ACTUARIAL MODELS FOR THE ANALYSIS OF DISABILITY INCOME INSURANCE

By

David G.W. Pitt

November 2004

A thesis submitted for the degree of Doctor of Philosophy at the Australian National University.

To my parents,

who have always supported me and provided me with the very best in educational opportunities.

DECLARATION

This thesis contains no material which has been accepted for the award of any other degree or diploma in any other university, and to the best of my knowledge and belief, this thesis contains no material published or written by another person, except where due reference is made in the thesis.

David G Pitt

David Pitt, November 2004

ABSTRACT

After a review of the actuarial literature on the analysis disability income insurance (DII), this thesis develops a number of different models for the description of claim incidence and claim termination rates for holders of DII policies. Models developed include generalised linear models, parametric mixture models with both accelerated failure components and long run return to work probabilities, semi-parametric Cox Regression models and censored regression quantile models. The results from the various proposed models and their suitability as models of the disability process are compared and contrasted. Chapter 6 brings together the results of the modelling work presented in the earlier chapters and proposes new models for premium rating and disabled life reserving using multiple state model theory and flowgraph analysis.

ACKNOWLEDGEMENTS

I would like to thank my PhD supervisor, Michael Martin. Michael helped me considerably with balancing the pressures of combining full-time university lecturing with a PhD enrolment. His advice and guidance have made my transition to academic work considerably easier. I am also grateful to Ross Maller for the useful suggestions and encouragement he gave me during the last year of my PhD enrolment. I would also like to thank the Head of the School of Finance and Applied Statistics at the Australian National University, Terry O'Neill, for the opportunity he gave me to work in his school and for the support he has given me during my time at the ANU.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

1.1 Introduction and Motivation

Between 1988 and 2001, the in-force annual premiums for Australian disability income insurance (DII) have increased from 180 million dollars to almost 1.1 billion dollars (Munich Re of Australia, Rice Kachor Research (2002)). This increase has occurred across both ordinary and superannuation linked lines of life insurance risk business. This sharp increase in the volume of DII business, and the inherent difficulty in calculating actuarially fair premium rates for this line of business, provided the motivation for the research presented in this thesis.

The calculation of a risk premium for an insurance contract requires an assessment of future expected cash outflows for an insurer that has entered into a portfolio of insurance contracts. In the case of DII, these cash outflows will be dominated by benefits payable to insured lives who subsequently become unable to work. They will also include a significant amount related to the expenses of selling and product administration. The actuary is required to assess the magnitude and timing of these uncertain future cash flows and, in turn, to recommend a premium rate, which will cover these future benefit payments and leave a contribution to the profit of the business, after allowance for investment income earned on premium income and associated reserves.

Expenses play a particularly important role in the determination of premium rates for DII business. DII business is predominantly sold through insurance brokers and agents, and hence large initial expenses related to commissions are common. Once a DII contract has been signed, the renewal expenses (payable generally annually) are significantly lower than the expenses that the office faces in the first year after initial sale. It is therefore of interest for insurers to have a measure of the persistency, or renewal rates, for those insured lives who have purchased DII in the past. Higher renewal rates will lead to lower expenses on average each year and in turn can justify lower premium rates.

The largest and most uncertain future cash outflow in the management of DII business arises due to the payment of the insured disability benefits. Uncertainty in the amount of these cash flows stems from two sources: first, the occurrence of future disability for an insured life is uncertain; and second, the duration of a claim under a DII contract can vary from as little as one week to many years. It is common actuarial practice (Booth, 2000) to estimate the rate of disability onset using a claim incidence rate and to estimate the duration of claims using a set of estimated claim termination rates.

A number of morbidity tables have been produced, based on the disability experience of life offices, in various countries. These tables provide practising actuaries with a starting point for the assessment of their future liabilities associated with in force DII business. These tables include (Munich Re, 2002):

- The Manchester Unity Tables (covering a lengthy nineteenth century investigation in England);
- The Commissioner's Disability Tables (CDT) 1964 (based on USA experience);

- The Commissioner's Individual Disability Tables (CIDA) 1985; and

- The Australian Table IAD89-93.

The construction of the Australian Table and also some of the experience analysis performed in the UK will be reviewed in Chapter 2 of this thesis.

1.2 Contribution of this Thesis

This thesis proposes the use of a number of recently developed statistical methodologies to the actuarial pricing of DII business. Experience analysis of DII business overseas and in particular in Australia has up until now relied on statistical methods developed over thirty years ago. Multiple state modelling, exposed-to-risk and graduation methods along with actual vs expected analysis have formed the historical basis for the modelling of claim incidence and claim termination rates. These forms of modelling do not readily capture interactions that exist between the rating variable information available to DII providers. Historically popular methods rely on subdividing DII data prior to performing analysis to enable faster computation of output. Faster computing speeds and increased statistical literature that makes use of these methods, enables considerably more data analysis to be performed on DII policy and claim data.

This thesis will highlight the role that a number of recent statistical procedures have to play in the analysis of DII business. In particular, mixture modelling in survival analysis, generalised linear modelling, censored regression quantiles and flowgraphs will all be considered and their efficacy and suitability for morbidity modelling will be investigated and discussed. The modelling will involve statistical estimation of underlying rates of claim incidence and claim termination. The results of this modelling will then be used with existing actuarial pricing theories to develop suitable premium rates and reserving strategies for DII business.

1.3 Structure of Thesis

Chapter Two will provide a review of the most important literature relating to the analysis of DII business from an actuarial and statistical viewpoint. This will place the research that follows in an appropriate context.

Chapter Three will report the results of two generalised linear model analyses of DII claim experience presented to the Institute of Actuaries of Australia membership in 2002 and 2003. The termination rate analysis in that chapter was published in the Australian Actuarial Journal and the incidence rate analysis formed part of the Proceedings of the 2003 Biennial Convention of the Institute of Actuaries of Australia.

Chapter Four continues the analysis of claim termination rates with the use of survival models for long term survivors (Maller and Zhou, 1996). Survival analysis provides a natural modelling framework for the analysis of disability claim durations. Both claim termination rates and probabilities of failure to recover are modelled in terms of various sets of underlying covariates or rating factors.

Chapter Five presents the results of a study based on the use of censored regression quantile analysis, (Portnoy, 2003). Regression quantiles provide a method for the diagnosis of heterogeneity in residuals resulting from regression analysis. In particular, censored regression quantiles enable the analyst to assess potentially varying impacts of covariates across different quantiles of the distribution of the dependent variable. In the case of DII analysis we can, for example, assess whether the impact of smoker status on claim termination varies according to the duration that a claimant has already been receiving DII benefits. This is an important feature of the analysis because the calculation of disabled life reserves depends critically on the impact of covariates at a range of claim durations.

Chapter Six ties together some of the work presented in earlier chapters and develops actuarial pricing strategies using both multiple state models and flowgraph models.

Chapter Seven concludes the thesis. A summary of the main findings of the research is given along with recommendations for future research.

Acknowledgement

The data that has been used in this thesis has been provided by the Institute of Actuaries of Australia Life and Risk Practice Committee and is used with their permission.

CHAPTER TWO

LITERATURE REVIEW

2.1 Disability Income Insurance Models

The development of a number of different models for the description of disability income insurance (DII) has taken place over the last two decades. The models proposed have varied considerably in terms of underlying assumptions, the amount of underlying statistical rigour and the extent of data used in their development. This section will review a number of the key papers on DII, and place the research work developed in this thesis in context.

2.1.1 Continuous Mortality Investigation Reports

The Continuous Mortality Investigation Bureau (CMIB) of the Institute of Actuaries and the Faculty of Actuaries in the UK has a sub-committee devoted to analysis of Permanent Health Insurance (PHI) data. PHI is a very similar product to DII products sold in Australia and Income Protection (IP) Insurance sold in the United States. The sub-committee has produced a number of reports (CMIR2 (1976), CMIR4 (1979), CMIR7 (1984), CMIR9 (1988), CMIR12 (1991), CMIR18 (1998) and CMIR20 (2001)). In terms of methodological development, CMIR12 (1991) was the most substantial of these reports. This report established the multiple state model as a key method for the pricing and reserving in respect of DII policyholders. The multiple state model used in the analysis is shown in Figure 2.1 below.

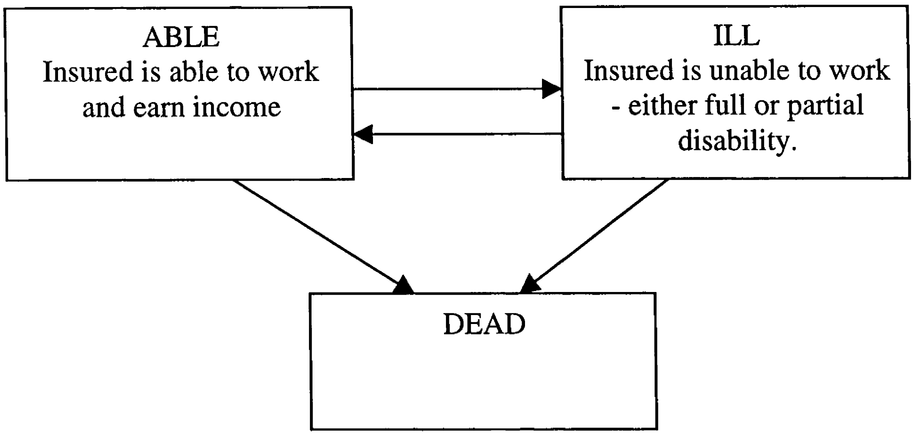


Figure 2.1: Multi-state model used for Disability Income Insurance Modelling. A policyholder may move between the 3 states, with death an absorbing state

CMIR 12 begins by explaining the link between transition intensities and conditional probabilities for the state occupied by policyholders. Discussion then turns to the graduation of claim recovery and mortality intensities. In order to simplify the analysis and without loss of generality, Male Standard lives are the only lives considered in the report. After considerable experimentation, the complete formula used for the graduation of recovery rates derived in CMIR 12 was

$$\rho_{y+z,z} = r.\left\{a+b\left(1+q.\max\left(4-wz,0\right)\right).\sqrt{Z}\left(Y-50\right)\right\}e^{-c\sqrt{Z}} \qquad (2.1)$$

where a, b, c, p, q and s are constants and

$\rho_{y+z,z}$ is the transition intensity (or hazard rate) for recovery of a life aged $y+z$ who has been in the disabled state for a continuous period of z years

y is exact age (in years) at the date of falling sick, and

$$Y = y \qquad \text{for } z \le 5,$$

$$Y = y + z - 5 \qquad \text{for } z > 5;$$

z is duration of sickness (in years) and

$$Z = z \qquad \text{for } z \le 1,$$

$$Z = 1 + s(z-1) \qquad \text{for } 1 < z \le 5,$$

$$Z = 1 + 4s \qquad \text{for } z > 5$$

and for deferred periods of 4, 13 and 26 weeks, we have $r = \min\left(p + \frac{1}{4}(wz - d)(1-p), 1\right)$ so that $r = 1$ if $wz > d + 4$ for deferred period d, expressed in weeks. Note also that w is a constant equal to the number of weeks in a year, taken as 52.18 in this analysis.

Equation (2.1) represents a stochastic model for the transition intensity from the disabled state to the able state. The model is fit using slices of data from the CMIR12 investigation where the slices contain recovery data for lives with specified age, deferred period and duration of disability characteristics. The above formulation was determined after various diagnostic checks on relationships present in the data. In particular, a plot of the log of the duration of claim effect (on claim termination rates) against the square root of duration produced a graph that was approximately linear and decreasing. This finding explains the $e^{-c\sqrt{Z}}$ term in the above graduation formula for termination rates. If all deferred periods are combined and the impact of increasing age on claim termination rates is plotted against age, an approximately linear, decreasing relationship is also found. There was some evidence that the observed reduction in claim termination rates for older policyholders varies depending on the deferred period of the underlying DII policy. This phenomenon suggests the need to model an interaction between age and deferred period in the description of claim termination rates – this interaction is explored more fully in Chapter 3, which focuses on generalised linear models and claim termination and incidence rates. This interaction is dealt with in equation (2.1) by the term $\sqrt{Z}\left(Y-50\right)$ included in the slope coefficient. There is also evidence of lower recovery rates during the periods soon after claim payments commence. These are dealt with in equation (2.1) by the reduction factor, r, which only applies for the period 4 weeks after the expiry of the deferred period. Evidence of an increased sensitivity of claim termination rates to age during the first four weeks of disability is dealt with by the $q.\max\left(4-wz,0\right)$ term in equation (2.1).

The graduation formula (2.1) was fitted using the method of maximum likelihood. A general discussion of maximum likelihood in the context of actuarial experience analysis is given in Forfar et al (1988). An experience analysis is the term used by actuaries to describe the investigation of a set of statistical data that is derived from business transactions over a long period. The most common experience analysis performed by actuaries is an investigation into mortality rates using data on deaths derived from payments of sums insured under endowment and term life insurance business. The use of maximum likelihood estimation means implies asymptotic standard errors for the estimated parameters in (2.1) are readily available. In addition, Bayliss (1991) tested the suitability of these standard errors by a method of simulating the experience and parameter-fitting process. His results were very close to those found using the asymptotic results based on maximum likelihood theory.

Next, we turn to the treatment of the able and ill to death transitions shown in Figure 2.1.The analysis of mortality rates using disability income insurance data is difficult due to the lack of data on deaths for insured lives who are initially in paid employment. However, the report does find evidence of a "hump-backed curve" relating the mortality experience to duration of disability. That is, mortality rates appear to be higher at shorter durations of disability claim (less than 20 weeks), regardless of the age of the claimant. The impact of increasing age on mortality rates is really only evident for ages greater than 50, where there is some evidence of the usual positive relationship between age and mortality rates.

The graduation of sickness inception intensities is also considered in CMIR12 (1991) by Waters. The method used is to separate data on exposure to the risk of claim inception and number of claims by age and deferred period. Separate graduations are performed for each of the four main deferred periods, namely 1 week, 4 weeks, 13 weeks and 26 weeks. Again experimentation with a range of functional models led to the use of

$$\sigma_x = \exp\{f(x)\}, \tag{2.2}$$

where σ_x represents the transition intensity (or hazard rate) for the transition from the able state to the disabled state, x represents age and f is usually a polynomial of degree three. This form of mathematical function and underlying model used for claim inceptions suggested the use of generalised linear modelling as a natural modelling approach. Higher sickness inception intensities were predicted for shorter deferred periods. In addition the impact of age was to increase the claim rates for both the youngest and oldest ages of the insured lives. That is, the sickness inception intensities were highest for those insured lives younger than 30 or older than 55. The impact of age on sickness inception intensities was greatest for the one week deferred period. The sickness inception intensities for the 26-week deferred period policies showed little evidence of a relationship with age – the graduated sickness inception intensities were about 10% across the entire working age range.

CMIR12 (1991) next provides details on the computational procedures for determining both probabilities of movement between, or occupancy of, various states in the three state model in Figure 2.1. In addition the approximation of monetary functions using numerical integration methods was also described. Chapter Six of this thesis considers numerical methods for dealing with the semi-Markov nature of the transitions from the disabled state and the associated calculations of premium rates and disabled life reserves. Chapter Six will also demonstrate a new use of flowgraphs to calculate DII premium rates.

2.1.2 Significant Contributions to the Disability Income Insurance Modelling Literature

Many other models have been proposed in the literature for the pricing and reserving of disability income insurance. Gregorius (1993) describes the methodology used in the Netherlands. The potential annuity of benefits payable up until retirement under a DII contract is broken into two separate parts, the A-cover and the B-cover. The A-cover includes all payments to insureds for the first year of disability. A simple multiplicative model is used to calculate the risk premium in respect of the A-cover. The risk premium is calculated as

$$P(i,j,k) = B.a(i).b(j).c(k), \tag{2.3}$$

where the functions a, b and c relate to the level of each of the three main risk factors, namely waiting period, age and class of profession. The premium rate calculation for the B-cover is more complicated due to the greater uncertainty in estimating appropriate tariffs for potentially long duration claims. Actuarial methods based on a probability model of future periods of disability for insureds are typically used. It is important to note that the model uses duration dependent recovery rates and also that mortality rates for active and disabled lives are not differentiated. Gregorius concedes that there are theoretical reasons for believing in differential mortality rates for active and disabled lives, however, there is inadequate data in the Netherlands DII market to assess this theory. In order to deal with the duration dependent recovery rates, Gregorius proposes a refinement of the usual three state model used in DII analysis. The disabled state is divided into six states where each state denotes disability for a different amount of time. To simplify the calculations, a model without return, where transitions out of the disabled state are ignored is proposed for the determination of single premiums. This simplification clearly introduces significant bias into the analysis proposed by Gregorius.

Taylor (1971) introduced a Markov model with multiple states for duration of disability. This paper was key to the development of stochastic analysis in morbidity studies. This work was developed further by Haberman (1984).

Segerer (1993) describes the actuarial treatment of the disability risk in Germany, Austria and Switzerland. In these countries, and a number of other European centres, the method of pricing and reserving for DII is based on multiple decrement tables (Neill, 1977). The decrements are state dependent. For active lives, the two decrements of primary interest are mortality and disability. For those lives in the active state, the two decrements are recovery and mortality. The formula used in Switzerland for a net single premium is

$$\frac{1}{N_x - N_{x+n}} \sum_{t=0}^{n-1} D_{x+t+1/2} . i_{x+t} . g_{x+t} . \overline{a}^{i}_{x+t+1/2:\overline{n-t-1/2|}}, \tag{2.4}$$

where N_x and D_x are standard actuarial commutation functions evaluated using national mortality rates and 3% per annum interest, i_x is the incidence rate of disability at age x, g_x is factor that reflects the degree of the disability and $\overline{a}^{i}_{x:\overline{n|}}$ represents the expected present value of an annuity of 1 dollar per annum payable while a life aged x remains disabled with payments continuing for a maximum period of n years. Interestingly, the disabled life reserves are calculated based on standard life annuity values, reduced by a factor to allow for the fact that payment is made contingent on both survival and continuing disability. The formula used for the calculation of disabled life reserves is

$$ {}_tV_x^i = 1.02 a_{x+t:\overline{n-t}|} \left(\left(1 - \frac{t}{5}\right) f(x) + \frac{t}{5} \right), \tag{2.5} $$

where

$$ f(x) = 0.3 + 0.7 \frac{(x-15)^2}{50}. $$

Here, $f(x)$ is the age-dependent deduction factor and ${}_tV_x^i$ represents the amount of money that an insurer should hold in reserve in respect of a DII policy that was bought t years ago by a life that has just become disabled and was aged x at the outset of the policy. At age 15, the factor is 0.3 and the factor increases continuously until it reaches unity at age 65. An age-dependent deduction factor of one implies equality between the standard life table annuity value and the disability annuity. The approaches adopted in Germany and Austria are slight variations on the approach in Switzerland. See Segerer (1993) for details. The underlying disability incidence and continuance rates used in Equations (2.4) and (2.5) are based on industry average statistics split by age and duration. Following tariff reform in parts of Europe in the early 1990s that impacted premium calculations, large loadings of up to 50% were added to previously used rates to bring them closer into line with industry experience.

The use of multiple state models in actuarial science has extended beyond just premium rating and reserving for DII. Hesselager (1994) developed a Markov model for loss reserving in general insurance. The model that he proposed, containing four states is shown below.

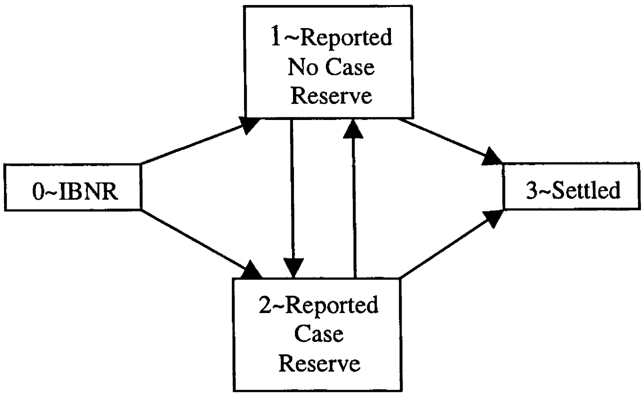


Figure 2.2 General Insurance Claim History Model

From the time when a claim occurs to the time that the final payment on the claim is made, the claim is thereafter assumed to move through the various states of the model shown in Figure 2.2. Claim occurrence leads to the introduction of the claim in state 0. Subsequently the insurer will be notified of the claim and the claim will move into either State 1 or State 2 depending on the anticipated size of the final claim. It is common for large case reserves (or case estimates) to be established in respect of potentially large claims soon after the claim is reported to the insurer. In such a case, the claim would move into State 2. If no case reserve is established when the insurer is notified of the claim, then the claim moves into State 1. It is likely that when the claim moves from the IBNR (incurred but not reported) state to one of the reported states that a (partial) claim payment will be made in respect of the claim. As the claim develops, particularly in the case of long-tail lines of general insurance business such as professional indemnity insurance or workers' compensation insurance, it is possible that the case reserve may change considerably and may even be reduced to zero. In these cases, transitions between states 1 and 2 are possible. Finally the claim will be settled, after which no subsequent payments will be made in respect of the claim. The movement of the claim to State 3 will involve a final partial payment to complete the claim payment process. Historical claim records of an insurer can provide the timing and amount of the various payments made on transition between the states of the model in Figure 2.2. These data can be used to derive both probability distributions for payments made on transitions between states and the transition intensities for movement between states. The expected future payments on a claim that occurred at time u in the IBNR state is

$$\sum_{m \neq n} \int_{u}^{\infty} y_{mn}(\xi)\, p_{0m}(u,\xi)\, \lambda_{mn}(\xi)\, d\xi, \tag{2.6}$$

where $y_{mn}(\xi)$ is the average payment made on transition from state m to state n, $p_{0m}(u,\xi)$ is the probability of a transition from state 0 to state m in the time between u and ξ and $\lambda_{mn}(\xi)$ denotes the transition intensity between states m and n at time ξ. Under the assumption of a Poisson process for claim arrivals and settlements, the transition intensities are constant and therefore the calculation of the transition probabilities between states can proceed by solving the usual Kolmogorov's equations, (see for example, Pitacco et al (1999)).

Originally, actuarial pricing and reserving theory was developed on the basis of the calculation of expected present values of cash flows. The argument put forward for relying solely on deterministic calculations was that this approach was justified for large portfolios of insurance business due to significant pooling. More recently, this approach has been justified, in some cases, by the use of stochastic models. Norberg (1996) develops ordinary differential equations for moments of present values of benefits commonly found in life insurance. Norberg (1991, 1992) defines a set of notation in relation to a multistate life insurance policy. The policyholder can reside in any of J states at time t. These J states may include healthy, sick, lapsed, totally and permanently disabled (TPD) and dead. Let I_t^j be an indicator variable for whether the policy is in state j or not at time t, and let N_t^{jk} denote the number of transitions from state j to state k during the time interval (0, t]. The payment function B generated by the policy is assumed to be of the form

$$dB_t = \sum_j I_t^j dB_t^j + \sum_{j \neq k} b_t^{jk} dN_t^{jk}, \tag{2.7}$$

where each B^j is a deterministic payment function specifying payments due during sojourns in state j and each b^{jk} is a deterministic function specifying payments due upon transition from state j to state k.

The present value of future benefits less premiums, that is the actuarial reserve, at time t is

$$\frac{1}{v_t}\int_t^n v_\tau dB_\tau, \tag{2.8}$$

where v_t is the standard discount factor. Norberg presents a theorem for finding the qth conditional moment of the present value in (2.8), given the information available at time t, that is

$$V_t^{q(j)} = E\left[\left(\frac{1}{v_t}\int_t^n vdB\right)^q \mid I_t^j = 1\right], \tag{2.9}$$

where $V_t^{q(j)}$ represents the qth quantile of the distribution of the present value of future payments under a DII contract, that is a contract that continues to make payments to this life while he resides in the disabled state, state j.

The key result is that the functions $V_t^{(q)j}$ are determined by the differential equations

$$\frac{d}{dt}V_t^{(q)j} = \left(q\delta_t^j + \mu_t^{j\cdot}\right)V_t^{(q)j} - qb_t^j V_t^{(q-1)J} - \sum_{k \neq j} \mu_t^{jk} \sum_{r=0}^{q} \binom{q}{r} \left(b_t^{jk}\right)^r V_t^{(q-r)k}. \qquad (2.10)$$

This thesis will further explore moments of present values of projected benefits less premiums in the case where the assumptions of the Norberg model do not apply – in particular in relation to disability income insurance where Markov transition intensities are invalid. Instead, semi-Markov transitions from the ill state to the able state and the associated complications for premium rating and reserving will be considered in this research.

2.1.3 Australian Industry Table

During 1995, the Graduation Sub-Committee of the Disability Committee of the Institute of Actuaries of Australia produced an industry table of claim incidence and claim termination rates based on the experience of Australian disability income insurance policyholders in the years between 1989 and 1993 inclusive. The table is now presented as an Excel spreadsheet model and is used to varying degrees by many of the large Australian life insurers which sell DII.

The Australian industry table (IAD89-93) was created for a number of reasons. The main reasons were to facilitate premium rate calculations, to automate the profit testing of a set of premium rates, to allow experience analyses, to facilitate the calculation of best estimate policy liabilities and solvency policy liabilities. Due to the relatively small amount of data available on this product during the period for which data was gathered, only aggregate (that is smoker, non-smoker and "unknown" smoker status combined) for 2 week and 1 month deferment period tables were created. For termination rates, male occupational classes A, B and C were combined, again because data from individual occupational classes at that time were too sparse to permit a reliable graduation.

The data was graduated by fitting formulae so that the chi-squared measure of goodness of fit (Benjamin and Pollard (1993)) was minimised. The formulae used for the graduation were chosen after considering the shapes apparent in the CIDA (US experience based) tables. Given that many Australian insurers, at the time of preparation of IAD89-93, used the CIDA tables as a base and then made appropriate adjustments to fit their own experience, it was decided to form the IAD89-93 table without making any reference to the CIDA tables. This approach would mean that the Australian derived table would not contain any systematic bias due to country-specific conditions and could be used by insurers independently of other internationally derived morbidity tables.

In the construction of IAD89-93, separate curves were fitted, varying with age, for male accident incidence rates and male sickness incidence rates. The formula used for accident incidence rates was a cubic function of age:

$$F(x) = A + Bx + Cx^2 + Dx^3, \tag{2.11}$$

where x represents the age last birthday at date of disablement. The accident incidence rate data is calculated using the number of accident induced new DII claims divided by the number of years that lives are observed in the able state, with a DII policy, and therefore are exposed to the risk of accident. Similarly for male sickness incidence rates, the formula adopted was

$$F(x) = A + Bx + CD^x. \tag{2.12}$$

Age was grouped into nine quinquennial age bands centred on ages 22, 27, 32, ..., 62. Separate graduations of claim incidence rates were performed for each of the four occupational classes A, B, C and D and for each of the 2 week and 1 month deferment periods. Other rating factors such as smoking status and other more unusual deferment periods had insufficient data, (for example six month deferred periods,) to warrant graduation. The report on the construction of the IAD table does provide some guidance to users on how to adjust the published rates for each of these factors, though this advice is more indicative than comprehensive.

For the determination of female incidence rates for both sickness and accident, sufficient data again was not available to warrant a reliable graduation of these rates. Instead a decision was made to adjust the male rates by a factor which varies according to age, deferment period, cause of claim (accident or sickness) and occupation class. After considerable experimentation, the formula chosen was

$$f_{c,X,o,w} = m_{c,X,o,w}.E_{c,X,o,w}, \tag{2.13}$$

where f denotes female incidence rates, m denotes male incidence rates, c is the claim cause, X is age last birthday at disablement, o represents the occupational class and w represents the waiting period (either 2 weeks or 1 month). The quantities $E_{c,X,o,w}$ are the fitted factors varying by cause, age, occupation class and deferment period. The $E_{c,X,o,w}$ factor is an adjustment factor that applies to predetermined accident incidence rates and was modelled using

$$E_{a,X,o,w} = F_{a,o,w} \text{ for } X \geq M_{o,w}, \tag{2.14}$$

where $F_{a,o,w}$ is a cubic polynomial as defined in (2.11) and for ages below $M_{o,w}$ the formula used was

$$\text{female rate} = \text{male rate at age } M_{o,w}.F_{a,o,w}\left[1-G_{a,w}\left(X-M_{o,w}\right)\right]. \qquad (2.15)$$

The rationale behind the use of Equation (2.15) for low ages was to account appropriately for the "accident bump" present in the male incidence rate curve and to allow a straight line to be fitted for ages below $M_{o,w}$. Similar adjustments were made to the male sickness incidence rates to estimate the female sickness incidence rates.

The monthly equivalent termination rates for each of the 2 week and 1 month deferment period claims were graduated using the formula

$$Y_{X,d} = A_d\left(1+B_d.X\right); d = 1 \text{ to } 6 \text{ or } 7, \qquad (2.16)$$

where $Y_{x,d}$ is the modelled value for the number of claims closing for lives aged x with claim duration d divided by the number of months that lives are aged x and disabled with duration d months and X is the age last birthday. The values of d from 1 to 6 or 7 correspond to the durations as presented in the 1995 Disability Committee Report of the Institute of Actuaries of Australia. These are the periods ending at 1, 3, 6, 12, 18, 24 and 36 months from disablement, with the 1 month duration only being relevant for the 14 day deferment period. Note that A_d values represent the monthly termination rates at age zero. Also note that the B_d factors were only fitted for duration 1 month and duration in excess of 1 month. The interaction effect between age and claim duration in determining claim termination rates was only modelled based on whether the duration was less than or greater than one month. The effect of this formula is to fit straight lines across ages at each duration.

Monthly equivalent rates are then converted to termination rates of the required duration by use of the formula

$$T_{X,d} = 1 - \left[1 - Y_{X,d}\right]^{P_d}, \tag{2.17}$$

where P_d is the number of months in the period to which the required termination rate relates. Equation (2.17) is used to convert monthly termination rates to annual termination rates, $T_{X,d}$. An annual termination rate for claim data can be derived by dividing the number of claims that close with duration d and age X into the total number of years that lives were disabled at age X and with duration d.

Partial claims are allowed for in the IAD rates by counting them as the appropriate fraction of a full claim based on the benefit amount paid as a proportion of the insured amount. The IAD table investigation indicated that there was some evidence of an anti-selection effect at younger ages and a selection effect at the middle ages. These effects were thought insignificant and were ignored in the construction of the table.

It should also be noted that a number of product actuaries for DII working in Australia have noted considerable divergence between their own experience and the rates published in the Australian IAD table over the last ten years. Waters (2004) presents a comparative analysis of the experience of UK DII policy experience using techniques from Credibility Theory. Many insurers adjust the IAD rates to conform to their own experience while other insurers continue to base their projections on the US-based CIDA tables adjusted by factors relating to age, duration and smoker status.

2.2 Actuarial Use of Generalised Linear Models

Actuaries working in both life risk and general insurance business have recently made extensive use of generalised linear models. General insurance actuaries now routinely use GLMs for pricing a range of both long and short-tail lines of business.

Brockman (1992) describes the use of GLMs in motor vehicle insurance pricing and illustrates some of the diagnostic procedures commonly employed to ensure underlying model assumptions are not seriously violated. Haberman (1996) summarises the use that actuaries have made of GLMs in a very accessible paper.

The first use of GLMs by actuaries in the life insurance domain appears to have been by Renshaw (1991). His paper, "Generalised Linear Models and Actuarial Science", describes how models traditionally used by actuaries in the graduation of mortality rates can be viewed as special cases of GLMs. The traditional Gompertz and Makeham functions for describing the variation in the force of mortality by age and the Wilkie model, for mortality, are recast as GLMs. For example, the Wilkie model for mortality is

$$\hat{q}_x = \frac{\exp(\hat{\eta}_x)}{1+\exp(\hat{\eta}_x)}, \text{ where } \hat{\eta}_x = \sum_{j=0}^{s-1} \hat{\beta}_j x^j, \qquad (2.18)$$

where $\hat{q}_x$ is the model predicted value of the one-year death probability for a life aged x. It is clear that this model related to modelling the death probability using a GLM with a binomial error distribution and logit link function – that is, to a logistic regression model.

One of the early significant mathematical models of disability rates was developed by Miller and Courant (1974). GLMs were first used in connection with disability income insurance by Renshaw. Renshaw (1995) describes the use of generalised linear modelling for graduating transition intensities in the well-known multiple state model used to describe the dynamics of disability income insurance. This model was shown in Figure 2.1.

Chapter 3 of this thesis will consider both claim incidence and claim termination rates for disability income insurance.

The 1997 Report of the Disability Committee of the Institute of Actuaries of Australia (IAAust) highlights the significance of rating factors in describing the incidence rates of disability. It was suspected that some or all of these rating factors might be significant in the explanation of termination rates, and GLMs have been employed in the next chapter to quantify this relationship in a multiple regression framework.

CHAPTER THREE

GENERALISED LINEAR MODELS AND DISABILITY INCOME INSURANCE

3.1 Motivation

[1]The common feature in all recent reports of disability experience has been the deterioration of claims termination rates. By this, we refer to lower claim termination rates leading to longer average duration for DII claims. Of course, the adequate analysis of termination rates requires very significant amounts of data. In this study, we have had access to claims data submitted to the Institute of Actuaries of Australia Disability Committee over the period 1980 to 1998, covering 106,000 claims.

This major dataset has been analysed using conventional approaches, (namely actual over expected) and using more modern Generalised Linear Modelling techniques. Also presented is a comparison of a stochastic approach to the setting of reserves for outstanding claims liabilities with the corresponding deterministic method.

---

[1] This chapter was the basis for a paper that appeared in the *Australian Actuarial Journal* (2002) – "Disability Claims – Does Anyone Recover?" This chapter also formed the basis of an invited presentation that was given at the Institute of Actuaries of Australia Biennial Convention in 2003.

3.2 Data

The Institute of Actuaries of Australia Life and Risk Committee supplied the Australian National University with their claims database for the period 1980 to 1998. This database had all identifiers relating to company or policy removed before access was given to the university for research. The database contained records relating to 106,000 individual claims and their various characteristics.

Each claim record contained the following information in relation to the claimant

Rating Variable	Rating Variable
Country of purchase of the DII contract	Benefit Rate
Disability Definition	Benefit Type
Gender	Medical Evidence
Deferment – Accident	Coverage Type
Deferment – Sickness	Contract Type
Date of Policy Commencement	Cause of Claim
Benefit Period – Accident	Smoker Status
Benefit Period – Sickness	Benefit Proportion
Occupation	Date Disability Commenced
Date of Birth	Date Claim Ceased
Expiry Age	

Table 3.1 Table of Covariates for GLM Analysis

Some of these characteristics were summarised to reduce the number of variables and without overly sacrificing data quality. The summarised characteristics were

Summarised Rating Variables
Deferment
Claim cause
Coterminus benefits
Proportionate benefit
Benefit rate
Age at claim

Table 3.2 Summarised Rating Variables

Note that these summarised rating variables provide a simpler way of dealing with the variables given in Table 3.1. When the data was prepared, a matrix was used where lives that have a DII contract with the same characteristics are grouped together. Total exposure and total number of recoveries for these lives are then calculated. The models that follow in this chapter are developed using the rating factors given in Table 3.1.

The detailed specifications used to determine these summarised characteristics are set out in Appendix 3.1. Certain records were excluded from the analysis due to missing data or in order to match the primary data selection criteria used by IAAust Disability Committee. The exclusion criteria are also set out in Appendix 3.1. After these records were excluded, 101,000 claims with 875,000 months of exposure to the risk of returning to work remained for subsequent analysis.

Not all claims remained in the original Life and Risk Committee database from commencement to termination. Some claims commenced prior to 1980, other claims remained open at the end of 1998 and some companies only contributed data during some of the years between 1980 and 1998. Nevertheless each claim contributed to the exposure for the months for which data was available for it and contributed to actual terminations only if it terminated in the time period when data was available for it. Claim cessation due to death, recovery or lump sum payment was defined as a claim termination. Claim cessations due to benefit expiry were not treated as a claim termination. This choice was made because our aim is to develop a model of claim termination rates. Claim durations were taken at monthly intervals from the date disability commenced regardless of the deferment period. The first month was labelled as duration zero. In the actual over expected analyses, only benefits classified as "Full" were included, except for the comparison of "Partial" and "Full" benefits.

3.3 Comparison of Actual vs Expected Claim Terminations

The IAAust Life and Risk Committee most comprehensively reports claim termination experience by using average duration of claims for the first three years of claims. This approach is reasonable given that each report concentrates on the four years covered. However, as noted in those reports, average claim duration needs to be interpreted with care as it can change due to fluctuations in the level of new claims volumes without concomitant change in underlying termination rates.

During this chapter, we have concentrated on actual over expected claim terminations (by claim numbers) as the measure of termination rate experience. For the calculation of expected terminations we have used IAD89-93.

The results are presented as an index which is the reciprocal of the actual over expected ratio so that as results deteriorate, the ratio increases.

The data used in subsequent analysis includes only claims that satisfied the following criteria:

- Individual coverage (excluding Business Overheads); and
- Contract type not Cancellable; and
- Not a partial benefit at any point in the claim payment history.

Detailed actual over expected results are set out in Appendix 3.2. The more important conclusions are summarised in the following. In this section only a one-dimensional view is taken of the experience, as there are simply too many possible cells to allow a reasonable multidimensional analysis to be conducted using actual over expected techniques. However, the section dealing with the generalised linear model does, of course, allow for a multivariate approach.

It will be noted that the data for the period 1989 to 1993 does not have an index of 100%; that is, the data does not agree with IAD89-93 for the same period. This discrepancy is due to three sources of difference between the data provided to the university and the data originally used by the IAD Graduation Sub-Committee of the Institute of Actuaries of Australia.

- The data provided to the authors was of later origin than that used to derive IAD89-93 owing to subsequent revisions and submission of additional data by contributors;

- The analysis in this paper has been based on summarised data for certain variates; and

- IAD89-93 uses an "artificial" approach to setting termination rates at longer durations, in particular a simple extrapolation of the trend in termination rates is used.

Since the comparisons used here are based on the relative values of the "index", the small difference between our 1989 to 1993 results and IAD89-93 is not significant in the interpretation of the results.

3.3.1 Experience over Time

In this analysis, the "year" refers to the year of exposure (and claim termination). The results from Table 3.3 are displayed using a scatterplot along with a lowess smooth of the points in Figure 3.1 below.

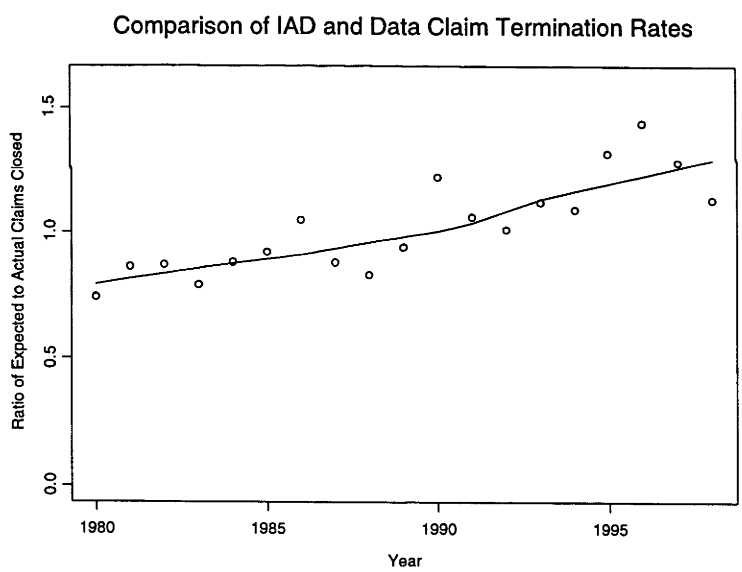


Figure 3.1: Comparison of Actual and Expected Experience over Time where expected experience is derived using the IAD89-93 table

This analysis confirms the general deterioration of experience over time but interestingly suggests that in the last two years under study (1997 & 1998) some improvement may be evident. Is this a real improvement or, like the periods 1986 to 1988 and 1990 to 1992, merely a random fluctuation before the presumed underlying deterioration resumes its course?

Because of this very material change in aggregate experience over time, the results of the analyses by each of seventeen covariates are presented in a two dimensional form to show the experience over time as well as how it varies with the value for each individual characteristic. The only exceptions to this presentation style are experience by benefit size, age at claim, duration and year of policy commencement as the number of individual values for each characteristic are too great to allow such an exposition.

3.3.2 Summary of Results

Year	Expected	Actual	Index
1980	19	26	74%
1981	79	92	86%
1982	226	261	87%
1983	515	651	79%
1984	629	715	88%
1985	763	826	92%
1986	1202	1146	105%
1987	2090	2381	88%
1988	2142	2580	83%
1989	2352	2513	94%
1990	3261	2681	122%
1991	4895	4625	106%
1992	5148	5082	101%
1993	6447	5736	112%
1994	6960	6369	109%
1995	7737	5840	132%

1996	8166	5658	144%
1997	7490	5862	128%
1998	5477	4830	113%
	65600	**57874**	**113%**

Table 3.3: Comparison of Actual and Expected Experience over Time

Table 3.2 shows the results for each characteristic presented as the Index for the total period ignoring differences by year. The experience by year for each characteristic is to be found in Appendix 3.2. This table identifies which values for which characteristics are material in impacting the termination experience.

Characteristic	**Value**	**Index**
Gender	Male	112%
	Female	125%
Occupation	A	126%
	B	111%
	C	107%
	D	111%
Deferment	7 days	66%
	14 days	108%
	30 days	125%
	90 days	245%
Definition	Own / Any 2 years	110%
	Own	126%
	Any	95%
Benefit Type	Level	105%
	Increasing	119%
Medical Evidence	Medical	83%
	Non Medical	101%

	Other	126%
Coverage	Individual	113%
	Business Overheads	126%
Contract Type	Level – Guaranteed	114%
	Level	109%
	Stepped – Guaranteed	97%
	Stepped	116%
	Cancellable – Level	85%
	Cancellable – Stepped	103%
No Claim Bonus	No	114%
	Yes	113%
Smoker Status	No Differentiation	107%
	Non Smoker – Checks	107%
	Non Smoker	113%
	Smoker	127%
Claim Cause	Unknown	115%
	W	69%
	X	147%
	Y	190%
	Accident	101%
CoTerminus	Yes	112%
	No	117%
Benefit Period	2 years	102%
	5 years	115%
	Expiry	127%
	Lifetime	123%
Benefit Proportion	Full	113%
	Partial	196%

Table 3.4 Index of Termination Rates by Characteristic

3.3.3 Some Initial Conclusions

The results summarised by Figure 3.1 and Tables 3.3 and 3.4 suggest that the characteristics used in IAD89-93 to predict claim termination rates, namely deferment, gender and occupation, do not capture significant differences in experience that are evident due to the presence of other rating factors. This suggests that the classification using the current industry table contains considerable heterogeneity. Indeed, usual industry practice in the application of the IAD89-93 table for Australian life insurers varies considerably. Some insurers apply a proportion (rating up or down) factor to the rates provided in IAD89-93. The proportion is calculated using a comparison of aggregate experience to the rates published in the IAD industry table. Other Australian insurers disregard the Australian table on the basis that it was constructed using insufficient experience. They instead base their rates on the CIDA85 tables produced in the United States and then apply a set of internally developed factors relating to age, smoker status and occupation class to bring the US table rates into line with the particular office experience.

In addition, not only is the experience deteriorating to an extent where IAD89-93 is materially overstating the likely termination rates, but also its shape for various characteristics may be materially and commercially significantly different to that shown in the experience.

3.4. GENERALISED LINEAR MODEL OF CLAIM TERMINATION RATES

3.4.1 Background to Generalised Linear Models

Generalised linear models (GLMs) were first developed by Nelder and Wedderburn (1972). GLMs extend the basic linear regression model in a number of critical aspects. The standard linear regression model, when used for the prediction of a dependent variable, Y (for example claim termination rate), with a number of independent variables, $X_1, X_2, \ldots, X_p$, (for example occupation class, duration of claim, age, smoker status) can be described as follows:

1. A random component: each value of Y is normally distributed with expected value μ (which may depend on covariates) and constant variance σ^2.

2. The systematic component: a set of independent variables $X_1, X_2, \ldots, X_p$ which combine to produce a linear predictor η given by $\eta = \sum_1^p x_j \beta_j$, where the β_j are regression coefficients, usually estimated by the least squares principle.

3. The "link" between the linear predictor and the mean of Y is $\mu = \eta$

The GLM extends this basic model in two significant ways:

1. The dependent variable, Y, may come from an exponential family distribution rather than only the Normal distribution. Exponential family distributions includes most of the common probability distributions used by actuaries in general insurance, and includes the Normal, Poisson, Binomial, Gamma and Inverse Gaussian distributions as special cases. The advantage of using this broader class of distributions for the response variable is that it allows the flexibility of a wider range of possible relationships between the variance and the mean of the dependent variable, and allows for features of the underlying error distribution such as asymmetry, heavy tails, etc.

2. The link function in (3) above can be any monotonic differentiable function. Common link functions include the identity link, the log link, the inverse link and the logit (or log-odds) link. The link function allows non-linearity in the relationship between Y and the covariates.

Having fit a GLM, the process of determining whether predictors are adding value to the model is also different to the process used in multiple regression or traditional linear modelling, insofar as the so-called deviance is used as a measure of fit rather than the more traditional "variation-explained".

To assess the adequacy of the fit of a GLM we need to define the deviance statistic. The deviance reflects the discrepancy between the actual values of the dependent variable and the fitted values relating to the dependent variable in a likelihood sense. The formula for the deviance is

$$D\left(\hat{Y},Y\right)=2\phi\left[l\left(Y,\phi\right)-l\left(\hat{Y},\phi\right)\right] \tag{3.1}$$

where ϕ is the dispersion parameter and $l\left(\hat{Y},\phi\right)$ is the log-likelihood function for the observed values.

The deviance is the scaled difference in log-likelihoods between a perfectly fitting (or saturated) model and the model for which the deviance is being calculated.

The addition of independent variables to a GLM inevitably reduces the deviance. The amount of this reduction in deviance (in other words the size of the step taken towards the perfectly fitting model) may be used as a measure of whether that particular independent variable should be retained in the model.

3.4.2 Advantages of Using Generalised Linear Models for Claim Termination Rates

Some advantages of using GLMs in describing termination rates are that:

- they enable the impact of changing the level of a rating factor to be quantified in a way that does not consider the change in other rating factors that will accompany such a change in rating variable. For example, the 1997 Disability Report considers the impact of occupation class on claim duration. The report states that claim duration for occupation class A is longer than that for occupation classes B to D. The GLM enables you to isolate out the impact of the change in occupation class from the changes in other variables which occur when you move from occupation class A to D. For example females are rarely in occupation class D, and secondly occupation class A lives are usually more severely disabled before they are unable to work than is the case for, say, occupation D lives; note, however, that such interpretations may struggle fo find real-world analogues, as changes in individual covariates often occur.

- they allow the calculation of predicted value of the termination rate for a particular life with a specific set of rating variables. GLMs also permit the calculation of the variance of fitted values;

- they enable suitable modelling of the variance of the expected termination rate (unlike conventional linear modelling) to more adequately reflect real-world experience;

- the predicted termination rates vary smoothly meaning that premiums and reserves calculated from the model will also vary smoothly for continuously varying covariates. They won't necessarily be smooth for, say, binary covariates.

- the results for termination rates can be summarised as the result of a single model rather than through the provision of a large set of different tables. This compactness allows for easy calculation and also provides a very efficient means of communicating such information.

3.4.3 Discussion

The GLM fitted in this research used data from the whole period 1980 to 1998 but using only a limited set of characteristics in order to facilitate the actual calculations by reducing the number of cells analysed. A GLM fit to the full set of characteristics using only data from the latest data period, 1995 to 1998, has been derived for claim incidence rates. The results of this analysis are shown later in this chapter.

The fitting of a suitable GLM requires multiple choices as regards the distribution of the errors, the link function, the predictors to use and whether any transformation of those predictors is desirable. In addition, it is worthwhile to also investigate interactions which may exist between the predictor variables.

A number of different GLMs were fit to the termination rate data. After considerable experimentation and assessment of residual plots, it was decided that the Poisson error structure was the most appropriate for describing the mean-variance relationship inherent in the data. It was comforting to note that this error family is the same one commonly used for claim frequencies when the claims experience of many short-tail lines of general insurance business are modelled.

The data selection for the final GLM chosen used claims which had a deferment of 14 or 30 days, did not have an Unknown claim cause and had Individual coverage. There were 83,000 such claims with 675,000 months of exposure. The following characteristics were retained in the final model:

Rating Variables	Rating Variables
Definition	Claim Cause
Gender	Deferment Period
Occupation Class	Benefit Rate
Smoker Status	Year of Exposure
Age at Claim	Claim Duration

Table 3.5 Retained Covariates

There were 275,000 cells in the data, after allowing for all multi-way classifications of the above ten rating variables.

It should be noted that unlike for the actual over expected analysis, the data used to fit the GLM included both "Partial" and "Full" benefit claims.

In order to give greater weight to those observations in which we have more confidence in, the GLMs were fit using the exposure for each rating factor combination as weights in the fitting algorithm.

The results in Appendix 3.3 show that all of the following rating factors aid significantly in the modelling of termination rates: age at date of claim, cause of claim, duration of claim, gender, occupation class, smoker status, deferment period, benefit rate and the calendar year at the date of possible claim termination. In addition a significant duration and age interaction was found. The dataset includes all claim terminations so that at every possible duration only certain claims will actually be observed to terminate.

From the output in Appendix 3.3, we can calculate the fitted values for claim termination rates. The prediction formula is

$$\text{Fitted Claim Termination Rate} = \exp(32.3 - .008\text{AgeClaim} + 0.371(\text{ClaimCauseW}) + \ldots - .00652(\text{AgeClaim}*\text{Sqrt}(\text{Duration})))$$

Standard errors of the fitted coefficients are provided in Appendix 3.3. It is clear from Appendix 3.3 that the model retains a significant residual deviance. This result is not surprising given that the nineteen years of data was included in the analysis and only a limited number of rating factors were included as explanatory variables. The interpretation of deviance residuals is explained in McCullagh (1989). A number of other variables and transformations of existing variables have been explored in the context of claim incidence rates.

Graphical and other comparisons of modelled termination rates versus actual data points are given for the claim incidence rates presented later in this chapter. This analysis will build a model for claim termination rates (and incidence rates) based only on the most recent five years of claim data. This reduced dataset will allow a considerably improved fit to the data to be achieved than was possible in the full analysis.

Other interesting results of the analysis include:

- The benefit rate has a statistically significant, negative impact on the rate of claim termination;

- there is a statistically significant interaction between duration and deferment period in determining the termination rate. At shorter durations the predicted termination rate is significantly higher for the shorter deferment period. After durations of approximately 8 months this difference becomes non-significant;

- there has been a statistically significant decline, over time, in termination rates when aggregated across all levels of the rating factors. The decline in termination rates is still statistically significant even after the impact of all other rating factors has been allowed for.

An alternative to the second method of fitting an interaction term between deferment period and duration is to use "break-point predictor terms". This has been employed successfully for the UK data by Renshaw(1995). The idea is to include terms of the form $(\text{Duration} - 3)^+$ which are only positive if the duration is greater than 3 and otherwise are zero. Such terms enable the rating factors to exhibit a non-constant linear relationship (after allowing for the link function) with the termination rate. They prove useful in modelling the lower termination rates that one observes at the very shortest durations.

3.5 STOCHASTIC CLAIMS RESERVING

Using the GLM output shown in Appendix 3.3, we aim to find the approximate distribution, using simulation, of the required reserves in respect of a DII portfolio to ensure relevant probabilities of adequacy. The analysis is for the average of 100 claims that are new at the date of valuation with a monthly disability payment of $2600. The graph below was generated for the case of a male aged 40, in occupation class A, with a deferred period of 2 weeks who became disabled because of an accident.

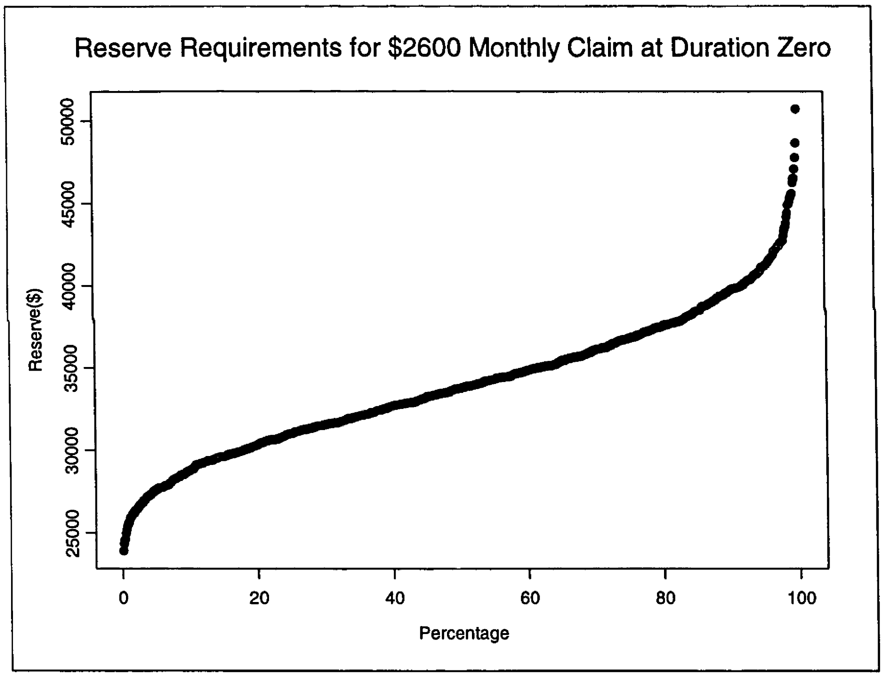


Figure 3.2: Reserve Requirements to obtain varying probabilities of adequacy

In the above analysis, allowance has been made for deteriorating claim termination rates and for interest at 6% per annum, compounding continuously.

The calculation above requires the use of two probability distributions. First, note that the fitted value for the natural log of the termination rate from the GLM is asymptotically normally distributed. This is because the fitted value is a linear combination of coefficient estimates and constant covariate values and the coefficient estimates are maximum likelihood estimates, which are themselves asymptotically normally distributed. The delta method was used to approximate the variance of the actual fitted values for the termination rates. The number of terminating claims, given the termination rate, was then simulated from a binomial distribution.

Table 3.6 shows the reserves required to give a particular probability of the reserve being adequate using the simulation approach described above.

Probability	**Claim Reserve at Start of Claim**	**Increase Over 50%**
50%	33,764	
75%	36,749	9%
90%	39,692	18%
95%	41,180	22%
99%	45,357	34%

Table 3.6: Reserve Requirements by Probability of Adequacy

The reserve calculated using a traditional deterministic approach would be approxiamtely at the 50% probability level in Table 3.6.

3.6 A GLM for Claim Incidence Rates

A GLM with Poisson error and logarithmic link function was fitted to the claim incidence rate data. The final model chosen was of the form

$$E\left[\log(\mathrm{Y})\right] = \log(\text{exposure}) + \beta_0 + \sum_{j=1}^{35} \beta_j x_j + \sum_{\text{selected } i,j;k=1}^{13} \beta_k x_i x_j, \tag{3.2}$$

where Y denotes actual number of claims multiplied by benefit percent, the x_j are covariates associated with holders of DII contracts and the β_j are regression coefficients derived using maximum likelihood estimation (McCullagh and Nelder (1989)). The exposure is the amount of time that disability income insurance holders are exposed to the risk of becoming disabled and is recorded in days. The $x_i x_j$ terms are interaction terms. These will be discussed further in this section.

This model is fit using an offset for log(exposure). The offset ensures that the model is fit with a regression coefficient for log(exposure) equal to one. This means that the model can be rewritten as

$$E\left(\log(\text{IncRate})\right) = \beta_0 + \sum_{j=1}^{35} \beta_j x_j + \sum_{\text{selected } i,j;k=1}^{13} \beta_k x_i x_j, \tag{3.3}$$

where IncRate denotes the incidence rate of claims measured on a per day basis.

The coefficients of the fitted model for each variable and for the statistically significant interaction terms are shown below in Table 3.4. A description of each covariate is given in Appendix 3.1. The fitted intercept for the model is –5.281746.

Covariate	Coefficient Estimate	Z-Score	Covariate	Coefficient Estimate	Z-Score
Gender	0.274745	2.93	Benamount6	0.211641	6.10
Age	0.246015	15.81	Benamount7	0.254199	5.52
$\sqrt{\text{Age}}$	-2.310697	-18.63	Smoker	0.156685	10.61
OccupationB	0.921944	7.49	Aids	0.222879	15.04
OccupationC	1.805061	22.77	Duration1	0.207573	12.56
OccupationD	2.000999	23.57	Duration2	0.083477	5.89
Definition2	-0.058871	-4.03	Ncb	0.067367	4.69
Definition3	-0.043319	-0.89	Contract1	-0.321532	-4.76
Definition4	-0.219713	-4.58	Contract2	-0.078040	-3.99
Definition5	-0.356431	-8.86	Contract3	0.149045	3.06
Definition6	-1.158268	-1.46	Medical	-0.205617	-6.11
Deferment2	0.906221	1.20	Age*OccupationB	-0.009219	-3.38
Deferment3	-0.048222	-0.06	Age*OccupationC	-0.019752	-11.39
Deferment4	-2.540355	-3.09	Age*OccupationD	-0.021182	-11.42
Deferment5	-2.612055	-3.20	Gender*Age	0.004870	2.40
Deferment6	-4.371066	-3.14	Gender*OccupationB	-0.175858	-2.95

Deferment7	-3.533438	-1.52	Gender*OccupationC	-0.474074	-9.73
Deferment8	-3.333802	-6.72	Gender*OccupationD	-0.563529	-7.08
Benperiod1	-1.078854	-13.62	Age*Deferment2	-0.025920	-1.96
Benperiod2	0.0196288	1.39	Age*Deferment3	-0.023097	-1.75
Benamount2	0.044054	2.20	Age*Deferment4	0.011881	0.80
Benamount3	0.174378	9.06	Age*Deferment5	0.002877	0.20
Benamount4	0.200430	9.88	Age*Deferment6	0.027720	1.08
Benamount5	0.235900	8.73	Age*Deferment7	-0.011860	-0.25

Table 3.7 Incidence Rate GLM Coefficients

The above table shows that thirteen interaction terms are included in the model for claim incidence rates. Interaction terms add to the flexibility of the model and enable a more realistic description of the underlying data. To consider an example, we see in the above table that the gender by age interaction is statistically significant. This means that the impact of age on incidence rates is different dependent on the value taken by gender in the model. The gender by age interaction term has a positive coefficient meaning that for a unit increase in age the resulting predicted incidence rate increases more when gender is 1 (female) than when gender is 0 (male).

The Poisson error structure employed in this model models the variance to increase in line with the size of the fitted value (mean). A diagnostic check was conducted on the model results to ensure that this proportional increase in variance with the fitted values was a reasonable assumption for this data. The test indicated that there was no evidence of overdispersion.

Another important characteristic of this analysis was the treatment of the smoker variable. In the original data, approximately 3.5% of the exposure was not classified as either smoker or non-smoker. One approach to the modelling is to ignore this 3.5% of the data. It is not possible to fit the GLM using a subset of the required covariates for certain data points unless a specific imputation technique is put into place. One imputation method is, therefore, to attempt to predict the value of the missing smoker value by employing logistic regression with smoker as the response. This analysis fits a regression model where the response variable is the probability that the policyholder was a smoker and the explanatory variables are the other covariates including claim information available in the data. The logistic regression methodology ensures that the fitted values for the probability of being a smoker are on the range from 0 to 1 and hence can be used as valid covariates in the fitted GLM.

3.7 Comparison of GLM Results with IAD89-93

An analysis of the main differences between both the smoothed incidence rates, fitted by the GLM from Section 3.6, and the existing rates from IAD89-93 is given in this section.

Table 3.8 compares the smoothed annual incidence rates using the GLM from Section 4 and the rates from IAD89-93 for 2-week deferment period policies.

	Occupation A	Occupation B	Occupation C	Occupation D
Age 22 (IAD)	1.8889%	4.0643%	5.4911%	5.7715%
Age 22 (GLM)	1.6661%	3.3402%	6.4850%	7.5635%
Age 27 (IAD)	1.7401%	3.4906%	4.9069%	5.3417%
Age 27 (GLM)	1.5625%	2.8803%	5.4560%	6.2806%
Age 32 (IAD)	1.8441%	3.2985%	4.8465%	5.4489%
Age 32 (GLM)	1.6318%	2.8408%	5.1082%	5.8771%
Age 37 (IAD)	2.1474%	3.4744%	5.1656%	5.9374%
Age 37 (GLM)	1.8429%	3.0734%	5.2040%	5.9217%
Age 42 (IAD)	2.6645%	4.0039%	5.8266%	6.7658%
Age 42 (GLM)	2.1955%	3.5029%	5.5956%	6.3472%
Age 47 (IAD)	3.4226%	4.8797%	6.7981%	7.8987%
Age 47 (GLM)	2.7603%	4.1802%	6.2835%	7.0909%
Age 52 (IAD)	4.4471%	6.1014%	8.0550%	9.3069%
Age 52 (GLM)	3.6175%	5.2749%	7.4288%	8.3044%
Age 57 (IAD)	5.7930%	7.6759%	9.5784%	10.9671%
Age 57 (GLM)	4.9656%	6.6699%	8.9480%	10.1073%
Age 62 (IAD)	7.5205%	9.6170%	11.3560%	12.8620%
Age 62 (GLM)	7.0301%	8.8306%	11.2966%	13.0054%
Age 67 (IAD)	9.4457%	11.7093%	13.2333%	14.8477%
Age 67 (GLM)	10.2979%	12.4523%	14.8078%	17.2813%

Table 3.8 IAD vs GLM incidence rates for two-week deferred period

Table 3.9 compares the smoothed incidence rates using the GLM from Section 3.6 and the rates from IAD89-93 for the one-month deferment period policies.

	Occupation A	Occupation B	Occupation C	Occupation D
Age 22 (IAD)	0.5856%	1.1248%	2.2204%	2.2662%
Age 22 (GLM)	0.6874%	1.3810%	2.6977%	3.2659%
Age 27 (IAD)	0.5358%	1.0037%	1.9808%	2.1374%
Age 27 (GLM)	0.6578%	1.2108%	2.2889%	2.7232%
Age 32 (IAD)	0.5947%	1.0333%	1.9486%	2.1758%
Age 32 (GLM)	0.6861%	1.1913%	2.1743%	2.5688%
Age 37 (IAD)	0.7290%	1.1581%	2.1002%	2.3628%
Age 37 (GLM)	0.7853%	1.2686%	2.2314%	2.6056%
Age 42 (IAD)	0.9400%	1.3428%	2.4340%	2.7153%
Age 42 (GLM)	0.9504%	1.4557%	2.4194%	2.8215%
Age 47 (IAD)	1.2517%	1.6247%	3.0144%	3.3107%
Age 47 (GLM)	1.2068%	1.7347%	2.7519%	3.1868%
Age 52 (IAD)	1.6995%	2.1345%	3.9507%	4.2741%
Age 52 (GLM)	1.5995%	2.1383%	3.2830%	3.7740%
Age 57 (IAD)	2.3454%	3.0658%	5.4235%	5.7822%
Age 57 (GLM)	2.2051%	2.7362%	4.0876%	4.6466%
Age 62 (IAD)	3.2603%	4.6743%	7.6233%	8.0478%
Age 62 (GLM)	3.1133%	3.4621%	5.1661%	5.9687%
Age 67 (IAD)	4.3157%	6.5442%	9.9974%	10.6651%
Age 67 (GLM)	4.4394%	4.6979%	6.9509%	7.4860%

Table 3.9 IAD vs GLM incidence rates for one month deferred period

The key message from the above comparison is that the incidence rates predicted by the GLM and the incidence rates in IAD are not materially different. Nevertheless, some differences are evident that are worthy of comment.

The difference between the IAD and the GLM rates is greatest at age 67 in the above table however, of course, the amount of exposure at this age group is very small. It should also be noted that a global comparison between the IAD 89-93 table and the rates predicted by the GLM cannot be made from the above table, since this would involve consideration of incidence rates for particular classes of business with different levels of exposure.

3.8 Goodness of Fit Analysis

In this section the fit of the GLM described in Section 3.6 is studied. A simple check of the goodness of fit is achieved by comparing the crude incidence rates with the rates predicted from the GLM. A chi-squared goodness of fit test on a three-way table of data is given. The table used includes age, occupation and is for males. The data in the table is aggregated across all other rating variables employed in the model. The chi-squared goodness of fit test provides a useful check as to the adequacy of the fit of the model. It is commonly used to assess the adherence of crude mortality rates to modelled mortality rates in the construction of life tables both in Australia and overseas.

Table 3.10 shows the value of the fitted minus the crude incidence rates for the GLM described in Section 3.6.

	Occupation A	Occupation B	Occupation C	Occupation D
Age 22	-0.73%	-1.52%	-1.47%	-1.80%
Age 27	-0.06%	-0.20%	-0.41%	0.10%
Age 32	-0.15%	-0.19%	0.16%	0.46%
Age 37	-0.18%	-0.14%	0.15%	0.01%
Age 42	-0.15%	-0.27%	0.12%	-0.42%
Age 47	-0.24%	0.03%	0.14%	0.51%
Age 52	-0.13%	0.17%	-0.04%	-0.58%
Age 57	-0.13%	0.12%	-0.07%	-0.24%
Age 62	-1.75%	-2.43%	-2.83%	2.03%
Age 67	2.75%	-2.06%	3.47%	2.45%

Table 3.10 Goodness of Fit Analysis

It is clear from the above table that the fitted rates at the higher ages do not adhere as closely to the crude rates as at younger ages. This phenomenon is partly due to the lack of data at the higher ages and therefore the greater volatility in the reported claim incidence rates at these ages. In the construction of IAD89-93, the authors note that the higher ages were often ignored in the fitting process and that constraints on the fitted values were used to ensure that the more volatile rates at these ages did not affect the fitted rates too much.

A global test of the fit of the model using a chi squared test indicates that the fitted rates generated by the GLM adhere sufficiently closely to the crude rates, though some differences may be of practical significance to businesses using IAD.

CHAPTER FOUR

ANALYSIS OF CLAIM TERMINATION RATES

This chapter investigates statistical models for the description of claim termination rates. A claim is considered to be terminated when the claimant returns to work. Other sources of claim termination include changes to benefit levels, expiry of the benefit period, the payment of a lump sum, and death. One of our aims is to derive premium rates and to develop appropriate reserving methodology, within a multiple state model framework, and so it is only the rate of return to work that is used when determining claim termination rates. The number of deaths in the dataset is very small because the population under consideration are mostly of working age. Mortality rates will therefore be based on standard life tables with suitable adjustments rather than the smoothed mortality rates implied by the dataset. This approach will be discussed further in Chapter Six.

4.1 Data

Claims which began in 1995 were extracted from the IAAust database. These claims were followed until termination or the end of calendar year 1998, whichever occurred first. The table below gives a description of the rating variables which were included in the data along with the name of S-Plus variables that were created for use in the statistical modelling.

Field	Description	Variables (S-Plus Names)
Duration	Duration of the claim (recorded in days). This is the number of days from when the sickness began until recovery (or censoring), less the deferment period.	durn2
Age	Age at the date of claim commencement	age
Terminate	An indicator of whether the claim was observed to terminate or was censored	terminate
Disability Definition	Own occupation for which the insured person is reasonably suited by education, training or experience, or any occupation after an initial period. (Indicator variable for any occupation after initial period)	poldesnew3
Sex	Indicator variable for gender; Male = 1.	sex1
Occupation Class	Occupation is grouped into four levels: A, B, C or D as described in IAAust Disability Reports	occupB, occupC, occupD
Frequency of Benefit Payment	Classified as (1) weekly, (2) monthly or (3) annually	benhp1, benhp2
Benefit Rate	Monthly benefit rate in dollars	benrate
Benefit Type	Level or Increasing Benefits. (Indicator variable for increasing benefits)	bentypnew2
Medical Evidence	Medical Exam required or Automatic Acceptance. (Indicator for medical exam required)	medevid1
Contract Type	Level Premiums or Stepped Premiums. (Indicator variable for Level Premiums)	conttypenew1
Smoker Status	Smoker or non-smoker. (Indicator variable for smoker)	smokernew
Sickness or Accident	Sickness claim or Accident related claim. (Indicator is for sickness)	sick
Deferred Period	Classified according to defpd0 (0 day), defpd1 (base level and deferment period between 1 and	defpd0 defpd2

	27 days), defpd2 (28 to 89 day deferment period) and defpd3 (deferment period in excess of 90 days)	defpd3

Table 4.1 Potential Covariates for Claim Termination Rate Analysis

Of the 8863 claims recorded, 7771 (88%) related to terminated claims, the remainder being censored. The most common cause of censoring was that the claim reached the end of 1998 and was continuing at that time. There were a small number of claims that were lost at the end of each of 1995, 1996 and 1997 and that are unable to be followed further. This issue arose due to changes in claim codes adopted by companies that provided this data to the IAAust Life and Risk Committee at the end of particular calendar years. Most of these claims were able to be traced by matching claims from one calendar year to the next on the basis of date of birth, date of entry to the policy, sex, occupation class and smoker status, however a small proportion (less than 1%) were unable to be successfully matched.

The age profile of claimants ranged from 17 to 70 with an average age of 40. The distribution of ages for new claimants was approximately bell shaped.

Of the 8863 claimaints included in the dataset, 2409 (27.2%) were in occupation class A, 667 (7.5%) were in occupation class B, 3165 (35.7%) were in occupation class C, and 2622 (29.6%) were in occupation class D. (Report of the IAAust Disability Committee, 1997).

Just over 50% of the claims related to disability definitions where any occupation applies in determining whether the claim can continue after an initial period.

Males account for 87% of the data, while monthly benefit payments are clearly the most common, also accounting for 87% of the data. Note also that 54.7% of the claimants had chosen benefits that increase in line with inflation. Only 5% of the claimants would have required thorough medical examinations before claim payments commenced. Level premiums accounted for 13.6% of the data, the remainder relating to stepped premiums. The smoker prevalence rate amongst claimants was 19.5%. Sickness caused 58.9% of the claims, the remainder being due to an accident.

4.2 Kaplan-Meier Analysis of Claim Durations

In order to understand the duration profile of disability claims, Kaplan-Meier (see Kaplan and Meier, 1958) survival curves have been created for the continuation of disability claims. Kaplan-Meier curves can be used to provide a non-parametric estimate of the survival function for claims. The event of interest in this survival analysis is clearly claim termination. The duration variable is used to measure time since claim onset, and not time since payment of disability benefits begins.

Immediately apparent from Figure 4.1 is the drop in claims in force after 730 days; that is, after two years. This issue was investigated and claims which cease due to the expiry of a two-year benefit period were not included in this analysis. It is suspected, therefore, that a small proportion of claims that cease after two years are recorded as recoveries, when in fact they relate to the expiry of the benefit period. The effect is negligible and subsequent analysis proceeds using the data as presented in Figure 4.1.

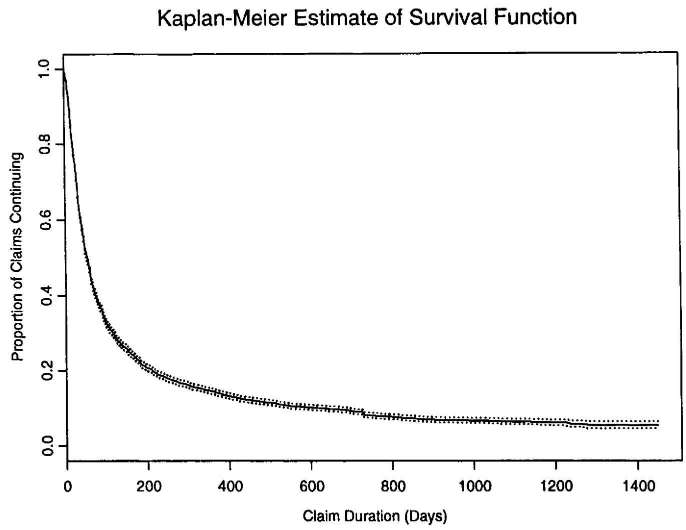


Figure 4.1 Kaplan-Meier Estimate of the Claim Duration Survival Function

The above graph includes 95% confidence intervals for the estimated survival function. From the Kaplan-Meier analysis we note that,

- there appears to be a non-zero long term survival probability of about 0.07. This probability relates to lives who do not recover from their disability; and

- the Kaplan-Meier estimate of the survival function is very smooth. This suggests that parametric survival function models may work well in this context.

The results of an initial investigation of the impact of the various rating factors outlined in Table 4.1 on claim termination rates are now presented. Again, Kaplan-Meier estimation is used. The Australian industry table for disability income insurance claim rates (IAD89-93) uses the following rating factors: age, sex, occupation class, deferment period and smoker status. Kaplan-Meier estimates of the survival function are created for each level of these rating variables, excluding age which is essentially a continuous variable.

Note that the Kaplan-Meier plots shown in Figures 4.2 to 4.5 represent one-way analyses of claim duration experience observed from 1995 to 1998 inclusive.

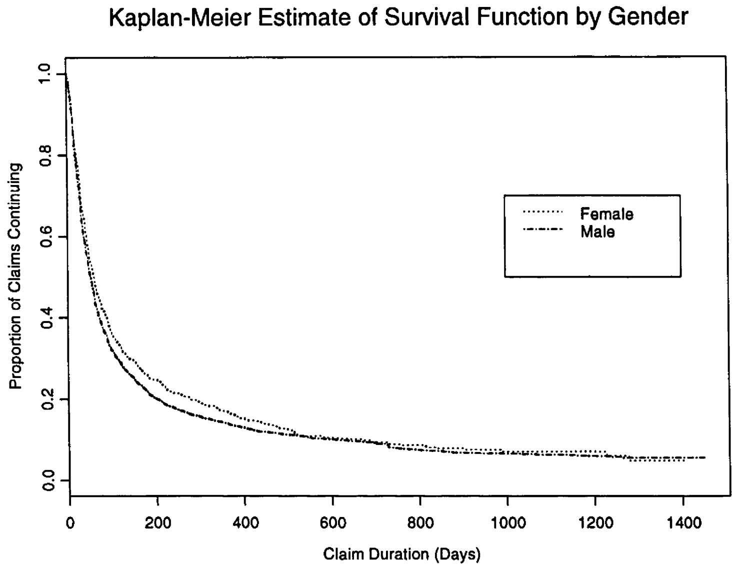


Figure 4.2 Kaplan-Meier Survival Function Split by Gender

The estimated survival functions for males and females are very close with mild evidence that males have higher recovery rates than females between six months and one and a half years after onset of disability, but that long term there is very little difference.

The Kaplan-Meier estimates by occupation class indicate that occupations can be grouped into two groups, "A and B" compared with "C and D". Occupation class definitions are given in the 1997 Report of the Disability Committee (TIAA, 1997).

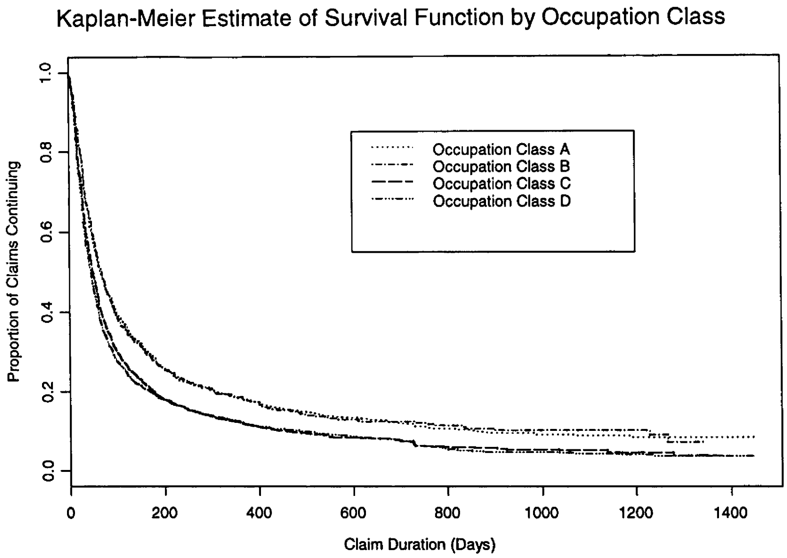


Figure 4.3 Kaplan-Meier Survival Function Split by Occupation Class

The most noticeable feature of the Kaplan-Meier estimates by deferred period is the significantly larger long term claim probability associated with the longest (greater than three months) deferred period group. There is also evidence of longer claim durations amongst

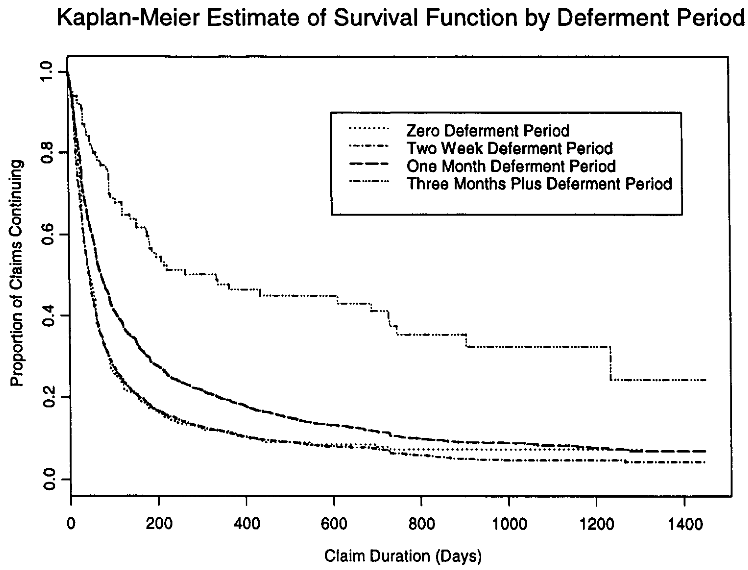


Figure 4.4 Kaplan-Meier Survival Function Split by Deferment Period

those claimants with policies that have deferred periods of one month across all claim durations. The three month deferred period has the longest predicted claim durations. Note that these durations exclude the deferred period itself. The initial three month continuous disability period that is required before claim payments commence under the relevant DII contract means that this group contains only more seriously disabled individuals than are present in the other deferred period groups.

Figure 4.5, shown below, demonstrates the effect of smoker status on claim duration.

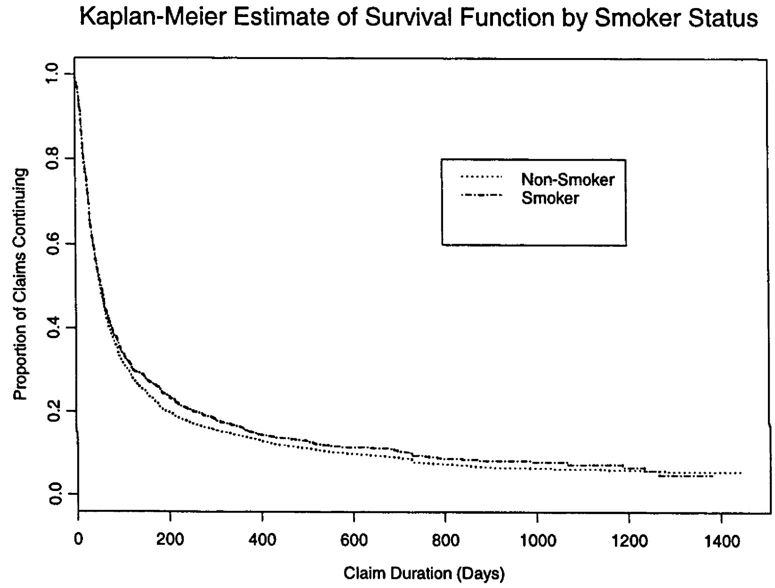


Figure 4.5 Kaplan Meier Survival Function Split by Smoker Status

Smoker status does not appear to have a significant impact on the longevity of claims. This conclusion is the same as reached by the IAAust Graduation SubCommittee of the Disability Committee found when constructing the IAD89-93 table. Of course, marginal analyses such as those presented above do not give a complete picture of how the covariates (jointly) relate the termination rates. A more complex modelling process, described in the next section, incorporates all the covariates into a single model.

4.3 Cox Regression and the Proportional Hazards Assumption

The most commonly used approach to model the effect of covariates on survival probabilities is the Cox Proportional Hazards Model, (Cox, 1972). The major theoretical development that this model provides is the ability to model covariate effects in the presence of censored observations.

The data for a Cox regression model, based on a sample of size *n*, consists of $(t_j, \delta_j, z_j), j = 1, 2, ..., n$ where t_j is the time on study for the *j*th individual, δ_j is the event indicator ($\delta_j = 1$ if the event has occurred and $\delta_j = 0$ if the lifetime is censored) and z_j is the *p*-vector of covariates or risk factors for the *j*th individual.

The relation between the distribution of event time and the covariates or risk factors *z* can be described in terms of a model, in which the hazard rate at time *t* for an individual is

$$\lambda(t; z) = \lambda_0(t)\exp(z\beta), \tag{4.1}$$

where $\lambda_0(t)$ is the baseline hazard rate, an unspecified function which outputs the hazard function for the standard set of conditions $z = 0$ and β is a *p*-vector of unknown coefficients. The parameters are estimated using the maximum (partial) likelihood technique. Importantly, the Cox model assumes that the hazards are proportional; in other words, the impact of covariates on the dependent variable, in this case rate of return to work, under the Cox model do not vary with time.

In the context of actuarial modelling of disability income insurance, this model has two major shortcomings. First, the Cox model does not produce a closed form mathematical formula for either the predicted hazard rate or the survival function. One of the primary aims of this work is to produce premium and reserve recommendations using multiple state modelling. In order for such work to be performed, it is preferable to have a mathematical model linking the various transitions in that modelling framework. The second possible limitation of the Cox model is the potential invalidity of the proportional hazards assumption.

A number of methods for testing the validity of the proportional hazards assumption in survival analysis have been proposed. Methods proposed based on statistical tests have included:

- Cox (1972) suggested testing the statistical significance of an interaction between time (or log(time)) and the various covariates specified in the model. If such an interaction term is statistically significantly different from zero then there is evidence that the impact of the covariate on survival duration varies with time; and

- Grambsch and Therneau (1994) and also Harrell (1986) have developed statistical tests based on the Schoenfeld partial residuals. These residuals are a measure of the difference between observed and expected values of the covariate at each time point. The idea of the tests is to detect a correlation between the Schoenfeld partial residuals (or some transformation thereof) and the rank order of the failure times.

Graphical procedures have also been proposed for testing the proportional hazards assumption. These have included:

- Andersen (1982) suggested a plot of cumulative baseline hazards in different groups;
- a plot of the difference of the log cumulative baseline hazard versus time; and

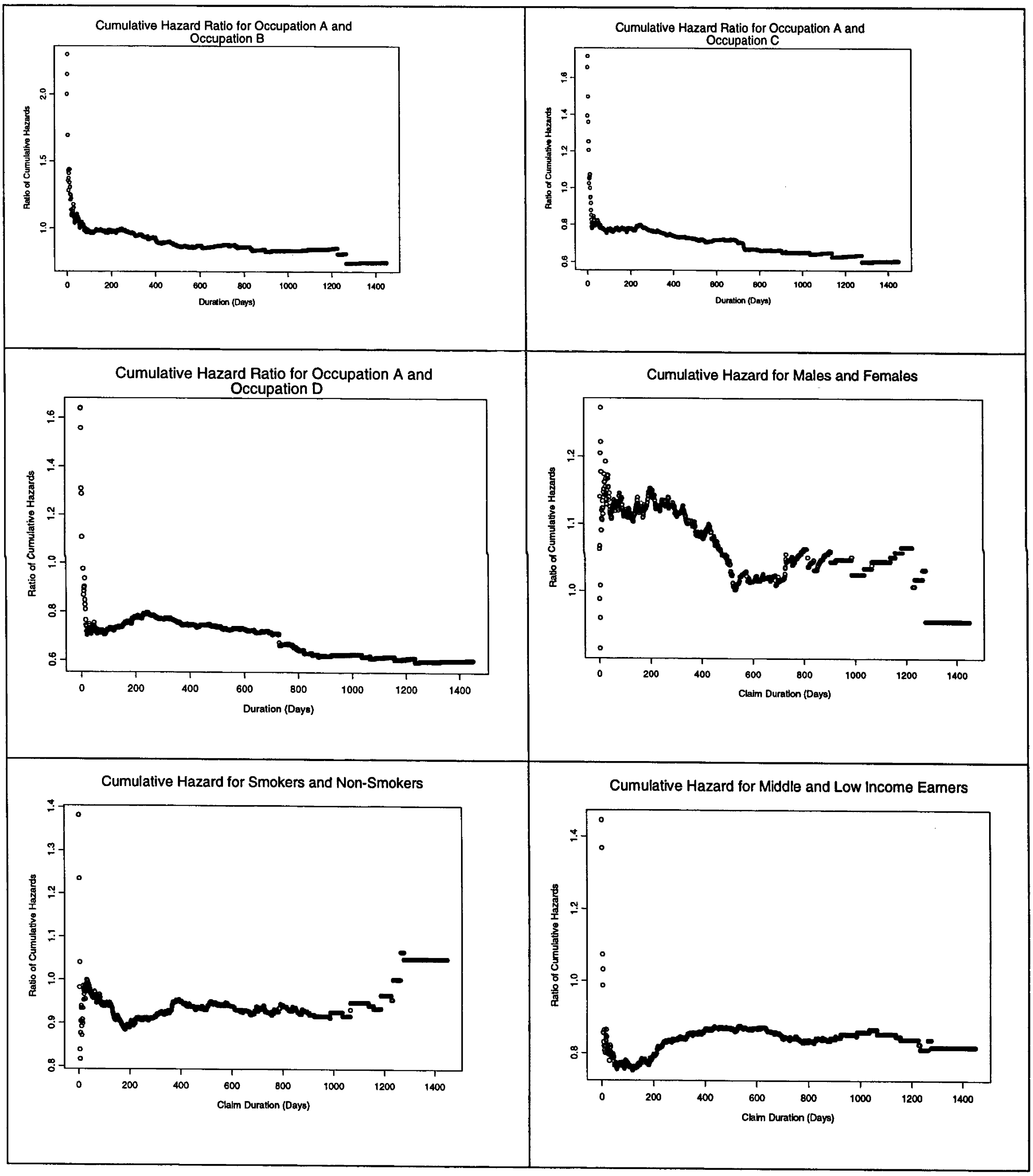
- Arjas (1988) suggested a plot of the estimated cumulative hazard versus number of failures.

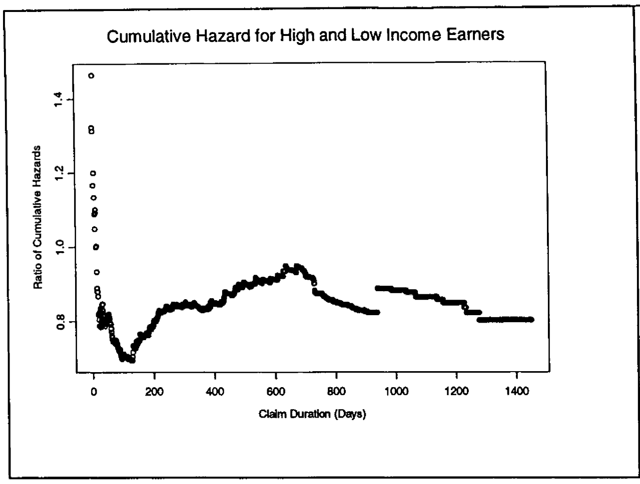
For covariates with only a small number of levels, graphical checks are more suitable. Based on the Kaplan-Meier survival function, plots of the ratios of cumulative hazard functions are shown below. If the proportional hazards assumption is reasonable, then the plots should be close to horizontal lines.

In order to test the validity of the proportional hazards assumption for the DII claim termination rate analysis, we adopt a graphical approach. The graphs in Figure 4.6 plot the ratio of the cumulative hazards for various levels of particular rating factors. From equation (4.1) integration of both sides leads to cumulative hazard rates, which are also proportional. Hence if the proportional hazards assumption is valid we would expect these graphs to depict horizontal lines with no clear upward or downward trend.

Note that the graphs which follow in Figure 4.6 show the ratio $\frac{\Lambda(\text{Group 1})}{\Lambda(\text{Group 2})}$, where Group 1 represents the first named classification in the graph title and Group 2 refers to the second named covariate classification in the graph title, and $\Lambda(x)$ is an empirical estimate of the cumulative hazard for disabled lives with characteristic set x. So, for example, in the first graph in Figure 4.6, we are considering the ratio $\frac{\Lambda(\text{Occupation A})}{\Lambda(\text{Occupation B})}$ as a function of claim duration. Again note that these cumulative hazard comparisons are one-way analyses.

The cumulative hazard ratio graphs for Occupation class show immediately that the cumulative hazard ratio seems to decrease with time. The occupation class graphs all show cumulative hazard ratios less than one. This indicates that the cumulative hazards are greater for occupation classes B, C and D than for class A. These graphs also indicate that the higher rate of return to work for claimants in Occupation Classes B, C and D compared to Occupation Class A becomes more significant as duration of claim increases.





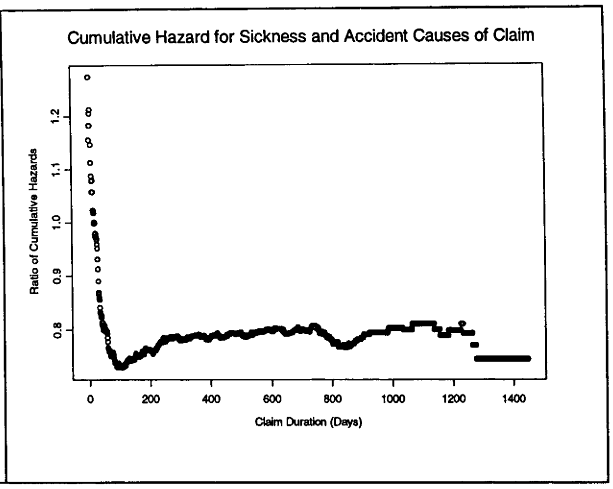


Figure 4.6 Cumulative Hazard Ratio Plots for Various Levels of Independent Variables

The cumulative hazard ratio graphs for benefit rate also show that middle and higher income earners have a lower rate of return to work. The effect of middle income compared to low level income is close to proportional across time. It is difficult to discern a pattern in the cumulative hazard ratio of high income earners compared to low income earners. There is certainly evidence of non-proportionality in the cumulative hazard ratio. The effect of smoking on the hazard rate is close to proportional. The cumulative hazard ratio graph for gender indicates that males have a higher rate of return to work than females but that the effect is reducing significantly with the duration of claim. Hence there is also evidence of non-proportionality in the effect of gender on the rate of return to work.

This graphical analysis shows clear violations of the assumption of proportional hazards for some of the key rating factors used in the proposed proportional hazards model. Extensions to the Cox regression model allowing for time varying regression coefficients have also been proposed (Therneau, 2000). These methods however will also not solve the problem of deriving a closed form mathematical expression for the predicted hazard rates. We therefore proceed with a parametric analysis of claim termination rates.

4.4 Log-Linear Mixture Models

In Section 4.2 we noted that the Kaplan-Meier estimates of the survival function were relatively smooth and also plateaued at long durations at a probability greater than zero, approximately 0.07. This feature of the survival data is referred to as an immune probability. This section describes survival analysis models, which take this feature of the data into account and therefore are suitable for describing claim termination rate data.

Maller and Zhou (1995) describe a statistical test for determining whether "immunes" are present in data. Immunes are long-term survivors and in the case of disability income insurance claim termination rate analysis, they refer to those individuals who become disabled and remain disabled for the long term. The Kaplan-Meier analysis in Section 4.2 suggests that about 7% of claimants are disabled for the long term. The method is described for the case of the exponential distribution and involves comparing the likelihood for a model where the immune probability is zero with the maximum likelihood achievable when the immune probability is allowed to vary on the range from zero to one. The test statistic is $d_n = -2\left\{l_n\left(\tilde{\theta}_{H_0}\right) - l_n\left(\tilde{\theta}\right)\right\}$ where $\tilde{\theta}$ are the maximum likelihood estimates (MLEs) obtained from fitting an exponential mixture model, $\tilde{\theta}_{H_0}$ is the corresponding MLE under the null hypothesis of no immunes, and $l_n(\theta)$ is the log-likelihood function. Maller and Zhou show that the asymptotic distribution of d_n is a 50-50 mixture of a chi-square random variable with one degree of freedom and a point mass at zero. Applying this test to the claim termination rate data, we get a test statistic of $-2(-17846.38 + 16357.45) = 2977.86$, highly significant under the chi-square point mass mixture distribution. This conclusion is not surprising after considering the Kaplan-Meier survival functions in Section 4.2. "Total and permanent disability" is also a commonly insured event and therefore long duration claims are well known phenomena and should also be expected to occur under DII.

Mixture models are based on fitting a parametric distribution to the claim durations for the lives that return to work. Define T to be a mixed random variable for the unknown claim duration of a disabled life that has just reached the end of the deferred period and is about to receive claim payments for the first time under this current period of disability. This distribution is then mixed with a point mass probability that the life will never return to work. For the case of the exponential mixture distribution, the density function is $f(t)=(1-\pi)\lambda e^{-\lambda t}, t\geq 0$ and the associated distribution function is $F(t)=(1-\pi)(1-e^{-\lambda t}), t\geq 0$, where π is the immune probability and λ is the usual exponential rate parameter. The survival function for the exponential mixture distribution is $\pi+(1-\pi)e^{-\lambda t}, t\geq 0$.

In order to achieve a good fit to the data, we will also consider a number of other potential mixture models from the Generalized F distribution family. The density functions, cumulative distribution functions and survival functions from the Generalised F family are summarised in the table below. All probability functions in the table are defined over $t\geq 0$.

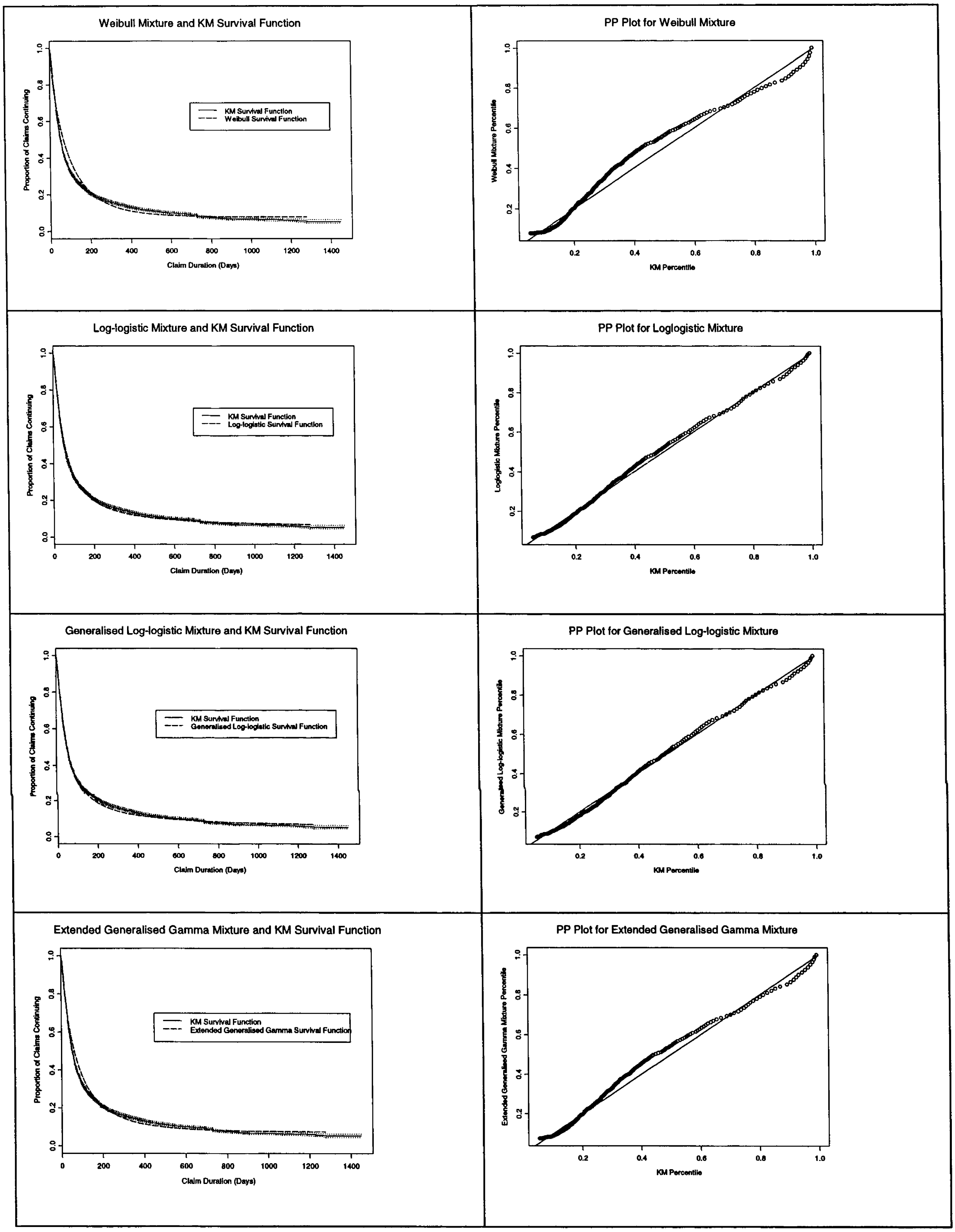
Model	**Density Function**	**Cumulative Distribution Function**	**Survival Function**
Weibull Mixture	$(1-\pi)(\lambda t)^{\alpha-1}\lambda\alpha\exp\{-(\lambda t)^{\alpha}\}$	$(1-\pi)\left[1-\exp\{-(\lambda t)^{\alpha}\}\right]$	$(1-\pi)\exp\{-(\lambda t)^{\alpha}\}$ $+\pi$
Log-Logistic Mixture	$(1-\pi)\lambda\alpha(\lambda t)^{\alpha-1}\{1+(\lambda t)^{\alpha}\}^{-2}$	$(1-\pi)\left\{1-\dfrac{1}{1+(\lambda t)^{\alpha}}\right\}$	$\dfrac{1-\pi}{1+(\lambda t)^{\alpha}}+\pi$

Generalised Log-logistic Mixture	$(1-\pi)\dfrac{(t\lambda)^{\alpha s-1}\alpha\lambda}{\{1+(t\lambda)^{\alpha}\}^{2s}B(s,s)}$	no simple form	no simple form
Extended Generalised Gamma Mixture	$(1-\pi)\dfrac{\alpha\lambda(\lambda t)^{\alpha s_1-1}}{\Gamma(s_1)}\left[s_1\exp\{-(t\lambda)^{\alpha}\}\right]^{s_1}$	no simple form	no simple form
Gamma Mixture	$(1-\pi)\dfrac{(s_1\lambda)^{s_1}t^{s_1-1}}{\Gamma(s_1)}\exp(-s_1\lambda t)$	no simple form	no simple form
Lognormal Mixture	$(1-\pi)\dfrac{\alpha}{t\sqrt{2\pi}}\exp\left[\dfrac{-\alpha^2\{\log(\lambda t)\}^2}{2}\right]$	$(1-\pi)\Phi\{\alpha\log(\lambda t)\}$	$(1-\pi)[1-\Phi\{\alpha\log(\lambda t)\}]$
Generalised F Mixture	$(1-\pi)\dfrac{\alpha}{t}B(s_1,s_2)^{-1}\left\{\dfrac{s_1(t\lambda)^{\alpha}}{s_2}\right\}^{s_1}\cdot\left\{1+\dfrac{s_1(t\lambda)^{\alpha}}{s_2}\right\}^{-(s_1+s_2)}$	no simple form	no simple form

Table 4.2 Summary of Potential Claim Duration Parametric Distributions

In order to determine which family of mixture densities is most appropriate, each of the models identified above was fit to the termination rate data. At this stage, covariate information was ignored in the analysis. The fitted claim survival function was then compared with the Kaplan-Meier estimate of the survival function from Section 4.2.

The results of fitting each of the mixture models to the claim duration data are given in Figure 4.7 and Table 4.3.


Weibull Mixture and KM Survival Function
Proportion of Claims Continuing
Claim Duration (Days)
KM Survival Function
Weibull Survival Function
PP Plot for Weibull Mixture
Weibull Mixture Percentile
KM Percentile
Log-logistic Mixture and KM Survival Function
Proportion of Claims Continuing
Claim Duration (Days)
KM Survival Function
Log-logistic Survival Function
PP Plot for Loglogistic Mixture
Loglogistic Mixture Percentile
KM Percentile
Generalised Log-logistic Mixture and KM Survival Function
Proportion of Claims Continuing
Claim Duration (Days)
KM Survival Function
Generalised Log-logistic Survival Function
PP Plot for Generalised Log-logistic Mixture
Generalised Log-logistic Mixture Percentile
KM Percentile
Extended Generalised Gamma Mixture and KM Survival Function
Proportion of Claims Continuing
Claim Duration (Days)
KM Survival Function
Extended Generalised Gamma Survival Function
PP Plot for Extended Generalised Gamma Mixture
Extended Generalised Gamma Mixture Percentile
KM Percentile

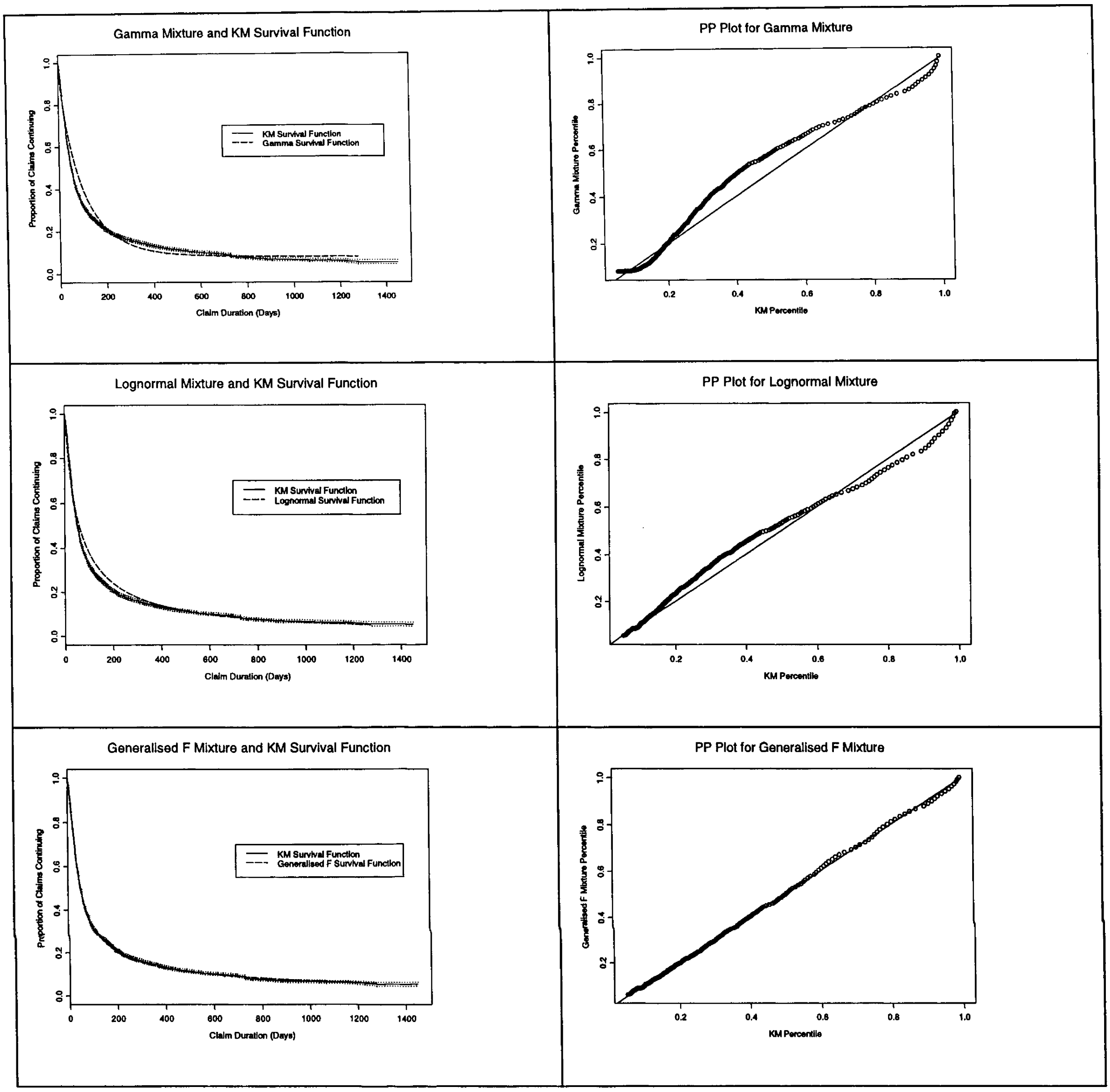


Figure 4.7 Assessment of Fit of Parametric Density to Claim Duration

Model	Maximised Log-Likelihood	R-squared for PP Plot	AIC
Weibull Mixture	-16 038.77	96.455%	32 083.54
Log-Logistic Mixture	-15 598.72	99.609%	31 205.44
Generalised Log-	-15 550.93	99.680%	31 109.86

logistic Mixture			
Extended Generalised Gamma Mixture	-15 841.87	98.370%	31 691.74
Gamma Mixture	-16 180.51	94.877%	32 367.02
Lognormal Mixture	-16 333.63	97.210%	32 673.26
Generalised F Mixture	-15 478.77	99.915%	30 967.54

Table 4.3 Assessment of Fit of Parametric Density to Claim Duration

It is clear that the three-parameter distributions, excluding the Log-logistic distribution, all significantly overestimate the survival function for claims of duration less than six months. The PP plots highlight this deficiency very clearly. This phenomenon occurs because the first six months after claim inception accounts for approximately 80% of claim terminations. The Extended Generalised Gamma fit exhibits similar properties to the Weibull, Gamma and Lognormal models. The Log-logistic distribution provides the best three-parameter distribution summary of the data. The Generalised log-logistic distribution provides only marginal improvement over the log-logistic distribution. The Generalised-F is clearly the best of the distributions considered in terms of fit. Note that the Generalised-F distribution leads to a very small estimated immune probability. However, the tail of the standard (non-mixed) Generalised F distribution is sufficiently long that the resulting model still predicts that a small percentage of claims will continue for a long period. The model predicts a 5.1% probability of claim continuation after ten years.

Based on the above findings, the analysis of the impact of covariates on claim duration will be performed using the log-logistic, generalised log-logistic and generalised F mixture distributions. We now describe the mixture models that are fitted and tested in this section. Assume that T is a random variable for the time (measured in days) it takes for a new disability claimant to return to work. We consider the transformation $Y = \log T$. The survival function for Y is

$$S(y) = (1-\pi) S_u(y) + \pi, \tag{4.2}$$

where $S_u(y)$ is the survival function of Y, given that the person returns to work. The density function for Y is

$$f(y) = (1-\pi) f_u(y), \tag{4.3}$$

where $f_u(y)$ is the density function for the time until return to work, conditional on the individual returning to work at some stage. The long term disability probability, π, is modelled using a logistic regression, $E(\pi \mid Z) = \frac{1}{1+\exp(Z'\gamma)}$, where Z is a covariate vector and γ is a vector of regression coefficients. The part of the model relating to return to work is often called the accelerated failure part of the survival model in the literature. The random variable T is said to have a generalised F distribution with μ and σ as location and scale parameters and s_1, s_2 as shape parameters, if $W = \frac{\log T - \mu}{\sigma}$ is the logarithm of a random variable having an F distribution with $2s_1$ and $2s_2$ degrees of freedom. The density of W is then

$$f(w;s_1,s_2)=\left(\frac{s_1e^w}{s_2}\right)^{s_1}\left(1+\frac{s_1e^w}{s_2}\right)^{-(s_1+s_2)}B(s_1,s_2)^{-1}, \quad (4.4)$$

and the survival function is

$$S(w;s_1,s_2)=\int_0^{s_2\left(s_2+s_1e^w\right)^{-1}} x^{s_2-1}(1-x)^{s_1-1}B(s_2,s_1)^{-1}\,dx, \quad (4.5)$$

where $-\infty<\mu<\infty,\sigma>0,s_1>0,s_2>0$ and $B(s_1,s_2)$ is the beta function evaluated at s_1 and s_2. For claimants who may return to work, we assume that the failure time T follows a generalised F distribution where the covariate vector X impacts the failure time through the relationship $\mu=X'\beta$, where β is a vector of regression coefficients. The model is fit using maximum likelihood estimation. The log-likelihood function for the model is

$$L(s_1,s_2,\sigma,\beta,\gamma)=\sum_{i=1}^{n}\left[\delta_i\log\{f(y_i;x_i,z_i,s_1,s_2,\sigma,\beta,\gamma)\}+(1-\delta_i)\log\{S(y_i;x_i,z_i,s_1,s_2,\sigma,\beta,\gamma)\}\right].$$

(4.6)

Note that if $s_1=s_2=s$ then the Generalised F distribution reduces to the Generalised log-logistic distribution. If in the Generalised Log-Logistic we have $s=1$, then the model further reduces to the log-logistic distribution.

The covariates in Table 4.1 along with all possible two-way interaction variables were tested in each of the three model families described above. Model selection was performed on the basis of the marginal significance of regression variables. Two-way interaction variables were also considered as possible regression variables. However, due most likely to the high correlation between the interaction variables and the underlying main effects, these interaction variables did not continue to have a significant effect throughout the model selection process and hence were not included in the final model.

The only continuous predictor used in the model was age. In order to properly model the effect of age on the return to work probability, three variables were used. The first variable was a simple linear predictor based on the age in years of the claimant at the time the disability commenced. The remaining two variables used were break-point predictor terms. These terms enable a different sensitivity of the return to work probability to increases in age at different levels of age. The terms were labelled ageind and ageind2. The variable ageind is equal to the age of the claimant if the claimant is 'young' and ageind2 is equal to the age of the claimant if the claimant is 'old'. The definitions of 'young' and 'old' were formed by maximising the log-likelihood of the resulting model. The definitions used in the final model are ageind is age for claimants below age 29. The variable ageind 2 is equal to age for claimants above age 44.

The likelihood ratio test and the Akaike's Information Criterion (AIC) were used to assess the models fitted. The results are summarised in Table 4.4 on the following page.

	Maximised Log-Likelihood	*Likelihood Ratio Test Statistic relative to Generalised F Model*	*AIC*
Accelerated Failure: No Covariates. No Logistic Model			
Generalised F	-15,478.44	-	30,964.9
Generalised Log-logistic	-15,631.24	305.6	31,268.5
Log-logistic	-15,701.86	446.8	31,407.8
Accelerated Failure: No Covariates. Logistic: No Covariates			
Generalised F	-15,478.77	-	30,967.5

Generalised Log-logistic	-15,550.93	144.3	31,109.9
Log-logistic	-15,598.72	239.9	31,203.4
Accelerated Failure: Covariates included. Logistic: No Covariates			
Generalised F	-15,297.81	-	30,627.6
Generalised Log-logistic	-15,343.79	92.0	30,717.6
Log-logistic	-15,403.06	210.5	30,834.12
Accelerated Failure: Covariates included. Logistic: Covariates included.			
Generalised F	-15,260.81	-	30,565.6
Generalised Log-logistic	-15,291.47	61.3	30,624.9
Log-logistic	-15,351.74	181.9	30,743.5

Table 4.4 Assessment of Accelerated Failure and Mixture Models for Claim Duration

Note also that these likelihood ratio test statistic values can be compared to critical values derived from the chi-squared distribution. This statistical test will be conservative because the true distribution of the likelihood ratio test statistic has greater density at zero and the shortest durations, than does a chi-square variable.

It is clear from the above table that the Generalised F mixture model with covariates for both the accelerated failure time part of the model and the logistic part of the model is optimal. A summary of this fitted model is given on the following page.

Generalized F mixture model				
The maximum loglikelihood is -15256.62				
Terms in the accelerated failure time model:				
	Coefficients	**Std.err**	**z-score**	**p-value**
Shape1	-1.27397			
Shape2	-1.59553			
Log(scale)	0.00278	0.003667	0.7579	0.4485249
(Intercept)	-0.00354	0.147557	23.4308	0.0000000
age	0.00278	0.003667	0.7579	0.4485249
ageind	-0.00354	0.002286	-1.5489	0.1213947
ageind2	0.00202	0.001226	1.6476	0.0994444
occupB	0.12864	0.063015	2.0414	0.0412076
occupC	-0.04742	0.042424	-1.1178	0.2636703
occupD	-0.12454	0.044126	-2.8224	0.0047672
benrate2	0.06753	0.040174	1.6809	0.0927729
benrate3	0.11961	0.041763	2.8639	0.0041843
benrate4	0.27448	0.056277	4.8774	0.0000011
benratetop2	0.12041	0.050253	2.3961	0.0165717
sick	0.04555	0.031021	1.4685	0.1419671
defpd2	0.35779	0.033275	10.7526	0.0000000
defpd3	1.02768	0.183896	5.5884	0.0000000

Table 4.5 Accelerated Failure Model Regression Coefficients

Terms in the logistic model:				
	Coefficients	**Std.err**	**z-score**	**p-value**
(Intercept)	10.00406	1.586363	6.3063	0.0000000
age	-0.07391	0.015437	-4.7875	0.0000017
smokernew	-0.92219	0.298505	-3.0894	0.0020058
conttypenew1	-0.58790	0.271987	-2.1615	0.0306560
sick	-2.51927	1.482109	-1.6998	0.0891704
defpd0	-2.52043	1.643859	-1.5332	0.1252161
defpd2	-0.83440	0.301255	-2.7697	0.0056100
defpd3	-2.54072	0.424056	-5.9915	0.0000000

Table 4.6 Logistic Model Regression Coefficients

The vast majority of the regressors shown in Table 4.6 have a statistically significant effect on the rate of return to work at the 5% significance level. For variables which are highly subdivided, for example, occupation which has four classes, the statistical significance of the variable is strongly affected by the amount of data for that particular class. For that reason, we note that occupation class C does not appear to have a significantly different rate of return to work than occupation class A, despite the contrasting results from the Kaplan-Meier analysis shown in Figure 4.3.

It is also of interest that there are regressors that are statistically significant predictors of the rate of return to work in the accelerated failure time part of the model which are not significant in the logistic part of the model. In particular the model shows that smoker status, which up until now in Australian studies has not been considered a significant determinant of claim termination rates, leads to a statistically significant increase in the probability of long term disability.

Apart from the likelihood ratio test, it is also possible to assess the quality of the fit of the model by dividing the data into groups according to the values of the covariates included in the final model. Out of the 8863 individuals in the study, 61 were found to possess all of the following characteristics: aged between 35 and 45, disability benefit of less than $2000 per month, disability caused by sickness, deferred period of two weeks, occupation class A and non-smoker. For these 61 lives, the Kaplan-Meier fit to the survival function is compared to the survival function predicted by the Generalised F model. The result of this comparison is shown in the graph below, where 95% confidence bands have been included around the Kaplan-Meier fit.

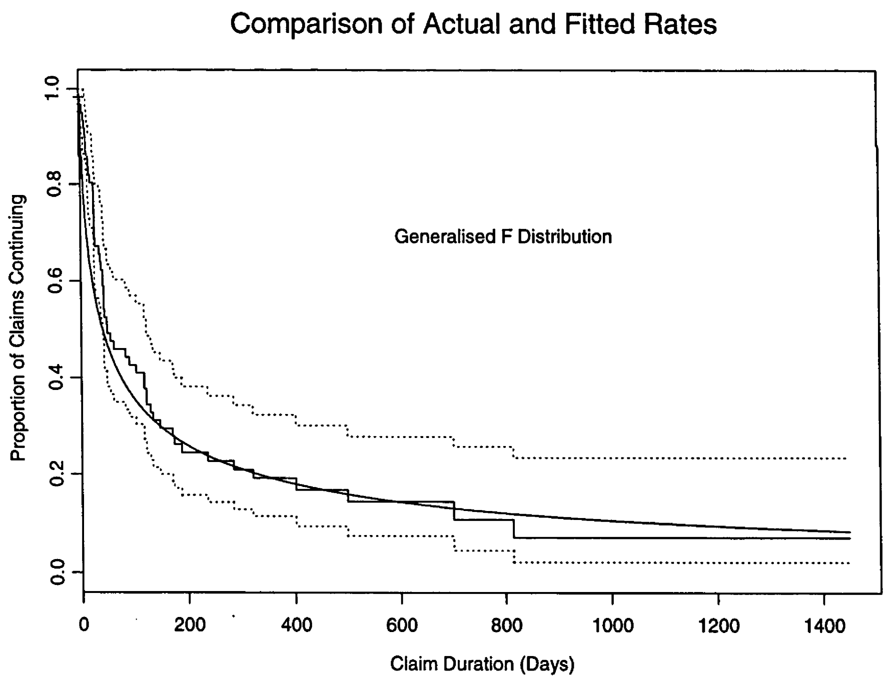


Figure 4.8 Comparison of Actual and Fitted Rates for the Generalised F Distribution

The fit of the Generalised F distribution is clearly very good except at the shortest durations where the model predicts higher rates of return to work than does the empirical Kaplan-Meier survival function. Since pricing and reserving for DII are impacted most by long duration claims, this lack of an accurate fit at the shorter durations has less financial consequence for a life office than would imprecise model fitting in the tail of the claim duration probability distribution, and so may not be of practical significance.

In Section 4.2 we demonstrated that the proportional hazards assumption of the Cox regression model was not satisfied by the covariates in the disability claim termination data. The impact of this assumption not being satisfied on the fit of the Cox regression model is shown in the graph below. This graph compares the same data as used in Figure 4.9 to compare the empirical Kaplan-Meier survival function with the survival function predicted using Cox regression.

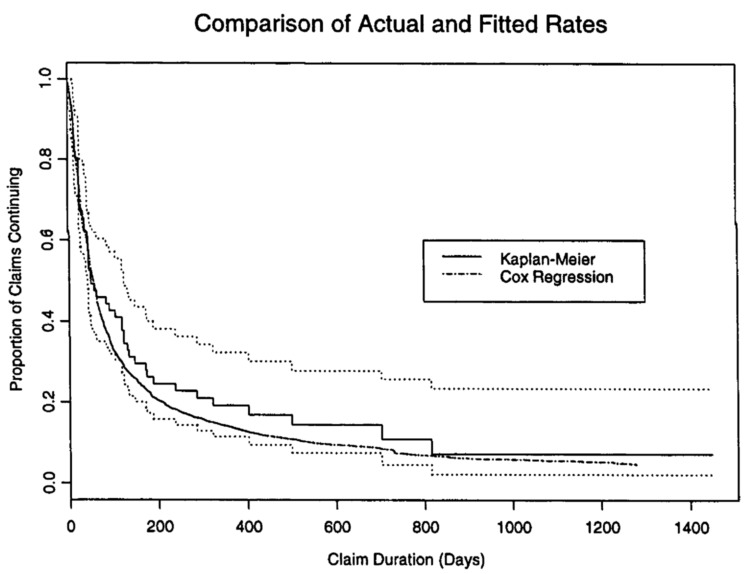


Figure 4.9 Comparison of Kaplan-Meier and Cox Regression Claim Duration Models

This graph shows clear evidence that the Cox regression model fits claim termination rates that are significantly higher than the Kaplan-Meier estimate between durations 6 months and 2 years.

A useful way to compare the fits of various models, given the aim of the modelling is premium rating, is to compare the predicted expected present value of an annuity payable to a disability annuitant throughout their period of disability. We consider a disability income insurance policy with a four year benefit period. The annuity is assumed to be payable continuously with the valuation performed at a force of interest of 5% per annum. Mortality is ignored, which is a reasonable assumption at this stage given that we are considering lives aged between 35 and 40 and also that our aim is to assess the relative merits of the Cox regression model and the Generalised F Mixture Model in describing claim durations. Table 4.7 gives the expected present value of an annual annuity of one dollar payable throughout the period of disability under each model. It is clear that the Generalised F Mixture model is preferable in this case to the Cox regression model as evidenced by a much closer estimate of the annuity value to the underlying annuity value.

Kaplan-Meier Survival Function	Cox Regression Model	Generalised F Mixture Model
0.6163	0.4739	0.5837

Table 4.7 Annuity Value Comparison for three Model Fitting Procedures

One of the most noteworthy features of the analysis in this chapter is the difference in statistically significant regressor variables between the accelerated failure time part of the model and the logistic regression for the immune probability component of the model. The next chapter extends this investigation to quantile regression, where the significance of rating variables is assessed at various quantiles of the distribution of claim durations, rather than just at the conditional mean.

CHAPTER FIVE

REGRESSION QUANTILE ANALYSIS OF CLAIM TERMINATION RATES

Actuarial interest in quantiles other than the median has increased considerably in recent years. Most notable is the Australian Prudential Regulatory Authority (APRA) standard for the valuation of general insurance liabilities, GPS210, introduced as part of the Australian General Insurance Reform (2001). This standard requires that a risk margin should be established "on a basis that is intended to secure the insurance liabilities of the insurer at a given level of sufficiency – that level is 75 per cent". Previously the General Insurance Act (1973) was considerably less prescriptive on the level of risk (or prudential) margin that insurers were required to hold.

Given that general insurance actuaries are now required to estimate a 75^{th} percentile of the distribution of outstanding claims for recording in profit and loss statements it becomes important that the impact of potential risk factors on various quantiles of the distribution of outstanding claims provisions be considered in addition to just the impact of risk factors on the mean of the outstanding claims provision.

In the context of disability income insurance claim termination rates, we have already seen an example of how various insured characteristics impact claim termination rates differently at different quantiles of the distribution of claim duration. For example, Chapter 4, the use of mixture models showed that the smoker status, which has not been included in previous Australian industry tables for claim termination, has a statistically significant impact on the probability that a DII claim will continue indefinitely and that the claimant will never return to work. This is evidence that smoker status is a statistically significant predictor for claim termination rates (leading to a reduction in claim termination rates) for the very longest duration claims. In other words, a traditional regression which considers only the impact of rating factors on the mean would not find that smoker status is statistically significant, however closer examination of the impact of smoker status in the *tail* of the probability distribution of claim durations indicates that smoker status is critical for long duration claims.

The importance of understanding the impact of potential rating factors and the different impacts they have across the claim duration distribution is of particular importance in reserving and pricing DII contracts. Failure to properly assess the impact of a rating factor in the tail of the probability distribution of claim durations will lead to serious underestimation of claim reserves in respect of disabled lives, particularly those lives who have been disabled for longer than, say, six months.

5.1 Regression Quantiles

One way of extending the linear model to allow for prediction of various quantiles of the distribution of the claim duration is the method of regression quantiles of Koenker and Bassett (1978). This methodology has recently been extended to allow for standard right censoring and therefore can provide an alternative to the Cox Model or mixture models; see Portnoy (2003).

Traditional statistical and actuarial analysis has focused on sample averages as estimates of the population mean. Variability has generally been considered using sample standard deviations and the assumption of normality or, more recently, other parametric assumptions have been made. It has long been argued, (Galton, 1889), that any complete analysis of the "full variety of an experience requires the entire distribution of a trait, not just a measure of its central tendency." We therefore consider the use of regression quantiles as a method for identifying heterogeneity among subpopulations by considering the behaviour of the percentiles as a function of their associated probability τ.

For a random variable Y of measurements from some population, the population quantile is defined to be the value $Q_Y(\tau)$ satisfying

$$P\{Y \leq Q_Y(\tau)\} = \tau \quad \text{for } 0 \leq \tau \leq 1. \tag{5.1}$$

Next, we describe the generalisation of this quantile to a regression context through the use of the conditional quantile. Specifically, the conditional quantile, $Q_{Y|X}(\tau; x)$, satisfies

$$P\{Y \leq Q_{Y|X}(\tau; x) \mid X = x\} = \tau. \tag{5.2}$$

Whereas traditional regression analysis provides a single regression curve, for example the conditional mean function, in this regression quantile context we can let τ vary, and therefore consider a family of conditional quantile curves to provide a clearer picture of the dependencies present in the data.

To simplify the analysis, Koenker and Bassett (1978) suggest the estimation of conditional quantile curves under the assumption that, after appropriate transformations, they are linear in the covariates. This assumption has the advantage of allowing easier interpretation of coefficient estimates and also permits significantly faster computation. The estimation of the conditional quantile functions involves finding the solution to the problem of choosing ξ to minimise

$$R_T(\xi) = \sum_{i=1}^{n} \rho_T(Y_i - \xi), \tag{5.3}$$

where ρ_T is the piecewise linear "check" function,

$$\rho_T(u) = u(\tau - I(u < 0)) = \tau u^+ + (1-\tau)u^-, \tag{5.4}$$

and where u^+ and u^- are the positive and negative parts of u taken positively, respectively.

If we next consider a general linear response model where $\{Y_i, \underline{x}_i\}$ denotes a sample of responses Y and explanatory variables $\underline{x}$ (in m dimensions), and suppose

$$Y_i = \underline{x}_i' \underline{\beta} + z_i, \quad i = 1, 2, \ldots, n, \tag{5.5}$$

where $\underline{\beta}$ is an m-dimensional parameter and z_i is the random error term. If we then minimise

$$R_\tau(\beta) \equiv \sum_{i=1}^{n} \rho_\tau(Y_i - \underline{x}_i' \underline{\beta}) \tag{5.6}$$

by varying $\underline{\beta}$ we obtain the regression quantiles. Note that the estimated regression quantile parameters implicitly depend on the probability, τ. In particular, the jth coordinate of $\underline{\hat{\beta}}(\tau)$

gives the marginal effect of a unit change in the *j*th explanatory variable, $x^{(j)}$, on the conditional τ' th-quantile of the response.

If the model predicts that the $\underline{\beta}$ coefficients change with τ, then we have evidence of heterogeneity in the population. This heterogeneity can take the form of unequal variances (heteroscedasticity) or it may represent the varying effect of heterogeneity among subpopulations.

Throughout the analysis which follows we will make use of the R library crq. This library contains a function which allows the user to fit censored regression quantile models and assess the extent of heterogeneity in the covariate effect over the range of claim durations.

5.2 Regression Quantiles and Claim Termination Rates

The aim of this section is to illustrate the application of censored regression quantiles (Portnoy 2003), to claim termination rates for DII. The heterogeneity of the effect of the covariates age, occupation and deferred period across the distribution of claim durations will be analysed. The benefits of using censored regression quantile analysis as compared to more traditional Cox regression in this context will also be explored.

The potential output from censored regression analysis can be extremely voluminous. This issue arsies due to the wide range of possible conditional quantile curves that can be estimated. In order to make the interpretation of results simpler, we restrict ourselves in this chapter to the consideration of the effect of age, occupation class (C or D vs A or B) and deferred period (greater than or equal to 28 days or less than 28 days).

A censored regression quantile model was fit to the entire dataset of claim durations described in Chapter 4. The R command used to fit the model is

```
crq(Surv(log(durn3),terminate) ~ age + occupnew + defpdnew, data = termrates2),
```
(5.7)

where occupnew is an indicator variable for occupation classes C and D, and defpdnew is an indicator variable for deferred period in excess of 27 days.

Mathematically, the form of the fitted censored regression model is

$$\log\left(\text{Time to return to work}\right) = \beta_0 + \beta_1(\text{Age}) + \beta_2\left(\text{Occupation Class}\right) + \beta_3\left(\text{Deferred Period}\right), \quad (5.8)$$

where separate models of the above form are fit to quantiles corresponding to breakpoints in the claim duration data.

To consider the impact of age on the log of claim duration we create a graph of the predicted censored regression quantile relationship between log of duration and age; see Figure 5.1.

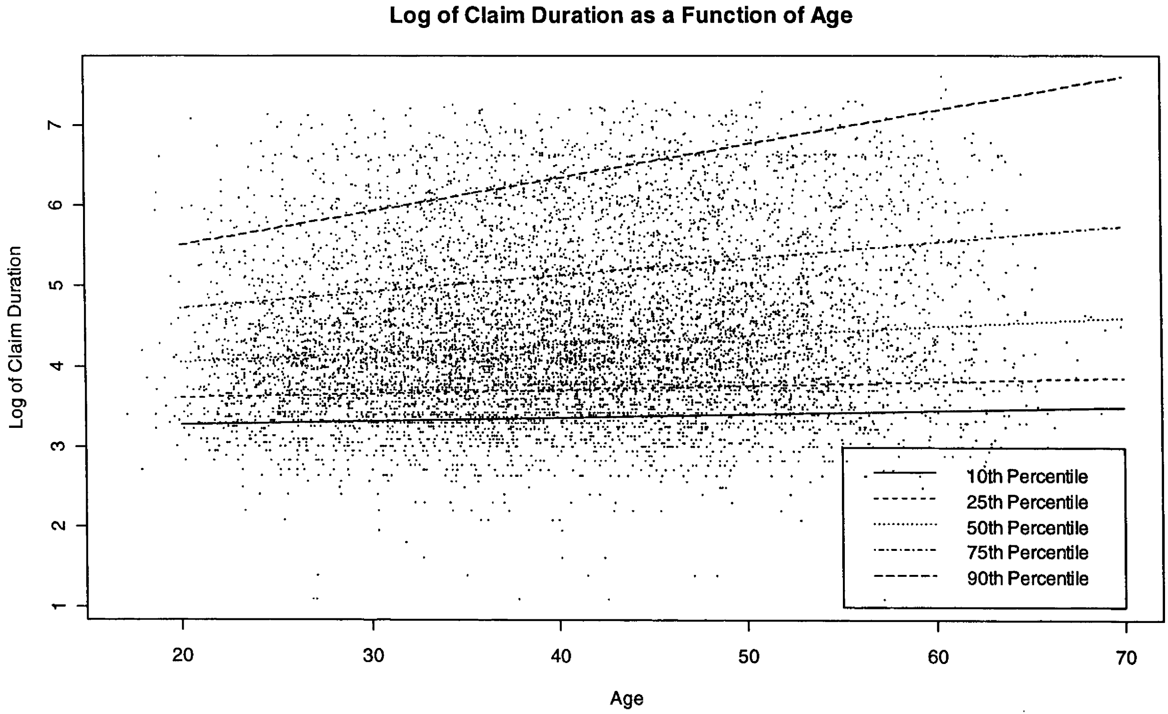


Figure 5.1 Log of Claim Duration as a Function of Age for Five Different Percentiles

Immediately obvious from Figure 5.1 is the increasing slope of the regression lines at higher percentiles of the distribution of log claim duration. This suggests that the age sensitivity of (the log of) claim duration is greater for longer duration claims. In particular, the effect of increasing age increases the 90th percentile of the distribution of claim durations for a given age much more than the same increase in age increases the 10th percentile of the distribution of claim durations. The regression coefficients for age, occupation class and deferred period for various percentiles are given below, in Table 5.1. Note that the effect of occupation class on claim duration also varies significantly with the percentile of the distribution being considered. In particular, the effect of being in occupation class C or D in reducing the predicted duration of disability is more pronounced at the higher percentiles of the distribution of the claims distribution.

Percentile	Censored Regression Quantile Coefficient (Age)	Censored Regression Quantile Coefficient (Occupation)	Censored Regression Quantile Coefficient (Deferred Period)
10th Percentile	0.00419	-0.01846	0.60475
25th Percentile	0.00494	-0.06772	0.59322
50th Percentile	0.01068	-0.12023	0.59366
75th Percentile	0.02048	-0.23410	0.71315
90th Percentile	0.04206	-0.29936	0.62723

Table 5.1 Censored Regression Quantile Coefficients

The above results have clear implications for the determination of the disabled life reserve (DLR). This quantity is the reserve held by an insurer in respect of an insured who is currently claiming benefits at the date of the valuation. Insurers will always have a material proportion of their portfolio relating to insured lives who are currently disabled and who have been disabled for a reasonable period of time at the date of valuation. The insurer is required to determine the amount of money that needs to be held in respect of these disabled and insured lives at a particular instant in time. Clearly the amount of money required depends on the future disability status of the insured life. Figure 5.1 shows that the effect of the insured being older or of being in occupation class A or B on the claim duration is more significant for longer duration claims; that is, for claims that have extended into higher percentiles of the claim duration probability distribution.

A model we have discussed earlier that is often used in survival analysis is the Cox Regression model, (Cox, 1972). This model estimates the impact of rating factors such as age and occupation class on the dependent variable, claim duration, by considering the impact of these rating factors collectively across all quantiles of the claim duration distribution. It is therefore of interest to assess the difference in the predicted sensitivities of claim duration to each of the insured characteristics from the Cox model and the censored regression quantile method.

5.3 Comparison of Cox Regression and Censored Regression Quantiles for Claim Termination Rates

To begin this section, we fit a Cox regression model to our data using age, occupation (class C or D indicator) and deferred period (greater than 27 days indicator) to the claim duration data. The dependent variable is the log of the claim duration and the usual right censoring in the data is used within the analysis. The output for this regression model is given below in Table 5.2.

coxph(formula = Surv(log(durn3), terminate) ~ age + occupnew + defpdnew, data = termrates2)					
	coef	**exp(coef)**	**se(coef)**	**z score**	**p-value**
age	-0.0124	0.988	0.00119	-10.40	0.0e+00
occupnew	0.1285	1.137	0.02524	5.09	3.5e-07
defpdnew	-0.4668	0.627	0.02482	-18.80	0.0e+00
Likelihood ratio test=631 on 3 df, p=0 n= 8863					

Table 5.2 Censored Regression Quantile Output

The three rating factors are clearly highly statistically significant and the overall model indicates that age, occupation class and deferred period are jointly statistically significant.

In order to compare the Cox regression model to the censored regression quantile model it is necessary to compare the predicted sensitivities of the quantiles of the claim duration distribution under the two models. For the Cox model, we have that the predicted hazard function for the ith individual in the sample, $h_i(t)$, is

$$h_i(t) = h_0(t) e^{\underline{x}_i \beta}, \quad i = 1, \ldots, n, \tag{5.9}$$

where $h_0(t)$ is the baseline hazard function. Given this form for the hazard function, the survival function can be written as

$$S_i(t) = \exp\left(-H_0(t) e^{\underline{x}_i \beta}\right), \tag{5.10}$$

where $H_0(t) = \int_0^t h_0(s)\, ds$.

So the conditional quantile for claim duration, T, at $\underline{x}$ becomes

$$Q_{\text{Cox}}(\tau \mid x) = H_0^{-1}\left(-\log(1-\tau) e^{-\underline{x}_i \beta}\right). \tag{5.11}$$

The quantity $Q_{\text{Cox}}(\tau \mid x)$ is therefore the predicted time since claim inception, under the Cox Regression model, when a proportion τ of those insureds who claim from their DII contract will have returned to work. The censored quantile regression coefficients give the predicted change in various quantiles of the distribution of the log of claim duration when various rating factors are increased by one unit. It is therefore possible to directly compare the coefficients estimated using censored regression quantiles with the derivative of the expression at (5.11). Consequently we compare the $\hat{\beta}(\tau)$ with the quantity

$$\frac{\partial}{\partial x} Q_{\text{Cox}}(\tau \mid x) = \frac{\partial}{\partial x} H_0^{-1}\left(-\log(1-\tau) e^{-\underline{x}_i \beta}\right). \tag{5.12}$$

To calculate the above derivatives we need to use numerical differentiation owing to the irregularities present in the inverse cumulative baseline hazard function, $H_0^{-1}(t)$. For the calculation of this derivative, we use the model in Table 5.2. From this model we calculate the fitted claim continuance probabilities at each of the times that a person in the sample returns to work for a life aged 40.53, the mean of the ages in the sample. Denote these values $S_1(t_i)$. We also use the model in Table 5.2 to calculate the fitted claim continuance probabilities for a life aged 41.53 (one plus the mean of the ages in the sample). Denote these values $S_2(t_i)$. Next we determine the log of claim duration that corresponds to each of the values of $S_1(t_i)$ for a life aged 41.53. These values are calculated using linear interpolation and the S-Plus function, out2, which performs this calculation (amongst other calculations) is given in full in Appendix 5.1. It is then straightforward to numerically estimate the predicted quantile sensitivity based on the Cox Regression model. The difference between the survival times for a given quantile of the log claim duration distribution estimates the sensitivity of various quantiles of the log of claim duration distribution under the Cox Model.

It is useful to compare the sensitivities of various quantiles of the log of claim duration distribution from the use of censored regression quantiles and the more conventional Cox regression model. Figures 5.2 and 5.3 compare these quantile sensitivities for changes in age and occupation class. The sensitivity labelled on the y-axis of the graphs in Figures 5.2 and 5.3 refers to the expression in (5.12).

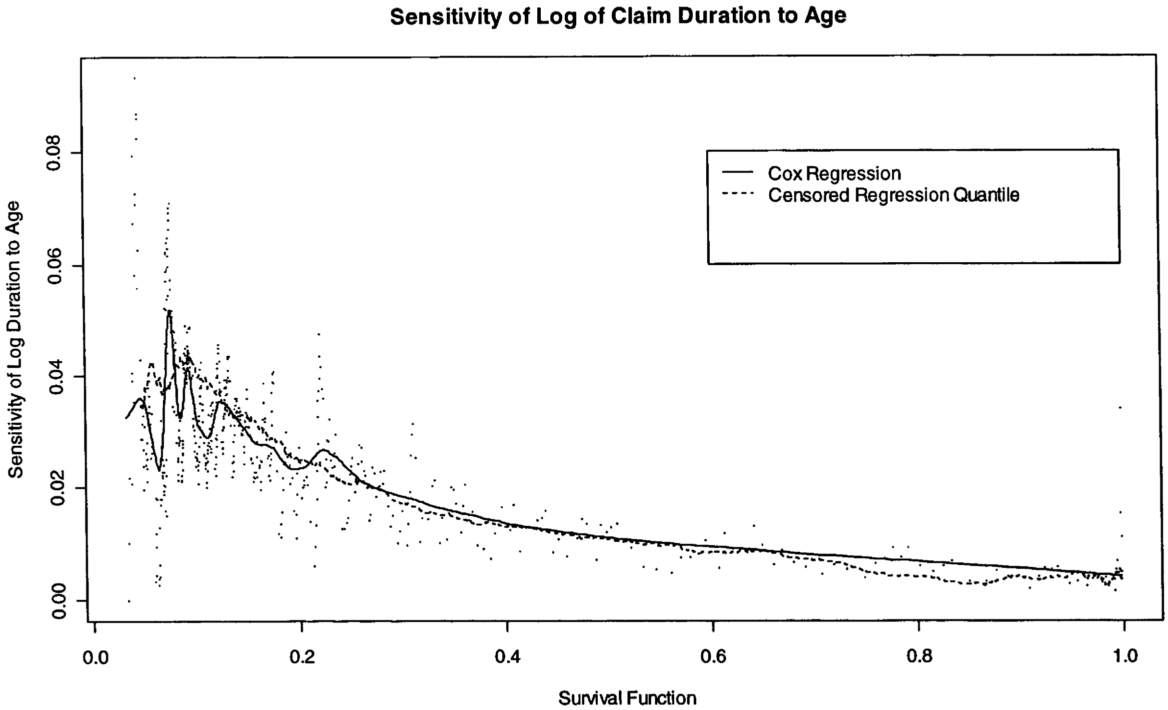


Figure 5.2 Comparison of Sensitivities of Log Duration to Age

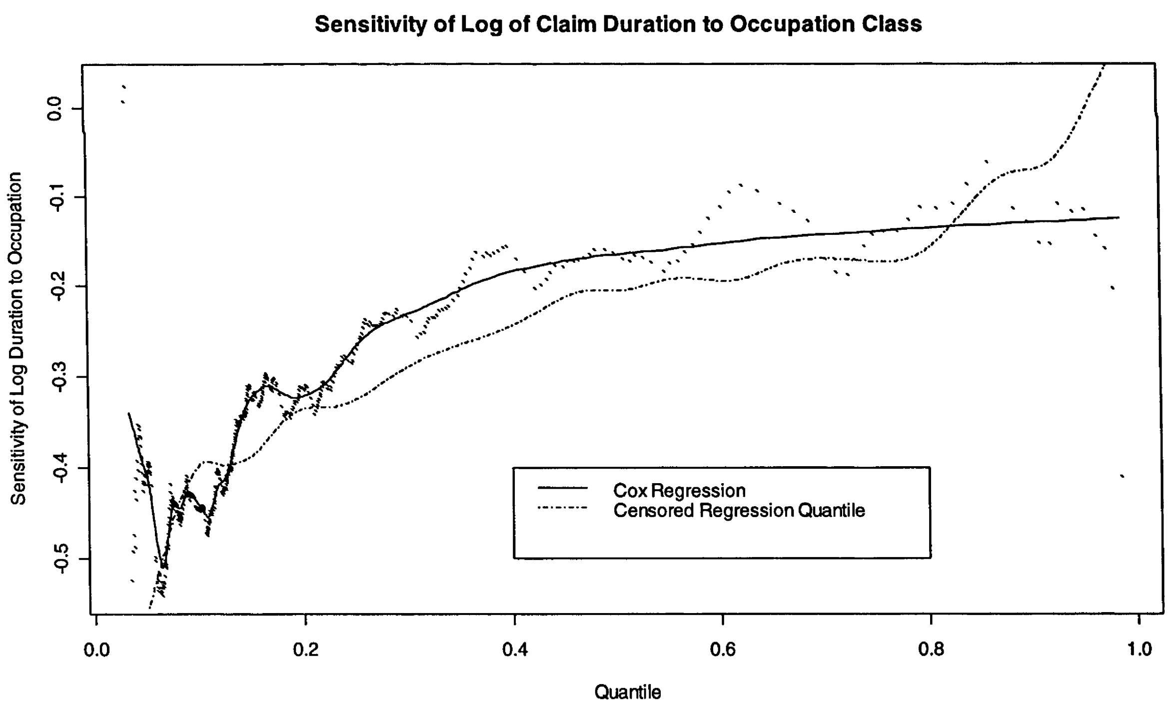


Figure 5.3 Comparison of Sensitivities of Log Duration to Occupation Class

From Figure 5.2 it is clear that the predicted age sensitivity of lower quantiles (at higher levels of the survival function) is higher under the Cox model than for the censored regression quantile analysis. There is considerable variability in the censored regression quantile coefficients for the higher quantiles (survival function between 0 and 0.2). This volatility is primarily due to the small number of claims that are still continuing at these claim durations.

From Figure 5.3, there is a clear bias in the estimation of the quantile sensitivities of log of claim duration to occupation class for the Cox regression model. The Cox regression model predicts a greater reduction in claim duration for occupation classes C and D than the censored regression quantile analysis over most of the range of the log of claim duration distribution. This result is driven in part by the inappropriateness of the proportional hazards assumption that underlies the Cox Regression model.

5.4 Assessing the Comparison between Cox Regression and Censored Regression Quantiles using Subsampling

Figure 5.2 also clearly demonstrates that using Cox regression alone can lead to flawed conclusions about the age sensitivity of the log of claim duration particularly for shorter duration claims. It is of interest to see whether the disparity between predicted age sensitivities of claim duration between the two approaches is likely to occur with most sets of disability income insurance data or whether the difference is more a feature of the particular set of Australian industry claim duration data that is being analysed.

To explore this, we consider a subsampling approach whereby 84 different datasets, each of size 400, chosen from the original set of 8863 data points. These 84 different datasets contain records 1 to 400, 101 to 500, ..., 8401 to 8800. Since the data is in no particular order, with respect to the variable of interest namely claim duration, this is similar to analyzing 84 different sets of randomly chosen disability income insurance claim duration data each of size 400 records.

For each of these datasets of size 400, we fit both a Cox regression model, equivalent to the model in Table 5.2, and also a censored regression quantile model. These models both use age, occupation class and deferment period as the only covariates.

We then compare the censored regression quantile age coefficient for each of the 400 models to the Cox regression age sensitivities for a range of quantiles. We are interested in assessing the absolute difference between the censored regression quantile coefficients and the Cox regression quantile derivative function. In order to make the comparison more straightforward, we averaged the Cox regression quantile sensitivities over survival function bands, namely [0,0.2), [0.2,0.4), [0.4,0.6),[0.6,0.8) and [0.8,1.0]. We also averaged the censored regression quantile function over the same bands for the survival function. The difference in the mean sensitivities for each of the 84 models were then calculated. A density of these differences was then created. The program which performs this subsampling is the S-Plus function, out3, shown in full in Appendix 5.1.

The density for the difference in mean sensitivities for age and Survival Function in excess of 0.8 is shown below. The mean of the average differences between the sensitivities is 0.0008570 which is 20.1% of the censored regression quantile sensitivity. Hence the Cox Regression predicts an age sensitivity of log of claim duration that is 20.1% higher than the censored regression quantile method for the shortest 20% of claim durations.

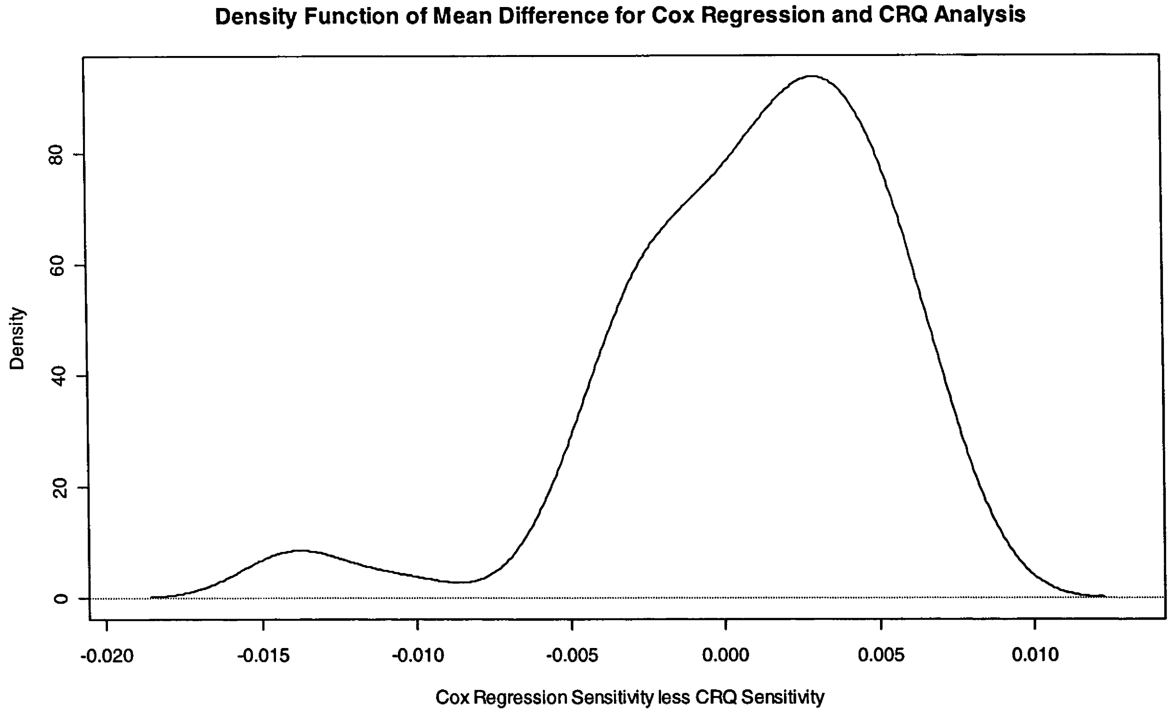


Figure 5.4 Density Function of Mean Difference for Cox Regression and CRQ Analysis for the Survival Function on the range [0.8,1.0]

Similarly, Figure 5.5 on the following page shows an empirical density function of mean differences in predicted quantile sensitivities from Cox Regression and censored regression quantiles over the [0.6,0.8) band of the survival function. The mean difference is 0.0024, or 35.6% of the censored regression quantile analysis. This finding indicates again that Cox Regression predicted sensitivities of the return to work hazard rate to covariates are consistently higher than their censored regression quantile counterparts over the 60% to 80% region for the claim duration survival function.

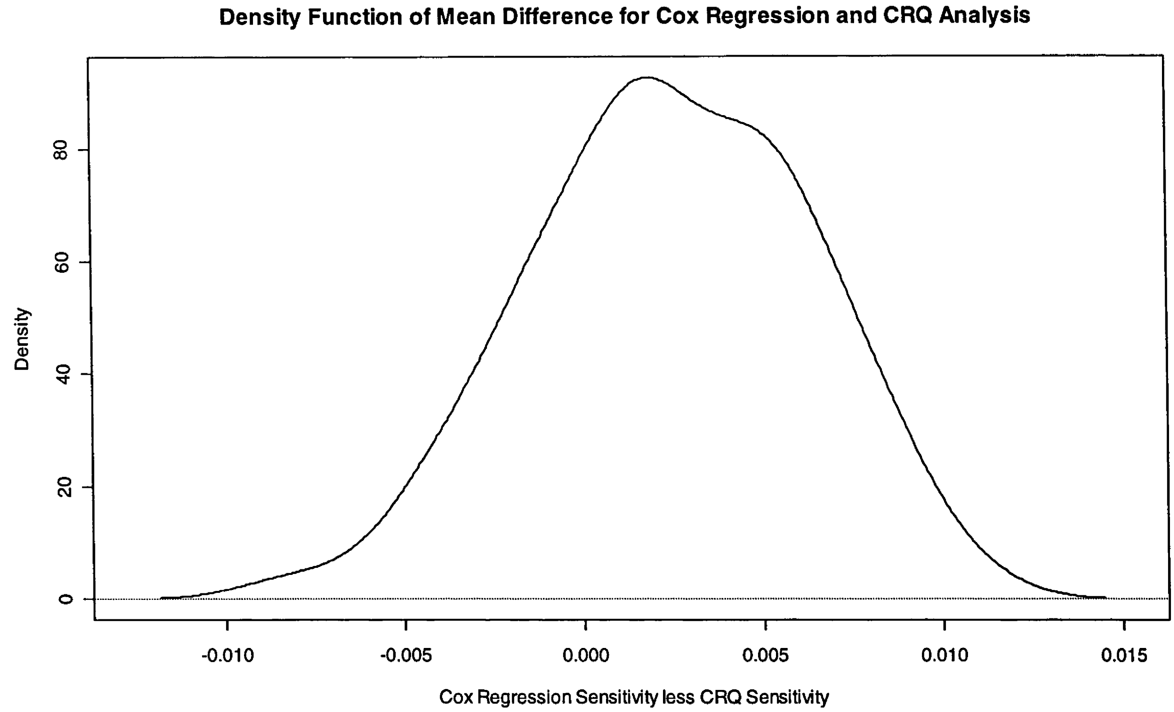


Figure 5.5 Density Function of Mean Difference for Cox Regression and CRQ Analysis for the Survival Function on the range [0.6,0.8)

This chapter has demonstrated an additional technique that can be used to detect heterogeneity in claim duration data. Censored regression quantiles therefore provide a more reliable method for assessing the impact of covariates in the tail of the probability distribution of claim durations than do other more commonly adopted methods from survival analysis, such as Cox Regression. In Chapter Six, we combine results for claim incidence and claim termination into a premium rating and valuation model for use by life insurers managing DII business.

CHAPTER SIX

MULTIPLE STATE MODELLING, FLOWGRAPHS AND DISABILITY INCOME INSURANCE

[1][2]This chapter investigates the use of multiple state models for pricing and reserving for disability income insurance products in light of results from Chapters 3 and 4 for both claim incidence and claim termination rates. Also, the piecewise constant intensity approach to premium determination of Jones (1993) is extended to allow for determination of premium rates for a range of DII contracts.

6.1 Multiple State Modelling

Haberman and Pitacco (1999) give the general multiple state model used for describing the transitions made by holders of DII policies. The model is shown below in Figure 6.1.

[1] The material in this chapter has been presented at an "Actuarial Models for Financing Disablement Benefits" seminar directed at actuarial practitioners.
[2] The multiple state model work in this chapter has also been used by the author along with Professor Richard Heaney in a publication, "Genetic Testing in Life Insurance", which appeared in *Agenda*, Vol.10, Issue 1.

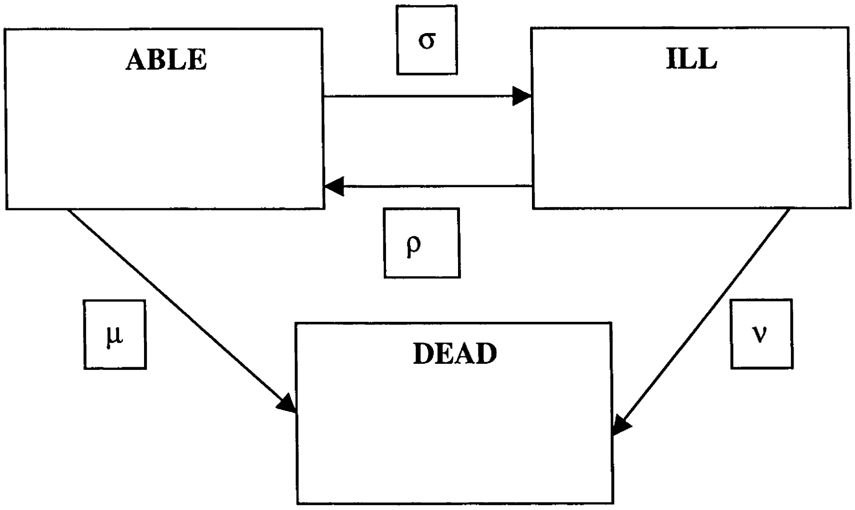


Figure 6.1 Multiple State DII Model

The transition intensities σ, ρ, μ and ν are the standard hazard rates for disability onset, recovery from disability, death from the able state and death from the disabled state respectively. These are the same as the force of sickness, the force of recovery, the force of mortality for able lives and the force of mortality for disabled lives, respectively.

The multiple state model in Figure 6.1, where transition intensities are assumed to be constant through time can be used to calculate conditional probabilities of the form $P_{ij}(0,t) = P(\text{life is in state } j \text{ at time } t \mid \text{life is in state } i \text{ at time } 0)$. Occupancy probabilities of the form $P_{\overline{ii}}(0,t) = P(\text{life is in state } i \text{ for all time from 0 to } t \mid \text{life is in state } i \text{ at time } 0)$ can also be determined using the multiple state model. An explicit formula for occupancy probabilities is

$$P_{\overline{ii}}(0,t) = \exp\left(-\int_0^t (\rho_r + \nu_r)\,dr\right) = \exp\left(-(\rho+\nu)t\right), \tag{6.1}$$

where the final equality only holds if transition intensities are constant through time.

The general transition probabilities are determined by solving a set of Kolmogorov forward differential equations. The derivation of these results, again for constant transition intensities, is given in Haberman and Pitacco (1999).

The modelling of the claim inception rate (or rate of disability onset) was described earlier in Chapter 3. A generalised linear model with a Poisson error and logarithmic link function was found to be suitable amongst the class of commonly used generalised linear models. This model highlighted a number of statistically significant dependencies between claim incidence and various rating factors or characteristics of insured lives. The statistically significant factors were listed in Appendix 3.3.

The modelling of claim duration was considered in depth in Chapter 4. Mixture models from survival analysis (Maller et al, 1995) based on the generalised F family of probability distributions (Peng et al, 1998) were considered, and again a number of rating factors were found to be statistically significant predictors of claim duration. These results were shown in Tables 4.5 and 4.6. The survival analysis from Chapter 4 also highlighted how the claim termination rate is very closely related to the duration of claim. The smoothed empirical claim termination rates are shown in Figure 6.2 and the reduction in claim termination rates with duration is very clear.

Most evident from Chapter 4 was the decreasing rate of return to work with increasing duration. For simplicity in this chapter, termination rates are described with a generalised linear model with indicator variables used to describe the differing rates of claim termination as duration of claim varies. The GLM was again fitted using a Poisson error structure and logarithmic link function.

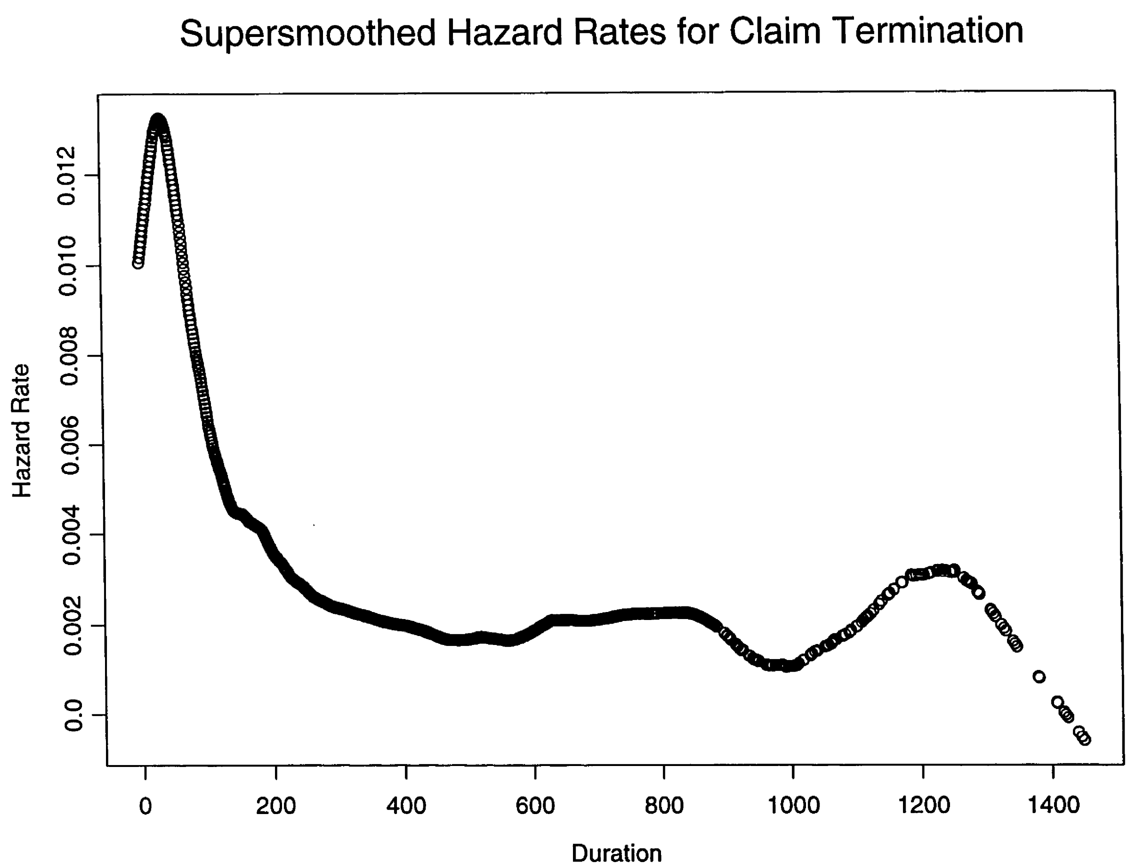


Figure 6.2 Supersmoothed Hazard Rates for Claim Termination

The fitted model for claim termination rates was therefore of the form

$$E(Y_i) = \exp\left(\beta_0 + \sum_{i=1}^{14} \beta_i X_i\right), \tag{6.2}$$

where Y_i denotes the claim termination rate for the ith life, X_i ($i = 1,2,\ldots,14$) are the independent predictor variables and the β coefficients are the estimated regression parameters. Table 6.1 gives the regression coefficients for each of the independent variables. Note that this model is a simplified version of the model presented in Chapter 3 for claim termination rates that is suited to the problem of DII premium rating and multiple state modelling.

Variable	GLM Regression Coefficient	Variable	GLM Regression Coefficient
Intercept	1.70011	Age greater than 50 years	-0.29768
Duration between 30 and 60 days	0.12094	Deferred Period between 28 and 89 days	-0.27829
Duration between 60 and 90 days	-0.22284	Deferred Period 90 days or greater	-0.98718
Duration between 90 and 180 days	-0.76698	Benefit Rate between 1500 and 2000 per month	-0.05445
Duration between 180 and 360 days	-1.46078	Benefit Rate between 2000 and 2500 per month	-0.07993
Duration greater than 360 days	-2.06876	Benefit Rate between 2500 and 3000 per month	-0.24049
Occupation either Class C or D	0.11541	Benefit Rate greater than 3000 per month	-0.10530
Age between 35 and 50 years	-0.15266		

Table 6.1 Estimated Regression Coefficients – Claim Termination Rate Analysis

Each of the variables included in this generalised linear model were statistically significant at the 5% level.

The duration dependent rates of claim termination imply that we are not able to use the results described above for constant transition intensities. Methods for calculation of transition probabilities and, in turn, risk premia for duration dependent recovery rates therefore need to be developed.

6.2 Duration Dependent Recovery Rates

Jones (1993) considers the problem of calculating transition probabilities for multiple state models, first where transition intensities are constant and then when transition intensities are assumed piecewise constant over various time intervals since policy inception but differ between time intervals. We first review the calculation of transition probabilities when transition intensities are constant. We first assume that $\mu_{ij}(t) = \mu_{ij}$. Define

$$P(S(u) = j \mid S(t) = i) = P_{ij}(u-t). \tag{6.3}$$

The transition intensities are then placed in an $N \times N$ matrix (in our case a 3×3 matrix). This matrix is denoted $M = \|\mu_{ij}\|$. Similarly, we define $P(z) = \|P_{ij}(z)\|$ with $P(0) = I$, the identity matrix. If we denote by $P'(z)$ the matrix with (i,j)'th entry $\frac{d}{dz}P_{ij}(z)$, the Kolmogorov forward equations can be written as $P'(z) = P(z)M$ with boundary condition $P(0) = I$.

Cox and Miller (1965) suggest a useful and quick method for solving the Kolmogorov forward differential equation in this case. The method involves writing

$$M = ADC, \tag{6.4}$$

where D is a diagonal matrix whose elements are the distinct eigenvalues of M, the i'th column of A is the eigenvector associated with the (i,i)'th element of A and C is the inverse of the matrix A. They then showed that

$$P(z) = A\text{diag}\left(e^{d_1 z},\dots,e^{d_N z}\right)C, \tag{6.5}$$

from which we obtain

$$P_{ij}(z) = \sum_{h=1}^{N} a_{ih} c_{hj} e^{d_h z}. \tag{6.6}$$

Hence, in order to find the transition probabilities between the three states of the multiple state model we need to be able to calculate, numerically, the eigenvalues and eigenvectors of the transition intensity matrix, M.

Jones (1993) extended the above methodology to allow for piecewise constant intensities. Under this proposal, the transition intensities are assumed to vary with time since the policy was sold. The transition intensities are assumed constant within set intervals of time but differ between separate intervals of time. To calculate the transition probabilities, Jones found that a suitable formula is

$$P_{ij}(t,u) = \sum_{h=1}^{N} b_{ih}^{(m_u)} c_{hj}^{(m_u)} e^{d_h^{(m_u)}\left(u-t_{m_u-1}\right)}, \tag{6.7}$$

where $b^{(m_u)}$ denote the elements of the eigenvector matrix formed under decomposition (6.4) based on the transition intensities that apply during time period (m_u). Similarly, $c^{(m_u)}$ denote the elements of the inverse of the eigenvector matrix formed using (6.4), and finally $d^{(m_u)}$ are the eigenvalues in the diagonal matrix D from the (6.4) decomposition applying to time period (m_u).

The above formulae only apply to transition intensities which vary with time since policy inception – for example they are useful in determining transition probabilities when transition intensities vary due to, for example, age. In the DII context, however, the recovery rate varies according to duration spent in the disabled state. This time differs from simply duration since policy inception for two reasons. First, the "policy inception" for the majority of policyholders does not mean that the policyholder becomes disabled. Second, the duration of most recent period of disability resets itself to zero each time that a life returns to the disabled state. It is possible for a life to be disabled for a period of three months and then recover and return to the disabled state one month later. The duration since disability onset returns to zero in this model when the second period of disability commences. The analysis by Jones therefore needs to be modified before it can be applied in this context.

Consider a tree diagram which indicates the health or disability of a life that has just bought a DII policy during the next one year period. The health or disability of the life is checked at each of 60, 120, 180 and 360 days after policy inception. Death is, of course, also possible and this possibility will mean that the sum of the probabilities for transitions from a particular node will sum up to slightly less than one. There are sixteen possible sequences of health and disability at the four time periods checked during the year after the policy is sold. The transitions for the first tree time periods are shown below in Figure 6.3.

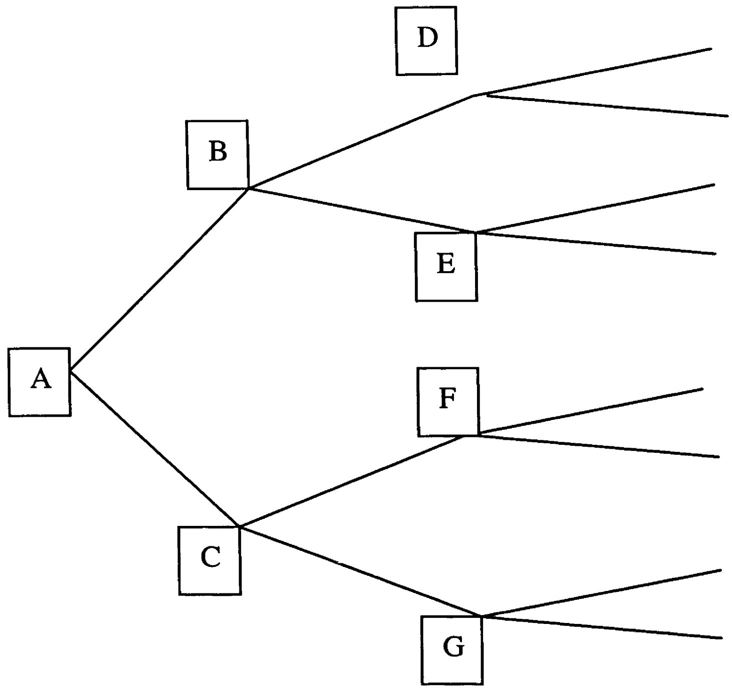


Figure 6.3 Three-step tree Diagram for Calculation of Premium Rates

An initial aim of this analysis is to determine the actuarially fair risk premium for a particular DII policy. We will consider a one-year DII policy sold to a life aged 35. The benefits provided under the policy are:

- if the life becomes disabled during the first year after purchasing the policy then benefits are paid during that year while the life is disabled equal to 75% of the salary immediately prior to paying the premium;

- if the life is disabled at the end of one year after purchasing the DII contract, then the life is paid at a rate equal to 75% of his salary while this disability continues. Once the life returns to the able state, payments under this insurance coverage cease.

For simplicity, in the determination of the risk premium, we will ignore the impact of expenses. Investment income will be assumed to earn on reserves held by the insurer at 5% per annum effective.

Define $P^*(0,t)$ to be the probability that the life is entitled to benefits at time *t*, where *t* is measured in days. Clearly for the first year (taken to be $t \leq 360$), this quantity is just the probability that the life is in the disabled state after *t* days given that the life was healthy at time zero. After the first year, $P^*(0,t)$ is the probability that the life was disabled at time one year and continues to be disabled for the period from exactly one year after policy purchase up until time *t*.

In order to use Figure 6.2 for determining risk premia we use the $P^*(0,t)$ values described above because these are the probabilities that a payment is made at time *t* under the contract. To derive the $P^*(0,t)$ values, we first determine the probabilities that the life reaches each of the 30 nodes in the tree. Consider the first probability p_1. This is the probability that a life who has just purchased a DII contract is healthy in sixty days' time. There may be any number of visits to the disabled state in the sixty days following purchase of the contract. This probability can be calculated using the method of Jones (1993), described above. The matrix of transition intensities for the first sixty days, where the able state is state 1, the disabled state is state 2 and the dead state is state 3, is

$$\begin{pmatrix} -\left(\lambda^{12}+\lambda^{13}\right) & \lambda^{12} & \lambda^{13} \\ \lambda^{21}_{(1)} & -\left(\lambda^{21}_{(1)}+\lambda^{23}\right) & \lambda^{23} \\ 0 & 0 & 0 \end{pmatrix}, \qquad (6.8)$$

where λ^{ij} is the transition intensity applicable to moving from state i to state j and, where the transition intensities from a particular state are dependent on time spent in a given state, we use the notation $\lambda^{ij}_{(m)}$ to denote the transition intensity from state i to state j that applies when the life has been in state i for an amount of time applicable to the m^{th} time grouping. For this study the groups of time are given in the Table 6.2 below. These groups are only relevant for states where exit transition intensities are duration dependent – in this case duration groupings are relevant only for the transition from the disabled to the able state.

Group Number	**Duration in Disabled State included in Group**
1	0 to 60 days
2	61-120 days
3	121 to 180 days
4	181 to 360 days
5	After 360 days

Table 6.2 Duration Bands for Recovery Transition Intensities

Using Jones (1993), we first construct the singular value decomposition of the matrix (6.8) and then use (6.7) to determine the probability that the life is in either the healthy or ill state after 60 days. This method is coded in S-Plus and the full code is given in Appendix 6.1.

This process is repeated to determine the probabilities for all other branches in Figure 6.3 in a recursive manner. At later branches, the probability calculation is more complicated due to allowance for varying durations of disability at the date when the probability calculation is required. For example, to determine the probability labelled p221 (that is the probability that a life who is disabled after both 60 and 120 days will be healthy after 180 days), we assume the life has been disabled for 90 days at time 120 – that is, we are implicitly assuming that disablement occurs uniformly over the interval from policy inception to duration 60 days for those who are disabled at duration 60 days. The required probability is therefore determined as the probability that a life who has been disabled for 90 days recovers in the next 30 days and then is still healthy after a subsequent 30 days plus the probability that the life who has been disabled for 90 days is disabled after a subsequent 30 days and then, having been disabled for 120 days recovers and is able after a further 30 days. This use of the law of total probability is required because different recovery rates apply for different durations of continuous disability.

Having found the probabilities of traversing each of the branches in the tree of Figure 6.3, we can easily determine, by simple multiplication of probabilities, the probability that the life reaches each of the nodes in the tree. These are determined by the node123 and node4 functions which are part of the major function shown in Appendix 6.1.

Next we need to determine the probabilities of payments that could be made more than one year after the policy is sold. These payments are restricted to those who are in the disabled state exactly one year after the policy is sold and continue in this state for a maximum of a further four years. That is, for this DII contract we assume a benefit period of five years. For a life who is healthy at each of 60, 120 and 180 days but is in the disabled state at time 360 days, we assume that this life has been in the disabled state for 90 days at the end of the year. The probability of ongoing disability is therefore calculated based on the recovery rates for a life who has been disabled for between 60 and 120 days for the first 30 days after one year, and then using recovery rates based on the period between 120 and 180 days for the next 60 days. The calculations of disability at these times are performed by the "prob" functions labelled L, M, N, O and P which also form part of the major function shown in Appendix 6.1.

If we define the probability of reaching each of the sixteen nodes after 360 days as, moving down the tree in Figure 6.3, as N411, N412, ..., N416, then the probability described in the previous paragraph for disability between 360 and 390 days can be written as

$$\Pr(\text{disabled for } t \in (361,390)) = \left(N42+N46+N410+N414\right)\exp\left(-\left(\lambda_{21}^{(2)}+\mu\right)(t-360)\right)$$
$$+\left(N44+N412+N48+N414\right)\exp\left(-\left(\lambda_{21}^{(4)}+\mu\right)(t-360)\right). \quad (6.9)$$

Note that for nodes N44, N412, N48 and N414, the life has been disabled for between 180 and 360 days at the end of the year, and hence we use the $\lambda_{21}^{(4)}$ recovery rates in the calculation.

6.3 Risk Premium Calculations

If we further define i as the annual effective rate of interest assumed to be earned on funds held by the insurance company, and B as the annual benefit amount paid while the life is in the disabled state, the risk premium is

$$\text{Risk Prem} = \sum_{i=1}^{1825} \frac{B}{365} \frac{P^*(0,t)}{(1+i)^{t/365}}. \quad (6.10)$$

This calculation is performed in the premium function, shown in Appendix 3.1. Note that the premiums are calculated for a \$10,000 annual benefit. Importantly, the premium increases for higher benefit levels are on a per dollar insured basis.

To illustrate the range of risk premiums applicable to lives with various characteristics, we perform the following calculations. The risk premium for a DII contract with a maximum allowable benefit of 5 years after the date that the policy is sold, where interest is assumed to earn at 5% per annum effective for the groups of lives shown in Table 6.2.

Age	20 to 60 at 5 year intervals
Occupation Class	Occupation Class A, B, C or D
Benefit Rate	<$1500, $1500-$2000, $2000-$2500, $2500-$3000, >$3000 per month
Deferred Period	<28 days, 28 to 89 days, 90 days or longer

Table 6.3 Groups of lives for risk premium calculations

The results of the risk premium calculations are shown in Figures 6.4, 6.5 and 6.6.

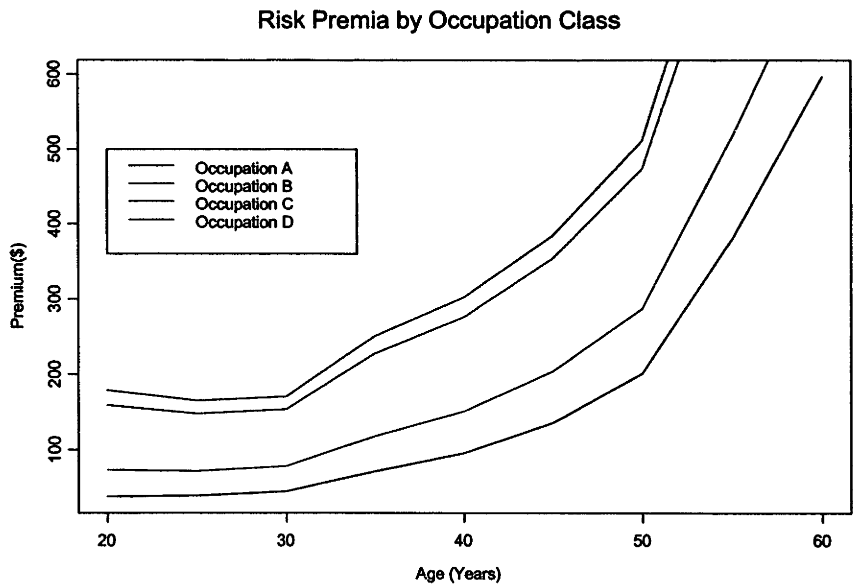


Figure 6.4 Risk Premia by Occupation Class

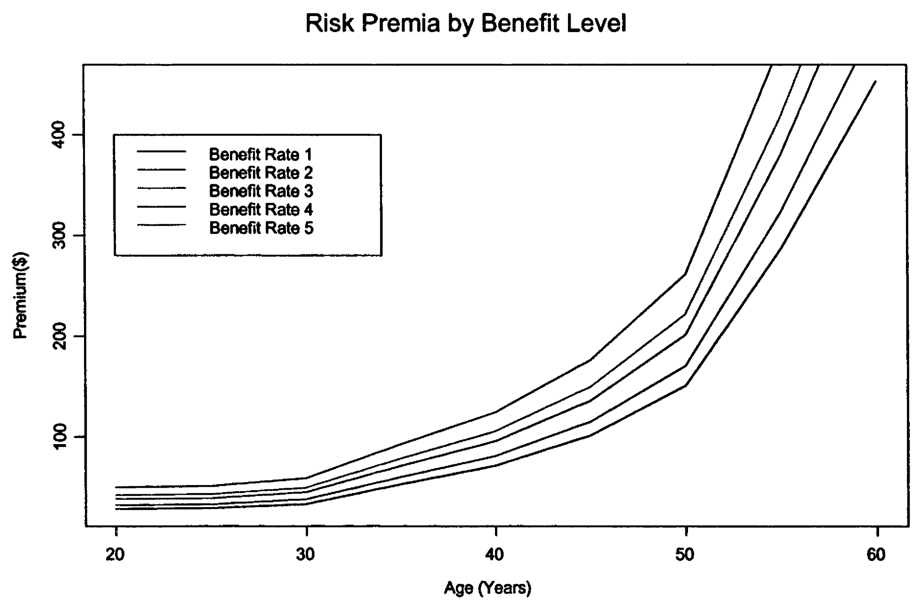


Figure 6.5 Risk Premia by Benefit Level

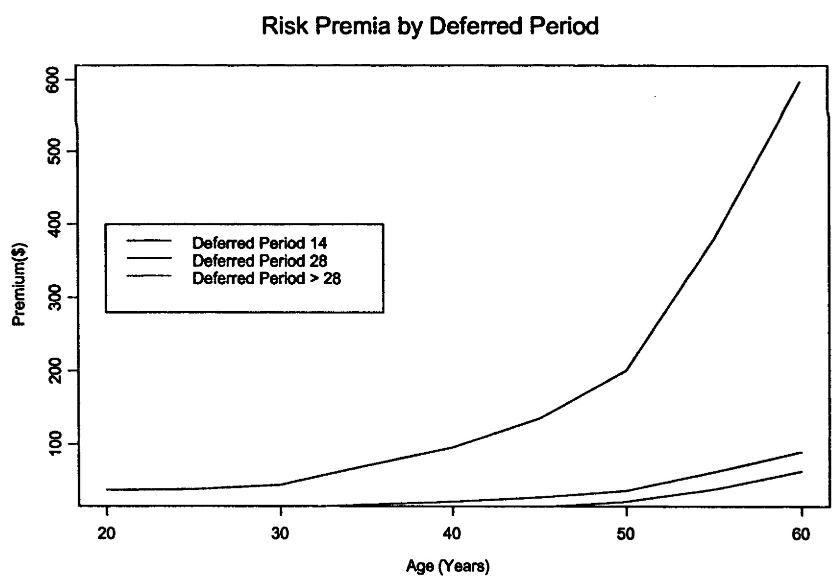


Figure 6.6 Risk Premia by Deferred Period

Salient features of this analysis include:

- A marked increase in premium rates with age across all levels of benefit, occupation class and deferred period
- a significant increase in the premium rate with age for the short deferred period. This is driven largely by the strong interaction between age and deferred period in the claim incidence model
- higher benefit amounts leading to higher per dollar of benefit premiums
- the fact that the higher claim incidence rates for occupation classes C and D more than offset the shorter duration claims for those in these occupation classes, resulting in higher premiums for occupation classes C and D compared to occupation classes A and B.

6.4 Flowgraph Models – An Alternative Pricing Methodology

Flowgraph models provide an alternative to multiple state model methodology for analysing systems where individuals move from one state to another through time. Diagramatically, flowgraphs look identical to multiple state models, however, they enable a greater flexibility in the choice of transition intensities and, in particular, they permit semi-Markov processes to be analysed more simply than is the case with traditional actuarial theory developed using multiple state Markov models.

A flowgraph consists of a number of states (outcomes) with (possibly bi-directional) arrows connecting those states between which transition is possible. These transitions are labelled with transmittances. A transmittance has two components: first, a probability that the transition will ultimately be made; and second, the moment generating function for the continuous random variable representing the time for transition between the two states. To fix ideas, we show in Figure 6.7 below a flowgraph that will be used for the pricing of DII contracts. Note that this model will be used to calculate a risk premium for a single one-year DII contract.

From Figure 6.7, it is clear that we assume five states labelled 1,2, 3, 4 and 5. We will label the transmittance between states i and j as $p_{ij}M_{ij}(t)$, where p_{ij} is the probability of transition between state i and state j and $M_{ij}(t)$ is the moment generating function for the random variable representing the amount of time spent from entry to state i until transition to state j.

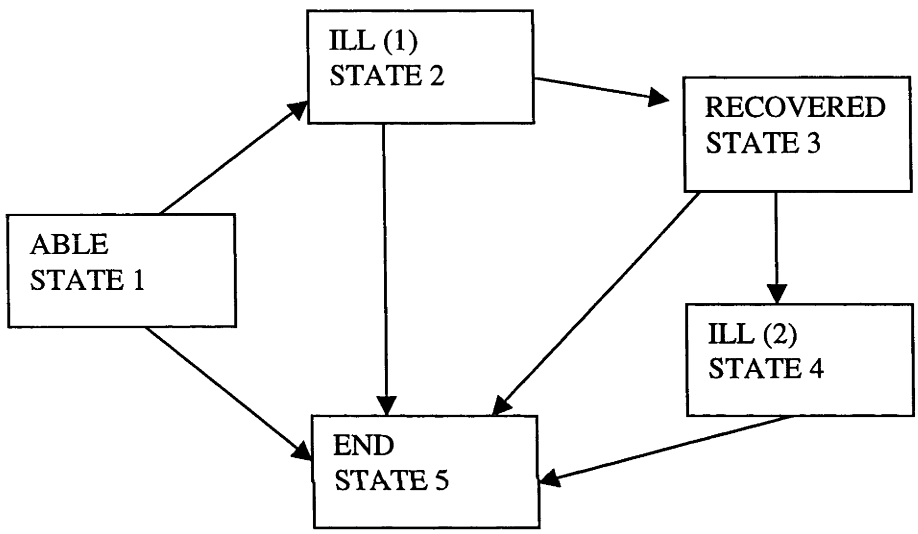


Figure 6.7 Flowgraph for Disability Income Insurance Contract

In order to use the above model for pricing a one-year DII contract, we aim to find a probabilistic description for the amount of time it takes an insured, after payment of the single upfront premium, to reach State 2 and State 4. These are clearly separate problems and flowgraph methodology (Huzarbazar, 2004) provides a solution to both problems.

Consider a flowgraph consisting of a number of states each connected with transmittances, as defined above. The flowgraph could consist of loops where it is possible for an individual to move between two states (or more) many times. Mason's Rule gives the calculation of the MGF (ie transmittance since transition from input to output is guaranteed) for the waiting time from an input state to an output state. Mason's rule states that the MGF from input to output, can be calculated using

$$M(s)=\frac{\sum_i P_i(s)\left[1+\sum_j(-1)^j L_j^i(s)\right]}{1+\sum_j(-1)^j L_j(s)}, \tag{6.11}$$

where the L terms refer to loops in the flowgraph. As depicted, Figure 6.7 contains no loops, and hence Mason's rule reduces to the sum of the products of the transmittances that link an input state to an output state. In this case, therefore, Mason's Rule reduces to the simple result that the moment generating function of the sum of a set of mutually independent random variables is just the product of the individual moment generating functions.

In order that we have a single input state and a single output state in our DII flowgraph model, we will temporarily consider a slightly altered flowgraph, shown below in Figure 6.8.

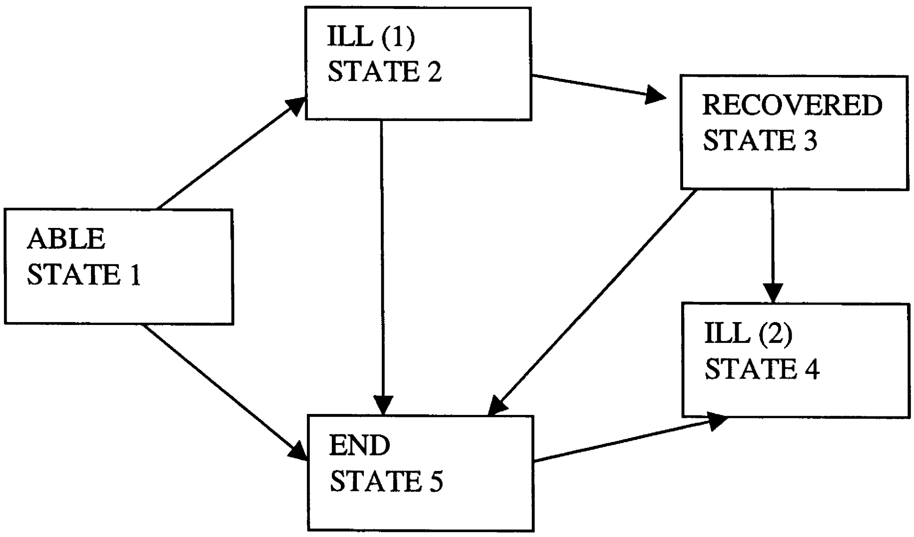


Figure 6.8 Revised Flowgraph with Single Output State

We now determine suitable transmittances for each of the arrows shown in Figure 6.8. This process will draw on previous investigations reported in this thesis.

In Chapter 3, it was found that a Poisson GLM with logarithmic link function was a suitable model for the description of claim incidence rates. Using the constant transition intensity implied by that model, we will use an exponential model for the transition between states 1 and 2, as well as and also for the transition between states 3 and 4 in Figure 6.7. Hence we assign the following moment generating functions:

$$M_{12}(t) = \frac{\lambda_{12}}{(\lambda_{12} - t)}, \tag{6.12}$$

and

$$M_{34}(t) = \frac{\lambda_{34}}{(\lambda_{34} - t)}, \tag{6.13}$$

where λ_{12} denotes the transition intensity for first onset of disability and λ_{34} denotes the transition intensity for onset of subsequent disability after a previous recovery.

Transition from state 2 to 3 represents recovery. This time, we refer to results from Chapter 4, of this thesis and, in particular, take note of the long-term survivor mixture models fitted there. The models developed in that chapter showed the reduction in the hazard rate for return to work for lives, particularly for long duration claims, exceeding six months. Given that in Figure 6.8 we are only interested in transitions during the first year from state 2 to state 3 (all other recoveries lead to a transition to state 5), we can again assume a constant transition intensity over this first year. This means that we will use the same form of moment generating function as given in (6.12) with λ_{12} replaced with λ_{23}.

We also define the following moment generating functions:

$$M_{15}(t) = M_{25}(t) = M_{35}(t) = e^{t}, \quad (6.14)$$

and

$$M_{54}(t) = e^{20t}. \quad (6.15)$$

Expression (6.14) gives the moment generating function for the transitions into the end state after one year for those lives that do not make the other possible transition during that period. Similarly, expression (6.15) gives the moment generating function for a certain transition twenty years after first moving into State 5. This transition is only included to ensure that State 4 is an output state – it will not affect the calculation of premiums for a DII contract under this model.

Using Mason's Rule, from (6.11), we obtain

$$\begin{aligned} M_{14}(t) = {} & p_{12} \cdot p_{23} \cdot p_{34} M_{12}(t) M_{23}(t) M_{34}(t) + p_{12} \cdot p_{23}(1 - p_{34}) M_{12}(t) M_{23}(t) M_{34}(t) M_{54}(t) + \\ & p_{12} \cdot p_{25} M_{12}(t) M_{25}(t) M_{54}(t) + p_{15} M_{15}(t) M_{54}(t). \end{aligned} \quad (6.16)$$

In order to calculate the single upfront premium we then use

$$\text{Premium} = \int_0^1 v^t f_{12}(t) \left(\frac{1 - M_{2(OUT)}(-\delta)}{\delta} \right) dt + \int_0^1 v^t f_{14}(t) \left(\frac{1 - M_{4(OUT)}(-\delta)}{\delta} \right) dt, \quad (6.17)$$

where $f_{12}(t)$ and $f_{14}(t)$ are the probability density functions for the time until transition from state 1 to state 2 and state 4, respectively, and $M_{2(OUT)}(-\delta)$ and $M_{4(OUT)}(-\delta)$ are the moment generating functions for transition from state 2 and state 4 to any other state in the flowgraph model. These moment generating functions, when evaluated at minus the continuously compounding rate of interest, represent the present value of \$1 payable at the time of transition out of state 2 or state 4.

The expression $\frac{1-M_{2(OUT)}(-\delta)}{\delta}$ therefore represents the present value of a stream of payments of \$1 payable continuously from when a life enters state 2 until the time of exit from state 2.

In order to be able to use (6.17) we need to be able to obtain approximations to $f_{12}(t)$ and $f_{14}(t)$. Using Mason's rule, we have the moment generating functions for the random variables with each of these probability density functions. In order to approximate these density functions, we need a suitable inversion technique to convert moment generating functions into approximate probability density functions. The method we use here is the saddlepoint approximation, see for example, Huzarbazar (2002), Reid (1988) or Goutis et al (1999).

If X is a random variable representing a waiting time, for example the amount of time for a life who has just purchased a one-year DII contract to become disabled for the first time, define $M_X(t)=E\left(e^{tX}\right)$, to be the moment generating function relating to an uknown density function $f_X(x)$. The cumulant generating function of the random variable X is defined as $K_X(t)=\log M_X(t)$. The saddlepoint approximation to the density of X is

$$\hat{f}_X(x)\approx\frac{\exp\left(K(\hat{s})-xs\right)}{\sqrt{2\pi K''(\hat{s})}}, \tag{6.18}$$

where $\hat{s}$ is the saddlepoint given by the solution to the equation $K'(\hat{s})=x$.

Consider specifically the approximation of $f_{14}(x)$, which is needed for our premium calculations. To facilitate the calculations involved, Mathematica software was used to find the saddlepoint approximation. The first step is to find a series of saddlepoints corresponding to durations for which we want to approximate the density function. For premium calculation, we are concerned with incidences of disability that occur in the first year after initial purchase of the policy. We therefore calculate a vector of saddlepoints (of length 100) evenly spaced over the year of DII coverage. Given these saddlepoints, we then apply (6.18) to approximate the density function, $f_{14}(x)$.

In order to evaluate expression (6.17), we also need formulae for each of $M_{2(OUT)}(t)$ and $M_{4(OUT)}(t)$. From Chapter 4, it is clear that a simple exponential model will not suffice here - we clearly need a probability distribution with a reducing hazard rate. For simplicity here we have adopted a piecewise exponential model. Srinivas and Pitt (2004) have applied regression trees to the problem of determining recovery transition intensities at various claim durations. They estimated separate transition intensities for the first 60 days post disability onset, 61 to 183 days, 184 days to 1.5 years, and greater than 5 years. We will assume that all claims which are continuing after five years cease at that time as a result of a five year benefit period.

The moment generating function for a piecewise exponential density with hazard rates given in Table 6.4 is

$$M_{2OUT}(t) = M_{45}(t) = \frac{\lambda_1}{t-\lambda_1}\left(e^{t_1(t-\lambda_1)}\right) + \frac{\lambda_2 e^{t_1(\lambda_2-\lambda_1)}}{t-\lambda_2}\left(e^{t_2(t-\lambda_2)} - e^{t_1(t-\lambda_2)}\right)$$

$$+\frac{\lambda_3 e^{-\lambda_1 t_1} e^{-\lambda_2(t_2-t_1)} e^{\lambda_3(t_1+t_2)}}{t-\lambda_3}\left(e^{t_3(t-\lambda_3)} - e^{t_2(t-\lambda_3)}\right) \tag{6.19}$$

$$+\frac{\lambda_4 e^{-(t_1\lambda_1+(t_2-t_1)\lambda_2+\lambda_3)} e^{1.5\lambda_4}}{t-\lambda_4}\left(e^{t_4(t-\lambda_4)} - e^{t_3(t-\lambda_4)}\right) + e^{5t} e^{-(\lambda_1 t_1+\lambda_2(t_2-t_1)+\lambda_3+3.5\lambda_4)}.$$

Time Period	**Hazard Rate**
$(0,t_1)$	λ_1
(t_1,t_2)	λ_2
(t_2,t_3)	λ_3
(t_3,t_4)	λ_4
t_4	Point Mass Probability of Claim Termination

Table 6.4 Piecewise Exponential Model Claim Recovery Intensities

The evaluation of the premium then proceeds using numerical integration techniques applied to (6.17). As an example, we consider the following set of parameters:

Parameter	**Explanation**	**Value**
p_{12}	Probability of initial incidence of disability in a one-year period	0.0578
p_{23}	Probability of recovery within a one-year period	0.8096
p_{34}	Probability of onset of subsequent disability after	0.5000

	recovery from previous disability	
λ_1	Recovery transition intensity applying for the first 60 days after onset of disability	1.491
λ_2	Recovery transition intensity applying between 61 and 183 days after onset of disability	0.608
λ_3	Recovery transition intensity applying between 184 and 1.5 years after onset of disability	0.784
λ_4	Recovery transition intensity applying between 1.5 years and 5 years after onset of disability	0.496
λ	Exponential distribution parameter for transitions from states 1 to 2, states 2 to 3 and states 3 to 4	2.00

Table 6.5 Example Flowgraph Model Parameters

Using Mathematica to perform the required calculations, we derive a premium of \$897.21 per annum for a \$50,000 annual income. Since the 5% incidence rate is at the higher end of incidence rates predicted in Chapter 3 and also high in the context of the current Australian industry disability table (IAD1989-93), it follows that the predicted premium is at the upper end of the scale for DII annual premia.

Flowgraph methods provide a number of opportunities for greater flexibility in modelling where multiple state models have been used frequently in the past. Their application is particularly suited to actuarial investigations because:

- they use moment generating functions to specify transmittances and moment generating functions intersect with expected present values which have been commonly used for premium rating and reserving in actuarial science for over a century;

- they enable easier manipulation of non-constant transition intensities in a multi-state framework. In particular, any probability distribution for which the moment generating function has a closed form can readily be used, and density functions for the duration until transition between non-adjacent states can be determined using saddlepoint approximations.

Further application of flowgraph modelling in actuarial science is certainly possible. In particular, Taylor (2002) considers a four state model for weekly compensation benefits. The model is depicted below in Figure 6.9.

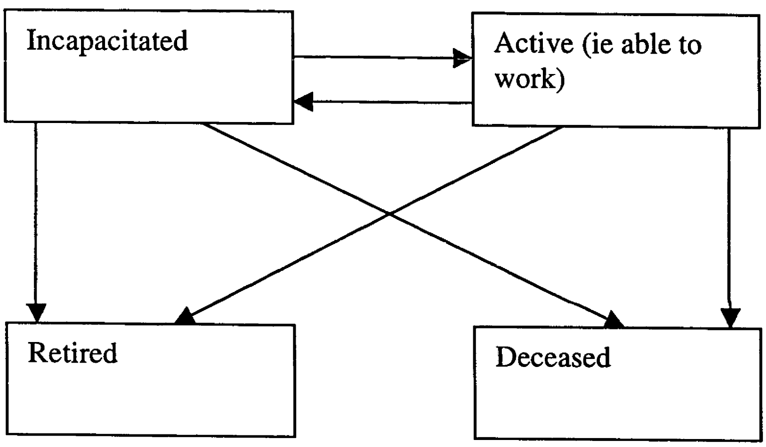


Figure 6.9 Flowgraph for transitions between statuses: weekly compensation

This model is again one that is amenable to flowgraph analysis, this is particularly because the random variable for the time spent in a particular state before transition from that state can potentially be modelled using a number of standard right-skewed variates commonly used in survival analysis.

CHAPTER SEVEN

SUMMARY AND CONCLUSIONS

This chapter will provide an overview of the main research findings of this thesis. It will also provide details of some avenues for further research.

7.1 Main Research Findings in this Thesis

The generalised linear modelling analysis presented in Chapter 3 has led to considerable discussion amongst actuaries working both in the DII market and in the reinsurance sector. This discussion has centred around the development of a new industry table for claim incidence rates and claim termination rates which uses generalised linear modelling as opposed to multi-way analysis used to construct the existing Australian disability table. The analysis from that chapter highlighted a number of significant rating variables that should be considered in the pricing and reserving of DII that have not been considered in the past. These include benefit amount, replacement ratio and the various interactions reported in that Chapter.

Chapter 4 introduced techniques from survival analysis to the disability claim termination rate discussion. Models that incorporate both hazard rates for return to work and a probability of not ever returning to work were developed. A Generalised F mixture model was found to provide a very good description of claim termination rates. The results from this chapter were extended to the problem of premium rating and reserving.

Chapter 5 demonstrated the use of censored regression quantile analysis, developed recently, to the DII claim duration arena. The most important result from that chapter was that the covariate sensitivities of claim termination rates do vary significantly for age, occupation class and deferred period. This finding has clear implications for actuaries who are required to recommend provisions that life insurers should hold in respect of existing disability claims. If an insurer were to rely on traditional mean regression analysis to assess the impact of increasing age on claim termination rates, their analysis would underestimate the true impact of age for a claim that has been in force for longer than, say, six months. Typically, insurers' balance sheets will contain material provisions in respect of DII claims for which the claimant has been receiving benefits for a long period before the date of valuation. The results of Chapter 5 and the method illustrated there are therefore commercially useful for insurers required to value the liabilities associated with a portfolio of diverse DII contracts.

Chapter 6 considered the actuarial pricing function in some depth. First a multiple state model, with piecewise exponential transition intensities for claim termination rates and constant claim incidence rates was developed. The duration-dependent transition intensities for claim termination meant that a tree based model was useful where nodes of the tree represent claimants at different claim durations. This numerical approach was useful for calculation of premium rates in the presence of semi-Markov claim termination rates, however the procedure described there quickly becomes very complex if greater refinement to the claim recovery rates is desired.

To overcome this problem, the second half of Chapter 6 provided an alternative solution to the problem of premium rating using flowgraph models. Flowgraphs are a very recent innovation in statistical science and their application to date has mainly been in the area of biostatistics. Importantly, innovative use of flowgraphs enables the user to calculate actuarially justified premium rates when transition intensities between the able, ill and dead states use probability models that have a closed form moment generating function.

7.2 Avenues for Further Research

The impact of economic variables, such as the level of interest rates and unemployment in particular, on the performance of DII business, have long been considered important by practising actuaries. The quantification of the impact of these variables on transition intensities and lapse rates for DII is an area that has not received much attention in Australia. This provides an interesting area for future research, and within the premium rating framework of flowgraph models, the impact of economic variables on both premium rates and recommended liability valuations could be quantified.

The use of generalised additive models which enable the use of a range of smoothers (for example, kernel smoothers and local linear loess estimates) provides another avenue for further research in this area.

The variability of predicted insurance outcomes is one that has received far less attention than it should have in the past in actuarial science. This thesis has taken a step towards solving this problem with its presentation of stochastic reserving results and pricing strategies. The bootstrap has recently been applied by general insurance actuaries in the determination of the variability of outstanding claims provisions. There is certainly scope to extend the work presented in Chapter 6 using flowgraphs and a nonparametric bootstrap based method. This work has the potential to give insurers a greater understanding of the extent of variability and probability of adequacy of the reserves held in their balance sheets.

The mortality rates of DII policyholders is another area that has received little attention. The reason for this is the lack of credible data that has been available on this context in Australia until recently. Research in the UK, based on limited amounts of data, has suggested that mortality rates for disabled lives are duration dependent, (Waters, 1991). In particular, rates of mortality are considerably higher than the equivalent level for insured lives, for those disabled lives that have only become disabled in the last few months. The Australian disability database continues to collect information on cause of claim. There is certainly scope to investigate mortality rates as a function of cause of claim and duration of disability. The results from such a study, would have clear implications for pricing and reserving of DII contracts and could also contribute to the pricing of life risk products in Australia.

APPENDIX 3.1 – Data Specifications

A1. Claims Excluded

Country Not Australia

Deferment Sickness Not Equal Deferment Period Accident

Sickness Only cover or Accident Only cover

A2. Summarised Characteristics

Deferment

The original data has deferment period in days. These were summarized as follows.

Actual Duration	Assumed Duration
0 – 10	7 days
11 – 24	14 days
25 – 31	30 days
32 – 61	60 days
62 – 92	90 days
93 – 183	180 days
184 +	365 days

CoTerminus

If Benefit Period Accident = Benefit Period Sickness, then CoTerminus.

Proportionate Benefit

If at any time a claim payment is not the full benefit then the claim is recorded as Partial

Age at Claim

This is recorded in quinquennial steps starting at age 22.

Benefit Rate

Recorded in the following bands ($ per month)

Actual Benefit Rate	Assumed Benefit Rate
0 – 999	500
1000 – 1999	1,500
2000 – 2999	2,500
3000 – 3999	3,500
4000 – 4999	4,500
5000 – 5999	5,500
6000 – 6999	6,500
7000 – 9999	8,500
10000 – 14999	12,500
15000 – 19999	17,500
20000 -	25,000

Claim Cause

All claims had an alpha cause coded. These were summarized as follows in order to reduce the number of possible cells in the data matrix. The choice of combinations was deliberately made based on the average claim duration as shown in the Disability Committee Reports so as to give three broad groups – short, medium and long.

Summarised Cause	**Original Causes**
V	None recorded
W	A, H, I, J, K, L
X	C, D, F, G, M. N, P, R, S
Y	B, E,
Z	Accident

The original causes are classified according to the WHO International Classification of Diseases as follows.

A	Infective and parasitic diseases
B	Neoplasms (MN = Malignant and BN = Benign)
C	Endocrine, Nutritional and Metabolic diseases
D	Diseases of the blood and blood forming organs
E	Mental disorders
F	Diseases of the Nervous system and sense organs
G	Diseases of the circulatory system
H	Diseases of the respiratory system

I	Diseases of the digestive system
J	Diseases of the genito-urinary system
K	Diseases of Pregnancy and childbirth
L	Diseases of the skin and subcutaneous tissue
M	Diseases of the musculoskeletal system and connective tissue
N	Congenital anomalies
P	Senility and ill defined conditions
Q	Accidents, poisoning and violence (external causes)
R	AIDS related complex and full blown AIDS
S	HIV+ and Lymphadenopathy

Appendix 3.2 – Detailed Actual / Experience Results

3.2.1 Experience by Calendar Year of Exposure

Year	Expected	Actual	Index
1980	19	26	74%
1981	79	92	86%
1982	226	261	87%
1983	515	651	79%
1984	629	715	88%
1985	763	826	92%
1986	1202	1146	105%
1987	2090	2381	88%
1988	2142	2580	83%
1989	2352	2513	94%
1990	3261	2681	122%
1991	4895	4625	106%
1992	5148	5082	101%
1993	6447	5736	112%
1994	6960	6369	109%
1995	7737	5840	132%
1996	8166	5658	144%
1997	7490	5862	128%
1998	5477	4830	113%
Total	**65600**	**57874**	**113%**

3.2.2. Experience by Gender and Year

	Male			Female		
Year	**Expected**	**Actual**	**Index**	**Expected**	**Actual**	**Index**
1980	19	26	72%	0	0	0%
1981	75	79	95%	5	13	35%
1982	202	240	84%	24	21	115%
1983	453	599	76%	62	52	119%
1984	561	646	87%	67	69	98%
1985	685	757	91%	77	69	112%
1986	1094	1030	106%	108	116	93%
1987	1870	2151	87%	220	230	96%
1988	1864	2295	81%	278	285	98%
1989	2058	2220	93%	294	293	100%
1990	2847	2384	119%	415	297	140%
1991	4330	4104	106%	565	521	108%
1992	4442	4467	99%	707	615	115%
1993	5596	4983	112%	850	753	113%
1994	6056	5628	108%	904	741	122%
1995	6695	5143	130%	1043	697	150%
1996	7088	4972	143%	1078	686	157%
1997	6444	5117	126%	1046	745	140%
1998	4742	4270	111%	736	560	131%
Total	**57120**	**51111**	**112%**	**8480**	**6763**	**125%**

3.2.3 Experience by Occupation and Year

	Occupation Class A			Occupation Class B		
Year	**Expected**	**Actual**	**Index**	**Expected**	**Actual**	**Index**
1980	15	17	86%	-	-	-
1981	30	27	111%	1	0	0%
1982	63	61	103%	6	6	108%
1983	138	215	64%	60	60	100%
1984	199	247	81%	70	53	131%
1985	242	260	93%	63	39	161%
1986	282	280	101%	178	70	254%
1987	445	457	97%	368	387	95%
1988	462	471	98%	365	433	84%
1989	547	549	100%	396	415	95%
1990	792	488	162%	444	366	121%
1991	1277	1117	114%	521	437	119%
1992	1486	1320	113%	484	517	94%
1993	1900	1572	121%	565	562	101%
1994	2042	1753	116%	602	606	99%
1995	2252	1564	144%	549	391	140%
1996	2348	1452	162%	409	360	114%
1997	2248	1594	141%	465	349	133%
1998	1642	1210	136%	255	190	134%
Total	**18410**	**14654**	**126%**	**5800**	**5241**	**111%**

3.2.3 Experience by Occupation and Year (Continued)

	Occupation Class C			Occupation Class D		
Year	**Expected**	**Actual**	**Index**	**Expected**	**Actual**	**Index**
1980	1	0	0%	4	9	42%
1981	13	19	68%	36	46	78%
1982	61	57	107%	96	137	70%
1983	201	248	81%	116	128	90%
1984	176	215	82%	184	200	92%
1985	185	193	96%	273	334	82%
1986	369	284	130%	373	512	73%
1987	868	1046	83%	409	491	83%
1988	913	1128	81%	402	548	73%
1989	957	1114	86%	452	435	104%
1990	1290	1143	113%	736	684	108%
1991	1692	1642	103%	1406	1429	98%
1992	1700	1835	93%	1479	1410	105%
1993	2078	1955	106%	1904	1647	116%
1994	2440	2290	107%	1876	1720	109%
1995	2758	2083	132%	2178	1802	121%
1996	2901	2242	129%	2508	1604	156%
1997	2802	2299	122%	1976	1620	122%
1998	1988	1978	101%	1593	1452	110%
Total	**23391**	**21771**	**107%**	**17999**	**16208**	**111%**

3.2.4 Experience by Deferment and Year

	7 Days			14 Days		
Year	**Expected**	**Actual**	**Index**	**Expected**	**Actual**	**Index**
1980	0	0	0%	4	11	38%
1981	2	0	0%	53	61	87%
1982	4	0	0%	140	169	83%
1983	42	93	46%	311	403	77%
1984	51	105	49%	372	394	94%
1985	53	83	64%	469	499	94%
1986	51	70	72%	846	756	112%
1987	41	78	53%	1590	1901	84%
1988	44	66	66%	1606	2054	78%
1989	47	69	68%	1763	1897	93%
1990	39	48	80%	2351	2021	116%
1991	44	61	72%	3508	3510	100%
1992	34	36	95%	3542	3708	96%
1993	20	30	68%	4245	4002	106%
1994	35	32	109%	4497	4329	104%
1995	24	31	76%	4849	3771	129%
1996	2	4	52%	5195	3667	142%
1997			0%	4392	3556	124%
1998	2	4	43%	2959	2706	109%
Total	**535**	**810**	**66%**	**42693**	**39415**	**108%**

3.2.4. Experience by Deferment and Year (Continued)

	30 Days			90 Days		
Year	**Expected**	**Actual**	**Index**	**Expected**	**Actual**	**Index**
1980	7	14	50%	8	1	766%
1981	17	30	58%	7	1	690%
1982	74	92	81%	8	0	0%
1983	149	151	99%	12	4	291%
1984	192	209	92%	13	7	190%
1985	224	241	93%	16	3	541%
1986	280	306	92%	25	14	181%
1987	411	389	106%	47	13	364%
1988	454	445	102%	38	15	254%
1989	502	532	94%	41	15	273%
1990	835	606	138%	37	6	611%
1991	1298	1041	125%	45	13	346%
1992	1516	1316	115%	56	22	256%
1993	2120	1679	126%	61	25	245%
1994	2366	1969	120%	62	39	160%
1995	2804	2010	140%	61	28	216%
1996	2910	1955	149%	60	32	187%
1997	3014	2268	133%	84	38	222%
1998	2453	2092	117%	64	28	228%
Total	**21627**	**17345**	**125%**	**745**	**304**	**245%**

3.2.5. Experience by Disability Definition and Year

	Own / Any 2			Own			Any		
Year	**Expected**	**Actual**	**Index**	**Expected**	**Actual**	**Index**	**Expected**	**Actual**	**Index**
1980	3	2	153%	15	22	66%	1	2	43%
1981	23	18	127%	50	73	68%	5	1	456%
1982	100	94	107%	101	130	78%	24	37	66%
1983	340	491	69%	147	116	127%	27	44	61%
1984	384	452	85%	207	211	98%	31	52	59%
1985	452	479	94%	224	239	94%	77	108	71%
1986	624	341	183%	303	345	88%	199	356	56%
1987	1281	1449	88%	494	483	102%	222	313	71%
1988	1320	1635	81%	544	540	101%	214	299	72%
1989	1431	1582	90%	550	488	113%	308	341	90%
1990	1953	1422	137%	956	774	124%	288	413	70%
1991	2891	2843	102%	1613	1317	122%	316	365	87%
1992	2908	3073	95%	1834	1604	114%	332	333	100%
1993	3587	3231	111%	2391	1965	122%	408	489	83%
1994	3649	3498	104%	2664	2266	118%	577	540	107%
1995	3454	2681	129%	3357	2426	138%	815	610	134%
1996	3751	2601	144%	3471	2409	144%	871	573	152%
1997	3108	2346	132%	3736	2826	132%	506	541	94%
1998	2139	2135	100%	2677	1890	142%	364	447	81%
Total	**33399**	**30373**	**110%**	**25332**	**20124**	**126%**	**5585**	**5864**	**95%**

3.2.6. Experience by Benefit Type and Year

	Level			Increasing			Level - Out of Working Hours		
Year	Expected	Actual	Index	Expected	Actual	Index	Expected	Actual	Index
1980	13	19	71%	6	7	81%	-	-	-
1981	61	71	85%	19	21	89%	-	-	-
1982	177	204	87%	50	57	88%	-	-	-
1983	404	544	74%	111	107	103%	-	-	-
1984	482	572	84%	143	143	100%	-	-	-
1985	595	661	90%	166	165	100%	-	-	-
1986	951	918	104%	248	228	109%	-	-	-
1987	1616	1860	87%	472	521	91%	-	-	-
1988	1522	1924	79%	617	656	94%	-	-	-
1989	1604	1697	94%	740	816	91%	1	0	0%
1990	2119	1843	115%	1112	838	133%	16	0	0%
1991	2849	2795	102%	1909	1700	112%	56	83	68%
1992	2692	2844	95%	2297	2097	110%	69	86	80%
1993	3006	2921	103%	3201	2659	120%	81	70	115%
1994	3018	2884	105%	3691	3285	112%	91	100	91%
1995	3092	2451	126%	4319	3251	133%	78	51	154%
1996	3275	2369	138%	4567	3145	145%	61	70	87%
1997	2563	2085	123%	4432	3621	122%	50	52	96%
1998	1701	1627	105%	2760	2554	108%	23	22	106%
Total	**31739**	**30289**	**105%**	**30859**	**25871**	**119%**	**527**	**534**	**99%**

3.2.7 Experience by Medical Evidence and Year

	Medical			Non Medical			Other		
Year	**Expected**	**Actual**	**Index**	**Expected**	**Actual**	**Index**	**Expected**	**Actual**	**Index**
1980	7	7	93%	9	19	46%	4	0	0%
1981	15	10	154%	51	82	62%	13	0	0%
1982	49	53	93%	146	208	70%	32	0	0%
1983	190	281	68%	284	370	77%	41	0	0%
1984	233	345	67%	304	370	82%	92	0	0%
1985	328	411	80%	321	399	80%	113	16	706%
1986	377	590	64%	517	463	112%	285	72	395%
1987	432	560	77%	961	1088	88%	659	711	93%
1988	343	518	66%	959	1227	78%	790	787	100%
1989	329	488	67%	963	1096	88%	1010	886	114%
1990	281	283	99%	1413	1453	97%	1525	910	168%
1991	340	390	87%	2217	2387	93%	2309	1824	127%
1992	266	270	99%	1526	1465	104%	3338	3347	100%
1993	231	268	86%	1877	1713	110%	4311	3753	115%
1994	281	268	105%	2280	2273	100%	4358	3828	114%
1995	332	287	116%	2455	2185	112%	4900	3368	145%
1996	240	148	162%	2440	1908	128%	5396	3602	150%
1997	160	155	103%	2339	2300	102%	4855	3407	143%
1998	78	85	91%	1388	1238	112%	3759	3505	107%
Total	**4511**	**5417**	**83%**	**22449**	**22244**	**101%**	**37788**	**30016**	**126%**

3.2.8 Experience by Coverage Type and Year

	Individual			Business Overheads		
Year	**Expected**	**Actual**	**Index**	**Expected**	**Actual**	**Index**
1980	19	26	74%	-	-	-
1981	79	92	86%	-	-	-
1982	226	261	87%	-	-	-
1983	515	651	79%	-	-	-
1984	629	715	88%	-	-	-
1985	763	826	92%	2	0	0%
1986	1202	1146	105%	6	0	0%
1987	2090	2381	88%	18	28	66%
1988	2142	2580	83%	29	21	138%
1989	2352	2513	94%	57	48	118%
1990	3261	2681	122%	86	49	175%
1991	4895	4625	106%	178	155	115%
1992	5148	5082	101%	216	200	108%
1993	6447	5736	112%	308	287	107%
1994	6960	6369	109%	226	94	241%
1995	7737	5840	132%	369	312	118%
1996	8166	5658	144%	328	265	124%
1997	7490	5862	128%	251	187	134%
1998	5477	4830	113%	195	158	123%
Total	**65600**	**57874**	**113%**	**2269**	**1804**	**126%**

3.2.9 Experience by Contract Type and Year

	Level - Guaranteed			Level - Non Guaranteed			Stepped - Guaranteed			Stepped - Non Guaranteed		
Year	**Exp.**	**Actual**	**Index**	**Exp.**	**Actual**	**Index**	**Exp.**	**Actual**	**Index**	**Exp.**	**Actual**	**Inde**
1980	12	6	199%	1	2	46%				6	18	35%
1981	11	6	181%	14	19	73%	4	0	0%	51	67	76%
1982	20	6	326%	24	36	68%	21	36	59%	161	183	88%
1983	86	138	62%	81	36	226%	25	44	58%	323	433	75%
1984	122	157	77%	106	134	79%	27	51	53%	374	373	100
1985	125	133	94%	94	132	72%	73	104	70%	470	457	103
1986	134	127	106%	153	159	96%	166	290	57%	749	570	131
1987	159	144	110%	331	342	97%	183	231	79%	1418	1664	85%
1988	147	144	102%	307	359	85%	191	298	64%	1497	1779	84%
1989	213	142	150%	310	286	108%	200	307	65%	1629	1778	92%
1990	226	294	77%	381	322	118%	194	220	88%	2461	1845	133
1991	119	103	116%	522	432	121%	316	365	87%	3938	3725	106
1992	160	89	180%	459	486	94%	331	333	100%	4198	4174	101
1993	189	134	141%	469	453	104%	411	486	85%	5377	4663	115
1994	183	138	133%	486	486	100%	585	540	108%	5705	5205	110
1995	185	112	165%	639	531	120%	830	610	136%	6084	4587	133
1996	74	48	155%	491	372	132%	894	583	153%	6707	4655	144
1997	50	23	218%	491	335	146%	514	542	95%	6435	4962	130
1998	1	0	0%	549	483	114%	364	447	81%	4563	3900	117
Total	**2215**	**1944**	**114%**	**5909**	**5405**	**109%**	**5332**	**5487**	**97%**	**52145**	**45038**	**116**

3.210 Experience by No Claim Bonus and Year

	No NCB			Yes NCB		
Year	**Expected**	**Actual**	**Index**	**Expected**	**Actual**	**Index**
1980	13	8	165%	6	18	33%
1981	34	16	213%	45	76	60%
1982	128	93	137%	99	167	59%
1983	417	479	87%	98	172	57%
1984	516	555	93%	113	160	71%
1985	649	700	93%	114	126	90%
1986	996	960	104%	206	186	111%
1987	1669	1930	86%	421	451	93%
1988	1613	1902	85%	529	678	78%
1989	1614	1633	99%	738	880	84%
1990	2227	1954	114%	1034	727	142%
1991	3371	3298	102%	1524	1327	115%
1992	3353	3363	100%	1796	1719	104%
1993	4418	3865	114%	2028	1871	108%
1994	4967	4445	112%	1993	1924	104%
1995	5639	4100	138%	2098	1740	121%
1996	5844	3954	148%	2322	1704	136%
1997	5244	4116	127%	2246	1746	129%
1998	3806	3565	107%	1672	1265	132%
Total	**46517**	**40936**	**114%**	**19083**	**16937**	**113%**

3.2.11 Experience by Smoker Status and Year

	Non Smoker - Periodic Checks			Non Smoker		
Year	Expected	Actual	Index	Expected	Actual	Index
1980	-	-	-			
1981	-	-	-	0	0	0%
1982	-	-	-	1	0	0%
1983	-	-	-	3	0	0%
1984	0	0	0%	2	0	0%
1985	4	0	0%	7	0	0%
1986	57	34	168%	107	108	99%
1987	110	137	80%	328	486	68%
1988	158	210	75%	478	680	70%
1989	187	178	105%	659	730	90%
1990	209	237	88%	1125	1005	112%
1991	147	108	136%	2234	2190	102%
1992	239	192	125%	2531	2709	93%
1993	427	301	142%	3295	3172	104%
1994	595	542	110%	3745	3577	105%
1995	400	355	113%	4660	3631	128%
1996	340	333	102%	4989	3564	140%
1997	221	208	106%	4865	3930	124%
1998	316	359	88%	3501	3040	115%
Total	**3411**	**3194**	**107%**	**32530**	**28822**	**113%**

3.2.12. Experience by Claim Cause and Year

	Unknown			Combined Cause W			Combined Cause X			Combined Cause Y		
Year	**Exp.**	**Actual**	**Index**	**Exp.**	**Actual**	**Index**	**Exp.**	**Actual**	**Index**	**Exp.**	**Actual**	**Index**
1980				2	8	31%	4	6	73%	5	0	0%
1981	6	0	0%	10	25	41%	19	17	111%	12	5	237%
1982	12	0	0%	44	63	69%	44	56	79%	11	10	108%
1983	8	1	0%	79	174	45%	125	134	93%	56	52	107%
1984	27	0	0%	107	219	49%	156	150	104%	66	55	121%
1985	20	0	0%	103	194	53%	215	208	103%	83	69	121%
1986	31	1	0%	178	285	62%	363	266	137%	97	77	126%
1987	29	0	0%	326	596	55%	621	539	115%	209	165	127%
1988	41	6	0%	290	629	46%	635	565	112%	257	174	148%
1989	99	0	0%	365	635	57%	763	721	106%	239	158	151%
1990	431	10	4313%	416	634	66%	944	648	146%	308	186	166%
1991	1414	1449	98%	555	773	72%	1066	790	135%	431	226	191%
1992	1706	1689	101%	570	872	65%	1053	797	132%	487	269	181%
1993	2229	1989	112%	628	873	72%	1351	941	144%	633	339	187%
1994	1866	1931	97%	688	1005	68%	1598	1029	155%	830	475	175%
1995	701	542	129%	878	1107	79%	2327	1289	181%	1161	549	211%
1996	515	322	160%	855	1041	82%	2452	1270	193%	1388	565	246%
1997	307	245	125%	854	1088	78%	2197	1328	165%	1414	644	220%
1998	8	9	85%	685	872	79%	1727	1240	139%	1023	568	180%
Total	**9451**	**8194**	**115%**	**7632**	**11093**	**69%**	**17660**	**11994**	**147%**	**8710**	**4586**	**190%**

3.2.13. Experience by CoTerminus Status and Year

	Yes			No		
Year	**Expected**	**Actual**	**Index**	**Expected**	**Actual**	**Index**
1980	16	18	87%	3	8	44%
1981	34	40	84%	46	52	88%
1982	90	94	96%	136	167	81%
1983	245	295	83%	270	356	76%
1984	323	373	87%	306	342	89%
1985	427	462	92%	336	364	92%
1986	780	740	105%	422	406	104%
1987	1510	1772	85%	580	609	95%
1988	1599	1976	81%	543	604	90%
1989	1814	2039	89%	538	474	114%
1990	2367	2236	106%	894	445	201%
1991	3115	2882	108%	1780	1743	102%
1992	3199	3144	102%	1950	1938	101%
1993	3940	3627	109%	2507	2109	119%
1994	4748	4318	110%	2212	2051	108%
1995	5379	3948	136%	2359	1892	125%
1996	5247	3859	136%	2919	1799	162%
1997	5083	4058	125%	2407	1804	133%
1998	3832	3305	116%	1645	1525	108%
Total	**43748**	**39186**	**112%**	**21853**	**18688**	**117%**

3.2.14. Experience by Benefit Period and Year

	2 Years			5 Years			Expiry		
Year	**Expected**	**Actual**	**Index**	**Expected**	**Actual**	**Index**	**Expected**	**Actual**	**Index**
1980	0	1	45%	0	1	6%	18	24	77%
1981	25	27	94%	3	8	37%	45	55	81%
1982	66	99	67%	18	9	204%	120	150	80%
1983	118	190	62%	63	53	119%	283	380	74%
1984	157	181	87%	58	79	73%	353	433	82%
1985	238	285	84%	95	82	116%	361	422	86%
1986	539	538	100%	89	121	74%	471	446	106%
1987	1090	1326	82%	171	227	75%	692	720	96%
1988	1169	1551	75%	181	235	77%	675	690	98%
1989	1335	1564	85%	203	217	94%	684	670	102%
1990	1725	1595	108%	339	267	127%	990	696	142%
1991	2553	2690	95%	581	571	102%	1424	1102	129%
1992	2384	2566	93%	687	721	95%	1660	1475	113%
1993	2789	2758	101%	1152	1056	109%	2019	1626	124%
1994	2832	2889	98%	1311	1266	104%	2264	1819	124%
1995	2776	2283	122%	1457	1078	135%	2562	1694	151%
1996	2944	2200	134%	1521	978	155%	2380	1574	151%
1997	2547	2119	120%	1291	1022	126%	2198	1385	159%
1998	1677	1677	100%	722	684	106%	2082	1420	147%
Total	**26965**	**26539**	**102%**	**9942**	**8675**	**115%**	**21281**	**16781**	**127%**

3.2.15. Experience by Benefit Proportion and Year

	Full Benefit			Partial Benefit		
Year	**Expected**	**Actual**	**Index**	**Expected**	**Actual**	**Index**
1980	19	26	74%	8	4	194%
1981	79	92	86%	26	13	196%
1982	226	261	87%	45	30	151%
1983	515	651	79%	115	73	158%
1984	629	715	88%	212	104	204%
1985	763	826	92%	307	181	170%
1986	1202	1146	105%	446	227	197%
1987	2090	2381	88%	658	448	147%
1988	2142	2580	83%	773	478	162%
1989	2352	2513	94%	896	557	161%
1990	3261	2681	122%	1133	539	210%
1991	4895	4625	106%	1135	524	217%
1992	5148	5082	101%	1031	428	241%
1993	6447	5736	112%	1507	851	177%
1994	6960	6369	109%	1423	742	192%
1995	7737	5840	132%	1671	722	231%
1996	8166	5658	144%	1598	534	299%
1997	7490	5862	128%	1428	663	215%
1998	5477	4830	113%	648	555	117%
Total	**65600**	**57874**	**113%**	**15060**	**7673**	**196%**

3.2.16. Experience by Age at Claim

Age At Claim	Expected	Actual	Index
22	1075	1144	94%
27	5485	5678	97%
32	9466	9134	104%
37	12345	10994	112%
42	12743	11131	114%
47	11233	9314	121%
52	7362	6115	120%
57	4110	3145	131%
62	1734	1186	146%
67	47	33	144%

3.2.17. Experience by Duration

Duration (Months)	Expected	Actual	Index
0	3554	3431	104%
1	17171	15581	110%
2	12891	16796	77%
3	6983	7644	91%
4	4501	3369	134%
5	3106	2159	144%
6	2970	1586	187%
7	2162	1075	201%
8	1469	811	181%
9	1038	664	156%
10	824	544	151%
11	688	464	148%
12	941	430	219%
13	775	334	232%
14	645	289	223%
15	520	249	209%
16	401	188	213%
17	334	179	187%
18	509	167	305%
19	409	152	269%
20	347	121	287%
21	294	98	300%
22	262	109	240%
23	229	146	157%
24-35	2023	873	232%
36-47	279	203	137%
48-59	127	106	120%
60-95	122	98	124%
96-119	27	8	335%

3.2.18. Experience by Year of Policy Commencement

Year of Entry	Expected	Actual	Index
1970	34	24	142%
1971	35	45	77%
1972	40	34	119%
1973	92	50	184%
1974	114	118	96%
1975	169	106	159%
1976	213	202	105%
1977	236	218	108%
1978	339	284	119%
1979	622	570	109%
1980	1228	1115	110%
1981	1587	1501	106%
1982	2218	1989	111%
1983	2011	1724	117%
1984	2270	2147	106%
1985	3496	3368	104%
1986	4049	4028	101%
1987	3590	3581	100%
1988	3909	3609	108%
1989	4765	4121	116%
1990	5236	4425	118%
1991	5600	4781	117%
1992	5722	4611	124%
1993	5692	4653	122%
1994	4954	3989	124%
1995	3795	3113	122%
1996	2201	2113	104%
1997	1107	1068	104%
1998	273	280	97%
Total	**65597**	**57867**	**113%**

3.2.19. Experience by Benefit Size

Benefit $/ Month	Expected	Actual	Index
500	6627	6117	108%
1500	27482	26447	104%
2500	20600	17271	119%
3500	5424	4159	130%
4500	2138	1594	134%
5500	1088	716	152%
6500	728	429	170%
8500	691	400	173%
12500	289	209	138%
17500	81	38	213%
25000	452	494	91%

APPENDIX 3.3 - GLM

Poisson model

Response: termrate

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev
NULL			265483	213643.9
ageclaim	1	3063.29	265482	210580.6
claimcauseW	1	4007.16	265481	206573.4
claimcauseX	1	809.86	265480	205763.6
claimcauseY	1	3349.97	265479	202413.6
duration	1	42495.61	265478	159918.0
sqrt(duration)	1	64.70	265477	159853.3
male	1	274.89	265476	159578.4
occupationA	1	235.89	265475	159342.5
occupationB	1	21.89	265474	159320.6
occupationC	1	5.86	265473	159314.8
smoker	1	0.66	265472	159314.1
deferment	1	728.86	265471	158585.2
benefitrate	1	28.57	265470	158556.7
APeriod	1	206.85	265469	158349.8
ageclaim:duration	1	22.76	265468	158327.1
ageclaim:sqrt(duration)	1	36.00	265467	158291.1

Call: glm(formula = termrate ~ ageclaim + claimcauseW + claimcauseX + claimcauseY + duration + sqrt(duration) + ageclaim + ageclaim * duration + ageclaim * sqrt(duration) + male + occupationA + occupationB + occupationC + smoker + deferment + benefitrate + APeriod, family = poisson(link = log), weights = COpen)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.938976	-0.5558429	-0.3008938	-0.06589518	7.022873

Coefficients:

	Value	Std. Error	t value
(Intercept)	3.23e+001	2.29e+000	14.11
ageclaim	-8.08e-003	1.08e-003	-7.49
claimcauseW	3.71e-001	1.08e-002	34.18
claimcauseX	-3.29e-001	1.04e-002	-31.72
claimcauseY	-5.61e-001	1.55e-002	-36.12
duration	-1.86e-001	1.03e-002	-18.06
sqrt(duration)	2.58e-001	4.51e-002	5.73
male	1.41e-001	1.36e-002	10.32
occupationA	-1.35e-001	1.29e-002	-10.51
occupationB	-8.56e-002	1.50e-002	-5.70
occupationC	-4.24e-003	9.99e-003	-0.42
smoker	-6.05e-002	8.75e-003	-6.92
deferment	2.38e-001	9.18e-003	25.96
benefitrate	-7.60e-006	1.86e-006	-4.09
APeriod	-1.67e-002	1.15e-003	-14.51
ageclaim:duration	1.82e-003	2.41e-004	7.55
ageclaim:sqrt(duration)	1.07e-003	-6.09	-6.52e-003

(Dispersion Parameter for Poisson family taken to be 1)

Null Deviance: 213643.9 on 265483 degrees of freedom

Residual Deviance: 158291.1 on 265467 degrees of freedom

Number of Fisher Scoring Iterations: 6

Appendix 5.1 -Selected S-Plus Functions – Chapter Five

```
> out2
function(a) {
        for(i in 3:(length(coxcrv$time)-1)) {
                A<<-max(coxcrv1$surv[coxcrv1$surv<coxcrv$surv[i]])
                B<<-min(coxcrv1$surv[coxcrv1$surv>coxcrv$surv[i]])
                C1<<-coxcrv1$time[coxcrv1$surv==A]
                D1<<-coxcrv1$time[coxcrv1$surv==B]
                timeout[i]<<-log(D1)*(coxcrv$surv[i]-A)/(B-A)+log(C1)*(B-coxcrv$surv[i])/(B-A)
                derivcvout[i]<<-timeout[i]-log(coxcrv$time[i])
                E1=length(coxcrv$surv[coxcrv$surv>0.8])
                F1=length(coxcrv$surv[coxcrv$surv>0.6])-E1
                G1=length(coxcrv$surv[coxcrv$surv>0.4])-(E1+F1)
                H1=length(coxcrv$surv[coxcrv$surv>0.2])-(E1+F1+G1)
                I1=length(coxcrv$surv[coxcrv$surv>0])-(E1+F1+G1+H1)
                E1mean<<-mean(derivcvout[1:E1])
                F1mean<<-mean(derivcvout[(E1+1):(E1+F1)])
                G1mean<<-mean(derivcvout[(E1+F1+1):(E1+F1+G1)])
                H1mean<<-mean(derivcvout[(E1+F1+G1+1):(E1+F1+G1+H1)])
                I1mean<<-mean(derivcvout[(E1+F1+G1+H1+1):(E1+F1+G1+H1+I1)])
        }}
```

```
> out3
function(a) {
        for(j in 1:84) {
                durn3a<<-durn3[(100*(j-1)+1):(100*j+300)]
                terminatea <<- terminate[(100*(j-1)+1):(100*j+300)]
                agea <<- age[(100*(j-1)+1):(100*j+300)]
                occupnewa<<-occupnew[(100*(j-1)+1):(100*j+300)]
                defpdnewa<<-defpdnew[(100*(j-1)+1):(100*j+300)]
                tempcox<<-
                coxph(Surv(durn3a,terminatea)~agea+occupnewa+defpdnewa,data=termrates2)
                tempcrq<<-
                crq(Surv(log(durn3a),terminatea)~agea+occupnewa+defpdnewa,data=termrates2)
                coxcrv<<-summary(survfit(tempcox,newdata=temp))
                coxcrv1<<-summary(survfit(tempcox,newdata=temp1))
                out2(5)
                E1means[j]<<-E1mean
                F1means[j]<<-F1mean
                G1means[j]<<-G1mean
                H1means[j]<<-H1mean
                I1means[j]<<-I1mean
                M1<<-length(tempcrq[tempcrq$sol[1,]<0.2])
                N1<<-length(tempcrq[tempcrq$sol[1,]<0.4])-M1
                O1<<-length(tempcrq[tempcrq$sol[1,]<0.6])-(M1+N1)
                P1<<-length(tempcrq[tempcrq$sol[1,]<0.8])-(M1+N1+O1)
                Q1<<-length(tempcrq[tempcrq$sol[1,]<1])-(M1+N1+O1+P1)
                M1mean<<-mean(tempcrq$sol[3,(1:M1)])
                N1mean<<-mean(tempcrq$sol[3,((M1+1):(M1+N1))])
```

```
        O1mean<<-mean(tempcrq$sol[3,((M1+N1+1):(M1+N1+O1))])
        P1mean<<-mean(tempcrq$sol[3,((M1+N1+O1+1):(M1+N1+O1+P1))])
        ifelse(M1+N1+O1+P1+1<=length(tempcrq$sol[3,]),Q1mean<<-
        mean(tempcrq$sol[3,((M1+N1+O1+P1+1):(M1+N1+O1+P1+Q1))]),Q1mean<<-0)
        M1means[j]<<-M1mean
        N1means[j]<<-N1mean
        O1means[j]<<-O1mean
        P1means[j]<<-P1mean
        Q1means[j]<<-Q1mean
        R1means[j]<<-E1means[j]-M1means[j]
        S1means[j]<<-F1means[j]-N1means[j]
        T1means[j]<<-G1means[j]-O1means[j]
        U1means[j]<<-H1means[j]-P1means[j]
        V1means[j]<<-I1means[j]-Q1means[j]
}}
```

Appendix 6.1 Selected S-Plus Functions – Chapter Six

> major

```
function(ra, rb, rc, rd, re, incrate, intrate, benefit)
{
	trp1(ra, incrate)
	trp2(rb, incrate)
	trp3(rc, incrate)
	trp4(rd, incrate)
	recrate5 <<- exp(sum(glmcoef * re))
	node123(p1, p2, p11, p12, p21, p22, p111, p112, p121, p122, p211, p212, p221, p222)
	node4(p1111, p1112, p1121, p1122, p1211, p1212, p1221, p1222, p2111, p2112, p2121,
p2122, p2211, p2212, p2221, p2222)
	probA(A1, C1, D1)
	probB(A1, C1, D1)
	probC(A2, C2, D2)
	probD(A1, C1, D1, A2, C2, D2)
	probE(A2, C2, D2, A5, C5, D5)
	probF(A1, C1, D1, A2, C2, D2, A5, C5, D5)
	probG(A1, C1, D1, A2, C2, D2, A5, C5, D5, A7, C7, D7)
	probH(A2, C2, D2, A5, C5, D5, A7, C7, D7)
	probI(A5, C5, D5, A7, C7, D7)
	probJ(A1, C1, D1, A5, C5, D5, A7, C7, D7)
	probK(A1, C1, D1, A5, C5, D5, A7, C7, D7)
	Ecalc(N42, N44, N46, N48, N410, N412, N414, N416)
	probL(E90, E210, E270, E330)
	probM(E90, E210, E270, E330)
	probN(E90, E210, E270, E330)
	probO(E90, E210, E270, E330)
	probP(5)
	premium(intrate, benefit)
	prem
}
```

> trp1

```
function(cov, incrate)
{
	recrate1 <<- exp(sum(glmcoef * cov))
	M1 <<- matrix(c( - (incrate + 0.000568), recrate1, 0, incrate, - (recrate1 + 0.000568), 0,
0.000568, 0.000568, 0), nrow = 3)
	A1 <<- eigen(M1)$vectors
	C1 <<- solve(A1)
	D1 <<- eigen(M1)$values
	diag1 <<- diag(c(exp((D1[1] * 60)/365), exp((D1[2] * 60)/365), exp((D1[3] * 60)/365)))
	temp1 <<- A1 %*% diag1 %*% C1
	p1 <<- temp1[1, 1]
	p2 <<- temp1[1, 2]
}
```

> probA

```
function(a, z, d)
{
	for(i in 1:60) {
		diagtemp1A <<- diag(c(exp((d[1] * i)/365), exp((d[2] * i)/365), exp((d[3] * i)/365)))
		temp1A <<- a %*% diagtemp1A %*% z
		prob12[i] <<- temp1A[1, 2]
	}
}
```

> node123

```
function(a, b, c, d, e, f, g, h, i, j, k, l, m, n)
{
	N11 <<- a
	N12 <<- b
	N21 <<- a * c
	N22 <<- a * d
	N23 <<- b * e
	N24 <<- b * f
	N31 <<- a * c * g
	N32 <<- a * c * h
	N33 <<- a * d * i
	N34 <<- a * d * j
	N35 <<- b * e * k
	N36 <<- b * e * l
	N37 <<- b * f * m
	N38 <<- b * f * n
}
```

> premium

```
function(intrate, benefit)
{
	for(i in 1:1825) {
		x[i] <<- ((benefit/365) * prob12[i])/((1 + intrate)^((intrate * i)/365))
	}
	prem <<- sum(x)
}
>
```

REFERENCES

Andersen, P.K. and Gill, R.D. (1982), "Cox's Regression Model Counting Process: a Large Sample Study," *Annals of Statistics*, 10, 1100 -1120.

Arjas, E. (1988) A graphical method for assessing goodness of fit in Cox's proportional hazards model, *Journal of the American Statistical Association,* Vol.83, pp. 204-212.

Australian Industry Table for Disability (1995). Produced by the *Graduation Sub-Committee of the Institute of Actuaries of Australia Disability Committee.*

Australian Prudential Regulation Authority, (2001) General Insurance Reform Act.

Australian Prudential Regulation Authority, (2001) General Insurance Prudential Standards and Guidance Notes, Liability Valuation for General Insurers.

Australian Prudential Regulation Authority, (2001) General Insurance Prudential Standards and Guidance Notes, Actuarial Opinions and Reports on General Insurance Liabilities

Australian Prudential Regulation Authority, (1973) Insurance Act.

Bayliss, P. (1991) CMIR12 Contribution See Continuous Mortality Investigation Report 12.

Benjamin, B and Pollard, J.H. (1993) The analysis of mortality and other actuarial statistics. Institute and Faculty of Actuaries, UK.

Booth, P.; Chadburn, R.; Cooper, D.; Haberman, S.; James, D. (2000) Modern Actuarial Theory and Practice, *Chapman and Hall, CRC*, London.

Brockman, M. J., and Wright, T. S. (1992). Statistical motor rating: making effective use of your data. *Journal of the Institute of Actuaries* 119, pp. 457-543.

Continuous Mortality Investigation Reports No 2, 4, 9,12,18 and 20, (1976 to 2001), CMIB, Institute and Faculty of Actuaries, UK.

Cox, D. (1972) Regression models and life tables (with discussion) *Journal of the Royal Statistical Society, B*. Vol.34, pp.187-220.

Cox, D. and Miller, H. (1965) The Theory of Stochastic Processes, *Spottiswoode, Ballantyine and Company*, London.

Efron, B. and Tibshirani, R. (1993) An Introduction to the Bootstrap, *Chapman and Hall, CRC*, New York.

Forfar, D. O.; McCutcheon and Wilkie, J. J. (1988) On graduation by mathematical formula, *Journal of the Institute of Actuaries*, Vol 115, pp.1-149.

Galton, F. (1889) Natural Inhertiance, *Macmillan*, London.

Goutis, C. and Casella, G. (1999) Explaining the Saddlepoint Approximation, *The American Statistician*, Vol. 53, pp. 216-232.

Grambsch, P.M. and Therneau, T.M. (1994) Proportional hazard tests and diagnostics based on weighted residuals, *Biometrika*, Vol.81, pp. 515-526.

Gregorius, F.K. (1993) Disability insurance in the Netherlands. *Insurance, Mathematics and Economics,* Vol. 13, pp. 101-116.

Haberman, S., and Renshaw, A. E. (1996). Generalised linear models and actuarial science. *The Statistician* 45, pp. 41-65.

Harrell, F.E. (1986), "The PHGLM Procedure," *SUGI Supplemental Library Guide, Version 5 Edition*, Cary, NC: SAS Institute Inc.

Hessalager, O. (1994) A Markov model for loss reserving. *Astin Bulletin*, Vol. 24, pp. 183-193.

Huzurbazar, Aparna V. (2004). "Multistate Models, Flowgraph Models and Semi-Markov Processes", *Communications in Statistics: Theory and Methods*, 33(3), 457-474.

Yau, C. Lillian, and Huzurbazar, Aparna V. (2002), ``Analysis of Censored and Incomplete Survival Data using Flowgraph Models", *Statistics in Medicine* , 21, 3727-3743.

The Institute of Actuaries of Australia Report of the Disability Committee (1997). *Transactions of the Institute of Actuaries of Australia 489-576.*

Jones, B.L. (1993) Modelling multi-state processes using a Markov assumption, *Actuarial Research Clearing House*, pp. 239-248.

Jones, B.L. (1994) Actuarial calculations using a Markov model, *Transactions of the Society of Actuaries,* Vol.46, pp.227-250.

Jones, B.L. (1996) Transient results for a high demand CCRC model, *Scandinavian Actuarial Journal,* pp.165-182.

Jones, B.L. (1995) A stochastic population model for high demand CCRCs, *Insurance Mathematics and Economics,* Vol. 16, pp.69-77.

Jones, B.L. (1997) Stochastic models for continuing care retirement communities, *North American Actuarial Journal*, Vol.1 (1), pp.50-73.

Kaplan, E.L. and Meier, P. (1958) Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Assocation*, Vol. 53, No. 282, pp. 457-481.

Kleiber, C.and Kotz, S. (2003) Statistical Size Distributions in Economics and Actuarial Sciences, *Wiley*, New York.

Koenker, R. and Bassett, G. (1978) Regression quantiles, *Econometrica*, Vol.46, pp. 33-50.

Kotz, S. and Nadarajah, S. (2000) Extreme Value Distributions: theory and applications, *Imperial College Press*, London.

Lee, E.T. and Wang, J.W. (2003) Statistical Methods for Survival Data Analysis, *Wiley*, New Jersey.

Louzada-Neto, F. (1999) Modelling Lifetime Data: A Graphical Approach, *Applied Stochastic Models in Business and Industry*, Vol.15, pp. 123-129.

Maindonald, J. and Braun, J. (2003) Data Analysis and Graphics Using R, *Cambridge University Press*, Cambridge.

MacKay, G. (1993) Permanent health insurance: Overviews and market conditions in the UK, *Insurance, Mathematics and Economics*, Vol.13, pp-123-130.

Maller, R. and Zhou, X. (1995) Survival Analysis with Long Term Survivors, *Wiley*, New York.

Mattsson, P. (1977) Some reflections on different disability models, *Scandinavian Actuarial Journal*, pp.110-118.

McCullagh P. and Nelder J. A. (1989) Generalised Linear Models, Second Edition, *Chapman and Hall/CRC.*

Miller, J.H. (1976) The underwriting and control of long-term disability insurance, *NRG Publications*, Amsterdam.

Miller, J.H. and Courant, S. (1974). A Mathematical Model of the Incidence of Disability. *Transactions of the Society of Actuaries 25*

Munich Re of Australia, Rice Kachor Research, (2002) Disability Analysis.

Neill, A. (1977) Life contingencies, *Heinemann*, London.

Nelder, J. A. and W. Wedderburn, R. W. M. (1972). Generalised linear models. *Journal of the Royal Statistical Society A 135*, pp. 370-384.

Norberg, R. (1991) Reserves in life and pension insurance. *Scandinavian Actuarial Journal*, pp. 3-24.

Norberg, R. and Hess, O. (1996) On probability distributions of present values in life insurance. *Insurance, Mathematics and Economics,* Vol.18, pp. 35-42.

Norberg, R. (1995) A time-continuous Markov chain interest model with applications to insurance, *Applied Stochastic Models and Data Analysis,* Vol.11, pp.245-256.

Peng, Y. and Dear, K.B. and Denham, J.W. A generalized F mixture model for cure rate estimation, *Statistics in Medicine*, Vol. 17(8), pp. 813-830.

Pitacco, E. (1995) Actuarial models for pricing disability benefits: towards a unifying approach, *Insurance Mathematics and Economics,* Vol. 16, pp.39-62.

Pitacco, E. (1999) Multistate Models for Long-Term Care Insurance and Related Indexing Problems, *Applied Stochastic Models in Business and Industry,* Vol.15, pp. 429-441.

Pitacco, E. and Haberman, S. (1999) Actuarial Models for Disability Insurance, *Chapman and Hall, CRC,* London.

Pitt, DGW and Service, DA. (2002) Disability Claims – Does Anyone Recover? *Australian Actuarial Journal*, pp. 615-659.

Portnoy, S. (2003) Censored regression quantiles, *Journal of the American Statistical Assocation*, pp.1001-1012.

Renshaw, A. E. (1991). Actuarial Graduation and Generalised Linear and Non-Linear Models. *Journal of the Institute of Actuaries* 118, II, pp. 295-312

Renshaw, A. E. and Haberman, S. (1995). On the graduation associated with a multiple state model for permanent health insurance. *Insurance, Mathematics and Economics, 17*, pp. 1-17.

Rolski, T. and Schmidli, H. and Schmidt, V.; Teugels, J. (1998) Stochastic Processes for Insurance and Finance, *Wiley*, New York.

Ross, S. (2000) Introduction to Probability Models (Seventh Edition), *Harcourt Academic Press*, San Diego.

Segerer, G. (1993) The actuarial treatment of disability risk in Germany, Austria and Switzerland. *Insurance, Mathematics and Economics*, Vol.13, pp131-140.

Srinivas, A. and Pitt, D.G.W. (2004) Premium and Reserve Calculations for Disability Insurance Using Classification Trees, *ANU Actuarial Studies Honours Thesis.*

Tableman, M. and Kim, J.S. and Portnoy, S. (2003) Survival Analysis Using S: Analysis of Time-to-Event Data, *Chapman and Hall, CRC,* London.

Taylor, G.C. (1971) Sickness: a stochastic process, *Journal of the Institute of Actuaries*, Vol.97, pp. 69-83.

Taylor, G.C. and Campbell, M. (2002) Statistical Case Estimation, University of Melbourne Centre for Actuarial Studies Research Paper Number 104.

Therneau, T.M. (2000) Modeling survival data: extending the Cox model. *Springer*, New York.

Tuomikoski, J. (1988) On modelling disability by mixed exponentials, *Transactions of the 23rd International Congress of Actuaries.*

Waters, H.R. (1989) Some aspects of the modelling of permanent health insurance, *Journal of the Institute of Actuaries*, Vol. 116, pp.611-624.

Waters, H.R. (2004) Actuarial Modelling of Income Protection Insurance in the UK - The Work of the C.M.I.B. Presented at the ANU and IAAust Seminar, *Actuarial Models for Financing Disablement Benefits.*

Wolthius, H. (1994) Life Insurance Mathematics (The Markovian model), *CAIRE Education Series*, No.2, Bruxelles.