

Demand modelling working party

James Tanser (Chair)

John Light

Sophia Mealy

Owen Morris

Special thanks to

Julie Fairbank

Matthew Barnes

1 Introduction

1.1 This working party was one of several suggested in the GRIP paper. Demand models are becoming increasingly important to the work of UK personal lines actuaries (and other pricing professionals) as they are a crucial part of the process which leads to price optimisation. Related topics have been covered in previous papers over the last 10 years or more, but demand models have not been explicitly examined in recent years.

1.2 Special thanks go to Julie Fairbank and Matthew Barnes for their hard work in helping the working party analyse the data in section 5.

1.3 So what do we mean by a demand model? A demand model is a model of customer behaviour, which seeks to predict future behaviour based on an analysis of past experience. In particular, it looks at the propensity of a customer to purchase a product (insurance, for example), and how the propensity changes based on the price of the product. The working party further concentrated on models which will return an accurate prediction of future behaviour (given as a probability of purchase), and not simply allow segmentation into high and low conversion cohorts.

1.4 The aims of the working party were:

- Provide an introduction to the topic describing the terms used
- Summarise the current methodologies used in the market
- Summarise possible alternate methodologies identified by a search of available literature
- Investigate several methods using agreed methodology to determine the descriptive and predictive power of the methods when applied to actual insurance data
- Provide a brief conclusion and highlight areas for further work.

1.5 As part of our research, the working party carried out a survey into the methods currently employed by those working in this area. Of the 32 individuals who started the survey, only 11 completed all the questions. Despite this relatively small level of participation, some interesting results were found, and are included in this paper.

2 Market practice

2.1 In this section we outline the types of model commonly used in demand modelling and which are being considered in this paper. A more detailed explanation of each model is given in section 3. We also provide selected results from the survey.

Main model types

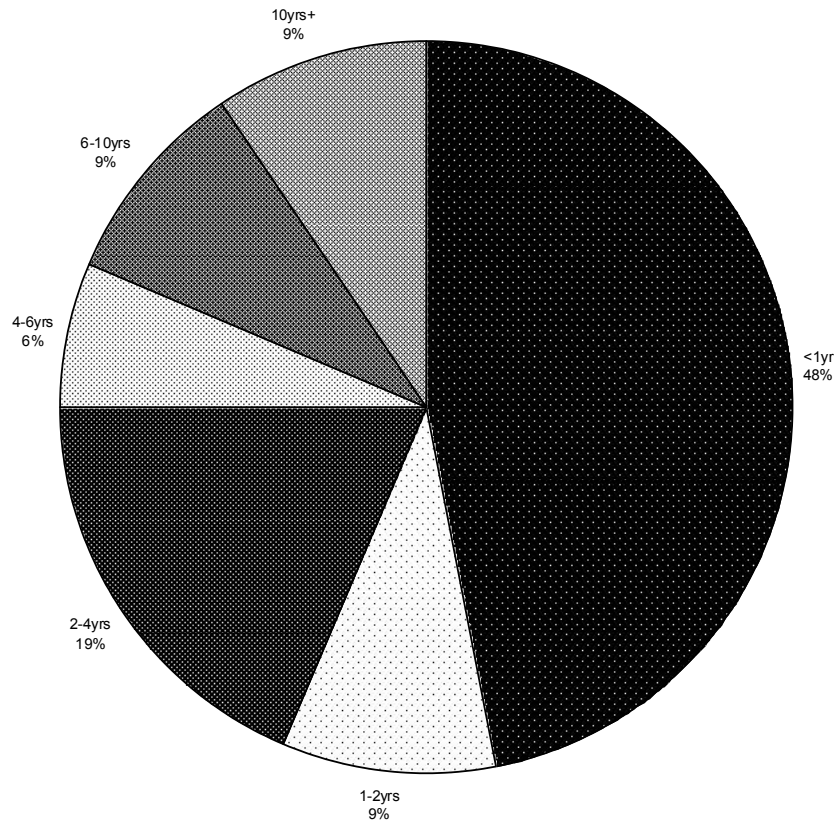
2.2 The models which are discussed are:

- 1** One-Way Analysis: By far the simplest analysis, the response variable to be modelled would be considered against each explanatory variable independently and in turn, e.g. looking at how 'demand' varies by consumer age.
- 2** Two-Way Analysis: The natural extension to One-Way Analysis, the response variable is considered against two explanatory variables at a time (e.g. looking at how 'demand' varies by consumer age and gender).
- 3** Generalised Linear Models (GLMs): GLMs provide a tool for considering many explanatory variables together. The GLMs family aims to address the problem of modelling correlations and interactions and can be used to model behaviour that is thought to depend on values of several other explanatory variables (e.g. demand depending on a combination of age, sex, location, price of competitors, price of other products etc). Simple linear regression is an example of the most basic GLMs.
- 4** Generalised Non-Linear Models (GNLMs): An extension of GLMs where the link function no longer needs to be applied directly to the linear predictor, but rather any function of the underlying explanatory variables. Although more complicated, these models may produce a more accurate predictive model or one that matches current rating structures more closely.
- 5** Neural Networks: Neural Networks are inspired by a simple understanding of the way the brain works. A network is trained (using a sample data set) how to discriminate between possible outcomes. The details of the calculations are often obscure, but they can cope with highly non-linear relationships. Discussion in this paper relies heavily on the report of the 1996 working party "Neural Networks v. GLMs in pricing general insurance", chaired by Julian Lowe.

2.3 A discussion on the use of sampling to reduce the size of the data set being analysed is also included. In particular, sampling procedures which take all of the successes and only a portion of the failures potentially offer a way to deal with very low probability events.

Survey results

2.4 The individuals who filled in the survey had a wide variety of experience. As can be seen from the chart, most of the respondents had been working on demand models for a relatively short length of time. Given the size of the survey (just 11 completed sets of questions) it is not possible to be certain that this is representative of the experience of those who are performing this work, but is probably indicative of the recent increase in interest in this topic.



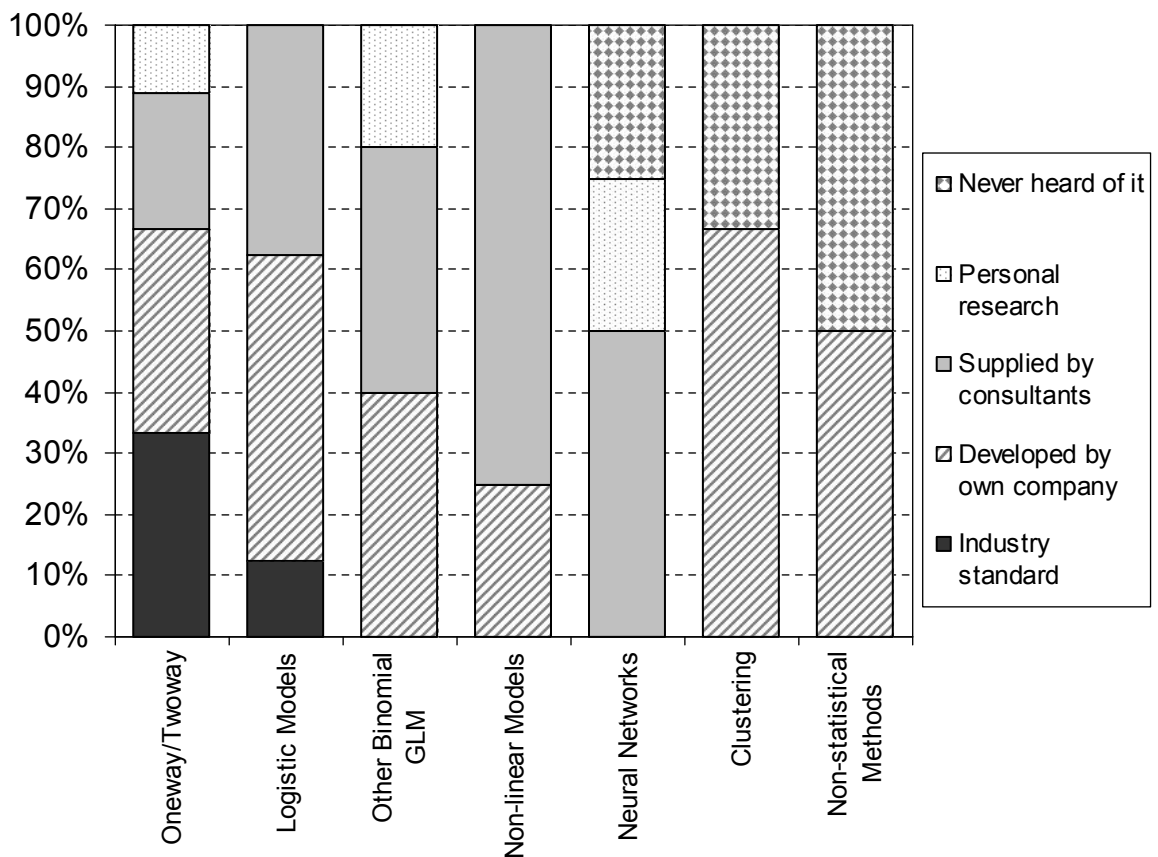
2.5 As might be expected, the usage of demand models is concentrated on the direct channels across all business lines. Given the small sample size, care needs to be taken with the following table, and similar comments apply to the other survey results included in this paper.

	Affinity	Broker	Direct Aggregator	Direct Phone	Direct Web	Don't use	Other*	Number of replies
PL Motor	0.0%	16%	21%	26%	32%	5%	0.0%	19
PL Household	12%	12%	12%	24%	24%	6%	12%	17
PL Other	0%	0%	0%	14%	0%	57%	29%	7
Commercial Lines	0%	13%	0%	13%	13%	63%	0%	8
Other insurance	0%	0%	0%	0%	0%	100%	0%	2

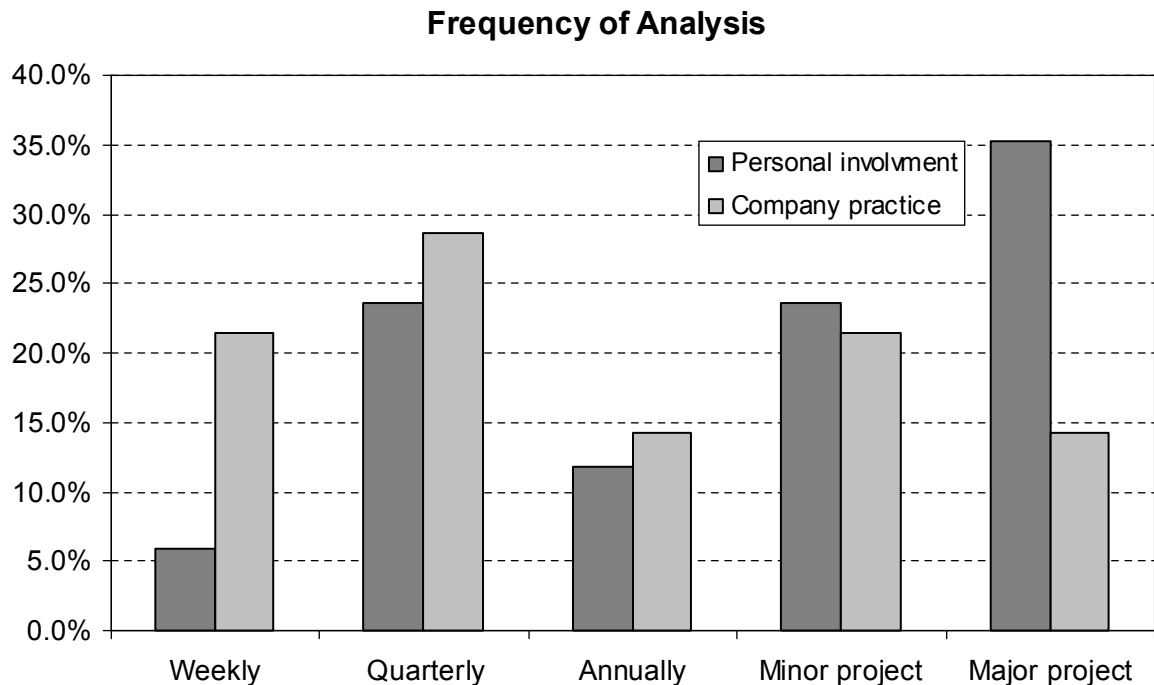
2.6 In the survey we saw that the most commonly used methods are the simple one and two way analyses together with various GLM approaches using a binomial model. In the following table "Regular" means the method is used when this type of analysis is performed, "Rejected" means it has been tried and the practitioner decided not to repeat the analysis using that method, "Considering" means that the respondent is aware of the method but has not yet made a decisions or researched it and "Not interested" means just that.

	Regular	Rejected	Considering	Not interested
Oneway/Two-way	8	1	1	1
Logistic Models	8	1	1	0
Other Binomial GLM	4	0	2	1
Non-linear Models	3	1	1	0
Neural Networks	0	2	2	1
Clustering	1	0	3	0
Non-statistical Methods	1	0	0	3
Other Method	0	0	1	3

2.7 These methods come from a variety of sources, with a healthy amount of research being done within organisations.



2.8 It appears from our survey that demand models are not yet fully integrated into the pricing process, as less than half of the respondent's companies perform regular studies.



3 Method detail

3.1 In this section we look at each of the main types of model in more detail.

One-Way Analysis and Two-Way Analysis

3.2 In one-way analysis the response variable is considered against one explanatory variable at a time, with the analysis repeated for each variable of interest. The explanatory variable may be categorical, continuous or discrete. For example, this may be done by plotting the response against the explanatory variable on a graph for each explanatory variable and examining these graphs by eye. Another method would be to tabulate a certain explanatory variable/response combination and examine the table.

3.3 This analysis ignores all correlations and interactions between explanatory variables (e.g. perhaps younger males are much more reactive to competitors prices and hence have a greater demand elasticity than younger females), and so can only give a simplistic view of any relationships. However, it is a very useful and easy technique and hence is usually used in some form as a starting point of any analysis to help decide which explanatory variables are important.

- 3.4 It may simply help to understand the situation being considered, or may lead to an obvious model, for example, a graph where a straight line appears to fit best would suggest that a linear regression model may be a good starting model. Alternatively, the shape of the graph of best fit may suggest that a transformation is necessary (e.g. to take log of the explanatory variable).
- 3.5 Two-way analysis is the natural progression from one-way analysis and is where we consider the relationship of the response to two explanatory variables. This can be done, for example, using a 3-D graph. This considers every combination of levels of the two variables separately, without any assumption as to the similarity of the patterns across the table.
- 3.6 One-way and two-way analyses can be used to show when things are definitely *not* dependent on an explanatory variable (in this case graphs should show no pattern at all), but if some effect is apparent they do not necessarily show exactly *how* things are related. When carrying out such simple analysis it is important to be careful to consider the relationships between explanatory variables. For instance, the demand may appear to be both positively correlated with age and number of years driving, but in fact, as age and number of years driving are very strongly correlated, it may be that we only need to include one of these explanatory variables in our model.
- 3.7 We could continue, considering three, four, five explanatory variables – but for this we need something more sophisticated than graphs and tables (which is all we need for one/two-way analysis). In addition, at every stage the amount of data in each cell of the resulting n-dimensional table will fall, and in practice the data volumes in many cells will quickly fall below a level at which they are credible.
- 3.8 The advantages of one-way and two-way analysis are that it is quick to do, does not require a lot of data manipulation, requires no assumptions about the data and are simple to understand.
- 3.9 There are a number of disadvantages of this type of analysis. It is only the first stage of the analysis – it requires a suitable model to be selected and to be fitted as a consequence. It can be time consuming to look at many explanatory variables and consider all the necessary combinations. It is also very difficult to apply this analysis to more than two explanatory variables at a time, and so is easily distorted by correlations in the data.

Generalised Linear Models (GLMs)

- 3.10 GLMs can be thought of as a generalisation of ordinary least squares regression. They were developed by John Nelder and Robert Wedderburn in the 1970s. GLMs bring together many statistical models (including linear regression, logistic regression and Poisson regression) into one framework, with a general algorithm for maximum likelihood estimation in all the models. GLMs relate the random distribution of the variable under consideration to the explanatory variables on which this variable depends. These (non-random) explanatory variables are combined using some sort of linear relationship (the *linear predictor*) and linked to the mean of the output variable distribution through a *link function*.

- 3.11** We will not go into the detail of GLMs in this paper, but rather concentrate on some specific attributes of GLMs relevant to demand modelling. For those who want to learn more about GLMs, there is a lot of literature available, and a good starting point is *Generalised Linear Models (2nd Edition) – P.McCullagh and J.A.Nelder (1989) (ISBN 0-412-31760-5)*.
- 3.12** The key assumptions for GLMs are (when predicting the response \underline{Y} using the explanatory variables \mathbf{X}):
- 1** Random component: Each component of \underline{Y} is independent and is from one of the exponential family of distributions:

$$f_x(x; \theta) = h(x) \exp(\eta(\theta)T(x) - A(\theta))$$
 (where $T(x)$, $h(x)$, $\eta(\theta)$, $A(\theta)$ are known functions)
 - 2** Systematic component: The p covariates are combined to give the linear predictor η :

$$\eta = \mathbf{X}\beta$$
 - 3** Link function: The relationship between the random and systematic components is specified via a link function, g , that is differentiable and monotonic such that:

$$E[Y] = \mu = g^{-1}(\eta)$$
- 3.13** This paper will concentrate on the choice of the random distribution and the choice of link function for demand modelling

The choice of random distribution

- 3.14** In demand modelling we are modelling retention and new business probabilities. We need to begin with a sensible distribution choice (Assumption **1**) to represent these. We are modelling a $[0,1]$ response (e.g. whether business is retained or not) and hence a suitable member of the exponential family of distributions is the binomial distribution. In fact, the binomial model is the only distribution with a $[0,1]$ response in the exponential family, although for very small probabilities like the ones we will see in demand modelling, a Poisson approximation to the binomial can be used¹.

The choice of link function

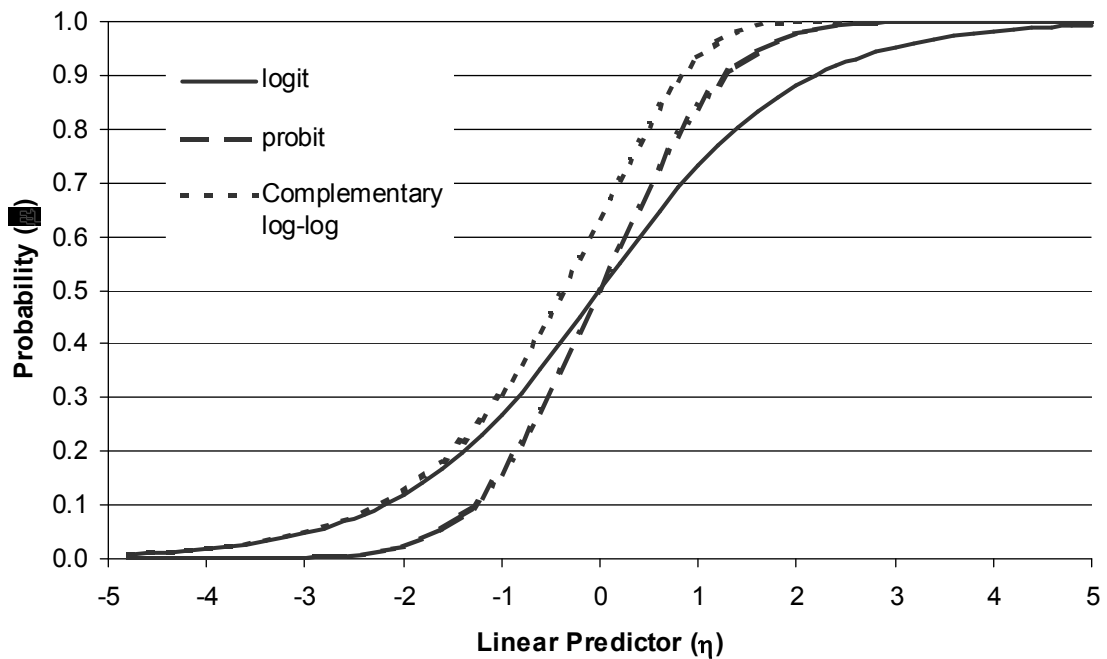
- 3.15** Once a distribution has been selected, it is then necessary to select a link function. Although some may argue that the choice of link function is somewhat arbitrary, there are a number of factors that can make some link function choices better than others. For example, each member of the exponential family has what we call a ‘canonical’ link function that has certain desirable statistical properties (in practice, modern GLM estimation techniques are sufficiently sophisticated that it is no longer necessary limit the choice of distribution function in this way).

¹ This can be seen by considering the probability that a sample from a Poisson distribution exceeds a given number. If $X \sim \text{Poi}(\lambda)$ then $P(X > x) = 1 - e^{-\lambda} \sum_{k=0}^x (\lambda^k/k!)$. This is of order λ^2 for $x = 2$ or more and so tends rapidly to zero for small λ .

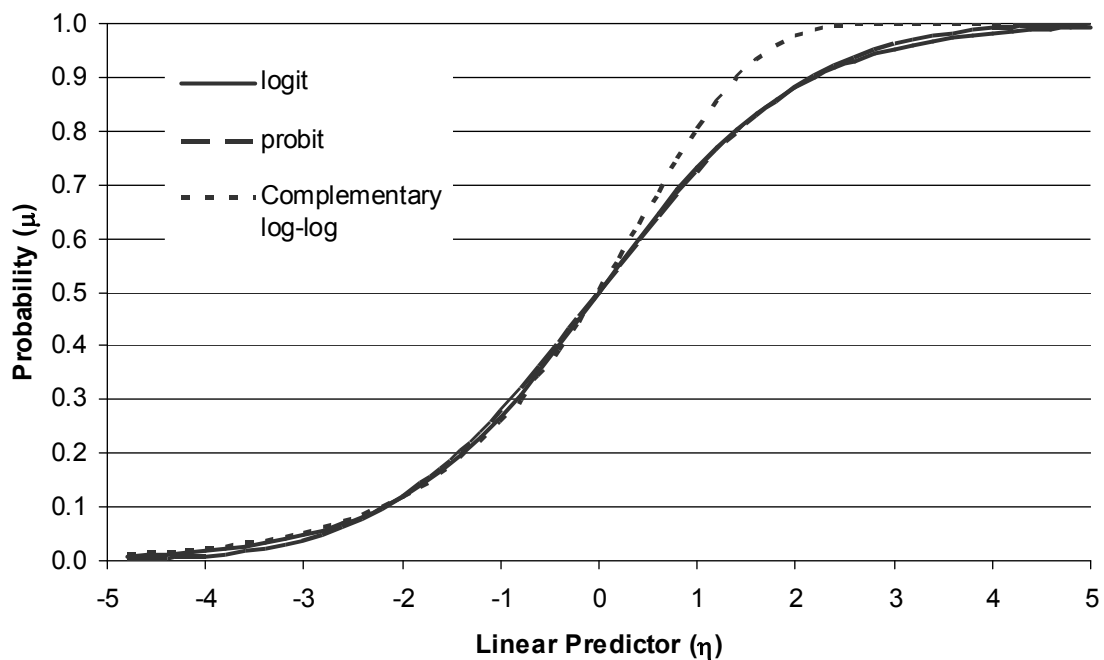
3.16 However, the link function does need to be *differentiable* and *monotonic* (so that a unique inverse can be calculated). It also should ideally map to the correct range (e.g. in this case the (0,1) probability space) and, unless we are going to place some restrictions on \mathbf{X} and $\boldsymbol{\beta}$ then the link function g actually should map (0,1) to (-infinity, +infinity). Common choices for the link function that satisfy these properties are:

- 1 Logit function (the combination of logit link and binomial error is commonly referred to as a logistic model):
 $\eta = g(\mu) = \log(\mu/(1-\mu))$
- 2 Probit or inverse Normal function:
 $\eta = g(\mu) = \text{inv Normal}(\mu)$
- 3 Complementary Log-Log function:
 $\eta = g(\mu) = \log(-\log(1-\mu))$

These are shown in the following graph for comparison.

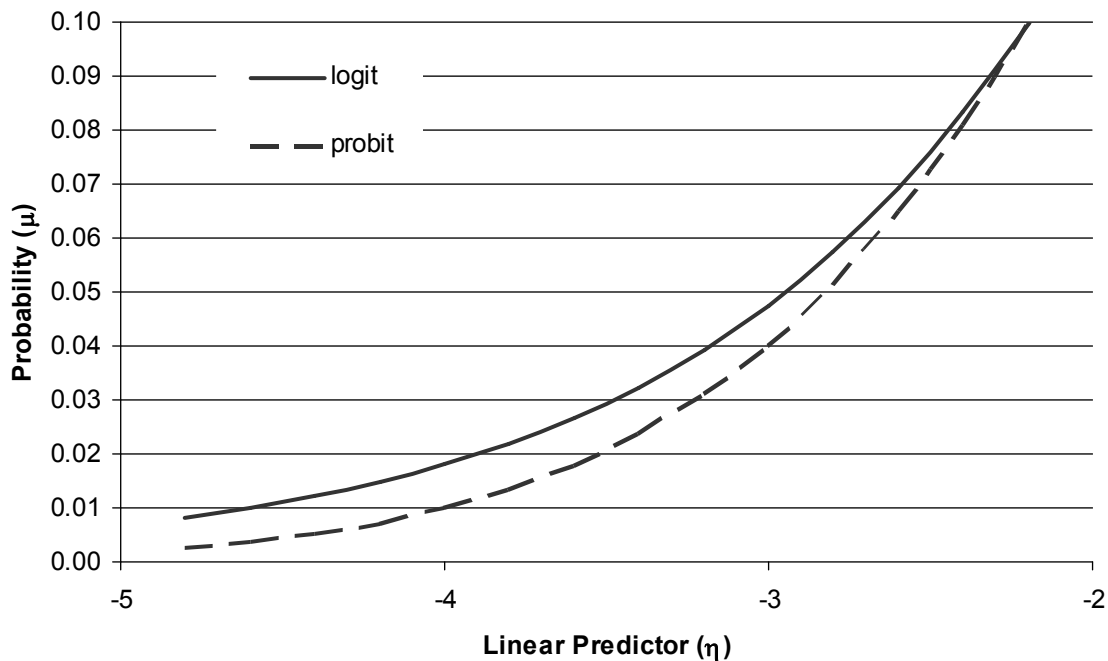


3.17 The differences between these link functions depend upon the size of the probability being modelled. For $0.1 \leq \mu \leq 0.9$ the logit and probit function are almost linearly related and it is usually difficult to discriminate between these on the grounds of goodness of fit. (The *Chambers and Cox (1967)* paper *Discrimination Between Alternative Binary Response Models* investigated this empirically and its findings often quoted – that it is only possible to discriminate between the two models when sample sizes are large and certain extreme patterns are observed in the data). In the following graph the curves have been transformed to match the logit curve at -2 and 0 by applying a simple transformation to the linear predictor. This transformation ($\eta = a \cdot \eta$ for the probit and $\eta = a \cdot \eta + b$ for the complementary log-log) does not affect the linear structure, but only the parameters from the GLM in an easy to understand way.



3.18 It can be seen from this that the logit and probit are indeed very similar over a wide range of values, but that the complementary log-log differs for probabilities larger than 0.5.

3.19 For small μ , complementary log-log is very close to the logit function (and both are close to $\log(\mu)$), but are quite different to the probit function. If we zoom in on small probabilities between 0 and 0.1, and rescale the probit so that it matches the logit at $p=0.1$, we can see that the probit falls faster than the logit.



3.20 All the link functions mentioned have the same asymptotic and approximate theory for the binomial function.

3.21 The most common choice for the link function for modelling retention and new business conversion is the logit link function. There are a number of reasons why this is a suitable choice:

- The logit link maps $(-\infty, +\infty)$ to $(0,1)$ – so no restriction needed on the linear predictor or adaptation needed to the MLE algorithm to find β .
- The logit link is the canonical link for the binomial model. This is theoretically the best choice for the link function in a number of ways, for example it makes the algebra easier, means that the sufficient statistic for β is $X^T Y$ and the variance function is of a prescribed form. However, today GLM software is advanced enough not to be restricted by using the canonical link function.
- Enables reasonably easy interpretation of model in terms of the odds ratio² (although it is arguable how easy the odds ratio itself is to interpret)

² Odds Ratio = $[p/(1-p)]/[q/(1-q)]$ for probabilities p and q . p and q are usually probabilities for the same event (e.g. purchase of policy) for two different population groups. An odds ratio greater than 1 indicates that the event is more likely for the first population group than the second (i.e. $p > q$).

- The logistic link function gives essentially the same results regardless of whether the data is sampled *prospectively* or *retrospectively*. Prospective samples involve selecting sample before event and then waiting to see outcome (e.g. if customer purchases or not). Retrospective samples are taken after event and explanatory variable values at time of event are identified (e.g. looking at the rating characteristics of all those customers who purchased policies vs those that did not). Retrospective design is much more efficient than prospective, especially when dealing with small probabilities. This property is not shared by any of the other link functions.

3.22 This final point is a very important one as means that, for example, in demand modelling investigation, we need to model with less data when we are modelling using a model with the logit link and binomial error. A prospective analysis would involve collecting data on a lot of customers in order to get a large enough sample of those that purchase policies (as the probability of purchase is usually very low), whereas retrospective sampling would mean that we could consider 100% of those customers that purchased together with a selection of those that did not. The analysis can then concentrate on those that purchased a policy rather than all those who were offered a policy.

3.23 So, the logit link function is a good choice. However, there are a number of reasons why we may use an alternative link function. As mentioned above, the logit link is not very intuitive in terms of pure probability. If the probability to be estimated is very small (as is often the case for demand modelling) and the results are to be used qualitatively rather than quantitatively then using the multiplicative Poisson model (using log link and modelling responses as Poisson variables) can produce more intuitive results. This is because in this model explanatory variables have a multiplicative effect on the response (as e.g.

$$\begin{aligned}\log(\mu) &= \eta = B(0) + B(1).x(1) + B(2).x(2) \\ \Rightarrow \mu &= \exp[B(0) + B(1).x(1) + B(2).x(2)] \\ &= \exp[B(0)] \cdot \exp[B(1).x(1)] \cdot \exp[B(2).x(2)]\end{aligned}$$

3.24 The advantages of GLMs are plentiful. GLMs are a good robust method to identify the effects of predictors on a response. It allows understanding of the effect of predictors and quantification of their effects. GLMs allow predictive models to be constructed. It is possible to model over a wide range of error structures using the GLMs framework. With the right software, GLMs are fast and allows a large number of different explanatory variables to be analysed easily.

3.25 However, GLMs also have their disadvantages. Some specialist software is needed: although it is possible to use free software, this generally requires a higher degree of technical knowledge that specialist professional packages. The final selection of variables may be subjective, and there is only a limited number of possible model forms which can be tried. Some link functions can make GLMs hard to interpret.

Generalised Non-Linear Models

3.26 Generalised Non-Linear Models (GNLMs) are an extension of GLMs. GNLMs tend to be those in which the link function uses a non-linear mixture of systematic components. An alternative way to describe this is that the linear predictor need no longer be a linear function. Examples include:

- Mixed additive multiplicative model – linear predictor replaced by a mixture of additive and multiplicative:

$$\eta = \mathbf{X}\cdot\boldsymbol{\beta} + e^{z\lambda}$$

- Alternative mixtures – linear predictor is replaced by a form which is a mixture of different effects:

$$\eta = \mathbf{X}\cdot\boldsymbol{\beta} + C_{\gamma} e^{z\lambda}$$

- Complicating the logit functions – e.g.

$$\eta = 1 / (1 + e^{-\mathbf{X}\cdot\boldsymbol{\beta} + C e^{z\lambda}})$$

- Using geometry functions (particularly useful to get seasonality in the model) – e.g.

$$\eta = 1 + \alpha + \cos(\beta - x(1)) + \sin(\gamma x(2))$$

3.27 Generalised Non-Linear Models can be fitted in a similar way to Generalised Linear Models. As for GLMs the mathematics of GLNMs is sufficiently challenging that an analytic solution is not possible, so a numerical process is used instead, using an iterative process and a design matrix. However, unlike GLMs, the design matrix is dependant on the parameter estimates and so it needs to be updated at the start of each of the iterations. This makes the whole process more complicated as the fitting algorithm needs to be ‘interrupted’ and the design matrix recalculated at each step.

3.28 The advantages of using Non-Linear models are that you can obtain a more accurate predictor model and something that fits the rating structure more closely.

3.29 The disadvantages are that it is more computationally intensive. Also, the extra flexibility of the model means that there is a greater risk of over-fitting the data and interpreting spurious features in fitted curves. Depending on the model fitted, the results can be more difficult to interpret than other models. Another problem is that some of the parameters within the non-linear predictor can be correlated (e.g. the intercept of the additive component and the intercept of the multiplicative component in a mixed additive model) making their estimation difficult – this can be overcome by adjusting the fitting process, but this may have to rely on manual (and hence subjective) selection of some of the parameters.

Neural networks

- 3.30** An excellent, and very readable, paper describing Neural Networks and a comparison with GLM modeling was written by the 1996 working party "Neural Networks v. GLMs in pricing general insurance" which was chaired by Julian Lowe. This can be found at the following link.

http://www.actuaries.org.uk/_data/assets/pdf_file/0003/26580/0417-0438.pdf

- 3.31** The following is in large part a summary extract of that paper.
- 3.32** Neural Networks ("NNs") are computer systems that are often said to operate in a similar fashion to the brain. A NN consists of a series of units called "neurons". These are usually simple processing units, which take one or more inputs and produce an output. Each input to a neuron has an associated "weight" which modifies the strength of the input. When a NN is trained, these weights are adjusted to bring the actual output as close as possible to the expected.
- 3.33** Very basic NNs typically consist of inputs X_1, \dots, X_n , each X_i being either 0 or 1. Each X_i is multiplied by a weight W_i , and the NN outputs either a 0 or a 1 depending on whether the sum was more or less than a certain amount. Neurons whose output depends on some function exceeding a certain amount are known as thresholding output neurons.
- 3.34** Although the example above is very simple, more complicated NNs are just variations on this basic theme. The variations stem from the way the component parts are connected (the NN topology), how the calculations of each neuron are translated to a suitable output (the transfer function) and how the weights are derived.

NN Topologies

- 3.35** A common structure for a NN consists of an input layer, one or more intermediate layers and an output layer. The inputs of a given neuron are fed from the outputs of neurons in the previous layer. Information flows from the input layer, through the hidden layer(s) and finally out through the output layer. This is known as a feed-forward network.
- 3.36** There are many other possible structures for NNs. These can include connections back to previous layers or even back to the neuron itself. NNs for which inputs to a given neuron are taken from outputs from its own or subsequent layers are called feedback networks. A network which has neurons that compete with each other, so that only one neuron responds to a particular input pattern, is known as a competitive network.

NN Transfer Functions

- 3.37** The transfer function is applied to the weighted sum of the inputs of a neuron to translate the inputs to an output. Good candidates for transfer functions are bounded, monotonic, continuous and differentiable everywhere. A commonly used function is the sigmoid function (so called because it is S-shaped).

Setting the Weights

- 3.38** For given input values, the NN looks at the difference between the calculated output and the desired output, and then, starting with the Output layer and working back to the Input layer, adjusts the weights according to their contribution to this difference. Usually all the weights are adjusted together - that is the changes to all the weights are calculated and then implemented simultaneously, rather than changing each weight one at a time.
- 3.39** This process can be more time consuming and produce less intuitive results than other modelling process due to the mechanical nature of the fitting process.

Sampling

- 3.40** One technique that is often discussed in text books is data sampling, and this is mentioned above in the discussion of link functions. Sampling refers to a process where the data is reduced in size by (randomly) using only a portion of the data. An obvious example might be to use only 50% of the records in an analysis. More interestingly, it is also possible to sample at different percentages depending on the y-variate, for example by using all records which resulted in a "success", but only a percentage of the records which resulted in a "failure". We will refer to this as differential sampling.
- 3.41** When faced with large datasets it is often attractive to find a method which will reduce the data volume, but still allow most of the information to be captured. Differential sampling certainly reduces the size of the data, but at what cost? Theory tells us that for some link functions the reduction in size is relatively cost free, but of course care is needed not to damage the predictive power of the final model.
- 3.42** A key element to consider when looking at differential sampling is the purpose of the investigation. In many purely statistical applications the purpose of the model is to explain the observed behaviour. In most cases, differential sampling will still allow the key drivers of a process to be observed, so you can identify a causal link between an explanatory variable and the outcome.
- 3.43** In insurance applications we are often more interested in predicting the outcome in a continuous way – we want the probability of conversion, for example. There are some exceptions, of course. We may be interested in an underwriting model which allocates each potential policy into a "accept" or "reject" pot. However, for the purposes of this paper we are concentrating on predictive models. Here differential sampling is only useful if it is possible to reverse the sampling procedure to return a probability.

4 Practical matters

Data

- 4.1 In any statistical exercise data is critical, and demand modelling is no exception.
- 4.2 Data quality is vital, especially if the demand propensities being modelled are low. One-way analyses should initially be carried out for all data fields to be included in the analysis, in order to identify missing or illegal data. Where illegal data is identified it must be cleaned. If cleaning is not possible then the practitioner should consider exclusion of individual records or exclusion of the entire data field from the analysis, with the relative merits of each being assessed according to the circumstances. Fields with missing data should be judgementally reviewed to identify any bias that may exist between the records with missing data and the outcome for that record. If there is bias and it is not possible to fill in from an alternative data source then the data field should be excluded from analysis.
- 4.3 Required data volumes depend on the method selected. A one-way or two-way analysis requires relatively little data, commensurate with the crudity of the approach. What *is* required is sufficient data in each cell to make the derived statistic significant. On the other hand, for a higher level of simple multivariate analysis the requirement to populate every cell with sufficient data leads to exponential increase in required volume with dimension. This is a major reason for selecting GLMs or other statistical methods when more than two factors interact. GLMs and non-linear approaches reduce the data requirement because any cells for which data is missing or insufficient ‘share’ the characteristics of neighbouring or linked cells.
- 4.4 Customer demand is affected by both generic trends and by changes over time in the direct competition. The validity of demand data will therefore fall off more quickly than data used for claims risk analysis. But this need for up-to-date data must be balanced against the volume required for a statistically significant result. In certain markets a generic demand trend is clear, as information becomes more readily available to customers and conversion rates fall. In such circumstances the results of the analysis must be adjusted for anticipated development of that trend into the time period to which the model will be applied, especially if the analysis is carried out on older data. For example, with GLMs the intercept might be adjusted to give a reduced average conversion rate, whilst maintaining the relativities relating to the explanatory variables.
- 4.5 It is important in demand modelling to capture failure data as well as success data. In some business situations this can be difficult. For example, an insurance company operating in the Broker market via EDI (Electronic Data Interchange) will have details of sales, but may not be passed details of quotations that do not lead to a sale. Predictive factors for customer buying behaviour are likely to be wider in scope than those used to assess claims risk. For example, the existence of a previous relationship with the company may be relevant, or some other personal characteristic that makes a customer more or less price-sensitive. It is necessary to consider all sources of such information that may be available, whether internally or in the market, and to arrange for data merges prior to statistical analysis.

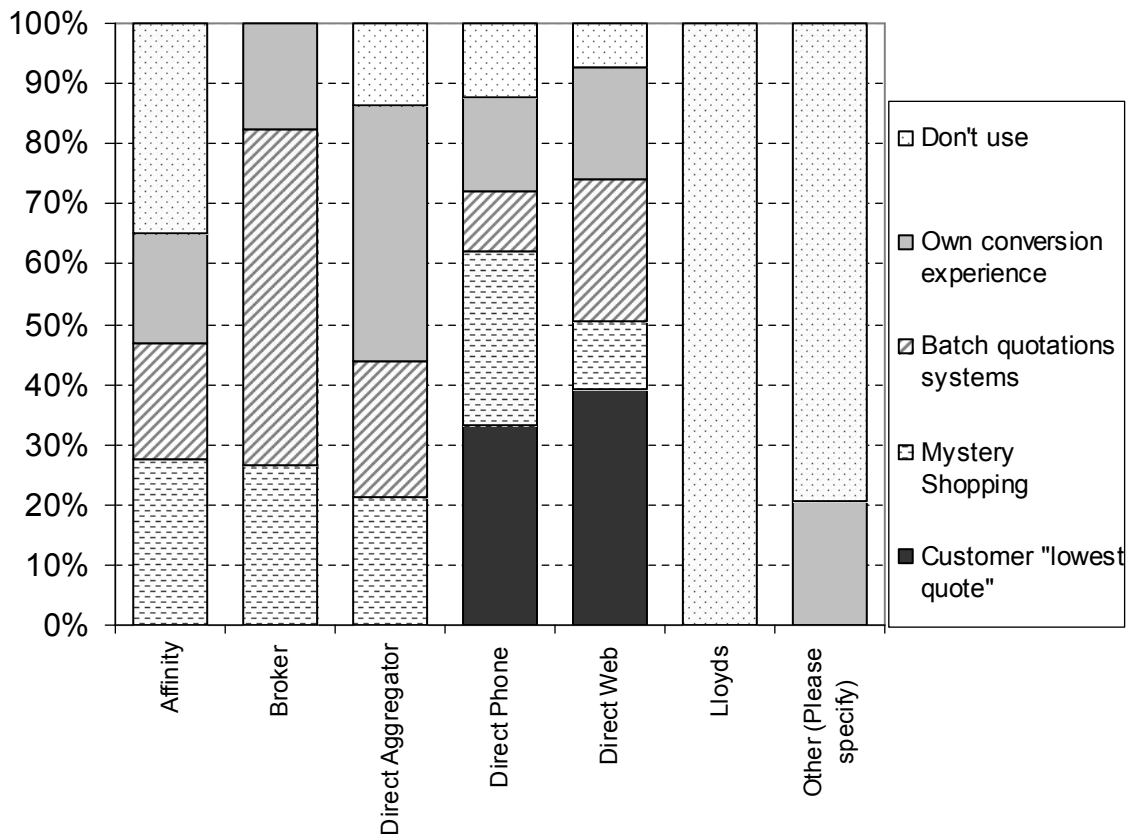
New business vs Renewal

- 4.6** The practical issues affecting new business and renewals modelling are often quite different.
- 4.7** For renewals, the modelling of customer behaviour, particularly in the personal lines market, can be characterised as modelling the *inertia* of the customer. Changes in lapse propensity with price can be interpreted as being partly determined by the price at which inertia fails and the customer seeks alternative quotations in the market, and only then by how that price will fare against those alternative quotations.
- 4.8** Conversely, for new business competition is the key driver, and the factors found to be significant in a demand model are likely to correspond to those policy characteristics where the pricing structure differs from the market. Deviations from the "norm" may not be a concern provided you have confidence in your risk models. By the same token those factors may change significantly if your risk models change.
- 4.9** The concept of price elasticity is to measure percentage change in propensity for percentage change in price, and this is fundamental to demand modelling. For renewal business the premium the customer paid in the previous year is a clear benchmark for the percentage premium change, especially if the inertia effect predominates. For new business on the other hand there may be no readily available equivalent. The most convenient alternative is to measure changes in expected conversion with deviations away from the standard company price and this can be assessed through price testing (random small deviations in price). However, this does not allow for changes in conversion with changes in market competitiveness.

Competition Data

- 4.10** We have already referred to the importance of competition on the conversion demand propensity, especially for new business and in rapidly moving markets. A major challenge is therefore how to allow for such changes in competitiveness and thereby ensure the demand model remains predictive for an acceptable length of time.
- 4.11** This requirement can be filled by the timely inclusion of competition data in the demand model.
- 4.12** The key practical issues with competition data are:
- Is such data readily available for your industry and channel?
 - Is the mix of risks in the data sufficiently broad in scope to mesh with the demand model?
 - Is the currency of the data and the speed of the upload process sufficient to give additional model value?
- 4.13** The practitioner should also be aware of any regulatory or legal implications if obtaining and applying such data were to be deemed anti-competitive.

- 4.14** Sources of competitive data will clearly vary by both market and distribution channel. There are companies in the UK personal lines arena who maintain and offer such information, especially for broker business, and for web-based business it is possible (although time-consuming) to obtain data from screen-scraping (using a program to generate quotations, for example) and other more formal sources.
- 4.15** Where no data is available a proxy for competitive position may be derived from analysis of in-house conversion rates, although in a dynamic market the delay in obtaining mature data and then analysing and processing it may mean that the overall implementation delay exceeds the point at which significant value is added.
- 4.16** The survey asked about the sources used for competitor data. As can be seen, batch quotation systems, where a computer program is used to generate a large number of quotations based on rating structures intended for the broker market, are used nearly universally. The customer lowest quote is unsurprisingly restricted to direct phone and internet channels, as it is typically not available in an intermediary based channel



Application

- 4.17** Demand modelling has a number of practical applications in the effective management of insurance portfolios, and also in the wider business environment.
- 4.18** A key application is in price optimisation. This paper has deliberately avoided exploring demand modelling techniques in price optimisation because of the difficulty of obtaining shared practical experience in a field that is universally regarded as commercially sensitive. However, it is clear that a sophisticated approach to demand modelling is a pre-requisite for any effective price optimisation approach.
- 4.19** A simpler, but related, use is to apply the results of demand models to examine the expected changes to the portfolio. For example, the methods could be applied to simulate the customer response to a proposed change to pricing structure, on either an existing renewal portfolio or on a typical lead profile, and hence derive the expected impact on profitability, volume, premium etc.
- 4.20** Developing this portfolio-modelling concept further leads to the creation of a model office, where all aspects of the financial structure of the current and future portfolio can be combined together to produce a view over the planning horizon, which can be applied to improve strategic decision-making.
- 4.21** With a slightly different focus, accurately modelled demand by customer characteristic is potentially a valuable tool in identifying and applying customer segments for targeted marketing. In this case, knowledge of the price elasticity of customers in any segment can be combined with the absolute levels of profitability and the marketing cost per sale for that segment to produce an optimum marketing strategy.
- 4.22** In all of the practical applications mentioned below, demand modelling is most powerful if combined with a good understanding of the manufacturing cost of the product on offer. In insurance terms this equates to a fairly robust risk premium basis and a good understanding of the fixed and variable expense base of the company, and other marginal costs. This understanding is not required in order to use demand modelling to predict buying behaviour, but it *is* required for a reliable assessment of the impact of such behaviour on the financial wellbeing of the portfolio.
- 4.23** In a wider context demand modelling can be applied to any business environment which satisfies the basic criteria. Those criteria are, a product with known manufacturing cost, a customer base with different propensities to purchase, and sufficient information on those customers as to be able to model and differentiate the factors driving those propensities.

5 Comparative study

- 5.1** The purpose of the comparative study was to determine, on real data, which of the methods considered resulted in the most predictive model. To do this it was decided to segregate the data into two parts, the training data on which the models would be fitted and the testing data on which the predictiveness of the models would be tested. The models were compared using standard data mining techniques to determine which gave the "best" results.

Data

- 5.2** The data used came from a variety of direct sales channels. The data was segregated into two - a "high" conversion data set and a "low" conversion set. A data sample was selected from the experience in early 2007, and the data split by time, so that the training data came from periods before the testing data. This was to ensure that the exercise was a reasonable proxy for actual usage, where models are fitted to past data and applied to future policies.

Modelling

- 5.3** The training data sets were initially modelled using Logistic GLMs. The modelling process was not exhaustive, but reflected models typically used in practice. This process was used to select a fixed set of variables and interactions which was then used in all subsequent models. It is possible that other combinations of variables and interactions would have resulted in a better fitting model for some of the models fitted, but time did not permit us to investigate this possibility.
- 5.4** The full list of models fitted using the common set of variables is:
- Binomial / logit link
 - Binomial / probit link
 - Binomial / complementary log-log link
 - Poisson / log link

Comparison of models

5.5 The testing data was then taken and, for each record, the probability of conversion for that record was calculated based on each of the four methods examined. This allowed the actual experience to be compared with the expected experience in order to determine the goodness of fit.

5.6 Several methods were considered to compare the models, but the final results shown here consider two methods:

- Lift curve: this looks at how good the models are at separating high and low probability segments
- Actual vs Expected: this is a general test of ordering of the fitted values.

Lift curve

5.7 This was defined as:

- The fitted values were placed in 100 pots of equal exposure on the basis of the fitted value. The first pot contained the 1% of the records where the fitted value is the smallest, the second contained the next 1% of exposure and the 100th pot the 1% of records with the highest fitted value.
- For each pot, the observed conversion rate in the pot was found
- The lift curve is a graph showing the pot number (1, 2, ..., 100) on the x-axis and the observed conversion rate on the y-axis.
- It is possible to plot the results from different models on the same graph.

5.8 A good model will be:

- Discriminatory, so that the lift (difference between first and last point) is large
- Diverse, so that each grouping has a different observed probability

5.9 Lift curves provide a good way to compare two or more models, by seeing which model does the best job of separating high and low risk segments, as the equal exposure rule means that each pot will have sufficient data to be credible. However, it does not compare fitted values (it looks only at the order). This means it tests descriptive power and not predictive power.

Actual vs Expected Graph

5.10 This was defined as:

- The fitted values were placed in 100 pots of equal exposure on the basis of the fitted value. The first pot contained the 1% of the records where the fitted value is the smallest, the second contained the next 1% of exposure and the 100th pot the 1% of records with the highest fitted value.
- For each pot, the expected number of conversions was calculated and divided by the observed conversion.
- The A/E curve is a graph showing the pot number (1, 2, ..., 100) on the x-axis and the Actual Conversions/Expected Conversions for each pot on the y-axis.
- It is possible to plot all the curves on the same graph.

5.11 If the model predicted the outcomes perfectly the result of this would be a flat line graph at the point Actual/Expected = 1. It is worth considering however the wide number of factors that could drive differences from this line – apart from the model chosen.

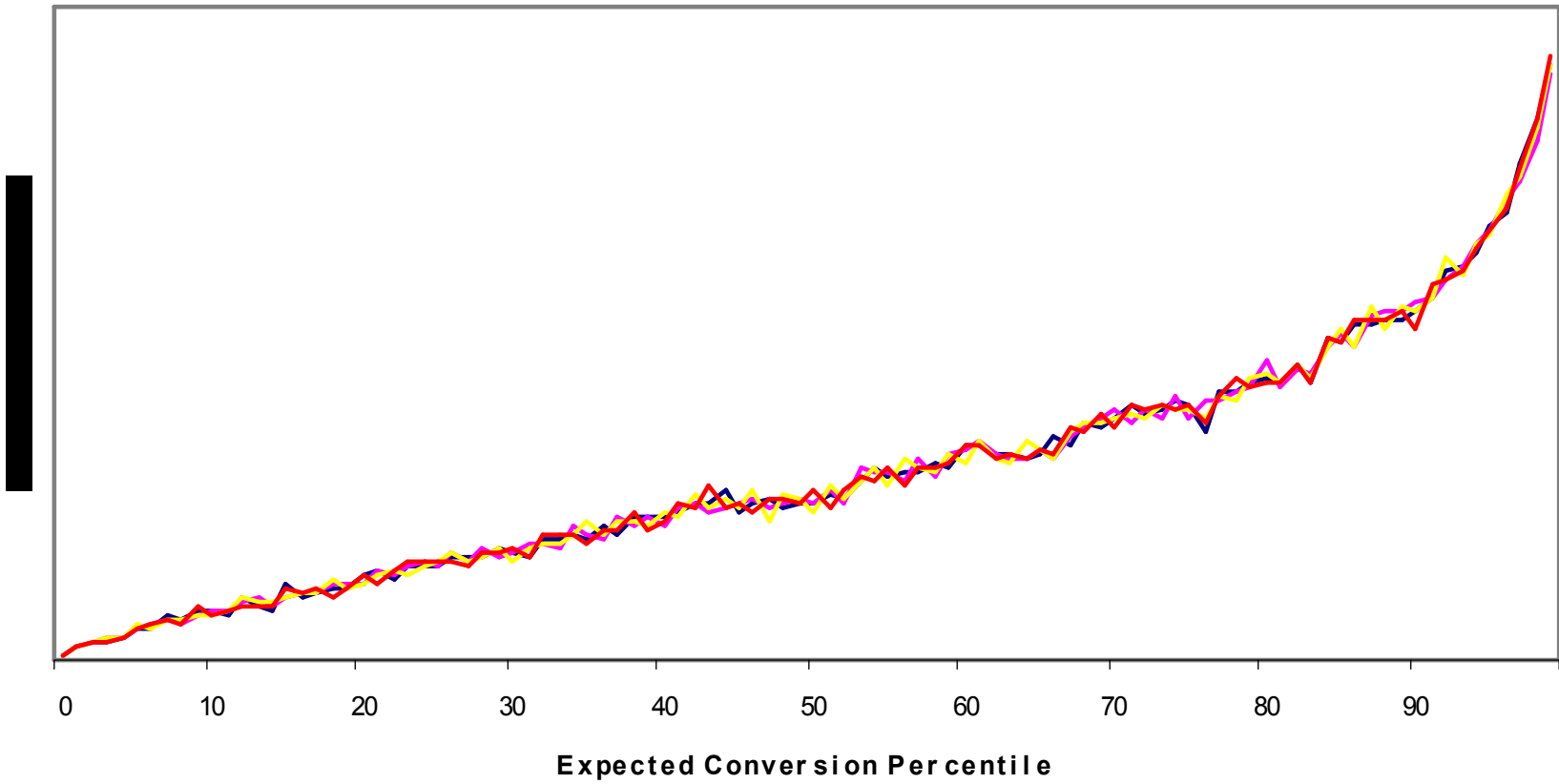
Results

5.12 The results of the analysis are given in four graphs below.

Lift curves

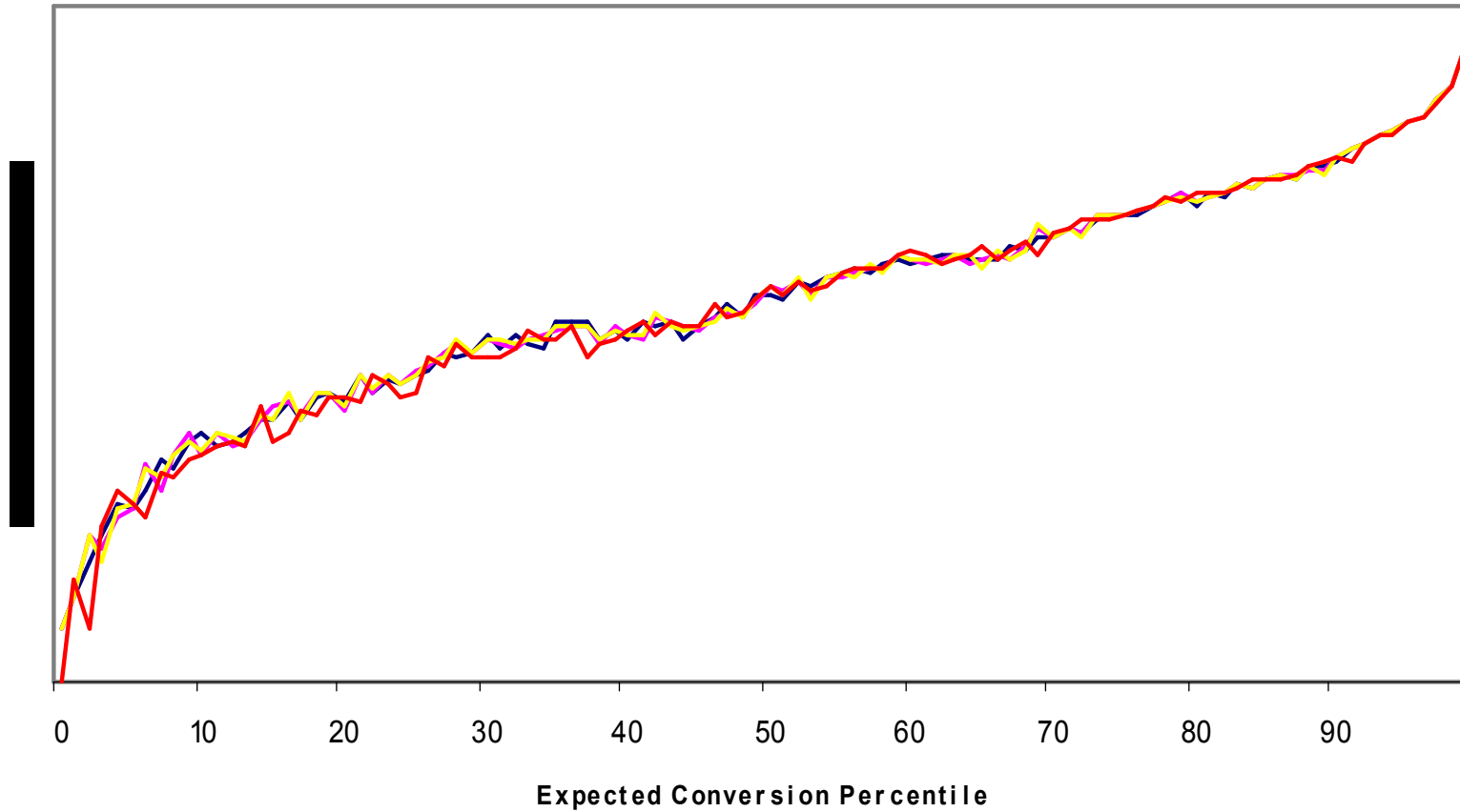
5.13 The first two graphs show the lift curves for the high and low conversion data. For reasons of commercial sensitivity the y-axis scale is not shown. To improve the appearance of the graph, the y-axis of the low probability graph is shown on a log scale. Neither of these comments change the conclusions drawn from the graphic. In both cases the lines lie largely on top of each other, suggesting that the models are all doing a roughly equally good job of separating high and low conversion segments within the modelled database.

Lift Curve - High Conversion Data Set



— LOGIT — POISSON — COMPLL — PROBIT

Lift Curve - Low Conversion Data Set



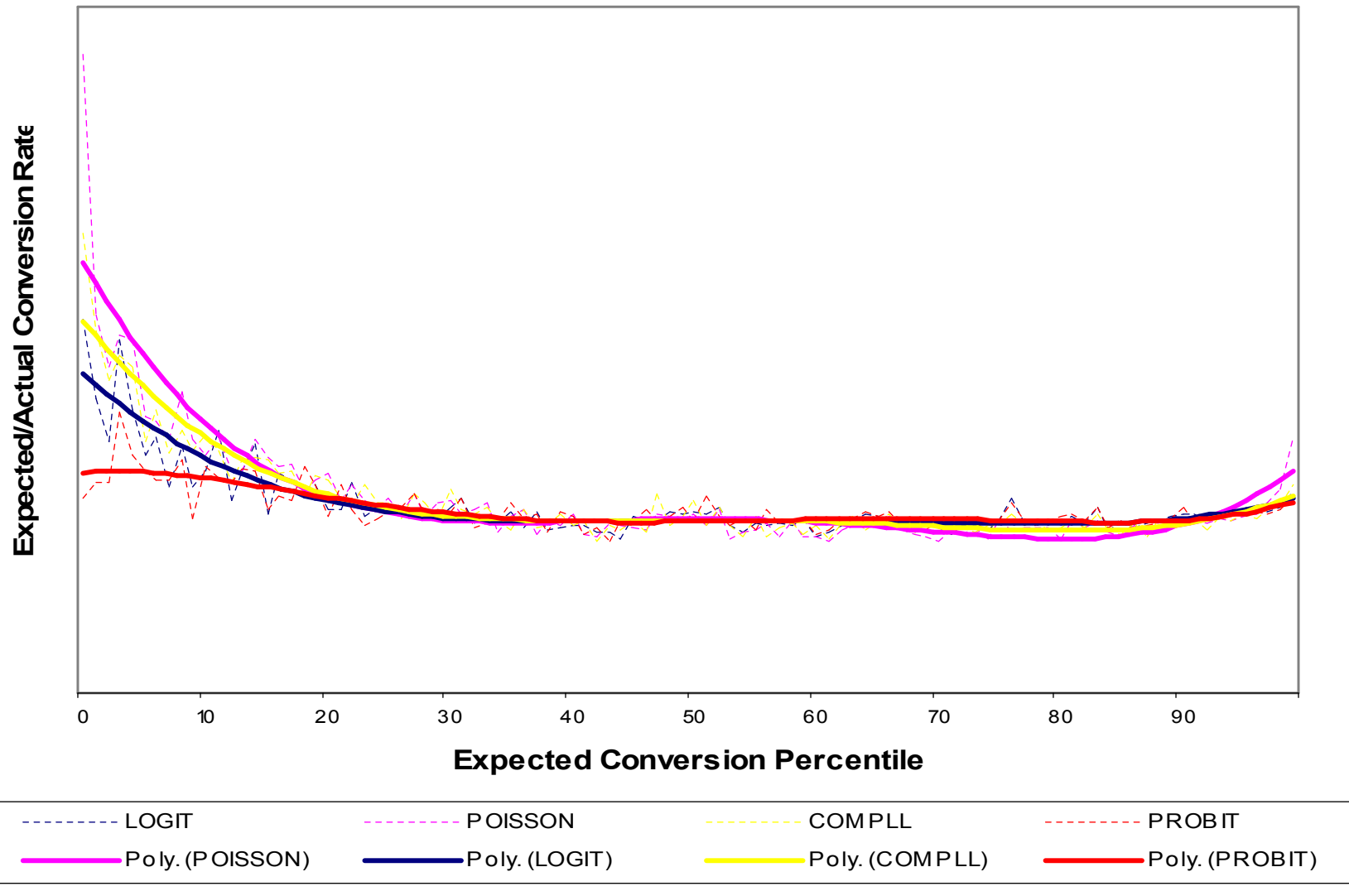
— LOGIT

— POISSON

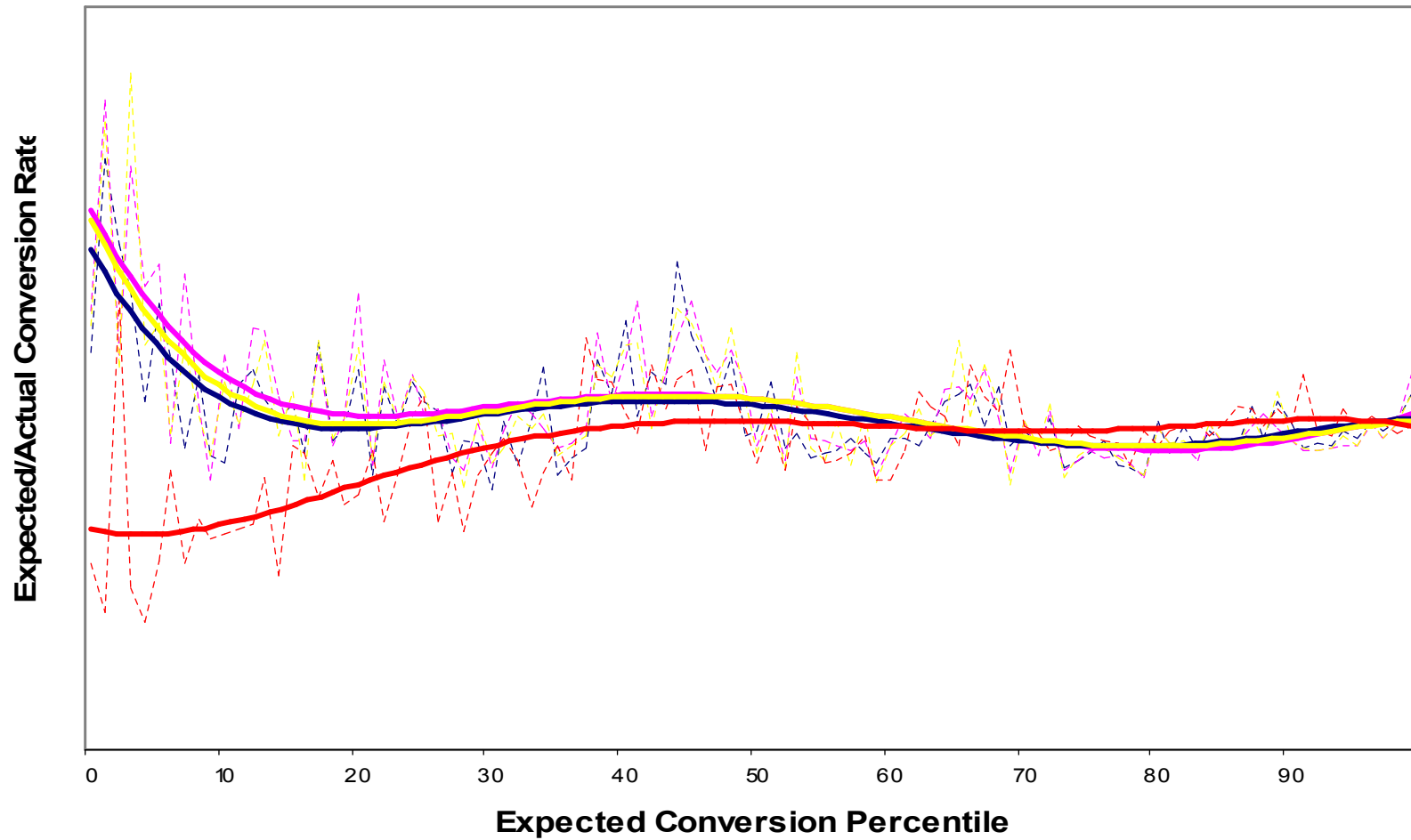
— COMPLL

— PROBIT

Actual Vs Expected - High Conversion Data Set



Actual Vs Expected - Low Conversion Data Set



----- LOGIT	----- POISSON	----- COMPLL	----- PROBIT
———— Poly. (POISSON)	———— Poly. (LOGIT)	———— Poly. (COMPLL)	———— Poly. (PROBIT)

Actual versus Expected

- 5.14** The last two graphs show the results of comparing the actual and expected for the high and low conversion data sets. The dotted lines show the actual experience and the thick lines shown are a smoothed version of the dotted lines included to allow the trends to be more easily identified. Although by definition each data point has the same number of quotations underlying it, the number of conversions increases from left to right. All other things being equal, we might expect more "noise" on the left of the graphs reflecting the lower level of actual conversions, and this pattern is indeed visible in both charts.
- 5.15** Once again, for reasons of commercial sensitivity, the y-axis scale on the graphs has been removed. It is worth commenting that if the overall experience on the testing data was the same as the overall experience on the training data we would expect the average of the observed ratios to be one. In this case, the overall level has changed and hence the average value is not one.
- 5.16** For the high conversion data there is little to separate the four methods considered between the 20th and 90th percentile. The Poisson model does particularly poorly for both high and low conversion rates, in both cases overestimating the probability of conversion. This would tend to reinforce the ideas that models should be selected which match the underlying process. For this data set, the probit model appears to do best at predicting behaviour on the low conversion segments.
- 5.17** For the low conversion data it is harder to draw any clear conclusions. We expect that the log and logit links give ever more similar answers as the probability modelled falls, and this graph indicates that the log, logit and complementary log-log links all result in very similar results. While it is hard to make a definitive statement, it would appear as if these three methods all overstate the conversion rate for the lowest conversion data, a result which was also seen on the high conversion data. The probit link function gives a noticeably different pattern, but it is not necessarily a better fit in this case.
- 5.18** There are some limitations to the methodology used. In particular, it may be that different results would be found if each model was iterated separately, rather than all using variables selected by using a logistic model. It is likely that, if anything, this method would tend to produce results which favour the logistic model over the other models, and it can be seen that the logistic consistently outperforms the log and complementary log-log models. The modelling was necessarily brief, and it is possible that a more exhaustive analysis would have resulted in better fitting models. Further, the results may be reflective of peculiarities in the factor selection or in the data used, rather than of the general case.
- 5.19** Bearing these limitations in mind, it is still notable that the probit link function gives different results from the logit, and could be argued to consistently outperform the logit. This would suggest that if the reader wished to perform their own analysis and was limited in time and resources, the combination of binomial model with the probit link would be the one most likely to produce interesting answers.

6 Summary

- 6.1** Anecdotal evidence suggests that interest in demand models is on the rise within the general insurance industry. There are a host of different methods in use, and most of these were not originally developed with insurance pricing in mind. The aim of the working party was to identify those methods which were both available and useful to analysts working on insurance pricing.
- 6.2** The results suggest that the method most commonly used, the logistic, performs well in most cases. In all cases we found that all the methodologies examined tend to struggle to predict the lowest conversion segments in a given dataset. The probit link function, used with the binomial error, outperformed the logistic model on one of the two datasets examined in this low conversion region. It is hard to draw general conclusions from this result, but this suggests that it may be worthwhile considering the use of this alternate model in applications where the predictiveness of the model over the whole of the range of probabilities is important.
- 6.3** Due to time constraints it was not possible for the working party to repeat the analysis on other data sets, and it was not possible to examine the use of sampling. The members of the working party would welcome further research into this area.

August 2008