

# INTRODUCING THE GENERALIZED LINEAR MODELS

Jong-Hwan Yoo

CEO and President of GLIS Consulting  
Owner of EMB Korea

**Abstract.** This paper deals with the Generalized Linear Models, an extension of the linear modeling process that allows models to be fit to data that follow probability distributions other than the Normal distribution, such as the Poisson, Binomial, Multinomial, and etc. Also, the Generalized Linear Models can be used for a non-linear distribution.

In the first chapter, we will begin by discussing the past problem of looking at the data using the simple linear regression models. We then introduce the Generalized Linear Models (GLM), and the characteristics of GLM to give a better understanding of the model. First, we discuss the components of GLM and the simple linear regression models. It is easier to compare both models when we understand the basic components of them. Then we discuss the Link Function, one of the major characteristics of GLM which allows the model to follow distributions other than Normal. Finally, we move to the Simpson's paradox and the Interaction test, yet another important aspect when we discuss GLM modeling. It helps us to understand the importance of GLM and multivariate tests when analyzing datasets with multiple parameters.

Chapter 3 discusses about The "Optimal" model. It is useful to understand what exactly the "Optimal" model is and what kind of procedures should be taken to achieve the "Best" model for GLM modeling process.

The auto insurance dataset is introduced for modeling example in chapter 4. GLM modeling software called EMBLEM was used for the modeling process. In EMBLEM, many useful applications including the statistics test for significance of each parameter, Interaction tests for multiple parameters, and many other functions are available for better and easier modeling. Since it is important to understand which application should be used for certain process, the process for GLM modeling will also be discussed in this chapter.

## 1. Introduction

### a. The data

Statistical models contain both systematic effect and random effect. The value of a model is that often it suggests a simple summary of the data in terms of the major systematic effects together with a summary of the nature and magnitude of the random variation. Regression analysis plays a central role in statistics being one of its most powerful and commonly used techniques for analyzing models and predicting the future trend. However, the simple linear regression is not always the best choice due to following issues. First of all, the dependent variable of interest may have a non-continuous distribution. Therefore, the predicted values should also follow the respective distribution. However, since the simple linear regression

is based on the assumption that the response variable follows the normal distribution only, it may not be the best method to analyze the model which follows the non-normal distribution. Another reason why the simple regression model might be inadequate to describe a particular relationship is that the effect of the predictors on the dependent variable may not be linear in nature. For example, the relationship between a person's age and various indicators of health is most likely not linear in nature. Therefore, the simple linear regression model may not be suitable for data which carries these characteristics.

## b. Introducing the Generalized Linear Models

The Generalized Linear Models is an extension of the linear modeling process. In other words, they extend the ideas of regression analysis to a wider class of problems involving the relationship between a response and one or more explanatory variables. For example, in a situation in which a patient takes a standard treatment we might consider how the probability of success depends on the patient's age, weight, blood pressure and so on. Also, they can be used for models which do not follow the normal distribution like Poisson, binomial, gamma and others. The *link function* is used to model responses when a dependent variable is assumed to be nonlinearly related to the predictors. In conclusion, the Generalized Linear Models can be used to predict responses both for dependent variable with non-continuous distribution and for dependent variables which are non-linearly related to the predictors.

## 2. Characteristics of GLM Model

### a. The difference between GLM and Simple Regression Models

It is important to understand the difference between simple linear regression models and generalized linear models. Even though the basic components for both models are similar, there are a few characteristics of GLM which make a big difference.

Let  $\mathbf{y}$  be a vector of observations, and let  $X$  be a matrix of covariates. The usual simple linear regression models (usually multiple regression models) take the form

$$E(Y) = \mu = \eta = \beta_0 + \sum_{j=1}^p \beta_j x_j \quad (1)$$

where  $\beta_j$  are vectors of regression coefficients. Typically we assume that the  $Y_i$  are normal and independent with standard deviation  $\sigma$ , so that we estimate  $\beta_j$  by minimizing the sum of squares which can be stated as  $(\mathbf{y} - \boldsymbol{\mu})^T(\mathbf{y} - \boldsymbol{\mu})$ .

In this case, the most important and common case is that in which the  $y_i$  and  $\mu_i$  are bounded. However, if we assume that  $y$  is binary,  $y = 1$  if an animal survives and  $y = 0$  if it does not say, then  $0 < \mu < 1$ . Then the simple linear model is inadequate in

these cases because complicated and unnatural constraints on  $\beta$  would be required to make sure that  $\mu$  stays in the possible range. Therefore, the simple linear regression model cannot be a good solution for those datasets. Generalized linear models instead assume a *link linear* relationship.

$$E(Y) = g(\mu) = \beta_0 + \beta_1 + \Lambda + \beta_j x_j \quad (2)$$

The link function  $g(\mu)$  serves to link the random or stochastic component of the model, the probability distribution of the response variable, to the systematic component of the model (the linear predictor). Typically  $g(\cdot)$  is used to transform the  $\mu_i$  to a scale on which they are unconstrained. As we can see, the Link Function plays an important role in GLM. We will not discuss the Link Function used in the generalized linear models.

### b. The Link Function

In simple linear models, the mean and the linear predictor are identical and the identity link is plausible in that both the linear predictor and the expected value can take any value on the real line. However, it does not fit well with the GLM because The link function in GLM specifies a nonlinear transformation of the predicted values so that the distribution of predicted values is one of several members of well-known distributions like Gamma, Poisson, binomial and so on. Therefore, it is used to model responses when a dependent variable is assumed to be nonlinearly related to the predictors.

Various link functions are commonly used, depending on the assumed distribution of the dependent variable ( $y$ ) values. Log link function is commonly used for Poisson, Gamma, and Normal and Inversed Normal distributions. However, Identity and Power link functions are available for them also.

Identity link:  $f(z) = z$

Log link:  $f(z) = \log(z)$

Power link:  $f(z) = z^a$ , for a given  $\mathbf{a}$

For Power link, each a value refers the following distributions:

$a=0$ , Normal error structure

$a=1$ , Poisson error structure

$a=2$ , Gamma error structure

For Binomial and other Multinomial distributions, Logit link function is preferred since the link should satisfy the condition that the expected value maps the

interval (0, 1) on to the whole real line. Also, other link functions are available along with Logit link.

Logit link:  $f(z)=\log(z/(1-z))$

Probit link:  $f(z)=\text{invnorm}(z)$  where *invnorm* is the inverse of the standard normal cumulative distribution function.

Complementary log-log link:  $f(z)=\log(-\log(1-z))$

Loglog link:  $f(z)=-\log(-\log(z))$

Usually, Poisson is used for the Frequency data since the Frequency states how many accidents are happening during the given period. For the Severity data, Gamma is commonly used. In the example which we will do the next section, we will use the Severity data to do the GLM modeling.

### c. One-way analysis

It is convenient to analyze One-way tables since they give an immediate view of where exposures are and may show trends in claims experience. However, by using one-way tables only, correlations between explanatory variables would not be taken into account. When there are multiple rating parameters in the model, many correlations can exist between those parameters. Therefore, one-way analysis may not be the best solution for large models.

### d. Simpson's Paradox

Another problem we face when analyzing the data with multiple parameters is Simpson's Paradox. It happens when we consider two parameters separately, and then consider them altogether. Suppose there are two hospitals A and B. Each hospital treats same kind of disease. Suppose there are two types of patients, Severe and Mild. If we look at the table down below, it is not hard to find out that the mortality rate of Hospital A and B for each type of patients are in fact same. However, when we look at the overall mortality rate, the mortality rate for each

Hospital is now very different. If we think the other way around, the data shows one hypothesis, but that hypothesis gets reserved when we add another parameter to it. So it is important to check the multivariate table and check if the interaction effect exists.

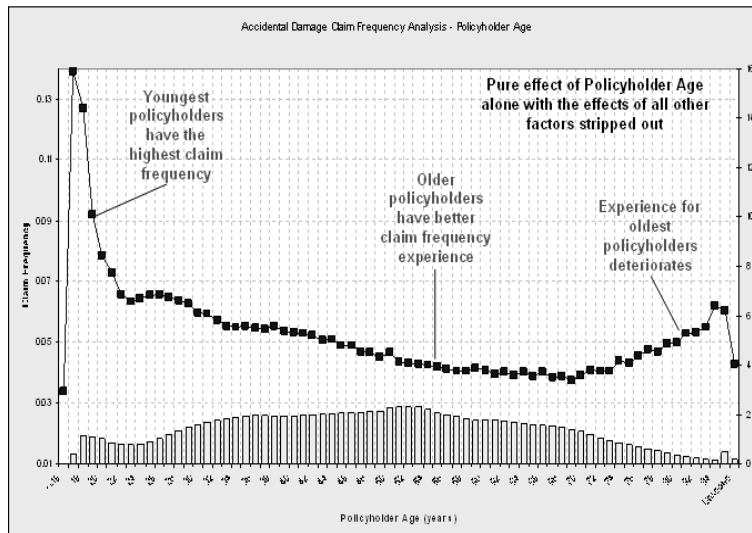
	<b>Severe</b>	<b>Mild</b>	<b>Total</b>
Hospital A	140/70	30/300	170/1,000
Hospital B	60/300	70/700	130/1,000
Total	200/1,000	100/1,000	300/2,000

### e. Example Trends - Risk Premium Modeling

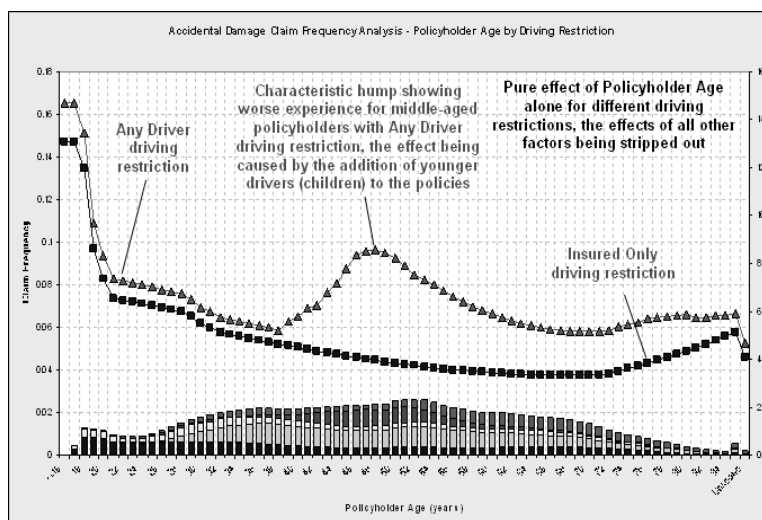
As mentioned in previous sections, it is important to identify not only the pure effect of one parameter but also the effect of multiple parameters altogether. One-

## INTRODUCING THE GENERALIZED LINEAR MODELS

way analysis is still important to identify the pure effect of one parameter. However, we also need to analyze the effect of multiple parameters simultaneously because correlations may exist between any of them.

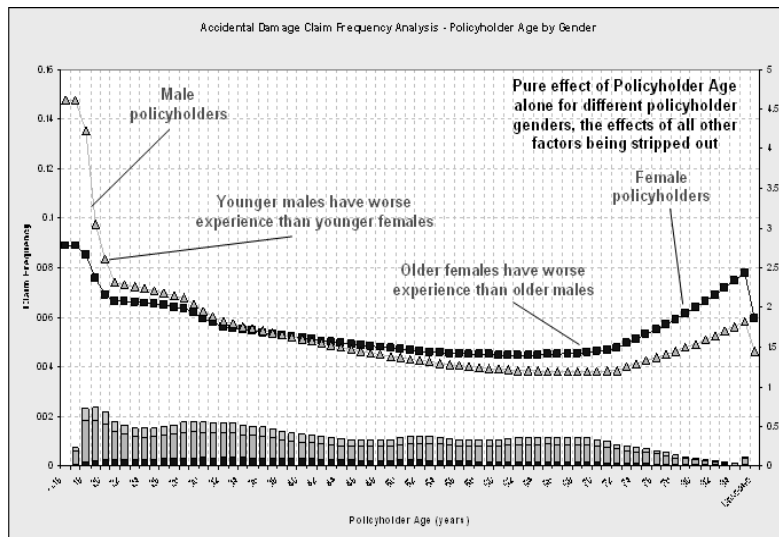


The table shown above is the frequency model of policyholder's age from auto insurance data. It shows a pure effect of the policyholder age parameter without considering any other parameter which means only the effect of the policyholder's age is considered for the result. As observed, young people tend to have higher claim frequency in this model. We can conclude that the young -aged group has higher claim frequency because young people have less driving experience and are also less careful. On the other hand, old-aged group (especially those who reached the retirement age) has higher claim frequency compared to the middle-aged group, because as people get older their ability to keep control decreases.



This result, however, gets reversed if we add the driving restriction parameter to the previous analysis. Young-aged group still have highest claim frequency. But there is a strange trend showing at the middle aged group. Why has this happened? By adding the driving restriction parameter, we now allow policyholders that their kids can be included in the insurance policy. Since they are young people themselves, they would have high claim frequency. It is important to understand that the model considering one parameter and the model comparing two parameters produces significant different results because the second parameter affects the first one.

Another correlation can be identified when looking at the multivariate table between policyholder age and gender parameters. Overall, men have lower claim frequency than women. However, if we only consider a young group, it is clear that the result is not different. We can conclude that young women have advantage against young men in claim frequency and should have lower risk because of it.



As we can see, more significant signals were revealed when we added different parameters to the original model with the pure effect of policyholder's age parameter. Therefore, more sophisticated techniques are required since we need to consider all factors simultaneously.

**f. Main Effect Model vs. The Interaction Model**

Sometimes, it is necessary to add interactions when building a model. Main effect model can be useful to find the major trends for each parameter, but may not be so when considering correlations between multiple parameters. For example, consider the Policyholder's age and Policyholder's sex parameters from the previous section. It is clear that there is a correlation between two parameters since the frequency between two sexes gets reversed as their age gets older. Therefore, in this case, adding interactions in the model may be necessary.

INTRODUCING THE GENERALIZED LINEAR MODELS

Age/Sex	Male	Female
10-19	$\beta_0$	$\beta_0 + \alpha_2$
20-29	$\beta_0 + \beta_2$	$\beta_0 + \alpha_2 + \beta_2$
30-39	$\beta_0 + \beta_3$	$\beta_0 + \alpha_2 + \beta_3$
40-	$\beta_0 + \beta_4$	$\beta_0 + \alpha_2 + \beta_4$

$$g(p_{ij}) = \beta_0 + \alpha_i + \beta_j$$

$$\alpha_1 = 0 \quad \beta_1 = 0$$

Age/Sex	Male	Female
10-19	$\beta_0$	$\beta_0 + \alpha_2$
20-29	$\beta_0 + \beta_2$	$\beta_0 + \alpha_2 + \beta_2 + (\alpha\beta)_{22}$
30-39	$\beta_0 + \beta_3$	$\beta_0 + \alpha_2 + \beta_3 + (\alpha\beta)_{23}$
40-	$\beta_0 + \beta_4$	$\beta_0 + \alpha_2 + \beta_4 + (\alpha\beta)_{24}$

$$g(p_{ij}) = \beta_0 + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

$$\alpha_1 = 0 \quad \beta_1 = 0$$

$$\alpha\beta_{11} = 0 \quad \alpha\beta_{12} = 0 \quad \alpha\beta_{13} = 0$$

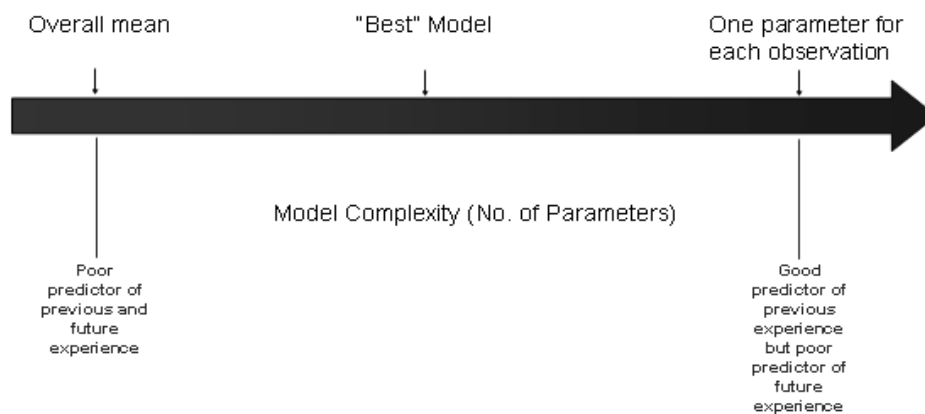
$$\alpha\beta_{21} = 0 \quad \alpha\beta_{31} = 0$$

**Main Effect Model (Above) vs. the Interaction Model (Below)**

We only need 5 variables to analyze the main effect model. However, we need 3 additional variables to analyze the interaction. Therefore, by adding interactions the model gets more complicated. It is important to keep a good balance between accuracy of the model and understandability. By adding more variables, the accuracy of the predictions can be improved but maybe spurious accuracy.

### 3. GLM Modeling

**a. What is a 'Good' Model?**



Our purpose of statistical modeling is to replace our data with a set of fitted values derived from a model to minimize the random noise and to be able to predict the future trend. When we use too few parameters when building a model, it will cause a poor predictor of previous and future experience. If that is true, one can simply assume that the 'Good' model is a model which fits the observed data as close as possible. However, if we use too many parameters to make a 'Perfect' fit, it can also be a bad choice. Yes, the model can show a good predictor of previous experience, but it can also show poor predictor of future experience. In conclusion, a 'Good' model is a model which do not only fit the observed data well, but also be able to predict the future trend. Not only that, the model should have some simple theoretical pattern for the raged data. Therefore, in order to create a good model, we need to exclude any parameter which is not significant to the data. Down below is graph of two models, one with too many parameters and one with too few parameters. Model with too many parameters cause an "Over-fitting" problem. The overall mean is small, but the variance is large. On the other hand, the model with too few parameters causes an "Under-fitting" problem. Which means the overall mean is large, but the variance is small. It is wise to avoid either problem. We can do this by using multiple methods.

#### **b. Process of finding the Best Model**

The main process of finding the Best Model can be defined by two ways. One is to simplify the model, and another is to complicate it. By using both methods, we can fit the model to the observed data, and also see for the future trend. There are two methods to check the significance for each parameter to simply the model. One is to check for a standard error. And another is to use a Chi-Square test or F-test statistics. Suppose there is one parameter we want to check for significance. First, we can run the test and find out the standard error. If the standard error is high, then we can say that the parameter is not significant for the model. Next, we can run either Chi-Square tests or F-tests. If the statistic is above the acceptable standard, (usually 5%) then we can conclude that the parameter is not significant enough to be used in the model. Of course, if the result is not clear, then we have to use our own judgment to make a decision whether the parameter should be included in the model or not. By test each parameters for significance, we can simplify the model by reducing the parameters used in the model. However, we also need to complicate the model so we can check the systematic effect better. Usually we create interactions so we can avoid the Simpson's paradox when analyzing multiple parameters altogether.

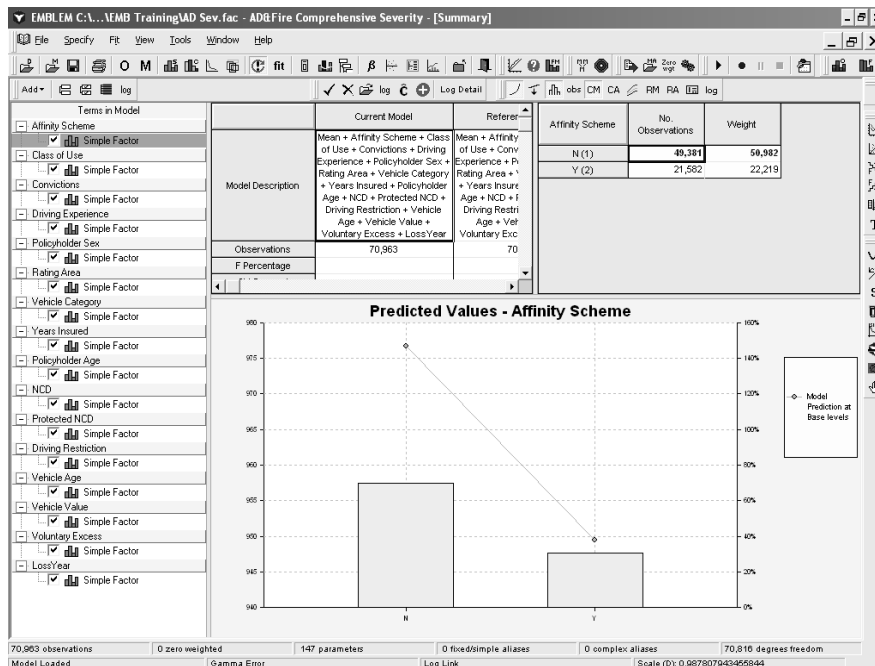
## **4. Practical Example: Auto Insurance Data Modeling**

### **a. Getting Started**

In this chapter, we used the auto insurance data as an example of GLM modeling. EMBLEM, a GLM statistical modeling program was used for analyzing the data. Both the Frequency and the Severity data were introduced and used to compare results between them. Poisson log link function was used for the Frequency data, and Gamma log link function was used for the Severity data.



## INTRODUCING THE GENERALIZED LINEAR MODELS



As shown above, all the rating factors were selected and used for fitting the model. In our next section, we will discuss how to simplify the model by eliminating parameters which do not play significant roles in the model.

### b. Simplifying the Model: Eliminate Parameters

The idea behind any statistical model is to fit a parsimonious model to the data, i.e. the model with fewest parameters that adequately describes the data. A model can be simplified in multiple ways. In this section, we will discuss about elimination of certain rating factors, the most straightforward model simplification method. Factors that are candidates for removal can be identified as follows:

- Factors where all the levels have small parameter estimates.
- Factors where the parameter estimates are all within two standard errors of zero, i.e. the standard error percentages are all greater than 50%.
- Factors which do not exhibit consistent trends for each time period.
- Factors which, when excluded from the model, result in models that are not significantly different from the original on a chi-squared basis.
- Factors which do not exhibit sensible trends (although judgment is required to determine what a sensible trend is).

Let's take a look at the factor "Class of Use" in our dataset.

JH Yoo

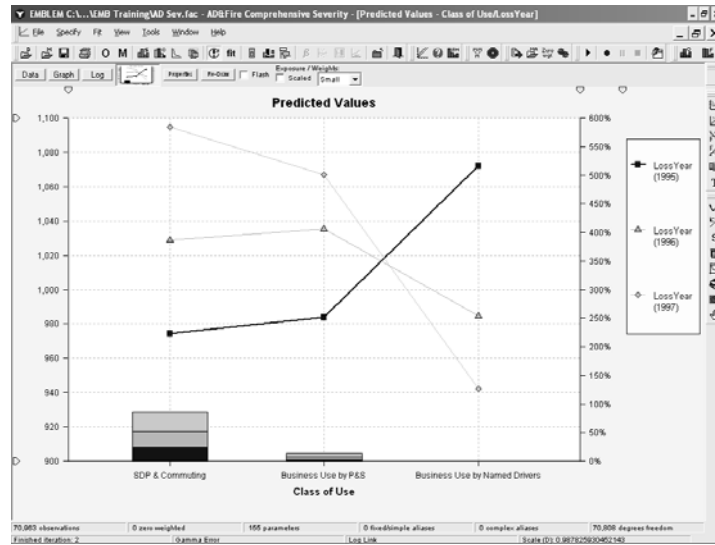
Parameter Number	Name	Value	Standard Error	Standard Error (%)	Alias Indicator (%)	Weight	Weight (%)	Exp(Value)
1	Mean	6.8842	0.02642	0.4		73,201	100.0	976.7263
-	Affinity Scheme (N)					50,982	69.6	
2	Affinity Scheme (Y)	-0.0284	0.00817	28.8		22,219	30.4	0.9720
-	Class of Use (SDP & Commuting)					62,936	86.0	
3	Class of Use (Business Use by P&S)	-0.0071	0.01096	154.8		10,012	13.7	0.9929
4	Class of Use (Business Use by Named Drivers)	-0.0490	0.06284	128.2		253	0.3	0.9522
-	Convictions (N)					60,505	82.7	
5	Convictions (Y)	0.0475	0.00985	20.7		12,696	17.3	1.0486
6	Driving Experience (0)	0.1222	0.03140	25.7		1,467	2.0	1.1299
7	Driving Experience (1)	0.0792	0.02656	33.5		2,505	3.4	1.0824
8	Driving Experience (2-4)	0.0540	0.01715	31.8		12,005	16.4	1.0555
9	Driving Experience (5-9)	0.0417	0.01317	31.6		20,056	27.4	1.0426
-	Driving Experience (10+)					37,168	50.8	

The value for the mean refers to a risk that is the base for all factors. Since this is the Severity dataset result, the value is the expected severity for the base risk on a linear predictor scale. The base level of each simple factor and custom factor included in the model is highlighted blue. In the screen above, the value of the mean is 6.8842. Since we used the Severity model using a gamma error structure with a log link function from the beginning, the expected Severity for overall mean is 976.7263, which can also be expressed as  $\text{Exp}(6.8842)$ . When a fitted model has a log link function, the table of fitted parameters includes an additional column showing the exponential of the linear predictor. This gives easy access to the relativities of different levels to the base level. The value for each non-base rating factor is the expected Severity on a log scale relative to the base risk. For the example above, a risk that is base for all rating factors other than "Affinity Scheme" (when "N" is base), with a value for "Affinity Scheme: Y" of -0.0284, will have an actual expected average cost  $\text{Exp}(6.8842-0.0284)$ . An alternative interpretation is that the non-standard risk will have an actual expected average cost that is 0.9720 times smaller than the base risk.

The Standard Error (%) field plays an important role to verify which parameters are significant and which not. The standard error percentages for "Class of Use" factor are over 50%. In EMBLEM, we can easily check the level of significance by looking at the color of font. If the color is green, it means the standard error percentage is below 50% and the parameter is therefore significant. If it is above 50%, then the color changes (green -> grey -> red).

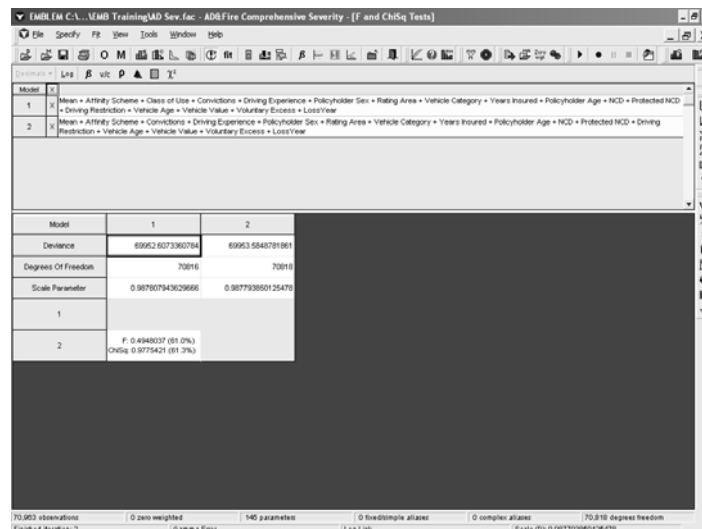
There is a non-statistical method for the exclusion/inclusion of a factor which is available to use in EMBLEM. It is to check how consistent any trends are for each time period. This can be achieved by fitting a time interaction with the factor being considered.

## INTRODUCING THE GENERALIZED LINEAR MODELS



From this time interaction graph, it is apparent that the factor "Class of Use" does not exhibit consistent trends for each time period.

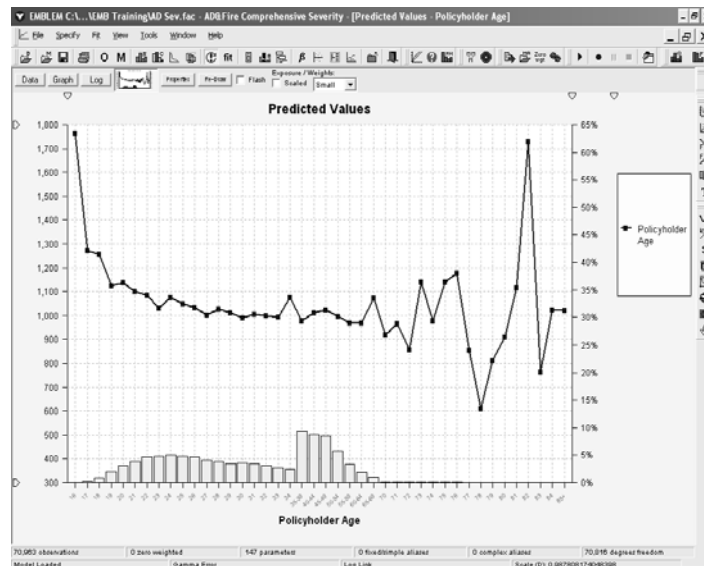
A good statistical test for the exclusion/inclusion of a factor is to see whether a model that includes the factor is significantly different from one that excludes the factor. This test can be carried out by fitting two models, one that includes the factor under consideration and another that excludes the factor. The results of the models can be compared using the F and chi-squared tests which is available in EMBLEM. Suppose we set a model which includes all parameters as a reference model. Then we exclude the "Class of Use" parameter, and compare the model with a reference model. Both chi-squared and F tests statistics are given to verify the significance between two models.



JH Yoo

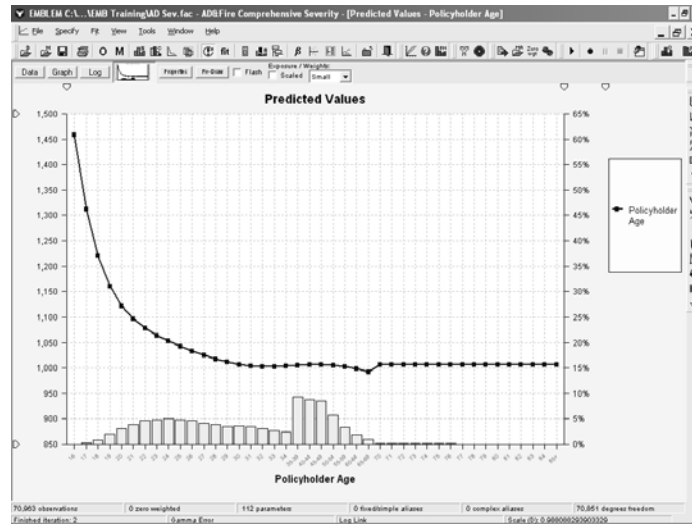
It is important to understand that the results of the F and chi-squared tests are shown in yellow boxes at the bottom of the lower grid. A low p-value for the tests indicates the two models are statistically different. However, as we can see, the results are shown in white boxes at the bottom. It indicates that models are not significantly different. Therefore, the exclusion of "Class of Use" is statistically justified. Bear in mind that a simplified model which is not significantly different from the unsimplified model is a more parsimonious model.

We can simply the model by editing parameters using grouping, smoothing, and other methods which are available in EMBLEM. By doing these methods, we are able to interpret the trend or effect for each parameter more easily and efficiently. For example, down below is an example of grouping and smoothing the policyholder age parameter. If we take a look at the graph without doing any editing, it is hard to see the trend of the parameter. However, we can use grouping and smoothing to simplify the parameter. By doing this, we are able to see the trend more clearly.



**Policyholder Age Parameter – Before grouping and smoothing**

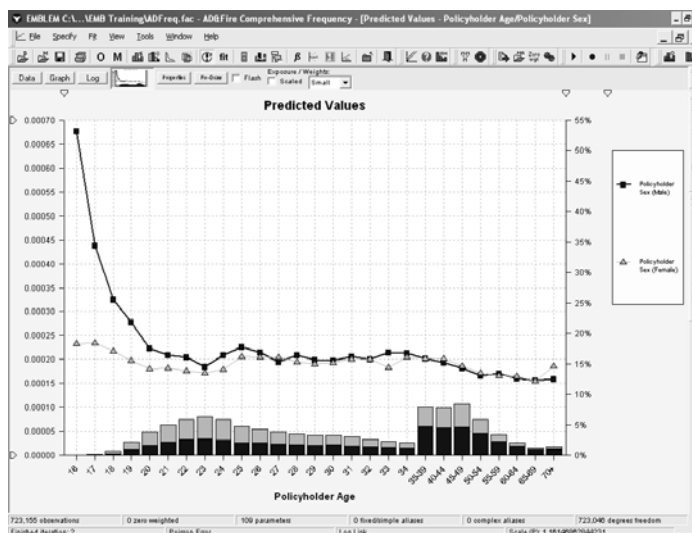
## INTRODUCING THE GENERALIZED LINEAR MODELS



**Policyholder Age Parameter – After grouping and smoothing**

### c. Interaction Tests

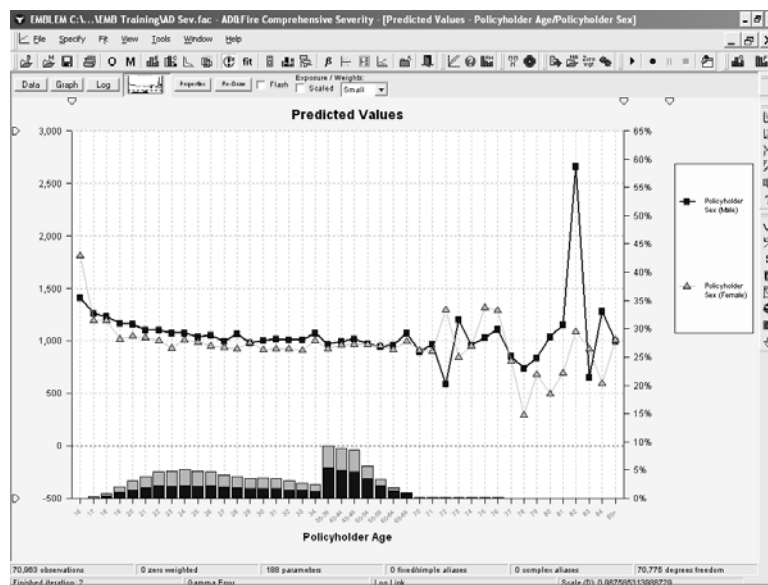
An interaction occurs in a dataset when the parameter estimates for a rating factor depend on the level of another rating factor. A common example found in motor insurance is that the experience of younger females is better than that for younger males, whereas the experience of older females is the same (or worse) than for older males. This time, we will look at the Frequency dataset.



JH Yoo

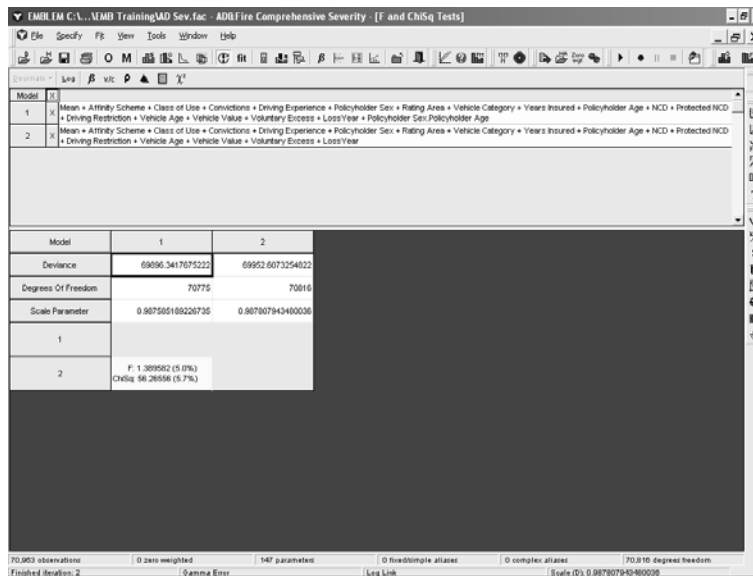
In this case, we should considering fitting interactions because there is a strong trend between two parameters. By fitting interactions we are essentially breaking the data down into finer and finer segments. This will generate more random fluctuations in experience, and a considerable amount of judgment is required to separate true underlying interactions from random noise.

However, more interactions are not always good when we want to find the “Best” model. To determine whether the interaction under investigation is sensible and should be retained in the final model, you should consider several factors. Remember that we need a model which is most simplified within same significance level. This means that we do not need any additional parameters which will not affect the significance of the model.

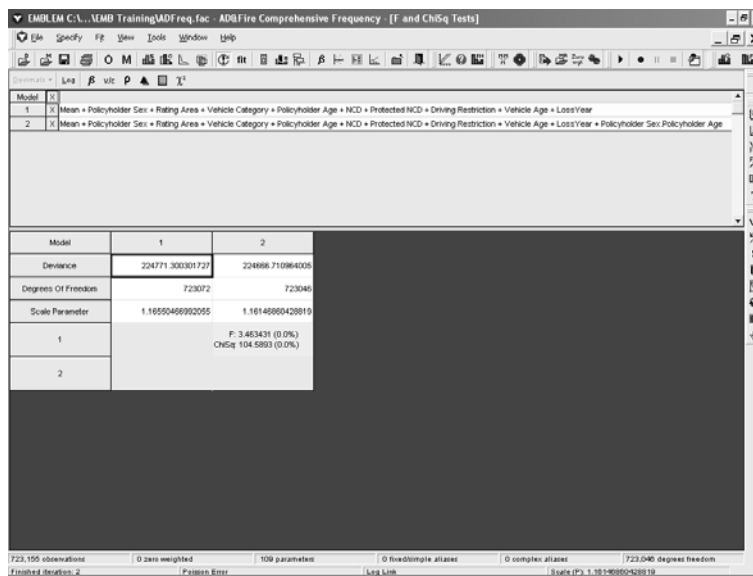


Let’s get back to the Severity model. If we do an interaction test for same parameters as the Frequency model, it is obvious that there does not seem to be any strong trend between two parameters. Yet, we are not sure whether this interaction is significant enough to include the model or not. Therefore, it is necessary to run the statistical test as before.

## INTRODUCING THE GENERALIZED LINEAR MODELS



When interpreting these tests, remember that adding an interaction term makes a model more complicated. If two models are not significantly different, the least complicated model (the model with the fewest parameters) is the model of choice. As we can see, the Chi-square statistics is 5.7% which is slightly over 5%. Therefore, we can say that there is not enough evidence to say these two models are significantly different. In conclusion, it is not really necessary to include this interaction to the model because we need the least complicated model.



This decision, however, can be reversed if we take a look at the Frequency model. As we found out earlier, the interaction between two parameters in the Frequency

model showed strong trend against each other. Clearly, the Policyholder sex parameter affects the Policyholder age parameter. If we run the statistics tests for this model, then it is clear that both the F and the Chi-Square statistics are below 5%. Even though our goal is to build the least complicated model, it is necessary to include the interaction because the trend is strong enough to affect the significance of the model.

#### **d. Process of GLM Modeling**

It is important to understand the whole process of GLM Modeling since a certain method can be used in a wrong period of processing. After we begin GLM modeling process, first of all, we have to look into each parameter and understand the importance and relevance. Sometimes, it is important to include a certain parameter even though it does not affect the significance of the model. After that, we can run a Univariate test. Test each parameter for significance and then decide which parameters will be included and which will not. It is quite useful to test parameters which were significant parameters from the past experience. After simplifying the model by reducing the number of parameters used in the mode, we then move to the Interaction tests for remain parameters. It is useful to check if there is any relationship between parameters which will eventually cause the Simpson's paradox if we only use univariate analysis. Then we repeated these procedures until we believe that our model is the "Best" model to use. EMBLEM not only allows users to do GLM Modeling by enabling many useful applications, but also allows them to do the modeling more easily and efficiently. In fact, applications of EMBLEM are easy to understand so even non-statisticians can use them without having to understand how GLM works in the mathematics term.

## **5. Summary and Conclusion**

By using the GLM modeling, we are able to analyze the dataset which does not follow the Normal distribution. We can even analyze the dataset which is not linearly distributed. Multiple useful yet practical distributions like Poisson, Binomial, Gamma and others are available to analyze the error structure. Interaction tests are available so that we can avoid the Simpson's Paradox while doing the multivariate analysis. Therefore, GLM helps us to analyze the large dataset with multiple response parameters more easily and efficiently. By using GLM, we can even add a newly founded response parameter to the model while verify it statistically.

EMBLEM makes the techniques of multivariate analysis accessible and approachable to both the statistician and non-statistician alike, thereby dramatically reducing the training required to model successfully and to produce results. More importantly, it allows the user to focus on the analysis itself and to impart his or her expertise in order to add value to the statistical process. EMBLEM not only provides us useful statistical or non-statistical test which were introduced throughout this paper, but also proves us more advanced techniques such as stepwise regression, underwriting scoring, and others. Also, by enabling VBA standard macros, EMBLEM saves us a lot of time and effort doing the modeling.

Even though these techniques are now available for us to use, we must not forget that they are not the ultimate solution for modeling and getting the "Best" result. GLM modeling and the EMBLEM tool only helps us to do our job easily. However,



they will only help us to get through the touch processing. In the end, our judgment is most important for modeling, and we will be able to improve it with more modeling experience.

## References

- [1] P. McCullagh and J.A. Nelder (1989), *Generalized Linear Models (Second Edition)*, Chapman & Hall, London.
- [2] Dobson, A. (1990), *An Introduction To Generalized Linear Models*, London: Chapman and Hall.
- [3] Firth, D. (1991), "Generalized Linear Models," in *Statistical Theory and Modelling*, ed. Hinkley, D.V., Reid, N., and Snell, E.J., London: Chapman and Hall.
- [4] Nelder, J.A. and Wedderburn, R.W.M. (1972), "Generalized Linear Models," *Journal of the Royal Statistical Society A*, 135, 370 -384.

YOO, JONG HWAN: Representative Consultant  
Seoul, Korea  
Cell Phone: 82 11 9263 5405  
Office : 82 2 518 1024  
          82 2 514 1026  
Fax: 82 2 514 1024  
E-mail: [jhyoo@glisconsulting.com](mailto:jhyoo@glisconsulting.com)