Abstract

This paper introduces an innovative approach to catastrophe (CAT) bond pricing, addressing the shortcomings of existing methods that don’t fully encompass catastrophe data characteristics or analyze the influence of both loss and inter-arrival time distributions on the CAT bond price. Our method combines the compound renewal process with the Cox-Ingersoll-Ross (CIR) process as a product measure to separately model uncertainties in insurance and financial risks. Valuation is performed in two steps, integrating risk-neutral measures for financial risks and a class of measures for insurance risks, preserving the structure of a renewal compound process. By using Bayesian inference, historical data, and capital market insights, we calibrate the pricing model effectively. Uniquely, our framework evaluates the impact of varying inter-arrival time distributions on CAT bond prices, an area previously unexplored in literature. This approach also separates market prices for claim frequency and severity under certain renewal process conditions. Empirical results...
reveal that inter-arrival time distribution notably influences the CAT bond price.

**JEL classification:** C11; C15; Q54; G12; G22
1 Introduction

As catastrophe risk stemming from climate change increases, insurers and re-insurers are turning to financial markets for capital adequacy. (Re-)insurers use insurance-linked securities (ILS) such as catastrophe (CAT) bonds to raise additional capital to increase their capacity to handle the increasing risk of natural disasters. For a financial investor, these bonds are particularly attractive due to their so-called zero-beta (or low correlation to the broad financial market) and relatively high yield. A CAT bond is an exotic derivative whose cash-flows are contingent on the occurrence of natural disasters. CAT bonds have become popular as an alternative risk transfer mechanism since their inception in 1994 with an estimated outstanding market capitalization of $37 billion as of 2023¹.

The increase in CAT bonds popularity as a viable alternative investment has drawn the attention of researchers and investors to understand how these instruments are priced. CAT bonds are particularly unique, since their underlying risk cover two distinct sources of risk (i.e., interest rate and catastrophe risk) which makes their valuation challenging. The pricing of CAT bond contracts is of great importance as it determines a fair value at which the insurer and the investor are willing to enter a deal. However, from a technical point of view, the approach of finding the price is not straightforward in light of the existence of both insurance and financial risks. To this end, many researchers have proposed different methods of pricing.

Some authors (see for example, Cox and Pedersen (2000), Lee and Yu (2002), Ma and Ma (2013)) have extended the risk-neutral approach used in the valuation of primitive financial derivatives to price CAT bonds. Traditional risk-neutral valuation principles are not directly applicable to CAT bonds. Since CAT bonds have an actuarial dimension, the use of pure financial valuation principles may under-price the inherent catastrophe risk (Domfeh et al. 2022). The catastrophe risk component of CAT bonds lies outside the financial market which naturally leads to an incomplete market setting for CAT bonds. Under an incomplete market, there is no unique price of a financial instrument. In other words, there is no unique equivalent martingale under risk-neutral valuation. Regardless, some researchers have taken inspiration from other markets such as credit derivatives in the face of incomplete markets. Under these market settings, risk-neutral is achieved by using a pricing kernel implied by the market. This means that the models used in valuation are calibrated to observed market prices. For instance, Beer and Braun (2022) developed a market-consistent pricing model for CAT bonds using observed secondary CAT bond prices and implied catastrophe intensity rate. Domfeh et al. (2022) utilized

¹https://www.artemis.bm/dashboard/
an entropy risk-neutral measure under a Bayesian framework to derive market-consistent risk premia for CAT bonds in the primary market.

The intractability of the catastrophe risk component in CAT bond pricing is often due to the inability to replicate payoffs or cash flows linked to catastrophes. Earlier researchers have made several strong assumptions about the CAT bond markets, allowing them to price these instruments as they would for traditional financial derivatives. However, these assumptions were quite restrictive and did not capture the inherent risk in pricing. For instance, Cox and Pedersen (2000), Lee and Yu (2002), Ma and Ma (2013) treat CAT bonds as zero-beta security. Under this assumption, investors should earn a zero-risk premium for holding CAT bonds in their portfolio because catastrophe is uncorrelated with the overall market. In other words, CAT bonds have idiosyncratic risk (Merton, 1976). This assumption has an immediate consequence on how the catastrophe risk component of CAT bonds is treated in valuation. Specifically, the aggregate loss process (i.e., loss intensity and severity) retains its physical distribution under the risk-neutral measure without any transformation or adjustment. The resultant risk-neutral price is only driven by the interest rate process. However, other researchers have attempted to relax this strong assumption by deriving an independent risk-neutral measure for the catastrophe risk component of CAT bonds. For example, Tang and Zhongyi (2019) apply a distortion measure known as the Wang Transform to the catastrophe risk component in their two-step valuation approach. The Wang Transform amplifies the catastrophe risk under the risk-neutral measure (see e.g, Hamada and Sherris (2003), Kijima and Murimachi (2008), Li et al. (2013)).

The main contribution of this paper is to generalize a family of equivalent risk-neutral measures for catastrophe risk and apply them to CAT bond valuation. We achieve this by using the product probability measure approach proposed by Tang and Zhongyi (2019). Under the product measure, CAT bonds’ sources of risk (financial and catastrophic) are independent. Therefore, the resultant pricing measure is a product of two independent measures, $Q_1$ and $Q_2$ respectively. We extend and generalize the characterization of equivalent measures inspired by the works of Macheras and Tzaninis (2020) in the context of CAT bond pricing. The authors introduced a class of equivalent measures such that a compound renewal process under the physical measure remains a compound renewal process under its corresponding equivalent measure. We utilize the Macheras and Tzaninis approach to transform the physical distribution of catastrophe risk into its equivalent counterpart.

In the CAT bond literature, many researchers treat the aggregate claim process as a compound Poisson process where the counting process follows an exponentially distributed inter-arrival
time. Burnecki et al. (2019) propose a time-inhomogeneous compound Poisson process to formalize a design for pricing contingent convertible catastrophe bonds (CocoCat). Burnecki and Giuricich (2019) in their attempt to address the heavy-tailed nature of catastrophe losses still maintains a non-homogeneous Poisson process with a deterministic intensity function of time. However, due to the un-anticipatory nature of catastrophic events, the assumptions that claims occur as (non)homogeneous Poisson process may be unsatisfactory. To address this issue, Ma et al. (2017) propose a doubly stochastic Poisson process (also known as a Cox process) where the Poisson intensity is stochastic, to model the claim arrival process. Their results reveal interesting patterns in the price formation of CAT bonds.

We argue that the aggregate loss process need not necessarily be a compound Poisson process, but rather a flexible compound process depending on the data generation process itself. For example, the inter-arrival process could follow a gamma distribution. The renewal process is a general process which includes a Poisson process. In this paper, we formalize the generalization of the compound renewal process to include other distributions inspired by the recent work of Macheras and Tzaninis (2020). We apply our framework to earthquake dataset and calibrate the model under several scenarios and test the sensitivity of CAT bond prices where the inter-arrival time follows either Exponential, Gamma, Weibull, or Mixed-exponential distributions.

We model the interest rate risk component of catastrophe bonds as another stochastic process that follows Cox-Ingersoll-Ross (CIR) model.

The remainder of the paper is organized as follows: Section 2 discusses the model assumptions and valuation framework. Section 3 provides details on the dynamics of interest rate risk under both physical and risk-neutral measures. Section 4 elaborates on the distribution of the aggregate loss process under both measures. In Section 5, the CAT bond prices associated with two scenarios and model assumptions thereof are derived. Section 6 presents a numerical illustration that includes parameter estimation, model calibration, and sensitivity analysis. Finally, Section 7 concludes the paper.

2 Model assumption and valuation framework

In this section, we build up our valuation framework for CAT bond pricing based on the product pricing measure proposed by Tang and Zhongyi (2019).

We introduce some notions on the probability space on which all random variables will be described. We model the uncertainty in the market by using a filtered probability space \((\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})\) where the flow of information is modeled by an increasing family of sub sigma-fields of \(\mathcal{F}\), \(\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}\), satisfying all usual conditions. We also assume the existence of the probability
measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}, \mathcal{F})$. As it is postulated that financial risks and actuarial risks are independent under both measures, a complete configuration of the probability space can then be represented by $
abla = \Omega_1 \times \Omega_2$, $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$, $\mathbb{P}(\omega) = \mathbb{P}_1(\omega_1) \times \mathbb{P}_2(\omega_2)$, and $\mathbb{Q}(\omega) = \mathbb{Q}_1(\omega_1) \times \mathbb{Q}_2(\omega_2)$, for $\omega = (\omega_1, \omega_2), \omega_1 \in \Omega_1, \omega_2 \in \Omega_2$, where the first component refers to the financial risk and the second component refers to the insurance risk. A detailed study of the exact form of the above-mentioned filtrations associated with insurance and financial risks under the independence assumption can be found in Cox and Pedersen (2000).

The classical pricing formula for asset pricing mentions that under arbitrage-free assumption, there exists an equivalent martingale measure $\mathbb{Q}$ (in the sense of absolute continuity) for the reference measure $\mathbb{P}$ such that any price process $\{V(t) : 0 \leq t \leq T\}$ discounted at a risk-free rate is a martingale. Hence, the present value of a claim at time $t$ can be written as follows:

$$V^\mathbb{Q}(t) = \mathbb{E}_\mathbb{Q}[D(t, T)V(T)|\mathcal{F}_t]$$

(2.1)

where $D(t, T) = \exp\{\int_t^T r_s ds\}$ is called the discount factor in which $r_s$ is a risk-free interest rate. Furthermore, $V(T)$ specifies the claim’s value at time $T$, which is a random variable. It is well-known that a complete market leads to a unique price while an incomplete market results in an infinite number of prices, in which case $V^\mathbb{Q}(t)$ is supposed to be in the interval $(\inf_{\mathbb{Q} \in \mathcal{M}} V^\mathbb{Q}(t), \sup_{\mathbb{Q} \in \mathcal{M}} V^\mathbb{Q}(t))$, where $\mathcal{M}$ is the class of all possible equivalent measures under which the price process becomes a martingale. In the sequel, we assume that the claim’s payoff at maturity time $T$ represented by $V(T)$ only consists of insurance risk. As the discount factor $D(t, T)$ is defined as a function of interest rate that itself is a macroeconomics element, it can be seen as a part connected to the financial risk. Therefore, using the independence assumption, we can rewrite the relation (2.1) in this way:

$$V^\mathbb{Q}(t) = \mathbb{E}_\mathbb{Q}^1[D(t, T)|\mathcal{F}_t^1] \times \mathbb{E}_\mathbb{Q}^2[V(T)|\mathcal{F}_t^2]$$

(2.2)

where $\mathcal{F}_t^1$ and $\mathcal{F}_t^2$ are natural filtrations associated with financial and insurance risks, respectively.

3 Equivalent measure for financial risk

We suppose that the dynamics of interest rate process $r_t$ under measure $\mathbb{P}_1$ is governed by the Cox-Ingersoll-Ross (CIR) model (Cox et al. 1985), of which stochastic differential equation
(SDE) is given by

\[ dr^{P_1}_t = \theta(m - r_t)dt + \sigma\sqrt{r_t}dW^{P_1}_t, \quad (3.1) \]

where \( P_1 \) is defined on measurable space \( (\Omega_1, \mathcal{F}_1) \), \( W^{P_1}_t \) is a Brownian motion process, \( \theta > 0 \) is the mean-reverting force measurement (i.e., the speed of mean-reverting), \( m > 0 \) is a long-run interest rate mean, and \( \sigma > 0 \) is the volatility parameter for the interest rate. The Feller condition \( 2m\theta > \sigma^2 \) guarantees that \( r_t \) is almost surely strictly positive (Feller 1951). Note here that we may want to use interchangeably, for instance, \( P_1 \) as a superscript or with a dash for a process in order to stress that the given process is described under measure \( P_1 \).

To price the financial risk contained in relation (2.2), the dynamics of interest rate \( r_t \) under a risk-neutral measure \( Q_1 \), namely \( r^{Q_1}_t \), should be derived. The well-known Girsanov’s Theorem shows that the dynamics of \( r^{Q_1}_t \) is given by:

\[ dr^{Q_1}_t = \theta^*(m^* - r^{Q_1}_t)dt + \sigma\sqrt{r^{Q_1}_t}dW^{Q_1}_t \quad (3.2) \]

characterized by new parameters \( \theta^* \) and \( m^* \), which are defined by

\[ \theta^* = \theta + \lambda, \quad m^* = \frac{\theta m}{\theta + \lambda} \quad (3.3) \]

where \( \lambda \) is a constant which determines the market price of risk. In the CIR model, the market price of risk has the form \( \lambda_t = \frac{-\lambda\sqrt{\pi t}}{\sigma^2} \), which is a popular choice used as a kernel function when applying Girsanov’s Theorem.

### 4 Equivalent measure for insurance risk

In order to price the insurance risk contained in the payoff function \( V(T) \) presented in (2.2), we first need to identify the underlying risk process that captures the respective insurance risk. For this purpose, we start with typical definitions of a CAT bond’s payoff function.

---

2The market price of risk here represents the market expectation about the rate evolution - a positive value of \( \lambda \) is interpreted as the situation where a significant decrease of rates may happen, while a negative value of which shows a significant increase of the rates.
### 4.1 CAT bond’s structure

We consider a simple framework for the CAT bond whose conditional payment takes place only once during the length of the contract. This is opposed to a multiple framework where multiple conditional payments are made periodically up to and including the maturity time $T$. For more information about the latter type, readers can refer to, for example, Burnecki et al. (2019). We follow a stylized structure (i.e., the former type) provided by Ma and Ma (2013).

A zero-coupon CAT bond delivers a face value $Z$ at maturity time $T$ if the condition $L_T \leq D$ happens, where $D$ stands for a contractually defined threshold level and $L_T$ is a loss index reflecting losses from natural catastrophes over the time interval $[0, T]$, and a proportion $q \in (0, 1)$ of the face value is paid to the bondholder if the trigger event $L_T > D$ occurs, that is,

$$ P_{\text{CAT}}(T) = \begin{cases} Z & \text{if } L_T \leq D \\ qZ & \text{if } L_T > D \end{cases} \quad (4.1) $$

According to the foregoing structure selected for the CAT bond contract\(^3\), the payment is contingent on the behaviour of the index-linked catastrophe loss $L_T$, which is modeled by a counting process and a sequence of positive independent and identically random variables. The counting process defines the number of claims over time interval $[0, T]$, and random variables denote the losses resulting from natural disasters. Such a CAT bond contract whose trigger mechanism is characterized by the underlying loss index $L_T$ is called a non-indemnity-type CAT bond. In many applications, the loss index $L_T$ is chosen to be an aggregate loss process. This model-based loss index plays a fundamental role when calculating the insurance premium in actuarial science, and one can make different assumptions for modeling purposes. Moreover, from (4.1), it is obvious that $L_T$ is the main information to determine the price of a CAT bond contract. In subsequence sections, we discuss further the assumptions that can be made to model $L_T$.

### 4.2 Underlying risk process model

Throughout this subsection, the probability space under consideration is $(\Omega_2, \mathcal{F}_2, P_2)$. We denote by $X = \{X_n\}_{n \in \mathbb{N}}$ a sequence of $\mathbb{P}_2$-i.i.d positive real-valued random variables named claim size process, $N = \{N_t\}_{t \in \mathbb{R}_+}$ a counting process or claim number, $T = \{T_n\}_{n \in \mathbb{N}_0}$ a claim arrival time process, and $W = \{W_n\}_{n \in \mathbb{N}}$ a claim interarrival process. In fact, $W$ is said to be inter-arrival times between consecutively arriving events, i.e., $W_n = T_n - T_{n-1}$. The counting

\(^3\)Also known as index-based binary CAT bond contract
process $N$ is called a renewal counting process with parameter $\delta$ if the interarrival process $W$ is $P_2$-i.i.d with distribution $K^\delta$. In particular, when $\delta > 0$ and $K^\delta = \text{EXP}(\delta)$, namely Exponential distribution with parameter $\delta$, the counting process $N$ becomes a homogeneous Poisson process with rate $\delta$. In particular, when $\delta > 0$ and $K^\delta = \text{EXP}(\delta)$, namely Exponential distribution with parameter $\delta$, the counting process $N$ becomes a homogeneous Poisson process with intensity $\delta$ where $\delta$ can be interpreted as the rate of arrivals. So, in a homogeneous Poisson process, $W_n$’s are independent and exponentially distributed with constant parameter $\delta$. In the situation where parameter $\delta$ varies with time, i.e., it is a deterministic function of time, the counting process $N$ is a non-homogeneous Poisson process for which $W_n$’s are not independent and exponentially distributed anymore. In case $\delta$ itself is a random variable, $N$ turns to a mixed renewal counting process. This process can be seen as a special case of a stochastic renewal counting process where $\delta$ is itself a stochastic process (also called Cox process). In this paper, we restrict our work to the renewal counting process with fixed parameter $\delta$.

The underlying risk process $L = \{L_t\}_{t \in \mathbb{R}^+}$ is said to be the aggregate claims process induced by $(N, X)$, which is defined as $L_t = \sum_{n=1}^{N_t} X_n$ for any $t \geq 0$. Accordingly, if $N$ is a renewal counting process with parameter $\delta$, and independent of $X$, the aggregate claims process is called a $P_2$-compound renewal process (CRP for short) specified by $K^\delta$ and $P_X^1$ (henceforth written as $P_2$-CRP$(K^\delta, P_X^1)$). An example of the CRP is called the compound Poisson process (CPP for short) in which the counting process $N$ is supposed to be a Poisson process, i.e., $L_t$ is a $P_2$-CPP$(\delta, P_X^1)$.

### 4.3 Equivalent measure under renewal risk model

The fact that the underlying risk process $L_t$ follows a compound renewal process induced by $(N, X)$, implies that we need to change the distribution of $(N, X)$ as a result of a change of measures. The new obtained pricing measure is characterized by market prices of insurance risks associated with the claim number and claim size, which reflect the risk averseness of an economic agent in the market. Characterization of such an equivalent measure can be conducted based on the information that $L_t$ provides for us. Macheras and Tzaninis (2020) introduced a class of all equivalent measures such that a compound renewal process under reference measure remains a compound renewal process under its corresponding equivalent measure. More formally, let $\mathcal{F}_t^2 = \mathcal{F}_t^L$ be the natural filtration generated by random process $L_t$, in symbol we write $\mathcal{F}_t^L = \sigma(L_s, s \leq t)$ which means sigma-algebra generated by process $L_s$. Denote by $\Lambda^\rho(\delta)$ the distribution function of interarrival process $W$ with parameter $\rho(\delta)$ under measure $\mathbb{Q}_2$, which is distributed as Gamma and general inverse Gaussian, respectively.

\footnote{Other well-known examples are negative binomial process and general inverse Gaussian process, of which $K^\delta$ is distributed as Gamma and general inverse Gaussian, respectively.}

\footnote{For simplicity in notation, we use $\mathbb{P}^2_{X_1}$ to denote the distribution of $X_1$ under measure $\mathbb{P}_2$.}

\footnote{Note here that a $\mathbb{P}_2$-CPP$(\delta, \mathbb{P}^2_{X_1})$ is in fact a $\mathbb{P}_2$-CRP$(K^\delta = \text{EXP}(\delta), \mathbb{P}^2_{X_1})$.}
where $\rho$ is an arbitrary function. We already know that the distribution function of $W$ under measure $\mathbb{P}_2$ is denoted by $K^\delta$. It can be shown that the Radon-Nikodym derivative satisfying the property of preserving the structure of a renewal compound process under both measures is of the following form

$$
\frac{d\mathbb{Q}_2}{d\mathbb{P}_2} \bigg|_{\mathcal{F}_t^2} = \left[ \prod_{j=1}^{N_t} h^{-1}(\gamma(X_j)) \times \frac{d\mathbb{Q}_2}{d\mathbb{P}_2} (W_j) \right] \times \frac{1 - \Lambda^\rho(\delta)(t - T_{N_t})}{1 - K^\delta(t - T_{N_t})}
$$

(4.2)

where $h$ and $\gamma$ are real-valued Borel measurable mapping from $(0, \infty)$ to $\mathbb{R}$ such that $\mathbb{E}^\mathbb{P}_2[h^{-1}(\gamma(X_j))] = 1$, $\mathbb{E}^\mathbb{P}_2[X_1^l h^{-1}(\gamma(X_j))] < \infty$ (for $l = 1, 2$), and $\frac{d\mathbb{Q}_2}{d\mathbb{P}_2}$ is the Radon-Nikodym derivative of distribution $W_1$ under measure $\mathbb{Q}_2$ with respect to distribution $W_1$ under measure $\mathbb{P}_2$. In practice, we set $h = \ln$ (natural logarithm) and $\gamma = h(f)$ with $f$ being a Radon-Nikodym derivative of $\mathbb{Q}_2^{\delta X_1}$ with respect to $\mathbb{P}_2^{\delta X_1}$, in symbol we write $f = \frac{d\mathbb{Q}_2^{\delta X_1}}{d\mathbb{P}_2^{\delta X_1}}$ where $\mathbb{Q}_2^{\delta X_1}$ and $\mathbb{P}_2^{\delta X_1}$ are assumed to be equivalent measures. To make the current paper self-contained, we provide a rough proof of (4.2) that is originally inspired by Macheras and Tzaninis (2020), see Appendix C. In the following examples, we explain more on the derivation of the Radon-Nikodym derivative using relation (4.2).

**Example 4.3.1** Let $\mathbb{P}_2$ and $\mathbb{Q}_2$ be probability measures such that $\{L_t\}_{t \in \mathbb{R}_+}$ is a $\mathbb{P}_2$-CPP($\delta$, $\mathbb{P}_2^{\delta X_1}$) and $\mathbb{Q}_2$-CPP($\rho(\delta)$, $\mathbb{Q}_2^{\delta X_1}$). This implies that the counting process $N$ is a Poisson process, and hence the inter-arrival process $W$, which is independent of $N$, contains a sequence of independent and exponentially distributed random variables under both measures. Therefore, we can write

$$
\frac{d\mathbb{Q}_2}{d\mathbb{P}_2} \bigg|_{\mathcal{F}_t^2} = \left[ \prod_{j=1}^{N_t} e^{\gamma(X_j)} \times \frac{\rho(\delta)e^{-\rho(\delta)W_j}}{\delta e^{-\delta W_j}} \right] \times \frac{e^{-\rho(\delta)(t - T_{N_t})}}{e^{\delta(t - T_{N_t})}}
$$

(4.3)

where the last line is due to the fact that $T_{N_t} = \sum_{j=1}^{N_t} W_j$.

Relation (4.3) can be reformulated through the following notation which was introduced by Macheras and Tzaninis (2020): Define a real-valued Borel measurable function $\beta_\delta(x) = \gamma(x) + \alpha_\delta$ such that $\alpha_\delta = \ln \rho(\delta) + \ln \mathbb{E}^\mathbb{P}_2[W_1]$. Knowing the fact that $W_1$ is $\mathbb{P}_2$-EXP($\delta$) and putting
\[ \rho(\delta) = \frac{e^{\alpha \delta}}{\mathbb{E}^\beta[W_1]} \] which itself leads to \( \alpha_\delta = \ln\left(\frac{\rho(\delta)}{\delta}\right) \) as \( \mathbb{E}^{\mathbb{P}_2}[W_1] = 1/\delta \), relation (4.3) turns into

\[
\frac{d\mathbb{Q}_2}{d\mathbb{P}_2}\bigg|_{\mathcal{F}_t^\mathbb{P}_2} = \exp\left\{\sum_{j=1}^{N_t} \beta(X_j) - \delta t \mathbb{E}^{\mathbb{P}_2}[e^{\beta(X_1)}] - 1\right\}, \quad (4.4)
\]

Relation (4.4) is the main result that was proved by Delbaen and Haezendonck (1989). This is however not surprising since a compound Poisson process is a special case of a compound renewal process, and one can generate subclasses by means of the general framework proposed by Macheras and Tzaninis (2020). We now intend to find the distribution of claim number and claim size inducing aggregate process \( \{L_t\}_{t \in \mathbb{R}_+} \) under new measure \( \mathbb{Q}_2 \). First, \( \mathbb{E}^{\mathbb{P}_2}[\exp\{\gamma(X_1)\}] = 1 \) together with \( \beta_\delta(x) - \alpha_\delta = \gamma(x) \) yield

\[
\mathbb{E}^{\mathbb{P}_2}[\exp\{\beta(X_1)\}] = \exp\{\alpha_\delta\} = \frac{\rho(\delta)}{\delta} \quad (4.5)
\]

and so we have that

\[
\rho(\delta) = \delta \mathbb{E}^{\mathbb{P}_2}[\exp\{\beta(X_1)\}] \quad (4.6)
\]

Second, for all \( A \in \mathcal{B}(\mathbb{R}_+) \) where \( \mathcal{B}(\mathbb{R}_+) \) is defined to be the Borel sets on \( \mathbb{R}_+ \), we can write

\[
Q_{X_1}^2(A) = \mathbb{E}^{\mathbb{P}_2}\left[I_A \frac{d\mathbb{Q}_2}{d\mathbb{P}_2}\bigg|_{\mathcal{F}_t^\mathbb{P}_2}\right] = \mathbb{E}^{\mathbb{P}_2}\left[I_A e^{\gamma(X_1)}\right] = \mathbb{E}^{\mathbb{P}_2}\left[I_A \frac{e^{\beta(X_1)}}{\mathbb{E}^{\mathbb{P}_2}[e^{\beta(X_1)}]}\right] = \int_A \frac{e^{\beta(x_1)}}{\mathbb{E}^{\mathbb{P}_2}[e^{\beta(X_1)}]} d\mathbb{P}_2^{\mathbb{P}_2}\bigg|_{\mathcal{F}_t^\mathbb{P}_2} X_1 \quad (4.7)
\]

From (4.6) and (4.7), we conclude that process \( L_t \) is a \( \mathbb{Q}_2\text{-CPP}\left(\delta \mathbb{E}^{\mathbb{P}_2}[e^{\beta(X_1)}], \frac{e^{\beta(x_1)}}{\mathbb{P}_2[e^{\beta(X_1)}]} \right) \).

As proposed by Muermann (2003), an alternative representation of relations (4.4) and (4.7) can be expressed in this way: Define \( \kappa = \mathbb{E}^{\mathbb{P}_2}[e^{\beta(X_1)}] \) and \( \nu(x_1) = \frac{e^{\beta(x_1)}}{\mathbb{P}_2[e^{\beta(X_1)}]} \). Then, we have that

\[
\frac{d\mathbb{Q}_2}{d\mathbb{P}_2}\bigg|_{\mathcal{F}_t^\mathbb{P}_2} = \exp\left\{\sum_{j=1}^{N_t} \ln(\kappa \nu(X_j)) + \delta t (1 - \kappa)\right\}, \quad (4.8)
\]

where \( \kappa \) and \( \nu(.) \) can be interpreted as the market prices of claim number risk and claim size risk, respectively. Hence, the characterization of Radon-Nikodym derivative (4.8) yields that under measure \( \mathbb{P}_2 \), \( L_t \) is a \( \mathbb{P}_2\text{-CRP}(\delta, \mathbb{P}_2^{\mathbb{X}_1}) \) while under measure \( \mathbb{Q}_2 \) is a \( \mathbb{Q}_2\text{-CPP}(\delta \kappa, \nu(x_1) \mathbb{P}_2^{\mathbb{X}_1}) \).

**Example 4.3.2** Let \( \mathbb{P}_2 \) and \( \mathbb{Q}_2 \) be probability measures such that \( \{L_t\}_{t \in \mathbb{R}_+} \) is a \( \mathbb{P}_2\text{-CRP}(K^\delta, \mathbb{P}_2^{\mathbb{X}_1}) \) and \( \mathbb{Q}_2\text{-CPP}(\rho(\delta), Q_{X_1}^2) \) with \( K^\delta = Ga(\delta)^7 \), where \( \delta = (\eta_1, \eta_2) \) and \( \eta_1 \) is assumed to be a

\[
f(x) = \frac{\eta_1^{\eta_1}}{\Gamma(\eta_1)} x^{\eta_1-1} e^{-\eta_2 x} \quad (x \geq 0)
\]

---

Electronic copy available at: https://ssrn.com/abstract=4599125
positive integer. Relation (4.2) gives us that

\[
\frac{dQ_2}{dP_2} \bigg|_{x^2} = e^{\sum_{j=1}^{N_t} \gamma(X_j)} \times \left( \frac{\rho(\delta) \Gamma(\eta_1)}{\eta_2^{\eta_1}} \right)^{\frac{N_t}{\eta_2^{\eta_1}}} \times \left( \prod_{j=1}^{N_t} \frac{1}{W_j^{\eta_1-1}} \right) \times \frac{e^{-t(\rho(\delta)-\eta_2)}}{\sum_{i=0}^{\eta_1-1} (\eta_2(t-T_{N_i}))^i}.
\]  

(4.9)

Analogously to the previous example, one can show that

\[
\frac{dQ_2}{dP_2} \bigg|_{x^2} = \exp \left\{ \sum_{j=1}^{N_t} \ln(\kappa \nu(X_j)) + N_t \ln \left( \frac{\eta_2 \Gamma(\eta_1)}{\eta_2^{\eta_1}} \right) - (\eta_1 - 1) \sum_{j=1}^{N_t} \ln(W_j) + \frac{t \eta_2}{\eta_1} (\eta_1 - \kappa) - \ln \left( \sum_{i=0}^{\eta_1-1} (\eta_2(t-T_{N_i}))^i \right) \right\}
\]

(4.10)

where \(\kappa\) and \(\nu(\cdot)\) are as before. In this example, \(L_t\) is \(Q_2\)-CPP\(\left( \frac{\eta_2}{\eta_1}, \kappa, \nu(x_1)P_2^{X_1} \right)\).

Example 4.3.3 Let \(P_2\) and \(Q_2\) be probability measures such that \(\{L_t\}_{t \in \mathbb{R}_+}\) is \(P_2\)-CRP(\(K^\delta\), \(P_2^{X_1}\)) and \(Q_2\)-CPP(\(\rho(\delta), Q_2^X\)) with \(K^\delta = WE(\delta)\), where \(\delta = (\eta_3, \eta_4).\) Then, we have that

\[
\frac{dQ_2}{dP_2} \bigg|_{x^2} = e^{\sum_{j=1}^{N_t} \gamma(X_j)} \times \left( \frac{\eta_4^{\eta_3} \rho(\delta)}{\eta_3} \right)^{\frac{N_t}{\eta_4^{\eta_3}}} \times \left( \prod_{j=1}^{N_t} \frac{1}{W_j^{\eta_3-1}} \right) \times \frac{e^{-\rho(\delta)W_j + \frac{t \eta_3}{\eta_4} W_j^{\eta_3}}}{e^{\eta_3^2}}
\]

(4.11)

Similarly, it can be shown that

\[
\frac{dQ_2}{dP_2} \bigg|_{x^2} = \exp \left\{ \sum_{j=1}^{N_t} \ln(\kappa \nu(X_j)) + N_t \ln \left( \frac{\eta_4^{\eta_3} \rho(\delta)}{\eta_4^{\eta_3}} \right) - (\eta_3 - 1) \sum_{j=1}^{N_t} \ln(W_j) + \frac{1}{\eta_4^{\eta_3}} \sum_{j=1}^{N_t} W_j^{\eta_3} - \frac{t \kappa}{\eta_4^{\eta_3}} + \frac{(t-T_{N_i})^{\eta_3}}{\eta_4^{\eta_3}} \right\}
\]

(4.12)

for which \(L_t\) is \(Q_2\)-CPP\(\left( \eta_4^{\eta_3} \rho(\delta), \kappa, \nu(x_1)P_2^{X_1} \right)\).

Example 4.3.4 Let \(P_2\) and \(Q_2\) be probability measures such that \(\{L_t\}_{t \in \mathbb{R}_+}\) is \(P_2\)-CRP(\(K^\delta\)).

\(WE(\delta)\) represents the distribution function of a Weibull distribution with the density given by

\[
f(x) = \frac{\eta_3}{\eta_4^{\eta_3}} x^{\eta_3-1} e^{-\left(\frac{x}{\eta_4}\right)^{\eta_3}} \quad (x \geq 0)
\]
and \(Q_2\)-CPP(\(\rho(\delta), Q_2^{\rho(\delta)}\)) with \(\mathbf{K}^{\delta} = M^{\text{EXP}}(\delta)^9\), where \(\delta = (\phi, \eta_5, \eta_6)\). Then, we have that

\[
\frac{dQ_2}{dP_2} \bigg|_{\mathcal{F}_t^2} = \exp \left\{ \sum_{j=1}^{N_t} \gamma(X_j) \right\} \times \left( \prod_{j=1}^{N_t} \frac{\rho(\delta)e^{-\rho(\delta)W_j}}{\phi \eta_5 e^{-\eta_5 W_j} + (1 - \phi) \eta_6 e^{-\eta_6 W_j}} \right) \times \frac{e^{\rho(\delta)(t-T_{N_t})}}{\Delta_1} \tag{4.13}
\]

where \(\Delta_1 = 1 - [\phi(1 - e^{-\eta_5(t-T_{N_t})}) + (1 - \phi)(1 - e^{-\eta_6(t-T_{N_t})})]\). Following the same procedure discussed earlier, we have that

\[
\frac{dQ_2}{dP_2} \bigg|_{\mathcal{F}_t^2} = \exp \left\{ \sum_{j=1}^{N_t} \ln(\nu(X_j)) - N_t \ln \left( \frac{\phi}{\eta_5} + \frac{(1 - \phi)}{\eta_6} \right) - \frac{\kappa t}{\phi \eta_5 + (1 - \phi) \eta_6} - \ln(\Delta_1) \right\} \tag{4.14}
\]

for which \(L_t\) is \(Q_2\)-CPP\(\left(\frac{\phi}{\eta_5} + \frac{(1 - \phi)}{\eta_6}\right)^{-1} \kappa, \nu(x)P_{X_1}^2\).

In the examples outlined above, our focus was on the situation where \(L_t\) follows a CPP under measure \(Q_2\). In this context, our examples showcased how the well-defined symbols introduced by Macheras and Tzaninis (2020) result in the Radon-Nikodym derivatives that are in terms of the parameters inherent in the distributions postulated within the framework of the real-world measure. This kind of representation purely based on the physical parameters enables us to fully calibrate the pricing formula using historical catastrophe data. However, finding such a representation appears to be more complicated when we assume a CRP for \(L_t\) under measure \(Q_2\). One reason is that the definition of \(\alpha_\delta\) can be only used in the case of one-dimensional space for parameter \(\rho(\delta)\). In the next example, we will show this hurdle.

**Example 4.3.5** Let \(P_2\) and \(Q_2\) be probability measures such that \(\{L_t\}_{t \in \mathbb{R}_+}\) is a \(P_2\)-CRP(\(\mathbf{K}^{\delta}\), \(P_{X_1}^2\)) and \(Q_2\)-CRP(\(\Lambda^{\rho(\delta)}, Q_2^{\rho(\delta)}\)) with \(\mathbf{K}^{\delta} = \mathbf{G}a(\delta)\) and \(\Lambda^{\rho(\delta)} = \mathbf{G}a(\rho(\delta))\), where \(\delta = (\xi_1, \xi_2)\) and \(\rho(\delta) = (\epsilon_1, \epsilon_2)\). Applying relation (4.2), we get that

\[
\frac{dQ_2}{dP_2} \bigg|_{\mathcal{F}_t^2} = \exp \left\{ \sum_{j=1}^{N_t} \gamma(X_j) \right\} \times \left( \prod_{j=1}^{N_t} \frac{\epsilon_1^j \Gamma(\xi_1)}{\xi_1^j \Gamma(\xi_1)} W_j^{\epsilon_1 j - 1} e^{-\epsilon_2 W_j} \right) \times \frac{\sum_{i=0}^{\epsilon_1 - 1} (\epsilon_2(t-T_{N_t}))^i}{\sum_{i=0}^{\epsilon_2 - 1} (\xi_2(t-T_{N_t}))^i} \times \frac{\sum_{i=0}^{\epsilon_1 - 1} (\epsilon_2(t-T_{N_t}))^i}{\sum_{i=0}^{\epsilon_2 - 1} (\xi_2(t-T_{N_t}))^i} \times \frac{\sum_{i=0}^{\epsilon_1 - 1} (\epsilon_2(t-T_{N_t}))^i}{\sum_{i=0}^{\epsilon_2 - 1} (\xi_2(t-T_{N_t}))^i} \tag{4.15}
\]

\(^9\text{MIX-EXP}(\delta)\) represents the distribution function of a mixture of two exponential distributions with the density given by

\[f(x) = \phi \eta_5 e^{-\eta_5 x} + (1 - \phi) \eta_6 e^{-\eta_6 x}\quad (x \geq 0),\]

with \(\phi \in (0, 1)\).
where $\xi_1$ and $\epsilon_1$ are said to be positive integers.

In this example, we observe that $\rho(\delta)$ is of two-dimension, which does not allow us to find an alternative representation for the Radon-Nikodym derivative in terms of only real-world parameters. Another interesting case is when we assume a mixture of exponential for the inter-arrival distribution under both measures.

**Example 4.3.6** Let $L_t$ be a $\mathbb{P}_2$-CRP($K^\delta$, $\mathbb{P}_2^X_1$) and $\mathbb{Q}_2$-CRP($\Lambda^{\rho(\delta)}$, $\mathbb{Q}_2^X_1$) with $K^\delta = \textit{MIX-EXP}(\delta)$ and $\Lambda^{\rho(\delta)} = \textit{MIX-EXP}(\rho(\delta))$, where $\delta = (\phi, \eta_5, \eta_6)$ and $\rho(\delta) = (\phi^*, \eta_7, \eta_8)$.

## 5 CAT bond price

Up to this point, we have dealt with two general scenarios: Scenario I presumes $L_t$ to be a CRP under measure $\mathbb{P}_2$ while assuming a CPP under measure $\mathbb{Q}_2$. Scenario II where $L_t$ follows a CRP under both measures. The intricacies posed by Scenario II necessitate a distinct discussion, given the challenges it entails. As a result, we opt to address Scenario II separately from the treatment of Scenario I.

### 5.1 Scenario I

According to (2.2), the price of a zero-coupon CAT bond with maturity time $T$, pay-off function (4.1), and model assumptions stated in examples 4.3.1, 4.3.2, 4.3.3, and 4.3.4 is given by

$$V_1^Q(t) = \mathbb{E}^Q_1[D(t, T)|\mathcal{F}^1_t] \times \mathbb{E}^Q_2[P_{\textit{CAT}}(T)|\mathcal{F}^2_t]$$

$$= P(t, T, r(t), \theta^*, m^*, \sigma)[Z - (Z - qZ)\mathbb{Q}_2(L_T > D)] \quad (5.1)$$

where the first term in (5.1) denotes the zero-coupon bond price expressed by

$$P(t, T, r(t), \theta^*, m^*, \sigma) = A(t, T)e^{-B(t, T)r(t)} \quad (5.2)$$

with $A(t, T)$ and $B(t, T)$ being defined as below:

$$A(t, T) = \left[\frac{2\vartheta e^{(\vartheta + \theta^*)(T-t)/2}}{(\vartheta + \theta^*)(e^{\theta(T-t)} - 1) + 2\vartheta}\right]^{\frac{2m^*\theta^*}{\sigma^2}},$$

$$B(t, T) = \frac{2(e^{\theta(T-t)} - 1)}{(\vartheta + \theta^*)(e^{\theta(T-t)} - 1) + 2\vartheta},$$

$$\vartheta = \sqrt{(\theta^*)^2 + 2\sigma^2}.$$
(see, e.g., Brigo and Mercurio (2007)), and that

\[ Q_2(L_T > D) = e^{-\delta \kappa T} \sum_{k=0}^{\infty} \frac{(\delta \kappa T)^k}{k!} Q_2 \left( \sum_{i=1}^{k} X_i > D \right) \]  

(5.3)

Determining \( \kappa \) and \( \nu(x) \) is based on the specification of function \( \beta(\cdot) \), which itself characterizes the distribution of claim frequency and claim severity under the risk-neutral measure. Following Delbaen and Haezendonck (1989), different choices are possible:

- \( \beta_1(x) = \alpha \), where \( \alpha \) is a constant (known as the expected value principle).
- \( \beta_2(x) = \ln(a + bx) \) with \( b > 0 \) and \( a = 1 - bE^F[X_1] > 0 \) (known as the variance principle).
- \( \beta_3(x) = cx - \ln(E^F[e^{cX_1}]) \) with \( c > 0 \) (known as the Esscher principle).

For example, if we choose \( \beta_1(x) = \alpha \), then we get \( \kappa = e^\alpha \) and \( \nu(x) = 1 \), which means that the distribution of claim severity remains unchanged. In contrast, the intensity parameter of the claim frequency distribution under the risk-neutral measure is scaled by \( e^\alpha \).

5.2 Scenario II

We now give our attention to examples 4.3.5 and 4.3.6 where \( L_t \) is assumed to be CRP under both measures. For such a situation, as we discussed earlier, it is not possible to acquire a representation based on \( \kappa \) and \( \nu(\cdot) \) and the CAT bond price will also rely on risk-neutral parameters contained in inter-arrival time distribution, which can not be estimated easily from information available in the capital market. More precisely, the probability of aggregate claim process \( L_t \) can be written as follows under the assumptions of examples 4.3.5 and 4.3.6, respectively; the proofs can be found in the Appendix:

\[ Q_2(L_T > D) = \sum_{k=0}^{\infty} Q_2(N_T = k) Q_2 \left( \sum_{i=1}^{k} X_i > D \right) \]

\[ = \sum_{k=0}^{\infty} \sum_{s=\kappa_1}^{\kappa_1+\epsilon_1-1} \frac{e^{-\epsilon_2 T(\epsilon_2 T)^s}}{s!} Q_2 \left( \sum_{i=1}^{k} X_i > D \right) \]  

(5.4)
and

\[ Q_2(L_T > D) = \sum_{n=0}^{\infty} \left\{ \sum_{s=0}^{n} \binom{n}{s} \pi_1^n \pi_2^{n-s} \text{CGDC}(s, \lambda_1, n-s, \lambda_2; T) \right\} - \sum_{s=0}^{n+1} \left( \frac{n+1}{s} \right) \pi_1^n \pi_2^{n+1-s} \text{CGDC}(s, \lambda_1, (n+1)-s, \lambda_2; T) \}

where the notation CGDC denotes the cumulative distribution function of Gamma Distribution Convolution (GDC), \( \pi_1 = \phi^*, \pi_2 = 1 - \phi^* \), \( \lambda_1 = \eta_7 \), and \( \lambda_2 = \eta_8 \). In this paper, we find the estimation of inter-arrival parameters under the risk-neutral measure by using the Bayesian framework.

5.2.1 Bayesian framework for risk-neutral parameters

The foundation of the Bayesian approach considers the specification of a sampling model from which our data is drawn and a marginal distribution of unknown parameters contained in the target distribution, called a prior distribution. Using Bayes’ rule, the information about unknown parameters is updated through the conditional model of observed data and the prior model of unknown parameters to achieve a new distribution called a posteriori. This transition from a prior distribution to a posterior distribution can be seen as moving from the physical measure to a risk-neutral measure. The inverse probability of unknown parameters under the physical measure (i.e., posterior distribution) can be utilized to find a reasonable estimation for corresponding risk-neutral parameters. The Bayesian estimator aims to reduce the expected loss post-analysis, determined using metrics such as squared error, absolute error, weighted squared error, or Kullback-Leibler divergence error, to name a few.

In our Bayesian analysis, the choice of loss function is influenced by the perspective that in an incomplete market, where there isn’t a single set of equivalent measures, the optimal measure is chosen to reduce the Kullback-Leibler divergence (KLD). By its standard definition, KLD defines the relative distance between two absolutely continuous measures \( \mathbb{P} \) and \( \mathbb{Q} \) in the entropy sense. The measure derived in this way is called the minimal entropy martingale measure, for more details see, e.g., Dhaene et al (2015). We consider the following Bayesian configuration.

Let \( f(x) \) be a density function for a continuous random variable \( X \), characterized by the parameter \( \Theta \). From the Bayesian perspective, a necessary form for a proper loss function under KLD
is then given by

$$KL(\Theta \parallel \hat{\Theta}) = KL(f(x; \Theta) \parallel f(x; \hat{\Theta})) = \int_{\mathcal{A}} \log \frac{f(x; \Theta)}{f(x; \hat{\Theta})} f(x; \Theta) dx, \quad (5.5)$$

where we call $KL(\Theta \parallel \hat{\Theta})$ as the Kullback error loss function (KEL) corresponding to the density function $f$. The Bayesian estimation is then defined as the estimator $\Theta^B$ that minimizes KLD between the actual parameter of interest $\Theta$ and its possible estimation $\hat{\Theta}$, that is,

$$\Theta^B = \arg \min_{\hat{\Theta}} \mathbb{E}_\Theta[KL(\Theta \parallel \hat{\Theta}) | x] \quad (5.6)$$

We revisit the scenario presented in example 4.3.5. Here, under the physical measure, the inter-arrival time distribution, represented by $W$, follows a Gamma distribution characterized by shape and rate parameters $\xi_1$ and $\xi_2$. Conversely, under the risk-neutral measure, it adheres to a Gamma distribution defined by shape and rate parameters $\epsilon_1$ and $\epsilon_2$. Let $W_m = (W_1, W_2, \cdots, W_m)$ be a complete sample from $W$. We specify the likelihood and prior of the model as follows:

$$W_m|\xi_1, \xi_2 \sim \text{Ga}(\xi_1, \xi_2), \quad \xi_1 > 0, \quad \xi_2 > 0,$$

$$\left(\xi_1, \xi_2\right) \sim \pi(\delta)$$

where

$$f(w_m|\xi_1, \xi_2) = \frac{\xi_2^{m\xi_2}}{\Gamma(m)(\xi_1)} \left(\prod_{i=1}^{m} w_i^{\xi_1-1}\right) \exp\left\{ -\xi_2 \sum_{i=1}^{m} w_i \right\} \quad (5.7)$$

$$\pi(\delta) \propto \frac{\xi_2}{\Gamma(\xi_1)} \exp\left\{ -\frac{\psi(\xi_1)}{\Gamma(\xi_1)} - \xi_1 \right\} \quad (5.8)$$

with $\psi(\xi_1) = \frac{\partial}{\partial \xi_1} \log \Gamma(\xi_1) = \frac{\Gamma'(\xi_1)}{\Gamma(\xi_1)}$ being the digamma function. The selected non-informative prior distribution above is derived by the justified maximal data information prior (JMDIP) which maximizes the prior average information in the data density minus the information in the prior density. A simulation study conducted by Moala et al. (2013) has shown that for Gamma distribution JMDIP outperforms in a class of non-informative priors such as Jeffrey’s Prior, Reference Prior, and Tibishirani’s Prior. Assuming the above likelihood and prior distribution,
the joint posterior distribution for parameters $\xi_1$ and $\xi_2$ is given by,

$$f(\xi_1, \xi_2|w_m) \propto f(w_m|\xi_1, \xi_2) \times \pi(\delta)$$

$$\propto \frac{\xi_2^{m\xi_1}}{\Gamma^m(\xi_1)} \left( \prod_{i=1}^{m} w_i^{\xi_1-1} \right) \exp \left\{ -\xi_2 \sum_{i=1}^{m} w_i \right\} \times \frac{\xi_1}{\Gamma(\xi_1)} \exp \left\{ (\xi_1 - 1) \frac{\psi(\xi_1)}{\Gamma(\xi_1)} - \xi_1 \right\}$$

(5.9)

Using relation (5.5), the KLD corresponding to $\text{Ga}(\xi_1, \xi_2)$ with density $f(x; \xi_1, \xi_2)$ is given by:

$$\text{KL}(\Theta \parallel \hat{\Theta}) = \int_0^\infty \log \frac{f(x; \xi_1, \xi_2)}{f(x; \hat{\xi}_1, \hat{\xi}_2)} f(x; \xi_1, \xi_2) dx,$$

$$= \log \frac{\xi_1}{\Gamma(\xi_1)} - \log \frac{\hat{\xi}_1}{\Gamma(\hat{\xi}_1)} + (\xi_1 - \xi_2) \frac{\xi_1 - \xi_1(\psi(\xi_1) - \log \xi_2)}{\xi_2}$$

(5.10)

According to relation (5.6), the Bayes estimate of $\xi_1$ and $\xi_2$ can be found by differentiating the following equation with respect to $\hat{\xi}_1$ and $\hat{\xi}_2$ and setting equal to zero,

$$\int_0^\infty \int_0^\infty \left[ \log \frac{\xi_1}{\Gamma(\xi_1)} - \log \frac{\hat{\xi}_1}{\Gamma(\hat{\xi}_1)} + (\xi_1 - \xi_2) \frac{\xi_1 - \xi_1(\psi(\xi_1) - \log \xi_2)}{\xi_2} \right] f(\xi_1, \xi_2|w_m) d\xi_1 d\xi_2,$$

(5.11)

which results in

$$\hat{\xi}_2 = \frac{\xi_1}{\mathbb{E}_p[\xi_1|\xi_2]_w}, \quad \psi(\hat{\xi}_1) = \mathbb{E}_p[\psi(\xi_1) - \log \xi_2|w] + \log \hat{\xi}_2.$$  

(5.12)

Applying the following difference equation,

$$\psi(x + h) - \psi(x) = \sum_{i=0}^{h-1} \frac{1}{x + i},$$

(5.13)

with $h = 1$ follows that

$$\hat{\xi}_2 = \frac{\xi_1}{\mathbb{E}_p[\xi_1|\xi_2]_w}, \quad \hat{\xi}_1 = \frac{1}{\mathbb{E}_p[\psi(\xi_1 + 1) - \psi(\xi_1)|w]}$$

(5.14)

To generate samples from the posterior distribution, one can employ the Markov Chain Monte Carlo (MCMC) simulation following the Gibbs sampling method. However, given the unavail-
ability of the precise analytical forms of both the joint and marginal posterior distributions, it becomes more practical to utilize the Metropolis-Hastings (MH) algorithm, as recommended by Moala et al. (2013). To generate samples from $\xi_1$ and $\xi_2$, we run Algorithm 1 where $f(\xi_1^{(t)}, \xi_2^{(t-1)}|w)$ is given by (5.9), and $Ga(\xi_1^{(t-1)}|c)$ and $Ga(\xi_2^{(t-1)}|d)$ are proposal distribution of which hype-parameters $c$ and $d$ are selected such that a good mixing of the chains and the convergence of the MCMC samples of parameters are obtained. The generated sample hereby is applied to find Bayes estimators of $\xi_1$ and $\xi_2$, denoted by $\xi_1^B$ and $\xi_2^B$, through (5.14). These Bayesian estimations of real-world parameters are then regarded as estimations of their corresponding risk-neutral parameters, that is, we set $\hat{\epsilon}_1 = \xi_1^B$ and $\hat{\epsilon}_2 = \xi_2^B$.

**Algorithm 1** Metropolis-Hasting sampling for Gamma distribution

**Initialize:** $\xi_1^{(0)}$ and $\xi_2^{(0)}$

**for** $t = 1, 2, \cdots$ **do**

Generate new value $\xi_1^{(t)}$ from $Ga(\xi_1^{(t-1)}|c)$, and accept it with the following probability known as the MH ratio:

$$u(\xi_1^{(t-1)}, \xi_1^{(t)}) = \min \left\{ 1, \frac{Ga(\xi_1^{(t-1)}|c) f(\xi_1^{(t)}, \xi_2^{(t-1)}|w)}{Ga(\xi_1^{(t)}|c) f(\xi_1^{(t-1)}, \xi_2^{(t)}|w)} \right\}$$

Generate new value $\xi_2^{(t)}$ from $Ga(\xi_2^{(t-1)}|d)$, and accept it with the following probability:

$$u(\xi_2^{(t-1)}, \xi_2^{(t)}) = \min \left\{ 1, \frac{Ga(\xi_2^{(t-1)}|d) f(\xi_1^{(t)}, \xi_2^{(t)}|w)}{Ga(\xi_2^{(t)}|d) f(\xi_1^{(t-1)}, \xi_2^{(t-1)}|w)} \right\}$$

**end for**

Let us consider the problem of example 4.3.6 in which the inter-arrival time distribution $W$ is a mixture of two exponential distributions characterized by parameters $(\phi, \eta_5, \eta_6)$ and $(\phi^*, \eta_7, \eta_8)$ under physical and risk-neutral measures respectively. For this scenario, we utilize the EM algorithm to find the corresponding estimate of $(\phi^*, \eta_7, \eta_8)$, for more details see Appendix D.
6 Numerical illustration

6.1 Parameter calibration of the CIR model

To calibrate the parameters of the CIR model (3.1), we take the daily historical yields on the
3-month US treasury bills from January 2, 1990, to May 25, 2022\(^\text{10}\) as depicted in Figure 1.

![Historical data on interest rates observed from January 2, 1990, to May 25, 2022.](image)

Figure 1: Historical data on interest rates observed from January 2, 1990, to May 25, 2022.

In this section, we use the maximum likelihood estimation (MLE) method for parameters of the
CIR model (3.1), of which the transition probability density is given by

\[ f(r_t | r_u) = c^* \chi^2_{(d_1, d_2)} (c^* r_t), \quad t > u, \]  

(6.1)

where \( c^* = \frac{4\theta}{\sigma^2 (1 - e^{-\theta(t-u)})} \) and \( \chi^2_{(d_1, d_2)} (.) \) denotes the probability density of a non-central chi-

square distribution with the degree of freedom \( d_1 = \frac{4\theta m}{\sigma^2} \) and the non-central parameter \( d_2 = \)

\(^{10}\)Note here that it is not essential to use the same time period for the catastrophe loss observations, as we already
assumed that insurance risk and financial risks are independent.
\[
\frac{4\theta e^{-\theta(t-u)}}{\sigma^2(1-e^{-\theta(t-u)})} r_u \text{ (see, e.g., Fergusson 2019). If the data } r^* = \{r_1, r_2, \ldots , r_n\} \text{ is given according to the historical data, then the log-likelihood function can be written as: }
\]

\[
l(\theta, m, \sigma; r^*) = \sum_{i=1}^{n} ln(c^*) + \sum_{i=1}^{n} ln(p_{\chi^2(d_1, d_2)}(c^* r_i | r_{i-1})) \quad (6.2)
\]

For the purpose of finding the MLE of (6.2), a numerical optimization technique can be applied. For that, initial values for parameters are essential to be specified to start the iteration process embedded in the algorithm. In this paper, the initial values are obtained using the ordinary least square estimation (OLSE) method. The main idea to achieve approximate estimates of parameters contained in the CIR model on the basis of OLSE method is to find a regression version that can describe a discretized version of the CIR model derived by the Euler discretization technique. According to the Euler scheme, the SDE associated with the CIR model (3.1) can be represented as follows:

\[
r_{t_{i+1}} - r_{t_i} = \theta (m - r_{t_i}) \Delta t_i + \sigma \sqrt{r_{t_i}} \Delta W_i \quad (6.3)
\]

where \( \Delta t_i = (t_{i+1} - t_i)/250 \) \(^{11} \) and \( \Delta W_i = W_{t_{i+1}} - W_{t_i}, \) for \( i = 0, 1, 2, \ldots , n - 1. \) After performing some simple algebraic manipulations, one can realize that the matrix representation of the regression model corresponding to the discretized version (6.3) is of the form:

\[
\begin{bmatrix}
    y_1 \\
    y_2 \\
    \vdots \\
    y_{n-1}
\end{bmatrix}
= \begin{bmatrix}
    \sqrt{\frac{\Delta t_1}{r_1}} & \sqrt{\Delta t_1}r_1 \\
    \sqrt{\frac{\Delta t_2}{r_2}} & \sqrt{\Delta t_2}r_2 \\
    \vdots & \vdots \\
    \sqrt{\frac{\Delta t_{n-1}}{r_{n-1}}} & \sqrt{\Delta t_{n-1}}r_{n-1}
\end{bmatrix}
\begin{bmatrix}
    \beta_1 \\
    \beta_2 \\
    \vdots \\
    \beta_{n-1}
\end{bmatrix}
+ \begin{bmatrix}
    \sigma N_1(0, 1) \\
    \sigma N_2(0, 1) \\
    \vdots \\
    \sigma N_{n-1}(0, 1)
\end{bmatrix}
\]

where \( y_i = \frac{r_{t_{i+1}} - r_{t_i}}{\sqrt{\Delta t_{t_i}}} \), \( \beta_1 = \theta m, \beta_2 = -\theta, \) and \( N_i(0, 1) \) are independent random variables, each having a standard normal distribution. Then, the OLSE of \( \beta \) and \( \sigma \) become

\[
\hat{\beta}_{OLS} = (Z^T Z)^{-1} Z^T Y, \quad \hat{\sigma}^2 = \frac{1}{n-2} ||Y - Z \hat{\beta}_{OLS}||^2 \quad (6.5)
\]

where \( ||.|| \) denotes the Euclidean distance. We use these estimates as initial values for a numerical optimization of the likelihood function (6.2). Table 1 represents the final results for the OLS and MLE of the CIR parameters in (3.1).

\(^{11}\)Since a daily base observations are used for the interest rate, the time step of the Euler scheme is computed as the time difference between consecutive points divided by the number of working days per year, which as a
<table>
<thead>
<tr>
<th>Method (Under physical measure)</th>
<th>$\hat{\theta}$</th>
<th>$\hat{m}$</th>
<th>$\hat{\sigma}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLSE (Initial values)</td>
<td>0.277</td>
<td>0.018</td>
<td>0.077</td>
</tr>
<tr>
<td>MLE (Optimal values)</td>
<td>0.254</td>
<td>0.011</td>
<td>0.074</td>
</tr>
</tbody>
</table>

Table 1: Estimated parameters of the CIR model under the physical measure.

The risk-neutralized parameters represented in (3.3) can be found by knowing the estimated physical parameters as well as the constant parameter $\lambda$ which determines the market price of interest rate risk. Following the same idea in Ahmad and Wilmott (2006), we find an estimation for parameter $\lambda$ empirically. We summarize our steps as follows:

For simplicity in notation, suppose that $P(t, T, r_t)$ denotes the zero-coupon bond price at time $t$ with maturity time $T$. In the first step, we consider the partial differential equation (PDE) associated with the CIR model, which is given by

$$\frac{\partial P}{\partial t}(t, T, r_t) + \left[\theta(m - r_t) + \lambda r_t\right]\frac{\partial P}{\partial r_t}(t, T, r_t) + \frac{1}{2}\sigma^2 r_t^2 \frac{\partial^2 P}{\partial r_t^2}(t, T, r_t) - r_t P(t, T, r_t) = 0 \quad (6.6)$$

In the second step, the Taylor series expansion of the zero-coupon bond price around $t = T$ with final condition $P(T, T, r_T) = 1$ is derived by the following polynomial representation:

$$P(t, T, r_t) = \sum_{j=0}^{\infty} c_j(r_t)(T - t)^j \quad (6.7)$$

Substituting (6.7) into (6.6) gives the following recursive relation for $c_j(r_t)$:

$$c_{j+1}(r_t) = \frac{1}{j + 1} \left\{ [\theta(m - r_t) - \lambda r_t] c_j'(r_t) + \frac{1}{2}\sigma^2 r_t c_j''(r_t) - r_t c_j(r_t) \right\}, \quad j = 0, 1, \cdots \quad (6.8)$$

where $c_j'(r_t)$ and $c_j''(r_t)$ are the first and second derivatives of function $c_j(r_t)$ with respect to $r_t$. Using (6.8), an approximation for the zero-coupon bond price is given by

$$P(t, T, r_t) \approx 1 - r_t(T - t) + \frac{1}{2} \left\{ (\lambda + \theta)r_t - \theta m - r_t^2 \right\}(T - t)^2 + \cdots, \quad \text{as} \quad t \rightarrow T \quad (6.9)$$

From this we have

$$-\frac{\log(P(t, T, r_t))}{(T - t)} \approx -r_t + \frac{1}{2} \left\{ \theta m - (\lambda + \theta)r_t \right\}(T - t) + \cdots, \quad \text{as} \quad t \rightarrow T \quad (6.10)$$

standard convention is said to be 250 days.
Relation (6.10) shows that the slope of the yield curve at the short end is equal to \( \frac{1}{2} \left\{ \theta m - (\lambda + \theta) r_t \right\} \). Observing the slope of the yield curve and short rate each day can be helpful in deriving a time series for the parameter \( \lambda \). By comparing the empirical slope with its corresponding analytical slope, we can generate this time series. We report the value of \( \lambda \) by taking an average of this time series. To end this, we consider a dataset consisting of daily rates on US treasury bills with maturity times of 3 months, 6 months, and 1 year. The 3-month maturity rate is considered the short rate, while the 6-month and 1-year maturity rates are used to calculate the yield curve’s slope per day. Using this approach, we have determined that \( \lambda = -0.033 \). This value is then applied to the relation (3.3) to find the risk-neutral parameters, see Table (2).

<table>
<thead>
<tr>
<th>Risk neutral parameters</th>
<th>( \theta^* )</th>
<th>( m^* )</th>
<th>( \sigma^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation</td>
<td>0.221</td>
<td>0.013</td>
<td>0.074</td>
</tr>
</tbody>
</table>

Table 2: Estimates of the CIR model parameters under risk-neutral measure.

Figure 2: \( \lambda \) time series
6.2 Parameters estimation of aggregate claim process

6.2.1 Loss data description

The data employed in this paper is the property damages (in dollars) resulting from earthquakes in the US, spanning the period from 1906 to 2005; more information can be found in Vranes and Pielke (2009). To take the inflation into account, we adjust the data to 2020 value using the Consumer Price Index\(^{12}\) (CPI). As the yearly observation for the CPI covers years between 1947 to 2021, a generalized additive model is used to predict CPI values from 1906 to 1946. Figure 3 shows the annual frequency of earthquake occurrence and the adjusted losses per year.

![Figure 3](https://example.com/earthquake_data.png)

Figure 3: Left plot illustrates the number of claims, per year, from April 1906 to September 2004, right plot illustrates the adjusted loss severity (on log-scale), per year, from April 1906 to September 2004.

6.2.2 Loss distribution

In the CAT bond pricing literature, a wide range of heavy-tailed distributions can be considered for the loss data. The following loss severity distributions, as discussed in Giuricich and Burnecki (2019), are selected to fit the data at our disposal:

\(^{12}\)Available at [https://fred.stlouisfed.org/series/CPIAUCSL](https://fred.stlouisfed.org/series/CPIAUCSL)
A Burr type XII distribution with shape parameters $a > 0$ and $b > 0$, and scale parameter $c > 0$ is given by:

$$Burr(a, b, c) = \frac{ab(x/c)^{a-1}}{(1+(x/c)^a)^{b+1}}$$ (6.11)

A generalized Pareto distribution with shape parameter $a \in \mathbb{R}$ and scale parameter $c > 0$ is given by:

$$GP(a, c) = \frac{1}{c}(1 + \frac{ax}{c})^{-(1+\frac{1}{a})}$$ (6.12)

A generalized Extreme Value distribution with shape parameter $a \in \mathbb{R} \setminus \{0\}$, location parameter $b \in \mathbb{R}$, and scale parameter $c > 0$ is given by:

$$GEV(a, b, c) = \frac{1}{c}exp\left(-\left(1 + \frac{a(x-b)}{c}\right)^{-\frac{1}{a}}\right) \times \left(1 + \frac{a(x-b)}{c}\right)^{-1-\frac{1}{a}}$$ (6.13)

Using the MLE method, we estimate all parameters contained in the above distributions, and then we perform the goodness-of-fit test for fitted distributions by means of well-known non-parametric tests, namely, the Kolmogorov-Smirnov (KS), Anderson-Darling (AD), and Cramér Von Mises (CVM) tests, each of which measures the discrepancy between empirical distribution and theoretical distribution in its own way (see, e.g., Ma and Ma 2013). Following Giuricich and Burnecki (2019), we approximate the $p$-values corresponding to each test via the bootstrap method as non-parametric tests basically require distributions to be fully specified while in our case we applied fitted distributions.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Parameters</th>
<th>MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burr type XII</td>
<td>$a, b, c$</td>
<td>$0.82, 0.70, 2.85 \times 10^7$</td>
</tr>
<tr>
<td>Generalized Pareto</td>
<td>$a, c$</td>
<td>$2.2, 2.6 \times 10^7$</td>
</tr>
<tr>
<td>Generalized extreme value</td>
<td>$a, b, c$</td>
<td>$2.6, 1.05 \times 10^8, 2.72 \times 10^8$</td>
</tr>
</tbody>
</table>

Table 3: Estimated parameters of loss severity distributions using MLE

The MLE of parameters and goodness-of-fit test results at the confidence level 0.05 have been provided in Tables 3 and 4, respectively. The results show that GEV cannot be fitted to our data. To check the performance of fitted distributions, we use Akaik information criteria (AIC) and Bayesian information criteria (BIC). The results of Table 5 show that GPD fits the best among
## 6.2.3 Frequency distribution

In regard to the distribution of the counting process $N_t$ which models the number of claims, we already considered different distributions for inter-arrival times, namely, Exponential, Gamma, Weibull, and a mixture of two Exponential, each resulting in different models for $N_t$. Our goal here is to fit these distributions to arriving time observations (in days) extracted from the recorded dates of earthquake occurrence and then assess their fit adequacy by implementing goodness-of-fit tests as elaborated in the previous subsection.

Non-parametric tests provided in Table 7 suggest that all distributions are suited for the arriving time observation. We observe from Table 8 that the Gamma distribution possesses the lowest AIC and BIC values, which means it fits the best among others. This provides evidence supporting the fact that in real applications, there exist cases where the exponential assumption for the inter-arrival times of catastrophe events may not be an optimal choice, and so one needs to work under a more general framework than the Poisson process.
Distribution Parameters MLE

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Parameters</th>
<th>MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>$\hat{\delta}$</td>
<td>0.002</td>
</tr>
<tr>
<td>Gamma</td>
<td>$\eta_1, \eta_2$</td>
<td>0.76, 0.002</td>
</tr>
<tr>
<td>Wiebull</td>
<td>$\eta_3, \eta_4$</td>
<td>0.84, 398.70</td>
</tr>
<tr>
<td>Mix-Exponential</td>
<td>$\phi, \eta_5, \eta_6$</td>
<td>0.87, 0.002, 0.02</td>
</tr>
</tbody>
</table>

Table 6: Estimated parameters of inter-arrival time distributions (under physical measure) using MLE, where the Expectation-Maximization algorithm is adopted for the case of the Mix-Exponential distribution.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>KS</th>
<th>AD</th>
<th>CVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>0.09(0.13)</td>
<td>1.18(0.08)</td>
<td>0.15(0.16)</td>
</tr>
<tr>
<td>Gamma</td>
<td>0.07(0.26)</td>
<td>0.25(0.74)</td>
<td>0.04(0.70)</td>
</tr>
<tr>
<td>Wiebull</td>
<td>0.07(0.13)</td>
<td>0.31(0.53)</td>
<td>0.05(0.46)</td>
</tr>
<tr>
<td>Mix-Exponential</td>
<td>0.06(0.42)</td>
<td>0.27(0.61)</td>
<td>0.03(0.65)</td>
</tr>
</tbody>
</table>

Table 7: Test statistic values of in-sample goodness-of-fit tests for fitted distributions with their corresponding $p$-values in the parentheses, each being estimated by performing 1000 Monte Carlo simulations.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Exponential</th>
<th>Gamma</th>
<th>Wiebull</th>
<th>Mix-Exponential</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>1175.830</td>
<td>1173.201</td>
<td>1173.815</td>
<td>1177.676</td>
</tr>
<tr>
<td>BIC</td>
<td>1178.249</td>
<td>1178.039</td>
<td>1178.653</td>
<td>1187.351</td>
</tr>
</tbody>
</table>

Table 8: Goodness-of-fit criterion for checking the performance of different fitted inter-arrival time distributions.

### 6.3 Sensitivity analysis

Let the CAT bond contract of our interest cover earthquake losses with face value $Z = 1$ USD and $q = 0.5$. The distribution for severity variables is considered to be a GPD as defined in (6.12) with estimated parameters depicted in Table 3. The threshold level $D$ is selected as a point beyond which the mean excess plot is linearly increasing according to the Extreme Value Theorem. We start by looking at scenario I, where under the physical measure $L_t$ is a CRP with inter-arrival distribution being Exponential, Gamma, Weibull, and Mixture of exponential,
while under risk-neutral measure $L_t$ is a CPP. We aim to investigate how the CAT bond price changes with respect to the risk premium $c$ and maturity time $T$ under the aforesaid distributions. The function $\beta_3(x)$ is chosen for transforming the severity distribution function once switching from the physical measure to the risk-neutral measure. It is easy to show that by choosing $\beta_3(x)$, we will end up with $\kappa = 1$ and $\nu(x) = \frac{\exp(cx)}{E\{\exp(cX_1)\}}$, where the parameter $c$ can be seen as an indicator of the earthquake risk premium, and $\nu(x)$ is the well-known Esscher transform, which is commonly employed as a kernel function for pricing in incomplete markets. Figure 4 represents the change of CAT bond price with respect to premium and maturity time, indicating the fact that a higher-risk premium (required by the bond investor) and a longer maturity time (i.e., the time at which the contract is expired) result in a lower bond price. While CAT bond prices with Gamma and Weibull distributions are deemed to behave similarly to that with exponential distribution (considered as a benchmark), the bond price with the mixture of exponential distribution tends to be higher for very small and very large values of premium $c$. As far as the change of maturity is concerned, the CAT bond price with the mixture of exponential tends to be consistently higher than that with exponential distribution, in contrast to the

Figure 4: Plots on the top depict the change of CAT bond price with respect to the risk premium $c$, assuming different inter-arrival time distributions with maturity $T = 2$, while those on the bottom show the change of CAT price with respect to maturity $T$ with $c = 0.6$.
case of Gamma and Weibull distributions in which the CAT bond price is either higher or lower.

In scenario II, we assumed that the inter-arrival time is distributed as a Gamma or a mixture of exponential under both measures, in that, \( L_t \) is a compound renewal process under \( P_2 \) and \( Q_2 \), in which case we applied Bayesian inference to find parameter estimates of inter-arrival time distributions under \( Q_2 \). Table 9 summarizes the results based on relations (5.14), (D.20), and (D.21):

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Parameters</th>
<th>Bayes estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma</td>
<td>( \epsilon_1, \epsilon_2 )</td>
<td>2.474, 0.01</td>
</tr>
<tr>
<td>Mix-Exponential</td>
<td>( \phi^*, \eta_7, \eta_8 )</td>
<td>0.85, 1.063, 0.063</td>
</tr>
</tbody>
</table>

Table 9: Estimate of parameters contained in Gamma and mixture of two exponential distributions using Bayesian framework when scenario II is under consideration.

Figure 5: Plot (a) and (b) show the change of CAT bond price with respect to premium \( c \) and maturity time \( T \), respectively, by considering scenario II for which the inter-arrival time distribution under both measures is assumed to be Gamma distribution (solid line) and a mixture of two exponential distribution (dash line).

We observe from Figure 5 that the CAT bond price with Gamma distribution is higher than the mixture of exponential for a small value of premium \( c \) while their price difference becomes negligible as the premium increases. On the other hand, the CAT bond price with Gamma distribution is higher compared with the mixture of exponential before maturity \( T = 2.3 \) (in...
the year), and as we move beyond this point the price for the mixture of exponential becomes higher. Based on our analysis, it is obvious that the choice of inter-arrival distribution impacts more significantly on the CAT bond price in scenario II.

7 Conclusion

We have developed a pricing model for a zero-coupon CAT bond, in which the aggregate loss process used in the CAT bond payoff function is allowed to follow a general compound renewal process, where the inter-arrival time distribution can be a different distribution, such as Gamma, Weibull, or a mixture of exponential, than the exponential distribution. Our pricing framework ensures that a compound renewal process under the physical measure remains a compound renewal process under a risk-neutral measure. We derived the pricing formula under two general scenarios, where the aggregate loss process either follows a compound renewal process under the physical measure but a compound Poisson process under the risk-neutral measure (scenario I), or follows a compound renewal process under both measures (scenario II).

Throughout the paper, we made the assumption that financial risk and insurance risk behave independently. Therefore, we modeled interest rates separately using the Cox-Ingersol-Ross model, and calibrated the market price of risk using market data. For the loss distribution, we fitted different heavy-tailed distributions to earthquake data, and selected the Generalized Pareto distribution based on the AIC and BIC criteria. Under scenario I, we used Maximum Likelihood Estimation and Expectation Maximization methods to estimate the parameters contained in the inter-arrival time distribution and severity distribution. In contrast, under scenario II, we solved a Bayes minimization problem based on the Kulback-Laibler divergence loss function. Finally, we conducted a sensitivity analysis to examine the impact of premium and maturity time on the CAT bond price.

The numerical experiments demonstrated that the choice of inter-arrival time distribution has a significant impact on the price of the CAT bond under both scenarios, with scenario II showing a more pronounced effect. Furthermore, the results indicate that an increase in the premium and maturity time of the CAT bond leads to a decrease in its price. To establish a benchmark price, we assumed that the aggregate loss process \( L_t \) follows a Compound Poisson process under both measures, implying that the counting process follows a Poisson process under both measures. In scenario I, we observed that the deviation of the bond price from our benchmark price was greater for the mixture of exponential distribution than for the Gamma and Weibull distributions at very large and small values of premium.
8 Funding Acknowledgments

This research was financed by Fundación MAPFRE.
References


Electronic copy available at: https://ssrn.com/abstract=4599125


Merton R C (1976) Option pricing when the underlying stock returns are discontinuous. Journal of Financial Economica, pp. 125-144


A Appendix

Proof. Recall that the distribution of renewal process $N_t$ under measure $Q_2$ is given by

$$Q_2(N_t = n) = Q_2(\{N_t \geq n\} \cap \{N_t \geq n + 1\}^c)$$

$$= Q_2(\{N_t \geq n\}) - Q_2(\{N_t \geq n + 1\})$$

$$= Q_2(T_n \leq t) - Q_2(T_{n+1} \leq t), \quad (A.1)$$

where $T_n = \sum_{i=1}^{n} W_i$ (i.e., the arrival time of $n$-th event) can be written as the sum of $n$ interarrival times $W_i$’s, each of which follows a gamma distribution with parameter $n\epsilon_1$ and $\epsilon_2$. Similarly, $T_{n+1}$ is a Gamma with parameters $(n+1)\epsilon_1$ and $\epsilon_2$. Hence, we have that

$$Q_2(N_t = n) = \sum_{s=n\epsilon_1}^{+\infty} \frac{e^{-\frac{\epsilon_2 t}{s}} s^n}{s!} - \sum_{s=(n+1)\epsilon_1}^{+\infty} \frac{e^{-\frac{\epsilon_2 t}{s}} s^{n+1-1}}{s!}$$

$$= \sum_{s=n\epsilon_1}^{+\infty} \frac{e^{-\frac{\epsilon_2 t}{s}} s^n}{s!}, \quad (A.2)$$

which means that the event that exactly $n$ events have occurred in the renewal process may be thought of as the event that between $n\epsilon_1$ and $n\epsilon_1 + \epsilon_1 - 1$ have occurred in the underlying Poisson process with intensity parameter $\epsilon_2$. Accordingly, in example 4.3.5, we can say that the problem of renewal process can be turned into the problem of Poisson process where the counting process $N_t$ takes its values between $n\epsilon_1$ and $n\epsilon_1 + \epsilon_1 - 1$ for $n = 0, 1, 2, \ldots$. ■

B Appendix

Proof. To find the distribution function associated with $L_t$ under the assumptions presented in example 4.3.6, we first obtain the probability function of claim number $N_t$, which itself follows a counting renewal process. To this purpose, it is then essential, according to relation (A.1), to derive the distribution function associated with the $n$-th arrival time $T_n$, which is a sum of i.i.d random variables following a mixture of two exponential distributions $\text{MIX-EXP}(\rho(\delta))$. We express our problem in the following with a general format.

Let random variable $Y$ follow a mixture of two exponential distributions that is

$$Y \sim \pi_1 \text{EXP}(\lambda_1) + \pi_2 \text{EXP}(\lambda_2), \quad (B.1)$$

where $\pi_i$ are mixing weights with the property that $\pi_1 + \pi_2 = 1$. Furthermore, assume that $\{Y_1, Y_2, \ldots, Y_n\}$ is a sample of i.i.d random variables drawn from $f_Y(y)$. Then, the probability
density function of $Z = \sum_{j=1}^{n} Y_j$ is given by

$$f_Z(z) = \sum_{k=0}^{n} \binom{n}{k} \pi_k^{n-k} \frac{\lambda_1^{n-k}}{\Gamma(n)} \frac{\lambda_2^{k}}{\Gamma(n-k)} n^{-1} F(n - k, n, (\lambda_1 - \lambda_2)z)$$ \hspace{1cm} (B.2)

with $1 F(a, b, c)$ being a confluent hypergeometric function of the first kind defined as below:

$$1 F(a; c; z) = \left\{ \begin{array}{ll}
\sum_{i=1}^{\infty} \frac{(a)_i}{(b)_i} \frac{t^i}{i!} & \text{hypergeometric series representation} \\
\frac{\Gamma(a + i)}{\Gamma(b)} \int_0^1 e^{ct} t^{a-1} (1 - t)^{b-a} \
\end{array} \right. \text{ Integral representation} \hspace{1cm} (B.3)

where the Pochhammer symbol $(a)_i = \frac{\Gamma(a + i)}{\Gamma(a)} = a(a + 1) \cdots (a + i - 1)$.

Define random vector $\Upsilon = (\Upsilon_1, \Upsilon_2, \cdots, \Upsilon_n)$ consisting of $n$ i.i.d Bernoulli variable $\Upsilon_i \sim \text{Ber}(\pi_1)$ which are associated with each $Y_i$ in this way:

$$Y_i|\Upsilon_i = k \sim \text{EXP}(\lambda_{2-k}); \quad \text{for} \quad k = 0, 1.$$ \hspace{1cm} (B.4)

We define $\Upsilon^* = \sum_{i=1}^{n} \Upsilon_i \sim \text{Bin}(n, \pi_1)$. Conditioning on $\Upsilon^*$, we can write

$$Z = \sum_{j=1}^{n} Y_j|\Upsilon^* = \left( \sum_{j: \Upsilon_j = 0} Y_j + \sum_{j: \Upsilon_j = 1} Y_j \right)|\Upsilon^*,$$ \hspace{1cm} (B.5)

from which one can conclude that $Z|\Upsilon^* \sim Z_1 + Z_2$, where

$$Z_1 \sim \text{Ga}(\Upsilon^*, \lambda_1) \quad \text{and} \quad Z_2 \sim \text{Ga}(n - \Upsilon^*, \lambda_2)$$ \hspace{1cm} (B.6)

This shows that the distribution of $Z$ given $\Upsilon^*$ can be written as the sum of two independent random variables distributed as gamma with different shape and rate parameters. Denote by $\text{GDC}(a, b, c; d; x)$ the gamma distribution convolution consisting of $\text{Ga}(a, b)$ and $\text{Ga}(c, d)$, which is given by:

$$\text{GDC}(a, b, c, d; x) = \left\{ \begin{array}{ll}
\frac{b^{a}c^{d}}{\Gamma(a+c)} e^{-bx} x^{a+c-1} F(c, a + c, (b - d)x), & x > 0 \\
0, & x \leq 0 \\
\end{array} \right.$$ \hspace{1cm} (B.7)
Finally, marginalizing in \( Z \) leads to

\[
f_Z(z) = \sum_{k=0}^{n} f(z|\Upsilon^*)p(\Upsilon^* = k) = \sum_{k=0}^{n} \binom{n}{k} \pi_1^k \pi_2^{n-k} \text{GDC}(k, \lambda_1, n-k, \lambda_2, z),
\]

which can be viewed as a mixture of \((n+1)\) components with mixing weights equal to \( \binom{n}{k} \pi_1^k \pi_2^{n-k} \).

Using (A.1) and (B.2), we derive the probability function of the renewal process \( N_t \) in example 4.3.6,

\[
Q_2(N_t = n) = Q_2(T_n \leq t) - Q_2(T_{n+1} \leq t) = \sum_{k=0}^{n} \binom{n}{k} \pi_1^k \pi_2^{n-k} \text{CGDC}(k, \lambda_1, n-k, \lambda_2, t) - \sum_{k=0}^{n+1} \binom{n+1}{k} \pi_1^k \pi_2^{(n+1)-k} \text{CGDC}(k, \lambda_1, (n+1)-k, \lambda_2, t)
\]

where the notation \( \text{CGDC} \) denotes the cumulative distribution function of \( \text{GDC} \), \( \pi_1 = \Phi^* \), \( \pi_2 = 1 - \Phi^* \), \( \lambda_1 = \eta_7 \), and \( \lambda_2 = \eta_8 \).

C Appendix

**Proof.** Assume that \( \mathcal{F}^W = \{ \mathcal{F}^W_n \}_{n \in \mathbb{N}_0} \) and \( \mathcal{F}^X = \{ \mathcal{F}^X_n \}_{n \in \mathbb{N}_0} \) are the natural filtration of \( W \) and \( X \), respectively. It can be shown that the following holds true (for unexplained details, interested readers can refer to Macheras and Tzaninis (2020)):

\[
\forall A \in \mathcal{F}^L_t \exists B_k \in \sigma(\mathcal{F}^W_t \cup \mathcal{F}^X_t) \text{ (for every } k \in \mathbb{N}_0) \text{ s.t. } A \cap \{N_t = k\} = B_k \cap \{N_t = k\}
\]

(C.1)

Using (C.1) we yield

\[
Q_2(A) = \sum_{k=0}^{\infty} Q_2(B_k \cap \{N_t = k\}) = \sum_{k=0}^{\infty} Q_2(B_k \cap \{T_k \leq t\} \cap \{W_{k+1} > t - T_k\})
\]

(C.2)
For a fixed but arbitrary $n \in \mathbb{N}_0$, we define

$$G_n = \bigcap_{j=1}^n (W_j^{-1}(E_j) \cap X_j^{-1}(F_j)) \cap \{W_{n+1} > t - T_n\}$$  \hspace{1cm} (C.3)

where $E_j, F_j \in \mathcal{B}(\mathbb{R}_+)$ for any $j \in \{1, 2, \cdots, n\}$. We then have that

$$Q_2(G_n) = \mathbb{E}_Q^n \left[ I_{E_1}(W_1) I_{F_1}(X_1) \cdots I_{E_n}(W_n) I_{F_n}(X_n) I_{\{W_{n+1} > t - T_n\}} \right]$$

$$= \left[ \prod_{j=1}^n \mathbb{E}_Q^n [I_{E_j}(W_j)] \right] \times \left[ \prod_{j=1}^n \mathbb{E}_Q^n [I_{F_j}(X_j)] \right]$$

$$= \left[ \prod_{j=1}^n \mathbb{E}_P^n \left[ dQ_{W_1}^2(W_j) I_{E_j}(W_j) \right] \right] \times \left[ \prod_{j=1}^n \mathbb{E}_P^n \left[ dQ_{X_1}^2(X_j) I_{F_j}(X_j) \right] \right]$$

$$\times \frac{\mathbb{E}_Q^n [I_{\{W_{n+1} > t - T_n\}}]}{\mathbb{P}_P^n [I_{\{W_{n+1} > t - T_n\}}]}$$

$$= \mathbb{P}_P^n \left[ I_{G_n} \prod_{j=1}^n \left( \frac{dQ_{W_1}^2(W_j)}{d\mathbb{P}_P^n(W_j)} \frac{dQ_{X_1}^2(X_j)}{d\mathbb{P}_P^n(X_j)} \right) \right] \times \frac{Q_2(W_{n+1} > t - T_n)}{\mathbb{P}_P^n (W_{n+1} > t - T_n)}$$  \hspace{1cm} (C.4)

A monotone class argument allows one to employ (C.4) in (C.2) for any $k \in \mathbb{N}_0$, which leads to the following Radon-Nikodym derivative.

$$\frac{dQ_2}{d\mathbb{P}_P^n} |_{x_t} = \left[ \prod_{j=1}^{N_t} \left( \frac{dQ_{W_1}^2(W_j)}{d\mathbb{P}_P^n(W_j)} \frac{dQ_{X_1}^2(X_j)}{d\mathbb{P}_P^n(X_j)} \right) \right] \times \frac{Q_2(W_{n+1} > t - T_{N_t})}{\mathbb{P}_P^n (W_{n+1} > t - T_{N_t})}$$  \hspace{1cm} (C.5)

---

**D Appendix**

**Proof.** There are two ways to specify the likelihood function of a mixture model. To illustrate this clearly, we start with the basic definition. Generally, a random variable $Y$ with a $g$-component mixture density can be generated in this way. Suppose that $Z$ be the set of $g$ latent variables where the $k$-th element of $Z$ is defined to be zero or one, according to whether the component of origin of $Y$ in the mixture is equal to $k$ or not. More precisely, consider $Z_j$ as the latent variable vector for the $j$-th sample point. Therefore, $Z_j$ can be a vector like this:

$$Z_j = (Z_{1j}, Z_{2j}, \cdots, Z_{gj}) \quad \text{or} \quad Z_j = (Z_{kj}); \quad k = 1, 2, \cdots, g$$  \hspace{1cm} (D.1)
where $Z_{kj} = 1$ if the $j$-th sample point belongs to the $k$-th component (or cluster), otherwise $Z_{kj} = 0$. We also know that each element of vector $Z_j$ occurs independently with probabilities $\pi_1, \pi_2, \cdots, \pi_g$. So, it is easy to see that $Z_j$ is distributed according to a multinomial distribution consisting of one draw on $g$ categories with probabilities $\pi_1, \pi_2, \cdots, \pi_g$, i.e., $Z_j \sim \text{Multi}(1, \pi)$, where $\pi = (\pi_1, \pi_2, \cdots, \pi_g)$. The probability density function is then given by:

$$p(z_j) = \mathbb{P}(Z_j = z_j) = \mathbb{P}(Z_{1j} = 1)^{z_{1j}} \times \mathbb{P}(Z_{2j} = 1)^{z_{2j}} \times \cdots \times \mathbb{P}(Z_{gj} = 1)^{z_{gj}} = \prod_{k=1}^{g} \pi_k^{z_{kj}}$$ (D.2)

For example, if we know that the $j$-th sample point belongs to the second component, then for the vector $Z_j$ we observe the vector $z_j = (z_{1j}, z_{2j}, \cdots, z_{gj}) = (0, 1, 0, \cdots, 0)$ with the probability function:

$$p(z_j) = \mathbb{P}(Z_j = (0, 1, 0, \cdots, 0)) = \pi_2$$ (D.3)

Let $z_j = k$ denotes the fact that the $j$-th sample point fall in the $k$-th component, i.e., the $k$-th element of vector $z_j$ is equal to one and the others are equal to zero. Based on this definition, we make another assumption which says that the conditional density of $y_j$ given $z_j = k$ is $f_k(y_j)$. Then, following the same logic used for the $p(z_j)$, we can conclude that

$$f(y_j | z_j) = \prod_{k=1}^{g} f_k(y_j)^{z_{kj}}$$ (D.4)

In order to obtain the mixture model of $f(y)$, we just need to apply the Bayes rule by summing up the terms on $z$ to get the probability density function of $f(y)$. That is,

$$f(y) = \sum_{k=1}^{g} f(y, z) = \sum_{k=1}^{g} f(y | z) p(z) = \sum_{k=1}^{g} \pi_k f_k(y)$$ (D.5)

where $f_k(y)$ are called the component densities of the mixture model and the $\pi_k$ are mixing weights that satisfy:

$$\sum_{k=1}^{g} \pi_k = 1 \quad \text{and} \quad 0 \leq \pi_k \leq 1 \quad (k = 1, 2, \cdots, g)$$ (D.6)
Suppose that \( \mathcal{Y} = (y_1, y_2, \ldots, y_n) \) is the vector of \( n \) observed data points. Using (D.5), the likelihood function is then given by

\[
f(\mathcal{Y}; \Theta) = \prod_{i=1}^{n} \left\{ \sum_{k=1}^{g} \pi_k f_k(y_i; \Theta_k) \right\}
\]  
(D.7)

Another useful decomposition of (D.7) is based on latent variables. We consider the vector \( \mathcal{X} = (\mathcal{Y}, \mathcal{C}) \) involving latent variables \( Z_i \) defined in (D.1), for which \( \mathcal{C} = (Z_1, Z_2, \ldots, Z_n) \), where \( Z_i = (Z_{ki}) \) with \( \mathbb{P}(Z_{ki} = 1) = \pi_k \), for \( i = 1, 2, \ldots, n \) and \( k = 1, 2, \ldots, g \). Therefore, the complete version of the likelihood function becomes

\[
f(\mathcal{X}; \Theta) = f(\mathcal{Y}, \mathcal{C}; \Theta) = \prod_{i=1}^{n} f(y_i, Z_i; \Theta) = \prod_{i=1}^{n} f(y_i|Z_i; \Theta)p(Z_i)
\]

\[
= \prod_{i=1}^{n} \left[ \prod_{k=1}^{g} f_k(y_i; \Theta_k)^{Z_{ki}} \right] \times \prod_{k=1}^{g} \pi_k^{Z_{ki}}
\]

\[
= \prod_{i=1}^{n} \prod_{k=1}^{g} [\pi_k f_k(y_i; \Theta_k)]^{Z_{ki}}
\]  
(D.8)

We adopt the above-mentioned setting to construct our conditional distribution for Bayesian framework under prior distributions selected for unknown parameters contained in the mixture model. For simplicity in notation and avoidance of confusion, suppose that \( z = (z_1, z_2, \cdots, z_n) \) is a vector of latent variables where \( z_i \), corresponding to the \( i \)-th sample point, takes value \( k \in \{1, 2, \cdots, g\} \) with probability \( p(z_i = k) = w_k \). The prior of mixing weights \( w = (w_1, w_2, \cdots, w_g) \) is chosen to be a standard Dirichlet prior, that is,

\[
w \sim \text{Dirichlet}_g(\alpha_1, \alpha_2, \cdots, \alpha_g) = \frac{\Gamma(\alpha)}{\prod_{k=1}^{g} \Gamma(\alpha_k)} \prod_{k=1}^{g} w_k^{\alpha_k-1}
\]  
(D.9)

where \( \alpha_1, \alpha_2, \cdots, \alpha_g \) are non-negative hyper-parameters, and \( \alpha = \sum_{k=1}^{g} \alpha_k \). Inspired by (D.2), one can write the conditional distribution for the cluster allocations in this way:

\[
p(z|w) = \prod_{k=1}^{g} w_k^{n_k}
\]  
(D.10)

where \( n_k = \sum_{i=1}^{n} I_{\{k\}}(z_i) \) is the number of samples attributed to the \( k \)-th cluster. By marginalising (D.10) with respect to \( w \), we get the marginal distribution of \( z \),

\[
p(z) = \frac{\Gamma(\alpha)}{\Gamma(\alpha + n)} \prod_{k=1}^{g} \frac{\Gamma(\alpha_k + n_k)}{\Gamma(\alpha_k)}
\]  
(D.11)
Motivated by (D.8), the joint distribution for $y = Y$ and $z$ is conveniently proportional with

$$p(y, z|\Theta) \propto \prod_{k=1}^{g} \left\{ \Gamma(\alpha_k + n_k) \prod_{i:z_i=k} f_k(y_i|\Theta_k) \right\}$$  \hspace{1cm} (D.12)

It is assumed that the vector $w$ is independent of $y$, which can be considered reasonable in a mixture model. Therefore, from (D.9), (D.10), (D.11), we are able to find the conditional distribution $p(w|z)$, i.e., the posterior distribution of $w$,

$$p(w|z) = \frac{p(z|w)p(w)}{p(z)} = \text{Dirichlet}_g(\alpha_1 + n_1, \alpha_2 + n_2, \cdots, \alpha_g + n_g)$$  \hspace{1cm} (D.13)

In addition, it is possible to find the posterior distribution of $\Theta$ by associating a conjugate prior with each parameter $\Theta_k$ in (D.12). In other words,

$$p(\Theta|y, z) \propto \prod_{k=1}^{g} \left\{ \prod_{i:z_i=k} f_k(y_i|\Theta_k) \right\} \pi(\Theta_k)$$  \hspace{1cm} (D.14)

We are now in a position to specify our non-parametric Bayesian framework for the mixture of two exponential distributions as follows:

$$y_i|z_i = k \sim \text{EXP}(\lambda_k)$$

$$w = (w_1, w_2) \sim \text{Dirichlet}_2(\alpha_1, \alpha_2)$$

$$\lambda_k \sim \text{Ga}(\tau_k, \Psi_k); \quad i = 1, 2, \cdots, n, \quad k = 1, 2.$$

where $(\alpha_k, \tau_k, \Psi_k)$ are known hyper-parameters. The corresponding Gibbs samplers with states $t = 1, 2, \cdots$ can be written as follows:

**Algorithm 2** Gibbs sampling for an exponential mixture

**Initialize:** $w^{(0)} = (w_1^{(0)}, w_2^{(2)})$ and $\Theta^{(0)} = (\lambda_1^{(0)}, \lambda_2^{(0)})$

for $t = 1, 2, 3, \cdots$ do

Generate $z_i^{(t)}$ ($i = 1, \cdots, n, k = 1, 2$) from

$$p(z_i^{(t)} = k|w_k^{(t-1)}, \lambda_k^{(t-1)}, y_i) \propto w_k^{(t-1)} \lambda_k^{(t-1)} y_i \exp\{-\lambda_k^{(t-1)} y_i\}$$

Compute $n_k^{(t)} = \sum_{i=1}^{n} I_{\{k\}}(z_i)$ and $(s_k^{y})^{(t)} = \sum_{i=1}^{n} I_{\{k\}}(z_i) y_i$

Generate $w^{(t)}$ from Dirichlet$_2(\alpha_1 + n_1^{(t)}, \alpha_2 + n_2^{(t)})$

Generate $\lambda_k^{(t)}$ from Ga$(\tau_k + n_k^{(t)}, \Psi_k + (s_k^{y})^{(t)})$

end for
To be able to find the Bayes estimate of mixture parameters \( \Theta = (w_1, w_2, \lambda_1, \lambda_2) \), we need to solve the minimization problem in (5.6), which can be easily shown to be equivalent to the following maximization problem:

\[
\Theta^B = \arg \min_{\hat{\Theta}} \mathbb{E}_\theta [KL(\Theta \parallel \hat{\Theta}) | y]
\]

\[
= \arg \min_{\hat{\Theta}} \mathbb{E}_\theta \left[ \mathbb{E}_Y [\log \left( \frac{f(Y; \Theta)}{f(Y; \hat{\Theta})} \right)] | y \right]
\]

\[
= \arg \max_{\hat{\Theta}} \mathbb{E}_\theta \left[ \mathbb{E}_Y [\log f(Y; \hat{\Theta})] | y \right]
\]

(D.15)

where the inner expectation is taken with respect to \( \mathbf{Y} \), while the outer expectation is computed under the posterior distribution \( \Theta | y \). Due to the complexity with which the objective function in (D.15) can be optimized, we apply the idea of the EM method for which a complete version of the likelihood function is considered, that is, we replace \( \log f(Y; \hat{\Theta}) \) with

\[
\log f(Y, \mathbf{Z}; \hat{\Theta}) = \sum_{j=1}^{n} \sum_{k=1}^{g} Z_{kj} \left[ \log \hat{w}_k + \log f_k(y_j | \hat{\lambda}_k) \right]
\]

(D.16)

The E-step suggests substituting the latent variable \( Z_{kj} \) with an expectation that is taken under probability distribution of the latent variable conditionally to the observed data and current value of the parameter, that is, we replace \( Z_{kj} \) with

\[
\mathbb{E}[Z_{kj}|y_j, \hat{\Theta}_k] = \mathbb{P}(Z_{kj} = 1|y_j, \hat{\Theta}_k) = \frac{f(y_j|Z_{kj} = 1, \hat{\lambda}_k)\mathbb{P}(Z_{kj} = 1)}{\sum_{l=1}^{g} f(y_j|Z_{lj} = 1, \hat{\lambda}_k)\mathbb{P}(Z_{lj} = 1)}
\]

\[
= \frac{f_k(y_j | \hat{\lambda}_k)\hat{w}_k}{\sum_{l=1}^{g} f_k(y_j | \hat{\lambda}_k)\hat{w}_l} = H_k(y_j)
\]

(D.17)

Given the value \( \hat{\Theta}^{(t-1)} \) at the \((t - 1)\)-th iteration, the E-step calculates

\[
Q(\hat{\Theta}, \hat{\Theta}^{(t-1)}) = \sum_{j=1}^{n} \sum_{k=1}^{g} (\log \hat{w}_k) \mathbb{E}_\theta \left[ \mathbb{E}_Y \left[ H_k^{(t-1)}(Y_j) \right] | y \right] + \mathbb{E}_\theta \left[ \mathbb{E}_Y \left[ H_k^{(t-1)}(Y_j) \log f_k(Y_j | \hat{\lambda}_k) \right] | y \right]
\]

(D.18)

After that the M-step finds the revised parameter \( \hat{\Theta}^{(t)} \) according to the following problem

\[
\hat{\Theta}^{(t)} = \arg \max_{\hat{\Theta}} Q(\hat{\Theta}, \hat{\Theta}^{(t-1)})
\]

(D.19)

The iteration between the E-step and M-step continues until a convergence criterion is satisfied.
Solving (D.19) yields

\[ \hat{w}_k^{(t)} = \frac{\sum_{j=1}^{n} \mathbb{E}_\Theta \left[ \mathbb{E}_Y \left[ H_k^{(t-1)}(Y_j) \right] | y \right]}{n} \]  \hspace{1cm} (D.20)

\[ \hat{\lambda}_k^{(t)} = \arg \max_{\lambda_k} \sum_{j=1}^{n} \mathbb{E}_\Theta \left[ \mathbb{E}_Y \left[ H_k^{(t-1)}(Y_j) \log f_k(Y_j|\lambda_k) \right] | y \right] \]  \hspace{1cm} (D.21)