



Institute of Actuaries of Australia

Statistical Case Estimation Modelling

- An Overview of the NSW WorkCover Model

Presented by Richard Brookes and Mitchell Prevett

Presented to the Institute of Actuaries of Australia
Accident Compensation Seminar 28 November to 1 December 2004.

This paper has been prepared for the Institute of Actuaries of Australia's (IAAust) Accident Compensation Seminar, 2004. The IAAust Council wishes it to be understood that opinions put forward herein are not necessarily those of the IAAust and the Council is not responsible for those opinions.

© 2004 Institute of Actuaries of Australia

The Institute of Actuaries of Australia

Level 7 Challis House 4 Martin Place

Sydney NSW Australia 2000

Telephone: +61 2 9233 3466 Facsimile: +61 2 9233 3446

Email: insact@actuaries.asn.au Website: www.actuaries.asn.au

Table of Contents

1	INTRODUCTION AND BACKGROUND.....	4
1.1	INTRODUCTION	4
1.2	BACKGROUND.....	4
1.2.1	What is an SCE?.....	4
1.2.2	How does an SCE model relate to standard actuarial techniques?.....	5
2	DATA, MODEL STRUCTURE AND TARGET VARIABLES.....	6
2.1	AVAILABLE DATA	6
2.2	MODEL STRUCTURE	6
2.3	TIME PERIOD AND PROJECTION	7
2.3.1	Short modelling periods	7
2.3.2	Long modelling periods	8
2.4	TARGETS	8
3	TESTING AND MODEL VALIDATION.....	10
3.1	DATA PARTITIONING.....	10
3.2	MODEL EVALUATION.....	10
3.2.1	Actual versus expected.....	10
3.2.2	Gains charts.....	11
3.2.3	Example.....	11
3.2.4	Other evaluation statistics	12
4	TECHNIQUES IN MODELLING	13
4.1	CLASSIFICATION AND REGRESSION TREES (CART®)	13
4.1.1	Description of CART	13
4.1.2	Example of CART.....	14
4.1.3	Potential drawbacks with CART	15
4.2	MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARST TM)	16
4.2.1	Description of MARS	16
4.2.2	Example of MARS spline function	16
4.2.3	Potential drawbacks with MARS	17
4.3	HYBRID CART, MARS AND GENERALISED LINEAR MODELS (GLMs)	17
5	THE WEEKLY AND MEDICAL MODELS.....	19
5.1	THREE YEAR PAYMENTS MODELS	19

5.1.1	Weekly CART Model	19
5.1.2	Weekly MARS Model.....	23
5.1.3	Medical CART Model.....	24
5.1.4	Medical MARS Model	28
5.1.5	Medical GLM.....	28
5.2	PAYMENT PATTERNS	29
5.2.1	Weekly Patterns	29
5.2.2	Medical Patterns.....	31
5.2.3	Issues with the pattern fitting approach.....	33
5.3	TAIL HAZARD FITTING	33
5.4	COMBINING THE THREE YEAR MODEL, THE PAYMENT PATTERN AND THE TAIL EXTRAPOLATION.....	34
6	PERFORMANCE OF THE NSW WORKCOVER MODEL	35
6.1	THREE YEAR PREDICTIONS	35
6.2	PATTERN PREDICTIONS.....	37
6.3	TAIL PREDICTIONS.....	37
6.4	RECENT PREDICTIONS	39
6.5	SCES VERSUS MANUAL CASE ESTIMATES	40
6.5.1	Predictiveness.....	40
6.5.2	Development Over Time of High Value Estimates.....	42
7	WHAT NEXT?.....	45
7.1	APPLICATIONS.....	45
7.2	REFINEMENTS TO THE MODEL	47
7.2.1	Data	47
8	BIBLIOGRAPHY	48
A.	APPENDIX.....	49
A.1.1.	Short Duration Claims (Less then 3 Months Developed)	49
A.1.2.	Medium Duration Claims (4 to 12 Months Developed)	50
A.1.3.	Long Duration Claims (Greater then 12 Months Developed).....	52
A.1.4.	Active Claims	53
A.1.5.	Inactive Claims	54

1 Introduction and Background

1.1 Introduction

In this paper we will describe the recent approach taken for developing the NSW WorkCover Statistical Case Estimate Model. The model has not yet been rolled out into Scheme operations so we do not discuss its uses in great detail.

This paper is intended as a case study illustrating the approach we have taken for this particular model. We tried several different modelling structures and methods before following the approach documented here. In our opinion, the features and limitations of the dataset were a significant factor in determining the final approach so this paper should not be taken as a technical exposition on the general approach to constructing such models.

Although building a Statistical Case Estimate Model is a relatively complex technical exercise we have endeavoured to keep the discussion of statistical issues to a practical level. Interested readers are recommended to refer to the papers in the bibliography for a technical discussion of the methods we have used.

1.2 Background

1.2.1 What is an SCE?

Statistical case estimates (SCE's) are individual estimates of the future claim related costs arising from existing, open claims. A statistical model produces the estimates on each individual claim, based on its risk characteristics such as:

- Claimant characteristics
 - ✓ Age, gender, occupation, marital and dependant status, wage rate etc
- Employer characteristics
 - ✓ Industry, wages, location, etc
- Claim status
 - ✓ Claim is open/closed/reopened/disputed, work status, etc
- Claim characteristics
 - ✓ Injury nature, location, etc
- Claim history
 - ✓ Payments and rates of payment, time lost, etc

1.2.2 How does an SCE model relate to standard actuarial techniques?

Standard actuarial modelling techniques concentrate on modelling the overall outstanding claim liabilities for a portfolio of claims in aggregate. Whilst there is generally some effort to subdivide the portfolio into more homogenous groups for modelling purposes, the approach can become unwieldy with a large number of subdivisions. For this reason, standard techniques cannot account for individual claim characteristics or adequately allocate total liabilities down to the individual claim level.

On the other hand, an SCE model is unlikely to give a good estimate of the overall outstanding claim liabilities for a portfolio. There is a variety of reasons for this, two being:

- There is no allowance for incurred but not reported (IBNR) claims
- Overall trends, such as superimposed inflation have not been appropriately allowed for.

Draft

2 Data, Model Structure and Target Variables

2.1 Available data

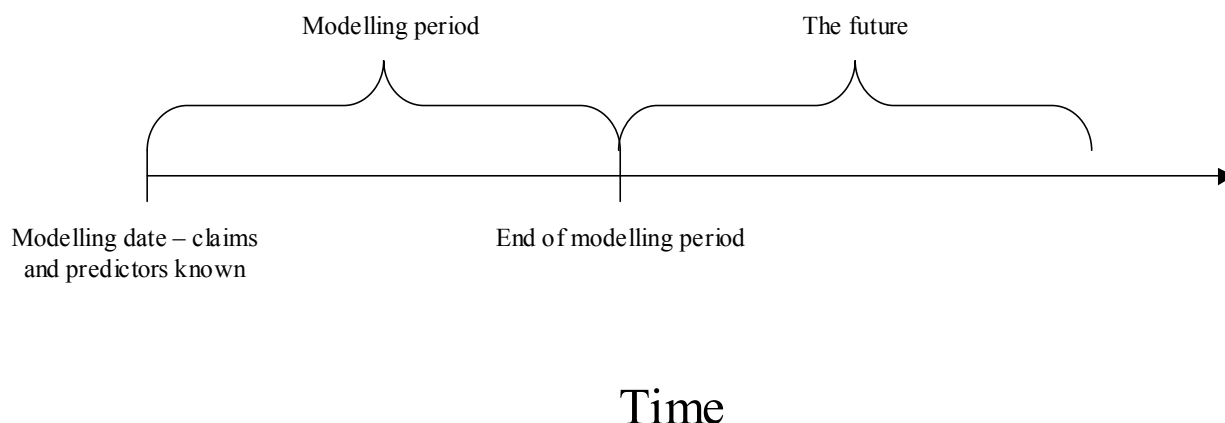
In general terms, an SCE is about taking all the information known regarding a claim at a point in time (the “valuation date”) and using it to project the future payments applicable to the claim, over its future lifetime. The information known about a claim in the NSW WorkCover database can be summarised under the headings in paragraph 1.2.1. We will refer to each item of data that is known and recorded on the database at the valuation date as a “predictor”. Significant effort was devoted to removing predictors where the data was clearly not robust and creating new ones where we felt that the transformation or combination of raw predictors might yield a better result than the raw predictor itself. It is not appropriate to list all the predictors after this process; suffice it to say that there were more than 200 of them!

One particular problem was whether to use case estimates at the valuation date as predictors. In NSW, the insurers currently set conventional case estimates for each claim in accordance with guidelines set out by NSW WorkCover. To use these as predictors seems circular, especially if the SCE model is going to be used to replace the conventional case estimates. However, the case estimates clearly contain information which is known about the claim and which would be useful to use for prediction. We settled for using case estimate binaries; for instance a variable which was set to Yes if there was case estimate for legal payments for a particular claim and No otherwise. Our reasoning was that, in the case of legal expenses for instance, the claim manager would know whether or not there was a lawyer involved in a particular claim with outstanding expenses. Even if the SCE was used in the future, the claim manager should still know this piece of information and the insurer should still collect it on their database.

2.2 Model structure

The model structure was based on all open claims at a valuation date in the past (the “modelling date”) and at which point we know the values of all the predictors. Over the subsequent modelling period we will track the actual costs for the claim and build statistical models that connect the predictors with these costs.

In diagrammatic terms, the situation is the following:



This structure is very similar to how the model will be used in practice.

2.3 Time Period and Projection

In our opinion, the most significant difficulty with this type of modelling is the incorporation of the time element:

- How to project payments over the remaining lifetime of a claim, which can be many years for serious claims;
- How to project any time-related trends into the future.

Since the main uses of the SCE model involve the allocation of the overall claim liability that is determined by standard actuarial methods, it is relative SCEs that are more important than absolute levels. Therefore, we made an early decision that the projection of time related trends into the future would be fairly crude. We aimed to determine the relative future cost of claims, based on the current environment, and maybe project forward some element of superimposed inflation for certain payment types.

In terms of the diagram above, the crucial decision to be made is:

- How long should the modelling period be?

2.3.1 Short modelling periods

One option is to choose a very short modelling period and chain the resulting model together in some way to get payments over a long period. At the extreme, one could build a daily model of incapacity. This is the approach used by Taylor and Campbell (2002) for their weekly compensation model. Less extreme is a typical annual actuarial model where the experience for claims is assumed to be similar for the same development year. The problem we found with a short modelling period is that it is very difficult to incorporate dynamic predictors such as payment history. Our investigative analysis led us to the conclusion that, for this dataset, the combination of weekly and medical payment histories is the best proxy for the severity and eventual outcome of a claim. We did try such models but found that, despite fitting good models over the (short) modelling period, the combination and projection process did not work well. The result was an inaccurate cost

projection over longer periods and poor differentiation between high and low cost claims.

2.3.2 *Long modelling periods*

Ideally, one would choose the longest possible modelling period – say twenty years. If the environment, claim profile and claim behaviour were stable, and one could build a statistical model which explained all the payment variation between claims then this would be the ideal SCE. Another way of saying this is that:

- A longer period will enable us to capture more of the ultimate claim cost in a single model; and
- The longer time period is more closely linked to the ultimate outcome of the claim.

However, one major disadvantage of a long period model is that a long period of reliable data and stable experience is required. The NSW Scheme has been subject of number of recent behavioural and legislative changes although we do not describe these here. In practice, we made quite extensive adjustments to the data with the intention of removing the effects of these changes.

Another disadvantage of a long modelling period is that, without a short period prediction, it is difficult to monitor actual versus expected outcomes and assess whether or not the model is still valid.

After taking all of these factors into consideration we decided to use a modelling period of three years and with a modelling date of 1 January 1999. We also fitted quarterly payment patterns for our modelled payments, to enable quarterly monitoring. This is described further in section 6.4..

2.4 Targets

Finally in this section we discuss the actual quantities to be modelled. In general terms there is a range of options. At one end of the range we could model a single quantity, total payments over the modelling period. A single model is simpler and often easier to interpret. It can also be less dependent on assumptions such as independence between payment types. However, it is also more difficult to monitor and adjust if it drifts out of calibration.

At the other end of the range, one could model a set of variables that build up to the total payments variable. For instance, one could model:

- Will a claimant be incapacitated?
- If so, then how many days compensation will he or she receive?
- What will be the paid rate per day?

The final model would combine all these sub-models. A set of sub-models is easier to monitor. It can also introduce some algebraic structure into the

problem that gives the statistical modelling techniques a better chance of finding robust predictive relationships.

For the NSW WorkCover SCE we tried a variety of options but ended up modelling cumulative payments over the three year modelling period for 13 payment types:

- Weekly Compensation
- Medical
- Rehabilitation
- Investigation
- Physiotherapy/Chiropractic
- Permanent Injury Lump Sums (Section 66)
- Pain and Suffering (Section 67)
- Section 66 and 67 Legal
- Miscellaneous Legal
- Death
- Other
- Recoveries
- Excess Recoveries

For some of these models, there were two sub-models; for instance did the claimant receive a Permanent Injury lump sum over the modelling period and, if so, how much (termed 2-stage modelling)?

For the remainder of this paper we will discuss in detail the weekly compensation and medical models since these are the most significant, comprising around 37% and 16% of total SCEs as at 30 June 2003.

3 Testing and Model Validation

3.1 Data partitioning

It is common practice in data mining and many statistical modelling exercises with large datasets, to randomly separate the data prior to modelling into a learning dataset and a testing dataset. The learning dataset is exclusively used for modelling and fitting purposes while the testing dataset is used to assess how well the model predicts on an independent dataset. This process is a safeguard against over-fitting and the evaluation against an independent test dataset is a better guide of how the model will fit to new data, going forward. In ideal circumstances one would also evaluate against a dataset from a different time period either from before the period used to fit the data or from afterwards. In this case, given our choice of modelling period, we would have needed another stable period of three years to evaluate our model. Such a period was not available although later in this paper we give the results of an evaluation for the year after the modelling period. For the NSW WorkCover model we have randomly split the dataset from the modelling period into 70% for learning and 30% for testing.

All of the models were built on the learning dataset, including the cross-validations used in some of the data-mining algorithms and all of our “in period” evaluations are based on the test dataset.

3.2 Model evaluation

In a project of this type, one can expect to build and compare lots of models. One will also be building models using a variety of different methods; for instance decision trees, neural nets, MARS models, regressions and GLMs. Therefore one needs an evaluation strategy that is independent of the modelling method.

3.2.1 *Actual versus expected*

The first evaluation method employed is a comparison of the actual and expected (predicted) values from the model on the test dataset. For a complete actual versus expected evaluation, one produces a graph or table for each important predictor that shows actual versus expected target values as the value of the predictor changes. A useful summary evaluation is to plot actual versus expected for values of the predicted target. The claims are ranked from lowest to highest, based on the expected values from the model and then divided into 10 to 100 equal size groups. The average actual and average predicted values are then compared for each group. For a well-fitting model the actual and expected means should match well across the entire range of the data. A better model can also be identified as one that predicts a

greater range of values (higher and lower prediction values) with no observable bias.

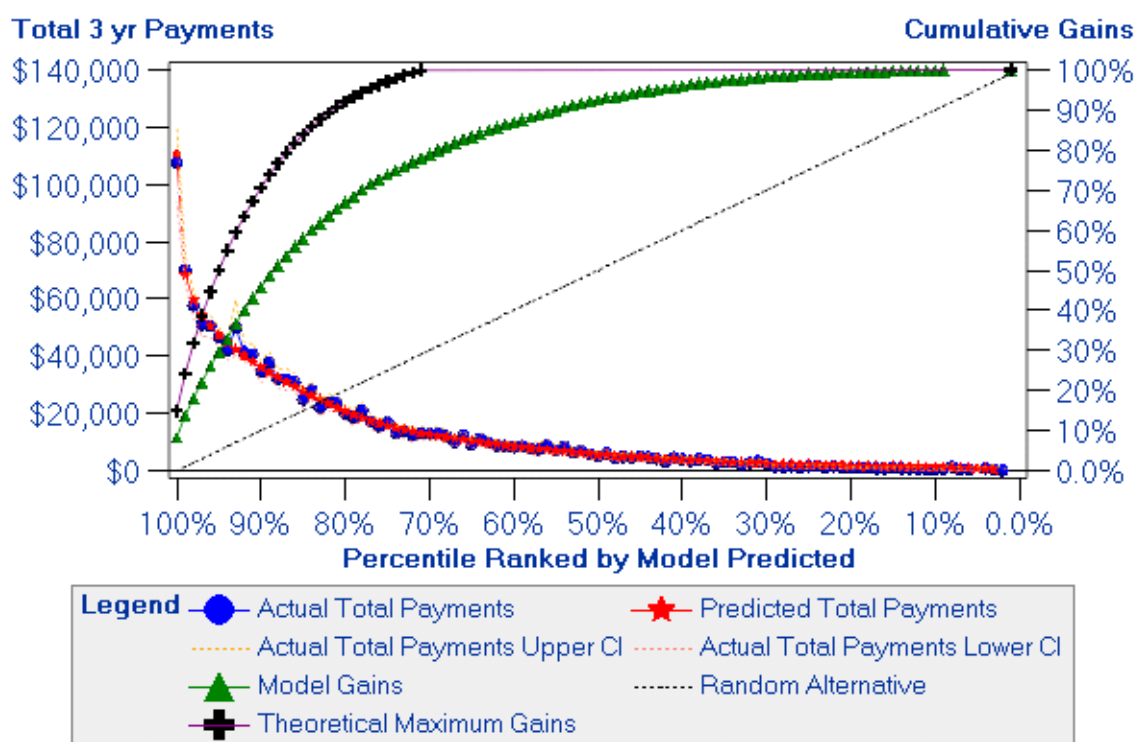
3.2.2 Gains charts

Another evaluation methodology we can employ involves calculating the percentage of the total cost captured by the predictions of the model. Firstly we can think of the baseline as a model with no information, in which ranking claims from highest to lowest results in a random ordering. For such a model the top 5% of predictions will capture only 5% of the total cost on average, the top 10% captures 10% of the cost, and so on. Alternatively, for any model with some degree of ranking, the total cost captured in the higher predictions will be higher than the percentage of observations and a better model can be identified as one that captures a significantly higher percentage of this cost.

3.2.3 Example

An example of the above evaluations is incorporated into the graph below.

Figure 3-1 Actual vs Expected and Gains Charts



The red and blue lines give the actual versus expected analysis. The percentile as ranked by the model predicted is presented on the horizontal axis. The red line contains 100 points identifying the mean prediction in each of the 100 percentiles while the blue line plots the mean actual value. The values of both are read off the left vertical axis.

The green line is read off the right vertical axis and shows the gains for the model e.g. the top decile (upper 10% of the predictions) captures around 46% of the total cost. The purple line demonstrates the theoretical best gains line that is attainable for this data. A perfect model would rank the data exactly from highest to lowest and hence this line plots the percentage of the total target cost captured in the upper percentiles of the data, ranked by the actual target.

3.2.4 *Other evaluation statistics*

Other model evaluation statistics are also helpful. We use the root average squared error (RASE) and R-square statistics on the testing dataset. The term RASE is used to distinguish it from the root mean squared error (RMSE) which is often adjusted to reflect the number of parameters used in the model. We have adopted the RASE over the RMSE because for some data mining models there is no agreed way to determine the number of parameters used in the model and the difference is insignificant when there is a large number of data observations. The natural interpretation of the RASE is that it represents the standard deviation of the raw residuals from the model and thus provides a good indication of the spread. Less spread in residuals indicates a better fitting model and hence a lower RASE is desirable. The R-square we employ is also not adjusted for the number of parameters in the model but with a large enough dataset, again the difference is insignificant. The R-square statistic has the natural interpretation that it gives the proportion of the response variable variation explained by the model. Both of these statistics are seriously affected by outliers and hence should not be considered in isolation from the other evaluations.

4 Techniques in Modelling

In this section we give very brief details of the less familiar modelling techniques we employed for the weekly and medical payment types. These are CART, MARS and a hybrid structure using CART, MARS and GLM (Generalised Linear Models) together. Interested readers are referred to the many more technical books and articles, some of which are given in the bibliography. We have not described GLMs since these are now part of the standard actuarial toolkit.

4.1 Classification and Regression Trees (CART®)

4.1.1 Description of CART

Salford Systems (the maker of CART and MARS) advertise that CART is a robust modelling tool that can be used to uncover important relationships in large datasets. These relationships can be used to develop accurate and reliable predictive models. The discovery process can include the identification of important predictors amongst possibly hundreds of potential predictors or the identification of complex but robust interactions between predictor variables.

The models are constructed through a process of binary recursive partitioning of the data. Each partition is determined using a splitting rule on the raw predictor variables which can take one of the following forms:

- If Age > 35 then split left, otherwise split right
- If Car = (sedan or hatch) then split left, otherwise split right

The potential splitting rules are generated through a process of brute force whereby every possible split (in most cases) is tested for each current partition (node) of the data. These splits are then ranked by the additional predictiveness they add to the model and the most predictive is chosen. After further partitioning the data for the chosen split, the process is repeated. Various methods are available for determining and ranking the quality of the splits.

CART employs a growing and pruning process to determine the optimal size tree. The dataset for modelling is randomly separated into learning and testing datasets (70%/30% is commonly used). The learning dataset is used to grow the tree to its maximal size, where no further splits are possible. CART then uses the testing dataset to prune back the maximal tree in order to minimize the model error on this data. There also exists a cross-validation option for determining the optimal tree size which is suitable for smaller datasets.

Some of the advertised strengths of CART are:

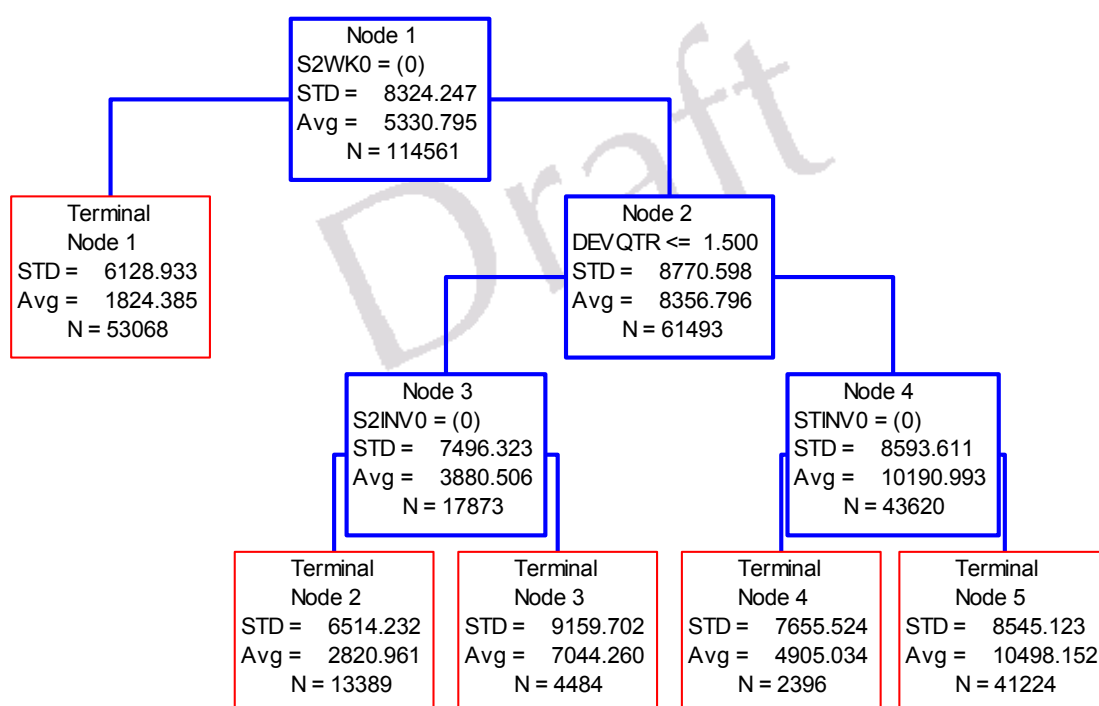
- Automatic variable selection amongst many predictors
- No need for transformation of predictors (splits are based on ranks)
- Very high level interactions are captured (each parent node is effectively an interaction on previous nodes)
- Resistant to outliers (outliers in the predictors will not result in outliers in the predictions)
- Resistant to missing values

For a more complete description of CART, readers are referred to Salford Systems [1].

4.1.2 Example of CART

An example CART tree is presented below.

Figure 4-1– Example CART Output



The upper section of a tree is presented above. Blue nodes are called parent nodes (or splitters) and red nodes are terminal nodes. At each parent node starting from the top, CART will determine the best splitting variable and split point for that variable, based on the explanatory power from this partition. After each split, all terminal nodes are assessed for their best partition and out of these competing splits the one with the most explanatory power is chosen.

Each parent node displays the splitting rule immediately following the node number, and the affirmative to this rule always leads to the left branch. Next are the target variable summary statistics for the sample at that node; standard

deviation, average and number of data points. Terminal nodes also include these summary statistics.

The example tree is the upper level of a tree with annual weekly payments as the target variable. Various predictors were given to CART, including injury nature and location, accident quarter and development quarter, age, gender, and also a range of variables defining active/inactive statuses by payment type.

S2WK0 is a variable defining the weekly payment status, taking a 1 when the claim has a positive weekly payment in the 3 months leading up to the 12 month period, and 0 otherwise. DEVQTR is development quarter. S2INV0 is the same as S2WK0 except based in the investigation payment type. STINV0 is also based on the investigation payment type taking a value of 1 when there is a positive case estimate at the modelling date and 0 otherwise.

The mean annual weekly payment for the whole population is \$5,331 and the standard deviation is \$8,324. A description of each splitting node is given below:

- Node 1. Claims that did not receive a weekly payment in the previous quarter go to the left. These, not surprisingly, have a much lower mean payment than the others.
- Node 2. Claims in development quarter zero or one go to the left. These have a lower mean payment than the others.
- Node 3. Claims that have not received an investigation payment go to the left. These have a low mean payment compared to the claims that go to the right; \$2,821 compared with \$7,044.
- Node 4. The claims that have no outstanding case estimate for future investigation payments go to the left. These have a low mean payment compared to the claims that go to the right; \$4,905 compared with \$10,498.

4.1.3 Potential drawbacks with CART

Despite CARTs many advantages, we have observed some potential drawbacks for the unwary:

- In some circumstances, a preference for selecting high level categorical predictors over other predictors even though the splits may test poorly. CART software incorporates a penalty for high level categorical predictors which partly counteracts this problem.
- Lower splits in any tree are heavily dependent on the early splits. This means that in some cases, a single different initial split could result in a significantly different tree.
- Finally, the criteria for ranking potential splits in a regression tree is based on least squares (although the least absolute deviation (LAD) method is also available in CART, the increase in run time for LAD generally renders it unsuitable for most large data modelling situations). Although the minimum terminal node size generally ensures that any individual outlier does not significantly affect the tree performance, least squares does mean the trees tend to

focus on the higher cost observations and as a result there is usually a low level of differentiation amongst small predictions.

4.2 Multivariate Adaptive Regression Splines (MARS™)

4.2.1 Description of MARS

Salford Systems state that the MARS technique builds regression models by fitting a series of optimal linear spline curves (termed basis functions) to each continuous predictor variable and optimally grouping each categorical variable. The technique employs a forward selection phase in order to select the most important predictor basis functions, followed by a backwards elimination phase to remove poor and over-fitting functions. Interactions between selected basis functions are tested and included in the model where appropriate during forward selection.

The derived basis functions for continuous variables are colloquially termed “Hockey Sticks” and take the following form:

$$BF_i = \max(X - k, 0)$$

or

$$BF_j = \max(k - X, 0)$$

where BF_k is the i^{th} selected forward hockey stick basis function in the model, X is the raw predictor variable upon which the basis function is derived, k is the optimal knot location selected, and BF_j is the j^{th} selected reverse hockey stick basis function in the model.

Optimal linear spline curves are constructed via a combination of forward and reverse hockey sticks.

The basis functions for categorical variables are simply indicator functions such as:

$$BF_i = \{1 \text{ if } X \text{ is in } (a, b, \dots), 0 \text{ otherwise}\}$$

The final model is a linear combination of basis functions.

For a more complete description of MARS readers are referred to Salford Systems [2].

4.2.2 Example of MARS spline function

As an example, a linear spline curve may take the following form:

$$BF_1 = \max(0, \text{WKLYC} - 1500)$$

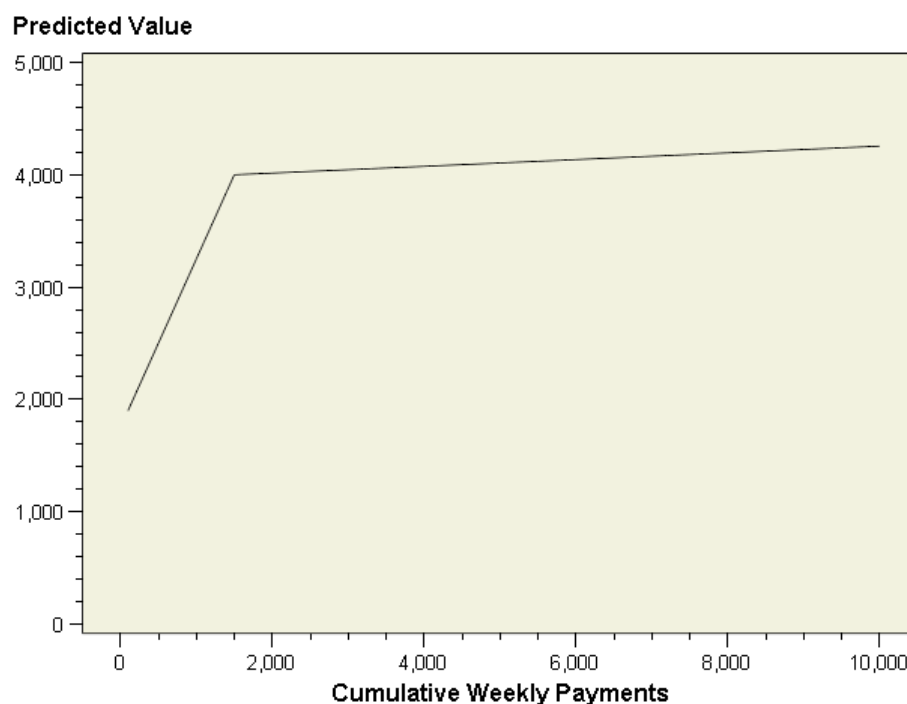
$$BF_2 = \max(0, 1500 - \text{WKLYC})$$

$$\text{Predicted} = 4000 + 0.030 * BF_1 - 1.5 * BF_2$$

Here WKLYC is the total cumulative weekly payment on a claim as at the modelling date. A single knot has been selected at \$1,500 and two basis functions are created on either side of this knot.

The dependent versus predicted relationship for the above example can be examined with the plot below. Here the slope is 1.5 from zero up to the knot and then 0.03 after the knot.

Figure 4-2 – MARS Example Dependent versus Predicted Plot



4.2.3 Potential drawbacks with MARS

MARS requires more care than CART. Some of the potential difficulties are:

- In contrast to CART, MARS is not resistant to outliers and has a limited ability to deal with missing values
- As for CART, differentiation for low values of the target can be poor
- In some circumstances, a well fitted and parameterised MARS model does not test well with an independent test dataset. We have not analysed this in any depth but, in our view, it is likely to be over-fitting, potentially due to the insufficient backwards elimination of poorly fitting basis functions.

4.3 Hybrid CART, MARS and Generalised Linear Models (GLMs)

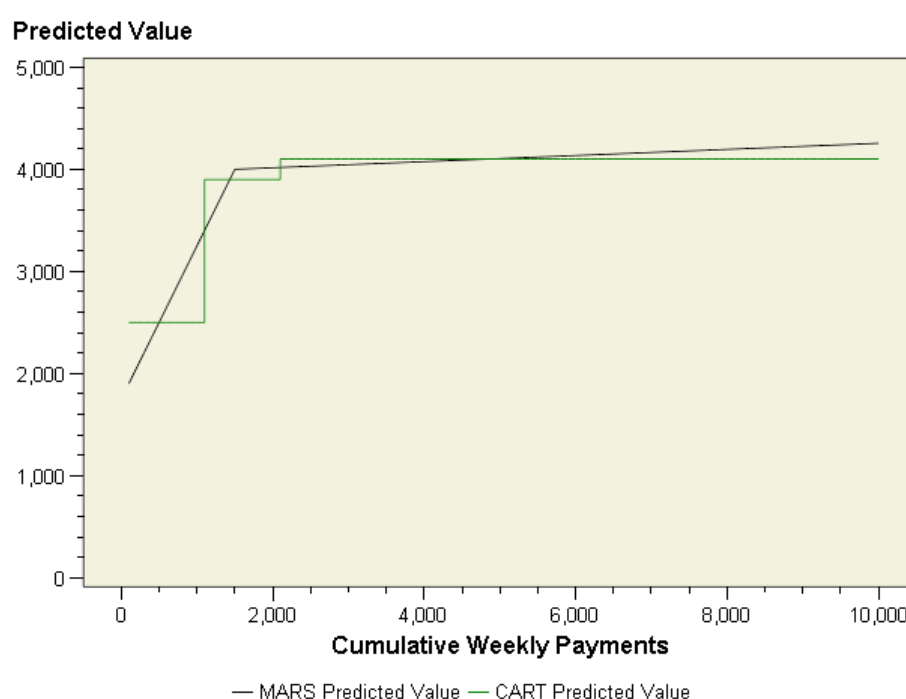
We have found that the CART and MARS algorithms complement each other in most modelling situations. Even for continuous targets we use CART as a first step and then generally use a CART/MARS hybrid model. The aim of using MARS after CART is to:

- Achieve smoother functional fits

- Identify weak continuous relationships that CART may not pick up.

The hybrid model consists of an initial CART tree followed by a MARS model to refine the tree. The CART tree takes all predictors available while the MARS model takes the CART terminal node number as a categorical predictor and some or all of the other predictors. The following plot demonstrates how a MARS function might improve the fit over a CART model.

Figure 4-3 – CART/MARS Comparison Dependent versus Predicted Plot



The observed over-fitting of MARS has sometimes led us to a refinement of this approach that seems to work well in practice.

This approach consists of the following steps:

- The terminal nodes are determined with the CART tree
- Then basis functions are created with the MARS model (incorporating the terminal nodes), with the MARS settings deliberately calibrated to avoid too much backwards elimination
- These basis functions are reduced and refined where appropriate using the GLM modelling process. This requires determination of the appropriate error distribution and link function for the GLM. Finally, using type 3 statistics, any poorly performing basis functions can be eliminated one after the other until all those remaining, are significant.

5 The Weekly and Medical Models

In this section we provide a commentary on the modelling approach for the weekly and medical payment types. This consists of three parts:

- The three year payment models
- The payment patterns for the three year payment models; and
- The fitting of a payment “tail” extending beyond three years.

5.1 Three Year Payments Models

5.1.1 Weekly CART Model

The weekly payment type includes all payments made in respect of sections 36, 37, 38 and 40 of the NSW Workers’ Compensation Act 1987. Payments under these sections are more commonly known as weekly payments for total incapacity (first 26 weeks), total incapacity (after 26 weeks), partial incapacity while unemployed, and partial incapacity while employed (make-up pay), respectively.

Summary statistics for the target variable are presented in the table below.

Table 5-1 – Summary Statistics for Weekly Target Payments

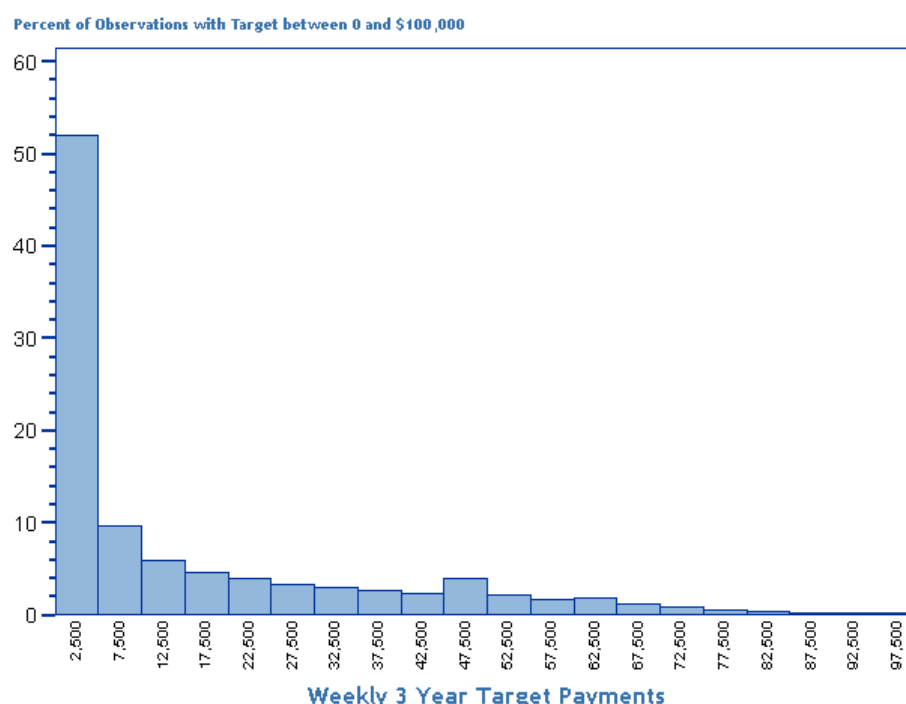
Number in learning	114,127
Mean	5,947
Standard Deviation	15,106
Skewness	3.21
Kurtosis	11.38
Quantiles	
100% Max	200,044
99%	69,489
95%	45,289
90%	22,739
75% Q3	1,469
50% Median	0
25% Q1	0
10%	0
5%	0
1%	0
0% Min	-31,776
Percentage Negative	0.30%
Percentage Equal to Zero	61.07%
Percentage Positive	38.63%

The learning sample consisted of 70% of the entire available dataset. The mean target value was almost \$6,000 and the coefficient of variation was 2.54. Only 39% of the claims in the dataset had a positive weekly payment over the 3 years. There were 354 observations with negative target payments

likely representing small reversals in previous payments (257 of these were for less than \$1,000). For simplification of the modelling process the target payments for these observations were set to zero.

The graph below presents a histogram of the weekly target between \$0 and \$100,000. The distribution generally appears reasonably right skewed and there is a concentration of payments around the \$47,500 region. This is likely to be a group on capped weekly payments for total incapacity over 26 weeks.

Figure 5-1 – Histogram of Weekly Target Payments between \$0 and \$100,000



In addition to the observations shown in Figure 5-1 – Histogram of Weekly Target Payments between \$0 and \$100,000 there are 99 observations where the weekly target is greater than \$100,000 and the highest is \$200,044.

Table 5-2 presents the important predictors table for the final CART model, along with the CART defined “Variable Importance”. Care is needed in the interpretation of the importance score. For instance, a variable can be used for only one split high up the tree and be given a lowish score. However, the variable is still an important predictor in the model. Nevertheless, examination of the table and the model (not presented) shows that:

- Total weekly payments in the past quarter are the best predictor of future weekly payments
- Past quarter payments in section 37, medical, rehabilitation, and investigation are also important
- Cumulative payments to date for weekly, physiotherapy/chiropractic, and section 36 are important

- The existence of case estimates for weekly, investigation is important
- The impairment level on paid section 66 (permanent injury) benefits is a predictor
- The severity score produced by the combination of injury nature and location is important, particularly for short duration claims where the payment history isn't fully developed
- The last payment period end date for weekly benefits is, indicating whether claimants have recently received weekly benefits, or how long ago they may have ceased, is a predictor.

Table 5-2 – Weekly CART Model Important Predictors

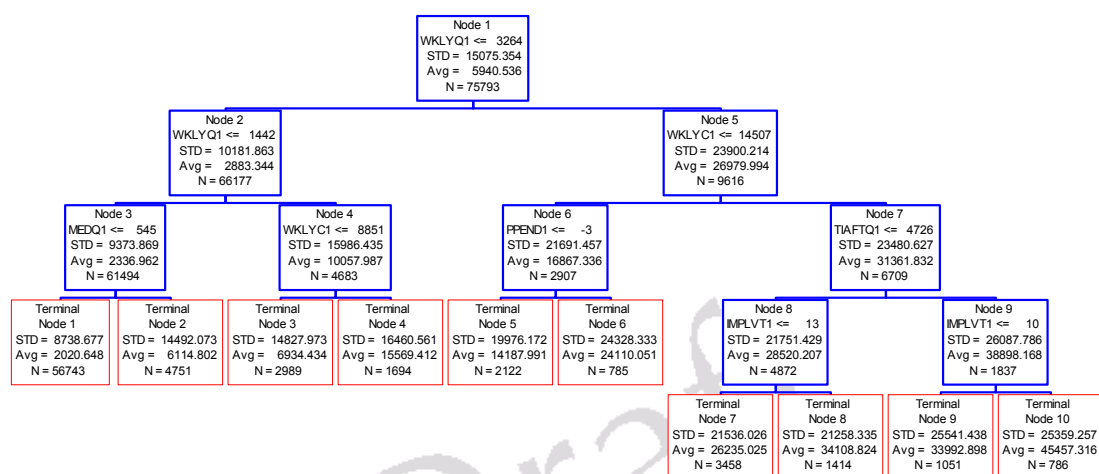
Variable	Importance
Weekly Payments Last Qtr	100
Weekly Payments Cumulative	10.34
Total Incapacity (after 26 wks) Payments Last Qtr	3.09
Impairment Level	2.48
Injury Severity Scale (Weekly)	1.63
Medical Payments Last Qtr	1.63
Days Since Initial Payment Date	1.47
Days Since Last Payment Period End Date	1.26
Weekly Case Estimate Binary	1.1
Interpreter Required Flag	0.74
Injury Location	0.65
Insurer	0.58
Investigation Case Estimate Binary	0.47
Physiotherapy Payments Cumulative	0.44
Policy Premium Experience Modifier	0.41
Rehabilitation Treatment Last Qtr	0.39
Other Payments Cumulative	0.37
Resumed Work Date Binary	0.3
Total Incapacity (first 26 wks) Payments Cumulative	0.27
Investigation Payments Last Qtr	0.27

Figure 5-2 presents the pruned weekly CART tree with 10 terminal nodes. This provides some interesting insights:

- Node 1: High quarterly weekly payments results in an increase in future average weekly costs by a multiple of 4.54 (\$26,980/\$5,940). This multiple is termed the “lift index” or just the “lift” of the split. Low quarterly weekly payments go left and have a lift of 0.49.
- Node 3: The low last quarter weeklies in this node are probably mostly inactive claims and relatively new claims. The subsequent split for these claims is on medical payments indicating that higher medical payments is a strong predictor of future weekly compensation (lift of 2.61) for inactive and short duration claims.
- Node 6: Active weeklies with less than \$14,500 cumulative weekly are split on the days since the last payment period end date. This split demonstrates that if the last payment period end date was more than 3 days ago, the claims are less likely to continue on benefits, resulting in a lift index of 0.84.

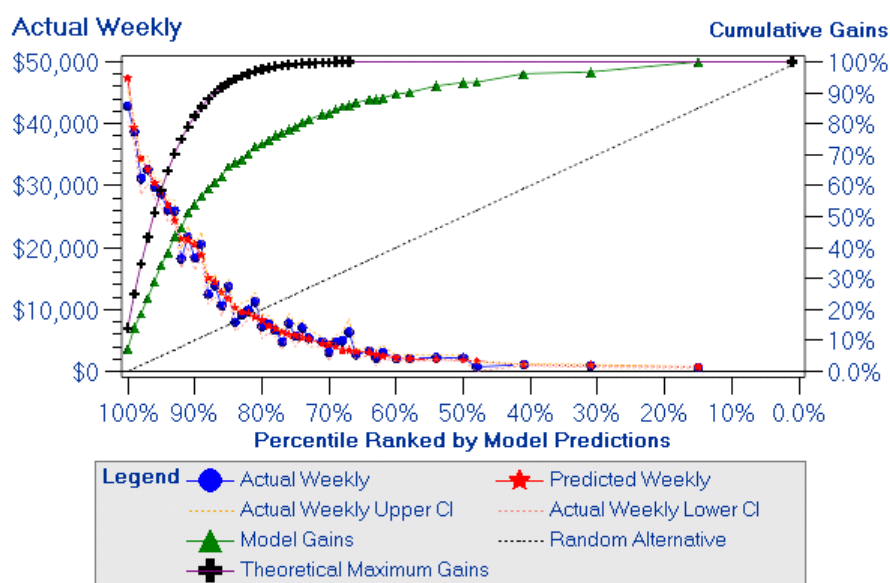
- Node 7: These 6,709 claims are those with high quarterly weekly payments and high cumulative weeklies (also a proxy for longer duration claims). Node 7 then splits on the level of section 37 quarterly payments.
- Nodes 8 and 9: Both of these splits are based on the paid impairment level for section 66 benefits. Higher impairment indicates higher future cost for both of these splits.

Figure 5-2– Weekly CART Tree with 10 Terminal Nodes



The full CART model has 68 terminal nodes and a differentiation in predicted values from a low of \$712 to a high of \$52,261. The actual versus expected and gains charts for this model are presented below.

Figure 5-3– Weekly CART Gains Actual vs Expected and Gains Charts 3 yr Payments Model

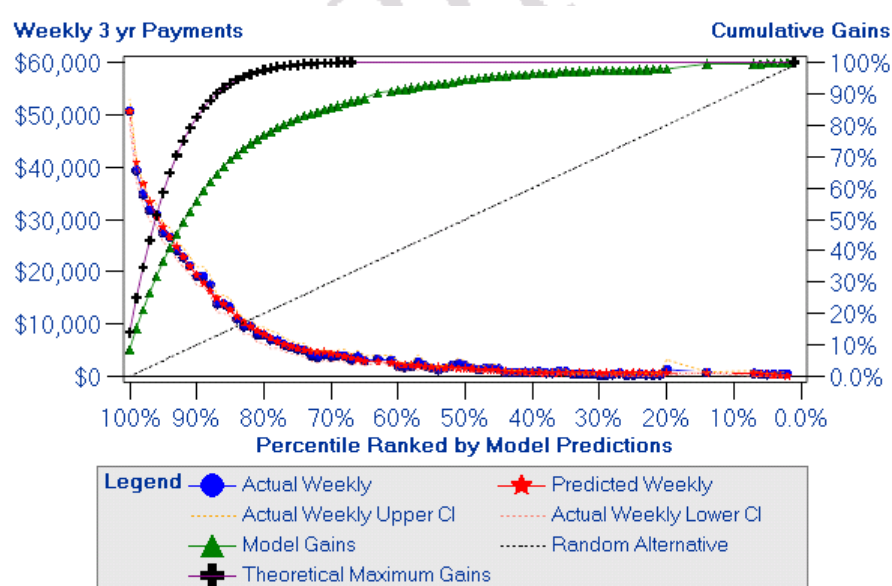


5.1.2 Weekly MARS Model

The MARS model was constructed using the weekly CART model terminal node number as a 68 level categorical predictor and all other continuous predictors. All class predictors were also tested in the model but were found not to add significantly to the predictiveness and so we dropped them for the final model. This was not unexpected since the CART node number ought to capture most of the information regarding the categorical predictors and MARS add to the predictiveness of the CART model by fitting functional forms to the continuous predictors.

Two way interactions with other basis functions already in the model were also allowed. MARS observes the hierarchy of including lower order interaction variables in the model even if they aren't significant but the higher order interaction is significant. 100 basis functions were initially selected in the forward selection phase and 63 of these subsequently eliminated in the backward elimination phase, leaving 37 in the final model. The most important predictor in the model was, not surprisingly, the weekly CART model node number.

Figure 5-4– Weekly MARS Actual vs Expected and Gains Charts 3 yr Payments Model



The actual versus expected chart for the MARS model is presented above and shows a marked improvement over the CART model. Firstly, the average actual target in the top percentile has increased from around \$44,000 in the CART model to around \$50,000. Secondly, the actual averages match the expected more closely and appear a great deal smoother. The gains for the MARS model are also substantially greater across range of the predictions. In the top decile of the predictions MARS captures 56% of the total cost and CART only captures 51%.

5.1.3 Medical CART Model

The medical payment type includes payments made for medical treatment, hospital treatment and ambulance services.

Some summary statistics for the target variable are below.

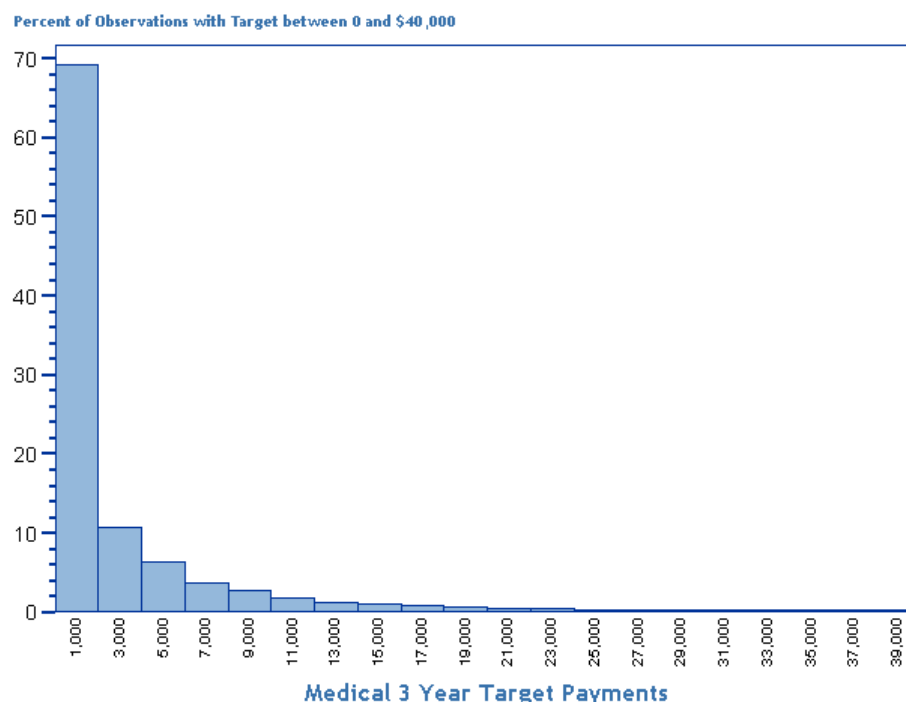
Table 5-3 – Summary Statistics for Medical Target Payments

Number in learning dataset	114,127
Mean	2,122
Standard Deviation	11,481
Skewness	41.43
Kurtosis	2,785.14
Quantiles	
100% Max	1,184,440
99%	29,029
95%	9,874
90%	5,136
75% Q3	955
50% Median	96
25% Q1	0
10%	0
5%	0
1%	0
0% Min	-55,686
Percentage Negative	0.60%
Percentage Equal to Zero	40.41%
Percentage Positive	58.98%

The mean medical target payment is \$2,122 and the coefficient of variation is 5.41, which is twice that of weekly target payments. The skewness is highly positive and the large kurtosis value indicates a heavy tail. Almost 58% of the observations have a positive payment for medical in the three year period which is around 20% higher than for weekly. There were 785 observations with negative target payments again, likely representing small reversals in previous payments (730 of these were for less than \$1,000). The target payments for these observations were set to zero.

The graph in Figure 5-5 shows the histogram for medical target payments between \$0 and \$40,000 which demonstrates high level of positive skewness.

Figure 5-5 - Histogram of Medical Target Payments between \$0 and \$40,000



Although not shown in Figure 5-5 - Histogram of Medical Target Payments between \$0 and \$40,000, the extreme observations for the medical target are significantly worse than for weekly (and any other target we modelled). There are 509 claims that have a medical cost between \$40,000 and \$100,000 and 124 above \$100,000. Although the number of claims above \$100,000 is similar to the weekly target, the spread is much more severe with the 13 observations above \$500,000 and the most extreme at \$1.2m.

The important predictors table for the final CART model is presented below. We note that:

- Cumulative medical payments are the best predictor of future medical payments.
- Medical, weekly, and medical treatment payments in the last quarter are also strong predictors.
- Days since initial payment and development month are both important and capture the effect of claim duration.
- The industry classification for the employer of the injured worker is an important predictor.
- The days from the injury date to cease work date, indicating whether or not there was a lag between the injury and the incapacity of the claimant, is a predictor. In general, gradual onset, latent and recurring claims will have longer lags.

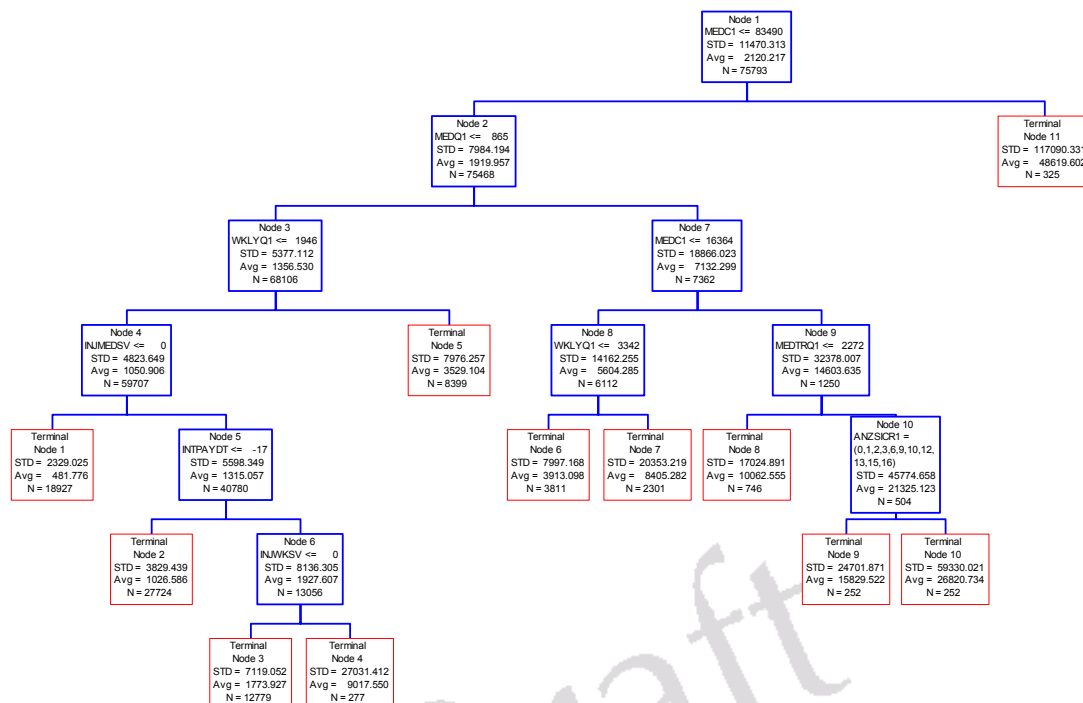
Table 5-4 – Medical CART Model Important Predictors

Variable	Importance
Medical Payments Cumulative	100
Medical Payments Last Qtr	30.56
Weekly Payments Last Qtr	9.56
Medical Treatment Payments Last Qtr	5.08
Days Since Initial Payment Date	2.79
Injury Severity Scale (Weekly)	2.74
Injury Severity Scale (Medical)	2.43
Medical Treatment Cumulative	1.57
ANZSIC (Level 1 Code)	1.08
Interpreter Required Flag	0.91
Development Month	0.9
Investigation Case Estimate Binary	0.54
Days Since Last Payment Period End Date	0.49
Days from Injury to Ceased Work Dates	0.49
Other Payments Last Qtr	0.39
Policy Premium Experience Modifier	0.37
Investigation Payments Last Qtr	0.26
Total Incapacity (first 26 wks) Payments Cumulative	0.23
Physiotherapy Payments Cumulative	0.21
Insurer	0.2

The pruned medical tree with 11 terminal nodes is presented below. We note that:

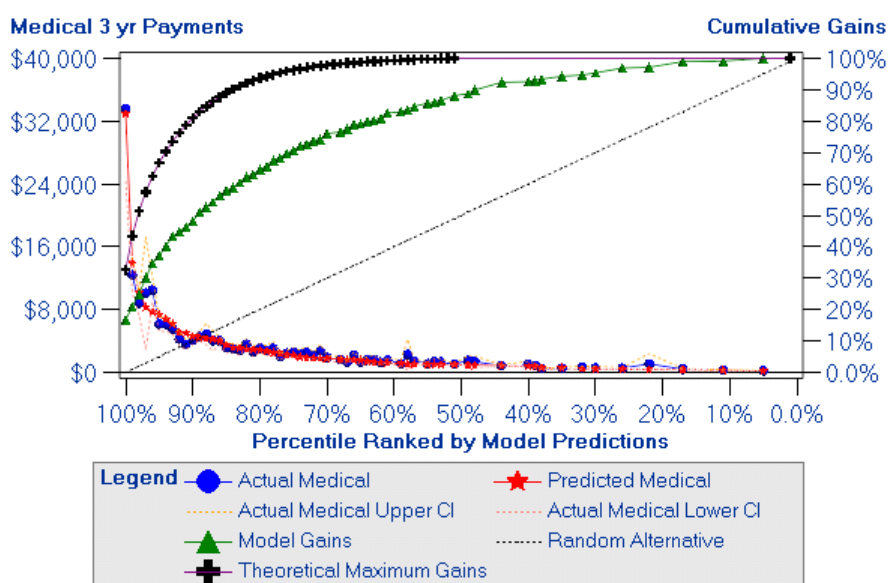
- Node 1: A small number of claims (325) with cumulative medical costs greater than \$83,490 are split right and have an average future medical cost of almost \$49,000 (a lift of 2.29). Quite a few of these claims would be the catastrophically injured.
- Node 2: High quarterly medical payments result in an increase in future average medical costs with a lift of 3.71.
- Node 3: High weekly payments in the last quarter result in a lift of 2.60.
- Node 4: Injury severity scale (Medical) equal to 0 results in a lift of 0.46.
- Node 5: Claims with initial payment date less than 17 days ago have a lift of 1.47.

Figure 5-6– Medical CART Tree with 11 Terminal Nodes



The full CART for medical has 97 terminal nodes and differentiates predictions between a low of \$179 and a high of almost \$49,000. The actual versus expected and gains charts for this model are presented below.

Figure 5-7– Medical CART Gains and A-v-E Charts 3 yr Payments Model



5.1.4 Medical MARS Model

Again the MARS model was fitted using the medical CART terminal node number as a categorical predictor and all other continuous predictors. Two way interactions were allowed and 100 basis functions were added in the forward selection phase. 42 basis functions were eliminated in the backward elimination phase leaving 58 basis functions in the final model. Again the most important predictor is the medical CART node number.

Examination of actual versus expected charts identified several issues with this model. In particular, several of the basis functions in the model were influenced significantly by a group of outliers that were identified as the very high cost, catastrophically injured claims. Several methods were employed to counter this problem including reducing the number of basis functions, transforming or capping the target and/or some predictors, and the exclusion of certain predictors. Finally, we adopted a solution of using a GLM procedure to review the selection of the basis functions using a separate cross validation dataset. The general method is described in paragraph 4.3 and particulars are given below.

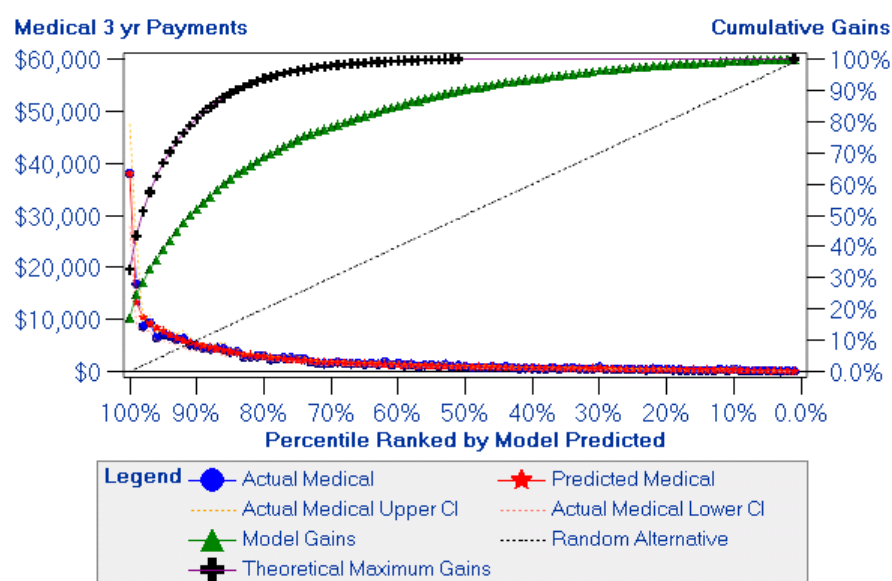
5.1.5 Medical GLM

The primary difference with the medical approach is that the MARS model is built on a random 50% of the learning data and the other 50% is used for cross validating the selected basis functions within a GLM. Type 3 tests were used to identify and eliminate the weakest basis functions, one at a time until all the remaining predictors were significant. This process resulted in the removal of a further 25 basis functions leaving 33 in the model.

After the final set of basis functions was selected the parameters were re-estimated based on the entire learning sample.

The final GLM model evaluations are presented below. The predictions are again much smoother than for CART and reach almost \$40,000 in the top percentile (compared with around \$33,000 for CART). The GLM model also captures 50% of the total cost in the top decile compared to 48% for CART.

Figure 5-8– Medical MARS Gains and A-v-E Charts 3 yr Payments Model



5.2 Payment Patterns

5.2.1 Weekly Patterns

After the three year payment model is built, the total predicted amount needs to be broken down into quarterly predictions. The superficial reason to do this is so that the cash flow can be inflated and discounted. From this point of view, the accuracy of the payment pattern is of secondary importance. However, the more important use of the quarterly cash flows is to monitor the validity of the model and to assess whether or not some recalibration is required. This latter purpose demands a reasonably accurate payment pattern.

Our general approach to this problem is to identify homogeneous groupings of claims by the pattern of payments over the three year period and fit a smooth curve to that pattern within each group.

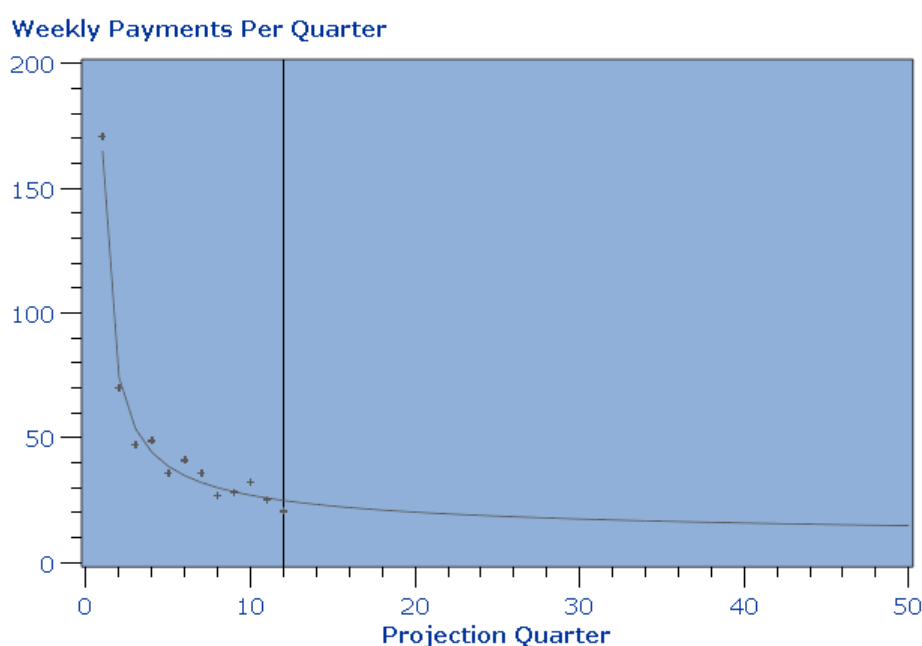
The pattern of payments for weekly compensation is broadly related to the rate of decay in active claims over the three year period which in turn is highly correlated with the level of cumulative weekly payments over the period. Using this reasoning we derived homogeneous groupings of claims for payment pattern fitting based on the three year payments CART model for the appropriate payment type where the terminal nodes of a CART tree can be thought of as reasonably homogeneous groups with respect to the target variable. The full tree for weekly compensation has 68 terminal nodes that, in order to simplify the analysis, we pruned back to 50 terminal nodes for pattern fitting.

Fitting a smooth curve to the quarterly payment patterns was carried out using regression modelling. The log of the mean quarterly payment amount

was modelled as a linear function of either the quarter number, the log of the quarter number, or the log log of the quarter number.

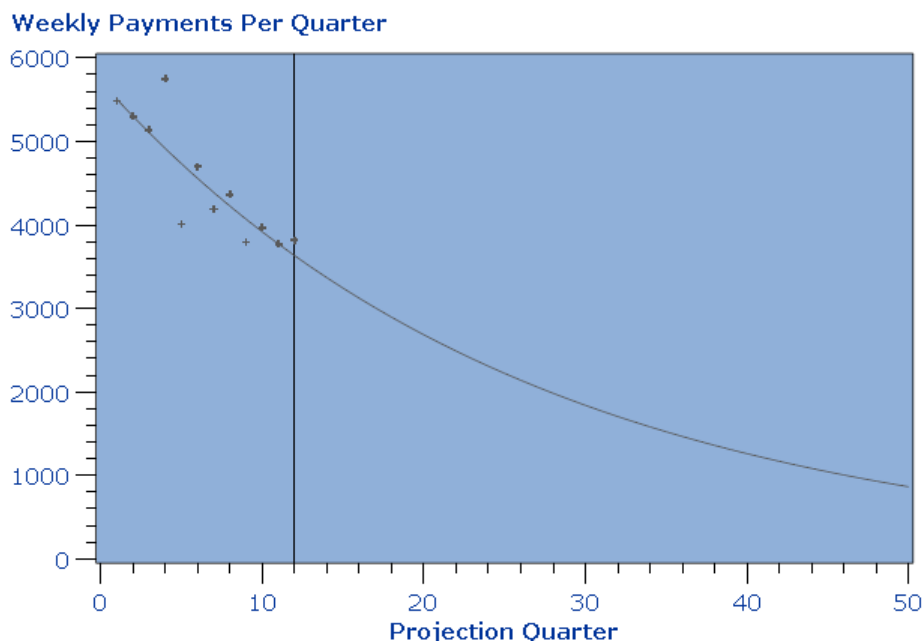
The graphs below demonstrate the curves fitted to the lowest and the highest cost nodes for the weekly compensation model. The “+” symbols represent the actual mean quarterly payments and the black line is the fitted regression curve. The curves are extrapolated up to projection quarter 50 to demonstrate how the tail pattern may look (although the curve is not actually used past the 12th quarter).

Figure 5-9 – Actual and Fitted Payment Pattern for the Lowest Cost Weekly Node



The lowest cost payment pattern (Figure 5-9) uses the log log of the quarter number as the only predictor and achieves almost 96% R-square. The graph shows quite clearly that the expected payments are very low with only around \$170 expected to be paid in the first quarter. This is almost 30% of the total three year prediction for this node. After projection quarter 1, the predicted pattern drops away rapidly to about \$25 (4% of 3 year prediction) by quarter 12.

Figure 5-10 – Actual and Fitted Payment Pattern for the Highest Cost Weekly Node



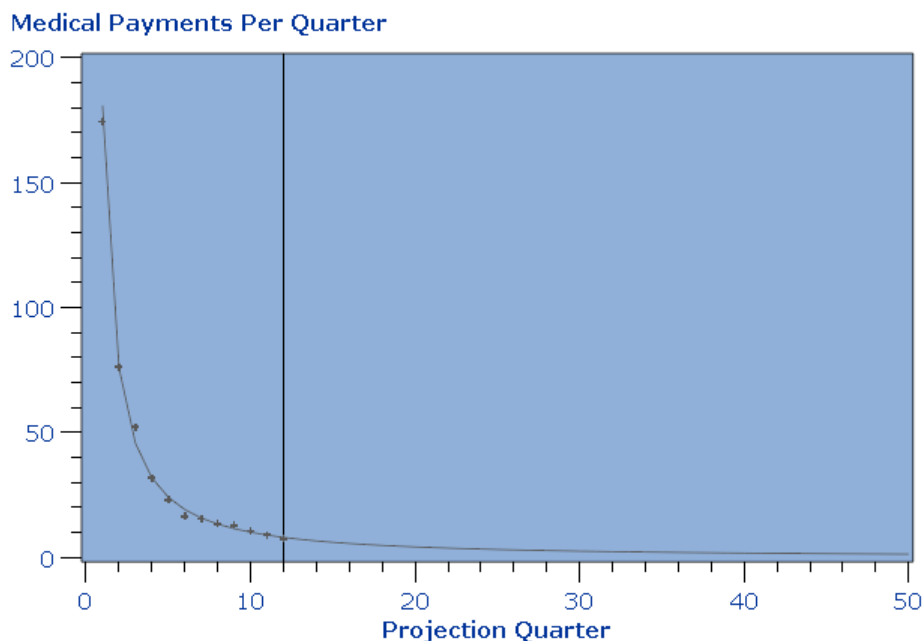
The highest cost node (Figure 5-10) uses the quarter number as the only predictor and achieves around 75% R^2 . Note the likely accelerated payments in quarter 4 and the subsequent low payments in quarter 5. The first quarter prediction for these claims is \$5,500 (10% of 3 year prediction) and payments subsequently drop away relatively slowly to \$3,600 (6.7% of 3 year prediction).

Across all of the patterns fitted for the weekly compensation payment type there were varying degrees of residual variability. The highest R-square reached up to 98.6% and in fact most of the patterns were above 95%. However, there were some patterns with high variability and even a couple where the R-square was below 70%.

5.2.2 Medical Patterns

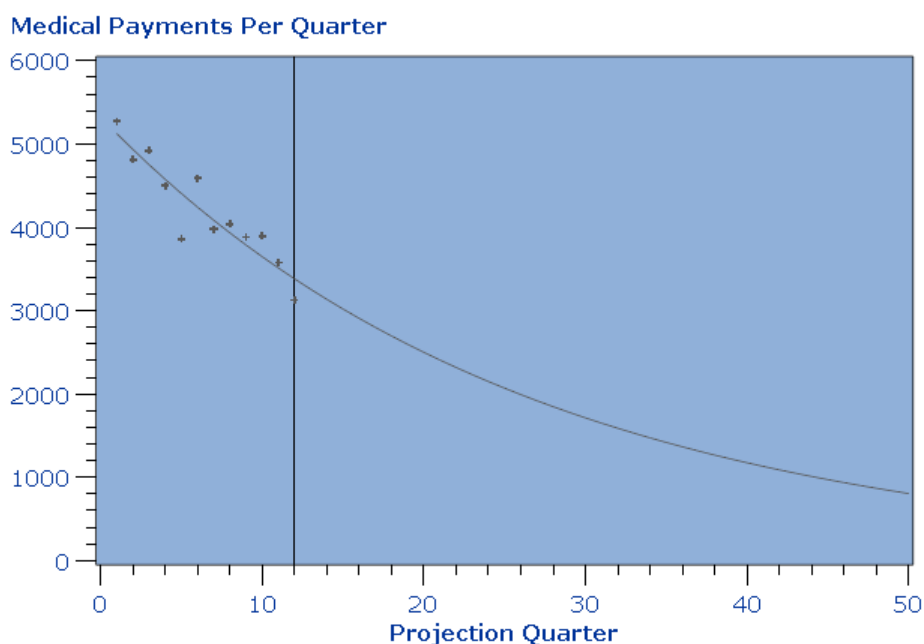
The graphs below demonstrate the curves fitted to the lowest and the highest cost nodes for the medical model.

Figure 5-11 – Actual and Fitted Payment Pattern for the Lowest Cost Medical Node



The lowest cost payment pattern (Figure 5-11) uses the log of the quarter number as the only predictor and achieves 99% R-square. The pattern demonstrates a steep decline in expected payments from \$170 in projection quarter 1 to only \$10 in projection quarter 12.

Figure 5-12 – Actual and Fitted Payment Pattern for the Highest Cost Weekly Node



The highest cost node (Figure 5-12) uses the quarter number as the only predictor and achieves 84% R^2 . The first quarter prediction for these claims is \$5,100 and predictions at the 12th quarter are still \$3,400.

5.2.3 *Issues with the pattern fitting approach*

This is one of the weaker parts of the methodology. The justification for using the nodes of the CART trees as homogeneous grouping with respect to the pattern of payments is not convincing and there is evidence in some of the model evaluations that these grouping are not, in fact, homogeneous. We are currently working on an SCE for another client where we are giving this issue more attention.

5.3 Tail Hazard Fitting

We also have to extend our primary model outside the three year period. For many claims the payments outside this period will be very small. However, for very serious claims as much as 70% of the liability can be in respect of payments outside the three year period. Therefore, this part of the modelling requires some care.

The general problem for fitting tail patterns is one of extrapolation of the quarterly pattern. The chosen extrapolation method was the exponential survival curve that takes the following form:

$$S_T(t) = \exp(-\lambda t)$$

where $S_T(t)$ is the probability of surviving longer than t periods, and λ is the constant hazard rate (or rate of decay).

This method was chosen because:

- The only parameter extrapolated is a constant hazard rate (trends would be more dangerous and problematic),
- The curve shape is monotonically decreasing (we can not imagine the expected quarterly payments for any claim actually increasing, after at least 3 years from projection) and generally satisfactory, and
- The estimation of the hazard parameter is relatively straightforward.

The hazard rates after the 12th projection quarter are likely to vary considerably from claim to claim. The solution is to divide the claims into groups that have very similar hazard rates and then estimate and project the future payments within these groups. Our first attempt at finding hazard rate groups was to use the payment pattern groups described in section 5.2 however, this produced far too many groups and some seemingly spurious hazard rates for some groups. The final method adopted was to build a CART tree on the hazard rate and use the terminal node number as the hazard rate grouping.

The final problem is to choose the period over which to estimate our hazard rates. For very short duration claims the hazard rate is not likely to be constant for some time however, for very long duration claims it is likely that the hazard rate is stabilising.

Our investigations suggested that most claims could be regarded as having a constant hazard rate from the end of the 8th quarter so we used the continuity of payments between quarters 8 and 12 as the basis for the hazard rate estimation.

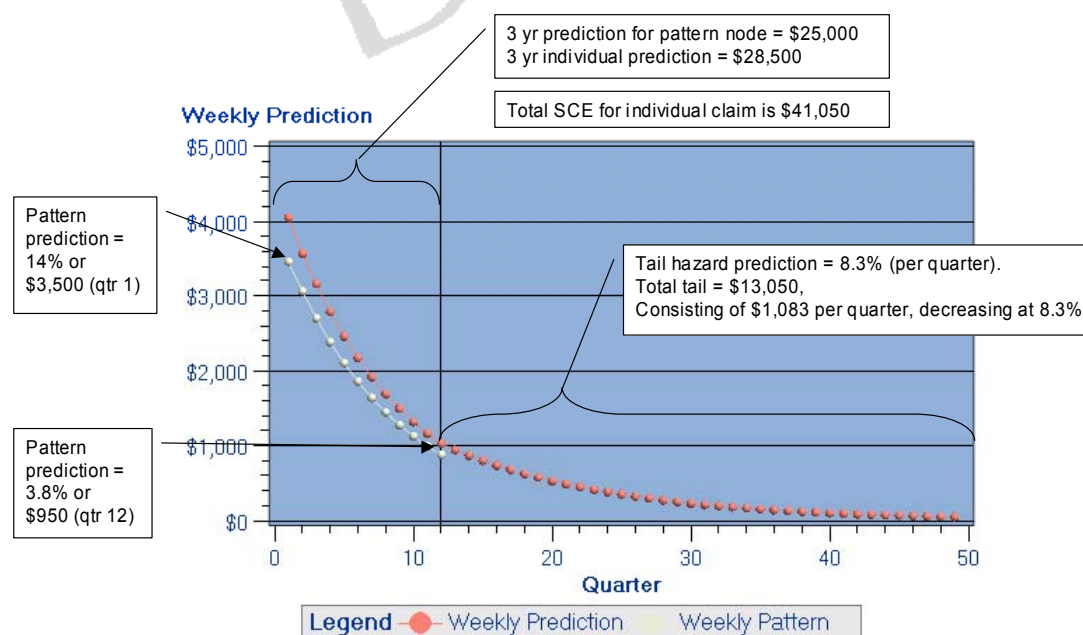
5.4 Combining the Three Year Model, the Payment Pattern and the Tail Extrapolation

Combining the various models into a “life-time” payment stream is reasonably straightforward. Broadly speaking, one:

- Expresses the payment pattern for a node as a percentage of actual three year payments for the node;
- Applies this pattern to the predicted three year payment amount for each claim in the node;
- Continues the predicted payment stream past the 12th quarter, using the predicted payment for the 12th quarter and the predicted hazard rate for the claim.

We illustrate this graphically for one claim below:

Figure 5-13 – Cash Flow Projection Example



6 Performance of the NSW WorkCover Model

In section 5 we provided a reasonably detailed explanation of the modelling approach for the weekly and medical payments. In this section we provide an overview of how the model performs for these payment types and for total payments.

6.1 Three Year Predictions

The raw fit statistics for the 13 payment types modelled, over the three year period, are presented in the table below. Here we have chosen to present the R-square value because it is the most widely understood (although misleading), the Root Average Square Error (RASE), and the Coefficient of Variation (CV) which is the RASE divided by the Mean Predicted (allowing that the RASE and RMSE are not significantly different).

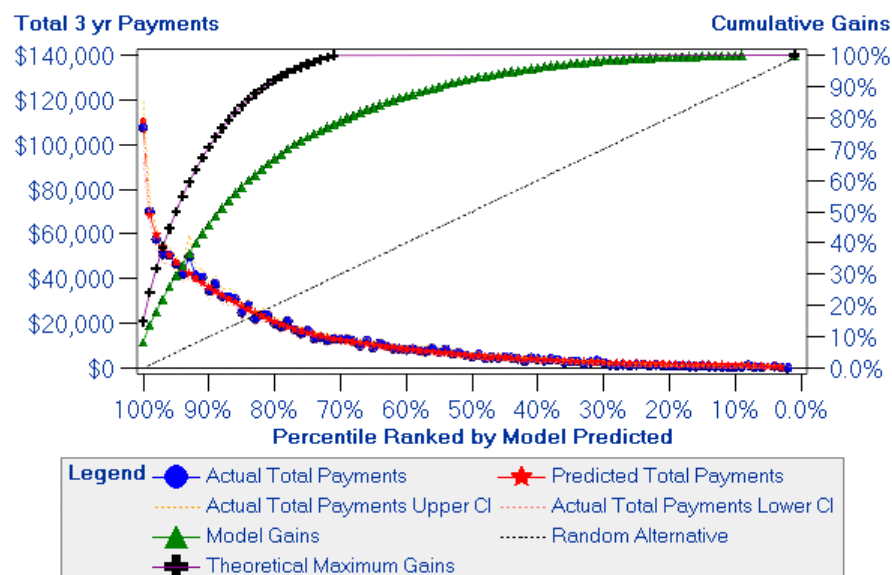
Table 6-1 – Fit Statistics with Payments Over 3 Years

Payment Type	Payment Period	Mean Predicted	Root Average Square Error	Coefficient of Variation	R-Square
Total Net Payments	3 Years	12,913	30,071	2.33	42.7%
Weekly	3 Years	5,871	12,317	2.10	53.1%
Medical	3 Years	2,217	10,957	4.94	50.0%

For all models the actual versus expected plots demonstrate no significant bias in the predictions. The final medical model has a CV of more than twice the weekly model. This is partly due to the more extreme nature of the medical payment type, particularly for the catastrophically injured.

The total net payments achieve a CV of roughly 2.3 and a R-square value of around 43%. The actual versus expected and gains chart evaluation for the total net payments is presented below.

Figure 6-1 – Total Net Payments Over 3 Years Actual versus Expected and Gains Charts



This graph compares the combination of all of the individual payment type model predictions to the actual net payments. The modelling approach employed effectively fits each payment type independently of the others. If there is any significant correlation between these payment types then the combination of the predictions may result in instability or bias. Figure 6-1 demonstrates that correlation between payment types has not resulted in considerable instability or bias.

Predicted three year costs reach up to around \$110,000 on average in the top percentile and the gains chart demonstrates that the top decile captures around 42.7% of the total cost (however the perfect model would capture around 70%).

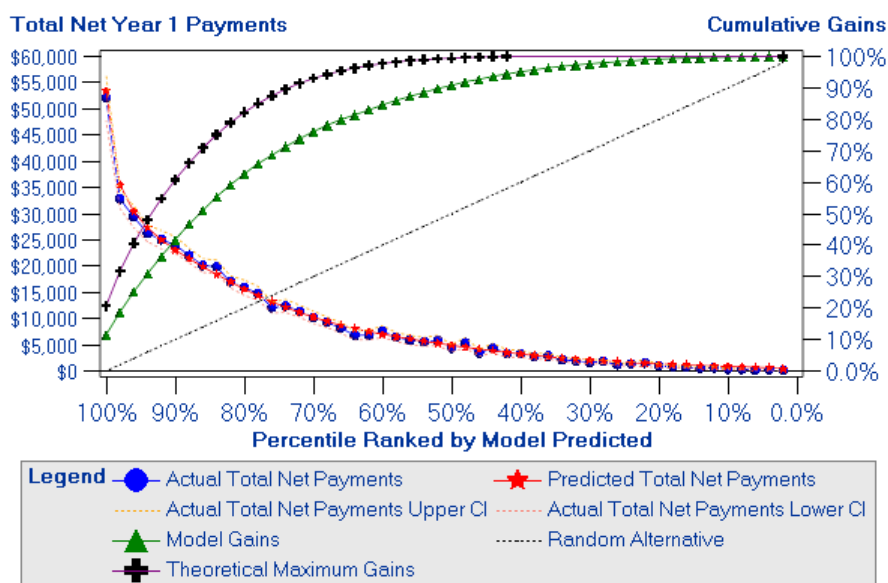
Final model evaluation charts for weekly and medical payment types were discussed in sections 5.1.2 and 5.1.5.

Some sample evaluations for claims of different durations and active versus inactive claims are given in Appendix A.

6.2 Pattern Predictions

Actual versus expected total payments for the first year of the three year modelling period are shown below. This shows a reasonably good match. Unfortunately, evaluations for the second and third year were not available for this paper.

Figure 6-2 – Total Net Payments Over 1 Year Actual vs Expected and Gains Charts



6.3 Tail Predictions

The tail predictions are probably the most uncertain component of this SCE model. This is because there is inevitably some degree of extrapolation into the future needed, where there is no longer any data to validate the results. As we were constructing the model and further data became available we were able to test the predictions against this data in the tail quarters 13 and 14 from the 01 Jan 1999 modelling date.

The graphs below exhibit the standard evaluation charts for the 13th and 14th projection quarters for the weekly compensation and medical payment types.

Figure 6-3 – Total Net Payments in Quarters 13 and 14 Actual versus Expected and Gains Charts

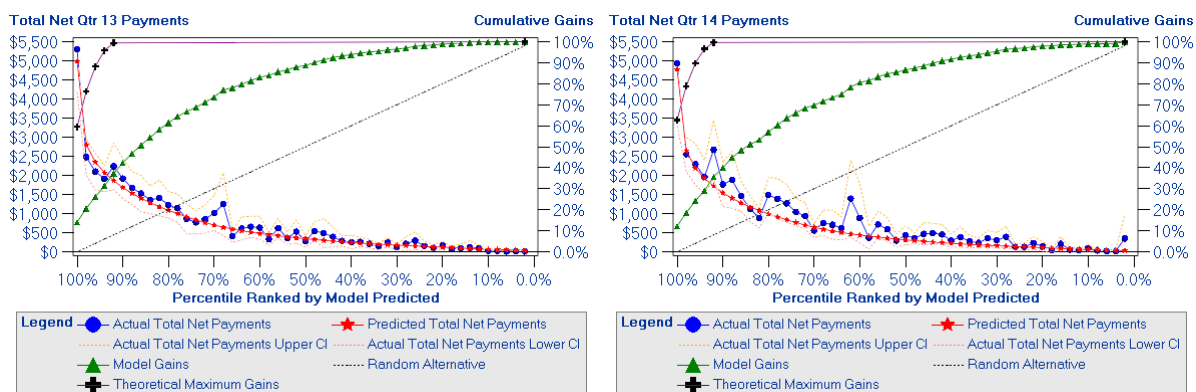


Figure 6-4 – Weekly Payments in Quarters 13 and 14 Actual versus Expected and Gains Charts

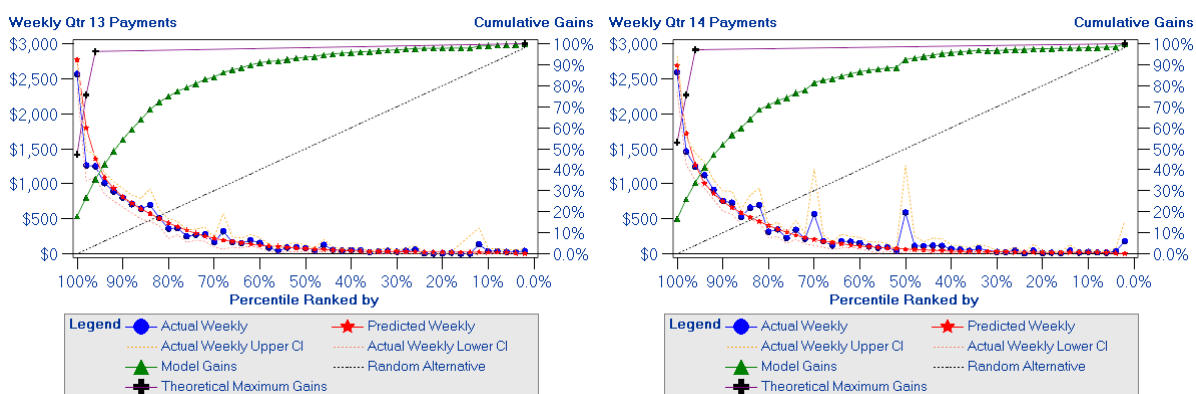
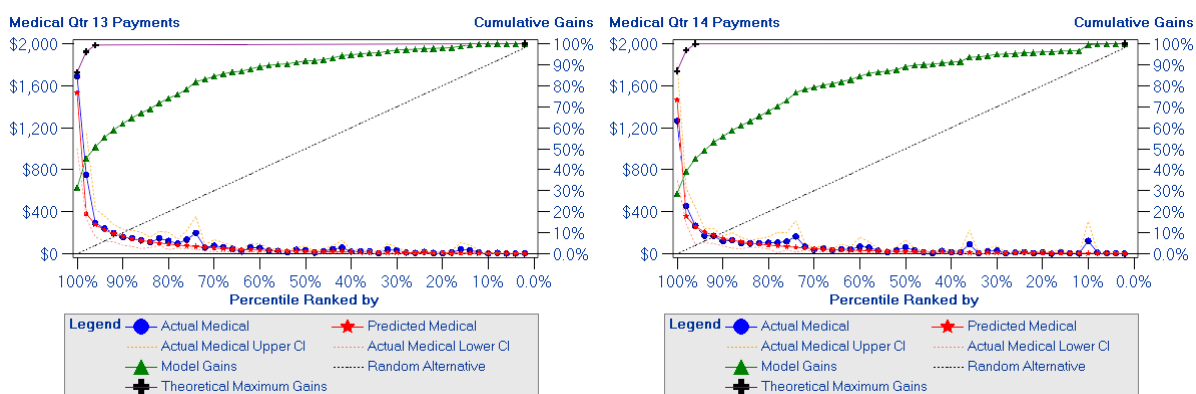


Figure 6-5 – Medical Payments in Quarters 13 and 14 Actual versus Expected and Gains Charts



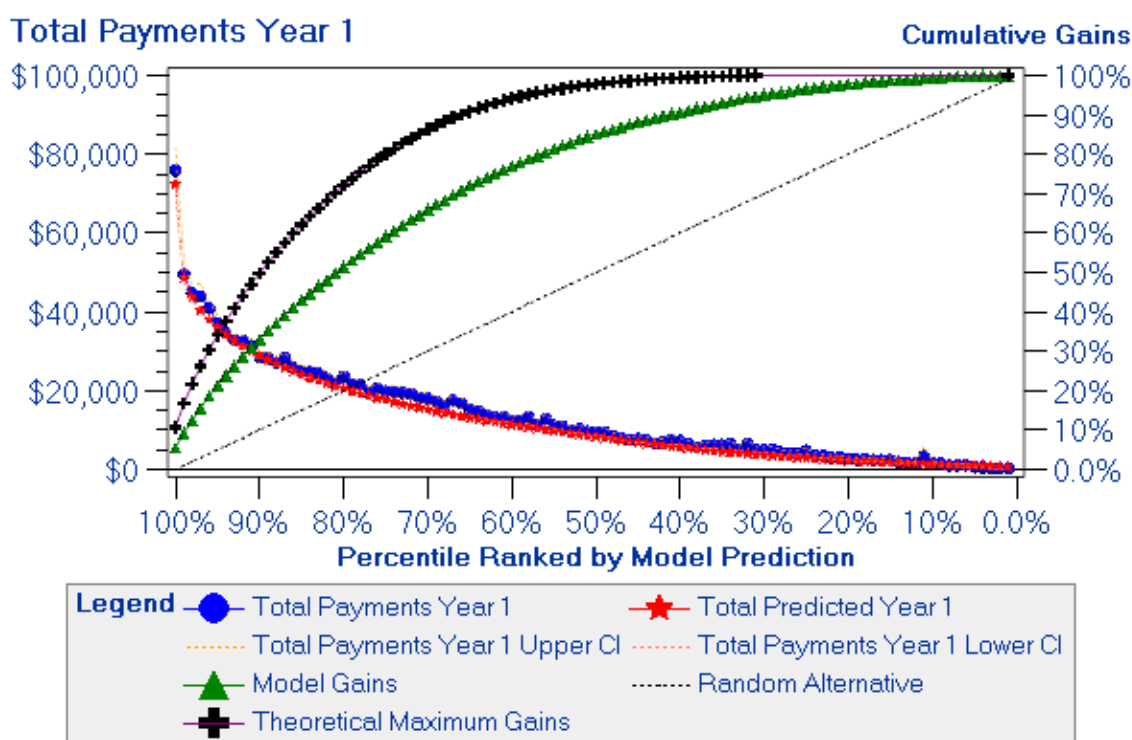
The graphs demonstrate the actual and expected payments match reasonably well even 3 or so years after projection although there is a sign of under-prediction for total payments in quarter 14. This could be due to the unusual payments and behaviour around that time, due to the recent legislative amendments in 2001. The graphs also suggest that the tail fitting procedure

employed is reasonably effective for the first two quarters that the method is used.

6.4 Recent Predictions

Although the model can be evaluated and shown to be robust as at the modelling date, the ultimate test is how it will perform today. It is paramount to the projection of sensible SCEs to be reasonably certain that the model is performing well and if it isn't, to understand where the inadequacies are. As such we have built in to the projection algorithm, a series of evaluations at each of the 4 quarters prior to the current projection quarter. Therefore when we projected total SCEs as at 30 June 2003, one of the evaluations automatically produced was actual versus expected one year total net payments, using the expected values from the model one year earlier. The graph below shows this comparison.

Figure 6-6 – Total Net Payments Over 3 Years Actual versus Expected and Gains Charts



The CV and R-square values for this comparison are 1.43 and 49.4%. Note that both the actual and expected values for the highest percentiles are considerably higher than those in section 6.2. Significant bias is present with actual payments exceeding expected after adjustment for known inflation. This bias is reasonable consistent across the entire range of the predictions and represents the effect of super imposed inflation between the dates when the model was built and the date of projection. The series of evaluations produced when the projection algorithm is run also serves the purpose of analysing the effect of super-imposed inflation by payment type. Through

analysis of these evaluations and comparison with the actuarial valuation for the scheme it is clear that there are several payment types with superimposed inflation over this period, notably rehabilitation (which increased by almost 50%) or accelerated payments, for instance statutory lump sums and their associated legal payments.

The SCE is parameterised in such a way that any observed superimposed inflation can be adjusted for by payment type and projected into the future. This should mean that the model can be recalibrated less frequently.

The fit statistics for total net, weekly and medical payment types as at 30 June 2002 are presented in the table below.

Table 6-2 - Fit Statistics as at 30 June 2002 with Payments Over 1 Year

Payment Type	Payment Period	Mean Predicted	Root Average Square Error	Coefficient of Variation	R-Square
Total Net Payments	Year 1	13,664	19,503	1.43	49.4%
Weekly	Year 1	4,210	6,625	1.57	51.0%
Medical	Year 1	1,642	6,198	3.77	41.3%

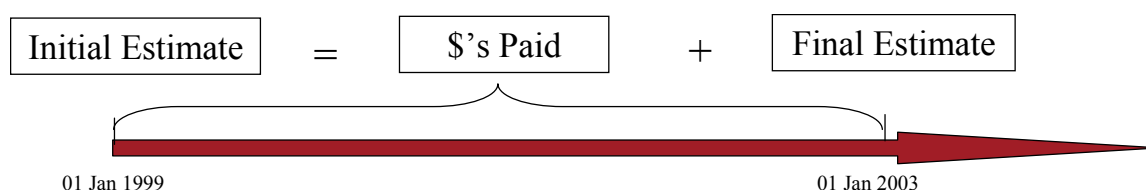
Both the weekly and medical payment types appear reasonable over 1 year. The CVs are down from the 3 year model (indicating a better fit) because the 1 year target is less uncertain and variable.

6.5 SCEs versus Manual Case Estimates

The comparison of SCEs and manual case estimates can be undertaken in many ways, the first of which is how effectively does each estimate predict future claim cost. The comparison of actual and expected for SCEs is relatively easy because a full payment projection is present and hence the corresponding cash flows can be compared. For manual case estimates the timing and pattern of expected cash flows is not produced and hence we need to employ another method for the comparison.

6.5.1 Predictiveness

We have compared the two estimates using outstanding cost development on a claim by claim basis. The initial estimates were collected/calculated as at 01 January 1999 and then compared to the final estimates as at 01 January 2003 (4 years later) plus payments in the intervening period.



The table below shows that the CV for weekly case estimates is considerably higher than for SCEs, indicating greater residual variation from case estimates and lower predictiveness. The R-square values show that roughly half of the variation is explained with SCEs compared with less than 20% for case estimates.

For the medical payment type, the CVs are generally lower than for weekly however, SCEs are still considerably better than case estimates. R-square values here indicate that SCEs explain around 45% of the variation while case estimates explain almost 30%.

The comparison for total case estimates versus total SCEs is not possible because some of the payment types modelled for SCE purposes, do not have respective manual case estimates and hence there would be a mismatch in total between the two estimates.

Table 6-3 – Predictiveness of SCEs and Manual Case Estimates

Payment Type	Estimate Type	Mean	Root Average Sq Error	Coefficient of Variation	R-square
Weekly	Case Estimate	14,712	51,865	3.53	18.9%
Weekly	SCE	8,564	21,095	2.46	49.0%
Medical	Case Estimate	4,206	36,860	8.76	28.2%
Medical	SCE	3,541	20,281	5.73	45.4%

Another method for comparing the 2 types of estimates is using gains charts. Figure 6-7 below compares the weekly SCEs' and case estimates' ability to rank the claims open as at 01 January 1999 by cost over the next 3 years. The lower black line ("+" symbols) represents the ranking by case estimates and demonstrates that the top 10% (decile) of case estimates capture 39% of the weekly payments over the next 3 years. The upper blue line plots the gains from ranking by SCEs and shows that 52% of the 3 year cost is captured in the top decile. The ability of the SCEs to capture more of the total cost at all points along the range of the data indicates that the SCE ranking of claims is superior.

Figure 6-7 – Gains Chart Comparisons for Weekly SCEs and Case Estimates

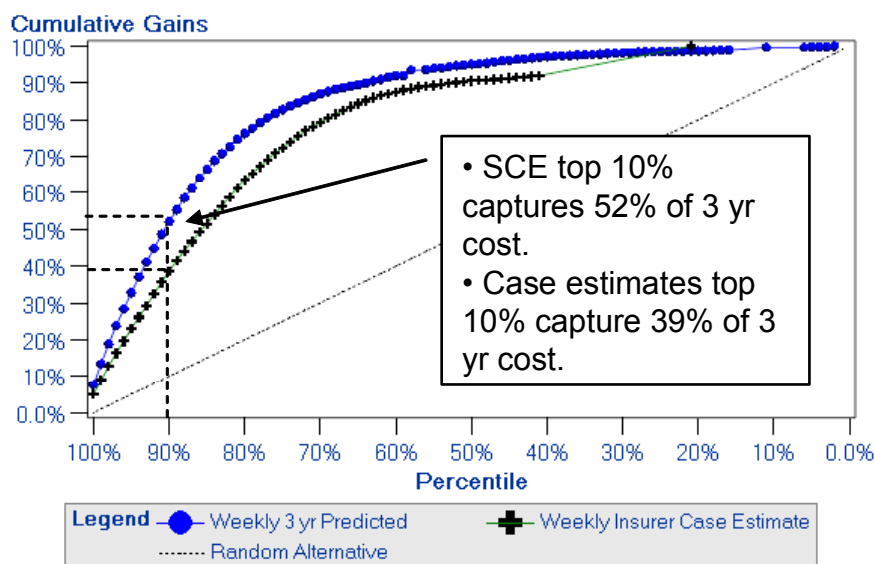
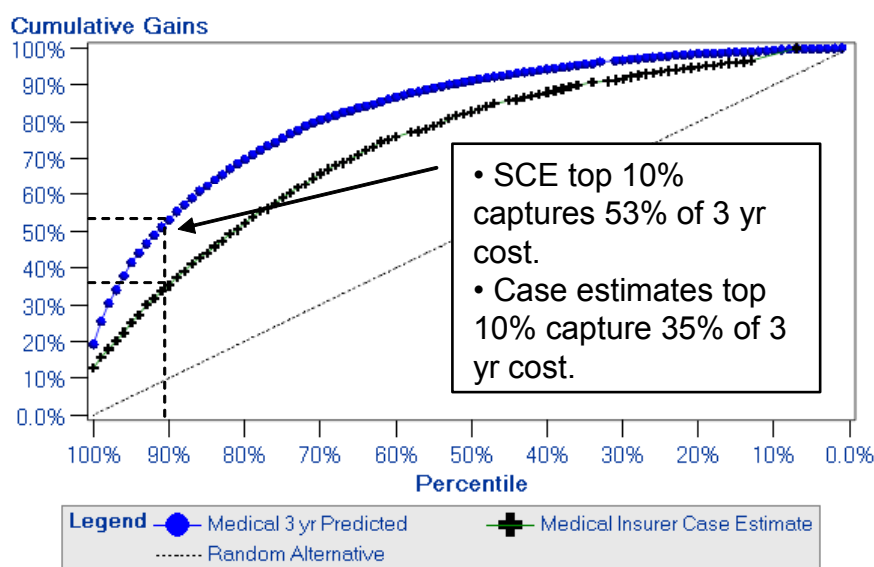


Figure 6-8 demonstrates the same comparison for the medical payment type. The top decile ranked by case estimates only captures 35% of the total cost while the top decile ranked by SCEs captures 53%. Again the ranking by SCEs is superior across the entire range of the data.

Figure 6-8 – Gains Chart Comparisons for Medical SCEs and Case Estimates



6.5.2 Development Over Time of High Value Estimates

Another comparison which can be made between SCEs and manual case estimates is the way in which they develop over time. Ideally, they should develop such that as payments are made through time, the estimates reduce

and the resulting total remains the same. We have assessed both the SCEs and manual case estimates in this regard for those claims with high value estimates.

Figure 6-9 is based on the top 1% of claims (874 claims) by manual case estimate as at 01 January 1999. The total case estimates on these claims as at 01 January 1999 was \$471m (an average of \$540,000 per claim), the left-most green bar. The total case estimates each quarter for these claims are represented by the declining green bars, while the cumulative payments on these claims since 01 January 1999 are represented by the yellow segment. Case estimates on these claims decline rapidly over the period to \$107m, \$28m has been made in payments, and 503 of the claims remain open.

Figure 6-9 – Development of Case Estimates for Top 1% of Weekly Claims by Case Estimate

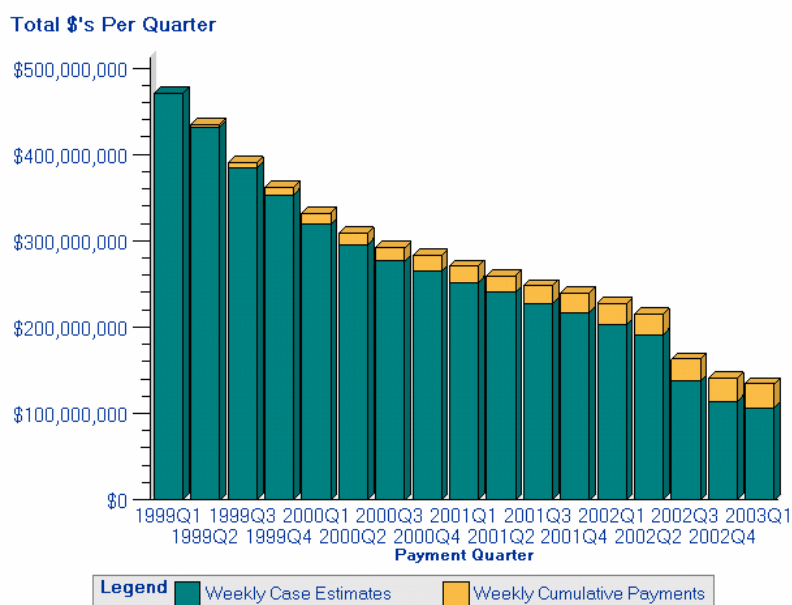
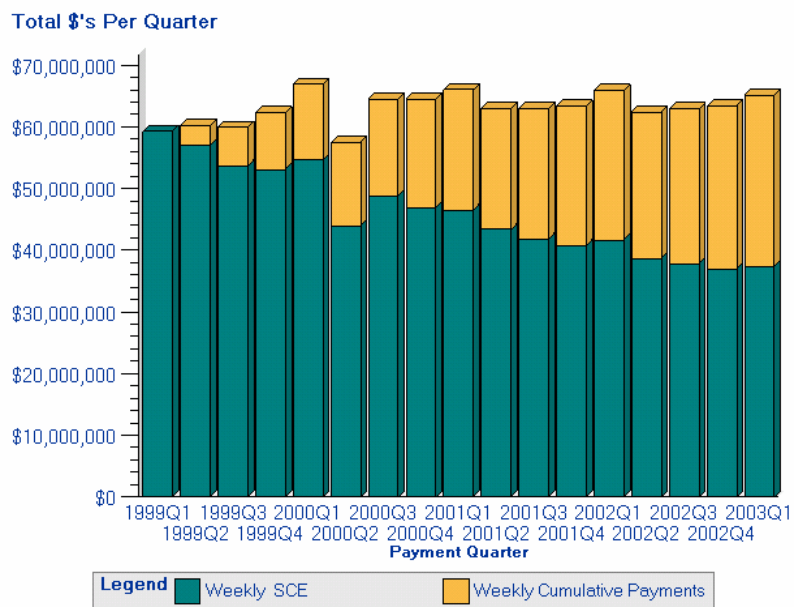


Figure 6-10 is based on the same 874 claims identified in the above graph however, the green bars now represent the SCEs on these claims. The total SCEs at the beginning of the period are \$59m (an average of \$68,000 per claim). In this case, the total SCEs plus payments do not noticeably decline or increase over the 4 year period. As at 01 January 2003 there is \$37m in outstanding SCEs, which combined with payments of \$28m results in \$55m of cost post 01 January 1999 (comparing reasonably with the original \$59m estimate).

Figure 6-10 – Development of SCEs for Top 1% of Weekly Claims by Case Estimate



7 What next?

7.1 Applications

To date, the SCE has been used for research into issues such as the premium formula but it has not yet been integrated into the operations of the Scheme. In this sense, other jurisdictions probably have more experience in the actual application of SCEs. However, in theory, the SCE ought to display the following features when compared with the two other standard methods of liability calculation:

Features	Conventional case estimates	Actuarial outstanding claims valuation	Statistical case estimates
Robust and objective	✗	✓	✓
Available at the individual level	✓	✗	✓
Automatically update	✗	✗	✓
Allow for IBNR	✗	✓	✗
Can be inflated and discounted	✗	✓	✓
Allow for trends in the claim profile	✓	?	✓
Allow for trends in the environment	?	✓	?

Some of this table may need a little explanation:

- The fact that we do not regard conventional case estimates as robust and objective should not be taken as a criticism of the people who set them. The fact is that with a large portfolio, a large number of people will be involved in the setting of conventional case estimates and it is difficult to standardise practice across this group. In addition, where the case estimates have an influence on an employer's premium rate there is the potential for the employer to attempt to influence the case estimate. Neither the actuarial calculation nor the SCE suffer from these difficulties.
- The SCE is an automated model that can run at any valuation date, reflecting information available at that date. Both the actuarial valuation and conventional case estimates require considerable resources and time to update, particularly the latter.
- Suppose there is a sudden change in the claim profile, say an increase of a particular injury type, that is not yet reflected in the long term payment experience of the portfolio. This will likely not be reflected in the actuarial calculation since most actuarial methods are based on overall payment levels and consider payment types independently. However, a claim manager setting a case estimate should take account of injury type and, to some extent, so should the

SCE, particularly if there is a combination of weekly and medical payments associated with the claim which indicates the severity.

- There are often changes in a workers compensation portfolio that are reflections of a change in culture, treatment protocols, predisposition to litigation and so on. Conventional case estimates might reflect these although probably not in a uniform fashion. To the extent that they are reflected in trends in payment levels, or the actuary allows for them explicitly then the actuarial liability will take them into account. However, the SCE probably will not, without recalibration. We have tried to incorporate some allowance for these trends in the NSW SCE by allowing the estimation and projection of super-imposed inflation for different payment types but, in our opinion, any SCE is unlikely to produce as good an estimate of the overall portfolio liability as standard actuarial techniques.

With these features in mind, the most straightforward application of an SCE is as a tool to allocate an overall outstanding claim liability, calculated by standard actuarial techniques, to sub-groups of claims. For instance:

- Pricing. The SCE ought to result in an accurate allocation of cost to small groups of claims and allow more accurate pricing by industry or employer.
- Insurer remuneration. The SCE could be used to allocate the outstanding claim liability between insurers for remuneration purposes. However, a standard actuarial valuation for each insurer would, arguably, do this just as well.
- Benchmarking the performance of service providers. In theory one could use the SCE to determine a benchmark cost for a group of claims and monitor the performance of, say, a rehabilitation service, in reducing the actual cost below the benchmark. Cost here includes weekly and other benefits so good performance includes improved return to work.

We believe that there are other potential applications also:

- As a monitoring tool to track the cost trends for sub-groups of claims. For instance, one could track the costs of claims of a particular injury type and monitor any trends.
- As a supporting tool for the actuarial valuation. Should the SCE show a different cost trend than the valuation, either for the whole portfolio, or for a sub-group of claims then this can be analysed and the information used to improve the actuarial assumptions.
- Formulation of lead indicators. Analysis of the drivers in the SCE can be used to define a set of lead indicators for the portfolio that can be monitored and used as an advance warning of any trends.
- Input to claim management. A robust indication of the likely cost outcome of a claim will be one of the inputs for deciding how the claim should be best managed and in the prioritisation of management resources between claims.

Finally, we discuss using the SCE as a replacement for conventional case estimates in their role as a tool for claim management. We believe that this

needs to be approached with caution. Part of the conventional case estimation process involves the gathering and analysis of information concerning a claim. We believe that it is vital that this process continues, even if the result is not formalised as a case estimate since:

- The information will be important in determining the appropriate management for the claim; and
- Some of the information is needed to support the SCE, for instance, whether or not there is legal involvement, whether there is a large outstanding recovery or the claimant's return to work status.

7.2 Refinements to the model

7.2.1 *Data*

We believe that the performance of the model is good, given the data that is available. However, there is no doubt that it could be much improved given more robust and extensive data. This is a much bigger issue than the SCE however we note that, strictly from the point of view of improving the SCE, the following would be helpful:

- The more robust coding of items such as injury nature, location and mechanism;
- A more regular and uniform payment regime. The existence of, for instance, employer reimbursement schedules is unhelpful since these delay knowledge of a claim's return to work status.
- The collection of other types of information. It is well documented that there are other (probably better!) indicators of a claim's likely outcome than the financial management information currently collected. For instance, the collection of claimant health status, psycho-social and attitudinal factors, and evidence based medicine flags would all improve the model performance markedly. This is especially true for claims that have not been open long at the valuation date so that past payment information is not yet a proxy for the claims severity and outcome.

8 Bibliography

Salford Systems, 2003. “An Overview of the CART® Methodology”, Salford Systems website, www.salford-systems.com

Salford Systems, 2003. “MARS® (Multivariate Adaptive Regression Splines)”, Salford Systems website, www.salford-systems.com

Salford Systems, 2003. “Hybrid CART-Logit Model in Classification & Data Mining”, Salford Systems website, www.salford-systems.com

Taylor G. and Campbell M., 2002. “Statistical Case Estimation” Centre for Actuarial Studies, Department of Economics, The University of Melbourne.

Draft

A. Appendix

A.1.1. *Short Duration Claims (Less than 3 Months Developed)*

Figure A-1 – Total Net Payments Over 3 Years Actual vs Expected and Gains Charts – Short Duration Claims

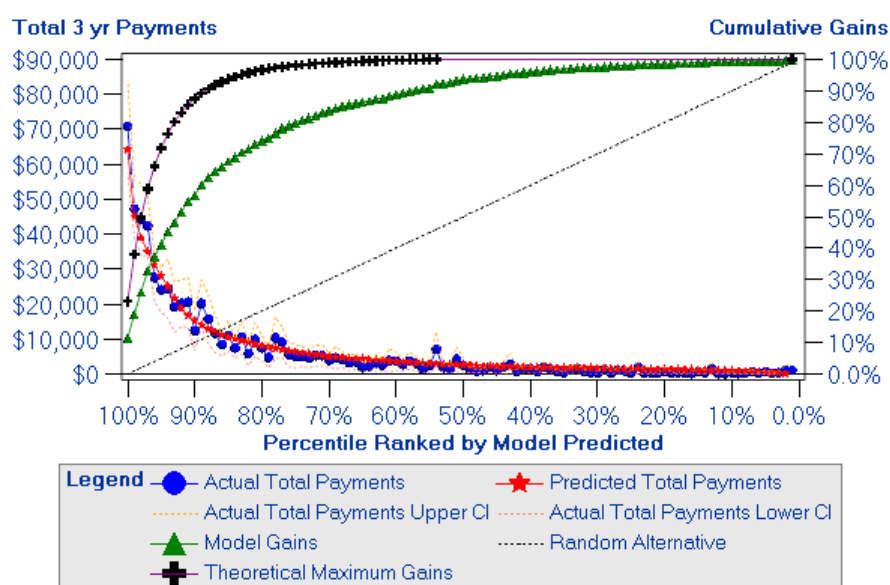


Figure A-2 – Weekly Payments Over 3 Years Actual vs Expected and Gains Charts – Short Duration Claims

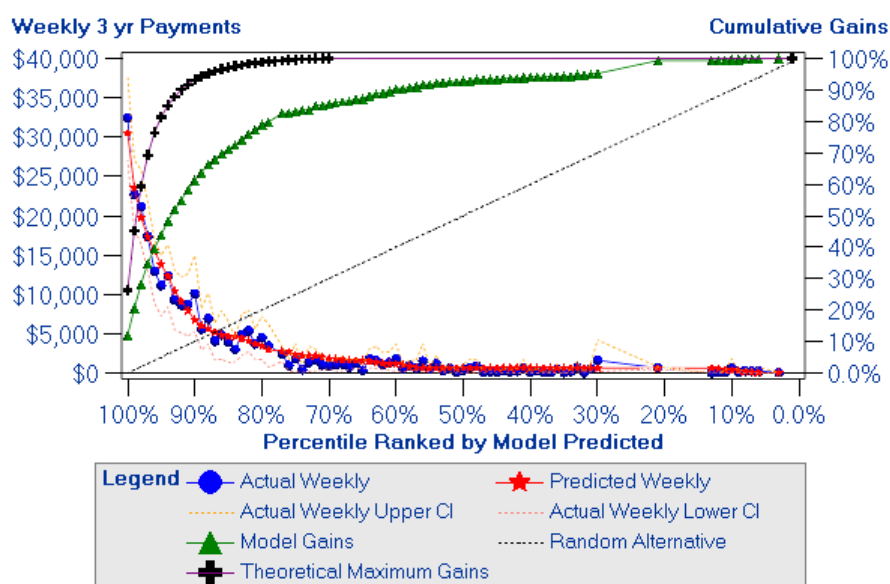
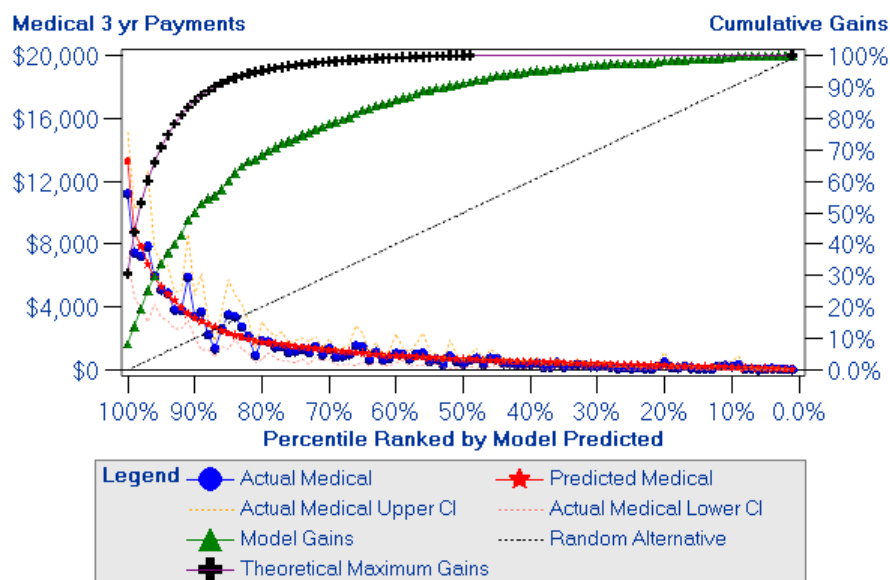


Figure A-3 – Medical Payments Over 3 Years Actual vs Expected and Gains Charts – Short Duration Claims



A.1.2. Medium Duration Claims (4 to 12 Months Developed)

Figure A-4 – Total Net Payments Over 3 Years Actual vs Expected and Gains Charts – Medium Duration Claims

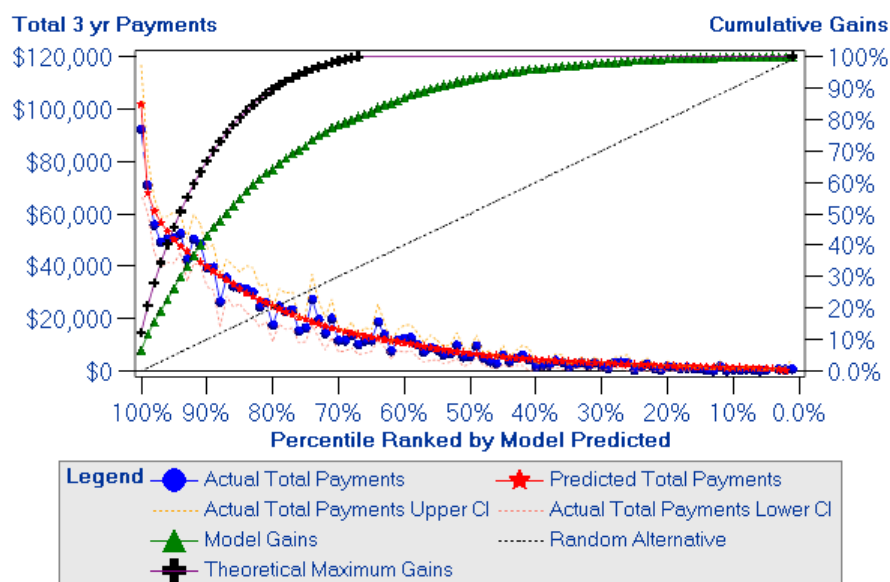


Figure A-5 – Weekly Payments Over 3 Years Actual vs Expected and Gains Charts – Medium Duration Claims

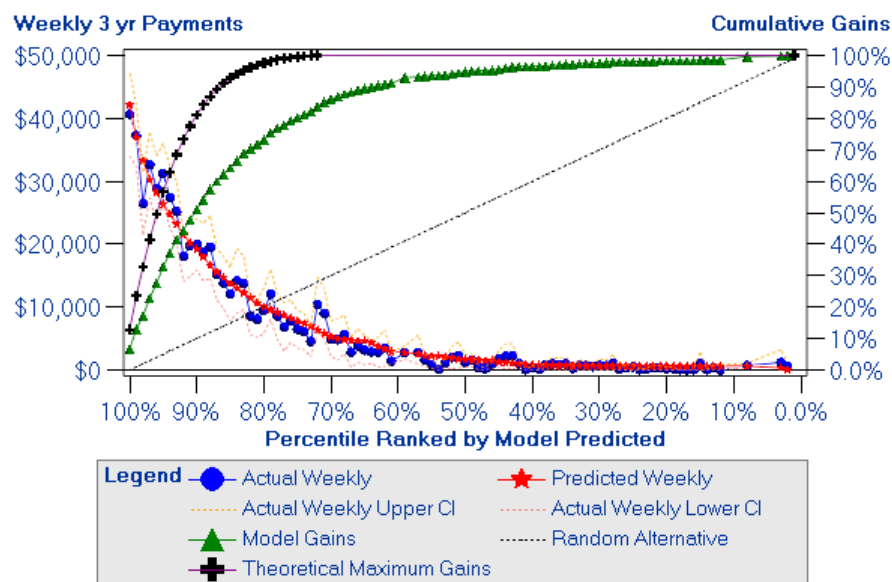
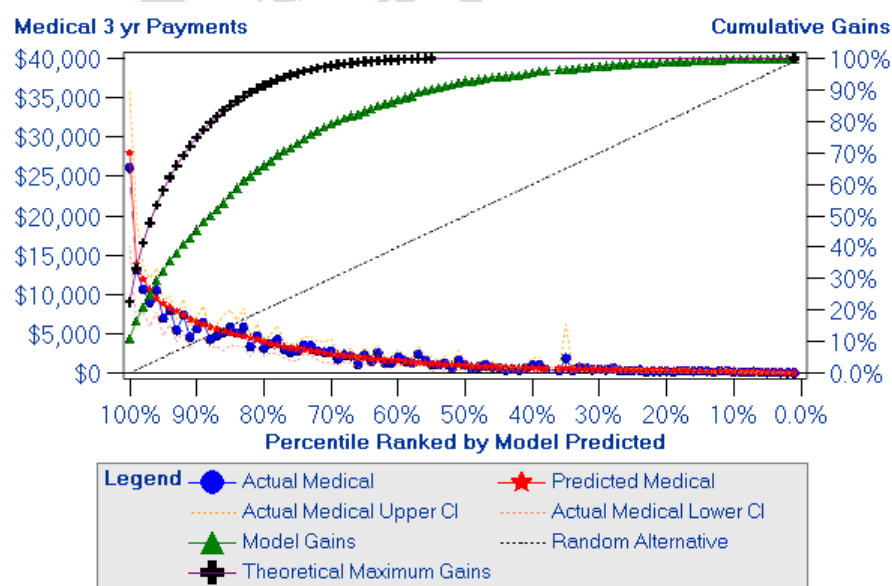


Figure A-6 – Medical Payments Over 3 Years Actual vs Expected and Gains Charts – Medium Duration Claims



A.1.3. Long Duration Claims (Greater than 12 Months Developed)

Figure A-7 – Total Net Payments Over 3 Years Actual vs Expected and Gains Charts – Long Duration Claims

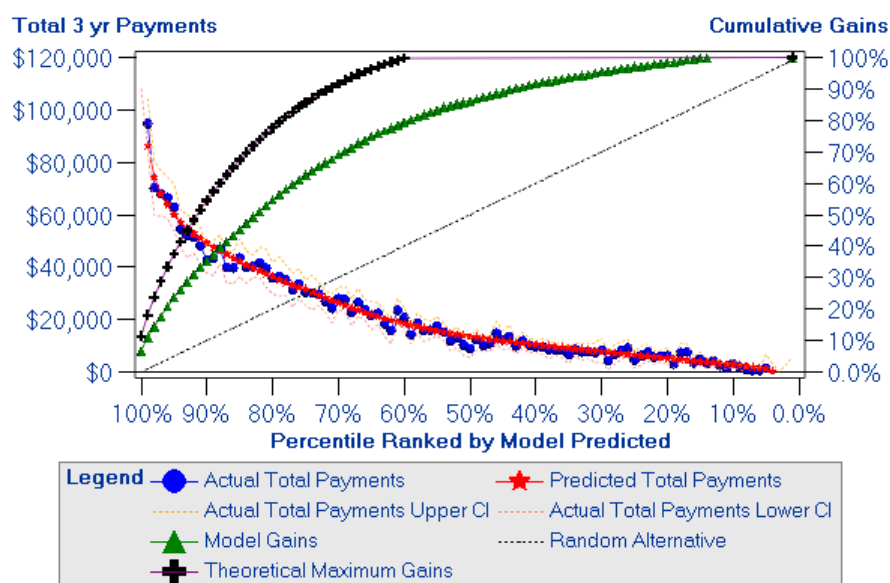


Figure A-8 – Weekly Payments Over 3 Years Actual vs Expected and Gains Charts – Long Duration Claims

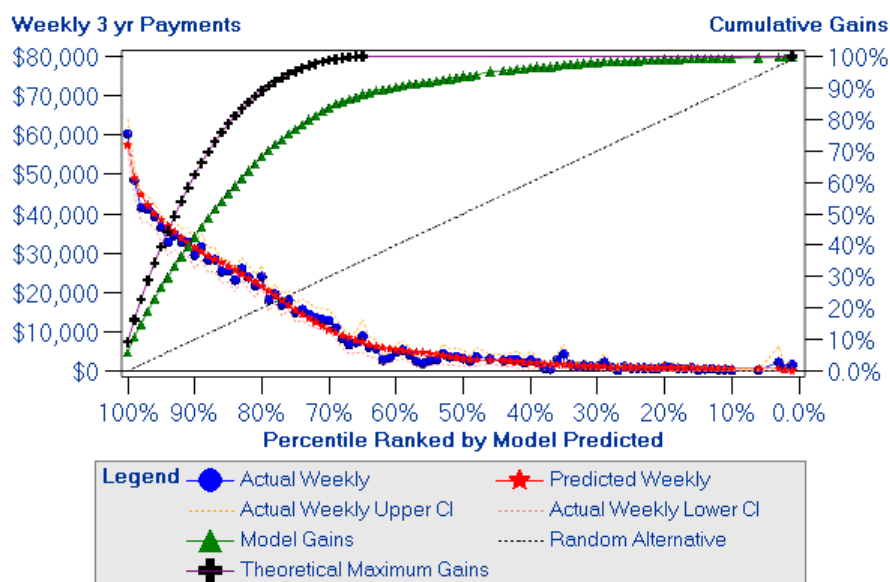
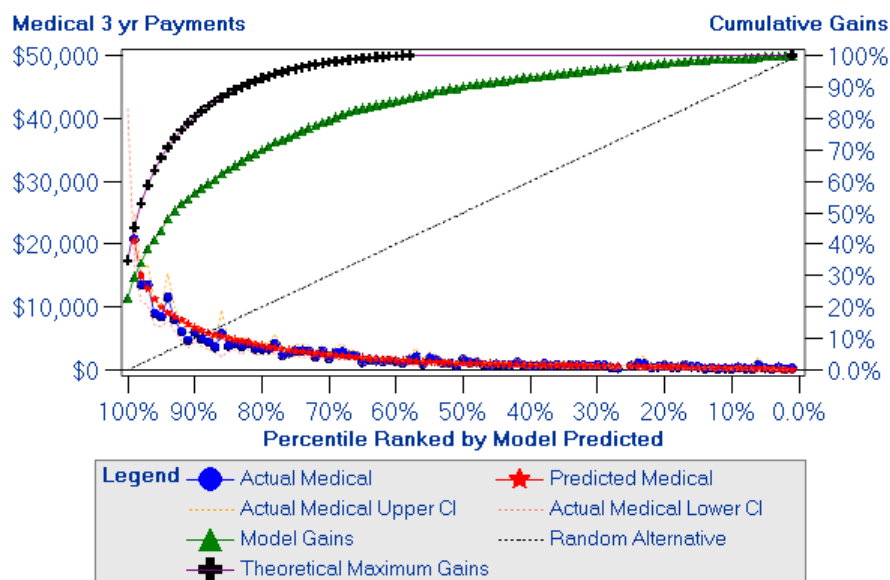


Figure A-9 – Medical Payments Over 3 Years Actual vs Expected and Gains Charts – Long Duration Claims



A.1.4. Active Claims

Figure A-10 – Weekly Payments Over 3 Years Actual vs Expected and Gains Charts – Active Weekly Claims

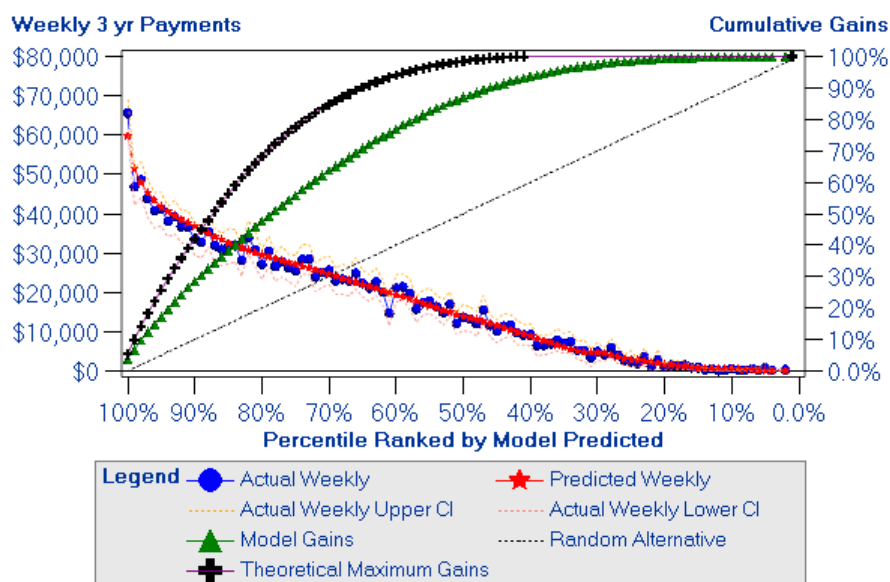
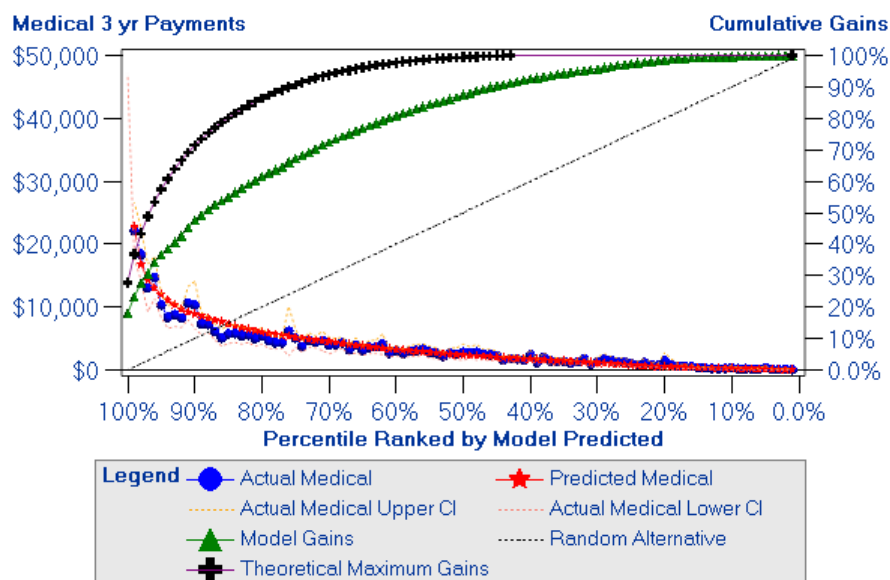


Figure A-11 – Medical Payments Over 3 Years Actual vs Expected and Gains Charts – Active Medical Claims



A.1.5. Inactive Claims

Figure A-12 – Weekly Payments Over 3 Years Actual vs Expected and Gains Charts – Inactive Weekly Claims

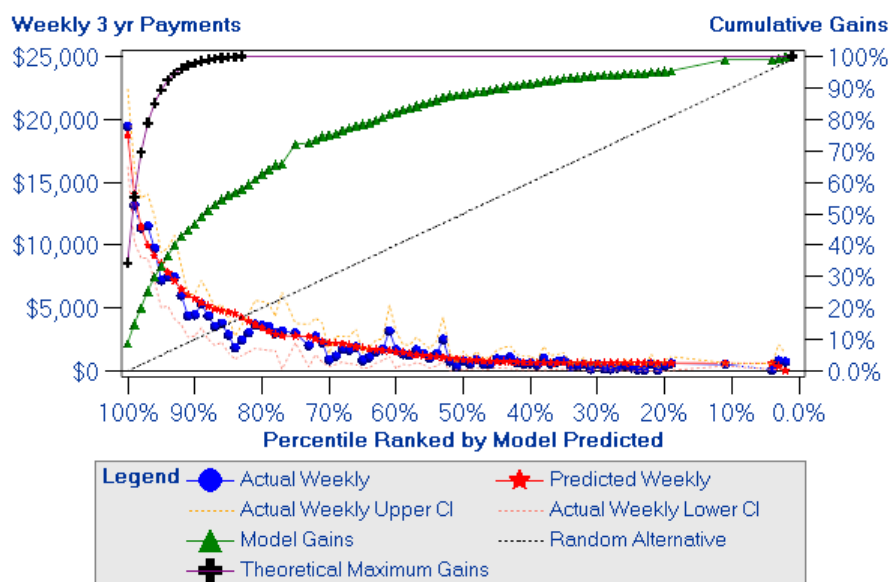


Figure A-13 – Medical Payments Over 3 Years Actual vs Expected and Gains Charts – Inactive Medical Claims

