Modelling Individual Claim Development Processes in Long Tail Insurance Products

Bo Wang

School of Mathematics and Statistics

University of New South Wales

A thesis in fulfilment of the requirements for the degree of

Doctor of Philosophy

June 2015

PLEASE TYPE

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: Wang

First name: Bo

Other name/s:

Abbreviation for degree as given in the University calendar: PhD

School: School of Mathematics and Statistics

Faculty: Science

Title: Modelling Individual Claim Development Processes in Long Tail Insurance Products

Abstract 350 words maximum: (PLEASE TYPE)

Insurance claims can have very long durations, from the date they are made known to the insurance company to the date the liability is settled. These reported but not yet finalised claims form a considerable part of the insurance liabilities of an insurance company, that is, the monies it need to set aside to pay for claims it is liable for. Traditionally, such analyses are undertaken on an aggregate basis, that is, the ultimate cost of a group of claims are projected in aggregate rather than finding the expected ultimate cost based on analysing and projecting individual claims.

This thesis is concerned with modelling the claims development behaviour of individual long tailed claims. The model developed herein (CDP model) is an individual claims reserving model, since it is concerned with individual claim trajectories and their ultimate costs.

While it may be possible to model the complete claim process in a multivariate framework with appropriate dependence structure, we have chosen to decompose the overall claim process into simpler components with a conditional hierarchical structure. The decomposition allows the complicated claims process to be represented with four simpler processes, turning each claim trajectory into four short time series.

The time series models of Generalised Linear Auto-Regression and Moving Average are used to analyse the component series. The results of these models are discussed; while some of the results conform to our intuition regarding the claim development behaviours, others have been surprising.

We use the CDP framework to project individual ultimate claims costs using simulation methods. We compared the accuracy of the predictions using the CDP model. We also aggregated the individual claims and investigated the usefulness of the method as a new valuation technique that insurance companies can use to project their claims liabilities.

We developed and investigated some important extensions to the basic CDP framework. They include making an allowance for censoring, where not all claims are observed to their finalisation; incorporating random effects, to allow for claimants differing attitudes towards the claims process

Declaration relating to disposition of project thesis/dissertation

……………………………………………………… Signature

……………………………………………………… Witness

29 June 2015 ……………………………… Date

FOR OFFICE USE ONLY

Date of completion of requirements for Award:

THIS SHEET IS TO BE GLUED TO THE INSIDE FRONT COVER OF THE THESIS

COPYRIGHT STATEMENT

‘I hereby grant the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all proprietary rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.
I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstract International (this is applicable to doctoral theses only).
I have either used no substantial portions of copyright material in my thesis or I have obtained permission to use copyright material; where permission has not been granted I have applied/will apply for a partial restriction of the digital copy of my thesis or dissertation.'

Signed ……………………………………………………........................

Date ……………………29 June 2015……………........................

AUTHENTICITY STATEMENT

‘I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.’

Signed ……………………………………………………........................

Date ……………………29 June 2015……………........................

Originality Statement

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed:____________________________

Some of the research contained in this thesis was previously published in

Wang, B. (2008). Modelling claim development processes. In *IAAust 16th General Insurance Seminar*

Abstract

Insurance claims can have very long durations, from the date they are made known to the insurance company to the date the liability is settled. These reported but not yet finalised claims form a considerable part of the insurance liabilities of an insurance company, that is, the monies it need to set aside to pay for claims it is liable for. Traditionally, such analyses are undertaken on an aggregate basis, that is, the ultimate cost of a group of claims is projected in aggregate rather than finding the expected ultimate cost based on analysing and projecting individual claims.

This thesis is concerned with modelling the claims development behaviour of individual long tailed claims, in particular, the New South Wales Compulsory Third Party insurance claims. These claims are compensation claims for bodily injuries caused by traffic accidents and can take many years to finalise, during which time they can undergo many revisions of the claims cost before finalisation. The model developed herein (the "Claims Development Process" model) is an individual claims reserving model, as it is concerned with individual claim trajectories and their ultimate costs. Individual Claims Reserving techniques are a relatively new area in actuarial science, with only an increasing presence in the statistical and actuarial literature over the last ten years.

Larsen [2007] denotes a claims process with $Z_i = (J_i, X_{i,J+1}, X_{i,J+2}, ..., X_{i,J+D_i}, G_i)$. That is, the claim can be represented by J_i, the reporting delay, its es-

timate, X_i's, observed at some regular time interval from the time it is reported to the time it is finalised, D_i, and a set of claim characteristics, G_i. Our framework makes some alterations to that of Larsen's. Firstly, we have allowed the claim characteristics to change throughout the duration of the claim; and secondly, we have changed the time scale from one based on calender time to one that is based on activity. This new time scale ticks over, or changes in value, whenever new information arrives regarding the claim causing the insurance company to revise the claims estimate; that is, this new time scale is a counter of how many revisions a claim has had. Under these changes, we rewrite the claim process as $Z_i = (X_{i,j=0}, X_{i,1}, ..., X_{i,m}, G_{i,j=0}, G_{i,1}, ..., G_{i,m})$, where j is a counter for the number of revisions the claim has had.

While it may be possible to model the X_i's in a multivariate framework with an appropriate dependence structure, we have instead chosen to decompose the overall claim process into simpler components with a conditional hierarchical structure. The claim process is decomposed into the delay (time between revisions), settlement (whether the current revision would be the final revision), direction (the direction of the revision) and size (the magnitude of the revision) component processes. The decomposition allows the complicated claims process to be represented with four simpler processes, turning each claim trajectory into four short time series. Each process is defined in terms of previous outcomes of itself and the other processes, giving a sequential conditional structure to the model.

Time series models with Generalised Linear Auto-Regression and Moving Average (GLARMA) structure are used to analyse the component series. While these trajectories are short, we still found significant serial dependence structure in the data. The results of these models are discussed; while some of the results conform to our intuition regarding the claim development behaviours, others have been surprising.

We use the CDP framework to project individual ultimate claims costs using numerical methods. Comparisons are made to actual claim settlement to show the accuracy of the individual claim projections. We also aggregated the individual claims and investigated the usefulness of the method as a new valuation technique that insurance companies can use to project their claims liabilities.

We developed and investigated some important extensions to the basic CDP framework. They include making an allowance for "censoring", where not all claims are observed to finalisation; incorporating random effects, to allow for claimants' differing attitudes towards the process of making a claim.

The thesis finishes with discussions on the contributions this method makes to the area of individual claims reserving techniques, as well as further research that can be undertaken. While the application of the new models and methods presented here are to CTP insurance claims they could also be applied in other situations such as workers' compensation, general liabilities and professional indemnity claims.

Acknowledgements

I would like to express my sincere gratitude to Professor William Dunsmuir, without whose patience, support and knowledge this thesis would not have been possible.

I would also like to thank the Motor Accidents Authority for the permission for me to use the NSW CTP claims dataset for this research project.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction

The underlying aim of this research is to contribute to the development of the analytical methods used for individual claims reserving. Individual claims reserving is an area that has enjoyed rapid development in recent years. This thesis focuses on long tailed insurance products, the claims of which can take more than ten years to settle.

In the insurance industry, the analytical modelling of long tailed insurance products is typically less sophisticated than their short tailed counterparts. However, long tailed products are intrinsically more risky to an insurer as well as, potentially, providing greater returns. Long tailed products refer to insurance products such as general liability insurance policies or injury compensation insurance policies. The statement regarding the general lack of analytical sophistication is true for both pricing, determining the premiums different policyholders should pay, and reserving, determining the pool of money insurance companies need to set aside to pay for claims.

The "reserves" of an insurance company refers to the claims liabilities the insurer has at a particular point in time. That is the value placed on the future payments the insurer needs to make for the claims that it knows about, claims that have occurred but it may not yet know about and claims that have not occurred but for which the insurer will be liable. The reserves for long tailed insurance products account for a large portion of the total reserves to a insurer. Due to the long tailed nature of the claims the liabilities can accumulate over time as opposed to short tailed products, where the claims are paid out relatively quickly. Because of this, insurers are exposed to the soundness of its reserving of its long tailed insurance liabilities. The collapse of one of Australia's largest financial institutions, HIH, comes to mind to demonstrate the need for robust reserving, amongst other factors.

The lack of sophistication in the research for long tailed products is typically a result of the following perceptions. Firstly, long tailed insurance products usually have fewer claims compared to their short tailed counterparts and this is seen as an impediment to complicated modelling. For example, in New South Wales (NSW) Australia, there are around 25,000 bodily injuries caused by motor vehicle accidents compared to around 500,000 damaged vehicles annually. In terms of volumes of data for statistical analysis, long tailed vehicle injury claims are small compared to short tailed vehicle damage claims. Secondly and more importantly, the analysis of long tailed insurance products is typically hampered by the considerable reporting delays. This makes the analysis of claims frequency difficult. For example, injuries sustained in a motor vehicle accident may not be obvious and only years later the injured party recognises the injury was due to the vehicular accident and lodges a claim. Any attempt to analysis claims occurrences would be based on censored (that is, incomplete) data. Thirdly, another censoring issue also hampers claim size analysis. As long tailed claims can take a considerable amount of time to settle; an extreme example would be that an injured child may need to be observed until high school age to establish the possibility of brain damage. This censoring issue is a key feature in long tailed claims data, where complete data, when all claims are settled, take a long time to emerge. To analyse complete claims data would mean the analysis would be carried out long after the time that the accidents occurred, which may render the results obsolete.

The current industry practice is to undertake any analysis of long tailed products at an aggregated level and even then the analysis is typically not sophisticated. When information pertaining to individual claims or individual policies is disregarded, important trends or changes in claims behaviour may not be observed.

This research aims to develop a framework that can be applied to the analysis of long tailed insurance claims. In particular, methodologies that enable a more structured analysis of the evolution of individual claims information are explored. This research explicitly allows the modelling and projection of individual claim trajectories; addressing the censoring issue mentioned previously. The estimate of ultimate claim cost at the individual level serves two purposes: firstly, when aggregated over all active claims, an estimate of the claim liabilities can be obtained in the traditional sense of reserving, and secondly, this would also implicitly aid the pricing of long tail insurance product by addressing the issue of censored data points as outlined above. These methods are applied to the NSW Compulsory Third Party (CTP) Insurance data and this dataset is used extensively throughout this thesis.

The focus of this research is on the development of reported claims through the analysis of their claim characteristics. As such, the methodologies are not applicable to the analysis of IBNR claims cost. Where necessary, we have adopted an existing aggregated claims model for the valuation of "incurred but not reported" (IBNR) claims. Hence, this research would not be able to be applied to Asbestos claims from a Workers' Compensation portfolio even though they are considered long tailed. This is because the focus of Asbestos claims is on the reporting delays of the IBNR claims.

1.2 General Insurance and Risk Transfer

General insurance (GI) is a social service that is used to transfer economic or financial risk from one party (the policyholder) to another party (the insurer). The maturity of an economy's insurance industry has been seen as an indicator of economic maturity and stability. That is, a sophisticated social being, be it an individual or an organisation, would utilise risk transfer products to manage financial uncertainties.

The underlying principle behind a general insurance product is that rational economic entities are, in the economic sense, risk averse (Hart et al. [1996]). When risk averse entities are faced with two options with the same expected value but where one provides a certain outcome (for example an expenditure or a cost) and the other option provides an uncertain outcome, they would choose the former option. Using motor vehicle insurance as an example, rational people might prefer paying a fixed cost (the insurance premium) as opposed to having an uncertain outcome (a significant loss should their vehicle be involved in an accident). In fact, risk averse economic entities are even willing to pay a little more to receive that certainty. Extending the example, the policyholder of the motor vehicle insurance policy is most likely to have paid more in insurance policy premiums than the claim payments the insurers have paid to him over a long period of time. Yet, the policyholder is willing to pay more than the expected loss to have the certainty so that they will not be financially disadvantaged if his vehicle was involved in an accident. This extra money the policyholder pays is known as a “premium” which forms the profit margin of the insurance company.

The general insurance company, on the other hand, accepts such “risks” from multiple individuals and through the “Law of Large Numbers” achieves a more certain outcome by aggregating uncertain outcomes. While the damage of a motor vehicle may represent a loss detrimental to the financial situation of an individual, by pooling a number of these risks together, the insurer is able to estimate its claims payment with reasonable certainty. Hence, the profit to the insurer lies in the extra premium the individuals are willing to pay to avoid the financial risk.

A general insurance policy is a contract between the insurer and the policyholder and this contract sets out the circumstances under which the insurer will pay the policyholder and how much the payment will be. The contract covers a period of time, usually a year, during which the insurance company is exposed to the risk of a claim payment arising from an accident. The policyholder would generally pay the insurer up front for the insurance cover. There are considerable risks involved for both the insurer and the policyholder. The risk for the insurer is not to have charged an adequate premium for the policy and is therefore unable to make a profit. The risk for the policyholder is that the insurer is not financially robust enough to be able to pay claims as required.

1.3 The Tasks of the Actuary

Actuaries are defined as managers of financial risk and uncertainty (Hart et al. [1996]). Actuaries perform two vital tasks for the insurer.

Firstly, the actuary needs to price the insurance product, that is, to determine an appropriate amount the insurer should charge for an insurance product that would cover the claim payments, the expenses the insurer incurs, as well as a profit to the shareholders of the insurance company for the risk they have borne. Pricing of GI products can take various approaches. The main "actuarial" approaches are: cost pricing, market pricing and pricing based on profit optimisation. Cost pricing is concerned with finding out exactly what a particular policy costs the insurance company to underwrite and the actual premium charged is that cost plus a profit margin. Market pricing, however, is centred around establishing customer profiles such that each policy is priced at what the customers are prepared to pay. The optimisation approach brings cost and customer behaviour together in a model that tries to maximise profit (or perhaps some other target measurement) over a set time frame.

Secondly, the actuary needs to reserve for the claim liabilities. Since the policyholder pays the premiums up front and expects to be indemnified should they suffer a loss, the insurance company needs to set a portion of the up front premiums aside to pay for claims that have not yet arisen. Due to the uncertain nature of when a claim may arise and the magnitude of the claim payment, the actuary analyses available past data and determines an appropriate amount for the insurer to set aside. This function of keeping an insurer financial sound by having enough funds set aside to cover future claims liability is mandated by law and heavily regulated in most developed financial markets.

1.4 Pricing and Reserving of Long Tailed Insurance Products

General insurance products are typically separated into short tailed business and long tailed business and the distinction between them is significant. Short tailed products tend to cover the material damage of properties where the economic loss of the policyholder is easily identifiable; for example, a damaged bumper bar due to a vehicle accident or lost luggage during a flight. To an experienced loss adjuster, the cost of rectifying the damages is clear cut and the claim is often settled quickly.

On the other hand, liability based claims, such as bodily injuries caused by a faulty product, are labelled as long tailed as they have a tendency to be finalised long after the actual event that lead to the claim. In fact, as discussed earlier, some claims may not even be known for decades after the actual event.

Reserving for long tailed products, due to its importance to the financial stability of an insurer, has received considerable input from researchers. Sophisticated methods have been developed for various specific situations and specific long-tailed products. This is not true for their short-tailed counterparts, as reserving for short tailed products is mostly straightforward. In contrast, reserving for long tailed products is a much trickier endeavour.

The main issue with long tailed claims is that during the lengthy period prior to settlement the circumstances of the claim can change considerably. For example, injuries can worsen, or liability can become denied or accepted based on litigation outcomes. As a result, the ultimate claim costs may bear no resemblance to the initial claim estimate. It is therefore up to the actuary to evaluate the information on hand and make an assessment of the value of the reserves required to pay for claims.

The more widely used reserving methods are discussed in detail in Chapter 4.

The pricing of long tailed products has always been more problematic for actuaries; and the level of rigour applied lags behind that of the short tailed products. This is due to a few reasons discussed below.

The claims data for long tailed insurance products is sparse compared to that of short tailed products. Take the example of motor vehicles in NSW; while comprehensive motor vehicle insurance has had a claim frequency of around 10% over the last ten years, Compulsory Third Party (CTP) bodily injury insurance has a claim frequency of 0.2%. In this case the volume of claims in a long tailed product is one-fiftieth of that a short tailed product, significantly reducing the volume of data available for analysing the often complex relationship between claim costs and risk factors.

Normally, claims are assumed to have independent errors; however, this may not be the case for long tailed claims. Bodily injury compensation insurance products (such as CTP claims or Workers' Compensation) are typically regulated and subject to various laws and legislations. That is, there are exogenous factors that may affect multiple claims simultaneously. For example, a precedent set in the court system which allows more generous compensation for a particular type of injury may affect the claims cost of other claims that are of a similar nature.

In this case, the assumption of independent errors may not be valid.

These features of long tailed insurance products have typically limited the pricing of these products to analysing the overall profitability of the product using valuation techniques. That is, the overall profitability of the product is examined by projecting the claims reserves "to ultimate" and comparing that to the premium received when the product is sold. If the premium is greater than the estimated claims costs plus other costs to operate the insurance operation then the product is profitable. The pricing exercise may be as simple as adjusting the premiums of the product at an overall level up or down so the desired level of profitability is achieved.

When pricing is carried out at an overall level, the profitability of individual segments of the portfolio cannot be guaranteed. As such, some policyholders may be losers and other policyholder may be winners under this overall pricing methodology. That is, there are cross-subsidies within the product; some policyholders are overcharged to pay for other policyholders that are undercharged. This makes the insurer vulnerable to anti-selection, if its competitors are more sophisticated in their pricing then the relatively better risks would leave the insurer and the relatively worse risks would stay.

In order to be able to price individual policies at a sub-product level, the projection of the claims reserve also needs to be at a sub-product level. While the claim projection for reserving is adequate at a portfolio level as reserving is mainly concerned with the overall financial strength of the insurer, claims projection for pricing purposes needs to be applied at a more granular level, preferably at an individual claim level. This is the issue on which this thesis focusses.

1.5 Aims of This Research

This research aims to examine methodologies in modelling long tail claim data and especially how claims develop over time. We have observed that notwithstanding the sparse data and the variability of the costs of long tailed claims, claims development of long tail claims still follows particular patterns based on claims characteristics.

We have firstly decomposed the complex claim development process that marks the entire development of the claim cost variable into its simpler components - the delay between the claims cost estimate being revised, whether at the current claims revision the claim is finalised, whether the current claims update results in a positive revision or negative revision and the magnitude of the revision. Various modelling methodologies are then applied to the component processes.

While understanding how claims develop and the patterns they exhibit is interesting and useful in itself, the natural extension is for reserving purposes. The estimated ultimate claim sizes of all open claims are the goal in reserving, that is, the future costs of all claims. The usefulness of the research extends further. By having estimated ultimate claim sizes at an individual claims level, this research also aids the pricing of long tailed products by removing one of the key barriers - censoring. This will allow more sophisticated pricing mythologies to be applied to the long tailed products, although pricing itself is not the focus of this thesis.

Actual data modelling is carried out on the NSW CTP dataset as maintained by the Motor Accidents Authority (MAA). We have used twelve years of complete data from 2001 to 2012 that records all claims in the NSW CTP scheme. The intention is to use nine years of data as a modelling dataset and three years as a hold out sample to check the predictive accuracy of the various models developed, assuming no change in the underlying processes and factors during this period.

1.6 Structure of This Thesis

This chapter provides an introduction to the research and articulates what it is trying to achieve in the field of individual claims reserving techniques. Chapters 2 and 3 provide a detailed discussion on the motivating dataset - NSW CTP claims.

The CTP insurance products is discussed and an overview of the NSW CTP scheme is provided in Chapter 2. Chapter 3 details the dataset the NSW Government regulator maintains for analytical purposes, the Personal Injury Register (PIR). This dataset is used throughout this thesis.

Chapter 4 applies a selection of actuarial reserving techniques (both aggregate models and individual claims models) currently used in practice to demonstrate the process of actuarial valuations of NSW CTP claims. Chapter 5 reviews the current literature in the area of individual claims reserving.

Chapter 6 details the framework for modelling claim development proposed in this research - the Claim Development Process framework. The models are then applied to the NSW CTP dataset and the results are provided and discussed in Chapter 7.

Chapter 8 uses a simulation approach to demonstrate how the models can be used to project the ultimate claim sizes for open claims. Chapter 9 extends the projection to be used as a tool for actuarial valuations. The results are compared with those from the panel current actuarial techniques.

Chapter 10 extends the model framework to incorporate random effects to allow for between claimant variations.

Chapter 11 concludes the thesis with possible further research directions.

Further analysis was carried out intending to incorporate spatial analysis into the CDP framework. However, this proved to be difficult due to the complexity of the likelihood functions to allow for spatial parameters. The spatial analysis based on the NSW CTP dataset can be found in Appendix H .

A glossary of actuarial acronyms and jargon is provided in Appendix A.

Chapter 2

CTP Insurance and NSW CTP Data

2.1 Introduction

In this research, the various modelling techniques have been applied to the NSW CTP data. This is a dataset that is compiled and maintained by the NSW CTP insurance regulator, as mandated by legislation, to aid analyses and policy making in relation to the NSW CTP scheme. In this chapter, CTP insurance and, in particular, the NSW CTP scheme are discussed in detail, then the Personal Injury Register and the data used for this research are discussed. This provides the necessary background information to the data that have motivated the model framework developed in Chapter 6. The discussion contained in this chapter also aids the interpretation of the modelling results.

2.2 Compulsory Third Party Insurance

Compulsory Third Party (CTP) Insurance in Australia refers to an insurance product that provides compensation to bodily injuries caused by motor vehicle accidents. It is a compulsory insurance mandated by law (along with Workers

Compensation Insurance) and regulated by government bodies. The insurance compensates the injured party, the claimant, with treatment expenses, loss of income and other damages due to the injuries sustained. CTP in Australia is unique in that it is sold separately to motor vehicle property damage policies, such as, Comprehensive Motor Vehicle Insurance. In other parts of the world the property damage and bodily injuries components are sold together as a packaged vehicle insurance product. The rationale to separate the two components into different policies is mainly the government's need to regulate the CTP product while wanting market forces to make the property damage component more efficient in an economic sense. The former is seen as a social necessity - the economic stress to care for a serious injury would create too much of a burden for both the claimant and the driver-at-fault. As such, the state governments of Australia have mandated the need for a CTP policy for every vehicle on the road, as a contingency in the event of an accident to protect the injured parties.

In Australia, while all states mandate CTP as a part of the vehicle registration process only three states (NSW, QLD and ACT) are privately underwritten. This means in these three states, the CTP policies are underwritten by private insurers; while the other states, the governments act as the insurer. Regardless of whether it is a privately underwritten or government administered scheme, actuaries are involved in both the pricing and valuation of the product. However, with a privately underwritten scheme the governments, or their agencies, need to regulate the insurers appropriately: it cannot allow the insurers to collude and extract a super profit from the policyholders (due to the compulsory nature of the product); at the same time, it cannot allow any of the insurers to significantly under-charge such that its financial health, and its ability to pay claim, is jeopardised. In NSW, the Motor Accidents Authority (MAA) is legislated as the regulator of the NSW CTP Scheme.

CTP schemes can be either "at-fault" or "no fault". An at-fault scheme requires the establishment of two parties in every accident - the injured party and the party at-fault. Once these two parties are established, the compensation of the injured party will be liable by the insurer of the party at-fault. Sometimes, the party at-fault can be hard to establish and legal actions are taken to establish the causal relationship of the injury. Hence, a significant portion of total costs of an at-fault CTP scheme is legal fees. On the other hand, a no fault scheme does not try to establish the party at-fault; basically all injuries due to motor vehicles are covered by the scheme. A no fault scheme can sometimes allow the injured party to resolve the claim faster due to the lack of need to establish who is at fault. This can also in turn reduce legal expenses. One key difference, however, is that in an at-fault scheme the driver at-fault is typically not covered by the CTP insurance and may sustain significant hardship if the injury is severe.

Usually, a government backed CTP scheme where the government charges the motorist a premium for the insurance and in turn pays for all of the claims would be a no-fault scheme as the government is the only insurer and would be paying for all claims regardless of who is at fault. A scheme where private insurers are the providers of the risk transfer service would typically be at-fault based; this allows the insurers to "select" better drivers to reduce claims costs. In New South Wales, a primarily at-fault scheme is employed (The Government of New South Wales [1999]). However, it can be argued that with the recent legislative changes discussed below the NSW scheme is gradually heading in the direction of a no-fault scheme.

In 2007, the Long Term Care and Support (LTCS) scheme was legislated in NSW where the long term care and treatment costs of catastrophic injuries will be managed by a central government agency. While the rationale is that the claimant does not have the expertise to manage the long term cost of the medical treatment and care of a catastrophic injury, it would be better for a centralised government agency to perform such duties especially with better bargaining power with the medical service providers. However, one important aspect of this scheme is that the admission is "no fault" based. That is, even drivers who are at-fault in an accident that leaves them catastrophically injured can receive care and support, which would otherwise be prohibitively expensive. Previously, under MACA 1999, drivers at-fault were not covered by CTP and would suffer great hardship when severely injured from the accident they were at fault for.

Also in 2007, the amended legislation (The Government of New South Wales [2007]) changed the definition of "Section 49" claims. Section 49 claims are very "minor" claims that do not need a police report and do not need to establish who is at fault. Prior to October 2008, anyone who was involved in an accident could fill out an Accident Notification Form (ANF) and receive medical treatment (usually precautionary treatment such as X-rays) up to $500. These minor claims are also referred to simply as "ANF's". A "full claim" on the other hand, required the formal lodgement of claim documentation, including police reports and other supporting evidence. Full claims are also referred to as "Section 74" claims. Under the 2007 Amendment, the injury party can now receive up to $5,000 of medical treatment as well as compensation for time unable to work due to the injuries. The rationale is to allow injured people to be able to access medical benefits quickly, without the need to formally lodge a full claim (an S74 claim) which will require fault to be established. The lodgement of a full claim can sometimes be a lengthy process which may cause some of the minor injuries to worsen if not treated promptly.

In 2013, the NSW State Government tried to introduce a "no-fault" scheme in NSW with the Motor Accidents Injuries Amendment Bill 2013 (The Government of New South Wales [2013]) but it was later withdrawn due to the Government unable to win support for the bill. At the time of writing, the NSW CTP scheme remains to be at-fault based with a no fault scheme for catastrophic injuries.

2.3 Brief History of NSW CTP

The NSW State Government first legislated that CTP insurance to be compulsory in 1942's Motor Vehicles (Third Party Insurance) Act. This was prompted by the severe financial hardship that accidents placed both the driver and injured person in. However, throughout the 1950's and 1960's a significant gap emerged from compensation amounts between Workers' Compensation Insurance injuries and CTP Insurance injuries. Following the high claim inflation during the 1970's the government had to act to contain the rising cost of CTP insurances.

The Motor Vehicles (Third Party Insurance) Amendment Act 1984 introduced some limits on damages in certain cases and allowed 5% discounting rather than 3%; however, this did not contain claims cost effectively. In 1988, the Transport Accidents Compensation Act legislated the TransCover scheme which placed limits on compensation in various situations. However, these restrictions also made it an unpopular scheme.

A change of government in 1988 brought about the Motor Accidents Act 1989, which can be considered as a major reform to the CTP industry. Firstly, indexed limits were placed on general damages (compensation for pain and suffering) and exclusions were made on general damages in small claims. Secondly, the industry was also opened up to private insurers, and the Motor Accidents Authority (MAA) was set up to regulate the insurers.

Many insurers jumped at the opportunity; however, the "tail" of CTP insurance was longer than some of the insurers anticipated. They did not expect that the cost of a claim could change dramatically (usually upwards) even many years after the date of the accident. This "escalation" of claims cost brought along with corresponding increases in insurance premiums. The government tried to step in and contain, unsuccessfully, the spiralling cost of CTP policies in 1995.

In 1999 the Motor Accidents Compensation Act (MACA) 1999 was introduced with the following changes to improve the scheme overall, not just from a cost perspective.

- Earlier notification of accident and treatment - by completing an Accident Notification Form (ANF), claimants are entitled to up to $500 of treatment they feel necessary.

- Earlier settlement of claims - clear timeframes for lodging claims and for insurers to respond to claims.

- Dispute resolution - the Claims Assessment and Resolution Service (CARS) was setup to resolve disputes in a non-adversarial environment. Claimants who disputes CARS assessment can appeal in court; however, they are not allowed to submit further evidence.

- Independent medical costs assessment - the Medical Assessment Service (MAS) was set up to determine the cost of treatment, rehabilitation and future care a claimant requires by an independent team of medical professionals. Their assessment is binding on the parties, CARS and the court system.

- Removal of non-economic loss from small claims - this is done through a whole person impairment threshold (10%) above which the claimant can claim for non economic loss.

- Regulation of fees and costs - this is done to cap the costs paid to legal and other professionals for their services.

Since then several amendments to the Act and other relevant Acts have been introduced. The most notable being the Long Term Care and Support (LTCS) Scheme. This enforced catastrophically injured persons from motor vehicle accidents to be collectively looked after by the State Government on a no-fault basis. Firstly, lump sum payments are no longer made to catastrophically injured persons due to their lack of investment strategies and wealth planning and the money simply ran out too quickly. Secondly, the benefits have been extended to at-fault parties to promote social welfare as it was deemed too costly for the family of the injured party if they received no support in the treatment and subsequent care of these persons.

Considerable amendment to the MACA has occurred in 2010 and this is partly why the modelling data was chosen to be 2001 to 2009. The change in question is that from 2010, a considerable number of smaller claims will no longer be full CTP claims, that is, they would be ANFs (The Government of New South Wales [2010]). The intention is for the minor injuries (such as minor bruising) to not clog up the CTP scheme resources and allow the more serious claims to proceed more quickly. As such, the claims arising under the new legislation will not be examined in this research. The other reason for only modelling claims from 2001 to 2009 is that the claim development during 2010 to 2012 have been reserved for validation of our models. That is, our models allow us to form a view on how the claims that are still open at the end of 2009 would develop in the future. We intend to use the actual developments of the claims during 2010 to 2012 as a way to validate our projections.

However, the non-inclusion of the new legislative framework does not invalidate our models and conclusions. If the aim was to understand how the claims under the new legislative environment would behave our modelling data can be "processed" to represent the new environment by removing the data relating to claims that are under $5,000 in size. A second and perhaps more interesting question is to use the framework developed in this thesis to answer how does the change in legislation alter claims development behaviour. We believe this question may be harder to answer. From prior experience, legislative changes takes a number of years to stabilise as all parties, claimants, regulator, insurers and lawyers, take some time to understand and adjust to these changes. Three years of data post change for a long tailed portfolio like CTP may not be enough to establish these new behaviours.

2.4 Market and Insurers

In NSW, private insurance companies are allowed to participate in the NSW CTP scheme by obtaining a license. There were close to 20 licenses issued since the industry was deregulated. However, a series of claim cost escalations led to consolidation in the market, there is now only 5 insurance companies with 7 licensed brands, they are as follows. Their respective market shares during 2011/2012 are contained in brackets (Motor Accidents Authority [2012]).

- AAMI (Suncorp) (8.8%)
- Allianz (12.6%)
- CIC (Allianz) (4.2%)
- GIO (Suncorp) (12.7%)
- NRMA (37.8%)
- QBE (16.8%)
- Zurich (7.1%)

The market can be classified as an oligopoly, in which "game theory" (Konstantinidis et al. [2007]) dominates the way prices are determined. This refers to insurer's desire to find out what the competitors are charging and then charge slightly less than the best competitor in profitable segments of the market and charge slightly more than the worst competitor in unprofitable segments. That is, each insurer would try to anticipate the premium levels of their competitors and make their decisions based on that. The employment of two or more licenses allows some insurers to segment the market and allows different brands to charge different premiums due to the different price elasticities of the segmented markets.

The NSW CTP Scheme has a “file and write” system where each insurer notifies the MAA on the premium rating structure up to three months in advance and once approved by the MAA the insurer can start issuing CTP insurance at the filed premium levels.

The CTP insurance in NSW is still largely community rating based; that is, every driver enjoys a relatively similar premium. Evidence suggest young male drivers can be sometimes three times more costly to insure compared to the overall average. However, the MAA will only allow approximately a 50% loading to be applied to the average premium on any one policy to address the issue of affordability (Motor Accidents Authority [2014a] and Motor Accidents Authority [2014b]). This heavy cross subsidy makes insurers purposefully avoid young drivers. However, being a compulsory insurance product, insurers cannot refuse to insure a driver. The main method in which insurers try to target different subset of the driving population is to design advertisement campaigns that attracts the 30 to 60 year old policyholders but not the very young drivers. The fact that insurers cannot charge the correct premium for some segment of the market means that the game theory and premium positioning for various segments of the market is very important in the setting of premiums. This aspect of the NSW CTP market is discussed in detail in Konstantinidis et al. [2007].

2.5 CTP Claims

When a motor vehicle accident occurs in NSW any injured persons may be eligible for compensation under the NSW CTP scheme for losses relating to their injuries, although the driver at-fault will have significant limitations to the entitled compensation. Such losses can be medical examinations, medical treatments, economic loss for not being able to work, medical aids or modifications performed to car or home or legal costs involved.

Since the police must be notified of all accidents resulting in bodily injury, each CTP claim must be accompanied by a police report detailing the accident and this allows the insurer to establish the party at-fault. The insurer of the at-fault vehicle is usually appointed as the managing insurer and will manage the claims of all injured persons from the accident. That is, this insurer is liable for all the claimants and their compensations. Each injured person of that accident has his or her own "file" and is allocated with a claim number.

When a party at-fault is difficult to establish or multiple parties contributed to the cause of the accident then the liability is said to be shared between the insurers of the vehicles in a determined ratio. In this case a managing insurer will be selected between these insurers and will make all the decisions about these claims; the other insurers need to share the cost in the determined ratio. When there is no party at-fault (accident caused by natural perils, etc) or the at-fault party cannot be found, then the claim is referred to as a "nominal defendant" claim, an insurer is appointed at random to manage the claims from that accident. However, all insurers in the NSW CTP scheme share the claims cost of such claims proportional to their market share.

Subsequent to the establishment of the claim file claim managers will look after the claim until the claim is finalised. This process is somewhat similar to other bodily injury claims (such as Workers Compensation Insurance or Public Liability insurance), where the claims manager may undertake the following actions.

- Gather information by talking to the claimant or the legal representative of the claimant.

- Monitor the injuries and conditions of the claimant through medical reports and work with various medical professionals to seek treatment and rehabilitation of the claimant.

- Make payments for expenditures of the claimant in terms of medical and other expenses.

- Work with the claimant or his or her legal representation to determine an ultimate compensation amount to settle the claim.

- If the claim progresses further, such as through CARS or the court system, then the claim manager may be required to prepare evidence and work with the legal counsels of the insurer.

Upon setting up the claim file, the claims manager makes an initial estimate of the cost of the claims from the information available. This cost would then vary up or down as the claims manager receives new information; this process is typically known as claims development. At the settlement of the claim, all of the cost of the claim would be paid out, either to the claimant or to the various professionals through the claim process, the claim will then be closed.

Since CTP insurance is heavily regulated and is a commoditised product, in that all insurers are essentially offering the same product as dictated by the government regulator, the claims cost for similar claims between different insurers are likely to be similar. However, different insurers do employ different claims settlement strategies to obtain a comparative advantage in reducing claims costs. Such strategies may include the following.

- Active claims management, which refers to the claims manager contacting lawyers and medical professionals over the phone rather than waiting for a response by traditional mail correspondence.

- Early settlement of claims, that is, settling claims earlier on, prior to the claimants' medical conditions deteriorating.

- Bulk conferencing, claims manager makes a trip to rural areas and personally visit lawyers looking after claims in one particular area. The claims manager presses the lawyers to settle claims while he or she is "on-site".

Due to the strategies involved, the claims development pattern may be different across different insurers.

2.6 Personal Injury Register Database

The MAA maintains the Personal Injury Register (PIR), a database of all claimants since the enactment of Motor Accidents Act 1989. It is required by the law that all insurers submit their claims data to this centralised database. A comprehensive set of claims information is to be submitted by the insurers on a quarterly basis. This information is combined across all insurers in the industry and available to each insurer a few weeks later with the data fields that may be used to identify the claimant, driver and managing insurer blanked out. The main purposes of the PIR are to:

- monitor the CTP scheme
- calculate premium levels and premium rating factors
- investigate the occurrence of fraud, and
- assist the MAA in deciding funding initiatives to reduce road accidents

When the insurers submit their information, a very stringent process of data validation is carried out. Each accident will be cross matched to a police accident report and an accident ID will be allocated. Each claimant will be given a claimant ID and linked to the accident. The "role" of each claimant is also recorded, for example, the driver at-fault or a pedestrian. This allows various quantitative studies, such as vehicle safety or road safety, to be undertaken. All the payments the insurer makes in regards to a claim is also carefully broken down into types of compensations, also known as, "Heads of Damage" or HoD. Such a breakdown can then gauge the effectiveness of various legislation in controlling the cost of providing CTP insurance.

The database is then made available to the insurers. Each insurer can obtain two sets of tables, their own claims (with claim identifiers to match with their own claim reference number) and the complete claims, with the managing insurer field masked. The two files when used together allow the insurers to directly compare their claims experience with that of the industry. The quarterly extract that the insurer can obtain is a subset of the complete database, and contains the following tables.

- Claims record information - this table includes information such as date of accident, litigation status, severity and nature of injury, claim status, etc at the date of data extraction. Historically, the first five injuries are recorded in this database; or if more than five are presented at the same time, then the most severe five injuries are chosen. The injury coding used in the earlier years is AIS (Abbreviated Injury Scale) 1985 Update. More recently the injury coding adopted is the AIS 2005 update. However, a mapping between the two updates is available and we can convert between the two coding systems reliably.

- Payment information - this table includes information on payments made during each quarter since the lodgement of the claim. The type of payment (such as medical treatment or legal costs) is also included.

- Case estimation information - this table includes information on what the managing insurer thinks is the amount left to be paid regarding a claim at the end of each of the quarters since the lodgement of the claim.

Notice that in each extract any one claim can only appear once in the qualitative claims information table yet can appear multiple times in the payment and case estimation files. A few randomly selected sample records of this database are provided in Tables 2.1 to 2.3.

obs_no	112012	122012	132012	152012	162012	172012	182012	192012	202012	212012	222012	232012	242012	252012	262012	272012	282012
acc_id	A98968	A108214	A118021	A51276	A103459	A125343	A133395	A140531	A149772	A157493	A165894	A173802	A183215	A190424	A198855	A207466	A215557
acc_date	30/08/1992	21/06/1997	25/02/1995	18/07/1998	31/07/1997	29/03/1999	21/06/2000	31/01/2001	14/07/2001	17/06/2002	27/02/2003	9/12/2003	3/05/2004	20/06/2005	29/04/2006	14/03/2007	18/02/2008
anf_flag							Y	Y			Y	Y		Y	Y	Y	Y
clm_flag	Y	Y	Y	Y	Y	Y	Y		Y	Y			Y	Y	Y		Y
clmreg_date	31/01/1994	30/11/1997	31/03/1995	28/02/1999	31/07/1999	31/03/2000	31/07/2000	28/02/2001	31/10/2001	30/06/2002	31/03/2003	31/12/2003	31/10/2004	31/07/2005	31/05/2006	30/04/2007	31/03/2008
clmrep_date	5/01/1994	16/11/1997	16/03/1995	8/02/1999	9/07/1998	27/03/2000	13/05/2004	.	15/10/2001	28/06/2002	.	.	6/10/2004	12/12/2005	3/07/2006	.	7/04/2008
anfrep_date	.	.	.	.	.	.	10/07/2000	6/02/2001	.	.	26/03/2003	29/12/2003	.	14/07/2005	24/05/2006	2/04/2007	5/03/2008
fin_date	13/11/1995	17/01/2003	18/01/1999	13/09/2005	12/07/1999	.	10/06/2004	1/05/2001	15/07/2005	26/09/2002	3/12/2003	26/07/2004	21/10/2004	.	9/01/2007	25/09/2007	.
reopen_date	.	.	.	13/09/2005	.	.	24/05/2004	.	.	.	.	.	.	.	.	.	.
status	F	F	F	F	F	O	F	F	F	F	F	F	F	O	F	F	O
injcd_1	Y02	K10	C07	B01	K12	B05	Y02		G05	U01			Y02	G01	J08		G01
injury_1	9	3	3	1	3	3	9	.	2	6	.	.	9	1	1	.	1
injcd_2		C06	C07	G01	K11	E09			G01					F01	A01		
injury_2	.	1	3	1	2	3	.	.	1	.	.	.	.	1	1	.	.
injcd_3		A01	K12	I01	A01	E09			K09					A01			
injury_3	.	1	3	1	1	3	.	.	1	.	.	.	.	1	.	.	.
injcd_4			K12	A01		E09			K13								
injury_4	.	.	3	1	.	3	.	.	1	.	.	.	.	.	.	.	.
injcd_5			B04			B07			E02								
injury_5	.	.	2	.	.	2	.	.	1	.	.	.	.	.	.	.	.
inj_sev	.	11	22	3	10	18	.	.	6	75	.	.	.	3	2	.	1
brain						Y											
spine																	
whiplash				Y					Y					Y			Y
MAIS	.	3	3	1	3	3	.	.	2	6	.	.	.	1	1	.	1
nom_def	N	N	N	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N
claimant	DRI\I	PAS\I	PAS\OWN\I	PAS\I	MCR\OWN\I	PAS\I	DRI\OWN\I	DRI\OWN\I	PAS\I	PAS\I	I\PED	DRI\OWN\I	DRI\I	PAS\I	DRI\I	PAS\I	DRI\OWN\I
workcomp							Y						Y				
shared	N	N	N	N	N	N	Y	N	N	N	N	N	N	Y	N	N	N
numshare	.	.	.	.	.	.	3	.	.	.	.	.	.	3	.	.	.
veh_rat	339	301	301	301	701	739	708	301	301	701	310	371	310	321	301	701	302
vehclass	04Aa	1	1	1	1	04Aa	06a	1	1	1	7	03e	7	03e	1	1	03c
dis_type	S	V	S	S	S		S	D	C	L	L	L	S		D	S	
lit_ind	N	Y	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	N
lit_lev		2	2	2													
leg_rep	N	Y	Y	Y	N	Y	N	N	Y	N	N	N	N	Y	N	N	N
offer_date	.	.	.	.	.	.	20/05/2004	.	25/03/2003	.	.	.	6/10/2004	10/01/2007	.	.	.
settle_date	.	4/10/2002	.	.	.	.	20/05/2004	1/05/2001	14/07/2005	26/09/2002	.	26/07/2004	11/10/2004	.	9/01/2007	25/09/2007	.
threshld		N					N	N	N	N	N	N	N	X	N	N	X
assesmnt																	
liabilty	N	R	R	F	F	F	F	A	F	F	A	A	F	F	X	A	F
carsref_date	.	.	.	.	.	.	.	.	9/07/2004	.	.	.	.	.	.	.	.
carsdec_date	.	.	.	.	.	.	.	.	15/06/2005	.	.	.	.	.	.	.	.
carsdcsn									A								
lastpay_date	28/06/1995	13/01/2003	11/01/1999	6/07/2004	.	5/12/2006	7/06/2004	.	13/07/2005	17/09/2002	23/04/2003	6/12/2004	11/10/2004	3/03/2006	1/11/2006	6/05/2007	11/06/2008
case_est	0	0	0	0	0	1309658	0	0	0	0	0	0	0	24716	0	0	16393
tpay_NET	843	20299	86687	42875	0	34675	4351	0	135099	12434	343	478	353	258	1420	69	2094
rehab	4	3	4	3	2	4	5	6	4	1	2	2	5	4	4	4	4
eco_flag	N	N	Y	N	N	N	Y	N	Y	N	N	N	N	N	N	N	N
emp_stat	E	B	E	H	E	A	E	E	E	C	E	E	E	H	E	B	E
prior	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N
fatality	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	N
postcode	2150	2145	2762	2485	2140	2000	2567	2000	2100	2850	2148	2144	2046	2756	2154	2480	2234
gender	M	M	F	M	M	F	F	M	F	F	M	M	F	M	F	F	F
garpcode	2112	2145	2765	4575	4000	2259	2570	2106	2100	2850	2761	2142	2015	2148	2145	2480	2234
occupation	6	.	6	.	4	.	2	1	2	.	4	4	6	.	2	.	2
weekly_earn	.	.	.	.	.	.	962	.	1412	.	.	.	.	.	.	.	0
YOB	1950	1980	1973	1971	1963	1995	1966	1959	1956	1925	1944	1950	1963	1977	1956	1989	1961

data has been transposed to fit the paper

Table 2.1: Sample of Claims Record Information Table

obs_no	acc_id	pay_year	pay_qtr	pay_type	pmt_GST	pmt_NET
112012	A98968	1994	3	M1	564	564
112012	A98968	1994	4	L1	279	279
112012	A98968	1995	1	L1	-279	-279
112012	A98968	1995	2	L1	279	279
122012	A108214	1997	4	L1	77	77
122012	A108214	1998	1	L3	38	38
122012	A108214	1999	1	L3	2239	2239
122012	A108214	2000	4	IA	50	46
122012	A108214	2001	2	LE	88	80
122012	A108214	2001	3	IA	1157	1051
122012	A108214	2002	1	LE	550	500
122012	A108214	2002	3	IA	1127	1025
122012	A108214	2002	4	LH	3061	2973
122012	A108214	2002	4	LF	12375	11250
122012	A108214	2003	1	IA	71	65
122012	A108214	2003	1	LH	955	955
132012	A118021	1995	2	A1	0	0
132012	A118021	1996	2	L1	227	227
132012	A118021	1997	3	L1	1200	1200
132012	A118021	1997	4	L1	2868	2868
132012	A118021	1998	1	L3	127	127
132012	A118021	1998	1	L1	850	850
132012	A118021	1998	2	L1	80	80
132012	A118021	1998	2	L3	3372	3372
132012	A118021	1998	3	L3	1500	1500
132012	A118021	1998	4	L2	27500	27500
132012	A118021	1998	4	M1	7528	7528
132012	A118021	1998	4	E2	5000	5000
132012	A118021	1998	4	G1	20600	20600
132012	A118021	1998	4	E1	11872	11872
132012	A118021	1999	1	L3	3963	3963
152012	A51276	1999	1	L1	167	167
152012	A51276	1999	2	L1	38	38
152012	A51276	2000	1	IA	1116	1116
152012	A51276	2000	2	TM	200	200
152012	A51276	2000	3	TM	415	415
152012	A51276	2000	4	IA	375	340
152012	A51276	2002	3	LH	81	74
152012	A51276	2002	3	LE	700	636
152012	A51276	2002	4	IA	797	737
152012	A51276	2003	1	LH	605	550
152012	A51276	2003	4	LE	1104	1037
152012	A51276	2003	4	LF	1650	1500
152012	A51276	2003	4	LG	1320	1200
152012	A51276	2003	4	LH	835	778
152012	A51276	2004	1	LH	781	730
152012	A51276	2004	2	LE	10450	9500
152012	A51276	2004	2	LF	4400	4000
152012	A51276	2004	2	NE	10500	10500
152012	A51276	2004	2	EP	11086	11086
152012	A51276	2004	2	LA	9000	9000
152012	A51276	2004	2	EF	3000	3000
152012	A51276	2004	2	TM	5000	5000
152012	A51276	2004	2	LH	1392	1271
152012	A51276	2004	3	VC	-20000	-20000
172012	A125343	2000	2	IA	1010	1010
172012	A125343	2000	4	RM	350	350
172012	A125343	2000	4	IA	170	158
172012	A125343	2001	1	IA	600	600
172012	A125343	2001	3	LG	330	300
172012	A125343	2001	4	TM	100	100
172012	A125343	2002	1	TM	1225	1225
172012	A125343	2002	2	LE	3967	3607
172012	A125343	2002	2	RO	645	625
172012	A125343	2002	3	RO	50	50
172012	A125343	2003	1	RO	265	265
172012	A125343	2003	2	IA	241	219
172012	A125343	2003	3	IA	125	113
172012	A125343	2003	3	RT	46	46
172012	A125343	2003	4	LG	1144	1040
172012	A125343	2005	1	LE	7645	6950
172012	A125343	2005	1	LG	241	219
172012	A125343	2005	1	LH	29	26
172012	A125343	2005	2	LG	1144	1040
172012	A125343	2005	3	LG	2200	2000
172012	A125343	2005	3	RX	1650	1650
172012	A125343	2005	4	LG	0	0
172012	A125343	2006	3	IA	15	14
172012	A125343	2006	4	LF	2200	2000
172012	A125343	2006	4	LE	11840	10763
172012	A125343	2006	4	LH	329	305
182012	A133395	2000	3	SP	0	0
182012	A133395	2004	2	IA	16	16
182012	A133395	2004	2	EP	2885	2885
182012	A133395	2004	2	TM	571	571
182012	A133395	2004	2	TP	639	639
182012	A133395	2004	2	HU	240	240
202012	A149772	2001	4	IA	120	109
202012	A149772	2002	1	TM	360	327
202012	A149772	2002	2	TM	88	80
202012	A149772	2002	3	TM	249	227
202012	A149772	2002	4	IA	385	350
202012	A149772	2003	1	TT	667	606
202012	A149772	2003	1	IA	880	800
202012	A149772	2003	2	IA	220	200
202012	A149772	2003	3	TT	236	214
202012	A149772	2003	4	TT	429	390
202012	A149772	2004	1	TT	442	401
202012	A149772	2004	2	TT	389	354
202012	A149772	2004	3	TT	307	279
202012	A149772	2004	4	IA	880	800
202012	A149772	2005	2	IA	44	40
202012	A149772	2005	2	LE	4928	4480
202012	A149772	2005	3	LA	19162	17420
202012	A149772	2005	3	IA	220	200
202012	A149772	2005	3	EF	100000	90909
202012	A149772	2005	3	TM	18604	16913
212012	A157493	2002	3	IA	394	358
212012	A157493	2002	3	HU	573	520
212012	A157493	2002	3	OB	12071	10974
212012	A157493	2002	3	TM	644	582
222012	A165894	2003	2	SP	68	61
222012	A165894	2003	2	SM	310	282
232012	A173802	2004	2	SP	55	55
232012	A173802	2004	4	SP	423	423
242012	A183215	2004	4	TT	353	353
252012	A190424	2006	1	IA	284	258
262012	A198855	2006	2	SP	100	90
262012	A198855	2006	3	RP	250	226
262012	A198855	2006	3	TM	606	550
262012	A198855	2006	3	IA	17	16
262012	A198855	2006	3	IA	121	110
262012	A198855	2006	3	TP	300	271
262012	A198855	2006	4	IA	25	22
262012	A198855	2006	4	RP	150	135
272012	A207466	2007	2	SM	58	53
272012	A207466	2007	2	IA	18	16
282012	A215557	2008	2	TM	60	55
282012	A215557	2008	2	SP	1572	1429
282012	A215557	2008	2	IA	671	610

Table 2.2: Sample of Payment Information Table

obs_no	acc_id	caseyear	case_qtr	case_Est	ce_nel
112012	A98968	1994	1	0	0
112012	A98968	1994	2	2943	0
112012	A98968	1994	4	0	0
112012	A98968	1995	1	0	0
112012	A98968	1995	2	559	0
112012	A98968	1995	3	280	0
112012	A98968	1995	4	280	0
112012	A98968	1996	1	0	0
122012	A108214	1997	4	0	0
122012	A108214	1998	1	119923	0
122012	A108214	1998	2	119885	0
122012	A108214	1998	3	119885	0
122012	A108214	1998	4	119885	0
122012	A108214	1999	1	119885	0
122012	A108214	1999	2	117646	0
122012	A108214	1999	3	117646	0
122012	A108214	1999	4	117646	0
122012	A108214	2000	1	117646	0
122012	A108214	2000	2	0	0
122012	A108214	2000	3	0	0
122012	A108214	2000	4	159818	0
122012	A108214	2001	1	159773	0
122012	A108214	2001	2	159773	0
122012	A108214	2001	3	159693	0
122012	A108214	2001	4	158642	0
122012	A108214	2002	1	158642	0
122012	A108214	2002	2	158142	0
122012	A108214	2002	3	158142	0
122012	A108214	2002	4	157117	0
122012	A108214	2003	1	157387	0
122012	A108214	2003	2	0	0
132012	A118021	1995	1	0	0
132012	A118021	1995	2	88000	0
132012	A118021	1995	3	88000	0
132012	A118021	1995	4	87715	0
132012	A118021	1996	1	87715	0
132012	A118021	1996	2	87715	0
132012	A118021	1996	3	87488	0
132012	A118021	1996	4	87488	0
132012	A118021	1997	1	87773	0
132012	A118021	1997	2	87773	0
132012	A118021	1997	3	87773	0
132012	A118021	1997	4	242433	0
132012	A118021	1998	1	111954	0
132012	A118021	1998	2	110977	0
132012	A118021	1998	3	103025	0
132012	A118021	1998	4	80275	0
132012	A118021	1999	1	10275	0
132012	A118021	1999	2	0	0
152012	A51276	1999	1	0	0
152012	A51276	1999	2	22333	0
152012	A51276	1999	3	22295	0
152012	A51276	2000	2	21178	0
152012	A51276	2000	3	20978	0
152012	A51276	2000	4	20563	0
152012	A51276	2001	1	20188	0
152012	A51276	2002	4	20406	0
152012	A51276	2003	1	68202	0
152012	A51276	2003	2	67597	0
152012	A51276	2003	4	87395	20000
152012	A51276	2004	1	82484	20000
152012	A51276	2004	2	81703	20000
152012	A51276	2004	3	76696	9500
152012	A51276	2004	4	0	0
152012	A51276	2005	1	1	0
152012	A51276	2005	4	0	0
162012	A103459	1999	3	0	0
162012	A103459	1999	4	0	0
172012	A125343	2000	1	0	0
172012	A125343	2000	2	716500	0
172012	A125343	2000	3	1240490	0
172012	A125343	2000	4	1240490	0
172012	A125343	2001	1	1239982	0
172012	A125343	2001	2	1239382	0

obs_no	acc_id	caseyear	case_qtr	case_Est	ce_nel
172012	A125343	2001	3	1239382	0
172012	A125343	2001	4	1239082	0
172012	A125343	2002	1	1238932	0
172012	A125343	2002	2	1237707	0
172012	A125343	2002	3	1243526	0
172012	A125343	2002	4	1338775	0
172012	A125343	2003	2	1338510	0
172012	A125343	2003	3	1338269	0
172012	A125343	2003	4	1338097	237000
172012	A125343	2004	1	1336953	237000
172012	A125343	2005	2	1329037	237000
172012	A125343	2005	3	1327893	237000
172012	A125343	2005	4	1324043	237000
172012	A125343	2006	1	1322723	237000
172012	A125343	2006	4	1322708	237000
172012	A125343	2007	1	1309658	237000
182012	A133395	2000	3	0	0
182012	A133395	2000	4	15984	0
182012	A133395	2001	2	0	0
182012	A133395	2004	3	0	0
192012	A140531	2001	1	0	0
192012	A140531	2001	2	600	0
192012	A140531	2001	3	0	0
202012	A149772	2001	4	0	0
202012	A149772	2002	1	2280	0
202012	A149772	2002	2	2240	0
202012	A149772	2002	3	2799	0
202012	A149772	2002	4	5800	0
202012	A149772	2003	1	5415	0
202012	A149772	2003	2	81453	0
202012	A149772	2003	3	86233	0
202012	A149772	2003	4	100997	65000
202012	A149772	2004	1	100568	65000
202012	A149772	2004	2	100126	65000
202012	A149772	2004	3	99737	65000
202012	A149772	2004	4	99430	65000
202012	A149772	2005	1	101150	65000
202012	A149772	2005	2	101150	65000
202012	A149772	2005	3	124783	0
202012	A149772	2005	4	0	0
212012	A157493	2002	2	0	0
212012	A157493	2002	3	0	0
212012	A157493	2002	4	0	0
222012	A165894	2003	1	0	0
222012	A165894	2003	2	4500	0
222012	A165894	2003	3	4157	0
222012	A165894	2003	4	4156	0
222012	A165894	2004	1	0	0
232012	A173802	2003	4	0	0
232012	A173802	2004	1	0	0
232012	A173802	2004	2	600	0
232012	A173802	2004	3	22	0
232012	A173802	2004	4	0	0
232012	A173802	2005	1	0	0
242012	A183215	2004	4	0	0
242012	A183215	2005	1	0	0
252012	A190424	2005	3	0	0
252012	A190424	2005	4	1500	0
252012	A190424	2006	1	48000	0
252012	A190424	2006	2	47716	0
252012	A190424	2007	1	17216	0
252012	A190424	2007	4	24716	0
262012	A198855	2006	2	0	0
262012	A198855	2006	3	500	0
262012	A198855	2006	4	3139	0
262012	A198855	2007	1	2964	0
262012	A198855	2007	2	0	0
272012	A207466	2007	2	0	0
272012	A207466	2007	3	444	0
272012	A207466	2007	4	0	0
282012	A215557	2008	1	0	0
282012	A215557	2008	2	5000	0
282012	A215557	2008	3	16393	0

Table 2.3: Sample of Case Estimation Information Table

Table 2.1 is known as a “Claim Header” file and provides the basic information regarding the claim, such as, “date of the accident”, “age of claimant - time of accident”, “legal representation”, etc. Some of the variables are static (e.g., age of claimant at the time of the accident) or dynamic (e.g., legal representation status). Table 2.2 is a “Payment Transaction” file and shows the date and amount of each payment made and also the nature of the payment. Table 2.3 is a “Balance Outstanding” file that shows the balance outstanding (or case estimate) of the claim at the end of each quarter. The variables contained in Tables 2.1 to 2.3 and their descriptions are provided in Appendix C.1.

2.7 Data Used for this Research and Issues with the PIR

Complete claims data has been collected and maintained since 1989. The data that the MAA collects is suitable for the analysis of claims development as every quarter a snapshot of all the claims in the system is collected from the insurers and then made available to the insurers for analytical purposes. Hence, by appending these quarterly data together a complete history of the claims is compiled and their development can be analysed. As noted above, while the historical payments and case estimates information for each claim are available in every extract, the other claims information is available as a snapshot. The concatenation of the historical claims data would provide a complete history of even the qualitative information. This will be discussed later on, but the qualitative information can be loosely termed as static or dynamic. A complete history of claim characteristics would allow the understanding of how the changes in these dynamic variables impact the claim development behaviours.

Since the implementation of the Motor Accidents Compensation Act in 1999 (MACA 1999) more than 14 years of data is available at the time of writing. A

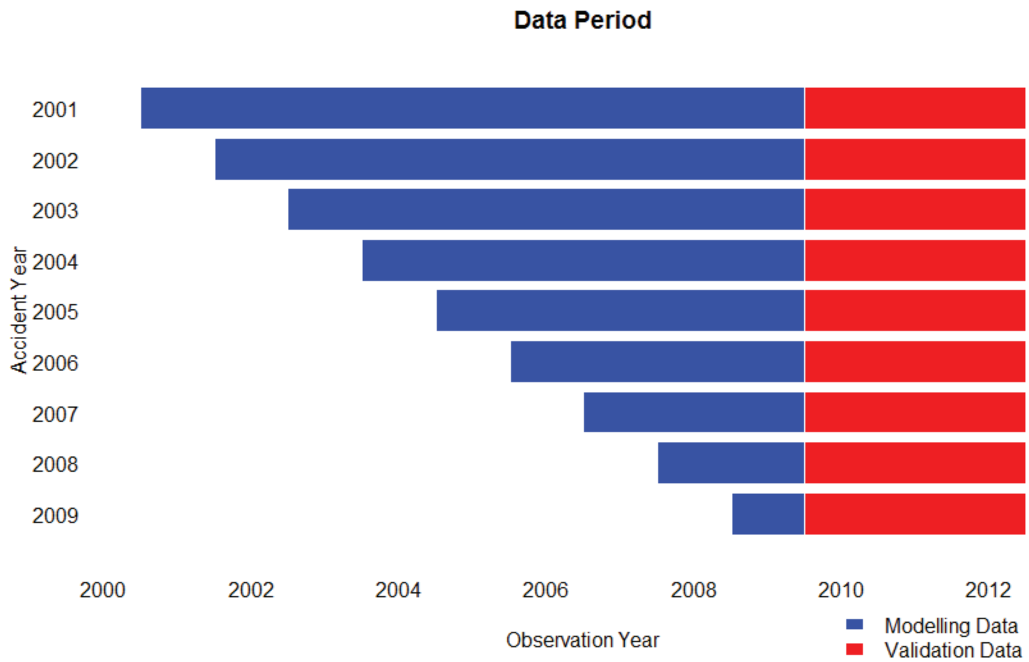

Figure 2.1: Time Period of Data Used

twelve year data period, 2001 to 2012, has been adopted for this research. The data is split into a modelling dataset and a validation dataset as represented in Figure 2.1. The data contains all full claims (ANFs are excluded) occurring from 1 January 2001 to 31 December 2009; the modelling dataset observes them until 31 December 2009 while the validation dataset observes these claims (which may contain claims lodged after 31 December 2009 arising from accidents that occurred prior) for a further three years.

The validation data allows comparison between the projected claims paths and claims costs to the actual realisations in the three year period. This use of the validation data serves two purposes. Firstly it enables an examination of the efficacy of the individual claims modelling approach when used as a valuation method, that is, compare between the total claims cost resulting from the predictive modelling and the actual observed claims costs in the validation dataset (see Chapter 8). Secondly, based on such comparisons, calibration can be performed to improve future predictions.

The choice of the data period was made after careful consideration of the following factors.

Firstly, The first 15 months of MACA data was not used as MACA started 1 October 1999. It is felt the behaviour of the scheme would be erratic under the new legislation and that it would take the participants (claimants, regulators, lawyers, insurers, etc.) some time to understand and get used to the new rules and features of the scheme. During this period, we observed claim frequency reduced sharply and feel the claims occurring from 2001 would be more stable and more reflective of the claims to be expected under the new scheme.

Secondly, under MACA Amendment 2007, effective from 1 October 2008 there has been a major change to the claim profile. After this date, minor claims (also known as ANFs) were extended from $500 to $5,000. Under the MACA Amendment 2007, claims of less than $5,000 would be greatly reduced as the lodgement of a full CTP claim is considerably more time consuming than the lodgement of an ANF. However, due to the unadvertised nature of this change, up until the end of 2009, very few ANFs were of a size greater than $500. That is, if the claimant's costs were greater than $500, a full claim were still lodged. Hence, the end of 2009 were chosen as a cut off point where the behaviour of the smaller claims (less than $5,000) were quite different before and after this date.

Thirdly, the split of 12 years of data into 9 years of modelling data and 3 years of validation data allows a significant portion of available data to be used for model development and model fitting while the three years of validation data allow the claims enough time to develop in order to gauge the accuracy of the claim size projections of the individual claim models.

Fourthly, ANFs have been excluded from our analyses. Since ANFs were designed to reimburse the injured's parties out of pocket expenses (up to a limit of $500) their claim sizes are almost certain when the ANFs are lodged. That is, there is almost no development in the sizes of ANFs. Also while numerous in number, the ANFs are of minor sizes and the insurers do not have a great deal of interest in them as their financial impact is minor. ANFs may make up around 20% of the total number of reported CTP matters (claims and ANFs in total) but only represent around 0.5% of the cost.

While the NSW CTP data is generally complete and relatively free of problems, we have identified several issues that can be dealt with straightforwardly.

- Injury Coding Change - Prior to September 2008, all injury coding was carried out on the AIS85 coding format; however, from September 2008, coding was done using AIS05 for new claims or claims that has had a revision (where MAA required the insurer to change the coding to the new format). The Abbreviated Injury Scale (both the 1985 and 2005 revisions) used the first four digits to denote the location and type of the injury and the last digit as an injury severity measure. Due to the coding change, only the location of the injury and the injury severity will be used in the modelling. While most of the injuries have their severities maintained from AIS85 to AIS05, some of injuries did have their severities revised. In particular stress related injuries which could have been rated as a severity 1, 2, 3 or 4 have been revised down to severity 1 injuries. There is no way to re-map them as we would not know whether the change is a genuine change or a change due to the coding change. Fortunately, this only applied to a very small percentages of the claims.

- Missing Variables - in the earlier stages of the PIR database, some data fields were collected but not provided by the insurers. For example, “Employment Status” was not available until 2002. However, the missing variables in the earlier extracts are typically static variables, and these blank fields were able to be “back filled” with the later extracts.

2.8 Usage of PIR

The NSW CTP PIR dataset is used extensively throughout this research. In Chapter 4, it is used to demonstrate the current mainstream actuarial valuation techniques as methods for projection the ultimate claim sizes on an aggregate basis. In Chapter 6, the individual claim characteristics are discussed and the framework for the Claim Development Processes (CDP) are developed. Chapter 7 applies the claim development processes modelling framework to the PIR and the results are discussed. Chapter 8 applies the CDP as a predictive modelling tool and predicts the ultimate claim sizes of the open claims at the end of the modelling dataset; comparisons are made to the traditional methods.

The MAA has kindly granted permission for the PIR to be used in this research.

Chapter 3

Detailed Description of the Data

3.1 Introduction

This chapter provides a detailed summary of the motivating dataset, the NSW CTP data. While the rationale underlying the construction of the Personal Injury Register (PIR) was discussed and an overview on the data structure was provided in the last chapter, this chapter aims to provide visual and numerical summaries of this dataset. This furthers the discussion regarding the features of the dataset and aids the development of the model framework in later chapters.

3.2 Actuarial Triangles

In this section, the actuarial concept of a "triangle" is introduced as well as a general discussion of data used in this section. As noted earlier, the NSW CTP data used in this research contains all claims from the accident periods 2001HY1 (1st half year of 2001) to 2009HY2 observed until the end of 2012. However, this dataset is split into a modeling dataset where the censoring date is at the end of 2009 and a validation dataset where the censoring date is at the end of 2012.

The data is then arranged in the format of a triangle. Table 3.1 shows the number of claims lodged with the insurers for the CTP dataset.

Accident		development period (j) - periods after the accident period																							
period	k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
2001HY1	1	2,343	5,338	5,668	5,851	5,951	6,035	6,099	6,167	6,218	6,231	6,253	6,264	6,268	6,272	6,277	6,282	6,284	6,284	6,283	6,283	6,283	6,283	6,283	6,283
2001HY2	2	2,234	4,800	5,185	5,317	5,413	5,497	5,583	5,651	5,663	5,672	5,678	5,685	5,691	5,693	5,697	5,698	5,699	5,699	5,699	5,703	5,703	5,703	5,703	
2002HY1	3	2,146	4,748	5,111	5,271	5,365	5,452	5,540	5,566	5,577	5,596	5,602	5,605	5,606	5,607	5,612	5,612	5,613	5,613	5,614	5,614	5,614	5,614		
2002HY2	4	2,012	4,366	4,666	4,796	4,889	4,979	5,039	5,064	5,075	5,089	5,095	5,102	5,105	5,110	5,110	5,112	5,112	5,112	5,115	5,117	5,117			
2003HY1	5	1,870	4,251	4,592	4,742	4,860	4,918	4,957	4,979	4,994	5,004	5,007	5,012	5,018	5,022	5,027	5,030	5,033	5,033	5,033	5,032				
2003HY2	6	1,916	4,159	4,504	4,679	4,787	4,837	4,885	4,897	4,916	4,919	4,933	4,939	4,946	4,948	4,951	4,951	4,951	4,951	4,953					
2004HY1	7	2,008	4,348	4,669	4,829	4,911	4,965	5,024	5,051	5,058	5,068	5,073	5,080	5,082	5,084	5,085	5,091	5,092	5,093						
2004HY2	8	2,029	4,224	4,513	4,644	4,708	4,758	4,786	4,801	4,815	4,825	4,835	4,847	4,851	4,855	4,863	4,870	4,874							
2005HY1	9	2,040	4,238	4,564	4,707	4,792	4,848	4,882	4,900	4,911	4,927	4,936	4,938	4,940	4,949	4,953	4,957								
2005HY2	10	1,922	3,885	4,180	4,297	4,367	4,405	4,440	4,456	4,472	4,483	4,484	4,485	4,498	4,504	4,512									
2006HY1	11	1,848	3,974	4,279	4,398	4,469	4,537	4,575	4,599	4,614	4,624	4,631	4,643	4,657	4,661										
2006HY2	12	1,903	3,866	4,171	4,276	4,339	4,387	4,436	4,449	4,456	4,468	4,489	4,503	4,510											
2007HY1	13	1,888	4,051	4,282	4,415	4,513	4,560	4,582	4,590	4,597	4,616	4,633	4,652												
2007HY2	14	1,908	3,843	4,154	4,309	4,397	4,439	4,453	4,471	4,501	4,513	4,520													
2008HY1	15	1,769	3,807	4,109	4,239	4,322	4,345	4,387	4,429	4,450	4,471														
2008HY2	16	1,731	3,777	4,096	4,210	4,276	4,323	4,378	4,402	4,428															
2009HY1	17	1,790	4,190	4,528	4,666	4,737	4,794	4,824	4,851																
2009HY2	18	1,756	4,140	4,525	4,682	4,772	4,810	4,849																	

modelling dataset validation dataset

Table 3.1: Reported Claims Numbers - $N_{k,j}$

The data are tallied by accident year (vertical axis) and development time (horizontal axis). Each cell represents the number of claims reported, $N_{k,j}$, for a combination of k, accident period, and j, development period. For example, $N_{1,0} = 2343$ means for accident half-year "2001HY1" there were 2,343 claims reported as at the end of the accident period; $N_{1,8} = 6218$ means after 8 half years of development, there were 6,218 claims reported for that accident half-year. The increase of the data along the horizontal direction is a result of reporting delay, that is, claims from accident period k may not be reported until some time later. Since at any point in time, the older accident periods would have had a longer period of time to develop relative to the newer accident periods and this gives actuarial data the "triangle" shape. Each diagonal in the triangle represents one half year of development of the reported claim numbers for all the accident half-years to date.

This reported claim numbers data triangle also highlights the issue of censoring. For example, for accident period "2001HY1" new claims continued to be lodged 8 years after the date of the accident, since $N_{1,16} = 6284$ and $N_{1,15} = 6282$. For long tailed insurance products like CTP, it can take a long time for claims to be reported and even longer for claims to be finalised. Unless examining accident periods from a significant period of time ago (e.g., 20-30 years ago), the data would be censored as claims would still be open and their final outcomes uncertain.

The data is also represented in two colours: the black numbers in the table represent the data used for modelling and the red numbers represent the data used for validation. The validation data represent an additional three years (six development periods) of observation for all accident periods. For example, accident period 2009HY2 only has one development period in the modelling data but has six development periods in the validation data. On the other hand, accident period 2001HY1 has 17 development periods observed in the modelling data and 6 more development periods observed in the validation data; that is, 23 observations tabulated in total.

Tables 3.2 to Table 3.4 show the data for incurred claims cost, $X_{k,j}$, payments to date, $P_{k,j}$, and case estimates, $CE_{k,j}$ used in this section. A more detailed discussion on notation is provided in Section 4.3. The layout of these summaries is the same as for Table 3.1. It is worth noting that the values in each cell may relate to different numbers of claims that have been reported up to that point in time, corresponding to the $N_{k,j}$ in the Table 3.1.

Accident		development period (j) - periods after the accident period																							
period	k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
2001HY1	1	96	251	286	323	372	395	432	448	451	449	457	463	463	463	465	464	466	463	457	459	459	460	460	460
2001HY2	2	101	229	262	307	337	358	416	426	432	431	424	420	420	423	420	421	423	424	425	424	423	424	423	
2002HY1	3	100	259	315	338	373	399	428	455	462	468	470	462	460	462	456	447	447	451	454	454	452	452		
2002HY2	4	93	255	276	302	339	362	415	418	418	422	422	416	414	415	420	418	420	420	434	434	434			
2003HY1	5	88	231	268	315	336	387	430	453	447	443	425	427	423	422	425	425	430	430	431	428				
2003HY2	6	136	312	358	398	452	485	520	517	501	482	490	484	487	481	487	487	506	501	499					
2004HY1	7	148	342	374	441	489	506	550	573	570	558	549	548	539	536	538	538	540	531						
2004HY2	8	144	304	362	417	452	467	499	494	479	480	478	470	468	473	473	475	477							
2005HY1	9	136	359	414	455	487	517	530	532	509	495	485	483	491	489	487	489								
2005HY2	10	173	390	435	476	515	523	558	543	532	517	511	509	505	506	506									
2006HY1	11	186	445	492	528	537	558	578	583	544	535	540	538	531	533										
2006HY2	12	172	414	456	481	507	517	535	503	493	486	484	479	482											
2007HY1	13	251	479	502	544	546	566	553	543	526	521	513	515												
2007HY2	14	178	386	448	482	526	530	547	540	524	513	510													
2008HY1	15	170	407	448	491	501	520	535	543	521	514														
2008HY2	16	181	416	483	503	535	552	584	577	566															
2009HY1	17	174	430	474	519	537	572	587	580																
2009HY2	18	188	419	483	513	562	588	627																	

modelling dataset validation dataset

Table 3.2: Incurred Claims Cost (\$m) - $X_{k,j}$

Accident		development period (j) - periods after the accident period																							
period	k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
2001HY1	1	1	10	20	33	52	98	127	184	240	292	323	347	371	390	402	415	421	425	432	434	436	437	439	441
2001HY2	2	1	10	19	35	48	76	117	178	231	271	311	332	348	365	378	385	390	392	397	398	400	402	404	
2002HY1	3	1	10	23	37	56	82	118	174	218	267	320	355	375	393	402	409	411	418	421	424	425	432		
2002HY2	4	2	10	19	31	47	74	119	174	229	270	303	338	353	364	377	384	388	399	400	404	408			
2003HY1	5	2	10	18	34	51	81	123	177	240	288	320	354	367	371	385	388	392	398	398	401				
2003HY2	6	2	10	23	42	75	105	164	235	305	354	391	404	416	425	432	435	442	456	459					
2004HY1	7	2	11	22	43	67	115	168	247	324	391	412	438	458	468	482	484	491	506						
2004HY2	8	2	11	22	42	74	121	187	248	315	343	378	392	402	408	417	421	428							
2005HY1	9	2	13	25	47	80	120	177	255	319	358	385	409	421	438	447	451								
2005HY2	10	3	14	29	54	93	133	210	277	338	374	403	415	429	443	451									
2006HY1	11	2	13	27	57	92	144	207	279	334	373	398	414	431	442										
2006HY2	12	3	13	30	57	99	145	198	276	331	358	392	401	414											
2007HY1	13	3	16	32	62	100	148	204	267	334	376	425	451												
2007HY2	14	3	14	35	66	106	156	217	290	351	383	408													
2008HY1	15	2	14	33	62	101	152	210	288	340	380														
2008HY2	16	3	17	39	73	117	170	247	316	375															
2009HY1	17	3	18	38	73	121	181	246	328																
2009HY2	18	3	17	42	75	133	202	276																	

modelling dataset validation dataset

Table 3.3: Payments Made (\$m) - $P_{k,j}$

Accident		development period (j) - periods after the accident period																							
period	k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
2001HY1	1	95	241	266	290	321	297	305	264	211	157	134	116	91	72	64	49	45	37	25	24	23	22	21	19
2001HY2	2	100	220	244	273	289	281	299	248	200	161	114	88	72	59	42	36	34	32	29	26	23	21	19	
2002HY1	3	98	249	292	301	317	317	311	282	244	202	149	108	85	69	54	38	36	33	33	30	27	20		
2002HY2	4	91	244	257	271	291	288	296	243	189	151	119	78	60	52	42	34	32	21	34	30	26			
2003HY1	5	86	221	250	282	285	306	307	276	207	156	105	73	56	50	41	38	37	33	32	27				
2003HY2	6	134	302	335	356	377	380	356	282	196	128	99	80	71	56	56	52	64	45	39					
2004HY1	7	146	332	351	398	422	391	382	326	246	168	137	111	81	68	56	54	49	26						
2004HY2	8	142	294	341	375	377	346	313	246	164	138	100	79	66	65	56	53	49							
2005HY1	9	134	346	389	408	407	397	353	277	190	137	100	73	70	51	41	37								
2005HY2	10	171	376	406	422	423	390	348	266	194	143	109	94	76	63	54									
2006HY1	11	183	433	466	471	445	414	371	304	210	161	142	124	100	91										
2006HY2	12	170	401	425	424	408	373	337	227	162	128	91	77	68											
2007HY1	13	249	463	470	482	446	418	349	275	192	145	89	64												
2007HY2	14	175	371	413	416	419	373	330	250	173	130	102													
2008HY1	15	168	393	415	429	400	368	325	255	181	134														
2008HY2	16	178	399	444	430	418	382	337	261	191															
2009HY1	17	171	412	436	446	416	391	341	252																
2009HY2	18	185	402	441	438	429	386	352																	

modelling dataset validation dataset

Table 3.4: Case Estimates (\$m) - $CE_{k,j}$

3.3 Sample Claim Paths

Tables 3.5 and 3.6 and Figures 3.1 and 3.2 present two "sample paths" of claim development. The tables show the concatenated quarterly data and the figures provide a simple visualisation of the data. "Incurred" in the following tables and figures refers to incurred claims cost and is the estimate of the final claims cost at that point in time.

Extract Date	Claim status	Legal Rep	Liability	Case Est	Paid to Date	Incurred	Inj Code 1	Inj Code 2	Inj Code 3	Inj Code 4	Inj Code 5	Inj Sev 1	Inj Sev 2	Inj Sev 3	Inj Sev 4	Inj Sev 5	Date of Accident	Date of Claim	Date of Settlement	Gender
30/06/2002	F	N	F	0	0	0	Y02					9	.	.	.	.	9/08/2001	22/04/2002	1/05/2002	F
30/09/2002	F	N	F	0	0	0	Y02					9	.	.	.	.	9/08/2001	22/04/2002	1/05/2002	F
31/12/2002	F	N	F	0	0	0	Y02					9	.	.	.	.	9/08/2001	22/04/2002	1/05/2002	F
31/03/2003	F	N	F	0	0	0	Y02					9	.	.	.	.	9/08/2001	22/04/2002	1/05/2002	F
30/06/2003	F	N	F	0	0	0	Y02					9	.	.	.	.	9/08/2001	22/04/2002	1/05/2002	F
30/09/2003	R	Y	X	8946	554	9500	G01					1	.	.	.	.	9/08/2001	22/04/2002	1/05/2002	F
31/12/2003	R	Y	R	8946	554	9500	G01					1	.	.	.	.	9/08/2001	22/04/2002	1/05/2002	F
31/03/2004	R	Y	R	8946	554	9500	G01					1	.	.	.	.	9/08/2001	22/04/2002	1/05/2002	F
30/06/2004	R	Y	R	8578	922	9500	G01	X01				1	1	.	.	.	9/08/2001	22/04/2002	1/05/2002	F
30/09/2004	R	Y	F	7828	1672	9500	G01	X01				1	1	.	.	.	9/08/2001	22/04/2002	1/05/2002	F
31/12/2004	R	Y	F	17874	3798	21672	G01	X01				1	1	.	.	.	9/08/2001	22/04/2002	1/05/2002	F
31/03/2005	R	Y	F	17554	5118	22672	G01	X01				1	1	.	.	.	9/08/2001	22/04/2002	1/05/2002	F
30/06/2005	R	Y	F	16316	6356	22672	G01	X01				1	1	.	.	.	9/08/2001	22/04/2002	1/05/2002	F
30/09/2005	R	Y	F	16016	6656	22672	G01	X01				1	1	.	.	.	9/08/2001	22/04/2002	1/05/2002	F
31/12/2005	R	Y	F	16016	6656	22672	G01	X01				1	1	.	.	.	9/08/2001	22/04/2002	1/05/2002	F
31/03/2006	R	Y	F	25000	7656	32656	G01	I01	X01	W01		1	1	1	1	.	9/08/2001	22/04/2002	1/05/2002	F
30/06/2006	F	Y	F	0	25436	25436	G01	I01	X01	W01		1	1	1	1	.	9/08/2001	22/04/2002	15/05/2006	F

data only shown to the last date of finalisation, data is static after that

Table 3.5: Sample Claims Data 1

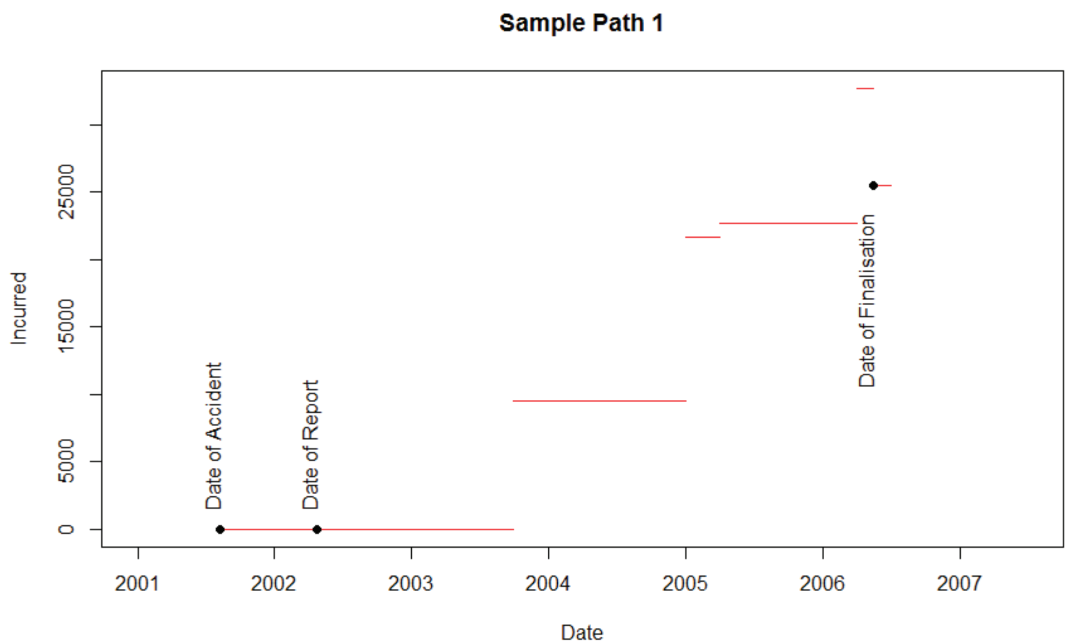

Figure 3.1: Sample Claim Development 1

For the claim shown in Table 3.5 and Figure 3.1, the accident occurred in Q4 (fourth quarter) 2001 but a claim was not lodged until Q3 2002. It then laid dormant for another six quarters; during this period the insurer believed the claim was finalized for $0, or a nil claim. At the end of 2003, the claim was "re-opened" and more information about the claim was known (including that the claimant obtained legal representation and some information regarding the injury) and immediately the estimate increased to around $10,000. About a year later, the liability status was changed from "Rejected" to "Accepted" and a further sizeable increase in the estimated claims cost was made. More than another year elapsed and the injury seemed to have worsened from affecting one body region to affecting two body regions. The claim was settled soon after with a saving, almost five years after the claim occurred.

The claim shown in Table 3.6 and Figure 3.2, however, has had very frequent revisions, almost quarterly. Yet, the main characteristics of the claim did not change throughout much of the life of the claim. Five years after the accident occurred, the claimant obtained legal representation and the claim size was revised upwards sharply. It seems the lawyer took a more active management of the claims and the claim was soon settled, at an amount, not too different to the estimate the insurer had prior to the lawyer was appointed.

Extract Date	Claim status	Legal Rep	Liability	Case Est	Paid to Date	Incurred	Inj Code 1	Inj Code 2	Inj Code 3	Inj Code 4	Inj Code 5	Inj Sev 1	Inj Sev 2	Inj Sev 3	Inj Sev 4	Inj Sev 5	Date of Accident	Date of Claim	Date of Settlement	Gender
30/06/2002	O	N	X	42303	179	42482	K11	Y02				2	9	.	.	.	20-Nov-01	22-Apr-02	.	F
30/09/2002	O	N	R	37813	624	38437	K11	Y02				2	9	.	.	.	20-Nov-01	22-Apr-02	.	F
31/12/2002	O	N	R	37813	624	38437	K11	K13	A01	A01		2	1	1	1	.	20-Nov-01	22-Apr-02	.	F
31/03/2003	O	N	F	97813	624	98437	K11	K13	A01	A01		2	1	1	1	.	20-Nov-01	22-Apr-02	.	F
30/06/2003	O	N	F	96217	2075	98292	K11	K13	A01	A01		2	1	1	1	.	20-Nov-01	22-Apr-02	.	F
30/09/2003	O	N	F	178622	2616	181238	K11	K13	A01	A01		2	1	1	1	.	20-Nov-01	22-Apr-02	.	F
31/12/2003	O	N	F	178622	2616	181238	K11	K13	A01	A01		2	1	1	1	.	20-Nov-01	22-Apr-02	.	F
31/03/2004	O	N	F	177632	3516	181148	K11	K13	A01	A01		2	1	1	1	.	20-Nov-01	22-Apr-02	.	F
30/06/2004	O	N	F	116346	4685	121031	K11	K13	A01	A01		2	1	1	1	.	20-Nov-01	22-Apr-02	.	F
30/09/2004	O	N	F	151836	4240	156076	K11	K13	A01	A01		2	1	1	1	.	20-Nov-01	22-Apr-02	.	F
31/12/2004	O	N	F	150801	5181	155982	K11	K13	A01	A01		2	1	1	1	.	20-Nov-01	22-Apr-02	.	F
31/03/2005	O	N	F	203000	5181	208181	K11	K13	A01	A01		2	1	1	1	.	20-Nov-01	22-Apr-02	.	F
30/06/2005	O	N	F	295000	5181	300181	K11	K13	A01	A01		2	1	1	1	.	20-Nov-01	22-Apr-02	.	F
30/09/2005	O	N	F	293467	6574	300041	K11	K13	A01	A01		2	1	1	1	.	20-Nov-01	22-Apr-02	.	F
31/12/2005	O	N	F	341707	8174	349881	K11	K13	A01	A01		2	1	1	1	.	20-Nov-01	22-Apr-02	.	F
31/03/2006	O	N	F	340280	9473	349753	K11	K13	A01	A01		2	1	1	1	.	20-Nov-01	22-Apr-02	.	F
30/06/2006	O	N	F	556379	17590	573969	K11	K13	A01	A01		2	1	1	1	.	20-Nov-01	22-Apr-02	.	F
30/09/2006	O	Y	F	553934	19815	573749	K11	K13	A01	A01		2	1	1	1	.	20-Nov-01	22-Apr-02	.	F
31/12/2006	O	Y	F	559533	19815	579348	K11	K13	A01	A01		2	1	1	1	.	20-Nov-01	22-Apr-02	.	F
31/03/2007	F	Y	F	0	390714	390714	K11	K13	A01	A01		2	1	1	1	.	20-Nov-01	22-Apr-02	27/02/2007	F

subset of complete data; data only shown to the last date of finalisation, data is static after tha

Table 3.6: Sample Claims Data 2

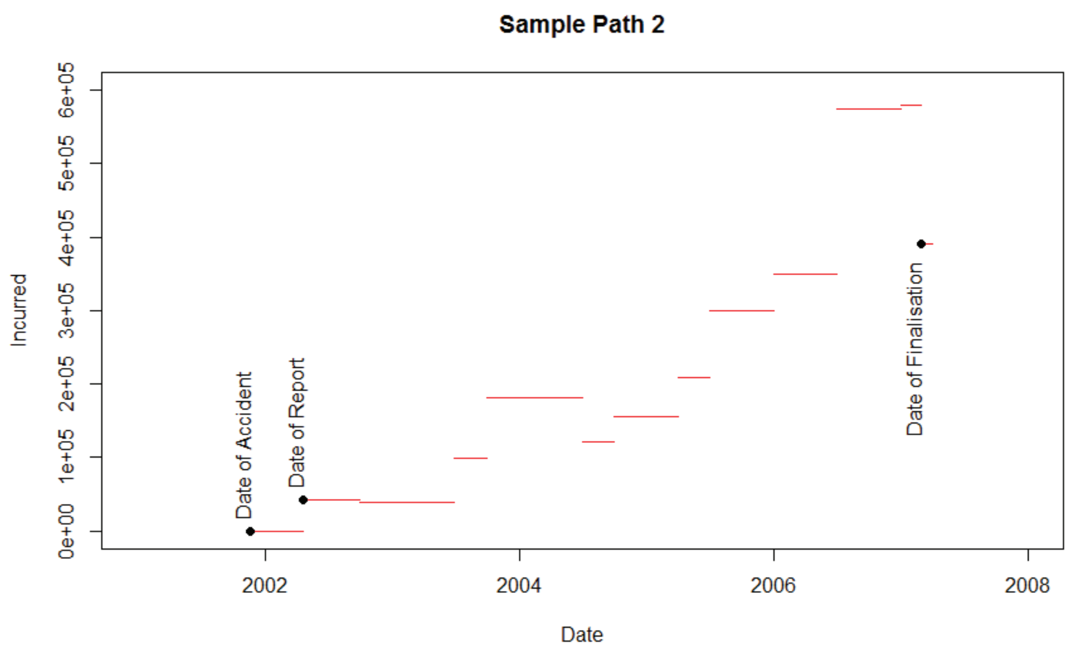

Figure 3.2: Sample Claim Development 2

The sample data and sample paths should provide a more tangible feel of the data that is been used throughout this research. In total, there are around 100,000 such claims to be analysed, each with their own trajectory and associated claim characteristics that varies over time. This NSW CTP dataset is a very complicated dataset. Further data processing is required in subsequent chapters and is documented as required.

3.4 Claim Development: Real Time vs. Number of Revisions

One of the features of the model proposed in this research is that the "duration" of a claim is not time elapsed on a real time scale, but on the number of revisions made to claims cost. This is somewhat analogous to the concept of "operational time", that is, it is an activity based concept of time rather than calender time based. For a more detailed discussion on operational time as a measurement of time passage refer to Taylor [1981].

Using the number of revisions as a measurement of time, impacts on the speed of claim management can be normalised. For example, in the NSW CTP scheme, the number of claims have reduced significantly since 1999, probably due both the introduction of the then new legislation MACA and advancement in vehicle safety technology such as stability control and air bag technologies. Amongst other causes, the reduction in the number of claims may have resulted in claims being attended to more promptly as the overall volume of claims has reduced. From the data, it is observed that the delays between revisions are shortening for the more recent years. This can be observed in the Table 3.7. It seems from the 2005 accident year onwards, the average duration between claim revisions has shortened across all revisions. As the delay between revisions shorten, it is reasonable to expect the development pattern would also hasten, such that most of the claims development would happen more promptly than before. Using an "activity" based time scale such as number of claim cost revisions would be able to better account for the more rapid claim revisions.

	Revision									
Accident Year	1	2	3	4	5	6	7	8	9	10
2001	3.5	3.0	2.7	2.4	2.4	2.3	2.3	2.2	2.2	2.3
2002	3.4	3.0	2.7	2.5	2.4	2.3	2.2	2.0	2.2	2.2
2003	3.5	3.0	2.6	2.5	2.4	2.3	2.3	2.2	2.3	2.2
2004	3.5	2.9	2.6	2.4	2.4	2.4	2.2	2.4	2.4	2.4
2005	3.3	2.9	2.6	2.4	2.3	2.2	2.0	2.2	2.3	2.0
2006	3.4	2.8	2.6	2.4	2.2	2.0	2.0	2.0	1.9	2.1
2007	3.2	2.7	2.3	2.2	2.0	2.0	2.0	2.0	1.8	1.8
2008	2.9	2.6	2.4	2.2	2.0	1.8	1.7	1.7	1.4	1.5
2009	3.2	2.5	2.1	1.9	1.8	1.7	1.5	1.4	1.3	1.2

Table 3.7: Average Delay per Revision

Secondly, using number of revisions as a time scale leads to more stable development patterns. In Tables 3.8 and 3.9, we compare the development ratios of the claims cost between the usage of development year and the usage of number of revisions as measurements of time. The former (Table 3.8) compares the incurred claims cost in one year to the previous year and the latter (Table 3.9) compares the claims incurred cost at one revision to that of the previous revision. The variability of the development factors, as measured by their standard deviation, is also shown.

Accident Year	Yearly Development Ratio Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8
2001	121%	120%	114%	101%	100%	99%	100%	100%
2002	125%	117%	116%	101%	100%	99%	99%	
2003	122%	120%	114%	97%	97%	99%		
2004	122%	116%	109%	96%	99%			
2005	120%	113%	105%	94%				
2006	116%	106%	103%					
2007	112%	106%						
2008	115%							
2009								
St. Dev.	4.3%	6.1%	5.2%	3.3%	1.4%	0.3%	1.0%	-

Table 3.8: Development Factors by Development Year

Accident Year	Development Ratio at each Revision Rev. 1	Rev. 2	Rev. 3	Rev. 4	Rev. 5	Rev. 6	Rev. 7	Rev. 8	Rev. 9
2001	114%	113%	106%	108%	109%	101%	99%	105%	102%
2002	117%	110%	110%	108%	106%	104%	98%	104%	97%
2003	121%	114%	107%	105%	102%	99%	99%	96%	109%
2004	117%	113%	107%	103%	101%	96%	101%	97%	100%
2005	110%	109%	104%	101%	100%	100%	98%	100%	94%
2006	106%	105%	100%	101%	100%	100%	95%	98%	101%
2007	102%	107%	105%	108%	97%	99%	99%	93%	97%
2008	110%	107%	104%	103%	100%	101%	93%	100%	106%
2009	112%	109%	109%	107%	105%	103%	102%	102%	100%
St. Dev.	5.8%	3.1%	2.9%	3.1%	3.6%	2.3%	2.8%	3.9%	4.7%

Table 3.9: Development Factors by Number of Revisions

While the standard deviations of the development ratios based on number of revisions seem to be smaller, the differences are minor. However, there is a subtle difference in the way the development ratios are devised. While the number of revisions based development ratios only includes claims that have had a change, the calender year based ratios would include all claims regardless whether they have had a claim revision or not. Hence, the ratios may be diluted and appear more stable than they really are. Consequently, it can be argued that the number of revisions as a time measure provides more stable development patterns in the NSW CTP portfolio.

Chapter 4

Current Approaches to Projecting Ultimate Claims Cost

4.1 Introduction

This chapter examines the current actuarial methods used in the projection of the ultimate claim costs for long tailed insurance products. The ultimate claim costs are of paramount interest to the insurer, both to assess to eventual profit or loss arising from a cohort of policies as well as to manage the cashflows required for claim payments.

Over time, actuaries have developed numerous methods for the projection of the ultimate claims cost for insurance products. Collectively, these methods are known as actuarial valuation techniques (Hart et al. [1996] and Taylor [2000]). Valuations are a legislated requirement for the management of insurance companies in developed countries. Several of the more well known and widely used techniques are discussed here. We have applied them to the NSW CTP dataset to demonstrate the methods as well as to evaluate their forecasting efficacy.

These methods can be categorised as either aggregate models or individual claims models (Kaas et al. [1988] and Taylor et al. [2008]). The former projects the ultimate claims cost for a portfolio of claims based on the underlying assumption that the development pattern for a sizeable pool of claims would be stable and historical patterns can be used to predict how claims would develop in the future. The latter attempts to provide an ultimate claims cost for each claim, and when summed over all claims in a portfolio, can also provide the ultimate claims cost for a portfolio of claims.

Aggregate models are much simpler to apply and explain to the management of insurance companies and, as a result, these models have become the current mainstream valuation techniques. With an overarching objective of maintaining their financial health insurance companies are more interested in whether they have set aside enough monies, as a total pool of funds, to pay for future claims rather than any particular individual claim. Hence, the valuation is primarily designed to ensure this total pool of funds is adequate.

However, the ability to produce claim cost estimates on an individual claim basis has certain merits. First of all, the insurance company may wish to analyse those claims that are ultimately settled for considerably higher than their current estimates; interesting insights may be learned as to why these claims cost more than originally expected, or this type of analysis may lead to fraud detection. Secondly, knowing the claims cost at an individual level may enable claims manager to deliver a better outcome, for example, by referring claims with very high projected claims cost to the appropriately experienced claim manager earlier on. Currently, the field of individual claim valuation models is in its infancy - a few models have been outlined but detailed implementation is difficult and results, as well as the processes in obtaining the results, are at times difficult to explain to the various stakeholders. This thesis contributes to this area by proposing a framework to model the process by which the claim estimates change over the duration of the claims.

Also, traditional actuarial methods only produce point estimates of the ultimate claims cost. These methods do not inherently produce any measurement of variability of the mean or "central estimate". With the professional and regulatory requirement for an actuary to examine claims cost variability (Institute of Actuaries of Australia [2013] and Australia Prudential Regulation Authority [2013]) as a part of the valuation process, a separate process based on the chain ladder framework (Wüthrich et al. [2009b] , England and Verrall [1999] and Pinheiro et al. [2003]) is typically used for the analysis of volatility regardless of the method adopted for the central point estimate. These analyses of volatility generally are not consistent with the actuarial methods themselves.

This chapter examines two mainstream aggregate actuarial valuation models and two individual claims valuation models. The aggregate models examined are the Incurred Chain-ladder Development (ICD) model and the Payments per Claim Finalised (PPCF) model. The ICD model is likely to be the most general model in its applicability and most widely used actuarial valuation method; the PPCF method is designed specifically for CTP portfolios. A third method, the Projected Case Estimate method while also popular with CTP portfolios is not discussed further in this chapter; further discussion is provided in Appendix B.2.

The individual claims valuation models discussed in this chapter are Generalised Linear Model (GLM) and Statistical Case Estimation (SCE). While the former are widely used in various fields as the de facto multivariate statistical tool to understand the relationship between the dependent and explanatory variables, its application in claims reserving, especially long tailed claims, is still relatively rare. Despite that, GLMs are the primary individual claim reserving tool currently used by actuaries for CTP reserving and is discussed in some detail in this chapter. SCE was introduced by Taylor and Campbell [2002] and it has been designed for claims with regular payments based on a set of defined claim "states". While used for CTP claims, this method is less useful for NSW CTP data, as the bulk of the claims payments are made in lump sums at finalisation.

Applications of the methods to the CTP modelling data (2001 to 2009) are shown and discussions on the pros and cons of each method are presented. The performance of each method is also assessed using the validation data (2010 to 2012). The overall effectiveness and shortcomings of these conventional methods are then discussed.

4.2 Actuarial Modelling of Long Tailed Claims

Traditionally, CTP claims are generally modelled at an aggregate level. As a part of the valuation process, the ultimate cost of a cohort of open claims is projected using historical claim development patterns. When the projected claims cost of the active claims is added to the known costs of the closed claims, the total claims cost of these claims can be determined. The issue with this approach is that claim cost is determined at an aggregate level and does not shed light on the relative costs of the insurer providing the product to different customers.

The preference to model claims on an individual basis, especially utilising the risk characteristics of the policyholder as explanatory variables, is recognised by the actuarial profession, claims managers and management of insurance companies. This detailed modelling approach by relating claims behaviour to policy characteristics has been used extensively for insurance products with a relatively high claims frequency such as home insurance or motor vehicle property damage insurance. However, its application to low claim frequency insurance products such as CTP insurance is less straight forward.

Taylor et al. [2008] has labelled these models "individual claims models" in the context of claims reserving, as opposed to aggregate level models. However, these models have their place in a pricing context as well, as they can be easily adapted to modelling how risk characteristics correlate with claims cost.

In the same paper (Taylor et al. [2008]), the authors laid the foundation work for the development in this area. Individual claims models were differentiated from aggregate claim models, and various types of covariates have been defined:

"Static covariates", "Dynamic covariates", "Time covariates" and "Unpredictable dynamic covariates". The last type is noted to be the most interesting but also the most difficult to model. To utilise in a predictive sense, to project the claim to its ultimate cost, there also needs to be a method of projecting the values of these unpredictable dynamic covariates to their ultimate values as well.

4.3 Notation Used in this Chapter

For this chapter, the following common notation is used. Let k denote the accident period in which the claim occurred; generally the accident period would be expressed in years or half years for long tailed insurance products. Half yearly accident periods are used in this chapter; this aligns with the financial reporting needs of the insurer, which is typically carried out on a half yearly basis.

Let $X_{k,i,j}$ be the reported incurred cost of the ith claim from accident period k at development time j, where j is defined as number of periods after the accident period k. Since not all claims occurring in accident period k would be reported during that period, $r_{k,i}$ is used to denote the reporting delay of the ith claim after the end of the accident period k. Consequently, $X_{k,i,j} \geq 0$ for $j \geq r_{k,i}$ and also define $X_{k,i} = X_{k,i,\infty}$ as the ultimate claims cost for the ith claim and is the variable of interest in this research.

Theoretically, development time periods do not need to be the same lengths as accident periods, e.g., development half-years can be used with accident years, however, it is more common to have them of the same length to make calculations slightly simpler. Using the same time intervals will also yield the "triangle" shaped data when claim numbers or claim incurred are tabulated by accidents periods and development periods.

We define the CTP data as with the following notation.

- $k = 1, 2, 3, ..., 18$ where k is measured in half years from the start of 2001, for example $k = 1$ refers to accident period 2001HY1, or the first half year of 2001

- $j = 0, 2, 3, ...18 - k$ since when $k = 1$ the maximum value j can take is 17
- $r = 0, 1, 2, ...18 - k$
- $i = 1, 2, ..., N_k$ where N_k is the number of claims occurred during accident period k

Since most of these actuarial techniques are aggregate methods and do not need to identify individual claims, the subscript i has been dropped to denote a summation over i. In particular, let $X_{k,j}$ be the total reported incurred cost from all claims from accident period k at development period j. Hence,

$$X_{k,j} = \sum_{i=1}^{N_{k,j}} X_{k,i,j} \tag{4.1}$$

where $N_{k,j}$ is the number of claims reported at time j for accident period k.

Similarly, Let $P_{k,i,j}$ and $CE_{k,i,j}$ denote, respectively, the amount of the claims cost that has been paid and the amount that is still outstanding from the ith claim of accident period k at development time j. Note that $X_{k,i,j} = P_{k,i,j} + CE_{k,i,j}$. Similar to $X_{k,j}$, let $P_{k,j}$ and $CE_{k,j}$ respectively be the total payments and outstanding amounts from all claims from accident period k at development time j.

We also define two time variables that will ease the notation. Let D be an upper bound for j, the development period, such that at development D, all claims from any accident periods are reported and settled. Secondly, let T be the development period that the data becomes censored; the censoring time for the modelling data is at the end of 2009 and the censoring time for the validation data is the end of 2012. Without mentioning it explicitly, T is taken to correspond to the censoring time for the modelling data, or in this case $T = 18 - k$, for $k = 1, 2, ..., 18$. T is the upper bound for j for each accident period, denoting the development period at which the claims become censored. For example, $k = 1$ corresponds to $T = 17$, that is, there are 17 half years of development after the accident period when the claims from this accident period becomes censored.

4.4 Chain-ladder Model

4.4.1 Incurred Chain Ladder

The Incurred Chain-ladder Development (ICD) Model is one of the earliest developed claim reserving methods and it is still widely used today (Hart et al. [1996]). This method has stood the test of time because it is simple to apply, easy to understand and explain and there is considerable scope to overlay the model with judgement, based on experience and knowledge.

Basically, the model hypothesises a fixed relationship between $X_{k,j}$'s for consecutive j's for all k's. Simply put, the reported claims cost from each accident year would "grow" similarly from one development period to the next period. Such development comes from two sources, IBNR (incurred but not reported) claims and IBNER (incurred but not enough reported) costs. The former refers to claims that have happened but are not yet known to the insurers and when they become known, the incurred costs for that accident period would grow; the latter refers to the change in claims cost over time, usually when new information becomes known about the claim and typically these changes are positive.

Let $\lambda_{k,j}$ be the ICD factor from development period $j-1$ to j for accident period k, defined as

$$\lambda_{k,j} = \frac{X_{k,j}}{X_{k,j-1}} \tag{4.2}$$

While we have used development periods of half years, in theory, the development periods can be of any lengths of time; but in practice, half years or quarters are used most frequently to meet the reporting requirements of insurers. For short tailed products, the insurer may wish to use quarterly or even monthly models, to capture any "seasonal" patterns the claims development may have. However, for long tailed claims, it is unclear whether modelling claims cost development more frequently than half yearly or quarterly would materially improve accuracy. Half year development periods have been used here to match the half yearly accident periods.

The ICD model postulates that the observed $\lambda_{k,j}$'s are identically and independently distributed for all k with mean λ_j and some unknown variance function. These true λ parameters can be estimated in the following fashion,

$$\hat{\lambda}_j = \frac{\sum_{k=1}^{18-j} X_{k,j}}{\sum_{k=1}^{18-j} X_{k,j-1}} \tag{4.3}$$

The upper bound of the summation is $18 - j$ for development period j; with a higher j, there are fewer accident periods that can be used to estimate λ_j. England and Verrall [2002] discusses the statistical properties of the ICD model in great detail and the reader is referred to their paper for further information.

Once the λ_j's are estimated, the projected ultimate claims cost for each accident period are calculated as

$$\hat{X}_k = X_{k,T} \prod_{j=T+1}^{D} \hat{\lambda}_j \tag{4.4}$$

where $X_{k,T}$ is the known incurred cost for accident period k as at the censoring date and $\hat{\lambda}_j$'s estimated from the data. The user can reasonably limit the calculation of the λ_j at a j when all the claims are reported and settled for an accident period, this is denoted by D in the equation above. This may be two years for most short tailed products while this may be ten or more for long tailed bodily injury products like CTP.

One disadvantage of this model is the fact that incurred claims costs are made up of payments and case estimates. While the former is factual and evidence based the latter is driven by claim management behaviour, which may change over time to adapt to the changing environment. This means, the λ's may change over time. For example, in the aim of having the most reliable estimates available, the claims manager learns from previous deviations and consciously corrects them, in this case, the ICD factors would change over time.

To overcome this disadvantage the ICD model does allow a significant level of control for the user to apply knowledge and experience through the selection of the λ's to be something other than the weighted average of the complete history. For example, it is common practice for actuaries to only use the most recent two years of data to calculate $\hat{\lambda}_j$, that is,

$$\hat{\lambda}_j = \frac{\sum_{k=max(1,15-j)}^{18-j} X_{k,j}}{\sum_{k=max(1,15-j)}^{18-j} X_{k,j-1}} \tag{4.5}$$

The summation range allows only 4 elements are summed, representing 2 years of data. The user can even select ICD factors completely ignoring the data based on other knowledge.

The ICD model is extremely easy to apply in practice. Appendix B.1 shows the ICD method applied to the NSW CTP data. The results of the modelling is show in Table 4.1.

While the accuracy of the various actuarial projection methods are compared later on, a few observations are made here. The total estimated claims cost using modelling data alone for the nine accident years is \$9,023m for the industry. The projected ultimate claims cost for the NSW CTP scheme ranges between around \$420m to \$600m per half year, hitting a low in 2002 and then steadily increasing until the 2009 accident periods. The increase in the claims cost would be needed to be passed onto the policyholders in the form of higher CTP premiums. By incorporating three extra years of development in the form of the validation data, we assume the projected ultimate claims cost is more accurate. The total projected cost over the 9 accident years is around \$9,045m. The change is only \$23m at the

Accident Period	Using 2001-2009 Modelling Data					Using 2010-2012 Validation Data			
	Number of Reported Claims	Incurred Cost to Date ($m)	Projected Ult Incurred ($m)	Proj. Ult. Number of Claims	Average Claim Size ($)	Projected Ult Incurred ($'m)	Proj. Ult. Number of Claims	Average Claim Size ($)	Over or Under Projection ($'m)
2001HY1	6,284	463	465	6,284	74,002	460	6,283	73,287	5
2001HY2	5,699	423	426	5,705	74,589	424	5,703	74,283	2
2002HY1	5,612	447	451	5,618	80,321	453	5,614	80,618	-1
2002HY2	5,110	420	421	5,115	82,292	435	5,117	85,026	-14
2003HY1	5,022	422	422	5,027	84,037	428	5,032	85,148	-6
2003HY2	4,946	487	489	4,951	98,736	498	4,953	100,541	-9
2004HY1	5,080	548	549	5,085	108,006	535	5,093	105,081	14
2004HY2	4,835	478	476	4,840	98,296	477	4,874	97,927	-2
2005HY1	4,927	495	487	4,932	98,789	496	4,957	100,018	-9
2005HY2	4,472	532	513	4,485	114,294	514	4,512	113,836	-1
2006HY1	4,599	583	548	4,624	118,578	542	4,661	116,220	7
2006HY2	4,436	535	500	4,478	111,725	491	4,510	108,922	9
2007HY1	4,560	566	551	4,642	118,596	525	4,652	112,793	26
2007HY2	4,397	526	526	4,527	116,112	518	4,520	114,601	8
2008HY1	4,239	491	511	4,445	114,895	520	4,471	116,290	-9
2008HY2	4,096	483	542	4,429	122,264	565	4,444	127,135	-23
2009HY1	4,190	430	554	4,872	113,720	561	4,891	114,726	-7
2009HY2	1,756	188	593	4,696	126,258	604	4,919	122,692	-11
Total	84,260	8,517	9,023	88,755	101,660	9,045	89,206	101,398	-23

Table 4.1: Summary of ICD Projection

total level, showing the ICD model provides very stable projection numbers.

4.4.2 Claim Number Chain Ladder

Table 4.1 also contains projected number of claims for each accident period. The projection method for the claims numbers is also the Chain Ladder method. Due to the statutory limitation of 12 months on reporting delays, with the exception of certain circumstances, the claims numbers are easier to project. The projection of the claim numbers are not shown here. However, the methodology is exactly the same as discussed above, replacing $X_{k,j}$ with $N_{k,j}$.

4.4.3 Projected Case Estimation Method

A related model to the ICD model is the Projected Case Estimates (PCE) model (Hart et al. [1996]). This model is designed to remedy two shortcomings of the ICD model.

Firstly, the development factors in the ICD model are derived from cumulative incurred costs, which incorporate both open and closed claims. Since closed claims can no longer change, they would dilute the development of the claims that are still open. The PCE method "removes" claims after they are finalised and the model only reflects the development of the open claims.

Secondly, the ICD model requires a separate, and often inconsistent, projection of claims payments. The PCE model incorporates the payments as a part of the analysis and projects payments at each future period as well as producing payments that are consistent with the incurred cost projection.

However, due to the model's similarity to ICD model the modelling process for PCE is not discussed further in this chapter. The reader is referred to Appendix B.2 for further details.

4.5 Payments per Claim Finalised in Operational Time

More recently, a group of methods loosely named "Payments per Claims methods" have come into popularity (Hart et al. [1996]). The assumption behind these methods is that claims sharing some kind of common trait would be similar in cost and the aim is to find that trait. For example, in householder property damage claims, when damage occurs the bigger claims tend to be reported promptly and the smaller and less obvious damages tend to be overlooked initially and reported later. Hence, a pattern can be established between claim size and reporting delay, which is measured from the date of the accident to the date that the claim is reported.

For CTP claims, a big determinant of claim size is the settlement delay. Many factors are correlated to that, either as a cause or as a result. A key determinant of claim size is the injury severity, that is, the more serious injuries need more time to stabilise and establish the extent of the damage and the course of treatment and rehabilitation. Secondly, the claimants with legal representation may be more litigious and may go through the court systems and possibly even appeals, which takes time and are more costly to the insurer. Thirdly, due to the phenomenon of claims inflation, the longer a claim stay opens, even though the circumstances may not have changed, it may cost more when it is eventually finalised.

Further, most CTP claims are paid in a lump sum payment at the time of claim finalisation. Although some ongoing treatment costs are paid for prior to the settlement of the claim, by far the majority of the total claims costs, including past and future economic loss, future treatment and care costs, compensation for pain and suffering, are paid after the terms of the settlement are agreed upon by the claimant and the insurer. Consequently, a connection can be established between the time taken to finalise the claim and the claim size, in particular, a "Payment per Claim Finalised" (PPCF) model is suitable for a CTP portfolio.

For the NSW CTP scheme, however, a further complication needs to be considered. In a previous section, it has been discussed that the finalisations in the scheme are becoming faster, that is, a claim that used to take three years to finalise and have a particular claim size may now only take two years to settle and have the same claim size. This means, over time, claim sizes at the same settlement delay may not be comparable, rather, the comparison needs to occur at points of time where the speed of settling claims is normalised. The concept of "operational time" has been developed. Operational times are points in real time where the same "proportion" of claims are finalised. For example, an operational time of 10% means the point of time when 10% of the claims for a particular accident period have been finalised. This means comparison of claims at the same operational time would occur between similar claims, as opposed to comparisons on a real time scale which may be misleading.

Before we apply the PPCF in operational time (PPCF(OT)) model to the NSW CTP data, we need to define some additional notation as well as discuss a further complication we face in the PPCF(OT) model.

Let j' now denote operational time in deciles, that is $j' \in (10, 20, ..., 100)$. $j' = 10$ refers to the first 10% of the claims that settled in a particular accident period. Numerically, we stratify the projected ultimate number of claims $\hat{N}_k$ into ten equally sized cells. When we sort those claims already finalised by the order of their date of finalisation, we can allocate claims into the ten cells.

Let $X_{k,i,j'}$ be the claim size of the ith claim from accident period k and finalised with operational time decile j'. The underlying assumption of the PPCF(OT) model is that claims finalised at similar durations (shares the same j') have similar claim sizes, even if from different accident periods. If this assumption is invalid then the PPCF(OT) model would not be appropriate.

However, this relationship we observe from the finalised claim sizes between claims from the same j' but from different k is complicated by claim inflation. Claim inflation is defined as the underlying rate of increase in the claim sizes over time and may stem from higher wages used in the calculation for compensation, increases in the cost of medical treatment, increases in the legal expenses, etc. Claims inflation would confuse the relationship we are trying to establish between the finalised claim sizes and we will neutralise the impact of inflation by expressing all payments in the dollar values at the censoring date. This way, the trend in the claim sizes are expressed in a common denomination and trends can be better established. We do this by indexing the claim costs at 4.3% p.a. from the date of settlement to the censoring date. The 4.3% used for the inflation of payment is the average wage inflation (for fulltime jobs without overtime earnings) observed in NSW for the period from 2001 to 2012, (Australian Bureau of Statistics [2014]). Note explicit inflation assumption is not needed for the ICD or PCE models, the ICD and CED factors implicitly allow for claims inflation when examining the amount of growth from one development period to the next.

Similar to the previous techniques, the details of applying the PPCF in OT can be found in the Appendix B.3. Projection was carried out modelling data alone as well as for modelling and validation data combined. The comparison is show in Table 4.2, we can see the PPCF(OT) method has also been quite stable, showing relatively small changes in the projected ultimate claims costs across the nine accident years, where the ultimate cost was assessed at 2009 ($8,999m) or at 2012 ($8,959m).

Accident Period	Using 2001-2009 Modelling Data					Using 2010-2012 Validation Data			
	Number of Reported Claims	Incurred Cost to Date ($m)	Projected Ult Incurred ($m)	Proj. Ult. Number of Claims	Average Claim Size ($)	Projected Ult Incurred ($m)	Proj. Ult. Number of Claims	Average Claim Size ($)	Over or Under Projection ($m)
2001HY1	6,284	463	456	6,284	72,105	453	6,283	72,545	3
2001HY2	5,699	423	417	5,705	73,081	410	5,703	71,871	7
2002HY1	5,612	447	441	5,618	78,434	443	5,614	78,857	-2
2002HY2	5,110	420	415	5,115	81,140	421	5,117	82,278	-6
2003HY1	5,022	422	420	5,027	83,472	416	5,032	82,602	4
2003HY2	4,946	487	490	4,951	98,945	488	4,953	98,625	1
2004HY1	5,080	548	524	5,085	103,016	529	5,093	103,862	-5
2004HY2	4,835	478	472	4,840	97,600	459	4,874	94,164	13
2005HY1	4,927	495	502	4,932	101,725	483	4,957	97,404	19
2005HY2	4,472	532	523	4,485	116,645	494	4,512	109,534	29
2006HY1	4,599	583	534	4,624	115,376	495	4,661	106,177	39
2006HY2	4,436	535	501	4,478	111,874	470	4,510	104,236	31
2007HY1	4,560	566	540	4,642	116,327	537	4,652	115,414	3
2007HY2	4,397	526	538	4,527	118,925	520	4,520	114,960	19
2008HY1	4,239	491	525	4,445	118,150	524	4,471	117,152	1
2008HY2	4,096	483	540	4,429	121,978	544	4,444	122,415	-4
2009HY1	4,190	430	593	4,872	121,704	624	4,891	127,625	-31
2009HY2	1,756	188	569	4,696	121,123	650	4,919	132,138	-81
Total	84,260	8,517	8,999	88,755	101,391	8,959	89,206	100,434	40

Table 4.2: Example of PPCF(OT) Model - Summary

4.6 Generalised Linear Modelling

Using GLMs to model claim sizes from an insurance portfolio has became standard practice for over 20 years since the publication of McCullagh and Nelder [1989]. However, its popularity has been generally limited to the pricing of short tailed portfolios, where the insurance premium is modelled based on policy characteristics such as age, gender, value of vehicle, etc. GLMs are extremely powerful in finding the relationships between the claim cost and these insurance “rating factors”.

4.6.1 GLM Modelling

While the use of GLMs in long tailed reserving is well documented in the literature (Taylor and McGuire [2004] and England and Verrall [2001]) and can be regarded as one of the first individual claim prediction models, it is still not widely adopted in practice. One of the main reasons is that while long tail claims can be related to claim characteristics, their claim characteristics change over time. For example, the GLM may find litigated claims tend to be very costly; however, litigation is a claim characteristic that may not be present when the claim is first reported and, instead, is developed as the claim evolves over its duration. When a valuation is carried out using GLMs at a particular censoring date, the claims to be "modelled" may include those recently reported and have not had a chance to be litigated yet. Hence, by modelling claims using only these pre-mature claim characteristics, the result would be biased Mulquiney and Actuaries [2004].

Another issue is the timing of the claims when they are incorporated into the GLM model. For any insurance product claims settled at various durations due to the varying complexities of the claim; this is more so for long tailed products. For NSW CTP, claims can settle as fast as within the first quarter after the accident or remain open after 10 years. The issue is at what duration are these claims incorporated into the GLM model. The preferred choice is when claims are finalised since the relationship of the ultimate claim cost and the claim characteristics at finalisation is the strongest. The length of time before a claim is settled is used as an explanatory factor as seen in the case of the PPCF method. This is demonstrated in Taylor and McGuire [2004] and Mulquiney and Actuaries [2004]; however, such a model is difficult to use for reserving as the ultimate claim characteristics are unknown at the reserving date as the open claims at that point of time becomes censored. The other option is to model all claims at a common duration such as at the time of report or at the end of the quarter or half year in which the claim was reported. GLM models based on claim characteristics at a common date can be used for reserving purposes since the ultimate size of a claim can be estimated once the common duration is reached. The downside to this approach is that the model is not as predictive, in the sense of explaining the variation in the ultimate claim sizes, as compared to the model based on ultimate claim characteristics.

The two modelling approaches are compared below to demonstrate the issues as well as forming a valuation estimate using the GLM approach. The theoretical framework of GLMs is not discussed here; for further details the reader is referred to McCullagh and Nelder [1989] and McGuire and Actuaries [2007]. The general structure of the two GLMs compared is listed below,

- ultimate claim sizes, X_i's, are assumed to have a gamma error distribution. Since insurance claims tend to become more expensive over time (inflation), the claim sizes have been inflated to the valuation date of December 2009 using 4.3% p.a., as discussed in Section 4.5. This is common practice to eliminate the impact of inflation when analysing claim behaviours,

- a log link function is used such that $\log \mu_i \propto \beta Z_i$, where μ_i is the mean parameter of X_i, β is the vector of coefficients and Z_i is a vector representing the claim characteristics,

- the dispersion parameter c is estimated from the data.

The variables used as explanatory factors are discussed below. They are chosen based on prior knowledge regarding the biggest drivers of claims cost in CTP claims.

- Gender - Male or Female

- Injury Severity - the maximum severity of the injuries sustained by the claimant, 0, 1, 2, ..., 6. 6 represents to a fatal injury and 0 represents an unknown injury. Many claims have an injury severity of 0 when they are first reported; however, they usually take on a different value as more information is received regarding the injury.

- Legal Representation - Y or N, whether the claimant has appointed a lawyer to act on his or her behalf
- Employment Status - Employed (by an employer or self employed) or Other (e.g. student or unemployed)
- Liability - Full, Partial or Other (which includes unknown or rejected liability statuses)
- Accident year - 2001, 2002, ..., 2009, the year the accident occurred
- Litigation level - Not litigated, Supreme Court, Local Court or CARS (CARS is a arbitration system where compensation is calculated by a team of specialist medical staff)
- Operational Time - Proportion of claims finalised chronologically prior to this claim, in bands of 5 percent

GLM Model 1 uses ultimate claim characteristics (the observed values of the explanatory variables when the claims were finalised). The model adopted in this section is simpler than the one shown in McGuire and Actuaries [2007] but it is still adequate to illustrate the modelling technique in this context. GLM Model 2, uses claim characteristics at the end of the quarter that it is notified in; and because of this, operational time is not included in this model. However, we have included the initial estimate of the claim size when the claim is reported in the model. Out of the 84,260 claims that has been lodged with the insurers at the end of 2009, 57,468 non-nil claims has been finalised and included in the GLM models. The results of the two GLMs fitted are presented in the following table.

Table 4.3 shows model parameters, estimated coefficients and their standard errors for the two models. Model 1 has considerably more terms and the coefficients make intuitive sense and are discussed briefly below.

Ultimate Claim Size Model				
Model	GLM Model 1		GLM Model 2	
	Claim Chars. at Finalisation		Claim Chars. at Report	
Number of claims	57,468		57,468	
Variables Used	38		12	
Log-likelihood	-648,981		-657,295	
AIC	1,298,041		1,314,617	
BIC	1,298,390		1,314,733	
Parameter	coeff	s.e.	coeff	s.e.
Intercept			**4.14**	0.04
acc_year = 2001	**11.75**	0.04		
acc_year = 2002	**11.78**	0.04		
acc_year = 2003	**11.88**	0.04		
acc_year = 2004	**11.88**	0.04		
acc_year = 2005	**11.96**	0.04		
acc_year = 2006	**12.13**	0.04		
acc_year = 2007	**12.21**	0.04		
acc_year = 2008	**12.29**	0.05		
acc_year = 2009	**12.64**	0.06		
optime = 5	**-2.15**	0.04		
optime = 10	**-2.12**	0.04		
optime = 15	**-2.17**	0.04		
optime = 20	**-1.93**	0.04		
optime = 25	**-1.81**	0.04		
optime = 30	**-1.63**	0.04		
optime = 35	**-1.54**	0.04		
optime = 40	**-1.35**	0.04		
optime = 45	**-1.22**	0.04		
optime = 50	**-1.06**	0.04		
optime = 55	**-0.97**	0.04		
optime = 60	**-0.84**	0.04		
optime = 65	**-0.73**	0.05		
optime = 70	**-0.66**	0.05		
optime = 75	**-0.65**	0.05		
optime = 80	**-0.52**	0.05		
optime = 85	**-0.37**	0.05		
optime = 90	**-0.25**	0.05		
optime = 95	**-0.12**	0.05		
optime = 100				
Rejected/Unknown Liability	**-0.41**	0.01	**-0.11**	0.01
Partial Liability	0.01	0.02	**-0.08**	0.03
Accepted Liability				
Minor Severity Not Represented	**-1.35**	0.01	**-0.60**	0.01
Moderate Severity	**0.33**	0.01	**-0.14**	0.02
Serious Severity	**1.31**	0.02	**0.19**	0.02
Minor Severity Represented				
Unemployed or Other	**-0.11**	0.01	**-0.16**	0.01
Employed at date of accident				
Female	**-0.08**	0.01	0.02	0.01
Male				
Claim proceeded to CARS	**0.59**	0.02	**-0.18**	0.05
Claim proceeded to Court	**1.00**	0.02	**-0.33**	0.09
Not Litigated				
Economic Loss not Compensable	**-0.33**	0.01	**0.05**	0.01
Economic Loss Compensable				
Initial Estimate at Report (log)			**0.67**	0.00
Scale	**0.78**	0.00	**0.63**	0.00

* Variables italicised represents baseline value for a categorical variable

Table 4.3: GLM Model Results

- The coefficients for accident years increases representing increasing claim costs overtime.

- The coefficients for operational times increases representing the sharp increase in the claim sizes for those claims taking longer to settle.

- Claims with rejected liability are on average smaller than those with liability fully or partially accepted.

- Claims with more severe injuries at finalisations cost considerably more than claims with a 0 or 1 maximum injury severity. Claims not legally represented is also considerable cheaper to settle.

- Claim sizes relating to female claimants are slightly, but statistically significantly, cheaper.

- Claims deemed to not have an economic loss component to the compensation are also considerably cheaper than those with an economic loss component.

Model 2 is more difficult to interpret. Firstly, it is a predictive model and hence accident years and operational times variables are not available to be used. The former is due to the fact that the model needs to be able to estimate the ultimate sizes of new claims arising from future accident years and the latter is due to the fact that finalisation time is unknown at the time the claim is reported. With the inclusion of the initial claim cost estimate in the model the coefficients of the other variables becomes how "conservative" the initial estimate is compared to the ultimate size. For example, the coefficients for injury severity suggests that for severe injuries the initial estimate is on average optimistic and the final claim sizes is relatively 21% higher; for moderate injuries the initial estimate is on average conservative and the final claim size is relatively 13% lower.

While Model 1 is insightful into the drivers of claims cost at an individual claims level, it does not lend itself to reserves projection. For example, Model 1 showed that for minor claims without legal representation, with a regression coefficient of -1.35, are around a quarter ($e^{-1.35}$) of the size of those that are represented; however, that is the claim characteristic at claim settlement. At some censoring date prior to that, an ultimately represented claim may still be unrepresented and by using a finalised claim model to predict the ultimate cost of censored claims will bias the prediction.

One approach is to create transition probabilities for legal representation variable. In the example above, if the probability for a claim becoming represented eventually is 20% then an expectation of the ultimate claim size can be the weighted average across all the possible values the claim characteristics can assume. However, this approach becomes complicated if a large number of those dynamic claim characteristics is considered.

A further issue is the usage of operational time Model 1. The results of GLM Model 1 suggest operational time is one of the key variables in determining ultimate claim sizes but this is a variable not available until the claim is actually finalised. Due the claim characteristics at settlement, which are unknown prior to settlement, are such strong drivers of ultimate claim size, we believe extensions to the GLM framework, such as hierarchical model structures, would not be effective in resolving this issue.

4.6.2 Prediction using GLMs

By using claim characteristics a common historical point of time, GLM Model 2 can be used for claim projections as these claim characteristics are available for all reported claims, including the open claims which is of the main interest for valuations. All the open claims at the end of December 2009 (modelling data) have had their ultimate claim sizes predicted using Model 2 and aggregated. The results are shown in the table below.

Accident Period	Number of Open Claims	Projected Cost Dec09 values ($m)	Projected Cost Future Values ($m)
2001HY1	590	43	48
2001HY2	494	38	42
2002HY1	445	37	42
2002HY2	420	41	47
2003HY1	442	42	48
2003HY2	501	46	54
2004HY1	482	62	72
2004HY2	489	61	72
2005HY1	578	80	94
2005HY2	677	118	139
2006HY1	910	176	199
2006HY2	1,103	210	237
2007HY1	1,590	293	330
2007HY2	1,896	320	364
2008HY1	2,274	378	434
2008HY2	2,799	434	505
2009HY1	3,280	424	501
2009HY2	1,571	179	216
Total	20,541	2,982	3,444

Table 4.4: Projections of Open Claims using GLM Model 2

In total 20,541 claims were projected costing \$2,982m. However, since the claim sizes were inflated to December 2009 values when fitted to the GLM, the predicted claims costs are also in December 2009 values. An inflation factor of 4.3% p.a. is applied to these values to reflect the actual costs of these claims when they are ultimate settled in the future. The rationale behind this selection is discussed in Section 4.5.

A key consideration when using GLMs (and other individual claim size reserving models) is the fact that these models can only project open claims. That is, the output of these models provide the user with the claim development on known claims - IBNER. Another method is needed to calculate claims that have occurred but yet unknown to the insurer - IBNR. For the purposes of calculating IBNR we have combined the IBNR claim count from the Chain Ladder projection with the yet-to-finalise claim sizes from the PPCF method.

Accident Period	Projected IBNR Number	Projected Average Size	IBNR Cost Dec09 values ($m)	IBNR Cost Future values ($m)
2001HY1	0	310,000	0	0
2001HY2	6	310,000	2	2
2002HY1	6	310,000	2	2
2002HY2	5	310,000	2	2
2003HY1	5	310,000	2	2
2003HY2	5	310,000	2	2
2004HY1	5	310,000	2	2
2004HY2	5	310,000	1	2
2005HY1	5	310,000	2	2
2005HY2	13	295,458	4	5
2006HY1	25	269,782	7	8
2006HY2	42	246,296	10	12
2007HY1	82	214,609	18	20
2007HY2	130	190,852	25	28
2008HY1	206	168,561	35	40
2008HY2	333	139,874	47	54
2009HY1	682	120,137	82	97
2009HY2	2,940	111,089	327	394
Total	4,495	125,960	566	671

Table 4.5: Calculating the Cost of IBNR Claims

Table 4.5 shows the cost of IBNR claims as at December 2009. The second column contains the expected IBNR claim numbers from the Chain Ladder model from Section 4.4. They are multiplied by the average outstanding claim sizes from the PPCF model. But since these claim sizes are also in December 2009 values, the 4.3% inflation factor has been applied to achieve the future value of these claims when they are finalised. The value of IBNR claims as at December 2009 is $671m.

The following table brings together the cost of claims already finalised, the projected cost of open claims using GLM Model 2 and the cost of IBNR claims from aggregate actuarial methods. The total cost using the GLM approach for the accident years 2001 to 2009 is around $8,640m.

Accident Period	Actual Cost of Finalised Claims ($m)	Proj. Cost of Open Claims ($m)	Proj. Cost of IBNR Claims ($m)	Total Claims Cost ($m)
2001HY1	412	48	0	461
2001HY2	382	42	2	426
2002HY1	402	42	2	446
2002HY2	365	47	2	414
2003HY1	370	48	2	419
2003HY2	408	54	2	464
2004HY1	423	72	2	497
2004HY2	368	72	2	442
2005HY1	342	94	2	438
2005HY2	316	139	5	460
2006HY1	266	199	8	472
2006HY2	186	237	12	434
2007HY1	124	330	20	474
2007HY2	85	364	28	478
2008HY1	43	434	40	517
2008HY2	22	505	54	581
2009HY1	10	501	97	608
2009HY2	1	216	394	610
Total	4,525	3,444	671	8,640

Table 4.6: Total Estimate Claims Cost using GLM

Similar to the previous aggregate methods, the modelling process is repeated using data to up December 2012, that is, including the validation data. The results of both projections are summarised in the table below. With the data up to 2012, the GLM method is producing a total cost of the 9 accidents years as $8,611m, very similar to the estimate using data up to 2009.

Accident Period	Using 2001-2009 Modelling Data: Number of Reported Claims	Incurred Cost to Date ($m)	Projected Ult Incurred ($m)	Proj. Ult. Number of Claims	Average Claim Size ($)	Using 2010-2012 Validation Data: Projected Ult Incurred ($m)	Proj. Ult. Number of Claims	Average Claim Size ($)	Over or Under Projection ($m)
2001HY1	6,284	463	461	6,284	73,297	470	6,283	74,879	-10
2001HY2	5,699	423	426	5,705	74,718	436	5,703	76,407	-10
2002HY1	5,612	447	446	5,618	79,347	461	5,614	82,159	-16
2002HY2	5,110	420	414	5,115	80,891	435	5,117	84,950	-21
2003HY1	5,022	422	419	5,027	83,415	435	5,032	86,394	-15
2003HY2	4,946	487	464	4,951	93,697	491	4,953	99,113	-27
2004HY1	5,080	548	497	5,085	97,750	534	5,093	104,913	-37
2004HY2	4,835	478	442	4,840	91,228	462	4,874	94,830	-21
2005HY1	4,927	495	438	4,932	88,763	476	4,957	96,006	-38
2005HY2	4,472	532	460	4,485	102,482	488	4,512	108,266	-29
2006HY1	4,599	583	472	4,624	102,178	492	4,661	105,472	-19
2006HY2	4,436	535	434	4,478	96,977	453	4,510	100,553	-19
2007HY1	4,560	566	474	4,642	102,025	506	4,652	108,759	-32
2007HY2	4,397	526	478	4,527	105,496	470	4,520	104,027	7
2008HY1	4,239	491	517	4,445	116,309	476	4,471	106,396	41
2008HY2	4,096	483	581	4,429	131,236	515	4,444	115,931	66
2009HY1	4,190	430	608	4,872	124,750	494	4,891	100,945	114
2009HY2	1,756	188	610	4,696	129,963	517	4,919	105,074	93
Total	84,260	8,517	8,640	88,755	97,345	8,611	89,206	96,534	28

Table 4.7: Summary of GLM Projection

While the above discusses the goodness of fit of the GLM model as a valuation tool at an overall level, it does not provide an indication on the goodness of fit at an individual claims level. This is discussed further in Section 8.5, where the GLM's accuracy of projecting ultimate claim sizes at an individual claims level is compared to that of the Claim Development Process, the framework proposed in this thesis.

4.7 Statistical Case Estimation

Statistical Case Estimate (SCE) is an individual claims reserving technique that is designed to model insurance payments related to bodily injury claims (Brookes and Prevett [2004] and Prevett and Gifford [2007]). It is based on a statistical model that finds the relationship between the claim characteristics and the periodic payments in a compensation claim. For example, an injured worker may receive two types of payments - wage replacement due to his inability to work and medical costs to treat his injuries. The wage replacement component may be modelled based on the state of the claimant ("injured") and his pre-injury earnings; while the medical costs may be modelled on the state of the claimant, age, gender and location of the claimant and the severity of the injury.

One key differentiating feature of the SCE is the prediction of dynamic variables. For workers compensation claims, the state of the claimant (such as "Undergoing Treatment", "Cannot Work", "Returned to Work at Reduced Capacity", etc) is a key explanatory variable. Transition models, such as transitional matrices, are used to project these dynamic variables on a periodic basis. Detailed discussions on the projection of dynamic variables are provided in Section 8.2.1. As a result, the projection of periodic cashflows and ultimate claims costs are carried out stochastically using simulation approaches.

Since SCE is a model for expected periodic payments for a claim, it is more suited for Workers Compensation products that have ongoing payments. It is not suitable for NSW CTP since most of the claims payments are lump sum payments made at the time the claim is finalised. However, it may be used to model the claimants that enter the previously mentioned LTCS scheme. Under this scheme, a centralised government agency (the LTCSA) cares for all catastrophically injured claimants and meets their ongoing medical and care costs. The SCE would be suited to model the claim payments to these long term claimants by modelling their claim characteristics.

As the SCE is designed for claims with periodic payments; it is not appropriate for the projection of the claims with lump sum payments and, hence, this technique is not applied to the NSW CTP data. However, the results of the SCE modelling carried out by Prevett and Gifford [2007] on the long-term claimants in the Victorian CTP scheme is discussed.

The SCE has been applied to the Transport Accident Commission's (TAC) Community Support Division claims by Prevett and Gifford, who presented the findings at the 2007 IAAust Accident Compensation Seminar. The Community

Support Division operates much like the LTCS scheme of NSW in that the TAC centrally cares for the long term claimants who has sustain extremely severe injuries. The costs of these claimants are mainly therapy and care costs.

The model incorporated 1 dynamic variable and 7 static variables. The dynamic variable is the state of the claimant with 5 levels, they are

i) Active High - High levels of therapy and care costs (costs $>$ 25,000 in the preceding 12 months)

ii) Active Low - low levels of therapy and care cost (500 $<$ costs $<$ 25,000 in the preceding 12 months)

iii) Inactive (Therapy) - no care costs and low level of therapy costs (therapy costs $>$ 50 in the preceding 12 months)

iv) Inactive - No care costs or undergoing therapy (total costs $<$ 500 in the preceding 12 months)

v) Death (not described further in the paper)

The 7 static variables relates to the level of injury and baseline factors when the claimant enters the long term care scheme, they are

i) Injury class (e.g., Quad C5, Acquired Brain Injury - Lv1, etc.)

ii) Functional code (e.g., minimum function, dependent in most tasks)

iii) Mobility Code (e.g., no use of arms/legs, some use of arms/legs, etc.)

iv) Functional Independence Measure (FIM)

v) Functional Assessment Measure (FAM)

vi) Age at injury

vii) Gender

Using the previous notation but changing the development period to 1 year rather than half years and further define $CS_{i,j}$ to be the claim status for the ith claim at development period j then

$$\hat{X}_i = P(i,T) + \sum_{j=T+1}^{D} \sum_{CS_{i,j}} (IP_{i,j}|CS_{i,j}) \times P(CS_{i,j}|CS_{i,j-1}) \tag{4.6}$$

where $P(i,T)$ is the cumulative payments made at T and $IP_{i,j}|CS_{i,j}$ is the incremental payment made in the period $j-1$ to j given claimant is in state $CS_{i,j}$. T and D take the meanings that have been defined in Section 4.3.

If the elements of $IP_{i,j}|CS_{i,j}$ and $P(CS_{i,j}|CS_{i,j-1})$ are simple then $\hat{X}_i$ may be evaluated in closed form; otherwise numerical methods need to be used.

In their modelling, the authors used GLMs to evaluate both terms and simulation methods were used for the projection, that is, by simulating a large number of potential claim trajectories and averaging them to provide an estimate of $\hat{X}_i$.

The authors have not provided the complete results of the GLM models or outstanding claims cost projections. The GLM model results for the "Active High" state are reproduced in Tables 4.8 and 4.9 for completeness. Note, the results are for long-term claimants in the Victorian CTP Scheme and the reader is referred to Prevett and Gifford [2007] for further details.

Variable	Function	Parameter Estimate	Log Odds Ratio	Signifcance (Pr > ChiSq)
Intercept		2.1888		<.0001
Development year	min((max(devyear-1,0)),7-1)	0.1704	118.6%	<.0001
Current age	min((max(currage-10,0)),25-10)	0.065	106.7%	<.0001
Injury class	Non Catastro	-2.7655	6.3%	<.0001
Injury class	Other Sev ABI	-2.4875	8.3%	<.0001
Injury class	Paraplegia - 100% disrptn of funct	-1.9375	14.4%	<.0001
Injury class	Sev ABI - 2, Mobility code gt 4	-0.8117	44.4%	0.0169
Injury class	Sev ABI - 3, Mobility code gt 5	-1.5747	20.7%	<.0001
Injury class	Sev ABI - 3, Mobility code le 5	-1.0048	36.6%	0.0008
Injury class	Sev ABI - 4, Other mobility code	-2.0766	12.5%	<.0001
Injury class	Base	0	100.0%	.

Table 4.8: Example of SCE - GLM Results of Transition from Active High to Active High (Prevett and Gifford [2007])

Variable	Function	Parameter Estimate	Multiple Effect	Significance (Pr > ChiSq)
Intercept		10.4165	33,406	<.0001
Injury class	Paraplegia - 100% disrptn of funct, Sev ABI - 4, Full use of arms & legs, Other Head - Other functional code	-0.2037	82%	<.0001
Injury class	Quad - C1 - C4	0.7103	203%	<.0001
Injury class	Quad - C7 - C8	-0.2904	75%	0.0004
Injury class	Sev ABI - 1, Sev ABI - 2, Mobility code gt 4	0.2832	133%	<.0001
Injury class	Sev ABI - 2, Mobility code le 4, Quad - C5	0.4984	165%	<.0001
Injury class	Sev ABI - 3, Mobility code gt 5	-0.1391	87%	0.0003
Injury class	Sev ABI - 4, Other mobility code	-0.4327	65%	<.0001
Injury class	Sev ABI - 5, Full use of arms & legs	-0.7126	49%	<.0001
Injury class	Other	0	100%	.
Current Age	Linear from 0 to 15	0.0498	105%	<.0001
Current Age	Linear from 40 to 60	-0.0098	99%	<.0001
Current Age	Linear from 60 to 70	-0.0215	98%	<.0001
Impairment %	< 50%	-0.1517	86%	<.0001
Impairment %	> 50%	0	100%	.

Table 4.9: Example of SCE - GLM Results of Payments of Active High State (Prevett and Gifford [2007])

The main advantage of the SCE is naturally the ability to associate periodic claim payments to claim characteristics and not just based solely on historical observed claim patterns. The observation of the strong link between claim payments and claim state allows very accurate prediction of claims payments by projecting the possible future states the claimant is in at future points of time. However, in this example the probability of transition between states is very low where over 90% of the claims do not change states from year to year. With this in mind, building a complicated model for the transition between claim states may not add commensurate value to the model.

4.8 Comparison of the Various Methods

Table 4.10 compares the various methods with a particular focus on whether they account for the features of the CTP data as discussed in the previous chapter.

Feature of Data	ICD	PPCI (OT)	GLM	SCE	CDP
	Traditional Methods				This Thesis
Use individual claim characteristics to improve valuation accuracy			Yes	Yes	Yes
Allow for dynamic covariates what changes over time				Yes	Yes
Able to predict time of claim finalisation for cashflow management	Yes[1]	Yes[1]		Yes	Yes
Allow for serial dependence structure in data				Yes	Yes
Provide claim size prediction for individual claims			Yes	Yes	Yes
Ability to use activity based time scale to allow for speed up of activities		Yes			Yes
Unified approach to understand variability of results	Yes[2]			Yes	Yes

1 predicts cashflows and not settlement of claims

2 A separate modelling process

Table 4.10: Comparison of the Current Methods to the Proposed Framework

While the aggregate methods (ICD and PPCF) can produce a valuation result they are not able to capture the more complicated features that are present in the CTP data, such as, the various claim characteristics that are recorded, the serial dependence structure and interim claim developments.

The last column in the above table shows the ability of the Claims Development Processes discussed in this thesis to allow for the various features observed in the CTP data. This may allow the practitioner to choose the appropriate technique under various circumstances.

4.9 Variability of the Central Estimate

The previous sections applied the various current mainstream actuarial techniques to the NSW CTP data; however, most of the models only produce central estimates, or point estimates, of the reserves. As discussed in Section 4.1, it is also important for actuaries to understand the variability of the estimates of the claim liabilities.

From Table 4.10, only the ICD and the SCE models have inherent measures of uncertainty. In the latter case, the Statistical Case Estimation model is essentially a numerical estimation method based on Monte-Carlo simulation; hence, by repeating the simulation a large number of iterations, an empirical distribution of the possible outcomes can be compiled and its variability measured.

For the ICD model, a suite of techniques have been developed to measure the variability of the valuation results produced. Hertig [1985] detailed a method to estimate the variability by assuming the ICD factors (λ_j's) follow lognormal distributions which are fitted using the data. The variability in the ultimate claims cost for each accident period can then be estimated, which when aggregated result in an estimate of the total variability of the claims reserve. Mack [1993] extended this without assuming the λ_j's follow a lognormal distribution which improves the useability of the method. There have been other methods of estimating variability including Bayesian and Bootstrapping methods, further details can be found in Taylor [2000] and Wüthrich et al. [2009b].

It is the current practice that, even when other valuation methods are used to produce the central estimate of the claim liabilities, for example, the PPCF(OT) model, the variability of the estimate is still based on the ICD model. This creates internal inconsistencies between the estimate of the mean and the estimate of its variability. This is particular the case for NSW CTP as the development pattern seems to be changing, increasing the variability of the ICD factors. The PPCF(OT) is designed to allow for the changing speed of claim finalisation and this is incorporated in the central estimate the method produces; however, the ICD model does not allow for the underlying change in claims management behaviour and this would artificially increase the estimated variability of the central estimate.

The CDP framework proposed in this thesis also uses Monte-Carlo simulation (see Chapter 8) similar to the SCE. Consequently, variability of the projection results can be readily calculated. Further details are contained in Section 9.5.

Chapter 5

Literature Review - Individual Claim Modelling

5.1 Individual Claim Modelling

The current practice in modelling CTP claims on an individual basis is still largely based on the Generalised Linear Models (GLM) framework. It is a well understood and proven tool in many general insurance contexts. The idea of using GLMs to model individual claims in the CTP insurance context is introduced by Taylor and McGuire [2004] where GLMs are used as an alternative claims reserving method when the more traditional methods may not appropriate. The traditional methods all assume stability in the claim behaviour which may not be true if the claim profile under consideration is changing. For example, the "tried and trusted" chain ladder models are not appropriate for CTP data at times when the incurred development pattern is unstable and exhibiting a trend. In these long tailed portfolios, the actuaries are typically required to make subjective changes to the ICD pattern to achieve a reasonable answer.

Taylor and McGuire [2004] undertook a theoretical introduction of the GLM and proceeded to construct a model based on finalised claims only. A model based on operational time, seasonal effects and finalisation quarter was constructed; interaction effects between the other covariates with operational time were also explored. Operational time denotes the order a claim was settled in, for example, if a claim was within the first 10% of claims to be settled then the operational time for that claim is in the 0% to 10% band. In this case, as well as the common practice, operational time has been measured in 10% bands. Operational time was deemed to be a major explanatory variable of claims due to the strong correlation between claim size and time required to settle. A seasonal effect was used as claims that occurred in the March quarter seemed to have a different experience, perhaps due to the holiday period. After the model was built, the paper did not detail how to use the GLM model to calculate reserving results. However, while the GLM allows the user to understand the drivers of claims cost at the individual level, considerable complexity is added to the valuation process. This would be equally true for all individual claims models, including the CDP model presented in this thesis, the amount of manual adjustments needed for the chain ladder would be inconsequential when compared to the effort needed to build and validate an individual claims model.

McGuire and Actuaries [2007] extends Taylor and McGuire [2004] by demonstrating how a GLM model can be used to incorporate legislative effects (namely the transition between MAA 1989 and MACA 1999) in the modelling of claim sizes. However, that paper also focuses on modelling finalised claims only. While such models extend the understanding of the drivers of finalised claim sizes, they do not lend themselves to be used for predictive modelling. That is, if operational time (or time of finalisation) is used in the GLM, then some method of predicting the time of finalisation for each claim would also be needed. This in itself would be a complicated endevour. Secondly, for a long tailed insurance class such as CTP, finalised claims represent only a portion of the total reported claims; and typically there is an over representation of smaller claims in the finalised pool of claims because they would be easier and quicker to settle. It is felt that modelling finalised claims only may not yield a conclusive view on the impact of injury severities as the dataset has an over representation of the less severe injuries.

Ayuso and Santolino [2008] presents a rather simple approach in estimating individual claim size. All claims were divided by injury severity and stage of reporting (initial, final or forensic report) and an average claim size were calculated assuming a lognormal distribution. The projected ultimate claim size of a claim is thus the relevant average claim size of the stratification the claim belongs to. To some extent, this model is similar to the PPCF (OT) model outlined in Section 4.5 with injury severity and stage of reporting acting as claim stratification criteria and a claim size is chosen for every stratification.

5.2 Statistical Case Estimation

Taylor and Campbell [2002] introduced the use of statistical case estimate (SCE) in the context of Workers Compensation insurance. They state that SCE have their greatest value in Workers Compensation, but may also have applications in other lines of business. The idea presented is that while a loss adjustor can place a "case estimate" based on the actual observation of the claim and personal experience, an actuary can place a "statistical case estimate" based on as many claim characteristics as available.

The SCE is most relevant in Workers Compensation, because this product provides a stream of regular income for injured workers. Also the level of the cashflow is heavily correlated with "the state" the claimant is in. In the typical Workers Compensation scenario, the states the injured worker can be in include "Injured - in receipt of compensation", "In Rehab - in receipt of compensation and medical expenses" and "Back to work - no compensation". Then the various claims characteristics such as pre-injury earning, time elapsed since injury, age and nature of injury are used to predict the likelihood of the injured worker moving between the various states. A Monte-Carlo approach can be used to calculate the likely cost for each worker. SCE, in essence, is to use the claimant characteristics to predict the elements of their transition matrix.

The Statistical Case Estimation method relies on the cashflow paid to the claimant to be highly correlated with the "state" the claimant is in; this makes SCE most applicable to income replacement products, such as workers compensation or salary continuance insurances. For other insurance products, it is more difficult to classify claims into a discrete number of "states" and have the ultimate claim outcome closely related to these claim states.

In 2004 Brookes and Prevett [2004] presented a paper where the practical application of the SCE to the NSW WorkCover Workers Compensation scheme was documented. They found that the method required 13 individual models to be built, one for each type of payment (such as treatment costs, income compensation, death benefits, etc). For each model, a choice of CART, MARS (Hastie et al. [2009]) and GLM was considered and the best model chosen, with different models chosen for different situations. They found SCE can forecast total cashflow quite well for up to 3 years and that one of the main advantages SCE has over traditional actuarial methods is that the reserves are available on an individual claims basis.

As discussed previously in 4.7, Prevett and Gifford [2007] applied the SCE to a group of claims in a CTP portfolio in 2007. The CTP portfolio is the Victorian CTP Scheme where Transport Accidents Commission, a government agency, operates a no-fault scheme. SCE is relevant and applicable in this case because payments are made on an ongoing basis rather than as lump sums as occurs in NSW or QLD. The claims of particular interest to the authors are the long term claimants, where a large part of the total claims cost is paid over a period of time at somewhat regular intervals. They find that SCE can explain up to 90% of claim size variation in the training dataset but only 32% of the variation in the validation dataset. One interesting point that the authors made was that the "stochastic dynamic variable" offered the greatest amount of information in regards to claim size, yet at the same time, most difficult to work with. The authors condensed a few of the most important dynamic variables into a single dimensional dynamic variable and built a transition model to predict the status of the claimant at each period.

In 2011 Greenfield et al. [2011] applied the principles of the SCE to New Zealand's welfare system. In its main model, the "Key Tier 1 Benefits" model, welfare recipient can be in one of five "states":

i) in receipt of Unemployment benefits

ii) in receipt of Sickness benefits

iii) in receipt of Domestic Purpose benefits (e.g., Sole Parents benefits)

iv) in receipt of Invalids' benefits

v) not in receipt of Key Tier 1 benefits

The first four are income replacement welfare benefits and recipients can only receive one of the four. The last state is a "Working" state and hence not in receipt of any of the Tier 1 Key benefits. The main model puts value on the future cost of Key Tier 1 welfare payments at an individual recipient level by modelling the likelihood of each individual moving between the five states, using the characteristics of each individual, e.g., age, gender, number of children, previous bouts on social welfare, etc. SCE is suited for the valuation of welfare payments as the payment sizes are highly correlated to the state the individual is in (there are specified payment rates for each benefit) and the probability of becoming unemployed or on disability benefits are predictable with the individual characteristics.

5.3 Claims Development Based Models

The research in this thesis would fall under a development based model, that is, the process of interest is how each individual claim changes its characteristics and incurred claims cost throughout its life. While most of the aggregate based reserving models such as the Incurred Cost Development (ICD) or Projected Case Estimation (PCE) models are development based, most of the individual claim based models are end result focused. That is, the modelling results are concerned about the ultimate cost of the claim, rather than modelling the changes a claim undergoes to reach that ultimate cost.

Wüthrich et al. [2009a] looks at claims development from a different angle. Insurers would need to set monies aside to pay for the ultimate cost of the claims from the policies they have insured. The pool of money set aside is called the claims liability. When the information changes through the course of the next accounting period, and the view of the ultimate cost changes, then based on the new information, the claims liability can be more or less than those set aside at the previous accounting period. Of course, this has a profit impact and may threaten the solvency of the insurance company if they need to contribute significantly more to the claims liability pool of funds.

Beside claims movement, there is another source of movements within a pool of claims, namely, the introduction of new claims. This can happen when the claim is reported late even thought it has already happened; they are known as Incurred But Not Reported (IBNR) claims. Schiegl [2010] approaches this by introducing a 3-D triangle, where accident period, reporting periods and development periods form the three axes of the claims "pyramid". Using this approach the modelling of claims development (IBNER) and IBNR can be separated.

Taylor et al. [2008] also discusses another class of models where the variable to be modelled is not the size of a finalised claim but the ultimate size to case estimate ratio at a particular stage of development. These types of model have been label as individual claims model conditioned on case estimates. These models would utilise the extra information contained in the case estimates of the claims and the ultimate size to case estimate ratio is also less volatile for companies with mature claims management departments.

5.4 An Individual Claims Reserving Model by Larsen (2007)

Throughout the author's literature review, "An Individual Claims Reserving Model" by Larsen [2007] is found to be the most similar to the author's research in terms of objectives and shares some of the modelling methodologies adopted. Despite the similarities, there are fundamental differences between the two approaches regarding model specification.

Larsen's ultimate aim is to project the claims liabilities reserve of an insurance company. While not explicitly mentioned, the framework is able to model IBNR (incurred but not reported) and IBNER (reported but not settled, hence, subject to further development) as well as Future Claims Liability (claims that have not yet occurred, but once they occur, the insurer is liable for). To this extent, the research in this thesis only concentrates on the modelling of IBNER claims, or the development of claims cost with the arrival of new information regarding the claim.

Larsen segments all claims into (i, m, j, g) categories, where i is the accident year, m is the season/quarter within the accident year, j is the reporting delay and g is a vector denoting baseline claim variables, such as business type or claim type. He defines $Z_\imath$, the complete claims process for claim $\imath$, as a Marked Poisson Process, formulated by, dropping the subscripts i and m,

$$Z_\imath = (J, Y_{t=J}, Y_{t=J+1}, ..., Y_{t=D}, U_{t=J}, U_{t=J+1}, ..., U_{t=D}, G) \tag{5.1}$$

where Y_t is the incremental claim size development that occurred at time t. J is defined as the reporting delay or the time elapsed between the event that give rise to the claim and the time the claim was lodged with the insurance company; as such, the first observed Y_t is Y_J. U_t marks the claims open/closed status as a binary variable, G is a set of claim characteristics and D is an arbitrary future time chosen such that all claims have been settled. When segmented by (i, m, j, g), the $Z_\imath$s are independent and identically distributed.

For the purposes of reserving, the outstanding claims reserve required in excess of incurred to date claims cost, at time V is

$$R_V = \sum_{i \leq V, V < k \leq D} Y_k \tag{5.2}$$

If the above summation is broken up into two segments over j, one with $j \leq V - i$ and the other with $j > V - i$ then IBNER and IBNR are accounted for more explicitly.

Since Larsen assumes $Z_\imath$'s are independent and identically distributed with respect to (i, m, j, g), he considers the processes that stratifies claims to (i, m, j, g) categories. With similar rationale as the CDP model developed in this thesis to decompose a complex process into simpler component process, Larsen decomposes the complete process into four component processes. The four decomposed processes are

i) The process by which claim $\imath$ arises in accident year i and season m - the number of claims from accident year i and season m is assumed to have a Poisson distribution with intensity $w = w_i \sigma_m$, where w_i is the claim intensity for accident year i and σ_m is the seasonal adjustment factor for season m.

ii) The process by which claim $\imath$ will adopt characters g - the probability that a claim would have characteristics, $G = g$, follows $P(G_\imath = g|i, m) = e_{i,m,g}/e_{i,m} f(i, m, g)$, where $e_{i,m}$ is the exposure (number of accident policies) in accident year i and season m and $e_{i,m,g}$ denotes the exposure with characteristics $G = g$. That is, the probability of a claim with claim characteristics g is proportional to the relative number of active policies having these characteristics. $f(i, m, g)$ is a claim frequency "adjustment factor" for policies with characteristics g as some policies are more likely to give rise to claims than other policies with different characteristics.

iii) The process by which claim ı will have a reporting delay of j - given characteristic g, the reporting pattern follows a function of j and g and is independent of the accident year i, i.e., $P(J = j|G = g, i, m) = r'(j, g)$ where $r'()$ is a non-negative function.

iv) Given (i, m, j, g), the incremental amounts, $(Y_J, Y_{J+1}, ..., Y_D)$ are assumed to follow a Markov process with respect to S_{k-1}, defined as $S_{k-1} = Y_{k-1} + Y_{k-2} + ... + Y_j$. With this assumption, $P(Y_k|Y_{k-1}, Y_{k-2}, ..., Y_j) = P(Y_k|S_{k-1})$ and consequently $P(Y_J, Y_{J+1}, ..., Y_D) = P(Y_D|S_{D-1})P(Y_{D-1}|S_{D-2})...P(Y_{J+1}|S_J)$. This allows the Y_k's to be modelled independently.

Larsen further suggests that this specification can easily be modelled using the GLM framework. By combining items i to iii, the result is that the number of claims in the segment (i, m, j, g) follows a Poisson distribution with mean

$$\lambda_{i,m,j,g} = e_{i,m,g}/e_{i,m} w_i \sigma_m f(i, m, g) r'(j, g) \tag{5.3}$$

In this parameterisation, and using appropriate functional forms of f and r' the parameters are easily estimated using the GLM framework with a log link, exposure as an offset and Poisson error structure.

Tackling item iv, the conditional distributions of the Y_k's, are more complex. Larsen again proposes a "divide-and-conquer" approach for the modelling of the incremental changes, where different models are adopted for Y_k's under different circumstances (determined by four binary processes). While Larsen primarily uses regression models with a gamma error structure to model the Y_k's, he has found the very large incremental changes follow a distribution with thicker tails and the

Pareto distribution is adopted.

Four binary processes are used to determine the circumstances and their probabilities are defined as

- $p_{Y=0} = P(Y_k = 0|Y_k \leq 0, S_{k-1})$
- $p_{Y>0} = P(Y_k > 0|S_{k-1})$
- $p_{Y>L} = P(Y_k > L|Y_k > 0, S_{k-1})$
- $p_{S>0} = P(S_k > 0|S_{k-1}) = P(-Y_k < S_k|Y_k \leq 0, S_{k-1})$

Using the above binary process, Larsen defined the situations that the process may experience at each time period and how the resulting change is modelled,

i) No movement at time k, $P(Y_k = 0|S_{k-1}) = p_{Y=0}(1-p_{Y>0})$. Note, two binary probabilities are used here, as the former denotes no movement and the latter denotes positive movement.

ii) A small positive movement at time k, $P(0 \leq Y_k < L|S_{k-1}) = p_{Y>0}(1-p_{Y>L})$, the distribution of Y_k is proposed to be a gamma distribution.

iii) A large positive movement at time k, $P(Y_k \geq L|S_{k-1}) = p_{Y>0}p_{Y>L}$, the distribution of Y_k is proposed to be a generalised Pareto distribution.

iv) A negative movement at time k not enough to make it a nil claim, $P(0 \leq -Y_k < S_{k-1}|S_{k-1}) = (1 - p_{Y=0})(1 - p_{Y>0})p_{S>0}$, the distribution of $(Y_k + S_{k-1})/S_{k-1}$ is proposed to be a gamma distribution.

v) A negative movement at time k enough to make it a nil claim, $P(Y_k = -S_{k-1}|S_{k-1}) = (1 - p_{Y=0})(1 - p_{Y>0})(1 - p_{S>0})$.

Larsen models the probabilities of the four binary processes $(p_{Y=0}, p_{Y>0}, p_{Y>L}, p_{S>0})$ with logistic regressions. The form of a negative movement, using the ratio of the change $(Y_k + S_{k-1})/S_{k-1}$ rather than the absolute value of the change, is adopted to ensure the magnitude of the change is less than S_{k-1}.

Larson proceeds to apply the framework on a Marine Insurance portfolio; while some sensible results were produce, Larsen did not carry out a validation of these results.

The CDP model framework proposed in this thesis bears some similarities to Larsen's approach to modelling individual claims but differs in several key aspects.

Firstly, our starting point is to understand claims development rather than strictly reserving. While the individual ultimate claim size projections, when aggregated, will result in the outstanding claims of all reported claims; the valuation of IBNR is not considered in this thesis. Larsen models IBNR claims by firstly generate the number of claims reported after the valuation date using Equation (5.3) and then generate the Z_1 for each and every IBNR claim.

Secondly, the time scale adopted in this research is the arrival of new information necessitating a change in the claims cost. Hence, nil changes are not modelled, instead, a delay process is used to count the number of time periods between the arrival of new information.

Thirdly, a binary process is used for the finalisation of the claim, i.e., to determine when the claim ceases further development. It is unclear from Larsen's paper how the U_t variables are used or modelled, and it seems all the Y_t's are modelled until time D rather than stopping once the claim is finalised.

Fourthly, since the CDP model uses past observed processes outcomes explicitly in the parameterisations, it is classified as an observation driven model. Larsen uses only the latest total incurred in his modelling of claim movements. The results of applying our CDP model to CTP data have shown how the claim moved in the past can influence how it will move in the future, and including past observation values may bring additional explanatory powers to the model.

5.5 Observation Driven Models

The inspiration of this research is based on the author's past research in the area of Observation Driven Models, Wang [2004]. In Cox et al. [1981] observation driven models were differentiated from parameter driven models - while the state equations for both classes of models are the same, the past observations explicitly enter the equations of the former. The observation driven models have the advantage of having likelihoods that are conditionally specified and hence easily calculated. However, stability of the model through time can be a problem for long series. In the CTP case, the series are quite short and stability is not at issue.

Davis et al. [2003] provides an in depth discussion on the two types of models for the Poisson count model, and provides the underlying mathematics of the observation driven case as well as a discussion of the properties of the model. Rydberg and Shephard [2003] applied the model to trade-by-trade share prices using a Generalized Linear Autoregressive Moving Average (GLARMA) model. The stock prices were first decomposed into three time series: the A series denoting the activity of the share price (where price moved at a trade), the D series denoting the direction of the movement and the S series denoting the size of the movement. The decomposition is beneficial as a complicated time-price process is decomposed into three sequentially conditioned models. These component models are relatively simple processes - two binary processes and a positive integer process, the latter is assumed to be a negative binomial.

Of particular interest is also the time scale used for these models. Traditionally, a process is observed at equally spaced time and an observation is recorded. This is certainly true for the actuarial modelling described in the next sections, whether the data is recorded monthly, quarterly or half-yearly. However, the observation driven models adopt an activity based time scale, that is, an observation of the process is recorded when something happens. In the example of Rydberg and Shephard [2003] each observation of the share price is made only when a trade of the stock occurs. Equally spaced observations of the share price are not made, for example, based on the most recent trade at the time of the observation. This is relevant to this research due to the fact that each change in the claims data can be thought of as the arrival of additional new information regarding the claim, and this can drive the incurred cost of the claim up or down. However, to be useful in the financial industry, the time of the changes are also of importance, as financial accounts need to be prepared half yearly or, at the minimum, yearly.

The CDP model presented in the next chapter uses the "decomposition" approach of Rydberg and Shephard [2003] but there are some important differences. Rydberg and Shephard [2003] are concerned with modelling a single, usually, long time series (i.e., with 100's or 1000's of events). For our application, there are around 100,000 very short "time series" (trajectories of the claims developments) that are to be modelled as an ensemble. Hence, issues of model stability through time, while a definite concern for Rydberg and Shephard's application, is not a concern for us.

As in Rydberg and Shephard [2003], the CDP model uses prior observations in each component process as regressors in future mean equations and it decomposes a complex process into a collection of time evolving conditional distributions which collectively model the current state of the claim given past history of the claim and any covariates associated with each claim. The CDP framework is described in detail in Chapter 6.

Chapter 6

The Claim Development Process

6.1 Introduction

The CDP model extends the framework of Larsen [2007] in a number of new directions. Firstly, with the exception of claim size $X_{j,i}$ and settlement indicator $U_{j,i}$, Larsen only considers "static" information as per the definition adopted by Taylor et al. [2008] which does not change throughout the duration of the claim. The CDP model, however, introduces dynamic covariate information in the modelling where these variables do change over the life time of the claim. Examples of dynamic covariates in the NSW CTP data include "legal representation", the level of rehabilitation deemed necessary for the claimant by the medical professionals and the severity of the injuries.

To allow for time varying information, the information set, $G_{j,i}$, is now indexed by claim i and time j, and the mark used in Larsen [2007] is generalised to the following form

$$Z_i = (X_{J,i}, X_{J+1,i}, X_{J+2,i}, ..., X_{D,i}, G_{J,i}, G_{J+1,i}, G_{J+2,i}, ..., G_{D,i}) \tag{6.1}$$

where $G_{j,i}$ is a set containing the information of the claim i up to time j, including the settlement indicator. As $G_{j,i}$ contains both dynamic and static information, the G_i as Larsen defined it can now be considered as a subset of $G_{j,i}$. In addition, T_i and J_i are now also incorporated into it, as the accident year and reporting delay can be interpreted as a part of the information set that describes the claim even though they do not change during the lifetime of the claim. While Larsen used incremental claim cost movements, $Y_{J+k,i}$, we have adopted the incurred claims cost, $X_{J+k,i}$, in the above formulation. This two formulations are equivalent with the substitution $X_{J+k,i} = \sum_{j=J}^{J+k} Y_{j,i}$.

Secondly, both the Larsen approach and the Statistical Case Estimation method discussed previously assume the claim process is a Markov process. In the case of Larsen, $X_{j,i}$ is used for the projection of $Y_{j+1,i}$ and previous claim size values of $X_{j-1,i}, X_{j-2,i}, \ldots$ are not considered. In the case of SCE, the current state of the worker ("Active-High","Active-Low", "Inactive", etc) along with other claims information defines the future behaviour of the claim. This research, however, incorporates the past values of the processes; that is, the behaviour of the claim development processes is conditioned on all available past data. This is discussed further below.

Thirdly, the CDP model proposes the usage of another time scale. Most of the current methodologies discussed in Chapter 4 and the literature reviewed in Chapter 5 adopt a calender time based time scale. However, for long tailed claims, the claim cost estimates are revised when new information is received regarding the claim. The speeding up or slowing down of claims managers' activities can influence the speed at which new information is received. The time scale used in this research is a counter of the number of times the claim has undergone a claim cost estimate revision. This time scale is "activity" based and can be thought as analogous to the operational time that has been discussed previously.

Fourthly, the CDP model takes inspiration from Rydberg and Shephard [2003], and developed in Wang [2004] to decompose a complex process into simpler component processes to ease the modelling complexity. This is a "divide-and-conquer" approach. This is further discussed in Section 6.3.

In this chapter the Claim Development Process model framework is discussed in detail. The key differences to the model developed by Larsen [2007] and the rationale behind them are discussed in this section. The next section provides an overview of the notation used in this and subsequent chapters. This is then followed by an in-depth discussion of the overall Claim Development Process, as well as the rationale to decompose it into component processes. Each component process is then discussed in turn as well as the modelling techniques adopted in the application of the CDP to the NSW CTP data.

6.2 Notation

The notation for this section differs slightly to Chapter 4 as we move away from the traditional reserving setting. The accident year variable k and development year j are no longer critical in the modelling of claims cost. In fact, this section argues that the number of revisions made to the claims estimate, driven by the arrival of new information regarding a claim, is a better time scale than the traditional development period. Hence, from this chapter onwards, the subscript j takes the definition of a different time scale - the number of revisions a claims has had rather than the calender time period represented by development period.

Let $T_{j,i}$ for $j = 0, 1, 2, ..., m_i$ be the (calender) time of the jth revision of the ith claim; where T_0, i is the time when the claim is reported and m_i is the total number of revisions for claim i. Also, let T_i denote the date of the accident on which the claim arose and further $t_{0,i} = T_{0,i} - T_i$ is the reporting delay - the period of time after the accident before the claim is reported to the insurer. $t_{0,i}$ is taken as covariate in the CDP model as this is a piece of information known when the claim is reported at $T_{0,i}$. More generally, $t_{j,i} = T_{j,i} - T_{j-1,i}$ is the length of time elapsed since the last revision until the jth revision.

Using j in the same fashion, define $X_{j,i}$ for $j = 0, 1, ..., m_i$ to be the incurred cost for claim i at its jth revision. Consistent with the notation used in Chapter 4, incurred cost, or claim size, refers to the amounts that has already been paid on the claim plus an estimate by the claims manager of future payments to be made. $X_{0,i}$ takes the similar meaning of the claim size when the claim is reported, at $T_{0,i}$, and let $X_{m_i,i}$, or simply X_i, be the ultimate claim size for claim i. The process $X_{j,i}$ is non-negative and maps the claim cost over its life.

Let n be the number of claims in the portfolio, hence, $i = 1, 2, ..., n$.

We also define some notation to deal with the censored nature of long tailed insurance claims data. Let T' be the calender time of the censoring time, for insurance liability valuations, this is the valuation date, e.g., 31 December 2009 in the case of our modelling data. Let j'_i be the largest j for the ith claim such that $T_{j,i}$ is less than T', in other words, the revision immediately preceding the censoring date, T'. Therefore, $X_{j',i}$ is the incurred cost of claim i at censoring date T'; and $\sum_i X_{j',i}$ is the total incurred cost of all the claims as at the censor date T'. In particular, an insurance company would be most interested to know the value of the difference $\sum_i X_i - \sum_i X_{j',i}$, which is by how much of the ultimate claims cost of a group of claims exceeds the currently incurred claims cost.

6.3 Decomposition into Component Processes

Using the notation introduced in Section 6.2 and allowing for the changes outlined in Section 6.1 in Equation (5.1), the complete claims process for all revisions to finalisation appear as

$$Z_i = (X_{0,i}, X_{1,i}, ..., X_{m,i}, T_{0,i}, T_{1,i}, ..., T_{m,i}, G_{0,i}, G_{1,i}, ..., G_{m,i}) \tag{6.2}$$

Equation (5.1) and Equation (6.2) above contain the same amount of information since the $U_{j,i}$, claim settlement information, is embedded in the time of the last revision, $T_{m,i}$; that is, $U_{j,i} = 0$ for $T < T_{m,i}$.

Omitting the claims information set, G_i, the complete claims incurred cost process for the ith claim, is now contained in the pair $(X_{j,i}, T_{j,i})$ for $j = 0, 1, ..., m_i$. While the former marks the claims cost at each revision the latter marks the times that the revisions occurred.

The joint distribution of the pair of processes $(X_{j,i}, T_{j,i})$ would be difficult to model. Instead, we have decomposed $X_{j,i}$ into two component processes using the equation $X_{j,i} = X_{j-1,i} * e^{(2D_{j,i}-1)Y_{j,i}}$, where $D_{j,i}$ denotes the direction of the jth claim revision and $Y_{j,i}$ is the magnitude of the jth revision. We also introduced the $S_{j,i}$ process that monitors the open/finalised status of the claim and $t_{j,i}$ denoting the delay from the $j-1$th revision to the jth revision. These four processes are more formally defined as,

- The Delay Component $(t_{1,i}, t_{2,i}, ..., t_{m,i})$ - this component process measures the time between consecutive changes to the incurred cost, $T_{j,i} - T_{j-1,i}$, and provides a time dimension to the incurred costs values. This process takes on strictly positive values and its unit of measurement depends on how the data is constructed.

- The Claim Status Component $(S_{1,i}, S_{2,i}, ..., S_{m,i})$ - this component process measures whether the current revision would be the final revision. The process will take on values of either 0 or 1. If a value of 1 is observed then the claim process terminates and the current change would be the last, i.e., $\min j$ such that $S_{j,i} = 1$ and $m_i = j$.

- The Direction Component $(D_{1,i}, D_{2,i}, ..., D_{m,i})$ - this component process measures whether the current change is a positive change or a negative change. The process can take on values of either 0 or 1, such that, $D_{j,i} = 1$ if $X_{j,i} > X_{j-1,i}$ and $D_{j,i} = 0$ if $X_{j,i} < X_{j-1,i}$. Under the definition of a revision used in the CDP framework $X_{j,i}$ cannot equal to $X_{j-1,i}$.

- The Size Component $(Y_{1,i}, Y_{2,i}, ..., Y_{m,i})$ - this component process, defined as $|\log \frac{X_{j,i}}{X_{j-1,i}}|$ measures the magnitude of the current claim revision. The process can only be positive real numbers.

Using the above decomposition, the joint processes $(X_{j,i}, T_{j,i})$ can now be expressed as $(t_{j,i}, S_{j,i}, D_{j,i}, Y_{j,i})$ and the claim process Z_i defined by Larsen in Equation (5.1) can be expressed as

$$Z_i = (X_{0,i}, t_{1,i}, ..., t_{m,i}, D_{1,i}, ..., D_{m,i}, Y_{1,i}, ..., Y_{m,i}, S_{1,i}, ..., S_{m,i}, G_{0,i}, G_{1,i}, ..., G_{m,i}) \tag{6.3}$$

where $j = 0$ denotes the time the claim is reported to the insurer and m, serving as a similar function to D used by Larsen, denotes the last revision to the claim such that $S_{m,i} = 1$, that is the ith claim is finalised on the mth revision.

The four component processes observed $(t_{j,i}, S_{j,i}, D_{j,i}, Y_{j,i})$ can be interpreted as one joint event and may be modelled together with a joint distribution and $G_{j,i}$ is a set of exogenous claim information available at the jth revision. However, dealing with joint distributions would add significant complexity to the model. Conditioning is introduced between the processes and covariates, effectively creating a hierarchy of the order in which the component processes occur. The conditioning is explicitly specified in the following fashion.

The hierarchical order of the claim processes is:

i) The delay component, this process "starts" straight after the last revision and its value will determine when on a real time scale when is the next revision taking place. This process is conditioned on only the prior information.

ii) The claim status component, it is assumed that this process denotes whether a claim is finalised at the current revision. This process is conditioned on prior information and the current value of the delay process.

iii) Once the claim status is determined, the next component is the direction of the revision. This process is conditioned on prior information as well as the current values of the delay and claim status processes.

iv) Lastly, the claim size component; this process is conditioned on past history as well as the current values of the preceding processes.

The above hierarchy was developed after discussions with CTP claim managers. An understanding was gained on the string of events that occurs requiring the claims manager to make a revision to the claim estimate. Typically, information that causes the claim estimate to be updated comes to the claim manager from the claimant or their lawyer. Then the claims manager assesses if the information would allow the claim to be settled. The claims manager will then proceed to evaluate the information as favourable or unfavorable to their company and the magnitude of the adjustment required to be made to the claim estimate. The process is adequately reflected in the hierarchical structure imposed in the framework above.

Defining $F_{j,i}$ as the past values of all the component process variables ($t_{j,i}$,$S_{j,i}$,$D_{j,i}$ and $Y_{j,i}$) up to the ith revision, then the joint processes $(t_{j,i}, S_{j,i}, D_{j,i}, Y_{j,i})$ can be expressed as four conditionally independent processes,

i) $t_{j,i}|F_{j-1,i}$

ii) $S_{j,i}|F_{j-1,i}, t_{j,i}$

iii) $D_{j,i}|F_{j-1,i}, t_{j,i}, S_{j,i}$

iv) $Y_{j,i}|F_{j-1,i}, t_{j,i}, S_{j,i}, D_{j,i}$

such that

$$\begin{aligned}
&P(t_{j,i} = t, S_{j,i} = s, D_{j,i} = d, Y_{j,i} = y|F_{j-1,i}) \\
=&P(t_{j,i} = t|F_{j-1,i})P(S_{j,i} = s|F_{j-1,i}, t_{j,i})P(D_{j,i} = d|F_{j-1,i}, t_{j,i}, S_{j,i}) \\
&P(Y_{j,i} = y|F_{j-1,i}, t_{j,i}, S_{j,i}, D_{j,i})
\end{aligned} \tag{6.4}$$

Introducing the other claims characteristics, let $G_{j,i}$ be the history of all the other covariates, both dynamic and static, up to the jth revision. Then specifically,

i) $t_{j,i}$ is now conditioned on $F_{j-1,i}$ and $G_{j-1,i}$

ii) $S_{j,i}$ is now conditioned on $t_{j,i}$, $F_{j-1,i}$ and $G_{j,i}$

iii) $D_{j,i}$ is now conditioned on $S_{j,i}$, $t_{j,i}$, $F_{j-1,i}$ and $G_{j,i}$

iv) $Y_{j,i}$ is now conditioned on $D_{j,i}$, $S_{j,i}$, $t_{j,i}$, $F_{j-1,i}$ and $G_{j,i}$.

Note, only the delay process will be reliant on the $G_{j-1,i}$ filter, as it will be conditioned only on the information available at the previous change. However, for the other component processes, the new information, $G_{j,i}$, is assumed to have become available at that point in time.

Hence forth, we will drop the conditioning in the notation to make it more concise, but the conditioning apply in all further discussion of the component processes. The complete claim process is now specified by $\{t_{j,i}$, $S_{j,i}$, $D_{j,i}$, $Y_{j,i}$, $G_{j,i}$ for $j = 1, 2, ..., m_i\}$. Revision $j = 0$ is taken to mean reporting of the claim, which is assumed be to exogenous to the claim revisions processes. We have taken the initial claims estimate, X_0, and the claim reporting delay, t_0, as exogenous to the CDP and are variables contained in the claim characteristics, $G_{j,i}$, as static covariates.

6.3.1 Likelihoods and Conditional Distributions

The likelihood for the claim development process would be very complex without the conditional hierarchy specified earlier. Using the conditioning specified above, the likelihood of the complete system of processes can be constructed in a sequential and time recursive fashion using products of conditional distributions. The sequence of events in the CDP framework is,

i) exogenous information is received regarding the claim when it is first reported, this is the information contained in $G_{0,i}$, in which, $X_{0,i}$ is also known.

ii) next item in the process is the time elapsed until the 1st revision, that is, $t_{1,i}$; at which time, new information $G_{1,i}$ becomes available.

iii) with this new information, the remaining three claims development processes is determined in succession - $S_{1,i}$, whether the new information allows the claim to be finalised; $D_{1,i}$, where the information is favourable or unfavourable; and $Y_{1,i}$, how much impact is the new information has on the claims estimate. Each of the three variables are conditioned on the information available to that point (Section 6.3)

iv) steps ii and iii are then repeated iteratively until the m_ith revision, at which time, $G_{m,i}$ does allow the claim to be finalised.

Hence, the likelihood for the jth revision of the ith claim can be expressed as (dropping $F_{j-1,i}$ from the conditioning,

$$L_{j,i} = P(t_{j,i}|G_{j-1,i})P(S_{j,i}|G_{j-1},i,t_{j,i})P(D_j|G_{j-1},t_{j,i},S_{j,i})f(y_{j,i}|G_{j-1,i},t_{j,i},S_{j,i},D_{j,i}) \tag{6.5}$$

and that the complete likelihood for the ith claim is

$$L_i = \prod_{j=1}^{m_i}(P(t_{j,i}|G_{j-1,i})P(S_{j,i}|G_{j-1},i,t_{j,i})P(D_j|G_{j-1},t_{j,i},S_{j,i})f(y_{j,i}|G_{j-1,i},t_{j,i},S_{j,i},D_{j,i})) \tag{6.6}$$

where m_i is the revision the claim is finalised on, that is, $S_{m_i,i} = 1$.

With the conditional distributions defined above, the likelihood for the ith claim can be segregated as

$$\begin{aligned} L_i = &\prod_{j=1}^{m_i}(P(t_{j,i}|G_{j-1,i}))\prod_{j=1}^{m_i}(P(S_{j,i}|G_{j-1},i,t_{j,i}))\prod_{j=1}^{m_i}(P(D_j|G_{j-1},t_{j,i},S_{j,i})) \\ &\prod_{j=1}^{m_i}(f(y_{j,i}|G_{j-1,i},t_{j,i},S_{j,i},D_{j,i})) \end{aligned} \tag{6.7}$$

That is, the four components are conditionally independent and can be modelled separately.

This is the "divide-and-conquer" approach we have adopted in the modelling of the complete claim development process - by decomposing it into simpler component processes and model each with relatively simpler models.

6.4 Modelling Individual Component Processes

A brief discussion on the overall model framework for the component processes is provided here and Sections 6.4.1 to 6.4.4 contain the full details for each component process.

The intention of this research is to model each component process with GLARMA type models. These models are discussed further in Davis et al. [2005], Rydberg and Shephard [2003] and Dunsmuir et al. [2014b]. In the case of NSW CTP claims the trajectories are not very long (90% of the claims are settled with in 5 revisions and 95% of the claims are settled within 10 revisions); hence, the NSW CTP data is more akin to longitudinal data as opposed to time series data (long series of relatively few variables) that GLARMA models are traditionally applied to.

Under the GLARMA model framework, a linear predictor of the form

$$\eta_{j,i} = Z_{j,i}^T \beta + W_{j,i} \tag{6.8}$$

is used, where $W_{j,i} = \sum_{k=1}^{p} \phi_k W_{j-k,i} + \sum_{k=1}^{q} \theta_k e_{j-k,i}$. $\eta_{j,i}$ is linked to the natural parameters of the discrete or continuous density selected for each of the component processes and $e_{j,i}$'s are the Pearson residuals for the selected density. For example, for a negative binomial distribution (chosen for the delay process), $\eta_{j,i} = \log \mu_{j,i}$, where $\mu_{j,i}$ is the mean of the distribution and $e_j = \frac{t_{j,i} - \mu_{j,i}}{\sqrt{\mu_{j,i} + \mu_{j,i}^2 / \alpha}}$. Z is the collection of covariates we use to model the "linear" component of $\eta_{j,i}$ and are different for each component process and incorporates current and past values of the other component processes based on the hierarchy defined above.

While the GLARMA framework allows the investigation of any lag structure appropriate, we have chosen to only fit an autoregressive (AR) structure with a lag of 1 (i.e., $p = 1$) for the NSW CTP data. That is $W_j = \phi_1 e_{j-1,i} + \phi_1^2 e_{j-2,i} + ...$ and the serial dependence structure gives an exponentially weighted smoothing of past Pearson residuals.

We felt this is appropriate because of two considerations. Firstly, while considered "long tailed" claims, the trajectories in the CTP dataset are still relatively short - most claims are settled within 5 revisions. These short trajectories do not allow the fitting of more complicated lag structure.

Secondly, we have tested more complicated lag structures using subsets of the CTP claims that have had longer trajectories. However, we have typically found that using lags greater than 1 led to problems with convergence and even when convergence occurred the coefficients at higher lags were typically not significant. In Chapter 7, the component models fitted are limited to a GLARMA structure limited to lag 1 in the AR component.

$Z_{j,i}$ is a component process specific vector of the variables of interest at revision j for the ith claim. Note that due to the hierarchical nature of the processes, the $Z_{j,i}$ vectors are different for different component processes and may include lagged values of the other component processes as well as current values of the other component process that are "earlier" in the hierarchy.

Three types of variables are used in $Z_{j,i}$ in addition to an initial column of 1's . These are slightly different from the three categories defined by Taylor and McGuire [2004].

i) Process variables themselves, that is, the values of the four claims development processes. By incorporating past values of these processes (the F filter) into the modelling of the current values, time series elements are introduced.

ii) Static variables, or baseline variables, do not change throughout the duration of the claim. These variables may include age of claimant at date of accident, gender, location of the claimant, etc.

iii) Dynamic variables, those that change during the evolution of the claim. This can be further divided into those that change deterministically (such as number of revisions, which increase incrementally) or those that change unpredictably. The changes in the latter are the most interesting in determining the outcome of the process variables. The dynamic variables include Liability Status, Injuries type and Injury Severities, Legal Representation, etc.

We examine each component process in turn, select an appropriate distribution with reference to the empirical distributions from the NSW CTP dataset and, based on this, define the relevant likelihood functions.

6.4.1 The Delay Component

The delay component process measures the time between consecutive changes in the claims estimate. This can also be interpreted as the interval of time that takes new information to arrive such that the claims manager needs to revise the previous claims estimate. Making a previously allowed for payment on the claim does not qualify to be a change in the claims estimate.

The preferred format of the delays between revisions would be the number of days and the exponential distribution is expected to fit the delay distribution well. The exponential distribution would also simplify the algebra involved.

For the NSW CTP PIR, however, claims information is provided as quarterly "snapshots" that contains the claims information as at that point in time. This means changes to claims information and claim size estimates can only be observed from quarter to quarter. Using the PIR, we can only ascertain whether a claim revision has been made sometime in the quarter and not any more accurately than that. It is possible for the managing insurer to have made a number of

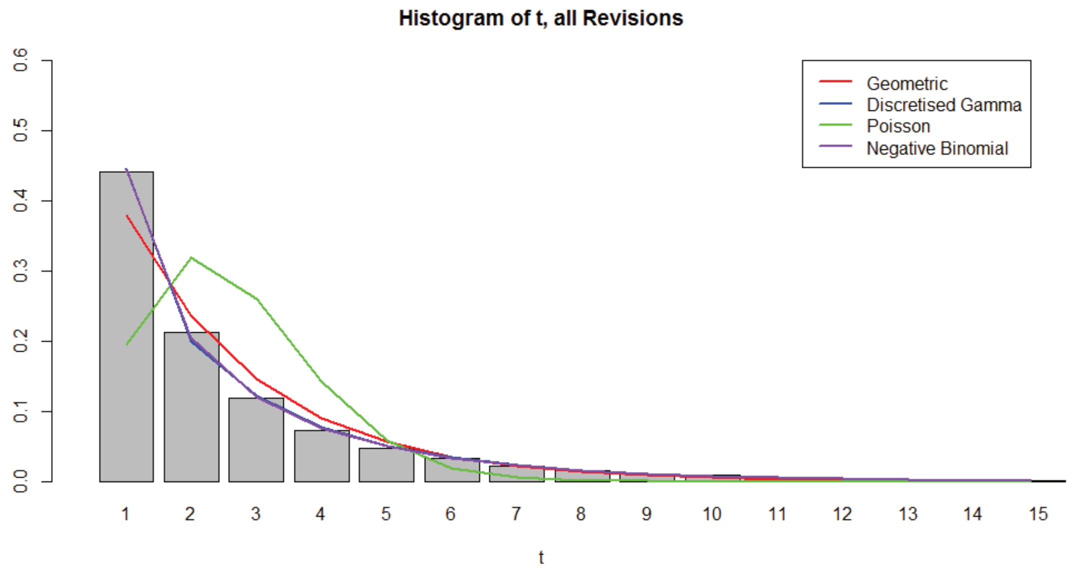

Figure 6.1: Distributions for Delay Process

changes within the same quarter, we would not be able to observe this using the PIR. The resulting implication is that when there is more than one change to the claims information and claim size estimate made between the quarterly extracts the changes are essentially rolled up and observed as one change.

This limitation in the data also means the delay process for the NSW CTP scheme is a discrete random variable, taking on values of 1, 2, 3, ... quarters. The geometric and negative binomial distributions are logical candidates from the pool of discrete distributions. On the other hand, "discretised" continuous distributions may also be possible candidates; this, however, is not ideal as these distribution would be difficult to work with. Figure 6.1 compares the empirical distribution of delays observed from the NSW CTP data to several fitted distributions.

From the Figure 6.1, the negative binomial and discretised gamma distributions appears to fit best and are equally good at representing the empirical distribution. The other distributions underestimated the volume of observations at the two ends of the distribution, that is, the extremely short and extremely long delays. We have not carried out any goodness of fit tests on the various fitted distribution as we are trying to visually gauge which distributions show a good representation of the data. The underlying data is an aggregation of a large number of individual variables each with a different mean and variance; the relationship between the claim characteristics and these parameters will be modelled and discussed in Chapter 7.

For the purposes of modelling the NSW CTP data, the negative binomial distribution is adopted for the delay process and is measured in quarters. For other applications other distributions may be more appropriate to be used in the general CDP framework.

For the negative binomial distribution, the following specification is adopted

$$P(t_{j,i} = t|F_{j-1,i}, G_{j-1,i}) = \frac{\Gamma(\alpha + t - 1)}{\Gamma(\alpha)\Gamma(t)}\Big(\frac{\alpha}{\mu_{j,i} + \alpha}\Big)^{\alpha}\Big(\frac{\mu_{j,i}}{\mu_{j,i} + \alpha}\Big)^{t-1} \tag{6.9}$$

where $t = 1, 2, 3, ...$,$\alpha > 0$ and $\mu_j = E[t_{j,i} - 1]$. Note this is a “shifted” negative binomial distribution as t starts at 1, the minimum delay in quarters. Under this specification, the mean parameter μ is specified explicitly in the probability density function which aids the modelling. It is also noted that when α approaches ∞ the negative binomial distribution approaches a Poisson distribution. That is, the α parameter can be considered as an over-dispersion parameter.

For the negative binomial distribution, the state equation for the GLARMA model is

$$\log \mu_{j,i} = Z^T_{\mu,j,i}\beta_\mu + W_{j,i} \tag{6.10}$$

where $W_{j,i} = \phi_{1,\mu}e_{j-1,i} + \phi^2_{1,\mu}e_{j-2,i} + ...$ and

$$e_{j,i} = \frac{(t_{j,i} - 1) - \mu_{j,i}}{\sqrt{\mu_{j,i} + \mu^2_{j,i}/\alpha}} \tag{6.11}$$

The log-likelihood function of the negative binomial function and its first and second derivatives can be found in Appendix F.1 as well as in Wang [2004]. The log-likelihood function and its derivatives are needed in the maximum likelihood estimation (MLE) of the parameters in Chapter 7. Note the likelihood and its derivatives are expressed using the "standard" negative binomial and a "shifting" transformation is carried out for model fitting, that is, $t_{j,i} - 1$ was fitted rather than $t_{j,i}$.

6.4.2 The Claim Status Component

This process models whether the claim is settled at a particular revision. Since the claim status can either be "Open" or "Settled" (complications of reopened claims are ignored), a Bernoulli process would be the natural selection. Let $S_{j,i}$ be the claim status process for the ith claim on its jth revision, for $j = 1, 2, ...$; then

$$S_{j,i} = \begin{cases} 1 & \text{if the } j\text{th revision is the last} \\ 0 & \text{if the } j\text{th revision is not the last} \end{cases} \tag{6.12}$$

and

$$P(S_{j,i} = 1 | F_{j-1,i}, G_{j,i}, t_{j,i}) = q_{j,i} \tag{6.13}$$

Based on this formulation, $S_{j,i}$ can be determined based on all the information available at time $T_{j-1,i}$, the new claims information as contained in filter $G_{j,i}$ and the delay that has taken place since the previous revision, $t_{j,i}$.

With the finalisation process variable, $S_{j,i}$, being a Bernoulli process, this can be used to infer the expected number of revisions. If all $q_{j,i}$ are the same for all i and all j, then the marginal distribution of the number of revisions the claims experiences before finalisation would follow a geometric distribution. However, since $q_{j,i}$ may be different for different is and js, the actual observed distribution of the number of revisions prior to finalisation is expected to have higher variability than the geometric distribution. Figure 6.2 shows such a comparison for the NSW CTP data across all claims.

The fitted geometric distribution compares well to the actual the number of

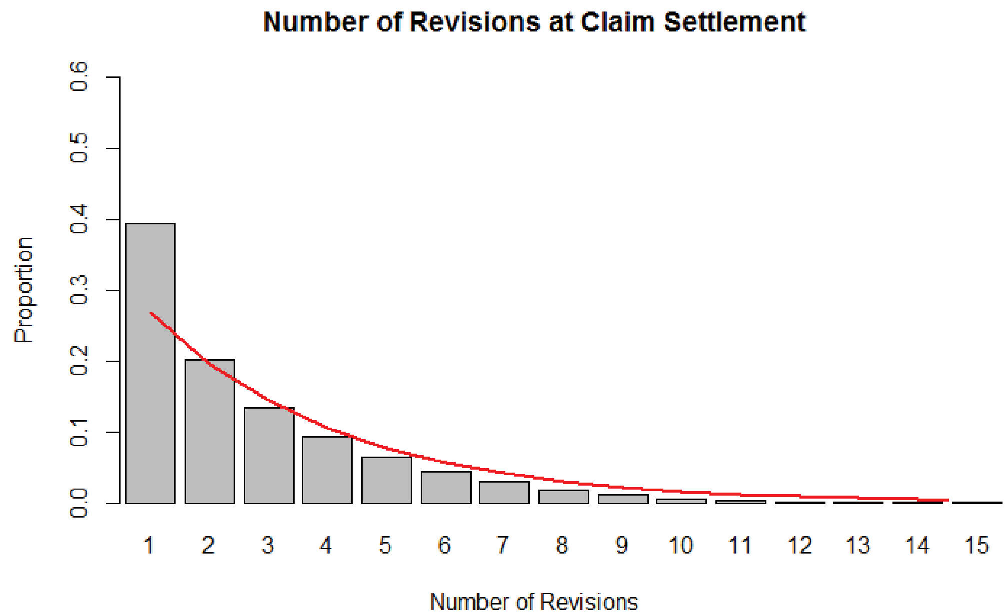

Figure 6.2: Observed Number of Revisions at Settlement vs. Fitted Geometric Distribution

revisions made at the settlement of a claim. As expected, there seems be a higher proportion of claims finalised with a delay of 1; this increases the variability of the observed distribution. This under-fitting of claims settled after 1 revision is due to the mixing of claims with different $q_{j,i}$ and this results in the observed distribution having a greater variability than the standard geometric distribution. Since the number of revisions is used as an explanatory variable in the model fitting (Chapter 7), this under-fitting should not be a concern.

For the binomial distribution, the state equation for the GLARMA model is

$$\text{logit}(q_{j,i}) = Z_{q,j,i}^T \beta_q + W_{j,i} \tag{6.14}$$

where $W_{j,i} = \phi_{1,q} e_{j-1,i} + \phi_{1,q}^2 e_{j-2,i} + ...$ and

$$e_{j,i} = \frac{S_{j,i} - q_{j,i}}{\sqrt{q_{j,i}(1 - q_{j,i})}} \tag{6.15}$$

The log-likelihood function of the binomial distribution can be found in Dunsmuir and Scott [2015] as well as in Wang [2004]. The log-likelihood function and its derivatives are needed in the maximum likelihood estimation (MLE) of the parameters in Chapter 7.

6.4.3 The Direction of Change Component

This process records whether a change is an upward change (on the receipt of unfavourable information, for the insurer) or a downward change (on the receipt of favourable information). Similar to the claim status process, the revision direction process is also binary, taking values up or down. Let $D_{j,i}$ be a Bernoulli variable such that

$$D_{j,i} = \begin{cases} 1 & \text{if the } j\text{th revision is a positive change} \\ 0 & \text{if the } j\text{th revision is a negative change} \end{cases} \tag{6.16}$$

and let

$$P(D_{j,i} = 1 | F_{j-1,i}, G_{j,i}, t_{j,i}, S_{j,i}) = p_{j,i} \tag{6.17}$$

Once again, Bernoulli seems to be the only logical choice for the direction of change process. However, one may ask why modelling direction of change and size of change separately and not model the change using a distribution on the real number scale. This further decomposition is partly driven by the aim to further simplify the distribution and also partly due to the behaviour of the claims managers. The claim managers seem to over-react to negative information so that an excessive claims estimate is built up over time. This usually leads to a reduction in the claim size at finalisation. By separating the claim direction from the claim size process; and allowing the claim size process to be conditioned on both claims finalisation and direction process, a more robust claims revision size processes can be modelled. This approach allows an asymmetry in the claim sizes for upward and downward revisions, as well as for final and prior revisions, to be reflected through the mean parameter. This is discussed further in Section 6.4.4.

Figure 6.3 shows the empirical proportions of upward revisions, $P(D_{j,i} = 1)$, by the number of revisions j, separately for the final revision and all prior revisions. The key observation is that the final revisions are far more likely to be downwards (suggesting the claim estimates prior to the settlement of the claim tend to be conservative) and the revisions prior to settlement are more likely to be upwards (incremental news received regarding the claims tend to be unfavourable news).

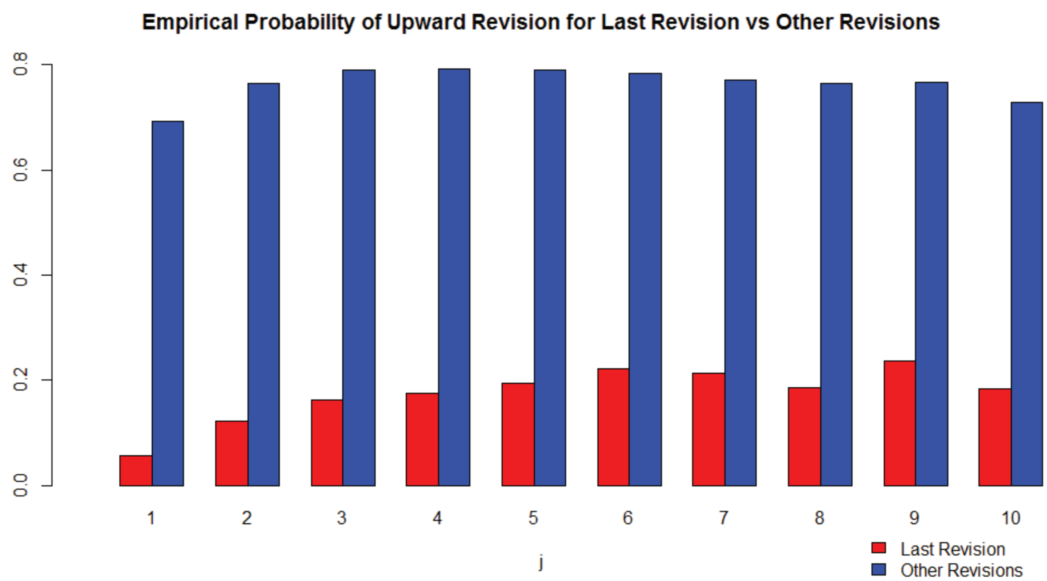

Figure 6.3: Proportion of Positive Revisions by Finalisation Status vs. Revision Number

Similar to the Settlement component process, the state equation for the GLARMA model is

$$\text{logit}(p_{j,i}) = Z^T_{\mu,j,i}\beta_p + W_{j,i} \tag{6.18}$$

where $W_{j,i} = \phi_{1,p}e_{j-1,i} + \phi^2_{1,p}e_{j-2,i}$ and

$$e_{j,i} = \frac{D_{j,i} - p_{j,i}}{\sqrt{p_{j,i}(1 - pj, i)}} \tag{6.19}$$

The log-likelihood function and its derivatives are needed in the maximum likelihood estimation (MLE) of the parameters in Chapter 7.

6.4.4 The Size of Change Component

This component process models the magnitude of the claim revision as the absolute difference between the logs of the pre and post revision claim costs, $Y_{j,i} = |\log(X_{j,i}) - \log(X_{j-1,i})| = \left|\log \frac{X_{j,i}}{X_{j-1,i}}\right|$. Such definition provides a magnitude of change that is a strictly positive continuous variable. Another added advantage of using this definition of size is that while CTP claim sizes, and incremental changes in size can be very long tailed (right skewed) in absolute dollar terms, this transforms reduces some of the skewness.

The NSW CTP data was examined to choose the most appropriate distribution to model the size of revision process, $Y_{j,i}$. Figure 6.4 shows a histogram of the observed magnitude of changes variable, $Y_{j,i}$, superimposed with several distributions fitted to the data. With the exception of lognormal, which showed the poorest fit, the other distributions all fitted the observed data well.

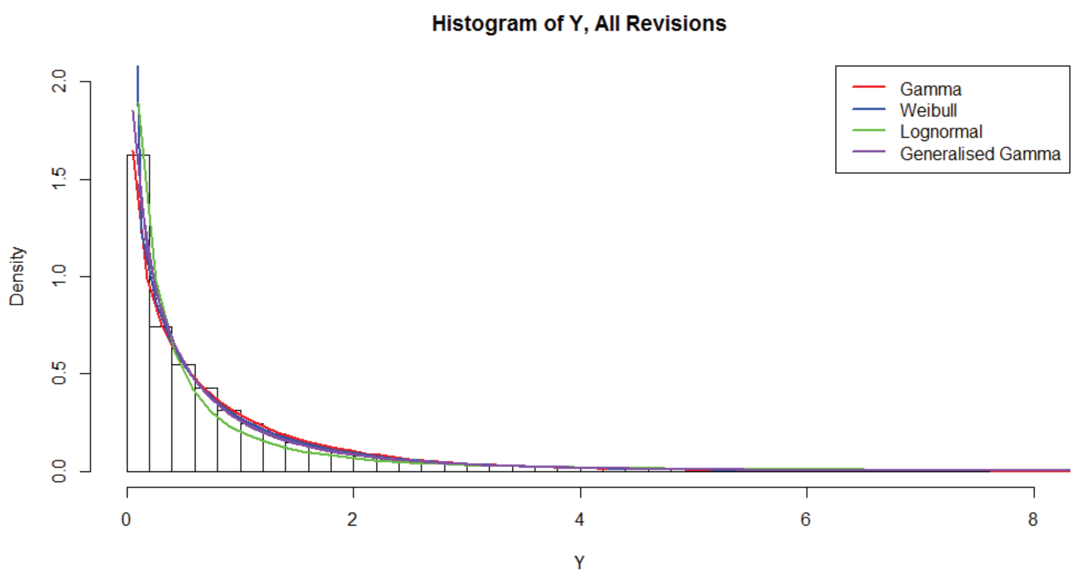

Figure 6.4: Observed Distribution of Size of Revision

In the above section, we saw that the revisions behave differently depending on whether it is the final revision or intermediate revisions. Figure 6.5 shows histograms of the observed size of changes stratified by Settlement flag and the

Direction of the revision.

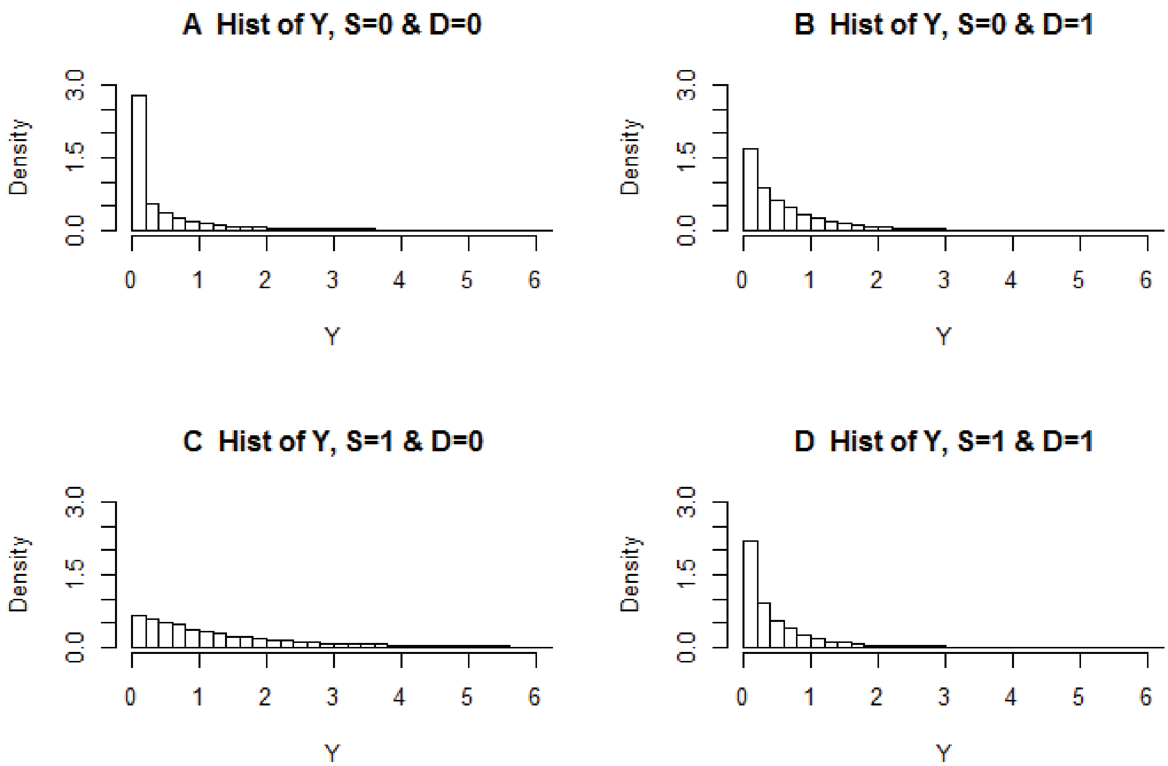

Figure 6.5: Observed Distribution of Size of Revision by Settlement and Direction

It can be seen from Figure 6.5 that negative revisions at finalisation (Panel C) behave quite differently. These "savings at finalisation" changes are considerably larger than other categories of revisions (Panels A, B and D). This difference reflects the level of conservatism the claims managers have build into the claim size estimates. It reveals claim managers typically reserve at a "worst-case" scenario which, frequently, result in a considerable reduction of the claim size when the claim is finalised. In fact, this "saving on finalisation" occurs for over 80% of the claims;

As a result, in the analysis of size of revisions, we consider the two "kinds" of revisions, "savings on finalisation" (revisions represented by Panel C) and the other revisions (as represented by Panels A, B and D), separately to determine the most appropriate distribution for their modelling. Figure 6.6 and Figure 6.7 show that the gamma, Weibull and generalised gamma distributions fit the size of change process quite well and almost indistinguishably. For its versatility, the generalised gamma distribution was proposed to model the Size component process, especially since Figure 6.5 shows the distributions of size of revisions are different for various subgroups of the dataset. Since the CDP framework attempts to model individual claim trajectories, this flexibility would be a useful attribute to have in the model.

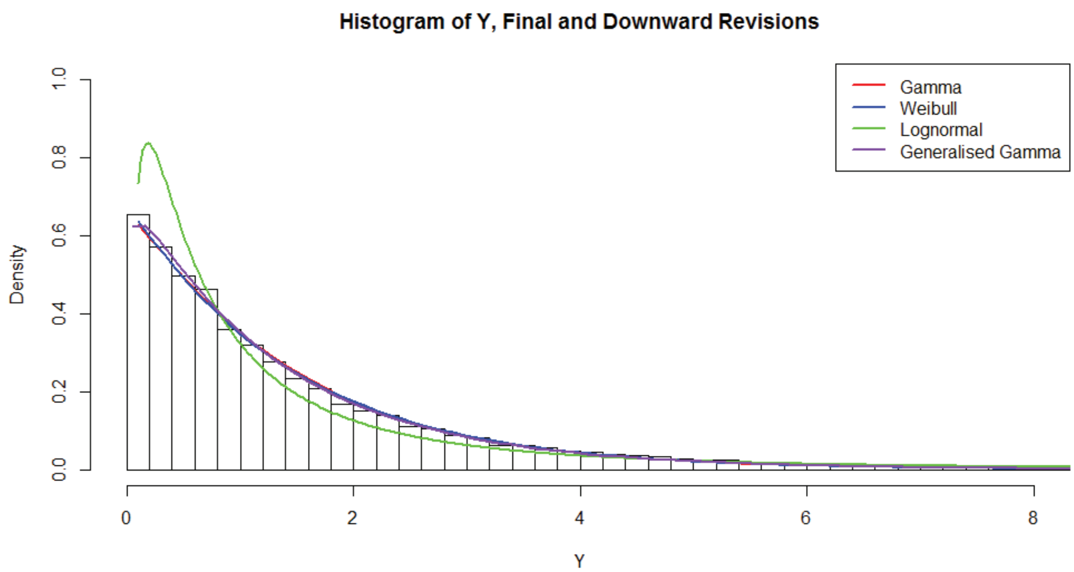

Figure 6.6: Observed vs Expected Magnitude of “Savings on Finalisation”

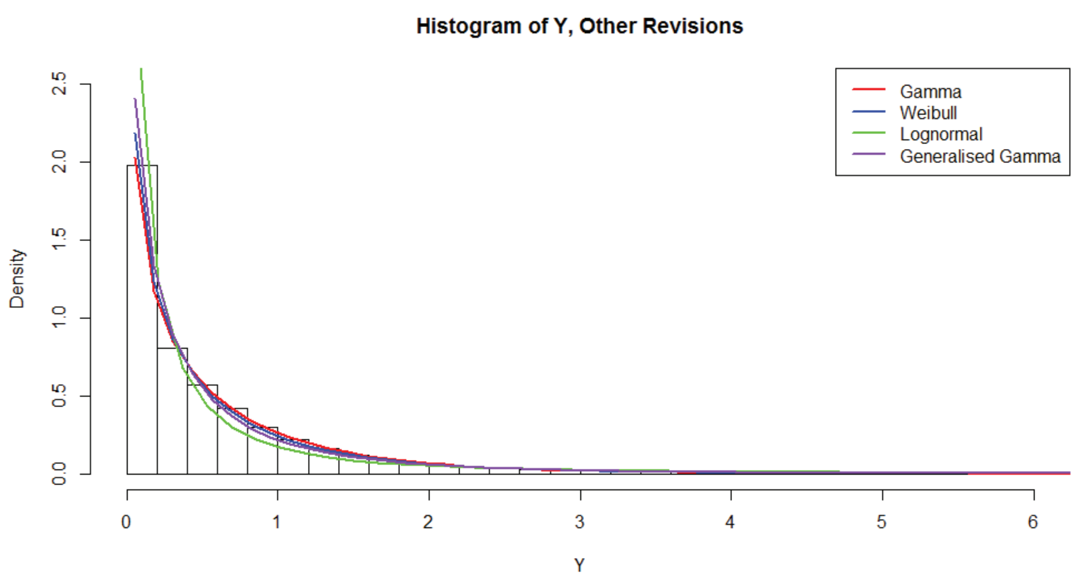

Figure 6.7: Observed vs Expected Magnitude of Other Revisions

By using various parameters, the generalised gamma can achieve a wide range of distributions.

However, the generalised gamma proved to be extremely difficult to achieve convergence. More details are provided in Section 7.8. Modelling suggests that for the NSW CTP data, the benefits of adopting the generalised gamma distribution is minimal and the increased time required to fit the generalised gamma distribution is significant. As a result, the modelling of the Size process in the research will adopt a standard gamma distribution.

The standard gamma distribution has the following density function:

$$f_{\lambda_{j,i},c}(y_{j,i}|F_{j-1,i},G_{j,i},t_{j,i},S_{j,i},D_{j,i}) = \frac{1}{\Gamma(c)\lambda_{j,i}}\left(\frac{y_{j,i}}{\lambda_{j,i}}\right)^{c-1} e^{-(\frac{y_{j,i}}{\lambda_{j,i}})} \tag{6.20}$$

where $\lambda_{j,i}$ is proportional to the mean of the variable $Y_{j,i}$.

For the standard gamma distribution, the state equation for the GLARMA model is

$$\log \lambda_{j,i} = Z^T_{\lambda,j,i}\beta_\lambda + W_{j,i} \tag{6.21}$$

where $W_{j,i} = \phi_{1,\lambda}e_{j-1,i} + \phi^2_{1,\lambda}e_{j-2,i} + ...$ and

$$e_{j,i} = \frac{y_{j,i} - c\lambda_{j,i}}{\sqrt{c\lambda^2_{j,i}}} \tag{6.22}$$

The log-likelihood function of the gamma function and the derivation of its first and second derivatives can be found in Appendix F.3. The log-likelihood function and its derivatives are needed in the maximum likelihood estimation (MLE) of the parameters in Chapter 7.

6.4.5 The Claims Cost Process and Likelihood

Let X_i be the ultimate cost of a claim, then using the notations defined above the ultimate claim cost can be expressed as the product of the revisions with the reported claim size.

$$X_i = X_{0,i} \prod_{j=1}^{m} e^{(2D_{j,i}-1)Y_{j,i}} \tag{6.23}$$

Or, more appropriately for the purposes of ultimate claim size projection,

$$X_i = X_{j',i} \prod_{j=j'+1}^{m} e^{(2D_{j,i}-1)Y_{j,i}} \tag{6.24}$$

where $t_{j+1,i} > T - T'$, that is, for open claims, the $j' + 1$th revision takes place after the censor time T. This is discussed in more detail in Section 6.5.

Using the above defined functional forms for the component processes, the log-likelihood of the complete process is (assuming the gamma distribution is used for the Size component)

$$\ell(\beta_\mu, \beta_q, \beta_p, \beta_\lambda, c, \alpha, \phi_{1,\mu}, \phi_{1,q}, \phi_{1,p}, \phi_{1,\lambda}) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}(\log P(t_{j,i})+\log P(S_{j,i})+\log P(D_{j,i})+\log f(y_{j,i})) \tag{6.25}$$

the conditioning has been dropped in the above equation to make it more concise.

The log-likelihood is segmented and the parameters are separately specified for each component process.

$$\begin{aligned}
\ell(\beta_\mu, \alpha, \phi_{1,\mu}) &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} \log P(t_{j,i}) \\
\ell(\beta_q, \phi_{1,q}) &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} \log P(S_{j,i}) \\
\ell(\beta_p, \phi_{1,p}) &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} \log P(D_{j,i}) \\
\ell(\beta_\lambda, c, \phi_{1,\lambda}) &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} \log f(y_{j,i})
\end{aligned} \tag{6.26}$$

In the modelling of these component processes, we further define δ as the vector of all the parameters of interest to be estimated using MLE. For example, for the Delay component, $\delta = (\beta_\mu \, \alpha \, \phi_{1,\mu})^T$. This usage is consistent in all the R programs and likelihood derivations in Appendices G and F.

6.5 Censoring

One important consideration, especially for a long tailed product such as CTP, is the censoring of claims development at the end of the data period. At time T', there would be claims that have occurred and reported prior to time T' but still not finalised, $T_{m,i} > T'$. These "censored" claims at the end of the data period will have a different form of likelihood to that shown in Equation (6.6).

$$L(\beta_\mu, \beta_q, \beta_p, \beta_\lambda, c, \alpha, \phi_{1,\mu}, \phi_{1,q}, \phi_{1,p}, \phi_{1,\lambda}) = \prod_{j=1}^{j'} P(t_j)P(S_j)P(D_j)f(y_j)P(t_{j'+1} > (T' - T_{j'})) \tag{6.27}$$

And the corresponding log-likelihood is

$$\begin{aligned} \ell_i(\beta_\mu, \beta_q, \beta_p, \beta_\lambda, c, \alpha, \phi_{1,\mu}, \phi_{1,q}, \phi_{1,p}, \phi_{1,\lambda}) =& [\sum_{j=1}^{j'} (\log P(t_{j,i}) + \log P(S_{j,i}) + \log P(D_{j,i}) + \log f(y_{j,i}))] \\ & + \log P(t_{j'+1} > (T' - T_{j'})) \end{aligned} \tag{6.28}$$

In Section 6.3 we have defined $P(t_j)$ in terms of all information available as at revision $j-1$. Hence, the quantity $\log P(t_{j'+1} < (T' - T_{j'}))$ can be defined from the various process variables and covariates known at $T_{j'}$, which occurred prior to the censoring date of T'. Also, due to the conditional hierarchy structure defined for the components, this log-likelihood can be decomposed as above. That is, Equation (6.26) can be adapted to accommodate censored claims as

$$\begin{aligned}
\ell(\beta_\mu, \alpha, \phi_{1,\mu}) &= \sum_{i=1}^{n}[\sum_{j=1}^{\min(m_i,j')} \log P(t_{j,i})] + \log P(t_{j'+1} > (T' - T_{j'})) \\
\ell(\beta_q, \phi_{1,q}) &= \sum_{i=1}^{n} \sum_{j=1}^{\min(m_i,j')} \log P(S_{j,i}) \\
\ell(\beta_p, \phi_{1,p}) &= \sum_{i=1}^{n} \sum_{j=1}^{\min(m_i,j')} \log P(D_{j,i}) \\
\ell(\beta_\lambda, c, \phi_{1,\lambda}) &= \sum_{i=1}^{n} \sum_{j=1}^{\min(m_i,j')} \log f(y_{j,i}) \qquad (6.29)
\end{aligned}$$

That is, apart from the Delay component, the other components are not affected by the issue of censoring. The adjustment required for the delay process is that the likelihood needs to allow for the fact that after the last revision, revision j', there has not been further revisions as at the census date. Equivalently, the delay of the next revision, revision $j'+1$, is longer than the length of time elapsed since the last revision at the census date, $t_{j'+1} > (T' - T_{j'})$.

In Chapter 7, the fitting the model considers the impact of accounting for the censoring issue. For the modelling of the censored claims, adjustments also need to be made to the log-likelihood of the delay component as well as its derivatives. These adjustments are detailed in Appendix F.2.

6.6 Model Fit and Inference

For inference, we assume, without proof, that the maximum likelihood estimates $\hat{\delta}$ for large samples have an approximate normal distribution, with mean at the true parameter, δ, and asymptotic covariance given by $\Omega = \left(-\frac{\partial^2 \ell}{\partial\delta\partial\delta^T}\right)^{-1}$. That is, $\hat{\delta} \sim N(\delta, \Omega)$. $\Omega(\delta)$ is estimated using $\Omega(\hat{\delta})$ and the standard errors for $\hat{\delta}$ are obtained from $\hat{\Omega}^{-\frac{1}{2}}$ in the usual fashion.

For model selection, we have used the Akaike Information Criterion (AIC) and

Bayesian Information Criterion (BIC). These model metrics trade off model fit (higher likelihood) while prevents overfitting by penalising the number of parameters used in the model. They are defined in Rice [2006] as

$$\begin{aligned} \text{AIC} &= -2\ell + 2k \\ \text{BIC} &= -2\ell + k\ln(n) \end{aligned} \tag{6.30}$$

where k is the number of parameters in the model, n is the number of observations, or revisions in the CDP framework and ℓ is the log-likelihood of the model of interest (from Equation (6.29)).

Where applicable, we have used the likelihood ratio test (LRT) to compare the goodness of fit of two models, where one is a nested model of the other. Model A is a nested model of Model B if Model A is a "special case" of Model B in the sense that Model A's parameters are a constrained subset of Model B. The LRT statistic, $\Lambda = 2(\ell_A - \ell_B)$ follows a χ^2 distribution with degrees of freedom as the difference in the number of parameters of the models; i.e., $\Lambda \sim \chi^2(\dim(\delta_B) - \dim(\delta_A))$, where δ_A and δ_B are the models' respective parameter vectors. If Λ is not in the critical region, then Model A and Model are deemed to have a similar fit.

The NSW CTP dataset has a large number of short trajectories - there are around 100,000 claims, with most claims finalises within 10 claim revisions; this dataset can be classified as a "longitudinal dataset". Justification for the use of asymptotic normality in this case can be found in the extensive literature on modelling longitudinal data such as Verbeke and Molenberghs [2009] and Diggle et al. [2002].

6.7 Summary

This chapter lays the theoretical foundation of the Claim Development Process framework. It introduces the overall claims development process (X_j, T_j) and introduces a hierarchical conditional structure to enable its decomposition into simpler component processes. The component processes are defined using the GLARMA structure and the GLARMA models were briefly discussed. Each component process were then discussed in detail and a distributional form is selected for each component. The issue of censoring and its impact on the Delay component is also raised and its resolution detailed.

While this thesis applies the CDP framework to NSW CTP data, it can be applied to other insurance products with long tailed claims. The application is split into three parts and discussed in the next three chapters. Firstly, in Chapter 7 the dataset is modelled to gain insights to how NSW CTP claims develop over time. Although the insights discussed are specific to the NSW CTP data, similar analysis is able to be undertaken for other portfolios as well. Secondly, in Chapter 8 the application of the claim development model as a mean of claims projection is examined, the projected claim sizes from using the modelling dataset are then compared to the actual claim sizes for those that were finalised within the validation dataset. Thirdly, in Chapter 9 the merits of using the Claims Development Processes as a method for valuation is examined.

Chapter 7

CDP Modelling Results

7.1 Introduction

This chapter details the implementation of the Claims Development Process model using the NSW CTP data. A brief discussion is provided for the software developed and used in the modelling process. The CTP data used for this section is briefly reviewed again along with the processing required to enable the models to be fit. Modelling results for each of the four component processes are tabulated and discussed. This chapter provides new insights into the development behaviour of NSW CTP claims that are not otherwise available with the previously discussed valuation techniques.

7.2 Model Fitting

The NSW CTP data are fitted to the CDP Component models specified Sections 6.4.1 to 6.4.4. The parameters of each of the component models (β, ϕ_1 and any shape parameters - α and c) are estimated using maximum likelihood estimation. As discussed in Section 6.3, the likelihoods of the four components can be separated due to their conditionally independent model specification, this allows the four component processes to be modelled separately. However, due to their longitudinal nature (multiple observations on the same claim over time) as discussed in Section 6.4, simple GLM software could not be used.

At the time of writing, Dunsmuir et al. [2014b] has recently released the GLARMA package in R (Dunsmuir and Scott [2015]) which enables the modelling of time series of count data in a GLARMA framework. Prior to this, for the purposes of this thesis, R programs were developed to carry out the model fitting of the CTP data. The model fitting follows the approach outlined in Davis et al. [2003] as well as Wang [2004]. Using custom written programs also allows the flexibility to fit more complicated models as discussed in the later chapters.

Maximum likelihood estimation of the model parameters was implemented using a Newton-Raphson iterative algorithm. Let δ be a vector of coefficients of interest, which is made up of β, the vector of coefficients, ϕ_1, the auto-regressive coefficient at lag 1, and other distributional parameters, such as α and c. Further, let ℓ be the log-likelihood of component process in question. Then the Newton-Raphson method estimates δ iteratively using the following relationship. The kth iterative estimate of δ is

$$\delta^k = \delta^{k-1} - \ell''(\delta^{k-1})^{-1}\ell'(\delta^{k-1}) \tag{7.1}$$

where $\ell'(\delta^{k-1})$ and $\ell''(\delta^{k-1})$ are the first and second derivative of ℓ with respect to δ evaluated at δ^{k-1}. Due to the functional forms of the log-likelihood functions, it can be quite difficult to obtain the first and second derivatives with respect to the parameters. However, we have done that and implemented these recursively in our R-code. The likelihoods have been discussed in Chapter 6.

Starting values adopted for the algorithm (δ^0) adopted were as suggested in Davis et al. [2003], where β^0 are derived from a simple GLM fit (i.e., without the ARMA structure) and ϕ_1^0 was initialised at 0. Using this approach the Newton-Raphson algorithm converges to the local maxima of the likelihood usually within 5 iterations. Some of the difficulties in model fitting are discussed in Section 7.12.

Two R programs were used to fit each of the component models

i) a generic program that performs the calculation of the likelihood of a particular distribution and its derivatives for one claim. Programs for the Poisson and binomial distributions were obtained from my supervisor. Programs for the negative binomial distribution were developed for Wang [2004] and programs for the gamma and generalised gamma distributions have been written for this thesis.

ii) a wrapper program that reads in the data, sets up the initial values of the δ vector, accumulates the likelihoods and its derivatives for each claim and then performs the Newton-Raphson algorithm iteratively to estimate the coefficients and parameters. The iterative process is terminated after convergence is obtained for the parameters estimates, when the incremental movements for all parameters is less than 10^{-6}.

A further complication for the delay process arises due to the issue of censoring, that is, claims not finalised at the end of the data period. Substantial modifications were applied to the negative binomial program to enable the model making allowances for censored claims. This is further discussed in Section 7.5.

R programs used for this chapter are contained in Appendix G,

- Calculating the likelihood contributions from the gamma distribution - Appendix G.1
- Calculating the likelihood contributions from the negative binomial distribution - Appendix G.2
- Calculating the likelihood contributions from the negative binomial distribution with allowance for censoring - Appendix G.3
- Calculating the likelihood contributions from the binomial distribution - Appendix G.4

- Example of a wrapper program - Appendix G.5.

For each component process, a "full" model is firstly fit with all the available covariates for revisions up to December 2009. An iterative process is then undertaken to eliminate successively the "weakest" covariate until an optimal model is achieved. Significance of the parameters as well as the various model selection criterion (AIC and BIC) are considered in choosing this optimal model. Once this model is selected, it was also fitted to different cohorts of claims (according to their year of accident) to examine the stability of the parameters across cohorts.

7.3 Data Preparation

The CTP data used in model fitting is briefly discussed again. Considerable processing needs to be made to change the previously used quarterly extracts from a time based dataset into an "revision" based dataset.

As before the data from accident years 2001 to 2009 have been used to produce the modelling results contained in the following sections. The results serve dual purposes - to understand the claims development behaviours and then use the results to project how open claims at the censoring date (or valuation date) will develop in the future.

For the claims that occurred during this period of time, their claim cost movements have been observed for a period of 9 years from January 2001 to December 2009. For each time the claim incurred cost changes by either \$100 or 1%, whichever is greater, it is marked to have had a "change". We have found changes smaller than this threshold may not be associated with the arrival of new information about the claim. Changes that are smaller than this threshold are also financially immaterial to a portfolio with annual claims cost of close around \$1,000m. These small changes are simply "rolled" into the next change.

The diagrams below show when "revisions" occur for the two claims used to demonstrate the format of the data from Chapter 3. The graphs are reproduced below with black "j" counters now superimposed to denote when changes occur. Typically, most changes to the claims incurred are recorded; however, the readers may notice that for the second claim, a small adjustment to the claims cost in 2006 Q4 was too small and was under the threshold to be classified as a revision.

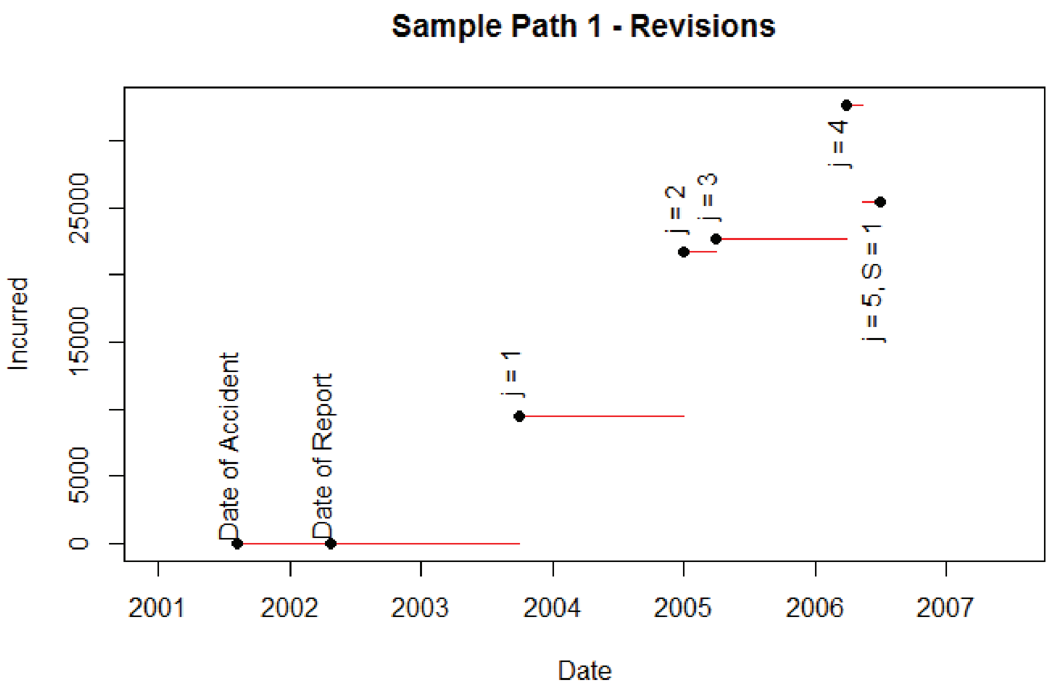

Figure 7.1: Sample 1 - Revision Based Observations

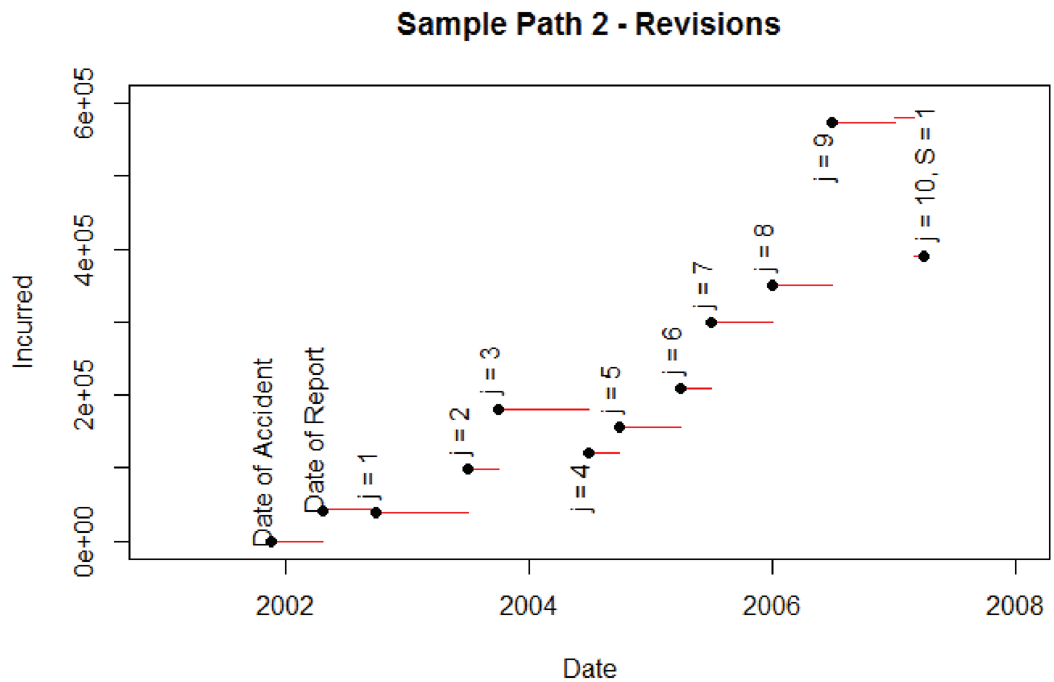

Figure 7.2: Sample 2 - Revision Based Observations

Table 7.1 provides a small subset of the data that corresponds to the claims shown. Only a subset of the covariates are shown, there are over 45 covariates in the dataset. While not modelled in the CDP, $X_{0,i}$, the initial estimate of the claim size when the claim is reported is also shown in Table 7.1. For modelling purposes, likelihoods are only calculated for actual claim revisions, that is, for $j = 1, 2, 3....$ The inclusion of $X_{0,i}$ is to facilitate the calculation of $Y_{1,i}$. Some claims are reported without an initial claim estimate, i.e., $X_{0,i} = 0$, this results in the $Y_{1,i}$ to be undefined. For the purposes of modelling, whenever a claim is reported at nil cost, a notional \$1000 is inserted. This can be seen for Sample Data 1 in Table 7.1.

			Process Variables - F_j,i					Covariates - G_j,i			
Sample (i)	Revision (j)	X_j,i	t_j,i	S_j,i	D_j,i	Y_j,i	Legal Rep	Severity	Liability	Gender
1	0	1,000								
	1	9,500	4	0	1	0.978	Y	1	F	F
	2	21,672	6	0	1	0.358	Y	1	F	F
	3	22,672	1	0	1	0.020	Y	1	F	F
	4	32,656	4	0	1	0.158	Y	1	F	F
	5	25,436	1	1	0	0.109	Y	1	F	F
2	0	42,482								
	1	38,437	1	0	0	0.043	N	2	R	F
	2	98,437	2	0	1	0.408	N	2	F	F
	3	181,238	2	0	1	0.265	N	2	F	F
	4	121,031	3	0	0	0.175	N	2	F	F
	5	156,076	1	0	1	0.110	N	2	F	F
	6	208,181	2	0	1	0.125	N	2	F	F
	7	300,181	1	0	1	0.159	N	2	F	F
	8	349,881	2	0	1	0.067	N	2	F	F
	9	573,969	2	0	1	0.215	N	2	F	F
	10	390,714	3	1	0	0.167	Y	2	F	F

Table 7.1: Revision Based Dataset (Subset)

To simplify the modelling of the interval valued covariates that are thought, a priori, to have a non-linear impact on the linear predictor, we have decided to categorise these into bands and treat them as factor type variables, with separate parameters for each category level.

i) Age of claimant has been banded into 0 to 9, 10 to 16, 17 to 25, 26 to 45, 46 to 65 and 66 and older.

ii) Injury severity score (ISS), this is a composite index constructed from the injuries severities sustained by the claimant. By design, ISS is made to be proportional to the probability of the claimant dying from the injuries sustain. ISS has been banded into 0, 1, 2, 3 to 5, 6 to 10, 11 to 30 and 31 to 75.

iii) Maximum injury severity (or simply known as just severity) ranges from 1 (minor) to 5 (severe); it can also take on values of 0 (unknown) and 6 (death). Severity has been used as a categorical variable in the model fitting.

The covariates are then encoded into a design matrix. Interval valued covariates are included as-is, or log transformed if appropriate, while a categorical variable with k categories require $(k-1)$ columns in the design matrix to code the contrasts of each level from the intercept. The following table shows the size of the dataset and the resultant design matrix. Table 7.2 shows the number of claims reported by accident year, the number of revisions made to those claims, the number of variables contained as well as the columns in the design matrix. In total 189,507 revisions data points were used; and the design matrix for each of the component processes typically has around 20 million elements in it.

Accident	Number of	Number of	Number of Covariates* (Columns in Design Matrix)			
Year	Claims	Revisions	Delay	Settlement	Direction	Size
2001	11,545	34,499	40 (86)	46 (100)	47 (101)	48 (102)
2002	10,298	29,483	40 (86)	46 (100)	47 (101)	48 (102)
2003	9,588	26,704	40 (86)	46 (100)	47 (101)	48 (102)
2004	9,528	26,366	40 (86)	46 (100)	47 (101)	48 (102)
2005	9,071	23,870	40 (86)	46 (100)	47 (101)	48 (102)
2006	8,695	19,666	40 (86)	46 (100)	47 (101)	48 (102)
2007	8,580	16,722	40 (86)	46 (100)	47 (101)	48 (102)
2008	8,078	10,236	40 (86)	46 (100)	47 (101)	48 (102)
2009	5,615	1,961	40 (86)	46 (100)	47 (101)	48 (102)
Total	80,998	189,507	40 (86)	46 (100)	47 (101)	48 (102)

* Include values of other Component Processes as allowed under the hierarchical conditioning

Table 7.2: Dimensions of the Modelling Dataset and Design Matrix

7.4 Delay Component - Ignoring Censoring

The results of the model fitting of the Delay Component allows the user to understand what are the factors that influence the lengths of time between claim revisions. As discussed above, two models were fit. Firstly a model using all the covariates is fitted to the modelling data (revisions observed to December 2009) and secondly an optimal model, where all the variables are significant and model selection criteria of AIC and BIC are minimised.

The following tables (Table 7.3 and Table 7.4) show the results of the fitting of the delay process. Both the parameter coefficients and their estimated standard errors are shown, refer to Section 6.6 for further details. If the parameter is significant (at the traditional 5% level, i.e., an absolute z-value above 1.96) then the covariate coefficient is expressed in bold. As the standard errors are used only to gauge the significance of the coefficients and the significant coefficients are highlighted with bold text, we have kept the parameters in the table to 2 decimal places. This is done to reduce the cluttering of the tables when more decimal places are presented. The results are split into two tables due to the large number of covariates used in the modelling. The covariates are further split into various sections based the "type" of the covariates. They are as follows.

- F - the F information set, the other component process variables from previous or current revisions as available based on the hierarchy specified in Section 6.3. For example, for the delay component process only past values of the other components are present in the modelling; while for the size component process, present and past observations from all the other components are present in the modelling.

- De - Determinist dynamic variables, these variables are dynamic, but behave in a predictable fashion. Such as $X_{j,i}$ can be calculated from $X_{j-1,i}$,$D_{j,i}$ and $Y_{j,i}$.

- S - Static variable, these variables do not change throughout the duration of the claim. For example, age of claimant at the date of the accident.

- D - Dynamic variables, these variables may change over the life time of the claim and behave in an uncertain fashion.

It should be noted that all the dynamic variables used for the modelling of the delay process are from the previous revision, i.e., as at $j-1$. As discussed previously, the information set available for the determination of the delay process is as at $F_{j-1,i}$ as the new information pertaining the current revision has not yet arrived. See Section 6.3 for further details.

Delay Model					
Model		All Variables		Significant Variables	
Number of Claims		80,998		80,998	
Number of Revisions		189,507		189,507	
Number of Parameters		88		72	
Log-likelihood		-267,972		-267,984	
AIC		536,116		536,112	
BIC		536,793		536,678	
Parameter	Type	coeff	s.e.	coeff	s.e.
Intercept		**6.48**	0.04	**6.46**	0.03
phi(1)		**0.25**	0.00	**0.25**	0.00
alpha		**4.58**	0.06	**4.58**	0.06
Previous Direction	F	**0.08**	0.01	**0.08**	0.01
Previous Size	F	**0.08**	0.00	**0.08**	0.00
Previous Incurred Cost (log)	De	**0.06**	0.00	**0.06**	0.00
Revision 1	De				
Revision 2	De	**0.38**	0.01	**0.38**	0.01
Revision 3	De	**0.43**	0.01	**0.43**	0.01
Revision 4	De	**0.34**	0.01	**0.34**	0.01
Revision 5	De	**0.17**	0.02	**0.17**	0.02
Revision 6 +	De	**-0.11**	0.02	**-0.12**	0.02
Revision Year = 2001	De	**-8.62**	0.04	**-8.62**	0.04
Revision Year = 2002	De	**-6.97**	0.02	**-6.96**	0.02
Revision Year = 2003	De	**-5.72**	0.02	**-5.72**	0.02
Revision Year = 2004	De	**-4.62**	0.02	**-4.62**	0.02
Revision Year = 2005	De	**-3.67**	0.01	**-3.67**	0.01
Revision Year = 2006	De	**-2.75**	0.01	**-2.75**	0.01
Revision Year = 2007	De	**-1.91**	0.01	**-1.91**	0.01
Revision Year = 2008	De	**-0.98**	0.01	**-0.98**	0.01
Revision Year = 2009	De				
Reporting Delay (log)	S	-0.00	0.01		
Insurer A	S				
Insurer B	S	**0.14**	0.01	**0.15**	0.01
Insurer C	S	**0.17**	0.01	**0.17**	0.01
Insurer D	S	**0.22**	0.01	**0.22**	0.01
Insurer E	S	**0.07**	0.01	**0.07**	0.01
Insurer F	S	**0.07**	0.01	**0.07**	0.01
Insurer Other	S	**0.21**	0.01	**0.21**	0.01
Gender = Female	S	0.00	0.01		
Employed	S				
Self Employed	S	-0.00	0.01		
Unemployed	S	0.01	0.03		
Other	S	**0.02**	0.01	**0.01**	0.01
Age between 0 - 9	S	**0.08**	0.02	**0.09**	0.02
Age between 10 - 16	S	**0.05**	0.02	**0.04**	0.02
Age between 17 - 25	S	**0.02**	0.01	**0.02**	0.01
Age between 26 - 45	S				
Age between 46 - 65	S	-0.00	0.01	-0.00	0.01
Age 66 +	S	**-0.03**	0.01	**-0.03**	0.01
Resides in Metro area	S				
Resides in Outer Metro area	S	**-0.02**	0.01	**-0.02**	0.01
Resides in Country area	S	-0.03	0.02		
Resides in Wollongong area	S	-0.02	0.02		
Resides in Newcastle area	S	**-0.02**	0.01	**-0.02**	0.01
Accident Year = 2001	S				
Accident Year = 2002	S	**-1.00**	0.01	**-1.00**	0.01
Accident Year = 2003	S	**-1.91**	0.01	**-1.91**	0.01
Accident Year = 2004	S	**-2.80**	0.01	**-2.80**	0.01
Accident Year = 2005	S	**-3.71**	0.02	**-3.71**	0.02
Accident Year = 2006	S	**-4.64**	0.02	**-4.65**	0.02
Accident Year = 2007	S	**-5.68**	0.02	**-5.69**	0.02
Accident Year = 2008	S	**-7.06**	0.03	**-7.06**	0.03
Accident Year = 2009	S	**-8.47**	0.05	**-8.47**	0.05
Legal Rep at report	S	**0.06**	0.01	**0.06**	0.01
Max. Severity at report	S	**0.03**	0.00	**0.03**	0.00

*italicised variables represents the baseline value of a categorical variable

Table 7.3: Delay Process - Coefficients1

Delay Model (cont.)					
Model		All Variables		Significant Variables	
Number of Claims		80,998		80,998	
Number of Revisions		189,507		189,507	
Number of Parameters		88		72	
Log-likelihood		-267,972		-267,984	
AIC		536,116		536,112	
BIC		536,793		536,678	
Parameter	Type	coeff	s.e.	coeff	s.e.
ISS at j - 1 = 0	D	**-0.41**	0.08	**-0.40**	0.08
ISS at j - 1 = 1	D				
ISS at j - 1 = 2	D	-0.01	0.01	-0.01	0.01
ISS at j - 1 between 3 and 5	D	**-0.03**	0.01	**-0.03**	0.01
ISS at j - 1 between 6 and 10	D	**-0.07**	0.02	**-0.08**	0.02
ISS at j - 1 between 11 and 30	D	**-0.11**	0.03	**-0.12**	0.02
ISS at j - 1 between 31 and 75	D	-0.09	0.05	**-0.10**	0.04
Maximum Severity at j - 1 is unknown	D	**0.42**	0.08	**0.42**	0.08
Maximum Severity at j - 1 = 1	D				
Maximum Severity at j - 1 = 2	D	-0.00	0.01	0.01	0.01
Maximum Severity at j - 1 = 3	D	-0.02	0.02	-0.01	0.02
Maximum Severity at j - 1 = 4	D	-0.02	0.03	-0.01	0.03
Maximum Severity at j - 1 = 5	D	**-0.11**	0.05	**-0.10**	0.04
Maximum Severity at j - 1 = 6	D	**-0.12**	0.06	**-0.11**	0.05
Body regions injured at j - 1 is unknown	D	0.01	0.01		
Body regions injured at j - 1 = 1	D				
Body regions injured at j - 1 = 2	D	0.02	0.01		
Body regions injured at j - 1 = 3	D	0.00	0.01		
Body regions injured at j - 1 = 4	D	**-0.05**	0.02		
Body regions injured at j - 1 = 5	D	0.02	0.05		
WPI at j - 1 is 10% or less	D				
WPI at j - 1 is more than 10%	D	**-0.22**	0.01	**-0.22**	0.01
WPI at j - 1 is unknown	D	**-0.02**	0.01	**-0.02**	0.01
Claim Size has been 5 times initial est.	D	**-0.10**	0.01	**-0.11**	0.01
Payment at j - 1 exceed 70% of Incurred	D	**-0.06**	0.01	**-0.06**	0.01
Claim duation at j - 1 (log)	D	**-1.30**	0.01	**-1.30**	0.01
Brain Injury at j -1	D	-0.00	0.02		
Spine Injury at j -1	D	**0.17**	0.07	**0.18**	0.06
Back injury at j - 1	D	0.03	0.02		
Whiplash Injury at j - 1	D	-0.04	0.02		
Legal Rep at j - 1	D	**0.06**	0.01	**0.07**	0.01
Economic Loss compensation at j - 1	D	0.01	0.01		
No Litigation at j - 1	D				
NSW Supreme Court at j - 1	D	**-0.60**	0.05	**-0.60**	0.05
NSW District Court at j - 1	D	**-0.77**	0.02	**-0.77**	0.02
NSW Local Court at j - 1	D	**-0.42**	0.05	**-0.42**	0.05
CARS at j - 1	D	**-0.51**	0.01	**-0.51**	0.01
Accepted Liability at j - 1	D				
Partially Accepted Liability at j - 1	D	**0.04**	0.01	**0.04**	0.01
Rejected Liabiliy at j - 1	D	**0.16**	0.01	**0.17**	0.01
Other at j - 1	D	**0.10**	0.01	**0.10**	0.01
Rehab Needed at j - 1	D				
Rehab Probably at j - 1	D	**0.04**	0.01	**0.05**	0.01
Rehab Possible at j - 1	D	-0.01	0.01		
Rehab not required at j - 1	D	**-0.04**	0.01	**-0.03**	0.01
Other at j - 1	D	**-0.12**	0.01	**-0.12**	0.01

*italicised variables represents the baseline value of a categorical variable

Table 7.4: Delay Process - Coefficients2

The log-likelihood of the model that uses all the covariates is -267,972 and that of the model that only has significant variables is -267,984. We have carried out an LRT (see Section 6.6 for further details) to compare the model fit between the model where all the variables are used and the model where only significant parameters are retained. The LRT statistic is 23.4 and the 5% critical region threshold for a χ^2 distribution with 16 degrees of freedom (the number of extra parameters used) is 26.3. Hence, the model with only significant parameters represents a similar fit with fewer parameters and this is also shown through the superior AIC and BIC (see Section 6.6 for details).

The parameter estimates provide insights to the behaviour of the delay component processes. Examining the Process (F) and Deterministic (De) variables, the following observations are made. Conditional on all the variables in the final model:

i) ϕ_1 is positive at 0.25 and strongly significant, this suggests a larger than predicted previous delay, t_{j-1}, would have an impact to increase the current delay.

ii) The direction of the previous revision, D_{j-1}, is positive and significant, suggesting an upward increase at the last revision would prolong the time before the next revision; while a downward revision would be more likely to reduce it.

iii) The size of the previous change, Y_{j-1}, is a key determinant of the current delay, and this factor is extremely significant. Larger sized previous changes would increase the current delay. This perhaps can be interpreted as a significant set of news regarding the claim, that caused a large claims revision, at the prior change may lead to a longer period of observation before a further revision. An interaction term between direction and size could also be used in the model framework to test asymmetry of the impact from the size of the previous change. While this may be of interest, we have only included main effects in the model.

iv) A large previous claims cost estimate (log transformed), $\log(X_{j-1,i})$, leads to a longer delay.

v) The counter of the number of revisions, j, is highly significant in predicting the delays between revisions. The data suggest that given the other factors are equal, subsequently revisions after the third gradually shortens.

Examining the "Static" explanatory variables, the following observations are made.

i) The insurer variable shows some interesting results. The 6 largest insurers were randomly allocated characters "A", "B", ..., "F" and the other insurers were grouped together. Insurer A (the baseline) seems to have the shortest intervals between delays as all the other coefficients are positive. It remains to be seen whether Insurer A updates its claims cost estimates more frequently and hence also makes more revisions before a claim is settled or that Insurer A is simply more pro-active in their claims management.

ii) Another claims factor impacting the delay between revisions is the age of the claimant. The pattern is generally consistent - younger claimants (10 - 16) have longer delays compared to older claimants. This again stands to reason as many injuries in children may require more time to observe before the claims are settled.

iii) The claim severity at the time the claim is reported as well as legal representation at the time the claim is reported are significant factors although the impact of each is modest.

The "Dynamic" variables by nature are the most interesting variables to examine, as they allow us to examine whether changes in claims behaviour are caused by the changing circumstances of the claimant or more aligned with the baseline characteristics (as implied by the model framework proposed by Larsen [2007]).

Once again, for the modelling of the delay factor, all the covariates used are as at the previous revision, $j-1$.

i) The "Large Change" variable is constructed to represent whether the claim at the previous revision is at a size that is more than 5 times the original estimate. This variable may be a proxy to indicate whether a significant claim estimate revision has already occurred, based on prior receipt of significant information pertaining the claim. This factor has a significant impact of the delay process; however, the impact is around -10%. That is, claims that have undergone significant increases have slightly shorter delays compared to those that have not had significant increases before.

ii) The injury severity based covariates are significant and seem to suggest the claims with more severe injuries have shorter delays between revisions.

iii) Claim duration at previous revision (log transformed) is a significant covariate. It suggests more mature claims have shorter delays between claim revisions.

iv) Once the claim enters the court or CARS system or meets the 10% Whole Person Impairment threshold, the delay also seems to be shorter, suggesting an acceleration of activities.

A second set of Delay models are constructed by fitting each accident year separately. These are used to examine the stability of the coefficients for different cohorts of claimants. We have carried out an LRT with the combined model and the cohort based models. The LRT statistic is 22,508 and the 5% critical region threshold for a χ^2 distribution with 445 degrees of freedom (the number of extra parameters used) is 495. Hence, the cohort based models represent a significantly better fit to the data, consistent with the AIC and BIC measures. This suggests there are changes to the parameters overtime or that there are material interaction effects not picked up by our modelling.

Delay Model											
Model		All	Data segmented by Accident Year								
		Years	2001	2002	2003	2004	2005	2006	2007	2008	2009
Number of Claims		80,998	11,545	10,298	9,588	9,528	9,071	8,695	8,580	8,078	5,615
Number of Revisions		189,507	34,499	29,483	26,704	26,366	23,870	19,666	16,722	10,236	1,961
Number of Parameters		72	63	62	61	60	59	58	57	53	44
Log-likelihood		-267,984	-49,658	-42,195	-38,551	-37,597	-32,845	-26,246	-19,424	-9,265	-949
Parameter	Type	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff
Intercept		**6.46**	**4.82**	**4.44**	**4.30**	**3.86**	**3.11**	**2.07**	**0.93**	**-1.73**	**-7.12**
phi_1		**0.25**	**0.15**	**0.14**	**0.15**	**0.18**	**0.14**	**0.15**	**0.14**	**-0.04**	**-0.14**
alpha		**4.58**	**5.90**	**8.16**	**8.99**	**10.18**	**12.56**	**50.74**	**162.18**	**6.00**	186.81
Previous Direction	F	**0.08**	**0.07**	**0.07**	**0.04**	**0.06**	**0.05**	-0.00	**-0.07**	-0.05	1.53
Previous Size	F	**0.08**	**0.09**	**0.08**	**0.08**	**0.07**	**0.04**	**0.04**	**0.08**	0.04	-0.40
Previous Incurred Cost (log)	De	**0.06**	**0.02**	**0.03**	**0.03**	**0.03**	**0.04**	**0.05**	**0.08**	**0.10**	**0.12**
Revision 1	De										
Revision 2	De	**0.38**	**0.22**	**0.26**	**0.24**	**0.23**	**0.22**	**0.19**	0.01	-0.08	0.13
Revision 3	De	**0.43**	**0.21**	**0.24**	**0.18**	**0.23**	**0.23**	**0.20**	**-0.09**	-0.10	
Revision 4	De	**0.34**	**0.12**	**0.14**	**0.07**	**0.10**	**0.14**	0.03	**-0.20**	**-0.36**	
Revision 5	De	**0.17**	0.04	0.01	-0.07	-0.02	-0.02	**-0.10**	**-0.43**	**-0.67**	
Revision 6 +	De	**-0.12**	**-0.08**	**-0.19**	**-0.25**	**-0.16**	**-0.24**	**-0.63**	**-0.64**		
Revision Year = 2001	De	**-8.62**	**-6.73**								
Revision Year = 2002	De	**-6.96**	**-4.73**	**-6.37**							
Revision Year = 2003	De	**-5.72**	**-3.30**	**-4.49**	**-6.05**						
Revision Year = 2004	De	**-4.62**	**-2.33**	**-2.99**	**-4.19**	**-5.60**					
Revision Year = 2005	De	**-3.67**	**-1.69**	**-2.08**	**-2.69**	**-3.80**	**-4.90**				
Revision Year = 2006	De	**-2.75**	**-1.12**	**-1.43**	**-1.79**	**-2.32**	**-2.96**	**-3.80**			
Revision Year = 2007	De	**-1.91**	**-0.63**	**-0.83**	**-1.19**	**-1.48**	**-1.59**	**-2.15**	**-3.16**		
Revision Year = 2008	De	**-0.98**	**-0.20**	**-0.39**	**-0.46**	**-0.83**	**-0.70**	**-0.78**	**-1.36**	**-1.90**	
Revision Year = 2009	De										
Insurer A	S										
Insurer B	S	**0.15**	**0.17**	**0.13**	**0.09**	**0.07**	0.03	**0.12**	0.03	**0.37**	**0.60**
Insurer C	S	**0.17**	**0.17**	**0.12**	0.01	-0.00	**0.05**	**0.14**	0.04	**0.36**	0.17
Insurer D	S	**0.22**	**0.26**	**0.14**	**0.17**	**0.10**	**0.08**	**0.18**	**0.13**	**0.47**	0.38
Insurer E	S	**0.07**	**0.14**	**0.11**	**0.08**	**0.06**	0.04	**0.10**	0.01	**0.27**	0.12
Insurer F	S	**0.07**	**0.15**	**0.16**	**0.09**	-0.04	-0.04	**0.10**	0.06	**0.16**	0.07
Insurer Other	S	**0.21**	**0.31**	**0.20**	**0.09**	0.01	0.04	**0.14**	0.06	**0.47**	0.16
Employed	S										
Self Employed	S										
Unemployed	S										
Other	S	**0.01**	0.02	-0.01	-0.01	-0.02	0.02	0.02	0.03	0.01	**0.42**
Age between 0 - 9	S	**0.09**	0.03	**0.17**	**0.08**	0.03	0.04	0.06	0.08	0.05	-0.54
Age between 10 - 16	S	**0.04**	0.01	0.05	0.04	-0.02	0.02	0.01	**0.12**	0.09	-0.78
Age between 17 - 25	S	**0.02**	-0.01	0.01	-0.00	0.00	0.03	-0.02	0.00	-0.04	0.13
Age between 26 - 45	S										
Age between 46 - 65	S	-0.00	-0.01	0.01	-0.01	-0.02	-0.02	-0.00	0.00	-0.01	0.07
Age 66 +	S	**-0.03**	**-0.07**	-0.02	0.01	-0.01	0.00	-0.05	-0.05	-0.10	-0.33
Resides in Metro area	S										
Resides in Outer Metro area	S	**-0.02**	**-0.09**	0.00	0.00	0.01	0.01	-0.03	0.02	0.08	-0.21
Resides in Country area	S										
Resides in Wollongong area	S										
Resides in Newcastle area	S	**-0.02**	0.01	-0.01	-0.03	0.00	**-0.05**	-0.04	-0.02	**0.07**	-0.07
Accident Year = 2001	S										
Accident Year = 2002	S	**-1.00**									
Accident Year = 2003	S	**-1.91**									
Accident Year = 2004	S	**-2.80**									
Accident Year = 2005	S	**-3.71**									
Accident Year = 2006	S	**-4.65**									
Accident Year = 2007	S	**-5.69**									
Accident Year = 2008	S	**-7.06**									
Accident Year = 2009	S	**-8.47**									
Legal Rep at report	S	**0.06**	0.02	0.01	**-0.05**	-0.01	-0.02	**0.09**	**0.22**	**0.16**	0.00
Max. Severity at report	S	**0.03**	0.00	-0.01	-0.01	0.01	0.02	0.01	**0.05**	**0.17**	0.62

*italicised variables represents the baseline value of a categorical variable

Table 7.5: Delay Process by Accident Year - Coefficients1

Delay Model (cont.)											
Model		All	Data segmented by Accident Year								
		Years	2001	2002	2003	2004	2005	2006	2007	2008	2009
Number of Claims		80,998	11,545	10,298	9,588	9,528	9,071	8,695	8,580	8,078	5,615
Number of Revisions		189,507	34,499	29,483	26,704	26,366	23,870	19,666	16,722	10,236	1,961
Number of Parameters		72	63	62	61	60	59	58	57	53	44
Log-likelihood		-267,984	-49,658	-42,195	-38,551	-37,597	-32,845	-26,246	-19,424	-9,265	-949
Parameter	Type	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff
ISS at j - 1 = 0	D	**-0.40**									
ISS at j - 1 = 1	D										
ISS at j - 1 = 2	D	-0.01	-0.01	-0.00	0.00	0.00	-0.02	-0.04	-0.02	0.01	0.04
ISS at j - 1 between 3 and 5	D	**-0.03**	**-0.04**	-0.00	-0.02	-0.02	-0.04	**-0.06**	-0.01	0.03	-0.10
ISS at j - 1 between 6 and 10	D	**-0.08**	-0.06	0.01	-0.07	-0.06	**-0.13**	-0.07	-0.05	**0.28**	-0.34
ISS at j - 1 between 11 and 30	D	**-0.12**	-0.07	-0.04	**-0.10**	-0.09	**-0.14**	**-0.10**	-0.12	**0.26**	-0.22
ISS at j - 1 between 31 and 75	D	**-0.10**	-0.06	0.11	0.00	-0.15	-0.12	-0.18	-0.08	0.10	
Maximum Severity at j - 1 is unknown	D	**0.42**	0.00	0.03	-0.01	0.03	0.03	-0.00	**0.09**	**0.14**	0.28
Maximum Severity at j - 1 = 1	D										
Maximum Severity at j - 1 = 2	D	0.01	0.03	0.02	0.00	-0.01	0.05	0.02	-0.03	**-0.14**	-0.43
Maximum Severity at j - 1 = 3	D	-0.01	0.01	-0.01	0.06	0.00	0.07	-0.02	-0.01	**-0.50**	-1.70
Maximum Severity at j - 1 = 4	D	-0.01	0.04	0.05	0.12	0.07	0.05	-0.02	-0.02	-0.18	-2.07
Maximum Severity at j - 1 = 5	D	**-0.10**	0.09	0.02	0.11	0.05	-0.06	-0.00	-0.03	**-1.02**	-1.62
Maximum Severity at j - 1 = 6	D	**-0.11**	-0.04	-0.11	-0.05	0.11	-0.09	0.17	-0.35	**-0.88**	-2.70
WPI at j - 1 is 10% or less	D										
WPI at j - 1 is more than 10%	D	**-0.22**	**-0.14**	**-0.15**	**-0.19**	**-0.18**	**-0.35**	**-0.13**	-0.08	**0.26**	1.72
WPI at j - 1 is unknown	D	**-0.02**	**-0.08**	-0.02	**-0.08**	-0.03	**-0.08**	0.04	0.03	**0.32**	1.52
Claim Size has been 5 times initial est.	D	**-0.11**	-0.03	**-0.04**	**-0.06**	-0.01	-0.02	0.02	-0.06	0.08	0.66
Payment at j - 1 exceed 70% of Incurred	D	**-0.06**	-0.03	**-0.06**	**-0.15**	**0.06**	0.05	**-0.14**	**-0.28**	**-0.40**	-1.17
Claim duation at j - 1 (log)	D	**-1.30**	**-1.19**	**-1.16**	**-1.18**	**-1.20**	**-1.18**	**-1.11**	**-1.08**	**-0.71**	
Spine Injury at j -1	D	**0.18**	0.02	0.21	-0.00	0.21	-0.02	-0.11	0.08	-0.21	
Legal Rep at j - 1	D	**0.07**	**0.05**	0.03	**0.07**	0.03	**0.05**	-0.01	-0.05	-0.02	0.19
No Litigation at j - 1	D										
NSW Supreme Court at j - 1	D	**-0.60**	**-0.49**	**-0.35**	-0.10	**-0.26**	**-0.30**	-0.17	**-1.66**		
NSW District Court at j - 1	D	**-0.77**	**-0.54**	**-0.52**	**-0.47**	**-0.35**	**-0.38**	**-0.50**	**-0.52**	**-1.04**	
NSW Local Court at j - 1	D	**-0.42**	**-0.24**	**-0.44**	-0.15	-0.18	0.00	**-0.57**	-0.39		
CARS at j - 1	D	**-0.51**	**-0.49**	**-0.44**	**-0.32**	**-0.24**	**-0.31**	**-0.25**	**-0.22**	-0.06	
Accepted Liability at j - 1	D										
Partially Accepted Liability at j - 1	D	**0.04**	0.04	**0.07**	-0.00	0.00	0.01	**0.12**	0.04	-0.03	-0.46
Rejected Liabiliy at j - 1	D	**0.17**	**0.14**	**0.11**	**0.13**	**0.11**	**0.14**	**0.10**	**0.11**	0.09	-0.14
Other at j - 1	D	**0.10**	**0.11**	**0.07**	**0.09**	**0.08**	**0.07**	**0.07**	**0.09**	**0.20**	0.33
Rehab Needed at j - 1	D										
Rehab Probably at j - 1	D	**0.05**	0.00	0.01	0.00	**0.05**	-0.02	0.02	-0.04	0.11	0.00
Rehab Possible at j - 1	D										
Rehab not required at j - 1	D	**-0.03**	0.02	0.03	-0.03	**0.06**	-0.05	-0.01	**-0.36**	0.04	-0.24
Other at j - 1	D	**-0.12**	**-0.06**	-0.02	0.01	-0.00	**-0.11**	**-0.13**	-0.04	**-0.15**	-0.02

*italicised variables represents the baseline value of a categorical variable

Table 7.6: Delay Process by Accident Year - Coefficients2

By examining the covariates for different cohorts of claims and appreciation on the stability of the parameters can be gained.

i) The ϕ_1 parameter is quite stable for different accident years

ii) The dispersion parameter α seems to increase for successive claim cohorts suggesting the volatility reduces.

iii) Most variables lose their levels of significance (due to less claims in each cohort) but the general pattern of the parameters is retained. These variables include revision counter, injury severity, insurer, etc.

iv) The claim duration variable and the litigation level (no litigation, court or CARS) have been very stable across the various accident years.

7.5 Delay Component - Allowing for Censoring

In Section 6.5 we discussed the impact of censored claims. In terms of fitting the model, all the finalised claims are modelled as above; however, each open claim will have an adjustment made to the likelihood calculation. The adjustment captures the probability that for a censored claim the next revision (revision $j'+1$) has not occurred on or before the censoring time T'. This adjustment allows for the fact that $t_{j'+1} > T' - T_{j'}$.

If ℓ_O is the likelihood of an open claims and $\ell_{T'}$ is the likelihood of the observed delays for revisions made prior to the censoring time of T' then $\ell_O = \ell_{T'} + \log P(t_{j'+1} > T' - T_{j'})$. $\ell_{T'}$ is also the complete likelihood for a claim that has finalised prior to the censoring time.

The adjustments to the log-likelihood and its derivatives calculation for the Newton-Raphson's iterative method can be found in Appendix F.2. This is then incorporated in the R code used to fit the model allowing for censoring, which is shown in Appendix G.3. As elements of $P(t_{j'+1} > T' - T_{j'})$ are additive rather than multiplicative, this makes the computation of its logarithm and derivatives more difficult.

In Tables 7.7 and 7.8 we compare the model results for ignoring the issue of censored claims and adjusting the likelihood for censored claims.

Delay Model with Censoring					
Model		Allowing for Censoring		Without Censoring	
Number of Claims		80,998		80,998	
Number of Revisions		189,507		189,507	
Number of Parameters		75		72	
Log-likelihood		-275,623		-267,984	
AIC		551,393		536,112	
BIC		552,050		536,678	
Parameter	Type	coeff	s.e.	coeff	s.e.
Intercept		**6.78**	0.03	**6.46**	0.03
phi_1		**0.24**	0.00	**0.25**	0.00
alpha		**3.76**	0.04	**4.58**	0.06
Previous Direction	F	**0.16**	0.01	**0.08**	0.01
Previous Size	F	**0.07**	0.00	**0.08**	0.00
Previous Incurred Cost (log)	De	**0.05**	0.00	**0.06**	0.00
Revision 1	De				
Revision 2	De	**0.40**	0.01	**0.38**	0.01
Revision 3	De	**0.44**	0.01	**0.43**	0.01
Revision 4	De	**0.32**	0.01	**0.34**	0.01
Revision 5	De	**0.14**	0.02	**0.17**	0.02
Revision 6 +	De	**-0.17**	0.02	**-0.12**	0.02
Revision Year = 2001	De	**-8.89**	0.04	**-8.62**	0.04
Revision Year = 2002	De	**-7.21**	0.02	**-6.96**	0.02
Revision Year = 2003	De	**-5.94**	0.02	**-5.72**	0.02
Revision Year = 2004	De	**-4.81**	0.02	**-4.62**	0.02
Revision Year = 2005	De	**-3.84**	0.02	**-3.67**	0.01
Revision Year = 2006	De	**-2.89**	0.01	**-2.75**	0.01
Revision Year = 2007	De	**-2.02**	0.01	**-1.91**	0.01
Revision Year = 2008	De	**-1.06**	0.01	**-0.98**	0.01
Revision Year = 2009	De				
Reporting Delay (log)	S	**-0.03**	0.01		
Insurer A	S				
Insurer B	S	**0.14**	0.01	**0.15**	0.01
Insurer C	S	**0.12**	0.01	**0.17**	0.01
Insurer D	S	**0.20**	0.01	**0.22**	0.01
Insurer E	S	**0.08**	0.01	**0.07**	0.01
Insurer F	S	**0.06**	0.01	**0.07**	0.01
Insurer Other	S	**0.20**	0.01	**0.21**	0.01
Employed	S				
Self Employed	S				
Unemployed	S				
Other	S			**0.01**	0.01
Age between 0 - 9	S	**0.08**	0.02	**0.09**	0.02
Age between 10 - 16	S	**0.04**	0.01	**0.04**	0.02
Age between 17 - 25	S	**0.02**	0.01	**0.02**	0.01
Age between 26 - 45	S				
Age between 46 - 65	S	0.00	0.01	-0.00	0.01
Age 66 +	S	-0.01	0.01	**-0.03**	0.01
Resides in Metro area	S				
Resides in Outer Metro area	S			**-0.02**	0.01
Resides in Country area	S				
Resides in Wollongong area	S				
Resides in Newcastle area	S			**-0.02**	0.01
Accident Year = 2001	S				
Accident Year = 2002	S	**-1.03**	0.01	**-1.00**	0.01
Accident Year = 2003	S	**-1.97**	0.01	**-1.91**	0.01
Accident Year = 2004	S	**-2.89**	0.01	**-2.80**	0.01
Accident Year = 2005	S	**-3.83**	0.02	**-3.71**	0.02
Accident Year = 2006	S	**-4.81**	0.02	**-4.65**	0.02
Accident Year = 2007	S	**-5.91**	0.02	**-5.69**	0.02
Accident Year = 2008	S	**-7.43**	0.02	**-7.06**	0.03
Accident Year = 2009	S	**-9.32**	0.06	**-8.47**	0.05
Legal Rep at report	S			**0.06**	0.01
Max. Severity at report	S	**0.03**	0.00	**0.03**	0.00

*italicised variables represents the baseline value of a categorical variable

Table 7.7: Delay Process - Censoring1

Delay Model with Censoring (cont.)					
Model		Allowing for Censoring		Without Censoring	
Number of Claims		80,998		80,998	
Number of Revisions		189,507		189,507	
Number of Parameters		75		72	
Log-likelihood		-275,623		-267,984	
AIC		551,393		536,112	
BIC		552,050		536,678	
Parameter	Type	coeff	s.e.	coeff	s.e.
ISS at j - 1 = 0	D	**-0.59**	0.07	**-0.40**	0.08
ISS at j - 1 = 1	D				
ISS at j - 1 = 2	D	-0.01	0.01	-0.01	0.01
ISS at j - 1 between 3 and 5	D	**-0.03**	0.01	**-0.03**	0.01
ISS at j - 1 between 6 and 10	D	**-0.06**	0.02	**-0.08**	0.02
ISS at j - 1 between 11 and 30	D	**-0.10**	0.03	**-0.12**	0.02
ISS at j - 1 between 31 and 75	D	-0.07	0.05	**-0.10**	0.04
Maximum Severity at j - 1 is unknown	D	**0.61**	0.07	**0.42**	0.08
Maximum Severity at j - 1 = 1	D				
Maximum Severity at j - 1 = 2	D	0.00	0.01	0.01	0.01
Maximum Severity at j - 1 = 3	D	-0.02	0.02	-0.01	0.02
Maximum Severity at j - 1 = 4	D	-0.02	0.03	-0.01	0.03
Maximum Severity at j - 1 = 5	D	**-0.11**	0.04	**-0.10**	0.04
Maximum Severity at j - 1 = 6	D	**-0.15**	0.05	**-0.11**	0.05
Body regions injured at j - 1 is unknown	D	**0.02**	0.01		
Body regions injured at j - 1 = 1	D				
Body regions injured at j - 1 = 2	D	0.01	0.01		
Body regions injured at j - 1 = 3	D	-0.01	0.01		
Body regions injured at j - 1 = 4	D	**-0.07**	0.02		
Body regions injured at j - 1 = 5	D	0.01	0.05		
WPI at j - 1 is 10% or less	D				
WPI at j - 1 is more than 10%	D	**-0.25**	0.01	**-0.22**	0.01
WPI at j - 1 is unknown	D	**-0.06**	0.01	**-0.02**	0.01
Claim Size has been 5 times initial est.	D	**-0.09**	0.01	**-0.11**	0.01
Payment at j - 1 exceed 70% of Incurred	D	**-0.15**	0.01	**-0.06**	0.01
Claim duation at j - 1 (log)	D	**-1.30**	0.01	**-1.30**	0.01
Spine Injury at j -1	D	**0.21**	0.06	**0.18**	0.06
Back injury at j - 1	D	**0.02**	0.01		
Legal Rep at j - 1	D	**0.14**	0.01	**0.07**	0.01
No Litigation at j - 1	D				
NSW Supreme Court at j - 1	D	**-0.62**	0.05	**-0.60**	0.05
NSW District Court at j - 1	D	**-0.79**	0.02	**-0.77**	0.02
NSW Local Court at j - 1	D	**-0.42**	0.05	**-0.42**	0.05
CARS at j - 1	D	**-0.53**	0.01	**-0.51**	0.01
Accepted Liability at j - 1	D				
Partially Accepted Liability at j - 1	D	**0.05**	0.01	**0.04**	0.01
Rejected Liabiliy at j - 1	D	**0.18**	0.01	**0.17**	0.01
Other at j - 1	D	**0.10**	0.01	**0.10**	0.01
Rehab Needed at j - 1	D				
Rehab Probably at j - 1	D	**0.04**	0.01	**0.05**	0.01
Rehab Possible at j - 1	D				
Rehab not required at j - 1	D	**-0.02**	0.01	**-0.03**	0.01
Other at j - 1	D	**-0.11**	0.01	**-0.12**	0.01

*italicised variables represents the baseline value of a categorical variable

Table 7.8: Delay Process - Censoring2

It is observed that most coefficients are very similar after allowing for censored claims. However, a few interesting observations are made below.

i) The α parameter is considerably smaller by allowing for censoring, that is, the volatility of the delay process is higher. This stands to reason as censoring would disproportionately "hide" delays that have taken longer than usual compared to those shorter than usual.

ii) The impact of the previous direction, $D_{j-1,i}$, is now much greater, suggesting a previous upward change has a stronger impact in the length the current delay. This is an effect that would not have otherwise being measured properly if censoring is not accounted for.

iii) There are some other minor differences in the dynamic covariates, however, the differences are not great.

7.6 Settlement Component

This section reports the results of modelling the Settlement Component Process. This process takes on the value of 1 if the claim is finalised at the current revision or 0 if the claim remains open at the current revision. By definition, there is only one single instance of $S_{j,i} = 1$ per claim, that is, when the claim is settled and all preceding $S_{j,i}$ takes on the value of 0. As such, the ARMA component of the model is not fit.

Table 7.9 and Table 7.10 show the results of the model fitting of the settlement process. Once again the results are split into two tables by the category of the variables. A new set of dynamic variables have been used in the modelling of the Settlement component process, when one of the dynamic variables changes, a set of indicator variables are used to denote the change. For example, when an injury severity worsens from severity 2 to a severity 3, both the current severity (3) and the indicator variable of "injury severity worsened" are used in the model fit.

We have carried out an LRT to compare the model fit between the model where all the variables are used and the model where only significant parameters are retained. The LRT statistic is 23.0 and the 5% critical region threshold for a χ^2 distribution with 17 degrees of freedom (the number of extra parameters used) is 27.6. Hence, the model with only significant parameters represents a similar fit with fewer parameters and this is also shown through the superior AIC and BIC, as show in the tables below.

Settlement Model					
Model		All Variables		Significant Variables	
Number of Claims		80,998		80,998	
Number of Revisions		189,507		189,507	
Number of Parameters		100		83	
Log-likelihood		-49,087		-49,098	
AIC		98,374		98,362	
BIC		99,160		99,015	
Parameter	Type	coeff	s.e.	coeff	s.e.
Intercept		**2.91**	0.15	**2.94**	0.15
phi(1)					
Current Delay (log)	F	**-0.45**	0.02	**-0.45**	0.02
Previous Size	F	**-0.14**	0.01	**-0.14**	0.01
Previous Direction	F	**-0.26**	0.03	**-0.26**	0.02
Previous Delay (log)	F	**-0.06**	0.02	**-0.06**	0.02
Previous Incurred Cost (log)	De	**-0.10**	0.01	**-0.10**	0.01
Revision 1	De				
Revision 2	De	**-0.66**	0.03	**-0.66**	0.03
Revision 3	De	**-0.91**	0.04	**-0.91**	0.04
Revision 4	De	**-0.98**	0.04	**-0.98**	0.04
Revision 5	De	**-0.99**	0.05	**-0.98**	0.05
Revision 6 +	De	**-0.90**	0.06	**-0.90**	0.06
Revision Year = 2001	De	**-0.57**	0.13	**-0.58**	0.13
Revision Year = 2002	De	-0.11	0.10	-0.12	0.10
Revision Year = 2003	De	**-0.19**	0.08	**-0.20**	0.08
Revision Year = 2004	De	**-0.21**	0.07	**-0.21**	0.07
Revision Year = 2005	De	-0.06	0.06	-0.06	0.06
Revision Year = 2006	De	-0.06	0.05	-0.06	0.05
Revision Year = 2007	De	0.06	0.04	0.06	0.04
Revision Year = 2008	De	**-0.10**	0.03	**-0.10**	0.03
Revision Year = 2009	De				
Reporting Delay (log)	S	**-0.21**	0.02	**-0.20**	0.02
Insurer A					
	S				
Insurer B	S	**-0.46**	0.03	**-0.48**	0.02
Insurer C	S	**-0.17**	0.04	**-0.17**	0.04
Insurer D	S	**-0.30**	0.04	**-0.32**	0.03
Insurer E	S	**0.32**	0.05	**0.29**	0.04
Insurer F	S	**-1.06**	0.03	**-1.09**	0.03
Insurer Other	S	**-0.48**	0.03	**-0.50**	0.03
Gender = Female	S	**-0.04**	0.02	**-0.04**	0.02
Employed	S				
Self Employed					
	S	**-0.08**	0.03	**-0.07**	0.03
Unemployed	S	-0.13	0.07		
Other	S	**0.13**	0.02	**0.13**	0.02
Age between 0 - 9	S	-0.02	0.05	-0.02	0.05
Age between 10 - 16	S	-0.02	0.05	-0.02	0.05
Age between 17 - 25	S	**0.15**	0.02	**0.15**	0.02
Age between 26 - 45	S				
Age between 46 - 65	S	0.03	0.02	0.03	0.02
Age 66 +	S	**0.40**	0.03	**0.40**	0.03
Resides in Metro area	S				
Resides in Outer Metro area	S	0.05	0.03		
Resides in Country area	S	-0.08	0.07		
Resides in Wollongong area	S	0.01	0.06		
Resides in Newcastle area	S	**0.07**	0.02	**0.06**	0.02
Accident Year = 2001	S				
Accident Year = 2002	S	0.05	0.03	0.06	0.03
Accident Year = 2003	S	**0.23**	0.04	**0.24**	0.04
Accident Year = 2004	S	**0.32**	0.05	**0.33**	0.05
Accident Year = 2005	S	**0.39**	0.06	**0.39**	0.06
Accident Year = 2006	S	**0.45**	0.07	**0.45**	0.07
Accident Year = 2007	S	**0.39**	0.08	**0.40**	0.08
Accident Year = 2008	S	**0.40**	0.10	**0.40**	0.10
Accident Year = 2009	S	**0.42**	0.15	**0.41**	0.15
Legal Rep at report	S	**-0.10**	0.03	**-0.09**	0.03
Max. Severity at report	S	**-0.09**	0.01	**-0.09**	0.01

*italicised variables represents the baseline value of a categorical variable

Table 7.9: Settlement Process - Coefficients1

Settlement Model (cont.)					
Model		All Variables		Significant Variables	
Number of claims		80,998		80,998	
Number of Revisions		189,507		189,507	
Number of Parameters		100		83	
Log-likelihood		-49,087		-49,098	
AIC		98,374		98,362	
BIC		99,160		99,015	
Parameter	Type	coeff	s.e.	coeff	s.e.
ISS at j = 0	D	**-0.49**	0.23	**-0.49**	0.23
ISS at j = 1	D				
ISS at j = 2	D	**-0.17**	0.03	**-0.17**	0.03
ISS at j between 3 and 5	D	**-0.27**	0.04	**-0.27**	0.04
ISS at j between 6 and 10	D	**-0.41**	0.06	**-0.40**	0.06
ISS at j between 11 and 30	D	**-0.38**	0.08	**-0.36**	0.08
ISS at j between 31 and 75	D	**-0.26**	0.12	-0.24	0.12
Maximum Severity at j is unknown	D	0.19	0.23	0.19	0.23
Maximum Severity at j = 1	D				
Maximum Severity at j = 2	D	0.06	0.04	0.05	0.04
Maximum Severity at j = 3	D	**0.23**	0.07	**0.21**	0.06
Maximum Severity at j = 4	D	**0.32**	0.10	**0.28**	0.09
Maximum Severity at j = 5	D	**0.27**	0.13	0.23	0.12
Maximum Severity at j = 6	D	**1.02**	0.16	**0.98**	0.16
Body regions injured at j is unknown	D	**-0.12**	0.03	**-0.12**	0.03
Body regions injured at j = 1	D				
Body regions injured at j = 2	D	-0.01	0.03	-0.01	0.03
Body regions injured at j = 3	D	**0.10**	0.03	**0.10**	0.03
Body regions injured at j = 4	D	**0.24**	0.05	**0.23**	0.05
Body regions injured at j = 5	D	0.19	0.11	0.18	0.11
WPI at j is 10% or less	D				
WPI at j is more than 10%	D	**-1.82**	0.03	**-1.82**	0.03
WPI at j is unknown	D	**-6.30**	0.04	**-6.29**	0.04
Claim Size has been 5 times initial est.	D	**0.26**	0.03	**0.25**	0.03
Payment at j exceed 70% of Incurred	D	**0.67**	0.04	**0.67**	0.04
Brain Injury Developed	D	0.03	0.10		
Litigation Level Increased	D	**-0.24**	0.06	**-0.23**	0.06
Legal Rep Appointed	D	**-0.46**	0.04	**-0.46**	0.04
WPI threshold met	D	**0.60**	0.03	**0.61**	0.03
Whiplash injury developed	D	**0.24**	0.03	**0.24**	0.03
Spine Injury developed	D	-0.46	0.38		
ISS increased	D	**0.26**	0.03	**0.23**	0.03
Liability Became Rejected	D	**0.18**	0.05	**0.18**	0.05
Maximum Severity Increased	D	-0.05	0.05		
Liability Became Accepted	D	0.03	0.10		
Rehab Needs Increased	D	0.09	0.10		
Number of body region injured Increased	D	**0.35**	0.03	**0.35**	0.03
Claim duation at j (log)	D	**0.63**	0.04	**0.62**	0.04
Brain Injury at j	D	**-0.18**	0.06	**-0.16**	0.05
Spine Injury at j	D	-0.11	0.16		
Back injury at j	D	-0.09	0.05	**-0.10**	0.02
Whiplash Injury at j	D	-0.00	0.05		
Legal Rep at j	D	**-0.50**	0.04	**-0.51**	0.04
Economic Loss compensation at j	D	**0.12**	0.02	**0.12**	0.02
No Litigation at j	D				
NSW Supreme Court at j	D	0.16	0.10		
NSW District Court at j	D	**-0.12**	0.04	**-0.11**	0.04
NSW Local Court at j	D	**0.32**	0.09	**0.32**	0.09
CARS at j	D	-0.04	0.03		
Accepted Liability at j	D				
Partially Accepted Liability at j	D	0.04	0.03		
Rejected Liabiliy at j	D	**0.21**	0.03	**0.21**	0.03
Other at j	D	**-0.18**	0.04	**-0.18**	0.04
Rehab Needed at j	D				
Rehab Probably at j	D	-0.02	0.03		
Rehab Possible at j	D	0.05	0.03		
Rehab not required at j	D	**-0.15**	0.03	**-0.17**	0.02
Other at j	D	0.00	0.03		

*italicised variables represents the baseline value of a categorical variable

Table 7.10: Settlement Process - Coefficients2

The observations regarding the process variables are as follows.

i) The length of the current delay, $t_{j,i}$, is a significant determinant. A longer delay results in the claim being less likely to be finalised. This suggests that for revisions that follow very closely with the previous revision, the likelihood of the claim being settled is relatively higher.

ii) A large previous revision, $Y_{j-1,i}$, reduces the chance of the claim being finalised at the current revision.

iii) A positive movement at the previous revision, $D_{j-1,i} = 1$, also reduces the chance of the claim being finalised at the current revision. A previous positive movement may mean the claim is still developing.

iv) A larger previous incurred claims cost estimate (log transformed), $\log X_{j-1,i}$, leads to the odds of the claim finalising at the current revision being lower.

v) The number of changes a claim has underwent, j, is also significant, and interesting. It seems the likelihood of the claim settled on the first revision is very high (the baseline value). The likelihood of the claim settled on the second revision drops sharply and then gradually increases as at j increases. Claim settling at the first revision may imply light injuries or simple claims. There is also likely to be a correlation with the current claim duration variable (a dynamic variable), which is highly positive, meaning the more mature the claim is, the more likely the claim is going to settle.

The "Static" or baseline covariates are discussed below.

i) Reporting delay (time between accident and claim lodgement), $t_{0,i}$, is significant, the longer the reporting delay the less likely the claim is settled at a given revision.

ii) Gender plays a small role in the probability of settlement. The small negative coefficient suggests Male claimants have a slightly lower probability of settlement at each revision.

iii) Employment status seems to have a significant impact, especially for the value of "Self-employed" and "Unemployed", where the likelihood of settlement is reduced materially. This may suggest the employed claimants may wish to settle the claim promptly to return to work.

iv) The insurer variable is again a key determinant in the likelihood of settlement. Insurer A is relatively less likely to settle as compared with the other insurers. All else being equal, Insurer E takes the lead as the insurer that is most likely to settle at a given revision and insurer F is the least likely to settle, the ratio of the odds is almost 4 times as great.

v) Age of the claimant and employments status do not offer a clear pattern apart from the retired claimants (age 66 and above). This suggests their claim may be simpler as there usually would not be a economic loss component.

The observations regarding the "Dynamic" variables are discussed below.

i) Claims have had major revisions (incurred more than 5 times the original estimate) are more likely to settle. This makes intuitive sense as these claims have already developed significant from their initial claims cost estimate.

ii) Claims with the Whole Person Impairment (a key criterion in receiving general damages) determined, whether the threshold is met or not, are far more likely to settle than if the WPI level remained unknown.

iii) The various injury related dynamic variables seem to suggest the claims with serious injuries (and worsening injuries) seem to be more likely to settle. This is against intuition as the claimant may be expected to see how the injury worsens before agreeing to a claim settlement.

iv) The Legal Representation appointed is material and significant in its impact on the likelihood of finalisation. Represented claims are far less likely (37%) to settle.

A second set of Settlement Component models are constructed by fitting each accident year separately. These are used to examine the stability of the coefficients for different cohorts of claimants. We have carried out an LRT with the combined model and the cohort based models. The LRT statistic is 1802 and the 5% critical region threshold for a χ^2 distribution with 539 degrees of freedom (the number of extra parameters used) is 594. Hence, the cohort based models represent a better fit to the data, suggesting there are changes to the parameters over time.

Settlement Model											
Model		All	Data segmented by Accident Year								
		Years	2001	2002	2003	2004	2005	2006	2007	2008	2009
Number of claims		80,998	11,545	10,298	9,588	9,528	9,071	8,695	8,580	8,078	5,615
Number of Revisions		189,507	34,499	29,483	26,704	26,366	23,870	19,666	16,722	10,236	1,961
Number of Parameters		83	74	73	72	71	70	69	68	67	58
Log-likelihood		-49,098	-10,654	-8,967	-7,241	-6,340	-5,583	-4,586	-3,163	-1,487	-176
Parameter	Type	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff
Intercept		**2.94**	**1.44**	**2.03**	**3.80**	**3.72**	**4.34**	**2.89**	**2.98**	**3.09**	**3.49**
phi(1)											
Current Delay (log)	F	**-0.45**	**-0.49**	**-0.52**	**-0.47**	**-0.36**	**-0.43**	**-0.24**	**-0.39**	-0.13	-1.04
Previous Size	F	**-0.14**	**-0.15**	**-0.17**	**-0.15**	**-0.09**	**-0.10**	**-0.23**	-0.06	-0.04	0.97
Previous Direction	F	**-0.26**	**-0.17**	**-0.22**	**-0.24**	**-0.45**	**-0.29**	**-0.27**	**-0.21**	-0.31	-0.92
Previous Delay (log)	F	**-0.06**	**-0.07**	**-0.08**	**-0.10**	0.02	**-0.11**	0.03	-0.04	0.30	1.15
Previous Incurred Cost (log)	De	**-0.10**	**-0.11**	**-0.13**	**-0.13**	**-0.13**	**-0.11**	-0.04	-0.03	-0.07	-0.08
Revision 1	De										
Revision 2	De	**-0.66**	**-0.69**	**-0.75**	**-0.71**	**-0.58**	**-0.71**	**-0.43**	**-0.52**	-0.35	**-2.20**
Revision 3	De	**-0.91**	**-1.03**	**-1.07**	**-0.87**	**-0.77**	**-0.84**	**-0.50**	**-0.69**	**-0.68**	**-3.35**
Revision 4	De	**-0.98**	**-1.10**	**-1.23**	**-0.82**	**-0.73**	**-0.92**	**-0.59**	**-0.84**	**-1.22**	
Revision 5	De	**-0.98**	**-1.10**	**-1.16**	**-0.82**	**-0.78**	**-1.00**	**-0.65**	**-0.66**	**-1.62**	
Revision 6 +	De	**-0.90**	**-0.97**	**-1.14**	**-0.61**	**-0.70**	**-0.87**	**-0.99**	**-1.02**	-1.27	
Revision Year = 2001	De	**-0.58**	**0.60**								
Revision Year = 2002	De	-0.12	**0.88**	**0.71**							
Revision Year = 2003	De	**-0.20**	**0.62**	**0.56**	-0.07						
Revision Year = 2004	De	**-0.21**	**0.47**	**0.37**	-0.07	0.03					
Revision Year = 2005	De	-0.06	**0.72**	**0.37**	0.08	0.05	-0.36				
Revision Year = 2006	De	-0.06	**0.58**	**0.55**	0.15	-0.17	**-0.47**	0.08			
Revision Year = 2007	De	0.06	**0.49**	**0.56**	**0.55**	0.02	**-0.42**	0.19	0.18		
Revision Year = 2008	De	**-0.10**	0.35	**0.46**	**0.43**	-0.04	**-0.27**	-0.07	-0.16	-0.27	
Revision Year = 2009	De										
Reporting Delay (log)	S	**-0.20**	**-0.21**	**-0.21**	-0.08	**-0.23**	**-0.33**	**-0.15**	**-0.19**	**-0.63**	**-2.83**
Insurer A	S										
Insurer B	S	**-0.48**	**-0.23**	**-0.51**	**-0.57**	**-0.46**	**-0.60**	**-0.47**	**-0.63**	**-0.78**	-0.03
Insurer C	S	**-0.17**	-0.16	**-0.32**	0.14	**-0.28**	**-0.38**	-0.03	0.15	-0.18	0.11
Insurer D	S	**-0.32**	**-0.47**	**-0.87**	0.00	0.12	-0.01	-0.13	0.01	0.15	-0.02
Insurer E	S	**0.29**	**0.59**	**0.52**	**0.42**	0.15	-0.15	0.11	-0.15	-0.41	1.34
Insurer F	S	**-1.09**	**-1.01**	**-0.94**	**-0.85**	**-1.05**	**-1.15**	**-1.22**	**-1.42**	**-1.65**	**-1.42**
Insurer Other	S	**-0.50**	**-0.34**	**-0.53**	**-0.64**	**-0.58**	**-0.67**	**-0.56**	**-0.46**	-0.24	1.17
Gender = Female	S	**-0.04**	-0.01	**-0.11**	-0.05	**-0.11**	0.07	-0.03	-0.09	0.11	0.41
Employed	S										
Self Employed	S	**-0.07**	-0.02	-0.08	-0.13	-0.17	-0.10	-0.01	0.10	**-0.51**	0.13
Unemployed	S										
Other	S	**0.13**	**0.18**	**0.11**	-0.02	0.06	0.00	0.15	**0.35**	-0.12	-0.09
Age between 0 - 9	S	-0.02	-0.11	0.12	-0.16	-0.12	0.15	-0.20	0.21	0.61	**3.38**
Age between 10 - 16	S	-0.02	-0.11	0.10	0.00	0.10	-0.05	-0.06	-0.25	-0.08	2.69
Age between 17 - 25	S	**0.15**	**0.10**	**0.27**	**0.24**	0.07	0.05	0.08	0.11	0.26	**1.05**
Age between 26 - 45	S										
Age between 46 - 65	S	0.03	0.01	0.08	0.09	-0.04	-0.01	0.01	**-0.19**	**0.36**	**0.84**
Age 66 +	S	**0.40**	**0.39**	**0.48**	**0.54**	**0.39**	**0.40**	**0.33**	0.15	**0.71**	1.27
Resides in Metro area	S										
Resides in Outer Metro area	S										
Resides in Country area	S										
Resides in Wollongong area	S										
Resides in Newcastle area	S	**0.06**	**0.09**	0.09	0.09	0.05	0.05	0.02	0.17	-0.08	0.39
Accident Year = 2001	S										
Accident Year = 2002	S	0.06									
Accident Year = 2003	S	**0.24**									
Accident Year = 2004	S	**0.33**									
Accident Year = 2005	S	**0.39**									
Accident Year = 2006	S	**0.45**									
Accident Year = 2007	S	**0.40**									
Accident Year = 2008	S	**0.40**									
Accident Year = 2009	S	**0.41**									
Legal Rep at report	S	**-0.09**	-0.05	-0.06	**-0.19**	-0.10	-0.11	0.01	-0.07	0.02	-1.20
Max. Severity at report	S	**-0.09**	**-0.07**	-0.05	**-0.14**	**-0.13**	-0.06	0.03	-0.13	-0.01	**-1.31**

*italicised variables represents the baseline value of a categorical variable

Table 7.11: Settlement Process by Accident Year - Coefficients1

Settlement Model (cont.)											
Model		All	Data segmented by Accident Year								
		Years	2001	2002	2003	2004	2005	2006	2007	2008	2009
Number of claims		80,998	11,545	10,298	9,588	9,528	9,071	8,695	8,580	8,078	5,615
Number of Revisions		189,507	34,499	29,483	26,704	26,366	23,870	19,666	16,722	10,236	1,961
Number of Parameters		83	74	73	72	71	70	69	68	67	58
Log-likelihood		-49,098	-10,654	-8,967	-7,241	-6,340	-5,583	-4,586	-3,163	-1,487	-176
Parameter	Type	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff
ISS at j = 0	D	**-0.49**									
ISS at j = 1	D										
ISS at j = 2	D	**-0.17**	**-0.14**	**-0.20**	**-0.28**	-0.02	-0.10	-0.14	-0.16	-0.08	-0.18
ISS at j between 3 and 5	D	**-0.27**	**-0.32**	**-0.22**	**-0.23**	-0.10	-0.18	**-0.33**	-0.23	-0.11	-0.41
ISS at j between 6 and 10	D	**-0.40**	**-0.40**	-0.27	**-0.39**	-0.14	-0.33	**-0.63**	**-0.52**	-0.06	-0.59
ISS at j between 11 and 30	D	**-0.36**	**-0.39**	-0.17	-0.43	-0.11	-0.32	**-0.69**	-0.28	-0.14	2.63
ISS at j between 31 and 75	D	-0.24	-0.41	0.00	-0.29	0.34	-0.20	-0.69	-0.03	0.51	
Maximum Severity at j is unknown	D	0.19	-0.18	-0.05	-0.18	**-0.62**	**-0.34**	-0.09	**-0.42**	**-0.78**	**-2.08**
Maximum Severity at j = 1	D										
Maximum Severity at j = 2	D	0.05	0.05	-0.09	0.10	-0.04	0.07	0.14	0.02	-0.05	0.67
Maximum Severity at j = 3	D	**0.21**	0.13	-0.01	0.27	0.26	0.22	**0.49**	0.01	-0.22	1.37
Maximum Severity at j = 4	D	**0.28**	0.25	0.11	**0.49**	0.24	0.33	0.54	-0.21	-0.04	1.29
Maximum Severity at j = 5	D	0.23	0.40	-0.23	**0.78**	-0.12	-0.34	0.54	-0.50	0.02	
Maximum Severity at j = 6	D	**0.98**	**1.29**	0.56	**1.45**	0.93	**0.97**	0.36	0.27	0.21	5.07
Body regions injured at j is unknown	D	**-0.12**	**-0.17**	**-0.17**	**-0.23**	0.09	**-0.20**	-0.11	-0.07	0.03	1.26
Body regions injured at j = 1	D										
Body regions injured at j = 2	D	-0.01	-0.03	-0.02	-0.00	-0.05	-0.10	0.03	0.07	-0.08	0.08
Body regions injured at j = 3	D	**0.10**	0.05	0.05	0.11	0.07	0.14	0.11	0.11	-0.11	0.04
Body regions injured at j = 4	D	**0.23**	0.20	0.23	0.15	0.17	**0.37**	0.19	0.28	0.13	1.53
Body regions injured at j = 5	D	0.18	**0.56**	0.22	**1.20**	-0.31	0.16	-0.44	0.36	-0.06	-0.14
WPI at j is 10% or less	D										
WPI at j is more than 10%	D	**-1.82**	**-1.91**	**-1.87**	**-1.69**	**-1.61**	**-1.47**	**-1.76**	**-2.04**	**-3.00**	**-5.44**
WPI at j is unknown	D	**-6.29**	**-5.66**	**-5.84**	**-6.20**	**-6.77**	**-6.70**	**-6.62**	**-7.02**	**-7.23**	**-10.57**
Claim Size has been 5 times initial est.	D	**0.25**	**0.23**	**0.42**	**0.27**	**0.17**	0.06	**0.37**	0.11	0.19	-1.42
Payment at j exceed 70% of Incurred	D	**0.67**	**0.64**	**0.58**	**0.63**	**0.68**	**0.84**	**0.71**	**1.01**	0.13	-1.43
Brain Injury Developed	D										
Litigation Level Increased	D	**-0.23**	-0.17	**-0.29**	-0.12	0.25	**-0.50**	**-0.74**	**-0.67**	-0.91	
Legal Rep Appointed	D	**-0.46**	**-0.40**	**-0.27**	**-0.53**	**-0.39**	**-0.68**	**-0.36**	**-0.58**	**-0.81**	-0.52
WPI threshold met	D	**0.61**	**0.86**	**0.86**	**0.43**	**0.51**	**0.40**	**0.28**	**0.37**	0.27	**2.58**
Whiplash injury developed	D	**0.24**	**0.25**	**0.28**	0.14	**0.44**	**0.26**	-0.01	0.05	0.39	0.44
Spine Injury developed	D										
ISS increased	D	**0.23**	**0.24**	**0.24**	**0.36**	**0.33**	**0.26**	0.04	0.17	-0.02	1.58
Liability Became Rejected	D	**0.18**	0.21	**0.34**	0.18	**0.42**	0.04	0.16	0.09	0.28	**-3.69**
Number of body region injured Increased	D	**0.35**	**0.53**	**0.41**	**0.44**	**0.42**	**0.24**	0.11	-0.02	0.17	0.16
Claim duation at j (log)	D	**0.62**	**0.92**	**0.96**	**0.41**	**0.61**	**0.38**	0.26	**0.50**	**0.61**	**1.77**
Brain Injury at j	D	**-0.16**	-0.12	-0.03	**-0.39**	0.01	-0.32	-0.01	-0.08	**-1.10**	-3.66
Back injury at j	D	**-0.10**	-0.09	-0.10	**-0.23**	**-0.13**	**-0.17**	0.02	-0.06	-0.02	**0.80**
Legal Rep at j	D	**-0.51**	**-0.43**	**-0.53**	**-0.49**	**-0.68**	**-0.51**	**-0.45**	**-0.60**	-0.61	0.37
Economic Loss compensation at j	D	**0.12**	0.08	**0.20**	**0.13**	-0.00	0.02	0.02	0.06	0.14	0.21
No Litigation at j	D										
NSW Supreme Court at j	D										
NSW District Court at j	D	**-0.11**	**-0.18**	-0.07	**-0.34**	**-0.24**	-0.09	**0.41**	0.32	**-1.39**	
NSW Local Court at j	D	**0.32**	**0.62**	**0.66**	0.25	**-0.56**	**0.63**	0.57	0.27	**2.08**	
CARS at j	D										
Accepted Liability at j	D										
Partially Accepted Liability at j	D										
Rejected Liabiliy at j	D	**0.21**	**0.22**	**0.23**	**0.46**	0.01	**0.36**	-0.09	0.07	0.04	2.53
Other at j	D	**-0.18**	**-0.47**	**-0.25**	**-0.56**	-0.10	0.02	0.10	0.22	0.26	0.04
Rehab Needed at j	D										
Rehab Probably at j	D										
Rehab Possible at j	D										
Rehab not required at j	D	**-0.17**	**-0.12**	**-0.14**	**-0.19**	**-0.24**	-0.14	**-0.43**	-0.06	0.10	1.32
Other at j	D										

*italicised variables represents the baseline value of a categorical variable

Table 7.12: Settlement Process by Accident Year - Coefficients2

7.7 Direction Component

Table 7.13 and Table 7.14 show the results of the fitting of the direction process. The results are split into two tables by the category of the variables due to the number of variables used in the modelling rather than the tables representing separate analyses.

We have carried out an LRT to compare the model fit between the model where all the variables are used and the model where only significant parameters are retained. The LRT statistic is 21.3 and the 5% critical region threshold for a χ^2 distribution with 14 degrees of freedom (the number of extra parameters used) is 23.7. Hence, the model with only significant parameters represents a similar fit with fewer parameters and this is also shown through the superior AIC and BIC.

Direction Model					
Model		All Variables		Significant Variables	
Number of Claims		80,998		80,998	
Number of Revisions		189,507		189,507	
Number of Parameters		101		87	
Log-likelihood		-85,333		-85,344	
AIC		170,866		170,859	
BIC		171,652		171,535	
Parameter	Type	coeff	s.e.	coeff	s.e.
Intercept		**1.20**	0.11	**1.20**	0.11
phi(1)		**0.10**	0.01	**0.10**	0.01
Current Settlement	F	**-2.63**	0.02	**-2.63**	0.02
Current Delay (log)	F	**0.29**	0.01	**0.29**	0.01
Previous Size	F	**-0.03**	0.01	**-0.03**	0.01
Previous Delay (log)	F	**0.05**	0.01	**0.05**	0.01
Previous Incurred Cost (log)	De	**-0.30**	0.01	**-0.30**	0.01
Revision 1	De				
Revision 2	De	**0.18**	0.03	**0.18**	0.03
Revision 3	De	**0.41**	0.03	**0.41**	0.03
Revision 4	De	**0.50**	0.04	**0.51**	0.04
Revision 5	De	**0.56**	0.04	**0.57**	0.04
Revision 6 +	De	**0.56**	0.05	**0.57**	0.05
Revision Year = 2001	De	**0.49**	0.10	**0.48**	0.10
Revision Year = 2002	De	**0.23**	0.08	**0.22**	0.08
Revision Year = 2003	De	**0.22**	0.07	**0.22**	0.07
Revision Year = 2004	De	**0.28**	0.06	**0.28**	0.06
Revision Year = 2005	De	**0.27**	0.05	**0.26**	0.05
Revision Year = 2006	De	**0.23**	0.04	**0.23**	0.04
Revision Year = 2007	De	0.02	0.03	0.02	0.03
Revision Year = 2008	De	-0.02	0.03	-0.03	0.03
Revision Year = 2009	De				
Reporting Delay (log)	S	**-0.05**	0.02	**-0.05**	0.02
Insurer A	S				
Insurer B	S	**0.24**	0.02	**0.24**	0.02
Insurer C	S	**0.36**	0.03	**0.36**	0.03
Insurer D	S	**0.27**	0.04	**0.28**	0.04
Insurer E	S	**-0.50**	0.03	**-0.50**	0.03
Insurer F	S	**-0.12**	0.03	**-0.12**	0.03
Insurer Other	S	**0.27**	0.03	**0.27**	0.03
Gender = Female	S	-0.01	0.01		
Employed	S				
Self Employed	S	**0.10**	0.03	**0.11**	0.03
Unemployed	S	0.07	0.06		
Other	S	-0.03	0.02		
Age between 0 - 9	S	**-0.38**	0.04	**-0.40**	0.04
Age between 10 - 16	S	**-0.30**	0.04	**-0.31**	0.04
Age between 17 - 25	S	**-0.14**	0.02	**-0.15**	0.02
Age between 26 - 45	S				
Age between 46 - 65	S	**-0.05**	0.02	**-0.05**	0.02
Age 66 +	S	**-0.11**	0.03	**-0.12**	0.03
Resides in Metro area	S				
Resides in Outer Metro area	S	**-0.11**	0.02	**-0.11**	0.02
Resides in Country area	S	-0.03	0.05		
Resides in Wollongong area	S	-0.02	0.04		
Resides in Newcastle area	S	**-0.08**	0.02	**-0.07**	0.02
Accident Year = 2001	S				
Accident Year = 2002	S	0.04	0.03	0.04	0.03
Accident Year = 2003	S	**0.10**	0.03	**0.09**	0.03
Accident Year = 2004	S	**0.22**	0.04	**0.22**	0.04
Accident Year = 2005	S	**0.41**	0.05	**0.41**	0.05
Accident Year = 2006	S	**0.55**	0.06	**0.55**	0.06
Accident Year = 2007	S	**0.71**	0.07	**0.71**	0.07
Accident Year = 2008	S	**0.99**	0.08	**0.99**	0.08
Accident Year = 2009	S	**1.66**	0.12	**1.66**	0.12
Legal Rep at report	S	-0.03	0.03		
Max. Severity at report	S	**0.02**	0.01	**0.02**	0.01

*italicised variables represents the baseline value of a categorical variable

Table 7.13: Direction Process - Coefficients1

Direction Model (cont.)					
Model		All Variables		Significant Variables	
Number of Claims		80,998		80,998	
Number of Revisions		189,507		189,507	
Number of Parameters		101		87	
Log-likelihood		-85,333		-85,344	
AIC		170,866		170,859	
BIC		171,652		171,535	
Parameter	Type	coeff	s.e.	coeff	s.e.
ISS at j = 0	D	-0.04	0.14	-0.04	0.14
ISS at j = 1	D				
ISS at j = 2	D	**0.10**	0.03	**0.10**	0.03
ISS at j between 3 and 5	D	**0.32**	0.03	**0.32**	0.03
ISS at j between 6 and 10	D	**0.46**	0.05	**0.47**	0.05
ISS at j between 11 and 30	D	**0.49**	0.07	**0.48**	0.06
ISS at j between 31 and 75	D	**0.43**	0.10	**0.43**	0.10
Maximum Severity at j is unknown	D	0.12	0.14	0.12	0.14
Maximum Severity at j = 1	D				
Maximum Severity at j = 2	D	-0.05	0.03	-0.05	0.03
Maximum Severity at j = 3	D	-0.06	0.05	-0.07	0.05
Maximum Severity at j = 4	D	0.04	0.08	-0.01	0.07
Maximum Severity at j = 5	D	-0.21	0.11	**-0.28**	0.10
Maximum Severity at j = 6	D	-0.20	0.13	-0.20	0.13
Body regions injured at j is unknown	D	**0.10**	0.03	**0.10**	0.03
Body regions injured at j = 1	D				
Body regions injured at j = 2	D	**-0.06**	0.02	**-0.07**	0.02
Body regions injured at j = 3	D	**-0.07**	0.03	**-0.09**	0.03
Body regions injured at j = 4	D	**-0.10**	0.04	**-0.13**	0.04
Body regions injured at j = 5	D	0.05	0.09	0.02	0.09
WPI at j is 10% or less	D				
WPI at j is more than 10%	D	**1.12**	0.03	**1.14**	0.02
WPI at j is unknown	D	**1.20**	0.02	**1.20**	0.02
Claim Size has been 5 times initial est.	D	**-0.15**	0.02	**-0.14**	0.02
Payment at j exceed 70% of Incurred	D	**2.05**	0.03	**2.04**	0.03
Brain Injury Developed	D	**0.35**	0.09	**0.29**	0.08
Litigation Level Increased	D	**0.41**	0.05	**0.41**	0.05
Legal Rep Appointed	D	**0.19**	0.03	**0.21**	0.03
WPI threshold met	D	0.05	0.04		
Whiplash injury developed	D	-0.04	0.03		
Spine Injury developed	D	0.27	0.31		
ISS increased	D	0.00	0.03		
Liability Became Rejected	D	**0.15**	0.04	**0.15**	0.04
Maximum Severity Increased	D	**0.08**	0.04	**0.08**	0.03
Liability Became Accepted	D	**-0.43**	0.08	**-0.43**	0.08
Rehab Needs Increased	D	**-0.36**	0.07	**-0.36**	0.07
Number of body region injured Increased	D	-0.05	0.03		
Claim duation at j (log)	D	**0.14**	0.03	**0.14**	0.03
Brain Injury at j	D	-0.10	0.05		
Spine Injury at j	D	-0.19	0.15		
Back injury at j	D	0.07	0.04	**0.10**	0.02
Whiplash Injury at j	D	0.03	0.04		
Legal Rep at j	D	**0.81**	0.03	**0.78**	0.02
Economic Loss compensation at j	D	**0.23**	0.02	**0.24**	0.01
No Litigation at j	D				
NSW Supreme Court at j	D	**-0.22**	0.09	**-0.24**	0.09
NSW District Court at j	D	**-0.13**	0.04	**-0.13**	0.04
NSW Local Court at j	D	**-0.27**	0.08	**-0.27**	0.08
CARS at j	D	**0.35**	0.02	**0.35**	0.02
Accepted Liability at j	D				
Partially Accepted Liability at j	D	**-0.11**	0.03	**-0.11**	0.03
Rejected Liabiliy at j	D	**-0.39**	0.03	**-0.40**	0.03
Other at j	D	**0.29**	0.03	**0.28**	0.03
Rehab Needed at j	D				
Rehab Probably at j	D	**0.06**	0.03	**0.06**	0.03
Rehab Possible at j	D	**0.14**	0.02	**0.15**	0.02
Rehab not required at j	D	**0.30**	0.02	**0.30**	0.02
Other at j	D	**-0.20**	0.03	**-0.19**	0.03

*italicised variables represents the baseline value of a categorical variable

Table 7.14: Direction Process - Coefficients2

Examining the Process (F) and Determined (DE) variables, the following observations are drawn.

i) ϕ_1 is slightly positive suggesting there is a small influenced from the direction of the previous movement in the same movement direction.

ii) The settlement indicator, $S_{j,i}$, is the single most powerful variable for the modelling of the direction of change. When the revision is the last, due to an agreement reached between the various parties, the direction is extremely likely to be negative. The odds is 14 times in the favour of a reduction in the claim size. This is may be likely due to the conservative nature of the claims managers, with many claims estimates set as the "worst case scenario" claim size rather than a "best estimate" claim size.

iii) The length of the delay since the last revision, $t_{j,i}$, is also a significant determinant of the likely direction of the current revision. The longer the delay, the more likely the change is an upward change.

iv) The previous incurred size (log-transformed), $\log X_{j-1,i}$, is significant which suggests larger claims are less likely to finalise.

v) The number of revisions a claim has had, j, is also an significant, and consistent, factor. The modelling suggests the later revisions will be more likely to be upward revisions than the earlier revisions.

Examining the "Static" explanatory variables, the following observations are drawn.

i) Reporting delay (log transform) is significant for the overall model, where a longer delay contributes to the probability of a negative revision.

ii) The insurer variable once again shows some interesting results. Relative to Insurer A, Insurers B,C and D seem to have relatively more positive changes and Insurer E and F seem to have relatively fewer upward revisions. This may be an indicator of the conservatism exhibited by the claim management team of the different insurers.

iii) Age of the claimant also exhibits some significant pattern in its impact on the direction process. The younger age groups (< 25) are less likely to exhibit positive movements as opposed to the working ages and close-to-retirement claimants. This may be a case of the more conservative estimates for the younger claimants since economic losses and long term medical care costs may be significant. When the actual outcome turns out better than assumed then a negative revision is likely.

iv) The region variable shows Metro claimants are more likely to have upwards movements.

The "Dynamic" variables as contained in Table 7.14 are discussed below.

i) The "Large Change" variable has a negative coefficient of around -0.15 suggesting then the claims costs is more likely to reduce if the claim has already experienced a "Large Change".

ii) Larger previous incurred costs also suggest the claims cost is more likely to reduced rather than increase.

iii) Legally represented claims tend to have more upward changes; the impact of this variable is relatively big (a coefficient of 0.81).

iv) Claims with economic loss component (correlated with the working ages claimants from the "Static" variables) and the Back/Spinal injury indicator (AIS codes starting with G) increase the likelihood of an upward revision and the pattern is consistent over time.

v) The ISS, or the Injury Severity Score, a composite score designed to allow for multiple injuries have a significant effect on the likelihood of an upward increase. That is, higher the ISS or the severity of the injuries sustained, the more likely it is to have an upward revision.

vi) Claims with a Rejected Liability status or the claims that have become a "rejected" claims at the current revisions are less likely to have an upward movement in the estimated claim size. This stands to reason as the claims cost for a claim that the insurer is not liable for can reduce sharply.

vii) The Rehab variable and the Number of Body Region Injured variable show some weak pattern at the overall level.

A second set of Direction models are constructed by fitting each accident year separately. These are used to examine the stability of the coefficients for different cohorts of claimants. We have carried out an LRT with the combined model and the cohort based models. The LRT statistic is 2342 and the 5% critical region threshold for a χ^2 distribution with 571 degrees of freedom (the number of extra parameters used) is 628. Hence, the cohort based models represent a better fit to the data, suggesting there are changes to the parameters over time.

Direction Model											
Model		All	Data segmented by Accident Year								
		Years	2001	2002	2003	2004	2005	2006	2007	2008	2009
Number of Claims		80,998	11,545	10,298	9,588	9,528	9,071	8,695	8,580	8,078	5,615
Number of Revisions		189,507	34,499	29,483	26,704	26,366	23,870	19,666	16,722	10,236	1,961
Number of Parameters		87	78	77	76	75	74	73	72	71	62
Log-likelihood		-85,344	-16,220	-13,779	-12,005	-11,744	-10,355	-8,422	-7,071	-3,928	-649
Parameter	Type	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff
Intercept		**1.20**	**2.35**	**1.92**	**1.28**	**1.22**	**1.08**	**1.22**	**2.15**	**5.42**	-0.38
phi(1)		**0.10**	**0.14**	**0.12**	**0.10**	**0.09**	**0.09**	**0.11**	**0.04**	**0.06**	0.02
Current Settlement	F	**-2.63**	**-2.64**	**-2.71**	**-2.70**	**-2.67**	**-2.55**	**-2.39**	**-2.30**	**-3.03**	**-2.75**
Current Delay (log)	F	**0.29**	**0.24**	**0.20**	**0.30**	**0.36**	**0.34**	**0.33**	**0.56**	**0.35**	0.08
Previous Size	F	**-0.03**	**-0.10**	**-0.09**	**-0.10**	-0.04	0.02	0.04	**0.17**	0.10	0.07
Previous Delay (log)	F	**0.05**	0.03	**0.08**	**0.13**	0.06	-0.01	0.03	0.04	-0.01	-0.64
Previous Incurred Cost (log)	De	**-0.30**	**-0.30**	**-0.30**	**-0.32**	**-0.28**	**-0.27**	**-0.32**	**-0.34**	**-0.59**	**-0.24**
Revision 1	De										
Revision 2	De	**0.18**	0.12	0.12	**0.36**	**0.45**	**0.21**	0.11	0.06	**-0.35**	0.14
Revision 3	De	**0.41**	**0.33**	**0.39**	**0.63**	**0.68**	**0.46**	**0.34**	**0.33**	-0.12	0.65
Revision 4	De	**0.51**	**0.38**	**0.52**	**0.82**	**0.74**	**0.50**	**0.45**	**0.51**	0.23	
Revision 5	De	**0.57**	**0.54**	**0.57**	**0.79**	**0.70**	**0.55**	**0.65**	**0.94**	0.62	
Revision 6 +	De	**0.57**	**0.48**	**0.71**	**0.75**	**0.77**	**0.73**	**0.80**	**0.86**	0.14	
Revision Year = 2001	De	**0.48**	-0.26								
Revision Year = 2002	De	**0.22**	-0.27	-0.25							
Revision Year = 2003	De	**0.22**	-0.09	-0.04	**0.42**						
Revision Year = 2004	De	**0.28**	0.12	0.24	**0.44**	-0.07					
Revision Year = 2005	De	**0.26**	-0.02	0.22	**0.43**	0.10	0.30				
Revision Year = 2006	De	**0.23**	0.09	0.15	**0.33**	0.02	**0.32**	0.10			
Revision Year = 2007	De	0.02	0.11	-0.03	-0.01	-0.17	0.14	0.11	-0.15		
Revision Year = 2008	De	-0.03	0.06	-0.09	-0.13	**-0.46**	0.11	0.01	**0.15**	-0.03	
Revision Year = 2009	De										
Reporting Delay (log)	S	**-0.05**	**-0.08**	-0.02	0.01	-0.06	**-0.10**	-0.05	0.08	0.18	0.50
Insurer A	S										
Insurer B	S	**0.24**	**0.30**	**0.32**	**0.25**	**0.46**	**0.40**	**0.26**	-0.21	**-0.46**	**1.42**
Insurer C	S	**0.36**	**0.36**	**0.40**	**0.32**	**0.62**	**0.60**	**0.53**	0.18	0.20	**1.82**
Insurer D	S	**0.28**	**0.27**	**0.67**	0.14	**0.49**	**0.26**	**0.35**	**-0.39**	-0.18	0.63
Insurer E	S	**-0.50**	**-0.43**	**-0.25**	**-0.59**	**-0.53**	**-0.59**	**-0.60**	**-0.59**	**-0.60**	-0.14
Insurer F	S	**-0.12**	**0.18**	0.14	-0.13	0.08	-0.16	-0.06	**-0.41**	**-0.93**	0.65
Insurer Other	S	**0.27**	**0.35**	**0.39**	**0.23**	**0.38**	**0.56**	**0.33**	**-0.39**	-0.40	**1.12**
Employed	S										
Self Employed	S	**0.11**	0.10	**0.15**	0.10	**0.16**	0.10	0.09	0.16	0.05	0.35
Unemployed	S										
Other	S										
Age between 0 - 9	S	**-0.40**	**-0.26**	**-0.33**	**-0.38**	**-0.48**	**-0.38**	**-0.32**	**-0.67**	**-1.04**	-0.21
Age between 10 - 16	S	**-0.31**	-0.07	**-0.37**	**-0.25**	**-0.28**	**-0.47**	**-0.28**	**-0.53**	**-0.47**	**-1.00**
Age between 17 - 25	S	**-0.15**	**-0.19**	**-0.12**	**-0.18**	**-0.13**	-0.10	-0.09	-0.13	**-0.28**	-0.11
Age between 26 - 45	S										
Age between 46 - 65	S	**-0.05**	-0.05	-0.05	-0.08	-0.06	-0.05	-0.04	0.03	-0.11	0.20
Age 66 +	S	**-0.12**	**-0.18**	-0.13	-0.06	**-0.20**	-0.14	-0.11	-0.02	**-0.36**	0.44
Resides in Metro area	S										
Resides in Outer Metro area	S	**-0.11**	0.02	-0.07	-0.08	**-0.15**	**-0.29**	**-0.23**	**-0.16**	0.11	-0.30
Resides in Country area	S										
Resides in Wollongong area	S										
Resides in Newcastle area	S	**-0.07**	-0.06	-0.08	-0.07	-0.08	-0.09	-0.10	-0.06	0.04	-0.04
Accident Year = 2001	S										
Accident Year = 2002	S	0.04									
Accident Year = 2003	S	**0.09**									
Accident Year = 2004	S	**0.22**									
Accident Year = 2005	S	**0.41**									
Accident Year = 2006	S	**0.55**									
Accident Year = 2007	S	**0.71**									
Accident Year = 2008	S	**0.99**									
Accident Year = 2009	S	**1.66**									
Max. Severity at report	S	**0.02**	0.05	0.05	0.00	0.01	**0.07**	-0.03	-0.10	-0.04	0.29

*italicised variables represents the baseline value of a categorical variable

Table 7.15: Direction Process by Accident Year - Coefficients1

Direction Model (cont.)											
Model		All				Data segmented by Accident Year					
		Years	2001	2002	2003	2004	2005	2006	2007	2008	2009
Number of Claims		80,998	11,545	10,298	9,588	9,528	9,071	8,695	8,580	8,078	5,615
Number of Revisions		189,507	34,499	29,483	26,704	26,366	23,870	19,666	16,722	10,236	1,961
Number of Parameters		87	78	77	76	75	74	73	72	71	62
Log-likelihood		-85,344	-16,220	-13,779	-12,005	-11,744	-10,355	-8,422	-7,071	-3,928	-649
Parameter	Type	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff
ISS at j = 0	D	-0.04									
ISS at j = 1	D										
ISS at j = 2	D	**0.10**	0.00	0.06	0.12	0.07	0.10	0.23	**0.31**	**0.24**	0.37
ISS at j between 3 and 5	D	**0.32**	**0.28**	**0.35**	**0.30**	**0.32**	**0.35**	0.47	**0.37**	**0.36**	0.44
ISS at j between 6 and 10	D	**0.47**	**0.43**	**0.47**	**0.55**	**0.39**	**0.49**	0.64	**0.61**	**0.56**	0.30
ISS at j between 11 and 30	D	**0.48**	**0.43**	**0.58**	**0.60**	**0.39**	0.33	0.89	**0.66**	**0.75**	0.50
ISS at j between 31 and 75	D	**0.43**	0.44	**0.55**	0.34	0.37	0.44	1.14	0.70	-0.15	-0.41
Maximum Severity at j is unknown	D	0.12	0.13	0.12	-0.11	0.08	**0.21**	0.31	-0.17	-0.15	0.09
Maximum Severity at j = 1	D										
Maximum Severity at j = 2	D	-0.05	-0.09	-0.06	-0.06	0.14	-0.10	-0.18	-0.02	0.03	0.19
Maximum Severity at j = 3	D	-0.07	-0.08	-0.19	-0.17	0.07	-0.06	-0.11	0.11	0.29	-0.07
Maximum Severity at j = 4	D	-0.01	-0.02	-0.10	0.12	0.22	-0.02	-0.32	0.01	0.41	-0.43
Maximum Severity at j = 5	D	**-0.28**	-0.42	**-0.52**	-0.08	0.09	-0.23	-0.65	-0.54	0.98	-1.78
Maximum Severity at j = 6	D	-0.20	-0.36	-0.39	0.09	-0.11	-0.54	-0.38	-0.10	-0.01	0.10
Body regions injured at j is unknown	D	**0.10**	0.08	0.09	**0.22**	**0.18**	0.01	-0.03	0.14	0.24	0.01
Body regions injured at j = 1	D										
Body regions injured at j = 2	D	**-0.07**	-0.07	-0.08	-0.07	**-0.16**	-0.01	-0.09	-0.06	-0.12	-0.15
Body regions injured at j = 3	D	**-0.09**	-0.08	-0.11	-0.06	**-0.19**	0.04	-0.20	-0.05	-0.08	-0.67
Body regions injured at j = 4	D	**-0.13**	-0.13	**-0.22**	-0.11	-0.17	**-0.21**	-0.10	0.04	-0.06	-1.00
Body regions injured at j = 5	D	0.02	0.01	0.13	-0.16	-0.41	-0.20	0.17	**0.73**	0.31	-0.99
WPI at j is 10% or less	D										
WPI at j is more than 10%	D	**1.14**	**1.37**	**1.28**	**1.17**	**1.12**	**0.77**	1.05	**1.20**	**1.39**	1.44
WPI at j is unknown	D	**1.20**	**1.07**	**1.04**	**1.14**	**1.22**	**1.42**	1.68	**1.64**	**1.40**	2.01
Claim Size has been 5 times initial est.	D	**-0.14**	**-0.10**	-0.10	-0.09	**-0.18**	**-0.26**	-0.23	**-0.23**	**-0.31**	0.01
Payment at j exceed 70% of Incurred	D	**2.04**	**1.52**	**1.62**	**1.76**	**2.17**	**2.53**	2.75	**2.90**	**3.33**	2.78
Brain Injury Developed	D	**0.29**	**0.65**	0.35	0.20	0.14	0.22	0.44	-0.10	-0.13	-0.17
Litigation Level Increased	D	**0.41**	0.19	**0.40**	**0.33**	**0.38**	**0.45**	0.90	**0.78**	0.52	
Legal Rep Appointed	D	**0.21**	**0.22**	**0.21**	**0.16**	**0.32**	**0.18**	0.18	**0.23**	0.04	0.59
Liability Became Rejected	D	**0.15**	0.07	**0.29**	0.06	0.15	0.07	0.22	0.07	0.09	-0.51
Maximum Severity Increased	D	**0.08**	0.09	0.03	0.08	-0.03	0.09	-0.01	0.06	**0.37**	-0.19
Liability Became Accepted	D	**-0.43**	-0.20	-0.39	**-0.55**	**-0.52**	**-0.52**	0.07	**-0.58**	**-0.76**	
Rehab Needs Increased	D	**-0.36**	0.87	-0.27	0.28	**-0.54**	-0.40	-0.30	**-0.85**	**-0.42**	-1.72
Claim duation at j (log)	D	**0.14**	-0.10	-0.07	0.11	0.11	0.08	0.29	0.03	**0.36**	0.50
Back injury at j	D	**0.10**	**0.14**	0.09	**0.10**	**0.23**	0.07	0.01	0.00	0.02	0.13
Legal Rep at j	D	**0.78**	**0.56**	**0.65**	**0.92**	**0.73**	**0.96**	0.99	**0.88**	**1.12**	0.79
Economic Loss compensation at j	D	**0.24**	**0.18**	**0.31**	**0.31**	**0.24**	**0.15**	0.20	**0.34**	**0.39**	0.15
No Litigation at j	D										
NSW Supreme Court at j	D	**-0.24**	0.09	-0.08	**-0.57**	-0.13	-0.31	-0.48	-0.52		
NSW District Court at j	D	**-0.13**	-0.16	-0.04	-0.00	0.01	**-0.25**	-0.51	0.04	0.27	
NSW Local Court at j	D	**-0.27**	**-0.36**	-0.15	**-0.41**	-0.18	-0.19	0.31	-0.32	0.67	
CARS at j	D	**0.35**	**0.43**	**0.46**	**0.48**	**0.26**	**0.22**	0.17	**0.26**	0.36	0.30
Accepted Liability at j	D										
Partially Accepted Liability at j	D	**-0.11**	**-0.11**	-0.10	**-0.19**	-0.00	-0.13	-0.06	-0.07	0.09	-0.05
Rejected Liabiliy at j	D	**-0.40**	**-0.29**	**-0.43**	**-0.48**	**-0.40**	**-0.45**	-0.32	**-0.35**	-0.11	-0.02
Other at j	D	**0.28**	**0.24**	**0.31**	**0.18**	**0.27**	**0.56**	0.37	**0.23**	**0.30**	0.61
Rehab Needed at j	D										
Rehab Probably at j	D	**0.06**	**0.21**	0.05	-0.10	0.04	0.17	0.07	-0.19	-0.25	0.46
Rehab Possible at j	D	**0.15**	0.11	**0.28**	0.04	**0.18**	0.11	0.17	-0.02	-0.38	0.41
Rehab not required at j	D	**0.30**	**0.20**	**0.21**	0.12	**0.15**	**0.28**	0.35	0.07	0.18	1.83
Other at j	D	**-0.19**	-0.07	-0.01	**-0.21**	**-0.25**	-0.09	-0.04	**-0.56**	**-1.16**	0.04

*italicised variables represents the baseline value of a categorical variable

Table 7.16: Direction Process by Accident Year - Coefficients2

7.8 Size Component - Gamma vs. Generalised Gamma

As discussed briefly in Section 6.4.4, the size process seems to be different depending on whether the revision is a final negative change ($S_{j,i} = 1$ and $D_{j,i} = 0$) or not. It was believed a generalised gamma distribution, which is a three-parameter distribution, may be more flexible than the standard gamma distribution and may prove to be a better fit to the data.

However, the generalised gamma distribution proved difficult to use in practice. Firstly, convergence is notoriously difficult. Lawless (Lawless [2011]) in his text book states that the generalised gamma distribution can appear to be similar for wildly different parameters. Poor starting values for the Newton-Raphson method would typically send the three parameters to positive infinity or negative infinity.

Secondly, for the ARMA component of the likelihood maximisation, the errors terms of the structure $\frac{(Y-E[Y])}{\sigma_Y}$ needs to be differentiated and used as a part of the derivatives of the log-likelihood. However, the generalised gamma distribution proves difficult mathematically for this purpose.

We have used numerical derivatives (R Package numDeriv) to apply the model using the generalised gamma distribution to examine whether the benefits of applying the generalised gamma distribution over a standard gamma distribution is significant. Sample code to fit the generalised gamma distribution can be found in Appendix G.7. Parallel computing (multi-thread processing) is used to speed up the calculations of the likelihood derivative using numerical methods.

Table 7.17 shows the coefficients and standard errors of the parameters for three models

i) Simple GLM without the ARMA component, this is the process of obtaining the initials values for the GLARMA modelling. A gamma error distribution with a log link function is used.

ii) GLARMA with a gamma distribution.

iii) GLARMA with a generalised gamma distribution. This model is fitted by carrying out grid search for appropriate starting values of the intercept, c and k parameters and then using numerical derivatives to achieve the Newton-Raphson algorithm.

Due to the dimensionality issue of using numerical derivatives only a subset of the data (claims from Accident Year 2003) and subset of the covariates (23 out of 102) are used for all three models.

	GLM - Gamma		GLARMA - Gamma		GLARMA - G.Gamma	
	Coeff.	SE.	Coeff.	SE.	Coeff.	SE.
Intercept	0.80	0.06	0.96	0.06	0.65	0.06
Saving on Finalisation	1.12	0.04	1.15	0.04	1.23	0.04
Current Settlement	-0.11	0.02	-0.10	0.01	-0.10	0.02
Current Direction	-0.29	0.04	-0.31	0.04	-0.32	0.04
Current Delay	0.21	0.01	0.20	0.01	0.21	0.01
Whiplash Injury at j	0.09	0.02	0.11	0.02	0.17	0.02
Large Change	-0.05	0.02	-0.11	0.02	-0.10	0.02
Litigation Increased at j	-0.09	0.05	-0.11	0.04	-0.11	0.04
Liability Rejected at j	0.66	0.04	0.63	0.03	0.64	0.04
Revision 1						
Revision 2	-0.09	0.02	-0.15	0.02	-0.15	0.02
Revision 3	-0.10	0.02	-0.17	0.02	-0.16	0.02
Revision 4	-0.13	0.03	-0.20	0.03	-0.20	0.03
Revision 5	-0.13	0.03	-0.20	0.03	-0.19	0.03
Revision 6 +	-0.25	0.03	-0.31	0.03	-0.31	0.03
Insurer A						
Insurer B	0.13	0.02	0.12	0.02	0.12	0.02
Insurer C	0.02	0.04	0.02	0.03	0.04	0.03
Insurer D	0.24	0.03	0.23	0.03	0.24	0.03
Insurer E	0.03	0.03	0.02	0.03	0.02	0.03
Insurer F	0.23	0.03	0.22	0.03	0.23	0.03
Other Insurer	0.15	0.03	0.14	0.03	0.14	0.03
Previous Incurred Cost (log)	-0.15	0.01	-0.14	0.01	-0.14	0.01
Reporting Delay (log)	0.10	0.01	0.10	0.01	0.09	0.01
phi_1	-	-	0.07	0.01	0.07	0.01
c	0.71	-	0.82	0.01	1.00	0.04
k	-	-	-	-	0.88	0.02

Table 7.17: Gamma vs. Generalised Gamma

It can be observed that the coefficients and the standard errors of the covariates are very similar for the two GLARMA models. If the generalised gamma distribution based GLARMA model has c and k parameters such that their product is not statistically different to the c in the gamma GLARMA model then the generalised gamma and gamma distributions have very similar shape and there is very little additional benefits in adopting a generalised gamma distribution. In this case the product of the c and k parameters from the generalised gamma distribution is 0.88 which is close to the c parameter of the gamma distribution (0.82), the two error distributions are mathematically similar. As such, the standard gamma distribution has been adopted in the modelling of the Size process.

7.9 Size Component - Gamma Response

From the discussion in the previous section, the standard gamma distribution is adopted for the modelling of the Size component process. The results from the modelling of the Size process are contained in Tables 7.18 and 7.19. Due to the hierarchical conditioning defined, the size process is the last to be modelled for the current revision and the outcomes of the settlement and direction processes are known. A "Saving on Finalisation" variable is defined to differentiate those revisions that are the final change with a downward revision and the other changes.

We have carried out an LRT to compare the model fit between the model where all the variables are used and the model where only significant parameters are retained. The LRT statistic is 11.8 and the 5% critical region threshold for a χ^2 distribution with 12 degrees of freedom (the number of extra parameters used) is 21.0. Hence, the model with only significant parameters represents a similar fit with fewer parameters and this is also shown through the superior AIC and BIC.

Size Model					
Model		All Variables		Significant Variables	
Number of Claims		80,998		80,998	
Number of Revisions		189,507		189,507	
Number of Parameters		104		92	
Log-likelihood		-112,339		-112,345	
AIC		224,882		224,870	
BIC		225,684		225,578	
Parameter	Type	coeff	s.e.	coeff	s.e.
Intercept		**1.31**	0.05	**1.31**	0.05
phi(1)		**0.08**	0.00	**0.08**	0.00
c		**0.86**	0.00	**0.86**	0.00
Saving on Finalisations	F	**1.22**	0.02	**1.22**	0.02
Current Direction	F	**0.30**	0.01	**0.30**	0.01
Current Settlement	F	**-0.69**	0.01	**-0.69**	0.01
Current Delay (log)	F	**0.15**	0.01	**0.15**	0.01
Previous Direction	F	**-0.08**	0.01	**-0.08**	0.01
Previous Delay (log)	F	**0.07**	0.01	**0.07**	0.01
Previous Incurred Cost (log)	De	**-0.15**	0.00	**-0.15**	0.00
Revision 1	De				
Revision 2	De	**-0.23**	0.01	**-0.23**	0.01
Revision 3	De	**-0.24**	0.01	**-0.24**	0.01
Revision 4	De	**-0.22**	0.01	**-0.22**	0.01
Revision 5	De	**-0.20**	0.02	**-0.20**	0.02
Revision 6 +	De	**-0.22**	0.02	**-0.22**	0.02
Revision Year = 2001	De	-0.04	0.04	-0.04	0.04
Revision Year = 2002	De	0.04	0.03	0.04	0.03
Revision Year = 2003	De	**0.15**	0.03	**0.15**	0.03
Revision Year = 2004	De	**0.16**	0.02	**0.16**	0.02
Revision Year = 2005	De	**0.16**	0.02	**0.16**	0.02
Revision Year = 2006	De	**0.16**	0.02	**0.16**	0.02
Revision Year = 2007	De	**0.12**	0.01	**0.12**	0.01
Revision Year = 2008	De	**0.17**	0.01	**0.17**	0.01
Revision Year = 2009	De				
Reporting Delay (log)	S	**0.05**	0.01	**0.05**	0.01
Insurer A	S				
Insurer B	S	**0.02**	0.01	**0.02**	0.01
Insurer C	S	**-0.04**	0.01	**-0.04**	0.01
Insurer D	S	**0.10**	0.01	**0.11**	0.01
Insurer E	S	0.01	0.01	0.01	0.01
Insurer F	S	**0.05**	0.01	**0.05**	0.01
Insurer Other	S	**0.04**	0.01	**0.04**	0.01
Gender = Female	S	**-0.05**	0.01	**-0.05**	0.01
Employed	S				
Self Employed	S	**-0.02**	0.01	**-0.02**	0.01
Unemployed	S	**0.08**	0.02	**0.08**	0.02
Other	S	0.00	0.01		
Age between 0 - 9	S	**-0.04**	0.02	**-0.05**	0.02
Age between 10 - 16	S	**-0.05**	0.02	**-0.05**	0.01
Age between 17 - 25	S	**-0.03**	0.01	**-0.03**	0.01
Age between 26 - 45	S				
Age between 46 - 65	S	**-0.03**	0.01	**-0.03**	0.01
Age 66 +	S	**-0.04**	0.01	**-0.04**	0.01
Resides in Metro area	S				
Resides in Outer Metro area	S	**0.03**	0.01	**0.03**	0.01
Resides in Country area	S	-0.01	0.02		
Resides in Wollongong area	S	0.01	0.02		
Resides in Newcastle area	S	0.01	0.01		
Accident Year = 2001	S				
Accident Year = 2002	S	**-0.02**	0.01	**-0.02**	0.01
Accident Year = 2003	S	0.01	0.01	0.00	0.01
Accident Year = 2004	S	0.01	0.02	0.01	0.02
Accident Year = 2005	S	0.03	0.02	0.03	0.02
Accident Year = 2006	S	**0.07**	0.02	**0.07**	0.02
Accident Year = 2007	S	**0.08**	0.03	**0.08**	0.03
Accident Year = 2008	S	**0.30**	0.03	**0.30**	0.03
Accident Year = 2009	S	**0.25**	0.04	**0.24**	0.04
Legal Rep at report	S	**-0.04**	0.01	**-0.04**	0.01
Max. Severity at report	S	**0.02**	0.01	**0.02**	0.01

*italicised variables represents the baseline value of a categorical variable

Table 7.18: Size Process - Coefficients1

Size Model (cont.)					
Model		All Variables		Significant Variables	
Number of Claims		80,998		80,998	
Number of Revisions		189,507		189,507	
Number of Parameters		104		92	
Log-likelihood		-112,339		-112,345	
AIC		224,882		224,870	
BIC		225,684		225,578	
Parameter	Type	coeff	s.e.	coeff	s.e.
ISS at j = 0	D	0.01	0.06	0.02	0.06
ISS at j = 1	D				
ISS at j = 2	D	**-0.02**	0.01	**-0.02**	0.01
ISS at j between 3 and 5	D	-0.01	0.01	-0.01	0.01
ISS at j between 6 and 10	D	**-0.06**	0.02	**-0.06**	0.02
ISS at j between 11 and 30	D	**-0.08**	0.03	**-0.09**	0.03
ISS at j between 31 and 75	D	**-0.16**	0.04	**-0.17**	0.04
Maximum Severity at j is unknown	D	0.07	0.06	0.07	0.06
Maximum Severity at j = 1	D				
Maximum Severity at j = 2	D	**0.05**	0.01	**0.05**	0.01
Maximum Severity at j = 3	D	**0.06**	0.02	**0.06**	0.02
Maximum Severity at j = 4	D	0.02	0.03	0.02	0.03
Maximum Severity at j = 5	D	0.07	0.05	0.07	0.05
Maximum Severity at j = 6	D	0.03	0.05	0.04	0.05
Body regions injured at j is unknown	D	**-0.03**	0.01	**-0.03**	0.01
Body regions injured at j = 1	D				
Body regions injured at j = 2	D	**0.03**	0.01	**0.03**	0.01
Body regions injured at j = 3	D	0.01	0.01	0.02	0.01
Body regions injured at j = 4	D	0.00	0.02	0.00	0.02
Body regions injured at j = 5	D	0.05	0.04	0.05	0.04
WPI at j is 10% or less	D				
WPI at j is more than 10%	D	**-0.68**	0.01	**-0.68**	0.01
WPI at j is unknown	D	**-0.65**	0.01	**-0.65**	0.01
Claim Size has been 5 times initial est.	D	**-0.08**	0.01	**-0.08**	0.01
Payment at j exceed 70% of Incurred	D	**-0.13**	0.01	**-0.13**	0.01
Brain Injury Developed	D	-0.04	0.04		
Litigation Level Increased	D	**0.21**	0.02	**0.21**	0.02
Legal Rep Appointed	D	**0.14**	0.01	**0.14**	0.01
WPI threshold met	D	**0.46**	0.02	**0.46**	0.02
Whiplash injury developed	D	0.01	0.01		
Spine Injury developed	D	0.11	0.14		
ISS increased	D	-0.01	0.01		
Liability Became Rejected	D	**0.07**	0.02	**0.07**	0.02
Maximum Severity Increased	D	**0.06**	0.01	**0.05**	0.01
Liability Became Accepted	D	**0.34**	0.03	**0.34**	0.03
Rehab Needs Increased	D	**0.36**	0.03	**0.36**	0.03
Number of body region injured Increased	D	**0.05**	0.01	**0.05**	0.01
Claim duation at j (log)	D	**-0.04**	0.01	**-0.04**	0.01
Brain Injury at j	D	**0.17**	0.02	**0.16**	0.02
Spine Injury at j	D	-0.04	0.07		
Back injury at j	D	0.02	0.02		
Whiplash Injury at j	D	**-0.08**	0.02	**-0.05**	0.01
Legal Rep at j	D	**0.06**	0.01	**0.06**	0.01
Economic Loss compensation at j	D	0.00	0.01		
No Litigation at j	D				
NSW Supreme Court at j	D	**-0.43**	0.04	**-0.42**	0.04
NSW District Court at j	D	**-0.43**	0.02	**-0.43**	0.02
NSW Local Court at j	D	**-0.30**	0.04	**-0.30**	0.04
CARS at j	D	**-0.07**	0.01	**-0.07**	0.01
Accepted Liability at j	D				
Partially Accepted Liability at j	D	0.02	0.01		
Rejected Liabiliy at j	D	**0.58**	0.01	**0.58**	0.01
Other at j	D	**0.28**	0.01	**0.27**	0.01
Rehab Needed at j	D				
Rehab Probably at j	D	**0.06**	0.01	**0.06**	0.01
Rehab Possible at j	D	**0.05**	0.01	**0.05**	0.01
Rehab not required at j	D	**0.07**	0.01	**0.07**	0.01
Other at j	D	**-0.07**	0.01	**-0.07**	0.01

*italicised variables represents the baseline value of a categorical variable

Table 7.19: Size Process - Coefficients2

The following observations are made for the Process variables.

i) ϕ_1 is 0.08 and suggests past larger movements would cause the current size of change to be bigger.

ii) The coefficient of Direction variable, $D_{j,i}$, is positive but small, suggesting positive movements are of a slightly bigger size.

iii) The coefficient for Settlement, $S_{j,i}$, is negative implying the final (upward) change is relatively smaller than the other changes.

iv) The "Saving on Finalisation" ($S_{j,i} = 1$ and $D_{j,i} = 0$) coefficient value is significant at 1.22 suggesting the final downward movement tends to be 3 times as big as other movements.

v) The lengths of the current and previous delay ($\log t_{j,i}$ and $\log t_{j-1,i}$) are also significant determinants of the size of the revision. The pattern is very clear and consistent across the various accident periods. Longer delays correlate with larger movements.

vi) The previous claim size variable (log-transformed) $\log X_{j-1,i}$ is negative and consistent for the various claim cohorts. This suggests larger previous claim incurred costs lead to smaller sizes of the current revision.

vii) The number of revisions a claim has had, j, is also an significant, and consistent, factor. The modelling suggests the initial changes tend to be larger by around 20% but the subsequent changes are similar in size.

Examining the "Static" explanatory variables, the following observations are made.

i) The coefficient for Reporting Delay is small and positive, suggesting claims with longer reporting delay tend to have somewhat larger claim revisions.

ii) Gender seems to have a small impact on the size of the revision; Male claimants' revisions are around 5% smaller than the Female counterpart.

iii) The insurer variable is a significant factor in the determination of the size of the movements. As compared to Insurer A, insurers B, D and F have slightly larger revisions.

The coefficients for the "Dynamic" variables as contained in Table 7.19 and are briefly discussed below.

i) The "Large Change" variable has a small negative impact, implying the sizes of revisions become smaller if the claims have already had significant revisions.

ii) Claims with a Rejected Liability status has larger revisions as the claims cost may reduce or increase sharply for rejected claims.

iii) The litigation levels are informative on the size of the revisions. Once the claim is being litigated, the movements are considerably smaller. This may due to the fact that at this late stage of the claim process the insurer has a good idea on what are the demands are of the claimant and how much the claims would cost.

iv) Some of the indicator variables denoting a change in the dynamic covariates, such as, WPI threshold is met and rehab needs of the claimant increased, increases the size of the claim revision.

A second set of Size Components models are constructed by fitting each accident year separately. These are used to examine the stability of the coefficients for different cohorts of claimants. We have carried out an LRT with the combined model and the cohort based models. The LRT statistic is 2948 and the 5% critical region threshold for a χ^2 distribution with 613 degrees of freedom (the number of extra parameters used) is 672. Hence, the cohort based models represent a better fit to the data, suggesting there are changes to the parameters over time.

Size Model											
Model		All				Data segmented by Accident Year					
		Years	2001	2002	2003	2004	2005	2006	2007	2008	2009
Number of Claims		80,998	11,545	10,298	9,588	9,528	9,071	8,695	8,580	8,078	5,615
Number of Revisions		189,507	34,499	29,483	26,704	26,366	23,870	19,666	16,722	10,236	1,961
Number of Parameters		92	83	82	81	80	79	78	77	76	69
Log-likelihood		-112,345	-19,361	-16,577	-15,812	-15,143	-13,716	-12,079	-9,784	-7,186	-1,212
Parameter	Type	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff
Intercept		**1.31**	**1.26**	**1.30**	**1.12**	**1.28**	**1.82**	**1.80**	**1.50**	**1.46**	**2.80**
phi(1)		**0.08**	**0.08**	**0.08**	**0.07**	**0.08**	**0.09**	**0.09**	**0.08**	**0.04**	0.07
c		**0.86**	**0.87**	**0.87**	**0.87**	**0.90**	**0.90**	**0.87**	**0.83**	**0.83**	**0.83**
Saving on Finalisations	F	**1.22**	**1.08**	**1.29**	**1.16**	**1.37**	**1.27**	**1.32**	**1.31**	**1.29**	**2.51**
Current Direction	F	**0.30**	**0.31**	**0.38**	**0.33**	**0.39**	**0.27**	**0.32**	**0.32**	**0.10**	**0.68**
Current Settlement	F	**-0.69**	**-0.56**	**-0.69**	**-0.59**	**-0.84**	**-0.76**	**-0.80**	**-0.74**	**-0.71**	**-1.67**
Current Delay (log)	F	**0.15**	**0.18**	**0.19**	**0.16**	**0.16**	**0.09**	**0.08**	**0.10**	**0.09**	-0.21
Previous Direction	F	**-0.08**	**-0.05**	**-0.08**	**-0.07**	**-0.05**	**-0.08**	**-0.13**	**-0.17**	**-0.22**	**-0.38**
Previous Delay (log)	F	**0.07**	**0.05**	**0.07**	**0.07**	**0.05**	**0.06**	**0.04**	**0.07**	0.06	**-0.54**
Previous Incurred Cost (log)	De	**-0.15**	**-0.14**	**-0.14**	**-0.13**	**-0.16**	**-0.18**	**-0.16**	**-0.15**	**-0.11**	**-0.26**
Revision 1	De										
Revision 2	De	**-0.23**	**-0.16**	**-0.23**	**-0.25**	**-0.18**	**-0.25**	**-0.27**	**-0.30**	**-0.26**	**-0.29**
Revision 3	De	**-0.24**	**-0.20**	**-0.26**	**-0.25**	**-0.19**	**-0.33**	**-0.33**	**-0.27**	**-0.20**	-0.41
Revision 4	De	**-0.22**	**-0.22**	**-0.24**	**-0.24**	**-0.20**	**-0.32**	**-0.35**	**-0.27**	-0.11	
Revision 5	De	**-0.20**	**-0.23**	**-0.22**	**-0.22**	**-0.19**	**-0.34**	**-0.37**	-0.12	-0.20	
Revision 6 +	De	**-0.22**	**-0.19**	**-0.22**	**-0.25**	**-0.23**	**-0.42**	**-0.40**	**-0.22**	0.40	
Revision Year = 2001	De	-0.04	-0.08								
Revision Year = 2002	De	0.04	-0.02	-0.12							
Revision Year = 2003	De	**0.15**	0.08	-0.04	0.13						
Revision Year = 2004	De	**0.16**	0.10	-0.02	0.12	0.07					
Revision Year = 2005	De	**0.16**	-0.04	0.01	**0.15**	**0.14**	-0.09				
Revision Year = 2006	De	**0.16**	-0.10	-0.05	**0.20**	**0.19**	-0.05	-0.10			
Revision Year = 2007	De	**0.12**	-0.10	-0.11	0.06	**0.22**	-0.00	**-0.11**	0.05		
Revision Year = 2008	De	**0.17**	-0.00	**-0.16**	0.05	**0.23**	**0.15**	-0.03	0.05	**0.42**	
Revision Year = 2009	De										
Reporting Delay (log)	S	**0.05**	**0.06**	0.00	0.03	**0.06**	0.02	**0.06**	**0.16**	**0.14**	0.40
Insurer A	S										
Insurer B	S	**0.02**	-0.04	0.02	0.01	0.04	0.06	**0.12**	**0.22**	-0.03	0.01
Insurer C	S	**-0.04**	**-0.11**	**-0.10**	-0.05	0.00	0.00	0.02	-0.01	-0.00	0.01
Insurer D	S	**0.11**	-0.02	-0.02	**0.17**	**0.24**	**0.22**	**0.16**	**0.16**	**0.15**	0.01
Insurer E	S	0.01	-0.05	0.03	0.04	0.06	0.02	0.01	**0.12**	0.02	-0.11
Insurer F	S	**0.05**	-0.03	**0.07**	**0.16**	**0.19**	**0.17**	0.07	-0.05	-0.09	-0.15
Insurer Other	S	**0.04**	0.01	0.01	0.04	**0.11**	**0.09**	**0.12**	**0.14**	-0.01	0.33
Gender = Female	S	**-0.05**	**-0.07**	**-0.05**	**-0.05**	**-0.04**	**-0.06**	**-0.05**	0.01	-0.05	-0.01
Employed	S										
Self Employed	S	**-0.02**	-0.03	-0.00	-0.03	-0.02	-0.06	0.03	-0.04	0.01	-0.03
Unemployed	S	**0.08**	0.07	0.08	0.08	0.06	0.07	**0.15**	-0.02	0.14	-0.07
Other	S										
Age between 0 - 9	S	**-0.05**	0.01	-0.03	-0.06	-0.03	**-0.21**	-0.07	-0.01	-0.12	0.04
Age between 10 - 16	S	**-0.05**	**-0.07**	-0.05	-0.04	-0.02	**-0.11**	0.09	**-0.18**	-0.05	0.02
Age between 17 - 25	S	**-0.03**	**-0.06**	-0.02	-0.03	**-0.04**	**-0.05**	0.00	0.04	-0.04	-0.11
Age between 26 - 45	S										
Age between 46 - 65	S	**-0.03**	**-0.06**	-0.03	-0.02	-0.01	-0.03	-0.03	0.02	-0.04	-0.09
Age 66 +	S	**-0.04**	-0.02	0.05	-0.06	-0.02	**-0.06**	**-0.07**	-0.04	**-0.09**	-0.06
Resides in Metro area	S										
Resides in Outer Metro area	S	**0.03**	0.01	0.02	**0.05**	0.04	-0.01	0.01	0.03	0.07	0.01
Resides in Country area	S										
Resides in Wollongong area	S										
Resides in Newcastle area	S										
Accident Year = 2001	S										
Accident Year = 2002	S	**-0.02**									
Accident Year = 2003	S	0.00									
Accident Year = 2004	S	0.01									
Accident Year = 2005	S	0.03									
Accident Year = 2006	S	**0.07**									
Accident Year = 2007	S	**0.08**									
Accident Year = 2008	S	**0.30**									
Accident Year = 2009	S	**0.24**									
Legal Rep at report	S	**-0.04**	0.00	-0.04	-0.03	**-0.09**	**-0.07**	**-0.12**	-0.05	-0.14	**-0.59**
Max. Severity at report	S	**0.02**	-0.01	0.01	0.02	**0.06**	**0.03**	**0.04**	**0.06**	0.01	0.16

*italicised variables represents the baseline value of a categorical variable

Table 7.20: Size Component by Accident Year - Coefficients1

Size Model (cont)											
Model		All	Data segmented by Accident Year								
		Years	2001	2002	2003	2004	2005	2006	2007	2008	2009
Number of Claims		80,998	11,545	10,298	9,588	9,528	9,071	8,695	8,580	8,078	5,615
Number of Revisions		189,507	34,499	29,483	26,704	26,366	23,870	19,666	16,722	10,236	1,961
Number of Parameters		92	83	82	81	80	79	78	77	76	69
Log-likelihood		-112,345	-19,361	-16,577	-15,812	-15,143	-13,716	-12,079	-9,784	-7,186	-1,212
Parameter	Type	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff	coeff
ISS at j = 0	D	0.02									
ISS at j = 1	D										
ISS at j = 2	D	**-0.02**	-0.01	-0.03	**-0.07**	0.02	0.01	-0.02	-0.04	-0.05	0.06
ISS at j between 3 and 5	D	-0.01	0.01	-0.01	-0.04	0.02	0.04	-0.02	-0.01	**-0.11**	-0.06
ISS at j between 6 and 10	D	**-0.06**	0.01	-0.03	**-0.12**	0.01	-0.04	-0.11	-0.12	**-0.32**	**0.18**
ISS at j between 11 and 30	D	**-0.09**	-0.06	-0.05	**-0.20**	0.08	-0.00	-0.16	-0.16	**-0.30**	**0.37**
ISS at j between 31 and 75	D	**-0.17**	-0.13	**-0.23**	-0.22	0.00	-0.05	-0.27	-0.11	**-1.04**	**-1.25**
Maximum Severity at j is unknown	D	0.07	0.06	**0.11**	0.03	**0.11**	0.07	**0.11**	**0.13**	**0.16**	**0.33**
Maximum Severity at j = 1	D										
Maximum Severity at j = 2	D	**0.05**	0.05	**0.08**	0.06	0.02	0.02	0.06	0.03	0.05	0.07
Maximum Severity at j = 3	D	**0.06**	**0.13**	0.08	**0.15**	-0.04	-0.02	0.06	0.03	0.11	**-0.31**
Maximum Severity at j = 4	D	0.02	0.10	0.06	0.13	**-0.20**	-0.00	0.05	-0.09	0.03	**-0.38**
Maximum Severity at j = 5	D	0.07	0.18	**0.23**	0.20	-0.17	-0.23	-0.15	0.30	0.37	-0.07
Maximum Severity at j = 6	D	0.04	**0.42**	0.11	0.06	**-0.41**	0.04	-0.05	**-0.40**	0.12	**-0.71**
Body regions injured at j is unknown	D	**-0.03**	**-0.07**	-0.04	-0.04	-0.02	0.01	-0.02	-0.00	-0.02	**-0.12**
Body regions injured at j = 1	D										
Body regions injured at j = 2	D	**0.03**	0.03	0.03	**0.06**	-0.02	0.00	0.05	0.04	0.02	**0.13**
Body regions injured at j = 3	D	0.02	-0.01	0.02	**0.06**	-0.02	-0.00	0.05	-0.00	0.06	**0.09**
Body regions injured at j = 4	D	0.00	0.05	0.01	0.01	-0.01	-0.03	0.01	0.03	-0.03	0.08
Body regions injured at j = 5	D	0.05	-0.02	0.17	0.10	0.12	-0.02	0.14	-0.11	0.01	**0.26**
WPI at j is 10% or less	D										
WPI at j is more than 10%	D	**-0.68**	**-0.63**	**-0.68**	**-0.77**	**-0.84**	**-0.62**	**-0.67**	**-0.63**	**-0.68**	**-0.62**
WPI at j is unknown	D	**-0.65**	**-0.62**	**-0.65**	**-0.60**	**-0.68**	**-0.63**	**-0.74**	**-0.79**	**-0.62**	**-1.07**
Claim Size has been 5 times initial est.	D	**-0.08**	**-0.10**	**-0.09**	**-0.08**	**-0.07**	**-0.06**	-0.00	0.03	**-0.16**	**0.15**
Payment at j exceed 70% of Incurred	D	**-0.13**	**-0.10**	**-0.15**	**-0.18**	**-0.22**	**-0.21**	**-0.10**	**-0.10**	-0.01	**0.21**
Litigation Level Increased	D	**0.21**	**0.19**	**0.29**	**0.10**	**0.18**	**0.21**	**0.23**	**0.26**	0.07	**0.26**
Legal Rep Appointed	D	**0.14**	0.05	0.04	**0.07**	**0.12**	**0.12**	**0.15**	**0.27**	**0.33**	**0.22**
WPI threshold met	D	**0.46**	**0.56**	**0.54**	**0.57**	**0.56**	**0.30**	**0.31**	**0.28**	**0.45**	**0.15**
Liability Became Rejected	D	**0.07**	**0.14**	0.07	0.07	**0.14**	-0.00	**0.15**	**0.13**	-0.06	**0.24**
Maximum Severity Increased	D	**0.05**	0.02	**0.06**	0.02	**0.07**	**0.11**	0.01	0.06	0.06	**0.23**
Liability Became Accepted	D	**0.34**	0.13	**0.31**	**0.35**	0.09	**0.40**	0.18	**0.58**	**0.48**	**-1.69**
Rehab Needs Increased	D	**0.36**	0.06	-0.11	-0.13	0.15	0.10	0.09	**0.72**	**0.37**	**0.50**
Number of body region injured	D	**0.05**	**0.06**	**0.05**	-0.01	0.01	0.00	**0.07**	**0.06**	**0.18**	**0.13**
Claim duation at j (log)	D	**-0.04**	-0.03	-0.01	0.01	-0.04	-0.03	-0.05	-0.06	**-0.11**	**-0.11**
Brain Injury at j	D	**0.16**	**0.10**	**0.11**	0.09	**0.21**	**0.22**	**0.27**	**0.33**	0.22	**0.27**
Whiplash Injury at j	D	**-0.05**	**-0.07**	**-0.05**	**-0.08**	**-0.07**	-0.03	**-0.05**	-0.05	-0.00	**-0.09**
Legal Rep at j	D	**0.06**	0.04	0.05	0.03	**0.12**	**0.10**	**0.13**	0.07	0.14	**0.69**
No Litigation at j	D										
NSW Supreme Court at j	D	**-0.42**	**-0.42**	**-0.47**	**-0.28**	**-0.31**	**-0.60**	**-0.42**	-0.40	-0.64	
NSW District Court at j	D	**-0.43**	**-0.44**	**-0.42**	**-0.43**	**-0.33**	**-0.48**	**-0.39**	**-0.31**	-0.21	
NSW Local Court at j	D	**-0.30**	**-0.29**	**-0.46**	**-0.33**	**-0.18**	-0.12	**-0.32**	0.00	-0.59	
CARS at j	D	**-0.07**	**-0.07**	**-0.10**	**-0.13**	-0.04	-0.04	0.00	-0.08	**-0.14**	**-0.10**
Accepted Liability at j	D										
Partially Accepted Liability at j	D										
Rejected Liabiliy at j	D	**0.58**	**0.66**	**0.65**	**0.65**	**0.48**	**0.61**	**0.42**	**0.42**	**0.49**	**0.45**
Other at j	D	**0.27**	**0.37**	**0.37**	**0.35**	**0.27**	**0.16**	**0.15**	**0.19**	**0.23**	**0.23**
Rehab Needed at j	D										
Rehab Probably at j	D	**0.06**	**0.07**	-0.02	0.01	0.05	0.06	**0.15**	**0.20**	0.03	**0.24**
Rehab Possible at j	D	**0.05**	**0.12**	0.04	0.01	**0.11**	**0.10**	**0.09**	**0.21**	0.11	**0.54**
Rehab not required at j	D	**0.07**	**0.05**	**0.06**	0.04	**0.09**	**0.12**	**0.19**	**0.22**	0.05	0.06
Other at j	D	**-0.07**	-0.03	**-0.12**	**-0.19**	**-0.09**	**-0.07**	-0.06	0.08	-0.02	**-0.10**

*italicised variables represents the baseline value of a categorical variable

Table 7.21: Size Component by Accident Year - Coefficients2

7.10 Inferences by Combining All Four Components Process

The above sections contain considerable information regarding the way NSW CTP claims develop. It can be overwhelming trying to comprehend all the drivers of the four component processes simultaneously. The fit of the modelling may also be improved by incorporating interaction terms between the covariates or designing splines to reflect the underlying shape of the relationship of the numerical variables.

However, the above modelling process does provide a method of testing inferences regarding the claims development behaviour for the insurance company. One key area of interest is to see how various claim managers handle claims differently. Although there is no such information in the PIR database for the NSW CTP claims information, the Insurer variable is used as demonstration of how the difference claim management behaviour of the various insurers can be investigated.

Naturally, each insurer has different internal claims management policies trying to achieve the best outcomes for the claimant and the insurer and these two seemingly opposite goals may have overlapping areas. For instance, if the claims managers can speed up the authorisation of advanced medical treatments, the injuries of the claimants may stabilise and improve sooner and cost the insurer less money in the long run. The following table compares the various coefficients of the four Claim Development Processes between Insurer A and Insurer D, the latter chosen at random, to see if any inference can be drawn. Insurer A is the baseline level in the models and hence has coefficients of 0; Insurer D's coefficients as well as their standard errors are shown to reflect the magnitude of the difference as well as the level of uncertainty in the estimates.

Process Modelled	Delay t_j,i	Settlement S_j,i	Direction D_j,i	Size Y_j,i
Insurer A	0.00	0.00	0.00	0.00
Insurer D	0.20	-0.32	0.28	0.11
S.E. - Insurer D	0.01	0.03	0.04	0.01

Table 7.22: Comparing Coefficients of Insurer Variable

From Table 7.22, it seems Insurer D, as compared to Insurer A, have longer delay between claim revisions but at the same time, is less likely to finalise the claim at a given revision. This may translate into Insurer D revising its claims less frequently and taking a more relaxed approach to claims management. This suggests the claims management style is quite different between the two insurers. Secondly, Insurer D is more likely to have an upward revision. Thirdly, Insurer D's size of adjustments tend to be larger than those of Insurer A. All of the coefficients are highly significant.

These kinds of analyses are extremely useful. By concentrating on one set of variables of interest across the various processes at one time allows a study of how the set of variables affect the claims development process overall. It offers considerable insights to how the claim characteristics can influence how claims may develop from its lodgement to its finalisation.

One major feature of this modelling framework is the adoption of the number of revisions as a pseudo activity based "time scale" for the modelling of the claim development. It is of interest, then, to examine whether the number of revisions influence the claims behaviour in a coherent manner. The coefficients of the revision variable has been compiled and presented in Table 7.23 below.

Process Modelled	Delay t_j,i	Settlement S_j,i	Direction D_j,i	Size Y_j,i
j = 1	0.00	0.00	0.00	0.00
j = 2	0.40	-0.66	0.18	-0.23
j = 3	0.44	-0.91	0.41	-0.24
j = 4	0.32	-0.98	0.51	-0.22
j = 5	0.14	-0.98	0.57	-0.20
j >= 6	-0.17	-0.90	0.57	-0.22

Table 7.23: Comparing Coefficients of Revision Variable

The number of revision influences the Delay variable significantly, later revisions have a shorter delay compared to the initial revisions. The second and third subsequent revision has the longest delays and it shortens after that. Using the first revision as a baseline, the subsequent revisions have lower probabilities of finalising the claim; however, that probability improves slightly as the claim has more revisions ($j \geq 6$). Later revisions also tend to have a higher chance of being a positive revision. For the size process, however, the data suggests the first revision is on average about 20% larger than the subsequent revisions.

The next chapter will use the modelling results from this chapter to predict the ultimate costs of individual claims. The output of this projection is the claims reserve for all reported claims.

7.11 Stratified Modelling Results

In the previous sections GLARMA models were built to investigate the impacts of the covariates on the component processes. However, any possible interactions between the various covariates were not investigated. The large number of covariates possible makes fitting interactions terms a very time consuming process.

However, industry knowledge regarding the NSW CTP claims suggests the claims behave differently based on injury severity. In this section, we devise separate models for the categories of “Minor”, “Moderate” and “Severe” claims. For Minor claims, the claims data is further split into Legally Represented (Minor R) and Not Represented (Minor NR). For the purposes of these definitions, all claims reported with a maximum injury severity of “3” to “6” are Severe; those with a maximum injury severity of “2” are Moderate and all other claims are “Minor”. Note that minor claims do have the possibly of deteriorating into a more serious injury; however, for the purposes of this modelling analyses, these claims remain in their initial allocated category. Typically, Minor claims consists of around 70% of the total number of claims; while Moderate claims and Severe claims make up

20% and 10% of the claims respectively and are tabulated below.

	Number of Claims and Revisions by Severity and Legal Rep				
Dimension	All	Minor R	Minor NR	Moderate	Severe
Number of Claims	80,998	26,758	30,118	15,036	9,086
Number of Revisions	189,507	70,410	58,356	36,948	23,793

Table 7.24: Tabulation of Number of Claims and Revisions

The process adopted in fitting the models remain the same. A full model with all covariates is fitted at first, then insignificant covariates are removed one by one until all remaining covariates are significant and the model selection criteria of AIC and BIC are optimised. The actual results of the models are contained in Appendix E.

It is clear that claims with different severities have different drivers for their claim development behaviours. Table 7.25 shows the log-likelihood of the fitted models.

	All	Stratified by Severity and Legal Rep.					Likelihood Ratio Test		
Component	Claims	Minor R	Minor NR	Moderate	Severe	Sum	Statistic	df	p-value
Delay with Censoring	-275,623	-103,993	-75,134	-55,968	-39,077	-274,171	2,904	180	0.000
Settlement	-49,098	-16,552	-13,922	-10,029	-7,399	-47,902	2,393	174	0.000
Direction	-85,344	-31,987	-24,548	-16,798	-11,229	-84,562	1,563	181	0.000
Size	-112,345	-37,242	-42,466	-20,858	-9,555	-110,122	4,445	202	0.000

Table 7.25: Tabulation of Log-likelihoods

Since, the model with all claims is “nested” within the aggregate of the stratified models, likelihood ratio tests can be performed compare their model fit. The LRTs are also shown in Table 7.25. In each case, the LRT statistic is significantly large and with a corresponding p-value of zero for the null hypothesis that the two models are similar in terms of model fit. These results suggest the model fit of the stratified models are superior and claims do develop differently based on the severity of the claimant’s injuries and whether the claimant is represented by a lawyer.

7.12 Model Convergence

While we have demonstrated that substantial and complex models can be fit using maximum likelihood estimation method to large datasets based on the Newton-Raphson algorithm, one of the key issues faced in this chapter was ensuring convergence of the GLARMA models. The identifiability issue of the generalised gamma and the gamma distribution was discussed above. Maximum likelihood estimates of the dispersion parameter in the negative binomial distribution (α) and the gamma distribution (c) were also difficult to get convergence for. The key is to select a starting value for the Newton-Raphson iterations that are close to the final parameter value. One strategy adopted is to hold the dispersion parameter constant and let the program optimise the other parameters against the likelihood. A "grid search" is perform to find the approximate value of the dispersion parameter that would maximise the likelihood. This value is then used as a starting value.

Another consideration is whether the model likelihood converged to a global maximum rather than a local maximum. While we cannot be sure the model converged to a global maximum as it would be difficult to prove the likelihood function is globally concave, we have carried a number of checks on the convergence and feel comfortable that the convergence obtained is at a global maximum. These checks include random starting values, stability in estimated parameters across cohorts of claims and the Hessian matrix of the likelihood function is negative definite (i.e., $\frac{\partial^2 \ell}{\partial \delta \partial \delta^T} < 0$) and we are able to estimate the standard errors of the parameter estimates.

The Delay component model allowing for censoring also proved very difficult to obtain converge. While the Delay process without allowing for Censoring may converge after 5 iterations the equivalent model allowing for Censoring may take over 50 iterations. Combined with the added complexity when Censoring is allowed, the modelling takes considerably longer to fit. In this case, we have found optimising the dispersion parameter α and the other parameters in alternating iterations (that is, on odd iterations hold α constant and optimise over the other parameters and vice versa on even iterations) have reduced the iterations required to achieve convergence.

Another issue faced is the computing time required, in particular, while using numerical derivatives to obtain convergence fitting the generalised gamma distribution. While R is quite efficient and performs all calculations by first loading datasets into the computer RAM, some models still took hours and even days to fit. In this case, multi-core (or multi-cpu) computers can reduce the computing time substantially by dividing the total datasets into smaller subsets and performing the calculations on the smaller datasets in parallel and then compiling and results and performing the likelihood optimisation calculations. Parallel processing was used in the fitting of the size component model with the generalised gamma distribution; the numerical derivatives calculations (for the likelihood contributions) used the snowfall package (Knaus [2013]) in R to speed up the calculations.

7.13 Conclusions

In this chapter, we have applied the CDP framework to the NSW CTP dataset. The model parameters for each of the four component processes were estimated using maximum likelihood estimation with Newton-Raphson's iterative method. Convergence was obtained for all the models even though this was more difficult to obtain for some of the models (e.g., Delay component with censoring and Size component process using the generalised gamma distribution) compared to the others. We have shown that while the generalised gamma distribution is more versatile in capturing different distribution "shapes" it did not improve model fit meaningfully for the Size component process of the NSW CTP data to warrant the increased computing time and problems with convergence.

The model parameters have been tabulated and discussed and offer considerable insights into the four component processes and, hence, the overall claim development behaviours for NSW CTP claims. While the discussion in this chapter focuses on the NSW CTP , the CDP framework is equally applicable to other long tailed insurance products and would yield valuable insights into the claim development behaviours of these products.

The next two chapters will use the modelling approach from Chapter 6 and the results from this chapter to develop simpler models from which to project open claims to their ultimate values.

Chapter 8

Individual Claims Projection using CDP

8.1 Introduction

This chapter continues with the discussion of the uses of the Claims Development Processes developed in Chapter 6. In Chapter 7 the model parameters of the component processes provide insights to the claims development behaviour for individual claims, as well as for groups of claims. This chapter explores the Claim Development Process framework as a method for the projection of individual claim trajectories. Chapter 9 discusses the viability of using the CDP as a technique for actuarial valuations, that is, calculate the claims cost a insurer needs to pay in the future for all claims aggregated.

The aim of this chapter is to adapt the results from the modelling carried out in Chapter 7 and apply these in a prediction of ultimate sizes on an individual claims basis. The many positive benefits from obtaining an estimated ultimate claim size at an individual claim level has been discussed in Chapter 4.

The chapter starts by discussing the methodology of ultimate claims cost projection. Numerical methods, not unlike the SCE method discussed in Chapter 4, are used for the projection. This chapter also validates the projection results using the modelling data with the actual claim outcomes in the validation dataset.

8.2 Projection

We have carried out the projection of individual ultimate claim sizes using numerical methods, in particular, Monte Carlo simulation. While it may be possible to algebraically evaluate Equation (6.24) the stochastic nature of the number of revisions until finalisation makes the computation difficult. The contribution to the expected value can only occur for the cases when m_i is known and the conditional hierarchical structure means the expectation needs to be evaluated in a step-wise fashion for each successive value of j. That is, only when $S_j = 1$ can the expected value contributions be calculated for the cases that the claim in question has j revisions; for the cases $S_j = 0$, the process continues to evaluate the other component processes.

The equation that needs to be evaluated to calculate the expected ultimate size of the ith claim is as follows (the subscript i has been dropped for clarity).

$$
\begin{aligned}
\hat{X} =& X_{j'}/P(t_{j'+1} > T - T'|G_{j'}) \sum_{k=T-T'+1}^{\infty} P(t_{j'+1} = k|G_{j'})\times \\
&\Bigg(P(S_{j'+1} = 1|G_{j'}, t_{j'}) \sum_{j=j'+1}^{j'+1} \sum_{l=0}^{1} P(D_j = l|G_{j-1}, t_{j-1}, S_{j-1}) \\
&\int_0^\infty \prod_{j=j'+1}^{j+1} e^{2(D_j-1)Y_j} f(Y_j|G_{j-1}, t_{j-1}, S_{j-1}, D_{j-1})dY_j \\
&+ P(S_{j'+1} = 0|G_{j-1}, t_{j-1}) \sum_{k=1}^{\infty} P(t_{j+1} = k|G_j)\times \\
&\big(P(S_{j'+2} = 1|G_{j'+1}, t_{j'+1}) \sum_{j=j'+1}^{j'+2} \sum_{l=0}^{1} P(D_j = l|G_{j-1}, t_{j-1}, S_{j-1}) \\
&\int_0^\infty \prod_{j=j'+1}^{j+2} e^{2(D_j-1)Y_j} f(Y_j|G_{j-1}, t_{j-1}, S_{j-1}, D_{j-1})dY_j \\
&+ P(S_{j'+2} = 0|G_{j'+1}, t_{j'+1}) \sum_{k=1}^{\infty} P(t_{j+1} = k|G_j)\times \\
&(...) \\
&\big)\Bigg)
\end{aligned}
\tag{8.1}
$$

where the various probability functions are defined in Chapter 6.

The above equation would be difficult to solve; however, the hierarchical structure set up for the component processes presents an easier alternative approach. It allows a step-wise simulation of the claim development revision by revision until the claim is settled; and for each revision the four component processes $\{t_{j,i}$, $D_{j,i}$, $S_{j,i}$ and $Y_{j,i}\}$ are simulated in turn until the claim is settled, i.e., $S_{j,i} = 1$. The projected ultimate claim size is then taken as the empirical mean of the ultimate sizes of the simulated trajectories. This approach is relatively straightforward and would allow practising actuaries to easily adapt this method for their own claims projection. An added benefit of using the simulation approach is that it provides an understanding of the variability of the projected ultimate claim sizes. That is, with a distribution of possible trajectories, prediction intervals can be constructed on a claim by claim basis. This is explored in the next chapter.

Prior to outlining the simulation algorithm in detail, the issue of forecasting "Dynamic Variables" is discussed in some detail.

8.2.1 Dynamic Variables

"Dynamic covariates" are claim characteristics that may change during the life of an insurance claim and typically they change in a stochastic fashion. These dynamic covariates are also some of the more powerful variables that explain claims behaviour; for example, whether the claimant decides to litigate against the insurer or whether the injuries of the claimant has worsened from the time the claims were reported. For the purpose of developing the model framework in Chapter 6 and model fitting in Chapter 7 we have treated these dynamic covariates as exogenous, that is, they are determined outside the system of processes under considerable.

For the purposes of ultimate claim size projection, we are required to project the future values of these dynamic covariates beyond the censoring date. There are a number of approach that can be taken to project dynamic variables into the future and a brief discussion on some of the approaches is provided below.

Firstly, we could treat all the dynamic covariates of interested as additional component processes. For example, we can model the likelihood of the claimant obtaining legal representation in the same fashion as we have modelled any of the four component processes. That is, by conditioning on all past information F_{j-1} and G_{j-1} we can model $P(\text{legrep}_j = 1)$ in the same way as we have modelled $P(D_j = 1)$. This essentially "endogenises" the dynamic covariates into the model framework. This approach, however, would make the modelling framework more complex. Currently, the claims development processes have four component processes; embedding the projection of dynamic variables into the model framework adds as many extra component processes as the number of dynamic covariates being considered. Each dynamic covariate will need to be considered in terms of the optimal response distribution and dependence structure as well as fitting the models to the observed data as we did in Chapters 6 and 7.

Secondly, we can build binomial or multinomial logistic regression models that explains the empirical transition probabilities of moving from one state to another state (e.g., No Legal Representation to Legally Represented) based on other covariates. This is quite a popular techniques used in the actuarial industry. This approach would be very similar to the first approach above, but without the various serial dependence structures that have been used for the component processes. For our projection purposes, this will still involve considerable work to develop models for all the dynamic covariates. This method has been used in Statistical Case Estimation (Greenfield et al. [2011] and Oryzak [2008]), where the claims payments in a workers compensation claim is closely related to a "status" variable that has values of "Working","Injured","In Treatment", etc. and usually a comprehensive GLM is built to explain the transition probabilities of the "status" variable between its various states.

Thirdly, transition matrices can be used to explain the probabilities of moving from one state to another state, from example, from the state "Legrep = 0" to "Legrep = 1" and vice versa. These transition matrices are assumed to be independent of other covariates and adopts the Markov assumption, where the probabilities of transitioning to future states are only dependent on the current state.

In our projection of dynamical variables, we have adopted the third approach with the following alterations. Firstly, we have not attempted to model all dynamic variables using transition matrices. We have selected the dynamic variables with the most explanatory power (from the z-values of the parameters in the previous chapter) to be modelled using transition matrices. That is, for the purposes of demonstrating how ultimate claim sizes are projected, we have adopted reduced models for the component processes. These models are fitted using a reduced set of covariates - namely, all the process variables and static variables but a limited number of dynamic covariates that we have selected based on their strength as explanatory variables for the component process models. The model results of these reduced models are shown in Appendix D. The retained and modelled dynamic covariates are Liability Status, Maximum Injury Severity, WPI Threshold, Legal Representation and Litigation Levels.

Secondly, it is observed that most of the transitions between the various levels of the dynamic variables occur earlier on in the life of the claims, i.e., when j is relatively small. We have tabulated the movements between the various values of each dynamic covariate in the form of transition probabilities by values of j, that is, the probability of a transition from the current state to another state at each claim revision. We have examined the empirical transition probabilities for data up to December 2009 and selected the following j-dependent transitions matrices to be used for projection. The selections are made by grouping the claim revisions which showed similar transition probabilities. The following tables (Table 8.1 to Table 8.5) show the transition matrices derived from the observed data and used in the ultimate size projection. These selected transition matrices do not necessarily have the stationary Markov property; however, given the projected trajectories are relatively short (most claims settle with in 10 revisions) this would not be a critical issue.

j = 1

Liability Status	From	To 0	1	2	3
Accepted	0	**98.0%**	0.0%	0.5%	1.5%
Partial	1	5.0%	**92.0%**	1.5%	1.5%
Rejected	2	8.0%	2.0%	**88.0%**	2.0%
Other/Unknown	3	60.0%	6.0%	16.0%	**18.0%**

j >= 2

Liability Status	From	To 0	1	2	3
Accepted	0	**99.0%**	0.0%	0.0%	1.0%
Partial	1	2.0%	**97.0%**	0.5%	0.5%
Rejected	2	4.0%	1.5%	**94.0%**	0.5%
Other/Unknown	3	36.0%	8.0%	18.0%	**38.0%**

Table 8.1: Transition Matrices for Liability

j <= 2

Max. Severity	From	To 0	1	2	3	4	5	6
Unknown	0	**52.0%**	31.0%	9.0%	5.0%	1.0%	1.0%	1.0%
Severity 1	1	1.0%	**93.0%**	5.0%	1.0%	0.0%	0.0%	0.0%
Severity 2	2	1.0%	3.0%	**84.0%**	11.0%	1.0%	0.0%	0.0%
Severity 3	3	0.0%	1.0%	2.0%	**92.0%**	4.0%	1.0%	0.0%
Severity 4	4	0.0%	1.0%	1.0%	4.0%	**90.0%**	4.0%	0.0%
Severity 5	5	0.0%	0.5%	0.5%	2.0%	4.0%	**92.0%**	1.0%
Severity 6	6	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	**100.0%**

j = 3 and j = 4

Max. Severity	From	To 0	1	2	3	4	5	6
Unknown	0	**59.0%**	27.0%	9.0%	5.0%	0.0%	0.0%	0.0%
Severity 1	1	0.0%	**94.0%**	5.0%	1.0%	0.0%	0.0%	0.0%
Severity 2	2	0.0%	2.0%	**92.0%**	6.0%	0.0%	0.0%	0.0%
Severity 3	3	0.0%	1.0%	1.0%	**97.0%**	1.0%	0.0%	0.0%
Severity 4	4	0.0%	0.0%	0.0%	2.0%	**95.0%**	3.0%	0.0%
Severity 5	5	0.0%	0.0%	0.5%	1.0%	1.0%	**97.0%**	0.5%
Severity 6	6	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	**100.0%**

j >= 5

Max. Severity	From	To 0	1	2	3	4	5	6
Unknown	0	**72.0%**	18.0%	8.0%	1.5%	0.5%	0.0%	0.0%
Severity 1	1	0.0%	**95.0%**	4.5%	0.5%	0.0%	0.0%	0.0%
Severity 2	2	0.0%	2.0%	**94.0%**	4.0%	0.0%	0.0%	0.0%
Severity 3	3	0.0%	0.0%	1.0%	**98.0%**	1.0%	0.0%	0.0%
Severity 4	4	0.0%	0.0%	0.0%	2.0%	**95.0%**	3.0%	0.0%
Severity 5	5	0.0%	0.0%	0.0%	0.5%	1.5%	**98.0%**	0.0%
Severity 6	6	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	**100.0%**

Table 8.2: Transition Matrices for Injury Severity

j = 1

WPI	From	To 0	1	2
More than 10%	0	**91.0%**	3.0%	6.0%
10% or Less	1	8.0%	**78.0%**	14.0%
Unknown	2	36.0%	3.0%	**61.0%**

j >= 2

WPI	From	To 0	1	2
More than 10%	0	**83.0%**	5.0%	12.0%
10% or Less	1	5.0%	**89.0%**	6.0%
Unknown	2	28.0%	6.0%	**66.0%**

Table 8.3: Transition Matrices for WPI

j <= 2

Legal Representation	From	To 0	1
No Representation	0	**86.0%**	14.0%
Legal Rep.	1	1.0%	**99.0%**

j >= 3

Legal Representation	From	To 0	1
No Representation	0	**82.0%**	18.0%
Legal Rep.	1	0.0%	**100.0%**

Table 8.4: Transition Matrices for Legal Representation

j <= 3

Litigation	From	To 0	1	2	3	4
No Litigation	0	**93.0%**	0.0%	1.5%	0.5%	5.0%
Supreme Court	1	1.5%	**98.5%**	0.0%	0.0%	0.0%
District Court	2	0.5%	0.1%	**99.2%**	0.1%	0.1%
Local Court	3	1.5%	1.0%	0.5%	**96.0%**	1.0%
CARS	4	4.0%	0.5%	14.5%	0.5%	**80.5%**

j >= 4

Litigation	From	To 0	1	2	3	4
No Litigation	0	**86.0%**	0.5%	2.0%	0.5%	11.0%
Supreme Court	1	0.5%	**99.0%**	0.3%	0.0%	0.2%
District Court	2	1.0%	0.2%	**98.6%**	0.0%	0.2%
Local Court	3	0.7%	0.2%	1.6%	**96.5%**	1.0%
CARS	4	2.8%	0.3%	7.9%	1.0%	**88.0%**

Table 8.5: Transition Matrices for Litigation Level

For example, there is a distinct behavioural pattern for the appointment of legal representation at the first two revisions, $j \leq 2$. It seems the bulk of the claimants who would ultimately appoint a lawyer to represent them would do so at the first revision. Perhaps this group of claimants has a genuine need for representation, whether it be language barrier or unfamiliarity with the legal framework of NSW CTP scheme and would have done so at an early stage.

These transition matrices allow the simulation of the future states of these dynamic variables by randomly sampling a $U(0,1)$ real number and then compare it to the cumulative probabilities of the transition probabilities to determine the state of the dynamic variable at the next revision. As the transition probabilities are calculated from the data, which have around 200,000 transitions, the levels of uncertainty of the transition probabilities are small. Uncertainty of the projected outcomes, however, will be introduced as a part of the Monte-Carlo simulation process.

As mentioned above, since transition matrices are constructed for a subset of the available dynamic covariates, albeit the more powerful ones, new component processes models are needed to be fit to the data. While the aim was to retain and incorporate the more powerful of the dynamic covariates, the model fit of these "reduced" models will nonetheless be inferior to those presented in Chapter 7. We felt this is appropriate in the context of ultimate claim size projection for two reasons. Firstly, we have chosen to incorporate those dynamic variables that are important to the component processes, this allows the focus of this chapter to remain on the projection methodology without dwelling too long on dynamic covariate projection. Secondly, the majority of the dynamic variables that have been dropped are injury related, such as, "Number of Body Regions Injured","Injury Severity Score","Spinal Injuries", etc. These covariates are likely to be correlated and the transition matrices approach adopted may not adequately capture the relationship between these covariates. For example, a claim may be simultaneously sampled to have a "Yes" for Spinal Injury and a low score for Injury Severity Score, an internal inconsistency. We have chosen to only model the "Maximum Injury

Severity" dynamic variable, which is the industry de facto measurement of injury severity for claim size modelling and it has proven to be a powerful covariate for all the component processes.

The model parameters of these reduced projection component models, along with the various selection criteria, can be found in Appendix D. Table 8.6 shows a summary of the reduced models by comparing the number of covariates used, log-likelihood and BIC between the full models and the reduced models used for projection. Due to the large sample sizes, all the reduced models have likelihoods that are significantly worse than the models fitted using all the covariates using the likelihood ratio test. This is also observed through the substantial deteriorated BIC's in Table 8.6. The deterioration in the Direction component model is particularly significant; investigation reveals it is caused by the inability to use the covariate "Payments exceeding 70% of Incurred at Previous Revision" which has a large impact on the direction of the revision.

Component		Full Models			Reduced Models		LRT
Model	No. of Var	Log-likelihood	BIC	No. of Var	Log-likelihood	BIC	p-value
Delay	73	-275,623	552,050	59	-275,817	552,284	0.0000
Settlement	83	-49,087	98,992	65	-49,562	99,766	0.0000
Direction	86	-85,344	171,535	62	-88,064	176,740	0.0000
Size	90	-112,345	225,578	72	-112,663	226,035	0.0000

Table 8.6: Summary of Reduced Models

8.2.2 Simulation Algorithm

This section describes the simulation algorithm used for the projection of the ultimate claims cost using the claim cost development component processes.

The following processes are carried out for each claim that is open at the censoring date. The claim development of revision j, where $j = j' + 1, j' + 2, ...$, and j' is the number of revisions each claim has had at the censoring date.

i) use $F_{j',i}$ and $G_{j',i}$ to determine the mean parameter of the delay process that occurs some time after the censoring date, that is, $t_{j'+1,i}$ given $t_{j'+1} > T - T'$.

The parameters are $\mu_{j'+1,i} = e^{\beta_\mu Z^T_{\mu,j'+1,i}}$ and α. A random observation is sampled from this truncated NegBin$(\mu_{j'+1,i}, \alpha)$ distribution.

ii) new values are sampled for each of the four dynamic variables using the above transition matrices. The new values are used to update the set of risk characteristics from $G_{j',i}$ to $G_{j'+1,i}$.

iii) use $F_{j',i}$, $G_{j'+1,i}$ and $t_{j'+1,i}$ to determine the $q_{j'+1,i}$ parameter of the finalisation process, $S_{j'+1,i}$, where $\text{logit}(q_{j,i}) = \beta_q Z^T_{q,j,i}$. A random observation from this Bernoulli$(q_{j'+1,i})$ distribution is sampled.

iv) similarly use $F_{j',i}$, $G_{j'+1,i}$, $t_{j'+1,i}$ and $S_{j'+1,i}$ to determine the $p_{j'+1,i}$ parameter of the finalisation process, $D_{j'+1,i}$. A random observation from the Bernoulli$(p_{j'+1,i})$ distribution is sampled as the value for $D_{j'+1,i}$.

v) use $F_{j',i}$, $G_{j'+1,i}$, $t_{j'+1,i}$, $S_{j'+1,i}$ and $D_{j'+1,i}$ to determine the parameters of the size of change process, $Y_{j'+1,i}$. The parameters are $\lambda_{j'+1,i}$ and c and a random observation from the Gamma$(\lambda_{j'+1,i}, c)$ is sampled as the value of $Y_{j'+1,i}$.

vi) the projected incurred claim size at revision $j' + 1$ is then calculated as $X_{j'+1,i} = (X_{j',i})e^{(2D_{j'+1,i}-1)Y_{j'+1,i}}$

vii) with all four process values at revision $j'+1$ now sampled, $F_{j',i}$ is now updated with the new component process values to $F_{j'+1,i}$.

viii) if $S_{j'+1,i} = 0$ then the process goes through the $j' + 2$ revision. $F_{j'+1,i}$ and $G_{j'+1,i}$ are used to determine the mean parameter of the delay component process $t_{j'+2,i}$. The parameters are $\mu_{j'+2,i} = e^{\beta_\mu Z^T_{\mu,j'+2,i}}$ and α. A random observation is sampled from this NegBin$(\mu_{j,i}, \alpha)$ distribution and steps (2) to (6) are repeated with $j = j' + 2$. if $S_{j'+1,i} = 1$ then $X_{j'+1,i}$ is the projected ultimate size for this trajectory. Its value is recorded.

ix) this constitutes one simulation of the ith claim. This processes is repeated for all open claims for the number of simulation required.

We have found a convergence of the mean of the projected ultimate claim sizes for individual claims after around 200 simulations. Around 100 simulation paths of all claims is needed for the total cost of all the open claims as at December 2009 to converge. An upper limit of 50 has also been placed on the number of further revisions a claim can experience after the censoring date. With the average revision taking three quarters, this limit represents the claim would remain open for more than 30 years. this limit has been rarely reached in the simulation.

8.3 Sample Projection Path

This section provides some sample projection paths of two randomly selected claims. These allow the reader to have a visual representation at how the projection takes place and also as a tool to see if the projected claim paths are well behaved or not. These two claims were chosen at random, their projection paths allow an investigation of the reasonableness of the individual projected claims trajectories. If there are any trajectories that were exceptionally high or low then this offers a chance for further investigation.

Figure 8.1 shows a few of the projected sample paths of a claim that occurred in 2006 for a male claimant who was legally represented. The injury sustained is severe, with an ISS of 14 at report which soon worsened to 27 at the time the claim is censored at the end of 2009. The first part of the figure shows the claims path prior to the censor date at the end of 2009 and the second part of the figure shows 4 possible projection paths. These projection paths are briefly discussed.

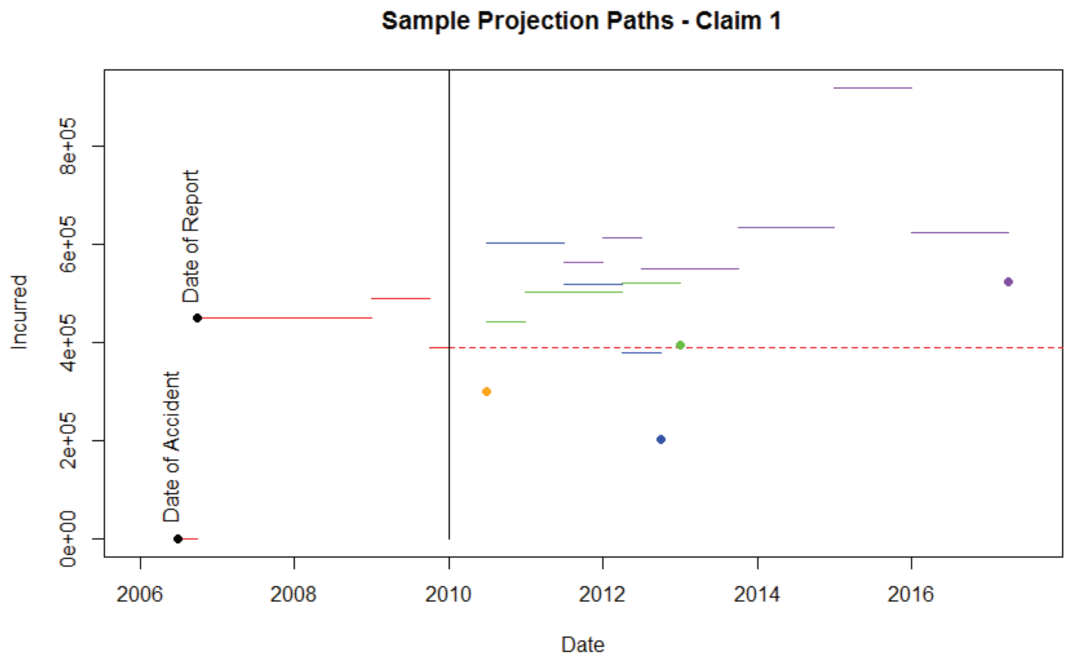

Figure 8.1: Sample Projection Path - Claim 1

Judging from these four projections, which is a extremely small sample for a long tailed CTP claim, the claim is expected to settle after around 3 or 4 revisions in around 3 years or so. A "saving" on finalisation is predicted in all cases. The average of the four estimates of X_i is around the \$350,000 mark, which represents around 30% saving from the initial estimate of \$490,000. The reason for projection path number 4 to rise sharply at the end of 2014 is due to the fact that the projection sampled an increase in the state of the ISS variable for the claim. The ISS was increased from the "11 to 30" level to the "31 and above" level. The change is the injury severity caused the claim estimate to be revised upwards with a magnitude of around 40%.

Using the validation data, however, this claim has not had a revision in the three years from the end of 2009 to the end of 2012, the claim estimate has remained at \$390,000 without any changes to the claims characteristics. Examining the claims further, the claimant has actually suffered a brain injury. Brain injuries are not modelled in this projection model (due to the added complexity of modelling the transitions of patients without brain injury developing a brain injury). From Chapter 7, the brain injury flag as a positive coefficient in the delay model (+0.10) and a negative coefficient in the settlement model (-0.40) suggesting brain injury claims takes longer between revisions and are less likely to settle at any given revision. Hence, brain injuries would take longer to settle than a comparable claim with an injury concerning a different body region. These effects would not be captured in this projection model as Brain Injury dynamic variable is not modelled.

The simulation for this claim was allowed to run for 2,000 times and the results are tallied in the chart below (Figure 8.2). This chart is essentially a heat map, showing the occurrences of various combinations the ultimate claim size and the delay until settlement. The simulation results were rounded, ultimate claim sizes to the nearest \$50,000 and time to settlement to the nearest 2 quarters.

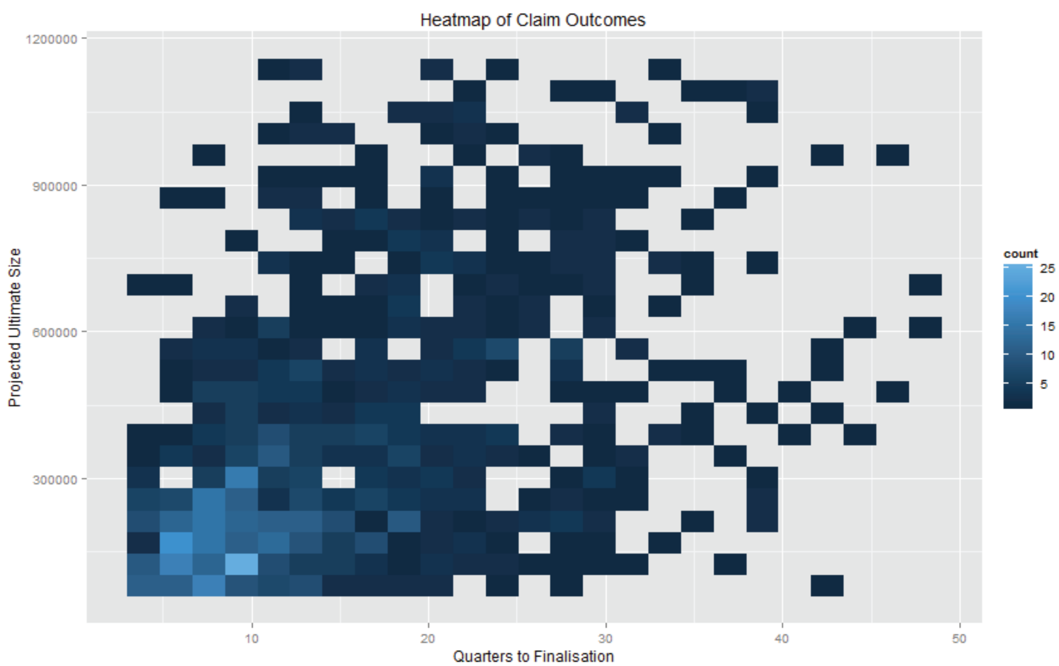

Figure 8.2: Heat Map of Simulation Outcomes - Claim 1

The simulation suggests the more likely outcome of this claim is to be settled within 15 quarters and also for a claim size less than \$300,000. In fact, around 50%

of the simulated scenarios fall into this quadrant. The average ultimate claims cost of these 2,000 simulations is $377,000 and the average delay until settlement is 15 quarters. However, the simulation produces a wide range of possible ultimate sizes and time to finalisation. This is unsurprising given the complexity and variability of CTP claims.

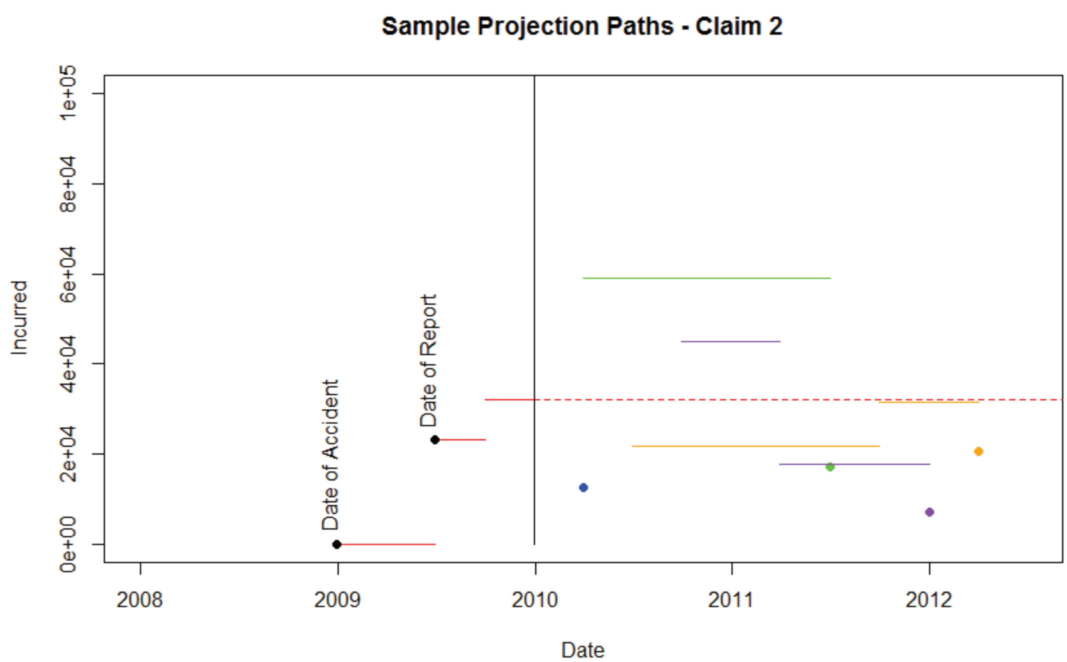

Figure 8.3: Sample Projection Path - Claim 2

Figure 8.3 shows a few of the projected sample paths of a different claim, this time the claim is less severe compared to the first claim. This is a claim that has an unknown severity when reported in April 2009. It had an estimate of $23,000 which was quickly escalated to $32,000. The four sample paths show that this claim is expected to be settled quickly and for a considerably saving compared to the current estimate. However, a 2,000 simulation average actually suggest this claim would increase to an ultimate claim size of $47,000. Coincidentally, this claim was settled within the 3 year validation data period, for an amount of $6,000. While the actual cost of $6,000 is lower than any of the four sample paths presented, it was within the 95% prediction interval for this claim.

8.4 Overall Projection Result

Using the modelling data, that is, claims occurred from 2001 to 2009, observed to December 2009, there were 17,554 open non-nil claims. The total incurred costs at the censoring date of 31 December 2009 was \$4,305m and the total projected ultimate claims cost is \$4,273m. The projected ultimate cost of these open claims are slightly lower than their incurred cost at the censoring date. Table 8.7 shows the projection results by accident year.

Accident Year	Number of Open Claims	Incurred as at Dec2009 (\$m)	Projected Ult. Claims Cost (\$m)
2001	325	150	133
2002	323	147	126
2003	467	209	186
2004	605	290	249
2005	1,010	444	395
2006	1,898	699	637
2007	3,354	873	870
2008	4,806	899	985
2009	4,766	595	691
Total	17,554	4,305	4,273

Table 8.7: Overall Projection Results

The projected claims cost for all accident years apart from the two most recent years are projected to reduce from the incurred cost at the censoring date. This most likely is a feature of the “saving on finalisation” behaviour observed in the previous chapter. The two most recent accident years will however increase considerably from the incurred cost at the censoring date.

The usage of these projection results as a mean of actuarial claims reserving is left until the next chapter. The remainder of this chapter focuses on validating the accuracy of the projection on an individual claims level.

8.5 Validation of Projected Ultimate Claim Size

In this section, the projected claim sizes from applying the Claim Development Process framework to the modelling dataset are compared to the actual results from the validation dataset. Out of the 17,544 open claims as at December 2009, 14,134 (around 80%) were finalised over the next 3 years, that is, within the validation dataset. One way to validate the accuracy of the projection is to compared it to the actual finalised claim size for those claims that did finalise. The comparison is shown in Table 8.8.

Accident Year	Number of Claims Finalised	Projected Ult. Claims Cost ($m)	Actual Finalised Claims Cost ($m)	Difference ($m)	Difference (%)
2001	224	99	101	2	2%
2002	247	102	113	12	11%
2003	328	144	153	8	6%
2004	465	188	189	1	0%
2005	829	322	299	-23	-7%
2006	1,645	542	434	-108	-20%
2007	2,869	754	622	-132	-18%
2008	3,978	788	635	-153	-19%
2009	3,549	495	401	-95	-19%
Total	14,134	3,435	2,947	-488	-14%

Table 8.8: Comparison of Projection Sizes vs. Actual Sizes

From Table 8.8 we can see for the claims that were finalised in the validation data period, the projected results were considerably higher than the actual results. This is especially so for the last 4 accident periods, where the projected results were consistently 20% higher than the actual finalised claims costs for those claims that did finalise from 2010 to 2012.

Another method to validate how predictive the projected ultimate claim sizes are of actual ultimate sizes is to use a gains chart. The idea is to sort the claims with the highest predicted claims cost to the lowest predicted claims cost and then chart cumulative actual cost for these claims. A straight diagonal lines means the projected claim sizes have no ability to predict actual claims cost while a line further away from the diagonal lines means the projected claim sizes is a good method to differentiate the smaller claims from larger claims. Gains charts only form a view on the ability to differentiate the claims with higher actual cost from those with lower actual cost. It does not consider any bias that may exist in the projections.

Figure 8.4 graphs the gains curve for the Claims Development Processes (CDP) projection, the projection of the GLM method from Chapter 4 and the usage of incurred cost at the censoring date as a predictor of claim size.

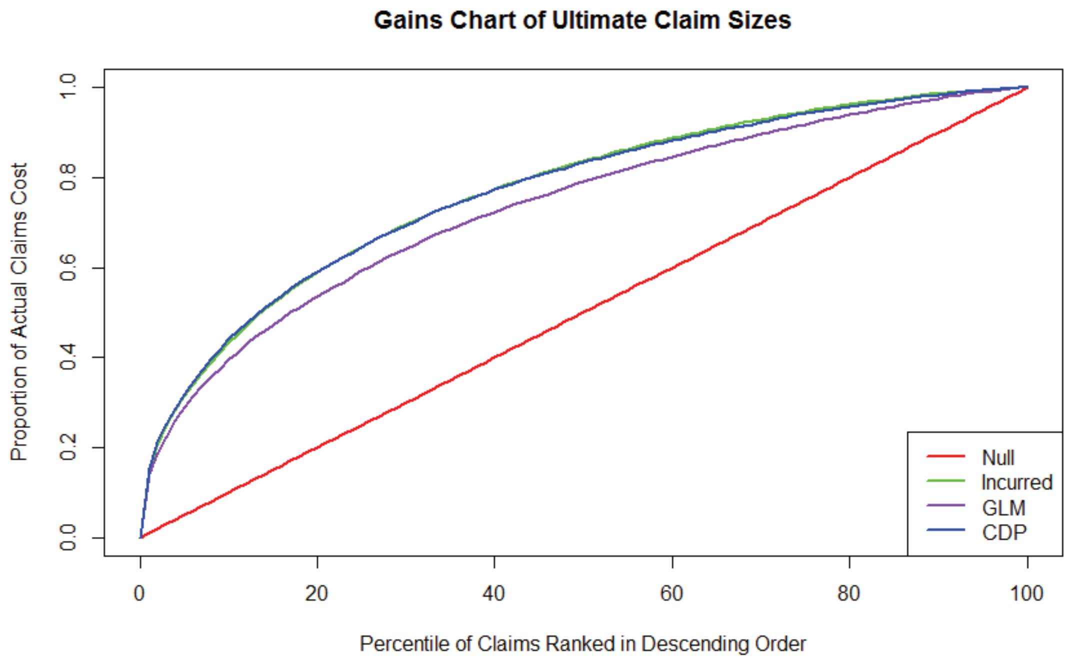

Figure 8.4: Gains Chart

From the diagram, it seems the CDP projections are not any better as a claims cost differentiater compared to using the incurred cost as at the censoring date for the open claims. This may be disheartening but there may be a few explanations to this.

- The NSW CTP claim managers employed by the insurers are experienced and very good at what they do in predicting the ultimate claim size. While the CDP framework does not yield additional benefit for the ultimate claim size prediction, it may work better in other portfolios.

- The cost of the settled claims within the validation period is dominated by a small number of very large claims from the older accident periods. Due to their maturity, the incurred cost is a good predictor of their ultimate claim size. This in turn shrouds the ability of the method in differentiating the claims cost of the relatively smaller and relatively newer claims.

- The CDP models need to be improved further and that can improve its predictive accuracy. This will be considered in a later section where the complete dataset is stratified by Injury Severity and Legal Representation and separate projection models are constructed.

The table below shows the correlation between each of the projection methods above and the actual finalised claim sizes. The correlations correspond well with the above gains chart, that the performances of using the CDP framework and the Incurred Cost as predictors of ultimate claim sizes are similar; while using GLMs lags behind.

Correlation with	CDP	GLM	Incurred
Actual Finalised Cost	0.87	0.77	0.85

Table 8.9: Comparison of Correlation with Actual Sizes

8.6 Projection by Injury Severity

In the previous chapter, we investigated the claim behaviour of different cohorts of claim, namely, stratifying using Injury Severity and Legal Representation at the end of the quarter the claims are reported. It was observed the claim behaviours were substantially different between the cohorts. In this section, we aim to carry out the ultimate size projection by breaking the claims into the same cohorts. The simulation methodology outlined in the previous chapters were repeated for each of the claim cohorts.

Table 8.10 shows the results of the projection when the claims were stratified and compared that to the projection results when all claims were combined. It can be observed, the total projected size is around 10% lower when the claims were stratified. In particular the Moderate and Severe injuries claims are projected to have significant savings while the minor claims were not projected to have material savings. That is, the projections are suggesting the claims managers are more conservative in the estimation of claims with more serious injuries. The examination of the projection results as a mean for actuarial valuations is left to the next chapter.

Claim Cohort	Number of Open Claims	Incurred as at Dec2009 ($m)	Projected Ult. Claims Cost ($m)
Minor Rep	8,058	1,266	1,247
Minor NR	3,396	302	297
Moderate	3,409	852	770
Severe	2,691	1,886	1,531
Total	17,554	4,305	3,844
All Claims Model	17,554	4,305	4,273

Table 8.10: Stratified Projection Results

The key question that remains is whether modelling claims in cohorts allowing for their specific claims development behaviour improved the predictive accuracy of finalised claim sizes. Figure 8.5 shows the gains chart of the Stratified CDP Projections compared with the other methods. While the stratified projection is improved and now is higher than the Incurred gains curve at points, the two curves are still indistinguishable. That is, to understand which claims will relatively cost more or less, the incurred cost as at the censoring date is as good an indication as any projection methods. The chart also includes a new gains curve, which is the "Perfect Foresight" curve, that is, the ultimate claim sizes for these claims would be known in advance. This curve is the maximum curve possible for the claims that were finalised in the validation data period and form an upper bound to the gains curves possible for any projection method.

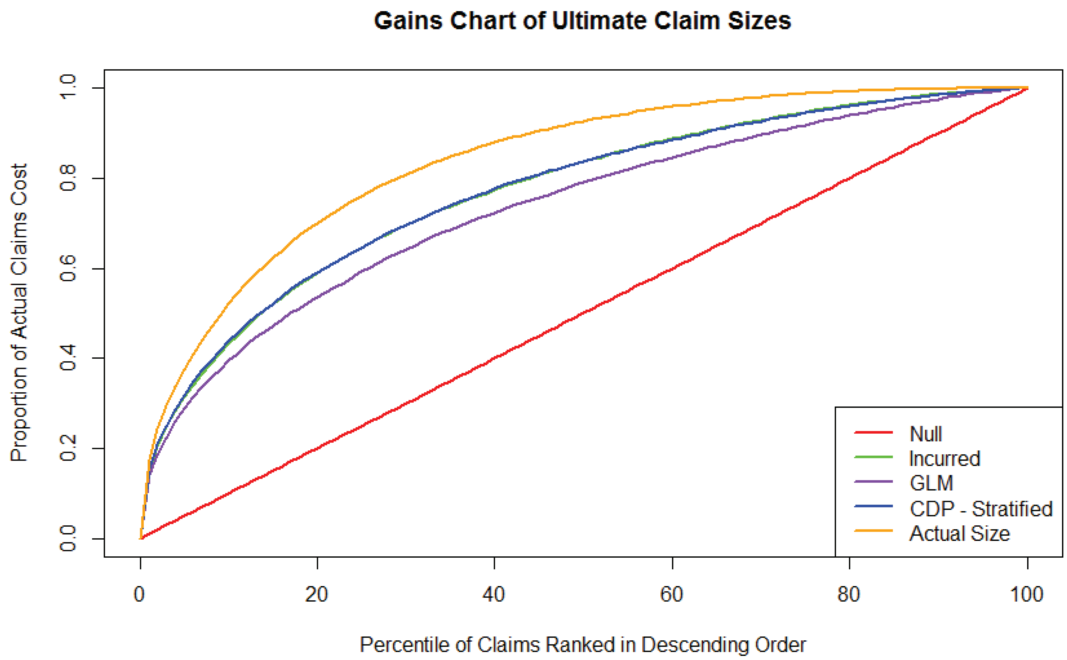

Figure 8.5: Gains Chart - Stratified by Injury Severity

8.7 Conclusion

In this chapter we have used a Monte-Carlo simulation approach to project open claims to their expected ultimate claim size. We have done this as algebraic calculation of the expected ultimate claim size is difficult.

Base on the findings from Chapter 7 that the "dynamic" covariates were very important explanatory variables for the four component process, we constructed a process of projecting future values of these dynamic covariates as well. This allows the "modelled" dynamic covariates to be used in the Monte-Carlo simulation.

The projection results were compared with the actual finalised claim sizes from the validation dataset, that is, data that was withheld from modelling for the purpose of accessing the efficacy of the model's predictiveness. We found that using an overall model to project all the claims was not as accurate as using separate models for different cohorts of claims. The categorisation of claims were based on claimant injury severity and legal representation, which are industry standard in segment NSW CTP claims.

In Chapter 9, we consider the appropriateness of the CDP framework as a valuation tool, including its ability to produce a consistent measure of variability of the projection results.

Chapter 9

Claims Projection as a Valuation Technique

9.1 Introduction

This chapter explores the Claim Development Process framework as a technique for actuarial valuations. By aggregating the expected ultimate sizes for a group of claims that are still outstanding at the censoring (or valuation) date, the future claims payments expected from these open claims can be calculated. By modelling the claim costs discussed in the previous chapters the insurer can understand the aggregate claim cost in a quantitative fashion and thus better plan for it.

The chapter starts by presenting the results from aggregating the individual claims results from Chapter 8. This chapter discusses some of the benefits when using the CDP framework as a valuation tool as well as some of its issues. Then, the results of the various valuation techniques discussed in Chapter 4 are then compared to that of Claims Development Processes.

9.2 Projection Results

Using the projection methodology described in Chapter 8 the results of the projection using the Claims Development Processes are tabulated below. Table 9.1 below shows the results in the same format as those for the GLM analysis. Similar to GLMs, the Claims Development Processes framework can only project open claims and does not provide an estimate for IBNR claims. The valuation results shown below incorporates the same IBNR estimate as adopted in the GLM methodology section, contained in Table 4.5. We have made the assumption that using a different methodology to analyse open claims would not impact the valuation of IBNR, or unreported, claims.

Two sets of results are presented - those from the overall projection model (Table 9.1) and those from the stratified models (Table 9.2) as discussed in the previous chapter.

Accident Year	Proj. Cost of Open Claims ($m)	Cost of Settled Claims ($m)	Cost of IBNR Claims ($m)	Total Claims Cost ($m)
2001	133	733	2	868
2002	126	719	4	849
2003	186	698	4	888
2004	249	734	4	987
2005	395	580	6	982
2006	637	398	19	1,055
2007	870	185	48	1,103
2008	985	54	94	1,134
2009	691	5	490	1,186
Total	4,273	4,107	671	9,052

Table 9.1: Projection Summary - Overall Model, CDP

Accident	Projected Ultimate Cost of Open Claims					Cost of	Cost of	Total
Year	Minor R ($m)	Minor NR ($m)	Moderate ($m)	Severe ($m)	Total ($m)	Settled Claims ($m)	IBNR Claims ($m)	Claims Cost ($m)
2001	29	11	17	75	132	733	2	867
2002	28	10	18	68	125	719	4	848
2003	57	14	25	86	183	698	4	885
2004	54	18	42	129	244	734	4	981
2005	91	29	80	172	372	580	6	958
2006	154	41	116	271	583	398	19	1,001
2007	246	50	174	299	769	185	48	1,002
2008	327	68	179	262	836	54	94	985
2009	249	54	116	166	584	5	490	1,079
Total	1,247	297	770	1,531	3,844	4,107	671	8,606

Table 9.2: Projection Summary - Stratified by Claim Cohorts, CDP

It is noted the results of the single overall projection model, the estimated claim cost for the 9 accident years is \$9,052m which is similar to the aggregate methods such as the ICD, PCE and PPCF methods. Using the stratified models, which projects claims separately for Minor Represented, Minor Not Represented, Moderate and Severe claims separately, the overall claim cost is \$8,606m, and this figure is similar to the GLM analysis. Intuitively, the stratified model projection results should be superior as it allows for the different claim development behaviours specific to each cohort of claims. The section below will compared the expected claim costs with the actual finalised claim cost during the validation data period to ascertain which set of results is superior.

9.3 Actual vs. Expected

One way to validate the projection model results is carry out what actuaries would call an "Actual versus Expected" (AvE) analysis. That is, from the 17,554 open claims that were modelled in the simulation process and find out which claims have been finalised during the three years covered by the validation data. For those claims, the expected ultimate claim cost from the projection model results can be compared to the actual costs these claims were finalised for. Out of the 17,544 claims that were open as at 31 December 2009, 14,134 were finalised as at 31 December 2012. Table 8.8 shows the AvE for the overall projection model and is reproduced below.

Accident Year	Number of Claims Finalised	Projected Ult. Claims Cost ($m)	Actual Finalised Claims Cost ($m)	Difference ($m)	Difference (%)
2001	224	99	101	2	2%
2002	247	102	113	12	11%
2003	328	144	153	8	6%
2004	465	188	189	1	0%
2005	829	322	299	-23	-7%
2006	1,645	542	434	-108	-20%
2007	2,869	754	622	-132	-18%
2008	3,978	788	635	-153	-19%
2009	3,549	495	401	-95	-19%
Total	14,134	3,435	2,947	-488	-14%

Table 9.3: Comparison of Projection Sizes vs. Actual Sizes

Table 8.8 showed the AvE for the stratified projection model results.

Accident Year	Number of Claims Finalised	Projected Ult. Claims Cost ($m)	Actual Finalised Claims Cost ($m)	Difference ($m)	Difference (%)
2001	224	97	101	4	4%
2002	247	100	113	14	14%
2003	328	141	153	12	8%
2004	465	183	189	6	3%
2005	829	302	299	-3	-1%
2006	1,645	493	434	-59	-12%
2007	2,869	666	622	-44	-7%
2008	3,978	664	635	-29	-4%
2009	3,549	411	401	-10	-2%
Total	14,134	3,057	2,947	-110	-4%

Table 9.4: Comparison of Projection Sizes (Stratified) vs. Actual Sizes

It is clear that the stratified projection results are superior. The overall difference is considerably smaller at -4% compared to -14%. Also, the accident year level differences are also smaller. However, the pattern of under-prediction and over-prediction is still present; that is, the projection of the older accident periods continue to be lower than the actual costs these claims are finalised at and vice-versa.

This is likely to be due to the "savings on finalisation" feature of the claims development behaviour that has been observed in the past chapters. A potential reason why the earlier years have been under-projected is due to changing behaviour - claims manager may have become more experienced and place less conservatism in their estimates over time. This may be especially so for the more severe claims, which are over-represented in the older accident years.

Typically, the AvE analysis on projection results are done annually comparing the actual claim sizes in the current year compared to the projection from the year before. That is,

- At the end of 2010, compared the claims finalised in 2010 to the projected claim sizes from a projection model built using data up to 2009;
- At the end of 2011, compared the claims finalised in 2011 to the projected claim sizes from a projection model built using data up to 2010;
- At the end of 2012, compared the claims finalised in 2012 to the projected claim sizes from a projection model built using data up to 2011;

This annual AvE allows the latest claims behaviour trends to be reflected in the model parameters and hence the projection results. By comparing three years of finalised claims places our projection results at a disadvantage. However, there is an easy way to check whether the claims behaviour regarding "saving on finalisation" has indeed changed during the validation period - leading to the incurred costs no longer contain as much conservatism. For the Size component model using data up to 2009, the fitted parameter for the "saving on finalisation" variable was 1.22; the same fitted parameter when data using up to 2012 (an additional three years of revisions) was 1.15. That is, the saving on finalisations during the validation data period was not as large as those during the modelling data period. This would be consistent with the observed AvE.

The table below compiles the projected claims cost of those claims that were finalised between 2010 and 2012 and compares them to their actual settled claims cost.

Accident Year	Number of Claims Finalised	Actual Settled Claims Cost ($m)	Projected Ultimate Claims Cost CDP ($m)	GLM ($m)	PPCF ($m)
2001	224	101	97	97	74
2002	247	113	100	85	81
2003	328	153	141	121	108
2004	465	189	183	139	153
2005	829	299	302	230	265
2006	1,645	434	493	379	448
2007	2,869	622	666	550	613
2008	3,978	635	664	656	643
2009	3,549	401	411	448	416
Total	14,134	2,947	3,057	2,704	2,800

Table 9.5: Comparison of AvE for the various methods

While the projected ultimate claim sizes from the CDP were the closest to the actual claim sizes on an aggregate basis, it also displays adequate comparisons at the accident year level. The other methods tended to severely under-project the earlier accident years. From an AvE perspective, the CDP seems to perform the closest to the actual results.

9.4 Comparison of Results Across Various Methods

Table 9.6 compares the four valuation techniques discussed in Chapter 4 (SCE was not applied due to its inappropriateness for the NSW CTP claims) to the results of the Claims Development Processes as a method for valuation.

Accident Year	2001 to 2009 Modelling Data ICD ($m)	PPCF ($m)	GLM ($m)	CDP ($m)	2010 to 2012 Modelling Data ICD ($m)	PPCF ($m)	GLM ($m)	CDP ($m)
2001	851	873	887	867	884	863	906	872
2002	842	856	860	848	888	864	896	848
2003	978	909	883	885	926	904	926	900
2004	951	996	939	981	1,012	988	997	988
2005	1,025	1,025	897	958	1,009	977	964	958
2006	1,001	1,034	907	1,001	1,033	965	945	1,014
2007	1,051	1,078	951	1,002	1,043	1,057	976	1,015
2008	1,083	1,065	1,098	985	1,085	1,068	991	997
2009	1,186	1,161	1,218	1,079	1,165	1,274	1,011	1,094
Total	8,968	8,998	8,640	8,606	9,045	8,959	8,611	8,686

Table 9.6: Comparison of Various Valuation Techniques

The following observations are made regarding the results of using the modelling data alone.

- The aggregate methods are very similar to each other in terms of the predicted ultimate claims cost for the industry over this nine year accident periods. All methods produced an total cost of around \$9,000m. Even at an individual accident year basis, the projected total is within \$20m of each other for most years (the difference is large for the 2009 accident year).

- The individual claims models are very similar to each other at around \$8,600m. The variations between the two methods at an accident year level are greater than the aggregate methods, they can be up to \$100m.

- Even with the addition of the validation data, the estimates of the ultimate claims cost for each method have changed very little. The aggregate methods are stubbornly staying at around \$9,000m and the individual claims project methods continue to remain at around \$8,600m. At the time of writing it is still uncertain which method is the closest to the true ultimate claims cost for NSW CTP.

9.5 Benefits of Valuation based on Individual Claims Projection

From the discussion in the previous section, the aggregate models seem to be adequate as valuation techniques for the NSW CTP portfolio. The overall projection results seem to be cross validated by all three methods giving similar answers and the answer remained stable even after three extra years of information is incorporated. There are still considerable benefits however to be using an individual claims model.

The individual claims model (GLM, SCE or CDP) can provide considerable information regarding the claims. For example

- Has the profile of claim characteristics changed over time?
- What is the cost differential of a legally represented claim versus a non-represented claims? Is there any benefit in employing a (costly) strategy to be more agreeable to the claimants' demands so they do not employ a lawyer?
- When incorporating claim manager identifiers, the behaviours of the claims management can be linked to claim cost outcomes and perhaps "leakages" can be identified. Leakages typically refer to a way of handling claims that may elevate claims cost for the insurer.

Secondly, a claim by claim projection provides the insurer the ability to identify costly claims. These may allow these claims to be managed by more experienced claims managers to achieve a better claims outcome for both the claimant and the insurer. Also, claims with the potential to "grow" the most from their initial estimate (at the time of reporting) can also be identified. This allows some learning to be propagated back to the claims team so they can better appreciate the potential cost of these types of claims.

The insurer is typically not only concerned with the actuary's best estimate at the ultimate claims cost of a particular portfolio, it would also like to know what is the variability of the actual outcome in relation to its mean. The Australian Prudential Regulation Authority asks insurers to reserve for the outstanding claims cost at the 75% percentile level out of a range of possible values the insurers has identified. Using an individual claims model such as CDP framework the standard deviation of the prediction can be readily measured.

Figure 9.1 is a histogram of the ultimate claims cost (for those claims that were open as at the end of 2009) from the 200 simulation runs. We have superimposed a normal distribution as a benchmark only; the fit of the simulation output to the normal distribution is adequate. The variability is actually surprisingly small, with a coefficient of variation of just 1%. The parameter variation can also be easily measured, if we assume the parameters follows a normal distribution with the fitted mean and fitted standard error, a simulation approach, that simulates possible claim trajectories, can also be used to measure the variation in the projected claims cost. These variations are difficult to measure using an aggregate method.

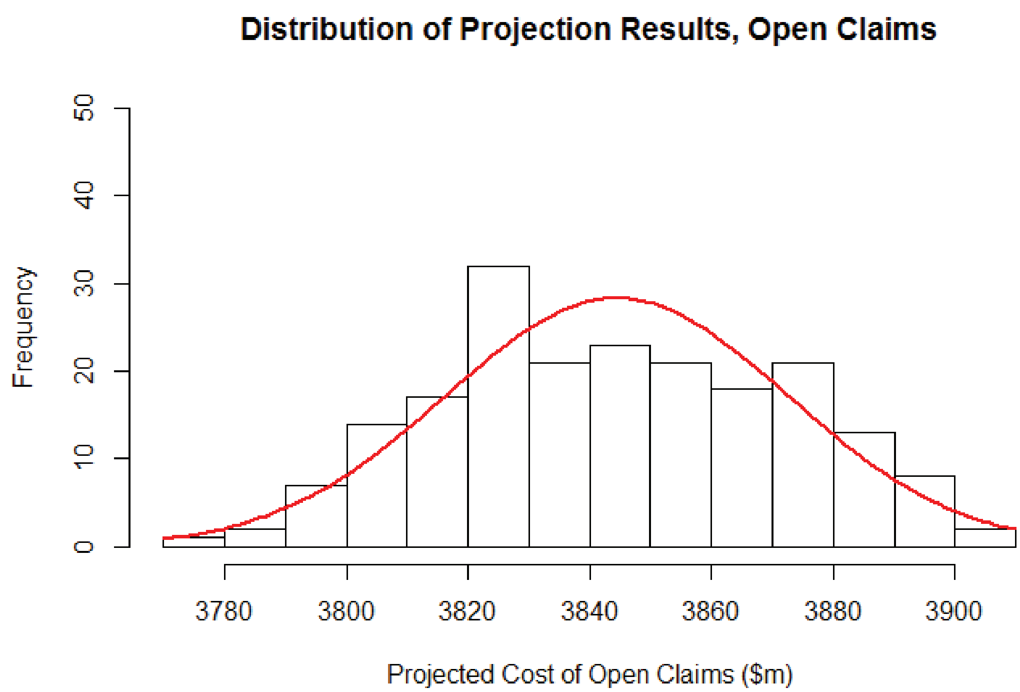

Figure 9.1: Distribution of Projection Results

9.6 Using Projection Results for Valuation

In this section a novel way of using the results of the projections modelling to carry out valuation is presented. For the purpose of claims reserve valuation, the insurer is interested in the ultimate size of all outstanding claim and the likely time the claims are to be settled. This information is used to enable financial performance analysis as in how much money is need to set aside to pay for future claims costs at each future period of time.

From Figure 8.2 we saw that for each claim, the results of the simulation can be expressed in a heat map format. This represents the distribution of the combinations of ultimate size and time to settlement outcomes. While the insights gained in Chapter 7 and the projection of individual claim paths earlier in this chapter are extremely valuable, for the purposes of valuation the insurer needs results fast.

Figure 9.2 portrays the same information as Figure 8.2; however, the y-axis has been log transformed. From the diagram, the $(\log(X_i), T_i)$ outcome pair seems to be a bivariate normal distribution.

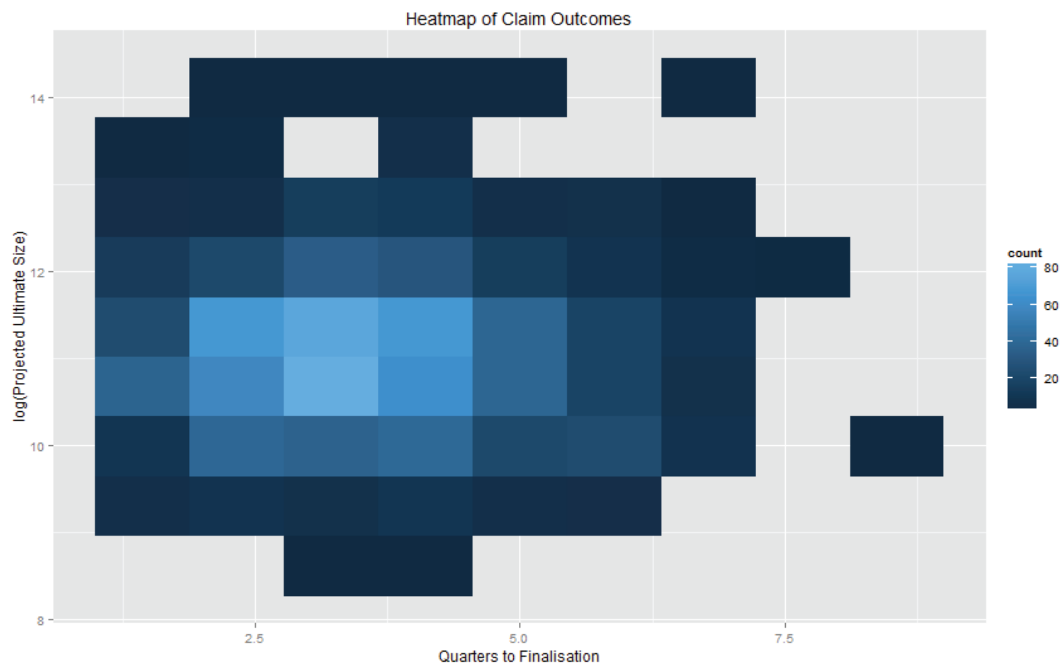

Figure 9.2: Heat Map of Claim Outcomes

Figures 9.3 and 9.4 show that the marginal distributions of $\log(X_i)$ and T_i for this claim indeed look adequately encapsulated by normal distributions. The correlation between the two outcome variables of interest is -0.02.

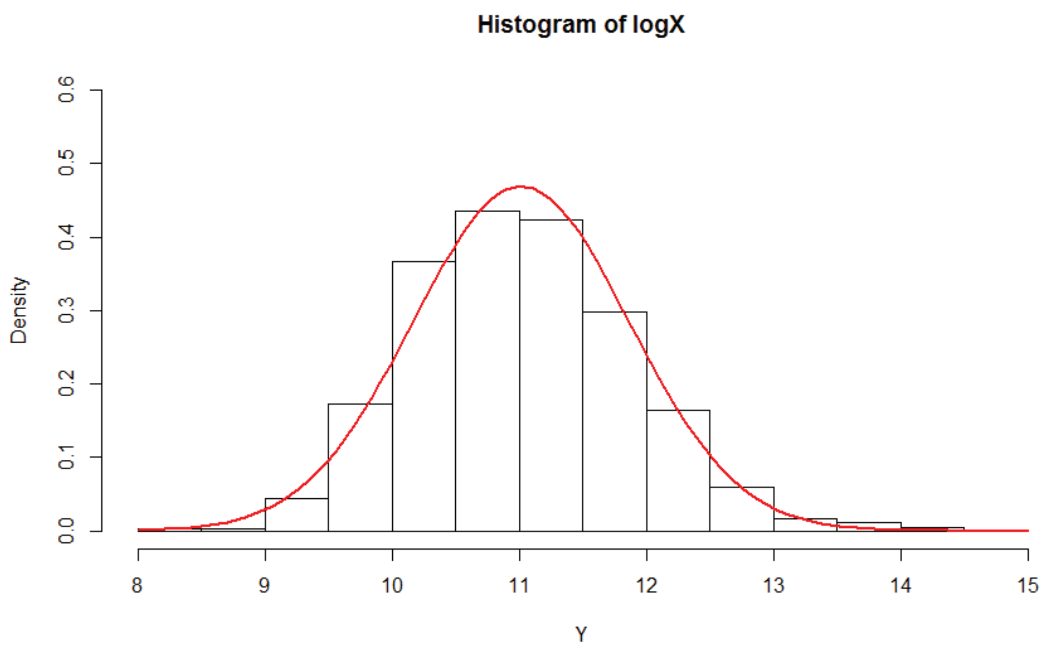

Figure 9.3: Marginal Distribution of $\log(X)$

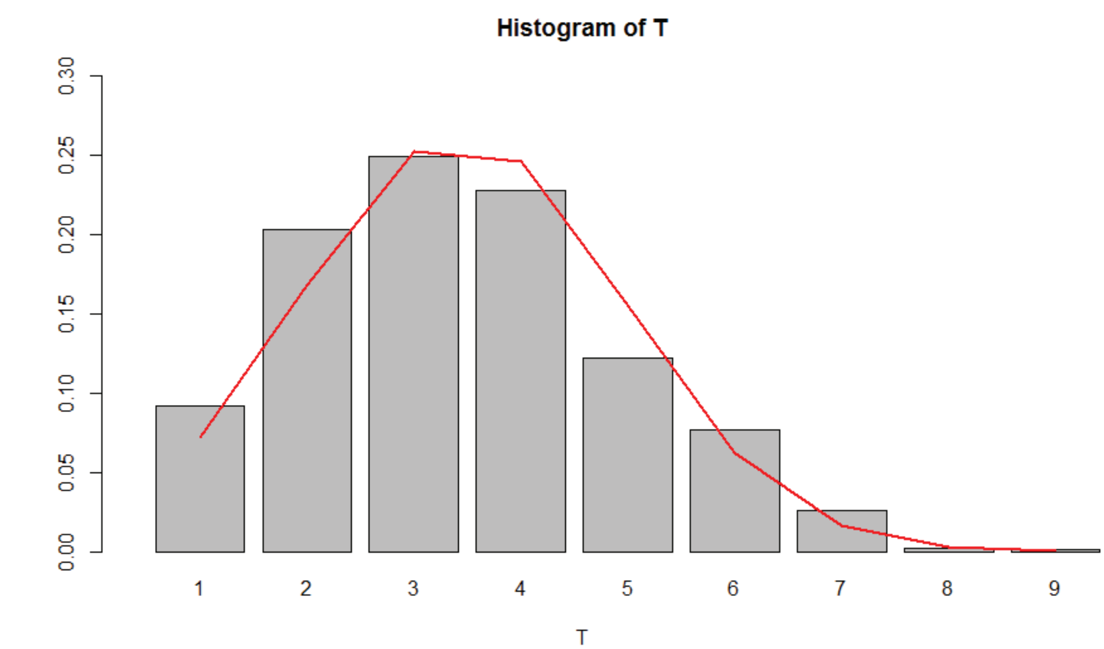

Figure 9.4: Marginal Distribution of T

This means for each claim, the range of likely outcomes can be represented by five parameters. They are presented by $\mu_{\log(X)}$, $\sigma_{\log(X)}$, μ_T , σ_T , ρ. These five parameters contain all the information required to understand the outcomes for each open claim at the censoring date. Hence, by designing a model relating the five parameters to the claim characteristics at the censoring time, the valuation results can be computed quickly.

Parameter	mu_ln(X)	sigma_ln(X)	mu_T	sigma_T	rho
Error Distribution	Normal	Gamma	Gamma	Gamma	Normal
Intercept	**1.207**	**1.888**	**2.944**	**1.391**	**0.093**
Scale	0.121	75.750	373.624	625.479	0.033
Charactersticts at Censoring Date					
Gender Male	**-0.014**	**0.032**	**0.114**	**0.040**	-0.003
Gender Female	0.000	0.000	0.000	0.000	0.000
Insurer A	**0.015**	**0.097**	**-0.724**	**-0.234**	**-0.016**
Insurer B	**-0.002**	0.004	-0.057	**-0.018**	-0.001
Insurer C	**0.067**	0.020	**0.615**	**0.182**	0.004
Insurer D	**-0.344**	0.021	**0.426**	**0.129**	**-0.017**
Insurer E	-0.025	0.031	**-0.317**	**-0.097**	**-0.012**
Insurer F	**-0.107**	**0.070**	**-0.116**	-0.038	**-0.011**
Insurer Other	0.000	0.000	0.000	0.000	0.000
Legal Rep No	**-0.518**	**-0.085**	**-0.052**	**-0.022**	**-0.027**
Legal Rep Yes	0.000	0.000	0.000	0.000	0.000
ISS of 0	**-0.298**	**0.152**	**-0.107**	**-0.027**	**-0.029**
ISS of 1	**-0.320**	**0.072**	-0.069	**-0.028**	**-0.025**
ISS of 2	**-0.195**	**0.066**	-0.057	**-0.027**	**-0.023**
ISS of 3 - 5	**-0.148**	**0.067**	-0.089	**-0.033**	**-0.016**
ISS of 6 - 10	-0.041	**0.087**	**-0.111**	**-0.036**	**-0.016**
ISS of 11 - 30	0.033	**0.079**	-0.052	**-0.017**	-0.007
ISS of 30+	0.000	0.000	0.000	0.000	0.000
No Litigation	-0.139	**0.190**	**-0.286**	**-0.076**	0.000
NSW Supreme Court	**0.386**	0.033	**-0.424**	**-0.118**	0.019
NSW District Court	**0.234**	0.041	**-0.711**	**-0.210**	0.007
NSW Local Court	0.022	0.020	**-0.341**	-0.098	0.012
Other Court	0.000	0.000	0.000	0.000	0.000
Accepted Liability	**0.408**	**-0.299**	**-0.334**	**-0.107**	-0.005
Partially Accepted Liability	**0.217**	**-0.263**	**-0.294**	**-0.099**	-0.011
Rejected Liabiliy	**-0.355**	-0.001	**-0.099**	**-0.046**	**-0.029**
Other	0.000	0.000	0.000	0.000	0.000
ln(Size at Censoring Date)	**0.873**	**-0.107**	**0.057**	**0.019**	**0.001**
Ln(Initial Estimate)	**0.014**	**0.018**	**0.024**	**0.008**	0.033
Deviance /DF	0.015	0.0136	0.003	0.002	0.001

Table 9.7: GLM of the 5 Parameters of Interest

GLM models were built for the 5 parameters and the model results are shown in Table 9.7. The models suggest the parameters can be predicted with a very high degree of accuracy. Use the claim characteristics as at censoring date the total claims cost can be easily compiled.

9.7 Issues

The Claims Development Process framework is not without its shortcomings. While being a novel technique that can be used to understand claims development behaviour and used for reserves valuation, it has some drawbacks.

Firstly, the time and effort employed in carrying out a valuation exercise using the CDP is considerable for the first time. While the ICD can be employed to carry out a valuation in a matter of days, including data extraction and preparation, analysis and documentation. The CDP on the other hand may take a considerable amount of time to setup initially. This is especially so given the complicated modelling involved; however, with modern multi-core/multi-thread processor computers, the time required to fit the models and project ultimate claim costs can be greatly reduced. This is possible as all claims are assumed independent and their likelihoods can be separately calculated. However, if anything goes wrong or the results are unexpected then considerable effort would be required to debug the process and find where the issue is. It may also take some time to convince the various stakeholders to adopt a new process, particularly one that involves complex modelling.

Secondly, and more importantly, individual claims modelling does not predict IBNR, or claims that have not reported yet. As these claims have not been reported, there is not claims information to project the claims with. In the analyses so far, an average claim size has been assumed based on the PPCF model, the same size as other unsettled claims from the accident period. While this approach does not fit in the CDP framework there is an alternative which allows the IBNR claims to be incorporated in the CDP models. IBNR claims can be "bootstrapped" from a pool of past IBNR claims. This method will allow the IBNR to be generated with claim characteristics and be projected to their ultimate size using the CDP framework.

9.8 Conclusion

As a valuation tool, the CDP framework has a few advantages. The main advantage is that it provides overall results similar to the aggregate methods such as the ICD and PPCF methods yet it also provides more granular results at an individual claims basis. This allows the insurance company to analyse the results in various perspectives to improve their claims management function or policy selection functions.

We have also compared the results of the CDP projection with the traditional actuarial methods. While the results are similar, at the time of writing we still cannot say with certainty which method is the most accurate for accident years 2001 to 2009. This is because of the long tailed nature of CTP claims and a considerable portion of the claims are still outstanding.

Some of the issues of using the CDP framework as a valuation method are also outlined. The main disadvantage is the length of time it needs to produce valuation results. However, in this chapter, a novel method was outlined to approximate the projection results using a bivariate gaussian distribution. Another disadvantage, which the CDP shares with other individual claims models, is that it does not project IBNR claims. We have used the IBNR estimation from the PPCF method for the purposes of presenting complete valuation results.

Chapter 10

Random Effects

10.1 Introduction

One of the difficulties discussed in Chapter 7 when applying the Claims Development Processes is the great heterogeneity of the CTP claims. Claims can be minor such as the cost of an X-ray taken "just-to-be-safe" or could be catastrophic spinal injury. Claimants can also vary greatly in age, social-economic background and ability to deal with the claim process themselves. The variability of claims characteristics compounded with the relativity short "trajectories" (on average, claims undergo 3 revisions before the claim is settled with 83% of claims finalised within 5 revisions) poses a few problems for model fitting, such as variability of coefficients as well as convergence of the estimates of the coefficients.

A few options exist in dealing with such variabilities. Firstly, the claims can be stratified into more homogenous groups - such as by injury severity or date of accident. Separate models may be fit to each of these groups and the parameters may be compared and the more homogenous groups may aid the model building. Secondly, the user may use industry knowledge and/or theory to identify variables whose coefficients are likely to vary across individual claims and use a random effects framework to capture such variation.

A random effects framework is the main focus of this chapter. The merits of a random effects model is discussed, in particular, considering the nature of the CTP data. The delay component process is first considered for the application of a random effects structure where we investigated the inclusion of both a random intercept as well as a random covariate. This chapter concludes with a comparison of the random effects model and the fixed effects model in the aspects of model fit and the predictive accuracy on an individual claim basis.

10.2 Random Effects vs. Fixed Effects

There are numerous benefits by adopting a random effects model. Firstly, the extent of the individual claim variations of the variables of interest may be investigated. By quantifying the individual effects, the random effects model may produce more accurate predictions. Secondly, by incorporating these random effects the significance of the other regression variables may improve. Thirdly, insights into which variable has claimant to claimant variability may be gained.

However, in relation to the CTP data, we consider a key issue - does the data heterogeneity impacts the individual claim predictability when a global model is used. The vast number of claims in addition to the short claim trajectory and significant claim heterogeneity make the CTP dataset more similar to a longitudinal dataset. In longitudinal data studies, random effects models play a key role to reduce between sample variability to help with inference making.

The methodology adopted for the analysis of random effects in the GLARMA framework is based on the research of Dunsmuir et al. [2014a] and Dunsmuir [2015]. In these papers, the authors develop the modelling tools for panel data of Poisson and binomial counts, in particular, using Adaptive Gaussian Quadratures to approximate the first and second order derivatives of the log-likelihood function. Their work has been extended to negative binomial and gamma responses in the following sections.

10.3 Delay Component Process

The Delay component process is a good candidate to consider random effects for. Intuitively, the claimant can be considered as a random effect due to the varying attitude of the individual towards the CTP claim process and the lawyer can also be considered as a random effect affecting only those claimants with legal representation. The former introduces a random intercept into the model, applying to each and every claim in the portfolio, while the latter only introduces a random effect only to the represented claims.

Rearranging Equation (6.9), and dropping the conditioning, we have the following

$$P(t_{j,i} = t) = \frac{\Gamma(\alpha + t - 1)}{\Gamma(\alpha)\Gamma(t)} \frac{(\mu_{j,i}/\alpha)^{t-1}}{(1 + \mu_{j,i}/\alpha)^{t-1+\alpha}} \tag{10.1}$$

where $t = 1, 2, 3, ...$

$\mu_{j,i}$ is the mean of the delay process and for the fixed effects only model is defined as

$$\log \mu_{j,i} = Z_{j,i}^T \beta + W_{j,i} \tag{10.2}$$

where $W_{j,i} = \phi_1 e_{j-1,i} + \phi_1^2 e_{j-2,i} + ...$

The overall likelihood function, using the subscript "FE" to denote the fixed effects model, is

$$\begin{aligned} L_{FE} &= \prod_{i=1}^{n} \prod_{j=1}^{m_i} L_{j,i} \\ L_{FE} &= \prod_{i=1}^{n} \prod_{j=1}^{m_i} \frac{\Gamma(\alpha + t_j - 1)}{\Gamma(\alpha)\Gamma(t_j)} \frac{(\mu_{j,i}/\alpha)^{t_{j,i}-1}}{(1 + \mu_{j,i}/\alpha)^{t_{j,i}-1+\alpha}} \end{aligned} \tag{10.3}$$

The corresponding log-likelihood function is

$$
\begin{aligned}
\ell_{FE}(\beta,\phi,\alpha) &= \sum_{i=1}^{n}\sum_{j=1}^{m_i} \log\left(\frac{\Gamma(\alpha+t_j-1)}{\Gamma(\alpha)\Gamma(t_j)}\right) \\
&\quad + \sum_{i=1}^{n}\sum_{j=1}^{m_i}((t_{j,i}-1)(\log\mu_{j,i}-\log\alpha)-(t_{j,i}-1+\alpha)\log(1+\mu_{j,i}/\alpha))
\end{aligned}
\tag{10.4}
$$

Now introduce the individual claimant random effect as a random intercept, u_i, to $\mu_{j,i}$,

$$
\log\mu_{j,i} = Z_{j,i}^T\beta + u_i + W_{j,i} \tag{10.5}
$$

where $u_i \overset{\text{iid}}{\sim} N(0,\sigma^2)$

The likelihood and the log-likelihood of the delay component with a random effects intercept are as follows.

$$
L_{RE} = \prod_{i=1}^{n}\int_{u_i}\prod_{j=1}^{m_i}\frac{\Gamma(\alpha+t_j-1)}{\Gamma(\alpha)\Gamma(t_j)}\frac{(\mu_{j,i}/\alpha)^{t_{j,i}-1}}{(1+\mu_{j,i}/\alpha)^{t_{j,i}-1+\alpha}}g(u_i)du_i \tag{10.6}
$$

where $g(u_i)$ is the normal density function for u_i.

And

$$
\begin{aligned}
\ell_{RE}(\beta,\phi,\alpha,\sigma) &= \sum_{i=1}^{n}\log\int_{u_i}\prod_{j=1}^{m_i}\frac{\Gamma(\alpha+t_j-1)}{\Gamma(\alpha)\Gamma(t_j)}\frac{(\mu_{j,i}/\alpha)^{t_{j,i}-1}}{(1+\mu_{j,i}/\alpha)^{t_{j,i}-1+\alpha}}g(u_i)du_i \\
&= \sum_{i=1}^{n}\log\left(\int_{u_i}\left(\prod_{j=1}^{m_i}\frac{\Gamma(\alpha+t_j-1)}{\Gamma(\alpha)\Gamma(t_j)}\frac{(\mu_{j,i}/\alpha)^{t_{j,i}-1}}{(1+\mu_{j,i}/\alpha)^{t_{j,i}-1+\alpha}}\right)\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{u_i^2}{2\sigma^2}}du_i\right)
\end{aligned}
\tag{10.7}
$$

The objective is to maximise the above log-likelihood function to estimate the parameters of the model. However, to do that we need to evaluate the integral in the above equation. In similar fashion to Dunsmuir [2015], we use the Laplace Approximation and Adaptive Gaussian Quadrature methods to approximate the integral in the above equation.

The following algebraic derivations of the likelihood function approximations and their derivatives are adapted from those found in Dunsmuir et al. [2014a] for the negative binomial distribution.

Let $u_i = \sigma z_i$ where $z_i \sim N(0,1)$ and to simplify notation let $h(\alpha) = \frac{\Gamma(\alpha+t_j-1)}{\Gamma(\alpha)\Gamma(t_j)}$, the above integral can be rewritten as

$$
\begin{aligned}
&\int_{z_i} \left(\prod_{j=1}^{m_i} h(\alpha) \frac{(\mu_{j,i}/\alpha)^{t_{j,i}-1}}{(1+\mu_{j,i}/\alpha)^{t_{j,i}-1+\alpha}} \right) \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}} dz_i \\
=& \frac{1}{\sqrt{2\pi}} \int_{z_i} \left(\prod_{j=1}^{m_i} \exp\left(\log h(\alpha) + (t_{j,i}-1)\log(\mu_{j,i}/\alpha) - (t_{j,i}-1+\alpha)\log(1+\mu_{j,i}/\alpha)\right) \right) e^{-\frac{z_i^2}{2}} dz_i \\
=& \frac{1}{\sqrt{2\pi}} \int_{z_i} \exp\left(\sum_{j=1}^{m_i} \left(\log h(\alpha) + (t_{j,i}-1)\log(\mu_{j,i}/\alpha) - (t_{j,i}-1+\alpha)\log(1+\mu_{j,i}/\alpha)\right) - \frac{z_i^2}{2} \right) dz_i
\end{aligned}
\tag{10.8}
$$

The integral above can be approximated using the Laplace Approximation method. The idea behind this method is that for an integral that can be expressed as $\int e^{f(z)} dz$ and that $f(z)$ is a concave function then the integrand can be approximated using a second order Taylor expansion of $f(z)$.

If $f(z)$ is expanded around z',

$$
f(z) = f(z') + f'(z')(z-z') - \frac{1}{2} f''(z')(z-z')^2 + O((z-z')^3) \tag{10.9}
$$

and if z^* is the global maximum of the function $f(z)$ and hence $f'(z^*) = 0$ and $f''(z^*) < 0$ then

$$
f(z) \approx f(z^*) + \frac{1}{2} |f''(z^*)| (z-z^*)^2 \tag{10.10}
$$

This leads to

$$
\begin{aligned}
\int e^{f(z)} dz &\approx \int e^{f(z^*) - \frac{1}{2}|f''(z^*)|(z-z^*)^2} dz \\
&= e^{f(z^*)} \int e^{-\frac{1}{2}|f''(z^*)|(z-z^*)^2} dz
\end{aligned}
\tag{10.11}
$$

Noticing the integral above is in the form of a normal density, the above can be written as the well known equation of the Laplace Approximation formula.

$$\int e^{f(z)}dz \approx \sqrt{\frac{2\pi}{|f''(z^*)|}}\exp(f(z^*)) \tag{10.12}$$

where z^* is the z that maximises $f(z)$.

Now using the transformations $\eta_{j,i} = \log\mu_{j,i} = Z_{j,i}^T\beta + \sigma z_i + W_{j,i}$ and $\alpha = e^{-d}$, we define $f(z)$ as follows

$$f(z) = \sum_{j=1}^{m_i}\left(\log h(d) + (t_{j,i}-1)(\eta_{j,i}+d) - (t_{j,i}-1+e^{-d})\log(1+e^{d+\eta_{j,i}})\right) - \frac{z_i^2}{2} \tag{10.13}$$

The terms in the summation are actually the log-likelihood contribution of the ith claim and jth revision, hence,

$$f(z) = \sum_{j=1}^{m_i}(\ell_{FE,j,i}) - \frac{z_i^2}{2} \tag{10.14}$$

Since $f(z)$ is a function of a log-likelihood function, it is a concave function with a defined global maxima and thus satisfies the requirement of the Laplace Approximation. Furthermore the first order and second order derivatives are derived as

$$\begin{aligned} f'(z) &= \sum_{j=1}^{m_i}\left((t_{j,i}-1)\sigma - (t_{j,i}-1+e^{-d})\frac{e^{d+\eta_{j,i}}}{1+e^{d+\eta_{j,i}}}\sigma\right) - z_i \\ &= \sum_{j=1}^{m_i}\left(\frac{d\ell_{FE,j,i}}{dz}\right) - z_i \end{aligned} \tag{10.15}$$

and

$$\begin{aligned} f''(z) &= \sum_{j=1}^{m_i} -(t_{j,i}-1+e^{-d})\sigma\left(\frac{e^{d+\eta_{j,i}}}{1+e^{d+\eta_{j,i}}}\sigma - \frac{e^{2(d+\eta_{j,i})}}{(1+e^{d+\eta_{j,i}})^2}\sigma\right) - 1 \\ &= \sum_{j=1}^{m_i}\left(\frac{d^2\ell_{FE,j,i}}{dz^2}\right) - 1 \end{aligned} \tag{10.16}$$

Since the objective is to find the z^* such that $f(z)$ is maximised, this problem is similar to that of maximising the log-likelihood over the parameter space which was carried in Chapter 7. A slight modification to the R code used allows the solution to the z^* to be readily calculated using the Newton-Raphson algorithm. This allows us to approximate the RE log-likelihood using the Laplace Approximation

$$\begin{aligned} \ell_{RE}(\beta, \sigma, \phi, d) &\approx \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{\pi 2}} \sqrt{\frac{2\pi}{f''(z^*)}} \exp(f(z^*)) \right) \\ &= \sum_{i=1}^{n} \log\left(\frac{\exp(f(z^*))}{\sqrt{f''(z^*)}} \right) \end{aligned} \tag{10.17}$$

While the Laplace Approximation allows the approximation of the log-likelihood of the random effects model, to maximise it for the estimation of the model parameters, the first and second order derivatives are needed. While the integrand in the log-likelihood function is a concave function that satisfies the requirements for the Laplace Approximation, the same cannot be said of the first and second order derivatives.

We turn our attention to another approximation method described in Pinheiro and Chao [2006], the Adaptive Gaussian Quadrature (AGQ). Where an integral is approximated by evaluation the integral at a set of predetermined values (the abscissas) using a particular density function (the kernel), then a weight average of these values is used to approximate the integral.

Pinheiro and Chao [2006] show that when only one Gaussian Quadrature point is used in the approximation then AGQ is equivalent to the Laplace Approximation. In the modelling of the CTP data in the later section, both Laplace Approximation and AGQ are carried out, if only to serve as a check on the proper function of the model fitting code.

From Equations (10.7) and (10.8), and using the definition of $f(z)$ from Equation (10.13),

$$\ell_{RE}(\beta,\sigma,\phi,d) = \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi}}\int_{z_i} \exp(f(z))dz\right) \tag{10.18}$$

In Laplace Approximation, the above equation is evaluate at z^*, which is the global maxima (or the mode) of $f(z)$. For AGQ estimate, we evaluate the above function using a number of quadrature points, determined by the convergence of the likelihood.

For the purpose of model fitting we have followed the approach discussed in Dunsmuir [2015]. The hermite quadrature points and the corresponding weights are obtained from the statmod package in R (Smyth [2014]). The hermite quadrature points and weights are for integration against the $\exp(-x^2)$ kernel rather than the standard normal density. Hence, a change of variable is required, where $\zeta = x/\sqrt{2}$ is the set of quadrature points centred around z^* .

Using AGQ, with q number of quadrature points,

$$\begin{aligned}\ell_{RE}(\beta,\sigma,\phi,d) &\approx \sum_{i=1}^{n} \log\left(\frac{\sqrt{2}}{\sqrt{2\pi}}\sum_{k=1}^{q} \sqrt{\sigma}\exp(f(z^*+\sqrt{2\sigma}\zeta_k))\exp(\zeta_k^2)w_k\right)\\ &= \sum_{i=1}^{n} \log\left(\sum_{k=1}^{q} \frac{\sqrt{\sigma}}{\sqrt{\pi}}\exp(\zeta_k^2)w_k\exp(f(z^*+\sqrt{2\sigma}\zeta_k))\right)\end{aligned} \tag{10.19}$$

where z^* has the same definition as under the Laplace Approximation approach.

Define $\tilde{z}_k = z^* + \sqrt{2\sigma}\zeta_k$, the "adaptive" abscissas then the equation used for the AGQ estimation is

$$\ell_{RE}(\beta,\sigma,\phi,d) \approx \sum_{i=1}^{n} \log\left(\sum_{k=1}^{q} \frac{\sqrt{\sigma}}{\sqrt{\pi}}\exp(\zeta_k^2)w_k\exp(\sum_{j=1}^{m_i} \ell(\tilde{z}_k) - \frac{\tilde{z}_k^2}{2})\right) \tag{10.20}$$

The derivatives of the log-likelihood is also numerically approximated using AGQ. To simply notation, let δ denote the vector of parameters $(\beta,\sigma,\phi,d)^T$

The first derivative of the log-likelihood can written as

$$
\begin{aligned}
\frac{\partial}{\partial \delta}\ell_{RE}(\delta) &= \sum_{i=1}^{n} \frac{1}{L_i(\delta)} \int_z \frac{\partial}{\partial \delta}\ell_i(\delta|z)\exp(\ell_i(\delta|z))g(z)dz \\
&= \sum_{i=1}^{n} \frac{1}{L_i(\delta)} \frac{1}{\sqrt{2\pi}} \int_z \frac{\partial}{\partial \delta}\ell_i(\delta|z)\exp(f(z))dz \\
&\approx \sum_{i=1}^{n} \frac{1}{L_i(\delta)} \left(\sum_{k=1}^{q} \frac{\partial}{\partial \delta}\ell_i(\delta|\tilde{z}_k)\frac{\sqrt{\sigma}}{\sqrt{\pi}}\exp(\zeta_k^2)w_k \exp(\sum_{j=1}^{m_i}\ell(\tilde{z}_k) - \frac{\tilde{z}_k^2}{2}) \right)
\end{aligned}
\tag{10.21}
$$

where $L_i = \int_z \exp(\ell_i|z)g(z)dz$.

The second derivative of the log-likelihood can written as

$$
\begin{aligned}
\frac{\partial^2}{\partial \delta \partial \delta^T}\ell_{RE}(\delta) &= \sum_{i=1}^{n} \frac{1}{L_i(\delta)} \int_z \frac{\partial}{\partial \delta}\frac{\partial}{\partial \delta^T}\ell_i(\delta|z)\exp(\ell_i(\delta|z))g(z)dz \\
&+ \sum_{i=1}^{n} \frac{1}{L_i(\delta)} \int_z \frac{\partial}{\partial \delta}\ell_i(\delta|z)\frac{\partial}{\partial \delta^T}\ell_i(\delta|z)\exp(\ell_i(\delta|z))g(z)dz \\
&- \sum_{i=1}^{n} \frac{1}{L_i(\delta)} \int_z \frac{\partial}{\partial \delta}\ell_i(\delta|z)\exp(\ell_i(\delta|z))g(z)dz \frac{1}{L_i(\delta)} \int_z \frac{\partial}{\partial \delta^T}\ell_i(\delta|z)\exp(\ell_i(\delta|z))g(z)dz
\end{aligned}
\tag{10.22}
$$

Using AGQ, the approximation equation is

$$
\begin{aligned}
\frac{\partial^2}{\partial \delta \partial \delta^T}\ell_{RE}(\delta) &\approx \sum_{i=1}^{n} \frac{1}{L_i(\delta)} \left(\sum_{k=1}^{q} \frac{\partial}{\partial \delta}\frac{\partial}{\partial \delta^T}\ell_i(\delta|\tilde{z}_k)\frac{\sqrt{\sigma}}{\sqrt{\pi}}\exp(\zeta_k^2)w_k \exp(\sum_{j=1}^{m_i}\ell(\tilde{z}_k) - \frac{\tilde{z}_k^2}{2}) \right) \\
&+ \sum_{i=1}^{n} \frac{1}{L_i(\delta)} \left(\sum_{k=1}^{q} \frac{\partial}{\partial \delta}\ell_i(\delta|\tilde{z}_k)\frac{\partial}{\partial \delta^T}\ell_i(\delta|\tilde{z}_k)\frac{\sqrt{\sigma}}{\sqrt{\pi}}\exp(\zeta_k^2)w_k \exp(\sum_{j=1}^{m_i}\ell(\tilde{z}_k) - \frac{\tilde{z}_k^2}{2}) \right) \\
&- \sum_{i=1}^{n} \frac{1}{L_i^2(\delta)} \left(\sum_{k=1}^{q} \frac{\partial}{\partial \delta}\ell_i(\delta|\tilde{z}_k)\frac{\sqrt{\sigma}}{\sqrt{\pi}}\exp(\zeta_k^2)w_k \exp(\sum_{j=1}^{m_i}\ell(\tilde{z}_k) - \frac{\tilde{z}_k^2}{2}) \right) \\
&\left(\sum_{k=1}^{q} \frac{\partial}{\partial \delta^T}\ell_i(\delta|\tilde{z}_k)\frac{\sqrt{\sigma}}{\sqrt{\pi}}\exp(\zeta_k^2)w_k \exp(\sum_{j=1}^{m_i}\ell(\tilde{z}_k) - \frac{\tilde{z}_k^2}{2}) \right)
\end{aligned}
\tag{10.23}
$$

We have adapted the wrapper program used to fit the models in Chapter 7 to allow random effects to be incorporated in the model fitting. We have maintained the program for calculation of the likelihood and its derivatives contributions from each claim and allowed for the AGQ approximation in the wrapper program to minimise the amount of changes required. An example of a wrapper code to allow for random effects can be found in G.6. Both the random effects models (random intercept model and random "legal representation" model) and the fixed effects model were fitted to the claims that arose from the 2003 accident year. All the models fitted in this chapter used 7 quadrature points in the estimate of the likelihood and its derivatives. We have found 5 quadrature points would be adequate for most applications as the incremental differences in the likelihood and its derivatives between 5 quadrature points and 7 quadrature points are less than 10^{-3}.

Table 10.1 shows the results of adopting a random effects GLARMA model for the delay process using a negative binomial response distribution. We have ignored the issue of censoring as the combined complexity of RE and censoring proved to be very challenging. Nevertheless, we can appreciate the impact of RE on the model without making an allowance of censoring. Comparison are made between the fixed effects only model, a Random Intercept model and a Random Effect on Legally Represented claimants model.

Variable	Fixed Effect		Random Intercept		Random LegRep	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Intercept	3.18	0.05	3.18	0.05	3.18	0.05
Legal Representation	-0.05	0.01	-0.05	0.01	-0.05	0.01
Previous Delay	0.05	0.02	0.05	0.02	0.05	0.02
Previous Change	0.08	0.01	0.08	0.01	0.08	0.01
Revision 1						
Revision 2	-0.70	0.02	-0.70	0.02	-0.70	0.02
Revision 3	-1.25	0.03	-1.25	0.03	-1.25	0.03
Revision 4	-1.58	0.03	-1.58	0.03	-1.58	0.03
Revision 5	-1.86	0.03	-1.86	0.03	-1.86	0.03
Revision 6 +	-2.22	0.03	-2.22	0.03	-2.22	0.03
No Litigation at j - 1						
NSW Supreme Court at j - 1	-0.21	0.11	-0.21	0.11	-0.21	0.11
NSW District Court at j - 1	-0.52	0.04	-0.52	0.04	-0.52	0.04
NSW Local Court at j - 1	-0.56	0.20	-0.56	0.20	-0.56	0.20
Other Court at j - 1	-0.08	0.09	-0.08	0.09	-0.08	0.09
Accepted Liability at j - 1						
Partially Accepted Liability at j - 1	-0.04	0.03	-0.04	0.03	-0.04	0.03
Rejected Liabiliy at j - 1	0.15	0.03	0.15	0.03	0.15	0.03
Other at j - 1	0.19	0.02	0.19	0.02	0.19	0.02
Insurer A						
Insurer B	0.07	0.02	0.07	0.02	0.07	0.02
Insurer C	0.05	0.03	0.05	0.03	0.05	0.03
Insurer D	0.24	0.02	0.24	0.02	0.24	0.02
Insurer E	0.12	0.02	0.12	0.02	0.12	0.02
Insurer F	0.14	0.02	0.14	0.02	0.14	0.02
Insurer Other	0.07	0.03	0.07	0.03	0.07	0.03
Reporting Delay (log)	-0.52	0.01	-0.52	0.01	-0.52	0.01
sigma (intercept)			0.0001	0.02		
sigma (legrep)					0.0001	0.03
phi_1	-0.02	0.00	-0.02	0.00	-0.02	0.00
alpha	2.80	0.08	2.80	0.08	2.80	0.08
Log-Likelihood	-41,788.93		-41,788.93		-41,788.93	

Table 10.1: Delay Results FE vs RE

Unfortunately, both random effects models have their respective σ parameters approach 0 and the model approaches the fixed effects model. The coefficients and log likelihood for both random effects models are the same as the fixed effects model. This is not unexpected as the negative binomial is already accounting for the extra dispersion that would be modelled through random effects models through the α parameter. In fact, the negative binomial distribution can be considered as a mixture of Poisson distributions with varying mean parameters. The next section attempts to model the Delay process using the Poisson distribution.

10.4 Delay Process using a Poisson response distribution

The Poisson distribution is used in the GLARMA modelling of the delay process with minimal changes in the R programs. The Poisson has the following likelihood

$$L = \prod_{i=1}^{n} \prod_{j=1}^{m_i} \frac{(\mu_{j,i})^{t_{j,i}-1} e^{-\mu_{j,i}}}{(t_{j,i}-1)!} \tag{10.24}$$

and the corresponding log-likelihood function is

$$\ell = \sum_{i=1}^{n} \sum_{j=1}^{m_i} ((t_{j,i}-1)(\log \mu_{j,i})) - \mu_{j,i} - \log((t_{j,i}-1)!) \tag{10.25}$$

Refer to Dunsmuir et al. [2014a] for further details.

Table 10.2 shows the results of adopting a random effects GLARMA model for the delay process using a Poisson response distribution. Comparison are made between the fixed effects only model, a Random Intercept model and a Random Effect on Legally Represented claimants model. Two additional models were fit - both random effects assumed independent and both random effects assumed correlated. Further details on the bivariate random effects models in the GLARMA framework can be found in Dunsmuir et al. [2014a].

Variable	Fixed Effect Coeff.	Fixed Effect S.E.	Random Intercept Coeff.	Random Intercept S.E.	Random LegRep Coeff.	Random LegRep S.E.	2 RE - Indep. Coeff.	2 RE - Indep. S.E.	2 RE - Correlated Coeff.	2 RE - Correlated S.E.
Intercept	3.17	0.03	3.25	0.04	3.22	0.04	3.25	0.04	3.20	0.04
Legal Representation	-0.05	0.01	-0.05	0.01	-0.04	0.01	-0.05	0.01	-0.01	0.01
Previous Delay	0.02	0.01	0.03	0.01	0.02	0.01	0.03	0.02	0.03	0.02
Previous Change	0.08	0.01	0.07	0.01	0.07	0.01	0.07	0.01	0.07	0.01
Revision 1										
Revision 2	-0.73	0.02	-0.76	0.02	-0.74	0.02	-0.76	0.02	-0.76	0.02
Revision 3	-1.28	0.02	-1.36	0.02	-1.32	0.02	-1.35	0.02	-1.35	0.02
Revision 4	-1.59	0.02	-1.68	0.02	-1.65	0.02	-1.68	0.02	-1.67	0.02
Revision 5	-1.87	0.03	-1.98	0.03	-1.94	0.03	-1.98	0.03	-1.96	0.03
Revision 6 +	-2.22	0.02	-2.31	0.03	-2.29	0.03	-2.32	0.03	-2.31	0.03
No Litigation at j - 1										
NSW Supreme Court at j - 1	-0.23	0.08	-0.28	0.10	-0.28	0.10	-0.29	0.10	-0.28	0.10
NSW District Court at j - 1	-0.52	0.03	-0.56	0.03	-0.57	0.03	-0.57	0.03	-0.57	0.03
NSW Local Court at j - 1	-0.49	0.16	-0.46	0.18	-0.47	0.18	-0.46	0.18	-0.45	0.18
Other Court at j - 1	-0.08	0.07	-0.10	0.08	-0.12	0.09	-0.12	0.09	-0.12	0.09
Accepted Liability at j - 1										
Partially Accepted Liability at j - 1	-0.04	0.02	-0.04	0.03	-0.04	0.03	-0.04	0.03	-0.05	0.03
Rejected Liabiliy at j - 1	0.14	0.02	0.14	0.02	0.12	0.02	0.13	0.02	0.14	0.02
Other at j - 1	0.18	0.01	0.19	0.01	0.18	0.01	0.19	0.01	0.19	0.02
Insurer A										
Insurer B	0.05	0.01	0.06	0.02	0.07	0.02	0.07	0.02	0.06	0.02
Insurer C	0.05	0.02	0.04	0.02	0.06	0.02	0.05	0.02	0.04	0.02
Insurer D	0.20	0.02	0.20	0.02	0.21	0.02	0.20	0.02	0.20	0.02
Insurer E	0.11	0.02	0.12	0.02	0.11	0.02	0.11	0.02	0.11	0.02
Insurer F	0.15	0.02	0.15	0.02	0.16	0.02	0.16	0.02	0.16	0.02
Insurer Other	0.05	0.02	0.06	0.02	0.06	0.02	0.06	0.02	0.05	0.02
Reporting Delay (log)	-0.49	0.01	-0.52	0.01	-0.51	0.01	-0.52	0.01	-0.50	0.01
sigma (intercept)			0.27	0.01			0.19	0.01	0.29	0.01
sigma (legrep)					0.31	0.01	0.26	0.01	0.30	0.01
correl									-0.12	0.00
phi_1	-0.02	0.00	-0.05	0.00	-0.04	0.00	-0.05	0.00	-0.05	0.00
Log-Likelihood	-43,296.54		-42,984.67		-42,974.33		-42,921.84		-42,845.95	

Table 10.2: Delay Results FE vs RE - Poisson

The Poisson distribution works well in the sense that all the parameters converge to a finite value in most of the models. The random effects model are very sensitive to starting values for the parameters. In the case of the "two random effects with correlation" model, convergence was difficult to achieve. We used a grid search approach, where for fixed combinations of the elements of the RE covariance matrix, we maximised the log-likelihood by carrying out the Newton-Raphson method on the other parameters. The log-likelihoods are recorded and the values of the covariance matrix that produced the largest log-likelihood was chosen as the starting values for the Newton-Raphson algorithm for all the parameters.

Two observations can be made. Firstly, the random effects models fit better than the fixed effects model. In fact, each successive, and more complicated, model would reject the null hypothesis under a likelihood ratio test. Secondly, the log-likelihood for all the models are smaller than that of the Fixed Effect model using the negative binomial distribution. This may suggest the random effects framework are not as efficient in dealing with individual subject variability as the negative binomial distribution without using random effects.

10.5 Prediction of the Delay Process with Random Effects

One of the key benefits expected of in the usage of random effects models are more accurate predictions on an individual claims basis. This section compares the predictions of the Delay Component between fixed effects only and random effects model. The Random Intercept (or the Claimant Effect) model using a Poisson response is being used for this section. The parameters of this model can be found in Table 10.2 above.

The predicted delay, $\mu_{j,i}$ can be calculated using Equation (10.5), however, it requires the estimation of u_i's. We attempt to find the posterior mean of u_i using

$$\tilde{u_i} = \frac{1}{P(t_i)} \int_z z P(t_i|z) g(z) dz \tag{10.26}$$

where t_i is the set of observed times between revisions for the ith claim.

We once again use the Adaptive Gaussian Quadrature technique to estimate the

$$\tilde{u_i} \approx \frac{1}{L_i(\delta)} \left(\sum_{k=1}^{q} \tilde{z}_k \frac{\sqrt{\sigma}}{\sqrt{\pi}} \exp(\zeta_k^2) w_k \exp(\sum_{j=1}^{m_i} \ell(\tilde{z}_k) - \frac{\tilde{z}_k^2}{2}) \right) \tag{10.27}$$

Once the individual u_i's are found, expected values of the individual delays can be computed under the random effects framework. Figure 10.1 and Figure 10.2 show the comparison of the predictions from a fixed effects only model and a Random Intercept model for two sample claims. The grey bars presents the actual observed delays for each revision. The red lines maps the population average and the green and purple lines map the expected value from fixed effects only and

Random Intercept models, respectively. The fit is considerably better with the inclusion of the random intercept terms, especially for Sample Claim 2.

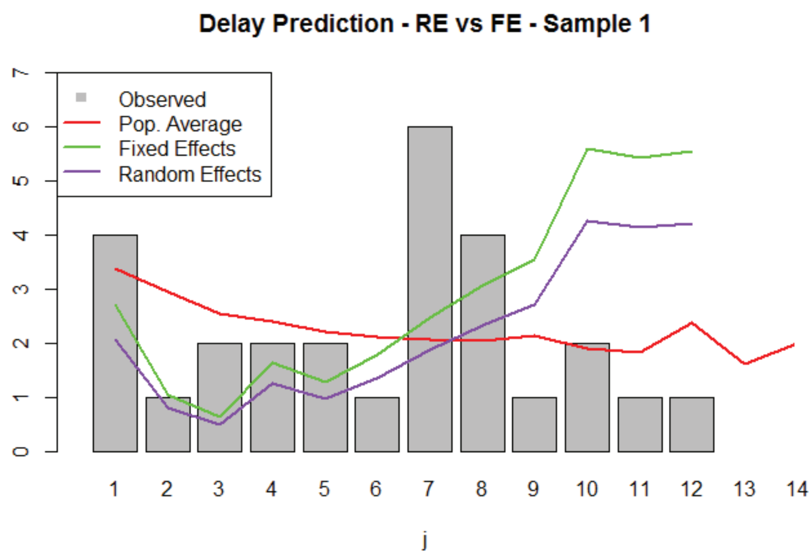

Figure 10.1: Predictions using Random Intercept - Sample Claim 1

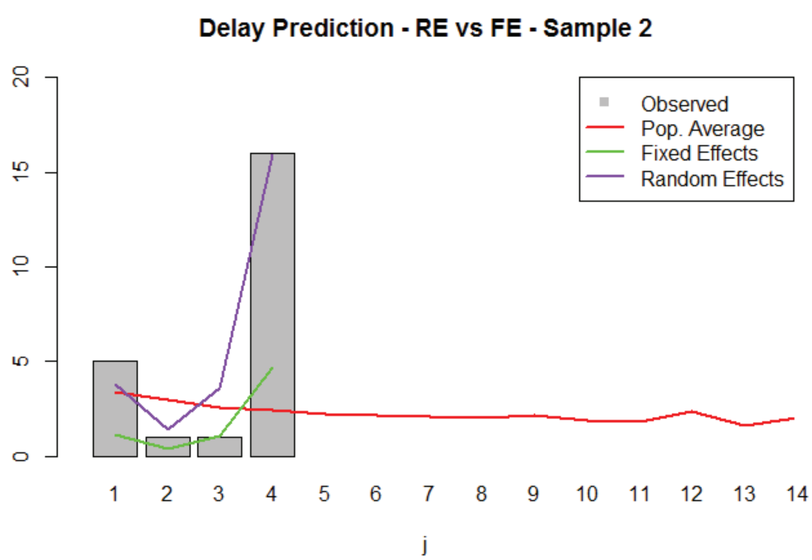

Figure 10.2: Predictions using Random Intercept - Sample Claim 2

10.6 Results of Random Effects Modelling - Size Process

The random effects model is also applied on the Size Process. Only minimal changes is required to adapt the R program to fit a gamma response distribution with RE terms. The results are shown in the table below. Only one random effect has been fit in each of the RE models. A bivariate RE could not be fit to the Size process as one of the σ's would converge to zero. Again, 7 quadrature points have been used in the model fitting.

Even though the random effects models are shown to be a better fit compared to the Fixed Effect model, the improvement in the log-likelihood, while significant, is not large. The results here confirm the initial belief that the delay process would be the most appropriate for the incorporation of random effects.

Variable	Fixed Effect		Random Intercept		Random LegRep	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Intercept	1.17	0.05	1.17	0.05	1.16	0.05
Saving on Finalisation	0.76	0.02	0.77	0.02	0.77	0.02
Legal Representation	0.06	0.02	0.06	0.02	0.05	0.02
Liability Rejected	0.16	0.04	0.17	0.04	0.16	0.04
Revision 1						
Revision 2	-0.22	0.02	-0.20	0.02	-0.21	0.02
Revision 3	-0.25	0.02	-0.23	0.02	-0.24	0.02
Revision 4	-0.32	0.03	-0.29	0.03	-0.30	0.03
Revision 5	-0.34	0.03	-0.31	0.03	-0.32	0.03
Revision 6 +	-0.48	0.03	-0.44	0.03	-0.46	0.03
Previous Incurred (log)	-0.15	0.00	-0.16	0.01	-0.16	0.00
Current Delay (Log)	0.20	0.01	0.20	0.01	0.20	0.01
Rejected Liabiliy at j	0.54	0.03	0.54	0.03	0.54	0.03
Reporting Delay (log)	0.06	0.01	0.06	0.01	0.05	0.01
sigma (intercept)			0.15	0.02		
sigma (legal rep)					0.13	0.02
phi_1	0.06	0.01	0.04	0.01	0.05	0.01
c	0.83	0.01	0.84	0.01	0.83	0.01
Log-likelihood	-16,782		-16,773		-16,777	

Table 10.3: Size Results FE vs RE

10.7 Conclusion

In this chapter we investigated the potential to incorporate random effects into the CDP framework. Due to the considerable complexity random effects add to the calculation of the derivatives of the log-likelihood, we have used Adaptive Gaussian Quadratures to estimate the derivatives.

We have applied the random effects model to the Delay component process with a negative binomial distribution as both a random intercept (claimant effect) and a random variable (lawyer effect, impacting only those claimants with legal representation). We have found the random effects were not needed when using a negative binomial distribution is used, it is suspected that the negative binomial distribution can already account for over-dispersion through the α parameter.

We then adopted a Poisson distribution for the delay component and fitted random effects models. The random effects parameters were significant when the Poisson distribution were used and the model fit was improved. The predictions (for delay) was compared across models with and without random effects and found the RE model's predictions were more accurate.

Chapter 11

Conclusions

11.1 Contributions to Individual Claims Modelling

This thesis makes contributions to the field of individual claims modelling. This area of actuarial science has received considerable attention in recent years. This research builds upon and extends some of the ideas and modelling framework establish by other researchers, in particular, that of Larsen [2007].

This thesis proposes a claims development framework that serves two main purposes. Firstly, the model fitting and resulting parameters allow the insurance companies to gain insights into the claims development processes of long tailed insurance claims. These insights will allow insurance companies to better manage claims and focus their attention on the types of claims that may deteriorate (worse outcomes for both the insurance company and the injured party).

Secondly, the framework can be used to project the ultimate claims cost of censored claims. Censoring occur due to the long tailed nature of the claims combined with the insurance companies' need to perform regular "valuations" of their active claims. Given the complicated nature of the model framework may prevent the insurance companies to perform valuations at an individual claims basis on a frequent basis, the ability to obtain an estimated ultimate claim size for every claim certainly has its merits.

This research also considers the ideas of serial dependence and random effects impacting on claims development. These areas are typically not accommodated in traditional methods of individual claims modelling. The ideas presented in this thesis may also be applied to non-valuation claims analysis; for example, in a rate making framework.

11.2 Contributions to Statistical Analysis

This thesis also makes contributions to the statistical field of longitudinal data analysis by using ideas with origins in time series and econometrics. GLARMA based modelling that incorporates the GLM based of analysis for covariates with the serial dependence structure of ARMA based models. While these types of models have been applied widely to long time series of count data, this thesis explores the application of this framework to a large number of shorter trajectories. The application extends easily to other panel data problems. This research also extends the application to continuous outcomes in the gamma and generalised gamma distributions.

This thesis also explores the incorporation of random effects into the GLARMA framework. Due to the complexity of calculating the derivatives of the log-likelihood function and Pearson residuals with respect to δ, we have used Adaptive Gaussian Quadratures to estimate the derivatives. Random effects have been applied to GLARMA models with negative binomial, Poisson and gamma error distributions.

Censoring for complicated model structures have been examined in this thesis as censoring is a feature common to long tail insurance claims. The adjustment to the maximum likelihood estimation process to allow for censoring incorporates additional complexity to the CDP framework. Furthermore, the projection of censored trajectories is undertaken using Monte-Carlo simulations.

Due to working with large datasets and complicated models, this thesis also developed strategies to make model fitting more effective. For example, parallel computing, strategies to aid convergence as well as using likelihood approximation techniques were adopted to speed up the fitting of the computation intensive models.

11.3 Limitations of the approach

While there are numerous benefits to modelling claims on an individual level using the Claims Development Process framework proposed in this research, there are also a number of limitations.

Firstly, the models are complicated. While the Chain Ladder projection models can be summarised using around ten to twenty ICD factors (one per development period) the CDP utilises around 300 parameters across the four component models. This complicated structure also makes the relationship of individual parameters and the ultimate claims projection less transparent. This may appear to be a "black box" to the average actuary in the sense that the impact of changes in the parameters over time does not have an explicit and obvious relationship with the ultimate claim sizes predicted. This may also create an issue with auditors who needs to understand the workings of the model from a corporate governance perspective.

Secondly, the modelling process is also very time consuming. Even with modern computing hours, the time required for model fitting of the component models as well as the projection methodologies outlined in this research is measured in days. In contrast, the time required to apply the aggregated models would be in hours or, in the case of the chain ladder models, possibly minutes. This may be a hindrance to actuaries that may wish to adopt this projection method on a frequent basis. However, we have proposed solutions to reduce the time required to yield projection results using a Gaussian approximation; which can be used on a regular basis with periodic calibration based on the full projection model.

Thirdly, the depth of statistical knowledge as well as coding knowledge may place this framework out of the reach of average general insurance actuaries. However, hopefully this thesis as well as continued research in the individual claims modelling area will allow a general appreciation of the benefits of these models and hasten the adoption of some of the techniques into practice.

11.4 Further areas of research

While the between claimant variability was examined and significant and informative results were obtained, the random effects framework have not been incorporated into ultimate claim size projection. This is partly a result of the computational intensiveness of the random effects model fitting; this makes the iterative process of debugging the projection process prohibitively lengthy. A further area of research would be to investigate coding algorithmic changes that allows the model fitting of random effects, in particular, making the adaptive gaussian quadrature fitting speedier. The model fitting in this research utilised multi-threaded processing, but the time required is still lengthy.

One of the original objectives of the research was to incorporate spatial features of CTP claims data into the modelling framework. While considerable investigation was undertaken to examine the spatial features of the data in Appendix H these features have not been incorporated into CDP framework. A possible extension of the framework is to incorporate spatial features which may further explain the between claimant variability as well as the serial dependance structure. A potential approach would be to use a hierarchical structure and model the parameters of the component processes (coefficients and dispersion parameters) using spatial modelling techniques. Using such a nested structure could allow for the geographical variations with in the CTP data while preserving the CDP model framework.

Another intention of the research was to investigate legislative changes and how such changes impact on the claim development behaviours observed. Such time structured changes would be feasible to be investigated using the GLARMA framework; and the results may be potentially interesting. During the data period, there has been two major legislative changes - the introduction of LTCS and a change to the ANF threshold. The former shifted the responsibility of ongoing medical and care requirements for catastrophically injured claimants from the claimant themselves to the government and the latter increased the threshold for ANFs which are a simplified claims process aimed to speed up access to minor medical procedures. Unfortunately, the LTCS only impacts a few hundred of claims in the modelling dataset and the ANF change only occurred towards the end of the data period. The timing and the impact of these changes made their analysis difficult; however, it would be of interest to gauge the impact of these legislative changes when the data matures for another few years.

Appendix A

Acronyms and Glossary

AIS - Abbreviated Injury Score, main injury coding methodology used to describe the injuries sustained from motor vehicle accidents for the NSW CTP scheme.

ANF - Accident Notification Form, a process to quickly claim for medical cost post an accident by submitting a simple form. This is formally referred to s49 claims.

AvE - Actual versus Expected, an actuarial tools that compared actual outcome with that of expected.

CARS - Claims Assessment and Resolution Service, a non-adversarial resolution to claim disputes that precedes any litigation through the court system.

CDP - Claim Development Processes, the modelling framework developed throughout this thesis.

CTP - Compulsory Third Party, insurance product in Australia that is compulsory by legislation for all vehicles to cover bodily injuries resulting from traffic accidents

IBNER - Incurred But Not Enough Reported, claims has been reported to the insurer, but the complete set of the claim circumstances are still not fully known.

IBNR - Incurred But Not Reported, claims that has happened but not yet known to an insurer.

ICD - Incurred Cost Development, an actuarial projection method that uses historically observed claims cost development ratios to project future claims incurred cost.

ISS - Injury Severity Scale, a 0 - 75 scale used to index the overall severity of the claimants injuries.

MAA - Motor Accidents Authority, the government regulator of NSW's CTP scheme

MAA 1989 - The previous CTP legislation that the MACA 1999 replace. Claims cost, and premiums, spiral under this legislation.

MACA 1999 - Motor Accident and Compensation Act 1999, the current legislation that regulates the NSW CTP scheme.

LTCS - Longterm Care and Support Act 2007, the legislation that mandates catastrophically injured NSW CTP claimants be looked after centrally by the LTCS scheme so the claimant does not need to manage the sizeable sum of money that is needed to pay for future medical and care costs.

NEL - Non-economic Loss, the compensation for pain and suffering. NEL is limited to the more severe claims post MACA 1999 to contain the sharp increases in CTP premiums in the 1990's.

PCE - Projected Case Estimate, an actuarial projection method that uses historically observed case estimate development patterns to project future claims cost.

PIR - Personal Injury Register, all insurers of the NSW CTP scheme are required by law to submit their claims information to the PIR on a quarterly basis.

s49 - Small claims as denoted by Section 49 of the Motor Accidents Compensation Act 1999. Once an Accident Notification Form is filled, injured party can claim up to $500 of damages without the need to demonstrate vehicle at-fault. These small claims are also known as ANFs. The $500 limit was raised to $5000 under MACA Amendment 2009.

s74 - Full claims as denoted by Section 74 of the Motor Accidents Compensation

Act 1999. Police reports needs to be submitted with the claim form so vehicle at-fault can be established. The driver of the vehicle at-fault is not eligible to claim an s74 claim.

SCE - Statistical Case Estimation, an actuarial model that predicts the ultimate size of workers compensation claim by aggregating the likely payments made to each individual claimant.

TAC - Transport Accident Commission, the equivalent of MAA for Victoria. Since Victoria adopts a government underwritten no-fault scheme, TAC is also the claims manager of CTP claims occurred in Victoria.

WPI - Whole person impairment, a medically determined index denoting the severity of the injury. In the NSW CTP scheme, a WPI score of more than 10% is required for the claimant to be eligible for compensation for non-economic loss (also know as, general damages or pain and suffering).

Appendix B

Traditional Actuarial Claim Projection Methods

B.1 ICD

The following tables demonstrate the ICD modelling applied to the NSW CTP dataset.

Accident		X_j, by development period (j) - periods after the accident period																	
period	k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2001HY1	1	96	251	286	323	372	395	432	448	451	449	457	463	463	463	465	464	466	463
2001HY2	2	101	229	262	307	337	358	416	426	432	431	424	420	420	423	420	421	423	
2002HY1	3	100	259	315	338	373	399	428	455	462	468	470	462	460	462	456	447		
2002HY2	4	93	255	276	302	339	362	415	418	418	422	422	416	414	415	420			
2003HY1	5	88	231	268	315	336	387	430	453	447	443	425	427	423	422				
2003HY2	6	136	312	358	398	452	485	520	517	501	482	490	484	487					
2004HY1	7	148	342	374	441	489	506	550	573	570	558	549	548						
2004HY2	8	144	304	362	417	452	467	499	494	479	480	478							
2005HY1	9	136	359	414	455	487	517	530	532	509	495								
2005HY2	10	173	390	435	476	515	523	558	543	532									
2006HY1	11	186	445	492	528	537	558	578	583										
2006HY2	12	172	414	456	481	507	517	535											
2007HY1	13	251	479	502	544	546	566												
2007HY2	14	178	386	448	482	526													
2008HY1	15	170	407	448	491														
2008HY2	16	181	416	483															
2009HY1	17	174	430																
2009HY2	18	188																	

Table B.1: Example of ICD Projection - Data

In Table B.1 the raw incurred cost triangle of the modelling data is presented again. From this table, the observed individual ICD factors can derived as well as the observed overall weighted average ICD factors and the observed weighted average ICD factors for latest two years. A set of $\hat{\lambda}_j$ has been selected, which is the weight average of the latest two years in most cases. These are show in Table

B.2.

Accident		ICD factors, X_j/X_j-1, by development period (j) - periods after the accident perioc																		
period	k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	D
2001HY1	1		2.60	1.14	1.13	1.15	1.06	1.09	1.04	1.01	1.00	1.02	1.01	1.00	1.00	1.01	1.00	1.00	0.99	
2001HY2	2		2.27	1.14	1.17	1.10	1.06	1.16	1.02	1.01	1.00	0.98	0.99	1.00	1.01	0.99	1.00	1.00		
2002HY1	3		2.59	1.22	1.07	1.10	1.07	1.07	1.06	1.02	1.01	1.00	0.98	1.00	1.00	0.99	0.98			
2002HY2	4		2.74	1.08	1.09	1.12	1.07	1.15	1.01	1.00	1.01	1.00	0.98	0.99	1.00	1.01				
2003HY1	5		2.64	1.16	1.18	1.07	1.15	1.11	1.05	0.99	0.99	0.96	1.00	0.99	1.00					
2003HY2	6		2.30	1.15	1.11	1.14	1.07	1.07	1.00	0.97	0.96	1.02	0.99	1.01						
2004HY1	7		2.32	1.09	1.18	1.11	1.04	1.09	1.04	0.99	0.98	0.98	1.00							
2004HY2	8		2.11	1.19	1.15	1.08	1.03	1.07	0.99	0.97	1.00	0.99								
2005HY1	9		2.64	1.15	1.10	1.07	1.06	1.03	1.00	0.96	0.97									
2005HY2	10		2.25	1.12	1.10	1.08	1.02	1.07	0.97	0.98										
2006HY1	11		2.40	1.10	1.07	1.02	1.04	1.03	1.01											
2006HY2	12		2.41	1.10	1.06	1.05	1.02	1.03												
2007HY1	13		1.90	1.05	1.08	1.00	1.04													
2007HY2	14		2.17	1.16	1.08	1.09														
2008HY1	15		2.39	1.10	1.10															
2008HY2	16		2.30	1.16																
2009HY1	17		2.47																	
2009HY2	18																			
Weighted Average			2.34	1.13	1.11	1.08	1.05	1.08	1.02	0.99	0.99	1.00	0.99	1.00	1.00	1.00	0.99	1.00	0.99	
Wghtd. Ave. (last 2 yrs)			2.33	1.12	1.08	1.04	1.03	1.04	0.99	0.98	0.98	0.99	0.99	1.00	1.00	1.00	0.99	1.00	0.99	
Selected			2.45	1.15	1.08	1.04	1.03	1.04	0.99	0.98	0.98	0.99	0.99	1.00	1.00	1.00	0.99	1.00	1.00	1.005

Table B.2: Example of ICD Projection - ICD Factors

Attention is drawn to the last λ selection in the table, it is the tail factor to incorporate the development from $j = 17$ to $j = D$. Equation (4.4) may also be written as

$$\hat{X}_k = X_{k,T} \left(\prod_{j=T+1}^{18-T} \hat{\lambda}_j \right) \lambda' \tag{B.1}$$

where $\lambda' = \prod_{j=18}^{D} \hat{\lambda}_j$, a tail ICD factor. Since the data only reaches $j = 17$ it does not offer an insight into how claims develop after this time. Hence, a somewhat arbitrary "tail" development factor, $\prod_{j=18}^{D} \hat{\lambda}_j = 1.005$ has been adopted.

Table B.3 shows the projection of the ultimate claims cost by applying the $\hat{\lambda}_j$'s consecutively. In particular the last column of Table B.3 is the series of $\hat{X}_k$ for each of the accident periods.

Accident		X_j, by development period (j) - periods after the accident period																		
period	k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	X_k
2001HY1	1	96	251	286	323	372	395	432	448	451	449	457	463	463	463	465	464	466	463	465
2001HY2	2	101	229	262	307	337	358	416	426	432	431	424	420	420	423	420	421	423	423	426
2002HY1	3	100	259	315	338	373	399	428	455	462	468	470	462	460	462	456	447	449	449	451
2002HY2	4	93	255	276	302	339	362	415	418	418	422	422	416	414	415	420	417	419	419	421
2003HY1	5	88	231	268	315	336	387	430	453	447	443	425	427	423	422	421	418	420	420	422
2003HY2	6	136	312	358	398	452	485	520	517	501	482	490	484	487	488	488	484	486	486	489
2004HY1	7	148	342	374	441	489	506	550	573	570	558	549	548	547	549	548	544	546	546	549
2004HY2	8	144	304	362	417	452	467	499	494	479	480	478	475	473	475	474	471	473	473	476
2005HY1	9	136	359	414	455	487	517	530	532	509	495	490	486	485	487	486	483	485	485	487
2005HY2	10	173	390	435	476	515	523	558	543	532	521	515	512	510	512	511	508	510	510	513
2006HY1	11	186	445	492	528	537	558	578	583	569	557	551	548	546	548	547	543	546	546	548
2006HY2	12	172	414	456	481	507	517	535	532	519	508	503	500	498	500	499	496	498	498	500
2007HY1	13	251	479	502	544	546	566	589	586	571	559	553	550	548	550	549	545	548	548	551
2007HY2	14	178	386	448	482	526	541	562	559	545	534	528	525	523	525	524	521	523	523	526
2008HY1	15	170	407	448	491	511	525	546	543	530	519	513	510	508	510	509	506	508	508	511
2008HY2	16	181	416	483	521	542	557	579	576	562	550	544	541	539	541	540	536	539	539	542
2009HY1	17	174	430	494	533	554	570	593	589	575	563	557	553	551	553	553	549	551	551	554
2009HY2	18	188	460	529	570	593	610	634	631	615	602	596	592	590	592	591	587	590	590	593

Table B.3: Example of ICD Projection - Projection

The same steps are then repeated after adding in the validation data. The details of the projection have been omitted but follow the same format as the ICD projection using only the modelling data. The results of this projection are then compared to the results of the projection based on the modelling data along. The comparison is shown in Table B.4.

Accident Period	Using 2001-2009 Modelling Data					Using 2010-2012 Validation Data			
	Number of Reported Claims	Incurred Cost to Date ($m)	Projected Ult Incurred ($m)	Proj. Ult. Number of Claims	Average Claim Size ($)	Projected Ult Incurred ($'m)	Proj. Ult. Number of Claims	Average Claim Size ($)	Over or Under Projection ($'m)
2001HY1	6,284	463	465	6,284	74,002	460	6,283	73,287	5
2001HY2	5,699	423	426	5,705	74,589	424	5,703	74,283	2
2002HY1	5,612	447	451	5,618	80,321	453	5,614	80,618	-1
2002HY2	5,110	420	421	5,115	82,292	435	5,117	85,026	-14
2003HY1	5,022	422	422	5,027	84,037	428	5,032	85,148	-6
2003HY2	4,946	487	489	4,951	98,736	498	4,953	100,541	-9
2004HY1	5,080	548	549	5,085	108,006	535	5,093	105,081	14
2004HY2	4,835	478	476	4,840	98,296	477	4,874	97,927	-2
2005HY1	4,927	495	487	4,932	98,789	496	4,957	100,018	-9
2005HY2	4,472	532	513	4,485	114,294	514	4,512	113,836	-1
2006HY1	4,599	583	548	4,624	118,578	542	4,661	116,220	7
2006HY2	4,436	535	500	4,478	111,725	491	4,510	108,922	9
2007HY1	4,560	566	551	4,642	118,596	525	4,652	112,793	26
2007HY2	4,397	526	526	4,527	116,112	518	4,520	114,601	8
2008HY1	4,239	491	511	4,445	114,895	520	4,471	116,290	-9
2008HY2	4,096	483	542	4,429	122,264	565	4,444	127,135	-23
2009HY1	4,190	430	554	4,872	113,720	561	4,891	114,726	-7
2009HY2	1,756	188	593	4,696	126,258	604	4,919	122,692	-11
Total	84,260	8,517	9,023	88,755	101,660	9,045	89,206	101,398	-23

Table B.4: Example of ICD Projection using Validation Data

B.2 Projected Case Estimation Model

While the ICD model is very easy to use and under some circumstances produces adequate results, actuaries soon developed other methods. One of them is the Projected Case Estimate (PCE) method and it addresses two flaws of the ICD model. These are discussed in turn.

Firstly, the ICD model only projects the ultimate claim size and not when these claims will be paid at each point in time. For cashflow management and to work out the time value of money (or discounting) aspect of the claim liabilities, a separate analysis needs to be done. Such a cashflow analysis examines past payment patterns and use that as a proxy for the future. But typically payment patterns developed in isolation to incurred patterns can be misleading.

The second flaw is that at a given point in time only a portion of the incurred are related to open claims, or the claims that are subject to further development. That is, for a given $X_{k,j}$, only the $CE_{k,j}$ portion (based on the estimation of claim managers) is subject to future development and volatility; the paid portion (most related to finalised claims) do not change in the future.

The PCE model addresses these two issues. It defines two sets of parameters - the Payout (PO) factors and the Case Estimation Development (CED) factors as opposed to the one set of ICD factors used in the ICD model. To ease notation, we further define the incremental payment variable $IP_{k,j} = P_{k,j} - P_{k,j-1}$

The CED factors (γ) and PO factors (ϕ) are defined as follows

$$\gamma_{k,j} = \frac{CE_{k,j} + IP_{k,j}}{CE_{k,j-1}} \tag{B.2}$$

and

$$\phi_{k,j} = \frac{IP_{k,j}}{CE_{k,j-1}} \tag{B.3}$$

And similarly to the ICD model, once the γ's and ϕ's are estimated, the future payments and future case estimates can be successively and alternatingly projected. The estimate of the parameters takes on a similar process as the ICD model, either use the complete history, use the latest number of years or manually select these factors by examining trends in the data or external knowledge. Actuaries in practice tend to use the a fixed number of the most recent years of data to estimate a set of parameters more reflective of recent trends. We have once again adopted selections that are weighted averages of the most recent two years.

$$\hat{\gamma}_j = \frac{\sum_{k=\max(1,15-j)}^{18-j} (CE_{k,j} + IP_{k,j})}{\sum_{k=\max(1,15-j)}^{18-j} CE_{k,j-1}} \tag{B.4}$$

and

$$\hat{\phi}_j = \frac{\sum_{k=\max(1,15-j)}^{18-j} IP_{k,j}}{\sum_{k=\max(1,15-j)}^{18-j} CE_{k,j-1}} \tag{B.5}$$

Once the set of $\hat{\gamma}_j$ and $\hat{\phi}_j$ factors is selected the case estimate and payments for a future development period, $j \in (T+1, D)$, are projected alternatingly and successively as follows

$$\hat{IP}_{k,j} = \hat{\phi}_j \hat{CE}_{k,j-1} \tag{B.6}$$

and

$$\hat{CE}_{k,j} = \hat{\gamma}_j \hat{CE}_{k,j-1} - \hat{IP}_{k,j} \tag{B.7}$$

starting from $CE_{k,T}$ which is the case estimate for accident period k as observed at the censoring date.

Again, the whole projection approach is repeated after incorporating the validation data and the two rounds of projection are summarised in Table B.5.

	Using 2001-2009 Modelling Data					Using 2010-2012 Validation Data			
Accident Period	**Number of Reported Claims**	**Incurred Cost to Date ($m)**	**Projected Ult Incurred ($m)**	**Proj. Ult. Number of Claims**	**Average Claim Size ($)**	**Projected Ult Incurred ($m)**	**Proj. Ult. Number of Claims**	**Average Claim Size ($)**	**Over or Under Projection ($m)**
2001HY1	6,284	463	459	6,284	73,041	460	6,283	73,287	-1
2001HY2	5,699	423	418	5,705	73,330	422	5,703	73,967	-4
2002HY1	5,612	447	443	5,618	78,936	450	5,614	80,140	-6
2002HY2	5,110	420	414	5,115	81,008	431	5,117	84,178	-16
2003HY1	5,022	422	416	5,027	82,801	424	5,032	84,206	-7
2003HY2	4,946	487	481	4,951	97,230	491	4,953	99,161	-10
2004HY1	5,080	548	540	5,085	106,192	529	5,093	103,817	11
2004HY2	4,835	478	470	4,840	97,072	467	4,874	95,768	3
2005HY1	4,927	495	482	4,932	97,640	485	4,957	97,814	-3
2005HY2	4,472	532	508	4,485	113,227	500	4,512	110,812	8
2006HY1	4,599	583	542	4,624	117,212	523	4,661	112,129	19
2006HY2	4,436	535	496	4,478	110,808	474	4,510	105,161	22
2007HY1	4,560	566	545	4,642	117,496	507	4,652	108,926	39
2007HY2	4,397	526	521	4,527	115,070	492	4,520	108,895	29
2008HY1	4,239	491	506	4,445	113,815	490	4,471	109,508	16
2008HY2	4,096	483	536	4,429	120,995	523	4,444	117,775	13
2009HY1	4,190	430	532	4,872	109,119	513	4,891	104,834	19
2009HY2	1,756	188	542	4,696	115,476	537	4,919	109,066	6
Total	84,260	8,517	8,852	88,755	99,738	8,716	89,206	97,709	136

Table B.5: Example of PCE Projection - Summary

The PCE model projected the total cost over the nine accident years to be \$8,852m, slightly lower than the ICD model. However, using the updated data, the projected ultimate cost has reduced by \$136m. The change is of a larger magnitude than the change witnessed in the ICD models.

The following tables demonstrate the basics of the PCE model by applying the methodology to the NSW CTP data. Tables B.6 and B.7 show the incremental payments (IP) and case estimate (CE) data in the triangle format.

Accident period	k	IP_j, by development period (j) - periods after the accident period																	
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2001HY1	1	1	8	10	13	19	47	29	57	56	52	31	24	24	19	11	13	6	4
2001HY2	2	1	8	9	16	13	28	41	61	53	39	40	21	16	17	13	7	5	
2002HY1	3	1	9	13	14	19	26	36	56	45	48	54	34	21	18	9	7		
2002HY2	4	2	9	9	12	17	27	45	55	54	42	33	35	15	10	14			
2003HY1	5	2	8	9	16	17	30	42	54	63	48	33	34	12	5				
2003HY2	6	2	8	14	19	33	30	58	71	70	49	37	13	12					
2004HY1	7	2	9	11	21	23	48	53	79	77	67	21	25						
2004HY2	8	2	9	11	20	32	47	66	61	67	28	35							
2005HY1	9	2	10	13	22	33	40	57	79	64	38								
2005HY2	10	3	11	15	25	39	41	77	67	61									
2006HY1	11	2	11	14	30	35	52	63	72										
2006HY2	12	3	11	17	27	41	46	54											
2007HY1	13	3	14	16	30	37	49												
2007HY2	14	3	12	21	31	40													
2008HY1	15	2	12	19	29														
2008HY2	16	3	14	22															
2009HY1	17	3	15																
2009HY2	18	3																	

Table B.6: Example of PCE Projection - $IP_{k,j}$ (\$m)

Accident	k	CE_j, by development period (j) - periods after the accident period																	
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2001HY1	1	95	241	266	290	321	297	305	264	211	157	134	116	91	72	64	49	45	37
2001HY2	2	100	220	244	273	289	281	299	248	200	161	114	88	72	59	42	36	34	
2002HY1	3	98	249	292	301	317	317	311	282	244	202	149	108	85	69	54	38		
2002HY2	4	91	244	257	271	291	288	296	243	189	151	119	78	60	52	42			
2003HY1	5	86	221	250	282	285	306	307	276	207	156	105	73	56	50				
2003HY2	6	134	302	335	356	377	380	356	282	196	128	99	80	71					
2004HY1	7	146	332	351	398	422	391	382	326	246	168	137	111						
2004HY2	8	142	294	341	375	377	346	313	246	164	138	100							
2005HY1	9	134	346	389	408	407	397	353	277	190	137								
2005HY2	10	171	376	406	422	423	390	348	266	194									
2006HY1	11	183	433	466	471	445	414	371	304										
2006HY2	12	170	401	425	424	408	373	337											
2007HY1	13	249	463	470	482	446	418												
2007HY2	14	175	371	413	416	419													
2008HY1	15	168	393	415	429														
2008HY2	16	178	399	444															
2009HY1	17	171	412																
2009HY2	18	185																	

Table B.7: Example of PCE Projection - $CE_{k,j}$ (\$m)

Tables B.8 and B.9 show the individual observed PO factors and CED factors, respectively. In red, the selected factors are shown, which is the observed weighted average of latest two years. Again, factors for $j = 18$ to $j = D$ have been condensed into one factor and by necessity $\phi_D = \gamma_D$ to ensure all claims are paid at development period D and there is no case estimates remaining.

Accident period	k	PO factors, IP_j/CE_j-1, by development period (j) - periods after the accident perioc 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	D
2001HY1	1		0.09	0.04	0.05	0.06	0.15	0.10	0.19	0.21	0.25	0.20	0.18	0.21	0.21	0.16	0.20	0.13	0.10	
2001HY2	2		0.08	0.04	0.06	0.05	0.10	0.14	0.21	0.21	0.20	0.25	0.18	0.18	0.24	0.22	0.17	0.13		
2002HY1	3		0.09	0.05	0.05	0.06	0.08	0.11	0.18	0.16	0.20	0.27	0.23	0.19	0.21	0.13	0.12			
2002HY2	4		0.10	0.04	0.05	0.06	0.09	0.15	0.19	0.22	0.22	0.22	0.30	0.19	0.17	0.27				
2003HY1	5		0.09	0.04	0.06	0.06	0.10	0.14	0.18	0.23	0.23	0.21	0.32	0.17	0.09					
2003HY2	6		0.06	0.04	0.06	0.09	0.08	0.15	0.20	0.25	0.25	0.29	0.13	0.15						
2004HY1	7		0.06	0.03	0.06	0.06	0.11	0.13	0.21	0.24	0.27	0.13	0.18							
2004HY2	8		0.06	0.04	0.06	0.09	0.12	0.19	0.20	0.27	0.17	0.25								
2005HY1	9		0.08	0.04	0.06	0.08	0.10	0.14	0.22	0.23	0.20									
2005HY2	10		0.06	0.04	0.06	0.09	0.10	0.20	0.19	0.23										
2006HY1	11		0.06	0.03	0.06	0.07	0.12	0.15	0.19											
2006HY2	12		0.06	0.04	0.06	0.10	0.11	0.14												
2007HY1	13		0.05	0.03	0.06	0.08	0.11													
2007HY2	14		0.07	0.06	0.07	0.10														
2008HY1	15		0.07	0.05	0.07															
2008HY2	16		0.08	0.05																
2009HY1	17		0.09																	
2009HY2	18																			
Weighted Average			0.07	0.04	0.06	0.08	0.11	0.15	0.20	0.23	0.22	0.23	0.22	0.18	0.19	0.19	0.17	0.13	0.10	
Wghtd. Ave. (last 2 yrs)			0.08	0.05	0.07	0.09	0.11	0.16	0.20	0.24	0.23	0.21	0.23	0.18	0.18	0.19	0.17	0.13	0.10	
Selected			0.08	0.05	0.07	0.09	0.11	0.16	0.20	0.24	0.23	0.21	0.23	0.18	0.18	0.19	0.17	0.13	0.10	0.900

Table B.8: Example of PCE Projection - PO Factors

Accident period	k	CED factors, (CE_j + IP_j)/CE_j-1, by development period (j) - periods after the accident period 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	D
2001HY1	1		2.62	1.15	1.14	1.17	1.07	1.12	1.05	1.01	0.99	1.05	1.04	0.99	1.00	1.04	0.97	1.04	0.93	
2001HY2	2		2.28	1.15	1.18	1.11	1.07	1.21	1.03	1.02	1.00	0.96	0.96	1.00	1.05	0.94	1.03	1.06		
2002HY1	3		2.62	1.23	1.08	1.12	1.08	1.09	1.09	1.02	1.03	1.01	0.95	0.98	1.02	0.91	0.84			
2002HY2	4		2.77	1.09	1.10	1.14	1.08	1.18	1.01	1.00	1.02	1.01	0.95	0.97	1.03	1.08				
2003HY1	5		2.67	1.17	1.19	1.07	1.18	1.14	1.08	0.98	0.98	0.88	1.02	0.94	0.98					
2003HY2	6		2.32	1.15	1.12	1.15	1.09	1.09	0.99	0.94	0.90	1.06	0.94	1.03						
2004HY1	7		2.33	1.09	1.19	1.12	1.04	1.11	1.06	0.99	0.95	0.95	0.99							
2004HY2	8		2.13	1.20	1.16	1.09	1.04	1.09	0.98	0.94	1.01	0.98								
2005HY1	9		2.66	1.16	1.11	1.08	1.07	1.03	1.01	0.92	0.93									
2005HY2	10		2.27	1.12	1.10	1.09	1.02	1.09	0.96	0.96										
2006HY1	11		2.42	1.11	1.08	1.02	1.05	1.05	1.02											
2006HY2	12		2.43	1.10	1.06	1.06	1.03	1.05												
2007HY1	13		1.91	1.05	1.09	1.00	1.05													
2007HY2	14		2.18	1.17	1.08	1.10														
2008HY1	15		2.41	1.11	1.10															
2008HY2	16		2.32	1.17																
2009HY1	17		2.50																	
2009HY2	18																			
Weighted Average			2.36	1.13	1.11	1.09	1.06	1.10	1.02	0.98	0.98	0.99	0.98	0.99	1.02	0.99	0.94	1.05	0.93	
Wghtd. Ave. (last 2 yrs)			2.35	1.12	1.08	1.05	1.03	1.05	0.99	0.95	0.95	0.96	0.97	0.98	1.02	0.99	0.94	1.05	0.93	
Selected			2.35	1.12	1.08	1.05	1.03	1.05	0.99	0.95	0.95	0.96	0.97	0.98	1.02	0.99	0.94	1.05	0.93	0.900

Table B.9: Example of PCE Projection - CED factors

By utilising the successive and alternating projection outlined above, Tables B.10 and B.11 show the projected case estimates and more importantly the projected incremental payments. The incremental payments by themselves reveal the projected ultimate cost as well as the timing when the claim costs are expected to be paid. The last column of Table B.11 is estimated claims cost for each accident period, X_k, and is the sum of the IP_j's.

Accident period	k	CE_j, by development period (j) - periods after the accident period 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	D
2001HY1	1	95	241	266	290	321	297	305	264	211	157	134	116	91	72	64	49	45	37	0
2001HY2	2	100	220	244	273	289	281	299	248	200	161	114	88	72	59	42	36	34	28	0
2002HY1	3	98	249	292	301	317	317	311	282	244	202	149	108	85	69	54	38	35	29	0
2002HY2	4	91	244	257	271	291	288	296	243	189	151	119	78	60	52	42	33	30	25	0
2003HY1	5	86	221	250	282	285	306	307	276	207	156	105	73	56	50	40	31	29	24	0
2003HY2	6	134	302	335	356	377	380	356	282	196	128	99	80	71	59	48	37	34	28	0
2004HY1	7	146	332	351	398	422	391	382	326	246	168	137	111	89	75	60	47	43	36	0
2004HY2	8	142	294	341	375	377	346	313	246	164	138	100	74	60	50	40	31	29	24	0
2005HY1	9	134	346	389	408	407	397	353	277	190	137	103	76	61	51	41	32	29	24	0
2005HY2	10	171	376	406	422	423	390	348	266	194	139	104	77	62	52	42	32	30	25	0
2006HY1	11	183	433	466	471	445	414	371	304	217	155	116	86	69	58	47	36	33	28	0
2006HY2	12	170	401	425	424	408	373	337	266	189	136	102	75	61	51	41	32	29	24	0
2007HY1	13	249	463	470	482	446	418	374	295	210	151	113	84	67	57	45	35	32	27	0
2007HY2	14	175	371	413	416	419	388	348	274	195	140	105	78	63	53	42	33	30	25	0
2008HY1	15	168	393	415	429	411	381	341	269	191	137	103	76	61	52	41	32	29	24	0
2008HY2	16	178	399	444	451	433	401	359	283	201	144	108	80	65	54	43	34	31	26	0
2009HY1	17	171	412	442	448	430	398	357	281	200	144	108	80	64	54	43	33	31	26	0
2009HY2	18	185	421	452	459	440	408	365	288	205	147	110	82	66	55	44	34	31	26	0

Table B.10: Example of PCE Projection - CE projection ($m)

Accident period	IP_j, by development period (j) - periods after the accident period 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	D	X_k
2001HY1	1	8	10	13	19	47	29	57	56	52	31	24	24	19	11	13	6	4	34	459
2001HY2	1	8	9	16	13	28	41	61	53	39	40	21	16	17	13	7	5	3	25	418
2002HY1	1	9	13	14	19	26	36	56	45	48	54	34	21	18	9	7	5	4	26	443
2002HY2	2	9	9	12	17	27	45	55	54	42	33	35	15	10	14	7	4	3	23	414
2003HY1	2	8	9	16	17	30	42	54	63	48	33	34	12	5	9	7	4	3	22	416
2003HY2	2	8	14	19	33	30	58	71	70	49	37	13	12	13	11	8	5	3	25	481
2004HY1	2	9	11	21	23	48	53	79	77	67	21	25	20	16	14	10	6	4	32	540
2004HY2	2	9	11	20	32	47	66	61	67	28	35	23	13	11	9	7	4	3	21	470
2005HY1	2	10	13	22	33	40	57	79	64	38	29	24	14	11	10	7	4	3	22	482
2005HY2	3	11	15	25	39	41	77	67	61	44	30	24	14	11	10	7	4	3	22	508
2006HY1	2	11	14	30	35	52	63	72	73	50	33	27	15	13	11	8	5	3	25	542
2006HY2	3	11	17	27	41	46	54	68	64	43	29	24	13	11	10	7	4	3	22	496
2007HY1	3	14	16	30	37	49	66	75	71	48	32	26	15	12	11	8	4	3	24	545
2007HY2	3	12	21	31	40	46	62	70	66	45	30	24	14	11	10	7	4	3	22	521
2008HY1	2	12	19	29	37	45	60	69	65	44	29	24	14	11	10	7	4	3	22	506
2008HY2	3	14	22	30	39	47	64	72	68	46	31	25	14	12	10	7	4	3	23	536
2009HY1	3	15	20	30	39	47	63	72	68	46	31	25	14	12	10	7	4	3	23	532
2009HY2	3	14	20	31	39	48	65	74	69	47	32	26	15	12	10	7	4	3	24	542

Table B.11: Example of PCE Projection - IP projection ($m)

B.3 Payments per Claim Finalised in Operational Time

The following tables apply the PPCF in operational time (PPCF(OT)) model to the NSW CTP data using the inflated payments.

Firstly, Table B.12 shows the number of claims finalised in each accident period and operational time decile. We observe for accident periods up to around 2005 more than 90% of the claims have already been finalised. All the claims that are still open from these accidents years are expected to be finalised in the last decile of operational times. Conversely, accident period second half-year of 2009 (2009HY2) has just finalised a small portion of claims as at the censoring date.

Accident Period	Operational Time (Deciles) 10	20	30	40	50	60	70	80	90	100
2001HY1	630	629	629	629	629	629	629	629	629	535
2001HY2	571	570	571	570	571	570	571	570	571	492
2002HY1	562	562	562	562	561	562	562	562	561	470
2002HY2	512	512	511	512	511	512	511	512	511	403
2003HY1	503	503	503	502	503	503	502	503	503	364
2003HY2	496	495	495	495	495	495	495	495	495	291
2004HY1	509	509	508	509	508	509	508	509	508	270
2004HY2	484	484	484	484	484	484	484	484	484	223
2005HY1	494	493	493	493	493	494	493	493	493	96
2005HY2	449	449	448	449	448	449	448	449	362	
2006HY1	463	462	463	462	463	462	462	463	89	
2006HY2	448	448	448	448	447	448	448	255		
2007HY1	465	464	464	464	465	464	238			
2007HY2	453	453	453	452	453	272				
2008HY1	445	444	445	444	281					
2008HY2	443	443	443	21						
2009HY1	488	338								
2009HY2	95									

Table B.12: Example of PPCF(OT) Model - Finalised Claims in Optime

Table B.13 shows the inflation adjusted payments made in each of the operational time deciles, which correspond to the claims finalised in each cell show in the table above.

Accident Period	Operational Time (Deciles) 10	20	30	40	50	60	70	80	90	100
2001HY1	2	3	7	10	24	39	65	93	102	131
2001HY2	1	2	4	12	17	32	56	71	92	137
2002HY1	1	2	5	9	18	33	52	64	99	149
2002HY2	1	2	4	9	19	30	41	69	101	117
2003HY1	2	2	6	11	22	31	44	63	103	99
2003HY2	1	2	5	10	20	44	53	71	126	82
2004HY1	2	2	6	15	20	43	48	68	120	110
2004HY2	1	2	4	12	20	38	55	73	98	69
2005HY1	1	3	4	12	18	38	45	65	104	33
2005HY2	1	3	5	13	21	36	60	71	84	
2006HY1	1	2	6	12	24	41	59	83	14	
2006HY2	2	2	5	12	24	31	52	43		
2007HY1	1	3	6	17	23	37	25			
2007HY2	2	4	8	16	27	18				
2008HY1	1	3	6	13	13					
2008HY2	1	5	10	1						
2009HY1	2	3								
2009HY2	0									

Table B.13: Example of PPCF(OT) Model - IP in Optime (\$m)

Table B.14 shows the Payment per Claim Finalised by performing a division of the claim payments made by number of claims finalised. It is clear that there is a distinct trend in the claim sizes by operational time - that they increase sharply with increases in the operational time. The claims in operational time 10 have an average size of around 2,000 to 3,000 while the claim in operational time decile 50 have an average size of around 50,000. Of course, the key is then to select a set of PPCF assumptions for the future. Generally, a weighted average size of the latest two years have been used and the selected value are shown in red.

Accident Period	Operational Time (Deciles) 10	20	30	40	50	60	70	80	90	100
2001HY1	4	6	11	16	38	61	104	148	162	246
2001HY2	2	4	7	21	30	57	99	124	161	278
2002HY1	2	4	9	16	31	58	93	114	176	316
2002HY2	2	5	8	18	38	59	81	135	198	291
2003HY1	4	4	11	22	44	62	88	126	205	271
2003HY2	2	4	10	20	41	88	108	143	255	283
2004HY1	3	5	11	30	39	85	95	134	237	407
2004HY2	2	5	9	25	42	78	113	152	202	310
2005HY1	3	6	9	24	37	77	92	132	211	347
2005HY2	2	6	11	30	47	80	134	157	233	
2006HY1	2	5	13	26	52	89	127	180	158	
2006HY2	3	5	11	28	54	70	116	167		
2007HY1	3	7	12	37	49	80	106			
2007HY2	3	8	18	35	61	65				
2008HY1	2	7	14	29	48					
2008HY2	2	11	24	37						
2009HY1	3	9								
2009HY2	3									
Selected	**2.6**	**8.8**	**17.0**	**35.0**	**53.0**	**76.0**	**120.0**	**160.0**	**220.0**	**310.0**

Table B.14: Example of PPCF(OT) Model - PPCF in Optime ($000)

The projection is rather easy from this point on. Table B.15 shows the number of open claims from each accident period being notionally allocated to an operational time. Table B.16 then shows the expected payments to be made for each of these operational time deciles that still contains open claims, by multiplying the number of open claims at each operational time decile by the selected claim size. Summing them along with payments already made provides the ultimate cost of the claims for each accident period. The last column of the Table B.16 shows the expected inflated cost of these payments; that is, inflating these payments from the modelling date of December 2009 to the time they are expected to be paid using 4.3% p.a..

Accident Period	Operational Time (Deciles) 10	20	30	40	50	60	70	80	90	100
2001HY1										87
2001HY2										78
2002HY1										92
2002HY2										108
2003HY1										138
2003HY2										204
2004HY1										238
2004HY2										261
2005HY1										397
2005HY2									86	448
2006HY1									373	462
2006HY2								194	447	447
2007HY1							226	464	464	464
2007HY2						183	452	452	452	452
2008HY1					166	444	444	444	444	444
2008HY2				421	443	443	443	443	443	443
2009HY1			630	488	488	488	488	488	488	488
2009HY2		469	469	469	469	469	469	469	469	469

Table B.15: Example of PPCF(OT) Model - Open Claims in Optime

Accident	Operational Time (Deciles)										Total OS	
Period	10	20	30	40	50	60	70	80	90	100	$Dec09	Inflated
2001HY1										27	27	30
2001HY2										24	24	27
2002HY1										28	28	32
2002HY2										34	34	38
2003HY1										43	43	48
2003HY2										63	63	74
2004HY1										74	74	86
2004HY2										81	81	95
2005HY1										123	123	144
2005HY2									19	139	158	185
2006HY1									82	143	225	255
2006HY2								31	98	139	268	302
2007HY1							27	74	102	144	347	392
2007HY2						14	54	72	99	140	380	432
2008HY1					9	34	53	71	98	138	402	463
2008HY2				15	23	34	53	71	97	137	431	501
2009HY1			11	17	26	37	59	78	107	151	486	575
2009HY2		4	8	16	25	36	56	75	103	145	469	565

Table B.16: Example of PPCF(OT) Model - OS Claims in Optime ($m)

Appendix C

NSW CTP Claims Database - PIR

C.1 Information Contained in the PIR Database

The quarterly PIR extracts contains the following information. The PIR Coding Manual (Motor Accidents Authority [2008]) provides more detailed descriptions for each field.

C.2 Table - Accident Details

Managing Insurer The insurer that has been allocated to management all claims for the accident. If a vehicle is at fault then that vehicle's CTP insurer is allocated to be the managing insurer.

Accident Number A unique identifier that has been assigned to the accident by the MAA. Only accidents with CTP claims are assigned an accident number

Accident Date The date of the accident.

Postcode The postcode corresponding to the suburb in which the accident occurred.

C.3 Table - Vehicle Details

Accident Number Used to join the Accident Details table

Insurance Rating Category This is to identify the vehicle class (motor car, goods moving vehicle, bus, etc) and the region (metro, country, etc)

Garage Postcode The postcode in which the vehicle is stored

Type of Ownership This can be either Private, Business or Government

C.4 Table - Claims Record

Observation Number Each claimant is assigned with a unique ID

Accident number Used to join the Accident Details table

Date of Birth Date of birth of this claimant

Gender Gender of this claimant

Employment Status Whether the claimant was employed, seeking work, no employed due to various reasons at the time of the accident.

Occupation Some broad division of employment into Managers, Professionals, Clerks, etc. under the Australian Standard Classifications of Occupations

Economic Loss Whether the claim includes a component for economic loss

Weekly Earning The weekly earnings of the claimant at the time of the accident. This field is subject to significant levels of missing values (for those that is employed or self employed)

Prior Injury Whether the claimant has made a prior injury claim (CTP or Workers Compensation)

Claim Status Whether the claim is still active or has finalised

Date of ANF Date the claimant submitted an ANF

Date of Claim Date the claimant lodged a full claim

Date Finalised Date the ANF or claim has been finalised

Rehab Indicator An indicator variable denoting the likelihood the claimant requires rehab

Litigation Whether court proceedings has commenced for this claim

Litigation Level The level of the court system the claim is at (Local, District or Supreme).

Fatality Whether the claimant has died of the injuries sustained.

Date of Death The date the claimant died

Liability Status Whether the insurer has accepted the liability of the claim.

Legal Representation Whether the claimant has appoint a lawyer to act on his/her behalf.

Shared Claim Whether the claim is shared between numerous insurers as the at-fault party cannot be determined or the fault is shared between vehicles

Number of Vehicle Shared The number of vehicles sharing liability of the accident

Claim Disposal Which avenue did the claim settle in (CARS, Court, Out of Court, etc)

Date of Referral to CARS The date the matter passed onto CARS.

CARS Outcome The decision of CARS, whether the claimant accepts the award made by CARS, rejected or is progressing further

Date of CARS Outcome The date CARS made an assessment on the matter.

Impairment Threshold Whether the claimant has exceeded the 10% Whole Body Impairment threshold in order for general damages to be awarded.

Date Settled The date that an agreement has been reached on the claim

C.5 Table - Injury Coding

Observation Number Used to join injuries to claimants

AIS Type Denote which Abbreviated Injury Scale system the injury code comes from, either 85 or 05. In 2008, MAA adopted the AIS 2005 system; previously the AIS 1985 has been used. For finalised claims, the AIS 85 codes are not revised. For active claims, the injury codes will be revised once an update is made to the claim.

Injury Code The 6 or 7 digit AIS code denote this particular injury. The first 5 or 6 digits denote the location and nature of the injury while the last digit typically denote the severity of the injury.

This table has a many to one relationship to the claimant table, that is, one claimant may sustain multiple injuries

C.6 Table - Quarterly Payment Summary

Observation Number Used to join payments to claimants.

Gross Payment The GST inclusive payment made in the quarter of the data extract.

Amount Outstanding The insurer's estimate of the payments to be made regarding this claim.

GST Input Tax Credit Whether the claimant can claim the GST from the government, precluding the insurance company doing so.

C.7 Table - Detailed Payment Information

Observation Number Used to join payments to claimants.

Date of Payment The data the payment was made

Payment Type The nature of the payment made, whether it is for treatment or legal expense, etc.

Gross Payment The GST inclusive amount of the payment made.

Appendix D

Modelling Results of the Reduced Models for Projection

Delay Model with Censoring					
Model		Model with all Covariates		Reduced Model for Projection	
Number of Claims		80,998		80,998	
Number of Revisions		189,507		189,507	
Number of Parameters		75		61	
Log-likelihood		-275,623		-275,817	
AIC		551,393		551,756	
BIC		552,050		552,306	
Parameter	Type	coeff	s.e.	coeff	s.e.
Intercept		**6.78**	0.03	**6.60**	0.03
phi_1		**0.24**	0.00	**0.24**	0.00
alpha		**3.76**	0.04	**3.74**	0.04
Previous Direction	F	**0.16**	0.01	**0.18**	0.01
Previous Size	F	**0.07**	0.00	**0.07**	0.00
Previous Incurred Cost (log)	De	**0.05**	0.00	**0.06**	0.00
Revision 1	De				
Revision 2	De	**0.40**	0.01	**0.41**	0.01
Revision 3	De	**0.44**	0.01	**0.44**	0.01
Revision 4	De	**0.32**	0.01	**0.32**	0.01
Revision 5	De	**0.14**	0.02	**0.14**	0.02
Revision 6 +	De	**-0.17**	0.02	**-0.18**	0.02
Revision Year = 2001	De	**-8.89**	0.04	**-8.87**	0.04
Revision Year = 2002	De	**-7.21**	0.02	**-7.19**	0.02
Revision Year = 2003	De	**-5.94**	0.02	**-5.91**	0.02
Revision Year = 2004	De	**-4.81**	0.02	**-4.79**	0.02
Revision Year = 2005	De	**-3.84**	0.02	**-3.82**	0.02
Revision Year = 2006	De	**-2.89**	0.01	**-2.87**	0.01
Revision Year = 2007	De	**-2.02**	0.01	**-2.00**	0.01
Revision Year = 2008	De	**-1.06**	0.01	**-1.05**	0.01
Revision Year = 2009	De				
Reporting Delay (log)	S	**-0.03**	0.01	**-0.03**	0.01
Insurer A	S				
Insurer B	S	**0.14**	0.01	**0.14**	0.01
Insurer C	S	**0.12**	0.01	**0.12**	0.01
Insurer D	S	**0.20**	0.01	**0.23**	0.01
Insurer E	S	**0.08**	0.01	**0.09**	0.01
Insurer F	S	**0.06**	0.01	**0.09**	0.01
Insurer Other	S	**0.20**	0.01	**0.20**	0.01
Employed	S				
Self Employed	S				
Unemployed	S				
Other	S			**0.02**	0.01
Age between 0 - 9	S	**0.08**	0.02	**0.07**	0.02
Age between 10 - 16	S	**0.04**	0.01	**0.04**	0.02
Age between 17 - 25	S	**0.02**	0.01	**0.02**	0.01
Age between 26 - 45	S				
Age between 46 - 65	S	0.00	0.01	0.00	0.01
Age 66 +	S	-0.01	0.01	-0.02	0.01
Accident Year = 2001	S				
Accident Year = 2002	S	**-1.03**	0.01	**-1.03**	0.01
Accident Year = 2003	S	**-1.97**	0.01	**-1.97**	0.01
Accident Year = 2004	S	**-2.89**	0.01	**-2.89**	0.01
Accident Year = 2005	S	**-3.83**	0.02	**-3.83**	0.02
Accident Year = 2006	S	**-4.81**	0.02	**-4.80**	0.02
Accident Year = 2007	S	**-5.91**	0.02	**-5.91**	0.02
Accident Year = 2008	S	**-7.43**	0.02	**-7.44**	0.02
Accident Year = 2009	S	**-9.32**	0.06	**-9.33**	0.06
Legal Rep at report	S				
Max. Severity at report	S	**0.03**	0.00	**0.03**	0.00

*italicised variables represents the baseline value of a categorical variable

Delay Model with Censoring (cont.)					
Model		Model with all Covariates		Reduced Model for Projection	
Number of Claims		80,998		80,998	
Number of Revisions		189,507		189,507	
Number of Parameters		75		61	
Log-likelihood		-275,623		-275,817	
AIC		551,393		551,756	
BIC		552,050		552,306	
Parameter	Type	coeff	s.e.	coeff	s.e.
ISS at j - 1 = 0	D	**-0.59**	0.07		
ISS at j - 1 = 1	D				
ISS at j - 1 = 2	D	-0.01	0.01		
ISS at j - 1 between 3 and 5	D	**-0.03**	0.01		
ISS at j - 1 between 6 and 10	D	**-0.06**	0.02		
ISS at j - 1 between 11 and 30	D	**-0.10**	0.03		
ISS at j - 1 between 31 and 75	D	-0.07	0.05		
Maximum Severity at j - 1 is unknown	D	**0.61**	0.07	**0.03**	0.01
Maximum Severity at j - 1 = 1	D				
Maximum Severity at j - 1 = 2	D	0.00	0.01	**-0.03**	0.01
Maximum Severity at j - 1 = 3	D	-0.02	0.02	**-0.09**	0.01
Maximum Severity at j - 1 = 4	D	-0.02	0.03	**-0.12**	0.02
Maximum Severity at j - 1 = 5	D	**-0.11**	0.04	**-0.19**	0.03
Maximum Severity at j - 1 = 6	D	**-0.15**	0.05	**-0.21**	0.03
Body regions injured at j - 1 is unknown	D	**0.02**	0.01		
Body regions injured at j - 1 = 1	D				
Body regions injured at j - 1 = 2	D	0.01	0.01		
Body regions injured at j - 1 = 3	D	-0.01	0.01		
Body regions injured at j - 1 = 4	D	**-0.07**	0.02		
Body regions injured at j - 1 = 5	D	0.01	0.05		
WPI at j - 1 is 10% or less	D				
WPI at j - 1 is more than 10%	D	**-0.25**	0.01	**-0.23**	0.01
WPI at j - 1 is unknown	D	**-0.06**	0.01	**-0.02**	0.01
Claim Size has been 5 times initial est.	D	**-0.09**	0.01	**-0.09**	0.01
Payment at j - 1 exceed 70% of Incurred	D	**-0.15**	0.01		
Claim duation at j - 1 (log)	D	**-1.30**	0.01	**-1.31**	0.01
Spine Injury at j -1	D	**0.21**	0.06		
Back injury at j - 1	D	**0.02**	0.01		
Legal Rep at j - 1	D	**0.14**	0.01	**0.15**	0.01
No Litigation at j - 1	D				
NSW Supreme Court at j - 1	D	**-0.62**	0.05	**-0.62**	0.05
NSW District Court at j - 1	D	**-0.79**	0.02	**-0.80**	0.02
NSW Local Court at j - 1	D	**-0.42**	0.05	**-0.43**	0.05
CARS at j - 1	D	**-0.53**	0.01	**-0.53**	0.01
Accepted Liability at j - 1	D				
Partially Accepted Liability at j - 1	D	**0.05**	0.01	**0.05**	0.01
Rejected Liabiliy at j - 1	D	**0.18**	0.01	**0.18**	0.01
Other at j - 1	D	**0.10**	0.01	**0.10**	0.01
Rehab Needed at j - 1	D				
Rehab Probably at j - 1	D	**0.04**	0.01		
Rehab Possible at j - 1	D				
Rehab not required at j - 1	D	**-0.02**	0.01		
Other at j - 1	D	**-0.11**	0.01		

Table D.1: Coefficients - Reduced Delay Component Model

Settlement Model					
Model		Model with all Covariates		Reduced Model for Projection	
Number of claims		80,998		80,998	
Number of Revisions		189,507		189,507	
Number of Parameters		83		65	
Log-likelihood		-49,098		-49,562	
AIC		98,362		99,255	
BIC		99,015		99,766	
Parameter	Type	coeff	s.e.	coeff	s.e.
Intercept		**2.94**	0.15	**3.22**	0.14
phi(1)					
Current Delay (log)	F	**-0.45**	0.02	**-0.44**	0.02
Previous Size	F	**-0.14**	0.01	**-0.10**	0.01
Previous Direction	F	**-0.26**	0.02	**-0.38**	0.02
Previous Delay (log)	F	**-0.06**	0.02	**-0.08**	0.02
Previous Incurred Cost (log)	De	**-0.10**	0.01	**-0.13**	0.01
Revision 1	De				
Revision 2	De	**-0.66**	0.03	**-0.72**	0.03
Revision 3	De	**-0.91**	0.04	**-0.97**	0.04
Revision 4	De	**-0.98**	0.04	**-1.04**	0.04
Revision 5	De	**-0.98**	0.05	**-1.03**	0.05
Revision 6 +	De	**-0.90**	0.06	**-0.94**	0.05
Revision Year = 2001	De	**-0.58**	0.13	**-0.64**	0.13
Revision Year = 2002	De	-0.12	0.10	-0.16	0.10
Revision Year = 2003	De	**-0.20**	0.08	**-0.24**	0.08
Revision Year = 2004	De	**-0.21**	0.07	**-0.26**	0.07
Revision Year = 2005	De	-0.06	0.06	-0.10	0.06
Revision Year = 2006	De	-0.06	0.05	-0.09	0.05
Revision Year = 2007	De	0.06	0.04	0.03	0.04
Revision Year = 2008	De	**-0.10**	0.03	**-0.08**	0.03
Revision Year = 2009	De				
Reporting Delay (log)	S	**-0.20**	0.02	**-0.20**	0.02
Insurer A	S				
Insurer B	S	**-0.48**	0.02	**-0.46**	0.02
Insurer C	S	**-0.17**	0.04	**-0.21**	0.04
Insurer D	S	**-0.32**	0.03	**-0.27**	0.03
Insurer E	S	**0.29**	0.04	**0.34**	0.04
Insurer F	S	**-1.09**	0.03	**-1.03**	0.03
Insurer Other	S	**-0.50**	0.03	**-0.47**	0.03
Gender = Female	S	**-0.04**	0.02		
Employed	S				
Self Employed	S	**-0.07**	0.03	**-0.07**	0.03
Unemployed	S			**-0.18**	0.07
Other	S	**0.13**	0.02	**0.04**	0.02
Age between 0 - 9	S	-0.02	0.05	-0.04	0.05
Age between 10 - 16	S	-0.02	0.05	-0.03	0.05
Age between 17 - 25	S	**0.15**	0.02	**0.16**	0.02
Age between 26 - 45	S				
Age between 46 - 65	S	0.03	0.02	0.03	0.02
Age 66 +	S	**0.40**	0.03	**0.38**	0.03
Resides in Metro area	S				
Resides in Outer Metro area	S				
Resides in Country area	S				
Resides in Wollongong area	S				
Resides in Newcastle area	S	**0.06**	0.02	**0.07**	0.02
Accident Year = 2001	S				
Accident Year = 2002	S	0.06	0.03	**0.07**	0.03
Accident Year = 2003	S	**0.24**	0.04	**0.25**	0.04
Accident Year = 2004	S	**0.33**	0.05	**0.35**	0.05
Accident Year = 2005	S	**0.39**	0.06	**0.42**	0.06
Accident Year = 2006	S	**0.45**	0.07	**0.47**	0.07
Accident Year = 2007	S	**0.40**	0.08	**0.40**	0.08
Accident Year = 2008	S	**0.40**	0.10	**0.39**	0.10
Accident Year = 2009	S	**0.41**	0.15	**0.40**	0.15
Legal Rep at report	S	**-0.09**	0.03	**-0.07**	0.03
Max. Severity at report	S	**-0.09**	0.01	**-0.10**	0.01
ISS at j = 0	D	**-0.49**	0.23		
ISS at j = 1	D				
ISS at j = 2	D	**-0.17**	0.03		
ISS at j between 3 and 5	D	**-0.27**	0.04		
ISS at j between 6 and 10	D	**-0.40**	0.06		
ISS at j between 11 and 30	D	**-0.36**	0.08		
ISS at j between 31 and 75	D	-0.24	0.12		
Maximum Severity at j is unknown	D	0.19	0.23	**-0.29**	0.04
Maximum Severity at j = 1	D				
Maximum Severity at j = 2	D	0.05	0.04	**-0.07**	0.02
Maximum Severity at j = 3	D	**0.21**	0.06	0.02	0.04
Maximum Severity at j = 4	D	**0.28**	0.09	0.08	0.06
Maximum Severity at j = 5	D	0.23	0.12	0.07	0.09
Maximum Severity at j = 6	D	**0.98**	0.16	**0.87**	0.10
Body regions injured at j is unknown	D	**-0.12**	0.03		
Body regions injured at j = 1	D				
Body regions injured at j = 2	D	-0.01	0.03		
Body regions injured at j = 3	D	**0.10**	0.03		
Body regions injured at j = 4	D	**0.23**	0.05		
Body regions injured at j = 5	D	0.18	0.11		
WPI at j is 10% or less	D				
WPI at j is more than 10%	D	**-1.82**	0.03	**-1.77**	0.03
WPI at j is unknown	D	**-6.29**	0.04	**-6.30**	0.04
Claim Size has been 5 times initial est.	D	**0.25**	0.03	**0.23**	0.03
Payment at j exceed 70% of Incurred	D	**0.67**	0.04		
Litigation Level Increased	D	**-0.23**	0.06	**-0.33**	0.05
Legal Rep Appointed	D	**-0.46**	0.04	**-0.46**	0.04
WPI threshold met	D	**0.61**	0.03	**0.55**	0.03
Whiplash injury developed	D	**0.24**	0.03		
ISS increased	D	**0.23**	0.03		
Liability Became Rejected	D	**0.18**	0.05	**0.20**	0.05
Maximum Severity Increased	D			**0.22**	0.04
Number of body region injured Increased	D	**0.35**	0.03		
Claim duation at j (log)	D	**0.62**	0.04	**0.66**	0.04
Brain Injury at j	D	**-0.16**	0.05		
Back injury at j	D	**-0.10**	0.02		
Legal Rep at j	D	**-0.51**	0.04	**-0.49**	0.04
Economic Loss compensation at j	D	**0.12**	0.02		
No Litigation at j	D				
NSW Supreme Court at j	D			**0.24**	0.10
NSW District Court at j	D	**-0.11**	0.04		
NSW Local Court at j	D	**0.32**	0.09	**0.41**	0.09
CARS at j	D				
Accepted Liability at j	D				
Partially Accepted Liability at j	D				
Rejected Liabiliy at j	D	**0.21**	0.03	**0.17**	0.03
Other at j	D	**-0.18**	0.04	**-0.20**	0.04
Rehab Needed at j	D				
Rehab Probably at j	D				
Rehab Possible at j	D				
Rehab not required at j	D	**-0.17**	0.02		
Other at j	D				

*italicised variables represents the baseline value of a categorical variable

Table D.2: Coefficients - Reduced Settlement Component Model

Direction Model					
Model		Model with all Covariates		Reduced Model for Projection	
Number of Claims		80,998		80,998	
Number of Revisions		189,507		189,507	
Number of Parameters		87		63	
Log-likelihood		-85,344		-88,064	
AIC		170,859		176,253	
BIC		171,535		176,740	
Parameter	Type	coeff	s.e.	coeff	s.e.
Intercept		**1.20**	0.11	**2.13**	0.11
phi(1)		**0.10**	0.01	**0.05**	0.01
Current Settlement	F	**-2.63**	0.02	**-2.38**	0.02
Current Delay (log)	F	**0.29**	0.01	**0.25**	0.01
Previous Size	F	**-0.03**	0.01	**0.06**	0.01
Previous Delay (log)	F	**0.05**	0.01		
Previous Incurred Cost (log)	De	**-0.30**	0.01	**-0.33**	0.01
Revision 1	De				
Revision 2	De	**0.18**	0.03	**0.37**	0.03
Revision 3	De	**0.41**	0.03	**0.66**	0.03
Revision 4	De	**0.51**	0.04	**0.78**	0.04
Revision 5	De	**0.57**	0.04	**0.87**	0.04
Revision 6 +	De	**0.57**	0.05	**0.93**	0.04
Revision Year = 2001	De	**0.48**	0.10	0.07	0.09
Revision Year = 2002	De	**0.22**	0.08	**-0.17**	0.07
Revision Year = 2003	De	**0.22**	0.07	**-0.14**	0.06
Revision Year = 2004	De	**0.28**	0.06	-0.03	0.05
Revision Year = 2005	De	**0.26**	0.05	0.02	0.04
Revision Year = 2006	De	**0.23**	0.04	0.02	0.04
Revision Year = 2007	De	0.02	0.03	**-0.13**	0.03
Revision Year = 2008	De	-0.03	0.03	**-0.08**	0.03
Revision Year = 2009	De				
Reporting Delay (log)	S	**-0.05**	0.02		
Insurer A	S				
Insurer B	S	**0.24**	0.02	**0.15**	0.02
Insurer C	S	**0.36**	0.03	**0.41**	0.03
Insurer D	S	**0.28**	0.04	**0.22**	0.03
Insurer E	S	**-0.50**	0.03	**-0.53**	0.02
Insurer F	S	**-0.12**	0.03	**-0.08**	0.02
Insurer Other	S	**0.27**	0.03	**0.16**	0.02
Employed	S				
Self Employed	S	**0.11**	0.03	**0.12**	0.02
Unemployed	S				
Other	S			**-0.17**	0.02
Age between 0 - 9	S	**-0.40**	0.04	**-0.48**	0.04
Age between 10 - 16	S	**-0.31**	0.04	**-0.34**	0.03
Age between 17 - 25	S	**-0.15**	0.02	**-0.14**	0.02
Age between 26 - 45	S				
Age between 46 - 65	S	**-0.05**	0.02	**-0.04**	0.02
Age 66 +	S	**-0.12**	0.03	**-0.11**	0.03
Resides in Metro area	S				
Resides in Outer Metro area	S	**-0.11**	0.02	**-0.12**	0.02
Resides in Country area	S				
Resides in Wollongong area	S				
Resides in Newcastle area	S	**-0.07**	0.02	**-0.08**	0.02
Accident Year = 2001	S				
Accident Year = 2002	S	0.04	0.03	0.00	0.02
Accident Year = 2003	S	**0.09**	0.03	0.03	0.03
Accident Year = 2004	S	**0.22**	0.04	**0.15**	0.04
Accident Year = 2005	S	**0.41**	0.05	**0.27**	0.04
Accident Year = 2006	S	**0.55**	0.06	**0.35**	0.05
Accident Year = 2007	S	**0.71**	0.07	**0.46**	0.06
Accident Year = 2008	S	**0.99**	0.08	**0.62**	0.07
Accident Year = 2009	S	**1.66**	0.12	**1.21**	0.10
Max. Severity at report	S	**0.02**	0.01	**0.05**	0.01

*italicised variables represents the baseline value of a categorical variable

Direction Model (cont.)					
Model		Model with all Covariates		Reduced Model for Projection	
Number of Claims		80,998		80,998	
Number of Revisions		189,507		189,507	
Number of Parameters		87		63	
Log-likelihood		-85,344		-88,064	
AIC		170,859		176,253	
BIC		171,535		176,740	
Parameter	Type	coeff	s.e.	coeff	s.e.
ISS at j = 0	D	-0.04	0.14		
ISS at j = 1	D				
ISS at j = 2	D	**0.10**	0.03		
ISS at j between 3 and 5	D	**0.32**	0.03		
ISS at j between 6 and 10	D	**0.47**	0.05		
ISS at j between 11 and 30	D	**0.48**	0.06		
ISS at j between 31 and 75	D	**0.43**	0.10		
Maximum Severity at j is unknown	D	0.12	0.14	0.02	0.03
Maximum Severity at j = 1	D				
Maximum Severity at j = 2	D	-0.05	0.03	**0.20**	0.02
Maximum Severity at j = 3	D	-0.07	0.05	**0.30**	0.03
Maximum Severity at j = 4	D	-0.01	0.07	**0.40**	0.05
Maximum Severity at j = 5	D	**-0.28**	0.10	**0.19**	0.07
Maximum Severity at j = 6	D	-0.20	0.13	0.09	0.08
Body regions injured at j is unknown	D	**0.10**	0.03		
Body regions injured at j = 1	D				
Body regions injured at j = 2	D	**-0.07**	0.02		
Body regions injured at j = 3	D	**-0.09**	0.03		
Body regions injured at j = 4	D	**-0.13**	0.04		
Body regions injured at j = 5	D	0.02	0.09		
WPI at j is 10% or less	D				
WPI at j is more than 10%	D	**1.14**	0.02	**1.29**	0.03
WPI at j is unknown	D	**1.20**	0.02	**1.17**	0.02
Claim Size has been 5 times initial est.	D	**-0.14**	0.02	**-0.22**	0.02
Payment at j exceed 70% of Incurred	D	**2.04**	0.03		
Brain Injury Developed	D	**0.29**	0.08		
Litigation Level Increased	D	**0.41**	0.05	**0.25**	0.04
Legal Rep Appointed	D	**0.21**	0.03	**0.20**	0.03
WPI threshold met	D			**-0.26**	0.04
Liability Became Rejected	D	**0.15**	0.04	**0.14**	0.04
Maximum Severity Increased	D	**0.08**	0.03		
Liability Became Accepted	D	**-0.43**	0.08	**-0.49**	0.08
Rehab Needs Increased	D	**-0.36**	0.07		
Claim duation at j (log)	D	**0.14**	0.03	**0.11**	0.02
Back injury at j	D	**0.10**	0.02		
Legal Rep at j	D	**0.78**	0.02	**0.82**	0.02
Economic Loss compensation at j	D	**0.24**	0.01		
No Litigation at j	D				
NSW Supreme Court at j	D	**-0.24**	0.09		
NSW District Court at j	D	**-0.13**	0.04		
NSW Local Court at j	D	**-0.27**	0.08		
CARS at j	D	**0.35**	0.02	**0.37**	0.02
Accepted Liability at j	D				
Partially Accepted Liability at j	D	**-0.11**	0.03	**-0.12**	0.03
Rejected Liabiliy at j	D	**-0.40**	0.03	**-0.42**	0.03
Other at j	D	**0.28**	0.03	**0.27**	0.03
Rehab Needed at j	D				
Rehab Probably at j	D	**0.06**	0.03		
Rehab Possible at j	D	**0.15**	0.02		
Rehab not required at j	D	**0.30**	0.02		
Other at j	D	**-0.19**	0.03		

Table D.3: Coefficients - Reduced Direction Component Model

Size Model					
Model		Model with all Covariates		Reduced Model for Projection	
Number of Claims		80,998		80,998	
Number of Revisions		189,507		189,507	
Number of Parameters		92		74	
Log-likelihood		-112,345		-112,663	
AIC		224,870		225,469	
BIC		225,578		226,035	
Parameter	Type	coeff	s.e.	coeff	s.e.
Intercept		**1.31**	0.05	**1.14**	0.04
phi(1)		**0.08**	0.00	**0.08**	0.00
c		**0.86**	0.00	**0.86**	0.00
Saving on Finalisations	F	**1.22**	0.02	**1.24**	0.02
Current Direction	F	**0.30**	0.01	**0.30**	0.01
Current Settlement	F	**-0.69**	0.01	**-0.70**	0.01
Current Delay (log)	F	**0.15**	0.01	**0.16**	0.01
Previous Direction	F	**-0.08**	0.01	**-0.07**	0.01
Previous Delay (log)	F	**0.07**	0.01	**0.07**	0.01
Previous Incurred Cost (log)	De	**-0.15**	0.00	**-0.13**	0.00
Revision 1	De				
Revision 2	De	**-0.23**	0.01	**-0.23**	0.01
Revision 3	De	**-0.24**	0.01	**-0.24**	0.01
Revision 4	De	**-0.22**	0.01	**-0.22**	0.01
Revision 5	De	**-0.20**	0.02	**-0.20**	0.02
Revision 6 +	De	**-0.22**	0.02	**-0.23**	0.02
Revision Year = 2001	De	-0.04	0.04	-0.03	0.04
Revision Year = 2002	De	0.04	0.03	0.06	0.03
Revision Year = 2003	De	**0.15**	0.03	**0.16**	0.03
Revision Year = 2004	De	**0.16**	0.02	**0.17**	0.02
Revision Year = 2005	De	**0.16**	0.02	**0.17**	0.02
Revision Year = 2006	De	**0.16**	0.02	**0.17**	0.02
Revision Year = 2007	De	**0.12**	0.01	**0.13**	0.01
Revision Year = 2008	De	**0.17**	0.01	**0.18**	0.01
Revision Year = 2009	De				
Reporting Delay (log)	S	**0.05**	0.01	**0.05**	0.01
Insurer A	S				
Insurer B	S	**0.02**	0.01	-0.01	0.01
Insurer C	S	**-0.04**	0.01	-0.01	0.01
Insurer D	S	**0.11**	0.01	**0.09**	0.01
Insurer E	S	0.01	0.01	-0.00	0.01
Insurer F	S	**0.05**	0.01	**0.05**	0.01
Insurer Other	S	**0.04**	0.01	0.00	0.01
Gender = Female	S	**-0.05**	0.01	**-0.05**	0.01
Employed	S				
Self Employed	S	**-0.02**	0.01		
Unemployed	S	**0.08**	0.02	**0.09**	0.02
Other	S			**0.01**	0.01
Age between 0 - 9	S	**-0.05**	0.02	-0.02	0.02
Age between 10 - 16	S	**-0.05**	0.01	**-0.03**	0.01
Age between 17 - 25	S	**-0.03**	0.01	**-0.02**	0.01
Age between 26 - 45	S				
Age between 46 - 65	S	**-0.03**	0.01	**-0.03**	0.01
Age 66 +	S	**-0.04**	0.01	**-0.03**	0.01
Resides in Metro area	S				
Resides in Outer Metro area	S	**0.03**	0.01	**0.03**	0.01
Resides in Country area	S				
Resides in Wollongong area	S				
Resides in Newcastle area	S			**0.01**	0.01
Accident Year = 2001	S				
Accident Year = 2002	S	**-0.02**	0.01	**-0.02**	0.01
Accident Year = 2003	S	0.00	0.01	0.01	0.01
Accident Year = 2004	S	0.01	0.02	0.02	0.02
Accident Year = 2005	S	0.03	0.02	0.04	0.02
Accident Year = 2006	S	**0.07**	0.02	**0.07**	0.02
Accident Year = 2007	S	**0.08**	0.03	**0.10**	0.03
Accident Year = 2008	S	**0.30**	0.03	**0.33**	0.03
Accident Year = 2009	S	**0.24**	0.04	**0.26**	0.04
Legal Rep at report	S	**-0.04**	0.01	**-0.04**	0.01
Max. Severity at report	S	**0.02**	0.01	**0.03**	0.01

*italicised variables represents the baseline value of a categorical variable

Size Model					
Model		Model with all Covariates		Reduced Model for Projection	
Number of Claims		80,998		80,998	
Number of Revisions		189,507		189,507	
Number of Parameters		92		74	
Log-likelihood		-112,345		-112,663	
AIC		224,870		225,469	
BIC		225,578		226,035	
Parameter	Type	coeff	s.e.	coeff	s.e.
ISS at j = 0	D	0.02	0.06		
ISS at j = 1	D				
ISS at j = 2	D	**-0.02**	0.01		
ISS at j between 3 and 5	D	-0.01	0.01		
ISS at j between 6 and 10	D	**-0.06**	0.02		
ISS at j between 11 and 30	D	**-0.09**	0.03		
ISS at j between 31 and 75	D	**-0.17**	0.04		
Maximum Severity at j is unknown	D	0.07	0.06	**0.08**	0.01
Maximum Severity at j = 1	D				
Maximum Severity at j = 2	D	**0.05**	0.01	**0.03**	0.01
Maximum Severity at j = 3	D	**0.06**	0.02	0.01	0.01
Maximum Severity at j = 4	D	0.02	0.03	0.03	0.02
Maximum Severity at j = 5	D	0.07	0.05	0.05	0.03
Maximum Severity at j = 6	D	0.04	0.05	**-0.19**	0.03
Body regions injured at j is unknown	D	**-0.03**	0.01		
Body regions injured at j = 1	D				
Body regions injured at j = 2	D	**0.03**	0.01		
Body regions injured at j = 3	D	0.02	0.01		
Body regions injured at j = 4	D	0.00	0.02		
Body regions injured at j = 5	D	0.05	0.04		
WPI at j is 10% or less	D				
WPI at j is more than 10%	D	**-0.68**	0.01	**-0.68**	0.01
WPI at j is unknown	D	**-0.65**	0.01	**-0.65**	0.01
Claim Size has been 5 times initial est.	D	**-0.08**	0.01	**-0.08**	0.01
Payment at j exceed 70% of Incurred	D	**-0.13**	0.01		
Litigation Level Increased	D	**0.21**	0.02	**0.21**	0.02
Legal Rep Appointed	D	**0.14**	0.01	**0.14**	0.01
WPI threshold met	D	**0.46**	0.02	**0.47**	0.02
Liability Became Rejected	D	**0.07**	0.02	**0.07**	0.02
Maximum Severity Increased	D	**0.05**	0.01	**0.08**	0.01
Liability Became Accepted	D	**0.34**	0.03	**0.36**	0.03
Rehab Needs Increased	D	**0.36**	0.03		
Number of body region injured Increased	D	**0.05**	0.01		
Claim duation at j (log)	D	**-0.04**	0.01	**-0.05**	0.01
Brain Injury at j	D	**0.16**	0.02		
Whiplash Injury at j	D	**-0.05**	0.01		
Legal Rep at j	D	**0.06**	0.01	**0.07**	0.01
No Litigation at j	D				
NSW Supreme Court at j	D	**-0.42**	0.04	**-0.43**	0.04
NSW District Court at j	D	**-0.43**	0.02	**-0.43**	0.02
NSW Local Court at j	D	**-0.30**	0.04	**-0.31**	0.04
CARS at j	D	**-0.07**	0.01	**-0.07**	0.01
Accepted Liability at j	D				
Partially Accepted Liability at j	D			**0.03**	0.01
Rejected Liabiliy at j	D	**0.58**	0.01	**0.59**	0.01
Other at j	D	**0.27**	0.01	**0.29**	0.01
Rehab Needed at j	D				
Rehab Probably at j	D	**0.06**	0.01		
Rehab Possible at j	D	**0.05**	0.01		
Rehab not required at j	D	**0.07**	0.01		
Other at j	D	**-0.07**	0.01		

Table D.4: Coefficients - Reduced Size Component Model

Appendix E

Stratified Modelling Results

E.1 Delay Component by Injury Severity

The following tables show the modelled coefficients for the all-claims model and also models stratified by injury severity for the Delay Component.

Delay Model with Censoring						
Model		All	Stratified by Injury Severity and Representation			
		Claims	Minor R	Minor NR	Moderate	Severe
Number of Claims		80,998	26,758	30,118	15,036	9,086
Number of Revisions		189,507	70,410	58,356	36,948	23,793
Number of Parameters		75	69	57	66	63
Log-likelihood		-275,623	-103,993	-75,134	-55,968	-39,077
Parameter	Type	coeff	coeff	coeff	coeff	coeff
Intercept		**6.78**	**6.75**	**6.93**	**7.05**	**6.48**
phi_1		**0.24**	**0.25**	**0.26**	**0.23**	**0.21**
alpha		**3.76**	**4.58**	**2.71**	**4.24**	**4.93**
Previous Direction	F	**0.16**	**0.08**	**0.09**	**0.15**	**0.23**
Previous Size	F	**0.07**	**0.08**	**0.08**	**0.08**	**0.06**
Previous Incurred Cost (log)	De	**0.05**	**0.06**	**0.08**	**0.04**	**0.04**
Revision 1	De					
Revision 2	De	**0.40**	**0.38**	**0.41**	**0.38**	**0.48**
Revision 3	De	**0.44**	**0.43**	**0.59**	**0.35**	**0.43**
Revision 4	De	**0.32**	**0.34**	**0.55**	**0.23**	**0.24**
Revision 5	De	**0.14**	**0.17**	**0.42**	0.04	**0.09**
Revision 6 +	De	**-0.17**	**-0.12**	0.05	**-0.25**	**-0.29**
Revision Year = 2001	De	**-8.89**	**-8.62**	**-9.51**	**-8.72**	**-7.79**
Revision Year = 2002	De	**-7.21**	**-6.96**	**-7.84**	**-6.95**	**-6.40**
Revision Year = 2003	De	**-5.94**	**-5.72**	**-6.52**	**-5.75**	**-5.35**
Revision Year = 2004	De	**-4.81**	**-4.62**	**-5.28**	**-4.65**	**-4.25**
Revision Year = 2005	De	**-3.84**	**-3.67**	**-4.19**	**-3.74**	**-3.43**
Revision Year = 2006	De	**-2.89**	**-2.75**	**-3.14**	**-2.78**	**-2.60**
Revision Year = 2007	De	**-2.02**	**-1.91**	**-2.22**	**-1.93**	**-1.77**
Revision Year = 2008	De	**-1.06**	**-0.98**	**-1.13**	**-1.02**	**-0.93**
Revision Year = 2009	De					
Reporting Delay (log)	S	**-0.03**		**-0.06**		
Insurer A	S					
Insurer B	S	**0.14**	**0.15**	**0.20**	**0.12**	0.01
Insurer C	S	**0.12**	**0.17**	**0.32**	0.03	**-0.07**
Insurer D	S	**0.20**	**0.22**	**0.21**	**0.17**	0.04
Insurer E	S	**0.08**	**0.07**	**0.08**	**0.08**	-0.00
Insurer F	S	**0.06**	**0.07**	**0.05**	**0.07**	-0.01
Insurer Other	S	**0.20**	**0.21**	**0.31**	**0.15**	0.03
Employed	S					
Self Employed	S					
Unemployed	S					
Other	S		**0.01**	**0.04**		
Age between 0 - 9	S	**0.08**	**0.09**	**0.17**	**0.10**	
Age between 10 - 16	S	**0.04**	**0.04**	**0.09**	0.05	
Age between 17 - 25	S	**0.02**	**0.02**	0.02	0.03	
Age between 26 - 45	S					
Age between 46 - 65	S	0.00	-0.00	0.01	0.01	
Age 66 +	S	-0.01	**-0.03**	-0.05	**-0.06**	
Resides in Metro area	S					
Resides in Outer Metro area	S					
Resides in Country area	S					
Resides in Wollongong area	S					
Resides in Newcastle area	S					**-0.04**
Accident Year = 2001	S					
Accident Year = 2002	S	**-1.03**	**-1.00**	**-1.16**	**-0.97**	**-0.92**
Accident Year = 2003	S	**-1.97**	**-1.91**	**-2.17**	**-1.90**	**-1.74**
Accident Year = 2004	S	**-2.89**	**-2.80**	**-3.21**	**-2.77**	**-2.56**
Accident Year = 2005	S	**-3.83**	**-3.71**	**-4.27**	**-3.68**	**-3.39**
Accident Year = 2006	S	**-4.81**	**-4.65**	**-5.40**	**-4.60**	**-4.25**
Accident Year = 2007	S	**-5.91**	**-5.69**	**-6.61**	**-5.69**	**-5.21**
Accident Year = 2008	S	**-7.43**	**-7.06**	**-8.12**	**-7.08**	**-6.54**
Accident Year = 2009	S	**-9.32**	**-8.47**	**-9.74**	**-9.08**	**-9.09**
Legal Rep at report	S		**0.06**			**-0.06**
Max. Severity at report	S	**0.03**	**0.03**	**0.18**	**-0.05**	

*italicised variables represents the baseline value of a categorical variable

Table E.1: Coefficients - Delay Component by Injury Severity 1

Delay Model with Censoring (cont.)						
Model		All	Stratified by Injury Severity and Representation			
		Claims	Minor R	Minor NR	Moderate	Severe
Number of Claims		80,998	26,758	30,118	15,036	9,086
Number of Revisions		189,507	70,410	58,356	36,948	23,793
Number of Parameters		75	69	57	66	63
Log-likelihood		-275,623	-103,993	-75,134	-55,968	-39,077
Parameter	Type	coeff	coeff	coeff	coeff	coeff
ISS at j - 1 = 0	D	**-0.59**	**-0.40**		**-0.43**	**-1.01**
ISS at j - 1 = 1	D					
ISS at j - 1 = 2	D	-0.01	-0.01		-0.06	
ISS at j - 1 between 3 and 5	D	**-0.03**	**-0.03**		**-0.12**	
ISS at j - 1 between 6 and 10	D	**-0.06**	**-0.08**		**-0.15**	
ISS at j - 1 between 11 and 30	D	**-0.10**	**-0.12**		**-0.20**	
ISS at j - 1 between 31 and 75	D	-0.07	**-0.10**		0.10	
Maximum Severity at j - 1 is unknown	D	**0.61**	**0.42**	**0.13**	**0.40**	**0.95**
Maximum Severity at j - 1 = 1	D					
Maximum Severity at j - 1 = 2	D	0.00	0.01		**-0.08**	0.02
Maximum Severity at j - 1 = 3	D	-0.02	-0.01		0.05	**-0.20**
Maximum Severity at j - 1 = 4	D	-0.02	-0.01		**-0.19**	**-0.19**
Maximum Severity at j - 1 = 5	D	**-0.11**	**-0.10**		-0.30	**-0.24**
Maximum Severity at j - 1 = 6	D	**-0.15**	**-0.11**		-0.16	**-0.63**
Body regions injured at j - 1 is unknown	D	**0.02**				0.02
Body regions injured at j - 1 = 1	D					
Body regions injured at j - 1 = 2	D	0.01				-0.01
Body regions injured at j - 1 = 3	D	-0.01				**-0.05**
Body regions injured at j - 1 = 4	D	**-0.07**				**-0.15**
Body regions injured at j - 1 = 5	D	0.01				0.27
WPI at j - 1 is 10% or less	D					
WPI at j - 1 is more than 10%	D	**-0.25**	**-0.22**	**-0.29**	**-0.23**	**-0.26**
WPI at j - 1 is unknown	D	**-0.06**	**-0.02**	**-0.09**	**-0.08**	**-0.08**
Claim Size has been 5 times initial est.	D	**-0.09**	**-0.11**	**-0.12**	**-0.09**	**-0.05**
Payment at j - 1 exceed 70% of Incurred	D	**-0.15**	**-0.06**	**-0.14**	**-0.17**	**-0.23**
Claim duation at j - 1 (log)	D	**-1.30**	**-1.30**	**-1.35**	**-1.27**	**-1.22**
Brain Injury at j -1	D					**0.08**
Spine Injury at j -1	D	**0.21**	**0.18**			**0.26**
Back injury at j - 1	D	**0.02**				
Legal Rep at j - 1	D	**0.14**	**0.07**	**-0.20**	**0.15**	**0.17**
Economic Loss compensation at j - 1	D			**0.04**		**0.03**
No Litigation at j - 1	D					
NSW Supreme Court at j - 1	D	**-0.62**	**-0.60**	**-0.33**	**-0.81**	**-0.50**
NSW District Court at j - 1	D	**-0.79**	**-0.77**	**-1.16**	**-0.70**	**-0.53**
NSW Local Court at j - 1	D	**-0.42**	**-0.42**	**-0.51**	**-0.63**	**-0.25**
CARS at j - 1	D	**-0.53**	**-0.51**	**-0.68**	**-0.48**	**-0.31**
Accepted Liability at j - 1	D					
Partially Accepted Liability at j - 1	D	**0.05**	**0.04**		**0.06**	
Rejected Liabiliy at j - 1	D	**0.18**	**0.17**	**0.14**	**0.18**	**0.12**
Other at j - 1	D	**0.10**	**0.10**	**0.13**	**0.07**	**0.06**
Rehab Needed at j - 1	D					
Rehab Probably at j - 1	D	**0.04**	**0.05**	**0.10**		
Rehab Possible at j - 1	D					
Rehab not required at j - 1	D	**-0.02**	**-0.03**	**-0.05**	**-0.04**	
Other at j - 1	D	**-0.11**	**-0.12**		**-0.15**	**-0.27**

*italicised variables represents the baseline value of a categorical variable

Table E.2: Coefficients - Delay Component by Injury Severity 2

A few interesting observations can be made. Typically, the more severe claims have fewer significant factors - the model for Severe claims is considerably simpler than the other models. Factors such as Age, Insurer and Rehabilitation Needs become less relevant for the more serious injuries. The significant variables common to all injury severities are consistent in magnitude and pattern, such as Number of Revisions and Year of Accident. In terms of the distribution, however, it seems the Severe claims' time between revisions are less volatile (a higher value α).

E.2 Settlement Component by Injury Severity

The following tables show the modelled coefficients for the all-claims model and also models stratified by injury severity for the Settlement Component.

Settlement Model						
Model		All	Stratified by Injury Severity and Representation			
		Claims	Minor R	Minor NR	Moderate	Severe
Number of claims		80,998	26,758	30,118	15,036	9,086
Number of Revisions		189,507	70,410	58,356	36,948	23,793
Number of Parameters		83	67	56	75	59
Log-likelihood		-49,098	-16,552	-13,922	-10,029	-7,399
Parameter	Type	coeff	coeff	coeff	coeff	coeff
Intercept		**2.94**	**1.37**	**4.28**	**1.57**	-0.70
phi(1)						
Current Delay (log)	F	**-0.45**	**-0.47**	**-0.32**	**-0.41**	**-0.46**
Previous Size	F	**-0.14**	**-0.11**	**-0.13**	**-0.16**	**-0.15**
Previous Direction	F	**-0.26**	**-0.16**	**-0.20**	**-0.28**	**-0.54**
Previous Delay (log)	F	**-0.06**	**-0.06**	**-0.09**		
Previous Incurred Cost (log)	De	**-0.10**	**-0.07**	**-0.17**	**-0.08**	
Revision 1	De					
Revision 2	De	**-0.66**	**-0.63**	**-0.79**	**-0.68**	**-0.54**
Revision 3	De	**-0.91**	**-0.91**	**-1.13**	**-0.83**	**-0.61**
Revision 4	De	**-0.98**	**-0.88**	**-1.39**	**-0.81**	**-0.74**
Revision 5	De	**-0.98**	**-0.97**	**-1.27**	**-0.88**	**-0.62**
Revision 6 +	De	**-0.90**	**-0.88**	**-1.15**	**-0.75**	**-0.63**
Revision Year = 2001	De	**-0.58**	-0.42	**-0.59**	0.17	-0.12
Revision Year = 2002	De	-0.12	0.27	-0.09	0.35	0.28
Revision Year = 2003	De	**-0.20**	**0.31**	-0.12	0.12	-0.02
Revision Year = 2004	De	**-0.21**	0.06	-0.17	0.08	0.15
Revision Year = 2005	De	-0.06	0.17	-0.02	0.19	0.25
Revision Year = 2006	De	-0.06	0.11	-0.09	0.17	0.24
Revision Year = 2007	De	0.06	**0.24**	-0.12	**0.28**	**0.27**
Revision Year = 2008	De	**-0.10**	-0.02	**-0.19**	0.14	-0.08
Revision Year = 2009	De					
Reporting Delay (log)	S	**-0.20**	**-0.16**	**-0.23**	**-0.13**	**-0.16**
Insurer A	S					
Insurer B	S	**-0.48**	**-0.67**	**-0.82**	**-0.39**	**0.16**
Insurer C	S	**-0.17**	**-0.52**	**-0.21**	-0.16	**0.42**
Insurer D	S	**-0.32**	**-0.43**	**-1.20**	-0.03	**0.76**
Insurer E	S	**0.29**	**0.29**	0.00	**0.35**	**0.57**
Insurer F	S	**-1.09**	**-1.40**	**-1.59**	**-0.70**	0.02
Insurer Other	S	**-0.50**	**-0.73**	**-0.81**	**-0.28**	0.11
Gender = Female	S	**-0.04**			**-0.10**	
Employed	S					
Self Employed	S	**-0.07**			**-0.18**	
Unemployed	S					
Other	S	**0.13**	**0.21**		**0.17**	**0.22**
Age between 0 - 9	S	-0.02	**0.23**	0.09	0.07	**-0.33**
Age between 10 - 16	S	-0.02	**0.17**	**0.24**	0.09	**-0.42**
Age between 17 - 25	S	**0.15**	**0.16**	**0.21**	**0.15**	0.07
Age between 26 - 45	S					
Age between 46 - 65	S	0.03	0.01	0.03	**0.13**	-0.03
Age 66 +	S	**0.40**	**0.33**	**0.26**	**0.52**	**0.66**
Resides in Metro area	S					
Resides in Outer Metro area	S		**0.12**			
Resides in Country area	S		**-0.27**			
Resides in Wollongong area	S					
Resides in Newcastle area	S	**0.06**	**0.19**			**-0.12**
Accident Year = 2001	S					
Accident Year = 2002	S	0.06	**0.11**	0.04	0.12	0.09
Accident Year = 2003	S	**0.24**	**0.24**	**0.41**	**0.31**	**0.28**
Accident Year = 2004	S	**0.33**	**0.37**	**0.42**	**0.41**	**0.56**
Accident Year = 2005	S	**0.39**	**0.50**	**0.41**	**0.61**	**0.63**
Accident Year = 2006	S	**0.45**	**0.65**	**0.48**	**0.64**	**0.86**
Accident Year = 2007	S	**0.40**	**0.75**	**0.49**	**0.65**	**0.63**
Accident Year = 2008	S	**0.40**	**0.90**	0.26	**1.02**	**0.91**
Accident Year = 2009	S	**0.41**	**0.82**	0.29	**0.93**	0.66
Legal Rep at report	S	**-0.09**			**-0.21**	
Max. Severity at report	S	**-0.09**	**-0.26**	**-0.14**		

*italicised variables represents the baseline value of a categorical variable

Table E.3: Coefficients - Settlement Component by Severity 1

Settlement Model (cont.)						
Model		All	Stratified by Injury Severity and Representation			
		Claims	Minor R	Minor NR	Moderate	Severe
Number of claims		80,998	26,758	30,118	15,036	9,086
Number of Revisions		189,507	70,410	58,356	36,948	23,793
Number of Parameters		83	67	56	75	59
Log-likelihood		-49,098	-16,552	-13,922	-10,029	-7,399
Parameter	Type	coeff	coeff	coeff	coeff	coeff
ISS at j = 0	D	**-0.49**	**-0.57**	**-0.39**		
ISS at j = 1	D					
ISS at j = 2	D	**-0.17**	**-0.11**	**-0.13**		
ISS at j between 3 and 5	D	**-0.27**	-0.08	**-0.16**		
ISS at j between 6 and 10	D	**-0.40**	**-0.24**	-0.09		
ISS at j between 11 and 30	D	**-0.36**	-0.17	-0.21		
ISS at j between 31 and 75	D	-0.24	-0.39			
Maximum Severity at j is unknown	D	0.19			-0.40	
Maximum Severity at j = 1	D					
Maximum Severity at j = 2	D	0.05			-0.14	
Maximum Severity at j = 3	D	**0.21**			0.16	
Maximum Severity at j = 4	D	**0.28**			**0.57**	
Maximum Severity at j = 5	D	0.23			0.35	
Maximum Severity at j = 6	D	**0.98**			**0.47**	
Body regions injured at j is unknown	D	**-0.12**			-0.04	
Body regions injured at j = 1	D					
Body regions injured at j = 2	D	-0.01			**-0.12**	
Body regions injured at j = 3	D	**0.10**			-0.01	
Body regions injured at j = 4	D	**0.23**			**0.26**	
Body regions injured at j = 5	D	0.18			-0.35	
WPI at j is 10% or less	D					
WPI at j is more than 10%	D	**-1.82**	**-1.75**	**-2.03**	**-2.06**	**-1.72**
WPI at j is unknown	D	**-6.29**	**-6.63**	**-6.56**	**-6.07**	**-5.99**
Claim Size has been 5 times initial est.	D	**0.25**	**0.17**	**0.35**	**0.37**	**0.24**
Payment at j exceed 70% of Incurred	D	**0.67**	**0.90**		**0.97**	**1.30**
Litigation Level Increased	D	**-0.23**			**-0.25**	
Legal Rep Appointed	D	**-0.46**	**-0.65**	**-0.43**	**-0.24**	**-0.35**
WPI threshold met	D	**0.61**	**0.43**	**0.47**	**0.63**	**0.68**
Whiplash injury developed	D	**0.24**	**0.23**			**0.52**
ISS increased	D	**0.23**	**0.25**	**0.21**	**0.36**	**0.17**
Liability Became Rejected	D	**0.18**	**0.37**		**0.23**	**0.32**
Maximum Severity Increased	D		**0.27**		**-0.25**	
Rehab Needs Increased	D			**0.47**		
Number of body region injured Increased	D	**0.35**	**0.34**	**0.42**	**0.40**	**0.42**
Claim duation at j (log)	D	**0.62**	**0.77**	**0.41**	**0.70**	**1.00**
Brain Injury at j	D	**-0.16**				**-0.24**
Spine Injury at j	D					**-0.45**
Back injury at j	D	**-0.10**			**-0.35**	
Whiplash Injury at j	D				**0.26**	
Legal Rep at j	D	**-0.51**		**-0.28**	**-0.45**	**-0.46**
Economic Loss compensation at j	D	**0.12**	**0.14**	**0.11**	**0.16**	**0.13**
No Litigation at j	D					
NSW Supreme Court at j	D				**0.82**	**-0.42**
NSW District Court at j	D	**-0.11**	**-0.37**		**-0.23**	**-0.37**
NSW Local Court at j	D	**0.32**				
CARS at j	D		**-0.21**			
Accepted Liability at j	D					
Partially Accepted Liability at j	D					
Rejected Liabiliy at j	D	**0.21**	**0.20**		**0.17**	**0.22**
Other at j	D	**-0.18**			**-0.20**	**-0.51**
Rehab Needed at j	D					
Rehab Probably at j	D					
Rehab Possible at j	D				**0.11**	**0.23**
Rehab not required at j	D	**-0.17**	**-0.12**			
Other at j	D				**-0.18**	

*italicised variables represents the baseline value of a categorical variable

Table E.4: Coefficients - Settlement Component by Injury Severity 2

Similar to the delay process, the more severe claims can be modelled using fewer covariates. It is also interesting to note that the Settlement probability at the higher severities is not impacted by the type of injury as all injury-related covariances (Number of Regions Injured, Whiplash injury, Spinal injury, ISS, etc) becomes insignificant. The insurer effect remains significant and the pattern of coefficients remain consistent for all injury severities. The coefficients at the higher Revisions are significantly higher and upon a previous negative change, the odds of finalising at the current revisions are considerably increased.

E.3 Direction Component by Injury Severity

The following tables show the modelled coefficients for all claims model and also models stratified by injury severity for the Direction Component.

Direction Model						
Model		All	Stratified by Injury Severity and Representation			
		Claims	Minor R	Minor NR	Moderate	Severe
Number of Claims		80,998	26,758	30,118	15,036	9,086
Number of Revisions		189,507	70,410	58,356	36,948	23,793
Number of Parameters		87	74	70	64	60
Log-likelihood		-85,344	-31,987	-24,548	-16,798	-11,229
Parameter	Type	coeff	coeff	coeff	coeff	coeff
Intercept		**1.20**	**1.68**	**1.72**	**1.76**	**2.61**
phi(1)		**0.10**	**0.09**	**0.20**	**0.10**	0.08
Current Settlement	F	**-2.63**	**-2.58**	**-2.52**	**-2.64**	**-2.72**
Current Delay (log)	F	**0.29**	**0.30**	**0.32**	**0.25**	**0.25**
Previous Size	F	**-0.03**	**-0.04**		**-0.08**	
Previous Delay (log)	F	**0.05**	**0.04**	**0.11**		
Previous Incurred Cost (log)	De	**-0.30**	**-0.31**	**-0.36**	**-0.32**	**-0.33**
Revision 1	De					
Revision 2	De	**0.18**	**0.13**	**0.17**	0.08	-0.08
Revision 3	De	**0.41**	**0.34**	**0.40**	**0.30**	0.15
Revision 4	De	**0.51**	**0.42**	**0.58**	**0.37**	0.19
Revision 5	De	**0.57**	**0.56**	**0.61**	**0.46**	0.10
Revision 6 +	De	**0.57**	**0.52**	**0.67**	**0.47**	0.14
Revision Year = 2001	De	**0.48**	0.21	0.25	0.04	**1.00**
Revision Year = 2002	De	**0.22**	-0.16	0.12	-0.16	**0.58**
Revision Year = 2003	De	**0.22**	-0.10	0.02	-0.10	**0.79**
Revision Year = 2004	De	**0.28**	0.03	0.13	0.08	**0.61**
Revision Year = 2005	De	**0.26**	**0.16**	0.10	0.12	**0.35**
Revision Year = 2006	De	**0.23**	**0.15**	0.07	0.09	**0.44**
Revision Year = 2007	De	0.02	0.02	-0.09	**-0.15**	0.10
Revision Year = 2008	De	-0.03	-0.08	-0.03	-0.10	0.05
Revision Year = 2009	De					
Reporting Delay (log)	S	**-0.05**	**-0.08**			
Insurer A	S					
Insurer B	S	**0.24**	**0.26**	**0.28**	**0.27**	**0.20**
Insurer C	S	**0.36**	**0.24**	**0.67**	**0.39**	**0.42**
Insurer D	S	**0.28**	**0.28**	**0.39**	**0.43**	**0.31**
Insurer E	S	**-0.50**	**-0.42**	**-0.64**	**-0.46**	**-0.32**
Insurer F	S	**-0.12**	**-0.19**	0.01	-0.07	**0.17**
Insurer Other	S	**0.27**	**0.23**	**0.27**	**0.44**	**0.28**
Employed	S					
Self Employed	S	**0.11**	**0.10**	**0.19**		
Unemployed	S					
Other	S		**-0.12**			
Age between 0 - 9	S	**-0.40**	**-0.59**	**-0.31**	**-0.67**	-0.00
Age between 10 - 16	S	**-0.31**	**-0.39**	**-0.35**	**-0.49**	-0.14
Age between 17 - 25	S	**-0.15**	**-0.18**	**-0.16**	**-0.15**	-0.09
Age between 26 - 45	S					
Age between 46 - 65	S	**-0.05**	**-0.08**	0.06	**-0.09**	**-0.14**
Age 66 +	S	**-0.12**	**-0.11**	0.09	**-0.25**	**-0.42**
Resides in Metro area	S					
Resides in Outer Metro area	S	**-0.11**	**-0.12**		**-0.13**	**-0.16**
Resides in Country area	S				**-0.27**	
Resides in Wollongong area	S					
Resides in Newcastle area	S	**-0.07**	**-0.11**		**-0.15**	
Accident Year = 2001	S					
Accident Year = 2002	S	0.04	0.05	-0.03	0.01	0.10
Accident Year = 2003	S	**0.09**	**0.15**	-0.12	0.03	**0.21**
Accident Year = 2004	S	**0.22**	**0.13**	0.06	**0.21**	**0.39**
Accident Year = 2005	S	**0.41**	**0.35**	**0.23**	**0.29**	**0.62**
Accident Year = 2006	S	**0.55**	**0.52**	**0.32**	**0.37**	**0.76**
Accident Year = 2007	S	**0.71**	**0.59**	**0.55**	**0.52**	**0.75**
Accident Year = 2008	S	**0.99**	**0.81**	**0.92**	**0.57**	**1.14**
Accident Year = 2009	S	**1.66**	**1.58**	**1.46**	**1.85**	**1.92**
Max. Severity at report	S	**0.02**			**0.17**	

*italicised variables represents the baseline value of a categorical variable

Table E.5: Coefficients - Direction Component by Severity 1

Direction Model (cont.)						
Model		All	Stratified by Injury Severity and Representation			
		Claims	Minor R	Minor NR	Moderate	Severe
Number of Claims		80,998	26,758	30,118	15,036	9,086
Number of Revisions		189,507	70,410	58,356	36,948	23,793
Number of Parameters		87	74	70	64	60
Log-likelihood		-85,344	-31,987	-24,548	-16,798	-11,229
Parameter	Type	coeff	coeff	coeff	coeff	coeff
ISS at j = 0	D	-0.04	0.12	-0.06	0.06	
ISS at j = 1	D					
ISS at j = 2	D	**0.10**	0.05	0.08	0.18	
ISS at j between 3 and 5	D	**0.32**	**0.21**	**0.28**	**0.41**	
ISS at j between 6 and 10	D	**0.47**	0.12	**0.37**	**0.51**	
ISS at j between 11 and 30	D	**0.48**	**0.32**	**0.43**	**0.49**	
ISS at j between 31 and 75	D	**0.43**	**1.16**	0.01	-0.03	
Maximum Severity at j is unknown	D	0.12	0.03			
Maximum Severity at j = 1	D					
Maximum Severity at j = 2	D	-0.05	**0.15**			
Maximum Severity at j = 3	D	-0.07	0.08			
Maximum Severity at j = 4	D	-0.01	-0.06			
Maximum Severity at j = 5	D	**-0.28**	**-0.78**			
Maximum Severity at j = 6	D	-0.20	-0.44			
Body regions injured at j is unknown	D	**0.10**		**0.18**		-0.02
Body regions injured at j = 1	D					
Body regions injured at j = 2	D	**-0.07**		-0.04		**-0.10**
Body regions injured at j = 3	D	**-0.09**		-0.02		**-0.18**
Body regions injured at j = 4	D	**-0.13**		**-0.20**		-0.10
Body regions injured at j = 5	D	0.02		0.27		0.22
WPI at j is 10% or less	D					
WPI at j is more than 10%	D	**1.14**	**1.04**	**1.27**	**1.42**	**1.38**
WPI at j is unknown	D	**1.20**	**1.19**	**1.41**	**1.23**	**1.26**
Claim Size has been 5 times initial est.	D	**-0.14**	**-0.20**	**-0.13**	**-0.20**	
Payment at j exceed 70% of Incurred	D	**2.04**	**1.92**	**2.48**	**1.90**	**1.42**
Brain Injury Developed	D	**0.29**		**1.06**		**0.30**
Litigation Level Increased	D	**0.41**	**0.44**		**0.25**	**0.43**
Legal Rep Appointed	D	**0.21**	**0.20**		**0.35**	
WPI threshold met	D			**0.24**		**-0.19**
ISS increased	D			**-0.11**		
Liability Became Rejected	D	**0.15**				
Maximum Severity Increased	D	**0.08**				
Liability Became Accepted	D	**-0.43**	**-0.38**			**-0.58**
Rehab Needs Increased	D	**-0.36**	**-0.27**	**-0.47**	**-0.57**	
Claim duation at j (log)	D	**0.14**	**0.13**	**0.16**		
Back injury at j	D	**0.10**		**-0.42**	**0.15**	
Whiplash Injury at j	D			**0.47**		**0.15**
Legal Rep at j	D	**0.78**	**0.92**	**0.79**	**0.69**	**0.66**
Economic Loss compensation at j	D	**0.24**	**0.25**	**0.32**	**0.15**	**0.12**
No Litigation at j	D					
NSW Supreme Court at j	D	**-0.24**				
NSW District Court at j	D	**-0.13**				**-0.23**
NSW Local Court at j	D	**-0.27**		**-0.48**		**-0.48**
CARS at j	D	**0.35**	**0.45**	**0.37**	**0.37**	**0.17**
Accepted Liability at j	D					
Partially Accepted Liability at j	D	**-0.11**	**-0.14**			
Rejected Liabiliy at j	D	**-0.40**	**-0.41**	**-0.35**	**-0.26**	**-0.24**
Other at j	D	**0.28**	**0.18**	**0.25**	**0.24**	**0.58**
Rehab Needed at j	D					
Rehab Probably at j	D	**0.06**				
Rehab Possible at j	D	**0.15**	**0.12**	**0.24**	**0.11**	
Rehab not required at j	D	**0.30**	**0.31**	**0.40**	**0.34**	**0.14**
Other at j	D	**-0.19**	**-0.37**	**-0.12**	**-0.22**	**-0.31**

*italicised variables represents the baseline value of a categorical variable

Table E.6: Coefficients - Direction Component by Injury Severity 2

Similar observations are made for the Direction Component as the previous component models. However, one interesting observation is that Severe claims are less likely to have a positive movements when compared to the other injury severities and that Insurer is not a significant covariates - this suggest that the various insurers are equally conservative about serious claims and all take a more conservative approach relative to the less serious injuries.

E.4 Size Component by Injury Severity

The following tables show the modelled coefficients for all claims model and also models stratified by injury severity for the Size Component.

Size Model						
Model		All	Stratified by Injury Severity and Representation			
		Claims	Minor R	Minor NR	Moderate	Severe
Number of Claims		80,998	26,758	30,118	15,036	9,086
Number of Revisions		189,507	70,410	58,356	36,948	23,793
Number of Parameters		92	75	73	74	72
Log-likelihood		-112,345	-37,242	-42,466	-20,858	-9,555
Parameter	Type	coeff	coeff	coeff	coeff	coeff
Intercept		**1.31**	**2.17**	**-0.22**	**1.99**	**2.58**
phi(1)		**0.08**	**0.08**	**0.09**	**0.08**	**0.09**
c		**0.86**	**0.86**	**0.91**	**0.86**	**0.85**
Saving on Finalisations	F	**1.22**	**1.09**	**1.70**	**1.07**	**0.89**
Current Direction	F	**0.30**	**0.26**	**0.75**	**0.09**	**-0.19**
Current Settlement	F	**-0.69**	**-0.65**	**-0.97**	**-0.63**	**-0.43**
Current Delay (log)	F	**0.15**	**0.14**	**0.19**	**0.14**	**0.13**
Previous Direction	F	**-0.08**	**-0.05**	**-0.08**	**-0.10**	**-0.13**
Previous Delay (log)	F	**0.07**	**0.07**	**0.06**	**0.07**	**0.06**
Previous Incurred Cost (log)	De	**-0.15**	**-0.22**		**-0.19**	**-0.26**
Revision 1	De					
Revision 2	De	**-0.23**	**-0.19**	**-0.30**	**-0.23**	**-0.20**
Revision 3	De	**-0.24**	**-0.19**	**-0.34**	**-0.21**	**-0.24**
Revision 4	De	**-0.22**	**-0.16**	**-0.35**	**-0.25**	**-0.21**
Revision 5	De	**-0.20**	**-0.14**	**-0.33**	**-0.20**	**-0.20**
Revision 6 +	De	**-0.22**	**-0.16**	**-0.36**	**-0.20**	**-0.27**
Revision Year = 2001	De	-0.04	-0.07	0.05	0.02	**0.23**
Revision Year = 2002	De	0.04	-0.01	**0.13**	0.07	**0.28**
Revision Year = 2003	De	**0.15**	**0.10**	**0.20**	**0.18**	**0.39**
Revision Year = 2004	De	**0.16**	**0.12**	**0.19**	**0.17**	**0.33**
Revision Year = 2005	De	**0.16**	**0.14**	**0.16**	**0.18**	**0.28**
Revision Year = 2006	De	**0.16**	**0.14**	**0.18**	**0.15**	**0.26**
Revision Year = 2007	De	**0.12**	**0.11**	**0.15**	**0.09**	**0.17**
Revision Year = 2008	De	**0.17**	**0.15**	**0.23**	**0.10**	**0.18**
Revision Year = 2009	De					
Reporting Delay (log)	S	**0.05**	**0.07**		**0.07**	**0.11**
Insurer A	S					
Insurer B	S	**0.02**	**0.05**	**-0.10**	**0.05**	**0.13**
Insurer C	S	**-0.04**	0.00	**-0.07**	-0.05	-0.03
Insurer D	S	**0.11**	**0.16**	-0.02	**0.13**	**0.19**
Insurer E	S	0.01	0.01	**-0.08**	0.06	**0.14**
Insurer F	S	**0.05**	**0.10**	**-0.09**	**0.12**	**0.22**
Insurer Other	S	**0.04**	**0.06**	**-0.05**	**0.08**	**0.08**
Gender = Female	S	**-0.05**	**-0.05**	**-0.04**	**-0.06**	
Employed	S					
Self Employed	S	**-0.02**				
Unemployed	S	**0.08**	**0.07**			
Other	S		**-0.03**			
Age between 0 - 9	S	**-0.05**	**-0.18**	**0.16**	-0.05	0.02
Age between 10 - 16	S	**-0.05**	**-0.11**	-0.00	-0.05	0.00
Age between 17 - 25	S	**-0.03**	-0.03	0.01	**-0.06**	-0.03
Age between 26 - 45	S					
Age between 46 - 65	S	**-0.03**	**-0.04**	-0.01	**-0.05**	**-0.06**
Age 66 +	S	**-0.04**	**-0.06**	-0.02	**-0.07**	**-0.10**
Resides in Metro area	S					
Resides in Outer Metro area	S	**0.03**		**0.04**	**0.05**	
Resides in Country area	S					
Resides in Wollongong area	S					
Resides in Newcastle area	S			**0.03**		
Accident Year = 2001	S					
Accident Year = 2002	S	**-0.02**	-0.01	-0.01	-0.02	0.02
Accident Year = 2003	S	0.00	0.02	**0.05**	0.02	**0.09**
Accident Year = 2004	S	0.01	0.03	0.05	0.01	**0.15**
Accident Year = 2005	S	0.03	**0.08**	**0.07**	0.05	**0.15**
Accident Year = 2006	S	**0.07**	**0.09**	**0.11**	**0.11**	**0.23**
Accident Year = 2007	S	**0.08**	**0.12**	**0.12**	0.09	**0.26**
Accident Year = 2008	S	**0.30**	**0.34**	**0.38**	**0.18**	**0.32**
Accident Year = 2009	S	**0.24**	**0.27**	**0.21**	**0.34**	**0.48**
Legal Rep at report	S	**-0.04**				**-0.10**
Max. Severity at report	S	**0.02**				

*italicised variables represents the baseline value of a categorical variable

Table E.7: Coefficients - Size Component by Severity 1

Size Model (cont.)						
Model		All	Stratified by Injury Severity and Representation			
		Claims	Minor R	Minor NR	Moderate	Severe
Number of Claims		80,998	26,758	30,118	15,036	9,086
Number of Revisions		189,507	70,410	58,356	36,948	23,793
Number of Parameters		92	75	73	74	72
Log-likelihood		-112,345	-37,242	-42,466	-20,858	-9,555
Parameter	Type	coeff	coeff	coeff	coeff	coeff
ISS at j = 0	D	0.02	**0.04**	**0.08**	0.16	
ISS at j = 1	D					
ISS at j = 2	D	**-0.02**	-0.00	**-0.03**	-0.10	
ISS at j between 3 and 5	D	-0.01	0.01	**-0.05**	-0.10	
ISS at j between 6 and 10	D	**-0.06**	-0.02	**-0.11**	-0.11	
ISS at j between 11 and 30	D	**-0.09**	**0.07**	**-0.17**	-0.12	
ISS at j between 31 and 75	D	**-0.17**	**0.19**	-0.07	**-0.20**	
Maximum Severity at j is unknown	D	0.07				**0.95**
Maximum Severity at j = 1	D					
Maximum Severity at j = 2	D	**0.05**				**0.24**
Maximum Severity at j = 3	D	**0.06**				0.14
Maximum Severity at j = 4	D	0.02				0.15
Maximum Severity at j = 5	D	0.07				**0.20**
Maximum Severity at j = 6	D	0.04				0.29
Body regions injured at j is unknown	D	**-0.03**				
Body regions injured at j = 1	D					
Body regions injured at j = 2	D	**0.03**				
Body regions injured at j = 3	D	0.02				
Body regions injured at j = 4	D	0.00				
Body regions injured at j = 5	D	0.05				
WPI at j is 10% or less	D					
WPI at j is more than 10%	D	**-0.68**	**-0.46**	**-0.79**	**-0.60**	**-0.65**
WPI at j is unknown	D	**-0.65**	**-0.65**	**-0.81**	**-0.59**	**-0.58**
Claim Size has been 5 times initial est.	D	**-0.08**	**-0.06**	**-0.13**	**-0.09**	**-0.06**
Payment at j exceed 70% of Incurred	D	**-0.13**	**-0.26**	**0.11**	**-0.25**	**-0.42**
Brain Injury Developed	D		**0.35**			**-0.11**
Litigation Level Increased	D	**0.21**	**0.28**	**0.23**	**0.22**	**0.12**
Legal Rep Appointed	D	**0.14**	**0.18**	**0.14**	**0.10**	
WPI threshold met	D	**0.46**	**0.38**	**0.56**	**0.49**	**0.28**
Whiplash injury developed	D			**0.07**		
Liability Became Rejected	D	**0.07**	**0.13**		**0.09**	**0.10**
Maximum Severity Increased	D	**0.05**	**0.05**	**0.05**	**0.06**	**0.08**
Liability Became Accepted	D	**0.34**	**0.28**	**0.25**	**0.42**	**0.33**
Rehab Needs Increased	D	**0.36**	**0.38**	**0.21**	**0.46**	**0.30**
Number of body region injured Increased	D	**0.05**		**0.05**	**0.05**	
Claim duation at j (log)	D	**-0.04**		**-0.08**		
Brain Injury at j	D	**0.16**				**0.25**
Spine Injury at j	D			**-0.91**		**0.18**
Whiplash Injury at j	D	**-0.05**	**-0.05**		**-0.06**	
Legal Rep at j	D	**0.06**	**-0.21**	**0.04**		**0.15**
Economic Loss compensation at j	D					
No Litigation at j	D					
NSW Supreme Court at j	D	**-0.42**	**-0.36**	**-0.59**	**-0.34**	**-0.35**
NSW District Court at j	D	**-0.43**	**-0.38**	**-0.45**	**-0.42**	**-0.38**
NSW Local Court at j	D	**-0.30**	**-0.34**	**-0.39**	**-0.21**	**-0.18**
CARS at j	D	**-0.07**	-0.03	**-0.08**	**-0.11**	**-0.13**
Accepted Liability at j	D					
Partially Accepted Liability at j	D			**0.06**		
Rejected Liabiliy at j	D	**0.58**	**0.55**	**0.51**	**0.57**	**0.58**
Other at j	D	**0.27**	**0.31**	**0.25**	**0.26**	**0.25**
Rehab Needed at j	D					
Rehab Probably at j	D	**0.06**	**0.06**		**0.08**	
Rehab Possible at j	D	**0.05**	**0.06**		**0.06**	
Rehab not required at j	D	**0.07**	**0.08**		**0.10**	**0.09**
Other at j	D	**-0.07**		**-0.07**	**-0.13**	**-0.14**

*italicised variables represents the baseline value of a categorical variable

Table E.8: Coefficients - Size Component by Injury Severity 2

The coefficients show that, at base levels, the size of the movements are materially bigger for the more serious injuries and that the movement sizes are more volatile as observed though a smaller c parameter.

Appendix F

Log-Likelihoods and Their Derivatives

F.1 Likelihoods for the Negative Binomial Response Distribution

The Delay component process is modelled using a negative binomial distribution. For this section, we transform t_j to a standard negative binomial distribution, i.e., $l_j = t_j - 1$, where $l_j = 0, 1, 2, \ldots$. This is done to retain the application of the following derivation to other negative binomial distributions. While the following derivation can be found in Wang [2004] for the analysis of stock data, it is reproduced below. Subscript i has been dropped to simplify notation

$$P(l_j|F_{j-1}) = \frac{\Gamma(\alpha + l_j)}{\Gamma(\alpha)\Gamma(l_j + 1)} \Big(\frac{\alpha}{\mu_j + \alpha}\Big)^{\alpha} \Big(\frac{\mu_j}{\mu_j + \alpha}\Big)^{l_j} \tag{F.1}$$

$$\log(\mu_j) = W_{l,j} = \beta_\lambda X_{\lambda,j}^T + Z_{l,j} \tag{F.2}$$

$$Z_{l,j} = \sum_{k=1}^{p} \phi_k Z_{j-k} + \sum_{k=1}^{q} \theta_k e_{j-k} \tag{F.3}$$

and

$$e_j = \frac{l_j - \mu_j}{\sqrt{\mu_j + \frac{\mu_j^2}{\alpha}}} \tag{F.4}$$

Note W_j, X_j and Z_j in this section take on different definition compared to those in the main body of the thesis. The notation in the likelihood derivation is consistent with the R programs. W_j in this section is the linear predictor, X_j is the design matrix of explanatory variables, and Z_j is the ARMA component of the linear predictor. Other notations are consistent.

The log-likelihood contribution of the jth revision is therefore,

$$\begin{aligned}\ell_l &= \log\left(\Gamma(\alpha + l_j)\right) - \log\left(\Gamma(\alpha)\right) - \log\left(\Gamma(l_j + 1)\right) + \alpha \log\left(\frac{\alpha}{\mu_j + \alpha}\right) + l_j \log\left(\frac{\mu_j}{\mu_j + \alpha}\right) \\ &= \log\left(\Gamma(\alpha + l_j)\right) - \log\left(\Gamma(\alpha)\right) - \log\left(\Gamma(l_j + 1)\right) + \alpha \log \alpha + l_j W_{l,j} - (\alpha + l_j) \log(e^{W_{l,j}} + \alpha)\end{aligned} \tag{F.5}$$

The first order derivatives of ℓ_l with respect to δ is

$$\frac{\partial \ell_l}{\partial \delta} = \frac{\partial \ell_l}{\partial W}\frac{\partial W}{\partial \delta} + u\frac{\partial \ell_l}{\partial \alpha} \tag{F.6}$$

where

$$\frac{\partial \ell_Y}{\partial W} = \frac{\alpha(l_j - e^{W_{t,j}})}{e^{W_{t,j}} + \alpha} \tag{F.7}$$

and

$$\frac{\partial \ell_Y}{\partial \alpha} = \digamma(\alpha + l_j) - \digamma(\alpha) + \frac{e^{W_{t,j}} - l_j}{e^{W_{t,j}} + \alpha} + \log\left(\frac{\alpha}{e^{W_{t,j}} + \alpha}\right) \tag{F.8}$$

where $\digamma$ is the digamma function, or the second derivative of the log of the gamma funcion.

The vector u takes the definition of a vector of 0's except the last element, which is a 1.

The second order derivatives of ℓ_l with respect to δ is

$$\begin{aligned}\frac{\partial^2 \ell_l}{\partial \delta^2} &= \frac{\partial \ell_l}{\partial W}\frac{\partial^2 W}{\partial \delta^2} + \frac{\partial^2 \ell_l}{\partial W^2}\frac{\partial W}{\partial \delta}\left(\frac{\partial W}{\partial \delta}\right)^T \\ &\quad + u\frac{\partial^2 \ell_l}{\partial W \partial \alpha}\frac{\partial W}{\partial \delta} + \left(\frac{\partial W}{\partial \delta}\right)^T \frac{\partial^2 \ell_l}{\partial W \partial \alpha} u^T + uu^T \frac{\partial^2 \ell_l}{\partial \alpha^2}\end{aligned} \tag{F.9}$$

where

$$\frac{\partial^2 \ell_l}{\partial W^2} = -\frac{(l_j + \alpha)\alpha e^{W_{l,j}}}{(e^{W_{l,j}} + \alpha)^2} \tag{F.10}$$

$$\frac{\partial^2 \ell_l}{\partial W \partial \alpha} = \frac{(l_j - e^{W_{l,j}})e^{W_{l,j}}}{(e^{W_{l,j}} + \alpha)^2} \tag{F.11}$$

and

$$\frac{\partial^2 \ell_l}{\partial c^2} = \frac{l_j \alpha + e^{2W_{l,j}}}{\alpha(e^{W_{l,j}} + \alpha)^2} + \text{trigamma}(\alpha + l_j) - \text{trigamma}(\alpha) \tag{F.12}$$

where the trigamma function is the third derivative of the log of the gamma function.

In the above derivatives, the calculations of $\frac{\partial W}{\partial \delta}$ and $\frac{\partial^2 W}{\partial \delta^2}$ are detailed in Davis et al. [2003]. These calculations also require the derivatives of the Pearson residuals e_j's against the δ.

The first order derivatives of e_j with respect to δ is

$$\frac{\partial e_j}{\partial \delta} = \frac{\partial e_j}{\partial W}\frac{\partial W}{\partial \delta} + u\frac{\partial e_j}{\partial \alpha} \tag{F.13}$$

where

$$\frac{\partial e_j}{\partial W} = \frac{-e^{W_{l,j}}(\alpha l_j + e^{W_{l,j}}(\alpha + 2l_j))}{2\alpha\left(\frac{e^{W_{l,j}}(e^{W_{l,j}}+\alpha)}{\alpha}\right)^{\frac{3}{2}}} \tag{F.14}$$

and

$$\frac{\partial e_j}{\partial \alpha} = \frac{e^{2W_{l,j}}(l_j - e^{W_{l,j}})}{2\left(e^{W_{l,j}} + \frac{e^{2W_{l,j}}}{\alpha}\right)^{\frac{3}{2}}\alpha^2} \tag{F.15}$$

The second order derivatives of e_j with respect to δ is

$$\begin{aligned}\frac{\partial^2 e_j}{\partial \delta^2} &= \frac{\partial e_j}{\partial W}\frac{\partial^2 W}{\partial \delta^2} + \frac{\partial^2 e_j}{\partial W^2}\frac{\partial W}{\partial \delta}\left(\frac{\partial W}{\partial \delta}\right)^T \\ &+ u\frac{\partial^2 e_j}{\partial W \partial \alpha}\frac{\partial W}{\partial \delta} + \left(\frac{\partial W}{\partial \delta}\right)^T \frac{\partial^2 e_j}{\partial W \partial \alpha}u^T + uu^T \frac{\partial^2 e_j}{\partial \alpha^2}\end{aligned} \tag{F.16}$$

where

$$\frac{\partial^2 e_j}{\partial W^2} = \frac{l_j \alpha^2 + e^{W_{l,j}}\alpha(2l_j - \alpha) + 2e^{2W_{l,j}}(2l_j + \alpha)}{4(e^{W_{l,j}} + \alpha)^2 \left(e^{W_{l,j}}\left(\frac{e^{W_{l,j}}+\alpha}{\alpha}\right)\right)^{\frac{1}{2}}} \tag{F.17}$$

$$\frac{\partial^2 e_j}{\partial W \partial \alpha} = \frac{\left(e^{W_{l,j}}\left(\frac{e^{W_{l,j}}+\alpha}{\alpha}\right)\right)^{\frac{1}{2}}(\alpha l_j - e^{W_{l,j}}(2l_j + 3\alpha))}{4(e^{W_{l,j}} + \alpha)^3} \tag{F.18}$$

and

$$\frac{\partial^2 e_j}{\partial \alpha^2} = \frac{\left(e^{W_{l,j}}\left(\frac{e^{W_{l,j}}+\alpha}{\alpha}\right)\right)^{\frac{1}{2}}(e^{W_{l,j}} - l_j)(e^{W_{l,j}} + 4\alpha)}{4\alpha(e^{W_{l,j}} + \alpha)^3} \tag{F.19}$$

F.2 Censoring - Delay Component Process

This section provides the adjustment required to allow to censored claim in the Delay component process. While the above log-likelihood is correct for claims observed to finalisation an adjustment needs to be made to allowed for those claims censored by the end of the data period. The below adjustments to the log-likelihood and also to its derivatives are required for the model fitting of the Delay component adjusting for the censoring of claims.

From Equation (6.29), the log-likelihood for the Delay component is

$$\ell(\beta_\mu, \alpha) = \sum_{i=1}^{n}[\sum_{j=1}^{\min(m,j')} \log P(t_{j,i}) + \log P(t_{j'+1} > (T' - T_{j'}))] \tag{F.20}$$

Similarly to the previous section, the transformation $l_j = t_j - 1$ was carried out.

Firstly, the log-likelihood is broken into those claims that have been finalised and those claims that are still open.

$$\begin{aligned}\ell(\beta_\mu, \alpha) = &\sum_{j'=m} \sum_{j=1}^{\min(m,j')} \log P(l_{j,i}) \\ &+ \sum_{j'<m} [\sum_{j=1}^{j'} \log P(l_{j,i}) + \log P(l_{j'+1} > (T' - T_{j'} - 1))]\end{aligned} \tag{F.21}$$

For the purposes of this section we will consider the latter part of the above equation and use ℓ_C to denote the last term in the above equation, the log-likelihood adjustments due to the censored claims.

$$\begin{aligned}\ell_C(\beta_\mu, \alpha) &= \sum_{j'<m} \log\left(P(l_{j'+1} > (T' - T_{j'} - 1))\right) \\ &= \sum_{j'<m} \log\left(1 - \sum_{k=0}^{T'-T_{j'}-1} P(l_{j'+1,i} = k)\right)\end{aligned} \tag{F.22}$$

To simplify the notation, let $A = 1 - \sum_{k=0}^{v} P(l = k)$ where $v = T' - T_{j'} - 1$.

The first order derivatives of A can be calculated as

$$\frac{\partial \log A}{\partial \delta} = \frac{\partial \log A}{\partial W}\frac{\partial W}{\partial \delta} + u\frac{\partial \log A}{\partial \alpha} \tag{F.23}$$

$$\begin{aligned}\frac{\partial \log A}{\partial W} &= \frac{1}{A}\frac{\partial A}{\partial W} \\ &= \frac{1}{A}\left(-\sum_{k=0}^{v} \frac{\partial P(l = k)}{\partial W}\right) \\ &= \frac{1}{A}\left(-\sum_{k=0}^{v} \frac{\partial e^{\ell_k}}{\partial W}\right) \\ &= \frac{1}{A}\left(-\sum_{k=0}^{v} e^{\ell_k}\frac{\partial \ell_k}{\partial W}\right)\end{aligned} \tag{F.24}$$

where ℓ_k is the log likelihood of the probability that the next revision has a delay of k and $\frac{\partial \ell_k}{\partial W}$ is an quantity that is already calculated in the model fitting code.

Similarly,

$$\begin{aligned}\frac{\partial \log A}{\partial \alpha} &= \frac{1}{A}\frac{\partial A}{\partial \alpha} \\ &= \frac{1}{A}\left(-\sum_{k=0}^{v}\frac{\partial P(l=k)}{\partial \alpha}\right)\end{aligned} \tag{F.25}$$

The second order derivatives can be calculated as

$$\begin{aligned}\frac{\partial^2 \log A}{\partial \delta^2} = &\frac{\partial \log A}{\partial W}\frac{\partial^2 W}{\partial \delta^2} + \frac{\partial^2 \log A}{\partial W^2}\frac{\partial W}{\partial \delta}\left(\frac{\partial W}{\partial \delta}\right)^T \\ &+ u\frac{\partial^2 \log A}{\partial W \partial \alpha}\frac{\partial W}{\partial \delta} + \left(\frac{\partial W}{\partial \delta}\right)^T\frac{\partial^2 \log A}{\partial W \partial \alpha}u^T + uu^T\frac{\partial^2 \log A}{\partial \alpha^2}\end{aligned} \tag{F.26}$$

$$\begin{aligned}\frac{\partial^2 \log A}{\partial W^2} &= -\frac{1}{A^2}\left(\frac{\partial A}{\partial W}\right)^2 + \frac{1}{A}\frac{\partial^2 A}{\partial W^2} \\ &= -\frac{1}{A^2}\left(-\sum_{k=0}^{v}\frac{\partial P(l=k)}{\partial W}\right)^2 + \frac{1}{A}\left(-\sum_{k=0}^{v}\frac{\partial^2 P(l=k)}{\partial W^2}\right)\end{aligned} \tag{F.27}$$

$$\begin{aligned}\frac{\partial^2 \log A}{\partial W \partial \alpha} &= -\frac{1}{A^2}\left(\frac{\partial A}{\partial W}\right)\left(\frac{\partial A}{\partial \alpha}\right) + \frac{1}{A}\frac{\partial^2 A}{\partial W \partial \alpha} \\ &= -\frac{1}{A^2}\left(-\sum_{k=0}^{v}\frac{\partial P(l=k)}{\partial W}\right)\left(-\sum_{k=0}^{v}\frac{\partial P(l=k)}{\partial \alpha}\right) + \frac{1}{A}\left(-\sum_{k=0}^{v}\frac{\partial^2 P(l=k)}{\partial W \partial \alpha}\right)\end{aligned} \tag{F.28}$$

$$\begin{aligned}\frac{\partial^2 \log A}{\partial \alpha^2} &= -\frac{1}{A^2}\left(\frac{\partial A}{\partial \alpha}\right)^2 + \frac{1}{A}\frac{\partial^2 A}{\partial \alpha^2} \\ &= -\frac{1}{A^2}\left(-\sum_{k=0}^{v}\frac{\partial P(l=k)}{\partial \alpha}\right)^2 + \frac{1}{A}\left(-\sum_{k=0}^{v}\frac{\partial^2 P(l=k)}{\partial \alpha^2}\right)\end{aligned} \tag{F.29}$$

F.3 Likelihoods for Gamma Response Distribution

From Section 6.4 we have specified the size component to take a gamma distribution with the follow parametrisation. The subscription i has been dropped to simplify the notation while j refers to the ordinal number of revisions a claim has had.

$$f(y_j|F_{j-1}, G_j, t_j, S_j, D_j) = \frac{1}{\Gamma(c)\lambda_j}\Big(\frac{y_j}{\lambda_j}\Big)^{c-1} e^{-(\frac{y_j}{\lambda_j})} \tag{F.30}$$

$$\log(\lambda_j) = W_{Y,j} = \beta_\lambda X_{\lambda,j}^T + Z_{Y,j} \tag{F.31}$$

$$Z_{Y,j} = \sum_{k=1}^{p} \phi_k Z_{j-k} + \sum_{k=1}^{q} \theta_k e_{j-k} \tag{F.32}$$

and

$$\begin{aligned} e_j &= \frac{y_j - c\lambda_j}{\sqrt{c\lambda_j^2}} \\ &= y_j c^{-1/2} e^{-W_{Y,j}} - c^{1/2} \end{aligned} \tag{F.33}$$

Similar to above W_j, X_j and Z_j take on different definition compared to those in the main body of the thesis. The notation in the likelihood derivation is consistent with the R programs. W_j in this section is the linear predictor, X_j is the design matrix of explanatory variables, and Z_j is the ARMA component of the linear predictor. Other notations are consistent.

The log-likelihood contribution of the jth revision is therefore,

$$\begin{aligned} \ell_Y &= -\log\Gamma(c) - \log\lambda_j + (c-1)(\log y_j - \log\lambda_j) - y_j/\lambda_j \\ &= -\log\Gamma(c) - W_{Y,j} + (c-1)(\log y_j - W_{Y,j}) - y_j e^{-W_{Y,j}} \end{aligned} \tag{F.34}$$

The first order derivatives of ℓ_Y with respect to δ is

$$\frac{\partial \ell_Y}{\partial \delta} = \frac{\partial \ell_Y}{\partial W}\frac{\partial W}{\partial \delta} + u\frac{\partial \ell_Y}{\partial c} \tag{F.35}$$

where

$$\frac{\partial \ell_Y}{\partial W} = -c + y_j e^{-W_{Y,j}} \tag{F.36}$$

and

$$\frac{\partial \ell_Y}{\partial c} = -\digamma(c) + \log y_j - W_{Y,j} \tag{F.37}$$

The vector u takes the previous definition of a vector of 0's except the last element, which is a 1.

The second order derivatives of ℓ_Y with respect to δ is

$$\begin{aligned}\frac{\partial^2 \ell_Y}{\partial \delta^2} = {} & \frac{\partial \ell_Y}{\partial W}\frac{\partial^2 W}{\partial \delta^2} + \frac{\partial^2 \ell_Y}{\partial W^2}\frac{\partial W}{\partial \delta}\left(\frac{\partial W}{\partial \delta}\right)^T \\ & + u\frac{\partial^2 \ell_Y}{\partial W \partial c}\frac{\partial W}{\partial \delta} + \left(\frac{\partial W}{\partial \delta}\right)^T \frac{\partial^2 \ell_Y}{\partial W \partial c}u^T + uu^T\frac{\partial^2 \ell_Y}{\partial c^2}\end{aligned} \tag{F.38}$$

where

$$\frac{\partial^2 \ell_Y}{\partial W^2} = -y_j e^{-W_{Y,j}} \tag{F.39}$$

$$\frac{\partial^2 \ell_Y}{\partial W \partial c} = -1 \tag{F.40}$$

and

$$\frac{\partial^2 \ell_Y}{\partial c^2} = -\text{trigamma}(c) \tag{F.41}$$

In the above derivatives, the calculations of $\frac{\partial W}{\partial \delta}$ and $\frac{\partial^2 W}{\partial \delta^2}$ are detailed in Davis et al. [2003]. These calculations also require the derivatives of the Pearson residuals e_js against the δ.

The first order derivatives of e_j with respect to δ is

$$\frac{\partial e_j}{\partial \delta} = \frac{\partial e_j}{\partial W}\frac{\partial W}{\partial \delta} + u\frac{\partial e_j}{\partial c} \tag{F.42}$$

where

$$\frac{\partial e_j}{\partial W} = -y_j c^{-1/2} e^{-W_{Y,j}} \tag{F.43}$$

and

$$\frac{\partial e_j}{\partial c} = -\frac{1}{2} y_j c^{-3/2} e^{-W_{Y,j}} - \frac{1}{2} c^{-1/2} \tag{F.44}$$

The second order derivatives of e_j with respect to δ is

$$\begin{aligned}\frac{\partial^2 e_j}{\partial \delta^2} = {} & \frac{\partial e_j}{\partial W}\frac{\partial^2 W}{\partial \delta^2} + \frac{\partial^2 e_j}{\partial W^2}\frac{\partial W}{\partial \delta}\left(\frac{\partial W}{\partial \delta}\right)^T \\ & + u\frac{\partial^2 e_j}{\partial W \partial c}\frac{\partial W}{\partial \delta} + \left(\frac{\partial W}{\partial \delta}\right)^T \frac{\partial^2 e_j}{\partial W \partial c} u^T + u u^T \frac{\partial^2 e_j}{\partial c^2}\end{aligned} \tag{F.45}$$

where

$$\frac{\partial^2 e_j}{\partial W^2} = y_j c^{-1/2} e^{-W_{Y,j}} \tag{F.46}$$

$$\frac{\partial^2 e_j}{\partial W \partial c} = \frac{1}{2} y_j c^{-3/2} e^{-W_{Y,j}} \tag{F.47}$$

and

$$\frac{\partial^2 e_j}{\partial c^2} = \frac{3}{4} y_j c^{-5/2} e^{-W_{Y,j}} + \frac{1}{4} c^{-3/2} \tag{F.48}$$

F.4 Likelihoods for the Generalised Gamma Response Distribution

We have also considered the generalised gamma as the response distribution for the Size component process. While we have found the generalised gamma distribution does not add material improvement to the fit of the data, we have provided the log-likelihood and its derivatives below.

$$f(y_j|F_{j-1}, G_j, t_j, S_j, D_j) = \frac{k}{\Gamma(c)\lambda_j}\left(\frac{y_j}{\lambda_j}\right)^{ck-1} e^{-(\frac{y_j}{\lambda_j})^k} \tag{F.49}$$

where

$$\log(\lambda_j) = W_{Y,j} = \beta_\lambda X_{\lambda,j}^T + Z_{Y,j} \tag{F.50}$$

$$Z_{Y,j} = \sum_{k=1}^{p} \phi_k Z_{j-k} + \sum_{k=1}^{q} \theta_k e_{j-k} \tag{F.51}$$

and

$$e_j = \frac{Y_j - \mu_j}{\sqrt{\frac{\lambda_j^2 \Gamma(c+2/k)}{\Gamma(c)} - \mu_j^2}} \tag{F.52}$$

where

$$\mu_j = \frac{\lambda_j \Gamma(c+1/k)}{\Gamma(c)} \tag{F.53}$$

The log-likelihood contribution of the jth revision is

$$\begin{aligned} \ell_Y &= \log k - \log\Gamma(c) - \log\lambda_j + (ck-1)(\log y_j - \log\lambda_j) - \left(\frac{y_j}{\lambda_j}\right)^k \\ &= \log k - \log\Gamma(c) - W_{Y,j} + (ck-1)(\log y_j - W_{Y,j}) - y_j^k e^{-kW_{Y,j}} \end{aligned} \tag{F.54}$$

Similar to above W_j, X_j and Z_j take on different definition compared to those in the main body of the thesis. The notation in the likelihood derivation is consistent with the R programs. W_j in this section is the linear predictor, X_j is the design matrix of explanatory variables, and Z_j is the ARMA component of the linear predictor. Other notations are consistent.

The first order derivatives of the log-likelihood with respect to δ is

$$\frac{\partial \ell_Y}{\partial \delta} = \frac{\partial \ell_Y}{\partial W}\frac{\partial W}{\partial \delta} + u_2 \frac{\partial \ell_Y}{\partial c} + u_1 \frac{\partial \ell_Y}{\partial k} \tag{F.55}$$

where

$$\frac{\partial \ell_Y}{\partial W} = -ck + k y_j^k e^{-kW_{Y,j}} \tag{F.56}$$

$$\frac{\partial \ell_Y}{\partial c} = -\digamma(c) + k(\log y_j - W_{Y,j}) \tag{F.57}$$

and

$$\frac{\partial \ell_Y}{\partial k} = \frac{1}{k} + c(\log y_j - W_{Y,j}) - \log(y_j e^{-W_{Y,j}}) y_j^k e^{-kW_{Y,j}} \tag{F.58}$$

The vector u_1 is defined as a vector of 0's except the last element, which is a 1; and the vector u_2 is defined as a vector of 0's except the second last element, which is a 1.

And the second order derivatives of the log-likelihood with respect to δ is

$$\begin{aligned}\frac{\partial^2 \ell_Y}{\partial \delta^2} = & \frac{\partial \ell_Y}{\partial W}\frac{\partial^2 W}{\partial \delta^2} + \frac{\partial^2 \ell_Y}{\partial W^2}\frac{\partial W}{\partial \delta}\left(\frac{\partial W}{\partial \delta}\right)^T \\ & + u_2 \frac{\partial^2 \ell_Y}{\partial W \partial c}\frac{\partial W}{\partial \delta} + \left(\frac{\partial W}{\partial \delta}\right)^T \frac{\partial^2 \ell_Y}{\partial W \partial c} u_2^T + u_2 u_2^T \frac{\partial^2 \ell_Y}{\partial c^2} \\ & + u_1 \frac{\partial^2 \ell_Y}{\partial W \partial k}\frac{\partial W}{\partial \delta} + \left(\frac{\partial W}{\partial \delta}\right)^T \frac{\partial^2 \ell_Y}{\partial W \partial k} u_1^T + u_1 u_1^T \frac{\partial^2 \ell_Y}{\partial k^2} \\ & + u_1 u_2^T \frac{\partial^2 \ell_Y}{\partial c \partial k} + u_2 u_1^T \frac{\partial^2 \ell_Y}{\partial c \partial k}\end{aligned} \tag{F.59}$$

where

$$\frac{\partial^2 \ell_Y}{\partial W^2} = -k^2 y_j^k e^{-kW_{Y,j}} \tag{F.60}$$

$$\frac{\partial^2 \ell_Y}{\partial W \partial c} = -k \tag{F.61}$$

$$\frac{\partial^2 \ell_Y}{\partial W \partial k} = -c + y_j^k e^{-kW_{Y,j}} + k \log(y_j e^{-W_{Y,j}}) y_j^k e^{-kW_{Y,j}} \tag{F.62}$$

$$\frac{\partial^2 \ell_Y}{\partial c^2} = -\mathrm{trigamma}(c) \tag{F.63}$$

$$\frac{\partial^2 \ell_Y}{\partial k^2} = -\frac{1}{k^2} - \log^2(y_j e^{-W_{Y,j}}) y_j^k e^{-kW_{Y,j}} \tag{F.64}$$

and

$$\frac{\partial^2 \ell_Y}{\partial c \partial k} = \log y_j - W_{Y,j} \tag{F.65}$$

Similarly to the gamma distribution derivations, the calculations of $\frac{\partial W}{\partial \delta}$ and

$\frac{\partial^2 W}{\partial \delta^2}$ are detailed in Davis et al. [2003]. The derivatives of the Pearson residuals e_j's against the δ are also required here. To simplify notation, the subscript j has also been dropped here.

$$
\begin{aligned}
e &= \frac{Y - \frac{\lambda\Gamma(c+1/k)}{\Gamma(c)}}{\sqrt{\frac{\lambda^2\Gamma(c+2/k)}{\Gamma(c)} - \frac{\lambda^2\Gamma^2(c+1/k)}{\Gamma^2(c)}}} \\
&= \frac{Y\Gamma(c) - \lambda\Gamma(c+1/k)}{\sqrt{\lambda^2\Gamma(c)\Gamma(c+2/k) - \lambda^2\Gamma(c+1/k)\Gamma(c+1/k)}} \\
&= (Y\Gamma(c)e^{-W} - \Gamma(c+1/k))A
\end{aligned} \tag{F.66}
$$

where $A = (\Gamma(c)\Gamma(c+2/k) - \Gamma(c+1/k)\Gamma(c+1/k))^{-\frac{1}{2}}$.

Unlike the gamma distribution, the derivatives of the Pearson residuals proved to be too difficult to derive. Due to this, we have proceeded to calculate the derivatives in R using numerical derivatives, namely, the numDeriv package.

Appendix G

R Code

G.1 Gamma Response Modelling Code

```
glarma.ll.gamma.claim.off <- function(Y, X, delta, r, phi.lags, theta
    .lags, offset = 0)
{
  #define parameters and variables
    n <- length(Y)
  p <- length(phi.lags)
  q <- length(theta.lags)
  s <- r + p + q + 1
  beta <- delta[1:r]
  phi <- delta[(r + 1):(r + p)]
  theta <- delta[(r + p + 1):(r + p + q)]
  c <- delta[s]
  u <- c(rep(0, s - 1), 1)
  mpq <- 0

  if((p + q) > 0) {
    mpq <- max(phi.lags[p], theta.lags[q])
  }

  nmpq <- n + mpq
```

```
  e <- array(0, nmpq)
  Z <- array(0, nmpq)
  W <- array(0, nmpq)
  lambda <- array(0, nmpq)
  e.d <- array(0, c(s, nmpq))
  Z.d <- array(0, c(s, nmpq))
  W.d <- array(0, c(s, nmpq))
  e.dd <- array(0, c(s, s, nmpq))
  Z.dd <- array(0, c(s, s, nmpq))
  W.dd <- array(0, c(s, s, nmpq))
  eta <- X %*% beta + offset

  ll <- 0
  ll.d <- matrix(0, ncol = 1, nrow = s)
  ll.dd <- matrix(0, ncol = s, nrow = s)

    #set up GLARMA structure
  for(time in 1:n) {
    tmpq <- time + mpq
    if(p > 0) {
      Z.d[(r + 1):(r + p), tmpq] <- Z[tmpq - phi.lags] + e[tmpq - phi.lags]
      Z.dd[(r + 1):(r + p), , tmpq] <- t((Z.d + e.d)[, (tmpq - phi.lags)])
      Z.dd[, (r + 1):(r + p), tmpq] <- Z.dd[, (r + 1):(r + p), tmpq] + (Z.d + e.d)[, (tmpq - phi.lags)]
      for(i in 1:p) {
        Z[tmpq] <- Z[tmpq] + phi[i] * (Z + e)[tmpq - phi.lags[i]]
        Z.d[, tmpq] <- Z.d[, tmpq] + phi[i] * (Z.d[, tmpq - phi.lags[i]] + e.d[, tmpq - phi.lags[i]])
        Z.dd[, , tmpq] <- Z.dd[, , tmpq] + phi[i] * (Z.dd[, , tmpq - phi.lags[i]] + e.dd[, , tmpq - phi.lags[i]])
      }
    }
```

```

    if(q > 0) {
      Z.d[(r + p + 1):(r + p + q), tmpq] <- e[tmpq - theta.lags]
      Z.dd[(r + p + 1):(r + p + q), , tmpq] <- Z.dd[(r + p + 1):(r + p + q), ,tmpq] + t(e.d[, tmpq - theta.lags])
      Z.dd[, (r + p + 1):(r + p + q), tmpq] <- Z.dd[, (r + p + 1):(r + p + q), tmpq] + e.d[, tmpq - theta.lags]
      for(i in 1:q) {
        Z[tmpq] <- Z[tmpq] + theta[i] * e[tmpq - theta.lags[i]]
        Z.d[, tmpq] <- Z.d[, tmpq] + theta[i] * e.d[, tmpq - theta.lags[i]]
        Z.dd[, , tmpq] <- Z.dd[, , tmpq] + theta[i] * e.dd[, , tmpq - theta.lags[i]]
      }
    }

   W[tmpq] <- eta[time] + Z[tmpq]
   W.d[, tmpq] <- matrix(c(X[time, ], rep(0, p + q + 1)), ncol = 1) + Z.d[, tmpq]
   W.dd[, , tmpq] <- Z.dd[, , tmpq]

   lambda[tmpq] <- exp(W[tmpq])
   lambdat<-lambda[tmpq]
   Yt<-Y[time]

     e.W<- -Yt * c^(-1/2) * lambdat^(-1)
   e.c<- -1/2 * Yt * c^(-3/2) * lambdat^(-1) - 1/2 * c^(-1/2)
   e.WW<- Yt * c^(-1/2) * lambdat^(-1)
   e.cW<- 1/2 * Yt * c^(-3/2) * lambdat^(-1)
   e.cc<- 3/4 * Yt * c^(-5/2) * lambdat^(-1) + 1/4 * c^(-3/2)

   e[tmpq]<- Yt*c^(-1/2)*lambdat^(-1) - c^(1/2)
   e.d[, tmpq]    <-e.W*W.d[,tmpq]+ e.c*u
```

```
      e.dd[,  , tmpq] <-  (e.W* W.dd[,  , tmpq]+ e.WW* W.d[, tmpq] %o
         % W.d[, tmpq]+e.cW*(W.d[,tmpq]%o%u+u%o%W.d[,tmpq])+e.cc*u%o%
         u)

 #update likelihood and derivatives.

   if(time > 1) {
 ll <- ll + -log(gamma(c)) - W[tmpq] + (c-1)*(log(Yt)-W[tmpq])-Yt*
     lambdat^(-1)
 ll.W<- -c + Yt * lambdat^(-1)
 ll.c<- -digamma(c) + (log(Yt) - W[tmpq])
 ll.WW <- -Yt * lambdat^(-1)
   ll.cc <- - trigamma(c)
   ll.cW <- -1

   ll.d <- ll.d + ll.W*W.d[, tmpq]+ ll.c*u

   ll.dd <- ll.dd + ( ll.W * W.dd[,  , tmpq]
        +(ll.WW )*W.d[, tmpq] %o% W.d[, tmpq]
        +(ll.cW)*(W.d[, tmpq] %o% u  + u%o%W.d[,tmpq])
              +(ll.cc * u %o% u) )
   }
 }
 list(delta = delta,ll = ll, ll.d = ll.d, ll.dd = ll.dd)
}
```

G.2 Standard Negative Binomial Response Modelling Code

```
glarma.ll.nb.claim.off <- function(Y, X, delta, r, phi.lags, theta.
    lags, offset = 0)
```

```
{
    #define parameters and variables
  n <- length(Y)
  p <- length(phi.lags)
  q <- length(theta.lags)
  s <- r + p + q + 1
  beta <- delta[1:r]
  phi <- delta[(r + 1):(r + p)]
  theta <- delta[(r + p + 1):(r + p + q)]
  alpha <- delta[s]
  u <- c(rep(0, s - 1), 1)
  mpq <- 0

  if((p + q) > 0) {
    mpq <- max(phi.lags[p], theta.lags[q])
  }

  nmpq <- n + mpq
  e <- array(0, nmpq)
  Z <- array(0, nmpq)
  W <- array(0, nmpq)
  mu <- array(0, nmpq)
  e.d <- array(0, c(s, nmpq))
  Z.d <- array(0, c(s, nmpq))
  W.d <- array(0, c(s, nmpq))
  e.dd <- array(0, c(s, s, nmpq))
  Z.dd <- array(0, c(s, s, nmpq))
  W.dd <- array(0, c(s, s, nmpq))
  eta <- X %*% beta + offset
  ll <- 0
  ll.d <- matrix(0, ncol = 1, nrow = s)
  ll.dd <- matrix(0, ncol = s, nrow = s)

    #set up GLARMA structure
```

```
  for(time in 1:n) {
    tmpq <- time + mpq
    if(p > 0) {
      Z.d[(r + 1):(r + p), tmpq] <- Z[tmpq - phi.lags] + e[tmpq - phi.lags]
      Z.dd[(r + 1):(r + p), , tmpq] <- t((Z.d + e.d)[, (tmpq - phi.lags)])
      Z.dd[, (r + 1):(r + p), tmpq] <- Z.dd[, (r + 1):(r + p), tmpq] + (Z.d + e.d)[, (tmpq - phi.lags)]
      for(i in 1:p) {
        Z[tmpq] <- Z[tmpq] + phi[i] * (Z + e)[tmpq - phi.lags[i]]
        Z.d[, tmpq] <- Z.d[, tmpq] + phi[i] * (Z.d[, tmpq - phi.lags[i]] + e.d[, tmpq - phi.lags[i]])
        Z.dd[, , tmpq] <- Z.dd[, , tmpq] + phi[i] * (Z.dd[, , tmpq - phi.lags[i]] + e.dd[, , tmpq - phi.lags[i]])
      }
    }

    if(q > 0) {
      Z.d[(r + p + 1):(r + p + q), tmpq] <- e[tmpq - theta.lags]
      Z.dd[(r + p + 1):(r + p + q), , tmpq] <- Z.dd[(r + p + 1):(r + p + q), ,tmpq] + t(e.d[, tmpq - theta.lags])
      Z.dd[, (r + p + 1):(r + p + q), tmpq] <- Z.dd[, (r + p + 1):(r + p + q), tmpq] + e.d[, tmpq - theta.lags]
      for(i in 1:q) {
        Z[tmpq] <- Z[tmpq] + theta[i] * e[tmpq - theta.lags[i]]
        Z.d[, tmpq] <- Z.d[, tmpq] + theta[i] * e.d[, tmpq - theta.lags[i]]
        Z.dd[, , tmpq] <- Z.dd[, , tmpq] + theta[i] * e.dd[, , tmpq - theta.lags[i]]
      }
  }

 W[tmpq] <- eta[time] + Z[tmpq]
```

```
W.d[, tmpq] <- matrix(c(X[time,  ], rep(0, p + q + 1)), ncol = 1) + Z.d[, tmpq]
W.dd[,  , tmpq] <- Z.dd[,  , tmpq]

mu[tmpq] <- exp(W[tmpq])
mut<-mu[tmpq]
Yt<-Y[time]
g<-mut+alpha
var<-mut+mut^2/alpha
h<-var

  e.W<-(-mut*(alpha*Yt+mut*(alpha+2*Yt)))/(2*alpha*var^(3/2))
e.a<- mut^2*(Yt-mut)/(2*alpha^2*var^(3/2))
e.WW<-(Yt * alpha^2 + mut * alpha * (2 * Yt - alpha) + 2 * mut^2 * (2 * Yt + alpha))/(4 * (mut+alpha)^2 * var^(1/2))
e.aW<-mut^3*(alpha*Yt-mut*(2*Yt+3*alpha))/(4*alpha^3*var^(5/2))
e.aa<-mut^3*(mut-Yt)*(mut+4*alpha)/(4*alpha^4*var^(5/2))

e[tmpq]<-(Yt-mut)/var^0.5
e.d[, tmpq]    <-e.W*W.d[,tmpq]+ e.a*u
e.dd[,  , tmpq] <-  (e.W* W.dd[,  , tmpq]+ e.WW* W.d[, tmpq] %o% W.d[, tmpq]+e.aW*(W.d[,tmpq]%o%u+u%o%W.d[,tmpq])+e.aa*u%o%u)

#update likelihood and derivatives.

  if(time > 1) {
ll <- ll + log(gamma(alpha+Yt)/(gamma(alpha)*gamma(Yt+1)))+alpha*log(alpha/g)+Yt*log(mut/g)
ll.W<-alpha*(Yt-mut)/(mut+alpha)
ll.a<-(digamma(alpha+Yt)-digamma(alpha)+ (mut-Yt)/(mut+alpha)+log(alpha/(mut+alpha)))
ll.d <- ll.d + ll.W*W.d[, tmpq]+ ll.a*u
ll.dd <- ll.dd + (alpha*(Yt-mut)/(mut+alpha)*W.dd[,  , tmpq]-
```

```
            +(alpha*mut*(Yt+alpha)/(mut+alpha)^2)*W.d[, tmpq]%o%W.d[,
                tmpq]
            +((Yt-mut)*mut/(mut+alpha)^2)*(W.d[, tmpq]%o%u + u%o%W.d[,
                tmpq])
            +(trigamma(alpha+Yt)-trigamma(alpha) + (Yt*alpha + mut^2)/(
                alpha*g^2))*u%o%u )
      }
  }

  list(delta = delta, ll = ll, ll.d = ll.d, ll.dd = ll.dd, mu = mu)
}
```

G.3 Negative Binomial Response with Allowance for Censoring Modelling Code

```
glarma.ll.nb.claim.off.cens <- function(Y, X, delta, r, phi.lags,
    theta.lags, offset = 0, censlag)
{
  #define parameters and variables
     n <- length(Y)
   p <- length(phi.lags)
   q <- length(theta.lags)
   s <- r + p + q + 1
   beta <- delta[1:r]
   phi <- delta[(r + 1):(r + p)]
   theta <- delta[(r + p + 1):(r + p + q)]
   alpha <- delta[s]
   u <- c(rep(0, s - 1), 1)
   mpq <- 0

   if((p + q) > 0) {
```

```
    mpq <- max(phi.lags[p], theta.lags[q])
  }

  nmpq <- n + mpq
  e <- array(0, nmpq)
  Z <- array(0, nmpq)
  W <- array(0, nmpq)
  mu <- array(0, nmpq)
  e.d <- array(0, c(s, nmpq))
  Z.d <- array(0, c(s, nmpq))
  W.d <- array(0, c(s, nmpq))
  e.dd <- array(0, c(s, s, nmpq))
  Z.dd <- array(0, c(s, s, nmpq))
  W.dd <- array(0, c(s, s, nmpq))
  eta <- X %*% beta + offset
  ll <- 0
  ll.d <- matrix(0, ncol = 1, nrow = s)
  ll.dd <- matrix(0, ncol = s, nrow = s)

    #set up GLARMA structure
  for(time in 1:n) {
    tmpq <- time + mpq
    if(p > 0) {
      Z.d[(r + 1):(r + p), tmpq] <- Z[tmpq - phi.lags] + e[tmpq - phi
          .lags]
      Z.dd[(r + 1):(r + p),   , tmpq] <- t((Z.d + e.d)[, (tmpq - phi.
          lags)])
      Z.dd[, (r + 1):(r + p), tmpq] <- Z.dd[, (r + 1):(r + p), tmpq]
         + (Z.d + e.d)[, (tmpq - phi.lags)]
      for(i in 1:p) {
        Z[tmpq] <- Z[tmpq] + phi[i] * (Z + e)[tmpq - phi.lags[i]]
        Z.d[, tmpq] <- Z.d[, tmpq] + phi[i] * (Z.d[, tmpq - phi.lags[
            i]] + e.d[, tmpq - phi.lags[i]])
```

```
        Z.dd[,  , tmpq] <- Z.dd[,  , tmpq] + phi[i] * (Z.dd[,  , tmpq - phi.lags[i]] + e.dd[,  , tmpq - phi.lags[i]])
      }
    }

    if(q > 0) {
      Z.d[(r + p + 1):(r + p + q), tmpq] <- e[tmpq - theta.lags]
      Z.dd[(r + p + 1):(r + p + q),  , tmpq] <- Z.dd[(r + p + 1):(r + p + q),  ,tmpq] + t(e.d[, tmpq - theta.lags])
      Z.dd[, (r + p + 1):(r + p + q), tmpq] <- Z.dd[, (r + p + 1):(r + p + q), tmpq] + e.d[, tmpq - theta.lags]
      for(i in 1:q) {
        Z[tmpq] <- Z[tmpq] + theta[i] * e[tmpq - theta.lags[i]]
        Z.d[, tmpq] <- Z.d[, tmpq] + theta[i] * e.d[, tmpq - theta.lags[i]]
        Z.dd[,  , tmpq] <- Z.dd[,  , tmpq] + theta[i] * e.dd[,  , tmpq - theta.lags[i]]
      }
  }

 W[tmpq] <- eta[time] + Z[tmpq]
 W.d[, tmpq] <- matrix(c(X[time,  ], rep(0, p + q + 1)), ncol = 1) + Z.d[, tmpq]
 W.dd[,  , tmpq] <- Z.dd[,  , tmpq]

 mu[tmpq] <- exp(W[tmpq])
 mut<-mu[tmpq]
 Yt<-Y[time]
 g<-mut+alpha
 var<-mut+mut^2/alpha
 h<-var

   e.W<-(-mut*(alpha*Yt+mut*(alpha+2*Yt)))/(2*alpha*var^(3/2))
 e.a<- mut^2*(Yt-mut)/(2*alpha^2*var^(3/2))
```

```
  e.WW<-(Yt * alpha^2 + mut * alpha * (2 * Yt - alpha) + 2 * mut^2 *
      (2 * Yt + alpha))/(4 * (mut+alpha)^2 * var^(1/2))
  e.aW<-mut^3*(alpha*Yt-mut*(2*Yt+3*alpha))/(4*alpha^3*var^(5/2))
  e.aa<-mut^3*(mut-Yt)*(mut+4*alpha)/(4*alpha^4*var^(5/2))

  e[tmpq]<-(Yt-mut)/var^0.5
  e.d[, tmpq]    <-e.W*W.d[,tmpq]+ e.a*u
  e.dd[,  , tmpq] <-  (e.W* W.dd[,  , tmpq]+ e.WW* W.d[, tmpq] %o% W.
      d[, tmpq]+e.aW*(W.d[,tmpq]%o%u+u%o%W.d[,tmpq])+e.aa*u%o%u)



 #update likelihood and derivatives.

   if(time > 1) {

   #if not last obs then carry on as normal
   if(time < n){
  ll <- ll + log(gamma(alpha+Yt)/(gamma(alpha)*gamma(Yt+1)))+alpha*
      log(alpha/g)+Yt*log(mut/g)
  ll.W<-alpha*(Yt-mut)/(mut+alpha)
  ll.a<-(digamma(alpha+Yt)-digamma(alpha)+ (mut-Yt)/(mut+alpha)+log(
      alpha/(mut+alpha)))
  ll.d <- ll.d + ll.W*W.d[, tmpq]+ ll.a*u
  ll.dd <- ll.dd + (alpha*(Yt-mut)/(mut+alpha)*W.dd[,  , tmpq]-
         +(alpha*mut*(Yt+alpha)/(mut+alpha)^2)*W.d[, tmpq]%o%W.d[,
             tmpq]
         +((Yt-mut)*mut/(mut+alpha)^2)*(W.d[, tmpq]%o%u + u%o%W.d[,
             tmpq])
         +(trigamma(alpha+Yt)-trigamma(alpha) + (Yt*alpha + mut^2)/(
             alpha*g^2))*u%o%u  )
  }

   #if last obs then do censoring calcs, which calculates
       probability of not having had a revision since previous
```

```
    revision to censoring time
else{
cens.ll <- 0
cens.ll.d <- matrix(0, ncol = 1, nrow = s)
cens.ll.dd <- matrix(0, ncol = s, nrow = s)

A <- 1
for(k in 0:censlag){ A <- A - (gamma(alpha+k)/(gamma(alpha)*
    gamma(k+1)))* exp(alpha*log(alpha/g)+k*log(mut/g)) }

cens.ll <- log(A)
temp.logA.W <- 0
temp.logA.a <- 0
temp.logA.WW <- 0
temp.logA.aa <- 0
temp.logA.Wa <- 0

for(k in 0:censlag){
    k.ll <- log(gamma(alpha+k)/(gamma(alpha)*gamma(k+1)))+alpha*
        log(alpha/g)+k*log(mut/g)
    k.ll.W<-alpha*(k-mut)/(mut+alpha)
    k.ll.a<-(digamma(alpha+k)-digamma(alpha)+ (mut-k)/(mut+alpha)
        +log(alpha/(mut+alpha)))
    k.ll.d <- k.ll.W*W.d[, tmpq]+ k.ll.a*u
    k.ll.WW <- (alpha*mut*(k+alpha)/(mut+alpha)^2)
    k.ll.Wa <- (k-mut)*mut/(mut+alpha)^2
    k.ll.aa <- (trigamma(alpha+k)-trigamma(alpha) + (k*alpha +
        mut^2)/(alpha*g^2))
    k.ll.dd <- (k.ll.W * W.dd[, , tmpq] + k.ll.WW*W.d[, tmpq]%o%
        W.d[, tmpq]
                    +(k.ll.Wa)*(W.d[, tmpq]%o%u + u%o%W.d[,tmpq])
                    +(k.ll.aa)*u%o%u )

    temp.logA.W <- temp.logA.W + (-exp(k.ll)*k.ll.W)
```

```
        temp.logA.a <- temp.logA.a + (-exp(k.ll)*k.ll.a)
        temp.logA.WW <- temp.logA.WW + (-exp(k.ll)*k.ll.WW)
        temp.logA.aa <- temp.logA.aa + (-exp(k.ll)*k.ll.aa)
        temp.logA.Wa <- temp.logA.Wa + (-exp(k.ll)*k.ll.Wa)
    }

    cens.ll.d <- 1/A * temp.logA.W *W.d[, tmpq]+ 1/A * temp.logA.a *u
    cens.ll.dd <- (1/A * temp.logA.W) * W.dd[, , tmpq] + (-1/A^2 *(temp.logA.W)^2 + 1/A * temp.logA.WW ) *W.d[, tmpq]%o%W.d[, tmpq]
                +(-1/A^2 * temp.logA.W * temp.logA.a + 1/A* temp.logA.Wa )*(W.d[, tmpq]%o%u + u%o%W.d[,tmpq])
                +(-1/A^2 * (temp.logA.a)^2 + 1/A*temp.logA.aa)*u%o%u

    ll <- ll + cens.ll
    ll.d <- ll.d + cens.ll.d
    ll.dd <- ll.dd + cens.ll.dd
    }
    }
  }
  list(delta = delta, ll = ll, ll.d = ll.d, ll.dd = ll.dd, mu = mu)
}
```

G.4 Binary Response Modelling Code

```
glarma.ll.bin.claim.off<-function(Y, X, delta, r, phi.lags, theta.lags, offset = 0)
{
  #define parameters and variables
     n <- length(Y)
   p <- length(phi.lags)
```

```
  q <- length(theta.lags)
  s <- r + p + q
  beta <- delta[1:r]
  phi <- delta[(r + 1):(r + p)]
  theta <- delta[(r + p + 1):(r + p + q)]
  mpq <- 0

    if((p + q) > 0) {
    mpq <- max(phi.lags[p], theta.lags[q])
  }

  nmpq <- n + mpq
  e <- array(0, nmpq)
  Z <- array(0, nmpq)
  W <- array(0, nmpq)
  pt <- array(0, nmpq)
  e.d <- array(0, c(s, nmpq))
  Z.d <- array(0, c(s, nmpq))
  W.d <- array(0, c(s, nmpq))
  e.dd <- array(0, c(s, s, nmpq))
  Z.dd <- array(0, c(s, s, nmpq))
  W.dd <- array(0, c(s, s, nmpq))
  eta <- X %*% beta
  ll <- 0
  ll.d <- matrix(0, ncol = 1, nrow = s)
  ll.dd <- matrix(0, ncol = s, nrow = s)

  #set up GLARMA structure
    for(time in 1:n) {
    tmpq <- time + mpq
    if(p > 0) {
      Z.d[(r + 1):(r + p), tmpq] <- Z[tmpq - phi.lags] + e[tmpq - phi
          .lags]
```

```
      Z.dd[(r + 1):(r + p), , tmpq] <- t((Z.d + e.d)[, (tmpq - phi.lags)])
      Z.dd[, (r + 1):(r + p), tmpq] <- Z.dd[, (r + 1):(r + p), tmpq] + (Z.d + e.d)[, (tmpq - phi.lags)]
      for(i in 1:p) {
        Z[tmpq] <- Z[tmpq] + phi[i] * (Z + e)[tmpq - phi.lags[i]]
        Z.d[, tmpq] <- Z.d[, tmpq] + phi[i] * (Z.d[, tmpq - phi.lags[i]] + e.d[, tmpq - phi.lags[i]])
        Z.dd[, , tmpq] <- Z.dd[, , tmpq] + phi[i] * (Z.dd[, ,tmpq - phi.lags[i]] + e.dd[, , tmpq - phi.lags[i]])
      }
    }
    if(q > 0) {
      Z.d[(r + p + 1):(r + p + q), tmpq] <- e[tmpq - theta.lags]
      Z.dd[(r + p + 1):(r + p + q), , tmpq] <- Z.dd[(r + p + 1):(r + p + q), ,tmpq] + t(e.d[, tmpq - theta.lags])
      Z.dd[, (r + p + 1):(r + p + q), tmpq] <- Z.dd[, (r + p + 1):(r + p + q), tmpq] + e.d[, tmpq - theta.lags]
      for(i in 1:q) {
        Z[tmpq] <- Z[tmpq] + theta[i] * e[tmpq - theta.lags[i]]
        Z.d[, tmpq] <- Z.d[, tmpq] + theta[i] * e.d[, tmpq - theta.lags[i]]
        Z.dd[, , tmpq] <- Z.dd[, , tmpq] + theta[i] * e.dd[, ,tmpq - theta.lags[i]]
      }
    }
    W[tmpq] <- eta[time] + Z[tmpq]
    #print(eta[time])
    #print(Z[tmpq])
    W.d[, tmpq] <- matrix(c(X[time, ], rep(0, p + q)), ncol = 1) + Z.d[, tmpq]
    W.dd[, , tmpq] <- Z.dd[, , tmpq]

      pt[tmpq] <- exp(W[tmpq])/(1 + exp(W[tmpq]))
```

```

      e[tmpq] <- (Y[time] - pt[tmpq])/sqrt(pt[tmpq]*(1-pt[tmpq]))
      e.d[,tmpq]<-((Y[time]-1)*exp(0.5*W[tmpq])/2 - Y[time]*exp((-0.5)*
          W[tmpq])/2)* W.d[, tmpq]
      e.dd[, ,tmpq]<-((Y[time]-1)*exp(0.5*W[tmpq])/4 + Y[time]*exp
          (-0.5*W[tmpq])/4)*W.d[,tmpq]%o%W.d[,tmpq]+
              ((Y[time]-1)*exp(0.5*W[tmpq])/2- Y[time]*exp(-0.5*W[tmpq])/
                  2)*W.dd[, , tmpq]

       #update likelihood and derivatives.

       if(time > 1) {
         ll <- ll + Y[time] * W[tmpq] - log(1+exp(W[tmpq]))
         ll.d <- ll.d + (Y[time]-pt[tmpq])*W.d[, tmpq]
         ll.dd <- ll.dd+(Y[time]-pt[tmpq])*W.dd[, ,tmpq]- pt[tmpq]*(1-pt[
             tmpq])* W.d[,tmpq]%o%W.d[,tmpq]
     }
   }
   list(delta = delta, ll = ll, ll.d = ll.d, ll.dd = ll.dd)
}
```

G.5 Example of a Wrapper Program

```
library(MASS)
source("glarma.ll.gamma.claim.off.excl.first.r")

data01 <- read.table("size09_full.csv",sep = ",", header=T)

k <- length(data01)
n <- length(data01[,14])
ones <- rep(1,n)
findn <- data01[,140]
```

```
data<-as.matrix(cbind(data01[,139],data01[,18],data01[,17],ones,findn
    ,data01[,c(
  5,6,7,9,10,12,13,14,16,20,24,27,28,29,30,31,32,33,34,35,36,37,41:45,47:54,55:62,67,69

  80:82,84:87,89:91,93:98,100:104,105,107:111,112,114:117,119:122,126,129:138)
      ]))

dataglm<-data[data[,2]>0,]
Yglm<-dataglm[,3]
Xnum<-length(dataglm[1,])
Xglm<-dataglm[,c(4:Xnum)]

init.glm<-glm(Yglm ~ -1 + Xglm,family=Gamma(log))
summary(init.glm)
numclaims <- length(unique(data[,1]))

beta <- init.glm$coefficients
phi.lags<-c(1)
phi.init<-c(0.08)
theta.lags<-rep(0,0)
theta.init<-rep(0,0)

c <- 0.83
delta<-c(beta,phi.init,theta.init,c)

for (iter in 1:6) {

  s = length(delta)

  ll <- 0
  ll.d <- matrix(0, ncol = 1, nrow = s)
  ll.dd <- matrix(0, ncol = s, nrow = s)

```

```
  for (l in 1:numclaims) {

  leni <- length(data[data[,1]==l,3])
  if(leni > 1) {
  datai <-  matrix(data[data[,1]==l,],nrow=leni)
 Y <- datai[,3];
 X <- as.matrix(datai[,c(4:Xnum)])

  temp <- glarma.ll.gamma.claim.off(Y,X,delta,Xnum-3, phi.lags, theta
      .lags)

 #update ll, ll.d, ll.dd

  ll <- ll + temp$ll
  ll.d <- ll.d + temp$ll.d
  ll.dd <- ll.dd + temp$ll.dd

  }
  }
  cov1 <- solve( - ll.dd)
  step <- cov1 %*% ll.d
  se1 <- sqrt(diag(cov1))
  delta.old <- delta
  delta <- delta.old + step

  list(iter = iter, delta.old = delta.old, delta = delta, ll = ll, ll
      .d = ll.d, ll.dd = ll.dd, cov1 = cov1, se1 = se1)
}
print(cbind(c(init.glm$coefficients,phi.init,theta.init,c),delta,se1,
    step))
```

G.6 Example of a Wrapper Program for Random Effects

```

library(MASS)
library(statmod)

source("glarma.ll.gamma.claim.off.excl.first.r")
data01 <- read.table("\size03.csv",sep = ",", header=T)

k <- length(data01)
n <- length(data01[,14])
ones <- rep(1,n)
findn <- data01[,140]

NGQ <- 5
d <- 1
startl <- 0.2
L<-matrix(c(startl),byrow=F,ncol=d,nrow=d)
lambda <- L[L != 0]

data<-as.matrix(cbind(data01[,145],data01[,18],data01[,17],ones,findn
    ,data01[,c(9,34,41:45,141,143,80,144)]))

dataglm<-data[data[,2]>0,]
Yglm<-dataglm[,3]
Xnum<-length(dataglm[1,])
Xglm<-dataglm[,c(4:Xnum)]

init.glm<-glm(Yglm ~ -1 + Xglm,family=Gamma(log))
summary(init.glm)

numclaims <- length(unique(data[,1]))

```

```

beta <- init.glm$coefficients
numbeta <- length(beta)

phi.lags<-c(1)
phi.init<-c(0.05)
theta.lags<-rep(0,0)
theta.init<-rep(0,0)

c <- 0.79
delta<-c(beta,lambda,phi.init,theta.init,c)
s = length(delta)

for (iter in 1:6) {

  #initialise ll, ll.d, ll.dd
  c<-delta[s]
  beta<-delta[1:numbeta]
  lambda <- delta[(numbeta+1):(numbeta+1)]
  numphi <- length(phi.lags)
  phi <- delta[(numbeta+2):(numbeta+1+numphi)]
  numtheta <- length(theta.lags)
  theta <- delta[(numbeta+2+numphi):(numbeta+1+numphi + numtheta) ]


  #initialise iteration vectors

  Tl.RE.Lap <- 0
  Tl.RE.AGQ <- 0
  Tl.d<-matrix(0, ncol = 1, nrow = s)
  Tl.dd<-matrix(0,s,s)


  for (l in 1:numclaims) {
```

```
#print(l)
  leni <- length(data[data[,1]==l,3])

if(leni > 1) {


 #initialise claim j vectors

 integrandLj<-rep(NA,NGQ)


 datai <-  matrix(data[data[,1]==l,],nrow=leni)
  Y <- datai[,3];
  X <- as.matrix(datai[,c(4:Xnum)])

 #set up laplace data
 offset  = X %*% beta

 R <- X[,1]
 RL <-R%*%L

 Zstar <- 0

 for (it in 1:10) {

   delta.lap<-c(Zstar,phi,c)

   temp <- glarma.ll.gamma.claim.off(Y,RL,delta.lap,1, phi.lags, theta.lags,offset)

   Fz <- temp$ll - 1/2*log(2 * pi) - Zstar^2 /2
   dFz<-temp$ll.d[1:d]-Zstar
   d2Fz<-temp$ll.dd[1:d,1:d]-diag(rep(1,d))
   Sigmastar<- solve(-d2Fz)
```

```
    Zstar<-Zstar+Sigmastar%*%dFz

  }

  Tl.RE.Lap <- Tl.RE.Lap + log(sqrt(2 * pi /-d2Fz)*exp(Fz))

  # AGQ

  Q<-chol(Sigmastar)
  detQ<-det(Q)

  # Gauss Quad points and weights for 1 dimension.

  GQ<-gauss.quad(NGQ,kind="hermite")
  zeta<-GQ$nodes
  w<-GQ$weights

  lj<-0
  Lj<-0

  lj.d<-matrix(0, ncol = 1, nrow = s)
  lj.dd<-matrix(0,s,s)

  ndata<-length(Y)

  zetaarray<-matrix(NA,ncol=NGQ^d,nrow=d)
  warray<-matrix(NA,ncol=NGQ^d,nrow=d)

  for (j in 1:d){
    zetaarray[j,]<-rep(rep(zeta,each=NGQ^(d-j)),NGQ^(j-1))
    warray[j,]<-rep(rep(w,each=NGQ^(d-j)),NGQ^(j-1))

  }

```

```
for (i in 1:NGQ^d){

  # Generate the z grid and weights from GQ values

  ztilde <- Zstar + 2^0.5*t(Q)%*%zetaarray[,i]
  weight <- detQ/pi^(d/2)*exp(sum(zetaarray[,i]^2))*prod(warray[,i
      ])

  # correlated random effects New Way

  ZR<-kronecker(t(ztilde),R)
  XallZR<-cbind(X,ZR)

  r = numbeta + 1
  delta.agq <- c(beta,lambda,phi,c)
  z.glimarma<-glarma.ll.gamma.claim.off(Y,XallZR,delta.agq,r,phi.
      lags,theta.lags,offs=0)

  lz<-z.glimarma$ll-sum(ztilde^2)/2
  lz.d<-z.glimarma$ll.d
  lz.dd<-z.glimarma$ll.dd

  integrandLj[i]<-exp(lz)*weight
  Lj<-Lj+integrandLj[i]
  lj.d<-lj.d+lz.d*integrandLj[i]
  lj.dd<-lj.dd+lz.dd*integrandLj[i]+lz.d%*%t(lz.d)*integrandLj[i]

}

lj <- log(Lj)
lj.d<-(1/Lj)*lj.d
lj.dd<-(1/Lj)*lj.dd - lj.d%*%t(lj.d)

T1.RE.AGQ <- T1.RE.AGQ + lj
```

```
    Tl.d<-Tl.d + lj.d
    Tl.dd<-Tl.dd + lj.dd

  }

  }

  cov1 <- solve( - Tl.dd)
  step <- cov1 %*% Tl.d
  se1 <- sqrt(diag(cov1))
  delta.old <- delta
  delta <- delta.old + step
  if(iter > 3) { delta[s-2]<-max(delta[s-2],0.0001)}
  else {delta[s-2]<-startl }
  delta[s]<-max(delta[s],0.01)

}

print(cbind(c(init.glm$coefficients,0.1,phi.init,theta.init,1),delta,
    se1,step))
Tl.RE.AGQ
```

G.7 Fitting generalised gamma using numerical methods

```
glarma.llonly.ggamma.claim.off <- function(Y, X, delta, r, phi.lags,
    theta.lags, offset = 0)
{
  n <- length(Y)
  p <- length(phi.lags)
  q <- length(theta.lags)
```

```
s <- r + p + q + 2
beta <- delta[1:r]
phi <- delta[(r + 1):(r + p)]
theta <- delta[(r + p + 1):(r + p + q)]
c <- delta[s-1]
  k <- delta[s]
mpq <- 0

if((p + q) > 0) {
  mpq <- max(phi.lags[p], theta.lags[q])
}

nmpq <- n + mpq
e <- array(0, nmpq)
Z <- array(0, nmpq)
W <- array(0, nmpq)
lambda <- array(0, nmpq)
e.d <- array(0, c(s, nmpq))
Z.d <- array(0, c(s, nmpq))
W.d <- array(0, c(s, nmpq))
e.dd <- array(0, c(s, s, nmpq))
Z.dd <- array(0, c(s, s, nmpq))
W.dd <- array(0, c(s, s, nmpq))
eta <- X %*% beta + offset

ll <- 0
ll.d <- matrix(0, ncol = 1, nrow = s)
ll.dd <- matrix(0, ncol = s, nrow = s)

for(time in 1:n) {
  tmpq <- time + mpq
  if(p > 0) {
      Z.d[(r + 1):(r + p), tmpq] <- Z[tmpq - phi.lags] + e[tmpq - phi.lags]
```

```
      Z.dd[(r + 1):(r + p), , tmpq] <- t((Z.d + e.d)[, (tmpq - phi.lags)])
      Z.dd[, (r + 1):(r + p), tmpq] <- Z.dd[, (r + 1):(r + p), tmpq] + (Z.d + e.d)[, (tmpq - phi.lags)]

      for(i in 1:p) {
          Z[tmpq] <- Z[tmpq] + phi[i] * (Z + e)[tmpq - phi.lags[i]]
      Z.d[, tmpq] <- Z.d[, tmpq] + phi[i] * (Z.d[, tmpq - phi.lags[i]] + e.d[, tmpq - phi.lags[i]])
      Z.dd[, , tmpq] <- Z.dd[, , tmpq] + phi[i] * (Z.dd[, , tmpq - phi.lags[i]] + e.dd[, , tmpq - phi.lags[i]])
      }
    }

    if(q > 0) {
      Z.d[(r + p + 1):(r + p + q), tmpq] <- e[tmpq - theta.lags]
      Z.dd[(r + p + 1):(r + p + q), , tmpq] <- Z.dd[(r + p + 1):(r + p + q), ,tmpq] + t(e.d[, tmpq - theta.lags])
      Z.dd[, (r + p + 1):(r + p + q), tmpq] <- Z.dd[, (r + p + 1):(r + p + q), tmpq] + e.d[, tmpq - theta.lags]
      for(i in 1:q) {
          Z[tmpq] <- Z[tmpq] + theta[i] * e[tmpq - theta.lags[i]]
      Z.d[, tmpq] <- Z.d[, tmpq] + theta[i] * e.d[, tmpq - theta.lags[i]]
      Z.dd[, , tmpq] <- Z.dd[, , tmpq] + theta[i] * e.dd[, , tmpq - theta.lags[i]]
      }
    }

    W[tmpq] <- eta[time] + Z[tmpq]
    W.d[, tmpq] <- matrix(c(X[time, ], rep(0, p + q + 1)), ncol = 1) + Z.d[, tmpq]
    W.dd[, , tmpq] <- Z.dd[, , tmpq]

```

```
    lambda[tmpq] <- exp(W[tmpq])
    lambdat<-lambda[tmpq]
    Yt<-Y[time]
    mean <- lambdat*gamma(c+1/k)/gamma(c)
      var <- lambdat^2*( (gamma(c + 2/k)/gamma(c)) - ( gamma(c + 1/k)
          /gamma(c))^2 )
    e[tmpq]<- (Yt - mean)/var^(1/2)

    #update likelihood and derivatives.
if(time > 1) {
  ll <- ll + log(k) - lgamma(c) - W[tmpq] + (c*k - 1)*(log(Yt)-W[tmpq
      ]) - (Yt*lambdat^(-1))^k
    }
  }
  list(delta = delta, ll = ll)
}

library(MASS)
library(statmod)
library(numDeriv)
library(parallel)
library(snow)
library(snowfall)

source("glarma.ll.ggamma.claim.off.excl.first.llonly.r")
data01 <- read.table("size03.csv",sep = ",", header=T)

k <- length(data01)
n <- length(data01[,14])
ones <- rep(1,n)
findn <- data01[,140]

data<-as.matrix(cbind(data01[,145],data01[,18],data01[,17],ones,findn
    ,data01[,c(7,14,16,20,31,34,41:45,92:97,141,143,144)]))
```

```
dataglm<-data[data[,2]>0,]
Yglm<-dataglm[,3]
Xnum<-length(dataglm[1,])
Xglm<-dataglm[,c(4:Xnum)]

init.glm<-glm(Yglm ~ -1 + Xglm,family=Gamma(log))
summary(init.glm)

numclaims <- length(unique(data[,1]))
beta <- init.glm$coefficients
phi.lags<-c(1)
phi.init<-c(0.06)
theta.lags<-rep(0,0)
theta.init<-rep(0,0)
betaphi <- c(beta,phi.init)

c <- 0.97
k <- 0.89

Splitll <- function(numsplit){

LLFunc <- function(betaphi){
      delta<-c(betaphi,c,k)
      s = length(delta)

      ll <- 0
      ll.d <- matrix(0, ncol = 1, nrow = s)
      ll.dd <- matrix(0, ncol = s, nrow = s)

      for (l in (1400*(numsplit-1)+1):(min(1400*numsplit,numclaims))) {
         leni <- length(data[data[,1]==l,3])

         if(leni > 1) {
```

```
        datai <-  matrix(data[data[,1]==l,],nrow=leni)
        Y <- datai[,3];
        X <- as.matrix(datai[,c(4:Xnum)])
        temp <- glarma.llonly.ggamma.claim.off(Y,X,delta,Xnum-3, phi.lags, theta.lags)
        ll <- ll + temp$ll
      }
    }
    return(ll)
  }

  ndll <-  LLFunc(betaphi)
  return(ndll)
}

Splitlld <- function(numsplit){

  LLFunc <- function(betaphi){
    delta<-c(betaphi,c,k)
    s = length(delta)

    ll <- 0
    ll.d <- matrix(0, ncol = 1, nrow = s)
    ll.dd <- matrix(0, ncol = s, nrow = s)

    for (l in (1400*(numsplit-1)+1):(min(1400*numsplit,numclaims))) {
      leni <- length(data[data[,1]==l,3])

      if(leni > 1) {
        datai <-  matrix(data[data[,1]==l,],nrow=leni)
        Y <- datai[,3];
        X <- as.matrix(datai[,c(4:Xnum)])
        temp <- glarma.llonly.ggamma.claim.off(Y,X,delta,Xnum-3, phi.lags, theta.lags)
```

```
        ll <- ll + temp$ll
      }
    }
    return(ll)
  }
  ndll.d <- grad(LLFunc, betaphi)
  return(ndll.d)
}

Splitlldd <- function(numsplit){

  LLFunc <- function(betaphi){
    delta<-c(betaphi, c, k)
    s = length(delta)

    ll <- 0
    ll.d <- matrix(0, ncol = 1, nrow = s)
    ll.dd <- matrix(0, ncol = s, nrow = s)

    for (l in (1400*(numsplit-1)+1):(min(1400*numsplit, numclaims))) {
      leni <- length(data[data[,1]==l,3])

      if(leni > 1) {
        datai <- matrix(data[data[,1]==l,], nrow=leni)
        Y <- datai[,3];
        X <- as.matrix(datai[,c(4:Xnum)])
        temp <- glarma.llonly.ggamma.claim.off(Y,X, delta, Xnum-3, phi.
            lags, theta.lags)
        ll <- ll + temp$ll
      }
    }
    return(ll)
  }
  ndll.dd <- hessian(LLFunc, betaphi)
```

```
  return(ndll.dd)
}

for (iter in 1:3) {

  #initialise ll, ll.d, ll.dd

  sfInit(parallel=TRUE, cpus=7, type="SOCK")
  sfExportAll()
  sfLibrary(numDeriv)
  test.ll <- sfLapply(1:7, Splitll)
  test.lld <- sfLapply(1:7, Splitlld)
  test.lldd <- sfLapply(1:7, Splitlldd)
  sfStop()

  sum(unlist(test.ll))

  sumll <- 0
  sumlld <- array(0,23)
  sumlldd <- array(0,c(23,23))

  for (m in 1:7) {
    sumll <- sumll + unlist(test.ll[m])
    sumlld <- sumlld + unlist(test.lld[m])
    sumlldd <- sumlldd + as.array(matrix(unlist(test.lldd[m]),ncol
        =23,byrow=TRUE))
  }
  step<-solve(-sumlldd,sumlld)
  print(iter)
  print(cbind(c(init.glm$coefficients,phi.init,theta.init),betaphi,
      step))
  betaphi <- betaphi + step
}

```

```
cov1 <- solve( - sumlldd)
se1 <- sqrt(diag(cov1))
print(cbind(c(init.glm$coefficients ,phi.init ,theta.init),betaphi,se1,
    step))
```

Appendix H

Spatial Features

H.1 Introduction

This appendix chapter explores the spatial aspect of NSW CTP data. In the context of valuation, perhaps the location aspect of the claims are less important, as establishing the total liabilities of a portfolio is the main objective. However, for long tailed products, valuation results flow directly into the pricing of the insurance products. Hence, correctly allowing for the location aspect in a valuation framework can enhance the pricing of the product.

Pricing is the other key aspect of an actuary's work, it is establishing the premium a policyholder needs to pay for the insurance product. In essence, the price is comprised of the cost of the insurance claims, the expenses of the insurance company as well as a profit margin for the insurance company. For long tailed products, valuation work is directly linked to the calculation of the first component - the cost of the insurance claims. Since the true cost of claims for a long tailed product will not be known until years later, valuation type techniques are employed to form an estimate of the ultimate claims cost.

One of the key objectives for a pricing project is to improve the granularity of the pricing work. More granular analyses allow the development of a more refined rating structure, and that is paramount in the selection of the "better" risks and leave the "worse" risks to the competitors. Take gender as an example, the insurer can reap great rewards if it can establish females are better risks than males by a magnitude of 10%, then offering a discount of slightly lower than 10%, say 8%, in a market that the other insurers do not differentiate gender in their pricing. On the other hand, if the competitors have developed a more sophisticated rating structure then the insurer is exposed to "anti-selection" and may stand to lose considerable money until its rating structure catches up to those of the competitors.

Location is an important rating factor in most insurance products. For home and contents insurance, location is highly correlated with weather and damages due to natural perils; for motor vehicle insurance, location through demographics infers considerable information regarding the driving behaviours of the policyholder.

H.2 Data

For NSW CTP insurance, location is a prescribed rating factor. The regulator, MAA, defines 5 regions,

- Metro - the Sydney Metropolitan Area
- Country - the rural areas in the state
- Outer Metro - the area immediately outside the Metro area
- Wollongong - the regional city of Wollongong and surrounding areas
- Newcastle - the regional city of Newcastle and surrounding areas.

the MAA also prescribe the premium "relativities" for these 5 regions, that is, how much an insurer has to charge for a Metro policy relative to a Country policy. The insurer is not allowed to incorporate further location based rating factors in addition to the regions as defined by the MAA.

From the modelling of the claim development processes, region, as defined by the MAA, has not been a significant variable in most of the processes. This is perhaps unsurprising as 60% of the vehicles and 65% of the CTP claimants are in the "Metro" region and the other regions are quite small in comparison, and they cover a vast areas of space.

For the purposes of this chapter, only the Sydney metro will be analysed. Firstly, the geography and demographics are easier to discuss due to the smaller region geographically and diverse demographics contained within. Secondly, the area contains 75% of the claims cost arising from the NSW CTP Scheme. We believe investigating the spatial elements within the Metro region would be of more interest.

The PIR data uses "postcode" as the measurement for location and provides three postcodes per claim,

i) Claimant Postcode - the location of the claimant's usual residence

ii) Garaged Postcode - the location of the vehicle's registered address, can be used as a proxy for the location of the driver-at-fault

iii) Accident Postcode - the location of the accident

While postcodes generally encapsulate a large area, it should be adequate when the area under analyses is limited to the metropolitan areas where population density is relatively high and each postcode comprises of a relatively small area.

The figures below provide a general overview of the areas of Sydney and the CTP claim rates. Figure H.1 shows the number of CTP "claims" per 1000 residents by postcode over the accident years 2004 to 2008. The population data is drawn from the 2006 Australian Census (Australian Bureau of Statistics [2006a]), and the data chosen is the 5 accident years centred around 2006. The average number of claims per 1000 residents across the whole state is around 1.2 claimants per 1,000 residents p.a.. The diagram shows much of the south side of the city (south of the Parramatta River and Parramatta Road/Great Western Highway) has a higher claimant per population rate.

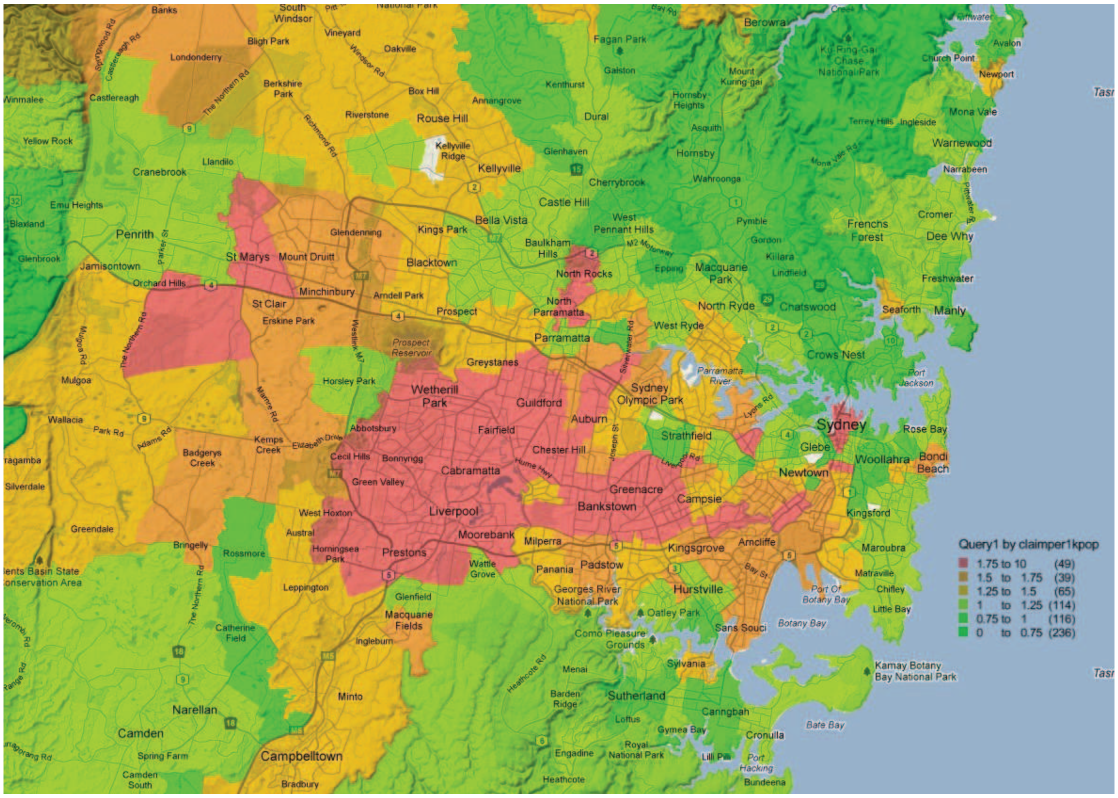

Figure H.1: Sydney Metro - Number of Claims per 1000 Population by Postcode

Figure H.2 shows the number of claims caused by 1000 vehicles garaged at each postcode. The number of vehicles data is drawn from the 2006 Vehicles Census (Australian Bureau of Statistics [2006b]) and once again is compared to claims from the accident years 2004 to 2008. The average number of claims per 1,000 garaged vehicles across the state is around 2.0. The diagram shows a similar pattern to Figure H.1, where the areas in the south reflect a higher number of claims per vehicle garaged.

Figure H.2: Sydney Metro - Number of Claims Caused by 1000 Vehicles Garaged by Postcode

Clearly, there is a clear pattern in the claim incidence in the NSW CTP scheme across the various postcodes in the Sydney Metro area. Attention is now turned to the Claims Development Process framework and the following figures (Figure H.3 to Figure H.7) map the average process variables by claimants' postcode of residence in the "Metropolitan" region as defined by the MAA (Motor Accidents Authority [2008]).

Upon examining the mapped raw averages, i.e., not accounting for covariates, the following observations are drawn.

- The area around "Richmond" in Sydney's northwest is labelled as "Outer Metro" region, creating a gap in the maps created. Other gaps include "The Pond" which is a large new development, and new postcode, near Kellyville and the University of NSW and Macquarie University.

- There does not seem to be a clear pattern for the delay between revision process variable, $t_{j,i}$.

- A clear pattern in the overall probability of finalisation ($\sum S_{j,i}/\sum j'$) by postcode. The city's South and Southwest areas seem to have a lower probability of finalisation than the other areas; in other words, a larger number of revisions is made to a claim before it finalises.

- There is also a clear pattern that is exhibited for the probability of an upward revision, ($\sum D_{j,i}/\sum j'$). The pattern is similar to that of the probability of settlement, that is, the South and Southwest stand out in having a higher probability of an upward revision. However, as observed earlier $D_{j,i}$ has a strong correlation with $S_{j,i}$ where the majority of final revision are also downward revisions. Hence, if the South and Southwest have a lower probability of finalisation, the observed final revisions are fewer and this leads to fewer downward revisions and the observed higher likelihood of an upward revision.

- The two maps that represent the average revisions sizes, Figure H.6 for the last and downward revision and Figure H.7 for the other revisions, show less clear pattern. However, there is still a definite degree of spatial correlation that is present in the maps.

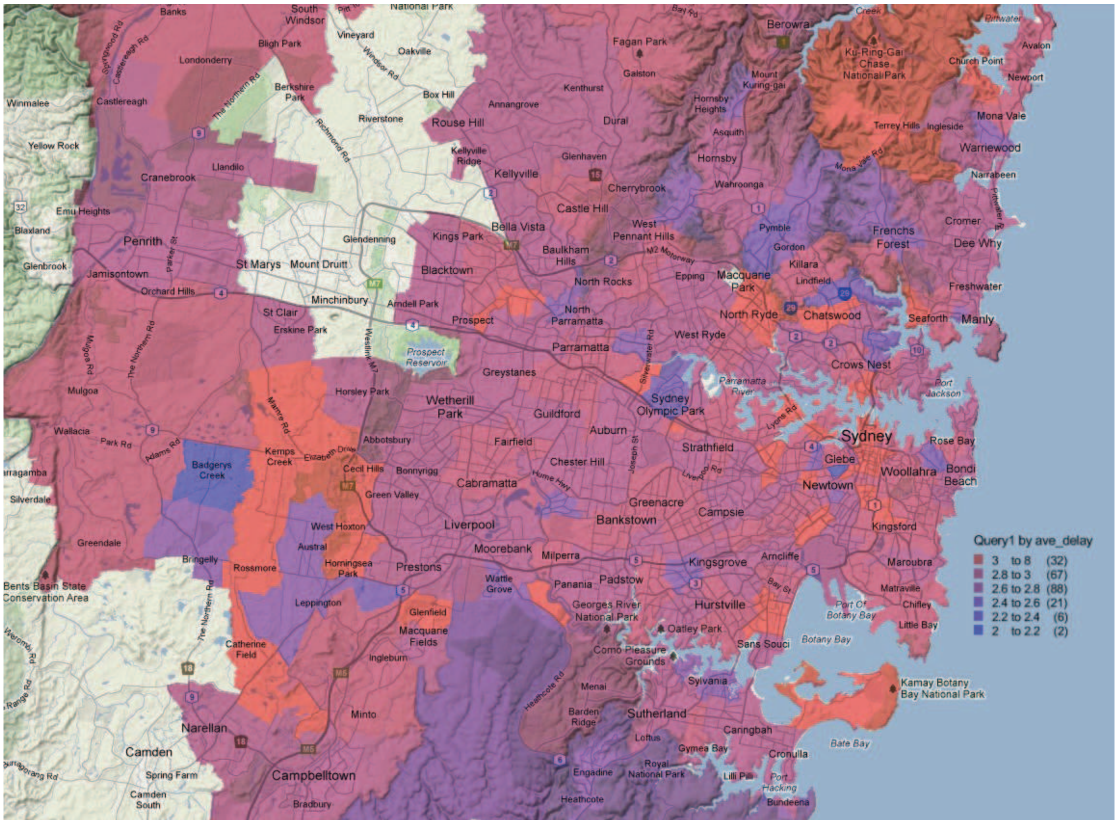

Figure H.3: Sydney Metro - Average Delay by Postcode

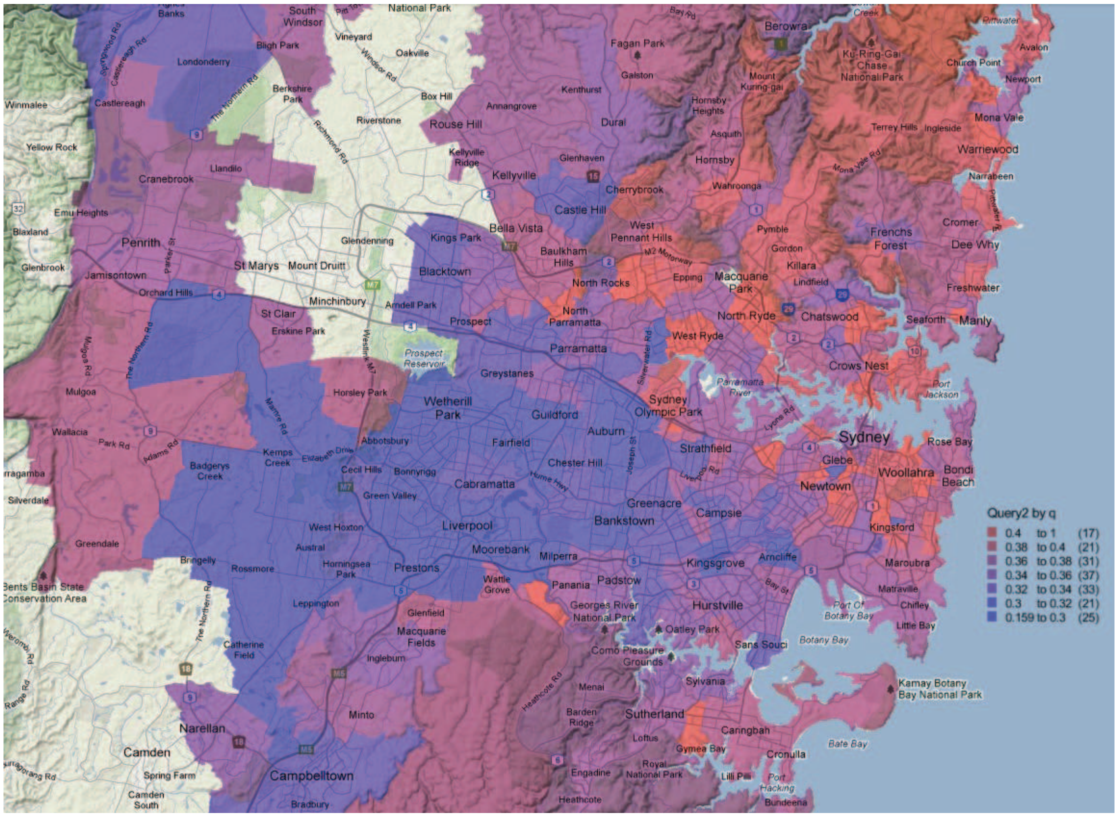

Figure H.4: Sydney Metro - Average Probability of Finalisation by Postcode

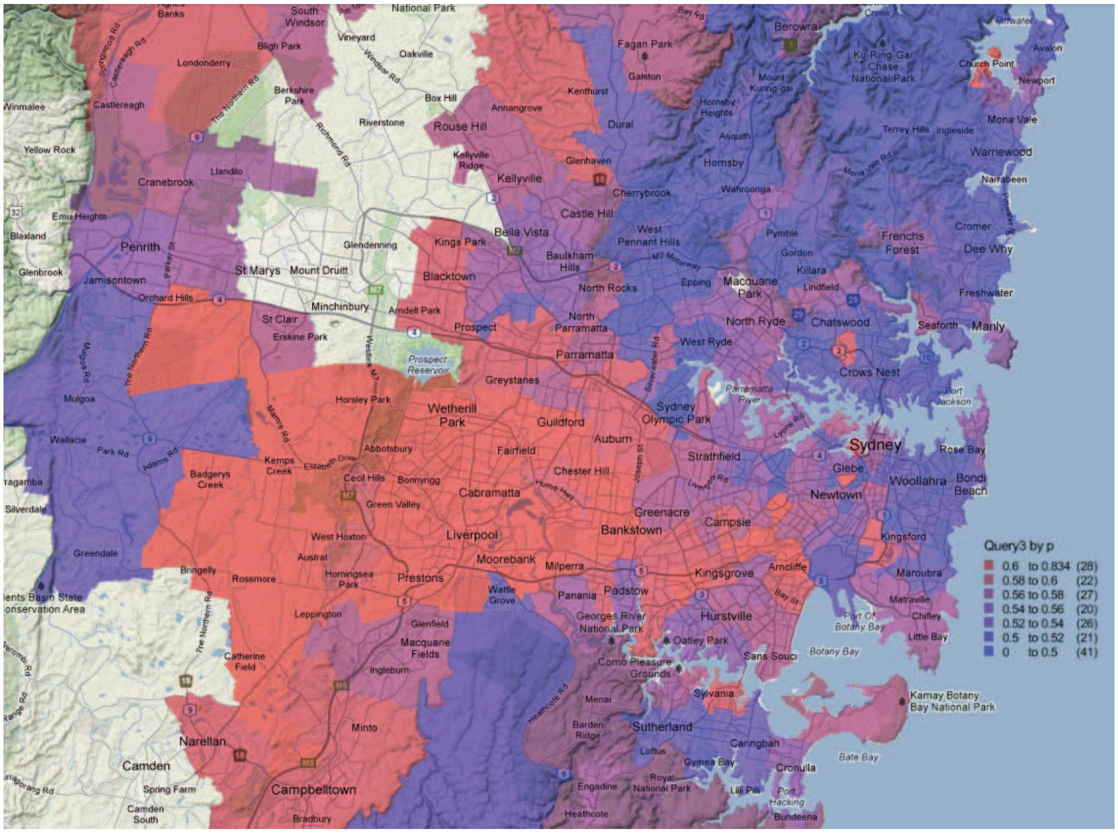

Figure H.5: Sydney Metro - Average Probability of Upward Revision by Postcode

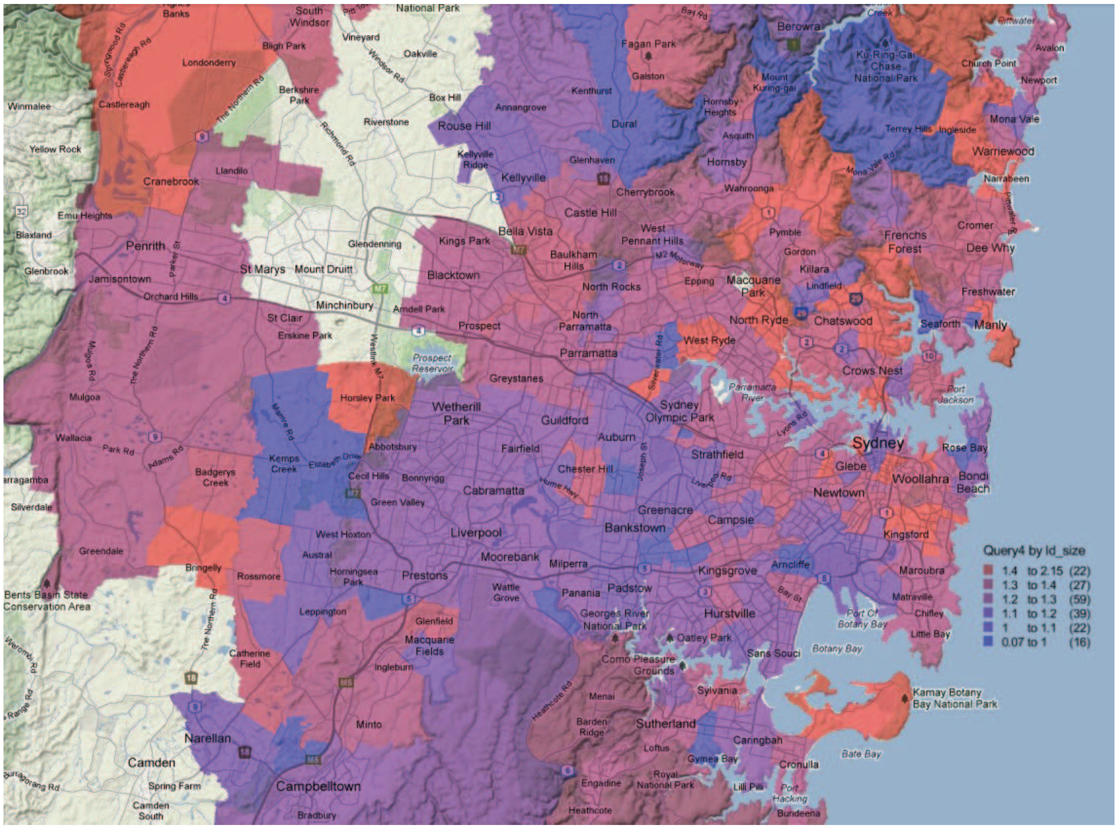

Figure H.6: Sydney Metro - Average Size of Last Downward Revision by Postcode

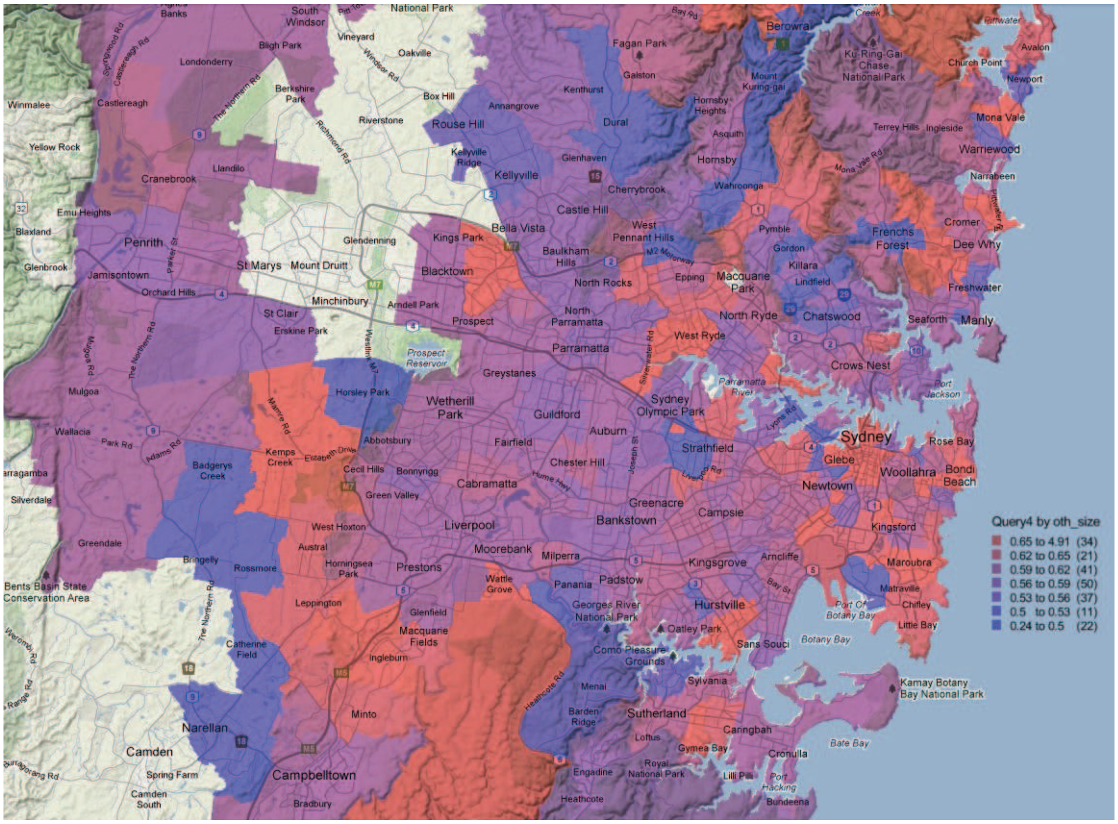

Figure H.7: Sydney Metro - Average Size of Other Revisions by Postcode

The following sections examine the location based information contained in the data. The modelling will concentrate on the settlement process as the data suggest considerable location based information is contained in the data. Note, the analyses contained in this chapter are purely for analytical and illustration purposes of the methods. The results are for academic purposes only.

H.3 Modelling Location Characteristics

The first method of allowing for location within the model framework is to use location characteristics as covariates. This does not deviate or complicate the framework developed in Chapters 6 and 7. It simply introduce more covariates that describes the characteristics of the location the claimant resides. In this section the following two characteristics are used.

i) Proportion of households that speak English only - this is used as a proxy for how familiar with the NSW CTP Scheme. As described earlier, the Australian Third Party Bodily Injury insurance product differs greatly to other parts of the world. Hence, the Census information on how many persons in an area speaks only English is used as a proxy as to how familiar is the claimant with the CTP scheme and how likely the claimant is to act proactively in settling the claim.

ii) Index of Socioeconomic Disadvantage - this variable, and other SEIFA indices, is widely used and have been significant in other insurance products in our experience. It can be a measure of income (important for the economic loss component of CTP claim), vehicle ownership rates, vehicle utilisation rates, etc.

The following two figures maps each of these variables across the Sydney Metro region. While the map shows similar pattern, the correlation between the two variables is actually quite low at -0.08. This can be loosely interpreted as a higher proportion of a postcode comprises of English speaking only families leads to a slightly smaller index of socioeconomic disadvantage (or more affluent area).

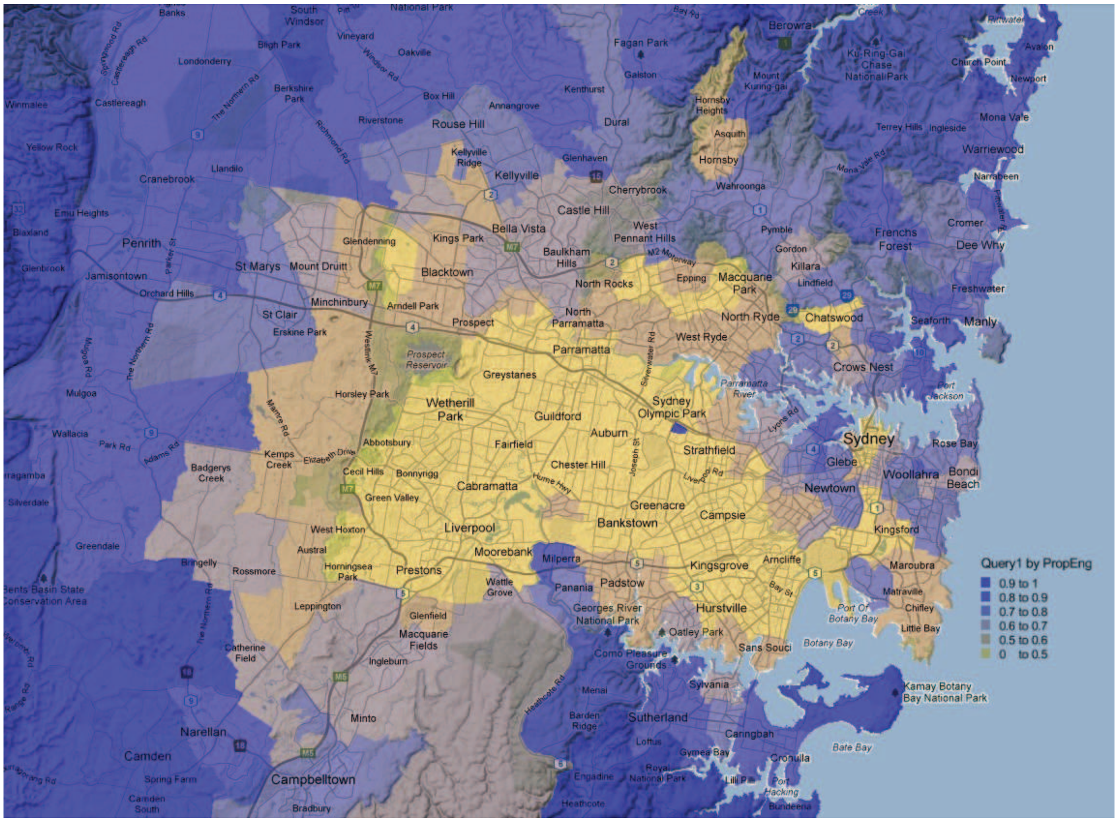

Figure H.8: Sydney Metro - Proportion of Households Speaks English Only

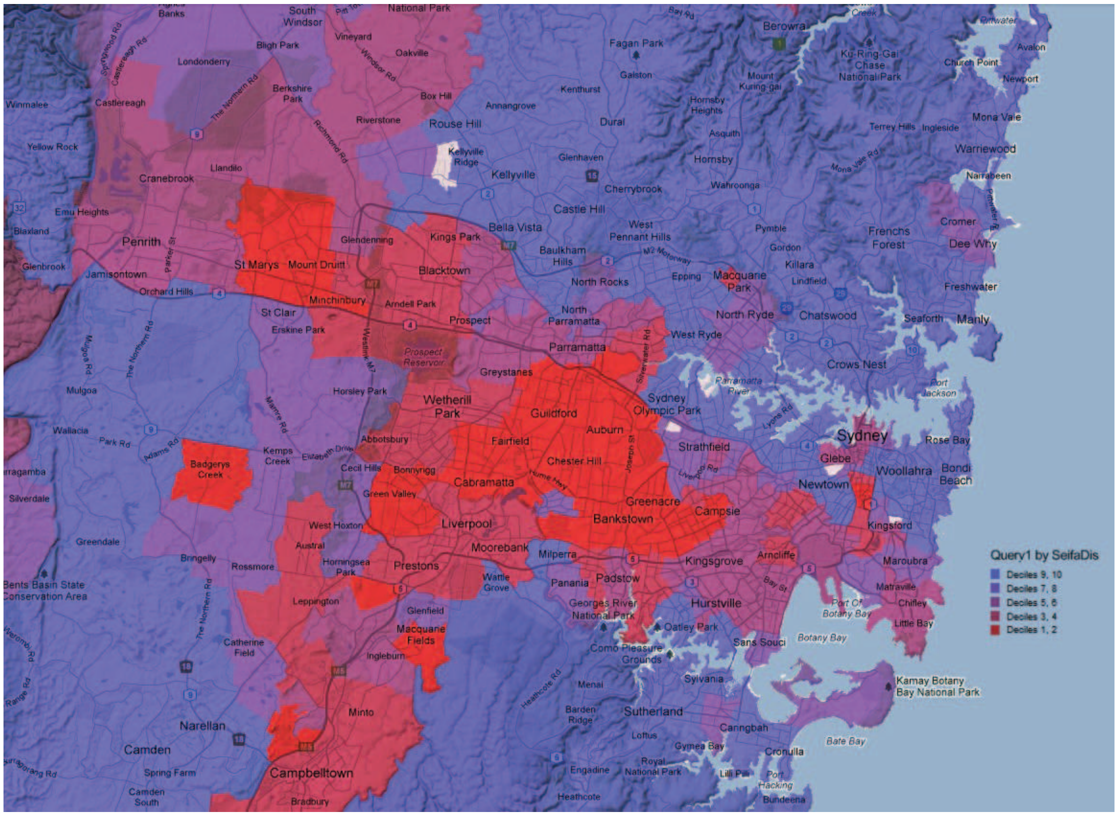

Figure H.9: Sydney Metro - Index of Socioeconomic Disadvantage, Deciles

Similar to the modelling from Chapter 7, two additional set of categorical variables are added to the covariates, the first bands the proportion of English speaking only families in a postcode, the seconds represents the pentiles of the index of socioeconomic disadvantage. Table H.1 shows the modelled coefficients of these location characteristics when applied to the settlement, $S_{j,i}$, process.

Variable	Coeff.	S.E.
Prop of English Only Spkers 0 - 50%	Baseline	
Prop of English Only Spkers 51 - 60%	-	
Prop of English Only Spkers 61 - 70%	-	
Prop of English Only Spkers 71 - 80%	-	
Prop of English Only Spkers 81 - 90%	-	
Prop of English Only Spkers 91 - 100%	-	
Index of Disadvantage bottom 1 - 20%	Baseline	
Index of Disadvantage bottom 21 - 40%	-0.02	0.02
Index of Disadvantage bottom 41 - 60%	**0.10**	0.02
Index of Disadvantage bottom 61 - 80%	**0.10**	0.03
Index of Disadvantage bottom 81 - 100%	**0.19**	0.02

Table H.1: Coefficients of the Socioeconomic Location Characteristics

The proportion of English Speaking Only householders in a postcode turned out not to be a significant explanatory factor in the modelling of settlement process; however, the index of socioeconomic disadvantage turned out to be a significant factor. The pattern for the latter is quite clear, the top 2 pentiles (or the 40% of the most disadvantaged postcodes) are similar, the next 2 pentiles have a 10% higher odds of finalising at a given revision and the bottom and the bottom pentile has the highest likelihood of finalisation with a coefficient of 0.19.

The next two figures (Figure H.10 and Figure H.11) show the average residual and the chi squared goodness fit test of the settlement process when no location information is included and when the socioeconomic index of disadvantage is used. Figure H.10 shows the model used in Chapter 6 still has considerable location based information with in the residuals, when observed actual number of finalisations is compared with the fitted number of finalisations. There are significant over-prediction of the number of final revisions in the South and Southwest parts of the city and under-prediction in the north and east parts. The bottom half of the Figure shows a chi-square goodness-of-fit test in a chloropleth map form. The darker reds shows a fitted value that deviates from the observed value significant while the lighter shades the fitted value is close to the observed value.

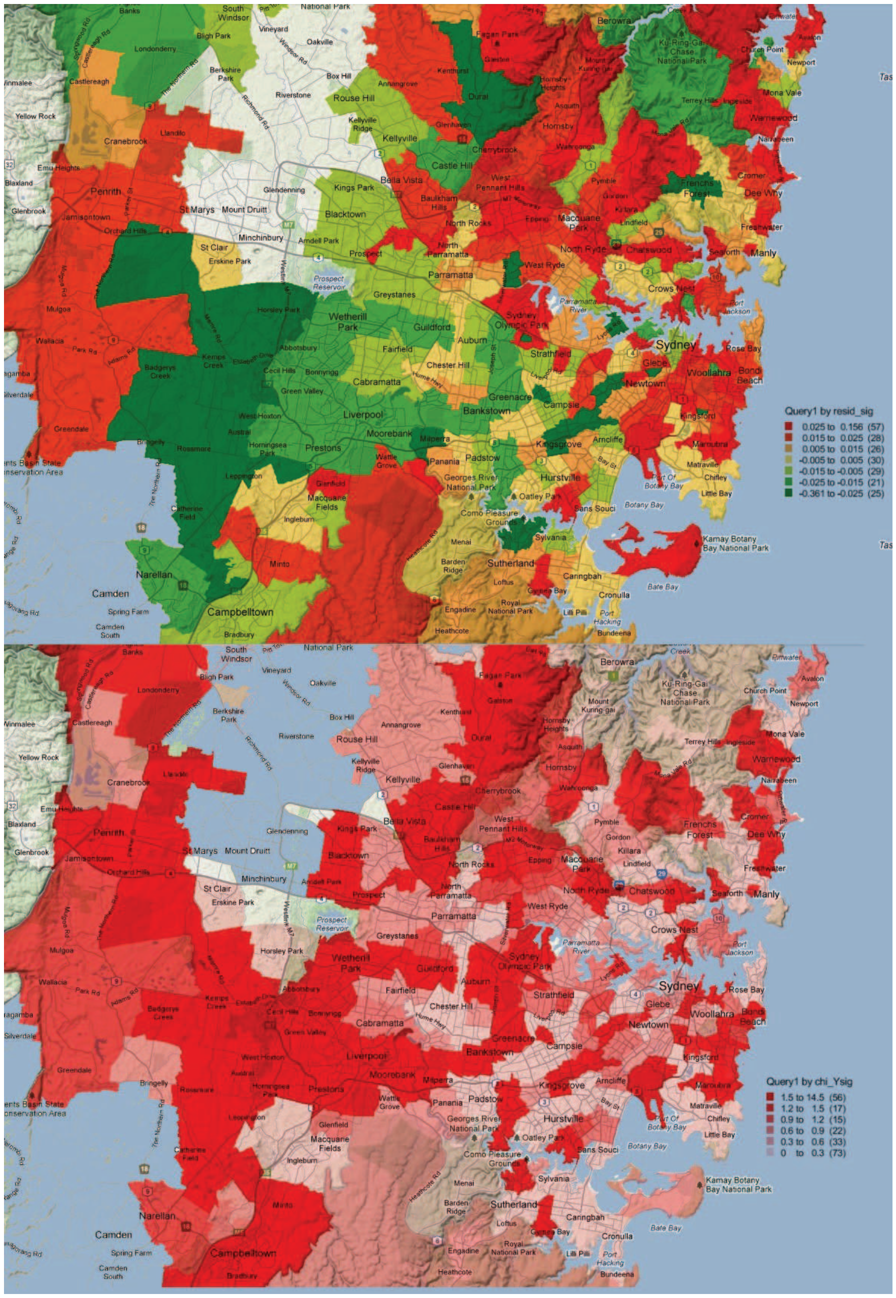

Figure H.10: No Location Information Used - Average Residual and Chisq Contribution

Figure H.11 presents the same information for the model with the socioeconomic index. While the colour patterns in the residual map is less obvious, there is still consistent over-prediction in the southwest parts of the Sydney Metro region. The χ^2 statistic for the model without location factors when summarised by the 217 postcodes is 272, which has a p-value of 0.007 against the null hypothesis that the model is a good fit by postcode. The χ^2 statistic for the model using the index of disadvantage is 210, which has a p-value of 0.621 against the null hypothesis of a good fit by postcode.

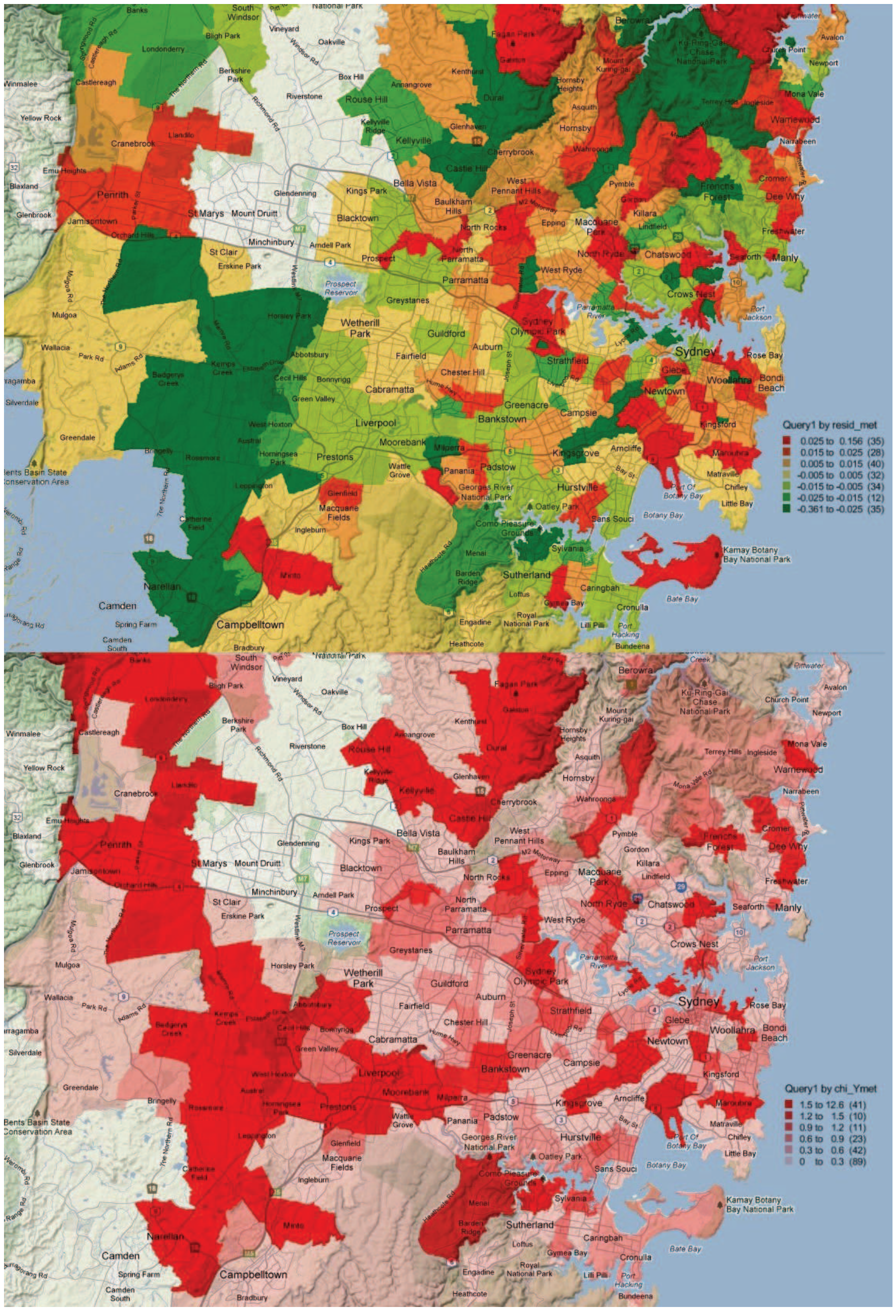

Figure H.11: Socioeconomic Characteristics - Average Residual and Chisq Contribution

Another method to measure the extent of the spatial information contained in the residuals is using a spatial correlogram. This is done through an iterative use of Moran's I, a measure of spatial covariance. Define distance intervals of equal length $(0, 1d], (1d, 2d], (2d, 3d]$... then define

$$I^l = \frac{\sum_i \sum_j w^l_{i,j}(Y_i - \hat{Y})(Y_j - \hat{Y})}{\sum_{i \neq j} \sum_i (Y_i - \hat{Y})^2} \tag{H.1}$$

where l is the multiple of the distance interval, i and j are pairs of postcodes and Y_i is the variable of interest for postcode i. $w^l_{i,j}$ is the weights used and takes on the value of 1 if the distance between i and j is between $((l-1)d, ld]$. The distance has been measured between the centroids of the postcodes using the Big Circle Distance (gnomonic projection). Plotting I^l against ld shows the autocorrelation against distance between postcodes. If there are spatial correlation, one would expect a significantly higher or lower correlation at shorter distance, which would decay towards 0 as distance increases.

Figure H.12 and Figure H.13 show the spatial correlation for the residuals from the model without location covariates and the model with the demographic characteristics. For the model without incorporating location based variables, there is a significant correlation between 5 and 15 km suggesting postcodes with centroids within 15km are correlated with each other. Note, within the Sydney Metro region, the pair of postcodes furthermost apart have a distance of 75km between them. The shaded region represent the 95% confidence interval, generated by simulation.

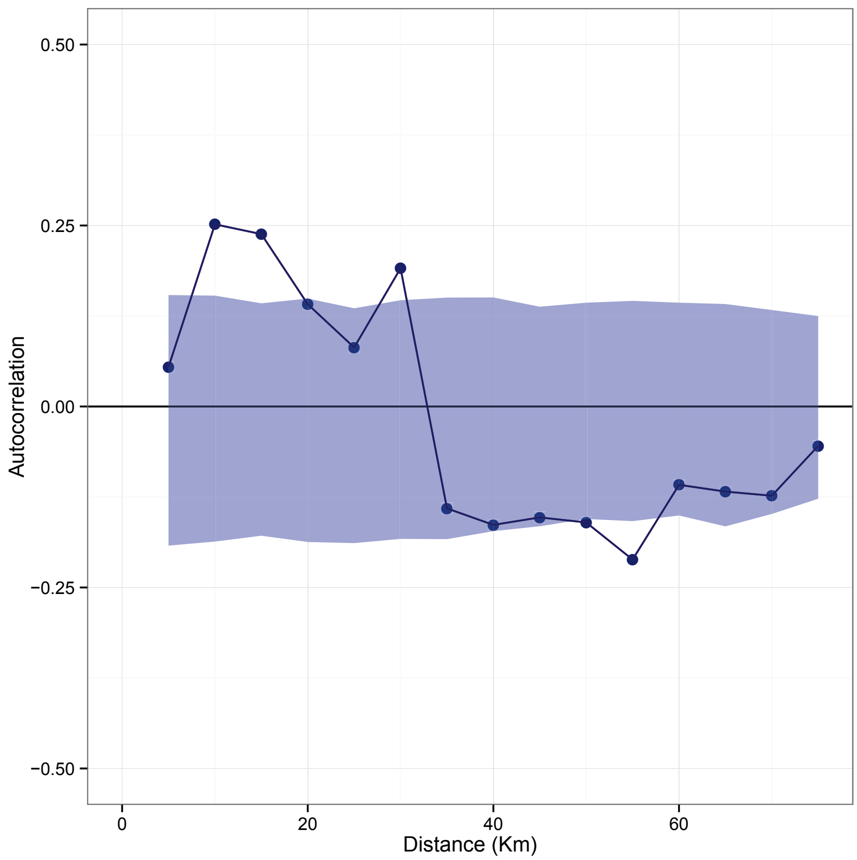

Figure H.12: Spatial Correlogram - No Location Information Model

By incorporating the demographics, the spatial autocorrelation is reduced. The spatial autocorrelation at all distances are now within the confidence interval; however, most of them marginally so.

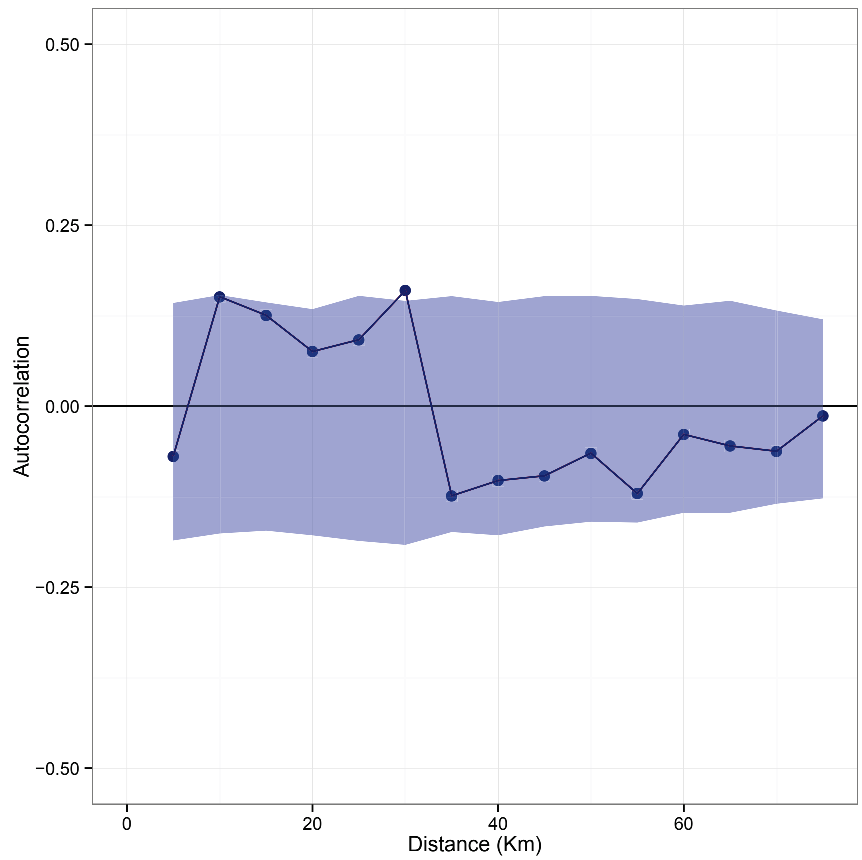

Figure H.13: Spatial Correlogram - Demographic Characteristics Model

H.4 Modelling Zones as Covariate

Another method to model location is to divide the area into an appropriate number of zones and use the zones directly in the modelling. The zoning adopted is the Australian Post's BSP zones (Barcoded Pre-sort). While this zoning is perhaps of little relevance to the NSW CTP claims finalisation probabilities, the zones are most readily available from the Australian Post website.

Within the Sydney Metro region there are 13 BSP Zones,

- Alexandria
- Burwood
- Frenchs Forest

- Illawarra
- Ingleburn
- Leightonfield
- Nepean
- Parramatta
- Pymble
- Seven Hills
- St Leonards
- Sydney Streets
- Waterloo

When the BSP zones are used as a covariates in the modelling of the probability of finalisation, it is a significant variable with coefficients having a range of 0.30, that is, the some locations has a 30% higher odds of finalisation a claim at a given revision compared to other locations. This is a wider spread compared to the coefficients for the socioeconomic disadvantage pentiles, which was around 0.2.

Figure H.14 shows the average residual and the chi-squared test for using zones directly in the model. The residual and the goodness of fit maps seem to be on par with the model that used the socioeconomic index of disadvantage. However, the χ^2 test statistic is 184, suggesting an improved fit by postcode.

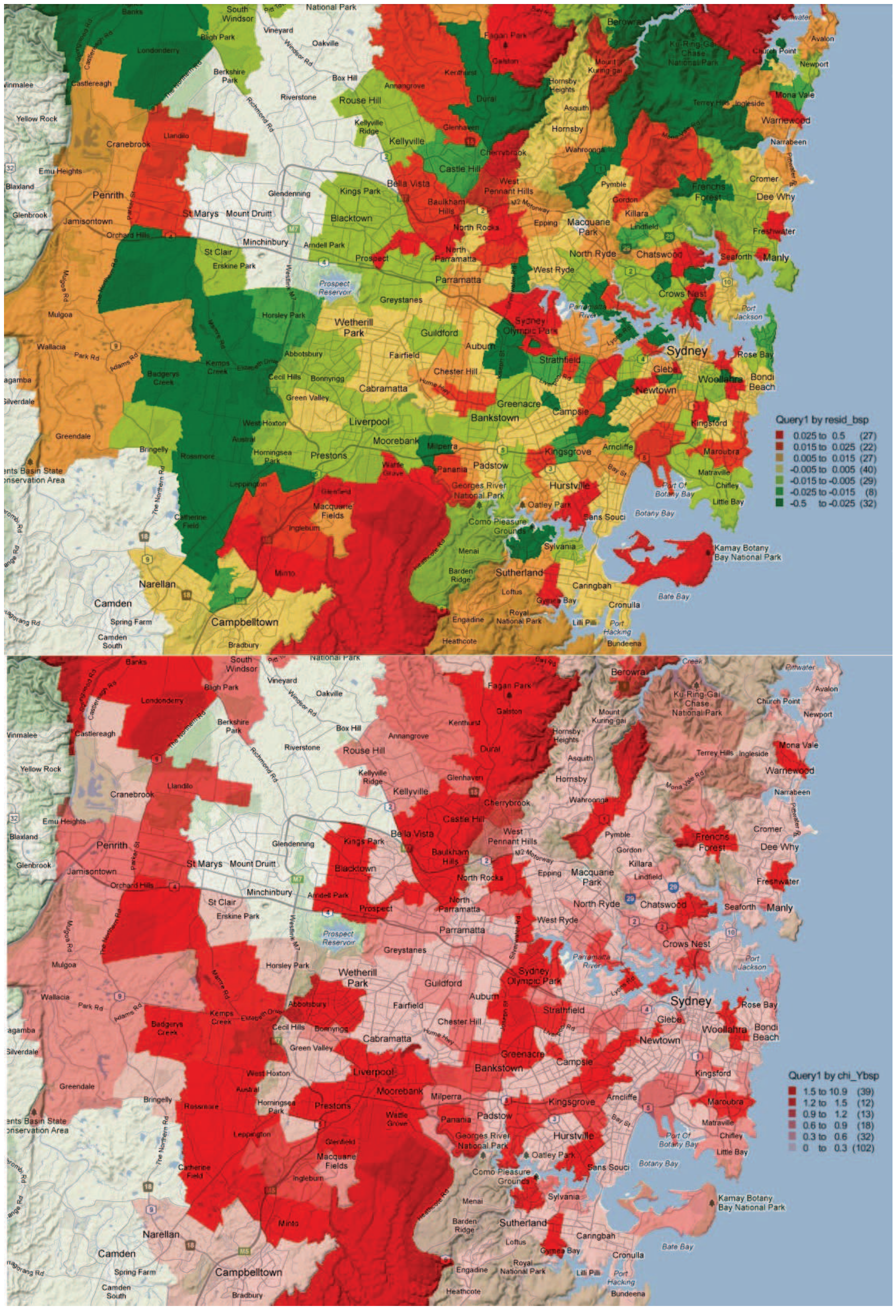

Figure H.14: BSP Zones - Average Residual and Chisq Contribution

Figure H.15 below shows the spatial correlogram for using the BSP zones in the modelling of the probability of settlement. All autocorrelations at various distances are within the confidence interval. It would appear the zones have removed all the spatial signals contained in the data.

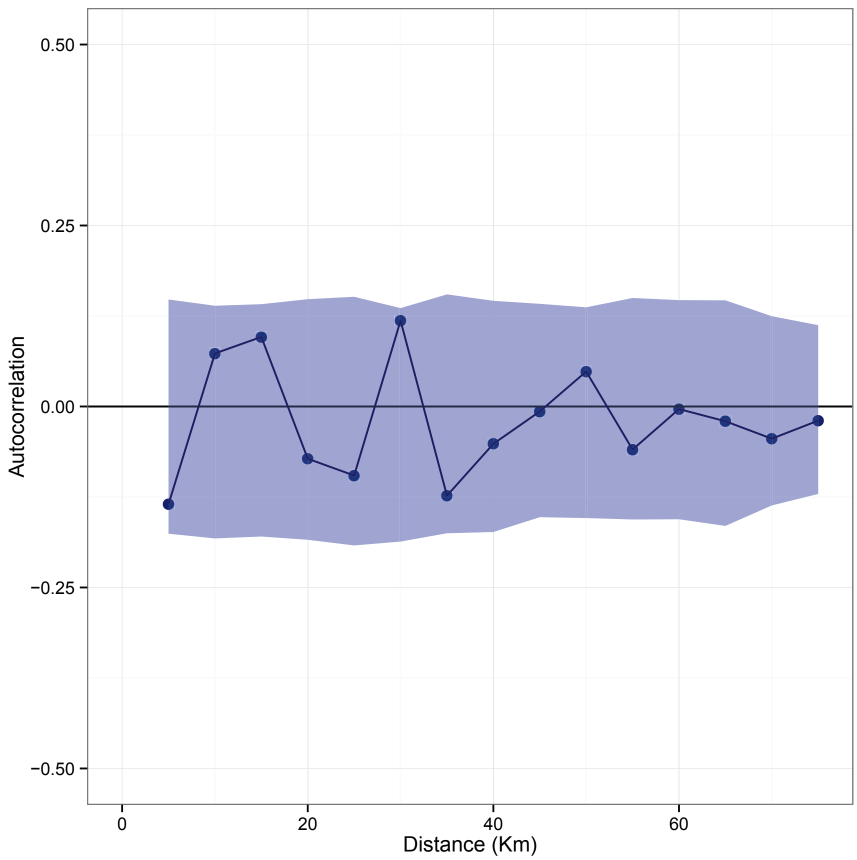

Figure H.15: Spatial Correlogram - BSP Zones Model

H.5 Spatial Smoothing

Kriging is a well known spatial modelling technique where interpolation between data points occur using a gaussian process with an assumed prior covariance structure. If the normality assumption is suitable, Kriging has been show to produce unbiased linear predictors. However, the normality assumption would not be appropriate in this situation.

Unlike Kriging, Thin Plate Splines (TP Splines) smoothing method does not require an assumed distribution. This makes TP splines a more versatile choice in modelling spatial patterns within the data. The residuals from the Settlement process model without any location factors were smoothed using TP splines and the resultant smoothed contour is shown below. The smoothed contour suggest the probability can differ up to 5%. Which also represents a 30% difference in the odds to finalise at a given revision.

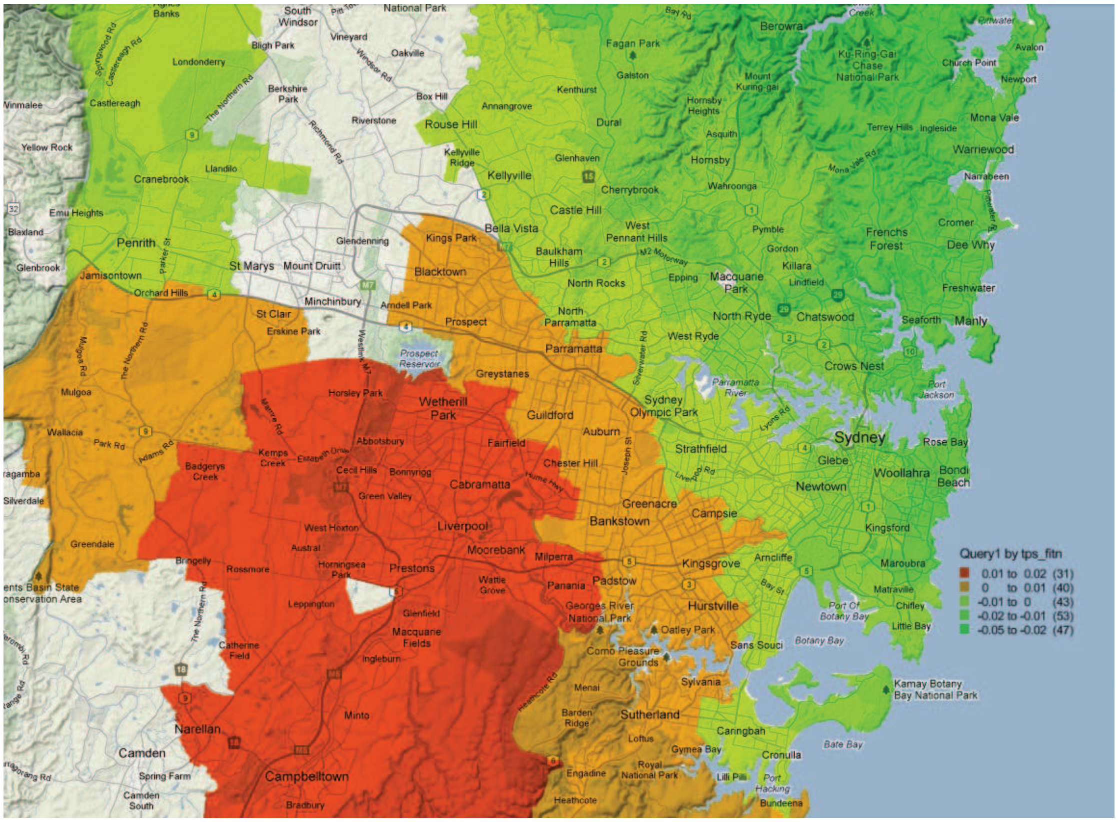

Figure H.16: TP Splines - Fitted Values Contour

Figure H.17 shows the average residual and the chi-squared test for fitting TP Splines to the residual of the no-location factor model. The residuals seems devoid of any clear pattern to the eye. The χ^2 statistic for a goodness of fit test using TP splines is 140, suggesting it removes location based information better than the other methods considered so far.

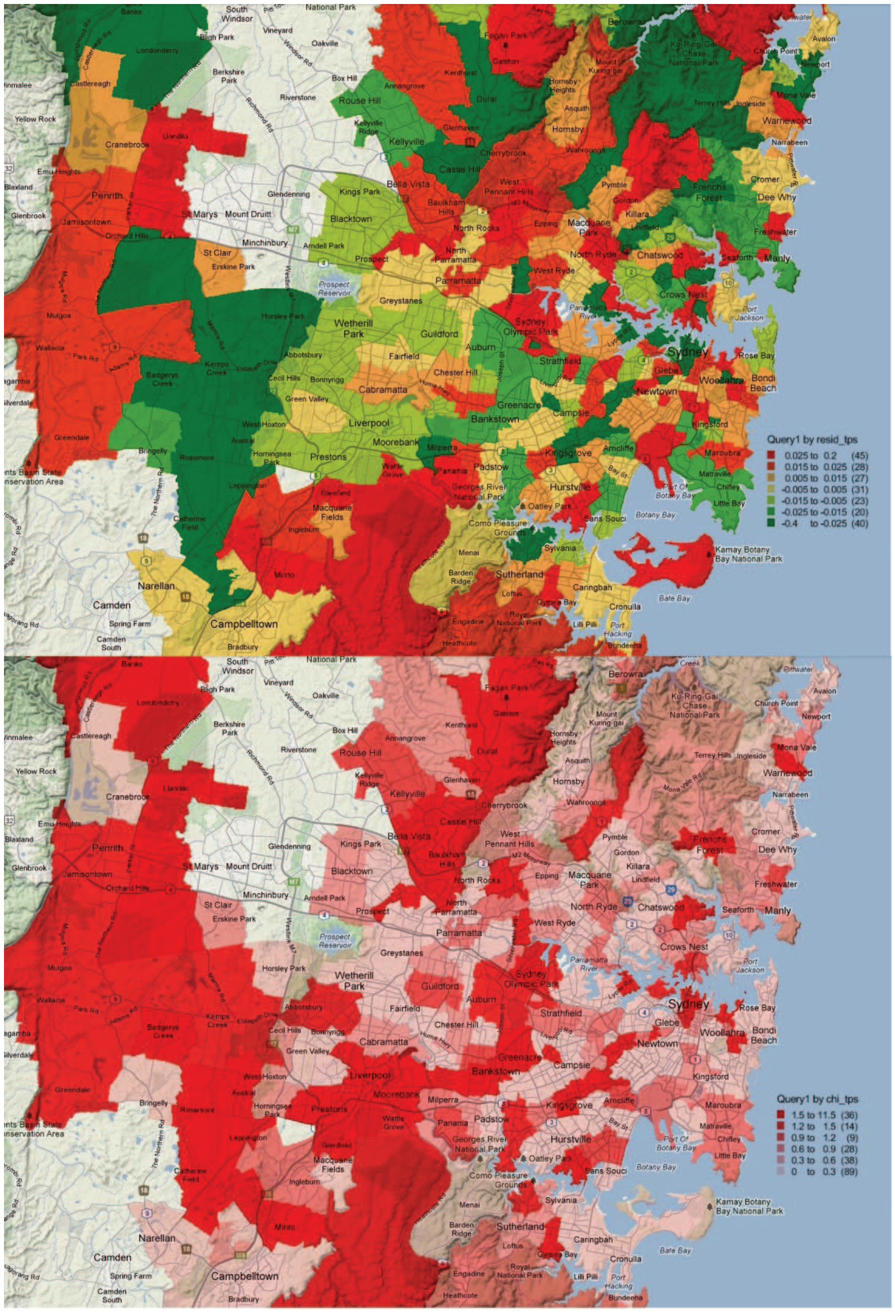

Figure H.17: TP Splines - Average Residual and Chisq Contribution

The spatial correlogram is very similar to the model that used BSP directly.

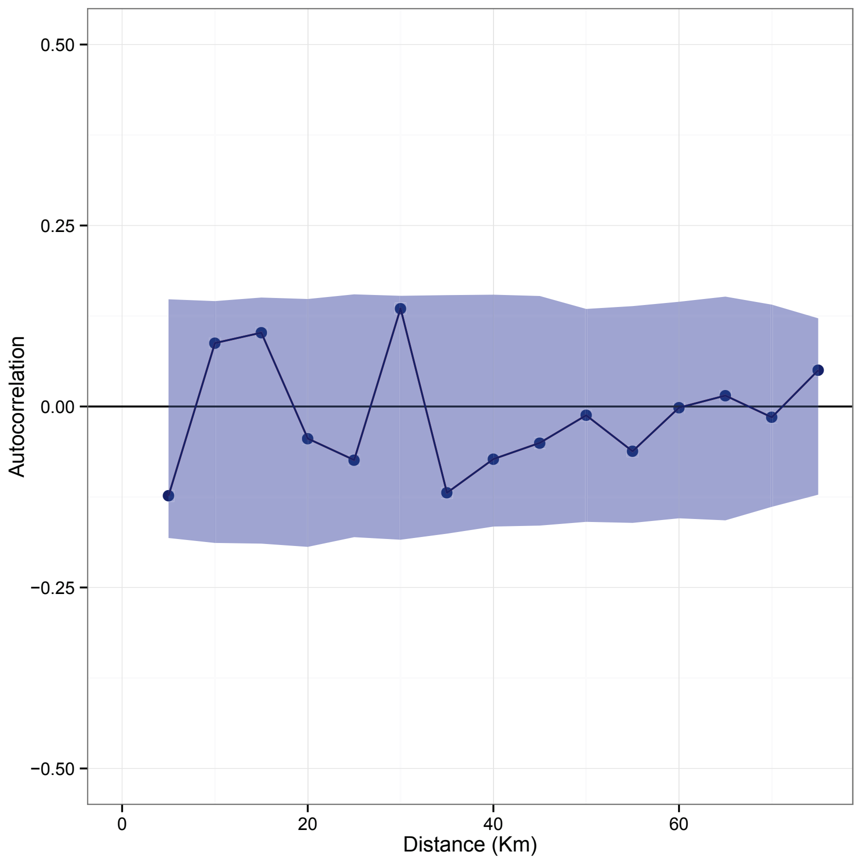

Figure H.18: Spatial Correlogram - TP Splines Model

While we have considered a few methods to allow for the spatial feature of the data, we have not able to incorporate these methodologies into the GLARMA structure in a unified framework. We have left such ambitions for potential future research.

References

Australia Prudential Regulation Authority (2013). Prudential standard gps310: Audit and actuarial valuation and reporting. 45

Australian Bureau of Statistics (2006a). 2006 Census. 315

Australian Bureau of Statistics (2006b). 9309.0 - motor vehicle census, Australia. 316

Australian Bureau of Statistics (2014). 6302.0 - average weekly earnings, Australia. 55

Ayuso, M. and Santolino, M. (2008). Prediction of individual automobile rbns claim reserves in the context of solvency ii. *Research Institute of Applied Economics, Working Paper*. 75

Brookes, R. and Prevett, M. (2004). Statistical case estimation modelling–an overview of the nsw workcover model. In *IAA Accident Compensation Seminar*. 66, 76

Cox, D. R., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., Juselius, K., and Lauritzen, S. L. (1981). Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics*, pages 93–115. 84

Davis, R. A., Dunsmuir, W. T., and Streett, S. B. (2003). Observation-driven models for Poisson counts. *Biometrika*, 90(4):777–790. 84, 116, 271, 276, 280

Davis, R. A., Dunsmuir, W. T., and Streett, S. B. (2005). Maximum likelihood estimation for an observation driven model for Poisson counts. *Methodology and Computing in Applied Probability*, 7(2):149–159. 96

Diggle, P., Heagerty, P., Liang, K.-Y., and Zeger, S. (2002). *Analysis of Longitudinal Data*. Oxford University Press. 112

Dunsmuir, W. (2015). *Handbook of Discrete Valued Time Series*, chapter Generalized linear autoregressive moving average models. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press. 206, 208, 212

Dunsmuir, W., McKendry, C., and Dean, R. (2014a). Modelling 'long' longitudinal data. to be submitted. 206, 209, 216

Dunsmuir, W. T., Li, C., and Scott, D. J. (2014b). *glarma: Generalized Linear Autoregressive Moving Average Models*. R package version 1.3-0. 96, 116

Dunsmuir, W. T. and Scott, D. J. (2015). The glarma package for observation driven time series regression of counts. *Journal of Statistical Software*. 102, 116

England, P. and Verrall, R. (1999). Analytic and bootstrap estimates of prediction errors in claims reserving. *Insurance: mathematics and economics*, 25(3):281–293. 45

England, P. D. and Verrall, R. J. (2001). A flexible framework for stochastic claims reserving. In *Proceedings of the Casualty Actuarial Society*, volume 88,1, pages 1–38. 57

England, P. D. and Verrall, R. J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8(03):443–518. 50

Greenfield, A., McGuire, G., and Miller, H. (2011). Actuarial valuation of the benefit system for working-age adults. Technical report, Taylor Fry. 77, 173

Hart, D. G., Buchanan, R. A., and Howe, B. A. (1996). *The Actuarial Practice of General Insurance*. Institute of Actuaries of Australia. 4, 5, 43, 49, 52, 53

Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., and Tibshirani, R. (2009). *The elements of statistical learning*, volume 2,1. Springer. 76

Hertig, J. (1985). A statistical approach to ibnr-reserves in marine reinsurance. *Astin Bulletin*, 15(02):171–183. 72

Institute of Actuaries of Australia (2013). Professional standard 300: Valuations of general insurance claims. 45

Kaas, R., Van Heerwaarden, A., and Goovaerts, M. J. (1988). Between individual and collective model for the total claims. *Astin Bulletin*, 18(02):169–174. 43

Knaus, J. (2013). *snowfall: Easier cluster computing (based on snow)*. R package version 1.84-6. 166

Konstantinidis, B., Mayo, R., and Priest, C. (2007). Game theory and australian's ctp markets. In *IAA Accident Compensation Seminar*. 18, 19

Larsen, C. R. (2007). An individual claims reserving model. *Astin Bulletin*, 37(1):113. iv, 79, 87, 89, 127, 223

Lawless, J. F. (2011). *Statistical models and methods for lifetime data*, volume 362. John Wiley & Sons. 151

Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *Astin Bulletin*, 23(02):213–225. 72

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Model*, volume 37. Chapman & Hall/CRC. 56, 58

McGuire, G. and Actuaries, T. F. C. (2007). Individual claim modelling of ctp data. In *IAAust XIth Accident Compensation Seminar*. 58, 59, 74

Motor Accidents Authority (2008). Coding manual - claims register and statistical database. 245, 317

Motor Accidents Authority (2012). *Annual Report 2012*. Motor Accidents Authority of NSW. 18

Motor Accidents Authority (2014a). Maa annual report 2013-14. 19

Motor Accidents Authority (2014b). Maa premium determination guidelines. 19

Mulquiney, P. and Actuaries, T. F. C. (2004). Application of soft-computing techniques in accident compensation. In *IAA Accident Compensation Seminar*. 57

Oryzak, D. (2008). Transition models underlying statistical case estimation. In *IAA General Insurance Seminar*. 173

Pinheiro, J. C. and Chao, E. C. (2006). Efficient laplacian and adaptive gaussian quadrature algorithms for multilevel generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 15(1). 211

Pinheiro, P. J., Andrade e Silva, J. M., and de Lourdes Centeno, M. (2003). Bootstrap methodology in claim reserving. *Journal of Risk and Insurance*, 70(4):701–714. 45

Prevett, M. and Gifford, D. (2007). Statistical case estimation for long term claimants–uncovering drivers of long term claims cost in accident compensation. In *IAA Accident Compensation Seminar*. xvi, 66, 67, 69, 70, 76

Rice, J. (2006). *Mathematical statistics and data analysis*. Cengage Learning. 112

Rydberg, T. H. and Shephard, N. (2003). Dynamics of trade-by-trade price movements: Decomposition and models. *Journal of Financial Econometrics*, 1(1):2–25. 84, 85, 88, 96

Schiegl, M. (2010). A three dimensional stochastic model for claim reserving. *arXiv preprint arXiv:1009.4146*. 78

Smyth, G. (2014). *statmod: Statistical Modeliog)*. R package version 1.4.20. 212

Taylor, G. (1981). Speed of finalization of claims and claims runoff analysis. *Astin Bulletin*, 12:81–100. 40

Taylor, G. (2000). *Loss Reserving: An Actuarial Perspective*. Springer US. 43, 72

Taylor, G. and Campbell, M. (2002). *Statistical Case Estimation*. Centre for Actuarial Studies, Department of Economics, University of Melbourne. 45, 75

Taylor, G., McGuire, G., and Sullivan, J. (2008). Individual claim loss reserving conditioned by case estimates. *Annals of Actuarial Science*, 3(1-2):215–256. 43, 46, 78, 87

Taylor, G. C. and McGuire, G. (2004). *Loss Reserving GLMs: A Case Study*. Centre for Actuarial Studies, Department of Economics, University of Melbourne. 57, 73, 74, 97

The Government of New South Wales (1999). Motor accdents compensation act 1999. 13

The Government of New South Wales (2007). Motor accdents compensation act 2007 amendments. 14

The Government of New South Wales (2010). Motor accdents compensation act 2010 amendments. 17

The Government of New South Wales (2013). Motor accident injuries amendment bill 2013. 14

Verbeke, G. and Molenberghs, G. (2009). *Linear mixed models for longitudinal data*. Springer Science & Business Media. 112

Wang, B. (2004). GLARMA Models and Stock Price Dynamics. Master's thesis, University of New South Wales, Sydney, Australia. 84, 88, 100, 103, 116, 117, 269

Wüthrich, M. V., Merz, M., Bühlmann, H., De Felice, M., Gisler, A., and Moriconi, F. (2009a). Modelling the claims development result for solvency purposes. *Astin Bulletin*. 78

Wüthrich, M. V., Merz, M., and Lysenko, N. (2009b). Uncertainty of the claims development result in the chain ladder method. *Scandinavian Actuarial Journal*, 2009(1):63–84. 45, 72