PREMIUM PRICING AND RISK ASSESSMENT FOR CLAIM AMOUNTS BASED ON GENERALIZED LINEAR MODELS (GLM)

Arş. Grv. Dr. Latife Sinem SARUL
İstanbul Üniversitesi
lasinem@istanbul.edu.tr

Prof. Dr. Mehmet Erdal BALABAN
İstanbul Üniversitesi
balaban@istanbul.edu.tr

Abstract

Actuarial Science is described as a mechanism that decreases the negative financial effects of random events which becomes obstacles to actualize reasonable expectations. It is important subject to make a fair share for the same amount of money which is paid by the people who has the same risk. It becomes even more important to be able to provide more effective methods with the reasonable prices on the customer retention and customer relationship management in the mutually competitive environment. In this case, it is expected to have methods which take into account customer's previous claim experience with high predictive powers by insurance companies. Today, a large number of assumptions which may be used in the classical methods of analysis and predictions of this analysis are not sufficient. The main purpose of this study is of great importance for sustainable customer relationships, just make up a portfolio of premium pricing to be able to create a model that takes into account risk factors for individuals. GLM is a powerful methodology to evaluate the non-normal data. In this reason, it is formed an effective model that takes into account risk factors for the individuals in the portfolio using GLM. As a result of this analysis, it is chosen Logarithmic Gamma Model which gives the best results of the analysis for the customers that forms the data set. Finally, risk assessment was made by evaluating coefficient of variation, max, min and average of the claim amounts. At the end, 0.1% customers of the portfolio forms high risk group with regard to the change in the coefficient of variation.

Keywords: Statistics, Insurance, Generalized Linear Models

JEL Classification: C31, G22

GENELLEŞTİRİLMİŞ LİNEER MODELLERE (GLM) DAYANARAK HASAR MİKTARLARI İÇİN RİSK DEĞERLENDİRME VE PRİM FİYATLAMA

Aktüerya bilimi normal olarak gerçekleşmesi beklenmeyen tesadüfî olayların olumsuz yöndeki finansal etkisini azaltmak için bir mekanizma olarak tanımlanmıştır. Aynı türden tehlikeyle karşı

karşıya olan kişilerin, prim olarak adlandırılan belirli bir miktar para ödemesi şeklinde toplanan bu tutarın, adil bir şekilde belirlenmesi sigorta şirketleri için önemli bir konudur. Karşılıklı rekabet ortamında müşteri bağlılığını sağlamak ve müşteri ilişkileri yönetimi açısından bakıldığında etkili yöntemler kullanılarak adil bir prim fiyatlama yapılması daha da önem kazanmaktadır. Bu durumda sigorta şirketleri için müşterinin geçmiş hasar tecrübelerini dikkate alan yüksek tahmin gücü olan yöntemlerin kullanılması oldukça önem taşımaktadır. Günümüzde çok sayıda varsayıma dayanan klasik yöntemler tahmin ve analiz için yeterli olmamaktadır. Bu çalışmada temel amaç, adil bir prim fiyatlama yapabilmek için portföyü oluşturan bireylere ilişkin risk faktörlerini dikkate alan matematiksel ve istatistiksel temellere dayanan bir model oluşturmaktır. GLM normal dağılmayan veri setlerinin analizinde kullanılan güçlü bir metodolojidir.Bu nedenle öncelikle prim fiyatlamaya temel oluşturan modeller incelenmiş daha sonra poliçe sahiplerinin risk faktörlerini de dikkate alan etkili bir model elde etmek için Genelleştirilmiş Lineer Modeller kullanılmıştır. Yapılan analizler sonucunda en iyi sonuç veren Logaritmik Bağlı Gama Model kullanılarak hasar tahminleri yapılmış ve veri setini oluşturan müşteriler için risk değerlendirmesi yapılmıştır. Bu analiz ile değişkenlik katsayısı, maksimum, minimum ve ortalama hasar miktarlarına dayanan risk değerlendirmesi yapılmıştır. Bu değerlendirme sonucunda portföyü oluşturan müşterilerin %0,1’ lik kısmının yüksek risk grubunu oluşturduğu görülmektedir.

Anahtar Kelimeler: İstatistik, Sigortacılık, Genelleştirilmiş Lineer Modeller

Jel Kodu: C31, G22

1. Introduction

Actuarial Science is a decision making mechanism based on mathematical and statistical basis for insurance related activities and incidental events that influence the presence of people or property for life. On the other hand insurance is the risk transfer system to meet the loss of people who suffered from the actual result of the realization of the claim by collecting certain amount of money so called premium from people who face the same kind of risk. The premium can not be applied equally to all the individuals that make up portfolio consisting of heterogeneous different level of risks. Fair pricing is of great importance to be able to compete in the market for the insurance companies. Pricing (Rate Making, Rating), is expressed as calculation of the premium which is paid to provision of insurance coverage. It is also defined as the credit rating given to companies by the evaluation companies. Process of determining the credit rating is made by rating among the weakest and the most powerful levels (Çuhacı, 2004).

Generalized Linear Models (GLM), which is used to model non-normally distributed data, is a methodology for modeling the relationship between variables. Development of the GLM began with the papers by Nelder and Wedderburn (1972). GLM has an important role to model non-normally distributed data sets. Beginning of the use of GLM in actuarial work is at early 1980s. McCullagh and Nelder (1989) have shown GLM's applicability to the different data sets. Haberman and Renshaw (1996) reviewed the applications of generalized linear models to actuarial problems. Nelder and Verral (1997) showed the relationship between Hierarchical Generalized Linear Models and Credibility Theory which is another useful tool for ratemaking. However Credibility Theory is out of the scope of this paper. Nelder and Verral demonstrated that how credibility theory can be included in the frame of Hierarchical Generalized Linear Models. GLM is more reasonable for pricing in which some monotone transformation of the mean is a linear function of x's while in linear models the mean is a linear function of the covariates x (Ohlsson, Johansson, 2000).

2. Generalized Linear Models

GLM theory is based on the family of exponential distribution. Exponential family puts the similar functions which are in the different mathematical form into a single re-characterized form as a more useful theoretical structure.

Exponential family is expressed in the form:

$$f(y;\theta,\phi) = exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\}$$

Here, y is the dependent variable, θ is the interest parameter or canonical parameter and φ is called the scale parameter. It is obtained the different members of the exponential family with specifying $a(.)$, $b(.)$ and $c(.)$ functions (Jong and Heller, 2008).

Below mean and variance functions are given respectively for the exponential family (McCullagh and Nelder 1989).

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

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

Normal (Gaussian), Poisson, Binomial, Beta, Multinomial, Dirichlet, Pareto, Gamma and Inverse Gaussian distributions are members of the exponential family (Gill, 2001). Generalized Linear Models are very significant to analize insurance data. Because Insurance data consists of claim sizes, claim frequencies and occurrence of a claim, the assumptions of normal model is generally inconvenient. Gamma and inverse gaussian models become more important for modeling continuous data which is also called claim data.

GLM consists of three main components that is random(stochastic) component, systematic component and link function which links the random and systematic component. Independent Y_i, $i = 1, \dots, N$ variables, assumed to come from the same distribution family, are called as random component. Covariates x_j , $j = 1, \dots, p$ produce the systematic component of GLM with η linear predictor given by $\eta = \sum_{j=1}^{p} x_j \beta_j$. The link function provides a connection between the systematic and random component. It is indicated by $\eta = g(\mu)$. It shows linear relation between expected value of the dependent variable and η linear predictor (McCullagh and Nelder 1989). Maximum Likelihood Method based on the likelihood function is commonly used to comply with non-normally distributed data sets for parameter estimation. Wald Tests and Score Tests are widely used to evaluate of the parameter significance. Goodness of fit statistics, are used to assess model fitting compares two models that best fit the data set. Likelihood Ratio Test forms the basis of goodness of fit tests are widely used in practice. Deviance, Pearson Chi-Square, McFadden R^2, Pseudo R^2 and Information Criteria are the other measurements to evaluate model fitting (Hoffmann, 2004). Deviance is a measure of distance between the saturated and fitted models. A large value of the deviance indicates a badly fitting model (Jong, Heller, 2008). Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are called Information Criteria which examine the complexity of the model. Residual, which is the difference between observed and fitted values, is an important tool used to measure the adequacy of the model. It is also used for determining for the new explanatory variables or the effects of non-linear trends in the actual covariates, identifying poorly fitting observations, evaluating the impact on the individual observations and revealing other trends such as heteroscedasticity (Frees, 2010). Pearson Residuals, Deviance and Anscombe Residuals are widely used in GLM. It is possible to examine GLM in the four major groups according to the distribution of the dependent variable, which is continuous dependent variable, integer, binomial and multinomial models. Continuous dependent variable models consist of Gamma Models, Inverse Gaussian Models and Linear Regression Models, which is a special case of GLM with normally distributed dependent variable. Models with discrete integer values of the dependent variable are the Poisson and Negative Binomial Models. Examples of the application of these models is the use of examining the effect of explanatory variables on the number of claims such as vehicle type, color, engine capacity in general and accident insurance or the examination of the number of accidents can be held in a city. Binomial and Multinomial Models include the models which dependent variable is discrete, proportional and categorical. These models are analyzed in logistic regression analysis.

Gamma model is used for situations where dependent variable is only zero and positive values. However, it is used for situations where the dependent variable is continuous; it can be applied to the discrete data sets which has a lot of integer values.

3. Research Methodology

In this study, it is intended to achieve a statistical model that is taking into account the risk factors for each insured vehicle to provide fair pricing policies for fleets using Generalized Linear Models. The model is based on the knowledge of the 20,664 policyholders in 2009 created to estimate the damage using Generalized Linear Models.

In this study analysis was performed using the R statistical software package*.

3.1 The variables:

It is attached a detailed description of the variables in Table 1 for used data. "ClAmnt" variable shown in Table 1 used as the dependent variable in the analyzes.

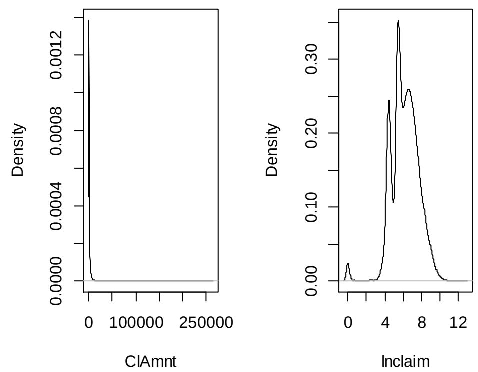
Table1: Explanations Related the Analysis of Variables

Variable Code	Description
ClAmnt	Claim Amount
VehUse	Vehicle Usage
MdlYr	Model Year
CylindVol	Cylinder Volume
MotPow	Motor Power
FlType	Fuel Type
VehType	Vehicle Type
CylindNum	Cylinder Number
VehWeight	Vehicle Weight
AxlsWeight	Axles Weight
VehLth	Vehicle Length

* R is a free software

Dependent variable "ClAmnt" refers to claim amount for each policyholder. Figure1 displays distribution of Claim Amount. Distribution of the Claim Amount appears in left panel and distribution of the log claim amount appears in right panel. It is observed that distribution of the logarithm of the claim amount is approximately normal.

Figure1: Distribution Graph of Claim Amount



Vehicle Type: There are 19 different vehicle type in data set. These are CAB(0,1%), CKK(0,4%), CKP(0,1%), CKU(2,4%), CPE(0,1%), CVA(4,6%), DSA(0,3%), EYP(0,3%), HBA(18,3%), KAM(1%), MIN(1,9%), PAN(20,9%), PCK(0,1%), ROA(0,0%), SED(38,3%), STV(9,6%), YCA(0,8%), YPV(0,8%) and OTHER(0,0%). SED type vehicles form 38% of all data set. However, STW type vehicles have the greatest claim amount.

Vehicle Usage: There are 9 different vehicle usage in data set. These are BUS(1,1%), CAB(0,4%), FUNERAL CAR(0,0%), MINIBUS(1,7%), PRIVATE CAR(34,1%), RENTED CAR(31,4%), RESQUER(0,2%), SMALL TRUCK (30,8%) and TRANSPORT VEHICLE(0,3%). PRIVATE CARs form approximately 34% of all data set and has the highest claim amount.

Model Year: Vehicles in the data set before and after the year 2007 have been categorized into two groups. Accordingly, for the year 2007 and later vehicles constitute nearly 69% of the data set and has the greatest claim amount.

Fuel Type: Vehicles in the data set have been categorized into two groups according to the fuel type. Accordingly, nearly 82% of vehicles in all data set are used diesel fuel.

Cylinder Volume: Vehicles are examined in three categories according to the volumes of cylinders, which are the range from 0 to 1500 m^3, from 1500 to 2500 m^3 and from 2500 to 6000

m^3. Accordingly, 59% of vehicles in the data set are between 0 to 1500 m^3. However, the vehicles which cylinder volume is in the range of 2500-6000 m^3 have the highest claim amount.

Motor Power: Vehicles are examined in four categories according to the engine powers, which are the range from 0 to 100, from 100 to 200, from 200 to 300 and from 300 to 400 horsepower. Accordingly, nearly 63% of vehicles in the data set have horsepower between 0 and 100.

Cylinder Number: Vehicles are evaluated in three categories according to the numbers of cylinders, which are the range from 0 to 4, from 5 to 7 and from 8 to 12. Approximately 96% of vehicles have “0-4” cylinder. However, the vehicles which the number of cylinder is in the range of 5-7 have the highest claim amount.

Vehicle Weight: Vehicles are examined in four categories according to the vehicle weight, which are the range from 0 to 1000, from 1000 to 2000, from 2000 to 3000 and from 3000 to 5000. Nearly 70% of vehicles are in the range of 1000-2000. However, the vehicles which the vehicle weight is in the range of 2000-3000 have the highest claim amount.

Axles Weight: Vehicles are evaluated in three categories according to the axles weight, which are the range from 0 to 2500, from 2500 to 3500 and from 3500 to 5000. Approximately 68% of vehicles are in the range of 2500-3500 and have the highest claim amount.

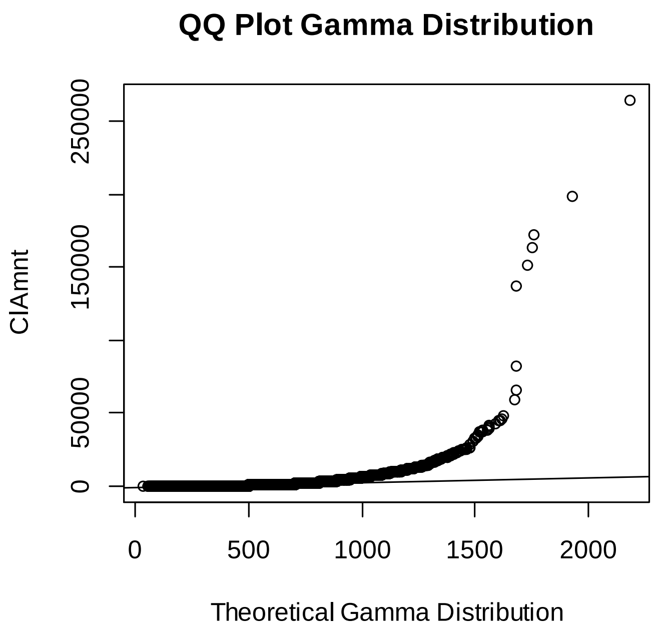
Vehicle Lenght: Vehicles are examined in three categories according to the vehicle lenght, which are the range from 0 to 4500, from 4500 to 5500 and from 5500 to 7500. Approximately 71% of vehicles are in the range of 0-4500. However, the vehicles which the number of cylinder is in the range of 4500-5500 have the highest claim amount.

4. Computer Results and Discussions

Non-normally distributed data sets do not provide the assumptions of normal linear models. GLM provide an important extention for normal linear models to be able to model non-normally distributed data sets. Figure 1 shows the claim amounts concentrate on positive values close to zero. This indicates that dependent variable claim amount is eligible to gamma distribution.

Figure 2 displays QQ Plot Gamma distribution

Figure2: Q-Q Plot Gamma distribution



First model includes all variables in data set, which are dependent variable claim amount and covariates Vehicle Usage, Vehicle Type, Model Year, Fuel Type, Cylinder Volume, Cylinder Number, Vehicle Weight, Axles Weight, Vehicle Lenght. As a result of analysis, Model Year and Vehicle Lenght affect claim amount at 1% significance level. Vehicle Type, Motor Power and Axles Weight affect claim amount at 5% significance level. AIC value which is used to compare different models is 333,504 and another indicator of the goodness of fit is the deviation appears 39,427 for Model1. In order to improve Model1, it is formed another model which is called Model2 with the covariates in different combinations. As a result of this analysis, Vehicle Type, Model Year and Axles Weight affect claim amount at 5%, Motor Power and Vehicle Weight affect claim amount at 1% significance level. AIC value is 333,659 and deviance is 39,704 for Model2. Another approach is that categorical variables which has the great number of levels is considered as a continuous variable. Vehicle Usage with 9 categories and Vehicle Type with 19 categories were considered as continuous independent variables in Model3. AIC value is 333,829 and deviation is 40,000 for Model3. Figure1 demonstrate that dependent variable is skewed to the right, which refers to Gamma and also Inverse Gaussian distribution. As an alternative, it has been tested Inverse Gaussian Model so called Model4. Link function is chosen as $\frac{1}{\mu^2}$. AIC value is 353,314 and deviance is 287,13.

It is given comparison table in Figure3 which represents four models comparatively. The values significant at 5% are shown in bold at the table. Figure3 is attached below.

Mathematical expression of Model1 used for predictions as follows;

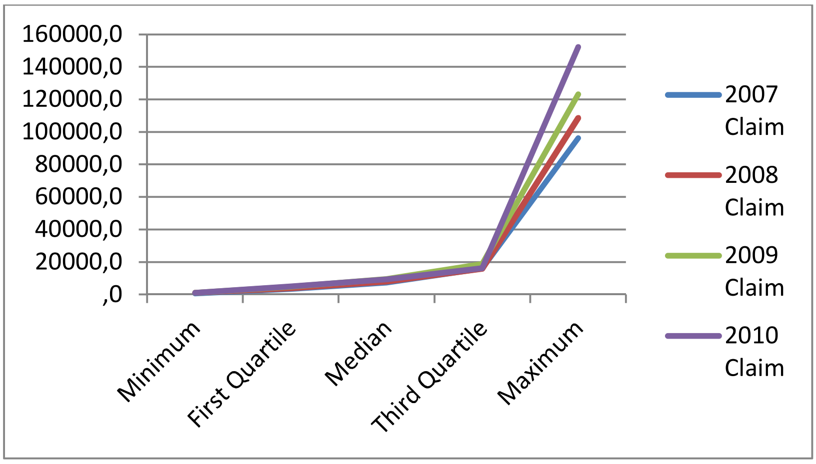
$$y \sim G(\mu, \nu)$$

$$\begin{aligned} ln\mu = {} & 8{,}5584 - 1{,}8407x_{VT1} - 1{,}1988x_{VT2} - 1{,}3240x_{VT3} - 1{,}5565x_{VT4} - 3{,}0376x_{VT5} \\ & - 1{,}3640x_{VT6} - 1{,}9534x_{VT7} - 1{,}4484x_{VT8} - 1{,}6507x_{VT9} - 1{,}8365x_{VT10} \\ & - 1{,}6057x_{VT11} - 1{,}9637x_{VT12} - 1{,}6820x_{VT13} - 1{,}4955x_{VT14} - 1{,}5032x_{VT15} \\ & - 1{,}4992x_{VT16} - 1{,}6743x_{VT17} + 0{,}0956x_{MY1} + 0{,}2056x_{MG1} + 1{,}1274x_{MG2} \\ & + 0{,}62x_{MG3} - 0{,}1086x_{DA1} + 0{,}2256x_{AU1} \end{aligned}$$

	Model1		Model2		Model3		Model4	
	Gamma Distr. Link Fonk: Log		Gamma Distr. Link Fonk: Log		Gamma Distr. Link Fonk: Log		Inv. Gauss. Distr. Link Fonk: $1/\mu^2$	
	Est.	t val.	Est.	t val.	Est.	t val.	Est.	t val.
(Intercept)	8,5584	16,28	8,5168	18,10	7,2131	39,62	0,0000	3,95
VehUseCAB	0,1229	0,43						
VehUseFUN.CAR	-1,6362	-1,2						
VehUseMINIBUS	-0,3013	-1,69						
VehUsePRVATECAR	-0,1088	-0,56						
VehUseRENTEDCAR	-0,3497	-1,82						
VehUseRESQUER	-0,1560	-0,41						
VehUseSMLLTRUCK	-0,2153	-1,16						
VehUseTRNS.VHICL	0,2623	0,84						
VehTypeCKK	-1,8407	-3,59	-1,7579	-3,42				
VehTypeCKP	-0,8461	-1,36	-0,7365	-1,18				
VehTypeCKU	-1,1988	-2,51	-1,0743	-2,25				
VehTypeCPE	-1,3240	-2,26	-1,2664	-2,13				
VehTypeCVA	-1,5565	-3,29	-1,5647	-3,30				
VehTypeDIGER	-3,0376	-2,13	-3,0476	-2,11				
VehTypeDSA	-1,3640	-2,54	-1,3006	-2,41				
VehTypeEYP	-1,9534	-3,64	-1,9177	-3,56				
VehTypeHBA	-1,4484	-3,12	-1,5156	-3,22				
VehTypeKAM	-1,6507	-3,38	-1,5152	-3,11				
VehTypeMIN	-1,8365	-3,69	-1,7180	-3,56				
VehTypePAN	-1,6057	-3,43	-1,6088	-3,42				
VehTypePCK	-1,9637	-3,32	-1,9256	-3,24				
VehTypeROA	-1,6820	-2,06	-1,7320	-2,09				
VehTypeSED	-1,4955	-3,22	-1,4425	-3,07				
VehTypeSTW	-1,5032	-3,23	-1,4634	-3,10				
VehTypeYCA	-1,4992	-3,05	-1,4904	-3,02				
VehTypeYPV	-1,6743	-3,38	-1,5218	-3,06				
MdlYr>=2007	0,0956	3,07	0,0824	2,76	0,0829	2,66		
CylindVol1500-2500	-0,0813	-1,8			0,0874	2,29		
CylindVol2500-6000	0,2001	1,59			0,6832	6,06		
MotPow100-200	0,2056	4,49	0,2427	6,77			0,0000	-10,25
MotPow200-300	1,1274	7,46	1,5901	15,75			0,0000	-25,19
MotPow300-500	0,6276	2,35	1,0632	4,42			0,0000	-11,33
FlTypeOIL	-0,0485	-1,21			0,0071	0,18		
VehWeight1000-2000	0,1982	1,31			0,2875	1,85	0,0000	-1,39
VehWeight2000-3000	0,2245	1,39			0,3393	2,11	0,0000	-1,79
VehWeight3000-5000	0,0420	0,24			0,0987	0,56	0,0000	-1,31
AxlsWeight2500-3500	-0,1086	-2,80	-0,0762	-2,05	-0,0952	-2,57	0,0000	4,19
AxlsWeight3500-4500	0,0590	0,22	-0,0449	-0,49	0,5608	2,04	0,0000	2,75
VehLth4500-5500	0,2256	5,07			0,2994	7,17		
VehLth5500-7500	0,1230	0,43			-0,5236	-1,89		
CylindNum5-7					0,3078	3,05		
CylindNum8-12					0,6805	3,02		
VehUse1					-0,0536	-4,63		
VehType1					-0,0122	-2,67		
AIC	**333.504**		**333.659**		**333.829**		**353.314**	
Deviance	**39.427**		**39.704**		**40.000**		**287,13**	

The estimation results obtained using Model 1 for the year 2009 includes each policy in data set. These estimates are evaluated in customer base with the pivot analysis. Figure4 represents a section of this analysis. There is prediction of sum of claim amounts for 2010 based on the data obtained the analysis of 2009 claim amounts using the logarithmic model. In this table, it is calculated standard deviation, variation of coefficient, minimum, maximum and average values for 2007, 2008, 2009 and 2010. Risks are grouped in three categories which are low-risk (1), moderate risk (2) and high risk (3). In this four years period, minimum of the claim amounts is called low-risk, maximum of the claim amounts is called high risk(3) and other values which are between minimum and maximum are called moderate risk(2). In addition, "Risk Change" column shows if there is an increase or decrease in claim amounts from 2009 to 2010. "Risk Change%" column shows the percent change in the transition from 2009 to 2010. Minimum, maximum, median, first quartile and third quartile values of claim amounts appear in Figure5 for four years.

Figure3: Claim Assessment Chart for 2007-2010 years



Claim Amounts appear for 2007, 2008, 2009 and 2010 respectively in Figure5. Claim Amounts did not show significant difference for smaller values than 20,000 while large claim amounts over 20,000 has been increasing passing by the year 2010 from 2007. This is the cause of large claim amounts over the 20,000 result from the realization of claim amounts due to accidents resulting in death.

Table3: Section of the Analysis of Claim Amounts

Claim Amounts				Standard Deviation	Coefficient of Variation	Minimum Claim Amounts	Average Claim Amounts	Maximum Claim Amounts	Claim Groups by years				Total Claim Amounts	Risk Variation	Risk Variation %
2007	2008	2009	2010						2007	2008	2009	2010			
96.294	3.384	1.675	5.035	40.257,1	1,51	1.675,0	26.597,1	96.294	3	2	1	2	101.353	1	2,01
37.326	1.056	2.767	2.648	15.243,5	1,39	1.056,0	10.949,3	37.326	3	1	2	2	41.149	0	-0,04
6.187	45.295	1.305	1.277	18.456,2	1,37	1.277,1	13.516,0	45.295	2	3	2	1	52.787	-1	-0,02
2.039	7.645	45.964	1.789	18.396,7	1,28	1.788,6	14.359,2	45.964	2	2	3	1	55.648	-2	-0,96
1.530	1.860	27.708	3.386	11.042,1	1,28	1.530,0	8.621,0	27.708	1	2	3	2	31.098	-1	-0,88
2.203	1.239	59.980	12.289	24.093,4	1,27	1.239,0	18.927,8	59.980	2	1	3	2	63.422	-1	-0,80
17.312	1.604	1.812	1.269	6.822,8	1,24	1.269,2	5.499,3	17.312	3	2	2	1	20.728	-1	-0,30
1.658	28.871	5.082	1.315	11.434,2	1,24	1.314,7	9.231,4	28.871	2	3	2	1	35.611	-1	-0,74
12.871	3.069	73.716	5.377	29.069,9	1,22	3.069,0	23.758,2	73.716	2	1	3	2	89.656	-1	-0,93
54.537	11.709	2.899	3.220	21.336,4	1,18	2.899,0	18.091,3	54.537	3	2	1	2	69.145	1	0,11
5.650	1.239	56.089	13.805	21.772,2	1,13	1.239,0	19.195,7	56.089	2	1	3	2	62.978	-1	-0,75
1.099	2.482	48.825	17.962	19.209,2	1,09	1.099,0	17.592,0	48.825	1	2	3	2	52.406	-1	-0,63
31.931	2.501	5.924	4.238	12.059,7	1,08	2.501,0	11.148,5	31.931	3	1	2	2	40.356	0	-0,28
4.818	1.601	25.289	3.698	9.560,2	1,08	1.601,0	8.851,4	25.289	2	1	3	2	31.708	-1	-0,85
56.493	3.827	3.712	16.324	21.633,8	1,08	3.712,0	20.089,1	56.493	3	2	1	2	64.032	1	3,40
2.611	9.254	31.739	1.592	12.162,2	1,08	1.592,3	11.299,1	31.739	2	2	3	1	43.604	-2	-0,95
64.673	4.528	8.216	13.405	24.434,2	1,08	4.528,0	22.705,6	64.673	3	1	2	2	77.417	0	0,63

Table4: Risk Assessment by Coefficient of Variation

Average	Coefficient of Variation	2009 Total Claim Amounts	2010 Total Claim Amounts	2010 Claim Percentages	Number of Customer for 2010 Claims
Low	0-0.5	5.626.894	5.573.156	48,5%	370
Medium	0.5-1	5.088.441	4.583.155	48,0%	366
Medium High	1-1.5	526.274	160.822	3,4%	26
High	>1.5	1.675	5.035	0,1%	1
General Total		11.243.284	10.322.168	100,0%	763

Table 4 shows the variation in the claim amounts according to the coefficients of variation. Accordingly, the coefficient of variation is between 0 and 0.5 for 48.5% of claim amounts and the coefficient of variation is between 0.5 and 1 for 48% of claim amounts in 2010. For this reason, 96.5% of customers in the portfolio assessed as low and moderate risk. On the other hand, 3.4% of customer's variability ranged from 1 to 1.5 while 0.1% of customers variability is over the 1.5. For this reason, customers are falling in this range should be carefully considered by the insurance company.

Table5: Risk Assessment Table

Coefficient of Variation	Variety					Total
	-2	-1	0	1	2	
0-0,5	2,9%	15,6%	9,2%	18,3%	2,5%	48,5%
0,5-1	1,0%	19,4%	8,9%	17,4%	1,2%	48,0%
1-1,5	0,3%	1,7%	0,7%	0,8%	0,0%	3,4%
1,5-2	0,0%	0,0%	0,0%	0,1%	0,0%	0,1%
Total	4,2%	36,7%	18,7%	36,7%	3,7%	

Table5 demonstrate the distribution of customers by the changes in the coefficients of variation and claim amounts. Variety numbers shows the changing ranges according to min, max and average of the claim amounts. The customers with large variety in the coefficient of variation can accept as high risk group. While pricing insurance companies pay attention to these customers.

5. Conclusion and Comments

The main purpose of this study is of great importance for sustainable customer relationships, just make up a portfolio of premium pricing to be able to create a model that takes into account risk factors for individuals. GLM is a powerful methodology to evaluate the non-normal data. In this reason, it is formed an effective model that takes into account risk factors for the individuals in the portfolio using GLM. As a result of this analysis, it is chosen Logarithmic Gamma Model which gives the best results of the analysis for the customers that forms the data set. Finally, risk assessment was made by evaluating coefficient of variation, variety according to ranges of claim amounts, max, min and average of the claim amounts. At the end, 0.1% customers of the portfolio forms high risk group with regard to the change in the coefficient of variation. This analysis can expand in future with expanded data set using Generalized Linear Mixed Models taking into account the random and fixed effects.

REFERENCES

Annette J. Dobson, A. G. (2008). *An Introduction to Generalized Linear Models.* Chapman&Hall, New York.

Charles E. McCulloch, S. R. (2008). *Generalized Linear Mixed and Mixed Models.* Wiley, Canada.

Edward W. Frees, V. R. (1999). A Longitudinal Data Analysis Interpretation of Credibility Models . *Insurance: Mathematics and Economics*, 229-247.

Faraway, J. J. (2006). *Extending the Linear Model with R:Generalized Linear, Mixed Effects and Nonpaametric Regression Models.* Chapman&Hall, New York.

Firth, D. (1988). Multiplicative Errors: Lognormal or Gamma? *Journal of the Royal Statistical Society,* 266-268.

Frees, E. W. (2004). *Longitudinal and Panel Data Analysis and Applications in the Social Sciences* . Cambridge University Press, New York.

Frees, E. W. (2010). *Regression Modeling with Actuarial and Financial Applications* . Cambridge University Press, New York.

Gil, J. (2001). *Generalized Linear Models: A Unified Approach.* Sage Publications, New York.

Hoffmann, J. P. (2004). *Generalized Linear Models: An Applied Approach.* Pearson Education, USA.

J.A.Nelder, R. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society*, 370-384.

James W. Hardin, J. M. (2012). *Generalized Linear Models and Extentions.* Stata Press, USA.

Katrien Antonio, J. B. (2008). Actuarial Statistics with Generalized Linear Models. *Insurance: Mathematics and Economics*, 58-76.

Lindsey, J. K. (1997). *Applying Generalized Linear Models.* Springer, USA.

P. McCullagh, J. (1989). *Generalized Linear Models.* Chapmann&Hall, New York.

Piet de Jong, Z. G. (2008). *Generalized Linear Models for Insurance Data.* .Cambridge University Press, New York.

Rob Kaas, M. G. (2008). *Modern Actuarial Risk Theory Using R.* Springer , Germany.

Steven Haberman, A. E. (1996). Generalized Linear Models and Actuarial Science . *Journal of the Royal Statistical Society* .

Bu sayfa bilerek boş bırakılmıştır

This page [is] intentionally left blank