

Article

# Estimating Major Risk Factor Relativities in Rate Filings Using Generalized Linear Models

Shengkun Xie <sup>1,\*</sup> and Anna T. Lawniczak <sup>2</sup><sup>1</sup> Ted Rogers School of Management, Ryerson University, Toronto, ON M5B 2K3, Canada<sup>2</sup> Mathematics and Statistics, University of Guelph, Guelph, ON N1G 2W1, Canada; alawnicz@uoguelph.ca

\* Correspondence: shengkun.xie@ryerson.ca; Tel.: +1-416-979-5000

Received: 26 July 2018; Accepted: 2 October 2018; Published: 11 October 2018



**Abstract:** Predictive modeling is a key technique in auto insurance rate-making and the decision-making involved in the review of rate filings. Unlike an approach based on hypothesis testing, the results from predictive modeling not only serve as statistical evidence for decision-making, they also discover relationships between a response variable and predictors. In this work, we study the use of predictive modeling in auto insurance rate filings. This is a typical area of actuarial practice involving decision-making using industry loss data. The aim of this study was to offer some general guidelines for using predictive modeling in regulating insurance rates. Our study demonstrates that predictive modeling techniques based on generalized linear models (GLMs) are suitable in auto insurance rate filings review. The GLM relativities of major risk factors can serve as the benchmark of the same risk factors considered in auto insurance pricing.

**Keywords:** rate Filings; auto insurance regulation; generalized linear models; rate-making; predictive modeling

**JEL Classification:** G22

## 1. Introduction

Modeling aggregate loss (Duan 2018; Frees 2014; Meyers 2007; Shi 2016) using insurance risk factors is a key aspect in the decision-making of rate change review application. In Duan (2018), a logistic regression model was proposed to classify the loss data of a Chinese company into different risk levels using the burden index. Multivariate negative binomial models for the insurance claim counts were proposed by Shi (2014) to capture the potential dependent structure among the different claim types. In David (2015), the Poisson regression and negative binomial models were applied to a French auto insurance portfolio to investigate the degree of risk using claim counts data. In Najafabadi (2017), a k-inflated negative binomial mixture regression model was proposed to model the frequency and severity of reported claims in an Iranian third-party insurance data set. Authors in Tang (2014) proposed a new risk factor selection approach based on EM algorithm and LASSO penalty, for a zero-inflated Poisson regression model. The analysis of car insurance data from the SAS Enterprise Miner database was then used to show the usefulness of the proposed method in rate-making. However, all those works focused on the study of the loss experience of an individual company, rather than on the total loss at the industry level. In auto insurance regulation, approval of the submitted rate change application by insurance companies requires significant statistical evidence from a company's past loss experience Chen (2009). Thus, a study of total loss behavior at the industry level becomes important in providing a constructive review of a company's rate change applications.

Auto insurance companies often consider various risk factors based on the available loss experience, such as age, number of years of license, gender, number of years without accidents, territory, and vehicle rate group. In addition, various socio-demographic factors are used to differentiate the risk levels of drivers. More recently, usage-based insurance (UBI) programs have been offered by many major insurance companies. The analysis of data associated with UBI (e.g., mileages, traffic conditions, and driving habits) are new types of rating factors (Cripe 2012; Husnjak 2015; Ma 2018). However, this is not the case for insurance regulators because the detailed individual loss information is not available for the review of a rate change. Of course, insurance regulators are not interested in reproducing company results when reviewing the rate change application, but they need an analysis based on the aggregated loss experience from major insurance companies or the whole industry. A review decision made by a regulator must be supported by key findings on the rating factors at the industry level. Since insurance companies use many risk factors in pricing, the question arises as to whether one can focus only on the major rating factors in the review process. Answering to this question calls for use of suitable predictive modeling techniques that can be applied in rate filings review, particularly in classification of risk (Hamid 2011; Watson 2004). The aim of such modeling techniques is to find better understanding of the characteristics of major risk factors. In Canada, among many other risk factors, Class and Driving Record (DR) are the most important factors. The DR is a factor that includes levels 0, 1, 2, ... 6+, where each level represents the accumulated non-accident years of an insured, while Class represents a combination of driver age, gender, and use of vehicle. The information that is used for carrying out relativity estimates is either loss cost or loss ratios. None of the current predictive modeling of insurance loss uses industry aggregate loss data, and focus on major risk factors. This motivated us to investigate if major risk factors are able to capture most of the variation of total loss data at an industry level.

In predictive modeling, to better capture the major variation of loss data, considerable effort has been spent on finding the optimal solution in terms of the minimum overall bias (Zhao 2011; David 2015; Frees 2015; Garrido 2016). Due to its significant impact on insurance rate regulation, loss cost or loss ratios modeling became popular (Harrington 1992; Tennyson 2009). When loss cost is used, the bias is defined as the difference between the predicted loss cost and its actual observed value. The minimum overall bias considers both the bias of estimating pure premium for each class of insured and its associated number of exposures. To estimate the risk relativities, minimum bias procedure (MBP) introduced by Bailey (1963) has often been used, and it has become a traditional approach in non-life insurance pricing (Feldblum 2003). Recent research on actuarial practice has demonstrated that generalized linear models (GLMs), an advanced statistical method, have become prevalent in the field of rate-making, risk modeling, or forecasting in most of the European and Asian countries (Ohlsson 2010). In Zhao (2011), GLM was used to model the loss frequency and loss severity of automobile insurance in China to analyze the impacts of automobile, human, and rating territory on loss frequency and severity. In David (2015), an overview of GLM in the calculation of the pure premium given the observable characteristics of the policyholders was presented. In Garrido (2016), GLM was fitted to the marginal frequency and the conditional severity components of the total claim cost with the introduction of dependence between the number of claims and the average claim size was studied. The main reason for the prevalence of GLM is that it enables a simultaneous modeling of all possible risk factors as well as determination of the retention of risk factors in the model.

In the risk classification system of auto insurance companies, rate-making is one of the most important aspects among many others, such as underwriting and marketing strategies (Outreville 1990). Its major goal is to develop a set of risk factor relativities that can be further used for pricing an insurance policy. Also, rate-making that uses industry-level data is a typical task in a rate filings review. Even though an insurance company has access to the transactional loss and claim data, it is clear that the rate-making process is based on the aggregate level of the company data. This is because it is not in the company's interest to evaluate the risk of a single exposure. The rate-making is done based on the average loss costs of each combination of the levels of risk factors. The average

loss cost (or simply the loss cost in the following discussion) is defined as the total loss divided by the total exposures within each possible combination of the levels of risk factors, where the total loss and the total exposures are the aggregate measures of loss and exposures for that particular combination. Traditionally, within actuarial practice, the estimates of rating factor relativities are conducted by using MBP (Feldblum 2003). In rate filings reviews, often the rating variables of Class and DR are considered. The relativities are estimated by MBP separately for each data set, that is, from each respective combination of different years, territories, and coverages. The problem with this approach as a rate-making technique is that the potential interactions among rating variables are not considered. Thus, the results may not be comparable. Also, MBP is a numerical approach that is unable to statistically evaluate the difference between relativities.

In this paper, we present the results of a comparative study employing MBP and GLM as modeling tools. We focus on the study of industry-level data used in rate filings reviews to decide if major data variation could be retained in the models, and how it is affected by the risk factors under consideration. We investigate the consistency of the results obtained from both MBP and GLM methods. It is expected that one could replicate the results obtained by MBP by applying GLM to the same data set. For ease of understanding the interaction of rating variables, we mainly focus on the predictive modeling of loss cost using DR, Class, and accident year as the major rating variables. The significance of this work is in providing general guidelines for the use of predictive modeling in an insurance rate filings review for auto insurance regulators. In particular, this work demonstrates the usefulness of GLM in rate-making for rate filing purposes. This approach can help us to understand how a decision is being made when focusing only on major risk factors. This paper is organized as follows. In Section 2, we summarize the data used in this work, and discuss the methodologies, including MBP and GLM, used for producing the results presented in this paper. In Section 3, comparative results obtained from MBP and GLM under different model settings are analyzed. Finally, in Section 4, we conclude our findings and provide summary remarks.

## 2. Methods

### 2.1. Minimum Bias Procedure

Let  $r_{ij}$  be a collection of observed values. It can be loss severity, claim counts, average loss cost, or loss ratio, depending on the interest of modeling, and how we define a response variable. Let  $x_{1i}$  be the value contributed by the  $i$ th DR level (i.e., DR takes values 0, 1, 2, 3, 4, 5, 6+), and let  $x_{2j}$  be the value contributed by the  $j$ th level of Class. Also, let  $w_{ij}$  be the number of exposures of risk in the  $(i,j)$ th combination of levels of the underlying risk factors. The objective of MBP is to find optimal solutions for  $x_{1i}$  and  $x_{2j}$ , subject to the following two sets of equations as constraints (Feldblum 2003):

$$\sum_j w_{ij}(r_{ij} - (x_{1i} + x_{2j})) = 0, \quad \forall i, \text{ where } i = 1, 2, \dots, 7,$$

$$\sum_i w_{ij}(r_{ij} - (x_{1i} + x_{2j})) = 0, \quad \forall j, \text{ where } j = 1, 2, \dots, 14.$$

Numerically, MBP iteratively solves for  $x_{1i}$  and  $x_{2j}$  using the following two equations, until  $x_{1i}^{(l)}$  and  $x_{2j}^{(l)}$  converge at the  $l$ th step, where

$$x_{1i}^{(l+1)} = \frac{\sum_j w_{ij}(r_{ij} - x_{2j}^{(l)})}{\sum_j w_{ij}},$$

$$x_{2j}^{(l+1)} = \frac{\sum_i w_{ij}(r_{ij} - x_{1i}^{(l)})}{\sum_i w_{ij}}.$$

The relativities are obtained by selecting  $k$ th levels as a base, which are defined as  $\frac{x_{1i}}{x_{1k}}, \forall i$  for DR, and  $\frac{x_{2j}}{x_{2k}}, \forall j$  for Class.

## 2.2. Generalized Linear Models for Rate-Making

The main idea of generalized linear models (GLMs) (De Jong 2001; Haberman 1996; Ohlsson 2010) in rate-making is to model a transformation of the expected value of a response variable, which in our case is  $r_{ij}$ . The transformation function  $g(\mu_{ij})$  is called the link function, where  $\mu_{ij} = E(r_{ij})$ . The model that is used to explain the transformation function  $g$  is a linear model, and it can be written as follows:

$$g(\mu_{ij}) = \mathbf{Y}^T \boldsymbol{\beta}, \text{ or } g(r_{ij}) = \mathbf{Y}^T \boldsymbol{\beta} + \epsilon_{ij},$$

where  $\mathbf{Y}^T = [y_{1i}, y_{2j}]$  and  $\boldsymbol{\beta}^T = [\beta_{1i}, \beta_{2j}]$ . Here,  $y_{1i}$  and  $y_{2j}$  are dummy variables. That is,  $y_{1i}$  takes the value 1 when  $i$  corresponds to the  $i$ th level of DR, otherwise it is zero.  $y_{2j}$  takes the value 1 when  $j$  corresponds to the  $j$ th level of Class, otherwise it is zero.  $\epsilon_{ij}$  is the error random variable assumed to have a certain probability distribution function.  $\mathbf{Y}^T \boldsymbol{\beta}$  is referred to as a systematic component. The variance of  $\epsilon_{ij}$  is assumed to have the following functional relationship with the mean response:

$$\text{Var}(\epsilon_{ij}) = \frac{\phi V(\mu_{ij})}{\omega_{ij}},$$

where  $V(x)$  is called a variance function. The parameter  $\phi$  scales the variance function  $V(x)$ , and  $\omega_{ij}$  is a constant assigning a weight. This result comes from a family of distributions called the exponential family (McCullagh 1989), which determines the parameters  $\phi$  and  $\omega_{ij}$ . The case when  $V(x) = 1$  implies a normal distribution. If  $V(x) = x$ , then the distribution is Poisson. If  $V(x) = x^2$ , then it is a gamma distribution, and if  $V(x) = x^3$ , then it is an inverse Gaussian distribution (Mildenhall 1999). These distributions are the distributions used in this work. In general, the following relationship between variance function and mean value of response is considered:

$$V(\mu_{ij}) \sim \mu_{ij}^p.$$

We discussed the cases when  $p=0, 1, 2$ , and  $3$ . For  $p < 0$  or  $p > 2$ , the distributions are called Tweedie distributions (note that the inverse Gaussian belongs to this class of distribution) (Dunn 2001; Tweedie 1957).

In the case when  $g$  is an identity function, GLM becomes a general linear model. Thus, the relativities are obtained by computing the exponential transformation of model coefficients, which are denoted by  $\exp\{\beta_{1i}\}$  and  $\exp\{\beta_{2j}\}, \forall i$  and  $j$  (Ohlsson 2010). In this work, the following model specifications are investigated:

- Systematic component:  $\mu(\text{Loss Cost}) \sim \text{Class} + \text{DR}$ ;
- Error distribution functions: Gaussian, Poisson, inverse Gaussian, and gamma;
- The multiplicative model, resulting in a log link function for GLM, that, is  $g = \log$ ;
- The number of exposures is used as a weight value.

From the way the systematic component for the given data is modeled, one can see that for each combination of risk factor levels, the loss cost has a unique distribution, and the model specifies a common variation captured by the error distribution function. Because the standard deviation estimate is available for each level of risk factors, one can easily construct their confidence intervals based on the normal approximation approach. To validate the choice of log-scale, a Box-Cox transform (Osborne 2010) can be applied. If the obtained parameter value is close to zero, this implies that a selection of the log link function is appropriate for the given data.

### 2.3. Handling Weights in GLM

In general, it is expected that GLM will be applied to either an industry-level data summary or a company's transactional data summary. There are no fundamental differences between these two cases, as they are only subject to different levels of data aggregation. In transactional data summary, the loss costs are calculated in exactly the same way as in the industry data summary. It is important to consider a suitable weight function for each cell data, so that the variance function of the pre-defined error distribution in GLM is appropriately modeled. The number of exposures associated with each cell of the two-way data summary (by DR and Class) is used as a weight for the loss cost approach. The main purpose is to offset the different levels of data variation among the cells. Under the definition of the weight function in GLM, the assumption of having an identical error distribution for each cell of loss cost becomes reasonable.

In practice, there is flexibility in specifying the weight function, but this depends on how each cell data within the summary is defined. In addition to the use of the number of exposures at each combination as a weight, a new type of weight function can be defined by

$$w_{ij} = \frac{E_i \times E_j}{E_{total}^2},$$

where  $E_i$  is the one-way summary of the number of exposures for the  $i$ th level of DR, and  $E_j$  is the one-way summary of the number of exposures for the  $j$ th level of Class.  $E_{total}$  is the total number of exposures. Also, another weight function can be defined by

$$w_{ij} = \frac{E_i}{E_{total}}$$

when one wants to emphasize the importance of DR, by ignoring the effect of the number of exposures from the Class. Similarly, a third weight function can be defined by

$$w_{ij} = \frac{E_j}{E_{total}}.$$

### 2.4. Inclusion of Year as a Major Risk Factor

The estimates of relativities obtained from multiple-year loss experience are often considered to be more robust. One approach is to combine the multiple-year data together and ignore their year labels. In this case, the credibility of loss cost for each combination is improved as the corresponding number of exposures is increased with the increase of number of years of data. Thus, the number of exposures is now based on multiple years of aggregation. However, the disadvantage of this approach is that the effect from different years on relativity estimates is transferred to the relativities of other risk factors. A better approach is to add *Year* as a major risk factor to the GLM, in which the systematic component is re-defined as

$$\log(\text{Loss Cost}) \sim \text{Class} + \text{DR} + \text{Year}.$$

Since we added *Year* as another variable,  $Y^T = [y_{1i}, y_{2j}, y_{3k}]$  and  $\beta^T = [\beta_{1i}, \beta_{2j}, \beta_{3k}]$ , where  $y_{3k}$  is a dummy variable taking the value 1 when  $k$  corresponds to the  $k$ th level of *Year*. The  $\beta_{3k}$  is the associated coefficient. In this approach, the effect from the factor *Year* can be estimated. Also, the data variation captured by the model is improved by the consideration of an additional risk factor. Under this approach, the estimate of relativity is less dependent on the choice of an error distribution function. This implies the robustness of estimating relativity using GLM. The relativity of *Year* reflects the level of loss cost for that particular year. Since loss costs of later years have not yet fully developed to the ultimate loss level, a smaller value of relativity of *Year* may only mean a lower loss cost for that particular year, based on the available reported loss amounts.

### 2.5. Fundamental Difference between MBP and GLM

The use of loss cost as a response variable is a common approach when applying MBP to industry-level data (Brown 1988; Garrido 2016; Ismail 2009). The loss cost approach considers both the relativities caused by loss frequency and loss severity. Within the loss cost approach, the relativities are obtained by a conditional iterative process, which implies that the  $l$ th step of the output of the algorithm becomes the input of the  $(l+1)$ th step, Brown (1988). The minimization of the total bias is based on the marginal balance principle that is applied to each column and each row of the loss cost data matrix, respectively, where the row and the column correspond to DR and Class. Essentially, MBP treats the bias within each cell of the data matrix as being independently and identically distributed, without actually measuring its distribution. This method is less practical when the total number of risk factors is large, due to the nature of the numerical approach. This approach does not give optimal estimate for base rate (Brown 1988), which is important for calculating pure premium. In terms of numerical convergence, the initial value condition plays an important role. The initial values are assigned based on the one-way summary of loss costs. Due to the nature of the iterative process, the final solution is often affected by the choice of initial values. Each value will be adjusted based on the result from the previous step.

Unlike MBP, GLM specifies an error distribution function for the bias of loss costs, and transforms observed values by link function to improve the linearity between the transformed response variable and the risk factors. The transformation is introduced to model a potential non-linear relationship via an appropriate mapping from the original space of the mean value of observations to a linear feature space. The selected error distribution function mainly captures the true distribution of underlying loss cost. Instead of placing a balance constraint for each row and column of data, GLM uses the assumption that within each row and each column, the error function is common to the bias. This constraint has a similar effect on determining a unique solution, as it was observed in MBP. However, in MBP, an iterative process is used to solve a non-linear optimization problem. The idea of using the GLM method is to fit the bias to a certain distribution and use a statistical approach to estimate the distribution parameters. In this case, an interval estimate of relativity becomes possible. Because of this, the use of GLM is more powerful than using MBP in terms of the statistical validation of test data. When specifying an equivalent constraint for both MBP and GLM, one can recover the same estimates of relativities from both MBP and GLM (Brown 1988; Mildenhall 1999).

With regards to making a prediction of loss costs, both MBP and GLM require a common base rate for all rating variables. The common base estimated using GLM corresponds to the intercept of the model. For MBP, it is calculated from the total loss divided by the total exposures for all risk factors. Therefore, the performances of MBP and GLM differ in minimizing overall bias: the base rate from GLM is a model estimate, but the base rate for MBP is just an empirical estimate. However, the estimated relativities can be the same for certain cases for both methods.

### 2.6. Some Discussions

In general, the use of GLM enables us to select variables according to their statistical significance. This is likely why GLM has been considered to be the most powerful tool for non-life insurance pricing (Ohlsson 2010). One of the important tasks of the pricing problem is ensuring that risk factors and their levels are included in the model. As the dimension of the risk factors and their associated number of levels increase, it becomes challenging to interpret the impact of particular combinations of the rating variables on the loss cost. Therefore, focusing on the most significant rating variables may be superior. However, we do not encounter this difficulty within the scope of this work, as we only consider a few important risk factors that appear in the rate filings review. The relativities estimates of major risk factors become a general guideline for rate filings reviewers to better understand the nature of insurance rate fluctuations when comparing the results from different companies. Thus, the selection of a proper approach makes rate regulation statistically more sound, as the results are less affected by reviewer experience and judgement.

Recently, the generalized additive model (GAM) has been proposed in predictive modeling for rate-making (Denuit 2004; Fuzi 2016; Klein 2014). From a technical perspective, GAM is an extension of GLM. In statistics, data are categorized into two types, numerical or categorical. When modeling data in practice, there is often no clear-cut rule about the data type. In most cases, it depends on expert opinion and the purpose of modeling. This means that for some of the variables within rate-making, we can have different types of treatments. For instance, the DR takes values 0, 1, 2, 3, 4, 5, 6+. If our goal is to estimate the average differential for each driving record, we assume that the collected data are categorical. However, since there is a natural ordering of the data, we expect that with the increase of the value of driving record, the differential decreases. It is reasonable to expect that there is a monotonic relationship within different driving records. Therefore, one can impose this relationship on a model by specifying the functional form of the Driving Record. This is the idea behind the extension of GLM to GAM.

### 3. Results

In this study, we considered an industry-level data set for accident years from 2007–2009, with third party liability (TPL) coverage and urban territory. The relativities of DR and Class at various levels were obtained for the accident year 2007 data set by both MBP and GLM. Additionally, a data set comprising these three accident years as a whole was used for relativity estimation for both cases of with and without *Year* as a rating factor. The modeling and analysis were based on the aggregate measure of each combination of levels among all factors that we considered. The summary of average loss costs is reported in Table 1. The value of each cell represents the loss cost of each given combination of levels from DR and Class. Notice that in the table the NA values mean that the loss costs were not available for these cells, which causes the problem of dealing with the missing values in the later computation. In some cases, zero values of the loss cost appeared in the summary data table. If this happens, then zero values were simply replaced by NAs. Ideally, we expected to have full information in the summary data table. Since we used industry-level data, the impact of missing values on some combinations of DR and Class was reduced when compared to the cases that dealt with company-level data. Thus, it makes sense to use the estimated relativities as a benchmark in rate filings review.

**Table 1.** Two-way summary of loss cost by Driving Record (DR) and Class. The values represent the dollar amount of average loss cost for the given combination of Class and DR. Note that there are no levels 4 and 14–17, because this is how the level is specified in the given data set.

	Class1	Class2	Class3	Class5	Class6	Class7	Class8	Class9	Class10	Class11	Class12	Class13	Class18	Class19
DR0	559	824	675	372	517	2006	2359	2513	3087	1109	1507	397	946	461
DR1	510	644	508	146	433	660	2398	1435	1300	995	1113	1951	644	436
DR2	395	709	580	90	149	697	290	1010	1533	992	1151	1068	558	827
DR3	563	573	648	137	187	470	509	544	2201	1218	881	1167	848	471
DR4	458	450	686	145	78	1288	147	787	936	849	679	578	532	776
DR5	388	444	585	62	224	329	708	214	NA	152	629	659	320	476
DR6	275	301	374	78	90	345	NA	657	NA	NA	295	479	NA	274

The obtained results of model coefficients and relativities are presented in Figure 1 for the case of using Gaussian distribution as an error function. The results include both the results of the general linear model and of the generalized linear models. From Figure 1, we can see that there were no significant large differences among the estimates of relativities. However, the general linear model gave slightly higher overall average relativities, which implies that the pure premium using results from the general linear model would be overestimated when compared to the generalized linear models. When GLM results were compared to the ones from MBP, a similar pattern was observed. This confirms our expectation that GLM would perform better in terms of reducing the effects of the interaction of risk factors, leading to lower estimates of relativities. However, this is only applicable for a Gaussian error function, which implies that the underlying assumption of the loss cost has to have a lighter tail distribution. This suggests that in modeling loss cost using GLM with a Gaussian error function, a large loss, corresponding to catastrophic events, must be removed before fitting the models to the data.

This is because the loss data without removing large loss often follow a heavy-tailed distribution (Bradley 2003; McNeil 1997).

	Gaussian		DRo and CLASS <sub>1</sub> AS a BASE		DR <sub>3</sub> and CLASS <sub>2</sub> AS a BASE		MBP	One-Way Relativity
	V <sub>1</sub> (Coef)		EXP(V <sub>1</sub> )		EXP(V <sub>1</sub> )			
	Generalized(log link)	Linear (Identity link)	Generalized(log link)	Linear (Identity link)	Generalized(log link)	Linear (Identity link)		
(Intercept)	6.5431	6.6070	694.45	740.28	790.85	947.03		
DRo	0.0000	0.0000	1.000	1.000	1.285	1.416	1.349	1.402
DR <sub>1</sub>	-0.2575	-0.2523	0.773	0.777	0.993	1.100	1.064	1.111
DR <sub>2</sub>	-0.2356	-0.3326	0.790	0.717	1.015	1.015	1.051	1.122
DR <sub>3</sub>	-0.2506	-0.3477	0.778	0.706	1.000	1.000	1.000	1.000
DR <sub>4</sub>	-0.4479	-0.5224	0.639	0.593	0.821	0.840	0.859	0.914
DR <sub>5</sub>	-0.5898	-0.6337	0.554	0.531	0.712	0.751	0.743	0.779
DR <sub>6</sub>	-0.9433	-0.9965	0.389	0.369	0.500	0.523	0.518	0.534
AVG of DR					0.576	0.601	0.596	
CLASS <sub>1</sub>	0.0000	0.0000	1.000	1.000	0.886	0.904	0.898	0.858
CLASS <sub>2</sub>	0.1206	0.1014	1.128	1.107	1.000	1.000	1.000	1.000
CLASS <sub>3</sub>	0.2778	0.3019	1.320	1.352	1.170	1.222	1.205	1.219
CLASS <sub>5</sub>	-1.2296	-1.3333	0.292	0.264	0.259	0.238	0.244	0.425
CLASS <sub>6</sub>	-0.8311	-0.9608	0.436	0.383	0.386	0.346	0.370	0.625
CLASS <sub>7</sub>	0.2886	0.2151	1.335	1.240	1.183	1.120	1.115	1.137
CLASS <sub>8</sub>	0.5864	0.1478	1.797	1.159	1.593	1.048	1.560	2.726
CLASS <sub>9</sub>	0.6564	0.4939	1.928	1.639	1.709	1.481	1.674	2.267
CLASS <sub>10</sub>	1.3519	1.3259	3.865	3.766	3.426	3.403	3.474	6.255
CLASS <sub>11</sub>	0.6990	0.6834	2.012	1.981	1.783	1.790	1.814	3.137
CLASS <sub>12</sub>	0.5532	0.4543	1.739	1.575	1.541	1.423	1.500	2.192
CLASS <sub>13</sub>	0.6096	0.5539	1.840	1.740	1.631	1.572	1.633	2.026
CLASS <sub>18</sub>	0.3414	0.3342	1.407	1.397	1.247	1.262	1.268	2.236
CLASS <sub>19</sub>	0.1236	0.0995	1.132	1.105	1.003	0.998	1.023	1.358
AVG of CLASS					0.973	0.976	0.978	

**Figure 1.** Model coefficients (i.e.,  $V_1$  columns) and relativities (i.e.,  $EXP(V_1)$  columns) obtained respectively under generalized linear model (GLM) with Gaussian error distribution function & log link function and under GLM with Gaussian error distribution function & the identity link function. The loss cost data from the accident year 2007 was used. MBP: minimum bias procedure.

The results obtained by using other error distributions, including gamma and inverse Gaussian, are presented in Figure 2. The estimates of relativities for both DR and Class were larger than the ones obtained from MBP or Poisson distribution. The MBP and Poisson distribution led to the same theoretical results—see Brown (1988). This may mean that the use of a heavy tail distribution in an error distribution function gives larger estimates of relativities. Risk increased when the tail of the loss distribution was heavier, resulting in larger relativities for the rating factors. When the relativities obtained from GLM or MBP were compared to one-way relativities, we saw that they were mostly underestimated. This suggests that more risk factors may need to be included in order to better capture data variation. Rate filings reviews do not aim at pricing, but focus instead on producing results that help to better understand the nature of major risk factors used by companies.

By looking at the loss cost or the one-way relativities of loss cost for DR1 and DR2 (i.e., the last column in both Figures 1 and 2), we realized that there was a reversal for the relativity of DR1 and DR2. We expected a higher relativity for the insured who had less experience. Unfortunately, the results from the one-way summary were not as good as the ones obtained by GLM or MBP because the one-way relativity based on a two-way summary loss cost ignores the differences of the number of exposures among cells. Therefore, in order to better model loss costs using GLM, a suitable choice of a weight function needs to be specified, reflecting the difference of data value in each cell caused by the different levels of loss aggregation.

The results of relativity estimates and their comparisons under different model assumptions are reported in Figures 3–6. In Figures 3 and 4, the overall pattern for DR was consistent for all cases. Relativity decreased with the increase of the number of non-accident years. This means that prior driving record proved to have validity in predicting a driver’s accident risk. This result coincides with the results of in Peck (1983), who demonstrated the significant predictive power of DR as a rating factor. In all cases, the relativities of DR obtained from GLM were close to the actual relativities

calculated from the one-way summary. For Classes 1 to 7, which contained the majority of the risk exposures, the estimated relativities for all considered models were close to the actual one-way estimates. The model estimates' departure from the actual one-way relativities became significant only for those Classes that contained a smaller number of risk exposures. The values of the coefficients and relativities of the considered methods are displayed in Figure 7. Overall, these values coincided with the actual values, but GLM underestimated the relativities at some levels of Class—namely, Class8, Class10, Class11, and Class18. We believe that this is due to both the missing values and the level of credibility of data from those classes. Combining Class with DR, we were able to see that Class was successful in explaining the loss cost data. A similar research outcome was reported in Lourens (1999). Unlike a single class with multiple characteristics as in our work, drivers' sex, age, and level of education were used in Lourens (1999).

	V <sub>1</sub> (Coef)			DR <sub>0</sub> and CLASS <sub>1</sub> AS A BASE EXP(V <sub>1</sub> )			DR <sub>3</sub> and CLASS <sub>2</sub> AS A BASE EXP(V <sub>1</sub> )			MBP	One-Way Relativity
	Gamma	Poisson	Inverse Gaussian	Gamma	Poisson	Inver.Gauss	Gamma	Poisson	Inverse Gaussian		
(Intercept)	6.66717	6.56395	6.90985	786	709	1002	1085	856	1749		
DR <sub>0</sub>	0.00000	0.00000	0.00000	1.000	1.000	1.000	1.526	1.344	1.922	1.349	1.402
DR <sub>1</sub>	-0.28298	-0.23742	-0.40088	0.754	0.789	0.670	1.150	1.060	1.287	1.064	1.111
DR <sub>2</sub>	-0.37133	-0.24853	-0.68348	0.690	0.780	0.505	1.052	1.048	0.970	1.051	1.122
DR <sub>3</sub>	-0.42242	-0.29582	-0.65332	0.655	0.744	0.520	1.000	1.000	1.000	1.000	1.000
DR <sub>4</sub>	-0.55469	-0.45061	-0.80964	0.574	0.637	0.445	0.876	0.857	0.855	0.859	0.914
DR <sub>5</sub>	-0.68957	-0.59600	-0.92040	0.502	0.551	0.398	0.766	0.741	0.766	0.743	0.779
DR <sub>6</sub>	-1.05704	-0.95752	-1.29788	0.347	0.384	0.273	0.530	0.516	0.525	0.518	0.534
AVG of DR							0.612	0.594	0.612	0.596	
CLASS <sub>1</sub>	0.00000	0.00000	0.00000	1.000	1.000	1.000	0.905	0.898	0.908	0.898	0.858
CLASS <sub>2</sub>	0.10006	0.10776	0.09652	1.105	1.114	1.101	1.000	1.000	1.000	1.000	1.000
CLASS <sub>3</sub>	0.30377	0.29452	0.30893	1.355	1.342	1.362	1.226	1.205	1.237	1.205	1.219
CLASS <sub>5</sub>	-1.30220	-1.26143	-1.36074	0.272	0.283	0.256	0.246	0.254	0.233	0.244	0.425
CLASS <sub>6</sub>	-0.89934	-0.85541	-0.97930	0.407	0.425	0.376	0.368	0.382	0.341	0.370	0.625
CLASS <sub>7</sub>	0.23256	0.25182	0.22531	1.262	1.286	1.253	1.142	1.155	1.137	1.115	1.137
CLASS <sub>8</sub>	0.44473	0.53990	0.23871	1.560	1.716	1.270	1.412	1.541	1.153	1.560	2.726
CLASS <sub>9</sub>	0.61420	0.62362	0.63693	1.848	1.866	1.891	1.672	1.675	1.717	1.674	2.267
CLASS <sub>10</sub>	1.34839	1.35100	1.32907	3.851	3.861	3.778	3.485	3.467	3.430	3.474	6.255
CLASS <sub>11</sub>	0.70853	0.70173	0.70663	2.031	2.017	2.027	1.838	1.811	1.841	1.814	3.137
CLASS <sub>12</sub>	0.46444	0.51273	0.40763	1.591	1.670	1.503	1.440	1.499	1.365	1.500	2.192
CLASS <sub>13</sub>	0.58335	0.59774	0.56892	1.792	1.818	1.766	1.621	1.632	1.604	1.633	2.026
CLASS <sub>18</sub>	0.34876	0.34417	0.33834	1.417	1.411	1.403	1.282	1.267	1.274	1.268	2.236
CLASS <sub>19</sub>	0.12555	0.12990	0.11738	1.134	1.139	1.125	1.026	1.022	1.021	1.023	1.358
AVG of CLASS							0.982	0.979	0.983	0.978	

Figure 2. Model coefficients (i.e., V<sub>1</sub> columns) and relativities (i.e., EXP(V<sub>1</sub>) columns) obtained respectively under GLM with log link function and various error distribution functions including gamma, Poisson, and inverse Gaussian. The loss cost data from the accident year 2007 was used.

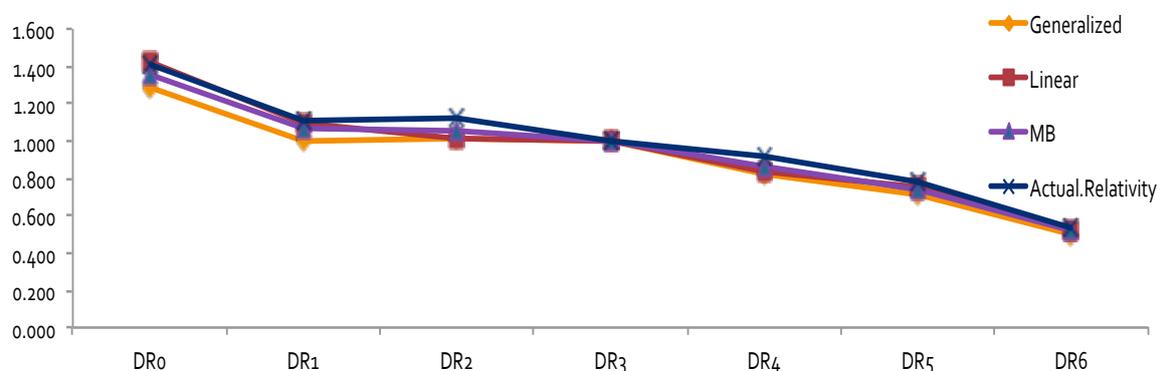


Figure 3. Results of relativities for DR obtained from GLM under Gaussian family.

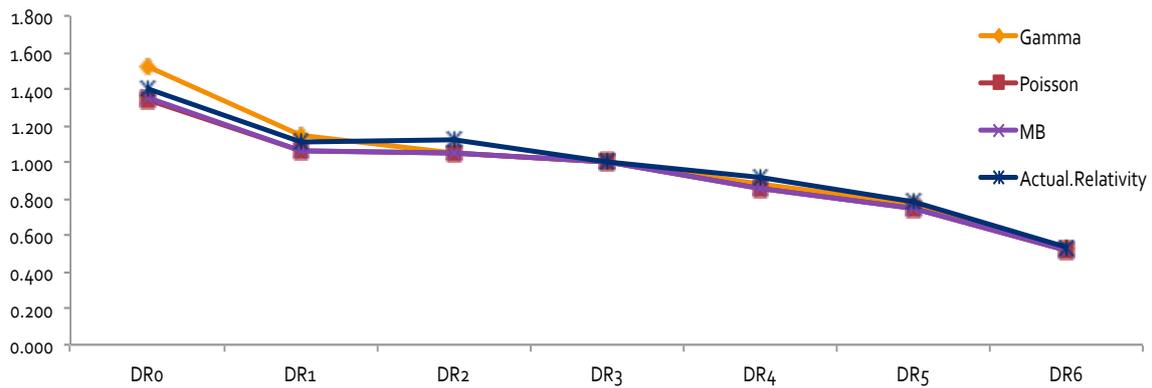


Figure 4. Results of relativities for DR obtained from GLM under gamma and Poisson models.

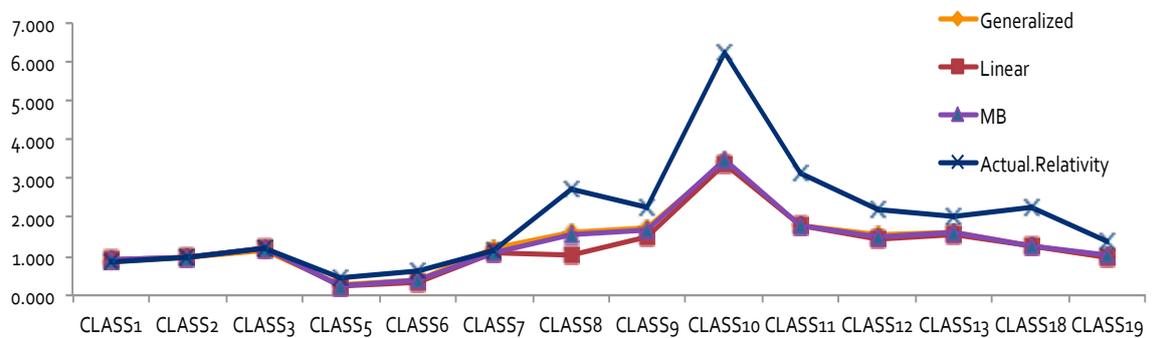


Figure 5. Results of relativities for Class obtained from GLM under Gaussian family.

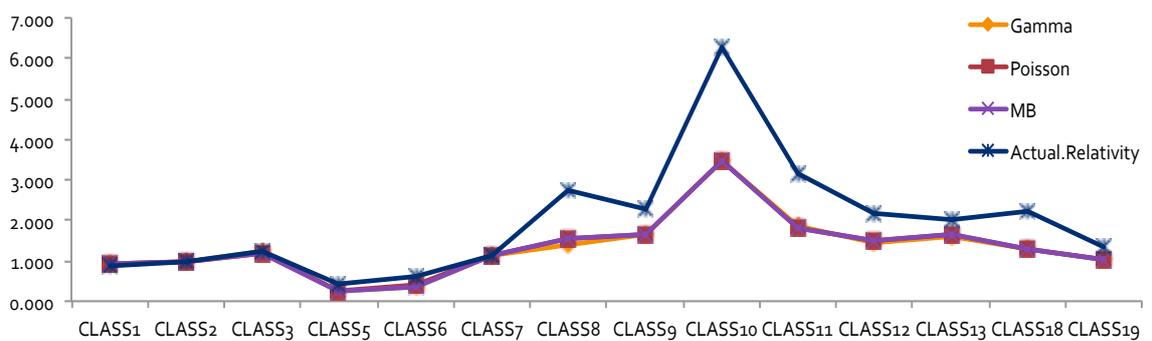


Figure 6. Results of relativities for Class obtained from GLM under gamma and Poisson models.

The values of the coefficients and relativities obtained by the GLM for the data with three accident years combined are reported in Figures 8 and 9. Recall that for the one-way relativities we observed the reversal of the values for DR1 and DR2 in both the 2007 and 2009 accident years (see Figures 1 and 2). However, after combining three years' data, this reversal of the values disappeared in one-way relativities. This implies that the use of the data of three accident years combined was better in terms of discovering the desired pattern. The results obtained by GLM for both cases with and without Year as a risk factor did not display a reversal phenomenon. Also, with the inclusion of Year as a risk factor, we could capture its effect on loss cost. From the results of Figures 8 and 9 we see that this effect was significant among different years.

From the results displayed in Figures 8 and 9, we can conclude that the model with *Year* as a risk factor was a good choice from the rate filings review perspective, as it could avoid extra uncertainty from single-year data. The results from the data of three years combined were more reliable and more suitable for acting as a benchmark for the rate filings review. The model can be easily extended to further include the territory as additional risk factor. However, we feel that from the rate filing review perspective it is not necessary, because it will significantly decrease the number of exposures for each combination of given levels. Also, we found that the definition of territory presents another difficulty, as it may change from time to time. Definitions of DR and Class are relatively more stable.

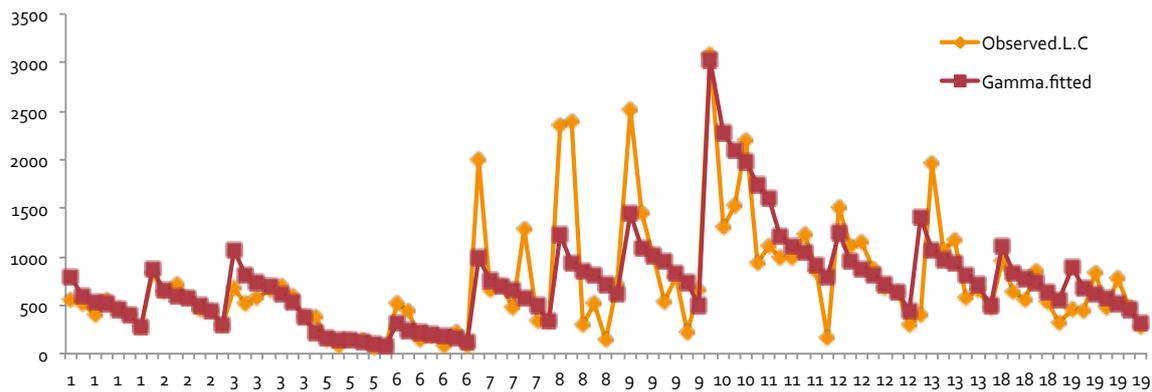


Figure 7. The predicted loss costs associated with different levels of Class under GLM with log link function and gamma error distribution function.

TPL-Urban 2007-2009	DR <sub>3</sub> and CLASS <sub>2</sub> AS A BASE			MBP	One-Way.Reliativity
	Gamma	Poisson	Inverse.Gauss		
(Intercept)	718	677	790		
DR <sub>0</sub>	1.279	1.231	1.378	1.231	1.267
DR <sub>1</sub>	1.074	1.037	1.134	1.037	1.096
DR <sub>2</sub>	0.950	0.962	0.923	0.962	1.048
DR <sub>3</sub>	1.000	1.000	1.000	1.000	1.000
DR <sub>4</sub>	0.897	0.862	0.956	0.862	0.956
DR <sub>5</sub>	0.779	0.760	0.821	0.760	0.842
DR <sub>6</sub>	0.525	0.513	0.555	0.513	0.562
AVG of DR				0.586	
CLASS <sub>1</sub>	0.930	0.927	0.931	0.927	0.889
CLASS <sub>2</sub>	1.000	1.000	1.000	1.000	1.000
CLASS <sub>3</sub>	1.231	1.214	1.238	1.214	1.221
CLASS <sub>5</sub>	0.292	0.283	0.312	0.283	0.467
CLASS <sub>6</sub>	0.338	0.334	0.345	0.334	0.550
CLASS <sub>7</sub>	1.167	1.173	1.165	1.173	1.158
CLASS <sub>8</sub>	1.468	1.525	1.433	1.525	2.658
CLASS <sub>9</sub>	1.786	1.741	1.854	1.741	2.289
CLASS <sub>10</sub>	2.740	2.701	2.883	2.701	4.860
CLASS <sub>11</sub>	1.738	1.727	1.803	1.727	2.969
CLASS <sub>12</sub>	1.349	1.391	1.322	1.392	2.021
CLASS <sub>13</sub>	1.435	1.434	1.445	1.435	1.755
CLASS <sub>18</sub>	1.260	1.255	1.302	1.255	2.201
CLASS <sub>19</sub>	1.126	1.127	1.130	1.127	1.490
AVG of CLASS				0.987	

Figure 8. Relativities obtained under GLM with log link function and various error functions including gamma, Poisson, and inverse Gaussian. The combined data of three accident years' loss cost were used.

Relativity Estimate using Three Year Data + Year as a Variable								
	TPL+Ubran							
	Generized Gaussian	Gamma	Poisson	Linear Gaussian	2007	2008	2009	
Base	676.5770	665.7869	665.9685	634.4699				
YEAR2007	1.0000	1.0000	1.0000	1.0000				
YEAR2008	0.8641	0.8917	0.8855	0.8910				
YEAR2009	0.7473	0.7819	0.7700	0.7804				
					One -way relativities			Avg.
DR0	1.2332	1.2744	1.2359	1.1962	1.4228	1.0081	1.4458	1.2922
DR1	1.0244	1.0755	1.0434	1.0389	1.1261	1.0946	1.0849	1.1019
DR2	0.9771	0.9517	0.9643	0.9211	1.1368	0.8533	1.1918	1.0606
DR3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
DR4	0.8340	0.9047	0.8628	0.8629	0.9269	0.8590	1.1467	0.9775
DR5	0.7529	0.7852	0.7638	0.7677	0.7916	0.8561	0.9105	0.8527
DR6	0.5146	0.5331	0.5190	0.5253	0.5424	0.5262	0.6618	0.5768
CLASS1	0.9165	0.9301	0.9262	0.9287	0.85762	0.93209	0.88218	0.8906
CLASS2	1.0000	1.0000	1.0000	1.0000	1.00000	1.00000	1.00000	1.0000
CLASS3	1.1893	1.2312	1.2171	1.2233	1.21875	1.31647	1.13487	1.2234
CLASS5	0.2800	0.3010	0.2859	0.2715	0.40807	0.49945	0.50632	0.4713
CLASS6	0.3392	0.3386	0.3373	0.3209	0.60465	0.55313	0.47658	0.5448
CLASS7	1.1818	1.1684	1.1708	1.1498	1.13667	1.14339	1.19736	1.1591
CLASS8	1.6123	1.5116	1.5556	1.1960	2.75350	2.64238	2.53476	2.6435
CLASS9	1.7069	1.8077	1.7381	1.5733	2.26476	1.91548	2.73710	2.3058
CLASS10	2.7508	2.7392	2.7205	2.3935	6.25818	3.16067	4.89364	4.7708
CLASS11	1.7626	1.7337	1.7412	1.5847	3.13759	3.54148	2.12313	2.9341
CLASS12	1.4525	1.3448	1.3948	1.3002	2.19119	2.05428	1.71716	1.9875
CLASS13	1.4533	1.4277	1.4356	1.3988	2.02595	1.54043	1.63193	1.7328
CLASS18	1.2626	1.2753	1.2641	1.2308	2.23503	2.04992	2.31546	2.2001
CLASS19	1.1188	1.1369	1.1298	1.1045	1.35728	1.59652	1.53518	1.4963

**Figure 9.** Relativities obtained under GLM with log link function and various error functions including gamma, Poisson, and inverse Gaussian. Three accident years’ loss cost data were used, and the model included *Year* as a rating variable. DR and Class relativities obtained from MBP and the one-way analysis are also reported.

#### 4. Conclusions

The GLM procedure with Poisson error distribution function is equivalent to MBP in estimating relativities. However, due to the different methods of estimating base rate, the GLM with Poisson error function outperformed MBP in terms of the overall bias. For the data considered, the gamma error function was the most appropriate error distribution. For the relativity estimation under multiple-year loss experiences, the gamma error distribution function was the only loss function that led to a monotonic decrease with the increase of driving experience. Because a monotonic decrease is desired for DR in auto insurance pricing, this suggests that the gamma error distribution function should be used. When GLM is applied to aggregate data, a weight function is needed if one believes that there is some bias from estimating average loss cost for each combination of risk factors, or if the level of data variation at each cell needs to be considered. The commonly used weight function is the number of exposures for each combination of the levels of the risk factors. However, GLM is able to reduce the effect of potential interaction of risk factors. Because of this, the obtained relativities for major risk factors are usually lower than those from the one-way analysis. GLM can be expanded by including additional rating variables, conducting a statistical test of significance, and evaluating the predictive power of the model. Since GLM is an efficient and reliable predictive tool, it is popular in other areas of predictive modeling.

The overall implication of the presented findings is that the pricing issues in a complex insurance system from the rate filings perspective is better understood. The estimates of major risk factors captured major pricing patterns. The results explain the natural variation in the process when reviewing a company’s rate change application. This enhances decision-making in operations using only major risk factors. because it is less affected by the details of the company’s results. Understanding how advanced statistical techniques work may also lead to a better communication with insurance companies in the process of discussing rate changes. Our study demonstrated that GLM is suitable as a predictive modeling technique in auto insurance rate filings review. The GLM relativities of major risk factors within rate filings review can be used as the benchmark for the same risk factors used in auto insurance pricing.

**Author Contributions:** Conceptualization, S.X.; Data curation, S.X.; Formal analysis, S.X.; Investigation, S.X. and A.L.; Methodology, S.X.; Resources, A.L.; Software, S.X.; Validation, S.X.; Visualization, S.X.; Writing – original draft, S.X.; Writing – review & editing, S.X. and A.L.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Bailey, Robert A. 1963. *Insurance Rates With Minimum Bias*. Arlington County: Casualty Actuarial Society, vol. 50, pp. 4–11.
- Bradley, Brendan O., and Murad S. Taqqu. 2003. Financial risk and heavy tails. In *Handbook of Heavy-Tailed Distributions in Finance*. Edited by ST Rachev. Amsterdam: Elsevier, pp. 35–103.
- Brown, Robert L. 1988. *Minimum Bias with Generalized Linear Models*. Arlington County: Casualty Actuarial Society, vol. 75, pp. 187–217.
- Chen, Shu-Ching, Min Chen, Na Zhao, Shahid Hamid, Kasturi Chatterjee, and Michael Armella. 2009. Florida public hurricane loss model: Research in multi-disciplinary system integration assisting government policy making. *Government Information Quarterly* 26: 285–94. [[CrossRef](#)]
- Cripe, Frederick F., and Stephen Fiete. 2012. Usage-Based Insurance Cost Determination System and Method. U.S. Patent 8,280,752, October 2.
- David, Mihaela, and Dănuț-Vasile Jemna. 2015. Modeling the frequency of auto insurance claims by means of poisson and negative binomial models. *Annals of the Alexandru Ioan Cuza University-Economics* 62: 151–68. [[CrossRef](#)]
- David, Mihaela. 2015. Auto insurance premium calculation using generalized linear models. *Procedia Economics and Finance* 20: 147–56. [[CrossRef](#)]
- De Jong, Piet, and Gillian Z. Heller. 2008. *Generalized Linear Models for Insurance Data*. Cambridge: Cambridge Books.
- Denuit, Michel, and Stefan Lang. 2004. Non-life rate-making with Bayesian GAMs. *Insurance: Mathematics and Economics* 35: 627–47. [[CrossRef](#)]
- Duan, Zhengmin, Yonglian Chang, Qi Wang, Tianyao Chen, and Qing Zhao. 2018. A Logistic Regression Based Auto Insurance Rate-Making Model Designed for the Insurance Rate Reform. *International Journal of Financial Studies* 6: 18. [[CrossRef](#)]
- Dunn, Peter K., and Gordon K. Smyth. 2001. Tweedie family densities: Methods of evaluation. Paper presented at 16th International Workshop on Statistical Modelling, Odense, Denmark, July 2–6.
- Feldblum, Sholom, and J. ERIC Brosius. 2003. *The Minimum Bias Procedure: A Practitioner'S Guide*. Arlington County: Casualty Actuarial Society, vol. 90, pp. 196–273.
- Frees, Edward W., and Gee Lee. 2015. Rating endorsements using generalized linear models. *Variance* 10: 51–74.
- Frees, Edward W. 2014. Frequency and severity models. *Predictive Modeling Applications in Actuarial Science* 1: 138–64.
- Fuzi, Mohd Fadzli Mohd, Abdul Aziz Jemain, and Noriszura Ismail. 2016. Bayesian quantile regression model for claim count data. *Insurance: Mathematics and Economics* 66: 124–37. [[CrossRef](#)]
- Garrido, José, Christian Genest, and Juliana Schulz. 2016. Generalized linear models for dependent frequency and severity of insurance claims. *Insurance: Mathematics and Economics* 70: 205–15. [[CrossRef](#)]
- Haberman, Steven, and Arthur E. Renshaw. 1996. Generalized linear models and actuarial science. *The Statistician* 45: 407–36. [[CrossRef](#)]
- Hamid, Shahid S., Jean-Paul Pinelli, Shu-Ching Chen, and Kurt Gurley. 2011. Catastrophe model-based assessment of hurricane risk and estimates of potential insured losses for the state of Florida. *Natural Hazards Review* 12: 171–76. [[CrossRef](#)]
- Harrington, Scott E. 1992. A Note on the Impact of Auto Insurance Rate Regulation. In *Foundations of Insurance Economics*. Dordrecht: Springer, pp. 724–28.
- Husnjak, Siniša, Dragan Peraković, Ivan Forenbacher, and Marijan Mumdziev. 2015. Telematics system in usage based motor insurance. *Procedia Engineering* 100: 816–25. [[CrossRef](#)]
- Ismail, Noriszura, and Abdul Aziz Jemain. 2009. *Comparison of Minimum Bias and Maximum Likelihood Methods for Claim Severity*. Arlington County: Casualty Actuarial Society, p. 243.

- Klein, Nadja, Michel Denuit, Stefan Lang, and Thomas Kneib. 2014. Nonlife ratemaking and risk management with Bayesian generalized additive models for location, scale, and shape. *Insurance: Mathematics and Economics* 55: 225–49. [CrossRef]
- Lourens, Peter F., Jan AMM Vissers, and Maaïke Jessurun. 1999. Annual mileage, driving violations, and accident involvement in relation to drivers' sex, age, and level of education. *Accident Analysis & Prevention* 31: 593–97.
- Ma, Yu-Luen, Xiaoyu Zhu, Xianbiao Hu, and Yi-Chang Chiu. 2018. The use of context-sensitive insurance telematics data in auto insurance rate making. *Transportation Research Part A: Policy and Practice* 113: 243–58. [CrossRef]
- McCullagh, Peter, and John A. Nelder. 1989. *Generalized Linear Models*. Boca Raton: CRC Press, vol. 37.
- McNeil, Alexander J. 1997. Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin: The Journal of the IAA* 27: 117–37. [CrossRef]
- Meyers, Glenn G. 2007. Estimating predictive distributions for loss reserve models. *Variance* 1: 248–72.
- Mildenhall, Stephen J. 1999. *A Systematic Relationship Between Minimum Bias and Generalized Linear Models*. Arlington County: Casualty Actuarial Society, vol. 86, pp. 393–487.
- Najafabadi, Amir T. Payandeh, and Saeed Mohammad Pour. 2017. A k-Inflated Negative Binomial Mixture Regression Model: Application to Rate-Making Systems. *Asia-Pacific Journal of Risk and Insurance* 12. [CrossRef]
- Ohlsson, Esbjörn, and Björn Johansson. 2010. *Non-Life Insurance Pricing With Generalized Linear Models*. Berlin: Springer, vol. 2.
- Osborne, Jason W. 2010. Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research & Evaluation* 15: 2.
- Outreville, J. Francois. 1990. Underwriting Cycles And Rate Regulation In Automobile Insurance Markets. *Journal of Insurance Regulation* 9: 274.
- Peck, Raymond C., and Jensen Kuan. 1983. A statistical model of individual accident risk prediction using driver record, territory and other biographical factors. *Accident Analysis & Prevention* 15: 371–93.
- Shi, Peng, and Emiliano A. Valdez. 2014. Multivariate negative binomial models for insurance claim counts. *Insurance: Mathematics and Economics* 55: 18–29. [CrossRef]
- Shi, Peng, and James Guszczka. 2016. Frameworks for General Insurance Ratemaking. In *Predictive Modeling Applications in Actuarial Science*. Case Studies in Insurance. Cambridge: Cambridge University Press, Vol. 2, p. 100.
- Tang, Yanlin, Liya Xiang, and Zhongyi Zhu. 2014. Risk factor selection in rate making: EM adaptive LASSO for zero—Inflated poisson regression models. *Risk Analysis* 34: 1112–27. [CrossRef] [PubMed]
- Tennyson, Sharon. 2009. The Relationship Between Auto Insurance Rate Regulation and Insured Loss Costs: An Empirical Analysis. Available online: <https://ecommons.cornell.edu/handle/1813/15118> (accessed on 11 October 2018).
- Tweedie, Maurice C. K. 1957. Statistical properties of inverse Gaussian distributions. II. *The Annals of Mathematical Statistics* 28: 696–705. [CrossRef]
- Watson, Charles C., Mark E. Johnson, and Martin Simons. 2004. Insurance rate filings and hurricane loss estimation models. *Journal of Insurance Regulation* 22: 39–64.
- Zhao, Hui-qing, and Han-zhang Wang. 2011. An Empirical Study of Rate Making of Automobile Insurance in China—Analysis Based on Generalized Linear Models. *Journal of Tianjin University of Commerce* 5: 8–12.

