UNSW AUSTRALIA | AGSM Business School

Individual claim liability analysis using Markov-modulated Poisson processes

School of Risk and Actuarial Studies
UNSW Business School
University of New South Wales

Alan Liang Xian

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

November 2, 2015

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

CONFIDENTIALITY

This thesis is submitted for the purposes of internal assessment by the School of Risk and Actuarial Studies, UNSW. The following document contains confidential information collected as part of the Linkage Project:"Modelling claim dependencies for the general insurance industry with economic capital in view: an innovative approach with stochastic processes", which is supported by the Australian Research Council. External circulation of this thesis without the express permission of the partners in the Linkage Project is strictly forbidden.

DECLARATION

I hereby declare that this submission of thesis 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 concept or in style, presentation and linguistic expression is acknowledged.

Signed:

Date:

ABSTRACT

The accurate estimation of outstanding claims is of key importance to insurers and is one of the primary roles of an actuary. Despite this, claims reserving methods that are commonly used in practice underutilise the vast amount of data that insurers collect from policyholders. Micro-level models have recently gained popularity as they allow claim analysis to be conducted at a more granular level and the higher computational requirements are considerably less of a problem given modern technological advances. Ultimately, this leads to a more accurate estimation of claims reserves which may reduce the additional risk margin insurers need to hold.

A further consideration when calculating reserves is that certain economic and environmental events can lead to claim intensity fluctuations, resulting in several distinct levels of varying claim arrival rates. The use of Markov-modulated Poisson processes to capture this type of stochasticity is well documented in the physical science and information technology literature. However, there is little literature in an insurance context and various issues regarding implementation are yet to be resolved before these models can be applied in practice.

This thesis applies a Markov-modulated Poisson process within a micro-level framework to model claim counts. Adjustments for insurance-related factors such as reporting delay, exposure, seasonality and other residual trends are discussed. Analysis is conducted to test the accuracy of the expectation maximisation algorithm that is used to calibrate the model and a real-world case study is then undertaken to demonstrate the theoretical findings. Careful analysis suggests that the MMPP model outperforms the commonly used chain-ladder algorithm in terms of prediction accuracy, highlighting the advantages of the proposed model.

ACKNOWLEDGEMENT

Firstly, I want to express my immeasurable appreciation to my supervisors Dr. Benjamin Avanzi, Dr. Greg Taylor and Dr. Bernard Wong for their support this year. Thank you for your invaluable advice, encouragement and patience. Your willingness to help gave me the confidence to tackle what was originally a hugely daunting Honours project and I am extremely grateful for your efforts in making this year such an enlightening and satisfying experience.

I would also like to thank the Australian Prudential Regulation Authority and the UNSW Business School for their financial support through the Brian Gray Scholarship and the UNSW Honours Year Scholarship. These scholarships have been a huge help as they have allowed me to concentrate on my thesis.

My sincerest thanks also goes to the staff of the School of Risk and Actuarial studies and the ARC Centre of Excellence in Population Ageing Research for their advice and support throughout this year. Further, I am very grateful to my fellow Honours students Andrew and Tiffany as well as the PhD students for all the good times we have had this year. In particular, thank you Xinda for letting me constantly barrage you with questions.

To my friends and family, thank you for driving me to work my hardest this year. Thank you, Michelle, for being such a great manager. And finally, I am deeply indebted to you, Calise, for your unwavering support. Thanks for patiently putting up with me despite all the rough times this year. We've done it!

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Background

The primary business of insurance is the coverage of contingent liabilities for a premium. In the event that these liabilities are ever realised, insurers must ensure that they have enough liquid funds available to adequately fulfil their obligations. However, the evaluation of exactly how much money should be held for this purpose is difficult due to the stochastic nature of the claims process. Consider the generalised development of a claim, shown below in Figure 1.1.

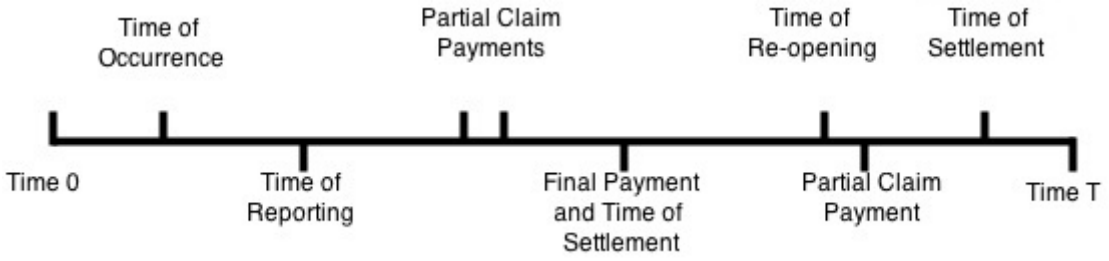

Figure 1.1: The Claims Process

At time 0, the policy begins and the policyholder is covered for certain types of losses until time T. If a claims event does occur at some point during this period, then the cost of the claim needs to be paid out by the insurer. We call the time that such an event occurs the "Time of Occurrence" and the time that it is reported to the insurer the "Time of Reporting". These two time points are not the same and can be quite far apart depending on the type of risk covered. After the claim is reported to the insurer and

some further processing/investigation time, payments start being made to the claimant. It is important to note that claims are generally not fully paid out in one lump sum but develop over a period of time. For example, in the case of a worker's compensation claim, rehabilitation can take a significant amount of time. This results in many partial claim payments that can vary in severity many years into the future, and the prediction of this claims process is thus very complicated.

Eventually, there is a final payment that is made and at that point, we consider the claim closed and the claims process is settled. However, in some cases, claims can be reopened due to the emergence of new information or an unforeseen development. If this occurs, then there is an associated "Time of Reopening", after which more partial claim payments are made. Finally, we have another settlement event and the case is closed again. We note that the time T does not have to fall at the end of this process and is only used to determine the claim eligibility in terms of the time of occurrence. If the time of occurrence falls between times 0 and T, the claim needs to be fully paid out regardless of how long the claim development process takes.

Each of the components in the described claims process can also differ significantly across different insurance lines of business. For example, private motor insurance generally has a short reporting delay (which is the time between the occurrence and reporting of the claim) while insurance claims that cover risks like asbestos exposure may not be reported for many years after the occurrence date. This creates complexities that have to be captured by models that are used for reserving these future liabilities. In addition, the reporting delay of claims introduces another complication. At the date that the reserve is determined (known as the valuation date), there may be claims that have been incurred but are yet to be reported to the insurer. These IBNR claims must be included in the valuation calculations on top of claims that have been reported but not settled (RBNS claims) in order to determine the total outstanding liability of the business.

Thus, accurate reserving calculations can be a very complex procedure. A natural question is "Why is accurate claims reserving important?". In terms of under-estimation of reserves, there are clear consequences when an insurer is not able to meet its obligations. Large losses over a short period that are not reserved could lead to insolvency. Even in the case where the business does not fail, there are costs associated with financial distress such as restricted access to capital sources and closer regulation by governing bodies. Reserves are also used as an indicator of whether products are priced correctly. Under-estimation could reflect inadequate pricing of a business's insurance products as claims are higher than expected. On the other hand, over-estimation of reserves raises its own set of issues. Following a similar train of logic to the previous statement, over-estimation could reflect over-pricing of an insurer's products, leading to less market competitiveness.

Also, reserves are the largest liability on a general insurer's balance sheet. For example, Suncorp's outstanding claims liabilities for financial year 2013 (incorporating a risk margin that satisfies a 90% probability of adequacy) was \$9,514m AUD, which is fairly large relative to their total general insurance liabilities of \$16,798m AUD (Suncorp Group, 2013). Thus, the over-estimation of claim liabilities (and hence the required reserves) can detriment the company's perceived financial position.

In practice, the prediction of future cash flows in general insurance largely utilise aggregated run-off triangles. These methods were originally adopted due to their computational simplicity and tractability, which also makes these methods very accessible to practitioners in the industry. The main idea behind these procedures is that by assuming that the average claim will follow a certain distribution over time, claims data can be analysed and projected in order to calculate future liability payments. However, difficulties arise in choosing an appropriate distribution and fitting correct parameters as well as choosing what types of data to use (for example, paid or incurred losses). The methods also do not provide a natural method of incorporating all of the extra covariate information that is collected by insurers. Further, there are other issues that can occur when the assumptions that are made by these procedures are violated (for example, Verdonck et al. (2009) highlights problems with the influence of outliers in triangular methods). Finally, in some cases, the modelling procedure can break down completely due to some of the characteristics of the data (see Kunkler, 2004). It is important to note that while fixes to these issues have been proposed, they are generally heuristic in nature and cannot be applied simultaneously. Hence, while these methods can be robust when the underlying assumptions are satisfied, they may not be adequate for situations when this is not the case. This is particularly relevant for general insurance, where policies can cover many different risks with differing properties.

A very imporant consideration that is also not taken into account by aggregate modelling techniques is sudden changes in the frequency of claims. From a realistic perspective, there are many climatic, political or economic variables that could cause these shifts, such as dry weather increasing fire risk. To complicate matters, these environmental variables may not be observable to analysts, either due to not being included in the data sets or due to their own inherent complexities. For example, it is well documented that La Niña episodes will greatly increase flood risk in Australia (see Australian Bureau of Meteorology, 2012). These hidden variables are being currently being investigated by various insurers and reinsurers, as well as by Lloyd's in the UK (see Lloyd's, 2010).

1.2 Research Motivations

The key motivation for the use of micro-level analysis is that it provides natural methods of incorporating potentially material micro-level information that would generally be missed or disregarded by macro-level models. Insurers already collect this information from their policyholders and several papers question the use of aggregate modelling when it is computationally feasible to make use of the extensive micro-level information (see England and Verrall, 2002). It is generally expected that these models will produce more accurate claims liability estimates due to the closer fit achieved by using the extra information available. Other advantages is that it relaxes some of the more unrealistic assumptions that are used in aggregate modelling, such as stochastic independence in the upper triangle of incremental claim amounts. The greatly increased sample space also accommodates modelling procedures and helps to avoid the issues encountered with aggregate techniques.

The majority of the literature on micro-level analysis is relatively old but has recently been reconsidered due to advancements in technology increasing the accessibility of these methods. However, the size of the literature is still quite small and generally theoretical in nature with few empirical case studies, particularly from an Australian non-life insurance perspective. Further, the focus is on the Poisson process whereas the generalisation to doubly stochastic Poisson processes, or Cox processes, provides a beneficial extension as it allows for stochasticity in the claims arrival intensity. One of the few papers that discuss Cox processes in this context is Avanzi et al. (2015) which employs a shot-noise process for this purpose.

Our primary extension to the micro-level analysis literature in this thesis will be through the use of Markov-modulated Poisson processes (MMPPs), which is a class of Cox processes that provide a realistic interpretation by allowing for sudden changes in the frequency of events. Extensive documentation demonstrates that MMPPs are suitable for processes that experience sudden bursts in activity such as web traffic flow (see Scott and Smyth, 2003). This is a property that is also observed in claim counting processes but the majority of the MMPP theory in financial literature has been focused in risk theory. My thesis will examine the use of this model in an insurance context to show that it will produce more accurate results when the risk frequency of claims is not homogeneous.

However, as is the case with micro-level analysis, the current literature is quite theoretical in this area with few papers that use empirical analysis to validate theoretical findings. Further, issues that will naturally arise in the practical application of such models are not discussed in the current literature (for example, the heterogeneity introduced by a non-constant exposure to risks). This is a significant barrier to the adoption of such

models by industry practitioners as it limits the accessibility of the method. My thesis will bridge this gap by presenting methods to deal with these issues. We also validate the theoretical findings by undertaking a comprehensive empirical case study using real world data. By doing so, we hope to justify the proposed advantages of MMPPs in providing realism to claims reserving models as well as provide an accessible procedure for industry actuaries that can be implemented in using standard software such as MATLAB. We note that this approach is not only applicable to a micro-level reserving framework but for any methodology that models claim counts under the assumption of independence between frequency and severity (and even this assumption can be relaxed, as discussed later on).

1.3 Thesis Structure

This thesis is structured in the following manner. Chapter 2 reviews the literature on both marked Poisson models and Markov-modulated Poisson processes, which are the main concepts used in our research. We outline both the theoretical frameworks for the model and the estimation procedures that exist for calibrating the parameters of our model.

Chapter 3 lays out the foundations for the model and introduces the main contributions of this thesis, which is a method for incorporating various real world insurance data characteristics in the proposed claims count model. We also detail a procedure for simulating and calculating the final outstanding claim numbers.

Chapter 4 analyses and discusses the chosen calibration method for the Markov-modulated Poisson process. Tests are conducted using simulated data which aim to consolidate the findings in the later empirical case study. Both the assumptions made by the model and the accuracy of the calibration method are examined.

Chapter 5 is the empirical case study and explicitly sets out our adjustments for each of the characteristics that are encountered in the data. The final claim count results are determined. The performance of the model is also compared with the popular chain-ladder algorithm

Finally, we conclude with a summary of the contributions of the thesis in Chapter 6 and outline interesting and important extensions for our claim counts model.

We also include an appendix which is split into two sections. The first gives a small discussion on the fitting of the severity component of the claims reserving model. The second provides the major MATLAB codes that were used in modelling the case study from Chapter 5.

CHAPTER 2

LITERATURE REVIEW

This chapter is split up into sections which describe key areas of research. The first section describes the current aggregate modelling techniques that are commonly used in practice. Section 2 discusses various issues that may be encountered when using these aggregate techniques. Section 3 will explore the current literature on marked Poisson processes while Section 4 introduces the theoretical background behind Markov-modulated Poisson processes and provides the fundamental results that we will use in our model. Finally, Section 5 introduces the concepts behind operational time adjustments to Poisson point processes.

2.1 Aggregate Modelling Algorithms

In this section, we introduce some of most popular aggregate reserving procedures for general insurance. The common procedure in all of these methods is that they begin by separating the claim payments into periods, usually accident years (denoted as i) and development years (denoted as j). Assuming that the data is complete, we will have an equal number of accident years and development years, and we write this number as n. It is implicitly assumed by these models that there are no more developments after the nth period.

We further denote the accumulated claims amount (either counts, paid or incurred) at accident year i and development year by $C_{i,j}$, for $i = 1, \ldots, n$ and $j = 1, \ldots, n$. We can then arrange these amounts into a triangle by accident year and development year along the columns and rows respectively. Clearly, we will not have data for the points

where $j > i$ and thus, the bottom right half of this triangle must be projected from the upper left of the triangle. The lower triangle is the outstanding claims amount that we are trying to estimate.

It is important to note that the algorithms described here are deterministic in nature. Various extensions to stochastic algorithms have been proposed. For example, a very popular stochastic model is the distribution-free chain ladder model described in Mack (1993). An important benefit of the extension is that the prediction uncertainty in the outstanding claims can be quantified and we can obtain estimates for the other moments, whereas the algorithms below can only provide central estimates. However, we introduce the deterministic methods below due to their simplicity and tractability, and our current analysis only requires the first moment for comparison purposes.

2.1.1 Chain-ladder algorithm

The chain-ladder algorithm assumes that the claims develop independently from their accident year, means that the factors that we apply only need to account for the development period over which we are projecting. It also assumes that cumulative claims amounts are proportional to the cumulative claim amount in the preceding period. We call this proportion a development factor between the two development years and write this formally as

$$\mathbb{E}\left[C_{i,j+1}|C_{i,1},\ldots,C_{i,j}\right] = f_j \times C_{i,j}, \tag{2.1}$$

where f_j are the development factors between each development year. The factors f_j are estimated by

$$\hat{f}_j^{CL} = \frac{\sum_{i=1}^{n-j} C_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}}, \tag{2.2}$$

which is essentially the sum of all known claims for development period $j+1$ divided by the claims from development period j, ignoring the years where there are unknown cells for either development period. Thus, we can apply the following formula across all the unknown cells in order to estimate their values:

$$C_{i,j+1} = f_j \times C_{i,j}. \tag{2.3}$$

An important (and somewhat limiting) characteristic of the above algorithm is that it doesn't incorporate any dependencies between accident years. Also, as implied above, there are various stochastic extensions to this algorithm such as Mack's chain-ladder model and the over-dispersed Poisson chain-ladder model (see Wüthrich (2015) for further details).

2.1.2 The Bornhuetter-Ferguson algorithm

The Bornhuetter-Ferguson (BF) algorithm (Bornhuetter and Ferguson, 1972) applies a similar method to Mack's chain-ladder method but also incorporates expected ultimate claims using an external factors such as expected loss ratios. The final aim is to calculate ultimate claims for each accident year and the relevant formula for the BF method is

$$\hat{C}_{i,n}^{BF} = C_{i,j} + \left(1 - \frac{1}{F_{i,j}}\right)\hat{C}_{i,n}, \tag{2.4}$$

where the factors $F_{i,j}$ are the development factors from period j until period n. These factors are calculated externally and could be determined using the chain-ladder factors from Section 2.1.1. In a similar vein, $\hat{C}_{i,n}$ is also calculated externally using quantities such as loss ratios.

Generally, the BF model is used for highly leveraged lines of business, where the standard chain-ladder approach would be to volatile to give reasonable estimates for immature years. In this sense, the model can also be seen as a credibility weighted model. However, the subjective nature of calculation of the various $\hat{C}$ terms here are a significant issue.

2.1.3 Other aggregate methods

There are other methods available that take different approaches to calculate reserves such as the Cape-Cod method which estimates the a-priori loss ratio in the previous Bornhuetter-Ferguson method using a factor for exposure and claims to date. There are also methods which combines the reserving procedures previously mentioned, such as the Benktander-Hovinen method which is essentially uses a weighting factor with the assumed proportion of reported and IBNR claims, which removes the subjectivity of the $\hat{C}$ estimates. For descriptions of these more complex methods, we refer to Dahl (2003).

2.2 Limitations of Macro-Level Modelling

While the relative simplicity and low processing cost of macro-level modelling makes it an attractive option when valuing future cash flows, there exist well documented issues with the procedure. We present some of the problems that can arise with these methods when the assumptions underlying the model are not satisfied. Again, while various works have attempted to adjust the models in order to overcome these concerns, the amendments can be complex and may not work simultaneously. This exposes a potential inability of these aggregate representations to capture the intricacies within claim liabilities. A major motivation of our research is that micro-level modelling will naturally avoid these complications by incorporating more information and relaxing some of the rigid assumptions that are made by macro-level modelling procedures.

2.2.1 Problems with prediction accuracy and precision

While macro-level modelling techniques will produce correct results under a homogeneous data structure, any deviations from homogeneity will result in inaccuracies in the final output. we detail some of the characteristics in macro-level models that can result from heterogeneous data below.

2.2.1.1 Lack of robustness

it is also know that the chain-ladder method is not robust and outliers can have a significant impact on the final outstanding claim amounts estimated by the model. The following example is borrowed from the Verdonck et al. (2009):

	1	2	3	4	5	6
1	12,000	6,000	600	300	150.0	15.00
2	13,000	6,500	650	325	162.5	.
3	10,000	5,000	500	250	.	.
4	12,000	6,000	600	.	.	.
5	11,000	5,500	.	.	.	.
6	10,000	.	.	.	.	.

Figure 2.1: Triangle of Claims Data

The upper triangle shown in the above table shows the observed data (including IBNR) and satisfies the chain-ladder assumptions because we can see that the ratios

between the columns are the same for each accident year. Figure 2.2 shows what the bottom triangle looks like after being projected using the appropriate chain-ladder factors. The total reserves calculated in this case is 7,482.5.

	1	2	3	4	5	6
1	12,000	6,000	600	300	150.0	15.00
2	13,000	6,500	650	325	162.5	16.25
3	10,000	5,000	500	250	125.0	12.50
4	12,000	6,000	600	300	150.0	15.00
5	11,000	5,500	550	275	137.5	13.75
6	10,000	5,000	500	250	125.0	12.50

Figure 2.2: Claims Data and Projections

Finally, the claim amount in accident year 1 and development period 2 (i.e. $C_{1,2}$) is multiplied by 10 so that it becomes an outlier. We project the triangle again using the chain ladder method and we obtain Figure 2.3:

	1	2	3	4	5	6
1	12,000	**60,000**	600	300	150.0	15.00
2	13,000	6,500	650	325	162.5	4.24
3	10,000	5,000	500	250	52.71	3.24
4	12,000	6,000	600	150.35	62.75	3.86
5	11,000	5,500	311.45	135.89	56.72	3.49
6	10,000	14,310	458.87	200.21	83.57	5.14

Figure 2.3: Claim Project with an Outlier at $C_{1,2}$

The total outstanding liability in this case is 15,842.49, which is more than twice of the our previous amount without the outlier. Thus, we can see that the chain-ladder method is not very robust in the presence of outliers. As an alternative, the use of micro-level data allows for the explanation and incorporation of the outlying using other external factors.

2.2.1.2 Large prediction errors

England and Verrall (2002) demonstrate several examples of stochastic claims reserving using various models such as Mack's chain-ladder model, the negative binomial model, the over-dispersed Poisson model. In these cases, they find that the prediction error of

these models can be "disappointingly large", and give many possible explanations for this observation. Firstly, the model that is used can be a poor fit of the data which would have clear consequences on the prediction error. However, the authors state that the size of these errors is

> "entirely consistent with the small sample of data that is available in run-off triangles."

Thus, there is an inherent issue with the aggregation of claims as it greatly reduces the sample size which can increases prediction error. There is a clear benefit here of using micro-level data to model claims reserves.

2.2.1.3 Mean-regressing bias

Halliwell (2007) highlights the problem of bias within the results the chain ladder method, as well as the Bornhuetter-Ferguson and additive methods for claims reserving. While there is no specification as to the direction of the bias, the author reference the simulations by Stanard (1985), which show that in general, the chain-ladder method over-predicts the outstanding liabilities in the run-off triangles. Their analysis concludes that the chain-ladder method is biased, with the bias most commonly being mean-regressing, due to earlier cumulative claim amounts being used as proxies for the exposure associated with the triangle. This is validated when the author adjusts for exposure in the Bornhuetter-Ferguson and additive models which results in reduced prediction error. This bias is also rigorously demonstrated in Taylor (2003). We note that this bias is specific to the chain ladder method and thus would not be present in micro-level models.

2.2.2 Problems with practical implementation

2.2.2.1 Issues with non-positive cells

Kunkler (2004) reveals a practical issue that can occur when modelling using an incremental claims triangle instead of the cumulative triangle presented previously. An assumption that is usually made is that each cell is positive or in some cases, non-negative. If there are a large number of zeroes in the data (as may be the case for later development periods), this will affect the calculation of the development factors. However, this many not be the case in practice due to considerations such as recovery values and over-payment of claims. This is especially relevant for reinsurers, where the incremental triangle of claims may have significant numbers of zero cells for excess layers. On the other hand, micro-level models do not encounter issues with non-positive transactions. Indeed, several proposed models (for example Larsen (2007)) specifically incorporate such transactions into the

modelling procedure.

2.2.2.2 Combination of different sources of information

There is a practical question of whether to use paid or incurred claims in the triangle representations. Incurred claims are calculated as the amount that has been paid with a case estimate of the outstanding amount added on top. These case estimates are done by case managers and thus, can be relatively subjective. Posthuma et al. (2008), Quarg and Mack (2004) and Merz and Wüthrich (2010) all propose methods to combine paid and incurred losses into a comprehensive model but their differing methodologies imply the inability of aggregate reserving methods to provide a natural solution to this problem. While this issue is still somewhat relevant for micro-level models, it is feasible to model both the paid and incurred claims by considering both payments and changes in case estimates within a single framework.

2.2.2.3 Inconsistency of reserve and premium calculations

Parodi (2014) points out that premium determinations are done using micro-level data. For example, when purchasing private motor insurance, insurers will use information such as the age of the policyholder to price their premiums. This is misaligned with reserving methodologies that are done by using aggregate techniques, where all the claims regardless of age would be summed into the triangle representations discussed previously. By using micro-level models, we can naturally incorporate the factors that are considered in premium calculations into reserving models and resolve this inconsistency.

2.2.3 Lack of specification of underlying processes

A final issue that is examined in Taylor and McGuire (2004) is that there is no theoretical framework behind the chain-ladder algorithm when applied in practice. The authors highlight that corrective action for the errors that are generated by this method may be difficult to diagnose and rectify as there is no specified underlying distribution. Because of this fact, the authors label the method as "phenomenological", meaning that the procedure reflects little of the underlying mechanisms that govern claim payments. The key observation here is that the chain ladder model proposed by Mack (1993) is distribution-less. Apart from the previous problems, this can also cause other issues, such as difficulty in calculating meaningful estimates of various statistical measures (for example, in the calculations of moments of order higher than 2). This is in stark contrast to the micro-level model that we present in this thesis which has a well defined theoretical foundation.

2.2.4 The effect of macro-level modelling issues

Ultimately, the issues presented above cause increased prediction error in the macro-level models. This error can arise in two forms, either bias or high variability in the final results. In the case of the former, this could lead to long-term under/over-estimation of the reserves. The latter case is a significant issue because reserving calculations are done relatively infrequently in practice. This is due to the large amount of work that is involved with practically implementing macro-level modelling procedures.

An alternative way to view the prediction error is to decompose it into model and parameter error. In the case of the latter, ad-hoc adjustments have been suggested and implemented in practice but this deviates from the theoretical framework that has been established and the resulting procedure has no root in a well-specified model (as addressed in Taylor and McGuire (2004)). In terms of model error, it seems like the natural (and perhaps singular) way to resolve the issues is to use micro-level data. This can bring many benefits such as a lower variability in the outputs of the model, which arises from the use of more information in the modelling procedure. This would mean that a company would be able to hold less capital as the quantile amount required from a regulatory standpoint would be lower. For example, from an Australian perspective, the Value at Risk amount required by the APRA Prudential Standards would be lower when the model produces results that are less variable. Thus, an example of an advantage provided by micro-level techniques may be lower regulatory capital requirements due to the increased precision in the model results.

2.3 Micro-level Analysis

The following section explores the current literature on micro-level analysis and investigates/compares the differing approaches proposed by various authors. The main method we will be implementing extends the framework established by Norberg (1993), which employs a marked Poisson process in order to conduct micro-level analysis. We begin with an introduction to the basic notation and definitions that will be used.

2.3.1 Counting processes

A counting process $\{N(t); t > 0\}$ is a stochastic process with values that are increasing positive integers. These processes are suitable for modelling claim counts. A well known counting process for this purpose is the Poisson process, described below. This process has several attractive properties such as independent and stationary increments as well as the Markov property. It is also fully defined by one intensity parameter and has a closed form solution when calibrating this parameter using maximum likelihood methods.

We are interested in a generalisation of the Poisson process: the doubly stochastic Poisson process, or Cox process. This has been applied extensively in many financial areas including credit risk and mortality (see Lando, 1998; Biffis, 2005). These processes allow for increased modelling flexibility due to the stochasticity of the claim intensity parameter, which allows inhomogeneity due to various factors such as seasonality to be captured by the model, leading to more robust and realistic results.

In the following sections, we provide formal definitions for the concepts discussed above.

2.3.1.1 Poisson distributions

Discrete distributions are a useful tool for modelling claim counts over a certain period of time. A popular distribution that is used for this purpose is the Poisson distribution, defined below.

Definition 2.3.1 *A discrete random variable X is Poisson distributed with intensity parameter λ if*

$$Pr(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots \tag{2.5}$$

It can further be shown that

$$\mathbb{E}[X] = \lambda \quad \textit{and} \quad Var[X] = \lambda. \tag{2.6}$$

2.3.1.2 Poisson processes

A Poisson process models the occurrence of a series of random observations over a specified time period using the above Poisson distributions. Let $\{N(t), t \geq 0\}$ be the number of observations by time t, and let $N(0) = 0$. Then for a time interval $(t, s]$, $N(s) - N(t)$ is the number of observations in the interval. It is assumed that claims cannot occur simultaneously. Independent and stationary increments are also assumed, as described below.

1. A Poisson process has independent increments if the increments $N(s) - N(t)$ and $N(u) - N(v)$ are independent random variables, given that the time intervals $(t, s]$ and $(u, v]$ do not overlap.
2. A Poisson process has stationary increments if the distribution of the increment $N(s + t) - N(t)$ is independent of t.

Using the above definitions, a Poisson process is formally defined as the following:

Definition 2.3.2 *Under the assumptions specified above, a process is a Poisson process if it has independent and stationary increments and the process has a constant intensity parameter $\lambda > 0$ such that for $s > t$, $N(s) - N(t)$ is Poisson distributed with intensity parameter $\lambda(s - t)$, i.e.*

$$Pr(N(s) - N(t) = k) = e^{-\lambda(s-t)} \frac{(\lambda(s-t))^k}{k!}, \quad k = 0, 1, 2, \ldots \tag{2.7}$$

The Poisson process also has the Markov/memoryless property, which means that given the present state of the process, future states of the process are independent of past states. This is expressed more formally as

Let $(\{\mathcal{F}_t\}, t > 0)$ be the filtration of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A Poisson process with filtration $\{\mathcal{F}_t\}$ is said to have the Markov property if for each $s > t$,

$$Pr(N(s) = k|\mathcal{F}_t) = Pr(N(s) = k|N(t) = l), \tag{2.8}$$

for some integers k, l where $k > l$.

2.3.1.3 Inhomogenous Poisson Processes

We can generalise the Poisson process by allowing the intensity parameter λ to change over time (i.e. $\lambda(t)$). In this case, the process is known as an inhomogenous Poisson

process. We define a mean value function as follows:

$$m(t) = \int_0^t \lambda(s)ds \tag{2.9}$$

Then the process $N(t)$ with the independence, stationarity and Markovian properties described above and a changing intensity parameter $\lambda(t)$ will be an inhomogeneous Poisson process and at each time t, $N(t)$ will have have a Poisson distribution with intensity $m(t)$, i.e.

$$\Pr(N(t) = k) = e^{-m(t)}\frac{m(t)^k}{k!}, \text{ where } k = 1, 2, 3... \tag{2.10}$$

Various literature demonstrates the benefits of using an inhomogeneous Poisson process to fit count data relative to the simple homogeneous Poisson case. For example, Mikosch (2009) demonstrates using Danish fire insurance data that a Poisson process with varying intensity fits the data significantly better than the constant intensity counterpart.

2.3.1.4 Doubly stochastic Poisson processes

A further generalisation can be made if we allow $\lambda(t)$ itself to be a stochastic process. In this case, the counting process is known as a doubly stochastic Poisson process or a Cox process. This is a very useful generalisation for insurance data, as it can be argued that factors that can cause inhomogeneity in claim arrivals (such as weather effects or seasonal impacts) may be considered stochastic in nature. Thus, the Cox process provides a realistic adaptation of the standard inhomogeneous Poisson process in the context of insurance data. Avanzi et al. (2015) provides an example of the benefits of such an approach through the use of a shot-noise stochastic intensity. We discuss this paper in greater depth in Section 2.3.8.4.

2.3.2 Claim information

We can represent the basic information necessary to implement a micro-level model as the space $\{(T_i, Z_i)\}_{i=1,\dots,n}$, where n is the total number of claims in the observed period. Looking at a single claim, we write

$$(T, Z) = \{T, U, \{v', Y(v'); 0 \leq v' < V\}, V, Y\}, \tag{2.11}$$

where

- T is the time of occurrence
- U is the notification delay
- v' are the times of the partial payments, beginning from the notification time
- $Y(v')$ are the cumulative payments up until time $T + U + v'$
- V is the time of the last partial payment, or the settlement date
- $Y = Y(V)$ is the ultimate claims cost

A visual representation of how the cumulative claim payments $Y(v')$ develop over time is shown in Figure 2.4.

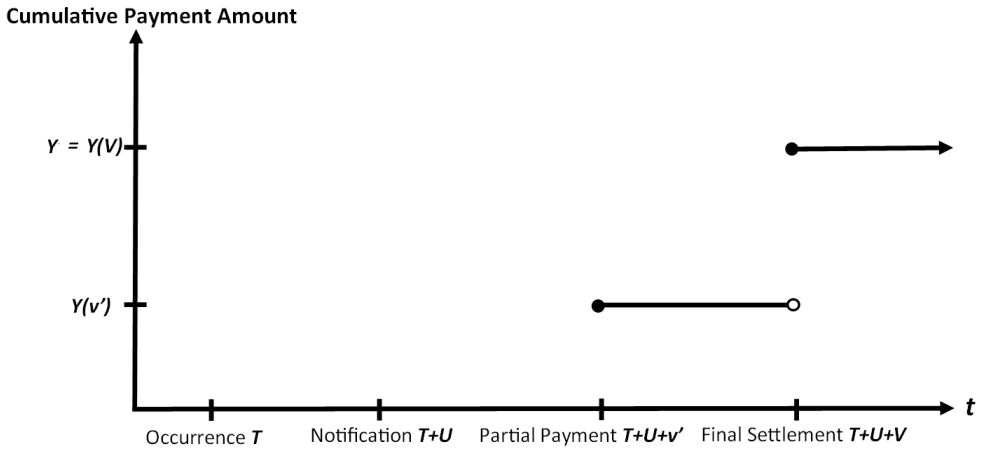

Figure 2.4: Claim occurrence and development

2.3.3 Position dependent marked Poisson processes

In our model, we build on the fundamental framework in Norberg (1993), which we outline in the following section. Assume that claims occur as a non-homogeneous Poisson process with intensity $w(t)$ at time $t \geq 0$, where $w(t)$ is the risk exposure per time unit at time t. This gives us independent increments in the claims counting process $N(t)$ which is determined by the occurrence epochs T_t, where $T_t = \inf\{t; N(t) \geq t\}$. Also, we have that

U is defined as the waiting time from occurrence until notification to the insurer, and V is the waiting time from notification until final settlement of the claim. We let random elements $\{Z(T_t)\}_{t\geq 0}$ be mutually independent and independent of the counting process $\{N(t)\}_{t\geq 0}$.

The distribution of $Z(t)$ is denoted as $P_{Z|t}$ and we use these elements to describe the individual claim developments $Z_t = Z(T_t)$ for $t = 1, ..., N$. Thus, we obtain the marked Poisson process with intensity $\{w(t)\}_{t\geq 0}$ and position dependent marking by $\{P_{Z|T_t}\}_{t\geq 0}$, i.e.

$$\{(T_t, Z_t)\}_{1\leq t\leq N} \sim \text{Poisson}(w(t), P_{Z|T_t}; t \geq 0). \tag{2.12}$$

Let Y_t be the individual final claim amount for a claim that arrives at time t. We derive the marginal distribution P_y from the joint distribution $P_{T,Z}(dt, dz)$ to obtain

$$P_Y(dy) = \frac{\int_{t>0} w(t) P_{Y|t}(dy)\, dt}{\int_0^\infty w(t)\, dt}. \tag{2.13}$$

We then define a position dependent marking of the marked Poisson process using the definition from Last and Brandt (1995):

Definition 2.3.3 *A marked Poisson process $\Phi = \{(T_n, X_n)\}_{n=1,...,N}$ is a position dependent marking of $\bar{\Phi} = \{T_n\}_{n=1,...,N}$ if $\{X_n\}_{n=1,...,N}$ are conditionally independent given $\bar{\Phi}$ and*

$$Pr\left(X_n|\bar{\Phi}\right) = G\left(T_n, dx\right), \tag{2.14}$$

where G is some stochastic kernel from $\mathbb{R}^+$ to $\boldsymbol{X}$.

Using these results, we have that the total claim amount X is a generalised Poisson variate distributed as Poisson$(\int_0^\infty w(t)dt, P_Y)$.

2.3.4 Claim decomposition by development category

2.3.4.1 The decomposition result

A key result provided by Norberg (1993) is the decomposition of the claim space into 4 mutually exclusive and exhaustive categories: covered but not incurred (cni), incurred but not reported (inr), reported but not settled (rns) and settled (s). The total outstanding amount X^O is the main quantity that we wish to predict each time τ, and comprises of the outstanding claims reported but not settled X^{orns} and the total of X^{cni} and X^{inr} which we define as X^{nr}. Using these definitions, we obtain from Norberg (1993) that the

component claim processes above are independent and that for $g = cni, inr, rns, s$,

$$\{(T_t^g, Z_t^g)\}_{1 \leq t \leq N^g} \sim \text{Poisson}\left(w^g(t), P_{Z|t}^g; t \geq 0\right) \tag{2.15}$$

with

$$w^g(t) = w(t) P_{Z|t}\left\{\xi_t^g\right\}, \tag{2.16}$$

and

$$P_{Z|t}^g\{dz\} = \frac{P_{Z|t}\{dz\}}{P_{Z|t}\left\{\xi_t^g\right\}}, z \in \xi_t^g, \tag{2.17}$$

where ξ_t^g denotes the set of developments that produce a claim at time t in category g.

2.3.4.2 Outstanding claims liability calculation

Norberg (1993) applies the decomposition result above to get an expression for the outstanding claims liability. At time τ, the outstanding RBNS claims are equal to

$$X^{orns} = \sum_{1 \leq i \leq N^{rns}} \left(Y_i - Y_i\left(\tau - T_i - U_i\right)\right), \tag{2.18}$$

while the not reported claims are equal to

$$X^{nr} = X^{inr} + X^{cni} = \sum_{1 \leq i \leq N^{nr}} Y_i^{nr}. \tag{2.19}$$

Our appropriate estimate for the expected outstanding claim liabilities $m_{X^O}$is thus the sum of the first moments $m_{X^{orns}|\mathcal{F}_\tau} + m_{X^{nr}}$, where $\mathcal{F}_\tau$ represents the information available by time τ. We have that

$$m_{X^{orns}|\mathcal{F}_\tau} = \sum_{1 \leq i \leq N^{rns}} \int_0^\infty (y - y_\tau) P_{Y_i^{rns}|T_i, u_i, y_i}(dy), \tag{2.20}$$

where $y_\tau = Y_i - Y_i\left(\tau - T_i - U_i\right)$. Also, we note that

$$w^{nr}(t) = w(t)\left(1 - P_{U|t}(\tau - t)\right) \tag{2.21}$$

and that

$$P_{Y|t}^{nr}(dy) = \frac{\int_{u > \tau - t} P_{U,Y|t}(du, dy)}{1 - P_{U|t}(\tau - t)}, \tag{2.22}$$

which gives that

$$m_{X^{nr}} = \int_{t>0} w^{nr}(t) \iint_{y>0} y P^{nr}_{Y|t}(dy)dt \tag{2.23}$$

$$= \iint_{t>0} w(t) \int_{u>\tau-t} \iint_{y>0} y P_{U,Y|t}(du, dy)dt. \tag{2.24}$$

2.3.5 General claim decomposition

An important generalisation to the decomposition result above is provided by Norberg (1999). Firstly, partitioning the claim space into mutually exclusive and collectively exhaustive categories will produce similar independence results to the decomposition by development categories. Formally, we obtain component g-claims which are independent with

$$\{(T^g_t,\, Z^g_t)\}_{1\le t\le N^g} \sim \text{Poisson}\left(w^g(t), P^g_{Z|t}; t \ge 0\right) \tag{2.25}$$

where

$$w^g(t) = w(t) P_{Z|t}\left\{\xi^g_t\right\}, \tag{2.26}$$

and

$$P^g_{Z|t}\{dz\} = \frac{P_{Z|t}\{dz\}}{P_{Z|t}\left\{\zeta^g_t\right\}}, z \in \xi^g_t, \tag{2.27}$$

for any finite set of categories $g = g_0, g_1, g_2, \ldots, g_q$. Other formulae corresponding to the results above can also be derived in a similar manner to the previous section. This general decomposition allows the model to analyse data using various factors of interest such as claim amount, franchise, reinsurance, calendar year and year of notification. It is also possible to measure the effect of interactions by decomposing across multiple factors, as demonstrated by Hesselager (1995) in the case of calendar year and year of notification.

2.3.6 General considerations for modelling claim developments

The time dependence within the generic mark $Z = (U, V, Y, Y')$ can be due to various trends in risk conditions which can be accounted for by parametrisation and scaling. Thus, it is reasonable to assume independence with respect to t, so that we can write the distribution of the generic mark Z as P_Z. The (U, V, Y) components can be thought of as the primary characteristics of the claim and the cumulative partial payments Y' are dependent on these characteristics. Thus, an intuitive approach to modelling P_Z is to first determine the marginals of (U, V, Y), which is made easier by the decomposition result above. We can then determine the conditional distribution of Y' given (U, V, Y).

2.3.7 Decomposition results using intensity measures

We can rewrite the results in Section 2.3.4.2 in a more interpretable format, as shown in Jin (2013). Using the notation from the previous sections, we can express the distribution of the mark in Equation (2.12) as

$$P_{Z|T} = P_{U|T} \times P_{D|T,U}, \tag{2.28}$$

where $Z = (U, D)$ with U as the random process for the reporting delay and $\mathcal{D}$ is the random development process associated with the claim (i.e. the partial payments). Using this expression, the claims are modelled as random elements in the claim space

$$\mathcal{C} = [0, \infty] \times Z = [0, \infty] \times [0, \infty] \times \mathcal{D} \tag{2.29}$$

with intensity measure

$$\lambda(dt) \times P_{U|T}(du) \times P_{X|T,U}(dx). \tag{2.30}$$

2.3.8 Implementation and extensions

There are a small number of papers in the literature on individual claims analysis that provide a demonstration of how to implement micro-level models in practice using the framework by Norberg, described above. In these papers, the various authors (Larsen, 2007; Antonio and Plat, 2013; Jin, 2013; Avanzi et al., 2015) also provide some useful extensions in terms of the distributions and covariates that are used. The following subsections will describe the key contributions of their papers.

2.3.8.1 Larsen (2007)

Larsen (2007) models the development of incurred losses by using the decomposition result in Section 2.3.5 in order to separate the claims space by characteristics such as line of business and claim type. Larsen firstly models the claim arrival times T by using parametric assumptions. In this step, he also incorporates seasonality by partitioning each period into seasonal units of time, and the parametric claim arrival times will be a function of a parameter representing the seasonality effect and a parameter representing the arrival rates for an accident year, i.e.

$$\lambda(t) = \lambda_i \sigma_m, \tag{2.31}$$

where m is the effect of seasonality in the mth seasonal period and λ_i is the risk frequency in accident year i.

The author then employs another parametric approach to specify the conditional distribution of the reporting delay U and then further iterates this approach by specifying the conditional distribution of the claims developments (including the partial payment frequency V and severity Y) given $\{T, U\}$. The severity of these partial payments are modelling using an extreme value approach, where losses smaller than some threshold L are modelled using a Gamma distribution while losses larger than L use a Generalised Pareto Distribution.

A key contribution of this paper is that Larsen relaxes independence assumptions between the various claim components $\{T, U, V, Y\}$ by using a regression approach and incorporating the components as covariates. Various other covariates of interest are also suggested such as case estimates but are not implemented.

2.3.8.2 Antonio and Plat (2013)

Antonio and Plat (2013) use a similar approach to Larsen (2007) but use probability distributions instead of fitting functions to some of the claim components such as arrival times. Different distributions are also explored and a key insight in this paper is that the piecewise combination of distribution functions may be necessary for components such as the reporting delay. In this specific case, the author fits a Weibull distribution with 9 degenerate components representing the initial 0-8 days. However, this paper does not allow for any dependency structures within the components of the claims process nor any variation in the claim frequency due to trends such as seasonality.

Another key contribution in this paper is the introduction of the three different event types that can occur in the claim development process for the partial payments: Payment without settlement, settlement without payment and payment with settlement. This mutually exclusive and collectively exhaustive set of events assists in the simulation of the final results. The use of hazard rates corresponding to these events allows for a relatively uncomplicated integration of this decomposition into the micro-level analysis framework by Norberg (1993) due to the Poisson nature of the partial claim arrivals. These hazard rates are modelled using a piece-wise constant function as this allows for closed form results but a Weibull distribution is also dicussed briefly.

2.3.8.3 Jin (2013)

Jin (2013) builds on the procedure used by Antonio and Plat (2013) and a key contribution of this paper is that it introduces two covariates that is commonly available in real world

general insurance data sets:

1. The claim status indicator
2. The initial case estimate

The inclusion of these covariates in the payment severity process through a linear model in the parameters of the chosen severity distribution allows the micro-level information to be incorporated into the claims development process. For example, the claim status indicator will serve as a proxy for the information that is incorporated in the model as the development has not yet occurred in the claims process. The initial case estimate would also be a good indicator as to the expected values that would be paid out in each partial payment.

Jin (2013) also extends the linear modelling approach to the hazard rates, so that they can be written as

$$h_{i,k}(t) = h_k^{(0)}(t) \exp(\boldsymbol{x}'\boldsymbol{\alpha}), \tag{2.32}$$

where the covariates are the accident years so that the effect between accident years on the hazard rates can be incorporated.

2.3.8.4 Avanzi et al. (2015)

Avanzi et al. (2015) uses a comparable approach to our proposed model by modelling claim frequency using a Cox process which assumes an inhomogeneous claim intensity over time. This intensity is modelled in this paper using a shot-noise process where there is an initial jump in the claim frequency before it decays at an exponential rate. This provides a realistic extension to the procedures by the authors in the previous subsections and is appropriate where there are events that cause sudden increases in the probability of claim occurrences or where there are sudden fluctuations in the intensity rate. The shot-noise approach also allows for autocorrelation in the claims process, which is not considered by the previously discussed papers.

A key contribution of this paper is the allowance for the stochastic reporting delay in the intensity parameters, in order to incorporate the fact that there are claims that have occurred but not yet been reported to the insurer (IBNR). These claims should be included in the modelling process but are unobservable at the time of valuation. The estimation of the unobservable intensity is done after employing a Reversible Jump Markov Chain Monte Carlo filtering algorithm to separate the effects of the reporting delay and the shot noise process. We aim to adapt this procedure to the Markov-modulated Poisson process, defined below.

2.4 Markov-modulated Poisson Processes

By assuming that the risk intensity parameter can be represented by an unobservable Markov chain, we can allow for changes in the frequency of claim occurrences over time. As previously stated, this provides realism to the claims reserving model because in practice, there are unobservable environmental variables that can cause regime-switching in the risk frequency of claims.

2.4.1 Hidden Markov models

A hidden Markov model (HMM) is a model where there exists a Markov chain with underlying states $\{M_i\}$ that are not directly observable. Instead, it is only possible to observe another stochastic process $\{Y_i\}$, which depends solely on the $\{M_i\}$ terms. We can define this more precisely as the following:

Definition 2.4.1 *A hidden Markov model is a bivariate discrete process* $\{M_i, Y_i\}_{i\geq 0}$ *where* $\{Y_i\}$ *is a sequence of independent random variables that have a conditional distribution that only depends on the unobservable Markov chain* $\{M_i\}$.

For a discrete time hidden Markov model, a simple representation is provided in Figure 2.5 below, where we assume the Markovian property on the states M at time t. The state space sequence over time $\{M(t); t \geq 0\}$ is not observable but we can observe their effect on a corresponding process $\{Y(t); t \geq 0\}$.

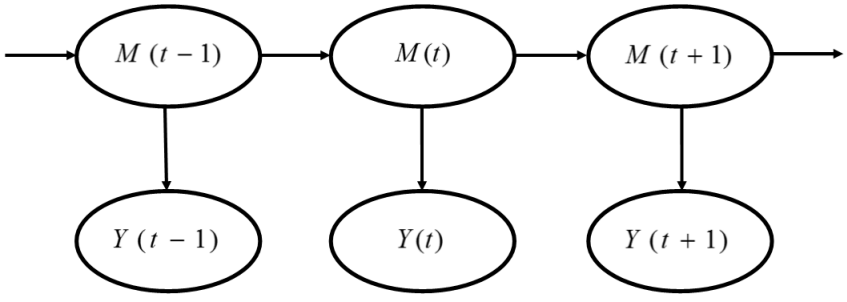

Figure 2.5: A simple hidden Markov model

2.4.2 Markov-modulated Poisson processes

A doubly stochastic Poisson process, or Cox process, is a Poisson process where the intensity parameter $\lambda(t)$ is also a stochastic process. We are interested in the Markov-modulated Poisson process (MMPP), where the intensity is controlled by a hidden continuous-time Markov chain with a finite state space. Let the Poisson arrival rates be $\{\lambda_i\}$, where $i = 1, 2, \ldots, r$ and let the generator matrix Q of the Markov chain be

$$\begin{bmatrix} -q_1 & q_{12} & \cdots & q_{1r} \\ q_{21} & -q_2 & \cdots & q_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ q_{r1} & q_{r2} & \cdots & q_r \end{bmatrix}, \tag{2.33}$$

where

$$q_i = \sum_{j \neq i} q_{ij}. \tag{2.34}$$

We also define

$$I_\lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r) = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_r \end{bmatrix} \tag{2.35}$$

and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_r)^{\mathrm{T}}$.

In the case where Q and Λ is independent of time t, the MMPP is said to be homogeneous and a steady state vector exists such that $\boldsymbol{\pi} Q = 0$ and $\boldsymbol{\pi e} = 1$.

2.4.2.1 MMPP's as Markov renewal processes

A Markov-modulated Poisson process is a Markov renewal process, as defined below.

Definition 2.4.2 *Consider a set of random variables (λ_t, X_t) where λ_t is the state of the Markov Chain at time t and X_t is the inter-arrival time between the $(t-1)$th and tth event. Then the Markov chain is a Markov renewal process if*

$$Pr(X_{t+1} \leq t, \lambda_t = j | \lambda_t, X_t, \lambda_{t-1}, X_{t-1}, \ldots, \lambda_1, X_1, \lambda_0) = Pr(X_t \leq t, \lambda_t = j | \lambda_t, X_t). \tag{2.36}$$

We can see that the process $\{(\lambda_t, X_t)\}$ can be generated by first choosing some starting state λ_0 according to some starting distribution s and then randomly choosing (λ_1, X_1) according to the equation above.

2.4.2.2 Fundamental results of Markov-modulated Poisson processes

Meier-Hellstern (1987) shows that the transition density matrix of a Markov renewal process is equal to

$$f(m) = \exp\{(Q - I_\lambda) m\} I_\lambda, \tag{2.37}$$

which gives the transition probability matrix of the unobservable Markov Chain with states $\{\lambda_i\}$ as

$$\int_0^\infty f(m)\, dm = (Q - I_\lambda)^{-1} I_\lambda. \tag{2.38}$$

Rydén (1994) demonstrates that for each (Q, Λ), we have that all states i in the transition probability matrix where $\lambda_i > 0$ form an irreducible, aperiodic and closed subclass. The remaining states are all transient. A consequence of this result is that the stationary distribution of the transition probability matrix $\boldsymbol{\pi}$ is unique and for each initial distribution s, we have that $\Pr(\lambda_t = i) \to \pi_i$. This simplifies the procedures for parameter estimation that are discussed later. In the following results, we will assume that the Markov chain is already stationary.

2.4.3 Switched Poisson Processes

The simplest non-trivial Markov-modulated Poisson process has two state spaces and this process is known as a switched Poisson process. In this case, we can obtain closed form expressions for the transition densities f given the state spaces $\{\lambda_1, \lambda_2\}$ and transition intensities $\{\mu_1, \mu_2\}$. Define a parameter $\phi = \{\lambda_1, \lambda_2, \mu_1, \mu_2\}$. Then, Rydén (1994) demonstrates the following results:

1. The transition densities of the observations $\{X_i\}_{i=1,2}$ are
   (a) $f_{11}(x) = -\frac{\lambda_1(S_2-\theta_1)}{D}\exp(-\theta_1 x) + \frac{\lambda_1(S_2-\theta_2)}{D}\exp(-\theta_2 x)$
   (b) $f_{12}(x) = -\frac{\lambda_2\mu_1}{D}\exp(-\theta_1 x) + \frac{\lambda_2\mu_1}{D}\exp(-\theta_2 x)$
   (c) $f_{21}(x) = -\frac{\lambda_1\mu_2}{D}\exp(-\theta_1 x) + \frac{\lambda_1\mu_2}{D}\exp(-\theta_2 x)$
   (d) $f_{22}(x) = -\frac{\lambda_2(S_1-\theta_1)}{D}\exp(-\theta_1 x) + \frac{\lambda_2(S_1-\theta_2)}{D}\exp(-\theta_2 x)$
2. The transition probability matrix of the hidden Markov chain $\{M_i\}_{i=1,2}$ is

$$P = \begin{pmatrix} \lambda_1\frac{S_2}{K} & \lambda_2\frac{\mu_1}{K} \\ \lambda_1\frac{\mu_2}{K} & \lambda_2\frac{S_1}{K} \end{pmatrix}$$

where

$$\begin{aligned} S_1 &= \lambda_1 + \mu_1, \\ S_2 &= \lambda_2 + \mu_2, \\ K &= \lambda_1\lambda_2 + \lambda_1\mu_2 + \lambda_2\mu_1, \\ D &= \sqrt{(S_1 + S_2)^2 - 4K}, \\ \theta_1 &= \frac{(S_1 + S_2 + D)}{2}, \\ \theta_2 &= \frac{(S_1 + S_2 = D)}{2}, \end{aligned}$$

from which it follows that the stationary probabilities are

$$\boldsymbol{\pi} = \begin{pmatrix} \frac{\lambda_1\mu_2}{\lambda_1\mu_2+\lambda_2\mu_1} & \frac{\lambda_2\mu_1}{\lambda_1\mu_2+\lambda_2\mu_1} \end{pmatrix} \Bigg($$

Thus, the only parameters left to estimate are the state spaces and transition intensities within our parameter $\phi = \{\lambda_1, \lambda_2, \mu_1, \mu_2\}$. It is also possible to derive expressions for the above in MMPPs of higher order but the complexity of the expressions would be best expressed in the matrix representation seen in Section 2.4.2.1.

2.4.4 Parameter Estimation in MMPPs

The literature on parameter estimation for the state spaces and transition intensities in Markov-modulated Poisson processes generally discusses two approaches: a moment based approach and a maximum likelihood approach. A key result proved in Rydén (1994) is that the maximum likelihood estimator is consistent so the estimator converges in probability to the true value. In our model, the convergence property is assisted by the large sample size available for modelling. Thus, we will use a likelihood approach to estimate the parameters in our MMPP. There are several documented methodologies for this procedure. In the following section, we will present the basic framework for maximum likelihood estimation and briefly detail the main methods that are available in the literature.

Firstly, denote $\phi = (Q, I_\lambda)$ as a parameter within the parameter space

$$\Phi = \{\phi : Q \text{ is irreducible and } \lambda_i > 0 \text{ for at least one } i\}. \tag{2.39}$$

Assuming that $\{X_i\}$ is a stationary process, we have that the likelihood of the parameter ϕ for the sequence of observations $\{X_i\}_{i=1,\ldots,n}$ is

$$\mathcal{L}(\phi; x_1, \ldots, x_n) = \boldsymbol{\pi}(\phi) \left\{ \prod_{i=1}^{n} \mathbf{f}(x_i; \phi) \right\} \mathbf{1}, \tag{2.40}$$

where $\mathbf{1}$ is a $n \times 1$ vector of ones (see Meier-Hellstern, 1984). The complete likelihood, which also assumes knowledge of the hidden state spaces $\lambda_0, \ldots, \lambda_r$, is given by Rydén (1994) as

$$\mathcal{L}^c(\phi; x_1, \ldots, x_n, \lambda_0, \ldots, \lambda_n) = \pi_{\lambda_0}(\phi) \prod_{i=1}^{r} f_{\lambda_{i-1}\lambda_i}(x_i; \phi). \tag{2.41}$$

2.4.4.1 The complete likelihood maximisation method of Meier-Hellstern (1984)

Meier-Hellstern (1984) chooses to maximise the complete likelihood with respect to I_λ and Q, as well as the hidden state spaces $\lambda_0, \ldots, \lambda_r$. An iterative optimisation, first with respect to (I_λ, Q) and then with respect to the hidden state spaces $\lambda_0, \ldots, \lambda_r$ is carried out in order to maximise the likelihood. An attractive feature of using the complete likelihood is that the expression for the standard likelihood contains the product of matrices which can be difficult to maximise, whereas the complete likelihood is the product of scalars. For simplicity, a transformation of the parameters can also be performed. However, it is shown that this method is likely to produce asymptotically biased results due to the misclassification of the hidden regime variables.

2.4.4.2 The likelihood maximisation method

We can also maximise the standard likelihood function from above. Ramesh (1995) demonstrates a procedure to do this for the parameters of a switched Poisson process but his technique can be extended to MMPPs of higher order. However, the product of matrices in the likelihood makes differentiation difficult, so instead, Ramesh uses a downhill simplex optimisation algorithm in order to calculate his parameters. Rydén (1996) proposes an alternative EM algorithm where the missing data is assumed to be the complete trajectory of the hidden Markov chain over the observation interval, meaning that there are no transitions in the Markov chain between the observations that we have. This results in a more computationally simple procedure for parameter estimation. We shall discuss both methods of optimisation in the following sections.

2.4.4.3 The Nelder-Mead downhill simplex algorithm

The Nelder-Mead downhill simplex algorithm is a general optimisation algorithm that is used to minimise functions. We define a simplex as a geometrical figure in N dimensions with $N+1$ vertices. In our case, define the function that we wish to minimise as $f(\boldsymbol{x}) = -\mathcal{L}(\phi; x_1, \ldots, x_N)$. At each of the $n+1$ corners of the figure, we use the following iterative formula:

$$\boldsymbol{P}_i = \boldsymbol{P}_0 + \Delta \boldsymbol{e}_i, \tag{2.42}$$

where the $\boldsymbol{e}_i$ terms are N unit vectors, Δ is a vector of constants that estimate the optimisation problem's characteristic length scale and $\boldsymbol{P}_0$ is the initial starting point. Then, the simplex undergoes transformations that result in a lower value for the function $f(\boldsymbol{x})$. These transformations can be reflections, expansions or contractions, and preserve the N-dimensional finite inner volume of the simplex. Upon getting close to the lowest point, the simplex will contract and converge on the final optimised set of values. We set a tolerance for the distance between steps of the convergence and stop the algorithm when the vector distance for a step is lower than this tolerance. An example of implementing this procedure is given in Press et al. (1989).

An unfortunate complication when using the Nelder-Mead downhill simplex algorithm in this context is that the likelihood we are attempting to maximise may take on very large or very small values that are outside the default floating-point implementations of standard computers. In our empirical study, this problem is compounded by the huge amounts of claims data that we use to calibrate the MMPP model. As a result, customised floating-point software is required and this limits the accessibility of such an algorithm.

2.4.4.4 The expectation maximisation (EM) algorithm for MMPPs

An alternative optimisation method is suggested by Rydén (1996). Suppose that there exist observable variables Y and additional hidden variables X. We denote the likelihood of a parameter ϕ given Y as $\mathcal{L}(\phi; y)$and the complete likelihood given all variables is denoted as $\mathcal{L}^c(\phi; y, x)$. For a given starting parameter estimate ϕ^0, it is known that

$$\hat{\phi} = \arg\max_{\phi} \mathcal{Q}\left(\phi; \phi^0\right) \tag{2.43}$$

satisfies $\mathcal{L}\left(\hat{\phi}; y\right) \geq \mathcal{L}(\phi^0; y)$, where

$$\mathcal{Q}\left(\phi;\, \phi^0\right) = \mathbb{E}_{\phi^0}\left[\log \mathcal{L}^c(\phi; y, x) \,|\, y\right]. \tag{2.44}$$

This means that as we iteratively produce new estimates for ϕ, the likelihoods of these estimates should improve. After enough steps, these estimates will ideally converge to the maximum likelihood estimate.

The maximisation step above is known as the M-step and the calculation of the condition expectation is known as the E-step. Generally, $\mathcal{L}(\phi; y)$ is non-linear in ϕ due to the complex expressions for the transition densities and thus, is difficult to maximise. The explicit E-step of the EM algorithm by Rydén (1996) makes this an attractive extension to the traditional EM algorithm.

Thus, the modified EM algorithm is very useful for calibrating hidden Markov models in general, and is widely used in the literature across many different research areas. We explicitly set out the procedure for applying this algorithm in the following section.

2.4.5 Parameter estimation using the modified EM algorithm

The EM algorithm in Section 2.4.4.4 is applied to the maximisation procedure by Rydén (1996) to demonstrate its use in estimating the parameters of the Markov modulated Poisson process. In the following sections, we will outline the E- and M-steps of the algorithm and provide a computational procedure to calibrate the parameters of our Markov-modulated Poisson process.

2.4.5.1 The explicit E-step in the EM Algorithm

For a sequence of observed inter-arrival times $y_1, \ldots, y_n$ from a stationary MMPP of order r, let t_k be the time of the kth event. We denote the times of the jumps in the hidden Markov chain $\{X(t)\}$ by $\{u_i\}_{i=1,\ldots,m}$ and let $u_0 = 0$ and $u_{m+1} = t_n$. We also write $I_k = [u_{k-1}, u_k)$ as the time interval between Markov chain jumps, s_k as the state of the Markov Chain during I_k and z_k as the number of events in I_k (not counting the event at $t = 0$). The standard counting process for the number of events up until time t (not including the one at $t = 0$) is denoted by $N(t)$. For convenience, we further define the following: $\Delta u_k = u_k - u_{k-1}$, $q_k = q_{kk}$ and $\pi = \pi(Q, I_\lambda)$. The complete likelihood of (Q, I_λ) can then be expressed as

$$\mathcal{L}^c = \pi_{s_1} \left\{ \prod_{k=1}^{m} \left(q_{s_k} \exp\left(-q_{s_k} \Delta u_k\right) \times \frac{q_{s_k, s_{k+1}}}{q_{s_k}} \right\} \exp\left(-q_{s_{m+1}} \Delta u_{m+1}\right) \times \left\{ \prod_{k=1}^{m+1} \left(\frac{\left(\lambda_{s_k} \Delta u_k\right)^{z_k}}{z_k!} \exp\left(-\lambda_{s_k} \Delta u_k\right) \times \frac{z_k!}{\left(\Delta u_k\right)^{z_k}} \right\} \right. \tag{2.45}$$

Given a parameter estimate (Q^0, I_λ^0), the E-step becomes

$$\mathcal{Q}\left(Q, I_\lambda; Q^0, I_\lambda^0\right) = \mathbb{E}_{\left(Q^0, I_\lambda^0\right)}\left[\log \mathcal{L}^c\left(Q, I_\lambda\right) | N\left(u\right), 0 \leq u \leq t_n\right] \tag{2.46}$$

$$= \sum_{i=1}^{r} \hat{\pi}_i \log \pi_i - \sum_{i=1}^{r} T_i q_i + \sum_{i=1}^{r} \sum_{j \neq i}^{r} m_{ij} \log q_{ij} + \sum_{i=1}^{r} \left(\hat{n}_i \log \lambda_i - \lambda_i \hat{T}_i\right), \tag{2.47}$$

where m_{ij} is the total number of jumps in the chain from state i to j, T_i is the total time in the chain spent in state i, which is equal to $\int_0^{t_n} I_{\{X(t)=i\}} dt$, and n_i is the total number of events that occurred while the Markov chain was in state i. The preliminary estimators for each of these are given below. We can drop the first term in the previous expression because the large sample size n makes the first time negligible (see Billingsley, 1961). The expressions for the remaining estimators are

$$\hat{m}_{ij} = \int_0^{t_n} \mathbb{P}_{\left(Q^0, I_\lambda^0\right)}\left[X\left(t-\right) = i, X\left(t\right) = j | \, N(u), 0 \leq u \leq t_n\right] dt, \tag{2.48}$$

$$\hat{T}_i = \int_0^{t_n} \mathbb{P}_{\left(Q^0, I_\lambda^0\right)}\left[X\left(t\right) = i | N(u), 0 \leq u \leq t_n\right] dt, \text{ and} \tag{2.49}$$

$$\hat{n}_i = \sum_{k=1}^{n} \mathbb{P}_{\left(Q^0, I_\lambda^0\right)}\left[X\left(t_k\right) = i | N(u), 0 \leq u \leq t_n\right]. \tag{2.50}$$

Using the above expressions, we can explicitly derive a formula for these preliminary expressions (see Rydén (1996) for the full derivations):

$$\hat{m}_{ij} = \frac{q_{ij}^0}{\pi\left(Q^0, I_\lambda^0\right)\left\{\prod_{k=1}^{n} f\left(y_k; Q^0, I_\lambda^0\right)\right\}\mathbf{1}} \times \int_0^{t_n} \pi\left(Q^0, I_\lambda^0\right) \left\{\prod_{k=1}^{N(t)} f\left(y_k; Q^0, I_\lambda^0\right)\right\} \exp\left\{\left(Q^0 - I_\lambda^0\right)\left(t - t_{N(t)}\right)\right\} \mathbf{1}_i \mathbf{1}_j^\intercal \times f\left(t_{N(t)+1} - t\right) \left\{\prod_{k=N(t)+2}^{n} f\left(y_k; Q^0, I_\lambda^0\right)\right\} \mathbf{1} dt, \tag{2.51}$$

$$\hat{T}_i = \frac{1}{\pi\left(Q^0, I_\lambda^0\right)\left\{\prod_{k=1}^{n} f\left(y_k;\, Q^0, I_\lambda^0\right)\right\}} \times \int_0^{t_n} \pi\left(Q^0, I_\lambda^0\right)\left\{\prod_{k=1}^{N(t)} f\left(y_k; Q^0, I_\lambda^0\right)\right\} \exp\left\{\left(Q^0 - I_\lambda^0\right)\left(t - t_{N(t)}\right)\right\} \mathbf{1}_i \mathbf{1}_j^\intercal \times f\left(t_{N(t)+1} - t\right)\left\{\prod_{k=N(t)+2}^{n} f\left(y_k; Q^0, I_\lambda^0\right)\right\} \mathbf{1} dt, \tag{2.52}$$

and

$$\hat{n}_i = \sum_{k=1}^{n}\left(\frac{1}{\pi\left(Q^0, I_\lambda^0\right)\left\{\prod_{j=1}^{n} f\left(y_j; Q^0, I_\lambda^0\right)\right\}\mathbf{1}} \pi\left(Q^0, I_\lambda^0\right) \times \left\{\prod_{j=1}^{N(t_k)} f\left(y_k; Q^0, I_\lambda^0\right)\right\} \mathbf{1}_i \mathbf{1}_i^\intercal \left\{\prod_{j=N(t_k)+1}^{n} f\left(y_k; Q^0, I_\lambda^0\right)\right\} \mathbf{1}\right). \tag{2.53}$$

where $\mathbf{1}_i$ is a vector of zeroes except for a one at the ith entry, and as defined previously, $f\left(y; Q, I_\lambda\right) = \exp\left\{\left(Q - I_\lambda\right) y\right\} \Lambda$. These estimators will provide the variables that are trying to calibrate in the Markov-modulated Poisson process model using the following expressions:

$$\hat{q}_{ij} = \frac{\hat{m}_{ij}}{\hat{T}_i}, \tag{2.54}$$

where $i, j = 1, \ldots, r, i \neq j$ and

$$\hat{\lambda}_i = \frac{\hat{n}_i}{\hat{T}_i}, \tag{2.55}$$

where $i = 1, \ldots, r$.

These final expressions are intuitively clear as the former estimates the transition probabilities from states i to j by the number of jumps from states i to j divided by the total time spent in state i while the latter estimates the risk frequency pertaining to state i by the number of events that occurred in state i divided by the total time spent in state i.

2.4.5.2 The EM algorithm

The EM algorithm uses forwards-backwards recursions to calculate parameter estimates. Using the above, one step in the EM algorithm is as follows:

Our forwards-backwards equations are set up as

1. Let $L(0) = \pi^0$ and for $k = 1, \ldots, n$, we let $L(k) = L(k-1) f(\Delta t_k)$.
2. Let $R(n+1) = \mathbf{1}$, and for $k = n, \ldots, 1$, we let $R(k) = f(\Delta t_k) R(k+1)$.

We then set starting values of $A_{ij} = 0$ and $B_i = 0$, and assign values according to the following for $k = 1, \ldots, n$:

1. $A_{ij} \leftarrow A_{ij} + L(k-1) \int_{t_{k-1}}^{t_k} \bar{F}(t - t_{k-1}) \mathbf{1}_i \mathbf{1}_j' f(t_{k-1} - t)\, dt\, R(k+1)$
2. $B_i \leftarrow B_i + L(k) \mathbf{1}_i \mathbf{1}_j' R(k+1)$

where f is the transition probability defined in Section 2.4.2.1 and

$$\bar{F}(t) = \exp\{(Q - I_\lambda) t\}. \tag{2.56}$$

Because we have $\mathcal{L}(\phi; y_1, \ldots, y_n) = L(n)\mathbf{1} = \pi^0 R(1)$, the new estimators for our transition rates and state spaces are

$$\hat{q}_{ij} = q_{ij}^0 \frac{A_{ij}}{A_{ii}} \text{ and } \hat{\lambda}_i = \frac{B_i}{A_{ii}}, \tag{2.57}$$

where $i = 1, \ldots, r$. AFter these estimates are obtained, the EM algorithm is repeated using the updated parameter estimates until convergence.

2.4.6 Computational simplifications of the modified EM algorithm

Similar to the case with the Nelder-Mead downhill simplex algorithm in Section 2.4.4.3, a difficulty arises in implementing the EM algorithm presented in the previous section due to the values of $A_{ij}$and B_i being either very large or very small. Again, in extreme cases, this could be outside the default floating-point implementations in standard software and thus, customised floating-point software is required. To circumvent this issue, Roberts et al. (2006) proposes a scaling procedure for the forwards-backwards recursions. Another complication of the EM algorithm presented by Rydén (1996) is that one of the parameters involves the integration of matrices, which is not straightforward to implement in MATLAB. Roberts et al. (2006) borrows a result from Van Loan (1977) and uses a matrix exponential approach, which is well defined in this case and is readily available in standard MATLAB. The algorithm is discussed in detail in Section 3.1.1.1.

2.5 Operational Time Scaling

One of the main contributions of this thesis is the application of operational time scaling to Markov-modulated Poisson processes within the marked Poisson model in order to solve some of the issues involved with the practical application of the model to real world data. One of the difficulties presented by the Markov-modulated Poisson process is that the calibration methods involve the use of claim inter-arrival times rather than the frequency of claims. Further, in order to accurately calibrate our MMPP parameters, we require a homogeneous Poisson process over time, as we assume the process is stationary over time. It is difficult to directly add a multiplicative factor to the intensities as this will change the fundamental results provided in Section 2.4.2.2. As a result, we scale the claim inter-arrival times instead. In the following section, we define the theoretical framework behind the time scaling of a Poisson process. We also provide a brief example at the end to demonstrate the scaling operation.

2.5.1 Operational time in Poisson processes

The use of operational time is a common approach in order to transform a Poisson process with varying intensity to a homogeneous Poisson process, where the intensities are independent of time (i.e. the intensity function $\lambda(t)$ can be written as λ. We use the definition from Bühlmann (1970):

Assume that there is a claim number process N_t and a monotonically increasing function $\rho(t)$ with $\rho(0) = 0$. We define a new process M_τ where

$$M_\tau = N_{\rho^{-1}(\tau)} \tag{2.58}$$

and

$$\rho^{-1}(\tau) = \inf_t \{t; \rho(t) = \tau\}. \tag{2.59}$$

From equations 2.58 and 2.59, it holds that $\tau_i = \rho(t_i)$ for any time points which we specified by the index i.

2.5.2 Insurance claim example

We provide a simple example here to demonstrate how we intend to use the operational time scaling in the Markov-modulated Poisson process. Assume that in period 1, we expect claims to arrive at a certain intensity λ while in period 2, we expect a claim intensity of 2λ. We want to homogenise the claim intensities over time. Thus, we want to

stretch out the time in the second period by a factor of two, so that the claim intensity over the entire time period is now the same as in period 1. Thus, we multiply the inter-arrival claims in the second period by two, with a minor adjustment to the inter-arrival time corresponding to the time between the last claim in period 1 and the first claim in period 2.

CHAPTER 3

MODEL AND METHODOLOGY

This thesis implements a Markov-modulated Marked Poisson process using micro-level data in order to model the arrival of insurance claims. In the following chapter, we will introduce the model and explicitly set out the scaled EM algorithm by Roberts et al. (2006) that is used to calibrate the parameters of the MMPP.

We start by providing the basic framework for the model. Using the notation in Chapter 2, the claims process distribution can be formally written as

$$\{(T, Z_i)\}_{i=1,\ldots,N} \sim \text{Poisson}(\lambda(t), P_{Z|t}; t \geq 0). \tag{3.1}$$

The intensity $\lambda(t)$ will be dictated by a hidden Markov chain. Using the results from Section 2.3, we essentially split the claims process into four distinct components:

1. Claim Occurrences
2. Reporting Delay
3. Partial Payment Occurrences
4. Partial Payment Severity

Our key area of interest in the micro-level model here is the arrival of the claims. In that respect, only the first two components are relevant. However, we will briefly discuss in Section 6.2 how the latter two components may be appropriately fitted. By doing so, the research methods in this thesis can be extended to model the outstanding claim amounts.

3.1 Claim Occurrences

The model assumes that the claim arrivals are determined by a Markov modulated Poisson process. We investigate multiple Markov chains of different orders to determine the best fit for the data set. Following the notation from Section 2.4.2, the two parameter matrices that must be estimated are the generator matrix Q and the jump intensity matrix Λ. Once these matrices have been determined, the Markov modulated Poisson process is fully specified and simulations of the Markov process and claim arrivals can be done using standard Markov chain and probability theory.

3.1.1 Parameter Estimation in the unobservable Markov chain

Within the Markov chain component of our model, we need to estimate the intensities λ_i corresponding to the state spaces as well as the transition parameters q_{ij}. We begin with a Markov chain of order 2, which is a Switched Poisson Process as defined in Section 2.4.3 but extensions to higher orders are also explored.

The parameter calibration is done using a maximum likelihood estimation approach due to the advantages discussed in Section 2.4.4 such as asymptotic unbiasedness. More specifically, we use the modified expectation maximisation algorithm by Rydén (1996) to calculate the parameter estimates. However, we adopt the scaling procedure described in Roberts et al. (2006) in order to prevent numerical underflow that is a prominent issue in the modified EM algorithm detailed in Section 2.4.5. This procedure removes the need for customised floating point software suggested by Rydén (1996), making the model much more accessible to practitioners. The algorithm has a second advantage in that it simplifies the calculations within the EM algorithm and reduces the computational time required. We outline the steps of the algorithm in the following sections.

3.1.1.1 The modified EM algorithm

From Section 2.4.5, we have that the likelihood of the parameters (Q, I_λ) is

$$\mathcal{L}^c = \pi_{s_1} \left\{ \prod_{k=1}^{m} \left(q_{s_k} \exp\left(-q_{s_k} \Delta u_k\right) \times \frac{q_{s_k, s_{k+1}}}{q_{s_k}} \right\} \exp\left(-q_{s_{m+1}} \Delta u_{m+1}\right) \times \left\{ \prod_{k=1}^{m+1} \left(\frac{\left(\lambda_{s_k} \Delta u_k\right)^{z_k}}{z_k!} \exp\left(-\lambda_{s_k} \Delta u_k\right) \times \frac{z_k!}{\left(\Delta u_k\right)^{z_k}} \right\} \right. \tag{3.2}$$

The intermediate expressions that we need to determine for the E-step of the EM algorithm are

$$\hat{m}_{ij} = \int_0^{t_n} \mathbb{P}_{\left(Q^0, I_\lambda^0\right)} \left[X\left(t-\right) = i, X\left(t\right) = j \middle| N(u), 0 \leq u \leq t_n\right] dt, \tag{3.3}$$

$$\hat{T}_i = \int_0^{t_n} \mathbb{P}_{\left(Q^0, I_\lambda^0\right)} \left[X\left(t\right) = i | N(u), 0 \leq u \leq t_n\right] dt, \text{ and} \tag{3.4}$$

$$\hat{n}_i = \sum_{k=1}^{n} \mathbb{P}_{\left(Q^0, I_\lambda^0\right)} \left[X\left(t_k\right) = i | N(u), 0 \leq u \leq t_n\right]. \tag{3.5}$$

3.1.1.2 Rescaled recursion equations

Firstly, the likelihood function of the sample path $\{N\left(u\right), 0 \leq u \leq T\}$ can be written in terms of the likelihood function of the sequence $Y^n = \{Y_1, \ldots, Y_n\}$

$$p\left(y^n\right) = \boldsymbol{\pi} \prod_{k=1}^{n} f\left(y_k\right) \mathbf{1}, \tag{3.6}$$

where y^n is a realisation of Y^n. We can also rewrite

$$p\left(y^n\right) = \prod_{i=1}^{n} p\left(y_i | y^{i-1}\right) \tag{3.7}$$

where $p\left(y_1 | y_0\right) = p\left(y_1\right)$. Letting $c_k = p\left(y_k | y^{k-1}\right)$, the rescaling of the backwards and forwards equations from Section 2.4.5.2 can be condensed into matrix notation and is expressed as follows:

$$L\left(k\right) = \boldsymbol{\pi} \prod_{i=1}^{k} \frac{f\left(y_i\right)}{c_i}, \text{ for } k = 1, \ldots, n \tag{3.8}$$

and

$$R\left(k\right) = \prod_{i=k}^{n} \frac{f\left(y_i\right)}{c_i} \mathbf{1}, \text{ for } k = n, \ldots, 1, \tag{3.9}$$

with $L\left(0\right) = \boldsymbol{\pi}$ and $R\left(n+1\right) = \mathbf{1}$. From the definition of c_k, we have that

$$c_k = \pi \prod_{i=1}^{k-1} \frac{f\left(y_i\right)}{c_i} f\left(y_k\right) \mathbf{1} = L\left(k-1\right) f\left(y_k\right) \mathbf{1}, \tag{3.10}$$

so finally, we can rewrite the rescaled forwards/backwards equations as

$$L\left(k\right) = \frac{L\left(k-1\right) f\left(y_k\right)}{L\left(k-1\right) f\left(y_k\right) \mathbf{1}}, \text{ and } R\left(k\right) = \frac{f\left(y_k\right) R\left(k+1\right)}{L\left(k-1\right) f\left(y_k\right) \mathbf{1}}. \tag{3.11}$$

Roberts et al. (2006) proves that the elements of the $L(k)$ and $R(k)$ are upper bounded by 1 and $\frac{1}{L_j(k-1)}$ respectively, where $L_j(k-1)$ is the corresponding element in the vector $L(k-1)$ for the jth element $R_j(k-1)$.

3.1.1.3 New parameter estimators using rescaled equations

Define a new operation $\bullet$ so that for two appropriately sized matrices A and B, $A \bullet B$ is the matrix consisting of the term by term multiplication of the two component matrices. We can now determine the expressions for the preliminary expressions from Section 2.4.5.1:

The matrix $\hat{m}$ of the $\{\hat{m}_{ij}\}$ terms is

$$\hat{m} = \quad Q' \bullet \sum_{k=1}^{n} \left(\frac{1}{c_k} \int_{t_{k-1}}^{t_k^-} f(t_k - t)\, R(k+1)\, L(k-1)\, \bar{F}(t - t_{k-1})\, dt \right)'. \tag{3.12}$$

The $r \times 1$ vector of the $\hat{n}_i$ terms, $\hat{n}$, is

$$\hat{n} = \sum_{k=1}^{n} \left(L(k)' \bullet R(k+1) \right). \tag{3.13}$$

Using the expression we obtained for $\hat{m}_{ij}$, we can obtain $\hat{T}_i$ as

$$\hat{T}_i = \frac{\hat{m}_{ii}}{q_{ii}}, \tag{3.14}$$

where q_{ii} is obtained from the starting values for each step in the EM algorithm. The estimators for the elements in Q and I_λ can now be obtained using the same expressions given at the end of Section 2.4.5.1.

3.1.1.4 Simplified evaluation of $\hat{m}$

In the previous section, we note that the evaluation of the estimator for m requires the computation of an integral of matrices. For notational simplicity, we write

$$\mathcal{I}'_k = \int_0^{y_k} \exp\left((Q - \Lambda)(y_k - y)\right) \Lambda R(k+1)\, L(k-1) \exp\left((Q - \Lambda)\, y\right) dy, \tag{3.15}$$

where $y_k = t_k - t_{k-1}$. The estimator $\hat{m}$ can then be rewritten as

$$\hat{m} = Q \bullet \sum_{k=1}^{n} \left(\frac{\mathcal{I}'_k}{c_k} \right). \tag{3.16}$$

In order to more easily evaluate $\mathcal{I}_k$, Van Loan (1977) proposes the use of the block triangular matrix

$$C_k = \begin{bmatrix} Q - \Lambda & \Lambda R\left(k+1\right) L\left(k-1\right) \\ 0 & Q - \Lambda \end{bmatrix} \Bigg(\tag{3.17}$$

By equating the terms in the equation

$$\frac{d}{dy} \exp\left(C_k y\right) = C_k \exp\left(C_k y\right), \tag{3.18}$$

it can be shown that $\mathcal{I}_k$ is simply the top right block of the matrix $e^{C_k y_k}$, which can be easily evaluated using the "expm" function in MATLAB, whic uses Padé's approximation with repeated squaring in order to calculate the matrix exponential, which is the suggested method in Van Loan (1977).

3.1.2 Accounting for discretised claim arrival records

In practice, the exact arrival time of a claim is not available. Instead, insurers generally only record the day that the claim arrives. However, for the purposes of parameter estimation using the modified EM algorithm, we require the exact inter-arrival times of claims. One natural approach to this issue is to assume that the claims arrive at equal intervals throughout the day. An alternative approach is to simply simulate the arrival times of the claims throughout the day using a uniform distribution. However, this approach introduces more randomness into the model and we show using simulations in Section 4.3 that the results for the latter case are on average equal to the case of equal inter-arrival times. Thus, we choose the first method to solve this issue as we do not want to introduce artificial randomness into our model calibration. This also simplifies our calculations as the inter-arrival times cannot extend over multiple periods. We discuss the impact of this choice on the final estimated parameters in Section 4.3 and empirically show that the final results are similar.

3.1.3 Accounting for heterogeneous intensities in the Markov-modulated Poisson process

There are multiple factors that can occur in practice that will result in the claim arrival frequency changing over time. In our model, we investigate the following factors that would naturally alter the intensity:

1. Reporting Delay
2. Exposure
3. Seasonality
   (a) Day of week
   (b) Weekend
   (c) Monthly
   (d) Quarterly
4. Residual Trend

We discuss the operational time adjustment that is applied for these factors in the following section. It is important to note that the time scaling procedure that we apply here would be able to include any heterogeneity and provides a great amount of flexibility in the modelling procedure. Due to the multiplicative nature of such adjustments, we can combine the impacts of multiple factors by simply computing a multiplicative aggregate of all the time scales in order to obtain a final scaling amount. In the following sections, we will express the inter-arrival adjustments for each of the components above as $\mathcal{A}_{RD}, \mathcal{A}_E, \mathcal{A}_S$ and $\mathcal{A}_T$ respectively, so the final adjustment $\mathcal{A}$ will be

$$\mathcal{A} = \mathcal{A}_{RD} \times \mathcal{A}_E \times \mathcal{A}_S \times \mathcal{A}_T. \tag{3.19}$$

3.1.3.1 Reporting delay

We first fit a distribution to the data by adopting a maximum likelihood approach to estimate parameters. Depending on the line of business, it may be appropriate to simply fit standard probability distributions such as the exponential, log-normal and Pareto distributions. However, as an alternative, Antonio and Plat (2013) demonstrates that it may also be necessary to have piecewise distributions. The authors fits a Weibull distribution mixed with nine degenerate components representing settlement days 0 to 8. An advantage of using this choice is that the flexibility of the Weibull distribution combined with the discrete empirical components allow the model to adequately capture both short

and long tailed characteristics of the data. Another strategy adopted by Jin (2013) is to separate the reporting delays into bins and use the empirical distribution function for the distribution of the delays. These approaches can be compared using the goodness-of-fit tests such as the Akaike Information Criterion and the Bayesian Information Criterion in order to choose the most appropriate distribution.

After a suitable distribution is chosen and calibrated, we scale the frequencies/inter-arrival times in order to include the unknown claims. An assumption is made here that the reporting delay distribution that we calibrate is the true reporting delay distribution and that all claims follow this distribution. Denote the cumulative density at time t as $F(t)$. Noting that the claim arrival records in practice are usually discretised, we set a unit of t to be equal to one discrete period in length. Then the proportion of reported claims relative to the total number of claims that occurred in the period is equal to $F(t) - F(t-1)$. In order to homogenise the claims arrival intensity, we scale the frequency by a multiplicative factor of $\frac{1}{F(t)-F(t-1)}$ which ultimately leads to an adjustment factor $\mathcal{A}_{RD}(t)$ for the claim inter-arrival times of

$$\mathcal{A}_{RD}(t) = F(t) - F(t-1) \tag{3.20}$$

We note that we make an approximation here in that claims that arrive in the same day will have the same adjustment factor despite technically arriving at different times. However, given that insurers will generally have micro-level daily data, this is approximation will not significantly impact the results of the calibration.

3.1.3.2 Exposure

Insurers can create a deterministic function $\epsilon(t)$ of the exposure over the investigation period, where $t = 1, 2, \ldots, T$. We set our reference point at the end of the investigation period. We would expect that if we had a constant rate of claims per policyholder, then we should multiplicatively adjust the inter-arrival times at time t by the multiplicative factor

$$\mathcal{A}_E = \frac{\epsilon(t)}{\epsilon(T)}. \tag{3.21}$$

The reasoning for this may be easier to see if we look at the claim intensity from a frequency perspective and then reciprocal the final results to get the final inter-arrival time adjustment. At day t, we scale the claim frequency by first dividing by the exposure to get the average claims per policyholder. We then multiply by the exposure on the final day to obtain the claims frequency scaled to the final day. By taking the reciprocal of this expression, we obtain the adjustment in 3.21 above.

3.1.3.3 Seasonality

We describe a general method for adjusting for seasonality here, as there are multiple period types that need to be investigated for seasonality trends. Denote total number of seasonal periods as p. We first set a reference period. For example, in the case of weekend seasonality, our reference period could be set to weekdays. We then calculate a vector $S = (s_1, \ldots, s_p)$ of the claim frequency in each seasonal period. If our reference period is set as the last seasonal period p, then we should adjust the inter-arrival times in season t by

$$\mathcal{A}_S = \frac{s_t}{s_p}, \tag{3.22}$$

using the same logic as in the case of exposure.

3.1.3.4 General trend

In practice, there will be many other factors other than exposure and seasonality that will have an impact on claim frequencies. Structural changes such as a widening of claim definitions or changes in policy cover can significantly change the number of claims. It is possible for insurers to collate this data from their own databases and filter out the most of the impact of such changes. However, other changes in other environmental factors over time such as weather severity may be difficult to account for, either due to the insurer not collecting the relevant information or the factor itself being difficult to observe or quantify. Instead of attempting to discern all the relevant causes of the change in claim intensity over time, we can simply fit a trendline to the data and adjust the inter-arrival times using the same operational time approach shown in the previous sections. We note that we should apply the adjustments for the known factors first so that the trendline here can capture any extra trends on top of the ones already investigated.

For simplicity, let us use the example of a linear trendline:

$$\text{Claim Frequency} = \alpha + \beta \times \text{Date}. \tag{3.23}$$

For consistency, we will keep the reference point as the last day of the investigation period. We specify our Date variable as the number of days until the last day (so that on the last day, Date = 0). Using least squares, we can obtain estimates for the coefficients α and β.

At any given day, the value of the trendline will equal to $\alpha + \beta \times$ Date. whereas the value on the last day will simply be α. Thus, for each day, the adjustment to inter-arrival

times in order to homogenise the claim intensity would be

$$\mathcal{A}_T = \frac{\alpha + \beta \times \text{Date}}{\alpha}. \tag{3.24}$$

Again, the reasoning behind this is clearly to see if we look at the adjustment from a frequency perspective. In order to homogenise the frequency relative to the last day in the investigation period where the trendline equation 3.24 has a value of α, we must first divide the frequency by $\alpha + \beta \times \text{Date}$. We then multiply the frequency by the value at which the claims data is meant to be homogeneous: α. Finally, by taking the reciprocal, we have the inter-arrival time adjustment given by Equation 3.24.

3.2 Model Output and Results

After both the Markov-modulated Poisson process parameters have been calibrated, we can use simulation procedures in order to determine the total number of outstanding claims in our data. We firstly need to determine the appropriate regimes at each time in the investigation period. The total number of claim arrivals can then be simulated using a Poisson distribution. This procedure is explicitly set out in the following sections.

3.2.1 Determining the regimes of the claims arrival process

Firstly, it is shown in Roberts et al. (2006) that the ith element of the vector L can be expressed using the notation in Section 3.1.1.2 as

$$L(k)_i = \left(\frac{L\,(k-1)\,f\,(y_k)}{L\,(k-1)\,f\,(y_k)\,\mathbf{1}} \right)_i = \Pr\left(X(t_k) = i | y^k\right) \tag{3.25}$$

where $X(t_k)$ is the regime of the Markov-modulated Poisson process at the time of the kth claim arrival. The latter expression is the posterior probability of the process being in state i at the time of the kth claim arrival t_k, given the claim inter-arrival times up until t_k. We assume that the regime of the process does not change between claim arrivals and thus, we can obtain the maximum posterior estimate of the regime at each point in time.

3.2.2 Simulating the number of claims in each inter-arrival period

We have that the total number of claims in each inter-arrival period is a conditional Poisson distribution, dependent on the state of the Markov-modulated Poisson process at the time. More formally, for the kth inter-arrival time y_k, we can write the claims counting process N for this period as

$$N(y_k) \sim \text{Poisson}\left(\lambda_k | X(t_k)\right), \tag{3.26}$$

where $X(t_k)$ is the state of the MMPP at the time of the k claim arrival and λ_k is the claim intensity corresponding to that state.

Thus, we can simulate the number of claims in each of these states by using the above conditional Poisson distribution as we have the most likely regimes at each claim inter-arrival period. We also note that we would need to simulate the claims under the new operational time and it is extremely convenient that we have the scaled inter-arrival times here. This allows us to simply use the relevant calibrated frequency and multiply by our adjusted inter-arrival times to obtain the final claim numbers.

Another advantage of this model is that the final claims amounts are also at the micro-level so that we retain the granularity of the original data. However, if required, we can easily aggregate the claim numbers into suitable periods for analysis.

3.2.3 Final results

From the above procedures, we determine the total number of claims that have occurred in the investigation period. When implementing the model in practice, this should be the period over which we want to estimate the outstanding claim numbers, which will be the claims that have been incurred but not reported (IBNRs). This value will then be the difference between the total number of claims that we have simulated and the observed number of claims. We note that our algorithm is stochastic so we can repeat the entire procedure multiple times to estimate the variance of the central estimate that we obtain (or even higher moments if necessary), which is an advantage of this model over the standard deterministic models described in Section 2.1.

CHAPTER 4

ANALYSIS OF THE PARAMETER ERROR OF THE EM ALGORITHM

The calibration algorithm presented in Section 2.4.6 is an attractive choice as it is computationally cheap (relative to the other estimation methods in the current Markov-modulated Poisson process literature) and is easily implementable in standard software such as MATLAB and R. However, we are not aware of any literature that investigates the accuracy and precision of this algorithm. Indeed, the paper in which this method is proposed (Roberts et al. (2006)) only investigates the improvements in terms of computational time and does not discuss the accuracy of the algorithm outputs. Thus, we fill this gap in the literature by conducting a simulation study in the following sections in order to justify the use of this calibration method in our model.

Further, we also make assumptions during the implementation of the algorithm in order to resolve the issue of the discretised claim arrival data. We show in Section 4.3 that both of the approaches we use provide similar results to demonstrate that the assumption choice does not impact the final model output.

4.1 Sensitivity of the modified EM algorithm

In the next two sections, we will test the sensitivity of the final estimations to the two manual inputs of the calibration algorithm:

1. Initial parameter estimates
2. Convergence tolerance

For simplicity (and given the final model choice in our real world case study), we will only conduct tests with MMPPs of order 2. However, we expect to achieve that the conclusions will also extend to higher orders, as the algorithm does not change significantly between different orders of MMPPs.

For the following sections, we use the following parameters for the simulated order 2 MMPP:

$$Q = \begin{bmatrix} -0.1 & 0.1 \\ 0.01 & -0.01 \end{bmatrix}, \quad \lambda = (10, 1) \tag{4.1}$$

Also, the time frame used to simulate the data is 1000 periods. As this is not a value that is adjustable by an actuary in practice, we do not test the impact of the chosen time frame on the final estimated results. However, we intuitively expect that a longer time frame will generate more observations, increasing the accuracy of the calibration.

4.1.1 Sensitivity to initial parameters

Roberts et al. (2006) notes that in the unscaled EM algorithm by Rydén (1996), the initial parameter estimates do not significantly impact the fit of the algorithm. This is seen in the log-likelihood function

$$\mathcal{L} = \log(\boldsymbol{\pi}(\boldsymbol{Q}, \boldsymbol{\lambda})) + \sum_{i=1}^{n} \log f(y_k) \tag{4.2}$$

that the initial parameter estimates $\boldsymbol{\pi}(\boldsymbol{Q}, \boldsymbol{\lambda})$ has negligible effects on the overall log-likelihood value, given a sufficiently large sample size.

The three manual starting parameters that are input into the algorithm are

1. A starting estimate for Q
2. A starting estimate for λ
3. A starting estimate for the stationary distribution of regimes a

The tolerance level for the minimum increase in the likelihood function is set to 10^{-5} and we can see from Section 4.1.2 that this should be more than sufficient. For comparison, the tolerance level used by Rydén (1996) is 10^{-1}.

4.1.1.1 Varying Q

Table 4.1 below shows the effects of changing the magnitude of the initial estimate of the generator matrix Q.

Initial a	**Initial $\boldsymbol{Q}$**	Initial λ	Estimated Q	Estimated λ
$(0.5, 0.5)$	$\begin{pmatrix} -0.05 & 0.05 \\ 0.005 & -0.005 \end{pmatrix}$	$(10, 1)$	$\begin{pmatrix} -0.0922 & 0.0922 \\ 0.0067 & -0.0067 \end{pmatrix}$	$(9.9060, 1.0093)$
$(0.5, 0.5)$	$\begin{pmatrix} -0.1 & 0.1 \\ 0.01 & -0.01 \end{pmatrix}$	$(10, 1)$	$\begin{pmatrix} -0.0922 & 0.0922 \\ 0.0067 & -0.0067 \end{pmatrix}$	$(9.9060, 1.0093)$
$(0.5, 0.5)$	$\begin{pmatrix} -0.5 & 0.5 \\ 0.05 & -0.05 \end{pmatrix}$	$(10, 1)$	$\begin{pmatrix} -0.0922 & 0.0922 \\ 0.0067 & -0.0067 \end{pmatrix}$	$(9.9060, 1.0093)$
(0.5,0.5)	$\begin{pmatrix} -1 & 1 \\ 0.1 & -0.1 \end{pmatrix}$	$(10, 1)$	$\begin{pmatrix} -0.0922 & 0.0922 \\ 0.0067 & -0.0067 \end{pmatrix}$	$(9.9060, 1.0093)$

Table 4.1: Parameter estimation when varying the magnitude of Q

Clearly, the magnitude of the initial estimates of Q do not significantly affect the final estimates from the EM algorithm. However, we will also investigate initial estimates of Q where the transition rates are in different ratios. The results are shown in Table 4.2 below.

Initial a	**Initial $\boldsymbol{Q}$**	Initial λ	Estimated Q	Estimated λ
$(0.5, 0.5)$	$\begin{pmatrix} -0.1 & 0.1 \\ 0.1 & -0.1 \end{pmatrix}$	$(10, 1)$	$\begin{pmatrix} -0.0922 & 0.0922 \\ 0.0067 & -0.0067 \end{pmatrix}$	$(9.9060, 1.0093)$
$(0.5, 0.5)$	$\begin{pmatrix} -0.01 & 0.01 \\ 0.1 & -0.1 \end{pmatrix}$	$(10, 1)$	$\begin{pmatrix} -0.0922 & 0.0922 \\ 0.0067 & -0.0067 \end{pmatrix}$	$(9.9060, 1.0093)$
$(0.5, 0.5)$	$\begin{pmatrix} -0.45 & 0.45 \\ 0.23 & -0.23 \end{pmatrix}$	$(10, 1)$	$\begin{pmatrix} -0.0922 & 0.0922 \\ 0.0067 & -0.0067 \end{pmatrix}$	$(9.9060, 1.0093)$
(0.5,0.5)	$\begin{pmatrix} -10 & 10 \\ 10 & -10 \end{pmatrix}$	$(10, 1)$	$\begin{pmatrix} -0.0922 & 0.0922 \\ 0.0067 & -0.0067 \end{pmatrix}$	$(9.9060, 1.0093)$

Table 4.2: Parameter estimation when varying the ratio of the transitions within Q

Again, it seems like the algorithm is very robust to variation in the initial parameter estimates of Q. Even in the last case of the table where the initial values are very far from the true transition rates, the algorithm converges to the same solution.

Thus, we conclude that the algorithm is very robust with respect to changes to the initial parameter estimate of Q.

4.1.1.2 Varying λ

We conduct the same series of tests as in Section 4.1.1.1. Table 4.3 varies the magnitude of the claim intensities λ while maintaining a ratio of 10. We can see that the algorithm performs consistently for the smaller claim intensity initial estimates. However, once the initial parameters become very large, the algorithm performs very poorly and produces nonsensical results. For the purposes of our case study, our final estimates are moderate and fall between the first and second example.

Initial a	Initial Q	**Initial $\boldsymbol{\lambda}$**	Estimated Q	Estimated λ
$(0.5, 0.5)$	$\begin{pmatrix} -0.1 & 0.1 \\ 0.01 & -0.01 \end{pmatrix}$	$(1, 0.1)$	$\begin{pmatrix} -0.0922 & 0.0922 \\ 0.0067 & -0.0067 \end{pmatrix}$	$(9.9060, 1.0093)$
$(0.5, 0.5)$	$\begin{pmatrix} -0.1 & 0.1 \\ 0.01 & -0.01 \end{pmatrix}$	$(10, 1)$	$\begin{pmatrix} -0.0922 & 0.0922 \\ 0.0067 & -0.0067 \end{pmatrix}$	$(9.9060, 1.0093)$
$(0.5, 0.5)$	$\begin{pmatrix} -0.1 & 0.1 \\ 0.01 & -0.01 \end{pmatrix}$	$(100, 10)$	$\begin{pmatrix} -2.638 & 2.638 \\ 4.02 \times 10^{-7} & -4.02 \times 10^{-7} \end{pmatrix}$	$(1.82 \times 10^{-80}, 1.6146)$

Table 4.3: Parameter estimation when varying the magnitude of the intensities λ

Further, Table 4.4 below shows the calibration estimates when we use different ratios between the claim intensities. We can see that we obtain exactly the same results, given a reasonable initial parameter estimate (although in the third case, we get a permuted version of the true parameters as the EM algorithm does not preserve the structure of the MMPP, as demonstrated in Guillou et al. (2015)). Thus, we have confidence from the results above that the calibration was not strongly influenced by our initial choice of lambda.

Initial a	Initial Q	**Initial $\boldsymbol{\lambda}$**	Estimated Q	Estimated λ
$(0.5, 0.5)$	$\begin{pmatrix} -0.1 & 0.1 \\ 0.01 & -0.01 \end{pmatrix}$	$(5, 1)$	$\begin{pmatrix} -0.0922 & 0.0922 \\ 0.0067 & -0.0067 \end{pmatrix}$	$(9.9060, 1.0093)$
$(0.5, 0.5)$	$\begin{pmatrix} -0.1 & 0.1 \\ 0.01 & -0.01 \end{pmatrix}$	$(10, 1)$	$\begin{pmatrix} -0.0922 & 0.0922 \\ 0.0067 & -0.0067 \end{pmatrix}$	$(9.9060, 1.0093)$
$(0.5, 0.5)$	$\begin{pmatrix} -0.1 & 0.1 \\ 0.01 & -0.01 \end{pmatrix}$	$(1, 10)$	$\begin{pmatrix} -0.0067 & 0.0067 \\ 0.0922 & -0.0922 \end{pmatrix}$	$(1.0093, 9.9060)$

Table 4.4: Varying the ratio of the claim intensities

4.1.1.3 Varying a

As in the previous sections, Table 4.5 shows the variation in the final estimated values for different initial values of a.

Initial $\boldsymbol{a}$	Initial Q	Initial λ	Estimated Q	Estimated λ
$(0.25, 0.75)$	$\begin{bmatrix} -0.1 & 0.1 \\ 0.01 & -0.01 \end{bmatrix}$	$(10, 1)$	$\begin{bmatrix} -0.0922 & 0.0922 \\ 0.0067 & -0.0067 \end{bmatrix}$	$(9.9060, 1.0093)$
$(0.5, 0.5)$	$\begin{bmatrix} -0.1 & 0.1 \\ 0.01 & -0.01 \end{bmatrix}$	$(10, 1)$	$\begin{bmatrix} -0.0922 & 0.0922 \\ 0.0067 & -0.0067 \end{bmatrix}$	$(9.9060, 1.0093)$
$(0.75, 0.25)$	$\begin{bmatrix} -0.1 & 0.1 \\ 0.01 & -0.01 \end{bmatrix}$	$(10, 1)$	$\begin{bmatrix} -0.0067 & 0.0067 \\ 0.0924 & -0.0924 \end{bmatrix}$	$(1.0093, 9.9060)$

Table 4.5: Parameter Varying the starting distribution

In this case, the robustness of the algorithm is very strong, as the 3 different starting distribution estimates all produce extremely similar results.

Thus, looking at all 3 sections, we conclude that the EM algorithm modified by Roberts et al. (2006) is very robust to the starting parameters when these initial parameters are within a reasonable range.

4.1.2 Sensitivity to convergence tolerance

An attractive property of the EM algorithm is that iterative steps of the algorithm will not decrease the value of the likelihood function. When applying this function in practice, we set a tolerance level for the increase of the log-likelihood function with each step so that when this increase falls below the tolerance, we stop the procedure. Table 4.6 below shows the impact of changing the tolerance level on the final estimated parameters.

Tolerance	Estimated Q	Estimated λ
10^{-1}	$\begin{bmatrix} -0.0924 & 0.0924 \\ 0.0067 & -0.0067 \end{bmatrix}$	$(9.9059, 1.0093)$
10^{-3}	$\begin{bmatrix} -0.0924 & 0.0924 \\ 0.0067 & -0.0067 \end{bmatrix}$	$(9.9059, 1.0093)$
10^{-5}	$\begin{bmatrix} -0.0922 & 0.0922 \\ 0.0067 & -0.0067 \end{bmatrix}$	$(9.9060, 1.0093)$
10^{-10}	$\begin{bmatrix} -0.0922 & 0.0922 \\ 0.0067 & -0.0067 \end{bmatrix}$	$(9.9060, 1.0093)$

Table 4.6: Varying the tolerance of the log-likelihood increase

4.2 Testing the accuracy of the EM algorithm using simulated data

In the following section, we investigate the general accuracy of the algorithm by Roberts et al. (2006) over different Q and λ values. From the results before, we can see that changing the tolerance value and the starting distirbution a does not significantly impact the final results so we set the tolerance to 10^{-5} and a to $(0.5, 0.5)$.

For each set of given 'true' parameters, we simulate the claim arrivals according to the Markov-modulated Poisson process that is determined by the parameter values. We then apply the EM algorithm to calibrate the parameters and record the results. This entire procedure is repeated three times for each set of parameters and the calibrated values are tabulated below in Table 4.7.

True Q	True λ	Estimated Q	Estimated λ
$\begin{bmatrix} -1 & 1 \\ 0.1 & -0.1 \end{bmatrix}$	$(10, 1)$	$\begin{bmatrix} -0.7556 & 0.7556 \\ 0.0836 & -0.0836 \end{bmatrix}$	$(9.5153, 0.9886)$
		$\begin{bmatrix} -0.7887 & 0.7887 \\ 0.0876 & -0.0876 \end{bmatrix}$	$(9.6234, 1.0127)$
		$\begin{bmatrix} -0.7632 & 0.7632 \\ 0.0786 & -0.0786 \end{bmatrix}$	$(9.2279, 1.0205)$
$\begin{bmatrix} -0.75 & 0.75 \\ 0.4 & -0.4 \end{bmatrix}$	$(8, 1)$	$\begin{bmatrix} -0.5942 & 0.5942 \\ 0.3277 & -0.3277 \end{bmatrix}$	$(7.551, 1.0407)$
		$\begin{bmatrix} -0.6046 & 0.6046 \\ 0.3338 & -0.3338 \end{bmatrix}$	$(7.3983, 1.0840)$
		$\begin{bmatrix} -0.6074 & 0.6074 \\ 0.3397 & -0.3397 \end{bmatrix}$	$(7.3255, 1.0648)$
$\begin{bmatrix} -0.5 & 0.5 \\ 0.4 & -0.4 \end{bmatrix}$	$(6, 2)$	$\begin{bmatrix} -0.3718 & 0.3718 \\ 0.3258 & -0.3258 \end{bmatrix}$	$(5.6363, 2.1011)$
		$\begin{bmatrix} -0.2895 & 0.2895 \\ 0.2577 & -0.2577 \end{bmatrix}$	$(5.7219, 2.1523)$
		$\begin{bmatrix} -0.3508 & 0.3508 \\ 0.3087 & -0.3087 \end{bmatrix}$	$(5.7397, 2.1641)$

Table 4.7: Testing the calibration performance of the EM algorithm using simulated data

We make several observations from the results above:

1. The estimates for the generator matrix Q are lower than the true values.
2. The ratios between the two transition rates in Q remains approximately the same.
3. The estimates for the claim intensities λ are reasonably accurate.

Thus, we conclude that the algorithm consistently underestimates the generator matrix while performing well on the jump intensities. This means that the algorithm underestimates the true number of regime switches in the model. However, the final central estimate of the algorithm will remain approximately the same as long as the second observation above holds true. This is shown in the following section.

4.2.1 On the consistent underestimation of Q

An important observation from the previous section is that while the values of Q may be underestimated, the ratio between the two transitions remains approximately the same. Consider the following Markov transition matrix:

$$\begin{bmatrix} -ar & ar \\ a & a \end{bmatrix} \Bigg(\tag{4.3}$$

Using standard Markov chain theory, the time spent in each state in the long term is equal to the stationary distribution, which can be expressed as

$$(\pi_1, \pi_2) = \left(\frac{1}{r+1}, \frac{r}{r+1} \right) \Bigg(\tag{4.4}$$

Note that this expression is independent of a, and instead depends on the ratio r between the transition rates. Thus, in our case, the stationary distribution of the MMPP is unaffected by the consistent underestimation in the generator matrix Q as the ratio between the transitions remains the same. We can then approximate the total number of claims using this stationary distribution and the claim intensities that we calibrate, which we noted in the previous section as reasonably accurate. As a result, we expect that the first moment of the MMPP (which leads to the central estimate of the outstand claims) will be unaffected by the underestimation.

4.3 Equally distributed vs simulated claim arrivals

An assumption needs to be made in order to solve the issue of discretised claims arrival data. There are two possible approaches here:

1. Assume that the claims arrive at equal intervals throughout the day
2. Assume that claims arrival times are uniformly distributed throughout the day

In the case of the second approach, we would simulate the exact claim arrival times using a uniform distribution before calculating the claim inter-arrival times. In the following section, we aim to show that the first method provides better results as there is less randomness that is introduced into the calibration and the assumption is more simplistic.

We first simulate the claims using an MMPP with the following parameters

$$Q = \left[\begin{pmatrix} -0.7 & 0.7 \\ 0.1 & -0.1 \end{pmatrix} \right. \left(\lambda = (7, 1) \right. \tag{4.5}$$

We then discretise our claim arrivals into days and then apply both of the methods listed above. This will provide the claim inter-arrival times that we use to calibrate our parameters.

4.3.1 Equal inter-arrival times

If we assume that claims arrive at equal intervals during the day, we obtain the following results. Note that there is no stochasticity in the calibration method as the claim inter-arrival times are the same and we use the same set of initial parameters for the EM algorithm. Thus, the final set of parameter estimates are

$$Q = \left[\begin{array}{cc} -0.5976 & 0.5976 \\ 0.1434 & -0.1434 \end{array} \right] \left(\lambda = (6.7538, 1.5648) \right. \tag{4.6}$$

4.3.2 Simulated inter-arrival times

We now simulate the inter-arrival times of the claims using a uniform distribution. We then obtain the inter-arrival times and calibrate our parameters using the EM algorithm. Table 4.8 shows the results of 3 simulations.

Q	λ
$\begin{bmatrix} -0.4925 & 0.4925 \\ 0.1211 & -0.1211 \end{bmatrix}$	$(6.8515, 1.2495)$
$\begin{bmatrix} -0.6879 & 0.6879 \\ 0.1067 & -0.1067 \end{bmatrix}$	(5.4935,2.0498)
$\begin{bmatrix} -0.6011 & 0.6011 \\ 0.1343 & -0.1343 \end{bmatrix}$	$(5.9789, 1.2648)$

Table 4.8: Calibrated parameters for the case of simulated claim inter-arrivals

4.3.3 Conclusion

As expected, we can see that the simulation of the exact arrival times of claims introduces significant randomness to the calibration and the parameters can vary by sizable amounts. However, on average, it seems that the simulated results are quite close to the results obtained by assuming equal inter-arrival times. Thus, we will use the latter assumption as both perform similarly in terms of accuracy but there is less parameter variability when assuming equally spaced claim arrivals. The simplistic approach brings additional benefits as the inter-arrival periods between claims can no longer go between one period and the next, which greatly simplifies the adjustments that are described in Section 3.1.3.

CHAPTER 5

CASE STUDY - PUBLIC LIABILITY INSURANCE

We use an empirical case study to demonstrate the benefits of our model in providing a realistic and accurate method to predict future claims. This study also demonstrates the accessibility of such a model can be implemented and because several components can be neatly separated into insurance-related factors, the methodology is also very tractable. For the purposes of this case study, we use the public liability data that comes from the AUSI data set which is briefly described below. Note that for the purposes of confidentiality, identifiable information such as axis scales have been removed.

5.1 The AUSI Data set

The Allianz, University of New South Wales, Suncorp and Insurance Australia Group (AUSI) dataset was developed as part of the Linkage Project: "Modelling claim dependencies for the general insurance industry with economic capital in view: an innovative approach with stochastic processes", which is supported by the Australian Research Council and several major industry partners, namely the Allianz Group, the Suncorp Group and the Insurance Australian Group. The dataset is comprised of micro-level general insurance claims data from 4 lines of business: Home, Private Motor, Compulsory Third Party and Public Liability. The data is in a standard format, consisting of

1. A policy file which describes the underwritten policy, containing information such as date of inception/expiry and sum insured.

2. A claim header file containing static claim information such as occurrence and notification dates, claim states and finalisation dates
3. A claim transaction file, containing information on each claim transaction such as transaction date, amount of transaction and claim status after the transaction.

For the purposes of the case study, we will be investigating the Public Liability line of business.

5.2 Preliminary data analysis

We begin with a preliminary examination of the data set. Some basic data cleaning was applied and claim records with clear errors were removed.

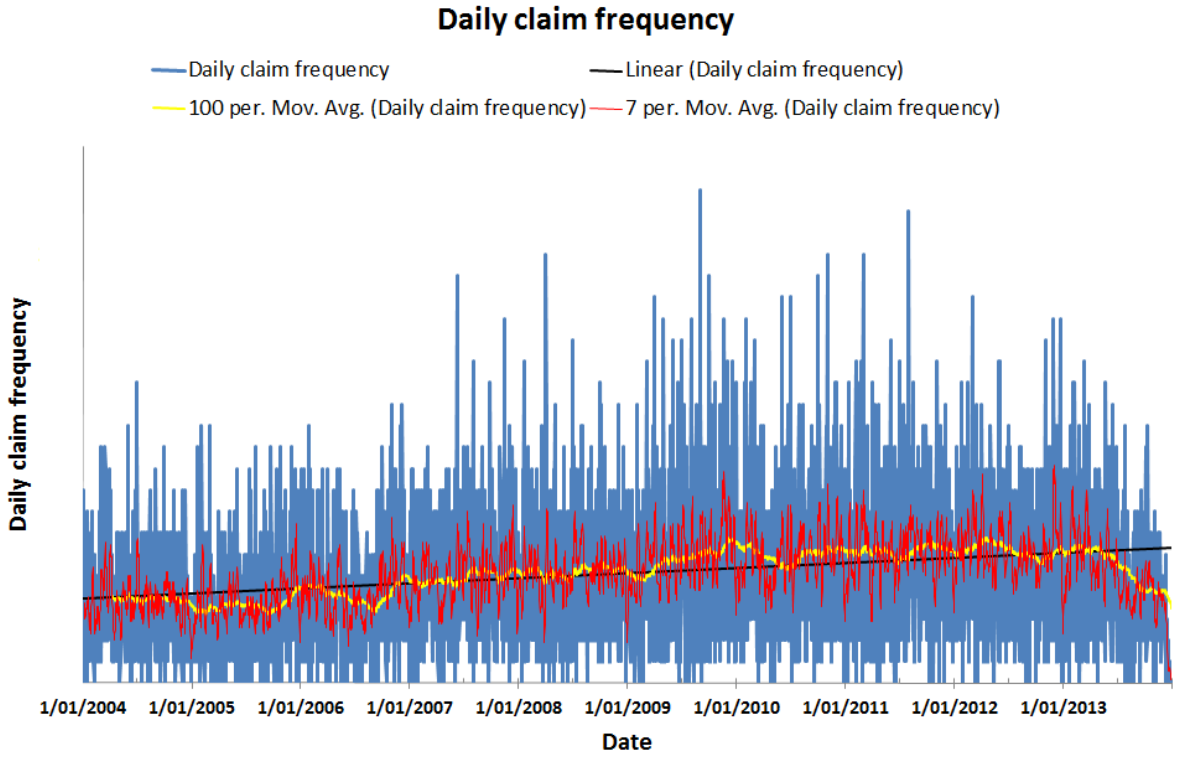

Figure 5.1: Daily Claim Frequency 1/1/2004 - 31/12/2013

Figure 5.1 above gives the daily claim frequency of claim events. As we can see, the claim frequencies tend to increase over time, apart from approximately the last one and a half years where there is a significant dip in the data. This is due to the reporting delay of the public liability line which can be quite long tailed, as discussed later in the chapter. Further, the claim intensity displays a strong upwards trend. We will investigate the source of this trend and use the operational time scaling procedure detailed in Section 2.5 to homogenise the first moment of the claim intensity so that the EM calibration

procedure can be carried out.

It is interesting to note that the second moment of the frequency also does not seem to change significantly over time, shown by the amplitude of the oscillating red moving average line in the figure above. We will also investigate the impact of our operational time scaling on the second moment, as we require stationarity in order to accurately calibrate our Markov-modulated Poisson process model.

Also, in the following sections, we plot the claim frequencies for ease of interpretation. However, the operational time adjustments that we apply will be the inverse of the frequency adjustments, as the frequency and inter-arrival times are inversely related.

We should also note that the adjustments that are applied should be analysed in a hierarchical manner that the effects of one factor do not overlap with the analysis of the effects of another. The most intuitively clear sequence to calculate these factors is the one provided in Section 3.1.3: Reporting Delay, exposure, seasonality and finally residual trend. However, we note that when we obtain the final adjustment factors, we can apply them in any order we prefer as the multiplicative adjustments are associative, as seen in (3.1.3).

5.3 Adjustment for reporting delay

In order to determine the appropriate adjustment for the reporting delay within the data, we first determine a suitable distribution to fit the reporting delay. Analysis of this delay n the data suggests that the main body of the distribution lies within the first year, although we have a very long (but very thin) tail that stretches out until slightly under 8 years. We fit the following probability distributions to the reporting delay data:

1. Pareto
2. Lognormal

Note that we also investigated other distributions such as the Weibull and exponential distributions but the fit was extremely poor and the inclusion of these fits in the following graphs distorted the axes significantly. This made the graphs quite difficult to read and thus, we remove these from the following analysis.

To smooth out the data slightly, we discretised the data into weeks and use the average times in each bin as the value of the bin. For example, the value of the bin for days 1 to 7 is 0.5.

The overall fit by the two probability distributions is shown in Figure 5.2 below.

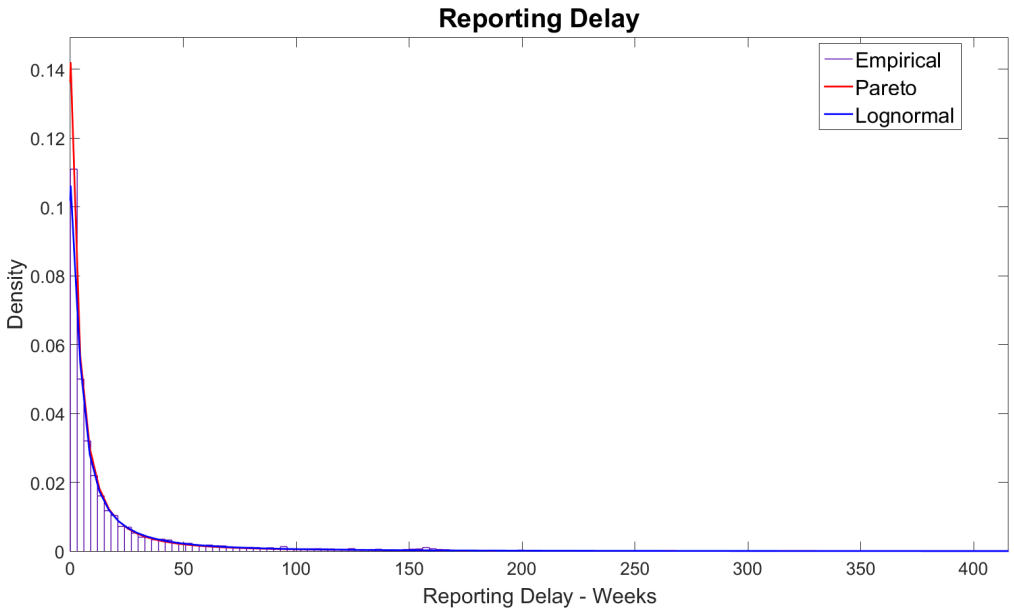

Figure 5.2: Reporting Delay fit

Looking at the empirical distribution, the tail of the distribution is indeed quite long but we can also see that the tail is very thin. Thus, we expect that most of the claims have a relatively short reporting delay (0-3 years) while there is a very small number of claims with very large delays. There is some structural influence here, as the statue of limitations is 3 or 6 years, dependent on the geographical state. There is also the consideration that the claims with extremely large reporting delays may be long-lasting court cases that may have a very high severity. It is reasonable to remove these from the analysis so that they can be modelled separately. However, given that the majority of the empirical distribution is captured within the first 3 years and the tail is very thin, we do not incorporate these impacts in our analysis.

In terms of the distributional fit to the data, we can see that the fit for the Pareto and log-normal distributions is generally quite good in the main body of the distribution. However, it seems that these distributions may overestimate the tail of the reporting delay data. We investigate both the body and the tail of the data separately in the following sections.

5.3.1 Reporting Delay - Body

Figure 5.3 below shows the main body of the reporting delay distribution which is the first 200 weeks. By observation, we see that the lognormal distribution initially underestimates the first bin of data while the Pareto distribution performs quite well. In the later weeks, we can see that the fit of the Pareto and log-normal distributions both perform reasonably.

We note that apart from the first bin, the Pareto and log-normal distributions produce very similar fits.

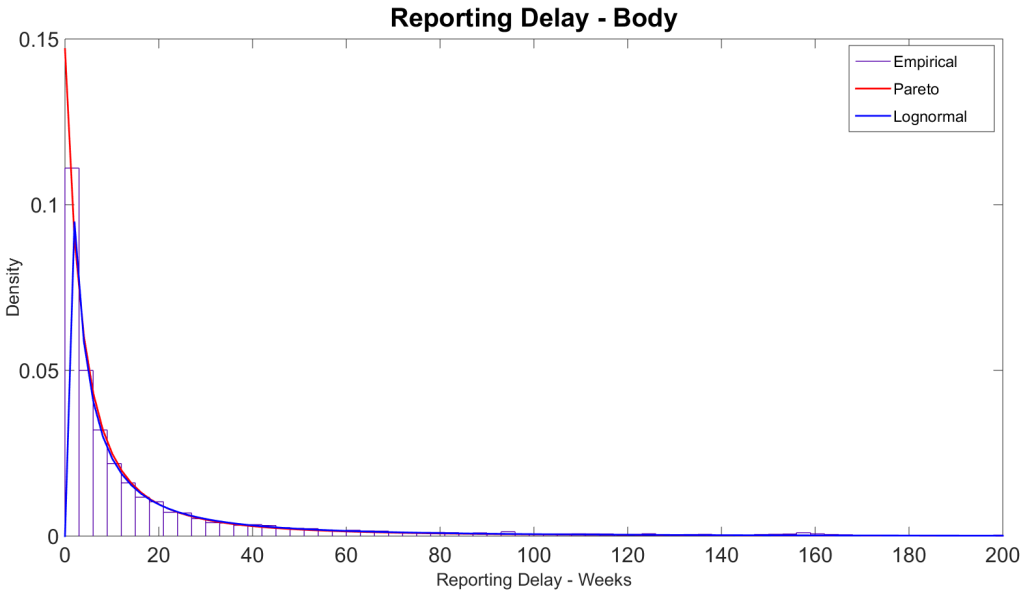

Figure 5.3: Reporting Delay fit - Body

5.3.2 Reporting Delay - Tail

Figure 5.4 below is the tail of the reporting delay distribution starting from 200 weeks.

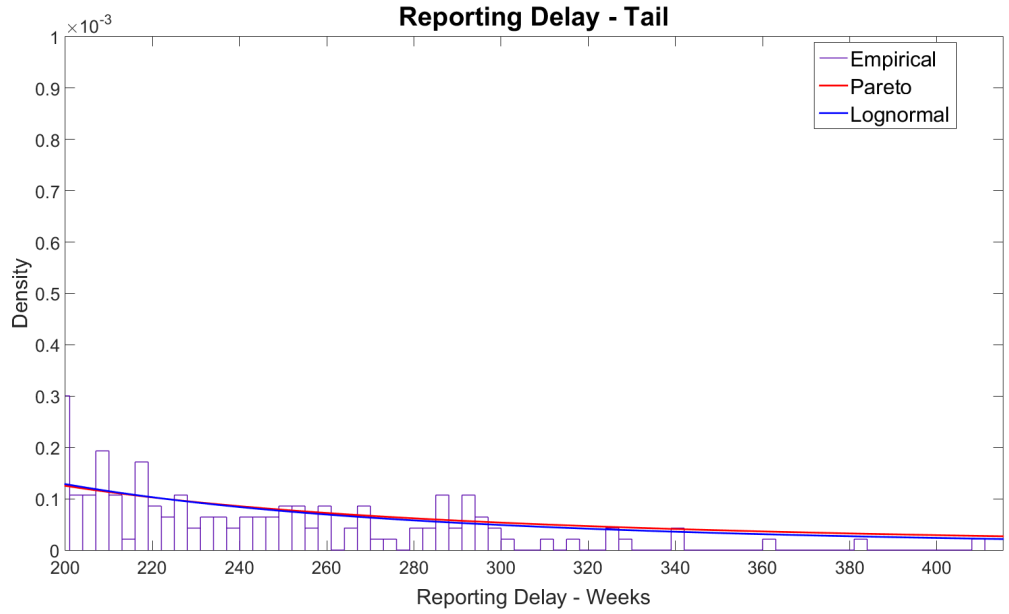

Figure 5.4: Reporting Delay fit - Tail

It is more difficult to assess the quality of the fit here compared to the body as there are significantly fewer observations. However, it seems that both the Pareto and

log-normal distributions provide a very reasonable fit to the data. The performance of both algorithms is also quite similar, but the second half of the data suggests that the log-normal provides a slightly better fit here.

5.3.3 Reporting Delay - model choice

From the qualitative analysis above, the choice of model between the Pareto and the log-normal distribution is relatively ambiguous. An alternative method to choose a suitable distribution is to use the Akaike/Bayesian information criterion in order to choose an appropriate distribution. The relevant statistics are provided in Table 5.1 below.

Distribution	Log-Likelihood	AIC	BIC
Pareto	-58668.9	117343.8	117367.4
Lognormal	-58233.9	116471.8	116487.5

Table 5.1: Reporting delay model choice statistics

All three statistics here corroborate our choice of the lognormal distribution. Thus, we will use this distribution to model the reporting delay of the claims in the public liability data.

5.3.4 Adjustment for reporting delay

Using the distribution specified above, we follow the methodology from Section 3.1.3.1 and scale the inter-arrival times of the data. The adjusted frequencies are shown in Figure 5.5 below. We note that the claim arrival intensities are still clearly not homogeneous as there is a very strong upwards trend in the data. In the following section, we investigate the impact of exposure on the claim intensities.

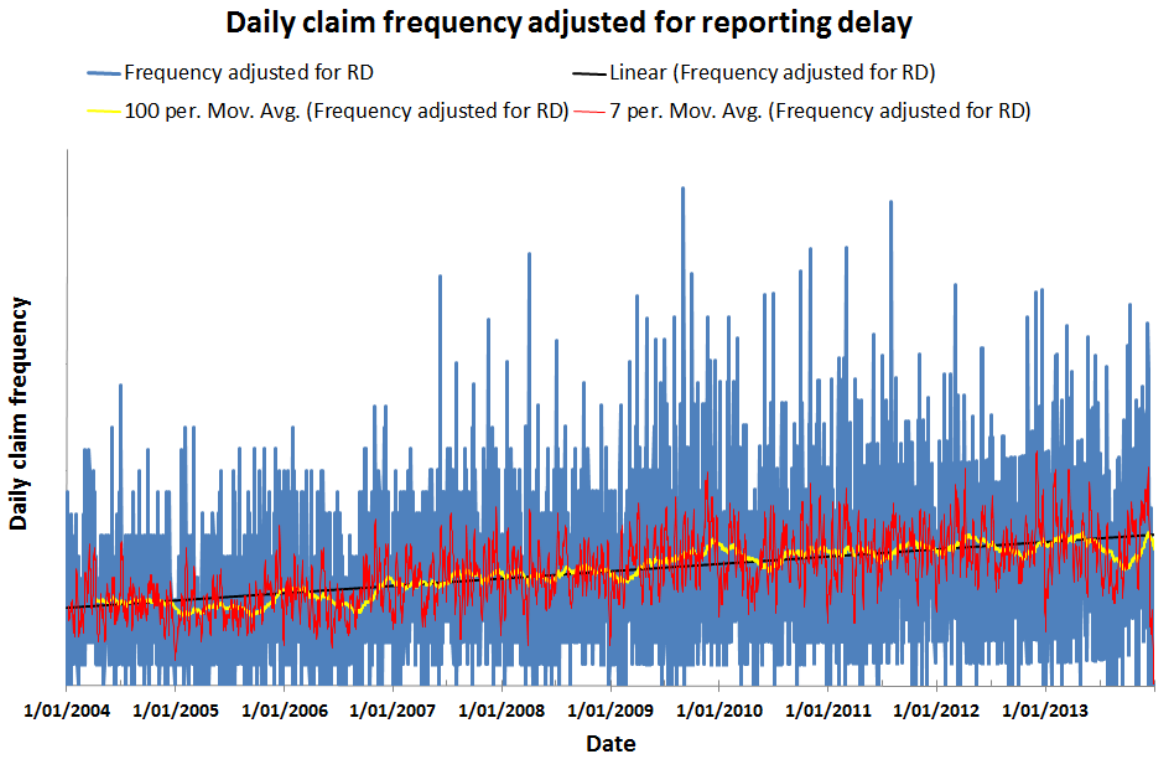

Figure 5.5: Frequency adjusted for reporting delay

5.4 Adjustment for exposure

Figure 5.6 below shows the daily number of policyholders during our investigation period from 1st of January, 2004 to the 31st of December, 2013.

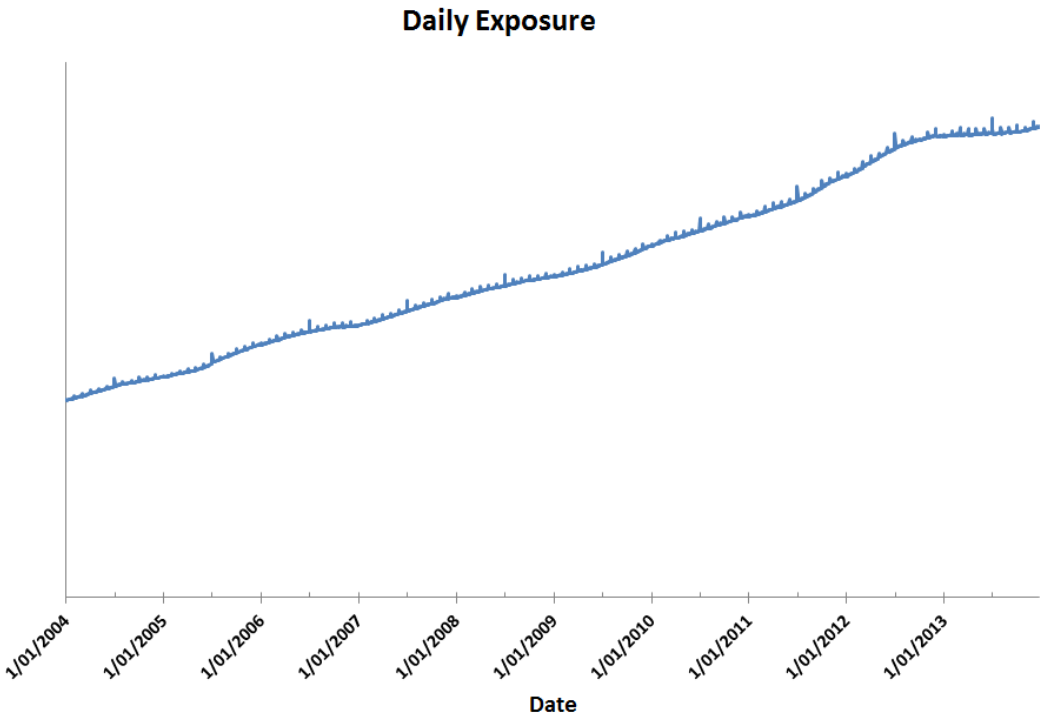

Figure 5.6: Daily exposure from 1/1/2004 - 31/12/2013

There is a similar upwards trend here and the exposure more than doubles over our investigation period. This will have a significant impact on our claims intensity, so we use the operational time scaling procedure detailed in Section 3.1.3.2. The reference date is set as the last day of the investigation period (31st of December, 2013). This date can be arbitrary designated but there are some benefits to our choice here. In practical terms, the last day of the investigation period is generally the day that the data was extracted and this day may be quite close to the current date. As a result, all the model outputs will be from the perspective of the current date (in terms of claims inflation). This allows for easy comparison with other models.

The multiplicative adjustment to the inter-arrival times are given in Figure 5.7 below. We can apply the following sense check to make sure the results are intuitively correct. In periods of lower exposure, we would expect lower numbers of claims. In order to make the intensity the same as our higher reference point, we need to shorten the inter-arrival times so that the operational "clock" of the claims process will run faster. This is what we see in Figure 5.7 below, where we have essentially scaled Figure 5.6 by the final exposure amount on the 31st of December, 2013.

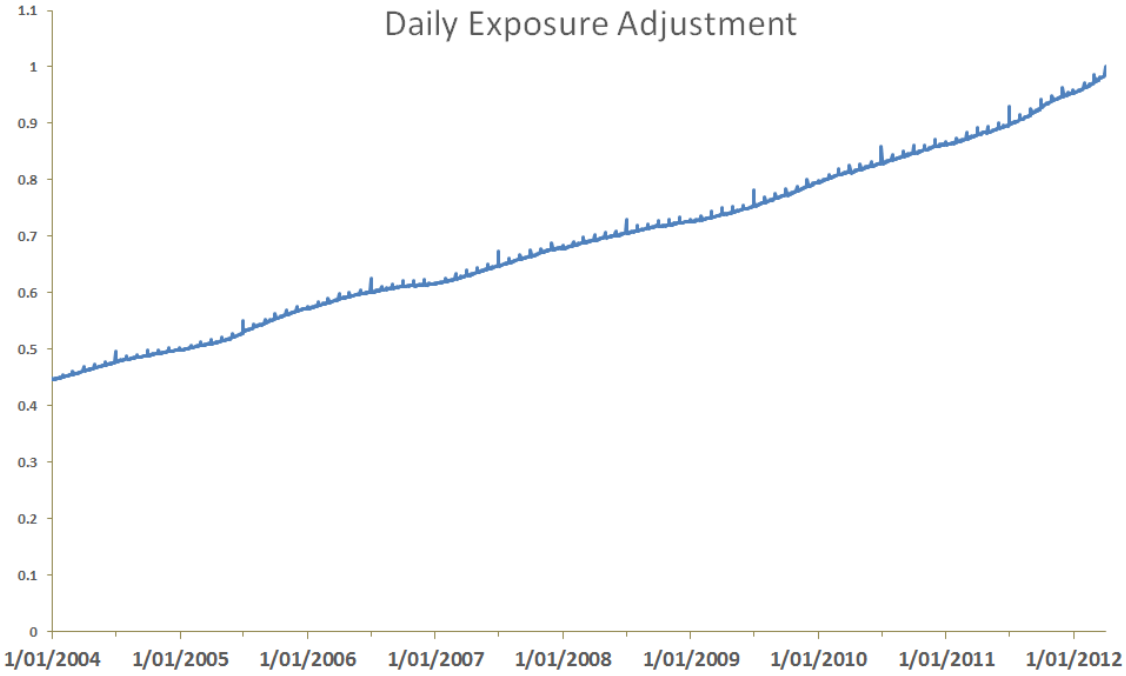

Figure 5.7: Daily exposure adjustment from 1/1/2004 - 31/12/2013

5.5 Adjustment for seasonality

We also need to adjust for any sort of seasonal trends that occur. This requires more analysis of the claims data. We investigate the following potential seasonal periods, chosen due to the nature of the public liability line:

1. Weekend
2. Quarterly

Note that we first need to adjust our frequencies/inter-arrival times by the exposure adjustment in order to remove the impact of exposure on our seasonality trends.

5.5.1 Weekend Seasonality

Figure 5.8 below gives the total amount of claims per week day over the entire investigation period.

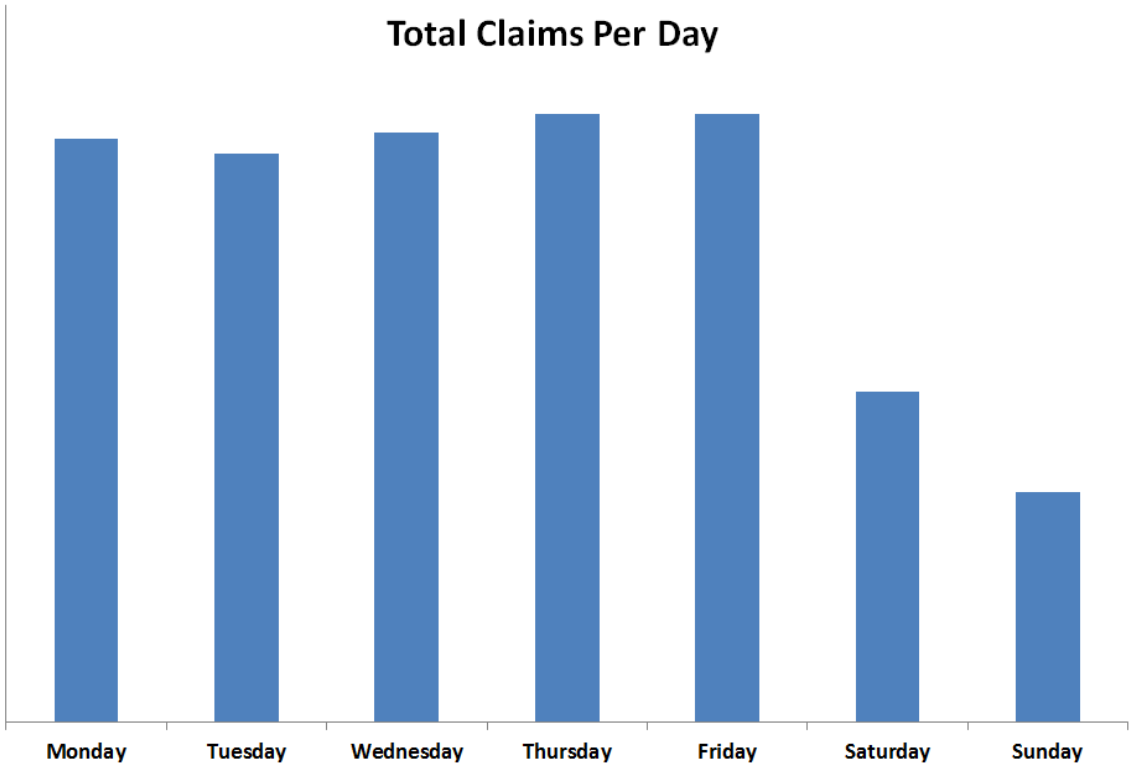

Figure 5.8: Total Claims per day adjusted for exposure

Over the weekdays, we can see a relatively homogeneous claim intensity. However, there is a clear dip over the weekends in the number of claims, which is consistent with what we would expect in the public liability line. We apply the operational time scaling outlined in Section 3.1.3.3 and set the reference point as a weekday. From our calculations, we have that

$$\frac{\text{Average Claims per weekend day}}{\text{Average Claim per weekday}} \approx 0.472. \tag{5.1}$$

Thus, we adjust the inter-arrival times on weekends by a multiplicative factor of 0.472. Doing a sense check, this means that the operational clock of the process runs faster on weekends so the claim intensity is increased to the same level as on weekdays.

5.5.2 Quarterly Seasonality

The following figure shows the total adjusted claims per quarter over the entire investigation period. These claim numbers have been adjusted for both exposure and weekend seasonality.

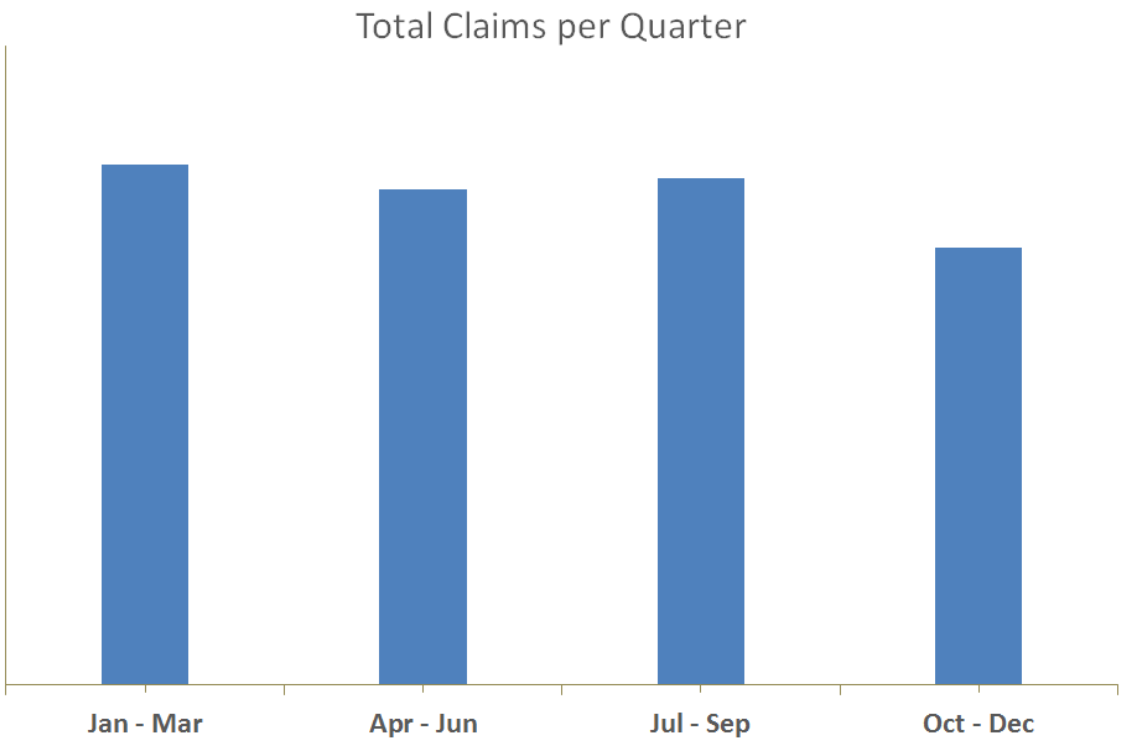

Figure 5.9: Total claims per quarter adjusted for exposure and weekend seasonality

From the graph, there doesn't seem to be a clear quarterly variation, as the total number of claims are relatively homogeneous over time. Thus, we do not adjust for quarterly seasonality.

5.5.3 Other seasonality periods

From the graph for weekend seasonality, there does not seem to be merit in implementing a time adjustment based on the day of the week, as any trend should be adequately captured by our weekend seasonality adjustment. We can also see from Figure 5.10 that there is no definitive monthly trend in the claim data. In order to avoid over-complicating the model, we also do not adjust for monthly seasonality.

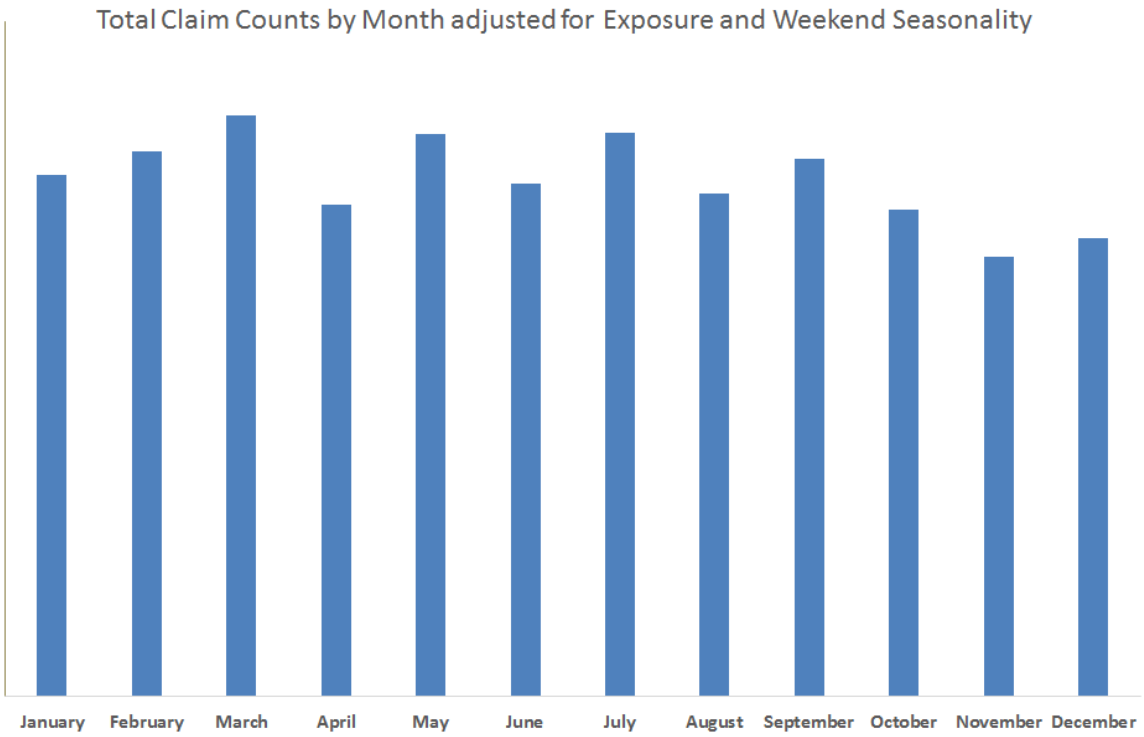

Figure 5.10: Total claims per month adjusted for exposure and weekend seasonality

However, if we wanted to incorporate these seasonal trends, the appropriate operational time adjustments are relatively straightforward and follow a similar procedure to the method presented in Section 5.5.1.

5.6 Adjustment for trend

Finally, after applying the appropriate adjustments, we can test for trends in the adjusted claim counts. Figure 5.11 below shows the claims frequency per day after adjusting for the three factors detailed in the previous sections.

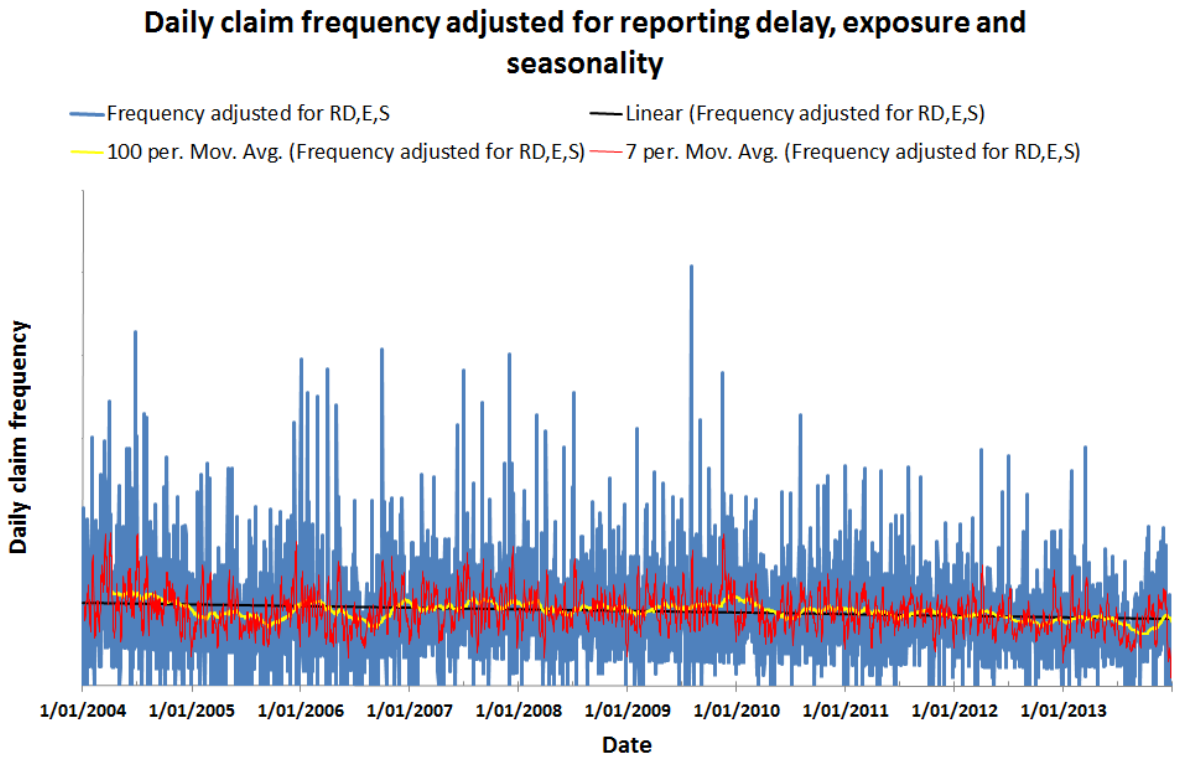

Figure 5.11: Frequency adjusted for reporting delay, exposure and weekend seasonality

We observe a slight downwards trend in the frequencies when fitting a line to the adjusted data. Using a least squares fit, we obtain the following p-values:

	p-value
(Intercept)	<2e-16
Date	2.15e-12

Table 5.2: Table: Linear fit of the adjusted frequency

Thus, we can see that the gradient of the line (which is the estimate for the Date variable) is very significant from the extremely low p-value of 2.15e-12. Because we use daily data, the gradient is equivalent to a decrease of about 2% per year. Over the ten year investigation period, this leads to a significant change in the claim arrival intensities. Thus, using the procedure from Section 3.1.3.4, we adjust the inter-arrival times with the linear fit. Figure 5.12 below plots the final adjusted frequency values.

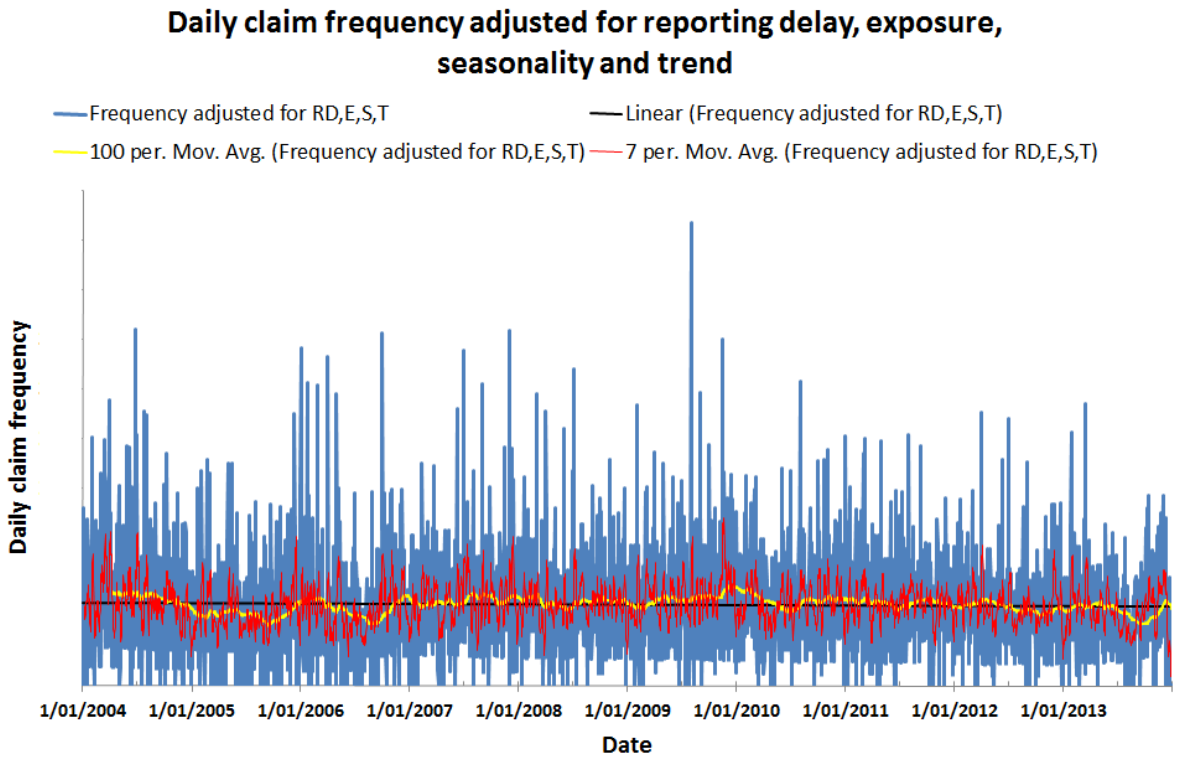

Figure 5.12: Frequency adjusted for exposure, weekend seasonality and trend

The trendline is significantly flatter relative to Figure 5.11. To statistically confirm this, we fit another linear model to the adjusted frequency and obtain the following p-values:

	p-value
(Intercept)	<2e-16
Date	0.996

Table 5.3: Table: Linear fit of the final adjusted frequency

The 0.996 p-value here clearly demonstrates that there is no trend over time in the average claims data. Thus, we have transformed the data so that the assumption of a constant first moment is satisfied. Also, if we look the second moment using the red 100 period moving average, the oscillations around the constant trendline seems to remain relatively similar over time (perhaps with a very slight decrease in variance towards the end of the investigation period). Thus, we conclude that the adjusted data now satisfies our requirement of stationarity (at least in terms of the first and second moments) and we can proceed with our Markov-modulated Poisson process parameter calibration.

5.7 Markov-modulated Poisson process calibration results

We apply the modified EM algorithm from Section 3.1.1 to calibrate the parameters of the MMPP. The required parameters are the generator matrix Q

$$\begin{bmatrix} \begin{pmatrix} -q_1 & q_{12} & \dots & q_{1r} \\ q_{21} & -q_2 & \dots & q_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ q_{r1} & q_{r2} & \dots & q_r \end{pmatrix} \end{bmatrix}, \tag{5.2}$$

and the claim intensity vector

$$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_r)^{\mathrm{T}}. \tag{5.3}$$

5.7.1 Choice of order

In the following sections, we give the estimated parameters as well as related statistics for multiple orders of the MMPP. The choice of order is based on observations of the estimates although standard AIC and BIC procedures are applicable here as well.

5.7.1.1 Order 2

$$Q = \begin{bmatrix} \begin{pmatrix} -0.3371 & 0.3371 \\ 0.4396 & -0.4396 \end{pmatrix} \end{bmatrix}, \quad \boldsymbol{\lambda} = (8.8094, 2.4123) \tag{5.4}$$

5.7.1.2 Order 3

$$Q = \begin{bmatrix} \begin{pmatrix} -0.3102 & 0.3102 & 0 \\ 0 & -0.8643 & 0.8642 \\ 0.6508 & 0 & -0.6508 \end{pmatrix} \end{bmatrix}, \quad \boldsymbol{\lambda} = (8.9227, 3.2681, 2.0016) \tag{5.5}$$

5.7.1.3 Order 4

$$Q = \begin{bmatrix} -0.5209 & 0 & 0.5209 & 0 \\ 0.6019 & -0.6019 & 0 & 0 \\ 0 & 0 & -0.8002 & 0.8002 \\ 0 & 0.6284 & 0 & -0.6284 \end{bmatrix}$$
$$\boldsymbol{\lambda} = (9.2525, 7.1322, 3.3307, 1.8720) \tag{5.6}$$

5.7.1.4 Order 5

$$Q = \begin{bmatrix} -0.6810 & 0.6810 & 0 & 0 & 0 \\ 0 & -0.8226 & 0 & 0.8226 & 0 \\ 0.7506 & 0 & -0.7507 & 0 & 0 \\ 0 & 0 & 0 & -0.8323 & 0.8323 \\ 0 & 0 & 0.7414 & 0 & -0.7414 \end{bmatrix}$$
$$\boldsymbol{\lambda} = (9.7549, 6.9618, 6.7195, 2.6364, 1.8355) \tag{5.7}$$

5.7.2 Estimation Discussion

If we examine the change in the claim intensities as we increase the order of the Markov-modulated Poisson process, we can see that the values seem to fluctuate around the two intensities initially estimated in the order 2 MMPP. In the order 3 chain, the lower two states have claim intensities that are around the lower regime in the order 2 chain. For the order 4 chain, we have essentially the order 2 chain where both the higher and lower regimes have split into two different regimes with claim intensities around the original values in the order 2 chain. Finally, looking at the order 5 chain, there is a very small difference between the claim intensities of state 2 and state 3 (6.9618 and 6.7195 respectively). It is difficult to realistically justify two different regimes with such close intensities. We also investigated chains of higher order but they also produced unrealistic or nonsensical results. Thus, it seems reasonable to use the order 2 chain in order to model our claim arrivals process for parsimony reasons.

An interesting characteristic of the calibrated Markov generators is that they imply a cyclical Markov chain process. For example, if we start from state 1 in the order 4 chain, we will then move to state 3 followed by state 4 and 2 before returning to state 1. Noting that the states are ordered from highest claim intensity to lowest claim intensity, we can interpret this as the process beginning in a high regime, moving to the lower regimes, eventually moving to the lowest regime before jumping up to the higher regimes. We

notice that movements between the two"main regimes" (high and low) only occurs from states 1 to 3 or states 4 to 2. This is most likely due to these regime shifts being more easily picked up by the calibration algorithm as the difference in intensity is much higher. This supports our hypothesis that the Markov-modulated Poisson process of order 2 is the correct model choice for our claims counting process.

5.7.3 Final Markov-modulated Poisson process parameters

Again, our final choice to model the claim arrivals is an order 2 Markov-modulated Poisson process with generator equal to

$$Q = \begin{bmatrix} -0.3371 & 0.3371 \\ 0.4396 & -0.4396 \end{bmatrix} \tag{5.8}$$

and claims intensities equal to the corresponding element of

$$\boldsymbol{\lambda} = (8.8094, 2.4123)\,. \tag{5.9}$$

5.8 Calculation of final outstanding claim counts

Finally, we apply the maximum posterior filtering algorithm described in Section 3.2 to determine the final number of outstanding claims. After determining the most likely regime at each point, we simulate the number of claims that occur during each inter-arrival period. For convenience, we aggregate the results into months. Figure 5.13 on the next page shows the comparison of the actual frequencies and the simulated frequencies from our model.

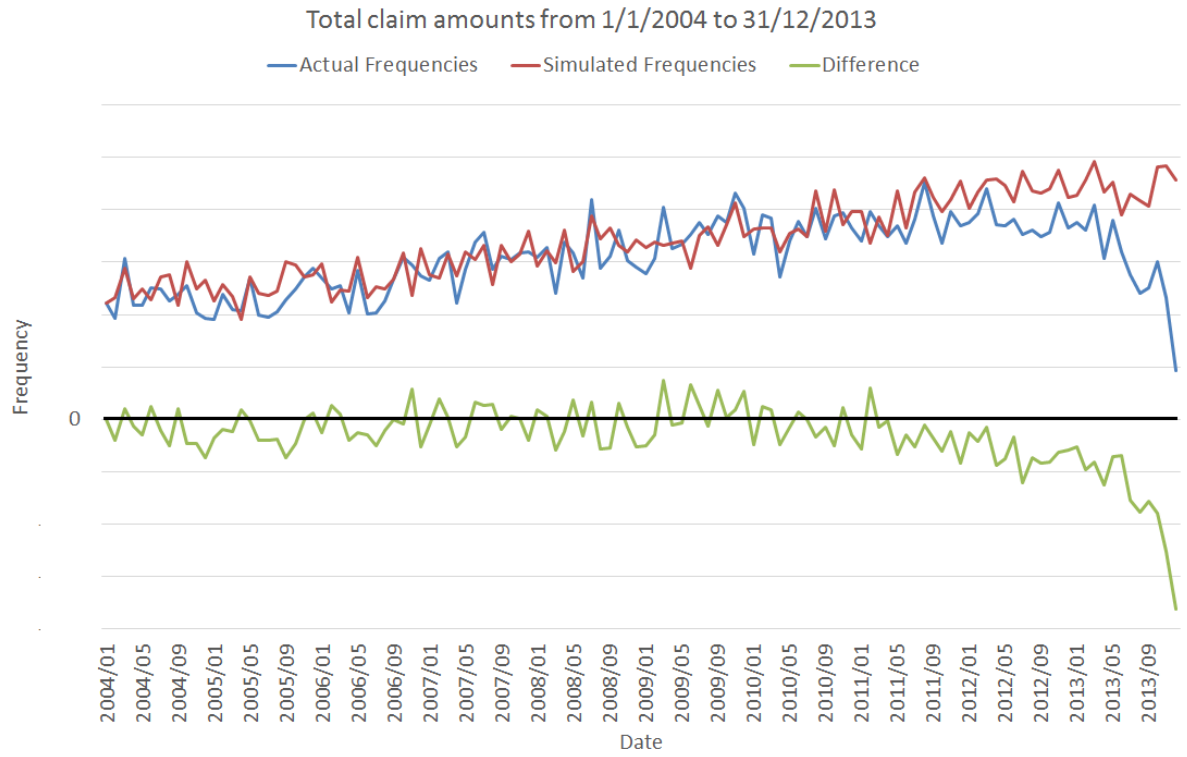

Figure 5.13: Actual vs simulated total public liability claim amounts for 2004 - 2013

For the first seven years of data (2004-2010), there is no bias on average for the difference between the simulated frequencies and the actual frequencies. Our simulated results follow the major trends displayed by the claim arrivals and generally provides a reasonable fit to the data. From the analysis on reporting delay in Section 5.3, we know that almost all claims (98.9%) are reported within three years of occurrence and thus, we would not expect a material difference between the actual and the simulated claim amounts in these first years. Also, as expected, the last three years demonstrate a significant deviation between the actual and simulated frequencies (shown by the green line in the above figure) due to the reporting delay of claims arriving between 2011 and 2013.

Our final IBNR claim amount is equal to the overall difference between these two frequencies and is calculated in Table 5.4 below.

Actual Frequency	Simulated Frequency	Outstanding Claims
18608	20504	1896

Table 5.4: Final outstanding claims reserve estimate

5.9 Comparative study of reserving methods

To demonstrate the benefits of our micro-level model, we conduct a comparison of the prediction errors between the MMPP model and the popular chain-ladder algorithm. Again, from Section 5.3, we know that there are very few unreported claims after three years. Our data set has an extraction date within the first few days of 2014 and thus, we expect that we would have almost all of the claims that occurred between 2004 and 2009. A reasonable approximation can be calculated from our reporting delay distribution, which gives that we would have 99.55% of the total number of claims. Thus, a test is conducted to compare the outstanding claims amount predicted by both models with the actual number of outstanding claims for which we have an accurate approximation.

Firstly, the claims that were reported after 31/12/2009 are removed. This essentially creates an artifical extraction date of the last day in 2009. We then fit our Markov-modulated Poisson process model to this subset of the data by following the same methodology demonstrated in the previous sections. The analysis follows a very similar structure and the results ended up being quite comparable so to avoid repetition, we do not provide detailed descriptions here. As the MMPP model is stochastic, we repeat the calibration and simulation components of the procedure one hundred times to obtain a central estimate of the total number of claims that occurred between 2004 and 2009.

The next step is to use the chain-ladder algorithm to calculate an estimate of the claim IBNRs. This method is discussed briefly in Section 2.1 and is very straightforward, so we also do not provide a detailed account of the analysis. However, we can compare the final outstanding claims in Table 5.5 below as a percentage of the actual IBNRs calculated from the full data set.

Actual IBNRs	Chain Ladder IBNRs	(Average) MMPP IBNRS
100%	96.29%	100.03%

Table 5.5: Outstanding claims estimate comparison

Thus, we can see that our Markov-modulated Poisson process model outperforms the chain-ladder algorithm and provides a much closer estimate of the outstanding claim amounts. We can further analyse the results on a quarterly basis. Figure 5.14 below shows a graph of the estimated total claims by the chain-ladder algorithm and the MMPP model, as well as the actual number of claims for one simulation. We can see that the stochastic micro-level model provides a much better fit in general to the actual number of claims than the deterministic chain-ladder model. This observation is especially true for the last year of data where the chain-ladder performs very poorly. Thus, our analysis indicates

that the Markov-modulated Poisson process model is a better model for the public liability claims.

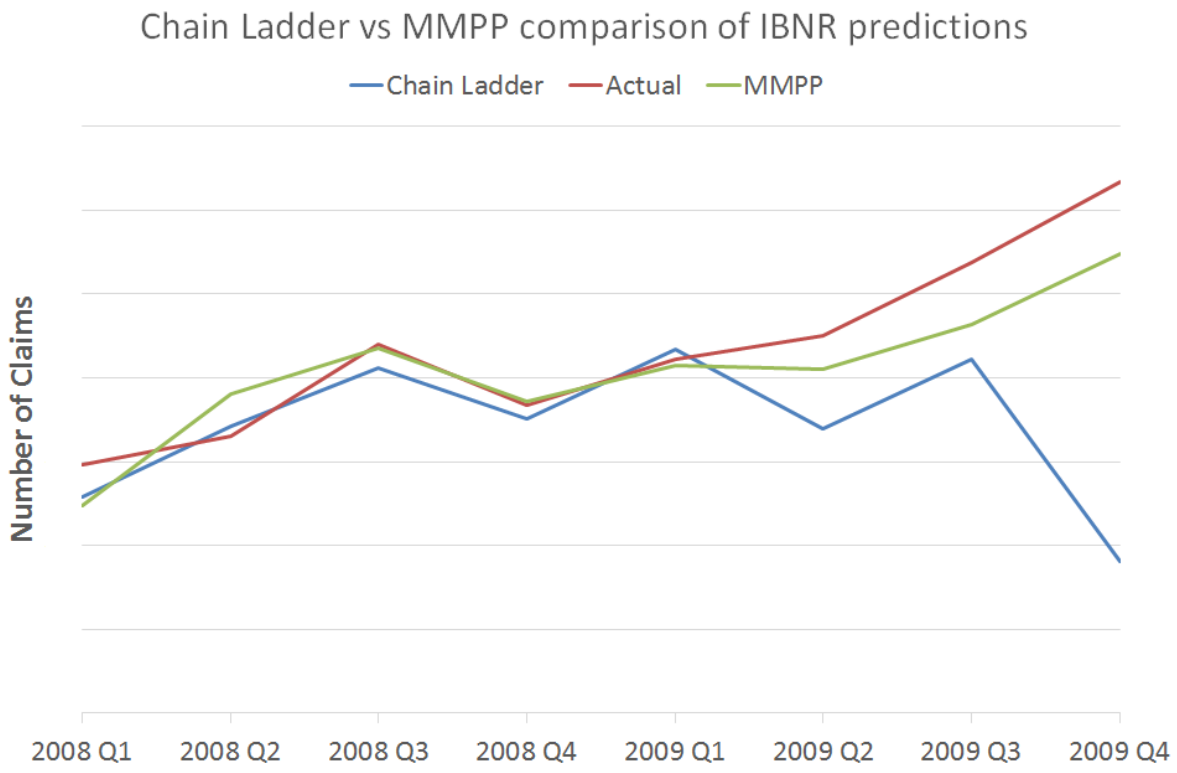

Figure 5.14: A comparison of the performance of the chain-ladder algorithm and the MMPP model for 2008-09

We note that for earlier periods, the difference between the actual total claims and the claim amounts simulated by the MMPP model does not have a strong bias. Indeed, it can fluctuate in both the negative and positive directions. In some cases, this can lead to a situation where the mode predicts total claim amounts that are lower than the actual observed claims at the date of extraction. An ad-hoc adjustment for this issue is to simply change the predicted value to be equal to the maximum of the observed amount and the predicted amount. We leave more theoretically rigorous remedies for this issue as future research.

Further, the chain-ladder in the case of claims early in the investigation period would have a total claims estimate that mainly consists of observed claims (due to the reporting delay having little impact in the earlier years). It is very possible for the chain-ladder to outperform the MMPP model in this scenario. However, this situation would only occur in the cases where almost all claims have been reported. These periods are of little interest to insurers for reserving purposes and thus, we conclude that the MMPP performs more favourably in situations that are more significant and relevant (i.e. when the claims are immature and have not fully developed.)

CHAPTER 6

CONCLUSION

6.1 Summary of results and main contributions

In this thesis, we investigate the use of Markov-modulated Poisson processes for a micro-level claims count model. The current literature on MMPPs mainly focuses on the areas of information technology and social sciences. While a small number of papers have explored these processes in an insurance context, the developments are theoretical and can make unrealistic assumptions such as a constant exposure amount over time and exact claim arrival records. Thus, the methods described in these papers cannot be directly applied to real world data. To resolve this issue and allow the MMPP models to be usable in practice, we create a new framework for adjusting the model so that the various factors that are encountered when applying real world insurance data can be incorporated into the analysis.

We first introduce the calibration algorithm from Roberts et al. (2006), which is a convenient extension to the EM algorithm proposed in Rydén (1996) as it resolves the issues of numerical underflow and computational complexity that were a significant barrier to the implementation of the model in practice. We note that there was no discussion in Roberts et al. (2006) about the actual accuracy of the new algorithm and previous papers (see Rydén (1994), Rydén (1996), Guillou et al. (2013)) only provide brief empirical examples of the EM algorithm in general. Also, there is no comprehensive analysis of the impact of the initial input parameter estimates on the final calibrated parameters. Thus, we fill this gap in the literature by providing an extensive simulation case study to verify the utility of the algorithm for calibrating the Markov-modulated Poisson process

parameters.

The second key contribution of this thesis is the operational time approach that is used to homogenise the claim arrival intensities. This approach is very flexible and allows for the incorporation of many important insurance characteristics such as reporting delay, exposure, seasonality and any other residual trends that may exist. Without these components, the MMPP model would not provide a realistic framework for modelling claim arrivals and thus, could not feasibly be used in practice. We also provide a detailed procedure to implement this operational time approach so that it is accessible to industry practitioners.

Finally, the academic literature is lacking empirical analyses that determine the performance of Markov-modulated Poisson processes using actual insurance data. We demonstrate the proposed calibration algorithm through a case study using real world public liability data. Further, we provide a comparison between the proposed micro-level model and the popular macro-level chain ladder model. It is shown that the prediction error of the outstanding claims is considerably lower when using the Markov-modulated Poisson processes approach, which validates the theoretical superior performance of our micro-level model.

6.2 Limitations and future research

The techniques developed in this thesis provides the framework for several extensions. Firstly, the order of the Markov-modulated Poisson process was chosen using qualitative analysis of the calibrated parameters Q and λ. We also noted that it is possible to apply model choice criteria such as the Akaike/Bayesian Information Criterion but these generally resulted in unrealistically high orders. Thus, a natural extension here is the application of statistical tests that can rigorously determine the most appropriate order. These tests could be in the following forms:

1. Test whether claim intensities in different regimes are genuinely different
2. Test the fit of the model over different orders

One possible procedure for the latter test is to adapt the reversible jump Markov chain Monte Carlo method in Brown and Buckley (2015) to compute a maximum a posteriori estimator for the number of components in the Markov chain. This paper applies this method to experience rating but we believe that the concepts involved are easily appropriated for our purposes. We give a brief description of the methodology below:

For some model M_i with a corresponding vector of parameters θ_i, we have a corresponding stationary distribution $\pi(M_i, \theta_i)$. The acceptance probability in our RJMCMC for a change in the order of the Markov chains (i.e. a new model M_j with corresponding parameter vector can then be expressed as

$$\min\left\{1, A\left(\theta_i, \theta_j\right)\right\} = \min\left\{1, \frac{\pi\left(M_j, \theta_j\right)}{\pi\left(M_i, \theta_i\right)} \frac{q\left(v\right)}{q\left(u\right)} \frac{r_{ji}}{r_{ij}} \left|\frac{\partial h_{ij}\left(\theta_i, u\right)}{\partial\left(\theta_i, u\right)}\right|\right\}\left(\right. \tag{6.1}$$

Another area of improvement is the use of the maximum posterior likelihood vector in order to obtain the most likely series of regimes in our observed data. However, when we try to interpret the results from a real world perspective, it is possible (as was the case in our empirical case study) that the regime switching is too frequent for any realistic interpretation. A simple method to combat this issue was to change the probability tolerance at which a change in regime occurs. This was applied in our case study but the reduction in the number of regime switches was not significant and complications can occur in choosing the correct level of this tolerance. Instead, we propose the use of statistical testing to determine whether the transitions are significant. Given that the time to a regime change is exponentially distributed (conditional on the state of the underlying Markov chain), a starting distribution for this test could be the normal distribution through the use of the central limit theorem.

We also note that the jump intensities of the MMPP that we calibrated using real world data seemed to suggest that there were two true regimes and the intensities were oscillating between the jumps. A natural extension here is to model the intensities with a Markov-modulated Poisson process where the claim intensities follow a diffusion process. The capture of the inherent intensity fluctuations could also feasibly reduce the order of the MMPP so that it is easier for statistical methods to pick the true number of regimes.

In terms of the reporting delay component of the micro-level model, we note that we achieved a reasonable fit with the standard log-normal distribution. However, an extension here is to fit a generalised linear model so that various other effects such as accident year influences can be incorporated into the model.

Further, the calibration procedure for the MMPP currently does not incorporate any adjustments for reporting delay, which limits the accuracy of the model. This will be especially true for the most recent accident years which generally the years of interest. Thus, an important extension would be to develop a method for incorporating the effect of reporting delay in the calibration of the MMPP.

As mentioned in Section 5.9, the MMPP filtering result can produce estimates that are unrealistic as the total predicted claims can sometimes be lower than the number of

observed claims. An ad-hoc adjustment is suggested but an interesting research topic would be the development of a theoretically rigorous solution to this problem.

Finally, a very important area of future research is the severity component of the micro-level model. Various papers (see Antonio and Plat (2013), Jin (2013)) have implemented models for severity but they make unrealistic assumptions such as independence between claim frequency and severity. We had also implemented a severity model in our case study but the fit was quite poor, especially for the distribution of incurred claim amounts. This is in part due to the non-continuous nature displayed by the case estimates. Examples of the severity fit can be seen in Appendix A. A straightforward extension is the application of GLMs to capture inherent characteristics and dependencies that exist in the data. We also suggest the use of mixture models that can include the impact of discrete case estimate distributions and the continuous claim payment distributions.

APPENDIX A

SEVERITY MODELLING

In this section, we depict some of the severity fits and demonstrate that while standard probability distributions may be reasonably fit to claim payments, they do not work well for fitting against case estimates. Figure A.1 below shows the fit of the severities associated with positive claim payments for bodily injury claims. From the AIC/BIC criteria, we conclude that the Weibull distribution is the best fit to the data. However, there is a clear underestimation of the density of payments between 150 and 200 dollars, which is problematic as this is a significant component in the main body of the distribution

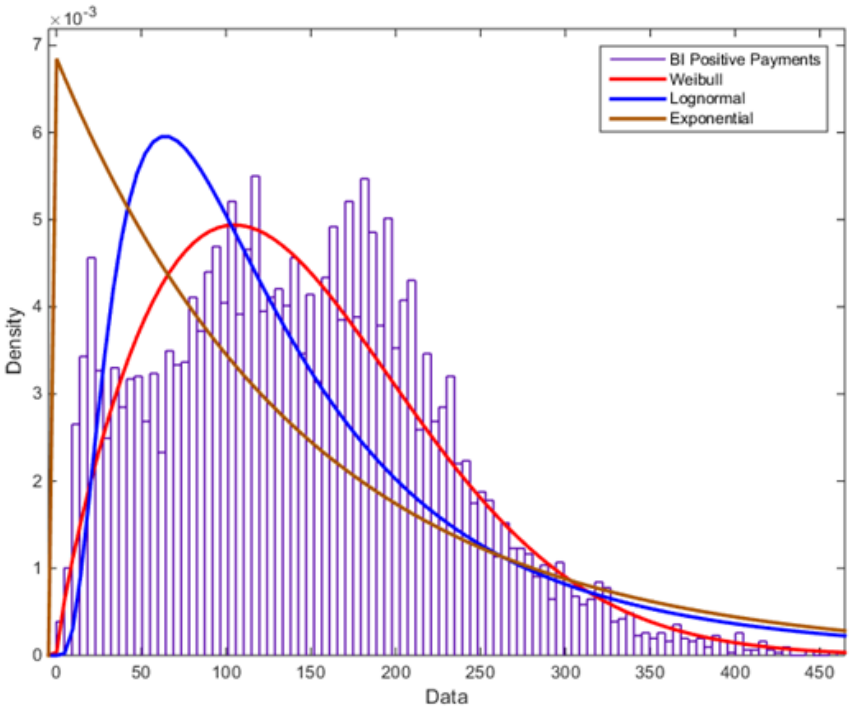

Figure A.1: BI - Postive payment severity density

Figure A.2 is similar to Figure A.1 but is instead for the negative payments. From an insurer's perspective, this would be the recoveries that are collected. The fit here is much worse and while the Weibull distribution again outperforms the other candidates in terms of the AIC and BIC, there is significant underestimation in the body of the distribution. This is also complicated by the fact that there are certain recovery amounts that seem to be standardised from some preliminary data analysis. This fact is unable to be incorporated in a standard probability distribution and thus presents another problem in modelling the claim severities.

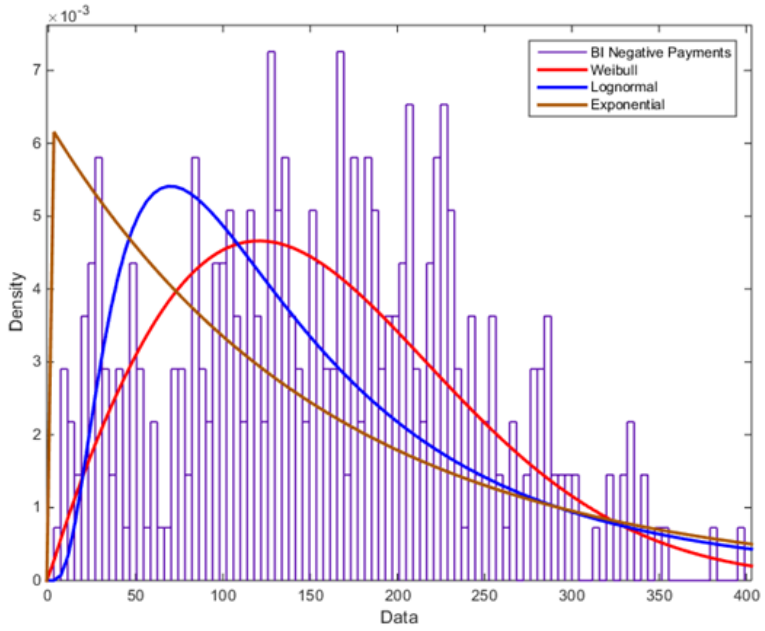

Figure A.2: BI - Recovery severity density

Finally, Figure A.3 shows the severity of transactions that occur when case estimates are revised. The main issue here is that while the main body of the distribution lies around zero, there is significant deviation on either side. These case estimates can reach magnitudes of millions of dollars and thus will greatly impact the modelling of incurred amounts. Further, the case estimate empirical distribution is not generally continuous as case managers will not evaluate a case exactly down to the dollar amount. Thus, standard probability distributions are not appropriate in this case and further research into this area would be greatly beneficial for the development of micro-level reserving models.

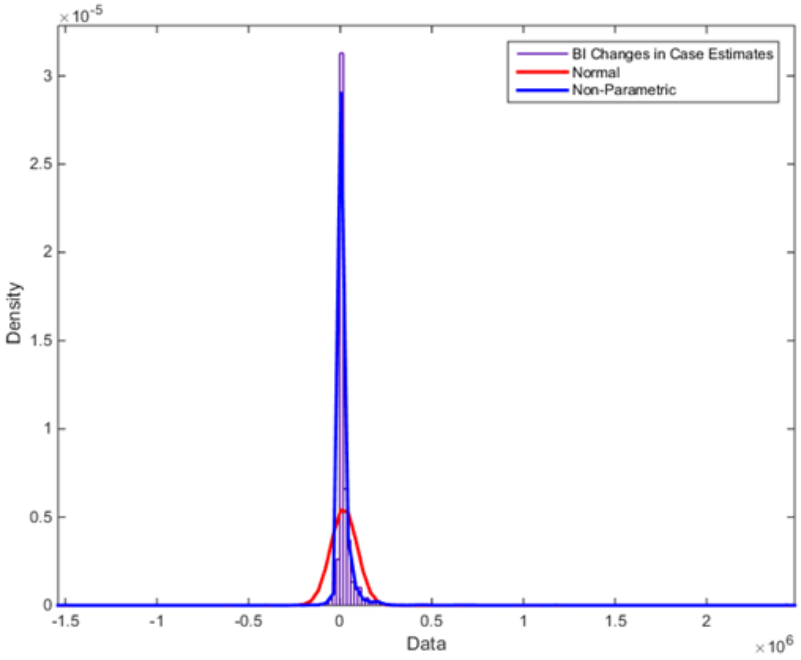

Figure A.3: BI - Severity of changes in case estimates

APPENDIX B

MATLAB CODE

B.1 Operational time adjustments

```
%% Operational Time


cnt = 1;
for i = 1 : length(DATE_OCC)
for j = 1 : Frequency(i)
DOI(cnt) = DATE_OCC(i) + j/Frequency(i) - DATE_OCC(1);
day_ind(cnt) = Weekday(i);
date_ind(cnt) = DATE_OCC(i) - DATE_OCC(1) + 1;
cnt = cnt + 1;
end
end
int_arr_real = diff(DOI);
int_arr_real(18608) = 9; %%end of period


%seasonality for weekends
weekend_seasonality = 0.472124365; %from excel investigation

%seasonality for exposure, base day is the last day (31/12/2013)
base_day_exposure = Exposure(end);
```

```
%inter-arrival time transformations from CSV



%%implement operational time

%for weekends
%first case is a Thursday
int_arr(1) = DOI(1);
for i = 2 : (length(DOI)-1)

%if stay in weekend
if and(or(day_ind(i-1) == 1, day_ind(i-1) == 7),or(day_ind(i) == 1,day_ind(i) == 7))
int_arr(i) = (DOI(i)-DOI(i-1))/weekend_seasonality;

%if stay in weekday
elseif and(and(day_ind(i-1) > 1, day_ind(i-1) < 7),and(day_ind(i) > 1,day_ind(i) < 7))
int_arr(i) = DOI(i) - DOI(i-1);

%if going into a weekend
elseif and(and(day_ind(i-1) > 1, day_ind(i-1) < 7),or(day_ind(i) == 1,day_ind(i) == 7))
int_arr(i) = (ceil(DOI(i-1)) - DOI(i-1)) + floor(DOI(i))
    - ceil(DOI(i-1)) + (DOI(i) - floor(DOI(i)))/weekend_seasonality;

%if going into a weekday
elseif and(or(day_ind(i-1) == 1, day_ind(i-1) == 7),and(day_ind(i) > 1,day_ind(i) < 7))
int_arr(i) = (ceil(DOI(i-1)) - DOI(i-1))/weekend_seasonality + floor(DOI(i))
    - ceil(DOI(i-1)) + (DOI(i) - floor(DOI(i)));
end;

end;
int_arr(length(DOI))=9; %from data until end

%for exposure
%create exposure vector corresponding to days of arrivals
for i = 1:(length(DOI))
arr_day_exp(i) = Exposure(find((DATE_OCC-DATE_OCC(1)+1)==date_ind(i)));
end;

%create vector for exposure adjustment
for i = 1:(length(date_ind)-1)
exp_adj(i) = mean([arr_day_exp(i),arr_day_exp(i+1)])/base_day_exposure;
end;
exp_adj(length(date_ind))=1;
```

```
%scale the interarrivals by the exposure adjustment
int_arr = int_arr.*exp_adj;

%%further adjust interarrival times
%intercept is at the end
gradient = -0.0004440;
intercept = 9.654 + gradient*(3652-1);

%Calculate interarrival time adjustments, where the base date is the end of
%each period
int_arr_adj = ones(1,length(int_arr));

for i = 1 : length(int_arr)
int_arr_adj(i) = (intercept-gradient*(date_ind(end)-date_ind(i)))/intercept;
end;

int_arr_final = int_arr .* int_arr_adj;

%reporting delay lognormal
ln_mean = 1.74031;
ln_std = 1.59611;

for i = 1:length(DOI)
rep_adj(i) = 1/logncdf((3652-ceil(DOI(i)))/7,ln_mean,ln_std);
end;

int_arr_final = int_arr_final .*rep_adj;
```

B.2 The scaled EM Algorithm by Roberts et al. (2006)

```
%Recursive scaled algorithm from Roberts and Ephraim 2006
%Assumed known ini_dist

function [L,R,loglikelihood, m_hat, n_hat, D_hat ,Phi_gen_new,Phi_jump_new] =
    EM_Algorithm_scaled_ndim(dim,int_arr,Phi_gen,Phi_jump,ini_dist)

n       =   length(int_arr);
%L is a column vector, R is stored as a column vector but is transposed for
%use
L       =   zeros(n+1,dim);
R       =   zeros(n+1,dim);
%Base case of recursion algorithms
```

```
L(1,:)      =   ini_dist;
R(n+1,:)    =   ones(1,dim);
%Recursion of the L and R cases
for i = 1:n
L(i+1,:) = ((L(i,:)*trans_f(int_arr(i),Phi_gen,Phi_jump))/(L(i,:)*
    trans_f(int_arr(i),Phi_gen,Phi_jump)*ones(dim,1)));
end;
for i = n:-1:1
R(i,:) = (((trans_f(int_arr(i),Phi_gen,Phi_jump)*R(i+1,:)')/(L(i,:)*
    trans_f(int_arr(i),Phi_gen,Phi_jump)*ones(dim,1)))');
end;
%calculation of n_hat which is a 2 by 1 column vector
n_hat = zeros(dim,1);
for i = 1:n
n_hat = n_hat + (L(i+1,:)').*(R(i+1,:)');
end;
%calculation of c_k for all k
c_vec = zeros(n,1);
for i = 1:n
c_vec(i) = L(i,:)*trans_f(int_arr(i),Phi_gen,Phi_jump)*ones(dim,1);
end;
%calculation of the integral of matrix exponentials using Van Loan I_k
I_k_sum = zeros(dim,dim);
for i = 1:n
I_k_sum = I_k_sum + (VL_I_k(i,dim,Phi_gen,Phi_jump,L,R,int_arr)/c_vec(i));
end;
%Calculation of m_hat
m_hat = Phi_gen.*(I_k_sum');
%Calculation of vector D_hat using the diagonal entries of m_hat
D_hat = zeros(2,1);
for i = 1 : dim
D_hat(i) = m_hat(i,i)/(Phi_gen(i,i));
end;
Phi_gen_new = zeros(dim,dim);
Phi_jump_new = zeros(dim,dim);
%for non-diagonal elements
for i = 1 : dim
for j = 1 : dim
if i ~= j
Phi_gen_new(i,j) = m_hat(i,j)/D_hat(i);
end;
end;
end;

%for diagonal elements
```

```
for i = 1 : dim
Phi_gen_new(i,i) = -sum(Phi_gen_new(i,:));
end;
%for intensity matrix
for i = 1 : dim
Phi_jump_new(i,i) = n_hat(i)/D_hat(i);
end;
loglikelihood = sum(log(c_vec));
end
```

B.3 MMPP order 2 calibration example

```
rng(1)

%specify size of matrix
dim = 2;

%Initialise the intensity parameters and the transition parameters
inten1 = 10;
inten2 = 2;
q1 = 0.1;
q2 = 0.1;

%Create the jump intensity and generator matricies
Phi_jump = diag([inten1, inten2]);
Phi_gen  = [ -q1 q1; q2 -q2];

%Create the starting stationary distribution of the states (two state
%model) which is assumed to be the initial distribution
ini_dist = [0.5 0.5];

%set tolerance for EM algorithm
tol = 1e-3;

%Run algorithm once
[L,R,loglikelihood_prev,m_hat,n_hat,D_hat,Phi_gen_new,Phi_jump_new] =
     EM_Algorithm_scaled_ndim(dim,int_arr_final,Phi_gen,Phi_jump,ini_dist);
loglikelihood = -inf;
Phi_gen = Phi_gen_new;
Phi_jump = Phi_jump_new;
count = 0;

while abs(loglikelihood - loglikelihood_prev) > tol
```

```
count = count + 1;
loglikelihood_prev = loglikelihood;
[L,R,loglikelihood,m_hat,n_hat,D_hat,Phi_gen_new,Phi_jump_new] =
     EM_Algorithm_scaled_ndim(dim,int_arr_final,Phi_gen,Phi_jump,ini_dist);
Phi_gen = Phi_gen_new;
Phi_jump = Phi_jump_new;
end;

%%filtering
%find maximum probability
for i = 1:length(L)
L_max(i) = max(L(i,:));
end;

%find state of maximum (remember to ignore first row as it is set as the
%initial distribution
for i = 1:length(L)
states(i,:) = (L(i,:) == L_max(i));
end;

for i = 1:(length(L))
for j = 1:dim
if states(i,j) == 1
state_vec(i) = j;
end;
end;
end;

%%count number of jumps
jump_count = 0;
for i = 1:(length(L)-1)
if state_vec(i) ~= state_vec(i+1)
jump_count = jump_count + 1;
end;
end;
```

B.4 Simulating from an MMPP model

```
%Simulate data
sim_Phi_gen = Phi_gen;
sim_Phi_jump = Phi_jump;

sim_q = [sim_Phi_gen(1,2), sim_Phi_gen(2,1)];
sim_lambda = [sim_Phi_jump(1,1), sim_Phi_jump(2,2)];

sim_stat_dist = [sim_q(2)/(sim_q(1) + sim_q(2)), sim_q(1)/(sim_q(1) + sim_q(2))];

%start in state one, simulate transition times, resimulate if first freqs
%are zero
time_cum = 0;
count = 1;
freq_cum = 0;
while time_cum < 3000
if mod(count,2) == 1
%simulate time until jump to other state
jump_time(count) = -(log(1-rand())/sim_q(1));
%simulate number of event s, assume uniformly distributed
freq(count) = poissrnd(sim_lambda(1)*jump_time(count));
else
jump_time(count) = -(log(1-rand())/sim_q(2));
freq(count) = poissrnd(sim_lambda(2)*jump_time(count));
end;
freq_cum = freq_cum + freq(count);
time_cum = time_cum + jump_time(count);
time_cum_vec(count) = time_cum;
count = count + 1;
end;

nz_freq = freq;
nz_time_cum_vec = time_cum_vec;
nz_jump_time = jump_time;

%remove zeros
for i = 1:length(freq)
if freq(i) == 0
nz_time_cum_vec(i) = 0;
nz_jump_time(i) = 0;
end;
```

```
end;

nz_freq = nz_freq(nz_freq ~= 0);
nz_time_cum_vec = nz_time_cum_vec(nz_time_cum_vec ~= 0);
nz_jump_time = nz_jump_time(nz_jump_time ~= 0);

%calculate exact arrival times
%first case

arr_time_index = 1;
arr_time(arr_time_index) = nz_time_cum_vec(1)/nz_freq(1);
arr_time_index = arr_time_index + 1;

for j = 2 : freq(1)
arr_time(j) = arr_time(j-1) + nz_time_cum_vec(1)/nz_freq(1);
arr_time_index = arr_time_index + 1;
end;


%other cases
for i = 2 : length(nz_freq)
%for the first case of each jump
arr_time(arr_time_index) = (nz_time_cum_vec(i) - nz_jump_time(i) +
    (nz_jump_time(i)/nz_freq(i)));
arr_time_index = arr_time_index + 1;
for j = 2 : nz_freq(i)
arr_time(arr_time_index) = arr_time(arr_time_index - 1) +
    (nz_jump_time(i)/nz_freq(i));
arr_time_index = arr_time_index + 1;
end;
end;

%Calculate the interarrival times
int_arr(1) = time_cum_vec(1)/freq(1);
int_arr(2:length(arr_time)) = diff(arr_time);
```

BIBLIOGRAPHY

Antonio, K., Plat, R., 2013. Micro-level stochastic loss reserving for general insurance. Scandinavian Actuarial Journal.

Australian Bureau of Meteorology, 2012. Record-breaking la niña events: An analysis of the la niña life cycle and the impacts and significance of the 201011 and 201112 la niña events in australia. Tech. rep.

Avanzi, B., Liu, C.-Y., Wong, B., Yang, X., 2015. Micro-level insurance claim count modelling with reporting delays: a shot noise cox process approach.

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

Billingsley, P., 1961. Statistical inference for Markov processes. Vol. 2. University of Chicago Press Chicago.

Bornhuetter, R., Ferguson, R., 1972. The actuary and ibnr. Proceedings of the Casualty Actuarial Society 59.

Brown, G. O., Buckley, W. S., 9 2015. Experience rating with poisson mixtures. Annals of Actuarial Science 9, 304–321.

Bühlmann, H., 1970. Mathematical methods in risk theory. Vol. 172. Springer Science & Business Media.

Dahl, P., 2003. Introduction to reserving. Stockholms Universitet Matematiska Institutionen.

England, P. D., Verrall, R. J., 2002. Stochastic claims reserving in general insurance. British Actuarial Journal 8 (3), 443–518.

Guillou, A., Loisel, S., Stupfler, G., 2013. Estimation of the parameters of a markov-modulated loss process in insurance. Insurance: Mathematics and Economics 53 (2), 388–404.

Guillou, A., Loisel, S., Stupfler, G., 2015. Estimating the parameters of a seasonal markov-modulated poisson process. Statistical Methodology 26 (0), 103 – 123.

Halliwell, L. J., 2007. Chain-ladder bias: Its reason and meaning. Casualty Actuarial Society 1 (2), 214–247.

Hesselager, O., 1995. Modelling of discretised claim numbers in loss reserving. ASTIN Bulletin 25 (2), 119–135.

Jin, X., 2013. Micro-level loss reserving models with applications in workers compensation insurance. University of Winsconsin-Madison, Empirical Paper.

Kunkler, M., 2004. Modelling zeros in stochastic reserving models. Insurance: Mathematics and Economics 34 (1), 23–35.

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

Larsen, C. R., 2007. An individual claims reserving model. Astin Bulletin 37 (1), 113.

Last, G., Brandt, A., 1995. Marked point processes on the real line. the dynamic approach, 1995.

Lloyd's, U. M. O., 2010. Forecasting risk: The value of long-range forecasting for the insurance industry. Tech. rep.

Mack, T., 1993. Distribution-free calculation of the standard error of chain ladder reserve estimates. Astin Bulletin 23 (02), 213–225.

Meier-Hellstern, K. S., 1984. A statistical procedure for fitting markov-modulated poisson processes. Technical report, Univ. of Delaware.

Meier-Hellstern, K. S., 1987. A fitting algorithm for markov-modulated poisson processes having two arrival rates. European Journal of Operational Research 29 (3), 370–377.

Merz, M., Wüthrich, M. V., 2010. Paid-incurred chain claims reserving method. Insurance: Mathematics and Economics 46 (3), 568–579.

Mikosch, T., 2009. Non-life insurance mathematics: an introduction with the Poisson process. Springer Science & Business Media.

Norberg, R., 1993. Prediction of outstanding liabilities in non-life insurance. ASTIN Bulletin 23, 95–115.

Norberg, R., 1999. Prediction of outstanding liabilities ii: Model variations and extensions. ASTIN Bulletin 29 (1), 5–25.

Parodi, P., 2014. Triangle-free reserving. British Actuarial Journal 19 (01), 168–218.

Posthuma, B., Cator, E., Veerkamp, W., Van Zwet, E., 2008. Combined analysis of paid and incurred losses. In: CAS E-Forum Fall. pp. 272–293.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P., 1989. Numerical Recipes, 3rd Edition. Cambridge University Press.

Quarg, G., Mack, T., 2004. Munich chain ladder. Blätter der DGVFM 26 (4), 597–630.

Ramesh, N. I., 1995. Statistical analysis on markov-modulated poisson processes. Environmetrics 6 (2), 165–179.

Roberts, W. J., Ephraim, Y., Dieguez, E., 2006. On rydén's em algorithm for estimating mmpps. Signal Processing Letters, IEEE 13 (6), 373–376.

Rydén, T., 1994. Parameter estimation for markov modulated poisson processes. Communication in Statistics. Stochastic Models 10 (4), 795–829.

Rydén, T., 1996. An em algorithm for estimation in markov-modulated poisson processes. Computional Statistics & Data Analysis 21 (4), 431–447.

Scott, S., Smyth, P., 2003. The markov modulated poisson process and markov poisson cascade with applications to web traffic modelling. Bayesian Statistics 7.

Stanard, J. N., 1985. A simulation test of prediction errors of loss reserve estimation techniques. Ph.D. thesis, New York University, Graduate School of Business Administration.

Suncorp Group, 2013. Suncorp group annual review. `http://www.suncorpgroup.com.au/annual-report-2014/`.

Taylor, G., 2003. Chain ladder bias. Astin Bulletin 33 (02), 313–330.

Taylor, G., McGuire, G., 2004. Loss reserving with glms: A case study. Casualty Actuarial Society Discussion Paper Program, Applying and Evaluating Generalised Linear Models.

Van Loan, C., 1977. Computing integrals involving the matrix exponential. Tech. rep., Cornell University.

Verdonck, T., Wouwe, M. V., Dhaene, J., 2009. A robustification of the chain-ladder method. North American Actuarial Journal 13 (2), 280–298.

Wüthrich, M. V., 2015. Non-Life Insurance: Mathematics and Statistics. Social Science Research Network.