CHAPTER 4

DATA BASE AND RESEARCH METHODOLOGY

The present study is an attempt to examine the survival of initial public offerings in India. The research design covers in detail the objectives, period and sample used in the study. It also provides the statistical techniques used to analyze the data along with the limitations of the study.

4.1 OBJECTIVES OF THE STUDY

The study has three objectives:

1. To examine the survival pattern of initial public offerings in India.
2. To examine the determinants of post-listing status of initial public offerings in India
3. To examine the determinants of survival duration of initial public offerings in India.

4.2 PERIOD OF THE STUDY

The study analyses the survival profile of IPOs in India that got listed on BSE (Bombay Stock Exchange) from January 1992 to December 2006. In 1992, the Controller of Capital Issues (CCI) was replaced by Securities and Exchange Board of India (SEBI) and a free pricing mechanism was introduced which is a major reform in Indian capital market. Further, a boom period in the Indian IPO market was witnessed during 1992-1996 wherein large number of issuers entered and formed part of the market. However, after 1996, the trend in IPO market became more or less bearish. Hence, in order to capture the IPOs during such fluctuations, the study examines the survival profile of IPOs from 1992 to 2006. The sample period is restricted upto 2006 so that each issue can be tracked for next 5 years, i.e. till the end of 2011. Hence, the overall time period of this study is from January 1992 till December 2011.
4.3 UNIVERSITY OF THE STUDY

The universe of the study comprises of all the Indian companies which raised capital for the first time since their inception at par or premium through equity issues and got listed on BSE during the period January 1992 till the end of 2006. The study is restricted to the companies which issue equity share capital only, excluding those which resorted to raise the funds through other instruments such as debt based issues or preference issues. The companies which got listed on the National Stock Exchange (NSE) do not form part of this study. Further, the companies which approached the market for the next time, i.e. through Follow on Public Offers (FPO) have also not been considered in this study.

4.4 SAMPLE OF THE STUDY

The initial data consists of 4018 common stocks issued on BSE from January 1992-December 2010. Out of the total data, certain corrections were made wherein the companies that got merged or amalgamated and the companies with insufficient information were removed. It resulted in 3607 IPOs issued on BSE during the period of 19 years. In order to track each firm for five years, the sample has been restricted till 2006. Finally, the study has been done on 3374 IPOs that got listed on BSE from January 1992 to December 2006 and they are analyzed till the end of 2011.

The sample selection is based upon the following criteria:

a) The firms that got listed on the Bombay Stock Exchange (BSE) have been included.

b) The initial public offering of common stock (equity) has been taken.

c) The issue has been made in the post-SEBI period and offered under free pricing era.

d) Share price data of each firm from the date of its listing till the end of 2011 is available.

e) Data on variables, such as issue size, subscription, lead manager, NIC code and year of incorporation, are available.
The BSE sensitivity index has been taken as representative for the sample IPOs. For Sensex, daily four prices are available, i.e. open, high, low and close. In order to compute the market returns and IPO returns, closing values of the index and closing values of the IPO firms have been taken respectively.

4.5 SOURCES OF DATA

Data for the variables, namely, issue date, issue size, issue price, times subscribed and lead manager’s names have been compiled from Prime database (commercial agency for monitoring and compilation of information on all primary public issues in India) and Capitaline database (maintained by Capital Market Publishers India Pvt. Ltd.). The incorporation year of each IPO and National Industrial Classification (NIC 2008) codes have been obtained from Prowess database (maintained by Centre for Monitoring Indian Economy Pvt. Ltd. (CMIE)). Based upon such NIC codes, IPOs have been classified into 10 major industries, namely Agriculture; Mining; Manufacturing; Construction; Wholesale and Retail; Transport and Storage; Accommodation and Food services; Information and Communication; Finance and insurance; Administration and support. Industries with less than 25 IPOs have been categorized into ‘others’ category.

The share price sheet of each IPO has been extracted from Capita chart database (maintained by Capital Market Publishers India Pvt. Ltd.). In order to manage the huge data, a program in ‘Visual Fox Pro’ has been designed which utilized the share price sheets of 3374 IPOs and extracted the data out of them. This was done to obtain the data on list date, last trading date, list price, and number of trading days for each IPO. Further, this program has been used to compute certain variables such as age of the firm, initial returns, market level, list delay, lead manager’s reputation and risk (standard deviation of the first 30 days) for each IPO. Also, each IPO has been tracked for five years from the date of its listing (for second objective) and coded as ‘1’ for survivors and ‘0’ for non-survivors through the program.

The data for post-listing IPO status, date and reason for delisting has been collected from official websites of BSE and Moneycontrol. Further, Capitaline database and offering prospectuses of issuing companies have been used to incorporate certain corrections in the data.
4.6 MEASUREMENT OF SURVIVORS AND NON-SURVIVORS

For evaluating the survivability of IPOs in India, it is important to correctly measure the survivors and non-survivors. The study defines the survivors as the firms that continue to list on the stock exchange and non-survivors that are delisted from the stock exchange due to liquidation, permanent suspension, compulsion by SEBI or any other reason except due to its merger or movement to another stock exchange. Although the fall in returns can also be taken as a failure, but as per Hensler et al. (1997) an upside potential for the stock price exists as long as it continues to be listed. Hence, the study takes delisting as a measure of failure and continuity of listing as that of success. This definition excludes merged or amalgamated IPOs from the sample which is consistent with Hensler et al., 1997, Boubakri et al., 2005; Bhattacharya et al., 2011 whereas the inclusion of suspended firms is consistent with Rath (2008) who supports that suspension from trading clearly foreshadows the company being delisted in the future.

4.7 STATISTICAL TECHNIQUES

This section explains the various statistical techniques used for the purpose of data analysis and testing of hypotheses. In order to achieve the three objectives, different techniques have been used. The data analysis has been done using Independent sample t test, Mann-Whitney Wilcoxon test, Chi-square test, Life table, Kaplan-Meier estimation, Log-Rank test, Logistic regression, ROC curve and Parametric survival model i.e. Log-logistic Accelerated Failure time model.

4.7.1 Descriptive Statistics

In order to empirically evaluate the survival profile of IPOs, it is important to examine their summary statistic. For this, certain moments have been calculated, i.e. Mean, Median, Standard deviation, Skewness and Kurtosis.

Common measure of central tendency, i.e. arithmetic mean gives the average value of entire data, whereas median gives the mid value. The spread of the data is given by standard deviation which shows that how close the data is to the average value. Skewness is the measure of deviation of data from the symmetry. A distribution is said to be
perfectly skewed if it looks similar to the left as well as right side from the center point. If it is less than zero, then the distribution is negatively skewed whereas greater than zero shows that the distribution is positively skewed. On the other hand, Kurtosis is a degree of peakedness in a distribution. Kurtosis equals to 3 is for normal distribution, i.e. mesokurtic, less than 3 is platykurtic and greater than 3 is leptokurtic distribution.

4.7.2 Chi Square test

In order to examine the survival pattern of IPOs it is important to know whether such pattern of survival and non-survival is significantly associated with the categories of independent variables or not. This association and its significance are tested through ‘Chi square test of independence’. The hypothesis under this test is:

H₀: IPO survival and covariate are independent.
H₁: IPO survival and covariate are dependent.

The Chi square test statistic is computed as follows:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Here, O= Observed frequency in each of the response categories
E = Expected frequency in each of the response categories

Hence, in order to ascertain whether the occurrence of survival and non-survival is associated with issue, market and company specific variables, this test would be very fruitful.

4.7.3 Independent Sample t test

To make the survivorship comparison of IPOs, ‘Independent sample t test’ has been used. This test compares the mean values of two groups and gives the indication of whether there exists any significant difference between the groups or not.
Under this, following hypothesis is tested:

H$_0$: There is no significant difference between the means of two groups.

H$_1$: There is a significant difference between the means of two groups.

The t statistic is computed as follows:

\[
    t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S^2_{x_1}}{n_1} + \frac{S^2_{x_2}}{n_2}}} \sqrt{n_1 - 1 + n_2 - 1}
\]

Hence, to compare the survivor and non-survivor across the issue, market and company specific variables this test is useful.

### 4.7.4 Mann-Whitney Wilcoxon Test

In order to overcome the underlying assumption of normality in parametric test, non-parametric tests are useful. Mann-Whitney (also known as Wilcoxon Rank-Sum Test or Mann-Whitney Wilcoxon (MWW) test) is a non-parametric test which tests the null hypothesis that two populations are same against the alternate that they are different, i.e. one population has larger values than another population. This technique is more efficient than the t test when the distribution is not found to be normal\(^1\). It ranks all the observations from both the groups and then sums the rank of one of the groups which is compared with the expected sum of ranks.

The Mann Whitney U statistic is defined as follows\(^2\):

\[
    U = n_1n_2 + \frac{n_2(n_2+1)}{2} - \sum_{i=1}^{n_1} R_i
\]

Here, the samples of size $n_1$ and $n_2$ are pooled and $R_i$ are the ranks. $U$ simply indicates the number of times observations in one sample precede observations in the other sample in ranking.

\(^1\) [http://en.wikipedia.org/wiki/Mann%E2%80%93Whitney_U_test](http://en.wikipedia.org/wiki/Mann%E2%80%93Whitney_U_test)

\(^2\) [http://www.statsdirect.com/help/default.htm#nonparametric_methods/mann_whitney.htm](http://www.statsdirect.com/help/default.htm#nonparametric_methods/mann_whitney.htm)
4.7.5 Logistic Regression

Logistic regression is a class of discrete choice models which is widely used in economics, social sciences and epidemiology (Chou et al., 2007). It is a multiple regression model where the dependent variable is categorical and independent variables are continuous or categorical (Field, 2005, p.218). Although logistic regression is very similar to linear regression but when the dependent variable is categorical, the linear regression is ill suited. The reason is that one of the assumptions of linear regression is the relationship between dependent and independent variable is linear, but when the dependent variable is categorical this assumption is violated. In order to solve this problem, the data can be transformed into logarithmic form which is done in ‘Logistic Regression’ (Field, 2005, p.220). The aim of this model is to assess how well the set of independent variables predicts the occurrence of the categorical dependent variable. Further, it also determines the effect size of each of the covariates on the dependent variable through ‘Odds ratio’.

Unlike OLS regression, logistic regression does not assume linearity of the relationship between independent variables and the dependent, does not require normal distributed variables, does not assume homoscedasticity and in general has less stringent requirements. The method used for estimation in logistic regression is ‘Maximum Likelihood’. This method yields the values of parameters that maximize the probability of obtaining the observed set of data which is done through ‘Likelihood function’. Hence the resultant values obtained from these maximum likelihood estimators are those which are closer to the observed data (Hosmer and Lemeshow, 2000, p.8).

In logistic regression, if Pi= probability that Yi=1 (i.e. an IPO continue to list on the exchange), and (1-Pi) = probability that Yi=0 (i.e. an IPO delist from the exchange), then the variable Yi has the following functional form (Gujarati, 2004, p.595):

$$\pi_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n + \epsilon_i)}}$$
Here, $\beta_1$, $\beta_2$, $\beta_3$ … $\beta n$ are the regression coefficients

$X_1$, $X_2$, $X_3$….Xn are the independent variables.

$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdot\cdot\cdot + \beta n X_n = Z$

The equation can also be written as:

$$P (Y) = \frac{1}{1 + e^{-z}} = \frac{e^z}{1 + e^z}$$

This equation is ‘Cumulative Logistic Distribution Function’. Here, $Zi$ ranges from $-\infty$ to $+\infty$, $Pi$ ranges between 0 and 1 and that $Pi$ is nonlinearly related to $Zi$ (i.e., $Xi$). The value of Beta obtained is known as maximum likelihood estimators.

Whereas, the probability of non-occurrence is:

$$1 - P (Y) = \frac{1}{1 + e^z}$$

From above two equations, it can be written as:

$$\frac{P (Y)}{1 - P (Y)} = e^{Zi}$$

This $\frac{P (Y)}{1 - P (Y)}$ is known as ‘Odds Ratio’, i.e. the probability of occurrence of an event to the probability of non-occurrence. Taking the natural log of this equation, logit can be obtained:

$$Li = \ln \left(\frac{P (Y)}{1 - P (Y)}\right) = Zi$$

$Li$ the log of the odds ratio, which is not only linear in $X$, but also linear in the parameters. The log of odds ratio, i.e. $L$ is known as logit, and hence the model is named as logit model.

In logistic regression the observed and predicted values are used to assess the model fit. For this, log-likelihood statistic is calculated by summing the predicted and actual outcomes (Field, 2005, p. 221):

$$Log\text{-}Likelihood = \sum_{i=1}^{n} \left( Y_i \ln (P(Yi)) + (1 - Y_i) \ln (1 - P(Yi)) \right)$$
It is similar to residual sum of square and indicates the unexplained variance in the model. In logistic regression the Log-Likelihood (LL) is computed for the baseline model with no covariates and the full model, i.e. with all covariates and then it calculates the improvement in the model as follows:

\[ \chi^2 = 2 \{LL \text{ (new)} - LL \text{ (baseline)} \} \]

The above equation is multiplied by 2 because it follows the chi square distribution with \((k-n)\) degree of freedom, where \(k\) is the number of parameters in the new model and \(n\) is the number of parameters of the baseline model. The p value of chi square statistic is used to determine whether such improvement is significant or not.

Another measure used to assess the goodness of fit is ‘Hosmer and Lemeshow’ test which is highly useful in the risk prediction model. It tests the hypothesis that whether the observed event rate matches with the expected event rate in sub groups of the data or not. In other words, it measures the lack of fit in a model. The model in which such observed and expected event rates are similar is known as well calibrated model (Hosmer and Lemeshow, 2000). This is computed as:

\[ H = \sum_{g=1}^{G} \frac{(O_g - E_g)^2}{N_g \pi_g (1 - \pi_g)}. \]

Here, \(O_g\), \(E_g\), \(N_g\) and \(\pi_g\) are the observed events, expected events, number of observations and the predicted risk respectively. \(G\) denotes the number of groups.

Once the model fitness has been determined, now it is essential to know how each individual variable contributes to the model. This is done through ‘Wald statistics’ which is similar to the \(t\) statistic in linear regression and follows the Chi square distribution. Wald statistic examines whether the beta coefficient of each independent variable is significantly different from zero. It is computed as follows:

\[ Wald = \frac{\beta^2}{SE^2} \]
Another measure used in logit model is ‘Pseudo R square’. It is pseudo because there is no direct measure like R square in linear regression. In linear regression, the R square measures the percentage of variance explained in the model, but when the dependent variable is dichotomous the variance would depend upon the frequency distribution of that variable. Hence, it is not measured as a percentage of variance explained rather it measures the weak, moderate or small effect size of the model (Garson, 2012b). Various measures of pseudo R square are available such as Cox and Snell’s, Nagelkerke’s and Mc fadden’s. Out of these, Nagelkerke’s pseudo R square is most acceptable.

The Nagelkerke’s pseudo R square is the adjusted Cox and Snell’s R square that adjusts the scale of statistics to capture the full range of 0 to 1. It is measured as follows (Field, 2005, p.223):

\[
\text{Cox and Snell’s R square } (R^2_{CS}) = 1 - e^{\left[-\frac{2(LL(new)) - (LL(baseline))}{n}\right]}
\]

\[
\text{Nagelkerke’s R square } (R^2_N) = \frac{R^2_{CS}}{1 - e^{\left[\frac{2(LL(baseline))}{n}\right]}}
\]

4.7.6 Receiver Operating Characteristic Curve

There are two ways of assessing the model fitness, i.e. Calibration and Discrimination. Calibration is a measure of how well the predicted probabilities matched with the actual. The Hosmer and Lemeshow assess the model calibration and gives the indication of model fit. On the other hand, discrimination is a measure of how well the model discriminates the groups who have the event from those who do not have the event. In order to measure the discrimination ability of logit model, ROC curve or c statistics is used (Cook, 2007).

Basically, ROC methodology has been developed from signal detection theory which is used to determine whether the electronic receiver is correctly discriminating between signal and the noise (Faggari and Reiser, 2002). It is a plot of sensitivity (true positive rate) against one minus specificity (false positive rate). These are closely related to type I and type II error. Sensitivity is the proportion of positive observations that are correctly classified to be positive, whereas specificity is the proportion of observation that has not
faced the event but they are falsely predicted to have the event. The area under this curve is used to interpret the discriminating power of the model. Higher the area under ROC curve, better is the model fit.

### 4.7.7 Survival Analysis

The concept of survival analysis draws its origins from the medical field, but nowadays it is being applied to predict different kind of events such as bank or corporate failure (Wheelock and Wilson, 1995), bond default (Moeller and Molina, 2003), bank rating transition (Louis et al., 2013) etc. Survival analysis, also known as event history analysis, is an umbrella of various statistical tools that focus on the occurrence of an event as well as the timings of event. Such duration till the event occurs is known as ‘Survival time’ whereas occurrence of an event is known as ‘Failure’ (Mills, 2010). The main aim behind survival analysis is not only to understand the pattern of events, but also the factors associated with the timings of such event (Williams, 2008). Hence, it is nothing but a type of regression model with different likelihood estimates where the dependent variable is survival time and independent variables may be fixed or time varying (Mills, 2010).

One of the distinguishing features of this methodology is that it is capable of dealing with the censored data. Censoring means when some information about the survival time of individual is known, but the exact survival time is not available. It may be because of certain reasons such as the individual has not faced the event before the study ends, the individual is lost to follow up in the study period, or individual is withdrawn from the study (Kleinbaum and Klein, 2005, p.5). This censoring in the data can be of different types such as left censored, right censored or interval censored. Usually, the data is right censored wherein the individual has not faced the event till the end of study time period (Clark et al., 2003). In the present study, the IPOs that continue to survive (i.e. listed till the end of 2011) are right censored. Since the logistic regression model is incapable of taking into account censored observations, hence such type of observations can be well captured through survival analysis (Ahmad, 2012).

There are three main functions that form part of survival analysis- survivor function, probability density function and hazard function (Lee and Wang, 2003, p.8). The survival
function, $S(t)$, refers to the probability that an individual will continue to survive from the time of origin till the end of the study period. The survival function is defined as (Kleinbaum and Klein, 2005, p.9):

$$S(t) = \Pr(T > t)$$

From the definition of cumulative distribution function $F(t)$ of $T$, survival function can also be written as (Lee and Wang, 2003, p.8):

$$S(t) = 1 - P(\text{an individual fails before } t) = 1 - F(t)$$

Here, $S(t)$ is the cumulative survival rate

- $T$ is the time until the firm experiences the event (trading months)
- $t$ is the study time period
- $F(t)$ is the cumulative density function $\Pr(T \leq t)$

The probability density function is the unconditional failure rate, which indicates the probability of failure in a small time interval (Lee and Wang, 2003, p.10). In other words, it means the unconditional (not conditioned on covariate) instantaneous (at any given instant $t$) probability of event (failure rate) (Mills, 2010). It is represented as:

$$f(t) = \lim_{\Delta t \to 0} \left( \frac{\text{an individual dying in the interval } (t,t+\Delta t)}{\Delta t} \right) = \lim_{\Delta t \to 0} \left( \frac{t \leq T \leq t+\Delta t}{\Delta t} \right)$$

Whereas, the hazard function, $h(t)$ or hazard rate refers to the instantaneous probability that an event will occur at time $t$, on the condition that the event has not already occurred (Pommet, 2011). The hazard function is defined as (Lee and Wang, 2003, p.11):

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}$$

This shows that hazard function is the conditional probability, i.e. $P(\text{A|B})$. This conditional failure rate means $P$ individual fails in the interval $[t, t + \Delta t]$ on the condition that it has survived up to time $t$, where $\Delta t$ means very small time interval (Pommet, 2011). Further, the limit has been taken which shows the instantaneous potential for
failing at time $t$ per unit time, given the survival up to time $t$ (Kleinbaum and Klein, 2005, p.12). It is also known as an instantaneous failure rate, conditional mortality rate, force of mortality or age specific failure rate. This hazard function is more intuitive to be used in survival analysis as it quantifies the instantaneous risk that company failure will take place at time $t$, given the company already survived to time $t$ (Noor and Iskandar, 2012). Hence, the survival function focuses on the non-occurrence of an event and hazard rate focuses on occurrence of an event within the study period (Clark et al., 2003). Finally, the probability function is the product of survival and hazard function:

$$ f(t) = S(t) \cdot h(t) $$

This relationship exhibits that when any one of the probability density function, survival function or hazard function is known, other two functions can easily be calculated (Kleinbaum and Klein, 2005, p.262).

There are three different models available for conducting survival analysis, i.e. Non-Parametric, Semi-Parametric and Parametric models. Non-parametric models cover ‘Life table’ and ‘Kaplan- Meier estimation’ which are excellent ways of preliminary analysis of survival data and for estimating as well as comparing the survivor function (Mills, 2010). The semi-parametric method for survival analysis is ‘Cox Proportional Hazard model’. This method describes the relation between the hazard rate and covariates which is written as follows (Bradburn et al., 2003a):

$$ h(t) = h_0(t) \cdot X \exp \{b_1x_1 + b_2x_2 + b_3x_3 + \ldots + bpxp\} $$

Here, $h(t)$ is the hazard function which depends upon set of covariates ($x_1$, $x_2$, $x_3$....xp) and its effect is measured by beta coefficients ($b_1$, $b_2$, $b_3$,....bp). The main advantage of using a semi-parametric partial likelihood approach is that there is no need to define the baseline hazard, the density function, or the survival function. However, the cost of using partial likelihood estimation is a certain loss of efficiency, because some important information might be left (Cox and Oakes, 1984; Chou et al., 2007). Further, this model is based upon the assumption that hazard of the two groups should be proportional and cannot cross each other. Hence, the fully parametric model for survival analysis, i.e.
‘Accelerated Failure Time’ model (AFT) has been used for examining the effect of set of covariates on the survival time.

All these survival models used in the study are explained as follows:

- **Life Table**: Life table is one of the oldest techniques for measuring mortality as well as for describing the survival experience of a population (Lee and Wang, 2003, p.87). Basically, it generates the duration distribution of the entire data that is used to understand the terminal behavior of observations. The output of life table is in the form of rows and columns wherein rows are the time intervals and columns include the counts, probabilities and cumulative probability of event of interest that occurred during the time intervals (Garson, 2012a).

- **Kaplan-Meier Estimation**: In 1958, Kaplan and Meier published a paper on how to deal with incomplete observations. They developed a non-parametric method for the same which is known as ‘Kaplan-Meier estimation method’ or ‘Product-limit method’. This method has now become one of the most acceptable ways of conducting survival analysis in the presence of censored data (Rich et al., 2010). Its aim is to estimate the proportion of population that would survive up to a given length of time (Hoon, 2008; Walters, 2009). Kaplan-Meier (KM) procedure provides the survival and the hazard rate of the data from survival and hazard function respectively. The output of KM method is the survival curve and hazard curve. The KM survival curve is a plot of survival probability of observation against time, whereas hazard curve plots the hazard probability against time. This provides the summary of the data which can be used to estimate the median survival time (Walters, 2009; Goel et al., 2010).

The survival probability $S(t_j)$ is calculated as follows (Clark et al., 2003; Hoon, 2008; Espenlaub et al., 2009; Goel et al., 2010):

$$S(t_j) = S(t_{j-1}) \left(1 - \frac{d_j}{n_j}\right)$$
Here, \( S(t_{j-1}) \) is the probability of being survived at \( t_{j-1} \); \( n_j \) is the number of observations survived just before \( t_j \) (number of IPOs listed); \( d_j \) is the number of events at \( t_j \) (number of IPOs delisted).

- **Log-Rank test**: The survival curves of two or more than two groups can be compared non-parametrically with the help of three tests, i.e. Log-Rank, Breslow and Tarone-Ware test. Out of these tests, Log-rank test is most widely used (Clark et al., 2003; Rich et al., 2010). Log rank test, also known as the ‘Mantel Cox test’ is a part of Kaplan-Meier estimates which tests the following hypothesis (Espenlaub et al., 2009):

  \[ H_0: \text{There is no significant difference between the survival curves} \]

  \[ H_1: \text{There is a significant difference between the survival curves} \]

  Log rank test statistic calculates the chi square for the event time for each group and sums the results (Rich et al., 2010; Goel et al., 2010):

  \[
  \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \sim \chi^2
  \]

  The p value provides the statistical significance of the difference between the survival curves of different groups.

- **Parametric Survival Analysis Model**: The analysis has been done using a special type of hazard model, i.e. Accelerated Failure Time (AFT) model. In AFT model, the survival time is assumed to follow some distribution. Such distributions include the Exponential, Weibull, Log-Normal, Log-Logistic and Gamma to represent the survival time data (Kleinbaum and Klein, 2005). The shape of hazard needs to be tested for selecting the best distribution for AFT model.

  AFT model states that the effect of covariates on the survival time increases or decreases by a constant factor, i.e. \( \phi \) which is known as ‘Accelerated Factor’ or ‘Time Ratio’ (Bradburn et al., 2003a).
This model can be written as:

\[ S(t) = S_0(\varphi t) \]

Here, \( S_0 \) is the baseline survivor function,

\[ \varphi = \exp \{ b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_px_p \} \]

If acceleration factor, \( \varphi \), is greater than 1 it shows that the length of survival time accelerates, whereas \( \varphi \) less than 1 means deceleration in survival time. In AFT model, the dependent variable is survival time on which the impact of set of covariates is determined.

The model is written in log-linear as follows (Bradburn et al. 2003a):

\[ \ln(T_j) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_p X_p + \varepsilon \]

Here, \( \ln(T_j) \) is the log of survival time, which is the dependent variable

\( \beta_0 \) is the constant

\( \beta_1, \beta_1, \beta_2, \ldots, \beta_p \) are the coefficient of covariates (\( X_1, X_2, X_3 \ldots X_p \))

\( \varepsilon \) is the residual term

The major advantage of survival analysis model is the ability to handle time-varying covariates as well as censored observations (Chancharat et al., 2008). That is why several researchers have employed different models for conducting survival analysis e.g. Weibull model (Audretsch and Lehmann, 2005), Log-Normal model (Woo et al., 1995; Ahmad, 2012), Log-Logistic (Hensler, et al., 1997; Hamza and Kooli, 2010; Goot et al., 2011; Raju and Prabhudesai, 2012) and Piecewise Exponential model (Yang and Sheu, 2006).

## 4.8 LIMITATIONS OF THE STUDY

The present study is based on the secondary data collected from various sources mentioned earlier. Hence, the conclusions drawn are subject to the correctness of data. Some other limitations of the study are as follows:

1. The study has taken only the equity shares and ignores the other instruments such as preference shares, debt etc.
2. The scope of the study is limited to IPOs that are listed on BSE during 1992 till 2006.

3. In the study, the returns (MAER and Market level) have been calculated using the BSE Sensitive Index (SENSEX). The results would have been different if other indices had been used for the same.

4. Due to non-availability of historic data for financial and corporate governance factors from 1992 till 2000, the number of independent variables taken is limited. The results would have been more conclusive if data for such variables was available.

5. Further, different measures are available for calculating the issue, market and company specific variables. Hence, the results would have been different, if other measures of variables had been applied.

6. During the study period, several policy variations and fluctuations have taken place in the Indian IPO market. However, the study has not covered the impact of such variations on the survival profile of IPOs.