INTRODUCTION

This paper deals with three main topics: Generalized Linear Model, Loss Reserving Techniques in Property-Casualty Insurance and how Generalized Linear Model can be applied to the Loss Reserving Problem in Property-Casualty Insurance in order to obtain a “better” result.

In past few decades there has been a large amount of research conducted in the field of modeling and prediction theory. Analyst and Scientist from different fields such as Investment, Credit Card and Insurance Industry, Medical and Biological Sciences and Image Processing conducted notable research on these topics. The most important reasons perhaps the massive data explosion over the past few years. It is questionable whether one should look at this data explosion as a curse or as a blessing. But with cutting-edge computing technology the Analysts and the Scientists of the above industry took this as a blessing and developed some advanced, mathematically sophisticated and practically reliable statistical methods. They have innovated techniques like Neural Networks, Genetic Algorithm and resurrected Generalized Linear Model. As we have mentioned at the beginning, one of the topics we will discuss in this paper is the Generalized Linear Model. In the Chapter-I, we will talk about the mathematical and statistical aspect of the Generalized

Linear Model at large. We will see the limitations of the classical linear model and the necessity of the invention of the generalized linear model.

Actuarial Science involves the science of financial uncertainty. The insurance industry is all about handling financial risk. Management of this risk, by predicting the future financial developments, is the main job of an Actuary. However, this act of prophecy is not quite easy and needs a combination of art and science. Precisely speaking, this prediction is dependent on several statistical tools along with the business knowledge and actuarial judgment from the Actuary. In Life Insurance the uncertainty is based on the life time of the insured or individual. On the other hand in Property-Casualty Insurance the uncertainty is emerged from the defined "failure" of the insured item or occurrence of a casualty. However, we will concentrate only on the P&CI part in this paper. An insured transfers the risk of the losing or damaging a property or affecting by a casualty to the insurer. The P&CI is liable to pay the insured an amount (based on the policy) of the loss occurred and this is called the liability in insurance terminology. As we have mentioned already, this loss is uncertain and the financial amount to be paid as liability is also unknown. This gives birth of the uncertainty in the all other components of the insurance e.g. pricing, rate-making and reserving. In the Chapter-II we will discuss the reserving techniques in the P&CI. We will define the terms and terminologies used and will provide an extensive idea about the P&CI loss reserving techniques.

In the last chapter, we will explore how the Generalized Linear Model can be used as a tool for P&CI loss reserving. In Chapter 2 we will see that the all traditional methods used for predicting the amount of necessary reserve, the ultimate claim value or the loss value are deterministic in nature. Due to simplicity they are popular in use but by using those, an Actuary will not have the provision for any diagnostic tests and estimating confidence intervals. The traditional stochastic methods e.g. two-way analysis or minimum biased method do not allow considering the interaction between two determining factors. The generalized linear model provides different statistical features including diagnostic tests and it captures the interaction effects of two determining factors. Unlike, classical linear model, the generalized linear model allows considering a non-Normal distribution as underlying model distribution. This is a very useful flexibility of the generalized linear model, as insurance severity data is believed to follow the Gamma distribution. In this chapter, we will mostly give a comparative discussion of the above techniques and will show the advantages of considering the GLM using a case study based on a paper published by Roosevelt Mosely (2004). Data analysis in this paper (Roosevelt, 2004) is based on 1992 Auto-Insurance data published by the Insurance Research Council (IRC). Guided by this paper we will discuss several possible explanatory variables which have potential to determine the ultimate settlement value of a reported loss. We will also discuss the diagnostic test of the final model.

Like any other model used in reserving technique, the Generalized Linear Model is nothing more than a tool used for predicting ultimate loss value or determining the reserve. Applying the generalized linear model can not replace the actuarial judgment and business knowledge (Roosevelt, 2004). However, a predictive data analysis such as the generalized linear model significantly improves any system and prediction procedure. Last but not the least due to its well-built mathematical back-bone it is easier for the reserving Actuary to explain his prediction to the management and underwriters.

CHAPTER I

GENERALIZED LINEAR MODEL

In this chapter we will discuss the statistical theory of the Generalized Linear Model (referred to as *GLM* in what follows). In order to present the topic in a more mathematically elegant way we will establish some required terminologies. We will represent the data by a two dimensional array, often known as *data matrix*, in which rows will be indexed as experimental units or *uncorrelated* observations. On the other hand, the columns of the data matrix are known as variates, some of which will be regarded as responses or dependents variates due to their sensitivity to the explanatory variables or *covariates* which is popularly known as *independent variables* (McCullagh, 1989). Covariates can be observed as a *quantitative value* or as a *qualitative value*. Qualitative variates are usually called *factors* in statistics and take on a finite set of values or labels (McCullagh, 1989). Dependent variables can be continuous or discrete in nature. They may be even observed as factors and take on values from a finite set of classes. GLM is essentially an extension of the *classical* concept of *multiple linear regression model*. So it is important to discuss the theory of multiple linear regression model before we proceed any further. The first section deals with definition and formation of the multiple linear regression model and statistical inference based on the same mainly for the expectations.

1.1 Multiple Linear Regression Models

The basic purpose of linear regression model is to find a *linear* relationship between a dependent variable and one or several independent variables. In most of the practical problems the number of independent variables is more than one in which case the model is called multiple linear regression model (Pearson, 1908) (McCullagh, 1989). Let us discuss a practical problem to understand the basic problem addressed by multiple linear regression model. In order to predicting the *stock price* of a insurance company one can consider different explanatory variables such as net written premium, loss ratio, combined ratio, GAAP earnings of the company along with different relevant market variables. The statistician should take care of different possibilities of the relationships. Usually stock price at a particular time point *t* has a relation with the value of the above independent factors at time point (*t-1)*. For example, the stock value assessed at the 4th quarter is likely to be related with the values of the above independent variables at the 3rd quarter. Different problem and purpose cultivated several kinds of linear models. But as we have mentioned before, the basic goal is to find a *linear combination* through which one can predict the dependent variable using the given value of explanatory variables. The coefficients of this linear combination are called model parameters and primary quantity to estimate. Some times it is required to study more than one *dependent* variable. In which case a *combination of the dependent variables* are considered. In this context, let us digress from our main topic for a while for short discussion of *general* linear model (StatSoft, 2003). The emergence of the theory of algebraic invariants in the 19^{th} century by outstanding work from mathematicians such as Gauss, Legendre, Boole, Cayle and Sylvester cultivated the root of the general linear model (StatSoft, 2004). The theory searches to identify those quantities in systems of equations which remain unchanged under linear transformations of the variables in the system. Invariants are extremely useful for classifying mathematical objects because they usually reflect intrinsic properties of the object of study. Similar to the most of the discoveries in theoretical mathematics, the success of the theory of invariants was far beyond the dreams of its originators. The development of the linear regression model in the late 19^{th} century and the development of correlation methods shortly thereafter, are clearly direct outcomes of the theory of algebraic invariants (StatSoft, 2004). Regression and correlational methods, in turn, serve as the basis for the general linear model. A very basic application of this theory is the result that the correlation between two statistical variables is remained unchanged under a linear transformation on those two variables. However in this paper we will not address this topic any further.

Our approach of discussing the multiple regression model is intended to make the extension to GLM natural. Guided by McCullagh and

Nelder (1989) we will represent the multiple regression model in the following form:

(C1) $Y_i \sim N(\mu_i, \sigma^2)$ *,*which means that the observations are normally and independently distributed with mean μ_i for the i^th^ observation and a common variance σ^2.

(C2) $\eta = \sum_{j=1}^{p} x_j \beta_j$ *,* which is linear predictor based on the covariates: $x_1, \ldots, x_p$.

(C3) $\eta = \mu$, which is identity *link.* Once we define the whole model in mathematical notations, immediately we will answer what we mean by the word *link.*

In this description, the data vector ***Y***, the mean vector μ, and the linear predictor η, all have *n* components. It is clear that the above set-up of the multiple linear model is a combination of the systematic or deterministic component and the stochastic component. As we have promised, we will now discuss what we have meant by the *link* η. In the above description we could simply avoid the concept of *link function*, but advent of this concept at this basic set up will ease the development of the *GLM* setup.

The *link function* defines a relation between the linear predictor η with the expected value μ of the data points ***Y*** ((McCullagh, 1989)). As we have seen in the above setup, the mean and the linear predictor are the identical and both of them can assume any real value. But this is not a plausible relationship in all cases. For an example, when we deal with binomial distribution we have 0 < μ < 1 and we need a link that should map the interval (0,1) onto the entire real line. For a counting distribution, such as Poisson distribution we must have μ > 0 while η could be negative. In this case we will have relation $\mu = e^{\eta}$ or $\log(\mu) = \eta$ which also implies that additive effect contributing to η will act as a multiplicative effect contributing to μ. So precisely speaking, link is a function that relates the expected value of the observations to the linear predictor (McCullagh, 1989). We will be back to this discussion and will state different link functions corresponding to different distributions of the exponential family later on in this chapter.

With the setup we have for multiple regression model, the very next job is to estimate the parameters $\beta_j,\ j = 1(1)p$. The theory of *least square estimation* answers this problem of estimating the parameters in a linear model. The basic foundation of this theory were founded (Rao, 2002) by Gauss (1809) and Markoff(1900) and later on significantly modified by different statisticians such as Aitken(1935), Bose(1950-51), Neyman(1938), Rao(1945) to name a few. At this point we will discuss the *least square estimation* method using the *Gauss-Markoff* setup $(\mathbf{Y}, \mathbf{X}\underset{\sim}{\beta}, \sigma^2 I)$.

Let us consider *uncorrelated* observations $y_1, \ldots, y_n$ such that:

$$\left.\begin{aligned} E(y_i) &= \sum_{j=1}^{p} x_{ij}\beta_j \\ V(y_i) &= \sigma^2 \end{aligned}\right\} i = 1(1)n$$

where $\underset{\sim}{\beta}$ and σ^2 are unknown parameters and $(x_{ij}) = \mathbf{X}$ is a matrix of known coefficients.

If $\boldsymbol{Y}$ and $\underset{\sim}{\beta}$ represents the column vector of the variables y_i and β_j, then we can re-write the above set of equations as:

$$\mathbf{Y} = \mathbf{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}, E(\underset{\sim}{\varepsilon}) = 0, D(\underset{\sim}{\varepsilon}) = \sigma^2\mathbf{I} \Rightarrow E(\mathbf{Y}) = \mathbf{X}\underset{\sim}{\beta}, D(\mathbf{Y}) = \sigma^2\mathbf{I} \qquad (0.1)$$

with the usual notations: residuals $\underset{\sim}{\varepsilon}$ are differences between the observed values and the corresponding values that are predicted by the model (StatSoft, 2004), σ^2 as the variance of the dependent variables or equivalently variance of the error terms $\underset{\sim}{\varepsilon}$ (this is often called the *residual variance* and they represent the variance that is not explained by the model; the better the fit of the model, the smaller the values of residuals (StatSoft, 2004)), D denotes the *dispersion matrix* (the dispersion matrix or the variance-covariance matrix consists of the variances of the variables along the main diagonal and the covariance between each pair of variables in the other matrix positions (Rao, 2002))and $\mathbf{I}$ identity matrix of order n. We need to estimate the β_j on the basis of the observed y_i. Aitken(1935) has given a more general set-up as follows:

$$E(\mathbf{Y}) = \mathbf{X}\beta, D(\mathbf{Y}) = \sigma^2\mathbf{G}, |\mathbf{G}| \neq 0 \qquad (0.2)$$

The idea of introducing a known n^th order matrix **G** is to capture any possible correlation among the observations of which $\mathbf{G} = \mathbf{I}$ is a special case when there is no correlation. The setup (1.2) can be reduced to the setup (1.1) with the transformation $\mathbf{Z} = \mathbf{G}^{-1/2}\mathbf{Y}$ which leads to:

$$\begin{aligned} E(\mathbf{Z}) &= \mathbf{G}^{-1/2}\mathbf{X}\beta = \mathbf{U}\beta \\ D(\mathbf{Z}) &= \sigma^2\mathbf{I} \end{aligned} \tag{0.3}$$

Setup (1.3) suggests that it is enough to discuss the parameter estimation based on the simple model stated at (1.1) which is denoted by $(\mathbf{Y}, \mathbf{X}\underset{\sim}{\beta}, \sigma^2\mathbf{I})$. We will avoid fathom into any further mathematical theory behind this estimation method; rather we will state the result of the *least square estimation* whenever needed in order to find the solution for $\hat{\underset{\sim}{\beta}}$. The *sum of squares of differences* between the observations and the expectations is called *normal equation* which can be stated as follows:

$$(\mathbf{Y} - \mathbf{X}\underset{\sim}{\beta})'(\mathbf{Y} - \mathbf{X}\underset{\sim}{\beta}) = \sum (y_i - x_{i1}\beta_1 - \ldots - x_{im}\beta_m)^2 \tag{0.4}$$

The equation in $\underset{\sim}{\beta}$ is obtained by minimizing (1.4) and given as:

$$\mathbf{X}'\mathbf{X}\underset{\sim}{\beta} = \mathbf{X}'\mathbf{Y} \tag{0.5}$$

The observational equation is in general inconsistent while the normal equation necessarily has a solution, as $\mathbf{X}'\mathbf{Y} \in \boldsymbol{M}\ (\mathbf{X}'\mathbf{X})$, which states that there exists a $\underset{\sim}{\beta}$**,** that satisfies the equation (1.5). We will denote such assured solution of (1.5) as $\hat{\underset{\sim}{\beta}}$ which results to:

$$(\mathbf{Y} - \mathbf{X}\underset{\sim}{\beta})'(\mathbf{Y} - \mathbf{X}\underset{\sim}{\beta}) \geq (\mathbf{Y} - \mathbf{X}\hat{\underset{\sim}{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\underset{\sim}{\beta}}) \tag{0.6}$$

So $(\mathbf{Y}-\mathbf{X}\hat{\underset{\sim}{\beta}})'(\mathbf{Y}-\mathbf{X}\hat{\underset{\sim}{\beta}})$ is the minimum and it is attained at $\underset{\sim}{\beta}=\hat{\underset{\sim}{\beta}}$ and unique for all such $\hat{\underset{\sim}{\beta}}$. An interesting statistical note can be mentioned at this point is $\hat{\underset{\sim}{\beta}}$ that minimizes (1.4) is also the *maximum likelihood*. This $\hat{\underset{\sim}{\beta}}$ is known as the *least square estimator* (Rao, 2002).

As we have seen so far for developing the theory of least square estimation and finding out the least square estimator, we do not require the Normality assumption what we have made at the very beginning of the section at *C1*. In order to develop the theory of least square estimation we need to use only first order and second order assumption what we have made. We do not require the normality assumption which is in fact an advantage as it is a quite hard to realize the actual underlying distribution. Precisely speaking our parameter estimation will remain valid even if the underlying distribution is not from normal distribution. So the most important assumption is the assumption of *constant variance* which should be checked using a statistical method or by simple graphical method. One more interesting point to note that thanks to *Central Limit Theorem* (*Linear Statistical Inference and Its Applications, Rao, 2002*), normality assumption can be relaxed in case we are working with the sum or the mean a large sample.

At this point we have the multiple linear regression model and we know how to estimate the parameters of this model given a set of observations. Let us now clarify some intuitive concept of this modeling approach and estimation problem. The parameters of the model actually represent the contribution of the corresponding explanatory variables in predicting the dependent variable. Here we will bring the concept of *partial correlation*. This concept comes from the fact that β_j gives the information of how much the corresponding explanatory variable X_j is related to the dependent variable Y, after suppressing the effect of all other present explanatory variables (StatSoft, 2004). However this cultivates the root of one of the most important limitations of this setup. We can give an example (StatSoft, 2004): how much do we agree with the fact the there exists a negative correlation between *length of hair* and *height of a person* in a population? Perhaps it seems that there is no logic as such for this to happen. But if we consider the fact that women are in general shorter than men in a population and accordingly introduce another explanatory variable *gender*, then the above phenomenon will be clarified.

We will close this section by discussing the concept of R^2 measure and *correlation coefficient* related to the multiple linear regression model. A small variability (compare to the overall variability) of residual values around the regression line depicts a good prediction or in other words states that there exists a good relation between X and Y. On the other hand if there is a lack of relation between X and Y, then the ratio of the residual variance to the overall variance will be found to be close to zero, whereas in the former case it was expected to be closer to 1. The term *coefficient of determination* or R^2 is defined to be 1 minus this ratio and we expect this value will be in the closed interval [0,1] (StatSoft, 2004). R^2 is a direct measure of how well the model fits the data and a value more than 60% is believed to be a good fitness. In practice, the positive square root of R^2 is named as *correlation coefficient* (and popularly used for detecting the degree of relation between the independent or explanatory variables *X's* and the dependent variable *Y*. Its value lies between *0* and *1* and this will give an overall dependency in the multiple linear regression model. On the other hand if we want to see the relationship between the individual independent variables and the dependent variable, we need to look at the individual β value. For example, a positive tells β us that there is positive relation between the dependent variable and the corresponding independent variable, whereas a negative value implicates a negative correlation. Understandably, if β assumes the value 0 then there is no relation between the dependent variable and the corresponding independent variable. As matter of fact the calculated β for the height of the people in the example we have given for predicting the length of hair is expected to be negative.

1.2 Assumptions and Limitations of Multiple Linear Regression Model

In this section we will clearly mention the assumptions have to be considered for multiple linear regression model and will show the consequent limitations of this setup due to the assumptions taken.

In this set up we need to assume that the relationship between dependent variable and the independent variables if exists, it would be a linear relationship. We can clarify this assumption in two ways. *First*, we can always approximate a second or higher order polynomial by a set of straight lines (remember how Archimedes found the area of a circle by approximating it by a polygon). So in a system, we can consider a complex process as a sum of several simple linear processes. But still we are *approximating* the model or true relation, which may be fatal for the prediction. The *second* approach to handle this limitation is to find out a proper *transformation* to change any possible non-linear relation into a linear relationship. At this point we should make two very important comments: one, before setting up the model the first and foremost thing we should do is to try to understand the relationship of the dependent variables, individually, with the dependent variable using the bivariate scatter plot. It will clearly depict the needed transformation in order to set up the linear model. Secondly, when we talk about a linear model it is linear in terms of the parameters, not in terms of the independent variables. If from scatter plot we need to imply that there is no linear relationship between a particular independent variable *X* and *Y*, rather there is a second order relationship, then we should consider another variable, *Z* with the transformation function *square* ($Z = X^2$) will give a linear relationship between *Y* and *Z* and we will consider *Z* in the model. Another set of assumption what we have discussed in the previous section is normality assumption and constant variance assumption on residuals (hence the dependent variable, as the residual is defined is to be predicted minus actual value of the dependent).Again, classical approach of linear regression transform the response variable to satisfy these conditions. But these transformations may not exist in reality (McCullagh, 1989). More importantly, the response variable may assume only positive values, which is very common in insurance related data but normal distribution assumes value in the entire real line. Also, if the response variable is strictly positive, then the variance of the response variable goes to zero as its mean goes to zero. This reasoning gives the idea that variance could be a function of the mean, not a constant (Anderson et al, 2004).

The way the multiple linear regression model has been set up in the previous section, assumption *C1* can not be a valid assumption for many insurance risks. The relation can be multiplicative not necessarily additive. In which case the assumption stated at *C2* is certainly violated (Anderson et al, 2004). Finally, we can have problem where the assumption *C3* should not be assumed. We have already mentioned example where the concept of *identity link* function does not hold. In insurance industry we come across several non-normal distributions (for example, *severity distribution* is usually *Gamma* or *Log-Normal* distribution) and most of these have strictly positive range. Let us recollect again that the residuals are predicted value of the dependent variables minus its actual value. This quantity could be any value around 0 and in that case if the underlying distribution of the dependent variable is strictly positive then we *can not* relate the expected value of the dependent variable with a identity function. This is a very severe drawback of multiple linear regression model setup that can be addressed by GLM set up. There is another conceptual limitation of all these regression techniques: what we can tell from these models is whether there exists a relation between dependent variable and the independent variables or not; what we can not tell is what the actual causal mechanism is. This can be a considerable problem when we actually choose explanatory variables to be present from the entire class of explanatory variables (McCullagh, 1989).

All most all of these limitations can be address by employing GLM technique. Before going detail into the theory of GLM we need to discuss a special class of distribution called the *exponential family of distributions*, which is a very important concept for developing the theory of GLM.

1.3 Family of Exponential Distribution

Exponential family of distribution is an extremely important and useful class of distributions in statistics. This family of distributions is mathematically elegant and has nice algebraic properties; at the same time all of these distributions occur naturally and quite justified to consider as the underlying model distribution. Distributions of this family can be continuous as well as discrete. The generic form of the density function of the exponential family can be written as (McCullagh, 1989):

$$f_y(y;\theta,\phi) = \exp\{(\frac{y\theta - b(\theta)}{a(\phi)}) + c(y,\phi)\} \tag{0.7}$$

Where $a(.)$ is positive and continuous, $b(.)$ is twice differentiable with the second derivative a positive function, precisely speaking a convex function and c(.) is independent of the parameter θ. With a known ϕ, this is an exponential-family model with *canonical parameter* θ and we are mainly interested in estimating θ. In case ϕ is known, it may be a two-parameter exponential family; ϕ is called *nuisance parameter* as σ^2 in regression analysis. Normal, gamma, inverse-Gaussian, exponential distribution are the example of continuous distributions of this family whereas Poisson, Bernoulli, binomial are example of discrete distributions within exponential family of distributions. Now we will clarify the expression (1.7) with the example of normal distribution, as a continuous distribution and Bernoulli as a discrete distribution (McCullagh, 1989)

Normal Distribution:

Suppose ***Y*** is normally distributed, i.e. ***Y~N(μ,σ²)***. Its density can be written as (Hardle et al, 2004):

$$f_y(y;\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{1}{2\sigma^2}(y-\mu)^2\} \qquad (0.8)$$

This can be rewritten as:

$$f_y(y;\mu,\sigma)=\exp\{y\frac{\mu}{\sigma^2}-\frac{\mu^2}{2\sigma^2}-\frac{y^2}{2\sigma^2}-\log(\sqrt{2\pi\sigma^2})\} \qquad (0.9)$$

From (1.9) we can readily observe that Normal distribution is a member of the exponential family of distributions with the parameters:

$$a(\phi)=\sigma^2, b(\theta)=\frac{\mu^2}{2}, c(y,\phi)=-\frac{y^2}{2\sigma^2}-\log(\sqrt{2\pi\sigma^2}),$$

where $\phi=\sigma, \theta=\mu$.

Now suppose, ***Y*** is Bernoulli distributed (The Bernoulli distribution can be described as the situations where a "trial" is made resulting in either "success" or "failure," such as when tossing a coin or when modeling the success or failure of a mechanical procedure (StatSoft, 2004)) with the probability function (Hardle et al, 2004):

$$P(Y=y)=\mu^y(1-\mu)^{1-y} \qquad (0.10)$$

which can be viewed as:

$$P(Y=y)=(\frac{\mu}{1-\mu})^y(1-\mu)=\exp\{y\log(\frac{p}{1-\mu})\}(1-\mu) \qquad (0.11)$$

using the *logit*, $\theta=\log(\frac{\mu}{1-\mu})\Leftrightarrow\mu=\frac{e^\theta}{1+e^\theta}$, we have an exponential family with:

$$a(\phi)=1, b(\theta)=-\log(1-\mu)=\log(1+e^{\theta}), c(y,\phi)=0 .$$

This distribution does not involve any nuisance parameter ϕ.

Weibull and Log-normal distributions are not in the class of exponential family of distributions.

In the next section we will see how this special class of distributions plays a very important role in the theory of GLM.

1.4 Generalized Linear Model – Setup and Parameter Estimation Techniques

As the name suggests, GLM extends the concept of the classical multiple linear regression model. As we have discussed in the section 1.1, the classical setup of the linear model assumes that the response variables is a linear combination of the independent variables and a normally distributed error term.

$$\mathbf{Y}=\mathbf{X}'\underset{\sim}{\beta}+\underset{\sim}{\varepsilon} \tag{0.12}$$

The least squares estimators $\hat{\underset{\sim}{\beta}}$ has been adapted under these assumption; But as we have discussed before these assumptions are sometimes too ideal to make in practical scenario. If the error term follows normal distributions, which is a continuous distribution, then that automatically restricts the response variable to be a continuous distribution. In that way, classical setup fails when it deals with binary data (Bernoulli distribution) or count data (Poisson distribution). Nelder and Wedderburn (1972) introduced the concept of GLM which is able to address these limitations and shows how linearity could be exploited to unify apparently diverse statistical techniques (McCullagh, 1989). Most important feature of GLM is that the regression function, i.e. the expectation μ of $\boldsymbol{Y}$, the response variable, is a monotonous function of $\eta = \mathbf{X}'\underset{\sim}{\beta}$. We will denote this function that relates μ and η by $\boldsymbol{G}$, which is called the link function, and usually we use the relation $G(\mu) = \eta$ (Hardle et al, 2004). In case of normal distribution, i.e. when we use the classical setup of the multiple linear regression model, $\boldsymbol{G}$ is taken to be identity function (Hardle et al, 2004). In GLM we consider that the response variable can assume any distribution in the family of the exponential distributions. It will allow $\boldsymbol{Y}$ to be a non-negative, even discrete distribution. The link function will be different for different distributions and it will relate the systematic component with the random component. In that way the link function plays a very important role in GLM set up.

So far we have seen, that GLM comprises of a wide range of underlying distributions for different models of which classical linear model is a special case. Further the linear model restrictions of normality, constant variance and additive forms of the covariates. In GLM setup variance is allowed to be function of the expected value of the response variable. Also, the additive form of the covariates will be kept only through a transformation. Having said that we need to answer what should be that transformation? Or in other words what should be the choice of the link function?

Let us consider the following relations (Hardle et al, 2004):

$$\eta = \mathbf{X}'\beta, G(\mu) = \eta \qquad (0.13)$$

The link is called *canonical link* in case we have $\eta = \mathbf{X}'\beta = \theta$, where θ is the canonical parameter mentioned in the previous section. With a canonical link it becomes easier to solve a theoretical or practical problem. Now what link function can we select besides the canonical link? There are several specified link functions for different models. For example, a binomial link with a canonical *logit* link is called a logit model and with a Gaussian link is called *probit* model (McCullagh, 1989). A very useful class of link functions is the class of power function popularly known as the Box-Cox transformations. It can be defined for all models for which we have a positive mean. This family of link functions has the form:

$$\eta = \begin{cases} \mu^{\lambda} & \text{if } \lambda \neq 0 \\ \log \mu & \text{if } \lambda = 0 \end{cases}$$

Let us give the main characteristics of some distribution functions in the exponential family (McCullagh, 1989)(Hardle et al, 2004).

Poisson distribution ($P(\mu)$)

Range	$0(1)\infty$
$b(\theta)$	$\exp(\theta)$
$\mu(\theta)$	$\exp(\theta)$
cannonical link	log
variance	μ
$a(\phi)$	1

Gamma distribution ($G(\mu,\nu)$)

Range	$(0,\infty)$
$b(\theta)$	$-\log(-\theta)$
$\mu(\theta)$	$-\frac{1}{\theta}$
cannonical link	reciprocal
variance	μ^2
$a(\phi)$	$\frac{1}{\nu}$

In this paper we will assume that severity distribution follows *Gamma distribution*. *Log-normal distribution* is another popular distribution that can be considered as severity distribution, which unfortunately is not a member of the exponential family.

Once we have a model set up we need to estimate the model parameters given observations or $\boldsymbol{Y}$ values. As we have mentioned in section 1, that least square estimator in classical setup of the multiple linear regression is the maximum likelihood estimator for normally distributed errors, in GLM set up we have $\boldsymbol{Y}$ that follows exponential family of distributions and we can still work on the maximum-likelihood for the GLM. The added advantage is that by using generic set up of exponential family of distributions, we can find the properties of different distribution at the same time. In the rest of this section we will discuss about the parameters estimation procedure in the GLM setup.

In order to develop the estimation technique we need to specify some properties of the density function f clearly. f will be considered as the density function with respect to *Lebesgue measure* in case of continuous distribution and *counting measure* in case of a discrete distribution (Rao, 2002)(Hardle et al, 2004). With these specifications we can write

$$\int_{-\infty}^{\infty} f(y,\theta,\phi)dy = 1 \tag{0.14}$$

With the exchange of differentiation and integration (which is possible for a *nice* function), we have,

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} f(y,\theta,\phi)dy = \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} f(y,\theta,\phi)dy = 0 \tag{0.15}$$

$$= \int_{-\infty}^{\infty} \{\frac{\partial}{\partial\theta}\log f(y,\theta,\phi)\} f(y,\theta,\phi)dy = E\{\frac{\partial}{\partial\theta} l(y,\theta,\phi)\}, \tag{0.16}$$

with the usual notation $\log(f(y,\theta,\phi)) = l(y,\theta,\phi)$, for the *log-likelihood*. The partial derivative of the log-likelihood wrt to θ, is called *score*. For score we have the following relation,

$$E\{\frac{\partial^2}{\partial\theta^2} l(y,\theta,\phi)\} = -E\{\frac{\partial}{\partial\theta} l(y,\theta,\phi)\}^2 \tag{0.17}$$

Now consider the density function as that of the exponential family of distributions, what we have specified at (1.7) and we will have,

$$E\{\frac{\mathbf{Y}-b'(\theta)}{a(\phi)}\}=0 \text{ and also, } E\{\frac{-b''(\theta)}{a(\phi)}\}=-E\{\frac{\mathbf{Y}-b'(\theta)}{a(\phi)}\}^2 \qquad (0.18)$$

Equations at (1.18) finally lead to the following expressions,

$$\begin{aligned} E(\mathbf{Y}) &= \mu = b'(\theta), \\ Var(\mathbf{Y}) &= b''(\theta)a(\phi) \end{aligned} \qquad (0.19)$$

We can readily observe that the expectation of **Y** depends on θ only, where as the variance of the response variable depends on the nuisance parameter as well. Usually we assume that the quantity $a(\phi)$ is the same with all observations. As we have said earlier, the estimated value of the model parameter beta will maximize the likelihood function of the response variable for given values of the effects (Hardle et al, 2004). Let us consider the observations vector, **Y** and their expectation $\boldsymbol{\mu}$ for given values of effects. Under the GLM set up, the log-likelihood of **Y** will be

$$\begin{aligned} l(\mathbf{Y},\mu,\phi) &= \sum_{i=1}^{n} l(Y_i,\theta_i,\phi) \\ \text{also, } \mu_i &= G(x_i^t\beta) \end{aligned} \qquad (0.20)$$

where $\theta_i=\theta(\eta_i)=\theta(x_i^t,\beta)$ and on the right hand side it is the summation of the individual log-likelihood contribution for each observation i. In order to clarify this theoretical description we will give example with $Y_i \sim N(\mu_i,\sigma^2)$, for which we have,

$$l(Y_i,\theta_i,\phi)=\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)-\frac{1}{2\sigma^2}(Y_i-\mu_i)^2 \qquad (0.21)$$

The sample likelihood will look like

$$l(\mathbf{Y},\mu,\sigma) = n\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i-\mu_i)^2 \tag{0.22}$$

Needless to mention, maximizing log-likelihood wrt $\underset{\sim}{\beta}$ under normal **Y** is equivalent to minimizing the least square criterion as the objective function.

We will discuss one more example in this context with a discrete distribution. Let us suppose **Y** follows Bernoulli distribution and has probability mass function as shown in (1.10) and (1.11). Then the individual log-likelihood will look like (Hardle et al, 2004)

$$l(Y_i,\theta_i,\phi) = Y_i\log(\mu_i) + (1-Y_i)\log_i(1-\mu_i) \tag{0.23}$$

That leads to:

$$l(\mathbf{Y},\mu,\phi) = \sum_{i=1}^{n}\{Y_i\log(\mu_i) + (1-Y_i)\log_i(1-\mu_i)\} \tag{0.24}$$

In case we do not know the distribution of **Y**, but the first two moments are known, we can use the quasi-likelihood instead of likelihood. We will only specify the form of the quasi-likelihood which is as follows

$$\begin{aligned} &E(\mathbf{Y}) = \mu, Var(\mathbf{Y}) = a(\phi)V(\mu) \\ &l(y,\theta,\phi) = \frac{1}{a(\phi)}\int_{\mu(\theta)}^{y}\frac{(s-y)}{V(s)}ds \end{aligned} \tag{0.25}$$

This concept was introduced by Wedderburn (1974) and is believed to be a major break-through in the development and applicability of GLM. This concept shows that we need to know how the variance of each observation changes with its mean value but it is not necessary to specify the distribution in its entirety.

We will close this section by discussing an algorithm for estimating $\underset{\sim}{\beta}$.

From the individual likelihood function, $\log(f(y,\theta,\phi)) = l(y,\theta,\phi)$ and for the generic form of the density function of the exponential family of distributions, we have (Anderson et al, 2004)

$$l(\mathbf{Y},\mu,\phi) = \sum_{i=1}^{n}\left\{\frac{Y_i\theta_i - b(\theta_i)}{a(\phi)} + c(Y_i,\phi)\right\} \tag{0.26}$$

In order to maximize (1.26) wrt $\underset{\sim}{\beta}$, we need to take the first order derivative wrt each β_j and set the equations to zero:

$$\frac{\partial l}{\partial \beta_j} = 0, \quad j = 1(1)p \tag{0.27}$$

In case we have explicit expressions for θ_i in terms of β_j, $j = 1(1)p$, we can simply substitute these expressions in the log-likelihood function and complete the differentiation. However, it is simpler to apply the chain rule instead, which gives (Anderson et al, 2004):

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n}\frac{\partial}{\partial \theta_j}\left\{\frac{Y_i\theta_i - b(\theta_i)}{a(\phi)} + c(Y_i,\phi)\right\}.\frac{\partial \theta_i}{\partial \mu_i}.\frac{\partial \mu_i}{\partial \eta_i}.\frac{\partial \eta_i}{\partial \beta_i} \tag{0.28}$$

Now with the relationships:

$$\mu_i = b'(\theta_i) \Rightarrow \frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i) \Rightarrow \frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{b''(\theta_i)}$$

$$\eta_i = g(\mu_i) \Rightarrow \frac{\partial \eta_i}{\partial \mu_i} = g'(\mu_i) \Rightarrow \frac{\partial \mu_i}{\partial \eta_i} = \frac{1}{g'(\mu_i)}$$

$$\eta_i = \sum_{j=1}^{p}\beta_j X_{ij} \Rightarrow \frac{\partial \eta_i}{\partial \beta_j} = X_{ij}$$

(1.28) will take the form:

$$\frac{\partial l}{\partial \beta_i} = \frac{\delta y}{\delta x}\sum_j \frac{(y_j - \mu_j)}{a(\phi)} \cdot \frac{1}{b''(\theta)} \cdot \frac{1}{g'(\mu_j)} \cdot x_{ij}, \qquad i = 1(1)p$$
$$= \sum_j \frac{\omega_j}{V(\mu_j) g'(\sum_{i=1}^{p} \beta_i x_{ji})}(y_j - \mu_j) x_{ij}, \qquad i = 1(1)p \tag{0.29}$$

Finding solution of these equations needs iterative numerical techniques like *Newton-Raphson algorithm* (Anderson et al, 2004).

We will discuss other modeling issues such as *selection of explanatory* variables, *goodness of fit* and *residual analysis* in later chapters in terms of a particular insurance (claim) data.

CHAPTER II
PROPERTY AND CASUALTY LOSS RESERVE

A Property and Casualty insurance company is different than any other financial institutions in terms of its liability for claims, which is the top-most priority for an insurance company. For an insured, the purpose of insurance is to get back a portion (depending on the particular policy) of the loss from the insurer and the insurer is liable for that *financial indemnification*. Measuring the exact financial indemnification is something truly complex and given the fact that claim investigation happens over a short or long period of time, it becomes even more complicated. *Loss Reserve* can be defined as an insurer's liability for claims (and/or future claims). More precisely the term describes the actuarial process of estimating an insurer's liabilities for loss and loss adjustment expenses (Wiser, 2001). Interestingly, there is no single method or procedure that can give a true estimation of loss reserve. An Actuary needs a combination of business sense, judgment and statistical techniques to estimate loss reserve. It is, in fact, very important to estimate with good accuracy as the financial condition and stability of an insurance company is not possible to assess without knowing the loss reserve estimate (Wiser, 2001). Our main topic of discussion in this chapter is *loss reserving procedure for a property-casualty insurance company*. In order to do that in systematic way, we need to define some terms and state some principles. With a common framework we will focus on different techniques, traditional and non-traditional, for loss reserve estimation. In this section we will compare different loss reserving techniques in property and casualty insurance industry. The first sections will clarify different definitions and principles.

2.1 Definitions and Principles of Loss Reserving in P&CI

"*The Statement of Principles Regarding Property and Casualty Loss and Loss Adjustment Expense Reserve*", adopted by the Board of Directors, Casualty Actuarial Society in May, 1998 is an excellent source for knowledge and information about the principles applicable to the loss reserve estimation and the related definitions. By saying loss, we will mean loss combined with loss adjustment expenses. As we have mentioned before there is no single strict formula or technique to find the actual amount of loss reserve; instead an Actuary needs to use different deterministic and stochastic techniques to capture the uncertainty of the loss reserve along with actuarial judgment and regular communication with the underwriters in order to predict an acceptably accurate loss reserve. We will present this section guided by the statement we have mentioned at the very outset of the section.

2.1.1 Definitions

Dates are very important in the analysis of reserves as losses develop over time. There are several stop gaps in the process of an accident, report of the accident to the insurer, it's entry in the data base and final settlement of the claim amount. We also excerpt definitions of some important dates from the aforesaid statement.

Accident Date is the date on which the accident occurs (CAS, 1998).
Report Date is the date on which the accident is reported to the insurer for the first time (CAS, 1998).
Recorded Date is the date on which the data is first recorded in the insurer's statistical database (CAS, 1998).
Accounting Date: Accounting date defines the group of claims for which liability may exist. It is a date specified for statistical and financial reporting purpose (CAS, 1998).
Valuation Date: Transactions made through this date will be included in the database for evaluating the reserve. Valuation date could be prior, identical or after the accounting date (CAS, 1998).

We need to define the following terms.
The Carried Loss Reserve is the carried loss reserve is defined as the amount shown in an already published statement or in an internal statement of financial condition (CAS, 1998).
Indicated Loss Reserve is the amount resulted from the application of a particular loss reserving evaluation procedure is called indicated loss reserve. This amount must be different for different valuation date with a given accounting date (CAS, 1998).
Case Reserve is the sum up value for specific known claims set by the claim adjusters or by formula (Wiser, 2001). Case reserve does not

allow future development. Adjuster's estimates aggregate of the estimates made by claims personnel for individual claims on the basis of facts of particular claims. On the other hand, formula reserves are established for a classified group of claims. This can be applied to the same homogeneous class of claims using the average claim values or representative statistical factors. Incidentally, in this paper we come across to this concept several times.

Development is the changes that happen to the value of the claim amount over time (Wiser, 2001). Incurred development is the difference between the estimated incurred loss costs at two different valuation dates. Paid development is defined to be the observed increase in the amount of claim payments for loss in succeeding valuation dates. There is a provision for future development of known claims for the incurred development of the claims that occurred on or before the accounting date and are still open. They could be both increasing and decreasing in nature. There is also a provision for the reopened claims reserve for future payments on closed claims that reopened for some reasons unseen at the time they were named closed.

We will discuss the concept of *Incurred But Not Reported* (IBNR) claims and reserve now. Usually IBNR consists of two components – one referred as pure IBNR, which are due to the claims that incurred but not reported. Other part is the provision for the claims that incurred as well as reported but not yet recorded, these are claims in transit.

The claim reporting procedure has an important effect on the IBNR; however practically it is not always possible to see these two components separately (Wiser, 2001) (CAS, 1998). The IBNR reserve should give the ultimate value of the IBNR claims with possible development that could happen after reporting.

Two other very important terms are *Allocated Loss Adjustment Expenses (*ALAE) and *Unallocated Loss Adjustment Expenses* (ULAE). In general, ALAE include the expenses such as attorney's fees along with other legal expenses associated with specific claims. ULAE is the adjustment expenses are the remaining claim expenses such as salaries, utilities, rent connected with the claim adjustment function but not readily assignable to specific claims.

2.1.2. Principles

As we have mentioned already, the Casualty Actuarial Society has published a statement of principles regarding P&C loss and loss adjustment expenses in the "*The Statement of Principles Regarding Property and Casualty Loss and Loss Adjustment Expense Reserve*". We excerpt the four principles from this statement:

> "......
>
> 1. An actuarially sound loss reserve for a defined group of claims as of a given valuation date is a provision, based on estimates derived from reasonable assumptions and appropriate actuarial methods for the unpaid amount

required to settle all claims, whether reported or not, for which liability exists on a particular accounting date.

2. An actuarially sound loss adjustment expense reserve for a defined group of claims as of a given valuation date is a provision, based on estimates derived from reasonable assumptions and appropriate actuarial methods, for the unpaid amount required to investigate, defend, and effect the settlement of all claims, whether reported or not, for which loss adjustment expense liability exists on a particular accounting date.

3. The uncertainty inherent in the estimation of required provisions for unpaid losses or loss adjustment expenses implies that a range of reserves can be actuarially sound. The true value of the liability for losses or loss adjustment expenses at any accounting date can be known only when all attendant claims have been settled.

4. The most appropriate reserve within a range of actuarially sound estimates depends on both the relative likelihood of estimates within the range and the financial reporting context in which the reserve will be presented.

Although specific reserve requirements may vary, the same basic principles apply in each context in which the reserves are stated, including statutory balance sheets, statements of opinion on loss reserves, and reports to shareholders or securities regulators.

… …”

> Loss reserve is defined to be a provision for its related liability and it can be treated as a combination of five components which may not be individually quantified. The five components are *case reserve, provision for future development on known claims, reopened claim reserve, provision for claim incurred but not reported, provision for claim in transit – which is incurred and reported but not recorded.*

A loss reserve can be divided into two parts – *known claims* and *unknown claims*.

The first three components are combined to generate the known claims which mean the amounts required for the future payments of claims that have already been reported.

The last two components are combined to create IBNR reserve. In practice often the strict definition of IBNR is violated and future developments of known claims, unreported claims, unrecorded claims and reopened claims are combined together to IBNR.

2.1.3 Considerations

We will close this section by discussing the important considerations required to be taken care of by reserve analysts. In order to present a statistically and actuarially sound reserve analysis it is necessity to understand the trends and changes that may affect the database. In the next section we will discuss exploratory data analysis and these considerations play very important role in that. These changes could occur from several ways: from changes in underwriting, data handling and processing, accounting to changes happen in social, political and legal ground (CAS, 1998). A sound knowledge of policy provisions is also very important factor for a good analysis. Organization of data with respect to time unit is called *data organization* which is a basic requirement. It is done on the basis of the dates that we have defined earlier. Defining IBNR is another important factor in the sense what to include and what not to include in IBNR (Wiser, 2001).

Classification – *Homogeneity*: Loss reserving analysis can be significantly improved by applying clustering techniques on the data base and divide the entire experiences in several subgroups on the basis of similar claim experience and settlement patterns, identical loss distribution. The reasoning behind such clustering is to minimize the distortion and deviation effects present in data (CAS, 1998). In next chapter we will explore this in a good extent. Another related concept is *Credibility* which is the measure of the accuracy of the predictive value attached to a prediction (CAS, 1998). Understandably, credibility is increased by increasing the homogeneous classes in the database. However dividing entire database into classes naturally reduces the number of data point in each class. Further division of those classes into smaller classes reduces the data point again. This may cause problem in prediction (Wiser, 2001). So there should be optimization between the number of homogeneous classes and the data points in each of them. In case the company data base is small, analyst should consider industry data before clustering.

Without data no statistical analysis is possible; so it is a sheer requirement to have proper financial data for a sound reserve analysis. Another very important factor to be considered carefully is the data emergence pattern: usually liability claims has greater delay in record than property claim (CAS, 1998). Delay between the accident date and recording date, however, depends on the line of business (LoB) as well the insurer.

We will now discuss *Settlement Pattern* which is perhaps, the most important concept for the purpose of this paper. Settlement pattern relates to the development occurs between the time of the report of a claim and the time when it is finally settled; precisely speaking the development occurs between reporting date and the settlement date. This development affects the reserving methods in a significant extent. Reserve Method for the LoBs for which claims are settled quickly, are less affected by this, however liability such as bodily injury may take a significant time before the settlement (Wiser, 2001). For the purpose of paper, we will see the possible complexities of settlement patterns in the next section as well as in the next chapter. *Development pattern* is required to be carefully studied as an insurer's claims method affects the way the case reserve develop for a group of claims. Also, any possible changes in claims practice may affect the consistency of historical developments. In case reserves have been established at the present value, the development history should be restated to remove the effect of discounting (For a detail discussion, see page 274, *Loss* Reserving, Wiser, 2001).

As we have mentioned already, the entire reserve analysis is based on data. More data we have, the analysis becomes easier. *Frequency* and *severity* play important roles here. The same amount of loss may be occurred due to few large claims or due to many small claims (Wiser, 2001). Evaluation of low frequency with high severity is challenging and should be done with extensive analysis. Adjustments should be made for never-before high expected severity claims.

Other affecting factors in reserving analysis are: *Aggregate Limits, Salvage, Subrogation and Collateral Sources, Presence of Reinsurance, Generally Accepted Accounting Principles (GAAP), Portfolio transfers, Commutations and Structured Settlements, Pools and Associations, Operational Changes, Changes in Contracts, External Influences, Reasonableness, Standard of Practice* (CAS, 1998). It is Actuary's or Reserving Analyst's duty to select the most appropriate reserving estimation method for a particular line of business or a group of claims. Usually they examine the indicators in more than one method for a group of claims.

In the next section we will discuss the *reserve estimation strategy* based on this definitions, principles and considerations.

2.2 P&CI Reserve Estimation Strategy

The entire reserve estimation strategy can viewed as four steps (Wiser, 2001) namely

- *Exploratory Data Analysis*
- *Application of different estimation techniques*
- *Evaluation of the resulted estimation*
- *Diagnostics – Monitoring results*

In this section, we will thoroughly discuss the above four steps. We will discuss several numerical examples to clarify the concepts and definitions.

Exploratory Data Analysis

Exploration of data is a very important to begin the reserving analysis; it gives the analyst the idea of which loss reserving method will be appropriate and what would be the interpretation of results. Exploration of data starts by finding the trends and changes that can potentially affect the database. As we have discussed in detail or sometimes mentioned, there are so many things that could affect the data. Changes in LoB, geographical territory, policy provision, purchasing reinsurance, reinsurance limits and attachment point, changes in external environment such as legal, political and social changes, economical changes such as increasing or decreasing inflation rate, modifications in companies data management system can be determining factors and should be considered carefully for reserving analysis (Wiser, 2001). Besides these considerations, the analyst should check the factors such as *rate of development*, *smoothness of development*, *presence of large loss*, *volume of data.* Conclusions should be made about *appropriate projection methodologies*, *anomalies in the data*, *questions to ask to the management.* As we have mentioned earlier the organization of data with respect to the time unit is important. Let us take the following *paid loss* data (Wiser, 2001) by the date of loss occurrence. From accounting exhibit the amount of loss paid in the year 2000 can be gathered. The following table shows how the loss is split over the accident years since 1994.

	(in 1,000)
Paid on 2000 Losses	$ 11,346
Paid on 1999 Losses	$ 16,567
Paid on 1998 Losses	$ 19,935
Paid on 1997 Losses	$ 11,956
Paid on 1996 Losses	$ 5,985
Paid on 1995 Losses	$ 3,211
Paid on 1994 Losses	$ 2,274
Total paid loss in 2000	$ 71,274

In the same way we could have a similar table for the year 1999, where it is known from accounting department that the total loss payment in 1999 is $ 73,972,000. The following table (Wiser, 2001) shows the payments made on different calendar years.

	(in 1,000)
Paid on 1999 Losses	$ 17,001
Paid on 1998 Losses	$ 22,343
Paid on 1997 Losses	$ 13,036
Paid on 1996 Losses	$ 9,098
Paid on 1995 Losses	$ 6,235
Paid on 1994 Losses	$ 4,693
Paid on 1993 Losses	$ 1,566
Total paid loss in 1999	$ 73,972

We will introduce a special kind of data organization method for these amounts over loss year or accident year. Construction of loss triangle helps understanding the comparability of loss amount over years. We will show accumulated loss payments (or cumulative) on a given loss year. The following triangle will show how the losses in each year are developed to its ultimate value. Naturally, if the number of development years is more, then the ultimate amount will found to be convergent, that means further development will be negligible. There are many other kind loss triangles we can construct; we will discuss some them in this section.

Cumulative Paid Loss Triangle (Wiser, 2001)

Accident Year	12	24	36	48	60	72	84
1994	$ 22,603	$ 40,064	$ 54,301	$ 64,114	$ 71,257	$ 75,950	$ 78,224
1995	$ 22,054	$ 43,970	$ 58,737	$ 71,841	$ 78,076	$ 81,287	
1996	$ 20,166	$ 39,147	$ 51,319	$ 60,417	$ 66,402		
1997	$ 19,297	$ 37,355	$ 50,391	$ 62,347			
1998	$ 20,555	$ 42,898	$ 62,832				
1999	$ 17,001	$ 33,568					
2000	$ 11,346						

Loss triangle based on cumulative incurred loss gives good idea of any possible inadequate case reserves, fluctuation of loss over the years. We will discuss this using the following table (Wiser, 2001):

Cumulative Incurred Loss Triangle

Accident Year	12	24	36	48	60	72	84
1994	$ 58,641	$ 74,804	$ 77,323	$ 77,890	$ 80,728	$ 82,280	$ 82,372
1995	$ 63,732	$ 79,512	$ 83,680	$ 85,366	$ 88,152	$ 87,413	
1996	$ 51,779	$ 68,175	$ 69,802	$ 69,694	$ 70,041		
1997	$ 40,143	$ 67,978	$ 75,144	$ 77,947			
1998	$ 55,665	$ 80,296	$ 87,961				
1999	$ 43,401	$ 57,547					
2000	$ 28,800						

If we compare the losses incurred from the first column of the above table, we can say that there was a low incurred loss in 1997 which developed substantially in 48 months. It could be resulted by inadequate case reserve or delayed loss processing. It will give analyst warning about what to do with the year 2000 losses, which is significantly low. If we compare this table with the previous one of cumulative paid loss triangle, we can see, 1997 low incurred loss is well anticipated by a low payment. But, the drop in incurred loss from 1996 to 1997 is 20%, where as this drop in paid loss is only 5%, on the basis of first hand reporting. In that way we may be interested in compare these changes in incurred losses or paid losses over the years. We introduce incremental incurred loss triangle and incremental paid loss triangle, based on the data on previous two loss triangle tables.

Incremental Incurred Loss Triangle

Accident Year	12	24	36	48	60	72	84
1994	$ 58,641	$ 16,163	$ 2,519	$ 567	$ 2,838	$ 1,552	$ 92
1995	$ 63,732	$ 15,780	$ 4,168	$ 1,686	$ 2,786	-739	
1996	$ 51,779	$ 16,396	$ 1,627	-108	$ 347		
1997	$ 40,143	$ 27,835	$ 7,166	$ 2,803			
1998	$ 55,665	$ 24,631	$ 7,665				
1999	$ 43,401	$ 14,146					
2000	$ 28,800						

From this table we can have an idea of increment in incurred loss in two successive periods. Interesting to note, the amount for 1997 during the period 12 months and 24 months, this increment is $27,835,000. 1998 has a similar comparably large annual development. These two facts automatically lead to the suspicion that there should be a processing delay in the company at the year end 1998. As we have discussed in beginning of this section, the analyst should consult the managements and underwriting team to find the reason of this anomaly (Wiser, 2001).

We can look at the incremental paid loss triangle in a same way. The following triangle shows the increment is paid loss in every successive 12-month period. We will notice that the payments during the second annual development period are close to the amount paid in the first annual development period.

Incremental Paid Loss Triangle (Wiser, 2001)

Accident Year	12	24	36	48	60	72	84
1994	$22,603	$17,461	$14,237	$9,813	$7,143	$4,693	$2,274
1995	$22,054	$21,916	$14,767	$13,104	$6,235	$3,211	
1996	$20,166	$18,981	$12,172	$9,098	$5,985		

1997	$19,297	$18,058	$13,036	$11,956			
1998	$20,555	$22,343	$19,934				
1999	$17,001	$16,567					
2000	$11,346						

In this way we can have several kind of loss triangles based on the different type of data. Before closing the discussion on exploratory data analysis, we will discuss few of them.

In order to understand inadequacy in case reserve, loss triangle based on the paid loss as a percent of incurred loss is very helpful to observe (Wiser, 2001). In this triangle the paid loss is divided by the corresponding reported loss for each development age. With the same data we have used so far we have the following triangle.

Loss Triangle Based on Paid Loss as a Percent of Incurred Loss

Accident Year	12	24	36	48	60	72	84
1994	38.54%	53.56%	70.23%	82.31%	88.27%	92.31%	94.96%
1995	34.60%	55.30%	70.19%	84.16%	88.57%	92.99%	
1996	38.95%	57.42%	73.52%	86.69%	94.80%		
1997	48.07%	54.95%	67.06%	79.99%			
1998	36.93%	53.42%	71.43%				
1999	39.17%	58.33%					
2000	39.40%						

The 1997 accident year has a high ratio at the first development year which indicated an anomaly with the other historical data.

Besides severity, claim counts are also very important factor for the reserving analysis. The following triangle shows the historical data for all reported claim counts by development period.

Reported Claim Counts Triangle (Wiser, 2001)

Accident Year	12	24	36	48	60	72	84
1994	32751	41201	41618	41755	41773	41774	41774
1995	33736	39528	39926	40044	40072	40072	
1996	27067	32740	33084	33183	33209		
1997	24928	29796	30074	30169			
1998	25229	31930	32281				
1999	17632	21801					
2000	15609						

From the above triangle we can conclude that for this particular book of business, all claims are found to be reported in a 24 months period. Another important observation is the decreasing number of claims from the year 1994 through the year 2000, which is the lowest in the year 2000. An analyst should fix the changes required to be made in order to anticipate this change. The triangle of similar kind can be formed on the basis of closed with-payment claim counts, closed without-payment counts. In case there is any unusual pattern in these statistics, analyst can make investigate further. Often it is crucial to observe the triangle based on closed claim as a percentage of reported claims and the triangle based on the open claims. These two triangles give excellent idea of trends and patterns in claim settlements for a particular line of business. If there is a large deviation on the number of open claims in two consecutive years or in two consecutive development years, the components of the business are required to be examined (Wiser, 2001). A high number of open claims could be a real challenge for estimating the reserve.

Another type of loss triangle can be formed using the average amount reserved on open claims and changes in average open claim. These kind of triangles tell the analyst, whether there is a movement of case reserves with reasonable inflationary increases. The following two triangles give some idea.

Average Open Claim Amount Triangle (Wiser, 2001)

Accident Year	12	24	36	48	60	72	84
1994	$ 5,339	$ 11,671	$ 16,499	$ 21,029	$ 28,782	$ 47,240	$ 60,722
1995	$ 5,254	$ 13,137	$ 19,405	$ 22,285	$ 29,820	$ 35,209	
1996	$ 5,894	$ 13,334	$ 17,939	$ 16,832	$ 14,722		
1997	$ 3,501	$ 14,190	$ 21,798	$ 28,896			
1998	$ 6,258	$ 13,941	$ 22,026				
1999	$ 8,206	$ 14,324					
2000	$ 4,629						

Changes in Average Open Claim

Accident Year	12	24	36	48	60	72	84
1994-1995	-1.59%	12.56%	17.61%	5.97%	3.61%	-25.47%	
1995-1996	12.18%	1.50%	-7.55%	-24.47%	-50.63%		
1996-1997	-40.60%	6.42%	21.51%	71.67%			
1997-1998	78.75%	-1.75%	1.05%				
1998-1999	31.13%	2.75%					
1999-2000	-43.59%						

In the same way, analyst should check the triangles based on average closed claims payments and triangle based on the changes in two consecutive accident years and consecutive development year. These triangles allow the analyst to decide whether the claim reserve is consistent with the inflationary increase in settlements.

Closed claims as a percent of open claims form triangle that gives the ratio of claims closed in the period to claims open at the beginning of the period. It is another important measure of the financial condition and liabilities of the claim department.

We will close our discussion about the data exploration at this point and move to the next step of reserving technique called loss reserve estimation procedures.

Loss Reserve Estimation Procedures

The entire purpose of the data exploration is to find out the suitable reserving techniques that will capture the trends and changing patterns of the historical data and will help to predict a good estimate of the reserve. Based on the triangles we have discussed previously, we will discuss different method of projection of reserves of the ultimate values. An important point need to remember that all this techniques are nothing but a tool for the projections and must be supplemented by the experience and business knowledge of the actuary and the reserve analyst. We will discuss three different methods for estimate the ultimate loss:

- Expected Loss Ratio Method
- Chain-Ladder Method
- Bornhuetter-Ferguson Method.

A very basic and simple method for ultimate loss estimation is the Expected Loss Ratio Method. Main advantages of this method are its simplicity in calculations and its applicability to a new line of business or a business with comparably less amount data (*Loss Development*

Model, Jim Sohenfelt, 2004). We describe the method using a simple example.

	(a)	(b)	(c)=(a)*(b)
Accident Years	Earned Premium	ELR	Ultimate Developed Loss
2000	$ 50,000	65%	32500
2001	$ 75,000	70%	52500
2002	$ 95,000	67%	63650
2003	$ 120,000	75%	90000

Now the basic question remained is how to determine the *ELR*? In most of cases it is determined from previous experiences, industry data for similar line of businesses and the selection need to be done carefully. Dependency to a subjective factor like *ELR* is a big disadvantage of this method. Also this method suggests no form solution in case *ELR* seems unreliable with the actual loss developments.

The most widely used method for loss reserving is *Chain-Ladder Method*. The chain-ladder method was emerged as a deterministic method and we will discuss that deterministic model. However several developments have been taken place on the stochastic version of the chain-ladder method. One good feature of this method is it intuitively captures the growth of the data flow. The data as organized earlier are considered to project the undeveloped loss year to its expected ultimate level. We need to assume that the data will be developed as previous years. We will mainly discuss the triangular method for paid loss development and the method is similar for the other triangles.

Let us consider the paid loss development triangle we have shown earlier.

Accident Year	12	24	36	48	60	72	84
1994	$ 22,603	$ 40,064	$ 54,301	$ 64,114	$ 71,257	$ 75,950	$ 78,224
1995	$ 22,054	$ 43,970	$ 58,737	$ 71,841	$ 78,076	$ 81,287	
1996	$ 20,166	$ 39,147	$ 51,319	$ 60,417	$ 66,402		
1997	$ 19,297	$ 37,355	$ 50,391	$ 62,347			
1998	$ 20,555	$ 42,898	$ 62,832				
1999	$ 17,001	$ 33,568					
2000	$ 11,346						

As we have mentioned earlier, one strong prerequisite of this kind of analysis is uniformity or homogeneity of the data over the years. The above triangle indicates significant dip in the recent years. So the analyst should investigate the reason behind this and need to take the right measure in order to bring uniformity. At the same time if we can some how normalize this fluctuations in total loss paid, then we can have a credible estimate. In the following table we have given that idea of *development factors* based on each accident year. The method is popularly known as *Chain-Ladder* Method. This will, in some extent, remove the volume difference effect over the years (Wiser, 2001).

Accident Year	12	24	36	48	60	72	84
1994	1.000	1.773	1.355	1.181	1.111	1.066	1.030
1995	1.000	1.994	1.336	1.223	1.087	1.041	
1996	1.000	1.941	1.311	1.177	1.099		
1997	1.000	1.936	1.349	1.237			
1998	1.000	2.087	1.465				
1999	1.000	1.974					
2000	1.000						

This triangle gives the development of each accident year. From the above triangle we have the complete development for the year 1994. But from the year 1995 onwards we have one less development factor for each year. Another thing to notice is the in the discount factors the fluctuations are not significant or in other words development factors are more uniform compare to the volume of the paid loss. We will now try to find a way to estimate the incomplete development factors e.g. for the 2000 the first development factor (12 months to 24 months). A common way is to consider different kind of average of the age-to-age factors and combine it with business knowledge of the actuary to finally settle into a development factor. We will one such example based on the example we have started in the following tables.

Accident Year	12	24	36	48	60	72	84
1994	1.000	1.773	1.355	1.181	1.111	1.066	1.030
1995	1.000	1.994	1.336	1.223	1.087	1.041	
1996	1.000	1.941	1.311	1.177	1.099		
1997	1.000	1.936	1.349	1.237			
1998	1.000	2.087	1.465				
1999	1.000	1.974					
2000	1.000						
Average		1.9508	1.3632	1.2046	1.0991	1.0535	1.0299
Avg last 3		1.9991	1.3749	1.2125	1.0991		
Avg last 4		1.9846	1.3651	1.2046			
Weighted Avg		1.9480	1.3640	1.2050	1.0990	1.0530	1.0300
Harmonic Mean		1.949	1.362	1.204	1.099	1.053	1.03

We define *weighted Average* as average of the development factors with weighted by the amount of incurred loss. *Harmonic Mean* is defined as the n^{th} root of n historical factors.

The way we define the above averages, they are combination of stability (i.e. using the entire data) and responsiveness (i.e. capturing more recent development). In the following table we will give a selection that the analyst has made in this particular case for the loss development factors keeping all judgments in mind.

Accident Year	12/24	24/36	36/48	48/60	60/72	72/84	84/ultimate
1994	1.7725	1.3554	1.1807	1.1114	1.0659	1.0299	1.053
1995	1.9937	1.3358	1.2231	1.0868	1.0411	1.03	
1996	1.9412	1.3109	1.1773	1.0991	1.06		
1997	1.9358	1.349	1.2373	1.1			
1998	2.087	1.4647	1.21				
1999	1.9745	1.35					
2000	1.96						

With the above choice of the development factors for the first incomplete development, the next job is to complete the entire range of the incomplete cells i.e. to forecast the development for each subsequent year. The following triangle shows these developments.

Accident Year	12/24	24/36	36/48	48/60	60/72	72/84	84/ult	DevToUlt
1994	1.7725	1.3554	1.1807	1.1114	1.0659	1.0299	1.053	1.053
1995	1.9937	1.3358	1.2231	1.0868	1.0411	1.03	1.053	1.085
1996	1.9412	1.3109	1.1773	1.0991	1.06	1.03	1.053	1.150
1997	1.9358	1.349	1.2373	1.1	1.06	1.03	1.053	1.265
1998	2.087	1.4647	1.21	1.1	1.06	1.03	1.053	1.530
1999	1.9745	1.35	1.21	1.1	1.06	1.03	1.053	2.066
2000	1.96	1.35	1.21	1.1	1.06	1.03	1.053	4.049

A detailed explanation of these development factors for different years can be found in the "*Loss Reserving*" chapter by *Ronald F Wiser*. There are several advantages and disadvantages of the chain-ladder method.

Advantages

- It is more objective; this method does not involve components like *expected loss ratio* or ELR and needs less actuarial judgment.

- As losses develop with the time, the estimate gets closer to the actual realization.

Disadvantages

- A logical flaw in the chain ladder method is, if there is no loss paid yet for a accident year, then the method will predict the ultimate loss for that year as zero.
- The method can be significantly misleading for a particular LoB or nature of an insurance company. If the payment of losses distributed mostly in the early years of the entire loss payment span, then this method will predict a very high ultimate loss. On the hand if loss payment is less in the early years then chain ladder will give an underestimation.
- As we have constructed the final triangle and filled up the incomplete cells, it has been assumed that the relationships between losses over different development periods is multiplicative, which need not to be true necessarily.

Third method namely *Bornhuetter-Ferguson (BF) Method* is kind of combination of the above two methods. More precisely the advantages of the ELR Method and Chain-Ladder Method are combined into this method (Shoenfelt, 2004). As we have mentioned before, the chain-ladder method is pretty much dependent on historical data. In that sense, this method sometimes produces inappropriate and unreliable results for a new LoB with small amount of historical data or a book of business that experiences occasional large losses. The chain-ladder method is even inappropriate for the business where losses are reported over a long period of time with mere amount of losses reported in first few development years (excess insurance and reinsurance). In order to project the ultimate losses of such businesses we need to have a method that addresses the stability as well as responsiveness. The BF method estimates ultimate loss by considering the sum of the actual reported loss and expected future incurred development. Expected future incurred development is dependent on the expected losses as well as selected loss development factors. The basic idea of BF Method can be described as (Shoenfelt, 2004):

Developed Loss = what is actually paid (incurred) + what we need to pay for a given ELR

The following example will clarify the BF Method:

	BF Method using Incurred Loss Development	Accident Year 2002
(a)	Earned Premium	$ 25,000,000
(b)	Expected Loss Ratio	73%
(c)=(a)x(b)	Expected Ultimate Loss	$ 18,250,000
(d)	Cumulative Loss Devel. Factor / Incur. Loss Devel.	1.214
(e)= 1 - 1/(d)	Unreported %	17.63%
(f)	Actual Incurred Loss as of Dec 31st 2002	$ 13,000,000
(f) + (c) x (e)	Estimated Ultimate Loss	16,217,051.07

The fascinating part of the BF estimated ultimate loss formula is that the expected loss ratio is less important as the experience develops and experience is more crucial as we walk down the line. However, a disadvantage remains as the method is still affected by the changes in claim practices such as speed of claim payments.

Evaluation of the resulted estimation

As we have clearly mentioned at the beginning that there is no single formula or technique to predict the ultimate loss or estimate the reserve; rather several techniques are employed and each of them provides a different estimates. An actuary needs to decide which one to believe or whether he needs to take the average or weighted average of these estimates. This selection, at least for the traditional methods, depends heavily on the Actuary's judgment and experience. We will draw another example from "*Loss Reserving*" chapter by Ronal F Wiser.

Estimated Ultimate Losses by Accident Years and Methods								
Year	Paid Dvlmnt	Incrd Dvlmnt	Avg Paid	Avg Incurred	BF Mathod	Rsrv Dvlmnt	Avg	Selected
1995	82370	83452	79092	82454	83188	82676	82205	83452
1996	88657	88323	84686	87989	88310	87643	87601	88323
1997	76888	70323	71454	69982	70821	70232	71617	70323
1998	75643	80573	74565	79456	80253	81563	78676	80573
1999	102342	93542	91399	92399	92056	93542	94213	93542
2000	69344	66345	65212	65898	66286	66328	66569	66345
2001	45938	44763	53757	53893	44821	46001	48196	44821
Total	541182	527321	520165	532071	525735	527985	529077	527379

In the example the authors showed that the analyst would choose the incurred loss development for all the years except the most recent. For the most recent year BF method has been chosen as incurred loss data is not matured yet. As a next step of the evaluation process the Actuary or the analyst needs to conduct some more tests and observe some more results. For an example they can compare the available

ELR with the projected ELR (i.e. ELR calculated using the projected loss and given earned premium). A significant disparity for any particular year seeks attention of the Actuary. Other measures like incurred loss as a percentage of ultimate loss, case reserve as a percentage of ultimate loss can be considered to verify the consistency in the estimation.

Diagnostics – Monitoring results

Once the Actuary selected an estimate from several possibilities, his immediate job is to monitoring his prediction. Depending on the book of business and company policies, the data is available in monthly basis, quarterly basis or yearly basis. The Actuary can compare his predicted result with the actual developments. In case there is significant abnormality and disparity between the actual initial development and the predicted result, an investigation is required. However, there is no statistical method for the diagnostic test of the above deterministic methods. A deterministic method does not have the provision of conducting diagnostic checks and calculating confidence intervals. So a significant amount of research has been conducted to create alternative methods which are equivalent to the above techniques and at the same time stochastic in nature. A publication by A. E. Renshaw and R. J. Verrall (1998) provides an excellent literature on this topic and this paper shows such equivalent stochastic model for the chain-ladder method. As discussed in this paper, we can see a stochastic claim reserving model as a combination of three components:

- Specification of a parameterized model structure
- A means to utilize the available loss data to fit into the model along with the diagnostic tests
- Finally, a mean to put back the estimated result to the loss triangle (precisely speaking, to the incomplete part of the run-off triangle).

We will close our discussion about loss reserving techniques and in the next chapter we will focus on the main question of this work. Using a published data analysis we will discuss how GLM could be used to predict the ultimate claim settlement value and hence can be used as a tool for loss reserving techniques. We will see some fascinating mathematical results, diagnostic tests of the model (goodness of fit) and discuss a case-study.