# Introducing credibility theory into GLMs for Ratemaking on Auto Portfolio.

Serdar Coskun

**Mémoire d'Actuariat**

# Confidentiality

Due to data confidentiality constraint, quantitative information has been intentionally modified in the following study. The modifications have been performed in such a way that they lead to sensible conclusions; yet they do not reflect actual figures.

For similar confidentiality purposes, we will call the insurance company on which we perform the study *Alpha Insurance*, and the country of implementation *Country A*.

# Acknowledgement

I would like to thank Albe Malherbe for supervising me during this thesis. My gratitude also goes to Paul-Andre St-Georges who gave me the opportunity to develop this study, and Olivier Lopez from the Centre d'Etudes Actuarielles.

# Contents

# Introduction

Auto insurance is the most advanced insurance product in term of pricing sophistication. In fact auto insurance covers are mandatory (at least Third-Party Liability), this usually lead to highly competitive environment. In many countries products are extremely segmented. The profit margin is usually low therefore the key to success is to be large and segmented in order to win the competition.

The most common used technique which is now largely accepted in the actuarial toolkit for auto insurance pricing is the use of Generalized Linear Models (GLMs). GLMs are very practical techniques to model insurance risk premium based on multivariate regression. GLMs estimate the price relativities of a number of rating factors in a multiplicative model. Yet they have at least one glaring shortcoming, hence they do not present credible results for segments with low exposure volumes. This is particularly true for categorical rating factors with many levels without an inherent ordering such as car models or geographical regions. For these so-called *multi-level factors* (MLF) we may observe little data on some levels, while there might be much data for others. For example, we may have a significant number of very common car models in our portfolio, but much less of very uncommon ones.

There is therefore a real challenge for a small player which has growth ambitions in a personal auto market, due to his lack of sufficient experience data to build accurate pricing models. While some insurers would investigate competitive premiums to replicate, one could argue that a small player is doomed to make loss until reaching a certain size. One the other side, a very poorly segmented product could lead to severe anti-selection, making harder for the company to be sustainable on a long-term.

The purpose of this study is to try to tackle the problem of low credibility brought by low exposure volume on MLF when performing a GLM, by introducing elements of credibility.

Theory of credibility has been developed in the 20$^{th}$ century, as to estimate how much weight to give to one particular risk segment's experience when performing a pricing review. In today's highly segmented environment GLMs have imposed itself as the main methodology for pricing. In this study, we therefore propose to incorporate credibility modelling within our GLM to combine both models. This application is based on the theory of Ohlsson and Johansson (2003).

In the first part of this thesis, we will present the results and limitations of our loss cost model using the traditional GLM modelling, after reminding the theory around it.

In the second part we will introduce the theory of credibility. We will present linear credibility models and Bayesian models. We will present Jewell's theorem who proposes a framework to make a bridge between the two approaches.

In the third part, we will present Ohlsson and Johansson's work on extending Jewell's theorem, enabling its wider application to insurance ratemaking. Based on this analysis, we will see that we can introduce credibility within the GLM.

In the last part, we will present the results of our GLM revisited on our auto portfolio and compare with our initial model. This will help us assess the improvement and limitations of using this methodology for pricing our auto insurance portfolio.

# Chapter 1: Pricing an auto insurance policy

## Section 1: Context of the study

In this introductory section, we will first present the definition of what personal auto insurance is and what type of covers it implies. We will then present the fundamental idea around insurance ratemaking, and explain what we are trying to achieve in our pricing study. We will then present a descriptive analysis and performance of the portfolio that we need to review.

### 1.1 What is Auto insurance?

Auto insurance is an insurance purchased for car which primary use is to provide financial protection against losses that would arise. The insured agrees to pay the premium and the insurance company agrees to pay the losses as defined in the policy.

Auto insurance is usually split in two main covers:

- Third Part Liability (TPL): liability coverage which pays for the insured legal responsibility to others in case of accident. It can be split into two main perils:
    - Third Party Bodily Injury (TPBI)
    - Third Party Property Damage (TPPD)
- Own Damage (OD): property damage which pays for the losses of the insured own vehicle. The perils can usually be split between Collision, Windscreen, Theft, Fire, Flood, Storm, etc.

In most countries, only the TPL cover is mandatory by law. One cannot go on the traffic without having an insurance to cover losses to a third-party. OD covers are more and more popular especially for recent vehicles. In some countries, insurance companies propose comprehensive covers that combine both TPL and OD covers.

Most insurance would also now include additional covers to protect insured in case of medical expenses or death.

### 1.2 Ratemaking for auto insurance

In a free market, an entity offering a product for sale should try to set a price at which the entity is willing to sell the product and the consumer is willing to purchase it. The simplest model focuses on the idea that the price should reflect the costs associated with the product as well as incorporate an acceptable margin or profit:

Price = Cost + Profit

For many non-insurance goods and services, the production cost is known before the product is sold. However for insurance product the Cost element is not known at the time of selling the product, given the insurance is a promise of indemnity in the future if certain events takes places (ex: damage to vehicle following collision).

Therefore we can tailor this general economic formula to the insurance industry, in which the premium is the “price” of an insurance product. The “cost” of an insurance product is the sum of the losses, claim-related expenses, and other expenses incurred in the acquisition and servicing of policies. Underwriting profit is the difference between income and outgo from underwriting policies:

Premium = Losses + Expenses + UW Profit

The operation that consists in setting insurance prices is referred to as “Ratemaking”. Therefore the goal of Ratemaking is to assure that the fundamental insurance equation (see above) is appropriately balanced. In this paper we will also use the terminology “Pricing” which is similar.

In this paper, we will only focus on the estimation of the expected Loss of the policy, referred as Loss Cost modelling, which is the main area of study of pricing actuaries and statisticians. In fact the estimation of the Expenses is highly dependent on the company’s expense management and resources allocation. The UW Profit assumption depends on the targeted profit from the company and implies larger considerations around capital requirement.

- How to estimate Losses?

The expected loss for each policy will vary depending on the risk insured. For example in our case study, the expected loss of a 20-year-old driving a high value vehicle in a large city is likely to be higher than a 50-year-old driving a small size vehicle in a country side town. Therefore it is important to adapt the level of premium for each specific risk. This means we need to adopt a segmented approach in our Loss Cost modelling.

Segmentation, also referred to as risk selection, is very important in a competitive and free market; hence if the level of segmentation is not as sophisticated as competitors, the insurance company is likely to be subject to anti-selection. This will happen as a result of being underpriced for higher risks, and overpriced for better risks compared to the competitors. Better risks will leave attracted by the competitors’ lower premiums while higher risks will stay given our price is lower; hence this will lead to further deteriorate the results of the company.

In our example, we intuitively understand that the higher expected loss of the younger driver is driven by its higher propensity to have an accident (less experimented and driving in a more dense area), and higher expected loss amount in case of an accident (higher value vehicle which parts will be more costly). In practice the Expected Loss (also known as Pure Premium, Loss Cost or Burning Cost) will be modelled as follow:

Pure Premium = Frequency × Severity

Frequency is a measure of the rate at which claims occur for the specific risk. Severity is a measure of the average cost of claims for the specific risk. Frequency and Severity are normally calculated as follow:

Frequency = Number of Claims / Exposure

Severity = Incurred loss amount / Number of Claims

Please note that for the rest of the analysis, the Exposure measure selected for personal auto is policy-year (one policy coverage over one year equals 1 exposure).

In order to model the Pure Premium, we therefore choose to model the frequency and severity components separately. The modelling based on GLMs will consist in estimating the expected value of the component (frequency or severity) as a linear combination of its covariates also referred to as rating factors (ex: driver age, vehicle make and model, etc.). (See Section 2: Generalized Linear Models). We will finally combine the two models to come up with a pure premium model.

This approach is consistent given the frequency and severity can have different covariates. For example we could potentially conclude that the vehicle make and model does not impact the frequency but only the severity. When possible, we will also prefer modelling each peril separately (TPBI, TPPD, etc.) and combine all the models at the end.

In our case study, we will be reviewing the frequency and severity loss cost models of the TPBI and TPPD perils of our auto insurance product.

### 1.3 Case Study: Pricing review of an auto portfolio

*Alpha Insurance* has been implanted in *Country A* for a few years selling different personal insurance products; however the size of its auto portfolio has remained relatively small due to historical growth strategy focusing on other product lines.

The product offer is exclusively made of TPL, which is split between TPPD and TPBI covers.

The current market share[1] of the company is 3%, on a total market size of €1.0 billion GPW in *Country A*. The following Table 1.1 presents the auto portfolio revenue in Accident Year (AY) 2014.

[1] In term of Gross Premium Written (GPW)

**Table 1.1 – Portfolio revenue in AY 2014**

| | AY 2014 |
|---|---|
| Total exposure | 88,747 |
| GPW (in $m) | 30.0 |
| Market share | 3.0% |
| LR | 70.2% |

The overall LR is at 70.2%. The LRs for TPPD and BI perils are respectively 66.3% and 79.4%. Table 1.2 present the revenue and performance of the book by peril.

**Table 1.2 – Portfolio revenue and performance by peril in AY 2014**

| | TPPD | BI | Total |
|---|---|---|---|
| Earned Exposure | 88,747 | 88,747 | 88,747 |
| GPE (in $m) | 21.0 | 9.0 | 30.0 |
| Average GPE | 237 | 101 | 338 |
| *Premium distrib* | *70%* | *30%* | *100%* |
| Claim frequency | 9.80% | 1.15% | 10.95% |
| Claim severity | 1,600 | 7,000 | 2,167 |
| Pure Premium | 157 | 81 | 237 |
| *Pure Premium distrib* | *66%* | *34%* | *100%* |
| LR | 66.3% | 79.4% | 70.2% |
| Deductible (in $m) | 0.0 | 0.0 | 0.0 |
| Limit (in $m) | 1.0 | 1.0 | 1.0 |

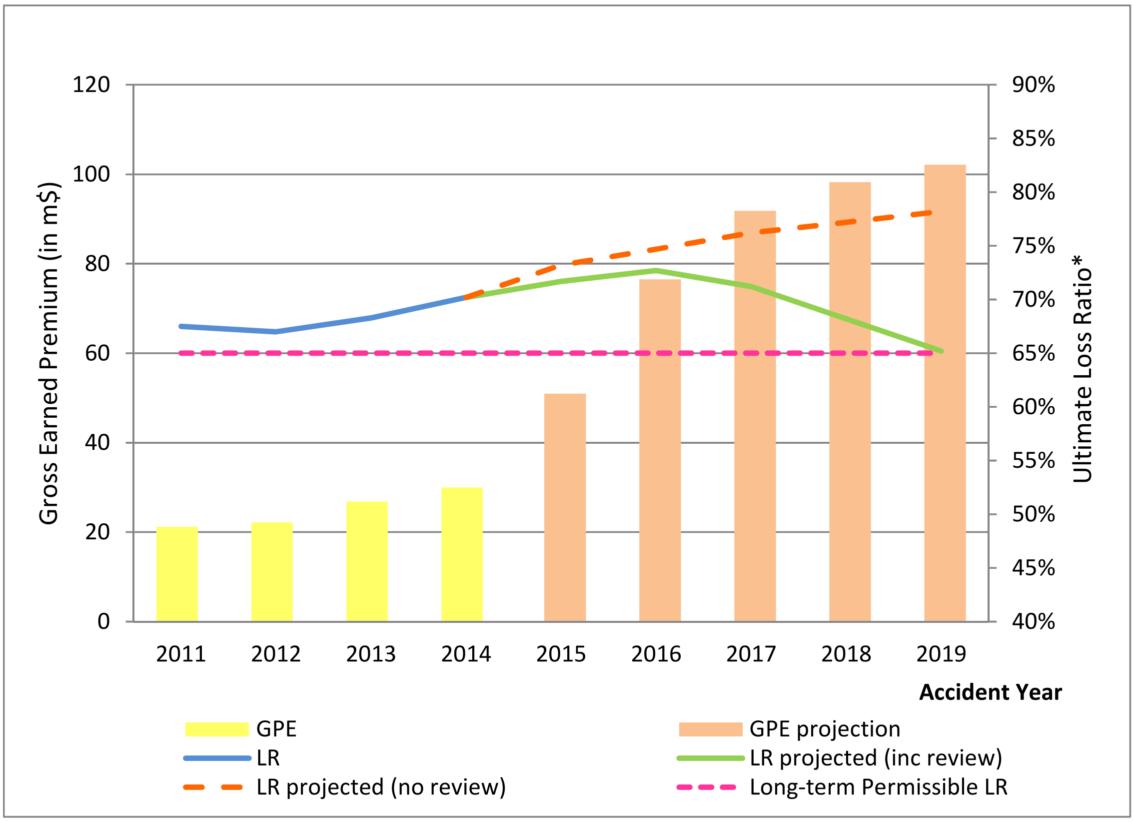
As expected the TPPD frequency is higher than BI which are less frequent claims with higher claim amounts. The TPPD premium represents 70% of the total premium (with BI representing 30%), while the TPPD pure premium represents 66% of the total pure premium (with BI representing 34%). This explains the gap in LR between the two perils.

From 2015, the strategy of *Alpha Insurance* changed with regard to Auto, hence the company decided to grow significantly over the next years in order to become a larger auto insurer in *Country A*, making the auto insurance its pull product for all personal lines products. The company therefore targets to increase its market share from 3% to 10% over the next five years.

This aggressive growth target represents a risk in term of profitability of the portfolio in the short and medium-term. In fact the auto market in *Country A* is a competitive market with big actors; hence the market is already relatively well segmented in term of pricing. One the other hand *Alpha Insurance*'s current auto rater is basic and has not been revised for a few years due to the low focus on the product historically. Therefore a rapid production growth with the current rating structure is likely to lead to severe anti-selection in the market.

For this reason, the company decided to invest in more resources to manage the portfolio profitability including an actuary to review the rates. Graph 1.1 below presents the portfolio volume and loss ratio (LR) over the past accident years, and their projections in the upcoming years with the new growth strategy.

**Graph 1.1 – Projection of the portfolio performance**




** Ultimate Loss Ratio includes IBNR (Incurred But Not Reported) claims.*

We estimate deterioration in the LR over the next two years due to the aggressive growth target. This is due to the competitive disadvantage given the auto rater is not well segmented; hence the portfolio will be facing antiselection. From 2017, the collected volume of data is expected to be enough to perform further pricing analysis to improve the risk selection, hence decreasing the LR. Without any rate review, we would expect the LR to keep increasing regularly over the next years.

Even though the volume of data is limited to reach the market segmentation from year one, there is still room to improve the risk selection in the short-term given the basic current rater. For this reason we will perform a loss cost model on the auto portfolio.

The following table presents the list of available rating factors, and lists the used variables.

**Table 1.3 – Available rating factors**

| Rating variables | Used |
|---|---|
| Age of insured | x |
| Gender of insured | x |
| Insured marital status | |
| Insured driver license age | |
| Conviction in the past 5 years (Yes/No) | x |
| Number of at-fault claims in the past 5 years | |
| Region | x |
| City | |
| Single payment (Yes/No) | |
| Vehicle model year | x |
| Vehicle value | |
| Vehicle make | x |
| Vehicle make-model | |
| Vehicle use (Social only/ Business) | |
| Total no of cars in the household | |
| Additional driver (Yes/No) | |
| Garage (locked garage / street) | |

While there is potential to use additional available variables, one key element to improve the risk selection is to introduce better vehicle risk segmentation. For this reason, *Alfa Insurance* also invested in an external database to add additional vehicle characteristic information. This data has been added via the vehicle make and model. Table 1.4 lists the additional vehicle information.

**Table 1.4 – Additional vehicle information**

| Additional external variables | Used |
|---|---|
| Vehicle type (City / Sport / Saloon / Cabriolet / 4x4) | |
| Vehicle automatic transmission (Yes/No) | |
| Vehicle Imported / Local | |
| Vehicle engine power | |
| Vehicle breaking assistance (Yes/No) | |
| Vehicle driving assistance (Yes/No) | |
| Vehicle no of doors | |
| Vehicle length | |
| Vehicle weight | |
| Alarm (Yes / No) | |
| Tracking device (Yes / No) | |

In order to perform a rate segmentation review of the portfolio, we will perform GLM modelling. This will be performed separately on the BI and TPPD perils; moreover frequency and severity models will be reviewed separately.

The next section presents the theoretical background of GLMs, before presenting the results of the modelling to our auto insurance portfolio.

## Section 2: Generalized Linear Models

As discussed in the previous section, segmented pricing allows the insurer to price individual risks more equitably by analyzing the loss experience of groups of similar risks. This protects the insurer against adverse selection, which can lead to unsatisfactory profits and loss in market share. Effective segmentation may provide insurers with a competitive advantage and may help expand the profile of risks the insurer is willing and able to write profitably.

The multivariate statistical technique that has quickly become the standard for loss cost modelling in many countries and for many lines of business is the Generalized Linear Model (GLM). This technique has been introduced in tashe 70's by Nelder & Wedderburn (1972) and further developed by McCullagh & Nelder (1989). The main benefit of this multivariate approach is that it considers all rating variables simultaneously and automatically adjust for exposure correlations between rating variables, which was the main shortcoming of univariate approaches used in the past.

In this section we will present the theoretical framework around GLMs. To do so, we will first remind the definition of Linear Models (LMs), and present exponential family of distributions used in GLMs. We will then present methods to estimate the quality of the fit and select the variables before validating the model. This section is largely based on "A Practitioner's Guide to Generalized Linear Models" (see Bibliography).

Given this paper is focused on introducing credibility, we will less emphasize the practical aspect of GLMs (which we assume the modeller to be familiar with) to mainly focus on its theoretical framework that will be used later in our revised model.

### 2.1 Linear models

The purpose of linear models (LMs) (such as generalized linear models) is to express the relationship between an observed response variable, $Y$ (for example claim frequency or severity), and a number of covariates (also called predictor variables), $X$. Both models view the observations, $Yi$, as being realizations of the random variable $Y$.

Linear models conceptualize $Y$ as the sum of its mean $\mu$, and a random variable, $\varepsilon$:

$$Y = \mu + \varepsilon \tag{1.1}$$

The model assumes that:

- the expected value of $Y$, $\mu$, can be written as a linear combination of the covariates, $X$.
- the error term, $\varepsilon$, is Normally distributed with mean zero and variance $\sigma^2$.

The linear model seeks to express the observed item $Y$ as a linear combination of a specified selection of predictor variables, plus a random variable $\varepsilon \sim N(0, \sigma^2)$.

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i \times X_\mathrm{i} + \varepsilon \tag{1.2}$$

- $(X_1, X_2, \dots, X_n)$ are the predictor variables, also called covariates.
- $(\beta_0, \beta_1, \dots, \beta_n)$ are the model parameters to estimate.

This model therefore assumes that the observe variable $Y$ is normally distributed with mean $\beta_0 + \sum_{i=1}^{n} \beta_i \times X_\mathrm{i}$ and variance $\sigma^2$.

Covariates correspond to indicator variables which take the value 0 or 1. For example if the covariate $X_1$ represents the gender of the driver, we will have $X_1$=1 if the driver is male and $X_2$=0 if the driver is not a male). In order for the model to be uniquely defined, we should not have any linear dependency between the covariates $X_\mathrm{i}$. For example if $X_1$ is the indicator of the driver being a male, we should not add any other covariate $X_2$ to indicate that the driver is a female; in which case there would be a linear dependency between $X_1$ and $X_2$ such as $X_2 = 1 - X_1$.

We introduce the following matrix notations:

$\underline{Y}$ is the column vector with components corresponding to the observed values for the response variable:

$$\underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_\mathrm{n} \end{pmatrix}$$

$\underline{\mathrm{X_i}}$ are the column vectors of the p covariates with components equal to the observed values for the respective indicator variables (e.g. the $\mathrm{i}^{\mathrm{th}}$ element of $X_1$ is 1 when the $\mathrm{i}^{\mathrm{th}}$ observation is male, and 0 if female).

$$\underline{X_1} = \begin{pmatrix} X_{11} \\ X_{12} \\ \vdots \\ X_{1\mathrm{n}} \end{pmatrix} \qquad \underline{X_2} = \begin{pmatrix} X_{21} \\ X_{22} \\ \vdots \\ X_{2\mathrm{n}} \end{pmatrix} \qquad \cdots \qquad \underline{X_p} = \begin{pmatrix} X_{\mathrm{p}1} \\ X_{\mathrm{p}2} \\ \vdots \\ X_{\mathrm{pn}} \end{pmatrix}$$

$\underline{\beta}$ is the column vector of the p+1 parameters, and $\underline{\varepsilon}$ be the vector of the n residuals:

$$\underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \quad \text{and} \quad \underline{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

To simplify the notation further the vectors $\underline{X_i}$ can be aggregated into a single matrix *X*. This matrix is called the design matrix and in the example above would be defined as:

$$\boldsymbol{X} = \begin{pmatrix} 1 & X_{11} & X_{21} & \cdots & X_{p1} \\ 1 & X_{12} & X_{22} & \cdots & X_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{pn} \end{pmatrix}$$

The system of equations takes the form:

$$\underline{Y} = \mathbf{X}.\underline{\beta} + \underline{\varepsilon} \qquad (1.3)$$

In the case of the linear model, the goal is to find values of the components of $\underline{\beta}$ which minimize the sum of squares of the components of $\underline{\varepsilon}$. If there are *n* observations and *p+1* parameters in the model, $\underline{\varepsilon}$ will have *n* components and $\underline{\beta}$ will have *p+1* components (*p+1*<*n*).

The linear model can be written as follow:

$$\underline{Y} = E[\underline{Y}] + \underline{\varepsilon} \quad \text{and} \quad E[\underline{Y}] = \mathbf{X}.\underline{\beta} \qquad (1.4)$$

McCullagh and Nelder (1989) outline the explicit assumptions as follows:

- *Random component:* Each component of $\underline{Y}$ is independent and is Normally distributed. The mean, $\mu_i$, of each component is allowed to differ, but they all have common variance $\sigma^2$.
- *Systematic component:* The *p* covariates are combined to give the "linear predictor" $\underline{\eta}$:
  $$\underline{\eta} = \mathbf{X}.\underline{\beta} \qquad (1.5)$$
- *Link function:* The relationship between the mean and the systematic components is specified via a link function. In the linear model the link function is equal to the identity function so that:
  $$E[\underline{Y}] = \underline{\mu} = \underline{\eta} \qquad (1.6)$$

Linear models pose quite tractable problems that can be easily solved with well-known linear algebra approaches; however they present few limitations for our modelling purposes. In fact it is difficult to assert Normality and constant variance for response variables. In fact if the response variable is strictly non-negative then intuitively the variance of *Y* tends to zero as the mean of *Y* tends to zero. That is, the variance is a function of the mean. Moreover the values for the response variable may be restricted to be positive. The assumption of Normality violates this restriction.

The use of Generalized Linear Models will enable to tackle those issues. GLMs consist of a wide range of models that include linear models as a special case. The LM restriction assumptions of Normality, constant variance and additivity of effects are removed. Instead, the response variable is assumed to be a member of the exponential family of distributions.

### 2.2 Exponential family of distributions

The exponential family of distributions is a 2-parameter family defined as:

$$f(y_i;\ \theta_i; \phi) = exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + \ c(y_i; \phi)\right\}$$

where *a(φ)*, *b(θ)*, and *c(y, φ)* are specified functions. Conditions imposed on these functions are that:

- $a_i$ is positive and continuous over ℝ.
- $b$ is a function defined over ℝ, twice differentiable with the second derivative a positive function (in particular *b(θ)* is a convex function)
- $c$ is independent of the parameter $\theta_i$.
- $\theta_i$ is a parameter related to the mean.
- $\phi$ is a scale parameter related to the variance.

These three functions are related by the simple fact that *f* must be a probability density function and so it must integrate to 1 over its domain. Different choices for $a_i(\phi)$, $b(\theta_i)$, and $c(y_i; \phi)$ define a different class of distributions and a different solution to the GLM problem. The parameter $\theta$ is called the canonical parameter and $\phi$ the scale parameter. A number of familiar distributions belong to the exponential family among which the Normal, Poisson and Gamma which are of interest for insurance application.

The chart below summarizes those useful distributions that are members of the exponential family:

**Table 1.5 – Main distributions from exponential family**

| | $\boldsymbol{a_i(\phi)}$ | $\boldsymbol{b(\theta_i)}$ | $\boldsymbol{c(y_i; \phi)}$ |
|---|---|---|---|
| *Normal* | $\phi/\ \theta_i$ | ${\theta_i}^2/2$ | $-\frac{1}{2}(\frac{\omega y_i^2}{\phi} + \ \ln(\frac{2\pi\phi}{\omega}))$ |
| *Poisson* | $\phi/\ \theta_i$ | $e^{\theta_i}$ | $-\ln\ y_i!$ |
| *Gamma* | $\phi/\ \theta_i$ | $-\ln(-\theta_i)$ | ${}^{\omega}/_{\phi} \ln\left({}^{\omega y_i}/_{\phi}\right) - \ln(y_i) - \ln\Gamma({}^{\omega}/_{\phi})$ |

It can be seen that the standard choice for $a_i(\phi)$ is

$$a_i(\phi) = \phi / \theta_i. \tag{1.8}$$

A member of the exponential family has the following two properties:

- the distribution is completely specified in terms of its mean and variance.
- the variance of $Y_i$ is a function of its mean.

This second property is emphasized by expressing the variance as:

$$Var(Y_i) = \frac{\phi V(\mu_i)}{\omega_i} \tag{1.9}$$

where V(x), called the variance function, is a specified function; the parameter $\phi$ scales the variance; and $\omega_i$ is a constant that assigns a weight, or credibility, to observation i.

In our analysis, we will only consider a subclass of the exponential family of distributions, the ones that have:

$$V(\mu_i) = \mu_i{}^p \tag{1.10}$$

with $p \geq 0$. This subclass includes the main distributions that are useful in insurance (such as Poisson and Gamma).

- *Prior weights:*

The prior weights allow information about the known credibility of each observation to be incorporated in the model. For example, if modelling claims frequency, one observation might relate to one month's exposure, and another to one year's exposure. There is more information and less variability in the observation relating to the longer exposure period, and this can be incorporated in the model by defining $\omega_i$ to be the exposure of each observation. In this way observations with higher exposure are deemed to have lower variance, and the model will consequently be more influenced by these observations.

- *Scale parameter:*

In some cases (e.g. the Poisson distribution) the scale parameter $\phi$ is identically equal to 1 and falls out of the GLM analysis entirely (which is the case in our case study). However in general and for the other familiar exponential distributions $\phi$ is not known in advance, and in these cases it must be estimated from the data.

For further details around how to estimate the scale parameter please refer to **Appendix I.1**.

The corresponding value of the variance function is summarized in the table below:

**Table 1.6 – Variance function examples**

| | *p* | *V(x)* |
|---|---|---|
| *Normal* | 0 | 1 |
| *Poisson* | 1 | $x$ |
| *Gamma* | 2 | $x^2$ |

For each observation $Y_i$, we assume a distribution defined as (1.7). Thus each observation has a different canonical parameter $\theta_i$ but the scale parameter $\phi$ is the same across all observations. It is further assumed that the functions $a(\phi)$, $b(\theta)$ and $c(y,\phi)$ are the same for all $i$. So each observation comes from the same class within the exponential family, but allowing $\theta$ to vary corresponds to allowing the mean of each observation to vary.

Moreover the parameters $\theta_i$ and $\phi$ encapsulate the mean and variance information about $Y_i$. It can be shown that for this family of distributions:

$$\mu_i = E(Y_i) = b'(\theta_i) \tag{1.11}$$

$$Var(Y_i) = b''(\theta_i).a(\phi) \tag{1.12}$$

With $b'$ the derivation of b is with respect to $\theta_i$.

For the demonstration of the relations (1.11) and (1.12) please refer to **Appendix I.2**.

(1.11) shows that the canonical parameter is essentially equivalent to the mean. (1.12) can be interpreted as establishing that the variance is a function of the mean times some scaling parameter $a(\phi)$. This is in line with the relation (1.9) that we've already seen above.

The chart below expresses the mean, the variance function and the scale parameters for the respective distributions:

**Table 1.7 – Parameters for main distributions from exponential family**

| | ***Notation*** | $\boldsymbol{\phi}$ | $\mu(\boldsymbol{\theta})$ | **V(μ)** |
|---|---|---|---|---|
| *Normal* | $N(\mu,\sigma^2)$ | $\sigma^2$ | $\theta$ | 1 |
| *Poisson* | $P(\mu)$ | 1 | $e^{\theta}$ | μ |
| *Gamma* | G(μ,ν) | $\nu^{-1}$ | $-1/\theta$ | $\mu^2$ |

Now we have defined the exponential family of distributions, we will present how these are useful in the GLMs.

## 2.3 Generalized Linear Models – theoretical framework

In LMs, the response variable is assumed to follow a Normal distribution, with constant variance for each observation. In GLMs, these limitations are removed. Instead the response variable is assumed to be a member of the exponential family of distributions. Therefore the variance is permitted to vary with the mean. We can outline the following assumptions for GLMs:

- *Random component:* Each component of $\underline{Y}$ is independent and is from one of the exponential family of distributions.
- *Systematic component:* The $p$ covariates are combined to give the "linear predictor" $\underline{\eta}$:
  $$\underline{\eta} = \boldsymbol{X}.\underline{\beta} \qquad (1.5)$$
- *Link function:* The relationship between the mean and systematic components is specified via a link function, $g$, that is differentiable and monotonic such that:

$$E[Y] = \underline{\mu} = g^{-1}(\underline{\eta}) \qquad (1.13)$$

This formulation is equivalent to say that: $g(\underline{\mu}) = \boldsymbol{X}.\underline{\beta}$

The choice of the link function is therefore very important depending on the type of response variable we want to model. The following table present the typical model form used in insurance pricing:

**Table 1.8 – Link function and error structure examples**

| $\underline{Y}$ | Claim frequency | Average claim amount | Probability of renewing |
|---|---|---|---|
| *Link function* | $\ln(x)$ | $\ln(x)$ | $\ln(x/(1-x))$ |
| *Error* | *Poisson* | *Gamma* | *Binomial* |

Once we have defined the model form (response variable, covariates, link function and error structure), we need to estimate the components of $\underline{\beta}$. To do so, we will maximize the likelihood function. By definition, this method seeks to find the parameters which, when applied to the assumed model form, produce the observed data with the highest probability.

The likelihood is defined to be the product of probabilities of observing each value of the y-variate. For continuous distributions such as the Normal and gamma distributions the probability density function is used in place of the probability. It is usual to consider the log of the likelihood since being a summation across observations rather than a product, this yields more manageable calculations (and any maximum of the likelihood is also a maximum of the log-likelihood).

Here is an illustration of the solving process:

From the relation (1.11) we can see that $\theta_i = b'^{-1}(\mu_i) = h(\mu_i)$, with $h = b'^{-1}$ .

Moreover from (1.13) we have $\mu_i = g^{-1}(\beta_0 + \sum_{k=1}^{p} \beta_k . X_{ik})$

We can therefore write that $\theta_i = h(g^{-1}(\beta_0 + \sum_{k=1}^{p} \beta_k . X_{ik}))$.

Let's write the maximum likelihood function:

$$l(y_1, y_2, \dots, y_n, \theta_1, \theta_2, \dots, \theta_n) = \prod_{i=1}^{n} f(y_i, \theta_i)$$

thus,

$$l(y_1, y_2, \dots, y_n, \beta_0, \beta_1, \dots, \beta_p) = \prod_{i=1}^{n} f(y_i, h(g^{-1}\left(\beta_0 + \sum_{k=1}^{p} \beta_k . X_{ik}\right)))$$

We can then estimate the parameters $\beta_i$ by solving the system of p equations:

$$\frac{\partial ln\, l(y_1, y_2, \dots, y_n, \beta_0, \beta_1, \dots, \beta_p)}{\partial \beta_k} = 0$$

We therefore obtain a set of parameter estimates for the model $\hat{\beta} = (\widehat{\beta_0}, \widehat{\beta_1}, \dots, \widehat{\beta_p})$.

In practice, the large number of observations that we are dealing with means that the resolution of the system of equation is rarely done using linear algebra. Instead numerical techniques (and in particular multi-dimensional Newton-Raphson algorithms) are used.

- *Offset term :*

An important feature that will be useful in our revisited GLM model is the introduction of an offset parameter. In some occasions the effect of an explanatory variable is known. In that case, rather than estimating parameters $\beta$ in respect of this variable, it is appropriate to include information about this variable in the model as a known effect. This can be achieved by introducing an "offset term" $\underline{\xi}$ into the definition of the linear predictor $\underline{\eta}$:

$$\underline{\eta} = \boldsymbol{X}.\underline{\beta} + \underline{\xi} \tag{1.14}$$

We will therefore obtain $E[Y] = \underline{\mu} = g^{-1}(\underline{\eta}) = g^{-1}(\boldsymbol{X}.\underline{\beta} + \underline{\xi})$

For example, the offset can be used when fitting a GLM on claim count (instead of claim frequency). Given the claim count is proportional to the exposure measure; we can set the offset term to be equal to the log of the exposure of each observation. This will result in frequency model multiplied by the exposure.

### 2.4 Goodness of fit

One key stake when performing a loss cost model using GLMs, is to select the right subset of covariates. This is an important phase of the analysis and requires a good sense of modelling from the analyst; the aim is to make a trade-off to choose the best model that will:

- the best explain (or fit) the response variable Y.
- have as less covariate as possible to be a robust prediction method.

In fact the more covariates we will introduce, the more precise the modelling of the response variable will be. However we will be faced to a risk of overfitting. In reality it is not possible for a model to perfectly describe the response variable given the experience of Y includes a random error, also called noise (see (1.1)). Overfitting occurs when a model describes random error or noise instead of the underlying relationship. This generally occurs when a model is excessively complex, such as having too many parameters relative to the number of observations. A model that has been overfit will generally have poor predictive performance. We say in that case that the model is not robust.

In order to estimate the quality of the model, the modeller will use different statistics, especially the Deviance and Pearson's Khi-square.

- *Deviance:*

A deviance is a measure of how much the fitted values differ from the observations. The definition of the deviance $D$ is as follow:

$$D = 2\phi \sum_{i=1}^{n} [ln\ l(y_i, y_i) - \ln l(y_i, \widehat{\mu_i})] \tag{1.15}$$

$\phi$ is the scale parameter.

$l(y_i, y_i)$ is the "saturated" likelihood which corresponds to the likelihood in case the parameters would be exactly equal to the observations; this implies that the number of parameters is equal to the number of observations, and the saturated likelihood is the maximum achievable likelihood.

$l(y_i, \widehat{\mu_i})$ is the likelihood with our estimated mean $\widehat{\mu_i}$.

Therefore the deviance represents the difference between the maximum achievable likelihood and the likelihood of our model. The better the model the smaller the value of the deviance.

- *Pearson's Khi-square:*

Pearson's Khi square $\chi^2$ is another statistic that assesses the overall quality of the fit of the model. It is defined as follow:

(1.16)

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{var(\hat{\mu}_i)}$$

Those statistics indicated how well the model fit the experience, yet they do not take into the number of parameter used. That is why other statistics have been created such as the *AIC* and *BIC* statistics. The logic behind them is that the inclusion of additional parameter should be penalised; which makes sense as we aim to select the best model which gives the best fit with the least variables possible.

- *AIC (Akaike Information Criterion):*

*AIC* statistic has the form:

$$AIC = -2 \ln l(y_i, \hat{\mu}_i) + 2p \tag{1.17}$$

with p is the number of parameters estimated in the model. We will also mention an alternative form of the *AIC*, which is the corrected *AIC* given by:

$$AICc = -2 \ln l(y_i, \hat{\mu}_i) + 2p \frac{n}{n-p-1} \tag{1.18}$$

where n is the total number of observations used. *AICc* is usually used for sample where the number of observation n is low, or the number of parameters p is large; hence *AICc* will further penalize the inclusion of an extra covariate.

- *BIC (Bayesian Information Criterion):*

*BIC* statistic is a similar measure as *AIC* and is defined as follow:

$$BIC = -2 \ln l(y_i, \hat{\mu}_i) + p.log(n) \tag{1.19}$$

These statistics are used to compare models to each other, hence the lower the statistic the better the model. This is used when it comes to select a subset of covariate by looking at the model including and excluding the variable.

### 2.5 Selection of variables

As a starting point of a new model, it is useful to make use of stepwise methods to select the subset of covariates. The most commonly used methods are the forward or backward method.

Step forward method is an iterative method which consists in testing the inclusion of a new variable in the model and looking at the impact on the deviance before and after inclusion to test its significance. The method follows these steps:

- Try all covariate separately in a single variable model.
- Select the variable which shows the best fit (lowest deviance).
- Select the second variable which associated with the first give the best fit overall.
- Repeat those steps until the inclusion of a new variable does not bring any more value to the fit.

In the step backward method we will start with a model including all variables. We will then remove the less significant covariate, and so on until the removal of a covariate is significantly affecting the quality of the overall fit.

In order to determine whether the inclusion of a new covariate is significant or not, we can perform different tests:

- Statistical tests, such as the Chi-squared test of equivalence. This methods consists in testing the null assumption "The two models (with and without the variable) are essentially the same". If the test rejects the assumption, we can use the more complex model including the variable.
  The details of the methodology can be found in **Appendix 1.3**.

- Judgement: While the use of the statistics is important for the selection of a variable, it is essential to use judgement and common sense. Therefore it is important to also check whether the pattern is logical and makes sense.

- Standard error of the parameter estimates:
  The estimation of the standard error of $\hat{\beta}$ is possible due to the asymptotic nature of the maximum likelihood. Under some regularity assumptions, the estimate of the maximum of likelihood is asymptotically normally distributed. Moreover the mean of the distribution is nil and its variance equal to the inverse of the Fisher Information matrix. We can write:

$$\sqrt{n}(\hat{\beta} - \beta) \longrightarrow N(0, \hat{I}^{-1}) \tag{1.20}$$

  with $\hat{I}$ being the estimator of the Fisher Information matrix:

$$\hat{I}_{j,k} = E[\frac{\partial \ln l}{\partial \hat{\beta}_j} . \frac{\partial \ln l}{\partial \hat{\beta}_k}]$$

  We therefore obtain the interval confidence at 95% for the parameter $\hat{\beta}_i$:

$$\left[\hat{\beta}_i - 1.96 \times \sqrt{\hat{I}_{i,i}^{-1}} \quad ; \quad \hat{\beta}_i + 1.96 \times \sqrt{\hat{I}_{i,i}^{-1}}\right]$$

  This interval is also called *Wald Confidence Interval*. The narrower the interval around the estimated curve, the more accurate the parameters and the better the model. A common test is to draw a horizontal line from the level base; if the line can pass through the confidence interval we may decide not to include the variable given the uncertainty around the estimated parameters is too high.

- Consistency over time: this test consists in checking the trend of the parameters when interacted with a time variable (ex: accident year). If the factor displays a consistent pattern over time, we can be more confident that the pattern is predictive of the future.

### 2.6 Model validation

The validation of the model is an important step of Generalized Linear Models. It consists in checking that the model is appropriate, which means that it is predictive and has good explanatory power. It therefore verifies that there is no overfitting. To do so the modeller will use different tests that we mention here below:

- *Residual analysis:*

Various measures of residuals can be derived to show, for each observation, how the fitted value differs from the actual observation. In practice, we will use the Deviance and Pearson residuals, which are linked to the Deviance and Pearson's Khi-square statistics. We will give a conceptual definition below:

The Deviance residual $r_i^d$ of an observation $i$ represents the contribution of the observation $i$ to the Deviance, hence we can write:

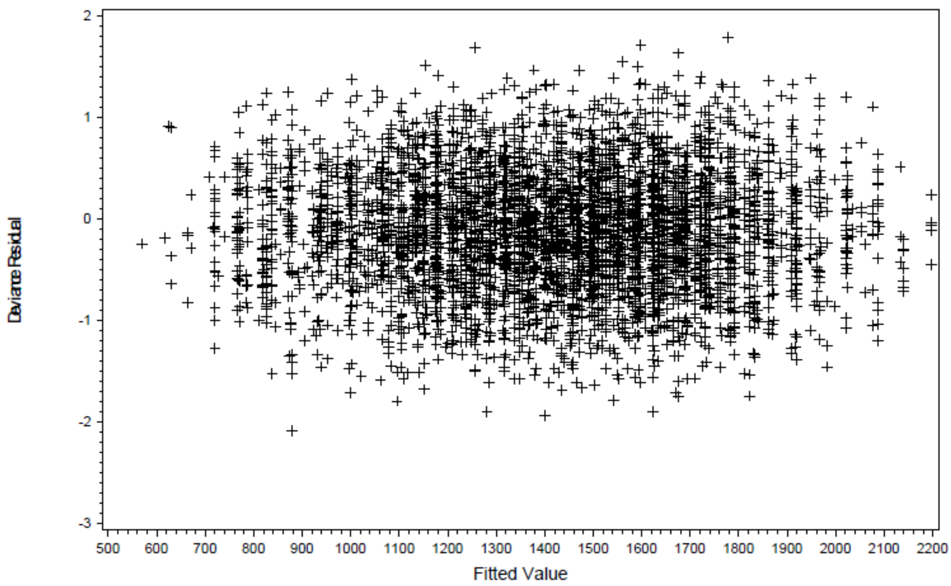
$$D = \sum_{i=1}^{n} r_i^d$$

Similarly, the Pearson's residual $r_i^p$ of an observation $i$ represents the contribution of the observation $i$ to the Pearson's statistic, hence we can write:

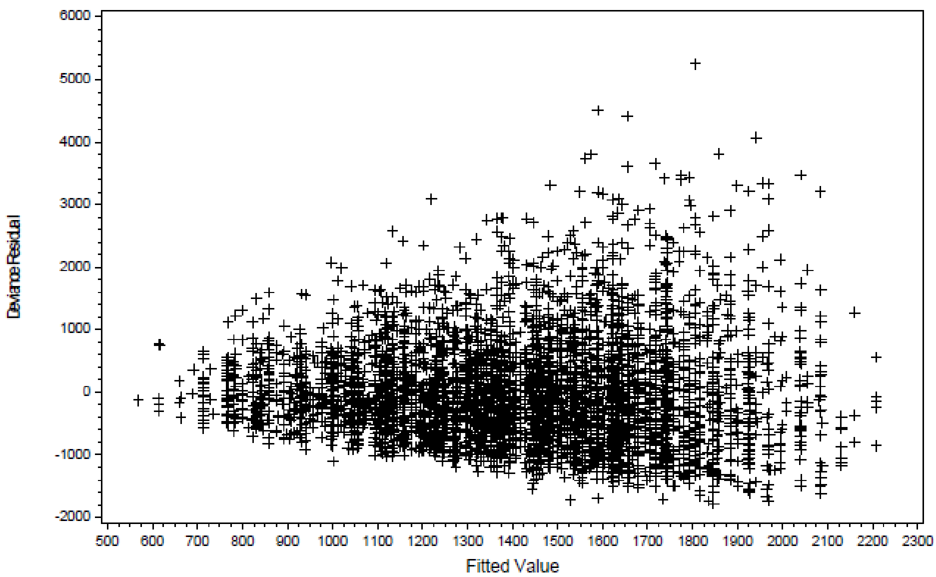
$$\chi^2 = \sum_{i=1}^{n} r_i^p$$

$r_i^d$ and $r_i^p$ have the same sign as the raw residual $y_i - \hat{\mu}_i$. Therefore the test consists in observing the scatter-plot of the residuals with regard to the estimated value $\hat{\mu}_i$ on the x-axis. If the error function that has been assumed is appropriate we would expect the residuals to be distributed symmetrically around the nil axis, with no obvious trend of the scatter-plot. We will then assume the error-term to be purely reflecting the noise.

For example, in the scatter plot below which shows the deviance residual of a hypothetical model: from the left to the right of the graph the general mean and variability of the residuals is reasonably constant, suggesting that the assumed variance function is appropriate




On the contrary in the graph below, we observe that the variability increases with the fitted value. This indicates that an inappropriate error function may have been selected and that the variance of the observations increases with the fitted values to a greater extent than has been assumed. We should not validate this model.




- *Hold out sample:*

A commonly used method is to compare the prediction of the model on a hold out sample. In fact given the model's estimates are derived from the data sample; we would expect the prediction to be close to the experience. By using a test dataset, we can check the predictiveness of the model, or whether it is overfitted.

We split the data in two subsets before the modelling:

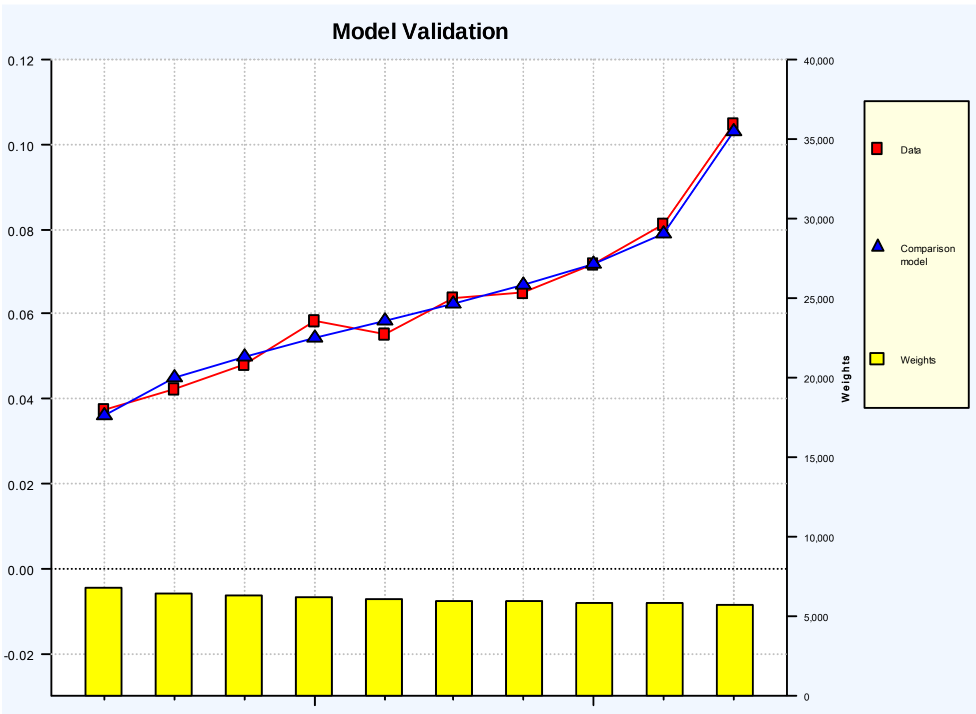
- modelling data, to build the model (80% to 90% of the total data).
- hold out data (or test data), used to compare predictions to actuals.

The test will consist in comparing the model predictions on the test dataset in comparison with its actuals. The test data is usually randomly selected through the sample. Another option is to retest the model on the future quarter data; in that case the hold out sample would be the future period data.

- *Lift Chart:*

A useful test to perform on the hold out data is the Lift Chart. It consists in sorting the observations by increasing estimated value $\hat{\mu}_i$. We then split the data in $n$ bands of equal population (we usually use 10 bands). We finally look at the predicted values versus the actual values by bands. A good model should have a good fit of the lift curve.

The following chart shows an example of a good fit:

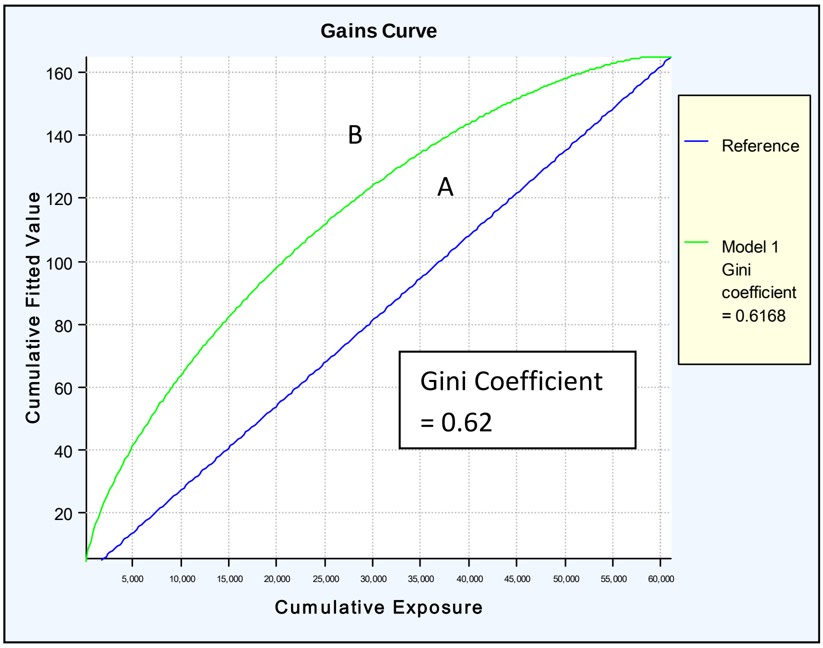



- *Gain curve:*

Gain curve is also good to compare the predictiveness of models. The gain curve is described as follow:

- On the x-axis, we have the exposure that we have sorted by decreasing estimated value $\hat{\mu}_i$.
- On the y-axis, we have the cumulative actual values.

The following charts present an example of Gain curves.




The blue line represents the mean model, in which all exposures have the same estimated value. The green line represents the model: the further away it is from the mean model, the more lift the model gives, hence the more predictive the model. The Gini coefficient corresponds to the proportion of the area between the curve green and blue (A) over half of the area of the chart (A+B). The higher the Gini coefficient the better the predictiveness of the model.

Now we have defined the theoretical background around Generalized Linear Models, we will use this method on our auto portfolio to perform loss cost modelling.

## Section 3: Results of the Loss Cost Model

### 3.1 Model overview

The loss cost model has been reviewed by individual perils for the frequency and severity. The final model, also called combined model, which represents the estimated pure premium by policy is defined as follow:

$$\boxed{\text{Combined TPL model}} = \boxed{\text{TPPD Freq}} \times \boxed{\text{TPPD Sev}} + \boxed{\text{TPBI Freq}} \times \boxed{\text{TPBI Sev}}$$

For our study, we will perform the GLM modelling using EMBLEM. EMBLEM is a software package developed by Tower Watson, and presents the advantage of having a user friendly interface to perform various analysis and tests. Alternatively we can use the SAS software; the results would be the same eventually.

An important diagnostic to look at is the correlation between variables. When analysing a variable to include in the model, it is important to consider variables which are correlated to this one; hence a highly correlated variable can capture the effect of the considered variable when introduced. We will therefore prefer to include first the variable we consider having the main effect. We then will try to add the other correlated factors to see if they bring additional prediction to the model.

After the analysis of correlation, we estimated the following buckets of variables to be analysed together.

**Table 1.9 – Groups of correlated variables**

| Bucket 1 | Bucket 2 | Bucket 3 |
|---|---|---|
| Driver license age<br>Conviction | Region<br>City | All vehicle related variables |

For further details around the variable correlation analysis, please refer to **Appendix I.4**.

Following a detailed GLM analysis by peril, we have estimated the predictive rating factors of our loss cost model. The following table summarizes the selected variables for the proposed rating model.

**Table 1.10 – List of selected rating factors**

| Variable | TPPD - Freq | TPPD - Sev | TPBI - Freq | TPBI - Sev |
|---|---|---|---|---|
| Age of insured | X | X | X | |
| Gender of insured | X | | X | |
| Insured marital status | | | | |
| Insured driver license age | | | | |
| Conviction in the past 5 years (Yes/No) | X | X | | |
| Number of at-fault claims in the past 5 years | X* | | X* | |
| Region | X | X | X | X |
| City | | | | |
| Single payment (Yes/No) | | | | |
| Vehicle model year | X | X | | |
| Vehicle value | | | | |
| Vehicle make | X | | X | |
| Vehicle make-model | | | | |
| Vehicle use (Social only/ Business) | | | | |
| Total no of cars in the household | | | | |
| Additional driver (Yes/No) | | | | |
| Garage (locked garage / street) | | | | |
| Vehicle type (City / Sport / Saloon / Cabriolet / 4x4) | | | | |
| Vehicle automatic transmission (Yes/No) | | | | |
| Vehicle Imported / Local | | X* | | |
| Vehicle engine power | X* | X* | X* | X* |
| Vehicle breaking assistance (Yes/No) | X* | X* | | X* |
| Vehicle driving assistance (Yes/No) | | | | |
| Vehicle no of doors | | | | |
| Vehicle length | | | | |
| Vehicle weight | | | | |
| Alarm (Yes / No) | | | | |
| Tracking device (Yes / No) | | | | |

Variables ticked with $\mathbf{X}^*$ correspond to variables that were not used before in the rater. We observe that by adding additional vehicle information, we are able to add 3 extra variables. This should improve the risk selection.

We can now present the results of the GLM by rating factor for each model.

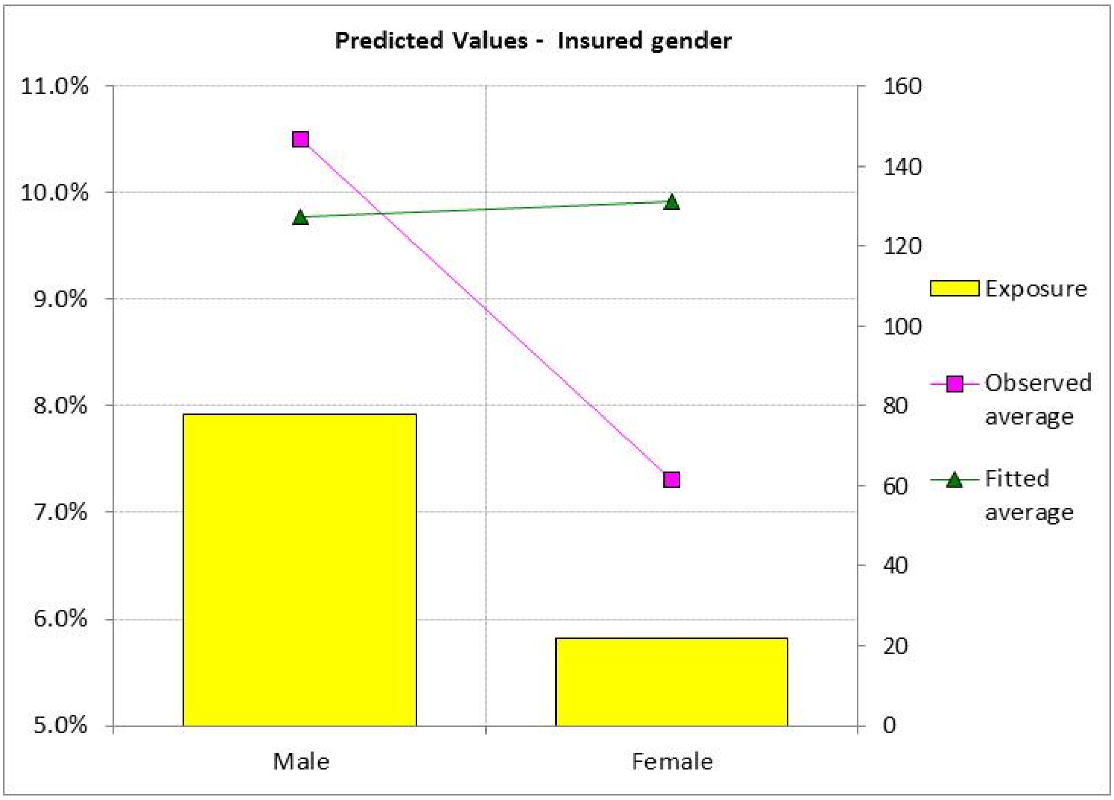
## 3.2 Rating factor analysis - TPPD

In this section, we will emphasize the approach we adopt to analyse whether the Gender[2] variable should be included in the frequency model. However we will not replicate the details for all factors as this will be redundant. For the other variables we will only present the final results.

[2] Country A is not part of the EU, therefore we can use the Gender variable as a rating factor.

- **Driver Gender**

Prior including the factor, we can compare the observed vs fitted frequency by segment. This is called the balance test.



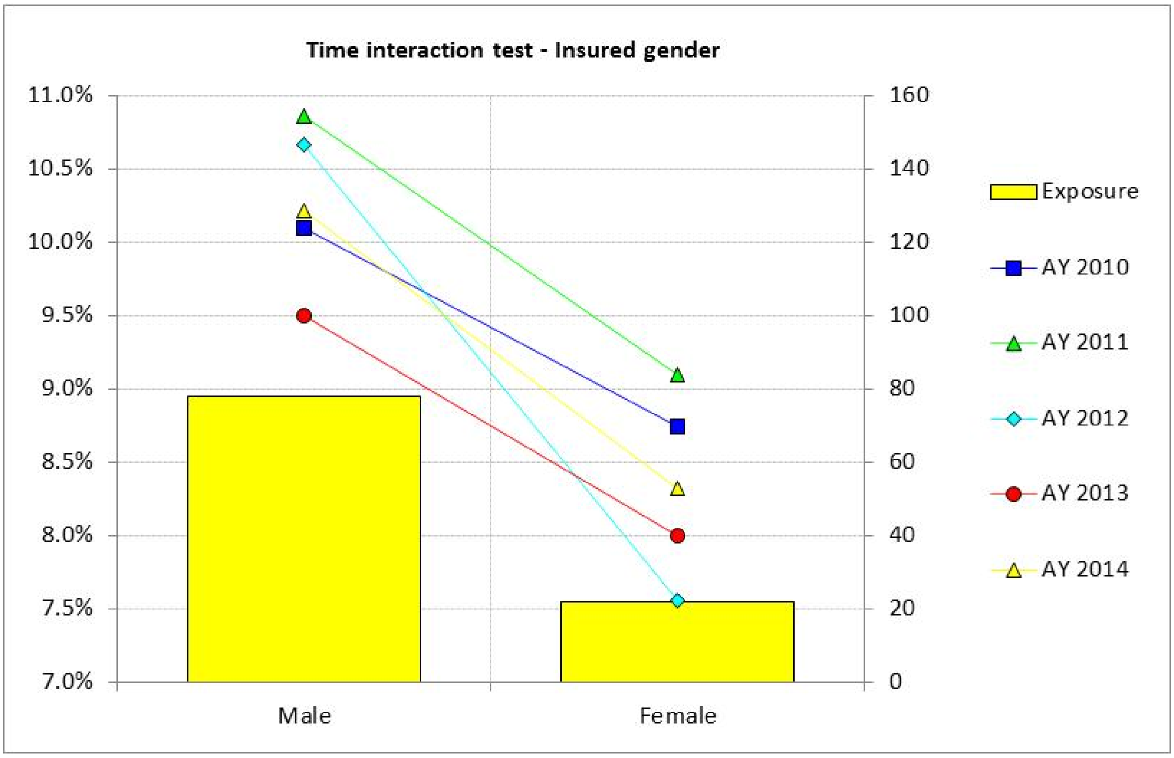

We observe that the fitted average curve is not well aligned with the observed. This means the other factors are not adjusting for the Gender factor, i.e. Gender could be a predictive factor.

We can select the model excluding Gender as a reference model, and compare the results of the Khi 2 test to the model including it.

| | Current Model | Reference Model | Difference |
|---|---|---|---|
| Model Label | Model including Gender | Model excluding Gender | |
| Sampling | Modelling | Modelling | |
| Zero Weighted | 914,530 | 914,530 | 0 |
| Fixed or Simple Alias | 0 | 0 | 0 |
| Complex Alias | 0 | 0 | 0 |
| Fitted Parameters | 27 | 26 | 1 |
| Deviance | 279,205.80 | 279,277.00 | -71.144 |
| Chi Squared Percentage | | Sub-Model | 0.00% |
| AIC | 422,998.58 | 423,000.15 | -1.56549 |
| Fitting Result | Converged OK | Converged OK | |

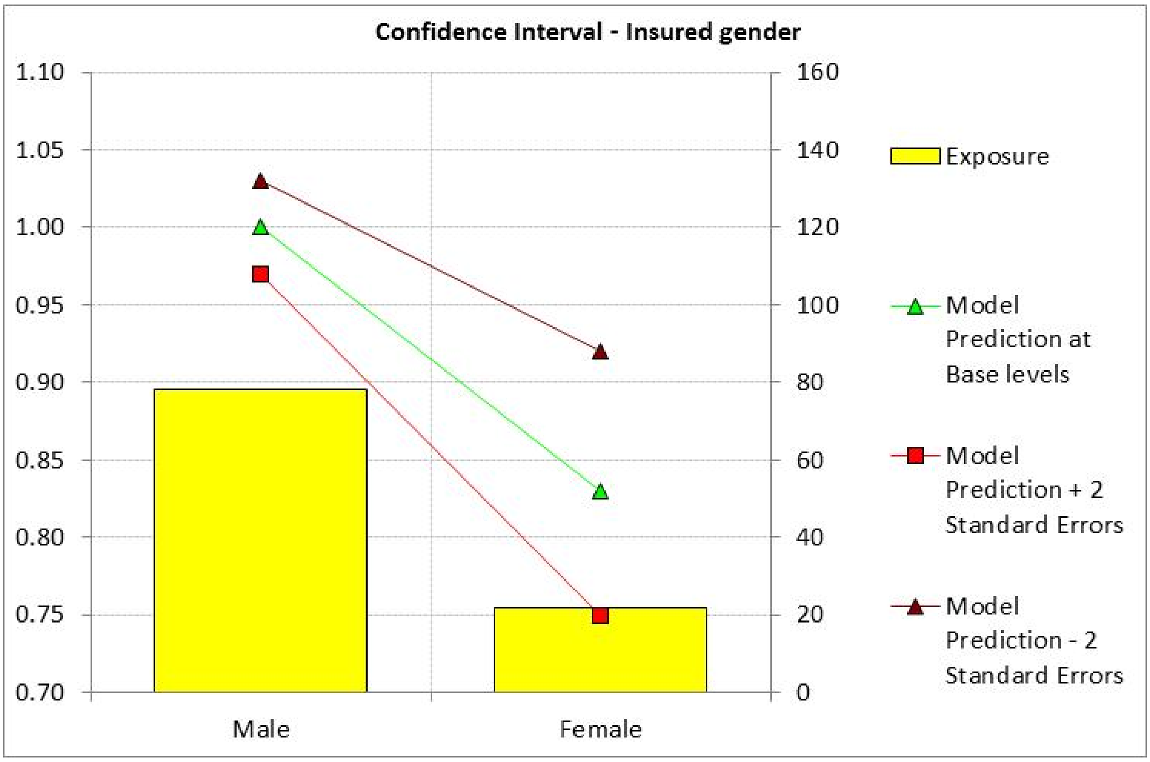
We observe that the Chi Squared percentage is nil, which means the two models can be considered different; hence Gender seems to be a predictive variable. We also observe that the AIC is lower when including Gender which is in line with our assumption.

Another important test is the time consistency test. For this test, we will 'interact' the factor with the Accident Year factor. This means that the model will fit a parameter for each Gender x Accident Year factor combination. If the variable is predictive, we would expect the relativity trend of each Gender x Accident Year factors to be in line. The following graph shows the results of the time interaction test.




We observe that the relativity curves are all showing a consistent trend, which shows that female drivers present a higher TPPD claim frequency.

The final test we will perform on that factor is the analysis of the confidence interval of the estimated relativities, also called the 'horizontal line' test. The following graph shows the estimated relativity of the gender frequency with its confidence interval at 95%.




We observe that the CI for Male and Female are distinct, in other words we cannot pass a horizontal line through the confidence interval parallel lines. The Gender factor therefore passes the horizontal line test.

Given Gender has passed all the tests; we will select the factor in our GLM model as we can conclude it is a predictive factor.

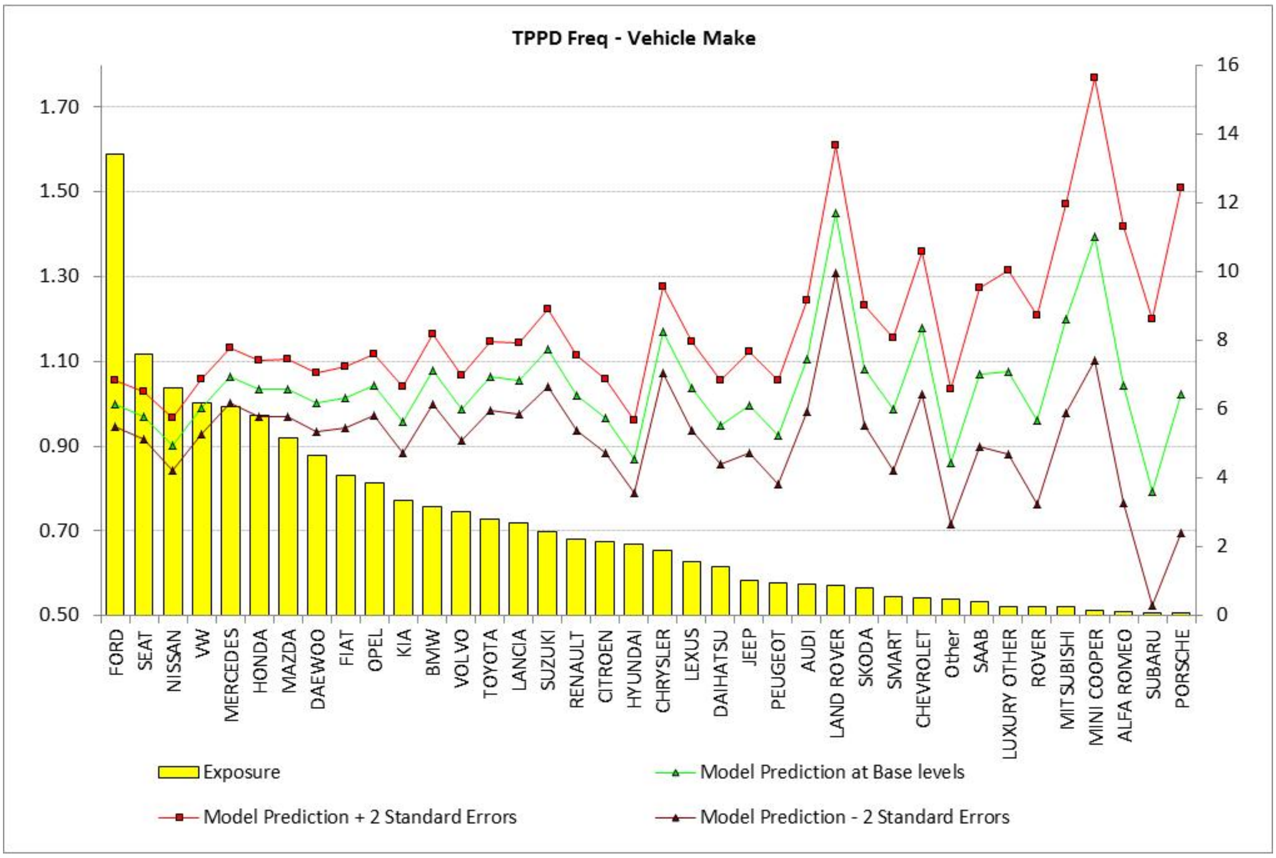
The results of the frequency model for the other factors are described in the **Appendix I.5**.

Overall the results of the GLM are satisfactory, especially for variables in which the number of levels is low enough to present enough volume on each of them (ex: Gender -> Male/Female). However an issue arises for factors with a high number of levels (ex: Make-Model which can have thousands of different levels). For these so-called multi-level factors, the result of the GLM will show a high degree of uncertainty on levels with low exposure. In our study, we will focus on the case of the Vehicle Make factor as it is an important area of improvement of our model.

- **Vehicle Make**

Prior to introducing the vehicle make factor, we first tested the vehicle characteristic factors, hence we've been able to include factors such as vehicle horsepower and vehicle transmission type. By doing this we are trying to explain the TPPD frequency as much as possible based on vehicle common characteristics. However, there may be residuals that still can't be explained; hence we will introduce the Make in order to capture the remaining 'Make specific' effect.

After introducing the Make factor, we observed that the factor is predictive, however given we are faced to a multi-level factor, the model result shows a high degree of uncertainty in term of confidence interval. In fact the vehicle make factor present 38 levels; some makes which are more popular will show a higher volume of exposure, while some other makes which are rarer will have low volume exposure. The following graph presents the model result when introducing the simple Make factor.



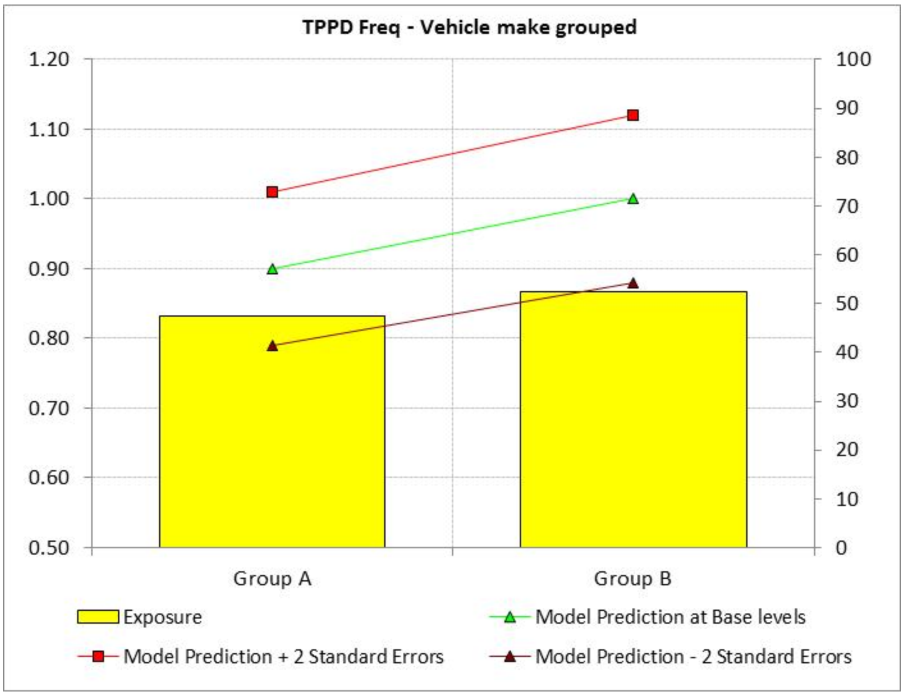

We observe that the confidence interval is relatively wide especially for make with few exposure volumes; hence the factor would not pass the horizontal line test through all make confidence interval. This is due to the fact that most of the levels of the factor do not have enough volume of exposure to be introduced in the GLM. In fact the GLM will maximize the likelihood of the past experience; therefore it will estimate $\hat{\beta}$ factors for each level. From (1.20) we know that the confidence interval of the estimated $\hat{\beta}$ is inversely proportional of the square of the exposure.

An alternative solution to cope with that limitation is to group levels together when they don't have enough exposure. Therefore we can create new grouped levels with higher exposure that can give credible results with the GLM.

In our study, we have split the makes between two groups, based on their individual frequency model result and underwriting and market inputs from the business (see table below).

| Group A | Group B |
|---|---|
| AUDI | ALFA ROMEO |
| BMW | CHEVROLET |
| DAIHATSU | CHRYSLER |
| FIAT | CITROEN |
| HONDA | DAEWOO |
| HYUNDAI | FORD |
| MAZDA | JEEP |
| MERCEDES | KIA |
| MITSUBISHI | LANCIA |
| NISSAN | LAND ROVER |
| OPEL | LEXUS |
| PEUGEOT | LUXURY OTHER |
| ROVER | MINI COOPER |
| SAAB | PORSCHE |
| SEAT | RENAULT |
| SKODA | SMART |
| SUZUKI | SUBARU |
| TOYOTA | |
| VOLVO | |
| VOLKSWAGEN | |
| Other | |

The following Graph shows the model result when introducing the grouped Make factor.




We can see that the model result by group is much more credible.

However by grouping levels together we are losing segmentation, which is our main objective in the GLM loss cost modelling. In fact in this model we would obtain similar rate for two different makes such as BMW and FIAT, while in reality those make are probably showing different TPPD frequency risk profiles. In fact the choice of grouping level together is here purely driven by the lack of exposure in each of them.

This limitation of GLM on small volume of exposure introduces a real problematic to our pricing study. In order to tackle this issue, we therefore propose to investigate alternative modelling methods to introduce credibility within our GLM.

### 3.3 Rating factor analysis – TPPD Severity

The results of the TPPD severity model by rating factors are shown in the **Appendix I.6**.

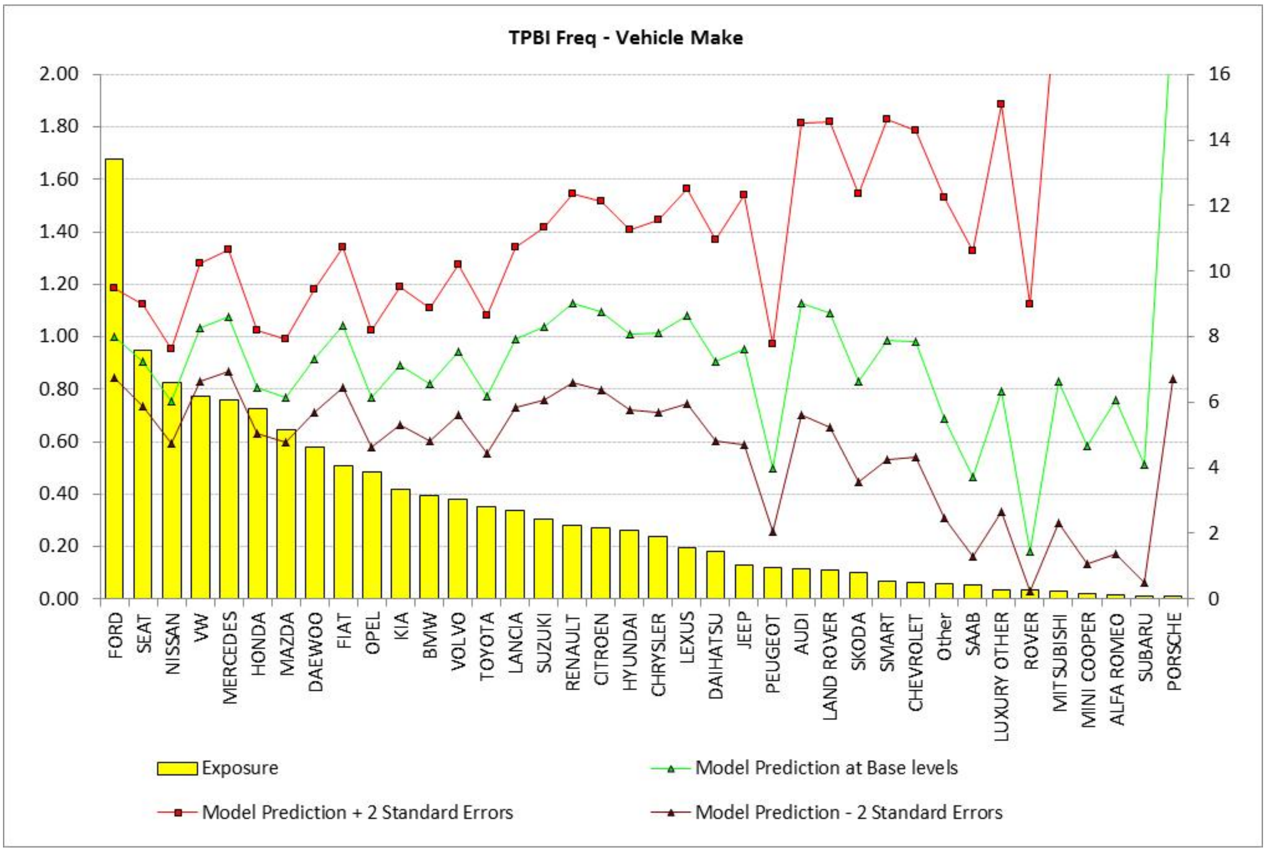
### 3.4 Rating factor analysis – TPBI Frequency

Modelling the BI peril is usually more difficult due to the nature of the risk. BI claims are less frequent claims which take longer to be settled; hence BI claims amounts are more volatile due to the uncertainty around the indemnity to be paid. For this reason, the number of factors that we can assure are predictive is lower in the BI peril modelling.

- **Vehicle Make**

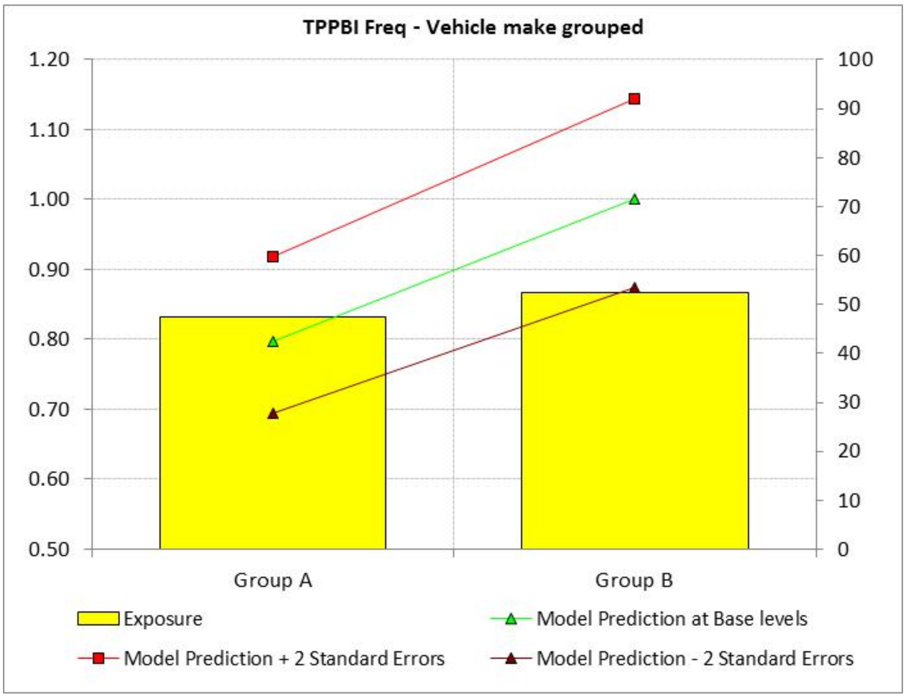
Similarly to the analysis we have performed for TPPD, we have tested the inclusion of auxiliary vehicle characteristic variables. In this case, only the Vehicle Horsepower category has passed the validation test to be included in the Model (see **Appendix I.7**).

We then introduced the Make simple factor. Such as for TPPD modelling, the model result is subject to a high degree of uncertainty (wide confidence interval). Therefore we cannot validate the inclusion of this factor.  The following graph presents the model result when introducing the simple Make factor.




We observe that the model result is subject to high degree of uncertainty. The confidence interval is even higher than the one obtained for TPPD. This makes sense given the BI peril has lower frequency experience.

Again, splitting the make into two groups of similar risk profile enables to reduce the confidence interval, however at a cost of a reduction in term of make segmentation. The following Graph shows the model result when introducing the grouped Make factor.



The results of the TPBI frequency model for the other factors are described in the **Appendix I.7**.

### 3.5 Rating factor analysis – TPBI Severity

The results of the TPBI severity model by rating factors are shown in the **Appendix I.8.**

**Conclusion**: The GLM is a powerful technique to determine predictive rating factors and model claim frequency and severity of auto insurance. However the main shortcoming that we observe is for rating variables which present a high number of levels. For those MLFs, there is a higher degree of uncertainty around the estimates of the levels with little exposure. The issue of credibility around risk group with little exposure volume is an important topic developed in the actuarial literature. We will therefore investigate alternative modelling techniques to try to tackle the issue observed for the vehicle make factor in our portfolio.

# Chapter 2: Credibility methods for Insurance Pricing

## Section 1: Introduction to Credibility

Even before the use of Generalized Linear Models developed in the 70's and 80's for insurance pricing, actuaries have investigated methods to segment risks appropriately within a heterogeneous portfolio. The concept of credibility has been used to developed methods to propose insurance premiums tailored for different groups of risks.

In this section we will first define what we mean by credibility and its use for insurance; we will then present the concept of linear credibility.

### 1.1 What is Credibility?

According to the dictionary definition, credibility is "the quality of being believable or worthy of trust". But how does this translate in insurance?

In reality everybody uses the concept of credibility when assessing the validity of a statement. For example, actuaries are often faced to a lack of statistics to infer conclusion in decision making process; in this case they will usually ask for expert judgement such as underwriters or portfolio managers. The actuary will then (unconsciously or not) either take the response of the expert as it is, or use it with caution depending on how much he trusts the expert's opinion. In fact if the actuary asked in the past for the expert's prediction at five different occasions and they all turned out to be largely false, it is likely that he would not give much credibility to his opinion this time. Credibility is therefore a matter of trust.

In insurance, the concept of credibility has been used to quantify how much credit we will give to the statement derived from the available data.

The theory arose at the beginning of the $20^{th}$ century. In the 1910s, the American multinational General Motors (GM) and other constructors including the small independent firm Tucker are insured at Allstate for worker's compensation cover. The experience across all insured constructors is computed and the overall rate is charged to everyone. However, GM computes its own rate and realizes that its experience is significantly better than what Allstate have been charging him. GM then asks Allstate to re-evaluate its premium purely based on its own experience, arguing that his volume is large enough to show stability across years. At the same time, Tucker asks for the same re-evaluation. While the actuaries of Allstate would agree with GM's argumentation they are facing a challenge with regard to this approach: if the number of employees of GM is high enough to fully trust its own experience and Tucker too small to do so, what is the minimum number of employees needed to be able to fully trust the company's experience? In other words what is the minimum number of employees to give full credibility to the insured experience?

Mowbray (1914) was the first to give an answer to that problematic, by setting a limit from which the insured experience is fully credible. This is called the credibility of stability or limited fluctuations. However his work brings a binary answer, depending on the exposure the insured is either rated on its experience or on the global experience. The limitation around this answer is that it is completely excluding insured which are below the minimum exposure level.

Whitney (1918) therefore introduced the concept of partial credibility. In his work he mentions the need to weight between the individual experience and the collective experience in order to estimate a fair premium.

The next development of this theory is brought by Buhlmann (1967-1969) who will formalize the concept of credibility. He will further develop the theory introducing more flexibility with the introduction of a weight factor (Buhlmann-Straub 1970) which increased the domain of application of the theory to insurance pricing.

### 1.2 The linear estimation of credibility

Insurance ratemaking is based on the Law of Large Numbers (LLN), hence by mutualising risks actuaries are able to estimate expected losses at a group level. It is obviously not possible to use the LLN on a single insured, therefore the insured will see his premium calculated based on the experience of the group.

For example, if we consider an individual $j$ and his historical claim amount for each year $i$ being $x_i$, with $i = 1$ to $n$. The observed risk premium of the individual is:

$$\bar{p}_n = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

The insurer could think of charging $\bar{p}_n$ to the insured, but what about insured who didn't report any claims, i.e. $x_i = 0$ for all risks $i = 1$ to $n$? Should we ask them a nil premium? This is obviously not a valid option as the insured would probably not have enough volume of experience on itself for his observed risk premium $\bar{p}_n$ to be credible, i.e. the LLN would not apply. As a result we will estimate the risk premium of the group $p_{group}$ and ask all insured this premium.

As illustrated in the example of Allstate, depending on the volume of the insured, actuaries can either decide to charge $\bar{p}_n$ or $p_{group}$.

American actuaries developed the theory of linear credibility which suggests taking a weighted average of these two risk premiums to achieve a fairer premium:

For an insured j, which belongs to a larger group of insured, his premium Π for the future year can be estimated as a linear function of $\bar{p}_n$ and $p_{group}$. The estimation of the linear credibility is defined as follow:

$$\Pi = Z\bar{p}_n + (1 - Z)p_{group} \tag{2.1}$$

with:

- $\bar{p}_n$ the observed risk premium of the insured
- $p_{group}$ the observed risk premium of the group
- $Z$ the credibility factor

$Z$ will depend on the individual risk and its evolution over time, it reflects how much credit we give to the insured own experience. Different cases can be possible:

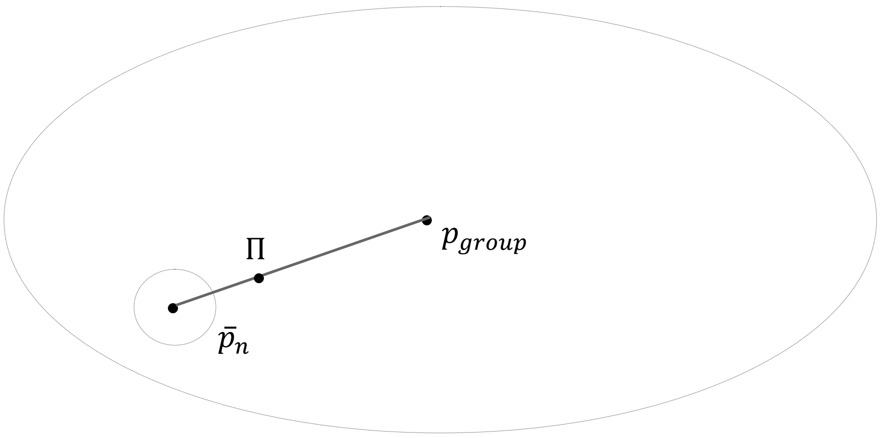
- If $Z = 1$, we have full credibility, i.e. the insured experience is large enough to justify only using it to estimate his risk premium.
- If $Z = 0$, we do not have any credibility, we will then use the risk premium derived from the group.
- If $0 < Z < 1$, we have partial credibility, in this case the insured premium will be a weighted average of the risk premium derived from his own experience and the one derived the from the group.

The linear credibility models developed by Buhlmann in the second half of the 20th century proposes to estimate the credibility factor $Z$. We present his major work in the following section.

## Section 2: Linear credibility models

The linear estimation of the credibility as defined in the prior section in (2.1) determines the future premium as a weighted average of the insured and the group observed risk premium.

This can be illustrated by the following chart.



The estimated risk premium of the insured for the future period is located on the segment between $\bar{p}_n$ and $p_{group}$.

The stake of the linear credibility is to estimate the credibility factor $Z$; hence we will present the main achievements in this area brought by Buhlmann. We will first introduce the mathematical framework, we will then present the Buhlmann model, followed by the Buhlmann-Straub which propose further developments.

### 2.1 Mathematical framework

Let's consider an insurance portfolio with $n$ policies (also called risks). Each policy is defined by a risk parameter $\Theta$ which contains all the information to describe the insured risk.

Let $X$ a random variable describing a statistic of the insured (such as frequency, severity, etc.). We will consider $X$ as the risk premium in this case. $X_{it}$ represents the observed risk premium of the insured $i$ over the period $t$, with $i \in [1, n]$ and $t \in [1, T]$. We define the following measures:

- A-priori mean of $X$: $E_X[X|\Theta = \theta]$. (2.2)
  For $i = 1$: it can be estimated as $E_X[X|\theta_1] = \sum_{t=1}^{T} x_t P[X = x_t |\, \theta_1]$

- Mean of the a-priori mean: $E[X] = E_X[X] = E_\Theta\big[E_X[X|\Theta]\big] = \sum_{i=1}^{n} P[\theta_i].E_X[X|\theta_i]$ (2.3)

- Variance of the a-priori mean: $Var_\Theta\big[E_X[X|\Theta]\big] = \sum_{i=1}^{n} P[\theta_i].\{E_X[X|\theta_i] - E_X[X]\}^2$ (2.4)

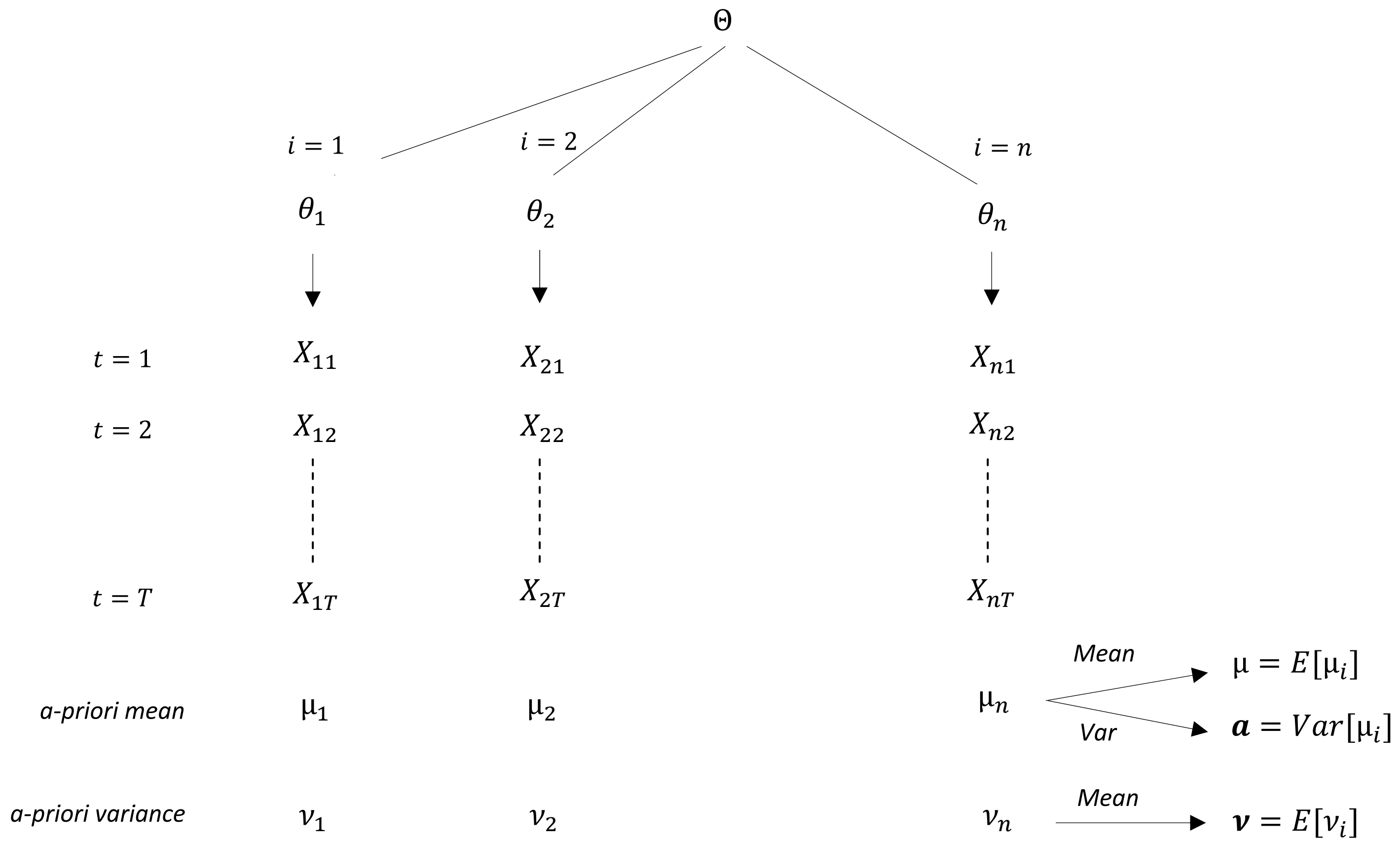
- A-priori variance: $Var_X[X|\Theta = \theta] = E_X[(X - E_X[X|\Theta = \theta])^2]$ (2.5)
  For $i = 1$: it can be estimated as $\sum_{t=1}^{T}(x_t - E_X[X|\theta_1])^2.P[X = x_t |\, \theta_1]$

- Mean of the a-priori variance: $E_\Theta\big[Var_X[X|\Theta = \theta]\big] = \sum_{i=1}^{n} P[\theta_i].Var_X[X|\Theta = \theta_i]$ (2.6)

We can simplify those notations:

- A-priori mean $\mu_i = \mu(\theta_i) = E_X[X|\Theta = \theta_i]$
- Mean of the a-priori mean: $\mu = E[\mu_i] = E[\mu(\theta_i)]$
- Variance of the a-priori mean: $a = Var(\mu_i) = Var[\mu(\theta_i)]$
- A-priori variance $v_i = v(\theta_i) = Var_X[X|\Theta = \theta_i]$
- Mean of the a-priori variance $v = E(v_i) = E(v(\theta_i))$

Scheme representation:




$\boldsymbol{\nu}$ is also called the variance "within" groups, while $\boldsymbol{a}$ is called the variance "between" groups.

## 2.2 Buhlmann model

We consider the insurance portfolio of $n$ policies as described above; however we do not know the structural parameters (derived from $\theta_i$). We have the observations of the risk premiums $X_{i,t}$ from which we will estimate the structural parameters, and the credibility factor.

The Buhlmann model aim to estimate the next period risk premium $X_{i,T+1}$, knowing the values of $X$ over the first $T$ periods: $E\left[X_{i,T+1} \middle| X_{i,1} = x_{i,1}, X_{i,2} = x_{i,2}, \dots, X_{i,t} = x_{i,T}\right]$. Buhlmann chose to model this Bayesian risk premium for the period $t + 1$ as a linear function of the past observations $(X_{i,1}\ ;\ X_{i,2}; \dots; X_{i,T})$.

Therefore the resolution of this model consists in solving the minimisation of the square error $E[(\mu(\theta_i) - a_0 - a_1 X_{i,1} - a_2 X_{i,2} - \cdots - a_n X_{i,T})^2]$

If we assume the following assumptions:

1. $E\left[X_{i,t}^2\middle|\theta_i\right] < \infty$ for all $i$ and all $t$, i.e. $X_{i,t}$ is quadratically integrable.
2. $E\left[X_{i,t}\middle|\theta_i\right] = \mu_i$ for all $i$ and all $t$, i.e. for a given risk $i$, all $X_{i,t}$ have the same mean.
3. $Cov\left[X_{i,t} \,.\, X_{i,t'}\middle|\theta_i\right] = \nu_i$ for all $i$ and all $t$, i.e. for a given risk $i$, $X_{i,t}$ are independent.
4. For all $i$ and $j$, $\Theta_i((X_{i,t})_{t=1\ldots T})$ $\Theta_j\left((X_{j,t})_{t=1\ldots T}\right)$ are i.i.d, i.e. for a given $t$, $X_i$ and $X_j$ are independent and $\theta_i$ and $\theta_j$ are i.i.d.

By considering a linear estimation of the credibility such as in (2.1), Buhlmann demonstrates that the expected value of the next period risk premium $X_{T+1}$ is defined as follow:

$$\Pi = E[X_{T+1}] = Z\bar{p}_T + (1-Z)p_{group}$$

with:

- $\bar{p}_T = \frac{1}{T}\sum_t X_{i,t}$ the observed risk premium of the insured $i$ over the first $T$ periods:
- $p_{group} = \frac{1}{T}\sum_{t,i} X_{i,t}$ the observed risk premium of the whole group of insured over the first $T$ periods.
- $Z$ the credibility factor defined as follow:

$$Z = \frac{1}{1 + \frac{1}{T} \cdot \frac{\nu}{a}} \tag{2.7}$$

We also often write $k = \frac{\nu}{a}$, hence we have $Z = \frac{1}{1+\frac{k}{T}}$.

$Z$ is therefore purely defined by the variance "within" and the variance "between". Moreover we can notice the following characteristics of $Z$:

- $Z$ is increasing with $T$: if $T \to \infty$ then $Z \to 1$, which means we have enough experience data therefore we can use the individual own experience.
- $Z$ is increasing with the variance 'between' $a$: if $a \to \infty$ then $Z \to 1$, hence if there is a high heterogeneity in the portfolio between groups, we will give more weight to the individual own experience. In the contrary if $a \to 0$ then $Z \to 0$, if the portfolio is homogenous across groups, we will not give any weight to the individual experience and use the group's experience.
- $Z$ is decreasing with the variance 'within' $\nu$: if $\nu \to \infty$ then $Z \to 0$, hence if there is a high heterogeneity within each groups, we will give less weight to the individual own experience. In the contrary if $\nu \to 0$ then $Z \to 1$, if the variance within the groups is very low, we will give more weight to the individual experience.

In general, we don't know the structural parameters therefore we need to estimate them. We will use the following estimator definition:

- $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \overline{x_i} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}$
- $\hat{v} = \frac{1}{n}\sum_{i=1}^{n} \overline{v_i} = \frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it} - \overline{x_i})^2$
- $\hat{a} = \frac{1}{T-1}\sum_{i=1}^{n}(\overline{x_i} - \overline{x})^2 - \frac{\hat{v}}{T}$

These estimators are non-biased. We highlight that $\hat{a}$ can be negative, which is inconsistent as it is a variance estimator. In this case we will use $\hat{a} = 0$.

We can illustrate the use of the Buhlmann model in the following example:

An insurance company covers employer's liability of two different companies. The annual claim amount over the past three years for each company is shown below.

| | Year 1 | Year 2 | Year 3 |
|---|---|---|---|
| Company A | 8 | 12 | 13 |
| Company B | 14 | 16 | 15 |

We want to estimate the pure premium for next year for each company. In this example, $n = 2$ and $T = 3$.

The a-priori mean of each company is estimated as follow:

$\overline{x_A} = \frac{1}{3}(8 + 12 + 13) = 11$

$\overline{x_B} = \frac{1}{3}(14 + 16 + 15) = 15$

Mean of the a-prior mean: $\hat{\mu} = \frac{1}{2}(11 + 16) = 13.5$

The a-priori variance of each company is estimated as follow:

$\widehat{v_A} = \frac{1}{3-1}[(8 - 11)^2 + (12 - 11)^2 + (13 - 11)^2] = 7$

$\widehat{v_B} = \frac{1}{3-1}[(14 - 15)^2 + (16 - 15)^2 + (15 - 15)^2] = 1$

Mean of the a-priori variance:

$\hat{v} = \frac{1}{2}(7 + 1) = 4$

Variance of the a-priori mean:

$\hat{a} = \frac{1}{3-1}[(11-13.5)^2 + (15-13.5)^2] - \frac{4}{3} = 2.92$

From (2.7), we can estimate the credibility factor:

$$Z = \frac{1}{1 + \frac{4}{3 \times 2.92}} = 0.457$$

From (2.1) we can estimate the pure premiums for each company for next year:

$\Pi_A = 0.314 \times 11 + (1 - 0.314) \times 13.5 = 12.36$

$\Pi_B = 0.314 \times 15 + (1 - 0.314) \times 13.5 = 14.19$

## 2.3 Buhlmann-Straub model

In the Buhlmann model, each risk is being attributed the same weight, no matter the exposure. This is the main shortcoming of the method. In reality each risk is likely to have a different exposure, for example if the number of employees of two companies is different. Moreover the exposure can also vary over time; hence a company may see his number of employee grow significantly between the year 1 to the year 3. Therefore we would expect risks or periods with more exposure to be taken more into account when evaluating the insured experience. The Buhlmann-Straub proposes to fix this bias.

The Buhlmann-Straub model is a generalization of the Buhlmann model, which introduces a weight factor for each risk and observed period. The assumptions of the model are the same as for Buhlmann classic model. Moreover we consider for each risk $i$ and period $t$, the exposure weight $\omega_{it}$. Therefore $\omega_{it}$ reflect how much weight we will to give to the experience of the risk $i$ over the period $t$.

We assume as in the classic Buhlmann model that each risk is characterized by its specific risk parameter $\theta_i$. We observe however that the initial assumption $Var_X[X_{it}|\Theta = \theta_i] = \nu(\theta_i)$ is no longer reasonable. In fact we would expect the variance to depend on the volume of exposure $\omega_{it}$.

By introducing this weight, we are now able to group together the experience of many risks over different time periods. In fact if we consider that the risk $X_{it}$ is made of different individual risks $S_{it}^{(k)}$ (e.g. employees within the company). $S_{it}^{(k)}$ corresponds to the aggregate claim amount of the k$^{\text{th}}$ risk in the class $i$ . We will have $X_{it} = \frac{1}{\omega_{it}} \sum_{k=1}^{\omega_{it}} S_{it}^{(k)}$. Given $X_{it}$ is an average of $\omega_{it}$ risks, which can be considered as, conditionally, independent of each other, we have: $Var_X[X_{it}|\Theta = \theta_i] = \frac{\nu(\theta_i)}{\omega_{it}}$ .

It is in fact reasonable to model the conditional variance as inversely proportional to the exposure measure.

The assumptions of the Buhlmann-Straub model are as follow:

The risk $i$ is characterized by an individual risk profile which is the realization of a random variable $\theta_i$, and we have that:

1. Conditionally given $\theta_i$, the $\{X_{it}: t = 1,2,\dots,T\}$ are independent with:

$$E_X[X_{it}|\Theta = \theta_i] = \mu(\theta_i) \tag{2.8}$$

$$Var_X[X_{it}|\Theta = \theta_i] = \frac{\nu(\theta_i)}{\omega_{it}} \tag{2.9}$$

2. The pairs $(\theta_1, X_1), (\theta_2, X_2), \dots, are\ independent$, and $\theta_1, \theta_2, \dots$ are independent and identically distributed.

In "A course in credibility theory and its application", Buhlmann describes the model as:

« A two-urn model. From the first urn we draw the risk profile $\theta_i$ , which determines the "content" of the second urn. In the second step, a random variable $X_{it}$ is drawn from the second urn. In this way a heterogeneous portfolio is modelled. The risks in the portfolio have different risks profiles (heterogeneity in the group).But the risks in the portfolio have also something on common: *a-priori, they cannot be recognized as being different (short terminology: they are a priori equal)*. This is modelled by the fact that the risk profiles $\theta_i$ are all drawn from the same urn. »

The resolution of the model consists in estimating, such as in the Buhlmann model, the a-priori mean $\mu(\theta_i)$ knowing the historical experience $(X_{i,1} = x_{i,1}, X_{i,2} = x_{i,2}, \dots., X_{i,T} = x_{i,T})$. Estimating the a-priori mean $\mu(\theta_i)$ is in fact equivalent to estimating the mean of $X_{i,T+1}$.

Buhlmann-Straub model is solved in the same way the Buhlmann model. We will consider the a-priori mean as a linear function of the past observations $(X_{i,1}\ ;\ X_{i,2}; \dots; X_{i,T})$. The demonstration will also consists in solving the minimisation of the square error $E[(\mu(\theta_i) - a_0 - a_1X_{i,1} - a_2X_{i,2} - \cdots - a_nX_{i,T})^2]$

The demonstration results in a same linear relation as for (2.1):

$$\Pi = E[X_{T+1}] = Z\bar{X}_T + (1 - Z)X_{group}$$

with:

- $\bar{X}_T = \frac{\sum_t \omega_{it}X_{it}}{\sum_t \omega_{it}}$ the observed risk premium of the insured $i$ over the first $T$ periods.
- $X_{group} = \frac{\sum_i \omega_i\bar{X_i}}{\omega}$ the observed risk premium of the whole group of insured over the first $T$ periods.

- $\omega_i = \sum_{t=1}^{T} \omega_{it}$ , the weight of the risk $i$ over the periods 1 to $T$.
- $\omega = \sum_{i=1}^{n} \omega_i$ ,the total weight of all risks.
- $Z$ the credibility factor defined as follow:

$$Z = \frac{1}{1 + \frac{1}{\omega_i} \cdot \frac{\nu}{a}} \tag{2.10}$$

If we put $k = \frac{\nu}{a}$, hence we have $Z = \frac{1}{1+\frac{k}{\omega_i}}$.

The structural parameters are unknown; therefore we will use the following estimators:

- $\hat{\mu} = \frac{1}{\omega} \sum_{i=1}^{n} \omega_i \overline{x_i} = \frac{1}{\omega} \sum_{i=1}^{n} \sum_{t=1}^{T} \omega_{it} x_{it}$
- $\hat{\nu} = \frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=1}^{T} \omega_{it} (x_{it} - \overline{x_i})^2$, with $\overline{x_i} = \frac{1}{\omega_i} \sum_{t=1}^{T} \omega_{it} X_{it}$
- $\hat{a} = \frac{\omega}{\omega^2 - \sum_{i=1}^{n} \omega_i^2} \{\sum_{i=1}^{n} \omega_i (\overline{x_i} - \overline{x})^2 - (n-1)\hat{\nu}\}$

These estimators are non-biased. We highlight that $\hat{a}$ can be negative, which is inconsistent as it is a variance estimator. In this case we will use $\hat{a} = 0$. We can easily obtain the Buhlmann model by putting $\omega_{it} = 1$, for all $i$ and $t$.

We can illustrate the use of the Buhlmann-Straub model in the following example:

An insurance company covers employer's liability of two different companies. The annual claim amount over the past three years for each company is shown below. The exposure figures correspond to the number of employee-year.

| | | Year 1 | Year 2 | Year 3 | Year 4 |
|---|---|---|---|---|---|
| Company A | Claim amount | 8000 | 12000 | 13000 | ? |
| | Exposure | 35 | 50 | 80 | 85 |
| Company B | Claim amount | 14000 | 16000 | 15000 | ? |
| | Exposure | 105 | 115 | 140 | 120 |

We want to estimate the pure premium for next year for each company. In this example, $n = 2$ and $T_1 = T_2 = 3$.

For the company A, we have the following parameters:

- Weight per year: $\omega_{A1} = 35$ , $\omega_{A2} = 50$, $\omega_{A3} = 80$
- Total weight of company A: $\omega_A = \sum_{t=1}^{3} \omega_{At} = 165$
- The observed risk premium per exposure for each year: $x_{A1} = \frac{8000}{35} = 228.6$ , $x_{A2} = \frac{12000}{50} = 240.0$ , $x_{A3} = \frac{13000}{80} = 162.5$
- A-priori mean: $\overline{x_A} = \frac{1}{165}(228.6 \times 35 + 240 \times 50 + 162.5 \times 80) = 200$

For the company B, we have the following parameters:

- Weight per year: $\omega_{B1} = 105$ , $\omega_{B2} = 115$, $\omega_{B3} = 140$
- Total weight of company B: $\omega_B = \sum_{t=1}^{3} \omega_{Bt} = 360$
- The observed risk premium per exposure for each year: $x_{B1} = \frac{14000}{105} = 133.3$ , $x_{B2} = \frac{16000}{115} = 139.1$ , $x_{B3} = \frac{15000}{140} = 107.1$
- A-priori mean: $\overline{x_B} = \frac{1}{360}(133.3 \times 105 + 139.1 \times 115 + 107.1 \times 140) = 125$

Total weight $\omega = 165 + 360 = 525$

Mean of the a-prior mean: $\hat{\mu} = \frac{1}{525}(165 \times 200 + 360 \times 125) = 148.6$

Mean of the a-priori variance:

$\hat{v} = \frac{1}{2\times(3-1)}\left(\sum_{t=1}^{T} \omega_{At}(x_{At} - \overline{x_A})^2 + \sum_{t=1}^{T} \omega_{Bt}(x_{Bt} - \overline{x_B})^2\right) = 73{,}991.98$

Variance of the a-priori mean:

$\hat{a} = \frac{\omega}{\omega^2 - (\omega_A^2 + \omega_B^2)}\{\omega_A(\overline{x_A} - \bar{x})^2 + \omega_B(\overline{x_B} - \bar{x})^2 - (n-1)\hat{v}\} = 2485.52$

From (2.10), we can estimate the credibility factor:

$$Z_A = \frac{1}{1 + \frac{73991.98}{165 \times 2485.52}} = 0.847$$

$$Z_B = \frac{1}{1 + \frac{73991.98}{360 \times 2485.52}} = 0.924$$

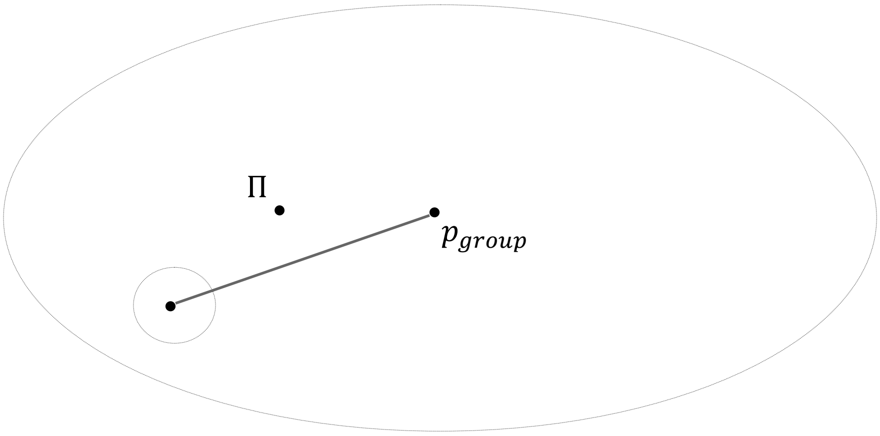
From (2.1) we can estimate the pure premiums for each company for next year:

$\Pi_A = 0.847 \times 200 + (1 - 0.847) \times 148.6 = 192.14$

$\Pi_B = 0.924 \times 125 + (1 - 0.924) \times 148.6 = 126.80$

# Section 3: Bayesian Credibility and Jewell's Theorem

The Bayesian credibility can be interpreted as a generalization of the linear estimation. In this approach the risk parameters will be considered as random variables; hence we will not necessarily obtain a linear estimation of the risk premium. Instead we will have a distribution of potential results. If we consider the illustration of the previous section, we can obtain the following best estimate for the insured risk premium:




The Bayesian risk premium is the best estimation of the insured individual risk premium $\mu(\theta_i) = E_X[X|\Theta = \theta_i]$. However it is really dependent on the chosen probability distributions and the implementation requires heavy calculus due to the computation of integrals.

We will first present the formalization of the Bayesian credibility, presenting the a-priori and a-posteriori distributions. We will then show that by selecting appropriate distribution families, the Bayesian approach is equivalent to the linear credibility.

### 3.1 Bayesian credibility model

If we consider $A$ an assumption, and $B$ an event which corresponds to the realization of a risk. We have:

- $P[A]$ is called the a-priori probability of $A$.
- $P[B|A]$ is the conditional probability of $B$ knowing $A$.
- $P[A|B]$ is called the a-posteriori probability of $A$.

The a-posteriori probability $P[A|B]$ is different than the a-priori probability $P[A]$, given we take into account the new information brought by the realization of $B$.

According to Bayes theorem we have:

$$P[A|B] = \frac{P[B|A] \times P[A]}{P[B]} \tag{2.10}$$

,with $P[B] > 0$. $P[B]$ is independent from $P[A]$, hence for the a-priori probability $P[A]$, $P[B]$ is considered as a constant. The a-posteriori probability $P[A|B]$ is proportional to the product of the a-priori probability $P[A]$ and the conditional probability of $B$, $P[B|A]$.

We can observe two situations: if the a-priori and a-posteriori probabilities belong to the same distribution family, or if they are different.

If we consider the scheme representation in the paragraph above (see 2.1 Mathematical framework), we can easily apply it to the Bayesian credibility model.

If we consider an individual risk $X_{it}$ which realizations at periods $t = 1, \dots, T$ correspond to claim amounts: $(X_{i1}, X_{i2}, \dots, X_{iT})$. If we are looking to estimate the pure premium for the next period $T + 1$. We make the assumption that all $X_i$ are i.i.d with probability distribution $f_{X_i}$.

In practice the distribution $f_{X_i}$ is unknown and varies by risk.

The individual risk premium for an insured which risk profile is $\theta_i$ is therefore:

$$\Pi_i = \Pi(\theta_i) = E[X_{iT+1} | \, \theta_i] = \mu(\theta_i) \tag{2.11}$$

Given that $\theta_i$ and $\mu(\theta_i)$ are unknown in practice, we will look for an estimator $\hat{\mu}(\theta_i)$. While we don't know the individual risk mean, we have information on the mean of the group. We will model the risk structure of the group by the random variable $\Theta$, which realization will correspond to the individual risk parameters $\theta_i$. The distribution of $\Theta$, $f_\Theta$ is called the structure function of the group; hence we can estimate the group risk premium as follow:

$$\Pi_{group} = \mu = E[\mu(\theta_i)] = \int_\Theta \mu(\theta_i) \, df_\Theta(\theta_i) \tag{2.11}$$

This can be considered like a two-urn model as mentioned by Buhlmann. We initially draw from the first urn the risk profile $\theta_i$ which defines the individual risk. This will determine the risk structure of the second urn from which we will draw the random variable $X_i$ which distribution depends on $\theta_i$.

Given we don't know explicitly the individual risk premium $\mu(\theta_i)$, our goal is to estimate it as much as we can. We will estimate it using the Bayes premium:

$$\Pi_i^{Bayes} = \tilde{\mu}(\theta_i) = E[\mu(\theta_i)|\mathrm{X}_i] = \int_\Theta \mu(\theta_i) \, df_{\Theta|X}(\theta_i|x_i)$$

with $\mathrm{X}_i = (\mathrm{X}_{i1}, \mathrm{X}_{i2}, \dots, \mathrm{X}_{iT})$ the prior realizations for that individual risk. (2.12)

In fact, Bayes premium $E[\mu(\theta_i)|\mathrm{X}_i]$ is the best estimator of $\mu(\theta_i)$ in term of the square error. For further details around the demonstration of this statement, please refer to the **Appendix II.1.**

$f_{\Theta|X}[\theta_i|x_i]$ is the a-posteriori distribution of $\Theta$, which is proportional to the product of the a-priori distribution of $\Theta$, $f_{\Theta}[\theta_i]$ and the conditional probability of $X$ knowing $\Theta$, $f_{X|\Theta}[x_i|\theta_i]$ (also known as the "likelihood function"). We will write:

$$f_{\Theta|X}[\theta_i|x_i] \infty f_{\Theta}[\theta_i] \, . f_{X|\Theta}[x_i|\theta_i] \tag{2.13}$$

$f_{\Theta}[\theta_i]$ is also called the *conjugate prior* for $f_{\Theta|X}[\theta_i|x_i]$.

For further examples of conjugate priors for common likelihood functions, please refer to the **Appendix II.2**.

We will see in the next paragraph that by choosing a-priori and a-posteriori functions that belongs to a specific distribution family, the Bayesian model can be similar to the linear Buhlmann model, as demonstrated by Jewell.

### 3.2 Jewell's theorem

Linear credibility estimators used in the Buhlmann model are derived under the assumption of linearity in the assumptions. Bayesian credibility on the other hand formalizes credibility using distributional assumptions without restrictions in term of class of function of the observations.

Exact credibility occurs when, under certain distributional assumption, the linear credibility estimator is optimal (in the sense of mean square error) for the Bayesian credibility model. In this case the linear credibility is therefore the best estimator of the Bayesian model.

In 1974, a theorem by Jewell states that when using a one-parameter exponential distribution to draw the observation ($X_{it}$), with its natural conjugate prior used to draw the risk parameter ($\Theta_i$), we obtain exact credibility, i.e. the optimal credibility estimator is equivalent to the linear credibility.

In fact, if we consider that $X_i$ follows a Poisson distribution: $X_i \sim P(\mu_i)$ (which is a distribution to model claim frequency) and the density function $\mu_i$ following a $Gamma$ distribution: $\mu_i \sim Gamma(\alpha, \beta)$ .

From (2.7) we can calculate the Buhlmann credibility factor $Z = \frac{1}{1+\frac{k}{T}}$, with:

$$k = \frac{\nu}{a} = \frac{E[\nu_i]}{Var(\mu_i)} = \frac{E_{\Theta}\big[Var_{\mathrm{X}}[X|\Theta = \theta_i]\big]}{Var_{\Theta}(E_X[X|\Theta = \theta_i])} = \frac{E_{\Theta}[\Theta]}{Var_{\Theta}[\Theta]} = \frac{\alpha/\beta}{\alpha/\beta^2} = \beta$$

We therefore obtain:

$$Z = \frac{1}{1 + \frac{\beta}{T}}$$

On the other hand, from (2.12) we know that the optimal estimate $\widetilde{\mu_i}$ of the parameter $\mu_i$ is $E[\mu_i|X_i]$, with $X_i = (X_{i1}, X_{i2}, \dots, X_{iT})$.

If we consider the density function of $\mu_i$ as the natural conjugate prior of the Poisson distribution $P(\mu_i)$, we have $\mu_i \sim Gamma(\alpha, \beta)$ and the a-priori distribution of $\mu_i$ is following a distribution $\sim Gamma\ (\alpha', \beta')$ , with $\alpha' = \alpha + \sum_{t=1}^{T} x_{it}$ and $\beta' = \beta + T$ (see **Appendix II.2**).

We therefore have:

$$E[\mu_i|X_i] = \frac{\alpha'}{\beta'} = \frac{\alpha + \sum_{t=1}^{T} x_{it}}{\beta + T}$$

$$= \frac{\alpha}{\beta + T} + \frac{T\overline{x_i}}{\beta + T}$$

,with $\overline{x_i} = \frac{1}{T}\sum_{t=1}^{T} x_{it}$

$$= \frac{\beta}{\beta + T} \times \frac{\alpha}{\beta} + \frac{1}{1 + \beta/_T} \times \overline{x_i}$$

$$= \frac{\beta/_T}{1 + \beta/_T} \times \frac{\alpha}{\beta} + \frac{1}{1 + \beta/_T} \times \overline{x_i}$$

$$= \left(1 - \frac{1}{1 + \beta/_T}\right) \times \frac{\alpha}{\beta} + \frac{1}{1 + \beta/_T} \times \overline{x_i}$$

$$= (1 - Z) \times \frac{\alpha}{\beta} + Z \times \overline{x_i}$$

$$= (1 - Z) \times E_\Theta[\Theta] + Z \times \overline{x_i}$$

, which is equivalent to Buhlmann linear credibility model.

Therefore we have demonstrated that for a Poisson distribution, when choosing its natural conjugate prior which is a Gamma distribution to model the risk parameter, the Bayesian approach is equivalent to the Buhlmann linear model.

Jewell's theorem, as we have demonstrated it here, is fundamental as it is a basis of the application to introduce credibility in our auto insurance pricing. Jewell's theorem enabled to make the bridge between the Bayesian model and Buhlmann model.

In the following chapter we will see how we can extend Jewell's theorem; therefore increasing the domain of application of the theory to insurance pricing. In particular we will show that we can use the outcomes of the theorem extension to incorporate credibility within our GLM auto pricing analysis.

# Chapter 3: Credibility theory and GLM revisited

The most common technique used nowadays to perform a loss cost modelling of an insurance portfolio is based on Generalized Linear Models. As we have seen in Chapter 1, the insurance industry uses GLMs to model risk premium as a multiplicative function of rating factors relativities. The success of GLMs is partly due to its practical implementation when considering many categorical rating factors; especially when those rating factors present few levels (e.g. gender), hence with enough exposure on each levels. Yet one of the main shortcomings of GLMs is that they do not cater for credibility.

Sophistication of insurance segmentation has in fact lead to the introduction of new rating variables with many levels without an inherent ordering, such as car model. For these variables, that we will call *multi-level factors*, GLMs does not give credible results for all levels. For example there is a number of a very common car models but also very uncommon ones. For levels with just a few exposures, the outputs of the GLMs will lead to non-credible results, as seen in Chapter 1.

In Chapter 2 we have seen how traditional credibility theory has historically been used in isolation of ordinary rating factors. In recent years, actuaries have worked to incorporate the theory of credibility within GLMs in order to cope with the uncertainty around multi-level factor.

Nelder and Verrall (1997) showed that credibility-like properties can be incorporated by introducing multi-level factors as random effects in GLMs. In our study we will use the theory of Ohlsson and Johansson (2003) which proposes to combine credibility theory and GLMs, by estimating random effects by mean of minimum square error (MSE) predictors.

In the first section we will present the formalization of introducing random effects in exponential family models that are used in GLMs. In the second section we will see how the extension of Jewell's theorem will enable us to solve the credibility estimator using MSE. We will finally present the process to estimate the parameters.

This chapter is largely based on Ohlsson and Johansson (2003) "Credibility theory and GLM revisited".

## Section 1: Random effect in exponential family models

In modern non-life insurance pricing, actuaries study the effect of rating factors on the risk premium based on GLMs techniques. In practice, this analysis is done separately on the frequency and severity as presented in the Chapter 1. Frequency and Severity models are then combined to create a final model to estimate the risk premium.

In our application, we will only focus on the modelling of the frequency component to incorporate credibility. Yet we will present the general formulation of the theory which can also be used to model the severity. We highlight that the severity modelling will require heavier calculation due to the different choice of distribution function.

If we consider an auto insurance portfolio. The rating factors divide the portfolio into *tariffs cells* and the key ratio $Y_i$ (which can be the frequency, severity or risk premium) is computed over the policies in cell $i$. (A tariff cell corresponds to a unique combination of rating factors: e.g. Male, 30, London, etc.).

From (1.7), we know that $Y_i$ is assumed to have a frequency function which belongs to the exponential family, i.e. of the form:

$$f_{Y_i}(y_i;\ \theta_i;\phi) = exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + \ c(y_i;\phi)\right\} \tag{3.1}$$

,with:

- $a_i(\phi) = {}^{\phi}/_{\omega_i}$ , $\phi$ being the dispersion parameter and $\omega_i$ the known exposure weight.
- $b(\theta)$ is twice differentiable with a unique inverse for the first derivative $b'(\theta)$.

From (1.11) and (1.12) we also know that:

$$\mu_i = E[Y_i] = b'(\theta_i) \tag{3.2}$$

$$Var[Y_i] = \frac{\phi b''(\theta_i)}{\omega_i} = \frac{\phi V(\mu_i)}{\omega_i} \tag{3.3}$$

If we consider $h(\mu)$ the inverse of $b'$, we have:

$$\theta_i = b'^{-1}(\mu_i) = h(\mu_i) \tag{3.4}$$

Moreover from Table 1.6, we know that we only consider the sub-class of exponential family models which is with a power function such as:

$$V(\mu) = \mu^p \tag{3.5}$$

The special values of $p$ that will be useful for insurance pricing are as follow:

- $p = 1$: Poisson distribution for claim frequency
- $p = 2$: Gamma distribution for claim severity
- $1 < p < 2$: Tweedie distribution for risk premium (compound Poisson with Gamma distributed summands)

Those models are continuous except for $1 < p < 2$ which shows a discontinuity at 0, where the probability is positive (probability of no claim).

Now we suppose that we have a number of ordinary rating factors (i.e. rating factors that are not multi-level factors) which divides the portfolio into $I$ tariff cells. Moreover we consider a multi-level factor which present $K$ levels. Let $\omega_{ik}$ be the exposure weight in the $i^{\text{th}}$ tariff cell with respect to the ordinary rating factors, and the $k^{\text{th}}$ level of the multi-level factor. We consider $Y_{ik}$ the key ratio (frequency in our study) that we want to model, $Y_{ik}$ is a random variable.

The model proposed by Ohlsson and Johansson assume that the effect of the multi-level factor is multiplicative. For level $k$ of the multi-level factor, the effect is considered to be the result of a random variable $U_k$. We therefore have:

$$E[Y_{ik}|U_k = u_k] = \mu_i u_k \tag{3.6}$$

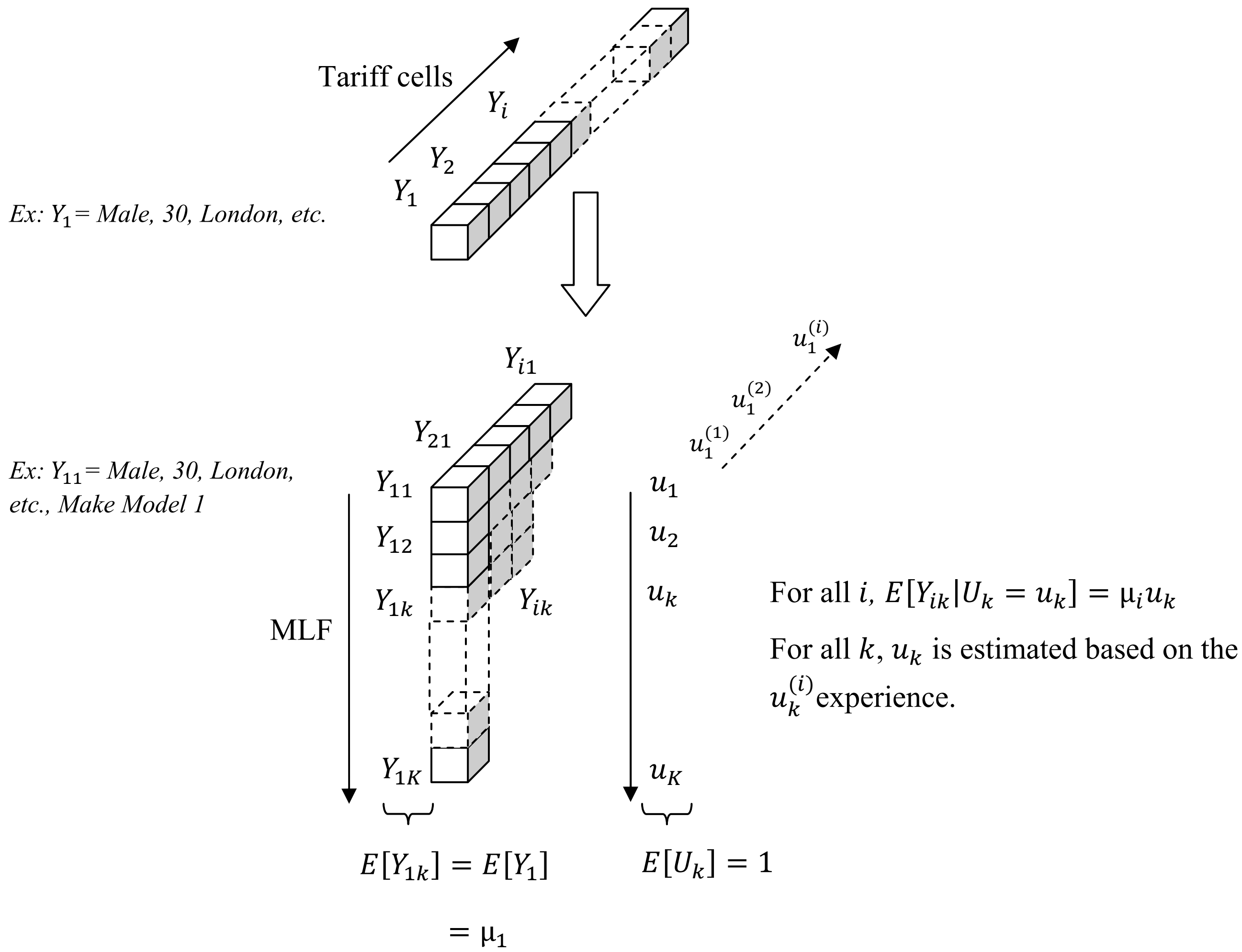
We know that the effects of the ordinary variables are random as well, yet this will not matter for the model. Since the systematic effects are captured by $\mu_i$, we can let the $U_k$s be purely random so that we have

$$E[U_k] = 1 \tag{3.7}$$

and hence:

$$E[Y_{ik}] = \mu_i \tag{3.8}$$

The following scheme illustrate the model of $Y_{ik}$ by introducing the random variable $U_k$:




Conditionally on $U_k = u_k$, we can assume that $Y_{ik}$ follows an exponential family distribution with expectation $\mu_i u_k$. We can therefore define the distribution of the key ratio $Y_{ik}$ in (3.1) in terms of the canonical parameter $\theta$. To do so we will use the relationship (3.2) which becomes $\theta'_{ik} = h(\mu_i u_k)$.

We can therefore write the conditional distribution of $Y_{ik}$ as follow:

$$f_{Y_{ik}|\Theta_k}(y_{ik}|\theta_k) = exp\left\{\frac{y_{ik}\theta'_{ik} - b(\theta'_{ik})}{\phi/\omega_{ik}} + \ c_1\right\} \tag{3.9}$$

Where $c_1$ is a constant that does not depend on $\theta_k$. We then want to re-write this distribution formulae with regard to $\theta_k$, hence Ohlsson and Johansson show that this formulation is equivalent to:

$$f_{Y_{ik}|\Theta_k}(y_{ik}|\theta_k) = exp\left\{\frac{\omega_{ik}}{\phi}\left[\frac{y_{ik}}{{\mu_i}^{p-1}}\theta_k - \frac{1}{\mu_i^{p-2}}b(\theta_k)\right] + \ c_2\right\} \tag{3.10}$$

Where $c_2$ is a constant that does not depend on $\theta_k$.

For further details around the demonstration of the formulae, please refer to **Appendix III.1**.

Conditional on $\Theta_k = \theta_k$, or equivalently on $\mathrm{U}_k = u_k$, the key ratios $Y_{ik}$'s are assume to be independent, and follow the distribution in (3.10). Moreover from (3.6), we have $E[Y_{ik}|U_k = u_k] = \mu_i u_k$. Therefore if we can estimate the $\mathrm{u}_k$, we can then run a standard GLM of the ordinary rating factors using $\log(u_k)$ as an offset variable (given the link function for frequency GLM modelling is log(.)).

From (1.14), we have indeed $E[Y_{ik}|U_k = u_k] = g^{-1}(\eta) = g^{-1}( X.\underline{\beta} + \underline{\xi} )$ ,

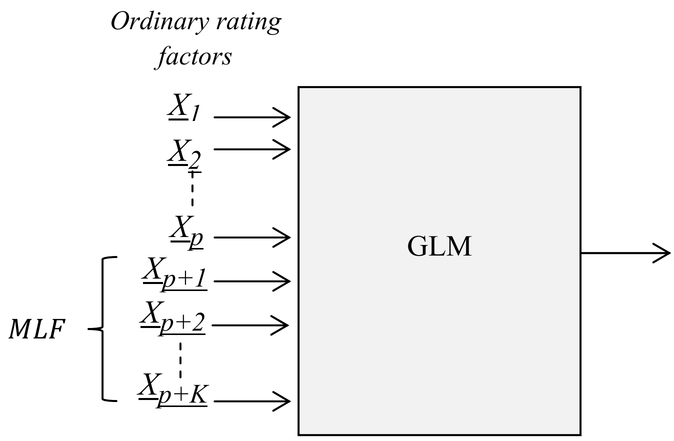
with the link function $g(x) = \log(x)$ and the offset term $\xi = log(u_k)$.

We highlight that given the $\mathrm{U}_k$'s are not known, we will need to predict them.

Therefore the point of modelling the MLF as a random variable $\mathrm{U}_k$ is then to find the best estimate of $\mathrm{U}_k$, which put as an offset term to our frequency model including the auxiliary variables gives us the model that fit our experience the best.

The following chart shows how the revisited GLM with random effect is run compared to a standard GLM.

**Standard GLM**




$$E[\,\underline{Y}\,] = g^{-1}(\,\boldsymbol{X}.\underline{\beta}\,)$$

with:

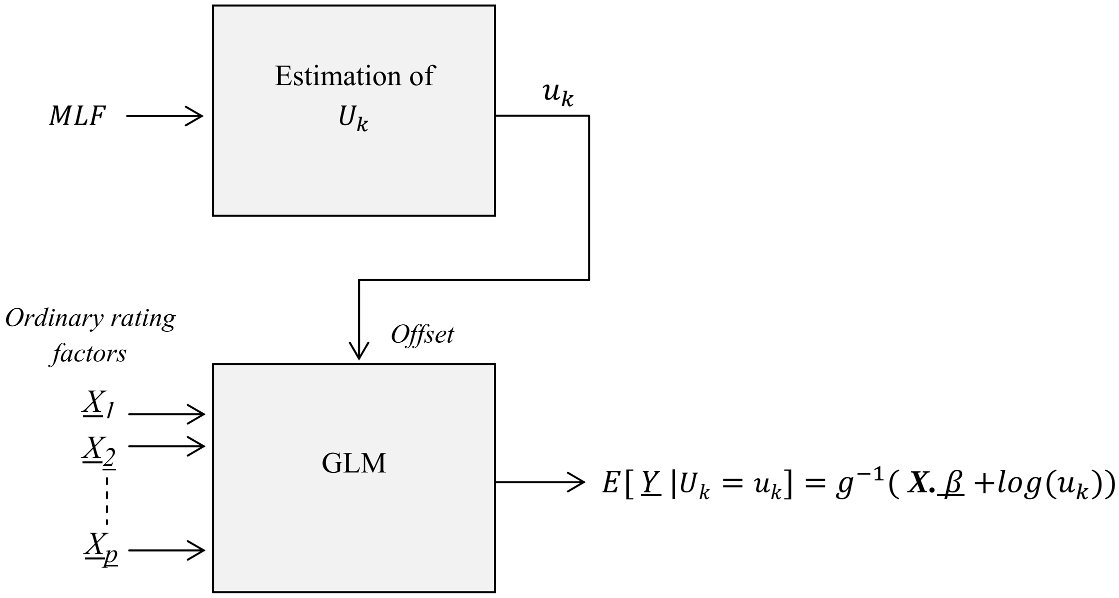
$$\underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{\mathrm{n}} \end{pmatrix} \qquad \underline{X_j} = \begin{pmatrix} X_{\mathrm{j},1} \\ X_{\mathrm{j},2} \\ \vdots \\ X_{\mathrm{j,n}} \end{pmatrix} \quad \underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{\mathrm{p+K}} \end{pmatrix}$$

$$\boldsymbol{X} = \begin{pmatrix} 1 & X_{1,1} & X_{2,1} & \cdots & X_{\mathrm{p+K},1} \\ 1 & X_{1,2} & X_{2,2} & \cdots & X_{\mathrm{p+K},2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1,\mathrm{n}} & X_{2,\mathrm{n}} & \cdots & X_{\mathrm{p+K,n}} \end{pmatrix}$$

In the standard GLM, we would introduce the MLF as an ordinary factor. This factor presents $K$ levels, hence the GLM will the estimate a $\beta_{\mathrm{j}}$ for $\mathrm{j} = \mathrm{p}+1$ to $\mathrm{p}+\mathrm{K}$. Yet due to the high number of levels of the MLF, the levels with few exposure will have an estimated β with little credibility.

**Revisited GLM**

We highlight that we will define the estimation of $U_k$ in the next sections.




with:

$$log(u_k) = \begin{pmatrix} log(u_{1,1}) = log(u_1) \\ log(u_{1,2}) = log(u_2) \\ \vdots \\ log(u_{1,K}) = log(u_K) \\ log(u_{2,1}) = log(u_1) \\ log(u_{2,2}) = log(u_2) \\ \vdots \\ log(u_{n,K}) = log(u_K) \end{pmatrix}$$

We observe that the first $K$ terms correspond to the subdivision of the tariff cell 1 into $K$ make models: $u_1 \rightarrow u_{1,1}\ ;\ u_{1,2}\ ;\ \dots\ ;\ u_{1,K}$

Given the offset term is defined by make model segment, we naturally observe the offset term repeated for the second tariff cell terms; hence $log(u_{1,1}) = log(u_{2,1}) = log(u_1)$

$$\underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{\mathrm{nK}} \end{pmatrix} \quad \underline{X_j} = \begin{pmatrix} X_{\mathrm{j},1} \\ X_{\mathrm{j},2} \\ \vdots \\ X_{\mathrm{j,nK}} \end{pmatrix} \quad \underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{\mathrm{p}} \end{pmatrix}$$

$$\boldsymbol{X} = \begin{pmatrix} 1 & X_{1,1} & X_{2,1} & \cdots & X_{\mathrm{p},1} \\ 1 & X_{1,2} & X_{2,2} & \cdots & X_{\mathrm{p},2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1,\mathrm{nK}} & X_{2,\mathrm{nK}} & \cdots & X_{\mathrm{p,nK}} \end{pmatrix}$$

In the revisited GLM, we do not introduce the MLF as an ordinary rating factor in the GLM. Instead we model it as a multiplicative random variable $U_k$ (see (3.6)). In fact we need to estimate $U_k$; hence once we know the estimated value of $U_k$ we can estimate the conditional mean $E[\, \underline{Y} \,|U_k = u_k]$. This is done by running a GLM on $\underline{Y}$ while setting $log(U_k)$ as an offset. We therefore expect the GLM that models the posterior mean to be better fit the experience compared to the standard GLM.

To run the revisited GLM, we initially need to estimate $U_k$; therefore the actual challenge here is to first be able to get a good estimate of the random variable $U_k$.

In fact we will see in the following sections that to estimate $U_k$, we will use the estimated mean $\mu_i$ of the tariff cells from the GLM; moreover we will see that by choosing a specific type of distribution function for $U_k$, its estimation is equivalent to the Buhlmann linear credibility method.

## Section 2: Extension of Jewell's theorem

To predict the $U_k{}'$s, we will first define its distribution function. Ohlsson and Johansson propose to follow Jewell (1974) and use the natural conjugate prior of the distribution in (3.9) that they define as:

$$f_\Theta(\theta) = \frac{1}{c(\delta, \alpha)} exp\left\{\frac{\theta\delta - b(\theta)}{1/_\alpha}\right\} \tag{3.11}$$

$c(\delta, \alpha)$ is just a normalising constant. The distribution is defined if $\alpha > 0$ and $\delta > 0$.

We will see in this section that by using this distribution function for the risk parameter in our Bayesian credibility model, this will be equivalent to use Buhlmann linear credibility model when estimating the MLF random variables.

Before demonstrating the extension of Jewell's theorem, Ohlsson and Johansson propose an important Lemma that will be useful:

Lemma 1: If we consider $\Theta = h(U_k)$, defined over the interval M in (III.2) and follows the distribution (3.7), we have:

(a) $\delta = E[U]$

(b) $\alpha = {E[U^p]}\big/{Var(U)}$ (3.12)

The demonstration of the Lemma is available in the **Appendix III.2**.

We can highlight that from (3.13) we can say that $\delta = 1$, hence we just have one parameter in the conjugate distribution which is $\alpha$.

Moreover we have the following assumptions:

- $\Theta_k$; $k = 1,2,\dots,K$ are independent and identically distributed random variables.
- For $k = 1,2,\dots,K$, the pairs $(Y_{ik};\Theta_k)$ are independent.
- Conditional on $\Theta_k$ the random variables $Y_{1k}, Y_{2k}, \dots, Y_{Ik}$ are independent.

From (3.7) we know that $E[\mathrm{Y_{ik}}|\mathrm{U_k}] = \mu_i \mathrm{U_k}$, with $\mu_i$ is the mean given by the ordinary rating factors that is estimated by GLM given the $\mathrm{U_k}$ as offset. Therefore in our case we first need to estimate a predictor of $\mathrm{U_k}$ using a credibility estimate for every $k$.

In analogy with classical credibility model, the resolution of this model consists in solving the minimisation of the square error $E[\mathrm{U_k} - g(\underline{Y})]^2$ with $\underline{Y} = \{Y_{1k}, Y_{2k}, \dots, Y_{Ik}\}$.

The solution of this minimization problem is $g(\underline{Y}) = E[\mathrm{U_k}|\underline{Y}]$. The demonstration of this has already been done in **Appendix II.2**.

By replacing $\mathrm{U_k}$ with the risk parameter $\Theta_k$, we obtain $g(\underline{Y}) = E[b'(\Theta_k)|\underline{Y}]$. We will try to re-write the expression of this a-posteriori mean using the extension of Jewell's theorem. We will re-write the Bayesian relation (2.13) as follow:

$f_{\Theta_k|Y_k}[\theta|y_k]$ is the a-posteriori distribution of $\Theta_k$, which is proportional to the product of the a-priori distribution of $\Theta_k$, $f_{\Theta_k}[\theta]$ and the conditional probability of $Y_k$ knowing $\Theta_k$, $f_{Y_k|\Theta_k}[y_k|\theta]$ (also known as the "likelihood function"). We will write:

$$f_{\Theta_k|Y_k}[\theta|y_k] \propto f_{\Theta_k}[\theta] \cdot f_{Y_k|\Theta_k}[y_k|\theta] \tag{3.13}$$

$$= f_{\Theta_k}[\theta] \prod_{i=1}^{I} f_{Y_{ik}|\Theta_k}[y_{ik}|\theta]$$

From (3.10) and (3.11), we have:

$$= \frac{1}{c(\delta,\alpha)} exp\left\{\frac{\theta\delta - b(\theta)}{1/\alpha}\right\} \prod_{i=1}^{I} exp\left\{\frac{\omega_{ik}}{\phi}\left[\frac{y_{ik}}{\mu_i{}^{p-1}}\theta - \frac{1}{\mu_i^{p-2}}b(\theta)\right] + c_2\right\}$$

Given $\delta = 1$, moreover we can get rid of the constant $c(\delta,\alpha)$ and $c_2$:

$$\propto exp\{\alpha(\theta - b(\theta))\} \prod_{i=1}^{I} exp\left\{\frac{\omega_{ik}}{\phi}\left[\frac{y_{ik}}{\mu_i{}^{p-1}}\theta - \frac{1}{\mu_i^{p-2}}b(\theta)\right]\right\}$$

$$= \; exp\left\{\alpha\big(\theta - b(\theta)\big) + \sum_{i=1}^{I}\frac{\omega_{ik}}{\phi}\left[\frac{y_{ik}}{\mu_i{}^{p-1}}\theta - \frac{1}{\mu_i^{p-2}}b(\theta)\right]\right\}$$

$$= \ exp\left\{\theta\left(\alpha + \frac{1}{\phi}\sum_{i=1}^{I}\omega_{ik}\frac{y_{ik}}{\mu_i^{p-1}}\right) - b(\theta)\left(\alpha + \frac{1}{\phi}\sum_{i=1}^{I}\omega_{ik}\mu_i^{2-p}\right)\right\}$$

$$= exp\left\{\frac{\theta\delta' - b(\theta)}{1/_{\alpha'}}\right\}$$

With

$$\alpha' = \alpha + \frac{1}{\phi}\sum_{i=1}^{I}\omega_{ik}\mu_i^{2-p} \tag{3.14}$$

$$\delta' = \frac{\left(\alpha + \frac{1}{\phi}\sum_{i=1}^{I}\omega_{ik}\frac{y_{ik}}{\mu_i^{p-1}}\right)}{\left(\alpha + \frac{1}{\phi}\sum_{i=1}^{I}\omega_{ik}\mu_i^{2-p}\right)} \tag{3.15}$$

We can notice that the posterior distribution (3.14) is a member of the same family as the prior distribution (3.11), given the conjugate prior.

Based on the Lemma 1 (a), we can therefore conclude that the expectation of $U_k$ in the posterior distribution is equal to $\delta'$. We can therefore re-write the optimal predictor of $E[\mathrm{U_k}|\underline{Y}]$, noted $\hat{u}_k$ as follow:

$$\hat{u}_k = \delta' = \frac{\left(\sum_{i=1}^{I}\omega_{ik}\frac{y_{ik}}{\mu_i^{p-1}}\right) + \phi\alpha}{\left(\sum_{i=1}^{I}\omega_{ik}\mu_i^{2-p}\right) + \phi\alpha} \tag{3.16}$$

We can rewrite this relation in the form of the classic credibility estimator. If we introduce the weighted average

$$\bar{u}_k = \frac{\sum_{i=1}^{I}(\omega_{ik}\mu_i^{2-p})\frac{y_{ik}}{\mu_i}}{\sum_{i=1}^{I}\omega_{ik}\mu_i^{2-p}} \tag{3.17}$$

$\bar{u}_k$ is the equivalent of the observed risk premium of the insured in Buhlmann model; it is the experience factor of the level $k$ of the MLF, indicating how one might adjust the expected values $\mu_i$ to take into account the experience $\mu_{ik}$.

We notice that for $p = 1$, we have $\bar{u}_k = \frac{\sum_{i=1}^{I}\omega_{ik}y_{ik}}{\sum_{i=1}^{I}\omega_{ik}\mu_i}$ (3.17.b)

We can then write the linear relation:

$$\hat{u}_k = z_k \bar{u}_k + (1 - z_k) \times 1 \tag{3.18}$$

with the credibility factor $z_k$ as follow:

$$z_k = \frac{\sum_{i=1}^{I} \omega_{ik} \mu_i^{2-p}}{\left(\sum_{i=1}^{I} \omega_{ik} \mu_i^{2-p}\right) + \phi\alpha} \tag{3.19}$$

We can rewrite $z_k$ in the classical credibility form by introducing the variance parameters $a$ and $\nu$ similar to the Buhlmann-Straub model which become:

$$a_i = Var(E[Y_{ik}|\Theta_k])$$

$$\nu_i = E[\omega_{ik} Var(Y_{ik}|\Theta_k)]$$

$a_i$ represents the variance between the levels of the MLF and $\nu_i$ represents the variance within the levels of the MLF.

From (3.6), we have $\quad a_i = \mu_i^2 Var(U_k) \quad$ (3.20)

Moreover from (3.6), (3.3) and (3.5), we have:

$$Var(Y_{ik}|\Theta_k) = \phi \frac{\mu_i^p u_k^p}{\omega_{ik}} \tag{3.21}$$

hence we obtain:

$$\nu_i = \phi \mu_i^p E[U_k^p] \tag{3.22}$$

From Lemma 1 (b), we can write $\frac{a_i}{\nu_i} = \frac{\mu_i^{2-p}}{\phi\alpha}$

then from (3.19) we can write

$$z_k = \frac{\phi\alpha \sum_{i=1}^{I} \omega_{ik} \, {}^{a_i}/_{\nu_i}}{\left(\phi\alpha \sum_{i=1}^{I} \omega_{ik} \, {}^{a_i}/_{\nu_i}\right) + \phi\alpha}$$

$$z_k = \frac{\sum_{i=1}^{I} \omega_{ik} \, {}^{a_i}/_{\nu_i}}{\left(\sum_{i=1}^{I} \omega_{ik} \, {}^{a_i}/_{\nu_i}\right) + 1} \tag{3.23}$$

We can recognize the formulation of the credibility factor similar as in Buhlmann-Straub model. We highlight that the credibility increases with the volume of exposure $\omega_{ik}$ or variance "between" $a_i$ , while it decreases with the variance "within" $\nu_i$ in line with Buhlmann linear credibility factor.

We have in fact demonstrated the extension of Jewell's theorem, i.e. by choosing the risk parameter a-prior distribution in (3.11), the estimation of the Bayesian a-posteriori risk parameter is equivalent to Buhlmann linear credibility model.

## Section 3: Estimation of the parameters

The Bayesian approach that we adopted in our model implies an iterative estimation approach of the parameters. In fact in order to run the standard GLM on ordinary factors, we first need to estimate the random variable $u_k$ to use it as an offset in the GLM. At the same time, in order to calculate the predictor of $u_k$, we need to estimate the GLM output $\mu_i$ as seen in the prior section. We will therefore use an iterative method, which is adequate for this Bayesian model.

We will first estimate the $\mu_i$; we will then use this estimate to calculate $u_k$. We will use the estimate of $u_k$ as an offset and re-estimate the $\mu_i$ based on the GLM. We will re-use the new estimated $\mu_i$ to re-calculate $u_k$, etc. The iteration can be stopped once the model converged, which means once we estimate that any additional iteration does not improve the fit of the model. Ohlsson and Johansson suggest the following iteration process:

0. Initially put $\hat{u}_k = 1$ for all $k$.

1. Estimate the $\mu_i$ in a GLM with all ordinary rating factors as explanatory variables, using a log-link and having $log(\hat{u}_k)$ as offset variable (offset will be nil in the first iteration).

2. Estimate $\phi\alpha$ using $\hat{\mu}_i$ from Step 1. This step requires to estimate $\alpha$, which will be described below.

3. Calculate $\hat{u}_k$ for all $k$, using estimates from Step 1 and 2.

4. Return to Step 1 with the offset variable $log(\hat{u}_k)$ from Step 1.

   Repeat Step 1 to 4 until convergence of the model.

In our case we will consider that the $u_k$ vector converges when we obtain for all $k = 1 \; to \; K$, $abs\left({u_k}^{(t+1)} - {u_k}^{(t)}\right) < 0.001$, with $t$ being the iteration indicator.

Ohlsson and Johansson suggest using in Step 2 a MSE approach to estimate unbiased estimator for $\hat{u}_k$. This approach is in analogy with the estimation procedure used in linear credibility.

The idea is to compute separate unbiased estimators of $\sigma^2 = \phi E[U^p]$ and $\sigma_U^2 = Var(U)$, whose ratio is equal to $\phi\alpha$. Ohlsson and Johansson propose the following estimators:

$$\hat{\sigma}^2 = \frac{\sum_{k=1}^{K}(I_k - 1)\hat{\sigma}_k^2}{\sum_{k=1}^{K}(I_k - 1)}$$

with $I_k$ the number of tariff cells where we have $\omega_{ik} > 0$. (3.24)

$$\hat{\sigma}_k^2 = \frac{1}{I_k - 1}\sum_{i=1}^{I} \omega_{ik} \mu_i{}^{2-p} \left(\frac{Y_{ik}}{\mu_i} - \bar{u}_k\right)^2$$

and the estimator of $\hat{\sigma}_U^2$ as follow:

$$\hat{\sigma}_U^2 = \frac{\sum_{k=1}^{K}\sum_{i=1}^{I} \omega_{ik} \mu_i{}^{2-p}(\bar{u}_k - 1)^2 - K\hat{\sigma}^2}{\sum_{k=1}^{K}\sum_{i=1}^{I} \omega_{ik} \mu_i{}^{2-p}} \qquad (3.25)$$

We can therefore estimate:

Variance within

$$\phi\alpha = \frac{\hat{\sigma}_k^2}{\hat{\sigma}_U^2}$$

Variance between

We can highlight that even though the GLM dispersion parameter $\phi$ is included in some equations, it does not enter in the estimation of the random variable, hence $\hat{u}_k$ does not depend on $\phi$.

For more details around the demonstration of how to obtain those estimators please refer to **Appendix III.3**.

We can illustrate the use of the revisited GLM model in the following example:

We have performed a GLM on an insurance product based on two rating factors $R_1$and $R_2$; hence we obtain the following loss cost model results:

| base |
|---|
| 10.0% |

| R1 | |
|---|---|
| r11 | 1.00 |
| r12 | 1.17 |
| r13 | 1.30 |

| R2 | |
|---|---|
| r21 | 1.00 |
| r22 | 1.07 |

We can translate these relativities into tariff cells estimated frequency $\mu_i$ as follow, with $i = 1\ to\ 6$:

| i | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\mu_i$ | 10.0% | 11.7% | 13.0% | 10.7% | 12.5% | 13.9% |

For example, $\mu_1$ is obtained as follow:

$\mu_1 = 10.0\%\ (base) * r11 * r21 = 10.0\% * 1.00 * 1.00$, and so on to calculate the other tariffs cells.

Meanwhile, we have a factor that we will consider a Multi-level factor, which presents six levels (from $A$ to $F$). The following charts show the exposure volumes $\omega_{ik}$ and observed frequency $y_{ik}$ for each tariff cells and MLF level combinations, with $i = 1\ to\ 6$, and $k = A\ to\ F$.

| $\boldsymbol{\omega_{ik}}$ | i | | | | | |
|---|---|---|---|---|---|---|
| k | 1 | 2 | 3 | 4 | 5 | 6 |
| A | 1000 | 1074 | 791 | 920 | 652 | 1350 |
| B | 589 | 158 | 655 | 408 | 974 | 358 |
| C | 2354 | 69 | 628 | 1864 | 2047 | 1265 |
| D | 974 | 977 | 458 | 1050 | 614 | 106 |
| E | 654 | 1810 | 602 | 1760 | 2013 | 900 |
| F | 175 | 97 | 105 | 40 | 35 | 36 |

| $\boldsymbol{y_i}$ | i | | | | | |
|---|---|---|---|---|---|---|
| k | 1 | 2 | 3 | 4 | 5 | 6 |
| A | 8.0% | 10.2% | 11.2% | 8.1% | 10.7% | 11.9% |
| B | 8.2% | 10.4% | 12.5% | 9.8% | 11.8% | 12.9% |
| C | 10.9% | 12.1% | 13.8% | 11.0% | 13.0% | 15.2% |
| D | 12.3% | 13.3% | 15.5% | 12.0% | 14.3% | 18.4% |
| E | 10.4% | 11.8% | 14.2% | 11.1% | 12.2% | 15.0% |
| F | 13.8% | 15.2% | 17.2% | 13.9% | 16.3% | 18.8% |

Let's calculate the estimated $\hat{u}_k$ factors based on the credibility theory. We will only do one iteration in this example.

We first will estimate $\bar{u}_k$ as mentioned in (3.17.b), hence we obtain:

| k | $\bar{u}_k$ |
|---|---|
| A | 0.837 |
| B | 0.920 |
| C | 1.061 |
| D | 1.166 |
| E | 1.024 |
| F | 1.335 |

For example for $\bar{u}_A$, we calculated as follow:

$$\bar{u}_A = \frac{1000 * 8.0\% + 1074 * 10.2\% + \cdots + 1350 * 11.9\%}{1000 * 10.0\% + 1074 * 11.7\% + \cdots + 1350 * 13.9\%} = 0.837$$

We can then estimate the variance within component based on (3.24):

| k | $\hat{\sigma}_k^2$ |
|---|---|
| A | 0.214 |
| B | 0.164 |
| C | 0.146 |
| D | 0.233 |
| E | 0.291 |
| F | 0.013 |

For example

$$\hat{\sigma}_A^2 = \frac{1}{(6-1)}(1000 * 10.0\% * (\frac{8.0\%}{10.0\%} - 0.837)^2 + \cdots + 1350 * 13.9\% * \left(\frac{11.9\%}{13.9\%} - 0.837)^2\right) = 0.214$$

And we finally obtain $\hat{\sigma}^2$as follow:

$$\hat{\sigma}^2 = 0.214 + 0.164 + \cdots + 0.013 = 1.060$$

To ease the calculations, we will calculate the factor $A_k = \sum_{i=1}^{6} \omega_{ik} \mu_i$

| k | $A_k$ |
|---|---|
| A | 696.3 |
| B | 377.9 |
| C | 956.8 |
| D | 475.2 |
| E | 920.9 |
| F | 56.2 |

For example $A_A = (1000 * 10.0\% + 1074 * 11.7\% + \cdots + 1350 * 13.9\%) = 696.3$

and we can calculate the factor $B_k = \sum_{k=A}^{F} A_k * (\bar{u}_k - 1)^2$

| k | $B_k$ |
|---|---|
| A | 18.423 |
| B | 2.417 |
| C | 3.545 |
| D | 13.076 |
| E | 0.529 |
| F | 6.305 |

This enables us to calculate the variance between as follow:

$$\hat{\sigma}_U^2 = \frac{\sum_{k=A}^{F} B_k - 6 * \hat{\sigma}^2}{\sum_{k=A}^{F} A_k} = \frac{44.3 - 6 * 1.060}{3483.4} = 44.3$$

We can then calculate

$$\phi\alpha = \frac{\hat{\sigma}^2}{\hat{\sigma}_U^2} = \frac{1.06}{44.3} = 97.4$$

We can finally estimate the credibility factors $Z_k$ as follow:

$$Z_k = \frac{A_k}{A_k + \phi\alpha}$$

| k | $Z_k$ |
|---|---|
| A | 0.877 |
| B | 0.795 |
| C | 0.908 |
| D | 0.830 |
| E | 0.904 |
| F | 0.366 |

and the estimated $\hat{u}_k$ factors based on (3.18):

| k | $\hat{u}_k$ |
|---|---|
| A | 0.857 |
| B | 0.936 |
| C | 1.055 |
| D | 1.138 |
| E | 1.022 |
| F | 1.123 |

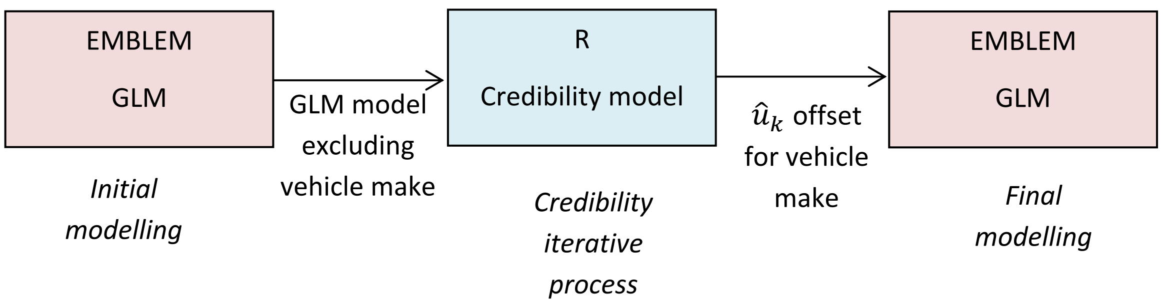
# Chapter 4: Application to our auto portfolio

## Section 1: Implementation of the model

The results of the GLM on the Vehicle make factor has shown the limitation of the method on MLFs (see Chapter 1 – Section 3). We therefore opted to test the credibility method on the vehicle make factor in order to improve the predictiveness of the frequency models.

While the EMBLEM software is very practical for GLM modelling, it is does not have the flexibility to implement alternative models. We therefore used R software which is more flexible and enables matrix calculations which we need to implement the credibility estimator calculation.

We therefore used the output from our initial GLM modelling for the estimation of the MLF factors $\hat{u}_k$ in R. We then re-plugged these $\hat{u}_k$ as offset of the vehicle make factor into Emblem for the final model validation. The following chart presents the different steps of the modelling.




The details of the R code used to estimate the $\hat{u}_k$ factors is shown in **Appendix IV.1.**

## Section 2: Results of the model

### 2.1 TPPD frequency model

The following table presents the results of the TPPD frequency credibility factors results.

**Table 4.1 – TPPD Frequency credibility estimators**

| Make memo | Weight | $\bar{u}_k$ | $z_k$ | $\hat{u}_k$ |
|---|---|---|---|---|
| FORD | 42,525 | 0.99 | 0.95 | 0.99 |
| SEAT | 24,016 | 0.96 | 0.91 | 0.96 |
| NISSAN | 20,937 | 0.94 | 0.89 | 0.95 |
| VW | 19,637 | 0.95 | 0.89 | 0.95 |
| MERCEDES | 19,198 | 1.05 | 0.89 | 1.04 |
| HONDA | 18,439 | 1.00 | 0.89 | 1.00 |
| MAZDA | 16,383 | 1.01 | 0.87 | 1.01 |
| DAEWOO | 14,700 | 0.99 | 0.86 | 0.99 |
| FIAT | 12,931 | 0.98 | 0.85 | 0.98 |
| OPEL | 12,249 | 1.02 | 0.84 | 1.02 |
| KIA | 10,558 | 0.99 | 0.80 | 0.99 |
| BMW | 10,045 | 1.05 | 0.81 | 1.04 |
| VOLVO | 9,580 | 0.96 | 0.80 | 0.97 |
| TOYOTA | 8,873 | 1.02 | 0.79 | 1.02 |
| LANCIA | 8,533 | 1.05 | 0.78 | 1.04 |
| SUZUKI | 7,731 | 1.11 | 0.76 | 1.09 |
| RENAULT | 7,064 | 0.98 | 0.74 | 0.98 |
| CITROEN | 6,829 | 0.98 | 0.74 | 0.99 |
| HYUNDAI | 6,635 | 0.87 | 0.72 | 0.91 |
| CHRYSLER | 5,987 | 1.12 | 0.72 | 1.09 |
| LEXUS | 4,938 | 1.04 | 0.66 | 1.03 |
| DAIHATSU | 4,533 | 0.90 | 0.66 | 0.93 |
| JEEP | 3,213 | 0.96 | 0.57 | 0.98 |
| PEUGEOT | 3,019 | 0.91 | 0.54 | 0.95 |
| AUDI | 2,890 | 1.08 | 0.55 | 1.04 |
| LAND ROVER | 2,773 | 1.41 | 0.55 | 1.22 |
| SKODA | 2,572 | 1.04 | 0.52 | 1.02 |
| SMART | 1,715 | 1.02 | 0.42 | 1.01 |
| CHEVROLET | 1,637 | 1.11 | 0.45 | 1.05 |
| Other | 1,521 | 0.86 | 0.37 | 0.95 |
| SAAB | 1,314 | 1.09 | 0.35 | 1.03 |
| LUXURY OTHER | 871 | 1.00 | 0.30 | 1.00 |
| ROVER | 819 | 0.88 | 0.27 | 0.97 |
| MITSUBISHI | 787 | 1.26 | 0.26 | 1.07 |
| MINI COOPER | 473 | 1.34 | 0.18 | 1.06 |
| ALFA ROMEO | 390 | 1.02 | 0.14 | 1.00 |
| SUBARU | 301 | 0.90 | 0.10 | 0.99 |
| PORSCHE | 259 | 1.05 | 0.10 | 1.00 |

We observe that the credibility factor increases with the weight, as shown in the graph below.

The result of the simulation gives us the following estimates:

- Variance factors: $\phi\alpha = 594.98$

- Overall estimated mean of the portfolio: $\mu = 9.80\%$

If we consider the definition of the credibility factor (3.19), with $p = 1$, we can estimate

$$z_k = \frac{\sum_{i=1}^{I} \omega_{ik} \mu_i}{(\sum_{i=1}^{I} \omega_{ik} \mu_i) + 594.98}$$

If we were to have all tariff cells at the same mean μ, we would obtain:

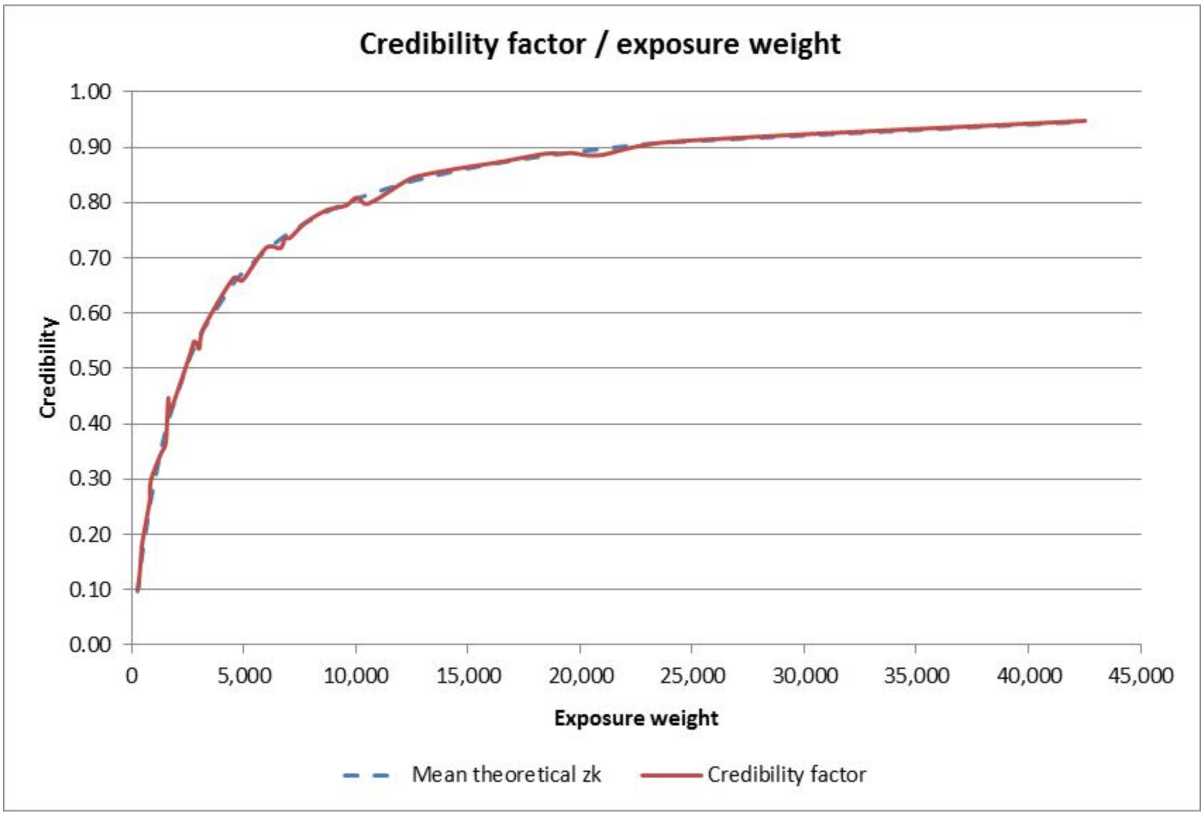
$$z_k = \frac{9.80\% * \sum_{i=1}^{I} \omega_{ik}}{9.80\% * (\sum_{i=1}^{I} \omega_{ik}) + 594.98}$$

This would correspond to the theoretical credibility factor in case the tariff cell mean $\mu_i$ does not deviate from the portfolio mean μ.

We highlight that to obtain 50% of credibility we would need 6071 exposure (based on a estimated mean $\mu_i$ equal to the portfolio mean μ).

We can plot both graph with regard to the exposure volume.

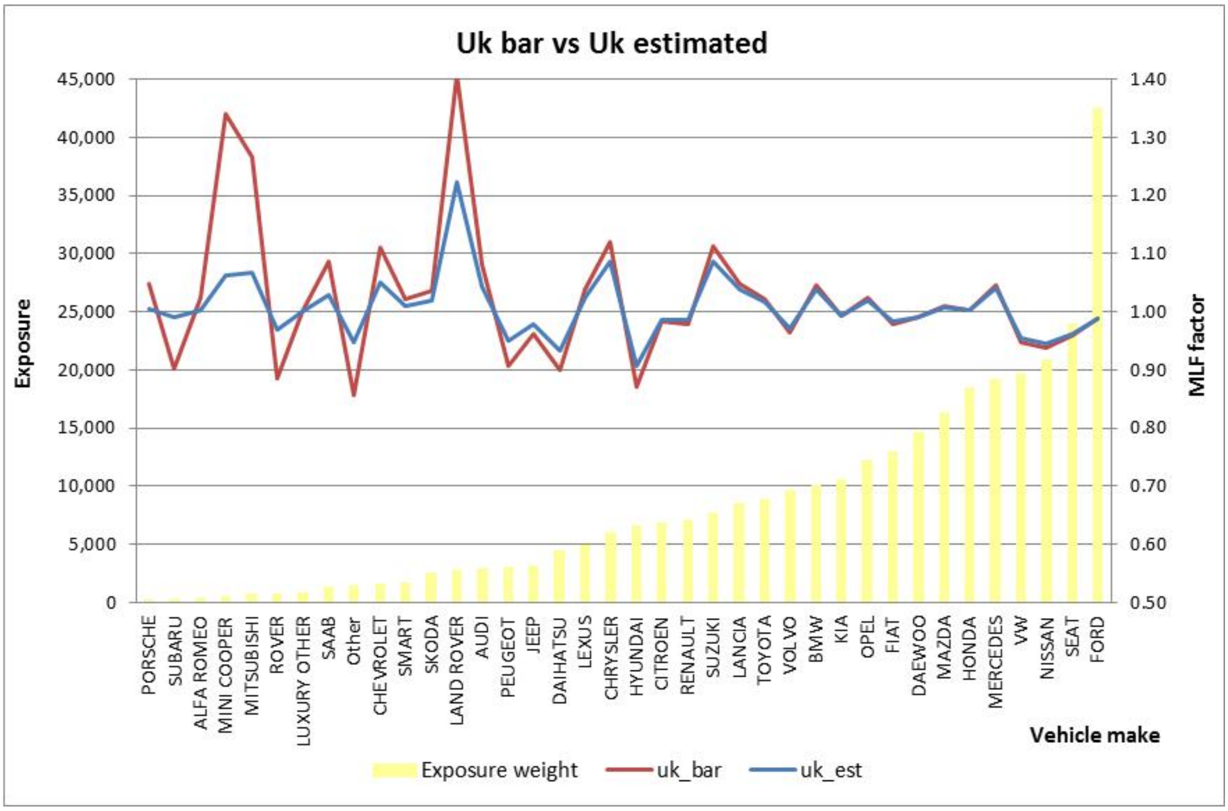
**Graph 4.1 – TPPD Frequency credibility factor by weight**




We observe that the credibility curve is not exactly monotonous, and fluctuates around the theoretical credibility curve. This is expected to be due to predicted mean term $\mu_i$ of the tariff cells that is present in the weight term of the credibility factor expression (see (3.19)); hence the credibility factor of the level $k$ can vary depending on its weight distribution by tariff cells. The model will therefore give more credibility to tariff cells with higher expected mean $\mu_i$, for similar exposure volume.

We can also observe the estimated factors $\hat{u}_k$ and $\bar{u}_k$ along the exposure weight.

**Graph 4.2 – TPPD Frequency MLF factors by weight**

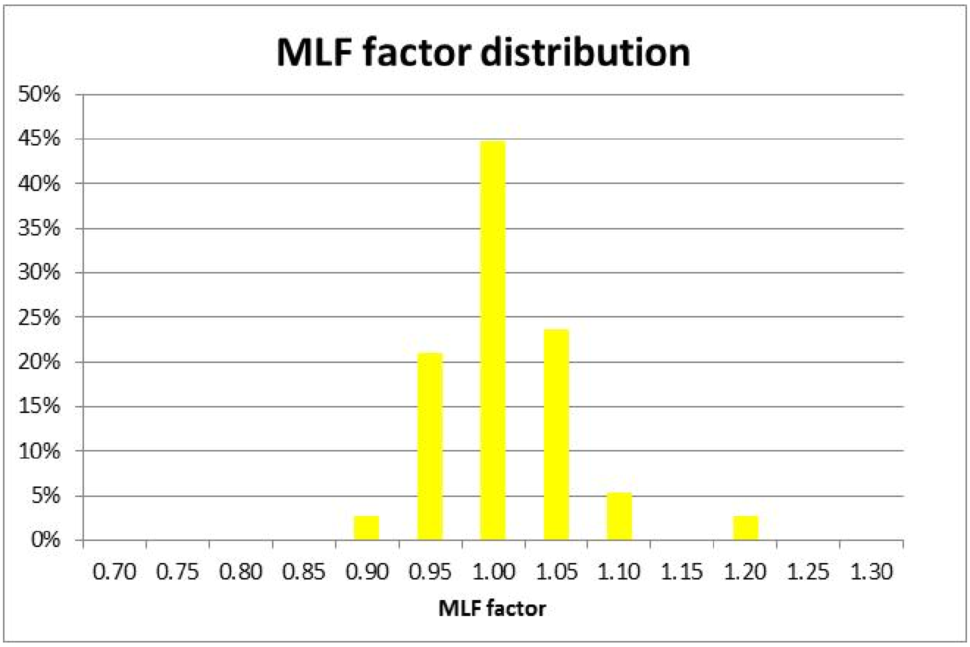



We observe that while the exposure increases, the credibility factor increases, the estimated $\hat{u}_k$ factors naturally converges to the observed factor $\bar{u}_k$. For vehicle make with few exposure we observe that the estimated $\hat{u}_k$ is weighted toward 1 from the initial observed $\bar{u}_k$.

Given we have included auxiliary factors in the GLM to estimate the tariff cell means $\mu_i$, they should capture the effects on the frequency; except the residual effect due to the MLF. Therefore the estimated $\hat{u}_k$ factors are here capturing the residual vehicle make effect once all other effects have been captured by the auxiliary rating factors selected in the initial GLM.

It is also interesting to analyse the distribution of the $\hat{u}_k$ estimates.

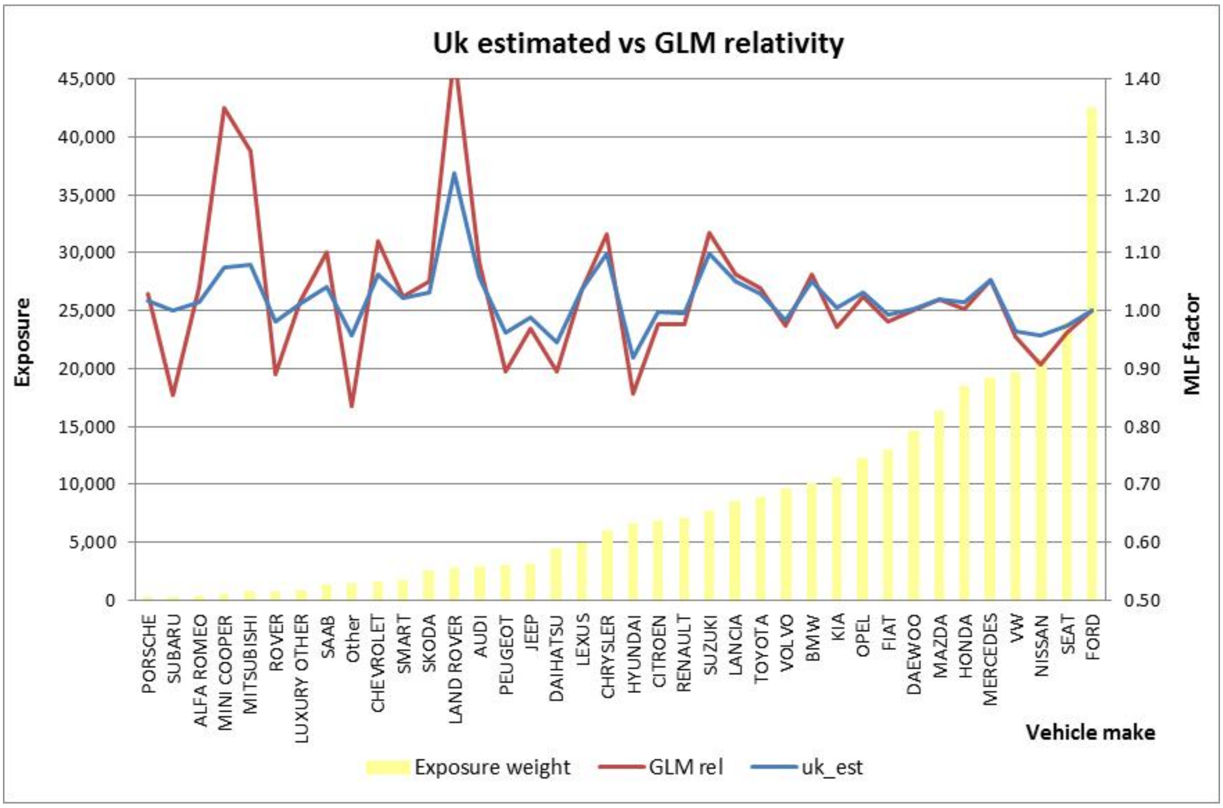
**Graph 4.3 – TPPD Frequency MLF factor distribution**




We observe that the MLF estimated factors are largely weighted toward 1, which means the auxiliary factors are already predicting a large part of the TPPD frequency.

We can also observe the re-based $\hat{u}_k$ factors along the estimated relativities if we had included the Vehicle Make as a regular rating factor in the GLM.

**Graph 4.4 – TPPD Frequency credibility factor vs GLM factor**



We observe that the credibility relativity $\hat{u}_k$ is converging toward the GLM result for levels with significant exposure volume. This actually makes sense; hence we can see that the main benefit of the credibility method is for the levels with few exposures, in which case the $\hat{u}_k$ factors are weighted more toward the portfolio mean (relativity of 1.00).

### 2.2 TPBI frequency model

We will analyse the similar outputs for the TPBI peril.

**Table 4.2 – TPBI Frequency credibility estimators**

| Make memo | Weight | $\bar{u}_k$ | $z_k$ | $\hat{u}_k$ |
|---|---|---|---|---|
| FORD | 42,525 | 1.08 | 0.10 | 1.01 |
| SEAT | 24,016 | 0.96 | 0.06 | 1.00 |
| NISSAN | 20,937 | 0.83 | 0.05 | 0.99 |
| VW | 19,637 | 1.12 | 0.05 | 1.01 |
| MERCEDES | 19,198 | 1.15 | 0.05 | 1.01 |
| HONDA | 18,439 | 0.87 | 0.04 | 0.99 |
| MAZDA | 16,383 | 0.83 | 0.04 | 0.99 |
| DAEWOO | 14,700 | 0.98 | 0.04 | 1.00 |
| FIAT | 12,931 | 1.12 | 0.03 | 1.00 |
| OPEL | 12,249 | 0.83 | 0.03 | 0.99 |
| KIA | 10,558 | 0.98 | 0.02 | 1.00 |
| BMW | 10,045 | 0.88 | 0.03 | 1.00 |
| VOLVO | 9,580 | 1.01 | 0.02 | 1.00 |
| TOYOTA | 8,873 | 0.82 | 0.02 | 1.00 |
| LANCIA | 8,533 | 1.07 | 0.02 | 1.00 |
| SUZUKI | 7,731 | 1.13 | 0.02 | 1.00 |
| RENAULT | 7,064 | 1.22 | 0.02 | 1.00 |
| CITROEN | 6,829 | 1.18 | 0.02 | 1.00 |
| HYUNDAI | 6,635 | 1.11 | 0.02 | 1.00 |
| CHRYSLER | 5,987 | 1.09 | 0.01 | 1.00 |
| LEXUS | 4,938 | 1.17 | 0.01 | 1.00 |
| DAIHATSU | 4,533 | 0.99 | 0.01 | 1.00 |
| JEEP | 3,213 | 1.05 | 0.01 | 1.00 |
| PEUGEOT | 3,019 | 0.55 | 0.01 | 1.00 |
| AUDI | 2,890 | 1.22 | 0.01 | 1.00 |
| LAND ROVER | 2,773 | 1.19 | 0.01 | 1.00 |
| SKODA | 2,572 | 0.89 | 0.01 | 1.00 |
| SMART | 1,715 | 1.06 | 0.00 | 1.00 |
| CHEVROLET | 1,637 | 1.06 | 0.00 | 1.00 |
| Other | 1,521 | 0.76 | 0.00 | 1.00 |
| SAAB | 1,314 | 0.51 | 0.00 | 1.00 |
| LUXURY OTHER | 871 | 0.87 | 0.00 | 1.00 |
| ROVER | 819 | 0.19 | 0.00 | 1.00 |
| MITSUBISHI | 787 | 0.90 | 0.00 | 1.00 |
| MINI COOPER | 473 | 0.66 | 0.00 | 1.00 |
| ALFA ROMEO | 390 | 0.83 | 0.00 | 1.00 |
| SUBARU | 301 | 0.58 | 0.00 | 1.00 |
| PORSCHE | 259 | 2.59 | 0.00 | 1.00 |

We observe that all vehicle makes are given a very low credibility; hence the level with the highest volume of exposure (FORD) only gets 0.10 credibility. This is due to low mean of the portfolio with regard to the variance in the portfolio.

The result of the simulation gives us the following estimates:

- Variance factors: $\phi\alpha = 11786.43$
- Overall estimated mean of the portfolio: $\mu = 1.15\%$

If we consider the definition of the credibility factor (3.19), with $p = 1$, we can estimate

$$z_k = \frac{\sum_{i=1}^{I} \omega_{ik} \mu_i}{(\sum_{i=1}^{I} \omega_{ik} \mu_i) + 11786.43}$$

If we were to have all tariff cells at the same mean μ, we would obtain:

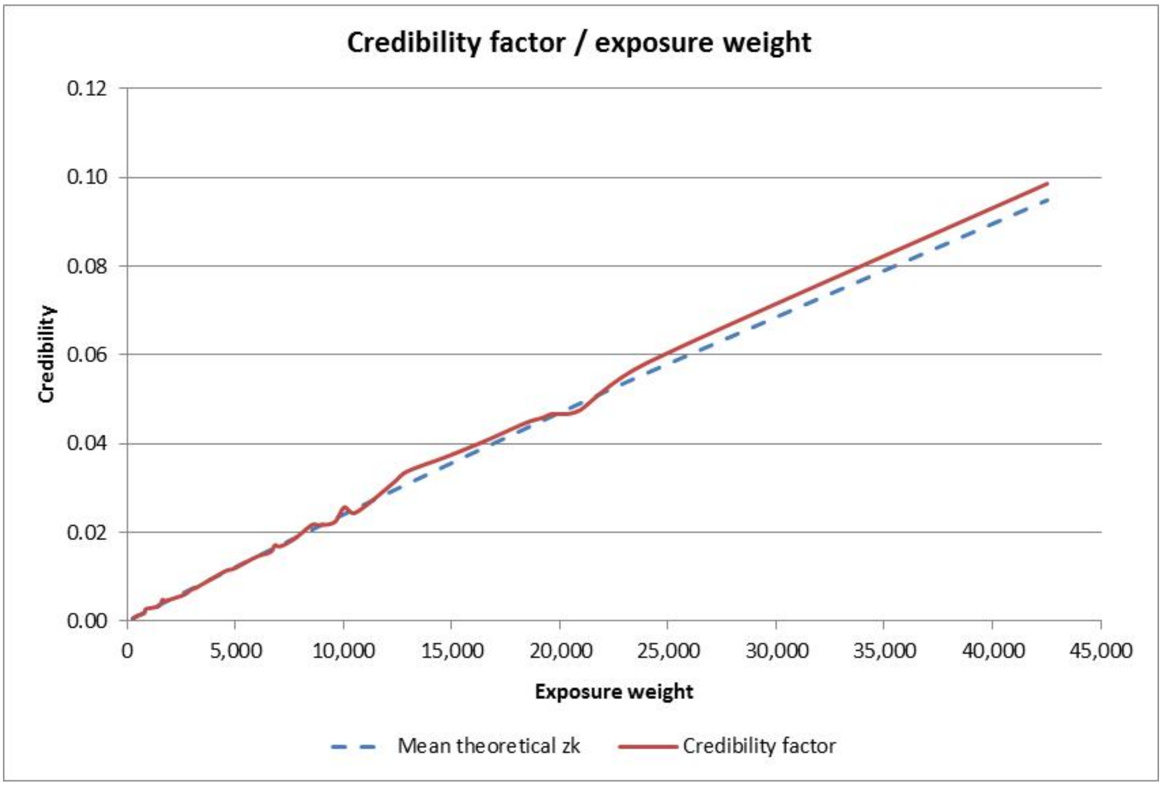
$$z_k = \frac{1.15\% * \sum_{i=1}^{I} \omega_{ik}}{1.15\% * (\sum_{i=1}^{I} \omega_{ik}) + 594.98}$$

This would correspond to the theoretical credibility factor in case the tariff cell mean $\mu_i$ does not deviate from the portfolio mean μ.

We highlight that to obtain 50% of credibility we would need 1,024,907 exposures (based on an estimated mean $\mu_i$ equal to the portfolio mean μ).

We can plot both graph with regard to the exposure volume.

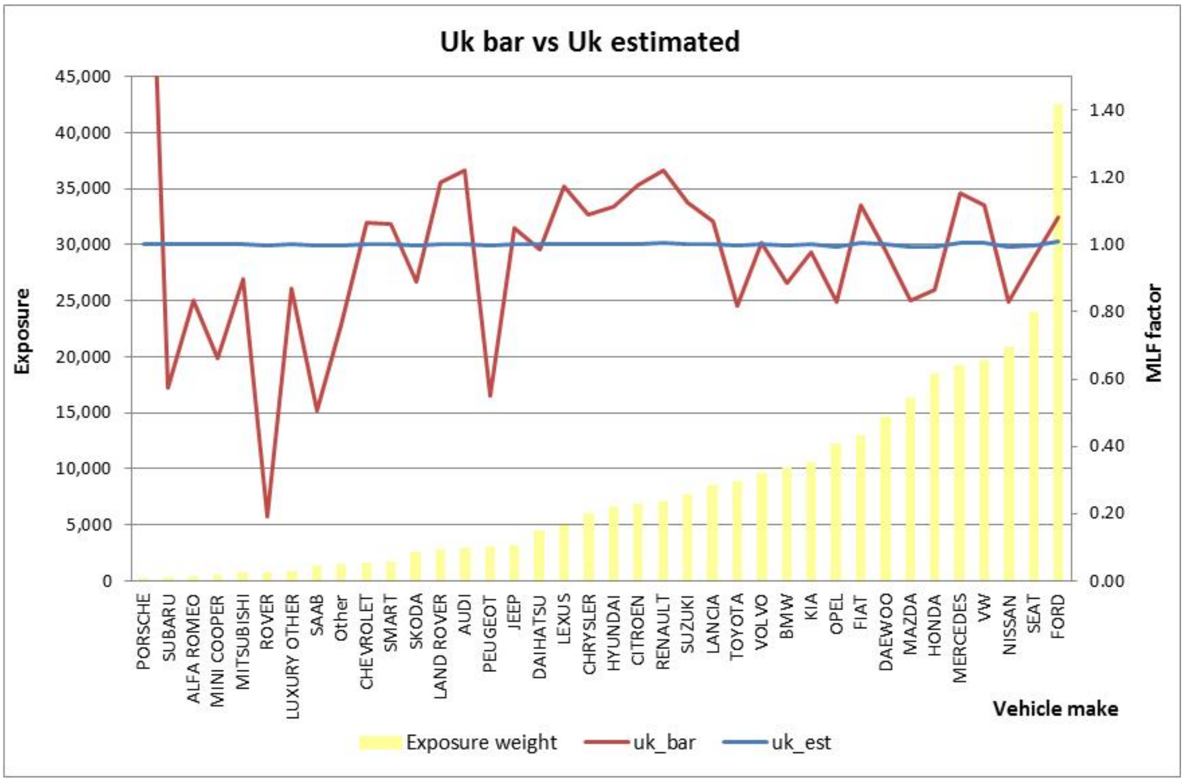
**Graph 4.5 – TPBI Frequency MLF factors by weight**




As observed, the mean of the portfolio is too low with regard to the variance within the portfolio, hence we do not have enough exposure volume by MLF level to get significant credibility.

We can also observe the estimated factors $\hat{u}_k$ and $\bar{u}_k$ along the exposure weight.

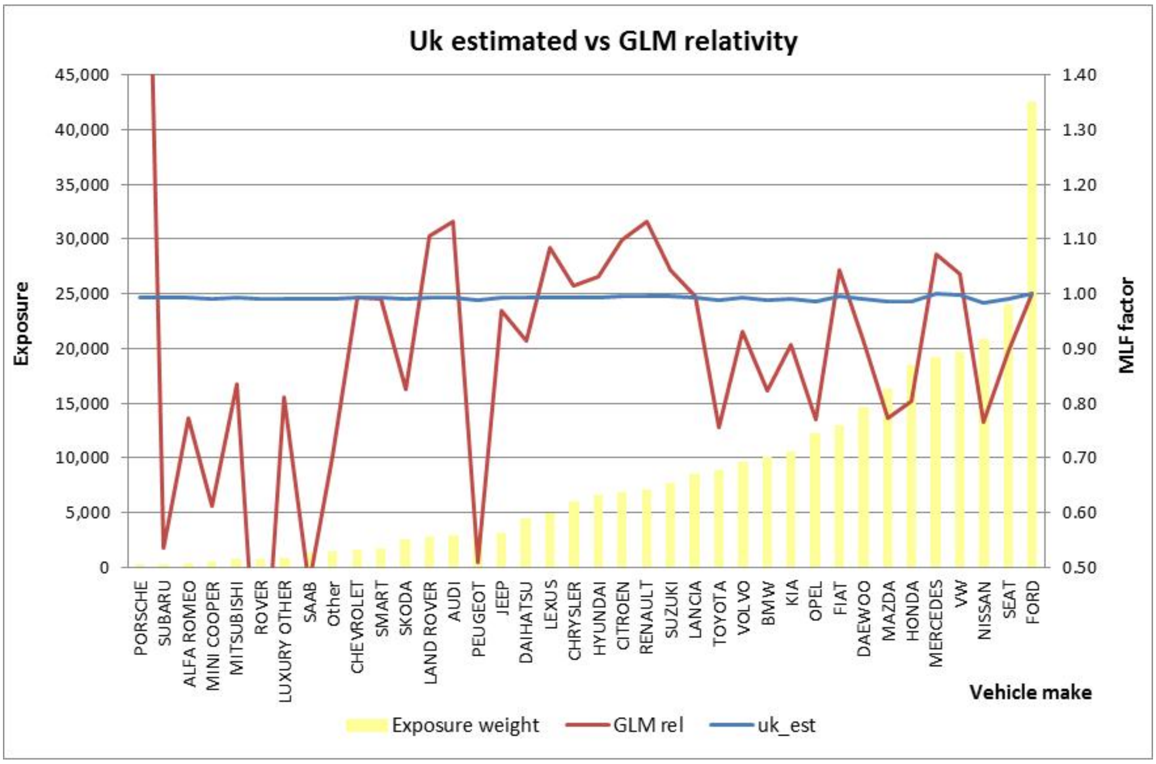
**Graph 4.6 – TPBI Frequency MLF factors by weight**




As a result the estimated MLF factors $\hat{u}_k$ are weighted very close toward 1, which means no credibility is given to the level experience.

The comparison with the GLM factor also shows the similar trend.

**Graph 4.7 – TPBI Frequency credibility factor vs GLM factor**



## Section 3: Validation of the model

In the initial GLM modelling, we have been able to validate the predictiveness of the selected rating variables based on various validation tests. We have then estimated the vehicle make factors based on the credibility methods. We now need to validate whether the revisited GLM using credibility approach gives better results in term of predictiveness.

For this, the best test is to compare the Lift curve and Gini coefficients diagnostics that we have presented in Chapter 1, for the models including and excluding the credibility approach. In fact performing one-way validation test on the MLF levels could lead to misleading interpretation given the lower volume of data by MLF level. We therefore prefer looking at those aggregate statistics.

We usually perform the validations on the test sample (ex: 20% of the data) that we didn't use for the GLM modelling. However a problem arises for portfolio with small volume of data such as in our case. Given we do not have a significant volume of data to perform an accurate GLM, it is even more detrimental to the model results to exclude 20% of the data on top. In order to tackle that issue, we will therefore perform the validation test based on an alternative method called *cross validation* or *k-fold* test.

The cross validation test consists in the following steps:

1. Perform a GLM modelling on 100% of the training dataset. From that step we are able to select our model structure.
2. Split the dataset into 5 randomly selected groups (Group 1, Group 2, …, Group 5).
3. Re-fit the model from step 1 on the data sample for Group 2 to 5 (only exclude Group 1).
4. Perform the validation test (Lift curve and Gini coefficient) of the fitted model from step 3 to the Group 1 data.
5. Perform step 3 and 4 for the other four groups.
6. Sum the 5 tests data together to produce an overall Lift curve and Gini curve.

Given we re-fit the model each time on four groups and perform the validation on the other $5^{th}$ group, the cross validation test gives consistent results.

For more details around the Cross validation test procedure, please refer to the **Appendix IV.2.**

In our study we have compared validation results for four different models:

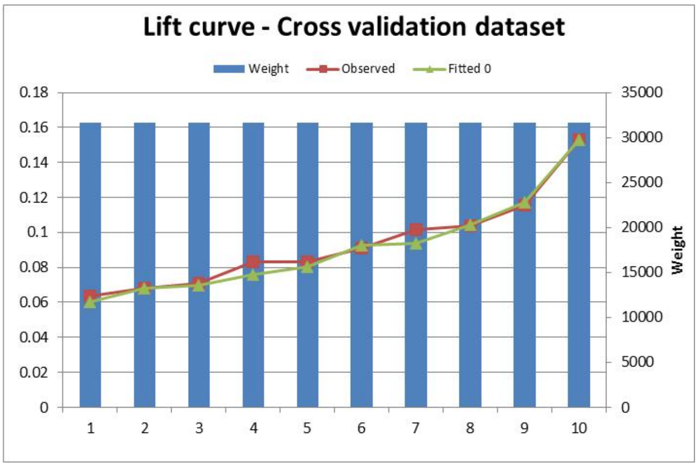
a. Model excluding the vehicle make factor
b. Model including the simple vehicle make factor in the GLM.
c. Model including the grouped vehicle make factor in the GLM.
d. Model based on the credibility approach.

For each model we will look at the following diagnostic:

- Gini coefficient
- Lift curve, for which we will estimate:
  - Model mean squared error.
  - Model $R^2$

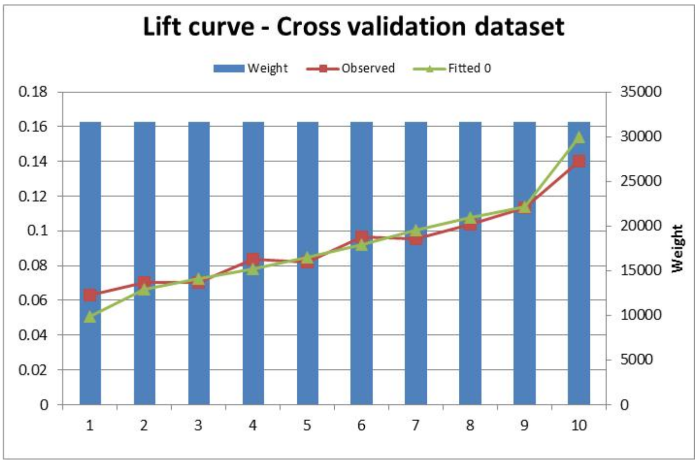
### 3.1 TPPD frequency model

a. Model excluding vehicle make



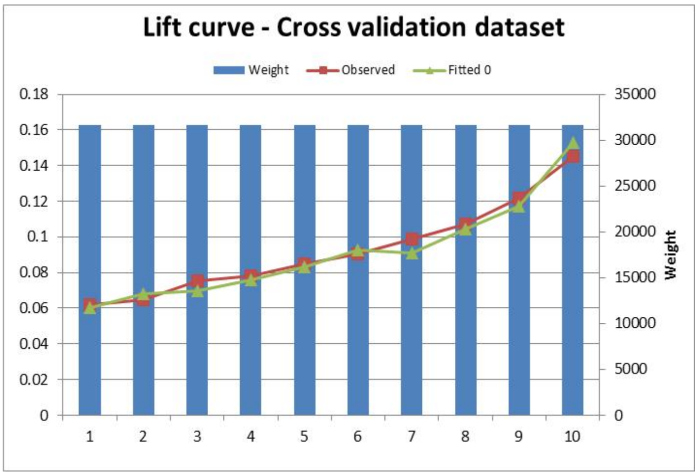

| | MSE | $R^2$ |
|---|---|---|
| By bucket | 9.6080E-06 | 0.9770 |

b. Model including vehicle make



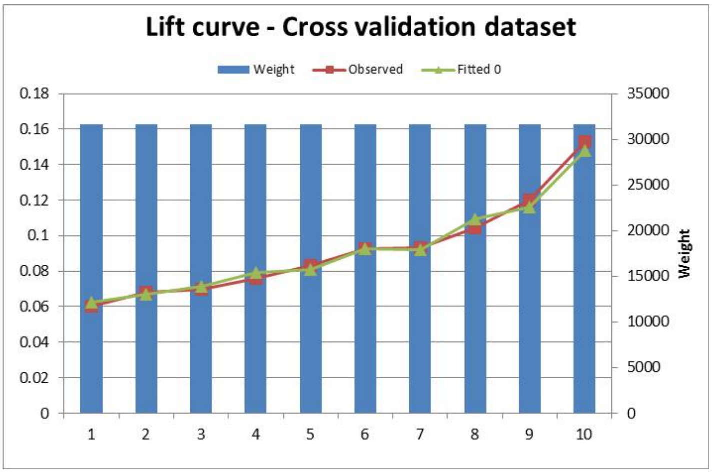

| | MSE | $R^2$ |
|---|---|---|
| By bucket | 1.2841E-05 | 0.9678 |

c. Model including grouped vehicle make




| | MSE | $R^2$ |
|---|---|---|
| By bucket | 5.6080E-06 | 0.9858 |

d. Model with credibility

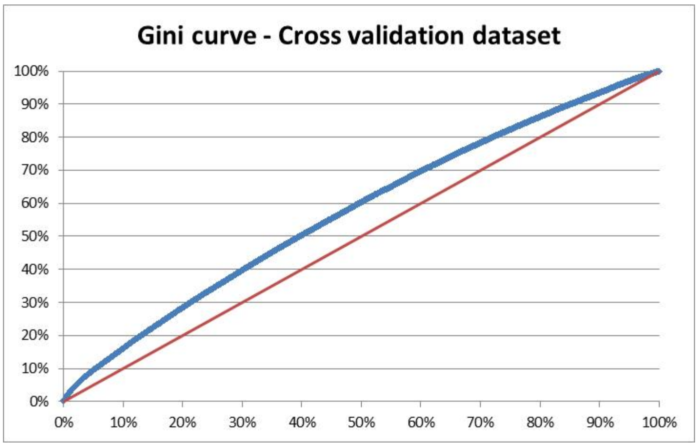



| | MSE | $R^2$ |
|---|---|---|
| By bucket | 3.2841E-06 | 0.9921 |

The results show that the model with the credibility method presents the lowest MSE and higher $R^2$ ; hence we can deduce that revisiting the GLM for MLFs improves the predictiveness of the model. In particular, we observe that the model including the simple factor gives the worse fit. This is driven by the fact that the model is likely to overfit, hence we observe the overfit on the lift curve extremity as the fitted curve diverge from the observed. Including the factor as grouped improves the fit, yet the credibility method enables to improve the fit even more.

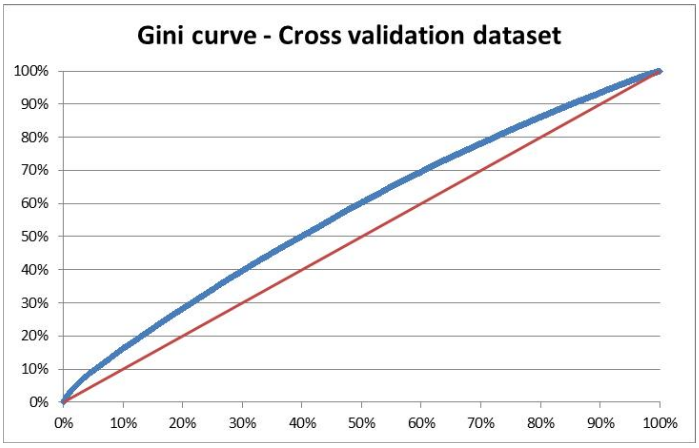
The Gini coefficients are shown similarly in the graphs below.

a. Model excluding vehicle make



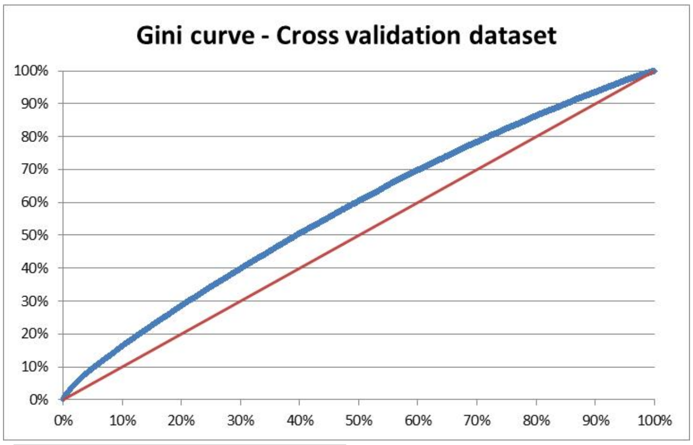

| | |
|---|---|
| Gini coeff | 15.10% |

b. Model including vehicle make



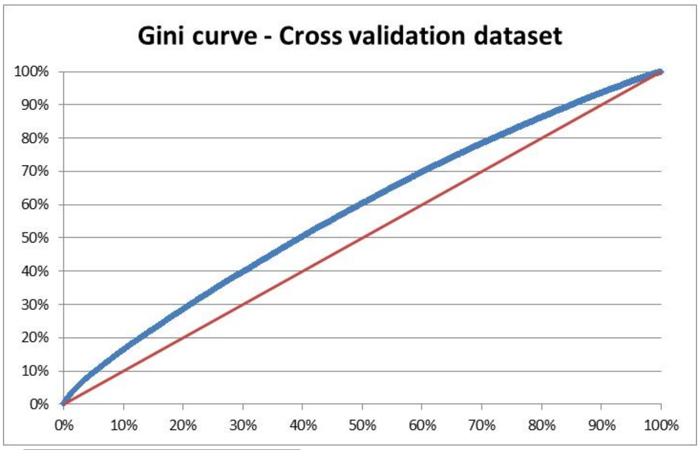

| | |
|---|---|
| Gini coeff | 14.67% |

c. Model including grouped vehicle make




| | |
|---|---|
| Gini coeff | 15.34% |

d. Model with credibility



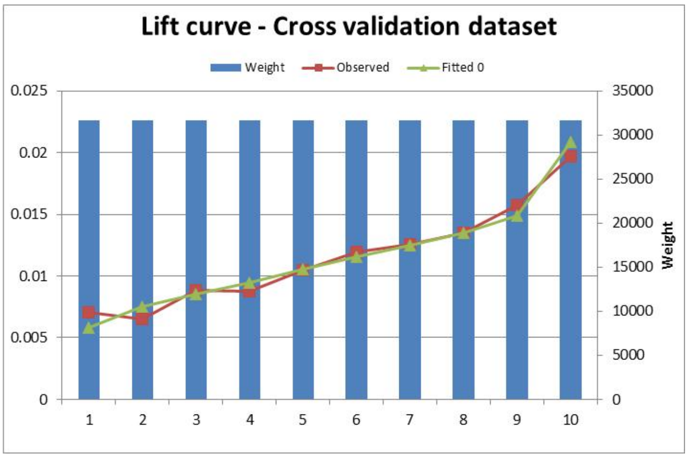

| | |
|---|---|
| Gini coeff | 15.53% |

We also observed that the Gini coefficient of the revisited GLM with credibility is higher; which means it is the model that best fit the TPPD frequency peril.

### 3.2 TPBI frequency model

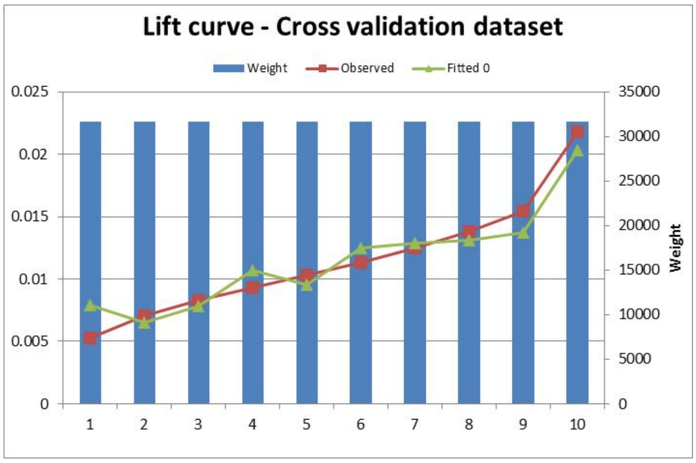
The following graphs show the lift curve of the cross validation model for the TPBI frequency.

a. Model excluding vehicle make



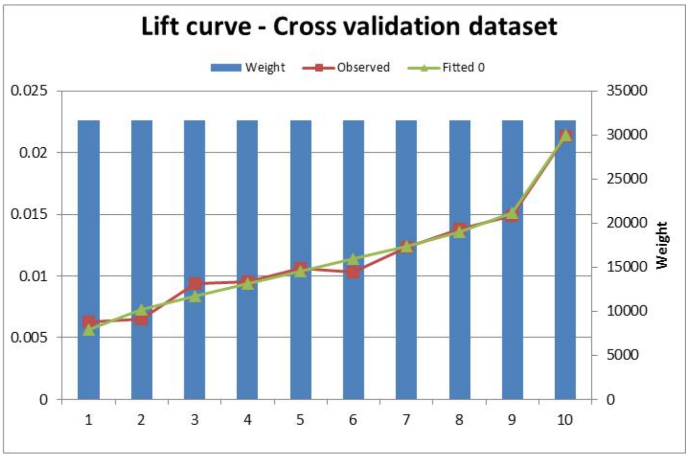

| | MSE | $R^2$ |
|---|---|---|
| By bucket | 2.0365E-08 | 0.9709 |

b. Model including vehicle make



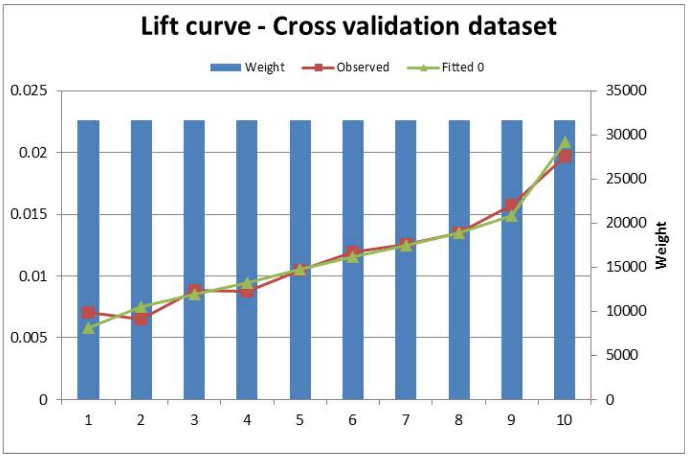

| | MSE | $R^2$ |
|---|---|---|
| By bucket | 8.0490E-08 | 0.8796 |

c. Model including grouped vehicle make




| | MSE | $R^2$ |
|---|---|---|
| By bucket | 2.4823E-08 | 0.9647 |

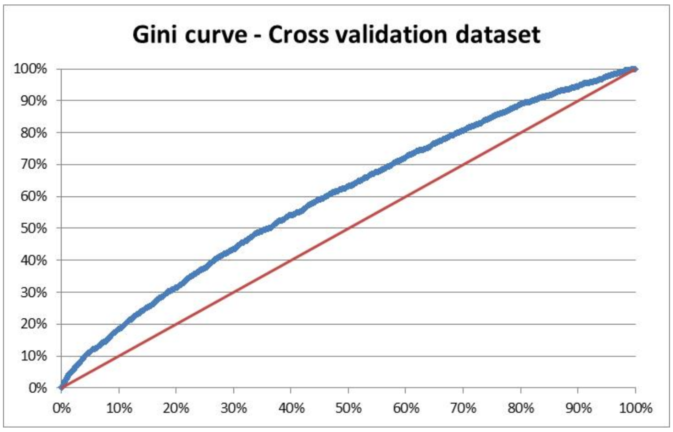
d. Model with credibility




| | MSE | $R^2$ |
|---|---|---|
| By bucket | 2.0165E-08 | 0.9722 |

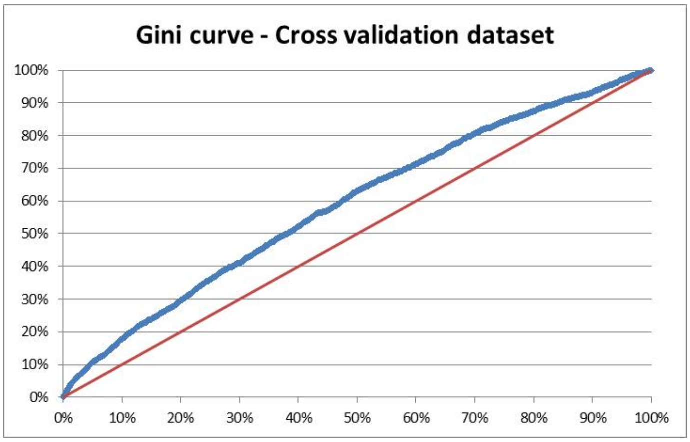
We observe that for TPBI peril as well the best lift curve statistics is achieved for the model including credibility. The Gini coefficients are shown similarly in the graphs below.

a. Model excluding vehicle make



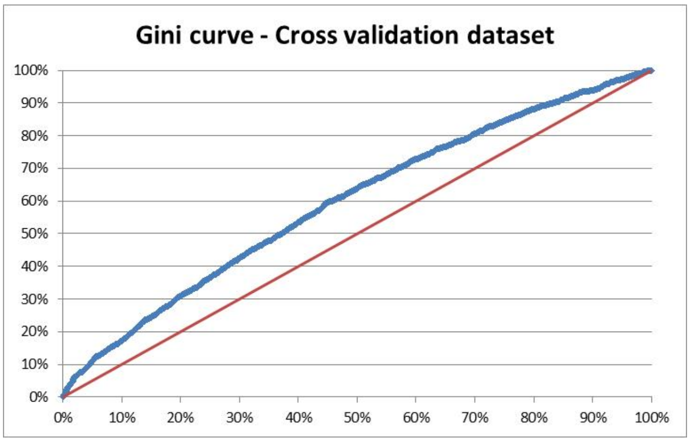

| | |
|---|---|
| Gini coeff | 19.82% |

b. Model including vehicle make



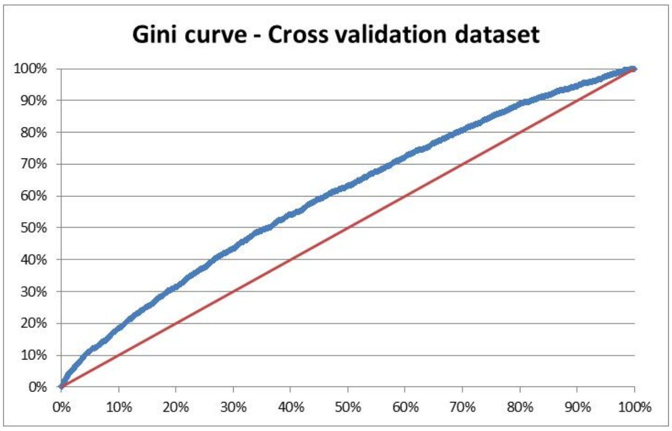

| | |
|---|---|
| Gini coeff | 17.71% |

c. Model including grouped vehicle make




| | |
|---|---|
| Gini coeff | 19.02% |

d. Model with credibility




| | |
|---|---|
| Gini coeff | 19.87% |

The higher Gini coefficient is also obtained by the model including credibility; which means it is more predictive than the other models. We notice that the model excluding the factor is closer to the result with credibility, which is in line with the low credibility factors which are weighted toward 1.

Overall the validation tests are satisfactory; hence we can see that introducing credibility for MLF increases the predictiveness of the frequency model. The improvement is more obvious for TPPD peril. For TPBI, due to the low estimated mean with regard to the variance within the portfolio, the credibility method marginally improves the model quality.

On the other side we can also highlight the limitation of the credibility approach. For some level of the MLF we may know the level of the expected relativity even with little exposure volume. In this case, the credibility method will pull the relativity toward the mean. For example, we may know that luxurious makes such as Porsche have a higher BI frequency.

When using credibility method the estimated $\hat{u}_k$ will be very close to 1.00 due to very low credibility for the level. For this reason it is important to review the final selections and potentially revise the estimated factors based on external market inputs, or combine it with expert judgement.

Another drawback of the method is that it is costly in term of implementation; hence the credibility modelling requires additional data manipulation and analysis. This has to be put in perspective with the level of modelling improvement we would obtain.

Nonetheless, the introduction of the credibility brought a solid answer to the issue of low exposure on MLF levels. Therefore we project for the next steps to investigate the introduction of additional factors to the model such as the Region or City; and develop the model to severity modelling from which we expect to get great benefit.

# Conclusion

Credibility theory has been developed early in the 20th century. It aimed to solve a fundamental question in P&C insurance: while insurance is based on mutualizing risks, in what extent are we able to split a portfolio in separate homogenous risk group in order to estimate more adequate premium for each group? The theory which is based on an analysis of the variance within the portfolio, achieved a breakthrough with Buhlmann model.

Later in the 20th century, the increase in segmentation in insurance pricing lead to the development of GLMs within the industry. Nowadays, GLMs are the most used technique to model insurance pure premiums, based on a multiplicative model. However they do present the drawback that they can't produce satisfactory results for multi-level factors, or at a cost of losing significant segmentation.

This is what we observed when performing a loss cost model on our TPL auto product. Prior to including credibility, we did not have enough exposure volume to get credible results for the vehicle make factor. Instead we have been able to split vehicle makes into two groups, such as 'higher risk' and 'lower risk'. However this simplification is a real limit in a competitive market such as auto insurance, given appropriate segmentation and risk selection is a key to the success.

For this reason, we decided to investigate how to incorporate credibility theory within our loss cost modelling. The paper of Ohlsson and Johansson enables to make the bridge between Buhlmann theory and the GLM modelling, making the use of credibility theory practical to our pricing analysis.

The results showed satisfactory results, especially for the TPPD peril modelling. In fact this model sophistication enabled us to give a different rate for each model. These results are also encouraging for us to develop the model further by including the region factor as an additional MLF, and apply it for severity modelling as well.

Finally, as mentioned by Ohlsson and Johansson, the study of credibility within GLMs could enable actuaries to better understand the estimators in pricing modelling, and eventually improve the communication between them and non-actuaries.

# Appendix I – Generalized Linear Models

## I.1 Estimation of the scale parameter

The estimation of the scale parameter is not necessary in order to solve the GLM parameter $\underline{\beta}$. Yet it is required to determine certain statistics. We can mention two approaches to estimate the scale parameter:

- $\phi$ can be treated as a parameter and estimated by maximum likelihood. The drawback is that it is not possible to derive an explicit formula for $\phi$ and the maximum likelihood process can take considerably long.

- Alternatively we can use an estimator of $\phi$ such as:

  - The moment estimator or Pearson's $\chi^2$ statistic:

$$\hat{\phi} = \frac{1}{n-p}\sum_{i=1}^{n}\frac{\omega_i(Y_i - \mu_i)^2}{V(\mu_i)}$$

  - The total deviance estimator:

$$\hat{\phi} = \frac{D}{n-p}$$

,with $D$ being the total deviance.

## I.2 Demonstration of the relation of the mean and variance for the exponential family of distribution

Let $Y$ be a variable from an exponential family of distribution. We will first define the moment-generating function of $Y$ :

$$M_Y(t) = E(e^{ty}) = \int_R e^{ty+\frac{y\theta-b(\theta)}{a(\phi)}-c(y,\theta)}dy$$

$$= e^{\frac{-b(\theta)}{a(\phi)}}\int_R e^{ty+\frac{y\theta}{a(\phi)}-c(y,\theta)}\,dy$$

$$= \frac{e^{\frac{-b(\theta)}{a(\phi)}}}{e^{\frac{-b(a(\phi)t+\theta)}{a(\phi)}}}\int_R e^{\frac{a(\phi)ty+\theta y}{a(\phi)}-c(y,\theta)-\frac{b(a(\phi)t+\theta)}{a(\phi)}}\,dy$$

$$= \frac{e^{\frac{-b(\theta)}{a(\phi)}}}{e^{\frac{-b(a(\phi)t+\theta)}{a(\phi)}}} \int_R e^{\frac{y(a(\phi)t+\theta)-b(a(\phi)t+\theta)}{a(\phi)} - c(y,\theta)} dy$$

$$= \frac{e^{\frac{-b(\theta)}{a(\phi)}}}{e^{\frac{-b(a(\phi)t+\theta)}{a(\phi)}}}$$

$$= e^{\frac{-b(\theta)+b(a(\phi)t+\theta}{a(\phi)}}$$

We can compute the mean of $Y$as follow:

$$E[Y] = M'_Y(t)|_{t=0}$$

,with $'$ corresponding to the derivation with regard to $t$.

$$= \left(e^{\frac{-b(\theta)+b(a(\phi)t+\theta)}{a(\phi)}}\right)'|_{t=0}$$

$$= e^{\frac{-b(\theta)+b(a(\phi)t+\theta)}{a(\phi)}} \left(\frac{b'(a(\phi)t+\theta).a(\phi)}{a(\phi)}\right)|_{t=0}$$

$$= e^{\frac{-b(\theta)+b(\theta)}{a(\phi)}} b'(\theta)$$

$$= b'(\theta).$$

We can compute the variance of $Y$as follow:

$$Var[Y] = (M''_Y(t) - M'_Y(t)^2)|_{t=0}$$

We've already calculated the $M'_Y(t)$ above; hence we will calculate:

$$M''_Y(t) = \left(e^{\frac{-b(\theta)+b(a(\phi)t+\theta)}{a(\phi)}}\right)''|_{t=0}$$

$$= e^{\frac{-b(\theta)+b(a(\phi)t+\theta)}{a(\phi)}} \{(b'(a(\phi)t+\theta))^2 + b''(a(\phi)t+\theta)a(\phi)\}|_{t=0}$$

$$= e^{\frac{-b(\theta)+b(\theta)}{a(\phi)}} \{b'(\theta)^2 + b''(\theta)a(\phi)\}$$

$$= b'(\theta)^2 + b''(\theta)a(\phi)$$

We can therefore calculate the Variance as follow:

$$Var[Y] = b'(\theta)^2 + b''(\theta)a(\phi) - b'(\theta)^2$$

$$Var[Y] = b''(\theta)a(\phi).$$

### I.3 Chi-squared test of equivalence

Chi-squared tests are statistical tests used to assess the goodness of fit of models. The chi-squared test of equivalence will test whether a model $M$ is statistically equivalent to a reference model $R$. The test's null assumption is as follow:

$H_0$: "The model $M$ is different from the model $R$".

The chi-squared consists in estimating the following test statistic:

$$\chi^2 = \sum_{i=1}^{n} \frac{(R_i - M_i)^2}{M_i}$$

with $R_i$ the estimated response variable of the reference model, $M_i$ the estimated response variable of the tested model $M$, and $n$ the number of observation.

The test is based on the statistic characteristic that it follows a Chi distribution of parameter $n - p - 1$.

We can therefore estimate the p-value of the model, i.e. lookup the probability that a variable following a $\chi^2_{n-p-1}$ distribution is higher than the statistic $\chi^2$. Depending on the level of acceptance we define (ex: 5%, 10%), we will accept $H_0$ (if p-value lower than the level) or reject it (if p-value higher than the level).

## I.4 Variable correlation analysis

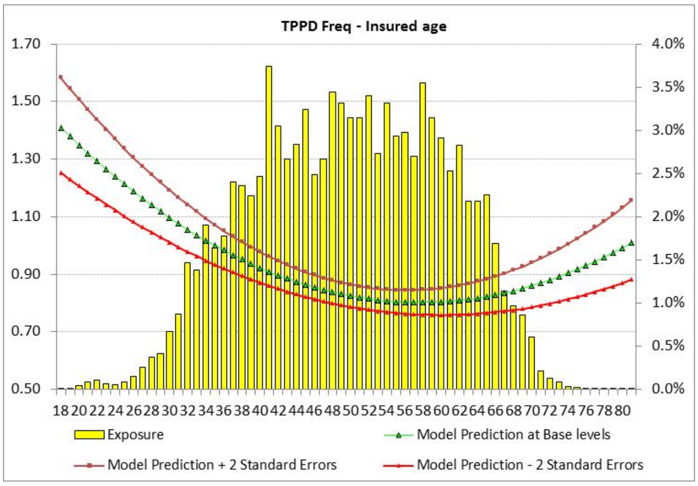
The following table shows Cramer V correlation matrix of the considered rating variables.

| Factor | Age of insured | Gender of insured | Insured marital status | Insured driver license age | Conviction | At-Fault claims | Region | City | Single payment | Vehicle model year | Vehicle value | Vehicle make | Vehicle make-model | Vehicle use | Total cars in household | Additional driver | Garage | Vehicle type | Vehicle automatic transmission | Vehicle imported / local | Engine power | Breaking assistance | Driving assistance | Vehicle no of doors | Vehicle length | Vehicle weight | Alarm |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Age of insured | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Gender of insured | 3% | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Insured marital status | 22% | 7% | | | | | | | | | | | | | | | | | | | | | | | | | |
| Insured driver license age | 10% | 4% | 10% | | | | | | | | | | | | | | | | | | | | | | | | |
| Conviction | 7% | 9% | 7% | 28% | | | | | | | | | | | | | | | | | | | | | | | |
| At-Fault claims | 4% | 3% | 9% | 29% | 28% | | | | | | | | | | | | | | | | | | | | | | |
| Region | 5% | 5% | 9% | 5% | 3% | 6% | | | | | | | | | | | | | | | | | | | | | |
| City | 4% | 5% | 5% | 3% | 5% | 8% | 99% | | | | | | | | | | | | | | | | | | | | |
| Single payment | 7% | 4% | 9% | 3% | 5% | 6% | 6% | 8% | | | | | | | | | | | | | | | | | | | |
| Vehicle model year | 4% | 6% | 5% | 5% | 5% | 7% | 7% | 4% | 6% | | | | | | | | | | | | | | | | | | |
| Vehicle value | 15% | 7% | 8% | 3% | 5% | 7% | 7% | 7% | 9% | 18% | | | | | | | | | | | | | | | | | |
| Vehicle make | 10% | 8% | 7% | 5% | 7% | 6% | 8% | 6% | 7% | 6% | 9% | | | | | | | | | | | | | | | | |
| Vehicle make-model | 7% | 3% | 8% | 7% | 8% | 9% | 8% | 3% | 7% | 3% | 9% | 70% | | | | | | | | | | | | | | | |
| Vehicle use | 9% | 6% | 8% | 6% | 7% | 4% | 7% | 8% | 9% | 5% | 9% | 10% | 17% | | | | | | | | | | | | | | |
| Total cars in household | 7% | 6% | 25% | 8% | 7% | 8% | 6% | 6% | 5% | 9% | 9% | 8% | 8% | 11% | | | | | | | | | | | | | |
| Additional driver | 7% | 8% | 31% | 4% | 8% | 4% | 4% | 7% | 8% | 3% | 8% | 12% | 18% | 10% | 10% | | | | | | | | | | | | |
| Garage | 9% | 9% | 9% | 6% | 6% | 5% | 12% | 11% | 7% | 4% | 12% | 22% | 14% | 16% | 7% | 9% | | | | | | | | | | | |
| Vehicle type | 7% | 6% | 8% | 5% | 8% | 8% | 5% | 4% | 7% | 9% | 20% | 14% | 14% | 6% | 12% | 7% | 14% | | | | | | | | | | |
| Vehicle automatic transmission | 6% | 7% | 9% | 5% | 5% | 4% | 4% | 5% | 7% | 4% | 17% | 15% | 11% | 11% | 8% | 13% | 7% | 20% | | | | | | | | | |
| Vehicle imported / local | 5% | 7% | 6% | 9% | 5% | 4% | 7% | 6% | 4% | 5% | 14% | 34% | 32% | 8% | 12% | 14% | 10% | 16% | 16% | | | | | | | | |
| Engine power | 9% | 8% | 6% | 7% | 5% | 9% | 6% | 8% | 5% | 11% | 13% | 19% | 17% | 14% | 5% | 9% | 13% | 16% | 38% | 42% | | | | | | | |
| Breaking assistance | 5% | 8% | 7% | 6% | 6% | 5% | 8% | 8% | 6% | 19% | 11% | 18% | 10% | 8% | 14% | 14% | 11% | 8% | 33% | 19% | 26% | | | | | | |
| Driving assistance | 7% | 8% | 6% | 7% | 5% | 7% | 4% | 6% | 9% | 18% | 14% | 21% | 17% | 11% | 10% | 13% | 11% | 16% | 35% | 22% | 16% | 18% | | | | | |
| Vehicle no of doors | 7% | 8% | 6% | 6% | 4% | 5% | 5% | 6% | 4% | 9% | 8% | 19% | 8% | 13% | 14% | 8% | 12% | 14% | 30% | 31% | 33% | 27% | 38% | | | | |
| Vehicle length | 8% | 6% | 5% | 7% | 6% | 8% | 5% | 5% | 8% | 9% | 9% | 18% | 12% | 8% | 12% | 14% | 12% | 16% | 28% | 39% | 33% | 27% | 18% | 18% | | | |
| Vehicle weight | 5% | 9% | 5% | 7% | 7% | 7% | 3% | 6% | 9% | 11% | 7% | 9% | 8% | 9% | 5% | 7% | 14% | 9% | 27% | 34% | 38% | 37% | 18% | 29% | 29% | | |
| Alarm | 7% | 8% | 6% | 3% | 3% | 8% | 7% | 6% | 8% | 22% | 15% | 8% | 26% | 13% | 6% | 11% | 12% | 12% | 15% | 22% | 27% | 18% | 44% | 35% | 41% | 40% | |
| Tracking device | 7% | 9% | 7% | 7% | 4% | 7% | 7% | 6% | 8% | 31% | 21% | 19% | 28% | 9% | 14% | 10% | 7% | 14% | 43% | 30% | 43% | 17% | 30% | 22% | 34% | 37% | 29% |

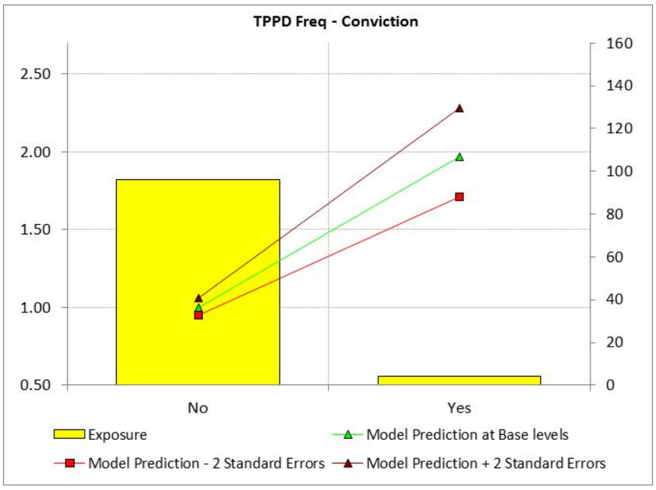
Note: Cramer's V is a coefficient of correlation between two variables. Buckets of correlated variables have been selected based on the criteria of a correlation factor higher or equal 30%.

## I.5 TPPD frequency model results

Insured Age

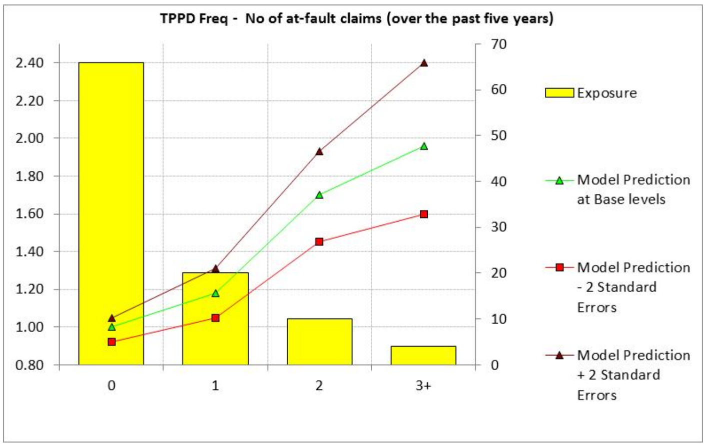



Conviction

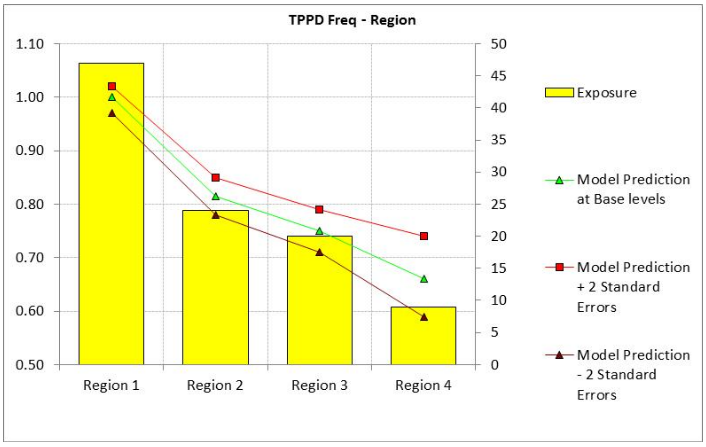



Driver age relativity has been fitted using a polynomial curve to smooth the model prediction.

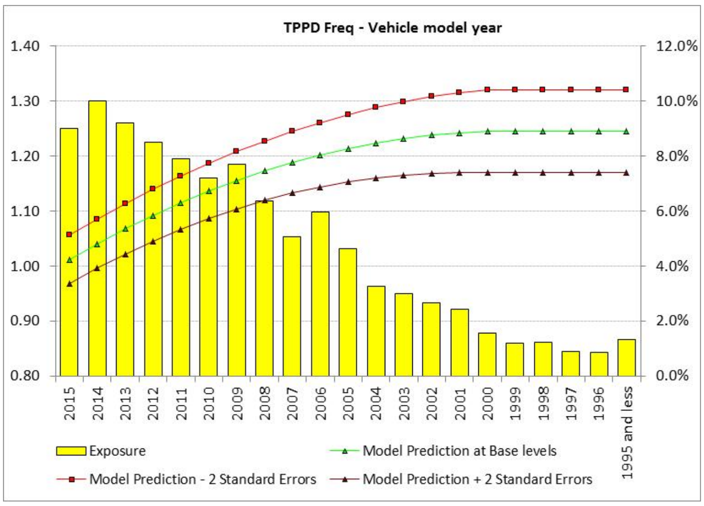
No of at-fault claims



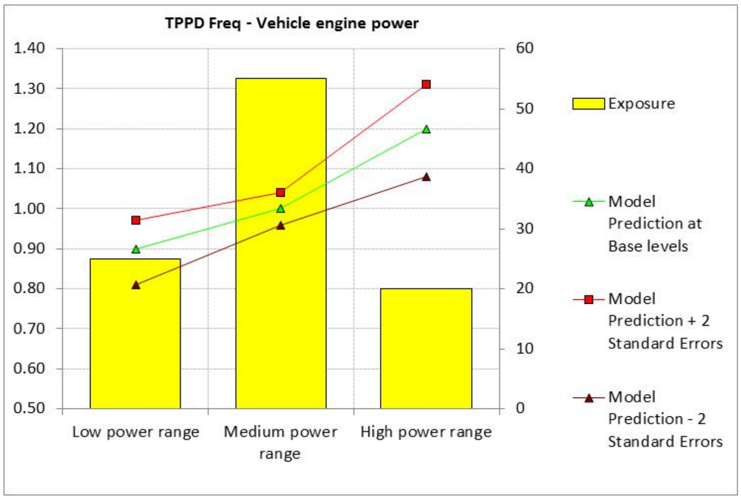

Region



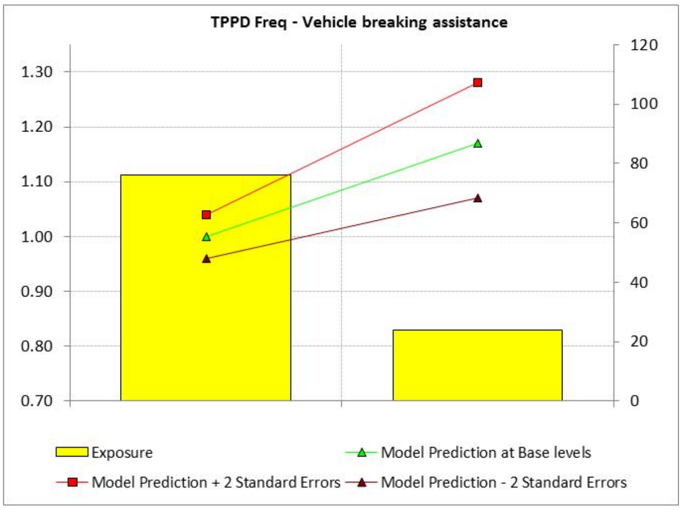

Vehicle model year




Vehicle engine power

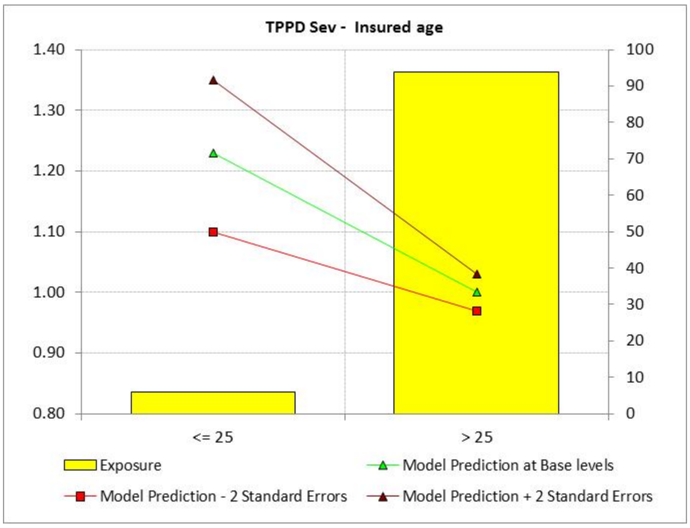


Vehicle breaking assistance

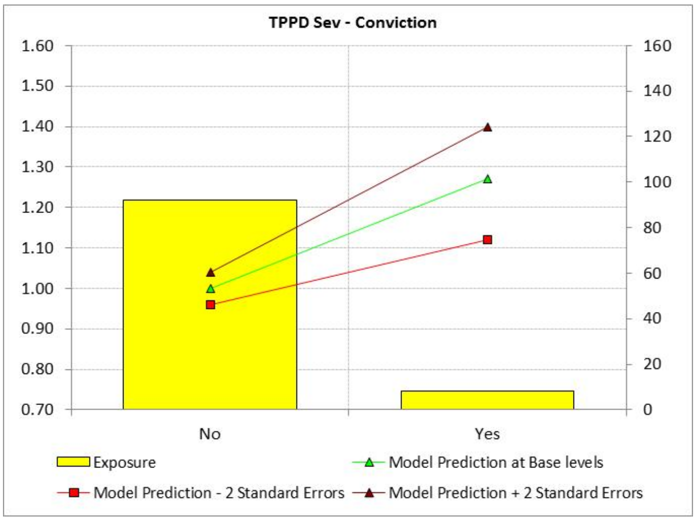



# I.6 TPPD severity model results

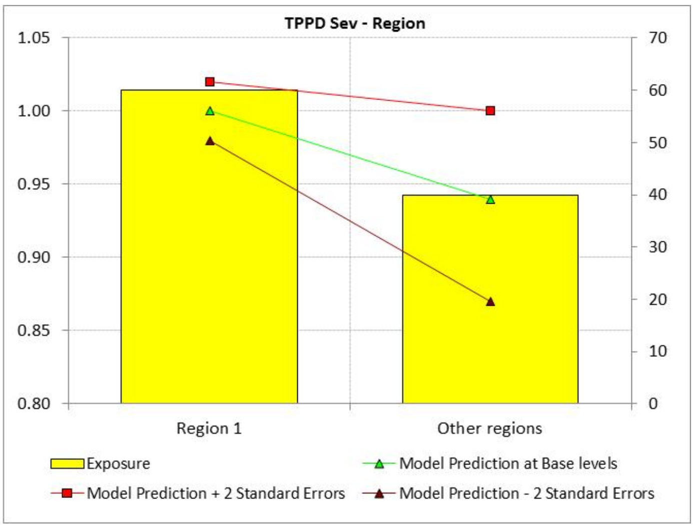
Insured Age



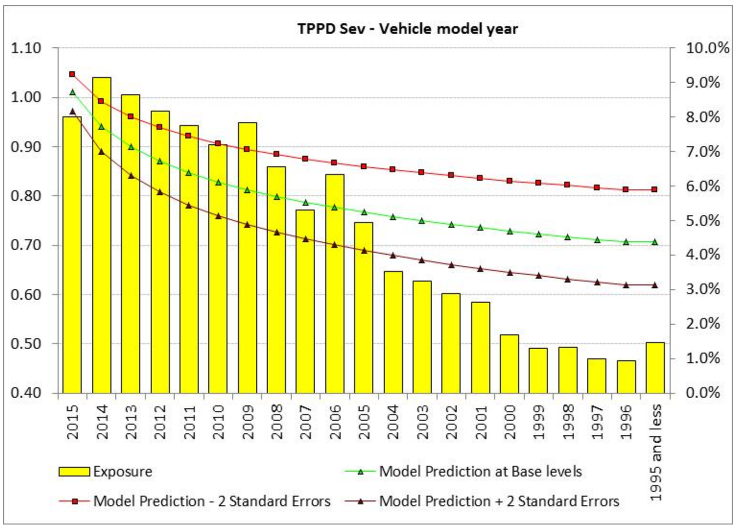

Conviction



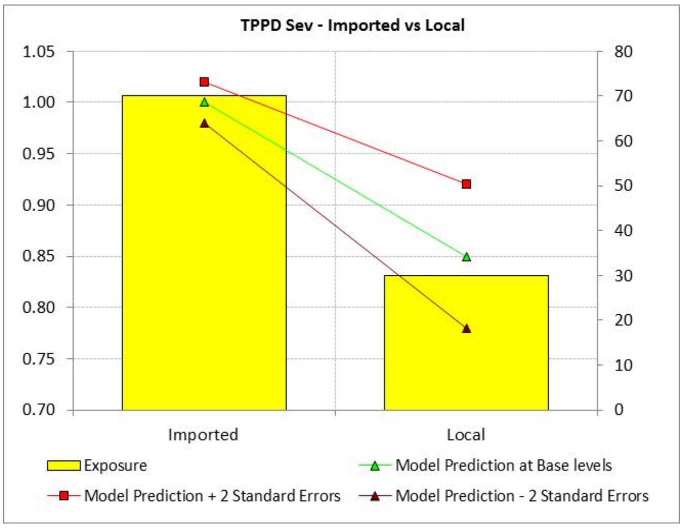

Region



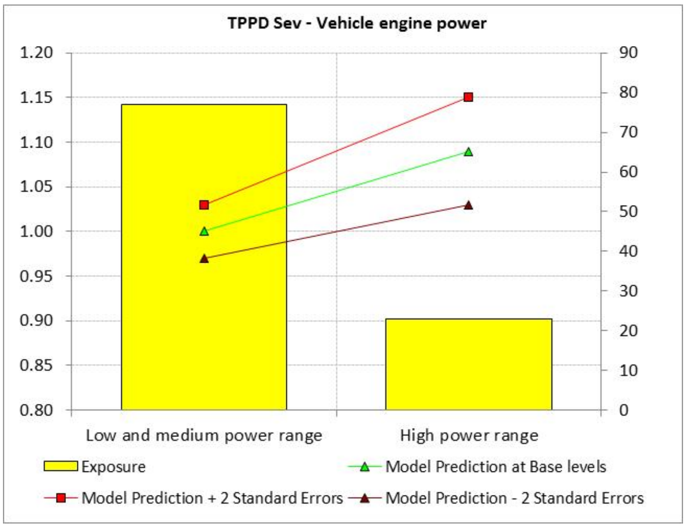

Vehicle model year



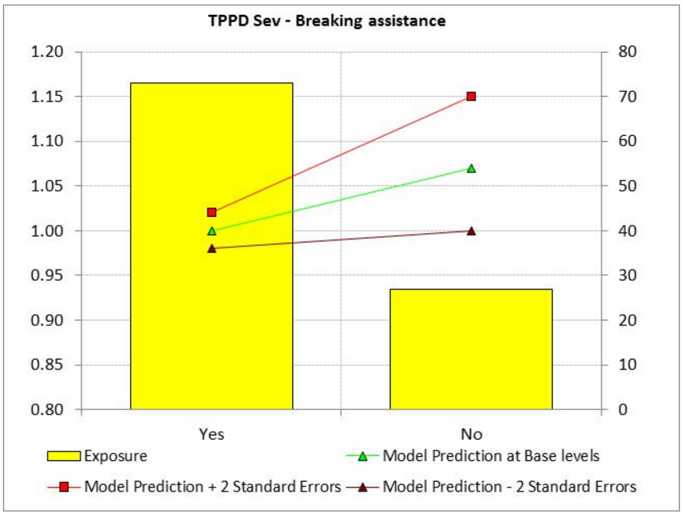
Imported / Local




Vehicle engine power

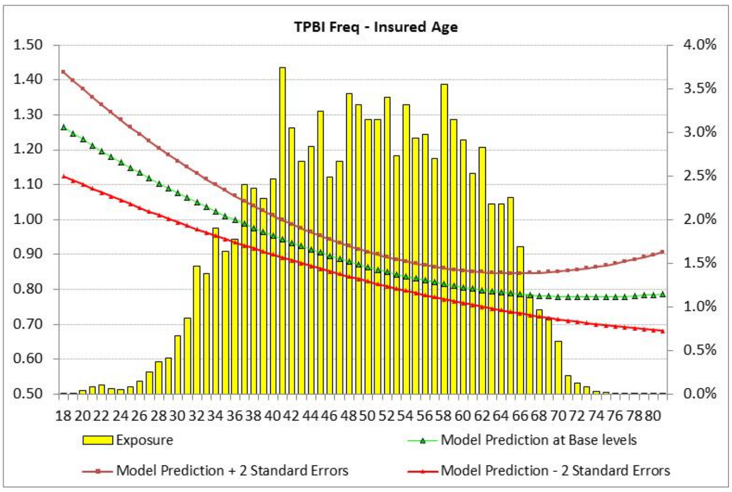



Breaking assistance

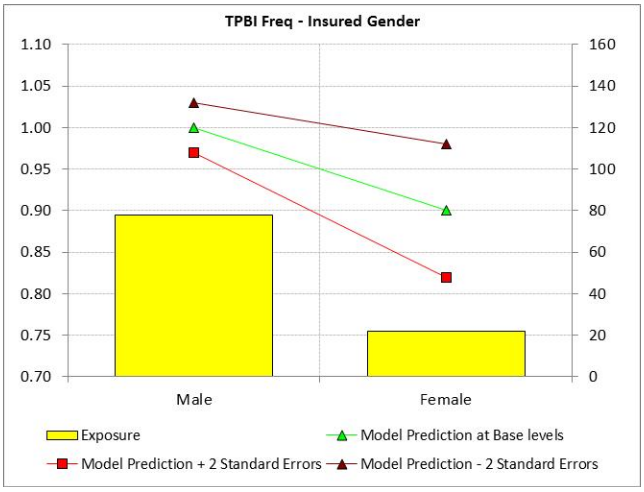



## I.7 TPBI frequency model results

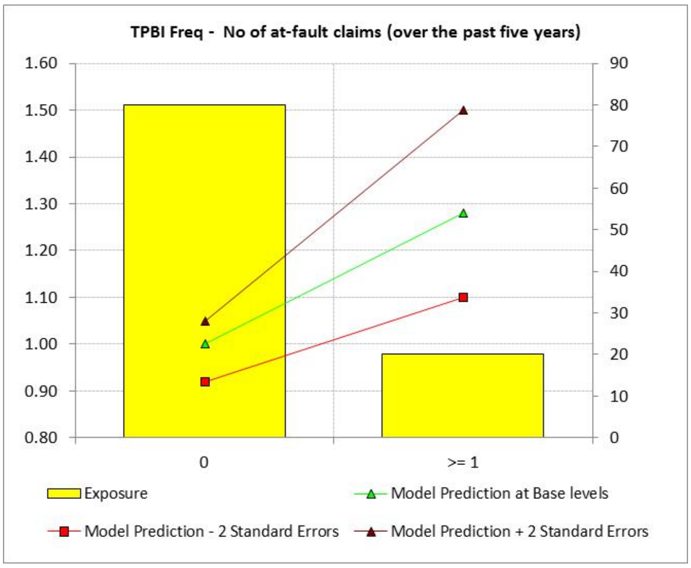
Insured age



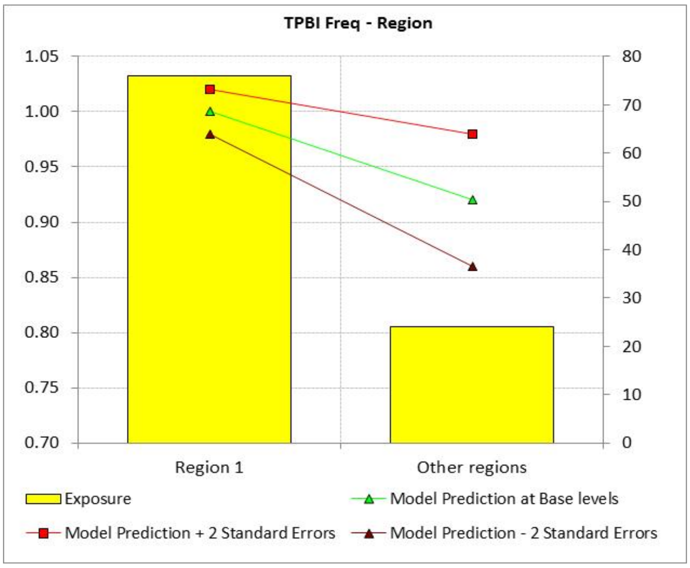

Insured gender



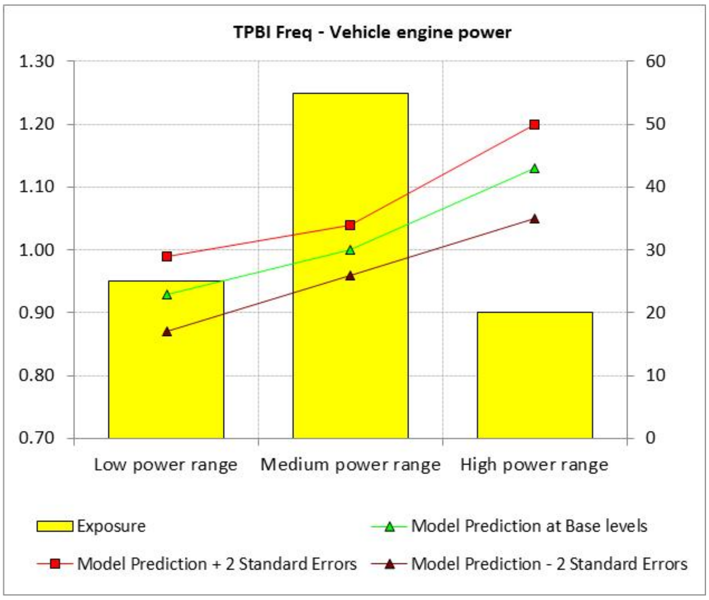
Not at-fault claims




Region

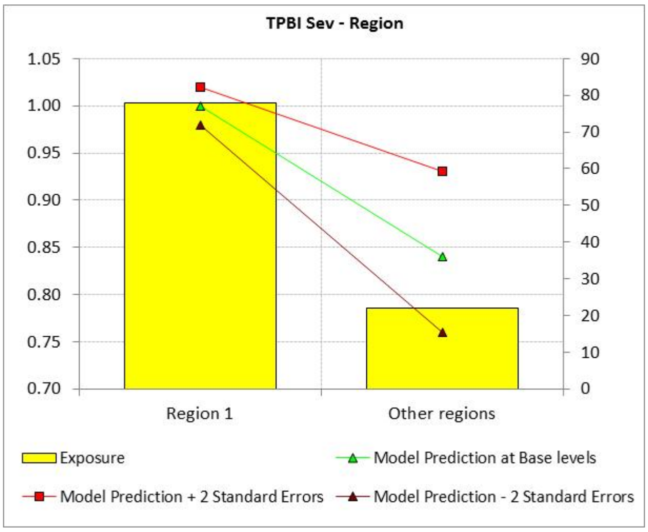



Vehicle engine power

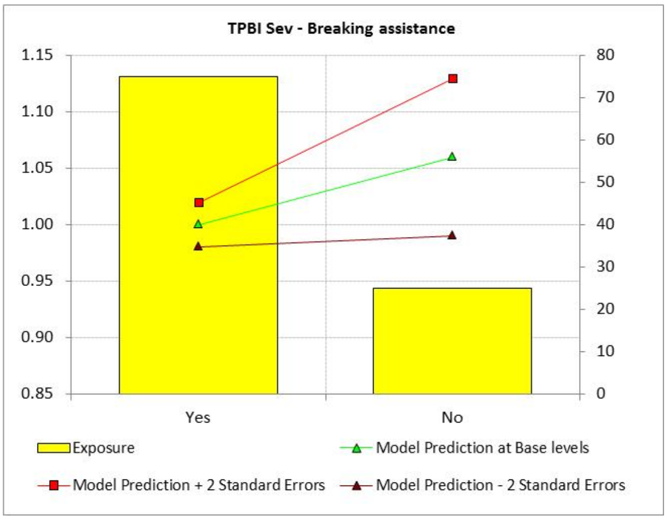


## I.8 TPBI severity model results

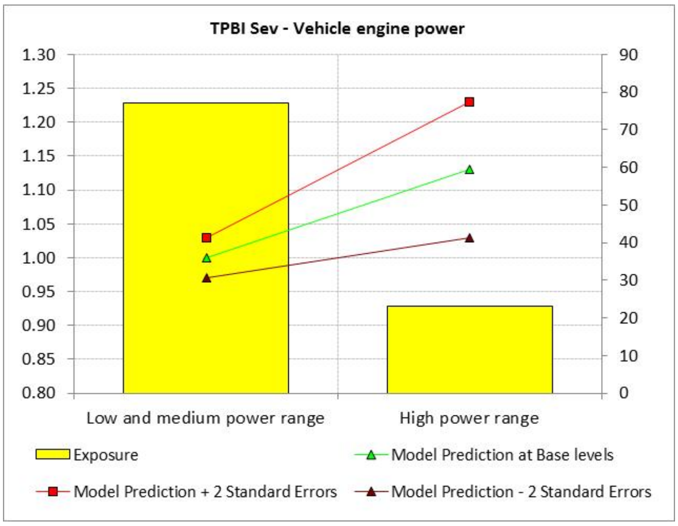
Region




Breaking assistance




Vehicle engine power



# Appendix II – Credibility methods

## II.1 Bayes Premium

Bayes premium $E[\mu(\theta_i)|X_i]$, also noted $\tilde{\mu}(\theta_i)$ , is the best estimator of $\mu(\theta_i)$ in the sense of the square error.

An estimator $\hat{\mu}(\theta_i)$ of $\mu(\theta_i)$ is at least as good as an estimator $\hat{\mu}^*(\theta_i)$ if:

$$E[(\hat{\mu}(\theta_i) - \mu(\theta_i))^2] \leq E[(\hat{\mu}^*(\theta_i) - \mu(\theta_i))^2]$$

, hence $E[(\hat{\mu}(\theta_i) - \mu(\theta_i))^2]$ is the mean square error of $\hat{\mu}(\theta_i)$.

*Demonstration*:

If we consider $\hat{\mu}(\theta_i)$ an estimator of $\mu(\theta_i)$ and $\tilde{\mu}(\theta_i) = E[\mu(\theta_i)|X_i]$ the a-posteriori mean of $\mu(\theta_i)$. We have:

$$E[(\hat{\mu}(\theta_i) - \mu(\theta_i))^2] = E\left[E[(\hat{\mu}(\theta_i) - \tilde{\mu}(\theta_i) + \tilde{\mu}(\theta_i) - \mu(\theta_i))^2|X_i]\right]$$

Moreover:

$$E[(\hat{\mu} - \tilde{\mu})(\tilde{\mu} - \mu)|X] = \tilde{\mu}E[\hat{\mu}|X] - E[\mu\hat{\mu}|X] - \tilde{\mu}^2 + \tilde{\mu}^2 = 0$$

Therefore:

$$E[(\hat{\mu}(\theta_i) - \mu(\theta_i))^2] = E[(\hat{\mu}(\theta_i) - \tilde{\mu}(\theta_i))^2] + E[(\tilde{\mu}(\theta_i) - \mu(\theta_i))^2]$$

$$\Rightarrow E[(\hat{\mu}(\theta_i) - \tilde{\mu}(\theta_i))^2] \leq E[(\hat{\mu}(\theta_i) - \mu(\theta_i))^2]$$

### II.2 Table of conjugate distributions

Let's consider the random variable $X_i$, with $i = 1, \dots, n$. If the likelihood function of $X_i$ belongs to the exponential family, then a conjugate prior exists.

The following table shows the natural conjugate prior for the most common likelihood functions.

| Likelihood function | Model parameters | Conjugate prior distribution | Prior hyperparameters | Posterior hyperparameters |
|---|---|---|---|---|
| $Binomial$ | $p$ | $Beta$ | $\alpha, \beta$ | $\alpha + \sum_{i=1}^{n} x_i \ , \beta + \sum_{i=1}^{n} N_i - \sum_{i=1}^{n} x_i$ |
| $Poisson$ | $\lambda$ | $Gamma$ | $\alpha, \beta$ | $\alpha + \sum_{i=1}^{n} x_i \ , \beta + n$ |
| $Normal$ with known variance $\sigma^2$ | μ | $Normal$ | $\mu_0, \sigma_0^2$ | $\frac{\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^{n} x_i}{\sigma^2}\right)}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)}, \frac{1}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)}$ |
| $Exponential$ | $\lambda$ | $Gamma$ | $\alpha, \beta$ | $\alpha + n \ , \beta + \sum_{i=1}^{n} x_i$ |

# Appendix III – Credibility theory and GLM revisited

### III.1 Demonstration of the conditional distribution of $Y_{ik}$

For the rest of the demonstration, it useful to explicit the following relations as presented by Ohlsson and Johansson. The function $b$:

$$b(\theta) = e^{\theta} \qquad \text{for } p = 1$$
$$= log(-\tfrac{1}{\theta}) \qquad \text{for } p = 2 \qquad \text{(III.1)}$$

These values are in line with Table 1.6, considering the relationship (3.2).

The canonical parameter space $M$ is defined as:

$$M = -\infty < \theta < +\infty \qquad \text{for } p = 1$$
$$= -\infty < \theta < 0 \qquad \text{for } p = 2 \qquad \text{(III.2)}$$

Moreover the derivative $b'$ is defined as follow:

$$b'(\theta) = e^{\theta} \qquad \text{for } p = 1$$
$$= -\tfrac{1}{\theta} \qquad \text{for } p = 2 \qquad \text{(III.3)}$$

And its inverse $h$, which corresponds to the canonical link function, as follow:

$$h(\mu) = log(\mu) \qquad \text{for } p = 1$$
$$= -\tfrac{1}{\mu} \qquad \text{for } p = 2 \qquad \text{(III.4)}$$

We can consider the random effect of the multi-level factor as a random variable $\Theta_k = h(U_k)$, which corresponds to the risk parameter we have in the credibility theory (see Chapter 2 – 2.1), with realization $\theta_k = h(u_k)$.

From (III.4) we have:

$$\theta'_{ik} = h(\mu_i u_k) = log(\mu_i u_k) = \log(\mu_i) + \log(u_k) = \log(\mu_i) + \theta_k \qquad \text{for } p = 1$$
$$= -\tfrac{1}{\mu_i u_k} = -\tfrac{1}{\mu_i} \times \tfrac{1}{u_k} = \tfrac{1}{\mu_i} \times \theta_k \qquad \text{for } p = 2 \qquad \text{(III.5)}$$

Moreover from (III.1) we have:

$$b(\theta'_{ik}) = b(h(\mu_i u_k))$$

$$= e^{\log(\mu_i)+h(u_k)} = \mu_i b\big(h(u_k)\big) = \mu_i b(\theta_k) \qquad \text{for } p = 1$$

$$= \log\left(-\frac{\mu_i}{h(u_k)}\right) = log(\mu_i) + \log\left(-\frac{1}{h(u_k)}\right) = log(\mu_i) + b(\theta_k) \qquad \text{for } p = 2 \qquad \text{(III.6)}$$

If we replace the values of $\theta'_{ik}$ and $b(\theta'_{ik})$ in (3.9) from (III.5) and (III.6), we can write:

$$f_{Y_{ik}|\Theta_k}(y_{ik}|\theta_k) = exp\left\{\frac{\omega_{ik}}{\phi}\left[\frac{y_{ik}}{\mu_i{}^{p-1}}\theta_k - \frac{1}{\mu_i^{p-2}}b(\theta_k)\right] + c_2\right\} \qquad \text{(III.7)}$$

Where $c_2$ is a constant that does not depend on $\theta_k$.

Demonstration:

For $p = 1$:

$$f_{Y_{ik}|\Theta_k}(y_{ik}|\theta_k) = exp\left\{\frac{\omega_{ik}}{\phi}[y_{ik}(\log(\mu_i) + \theta_k) - \mu_i b(\theta_k)] + c_1\right\}$$

$$= exp\left\{\frac{\omega_{ik}}{\phi}[y_{ik}\theta_k - \mu_i b(\theta_k)] + \underbrace{\frac{\omega_{ik}}{\phi}y_{ik}\log(\mu_i) + c_1}_{c_2}\right\}$$

For $p = 2$:

$$f_{Y_{ik}|\Theta_k}(y_{ik}|\theta_k) = exp\left\{\frac{\omega_{ik}}{\phi}\left[\frac{y_{ik}\theta_k}{\mu_i} - b(\theta_k)\right] - \frac{\omega_{ik}}{\phi}\log(\mu_i) + c_1\right\}$$

$$= exp\left\{\frac{\omega_{ik}}{\phi}[y_{ik}\theta_k - \mu_i b(\theta_k)] + \underbrace{\frac{\omega_{ik}}{\phi}y_{ik}\log(\mu_i) + c_1}_{c_2}\right\}$$

### III.2 Demonstration of Lemma 1

(a) (Ha): We suppose that $f_\Theta(\inf \mathrm{M}) = f_\Theta(\sup \mathrm{M}) = 0$, with M the interval of the canonical parameter.

From (3.11) we derive $f_\Theta$:

$$f'_\Theta(\theta) = \alpha(\delta - b'(\theta))f_\Theta(\theta)$$

If we integrate $f'_\Theta$ over M:

$$\int_M f'_\Theta(\theta)d\theta = f_\Theta(\sup \mathrm{M}) - f_\Theta(\inf \mathrm{M}) = 0$$

from (Ha).

$$\Leftrightarrow \alpha \int_M (\delta - b'(\theta))f_\Theta(\theta)d\theta = 0$$

$$\Leftrightarrow \alpha(\delta - E[b'(\Theta)]) = 0$$

$$\Leftrightarrow \delta = E[b'(\Theta)] = E[U]$$

given we have $\alpha \neq 0$.

(b) (Hb): We suppose that $f_\Theta{}'(\inf \mathrm{M}) = f_\Theta{}'(\sup \mathrm{M}) = 0$

From (3.11) we derive $f_\Theta$ twice:

$$f''_\Theta(\theta) = \alpha^2(\delta - b'(\theta))^2 f_\Theta(\theta) - \alpha b''(\theta)f_\Theta(\theta)$$

If we integrate $f'_\Theta$ over M:

$$\int_M f''_\Theta(\theta)d\theta = f_\Theta{}'(\sup \mathrm{M}) - f_\Theta{}'(\inf \mathrm{M}) = 0$$

from (Hb).

$$\Leftrightarrow \alpha^2 \int_M (b'(\theta) - \delta)^2 f_\Theta(\theta)d\theta - \alpha \int_M b''(\theta)f_\Theta(\theta)d\theta = 0$$

$$\Leftrightarrow \alpha^2 Var\big(b'(\Theta)\big) - \alpha E[b''(\Theta)] = 0$$

$$\Leftrightarrow \alpha = \frac{E[b''(\Theta)]}{Var\big(b'(\Theta)\big)} = \frac{E[U^p]}{Var(U)}$$

given we have $\alpha \neq 0$.

To demonstrate (Ha) and (Hb), Ohlsson and Johansson present a second Lemma that we will demonstrate as follow:

(Ha) is valid provided that $\alpha > 0$:

- $p = 1$:

$$f_\Theta(inf M) = f_\Theta(-\infty) = \frac{1}{c(\delta,\alpha)} exp\left\{\frac{\theta\delta - exp(\theta)}{1/\alpha}\right\} \xrightarrow[\theta \to -\infty]{} \frac{1}{c(\delta,\alpha)} \times 0 = 0 \text{ , if } \alpha > 0$$

$$f_\Theta(sup\ M) = f_\Theta(+\infty) = \frac{1}{c(\delta,\alpha)} exp\left\{\frac{\theta\delta - exp(\theta)}{1/\alpha}\right\} \xrightarrow[\theta \to +\infty]{} \frac{1}{c(\delta,\alpha)} \times 0 = 0 \text{ if } \alpha > 0$$

- $p = 2$:

$$f_\Theta(inf M) = f_\Theta(-\infty) = \frac{1}{c(\delta,\alpha)} exp\left\{\frac{\theta\delta - log(-\frac{1}{\theta})}{1/\alpha}\right\} \xrightarrow[\theta \to -\infty]{} \frac{1}{c(\delta,\alpha)} \times 0 = 0 \text{ , if } \alpha > 0$$

$$f_\Theta(sup\ M) = f_\Theta(0) = \frac{1}{c(\delta,\alpha)} exp\left\{\frac{\theta\delta - log(-\frac{1}{\theta})}{1/\alpha}\right\} \xrightarrow[\theta \to 0]{} \frac{1}{c(\delta,\alpha)} \times 0 = 0 \text{ if } \alpha > 0$$

(Hb):

- $p = 1$:

$$f_\Theta{}'(inf M) = f_\Theta{}'(-\infty) = \alpha(\delta - e^\theta) f_\Theta(\theta) \xrightarrow[\theta \to -\infty]{} 0$$

$$f_\Theta{}'(sup\ M) = f_\Theta{}'(+\infty) = \alpha(\delta - e^\theta) f_\Theta(\theta) \xrightarrow[\theta \to +\infty]{} 0$$

- $p = 2$:

$$f_\Theta{}'(inf M) = f_\Theta{}'(-\infty) = \alpha(\delta + \tfrac{1}{\theta}) f_\Theta(\theta) \xrightarrow[\theta \to -\infty]{} 0$$

$$f_\Theta{}'(sup\ M) = f_\Theta{}'(0) = \alpha(\delta + \tfrac{1}{\theta}) f_\Theta(\theta) \xrightarrow[\theta \to 0]{} 0$$

### III.3 Demonstration of unbiased estimators $\hat{\sigma}_k^2$ and $\hat{\sigma}_U^2$

- Estimation of $\hat{\sigma}_k^2$:

From (3.6) we can note that conditionally on $U_k$, the variables $X_{ik} = \frac{Y_{ik}}{\mu_{ik}}$ are independent with common expectation $U_k$. Moreover from (3.21) we have:

$$Var(Y_{ik}|\mathrm{U}_k = \mathrm{u}_k) = \phi \frac{\mu_i^p u_k^p}{\omega_{ik}}$$

$$\Leftrightarrow Var(X_{ik}|\mathrm{U}_k = \mathrm{u}_k) = \phi \frac{\mu_i^{p-2} u_k^p}{\omega_{ik}}$$

$$\Leftrightarrow Var(X_{ik}|\mathrm{U}_k = \mathrm{u}_k) = \phi \frac{u_k^p}{\widetilde{\omega}_{ik}}$$

(III.8)

with $\widetilde{\omega}_{ik} = \omega_{ik}\mu_i^{2-p}$. We highlight that $\bar{u}_k$ in (3.17) is a $\widetilde{\omega}_{ik}$-weighted average of the $X_{ik}$s.

Based on the Lemma III.3, we can estimate the following unbiased estimator of $\sigma_k^2 = \phi u_k^p$:

$$\hat{\sigma}_k^2 = \frac{1}{I-1}\sum_{i=1}^{I} \omega_{ik}\mu_i{}^{2-p} \left(\frac{Y_{ik}}{\mu_i} - \bar{u}_k\right)^2$$

For each level, we get a separate estimator and it is then natural to weigh them together with weights $I-1$ to the overall estimator:

$$\hat{\sigma}^2 = \frac{\sum_{k=1}^{K}(I-1)\hat{\sigma}_k^2}{\sum_{k=1}^{K}(I-1)}$$

- Estimation of $\hat{\sigma}_U^2$:

From (3.17) we have

$$\bar{u}_k = \frac{\sum_{i=1}^{I} \omega_{ik}\mu_i{}^{2-p} \frac{y_{ik}}{\mu_i}}{\sum_{i=1}^{I} \omega_{ik}\mu_i{}^{2-p}}$$

Therefore $\bar{u}_k$ is a linear function of the $y_{ik}$s, hence by (3.6) we have:

$$E[\bar{U}_k|U_k = u_k] = u_k$$

and

$$E[\overline{U}_k] = 1$$

We also have

$$E[(\overline{U}_k - 1)^2] = Var[\overline{U}_k] = Var[E(\overline{U}_k|\mathrm{U}_k)] + E[Var(\overline{U}_k|\mathrm{U}_k)]$$

$$= Var(\mathrm{U}_k) + E[Var(\overline{U}_k|\mathrm{U}_k)]$$

Moreover we know from the assumptions of the model that conditional on $\Theta_k$, i.e. on $\mathrm{U}_k$, the random variables $Y_{1k}, Y_{2k}, \dots, Y_{Ik}$ are independent. Moreover $\bar{u}_k$ is a $\widetilde{\omega}_{ik}$-weighted average of the $X_{ik}$s, with $\widetilde{\omega}_{ik} = \omega_{ik}\mu_i^{2-p}$. We can therefore write from (III.8):

$$Var(\widetilde{\omega}_{.k}\overline{\mathrm{U}}_k|\mathrm{U}_k) = \sum_{i=1}^{I} \widetilde{\omega}_{ik}{}^2 Var\left(\frac{Y_{ik}}{\mu_i}|\mathrm{U}_k\right) = \sum_{i=1}^{I} \widetilde{\omega}_{ik}\phi U_k^p = \widetilde{\omega}_{.k}\phi U_k^p$$

with $\widetilde{\omega}_{.k} = \sum_{i=1}^{I} \widetilde{\omega}_{ik}$

Hence we have

$$E(\widetilde{\omega}_{.k}(\overline{\mathrm{U}}_k - 1)^2) = \widetilde{\omega}_{.k}Var(\mathrm{U}_k) + \phi E[U_k^p]$$

Moreover we know from the assumption of the model that $\Theta_k$, i.e. $\mathrm{U}_k$; $k = 1,2,\dots,K$ are independent and identically distributed random variables. We can therefore drop the $k$ in $Var(U_k)$ and $E[U_k^p]$, summing over the index $k$ we get:

$$E\left[\sum_{k=1}^{K} \widetilde{\omega}_{.k}(\overline{\mathrm{U}}_k - 1)^2\right] = \widetilde{\omega}_{..}Var(U) + K\phi E[U^p]$$

with $\widetilde{\omega}_{..} = \sum_{k=1}^{K} \widetilde{\omega}_{.k}$, and $K$ the number of level of the MLF. We conclude that the following is an unbiased estimator of

$$\hat{\sigma}_U^2 = \frac{\sum_{k=1}^{K} \widetilde{\omega}_{.k}(\bar{u}_k - 1)^2 - K\hat{\sigma}^2}{\widetilde{\omega}_{..}} = \frac{\sum_{k=1}^{K}\sum_{i=1}^{I} \omega_{ik}\mu_i{}^{2-p}(\bar{u}_k - 1)^2 - K\hat{\sigma}^2}{\sum_{k=1}^{K}\sum_{i=1}^{I} \omega_{ik}\mu_i{}^{2-p}}$$

*Lemma III.3*: Let $X_1, X_2, \dots, X_n$ be a sequence of uncorrelated random variables with common mean μ and variance inversely proportional to weights $\omega_i$, i.e. $Var(X_i) = \frac{\sigma^2}{\omega_i}$; $i = 1,2,\dots,n$. With $\omega_. = \sum_{i=1}^{n} \omega_i$ we let

$\bar{X} = \frac{1}{\omega_.}\sum_{i=1}^{n} \omega_i X_i$ and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n} \omega_i (X_i - \bar{X})^2$ then $s^2$ is an unbiased estimator of $\sigma^2$.

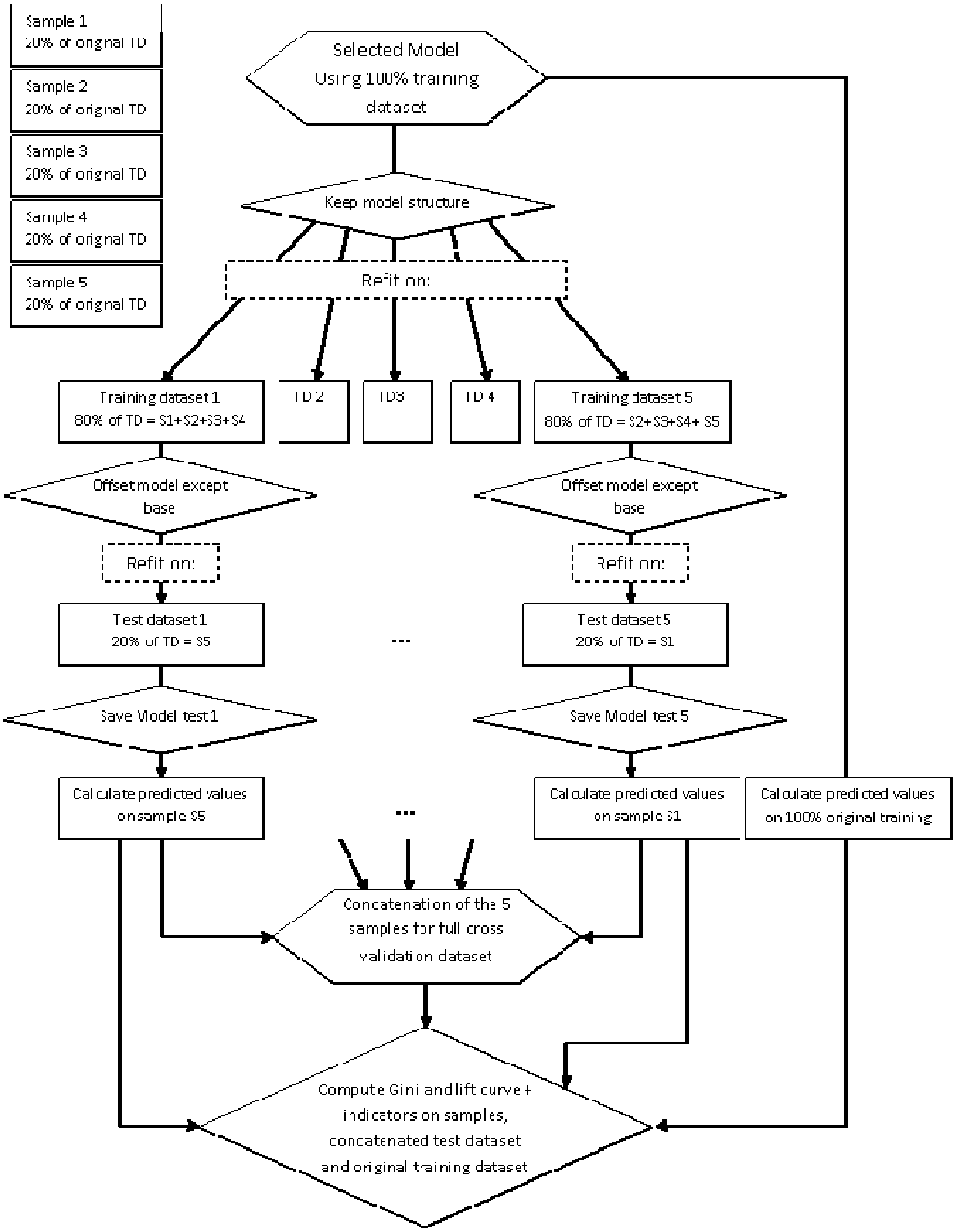
# Appendix IV – Application to our portfolio

### IV.1 Credibility program implemented in R

```
#Yik freq
Yik_freq <- Claim_count / weight
Yik_freq[is.na(Yik_freq)] <- 0

#Estimation of uk_bar
for (j in 1:38)
{  uk_bar[1,j]<-sum(Claim_count [,j])/sum(weight[,j]*mui_freq[]) }

I_k <- matrix(,1,38)
a_wgt <- matrix(,1,38)

for (j in 1:38)
{
  #Ik estimation (number of tariff cells where wik>0)
  Ik<-sum(weight[,j]>0)

  c<- Yik_freq[,j]/mui_freq[]

  sigmak <- (1/(Ik-1))*sum(weight[,j]*mui_freq[]*(c - uk_bar[,j])^2)
  a_wgt_k <- sum(weight[,j]*mui_freq[])

  sigma_k[1,j]<-sigmak
  a_wgt[1,j]<-a_wgt_k

  I_k[1,j]<-Ik
}

#Calculation of sigma and sigmauk
sigma <- sum((I_k-1)*sigma_k)/sum(I_k-1)
#Calculation of sigma_u
sigma_u <- (sum(a_wgt*((uk_bar-1)^2)) - 38*sigma)/sum(a_wgt)

#Calculation of u estimated and zk
for (j in 1:38)
{
  zk[1,j] <- a_wgt[1,j]/(a_wgt[1,j]+sigma/sigma_u)
  uk_est[1,j] <- zk[1,j]*uk_bar[1,j]+(1-zk[1,j])
}

#update uk_offset vector with new uk estimates
a <- t(uk_est)
b <- t(zk)
#First iteration to create first uk_est vector
uk_offset$u_offset <- a
```

### IV.2 Cross validation process

The following chart presents the process of the cross validation method.



# References

ANDERSON, FELDBLUM, MODLIN, SCHIRMACHER, THANDI (2007): *A practitioner's Guide to Generalized Linea Models*

WERNER, MODLIN (2010): *Basic Ratemaking*

DOBSON (2002): *An introduction to Generalized Linear Models – second edition*

THEROND (2013): *Théorie de la crédibilité –*
http://www.therond.fr/wp- content/uploads/cours/Credibilite.pdf

EL JERDY (2008): *Tarification des groupes en assurance santé* - Mémoire d'actuariat.

BUHLMANN, GISLER (2005): *A Course in Credibility Theory and Its Applications*

OHLSSON, JOHANSSON (2003): *Credibility Theory and GLM revisited*

SAS SUPPORT: *http://support.sas.com/index.html*