ways to capture this concept in an equation depending on whether our agent is from the "buy" or "sell" side. Taking as given that the agent is a buyer, the problem of deciding whether to take the risk or not will be reduced in finding the maximal price for which she is ready to pay for it. The indifference price $\pi(X)$ is given by the constraint:

$$\mathbb{E}_{\mathbb{P}}[U(W + X - \pi(X))] = \mathbb{E}_{\mathbb{P}}[U(W)]$$

Then:

$$\mathbb{E}_{\mathbb{P}}[\exp((-\frac{1}{\gamma})(W + X - \pi(X))] = \mathbb{E}_{\mathbb{P}}[\exp(-\frac{1}{\gamma})W]$$

$$\Leftrightarrow \pi(X|W) = e_{\gamma}(W) - e_{\gamma}(W + X)$$

where e_{γ} is the opposite of the guaranteed return that someone would accept rather than taking a chance on a higher, but uncertain, return i.e the opposite of certainty equivalent defined for any bounded random variable Ψ as $e_{\gamma}(\Psi) \triangleq \gamma \ln \mathbb{E}_{\mathbb{P}}[\exp(-\frac{1}{\gamma}\Psi)]$.

Based on the above, it is apparent that $\pi(X|W)$ satisfies the properties of increasing monotonicity, convexity, and translation invariance $\pi(X+m|W) = \pi(X|W)+m$ whilst the functional $e_{\gamma}(X) = -\pi(X|W = 0)$ is also monotonous decreasing, convex and translation invariant $e_{\gamma}(\Psi + m) = e_{\gamma}(\Psi) - m$, and thus it has similar properties.

An important remark on $\pi(X)$ is that it is dependent on the agent's initial exposure. Consequently,if $\pi^s(X|W)$ is the indifference seller's price and $\pi^b(X|W)$ is the indifference buyer's price it holds that there is a strong relationship between the seller's and buyer's reservation price rules in a way that the one is the opposite of the other, given that the presented framework is symmetric.

$$\pi^s(X|W) = -\pi^b(-X|W)$$

As we can see the functional e_{γ} is nothing else than the entropic risk measure described in the previous section. So having started from a utility maximisation problem, we concluded with an equivalent risk measure minimisation problem in order to price this contingent claim X. The entropic functional is the most appropriate to work with given the desired properties that satisfies as a convex measure of risk and not the one of utility as such, allowing to focus only to the notion of price. Consequently, we see that the entropic risk measure cannot be used only for quantifying reserves for meeting the capital requirement set by a regulator but it also constitutes a very powerful pricing tool.

Chapter 5

A finite mixture modelling perspective for combining experts' opinions with an application to quantile-based risk measures

This chapter is mostly dedicated to our article published on an open access basis at journal Risks entitled "A finite mixture modelling perspective for combining experts' opinions with an application to quantile-based risk measures"[1], see Makariou, Barrieu & Tzougas (2021). The model risk associated with the computation of the reserve is described as a traditional opinion combination problem which can be effectively captured by using a finite mixture modelling approach to allow for both sources of model risk described in Section 4.3.1 to be accounted. Here, the article is presented in the exact format in which it has been published followed by some supplementary material.

5.1 A finite mixture modelling perspective for combining experts' opinions with an application to quantile-based risk measures

[1]The article can be accessed via the following link: https://doi.org/10.3390/risks9060115

risks

MDPI

Article

A Finite Mixture Modelling Perspective for Combining Experts' Opinions with an Application to Quantile-Based Risk Measures

Despoina Makariou *, Pauline Barrieu * and George Tzougas *

Department of Statistics, London School of Economics and Political Science, London WC2A 2AE, UK
* Correspondence: d.makariou@lse.ac.uk (D.M.); P.M.barrieu@lse.ac.uk (P.B.); g.tzougas@lse.ac.uk (G.T.)

Abstract: The key purpose of this paper is to present an alternative viewpoint for combining expert opinions based on finite mixture models. Moreover, we consider that the components of the mixture are not necessarily assumed to be from the same parametric family. This approach can enable the agent to make informed decisions about the uncertain quantity of interest in a flexible manner that accounts for multiple sources of heterogeneity involved in the opinions expressed by the experts in terms of the parametric family, the parameters of each component density, and also the mixing weights. Finally, the proposed models are employed for numerically computing quantile-based risk measures in a collective decision-making context.

Keywords: opinion pooling; finite mixture models; expectation maximization algorithm; quantile-based risk measures

Citation: Makariou, Despoina, Pauline Barrieu, and George Tzougas. 2021. A Finite Mixture Modelling Perspective for Combining Experts' Opinions with an Application to Quantile-Based Risk Measures. *Risks* 9: 115. https://doi.org/10.3390/risks9060115

Academic Editor: Mogens Steffensen

Received: 27 April 2021
Accepted: 2 June 2021
Published: 9 June 2021

1. Introduction

"Opinion is the medium between knowledge and ignorance" is an expression that is ascribed to Plato. Indeed, due to the growing uncertainty in an abundance of contemporary societal settings, we often come across circumstances when an agent, who acts on behalf of another party, is called to make a decision by combining multiple and sometimes diverging sources of information that can be described as opinions. Moreover, the latter may take any form; from experts to forecasting methods or models (see Clemen and Winkler (2007)), and from now on, we may use these terms interchangeably when referring to an opinion. Opinions communicated to an agent can differ to varying degrees, and the level of confidence that an agent allocates to any given viewpoint is subjective.

Some examples where an idiosyncratic combination of opinions is required for a decision to be made at an individual, corporate, and policy level follow. In the private sphere, consider an individual who plans to sell their house, and in doing so consults property experts to determine an appropriate selling price. While the latter may be influenced by some "standard" factors, such as the number of bedrooms in a given postcode, various experts may additionally examine different price determinants such as the proximity of the property to a good school or a park. That said, the seller may want to incorporate all this diverse information in an effort to achieve a better financial outcome for themselves, but the weight that each reported opinion has in this process lies mostly on the seller's perception. In a financial corporate environment, consider the case where an investment manager asks a number of quantitative analysts to evaluate the return on a stock. Disagreement in opinions here could arise from the fact that some analysts may be more optimistic than others about the future. As mentioned in Peiro (1999), aggregate stock market returns are asymmetrically distributed; the largest movements in the market usually refer to decreases rather than increases in returns. As a result, one can say that an analyst foreseeing a regime shift, let us say, close to a firm's earnings announcement period (see McNichols (1988)) would possibly choose a more heavy-tailed distribution to model the returns compared to others who did not have such a negative expectation. Once again, an investment manager decides on the level of trust to show to any given opinion, based on their own subjective criteria. Finally, at a public policy level and in light of the COVID-19 pandemic, policy makers consult experts from a variety of disciplines, such as anthropology, mathematics, statistics, epidemiology, and engineering, to name a few (see Government Office for Science (2020)) to enable them to build the strategy for its effective management. Reported opinions may not always align, as each specialist sees the problem from a different angle. Government officials though, regardless of divergent opinions, need to combine and evaluate the weight of each view for policy decisions. The subjective character of how much emphasis is given to each opinion by a given policy maker is apparent by the recorded observations of so many different responses related to handling the pandemic across different countries. Similarly to the aforementioned quandaries, decision dilemmas have long been investigated in the particularly rich literature concerning combinations of opinions, which, as Clemen and Winkler (1999) indicate, embraces a number of behavioural and quantitative approaches. See Section 2 for a detailed literature review.

Let us now discuss our motivation behind this study. As is well known, in quantitative risk management, the process of defining, measuring and managing operational risk is crucial since it formalizes the financial institutions' approaches to comply with the qualifying qualitative criteria of the Basel Capital Accord and Solvency Directive. This approach relies on the knowledge of experienced enterprise agents and risk management experts who are asked to provide opinions regarding plausible high-severity events. For instance, these opinions can be expressed as parameters of an assumed loss distribution. However, the company's risk profile, which could accord to a consensus of experts' individual judgements regarding the severity distribution, might often not be robustly estimated. The main reason for this is that when experts are presented with internal data and need to express probabilistic opinions about the same uncertain quantity of interest, there may be multiple sources of heterogeneity in their responses concerning the choice of models and their parameters and, in addition to these, the allocation of weights from the agent that are not considered as being embedded in the data-generative process of the uncertain quantity of interest based on which the agent needs to make a decision. In particular, each expert reports their opinion based on what their focus is, and if we assume that they report their opinions honestly, each believes that their opinion reflects best the true data-generative process. Therefore, since a major challenge in operational risk management is to evaluate the exposure of severe losses based on a weighted combination of a variety of opinions in the first place, it appears that it would make more sense to employ probabilistic models that reflect group structures.

In this paper, we present an alternative perspective for modelling of operational risk in an enterprise context by combining expert opinions based on finite mixture models. Finite mixtures models can provide a formal framework for clustering and classification that can be effectively used within the opinions combination research setting. In particular, this versatile and easily extensible class of models can accommodate different sources of unobserved heterogeneity in the data-generative process of the uncertain quantity of interest by allowing for the mixture components to represent groups within which there is a concurrence of judgements. At this point, it is worth noting that finite mixtures models have not been applied in the area of opinion combinations, with the exception of Rufo et al. (2010), who employed Bayesian hierarchical models based on mixtures of conjugate prior distributions for merging expert opinions. Furthermore, it should be noted that Shevchenko and Wüthrich (2006) employed the Bayesian inference method for quantifying frequency and severity distributions in the context of operational risk. Their approach was based on specifying the prior distributions for the parameters of the frequency and severity distributions based on expert opinions or external data. Furthermore, Lambrigger et al. (2009) extended the framework of the previous paper by developing a Bayesian inference model that permits combining internal data, external data, and expert opinions simultaneously. The setup they proposed enlarged the Bayesian inference models of the exponential dispersion family (EDF) and their corresponding conjugate priors; see, for instance, Bühlmann and Gisler (2006), Chapter 2. However, to the best of our knowledge, the use of finite mixture models within the traditional frequentist approach for combining diverging opinions remains a largely uncharted research territory. Our main contribution is that we consider that the component distributions can stem from different parametric families. The advantage of this formulation is that it allows the agent to obtain the aggregated opinion of a group of experts, based on a linear opinion pool, and account for the various sources of unobserved heterogeneity in the decision-making process in the following ways: (i) by assuming that the data are drawn from a finite mixture distribution with components representing different opinions about both the distribution family and its parameters regarding the uncertain quantity of interest, and (ii) via the mixing weights that reflect the quality of each opinion. Furthermore, when the proposed family of models is applied to internal data, it can enable the agent to utilize all the available information for accurately assessing the effectiveness of (i) the combination of the expert judgements and (ii) their own judgement about the weights that they intended to allocate to each expert—a concept not so dissimilar to the the main idea behind the long-established weights allocation approach of Cooke (1991) and the scoring rules in general. Finally, the proposed family of models is used for numerically computing quantile-based risk measures, which are of interest in a variety of different types of insurance problems, such as setting premiums, insurance deductibles, and reinsurance credance levels and determining reserves or capital and ruin probabilities.

The rest of this paper proceeds as follows. Section 2 provides a brief literature review on some traditional approaches for combining diverging opinions. Section 3 explores the topic of combining diverging opinions using finite mixture models. Section 4 describes the calculation of quantile-based risk measures based on the finite mixture modelling methodology. In our numerical application, we focus on quantile-based risk measures. Finally, concluding remarks can be found in Section 5.

2. Traditional Approaches for Combining Expert Judgements

In this section, we briefly present some famous approaches in aggregating expert judgements. The latter topic can be seen from different perspectives, and in recent decades, several quantitative and behavioural methods have been used for its study. That said, no method can be considered superior to another because for each opinion combination problem, a whole process should be established to identify the most appropriate combination strategy; see Clemen and Winkler (1999). In doing so, factors such as experts' availability, degree of divergence in opinions, past experience regarding the experts, and the random quantity of interest among others should all be considered. As one would expect, such diversity in approaches to combining judgements has resulted in a rich and interdisciplinary literature that would be impossible to cover in its entirety in this article; however, we provide a short review of some important works.

2.1. Behavioural Approaches

Behavioural approaches to opinion aggregation typically involve sources of information, commonly referred to as experts, interacting with each other in order to reach some conclusions. This interaction between experts can happen in a direct or indirect manner. For instance, in an approach called feedback and re-assessment (see Winkler (1968)) no direct communication is allowed among the experts. The agent first collects the views of each individual expert, and then each of them is presented with the other expert opinions and given the opportunity to revise their own view and re-submit it to the agent. Multiple rounds of this process may be required to reach a consensus, or at least to decrease the number of diverging views, thus simplifying decision making. Subsequently, these views may need to be quantitatively combined. One of the earliest methods associated with the feedback and re-assessment approach is known as the Delphi Method; see Linstone and Turoff (1975), Dalkey (1969), and Parenté and Anderson-Parenté (1987).

Additionally, another behavioural aggregation approach, which is known as group re-assessment (Winkler (1968)), allows for direct discussion between experts, after they have individually shared their view, in search of a group opinion consensus. Examples of methods falling into this category are the Nominal Group technique and Kaplan's approach; see Delbecq et al. (1975) and Kaplan (1992), respectively. An advantage of such group reassessment approaches is that experts, when given the opportunity to discuss, may find that there are other factors to consider that would have been otherwise overlooked. However, the fact that the experts need to make a decision as a group comes with certain complications, which we discuss briefly below.

In particular, if we assume that the initial individual opinions are expressed in terms of probability distributions, the moment that discussion between experts starts, each expert also brings considerations about their individual utility function. Furthermore, psychological factors have a role to play in reaching common agreement; some experts may have more advanced leadership skills than others, which may result in the latter adjusting their views merely to reach consensus without necessarily agreeing on the outcome. Last but not least, a phenomenon called polarisation may happen when the group takes riskier decisions as a whole compared to if an individual were to make a decision alone; see Plous (1993) and Wallach et al. (1962). However, this is certainly not to say that the choice to use group decisions is flawed; direct group interactions can be functional in certain circumstances; see Hogarth (1977). Having discussed some behavioural methods, we continue with the presentation of a few quantitative approaches for combining diverging expert opinions.

2.2. Quantitative Approaches

Addressing the problem of combining opinions quantitatively often involves analytical models and procedures operating on individual probability distributions to yield a single combined distribution; see Winkler (1968), French (1983), Genest (1992), Cooke (1991), Clemen (1989), and Clemen and Winkler (1999) for an overview. Focusing on the field of quantitative combination of probability distributions, we see that the linear, logarithmic, and Bayesian pooling methods are typical approaches—a summary of the general ideas behind these methods follows.

The linear pool (see Stone (1961)) and logarithmic pool (see Genest et al. (1984) and Clemen and Winkler (1999)) involve, respectively, a weighted linear or multiplicative combination of the expert probabilities. Out of the two, the linear pool is often perceived as a more attractive combination method because of its intuitiveness and the fact that it satisfies a number of convenient properties; see Cooke (1991) and Clemen and Winkler (1999). In the Bayesian framework, the agent firstly determines a prior distribution over the values of the examined random variable, and then information provided by other sources, say experts, is merely seen as observed data. These "data" are then inserted into a likelihood function along with the prior distribution of the agent to derive a posterior distribution. Although interesting, the implementation of this approach can be challenging in practice; see Bolger and Houlding (2016).

Furthermore, an important note is the meaning of the word "probabilities" in the quantitative opinion combination literature. Whilst traditionally, "probabilities" means mass or density functions for the discrete case and continuous case, respectively Clemen (1989), in recent years, there has been some evidence that combining quantiles, first suggested by Vincent (1912), might be at least as good as combining probability densities (see Lichtendahl et al. (2013), Busetti (2017), Bansal and Palley (2017), Hora et al. (2013), Bogner et al. (2017), and Jose et al. (2013)), despite some criticism from Colson and Cooke (2017). Quantiles combination was also found to be preferable when individual forecasts are biased; see Bamber et al. (2016) and Lichtendahl et al. (2013). Next, we discuss the topic of weights allocation, which, as we will see, is once again a subjective matter depending on the opinions combination problem in question.

2.3. Weights Determination

When combining competing views, the determination of weights is difficult because there are no methods for weights allocation obtained straight from first principles; see Clemen (2008). Nevertheless, the interpretation of weights is flexible, and as Genest and McConway (1990) mention, based on the meaning chosen, one can direct oneself in selecting an appropriate method for their computation. Generally speaking, the weights should somehow reflect the quality of expert opinions; see Bolger and Houlding (2016). When weights are interpreted in this way, the evaluation of the quality of probabilistic forecasts entails the computation of performance measures that account for what has happened in reality; see Winkler et al. (1996) and Gneiting and Raftery (2007).

Such measures, known as scoring rules, play an ex post and ex ante role in the evaluation of probabilities reported; see Winkler et al. (1996): an ex post role because the decision maker needs to first observe what happens in reality before they can truly assess the quality of probabilities experts have reported, and an ex ante role because the experts anticipate the ex post evaluation from the agent and thus have an incentive to be honest when they are expressing their opinions. There are many scoring rules, even though the most preferred are those called strictly proper, meaning that an expert can only maximise their score for an expressed opinion by reporting their forecast honestly; see Winkler et al. (1996) and Gneiting and Raftery (2007). Overall, the choice of scoring rule would in turn lead to different weights; see Genest and McConway (1990).

That said, probably one of the most famous approaches for weights determination, often referred to as the "classical" method, is the one presented in Cooke (1991). There, before making a decision, the agent requests experts to provide their views on quantities whose values are known to the agent but totally unknown to the experts. See Cooke and Goossens (2008) and Eggstaff et al. (2014b) for merits of the "classical" approach, Eggstaff et al. (2014a) for a novel way to make it account for sequential weight updating, and Flandoli et al. (2011) for shortcomings and some alternatives to the "classical" approach. Given the complication involved in weights calculation, the simple averaging scheme is popular in practice because of its perceived robustness and simplicity (see O'Hagan et al. (2006), Lichtendahl et al. (2013)), whilst there is no clear indication that it performs worse than Cooke's approach; see Clemen (2008).

All in all, it should be mentioned that there is limited literature on determining the opinion weights, but for the interested reader, the recent work of Koksalmis and Kabak (2019) provides a comprehensive literature review across various disciplines, suggesting a classification system with the following categories: similarity-based approaches, index-based approaches, clustering-based approaches, integrated approaches, and other approaches. Moving forward to Section 3, we recommend an alternative approach in the area of quantitative combination of probabilistic opinions based on finite mixture models. Such an approach has the benefit of accounting for various forms of heterogeneity among expert views, and straightforward weights computation being able to deal with even a large number of experts.

3. A Finite Mixture Modelling Viewpoint for Opinions Combination

In this section, we present a different approach and incorporate finite mixture models into the diverse set of methodologies for aggregating different opinions. We start by explaining the motivation behind our proposal, followed by a formal presentation of finite mixture models. We then explain how the finite mixture model framework is interpreted for the purposes of combining judgements.

3.1. Motivation Behind the Suggested Approach

Thinking about opinions in the context of distributions or models, in the traditional framework of combining judgements, described in Section 2, each expert gives the agent a different model, and then these individual models are combined into a single model with weights being decided by the agent—most often by taking the weighted average of the individual probability density functions or quantiles. Our proposed finite mixture modelling perspective provides a platform for opinion pooling in two stages. Firstly, we cluster expert opinions of the same kind, and secondly, we perform a convex combination of different clusters using mixing weights that represent the quality of each opinion as this is perceived by the agent. Using historical data, the maximum likelihood (ML) estimation of the parameters and weights can reveal whether both parties are rigorous in their judgements.

3.2. Finite Mixture Models

We start by giving some background on finite mixture models and their use across multiple disciplines and then provide their mathematical definition.

3.2.1. Overview

Finite mixtures is a flexible and easily extensible class of models that account for unobserved heterogeneity; see, for instance, Newcomb (1886), Pearson (1894), Everitt and Hand (1981), Titterington et al. (1985) and McLachlan and Basford (1988) and McLachlan et al. (2019). In particular, starting from a sample of observations, which are assumed to come from a number of underlying classes with unknown proportions, the density of the observations in each of these classes is determined for the purpose of decomposing the sample into its mixture components; see Wedel and DeSarbo (1994). It should be noted that the popularity of mixture models has spread substantially in works of applied and methodological interest across various disciplines such as insurance, economics, finance, biology, genetics, medicine, and most recently in the sphere of artificial intelligence. A few notable works across the aforementioned disciplines include these of Titterington (1990), Samadani (1995), Yung (1997), Allison et al. (2002), Karlis and Xekalaki (2005), McLachlan et al. (2005), Grün and Leisch (2008), Efron (2008), Schlattmann (2009), Bouguila (2010), Mengersen et al. (2011), Elguebaly and Bouguila (2014), Tzougas et al. (2018), Henry et al. (2014), Miljkovic and Grün (2016), Gambacciani and Paolella (2017), Oboh and Bouguila (2017), Tzougas et al. (2014), Miljkovic and Grün (2016), Punzo et al. (2018), Blostein and Miljkovic (2019), Chai Fung et al. (2019), Caravagna et al. (2020), and Bermúdez et al. (2020), though this list is certainly not exhaustive. A short summary of the main characteristics of the class of finite mixture models with component distributions stemming from different parametric families, which we consider in this study follows. The interested reader can also refer to McLachlan and Peel (2000b) for a more detailed treatment of finite mixture models and to McLachlan et al. (2019) for an up-to-date account of the theory and methodological developments underlying their applications.

3.2.2. Definition

Consider that $\mathbf{X} \triangleq \{X_i\}_{i=1}^{\nu}$ is a sample of independent and identically distributed (i.i.d.) random variables from an n-component finite mixture distribution with density function

$$f(x|\Xi) = \sum_{z=1}^{n} \pi_z f_z(x|\boldsymbol{\theta}_z), \tag{1}$$

where $\Xi = (\boldsymbol{\theta}, \boldsymbol{\pi})$, with $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_n)$, where $\boldsymbol{\theta}_z$ denotes the parameters of the zth density function $f_z(\cdot)$, and where $\pi^T = (\pi_1, \pi_2 ..., \pi_n)$ is the vector of component weights, with π_z the prior (or mixing) probability of the component z, where $0 < \pi_z \leq 1$ $\forall z \in \{1, 2, ..., n\}$ and $\sum_{z=1}^{n} \pi_z = 1$ holds. Furthermore, assume that the density functions f_z are absolutely continuous with respect to the Lebesgue measure and are elements from univariate parametric families with a d-dimensional parameter vector $\boldsymbol{\theta}_z$, $\mathfrak{F} = \{f_z(\cdot|\boldsymbol{\theta}_z), \boldsymbol{\theta}_z \in \Theta \subset \mathbb{R}^d\}$.

At this point, it is worth noting that, under the proposed modelling framework, the component distributions $f_z(\cdot)$ in Equation (1) do not necessarily arise from the same parametric family. Therefore, our general approach allows for the design of more flexible models to include a large number of alternative convex combinations of heavy-tailed and light-tailed distributions. Moreover, with this formulation, this class of models can take into account heterogeneity in the data arising from three different sources, differing parameters, differing parametric families, and mixing weights.

3.2.3. Estimation via the Expectation Maximisation Algorithm

Consider the finite mixture model with the associated log-likelihood

$$l(\Xi) = \sum_{i=1}^{\nu} \log(f(x_i|\Xi)),$$

where $f(x_i|\Xi)$ is given by Equation (1). The direct maximization of the above function with respect to the vector of parameters $\Xi = (\boldsymbol{\theta}, \boldsymbol{\pi})$, is complicated. Fortunately, such a task can be easily achieved via the Expectation Maximization (EM) algorithm, which is the standard iterative method that is used for finding ML estimates for models with latent variables; see Dempster et al. (1977). In particular, the popularity of EM algorithm for fitting mixture models to data is such that, as stated in McLachlan et al. (2019), all research works on this topic after 1977 use this method because it unifies the ML estimation from data that can be viewed as being incomplete. For more details regarding the EM algorithm, the interested reader can, for instance, refer to the works of Titterington et al. (1985), McLachlan and Basford (1988), Couvreur (1997), and Karlis and Xekalaki (1999).

Regarding the implementation of the EM algorithm for ML estimation in the context of finite mixture models, we follow the standard approach of combining the observed data, which are represented by the random variable $\boldsymbol{X}$, with the set of unobserved latent random variables $\boldsymbol{w} = (w_{i1}, w_{i2}, ..., w_{in})$, where $w_{iz} = 1$ if the i-th observation belongs to the z-th component, and 0 otherwise, for $i = 1, ..., \nu$ and $z = 1, ..., n$.

Then, the complete data log-likelihood of the model is given by

$$l_c(\Xi) = \sum_{i=1}^{\nu} \sum_{z=1}^{n} w_{iz}[\log(\pi_z) + \log(f_z(x_i|\boldsymbol{\theta}_z)]. \tag{2}$$

In what follows, at the E-Step of the algorithm, it is necessary to compute the Q-function, which is the conditional expectation of the complete data log-likelihood given by Equation (2), while the M-Step consists of maximizing the Q-function with respect to $\Xi = (\boldsymbol{\theta}, \boldsymbol{\pi})$. A generic algorithm is formally described in what follows.

E − Step : Using the current estimates $\pi_z^{(r-1)}$ and $\theta_z^{(r-1)}$ at iteration $r - 1$, calculate the "membership weights":

$$\pi_{iz}^{(r)} = E(w_{iz}|x_i, \Xi^{(r-1)}) = \frac{\pi_z^{(r-1)} f_z(x_i|\boldsymbol{\theta}_z^{(r-1)})}{\sum_{z=1}^{n} \pi_z^{(r-1)} f_z(x_i|\boldsymbol{\theta}_z^{(r-1)})}, \tag{3}$$

for $i = 1, ..., \nu$ and $z = 1, ..., n$. Note that $\pi_{iz}^{(r)}$ is the posterior probability that x_i comes from the mixture component z, calculated at the rth iteration of the EM algorithm. Thus, the Q-function is given by

$$Q(\Xi|\Xi^{(r-1)}) = \sum_{i=1}^{\nu} \sum_{z=1}^{n} \pi_{iz}^{(r)} [\log(\pi_z) + \log(f_z(x_i|\boldsymbol{\theta}_z)).$$

M − Step : Obtain new estimates for $\boldsymbol{\pi}$ and $\boldsymbol{\theta}$ by maximizing the Q-function:

- The updated estimates $\hat{\pi}_z^{(r)}$ are given by:

$$\hat{\pi}_z^{(r)} = \frac{\sum_{i=1}^{\nu} \pi_{iz}^{(r)}}{\nu}, \quad z = 1, \ldots, n.$$

- The updated estimates $\hat{\theta}_z^{(r)}$ are obtained using a weighted likelihood approach for each of the different component distributions with weights $\pi_{iz}^{(r)}$ given by Equation (3). It is clear that ML estimation can be accomplished relatively easily when the M-Step is in closed form. On the contrary, when this is not the case, numerical optimization methods are required for maximizing the the weighted likelihood.

Finally, initialization of parameters can be done using the following data partition methods: (i) means clustering method (see Forgy (1965) and MacQueen (1967)), (ii) Euclidean distance-based initialization (see Maitra (2009)), and (iii) random initialization (see McLachlan and Peel (2000a)).

3.3. Opinions Combination Problem in a Finite Mixture Model Setting

When considering the application of finite mixtures in the area of opinions combination, the framework described in Section 3.2 can be adjusted as follows. A decision maker, otherwise called an agent, needs to make a decision about an $\boldsymbol{X}$ random quantity of interest. Since this decision is made under circumstances of uncertainty, the agent seeks for the opinion of an arbitrary number of consultants $z = 1, 2, \ldots n$ and the combined opinion is seen as a finite mixture model of the type described in 2.3.1 allowing for divergence in expert opinions, both in the class of f_z and in components parameters $\boldsymbol{\theta_z}$. The mixing weights π_z show the level of trust that the agent has to each expert. As in traditional approaches to expert opinions combination, the weights is up to the agent to determine. If the agent has access to older data about $\boldsymbol{X}$, the decision process can be made in two stages to ensure that weights allocation is right.

In the first stage, the agent fits alternative finite mixture models to the available internal data and identifies the mixing probabilities π_z, the class of $f_z(\cdot)$, and the parameters $\boldsymbol{\theta_z}$ that lead to a robust estimation of the company's risk profile. Then, the experts are asked to provide their views on $f_z(\cdot)$ class and $\boldsymbol{\theta_z}$ given the old data without knowing that the agent knows the real answer. The agent checks the reply of the zth expert by comparing it to the correct answer, which, as mentioned previously, it is known to the agent but unknown to the zth expert. In the second stage, the agent needs to make a decision on a totally unknown situation and thus provides the data of real interest to Z experts. Assuming that past experts' performance in getting a good answer indicates their future ability in providing reliable advice, the agent has an indication of how much trust should be given to the zth consultant. Assessing the quality of a probabilistic forecast on an ex post and ex ante basis using real data is not much different from the rationale of using scoring rules as mentioned in Section 2.1. In what follows, we present an application of finite mixture models to combine expert views in a financial setting and in particular when multiple experts are given the task to compute the financial risk measure Value at Risk ($V@R$).

4. Application to a Quantile-Based Financial Risk Measures Setting

In this section, we apply a finite mixture methodology to address the issue of combining diverging expert opinions in an insurance context. We assume that the experts are actuaries and that the opinions expressed by each of them refer to the reserve, or otherwise risk measure, that the institution needs to report to the financial regulator. In particular, without loss of generality, we focus on the popular risk measure called Value at Risk ($V@R$) having the quantile as core ingredient; however our general approach can be applied to any quantile-based risk measure. Any discussion from now on is focused on quantiles because, as we will later on see, the latter is the core ingredient of $V@R$. We start by giving a brief presentation of risk measures and $V@R$ using a general notation, and then the application using simulated finite mixture data follows.

4.1. Motivation Behind the Application

Financial institutions are subject to a number of economic capital requirements following Basel II and Basel III directives in the banking sector and Solvency II and the Swiss Solvency Test in the insurance industry. Since the regulators do not instruct the use of a specific model for the calculation of the reserve, otherwise called risk measure, the choice of any probabilistic model that is used internally by a financial institution for calculating risk measures is crucial.

The above-mentioned challenge known as model risk (see Barrieu and Scandolo (2015) and Barrieu and Ravanelli (2015), among others) is of paramount importance for the health of the financial system along with the choice of the risk measure itself by the regulator; see Danielsson et al. (2001) and Embrechts et al. (2014). The multiple model alternatives for computing a given risk measure can be seen through the prism of an opinions' combination problem. A financial institution, being an agent, instructs actuaries to present alternative internal models for the computation of a risk measure such as $V@R$. In presence of model risk, the agent prefers to use a combination method to take into account the different opinions, i.e., models, prior to reaching a capital reserve decision.

In the context of combining expert opinions for computing quantile-based risk measures, such as $V@R$, there is a clear advantage that the suggested finite mixture modelling approach enjoys over the classical approach of calculating quantiles such as the weighted average of individual quantiles coming from the expert judgements; see, Lichtendahl et al. (2013). This is that it provides a way to assess if the information from the experts that determines the decision-making process of the agent and the data-generative process are highly "synchronous" under a single chosen model in order to ensure that the resulting risk measure value can, as accurately as possible, determine the minimum cushion of economic liquidity.

Finally, under our general approach, which allows for flexibility in the choice of the component distributions which reflect different expert opinions, the resulting risk measures can be calculated using a convex combination of an abundance of alternative heavy-tailed and light-tailed distributions. Thus, since risk measures are equal or proportional to solvency capital requirements, the adopted modelling framework allows us to strike the right balance between calculating risk measures that are not too conservative and hence are preferred by financial institutions and insurance companies who wish to minimise the level of their reserves, since there are many restrictions on how this money can be invested, and computing stricter risk measures that would rather be imposed by regulators who wish to protect consumers. Moving forward, we start by defining financial risk measures in general before narrowing down to Value at Risk, which we use in our application.

4.2. Risk Measures

Financial institutions want to know the minimum amount of capital to add to a position they take in the market to make it acceptable from a regulatory viewpoint. From now on, our random quantity of interest $\boldsymbol{X}$ is a financial position. More precisely, a financial position is a mapping

$$\boldsymbol{X} : \Omega \longrightarrow \mathbb{R} \cup \{+\infty\}$$

where Ω is an non-empty set representing a fixed set of possible scenarios. Let ω be a scenario that is part of Ω. Then $\boldsymbol{X}(\omega)$ reflects the terminal value of the position (profit or loss) at the end of the trading period if the scenario $\omega \in \Omega$ is observed. Assuming that $\mathscr{X}$ is a set of financial positions, we let the financial position $\boldsymbol{X}$ belong to it. Whilst from an economic perspective $\mathscr{X}$ would have to be of a very large size, preferably the space of all $\boldsymbol{X} : \Omega \longrightarrow \mathbb{R} \cup \{+\infty\}$, it is quantitatively convenient to introduce the restriction of boundedness. Furthermore, $\mathscr{X}$ is a linear space containing the constants. At this point, we do not fix a probability measure in Ω.

To calculate the capital requirement, an actuary finds some number $\rho(\boldsymbol{X})$ that quantifies the risk of taking the financial position $\boldsymbol{X}$. In particular, a monetary risk measure ρ is a mapping

$$\rho : \mathscr{X} \longrightarrow \mathbb{R}$$

which satisfies the following conditions

$$\text{Monotonicity: if } \boldsymbol{X} \leq \boldsymbol{Y}, \text{ then } \rho(\boldsymbol{X}) \geq \rho(\boldsymbol{Y})$$

$$\text{Cash invariance: } \rho(\boldsymbol{X} + m) = \rho(\boldsymbol{X}) - m \quad \forall\, m \in \mathbb{R} \text{ for all } \boldsymbol{X}, \boldsymbol{Y} \in \mathscr{X}$$

$$\text{Normalisation: } \rho(0) = 0$$

The condition of monotonicity simply reflects the fact that a position yielding a higher payoff in all scenarios, i.e., in the whole Ω, carries less risk. The cash invariance property demonstrates that risk is measured in monetary units, meaning when an amount m is added to a risky position, its risk will decrease by the same amount m. Normalisation implies that if one has nothing, there is no need to put aside any reserve.

There is a variety of risk measures (see, for instance Barrieu and El Karoui (2005), Föllmer and Schied (2010), Acciaio and Penner (2011), Föllmer and Schied (2016)), and in many cases, quantiles are a key ingredient. That said, for an $\alpha \in (0,1)$, the α-quantile of a random variable on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, where $\mathbb{P}$ is a probability measure on a measurable space $(\Omega, \mathscr{F})$, is any real number Q that satisfies the property

$$\mathbb{P}[\boldsymbol{X} \leq Q] \geq \alpha \text{ and } \mathbb{P}[\boldsymbol{X} < Q] \leq \alpha.$$

The set of all α-quantiles of $\boldsymbol{X}$ is an interval $[Q^-_{\boldsymbol{X}}(\alpha), Q^+_{\boldsymbol{X}}(\alpha)]$ where the lower quantile function of $\boldsymbol{X}$ is

$$Q^-_{\boldsymbol{X}}(t) = \sup\{x | \mathbb{P}[\boldsymbol{X} < x] < t\} = \inf\{x | \mathbb{P}[\boldsymbol{X} \leq x] \geq t\}$$

and the upper quantile function of $\boldsymbol{X}$ is

$$Q^+_{\boldsymbol{X}}(t) = \inf\{x | \mathbb{P}[\boldsymbol{X} \leq x] > t\} = \sup\{x | \mathbb{P}[\boldsymbol{X} < x] \leq t\}$$

A very famous risk measure upon which the financial and insurance industry heavily relies is Value at Risk ($V@R$). If we fix some level $\alpha \in (0,1)$, the $V@R$ of a financial position $\boldsymbol{X}$ at level α is defined as

$$V@R_\alpha(\boldsymbol{X}) := -Q^+_{\boldsymbol{X}}(\alpha) = \inf\{m | \mathbb{P}[\boldsymbol{X} + m < 0] \leq \alpha\}$$

where $Q^+_{\boldsymbol{X}}(\alpha)$ is the upper quantile function of $\boldsymbol{X}$. The financial interpretation of $V@R_\alpha(\boldsymbol{X})$ is the smallest amount of capital, which, if added to the position $\boldsymbol{X}$ and invested in a risk-free manner, ensures that the probability of a negative outcome is below the level α. In the following subsection, we discuss how quantile-based risk measures, such as $V@R$, can be numerically computed in the case of finite mixtures utilising the EM algorithm.

4.3. Computation of V@R Using Finite Mixtures Models

In the context of computing quantile-based risk measures using finite mixtures models, one should take into account that there is no closed-form solution and numerical estimation is required. For expository purposes, we present the numerical calculation of the $V@R$ using a finite mixture modelling methodology in the context of combining diverging expert opinions. However, note that the computation of other quantile risk measures, with many more interesting properties than the $V@R$, such as the Tail Value at Risk ($TV@R$) is straight forward using finite mixture models. For more details, one can refer to Miljkovic and Grün

(2016). Since under the modelling framework we propose the component distributions can stem from different parametric families further interesting results can be obtained.

Let $\boldsymbol{X}$, presented in Section 3.2, be the random vector of ν financial positions of a financial institution introduced in Section 4.2. As we have seen in Section 4.2, $V@R_a(\boldsymbol{X})$ is the α-quantile of the distribution of financial position $\boldsymbol{X}$, and it satisfies the following property

$$\mathbb{P}(\boldsymbol{X} > V@R_a(\boldsymbol{X})) = 1 - \alpha$$

Since in the context of finite mixture models, the $V@R_a(\boldsymbol{X})$ does not have a closed-form solution, we compute it numerically by solving Equation (4)

$$F_{\boldsymbol{X}}(V@R_a(\boldsymbol{X})) = \alpha \tag{4}$$

where $F_{\boldsymbol{X}} := \mathbb{R} \longrightarrow [0,1]$ is the cumulative distribution function of the random financial position $\boldsymbol{X}$.

In particular, the numerical computation of $V@R_a(\boldsymbol{X})$ can be achieved easily using the R programming language in a two-step process. Firstly, we create an R function according to Equation (1) with the only difference that now f_z is replaced by $F_{\boldsymbol{X}_z}$ as follows

$$F_{\boldsymbol{X}}(x|\Xi) = \sum_{z=1}^{n} \pi_z F_{\boldsymbol{X}_z}(x|\boldsymbol{\theta}_{\boldsymbol{z}})$$

where $\Xi = (\boldsymbol{\theta}, \boldsymbol{\pi})$, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_n)$, and $\pi^T = (\pi_1, \pi_2 ..., \pi_n)$ represents the vector of unknown parameters, π_z is the prior (or mixing) probability of the component z where $0 < \pi_z \leq 1 \; \forall z \in \{1, 2, ..., n\}$ and $\sum\limits_{z=1}^{n} \pi_z = 1$ holds. Secondly, we create the inverse function of $F_{\boldsymbol{X}_z}$ denoted as $F_{\boldsymbol{X}_z}^{-1}$, which is the $V@R_a(X)$. The function $F_{\boldsymbol{X}_z}^{-1}$ is derived in R by returning the $uniroot()$ argument in the package **stats** in R of $(F_{\boldsymbol{X}_z} - \alpha)$ for a pre-determined quantile bracket. It should be mentioned that in order to evaluate $F_{\boldsymbol{X}_z}$ at the point x, one needs to utilise the EM algorithm to estimate the parameters and mixing probabilities of $F_{\boldsymbol{X}}$. In the end, in order to calculate the quantile using the function $F_{\boldsymbol{X}_z}^{-1}$, one just needs to insert as arguments the the percentile α upon which $V@R_a(\boldsymbol{X})$ will be calculated as well as the vectors of estimated parameters and mixing probabilities. In Section 4.4, we present our numerical application.

4.4. Numerical Application

In this subsection, a numerical example is presented to illustrate the proposed approach for combining expert opinions. In particular, without loss of generality, we assume that the experts can be classified into two groups within each of which there is a consensus of opinions. In this context, the components represent different expert opinions about the distribution family and its parameters, whilst the weights reflect the quality of each opinion as this may be assessed by the agent.

In what follows, we generate multiple samples from two-component mixtures of some classical distributions where the components of the mixture do not necessarily belong to the same parametric family. In particular, we consider the two-component (2C) Normal, 2C Gamma, 2C Lognormal, 2C Pareto mixtures, and also the 2C Lognormal-Gamma, and 2C Pareto-Gamma mixtures. Note that when using real data, one can distinguish between the competing models by employing the Deviance (DEV), Akaike information criterion (AIC), and the Schwartz Bayesian criterion (SBC). Furthermore, the prediction performances of the models can be assessed via out-of-sample validation. The prediction performances can be measured using the root-mean squared error (RMSE) and the deviance statistic. To provide a potential practical application of the proposed perspective, note that the convex combination of moderate and heavy-tailed distributions, similar to the aforementioned ones, can be used for efficiently approximating positive insurance loss data with right skewness, which can often be represented as an amalgamation of losses of different magnitudes. For example, Tzougas et al. (2014) and Miljkovic and Grün (2016) proposed the use of mixtures of finite mixture claim severity models in an actuarial setting, whilst Tzougas et al. (2018) and Blostein and Miljkovic (2019) considered finite mixture models where all components of the mixture are not necessarily assumed to be from the same parametric family.

The probability density functions (pdfs) of the component distributions, denoted by f_z in Equation (1), are given by Equations (5)–(8) below.

- Normal distribution: the pdf of the Normal distribution is given by:

$$f_z(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}e^{---\frac{(x-\mu)^2}{2\sigma^2}} \tag{5}$$

for $-\infty < x < \infty$ where $-\infty < \mu < \infty$ and $\sigma > 0$. The mean of $\boldsymbol{X}$ is given by $\mathbb{E}(\boldsymbol{X}) = \mu$ and the variance of X by $Var(\boldsymbol{X}) = \sigma^2$. This parametric family is chosen for relatively symmetric insurance loss data, which take either positive or negative values.

- Lognormal distribution: the pdf of the Lognormal distribution is as follows:

$$f_z(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{x}e^{-\frac{[\log(x)-\mu]^2}{2\sigma^2}} \tag{6}$$

for $x > 0$ where $\mu > 0$ and $\sigma > 0$. Here, $\mathbb{E}(\boldsymbol{X}) = c^{\frac{1}{2}}e^{\mu}$ and $Var(\boldsymbol{X}) = c(c-1)e^{2\mu}$ where $c = e^{\sigma^2}$.

- Gamma distribution: the density of the Gamma is given by:

$$f_z(x|\mu,\sigma) = \frac{1}{(\sigma^2\mu)^{\frac{1}{\sigma^2}}}\frac{x^{\frac{1}{\sigma^2}-1}e^{\frac{-x}{(\sigma^2\mu)}}}{\Gamma(\frac{1}{\sigma^2})} \tag{7}$$

for $x > 0$, where $\mu > 0$ and $\sigma > 0$. This is a re-parameterisation, which was given in Equation (17.23) of Johnson et al. (1994) in p.343, and it can be obtained by setting $\sigma = \frac{1}{\alpha}$ and $\mu = \alpha\beta$. Moreover, $\mathbb{E}(\boldsymbol{X}) = \mu$ and $Var(\boldsymbol{X}) = \sigma^2\mu^2$. The Gamma has a less heavier tail than the Lognormal one.

- Pareto distribution: the pdf of the Pareto distribution is as follows:

$$f_z(x|\mu,\sigma) = \frac{1}{\sigma}\mu^{\frac{1}{\sigma}}(x+\mu)^{-\frac{1}{\sigma+1}} \tag{8}$$

for $x \geq 0$, where $\mu > 0$ and $\sigma > 0$. Furthermore, $\mathbb{E}(\boldsymbol{X}) = \mu(\frac{1}{\sigma}-1)$, and $Var(\boldsymbol{X}) = (\mu(\frac{1}{\sigma}-1))^2(\frac{1}{1-2\sigma})$ exists only if $\sigma < 1/2$. This is an alternative distributional class choice that may be preferred to model more heavily right-skewed insurance loss data than the previous two distribution choices.

As described in Section 3.3, the fitting of such mixture distributions can be achieved via the EM algorithm, which is implemented for estimating both the parameters of each mixture component distribution and mixing weights. Subsequently, using these estimated values, we proceed with calculating the quantiles of the mixture models across all estimated weights combinations and for various probability levels $(1-\alpha)$. At this point, it should be noted that we choose to compute quantiles directly from the finite mixture models, but for comparison purposes, we also combine quantiles for each expert view as it has often been encountered in the literature; see, for instance, Lichtendahl et al. (2013). Note also that the calculation of risk measures using finite mixture models has also been addressed by Miljkovic and Grün (2016) and Blostein and Miljkovic (2019). However, we would like to emphasise that this is the first time that the 2C Pareto and 2C Pareto-Gamma models are used for computing quantile based risk measures. Therefore, this constitutes one more novelty of our work in addition to proposing the finite mixture modelling approach as an efficient tool for combining expert opinions.

The results of our numerical application for each of the 2C component mixture models we consider in this study, namely the Normal, Gamma, Lognormal, Pareto, Lognormal-Gamma, and Pareto-Gamma, are presented in the following manner. In Table 1, for each of the previously described 2C mixture models, we show the parameters estimates across all estimated weights combinations derived using the EM algorithm. Then, in Table 2, we present the 2C mixture model-based quantiles which are computed by utilising the EM algorithm parameter and weight estimates, which are presented in Table 1, as well as the quantiles derived by using the weighted average approach across all weights combinations that are used to generate the data. Both quantile types are calculated at two widely used, in a financial context, probability levels $(1 - \alpha)$, i.e., 0.950 and 0.990. Finally, in Figures 1–6, we plot the mixture-model-based quantiles and the weighted average-based quantiles computed at a more extended range of $(1 - \alpha)$ probability levels ranging from 0.950 to 0.995. It is important to mention that the values for the two quantile types of interest appear to be substantially different, and therefore we deemed it necessary to have two distinct y axes in each plot of Figures 1–6 to allow for an easier comparison.

As we observe, the quantile values in the case of the 2C Normal, 2C Gamma, 2C Lognormal, and 2C Pareto mixtures, and also the 2C Lognormal-Gamma and 2C Pareto-Gamma mixtures, are higher than the weighted-average-based ones. Regarding the decision-making problem we address, as was previously mentioned, the approach we consider is more flexible because it provides a two-fold benefit to the decision maker, since, in addition to enabling them to evaluate the efficacy of the expert views aggregation process, it allows them to test how the weights that they were intending to allocate to each expert opinion based on their personal judgement compared to the ones estimated by the model.

Table 1. EM algorithm estimates for various two-component (2C) finite mixture models. Estimates refer to the parameters mean ($\hat{\mu_1}, \hat{\mu_2}$) and standard deviation ($\hat{\sigma_1}, \hat{\sigma_2}$) and the mixing weights of each mixture component ($\hat{\pi_1}, 1 - \hat{\pi_1}$) across all plausible mixing weight combinations (π_1) in the true data generative process. All estimates provided are statistically significant at a 5% threshold or below.

Parametric Family	$\hat{\pi}_1$	$\hat{\pi}_2$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$
	0.098	0.902	5271.210	1333.022	0.228	0.442
	0.203	0.797	5273.012	1342.762	0.231	0.448
	0.300	0.700	5272.341	1341.234	0.225	0.448
	0.400	0.600	5271.032	1340.012	0.221	0.446
2C Normal	0.500	0.500	5272.002	1342.569	0.230	0.447
	0.600	0.400	5272.321	1343.812	0.238	0.449
	0.700	0.300	5270.921	1341.989	0.236	0.444
	0.800	0.200	5271.981	1345.091	0.239	0.449
	0.901	0.099	5273.182	1343.991	0.240	0.447
	0.090	0.910	9.538	8.042	0.723	0.884
	0.200	0.800	9.521	8.025	0.717	0.879
	0.299	0.701	9.539	8.042	0.772	0.883
	0.398	0.602	9.537	8.041	0.771	0.881
2C Lognormal	0.498	0.502	9.538	8.064	0.778	0.899
	0.600	0.400	9.548	8.053	0.766	0.896
	0.700	0.300	9.528	8.035	0.741	0.873
	0.802	0.198	9.508	0.722	8.016	0.858
	0.901	0.099	9.511	0.733	8.023	0.867
	0.086	0.914	6786.348	3162.126	0.625	0.352
	0.207	0.793	6737.558	3127.165	0.629	0.340
	0.307	0.693	6738.557	3124.408	0.635	0.342
	0.400	0.600	6739.659	3127.512	0.629	0.344
2C Gamma	0.499	0.501	6784.309	3171.627	0.630	0.357
	0.601	0.399	6754.742	3123.512	0.638	0.341
	0.700	0.300	6783.127	3170.006	0.636	0.346
	0.799	0.201	6783.021	3172.871	0.621	0.341
	0.902	0.098	6786.735	3172.513	0.599	0.343
	0.088	0.912	1364.138	3148.568	3.354	2.439
	0.204	0.796	1329.177	3099.778	3.342	2.442
	0.295	0.705	1326.426	3100.769	3.344	2.448
	0.405	0.595	1329.524	3101.871	3.346	2.443
2C Pareto	0.494	0.506	1373.639	3146.521	3.359	2.444
	0.605	0.395	1325.524	3116.954	3.343	2.452
	0.694	0.306	1372.018	3145.339	3.348	2.450
	0.805	0.195	1374.883	3145.233	3.343	2.434
	0.896	0.104	1374.525	3148.947	3.345	2.422
	0.088	0.912	1902.904	2393.673	2.085	0.604
	0.204	0.796	1867.943	2344.883	2.073	0.608
	0.295	0.705	1865.186	2345.882	2.075	0.614
	0.404	0.596	1868.129	2346.984	2.077	0.608
2C Lognormal-Gamma	0.494	0.506	1912.405	2391.634	2.079	0.609
	0.605	0.395	1864.289	2362.067	2.074	0.617
	0.694	0.306	1910.784	2390.452	2.079	0.615
	0.805	0.195	1913.649	2394.123	2.074	0.602
	0.896	0.104	1912.018	2396.106	2.076	0.599
	0.088	0.912	9.538	3175.526	0.725	0.737
	0.197	0.803	9.522	3140.588	0.722	0.741
	0.297	0.703	9.543	3139.065	0.781	0.747
	0.396	0.604	9.547	3142.169	0.777	0.742
2C Pareto-Gamma	0.496	0.504	9.547	3186.286	0.784	0.743
	0.598	0.402	9.558	3138.171	0.772	0.751
	0.698	0.302	9.537	3184.665	0.765	0.749
	0.800	0.200	9.517	3187.353	0.728	0.734
	0.899	0.101	9.520	3189.272	0.739	0.729

Table 2. Comparison between the finite mixture model-based $(1-\alpha)$ quantile $Q_{mix.}$ and the weighted average-based $(1-\alpha)$ quantile $Q_{w.a.}$ derived for the various parametric families considered in this study. Note that $(1-\alpha)$ denotes the probability level at which the quantile is computed. The quantile $Q_{mix.}$ is calculated by using the derived EM parameters and weights estimates shown in Table 1. For the computation of quantile $Q_{w.a.}$, no model estimation is involved, and it is calculated as the weighted average of two individual quantiles, each coming from a distribution family with parameters and weights as those used to generate the data.

Finite Mixture Model-Based Quantile $Q_{mix.}$						
$1-\alpha$	2C Normal	2C Lognormal	2C Gamma	2C Pareto	2C Lognormal-Gamma	2C Pareto-Gamma
0.950	5271.204	18,674.230	6254.920	7221.820	13,053.610	5189.413
0.990	5271.499	37,254.660	11,911.220	16,895.230	33,302.150	7286.269
0.950	5273.171	24,144.120	9057.693	6569.990	22,039.780	5157.749
0.990	5273.394	45,694.640	14,982.950	15,542.580	44,546.370	7797.691
0.950	5272.559	30,361.480	10,637.600	6108.752	29,534.510	5231.247
0.990	5272.754	57,847.270	16,443.240	14,589.80	58,206.920	8493.526
0.950	5271.286	34,225.830	11,600.840	5564.221	34,066.590	5273.603
0.990	5271.465	63,167.590	17,229.970	13,503.800	63,994.280	9481.466
0.950	5272.297	38,063.000	12,505.360	5147.291	38,106.660	5449.650
0.990	5272.474	68,905.430	18,108.690	12,655.920	69,880.110	10,662.670
0.950	5272.650	40,758.710	13,203.380	4409.509	41,120.820	5487.197
0.990	5272.827	71,836.230	18,826.360	11,044.520	73,116.860	11,643.710
0.950	5271.267	40,870.610	13,793.70	3904.314	42,478.650	5662.312
0.990	5271.438	69,767.140	19,373.060	9889.721	73,938.670	12,784.550
0.950	5272.348	40,907.890	14,083.190	3191.665	41,516.230	5810.997
0.990	5272.517	68,097.500	19,440.430	8118.793	69,476.320	13,964.470
0.950	5273.564	43,519.640	14,211.730	2596.741	44,222.170	5958.190
0.990	5273.731	72,328.780	19,271.110	6335.289	73,824.540	14,757.320

Weighted average-based quantile $Q_{w.a.}$						
$1-\alpha$	2C Normal	2C Lognormal	2C Gamma	2C Pareto	2C Lognormal-Gamma	2C Pareto-Gamma
0.950	1713.689	15,847.350	6130.278	6974.334	9075.481	5234.196
0.990	1713.975	28,081.790	7666.813	16,133.486	12,933.270	7791.443
0.950	2106.655	19,018.620	7108.780	6415.912	12,999.186	5328.475
0.990	2106.926	33,072.860	9074.596	14,783.290	19,607.500	8635.935
0.950	2499.620	22,189.900	8087.282	5857.489	16,922.892	5422.755
0.990	2499.877	38,063.930	10,482.379	13,433.094	32,955.980	9480.427
0.950	2892.586	25,361.170	9065.785	5299.066	20,846.597	5517.035
0.990	2892.828	43,054.990	11,890.162	12,082.899	26,281.740	10,324.919
0.950	3285.551	28,532.450	10,044.287	4740.644	24,770.303	5611.315
0.990	3285.779	48,046.060	13,297.946	10,732.703	39,630.220	11,169.410
0.950	3678.516	31,703.730	11,022.790	4182.221	28,694.008	5705.594
0.990	3678.730	53,037.130	14,705.729	9382.508	46,304.450	12,013.902
0.950	4071.482	34,875.000	12,001.292	3623.798	32,617.714	5799.874
0.990	4071.682	58,028.200	16,113.512	8032.312	52,978.690	12,858.394
0.950	4464.447	38,046.280	12,979.794	3065.376	36,541.419	5894.154
0.990	4464.633	63,019.270	17,521.295	6682.116	59,652.930	13,702.886
0.950	4857.413	41,217.550	13,958.297	2506.953	40,465.125	5988.433
0.990	4857.584	68,010.340	18,929.078	5331.921	66,327.170	14,547.377

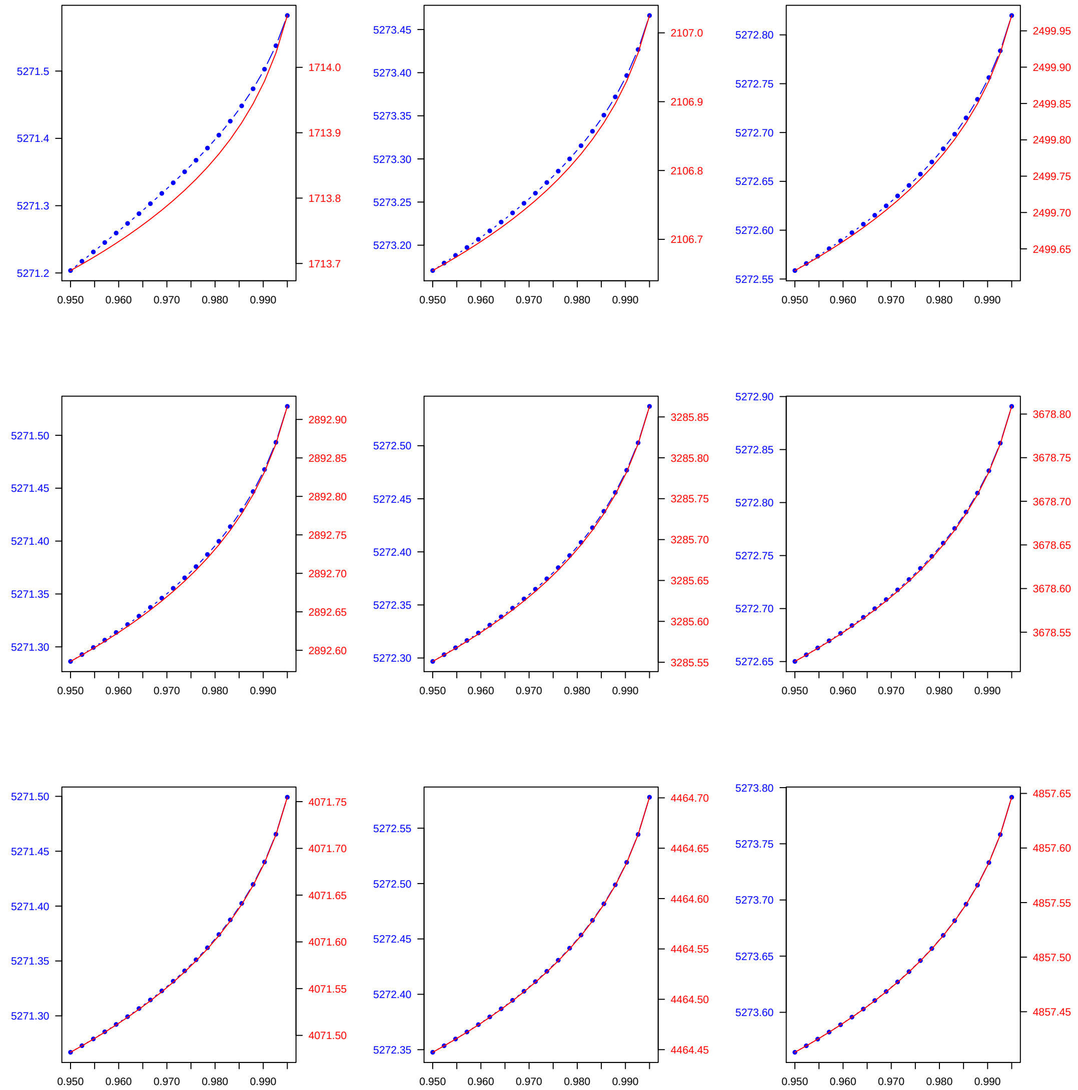


Figure 1. Two-component (2C) Normal finite mixture model-based $(1 - \alpha)$ quantile $Q_{mix.}$ (blue colour) across all $(\hat{\pi}_1, \hat{\pi}_2)$ combinations versus two-component (2C) Normal weighted average-based $(1 - \alpha)$ quantile $Q_{w.a.}$ (red colour) across all (π_1, π_2) combinations, where $(1 - \alpha)$ takes values in the range of 0.950-0.995. Note that, due to a considerable discrepancy between $Q_{mix.}$ and $Q_{w.a.}$ values, each given plot has two different y axes—one for each quantile type.

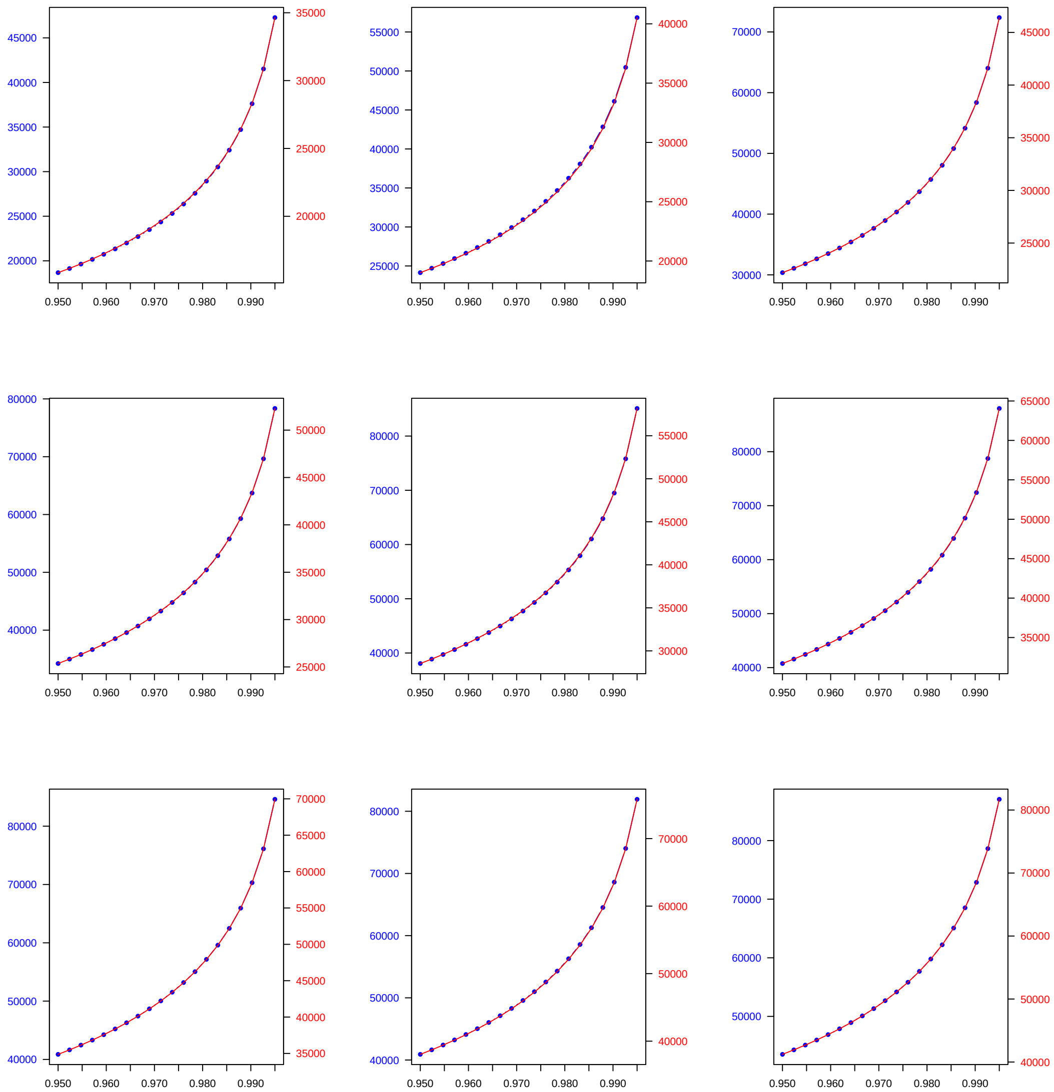


Figure 2. Two-Component (2C) Lognormal finite mixture model-based $(1-\alpha)$ quantile $Q_{mix.}$ (blue colour) across all $(\hat{\pi}_1, \hat{\pi}_2)$ combinations versus two-component (2C) Lognormal weighted average-based $(1-\alpha)$ quantile $Q_{w.a.}$ (red colour) across all (π_1, π_2) combinations, where $(1-\alpha)$ takes values in the range of 0.950-0.995. Note that, due to a considerable discrepancy between $Q_{mix.}$ and $Q_{w.a.}$ values, each given plot has two different y axes—one for each quantile type.

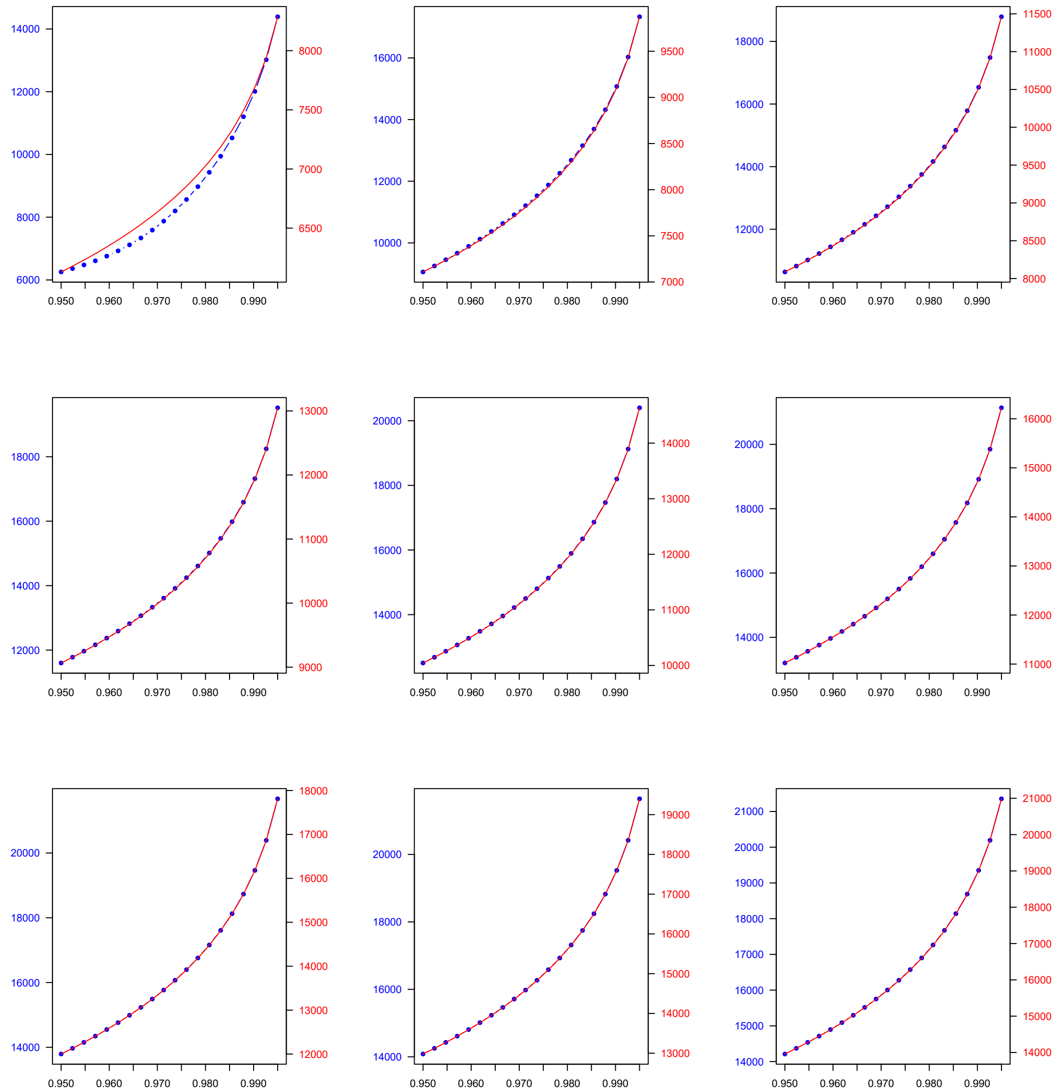


Figure 3. Two-Component (2C) Gamma finite mixture model-based $(1-\alpha)$ quantile $Q_{mix.}$ (blue colour) across all $(\hat{\pi}_1, \hat{\pi}_2)$ combinations versus two-component (2C) Gamma weighted average-based $(1-\alpha)$ quantile $Q_{w.a.}$ (red colour) across all (π_1, π_2) combinations, where $(1-\alpha)$ takes values in the range of 0.950-0.995. Note that, due to a considerable discrepancy between $Q_{mix.}$ and $Q_{w.a.}$ values, each given plot has two different y axes—one for each quantile type.

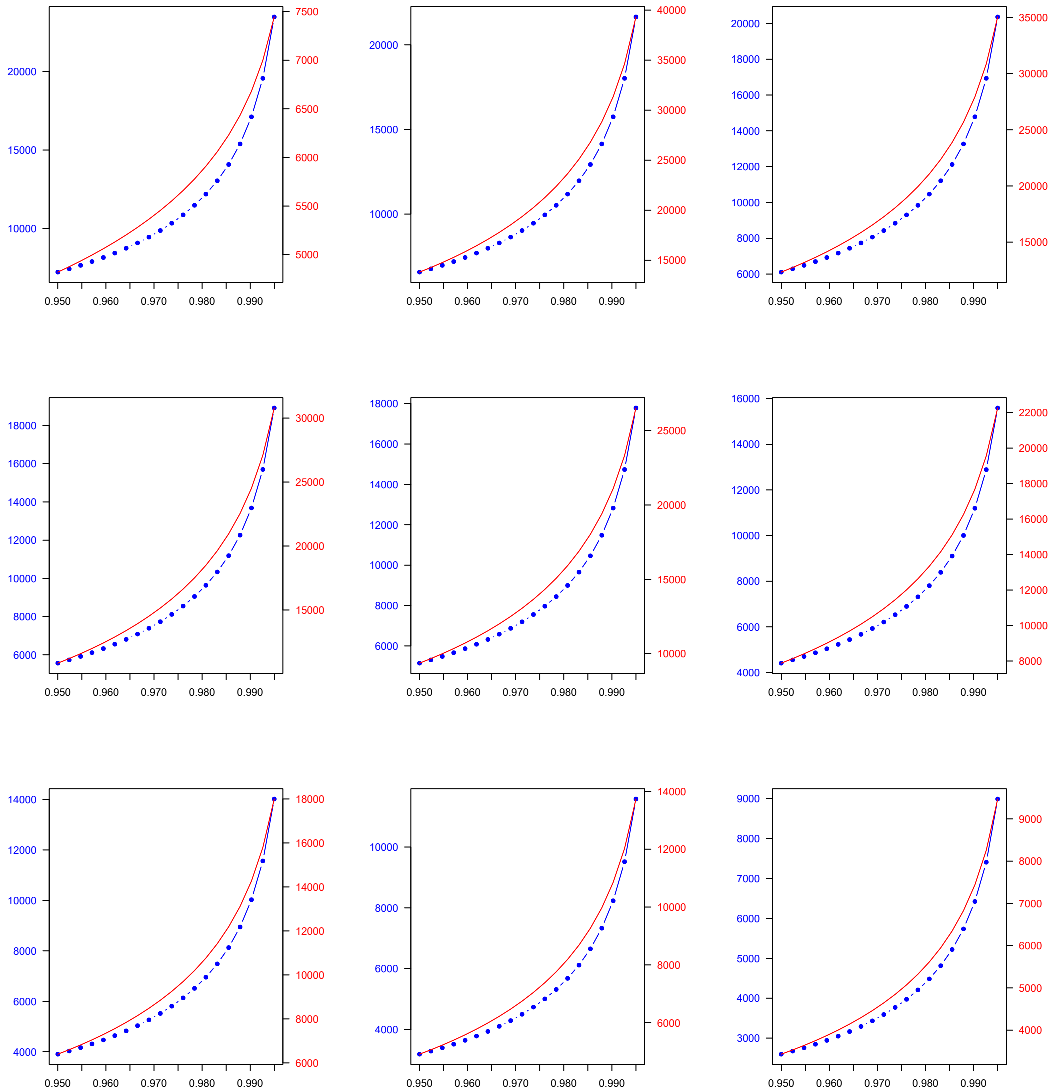


Figure 4. Two-Component (2C) Pareto finite mixture model-based $(1-\alpha)$ quantile $Q_{mix.}$ (blue colour) across all $(\hat{\pi}_1, \hat{\pi}_2)$ combinations versus two-component (2C) Pareto weighted average-based $(1-\alpha)$ quantile $Q_{w.a.}$ (red colour) across all (π_1, π_2) combinations, where $(1-\alpha)$ takes values in the range of 0.950-0.995. Note that, due to a considerable discrepancy between $Q_{mix.}$ and $Q_{w.a.}$ values, each given plot has two different y axes—one for each quantile type.

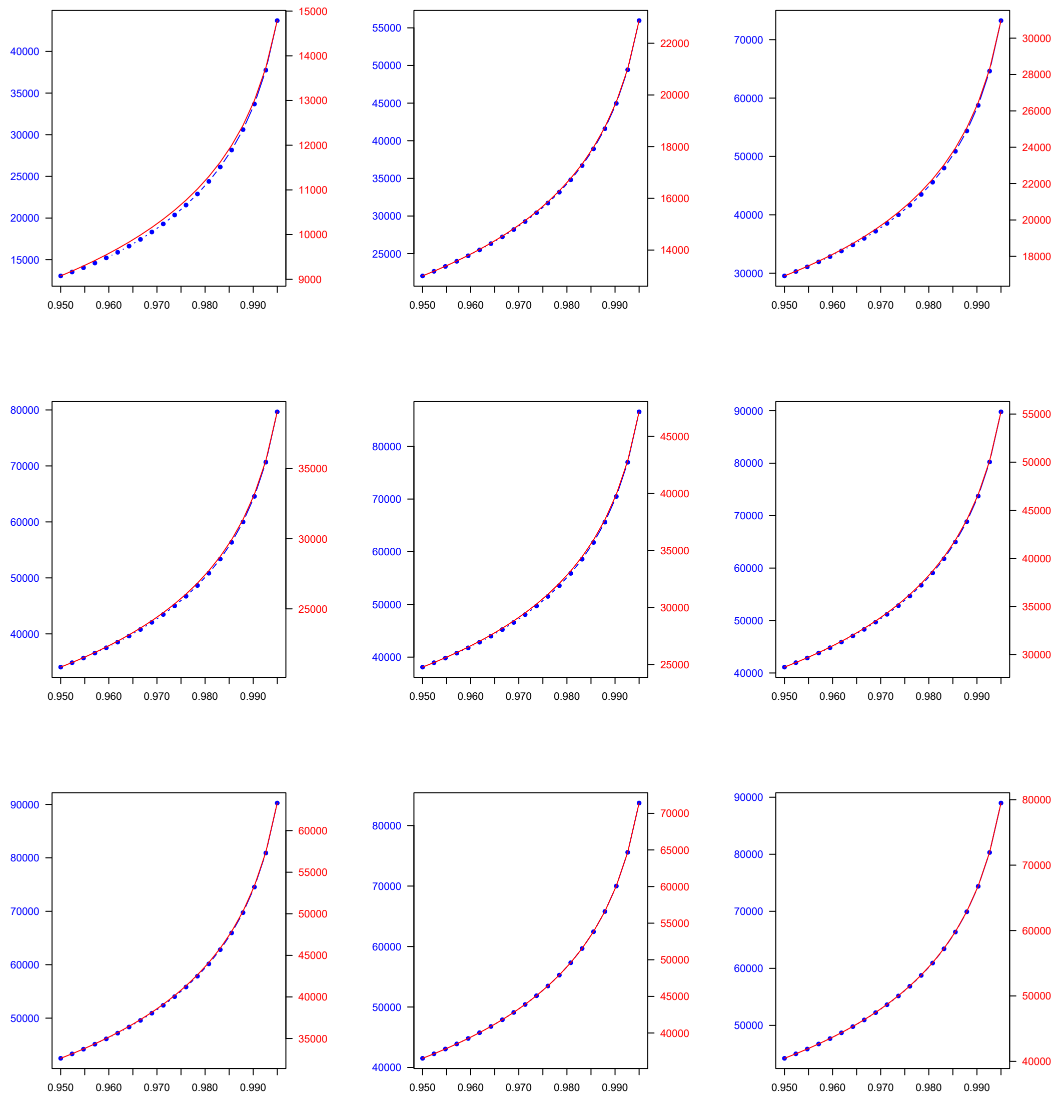


Figure 5. Two-Component (2C) Lognormal-Gamma finite mixture model-based $(1-\alpha)$ quantile $Q_{mix.}$ (blue colour) across all $(\hat{\pi}_1, \hat{\pi}_2)$ combinations versus two-component (2C) Lognormal-Gamma weighted average-based $(1-\alpha)$ quantile $Q_{w.a.}$ (red colour) across all (π_1, π_2) combinations, where $(1-\alpha)$ takes values in the range of 0.950-0.995. Note that, due to a considerable discrepancy between $Q_{mix.}$ and $Q_{w.a.}$ values, each given plot has two different y axes-one for each quantile type.

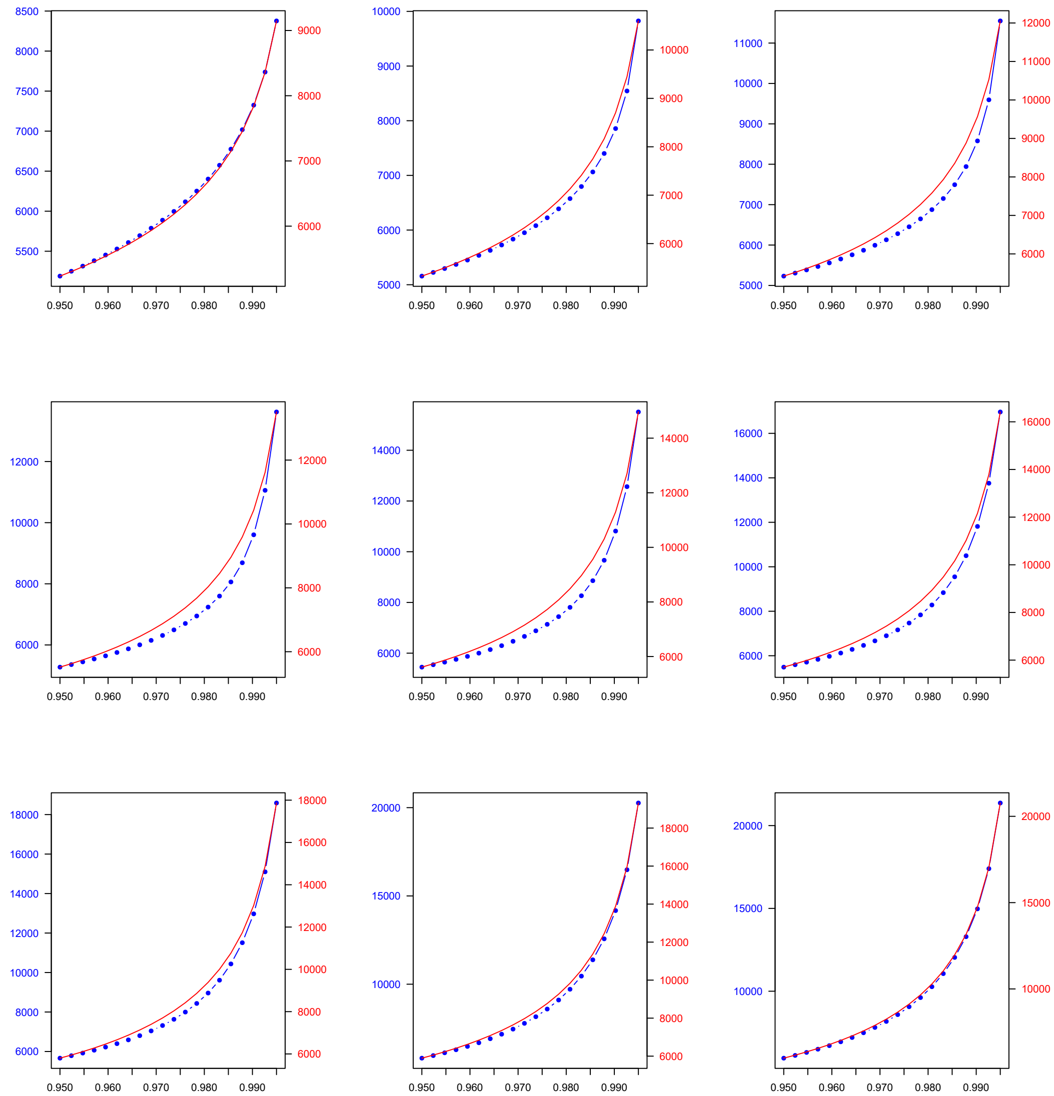


Figure 6. Two-Component (2C) Pareto-Gamma finite mixture model-based $(1-\alpha)$ quantile $Q_{mix.}$ (blue colour) across all $(\hat{\pi}_1, \hat{\pi}_2)$ combinations versus two-component (2C) Pareto-Gamma weighted average-based $(1-\alpha)$ quantile $Q_{w.a.}$ (red colour) across all (π_1, π_2) combinations, where $(1-\alpha)$ takes values in the range of 0.950-0.995. Note that, due to a considerable discrepancy between $Q_{mix.}$ and $Q_{w.a.}$ values, each given plot has two different y axes—one for each quantile type.

5. Concluding Remarks

When making a decision in an uncertain environment, an agent may consult multiple experts. In such a scenario, the aggregation of individual opinions before reaching a decision is required. In this study, we contribute to the plethora of interdisciplinary literature on this topic by proposing a finite mixture modelling approach that can enable the agent to combine the component distributions in order to obtain a single distribution of the quantity of interest that is a quantile-based risk measure. The component distributions we consider in this study can be used in practice to model various quantities of interest in financial and insurance applications such as financial returns and insurance losses with light and heavy tails. The suggested method allows for considerable flexibility in expert opinions regarding the distribution class of the random quantity of interest and its parameters, and it also provides an efficient way for weights computation—a task recognised as being particularly strenuous in this segment of literature. By employing the perspective that opinions take the form of quantiles, we compare our approach to the traditional weighted average one, and we find that they lead to different results. Furthermore, the proposed models can be used for carrying out different tasks in insurance such as calculating premiums and reserves and measuring tail risk.

A compelling direction of further research would be to use combinations of finite mixtures and composite models that can mitigate instabilities of tail index estimations inherited by finite mixture models; see, for instance, Fung et al. (2021). Furthermore, a natural extension of our study is to employ Bayesian inference for mixtures, which will allow us to combine internal data, external data, and expert opinions proceeding along similar lines as in Lambrigger et al. (2009). Additionally, in this paper, we have focused only on the opinion aggregation process without considering how experts have elicited their views; therefore, it would be interesting to examine ways in which this aspect is also taken into account. Finally, another potential topic of interest, with regards to weights allocation this time, is for the weights to reflect the risk aversion level of the agent as well as the quality of a given expert's judgement, and the level of disagreement between experts.

Author Contributions: Conceptualization, D.M. and P.B.; methodology, D.M., P.B. and G.T.; software, D.M. and G.T.; formal analysis, D.M., P.B. and G.T.; investigation, D.M., P.B. and G.T.; writing—original draft preparation, D.M.; writing—review and editing, D.M. and G.T.; supervision, P.B. and G.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EM	Expectation Maximization
ML	Maximum Likelihood
V@R	Value at Risk
TV@R	Tail Value at Risk
2C	Two component
pdf	Probability density function

References

Acciaio, Beatrice, and Irina Penner. 2011. Dynamic risk measures. In *Advanced Mathematical Methods for Finance*. Berlin/Heidelberg: Springer, pp. 1–34.

Allison, David B., Gary L. Gadbury, Moonseong Heo, José R Fernández, Cheol-Koo Lee, Tomas A Prolla, and Richard Weindruch. 2002. A mixture model approach for the analysis of microarray gene expression data. *Computational Statistics & Data Analysis* 9: 1–20.

Bamber, Jonathan L., Willy P. Aspinall, and Roger M. Cooke. 2016. A commentary on "how to interpret expert judgment assessments of twenty-first century sea-level rise" by Hylke de Vries and Roderik SW van de Wal. *Climatic Change* 137: 321–28. [CrossRef]

Bansal, Saurabh, and Asa Palley. 2017. Is it better to elicit quantile or probability judgments? A comparison of direct and calibrated procedures for estimating a continuous distribution. In *A Comparison of Direct and Calibrated Procedures for Estimating a Continuous Distribution (6 June 2017). Kelley School of Business Research Paper*. Bloomington: Indiana University, pp. 17–44.

Barrieu, Pauline, and Claudia Ravanelli. 2015. Robust capital requirements with model risk. *Economic Notes: Review of Banking, Finance and Monetary Economics* 44: 1–28. [CrossRef]

Barrieu, Pauline, and Giacomo Scandolo. 2015. Assessing financial model risk. *European Journal of Operational Research* 242: 546–56. [CrossRef]

Barrieu, Pauline, and Nicole El Karoui. 2005. Inf-convolution of risk measures and optimal risk transfer. *Finance and Stochastics* 9: 269–98. [CrossRef]

Bermúdez, Lluís, Dimitris Karlis, and Isabel Morillo. 2020. Modelling unobserved heterogeneity in claim counts using finite mixture models. *Risks* 8: 10. [CrossRef]

Blostein, Martin, and Tatjana Miljkovic. 2019. On modeling left-truncated loss data using mixtures of distributions. *Insurance: Mathematics and Economics* 85: 35–46. [CrossRef]

Bogner, Konrad, Katharina Liechti, and Massimiliano Zappa. 2017. Combining quantile forecasts and predictive distributions of streamflows. *Hydrology and Earth System Sciences* 21: 5493–502. [CrossRef]

Bolger, Donnacha, and Brett Houlding. 2016. Reliability updating in linear opinion pooling for multiple decision makers. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 230: 309–22. [CrossRef]

Bouguila, Nizar. 2010. Count data modeling and classification using finite mixtures of distributions. *IEEE Transactions on Neural Networks* 22: 186–98. [CrossRef]

Bühlmann, Hans, and Alois Gisler. 2006. *A Course in Credibility Theory and Its Applications*. Zurich: Springer Science & Business Media.

Busetti, Fabio. 2017. Quantile aggregation of density forecasts. *Oxford Bulletin of Economics and Statistics* 79: 495–512. [CrossRef]

Caravagna, Giulio, Timon Heide, Marc J. Williams, Luis Zapata, Daniel Nichol, Ketevan Chkhaidze, William Cross, George D. Cresswell, Benjamin Werner, Ahmet Acar, and et al. 2020. Subclonal reconstruction of tumors by using machine learning and population genetics. *Nature Genetics* 52: 898–907. [CrossRef]

Chai Fung, Tsz, Andrei L. Badescu, and X. Sheldon Lin. 2019. A class of mixture of experts models for general insurance: Application to correlated claim frequencies. *ASTIN Bulletin* 49: 647–88. [CrossRef]

Clemen, Robert T. 1989. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* 5: 559–83. [CrossRef]

Clemen, Robert T. 2008. Comment on cooke's classical method. *Reliability Engineering & System Safety* 93: 760–5.

Clemen, Robert T., and Robert L. Winkler. 1999. Combining probability distributions from experts in risk analysis. *Risk Analysis* 19: 187–203. [CrossRef]

Clemen, Robert T., and Robert L. Winkler. 2007. *Aggregating Probability Distributions*. Cambridge: Cambridge University Press, pp. 154–76. [CrossRef]

Colson, Abigail R., and Roger M. Cooke. 2017. Cross validation for the classical model of structured expert judgment. *Reliability Engineering & System Safety* 163: 109–20.

Cooke, Roger. 1991. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford: Oxford University Press on Demand.

Cooke, Roger M., and Louis L. H. J. Goossens. 2008. Tu delft expert judgment data base. *Reliability Engineering & System Safety* 93: 657–74.

Couvreur, Christophe. 1997. The EM algorithm: A guided tour. In *Computer Intensive Methods in Control and Signal Processing*. Boston: Birkhäuser, pp. 209–22.

Dalkey, Norman C. 1969. *The Delphi Method: An Experimental Study of Group Opinion*. Technical Report. Santa Monica: RAND Corporation.

Danielsson, Jon, Paul Embrechts, Charles Goodhart, Con Keating, Felix Muennich, Olivier Renault, and Hyun Song Shin. 2001. *An Academic Response to Basel II*. Zurich: FMG.

Delbecq, Andre L., Andrew H. Van de Ven, and David H. Gustafson. 1975. *Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes*. Glenview: Scott Foresman.

Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* 39: 1–22.

Efron, Bradley. 2008. Microarrays, empirical bayes and the two-groups model. *Statistical Science* 23: 1–22.

Eggstaff, Justin W., Thomas A. Mazzuchi, and Shahram Sarkani. 2014a. The development of progress plans using a performance-based expert judgment model to assess technical performance and risk. *Systems Engineering* 17: 375–91. [CrossRef]

Eggstaff, Justin W., Thomas A. Mazzuchi, and Shahram Sarkani. 2014b. The effect of the number of seed variables on the performance of cooke's classical model. *Reliability Engineering & System Safety* 121: 72–82.

Elguebaly, Tarek, and Nizar Bouguila. 2014. Background subtraction using finite mixtures of asymmetric gaussian distributions and shadow detection. *Machine Vision and Applications* 25: 1145–62. [CrossRef]

Embrechts, Paul, Giovanni Puccetti, Ludger Rüschendorf, Ruodu Wang, and Antonela Beleraj. 2014. An academic response to basel 3.5. *Risks* 2: 25–48. [CrossRef]

Everitt, Brian S., and David J. Hand. 1981. Finite mixture distributions. In *Monographs on Applied Probability and Statistics*. London and New York: Chapman and Hall.

Flandoli, Franco, Enrico Giorgi, William P. Aspinall, and Augusto Neri. 2011. Comparison of a new expert elicitation model with the classical model, equal weights and single experts, using a cross-validation technique. *Reliability Engineering & System Safety* 96: 1292–310.

Föllmer, Hans, and Alexander Schied. 2010. Convex and Coherent Risk Measures. *Journal of Quantitative Finance* 12. Available online: https://www.researchgate.net/publication/268261458_Convex_and_coherent_risk_measures (accessed on 8 June 2021)

Föllmer, Hans, and Alexander Schied. 2016. *Stochastic Finance: An Introduction in Discrete Time*. Berlin: De Gruyter.

Forgy, Edward W. 1965. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics* 21: 768–69.

French, Simon. 1983. *Group Consensus Probability Distributions: A Critical Survey*. Manchester: University of Manchester.

Fung, Tsz Chai, George Tzougas, and Mario Wuthrich. 2021. Mixture composite regression models with multi-type feature selection. *arXiv* arXiv:2103.07200.

Gambacciani, Marco, and Marc S. Paolella. 2017. Robust normal mixtures for financial portfolio allocation. *Econometrics and Statistics* 3: 91–111. [CrossRef]

Genest, Christian. 1992. Vincentization revisited. *The Annals of Statistics* 20: 1137–42. [CrossRef]

Genest, Christian, and Kevin J. McConway. 1990. Allocating the weights in the linear opinion pool. *Journal of Forecasting* 9: 53–73. [CrossRef]

Genest, Christian, Samaradasa Weerahandi, and James V. Zidek. 1984. Aggregating opinions through logarithmic pooling. *Theory and decision* 17: 61–70. [CrossRef]

Gneiting, Tilmann, and Adrian E Raftery. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association* 102: 359–78. [CrossRef]

Government Office for Science. 2020. *Transparency Data List of Participants of Sage and Related Sub-Groups*. London: Government Office for Science.

Grün, Bettina, and Friedrich Leisch. 2008. Finite mixtures of generalized linear regression models. In *Recent Advances in Linear Models and Related Areas*. Heidelberg: Physica-Verlag, pp. 205–30.

Henry, Marc, Yuichi Kitamura, and Bernard Salanié. 2014. Partial identification of finite mixtures in econometric models. *Quantitative Economics* 5: 123–44. [CrossRef]

Hogarth, Robert M. 1977. Methods for aggregating opinions. In *Decision Making and Change in Human Affairs*. Dordrecht: Springer, pp. 231–55.

Hora, Stephen C., Benjamin R. Fransen, Natasha Hawkins, and Irving Susel. 2013. Median aggregation of distribution functions. *Decision Analysis* 10: 279–91. [CrossRef]

Johnson, Norman L., Samuel Kotz, and Narayanaswamy Balakrishnan. 1994. *Continuous Univariate Distributions. Models and Applications*, 2nd ed. New York: John Wiley & Sons.

Jose, Victor Richmond R., Yael Grushka-Cockayne, and Kenneth C. Lichtendahl, Jr. 2013. Trimmed opinion pools and the crowd's calibration problem. *Management Science* 60: 463–75. [CrossRef]

Kaplan, Stan. 1992. 'expert information'versus 'expert opinions'. another approach to the problem of eliciting/combining/using expert knowledge in pra. *Reliability Engineering & System Safety* 35: 61–72.

Karlis, Dimitris, and Evdokia Xekalaki. 1999. Improving the em algorithm for mixtures. *Statistics and Computing* 9: 303–7. [CrossRef]

Karlis, Dimitris, and Evdokia Xekalaki. 2005. Mixed poisson distributions. *International Statistical Review/Revue Internationale de Statistique* 73: 35–58. [CrossRef]

Koksalmis, Emrah, and Özgür Kabak. 2019. Deriving decision makers' weights in group decision making: An overview of objective methods. *Information Fusion* 49: 146–60. [CrossRef]

Lambrigger, Dominik D., Pavel V. Shevchenko, and Mario V Wüthrich. 2009. The quantification of operational risk using internal data, relevant external data and expert opinions. *arXiv* arXiv:0904.1361.

Lichtendahl, Kenneth C., Yael Grushka-Cockayne, and Robert L. Winkler. 2013. Is it better to average probabilities or quantiles? *Management Science* 59: 1594–611. [CrossRef]

Linstone, Harold A., and Murray Turoff. 1975. *The Delphi Method*. Reading: Addison-Wesley Reading.

MacQueen, James. 1967. Some methods for classification and analysis of multivariate observations. Paper presented at the fifth Berkeley Symposium on Mathematical Statistics and Probability, Oakland, CA, USA, January 1; Volume 1, pp. 281–97.

Maitra, Ranjan. 2009. Initializing partition-optimization algorithms. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 6: 144–57. [CrossRef] [PubMed]

McLachlan, Geoffrey J., and David Peel. 2000a. *Finite Mixture Models*. New York: John Wiley & Sons.

McLachlan, Geoffrey J., and David Peel. 2000b. *Finite Mixture Models*. New York: Wiley Series in Probability and Statistics.

McLachlan, Geoffrey J., and Kaye E. Basford. 1988. *Mixture Models: Inference and Applications to Clustering*. New York: M. Dekker, Volume 38.

McLachlan, Geoffrey J., Kim-Anh Do, and Christophe Ambroise. 2005. *Analyzing Microarray Gene Expression Data*. Hoboken: John Wiley & Sons, Volume 422.

McLachlan, Geoffrey J., Sharon X. Lee, and Suren I. Rathnayake. 2019. Finite mixture models. *Annual Review of Statistics and Its Application* 6: 355–78. [CrossRef]

McNichols, Maureen. 1988. A comparison of the skewness of stock return distributions at earnings and non-earnings announcement dates. *Journal of Accounting and Economics* 10: 239–73. [CrossRef]

Mengersen, Kerrie L., Christian Robert, and Mike Titterington. 2011. *Mixtures: Estimation and Applications*. Chichester: John Wiley & Sons, Volume 896.

Miljkovic, Tatjana, and Bettina Grün. 2016. Modeling loss data using mixtures of distributions. *Insurance: Mathematics and Economics* 70: 387–96. [CrossRef]

Newcomb, Simon. 1886. A generalized theory of the combination of observations so as to obtain the best result. *American journal of Mathematics* 8: 343–66. [CrossRef]

Oboh, Bromensele Samuel, and Nizar Bouguila. 2017. Unsupervised learning of finite mixtures using scaled dirichlet distribution and its application to software modules categorization. Paper presented at the 2017 IEEE International Conference on Industrial Technology (ICIT), Toronto, ON, Canada, March 22–25; pp. 1085–90.

O'Hagan, Anthony, Caitlin E. Buck, Alireza Daneshkhah, J. Richard Eiser, Paul H. Garthwaite, David J. Jenkinson, Jeremy E. Oakley, and Tim Rakow. 2006. *Uncertain Judgements: Eliciting Experts' Probabilities*. Chichester: John Wiley & Sons.

Parenté, Frederik J., and Janet K. Anderson-Parenté. 1987. Delphi inquiry systems. *Judgmental Forecasting* 129–56.

Pearson, Karl. 1894. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A* 185: 71–110.

Peiro, Amado. 1999. Skewness in financial returns. *Journal of Banking & Finance* 23: 847–62.

Plous, Scott. 1993. *The Psychology of Judgment and Decision Making.* New York: Mcgraw-Hill Book Company.
Punzo, Antonio, Angelo Mazza, and Antonello Maruotti. 2018. Fitting insurance and economic data with outliers: a flexible approach based on finite mixtures of contaminated gamma distributions. *Journal of Applied Statistics* 45: 2563–84. [CrossRef]
Rufo, M. J., C. J. Pérez, and Jacinto Martín. 2010. Merging experts' opinions: A bayesian hierarchical model with mixture of prior distributions. *European Journal of Operational Research* 207: 284–89. [CrossRef]
Samadani, Ramin. 1995. A finite mixtures algorithm for finding proportions in sar images. *IEEE Transactions on Image Processing* 4: 1182–86. [CrossRef] [PubMed]
Schlattmann, Peter. 2009. *Medical Applications of Finite Mixture Models.* Berlin/Heidelberg: Springer.
Shevchenko, Pavel V., and Mario V. Wüthrich. 2006. The structural modelling of operational risk via bayesian inference: Combining loss data with expert opinions. *The Journal of Operational Risk* 1: 3–26. [CrossRef]
Stone, Mervyn. 1961. The linear opinion pool. *The Annals of Mathematical Statistics* 32: 1339–42. [CrossRef]
Titterington, D. M. 1990. Some recent research in the analysis of mixture distributions. *Statistics* 21: 619–41. [CrossRef]
Titterington, D. Michael, Adrian F. M. Smith, and Udi E. Makov. 1985. *Statistical Analysis of Finite Mixture Distributions*. Chichester: Wiley.
Tzougas, George, Spyridon Vrontos, and Nicholas Frangos. 2014. Optimal bonus-malus systems using finite mixture models. *Astin Bulletin* 44: 417–44. [CrossRef]
Tzougas, George, Spyridon Vrontos, and Nicholas Frangos. 2018. Bonus-malus systems with two-component mixture models arising from different parametric families. *North American Actuarial Journal* 22: 55–91. [CrossRef]
Vincent, Stella Burnham. 1912. *The functions of the Vibrissae in the Behavior of the White Rat.* Chicago: University of Chicago, Volume 1.
Wallach, Michael A., Nathan Kogan, and Daryl J. Bem. 1962. Group influence on individual risk taking. *The Journal of Abnormal and Social Psychology* 65: 75. [CrossRef]
Wedel, Michel, and Wayne S. DeSarbo. 1994. A review of recent developments in latent class regression models. In *Advanced Methods of Marketing Research*. Edited by Richard Bagozzi. Oxford: Blackwell Publishing Ltd., pp. 352–88.
Winkler, Robert L. 1968. The consensus of subjective probability distributions. *Management Science* 15: B-61. [CrossRef]
Winkler, Robert L., Javier Munoz, José L. Cervera, José M. Bernardo, Gail Blattenberger, Joseph B. Kadane, Dennis V. Lindley, Allan H. Murphy, Robert M. Oliver, and David Ríos-Insua. 1996. Scoring rules and the evaluation of probabilities. *Test* 5: 1–60. [CrossRef]
Yung, Yiu-Fai. 1997. Finite mixtures in confirmatory factor-analysis models. *Psychometrika* 62: 297–330. [CrossRef]

5.2 Supplementary material

In our study, for expository purposes of our suggested method, we generate the data ourselves thus there is not ambiguity regarding which model is more appropriate for the data in hand. Nevertheless, if the introduced method were to be applied to real data, there may be multiple competing models out of which we need to choose the most "appropriate" one. Usually, the fitness of these models alternatives is judged based on several criteria such as the global deviance (DEV), Generalized Akaike Information Criterion (GAIC), and Bayesian Information Criterion (BIC) which are defined below as follows.

Definition (Global deviance).

$$\hat{D} = -2\log(\hat{L})$$

where $\hat{D}$ is the fitted DEV and L denotes the probability of observing the data sample.

Definition (Generalized Akaike Information Criterion).

$$AIC = \hat{D} + 2df$$

where df are the degrees of freedom which can be defined as the number of fitted parameters in the model.

Definition (Bayesian Information Criterion).

$$BIC = \hat{D} + \log(n)df$$

where n is the number of independent observations.

It is evident that both AIC and BIC criteria add a loading on $\hat{D}$ which is generally equal to the degrees of freedom multiplied by a quantity whilst their difference lies on this quantity. The basis for both AIC and BIC is highly theoretical and each of them have a different goal. For instance, as Kuha (2004) mentions, BIC implies a "penalty" on large models and it may be more appropriate criterion if the aim is to find which model out of those examined is more probable to be true whilst with AIC there is no assumption that there must be an identifiable model. For a good overview of some aspects of model selection criteria we direct the interesting reader to the works of Sclove (1987) and Stone (1979).

A further extension of our work is for the components of the mixture distribution to be coming from different parametric families than those exhibited already. For instance, we could consider that for one of the components of the mixture distribution, individual claim sizes $X_{i,k}$ come from the Weibull distribution which can be parameterised as follows according to Johnson et al. (1994)

$$f(x) = \frac{sx^{s-1}}{m^s} e^{-\left(\frac{x}{m}\right)^s}$$

where $X_{i,k} > 0$ where $m > 0$ and $s > 0$. The mean and variance of $X_{i,k}$ are given by $\mathbb{E}(X_{i,k}) = m\Gamma\left(\frac{1}{s} + 1\right)$ and $\text{Var}(X_{i,k}) = m^2\{\Gamma\left(\frac{2}{s} + 1\right) - \left[\Gamma\left(\frac{1}{s} + 1\right)\right]^2\}$ respectively where m is a scale parameter which also has an effect on the mean of $X_{i,k}$. It is worth mentioning that for $s = 1$, the Weibull distribution reduces to the exponential distribution whilst for $s = 2$ to the Rayleigh distribution. Finally, the reason why we suggest the Weibull distribution is because it is a beneficial option for modelling truncated data, i.e. a data type which is seen extensively across the non-life insurance sector especially in the case of excess of loss reinsurance treaty contracts.

On a separate note, we remark on the fact that in our application we considered finite mixture models with two components for expository purposes. We acknowledge that in real life applications, we expect that the number of components, as well as the experts, is more realistic to be greater than two. Whilst this is computationally feasible, it means that the complexity of the problem naturally increases. The degree at which complexity rises cannot be quantified easily as it depends on many parameters such as the type and volume of data in hand, and the power of the computers owned by the company which would like to solve the opinions combination problem. That said, there is a way to reduce the computational times by using parallel computing, see Ferrall (2005) for an application of the later to finite mixture models.

In Chapter 6, we examine at a general level the topic of multivariate data with dependence in non-life insurance.

Chapter 6

On modelling multivariate data with dependence in non-life insurance

Nowadays, insurance companies often underwrite several lines of insurance contracts, whose claim counts and size may exhibit some dependence structures. As a result, the actuarial modelling exercise may involve multivariate data. In Chapter 6, we present some situations indicated by recent actuarial literature in which multivariate data modelling with dependence is of particular importance for rate-making and claim prediction purposes.

As a starting point, it is worth mentioning that actuarial pricing or claims modelling in insurance is usually made under the assumption that losses arising out of different coverage types are independent. Nevertheless, there are several circumstances when it would be beneficial to relax such an assumption when modelling jointly different types of claims and their associated counts.

In what follows, we briefly present some relevant studies which are classified per reason why such joint modelling methodologies are important in the current non-life insurance landscape. That said, some of these previous works are general enough to be used for multiple applications, thus we note that the classification that we deploy here relies mostly on the examples that the authors of these works have used with respect to the application potential of their approaches.

6.1 Multiple types of coverage under a single policy for one insured

We recognise that sometimes multiple types of coverage may exist under the same insurance policy for a given policyholder. If a qualifying insurance event happens, the insurer may have to deal with more than one claim types under a single policy for the same insured. A prime example of an insurance business class where this problem can arise is motor. There, a car crash may result in an insurer having to pay both for claims towards fixing the damaged vehicle and potential bodily injury claims arising out of the accident. Another instance is property insurance as the total sum insured is usually divided among three different elements, i.e. building, contents, and business interruption values. This means that if, as an illustration, a fire is a qualifying event for the insurance to be activated, and we assume that the property is used for commercial purposes, the insured may claim for any of the aforementioned coverage types.

One of the works addressing the aforementioned issue is of Bermúdez & Karlis (2011) which challenges the assumption that different claims types are independent and in doing so several multivariate Poisson regression models are introduced, namely the common covariance model and the full covariance model alike which are then extended by their zero inflated variants. The aforementioned models are fitted by using a Bayesian approach. The reasons behind such choice is to ease the estimation process and allow to derive posterior quantities of interest which are not merely point estimates but they are also accompanied with their posterior distribution. A real data application showcases that the proposed approaches are insightful for rate-making purposes in non-life insurance.

Along similar lines regarding the recognition of dependence structures existing between different types of claims, Bermúdez & Karlis (2012) employ bivariate Poisson (BP) regression models for premium derivation in motor insurance taking into account correlation between two different types of claims, i.e. third party liability versus all other car insurance claims types which are often observed. It is shown that even a small correlation between these claims types can lead to considerably higher premiums as opposed to those that would have been derived if the dependence structure would not have been captured.

Furthermore, an interesting and slightly different approach comes from Abdallah et al. (2016) having as motivation the fact that claims that have happened lately as opposed to more past ones may be more predictive of future claims. It is recognised that the multivariate negative binomial distribution, whilst being a famous modelling option for panel data, does not allow for imposing weights on past claims based on time. Whilst there are models that can capture this time element, these usually require very intricate numerical procedures for estimating the parameters and the whole modelling procedure becomes even more complicated when one also considers the dependence that exist between multiple types of claims. That said, a bivariate dynamic distribution for claim counts is developed using as basis random effects originating from the Sarmanov family of multivariate distributions and an approximation of the posterior distribution of such random effects is recommended as a way of retrieving a dynamic distribution based on such types of bivariate priors. The proposed model is then applied on real insurance claims data exhibiting the flexibility of the method in deriving predictive premiums.

Furthermore, Bermúdez & Karlis (2017) extend various bivariate Poisson regression models which have been considering correlation between different claims types to deduce premiums on an a priori basis, to now account for posteriori, i.e. experience based, rating which also accounts for dependence between different types of claims. Then, two bivariate posteriori rating models are introduced to obtain posterior premiums and a posteriori risk factors. By applying the models on a real motor insurance claims data set, it is found that the deduced posteriori risk factors are considerably lower than those factors deduced under the independence assumption.

Moreover, Bermúdez et al. (2018) introduces a novel approach recognising that in the process of rate making, an insurer may want to take into account simultaneously both time and cross dependence between different types of policies or claim types rather than treating these elements as two distinct cases. This is important as the correlation may exist not only between different coverage types for a given policyholder but also in between observations of the same policyholder experienced over time impacting the rates charged by the insurer. The author captures this effect by developing a bivariate INAR(1) regression model which is followed by a numerical application in motor insurance field and finds that the suggested approach leads to better performance compared to simpler models which do not take into account simultaneously the elements of time and cross correlation.

Finally, Bolancé & Vernic (2019) analyses three models which rely on the multivariate Sarmanov distribution to account for inter-dependencies across various claims data. The work uses real motor and home insurance data on which are fitted three trivariate Sarmanov distributions with generalised linear models (GLMs) for marginals. The parameters of the presented models are estimated using a method to approach the the maximum likelihood (ML) estimators. A numerical study based on real motor and home insurance data indicates that there is a positive and long run dependence between the accident frequency in auto and home insurance classes of business.

6.2 Multiple types of coverage for multiple perils under a single policy for one insured

An additional categorisation of the literature involve modelling methods accounting for a policyholder who buys an insurance policy which includes multiple types of coverage but for multiple perils. That said, this is a fairly often phenomenon within property insurance where the total insured value for a given location may be split among the sum of buildings, contents, and business interruption insured values, as we have also seen in Chapter 2, Section 2.3.2.2.

Under such circumstances, claims may arise if a loss is experienced in any of the aforementioned coverage types. However, as an example, the building element may be covered against earthquake, fire, or hurricane losses whilst for the contents element may exclude earthquake related losses. In this context, an interesting approach is coming from Jeong & Dey (2021) where a shared random effects model is presented to account for the unobserved heterogeneity across different types of claims considering also the correlation among the claims from multiple perils.

6.3 Multiple types of coverage under a single policy for multiple insureds in the same household

An actuary, who is pricing an insurance policy with several types of coverage for a given insured, may want to take into account that other members of their same household may have bought an insurance policy too. This information is important because dependence structures may exist between these "seemingly" independent policies.

The research studies of Pechon et al. (2018) and Pechon et al. (2019) acknowledge that there may be correlation between different coverage types bought by different members of the same household. In the light of big data and an ongoing trend for customising insurance products, it is suggested that there is a need to build multivariate count models that take into account such inter-dependencies. In particular, this work aims to bring this household-wide perspective in the actuarial pricing. Similarly, Pechon et al. (2021) proposes a multivariate Poisson mixture, with random effects correlated using a hierarchical structure, capturing the correlation between unobserved risk factors across home and motor insurance and among insureds from the same household allowing to identify which households should be considered as "riskier" by the insurer.

6.4 Insurance bundling

Another trend in non-life insurance is the offering of multiple insurance contracts bundled together by an insurer to meet several coverage requirements of a policyholder. Several advantages to the insured and the insurer related to bundled insurance products are provided in Chapter 7, Section 7.1.

Looking at the literature, Shi & Valdez (2014) addressed the need for multivariate count models taking into account the dependency among different claims types when pricing bundled insurance policies. In this context, various approaches in developing multivariate count models using the negative binomial distribution are explored. It is acknowledged that, traditionally, common shock variables could be used for introducing correlation, however such a choice would rely on the Negative Binomial type I distribution and it would not be appropriate for dispersion modelling.

To resolve such limitations, it is suggested to use copulas for the modelling multivariate claim counts. Within this perspective two approaches are explored. Firstly, a mixture max-id copulas applied on discrete count data to account for pair-wise association in addition to tail and global dependence. Secondly, elliptical copulas which regard unstructured dependence and both positive and negative correlation between different claim types. The suggested models are compared to the common shock model in an application using real motor insurance data and results indicate a better performance for the copula-based models.

6.5 Data analytics-based pricing

In the era of big data, insurers may want to utilise the wealth of information that they can nowadays gather about their policyholders from various sources in order to produce more fair premiums and expand their business line offerings.

With this in mind, Denuit et al. (2019) examines the topic of Pay-How-You-Drive (PHYD) or Usage-Based (UB) systems for automobile insurance where actuaries are provided with behavioural risk factors of the policyholders, such as the time of the day in which they commute and the average speed in which they drive among other driving patterns. Normally these data are collected with the assistance of telematic devices located in policyholders' automobiles and it is beneficial for actuarial purposes for such data to be incorporated for rate making purposes. Towards this direction, multivariate mixed models are introduced for the description of the joint dynamics of telematics data and claim frequencies.

Further to the aforementioned literature review, we transition to Chapter 7, where our contribution entitled "The Multivariate Poisson-Generalized Inverse Gaussian Claim Count Regression Model with Varying Dispersion and Shape" is presented.

Chapter 7

The Multivariate Poisson-Generalized Inverse Gaussian Claim Count Regression Model with Varying Dispersion and Shape

This chapter is primarily dedicated to our article entitled "The Multivariate Poisson-Generalized Inverse Gaussian Claim Count Regression Model with Varying Dispersion and Shape", which is currently under minor revision in a peer reviewed journal. The article is presented in the exact format in which it has been submitted for review. Some supplementary material is included at the end.

7.1 The Multivariate Poisson-Generalized Inverse Gaussian Claim Count Regression Model with Varying Dispersion and Shape

The Multivariate Poisson-Generalized Inverse Gaussian Claim Count Regression Model with Varying Dispersion and Shape

George Tzougas[1,*] and Despoina Makariou[2]

[1] Department of Actuarial Mathematics and Statistics, Heriot-Watt University, UK

[2] Department of Statistics, London School of Economics and Political Science, UK

May 3, 2022

We introduce a multivariate Poisson-Generalized Inverse Gaussian regression model with varying dispersion and shape for modelling different types of claims and their associated counts in non-life insurance. The multivariate Poisson-Generalized Inverse Gaussian regression model is a general class of models which, under the approach adopted herein, allows us to account for overdispersion and positive correlation between the claim count responses in a flexible manner. For expository purposes, we consider the bivariate Poisson-Generalized Inverse Gaussian with regression structures on the mean, dispersion, and shape parameters. The model's implementation is demonstrated by using bodily injury and property damage claim count data from a European motor insurer. The parameters of the model are estimated via the Expectation-Maximization algorithm which is computationally tractable and is shown to have a satisfactory performance.

Keywords: Multivariate Poisson-Generalized Inverse Gaussian Distribution; EM Algorithm; Regression Models for the Marginal Means, Dispersion, and Shape Parameters; Bonus-Malus Premiums; Non-Life Insurance

1 Introduction

The regression analysis of multivariate count data for capturing the dependence structures between multiple count response variables based on explanatory variables is encountered across several disciplines such as biology, biometrics, genetics, medicine, marketing, ecology, sociology, econometrics, and insurance. In general, multivariate count

*Heriot Watt, Edinburgh, EH14 4AS, Email: G.Tzougas@hw.ac.uk

data models can be classified into the following three classes: multivariate Poisson models, multivariate mixed Poisson (MVMP) models, and copula-based models. For more details, the interested reader can refer to the works of M'Kendrick (1925), Stein and Juritz (1987), Stein et al. (1987), Kocherlakota (1988), Aitchison and Ho (1989), Jung and Winkelmann (1993), Joe (1997), Johnson et al. (1997), Krummenauer (1998), Lakshminarayana et al. (1999), Lee (1999), Munkin and Trivedi (1999), Gurmu and Elder (2000), Chib and Winkelmann (2001), Ho and Singer (2001), Kocherlakota and Kocherlakota (2001), Cameron et al. (2004), Karlis and Meligkotsidou (2005), Zimmer and Trivedi (2006), Genest and Nešlehová (2007), Park and Lord (2007), Ma et al. (2008), Winkelmann (2008), Aguero-Valverde and Jovanis (2009), El-Basyouny and Sayed (2009), Famoye (2010), Nikoloulopoulos and Karlis (2010), Ghitany et al. (2012), Cameron and Trivedi (2013), Nikoloulopoulos (2013), Rüschendorf (2013), Zhan et al. (2015), Marra and Wyszynski (2016), Nikoloulopoulos (2016), Chen and Hanson (2017), Silva et al. (2019), and Chiquet et al. (2020).

In a non-life insurance setting, the actuary may be concerned with modelling jointly different types of claims and their associated counts. In this market segment, there are several circumstances where the interest lies in developing models which can accommodate for positively correlated claims whilst accounting for overdispersion which is a direct consequence of unobserved heterogeneity due to systematic effects in the data. Furthermore, these dependence structures between different claim types may be observed within the same insurance policy, such as property damage and bodily injury claims in motor third party liability (MTPL) insurance, or in alternative types of coverage, such as home and auto insurance, bundled together under a single policy. Regarding the latter, some of the advantages for the policyholder are multi-product premium discounts, straightforward tracking of policy renewal dates, easy claims reporting, and a more "personal" relationship between the insured and their insurer where the latter closely identify their needs and mitigate possible insurance coverage gaps. From the insurer's perspective though, bundling multiple types of insurance for the same policyholder translates into a need to develop predictive models which can efficiently capture the joint dynamics of different claims types associated with various insurance business lines. With regards to the use of alternative multivariate count models in non-life insurance, see for instance, Bermúdez and Karlis (2011), Bermúdez and Karlis (2012), Shi and Valdez (2014a), Shi and Valdez (2014b), Abdallah et al. (2016), Bermúdez and Karlis (2017), Bermúdez et al. (2018), Pechon et al. (2018), Pechon et al. (2019), Bolancé and Vernic (2019), Denuit et al. (2019), Fung et al. (2019), Bolancé et al. (2020), Pechon et al. (2021), Jeong and Dey (2021), Gómez-Déniz and Calderín-Ojeda (2021) and Tzougas and di Cerchiara (2021).

In the current study, we develop a multivariate Poisson-Generalised Inverse Gaussian (MVPGIG) regression model with varying dispersion and shape for modelling positively correlated and overdispersed claim counts from different types of coverage in a flexible manner. In particular, within the adopted modelling framework, in addition to the marginal mean parameters, which are traditionally modelled using risk factors, regressors are allowed on the dispersion and shape parameters. The proposed approach allows us to model the skewness and kurtosis of the model explicitly as a function of the explanatory variables for the mean, dispersion and shape parameters. Instead, if only the mean parameter is modelled in terms of explanatory variables then this can lead to a misclassification of policyholders with a high number of claims due to the unobserved heterogeneity changes with covariates.Furthermore, the MVPGIG, is a broad family of models including many MVMP models considered in the aforementioned literature ones as special and/or limiting cases, such as, for example, the multivariate Negative Binomial (MVNB), or multivariate Poisson-Gamma, multivariate Poisson-Inverse Gaussian (MVPIG), multivariate Poisson–Inverse Exponential, multivariate Poisson–Inverse Chi Squared, and multivariate Poisson–Scaled Inverse Chi Squared distributions, depending on the estimated values of the dispersion and shape parameters which are modelled based on covariate information, hence enabling us to account for the tail behaviour of observed data in versatile manner. The latter can be regarded as an important property for capturing overdispersion since this phenomenon is not necessarily attributed to an excess of zeros but it may be also caused by a heavy tail in the claim count data, see Shared (1980). For illustrative purposes, the bivariate Poisson-Generalised Inverse Gaussian (BPGIG) regression model with varying dispersion and shape is fitted on Motor Third Party Liability (MTPL) insurance bodily injury and property damage claim count data using a novel Expectation-Maximization (EM) type algorithm. The proposed maximum likelihood (ML) estimation scheme takes advantage of the stochastic mixture representation of the BPGIG model in order to reduce the problem of maximizing its cumbersome likelihood function which is expressed in terms of the modified Bessel function of the third kind to the simple problem of maximising the likelihood function of its mixing densivty.

The remainder of this article is organized, as follows: Section 2 deals with the construction of the proposed MVPGIG regression model with varying dispersion and shape parameters. In Section 3, we describe the ML estimation procedure for the BPGIG model via the EM algorithm. A real data application based on the two dimensional MTPL data set is presented in Section 4 and the fitting performance of the BPGIG regression model with varying dispersion and shape parameters is compared to that of the bivariate Negative Binomial (BNB) and Poisson-Inverse Gaussian (BPIG) regression models with varying dispersion of Tzougas and Pignatelli di Cerchiara (2021) that we use as a benchmark for comparison. In Section 5, the a posteriori, or Bonus-Malus, premiums determined by the from the BNB and BPIG models are compared to those resulting from the proposed BPGIG model using the expected value principle. Finally, concluding remarks can be found in Section 6.

2 The Multivariate Poisson-Generalized Inverse Gaussian Regression Model with Varying Dispersion and Shape Parameters

Consider that in a non-life insurance policy of an insured j, where $j = 1, ..., n$, we observe multi-peril claim frequencies $K_{i,j}$, for $i = 1, ..., m$ types of coverage. Assume that given the random variables $Z_j > 0$, $K_{i,j}|Z_j$ per claim type i are distributed according to a Poisson distribution with probability mass function (pmf) given by

$$P\left(k_{i,j}|z_j\right) = \frac{\exp[-\mu_{i,j} z_j](\mu_{i,j} z_j)^{k_{i,j}}}{k_{i,j}!}, \tag{1}$$

for $k_{i,j} = 0, 1, 2, 3, ...,$ where $\mu_{i,j} > 0$, with mean and variance $\mathbb{E}(k_{i,j}|z_j) = \mu_{i,j} z_j$ and $\text{Var}(k_{i,j}|z_j) = \mu_{i,j} z_j$.

Also, suppose that Z_j are random variables from a Generalized–Inverse Gaussian (GIG) distribution with probability density function (pdf) given by

$$g(z_j; \sigma_j, \nu_j) = \frac{c_j^{\nu_j}}{2K_{\nu_j}\left(\frac{1}{\sigma_j}\right)} z_j^{\nu_j - 1} \exp\left[-\frac{1}{2\sigma_j}(c_j z_j + \frac{1}{c_j z_j})\right], \tag{2}$$

for $\sigma_j > 0$ and $-\infty < \nu_j < \infty$, where $c_j = \frac{K_{\nu_j+1}(\sigma_j^{-1})}{K_{\nu_j}(\sigma_j^{-1})}$ and

$$K_{\nu_j}(\omega) = \int_0^\infty x^{\nu_j - 1} \exp\{-\frac{1}{2}\omega(x + \frac{1}{x})\}dx$$

is the modified Bessel function of the third kind of order ν_j and argument ω. This parameterization ensures that the model is identifiable since $\mathbb{E}(Z_j) = 1$. Furthermore, note that

$$Var(Z_j) = \frac{K_{\nu_j+2}(\frac{1}{\sigma_j})K_{\nu_j}(\frac{1}{\sigma_j})}{K_{\nu_j+1}(\frac{1}{\sigma_j})^2} - 1.$$

Thus, considering the assumptions in Eqs (1 and 2) it is easy to see that the unconditional distribution of $K_{i,j}$ will be a multivariate Poisson-Generalized Inverse Gaussian (MVPGIG) distribution with joint probability mass function (jpmf) given by

$$P(k_{1,j}, k_{2,j}, ..., k_{m,j}) = \frac{\prod_{i=1}^{m} \mu_{i,j}^{k_{i,j}}}{\prod_{i=1}^{m} k_{i,j}!} \frac{c_j^{\nu_j}}{2K_{\nu_j}(\frac{1}{\sigma_j})} \left[(2\sum_{i=1}^{m} \mu_{i,j} + \frac{c_j}{\sigma_j})c\sigma_j\right]^{-\frac{\sum_{i=1}^{m} k_{i,j} + \nu_j}{2}} 2K_{\sum_{i=1}^{m} k_{i,j} + \nu_j}\left[\sqrt{\frac{1}{c\sigma_j}(2\sum_{i=1}^{m} \mu_{i,j} + \frac{c}{\sigma_j})}\right]. \tag{3}$$

Note that if we let $\nu_j = -0.5$ in Eq. (3) the MVPGIG distribution reduces to a multivariate Poisson-Inverse Gaussian (MVPIG) distribution. Further, the multivariate Negative Binomial (MVNBB) distribution is a limiting case of Eq. (3), obtained by letting $\sigma_j \to \infty$ for $\nu_j > 0$ and $\nu_j < -1$ respectively.

Henceforth, for expository purposes, we will restrict attention to the bivariate case m=2. We assume that the mean, dispersion and shape parameters of the bivariate Poisson-Generalized Inverse Gaussian (BPGIG) are modelled as functions of explanatory variables with parametric linear functional forms:

$$\mu_{1,j} = \exp(\boldsymbol{x_{1,j}^T \beta_1}), \tag{4}$$

$$\mu_{2,j} = \exp(\boldsymbol{x}_{\mathbf{2},\boldsymbol{j}}^T \boldsymbol{\beta}_{\mathbf{2}}), \tag{5}$$

$$\sigma_j = \exp(\boldsymbol{x}_{\mathbf{3},\boldsymbol{j}}^T \boldsymbol{\beta}_{\mathbf{3}}) \tag{6}$$

and

$$\nu_j = \boldsymbol{x}_{\mathbf{4},\boldsymbol{j}}^T \boldsymbol{\beta}_{\mathbf{4}}, \tag{7}$$

where $\boldsymbol{x}_{\mathbf{1},\boldsymbol{j}}$, $\boldsymbol{x}_{\mathbf{2},\boldsymbol{j}}$, $\boldsymbol{x}_{\mathbf{3},\boldsymbol{j}}$ and $\boldsymbol{x}_{\mathbf{4},\boldsymbol{j}}$ are vectors of covariates with dimensions $p_1 \times 1$, $p_2 \times 1$, $p_3 \times 1$ and $p_4 \times 1$ respectively, with $(\beta_{1,1}, ..., \beta_{1,p_1})^T$, $(\beta_{2,1}, ..., \beta_{2,p_2})^T$, $(\beta_{3,1}, ..., \beta_{3,p_3})^T$ and $(\beta_{4,1}, ..., \beta_{4,p_4})^T$ the corresponding parameter vectors and where it is considered that the matrices $\mathbf{X_1}$, $\mathbf{X_2}$, $\mathbf{X_3}$ and $\mathbf{X_4}$ with rows given by $x_{1,i}$, $x_{2,i}$, $x_{3,i}$ and $x_{4,i}$ respectively, are of full rank.

Finally, the following desirable properties ensure the flexibility of the proposed model for capturing overdispersion and accommodating for positive correlation structures[1] between the different claim count response variables.

1. The marginal distribution of $K_{i,j}$, for $i = 1, ..., m$ and $j = 1, ..., n$, is a Poisson-Generalized Inverse Gaussian, or Sichel, distribution. Also, the mean and the variance of $K_{i,j}$ are given by
$$\mathbb{E}\left(K_{i,j}\right) = \mu_{i,j} \tag{8}$$
and
$$\mathrm{Var}\left(K_{i,j}\right) = \mu_{i,j} + \mu_{i,j}^2 \left(c_j^{-2} + \frac{2(\nu_j + 1)}{c_j} \sigma_j - 1 \right). \tag{9}$$

2. Let $K_{i,j}|z_j$ and z_j, for $i = 1, ..., m$ and $j = 1, ..., n$, be distributed according to the Poisson and GIG distributions which are given Eqs (1 and 2) respectively. Also, consider that the cumulative generating function of z_j is denoted by $C_{_j}(t)$, then the cumulative generating function of the marginal distribution of $K_{i,j}$, which is denoted by $C_{K_{i,j}}(t)$, is given by
$$C_{K_{i,j}}(t) = C_{_j}[\mu_{i,j}(e^t - 1)] \tag{10}$$
and hence, since $_j$ has a unit mean, the third and fourth cumulants of $K_{i,j}$ and z_j are related by
$$C_{3K_{i,j}} = \mu_{i,j} + 3\mu_{i,j}^2 Var(z_j) + \mu_{i,j}^3 C_{3z_j}, \tag{11}$$
and

[1] A limitation of the proposed model is that it cannot allow for negative correlation between the claim count response variables. However, regarding MTPL data, such as those we use in this study, positive correlation between MTPL bodily injury and property damage claim counts is what we expect.

$$C_{4K_{i,j}} = \mu_{i,j} + 7\mu_{i,j}^2 Var(z_j) + 6\mu_{i,j}^3 C_{3z_j} + \mu_{i,j}^4 C_{4z_j}, \tag{12}$$

where $C_{3K_{i,j}}$ and $C_{4K_{i,j}}$ are the third and fourth cumulants of $K_{i,j}$.
The skewness and kurtosis of $K_{i,j}$ are $\sqrt{\beta_1} = \kappa_{3K_{i,j}}/[Var(K_{i,j})]^{1.5}$ and $\beta_2 = 3 + \{\kappa_{4K_{i,j}}/[Var(K_{i,j})^2]\}$ respectively, where the cumulants of the mixing distribution are given by

$$C_{3_j} = [g_2 - 3g_1] \tag{13}$$

and

$$C_{4\gamma} = (g_3 - 4g_2 + 6g_1 - 3g_1^2), \tag{14}$$

where
$g_1 = [1/c_j^2 + 2\sigma_j(\nu_j + 1)/c_j - 1]$,
$g_2 = 2\sigma_j(\nu_j + 2)/c_j^3 + [4\sigma_j^2(\nu_j + 1)(\nu_j + 2) + 1]/c_j^2 - 1$
$g_3 = [1 + 4\sigma_j^2(\nu_j + 2)(\nu_j + 3)]/c_j^4 + [8\sigma_j^3(\nu_j + 1)(\nu_j + 2)(\nu_j + 3) + 4\sigma(\nu_j + 2)]/c_j^3 - 1.$

3. The covariance (Cov) between $K_{1,j}$ and $K_{2,j}$ is given by

$$\text{Cov}\,(K_{1,j}, K_{2,j}) = \mu_{1,j}\mu_{2,j}\left(c_j^{-2} + \frac{2(\nu_j + 1)}{c_j}\sigma_j - 1\right). \tag{15}$$

3 Statistical Inference: The EM Algorithm

In this section, an Expectation-Maximization (EM) type algorithm (Dempster et al., 1977; McLachlan and Krishnan, 2007) is developed to facilitate maximum likelihood (ML) estimation of the BPGIG regression model with varying dispersion and shape.

Furthermore, assume that $(k_{1,j}, k_{2,j}, \boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}}, \boldsymbol{x_{4,j}})$, $j = 1, ..., n$, is a sample of independent observations, where $K_{1,j}$ and $K_{2,j}$ are the claim count variables and $\boldsymbol{x_{1,j}}, \boldsymbol{x_{2,j}}, \boldsymbol{x_{3,j}}$ and $\boldsymbol{x_{4,j}}$ are the vectors of covariates with dimensions $p_1 \times 1$, $p_2 \times 1$, $p_3 \times 1$ and $p_4 \times 1$ respectively. Also, assume that Z_j, $j = 1, ..., n$, are the random effects which are non-observable and are considered to produce missing data. By augmentation of the unobserved Z_j one can write the complete log-likelihood as follows:

$$\begin{aligned} \ell_c(\boldsymbol{\theta}) \propto & \sum_{i=1}^{2}\sum_{j=1}^{n} [-\mu_{i,j} z_j + k_{i,j} \log(\mu_{i,j})] \\ & + \sum_{j=1}^{n} \left[\nu_j \log(c_j) + (\nu_j - 1)\log(z_j) - \log(K_{\nu_j}(\sigma_j^{-1})) - \frac{1}{2\sigma_j}\left(c_j z_j + \frac{1}{c_j z_j}\right)\right], \end{aligned} \tag{16}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4)$ is the vector of the parameters.

We present below the E- and the M-Steps of our EM type algorithm. At the E-Step, we compute the Q-function, which is the conditional expectation of the complete log-likelihood function given by Eq. (16), given $\boldsymbol{\theta}^r$, which is the estimated value of $\boldsymbol{\theta}$ at the r-th iteration. The M-Step consists in maximizing the Q-function. In particular, we want to find the updated parameters $\boldsymbol{\theta}^{r+1}$ such that the Q-function is increased with respect to $\boldsymbol{\theta}$.

- **E-Step:**

  The Q-function at the r-th iteration can be written as

$$
\begin{aligned}
Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}) &\equiv \mathbb{E}_z\left(\ell_c(\boldsymbol{\theta})|k_{1,j},k_{2,j};\boldsymbol{\theta}^{(r)}\right) \\
&\propto \sum_{i=1}^{2}\sum_{j=1}^{n}\left[-\mu_{i,j}^{(r)}w_{1,j}+k_{i,j}\log(\mu_{i,j}^{(r)})\right] \\
&+\sum_{j=1}^{n}\left[\nu_j^{(r)}\log(c_j^{(r)})+(\nu_j^{(r)}-1)w_{3,j}-\right. \\
&\left.\log(K_{\nu_j^{(r)}}(1/\sigma_j^{(r)}))-\frac{1}{2\sigma_j^{(r)}}\left(c_j^{(r)}w_{1,j}+\frac{w_{2,j}}{c_j^{(r)}}\right)\right],
\end{aligned}
\tag{17}
$$

  where we have defined the pseudo-values $w_{1,j}=\mathbb{E}_{z_j}[z_j|k_{i,j};\boldsymbol{\theta}^{(r)}]$, $w_{2,j}=\mathbb{E}_{z_j}[z_j^{-1}|k_{i,j};\boldsymbol{\theta}^{(r)}]$ and $w_{3,j}=\mathbb{E}_{z_j}[\log(z_j)|k_{i,j};\boldsymbol{\theta}^{\boldsymbol{r}}]$.

- **M-Step:**

  – Firstly, differentiate the $Q-$function with respect to $\boldsymbol{\beta_1}$:

$$
h_1\left(\boldsymbol{\beta_1}\right)=\frac{\partial Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right)}{\partial\beta_{1,l}}=\sum_{j=1}^{n}\left(k_{1,j}-\mu_{1,j}^{(r)}w_{1,j}\right)x_{1,j,l}, \tag{18}
$$

  and

$$
H_1\left(\boldsymbol{\beta_1}\right)=\frac{\partial^2 Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right)}{\partial\beta_{1,l}\partial\beta_{1,l}^T}=\sum_{j=1}^{n}\left(-\mu_{1,j}^{(r)}w_{1,j}\right)x_{1,j,l}x_{1,j,l}^T, \tag{19}
$$

  for $j=1,...,n$ and $l=1,...,p_1$.

  Then, the Newton-Raphson iterative algorithm for $\boldsymbol{\beta_1}$ is as follows:

$$
\boldsymbol{\beta_1}^{(r+1)}\equiv\boldsymbol{\beta_1}^{(r)}-\left[H_1\left(\boldsymbol{\beta_1}^{(r)}\right)\right]^{-1}h_1\left(\boldsymbol{\beta_1}^{(r)}\right). \tag{20}
$$

  – Secondly, differentiate the $Q-$function with respect to $\boldsymbol{\beta_2}$:

$$
h_2\left(\boldsymbol{\beta_2}\right)=\frac{\partial Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right)}{\partial\beta_{2,l}}=\sum_{j=1}^{n}\left(k_{2,j}-\mu_{2,j}^{(r)}w_{1,j}\right)x_{2,j,l}, \tag{21}
$$

  and

$$H_2\left(\beta_2\right) = \frac{\partial^2 Q\left(\theta;\theta^{(r)}\right)}{\partial\beta_{2,l}\partial\beta_{2,l}^T} = \sum_{j=1}^{n}\left(-\mu_{2,j}^{(r)} w_{1,j}\right) x_{2,j,l} x_{2,j,l}^T, \tag{22}$$

for $j = 1, ..., n$ and $l = 1, ..., p_2$.

Then, the Newton-Raphson iterative algorithm for $\boldsymbol{\beta_2}$ is as follows:

$$\boldsymbol{\beta_2}^{(r+1)} \equiv \boldsymbol{\beta_2}^{(r)} - \left[H_2\left(\boldsymbol{\beta_2}^{(r)}\right)\right]^{-1} h_2\left(\boldsymbol{\beta_2}^{(r)}\right). \tag{23}$$

– Thirdly, differentiate the $Q-$function with respect to $\boldsymbol{\beta_3}$:

$$\begin{aligned} h_3\left(\boldsymbol{\beta_3}\right) = \frac{\partial Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right)}{\partial\beta_{3,l}} \quad = \quad & \sum_{i=1}^{n}\sigma_j^{(r)}\left[\frac{\nu_j^{(r)}}{\sigma_j^{(r)}} - \frac{c_j^{(r)}}{\left(\sigma_j^2\right)^{(r)}}\right. \\ & + \frac{1}{2\left(\sigma_j^2\right)^{(r)}}\left(c_j^{(r)} w_{1,j} + \frac{w_{2,j}}{c_j^{(r)}}\right) + \frac{\nu_j^{(r)}}{c_j^{(r)}}\frac{\partial c_j^{(r)}}{\partial\sigma_j^{(r)}} \\ & \left. - \frac{1}{2\sigma_j^{(r)}}\frac{\partial c_j^{(r)}}{\partial\sigma_j^{(r)}}\left(w_{1,j} - \frac{w_{2,j}}{\left(c_j^2\right)^{(r)}}\right)\right] x_{3,j,l}, \end{aligned} \tag{24}$$

and

$$H_3(\boldsymbol{\beta}_3) = \frac{\partial^2 Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right)}{\partial\beta_{3,l}\partial\beta_{3,j}^T} =$$

$$\begin{aligned}
= & \sum_{i=1}^{n}\sigma_j^{(r)}\left\{\frac{\nu_j^{(r)}}{\sigma_j^{(r)}} - \frac{c_j^{(r)}}{\left(\sigma_j^2\right)^{(r)}} + \frac{1}{2\left(\sigma_j^2\right)^{(r)}}\left(c_j^{(r)}w_{1,j} + \frac{w_{2,j}}{c_j^{(r)}}\right)\right. \\
& + \frac{\nu_j^{(r)}}{c_j^{(r)}}\frac{\partial c_j^{(r)}}{\partial\sigma_j^{(r)}} - \frac{1}{2\sigma_j^{(r)}}\frac{\partial c_j^{(r)}}{\partial\sigma_j^{(r)}}\left(w_{1,j} - \frac{w_{2,j}}{\left(c_j^2\right)^{(r)}}\right) \\
& + \sigma_j^{(r)}\left[\left(\frac{\nu_j^{(r)}}{c_j^{(r)}} - \frac{1}{2\sigma_j^{(r)}}\left(w_{1,j} - \frac{w_{2,j}}{\left(c_j^2\right)^{(r)}}\right)\right)\frac{\partial^2 c_j^{(r)}}{\partial\left(\sigma_j^2\right)^{(r)}}\right. \\
& + \frac{1}{\left(\sigma_j^2\right)^{(r)}}\left(w_{1,j} - \frac{w_{2,j}}{\left(c_j^2\right)^{(r)}} - 1\right)\frac{\partial c_j^{(r)}}{\partial\sigma_j^{(r)}} \\
& - \left(\frac{\nu_j^{(r)}}{\left(c_j^2\right)^{(r)}} + \frac{w_{2,j}}{\left(c_j^3\right)^{(r)}\sigma_j^{(r)}}\right)\left(\frac{\partial c_j^{(r)}}{\partial\sigma_j^{(r)}}\right)^2 - \frac{\nu_j^{(r)}}{\left(\sigma_j^2\right)^{(r)}} \\
& \left.\left. + \frac{2c_j^{(r)}}{\left(\sigma_j^3\right)^{(r)}} - \frac{1}{\left(\sigma_j^3\right)^{(r)}}\left(c_j^{(r)}w_{1,j} + \frac{w_{2,j}}{c_j^{(r)}}\right)\right]\right\}x_{3,j,l}x_{3,j,l}^T,
\end{aligned} \tag{25}$$

for $j = 1, .., n$ and $l = 1, ..., p_3$, where

$$\frac{\partial c_j^{(r)}}{\partial\phi_j^{(r)}} = \frac{c_j^{(r)}(2\nu_j^{(r)} + 1)}{\phi_j^{(r)}} + \frac{1 - \left(c_j^2\right)^{(r)}}{\left(\phi_j^2\right)^{(r)}} \tag{26}$$

and where

$$\frac{\partial^2 c_j^{(r)}}{\partial\left(\phi_j^2\right)^{(r)}} = \left(\frac{2\nu_j^{(r)} + 1}{\phi_j^{(r)}} - \frac{2c_j^{(r)}}{\left(\phi_j^2\right)^{(r)}}\right)\frac{\partial c_j^{(r)}}{\partial\phi_j^{(r)}} - \frac{c_j^{(r)}(2\nu_j^{(r)} + 1)}{\left(\phi_j^2\right)^{(r)}} + \frac{2(\left(c_j^2\right)^{(r)} - 1)}{\left(\phi_j^3\right)^{(r)}}. \tag{27}$$

Then, the Newton-Raphson iterative algorithm for $\boldsymbol{\beta_3}$ is as follows:

$$\boldsymbol{\beta_3}^{(r+1)} \equiv \boldsymbol{\beta_3}^{(r)} - \left[H_3\left(\boldsymbol{\beta_3}^{(r)}\right)\right]^{-1} h_3\left(\boldsymbol{\beta_3}^{(r)}\right). \tag{28}$$

– Finally, differentiate the $Q-$function with respect to $\boldsymbol{\beta_4}$:

$$
\begin{aligned}
h_4\left(\boldsymbol{\beta}_4\right)=\frac{\partial Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right)}{\partial \beta_{4,l}} \quad &= \quad \frac{\partial}{\partial \beta_{4,j}}\left\{\sum_{i=1}^{n}\left[\nu_j^{(r)}\log\left(c_j^{(r)}\right)\right.\right.\\
&+\left(\nu_j^{(r)}-1\right)w_{3,j}-\log\left(K_{\nu_j^{(r)}}\left(\frac{1}{\sigma_j^{(r)}}\right)\right)\\
&\left.\left.-\frac{1}{2\sigma_j^{(r)}}\left(c_j^{(r)}w_{1,j}+\frac{w_{2,j}}{c_j^{(r)}}\right)\right]\right\}x_{4,j,l}, \qquad (29)
\end{aligned}
$$

and

$$
\begin{aligned}
H_4\left(\boldsymbol{\beta}_4\right)=\frac{\partial^2 Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(r)}\right)}{\partial \beta_{4,l}\partial \beta_{4,j}^T} \quad &= \quad \frac{\partial^2}{\partial \beta_{4,j}\partial \beta_{4,j}^T}\left\{\sum_{i=1}^{n}\left[\nu_j^{(r)}\log\left(c_j^{(r)}\right)\right.\right.\\
&+\left(\nu_j^{(r)}-1\right)w_{3,j}-\log\left(K_{\nu_j^{(r)}}\left(\frac{1}{\sigma_j^{(r)}}\right)\right)\\
&\left.\left.-\frac{1}{2\sigma_j^{(r)}}\left(c_j^{(r)}w_{1,j}+\frac{w_{2,j}}{c_j^{(r)}}\right)\right]\right\}x_{4,j,l}x_{4,j,l}^T, (30)
\end{aligned}
$$

for $j=1,..,n$ and $l=1,...,p_4$.

Thus, the Newton-Raphson iterative algorithm for $\boldsymbol{\beta_4}$ is as follows:

$$
\boldsymbol{\beta_4}^{(r+1)} \equiv \boldsymbol{\beta_4}^{(r)} - \left[H_4\left(\boldsymbol{\beta_4}^{(r)}\right)\right]^{-1} h_4\left(\boldsymbol{\beta_4}^{(r)}\right). \qquad (31)
$$

4 Numerical Illustration

We conducted an empirical analysis using a sample of claim frequency data which was randomly selected from a larger pool of MTPL insurance policies observed during the year 2017 from a major European insurance company. We are interested in modelling the MTPL bodily injury and property damage claims with their associated claim counts denoted by $K_{1,j}$ and $K_{2,j}$ respectively, for $j=1,...,n$. For each policy, the total number of claims for each type of claim were reported within this yearly period. The sample comprised insured parties with complete records; i.e., with the availability of all a priori rating variables which affect both $K_{1,j}$ and $K_{2,j}$. Furthermore, an exploratory analysis was carried out in order to accurately select the subset of explanatory variables with the highest predictive power for both $K_{1,j}$ and $K_{2,j}$. There were $n=5186$ observations and three explanatory variables that met our criteria. Table 1 summarizes the explanatory variables whilst Table 2 depicts some standard descriptive statistics for $K_{1,j}$ and $K_{2,j}$, along with the values of Kendall's τ and Spearman's ρ correlation coefficients. As it was expected, Table 2 shows the existence of positive correlation between $k_{1,j}$ and $k_{2,j}$

as well as their marginal overdispersion. Furthermore, we would like to call attention to the fact that, as is well known, the range of Kendall's τ and Spearman's ρ for discrete random variables is narrower than $[-1, 1]$, see Denuit and Lambert (2005), Mesfioui and Tajar (2005) and Mesfioui et al. (2021). Furthermore, Nikoloulopoulos and Karlis (2010) and Safari-Katesari et al. (2020) showed how to compute the population versions of Kendall's τ and Spearman's ρ by pairing copulas with discrete marginal distributions respectively. Following their setup, we investigated the variability of the population versions of Kendall's τ and Spearman's ρ from lowest to highest attainable values for our data by pairing two marginal Poisson distributions with varying mean parameter μ from 1 up to 20 using the Normal copula. Also, we considered that the copula parameter θ can vary from -1 to 1. We observed that the values of Kendall's τ and Spearman's ρ stabilize close to 1 for the values of μ which are greater than 10. Therefore, the bivariate Negative Binomial (BNB) and Poisson-Inverse Gaussian (BPIG) regression models with varying dispersion and the bivariate Poisson-Generalized Inverse Gaussian (BPGIG) regression model with varying dispersion and shape which allow for positive correlation between the two types of claims are better assumptions than the bivariate Poisson model, as the latter is not equipped for handling overdispersion. Moreover, Table 3 presents the estimated regression coefficients for the BNB and BPIG regression models with varying dispersion and the BPGIG regression model with varying dispersion and shape[2].

Furthermore, we compare the fit of the BNB and BPIG regression models with varying dispersion to that of the BPGIG regression model with varying dispersion and shape based on the standard specification tests DEV, AIC and SBC. The DEV is given by

$$\text{DEV} = -2\hat{l}\left(\hat{\boldsymbol{\theta}}\right), \tag{32}$$

with $\hat{l}$ being the maximum of the log-likelihood and $\hat{\boldsymbol{\theta}}$ the vector of estimated parameters of the model. Moreover, the AIC is defined as

$$\text{AIC} = \text{DEV} + 2 \times df \tag{33}$$

and the SBC is given by

$$\text{SBC} = \text{DEV} + \log(n) \times df, \tag{34}$$

where df are the degrees of freedom which correspond to the number of fitted parameters in the model and n is the number of observations in the sample. The values of the DEV, AIC and SBC for the competing models are provided in Table 4. As is well known, two models can be considered to be significantly different if the difference in the log-likelihoods exceeds five, corresponding to a difference in their respective AIC and SBC values of greater than ten and five respectively. Thus, this case we see that the best fitting performances are provided by the BPGIG regression model with varying dispersion and shape[3].

[2]All the parameters were statistically significant at a 5% threshold.

[3]Note that the stopping criterion for the EM algorithm was rather strict as the algorithm iterated between the E and the M-steps until the relative change in the log-likelihood between two successive iterations was smaller than 10^-12.

Finally, we compare the forecasting performance of the proposed model and the benchmark models using both in-sample estimation and out-of sample validation. For this purpose, we split the data into training and test data at the ratio of 9 : 1. Therefore, the training data for the re-estimation of the parameters of the models contains 4149 data points. The remaining 1037 data points are used for testing purposes. To measure the prediction performances the deviance statistic is used. The deviance value for the BNB, BPIG, and BPGIG models are 490.80, 475.25, and 472.35 respectively. Thus, the BPGIG model outperforms the two competing bivariate mixed Poisson models.

5 Calculation of the A Posteriori Premiums

In this subsection, the expected value premium principle is used to compute the a posteriori, or Bonus-Malus, premium rates determined by the BNB, BPIG, and BPGIG models for $t = 1$ for three risk class profiles that we classify as Best, Average, and Worst according to the values of the mean claim frequencies $\mu_{1,j}$ and $\mu_{2,j}$, which are calculated using the same set of explanatory variables per claim type $i = 1, 2$. The results are depicted in Table 5.

6 Conclusions

In this article, we presented the MVPGIG claims count regression model with varying dispersion and shape for modelling different types of claims in non-life insurance. The MVPGIG is a wide family of models which, under the proposed modelling framework, can provide sufficient flexibility for capturing overdispersion and positive correlation structures in highly-dimensional claim count data. For demonstration purposes, the bivariate version of the model, namely the BPGIG model, with regression specifications for the mean, dispersion, and shape parameters was fitted on MTPL property damage and bodily injury claim count data. The ML estimates of the parameters of the model were obtained via a novel EM type algorithm. However, it should be noted that a shortcoming of the proposed approach is that there is a strong discrepancy between the flexibility within the equations of the random effect distribution, and the rigidity between these equations. In order to relax this rigidity, the BPGIG model can be constructed either by using the so-called trivariate reduction method or by considering correlated GIG random effects (say $z_{1;j}$ and $z_{2;j}$) paired via a Gaussian copula following the approaches of Bermúdez and Karlis (2017) and Pechon et al. (2018) in the former and latter case respectively. Both approaches are very efficient when modelling different types of claims from different types of coverage or household claim frequencies in MTPL insurance. Finally, in a forthcoming paper, time series components will be included to accommodate for both cross dependence between different types of claims and time dependence, proceeding along similar lines as in Bermúdez et al. (2018) among others.

References

Abdallah, A., Boucher, J.-P., and Cossette, H. (2016). Sarmanov family of multivariate distributions for bivariate dynamic claim counts model. *Insurance: Mathematics and Economics*, 68:120–133.

Aguero-Valverde, J. and Jovanis, P. P. (2009). Bayesian multivariate Poisson lognormal models for crash severity modeling and site ranking. *Transportation research record*, 2136(1):82–91.

Aitchison, J. and Ho, C. (1989). The multivariate Poisson-log normal distribution. *Biometrika*, 76(4):643–653.

Bermúdez, L., Guillén, M., and Karlis, D. (2018). Allowing for time and cross dependence assumptions between claim counts in ratemaking models. *Insurance: Mathematics and Economics*, 83:161–169.

Bermúdez, L. and Karlis, D. (2011). Bayesian multivariate Poisson models for insurance ratemaking. *Insurance: Mathematics and Economics*, 48(2):226–236.

Bermúdez, L. and Karlis, D. (2012). A finite mixture of bivariate Poisson regression models with an application to insurance ratemaking. *Computational Statistics & Data Analysis*, 56(12):3988–3999.

Bermúdez, L. and Karlis, D. (2017). A posteriori ratemaking using bivariate poisson models. *Scandinavian Actuarial Journal*, 2017(2):148–158.

Bolancé, C., Guillen, M., and Pitarque, A. (2020). A Sarmanov distribution with beta marginals: An application to motor insurance pricing. *Mathematics*, 8(11).

Bolancé, C. and Vernic, R. (2019). Multivariate count data generalized linear models: Three approaches based on the Sarmanov distribution. *Insurance: Mathematics and Economics*, 85:89–103.

Cameron, A. C., Li, T., Trivedi, P. K., and Zimmer, D. M. (2004). Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts. *The Econometrics Journal*, 7(2):566–584.

Cameron, A. C. and Trivedi, P. K. (2013). *Regression analysis of count data*, volume 53. Cambridge university press.

Chen, Y. and Hanson, T. (2017). Copula regression models for discrete and mixed bivariate responses. *Journal of Statistical Theory and Practice*, 11(4):515–530.

Chib, S. and Winkelmann, R. (2001). Markov chain Monte Carlo analysis of correlated count data. *Journal of Business & Economic Statistics*, 19(4):428–435.

Chiquet, J., Mariadassou, M., and Robin, S. (2020). The Poisson-lognormal model as a versatile framework for the joint analysis of species abundances. *bioRxiv*.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

Denuit, M., Guillen, M., Trufin, J., et al. (2019). Multivariate credibility modelling for usage-based motor insurance pricing with behavioural data. *Annals of Actuarial Science*, 13(2):378–399.

Denuit, M. and Lambert, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis*, 93(1):40–57.

El-Basyouny, K. and Sayed, T. (2009). Accident prediction models with random corridor parameters. *Accident Analysis & Prevention*, 41(5):1118–1123.

Famoye, F. (2010). On the bivariate negative binomial regression model. *Journal of Applied Statistics*, 37(6):969–981.

Fung, T. C., Badescu, A. L., and Lin, X. S. (2019). A class of mixture of experts models for general insurance: Application to correlated claim frequencies. *ASTIN Bulletin: The Journal of the IAA*, 49(3):647–688.

Genest, C. and Nešlehová, J. (2007). A primer on copulas for count data. *ASTIN Bulletin: The Journal of the IAA*, 37(2):475–515.

Ghitany, M., Karlis, D., Al-Mutairi, D., and Al-Awadhi, F. (2012). An EM algorithm for multivariate mixed poisson regression models and its application. *Applied Mathematical Sciences*, 6(137):6843–6856.

Gómez-Déniz, E. and Calderín-Ojeda, E. (2021). A priori ratemaking selection using multivariate regression models allowing different coverages in auto insurance. *Risks*, 9(7):137.

Gurmu, S. and Elder, J. (2000). Generalized bivariate count data regression models. *Economics Letters*, 68(1):31–36.

Ho, L. L. and Singer, J. d. M. (2001). Generalized least squares methods for bivariate Poisson regression. *Communications in Statistics-Theory and Methods*, 30(2):263–277.

Jeong, H. and Dey, D. K. (2021). Multi-peril frequency credibility premium via shared random effects. *Available at SSRN 3825435*.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997). *Discrete multivariate distributions*, volume 165. Wiley New York.

Jung, R. C. and Winkelmann, R. (1993). Two aspects of labor mobility: a bivariate Poisson regression approach. *Empirical economics*, 18(3):543–556.

Karlis, D. and Meligkotsidou, L. (2005). Multivariate Poisson regression with covariance structure. *Statistics and Computing*, 15(4):255–265.

Kocherlakota, S. (1988). On the compounded bivariate Poisson distribution: A unified treatment. *Annals of the institute of Statistical Mathematics*, 40(1):61–76.

Kocherlakota, S. and Kocherlakota, K. (2001). Regression in the bivariate Poisson distribution.

Krummenauer, F. (1998). Representation of multivariate discrete distributions by probability generating functions. *Statistics & probability letters*, 39(4):327–331.

Lakshminarayana, J., Pandit, S. N., and Srinivasa Rao, K. (1999). On a bivariate Poisson distribution. *Communications in Statistics-Theory and Methods*, 28(2):267–276.

Lee, A. (1999). Applications: Modelling rugby league data via bivariate negative binomial regression. *Australian & New Zealand Journal of Statistics*, 41(2):141–152.

Ma, J., Kockelman, K. M., and Damien, P. (2008). A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using bayesian methods. *Accident Analysis & Prevention*, 40(3):964–975.

Marra, G. and Wyszynski, K. (2016). Semi-parametric copula sample selection models for count responses. *Computational Statistics & Data Analysis*, 104:110–129.

McLachlan, G. J. and Krishnan, T. (2007). *The EM algorithm and extensions*, volume 382. John Wiley & Sons.

Mesfioui, M. and Tajar, A. (2005). On the properties of some nonparametric concordance measures in the discrete case. *Nonparametric Statistics*, 17(5):541–554.

Mesfioui, M., Trufin, J., and Zuyderhoff, P. (2021). Bounds on spearman's rho when at least one random variable is discrete. *European Actuarial Journal*, pages 1–28.

M'Kendrick, A. (1925). Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98–130.

Munkin, M. K. and Trivedi, P. K. (1999). Simulated maximum likelihood estimation of multivariate mixed-poisson regression models, with application. *The Econometrics Journal*, 2(1):29–48.

Nikoloulopoulos, A. K. (2013). Copula-based models for multivariate discrete response data. In *Copulae in Mathematical and Quantitative Finance*, pages 231–249. Springer.

Nikoloulopoulos, A. K. (2016). Efficient estimation of high-dimensional multivariate normal copula models with discrete spatial responses. *Stochastic environmental research and risk assessment*, 30(2):493–505.

Nikoloulopoulos, A. K. and Karlis, D. (2010). Regression in a copula model for bivariate count data. *Journal of Applied Statistics*, 37(9):1555–1568.

Park, E. S. and Lord, D. (2007). Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. *Transportation Research Record*, 2019(1):1–6.

Pechon, F., Denuit, M., and Trufin, J. (2019). Multivariate modelling of multiple guarantees in motor insurance of a household. *European Actuarial Journal*, 9(2):575–602.

Pechon, F., Denuit, M., and Trufin, J. (2021). Home and motor insurance joined at a household level using multivariate credibility. *Annals of Actuarial Science*, 15(1):82–114.

Pechon, F., Trufin, J., Denuit, M., et al. (2018). Multivariate modelling of household claim frequencies in motor third-party liability insurance. *Astin Bulletin*, 48(3):969–993.

Rüschendorf, L. (2013). Copulas, Sklar's theorem, and distributional transform. In *Mathematical risk analysis*, pages 3–34. Springer.

Safari-Katesari, H., Samadi, S. Y., and Zaroudi, S. (2020). Modelling count data via copulas. *Statistics*, 54(6):1329–1355.

Shared, M. (1980). On mixtures from exponential families. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(2):192–198.

Shi, P. and Valdez, E. A. (2014a). Longitudinal modeling of insurance claim counts using jitters. *Scandinavian Actuarial Journal*, 2014(2):159–179.

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

Silva, A., Rothstein, S. J., McNicholas, P. D., and Subedi, S. (2019). A multivariate Poisson-log normal mixture model for clustering transcriptome sequencing data. *BMC bioinformatics*, 20(1):1–11.

Stein, G. Z. and Juritz, J. M. (1987). Bivariate compound Poisson distributions. *Communications in Statistics-Theory and Methods*, 16(12):3591–3607.

Stein, G. Z., Zucchini, W., and Juritz, J. M. (1987). Parameter estimation for the Sichel distribution and its multivariate extension. *Journal of the American Statistical Association*, 82(399):938–944.

Tzougas, G. and di Cerchiara, A. P. (2021). Bivariate mixed poisson regression models with varying dispersion. *North American Actuarial Journal*, pages 1–31.

Tzougas, G. and Pignatelli di Cerchiara, A. (in press 2021). *North American Actuarial Journal.*

Winkelmann, R. (2008). *Econometric analysis of count data.* Springer Science & Business Media.

Zhan, X., Aziz, H. A., and Ukkusuri, S. V. (2015). An efficient parallel sampling technique for multivariate Poisson-lognormal model: Analysis with two crash count datasets. *Analytic methods in accident research*, 8:45–60.

Zimmer, D. M. and Trivedi, P. K. (2006). Using trivariate copulas to model sample selection and treatment effects: application to family health care demand. *Journal of Business & Economic Statistics*, 24(1):63–76.

List of tables

Table 1: The explanatory variables and their description

Variables	Categories C1	C2	C3
City population (v1)	$\leq$ 1,000,000	1,000,001-2,000,000	$\geq$ 2,000,001
Number of years that the policyholder has been registered with the insurance company (v2)	< 5 years	> 5 years	-
Horsepower of the vehicle (v3)	0-1400 cc	1400-1800 cc	$\geq$ 1800 cc

Table 2: Descriptive statistics for the two responses

K_1		K_2	
statistic	value	statistic	value
Minimum	0	Minimum	0
Median	0	Median	0
Mean	0.0954	Mean	0.0618
Variance	0.1375	Variance	0.0644
Maximum	4	Maximum	3
Kendall's τ: 0.1760			
Spearman's ρ: 0.1777			

Table 3: Parameter estimates of the BNB and BPIG regression models with varying dispersion and the BPGIG regression model with varying dispersion and shape

	BNB			BPIG			BPGIG			
Variable	Coeff. β_1	Coeff. β_2	Coeff. β_3	Coeff. β_1	Coeff. β_2	Coeff. β_3	Coeff. β_1	Coeff. β_2	Coeff. β_3	$\nu = -0.53$
Intercept	-2.3933	-2.9262	-1.1296	−2.3950	−2.9279	−0.5908	-2.3964	-2.9293	-0.5893	-
v1 C2	0.0524	0.1504	-0.1912	0.0535	0.1518	−0.1157	0.0543	0.1529	-0.1149	-
v1 C3	0.1556	0.1770	-0.2303	0.1587	0.1793	−0.1364	0.1615	0.1813	-0.1352	-
v2 C2	0.0452	0.1780	0.3627	0.0465	0.1790	0.1959	0.0475	0.1797	0.1968	-
v3 C2	-0.1216	-0.0203	-0.3144	−0.1203	-0.0190	−0.1769	-0.1187	-0.0174	-0.1756	-
v3 C3	0.1767	0.1712	-0.0883	0.1784	0.1731	−0.0716	0.1798	0.1747	-0.0711	-

Table 4: BNB, BPIG and BPGIG Models Comparison based on Global Deviance, AIC and SBC

Model	*df*	Global Deviance	AIC	SBC
BNB	18	4388	4424	4542
BPIG	18	4249	4285	4403
BPGIG	19	4223	4261	4386

Table 5: Comparison of the A Posteriori, or Bonus-Malus, Premium Rates for $t = 1$

	$t=1$ BNB regression model Best profile				$t=1$ BPIG regression model Best profile				$t=1$ BPGIG regression model Best profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	76.36	312.65	548.94	785.23	82.05	239.47	500.38	800.56	82.05	254.83	556.90	917.13
1	312.65	548.94	785.23	1021.52	239.47	500.38	800.56	1110.36	1 254.82	556.90	917.13	1297.67
2	548.94	785.23	1021.52	1257.81	500.38	800.56	1110.36	1422.85	556.91	917.13	1297.67	1688.07
3	785.23	1021.52	1257.81	1494.10	800.56	1110.36	1422.85	1736.37	917.13	1297.67	1688.07	2084.96
	$t=1$ BNB regression model Average profile				$t=1$ BPIG regression model Average profile				$t=1$ BPGIG regression model Average profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	73.18	334.41	595.64	856.88	80.00	247.48	528.29	849.50	80.00	264.20	590.75	978.13
1	334.41	595.64	856.88	1118.11	247.48	528.29	849.50	1179.87	264.20	590.75	978.13	1386.04
2	595.64	856.88	1118.11	1379.35	528.29	849.50	1179.87	1512.72	590.75	978.13	1386.04	1804.03
3	856.88	1118.11	1379.35	1640.58	849.50	1179.87	1512.72	1846.48	978.13	1386.04	1804.03	2228.75
	$t=1$ BNB regression model Worst profile				$t=1$ BPIG regression model Worst profile				$t=1$ BPLN regression model Worst profile			
$k_{1,j}/k_{2,j}$	0	1	2	3	0	1	2	3	0	1	2	3
0	71.49	283.19	494.88	706.58	79.05	210.99	425.43	674.37	79.05	223.39	470.01	766.39
1	283.19	494.88	706.58	918.27	210.99	425.43	674.37	932.83	223.39	470.01	766.39	1081.27
2	494.88	706.58	918.27	1129.97	425.43	674.37	932.83	1194.13	470.01	766.39	1081.27	1405.04
3 1734.49	706.58	918.27	1129.97	1341.66	674.37	932.83	1194.13	1456.54		766.39	1081.28	1405.04

Acknowledgements: The authors would like to thank the two anonymous referees for their very helpful comments and suggestions which have significantly improved this research paper.

7.2 Supplementary material

We briefly consider a potential extension of our work based on the research of Bermúdez & Karlis (2017). That said, we deem helpful to provide the following background information. In particular, there, the number of claims for two coverage types for an insured i at time j are random variables which are represented by N_{1ij} and N_{2ij}. The values of these random variables are denoted as n_{1ij} and n_{2ij} respectively. The bivariate Poisson distribution is presented as a tool that can be used to capture the underlying correlation between two types of claims arising from the same insurance contract. That said, $(N_1, N_2) \sim \text{BP}(\lambda_1, \lambda_2, \lambda_3)$ reflecting that the pair of random variables $(N1, N2)$ are following the bivariate Poisson distribution with parameters λ_1, λ_2 and λ_3.

Then, an alternative definition of the bivariate Poisson distribution using the so-called trivariate reduction method is presented. It starts by considering that the independent random variables X_k for $k = 1, 2, 3$ follow the Poisson distribution with respective parameters λ_k. Consequently, the random variables $N_1 = X_1 + X_3$ and $N_2 = X_2 + X_3$ also follow a bivariate Poisson distribution jointly. Then, the joint probability function for the i^{th} policyholder, when ignoring the subscript j for the sake of simplicity is given by

$$P(n_{1i}, n_{2i}) = P(N_{1i} = n_{1i}, N_{2i} = n_{2i}; \boldsymbol{\lambda}_{\mathbf{i}}) =$$

$$= e^{-(\lambda_{1i}+\lambda_{2i}+\lambda_{3i})} \frac{\lambda_{1i}^{n_{1i}}}{n_{1i}!} \frac{\lambda_{2i}^{n_{2i}}}{n_{2i}!} \sum_{s=0}^{min(n_{1i}, n_{2i})} \binom{n_{1i}}{s} \binom{n_{2i}}{s} s! \left(\frac{\lambda_{3i}}{\lambda_{1i}\lambda_{2i}} \right)^s$$

where $\lambda_{\mathbf{i}} = (\lambda_{1i}, \lambda_{2i}, \lambda_{3i})$.

That said, it would be fruitful for future research to consider an extension of our problem using a trivariate reduction scheme in order to incorporate flexibility in the tail of claims distribution, and also model the correlation between the two response variables by incorporating risk factors in λ_3.

Finally, in our application we focus on the widely used Kendall's τ and Spearmann's ρ measures of correlation between the different claims types, and this means that in our bivariate case, we associate the entire distribution of the two claims types. Nevertheless, it is worth noting that there may be differences in the dependence between the upper part of the distribution and the mid-range and/or lower part of the distribution and this can impact claims modelling, see Embrechts et al. (2002). Therefore, a potential extension of our research could focus on the tail dependence aspect, especially if we were to model claim sizes instead of claims counts as in the former case tail dependence is perhaps even more relevant.

References

Abdallah, A., Boucher, J.-P. & Cossette, H. (2016), 'Sarmanov family of multivariate distributions for bivariate dynamic claim counts model', *Insurance: Mathematics and Economics* **68**, 120–133.

Accenture (2018), 'Machine learning in insurance'. Accessed = 2021-09-02.
URL: *https://www.accenture.com/_acnmedia/pdf-84/accenture-machine-leaning-insurance.pdf*

Aguero-Valverde, J. & Jovanis, P. P. (2009), 'Bayesian multivariate Poisson lognormal models for crash severity modeling and site ranking', *Transportation research record* **2136**(1), 82–91.

Aitchison, J. & Ho, C. (1989), 'The multivariate Poisson-log normal distribution', *Biometrika* **76**(4), 643–653.

Alexander, C. & Sarabia, J. M. (2012), 'Quantile uncertainty and value-at-risk model risk', *Risk Analysis: An International Journal* **32**(8), 1293–1308.

Anzai, Y. (2012), *Pattern recognition and machine learning*, Elsevier.

Artzner, P., Delbaen, F., Eber, J.-M. & Heath, D. (1999), 'Coherent measures of risk', *Mathematical finance* **9**(3), 203–228.

Bank of England (2021), 'What is the Prudential Regulation Authority (PRA)?'. Accessed = 2021-09-10.
URL: *https://www.bankofengland.co.uk/knowledgebank/what-is-the-prudential-regulation-authority-pra*

Banks, E. (2014), *Alternative Risk Transfer - Integrated Risk Management through Insurance, Reinsurance, and the Capital Markets*, John Wiley and Sons Ltd.

Barrieu, P. & El Karoui, N. (2005), 'Inf-convolution of risk measures and optimal risk transfer', *Finance and stochastics* **9**(2), 269–298.

Barrieu, P. & El Karoui, N. (2007), ‘Pricing, hedging and optimally designing derivatives via minimization of risk measures’, *arXiv preprint arXiv:0708.0948* .

Barrieu, P. & Scandolo, G. (2015), ‘Assessing financial model risk’, *European Journal of Operational Research* **242**(2), 546–556.

Bermúdez, L., Guillén, M. & Karlis, D. (2018), ‘Allowing for time and cross dependence assumptions between claim counts in ratemaking models’, *Insurance: Mathematics and Economics* **83**, 161–169.

Bermúdez, L. & Karlis, D. (2011), ‘Bayesian multivariate Poisson models for insurance ratemaking’, *Insurance: Mathematics and Economics* **48**(2), 226–236.

Bermúdez, L. & Karlis, D. (2012), ‘A finite mixture of bivariate Poisson regression models with an application to insurance ratemaking’, *Computational Statistics & Data Analysis* **56**(12), 3988–3999.

Bermúdez, L. & Karlis, D. (2017), ‘A posteriori ratemaking using bivariate Poisson models’, *Scandinavian Actuarial Journal* **2017**(2), 148–158.

Bertram, P., Sibbertsen, P. & Stahl, G. (2015), ‘The impact of model risk on capital reserves: A quantitative analysis’, *Journal of Risk* **17**(5).

Birds, J. (2010), *Birds’ Modern Insurance Law*, 8 edn, Sweet & Maxwell.

Bishop, C. (2016), *Pattern Recognition and Machine Learning*, Information Science and Statistics, Springer New York.

Bolancé, C. & Vernic, R. (2019), ‘Multivariate count data generalized linear models: Three approaches based on the sarmanov distribution’, *Insurance: Mathematics and Economics* **85**, 89–103.

Boucher, C. M., Danielsson, J., Kouontchou, P. S. & Maillet, B. B. (2014), ‘Risk models-at-risk’, *Journal of Banking & Finance* **44**, 72–92.

Box, G. E., Jenkins, G. M., Reinsel, G. C. & Ljung, G. M. (2015), *Time series analysis: forecasting and control*, John Wiley & Sons.

Boyd, J. (2016), ‘Modeling fundamentals: So you want to issue a cat bond’. Accessed = 2016-08-26.
URL: *http//www.air-worldwide.com/Publications/AIR-Currents/2016/Modeling-Fundamentals--So-You-Want-to-Issue-a-Cat-Bond/*

Boyd, S. & Vandenberghe, L. (2004), *Convex Optimization*, Cambridge University Press.

Braun, A., Eling, M., Schmeiser, H. & Schanz, K.-U. (2020), ‘An investigation into the insurability of pandemic risk’. Accessed = 2021-12-03.
URL: *https://www.genevaassociation.org/sites/default/files/research-topics-document-type/pdf_public/insurability_report_web.pdf*

Breiman, L., Cox, D. & Breiman, L. (2001), ‘Comment-statistical modeling: The two cultures’, *Statistical Science* **16**(3), 199–231.

Bugmann, C. (1997), ‘Proportional and non-proportional reinsurance’. Accessed = 2021-10-12.
URL: *https://www.swissre.com/dam/jcr:1ec8b4cd-dafc-411f-aafd-385d6517ccfd/pub_proportional_non_proportional_en.pdf*

Bühlmann, H. (2005), *Mathematical Methods in Risk Theory*, Grundlehren der mathematischen Wissenschaften, Springer Berlin Heidelberg.

Cairns, A. J. (2000), ‘A discussion of parameter and model uncertainty in insurance’, *Insurance: Mathematics and Economics* **27**(3), 313–330.

Cameron, A. C., Li, T., Trivedi, P. K. & Zimmer, D. M. (2004), ‘Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts’, *The Econometrics Journal* **7**(2), 566–584.

Cameron, A. C. & Trivedi, P. K. (2013), *Regression analysis of count data*, Vol. 53, Cambridge university press.

Casualty Actuarial Society (2020), ‘Exceedance probability in catastrophe modeling’.
URL: *https://www.casact.org/sites/default/files/2021-03/02_humphreys.pdf*

Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M. & Montrucchio, L. (2011), ‘Risk measures: rationality and diversification’, *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics* **21**(4), 743–774.

Chen, Y. & Hanson, T. (2017), ‘Copula regression models for discrete and mixed bivariate responses’, *Journal of Statistical Theory and Practice* **11**(4), 515–530.

Chib, S. & Winkelmann, R. (2001), ‘Markov chain monte carlo analysis of correlated count data’, *Journal of Business & Economic Statistics* **19**(4), 428–435.

Chiquet, J., Mariadassou, M. & Robin, S. (2020), ‘The Poisson-lognormal model as a versatile framework for the joint analysis of species abundances’, *bioRxiv* .

Cont, R. (2006), 'Model uncertainty and its impact on the pricing of derivative instruments', *Mathematical finance* **16**(3), 519–547.

Cummins, J. D. (2008), 'Cat bonds and other risk-linked securities: state of the market and recent developments', *Risk management and insurance review* **11**(1), 23–47.

Cummins, J. D., Dionne, G., Gagné, R. & Nouira, A. (2021), 'The costs and benefits of reinsurance', *The Geneva Papers on Risk and Insurance-Issues and Practice* **46**(2), 177–199.

Cummins, J. D., Harrington, S. E. & Klein, R. W. (1991), 'Cycles and crises in property/-casualty insurance: a background discussion', *Cycles and Crises in Property/Casualty Insurance: Causes and Implications for Public Policy, National Association of Insurance Commissioners, Kansas City, MO* pp. 1–51.

Cummins, J. D. & Outreville, J. F. (1992), An international analysis of underwriting cycles in property-liability insurance, *in* 'Foundations of Insurance Economics', Springer, pp. 609–625.

Cummins, J. & Mahul, O. (2009), *Catastrophe Risk Financing in Developing Countries: Principles for Public Intervention*, World Bank e-Library, World Bank.

Delong, L. (2013), *Backward Stochastic Differential Equations with Jumps and Their Actuarial and Financial Applications: BSDEs with Jumps*, EAA Series, Springer London.

Denuit, M., Guillen, M., Trufin, J. et al. (2019), 'Multivariate credibility modelling for usage-based motor insurance pricing with behavioural data', *Annals of Actuarial Science* **13**(2), 378–399.

Deprez, O. & Gerber, H. U. (1985), 'On convex principles of premium calculation', *Insurance: Mathematics and Economics* **4**(3), 179–189.

El-Basyouny, K. & Sayed, T. (2009), 'Accident prediction models with random corridor parameters', *Accident Analysis & Prevention* **41**(5), 1118–1123.

Embrechts, P., McNeil, A. & Straumann, D. (2002), 'Correlation and dependence in risk management: properties and pitfalls', *Risk management: value at risk and beyond* **1**, 176–223.

Embrechts, P., Puccetti, G. & Rüschendorf, L. (2013), 'Model uncertainty and var aggregation', *Journal of Banking & Finance* **37**(8), 2750–2764.

Embrechts, P. & Wuttrich, M. (2022), 'Recent challenges in actuarial science', *Annual Review of Statistics and Its Application* **9**.

European Insurance and Occupational Pensions Authority (2015), 'Directive 138/2009/EC (Solvency II Directive)'. Accessed = 2021-09-15.
URL: *https://www.eiopa.europa.eu/rulebook-categories/directive-1382009ec-solvency-ii-directive_en*

Ferrall, C. (2005), 'Solving finite mixture models: Efficient computation in economics under serial and parallel execution', *Computational Economics* **25**(4), 343–379.

Financial Conduct Authority (2013), 'FCA-Lloyd's co-operation arrangements'. Accessed = 2021-10-15.
URL: *https://assets.lloyds.com/assets/pdf-regulation-of-lloyds-fca-cooperation-agreement20130717/1/pdf-regulation-of-lloyds-FCA-cooperation-agreement20130717.pdf*

Financial Conduct Authority (2019), 'Machine learning in UK financial services'. Accessed = 2021-12-01.
URL: *https://www.bankofengland.co.uk/-/media/boe/files/report/2019/machine-learning-in-uk-financial-services.pdf*

Finken, S. & Laux, C. (2009), 'Catastrophe bonds and reinsurance: the competitive effect of information-insensitive triggers', *Journal of Risk and Insurance* **76**(3), 579–605.

Föllmer, H. & Schied, A. (2002), 'Convex measures of risk and trading constraints', *Finance and stochastics* **6**(4), 429–447.

Föllmer, H. & Schied, A. (2010), 'Convex and Coherent Risk Measures', *Journal of Quantitative Finance* p. 12.
URL: *http://alexschied.de/Encyclopedia6.pdf*

Föllmer, H. & Schied, A. (2016), *Stochastic Finance: An Introduction in Discrete Time*, De Gruyter Textbook, De Gruyter.

Friedman, J., Hastie, T., Tibshirani, R. et al. (2001), *The elements of statistical learning*, Vol. 1, Springer series in statistics New York.

Gareth, J., Daniela, W., Trevor, H. & Robert, T. (2013), *An introduction to statistical learning: with applications in R*, Spinger.

Gatzert, N. & Kellner, R. (2011), 'The influence of non-linear dependencies on the basis risk of industry loss warranties', *Insurance: Mathematics and Economics* **49**(1), 132–144.

Gatzert, N. & Schmeiser, H. (2012), 'Industry loss warranties: contract features, pricing, and central demand factors', *The Journal of Risk Finance* .

Genest, C. & Nešlehová, J. (2007), 'A primer on copulas for count data', *ASTIN Bulletin: The Journal of the IAA* **37**(2), 475–515.

Gerber, H. (1979), *An introduction to mathematical risk theory*, S.S. Huebner Foundation monograph series, S. S. Huebner Foundation for Insurance Education, Wharton School, University of Pennsylvania.

Ghitany, M., Karlis, D., Al-Mutairi, D. & Al-Awadhi, F. (2012), 'An EM algorithm for multivariate mixed poisson regression models and its application', *Applied Mathematical Sciences* **6**(137), 6843–6856.

Gorge, G. (2009), Reinsurance vs securitisation, *in* P. Barrieu & L. Albertini, eds, 'The handbook of insurance-linked securities', 1 edn, John Wiley & Sons, chapter 3B, p. 29.

Greenwald, J. (2006), 'Reinsurers attach sidecars to expand catastrophe capacity'. Accessed = 2021-08-20.
URL: *https://www.businessinsurance.com/article/20060514/ISSUE01/10001891 5/reinsurers-attach-sidecars-to-expand-catastrophe-capacity*

Guo, Y., Zhou, Y., Hu, X. & Cheng, W. (2019), Research on recommendation of insurance products based on random forest, *in* '2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)', IEEE, pp. 308–311.

Gurmu, S. & Elder, J. (2000), 'Generalized bivariate count data regression models', *Economics Letters* **68**(1), 31–36.

Guy Carpenter (2005), 'The growing appetite for catastrophe risk: The catastrophe bond market at year-end 2004', *New York* .

Henderson, V. & Hobson, D. (2009), *Utility Indifference Pricing: An Overview*, Princeton University Press, pp. 44–74.
URL: *http://www.jstor.org/stable/j.ctt7rn0j.5*

HM Treasury (2016), 'Insurance linked securities: consultation'. Accessed = 2021-08-25.
URL: *https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/504046/Insurance_linked_securities_consultation.pdf*

Ho, L. L. & Singer, J. d. M. (2001), 'Generalized least squares methods for bivariate poisson regression', *Communications in Statistics-Theory and Methods* **30**(2), 263–277.

Imbens, G. & Athey, S. (2021), 'Breiman's two cultures: A perspective from econometrics', *Observational Studies* **7**(1), 127–133.

Insurance Information Institute (2021), 'Background on: Insurance fraud'. Accessed = 2021-12-03.
URL: *https://www.iii.org/article/background-on-insurance-fraud*

Ishaq, A. (2005), Reinsuring for catastrophes through industry loss warranties–a practical approach, *in* 'Casualty Actuarial Society Forum', Citeseer, pp. 75–91.

Jain, P. & Kar, P. (2017), 'Non-convex optimization for machine learning', *arXiv preprint arXiv:1712.07897* .

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013), *An introduction to statistical learning*, Springer.

Jeong, H. & Dey, D. (2021), 'Multi-peril frequency credibility premium via shared random effects', *Available at SSRN 3825435* .

Jeong, H., Tzougas, G. & Fung, T. C. (2021), 'Multivariate claim count regression model with varying dispersion and dependence parameters'.

Jerry, R. H. & Richmond, D. R. (2012), *Understanding insurance law*, LexisNexis.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994), *Continuous univariate distributions, volume 1*, John wiley & sons.

Jung, R. C. & Winkelmann, R. (1993), 'Two aspects of labor mobility: a bivariate poisson regression approach', *Empirical economics* **18**(3), 543–556.

Karlis, D. & Meligkotsidou, L. (2005), 'Multivariate Poisson regression with covariance structure', *Statistics and Computing* **15**(4), 255–265.

Kerkhof, J., Melenberg, B. & Schumacher, H. (2010), 'Model risk and capital reserves', *Journal of Banking & Finance* **34**(1), 267–279.

Kocherlakota, N. R. (1992), 'Bubbles and constraints on debt accumulation', *Journal of Economic theory* **57**(1), 245–256.

Kocherlakota, S. (1988), 'On the compounded bivariate poisson distribution: A unified treatment', *Annals of the institute of Statistical Mathematics* **40**(1), 61–76.

Kocherlakota, S. & Kocherlakota, K. (2001), 'Regression in the bivariate poisson distribution'.

Kopf, E. (1871), 'Notes on the origin and development of reinsurance'. Accessed = 2021-10-18.

URL: *https://www.casact.org/sites/default/files/database/proceed_proceed29_29022.pdf*

Kuha, J. (2004), 'Aic and bic: Comparisons of assumptions and performance', *Sociological methods & research* **33**(2), 188–229.

Laeven, R. J. A. & Stadje, M. (2012), 'Entropy Coherent and Entropy Convex Measures of Risk', *Mathematics of Operations Research* **38**(2), 265–293.
URL: *http://dx.doi.org/10.1287/moor.1120.0559*

Lakshminarayana, J., Pandit, S. N. & Srinivasa Rao, K. (1999), 'On a bivariate Poisson distribution', *Communications in Statistics-Theory and Methods* **28**(2), 267–276.

Lane, M. (2007), 'Of sidecars and such', *Lane Financial, Wilmette, IL* .

Lee, A. (1999), 'Applications: Modelling rugby league data viabivariate negative binomial regression', *Australian & New Zealand Journal of Statistics* **41**(2), 141–152.

Li, Y., Yan, C., Liu, W. & Li, M. (2018), 'A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification', *Applied Soft Computing* **70**, 1000–1009.

Lloyd's Market Association (2013), Catastrophe modelling guidance for non-catastrophe modellers, Technical report.

Lloyd's Market Association (2017), Understanding uncertainty in catastrophe modelling for non-catastrophe modellers, Technical report.

Lloyd's (2019), 'Lloyd's pocket guide'. Accessed = 2021-08-20.
URL: *https://assets.lloyds.com/assets/pdf-history-lloydspocketguidev7digital/1/pdf-history-Lloydspocketguidev7digital.pdf*

Ma, J., Kockelman, K. M. & Damien, P. (2008), 'A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using bayesian methods', *Accident Analysis & Prevention* **40**(3), 964–975.

Maes, S. & Schoutens, W. (2012), 'Contingent capital: An in-depth discussion', *Economic notes* **41**(1-2), 59–79.

Makariou, D., Barrieu, P. & Chen, Y. (2021), 'A random forest based approach for predicting spreads in the primary catastrophe bond market', *Insurance: Mathematics and Economics* **101**, 140–162.

Makariou, D., Barrieu, P. & Tzougas, G. (2021), 'A finite mixture modelling perspective for combining experts' opinions with an application to quantile-based risk measures', *Risks* **9**(6).

McCullagh, J. & Nelder, P. (1983), 'Generalized linear models, chapman & hall'.

McKinsey & Company (2021), 'Insurance 2030— the impact of AI on the future of insurance'. Accessed = 2021-11-02.
URL: *https://www.mckinsey.com/industries/financial-services/our-insights/insurance-2030-the-impact-of-ai-on-the-future-of-insurance#*

Meier, U. B. & Outreville, J. F. (2003), The reinsurance price and the insurance cycle, *in* 'Presentado al Seminar of the European Group of Risk and Insurance Economists (EGRIE), Zurich (Suiza)'.

Michel-Kerjan, E. & Morlaye, F. (2008), 'Extreme events, global warming, and insurance-linked securities: How to trigger the "tipping point"', *The Geneva Papers on Risk and Insurance-Issues and Practice* **33**(1), 153–176.

Mobley, W., Sebastian, A., Blessing, R., Highfield, W. E., Stearns, L. & Brody, S. D. (2021), 'Quantification of continuous flood hazard using random forest classification and flood insurance claims at large spatial scales: a pilot study in southeast texas', *Natural Hazards and Earth System Sciences* **21**(2), 807–822.

Munkin, M. K. & Trivedi, P. K. (1999), 'Simulated maximum likelihood estimation of multivariate mixed-Poisson regression models, with application', *The Econometrics Journal* **2**(1), 29–48.

Murphy, K. P. (2012), *Machine learning: a probabilistic perspective*, MIT press.

NAIC (2021), 'Sidecars'. Accessed = 2021-09-10.
URL: *https://content.naic.org/cipr_topics/topic_sidecars.html*

Neufeld, A. & Witten, D. (2021), 'Discussion of breiman's" two cultures": From two cultures to one', *Observational Studies* **7**(1), 171–174.

Nikoloulopoulos, A. K. (2013), Copula-based models for multivariate discrete response data, *in* 'Copulae in Mathematical and Quantitative Finance', Springer, pp. 231–249.

Nikoloulopoulos, A. K. (2016), 'Efficient estimation of high-dimensional multivariate normal copula models with discrete spatial responses', *Stochastic environmental research and risk assessment* **30**(2), 493–505.

Nikoloulopoulos, A. K. & Karlis, D. (2010), 'Regression in a copula model for bivariate count data', *Journal of Applied Statistics* **37**(9), 1555–1568.

Nowak, T., Richter, A. & Hartung, T. (2014), *Non-Life Insurance-Linked Securities: Risk and Pricing Analysis*, Münchener Reihe, Verlag Versicherungswirtschaft.

Park, E. S. & Lord, D. (2007), 'Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity', *Transportation Research Record* **2019**(1), 1–6.

Parsons, C. (2003), 'Moral hazard in liability insurance', *The Geneva Papers on Risk and Insurance-Issues and Practice* **28**(3), 448–471.

Pechon, F., Denuit, M. & Trufin, J. (2019), 'Multivariate modelling of multiple guarantees in motor insurance of a household', *European Actuarial Journal* **9**(2), 575–602.

Pechon, F., Denuit, M. & Trufin, J. (2021), 'Home and motor insurance joined at a household level using multivariate credibility', *Annals of Actuarial Science* **15**(1), 82–114.

Pechon, F., Trufin, J., Denuit, M. et al. (2018), 'Multivariate modelling of household claim frequencies in motor third-party liability insurance', *Astin Bulletin* **48**(3), 969–993.

Peel, D. & MacLahlan, G. (2000), 'Finite mixture models', *John & Sons* .

Prudential Regulation Authority and Society of Lloyd's (2013), 'The PRA and the Society of Lloyd's co-operation agreement'. Accessed = 2021-10-18.
URL: *https://assets.lloyds.com/assets/pdf-regulation-of-lloyds-pra-and-lloyds-cooperation-agreementdecember2013/1/pdf-regulation-of-lloyds-PRA-and-Lloyds-cooperation-agreementDecember2013.pdf*

Punter, A. (2003), *Risk Financing and Management*, Institute of Financial Services.

PwC (2021), 'Seizing the ILS opportunity'. Accessed = 2021-10-02.
URL: *https://www.pwc.co.uk/services/risk/insights/seizing-ils-opportunity.html*

Rheinlander, T. & Sexton, J. (2011), *Hedging Derivatives*, Advanced series on statistical science & applied probability, World Scientific.

Rigby, R. & Stasinopoulos, D. (2010), 'A flexible regression approach using gamlss in r (2010)'.

Risk Management Solutions (2012), 'Cat bonds demystified: RMS guide to the asset class'.
URL: *https://forms2.rms.com/rs/729-DJX-565/images/cm_cat_bonds_demystified.pdf*

Roussos, J., Bradley, R. & Frigg, R. (2021), 'Making confident decisions with model ensembles', *Philosophy of science* **88**(3), 439–460.

Russell, S. & Norvig, P. (2010), 'Artificial intelligence: A modern approach. third edit', *Prentice Hall. doi* **10**, B978–012161964.

Rustam, Z. & Saragih, G. (2021), Prediction insolvency of insurance companies using random forest, *in* 'Journal of Physics: Conference Series', Vol. 1752, IOP Publishing, p. 012036.

Sclove, S. L. (1987), 'Application of model-selection criteria to some problems in multivariate analysis', *Psychometrika* **52**(3), 333–343.

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

Shmueli, G. (2010), 'To explain or to predict?', *Statistical science* **25**(3), 289–310.

Silva, A., Rothstein, S. J., McNicholas, P. D. & Subedi, S. (2019), 'A multivariate Poisson-log normal mixture model for clustering transcriptome sequencing data', *BMC bioinformatics* **20**(1), 1–11.

Smith, M. L. & Kane, S. A. (1994), The law of large numbers and the strength of insurance, *in* 'Insurance, risk management, and public policy', Springer, pp. 1–27.

Staudt, Y. & Wagner, J. (2021), 'Assessing the performance of random forests for modeling claim severity in collision car insurance', *Risks* **9**(3).
URL: *https://www.mdpi.com/2227-9091/9/3/53*

Stein, G. Z. & Juritz, J. M. (1987), 'Bivariate compound Poisson distributions', *Communications in Statistics-Theory and Methods* **16**(12), 3591–3607.

Stein, G. Z., Zucchini, W. & Juritz, J. M. (1987), 'Parameter estimation for the Sichel distribution and its multivariate extension', *Journal of the American Statistical Association* **82**(399), 938–944.

Stone, M. (1979), 'Comments on model selection criteria of akaike and schwarz', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 276–278.

Sutton, R. S. & Barto, A. G. (2018), *Reinforcement learning: An introduction*, MIT press.

Swiss Re (2019), 'Confronting the cost of catastrophe'. Accessed = 2021-09-05.
URL: *https://www.swissre.com/risk-knowledge/mitigating-climate-risk/natural-catastrophes-affecting-more-people-what-can-do-to-help.html*

Swiss Re (2021), 'The essential guide to reinsurance'. Accessed = 2021-08-30.
URL: *https://www.swissre.com/dam/jcr:d06472ab-2625-48cf-8b4e-7c7ac8aa63f0/The-essential-guide-to-reinsurance.pdf*

Swiss Re Institute (2020), 'Machine intelligence in insurance: insights for end-to-end enterprise transformation'. Accessed = 2021-09-02.
URL: *https://www.swissre.com/dam/jcr:dbf95c92-4f1a-496a-aefc-0b9807178d59/sigma5_2020_en.pdf*

Swiss Re Institute (2021), 'sigma 1/2021 - natural catastrophes in 2020'. Accessed = 2021-09-02.
URL: *https://www.swissre.com/dam/jcr:ebd39a3b-dc55-4b34-9246-6dd8e5715c8b/sigma-1-2021-en.pdf*

The Economist (2009), 'The advantages of contingent capital'. Accessed = 2021-07-11.
URL: *https://www.economist.com/free-exchange/2009/10/14/the-advantages-of-contingent-capital*

Tzougas, G., Vrontos, S. & Frangos, N. (2014), 'Optimal bonus-malus systems using finite mixture models', *Astin Bulletin* **44**(2), 417–444.

Vermaelen, T. & Wolff, C. C. (2010), *Contingent capital: the case for COERCs*, Citeseer.

Warr, C. (2016*a*), *LM1 London Market insurance essentials*, The Chartered Insurance Institute.

Warr, C. (2016*b*), *LM2 London Market insurance principles and practices*, The Chartered Insurance Institute.

Winkelmann, R. (2008), *Econometric analysis of count data*, Springer Science & Business Media.

Wuthrich, M. V. & Merz, M. (2021), 'Statistical foundations of actuarial learning and its applications', *Available at SSRN* .

Zhan, X., Aziz, H. A. & Ukkusuri, S. V. (2015), 'An efficient parallel sampling technique for multivariate Poisson-lognormal model: Analysis with two crash count datasets', *Analytic methods in accident research* **8**, 45–60.

Zhou, Z.-H. (2019), *Ensemble methods: foundations and algorithms*, Chapman and Hall/CRC.

Zimmer, D. M. & Trivedi, P. K. (2006), 'Using trivariate copulas to model sample selection and treatment effects: application to family health care demand', *Journal of Business & Economic Statistics* **24**(1), 63–76.