CHAPTER – 3
Data and Methodology

3.1 Data and Methodology for Primary vs Sectoral Indices of BSE and NSE
3.2 Data and Methodology for Sensex vs International Indices
3.3 Data and Methodology for Stock market Indices vs Commodity Prices
3.1 Data and Methodology for Primary vs Sectoral Indices of BSE and NSE

At the outset it needs to be clarified that the term primary index has been used to refer to the main index that is widely followed by investors as well as the wider community of interested stakeholders in order to make sense of the stock market. Thus, in the present case, the Sensex could be considered as the primary index of the BSE while the Nifty would fulfil a similar role for the NSE. The study has been performed for discerning any possible relationship between the primary index on one hand, and the sectoral components on the other. As some of the constituent scrips of the sectoral indices will obviously feature as components of the primary index, an alternative approach has been considered where instead of operating with the sectoral indices per se, we consider the constituent scrips corresponding to each sectoral index which do not simultaneously feature as a component of the primary index. It is obvious that the number of such scrips may be substantial. Therefore we have chosen to consider a maximum of three leading scrips in each case on the basis of market capitalisation. Thus, in case of the BSE we attempt to regress the values of the Sensex against those of the three leading scrips each from the auto, bank, consumer durables, consumer goods, healthcare, IT, metals, oil & gas and power sectors. Similarly, in case of the NSE we try to relate the values of the S&P CNX Nifty with those of the non-Nifty scrips from the bank, infrastructure and IT sectors.

The data utilised for the purpose includes values of the Sensex, the Nifty and the selected scrips on the respective exchanges from 2000 to 2009. Care has been taken to ensure that only those instances are considered where data is concurrently available for all components.

A brief introduction to the three major statistical tools that constitute the bedrock of the present effort would be very relevant here.
(a) Unit Root Test

Whenever a sequence of data is to be used for time series analysis, it is necessary to evaluate whether such data may give rise to spurious phenomena that result in localised short-term effects that die down with time. The property of data that is free from such spurious phenomena is known as stationarity. It has been said that a stochastic process is said to be stationary if its mean and variance are constant over time and the value of the covariance between the two time periods depends only on the lag between these two periods and not the actual time at which the covariance is computed. The difficulty in dealing with a non-stationary time series is that any conclusion regarding the behaviour of such a time series will be applicable only for a limited period of time. In other words, all such conclusions cannot be applied to other time periods. As a result forecasts made on the basis of these conclusions will have little practical value.

For the purpose of testing the existence of stationarity in a data series we use the following equation:

\[ Y_t = \rho Y_{t-1} + u_t \]

When \( \rho = 1 \), i.e., when a unit root is present, it indicates a non-stationary process. The above equation can be alternatively written as follows:

\[ \Delta Y_t = \delta Y_{t-1} + u_t \]

where \( \delta = \rho - 1 \); obviously the existence of non-stationarity is now denoted by \( \delta \neq 0 \). While Dickey and Fuller have proposed number of tests for stationarity, in this work we have utilised the Augmented Dickey-Fuller (ADF) test, which also happens to be the most popular variant of the test. In a few cases, when the ADF test has not proved conclusive, we have opted for the Philips-Perron test.
(b) Cointegration Analysis

For the next step, i.e., evaluation of existence of long-term relationship, we have adopted the cointegration analysis technique proposed by Johansen. The phenomenon of cointegration was introduced by Granger. At a broad level it can be said that cointegration is applicable when several variables are driven by a common stochastic trend; such variables have a particularly strong link that may be of interest. If \( y_t \) and \( x_t \) are two I(1) variables, then they are said to be cointegrated if a linear combination of these variables is I(0), i.e., if there exists a \( \beta \) such that \( y_t - \beta x_t \) is I(0). This can also be denoted as CI(1,1).

The estimation and testing of cointegration is performed at two levels, viz., single equation methods and system methods. The simplest form of the single equation is the two-variable model, where each variable is I(1), so that the cointegrating relationship is denoted by

\[
y_t = \beta x_t + u_t, \quad t = 1,2,\ldots,T
\]

where \( u_t \) is I(0).

Since the present work involves use of the system methods for estimation and testing of cointegration (besides the disadvantages of the single equation methods compared the system methods) it would be more relevant to look into the cointegration analysis from the system aspect. One of the most popular methods in this respect is the one attributed to Johansen. This method is based on a Vector Autoregressive (VAR) model. If we have a set of time series variables \( y_t = (y_{1t}, y_{2t}, \ldots, y_{Kt})' \) where each variable is I(1) the basic model of order \( p \) will have the form

\[
y_t = A_1 y_{t-1} + A_2 y_{t-2} + \ldots + A_p y_{t-p} + u_t, \quad t = 1,\ldots,T
\]

where \( A_i \) (\( i = 1,2,\ldots, p \)) are the \( (K \times K) \) coefficient matrices while \( u_t = (u_{1t}, u_{2t}, \ldots, u_{Kt}) \) is the unobservable error term. It is generally assumed that \( u_t \) is an independent white noise process with zero mean and is in invariant with time, having a positive definite covariance matrix \( E(u_t u_t') = \Sigma_u \) i.e., \( u_t \)'s are independent stochastic vectors with \( u_t \sim (0, \Sigma_u) \).

---

2 A time series variable is said to be I(1) if the first difference of its values \((y_t - y_{t-1})\) gives rise to stationary values. On the other hand, two variables \( x_t \) and \( y_t \) are said to be I(0), i.e., integrated of the order 0 if for a particular value of \( \beta \) the linear combination \( y_t - \beta x_t \) is found to be stationary, though the \( x_t \) and \( y_t \) themselves are non-stationary or I(1).
The Johansen procedure uses the method of maximum likelihood in case of the VAR(p) model on the basis of the assumption that the error terms are Gaussian. By subtracting $y_{t-1}$ from the earlier equation we get

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \ldots + \Gamma_p \Delta y_{t-p+1} + u_t$$

where $\Pi = -\left(I_k - \sum_{i=1}^{p} A_i \right)$ and $\Gamma_i = -\sum_{j=p+1}^{p} A_j, i = 1, 2, \ldots, p - 1$.

As $y_t$ is a $K \times 1$ vector of I(1) variables, $\Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$ are all $K \times 1$ vector of I(0) variables, while on the right hand side $y_{t-1}$ is I(1). This means that the system of equations will be consistent if $\Pi y_{t-1}$ must also be I(0) and hence will contain the cointegrating relations. $\Gamma_i, i = 1, 2, \ldots, p-1,$ are referred to as short-run parameters, while $\Pi y_{t-1}$ is sometimes referred to as the long-run part.

In order that $\Pi y_{t-1}$ should be I(0), $\Pi$ should not be of full rank. If its rank is $r$ then $\Pi$ can be expressed as the product of two matrices $\alpha$ and $\beta$ i.e., $\Pi = \alpha \beta'$ where both $\alpha$ and $\beta$ is a $K \times r$ matrices of rank $r$. In that case $\beta' y_{t-1}$ denoted the $r$ cointegrating equations. Therefore the rank of $\Pi$ is known as the cointegrating rank of the system. $\beta'$ is known as cointegrating matrix whereas $\alpha$ is the matrix of weights relating to the cointegrating relations. As $\alpha$ and $\beta$ are not unique, there can be many $\alpha$ and $\beta$ matrices containing the cointegrating relations or their linear transformations.

In order to test for cointegration with system method we have used two kinds of tests, viz., the trace test and max eigenvalue test.

In case of the trace test the null hypothesis is taken as ‘at most $r$ cointegrating vectors’ where $r$ is the rank of $\Pi$. Accordingly the hypotheses are considered in the following sequence:

- $H^0_0 : r = 0$ versus $H^0_1 : r > 0$
- $H^1_0 : r = 1$ versus $H^1_1 : r > 1$
- $H^2_0 : r = 2$ versus $H^2_1 : r > 2$
- $\ldots\ldots\ldots\ldots$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$
- $H^{k-1}_0 : r = K - 1$ versus $H^{k-1}_1 : r = K$
The testing sequence terminates and the corresponding cointegrating rank in the null hypothesis is selected when the null hypothesis cannot be rejected for the first time. In case the first null hypothesis in the sequence i.e., $H^0_0$ cannot be rejected, then it means that there is no cointegrating relationship involving the $K_I(1)$ variables, and hence a VAR process in first difference is then considered for studying relationships involving the $K$ variables. If, on the other hand, all the null hypotheses starting from $H^0_o$ upto $H^{K-1}_0$ are rejected, VAR process in level values should then be considered for subsequent analysis.

An alternative method is provided by the max eigenvalue test, where instead of testing for a pair of hypotheses as in the trace test, the null hypothesis ($H_0$) is that there are $r$ cointegrating vectors against the alternative ($H_1$) that there are $r+1$ cointegrating vectors. Test statistic, denoted as $\lambda_{\text{max}}$, can easily be seen to be equal to

\[
\lambda_{\text{max}} = -T \ln \left( 1 - \hat{\lambda}_{r+1} \right)
\]

(c) Granger Causality Test

The third component of the analytic portion of the study is made up of the Granger Causality test. This kind of causality study enters into the picture as cointegration analysis cannot throw complete light on the nature of relationship between specific pairs of variables. However, in the light of the present study, the focus is obviously on the latter as we try to examine the interactions among the primary index for each exchange and its diverse sectoral scrips.

At first each data series is to be tested for stationarity using the ADF test. Once it has been ascertained that each of these series is non-stationary, i.e., all of them are $I(1)$, the Johansen cointegration test is carried out to evaluate the existence of a long-term relationship. Lastly, we have again resorted to the Granger Causality test to identify the contours of the relationship.
3.2 Data and Methodology for Sensex vs International Indices

We have utilised the above approach for discerning any possible relationship between the Sensex and the global indices. The data utilised for this purpose consists of the values of the eight indices specified earlier over a period from 2000 to 2009. Care has been taken to ensure that only those instances are considered where data is concurrently available for all eight indices.

At first each of the eight data series is to be tested for stationarity using the ADF test. Once it has been ascertained that each of these series is non-stationary, i.e., all of them are I(1), the Johansen cointegration test is carried out to evaluate the existence of a long-term relationship between the Sensex and each of the seven foreign indices. Finally, the analysis is sought to be completed through the application of the Granger Causality test, the motive being to establish the nature of the relationship existing between the bellwether\(^3\) Indian index, viz., the Sensex and each of the international indices considered for the study.

3.3 Data and Methodology for Stock market Indices vs Commodity Prices

In order to add another dimension to this work, we have included a third component where we have sought to apply the methodology established in the preceding cases to another type of comparison where we are looking across markets. Specifically we have attempted to decipher whether there are commonalities in the movements of the stock market (as exemplified by the Sensex) and those of the commodity prices (as exemplified by the international prices of gold and oil). As before we use the ADF test for determining the existence of stationarity or otherwise among the three series that constitute the data for this study. Once non-stationarity is confirmed for all the series, we proceed for Johansen Cointegration analysis. The analysis is rounded out by utilising the Granger Causality test for determining the nature of relationship between indices.

---

\(^3\) According to the Merriam-Webster dictionary, the word bellwether refers to "one that takes the lead or initiative: leader; also: an indicator of trends". As per Wall Street Words: An A to Z Guide to Investment Terms for Today's Investor by David Logan Scott, the word implies "a security that tends to lead the market and signal the general direction of future price movements. An increasing price for a bellwether stock is considered a bullish signal for the overall stock market." Consequently, a bellwether index could be regarded as one that sets the direction for the entire market.