Institute and Faculty of Actuaries

Generalized Linear Models for Non-life Pricing - Overlooked Facts and Implications

A Report from GIRO Advanced Pricing Techniques (APT) Working Party

- Ji Yao

August 2013

Abstract

This paper examines in details six overlooked facts of using generalized linear models (GLMs) for non-life pricing and discusses whether the use of GLMs is still fit for purpose in a competitive market. The six facts of GLMs discussed are

1. Either zero or full credibility is given to the data and there is no way to do blending

2. Prediction of a risk depends on data in other completely independent segments

3. Model predictions depend on the mixture of rating factors in the data

4. Maximum likelihood estimate of prediction is lower than mean of prediction distribution

5. Link function could bias the model prediction and significantly change the lower and upper bound of prediction

6. Model diagnostics is only relevant in the segments where the model is used

These facts indicate GLMs' technical limitations in pricing practice, particularly in a competitive market. It is critically important to understand these limitations when apply GLMs in pricing.

Key words

Generalized linear models (GLMs), non-life pricing, credibility theory, dependency, maximum likelihood estimation, link function, model diagnostics,

Content

1 Introduction

1.1 The working party and membership

The GIRO Advanced Pricing Techniques (APT) Working Party was created in Oct 2012 and its membership was growing over time. As at 14th May 2013, members of the working party are

Ji Yao (Chair)
Shamim Ashraf
Zhen Chen
Jacky Cheung
Alberto Chierici
Rebecca Evans
Pulkit Gupta
Peter Jones
Dani Katz
Jeremy Keating
Jing Li
Xavier Lo
Aldo Monteforte
Vickraaz Narrayen
Elena Papadopoulou
Nick Porter
Tim Rourke
Mark Schaffer
Tom Taverner
Simon Yeung
Simon Young

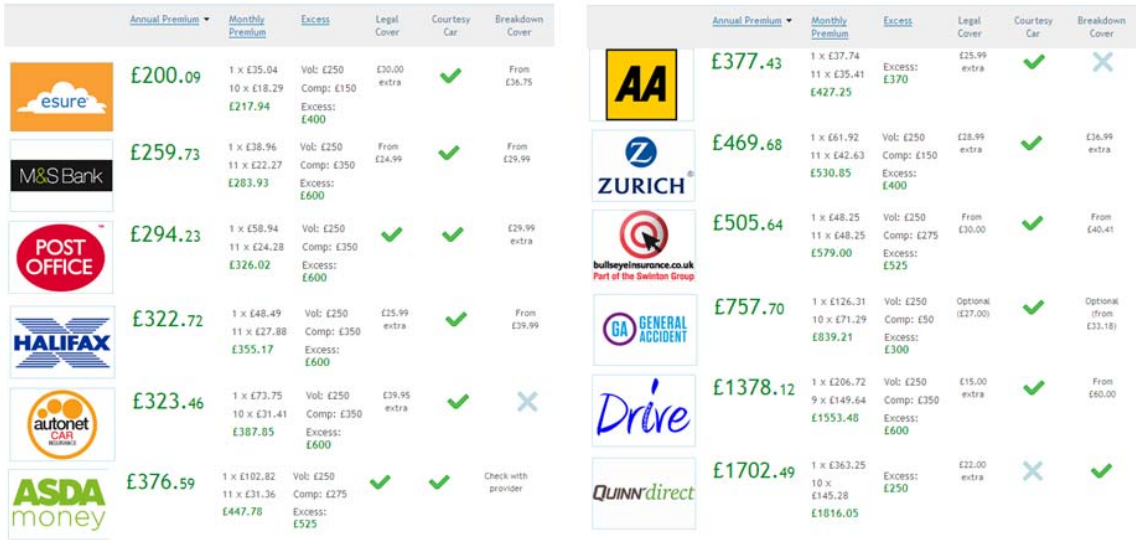
The working party works in three parallel work streams, and this report is an output from one of the work streams. So its findings and views might not represent the view member of the working party.

1.2 Background

Generalized linear models (GLMs) have been widely used as the main pricing technique in the insurance industry for more than a decade in the UK. The approach of using GLMs to set price is well established and standardised [1] [2]. However, the market has changed rapidly recently and in particular price comparison websites have changed the distribution dynamic of the market and have increased competition. The fierce competition in the market raises the question of whether the current pricing technique is still fit for purpose.

Another key feature of insurance pricing is the uncertainty of production costs. Changes in claims procedures, legal systems, regulatory requirements and even pure randomness contribute to the pricing uncertainty. So arguably, motor insurance in the UK is the only market in the world which is close to a perfect competitive market and where the production costs, i.e. costs to pay claim, are hugely uncertain. The combination of competitiveness and uncertainty is a unique challenge, which requires robust pricing techniques.

There are also some evidences in the market that the GLMs pricing technique does not work in certain circumstance. For example, although insurance companies use similar GLMs approach, the key output from models – quoted premiums – are significantly different between companies. Even for an ordinary risk, the quoted premium could range from £200 to £1000+, as shown in Figure 1. There are many other factors that affect the quoted premiums, however, as the technique drives the main component of quoted premium, GLMs worth investigating.



	Annual Premium	Monthly Premium	Excess	Legal Cover	Courtesy Car	Breakdown Cover
esure	£200.09	1 x £35.04 10 x £18.29 £217.94	Vol: £250 Comp: £150 Excess: £400	£30.00 extra	✓	From £36.75
M&S Bank	£259.73	1 x £38.96 11 x £22.27 £283.93	Vol: £250 Comp: £350 Excess: £600	From £24.99	✓	From £29.99
POST OFFICE	£294.23	1 x £58.94 11 x £24.28 £326.02	Vol: £250 Comp: £350 Excess: £600	✓	✓	£29.99 extra
HALIFAX	£322.72	1 x £48.49 11 x £27.88 £355.17	Vol: £250 Comp: £350 Excess: £600	£25.99 extra	✓	From £39.99
autonet CAR INSURANCE	£323.46	1 x £73.75 10 x £31.41 £387.85	Vol: £250 Comp: £350 Excess: £600	£39.95 extra	✓	✗
ASDA money	£376.59	1 x £102.82 11 x £31.36 £447.78	Vol: £250 Comp: £275 Excess: £525	✓	✓	Check with provider
AA	£377.43	1 x £37.74 11 x £35.41 £427.25	Excess: £370	£25.99 extra	✓	✗
ZURICH	£469.68	1 x £61.92 11 x £42.63 £530.85	Vol: £250 Comp: £150 Excess: £400	£28.99 extra	✓	£36.99 extra
bullseyeinsurance.co.uk Part of the Swinton Group	£505.64	1 x £48.25 11 x £48.25 £579.00	Vol: £250 Comp: £275 Excess: £525	From £30.00	✓	From £40.41
GA GENERAL ACCIDENT	£757.70	1 x £126.31 10 x £71.29 £839.21	Vol: £250 Comp: £50 Excess: £300	Optional (£27.00)	✓	Optional (from £33.18)
Drive	£1378.12	1 x £206.72 9 x £149.64 £1553.48	Vol: £250 Comp: £350 Excess: £600	£15.00 extra	✓	From £60.00
QUINN direct	£1702.49	1 x £363.25 10 x £145.28 £1816.05	Excess: £250	£22.00 extra	✗	✓

Figure 1: Quotes for a 40 year old married female with a clean license held for 15 years for a 59 diesel Golf GTD 2.0L 3 door hatchback. Car is kept at home and parked on a driveway for social use only, approx 7000 miles. Quotes were made in early June 2013.

1.3 Objectives

This paper aims to provide evidence for discussion about whether the use of GLMs is still fit for purpose for non-life pricing in the current competitive market. Details of the technical facts relating to GLMs, which have been largely overlooked so far, and their implications will be discussed.

1.4 Disclaimer

While this paper is the product of a GIRO working party, its findings do not represent the official view of the Institute and Faculty of Actuaries, or that of the employers of the working party's members.

1.5 Acknowledgement

A number of people have helped the working party members produce this paper. These include Lee Harris, Jan Iwanik, Chris Pointon and Jenny Xu. The working party would like to thank all of these people for the help and support that they gave to the working party.

A thank you is also made to Rob Barritt for coordinating the working party with the profession.

1.6 Outline

The remainder of the paper proceeds as follows. Section 2 will discuss key feature of the current market, which is dominated by price comparison websites. Section 3 – 8 will examine six facts of GLMs in turn. Each section will start with a discussion of the fact, followed by numerical examples to illustrate the idea, and closed by implications and possible solutions. To simplify the calculation and illustrate the idea clearly, one of key features of this paper is that all numerical examples are based on a simple dataset and can be reproduced easily in SAS. Section 9 will summarise the facts of GLMs and conclude the paper.

Throughout the paper, if a notation is not explicitly defined, the notation in [1] is used.

2 Features of price comparison market

With the introduction of price comparison websites in the UK, the motor insurance market in the UK is essentially very close to a perfect competitive market. Some of its features are:

- Large number of buyers and sellers
- Homogeneous products
- Perfect knowledge about the market
- Absence of transport cost
- No attachment between the buyers and sellers

As a result of these features, the following observations can be made:

- Quoted premium has to be among the cheapest 5-10, although not necessarily the cheapest, in order to convert
- Over-priced quotes are almost never converted. This is different from traditional channels, where over-priced premium can be used to offset losses made from under-priced premium
- Overall conversion rate of a company is very low at around 0.1%, which indicates that
    - Company quotes for the whole market, but converting only a small segment
    - Mixture of sales is different from mixture of quotes
    - Conversion modelling is difficult due to lack of sales
- The huge volume of traffic on price aggregator websites means that sales volume can pick up very quickly. This may happen when
    - An error is made in a company's own pricing structure
    - Major competitors change pricing structure
- Written business can concentrate in small segments where price is very competitive, which creates concentration risk. These segments could be due to a combination of multiple rating factors and therefore hard to be picked up by any management information (MI) reports.

These features and observations mean more sophisticated pricing techniques are required. If over-priced, there will be no sales at all. If under-priced, a lot of loss-making business will be acquired. So the traditional approach to offset under-priced business with over-priced business no longer works. Furthermore, pricing need to be responsive to market change in a specific segment. It is important to bear these features and observations in mind in the following discussion on GLMs.

3 Either zero or full credibility is given to the data and there is no way to do blending

3.1 Discussion of fact

When choosing factors in the GLM, the standard practice is to test all available factors in the data by fitting a factor into the model and using statistical measure to decide whether such a factor is statistically significant or not. If it is, the factor will be kept in the model; otherwise, the factor will be excluded from the model. In making this decision, a significant level is usually arbitrarily chosen. When a factor is considered as significant and kept in the model, full credibility will be given to the estimated parameter of that factor.

However, the confidence level could never be 100%. Even if a confidence level of 99% is chosen, statistical significance means that there is still 1% chance that the factor is the same as the base level. So the estimated parameter should not be given full credibility.

On the other hand, if a factor does not pass the significance test, the estimated parameter still contain some information in the data and therefore should not be discarded completely from the model. This could happen when modelling a small segment, for example left hand drive, or factors with many levels, for example postcode, vehicle type and occupation, as the limited volume reduce the significance of the results.

3.2 Examples

Consider the following problem: there is a bag of blue and red balls, each has a number on it. After sample 6 balls from the bag with 3 for each colour, you are asked to guess the average number on blue and red balls in the bag, respectively. The sample balls are

Table 1. List of colour and number of balls

colour	number
Blue	2
Blue	3
Blue	4
Red	1
Red	2
Red	3

In a standard GLMs approach, after testing the ‘colour’ factor in a GLM (normal distribution with identical link function), it shows that Blue is not significantly different from Red at 95% confidence level and therefore ‘colour’ factor should not be kept in the model. SAS output is shown in Table 2. So the estimated average number of blue and red balls is 2.5, the average of the all observations.

Table 2. SAS output

Parameter	Level1	DF	Estimate	StdErr	Lower WaldCL	Upper WaldCL	ChiSq	ProbChiSq
Intercept		1	2	0.4714	1.0761	2.9239	18	<.0001
colour	Blue	1	1	0.6667	-0.3066	2.3066	2.25	0.1336
colour	Red	0	0	0	0	0	.	.
Scale		1	0.8165	0.2357	0.4637	1.4377	_	_

If another 6 balls with exactly same number are sampled, i.e. 12 balls in total, the same analysis could be done.

Table 3. List of colour and number of balls

colour	number
Blue	2
Blue	2
Blue	3
Blue	3
Blue	4
Blue	4
Red	1
Red	1
Red	2
Red	2
Red	3
Red	3

As shown in Table 4, testing the 'colour' factor in a GLM shows that Blue is significantly different from Red at 95% confidence level and therefore 'colour' factor is kept in the model. So the average number for Red ball is estimated to be 2, and that for Blue ball is 3, which is the average of the observation of each segment.

Table 4. SAS output

Parameter	Level1	DF	Estimate	StdErr	Lower WaldCL	Upper WaldCL	ChiSq	ProbChiSq
Intercept		1	2	0.3333	1.3467	2.6533	36	<.0001
colour	Blue	1	1	0.4714	0.0761	1.9239	4.5	0.0339
colour	Red	0	0	0	0	0	.	.
Scale		1	0.8165	0.1667	0.5473	1.2182	_	_

So GLMs take an approach that using either the total average or the segment average. The prediction has been changed suddenly just because of six extra sample balls. Intuitively, it makes sense to blend the total average and the segment average. Taking a Bayesian view, without any prior knowledge that blue ball is different from red one (i.e. non-informative prior), some credibility should always be given to the total average to take full advantage of all available information. Mathematically, a blended prediction y_i' for the i^{th} segment takes the form of

$$y_i' = w \cdot y_t + (1 - w) \cdot y_i$$

where y_t is the total average, y_i is the i^{th} segment average and w is weight. w could depend on the size of segment, size of total portfolio, volatility of experience etc. A method to calculate w is suggested in Section 3.4.

3.3 Key implications

The blended prediction y_i' is always between the total average y_t and the segment average y_i, in other word, y_t is further away from y_t than y_i'. This 'mean aversion' behaviour means that GLM tends to push prediction toward extreme level, either upward or downward. For example, as in the 12 blue/red balls example in Section 3.2, the blended prediction for blue ball is 2.87 (how this is calculated will be discussed in Section 3.4). If the 'colour' factor is chosen in the model, the prediction will be 3. Similarly for red ball, the blended prediction is 2.13 and the GLM prediction is 2. So GLM prediction tends to be more extreme, i.e. more policies are priced at lower and higher end. Figure 2 shows a comparison of typical distribution of model prediction from two models.

In a competitive market, only cheap quote can be converted to sales and over-price policies never get converted. So GLMs prediction leads to lower overall written premium.

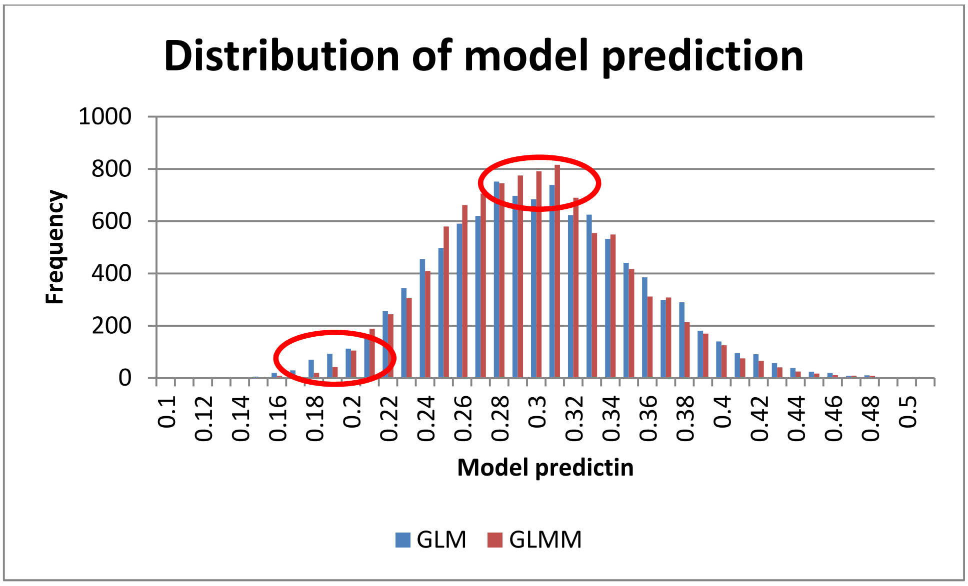


Figure 2. Comparison of typical distribution of model prediction from GLM and GLMM. The mean prediction is 0.3, but GLM has more prediction to the left tail, and GLMM has more prediction around 0.3.

3.4 Possible solutions

It is possible to apply credibility theory with GLM by using standard deviation of estimate [3] [4], but Generalized Linear Mixed Model (GLMM) provides a convenient and standard way to implement it [4] [5]. In this section, GLMM is briefly introduced and the focus is to compare GLMM with GLM in a numerical example to show how GLMM could reduce the volatility of prediction. More technical details of GLMM are available in [5] [6].

GLMM is an extension of GLM to mixed model. The key difference between them is a random effect component in the linear predictor η for GLMM, i.e.

$$\eta = \mathbf{X}\beta + \mathbf{Z}\gamma$$

where $\mathbf{X}$ is a design matrix of fixed-effect predictor variables, β is fixed-effect coefficients, $\mathbf{Z}$ is a design matrix of random effect and γ is random effect coefficients. The random effects are assumed to be normally distributed with mean zero and variance matrix $\mathbf{G}$, that is

$$\gamma \sim \mathrm{N}\left(0, \mathbf{G}\right).$$

By this normality assumption, random effect coefficient estimated is automatically blending with base level.

Other components of GLMM, such as link function and error structure, are same as that of GLM. The major outputs from GLMM are the estimate of β and $\mathbf{G}$. GLMM can also provide prediction $\hat{\mu}_{GLMM,i}$ for the i^{th} row.

One of main measures to quantify the difference between prediction and the true values of the quantity being estimated is mean square error (MSE). MSE can be calculated for every row, i.e.

$$MSE_{GLMM,i} = E\left[\left(\hat{\mu}_{GLMM,i} - \mu_i\right)^2\right].$$

and for the whole portfolio

$$MSE_{GLMM} = \sum_i MSE_{GLMM,i}$$

$$= \sum_i \left(E\left[\left(\hat{\mu}_{GLMM,i} - \mu_i \right)^2 \right] \right)$$

The same calculation could be done for GLM model, MSE_{GLM}, and compared it with MSE_{GLMM}.

3.4.1 Numerical examples

A dataset was created by repeating the data in Table 5 by 50 times, i.e. a dataset with 1000 rows, to illustrate the idea in Section 3.4.

Table 5. Data for GLMM

ID	Age	NCD Yr	Claim 3 Yrs	R_Age	R_NCD	R_Claim	μ_i
1	Young	0	1	2	1	1.5	0.3
2	Young	0	1	2	1	1.5	0.3
3	Young	1	0	2	0.9	1	0.18
4	Young	1	1	2	0.9	1.5	0.27
5	Young	2	1	2	0.8	1.5	0.24
6	Young	2	0	2	0.8	1	0.16
7	Old	0	2	1	1	2	0.2
8	Old	1	1	1	0.9	1.5	0.135
9	Old	2	1	1	0.8	1.5	0.12
10	Old	3	0	1	0.7	1	0.07
11	Old	3	1	1	0.7	1.5	0.105
12	Old	4	0	1	0.5	1	0.05
13	Old	4	1	1	0.5	1.5	0.075
14	Old	5	1	1	0.3	1.5	0.045
15	Old	5	1	1	0.3	1.5	0.045
16	Old	5	0	1	0.3	1	0.03
17	Old	5	0	1	0.3	1	0.03
18	Old	5	0	1	0.3	1	0.03
19	Old	5	0	1	0.3	1	0.03
20	Old	5	0	1	0.3	1	0.03

Column 'Age', 'NCD Yr' and 'Claim 3 Yr' are rating factors. Column 'R_Age', 'R_NCD' and 'R_Claim' are the relativities of these rating factors, and column μ_i is the true claim frequency, which is the production of relativity and a base frequency 0.1.

The following steps were taken to compare the GLM and GLMM predictions:

1) Simulate claim number for each row by Poisson distribution with parameter μ_i;
2) Build GLM and GLMM, and calculate the prediction for each row. In GLMM, all rating factors are modelled as a random effect;
3) Repeat step 1) and 2) 10000 times;
4) Calculate the mean and MSE of prediction for each row;
5) Calculate the total MSE.

GLMM is available in SAS as **`proc glimmix`**. The SAS code to run GLMM is

```
proc glimmix data=data4;
      class age ncd claim;
      model NClaim=/dist=p solution;
      random age ncd claim/solution;
      output out=temp1 pred(ilink)=pred;
run;
```

More details about **proc glimmix** are available in [5].

Figure 3 shows the MSE of each ID and average MSE. Figure shows that GLMM reduce SME significantly for segments where GLM result is highly uncertainty (i.e. high SME) and increase SME slightly for segment where GLM result is stable. And on average SME of GLMM is lower than that of GLM by about 10%.

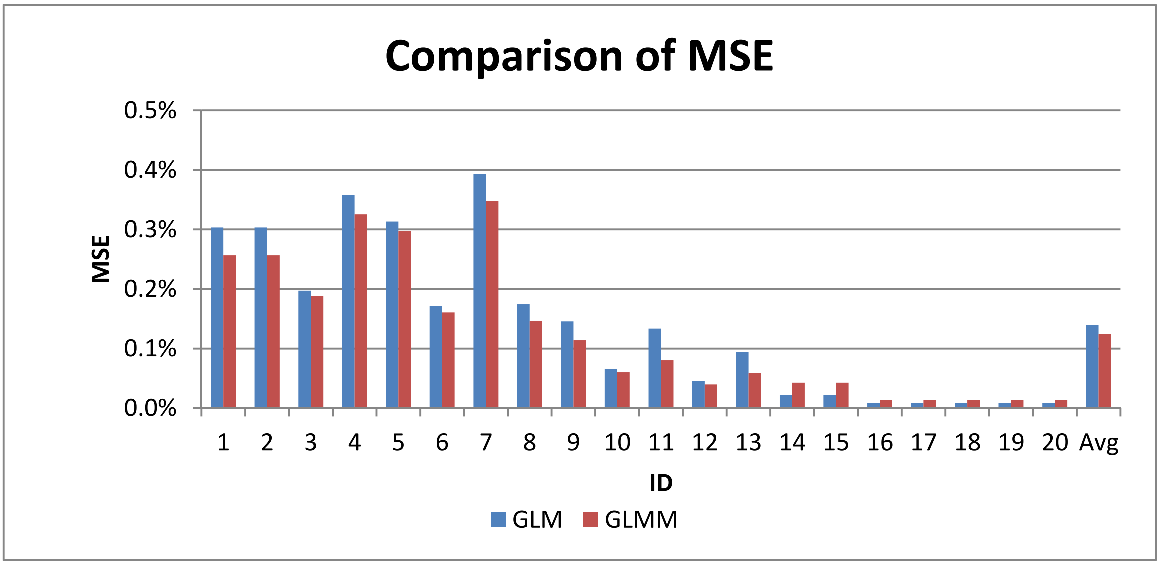


Figure 3. Comparison of MSE between GLM and GLMM.

4 Prediction of a risk depends on data in other completely independent segments

4.1 Discussion of fact

Linear predictor $\eta = \mathbf{X}\boldsymbol{\beta}$ is a key assumption in GLMs. It is assumed that the effect of any factor in the model is independent of other factors in the model (interactive term is considered as one factor). This independency was considered as an advantage of GLMs, because this technique can be used to extrapolate model to segments with limited historical data.

For example, if a company wants to develop into a new market of ‘young driver with low license year’, it can use its historical experience of ‘old driver with low license year’ and a loading of young driver. This ‘young driver’ loading can be derived by comparing ‘young driver with high license year’ with ‘old driver with high license year’.

Table 6. Illustration of segments. The black-out segment has limited historical exposure.

Age\License Year	Low	High
Young		
Old		

This extrapolation is particularly useful and almost essential for current personal line pricing. With dozens of rating factors and each rating factor with 2-100s possible values, it is impossible to have all combination of rating factors in historical data. Pricing has to extrapolate from experience of existing combination of rating factors. Particularly for personal motor insurance, as there are thousands of millions of possible combination, extrapolation is essentially used in pricing for most risks, if not all.

However, a side effect of this extrapolation is that any risk in the portfolio affects the model predictions of other risks in the portfolio, or in other words, the model prediction of any specific risk depends on the whole dataset, not just risk with relevant rating factors. Mathematically, it is

$$\frac{\partial \hat{\mu}_i}{\partial y_j} \neq 0$$

where $\hat{\mu}_i$ is model prediction $\hat{\mu}_i = g^{-1}(\hat{\eta}_i) = g^{-1}\left(X_i \hat{\boldsymbol{\beta}}\right)$ for the i^{th} risk and y_j is the observed data of the j^{th} risk.

It works in a subtle way as explained in the following example. Consider a personal motor portfolio. One of the insured is Mr. Drivelikeaboy, who is a young student living in Great London with a second-hand old car. If Mr. Drivelikeaboy has one more claim and claim experiences of all the others keep same, GLMs will increase the relativities of levels of rating factors that relevant to Mr. Drivelikeaboy, such as ‘age - young’, ‘area – Great London’ and ‘car age – old’. Because claim experiences of all the others keep same, other levels of these rating factors will change (most will decrease), which in turn changes predictions of other risks.

4.2 Examples

Consider a simple regression model on claim frequency using age and car age as explanatory factors. There are four possible type of risk and associated observed claim frequency are shown in Table 7.

Table 7. Initial data

age	carage	claim
Old	Old	0.2
Old	New	0.3
Young	Old	0.4
Young	New	0.7

By using Poisson distribution and log link function, the model prediction is

Table 8. GLM prediction of initial data

age	carage	claim	pred
Old	Old	0.2	0.1875
Old	New	0.3	0.3125
Young	Old	0.4	0.4125
Young	New	0.7	0.6875

If claim of the risk of 'old –old' increase from 0.2 to 0.25, the model prediction will change, which is shown in Table 9. It shows that the model prediction of risk 'Young – New' reduces 0.02083. This risk is completely irrelevant with the risk 'Old – Old', of which the claim experience changes.

Table 9. GLM prediction of updated data

Age	carage	claim	pred	Change
Old	Old	0.25	0.21667	0.02917
Old	New	0.3	0.33333	0.02083
Young	Old	0.4	0.43333	0.02083
Young	New	0.7	0.66667	-0.02083

This is a very simple example of two factors with two levels each, but same result could be observed when the model is much more complicated. Model predictions of almost every risks are subtly affected by claim experience of many irrelevant policies.

4.3 Key implications

While it makes sense to price using similar but not exactly same risk type, e.g. price 'Young – New' using experience of "Young – Old' and 'Old – New' risks in the previous example, it is surprising (at least to customer) that price depends on completely irrelevant risks. This might have some implication to customer and regulator about treating customer fairly (TCF).

To company, the dependency essentially introduce unwanted noise into pricing. Although at portfolio level, underestimation and overestimate might be cancelled out, it would not be cancelled out in written business. This is because customers are more likely to buy when price is underestimated, and overestimated price might never be converted. This is particularly relevant in a competitive market, where price is the key drive of decision.

4.4 Possible solutions

Linear predictor is a basic assumption of linear model, so to solve this problem fundamentally, new model structure, such as non-linear model or credibility approach, need to be explored.

Adding more interactive term might partly solve the problem, as it makes the model prediction less dependent between segments. However, it is difficult to manage many interactive terms in GLMs. Building segmented models is another approach. It is easier to manage the set of models, but it is difficult to find an appropriate segmentation of portfolio.

Another consideration is to understand how strong this dependency $\frac{\partial \hat{\mu}_i}{\partial y_j}$ is. Sensitivity test provide a possible approach. By change the claim experience slightly, sensitivity test helps to find how this affect other segment and by how much.

5 Model predictions depend on the mixture of rating factors in the data

5.1 Discussion of fact

Estimating parameters in GLMs is like a battle between of all data points. Every data point tries to drag the parameter towards the outcome that fits itself best.

If the model structure perfectly reflects all underlying drivers of the data, all data points will share the same 'best' outcome, the true parameters. The mixture of rating factors in data doesn't matter and the model prediction is always the same. The more the data, the closer the estimated parameter to the truth.

However, if the model structure is not perfect, every data point has its own version of 'best' outcome, which is an approximation to the truth. Adding more data into a segment will drag the estimated parameters towards the 'best' outcome represent that segment's 'interest'.

5.2 Examples

Consider another simple regression model on claim frequency using age and car age as explanatory factors. As shown in Table 10, there are four possible type of risk and associated observed claim frequency

Table 10. Initial data

age	carage	claim
Old	Old	0.2
Old	New	0.3
Young	Old	0.4
Young	New	0.6

If we use the perfect model structure of log link function and Poisson distribution, the output parameters of GLMs are shown in Table 11.

Table 11. GLM output of initial data

Parameter	Level1	Estimate	StdErr
Intercept		-0.9163	1.4142
age	Old	-0.6931	1.7321
age	Young	0	0
carage	New	0.4055	1.6667
carage	Old	0	0
Scale		1	0

Adding more data into any of these four segments won't change the parameter estimation, but will only reduce the standard deviance of the estimation. For example, if the 'Old/Old' combination appears twice, the data will become

Table 12. Updated data

age	carage	claim
Old	Old	0.2
Old	Old	0.2
Old	New	0.3
Young	Old	0.4
Young	New	0.6

and the estimated parameters are

Table 13. GLM output of updated data

Parameter	Level1	Estimate	StdErr
Intercept		-0.9163	1.3693
age	Old	-0.6931	1.5811
age	Young	0	0
carage	New	0.4055	1.559
carage	Old	0	0
Scale		1	0

However, if a non-perfect model structure is used, say identity link function and Poisson distribution, the result will be different. Table 14 shows the estimated parameters for initial four data points

Table 14. GLM output of initial data – identity link

Parameter	Level1	Estimate	StdErr
Intercept		0.4286	0.565
age	Old	-0.2411	0.6061
age	Young	0	0
carage	New	0.1339	0.5836
carage	Old	0	0
Scale		1	0

and estimated parameters for updated five data points are

Table 15. GLM output of updated data – identity link

Parameter	Level1	Estimate	StdErr
Intercept		0.4305	0.5594
age	Old	-0.2374	0.5827
age	Young	0	0
carage	New	0.1297	0.552
carage	Old	0	0
Scale		1	0

The predicted values from the two models are different as well. For the first model it is

Table 16. GLM prediction of initial data – identity link

age	carage	claim	prediction
Old	Old	0.2	0.1875
Old	New	0.3	0.32143
Young	Old	0.4	0.42857
Young	New	0.6	0.5625

and for the second mode, it is

Table 17. GLM prediction of updated data – identity link

age	carage	claim	prediction
Old	Old	0.2	0.19315
Old	Old	0.2	0.19315
Old	New	0.3	0.3229
Young	Old	0.4	0.43053
Young	New	0.6	0.56027

Because there are more ‘Old/Old’ combination in the updated data, the model is dragged closer to the ‘Old/Old’ observed claim frequency from 0.1875 to 0.19315. However, the fit deteriorates for the rest of the data.

So this example show that how the model results could be different just because the mixture of rating factor in data is different, and the difference could quite material.

5.3 Key implications and possible solutions

GLMs should be trained on expected future mixture of portfolio, rather than historical portfolio. Current practice is to train GLMs on previous 2-4 years of claim experience with no adjustment to the exposure. So the results that are best fit to the historical portfolio and the large segments in the history. However, because of change of market condition, the sale actually made in the future could be in those smaller segments in history where the model fit is poorer.

An iterative modelling approach can be taken. It might work in following steps:

1) Build GLMs using all historical claim experience data
2) Build conversion model by using GLMs' output and current market condition
3) Score a conversion rate to every historical data
4) Re-build GLMs using all historical claim experience data and the scored conversion rate as a weight
5) If GLMs in step 4) is similar to the model in step 1), then finish, otherwise go to step 2) to update the conversion model by the GLMs in step 4)

6 Maximum likelihood estimate of prediction is lower than mean of prediction distribution

6.1 Discussion of fact

6.1.1 Uncertainty of estimates

Usually GLMs use maximum likelihood (ML) to estimate the parameters β_i in the model. ML gives a point estimation of the parameter, denoted as $\hat{\beta}_i$, which is just one single value. The predicted mean of each row, which is $g^{-1}\left(\mathbf{X}\hat{\boldsymbol{\beta}}\right)$, is also a ML estimate, because of the functional invariant property of ML estimate.

However, other values could still be the true parameter because data observed is just one sample of the true distributions, i.e. there is uncertainty over the ML estimate. Likelihood of the other values can be calculated, which should be lower than that of the ML estimate. All these possible values with associated likelihood will define a probability distribution: the distribution of estimate. The ML estimate $\hat{\beta}_i$ is the mode of the estimator distribution. Except for Normal distribution, generally the mode does not equal to the mean or median of the distribution.

6.1.2 Calculation of distribution of estimate

Bayesian and Frequentist have different way to reflect this uncertainty of parameter. From a Bayesian point of view, the distribution of estimate is essentially a posterior distribution of the parameter. So the standard Bayesian approach could be used to calculate the estimator distribution. However, typically the posterior distribution does not have a close-form solution and need a numerical solution, such as Markov Chain Monte Carlo (MCMC) method. An assumption of prior distribution is also required in this calculation.

From a Frequentist point of view, this uncertainty over estimator is reflected in the standard error or confidence interval of estimate, and the estimator distribution is essentially a sampling distribution, i.e. how much would the estimate change if data is re-sampled? Usually the sampling distribution is an unknown distribution and approximation has to be made to calculate it. Because there is a large sample theory for MLE that the estimate is asymptotically normal distributed, one common assumption is that the estimate is normal distributed with ML estimate of mean and standard deviation.

However, convergence of the sampling distribution to Normal distribution varies significantly. Particularly, when the link function is not identical function, the linear predictor $\mathbf{X}\hat{\boldsymbol{\beta}}$ and the prediction $g^{-1}\left(\mathbf{X}\hat{\boldsymbol{\beta}}\right)$ cannot both be normal distributed. For example, when log link is used, if $\mathbf{X}\hat{\boldsymbol{\beta}}$ is Normal distributed, $g^{-1}\left(\mathbf{X}\hat{\boldsymbol{\beta}}\right)$ will be log-Normal distributed.

6.2 Examples

A dataset of claim number is created to illustrate the idea. It consists of 13 policies of 0 claim, 6 of 1 claim and 1 of 2 claims:

Table 18. Data for calculation of distribution of estimate

# of claims (y_i)	# of policies (n_i)
0	13
1	6
2	1
Total	20

Consider a GLM using Poisson distribution and log link function with only one intercept parameter, i.e. a blank model.

$$Y \sim Poisson\left(\exp(\beta)\right)$$

With this simplification, it is much easier to calculate the Bayesian approach posterior distribution to illustrate the idea.

By definition, the posterior distribution of β is

$$p(\beta \mid \mathbf{y}) \propto p(\beta) \cdot p(\mathbf{y} \mid \beta)$$

By assuming data y_i's are independent to each other and using a non-informative prior distribution, which assigns equal probabilities to all possibilities, the posterior distribution becomes

$$\begin{aligned} p(\beta \mid \mathbf{y}) &\propto \prod_i p(y_i \mid \beta) \\ &\propto \prod_i \exp\left[\beta y_i - \exp(\beta)\right] \\ &= \exp\left\{\sum_i \left[\beta y_i - \exp(\beta)\right]\right\} \\ &= \exp\left\{\beta \sum_i y_i - n \cdot \exp(\beta)\right\} \end{aligned}$$

where the second step use definition of Poisson distribution and ignore the term $1/(y_i!)$ as it is a constant.

The model prediction is calculated through the link function $\mu = g^{-1}(\mathbf{X}\boldsymbol{\beta}) = \exp(\beta)$. Its posterior distribution can be derived as

$$\begin{aligned} p(\mu \mid \mathbf{y}) &\propto \prod_i p(y_i \mid \mu) \\ &\propto \prod_i \exp(y_i \ln \mu - \mu) \\ &= \exp\left[\sum_i (y_i \ln \mu - \mu)\right] \\ &= \exp\left[\ln \mu \sum_i y_i - n\mu\right] \end{aligned}$$

The results are shown in Figure 4 and 5. On both figures, the result of GLM's approximation is also plotted. GLM estimates the mean of parameter β is -.0.9163 and standard error is 0.3536.

Figure 4 shows that the posterior distribution of β shifted toward left compared to Normal approximation in GLM, although the two distributions share the same mode at around -0.9.

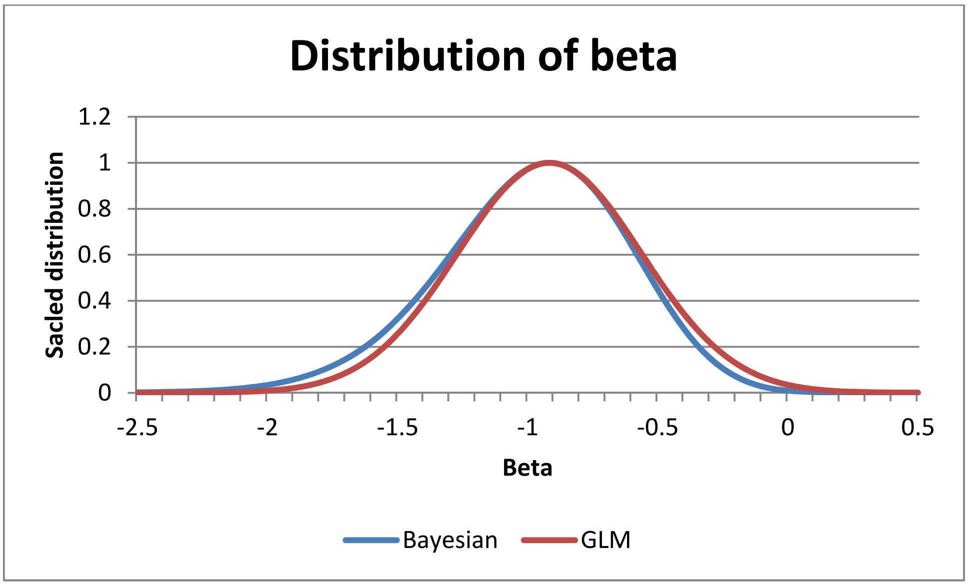


Figure 4. Distribution of β

Figure 5 shows that the posterior distribution of prediction μ shifted toward left compared to log-Normal distribution, and what's more, it is far from Normal approximation. Both posterior distribution and log-Normal distribution has mode at 0.4. However, the mean of the distribution is higher, as show in Table 19.

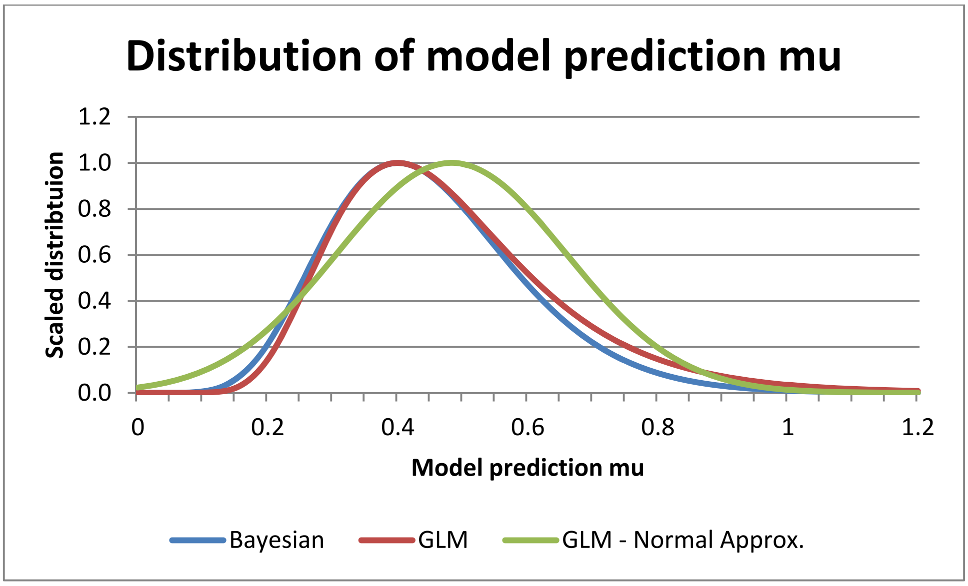


Figure 5. Distribution of μ

Table 19. Mode and mean of distribution of estimate

	Mode	Mean	Mean vs Mode
Posterior	0.4	0.45	12.5%
GLM	0.4	0.483	20.8%

Table 19 shows that the mode, or maximum likelihood estimate, is much lower than the mean of the distribution of estimation. This is mainly because in this illustrative example, the number of data is very low and hence the uncertainty of β is high at 0.3536. Usually in practice we observe that the standard error of linear predictor $\mathbf{X}\beta$ ranges 0.01-0.1 and the more complicated is the model the higher is β.

To give a more realistic scenario, the same 20 policies as in table 18 are repeated 15 times. GLM based on this new dataset has standard error of β of 0.0913. And the distribution of parameter β and prediction μ are plotted in Figure 6 and 7. Figure 6 shows that posterior distribution of parameter β is very close to Normal distribution. However, the distribution of prediction μ in Figure 7 is still not Normally distributed. Table 20 shows that the maximum likelihood estimate is about 1% lower than the mean.

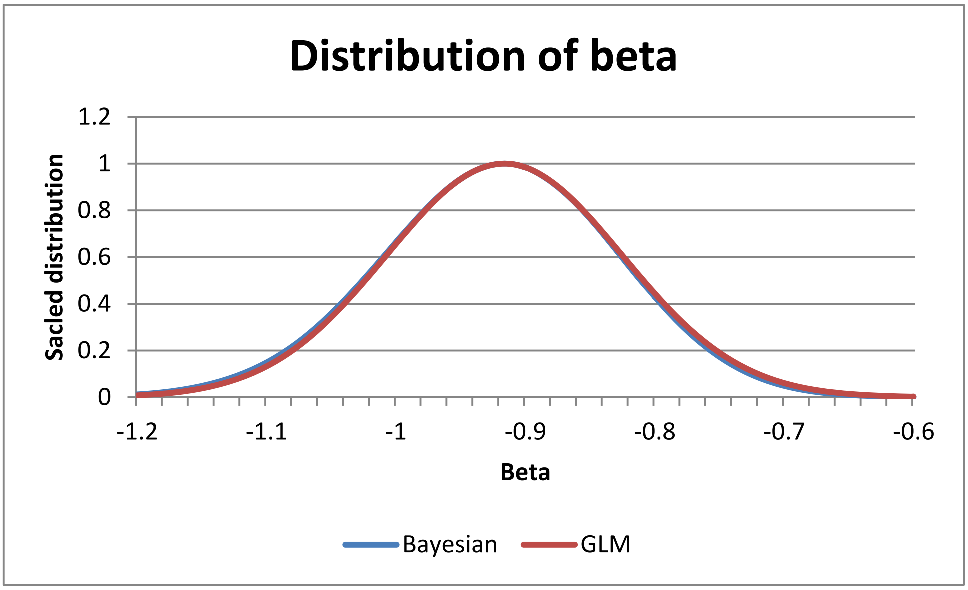


Figure 6. Distribution of β

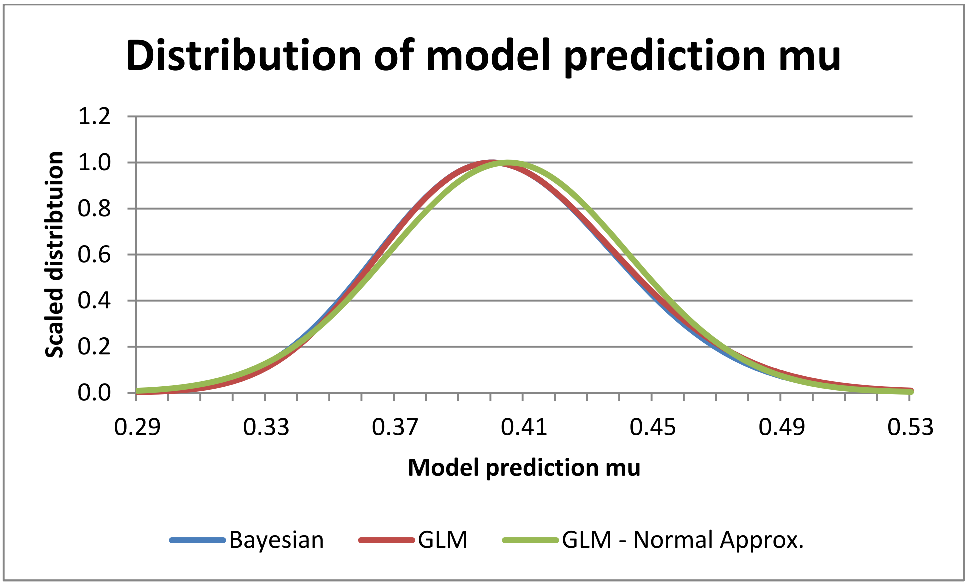


Figure 7. Distribution of μ

Table 20. Mode and mean of distribution of estimate

	Mode	Mean	Mean vs Mode
Posterior	0.4	0.403	0.8%
GLM	0.4	0.405	1.3%

6.3 Key implications

When the distribution of estimate is positively skewed, usually the mean of distribution is higher than mode, which indicates the ML estimate is lower than the probability weighted average of all possible predictions, i.e. the mean. So the current practice to use ML estimate might not be ideal, as it underestimate cost on average.

The extent of underestimation depends on uncertainty in parameter estimation, and ultimately on volume and quality of data. Model has higher uncertainty, for example severity model, large claim frequency model, the impact could be quite big. As a rule of thumb, if the standard error of β is about 0.1, ML estimate underestimate the mean about 1%.

6.4 Possible solutions

The uncertainty of parameter and prediction need to be better understood and evaluated. Although both parameters β and prediction μ are asymptotically Normal distributed, the speeds of convergence are different. The parameters β usually converge faster, so it can be assumed that

$$\beta_i \sim N\left(\hat{\beta}_i, s_i^2\right)$$

and β_i s are mutually independent. This assumption gives that the linear predictor is also Normal distributed with mean $\mathbf{X}\hat{\boldsymbol{\beta}}$ and variance $\mathbf{X}^T\hat{\boldsymbol{\Sigma}}\mathbf{X}$, i.e.

$$\eta = X\beta \sim N\left(\mathbf{X}\hat{\boldsymbol{\beta}}, \mathbf{X}^T\hat{\boldsymbol{\Sigma}}\mathbf{X}\right)$$

where $\hat{\boldsymbol{\Sigma}}$ is covariance matrix of $\boldsymbol{\beta}$. Because it is a Normal distribution, the mode equals mean, so the ML estimate is a good approximation mean.

The uncertainty of μ depends on the link function. But generally there is

$$E\left[\mu\right] = E\left[g^{-1}\left(\hat{\eta}\right)\right] \neq g^{-1}\left(E\left[\hat{\eta}\right]\right).$$

7 Link function could bias the model prediction and significantly change the lower and upper bound of prediction

7.1 Discussion of fact

In using canonical link function, the model prediction always equal to the average in the data. However, this desired feature is not guaranteed when non-canonical link function is used [7].

Usually log link function is used in risk premium with the main aim to simplify the implementation as multiplicative rating structure is preferred. Log is a canonical link function for Poisson distribution but not for Gamma distribution. This explains sometime in the severity model, the modelled prediction does not equal to the historical average in the data exactly. However, based on our previous experience, this difference usually is not big with maximum of 1%. Theoretically reason for this will not be explored further here but more details are in [7].

A more serious issue is the impact on the confidence interval of the prediction. As discussed in Section 6, the uncertainty of parameters and prediction are important element in the model and are asymptotically Normal distributed. Although the parameter β usually converges to Normal distribution quickly, the prediction after link function transfer $g^{-1}(\mathbf{X}\boldsymbol{\beta})$ does not. Because the link function is applied on the linear predictor, it could change the confidence intervals significantly.

7.2 Examples

We build a set of severity model with Gamma distribution using the dummy dataset in Table 21

Table 21. Data for severity model with Gamma distribution

Factor	Value
A	100
A	400
A	500
A	1000
A	10000
B	100
B	200
B	300
B	400
B	500

'Factor' is the only rating factor used in the model, and the claim value is in the 'Value' column. Table 22 – 24 show the results of GLM with different link functions.

Table 22. GLM output using canonical link (Inverse link)

factor	prediction	bottom	top	xbeta	stdbeta
A	2400	1239.5	37656.24	0.000416667	0.00019904
B	300	154.94	4707.03	0.003333333	0.001592318

Table 23. GLM output using log link

factor	prediction	bottom	top	xbeta	stdbeta
A	2400	941.014	6121.06	7.78322	0.4777
B	300	117.627	765.13	5.70378	0.4777

Table 24. GLM output using identical link

factor	prediction	bottom	top	xbeta	stdbeta
A	2399.99	152.968	4647.02	2399.99	1146.46
B	300	19.12	580.88	300	143.31

These examples show that the top and bottom bound could be very different. The canonical link gives a very wide confidence interval. Also the prediction is not necessarily sitting in the middle of the range. Figure 8-10 further illustrate the idea.

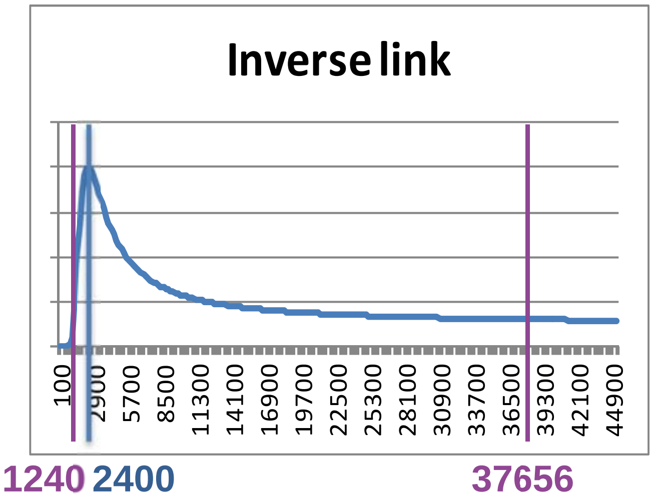


Figure 8. Confidence interval of GLM prediction using canonical link (inverse link)

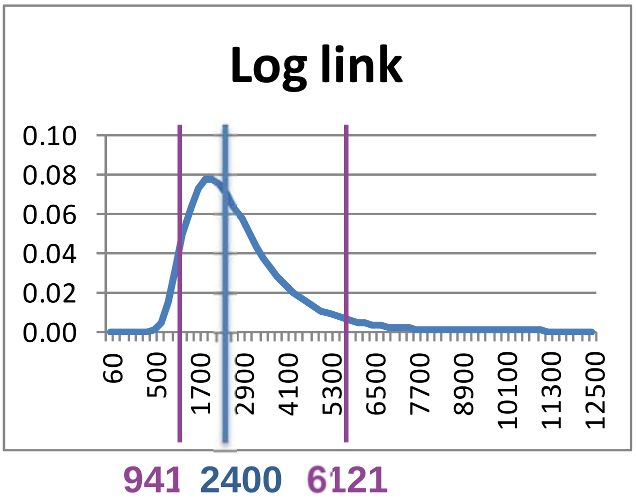


Figure 9. Confidence interval of GLM prediction using log link

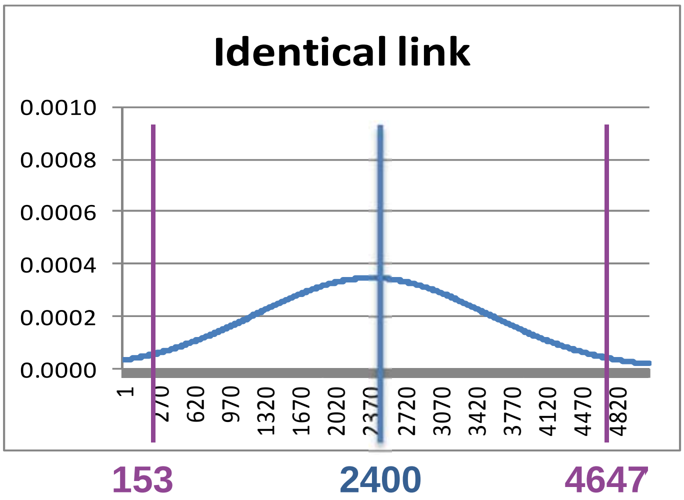


Figure 10. Confidence interval of GLM prediction using identical link

7.3 Key implications

As currently not much focus is put on the uncertainty of prediction, the choice of link function does not affect the results much. However, when the technique in Section 6 is used, different link will give different uncertainty of parameter and predictions.

8 Model diagnostics is only relevant in the segments where the models is used

8.1 Discussion of fact

In current GLMs practice, most of model diagnostic, such as residual check, lift curve and gain curve, are applied to whole modelling dataset. However, in a competitive market, sales are only made in the segments where model under-prices the risks. So only this part of the model needs to be checked and the other segments are irrelevant. The most relevant and important model diagnostic is: if the model under-estimate the cost, how wrong it is?

8.2 Examples

In Figure 11, it compares two models, Model 1 and 2, by showing the ratio of the modelled risk premium to actual risk premium minus one. Because on average the modelled risk premium equals to actual risk premium, the distribution is centred around 0%. As the model is not perfect, there are some risks that are underestimated (i.e. the ratio is less than 0%) and some are over-estimated.

Compare to Model 1, Model 2 has a much higher peak around 0%, which mean it is more likely to get price right. However, Model 2 has a much fat left tail. In a traditional market where the portfolio of sales well spread over the whole distribution, the loss made from the under-estimation segments could be offset by the profit in the over-estimation segments. So Model 2 might perform similar to Model 1 (or might be better, as the peak around 0% is higher)

However, in a competitive market, only the left tail will be converted to sales, which means the Model 2 will write much more loss-making risk.

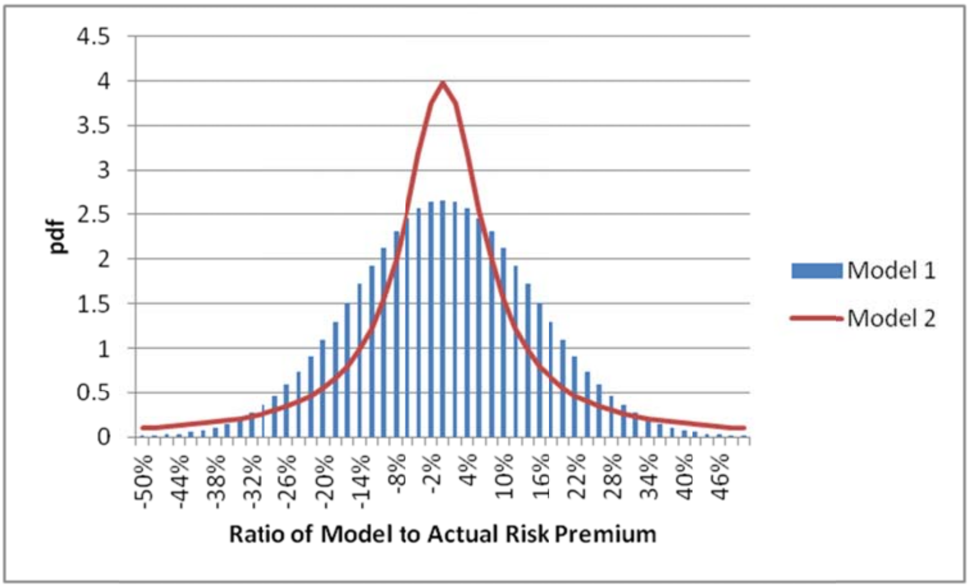


Figure 11. Comparison of model error

8.3 Key implications and possible solutions

Historically on the Phone and Web channel, the portfolio well spread over the whole distribution, and the loss made from the under-estimation segments is offset by the profit in the over-estimation segments.

However, in the current competitive market of aggregator, there are many bidders in the market and the conversion rate could be as low as at 0.1%, which means only small part of the left tail is actually written.

So the current practice of model diagnostics might give an optimistic message of model fitness. With conversion rate as low as 0.1% on price comparison website, only this 0.1% of model really matters. Rather than how good is the model, the question really should be asked is how bad the model is when the model is wrong.

The current model diagnostics methods need to be applied to sales portfolio. However, the sales portfolio will only happen in the future after price is set, so it has to be estimated. So it has to be estimated. One approach is to use the conversion model. So the process to build risk premium GLM might work in following steps:

1) Build GLMs using all historical claim experience data
2) Build conversion model by using GLMs' output and current market condition
3) Score a conversion rate to every historical data
4) Model diagnostic on the historical data with scored conversion rate as weight
5) If the model diagnostic show good fit, then finish, otherwise go to step 1) and use scored conversion rate as weight in GLMs training

9 Conclusion

Six facts of GLMs are discussed in this paper. From a statistical point of view, nothing is really new. However their applications and implications in non-life pricing have not been fully explored before.

Overall, it has been shown that GLMs might underestimate premium in segments where the model is highly uncertain. GLMs also tend to push premium towards extreme level, which creates downside risk in a competitive market. GLMs might also need to be tailored to specific segments that relevant to future sales.

It is pre-mature to conclude whether the use of GLMs is still fit for purpose in current market. However, it is clear that GLMs have some limitations when used for pricing. It is important to understand and communicate these limitations so business management could make informed decision.

Reference

[1] McCullagh, P. and J. A. Nelder, "Generalized Linear Models", 2nd Ed., Chapman & Hall/CRC, 1989

[2] Anderson, D., S. Feldblum, C. Modlin, D. Schirmacher, E. Schirmacher, and N. Thandi. “A practitioner’s guide to generalized linear models.” Casualty Actuarial Society Discussion Paper Program, 2004, 1-116

[3] Garrido, J. and J. Zhou, “Full Credibility with Generalized Linear and Mixed Model”, ASTIN Bulletin, Vol 39 No 1, 61-80.

[4] Klinker, F., “Generalized Linear Mixed Models for Ratemaking: A Means of Introducing Credibility into a Generalized Linear Model Setting”, 2011 Winter CAS E-forum Volume 2.

[5] The Glimmix procedure, SAS/STAT(R) 9.2 User's Guide, Second Edition.

[6] McCulloch, C., “Generalized Linear Mixed Models”, Encyclopaedia of Envirometrics, Sep 2006.

[7] Klinker, F.,”GLM invariants”, 2011 Summer CAS E-Forum.