CHAPTER 3
RESEARCH METHODOLOGY

The current chapter defines the scope of the study that is the research question which describes the aim of the study; the time period of the study; the data collection sources; description of the sample set and details of different statistical tools and techniques used for data analysis. Besides, the limitations of the study have also been detailed at the end of the chapter.

RESEARCH QUESTION

The present study, an application of the Resource Based Theory, attempts to study various aspects related to resources in their relation to competitive advantage in the Indian corporate sector. The study seeks to identify the resources, to establish the interactions that exist among these resources and to explore the deployment aspect of these resources. The past research highlights the various aspects of the Resource Based Theory and their relationship with competitive advantage. However, very few studies are found to explore the various aspects of the Resource Based Theory and their relationship with competitive advantage in the Indian corporate sector. The present study tries to bridge the gaps as it tries to identify the resources, together with the interactions between them that lead to competitive advantage. The study recognizes the importance of each firm’s industrial context and thus, identifies resources for each sector specifically, and brings forth the influence of various industrial characteristics on the relationship among resources and competitive advantage. Further, the study explores the deployment issue by studying the role played by competence and governance in utilizing a firm’s resources and their relationship with competitive advantage.

STUDY PERIOD

To attain the objectives of the study, the data has been collected for seven years ranging from financial year 2004-05 to 2010-11.

SOURCES OF INFORMATION

The data required to carry out the study has been collected from databases, websites and company annual reports. The company level data has been collected using the Prowess
database of the Centre for Monitoring Indian Economy (CMIE). This database has been utilized by a number of researchers in various fields studying the Indian corporate sector (Pattnaik et al., 2011). The database is comprehensive, providing data on a range of variables related to the companies. Additionally, the annual reports of the companies, the official company websites, and the websites: moneycontrol.com and bloomberg.com, have been used to collect data on the various CEO characteristics. The industry level data has been collected from Capitaline Plus database of the Capital Market Publishers India Pvt. Ltd. and the Economic Intelligence Service documents released by the CMIE. The study uses the Thomson Reuters business classification: classification structure 2012 to define the sectors of various industries.

SAMPLE DESIGN

To carry out the study to explore the identification, deployment and interaction of resources creating competitive advantage in India the sample for the study comprises of the companies forming BSE 500 index as on 31st March, 2011. These 500 companies were chosen as the sample for the reasons: 1) these 500 companies account for more than 90 percent of the market capitalization and thus provide with a representative sample, 2) The firms that are better performers in each of the industries form a part of the sample and thus enable to gain understanding of the reasons for their attaining superior performance. 3) There is a higher probability for these 500 companies reporting the data on the variables under study.

DETAILED METHODOLOGY

The following discussion explains in detail the methodology adopted to carry out each objective of the thesis.

Identification of resources

The first section, which identifies the resource investments that lead to competitive advantage in the Indian corporate sector, is divided into two sub-sections. In the first sub-section, the resource investments that lead to competitive advantage are studied extensively taking the entire Indian corporate sector. It is an attempt to gain preliminary understanding of the resources that lead to competitive advantage in the Indian corporate sector, in general. The initial sample comprises of the entire list of companies
forming the BSE 500, as discussed above. However, since a balanced panel seems more appropriate as it provides more reliable estimates than those provided by an unbalanced one (Jimenez-Martin, 1998), for the present analysis only those companies were included in the final sample that report the data for all the seven years. Since the data was not available for all the 500 companies that formed the initial sample, the screening of the data provided with a final balanced panel data-set of 340 companies, with a total of 2380 firm years.

The data is standardized to avoid the adverse impact of the variation in the units used to express the independent and dependent variables (Gujarati, 2004). The data is analyzed to establish the influence of various resource investments on the firm performance. Thus, we identify the resources that influence competitive advantage. Specifically, the influence of investments in research and development, marketing, advertising, human capital and physical capital and the interactions among them is established on the Tobin’s q, by means of panel data random effects generalized least squares regression. The description of this technique has been provided later in this chapter.

In the second sub-section, each business sector is analyzed separately to provide a deeper understanding of the resources that lead to competitive advantage in the Indian corporate sector. The initial sample comprises of the entire list of companies forming the BSE 500 as discussed above. The companies are segregated into 21 business sectors based on the Thomson Reuters business classification: classification structure 2012. After segregation, only those companies are included in each of the business sectors that have the data for all the seven years. Thus, a balanced panel data-set is utilized for each of the business sectors. However, for the final analysis 16 business sectors are used as for the remaining 5 business sectors viz. healthcare, real estate, retailers, telecommunication services and transportation, the final number of companies that accomplished data requirements was too small to be considered for statistical analysis. There were 7 companies which did not fit into any of the business sectors and these were grouped as “miscellaneous”, as shown in Table 3.1 and thus, this set is not considered for the final analysis.
Table 3.1: Distribution of sample across industries and business sectors

<table>
<thead>
<tr>
<th>Sector</th>
<th>Number of Industries</th>
<th>Number of Companies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applied Resources</td>
<td>2</td>
<td>8 (1.6%)</td>
</tr>
<tr>
<td>Automobiles and Auto Parts</td>
<td>6</td>
<td>22 (4.4%)</td>
</tr>
<tr>
<td>Banking and Investment services</td>
<td>5</td>
<td>65 (13%)</td>
</tr>
<tr>
<td>Chemicals</td>
<td>5</td>
<td>22 (4.4%)</td>
</tr>
<tr>
<td>Cyclical Consumer Products</td>
<td>8</td>
<td>44 (8.8%)</td>
</tr>
<tr>
<td>Cyclical Consumer Services</td>
<td>6</td>
<td>23 (4.6%)</td>
</tr>
<tr>
<td>Energy Fossil Fuels</td>
<td>4</td>
<td>23 (4.6%)</td>
</tr>
<tr>
<td>Food and Beverages</td>
<td>5</td>
<td>20 (4%)</td>
</tr>
<tr>
<td>Healthcare</td>
<td>1</td>
<td>2 (0.4%)</td>
</tr>
<tr>
<td>Industrial and Commercial services</td>
<td>4</td>
<td>66 (13.2%)</td>
</tr>
<tr>
<td>Industrial Goods</td>
<td>6</td>
<td>31 (6.2%)</td>
</tr>
<tr>
<td>Mineral Resources</td>
<td>4</td>
<td>28 (5.6%)</td>
</tr>
<tr>
<td>Personal and Household Products</td>
<td>1</td>
<td>11 (2.2%)</td>
</tr>
<tr>
<td>Pharmaceutical and Medical Research</td>
<td>1</td>
<td>32 (6.4%)</td>
</tr>
<tr>
<td>Real Estate</td>
<td>1</td>
<td>3 (0.6%)</td>
</tr>
<tr>
<td>Retailers</td>
<td>1</td>
<td>4 (0.8%)</td>
</tr>
<tr>
<td>Software and Information Technology</td>
<td>3</td>
<td>35 (7%)</td>
</tr>
<tr>
<td>Technology Equipment</td>
<td>5</td>
<td>10 (2%)</td>
</tr>
<tr>
<td>Telecommunication Services</td>
<td>1</td>
<td>8 (1.6%)</td>
</tr>
<tr>
<td>Transportation</td>
<td>2</td>
<td>12 (2.4%)</td>
</tr>
<tr>
<td>Utilities</td>
<td>2</td>
<td>24 (4.8%)</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>4</td>
<td>7 (1.4%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4</strong></td>
<td><strong>500 (100%)</strong></td>
</tr>
</tbody>
</table>

The numbers in parentheses represent the percentage of companies forming the total sample.
This screening of the data reduced the balanced panel data set to 320 companies, which comprises the final sample.

The resources that lead to competitive advantage are identified for each of the business sectors. The influence of investments in research and development, marketing, advertising, human capital and physical capital and the interactions among them is established on the Tobin’s q, in each of the business sectors, by means of panel data random effects generalized least squares regression.

**Industry context and the relationship between resource investment and competitive advantage**

In the second objective, the resource investment decisions which form a key part of a firm strategy are examined within the purview of industrial conditions. The industry conditions that are considered in this chapter are 1) industry dynamism and 2) industry structure features viz. industry concentration, industry growth and industry capital intensity. The data pertaining to the companies is collected through the PROWESS database of the Centre for Monitoring Indian Economy (CMIE) and the annual reports of the companies. The industry-level data is collected through the Capitaline Plus database and the Economic Intelligence Service documents released by the CMIE. Since the data was not available for all the 500 companies that formed the initial sample, the screening of the data provided with a balanced panel data-set of 340 companies.

**Management competence, corporate governance and competitive advantage**

To carry out the third objective, we establish the influence of governance mechanisms as: board of directors and ownership structure together with the managerial competence: CEO characteristics on firm performance.

This chapter investigates the issue of corporate governance of firms in the Indian corporate sector. The initial sample comprises of the BSE 500 companies for a period of 7 years beginning from 2004-05 to 2010-11. As the data for all variables is not available for all the companies, hence those companies are retained for which data for at least 4 out of the 7 years is available. The 4 year cut-off is used as it helps to generate a sample
on which panel data techniques can be applied and efficient estimates can be obtained. After the screening process, the final sample has 326 companies, with 2252 observations.

The influence of managerial competence and governance mechanisms is established on Tobin’s q for the entire set of 326 firms. For further analysis, the companies are grouped into two categories on the basis of concentration of insider ownership. The two kinds of companies are: ones that have dominant insider ownership and the others that do not have dominant insider ownership. Firms with insiders owning greater than 50 percent of shareholding are grouped as the companies are dominated by insiders, and firms with insiders owning less than 50 percent are grouped as the companies that are not dominated by insiders. Of the final sample the firms that have dominant insider ownership comprise of 1182 firm years, and that do not have dominant insider ownership comprise of 1070 firm years.

**Relationship between management competence, corporate governance, resource investments and competitive advantage**

The next chapter explores the relationship between management competence, corporate governance, resource investments and competitive advantage. To analyze the issue the data-set is the same as in the previous section however, the values for all variables are taken as average of seven years. The structural equation modeling technique, which is explained later in this chapter, is used to establish the relationship between management competence, corporate governance, resource investments and competitive advantage.

**STATISTICAL TECHNIQUES**

**Panel data regression analysis**

According to Baltagi (2005) using a panel data-set provides certain benefits: 1) It helps in controlling for individual heterogeneity 2) it provides data that are more informative, have more variability, have less collinearity among the variables, and provide more degrees of freedom and more efficiency, 3) Panel data are better able to study the dynamics of adjustment, 4) Panel data are better able to identify and measure the effects that are simply not detectable in pure cross-section or pure time-series data, 5) Panel
data models allow one to construct and test more complicated behavioral models than purely cross-section or time-series data, and 6) Micro panel data reduces or eliminates the biases resulting from aggregation over firms or individuals gathered on units of analysis, for example individuals, firms and households.

Panel data, also known as longitudinal data, follows a particular set of cross-sectional unit over a period of time (Hsiao, 2003). Thus, both space and time dimensions are present in a panel data set (Gujarati, 2004). In the form of equation, a panel data regression differs from a regular time-series or cross-section regression as it has a double subscript on its variables, that is,

\[ Y_{it} = \alpha + X'_{it}\beta + u_{it} \]

\[ i = 1, \ldots, N ; t = 1, \ldots, T \]

here, \( i \) denotes the cross-section dimension and \( t \) denotes the time-series dimension (Baltagi, 2005). In the present study cross-section dimension is the Indian companies and the time-series dimension is 2004-05 to 2010-11. \( \alpha \) is a scalar, \( \beta \) is \( K \times 1 \) and \( X_{it} \) is the \( it \)th observation on \( K \) explanatory variables. Most of the panel data applications utilize a one-way error component model for the disturbances, with

\[ u_{it} = \mu_i + \nu_{it} \]

where \( \mu_i \) is the unobservable individual-specific effect that is time-invariant and \( \nu_{it} \) is the remainder disturbance and varies with individuals and time. \( \mu_i \) accounts for any individual-specific effect that is not included in the regression; \( \nu_{it} \) is the usual disturbance in the regression (Baltagi, 2005).

The panel data set can be distinguished as: balanced and unbalanced panel data-set. A data-set is called balanced panel when the same time periods are available for all cross section units. While, a data-set is unbalanced when some time periods are missing for some units in the population of interest (Wooldridge, 2002). The present research, both balanced and unbalanced panel data-sets have been utilized. To carry out objectives 1 and 2 the data sets used for analyses are balanced panel data-sets while for objective 3 the data set used for analyses is unbalanced panel data-set.
Once the equation to be estimated is known, the model can be estimated as using two models—random effects model and fixed effects model. The random effects model is applicable when two conditions are fulfilled, first, the unobserved individual effects, that is, $\alpha_i$ are random draws from a common population. Second, when the explanatory variables are strictly exogenous, that is, the error term $\mu_i$ is uncorrelated with the past, current and future value of the regressors (Hsiao, 2003). Thus, in random effects model $\mu_i$ are uncorrelated with everything else in the model, the individual-level effects are simply parameterized as additional random disturbances. Hence, the term “Random effects”. But, if the $\mu_i$ are correlated with the regressors, then fixed effects model is to be used (Baum, 2006).

To establish whether random effects model or fixed effects model is to be used for analysis, a formal test known as the Hausman test developed by Hausman in 1978 can be utilized. The null hypothesis in the Hausman test is that the fixed effects model and the random effects model estimators do not differ substantially. The test statistic in the Hausman test has an asymptotic $\chi^2$ distribution. If the null hypothesis is rejected, the conclusion is that random effects model is inappropriate and fixed effects model is a better alternative, while if the null hypothesis is not rejected, the random effects model is appropriate (Gujarati, 2004).

In the present study, the panel data-sets corresponding to time-period 2004-05 to 2010-11 are utilized to carry out the objectives 1, 2 and 3. The Hausman test is used to determine the application of fixed effects model or random effects model. While, for objectives 1 and 2 the test indicates the use of random effects model, for objective 3, the test indicates fixed effects model as the more appropriate of the two models.

Further, the present study uses the generalized least squares technique of estimation under the random effects model. The generalized least squares technique provides corrections for the presence of auto-correlation and heteroscedasticity (Kmenta, 1986). It assumes that regression parameters do not change over time and do not differ between various cross-sectional units, enhancing the reliability of the coefficient estimates (Singh et al., 2010).
STATA 11 software has been used to carry out the data analysis using the panel data regression.

Before carrying out the panel data regression analysis, the data is tested to ensure it satisfies certain assumptions. The assumptions are:

**Homoscedasticity:** It means that the conditional variances of each error term are identical across all observations. The opposite of homoscedasticity is when the variances of the error term become variable, this condition is known as heteroscedasticity. The reasons for the presence of heteroscedasticity are: poor data collecting techniques, presence of outliers, incorrect specification of the model to be tested, skewness in the distribution of one or more regressors included in the model, incorrect data transformations and through incorrect functional forms. The problem of heteroscedasticity is more likely to be present in case of cross-sectional data than time series data (Gujarati, 2004). There are a number of tests available for identifying the problem of heteroscedasticity, namely, Glejser’s (1969) Test, The Goldfeld and Quandt (1965) Test Breusch and Pagan (1979) Test, White’s (1980) Test, etc. (Baltagi, 2008).

As stated above, the chances for the presence of heteroscedasticity in time series data are less, nevertheless, in the present study the results have been computed using the White’s robust standard errors using the Random Effects generalized least squares technique which takes care of any potential problem of heteroscedasticity.

**No Autocorrelation:** Autocorrelation is present in the data when there exists correlation between the members of series of observations ordered in time (as in time series data) or in space (as in cross-sectional data). When carrying out regression analysis it is assumed that such autocorrelation is not present in the error terms. In other words the model assumes that the error term related to any observation is not correlated to the error term of any other observation. The reasons for the presence of autocorrelation are: inertia, excluded variable specification bias, Cobweb phenomenon, lags, manipulation of data, data transformation, non-stationarity (Gujarati, 2004).

The Durbin-Watson Statistic is calculated to check if the problem of autocorrelation is present in the data. The value of Durbin-Watson statistic ranges from 0-4. If the statistic is close to 4 the data has negative correlation, if it is close to 0 the
data has positive correlation, while a score near 2 reveals no auto correlation. For objectives 1 and 2, The Durbin-Watson statistic is around 2 pertaining to the data corresponding to various objectives, hence, no problem of autocorrelation seems to be affecting the data corresponding to the present study. Whereas for objective 3 the score is not clearly close to 2, it raises some doubt regarding the presence of auto-correlation. The Wooldridge test of auto-correlation (Drukker, 2003) is used to address our doubt. The null hypothesis “no first order autocorrelation” was accepted, thus alleviating the doubt.

No Multicollinearity: Multicollinearity exists when the independent variables are highly correlated with each other. Thus, a condition of no multicollinearity exists when none of the regressors can be written as exact linear combinations of the remaining regressors in the model. The presence of multicollinearity reduces the predictive power of the independent variable to the extent that it is correlated with the other independent variable. The reason for the same is that, in such a case, it becomes difficult to disentangle the separate influence of the variables that are collinear (Gujarati, 2004). The possible reasons for the presence of multicollinearity are: problems with the data collection methods, constraints on the model or in the population being sampled, model specification or due to an overdetermined model.

To check for the presence of multicollinearity, one can have a general idea by looking at the correlation matrix. Variable Inflation Factors (VIF) are computed to ascertain the exact magnitude of multicollinearity. VIF is defined as the extent to which the variance of the regression coefficients is inflated or affected due to the presence of multicollinearity. A score below 10 is considered to rule out any possibility of multicollinearity (Berman, 2002). The score for each of the variables forming a part of the study was found to be less than 10, thus, ruling out the possibility of multicollinearity. The VIFs have been calculated through the process as carried out by Hitt et al. (2006), using the OLS model, as Random Effects (RE) regression does not support the calculation of VIFs.

After satisfying all the assumptions, the panel data regression analysis has been conducted and the results have been interpreted. The influence of individual
independent variable on the dependent variable has been assessed with the help of beta coefficients and their significance level. The model fit has been established, in case of Random Effects Models by means of the Wald-Chi statistic, and in case of the Fixed Effects Models by means of the F-Statistic. In each of the models that have been tested the Wald-Chi and the F-Statistic have been found to be highly significant \((p<.01)\). The \(R^2\) squares have not been interpreted because of problems with their interpretation in generalized least squares regressions (Kmenta, 1986).

**Independent sample t test**

This test is used to compare the means of two different sets where the scores of one set are independent from those of the other set. Thus, the essential condition for application of independent sample t test is the lack of complete relationship between the two sets (Harris, 1995). In t test, the null hypothesis is that the means of the two sets are equal and the alternate hypothesis is that there is a difference in the means of these two sets.

In the present study, to carry out objective 3, t test has been applied to test the difference between the governance mechanisms of companies with dominant insider ownership and companies without dominant insider ownership. The t test establishes whether there exist significant differences among the two groups of firms (with dominant insider ownership and without dominant insider ownership) as far as their performance, corporate governance mechanisms and managerial competence are concerned.

The t test has been carried out using the STATA 11 statistical package.

**Two stage least squares**

There arises, at times, a situation when the regressors are endogenous. Endogeneity of the right-hand regressors poses a serious problem in econometrics. Endogeneity means the correlation of the right-hand side regressors and the disturbances. Endogeneity leads to inconsistency of the estimates and requires instrumental variable (IV) methods like two-stage least squares to obtain consistent parameter estimates (Baltagi, 2005). The methods of two-stage least squares give estimators that are consistent and efficient.
The two-stage least squares method, as the name indicates, is a method involving two successive applications of OLS. In the first stage of the process, the dependent variable is regressed on all the predetermined variables in the whole system, not just that equation. This is done to get rid of the likely correlation between the dependent variable and the error term. In the second stage the values, as obtained in the first stage, are used as the values for right hand side endogenous variables. The estimators thus obtained are consistent; that is, they converge to their true values, as the sample size increases indefinitely (Wooldridge, 2002).

In the present study the two-stage least squares technique has been applied to care of the potential problem of the presence of endogeneity while studying corporate governance mechanism of the board of directors (objective 3). The current research shares the concern raised by Barnhart and Rosenstein (1998) that board characteristics and firm performance may be jointly determined. Researchers have considered that firm performance might be affecting board characteristics, to this tune, Hermalin and Weisbach (1991) in their study report that poor performance increases the chances of more outside directors joining the board. Thus, the relation cannot just be unidirectional. Hence, researchers studying board characteristics take into account the issue of endogeneity, as otherwise the analysis might produce biased coefficients. Many studies in the past have taken into consideration the issue of endogeneity of the board of director characteristics and firm performance like, board composition (Hermalin and Weisbach, 1991; Barnhart et al., 1994; Jackling and Johl, 2009), board size (Yermack, 1996; Ghosh, 2006; Coles et al., 2008), and board meetings (Brick and Chidambaram, 2010). Thus, in conformity with the past research, to address the concern of endogeneity, the simultaneous equations are used and its effect of possible endogeneity on the results is obtained. The two-stage least squares is used to obtain the estimates. Following Agrawal and Knoeber (1996), the two-stage least squares regression takes firm performance and board characteristics as endogenous allowing each to affect the others.

The simultaneous system of equations take board characteristics that is, board meetings, board size, board composition and firm performance as endogenous variables, while the remaining are considered exogenous. The instruments used are the lagged
values of the exogenous variables. The two stage least squares technique has been applied using the STATA 11 statistical package.

**Structural equation modeling**

Structural equation modeling (SEM) is a unique combination of dependence and interdependence techniques. It can be called a hybrid of factor analysis and multiple regression analysis. It estimates a series of interdependent multiple regression equations simultaneously by specifying the structural model used by the statistical program (Hair et al. 2009).

It explains all of the relationships among multiple constructs by means of equations. Constructs are unobservable or latent factors which are represented by observable or measurable multiple variables. These measurable variables are called manifest variables or indicators, which are gathered through various data collection methods. In the present study the SEM technique is used to achieve objective 3.

In SEM, endogenous and exogenous constructs ought to be differentiated. The exogenous constructs are latent, multi-item equivalent of independent variables. While, the endogenous constructs are latent, multi-item equivalent to dependent variables. The endogenous constructs are theoretically determined by factors within the model and by means of a path diagram. The dependence is represented by a path to an endogenous construct from an exogenous construct.

SEM rests on a strong theoretical base. To distinguish which independent variable predicts each dependent variable one needs to draw upon theory, prior experience and research objectives. The model is to be developed only if it is strongly supported by an underlying theory. Theory is of immense importance in SEM, as it is considered to be a confirmatory analysis, that is, it is useful test and potentially confirm. Theory is required as it helps to identify the relationships in both measurement and structural models, to suggest modifications to the proposed relationships and various other facets of model estimation (Hair et al. 2009).

Once a researcher is able to express a theory in terms of relationships among indicators and latent constructs, then SEM will assess how well the theory fits reality, as represented by the data.
The SEM process begins by providing a good definition of the individual constructs. This definition is important as it provides a basis for selecting and designing individual indicator items. The selection of indicator variables is a very important step as it sets the foundation for the remainder of the SEM analysis.

In the second stage, each latent construct, to be included in the model, is identified and the measured indicator variables are assigned to latent constructs. This identification and assignment is generally represented by means of a diagram.

The third stage concerns with designing of the study to produce empirical results. After the basic model, in terms of constructs and indicators, has been specified, in this stage certain issues concerning the research design and model estimation are addressed (Hair et al. 2009).

Firstly, as regards the research design, the type of data to be analyzed is decided. There are two types of data: correlation and covariance. Though most of the SEM programs analyze the raw data and a researcher is not required to compute a correlation or a covariance matrix, as was previously done. However, the researcher’s choice as regards the correlation and covariance matrix has a huge bearing on the interpretive and statistical issues. The experts recommend the use of covariance matrix, as it provides increased flexibility due to the presence of a relatively higher information content.

Secondly, the research design decisions regarding the missing data need to be taken. It is required to determine the percentage of missing data and whether the missing data is random or non-random. It is recommended that when the sample sizes exceed 250 and the total amount of missing data is below 10 percent, then the pair-wise approach is a good solution to the missing data problem. With small samples and large amount of missing data, the model based approach (EM/ML) is a better alternative (Hair et al. 2009).

Thirdly, as far as the issue of sample size is concerned though it is recommended that large samples generally produce stable results that are replicable the sample decisions need to be based on a set of factors which include: the number of constructs comprising the study, the number of indicators for each of the constructs, the communalities and the amount of missing data etc.
The issues in model estimation involve firstly determining the model structure. It is the step where the theoretical model structure is determined and communicated, that is, assigning indicator variables to the constructs they should represent. Generally, the path diagrams are used for this purpose. Once the model has been specified the researcher needs to choose the estimation technique to identify estimates for each free parameter.

The options available are: ordinary least squares, maximum likelihood estimation, weighted least squares, generalized least squares and asymptotically distribution free estimation. The maximum likelihood estimation is the most widely used approach which has proven to be fairly robust to violation of normality assumption. In the present study, to achieve objective 3, the maximum likelihood estimation has been used (Hair et al. 2009).

After the estimation technique has been decided the computer program to be utilized for performing SEM needs to be determined. There are a range of software programs available. However, LISREL (LInear Structural RELations) is the most widely used program. It is a flexible program that can be applied in a variety of situations, that is, cross-sectional, experimental, quasi-experimental and longitudinal. The present study utilizes LISREL 8.7 for performing SEM.

The fourth step after taking care of all the issues regarding the research design is to assess the validity of the measurement model. Measurement model validity depends on the goodness-of-fit for the measurement model and specific evidence of construct validity. Goodness-of-fit indicates how well the specified model reproduces the covariance matrix among the indicator items, that is, the similarity of the observed and estimated covariance matrices. Researchers have developed a range of measures that reflect various facets of the models’ ability to represent data. Here, only those measures have been discussed which have been utilized in the study. These are:

1) Chi-square: the extent to which the chi-square value increases implies that covariance matrices of the observed sample and that estimated by SEM are not equal, thereby, meaning a lack of fit. Since the critical values of the Chi-square distribution are known, the probability that any observed sample and SEM
estimates are actually equal in a given population can be found. This probability is the traditional p-value associated with parametric statistical tests. SEM programs provide both the Chi-square value and p-value. To represent goodness-of-fit, the Chi-square values should be small and the p-value should not be statistically significant, that is, the p-value should be large.

The Chi-square statistic has two mathematical properties that are problematic in its use as a goodness-of-fit measure. Chi-square statistic being a mathematical function of the sample size and the difference between observed and estimated covariance matrices. As the sample size increases so does the value of Chi-square even if the differences in the matrices are identical. Second, the Chi-square statistic is also likely to be greater as the number of observed variables increases. Thus, just adding indicators to a model will cause the Chi-square statistic to increase (Hair et al. 2009).

2) **Goodness-of-fit index (GFI):** The GFI is a measure of the relative amount of variance and covariance is the sample data that is jointly explained by the sample data. The Adjusted Goodness-of-fit index (AGFI) differs from Goodness-of-fit index only in the fact that it adjusts for the number of degrees of freedom in the specified model. As such, it also addresses the issue of parsimony by incorporating a penalty for the inclusion of additional parameters. The GFI and AGFI can be classified as absolute indices of fit because they basically compare the hypothesized model with no model at all. Both the indices range from zero to 1, with values close to 1 being indicative of good fit (Byrne, 2009).

3) **Normed Fit Index (NFI):** it is a ratio of the difference in the Chi-square value for the fitted model and a null model divided by the Chi-square value for the null model. it ranges between 0 and 1. A value of 1 represents perfect fit (Byrne, 2009).

4) **Comparative Fit Index (CFI):** CFI is an incremental fit index that is an improved version of the NFI. CFI has many desirable qualities including its relative, but not complete, insensitivity to model complexity; it is among the
Research Methodology

most widely used indices. Its value ranges between 0 and 1, a value above .90 is associated with a model that fits well (Byrne, 2009).

5) **Root Mean Square Error of Approximation (RMSEA):** It is a measure that attempts to rectify the tendency of the Chi-square Goodness-of-fit test statistic to reject models with a large samples or a large number of observed variables. It explicitly tries to correct for both model complexity and sample size by including both in its computation. Lower the RMSEA the better it is, typically values below .10 are found for most accepted models (Hair *et al.*, 2009).

After the validity of the measurement model has been established the next step involves the specifying the structural model by assigning relationships from one construct to another based on the proposed theoretical model. The path diagram now represents both the measurement and structural part of SEM in one overall model. At this stage the model is ready for estimation, that is, the overall theory is about to be tested, including the hypothesized dependence relationships among constructs.

The final stage in the SEM includes testing the validity of the structural model and its corresponding hypothesized structural relationships. There are two key differences in assessing the fit of a structural model form that of a measurement model. Firstly, the model can be compared to a competing model in addition to establishing the overall model fit. Secondly, particular emphasis is placed on the estimated parameters for the structural relationships because they provide direct empirical evidence relating to the hypothesized relationships depicted in the structural model. The process of establishing the structural model’s validity follows the general guidelines as outlined for assessing validity of the measurement model. The overall fit can be measured using the same criteria as used for the measurement model.

Comparison with null model: in the present study, as suggested by Bentler and Bonett (1980), the proposed theoretical model is compared against a null model.

A good model fit is insufficient in itself to support a proposed structural theory. The individual parameter estimates that represent each specific hypothesis should also be examined. A theoretical model is considered valid to the extent that the parameter estimates are: 1) statistically significant and in the proposed direction and nontrivial.
Hence, the structural model is considered to be acceptable only when it demonstrates acceptable model fit and the path estimates representing each of the hypotheses are significant and in the predicted direction (Hair et al. 2009).

Summarizing, structural equation modeling is a hybrid of factor and path analysis (Hoskisson et al., 1994). In SEM, after representing the causal processes by means of structural equations, to aid clearer conceptualization of the theory, these structural equations are then pictorially modeled. Then, to determine the extent to which the hypothesized model is consistent with the data, the model is statistically tested by a simultaneous analysis of the entire system of variables (Byrne, 2007).

**THE DEPENDENT VARIABLE: COMPETITIVE ADVANTAGE**

The dependent variable is competitive advantage. Operationalizing competitive advantage has been the subject of much debate in strategic management research. For the present research competitive advantage has been operationalized as the firm performance, because a firm with competitive advantage is expected to have higher performance than if it does not. A number of researchers have measured it in conventional ways such as market-based performance and financial based performance (Fahy and Smithee, 1999). Though researchers in the past have argued that performance alone is not a sufficient parameter for competitive advantage it can however be said in favor of using performance as a measure of competitive advantage, that firstly, performance is the real and measurable part of competitive advantage and secondly, a higher performing firm is more likely to have a competitive advantage over a low performing firm than the other way around. Thus, in the present study the firm performance is used to measure competitive advantage. The terms firm performance and competitive advantage have been used interchangeably.

The measure used to capture firm performance is Tobin’s q. It reflects the market expectations of the firm’s future growth and profit potential (Lindenberg and Ross, 1981, Montgomery and Wernerfelt, 1988). It is a stock market based measure. Tobin’s q is a measure of investor confidence in the future prospects of the firm, in its growth, profit potential (Wang et al., 2009) which are related to the strategy future of the firm. It is also a measure of investor confidence, which in turn is an indicator of the
effectiveness of the corporate governance mechanisms of the firm (Dwivedi and Jain, 2005). Further, Tobin’s q has the ability to capture the value of long term investments (Surroca et al., 2010), measuring the management’s ability to generate a certain income stream from the existing asset base (Pant and Pattanayak, 2007).

In the context of resource based theory, Ricardian and monopoly rents are essential conditions for a firm to sustain its competitive advantage. Tobin’s q is the capitalized value of the aggregate Ricardian and monopoly rents (Villalonga, 2004). It is a measure that captures the intangibility at the firm level. Thus, using Tobin’s q seems to be better justified as a measure of competitive advantage than any of the accounting based measures. It overcomes the limitations of accounting based measures along with a strong empirical and theoretical base (Bharadwaj et al., 1999).

Tobin’s q is the calculated as a ratio of a firm’s market value to the replacement cost of its assets (Acquaah and Chi, 2007). We use Tobin’s q which has been calculated as:

\[
\frac{MVE+ Pref \ Cap + BVD}{BV \ of \ Assets}
\]

Where \(MVE\) is market value of equity calculated by multiplying the firm’s share price by the number of shares outstanding, \(Pref \ Cap\) is the liquidating value of the firm’s preference share capital, \(BVD\) is the book value of long term liabilities added to the net of short term assets deducted from short term liabilities and \(BV \ of \ Assets\) is the book value of assets. This method of considering the book value of debt and assets rather than market value of debt and replacement cost of assets is consistent with Sarkar and Sarkar (2000) and Pant and Pattanayak (2007), as in India, debt is not actively traded and is largely institutional, while the replacement cost is not provided for the assets.

The above discussion provides a detailed description of the research methodology adopted to carry out the study.

**LIMITATIONS OF THE STUDY**

The limitations of the study are detailed as follows:

The first limitation of the study is that due to lack of disclosure by a sizeable number of companies the sample size has to be reduced while carrying out the analysis. This
limitation of non-availability of data has a more drastic effect while carrying out the third objective. As due to the non-availability of data an unbalanced panel data-set has to be used, as if we would have otherwise resorted to a balanced panel data set for carrying out this objective, the final sample obtained would have been much smaller in size.

The second limitation is that some of the measures of resource investments as used by researchers working in the context of developed economies could not be utilized in the present study. This is because most of the companies do not report data on certain variables as do their counterparts in developed economies. This inability to secure data, in effect, limits the understanding we provide about the resource-competitive advantage relationship in the Indian context.

The third limitation is that the study includes only 500 companies. A higher sample size may provide a better picture of the influence of resource investments and their deployment on competitive advantage in the Indian scenario.