A Bayesian Approach to Assessing the Risk Premium on Catastrophe Bond Derivatives at Issuance

Submitted by

Dickson K. Nkwantabisa

Doctor of Business Administration in Finance Program

In partial fulfillment of the requirements

For the degree of Doctor of Business Administration in Finance

Sacred Heart University, Jack Welch College of Business and Technology

Fairfield, Connecticut

Date: September 24, 2021

Dissertation Supervisor: Dr. Loran Chollete

Signature: ____________________________

Committee Member: Dr. Arpita Chatterjee

Signature: ____________________________

Committee Member: Dr. W. Keener Hughen

Signature: ____________________________
A Bayesian Approach to Assessing the Risk Premium on Catastrophe Bond Derivatives at Issuance

Submitted by

Dickson K. Nkwantabisa

Doctor of Business Administration in Finance Program

In partial fulfillment of the requirements

For the degree of Doctor of Business Administration in Finance

Sacred Heart University, Jack Welch College of Business and Technology

Fairfield, Connecticut

Date: July 22, 2021

Dissertation Supervisor: Dr. Loran Chollete Signature:

Committee Member: Dr. Arpita Chatterjee Signature:

Committee Member: Dr. W. Keener Hughen Signature:

Committee Member: Dr. Vasudev Das Signature:
A Bayesian Approach to Assessing the Risk Premium on Catastrophe Bond Derivatives at Issuance

Dickson K. Nkwantabisa [July 2021]

Abstract

Catastrophe (CAT) bond pricing is a challenging task due to the uncertainty inherent in the incomplete market setting in which they operate as such various pricing approaches have been proposed. In this paper, we offer an alternative Bayesian methodology which is a natural approach in the context of uncertainty. Our Bayesian model is highly flexible and can be implemented under different model assumptions without losing generalization. We develop an entire Bayesian framework to model the two fundamental sources of risks in CAT bond pricing – catastrophe and interest rate risks. Using a Hierarchical Dirichlet Process model (Teh et al., 2006), we model the collective catastrophe risk via a model-based clustering approach. Interest rate risk is modeled as a CIR process via the Bayesian approach. We can account for parameter and model uncertainties through these models, which leads to more reliable CAT bond prices. Finally, we use the models to find prices, present values, and risk premia of CAT bond contracts corresponding to different grouped risk profiles via several numerical examples.

Keywords and phrases: CAT bond, Bayesian, Hierarchical Dirichlet, Markov chain Monte Carlo, risk-neutral valuation, Cox Ingersoll Ross (CIR) model, entropy maximization

JEL classification: C11, G13, C14, E43, D81

This version: July 22, 2021
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>1</td>
</tr>
<tr>
<td>1.0 INTRODUCTION</td>
<td>4</td>
</tr>
<tr>
<td>2.0 LITERATURE REVIEW</td>
<td>7</td>
</tr>
<tr>
<td>2.1 CAT BOND MECHANISM</td>
<td>10</td>
</tr>
<tr>
<td>3.0 BAYESIAN COLLECTIVE RISK MODEL</td>
<td>11</td>
</tr>
<tr>
<td>3.1 A BAYESIAN PRIMER</td>
<td>11</td>
</tr>
<tr>
<td>3.2 GENERAL COLLECTIVE RISK MODEL</td>
<td>12</td>
</tr>
<tr>
<td>3.3 BAYESIAN COLLECTIVE RISK MODEL</td>
<td>13</td>
</tr>
<tr>
<td>3.3.1 HIERARCHICAL COMPOUND COLLECTIVE RISK MODEL</td>
<td>17</td>
</tr>
<tr>
<td>3.3.2 HIERARCHICAL DIRICHLET PRIOR MODEL SPECIFICATION</td>
<td>17</td>
</tr>
<tr>
<td>3.4 PARAMETER CALIBRATION OF THE BAYESIAN CATASTROPHE MODEL</td>
<td>20</td>
</tr>
<tr>
<td>3.4.1 DATA</td>
<td>20</td>
</tr>
<tr>
<td>3.4.2 BAYESIAN INFERENCE VIA MCMC</td>
<td>21</td>
</tr>
<tr>
<td>3.4.3 CONVERGENCE DIAGNOSTICS</td>
<td>23</td>
</tr>
<tr>
<td>3.4.4 BAYESIAN PREDICTIVE INFERENCE VIA MCMC</td>
<td>24</td>
</tr>
<tr>
<td>4.0 STOCHASTIC INTEREST MODEL</td>
<td>26</td>
</tr>
<tr>
<td>4.1 CIR MODEL</td>
<td>26</td>
</tr>
<tr>
<td>4.1.1 DATA AUGMENTATION</td>
<td>27</td>
</tr>
<tr>
<td>4.2 BAYESIAN CIR MODEL</td>
<td>28</td>
</tr>
<tr>
<td>4.2.1 BAYESIAN CIR MODEL SPECIFICATION</td>
<td>28</td>
</tr>
<tr>
<td>4.2.2 SAMPLING ALGORITHM</td>
<td>31</td>
</tr>
<tr>
<td>4.2.3 FORECASTING</td>
<td>32</td>
</tr>
<tr>
<td>4.3 PARAMETER CALIBRATION OF THE BAYESIAN CIR MODEL</td>
<td>32</td>
</tr>
<tr>
<td>5.0 CONTINGENT CLAIM PRICING MODEL</td>
<td>35</td>
</tr>
<tr>
<td>5.1 PRICING MODEL</td>
<td>36</td>
</tr>
<tr>
<td>5.2 PAYOFF FUNCTION</td>
<td>37</td>
</tr>
<tr>
<td>5.2.1 AN ILLUSTRATIVE EXAMPLE</td>
<td>38</td>
</tr>
<tr>
<td>6.0 RISK-NEUTRALIZATION OF CAT BOND PRICES</td>
<td>42</td>
</tr>
<tr>
<td>6.1 PRICING MECHANISM UNDER THE ENTROPY PRINCIPLE</td>
<td>42</td>
</tr>
<tr>
<td>6.1.1 AN ILLUSTRATIVE EXAMPLE</td>
<td>44</td>
</tr>
</tbody>
</table>
6.2 Risk Premium Assessment ................................................................. 46

7.0 Concluding Remarks .................................................................... 48

Bibliography .................................................................................... 49

Appendix A ...................................................................................... 51

Finding the Maximum Entropy by the Process of Lagrange Multipliers ........................................................................ 51
1.0 Introduction

Natural catastrophes are on the rise in recent decades, and (re-) insurance companies and countries have limited capacity to digest a fraction of the catastrophe risk. Using insurance-linked securities (ILS) as an alternative risk transfer (ATR) mechanism helps insurers raise additional capital. One of such ILS is catastrophe bonds. Catastrophe (CAT) bonds are ILS securities with a payoff linked to insurance risk. In other words, they are derivatives with the underlying claims process contingent on a set of specified trigger mechanisms. This new asset class is still young. Since its inception in 1994, CAT bonds and other ILS have grown with remarkable success. With outstanding capital increasing steadily, the ILS market capitalization is about $41.8 billion as of 2020\(^1\). High demand and supply have ensured the exceptional high market. Insurers, countries, and regional governments on the sell side; see CAT bonds issuance as a viable option to raise capital to expand their capacity to cope with the ever-increasing catastrophe risk due to climate change. CAT bonds appeal to investors on the buy-side due to their zero beta (see, e.g., Froot et al., 1995, Cummins, 2009). Usually, the underlying catastrophe risk in CAT bonds has a low correlation with the market, although this may not always be the case (see, e.g., Gurtler et al., 2016, Froot K., 2001). Moreover, in a historically low-interest-rate environment, CAT bonds have exceptional appeal to investors due to their high yield.

As the CAT market expands as an alternative investment for investors, it has become increasingly important for researchers and investors to understand how these financial instruments are priced by the (re-) insurer and the secondary market. The methods prescribed in the literature to price these instruments vary, far from unified and sometimes contradicting one another. One of the main challenges in pricing is that CAT bonds operate in an incomplete market. These bond prices do not admit a unique equivalent martingale measure. To be able to price CAT bonds like any other traditional derivative requires a complete market. An incomplete market presents many possible prices for an asset corresponding to different risk-neutral measures (Cox and Pedersen, 2000).

For this reason, statistically estimating the underlying physical risk metric using observed data becomes the preferred approach. Furthermore, traditional assumptions applicable to derivatives are not generalizable to CAT bonds due to the underlying stochastic contingent process

\(^1\) See https://www.artemis.bm/dashboard/
(Burnecki et al., 2005). The premises of asymptotic normality as applied to derivatives do not hold here. There are further statistical challenges associated with the moments of power-law distributions making it impossible to use traditional pooling methods and central limit theorems (Burnecki et al., 2005). Throughout the literature, the typical approach to modeling the catastrophe risk component of CAT bonds has been through numerical analysis. Although the numerical method is generally the best to portray the behavior of complex systems, incorrectly specifying the interdependencies among the variables or type of distribution can lead to incorrect results. Even simpler approaches by necessity neglecting interdependence of some variables involved are challenging. Fitting distributions to variables such as the frequency of major hurricanes is a common approach. It still leaves room for uncertainty even as far as the choice of the probability distribution to be fitted. Extreme events are highly unpredictable, so fitting a specific probability distribution to its occurrence without allowing for flexibility can lead to an incorrect model prediction of their future occurrence. For instance, Ma and Ma (2013) proposed a valuation framework for CAT bonds in which the catastrophe event intensity (frequency) and aggregate loss were is fitted with a nonhomogeneous Poisson process (NHPP) and General Extreme Value (GEV) distributions separately. Ma, Ma, and Xiao (2017) reject the idea of using NHPP to model the arrival process (intensity) of catastrophic events and instead propose a doubly stochastic Poisson process. They argue that NHPP with a deterministic intensity function is insufficient to model unanticipated catastrophic risk events due to their stochastic nature.

The pricing of CAT bonds can be thought of as a contingent claim. Their payoff is contingent on an underlying catastrophe process not exceeding a given threshold stipulated in the bond contract. When a payoff on ILS is contingent on a possible occurrence of insured catastrophe losses, catastrophe modeling is an invaluable tool for investors in analyzing the risk of the securities to determine the price at which they would be willing to assume this risk. Superior modeling skills become a competitive advantage in a market that remains inefficient and suffers from information asymmetry. The CAT bond market is opaque, as insurers make available scanty information during the issuance of bonds. Furthermore, most of these bond issuance are private placements, and information on them is not readily available. The lack of information can mask potential layers of high risk and unfair pricing of the risk transferred to the investor.

Under a frequentist approach, some researchers (e.g., Cox and Pedersen (2000), Lee and Yu (2002), Ma and Ma (2013), Tang and Zhongyi, 2019) have attempted to derive CAT bond
pricing models with a link to industry loss trigger indices. These proposed models have little or no use to a buy-side investor. The former approaches assume that all types of catastrophe risk data recorded on an industry loss index have the same characteristics. While this assumption simplifies the catastrophe modeling approach, the results have no practical implementation. CAT bond contracts are peril/catastrophe specific. For example, a CAT bond contract may cover single peril (like Tornados) or multiple perils (like Tornados, Severe storms, and Winter storms). The approach of considering all types of catastrophes on an industry loss index to bear the same risk profile places a practical limitation. We argue that even though different type perils are listed on a loss index due to their having passed a minimum catastrophic loss cost, they may exhibit distinct or grouped characteristics, hence different risk profiles.\(^2\) A catastrophe loss index is akin to any stock index such as the S&P 500, which has heterogeneous groups in terms of sectors. We hope to price CAT bonds based on their individual or group risk profiles by accounting for the natural groupings. Consequently, we can assess the fair prices, present values of bond contracts, and their expected risk premia based on their risk profile.

CAT bonds modeling involves contingent cash flows based on a probability that a rare event will occur within the life term of a bond contract. Such a probabilistic problem naturally yields itself Bayesian modeling techniques. However, as at the writing of this study, there are no published papers that apply Bayesian methods to CAT bond pricing to the best of our knowledge. While past research has investigated the effects of catastrophe risk on CAT bond prices in the secondary market (e.g., Lane (2000), Papachristou (2011), Ahrens et al. (2014), Gürtler et al. (2016), Braun (2016)), no empirical research has yet examined the risk-premium of CAT bonds at the time of issue. This study hopes to fill the lacuna in the literature.

CAT bonds have two independent sources of risk – catastrophe and interest risks. We develop two separate models for the two sources of risk and unify the results in our asset pricing framework. Notably, we apply a Hierarchical Dirichlet Process (HDP) model to model the catastrophe risk and Cox Ingersoll Ross (CIR) model for the interest rate process. To assess the risk premia and present value of CAT contracts corresponding to different risk profiles, we utilize the entropy maximization principle to derive risk-neutral measures. We demonstrate the use of our models through several numerical examples. Our models reveal four distinct groups of risk.

\(^2\) Usually loss indices only include catastrophic losses to the industry of at least $25 million and affecting a significant number of insurers and policyholders.
profiles. Based on these groupings, we assess the pricing evolution of CAT bond prices for increasing maturity and different trigger thresholds. We observe that CAT bonds value decreases as the threshold level decreases and time to maturity increases within all four risk groups. We observe higher CAT bond values across the distinct groups of perils for groups with a relatively lower number of claims. We also notice that risk groups with a higher probability of breaching a set threshold have higher expected risk premiums.

The remaining of this paper consists of six sections. Section 2 outlines some of the modeling approaches proposed in the literature for CAT pricing and the general pricing mechanism. We get a detailed outline of the Bayesian theory and modeling approach to the catastrophe risk in Section 3. The stochastic interest rate model methodology is discussed in Section 4. We introduce the product probability measure model for contingent claim pricing in Section 5 with some illustrative examples. To assess the risk premia, we thoroughly discuss and implement the entropy maximization principle and show some numerical examples in Section 6. Finally, we conclude this paper with some closing remarks in Section 7.

2.0 Literature Review

In the literature, various pricing frameworks have been proposed to design CAT bonds in the primary and secondary markets. Some early works treat CAT bonds as zero-beta security. In other words, the correlation between the underlying catastrophe risk and the market is zero. Therefore, investors should earn a zero risk premium (see, e.g., Froot (1995), Cummins and Geman (1995), Cox and Pedersen (2000), Lee and Yu, (2002) and Ma and Ma (2013). Proponents of this valuation approach follow Merton (1976), who argues that localized jumps in an asset price may be due to asset-specific events and are uncorrelated with the market. This assumption allows for the valuation of CAT bond prices under risk-neutral measures. The implication is that the aggregate loss processes (i.e., intensity and severity of losses) retain their original distributional characteristics after been transformed from the physical probability measure to the risk-neutral measure (Lee and Yu, 2002). A similar argument supported by the zero risk premium is that catastrophic risk has a marginal influence on the overall economy and therefore does not pose a
“systematic risk” to the market (Tang and Zhongyi, 2019). However, there is empirical evidence to suggest that catastrophic events may have a substantial systematic effect on the market (see, e.g., Gurtler et al., 2016). The current COVID-19 outbreak (which can be considered a natural catastrophe) had U.S. stocks tumble 11% in five days\(^3\). Pandemic Emergency Financing bonds issued worth $425 million by the World Bank went into default when the World Health Organization declared the COVID-19 as a pandemic\(^4\).

Vaugirard (2003) develops an arbitrage model (based on the *Arbitrage Portfolio Theory (APT)*) to price insurance-linked securities which incorporate catastrophe events in a stochastic interest rate environment. Notwithstanding the incomplete market, Vaugirard (2003) attempts to vindicate the arbitrage approach to pricing CAT bonds by asserting that the catastrophic jump risk index can be mimicked by instruments such as energy and power derivatives. While the replication of the interest rate dynamics is possible, the analogous assumption concerning catastrophe risk requires more justification. An extension of this pricing approach is afforded by Nowak and Romaniuk (2013), who price payoffs is contingent on only catastrophe risk. Jarrow (2010) proposes a closed-form model consistent with APT on the assumption of an arbitrage-free LIBOR term structure of interest rates.

Not so many studies have been done using this econometric approach. The econometric approach is usually applied to issue CAT bonds already trading in the secondary market. The earliest work can be attributed to Lane (2000), who presents a two-factor model using 16 catastrophe bonds issued in 1999. He models the expected excess return (EER) as a function of the probability of first loss (PFL) and the conditional expected loss (CEL). The EER or risk premium is modeled with a Cob-Douglas function to capture the asymmetrical nature of the catastrophe losses and PFL. The spread on the bond is expressed as the sum of EER and expected loss (EL). The price of a CAT bond will be LIBOR plus the spread. This is the first model developed to understand the behavior of the CAT bond market. Papachristou (2011) utilizes a generalized additive model to examine the factors that affect the CAT bond premiums on 192 CAT bonds launched between 2003 and 2008. Ahrens et al. (2014) examine the impact of hurricane Katrina on CAT bond prices using the theoretical framework of Lane (2000). They utilize a\


Bayesian treed approach to estimate the parameters of Lane’s pricing framework. Their results show that during the 2005 hurricane season, the investment-grade rating increases the impact of the conditional expected loss. They posit that investors who demand highly rated bonds are more concerned with possible losses than subinvestment grade bond investors. Gürtler et al. (2016) examine how financial crises and natural catastrophes affect CAT prices based on bond-specific information and macroeconomic factors. Braun (2016) tests several hypotheses on the factors that affect CAT premiums in the primary market and develops a robust forecasting model for predicting the CAT bond spreads. Braun (2016) points out other factors such as insurance underwriting cycles, rating class, issuer, catastrophe risk modeler, territory covered, peril, and trigger type in addition to expected loss as relevant drivers of CAT premiums in the primary market. Gomez and Carcamo (2014) develop a multifactor spread model in a panel data setting to evaluate the drivers of CAT spreads in the secondary market.

Probability transforms are distortion operators that combine both actuarial and financial pricing theory. The Wang transform has been widely used in ILS pricing (see, e.g., Hamada and Sherris, 2003, Kijima and Murimachi, 2008, Li et al., 2013). However, Pelsser (2008) shows that the Wang transform cannot lead to consistent prices with arbitrage-free prices for a general stochastic process. As a result, he argues that it cannot be used as a universal framework for pricing insurance and financial risk. Wang (2000) introduces a universal framework (Wang transform) for pricing financial and insurance risk through a transfer and correlation measure that extends the CAPM to price all kinds of assets and liabilities. He broadens the CAPM to risk with non-normal distribution to obtain a new parameter called the market price of risk, which is analogous to the Sharpe ratio in the case of normally distributed returns. The two-factor Wang transform inflates probability densities to cope with adverse expectations while deflating them to accommodate favorable outcomes (Wang, 2000, 2004). In other words, it accounts for extreme tail risk in the probability distribution. This is akin to the “volatility smile” in option prices. As a result, it incorporates a form of risk adjustment.

The two-step valuation approach is more recent, and it can be attributed to Pelsser and Stadje (2014). They develop a general valuation framework that is consistent with market prices and actuarial risk pricing principles. Dhaene et al. (2017) propose a “fair valuation” approach (analogous to Pelsser and Stadje’s system) which is both market-consistent and actuarial. Tang and Zhongyi (2019) develop a CAT bond pricing model based on the two-step valuation approach.
They propose a product pricing measure that combines a distorted probability measure related to the underlying catastrophe risk and a risk-neutral probability measure incorporating interest rate risk. In their product measure model, the two sources of risks (i.e., catastrophe and interest rate risks) are modeled separately and integrated to form the pricing framework for the CAT bond. Tang and Zhongyi’s (2019) approach bears some semblance to the APT and probability transform approach.

2.1 CAT Bond Mechanism

Financial institutions structure CAT bonds under the typical characteristics of traditional bonds. However, the probabilistic scenario is given by the trigger event, and the insurance purpose of the bond changes the way transactions are conducted.

![Diagram of CAT bond transaction]

Figure 1: Structure of a CAT bond transaction

Due to the complexity and novelty of these instruments, all current investors are sophisticated institutional investors such as government agencies, pension, and hedge funds. A sponsor has to establish a Special Purpose Vehicle (SPV) to raise money on the bond market. The SPV is responsible for receiving all the income from the sponsor’s CAT bond issue and premiums. In return, it pays back principal and interest to the investor and loss payments to the sponsor should the stipulated catastrophe event be triggered (as shown in Figure 1 above). The sole purpose of the SPV is to eliminate credit default risk and serves to protect the investor and the risk carrier. The
risk carrier does not issue the CAT bond itself but instead sets up a particular company for this purpose (i.e., the SPV). To the investor, the SPV means that other risks, such as the risk carrier going bankrupt, are limited. For the sponsor, the SPV implies that the funds are earmarked and available if the natural catastrophe the bonds cover occurs. In Figure 1, the investor places capital in the SPV, which is in turn invested in low-risk money market instruments. The most common are short-term government bonds. For the term maturity of the bond, the investor receives income in the form of a coupon that is usually paid quarterly. It consists of a small return on the investment plus the premium paid by the sponsor. At maturity, the principal is paid back to the investor, assuming the conditions for the trigger event have not occurred. If the trigger event occurs within the bond term, the trust account is liquidated, and the proceeds are used to compensate the risk carrier.

3.0 Bayesian Collective Risk Model

3.1 A Bayesian Primer

Probabilistic modeling often involves some data $x_1, x_2, \ldots, x_n$, latent variables $y_1, y_2, \ldots, y_n$ and a parameter of interest, $\theta$, that might help establish the relationship between variables and data. They are typically generative models of data that draw inferences on latent variables given observed data such as

$$P(y_1, y_2, \ldots, y_n | x_1, x_2, \ldots, x_n, \theta) = \frac{P(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n | \theta)}{p(x_1, x_2, \ldots, x_n | \theta)}.$$ 

The prediction of some future data point given observed data and parameter can be obtained as $P(x_{n+1}, y_{n+1} | x_1, x_2, \ldots, x_n, \theta)$. The Bayesian approach to probabilistic modeling differs slightly (in terms of formulation) with the introduction of a prior distribution, $\pi(\theta)$ based on our knowledge about the data distribution. Both inference and learning about parameters are carried in one step via the posterior distribution

$$P(y_1, y_2, \ldots, y_n, \theta | x_1, x_2, \ldots, x_n) = \frac{P(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n | \theta) \pi(\theta)}{p(x_1, x_2, \ldots, x_n)}.$$
The posterior distribution can be obtained by sampling algorithms such as Expectation-Maximization (EM), Markov Chain Monte Carlo (MCMC), or variation inference. We can predict new data points by marginalizing over our latent variables and parameters $P(x_{n+1}|x_1, x_2, ..., x_n) = \int P(x_{n+1}|\theta)P(\theta|x_1, ..., x_n)d\theta$.

The Bayesian approach yields both point estimates (i.e., expected values) and a whole distribution of the estimated parameter. The benefits of such an approach are immediately evident. Using prior distributions, one can include several layers of randomness in a Bayesian model, which highlights uncertainty regarding parameter estimates. In addition, one can update estimates of parameters when new data becomes available, thereby enhancing the reliability of our estimates.

3.2 General Collective Risk Model

In the classical actuarial literature, catastrophe risks are characterized by two independent stochastic processes: the number of claims (intensity) – which counts the claims, and the claim amounts (severity) – which determine the amount when a claim is reported. Consider an industry trigger loss index as a portfolio of catastrophe risks that need to be modeled. This portfolio has two primary sources of randomness – the claim number process and the claim size process. The aggregate claims process is, therefore, a compound process of these two sources of risks. Let $i$ be peril types, $i = 1, 2, ..., I$ and denote $t = 1, ..., T$, the time interval for the arrival of a catastrophe event (quarterly observations). Assume $N_{i,t}$ (with $N(0)=0$) to be the random number of claims within an arbitrary time interval $[0, t]$ with corresponding $k$th claim sizes as, $X_{i,k}$ $k = 1, 2, ...$ for each peril type. We assume that sequences $\{X_{i,1}, X_{i,2}, \ldots\}$ are independent and identically distributed nonnegative random variables with a common distribution $F(x) = Pr\{X_1 \leq x\}$. Further, assume that the two processes $N_{i,t}$ and $X_{i,t}$ are independent for all $i$ and $t$. For each of these $N_{i,t}$ claims, an insurance company also records the claim amounts, $\{X_{i,1}, X_{i,2}, \ldots\}$ generated. Therefore, the aggregate claims up to a fixed time $t \geq 0$ constitute the collective risk

$$S_{i,t} = X_{i,1} + \cdots + X_{i,N(t)} = \sum_{j=1}^{N_{i,t}} X_{i,j}, \quad t \geq 0, i = 1, 2, ..., I, j = 1, ..., N$$ (3.1)
with \( S(t) = 0 \) if \( N(t) = 0 \). The sequence of \( \{S_t\}_{t \geq 0} \) forms another stochastic process which is called a compound or random sum. The moment generating function of \( S \) conditioning on \( N \) is \( m_S(t) = E[e^{tS}] = m_N(\log X(t)) \). For the distribution of \( S \), one can find

\[
F_S(x) = Pr[S \leq x] = \sum_{n=0}^{\infty} Pr[S \leq x | N = n] Pr[N = n] = \sum_{n=0}^{\infty} F_X^n Pr[N = n] \tag{3.2}
\]

where \( F_X^n = Pr(X_1 + X_2 + \cdots + X_n \leq x) \) is the n-fold convolution of the distribution of claim sizes.

### 3.3 Bayesian Collective Risk Model

The independence assumption in the catastrophe risk process may be restrictive in many applications. For instance, some catastrophes such as seasonal tornados in North America follow a pattern, and as such, claim amount and claim numbers may not necessarily be independent. In such a situation, the i.i.d assumption needs to be relaxed. For example, Shao et al. (2017) show that allowing for dependency between the claim amounts and the inter-arrival times in the claims process via a semi-Markov risk model can obtain fairer CAT bond prices.

We argue that the catastrophe risk measure should include additional risk information about the process other than intensity and severity. CAT bond derivatives are often priced using historical data on insured catastrophe losses indices that contain different peril types. An example of such a data set is the catastrophe loss index of the Property Claim Service (PCS) which comprises all catastrophe events in the U.S. with an insured cost of more than USD 25M. Such a loss index has heterogeneous aggregate losses due to the natural grouping of the different peril types. Modeling would serve practical purposes if one could price CAT bonds based on their individual or grouped risk profiles.

Furthermore, CAT bonds contract designs have become sophisticated. Insurers and reinsurers bundle multiple catastrophe risks (e.g., tropical cyclones, earthquakes, wildfire) in one bond contract. The assessment of the probability of a trigger event of such multi-peril bonds is an uneasy feat. These bond contracts may mask high layers of risk in CAT bond prices. It is increasingly difficult for investors to assess the fair value of risk of buying such financial instruments.
Therefore, it becomes imperative that a collective risk model (Migon and Moura, 2006) usually be applied in practice. The collective risk model under a fully hierarchical Bayesian framework with Dirichlet prior allows us to incorporate two more covariates to the catastrophe risk model -- seasonality and peril type. This hierarchical Dirichlet model architecture provides a practical, flexible, and broad generalization of the collective risk model in pricing CAT bonds. Under this framework, one can disentangle the effects of single and multi-peril risk profiles and assess how they affect bond contracts corresponding to different risk exposures without losing generalization.

*Claim Number Process:* several distributions can be used to model $N_{i,t}$ such as the Poisson, Negative Binomial, and Gamma. This paper employs an extension of the Poisson process (nonhomogeneous Poisson process) to accommodate dependency peril-specific characteristics in catastrophe frequency. We explicitly incorporate peril- and seasonal-dependency in the Poisson process via a likelihood specification as

$$N_{i,s} | \lambda_{i,s} \sim \text{Poisson}(\lambda_{i,s}), \quad \lambda_{i,s} > 0$$

$$\log(\lambda_{i,s}) | \alpha_i, \beta_2, X_s = \alpha_i + \beta_2 X_s$$

where $\lambda_{i,s}$ is the average number of claims per quarter for particular perils for a fixed unit of time. The intensity of the arrival process, $\lambda_{i,s}$ is modeled as a Poisson regression. The deep reasoning is that we want to capture the seasonality and peril-specific characteristics that may influence the final aggregate loss for different perils. Notably, the random intercept, $\alpha_i$ allows claims intensity to be correlated through peril type$^5$. The quarterly seasonal patterns of catastrophe are directly built into the model via, $X_s$. This approach is appropriate since the underlying risk in our portfolio of catastrophes shows discernable patterns. Some perils exhibit higher occurrence relative to others during June through August, as presented in Figure 3.1.

$^5$ $\alpha_i$ is a peril-specific parameter which indicates the average intensity of each peril type $i$. 
Claim Size Process: to model the claim amount involves selecting an appropriate distribution for $F(x)$. One has to fit several probability density functions to the observed loss data and choose the distributions that best describe the data generating process. In actuarial science, an extreme or rare event occurs with a low probability but causes severe damage. Distributions with long- and heavy-tails are typically used to model these phenomena to assess tail risks. Since the choice of distribution can influence the CAT bond prices, we consider several heavy-tailed distributions (within the domain $\mathbb{R}_+$) via nonparametric tests. In this paper, we utilize the Kolmogorov-Smirnov (K-S), and Anderson-Darling (A-D) tests for our model selection (see, e.g., Ma and Ma (2013)).

For a random sample, the goodness of fit exercise is to test, $H_0$: sample comes from a population with a theoretical distribution $G(x)$.

The K-S (Chakravarti et al., 1967) test is technically distribution-free and nonparametric. In other words, K-S does not assume the distribution of the data. The K-S is used to test if a sample comes from a population of a specified distribution. Given $N$ ordered random samples $Y_1, Y_2, \ldots, Y_N$, the empirical cumulative distribution function (ECDF) is defined as $F_N(x) = N^{-1} \times \left[\text{number of observations} \leq x\right]$. The K-S statistic $D$, is based on the maximum distance between $F_N(x)$ and $G(x)$ for all values of $x$. The closer the $D$ statistic is to 0, the more likely the two samples are drawn from the same distribution. When comparing across several distributions, the distribution with the largest p-value is the best fit. The A-D (Stephens, 1974) test modifies the K-
S test and gives more weight to the distribution’s tails than the K-S test. Further, we assess the performance of the selected distributions, with the best fit model having the lowest AIC and BIC values. Tables 3.1 – 2 report the test statistic and performance for the different distributions tested.

Table 3.1: Test statistics for different distributions (p-values are in parenthesis)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Weibull</th>
<th>Inv. Gamma</th>
<th>Pareto</th>
<th>Lognormal</th>
<th>Gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test statistic (critical value = 0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D_n$</td>
<td>0.69271</td>
<td><strong>0.088842</strong></td>
<td>0.21191</td>
<td>0.14371</td>
<td>0.23535</td>
</tr>
<tr>
<td></td>
<td>(2.2 × 10^{-16})</td>
<td>(0.2687)</td>
<td>(2.22 × 10^{-6})</td>
<td>(0.01054)</td>
<td>(1.55 × 10^{-6})</td>
</tr>
<tr>
<td>A-D</td>
<td>Inf</td>
<td><strong>3.0864</strong></td>
<td>2.4579</td>
<td>2.3471</td>
<td>3.6679</td>
</tr>
<tr>
<td></td>
<td>(0.000549)</td>
<td>(0.2472)</td>
<td>(0.4522)</td>
<td>(0.4979)</td>
<td>(0.1356)</td>
</tr>
</tbody>
</table>

Table 3.2: Performance of different distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Criteria</th>
<th>Weibull</th>
<th>Inv. Gamma</th>
<th>Pareto</th>
<th>Log-normal</th>
<th>Gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AIC</td>
<td>984.5139</td>
<td><strong>889.1509</strong></td>
<td>955.1093</td>
<td>917.1459</td>
<td>993.571</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td>990.2023</td>
<td><strong>894.8393</strong></td>
<td>960.7977</td>
<td>922.8343</td>
<td>999.2594</td>
</tr>
</tbody>
</table>

Note: optimum values are in boldface

From Table 3.1, we fail to reject the $H_0$ that our data is drawn from an Inverse-Gamma distribution for the K-S test. The A-D test, with its emphasis on the tails of the distributions, suggests that all distributions can be a good fit based on all p-values > 0.05. However, in terms of performance, the Inverse-Gamma distribution shows the best fit with the lowest AIC and BIC values (see Table 3.2). This can be easily observed from the empirical vs. theoretical CDFs of the selected distributions in Figure 3.2.

Figure 3.2: Empirical vs. fitted theoretical cumulative distribution functions
Based on the model selection choice, the likelihood for claim size can be written as

\[ X_i | \kappa_i, \theta_i \sim \text{Inv. Gamma} (\kappa_i, \theta_i), \quad \theta_i > 0, \kappa_i > 0, \]

**Aggregate Claim Process:** the random sum of individual claim sizes, \( X_i \) Conditioned on \( N_i \) as shown in Eq. 1, constitutes the total claims cost. For a fixed unit of time \( t \), peril type \( i \) and season-quarter \( s \), knowing that \( N_{i,s} = n_{i,s} \), the claim sizes \( X_{i,s,j}, j = 1,2,...,n_{i,s} \) are independent and identically distributed. It follows that the sum of these inverse gamma distributions forms another inverse gamma distribution as

\[ S_{i,s} | n_{i,s}, \kappa_i, \theta_i \sim \text{Inv. Gamma} (n_{i,s} \cdot \kappa_i, \theta_i), \quad \theta_i > 0, \kappa_i > 0 \]

### 3.3.1 Hierarchical Compound Collective Risk Model

A hierarchical Bayesian model shares information or “borrows statistical strength” from all levels of the data used. In other words, our model parameters are not estimated from only perils belonging to a similar type, but also from different perils, leveraging the hierarchical structure. This multilevel interaction and dependency throughout the hierarchy can ensure more reliable and robust estimates of parameters. We create a hierarchical model by placing shared priors over parameters and estimate them directly from data instead of assigning the parameters of our priors to some fixed constant. These high-level priors are known as **hyper-priors** and their parameters as **hyper-parameters**.

\[ S_{i,s} | n_{i,s}, \kappa_i, \theta_i \sim \text{Inv. Gamma} (n_{i,s} \cdot \kappa_i, \theta_i), \quad \theta_i > 0, \kappa_i > 0 \]

\[ N_{i,s} | \lambda_{i,s} \sim \text{Poisson} (\lambda_{i,s}), \quad \lambda_{i,s} > 0 \]

\[ \log (\lambda_{i,s}) | \alpha_i, \beta_2, X_s = \alpha_i + \beta_2 X_s \]

### 3.3.2 Hierarchical Dirichlet Prior Model Specification

We believe that some catastrophes may share similar characteristics in terms of their frequency of occurrence, thus, forming clusters or groups. In this setting, it is natural to consider a Dirichlet Process (DP) prior, which induces partitions in the parameter space. As introduced by
Ferguson (1973), a Dirichlet process is a prior measure on the space of probability measures. The Dirichlet process $DP(\gamma, G_0)$ has two parameters, a scale parameter $\gamma_0 > 0$ and a base probability measure, $G_0$.

Consider $B$ as a measurable subset under $G_0$ and $G \sim DP(\gamma, G_0)$. The DP is formally defined by the property that, for any finite partition, $\{B_1, \ldots, B_k\}$ of the base measure $G_0$, the joint distribution of the random vector $(G(B_1), \ldots, G(B_k))$ is the Dirichlet distribution with parameters $(\gamma G_0(B_1), \ldots, \gamma G_0(B_k))$. The expectation and variance of $B$ is given as

$$\mathbb{E}\{G(B)\} = G_0(B) \quad \text{and} \quad \text{Var}\{G(B)\} = \frac{G_0(B)(1 - G_0(B))}{\gamma + 1}$$

This implies that, $G_0$ is the mean, and $\gamma$ controls the concentration around the mean (Hong and Martin, 2017). An important property of DP is its large weak support (Müller, Quintana, Jara and Hanson, 2015, pp. 8). The weak support suggests that any distribution having the same support as $G_0$ can be approximated weakly by a DP probability measure. This property makes the DP prior one of the desirable processes in nonparametric Bayesian analysis. Another unique property of the DP is the discrete nature of $G$. With probability 1, we can always write $G \sim DP(\gamma, G_0)$ as a weighted sum of point-masses,

$$G = \sum_{h=1}^{\infty} w_h \delta_{m_h}(\cdot),$$

where $w_1, w_2, \ldots$ are random weights and $\delta_x(\cdot)$ is the point-mass distribution at $x$ for random locations, $m_1, m_2, \ldots$. The random weights, $w_1, w_2, \ldots$ can be explicitly represented in the stick-breaking construction outlined by Sethuraman (1994) as

$$w_1 = v_1 \quad w_h = v_h \prod_{k<h} (1 - v_k) \quad \text{where} \quad v_1, v_2, \ldots \sim \text{Beta}(1, \gamma).$$

The above expression satisfies $\sum_{j=1}^{\infty} w_h = 1$ with probability 1. Thus, the random weight, $w_h$ is a probability measure with positive integers.

We specify two DP priors with different base distributions to induce separate partitioning of the parameter space by aggregate claims and claim numbers\textsuperscript{6}. Specifically, to ensure that model parameter are grouped for multiple perils which share common aggregate claim amounts, but different claim count characteristics, we introduce

\textsuperscript{6} Two DP priors are needed due to the independent relationship between claim amount and claim number processes.
\[(\kappa_i, \theta_i) \sim DP(\gamma_1, G_0(\cdot))\]
\[\alpha_i \sim DP(\gamma_2, H_0(\cdot))\]
\[\beta_2 \sim Normal(\mu_0, \sigma_0^2)\]
\[G_0(\xi_1, \xi_2, \psi_1, \psi_2) = Gamma(\xi_1, \xi_2) \times Gamma(\psi_1, \psi_2)\]
\[H_0(\eta_1, \eta_2) = Gamma(\eta_1, \eta_2)\]

where \(\gamma_1, \gamma_2\) are the concentration parameters and \(G_0(\cdot), H_0(\cdot)\), the base distribution of the Dirichlet prior and \(\xi_1, \xi_2, \psi_1, \psi_2, \eta_1, \eta_2\) are hyper-parameters. The shape and scale parameters \((\kappa_i, \theta_i)\) are modeled as a bi-variate gamma distribution with \((\xi_1, \xi_2, \psi_1, \psi_2)\) as hyper-parameters. \(Gamma(\xi_1, \xi_2)\) controls the shape while \(Gamma(\psi_1, \psi_2)\) controls the scale. The approach provides much flexibility in the model to account for the fact that different perils may exhibit other characteristics. Further, we induce clusters in the claims number process by introducing \(Gamma(\eta_1, \eta_2)\) over \(\alpha_i\). The prior for the seasonal dependency parameter \(\beta_2\) is \(Normal(\mu_0, \sigma_0^2)\). The reason for not placing a DP over \(\beta_2\) is that, we believe the claim numbers are driven more by peril type than seasonal patterns. The summarized Hierarchical Dirichlet model can be written as

\[
\begin{align*}
S_{t, t} | n_{i, t}, \kappa_i, \theta_i & \sim Gamma(n_{i, s} \cdot \kappa_i, \theta_i), \quad \theta_i > 0, \quad \kappa_i > 0 \\
N_{i, t} | \lambda_{i, s} & \sim Poisson(\lambda_{i, s}), \quad \lambda_{i, s} > 0, \\
\log(\lambda_{i, s}) | \alpha_i, \beta_2, X_s & = \alpha_i + \beta_2 X_s \\
(\kappa_i, \theta_i) & \sim DP(\gamma_1, G_0(\cdot)) \\
\alpha_i & \sim DP(\gamma_2, H_0(\cdot)) \\
\beta_2 & \sim Normal(\mu_0, \sigma_0^2) \\
G_0(\xi_1, \xi_2, \psi_1, \psi_2) & = Gamma(\xi_1, \xi_2) \times Gamma(\psi_1, \psi_2) \\
H_0(\eta_1, \eta_2) & = Gamma(\eta_1, \eta_2)
\end{align*}
\]

with the following distributions for its hyperparameters

\[
\begin{align*}
\xi_1 & \sim Gamma(0.01, 0.01) \\
\xi_2 & \sim Gamma(0.01, 0.01) \\
\psi_1 & \sim Gamma(0.01, 0.01) \\
\psi_2 & \sim Gamma(0.01, 0.01) \\
\beta_2 & \sim Normal(0, 0.01)
\end{align*}
\]

The explicit solution for the likelihood function and posterior density for the above expressions is mathematically intractable in closed form due to the introduction of the DP prior.
Therefore, we utilize the Markov Chain Monte Carlo (MCMC) techniques to simulate random samples. We use OpenBUGS (Spiegelhalter et al., 2007) to implement the model.

3.4 Parameter Calibration of the Bayesian Catastrophe Model

3.4.1 Data

Practically, an insurance company keeps records on claims reported by policyholders, such as the number of claims, peril type, claim amount, and peril event dates. Therefore, the aggregate claim $S$, the total number of claims $N$ are observed, hence treated as the model's inputs. This empirical study is conducted using data provided by the Canadian company – Catastrophe Indices and Quantification Inc. (CatIQ), a subsidiary of PERILS AG in Switzerland. CatIQ is an insurance service company that collates catastrophe loss data from (re)insurers and builds comprehensive industry insured loss and exposure indices to serve the needs of the (re)insurance industries and other stakeholders. CatIQ only includes catastrophe losses to the industry of at least CA$25 million and affecting a significant number of insurers and policyholders. Our dataset on insured catastrophe consists of 130 catastrophes in Canada within the period 2008 – 2020. CatIQ is an authority on insured property losses for Canada. As such, most underwriters use the CatIQ loss index as a veritable source. Figure 3.3 shows the adjusted total annual CatIQ loss and number of qualified catastrophes between 2008 and 2020.

---

7 https://public.catiq.com
3.4.2 Bayesian Inference via MCMC

We run 40,000 steps of MCMC sampling, discard the first 10,000 steps of the chain as burn-in, and use the remaining 30,000 steps for our estimation. Table 3.3 below reports the posterior means of parameters.

Table 3.3: Posterior means of the Hierarchical Dirichlet Prior model

<table>
<thead>
<tr>
<th>Peril</th>
<th>Posterior means</th>
<th>( S_i )</th>
<th>( \theta_i )</th>
<th>( \lambda_{i1} )</th>
<th>( \lambda_{i2} )</th>
<th>( \lambda_{i3} )</th>
<th>( \lambda_{i4} )</th>
<th>( \alpha_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Windstorm)</td>
<td></td>
<td>( \kappa_i )</td>
<td>( \theta_i )</td>
<td>( \lambda_{i1} )</td>
<td>( \lambda_{i2} )</td>
<td>( \lambda_{i3} )</td>
<td>( \lambda_{i4} )</td>
<td>( \alpha_i )</td>
</tr>
<tr>
<td>2 (Severe Storm)</td>
<td></td>
<td>( \kappa_i )</td>
<td>( \theta_i )</td>
<td>( \lambda_{i1} )</td>
<td>( \lambda_{i2} )</td>
<td>( \lambda_{i3} )</td>
<td>( \lambda_{i4} )</td>
<td>( \alpha_i )</td>
</tr>
<tr>
<td>3 (Hailstorm)</td>
<td></td>
<td>( \kappa_i )</td>
<td>( \theta_i )</td>
<td>( \lambda_{i1} )</td>
<td>( \lambda_{i2} )</td>
<td>( \lambda_{i3} )</td>
<td>( \lambda_{i4} )</td>
<td>( \alpha_i )</td>
</tr>
<tr>
<td>4 (Winter Storm)</td>
<td></td>
<td>( \kappa_i )</td>
<td>( \theta_i )</td>
<td>( \lambda_{i1} )</td>
<td>( \lambda_{i2} )</td>
<td>( \lambda_{i3} )</td>
<td>( \lambda_{i4} )</td>
<td>( \alpha_i )</td>
</tr>
<tr>
<td>5 (Flood)</td>
<td></td>
<td>( \kappa_i )</td>
<td>( \theta_i )</td>
<td>( \lambda_{i1} )</td>
<td>( \lambda_{i2} )</td>
<td>( \lambda_{i3} )</td>
<td>( \lambda_{i4} )</td>
<td>( \alpha_i )</td>
</tr>
<tr>
<td>6 (Tornado)</td>
<td></td>
<td>( \kappa_i )</td>
<td>( \theta_i )</td>
<td>( \lambda_{i1} )</td>
<td>( \lambda_{i2} )</td>
<td>( \lambda_{i3} )</td>
<td>( \lambda_{i4} )</td>
<td>( \alpha_i )</td>
</tr>
<tr>
<td>7 (Hurricane)</td>
<td></td>
<td>( \kappa_i )</td>
<td>( \theta_i )</td>
<td>( \lambda_{i1} )</td>
<td>( \lambda_{i2} )</td>
<td>( \lambda_{i3} )</td>
<td>( \lambda_{i4} )</td>
<td>( \alpha_i )</td>
</tr>
<tr>
<td>8 (Tropical Storm)</td>
<td></td>
<td>( \kappa_i )</td>
<td>( \theta_i )</td>
<td>( \lambda_{i1} )</td>
<td>( \lambda_{i2} )</td>
<td>( \lambda_{i3} )</td>
<td>( \lambda_{i4} )</td>
<td>( \alpha_i )</td>
</tr>
<tr>
<td>9 (Fire)</td>
<td></td>
<td>( \kappa_i )</td>
<td>( \theta_i )</td>
<td>( \lambda_{i1} )</td>
<td>( \lambda_{i2} )</td>
<td>( \lambda_{i3} )</td>
<td>( \lambda_{i4} )</td>
<td>( \alpha_i )</td>
</tr>
</tbody>
</table>

Figure 3.3: CatIQ annual catastrophe cost incurred (left) and the number of catastrophes (right) in Canada from 2008 – 2020.
The induced clusters in the aggregate claims parameters are pretty evident. For instance, perils 1, 2, 3, and 4 have somehow similar parameter estimates. Perils 5, 6, 7 also show some similarities in terms of their parameter estimates, while perils 8 and 9 exhibit very different parameter estimates. Peril types show a strong differentiation in the number of claims. We note that the number of claims is greatly influenced $\alpha_i$, the random intercept in nonhomogenous Poisson regression. The seasonality component $\beta_2 = 0.0615$ is not zero; hence it contributes significantly to the claim number process. Next, we report the group indicator and group frequency for each peril Table 3.4. The Bayesian outputs are obtained based on 30,000 MCMC values after a burn-in of 10,000. Therefore, the total frequency is 30,000.

Table 3.4: Distribution of 30,000 MCMC samples under various aggregate claim ($S_i$) groups induced by the Hierarchical Dirichlet Prior model

<table>
<thead>
<tr>
<th>Peril (Number)</th>
<th>Group indicators and corresponding frequencies.</th>
<th>Highest Frequency Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Windstorm)</td>
<td><strong>Group Indicator</strong> 2 6 <strong>Frequency</strong> 29966 34</td>
<td>2</td>
</tr>
<tr>
<td>2 (Severe Storm)</td>
<td><strong>Group Indicator</strong> 2 6 <strong>Frequency</strong> 28208 1792</td>
<td>2</td>
</tr>
<tr>
<td>3 (Hailstorm)</td>
<td><strong>Group Indicator</strong> 2 <strong>Frequency</strong> 30000</td>
<td>2 (Perfect/Pure Group)</td>
</tr>
<tr>
<td>4 (Winter Storm)</td>
<td><strong>Group Indicator</strong> 2 <strong>Frequency</strong> 30000</td>
<td>2 (Perfect/Pure Group)</td>
</tr>
<tr>
<td>5 (Flood)</td>
<td><strong>Group Indicator</strong> 2 6 <strong>Frequency</strong> 28653 1347</td>
<td>2</td>
</tr>
<tr>
<td>6 (Tornado)</td>
<td><strong>Group Indicator</strong> 2 3 6 <strong>Frequency</strong> 29305 73 622</td>
<td>2</td>
</tr>
<tr>
<td>7 (Hurricane)</td>
<td><strong>Group Indicator</strong> 2 6 <strong>Frequency</strong> 29283 717</td>
<td>2</td>
</tr>
<tr>
<td>8 (Tropical Storm)</td>
<td><strong>Group Indicator</strong> 2 3 6 <strong>Frequency</strong> 6814 21817 1369</td>
<td>3</td>
</tr>
<tr>
<td>9 (Fire)</td>
<td><strong>Group Indicator</strong> 2 3 5 6 9 <strong>Frequency</strong> 1772 25305 4 2699 220</td>
<td>3 (Most Impure Group)</td>
</tr>
</tbody>
</table>

Perils 3 (Hailstorm) and 4 (Winter Storm) show pure grouping by assigning the same group indicator to all MCMC samples. For example, in the case of Hailstorm, we observe 100% occupancy for Group 2. This is also true for Winter Storm. Therefore, according to the HDP model, Hailstorm and Winter Storm can be grouped. Windstorm and Severe Storm, Flood, Tornado, and Hurricane also report more than 98% occupancy for Group 2, leading them to be grouped in Group
2. Tropical Storm and Fire seem to show higher occupancy in Group 3 (about 72.7% for peril 8 and 84.3% for peril 9). However, we get these two perils separate due to the marked difference observed in their claim numbers (see Table 3). Our data shows that fire occurs with a relatively low frequency but has the highest claim cost per occurrence. The HDP model captured this behavior. Based on the above analysis, we group the aggregate claim parameters in four clusters, as reported in Table 3.5. The grouping on the number of claims, $N_l$ is presented in Table 1A in Appendix A.

Table 3.5: Distinct groups of aggregate claim parameters

<table>
<thead>
<tr>
<th>Claims Group</th>
<th>Perils</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Windstorm, Severe storm, Flood, Tornados, Hurricane</td>
</tr>
<tr>
<td>Group 2</td>
<td>Hailstorm, Winter storm,</td>
</tr>
<tr>
<td>Group 3</td>
<td>Tropical storm</td>
</tr>
<tr>
<td>Group 4</td>
<td>Fire</td>
</tr>
</tbody>
</table>

3.4.3 Convergence Diagnostics

Markov chains need to converge before we can get meaningful results from them. Due to the stochasticity of the HDP model, we cannot directly assess the convergence of the aggregate claims per se. Instead, we test for convergence on the hyperparameters, $(\xi_1, \xi_2, \psi_1, \psi_2)$ of the base distribution $G_0$. The assumption that, when parameters under the base measure converges, then the target distribution converges can be supported by the weak approximation property of the DP probability measure.

We perform two checks based on graphical techniques. We ran three parallel chains Markov chains with 50,000 steps each, including a 10,000 burn-in period. Each of the chains simulated in parallel has different starting points, permitting verification on whether the sequences mix, indicating convergence. As reported in Figure 3.4, the three chains mix well, indicating convergence. The kernel density plot shows no multimodal distributions, further corroborating convergence. A more accurate way to verify convergence is to perform the Brooks-Gelman-Rubin (BGR) test (Gelman and Rubin, 1992). The BGR test compares the variance with and between the chains. We report the BGR test in Figure 3.5 below.
3.4.4 Bayesian Predictive Inference via MCMC

The goal of the catastrophic risk model developed is to predict the probability that a catastrophe would occur with aggregate claim cost being more than a stipulated threshold. To put it more formally, consider the number of the claims process, \( \{N(t) : t \in [0, T]\} \) using the nonhomogeneous Poisson process (NHPP) with parameters \( \lambda(t) > 0 \). Then the probability of aggregate claims, \( S(t) \) being less than or equal to a threshold, \( D \) at time \( t \) is the marginal distribution of aggregate claim obtained by convolution as
\[ S(t, D) = \sum_{n=0}^{\infty} e^{-\lambda(T-t)} \left( \frac{\lambda(T-t)}{n!} \right)^n p^{n*}(D) \]  

(3.3).

We should note that Eq. 3.3 is the explicit form of Eq. 3.2. In classical approaches, to approximate the density function of \( S(t) \) involves convolution methods that often require complicated and lengthy numerical integration. An alternative approach under the Bayesian method is to use MCMC simulation.

We learn about the distribution of \( S(t) \), by using all the posterior samples of its parameters, \( \phi = (\lambda, \kappa, \theta) \). Consider the posterior predictive distribution of some variable \( z \) (which is dependent on \( \phi \)) given observed data \( Y \) defined by

\[
\pi(z \mid Y) = \int f(z \mid \phi)\pi(\phi \mid Y) d\phi
\]

where \( \pi(\phi \mid Y) \) is the posterior distribution of \( \phi \) given the data \( Y \), and \( f(z \mid \phi) \) is the density of \( z \) given \( \phi \). By integrating over all possible values of \( \phi \), we account for variability in parameter estimates, often ignored in classical approaches. Consequently, we obtain better estimates than if we had used point estimates such as the posterior mean. Since we cannot integrate directly over \( \phi \), we instead use the MCMC method to obtain posterior predictive distribution estimates. Let \((t, T), t \geq 0\) be a time interval of interest for which we want to predict aggregate claims. Given \( N \) number of past claims, we can predict the future number of claims, \( N_f \) as

\[
\pi(N_f = n \mid N) = \int_0^{\infty} p(N_f \mid \theta, N)\pi(\theta \mid N) d\theta
\]

\[
= E_{\theta \mid N}\{ p(N_f = n \mid \theta) \}
\]

\[
= E_{\theta \mid N}\left\{ \frac{e^{-\theta \theta^n}}{n!} \right\} = \sum_{n=0}^{\infty} \frac{e^{-\lambda(T-t)} \left( \lambda(T-t) \right)^n}{n!} \text{ as } N_f \sim Poisson\left( \lambda_s^{(i)} \right)
\]

\[
\approx \frac{1}{m} \sum_{i=1}^{m} e^{-\lambda(T-t)} \left( \lambda_i(T - t) \right)^n
\]

where \( m \) is the number of MCMC simulated samples and \( \lambda(i) \) is the \( i^{th} \) simulated value of \((\theta \mid N)\). Similarly, given \( S \) past aggregate claims, we predict future claims, \( S_f \) as follows
\[
\pi(S \leq S_f | n_f, u) = \int_u p(S \leq S_f | U = u, N_f = n_f) du \\
= E_{u|S}[p(S \leq S_f | \kappa, \theta, n_f)] \approx \frac{1}{m} \sum_{i=1}^{m} \left\{ 1 - \left( \frac{\theta^{(i)}}{S_f} \right)^{n_f \kappa^{(i)}} \right\}^{-1}
\]

where \( u \sim \text{inv. gamma}(\kappa, \theta) \), \( \kappa^{(i)} \) and \( \theta^{(i)} \) are the \( i \)th simulated values \( \kappa \) and \( \theta \) respectively. We simulate \( F_S(t, D) \) or \( Pr[S \leq D] \) at time interval \((t, T), t \geq 0\) in the following steps:

- Obtain estimates \( \lambda^{(i)}, \kappa^{(i)}, \theta^{(i)} \) for the parameters \( \lambda, \kappa, \theta \)
- Simulate \( N_f \sim \text{Poisson}(\lambda^{(i)}) \)
- Simulate \( N_f \) values of \( S \sim \text{Inv. Gamma}(N_f \cdot \kappa^{(i)}, \theta^{(i)}) \)
- Evaluate the probability that \( S^{(i)} \leq D \).

### 4.0 Stochastic Interest Model

#### 4.1 CIR Model

The instantaneous interest (short rate) dynamics proposed by Cox, Ingersoll, and Ross (Cox et al., 1985) assume a mean-reverting process. In this model, the diffusion term has a ‘square root.’ The model is a benchmark for obtaining analytical bond prices. Our choice for using the CIR model over others such as the Vasicek model is that our interest rate data have positive interest rates. The latter admits negative interest rates, which is not suitable in this paper. The short rate dynamics \( \{r(t): t \in [0, T]\} \) is defined by the following stochastic differential equation

\[
dr(t) = (\alpha - \beta r(t))dt + \sigma \sqrt{r(t)}.dW(t) \tag{4.1}
\]

with the condition \( \alpha > \sigma^2/2 \) to ensure that \( r(t) \) remains positive and the origin is never accessed by the process, where \( \{W(t), t \geq 0\} \) is a standard Brownian motion and \( \alpha, \beta, \sigma \geq 0 \) are parameters to be estimated. The estimation of the continuous-time model in Eq. (4.1) from discretely observed data can be problematic since the likelihood is difficult to obtain (see, e.g. (Keener et al., 2013). It is a well known fact that the conditional density of the CIR process follows
a non-central chi-squared distribution. However, it may not always be possible to get this distribution in practice. In this paper, we discretize the continuous-time model using the Euler-Maruyama approximation scheme as

\[ r_{t+\Delta t} = r_t + (\alpha - \beta r_t)\Delta + \sigma \sqrt{\Delta t} \epsilon_t, \quad \Delta t = \frac{T}{N}, \quad \Delta t > 0 \]

where \( \Delta t \) is the time interval and \( \epsilon_t \sim \text{Normal}(0,1) \).

4.1.1 Data Augmentation

Some difficulty arises in the use of discretization when the magnitude of the observation interval \( \Delta t = t_{i+1} - t_i \) is large. We overcome this difficulty in the Bayesian approach by introducing augmented data between each pair of observations. Suppose we have \( T \) observations, and \( M \) augmented data between each pair of consecutive observations available:

\[ t_i = \tau_{0,i} < \tau_{1,i} < \cdots < \tau_{M,i} = t_{i+1} \]

The stepsize of the augmented data \( \delta_t = \tau_{j+1,i} - \tau_{j,i} \) is kept small enough to ensure the accuracy and stability of the Euler-Maruyama scheme. Let \( Y = (r_1, r_2, ..., r_T) \) denote all observations and \( Y^* = (r_1^*, r_2^*, ..., r_{T-1}^*) \) be augmented data where \( r_t^* = (r_{t,0}^*, r_{t,1}^*, ..., r_{t,M}^*) \) and \( r_{t,0}^* = r_t \). For each time \( t \geq 0 \), we define \( \Delta^* = \frac{\Delta}{M+1} \) and assume that \( r_{t,j}^* \) is a Markov process for \( j = 0, 1, ..., M \).

We assume that the transition density of the CIR model follows a Markov process:

\[
P(r_t^* | Y, \alpha, \beta, \sigma) = P(r_{t,0}^*, r_{t,1}^*, ..., r_{t,M}^* | Y, \alpha, \beta, \sigma) = \prod_{j=1}^M P(r_{t,j}^* | r_{t,j-1}^*, \alpha, \beta, \sigma),
\]

where \( r_{t,0}^* = r_t \). In this case the full conditional transition density of the process is

\[
r_{t,j+1}^* | r_{t,j}^*, \alpha, \beta, \sigma \sim N(r_{t,j}^* + (\alpha - \beta r_{t,j}^*) \Delta, \sigma^2 \Delta r_{t,j}^*)
\]
4.2 Bayesian CIR Model

4.2.1 Bayesian CIR Model Specification

To achieve parameter estimates, we need to find the likelihood and full conditional posterior distribution of our in closed form. We derive the fully conditional posterior distribution by dividing parameters \((\alpha, \beta, \sigma)\) into \((\psi, \sigma^2)\) where \(\psi = (\alpha, \beta)^T\). This is possible since \((\alpha, \beta)\) is linear in the drift component so that we can estimate them jointly as \(\psi = (\alpha, \beta)^T\) (see, e.g., Feng and Xie (2012)). We specify the likelihood and priors of the model as follows:

\[
Y^*, Y | \psi, \sigma^2 \sim MVN(\mu, \Lambda^{-1})
\]

\[
\psi \sim Truncated \ MVN(\mu_0, \Sigma^{-1})
\]

\[
\sigma^2 \sim IG(\nu_0, \beta_0)
\]

where the likelihood follows a multi-variate normal (MVN) distribution due to the two parameters \(\psi, \sigma^2\). The prior of \(\psi\) is drawn from a truncated MVN to ensure \(\alpha, \beta\) remain positive. Finally, the prior of \(\sigma^2\) is drawn from an inverse gamma (IG) distribution. Due to the independent structure between \((\psi, \sigma^2)\), the likelihood function can be written as

\[
P(Y, Y^* | \psi, \sigma^2) = \prod_{t=1}^{T-1} \prod_{j=0}^{M} P(r_{t,j+1}^* | r_{t,j}^*, \psi, \sigma^2)
\]

\[
\propto \exp \left\{ \sum_{t=1}^{T-1} \sum_{j=0}^{M} \left[ \frac{r_{t,j+1}^* - \left[ r_{t,j}^* + (\alpha - \beta r_{t,j}^*)\Delta \right]}{2\sigma^2 \Delta r_{t,j}^*} \right]^2 \right\}
\]

\[
\propto \exp \left\{ -\sum_{t=1}^{T-1} \sum_{j=0}^{M} \left[ (\alpha - \beta r_{t,j}^*)\Delta + r_{t,j}^* \right]^2 - 2r_{t,j+1}^* (\alpha - \beta r_{t,j}^*)\Delta \right\}
\]

After expanding and simplifying, we get

\[
\propto \exp \left\{ -\frac{\Delta A \alpha^2 + \Delta B \beta^2 - 2\Delta(T - 1)(M + 1)\alpha \beta - 2C\alpha - 2D\beta}{2\sigma^2} \right\} \tag{4.2}
\]

where

\[
A = \sum_{t=1}^{T-1} \sum_{j=0}^{M} \frac{1}{r_{t,j}^*}
\]

\[
B = \sum_{t=1}^{T-1} \sum_{j=0}^{M} r_{t,j}^*
\]
\[ C = \sum_{t=1}^{T-1} \sum_{j=0}^{M} \frac{r_{t,j} - r_{t,j+1}^*}{r_{t,j}^*} \]

\[ D = \sum_{t=0}^{T-1} \sum_{j=0}^{M} (r_{t,j}^* - r_{t,j+1}^*) \]

Since the likelihood, \( P(Y, Y^* | \Psi, \sigma^2) \) is a bivariate normal distribution of the in terms of \( \Psi, \sigma^2 \), it can be written as:

\[ P(Y, Y^* | \Psi, \sigma^2) = \text{MVN}(\mu, \Lambda^{-1}) \propto |\Lambda_{\Psi}|^{\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\Psi - \mu_{\Psi})^T \Lambda_{\Psi}^{-1} (\Psi - \mu_{\Psi}) \right\} \]  \hspace{1cm} (4.3)

where \( \mu_{\Psi} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \) and \( \Lambda_{\Psi} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \). Then it follows that

\[
(\Psi - \mu_{\Psi})^T \Lambda_{\Psi} (\Psi - \mu_{\Psi}) = (\alpha - \mu_1, \beta - \mu_2) \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \alpha - \mu_1 \\ \beta - \mu_2 \end{pmatrix} = \\
\left( a_{11} \alpha^2 + a_{22} \beta^2 + 2a_{12} \alpha \beta - 2(a_{11} \mu_1 + a_{12} \mu_2) \right) \alpha - 2(a_{22} \mu_2 + a_{12} \mu_1) \beta \\
+ (2a_{12} \mu_1 \mu_2 + a_{11} \mu_1^2 + a_{22} \mu_2^2) 
\]  \hspace{1cm} (4.4)

Comparing formula (4.3) and (4.4), we get

\[ a_{11} = \frac{\Delta}{\sigma^2} A; \quad a_{22} = \frac{\Delta}{\sigma^2} B; \quad a_{12} = \frac{\Delta}{\sigma^2} (T - 1)(M + 1), \]

\[ a_{11} \mu_1 + a_{12} \mu_2 = C; \quad a_{22} \mu_2 + a_{12} \mu_1 = D \]

therefore,

\[
\mu_{\Psi} = \left( \frac{a_{22} C - a_{12} D - a_{12} C + a_{11} D}{a_{11} a_{22} - a_{12}^2}, \frac{-a_{12} C + a_{11} D}{a_{11} a_{22} - a_{12}^2} \right)^T 
\]

and

\[
\Lambda_{\Psi} = \begin{pmatrix}
\frac{\Delta}{\sigma^2} A & -\frac{\Delta}{\sigma^2} (T - 1)(M + 1) \\
-\frac{\Delta}{\sigma^2} (T - 1)(M + 1) & \frac{\Delta}{\sigma^2} B
\end{pmatrix}
\]

The full conditional posterior distribution for \( \Psi \) is

\[ P(\Psi | Y, Y^*, \sigma^2) = P(Y, Y^* | \Psi, \sigma^2) \pi(\Psi) \]  \hspace{1cm} (4.5)

with \( \pi(\Psi) \sim \text{Truncated MVN}_{(0, \infty)}(\mu_0, \Sigma_0^{-1}) \)
\[ \pi(\Psi) = |\Sigma_0|^{\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\Psi_0 - \mu_0)^T \Sigma_0^{-1} (\Psi_0 - \mu_0) \right\} \propto \exp\left\{ -\frac{1}{2} (\Psi_0 - \mu_0)^T \Sigma_0^{-1} (\Psi_0 - \mu_0) \right\} \]

\[ = -\frac{1}{2} \exp\left\{ (\Psi_0 - \mu_0)^T (\Psi_0 - \mu_0) \Sigma_0^{-1} \right\} \]

\[ \propto -\frac{1}{2} \exp\left\{ \Sigma_0^{-1} \Psi_0 \Psi_0 - \Sigma_0^{-1} \Psi_0 \mu_0 - \Sigma_0^{-1} \mu_0 \Psi_0 + \Sigma_0^{-1} \mu_0 \mu_0 \right\} \]

\[ \propto -\frac{1}{2} \exp\left\{ \Psi_0^T \Sigma_0^{-1} \Psi_0 - 2 \Psi_0^T \Sigma_0^{-1} \mu_0 + \mu_0^T \Sigma_0^{-1} \mu_0 \right\} \]

\[ = -\frac{1}{2} \left\{ \Psi_0^T \Sigma_0^{-1} \Psi_0 - 2 \Psi_0^T \Sigma_0^{-1} \mu_0 \right\} \]

\[ \propto -\frac{1}{2} \exp\left\{ \Psi_0^T A_0 \Psi_0 - 2 \Psi_0^T b_0 \right\} \quad (4.6) \]

where \( A_0 = \Sigma_0^{-1} \) and \( b_0 = \Sigma_0^{-1} \mu_0 \).

The likelihood can multi-variate likelihood can be rewritten as

\[ P(Y, Y^* | \Psi, \sigma^2) \propto |\Lambda|^{\frac{1}{2}} \exp\left\{ -\frac{1}{2} (r_i - \mu)^T \Lambda^{-1} (r_i - \mu) \right\} \]

\[ = -\frac{1}{2} \exp\left\{ \sum_i r_i^T \Lambda^{-1} r_i - 2 \sum_i \mu^T \Lambda^{-1} r_i + \sum_i \mu^T \Lambda^{-1} \mu \right\} \]

\[ \propto -\frac{1}{2} \exp\left\{ -2 \mu^T \Lambda^{-1} n \bar{r} + n \mu^T \Lambda^{-1} \mu \right\} \]

\[ = -\frac{1}{2} \exp\left\{ -2 \mu^T b_1 + \mu^T A_1 \mu \right\} \quad (4.7) \]

where \( b_1 = n \bar{r} \Lambda^{-1} \) and \( A_1 = n \Lambda^{-1} \).

Therefore the full conditional posterior for \( \Psi \) is just the product of Eq. (4.6) and Eq. (4.7) as

\[ P(Y, Y^* | \Psi, \sigma^2) \pi(\Psi) \propto -\frac{1}{2} \exp\left\{ -2 \mu^T b_1 + \mu^T A_1 \mu \right\} \times -\frac{1}{2} \exp\left\{ \Psi_0^T A_0 \Psi_0 - 2 \Psi_0^T b_0 \right\} \]

\[ \propto \exp\left\{ \mu^T b_1 - \frac{1}{2} \mu^T A_1 \mu - \frac{1}{2} \mu^T A_0 \mu + \frac{1}{2} \mu^T A_0 A_1 \right\} \]

with \( A_n = A_0 + A_1 = \Sigma_0^{-1} + n \Lambda^{-1} \); \( b_n = b_0 + b_1 = \Sigma_0^{-1} \mu_0 + n \bar{r} \Lambda^{-1} \), the full conditional posterior can be simplified as

\[ P(\Psi | Y, Y^*, \sigma^2) \propto \exp\left\{ \mu^T b_n - \frac{1}{2} \mu^T A_n \right\} \]

\[ \propto MVN(\Psi | \Sigma_0^{-1} \mu_0 + n \bar{r} \Lambda^{-1}, \Sigma_0^{-1} + n \Lambda^{-1}) \].
The full conditional posterior distribution for $\sigma^2$ on the other hand can be expressed as

\[
P(\sigma^2|Y, Y^*, \Psi) \propto P(Y, Y^*|\Psi, \sigma^2) \pi(\sigma^2) \quad \text{with} \quad \pi(\sigma^2) \sim \text{Inverse Gamma}(v_0, \beta_0).
\]

Therefore,

\[
P(\sigma^2|Y, Y^*, \Psi) \propto \exp\left\{ \frac{1}{2} \sum_{t=1}^{T-1} \sum_{j=0}^{M} \left[ \frac{r_{t,j+1}^* - \left[ r_{t,j}^* + (\alpha - \beta r_{t,j}^*)\Delta \right]}{2\sigma^2 r_{t,j}^*} \right]^2 \right\} (\sigma^2)^{-(v_0+1)} \exp \left\{ -\frac{\beta_0}{\sigma^2} \right\}
\]

\[
\propto (\sigma^2)^{-\left(v_0 + \frac{(T-1)(M+1)}{2} + 1\right)} \exp \left\{ -\frac{1}{2} \sum_{t=1}^{T-1} \sum_{j=0}^{M} \left[ \frac{r_{t,j+1}^* - \left[ r_{t,j}^* + (\alpha - \beta r_{t,j}^*)\Delta \right]}{r_{t,j}^*} \right]^2 + \frac{\beta_0}{2} \right\}
\]

\[
\propto \text{Inverse Gamma} \left( \sigma^2 | v_0 + \frac{(T-1)(M+1)}{2}, \beta_0 + \sum_{t=1}^{T-1} \sum_{j=0}^{M} \left[ \frac{r_{t,j+1}^* - \left[ r_{t,j}^* + (\alpha - \beta r_{t,j}^*)\Delta \right]}{2r_{t,j}^*} \right]^2 \right)
\]

### 4.2.2 Sampling Algorithm

We obtain samples of the model parameters using a Gibbs sampler. The Gibbs sampler is an MCMC algorithm that generates realizations from an assumed distribution for each parameter in turns, conditional on the current values of the other parameters and observed data points. The sequence of samples generated constitutes a Markov chain whose stationary distribution is the joint posterior distribution we desire. It is relatively simpler to sample from the fully conditional posterior distribution (if available in closed form) than to marginalize over a joint distribution by integration. Our goal is to simulate samples for $\alpha$, $\beta$ and $\sigma^2$ via a Gibbs sampler from the fully conditional posterior distribution derived above, using the following steps:

**Step 1:** Initialize $\Psi, \sigma^2, r_{1,0}^*$

**Step 2:** Use data augmentation to generate samplings of $Y^* = r_1^*, r_2^*, \ldots, r_{T-1}^*$

**Step 3:** Use Gibbs sampler to

a. update $\Psi$ from $P(\Psi|Y, Y^*, \sigma^2)$ where $Y^* = r_1^*, r_2^*, \ldots, r_{T-1}^*$ are from the previous iteration

b. update $\sigma^2$ from $P(\sigma^2|Y, Y^*, \Psi)$ where $Y^* = r_1^*, r_2^*, \ldots, r_{T-1}^*$ are from the prior iteration, and $\Psi$ is from (a)

**Step 4:** Update $r_1^*, r_2^*, \ldots, r_{T-1}^*$ from $P(r_i^*|\Psi, \sigma^2)$

**Step 5:** Repeat Step 3 and Step 4 until the sampling size $N$ is reached.
4.2.3 Forecasting

Given the MCMC samples of each parameter \{\alpha^{(i)}, \beta^{(i)} and \sigma^{(i)}, i = 1, ..., N\} and an initial interest, \(r_0\), we can forecast future interest rate \{\(r_1^{(i)}, r_2^{(i)}, ..., r_T^{(i)}, i = 1, ..., N\}\) recursively as

\[
r_1^{(i)} = r_0 + (\alpha^{(i)} - \beta^{(i)}r_0)\Delta + \sigma^{(i)}\sqrt{\Delta}r_{0}\varepsilon_1^{(i)}
\]

\[
r_2^{(i)} = r_1^{(i)} + (\alpha^{(i)} - \beta^{(i)}r_1^{(i)})\Delta + \sigma^{(i)}\sqrt{\Delta}r_{1}\varepsilon_2^{(i)}
\]

\[
\vdots
\]

\[
r_T^{(i)} = r_{T-1}^{(i)} + (\alpha^{(i)} - \beta^{(i)}r_{T-1}^{(i)})\Delta + \sigma^{(i)}\sqrt{\Delta}r_{T-1}\varepsilon_T^{(i)}
\]

where \(\varepsilon_t^{(i)} \sim N(0, 1), i = 1, ..., N, \ t = 1, ..., T\). In this same step, we also estimate \(B_{CIR}(0, t)\), the time-0 yield of a risk-free zero-coupon bond which pays 1 at time \(t\) (in years). The bond yield is an exponential affine form with the corresponding formula

\[
P(t, T) = A(t, T)\exp(-B(t, T)r_t)
\]

where

\[
A(t, T) = \left(\frac{2h \exp((a + h)(T - t)/2)}{2h + (a + h)(\exp((T - t)h) - 1)}\right)^{2ab/\sigma^2}
\]; \(a = \alpha, \ b = \frac{\alpha}{\beta}\)

\[
B(t, T) = \frac{2(\exp((T - t)h) - 1)}{2h + (a + h)(\exp((T - t)h) - 1)}
\]

\[
h = \sqrt{a^2 + 2\sigma^2}
\]

4.3 Parameter Calibration of the Bayesian CIR Model

We calibrate the parameters of the CIR model using 12 years of weekly data on 3-months Canadian Treasury bills from January 1, 2008, to December 12, 2020. Figure 4.1 shows the historical yields on the Canadian 3-month Treasury bills.

---

8 Due to the independence between interest and catastrophe risk it not necessary to use a dataset with the same time period.
Figure 4.1: Time series of Canadian 3-month Treasury bills over from 2008 to 2020

We run 15,000 steps of MCMC sampling with $M = 20$ and $\Delta = 1/252$. We discard the first 5,000 steps of the chain and use the remaining 10,000 steps for our estimation. Table 4.1 below reports the specifications of hyperparameters used in our model.

Table 4.1: Hyperparameters of the Bayesian CIR model

<table>
<thead>
<tr>
<th>Hyperparameter</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_0$</td>
<td>2.1</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>3.0</td>
</tr>
<tr>
<td>$\mu^1_0$</td>
<td>0.0</td>
</tr>
<tr>
<td>$\mu^2_0$</td>
<td>0.0</td>
</tr>
<tr>
<td>$\Sigma_0^{-1}$</td>
<td>10.0</td>
</tr>
</tbody>
</table>

The posterior mean, posterior standard deviation, 95% credible intervals of the highest posterior density (HPD), and Geweke convergence diagnostics are reported in Table 4.2.

Table 4.2: Summary statistics of MCMC samples

<table>
<thead>
<tr>
<th>Params</th>
<th>Posterior Mean</th>
<th>SD</th>
<th>95% HPD</th>
<th>Geweke</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>3.0299</td>
<td>$2.07 \times 10^{-3}$</td>
<td>(3.0258, 3.0339)</td>
<td>-0.3852</td>
</tr>
<tr>
<td>$\beta$</td>
<td>3.2694</td>
<td>$3.56 \times 10^{-3}$</td>
<td>(3.2625, 3.2765)</td>
<td>-1.3129</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.00171</td>
<td>$1.90 \times 10^{-5}$</td>
<td>(0.001673, 0.001747)</td>
<td>0.8433</td>
</tr>
</tbody>
</table>

The Geweke diagnostic seems to support the null hypothesis of convergence since all the z-score values are well within two standard deviations of zero. Figure 4.2 below displays the trace plots.
of the MCMC sampling sequence for $\alpha, \beta$ and $\sigma^2$ which may lend additional support to the convergence of each sequence. The long-term interest rate in the CIR model is modeled as $\alpha / \beta$. Our estimated value for the long-term interest rate is \( \frac{3.0299}{3.2694} = 0.9267\% \) with a deviation of only about 0.0074 from the historical mean of the observed data, 0.9342%. The posterior distribution of $\alpha, \beta$ and $\sigma^2$ shown in Figure 4.3, all look reasonably normal.

![Trace plots of parameters $\alpha, \beta$ and $\sigma^2$](image1)

Figure 4.2: Trace plots of parameters $\alpha, \beta$ and $\sigma^2$ from the fully conditional posterior distribution with $M = 20, N = 15,000$

![Posterior distribution of Bayesian CIR parameters](image2)

Figure 4.3: Posterior distribution of Bayesian CIR parameters

34
5.0 Contingent Claim Pricing Model

CAT bonds are typically priced under certain assumptions. First, there exists an arbitrage-free market with an equivalent martingale measure. Second, financial markets behave independently of catastrophe occurrence. Third, interest rate changes can be replicated using existing financial instruments. Let $0 < T < \infty$ represents the maturity date for the continuous-time trading interval $[0, T]$. Then the market uncertainty can be defined by a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0,T]}, \mathbb{P})$, where $\mathcal{F}_t$ is a family of $\sigma$-algebras containing all available information about the underlying processes. We consider the process that could trigger payment of CAT bond to be a nonnegative non-decreasing, right-continuous stochastic process $Y = \{Y_t, t \geq 0\}$ defined on a filtered physical probability space $(\Omega^{(1)}, \mathcal{F}^{(1)}, \{\mathcal{F}_t^{(1)}\}_{t \in [0,T]}, \mathbb{P}^{(1)})$, where $\mathcal{F}_t^{(1)}$ represent all available catastrophic risk information (e.g., claim amount, number of claims, peril type). Consider an arbitrage-free financial market, defined by another filtered physical probability space $(\Omega^{(2)}, \mathcal{F}^{(2)}, \{\mathcal{F}_t^{(2)}\}_{t \in [0,T]}, \mathbb{P}^{(2)})$, where $\mathcal{F}_t^{(2)}$ represents available investment information (e.g., interest rates, inflation or securities prices). Following the argument of Cox and Pedersen (2000) about the independence of the two stochastic processes; we can define a product probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0,T]}, \mathbb{P})$, with $\mathcal{F}_t = \mathcal{F}_t^{(1)} \times \mathcal{F}_t^{(2)} \subset \mathcal{F}$ for $t \geq 0$, and $\mathbb{P}^{(1)} \times \mathbb{P}^{(2)} \subset \mathbb{P}$. The independence of the two probability spaces is reasonable considering the low correlation between the underlying catastrophic risks and financial risks. While the latter can be replicated with an available portfolio, the former cannot. This naturally leads us to consider an incomplete market setting for evaluating catastrophe bonds. There is no unique price for a security in an incomplete market setting as there could be infinitely many prices corresponding to different equivalent martingale measures. Merton (1976) posits that under the risk-neutral pricing measure $\mathbb{Q}$, the overall economy depends only on financial variables. There is, however, evidence to the contrary that catastrophe risk may pose a substantial systematic risk (see, e.g., Gürşler et al., 2016). Merton (1976) argues that the aggregate loss process (i.e., intensity and severity of losses) retain their original distributional characteristics $\mathbb{P}^{(1)}$ after been transformed from the physical probability measure to the risk-neutral measure $\mathbb{Q}^1$ (see for example, Doherty (1997); Cox and Pedersen (2000); Lee and Yu (2002); Ma and Ma (2013)). In other words, the expectation of a random variable $X$ on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0,T]}, \mathbb{P})$ is the same under $\mathbb{Q}$. 

35
\( E^\mathbb{P}[X] = E^\mathbb{Q}[X] \)

To develop a fair pricing model that is consistent with both financial and actuarial valuation, we follow Tang and Zhongyi (2019) and apply a distortion function to both physical probability measures \( \mathbb{P}^{(1)} \) and \( \mathbb{P}^{(2)} \) to transform them to their risk-neutral counterparts \( \mathbb{Q}^{(1)} \) and \( \mathbb{Q}^{(2)} \). Tang and Zhongyi (2019) show that following a distorted probability measure \( \mathbb{Q}^{(1)} \), \( Y \) becomes heavier tailed than \( \mathbb{P}^{(1)} \), hence the riskiness is amplified under \( \mathbb{Q}^{(1)} \). One of the most popular risk-neutral approaches is to use the Wang transform proposed by Wang (2000, 2003, 2004). In this paper, to be consistent with our Bayesian pricing framework, we convert the physical distributions \( \mathbb{P}^{(1)} \) and \( \mathbb{P}^{(2)} \) into their risk-neutral forms using the maximum information entropy proposed by Stutzer (1996) and Li (2010).

### 5.1 Pricing Model

Let us consider the trigger process of a CAT bond to be an index, \( Y(t) \). Then the CAT bond is a derivative product that pays off \( P_{CAT}(Y(t)) \) for \( t = 1, 2, ..., T \) according to a payoff function \( P_{CAT}(\cdot) \), where \( P_{CAT}(\cdot) \) represents a CAT bond whose payments are contingent on \( Y(t) \)'s process. Let \( \{ r_t, t = 1, ..., T \} \) be an interest rate process. Conditioning on available information on catastrophe risk and financial market performance, the risk-neutral price for a CAT bond which pays \( P_{CAT}(Y(t)) \) at time \( t, t = 1, ..., T \) is

\[
V_t = K \left[ E^{\mathbb{Q}^{(2)}} \left( e^{-\int_t^T r_s ds} | \mathcal{F}_t^{(2)} \right) \right] * E^{\mathbb{Q}^{(1)}} \left( P_{CAT}(T) | \mathcal{F}_t^{(1)} \right)
\]

\[
V_t = K E^{\mathbb{Q}} \left[ \left( e^{-\int_t^T r_s ds} \cdot P_{CAT}(T) | \mathcal{F}_t \right) \right]
\]

where \( K \) is the face value, \( \mathcal{F}_t = \mathcal{F}_t^{(1)} \times \mathcal{F}_t^{(2)} \subset \mathcal{F} \) and \( \mathcal{Q} = \mathbb{Q}^{(1)} \times \mathbb{Q}^{(2)} \). \( E^{\mathbb{Q}}[\cdot | \mathcal{F}_t] \) is the expectation of the bond price for the risk-neutral measure. Our task is to statistically model the two independent processes \( Y(t) \) and \( r_t \) and convert them to their equivalent risk-neutral measures.
5.2 Payoff Function

According to the CAT bond payment structure, investors receive premiums if the bond is not triggered. There are several trigger approaches; however, we utilize the industry index trigger to price CAT bonds in this study. This means that investors may lose their principal if the estimated aggregated loss for the whole industry exceeds a predetermined threshold. CAT bonds can be designed with different payoff functions. In this paper, we consider two payoff functions for a CAT bond contract. Assume a zero-coupon CAT bond with maturity date $T$ has a payoff function as follows:

$$P_{\text{CAT}}^{(1)} = \begin{cases} K, & \text{if } L(T) \leq D, \\ a \cdot K, & \text{if } L(T) > D, \end{cases}$$

(5.2)

where $L(T)$ is a total insured loss at the expiry date $T$, $D$ is the industry index threshold value pre-specified in the bond contract, and $a \in [0, 1)$ is a fraction of the principal $K$, which the bondholders must pay when the bond is triggered. Consider another payoff function with a coupon payment at the maturity date, if the trigger event has not occurred, of the form

$$P_{\text{CAT}}^{(2)} = \begin{cases} K + c, & \text{if } L(T) \leq D, \\ K, & \text{if } L(T) > D, \end{cases}$$

(5.3)

where $c > 0$ is the coupon payment.

Insert payment function of Eq. (5.2) into Eq. (5.1) for the price of a zero-coupon bond, and we get

$$V_t = KE^Q\left[\left(e^{-\int_t^T r_s ds} \cdot P_{\text{CAT}}(T)|F_t \right)\right]$$

Since we are using the CIR bond prices $e^{-\int_t^T r_s ds} = B_{\text{CIR}}(t, T)$

$$V_t^{(1)} = KE^Q\left[\left(B_{\text{CIR}}(t, T) \cdot P_{\text{CAT}}(T)|F_t \right)\right]$$

$$= E^Q\left[B_{\text{CIR}}(t, T) \cdot \left(K \cdot P(L(T) \leq D) + a \cdot K \cdot P(L(T) > D)\right)|F_t \right]$$

$$= E^Q\left[B_{\text{CIR}}(t, T) \cdot \left(K \cdot F(T, D) + a \cdot (1 - F(T, D))\right)\right]$$

(5.4)

For the case where the CAT bond pays principal plus a coupon at maturity if the event has not been triggered, the value of the bond can be evaluated by inserting Eq. (5.3) into Eq. (5.1):

$$V_t^{(2)} = KE^Q\left[\left(B_{\text{CIR}}(t, T) \cdot P_{\text{CAT}}(T)|F_t \right)\right]$$
\begin{align*}
  &= E^Q[B_{CIR}(t, T) \cdot (K+c \cdot \mathbb{1}_{L(T) \leq D} + K\mathbb{1}_{L(T) > D}) | \mathcal{F}_t] \\
  &= E^Q[B_{CIR}(t, T) \cdot (K + c \cdot Pr(L(T) \leq D) + K \cdot Pr(L(T) > D)] \\
  &= E^Q \left[ B_{CIR}(t, T) \cdot \left((K + c) \cdot F(T, D) + K \left(1 - F(T, D)\right)\right) \right] \\
  \end{align*}  

(5.5)

5.2.1 An Illustrative Example

We investigate the effects of interest rate and catastrophic risks on single- and multi-peril CAT bond prices with a face value $K = CA$100. We utilize the bond price formula in Eq. 5.4–5, and the parameter estimate for the Bayesian CIR model and Hierarchical Dirichlet Process model (HDP) derived in Sections 3 and 4. Specifically, we illustrate how bond prices evolve with changing maturity times and varying catastrophe thresholds. We price a CAT bond using the spot interest rate process outlined in Section 3. From our estimation, we conclude that both the initial short-term interest rate, $r_0$ and the long-term average annual interest rate, $\alpha / \beta$ is 0.9267%. We assume a maturity period $T \in [0, 2]$ years, a fraction of capital lost, $\alpha = 0.5$ when the aggregate loss $L(T)$ exceeds the threshold level $D \in [13.25, 193.60]$ ten million CAD (i.e., the threshold interval of 50th to 99th quantile of the annual loss). Lastly, the coupon rate payment in Eq. 5.5 for a coupon-bearing bond is $c = CA$10. Figures 5.1 – 2 report the CAT bond prices for payoff function $P_{CAT}^{(1)}$ for the four distinct groupings of catastrophe risk profiles under the HDP model.

---

9 Bond prices are simulated using the physical or real-world valuation for both zero-coupon and coupon bonds (i.e., $V_1^{(1)} = E^P[B_{CIR}(t, T) \cdot (K( F(T, D) + a(1 - F(T, D)))$ and $V_1^{(2)} = E^P \left[ B_{CIR}(t, T) \cdot \left((K + c) \cdot F(T, D) + K \left(1 - F(T, D)\right)\right) \right]$.)
Figure 5.1: Zero-coupon CAT bond price simulation for groups 1 and 2 under the HDP and stochastic interest rate assumptions.

(a) $V^{(1)}(t)$ for multi-peril bond group 1

(b) $V^{(1)}(t)$ for multi-peril bond group 2

Figure 5.2: Zero-coupon CAT bond price simulation for groups 3 and 4 under the HDP and stochastic interest rate assumptions.

(a) $V^{(1)}(t)$ for single-peril bond group 3

(b) $V^{(1)}(t)$ for single-peril bond group 4

10 Multi-Peril Group 1: Windstorm, Severe storm, Tornadoes and Hailstorm
11 Multi-Peril Group 2: Flood, Tornado and Hurricane
12 Single-Peril Group 3: Tropical storm
13 Single-Peril Group 4: Fire
Figures 5.1 – 2 show that the price of CAT bonds, $V(t)$, decreases as the threshold level, $D$, decreases, and as the time to maturity, $T$, increase. We can also observe from the bond price surfaces that single-peril bond prices show some difference in terms of shapes from the multi-peril bonds. Notably, for single-peril bonds (i.e., Groups 3 and 4), there’s a sharp drop in prices in the lower end of threshold levels. In comparison, multi-peril bonds (i.e., Groups 1 and 2) show a gradual decrease in price as threshold levels decrease. This difference may explain the price difference in single- and multi-peril bonds and differences in the risk premia. In Figure 5.3, we show the price difference between Single-Peril Group 4 and Multi-Peril Group 1.

The observed positive price difference implies that indeed Single-Peril Group 4 is priced higher than Multi-Peril Group 1. While this may not be a rule or not always the case for multi- and single-peril CAT bonds, it underscores the relevance of exploring the individual and grouped risk profiles when assessing CAT bond prices designed around an industry trigger index. There’s a significant price difference at lower threshold levels, but this difference reduces to zero at higher threshold levels. For a fixed threshold interval and time to maturity, the prices of CAT bonds are affected mainly by the rate of predicted claims to occur according to the HDP model.

Figure 5.3: Zero-coupon simulated bond price difference between Multi-Peril Group 1 and Single Peril Group 4 under the HDP and stochastic interest rate assumptions.

Therefore, the aggregate loss, conditional on the number of claims, will be high when there’s an expected high number of claims and vice-versa (see Table 3.3 for this difference). When aggregate claims are high enough, there is a relatively higher probability of breaching the pre-specified
industry threshold, triggering the bond to be paid out to the re-insurer or insurer. Consequently, CAT bonds with higher predicted aggregate claims such as Multi-Peril Group 1 have lower expected prices relative to Single-Peril Group 4 with lower predicted aggregate claims. In Section 6, we will examine how these price differences affect the risk premia on bonds. In Figures 5.4 – 5 the CAT bond prices for payoff function $P_{CAT}^{(2)}$ is presented. As illustrated below, the bond prices are all higher than the zero-coupon bond in all groupings. This indicates that the choice of payoff function has a significant impact on the CAT bond prices.

(a) $V^{(2)}(t)$ for multi-peril bond group 1  
(b) $V^{(2)}(t)$ for multi-peril bond group 2

Figure 5.4: Coupon-bearing CAT bond price simulation for groups 1 and 2 under the HDP and stochastic interest rate assumptions.
Figure 5.5: Coupon-bearing CAT bond price simulation for groups 3 and 4 under the HDP and stochastic interest rate assumptions.

6.0 Risk-Neutralization of CAT Bond Prices

One of the advantages of the Bayesian approach to risk-neutralization is that we can account for the uncertainty in our pricing model that arises from estimating model parameters and possibly uncertainty in choosing between models. For a more practical appeal of the Bayesian approach, see Li (2014), Kogure and Fushimi (2018), Li, Kogure, and Liu (2019). From a Bayesian perspective, the physical probability measure can be considered our prior distribution of CAT bond prices. This prior knowledge gets updated to the posterior (risk-neutral) distribution, when we observe issue prices at time-0 via the maximum entropy principle.

6.1 Pricing Mechanism Under the Entropy Principle

We simultaneously generate $N$ states of the MCMC sampling for the CIR bond prices \{$B_{CIR}(t), t = 1,2,\ldots,T\}$ and contingent payoffs \{$P_{CAT}(Y_t), t = 1,2,\ldots,T\}$, and denote them by
\{\left( tB_{CIR}^{(i)}, tP_{CAT}^{(i)} \right), \left( 2B_{CIR}^{(i)}, 2P_{CAT}^{(i)} \right), \ldots, \left( tB_{CIR}^{(i)}, tP_{CAT}^{(i)} \right) \}_{i = 1, 2, \ldots, N}\). We let \( \pi \) denote the physical distribution of the \( N \) states of the MCMC sampling with equal probability of \( \frac{1}{N} \) in each state. To implement risk-neutralization, we need at least one market price constraint. There’s currently no active CAT bond designed around the CATIQ index. For illustrative purposes, we use a hypothetical risk premium \( \delta \) of 250 basis points above LIBOR to discount future payoffs to the present as

\[
V_0 = \exp(-\delta t)\mathbb{E}^P[P_{CAT}(t)|\mathcal{F}_t] = \sum_{i=1}^{N} \sum_{t=1}^{T} \exp(-\delta t) \left( tB_{CIR}^{(i)} tP_{CAT}^{(i)} \pi_i \right) \text{ for } i = 1, \ldots, N \tag{6.1}
\]

where \( V_0 \) is the issue price of the bond at time \( t = 0 \). We acknowledge that using a hypothetical risk premium for the price constraint does not reflect the accurate market price of risk. However, for this numerical exercise, it may suffice. We convert \( \pi \) into the risk-neutral version \( \pi^* \) using the market constraint such that

\[
\sum_{i=1}^{N} \sum_{t=1}^{T} \left( tB_{CIR}^{(i)} tP_{CAT}^{(i)} \pi_i^* \right) = V_0 \text{ for } i = 1, \ldots, N, \quad t = 1, \ldots, T \tag{6.2}
\]

where \( \sum_{i=1}^{N} \sum_{t=1}^{T} \left( tB_{CIR}^{(i)} tP_{CAT}^{(i)} \right) \pi_i^* = \mathbb{E}_Q[\cdot|\mathcal{F}_t] \)

The intuition is that the expected price at time \( t = 0 \) must equal the payoff at \( t = T \), but calculated using the risk-neutral measure instead of the physical probability measure and discounted back using the risk-free rate. Therefore, the risk-neutral approach corrects for risk by adjusting probabilities rather than adjusting the risk-free rate. It corrects physical probabilities to overweight states in which aggregate outcomes are particularly bad\(^{14}\). This is the reason why martingales or risk-neutral methods are used to price risky payoffs. Based on the maximum entropy principle, the risk-neutral distribution \( \pi^* \) should minimize the Kullback-Leibler information divergence

\[
\text{KL}(P \parallel Q) : \arg \min \sum_{i=1}^{N} \pi_i^* \ln \left( \frac{\pi_i^*}{\pi_i} \right)
\]

with the following additional constraints \( \pi^* > 0, \text{ for } i = 1, \ldots, N, \sum_{i=1}^{N} \pi_i^* = 1 \). This minimization problem can be solved by the method of Lagrange multipliers as

\(^{14}\) There is a close linkage between the risk-neutral measure and the Arrow-Debrau theory under both its no-arbitrage argument and equilibrium conditions. Interested reader can refer to (Duffie, 2010).
\[
\pi_i^* = \frac{\pi_i \exp \left( \lambda \sum_{t=1}^{T} \left( t^{(i)}_{\text{CIR}} \cdot t^{(i)}_{\text{CAT}} \right) \right)}{\sum_{i=1}^{N} \pi_i \exp \left( \lambda \sum_{t=1}^{T} \left( t^{(i)}_{\text{CIR}} \cdot t^{(i)}_{\text{CAT}} \right) \right)} \text{ for } i = 1, ..., N, t = 1, ..., T. \tag{6.3}
\]

For brevity of expression, let \( \alpha_i \triangleq \sum_{t=1}^{T} \left( t^{(i)}_{\text{CIR}} \cdot t^{(i)}_{\text{CAT}} \right) \) then Eq. 6.3 can be simplified as

\[
\pi_i^* = \frac{\pi_i \exp (\lambda \alpha_i)}{\sum_{i=1}^{N} \pi_i \exp (\lambda \alpha_i)} \text{ for } i = 1, ..., N \tag{6.4}
\]

The \( \lambda \) is the Lagrange multiplier happens to be the minimizer which can be calculated by solving the minimization problem

\[
\arg\min_{\lambda} \sum_{i=1}^{N} \exp \left[ \lambda (\alpha_i - V_o) \right] \tag{6.5}
\]

See Appendix A for the solution and full proofs. Therefore the risk-neutral price for a zero-coupon can be approximated as

\[
V^{(1)}_i = K E^Q \left[ \left( B^{(i)}_{\text{CIR}}(t, T) \cdot P^{(i)}_{\text{CAT}}(T) \right) | \mathcal{F}_t \right] \approx \left[ \sum_{i=1}^{N} \sum_{t=1}^{T} \left( t^{(i)}_{\text{CIR}} \cdot t^{(i)}_{\text{CAT}} \right) \pi_i^* \right] \tag{6.7}
\]

6.1.1 An Illustrative Example

We drew 10,000 samples from the predictive distributions of HDP and CIR models. Using the expression in Eq. 6.7, we calculate the present values of a two-year zero-coupon CAT bond with a face value of \( K = \text{CA}\$100 \) belonging to groups 1 – 4. We use a single threshold \( D \in [75.93] \text{ ten million CAD (3*average annual loss)} \), an initial short-term interest rate, \( r_0 \) of 0.9267%. The results are shown in Figure 6.1 for multi-peril Group 1—2 and Figure 6.2 for single-peril Group 3—4. The summary statistic of the present values is presented in Table 6.1.
Multi-peril bond group 1

Multi-peril bond group 2

Figure 6.1: Distributions of the present values for multi-peril CAT bond

Single-peril bond group 3

Single-peril bond group 4

Figure 6.2: Distributions of the present values for single-peril CAT bond

It may be interesting to note that the expected theoretical price of the CAT bonds across all four groups is relatively higher in the risk-neutral distribution than the physical one. Our results also show that the risk-neutral distribution is skewed to the right of the physical distribution for all
cases, reflecting the risk adjustment for interest rate and catastrophe risk. The risk-neutral prices may compensate for the additional risk that may arise from model and parameter uncertainties on a risk-adjusted basis. From a Bayesian viewpoint, the posterior (risk-neutral) distribution yields more information. This is because we update our beliefs about prices after having observed the market prices.

6.2 Risk Premium Assessment

Due to the sophisticated pricing mechanism of CAT bonds, they are limited to institutional investors such as hedge funds, pension funds rather than the open market. A buy-side investor going into a private placement will be better equipped if they have an idea of the expected risk premium based on the prevailing interest rate environment and catastrophe risk characteristics. Consequently, a more helpful expression with the straightforward interpretation would be to calculate the risk premium per annum, which can easily be derived from solving the root of the Eq. 6.8 for $\delta$. We want to find a unique risk premium that can make the two discounted payoffs equal. Eq. 6.8 is essentially the same as Eq. 6.1, where we specify the market constraint.

$$\sum_{t=1}^{T} (E^{Q}[ (B_{CIR}(t,T) \cdot P_{CAT}(T)|\mathcal{F}_t ) ] - \exp (-\delta t)E^{P}[ (B_{CIR}(t,T) \cdot P_{CAT}(T)|\mathcal{F}_t ) ]) $$ (6.8)

We calculate the risk premia with maturities ranging from 1 to 10 years, as presented in Figure 6.2.
To ensure that CAT bonds are priced correctly, we assume the same risk premium for all the peril groups. However, there may be differences in risk premia for different perils and maturities. From Figure 6.2, we can observe that for a fixed index trigger threshold, peril-specific CAT bonds exhibit different risk premium profiles. First, all CAT bond groups show an increase in risk premium as their maturity time increases. This is expected since an investor’s risk exposure increase with higher holding periods; hence they have to be compensated with a higher risk premium. Risk premia at lower maturities are flat (1 – 5 years) with no discernable differences across peril groups but increase sharply afterward. There exist a non-linear and somewhat exponential relationship between maturity and risk premia. This may explain why most CAT bonds have short maturity dates (typically between 2 – 5 years). However, there are some with longer maturities. The short maturity of the bonds somewhat mitigates the risk of losing principal invested. Riskier bonds (i.e., perils with a higher probability of breaching threshold) show higher risk premiums such as Multi-peril group 1 and vise versa such as Single-peril group 4.

Figure 6.2: Annualized risk premia under stochastic interest rate and HDP catastrophe models
7.0 Concluding Remarks

This study presents a fully Bayesian approach as an alternative to modeling CAT bond prices. Mainly, the focus has been on assessing the catastrophe risk profiles of peril-specific bond contracts and the consequently expected risk premia. This study may yield additional insights into our understanding of how multi-peril and single-peril CAT bond contracts are priced. 15

This novel approach is consistent with financial asset pricing and actuarial valuation theories. The method is flexible, robust and takes care of all uncertainties in the estimation process. The methodology can offer practical uses. The architecture of the HDP model can accommodate several characteristics of risk in the calculation of aggregate claims, hence yielding robust estimates. For example, in constructing the non-homogenous Poisson process, one may include location-specific factors can influence aggregate claims. There may be specific geographies that may be more susceptible to hurricanes, for example. Another advantage is the dynamic grouping induced in the posterior samples of parameters. It ensures that when new information or data is available, these groupings can adjust/change to measure risk correctly. The Bayesian CIR model offers more precision in the estimation of future interest rates. The Euler-Maruyama discretization scheme and data augmentation address the difficulty in approximating a continuous process such as interest rate with discretely observed data.

The maximum entropy approach for calculating the risk premia requires no subjective input from the user as, is the case with other techniques such as the Wang Transform (Li, 2010). The maximum entropy approach to risk-neutral pricing can incorporate different market prices, which is pertinent is the CAT budding bond market. The numerical methods outlined in this study may be computationally expensive, especially in calculating catastrophe probabilities. However, with the availability of cheap computing power, the benefit of this methodology far outweighs the cost.

An avenue for future research is to consider a three-parameter family of continuous probability distributions. We believe that it may enhance the flexibility of our HDP model for catastrophe risk. A generalized inverse gamma density \( f(x; a, d, p) \) may show a different rate of decay for different peril, which can enhance our understanding.

15 Current ratio of single peril to multi-peril CAT bond outstanding is 4:6 (see, e.g., Herrmann and Hibbeln, (2021)).
Bibliography


### Appendix A

#### Table 1A: Grouping based on catastrophe counts ($N_i$)

<table>
<thead>
<tr>
<th>Peril</th>
<th>Group indicators and corresponding frequencies.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Windstorm)</td>
<td>Group Indicator 3</td>
</tr>
<tr>
<td></td>
<td>Frequency 30000</td>
</tr>
<tr>
<td>2 (Severe Storm)</td>
<td>Group Indicator 1</td>
</tr>
<tr>
<td></td>
<td>Frequency 30000</td>
</tr>
<tr>
<td>3 (Hailstorm)</td>
<td>Group Indicator 9</td>
</tr>
<tr>
<td></td>
<td>Frequency 30000</td>
</tr>
<tr>
<td>4 (Winter Storm)</td>
<td>Group Indicator 7</td>
</tr>
<tr>
<td></td>
<td>Frequency 30000</td>
</tr>
<tr>
<td>5 (Flood)</td>
<td>Group Indicator 6</td>
</tr>
<tr>
<td></td>
<td>Frequency 30000</td>
</tr>
<tr>
<td>6 (Tornado)</td>
<td>Group Indicator 5</td>
</tr>
<tr>
<td></td>
<td>Frequency 30000</td>
</tr>
<tr>
<td>7 (Hurricane)</td>
<td>Group Indicator 7</td>
</tr>
<tr>
<td></td>
<td>Frequency 30000</td>
</tr>
<tr>
<td>8 (Tropical Storm)</td>
<td>Group Indicator 8</td>
</tr>
<tr>
<td></td>
<td>Frequency 30000</td>
</tr>
<tr>
<td>9 (Fire)</td>
<td>Group Indicator 2, 4</td>
</tr>
<tr>
<td></td>
<td>Frequency 10560, 19440</td>
</tr>
</tbody>
</table>

Finding the maximum entropy by the process of Lagrange multipliers

Based on the maximum entropy principle, the risk-neutral distribution $\pi^*$ should minimize the Kullback-Leibler information divergence

$$KL(\mathbb{P} \parallel \mathbb{Q}) = \arg \min \sum_{i=1}^{N} \pi^*_i ln \left( \frac{\pi^*_i}{\pi_i} \right)$$

with the following additional constraints $\pi^*_i > 0$, for $i = 1, ..., N$, $\sum_{i=1}^{N} \pi^*_i = 1$.

This minimization problem can be solved by the method of Lagrange multipliers as follows:

$$\mathcal{L} = \sum_{i=1}^{N} \pi^*_i ln \left( \frac{\pi^*_i}{\pi_i} \right) - \gamma \left( \sum_{i=1}^{N} \pi^*_i - 1 \right) - \lambda \left[ \sum_{i=1}^{N} \sum_{t=1}^{T} \left( tB^{(i)}_{\text{CIR}} tB^{(i)}_{\text{CAT}} \pi_i^* \right) - V_0 \right]$$

$\gamma \in \mathbb{R}, \lambda \in \mathbb{R}$
Taking first order conditions with respect to $\gamma$ and $\lambda$, we get

$$\frac{\partial \mathcal{L}}{\partial \gamma} = -\gamma (\sum_{i=1}^{N} \pi_i^* - 1) = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{i=1}^{N} \sum_{t=1}^{T} (t_{B_{CIR}^{(i)}} t_{P_{CAT}^{(i)}}) \pi_i^* = 0$$

Taking first order conditions with respect to $\pi_i^*$, we get

$$\frac{\partial \mathcal{L}}{\partial \pi_i} = \ln \frac{\pi_i^*}{\pi_i} + 1 - \gamma - \lambda \left[ \sum_{t=1}^{T} (t_{B_{CIR}^{(i)}} t_{P_{CAT}^{(i)}}) \pi_i^* \right] = 0$$

$$\exp \left( \ln \frac{\pi_i^*}{\pi_i} + 1 - \gamma - \lambda \left[ \sum_{t=1}^{T} (t_{B_{CIR}^{(i)}} t_{P_{CAT}^{(i)}}) \pi_i^* \right] \right) = 1$$

$$\exp \left( \ln \pi_i^* - \ln \pi_i + 1 - \gamma - \lambda \left[ \sum_{t=1}^{T} (t_{B_{CIR}^{(i)}} t_{P_{CAT}^{(i)}}) \pi_i^* \right] \right) = 1$$

$$\pi_i^* = \pi_i \exp \left( -1 + \gamma + \lambda \left[ \sum_{t=1}^{T} (t_{B_{CIR}^{(i)}} t_{P_{CAT}^{(i)}}) \pi_i^* \right] \right)$$

Expanding to get

$$\pi_i^* = \exp(-1 + \gamma) \pi_i \exp \left( \lambda \sum_{t=1}^{T} (t_{B_{CIR}^{(i)}} t_{P_{CAT}^{(i)}}) \right)$$

(2)

Since $\sum_{i=1}^{N} \pi_i^* = 1$ and $\sum_{i=1}^{N} \pi_i = 1$, it follows that

$$\sum_{i=1}^{N} \pi_i^* = \exp(-1 + \gamma) \sum_{i=1}^{N} \pi_i \exp \left( \lambda \sum_{t=1}^{T} (t_{B_{CIR}^{(i)}} t_{P_{CAT}^{(i)}}) \right) = 1$$

Dividing through, we get

$$\exp(-1 + \gamma) = \frac{1}{\sum_{i=1}^{N} \pi_i \exp \left( \lambda \sum_{t=1}^{T} (t_{B_{CIR}^{(i)}} t_{P_{CAT}^{(i)}}) \right)}$$

(3)

Insert equation (3) in to equation (2) to get

$$\pi_i^* = \frac{\pi_i \exp \left( \lambda \sum_{t=1}^{T} (t_{B_{CIR}^{(i)}} t_{P_{CAT}^{(i)}}) \right)}{\sum_{i=1}^{N} \pi_i \exp \left( \lambda \sum_{t=1}^{T} (t_{B_{CIR}^{(i)}} t_{P_{CAT}^{(i)}}) \right)} \text{ for } i = 1, ..., N, \quad t = 1, ..., T. \quad (4)$$

For brevity of expression, let $\alpha_i = \sum_{t=1}^{T} \left( t_{B_{CIR}^{(i)}} t_{P_{CAT}^{(i)}} \right)$, then (4) can be simplified as

52
\[
\pi_i^* = \frac{\pi_i \exp(\lambda \alpha_i)}{\sum_{i=1}^{N} \pi_i \exp(\lambda \alpha_i)} \text{ for } i = 1, \ldots, N \quad (5)
\]

To avoid overflow in the numerical calculation of exponents, let \( \Gamma = \max \{\lambda \alpha_i\} \). Then (5) is equivalent to

\[
\pi_i^* = \frac{\pi_i \exp(\lambda \alpha_i - \Gamma)}{\sum_{i=1}^{N} \pi_i \exp(\lambda \alpha_i - \Gamma)} \text{ for } i = 1, \ldots, N
\]

the \( \lambda \) is the Lagrange multiplier and can be found in the following steps.

First, we simply (1) as

\[
\sum_{i=1}^{N} \pi_i^* \alpha_i = V_o \quad (6)
\]

Substituting (5) into (6), we get

\[
\sum_{i=1}^{N} \frac{\pi_i \exp(\lambda \alpha_i)}{\sum_{i=1}^{N} \pi_i \exp(\lambda \alpha_i)} \alpha_i = V_o
\]

Remembering that \( \pi_i = \frac{1}{N} \), we can simplify to

\[
\frac{\sum_{i=1}^{N} \alpha_i \exp(\lambda \alpha_i)}{\sum_{i=1}^{N} \exp(\lambda \alpha_i)} = V_o \quad (7)
\]

This happens to be the minimizer of the following minimization problem

\[
\arg \min_{\lambda} \sum_{i=1}^{N} \exp [\lambda(\alpha_i - V_o)] \quad (8)
\]

And it can be proved as follows:

Let \( f(\lambda) = \sum_{i=1}^{N} \exp [\lambda(\alpha_i - V_o)] \), we find the first order conditions and equate the result to zero, then

\[
\frac{df}{d\lambda} = \sum_{i=1}^{N} (\alpha_i - V_o) \exp [\lambda(\alpha_i - V_o)]
\]

\[
\exp[-\lambda V_o] \sum_{i=1}^{N} (\alpha_i - V_o) \exp[\lambda \alpha_i] = 0
\]
That is

\[
\sum_{i=1}^{N} \{\alpha_i - V_o\} \exp[\lambda \alpha_i] = 0
\]

\[
\sum_{i=1}^{N} \alpha_i \exp[\lambda \alpha_i] = V_o \sum_{i=1}^{N} \exp[\lambda \alpha_i]
\]

Therefore,

\[
\frac{\sum_{i=1}^{N} \alpha_i \exp(\lambda \alpha_i)}{\sum_{i=1}^{N} \exp(\lambda \alpha_i)} = V_o \tag{9}
\]

Equation (9) is the same form as (7). So solving the minimization problem in eqn. (8) is identical to solving eqn. (7).