CHAPTER 3

OBJECTIVES AND RESEARCH METHODOLOGY

3.1 DATA DESCRIPTION

3.1.1 INTRODUCTION

The study is principally intended to examine the explanatory power of different variables on security returns. A sample of 30 companies whose scrip are actively traded, belonging to different industry categories (table) during the period 1995 to 2014 has been considered.
In order to identify the influence of exogenous variables of economy and companies financial on NIFTY 50 on yearly basis for the period constituted the data base.

The study is based on secondary data from the reports of the National Stock Exchange and individual quoted firms. The final sample set consists of a balanced panel of 30 firms companies that becomes the study's Index representative. The length of the time series for all the macro variables and the five financial ratios may not be the same as of the study period as it depends on the availability from various secondary sources.

**Criterion for selecting companies**

A company has been regarded as eligible for inclusion in sample if it satisfies following conditions:

- The sample companies were selected on the basis of their industry weightage on the index and their availability.
- The companies did not skip dividend for any two successive years between the periods of study.
- The average earning per share of any three successive years is not zero or negative during the period 1995-2014.
- Further, only those companies whose price data is available are retained in the sample size.
- It is listed in the National stock exchange.

**3.1.2 STUDY POPULATION AND SAMPLE SIZE**

From the population of 50 listed firms that constitute NIFTY 50 on the National stock Exchange (NSE), a sample of 30 financial and non financial quoted companies were purposely selected for analysis (Table 3.1.2).

**3.1.2 Sample Distribution by sector classification**

<table>
<thead>
<tr>
<th>Industry</th>
<th>Number of Companies</th>
<th>Companies Covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heavy &amp; Manufacturing</td>
<td>5</td>
<td>Ambuja cements, ACC ltd, Tata steel, Sterlite Industries India.</td>
</tr>
<tr>
<td>Pharmaceutical</td>
<td>3</td>
<td>Cipla ltd, Dr. Reddy's lab, Ranbaxy ltd.</td>
</tr>
<tr>
<td>IT &amp; ITES</td>
<td>2</td>
<td>HCL Infosystem ltd., Infosys ld.</td>
</tr>
<tr>
<td>Industry</td>
<td>Count</td>
<td>Companies</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Energy</td>
<td>5</td>
<td>ONGC ltd., Reliance industries Ltd., BPCL, Tata power, GAIL</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>2</td>
<td>Reliance Infra.</td>
</tr>
<tr>
<td>Commercial Banks</td>
<td>5</td>
<td>ICICI bank ltd., SBI, BOB, HDFC Bank Ltd. Indus Ind Bank</td>
</tr>
<tr>
<td>General Engineering</td>
<td>4</td>
<td>Hero ltd, Larsen &amp; Tubro, Mahindra &amp; Mahindra, Tata Motor ltd.</td>
</tr>
<tr>
<td>Cotton textile</td>
<td>1</td>
<td>Grasim ltd.</td>
</tr>
<tr>
<td>FMCG</td>
<td>2</td>
<td>ITC ltd., HUL</td>
</tr>
<tr>
<td>Finance</td>
<td>1</td>
<td>Housing Development Finance Corporation Ltd.</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>1</td>
<td>Asian Paints ltd.</td>
</tr>
</tbody>
</table>

### 3.2 DATA COLLECTION AND INSTRUMENT

The various economic variables selected for the study are divided into two categories, viz. Macro variables of the country and Micro variables for the companies. Secondary data were used for this study. The data for macroeconomic variables were sourced from relevant issues of RBI Handbook of statistics on Indian Economy, Annual publication of the Reserve Bank of India and trading economics website. Average closing values for Nifty are obtained from NSE and yahoo finance websites.

For firm's financial variables (ratios) the data were collected from CMIE databank PROWESS. Data have also been derived from the income statements and the balance sheets of the listed companies’ publications, moneycontrol. In addition, data was gathered from Books, papers, articles, and Specialized National, International Journals.

#### 3.2.1 DATA SETS (General Description)

For the financial development measure, considering various dimensions of financial sector and financial development measure suggested by existing literature, ten prominent indicators of stock market development are identified for the existence of study.

Financial theory asserts that movement in stock prices is related to economic variables. As economic variables contain important information for investors it is hypothesized that the stock market participants take these factors into account for estimating appropriate discount rate and the expected flow of dividends from stocks.
The empirical evidence, however, are not uniform in providing support to stock returns predictability using macroeconomic variables in the sense that while some studies have found that certain macroeconomic variables have significant effects in explaining returns of some stock indices, others have found no such evidence i.e., significant effects of the same variables for some other stock returns (see, for instance, Balvers et al. (1990) and Flannery and Protopapadakis (2002) for details of such findings). Further, it has been observed that predictive ability of some macro variables with respect to equity returns is quite uneven over time. Durham (2001), for example, has found this for some variables concerning monetary policy.

Researchers have also identified a number of financial variables that appear to predict future stock returns. These include price-earnings ratio, price-book value ratio, dividend yield, short term interest rate etc. (see Campbell and Shiller (1988a, b, 1998), Fama and French (1988) and Ang and Bekaert (2001) to cite a few). Keeping this in mind, I studied the impact on Indian stock returns using standard macroeconomic and micro variables relevant for India.

**Gross Domestic Product at factor cost is used as a proxy of Real output (at constant price with base year: 2004-2005)**

A measure of real output or real economic activity often used is Gross domestic Product (GDP) (Birajdar et al., 2007). As GDP numbers present a measure of overall economic activity in the economy and affect stock prices through its influence on expected future cash flows there exists a positive relationship between stock prices and GDP. To start with, an increase in current real activity increases demand on existing capital stock, which ultimately induces increased capital investment in the future, and the stock market is very likely to anticipate this (see Gallinger (1994)).

**Broad Money Supply as a proxy of Liquidity (based on M3)**

Money supply is another fundamental macroeconomic variable which widely used in the literature to determine the stock prices. Beside the extensive empirical investigation, the relationship between money supply and stock price is still ambiguous. According to the portfolio theory, an increase in the money supply may results in a portfolio change from noninterest bearing money assets to financial assets like stock.

Money supply has a direct effect on stock prices by changing liquidity. Further, as noted by Musilek (1997), money supply also has an indirect effect on
stock prices through corporate dividends by increasing or decreasing interest rates. The Money supply variable used in this thesis is M3 which will capture the percentage rate of changes in Indian money.

**Prime lending rate (short term)**

Stock prices are also influenced by changes in interest rates. Since interest rate is an opportunity cost of holding stock, an increase in interest rate is likely to lead to a substitution effect between stocks and other interest bearing assets. It is, therefore, expected that as interest rate declines stock price would rise (cf. Musilek (1997)). The interest rate variable employed in this thesis is the short-term interest rate rather than long-term interest on due to the modeling criterion conducted.

**Wholesale price Index is used in order to incorporate the inflation rate (with base year: 1993-94)**

Inflation is an increase in the general level of prices, or, alternatively, it is a decrease in the value of money. Inflation is one of those macroeconomic variables that affect every Indian citizen, irrespective of an investor, borrower or lender, almost every day. Inflation is seen as negative news by the stock markets, because it tends to curb consumer spending and therefore corporate profits. It also affects the value of the domestic currency adversely in the foreign exchange markets. The two frequently used measures of inflation in India are based on the WPI and the Consumer Price Index (CPI). Unfortunately, in India we do not have an aggregate CPI appropriate for use as an indicator of aggregate prices and demand pressures. Thus in this study WPI is used as a proxy to Indian domestic inflation.

**Real effective exchange rate (36 currency Bilateral weight with base year: 1993-94, rupees per $US)**

The other important macroeconomic variable used in this study has been the exchange rate, which represents the bilateral nominal rate of exchange of the Indian Rupee (Rs.) against one unit of a foreign currency namely US Dollar ($) has been taken to be the foreign currency against which the Indian Rupee exchange rate is considered. This is because the US Dollar has remained to be the most dominating foreign currency used for trading and investment throughout the period of this study. On an average, export-oriented companies are adversely affected by a stronger domestic currency while import-oriented firms benefit from it.
Dividend Policy

Dividend announcements are one of the most important events and the studies on stock market reaction to earnings information are included in the semi-strong form of efficient market hypothesis (EMH). The semi-strong form of efficient market hypothesis states that stock prices reflects all the publicly available information instantaneously and accurately. In this study an attempt is on the stock market reaction to dividend announcements in India in the light of various previous studies conducted in various developed countries of the world such as the USA, the UK, Australia, etc. Theoretically, stock dividends have no impact on financial position of the announcing company as net worth and total assets remain the same, through empirical evidence across the globe shows that markets react to stock dividends announcements. Dividends may convey information about the company, so it suggests the possibility of its influence on the stock market.

Economic Value added to measure firm's value addition (calculated as per CAPM)

Proponents of EVA claim that EVA is highly correlated with stock returns. EVA derives stock prices (Stewart, 1995; Medeiros, 2005) better than other accounting based performance indicators. Lefkowitz (1999) analyzed the US companies and results of the study supported Stern- Stewart hypothesis, i.e., EVA is better correlated with stock returns as compared to traditional performance measures. They found that EVA is reasonably reliable guide to understand the firm’s value. Machuga et al. (2002) in their study highlighted that EVA can be used to enhance future earnings predictions. Lehn & Makhija (1997) investigated the degree of correlation between different performance measures and stock market returns. The results indicate that EVA is the most highly correlated measure with stock returns.

Earnings per Share

Earnings per share is the amount of profit after tax divided by the total number of shares outstanding. This is very important parameter. The impact of an announcement of Earnings per share on stock prices had often been the centre of interest to researchers, shareholders and investors. This is because; EPS is one of the investment tools to evaluate a company's performance either in the short or long run. The estimated earnings can be used to measure the financial health and
prospect of a company. Therefore, in this study an investigation and evaluation has been performed to indicate the impact of EPS on the stock prices. In a way, EPS can be used as a performance indicator of the financial standing of the company.

**Debt Equity ratio to incorporate financial leverage**

Financial leverage measures firm's exposure to the financial risk. The use of fixed-charges sources of funds, such as debt and preference capital along with the owners' equity in the capital structure, is described as financial leverage, gearing, or trading on equity. The financial leverage employed by a company is intended to earn more return on the fixed-charge funds than their costs. The surplus (or deficit) will increase (or decrease) the return on the owners' equity. The percentage change in EPS occurring due to a given percentage change in EBIT is referred to as the DFL.

\[
DFL = \frac{\text{Percentage changes in EPS}}{\text{Percentage change in EBIT}}.
\]

### 3.3 OBJECTIVES OF THE RESEARCH

The objectives of this study have been decided after discussing the various issues and challenges faced by the stock market and real economy. The main objective of this research study is better understanding of the integration of stock market and real economy at the basic level. The scope and objectives had to be kept in fewer but certainly with the purpose of fulfilling the basic rationale and motive of a research project. Particularly, the study is intended to answer underneath objectives:

1. The study has been undertaken to examine the empirical relationship between stock prices and selected quantitative macro and company specific intrinsic factors (Gross domestic product, Money supply, Interest rate, Whole sale price index, Exchange rate, Dividend per share, Earning per share, Dividend payout, Economic value added, and Debt equity proportion) for the period 1995-2014.
2. To analyze if the security market properly discount corporate fundamentals to enable the share prices to represent intrinsic value i.e. to identify the contribution made by each of them on market stock price.

3. To enhance the investors understanding and evaluation in terms of sensitivity of the respective stock market index to the systematic impact of macroeconomic and company's financial factors.

3.4 RESEARCH HYPOTHESES

From literature, it is evident that a firm's stock price performance is affected by the certain external and fundamental variables. If these factors do affect a firm's performance and value, then a strong correlation between firm's stock returns and a set of variables is expected. This study therefore offers that a mix of fundamental and firm's financial affects its performance positively/negatively. Hence, hypothesis can be stated as follows:

1. $H_0$: A firm's financial variables as measured by value addition, earnings, dividend yield and financial leverage does not have significant influence on its stock returns.

$H_1$: A firm's financial variables as measured by value addition, earnings, dividend yield and financial leverage have significant influence on its stock returns.

2. $H_0$: A country's macro variables does not have significant influence on the listed companies market prices in its stock exchanges as measured by real output, exchange rate, Inflation, short term interest rate and monetary policy.

$H_1$: A country's macro variables have significant influence on the listed companies market prices in its stock exchanges as measured by real output, exchange rate, Inflation, short term interest rate and monetary policy.

3.5 METHODOLOGICAL APPROACH

The study deploys correlation and a linear multiple regression models to measure the effect of the independent variables on the dependent variable. Also to avoid the problem of multi-collinearity, Variance Inflation Factor has been estimated. Following the identification of the possible econometric models/estimators, the stationarity of all the variables are checked by applying the Panel unit root test such as Im Pesaran, Shin, Levin, Lin, Chu and Breitung. If any series is found to be nonstationary, then the series is made stationary by taking differences.
The preliminary modeling undertaken to arrive at a preferred econometric model and estimator is described. The use of effects-based models is important to allow for the heterogeneity in the dataset. The choice between fixed and random effects was based on extensive statistical analysis, which suggested that fixed effects were preferable.

The preliminary analysis revealed that static models result in unreasonably large parameter estimates and very high t-statistics, and hence a dynamic specification is preferred. There are two possible alternatives for dynamic models: an ECM or an ARDL model. The difference lies in whether the long-run relationships are modelled explicitly (in an ECM) or implicitly (in an ARDL model). Following advice from academic advisers and earlier literature, an ARDL model was chosen for use in this study due to the relatively small number of time-series observations available in the dataset, which is likely to make identification of explicit long-run elasticities challenging.

This leads to the final decision required in order to arrive at the preferred econometric estimator: which of the two common dynamic panel data estimators is preferred? On the basis of its theoretical advantages, the Arellano Bond estimator has been chosen in favour of the Blundell–Bond estimator. To address the possibility of endogeneity in the models Arellano and Bond develop the difference GMM model by differencing all regressors and employing Generalized Method of Moments [(Hansen, 1982) Arellano and Bond (1991) have suggested a consistent and efficient estimator for short panels based on the first difference of the dynamic model)].

Further, the approach also envisages appropriate specification like the diagnosis test of the final model so that the data analysis does not produce misleading inferences owing to any probable inappropriate specification. This leads to the use of Sargan test and Arellano - Bond Serial correlation test.

3.5.1 LAG-LENGTH CRITERIA

In applying econometrics techniques determination of lag length of an autoregressive process is a difficult task. This will prevent the inclusion of irrelevant variables and will make the estimation process more efficient. To overcome this problem various lag length selection criteria such as Akaike
Information Criterion, Schwarz Information Criterion, Hannan-Quinn Criterion, Final Prediction Error, and Corrected version of AIC have been suggested in the literature. Literature shows that with an increase in the sample size the performance of these criteria also improves. For the sample size of 30, although all the criterion performs well but AIC and FPE have the maximum probability of error free estimation. A sample size of 60, could be best estimated by HQC, however, AIC and SIC also give close estimation to that of HQC. Likewise, for a large sample (120 or more) SIC has the highest probability of correct estimation. Hence, AIC and FPE are effectual but not asymptotically stable where as AIC, SIC and HQC are free from such flaws.

INSTRUMENTATION

ECONOMETRIC PROCEDURE SELECTION

3.6 INTRODUCTION

The purpose of this section is to describe the econometric techniques that are available for use in the study. Using Panel Data analysis (PDA), the models like Fixed Effects, Random Effects for Static PDA and Generalized Method of Moments (GMM) technique for Dynamic PDA have been used for the estimation of the data. Hsiao and Tahir (1997) argue that pooling data, using appropriate estimation techniques, and grouping individuals according to certain a priori criteria can help overcome heterogeneity problem. However it is rather difficult to establish exogeneity between the regressors and error term especially in company financial data and therefore the direction of causality between variables might be ambiguous because of the potential endogeneity. Consequently, the contemporaneous data for both dependent variable and its determinants may cause spurious results. In financing literature the endogeneity problem is either largely ignored or corrected for only using
fixed effects or control variables approach. The researcher controls this important problem by employing GMM technique to avoid significant bias in estimates.

The DPD approach

The DPD (Dynamic Panel Data) approach is usually considered as the work of Arellano and Bond (AB) (Rev. Ec. Stud., 1991), but they in fact popularized the work of Holtz-Eakin, Newey and Rosen (Econometrica, 1988). It is based on the notion that the instrumental variables approach does not exploit all of the information available in the sample. By doing so in a Generalized Method of Moments (GMM) context, we may construct more efficient estimates of the dynamic panel data model. (See Appendix)

The Arellano–Bond estimator

The Arellano–Bond estimator sets up a generalized method of moments (GMM) problem in which the model is specified as a system of equations, one per time period, where the instruments applicable to each equation differ (for instance, in later time periods, additional lagged values of the instruments are available).

3.6.1 APPROACH SELECTION

<table>
<thead>
<tr>
<th>Fixed or Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static or Dynamic model</td>
</tr>
<tr>
<td>Dynamic Panel data or Error correction</td>
</tr>
<tr>
<td>Arellano - Bond or Blundell - Bond</td>
</tr>
</tbody>
</table>

Autoregressive distributed lag models

Description of method

In an auto-regressive distributed lag (ARDL) model, the variable of interest is assumed to be a function of the past values of itself (auto-regressive) and the current
and past values of other variables (distributed lag). They can relatively easily be extended to incorporate panel data. ARDL models can accommodate a variety of lag structures and include well-known models such as static regressions as special cases.

The general form is:

\[ y_t = \mu + \sum_{k=1}^{p} \rho_k y_{t-k} + \sum_{j=0}^{q} \beta_j x_{t-j} + \varepsilon_t \]

where \( y \) is the dependent variable and \( x \) is an explanatory variable. The static model is the case where \( p = 0 \) and \( q = 0 \).

**Advantages**

This type of model can accommodate very general lag structures and can easily be extended to incorporate panel data.

**Limitations**

Models of this type are likely to have difficulties in successfully identifying the ‘correct’ relationships between the variables in data which contain a unit root (see Box 3.6.1 below), as issues of spurious correlation may arise. As also noted in Box 3.6.1, economic time series often display evidence of a unit root and therefore it is important to consider this issue. However, the existence, or otherwise, of a unit root needs to be analysed on a case-by-case basis. Should a unit root be found in the data, the issue can be resolved by modelling in differences (i.e., using the difference between two years of data as the dependent variable in the regression).

One criticism that has been levelled at ARDL models is that, if there is a stochastic (random) trend present in the data, the dynamics in an ARDL model will be approximating this trend rather than modelling ‘real’ dynamics. However, if there is not a stochastic trend in the data, this criticism is not valid. The presence, or otherwise, of such a stochastic trend is an empirical issue and is difficult to identify given the small number of time-series observations in the dataset available for this study. (There are a maximum of 18 time-series observations available in the dataset).

**Unit roots (Box 3.6.1)**

A unit root is defined as where the value of a variable is equal to its previous level plus or minus a random error. Unit roots are often present in data which trends over time. More technically:

\[ y_t = y_{t-1} + \varepsilon_t \]

Where \( y_t \) is variable of interest at time \( t \) and \( \varepsilon_t \) is a random error term.
The ARDL models discussed in this section implicitly pool the data (note that the intercept does not vary across cross-sectional units). Pooling the data constrains all the parameters in a model (including the intercept) to be the same across all cross-sectional units.

**Feasibility**
This type of model is straightforward to estimate, and standard diagnostic tests can be used to identify errors in specification. However, the model may need to be enhanced to an ECM (see below), or estimated in differences to adequately incorporate data which may contain a unit root

**Error correction models**

**Description of method**

When data series move together over time (as is common in economic variables, such as demand, income, etc), standard statistical techniques such as OLS may find a spurious relationship between the variables. To counteract this, an ECM identifies a long-run relationship between the variables, while allowing for short-run deviations from this relationship. As with the ARDL model above, ECMs can be extended in a relatively straightforward way to allow for panel data. (See Appendix)

Time-series data often contains a unit root (see Box 3.6.1). An ECM allows for this by identifying a long-run relationship between variables such as demand and income, often based on economic theory, but allowing for short-term deviations from this long-run relationship.

In the equation below, there is a long-run relationship between the variables $y$ and $w$, which both contain a unit root (by assumption), but the short-run relationship is
affected by \( w \) and another variable, \( x \), which does not contain a unit root.

\[
\Delta y_t = \sum_{k=1}^{p} \rho_k \Delta y_{t-k} + \sum_{j=0}^{q} \beta_j x_{t-j} + \sum_{l=0}^{r} \gamma_l \Delta w_{t-l} + \lambda (y_{t-1} - \theta w_{t-1}) + \varepsilon_t
\]

Where \( \rho, \beta, \gamma, \) and \( \lambda \) are parameters to be estimated and \( \varepsilon_t \) is a random error term.

**Advantages**

The model is flexible and provides both short- and long-run elasticities, in addition to being consistent in dealing with data which contains a unit root.

**Limitations**

These models can be relatively complex to specify due to the two stage conceptual framework which requires the long-run relationship to be identified before the short-run dynamics are determined.

**Data requirement**

This type of model requires no data beyond the standard requirements for model estimation, although a longer time series may be required to identify the lag structure.

**Feasibility**

This type of model has similar data requirements to that of dynamic panel effects-based models, such as those set out below. However, using an ECM (and hence modeling in differences) results in the loss of one degree of freedom per flow.

**Dynamic panel effects-based models**

**Description of method**

Panel data often contains data with a substantial degree of heterogeneity in both levels and responses to changes in explanatory variables. An effects-based model compensates for the differences in the levels of variable by allowing each flow to have a separate intercept. Effects-based models allow for some heterogeneity in levels and omitted factors, and the market segmentation accounts for differences in responses. Hence, the elasticities are assumed to be the same across all flows within each model.
There are two main types of panel effects models: Fixed and Random effects. Both models contain an intercept which varies between cross-sectional units (flows), but are constant over time. However, fixed-effects models contain a term which is fixed but differs across cross-sectional units—i.e. flows (μ)—while, for a random-effects model, the intercept is assumed to be the same in expectation, but varies according to a random variable (u_i). The key distinction between the two types of model is whether or not the unobserved effects (which the effects, fixed or random, are allowing for) are correlated with the observed effects. It is possible to use a Hausman test to determine whether the data supports fixed or random effects.

The estimated elasticities are constrained to be the same across all flows within a market segment. This constraint is imposed because the market segmentation aims to group together flows with similar behavioural responses.

**Fixed effects:**

\[ y_{it} = \sum_{k=1}^{p} \rho_k y_{i,t-k} + \sum_{j=0}^{q} \beta_j x_{i,t-j} + \mu_i + \epsilon_t \]

**Random effects:**

\[ y_{it} = \sum_{k=1}^{p} \rho_k y_{i,t-k} + \sum_{j=0}^{q} \beta_j x_{i,t-j} + (\mu + u_i) + \epsilon_t \]

Where ρ and β are parameters to be estimated and ε_t is a random error term. It is possible to use a Hausman test to determine whether the data supports fixed or random effects (see box 3.6.3).

**Advantages**

These models allow for heterogeneity in the data, which provides for more accurate identification of the elasticities of interest. Models of this type can also estimate a wide range of lag structures.

**Limitations**

These models (dynamic panel models specified as an ARDL model) may not be adequate to deal with variables which contain unit roots, as is often the case in time-series and panel data models (see Box 3.6.1). Where this is the case, these models will need to be enhanced to an ECM or other approach (such as modelling the differences...
in the variables).

**Data requirements**

These types of models require the dependent variable to be in a panel data format, as per the data for this study. The structure of the explanatory variables is less important since they can usually be linked to the dependent variable by making assumptions about how the variable differs across the panel, although explanatory variables which match the level of aggregation of the dependent variable are preferred. The assumptions required are discussed in the Model specification report.

**Feasibility**

The issue with this class of models is not whether it is feasible to estimate them (given the data available for this study), but whether they are the most appropriate means of estimating the elasticities to be used in a forecasting framework (i.e., whether, given the dataset available for use in this study, this class of models is optimal).

**Others**

A number of other types of econometric models were discounted after an initial shift, for a number of reasons, including practicality and the ability to produce meaningful long-term forecasts. The two models considered in this sub-section are structural time series and almost ideal demand systems models.

**Structural time series**

Structural time series models are very flexible and allow for elasticities, together with a trend, which can change over time.

More technically, these models allow for varying coefficients and a stochastic slope, and as such provide a very flexible way to model time-series data. This methodology requires a long time dimension to the data in order to identify the dynamics. The time dimension of the dataset available for this study is relatively short, and for this reason this methodology has not been used.

**Summary**

This section has reviewed several possible econometric techniques for use in this study. Table 3.6.1 summarizes the methods reviewed and their theoretical suitability and feasibility for use within this study.

**Table 3.6.1 Summary of econometric model**
As can be seen from the table, a number of techniques may have been useful for estimating the models in this study. Almost ideal demand systems and structural time series were not investigated further because of the practical considerations set out above. The remaining three suitable techniques all have strengths and weaknesses, and the preferred option has been determined both by the data as well as theoretical considerations. It is possible to use combinations of these techniques; for example, ARDL and effects-based models, or effects-based and an ECM. These combinations have also been considered.

The next section considers the initial modelling undertaken in order to decide which option, or combination of options (ARDL, ECM, or panel effects), was preferred.

### 3.6.2 INITIAL MODELLING

Following the identification of the possible econometric models/estimators, the preliminary modeling undertaken to arrive at a preferred econometric model and estimator is described in this section.

The use of effects-based models is important to allow for the heterogeneity in the dataset. The choice between fixed and random effects was based on extensive statistical analysis, which suggested that fixed effects were preferable.

The preliminary analysis revealed that static models result in unreasonably large parameter estimates and very high t-statistics, and hence a dynamic specification is preferred. There are two possible alternatives for dynamic models: an ECM or an ARDL model. The difference lies in whether the long-run relationships are modeled explicitly (in
an ECM) or implicitly (in an ARDL model).

Following advice from academic advisers and earlier literature, an ARDL model was chosen for use in this study due to the relatively small number of time-series observations available in the dataset, which is likely to make identification of explicit long-run elasticities challenging.

This leads to the final decision required in order to arrive at the preferred econometric estimator: which of the two common dynamic panel data estimators is preferred? On the basis of its theoretical advantages, the Arellano Bond estimator has been chosen in favour of the Blundell–Bond estimator. The preliminary modeling also suggests that the data cleaning process has not had an unduly large effect on the estimated elasticities.

This section discusses the initial modeling undertaken for the study, based on the diagnostics available for use with the preferred model.

The aims of the initial modeling were to investigate:

- The optimal formulation of the stock prices determinant variable;
- Whether the data cleaning process applied to the data makes an undue difference to the estimated elasticities;
- The preferred estimator.

The results presented in this section are for descriptive purposes only and should not be interpreted as robust models.

The next section discusses static and dynamic models.

**Dynamics**

One of the most important issues investigated during the preliminary modeling phase is that of dynamics—i.e., whether a model is required that allows for adjustments to changes in the explanatory variables after the current period.

A static model assumes that all effects on the dependent variable from changes in the explanatory variables are completed within the same period as the change occurs (i.e., within one year for this dataset). Dynamic models include lagged values of the dependent variable as an explanatory variable in the regression equation, together with lagged values of the explanatory variables. In Static Pooled Data Models, individual specific as well as time effects are taken care of, but such models cannot explain the impact of adjustment cost and floatation costs on firms financing and dividend distribution decisions.

Estimating parameters using a pure static model, if all the coefficients of possible lagged
variables are not different from zero, restricts the previous periods so that they have no impact at all on current adjustments.

Thus a dynamic econometric model is specified. The use of dynamic model is attributed to couple of other reasons. Firstly, since the firms cannot offset the adjustment and miscellaneous costs immediately, it is expected that there is a role of the lagged values of both the dependent and independent variables. Secondly, a dynamic model is more general than a static model. A firms’ dynamic adjustment of the effect of various factors may take several years to complete. The significance of explanatory variables can change considerably in the dynamic analysis. Applicability of such a procedure will help identify broad group of factors influencing share price, increase their robustness pattern, and establish a degree of generalizability over cross section of industry, and time over the period of study. Therefore a partial adjustment model is specified to find out the effect of these factors on the market price of stock.

3.6.3 INITIAL MODEL RESULTS

An important consideration in the estimation of panel data models is whether to use fixed or random effects, or no panel effects. Comparisons were made between random and fixed-effects models and, for most of the models, random effects were rejected in favour of fixed effects, based on the Hausman test statistic. This is in addition to the theoretical rationale for using fixed effects in preference to random effects; namely that the unobserved variables are likely to be correlated with the observed variables and, therefore, the random-effects models are likely to be inconsistent (i.e., they will not produce accurate estimates of the ‘true’ elasticity). Accordingly, the modeling proceeded with fixed-effects models.

**Hausman test (Box 3.6.3)**

The Hausman test can be used to test whether fixed or random effects are preferred. In essence, it tests whether it is valid to assume that the unobserved effects are uncorrelated with the observed variables.

The basis of the test is that, if the unobserved effects are correlated with the observed effects, the random-effects estimator is inconsistent, but the fixed-effects estimator is not. However, if the unobserved effects are not correlated with the observed effects then the fixed-effects estimator is still consistent, while the random-effects model is both consistent and efficient.

Source: Oxera Consulting ltd.

The Hausman test can be used to test whether fixed or random effects are preferred. In essence, it tests whether it is valid to assume that the unobserved effects are
3.7 MODELLING APPROACH

3.7.1 Choosing between ARDL and ECM

After deciding between fixed and random effect models one need to choose between dynamic models as there are two main classes of dynamic panel data estimators: ARDL models, or ECMs. The relatively short time series of the dataset available for use in this study (a maximum of 19 annual observations) may prevent a single pooled model being estimated for each market segment. Panel data can improve the difficulties associated with the short time series, but introduces other complexities into the analysis. Both ARDL and ECMs can be constructed as time-series or panel data approaches.

Following consultation with experts and earlier literature, has determined that the time-series element of the dataset available for this study may pose problems for identifying the explicit long-run dynamics required to estimate an ECM. In light of this, ARDL-type dynamic panel data models, which have been developed specifically for datasets with a small time dimension, have been used.

A dynamic model can be achieved by using a lagged dependent variable as an explanatory variable. However, since the lagged dependent variable is endogenous, it is correlated with the unobserved panel effects by construction, which makes standard fixed- or random-effects estimators inconsistent. Therefore, more advanced dynamic panel data estimators are required to estimate consistent elasticities. Two main estimators are available to use in the case of large N, small T datasets (such as that currently available for use in this study). The two techniques are known as Arellano & Bond, and Blundell & Bond. The next section details the analysis that has been undertaken to determine which of these estimation techniques is preferred.
3.7.2 DYNAMIC PANEL DATA MODELS

The models set out below have been developed specifically for datasets with a large cross-sectional component (N) and a small time-series component (T). The difference between the two lies in the assumptions which need to be made to use the estimators. The Blundell–Bond estimator requires more assumptions than the Arellano–Bond estimator, but is more efficient and is still consistent when the data is highly dependent on previous data.

Process
Initial models were computed using the estimation techniques outlined by Arellano & Bond. Arellano & Bond suggest an estimator constructed to deal with the endogenous nature of the lagged dependent variable. This approach is broadly similar to standard instrumental variables (IV) estimators, with different variables being used as instruments (or proxies) for the lagged dependent variable. The main requirement for applying these estimators is that there should be no serial correlation in the residuals otherwise the instrumentation is not valid and hence the estimator is inconsistent.

The next section considers the diagnostic tests available to test the assumptions and conditions underlying the chosen modeling technique.

3.8 DIAGNOSTIC TESTING

Diagnostic testing in panel data is not as advanced as diagnostic testing in cross-sectional or time-series analysis. However, there are a number of diagnostic tests which can be used, including tests for:

- Autocorrelation;
- Cointegration;
- Instrument validity.

3.8.1 AUTOCORRELATION

A test for autocorrelation in dynamic panel data was developed by Arellano & Bond. Both the Arellano–Bond estimators include an assumption that there is no residual autocorrelation in the model.

The Arellano–Bond test for autocorrelation is a statistical test for correlation in the first-differenced errors. There will be autocorrelation in the first differences because
of the construction of the model, and hence it is to be expected that the null hypothesis of no autocorrelation will be rejected for the first differences. However, rejection of the null hypothesis of no autocorrelation at higher orders (e.g., at second or third differences) implies that the moment conditions are not valid and hence that the estimator is not valid.

3.8.2 COINTEGRATION

There is a developed literature on testing for cointegration in panel data. However, one of the challenges with testing the stationarity of residuals in panel data is the interpretation of a rejection of the null hypothesis that ‘all cross-sectional units are stationary’. This implies that at least one of the cross-sectional units is non-stationary. If the cross-sectional units in question are the residuals from a regression, this might imply that at least one of the cross-sectional units has a spurious regression with the others. In practice, this means that the testing for spurious relationships between variables in a panel dataset is not straightforward.

It is important to note that the power of these tests is also expected to be weak in this current study due to the low number of time-series observations.

3.8.3 MOMENT VALIDITY

The estimators mentioned above are only valid if the moment conditions are valid—i.e., the instruments used in the estimation fulfill the assumptions of being uncorrelated with the error term and correlated with the lagged dependent variable. However, there are no direct tests for this assumption. A second-best alternative is to test whether the instruments are valid using the Sargan test. As stated above, the validity of the instruments is important for consistent estimates of the parameters of interest. However, this test can be rejected for a number of reasons, and rejection of the null hypothesis does not necessarily imply that the model is misspecified. If the error term is not homoscedastic, the test statistic does not have a standard asymptotic distribution—i.e., critical values cannot be tabulated. This does not mean that the model is misspecified, but that the Sargan test will be failed, and hence, under these circumstances, the test is not a useful test of model specification.

Model fit

The fit of the model to the data is an important consideration for all models,
but particularly so when using models for forecasting. However, it is important to avoid ‘tweaking’ the models solely to improve the fit to the available data since this may result in models that provide a good fit to the dataset, but do not forecast well.

The approach to measuring model fit in panel data models is still an area of active research. On the advice of academic advisers, the squared correlation coefficient between actual and fitted values has been used as the measure of model fit for this study. This provides a number—for each model which is bounded by zero and one—where a result closer to one implies a better match between the actual and fitted results.

The previous section discussed the modelling undertaken to determine the preferred econometric approach and this section has discussed how the models can be tested for possible errors in specification, or breaches of the underlying model assumptions. The next two sections consider whether the data cleaning process is likely to have had an unduly large effect on the parameter estimates, and hence the estimated elasticities, and how the car ownership variable could enter into the analysis.

This chapter has outlined how the preferred econometric estimator has been selected. Below illustrates this process, and the decisions made.
The preferred econometric estimator is the Arellano - Bond estimator, which has been derived for the type of modeling that has been undertaken for this study. This has been selected after considering a substantial range of econometric estimators on theoretical grounds, which led to the rejection of some possibilities as being unsuitable for use in this study. Following this theoretical consideration, preliminary modeling was undertaken to determine which of a number of techniques were most appropriate (i.e., whether the preferred model was static or dynamic, etc). The decision to use a dynamic estimator on a dataset with a relatively small time-series component has resulted in the use of an estimator that has been derived for just such ‘large N small T’ datasets. The justification and selection of the preferred econometric estimator is a key step in estimating a new elasticities-based forecasting framework. The chosen econometric approach will be used to estimate a number of functional forms for each market segment.

**Software used:** Microsoft Office 2007 and Eviews 8.0 version
REFERENCES


