CHAPTER 4

RESEARCH METHODOLOGY

4.1 SAMPLING

The sample for the study was derived from 521 companies that went public between 1997 and 2007. Company level data is compiled from Prowess, a database provided by the Centre for Monitoring the Indian Economy (CMIE) and Prime database, a database on primary market.

4.1.1 Method of Sampling

The study is based on the analysis of entire population of companies that went for IPO between 1997 and 2007. The IPOs before 1997 is not included in the study because of two main reasons: (1) many researchers have documented the instances of “fly-by-night” entrepreneurs who eroded investor wealth during 1992-1996, and (2) during this period SEBI (Securities and Exchange Board of India) introduced more stringent regulations. The methodology required data from one year before IPO to two years after the IPO. Therefore firms that went public after 2007 were also dropped from the sample as for them the data for next two years would not be available.

4.1.2 Data Collection

Our final sample consisted of 306 public companies. Total number of companies dropped during the data collection were 215. 148 companies were dropped because data for variable
like promoter ownership was not readily available before 2001. The reason being Security Exchange Board of India (SEBI) made the disclosure of information related to promoter’s ownership mandatory after 2001 only. 36 companies were dropped because for some firms the data for all the variables were not available. 31 companies were dropped because for some firms the data for variables were not available for all the time windows. The data was collected for all the variables used in the study for all the 306 public firms.

A sample of private companies who were eligible for IPO but did not come with IPO between 1997-2007 was also selected. The number of such firms were 6000 as identified by Prowess database.

4.2 DATA ANALYSIS

The next section discusses the methods used in the present thesis. (See table 4.1 below for a quick summary of statistical methods employed in the thesis)

<table>
<thead>
<tr>
<th>Research Issues</th>
<th>Methods Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Determinants of Going Public Decision</td>
<td>Panel Probit Regression and Wilcoxon Sign Rank Test</td>
</tr>
<tr>
<td>Impact of IPO on Financial Performance</td>
<td>Wilcoxon Sign Rank Test</td>
</tr>
<tr>
<td>Impact of Change in Promoter’s Ownership Around IPO on Post IPO Financial Performance</td>
<td>Panel Data Analysis</td>
</tr>
<tr>
<td>Impact of Underpricing on Post IPO Financial Performance.</td>
<td>Wilcoxon Sign Rank Test</td>
</tr>
<tr>
<td>Impact of Market Timing on Post IPO Financial Performance.</td>
<td>Wilcoxon Sign Rank Test &amp; Ordinary Least Square Regression (OLS)</td>
</tr>
</tbody>
</table>
4.2.1 Methods used

1. Panel probit regression analysis

The focus of this method is on analyzing the effect of independent variables on the probability of going public. The outcome we are interested in understanding is not a continuous variable, but a binary outcome. Therefore we could not use general linear estimation of the parameters. There is one way through which the problem can be addressed and that is to use random effects probit model.

The key assumptions underlying this estimator are:

- \( c_i \) and \( x_{it} \) are independent.

- The \( x_{it} \) are strictly exogenous (this will be necessary for it to be possible to write the likelihood of observing a given series of outcomes as the product of individual likelihoods).

- \( c_i \) has a normal distribution with zero mean and variance \( \sigma^2_c \) (note: homoskedasticity).

- \( y_{i1}, \ldots, y_{iT} \) are independent conditional on \( (x_i; c_i) \) - this rules out serial correlation in \( y_{it} \), conditional on \( (x_i; c_i) \). This assumption enables us to write the likelihood of observing a given series of outcomes as the product of individual likelihoods.

- Clearly these are restrictive assumptions, especially since endogeneity in the explanatory variables is ruled out. The only advantage (which may strike you as rather marginal) over a simple pooled probit model is that the RE model allows for serial correlation in the unobserved factors determining \( y_{it} \), i.e. in \( (c_i + u_{it}) \).
Since our dependent variable is either public or private we were required to have data for both public companies and private companies. Data characteristics of public companies are already discussed in research design section. Data for private companies are taken from the sample of companies who were eligible to go public but not listed till that point of time.

The following probit regression model was formulated based on our hypotheses:

$$\pi(\text{IPO}_t = 1) = f(\beta_1 \text{Size}_{t-1} + \beta_2 \text{Age}_{t-1} + \beta_3 \text{A}_{t-1} + \beta_4 \text{Beta}_{t-1} + \beta_5 \text{SG}_{t-1} + \beta_6 \text{PBDIT}_{t-1}$$

$$+ \beta_7 \text{Disc}_{t-1} + \beta_8 \text{Leverage}_{t-1} + \beta_9 \frac{M}{\text{B}_{t-1}} )$$

(1)

Further, specification (2) and specification (3) are estimated by one way random effect probit regression. Specification (2) is estimated to control industry effects and specification (3) is estimated to control year effects.

$$\pi(\text{IPO}_t = 1) = f(\beta_1 \text{Size}_{t-1} + \beta_2 \text{Age}_{t-1} + \beta_3 \text{A}_{t-1} + \beta_4 \text{Beta}_{t-1} + \beta_5 \text{SG}_{t-1} + \beta_6 \text{PBDIT}_{t-1}$$

$$+ \beta_7 \text{Disc}_{t-1} + \beta_8 \text{Leverage}_{t-1} + \beta_9 \frac{M}{\text{B}_{t-1}} + \beta_{10} \text{Industry})$$

(2)

$$\pi(\text{IPO}_t = 1) = f(\beta_1 \text{Size}_{t-1} + \beta_2 \text{Age}_{t-1} + \beta_3 \text{A}_{t-1} + \beta_4 \text{Beta}_{t-1} + \beta_5 \text{SG}_{t-1} + \beta_6 \text{PBDIT}_{t-1}$$

$$+ \beta_7 \text{Disc}_{t-1} + \beta_8 \text{Leverage}_{t-1} + \beta_9 \frac{M}{\text{B}_{t-1}} + \beta_{10} \text{Time})$$

(3)

The methodology is derived from the study of Pagano et. al. (1998). They were the first to apply probit regression model to find out the factors that can determine a company to do an IPO in Italy. A few other studies like Chun et. al. (2002) and Albornoz and Pope (2004)
also used the similar construct. Our study uses the already developed model by these studies and adds a few Indian specific variables which were not used by them.

The model variables are explained below:

The dependent variable

The nomenclature of dependent variable is ‘IPO,’ a dummy variable, which equals 1 if the company is publicly held and 0 if the company is a private in a particular year. Individual companies are indexed i: for each year t, in the sample. At any time t, the sample includes all companies which are private at that point in time, and the companies which go public (had an IPO) in that year. After a company goes public, that company is dropped from the private sample.

The independent variables

The selection of the independent variables in the model is based on hypotheses. The brief justifications for inclusion of each of the independent variables in the model are as follows:

Size is included in the model as review of literature indicates that bigger size firms (a) tend to attain more liquidity benefits than the smaller size firms; (b) tend to face lower information asymmetry and adverse selection costs than the smaller firms; and (c) have relative ease in bearing the initial and subsequent expenses owing to firm becoming public. Based on above discussion a positive relationship between size and probability of going public is anticipated.
Age is included in the model as review of literature indicates that younger firms are perceived to be riskier than the older firms. Therefore, a negative relationship between age and probability of going public is anticipated.

IA (intangible assets to total assets ratio) is included in the model as review of literature indicates that firms with higher IA are perceived to be riskier than the firms with lower IA. Therefore, a positive relationship between IA and probability of going public is anticipated.

Beta is included in the model as a measure of the systematic risk of the particular firm. Firms with higher beta are riskier and have higher cost of capital than firms with lower beta. The review of literature indicates that firms go public to facilitate the risk diversification by the initial owners and to bring down their cost of capital. Therefore, a positive relationship between the firm’s beta and probability of going public is anticipated. To the best of the authors’ knowledge, this is the first ever study to gauge the influence of firm’s beta on its going public decision.

SG (growth in sales) is included in the model to examine whether firms go public to raise the capital required for their planned growth and expansion. Firms experiencing higher sales growth require more capital than firms experiencing lower sales growth. Therefore, a positive relationship between SG and probability of going public is anticipated.

PBDIT (PBDIT/Total Assets) is included in the model as review of literature indicates that more profitable firms tend to face lower information asymmetry and adverse selection costs than the less profitable firms. Therefore, a positive relationship between PBDIT and probability of going public is anticipated.
Disc (corporate current taxes to sales ratio) is included in the model as a measure of the firm’s level of transparency. Firms with higher ‘Disc’ are more transparent than firms with lower ‘Disc’. The review of literature indicates that firms which are already more transparent should experience lesser loss of confidentiality than those which are relatively less transparent. Therefore, a positive relationship between ‘Disc’ and probability of going public is anticipated.

Leverage (Debt to Equity ratio) is included in the model to measure financial leverage of the firms. Literature review indicates firms do IPO to rebalance their capital structure. Therefore, a positive relationship between Leverage and probability of going public is anticipated.

M/B (Industry market to book ratio) is included in the model to access whether firms take advantage of windows of opportunity or not. Literature review indicates that firm belonging to overvalued industry is more likely to go public. Therefore, a positive relationship between M/B and probability of going public is expected.

2. Wilcoxon Sign Rank Test

This technique is used to analyze the impact of firms’ IPOs on their performance. The impact was analyzed by comparing the post-IPO performance of Indian public firms with their pre-IPO performance using a three-step procedure. First, the firm-wise percentage changes in the performance variables for the following three time windows were calculated: (i) one year before IPO (Y-1) to IPO year (Y+0), (ii) one year before IPO (Y-1) to one year after IPO (Y+1), and (iii) one year before IPO (Y-1) to two year after IPO
It is to be noted that the percentage changes in performance variables are calculated separately for every firm in our sample. Second, the median values of the firm-wise percentage changes in the performance variables are separately calculated and reported for each of the three time windows considered above. Finally, the ‘Wilcoxon two sample signed-rank test’ is used to assess whether the median value of a performance variable in the post-IPO period is significantly different from its value in the pre-IPO period.

The Wilcoxon signed-rank test is a non-parametric statistical hypothesis test used when comparing repeated measurements on a single sample to assess whether their population mean ranks differ (i.e. it is a paired difference test). It can be used as an alternative to the paired Student's t-test, when the population cannot be assumed to be normally distributed.

Assumptions underlying a Wilcoxon test

- Each pair of observations must represent a random sample from a population and must be independent of every other pair of observations.
- The underlying variable of interest is continuous.
- The Distribution of the differences scores is continuous and symmetrical in the population.

Jain and Kini (1994) were the pioneer in applying Wilcoxon Sign-Rank Test to analyse the impact of IPO on operating performance of US firms. A few other studies like Faccio and Lasfer (2000) and Balatbat et. al. (2004) adopted same methodology for UK and Australia.
In the present study already developed construct has been used to analyse the impact of IPO on operating performance of firms for a sample of emerging market firms.

In the present study changes in the variables are analyzed for the following three time windows: (i) one year before IPO (Y-1) to IPO year (Y+0), (ii) one year before IPO (Y-1) to one year after IPO (Y+1), and (iii) one year before IPO (Y-1) to two year after IPO (Y+2). The test for differences between the two groups is performed using the Wilcoxon two sample signed rank test. Following three issues have been analysed through Wilcoxon Sign-Rank Test:

1. IPO determinants are also analysed by examining the changes in following variables from pre IPO to post IPO period:
   - Promoter’s ownership: Promoter’s ownership has been analysed to examine if the ownership declined significantly in post IPO period.
   - Debt to equity: Debt to equity ratio has been analysed to examine if the financial leverage declined significantly in post IPO period.
   - Capital expenditure: This variables has been analysed to examine if the level of capital expenditure increased significantly in post IPO period or not?
   - Investments: The purpose for analysing investments is to examine if the level of investments increased in post IPO period or not?
   - Cost of credit: This variable is used to analyse if the cost of capital declined significantly in post IPO period or not?
   - Beta: Beat has been used as a measure of risk. It has been analysed to examine if the overall risk declined in post IPO period or not?
2. The financial impact of IPO has been studied by examining the trend in following variables:
   - Asset utilization: In order to analyse the asset utilization efficiency this ratio has been examined and compared with the pre IPO ratio.
   - Operating performance: Operating performance measured by operating cash flow to total assets has been analysed to examine if the performance of firms declined in post IPO period or not?

3. The market sentiment and investors’ expectation of the future earnings from the firms is captured by P/E, M/B and EPS in the present study. The trend in the variables are analysed to see if the sentiment and expectation weakened in post IPO period or not?

3. Panel Data Analysis

Panel data analysis has been done to establish the relationship between promoters’ ownership and operating performance of firms around their IPOs. The data taken for the variables have both a cross-sectional unit of observation (the firm) and a temporal reference. Therefore we have utilize econometric model for both time series and cross-section variation in the data. Indeed, the results of the Breusch-Pagan Lagrange Multiplier test clearly indicate that the traditional OLS regression is inappropriate. The evidence of firm and time effects is found by performing F-test (see Table 5.8 & 5.9 in chapter 5). Breusch-Pagan Lagrange Multiplier test and the Hausman test confirm that the suitable model should be the fixed effects model and not the random effect model (see Table 5.8 &
Further Wald test confirm the presence of time effect. Hence the study used both fixed effect model and fixed effect and time effect model.

The use of panel data estimations gave following benefits:

- Panel data suggests that individuals, firms, states or countries are heterogeneous. Time-series and cross-section studies not controlling this heterogeneity run the risk of obtaining biased results.
- Panel data give more informative data, more variability, less collinearity among the variables, more degrees of freedom and more efficiency. Time-series studies are plagued with multicollinearity. This is less likely with a panel since the cross-section dimension adds a lot of variability.
- Panel data are better able to study the dynamics of adjustment. Cross-sectional distributions that look relatively stable hide a multitude of changes.
- Panel data are better able to identify and measure effects that are simply not detectable in pure cross-section or pure time-series data.

Initially following specification was formulated:

\[
\text{Performance}_{i,t} = \alpha + \beta_1 \text{PROM}_{i,t} + \beta_2 \text{CURR}_{i,t} + \beta_3 D/E_{i,t} + \beta_4 \text{GRW}_{i,t} + \beta_5 \text{CAPEX}_{i,t} + w_{i,t}
\]  
(4)

In order to access any nonlinear relationship between ownership and performance following alternative model were formulated:
\[ Performance_{i,t} = \alpha_i + \beta_1 PROM_{i,t} + \beta_2 (PROM)_{i,t}^2 + \beta_3 CURR_{i,t} + \beta_4 D/E_{i,t} + \beta_5 GRW_{i,t} + \beta_6 CAPEX_{i,t} + w_{i,t} \]

\[ Performance_{i,t} = \alpha_i + \beta_1 PROM_{i,t} + \beta_2 (PROM)_{i,t}^2 + \beta_3 (PROM)_{i,t}^3 + \beta_4 CURR_{i,t} + \beta_5 D/E_{i,t} + \beta_6 GRW_{i,t} + \beta_7 CAPEX_{i,t} + w_{i,t} \]

where, \( i = 1, 2, \ldots, 306 \)

\( t = IPO-1, IPO, IPO+1, IPO+2 \)

The methodology is derived from previous studies on similar research issues. Particularly, the present technique relies heavily on the study by Kim et al. (2004). Other studies who adopted the same methodology are Mikkelson et al. (1997), Wang (2005) and Christina (2005).

The dependent variable

The dependent variable is performance variable, which is either operating return on total asset \((PBDIT/TA)\) or cash flow from operating activities divided by total assets \((CF/TA)\). Operating return on total assets is used as a measure of efficiency of assets utilization. Cash flow from operating activities are a primary component in net present value (NPV) calculations used to value a firm.

The independent variables

Promoter ownership \((PROM)\) is the ownership of promoters in percentage. \((PROM)^2\) and \((PROM)^3\) represent the quadratic and cubic forms, respectively.
Controlling variables

Liquidity is a general measure of financial stability and hence included as a determinant of performance. Researchers expect a positive relationship between liquidity and performance. Our study included Current ratio (CURR) to measure the liquidity of firms.

In the present study Debt-Equity ratio (D/E), calculated as ratio of total borrowings and net worth, is taken as controlling variable. Debt creates discipline and contributes to less agency conflicts inside a firm. Hence according to agency theories, the relationship of performance with debt should be positive. Pecking order however argues for a negative relationship. Hence the exact relationship is a matter of debate.

In order to capture firm’s growth, percentage growth in sales in last three years (GRW) is used as a controlling variable. A positive relationship is expected between growth and performance.

Capital expenditure (CAPEX) is included as a proxy for level of investment as a controlling variable. A positive relationship is expected between level of investment and performance (for the calculation and definition of variables see Annexure).

4. Ordinary Least Square Regression (OLS)

To assess whether issues were timed with the market or not following regression model was estimated using ordinary least square method (OLS). The regression equation
examines the relationship between dummy variable HOT and amount of proceeds raised through IPO. Here \( t \) is the IPO year.

\[
Y_t = \alpha + \beta_1 HOT + \beta_2 \frac{M}{B_t} + \beta_3 \frac{PBDIT}{TA_{t-1}} + \beta_4 Size_{t-1} + \beta_5 Intang_{t-1} + \beta_6 \frac{B}{E} + \beta_7 Sgrw_{t-1}\epsilon_t \quad \ldots (7)
\]

The dependent variable

The dependent variable \( Y_t \) is calculated as ratio of total proceeds from IPO and yearend total assets.

Independent variable

The dummy variable \( HOT \) captures the market timing effect, if any, by the firm belonging to hot issue market rather than cold issue market.

Controlling variables

Market-to-book ratio (M/B) ratios has been taken to control the effect of growth opportunities. We expect firms with more growth opportunities to raise more capital.

Second controlling variable is profitability which is measured as operating return on total assets. We expect an opposite relationship here. The logic is more profitability a firm is less will be its requirement of funds form the market.

Third controlling variable included is size. The expectation is bigger a firm is more is the chances that it will raise more fund through IPO.
Forth controlling variable included is intangibility which proxies asset risk of a firm. The expectation is more risky a firm is more is the chance that it will take benefit of favourable market i.e. more proceed will be raised.

Debt-to-equity (D/E) is included to control for capital structure related relationship. We expect companies with higher debt to raise more proceeds at the time of IPO.

Companies with more growth opportunities are likely to raise more capital at the time of IPO. In order to control this effect growth in sales is included as one of the controlling variable. Relationship expected is positive.

The methodology is adopted from Alti (2006) who asserts that favourable market condition should reflect the timing attempt by issuers. In order to quantify the timing effects we have hypothesized that firms that go public when the market is favourable are likely to raise more amount of proceeds than if they had gone public when the market conditions were unfavourable. The same preposition was also used by Alti (2006).

In order to estimate regression model using OLS following assumptions were made:

**Heteroscedasticity:** The problem of heteroscedasticity is more prevalent in cross-sectional data because they (cross sectional data) involve units or groups that are heterogeneous in nature. Heteroscedasticity was suspected in the data set used in this study due to the crosssectional nature of the data. Two diagnostic tests, Breusch-Pagan and Cook-
Weisberg, were employed in order to check for the presence of heteroscedasticity (see Table 5.13 in chapter 5). The presence of heteroscedasticity was confirmed by both tests. The Breusch-Pagan test indicated that heteroscedasticity was significant at the 0.01 level of significance; the Cook-Weisberg test indicated the presence of heteroscedasticity at the 0.10 level of significance. At this juncture, estimation with OLS was rejected, and an alternative estimation techniques capable of correcting for heteroscedastic errors were rendered: the robust robust regression method means OLS method with standard errors corrected for heteroscedasticity by White’s method.

**Multicollinearity:** One of the problems that may arise in regression analysis is multicollinearity. The presence of a high level of multicollinearity makes it difficult to disentangle the separate influences of independent variables on the dependent variable. Technically, the OLS assumption pertaining to the absence of exact linear relationships among some of the independent variables is violated only in the case of exact multicollinearity. Nonetheless, we decided to conduct a number of diagnostic tests to check for the presence of high multicollinearity. These tests include: (1) Analyzing the presence of High R² with few significant t-ratios, (2) Checking for high zero-order correlations between independent variables, (3) Checking for the presence of correlations between estimated regression coefficients, and (4) use of Variance-Inflation Factor (VIF) method. As a result of these diagnostic tests, we conclude that multicollinearity among the predictors in our model is not a concern.
**Serial Correlation:** serial correlation may be defined as “correlation between members of series of observations ordered in time (as in time-series data) or space (as in cross-sectional data)”. The model assumes the absence of serial correlation, meaning the disturbance term relating to any observation is not influenced by the disturbance term relating to any other observation. The OLS estimators in the presence of serial correlation are still linear-unbiased and consistent, but are not efficient, meaning they do not have minimum variance. In other words, in the presence of serial correlation we are likely to declare a coefficient statistically insignificant even though in fact it may be significant. Serial correlation is a common problem for the data involving multiple time periods. For the purpose of this study, a regression model \( u_t = ru_{t-1} + e_t \), where \(-1 < r < 1\) was estimated to see if the value of \( r \), the first-order coefficient of serial correlation, is significant. It was found that the value of \( r \) (-0.023) was not significant (\( t = -0.464 \)) indicating an absence of first-order serial correlation in the data.

For testing normality Jarque-Bera test, is performed. The Jarque-Bera test is a diagnostic of departure from normality, based on the sample kurtosis and skewness. The null hypothesis is that both excess kurtosis and skewness are 0 that is, the data are from a normal distribution.