Other Liability Insurance (Mercury Insurance Group)

Jenny Liang (Wen Liang)
Email: wliang88@berkeley.edu

1. Introduction

Glenn Meyers and Peng Shi provide databases that include major personal and commercial lines of business from U.S. property casualty insurers in 2011. The datasets included six lines: (1) Private Passenger Auto Liability/Medical; (2) Commercial Auto/Truck Liability/Medical; (3) Workers' Compensation; (4) Medical Malpractice; (5) Other Liability; (6) Product Liability.

For the purpose of this project, only the data involving Mercury Insurance Group will be used. Unfortunately, the datasets that contain information for Mercury Insurance Group are only the Workers' Compensation and Other Liability datasets. I choose Other Liability line of business dataset because Mercury's information from Workers' Compensation mostly is zero.

The Claim Paid Losses in Other Liability Insurance for accident years in 1988-1997 form a development triangle as below. Presenting the data in a triangle structure is because

- It shows the development of claims over time for each origin period
- Most reserving methods of the "ChainLadder" package in R expect triangles as input data sets with development periods along the columns and the origin period in rows

Accident	Development year									
Year	1	2	3	4	5	6	7	8	9	10
1988	233	887	2548	4353	6210	7542	8133	8177	8295	8706
1989	142	828	2758	4412	6060	7448	8208	8555	8622	
1990	189	927	3304	6019	7463	8942	9366	9506		
1991	142	1043	2299	3764	4799	5380	5842			
1992	98	1284	3813	5995	7117	7846				
1993	164	824	2723	4090	6216					
1994	164	1238	3010	4603						
1995	285	1227	3149							
1996	213	858								
1997	163									

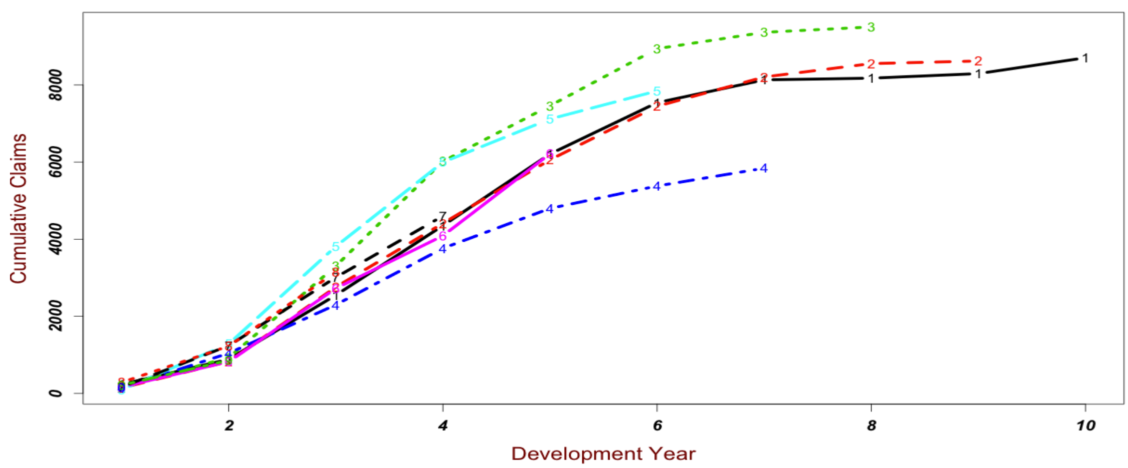
2. Objective: I aim to forecast the future claims development for the bottom right half of the triangle. In other words, I am estimating unobserved values and quantify the outstanding loss liabilities for each origin year.

For the first part of this paper, I will be using the loss development factor deterministic method and applying the implemented Mack and Boot Strap Chain Ladder Stochastic Reserving models in R. I then approach the same problem with statistical approaches, building a log linear model and a log-linked GLM

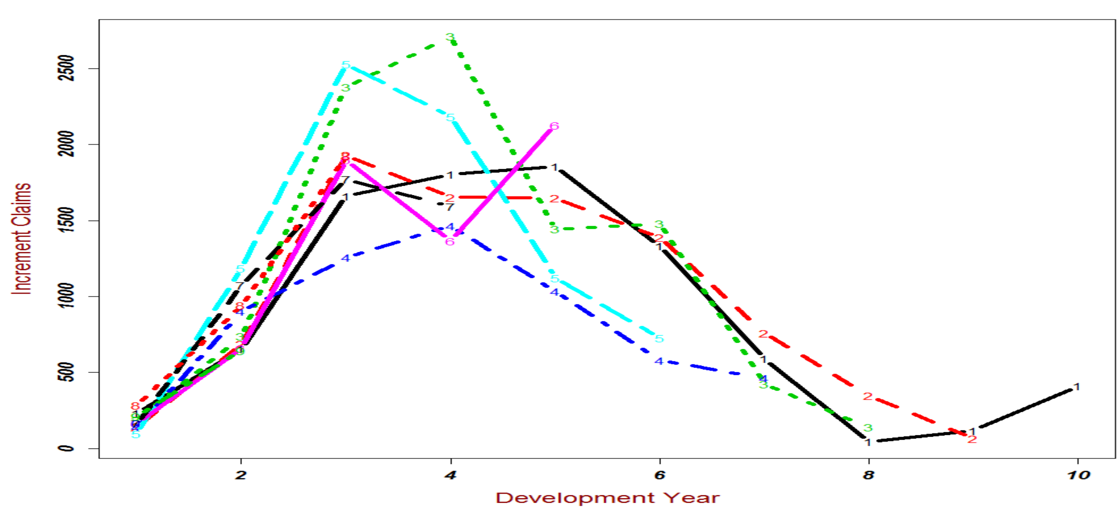
Poisson model. In conclusion, I summarized and discuss my achieved results for all the applied methods.

3. Data Overview

The cumulative claims development is visualized in the below Figure. Most of the insured incidents take 3-6 years to settle down and close. The growth rate is the highest after 2 - 3 years from the claims origin point but it does slow down with the increasing development period.



Incremental claims for the development years are displayed in the figure below. For this data set, the highest incremental claims come from the third, fourth and fifth development years. In general, the decreasing trend in claims development is visible.



4. Modeling

4. 1 Chain Ladder -- Reserving Methods

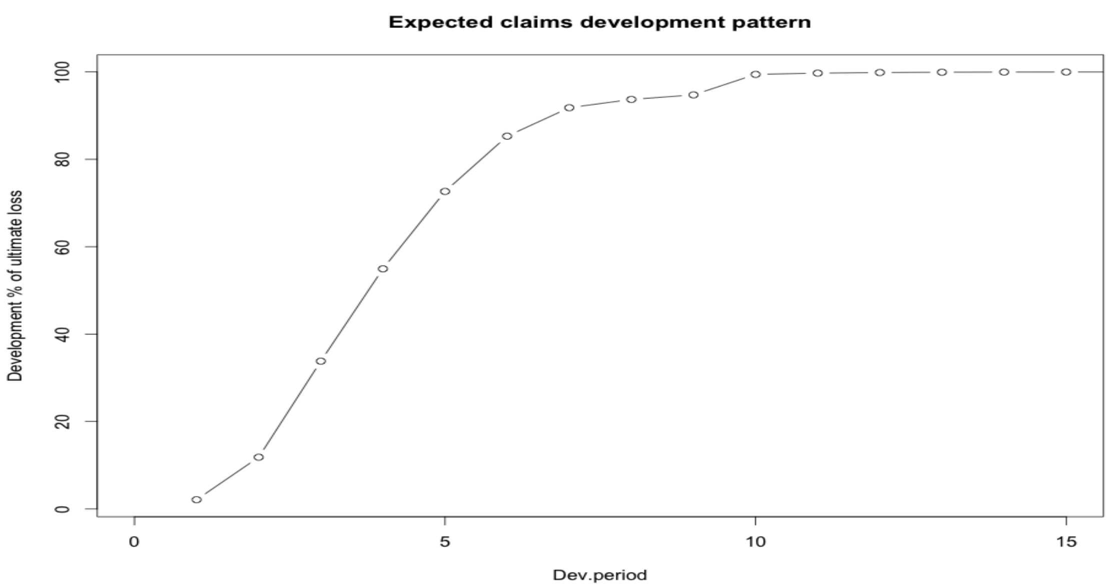
Chain Ladder methods use algorithms to forecast outstanding claims on the basis of historical data. They assume the cumulative claims losses for each business year develop similarly by delay year. Deterministic reserving method, Loss Development Factor method, uses the most basic chain ladder function. The Chain Ladder package in R has implementations of stochastic reserving models such as Mack, Munich, and Boot Strap Chain Ladder.

4.1.1 Loss Development Factor method -- Deterministic Reserving

Loss Development Factor method uses the basic chain ladder function, which link ratios are calculated as the volume weighted average development ratios of a cumulative loss development triangle from one development period to the next.

$$f_k = \frac{\sum_{i=1}^{n-k} C_{i,k+1}}{\sum_{i=1}^{n-k} C_{i,k}}$$

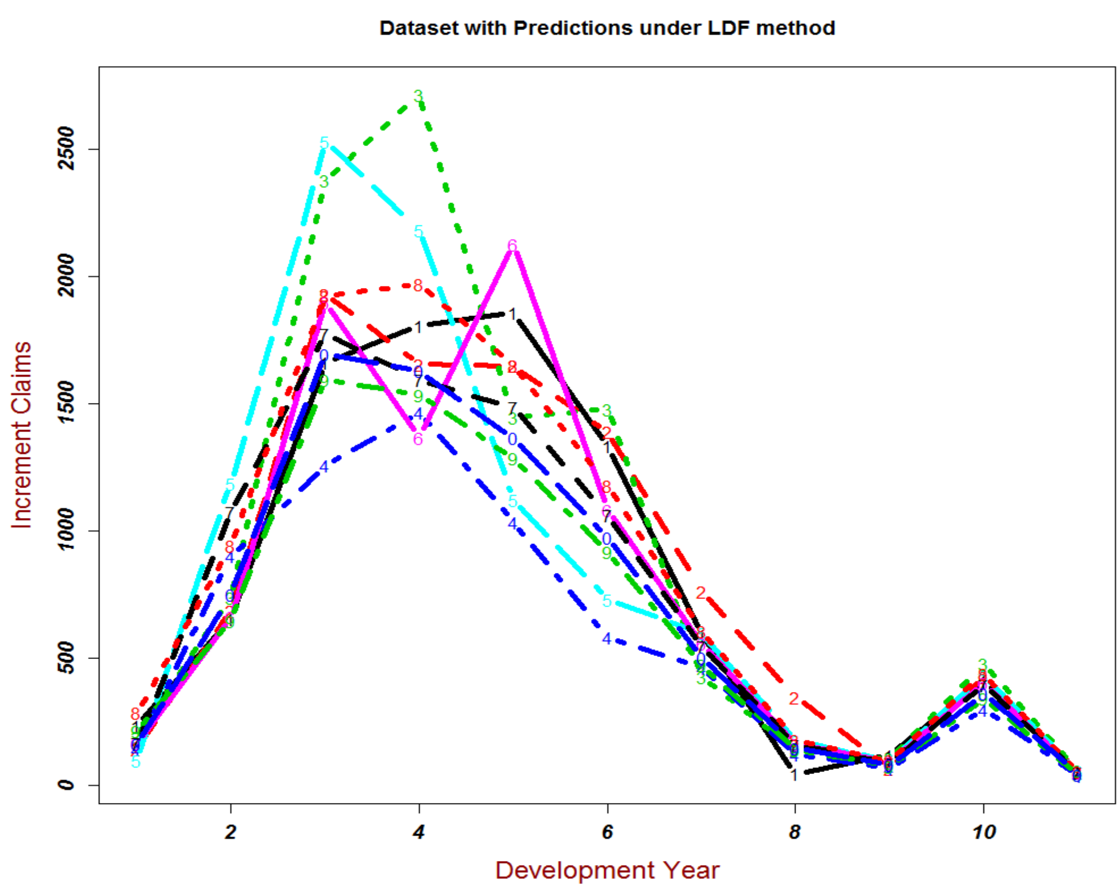
Since the oldest origin year is not fully developed, I extrapolate anther 100 development periods assuming a log-linear model. The link ratios then allow me to plot the expected claims development patterns.



The link ratios are then applied to the latest known cumulative claims amount to forecast the next development period. An ultimate column is appended to the right to accommodate the expected development beyond the oldest year (10) of the triangle due to the tail factor (1.005696) being greater than unity.

Accident Year	Development Year										
	1	2	3	4	5	6	7	8	9	10	Ult
1988	233	887	2548	4353	6210	7542	8133	8177	8295	8706	8756
1989	142	828	2758	4412	6060	7448	8208	8555	8622	9049	9101
1990	189	927	3304	6019	7463	8942	9366	9506	9611	10087	10145
1991	142	1043	2299	3764	4799	5380	5842	5963	6029	6327	6363
1992	98	1284	3813	5995	7117	7846	8445	8619	8715	9146	9198
1993	164	824	2723	1090	6216	7298	7855	8017	8106	8507	8556
1994	164	1238	3010	4603	6087	7147	7692	7851	7938	8331	8379
1995	285	1227	3149	5117	6766	7944	8550	8727	8823	9251	9313
1996	213	858	2452	3985	5270	6187	6659	6797	6872	7212	7253
1997	163	912	2606	4234	5599	6573	7075	7221	7301	7663	7706

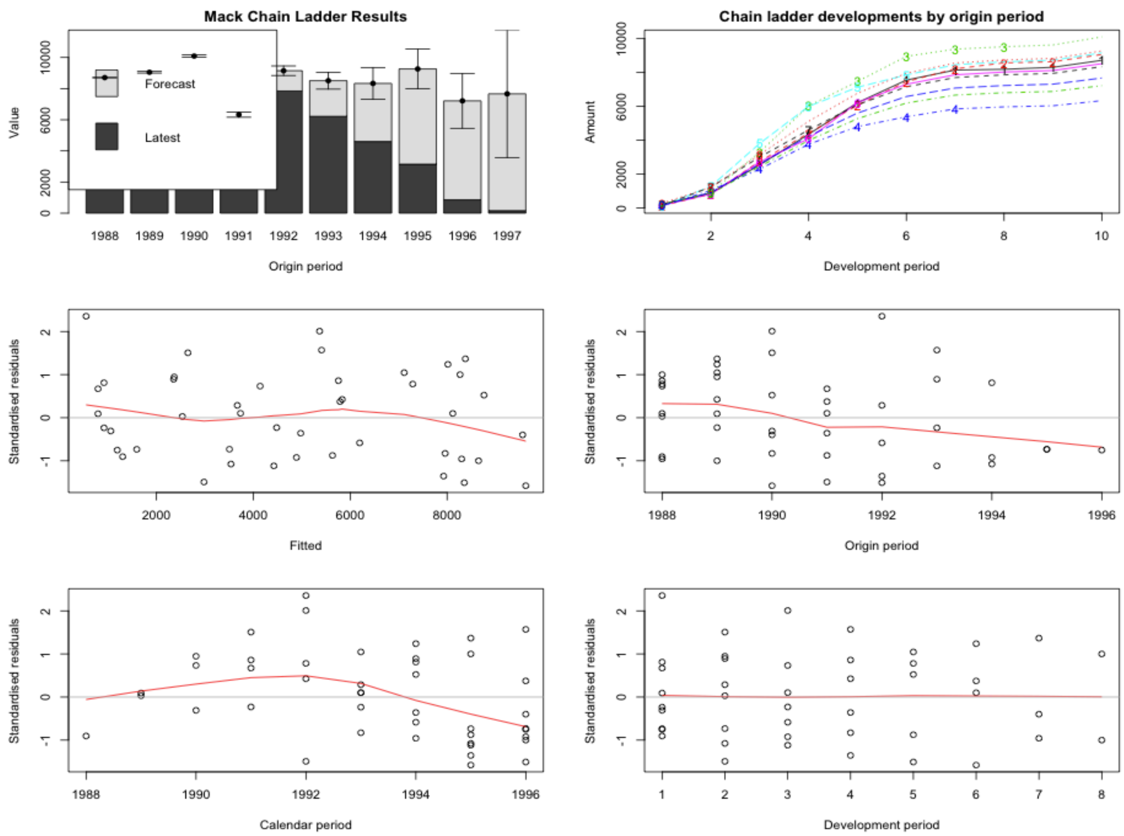
Based on the estimated unobserved values, I can then draw the complete graph of Incremental claims for each origin year. The total estimated outstanding loss under this method is about 28778.



4.1.2 Mack chain ladder Implement – Stochastic Reserving

The Mack Chain Ladder model forecasts IBNR (Incurred But Not Reported) claims based on a historical cumulative claim triangle and calculates the standard error for the reserves estimates. It can be regarded as a weighted linear regression through the origin for each development period: lm(y ~ x + 0, weights=w/x^(2- alpha)), where y is the vector of claims at development period k + 1 and x is the vector of claims at development period k.

The Mack's method is implemented in the ChainLadder package via the function MackChainLadder. *But* this method will only works if accident years are independent. To ensure Mack's Method is applicable for the dataset, we can check whether there are trends in the residual plots below.



These residuals plot show the standardized residuals against fitted value, origin period, calendar period, and development period. The bottom left plot looks perhaps more like a level drop in calendar year immediately after 1992. However, the fit to the most recent years of data isn't bad, so it might not be too problematic to use that forecast for the next year. I then access the loss development factors and the full triangle. Notice, not only that that the total amount of reserves is the same as using the deterministic method, but also the predict triangle.

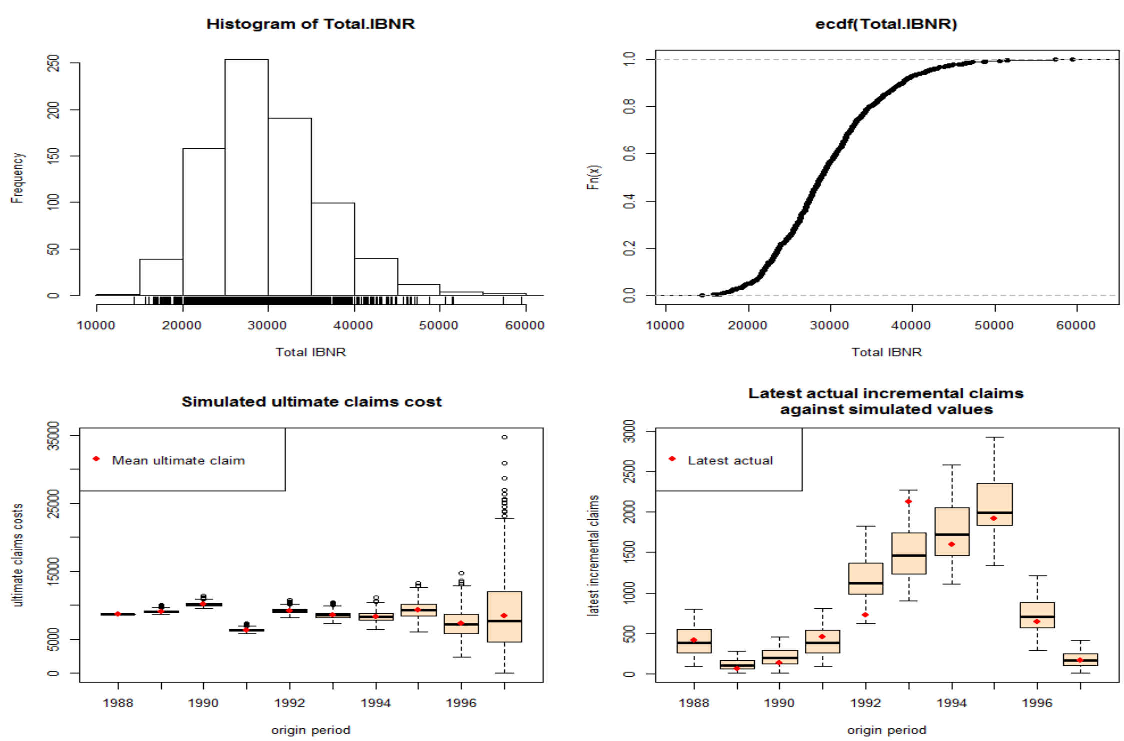
Accident Year	Development Year									
	1	2	3	4	5	6	7	8	9	10
1988	233	887	2548	4353	6210	7542	8133	8177	8295	8706
1989	142	828	2758	4412	6060	7448	8208	8555	8622	9049
1990	189	927	3304	6019	7463	8942	9366	9506	9611	10087
1991	142	1043	2299	3764	4799	5380	5842	5963	6029	6327
1992	98	1284	3813	5995	7117	7846	8445	8619	8715	9146
1993	164	824	2723	1090	6216	7298	7855	8017	8106	8507
1994	164	1238	3010	4603	6087	7147	7692	7851	7938	8331
1995	285	1227	3149	5117	6766	7944	8550	8727	8823	9261
1996	213	858	2452	3985	5270	6187	6659	6797	6872	7212
1997	163	912	2606	4234	5599	6573	7075	7221	7301	7663

4.1.3 Boot Strap Chain Ladder Implement– Stochastic Reserving

Boot Chain Ladder uses a two-stage approach. 1.Calculate the scaled Pearson residuals and bootstrap R times to forecast future incremental claims payments via the standard chain-ladder method. 2.Simulate the process error with the bootstrap value as the mean and using an assumed process distribution.

This two-stage boot strap approach is implemented in the Boot Chain Ladder function as part of the Chain Ladder package. As input parameters we provide the cumulative triangle, the number of bootstraps and the process distribution to be assumed:

```
BootCL=BootChainLadder(top_tri,R=800,process.distr="od.pois")
```



- Top left: Histogram of simulated total IBNR
- Top right: Empirical distribution of total IBNR
- Bottom Left: Box-whisker plot of simulated ultimate claims cost by origin period
- Bottom right: Box-whisker plot of simulated ultimate claims cost by origin period

The set of reserves obtained in this way forms the predicted distribution, from which summary statistics such as mean, prediction error and quantiles can be derived. The distribution of the IBNR appears to follow a log-normal distribution, so let's keep that in mind.

4. 2. Statistical Modeling

There are three main categories in statistical predictive modeling: Classical Linear Models, Generalized Linear Models(GLMs), and Data Mining. The differences between Linear Models and GLMs is as followed.

- Regression:

$$Y_i = \mu_i + \varepsilon$$
$$\mu_i = X'\mathrm{B}$$
$$\varepsilon \sim N(0,\sigma^2)$$

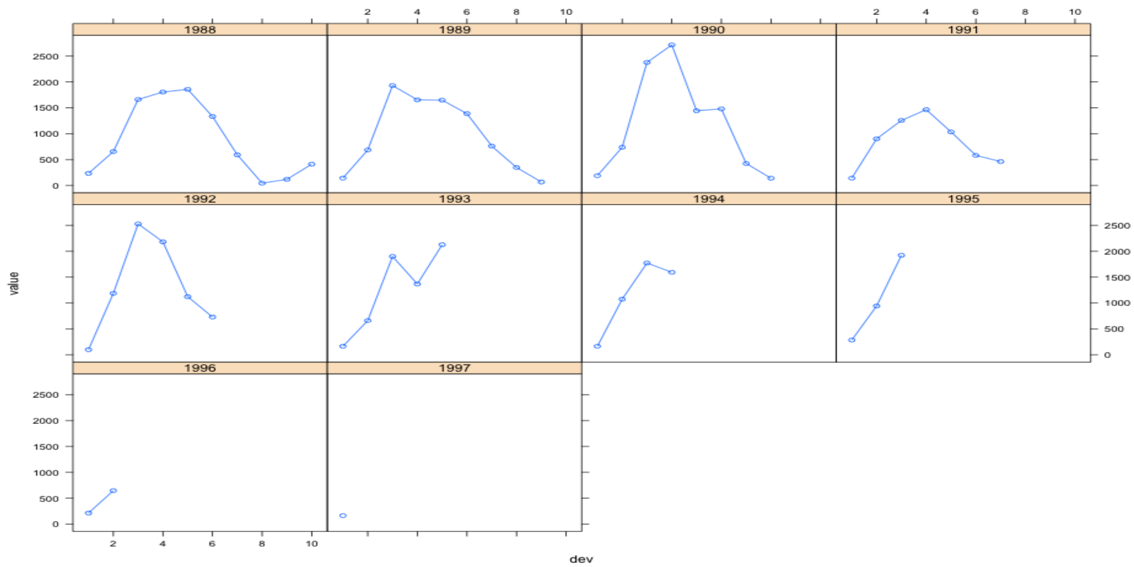
- GLM:

$$Y = h(\mu_i) + \varepsilon$$
$$h(\mu) = X'\mathrm{B}$$
$$\varepsilon \sim \text{exponential family}$$
$$\text{h is a link function}$$

An easy way to think about GLMs is as models that generalize the error term distribution to a family of distributions, called exponential family. It includes normal, binomial, Poisson, and gamma distributions among others. In addition, the response variable in GLM is related to linear regression through a link function. Common used link functions are Identity, Inverse, Inverse Squared, Log and Logit.

4.2.1 Pre-Analysis

The chain ladder methods uses cumulative claims, but statistical approaches uses the incremental claims. The R package ChainLadder comes with two helper functions, cum2incr and incr2cum. They can transform cumulative triangles into incremental triangles and vice versa. The development of the incremental claims is shown in below figure individually for each origin period. we can see that the outcome is a continuous outcome, but is right skewed and always positive.

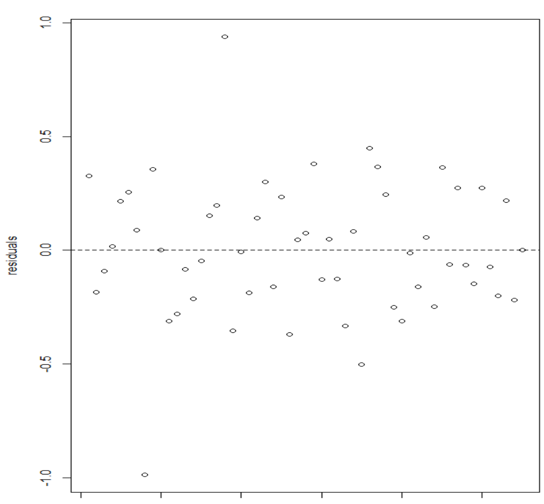


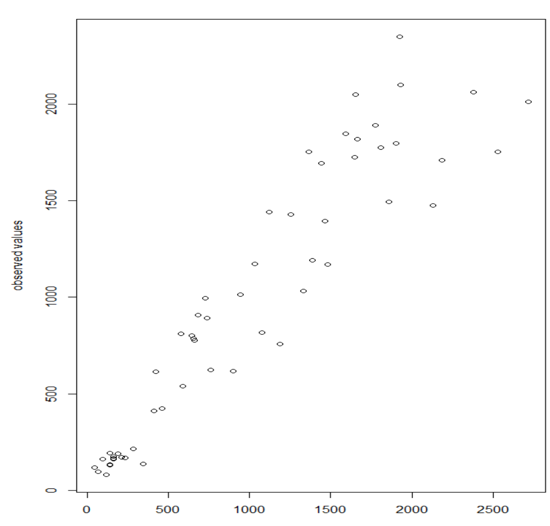
4.2.2 Linear Model with log-transformed outcome

Since the distribution has a positive skew, taking a natural logarithm of the variable helps fitting variable into a model. Thus, that is the first model I build, and carry out the linear regression with

lm(log(inc_loss) ~ as.factor(dev) + as.factor(ay), data=inc_data)

Despising a few outliers, the residual plot below look quite well behaved. In an ideal case, the observed values vs. the fitted values plot should also be distributed along the diagonal. The total correlation coefficient between the fitted and observed value is 0.8827, thus I decided to investigate further.



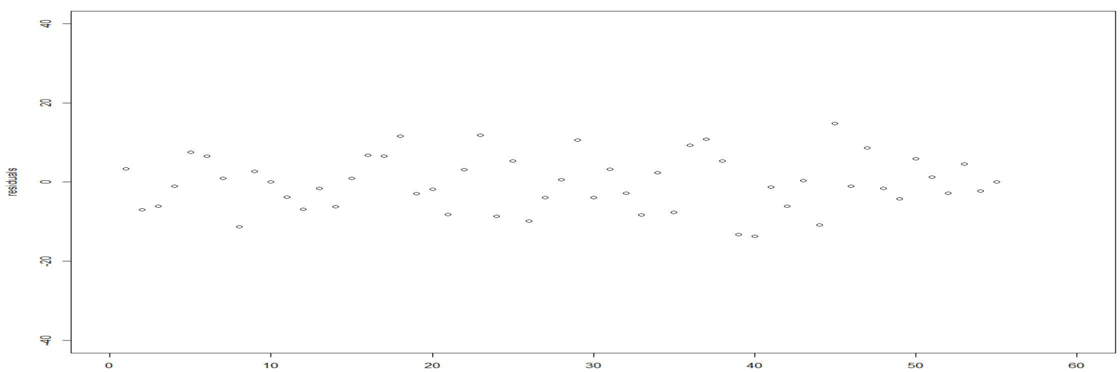


4.2.3 Log-linked GLM Poisson

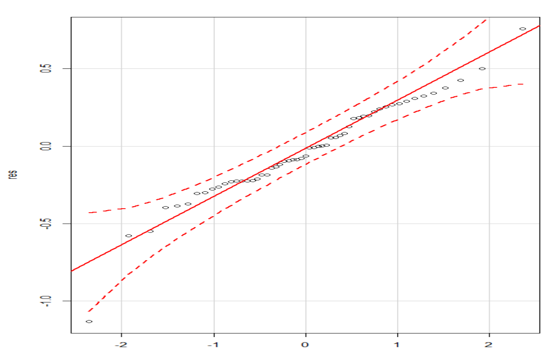
Since the outcome is right skewed and always positive, the incremental losses seem to be Poisson distributed. Thus I choose using the Poisson family in GLM. The link function is log() to be consistent with the previous linear model, thus the model is modeling the following:

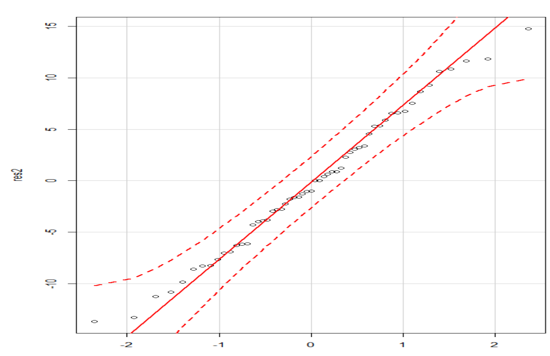
glm(inc_loss ~ factor(ay) + factor(dev), data=inc_data, family=poisson("log"))

Looking at the residual plot below, we can see that I was able to successfully get rid of the residual outlines from the previous model. In fact, the residuals is stationary with zero mean, and constant variance.



The observed values vs. the fitted values plot is distributed along the diagonal. The relationship between observed and fitted value is satisfying, the correlation coefficient is 0.9497. For model fit checking purpose, I also compare the qqPlot(s) from these two models. And the log-linked GLM Poisson indeed fits the data better.





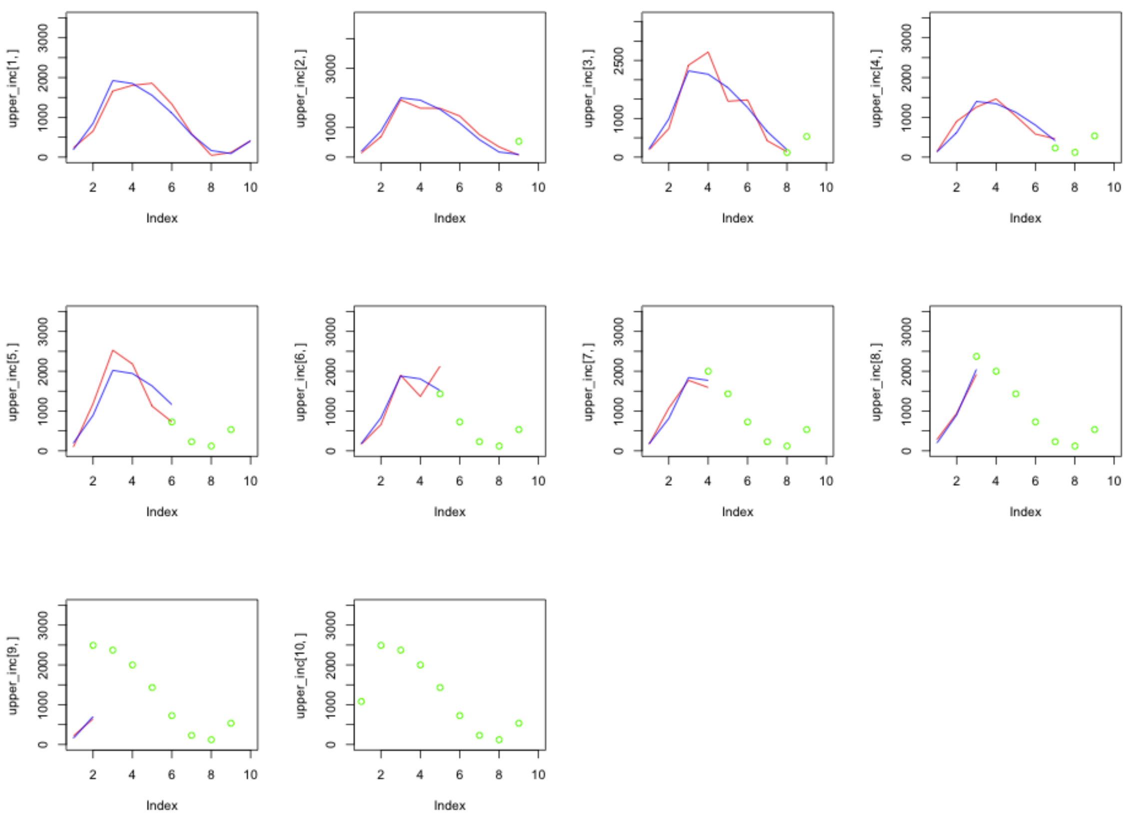
Prediction of the claims

The intercept term estimates the first log-payment of the first origin period. The other coefficients are then additive to the intercept. Thus, the predictor for the second payment of 1989 would be exp(5.22139 + 0.03866 + 1.52445)=884.038.

The second column in the output above gives us immediate access to the standard errors. Based on those estimated coefficients, we can predict the incremental claims payments. The total amount of reserves is the sum of incremental predicted payments beyond year 1997.

```
sum(predict(gl,type="response", newdata=subset(Claims, cal > 1997)))
```

For a better illustration of how fitted my model capture the observed claims development. The following graph is provided. The red lines represent the observed incremental losses. The blue lines represent the model fitted values, the green lines stand for the predicted incremental losses.



There are some overestimations and underestimations and underestimations at the middle year of the claim development. But the fit is satisfying. The total amount of financial instruments that need to be held in claims reserve under this model is 28779.

5. Conclusion:

The main goal of this paper was to learn the preliminary reserving techniques for insurance companies like Mercury Insurance Group. Once an appropriate model is built, the predicted claim reserve is approximately around 28778, whether I was using the deterministic method, applying the implemented stochastic reserving models, or building statistical models.

Bibliography

Glenn Meyers and Peng Shi,
http://www.casact.org/research/index.cfm?fa=loss_reserves_data

Package 'ChainLadder'
http://cran.r-project.org/web/packages/ChainLadder/ChainLadder.pdf

Arthur Charpentier, *Computational Actuarial Science With R,* Chapter 14

Zuzana Kaderjakova, *Modeling Dependencies in claims reserving*