Chapter 2

Data and Methodology

An attempt has been made to investigate the energy sector profile in the country in terms of the available resources i.e., Commercial: Coal, Oil, Natural Gas, Electric Power: Thermal, Hydro and Nuclear; Non-Commercial: Firewood, Animal waste, Crop residue and animal drought power; Renewable: Solar, Wind, Bio-gasification and Small-hydro Power; their production, consumption and imports during the period from 1973 to 2005. In order to see impact of energy consumption on the economy growth of the country, the relationship between energy consumption and economic growth of the country has been identified. The relationship between energy consumption and GDP has been evaluated both, in totality (total energy consumption) and component-wise (coal, oil, natural gas, electricity etc.). In addition, an attempt has also been made to forecast overall commercial energy demand and the resource-wise energy demand for the period 2010-2030 A.D.

Variables

For the purpose of analyzing the aforementioned objectives, variables listed in Table 2.1 have been used. The study has used time-series data pertaining to energy variables (production and consumption) and other economic variables (GDP, Per Capita Net National Product, and Population). The time series data was converted into logarithms base before applying the tests.

Data Collection

The secondary data for the above variables has been collected for the period 1973 to 2005. The information and data pertaining to energy variables has been collected from various sources such as:
Table 2.1

Variables

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
<th>Unit: Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cp</td>
<td>Production of Coal</td>
<td>MTOE</td>
</tr>
<tr>
<td>Op</td>
<td>Production of Oil</td>
<td>MTOE</td>
</tr>
<tr>
<td>NGp</td>
<td>Production of Natural Gas</td>
<td>MTOE</td>
</tr>
<tr>
<td>Ge</td>
<td>Generation of Electricity</td>
<td>MTOE</td>
</tr>
<tr>
<td>TCEp</td>
<td>Total Commercial Energy Production</td>
<td>MTOE</td>
</tr>
<tr>
<td>Cc</td>
<td>Consumption of Coal</td>
<td>MTOE</td>
</tr>
<tr>
<td>Oc</td>
<td>Consumption of Oil</td>
<td>MTOE</td>
</tr>
<tr>
<td>NGc</td>
<td>Consumption of Natural Gas</td>
<td>MTOE</td>
</tr>
<tr>
<td>Ec</td>
<td>Consumption of Electricity</td>
<td>MTOE</td>
</tr>
<tr>
<td>TCEc</td>
<td>Total Commercial Energy Consumption</td>
<td>MTOE</td>
</tr>
<tr>
<td>PCEC</td>
<td>Per Capita Energy Consumption</td>
<td>kWh</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross Domestic Product</td>
<td>Rs. crore</td>
</tr>
<tr>
<td>PCNNP</td>
<td>Per capita Net National Product</td>
<td>Rupees</td>
</tr>
</tbody>
</table>


2. Annual Reports, Ministry of Power, Government of India, New Delhi;

3. Annual Report, Ministry of Petroleum and Natural Gas, Government of India, New Delhi;


7. Five Year Plans (Seventh to Eleventh) published by Planning Commission, Government of India, New Delhi;

8. India Energy Book published by World Energy Council-India Member Committee, New Delhi; and


In addition to the above, the data on macro-economic variables has been collected from Handbook of Statistics on Indian Economy (annually) published by Reserve Bank of India (RBI) and the online Database on Indian Economy, accessed at (http://dbie.rbi.org.in/ DBIE/ dbie. rbi? site=home), an online service provided by RBI.

The information on energy sector and various energy resources has been collected from the documents published by various agencies, namely, Central Electricity Agency, Central Electricity Regulatory Commission, State Electricity Regulatory Commissions, Geological Survey of India, Coal India Limited (CIL), Advisory Board on Energy, Petroleum and Natural Gas Regulatory Board, Bureau of Energy Efficiency, Energy Information Administration (official website of USA Government), Power Grid Corporation of India, Directorate General of Hydrocarbons, World Bank, Rural Electrification Corporation, Power Finance Corporation etc. The information on the Indian Economy has
been collected from different Government of India published documents and web search.

**Software Used**

There are many types of software for making the Cross-Tabulation, Trends, Growth Rates, Compound Annual Growth Rates (CAGR) T-test, Regression and so on. In this study, for analysis and estimation, the software’s Excel, Eviews and SPSS have been applied.

**Methodology**

**a. Trend Analysis:** For the purpose of understanding the trend and pattern of production and consumption of energy resources (commercial, non-commercial and renewable forms of energy) in the last 33 years, the energy produced or consumed in different forms has been aggregated and trend analysis has been conducted using general least square technique. To examine the structural stability of the technique, the sample period was divided into two parts; 1973-74 to 1990-91 and 1991-92 to 2005-06.

The growth rate is calculated using the least squares growth rate equation which is as follows. The least-squares growth rate, r, is estimated by fitting a linear regression trend line to the annual values of the variable in the relevant period. The regression equation takes the form:

\[ Y = a + bt \]

where, Y, is the dependent variable for a selected value of t. a is the estimated value of Y where the line crosses the Y-axis when t is zero. b is the slope of the line, or the average change in Y for each change of one unit in t. t is any value of time that is selected.
Further, the compound annual growth rate has been calculated by taking the nth root of the total percentage growth rate, where n is the number of years in the period being considered. This can be written as follows:

\[
CAGR = \left( \frac{\text{Ending Value}}{\text{Beginning Value}} \right)^\left(\frac{1}{\text{n of years}}\right) - 1
\]

\[b. \text{ Vector Auto Regression (VAR) Analysis:} \]
For the purpose of evaluating the relationship of energy consumption and the economic growth of the country, VAR Model has been applied. VAR is an econometric device to model multivariate time series.

The Vector Auto Regressive (VAR) model of Sims (1980) is one of the most successful, flexible and easy to use model for the analysis of multivariate time series. VAR models are widely used in econometrics. Their popularity is due to the flexibility of the VAR framework and the ease of producing macroeconomic models with useful descriptive characteristics, within statistical tests of economically meaningful hypothesis can be executed. Over the last two decades, VARs have been applied to numerous of macroeconomic data sets providing an adequate fit of the data and fruitful insight on the interrelation between economic data.

A VAR is an equation, n-variable linear model in which each variable is in turn explained by its own lagged values, plus current and past values of the remaining n-1 variables. VAR is simply an autoregressive model with at least two variables time series having p as the number of lags used in the model. Thus, it is a multivariate Auto Regressive model in which n variables are specified as linear functions of
p of their own lags, p lags of the other n - 1 variable. The lag length (p) for the model may be determined using model selection criteria such as the Akaike Information Criterion (AIC) and the Schwartz Bayesian Criterion (SBC).

It has been argued that it is better to pursue macroeconomic modeling within a VAR framework than within a simultaneous equations framework. The proponents of VAR modeling argue that simultaneous equations modeling places too many dubious a priori restrictions on structural parameters and that VARs provide a flexible framework free from the constraints of our questionable inadequate macroeconomic theory (Statistical Electronic Textbook, 1984-2003).

In VAR analysis, there is no division between endogenous and exogenous variables. It is also assumed that no instantaneous feedback exists among the variables and so the system is not simultaneous. Thus, a VAR is a system of equations where current values of each variable depend on past values of itself and past values of the other variables.

This model (VAR) is used for analyzing the interrelation of time series and the dynamic impacts of random disturbances (or innovations) on the system of variables. Following Enders (1995), consider a simple bivariate first order VAR, i.e. VAR(1), model

\begin{align*}
y_t &= \beta_{10} - \beta_{12} x_t + \alpha_{11} y_{t-1} + \alpha_{12} x_{t-1} + u_{yt} \\
x_t &= \beta_{20} - \beta_{21} y_t + \alpha_{21} y_{t-1} + \alpha_{22} x_{t-1} + u_{xt}
\end{align*}

Where it is assumed that both \( y_t \) and \( x_t \) are stationary; \( u_{yt} \) and \( u_{xt} \) are white noise with standard deviations of \( \sigma_y \) and \( \sigma_x \), respectively; \( u_{yt} \) and \( u_{xt} \) are uncorrelated.
Thus, the VAR(1) model captures the feedback effects allowing current and past values of the variables in the system. The coefficients $\beta_{12}$ and $\beta_{21}$ represent the contemporaneous effects of a unit change of $x_t$ on $y_t$ and of $y_t$ on $x_t$, respectively;

$\alpha_{12}$ is the effect of a unit change of $x_{t-1}$ on $y_t$,

$\alpha_{21}$ is the effect of a unit change of $y_{t-1}$ on $x_t$.

Hence $y_t$ and $x_t$ have mutually contemporaneous effects on each other in the system.

The disturbance terms $u_{yt}$ and $u_{xt}$ are shocks or innovations in $y_t$ and $x_t$. The term $u_{yt}$ has an indirect contemporaneous influence on $x_t$ if $\beta_{21} \neq 0$, and $u_{xt}$ has an indirect contemporaneous effect on $y_t$ if $\beta_{12} \neq 0$.

The first step is to check for stationary of the original variables and then test co-integration between them. The results are very sensitive to the selection of lag length. If the chosen lag length is less than the true lag length, the omission of relevant lags can cause bias. If the chosen lag length is more, the irrelevant lags in the equation cause the estimates to be inefficient.

If the time series are not stationary then the VAR framework needs to be modified to allow consistent estimation of the relationships among the series. The Vector Error Correction (VEC) model is just a special case of the VAR for variables that are stationary in their differences (i.e., $I(1)$). The VEC can also take into account any co-integrating relationships among the variables.

**Tests for Presence of Unit Roots**

The standard practice in the time series literature obliges a check for unit roots in each series before estimating any equation. According to
Granger and Newbold (1974) as cited by Wijeweera et al. (2005), if there is a unit root then that particular series is considered to be non-stationary.

Empirical analysis using time-series data requires that each variable $x_t$ is stationary, meaning that:

i. Its mean $E(x_t)$ and variance $V(x_t)$ are constant over time $t$.

ii. The value of covariance $Cov(x_t, x_{t-k})$ between two time periods $t$ and $t-k$ depends only on the lag $k$ and not on the actual time $t$ at which the covariance is computed\(^{17}\).

The problem of spurious regression is likely to arise, when two time-series exhibit both strong upward or downward trends. In this case the high $R^2$ could be due to the presence of the trend, not due to any meaningful relationship between the two variables. Moreover, estimation based on non-stationary variables may lead to spurious results, which will produce high $R^2$ and $t$-statistics, but without any coherent economic meaning or has insignificant results. In accordance with standard practice, this study has checked whether or not the variables are stationary.

A unit root test examines whether a time series variable is non-stationary using an autoregressive model. The most famous test is the Augmented Dickey-Fuller test. Another test is the Philips-Perron test. Both these tests use the existence of a unit root as the null hypothesis. The Augmented Dickey-Fuller (ADF) test for checking unit roots was employed in this study.

\(^{17}\)Gujarati\(1995\), p.713
The ADF Unit Root Test is based on the following three regression forms:

1. Without Constant and Trend  
   $$\Delta Y_t = \delta Y_{t-1} + u_t$$

2. With Constant  
   $$\Delta Y_t = \alpha + \delta Y_{t-1} + u_t$$

3. With Constant and Trend  
   $$\Delta Y_t = \alpha + \beta T + \delta Y_{t-1} + u_t$$

The hypothesis is:  
$$H_0: \delta = 0 \text{ (Unit Root)}$$  
$$H_1: \delta < 0$$

Decision rule:
If \( t^* > \) Mackinnon critical value, \( ==> \) not reject null hypothesis, i.e., unit root exists.
If \( t^* < \) Mackinnon critical value, \( ==> \) reject null hypothesis, i.e., unit root does not exist.

The result of the ADF test indicated that the series were non-stationary when the variables were defined in levels. By first-differencing the series, in all cases, the null hypothesis of non-stationary was rejected at the 1 per cent significance level.

Co-integration Test

Once the order of integration of each variable is established, we then evaluate whether the variables under consideration are co-integrated. According to Engle and Granger (1987), a linear combination of two or more non-stationary series (with the same order of integration) may be stationary. If such a stationary linear combination exists, the series are considered to be co-integrated and long-run equilibrium relationships exist. Having recognized the existence of co-integration, although the series are individually non-stationary, they cannot drift farther away from each other arbitrarily. Co-integration implies that causality exists between the two variables, but it does not indicate the direction of the causal relationship. The presence of co-integration...
among the variables rules out the possibility of ‘spurious’ regression (Belloumi, 2010). There are various approaches to test for co-integration, say, Engle and Granger approach, Johansen approach, ARDL bounds testing approach (by Pesaran et al., 2001), and Gregory and Hansen approach.

According to Belloumi (2010), the bivariate approach of Engle and Granger is very restrictive because it can be applied only if there is one co-integrating relation. And the most commonly used method is the Johansen co-integration test based on the autoregressive representation discussed by Johansen (1988) and Johansen and Juselius (1990). This test determines the number of co-integrating equations for any normalization used. It provides two different likelihood ratio tests; one is based on the trace statistic and the other on the maximum eigenvalue.

Since the variables are integrated of order one, the test for co-integration was applied. The co-integration test, formulised by Engle and Granger (1987), was further improved by Johansen (1988). The test is given by the following equation:

$$\lambda_{\text{trace}} \left( \frac{r}{n} \right) = T \sum_{i=r+1}^{n} \log(1 - \lambda_i)$$

Where $r$ is the number of co-integrating relations, and $n$ is the number of variables. The null hypothesis is that the number of co-integrating vectors is less than or equal to $r$ against the alternative hypothesis of $r > 0$.

**Lag Length Determination**

The over parameterization and loss of degrees of freedom problems must be avoided to capture the important information in the system. The
appropriate lag length must be determined by allowing a different lag length for each equation at each time and choosing the model with the lowest AIC and SBC values. The same sample period must be considered for different lag lengths. If the lag length is too small, the model will be mis-specified; if it is too large, the degrees of freedom will be lost. It was decided to use test criteria such as the Akaike Information Criterion, the Schwartz Bayesian Criterion and Hannan and Quinn. The AIC, SBC and HQ, with respect to the different lag order, are defined as:

\[
\text{AIC} = -2 \log L/T + 2 n/T \\
\text{SBC} = -2 \log L/T + n \log (T)/T \\
\text{HQ} = -2 \log L/T + 2n \log (\log (T))/T
\]

Where \( L \) is the maximized likelihood, \( n \) is the number of parameters and \( T \) is the sample size: see Akaike (1985), Schwartz (1978), and Hannan and Quinn (1979); we will chose a model with the least AIC and SBC. The decision rule is to choose the lag order with the minimum value of the criteria.

The results of the VEC model have been discussed in Chapter 5.

c. **Forecasting Technique:** An attempt has been made in this study to project the demand for commercial energy mainly coal, oil, natural gas and electricity for 20 year period beginning from 2010 to 2030 based on economic growth. The economic growth rate has been taken as a reference for convenience to simplify calculation mechanism in developing the relationship between energy growth and economic growth. The projection is based on the assumption that the energy elasticities remain stable over the next 20 years and that the economy is expected to grow at a rate of 7.81 per cent which was calculated based on past data for 20 years beginning from the period 1991. Further, an attempt has also been made to show a more optimistic
economic growth rate scenario of 10 per cent for the next 20 years till 2030. The energy growth rate has been calculated by using the formula:

\[ \frac{\Delta E}{E} = \frac{\Delta Y}{Y} \left( \frac{\Delta Y}{\Delta E} \right) \cdot \frac{E}{Y} \]

\[ \frac{\Delta E}{E} = \frac{\Delta Y}{Y} \left( \frac{\Delta Y}{\Delta E} \right) \cdot \frac{E}{Y} \]

The results of the forecasting of commercial energy demand have been discussed in Chapter 6.