Claim count modelling with shot noise Cox processes

Chung-Yu Liu

School of Risk and Actuarial Studies

Australian School of Business Thesis

Disclaimer and Copyright

University of New South Wales

School of Risk and Actuarial Studies
Australian School of Business

Thesis

Claim count modelling with shot noise Cox processes

Chung-Yu Liu

Under the supervision of:
Dr. Benjamin Avanzi
and
Dr. Bernard Wong

A thesis submitted in partial fulfilment of the requirements for the degree of Bachelor of Commerce (Honours in Actuarial Studies)

DECLARATION

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

Signed:

_____

Date:

_____

ABSTRACT

Environmental and economic events may lead to sudden changes in the claim arrival rate for an insurer. These random fluctuations cannot be captured by the homogeneous Poisson process. Hence doubly stochastic Poisson processes, or Cox processes, have been introduced as a tool which allows for stochasticity in the claim intensity. In particular, the shot noise Cox process has analytically tractable properties as well as a nice physical interpretation from an insurer's perspective. Despite extensive developments in the theory of doubly stochastic Poisson processes and applications in physical sciences, finance and mortality, very little has been done in using the model in an insurance context for claim counts.

This research calibrates and compares two methods to fit shot noise Cox processes to claim insurance data. The main issue in modelling lies in fitting parameters to an unobservable intensity process. We propose a framework for applying the Kalman filter based on a Gaussian approximation of the shot noise process. We also calibrate a Markov Chain Monte Carlo filtering method previously applied on high frequency trading data to use in a non-life insurance context. In particular, we improve the stochastic expectation maximisation method by reducing the dimension of the optimisation problem.

The proposed methods are then calibrated and validated through simulation studies which reflect the nature of insurance data. Computational challenges in implementation of the procedure are addressed in order to improve the accuracy and efficiency of the methods. A comprehensive study of modelling the shot noise Cox process on real general insurance claims data is undertaken where practical issues inherent in insurance claims data such as impact of insurer exposure are addressed.

ACKNOWLEDGEMENTS

Firstly, I wish to express my deepest appreciation and gratitude to my supervisors Dr Benjamin Avanzi and Dr Bernard Wong for their constant guidance and support throughout the year. Your knowledgeable advice, patience and constant encouragement have given me the skills and the confidence to complete this Honours project. Thank you to you both for an enriching and enjoyable Honours year. We made it!

I am also very grateful for the financial support provided by the Donors of EJ Blackadder Honours Scholarship and to the Australian Prudential Regulation Authority through the Brian Gray Scholarship. The financial support you have provided has been invaluable as it allowed me to dedicate my time to my studies.

To the staff of the School of Risk and Actuarial studies, I wish to thank you for your advice and support through my Honours year as well as your guidance throughout my five years of study within the school. A very special thank you goes to my fellow Honours students; Vincent, Andy, Daniel and Qiming, and PhD students for all the good times and for keeping me smiling despite some tough times through this year. I wish you all the best in your future endeavours whether it be in the workforce or in pursuit of further academic studies.

To all my friends, thanks for putting up with me this year. Finally I would like to thank my parents Kevin and Jane and my sister Tina for their endless patience and understanding. Your constant support has kept me motivated and I will always be indebted to you.

CONTENTS

CHAPTER 1

INTRODUCTION

The primary objectives for general insurance companies include ensuring they are able to meet their financial obligations to policyholders while being able to deliver profits for their shareholders. In order to meet these objectives, products need to be priced to accurately reflect the risk the insurer has undertaken while reserves and capital need to be held to ensure a certain level of safety over a certain time horizon. Hence, developing accurate methods of modelling the number of claims incurred over a specified time interval is important as this quantity is directly linked to the capital and reserving requirements for an insurer.

The classical method for modelling claim counts is the distributional approach, where it is assumed that the number of claims over a certain length of time follows a particular discrete distribution. The popular distributions used in this method include the Poisson and negative binomial distributions. Despite its simplicity and accessibility to practitioners, this approach has several shortcomings. For instance, in order to reliably model the number of claims per year using the distributional approach, historical data on the aggregate yearly number of claims for several decades would be required. The insurer would usually have claims data at a finer level than aggregate number of claims per year which is not utilised with this approach. This means that the distributional approach does not utilise the data insurers have efficiently.

These shortcomings motivate the use of stochastic processes to model claims counts. In particular, count processes are used as they are able to capture the evolution of claim frequency over time as well as utilise data more efficiently. A popular process in modelling claim counts over time is the Poisson process. This process, however, implies that the claim rate for the insurer does not change over time whereas in reality, this rate at which claims are incurred changes according the random variations in environmental and financial conditions. For example, in periods of extreme weather such as heavy rain may lead to greater number of road accidents and hence lead to a significantly higher number of claims for the insurer during that time. The inhomogeneous Poisson processes, which allows for intensity to change deterministically with time, does cover some these shortcomings. Unfortunately, the homogeneous Poisson process are still unable to capture the randomness inherent in the changes in environmental conditions and their impact upon claim frequency for the insurer.

1.1 Research Motivation

The primary motivation for this research is to develop a robust modelling framework for insurance claim counts which is able to capture the impact of random nature of environmental and economic conditions on the claim intensity. This can be done so by allowing the claim intensity of the Poisson process to also be a stochastic process. This forms what is known as the doubly stochastic Poisson process, or the Cox process. The idea was first contrived by Cox (1955) with extensive applications in physical sciences such as in Diggle and Diggle (1983) and Møller (2003). It has been also been applied in credit risk modelling in Lando (1998) and mortality modelling in Biffis (2005). It is only recently that the idea has been considered in a general insurance context to model claim counts and inter-arrival times. From an insurer's perspective, allowing for stochasticity in the claim intensity means that the counts process can capture the randomness of variations in claim intensity, giving a more realistic and robust model for claim counts. As this model is able to capture the dynamics of claim rates over time more accurately, it will potentially allow insurers to accurately calculate reserves and capital required as well as calibrate their pricing strategy.

Despite the fact the model has been empirically applied in financial areas such as modelling default rates in credit risk and mortality rates in a life insurance context, the majority of the current literature for doubly stochastic Poisson processes in a general insurance context focuses heavily on development of theoretical results. There has been no empir-

ical studies of doubly stochastic Poisson process and recent applications of the model to pricing insurance contracts has been done under a purely theoretical setting. When using doubly stochastic Poisson processes to model claim counts in an insurance setting, a variety of practical issues such as modelling an unobserved intensity process need to be dealt with. Through exploring, comparing and calibrating alternative modelling methods through a case study on real data, the research aims to address these issues in order to make the doubly stochastic Poisson process more accessible for practitioners to use for pricing, reserving and solvency calculations.

1.2 Thesis Outline

The thesis is structured in the following way. Chapter 2 provides a background of various stochastic processes used to model counts data and review general features of doubly stochastic Poisson processes. Methods used to fit doubly stochastic Poisson processes to fit counts data in various fields outside of insurance are discussed in more detail.

Theoretical features of the shot noise Cox process are explored in Chapter 3. Distributional properties and moments of both the shot noise intensity with exponential decay and the shot noise Cox process are derived and discussed as they will be utilised in the model fitting procedure. Qualitative criteria for selecting an appropriate intensity process for the doubly stochastic Poisson process will also be included which motivated the selection of the shot noise process for the research.

In Chapter 4, we discuss the issues involved in fitting shot noise Cox processes to counts data which mainly revolve about modelling an unobserved intensity process. We propose and discuss in detail two different methods for estimating parameters via filtering of the intensity which are based on widely used techniques for modelling latent variables. Both heuristic and formal goodness of fit tests are also provided. The methods described are calibrated and validated through a simulation study conducted in Chapter 5. The simulation study is done for various parameter sets where the performance of both methods proposed in Chapter 4 are analysed and compared.

The thesis concludes in Chapter 6 with a summary of the contributions of the thesis in the context of the current theoretical literature and application of claim count modelling. Suggestion for areas of further research are also included.

CHAPTER 2

LITERATURE REVIEW

This chapter provides a literature review on key methods in relation to modelling insurance claim frequency over time. Section 2.1 will provide a brief overview of properties of count processes which have been used extensively to model counts data. In particular, the shortcomings of the homogeneous Poisson process are highlighted in order to motivate the use of doubly stochastic Poisson processes. Section 2.2 focuses on the theoretical framework of doubly stochastic Poisson processes which includes its definition and key distributional properties. The key features of the intensity processes will be outlined in Section 2.3, where two specific affine processes will be analysed. Current methodology of fitting stochastic processes and model selection procedures will be explored in Section 2.4. In particular, different methods for filtering out latent variables will be discussed. As a potential application of the doubly stochastic Poisson processes lies in reserve calculations, the current methods and techniques in stochastic reserving will also be reviewed in Section 2.5.

2.1 Background of count processes

This section aims to provide a general overview of features in count processes as well as major ones which have been used to model counts data. The shortcomings of the described processes are also outlined and they are used to motivate the development of

doubly stochastic Poisson processes.

Traditionally, discrete distributions have been used to model counts data over a specified time frame where the parameters of the distribution is usually estimated through means of maximised likelihood. In this approach, the number of claims over a length of time (such as 1 year) is a random variable X with a discrete distribution. Popular distributions include the Poisson distribution with parameter λ, whose probability distribution function and first two central moments are stated by Klugman et al. (1998) as:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, ... \tag{2.1.1}$$
$$\mathbb{E}[X] = \lambda, \quad \text{Var}(X) = \lambda$$

as well as the negative binomial distribution with parameters n and p, whose probability distribution function and first two central moments are also stated by Klugman et al. (1998) as:

$$P(X = k) = \frac{(n+k-1)!}{k!(n-1)!}(1-p)^n p^k, \quad k = 0, 1, 2, ... \tag{2.1.2}$$
$$\mathbb{E}[X] = \frac{np}{1-p}, \quad \text{Var}(X) = \frac{np}{(1-p)^2}$$

This method, despite its simplicity, does not capture how the counts evolve over time. In order to do so, a *stochastic process* $\{X(t), t \geq 0\}$ will be considered as an alternative model. In an insurance context, using stochastic processes to model claim frequencies allows one to take advantage of knowledge of claim arrival times from data in order to produce more accurate calculation of reserves over time as well as accurately capture dependencies between lines of business over time.

Denote $X(t) - X(s)$ where $t > s$ as the *increment* of the stochastic process from s to t. A stochastic process is said to have *stationary increments* if the distribution of the increment $N(t) - N(s)$ is dependent only on the length of the interval $t - s$. This property is expressed in the following way by Denuit et al. (2007):

$$\Pr(X(t) - X(s)) = \Pr(X(t-s) - X(0)) = f(h)$$

where $f(\cdot)$ is a function of $h = t - s$. Two intervals (a, b) and (s, t) are said to be be *disjoint* if the intersection of the two intervals is empty, or in other words, $(a, b) \cap (s, t) = \emptyset$. A stochastic process is said to have *independent increments* is for any set of disjoint inter-

vals, the increments over these intervals are independent. In other words, for any 2 disjoint intervals (a, b) and (s, t), the increments $X(t) - X(s)$ and $X(b) - X(a)$ are independent.

A stochastic process is said to have the Markov property if the conditional distribution of future projections of the process given the whole history of the process is the same as the conditional distribution of future projections of the process given most recent information on the current state. This is expressed mathematically as:

$$\Pr(X(t)|X(u), 0 \leq u \leq s) = \Pr(X(t)|X(s))$$

A special type of stochastic process called a *counting process* is used to model event frequency over time. A count process is defined by Klugman et al. (1998) to be a non-negative integer valued stochastic process $\{N(t), t \geq 0\}$ and for any $t > s$, $N(t) \geq N(s)$ almost surely.

2.1.1 Poisson Process

One of the first count processes used for modelling counts data is the Poisson process. It's popularity is sourced from its various desirable theoretical properties. This is defined in Mikosch (2009) in the following way:

Definition 2.1. *A stochastic process $\{N(t), t \geq 0\}$ is said to be a Poisson process if the following conditions hold:*

1. *The process starts at zero: $N(0) = 0$ almost surely.*
2. *The process has independent increments for any $t_i, i = 0, ..., n$ and $n \geq 1$ such that $0 = t_0 < t_1 < ... < t_n$, the increments $N(t_i) - N(t_{i-1}), i = 1, ..., n$ are mutually independent.*
3. *There exists a non-decreasing right continuous function $m : [0, \infty) \to [0, \infty)$ with $m(0) = 0$ such that the increments $N(t) - N(s)$ for $0 < s < t < \infty$ have a Poisson distribution $Pois(m(t) - m(s))$ where m is denoted as the mean value function of $N(t)$.*
4. *With probability 1, the sample paths of the process $N(t)$ are right-continuous for $t \geq 0$ and have limits from the left for $t > 0$. This implies that $N(t)$ has cadlag sample paths.*

Note that combining the first and third properties implies that $N(t) = N(t) - N(0) \sim$

$Pois(m(t))$. If m is continuous, another function $\lambda(t)$ is denoted as the *intensity function* such that:

$$m(t) = \int_s^t \lambda(y)dy, s < t \tag{2.1.3}$$

For a Poisson process, the intensity function can be thought of changes in flow of time, where higher intensities corresponds to speeding up of arrival times and lower intensities corresponds to the slowing down of arrival times.

2.1.1.1 Homogeneous Poisson process

Out of all the Poisson processes, the most popular one is the homogeneous Poisson process, denoted as $\{\tilde{N}(t), t \geq 0\}$ where the intensity function is a constant $\lambda(t) = \lambda$. A *standard homogeneous Poisson process* has an intensity of $\lambda = 1$. The homogeneous Poisson process has the following definition based from the one in Mikosch (2009):

Definition 2.2. *The homogeneous Poisson process with intensity λ has the following properties:*

- *It has independent and stationary increments.*
- *For small increments $\Delta t > 0$, then:*

$$\Pr(\tilde{N}_{t+\Delta t} - \tilde{N}_t = k) = \begin{cases} 1 - \lambda\Delta t + o(\Delta t) & \text{if } k = 0 \\ \lambda\Delta t + o(\Delta t) & \text{if } k = 1 \\ o(\Delta t) & \text{if } k \geq 2 \end{cases}$$

where $o(\Delta t)$ is a function such that $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$.

From the above definition, the probability distribution of increments of the homogeneous Poisson process is:

$$\Pr(\tilde{N}_t - \tilde{N}_s = k) = \frac{e^{-\lambda(t-s)}(\lambda(t-s))^k}{k!} \tag{2.1.4}$$

which has the probability generating function (from Denuit et al. (2007)):

$$G_t(z) = \exp(\lambda t(z-1)) \tag{2.1.5}$$

From the probability generating function, it can be inferred that the relationship between

the first 2 central moments of a Poisson process to be:

$$\mathbb{E}[\tilde{N}_t] = \text{Var}[\tilde{N}_t] = \lambda t \tag{2.1.6}$$

Denote T_i as the arrival time of the i^{th}. For a homogeneous Poisson process, it is assumed that the interarrival time for the i^{th} claim $T_i - T_{i-1}$ is exponentially distributed. This results from the assumption of stationary and independent increments of the homogeneous Poisson process.

The popularity of the homogeneous Poisson process stems from its simplicity in modelling and its range of very attractive statistical properties such as stationarity and independence of its increments. In fitting a homogeneous Poisson process, the only parameter which needs to be estimated is λ and since there exists a closed form for the probability distribution of increments, likelihood methods can be implemented in order to find an estimate of λ.

Due the stationarity and independence of increments for a homogeneous Poisson process, its relies on the assumption that intensity which claims arrive remain constant over time. This means that impact of changes in meteorological and safety conditions on the probability of an accident occurring. The assumption has been shown in various empirical studies to be an erroneous. In fact, Seal (1983) first identifies that there is little empirical evidence that the Poisson process is a good fit for insurance claim frequency data. Another major shortcoming of the homogeneous Poisson process is that it does not capture a phenomenon known as *overdispersion* which is prevalent in counts data, particularly in claim frequency in an insurance setting.

2.1.1.2 Inhomogeneous Poisson Process

An alternative model to the homogeneous Poisson process is the inhomogeneous Poisson process, which follows Definition 1.1 where the intensity function $\lambda(t)$ is a non-constant deterministic function with respect to time. The deterministic intensity function aims to capture inhomogeneity in arrival time of events due to a variety of factors. In an insurance context, inhomogeneity in claim arrival times arises due to seasonal weather effects resulting in clusters of accidents occurring in particular times in the year. The inhomogeneous Poisson process is defined in Denuit et al. (2007) as:

Definition 2.3. *The inhomogeneous Poisson process with intensity function $\lambda(t)$ has the*

following properties:

- *It has independent increments.*
- *For small increments $\Delta t > 0$, then:*

$$\Pr(\tilde{N}_{t+\Delta t} - \tilde{N}_t = k) = \begin{cases} 1 - \lambda(t)\Delta t + o(\Delta t) & \text{if } k = 0 \\ \lambda(t)\Delta t + o(\Delta t) & \text{if } k = 1 \\ o(\Delta t) & \text{if } k \geq 2 \end{cases}$$

where $o(\Delta t)$ is a function such that $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$.

Mikosch (2009) used the inhomogeneous Poisson process to model accident claims frequency on the Danish fire insurance data and showed the fit is a significant improvement from using a homogeneous Poisson process. It can be argued, however, that some of the factors which results in the inhomogeneity in arrival times of events are random by nature. Once again, in an insurance context, weather and its impact on claim arrival times is considered to be stochastic. This will result in stochasticity in the intensity of the process which is not fully captured with a deterministic intensity function. Although the inhomogeneous Poisson process allows the stationary increments property to be relaxed, it still doesn't account for overdispersion of counts data as the mean and variance of the increments is still assumed to be equal.

2.1.2 Overdispersion

One of the main properties of both the homogeneous and inhomogeneous Poisson process is that over a specified time interval, the mean of the process is the same as the variance of the process. Denuit et al. (2007) describes the phenomenon overdispersion to be when the variance of the process exceeds the mean. This phenomenon arises due to various reasons. One of these is the heterogeneity of the population resulting in difference in the risk of claim for each policyholder. This fact is not accounted for when assuming a constant intensity λ.

Hougaard et al. (1997) attempts to address this by introducing a different class of Poisson processes known as the *mixed Poisson process*. The intensity λ is assumed to be a non-negative random variable with a known distribution such as Gamma and Inverse Gaussian. An interpretation for randomizing the intensity is that the riskiness of the policyholders is captured through an assumed distribution of the population.

Although this method does account for overdispersion of claim frequencies, the assumed distribution of the population does not change over time as the population changes. Also, the mixed Poisson process does not account for common changes in the claim intensity of each policyholder due to seasonal effects. In the study of accident statistics by Bartlett (1986) suggests that although there is strong evidence for heterogeneity between drivers, there is also significant evidence which suggests changes in the claim intensity could be caused by shifts in environmental factors. A stochastic process would be more realistic in capturing random fluctuations in the intensity and this leads to the development of the doubly stochastic Poisson process.

2.2 Doubly stochastic Poisson processes

Doubly stochastic Poisson process, also known as the Cox process, was first introduced in Cox (1955) who considered stochastic variations in the intensity for a Poisson process. It is able to account for both overdispersion which is prevalent in counts data as well as capture dynamics in claim intensity resulting from seasonal trends and environmental factors. This section aims to give a detailed description of the theoretical framework for doubly stochastic Poisson processes and how certain statistical properties can be used in model fitting and selection.

Doubly stochastic Poisson processes have been applied extensively in physical sciences such as population dynamics in Diggle and Chetwynd (1991) and in rainfall modelling such as Rodriguez-Iturbe et al. (1987). They mainly focus on modelling clustered spatial point patterns which arrive in different rates according to the position on the plane. Temporal Cox processes have also been used extensively in finance and longevity modelling. In credit risk, Lando (1998) used the arrival time of the first event of a doubly stochastic Poisson process to represent default time of a company and hence determine probability of default. Biffis (2005) and Schrager (2006) extended the idea of modelling the first event arrival time of a doubly stochastic Poisson process with affine intensity to model survival probability.

More recently, theoretical developments have been made for using doubly stochastic Poisson processes in a non-life insurance setting. In ruin probability, for the following surplus

process:

$$U_t = u + ct - \sum_{j=1}^{N(t)} X_j$$

traditionally the claim arrive rate $N(t)$ was modelled using a homogeneous Poisson process to derive ruin probabilities. Björk and Grandell (1988) and Albrecher and Asmussen (2006) extended this framework by deriving estimates of ruin probabilities in the case where $N(t)$ is a doubly stochastic Poisson process. Dassios and Jang (2003) derived the price of catastrophe reinsurance contracts and derivatives under the assumption that claims arrive according to a doubly stochastic Poisson process.

2.2.1 Definitions

Consider a homogeneous Poisson process with intensity rate 1 $\{\tilde{N}_t, t \geq 0\}$ and a stochastic process $\{\Lambda_t, t \geq 0\}$ which is non-negative for all $t \geq 0$. Then Grandell (1976) defines the doubly stochastic Poisson process $\{N(t), t \geq 0\}$ as $N(t) \equiv \tilde{N} \circ \left(\int_0^t \Lambda(s)ds\right)$ which is a homogeneous Poisson process with a underlying stochastic subordinating process $\left(\int_0^t \Lambda(s)ds\right)$ which is the integral of the intensity process. We define the integrated process $M(t) = \int_0^t \Lambda(s)ds$ as the *mean value process*. It should be noted that the homogeneous Poisson process $\tilde{N}_t$ is independent of the subordinator $\left(\int_0^t \Lambda(s)ds\right)$ and hence independent of the intensity process Λ_t.

The characteristic functional of a stochastic process allows one to determine the distribution due to the bijective relationship between distributions and generating functionals. In his discussion paper of Cox (1955), Bartlett defines the characteristic functional of the doubly stochastic Poisson process to be:

$$\Phi_N(t)(\theta(t)) = \mathbb{E}_\Lambda \left[\exp\left(\int_0^t \Lambda(s)(e^{\theta(s)} - 1)dt\right)\right] = \Phi(e^{\theta(t)} - 1) \tag{2.2.1}$$

where $\Phi_{M(t)}$ is the characteristic functional of the mean value process $M(t) = \int_0^t \Lambda(t)dt$. Hence, an explicit expression for the characteristic functional for the intensity process is all that is required in order to determine an explicit expression for the characteristic functional of the counts process.

The probability generating functional is useful in obtaining probabilities for the doubly

stochastic Poisson process. Cox and Isham (1980) defines this to be:

$$G(\theta(t)) = \mathbb{E}_{\Lambda}\left[\exp\left(\int_{-\infty}^{\infty} \Lambda_t(1-\theta(t))dt\right)\right] \tag{2.2.2}$$

The probability distribution function of the increments of a doubly stochastic Poisson process can then be determined from the above. Grandell (1976) defines the distribution function of the doubly stochastic Poisson process as follows:

Definition 2.4. *(Probability distribution of the increment of a doubly stochastic Poisson process)*
$\{N(t), t \geq 0\}$ *is a doubly stochastic Poisson process with underlying intensity process* $\{\Lambda(t), t \geq 0\}$ *if for a realised sample path of the intensity process on the interval* (s,t)*, the probability distribution of the increment* $N(t) - N(s)$ *is given by:*

$$\Pr(N(t) - N(s) = n | \Lambda(u) = \lambda(u), s \leq u \leq t) = \frac{\left(\int_s^t \lambda(u)du\right)^n e^{-\left(\int_s^t \lambda(u)du\right)}}{n!}$$

where the intensity process is non-negative for all t*.*

Hence, by taking expectations with respect to the intensity process Λ_t, the unconditional probability distribution of $N(t)$ can be determined to be:

$$\begin{aligned}\Pr(N(t) - N(s) = n) &= \mathbb{E}_{\Lambda}[P(N(t) - N(s) = n | \Lambda(u) = \lambda(u), s \leq u \leq t)] \\ &= \mathbb{E}_{\Lambda}\left[\frac{\left(\int_s^t \Lambda(u)du\right)^n e^{-\left(\int_s^t \Lambda(u)du\right)}}{n!}\right]\end{aligned} \tag{2.2.3}$$

Both the inhomogeneous and homogeneous Poisson process are special cases of the doubly stochastic Poisson process, where the intensity process is a purely deterministic function with respect to time (for inhomogeneous Poisson process) or a constant (for homogeneous Poisson process). It can be seen that for $0 \leq a \leq b$, the distribution of $N(t)$ conditional on a sample path of Λ_t will be the same as that of the inhomogeneous Poisson process. Hence it can be said that the disjoint increments of the doubly stochastic Poisson process are conditionally independent. Like the inhomogeneous Poisson process through, the increments may not necessarily be stationary. The mixed Poisson process described in Hougaard et al. (1997) is also a special case of the doubly stochastic Poisson process, where the intensity process is just a random variable.

Grandell (1976) provides an alternative definition based on the interarrival times of the

process. Consider $\tilde{T}_i$ and T_i which are the i^{th} interarrival times of the $\tilde{N}(t)$ and $N(t)$ respectively. For all i:

$$T_i \equiv M^{-1}(\tilde{T}_i)$$
$$\text{where } M^{-1}(x) = \sup(y; M(y) \leq x) \tag{2.2.4}$$

This definition of the doubly stochastic Poisson process is used to model time to default for credit risky assets as demonstrated in Lando (1998), where the arrival time of the first event can be thought of as the default time. The arrival time of the first event of a doubly stochastic Poisson process is then extended to model mortality as demonstrated in Biffis (2005) and Schrager (2006). Cox and Isham (1980) then presents another definition the doubly stochastic Poisson process similar to the definitions of the homogeneous and inhomogeneous Poisson processes stated in the previous section:

$$\frac{\Pr(N_{t+\Delta t} - N(t) > 0 | \Lambda(s) = \lambda(s), 0 \leq s \leq t)}{\delta} = \lambda(t) \tag{2.2.5}$$

$$\Pr(N_{t+\Delta t} - N(t) > 1) = o(\delta) \tag{2.2.6}$$

where Δt is a small increment in time t and $\lambda(t)$ is the realisation of the intensity process at time t.

2.2.2 Thinning

One of the attractive features of the homogeneous Poisson process is that it can be decomposed into a sum of independent homogeneous Poisson processes whose intensity is a proportion of the intensity for the aggregate process. This is known as thinning of the Poisson process. This is desirable as it allows for categorisation of events into different types. From a general insurance perspective, this means the overall claim process for the company can be decomposed into claim processes for different lines of businesses.

Like the Poisson process, the doubly stochastic Poisson process can also be thinned. Karr 1985 defines the thinned Cox process in the following:

Definition 2.5. *(The thinned doubly stochastic Poisson process)*
Consider the doubly stochastic Poisson process in the form:

$$N(t) = \sum_i \mathbf{1}_{\{T_i \leq t\}}$$

Let U_i be a random variable which is independent to $N(t)$ where $\Pr(U_i = 1) = p$. Then

the thinned doubly stochastic Poisson process $N'(t)$ is of the form:

$$N'(t) = \sum_i U_i \mathbf{1}_{\{T_i \leq t\}}$$

Karr 1985 also provided statistical properties of the thinned process $\{N'(t), t \geq 0\}$ which lines the thinned process with the original process:

$$\Phi_{N'(t)}(f(t)) = \Phi_{N(t)}(-\log(1 - p + pe^{-f(t)})) \tag{2.2.7}$$

$$P(N'(t) = 0) = \Phi_{N(t)}(-\log(1 - p)) \tag{2.2.8}$$

$$\mathbb{E}[N'(t)] = p\mathbb{E}[N(t)] \tag{2.2.9}$$

which are analoguous to the results for the homogeneous Poisson process. Karr 1985 also showed that $N(t) - N'(t)$ and $N'(t)$ are independent doubly stochastic Poisson process processes with intensity processes $(1 - p)\Lambda(t)$ and $p\Lambda(t)$ respectively.

2.2.3 General statistical properties

Denote the filtration at time t of the possible paths of the underlying intensity process $\mathcal{F}_t^{\Lambda}$. Cox and Isham (1980) then defines the first 2 conditional central moments to be:

$$\mathbb{E}[N(t) - N(s)|\Lambda(u) = \lambda(u), , s \leq u \leq t] = \int_s^t \lambda(u)du \tag{2.2.10}$$

$$\mathrm{Var}[N(t) - N(s)|\Lambda(u) = \lambda(u), s \leq u \leq t] = \int_s^t \lambda(u)du \tag{2.2.11}$$

This is analogous to the inhomogeneous Poisson process where the first 2 central moments are equal. Then by law of iterated expectation, the unconditional moments can be determined to be:

$$\mathbb{E}[N(t) - N(s)] = \int_s^t \mathbb{E}[\Lambda(u)]du \tag{2.2.12}$$

$$\mathrm{Var}[N(t) - N(s)] = \int_s^t \mathbb{E}[\Lambda(u)]du + \mathrm{Var}(\int_s^t \Lambda(u)du) \tag{2.2.13}$$

where we can see that the extra variance term of the stochastic integral of the underlying process $\int_a^b \Lambda du$ accounts for overdispersion in claim counts. This extra term in the variance arises from the uncertainty in the estimate of the claim intensity which increases the variability of the number of events in the time interval $[a, b]$.

Many of the statistical properties of the doubly stochastic Poisson process is dependent on the stastitical properties of intensity process. Firstly, Grandell (1976) noted that $N_A(t) \equiv N_B(t)$ if and only if $\Lambda_A(t) \equiv \Lambda_B(t)$ which gives the uniqueness of the intensity process for each doubly stochastic Poisson process. This implies that the doubly stochastic Poisson process is completely characterised by the intensity process. Grandell (1976) also noted that the doubly stochastic Poisson process $N(t)$ has stationary increments *if and only if* the intensity process $\Lambda(t)$ has stationary increments. This is apparent when considering homogeneous Poisson process, which is a very special cases of the doubly stochastic Poisson process, where the constant intensity is clearly stationary. As mentioned previously, the doubly stochastic Poisson process has conditionally independent increments, however, the increments may not be unconditionally independent. In fact, Grandell (1976) found be shown however, that the correlation of number of events in two different intervals (s,t) and (a,b) to be given by the equation below:

$$\text{Cov}(N(t)-N(s),N(b)-N(a)) = E\left[\int_c^d \Lambda(u)du\right] + \text{Cov}\left(\int_s^t \Lambda(u)du, \int_a^b \Lambda_u du\right) \quad (2.2.14)$$

where $(c,d) = (s,t) \cap (a,b)$ is the overlap of the two intervals (s,t) and (a,b). Suppose the two intervals were disjoint, then the above equation reduces down to:

$$\begin{aligned}\text{Cov}(N(t)-N(s),N(b)-N(a)) &= \text{Cov}\left(\int_s^t \Lambda(u)du, \int_a^b \Lambda(u)du\right) \\ &= \text{Cov}\left(M(t)-M(s),M(a)-M(b)\right) \qquad (2.2.15)\end{aligned}$$

Hence the doubly stochastic Poisson process $N(t)$ will have uncorrelated increments *if and only if* the mean value process $M(t) = \int_0^t \Lambda(u)du$ has uncorrelated increments.

Consider the compensated Cox process $C(t) = N(t) - \int_0^t \Lambda(s)ds$, which is analogous to the compensated Poisson process. A consequence of the correlation structure is that $C(b) - C(a)$ and $\int_s^t \Lambda(u)du$ are uncorrelated. Furthermore, if the intervals (a,b) and (s,t) are disjoint, then $C(b) - C(a)$ and $C(t) - C(s)$ are uncorrelated.

2.2.4 Some examples

Much of the properties of the doubly stochastic Poisson process are largely dictated by the characteristics and structure of the underlying intensity and mean value processes. This implies that the doubly stochastic Poisson process is completely characterised by the underlying intensity process. Thus the problem of modelling claim counts using dou-

bly stochastic Poisson processes reduces down to modelling the claim intensity using a stochastic process.

Aside from the trivial cases where $\Lambda(t)$ is a constant or deterministic, resulting in either a homogeneous or inhomogeneous Poisson process, another special case of the doubly stochastic Poisson process is to consider the intensity process as a continuous time Markov chain driven by another continuous time Markov chain $B(t)$ with finite state space S which satisfies the following differential equations:

$$p_k'(t) = -q_k p_k(t) + \sum_{i \neq k} p_i(t) q_{ik}$$

where $p_k(t) = \Pr(B(t) = k)$, q_{ik} and q_k are elements from the generator matrix such that:

$$q_{ik} = \lim_{h \to 0} \frac{P(B(t+h) = k | B(t) = i)}{h}$$
$$q_i = \sum_{k \neq i} q_{ik}$$

So when $B(t) = k$, we have $\Lambda(t) = \lambda_k$. This process is considered in Rudemo (1972) and Neuts (1971) with applications in queuing and replacement theory. As the underlying Markov process $B(t)$ has finite states, the intensity process $\Lambda(t)$ will also have finite states meaning that the values which the intensity process can take will be limited. This is not realistic in insurance, where the claim arrival rate should be able to reach any non-negative values. In order to increase the set of values the intensity can take, the number of states will need to be increased. In relation to model fitting, one extra state would lead to one extra parameter needed to be estimated and hence for a large state space, this model is not parsimonious.

The intensity process of the doubly stochastic Poisson process does not necessarily need to have a parametric form. Bouzas et al. (2009) and Bouzas et al. (2010) uses functional principal component analysis to fit the intensities of doubly stochastic Poisson processes via piecewise cubic interpolation. The advantage of non-parametric intensities is that it does not require any distributional assumptions on the intensity and hence it is more . However forecasting future claim arrivals with a non-parametric intensity constructed through interpolation may be restricted. As this research focuses on fitting insurance claims data where applications of the model includes forecasting, we shall focus on parametric models for the intensity process.

2.3 Affine intensity processes

Although there are infinitely many stochastic processes which could be used to model the intensity, in order to determine analytical expressions for key statistical properties such as distributions and moments of the doubly stochastic Poisson process, the intensity process needs to be analytically tractable.

A key criteria for the intensity process $\Lambda(t)$ is that $\Lambda(t) \geq 0$ for all $t > 0$ with probability 1 . This is necessary in the insurance context as the rate claims are incurred should be positive. Also it ensures that the mean value process $\int_0^t \Lambda(s)ds$ is monotonically increasing over time, which is once again realistic from an insurer's perspective.

One family of stochastic processes to consider is the affine intensity processes. The main advantage of affine jump diffusion process lies in its analytical tractability which has been utilised in modelling default times in credit risk in Lando (1998) and mortality in Biffis (2005). This section aims to give a general overview of affine processes and distributional properties for two particular affine processes which are also positive with probability 1; the shot-noise process and the Cox-Ingersoll-Ross process. Other potential intensity processes will be reviewed as well.

2.3.1 Definition

The definition from Biffis (2005) of an n-dimensional affine jump diffusion processes is given below:

Definition 2.6. *Consider the probability space* $(\Omega, \mathcal{F}, \mathbb{P}$. *An* n*-dimensional affine jump diffusion process* Λ_t *is a* $\mathcal{F}$*-Markov process which is the solution to the following stochastic differential equation:*

$$d\Lambda(t) = \mu(t, \Lambda(t))dt + \sigma(t, \Lambda(t))dB(t) + dS(t) \tag{2.3.1}$$

where $B(t)$ *is a standard Brownian motion and* $S(t)$ *is a point process with jump intensity* $\{\nu(t, \Lambda_t), t \geq 0\}$. *In particular, the drift* $\mu(t, \Lambda_t)$, *instantaneous covariance matrix* $\sigma(t, \Lambda_t)\sigma(t, \Lambda_t)^T$ *and jump intensity* $\kappa(t, \Lambda_t)$ *are affine functions of* Λ_t.

Another desirable quality for the underlying intensity is mean reversion. In an insurance context, this implies that the claim rate has a tendency to move towards an average rate

over time. This implies that any shifts from this average are temporary shocks caused from seasonal effects and will be absorbed over time. Mean reversion of the intensity also ensures that the mean value process does not diverge, or in other words, $\int_0^t \Lambda_t < \infty$. In particular, a basic affine jump diffusion process with mean reversion will satisfy the following stochastic differential equation stated in Bluhm et al. (2003):

$$d\Lambda(t) = \kappa(\theta - \Lambda(t))dt + \sigma\sqrt{\Lambda(t)}dB(t) + dS(t) \tag{2.3.2}$$

where κ is the rate of mean reversion.

Noting that one of the necessary conditions for the intensity process $\Lambda(t)$ is that $\Lambda(t) \geq 0$ for all t and that mean reversion is desirable, two main processes which hold both of these properties are the shot-noise process and the Cox-Ingersoll-Ross process.

2.3.2 Shot noise process

One of the more well known shot noise processes is where the spikes in the intensity decays exponentially. This is also known as the classical shot noise process and arrives from letting the function . Cox and Isham (1980) defines the classical shot noise process which is the explicit solution to the above equation as follows:

$$\Lambda(t) = \Lambda(0)e^{-kt} + \sum_{j=1}^{J(t)} X_j e^{-k(t-\tau_j)} \tag{2.3.3}$$

where τ_j are arrival times from the homogeneous Poisson process $J(t)$ with intensity ρ and $\{X_j\}$ are independent and identically distributed non-negative jump sizes with density g. The term $k > 0$ determines the rate at which the process decays exponentially from jumps. These aspects are characterised in Figure 3.1, which shows a sample path of the shot noise process. It can be seen that the stochasticity of this process lies purely with the jumps, where the exponential decay of these jumps are completely deterministic.

An interpretation of the above in an insurance context is analogous to the one given by Dassios and Jang (2003), where certain environmental or financial events will trigger an increase in the rate at which people claim. The timing of these events is given by $\tau_1, ..., \tau_j$ and the severity of the impact of these events is given by $X_1, ..., X_j$. Eventually, the impact of these events will recede and the rate at which this occurs is determined by the constant k. In Dassios and Jang (2003), this was used to price stop loss reinsurance contracts,

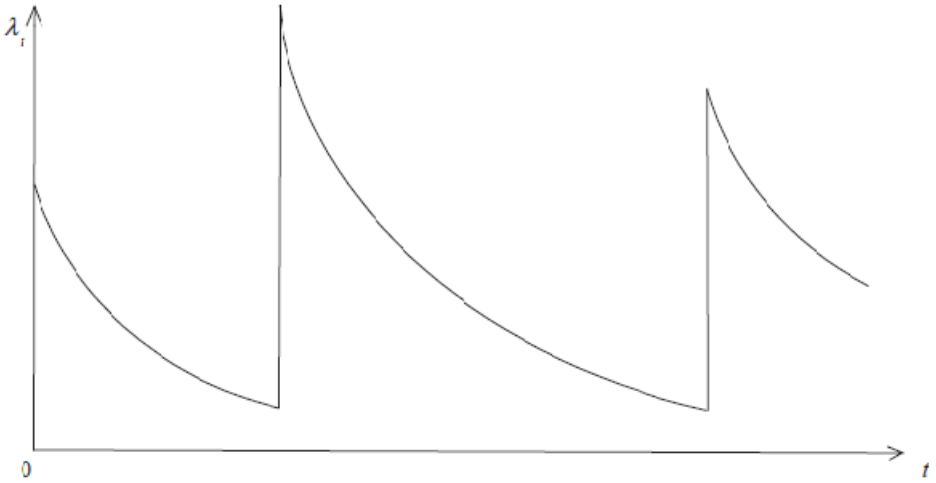

Figure 2.1: Illustration of sample path of the shot noise process

where the jump times are based on periods where natural disasters such as earthquakes occur.

2.3.3 Cox-Ingersoll-Ross process

The Cox-Ingersoll-Ross process from Cox et al. (1985) is the solution to the following stochastic differential equation:

$$d\Lambda(t) = \kappa(\theta - \Lambda(t))dt + \sigma\sqrt{\Lambda(t)}dB_t \tag{2.3.4}$$

This can be seen as a basic affine diffusion process without a jump component. Consequently the process will be continuous for all t. In the long run, this process reverts to the mean denoted as θ where κ is the rate of the adjustment. In particular, given the condition $\sigma^2 \leq 2\kappa\theta$, then the process $\Lambda(t)$ will always be positive. Cox et al. (1985) also notes that the distribution of the process is non-central chi-squared. A diagram of the sample path is shown in Figure 2.2.

2.3.4 Other affine processes

Other potential affine processes have been considered for various applications in finance and physical sciences. Basu and Dassios (2002) prices stop-loss reinsurance contracts where the underlying intensity process considered is a log-normal process $\Lambda(t) = ce^{\sigma Y(t)}$ such that $Y(t)$ is a Gaussian process. In particular, the Gaussian process considered in

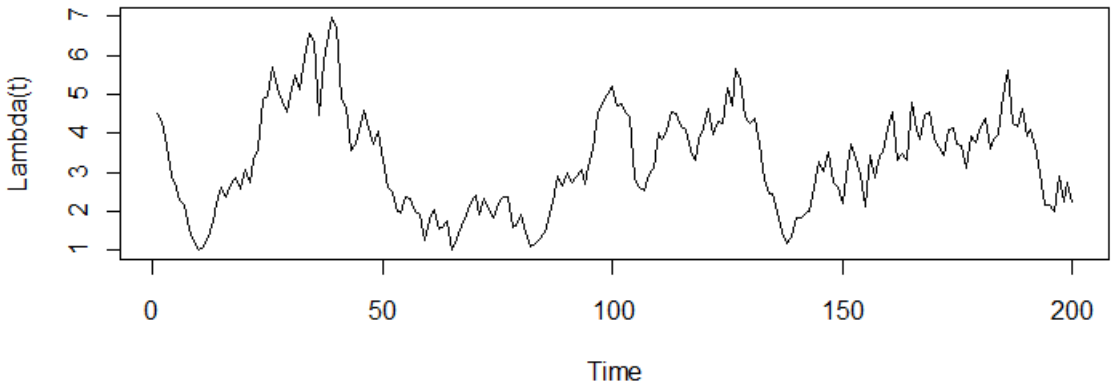

Figure 2.2: Illustration of sample path of the Cox-Ingersoll-Ross process

the paper is the Ornstein-Uhlenbeck process with the stochastic differential equation:

$$dY(t) = -aY(t)dt + dB(t)$$

Another process which has been used is the Hawke's process, which is a generalised version of the shot noise process. This is defined in Hawkes (1971) to be:

$$\Lambda(t) = \Lambda(0) + \sum_{\tau_i < t} \sum_{j=1}^{P} \alpha_j e^{-\beta_j (t-\tau_j)} \tag{2.3.5}$$

where the jumps occur according to a homogeneous Poisson process with rate ρ. The above equation implies that for each event, multiple jumps in the intensity would occur and for each jump, there is a different rate of decay. For $P = 1$, the Hawke's process corresponds to the shot-noise process.

2.4 Model fitting and selection

In order to model with a doubly stochastic Poisson process, the parameters of the intensity process needs to be estimated. This section aims to review methods to estimate parameters as well as overcoming the issues to fit the doubly stochastic Poisson process. The alternative non-parametric estimation techniques for the intensity will also be outlined. Several model selection methodology will also be reviewed to analyse the goodness of fit for stochastic processes.

One potential method in estimating parameters for the intensity process for a doubly

stochastic Poisson process is the maximum likelihood method. For a given intensity process, an expression for the probability distribution of the count process is obtained and hence the likelihood function can be derived. This was attempted in Konecny (1986) for the case where the stochastic intensity is driven by a Markov chain, but this was only for the case where there were two states in the Markov chain. There is, however, limited literature on the direct application of the maximum likelihood method on data, especially insurance claim data. This may be due to the difficulty in deriving closed form expressions for the likelihood function and verifying that estimators possess attractive properties such as asymptotic efficiency and normality, consistency. This is especially for non-stationary intensity processes, which are the processes the research is focused on, as the properties listed earlier was only shown to hold in Konecny (1986) for stationary intensity processes. The likelihood function may not necessarily be in a closed form and hence difficult to apply to real data. Another approach which is similar to the maximum likelihood method is the least squares estimation method. This method is adopted in Luciano and Vigna (2005) in modelling mortality rates with the Cox-Ingersoll-Ross process where the squared difference between the theoretical and actual mortality probability is minimised. The time to death is measured as the arrival time for the first event of a doubly stochastic Poisson process for a given intensity. Like the maximum likelihood estimation method, there also needs to be an analytic expression for the distribution of the doubly stochastic Poisson process.

The main practical issues which arises in the fitting procedure through maximum likelihood or least squares method directly. This is where the stochastic claim intensity process is unobservable in the counts data. Rather, only claim arrival times and frequency history is known. If the expression for moments and distribution of the doubly stochastic Poisson process cannot be attained in a closed form for a particular intensity, direct application of maximum likelihood or least squares would not be possible.

The filtering problem is akin to the issue of fitting unobservable stochastic volatility in financial markets, where this relationship is highlighted in both Barndorff-Nielsen and Shephard (1998) and Frey and Runggaldier (2001). Hence, methods of dealing with unobservable intensity process in the doubly stochastic Poisson process are similar to methods of dealing with unobservable stochastic volatility in finance. The main method used to estimate parameters for unobservable intensities involves *filtering* . This procedure involves the estimation of the conditional intensity process given an observed process and conducting statistical inference on the conditional intensity.

In physical sciences, the parameter estimation for doubly stochastic Poisson processes has been done using without likelihood estimation. Diggle et al. (1976) and Stoyan (1992) utilises second order summary statistics such as the Ripley's K-function and pair correlation function to apply the method of minimal constrast on clustered point patterns. These methods, however, have been developed for modelling spatial point process and hence are not applicable to model temporal data such as claim intensity. Hence we investigate the two popular filtering methods used in filtering latent variables: the Kalman filter and Markov Chain Monte Carlo methods.

2.4.1 Kalman Filter

One type of filtering involves using the Kalman-Bucy filter introduced in Kalman (1960). The rationale behind the Kalman filter is as follows. Consider the conditional mean square error of the estimate of the intensity $\hat{\Lambda}_t$ from the true intensity $\Lambda(t)$ denoted below:

$$MSE = \mathbb{E}[(\hat{\Lambda}_t - \Lambda(t))^2 | N(s), 0 \leq s \leq t]$$

Kalman (1960) noted that the above equation is minimised when the estimated intensity process $\hat{\Lambda}_t$ is the conditional expectation of the intensity $\mathbb{E}[\Lambda(t)|N(s), 0 \leq s \leq t]$. Hence the intensity process is adjusted recursively in order to minimise the conditional mean square error via several recursive differential equations which describes links the current state of the intensity process to the future state as well as the change in the mean square error.

The state equations for the Kalman filter from Kalman (1960) is as follows:

$$X_t = A_t X_{t-1} + B_t u_t + W_t \tag{2.4.1}$$

$$Y_t = H_t X_t + V_t \tag{2.4.2}$$

where W_t and V_t are zero mean Gaussian noise with covariance matrix Q_t and R_t respectively and u_t is another control input process. X_t is the unobserved process which is filtered, u_t refers to an external control variable which the user inputs and Y_t is the measured or observed process which is driven by the unobserved process. In a insurance context, the unobserved process would be the claim intensity which drives the observed claim counts process. After initialisation at time 0, there are two main steps to this algorithm: the time update and the prediction update. Then the general algorithm given in

Kalman (1960) is as follows:

1. Time update which predicts the value of X_t for the next time step:
   (a) Predict (a priori) the estimate for the unobserved process for next time step: $X_{t|t-1} = A_t X_{t-1} + B_t u_t$
   (b) Predict (a priori) the estimate for the covariance for next time step: $P_{t|t-1} = A_t P_{t-1} A_t^T + Q_t$
2. Prediction update which calibrates the value of X_t based on the measured process Y_t:
   (a) Update the covariance: $S_t = H_t P_{t|t-1} H_t^T + R_t$
   (b) Derive the Kalman gain: $K_t = P_{t|t-1} H_t^T S_t^{-1}$
   (c) Update the unobserved process estimate (a posteriori): $X_t = X_{t|t-1} + K_t(Y_t - H_t X_{t|t-1})$ where Y_t is the observed data.
   (d) Update the covariance estimate (a posteriori): $P_t = (I - K_t H_t) P_{t|t-1}$

The filter has been in finance used to estimate stochastic volatility in Rydberg and Shephard (1999) for stocks and in longevity for estimating mortality rates in Schrager (2006) which is modelled by an affine intensity process. The Kalman-Bucy filter, however, is applicable only for Gaussian processes in general. Both the Cox-Ingersoll-Ross and shot-noise processes are not a Gaussian process as it is a discontinuous process, and hence a direct application of the Kalman filter cannot be applied to the process. To overcome this issue, Dassios and Jang (2005) approximated the shot-noise process by a Gaussian process in order to develop the Kalman-Bucy filter, which is the continuous time extension of the Kalman filter, in order to price stop-loss reinsurance contracts. The Kalman-Bucy filter also relies on a assumption that the system is linear, which is not necessarily true for many non-Gaussian intensity processes.

2.4.2 Markov Chain Monte Carlo methods

An alternative class of methods which are used to filter out an unobserved intensity are Markov chain Monte Carlo methods which involves resampling a conditional intensity similar to that of moving between states in a Markov chain. The idea behind this is that eventually the sampled intensity will move towards a stationary distribution. Examples of Markov Chain Monte Carlo methods include the Metropolis-Hastings algorithm (see Metropolis et al. (1953) and Hastings (1970)), which is based on the random walk, and

the Gibbs sampler proposed in Geman and Geman (1984). The general algorithm in the Markov chain Monte Carlo method is:

1. Propose a type of transition.
2. Given the type of transition, generate a new state to transition into.
3. Accept the new state with a certain probability.

Given that the sampling comes from a Markov chain, a stationary distribution of the sampled intensity will be reached with enough iterations of the above procedure. Bayesian likelihood inference through expectation maximisation is then conducted on the sampled intensity in order to obtain parameter estimates. Markov chain Monte Carlo methods have been widely used as the method to model stochastic volatility in finance. Chib et al. (2002) modelled the stochastic volatily from S&P 500 daily returns via a Markov chain Monte Carlo method based on the Metropolis Hastings algorithm while Barndorff-Nielsen and Shephard (2001) considered non-Gaussian stchastic volatility models using Monte Carlo techniques with particle filters.

In particular for the shot-noise Cox process, Centanni and Minozzo (2006) proposes an extension of the MCMC filter known as the reverse jump Markov Chain Monte Carlo method in order to model the dynamics in the S&P 500 futures index. Monte Carlo simulation in this method is required as the system is non-linear. This method is based on the Metropolis-Hastings algorithm where the estimate for the intensity process is constantly updated based on the observed events. While the Metrotroplis-Hastings algorithm requires the intensity process to be Markovian, the reverse jump Markov Chain Monte Carlo method allows this assumption to be relaxed. For the simulated conditional intensity, expectation maximisation is used in order to predict the parameters for the model and Another advantage of this procedure is that the intensity process can be recursively updated as new information arrives. There are however issues on convergence and As this method is calibrated in this research to filter the unobserved claim intensity, more details about the procedure is provided in 5.2.1.

2.4.3 Goodness of fit tests for stochastic processes

As mentioned in previous sections, the intensity process is unique for every doubly stochastic Poisson process. So in order to test for the goodness in fit for the counts process, it is suffice to test the goodness of fit for the underlying intensity process. Hence, the hypothesis that the true intensity process $\Lambda(t)$ matches the fitted intensity process $\tilde{\Lambda}_t$ can

expressed in Dachian and Kutoyants (2008) to be:

$$\mathcal{H}_0 : \Lambda(t) = \tilde{\Lambda}_t \quad t \geq 0 \quad \text{vs} \quad \mathcal{H}_1 : \Lambda(t) \neq \tilde{\Lambda}_t \quad t \geq 0 \tag{2.4.3}$$

One of the classical test statistic used for this test is the Kolmogorov Smirnov test statistic from Kolmogorov (1933). Denote $M_t = \int_0^t \hat{\Lambda}_s ds$ as the mean value process. Then the count process version is expressed in Dachian and Kutoyants (2008) to be:

$$D_n = \frac{1}{\tilde{M}_t} \sup_{0 \leq s \leq t} |\hat{M}_t^{(n)} - \tilde{M}_t| \tag{2.4.4}$$

where $\tilde{\Lambda}_t$ denotes the fitted mean value process and $\hat{M}_t^{(n)}$ is the empirical mean value process. The empirical mean value process can be determined by considering the number of events in n sub-intervals of length t. The empirical mean value process can be found by:

$$\hat{M} = \frac{1}{n-t} \sum_{i=1}^{n-t+1} \#\text{ of events in the time interval } [i, i+t-1)$$

For the Kolmogorov Smirnov test statistic, it can be shown that for a large enough sample size, the statistic converges to the following:

$$\sqrt{n} D_n \to \sup_{0 \leq s \leq 1} |W_s|$$

where $\{W_s, 0 \leq s \leq 1\}$ is a Brownian bridge process. This can be used to find the approximate the probability $\Pr(D_n > \hat{D}_n)$ as well as approximating the alpha level rejection region R such that $\Pr(D_n > R) = \alpha$.

The other classical test statistic used to test the hypothesis is based on the Cramer-von Mises (or Anderson-Darling) test statistic in Anderson and Darling (1952). This is again expressed in Dachian and Kutoyants (2008) to be:

$$V_n^2 = \frac{n}{\tilde{M}_t^2} \int_0^t (\hat{M}_t - \tilde{M}_t)^2 d\tilde{M}_t \tag{2.4.5}$$

Like the Kolmogorov Smirnov test statistic, the Cramer-von Mises test statistic also convergences for large enough n in the following way:

$$V_n^2 \to \int_0^1 W_s^2 ds$$

Similarly to the Kolmogorov Smirnov case, this approximation can be used to determine p-values as well as rejection regions.

CHAPTER 3

FEATURES OF THE SHOT NOISE COX PROCESS

The purpose of this chapter is to analyse and develop statistical features of the shot noise Cox process which will be used in the model fitting procedure in the next chapter. In Section 3.1, we analyse the properties of the shot noise intensity process with exponential decay, in particular focusing on the derivation of the distribution and moments of $\Lambda(t)$. From Chapter 2, Grandell (1976) noted that doubly stochastic Poisson processes are uniquely defined by the intensity process. As we shall use the results in the modelling number of claims over a time increment $[s, t]$ in the following chapters, we focus the development of our results on the increments of the shot noise Cox process $N(t) - N(s)$. Hence we then can use these properties to derive further distributional properties and moments for the increments of the shot noise Cox process in section 3.2.

3.1 Shot noise intensity

In this section, we present the definitions and features of the shot noise intensity in an insurance context. The distributional properties, moments and time correlation structure of the shot noise intensity process are derived.

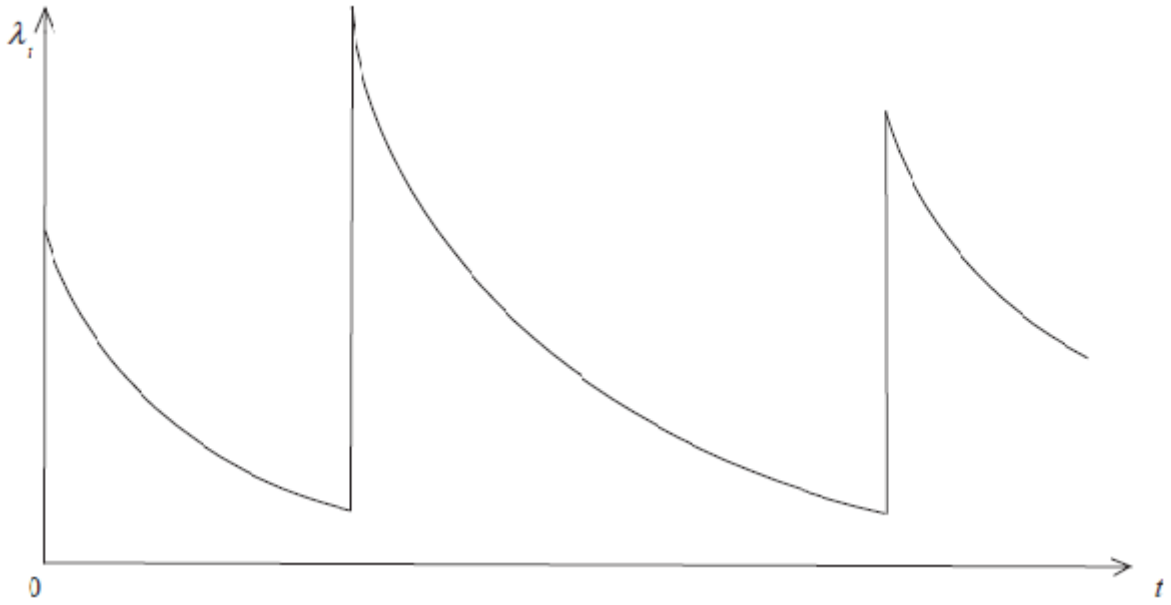

Figure 3.1: Sample path of the shot noise process

As in Cox and Isham (1980), the shot noise intensity process with exponential decay is defined in the following way:

Definition 3.1. *(Shot noise process with exponential decay) Consider the following stochastic differential equation of the process* $\{\Lambda(t), t \geq 0\}$*:*

$$d\Lambda(t) = -k\Lambda(t)dt + dK(t) \tag{3.1.1}$$

where $\{K(t), t \geq 0\}$ *is a point process with underlying Poisson process* $\{J(t), t \geq 0\}$ *with rate* ρ*. The solution of* (3.1.1) *is the shot noise process of the form:*

$$\Lambda(t) = \Lambda(0)e^{-kt} + \sum_{j=1}^{J(t)} X_j e^{-k(t-\tau_j)}, \quad t \geq 0 \tag{3.1.2}$$

where $\Lambda(0) > 0$*, the shot size* X_j*,* $j = 1, ..., J(t)$*, are independent and identically distributed exponential random variables with parameter* η *while* τ_j *represents the arrival time of the shots.*

A typical sample path of the shot noise process is shown in Figure 3.1. For $\{\Lambda(t), t \geq 0\}$ to be a valid underlying intensity process for the Cox process, it must be almost surely positive for all $t \geq 0$. This is due to the fact a counts process cannot have a negative intensity. For the shot noise intensity, since $\Lambda(0) > 0$ and the intensity $\Lambda(t)$ is the sum of

positive jumps with exponential decay with $\Lambda(0)$, it can be seen that $\Lambda(t)$ is almost surely positive for all time $t > 0$. Hence the shot noise process is a valid underlying intensity for the Cox process.

There is an alternative for (3.1.2). From Proposition 2.1.16 in Mikosch (2009) which states that:

$$S(t) = \sum_{j=1}^{J(t)} X_j e^{-k(t-\tau_j)} \stackrel{d}{=} \sum_{j=1}^{J(t)} X_j e^{-kt(1-U_j)} \stackrel{d}{=} \sum_{j=1}^{J(t)} X_j e^{-ktU_j} \tag{3.1.3}$$

where U_j are independent and identically distributed uniform $[0, 1]$ random variables which are independent of X_j and $J(t)$. The second inequality arrives from the fact for a uniform $[0, 1]$ random variable U_j has the same distribution as $1 - U_j$ to show that. This leads to the following reformulation:

Lemma 3.1. *(Alternative representation of shot noise intensity) The shot noise intensity can be expressed in the following form which has an equivalent distribution for all $t \geq 0$:*

$$\Lambda(t) = \Lambda(0)e^{-kt} + S(t) \stackrel{d}{=} \Lambda(0)e^{-kt} + \sum_{j=1}^{J(t)} X_j e^{-ktU_j} \tag{3.1.4}$$

where U_j are independent and identically distributed $[0, 1]$ random variables which are independent of X_j and $J(t)$.

The above representation of the shot noise process immediately follows from the fact that $\Lambda(0)$ is independent to $S(t)$. This result can be explained from the fact the ordered jump times τ_j in $S(t)$ is just the permutation of the jump times tU_j in ascending order. This result will be useful for derivations of results in this chapter.

The mean value process, which is the integrated intensity process, can be then be defined in the following:

Definition 3.2. *(Mean value process) The mean value process $\{M(t), t \geq 0\}$ for the shot noise Cox process is defined as follows:*

$$M(t) = \int_0^t \Lambda(t)dt = \frac{1}{k}\left[\Lambda(0)(1 - e^{-kt}) + \sum_{j=1}^{J(t)} X_j(1 - e^{-k(t-\tau_j)})\right] \tag{3.1.5}$$

where $M(0) = 0$. The process $M(t)$ is both increasing and non-negative.

The above definition is derived in the following:

$$\begin{aligned}
M(t) &= \int_0^t \Lambda(s)ds \\
&= \int_0^t \Lambda(0)e^{-ks} + \sum_{j=1}^{J(s)} X_j e^{-k(s-\tau_j)} ds \\
&= \frac{1}{k}\Lambda(0)(1-e^{-kt}) + \int_0^t \sum_{j=1}^{J(s)} X_i e^{-k(s-\tau_j)} ds \\
&= \frac{1}{k}\Lambda(0)(1-e^{-kt}) + \sum_{j=1}^{J(s)} X_i \int_{\tau_j}^t e^{-k(s-\tau_j)} ds \\
&= \frac{1}{k}\left[\Lambda(0)(1-e^{-kt}) + \sum_{j=1}^{J(t)} X_j (1-e^{-k(t-\tau_j)})\right]
\end{aligned}$$

where in the second last line arrives from the fact that for the shots component $X_j e^{-k(s-\tau_j)} = 0$ for $s < \tau_j$.

As mentioned in the previous chapter, (3.1.2) can be interpreted from an insurance perspective in the following way. In the occurrence of an event such as a hailstorm, the insurer will face a sudden increase in the number of claims incurred. The claim arrival rate will decrease as impact of the event diminishes with time, which is reflected by exponential decay, until the next event occurs. These events arrive randomly according to a Poisson process where the parameter ρ reflects the average number of events per time unit which will impact the claim arrival rate. The exponential random variables X_i with parameter η reflects the immediate impact of the event on the claim arrival rate such that the average jump size is $\frac{1}{\eta}$. The parameter k in the decay reflects the rate of absorption of the event. For rare event or catastrophe insurance applications such as in Dassios and Jang (2003), typically there would be a few events with large impact which is indicated by a small ρ and small η while for high frequency insurance businesses, there would be a large number of primary events with smaller impact which is indicated by relatively larger ρ and η. Hence the process is also adaptable to the characteristics of different lines of business which the insurer may wish to model.

Based on (3.1.1), it can be seen that the shot noise process is in fact mean reverting. The interpretation of this feature in an insurance context is that the claim arrival rate will generally be at a mean rate until events generate sudden increases in the rate. As the impact of the event decays, the intensity will revert to the mean level. This also implies

that as $t \to \infty$, the process will remain finite which is reasonable as one would expect that in the long run, the claim arrival rate for an insurer should not diverge.

From a forecasting perspective, it is also reasonable to assume that in order to predict future claim arrivals, only the current or most recent claim arrival information is needed. This feature can be mathematically translated into the Markovian property as described in the previous chapter. In order to show that the shot noise process is Markovian, we introduce the alternative formulation of the Markov property:

Definition 3.3. *A stochastic process* $\{\Lambda(t), t \geq 0\}$ *is said to be Markovian if and only if for all* $s < t$*, we have:*

$$\mathbb{E}[\Lambda(t)|\mathcal{F}_s] = \mathbb{E}[\Lambda(t)|\Lambda(s)] \tag{3.1.6}$$

Then considering the form of the shot noise process in (3.1.2), for any $s \leq t$, we have:

$$\begin{aligned}
\Lambda(t) &= \Lambda(0)e^{-kt} + \sum_{j=1}^{J(t)} X_j e^{-k(t-\tau_j)} \\
&= \Lambda(0)e^{-ks}e^{-k(t-s)} + \sum_{j=1}^{J(s)} X_j e^{-k(s-\tau_j)} + \sum_{j=J(s)+1}^{J(t)} X_j e^{-k(t-\tau_j)} \\
&= (\Lambda(0)e^{-ks} + \sum_{j=1}^{J(s)} X_j e^{-k(s-\tau_j)})e^{-k(t-s)} + \sum_{j=1}^{J(t)-J(s)} X_{J(s)+j} e^{-k(t-\tau_{J(s)+j})} \\
&= \Lambda(s)e^{-k(t-s)} + \sum_{j=1}^{J(t)-J(s)} X_{J(s)+j} e^{-k(t-\tau_{J(s)+j})}
\end{aligned}$$

Then taking the conditional expectations of both sides with the filtration of the intensity process $\mathcal{F}_t$, we have:

$$\begin{aligned}
\mathbb{E}[\Lambda(t)|\mathcal{F}_s] &= \mathbb{E}[\Lambda(s)e^{-k(t-s)} + \sum_{j=1}^{J(t)-J(s)} X_{J(s)+j} e^{-k(t-\tau_{J(s)+j})}|\mathcal{F}_s] \\
&= \Lambda(s)e^{-k(t-s)} + \mathbb{E}[\sum_{j=1}^{J(t)-J(s)} X_{J(s)+j} e^{-k(t-\tau_{J(s)+j})}|\mathcal{F}_s]
\end{aligned}$$

Since $\{J(t), t \geq 0\}$ is a Poisson process with arrival times τ_j, we can use the stationary and independent increment property of the homogeneous Poisson process $J(t)$. Hence, we

can see that:

$$\mathbb{E}[\sum_{j=1}^{J(t)-J(s)} X_{J(s)+j}e^{-k(t-\tau_{J(s)+j})}|\mathcal{F}_s] = \mathbb{E}[\sum_{j=1}^{J(t-s)} X_j e^{-k(t-s-\tau_j)}] = \mathbb{E}[\sum_{j=1}^{J(t-s)} X_j e^{-k(t-s-\tau_j)}|\Lambda(s)]$$

where τ_j are arrival times for the Poisson process $J(t-s)$. We note that τ_j, $J(t-s)$ and X_j are independent of $\Lambda(u)$ for $0 \leq u \leq s$ (and hence $\Lambda(s)4$). Thus this implies that:

$$\begin{aligned}\mathbb{E}[\Lambda(t)|\mathcal{F}_s] &= \Lambda(s)e^{-k(t-s)} + \mathbb{E}[\sum_{j=1}^{J(t-s)} X_j e^{-k(t-s-\tau_j)}|\Lambda(s)] \\ &= \mathbb{E}[\Lambda(t)|\Lambda(s)]\end{aligned}$$

Hence, we can see that the shot noise process with exponential decay is Markovian.

3.1.1 Moments of the shot noise intensity

We now derive the moment generating function of $\Lambda(t)$ in order to obtain the moments of $\Lambda(t)$. These properties will be used. We shall first consider the moments of $\Lambda(t)$ in the case the initial value $\Lambda(0)$ is a known constant. It can then be shown that if $\Lambda(0)$ is Gamma distributed, the stationary distribution of $\Lambda(t)$ is also Gamma. This is important from a modelling perspective, since $\Lambda(0)$ is not observable and hence an assumption needs to be made on its distribution.

Noting that in Equation (3.1.2), $\Lambda(0)$ and $S(t) = \sum_{i=1}^{J(t)} X_j e^{-k(t-\tau_j)}$ are independent, then see that in the case $\Lambda(0)$ is a constant:

$$\begin{aligned}\Phi_{\Lambda(t)|\Lambda(0)}(u) &= \Phi_{\Lambda(0)e^{-kt}+S(t)|\Lambda(0)}(u) \\ &= e^{\Lambda(0)e^{-kt}u}\Phi_{S(t)}(u)\end{aligned}$$

From the reformulation of the shot noise process in (3.1.3), let $Y_j = X_j e^{-ktU_j}$. Note that since X_j and U_j are independent and identically distributed, Y_j are also independent and identically distributed. Since $X_j \sim Exponential(\eta)$ and $U_j \sim Uniform(0,1)$ consider the

probability distribution function of Y_j:

$$\begin{aligned} P(Y_j \leq y) &= \int_0^1 P(e^{-ktU_j} X_j \leq y | U_j = u) du \\ &= \int_0^1 P(X_j \leq y e^{-ku}) du \qquad \text{since } X_j \text{ and } U_j \text{ are independent} \\ &= \int_0^1 1 - e^{-\eta y e^{ktu}} du \end{aligned}$$

Since the range of X_j is $(0, \infty)$ and the range of U_j is $(0, 1)$, then the range of Y_j will be $(0, \infty)$. We then differentiating both sides with respect to y to obtain the following probability density function for $y > 0$:

$$\begin{aligned} f_{Y_i}(y) &= \frac{d}{dy} \int_0^1 1 - e^{-\eta y e^{ktu}} du \\ &= \frac{1}{ykt} \int_0^1 \eta ykt e^{-\eta y e^{ktu}} e^{ktu} du \\ &= \frac{1}{ykt} (e^{-\eta y} - e^{-\eta y e^{kt}}) \end{aligned}$$

Note for $y \leq 0$, $f_{Y_i}(y) = 0$. Then the moment generating function of Y_j, $\Phi_{Y_j}(u)$ is given by:

$$\begin{aligned} \Phi_{Y_j}(u) = \mathbb{E}[e^{uY_j}] &= \int_0^\infty e^{-yu} f_{Y_j}(y) dy \\ &= \int_0^\infty \frac{1}{ykt} (e^{-\eta y} - e^{-\eta y e^{kt}}) e^{-yu} dy \end{aligned}$$

By letting:

$$\begin{aligned} \eta^* &= \eta - u \\ k^* &= \frac{1}{t} \ln \left(\frac{\eta e^{kt} - u}{\eta - u} \right) \end{aligned}$$

the above can be expressed as:

$$\Phi_{Y_j}(u) = \frac{k^*}{k} \int_0^\infty \frac{1}{yk^*t} (e^{-\eta^* y} - e^{-\eta^* y e^{k^* t}}) dy$$

Since $f^*(y) = \frac{1}{yk^*t}(e^{-\eta^* y} - e^{-\eta^* y e^{k^* t}})$ is a probability density function of Y_j with new parameters η^* and k^*, hence $\int_0^\infty \frac{1}{yk^*t}(e^{-\eta^* y} - e^{-\eta^* y e^{k^* t}}) = 1$ which then implies that:

$$\Phi_{Y_j}(u) = \frac{k^*}{k} = \frac{1}{kt} \ln\left(\frac{\eta e^{kt} - u}{\eta - u}\right) \tag{3.1.7}$$

Finally, we can derive the moment generating function of $\Lambda(t)$ conditional on $\Lambda(0)$. Note that $\Phi_{S(t)}(u) = G_{J(t)}(\Phi_{Y_i}(u))$ where $G_{J(t)}$ is the probability generating function for $J(t)$. Since $J(t)$ is Poisson distributed with parameter ρt, then using (3.1.7):

$$\begin{aligned}\Phi_{S(t)}(u) &= \exp\left(\rho t\left[\frac{1}{kt}\ln\left(\frac{\eta e^{kt} - u}{\eta - u}\right) - 1\right]\right) \\ &= e^{-\rho t}\left(\frac{\eta e^{kt} - s}{\eta - u}\right)^{\frac{\rho}{k}}\end{aligned}$$

Hence the moment generating function of $\Lambda(t)$ conditional on $\Lambda(0)$ is given by:

$$\Phi_{\Lambda(t)|\Lambda(0)}(u) = e^{\Lambda(0)e^{-kt}u} e^{-\rho t}\left(\frac{\eta e^{kt} - u}{\eta - u}\right)^{\frac{\rho}{k}} \tag{3.1.8}$$

Consider the cummulant generating function of $\Lambda(t)|\Lambda(0)$:

$$\kappa_{\Lambda(t)|\Lambda(0)}(u) = \log\left(\Phi_{\Lambda(t)|\Lambda(0)}(u)\right) = \Lambda(0)e^{-kt}u - \rho t + \frac{\rho}{k}\log\left(\frac{\eta e^{kt} - u}{\eta - u}\right) \tag{3.1.9}$$

By differentiating the cummulant generating function, the conditional mean of $\Lambda(t)$ can be shown to be:

$$\mathbb{E}[\Lambda(t)|\Lambda(0)] = \kappa'_{\Lambda(t)|\Lambda(0)}(0) = (\Lambda(0) - \frac{\rho}{\eta k})e^{-kt} + \frac{\rho}{\eta k} \tag{3.1.10}$$

Similarly, the conditional variance of $\Lambda(t)|\Lambda(0)$ can be also shown to be:

$$\mathrm{Var}(\Lambda(t)|\Lambda(0)) = \kappa''_{\Lambda(t)|\Lambda(0)}(0) = \frac{\rho}{\eta^2 k}(1 - e^{-2kt}) \tag{3.1.11}$$

Aside from the moments of the shot noise Cox process, it is also of interest to observe the correlation structure of the shot noise Cox process.

Proposition 3.1. *(Covariance of shot noise process) For all $s < t$, the covariance of the shot noise process conditional on $\Lambda(0)$ is given by:*

$$Cov(\Lambda(t), \Lambda(s)|\Lambda(0)) = \frac{\rho}{\eta^2 k}(e^{-k(t-s)} - e^{-k(t+s)}) \tag{3.1.12}$$

Proof. To prove this result, we utilise the following result from Klüppelberg and Mikosch (1995):

Lemma 3.2. *Consider the shot noise process of the following form:*

$$S(t) = \sum_{i=1}^{J(t)} H(t - \tau_j)$$

where $J(t)$ is a homogeneous Poisson process with rate ρ. Then given that $Var(S(t)) < \infty$, for any $s < t$, the covariance of $S(t)$ is given by:

$$Cov(S(t), S(s)) = \rho \int_0^s \mathbb{E}[H(u)H(u + t - s)]du \tag{3.1.13}$$

Noting that for the classical shot noise process, $H(t) = X_j e^{-k(t-\tau_j)}$ where X_j is an exponential random variable with parameter η. Then we use (3.1.13) to see that:

$$\begin{aligned}
\text{Cov}(\Lambda(t), \Lambda(s)|\Lambda(0)) &= \text{Cov}\left(\sum_{i=1}^{J(t)} X_j e^{-k(t-\tau_j)}, \sum_{i=1}^{J(s)} X_j e^{-k(s-\tau_j)}\right) \\
&= \rho \int_0^s \mathbb{E}[X^2 e^{-ku} e^{-k(u+t-s)}]du \\
&= \rho \int_0^s \mathbb{E}[X^2] e^{-k(2u+t-s)}]du \\
&= \frac{\rho}{\eta^2 k}(e^{-k(t-s)} - e^{-k(t+s)})
\end{aligned}$$

□

3.1.2 Stationary distribution of the shot noise intensity

It is interesting to observe the long run behaviour of the shot noise intensity. The stationary distribution of $\Lambda(t)$ reflects the distribution of $\Lambda(t)$ in the long run (as $t \to \infty$). It can be shown that the stationary distribution of . This is done by consider the limit of (3.1.8)

as $t \to \infty$:

$$\begin{aligned}\Phi_{\Lambda(\infty)}(u) &= \lim_{t\to\infty} \Phi_{\Lambda(t)}(u) \\ &= \lim_{t\to\infty} e^{\Lambda(0)e^{-kt}u} \left(\frac{\eta - ue^{-kt}}{\eta - u}\right)^{\frac{\rho}{k}} \\ &= \left(\frac{\eta}{\eta - u}\right)^{\frac{\rho}{k}}\end{aligned}$$

which resembles the moment generating function of a Gamma distribution with parameters $\frac{\rho}{k}$ and η. This means that in the long run, the realisation of $\Lambda(t)$ should resemble a Gamma distribution.

The mean of the stationary distribution of $\Lambda(t)$ can also be determined by taking the limit of (3.1.10) as $t \to \infty$:

$$\mathbb{E}[\Lambda(t)] = \frac{\rho}{\eta k} \tag{3.1.14}$$

Similarly, the variance of $\Lambda(t)$ as $t \to \infty$ is given by taking the limit of (3.1.11):

$$\mathrm{Var}(\Lambda(t)) = \frac{\rho}{\eta^2 k} \tag{3.1.15}$$

Remark 3.1. *So far, the moments and correlation structure of the shot noise intensity process $\Lambda(t)$ have been derived under the assumption that $\Lambda(0)$ is a known constant. Suppose instead of treating $\Lambda(0)$ as a constant, we assume that $\Lambda(0)$ is Gamma distributed with parameters $\frac{\rho}{k}$ and η. Then it can be shown that the distribution of $\Lambda(t)$ coincides with the stationary distribution of $\Lambda(t)$.*

To show this, consider that the moment generating function of $\Lambda(t)$ under this assumption using (3.1.8):

$$\begin{aligned}\Phi_{\Lambda(t)}(u) &= \Phi_{\Lambda(0)}(e^{-kt}u)\Phi_{S(t)}(e^{-kt}u) \\ &= \Phi_{\Lambda(0)}(e^{-kt}u)e^{-\rho t}\left(\frac{\eta e^{kt} - u}{\eta - u}\right)^{\frac{\rho}{k}}\end{aligned}$$

Since $\Lambda(0)$ is Gamma distributed, then the moment generating function of $\Lambda(0)$ is:

$$\Phi_{\Lambda(0)}(e^{-kt}u) = \left(\frac{\eta}{\eta - e^{-kt}u}\right)^{\frac{\rho}{k}} = \left(\frac{e^{kt}\eta}{e^{kt}\eta - u}\right)^{\frac{\rho}{k}}$$

Hence:

$$\begin{aligned}\Phi_{\Lambda(t)}(u) &= \left(\frac{e^{kt}\eta}{e^{kt}\eta - u}\right)^{\frac{\rho}{k}} e^{-\rho t}\left(\frac{\eta e^{kt} - u}{\eta - u}\right)^{\frac{\rho}{k}} \\ &= \left(\frac{\eta}{\eta - s}\right)^{\frac{\rho}{k}}\end{aligned}$$

which resembles the moment generating function of the Gamma distribution. This also shows that the marginal distribution of $\Lambda(t)$ is the same as the assumed distribution of $\Lambda(0)$.

Under this additional assumption, the first and second moments of $\Lambda(t)$ under this assumption is the same as the long run mean and variance in (3.1.14) and (3.1.15) respectively. This additional assumption changes the correlation structure of $\Lambda(t)$ however due to the uncertainty provided by assuming $\Lambda(0)$ is a random variable.

Proposition 3.2. *Suppose $\Lambda(0)$ is Gamma distributed with parameters $\frac{\rho}{k}$ and η, then the covariance structure of $\Lambda(t)$ for all $s \leq t$ is given by:*

$$Cov(\Lambda(t), \Lambda(s)) = \frac{\rho}{\eta^2 k} e^{-k(t-s)} \tag{3.1.16}$$

Proof. By using the property of conditional covariance, $\mathrm{Cov}(X, Y) = E[\mathrm{Cov}(X, Y|Z)] + \mathrm{Cov}(E[X|Z], E[Y|Z])$, on (3.1.12), the covariance of $\Lambda(t)$ for $s < t$ is given by:

$$\begin{aligned}\mathrm{Cov}(\Lambda(t), \Lambda(s)) &= \mathbb{E}\left[\mathrm{Cov}(\Lambda(t), \Lambda(s)|\Lambda(0))\right] + \mathrm{Cov}(\mathbb{E}[\Lambda(t)|\Lambda(0)], \mathbb{E}[\Lambda(s)|\Lambda(0)]) \\ &= \mathbb{E}\left[\frac{\rho}{\eta^2 k}(e^{-k(t-s)} - e^{-k(t+s)})\right] \\ &+ \mathrm{Cov}\left((\Lambda(0) - \frac{\rho}{\eta k})e^{-kt} + \frac{\rho}{\eta k}, (\Lambda(0) - \frac{\rho}{\eta k})e^{-ks} + \frac{\rho}{\eta k}\right) \\ &= \frac{\rho}{\eta^2 k}(e^{-k(t-s)} - e^{-k(t+s)}) + \mathrm{Var}(\Lambda(0)e^{-k(t+s)} \\ &= \frac{\rho}{\eta^2 k}(e^{-k(t-s)} - e^{-k(t+s)}) + \frac{\rho}{\eta^2 k}e^{-k(t+s)} \\ &= \frac{\rho}{\eta^2 k}e^{-k(t-s)}\end{aligned}$$

□

Based on the above result, it can be seen that under the assumption that $\Lambda(0)$ has a Gamma distribution, the covariance between $\Lambda(t)$ and $\Lambda(s)$ only depends on the length of the time interval $t - s$. Along with Equation (3.1.14), this implies that the shot noise

intensity at the very least exhibits weak stationarity, which implies that the shot noise Cox process and its increments should also exhibit weak stationarity. In fact, we show in the next section that the increments of the shot noise Cox process exhibits strict stationarity.

3.2 The increment of the shot noise Cox process

In this section, we extend the statistical properties derived and discussed in the previous section to shot noise Cox processes. We first derive the moment generating function and the probability mass function for the increment of the shot noise Cox process $N(t)-N(s)$. Using the results introduced in the previous section, the moments and correlation structure of the Cox process with shot noise intensity also are derived.

First, we define the shot noise Cox process based on the definition in Grandell (1976):

Definition 3.4. *(Shot noise Cox process)*
Denote $\{N(t), t \geq 0\}$ *as a shot noise Cox process with the shot noise intensity process* $\{\Lambda(t), t \geq 0\}$ *from Definition 3.1. The probability mass function of the increment* $N(t) - N(s)$ *is given by:*

$$P(N(t) - N(s) = n) = \mathbb{E}\left[\frac{e^{-(M(t)-M(s))}(M(t) - M(s))^n}{n!}\right] \tag{3.2.1}$$

where $M(t)$ *is the mean value process given in Equation* (3.1.5)*. It can be said that the shot noise Cox process is a homogeneous Poisson process* $\{\tilde{N}(t), t \geq 0\}$ *subordinated by the mean value process* $\{M(t), t \geq 0\}$*, or in other words* $N(t) \stackrel{d}{=} \tilde{N}(M(t))$*.*

The above definition of the probability mass function is expressed in terms of an increment of the process $\{N(t), t \geq 0\}$ as the research focuses on modelling number of events within a time interval. Hence the statistical properties derived in this section will be based on the increment $N(t) - N(s)$ as they are used in the modelling procedure described in the following chapters.

3.2.1 Moment generating function of $N(t) - N(s)$

In this section, we shall derive the moment generating function and the pr. Before we can derive the Conditional on $\Lambda(0)$, we first derive the conditional moment generating function

of the increment of the mean value process $M(t) - M(s)$:

Proposition 3.3. *(Conditional moment generating function of $M(t) - M(s)$)*
For $0 \leq s \leq t$, the moment generating function of $M(t) - M(s)$ conditional on $\Lambda(0)$ is given by:

$$\mathbb{E}[e^{\nu(M(t)-M(s))}|\Lambda(0)] = e^{\frac{\nu}{k}(e^{-ks}-e^{-kt})\Lambda(0)} \left(\frac{\eta - \frac{\nu}{k}(e^{-ks} - e^{-k(t)})}{\eta - \frac{\nu}{k}(1 - e^{-k(t-s)})}\right)^{\frac{\rho}{k}} \times \exp\left(\frac{\rho\nu}{-\nu + \eta k}(t-s)\right)\left(\frac{-\nu(1-e^{-k(t-s)}) + \eta k}{\eta k}\right)^{\frac{\rho\eta}{-\nu+\eta k}} \tag{3.2.2}$$

Proof. In order to prove this result, we use the following result from Dassios and Jang (2003):

Lemma 3.3. *(Dassios and Jang, 2003) For the mean value process of the form in* (3.1.5)*, we have the Laplace transform of the increment $M(t) - M(s)$:*

$$\mathbb{E}[e^{-\nu(M(t)-M(s))}|\Lambda(s)] = \exp\left(-\frac{\nu}{k}(1-e^{-k(t-s)})\Lambda(s)\right)\exp\left(-\int_s^t \rho\left(1 - \hat{g}\left(\frac{\nu}{k}(1-e^{-k(t-s)})\right)\right)\right) \tag{3.2.3}$$

where $\hat{g}(\cdot)$ is the moment generating function for X_j.

Noting that since X_j is exponentially distributed with parameter η, then $\hat{g}(u) = \mathbb{E}[e^{uX_i}] = \frac{\eta}{\eta+u}$. Then by considering the integral in Equation (3.2.3):

$$\begin{aligned} I &= \int_s^t \rho\left(1 - \frac{\eta}{\eta + \frac{\nu}{k}(1-e^{-k(t-u)})}\right)du \\ &= \rho\left(t - s - \frac{\eta}{\nu+\eta k}\int_{e^{ks}}^{e^{kt}} \frac{1}{u} + \frac{\nu e^{-kt}}{\nu + \eta k - \nu e^{-kt}}du\right) \\ &= \rho\left(t - s - \frac{\eta}{\nu+\eta k}\left(kt - ks + \ln\left(\frac{\nu + \eta k - \nu e^{-k(t-s)}}{\eta k}\right)\right)\right) \end{aligned}$$

By substituting the above back into Equation (3.2.3) gives:

$$\mathbb{E}[e^{-\nu(M(t)-M(s))}|\Lambda(s)] = \exp\left(-\frac{\nu}{k}(1-e^{-k(t-s)})\Lambda(s)\right)\exp\left(-\frac{\rho\nu}{\nu+\eta k}(t-s)\right) \times \left(\frac{\nu(1-e^{-k(t-s)}) + \eta k}{\eta k}\right)^{\frac{\rho\eta}{\nu+\eta k}}$$

We can rewrite Equation (3.1.5) in the following way:

$$\begin{aligned} M(t) &= \frac{1}{k}\left(\Lambda(0) + \sum_{j=1}^{J(t)} X_j - (\Lambda(0)e^{-kt} + \sum_{j=1}^{J(t)} X_j e^{-k(t-\tau_j)}\right) \\ &= \frac{1}{k}\left(\Lambda(0) + \sum_{j=1}^{J(t)} X_j - \Lambda(t)\right) \end{aligned}$$

Then considering $M(t) - M(s)$ gives:

$$\begin{aligned} M(t) - M(s) &= \frac{1}{k}\left(\Lambda(0) + \sum_{j=1}^{J(t)} X_j - \Lambda(t)\right) - \frac{1}{k}\left(\Lambda(0) + \sum_{j=1}^{J(s)} X_j - \Lambda(s)\right) \\ &= \frac{1}{k}\left(\sum_{j=J(s)+1}^{J(t)} X_j - \Lambda(t) + \Lambda(s)\right) \\ &= \frac{1}{k}\left(\sum_{j=1}^{J(t)-J(s)} X_j - \Lambda(t) + \Lambda(s)\right) \\ &= \frac{1}{k}\left(\sum_{j=1}^{J(t-s)} X_j - \Lambda(t) + \Lambda(s)\right) \end{aligned}$$

where we use the fact $J(t)$ has stationary and independent increments in the above to see that $\sum_{j=1}^{J(t-s)} X_j$ is independent of $\Lambda(s)$. Since we have shown that $\Lambda(t)$ is Markovian, see that for any deterministic function $f : \mathbb{R} \to \mathbb{R}$, $\mathbb{E}[f(\Lambda(t))|\mathcal{F}_s] = \mathbb{E}[f(\Lambda(t))|\Lambda(s)]$. Hence:

$$\begin{aligned} \mathbb{E}[e^{-\nu(M(t)-M(s))}|\Lambda(s), \Lambda(0)] &= \mathbb{E}[e^{\frac{1}{k}\left(\sum_{j=1}^{J(t-s)} X_j - \Lambda(t) + \Lambda(s)\right)}|\Lambda(s), \Lambda(0)] \\ &= \mathbb{E}[e^{\frac{1}{k}\left(\sum_{j=1}^{J(t-s)} X_j - \Lambda(t) + \Lambda(s)\right)}|\Lambda(s)] \\ &= \mathbb{E}[e^{-\nu(M(t)-M(s))}|\Lambda(s)] \end{aligned}$$

where we have used the Markov property of $\Lambda(t)$ and the independence of $\sum_{j=1}^{J(t-s)} X_j$ from $\Lambda(s)$. Then using the following property of conditional expectation:

$$\mathbb{E}[\mathbb{E}[e^{-\nu(M(t)-M(s))}|\Lambda(s), \Lambda(0)]|\Lambda(0)] = \mathbb{E}[e^{-\nu(M(t)-M(s))}|\Lambda(0)]$$

which implies that:

$$\mathbb{E}[e^{-\nu(M(t)-M(s))}|\Lambda(s)]|\Lambda(0)] = [\mathbb{E}[e^{-\nu(M(t)-M(s))}|\Lambda(0)]$$

then we take the expectation of (3.2.3) conditional on $\Lambda(0)$ to obtain:

$$[\mathbb{E}[e^{-\nu(M(t)-M(s))}|\Lambda(0)] = \mathbb{E}[\exp\left(-\frac{\nu}{k}(1-e^{-k(t-s)})\Lambda(s)\right)|\Lambda(0)] \times \exp\left(-\frac{\rho\nu}{\nu+\eta k}(t-s)\right)\left(\frac{\nu(1-e^{-k(t-s)})+\eta k}{\eta k}\right)^{\frac{\rho\eta}{\nu+\eta k}}$$

From Equation (3.1.8), it can be deduced that:

$$\mathbb{E}\left[\exp\left(-\frac{\nu}{k}(1-e^{-k(t-s)})\Lambda(s)\right)|\Lambda(0)\right] = e^{-\frac{\nu}{k}(1-e^{-k(t-s)})\Lambda(0)e^{-kt}}e^{-\rho s}\left(\frac{\eta e^{ks}+\frac{\nu}{k}(1-e^{-k(t-s)})}{\eta+\frac{\nu}{k}(1-e^{-k(t-s)})}\right)^{\frac{\rho}{k}}$$
$$= e^{-\frac{\nu}{k}(e^{-ks}-e^{-kt})\Lambda(0)}\left(\frac{\eta+\frac{\nu}{k}(e^{-ks}-e^{-k(t)})}{\eta+\frac{\nu}{k}(1-e^{-k(t-s)})}\right)^{\frac{\rho}{k}}$$

Thus recombining the expressions gives:

$$[\mathbb{E}[e^{-\nu(M(t)-M(s))}|\Lambda(0)] = e^{-\frac{\nu}{k}(e^{-ks}-e^{-kt})\Lambda(0)}\left(\frac{\eta+\frac{\nu}{k}(e^{-ks}-e^{-k(t)})}{\eta+\frac{\nu}{k}(1-e^{-k(t-s)})}\right)^{\frac{\rho}{k}} \times \exp\left(-\frac{\rho\nu}{\nu+\eta k}(t-s)\right)\left(\frac{\nu(1-e^{-k(t-s)})+\eta k}{\eta k}\right)^{\frac{\rho\eta}{\nu+\eta k}}$$

Substituting $-\nu$ with ν will give the result in the proposition. □

Remark 3.2. *The above gives the moment generating function for the mean value process $M(t)$ given $\Lambda(0)$ as:*

$$\Phi_{M(t)|\Lambda(0)}(\nu) = \mathbb{E}[e^{\nu M(t)}|\Lambda(0)]$$
$$= e^{\frac{\nu}{k}(e^{-ks}-e^{-kt})\Lambda(0)}\left(\frac{\eta-\frac{\nu}{k}(1-e^{-k(t)})}{\eta-\frac{\nu}{k}(1-e^{-kt})}\right)^{\frac{\rho}{k}}$$
$$\times \exp\left(\frac{\rho\nu}{-\nu+\eta k}t\right)\left(\frac{-\nu(1-e^{-kt})+\eta k}{\eta k}\right)^{\frac{\rho\eta}{-\nu+\eta k}}$$

Then using Equation (2.2.1) *from Cox (1955), the moment generating function of the Cox*

process $N(t)$ given $\Lambda(0)$ is given by:

$$\begin{aligned}\Phi_{N(t)|\Lambda(0)}(\nu) &= \Phi_{M(t)|\Lambda(0)}(e^{\nu}-1) \\ &= e^{\frac{e^{\nu}-1}{k}(e^{-ks}-e^{-kt})\Lambda(0)}\left(\frac{\eta-\frac{e^{\nu}-1}{k}(1-e^{-kt})}{\eta-\frac{e^{\nu}-1}{k}(1-e^{-kt})}\right)^{\frac{\rho}{k}} \\ &\times \exp\left(\frac{\rho(e^{\nu}-1)}{-(e^{\nu}-1)+\eta k}t\right)\left(\frac{-(e^{\nu}-1)(1-e^{-kt})+\eta k}{\eta k}\right)^{\frac{\rho\eta}{-(e^{\nu}-1)+\eta k}}\end{aligned}$$

Suppose we assume $\Lambda(0)$ has a Gamma distribution with parameters $\frac{\rho}{k}$ and η. Then we can derive the moment generating function for the increments $M(t)-M(s)$ and $N(t)-N(s)$:

Proposition 3.4. *(Moment generating function of $N(t)-N(s)$)*
Assuming $\Lambda(0)$ has a Gamma distribution with parameters $\frac{\rho}{k}$ and η, the moment generating function for $M(t)-M(s)$ is given by:

$$\mathbb{E}[e^{\nu(M(t)-M(s))}] = \exp\left(\frac{\rho\nu}{-\nu+\eta k}(t-s)\right)\left(\frac{-\nu(1-e^{-k(t-s)})+\eta k}{\eta k}\right)^{\frac{-\rho\nu}{(-\nu+\eta k)k}} \tag{3.2.4}$$

and hence the moment generating function for $N(t)-N(s)$ is given by:

$$\mathbb{E}[e^{\nu(N(t)-N(s))}] = \exp\left(\frac{\rho(e^{\nu}-1)}{-(e^{\nu}-1)+\eta k}(t-s)\right)\left(\frac{-(e^{\nu}-1)(1-e^{-k(t-s)})+\eta k}{\eta k}\right)^{\frac{-\rho(e^{\nu}-1)}{(-(e^{\nu}-1)+\eta k)k}} \tag{3.2.5}$$

Proof. By taking the expectation of Equation (3.2.2) with respect to $\Lambda(0)$, by the law of total expectation, we obtain:

$$\begin{aligned}\mathbb{E}[e^{\nu(M(t)-M(s))}] &= \mathbb{E}[\mathbb{E}[e^{\nu(M(t)-M(s))}|\Lambda(0)]] \\ &= \mathbb{E}\left[e^{\frac{\nu}{k}(e^{-ks}-e^{-kt})\Lambda(0)}\right]\left(\frac{\eta-\frac{\nu}{k}(e^{-ks}-e^{-k(t)})}{\eta-\frac{\nu}{k}(1-e^{-k(t-s)})}\right)^{\frac{\rho}{k}} \\ &\times \exp\left(\frac{\rho\nu}{-\nu+\eta k}(t-s)\right)\left(\frac{-\nu(1-e^{-k(t-s)})+\eta k}{\eta k}\right)^{\frac{\rho\eta}{-\nu+\eta k}}\end{aligned}$$

From Klugman et al. (1998), the moment generating function for a $Gamma(\frac{\rho}{k},\eta)$ random variable $\Lambda(0)$ is given by:

$$\Phi_{\Lambda(0)}(u) = \left(\frac{\eta}{\eta-u}\right)^{\frac{\rho}{k}}$$

Hence we have:

$$\begin{aligned}\mathbb{E}[e^{\nu(M(t)-M(s))}] &= \left(\frac{\eta}{\eta-\frac{\nu}{k}(e^{-ks}-e^{-kt})}\right)^{\frac{\rho}{k}}\left(\frac{\eta-\frac{\nu}{k}(e^{-ks}-e^{-k(t)})}{\eta-\frac{\nu}{k}(1-e^{-k(t-s)})}\right)^{\frac{\rho}{k}} \\ &\quad\times\exp\left(\frac{\rho\nu}{-\nu+\eta k}(t-s)\right)\left(\frac{-\nu(1-e^{-k(t-s)})+\eta k}{\eta k}\right)^{\frac{\rho\eta}{-\nu+\eta k}} \\ &= \exp\left(\frac{\rho\nu}{-\nu+\eta k}(t-s)\right)\left(\frac{-\nu(1-e^{-k(t-s)})+\eta k}{\eta k}\right)^{\frac{-\rho\nu}{(-\nu+\eta k)k}}\end{aligned}$$

We then substitute $\nu = e^{\nu} - 1$ into the above to obtain Equation (3.2.5). □

Remark 3.3. *By letting $\nu = -\nu$ in Equation* (3.2.4)*, we obtain*

$$\mathbb{E}[e^{-\nu(M(t)-M(s))}] = \exp\left(-\frac{\rho\nu}{\nu+\eta k}(t-s)\right)\left(\frac{\nu(1-e^{-k(t-s)})+\eta k}{\eta k}\right)^{\frac{\rho\nu}{(\nu+\eta k)k}} \tag{3.2.6}$$

Hence in order to obtain the probability mass function for n claims for the Cox process, the n^{th} derivative of the above equation is taken such that:

$$P(N(t)-N(s)=n) = \frac{(-1)^n}{n!}\frac{d^{(n)}}{d\nu^n}\mathbb{E}[e^{-\nu(M(t)-M(s))}]\Big|_{\nu=1}$$

In fact, (3.2.6) is the probability of no events in the time interval $[s,t]$ which can be used in determining survival probabilities in mortality or default times in credit risk. Since the derivative of Equation (3.2.6) of any order will only depend on the length of the time interval $t-s$, this implies that the increment of the shot noise Cox process $N(t)-N(s)$ are stationary.

However, from large general insurance companies, there usually are a large number of claims incurred in a small interval of time such as in a day, meaning we need to take the large order of the derivative of (3.2.6). This implies that even over a small time increment, there will be computational issues in attaining higher order derivatives required. This issue is exacerbated when considering number of claims over larger time increments such as weeks, fortnights and months. Hence the computational difficulty involved in finding exact probabilities prevents this method from being used efficiently in an insurance context.

3.2.2 Moments and correlation structure of the Shot noise Cox process

Through finding the first and second derivative of the logarithm of Equation (3.2.5) with respect to ν, one can deduce the first two central moments of $N(t) - N(s)$ (under the assumption $\Lambda(0)$ is Gamma distributed) as:

$$E[N(t) - N(s)] = \frac{\rho(t-s)}{\eta k} \tag{3.2.7}$$

$$\mathrm{Var}(N(t) - N(s)) = \frac{2\rho}{\eta^2 k^2}\left(t - s - \frac{1 - e^{-k(t-s)}}{k}\right) + \frac{\rho(t-s)}{\eta k} \tag{3.2.8}$$

Then by dividing Equation (3.2.8) by (3.2.7), the coefficient of variation can be deducted to be:

$$\text{Coefficient of variation: } = \frac{2}{\eta k}\left(1 - \frac{1 - e^{-k(t-s)}}{k(t-s)}\right) + 1 \tag{3.2.9}$$

From the above, since the coefficent of variation is always greater than 1, the shot noise Cox process is overdispersed compared to the homogeneous Poisson process. It is actually invariant on the value of ρ. However, as η decreases, which implies the expected jump size in the intensity increases, the level of overdispersion increases.

Since the increments are stationary, then there should be an autocovariance function for the increments. To observed the correlation structure of the shot noise Cox process, the covariance of the shot noise Cox process can then be found as:

Proposition 3.5. *The covariance structure of the shot noise Cox process $N(t)$ is given for any $s < t$:*

$$Cov(N(t), N(s)) = \frac{2\rho s}{\eta^2 k^2} + \frac{\rho(e^{-kt} + e^{-ks} - 1 - e^{-k(t-s)})}{\eta^2 k^3} + \frac{\rho s}{k} \tag{3.2.10}$$

Proof. Using the fact $\mathrm{Cov}(X, Y) = -\frac{1}{2}(\mathrm{Var}(X - Y) - \mathrm{Var}(X) - \mathrm{Var}(Y))$, then see that:

$$\begin{aligned}
\mathrm{Cov}(N(t), N(s)) = -\frac{1}{2}\Bigg(&\frac{2\rho}{\eta^2 k^2}\left(t - s - \frac{1 - e^{-k(t-s)}}{k}\right) + \frac{\rho(t-s)}{\eta k} - \frac{2\rho}{\eta^2 k^2}\left(t - \frac{1 - e^{-kt}}{k}\right) \\
&- \frac{\rho t}{\eta k} - \frac{2\rho}{\eta^2 k^2}\left(s - \frac{1 - e^{-ks}}{k}\right) - \frac{\rho s}{\eta k}\Bigg) \\
= &\frac{2\rho s}{\eta^2 k^2} + \frac{\rho(e^{-kt} + e^{-ks} - 1 - e^{-k(t-s)})}{\eta^2 k^3} + \frac{\rho s}{k}
\end{aligned}$$

□

Then we can find the autocovariance function of $N(t) - N(t-1)$ for $h = 1, 2, 3, ...$ as:

$$\begin{aligned}\gamma(h) &= \text{Cov}(N(t+h) - N(t-1+h), N(t) - N(t-1)) \\ &= \text{Cov}(N(t+h), N(t)) - \text{Cov}(N(t+h), N(t-1)) \\ &- \text{Cov}(N(t-1+h), N(t)) + \text{Cov}(N(t-1+h), N(t-1)) \\ &= \frac{\rho e^{-kh}(e^{\sqrt{k}} - e^{-\sqrt{k}})^2}{\eta^2 k^3} \end{aligned} \tag{3.2.11}$$

The above implies a couple of features about the shot noise Cox process. Firstly, despite having stationary increments, the shot noise Cox process does not have independent increments. Secondly, for, the autocovariance function (and hence the autocorrelation function) of $N(t) - N(t-1)$ is always postive for lags greater than zero and exponentially decays as $h \to \infty$ where k is the measure of how quickly the autocovariance decays with h. This provides a quick method to check the validity of using the Cox process on claims data by checking the autocorrelation of the increments are positive.

CHAPTER 4

METHODS ON FITTING THE SHOT NOISE COX PROCESS

Given that the intensity process completely characterises the doubly stochastic Poisson process, the problem of fitting the counts process is reduced to fitting the intensity process. This is done so by estimating the parameters ρ, η and k. The main issue, however, is that only the claim counts process is actually observed while the intensity process are latent variables. One method that could have been used to bypass the filtering problem is to utilise the explicit expression for distribution for the number of counts over a time increment, $N(t + \Delta t) - N(t)$, derived in the previous chapter and used to construct a likelihood function as follows:

$$L(\rho, \eta, k) = \prod_{i=1}^{m} P(N(\Delta t(i+1)) - N(\Delta ti) = n_i)$$

where n_i represents the number of counts in the n^{th} increment. There are two main problems with this approach though:

- Unlike the homogeneous and inhomogeneous Poisson process, the Cox process does not necessarily have independent increments property. In fact, we have shown that the covariance between disjoint increments is not always zero for shot noise Cox processes in Chapter 2. Hence the joint distribution of the claim counts in all

increments cannot be simply expressed as the product of the marginal distribution of each increment.

- Even if the independent increment assumption was valid for the Cox process, calculating the probability of n claims in a time interval relies on taking the n^{th} derivative of the expression derived in the previous chapter. In high volume business lines in general insurance such as domestic and commercial motor industries, the insurer typically faces up to tens to hundreds of claims per day. This implies that the high order derivatives will be required to evaluation $P(N(\Delta t(i+1)) - N(\Delta ti) = n_i)$, which is computationally intensive and inefficient.

As this problem is also prevalent in areas such as finance in modelling stochastic volatility, methods used in those fields can be borrowed. In this chapter, we will investigate two potential methods for filtering of the intensity process. The first of these is the Kalman filter which was first developed by Kalman (1960) and has been widely applied in finance to model stochastic volatility. The filter, however, requires the latent intensity to follow a Gaussian distribution which allowed for the implementation for an Ornstein-Uhlenbeck intensity process in Rydberg and Shephard (1999). Hence, Gaussian approximations of the shot noise intensity and claim counts process as explored in Dassios and Jang (2003) will be required. The other popular method involves Markov Chain Monte Carlo simulations. In particular, we will be discussing the method developed in Centanni and Minozzo (2006) for filtering shot noise intensity from a shot noise Cox process.

In Section 4.1, methods of obtaining initial estimates for the parameters are developed via a distributional approximation of shot noise Cox process along with method of moments. Obtaining initial estimates which are close enough to the true value of the parameters helps improve the efficiency and accuracy of the final estimate the methods converge to. In Section 4.2, we develop a method based on the Kalman filter where we utilises the Kalman Bucy approximation of both the shot noise intensity and Cox process in Dassios and Jang (2005) to estimate parameters. This method is relatively quick to implement and. Hence, in section 4.3, an alternative method based on the reverse jump Markov Chain Monte Carlo expectation maximisation algorithm in Centanni and Minozzo (2006) is proposed. As $\Lambda(0)$ is unobservable, for both methods, it shall be assumed that $\Lambda(0)$ has a Gamma distribution with parameters $\frac{\rho}{k}$ and η as in Remark 3.1. Finally, in Section 4.4, both heuristic and formal goodness of fit tests are provided.

4.1 Negative Binomial Approximation of Shot Noise Cox process

In the filtering methods for estimating parameters, the final estimates are usually dependent on the initial input estimates of the parameters. Obtain good initial estimates also allows for quicker and more accurate convergence of the estimation process to the true parameters. Hence, a method is developed in order to obtain initial estimates of the parameters.

From Grandell (1976), the doubly stochastic Poisson process is a homogeneous Poisson process subordinated by the mean value process. In other words, $N(t) \stackrel{d}{=} \tilde{N} \circ M(t)$ where $\tilde{N}$ is a homogeneous Poisson process and $M(t)$ is the mean value process. Then from the following first order Taylor expansion of $M(t)$ and noting that $\Lambda(t) = \frac{dM(t)}{dt}$ is the underlying intensity process:

$$\begin{aligned} M(t+\Delta) - M(t) &\approx \frac{dM(t)}{dt}\Delta \\ \Rightarrow M(t+\Delta) - M(t) &\approx \Delta\Lambda(t) \end{aligned}$$

Consider the following discretisation of the Cox process:

$$\begin{aligned} N(t+\Delta) - N(t) &\stackrel{d}{=} \tilde{N}(M(t+\Delta)) - \tilde{N}(M(t)) \\ &\stackrel{d}{=} \tilde{N}(M(t+\Delta) - M(t)) \quad \text{by stationary increments of homogeneous Poisson} \\ &\approx \tilde{N}(\Lambda(t)\Delta) \end{aligned}$$

From the derivation of the moment generating function for $\Lambda(t)$ in the previous chapter, we can see that the stationary distribution is Gamma distributed with parameters $\frac{\rho}{k}$ and η. A homogeneous Poisson process with a Gamma (a, b) distributed subordinator results in a negative binomial process with parameters $r = a$ and $p = \frac{b}{b+1}$. Hence, one simple way to approximately estimate the parameters is to assume the increments of the counts process $N(\Lambda(t)\Delta)$ follow a negative binomial (r, p) distribution. Noting that $a = \frac{\rho}{k}$ and $b = \eta$, we have the following equations which give the initial estimates ρ_0, k_0 and η_0:

$$\frac{\rho_0}{k_0} = \hat{r} \qquad \eta_0 = \frac{\hat{p}}{1-\hat{p}}$$

where $\hat{r}$ and $\hat{p}$ are the estimated parameters from the negative binomial approximation. This however presents the problem of having only two equations to solve for three pa-

rameters. From the properties derived in Chapter 3 are used to obtain separate initial estimates for both ρ and k. This could either be done using the coefficient of variation in Equation .

4.2 Kalman Filter Approximation

In this section, a method of filtering for the unobservable intensity based on the Kalman filter is proposed. The Kalman filter, as described in 2, is can only be applied where both the measurable and unobservable processes are Gaussian. Hence, a Gaussian approximation of both the shot noise intensity and the counts process in Dassios and Jang (2005) in order to apply the Kalman Filter to estimate the conditional intensity process. The Kalman Bucy differential equations derived in Dassios and Jang (2005) for the Gaussian approximations are then discretised in order to obtain the equations for the measurable and unobservable processes. Then the algorithm to apply the Kalman filter to estimate the parameters of the shot noise process is developed.

4.2.1 The state equations

In order to set up the state equations, we will discretise the Kalman Bucy approximation of the shot noise intensity and shot noise Cox processes. Assuming that $\Lambda(0)$ is Gamma distributed with parameters $\frac{\rho}{k}$ and η and for large ρ, from Dassios and Jang (2005), $\Lambda(t)$ and $N(t)$ are approximated with the following Gaussian processes:

$$Z(t) \approx \frac{\Lambda(t) - \frac{\rho}{\eta k}}{\sqrt{\frac{\rho}{\eta^2 k}}} \tag{4.2.1}$$

$$W(t) \approx \frac{N(t) - \frac{\rho}{\eta k} t}{\sqrt{\frac{\rho}{\eta^2 k}}} \tag{4.2.2}$$

So we have two new processes with $W(t)$ being the measured process and $Z(t)$ being the unobserved process. The large ρ assumption is necessary for the approximation. This is because a large ρ implies a high number of jumps in the intensity process and as ρ tends towards infinity, the shot noise process will have jumps at almost every instant which means the process will more closely resemble a diffusion process.

Dassios and Jang (2005) showed that the Kalman Bucy equations for the above processes

are given by:

$$dZ(t) = -kZ(t)dt + \sqrt{2k}dB_1(t) \tag{4.2.3}$$

$$dW(t) = Z(t)dt + \sqrt{\eta}dB_2(t) \tag{4.2.4}$$

where $B_1(t)$ and $B_2(t)$ are independent standard Brownian motion. Then Equations (4.2.3) and (4.2.4) are then discretised (via Euler's scheme) to the following:

$$\begin{aligned} Z_{t+\Delta} - Z_t &= -kZ_t + \sqrt{2k}\Delta B_1 \\ W_{t+\Delta} - W_t &= Z_t + \sqrt{\eta}\Delta B_2 \end{aligned}$$

where B_1 and B_2 are independent standard normal random variables. In the left hand side of the second equation in the above, the increment $W_{t+\Delta} - W(t)$ can be seen as the Gaussian approximation of the increment in the Cox process $N(t+\Delta) - N(t)$ where we let:

$$\begin{aligned} Y(t+\Delta) &= W_{t+\Delta} - W_t \\ &\approx \frac{N(t+\Delta) - \frac{\rho}{\eta k}(t+\Delta)}{\sqrt{\frac{\rho}{\eta^2 k}}} - \frac{N(t) - \frac{\rho}{\eta k}t}{\sqrt{\frac{\rho}{\eta^2 k}}} \\ &= \frac{N(t+\Delta) - N(t) - \frac{\rho}{\eta k}\Delta}{\sqrt{\frac{\rho}{\eta^2 k}}} \end{aligned}$$

Thus the measured process is the process of the increment of the shot noise Cox process $N(t+\Delta) - N(t)$ rather than the process $N(t)$ itself.

So by setting $\Delta = 1$ in the discretised process, the Kalman filter has the following state equations:

$$Z_{t+1} = (1-k)Z_t + \sqrt{2k}B_1 \tag{4.2.5}$$

$$Y_{t+1} = Z_t + \sqrt{\eta}B_2 \tag{4.2.6}$$

The unobservable shot noise intensity is represented by Z_{t+1} while the observed process Y_t represents the standardised increments in the count process $N(t+1) - N(t)$.

4.2.2 Kalman Filter algorithm

In this subsection, the Kalman filter algorithm to filter the shot noise intensity process from a measured process at times $1, 2, .., T$ is given. We first propose initial estimates for the unobserved process Z_0 and its covariance P_0. Then equations for both the time update and prediction update are given which are applied recursively. Finally, the likelihood function which is to be maximised to estimate the parameters is given. A summary of the procedure is as follows:

1. Based on the parameter values ρ, eta and k, transform the observed number of claims in each increments $N(t+1) - N(t)$ into a measurable process.
2. Obtain initial estimates of Z_0 and P_0.
3. Apply the time update and prediction update to obtain the filtered intensity.

For a particular set of parameters ρ, η and k, the number of claims in the time interval $[t-1, t]$ are aggregated for $t = 0, ..., T$. The first step of the procedure is to obtain initialise estimates for the unobserved state Z_0 and it's covariance at time 0, P_0. These estimates are based on the initial estimate of $\Lambda(0)$. Noting that:

$$\mathbb{E}[\Lambda(0)] = \frac{\rho}{\eta k} = \mathbb{E}[N(t+1) - N(t)] \quad \text{for all } t$$

an appropriate estimate for $\Lambda(0)$ is the sample mean of the aggregated claims in an increment given by:

$$\hat{\Lambda(0)} = \frac{\sum_{t=1}^{T} N_t}{T}$$

where N_t refers to the number of claims in the interval $[t-1, t]$. Also, noting that:

$$Var(\Lambda(0)) = \frac{\rho}{\eta^2 k}$$

which implies that:

$$Var(Z(0)) \approx \frac{Var(\Lambda(0))}{\frac{\rho}{\eta^2 k}} \approx 1$$

So we set $P_0 = 1$ as an initial estimate for the covariance of the unobserved process Z_t.

After initialisation, the time update is applied to Z_t. From the generalised algorithm provided in 2, the equations for the time update are the following:

1. The prior estimate of the unobserved process for next time step:

$$Z_{t|t-1} = (1-k)Z_{t-1}$$

2. The prior estimate for the covariance of the standardised intensity for next time step:

$$P_{t|t-1} = (1-k)^2 P_{t-1} + 2k$$

Similarly, the following equations for the prediction update can be obtained:

1. The Kalman gain:

$$K_t = \frac{P_{t|t-1}}{P_{t|t-1} + \eta}$$

2. The posterior estimate of the unobserved process:

$$Z_t = Z_{t|t-1} + K_t(Y_t - Z_{t|t-1})$$

where $Y_t - Z_{t|t-1}$ is known as the prediction residual.

3. The posterior covariance estimate:

$$P_t = (1 - K_t)P_{t|t-1}$$

The above time updates and prediction updates are applied for $t = 1, ...T$ to obtain estimates for the unobserved process $Z_1, Z_2, ..., Z_T$. The values of $P_{t|t-1}$ are stored as they will be used in the likelihood estimation of the process.In order to obtain the filtered estimates of the shot noise intensity process at $t = 1, ...T$, we substitute the $Z_1, Z_2, ..., Z_T$ into the following:

$$\Lambda(t) = \frac{\rho}{\eta k} + Z(t)\sqrt{\frac{\rho}{\eta^2 k}}$$

where the above arrives from rewriting Equation (4.2.1).

In order to estimate the parameters ρ, η and k, the method of maximum likelihood is used. Given that if the Gaussian assumption for Z_t and Y_t are valid, then the prediction residual $Y_t - Z_{t|t-1}$ is Gaussian with mean 0 and variance given by the prior estimates

$P_{t|t-1}$, Then the log likelihood function is given:

$$l(\rho, \eta, k) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log)(|P_{t|t-1}|) - \frac{1}{2}\sum_{t=1}^{T}\frac{(Y_t - Z_{t|t-1})^2}{P_{t|t-1}} \tag{4.2.7}$$

The above is then maximised with respect to the parameters ρ, η and k in order to obtain the maximum likelihood estimates.

4.2.3 Validity Test for the Kalman Filter approximation

As noted before, the validity of the parameter estimates depends on the validity of the Gaussian approximation of the intensity. One of the assumptions made for the likelihood estimation is the normality of the prediction residual $Y_t - Z_{t|t-1}$. So one way to test the validity of the parameter estimates is to test the normality of the prediction residual. Denote A_t as the standardised prediction residual. Then we test the null hypothesis:

$$\mathcal{H}_0 : A_t \sim N(0, 1)$$

The test statistic is given by the following Pearson correlation coefficient which compares the correlation between the sample quantiles $A(i)$ and theoretical quantiles $q(i)$ for $i = 1, ..., n$ of the standard normal distribution:

$$r_Q = \frac{\sum_{i=1}^{n}(A(i) - \bar{A})(q(i) - \bar{q})}{\left(\sum_{i=1}^{n}(A(i) - \bar{A})^2 \sum_{i=1}^{n}(q(i) - \bar{q})^2\right)}$$

where $\bar{A} = \frac{1}{n}\sum_{i=1}^{n} A(i)$ and $\bar{q} = \frac{1}{n}\sum_{i=1}^{n} q(i)$. For $\alpha = 0.05$, we have the following critical values for $\alpha = 0.05$:

n	50	100	200
$r_{Q,0.05}$	0.9768	0.9873	0.9931

Table 4.1: Critical values for r_Q at 5% level of significance

If A_t is standard normal, then the correlation between the sample quantiles and theoretical quantiles should be very close to 1. Denoting the α level critical value for the test statistic as $r_{Q,\alpha}$, we reject $(H)_0$ if $r_Q < r_{Q,\alpha}$ and the assumption and parameter estimates is invalid. Otherwise, the parameter estimates are valid.

The Kalman filter approximation provides a quick method to filter out the unobserved intensity to estimate the parameters. It is optimal if the Gaussian approximations for the

shot noise intensity and Cox process are valid. However, as Dassios and Jang (2005) points out, the approximation for the shot noise process with a Gaussian process is only valid for large ρ where the shot noise intensity will start to behaving like a diffusion process. Hence, a more robust and exact method using Markov Chain Monte Carlo method is proposed in the next section.

4.3 Reverse Jump Markov Chain Monte Carlo method

The proposed Reverse Jump Markov Chain Monte Carlo estimation method is based on the algorithm developed in Centanni and Minozzo (2006). It consists of two steps: filtering the sample intensity process and maximisation of the conditional likelihood function given the intensity process with respect to parameters via stochastic expectation maximisation.

In this section, we describe the Markov Chain Monte Carlo filtering algorithm from Centanni and Minozzo (2006) in further detail. For the stochastic expectation maximisation step, we derive an explicit form for the conditional log likelihood function. By deriving the gradient function and Hessian of the log likelihood function, we are able to reduce the three parameter optimisation problem to a two parameter optimisation problem.

4.3.1 Filtering of the intensity process

Based on Equation (3.1.2) the conditional intensity is characterised by the following terms:

- The number of jumps in the shot noise process n.
- The timing of the jumps of the shot noise process $\{\tau_1, ..., \tau_n\}$
- The size of the jumps of the shot noise process $\{X_1, ..., X_n\}$
- The starting position of the shot noise process $\Lambda(0)$

These terms will define the state which the Markov Chain is currently in. The states will change depending on the observed number of claim counts within each time interval. Typical of Markov Chain Monte Carlo methods, The algorithm can be decomposed into 3 steps:

1. Choosing a type of transition.
2. Simulating a transition to a new state.
3. Accepting the new state based on the acceptance probability A.

The above algorithm will be iterated over a large amount of times to ensure the Markov chain converges to a stationary distribution. There will also be a burn in period to ensure the Markov chain reaches a steady state. We will discuss choosing the number of iterations in more detail in the next chapter through the simulation study.

4.3.1.1 Choosing a type of transition

As the states are characterised by n, $\{\tau_1, ..., \tau_n\}$ and $\{X_1, ..., X_n\}$, the transitions will be based on changing one of these values. There are 5 types of transitions:

- Birth of new jump
- Shift in starting position
- Death of a current jump
- Jump position shift
- Jump height shift

For simplicity, it is assumed that if the number of jumps in the current state is 0, then $P(Birth) = P(Start) = 0.5$ while the probability of other transitions is zero. Otherwise, it shall be assumed the probability of entering into any state is 0.2.

4.3.1.2 Simulate a transition to a new state

We now describe the algorithm used to simulate a transition into a new state.

- For the birth of a new jump,
  1. Simulate a new jump time τ^* on the interval $[0, t]$ where $\tau* \sim U[0, t]$
  2. Simulate a new jump size X^* with an exponential distribution with parameter η
  3. Then noting that if $\tau_i \leq \tau^* \leq \tau_{i+1}$, we have the transitions:

$$\{\tau_1, ..., \tau_i, \tau_{i+1}, ...\tau_n\} \longrightarrow \{\tau_1, ..., \tau_i, \tau^*, \tau_{i+1}, ...\tau_n\}$$
$$\{X_1, ..., X_i, X_{i+1}, ...X_n\} \longrightarrow \{X_1, ..., X_i, X^*, X_{i+1}, ...X_n\}$$

- For the shift in starting position, simulate a gamma random variable with parameters $\frac{\rho}{k}$ and η as the new starting position $\Lambda^*(0)$

- For the death of a current jump:
  1. Given that there are currently m jumps, choose an integer l from 1 to m.
  2. Then:

$$\{\tau_1, ..., \tau_{l-1}, \tau_l, \tau_{l+1}...\tau_n\} \longrightarrow \{\tau_1, ..., \tau_{l-1}, \tau_{l+1}, ...\tau_n\}$$
$$\{X_1, ..., X_{l-1}, X_l, X_{l+1}...X_n\} \longrightarrow \{X_1, ..., X_{l-1}, X_{l+1}, ...X_n\}$$

- For jump position shift
  1. Given that there are currently m jumps, choose an integer l from 1 to m.
  2. Simulate a uniform random variable on $[\tau_{l-1}, \tau_{l+1}]$ to represent the new jump time τ^*.
  3. Then:

$$\{\tau_1, ..., \tau_{l-1}, \tau_l, \tau_{l+1}...\tau_n\} \longrightarrow \{\tau_1, ..., \tau_{l-1}, \tau^*, \tau_{l+1}, ...\tau_n\}$$

- For jump height shift
  1. Given that there are currently m jumps, choose an integer l from 1 to m.
  2. Simulate a new jump size X^* with an exponential distribution with parameter η.
  3. Then:

$$\{X_1, ..., X_{l-1}, X_l, X_{l+1}...X_n\} \longrightarrow \{X_1, ..., X_{l-1}, X^*, X_{l+1}...X_n\}$$

4.3.2 Acceptance of the new state

Let β as the vector representing the current state $\{\Lambda_0, \tau_1, ..., \tau_n, X_1, ..., X_n\}$ where n is the number of jumps in the shot noise process and let β' be the vector representing the next state with n' jumps. Then Centanni and Minozzo (2006) stated that the acceptance probability can be decomposed in the following way:

$$A = \underbrace{\frac{p(T_0^t|n', \beta')}{p(T_0^t|n, \beta)}}_{\text{likelihood ratio}} \times \underbrace{\frac{p(n', \beta')}{p(n, \beta)}}_{\text{prior ratio}} \times \times \text{ proposal ratio} \tag{4.3.1}$$

where T_0^t denotes the observed claim times on the interval $[0, t]$.

For the shot noise Cox process where it is assumed $\Lambda(0)$ is Gamma distributed, Centanni and Minozzo (2006) expresses the conditional distribution of T_0^t as the following:

$$\begin{aligned} p(T_0^t|n,\beta) &= \left[\prod_{i=1}^{N_t} \Lambda(T_i) \exp\left(-\int_{T_{i-1}}^{T_i} \Lambda(s)ds\right)\right] \exp\left(-\int_{T_n}^{t} \Lambda(s)ds\right) \\ &= \left[\prod_{i=1}^{N_t} \Lambda(T_i)\right] \exp\left(-\int_{0}^{t} \Lambda(s)ds\right) \end{aligned}$$

where N_t denotes the number of claims on the interval $[0, t]$ and noting that $T_0 = 0$. Given the distributional assumptions of jump times and sizes of the shot noise process stated in the definition of the shot noise process, the prior distribution can also be expressed as:

$$\begin{aligned} &\frac{\eta^{\frac{\rho}{k}} \Lambda(0)^{\frac{\rho}{k}-1} \exp(-\eta\Lambda(0))}{\Gamma(\frac{\rho}{k})} \left[\prod_{j=1}^{n} \eta \exp(-\eta X_j) \rho \exp(-\rho(\tau_j - \tau_{j-1}))\right] \exp(-\rho(t-\tau_n)) \\ &= \frac{\eta^{\frac{\rho}{k}} \Lambda(0)^{\frac{\rho}{k}-1} \exp(-\eta\Lambda(0))}{\Gamma(\frac{\rho}{k})} \left[\prod_{j=1}^{n} \eta \exp(-\eta X_j)\right] \rho^n \exp(-\rho t) \end{aligned}$$

From the equation of the prior distribution, for each transition type, we can deduce the following prior ratios:

- Birth of new jump - Since the only change lies in the extra jump size X^* in the product, then it can be deduced that the prior ratio is given by:

$$\begin{aligned} \frac{p(n',\beta')}{p(n,\beta)} &= \frac{\frac{\eta^{\frac{\rho}{k}} \Lambda(0)^{\frac{\rho}{k}-1} \exp(-\eta\Lambda(0))}{\Gamma(\frac{\rho}{k})} \left[\prod_{j=1}^{n+1} \eta \exp(-\eta X_j)\right] \rho^{n+1} \exp(-\rho t)}{\frac{\eta^{\frac{\rho}{k}} \Lambda(0)^{\frac{\rho}{k}-1} \exp(-\eta\Lambda(0))}{\Gamma(\frac{\rho}{k})} \left[\prod_{j=1}^{n} \eta \exp(-\eta X_j)\right] \rho^n \exp(-\rho t)} \\ &= \frac{\left[\prod_{j=1}^{n+1} \eta \exp(-\eta X_j)\right] \rho}{\left[\prod_{j=1}^{n} \eta \exp(-\eta X_j)\right]} \\ &= \rho\eta \exp(-\eta X^*) \end{aligned}$$

- Shift in starting position - Since the size and timing of the jumps don't change, we

have:

$$\begin{aligned}\frac{p(n',\beta')}{p(n,\beta)} &= \frac{\frac{\eta^{\frac{\rho}{k}}\Lambda^*(0)^{\frac{\rho}{k}-1}\exp(-\eta\Lambda^*(0))}{\Gamma(\frac{\rho}{k})}\left[\prod_{j=1}^{n+1}\eta\exp(-\eta X_j)\right]\rho^n\exp(-\rho t)}{\frac{\eta^{\frac{\rho}{k}}\Lambda(0)^{\frac{\rho}{k}-1}\exp(-\eta\Lambda(0))}{\Gamma(\frac{\rho}{k})}\left[\prod_{j=1}^{n}\eta\exp(-\eta X_j)\right]\rho^n\exp(-\rho t)} \\ &= \exp(-\eta(\Lambda^*(0)-\Lambda(0))\left(\frac{\Lambda^*(0)}{\Lambda(0)}\right)^{\frac{\rho}{k}-1}\end{aligned}$$

- Death of a current jump - Since the only change lies in the one less jump size X_k in the product, then it can be deduced that the prior ratio is given by

$$\begin{aligned}\frac{p(n',\beta')}{p(n,\beta)} &= \frac{\frac{\eta^{\frac{\rho}{k}}\Lambda(0)^{\frac{\rho}{k}-1}\exp(-\eta\Lambda(0))}{\Gamma(\frac{\rho}{k})}\left[\prod_{j=1,j\neq k}^{n}\eta\exp(-\eta X_j)\right]\rho^{n-1}\exp(-\rho t)}{\frac{\eta^{\frac{\rho}{k}}\Lambda(0)^{\frac{\rho}{k}-1}\exp(-\eta\Lambda(0))}{\Gamma(\frac{\rho}{k})}\left[\prod_{j=1}^{n}\eta\exp(-\eta X_j)\right]\rho^n\exp(-\rho t)} \\ &= \frac{\left[\prod_{j=1,j\neq k}^{n}\eta\exp(-\eta X_j)\right]}{\left[\prod_{j=1}^{n}\eta\exp(-\eta X_j)\right]\rho} \\ &= (\rho\eta\exp(-\eta X^*))^{-1}\end{aligned}$$

- Jump position shift - Since the jump height and starting position both remain unchanged, then we have:

$$\frac{p(n',\beta')}{p(n,\beta)} = \frac{\frac{\eta^{\frac{\rho}{k}}\Lambda(0)^{\frac{\rho}{k}-1}\exp(-\eta\Lambda(0))}{\Gamma(\frac{\rho}{k})}\left[\prod_{j=1}^{n}\eta\exp(-\eta X_j)\right]\rho^n\exp(-\rho t)}{\frac{\eta^{\frac{\rho}{k}}\Lambda(0)^{\frac{\rho}{k}-1}\exp(-\eta\Lambda(0))}{\Gamma(\frac{\rho}{k})}\left[\prod_{j=1}^{n}\eta\exp(-\eta X_j)\right]\rho^n\exp(-\rho t)} = 1$$

- Jump height shift - Since the only change lies in the changed jump size X_k to X^* in the product, we have:

$$\begin{aligned}\frac{p(n',\beta')}{p(n,\beta)} &= \frac{\frac{\eta^{\frac{\rho}{k}}\Lambda(0)^{\frac{\rho}{k}-1}\exp(-\eta\Lambda(0))}{\Gamma(\frac{\rho}{k})}\left[\prod_{j=1,j\neq k}^{n}\eta\exp(-\eta X_j)\right]\eta\exp(-\eta X^*)\rho^n\exp(-\rho t)}{\frac{\eta^{\frac{\rho}{k}}\Lambda(0)^{\frac{\rho}{k}-1}\exp(-\eta\Lambda(0))}{\Gamma(\frac{\rho}{k})}\left[\prod_{j=1}^{n}\eta\exp(-\eta X_j)\right]\rho^n\exp(-\rho t)} \\ &= \frac{\exp(-\eta X^*)}{\exp(-\eta X_j)} = \exp(-\eta(X^*-X_j))\end{aligned}$$

For each types of transition, Centanni and Minozzo (2006) provides the following proposal ratios:

- Birth of new jump:

$$\frac{tp(d|n+1)}{(n+1)p(b|n)\eta\exp(-\eta X^*)}$$

- Shift in starting position

$$\exp(-\eta(\Lambda(0)-\Lambda^*(0))\left(\frac{\Lambda(0)}{\Lambda^*(0)}\right)^{\frac{\rho}{k}-1}$$

- Death of a current jump

$$\frac{np(b|n-1)\eta\exp(-\eta X^*)}{tp(d|n)}$$

- The proposal ratio of the jump position shift is equal to 1.
- Jump height shift

$$\exp(-\eta(X_l - X_l^*))$$

Hence the acceptance probability A for each transition state is given by:

- Birth of new jump:

$$\begin{aligned} A &= \frac{\left[\prod_{i=1}^{N_t}\Lambda^*(T_i)\right]\exp\left(-\int_0^t\Lambda^*(s)ds\right)}{\left[\prod_{i=1}^{N_t}\Lambda(T_i)\right]\exp\left(-\int_0^t\Lambda(s)ds\right)}\times\eta\rho\exp(-\eta X^*)\times\frac{tp(d|n+1)}{(n+1)p(b|n)\eta\exp(-\eta X^*)} \\ &= \frac{\left[\prod_{i=1}^{N_t}\Lambda^*(T_i)\right]\exp\left(-\int_0^t\Lambda^*(s)ds\right)}{\left[\prod_{i=1}^{N_t}\Lambda(T_i)\right]\exp\left(-\int_0^t\Lambda(s)ds\right)}\times\frac{t\rho}{n+1} \end{aligned}$$

  where X^* refers to the size of the new jump and $\Lambda^*(t)$ represent the estimator for the intensity function given the new jump size and jump time.

- Shift in starting position

$$\begin{aligned} A &= \frac{\left[\prod_{i=1}^{N_t} \Lambda^*(T_i)\right] \exp\left(-\int_0^t \Lambda^*(s)ds\right)}{\left[\prod_{i=1}^{N_t} \Lambda(T_i)\right] \exp\left(-\int_0^t \Lambda(s)ds\right)} \times \exp(-\eta(\Lambda^*(0) - \Lambda(0)) \left(\frac{\Lambda^*(0)}{\Lambda(0)}\right)^{\frac{\rho}{k}-1} \\ &\times \exp(-\eta(\Lambda(0) - \Lambda^*(0)) \left(\frac{\Lambda(0)}{\Lambda^*(0)}\right)^{\frac{\rho}{k}-1} \\ &= \frac{\left[\prod_{i=1}^{N_t} \Lambda^*(T_i)\right] \exp\left(-\int_0^t \Lambda^*(s)ds\right)}{\left[\prod_{i=1}^{N_t} \Lambda(T_i)\right] \exp\left(-\int_0^t \Lambda(s)ds\right)} \end{aligned}$$

  where $\Lambda^*(t)$ represent the estimator for the intensity function given the new jump size and jump time.

- Death of a current jump:

$$\begin{aligned} A &= \frac{\left[\prod_{i=1}^{N_t} \Lambda^*(T_i)\right] \exp\left(-\int_0^t \Lambda^*(s)ds\right)}{\left[\prod_{i=1}^{N_t} \Lambda(T_i)\right] \exp\left(-\int_0^t \Lambda(s)ds\right)} \times (\eta\rho \exp(-\eta X^*))^{-1} \times \frac{np(b|n-1)\eta \exp(-\eta X^*)}{tp(d|n)} \\ &= \frac{\left[\prod_{i=1}^{N_t} \Lambda^*(T_i)\right] \exp\left(-\int_0^t \Lambda^*(s)ds\right)}{\left[\prod_{i=1}^{N_t} \Lambda(T_i)\right] \exp\left(-\int_0^t \Lambda(s)ds\right)} \times \frac{n}{t\rho} \end{aligned}$$

  where X^* refers to the size of the suppressed jump and $\Lambda^*(t)$ represent the estimator for the intensity function given the new jump size and jump time.

- Jump position shift:

$$\begin{aligned} A &= \frac{\left[\prod_{i=1}^{N_t} \Lambda^*(T_i)\right] \exp\left(-\int_0^t \Lambda^*(s)ds\right)}{\left[\prod_{i=1}^{N_t} \Lambda(T_i)\right] \exp\left(-\int_0^t \Lambda(s)ds\right)} \times 1 \times 1 \\ &= \frac{\left[\prod_{i=1}^{N_t} \Lambda^*(T_i)\right] \exp\left(-\int_0^t \Lambda^*(s)ds\right)}{\left[\prod_{i=1}^{N_t} \Lambda(T_i)\right] \exp\left(-\int_0^t \Lambda(s)ds\right)} \end{aligned}$$

  where $\Lambda^*(t)$ represent the estimator for the intensity function given the new jump size and jump time.

- Jump height shift:

$$A = \frac{\left[\prod_{i=1}^{N_t} \Lambda^*(T_i)\right] \exp\left(-\int_0^t \Lambda^*(s)ds\right)}{\left[\prod_{i=1}^{N_t} \Lambda(T_i)\right] \exp\left(-\int_0^t \Lambda(s)ds\right)} \times \exp(-\eta(X_l^* - X_l)) \times \exp(-\eta(X_l - X_l^*))$$
$$= \frac{\left[\prod_{i=1}^{N_t} \Lambda^*(T_i)\right] \exp\left(-\int_0^t \Lambda^*(s)ds\right)}{\left[\prod_{i=1}^{N_t} \Lambda(T_i)\right] \exp\left(-\int_0^t \Lambda(s)ds\right)}$$

  where X^* refers to the changed size of the jump and $\Lambda^*(t)$ represent the estimator for the intensity function given the new jump size and jump time.

In order to simplify the calculations in R, the logarithms of the above expressions were taken. Hence the acceptance algorithm becomes:

1. Generate $\log(U)$ where $U \sim U[0,1]$
2. If $\log(U) \leq \log(A)$, the new state is accepted. Otherwise the old state is retained.

4.3.3 Stochastic Expectation Maximisation Algorithm

Once a sample of the intensity function is produced from a large number of iterations of the Markov Chain Monte Carlo algorithm, the expectation maximisation algorithm is used to determine the parameters. Centanni and Minozzo (2006) states the algorithm is as following way:

1. Generate a S_r sample from the RJMCMC algorithm
2. Calculate the likelihood function to be maximised is:

$$L(\theta) = \frac{1}{S_r} \sum_{i=1}^{S_r} f(\mathbf{T}, N_t, \mathbf{X}, \tau.J_t : \theta)$$
$$= \frac{1}{S_r} \sum_{i=1}^{S_r} P(T_{N+1} > t | \mathbf{T}, \mathbf{X}, \tau.J_t : \theta) \times f(\mathbf{T}|\mathbf{X}, \tau.J_t : \theta) \times f(\mathbf{X}, \tau.J_t : \theta)$$

3. Maximise the above functions with respect to the parameters, which are η, ρ and k.

We now provide the explicit form for the log-likelihood function. From the previous section, we can see that:

$$P(T_{N+1} > t|\mathbf{T}, \mathbf{X}, \tau.J_t : \theta) f(\mathbf{T}|\mathbf{X}, \tau.J_t : \theta) = p(\mathbf{T}_0^t|n, \beta)$$
$$= \left[\prod_{i=1}^{N_t} \Lambda(T_i)\right] \exp\left(-\int_0^t \Lambda(s)ds\right)$$

where the first part is the density of a gamma random variable with parameters ρ/k and η

$$f(\mathbf{X}, \tau.J_t : \theta) = p(n, \beta)$$
$$= \frac{\eta^{\frac{\rho}{k}} \Lambda(0)^{\frac{\rho}{k}-1} \exp(-\eta\Lambda(0))}{\Gamma(\frac{\rho}{k})} \left[\prod_{j=1}^{n} \eta \exp(-\eta X_j)\right] \rho^n \exp(-\rho t)$$

Hence for claim arrival times T_i such that $0 = T_0 < T_1 < ... < T_{N(t)} < t$, consider the log-likelihood process:

$$\begin{aligned}
\log(L(\rho, \eta, k)) = \log &\left[\left(\prod_{i=1}^{N(t)} \Lambda(T_i)\right) \exp\left(-\int_0^t \Lambda(s)ds\right)\right. \\
&\left. \times \frac{\eta^{\frac{\rho}{k}} \Lambda(0)^{\frac{\rho}{k}-1} \exp(-\eta\Lambda(0))}{\Gamma(\frac{\rho}{k})} \left[\prod_{j=1}^{n} \eta \exp(-\eta X_j)\right] \rho^n \exp(-\rho t)\right] \\
&\approx \sum_{i=1}^{N(t)} N_i \left[-kT_i + \log\left(\Lambda(0) + \sum_{j=1}^{n} X_j e^{k(\tau_i)} \mathbf{1}_{\tau_j \leq T_i}\right)\right] - \frac{\Lambda(0)(1 - e^{-kt})}{k} \\
&- \sum_{j=1}^{n} \frac{X_j(1 - e^{-k(t-\tau_j)})}{k} + \frac{\rho}{k}\log(\eta) + \left(\frac{\rho}{k} - 1\right)\Lambda(0) - \eta\Lambda(0) - \log(\Gamma\left[\frac{\rho}{k}\right]) \\
&+ n\log(\eta) - \eta\sum_{j=1}^{n} X_j + n\log(\rho) - \rho t \qquad (4.3.2)
\end{aligned}$$

where N_i denotes the number of claim counts in the interval $[T_{i-1}, T_i]$. This approximation is required as in insurance data, there are multiple claims in the same day as we cannot observe the exact time that the claim has incurred. So we let $\Lambda(T_i)$ represents the intensity process from T_{i-1} to T_i).

For $S_r = 1$, this represents the stochastic expectation maximisation algorithm as noted in Centanni and Minozzo (2006). For large values of S_r, this algorithm becomes similar to that of the deterministic expectation maximisation algorithm. Although Centanni and

Minozzo (2006) states that the stochastic expectation maximisation algorithm does not guarantee pointwise convergence, applying a deterministic expectation maximisation algorithm means that multiple samples of the jump sizes, jump times and starting values which represent the state the Markov chain is in need to be stored. As S_r increases, storing multiple samples of the becomes unfeasible while computation time becomes unreasonable, especially with lengthy data where the number of jump sizes and jump times are large. Hence, we choose to sacrifice some accuracy to ensure that the method remains efficient.

4.3.4 Reduction to two parameter optimisation

In this section, we derive the gradient vector and Hessian matrix of the log-likelihood function. The main reason behind this is that it allows the use of computationally quicker optimisation methods such as the quasi-Newton method which requires the gradient and Hessian matrices. In the process, we are able to derive the conditional maximum likelihood estimate of the jump size parameter η with respect to the other 2 parameters and other simulated variables. Hence, this reduces the optimisation problem from three parameters to two parameters.

Then the gradient vector will be $\left(\frac{\partial \log(L)}{\partial \rho}, \frac{\partial \log(L)}{\partial \eta}, \frac{\partial \log(L)}{\partial k}\right)^T$ where:

$$\frac{\partial \log(L)}{\partial \rho} = \frac{\log(\eta)}{k} + \Lambda(0) - \frac{1}{k}\psi\left(\frac{\rho}{k}\right) + \frac{n}{\rho} - t \tag{4.3.3}$$

$$\frac{\partial \log(L)}{\partial \eta} = \frac{\rho}{k\eta} - \Lambda(0) + \frac{n}{\eta} - \sum_{j=1}^{n} X_j \tag{4.3.4}$$

$$\begin{aligned}\frac{\partial \log(L)}{\partial k} = &\sum_{i=1}^{N(t)} N_i \left(-T_i + \frac{\sum_{j=1}^{n} X_j \tau_j e^{k\tau_j} \mathbf{1}_{\tau_j \leq T_i}}{\Lambda(0) + \sum_{j=1}^{n} X_j e^{k\tau_j} \mathbf{1}_{\tau_j \leq T_i}}\right) + \frac{\Lambda(0)(1 - e^{-kt} + kte^{-kt})}{k^2} \\ &+ \sum_{j=1}^{n} \frac{\Lambda(0)(1 - e^{-k(t-\tau_j)} + k(t-\tau_j)e^{-k(t-\tau_j)})}{k^2} - \frac{\rho(\log(\eta) + \Lambda(0))}{k^2} + \frac{\rho}{k^2}\psi\left(\frac{\rho}{k}\right)\end{aligned} \tag{4.3.5}$$

where $\psi(x) = \frac{d}{dx}\log(\Gamma(x))$. Note that equating Equation (4.3.5) to zero implies that:

$$\eta = \frac{\frac{\rho}{k} + n}{\Lambda(0) + \sum_{i=1}^{n} X_j} \tag{4.3.6}$$

which will always be positive provided that $\rho > 0$ and $k > 0$. We note that the expected value of the number of shots n is given by ρt, the expected value sum of the jumps $\sum_{i=1}^{n} X_j$ is $\frac{\rho t}{\eta}$ and under the Gamma assumption, the expected value of $\Lambda(0)$ is $\frac{\rho}{k}$. It can be seen

that by substituting these values into (4.3.6), the maximum likelihood estimate of η is actually consistent with these estimates of the simulated parameters.

Then the Hessian matrix can then be derived as following:

$$H = \begin{pmatrix} \frac{\partial^2 \log(L)}{\partial \rho^2} & \frac{\partial^2 \log(L)}{\partial \rho \partial \eta} & \frac{\partial^2 \log(L)}{\partial \rho \partial k} \\ \frac{\partial^2 \log(L)}{\partial \rho \partial \eta} & \frac{\partial^2 \log(L)}{\partial \eta^2} & \frac{\partial^2 \log(L)}{\partial \eta \partial k} \\ \frac{\partial^2 \log(L)}{\partial \rho \partial k} & \frac{\partial^2 \log(L)}{\partial \eta \partial k} & \frac{\partial^2 \log(L)}{\partial k^2} \end{pmatrix}$$

where:

$$\begin{aligned}
\frac{\partial^2 \log(L)}{\partial \rho^2} &= -\frac{n}{\rho^2} - \frac{1}{k^2}\psi_1\left(\frac{\rho}{\gamma}\right) \\
\frac{\partial^2 \log(L)}{\partial \rho \partial \eta} &= \frac{1}{k\eta} \\
\frac{\partial^2 \log(L)}{\partial \rho \partial k} &= -\frac{\log(\eta)}{k^2} + \frac{1}{k^2}\psi\left(\frac{\rho}{k}\right) + \frac{\rho}{k^3}\psi_1\left(\frac{\rho}{k}\right) \\
\frac{\partial^2 \log(L)}{\partial \eta^2} &= -\frac{\rho}{k\eta^2} - \frac{n}{\eta^2} \\
\frac{\partial^2 \log(L)}{\partial \eta \partial k} &= -\frac{\rho}{k^2\eta} \\
\frac{\partial^2 \log(L)}{\partial k^2} &= \sum_{i=1}^{N(t)} N_i \left[\frac{\sum_{j=1}^n X_j \tau_j^2 e^{k\tau_j} \mathbf{1}_{\tau_j \le T_i}}{\Lambda(0) + \sum_{j=1}^n X_j e^{k\tau_j} \mathbf{1}_{\tau_j \le T_i}} + \left(\frac{\sum_{j=1}^n X_j \tau_j e^{k\tau_j} \mathbf{1}_{\tau_j \le T_i}}{\Lambda(0) + \sum_{j=1}^n X_j e^{k\tau_j} \mathbf{1}_{\tau_j \le T_i}} \right)^2 \right] \\
&\quad + \frac{\Lambda(0)(2e^{-kt} - k^2 t^2 e^{-kt} - 2)}{k^3} + \sum_{j=1}^n \frac{X_j(2e^{-k(t-\tau_j)} - k^2(t-\tau_j)^2 e^{-k(t-\tau_j)} - 2)}{k^3} \\
&\quad + \frac{2\rho(\log(\eta) + \Lambda(0))}{k^3} - \frac{2\rho}{k^3}\psi\left(\frac{\rho}{k}\right) - \frac{\rho^2}{k^4}\psi_1\left(\frac{\rho}{k}\right)
\end{aligned}$$

where $\psi_1(x) = \frac{d^2}{dx^2}\log(\Gamma(x))$ Noting that:

$$\frac{\partial^2 \log(L)}{\partial \eta^2} = -\frac{\rho}{k\eta^2} - \frac{n}{\eta^2} < 0$$

which implies that (4.3.6) maximises L with respect to η.

4.4 Goodness of fit tests

Generally, likelihood is the one of the main comparison tool of goodness of fit from various models or various fitting methods. However, the likelihood function for the Markov

Chain Monte Carlo method is based on the conditional probability of observing the claim process, while the likelihood function for the Kalman filter looks at the distribution for the prediction residual. Hence using the likelihood value to compare the goodness of fit for both methods is not valid. In this section, we propose some heuristic methods to check the goodness of fit of parameters as well as provide more formal goodness of fit tests based on distance tests Dachian and Kutoyants (2008).

4.4.1 Comparing moments and autocovariance functions

We first provide some heuristic methods to check the goodness of fit of the shot noise Cox process based on the parameter estimates. Although these methods have not been proven in a formal, they still provide useful information for the modeller. A quick diagnostic for goodness of fit is to compare the implied mean and variance of number of claims in an increment based on the parameter estimates compared to sample mean and variance from the dataset. Another heuristic method which is inspired to the minimal contrast estimation method applied in physical sciences is to compare the implied autocovariance function derived in Equation (3.2.11) with the empirical autocovariance in the time interval $[0, T]$. The empirical autocovariance function is given by the following:

$$\hat{\gamma}(h) = \frac{1}{T} \sum_{t=1}^{T-|h|} (N_{t+|h|} - \bar{N})(N_t - \bar{N})$$

where N_t represents the number of counts in the time interval $[t-1, t]$ while $\bar{N}$ represents the average number of counts over all increments. If the model fits well, the model should exhibit a similar covariance structure to the covariance structure of the data.

4.4.2 Distance statistics

More formally, we wish to test the following:

$$\mathcal{H}_0 : \Lambda(t) = \tilde{\Lambda}_t \quad t \geq 0 \quad \text{vs} \quad \mathcal{H}_1 : \Lambda(t) \neq \tilde{\Lambda}_t \quad t \geq 0$$

Then consider the empirical mean value process for a dataset divided into n sub-intervals defined in the following way:

$$\hat{M}(t) = \frac{1}{n-t} \sum_{i=1}^{n-t+1} \#\text{ of events in the time interval } [i, i+t-1) \tag{4.4.1}$$

Then the two statistics proposed in Dachian and Kutoyants (2008) which can be used to compare goodness of fit between the models are:

$$D_n = \frac{1}{\tilde{M}(t)} \sup_{0 \le s \le t} |\hat{M}(s) - \tilde{M}(s)| \tag{4.4.2}$$

$$V_n^2 = \frac{n}{\tilde{M}(t)^2} \int_0^t (\hat{M}(s) - \tilde{M}(s))^2 d\tilde{M}(s) \tag{4.4.3}$$

where $\tilde{M}(t)$ denotes the fitted mean value process. Noting that the that for both the Markov Chain Monte Carlo simulation and the Kalman filter, we can consider the fitted mean

These statistics is analoguous to the Kolmogorov Smirnov and Anderson Darling test statistics and hence have a similar interpretation. A smaller value in the distance statistics indicate a smaller discrepancy of the fitted value from the model and hence a better fit. The Kolmogorov Smirnov test however focuses only on the largest discrepancy of the fitted mean value process while the Anderson Darling test statistic focuses on the overall discrepancy of the fitted mean value process.

CHAPTER 5

COMPARATIVE STUDY OF MODEL FITTING METHODS

This chapter provides a simulation study in order to compare the methods discussed in the previous chapter to fit claim counts data. The study also highlights computational issues involved in applying both methods which are discussed and any calibrations made to improve the efficiency and accuracy of the methods are justified through this study. More specifically, for the quicker Kalman filter approximation method, the study also serves to observe the importance of the large ρ assumption on the validity of the estimates. For the more computationally intensive Markov Chain Monte Carlo method, we propose methods to overcome convergence issues in implementation as well as ways to improve efficiency of the algorithm without sacrificing accuracy of estimates.

Section 5.1 proposes an algorithm for simulating sample paths of the shot noise intensity for given parameter values ρ, η and k which is then used to simulate the shot noise Cox process. For the comparative study, we consider two cases based on the relative frequency of arrival of the shots in the intensity process. Since the parameter ρ represents the average number of shots in 1 time unit, the frequency is reflected in the value of ρ. The relatively low frequency case is reflected by a small ρ while a higher frequency case is reflected by a higher ρ. In Section 5.2, the Kalman filter is used to fit the shot noise Cox process to both cases while in section 5.3, the Markov Chain Monte Carlo method is

used. A summary of the comparative performance of both methods is then provided in Section 5.4 using goodness of fit tests provided in Chapter 5.2.1. Implementation of this case study is all done in the statistical software package R.

5.1 Simulating the shot noise Cox process

In this section, we propose a method for simulating a sample path of the shot noise intensity process. Then arrival times of a inhomogeneous Poisson process can be simulated from the sample path in order to generate a sample of arrival times for the shot noise Cox process. Then the arrival times are aggregated to represent the number of events per time interval.

In order to simulate the claim count process $N(t)$ over a time interval $[0, T]$ with parameters ρ, η and k, the intensity process $\Lambda(t)$ is generated in the following way:

1. Simulate the number of jumps $J(T)$ with a Poisson (ρT) random variable.
2. For the j^{th} jump, simulate the jump size X_j with an Exponential random variable with parameter η.
3. Conditional on the number of jumps in a homogeneous Poisson process in the time interval $[0, T]$, the arrival times have a uniform $(0, T)$ distribution. Hence for the j^{th} jump, simulate the conditional jump time τ_j with a Uniform $(0, T)$ distribution.
4. For $t = 0, t_1, t_2, ..., T$, a realisation of the intensity can be found by the following formula:

$$\Lambda(t) = \Lambda(0)e^{-kt} + \sum_{j=1}^{J(T)} X_j e^{-k(t-\tau_j)} \mathbf{1}_{t>\tau_j}$$

Once the realisation for the intensity $\Lambda(t)$ is found, the number of claims per period can be simulated by adapting the simulation algorithm for claim arrival times described in Čížek et al. (2011):

1. Set $\bar{\Lambda} = \max(\Lambda(t))$
2. Set $T^* = 0$
3. While $T^* < T$:
   (a) Generate an exponential random variable E with intensity $\bar{\Lambda}$.

(b) Set $T^* = T^* + E$

(c) Generate a uniform $[0, 1]$ random variable U

(d) If $U \leq \frac{\Lambda(T^*)}{\bar{\Lambda}}$, T^* is recorded as a claim arrival time.

The above algorithm generates the arrival times of the shot noise Cox process. For modelling purposes however, the aggregate number of events over various time intervals are needed. Hence we partition the time interval $[0, T]$ into n intervals $[t_i, t_{i+1}]$, $i = 0, .., n-1$. Then the number of claim arrival times lying in $[t_i, t_{i+1}]$ will represent the number of claims in that time interval denoted by the increment $N(t_{i+1}) - N(t_i)$.

For the comparative study, the parameters for the low shot frequency case are $\rho = 1$, $\eta = 1$ and $k = 1$ while for the high frequency case, the parameters are $\rho = 100$, $\eta = 1$ and $k = 1$. Using the above algorithm, sample paths for the shot noise intensity $\Lambda(t)$ and Cox process $N(t)$ for the times $t = 1, 2, .., 100$. The simulated sample paths of the shot noise intensity $\Lambda(t)$ and Cox process $N(t)$ for the low shot frequency case are shown in Figure 5.1 and in Figure 5.2 for the high shot frequency case. It can be seen that for the low frequency case, the evolution of the Cox process $N(t)$ is much less consistent compared to when ρ is large. This is due to the fact for smaller ρ, the shot noise intensity will decay for longer before another shot in the intensity. Hence the sudden acceleration in the growth of the count process $N(t)$ due to peaks in the shot noise intensity $\Lambda(t)$ is much more apparent in the case when ρ is small than when ρ is large.

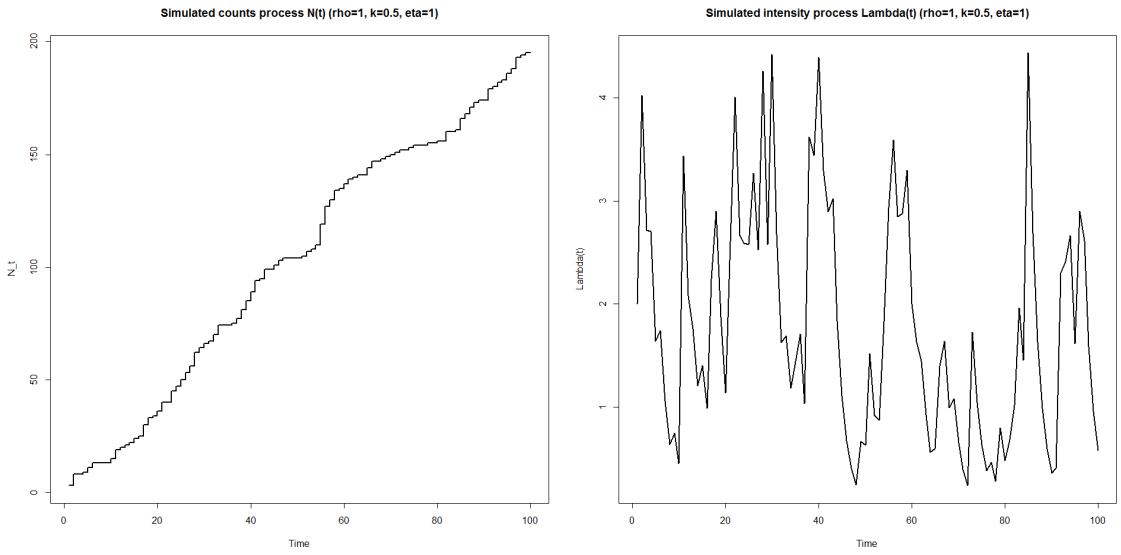

Figure 5.1: Simulated Intensity process $\Lambda(t)$ and counts process $N(t)$ ($\rho = 1$, $\eta = 1$ and $k = 1$)

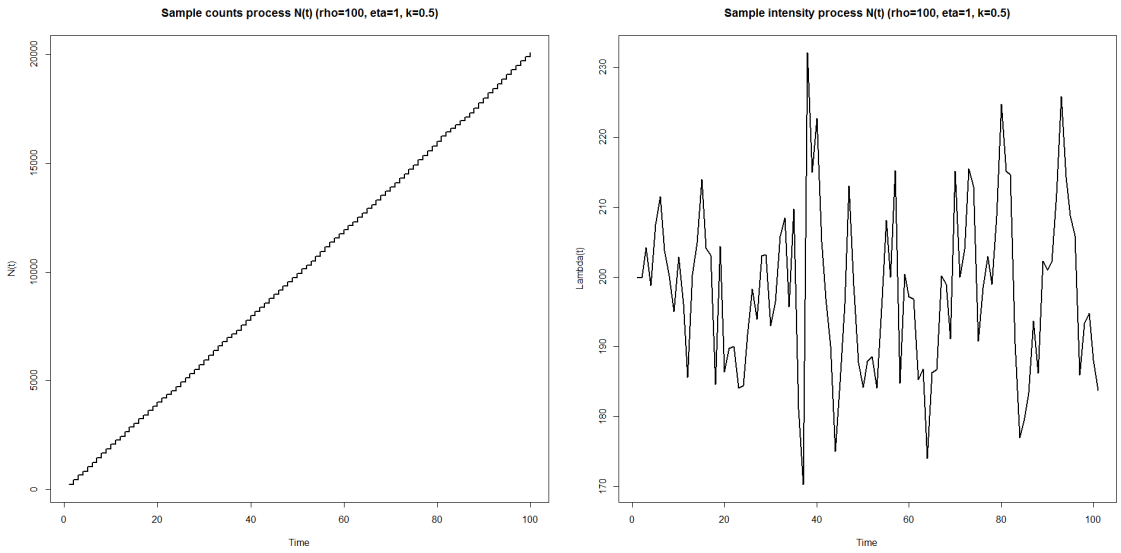

Figure 5.2: Simulated Intensity process $\Lambda(t)$ and counts process $N(t)$ ($\rho = 100$, $\eta = 1$ and $k = 1$)

5.2 Kalman filter

In this section, the shot noise Cox process is fitted to the simulated aggregated claims datasets using the Kalman filter method developed in 5.2.1. In both cases, the validity of the Gaussian approximation for the Kalman filter is tested using the method described in Chapter 5.2.1, where the importance of the large ρ assumption becomes apparent.

5.2.1 Low frequency case

In the case where $\rho = 1$, $\eta = 1$ and $k = 0.5$, we first obtain the initial estimates using the method described in Section 4.1 in Chapter :

$$\rho_0 = 1.12, \eta_0 = 1.23, k_0 = 0.6$$

For the Kalman filter, the final estimates are senstive to the initial estimates. To see this, the method is applied for different initial values of the parameters as shown in 5.1: It also shows that parameter estimates for ρ is sensitive to it's initial estimate. This highlights the importance of good initial estimates for the parameters and hence the developments made in Section 4.1 in 5.2.1 play a significant role in the estimation procedure.

Table 5.2 shows the estimates of each parameter with the standard errors of the parame-

Initial parameters	$\hat{\rho}$	$\hat{\rho}$	$\hat{\rho}$
$\rho_0 = 10, \eta_0 = 1, k_0 = 0.5$	6.50	1.65	0.65
$\rho_0 = 50, \eta_0 = 2, k_0 = 0.05$	38.10	8.29	1.85
$\rho_0 = 100, \eta_0 = 0.2, k_0 = 2$	78.11	9.56	2.62

Table 5.1: Final values for Kalman filter for different initial estimates

ters. The log-likelihood for the parameter estimates is $-\mathbf{1082.31}$. The filtered intensity process at the estimated parameters can be compare with the actual simulated intensity process. This is shown in Figure 5.3. The fitted intensity appears to match the true intensity process quite well.

	ρ (s.e)	η (s.e)	k (s.e)
Actual	1	1	0.5
Kalman fit	1.15 (0.093)	0.83 (0.045)	0.56 (0.027)

Table 5.2: Kalman filter parameter estimates for low shot frequency case

Now we address the validity of these estimates based on the assumption. Based from the QQ-plot in 5.4, it can be seen that the standardised prediction residuals of the Kalman filter does not appear to follow a standard normal distribution. More formally, the null hypothesis $\mathcal{H}_0 : r_Q \sim N(0, 1)$ is tested where r_Q is the Pearson correlation coefficient between the sample and theoretical quantiles. The test statistic is evaluated to be $r_Q = 0.77$ which is less than the critical value $r_{Q,\alpha} = 0.9873$ (see Table 4.1). Thus the null hypothesis is rejected and there is significant evidence that the prediction residuals do not follow a normal distribution. Hence, in the case for $\rho = 1$, the Gaussian approximation is not reliable and hence it can be seen that the estimates for a Kalman filter are not reliable.

5.2.2 High frequency case

Table 5.3 shows the estimates of each parameter with the standard errors of the parameters. The log-likelihood for the estimates is -2122.31. The filtered intensity process at the estimated parameters can be compare with the actual simulated intensity process. This is shown in Figure 5.9. Once again the fitted intensity appears to match the true intensity process quite well.

Once again, the validity of the Gaussian approximation will be tested in the case where we have large ρ, which is the condition Dassios and Jang (2005) stated the approximation is valid. Based from the QQ-plot in 5.6, apart from the upper tail, it can be seen that the

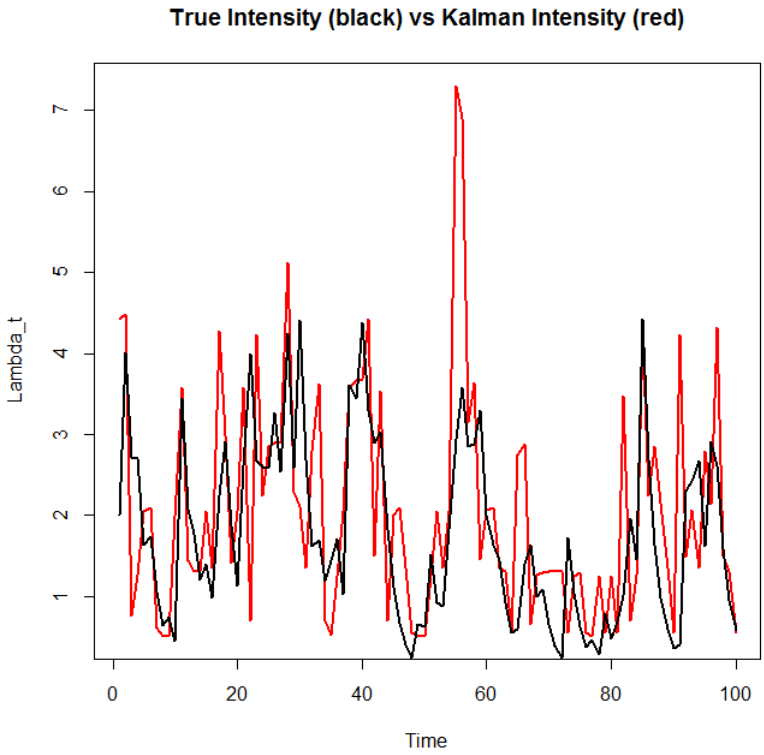

Figure 5.3: Fitted Kalman (red) vs Actual (black) intensity process ($\rho = 1$)

	ρ (s.e)	η (s.e)	k (s.e)
Actual	100	1	0.5
Kalman fit	91.84 (5.13)	0.75 (0.071)	0.59 (0.03)

Table 5.3: Kalman filter parameter estimates for high shot frequency case

standardised prediction residuals of the Kalman filter seem to follow a standard normal distribution. Once again, we formally test the null hypothesis $\mathcal{H}_0 : r_Q \sim N(0, 1)$ where r_Q is the Pearson correlation coefficient between the sample and theoretical quantiles. The test statistic is evaluated to be $r_Q = 0.9881$ which is greate than the critical value $r_{Q,\alpha} = 0.9873$ (see Table 4.1). Thus the null hypothesis is accepted and the prediction residuals appear to follow a normal distribution. Hence, in the case for $\rho = 1$, the Gaussian approximation is valid and hence it can be seen that the estimates for a Kalman filter are reliable.

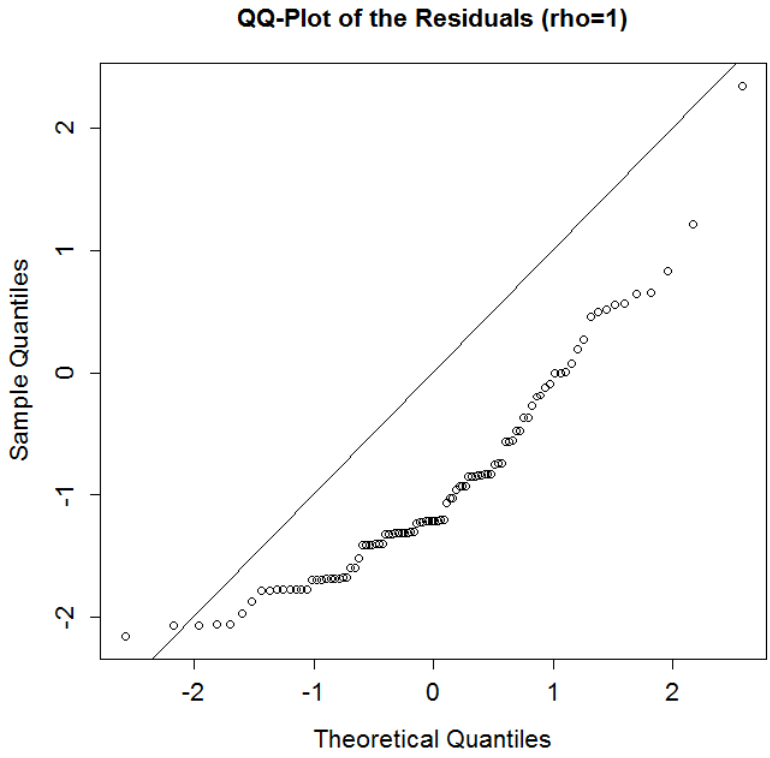

Figure 5.4: QQ-plot for the standardised prediction residuals from Kalman filter ($\rho = 1$)

5.3 Reverse Jump Markov Chain Monte Carlo method

We now apply the Markov Chain Monte Carlo method describe in Section 4.3 in Chapter 5.2.1 to the dataset. In this section, issues regarding computational efficiency and accuracy of the algorithm will be explored and methods are developed to overcome these issues.

5.3.1 Minimum number of iterations required

The first step in implementing the Markov Chain Monte Carlo algorithm is to determine the number of stochastic expectation maximisation iterations and also the number of Markov Chain Monte Carlo filtering steps for each expectation maximisation iteration. As noted previous, the parameter ρ reflects the expected number of shots in the intensity for a unit of time which implies there should be on average ρT shots in the intensity process over the time interval $[0, T]$. Since for each transition in the Markov Chain Monte Carlo algorithm, the number of shots in the filtered intensity can only go up by 1 if the chosen transition is a birth transition. In order for the filtered intensity to converge, a large enough period is required in order for the number of shots to reach a stable level at

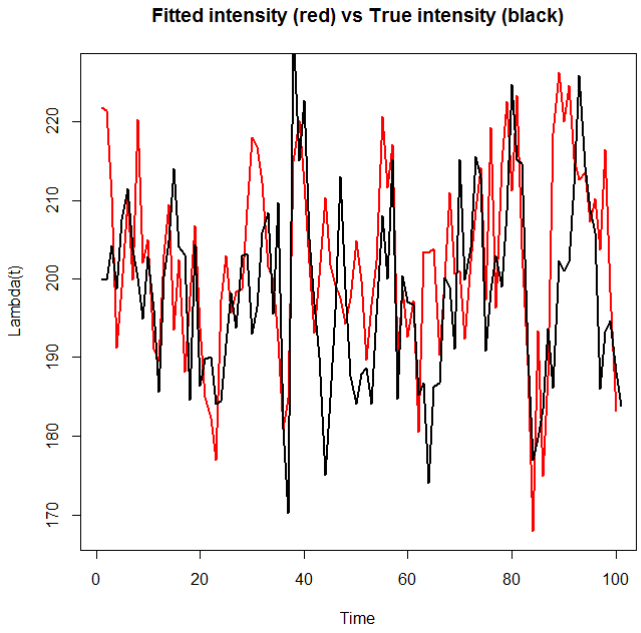

Figure 5.5: Fitted Kalman (red) vs Actual (black) intensity process

around ρT. Too many iterations of the Markov Chain Monte Carlo algorithm makes its implementation needlessly inefficient.

In order to improve the efficiency of the MCMC algorithm without sacrificing accuracy, one can estimate the number of iterations required for the Markov Chain Monte Carlo algorithm to burn in. Given the transition probability assumptions in Chapter 5.2.1, for each Markov Chain Monte Carlo step, the probability of choosing a birth transition 20% of an intensity jump occurring. Thus, for the filter to be able to fully converge, sufficient amount of iterations for the algorithm is required such that:

$$\begin{aligned} \text{Minimum number of iterations} \times P(\text{birth transition}) &= \rho T \\ \Rightarrow \text{Minimum number of iterations} &= 5\rho T \end{aligned}$$

The ρ in the above will be the initial estimate obtained from the method described in Section 4.1. Hence, for larger ρ, one expects a larger burn in period as compared to a smaller ρ. For $T = 100$ and $\rho = 1$, the minimum number of iterations required for the Markov chain to start to stabilise is around 500. In contrast, for $\rho = 100$, the minimum number of iterations required for the Markov chain to start to stabilise is around 50000. However, considering that there is also a chance of suppression of a shot through a death transition, the number of iterations actually used should be higher than the minimum

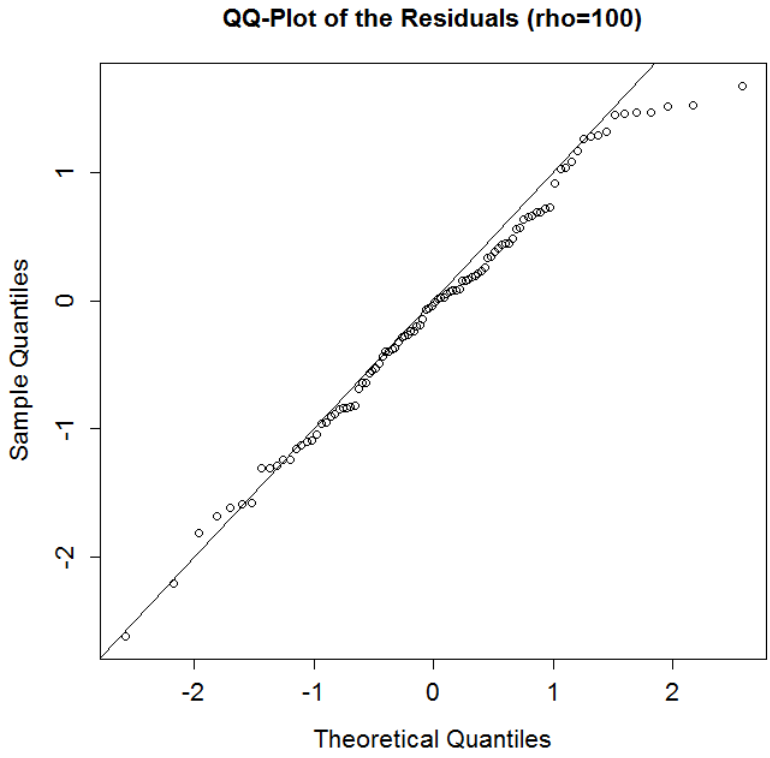

Figure 5.6: QQ-plot for the predictive residuals from Kalman filter ($\rho = 100$)

number of iterations.

5.3.2 Reparameterisation of the Likelihood

For the MCMC algorithm, 100 iterations of the stochastic expectation maximisation algorithm were ran. For each EM iteration, there were 5000 iterations of the MCMC algorithm which included a burn in phase of 1000 iterations. The corresponding estimates from the simulation study can be seen in the table:

Method	$\hat{\rho}$	$\hat{\eta}$	$\hat{k}$	Neg log likelihood	$\frac{\rho}{k\eta}$
Original MCMC estimates	1.4×10^{-14}	1540	4.77×10^{-18}	34.7	1.92

Based from the above table, the parameters estimates are shown to be very unrealistic and inaccurate. A very small *rho* with a very small k implies that the shot noise process will have almost no jumps with very slow decay. This means the filtered process behaves like a decreasing exponential function and as implies the number of claims per time unit will decrease with time. The Hessian matrix was almost singular which implied that there

was a lack of convergence in the algorithm and hence the parameter estimates are invalid. Despite the divergence of ρ and k from the true values, see that the expected number of counts per time unit implied by the parameters are given by:

$$E[N(t+1) - N(t)] = \frac{\rho}{\eta k} = 1.92$$

which is close to the value implied by the true parameters. This can also be explained by the fact ρ and k have a tendency to move in a similar pattern. As ρ increases, there would many more jumps in the intensity. In order to revert the intensity back to its mean level, k would also need to increase as well decay the jumps more quickly. Hence the ratio of ρ and k is generally preserved. This motivates the following reparametrisation of the likelihood with respect to α and β such that:

$$\alpha = \rho \qquad \beta = \frac{\rho}{k}$$

5.3.3 Low frequency case study

Based on the parametrisation, the fitted parameters with the standard errors in brackets are shown in Table 5.4. Figure 5.7 shows the fitted intensity process filtered by the MCMC algorithm against the true intensity process.

	$\hat{\rho}$	$\hat{\eta}$	$\hat{k}$	Log likelihood	$\frac{\rho}{k\eta}$
MCMC alternative parametrisation	0.9212	0.9513	0.4315	-986.72	2.24

Table 5.4: MCMC parameter estimates for low shot frequency case

The convergence of the parameters can be shown in the graph below: From the above, there is some semblance of convergence of the stochastic expectation maximisation algorithm for a large enough number of iterations. As mentioned in Centanni and Minozzo (2006), the stochastic expectation maximisation algorithm does not guarantee pointwise convergence, however, there is convergence to a small enough parameter space in which the true parameters lie.

5.3.4 High frequency case study

The log-likelihood for the estimates is -24122.31. The following table gives the estimates for the parameters.

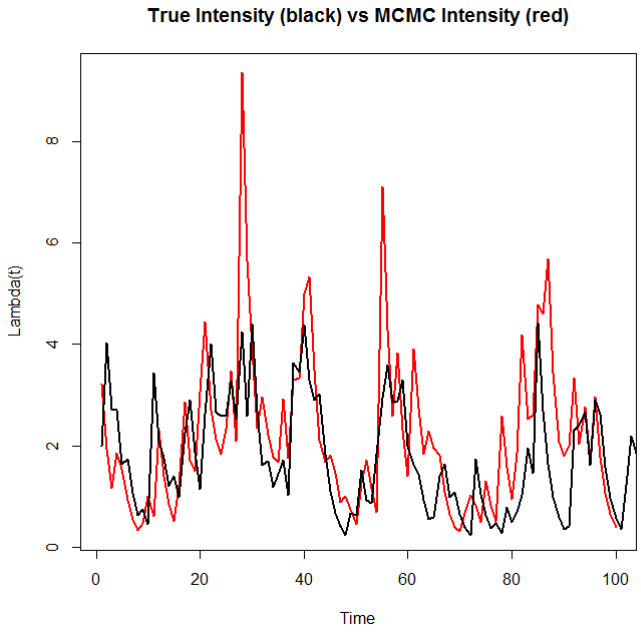

Figure 5.7: Fitted Intensity from MCMC (red) vs Actual (black) intensity process ($rho = 1$)

	ρ (s.e)	η (s.e)	k (s.e)
Actual	100	1	0.5
MCMC fit	93.15 (4.23)	0.78 (0.082)	0.53 (0.01)

Table 5.5: MCMC parameter estimates for low shot frequency case

5.4 Comparison of the two methods

As we have demonstrated that for a low frequency data, the Kalman filter estimates are not valid as the prediction residual are not normally distributed, we shall only compare the methods in the case of high frequency data. Based on Table 5.6, despite the fact

	ρ (s.e)	η (s.e)	k (s.e)	Likelihood
Actual	100	1	0.5	
Kalman fit	91.84 (5.13)	0.75 (0.071)	0.59 (0.03)	-2122.31
MCMC fit	93.15 (4.23)	0.78 (0.082)	0.53 (0.01)	-24122.31

Table 5.6: Comparison of the methods in high frequency case

that the Kalman filter method has a better log likelihood (as expected for large ρ), the MCMC parameter estimates were closer to the true parameter estimates. As mentioned in the previous chapter, both methods utilise different likelihood functions the log likelihood may not be the best measure of goodness of fit. Hence, we use the goodness of fit tests

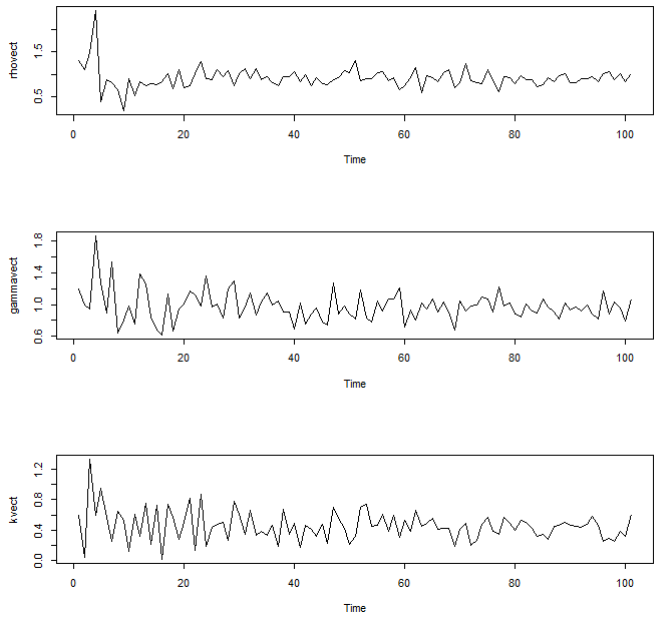

Figure 5.8: Convergence rate of parameters where $\rho = 1$, $\eta = 1$, $k = 0.5$

	High frequency KS	High frequency AD
Kalman fit	0.29	0.36
MCMC fit	0.28	0.37

Table 5.7: Goodness of fit statistics for high frequency case

described in Chapter 5.2.1 to compare the two methods in the high frequency case.

From Table 5.7, for the high frequency data indicated with a high value of ρ, both data sets yield distance statistics which are very close together. This further demonstrates that for large ρ, the Gaussian approximation of the shot noise process becomes more valid.

In this chapter, we have shown for low frequency data which exhibits a small ρ, We have demonstrated that if ρ is expected to be large, then since the results from the Kalman filter does not differ from the Markov Chain Monte Carlo method by much, the Kalman filter would be preferable due to its quick implementation. There are other practical issues inherent in insurance claims data which need to be considered when fitting the shot noise Cox process.

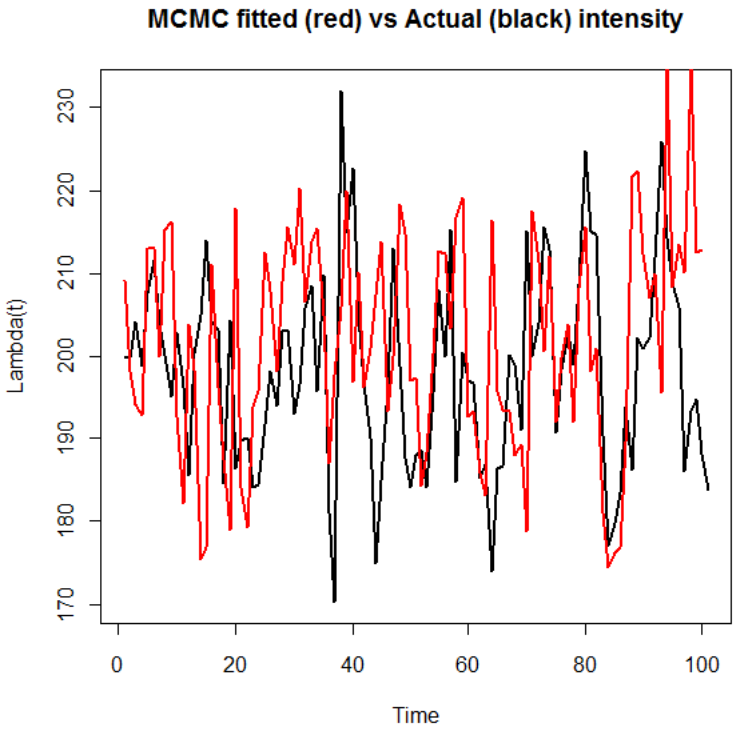

Figure 5.9: Fitted MCMC (red) vs Actual (black) intensity process ($\rho = 100$

CHAPTER 6

CONCLUSION

6.1 Summary of main contributions

In this thesis, we explore the use of the shot noise Cox process as a robust claims model which allows for stochasticity in the claim arrival rate. In particular, we have investigated and applied methods to fit these processes to real insurance claims data. The majority of the current literature focuses on using doubly stochastic Poisson processes in a non life insurance context on pricing and ruin probability calculations in a purely theoretical setting. Although doubly stochastic Poisson processes have been fitted in other areas such as finance, longevity and physical sciences, the methods used in these fields cannot be directly applied to non-life insurance due to the different nature of the problem. Hence a new framework for fitting shot noise Cox processes to non-life insurance data is required.

Firstly, we have explored and derived some statistical features of the shot noise Cox process. In particular, the moment generating function and the correlation structure of the increments of shot noise Cox process are derived which are required in the fitting procedure of real life insurance data.

We have provided details of two methods for fitting shot noise Cox processes through extending popular methods in other relevant fields for filtering latent variables. Based on a Gaussian approximation of the shot noise intensity and Cox process, we have developed

a framework to use Kalman filter as a quick and efficient method to estimate parameters for the shot noise Cox process. We have also comment on the validity of the approximation and proposed a method to check the validity of these parameter estimates by testing the normality of the prediction residuals. The other method we use is the reverse jump Markov Chain Monte Carlo algorithm with stochastic expectation maximisation. Through deriving the conditional log-likelihood, its gradient vector and Hessian matrix, we are able to derive the conditional maximum likelihood estimate for η based on the filtered intensity and other two parameters. This reduces the dimension of the optimisation problem and hence has improved the efficiency of the stochastic expectation maximisation procedure.

Through a simulation study, we have calibrated the model fitting procedure to improve the efficiency and accuracy of the methods. Reliable initial estimates are obtained through a combination of method of moments and approximating the distribution of the shot noise Cox process. In particular for the Markov Chain Monte Carlo method, we deal with convergence issues in the algorithm through reparameterisation of the problem as well as discuss the number of iterations to balance computational efficiency and accuracy of the parameter estimates. Finally we further explore practical considerations for fitting shot noise Cox processes through a comprehensive demonstration on real general insurance data.

6.2 Areas for further research

Through developing methods to fit shot noise Cox processes, this research has laid the foundation for various extensions. This thesis consider fitting the doubly stochastic Poisson process with a shot noise intensity where the stochasticity in the claim intensity only arrives as positive jumps. An immediate extension on this work would be to explore methods to fit doubly stochastic Poisson processes with other stochastic processes for the intensity such as affine diffusion processes. This would increase the flexibility of the doubly stochastic Poisson process for the modeller and from an insurer's perspective, they would be able to select the intensity process which reflects the nature of the claim arrivals for each line of business most accurately. This also means that methods to compare the fits of doubly stochastic Poisson processes with different intensities would need to be expanded as well as both qualitative and quantitative selection criteria.

As mentioned in the introduction, from an insurer's perspective, a more accurate claim count model would allow for more accurate pricing, reserving and economic capital calcu-

lations. Although there has already been theoretical developments on using Cox processes in pricing and probability of ruin, stochastic reserving using doubly stochastic Poisson process is a potential area for future research. Doubly stochastic Poisson process can also be used to model the claim settlement process for the insurer. For each policy, the claim settlement rate is generally decaying with time as more claims are settled until another event causing the need for more settlement payments occur. This means that shot noise Cox processes are natural models for the claim settlement process as the shot noise intensity reflects the behaviour of rate at which claims are settle for a policy.

BIBLIOGRAPHY

Albrecher, H., Asmussen, S., 2006. Ruin probabilities and aggregrate claims distributions for shot noise cox processes. Scandinavian Actuarial Journal 2006 (2), 86–110.

Anderson, T., Darling, D., 1952. Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes. The Annals of Mathematical Statistics, 193–212.

Barndorff-Nielsen, O., Shephard, N., 1998. Aggregation and model construction for volatility models.

Barndorff-Nielsen, O., Shephard, N., 2001. Non-gaussian ornstein–uhlenbeck-based models and some of their uses in financial economics. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63 (2), 167–241.

Bartlett, M., 1986. Mixed cox processes, with an application to accident statistics. Journal of Applied Probability, 321–333.

Basu, S., Dassios, A., 2002. A cox process with log-normal intensity. Insurance: mathematics and economics 31 (2), 297–302.

Biffis, E., 2005. Affine processes for dynamic mortality and actuarial valuations. Insurance: Mathematics and Economics 37 (3), 443–468.

Björk, T., Grandell, J., 1988. Exponential inequalities for ruin probabilities in the cox case. Scandinavian Actuarial Journal 1988 (1-3), 77–111.

Bluhm, C., Overbeck, L., Wagner, C., 2003. An introduction to credit risk modeling. CRC Press.

Bouzas, P., Aguilera, A., Ruiz-Fuentes, N., 2009. Functional estimation of the random rate of a cox process. Methodology and Computing in Applied Probability, 1–13.

Bouzas, P., Ruiz-Fuentes, N., Matilla, A., Aguilera, A., Valderrama, M., 2010. A cox model for radioactive counting measure: Inference on the intensity process. Chemometrics and Intelligent Laboratory Systems 103 (2), 116–121.

Centanni, S., Minozzo, M., 2006. A monte carlo approach to filtering for a class of marked doubly stochastic poisson processes. Journal of the American Statistical Association 101 (476), 1582–1597.

Chib, S., Nardari, F., Shephard, N., 2002. Markov chain monte carlo methods for stochastic volatility models. Journal of Econometrics 108 (2), 281–316.

Čížek, P., Härdle, W., Weron, R., 2011. Statistical tools for finance and insurance. Springer Verlag.

Cox, D., 1955. Some statistical methods connected with series of events. Journal of the Royal Statistical Society. Series B (Methodological), 129–164.

Cox, D., Isham, V., 1980. Point processes. Vol. 12. Chapman & Hall/CRC.

Cox, J., Ingersoll Jr, J., Ross, S., 1985. A theory of the term structure of interest rates. Econometrica: Journal of the Econometric Society, 385–407.

Dachian, S., Kutoyants, Y., 2008. On the goodness-of-fit tests for some continuous time processes. Statistical models and methods for biomedical and technical systems, 385–403.

Dassios, A., Jang, J., 2003. Pricing of catastrophe reinsurance and derivatives using the cox process with shot noise intensity. Finance and Stochastics 7 (1), 73–95.

Dassios, A., Jang, J., 2005. Kalman-bucy filtering for linear systems driven by the cox process with shot noise intensity and its application to the pricing of reinsurance contracts. Journal of applied probability 42 (1), 93–107.

Denuit, M., Maréchal, X., Pitrebois, S., Walhin, J., 2007. Actuarial modelling of claim counts. Wiley Online Library.

Diggle, P., Besag, J., Gleaves, J., 1976. Statistical analysis of spatial point patterns by means of distance methods. Biometrics, 659–667.

Diggle, P., Chetwynd, A., 1991. Second-order analysis of spatial clustering for inhomogeneous populations. Biometrics, 1155–1163.

Diggle, P., Diggle, P., 1983. Statistical analysis of spatial point patterns.

Frey, R., Runggaldier, W., 2001. A nonlinear filtering approach to volatility estimation with a view towards high frequency data. International Journal of Theoretical and Applied Finance 4 (2), 199–210.

Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. Pattern Analysis and Machine Intelligence, IEEE Transactions on (6), 721–741.

Grandell, J., 1976. Doubly stochastic Poisson processes. Springer-Verlag Berlin, Heidelberg, New York.

Hastings, W., 1970. Monte carlo sampling methods using markov chains and their applications. Biometrika 57 (1), 97–109.

Hawkes, A., 1971. Spectra of some self-exciting and mutually exciting point processes. Biometrika 58 (1), 83–90.

Hougaard, P., Lee, M., Whitmore, G., 1997. Analysis of overdispersed count data by mixtures of poisson variables and poisson processes. Biometrics, 1225–1238.

Kalman, R., 1960. A new approach to linear filtering and prediction problems. Journal of basic Engineering 82 (1), 35–45.

Klugman, S., Panjer, H., Willmot, G., Venter, G., 1998. Loss models: from data to decisions. Vol. 2. Wiley New York.

Klüppelberg, C., Mikosch, T., 1995. Explosive poisson shot noise processes with applications to risk reserves. Bernoulli, 125–147.

Kolmogorov, A., 1933. Sulla determinazione empirica di una legge di distribuzione. Giornale dell Istituto Italiano degli Attuari 4 (1), 83–91.

Konecny, F., 1986. Maximum likelihood estimation for doubly stochastic poisson processes with partial observations. Stochastics: An International Journal of Probability and Stochastic Processes 16 (1-2), 51–63.

Lando, D., 1998. On cox processes and credit risky securities. Review of Derivatives research 2 (2), 99–120.

Luciano, E., Vigna, E., 2005. Non mean reverting affine processes for stochastic mortality. ICER Applied Mathematics Working Paper No. 4-2005.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E., 1953. Equation of state calculations by fast computing machines. The journal of chemical physics 21, 1087.

Mikosch, T., 2009. Non-Life Insurance Mathematics: An Introduction with the Poisson Process. Springer Verlag.

Møller, J., 2003. Shot noise cox processes. Advances in Applied Probability 35 (3), 614–640.

Neuts, M., 1971. A queue subject to extraneous phase changes. Advances in Applied Probability, 78–119.

Rodriguez-Iturbe, I., Cox, D., Isham, V., 1987. Some models for rainfall based on stochastic point processes. Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences 410 (1839), 269–288.

Rudemo, M., 1972. Doubly stochastic poisson processes and process control. Advances in Applied Probability, 318–338.

Rydberg, T., Shephard, N., 1999. A modelling framework for the prices and times of trades made on the new york stock exchange.

Schrager, D., 2006. Affine stochastic mortality. Insurance: Mathematics and Economics 38 (1), 81–97.

Seal, H., 1983. The poisson process: its failure in risk theory. Insurance: Mathematics and Economics 2 (4), 287–288.

Stoyan, D., 1992. Statistical estimation of model parameters of planar neyman-scott cluster processes. Metrika 39 (1), 67–74.

R CODE

Kalman Filter

Kalman filter code:

```
ShotKalman <- function(parameter) {
#Initialise parameters
rho = parameter[1]
eta = parameter[2]
k = parameter[3]

#Transform observed data (Dassios Jang 2005)
X = (data-rho/(k*eta))/sqrt(rho/(k*eta^2))

Z=rep(0, length(data))
Z[1]=(mean(data)-rho/k/eta)/sqrt(rho/(k*eta^2))
P=1
Pvect=P
Vvect=X[1]-Z[1]

#Filter part
for (t in 1:(length(data)-1)) {

#Time Update
Z[t+1] = Z[t]*(1-k)
P = (1-k)^2*P +2*k
Pvect<-c(Pvect,P+eta)
#Measurement Update
K = P/(P+eta)
Vvect=c(Vvect, X[t+1]-Z[t+1])
```

```
Z[t+1] = Z[t+1] + K*(X[t+1]-Z[t+1])
P = (1-K)*P
}

#Retransform data
Lambda = Z*sqrt(rho/(eta^2*k)) + rho/(eta*k)


#EM Part
nlloglik <- length(data)*0.5*log(2*pi)+0.5*sum(log(Pvect))+
0.5*sum(log(Vvect^2/Pvect))
return(nlloglik)
#return(Lambda)
#return(rbind(Vvect, Pvect))
}
```

MCMC

```
for (ites in 1:100) {

#Initialise states
jumpT<-c() #Jump times
jumpS<-c() #Jump size
n<-0 #Number of jumps

#Loop starts where we observe 1 year

for(count in 1:5000) {

#Choose transition
randU<-runif(1)
#In case where there are no jumps:
if(n==0) {
lambda<-function(t) { #Use log likelihood
log(lambda0)-k*t
}
intlambda<-function(t) {
1/k*lambda0*(1-exp(-k*t))
}
if(randU<0.5) { #Choose birth
#Generate new state
newTime=runif(1,0,length(data))
newSize = rexp(1, gamma)
jumpTnew=c(jumpT, newTime)
jumpSnew = c(jumpS, newSize)
```

```
#Create function for changed
lambdanew<-function(t) {
lambda0*exp(-k*t)+sum((t>jumpTnew)*jumpSnew*exp(-k*(t>jumpTnew)
*(t-jumpTnew)))
}
intlambdanew<-function(t) {
1/k*(lambda0*(1-exp(-k*t))+sum((t>jumpTnew)*jumpSnew*(1-exp(-k*(t>jumpTnew)
*(t-jumpTnew)))))
}

evallambda<-c()
for(count1  in 1:length(data)) {
evallambda[count1]<-lambda(count1)
}
evallambdanew<-c()
for(count1  in 1:length(data)) {
evallambdanew[count1]<-log(lambdanew(count1))
}
likelihood_rat=sum(data*evallambdanew)-intlambdanew(length(data))-
sum(data*evallambda)+intlambda(length(data))
prior_rat = log(dexp(newSize, gamma))+log(rho)
proposal_rat = log(length(data)/(n+1))-log(gamma)+gamma*newSize
accept_rat = likelihood_rat+ prior_rat+proposal_rat

#Acceptance of new state
accept_rand= log(runif(1))
if(accept_rand<accept_rat) {
jumpT = jumpTnew
jumpS = jumpSnew
n=n+1
}
}

else {#Choose shift
#Generate new state
lambda00 = rgamma(1, rho/k, gamma)

#Create function for changed
lambdanew<-function(t) { #Use loglikelihood in case k gets large
log(lambda00)-k*t
}
intlambdanew<-function(t) {
1/k*lambda00*(1-exp(-k*t))
}

#Compute likelihood ratios
```

```
evallambda<-c()
for(count1  in 1:length(data)) {
evallambda[count1]<-lambda(count1)
}
evallambdanew<-c()
for(count1  in 1:length(data)) {
evallambdanew[count1]<-lambdanew(count1)
}
likelihood_rat=sum(data*evallambdanew)-intlambdanew(length(data))-
sum(data*evallambda)+intlambda(length(data))
accept_rat= likelihood_rat #Prior and proposal multiply to 1

#Acceptance of new state
accept_rand= log(runif(1))
if(accept_rand<accept_rat) {
lambda0=lambda00
}
}
}
else { #In case where there is a jump

#Create function for lambda
lambda<-function(t) {
lambda0*exp(-k*t)+sum((t>jumpT)*jumpS*exp(-k*(t>jumpT)*(t-jumpT)))
}
intlambda<-function(t) {
1/k*(lambda0*(1-exp(-k*t))+sum((t>jumpT)*jumpS*(1-exp(-k*(t>jumpT)
*(t-jumpT)))))
}

#Choose transition type
if(randU<0.2) { #Choose start
#Generate new state
lambda00 = rgamma(1, rho/k, gamma)

#Create function for changed
lambdanew<-function(t) {
lambda00*exp(-k*t)+sum((t>jumpT)*jumpS*exp(-k*(t>jumpT)*(t-jumpT)))
}
intlambdanew<-function(t) {
1/k*(lambda00*(1-exp(-k*t))+sum((t>jumpT)*jumpS*(1-exp(-k*(t>jumpT)
*(t-jumpT)))))
}

#Compute likelihood ratios
evallambda<-c()
for(count1  in 1:length(data)) {
```

```
evallambda[count1]<-log(lambda(count1))
}
evallambdanew<-c()
for(count1  in 1:length(data)) {
evallambdanew[count1]<-log(lambdanew(count1))
}
likelihood_rat=sum(data*evallambdanew)-intlambdanew(length(data))
- sum(data*evallambda)+intlambda(length(data))
accept_rat= likelihood_rat

#Acceptance of new state
accept_rand= log(runif(1))
if(accept_rand<accept_rat) {
lambda0=lambda00
}
}

else if (randU<0.4) { #Choose position
#Generate new state
A=sample(1:n, 1)
min=0
max=length(data)
if(A>1) {
min=jumpT[A-1]
}
if(A<n) {
max=jumpT[A+1]
}
Timeshift=runif(1, min, max)
jumpTnew = jumpT
jumpTnew[A]=Timeshift

#Create function for changed
lambdanew<-function(t) {
lambda0*exp(-k*t)+sum((t>jumpTnew)*jumpS*exp(-k*(t>jumpTnew)
*(t-jumpTnew)))
}
intlambdanew<-function(t) {
1/k*(lambda0*(1-exp(-k*t))+sum((t>jumpTnew)*jumpS*(1-
exp(-k*(t>jumpTnew)*(t-jumpTnew)))))
}

#Compute likelihood ratios
evallambda<-c()
for(count1  in 1:length(data)) {
evallambda[count1]<-log(lambda(count1))
}
```

```
evallambdanew<-c()
for(count1  in 1:length(data)) {
evallambdanew[count1]<-log(lambdanew(count1))
}
likelihood_rat=sum(data*evallambdanew)-intlambdanew(length(data))
- sum(data*evallambda)+intlambda(length(data))
accept_rat= likelihood_rat
#Acceptance of new state
accept_rand= log(runif(1))
if(accept_rand<accept_rat) {
jumpT=jumpTnew
}
}

else if (randU<0.6) { #Choose height
#Generate new state
A=sample(1:n, 1)
Positionshift=rexp(1, gamma)
jumpSnew = jumpS
jumpSnew[A]=Positionshift

#Create function for changed
lambdanew<-function(t) {
lambda0*exp(-k*t)+sum((t>jumpT)*jumpSnew*exp(-k*(t>jumpT)*(t-jumpT)))
}
intlambdanew<-function(t) {
1/k*(lambda0*(1-exp(-k*t))+sum((t>jumpT)*jumpSnew*(1-exp(-k*(t>jumpT)
*(t-jumpT)))))
}

#Compute likelihood ratios
evallambda<-c()
for(count1 in 1:length(data)) {
evallambda[count1]<-log(lambda(count1))
}
evallambdanew<-c()
for(count1  in 1:length(data)) {
evallambdanew[count1]<-log(lambdanew(count1))
}
likelihood_rat=sum(data*evallambdanew)-intlambdanew(length(data))-
sum(data*evallambda)+intlambda(length(data))
accept_rat= likelihood_rat
#Acceptance of new state
accept_rand= log(runif(1))
if(accept_rand<accept_rat) {
jumpS=jumpSnew
}
```

```
}

else if (randU<0.8) { #Choose birth
#Generate new state
newTime=runif(1,0,length(data))
newSize = rexp(1, gamma)
jumpTnew=c(jumpT, newTime)
jumpSnew = c(jumpS, newSize)
#Sorting the jumps in order
jumpdata<-data.frame(jumpTnew,jumpSnew)
datasort<-jumpdata[with(jumpdata, order(jumpTnew)),]
jumpTnew= datasort$jumpTnew
jumpSnew= datasort$jumpSnew

#Create function for changed
lambdanew<-function(t) {
lambda0*exp(-k*t)+sum((t>jumpTnew)*jumpSnew*exp(-k*(t>jumpTnew)*(t-jumpTnew)))
}
intlambdanew<-function(t) {
1/k*(lambda0*(1-exp(-k*t))+sum((t>jumpTnew)*jumpSnew*(1-exp(-k*(t>jumpTnew)
*(t-jumpTnew)))))
}

evallambda<-c()
for(count1  in 1:length(data)) {
evallambda[count1]<-log(lambda(count1))
}
evallambdanew<-c()
for(count1  in 1:length(data)) {
evallambdanew[count1]<-log(lambdanew(count1))
}
likelihood_rat=sum(data*evallambdanew)-intlambdanew(length(data))-
 sum(data*evallambda)+intlambda(length(data))
prior_rat = log(rho)
proposal_rat = log(length(data)/(n+1))
accept_rat = likelihood_rat+ prior_rat+proposal_rat

#Acceptance of new state
accept_rand= log(runif(1))
if(accept_rand<accept_rat) {
jumpT = jumpTnew
jumpS = jumpSnew
n=n+1
}
}

else { #Choose death
```

```
#Generate new state
A=sample(1:n, 1)
jumpTnew=jumpT[-A]
jumpSnew = jumpS[-A]

#Create function for changed
lambdanew<-function(t) {
if(n==1) { #give log lambda
lambdanew<-log(lambda0)-k*t
}
else {
lambdanew<-lambda0*exp(-k*t)+(n>1)*sum((t>jumpTnew)*jumpSnew*exp(-k*
(t>jumpTnew)*(t-jumpTnew)))
}
return(lambdanew)
}
intlambdanew<-function(t) {
1/k*(lambda0*(1-exp(-k*t))+(n>1)*sum((t>jumpTnew)*jumpSnew*(1-exp(-k
*(t>jumpTnew)*(t-jumpTnew)))))
}

evallambda<-c()
for(count1  in 1:length(data)) {
evallambda[count1]<-log(lambda(count1))
}
evallambdanew<-c()
for(count1  in 1:length(data)) {
if(n==1) {
evallambdanew[count1]<-lambdanew(count1)
}
else {
evallambdanew[count1]<-log(lambdanew(count1))
}
}
likelihood_rat=sum(data*evallambdanew)-intlambdanew(length(data))-
sum(data*evallambda)+intlambda(length(data))
prior_rat = -log(rho)
proposal_rat = log(n/length(data))
accept_rat = likelihood_rat+ prior_rat+proposal_rat

#Acceptance of new state
accept_rand= log(runif(1))
if(accept_rand < accept_rat) {
jumpT = jumpTnew
jumpS = jumpSnew
n=n-1
}
```

```
}
}
}

#Optimising with respect to the parameters
llhoodfn<- function(x){
 alphaest = x[1] #rho
 betaest = x[2] #rho/k
 gammaest = (alphaest+n)/(lambda0 + sum(jumpS))
 evallambda<-c()
if(n==0) {
 lambda<-function(t) {
log(lambda0)-(alphaest/betaest)*t
}
 intlambda<-function(t) {
1/(alphaest/betaest)*lambda0*(1-exp(-(alphaest/betaest)*t))
 }
 jumpSlike = 0
for(count1  in 1:length(data)) {
evallambda[count1]<-lambda(count1)
}
}
else {
 lambda<-function(t) {
lambda0*exp(-(alphaest/betaest)*t)+sum((t>jumpT)*jumpS*exp(
-(alphaest/betaest)*(t>jumpT)*(t-jumpT)))
}
 intlambda<-function(t) {
1/(alphaest/betaest)*(lambda0*(1-exp(-(alphaest/betaest)*t))
  +sum((t>jumpT)*jumpS*(1-exp(-(alphaest/betaest)*(t>jumpT)*(t-jumpT)))))
 }
 jumpSlike = log(dexp(jumpS, gammaest))
for(count1  in 1:length(data)) {
evallambda[count1]<-log(lambda(count1))
}
}
 llhood = -(log(dgamma(lambda0,alphaest/(alphaest/betaest), gammaest))
  +sum(jumpSlike)+n*log(alphaest)-alphaest*length(data) - intlambda(
  length(data)) + sum(data*evallambda))
 return(llhood)
}

llgrad<- function(x) {
rhoest<-x[1]
betaest<-x[2]
k<-rhoest/betaest
gamma<-(betaest+n)/(lambda0 + sum(jumpS))
```

```
llgrad<-c()
t=length(data)
T=1:length(data)
llgrad[1] <- log(gamma)/k + lambda0 - 1/k*digamma(betaest) + n/rhoest - t
A<-c()
B<-c()
for (count in 1:length(data)) {
A[count]<-sum(jumpS*exp(k*jumpT)*(count>= jumpT))
B[count]<-sum(jumpS*jumpT*exp(k*jumpT)*(count>= jumpT))
}
llgrad[2] <- -k/beta*(sum(data*(B/(lambda0+A) - T)) + lambda0*(1-
exp(-k*t)-k*t*exp(-k*t))/k^2 + lambda0/k^2*(n - sum(exp(-k*(t-jumpT))-k*
(t-jumpT)*exp(-k*(t-jumpT))))-betaest*(log(gamma)+lambda0)/k +
betaest/k*digamma(betaest))
return(llgrad)
}

shotoptim<-constrOptim(c(alpha, beta), llhoodfn, llgrad, ui=rbind(c(1,0),
c(0,1)), ci=c(0,0))
alpha=shotoptim$par[1]
beta=shotoptim$par[2]
gamma = (rho/k+n)/(lambda0 + sum(jumpS))
rho=alpha
k = alpha/beta
rhovect<-c(rhovect, rho)
gammavect<-c(gammavect, gamma)
kvect <- c(kvect, k)
nvect <-c(nvect, n)
lvect <- c(lvect, shotoptim$value)
}
```