

A man with short, graying hair and glasses, wearing a dark pinstriped suit jacket, a light-colored striped shirt, and a dark blue tie with white polka dots. He is sitting in an office, looking slightly to the left of the camera. In the background, there is a glass partition, a framed picture on the wall, and a ceiling light fixture.

When GLMs go wrong – or how to tell when your models are not working

Generalised Linear Models are at the heart of successful personal lines pricing. Richard Bland considers the risk that these powerful tools might misbehave.

Generalised Linear Models (GLMs) often fail during a typical rating exercise and it is important that, as a modeller, you are able to identify when this happens. Failure to do so will result in inaccurate models, and hence wrong premiums being charged. Fortunately, many (but not all) rating software tools provide automatic messages to warn the user.

Some basic theory about GLMs

Before examining how GLMs fail, it is first necessary to run quickly through the theory of how GLMs work and the numerical techniques which are used to derive the solution.

The general concept of a linear model is that a statistic (the Y-variate) can be represented as a linear combination of some explanatory variables, where each variable is multiplied by a parameter and then added together. Some error term is then added to this formula to allow for variation in all the observed values of the Y-variate.

In the case of generalised linear models, the linear combination is transformed by the inverse of a link function (often log()) before the error term is added. This enables us to model a wider range of relationships, in particular the multiplicative combination often used in insurance models. A further generalisation allows us to use categorical variables in the formula – rather than using a

parameter multiplying the variable itself, we use multiple parameters, selecting the appropriate parameter for a data point based on the category into which the variable falls. This produces the GLM formula with which many will be familiar:

$$y_i = g^{-1}(\sum x_{ij} \cdot \beta_j) + \varepsilon_i$$

How are the appropriate parameters selected? The error term is deemed to come from a particular distribution, for example, Poisson or gamma, and so for a given set of parameters and a given set of data points, we can calculate a likelihood (more usually expressed as a log likelihood for scaling reasons) that the observed data would occur, given the parameters used. A maximum likelihood solution occurs when we alter the parameters to maximise the likelihood of the observed data, given the parameters.

A convenient way to maximise a differentiable function $f(y_i)$ is to use a Newton-Raphson iteration. If we take the probability density function, substitute in the formula above and differentiate with respect to β_j , we then get:

$$\beta_j' = \beta_j - \frac{f'(\beta_j)}{f''(\beta_j)}$$



and because the error functions have been defined, they can be analytically differentiated with respect to the parameters. The good news is that exponential error functions have well-behaved differentials which lead to a single maximum for the log-likelihood function; the bad news is that the β_j in the formula above is not a single variable, but a vector of parameters. So $f'(\beta_j)$ is a vector, and $f''(\beta_j)$ is a symmetrical square matrix, this means that we need to invert $f''(\beta_j)$ and then multiply $f'(\beta_j)$ by it to perform the iteration.



Figure 1 | Example showing correct aliasing (aliasing – groups 15–20 are aliased)

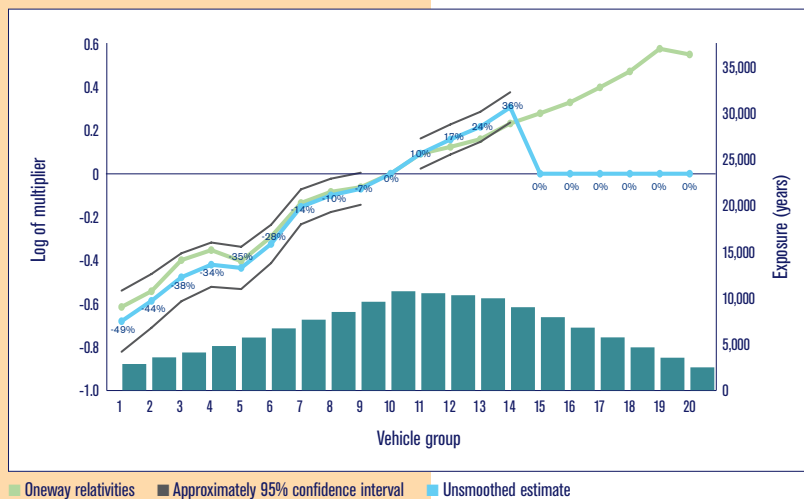
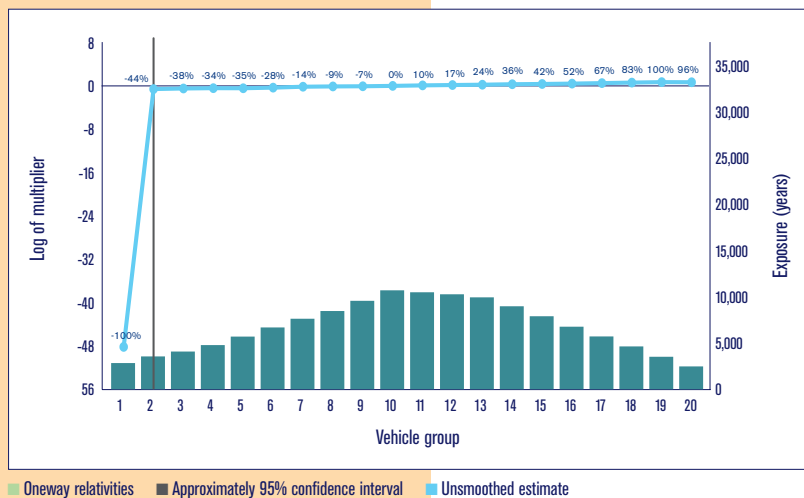


Figure 2 | Example showing an infinite parameter (non-convergent – group 1 has no claims)



Singularities in the Hessian

It is now clear what can go wrong with this method. Suppose we cannot invert the matrix of 2nd partial differentials (often known as the Hessian) because it is singular? (If a matrix is singular, attempting to invert it would involve dividing by zero.)

Aliasing

The most common cause of singularity is linear dependence between parameters – the value of parameter **a** always tells us the value of parameter **b**. This will immediately result in a singular Hessian and prevent us from iterating the Newton-Raphson formula. An alternative way of looking at this is that it would not be possible to find a unique solution for the parameters if two or more were completely correlated. Fortunately, there is an easy way of detecting linear dependence in the data. By multiplying the data matrix of the explanatory variables by its transpose, we create a symmetrical square matrix of the order of the parameters, and we can apply a decomposition to this which will identify any columns in the data which are linearly dependent. (A decomposition is a transform which breaks a matrix down into useful component parts – in this case, one of the parts will have zeros where columns in the original matrix are linearly dependent.) The parameters corresponding to these columns are known as ‘aliased’ and are removed from the model in order to ensure a unique solution (see Figure 1).



“ It is unlikely to be acceptable for you to charge near-zero premiums for some of your risks. ”

Most good GLM programs do this for you automatically and have a (sometimes hidden) option which allows the cut-off point at which the aliasing algorithm is triggered to be tuned – since the data may consist of many millions of records, the linear dependence does not have to be exact to cause a problem; the aliasing algorithm defines a threshold beyond which two variables are close enough to being linearly dependent to be aliased.

Estimating infinity

A more serious failure which cannot be handled automatically by the software is what happens when the maximum likelihood solution occurs at infinity for one or more parameters. This commonly happens for models with a log link solution when the correct estimated value for the Y-variate is zero, and it will occur, for example, in models of claim frequency if a categorical factor contains a category for which there are no claims. It may be that the GLM algorithm will fail immediately because of an inability to invert the Hessian, or it may be that it will be able to iterate – but it will never converge, because the convergence point is at infinity for one of the parameters (see Figure 2).

A similar failure can occur in a situation known as near-aliasing, where two or more variables are not quite linearly dependent. The variables are perfectly correlated for most data points, but there are a small number of points – typically data errors – where one of the variables has a different value. This is problematic if the data

points for which this occurs also form a category in which the Y-variate is zero (or infinite, after applying the link function). Again, if it is possible to iterate the model, it will never converge because the optimal solution requires the parameters to be at positive and negative infinity.

What happens if you fail to identify these failures?

Some different GLM systems react to this situation in different ways. If you define convergence as obtaining a set of unique, stable parameter estimates, then your model will never converge in these error situations and this will be reported as model failure by the software. If you define convergence as having obtained a stable value of the deviance (a definition of the difference between the data and the model fit), then you may accept this as a solution since the wayward parameters are unlikely to make much difference to (most of) the modelled values – but the resulting model does not have a uniquely defined set of parameter estimates and you might very well produce a different result with the same data if you were to vary the starting point of the iteration. There is also one very dangerous consequence in terms of insurance rating – some of the modelled values will be close to zero. It is unlikely to be acceptable for you to charge near-zero premiums for some of your risks.

Differences in the behaviour of GLM systems also arise because, although this article has described the classical mathematics of fitting GLMs, there are also some mathematical

shortcuts which can produce reasonably good approximate solutions for well-behaved GLMs using much less computational time, but since they do not involve constructing the matrices described above, the error detection methods described here are not applicable and there may not be any easy way of determining when a model has failed.

Some actuaries are happy to accept non-convergent models as solutions to their rating processes. One modeller was heard to comment that they were much happier with their new modelling software because models which previously did not converge were now being defined as convergent. However, they could equally well have commented that models which were faulty were now not being correctly identified, and which attitude is correct depends on the purpose for which the models are being used. If the purpose of the model is to produce modelled values for internal purposes which are then used as part of a further analysis, then it may not matter that the underlying parameter estimates are not a unique solution. However, if the parameters are then used as part of a rating formula which is engineered into a front-end rating system or published as premium tables, then it is vitally important that the model used produces stable parameters and does not produce outlying values when fitted to data.

For more information, contact:

Richard Bland

+44 (0) 1737 274541

richard.bland@watsonwyatt.com